

# GFaI Tagungsband 2024

AI4EA Workshop (Berlin Workshop on Artificial Intelligence for Engineering Applications)

3D-iSA (3D in Science & Applications)

# INHALTSVERZEICHNIS

GFal-Tagungsband 2024

## VORWORT

- 2 Vorwort der Workshopleitung

## PUBLIKATIONEN 3D-iSA

- 6 Reconstruct-HM3D: Providing CAD models as ground truth for the Habitat-Matterport 3D Research Dataset
- 13 3D integration of multispectral Image Data in forensic trace Recording
- 19 Automated Re-Photography: Localization of Historical Photographs for Urban Environments using Structure-Feature Based Camera Pose Estimation
- 26 Modular Target and Approach for Geometric Calibration of Multimodal Imaging Systems
- 31 Depth Completion by Colorization for Solid-State LiDAR Sensors
- 40 Digitization of the Measurement process of Staircases and Landings
- 45 Lidar-based Missing Plant Detection in Steep Vineyards
- 49 Alpha Shape For 3D Point Cloud Analysis
- 53 Automation of painting Processes for small batch Sizes: Development of a multi-camera System for real-time recording of manual spray Paths
- 58 Additively Manufactured Forging Dies with Contour-Optimized Cooling
- 66 Towards a robust automated surface inspection Method for CT-scanned Cannulas
- 71 Deflection Shape Projection for calculating similarity Measures
- 75 Online Optimization of Stereo Camera Calibration Accuracy

## PUBLIKATIONEN AI4EA

- 82 Testing ML-based Systems Using Probabilistically Extended Ontologies
- 89 Finding Predictive Features for Energy Consumption of CNC Machines
- 96 Feedback-driven Object Detection and iterative Model Improvement for accurate Annotations
- 102 Enhancing Investment Casting Quality Control: AI-Driven Defect Detection on Ceramic Shells in Investment Casting
- 111 Humanity-Centered Production – a Spotlight on Future Productivity
- 115 Optimized test case selection in a Continuous Integration process using Reinforcement Learning
- 121 Low-Cost Smart Sensor IIoT Platform for Real-Time Monitoring and Predictive Maintenance
- 130 AI-based Real-Time Analysis of Sawing Processes for the Identification of Materials
- 138 An Object-Based Approach for Classifying Flake Graphite Cast Iron Micrographs through Semantic Segmentation

## 3D-iSA

Mit großer Freude blicken wir auf die 3D in Science & Applications (3D-iSA) 2024 zurück, die am 26. und 27. November in Kombination mit der AI4EA stattfand. Als 26. Veranstaltung der Workshop-Reihe der GFal setzte die 3D-iSA die langjährige Tradition des interdisziplinären Austauschs fort und präsentierte sich dabei in einem neuen Gewand. Der neue Titel „3D in Science & Applications“ spiegelt den Anspruch wider, aktuelle wissenschaftliche Entwicklungen und anwendungsorientierte Lösungen in den Bereichen der 3D-Technologien miteinander zu verknüpfen.

Mit 13 Vorträgen in fünf thematischen Sessions bot die 3D-iSA 2024 einen umfassenden Überblick über die Vielfalt und Relevanz moderner 3D-Technologien. Rund 50 Teilnehmende aus Wissenschaft, Industrie und Praxis kamen zusammen, um sich über innovative Ansätze, methodische Fortschritte und praxisnahe Anwendungen auszutauschen. Die Beiträge spannten dabei einen weiten thematischen Bogen – von der 3D-Datenerfassung und Rekonstruktion über die Kalibrierung hochpräziser Systeme bis hin zur additiven Fertigung, Datenanalyse und Qualitätssicherung.

Ein zentrales Anliegen der 3D-iSA war es, die interdisziplinäre Vernetzung zu fördern. Insbesondere die Verbindung mit der AI4EA ermöglichte es, Synergien zwischen den Bereichen 3D-Technologien und Künstliche Intelligenz zu schaffen. Der intensive Austausch während der Vorträge und Diskussionen zeigte, wie eng diese Felder miteinander verbunden sind und welche Potenziale für die Zukunft in ihrer Zusammenarbeit liegen.

Unser besonderer Dank gilt allen, die diese Veranstaltung möglich gemacht haben: den Vortragenden für ihre spannenden und inspirierenden Beiträge, dem Programmkomitee und Organisationsteam für die engagierte Planung sowie allen Teilnehmenden, die durch ihre Fragen, Diskussionen und Anregungen den Workshop zu einem lebendigen Ort des Austauschs gemacht haben. Dank Ihnen konnte die 3D-iSA 2024 erneut zeigen, dass 3D-Technologien weit über ihre traditionellen Anwendungsfelder hinausreichen und ein wesentlicher Treiber für Innovation und Fortschritt sind.

Januar 2025

Felicitas Böhm



### Programmkomitee

Prof. Alfred Iwainsky (AI4Tech / GFal e.V.)  
Felicitas Böhm (GFal e.V.)  
Benjamin Hohnhäuser (GFal e.V.)  
Daniel Krueger (GFal e.V.)  
Prof. Frank Neumann (Hochschule für Technik und Wirtschaft Berlin)  
Prof. Holger Schlingloff (Humboldt-Universität zu Berlin)



## AI4EA

Künstliche Intelligenz (KI) ist heute ein zentraler Treiber für Innovationen in Wissenschaft, Industrie und Technik. Sie durchdringt nahezu alle Bereiche moderner Technologien und schafft Lösungen, die noch vor wenigen Jahren undenkbar erschienen. Von der Produktentwicklung über die Produktion bis hin zur Wartung und Zustandsüberwachung revolutioniert KI Prozesse und schafft neue Möglichkeiten für Effizienz und Präzision.

Der Berlin Workshop on Artificial Intelligence for Engineering Applications 2024 (AI4EA24), welcher am 27. und 28. November 2024 stattfand, greift genau diese Themen auf und präsentiert KI-basierte Lösungen für industrielle Herausforderungen. In diesem Jahr fand der Workshop bereits zum dritten Mal statt, und wir freuen uns sehr über das stetig wachsende Interesse an dieser Veranstaltung. AI4EA hat sich als eine bedeutende Plattform für den Wissenstransfer zwischen Forschung und Industrie etabliert. Der Workshop legt bewusst den Fokus auf den Austausch zwischen diesen beiden Welten und bietet Wissenschaftlern sowie Industrieexperten die Möglichkeit, ihre Ideen, Forschungsergebnisse und praktischen Herausforderungen zu diskutieren und zu veröffentlichen.

Im Jahr 2024 standen wieder die Anwendungsbereiche Qualitätssicherung in der Produktion, Predictive Maintenance, Prozesssteuerung und -optimierung sowie Produktentwicklung im Mittelpunkt des Workshops. Diese Themen spiegeln die aktuellen Bedürfnisse und Entwicklungen in der Industrie wider und zeigen, wie KI-basierte Ansätze reale Probleme lösen können.

Insgesamt gingen 13 Einreichungen ein, von denen alle für eine Präsentation im Rahmen des Workshops akzeptiert wurden. Da die Autoren die Möglichkeit hatten, ihre Beiträge alternativ als Poster zu präsentieren, und nicht alle eingereichten Arbeiten zur Veröffentlichung zugelassen wurden, umfasst dieser Tagungsband fünf ausgewählte Beiträge. Jeder zur Veröffentlichung eingereichte Beitrag durchlief ein Begutachtungsprozess im Peer Review-Verfahren mit mindestens zwei, in den meisten Fällen jedoch drei oder vier anonymen Reviews. Dies gewährleistet die hohe Qualität der in diesem Band präsentierten Arbeiten.

Unser besonderer Dank gilt dem Programmkomitee für die engagierte und sorgfältige Begutachtung der Einreichungen. Darüber hinaus möchten wir uns herzlich bei Frau Beuster vom GRW-Netzwerk NET4AI und beim Netzwerkmanagement vom ZIM-Innovationsnetzwerk AI4Tech bedanken, die als Mitveranstalter des Workshops maßgeblich zur Moderation und Unterstützung beigetragen haben. Nicht zuletzt gilt unser Dank den Autoren, die mit ihren Beiträgen und Präsentationen die Veranstaltung bereichern haben.

Wir hoffen, dass dieser Tagungsband nicht nur die vielfältigen Anwendungsmöglichkeiten von KI in der Industrie aufzeigt, sondern auch als Inspiration für zukünftige Forschung und Zusammenarbeit dient.

Januar 2025

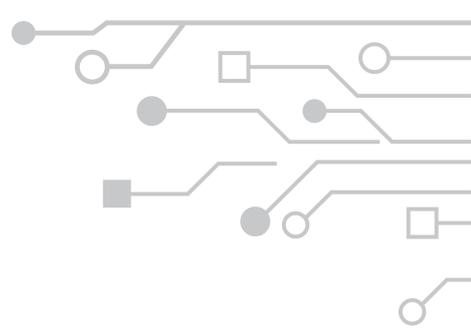
Benjamin Hohnhäuser

### Programmkomitee

Prof. Alfred Iwainky (GFal e.V. / AI4Tech)  
 Prof. Benjamin Noack (Otto-von-Guericke-Universität  
 Magdeburg)  
 Prof. Erik Rodner (Hochschule für Technik und Wirtschaft Berlin)  
 Prof. Gerd Stanke  
 Prof. Holger Schlingloff (GFal e.V.)  
 Prof. Jörg Reiff-Stephan (Technische Hochschule Wildau)  
 Prof. Wolfgang Halang (FernUniversität in Hagen)  
 Mario Koddenbrock (Hochschule für Technik und Wirtschaft Berlin)

Gregor Wrobel (GFal e.V.)  
 Dr. Christof Puhle (GFal e.V.)  
 Shaunak Atul Kanikar (GFal e.V.)  
 Immanuel Rettig (GFal e.V.)  
 Benny Botsch (GFal e.V.)  
 Miriam Schneider (GFal e.V.)  
 Benjamin Hohnhäuser (GFal e.V. / AI4Tech)  
 Hannes Diethelm (GFal e.V.)  
 Dr. Daniel Herfert (GFal e.V.)





**Workshop 3D-iSA (3D in Science & Applications)**



26. Anwendungsbezogener Workshop zur  
Erfassung, Modellierung, Verarbeitung und Auswertung von 3D-Daten

Berlin  
26.11./ 27.11.2024



# Reconstruct-HM3D: Providing CAD models as ground truth for the Habitat-Matterport 3D Research Dataset

Marian Bookhahn<sup>1</sup>, Tara Lorenz<sup>2</sup>, Felicitas Böhm<sup>2</sup>, Frank Neumann<sup>1</sup>

<sup>1</sup>HTW – Hochschule für Technik und Wirtschaft Berlin, Fachbereich 2: Ingenieurwissenschaften - Technik und Leben, Studiengang Ingenieurinformatik, Wilhelminenhofstr. 75a, 12459 Berlin

<sup>2</sup>GFal e.V., Volmerstraße 3, 12489 Berlin

eMail: frank.neumann@htw-berlin.de

**Abstract:** The research project ServiceTechNavigator aims at supporting service technicians, who need to reach an installation site on a vast factory premise. The project adopts a novel approach in indoor navigation employing the existing 3D Building Information Models (BIM) as an absolute position reference for this purpose. Such models are widely adopted and utilized in recent years for a variety of tasks in construction planning and facility management. In 3D BIM data, real geometries are represented by parametric models, like lines or planes, which are enriched by different kinds of semantic categories, like walls, windows, doors, etc. Compared to huge memory demanding 3D scans, such condensed representations are easier to work with. To enable research in the area of 3D BIM data reconstruction and BIM based indoor navigation, large datasets with real sensor data along with parametric representations are needed.

Thus, we present Reconstruct-HM3D, which is an extension to the Habitat-Matterport 3D Research Dataset (HM3D) and provides semantically enriched parametric models for the real scan data. Considering HM3D is the largest-ever dataset of real 3D indoor spaces, our ground truth CAD models enable research in a variety of areas, such as Room Plan Estimation (RPE) (cf. RoomPlan Swift API) and BIM based indoor navigation. We reconstructed all indoor spaces of HM3D programmatically and revised them manually. Ongoing evaluations of the Reconstruct-HM3D dataset show its high benefits for the training and evaluation of the trajectory reconstruction, various types of features, segmentation, monocular depth estimation, place recognition and vendor neutral extraction of room plans.

**Keywords:** BIM, HM3D, Reconstruct-HM3D, trajectory reconstruction, place recognition, room plan estimation

## 1 INTRODUCTION

If service technicians from a service company are commissioned to replace a spare part or to carry out maintenance for production, transport (e.g. escalators, lifts) or technical building systems (e.g. heating, ventilation, sanitation), they must be enabled by the client to reach the installation site of the respective systems. In addition to the exact address and room number, they also need the route to the place of performance. Particularly in the case of large companies with a central entrance and extensive building and area structures (e.g. manufacturing industry, hospitals, logistics), the challenges are great, so that valuable time is often lost until the service provider finds their way around an unfamiliar site and arrives at the installation location. Due to the high frequency of such appointments, the client's specialist departments generally do not have sufficient personnel capacity to accompany the service providers and guide them to the site. The same applies to security and security guards, who are tied to a specific location and are not allowed to leave their work area.

### 1.1 OBJECTIVES OF THE SERVICETECHNAVIGATOR RESEARCH PROJECT

The ServiceTechNavigator R&D project aims to navigate a service technician within an industrial property to and from their work location using an existing BIM model. For

this purpose, he will be given a tablet at reception, on which the ServiceTechNavigator software runs and which he must carry with him during his entire assignment. The app will display all the important information that will guide him on his way to the site, during his work and back to reception (cf. Fig. 1). To enable the service technician to navigate, his trajectory within the site must be determined during his entire assignment.

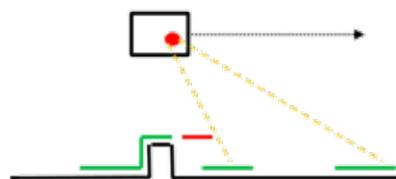


**Table 1.** The ServiceTechNavigator app guides service technicians on the way to their work location

## 1.2 NAVIGATION APPROACH OF THE SERVICETECHNAVIGATOR RESEARCH PROJECT

The ServiceTechNavigator app adopts a novel approach in indoor navigation employing the existing 3D Building Information Models (BIM) as an absolute position reference and uses its geometric features for orientation. Unlike today's radio-based solutions (active location determination via trigonometry), the project idea is based on a passive approach and therefore does not require any intervention in the infrastructure or the equipping of radio anchors or beacons. The optical navigation approach is to be implemented in an app environment that utilises the internal sensors of modern mobile devices and can draw on existing CAD/BIM models (reference points from the model and optical detection) for orientation. Another advantage is the use of enriched information on the walking surfaces and features, so that the envisaged solution also supports barrier-free use by recognising obstacles and elevations.

Besides the use of the BIM model, the featured approach calculates the base trajectory from the IMU integration, where the lateral drift will be minimised by comparing detected areas with a wall map extracted from the BIM model (cf. Fig 2).



**Table 2.** Wall map comparison: correctly matched wall segments are depicted in green in contrast to the red segment contained in the wall map but missing at the location

In addition, the global position will be synchronised through machine-learned feature for distinct elements (e.g. pillars, windows, doors). Fig. 3 depicts the matching of these recognised features with locations in the BIM model. The left image shows the semantic segmentation of the current scene, whereas the right side depicts the association between such features with elements in the BIM model providing an absolute position.



**Table 3.** Left: semantic segmentation of the current scene, Right: association between features and elements in the BIM model providing an absolute position

## 1.3 RESEARCH OBJECTIVES FOR RECONSTRUCT-HM3D

The previously laid-out indoor navigation approach requires extensive training and evaluation data for the trajectory reconstruction, various types of features, segmentation, place recognition and the extraction of room plans. For this purpose, large datasets with real sensor data along with parametric representations are needed.

In particular, we are interested in scans of buildings interiors with sensors whose noise characteristic is close to the sensors used in our settings. This includes depth, optical, as well as IMU sensors.

Furthermore, we need to have a BIM plan available,

against which we can match. On the other hand, there should be a lot of (geometric) obstacles placed following a realistic distribution.

By keeping the essence of the original HM3D dataset, our dataset should permit the generation of arbitrary passages through the dataset's scenes resulting in an unlimited number of camera viewpoints and trajectories.

## 2 RELATED WORK

In the following, the research literature will be reviewed regarding datasets that feature the interior of buildings. Such datasets are typically categorised into synthetic, 3D reconstruction-based, floorplan-based and panorama-based datasets [1]. Floorplan-based and panorama-based datasets do not contain realistic geometric models, which are required for the employed approach for indoor navigation. Therefore, we will focus only on synthetic and 3D reconstruction-based datasets in the following.

### 2.1 SYNTHETIC 3D SCENE DATASETS

The SUNCG dataset [2] contains synthetic data featuring millions of 3D indoor scenes that comprise annotated layouts and object labels. Due to its synthetic character, SUNCG contains unrealistic textures and fails to capture noise as well as irregularities and, consequently, can affect the transferability of trained models to real-world scenarios.

In 2023, Meta Research published the Aria Synthetic Environments dataset [3] that contains 100,000 synthetically generated indoor scenes spanning over multiple rooms considering the device characteristics of the Aria glasses. It aims at the training and evaluation of machine learning algorithms for scene understanding and embodied AI tasks and offers comprehensive ground truth data for the camera trajectory, the floorplans, the instance segmentation and the depth maps. Nevertheless, the synthetic nature of the data, in particular the textures, and the rather small scene size make it unsuitable for the task of indoor navigation.

### 2.2 3D RECONSTRUCTION DATASETS

The Habitat-Matterport 3D (HM3D) Research Dataset is a large-scale dataset containing 1,000 high-resolution 3D scans generated from commercial, residential and civic spaces [1]. It is the largest dataset in this category and comprises a high number of scenes and floors. Each scene in the dataset is represented with photorealistic 3D meshes (cf. Fig. 4 and 5), allowing for realistic

interaction and perception tasks in machine learning, robotics and computer vision. The dataset is compatible with simulators like Habitat, enabling access to RGB-D images, IMU data, semantic segmentation maps, and agentcentric views. It has distinct advantages over other datasets like Gibson [4], Replica [5], RoboTHOR [6], MP3D [7] and ScanNet [8], which is its high visual fidelity, when real images are being compared to images rendered from synthetic viewpoints and its low incidence of reconstruction artifacts like missing surfaces, holes or untextured surface regions [1], enabling the large-scale generation of sensor data suitable for monocular depth estimation and other image based machine learning techniques.



**Table 4.** Real vs. rendered images from the HM3D dataset [1]



**Table 5.** Example scenes from HM3D dataset [1].

### 3 PROBLEM ANALYSIS

A dataset suitable for our approach of indoor navigation should comprise sensor data of indoor scenes from an IMU, a camera and ideally a depth sensor, as well as the corresponding floorplans in addition. As gathering a statistically representative dataset is a rather daunting task, it seems wise to use existing datasets. To the best of our knowledge, no such dataset exists. While the HM3D [1] dataset can be regarded as statistically representative, it has the following shortcomings with regards to our investigations:

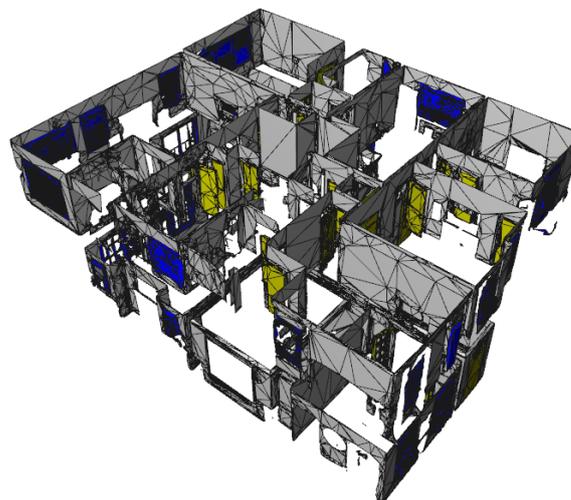
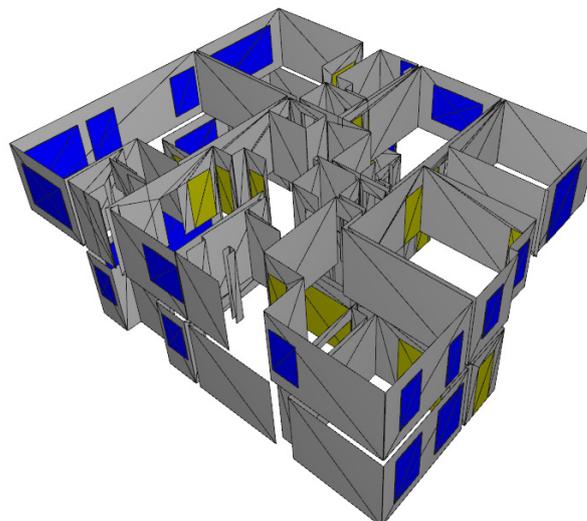
- There are no BIM data or floorplan available.
- The employed depth sensors are not comparable to the relevant sensors in our settings. Their noise characteristic might be very different from ours.
- The dataset contains no notion of trajectories. All measurements were taken statically from locations chosen with the mean to have reconstruction. It lacks IMU data altogether.

However, it still seems worthwhile to consider the HM3D dataset as a source of a representative sample of geometric object distributions. In order to make use of the dataset, we plan to undertake the following actions:

- **Trajectories:** It should make no big difference whether a trajectory was simulated or recorded from human motion. Human trajectories typically comprise rather low frequencies. Thus, they can simply be described by handdrawn or programmatically generated spline curves.
- **IMU:** The typical noise characteristics of IMU data is generally known and relatively independent of the surroundings [9] [10].
- **Depth Measurements:** Instead of relying on dedicated depth sensors, we make use of HM3D's high visual fidelity and employ monocular depth estimation. Thereby, we minimize the domain gap between synthetic and real sensor data.

### 4 PROPOSED SOLUTION

We choose to use a staged, semi-automatic process based on heuristics to reconstruct CAD-like parametric models from the original 3D data of the HM3D dataset. Unlike raw 3D scans, which are prone to noise and lack structured geometric representations, our approach produces clean, structured models (cf. Fig. 6) by fitting parametric models like planes, detecting wall openings, and validating structural features.



**Table 6.** Above: Reconstructed model visualized as a triangle mesh | Below: Raw mesh from HM3D Gray: Walls | Green: Doors | Blue: Windows

The staged process can be summarized by the steps depicted in Fig. 7. Due to the nature of the original HM3D data with sensor and process noise as well as inaccuracies of human annotations, the automatic reconstruction process needs a vast amount of specifically tweaked and tuned heuristics, which details cannot be depicted entirely in this work. Therefore, we pick a few examples to illustrate which strategies have been chosen.

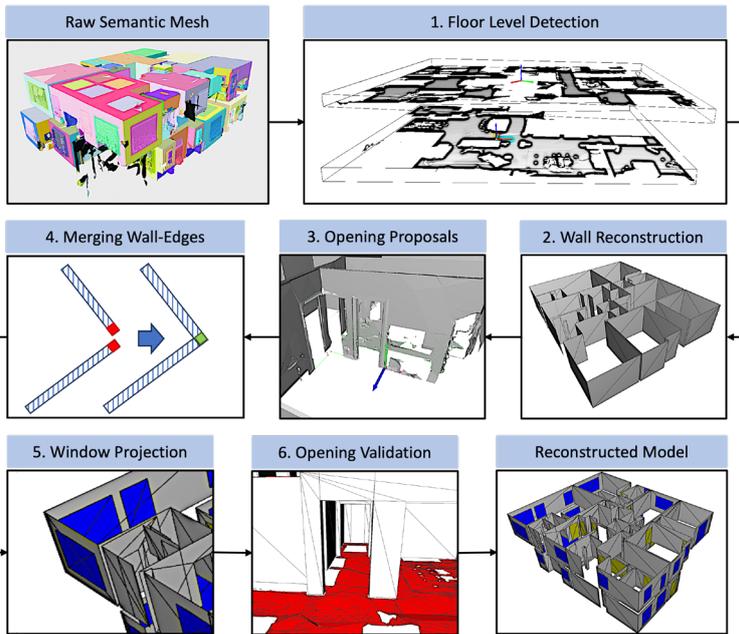


Fig. 7. Staged reconstruction process

## 5 EVALUATIONS AND DISCUSSION

### 5.1 ANNOTATION

The original dataset was manually annotated by assigning instance-labels to the triangles of the reconstructed polygon-meshes. This process is prone to inconsistencies and noise, like multiple, in reality separated, mesh-segments having the same instance ID (cf. Fig. 8), as well as small mesh-fragments, which should be regarded as outliers (cf. Fig. 9). The initial wall-reconstruction process is mainly based on RANSAC plane-fitting, followed by density-based clustering, to find aligned but separated wall-segments. During this process we do consider metrics like the mean angular deviation between triangle-normals and between plane and floor-normals for our heuristics.

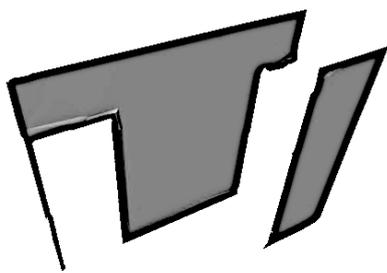


Table 8. Separate wall-segments with the same instance ID



Table 9. Small outlier mesh-segment

### 5.2 WALL-OPENINGS AND DOORS

Considering all types of wall-openings, only openings being passable for humans are typically captured by floorplans and thus, are relevant to indoor navigation tasks. Reconstructing wall-openings from the given 3D scans can be difficult, regarding wall-openings can result from different types of occlusions like furniture and open doors. (cf. Fig. 10).

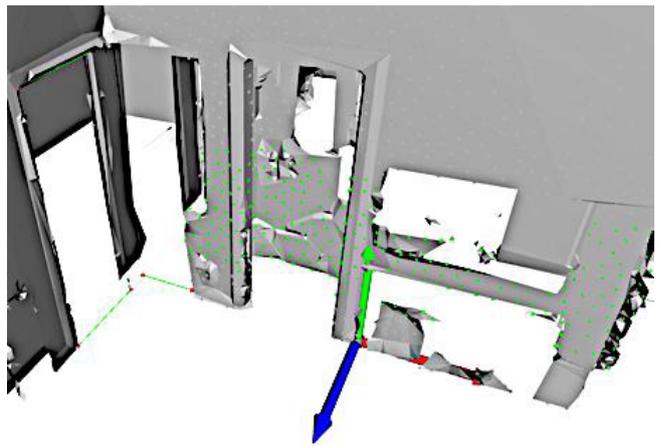


Table 10. Wall-segment with different wall-openings. Green lines with red dots indicate detected opening-proposals.

In the given example in Fig. 10, we sample points at a certain height along the wall-mesh (Poisson-Disk-Sampling) and calculate density-based clusters to extract proposals for wall-openings, which are going to be validated in a following stage.

### 5.3 MERGING WALL-EDGES

After reconstructing the wall-segments, gaps and intersections, resulting from noisy data, need to be fixed (cf. Fig. 11). Consequently, we calculate possible intersection points between walls and apply a cost-function to match corresponding edges. Afterwards, wall-dimensions are being corrected accordingly (cf. Fig. 12).

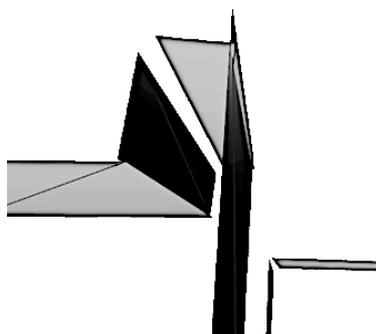
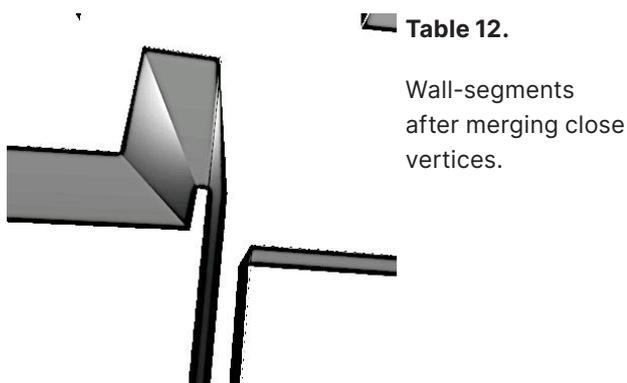


Table 11.

Wall-segments before merging close vertices.

**Table 12.**

Wall-segments after merging close vertices.

#### 5.4 VALIDATION OF WALL-OPENINGS AND DOORS

To validate or invalidate the previously introduced proposal for wall-openings, different tests need to be performed. An opening is considered to be valid if one of the following criteria is true:

1. Parts of the floor mesh passes through the opening without interruptions.
2. A door-mesh is closely aligned with the wall-opening (closed door).
3. Two opening-proposals with close proximity are parallel and have similar dimensions.

#### 5.5 MANUAL REVISION

The manual revision process is programmatically guided, such that detected inconsistencies, like unmatched doors and windows, intersecting walls as well as freestanding walls without any merged edges are highlighted to the revisor. Due to our highly effective programmatic reconstruction pipeline, we were able to manually revise all models within less than 200 working hours. Considering the corrections needed during revision, we achieved a very low false-positive rate of 3% and false-negative rate of less than 1% for wall-openings.

#### 6 CONCLUSIONS AND OUTLOOK

We provide semantically annotated triangle-meshes in PLY format as well as mapping files, to assign the instance colours of our models to the original annotation colours. This step is necessary, since we split the original instances into distinct segments, according to the parametric models used. Additionally, we built an exporter to convert our data structure into GeoJSON format for 2D floorplans and CityJSON format [11] for a parametric 3D representation, which ensures a lightweight encoding suitable for indoor navigation.

The proposed solution effectively reconstructs raw 3D scans from the HM3D dataset into structured, CAD-like parametric models. The semi-automatic methodology ensures accuracy and scalability. Based on our enhanced dataset, we are able to overcome the issues with the original HM3D data outlined in chapter 3.

With our complementing work to the HM3D dataset, we enhance its value to many research areas like indoor-navigation, 3D feature extraction, place recognition as well as scene understanding and reconstruction. Ongoing evaluations of our dataset within our approaches to indoor-navigation show its high benefits for feature extraction and trajectory reconstruction. We think it can be worthwhile to apply our reconstruction pipeline to other datasets like ARKitScenes [12].

#### ACKNOWLEDGEMENTS

This research has been funded by the Federal Ministry for Economic Affairs and Climate Action (BMWK) of Germany in the Central Innovation Programme for small and medium-sized enterprises (German Zentrales Innovationsprogramm Mittelstand) (project number 16KN117720).

#### REFERENCES

- [1] S. K. Ramakrishnan, A. Gokaslan, E. Wijmans, O. Maksymets, A. Clegg, J. M. Turner, E. Undersander, W. Galuba, A. Westbury, A. X. Chang, M. Savva, Y. Zhao and D. Batra, "Habitat-Matterport 3D Dataset (HM3D): 1000 Large-scale 3D Environments for Embodied AI," in Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track, 2021.
- [2] S. Song, F. Yu, A. Zeng, A. X. Chang, M. Savva and T. Funkhouser, "Semantic Scene Completion from a Single Depth Image," Proceedings of 30th IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [3] J. Engel, K. Somasundaram, M. Goesele, A. Sun, A. Gamino, A. Turner, A. Talattof, A. Yuan, B. Souti, B. Meredith, C. Peng, C. Sweeney, C. Wilson, D. Barnes, D. DeTone, D. Caruso, D. Valleroy, D. Ginpall, D. Frost, E. Miller, E. Mueggler, E. Oleinik, F. Zhang, G. Somasundaram, G. Solaira, H. Lanaras, H. Howard-Jenkins, H. Tang, H. J. Kim, J. Rivera, J. Luo, J. Dong, J. Straub, K. Bailey, K. Eickenhoff, L. Ma, L. Pesqueira, M. Schwesinger, M. Monge, N. Yang, N. Charron, N. Raina, O. Parkhi, P. Borschowa, P. Moulon, P. Gupta, R. Mur-Artal, R. Pennington, S. Kulkarni, S. Miglani, S. Gondi, S. Solanki, S. Diener, S. Cheng, S. Green, S. Saarinen, S. Patra, T. Mourikis, T. Whelan, T. Singh, V. Balntas, V. Baiyya, W. Dreewes, X. Pan, Y. Lou, Y. Zhao, Y. Mansour, Y. Zou, Z. Lv, Z. Wang, M. Yan, C. Ren, R. D. Nardi and R. Newcombe, Project Aria: A New Tool for Egocentric Multi-Modal AI Research, 2023.
- [4] F. Xia, A. R. Zamir, Z. He, A. Sax, J. Malik and S. Savarese, "Gibson env: Real-world perception for embodied agents," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

- [5] J. Straub, T. Whelan, L. Ma, Y. Chen, E. Wijmans, S. Green, J. J. Engel, R. Mur-Artal, C. Ren, S. Verma, A. Clarkson, M. Yan, B. Budge, Y. Yan, X. Pan, J. Yon, Y. Zou, K. Leon, N. Carter, J. Briales, T. Gillingham, E. Mueggler, L. Pesqueira, M. Savva, D. Batra, H. M. Strasdat, R. De Nardi, M. Goesele, S. Lovegrove and R. Newcombe, "The Replica Dataset: A Digital Replica of Indoor Spaces," June 2019.,
- [6] M. Deitke, W. Han, A. Herrasti, A. Kembhavi, E. Kolve, R. Mottaghi, J. Salvador, D. Schwenk, E. Vanderbilt, M. Wallingford, L. Weihs, M. Yatskar and A. Farhadi, "RoboTHOR: An Open Simulation-to-Real Embodied AI Platform," April 2020.
- [7] A. Chang, A. Dai, T. Funkhouser, M. Halber, M. Nießner, M. Savva, S. Song, A. Zeng and Y. Zhang, "Matterport3D: Learning from RGB-D Data in Indoor Environments," September 2017.
- [8] A. Dai, A. X. Chang, M. Savva, M. Halber, T. Funkhouser and M. Nießner, "ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes," February 2017.
- [9] A. D. Young, M. J. Ling and D. K. Arvind, "IMUSim: A simulation environment for inertial sensing algorithm design and evaluation," in Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks, 2011.
- [10] M. Sharifi Renani, A. M. Eustace, C. A. Myers and C. W. Clary, "The Use of Synthetic IMU Signals in the Training of Deep Learning Models Significantly Improves the Accuracy of Joint Kinematic Predictions," Sensors, vol. 21, p. 5876, August 2021.
- [11] H. Ledoux, K. Arroyo Otori, K. Kumar, B. Dukai, A. Labetski and S. Vitalis, "CityJSON: a compact and easy-to-use encoding of the CityGML data model," Open Geospatial Data, Software and Standards, vol. 4, June 2019.
- [12] G. Baruch, Z. Chen, A. Dehghan, T. Dimry, Y. Feigin, P. Fu, T. Gebauer, B. Joffe, D. Kurz, A. Schwartz and others, "ArkitScenes: A diverse real-world dataset for 3d indoor scene understanding using mobile rgb-d data," arXiv preprint arXiv:2111.08897, 2021.

# 3D-Integration multispektraler Bilddaten bei der kriminaltechnischen Spurenaufnahme

Stefan Holtzhausen<sup>1</sup>, Arthur Hilbig<sup>1</sup>, Andreas Jalowi<sup>2</sup>,  
Andreas Engel<sup>3</sup>, Robert Pfannenberger<sup>2</sup> und Kristin Paetzold-Byhain<sup>1</sup>

<sup>1</sup> TU Dresden, Professur für Virtuelle Produktentwicklung

<sup>2</sup> Landeskriminalamt Sachsen, Tatortgruppe

<sup>3</sup> Institut für Rechtsmedizin, Medizinische Fakultät, TU Dresden

**Zusammenfassung:** In der heutigen Kriminaltechnik sind 2D- und 3DAufnahmeverfahren wichtige Werkzeuge zur Spurenaufnahme und -dokumentation. Insbesondere die forensische Fotografie nutzt unterschiedliche Wellenspektren zur Sichtbarmachung von Spuren, die mit bloßem Auge im sichtbaren Wellenlängen nicht erkennbar sind. Mithilfe eines dedizierten optischen Aufnahmesystems (Sensor, Objektiv, Filter, Strahlenquelle) werden Aufnahmen mit unterschiedliche Spektren, wie im Infrarot- (IR) und Ultraviolettbereich (UV), angefertigt, um latente Spuren, etwa Fingerabdrücke oder biologische Substanzen, sichtbar zu machen. Ein besonderer Anwendungsbereich ist die Visualisierung von Hämatomen und subkutanen Einblutungen an Leichen, die äußerlich oft unsichtbar sind. Im Bereich der 3D-Dokumentation werden hochpräzise Aufnahmesysteme, von Mikroskopen bis zu terrestrischen Laserscannern, bereits seit einiger Zeit eingesetzt. Jedoch ist zurzeit eine direkte, messtechnische Auswertung von forensischen 2D-Aufnahmen bisher nicht möglich. Im Beitrag wird ein Ansatz vorgestellt, bei dem ohne zusätzliche Hardwareanforderungen an das Aufnahmesystem, forensische 2D-Aufnahmen und 3D-Scans mittels photogrammetrischer Referenzkörpern aufeinander referenziert werden. Ein nachgelagerter Texturierungsprozess ermöglicht eine digitale Analyse der 3D-Daten. Gestützt wird die Methode durch eine Untersuchung zur Abbildungsgenauigkeit, um die Aussagefähigkeit der erzeugten, texturierten 3D Daten zu stützen.

**Schlüsselwörter:** Photogrammetrie · Forensische Fotografie · Texturierung · 3D-Auswertung

## 1 Stand der Technik und Motivation

### 1.1 Motivation

Das Erkennen charakteristischer Hautmerkmale von Geschädigten durch die Analyse von Einblutungsmustern und Hämatomen ist für die Aufklärung schwerer Straftaten und die Identifizierung von Tätern entscheidend. Hierfür ist es unerlässlich, die Verletzungen maßstabsgetreu zu dokumentieren, um Rückschlüsse auf Art und Intensität der Gewalteinwirkung zu ermöglichen und potenzielle Verursacher sowie eingesetzte Objekte zu identifizieren oder einzugrenzen.

Auch zur Identifizierung eines potenziellen Täters bei Verdacht auf Misshandlung, etwa durch Bissverletzungen (siehe Abbildung 1, Links), ist eine Dokumentation und vollständige Beschreibung der Verletzungen essenziell, um Merkmale wie Zahnlücken oder Unregelmäßigkeiten im Gebiss darzustellen.



**Abb. 1:** Links: Bissspuren auf der Haut eines Kindes, Rechts: Schuhspuren, welche sich in Form von Einblutungen absetzen (Quelle: RM Dresden)

In Abbildung 1 Rechts sind beispielsweise Verletzungen abgebildet, die vermutlich durch Tritte entstanden sind. Trotz erkennbarer Strukturen ist die genaue Bestimmung des Musters in zweidimensionalen Bildern aufgrund von Verzerrungen und gewölbter Haut schwierig und oft unzureichend für die Zuordnung eines Tatmittels.

Würde hier eine 3D-Erfassung der Daten durchgeführt, könnte die Ausdehnung der Verletzungsmuster auf der Hautoberfläche oder in tieferliegenden Strukturen jederzeit reproduzierbar und maßstabsgetreu bestimmt werden, insbesondere durch die Kombination von multispektraler Fotografie. Ziel ist das Aufstellen eines solchen Prozesses sowie die Fusion der dabei aufgenommenen Daten. Ferner muss eine Aussage zur Methodik und der damit erreichbaren Genauigkeit getroffen werden.

## 1.2 STAND DER WISSENSCHAFT

Als bewährte Methode der Spurensicherung gilt die forensische Fotografie, die mit Standard-, IR- und UV-Fotografie durchgeführt wird [5,4,11]. Spezielle Lichtquellen machen Blutungen unter der Haut sichtbar, die bei normalem Licht unsichtbar bleiben. Das Hautkolorit und Melanin beeinflussen die Lichtabsorption, die je nach Wellenlänge variiert und so eine tiefere Durchdringung der Haut ermöglicht [9]. Die Absorption der Haut ist wellenlängenabhängig, wobei das Maximum um 500nm liegt und damit dem Absorptionsmaxima von Melanin und Hämoglobin (Blutfarbstoff) entspricht. Bei längeren Wellenlängen (700 bis 1300 nm) reduziert sich die Streuung in der Haut. Damit wird hier die größte Eindringtiefe in die Haut erreicht [9]. Hämatome entstehen durch stumpfe Gewalt, welche die Kapillaren unter der Haut verletzt und Blutungen verursacht [4]. Die äußere Sichtbarkeit hängt von Faktoren wie der Menge des Hämoglobins und dem Abheilungsstadium ab. Forensische Lichtquellen können Hämatome klarer von der unverletzten Haut abgrenzen und ermöglichen so eine präzisere Dokumentation (Abb. 2) [5,4,12,18].



**Abb. 2:** Sichtbare Hämatome bei Verwendung von forensischer Fotografie [12]

Neben der 2D-Verletzungsdokumentation ist die 3D-Erfassung ein wichtiger Ermittlungsschritt. Dabei kommen zunehmend Verfahren wie Laserscanning, Streifenprojektionsverfahren oder photogrammetrische Methoden zum Einsatz [3,2,17]. Auch die Computertomographie (CT) wird zunehmend als postmortale CT in der forensischen Medizin eingesetzt [15]. Sie ergänzt Obduktionen durch 3D-Darstellungen innerer Bereiche, die Verletzungsmuster auch für medizinische Laien verständlicher machen können. Oberflächliche Blutungen sind jedoch aufgrund der CT-Auflösung schwer zu erkennen.

Die Kombination von 2D-Bilddaten und 3D-Geometrien zur gemeinsamen geometrischen und farblichen Beschreibung in einer virtuellen Szene spielt in computergrafischen Anwendungen seit Jahren eine wichtige Rolle [1]. Hierfür werden Verfahren der Photogrammetrie und Computergrafik genutzt. Um ein 3D-Oberflächenmodell (Objektraum) mit der Textur eines 2D-Fotos (Bildraum) zu versehen, muss ein Objektpunkt vom Objektkoordinatensystem in das Kamerakoordinatensystem (extrinsische Orientierung, EOR) und anschließend in das 2D-Koordinatensystem des Bildes (intrinsische Orientierung, IOR) transformiert werden. Klassische Weitwinkel- und Teleobjektive verwenden ein zentralperspektivisches Abbildungsmodell, das den Einsatz photogrammetrischer Methoden auch auf herkömmlichen Smartphone-Kameras ermöglicht [8]. In diesem Modell wird die innere Orientierung durch eine vorherige Kalibrierung bestimmt, bei der mithilfe einer Bündelblockausgleichung verschiedene Verzeichnungsparameter [6], der Bildhauptpunkt der Kamera und die Brennweite berechnet werden [14]. Die äußere Orientierung kann anschließend durch einen iterativen Lösungsansatz der Kollinearitätsgleichung im sogenannten räumlichen Rückwärtsschnitt ermittelt werden [14]. Dabei sind Punktkorrespondenzen zwischen 2D-Bildpunkten und 3D-Objektpunkten erforderlich, die beispielsweise durch Referenzmarken auf dem realen Objekt markiert sind. Häufig kommen hierfür runde Marker mit binären Codes zum Einsatz, die mittels Bildverarbeitungsmethoden [20,10] erkannt und ausgewertet werden können.

Mit den bekannten Abbildungsparametern (EOR, IOR) lässt sich jedem 3D-Punkt auf der Oberfläche ein entsprechender Punkt im Kamerabild zuweisen. Die Oberfläche des digitalen Modells wird durch Dreiecke beschrieben, und das Hinzufügen sogenannter Texturkoordinaten  $u$ ,  $v$  an den Dreiecksvertices ermöglicht die Projektion eines Bildausschnitts auf die Dreiecksfläche. Der Bildausschnitt wird durch die entsprechenden Punkte im Kamerabild definiert [19,7].

Der derzeitige Standard zur objektiven Sicherung von Spurenbildern und -lagen für gerichtliche Verfahren erfolgt bisher in 2D und meist unter Tageslichtbedingungen. Der Einsatz multispektraler Techniken, wie im UV- oder IR-Bereich, erfordert ferner spezialisierte Fachkenntnisse in der Spezialfotografie und entsprechende technische Ausstattung. Dies führt dazu, dass diese berührungsfreie, spurenschonende Methode nur selten angewendet wird. Es gibt bisher keine abgestimmte Prozessbeschreibung, welche rechtsmedizinische (forensische 2D-Fotografie) und kriminaltechnische, messtechnische Methoden (3D-Scans) für die multispektrale Spurensicherung vereint.

## 2 METHODE

Im angestrebten Prozess kommen 3D-Daten zum Einsatz, welche meist durch handgeführte 3D-Scanner aufgenommen werden. Zu nennen sind hier insbesondere die marktverfügbaren Systeme *Artec EVA*, *Artec LEO* und *Artec Spider* welche auch beim Landeskriminalamt Sachsen im Einsatz sind. Diese Systeme arbeiten mittels strukturierter Beleuchtung und liefern 3D-Oberflächendaten in Form von Dreiecksnetzen. Die multispektrale Fotografie erfolgt ebenfalls zur Spurensicherung und arbeitet mit digitalen Kamerasystemen (Nikon Z7II), Festbrennweiteobjektiven (Nikkor Z 50mm F1.2 S) und verschiedenen Filtern, welche nur enge Wellenlängenbereiche des Lichtes passieren lassen. Gleichzeitig werden verschiedene Lichtquellen in die Aufnahmekonfiguration integriert. Ziel der Untersuchung ist es, zu beurteilen, mit welchem Abbildungsfehler Texturen, welche auf Basis forensischer Fotografien aufgenommen wurden, auf das 3D-Modell aufgebracht werden können. Dies ist nötig um Aussagen zur Formspur und deren möglichen Verursacher gerichtsfest treffen zu können.

### 2.1 KAMERA-KALIBRIERUNG

Die photogrammetrische Kamerakalibrierung ist ein wesentlicher Schritt, um die 3D-Rekonstruktion der Oberflächentextur anhand von 2D-Bildaufnahmen sicherzustellen. Ziel der Kalibrierung ist es, die IOR (wie Brennweite, Lage des Bildhauptpunkts sowie Verzeichnungsparameter) sowie die EOR (Lage Kamera zu Objekt - Rotation, Translation) präzise zu bestimmen. Damit ist jedem 3D-Punkt ein Bildpunkt zuordenbar. [14,6]

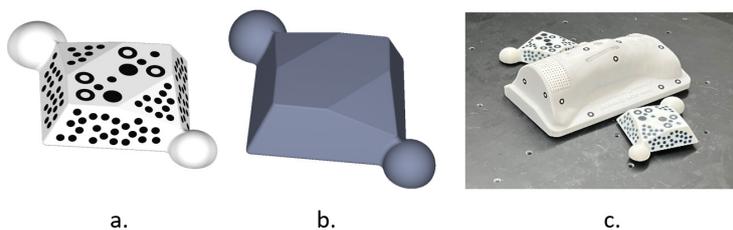
Die Kalibrierung erfolgt üblicherweise mithilfe einer geometrischen Referenzstruktur (Testfeld) mit präzise eingemessenen Punkten oder Markern. Durch das Fotografieren des Testfelds aus verschiedenen Winkeln können 2D-Bildpunkte und die entsprechenden 3D-Objektpunkte in Bezug gesetzt werden. Diese

Punktkorrespondenzen ermöglichen eine iterative Berechnung der IOR-Parameter im sogenannten Bündelblockausgleich, einem Verfahren, das die Abweichungen zwischen den gemessenen und berechneten Bildkoordinaten minimiert. In der Praxis ist das Modell der zentralperspektivischen Abbildung von Bedeutung, welches die Kamera als Projektionszentrum darstellt und eine genaue Abbildung zwischen Bild- und Objektkoordinaten erlaubt. Sind IOR und EOR bekannt, lassen sich Kollinearitätsgleichungen aufstellen, mit dem eine Transformation eines 3D-Punktes in den Bildraum ermöglicht wird. Um den photogrammetrischen Ablauf bei Nahbereichsaufnahmen zu vereinfachen und den Zeit- sowie Arbeitsaufwand zu reduzieren, wird statt der Testfeldkalibrierung eine Simultankalibrierung eingeführt, die es ermöglicht, auf vorherige Kalibrierungen zu verzichten oder diese nur bei Bedarf zu ergänzen. Dabei soll die jeweilige IOR direkt während der Datenaufnahme berechnet werden [14].

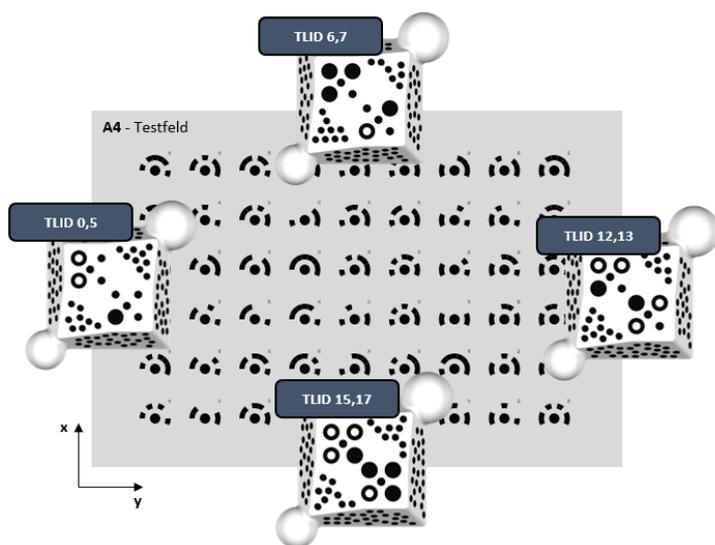
Üblicherweise ist eine vollständige Abdeckung im Bild mit Passpunkten zur Simultankalibrierung im hier vorliegenden forensischen Anwendungsfall nicht möglich. Vielmehr können nur Referenzmarker um den Bereich der aufzunehmenden Spur (bspw. Hämatom) platziert werden. Daher wird nachfolgend untersucht, welchen Einfluss die Positionierung der Referenzmarker auf die Abbildungsgenauigkeit der Textur, auch unter verschiedenen forensischen Lichtbedingungen hat. Als Referenzkörper dienen eigens entwickelte Geometrien (Abb. 3), welche neben codierten Referenzmarkern und Detektion im Bild auch markante geometrische Eigenschaften für die spätere Registrierung im 3D-Raum aufweisen. Die monochromatische Referenzmarken in Form eines 2x2-Kreisgitters ähnlich [13] stellen über die fünf kreisförmigen Markenbestandteilen unmittelbar ausreichend Korrespondenzpunkte für die zu initiale EOR-Ausrichtung bereit, wobei die Indexierung durch eine ternäre Kodierung der einzelnen Passpunkte in drei verschiedene Zustände (0=kleiner Passpunkt, 1=großer Passpunkt, 2=konzentrischer hohler Passpunkt) erfolgt. Diese Kodierung erzeugt eine eindeutige Codekombination auf der Grundlage einer Lyndon-Faktorisierung [16]. Die Referenzkörper, im folgenden Ternary-Lyndon-Referenzkörper (TLRK) genannt, sind mit zusätzlichen uncodierte kreisrunde Umgebungspasspunkte (schwarz auf weiß) versehen, um eine ausreichende Anzahl >10 an korrespondierenden Punkten zu erhalten, die für die Simultankalibrierung unerlässlich sind. Die TLRK wurden mittels Mehrfarben-3D-Druck (FFF-Verfahren, Drucker Bambu Lab P1S) additiv gefertigt und 3D Vermessen (Zeiss ATOSQ, MF 100x100). Die Abweichung liegt dabei unter 0.12mm.

## 2.2 VERSUCHSAUFBAU UND DURCHFÜHRUNG

Der geplante Testaufbau soll grundlegend mit der angestrebten Anwendung im rechtsmedizinischen Kontext übereinstimmen (Dokumentation und Vermessung von Hämatomen). Es wurde ein Messbereich in etwa der Größe eines A4-Blattes vorgesehen. In die Ecken dieses Bereiches werden die genannten TLRK platziert (Abb. 4). Das Testfeld besteht weiterhin aus 14bit-codierten Gray-Code-Ringreferenzmarken, die in der Analyse als Kontrollpunkte eines Soll-Ist-Vergleichs dienen sollen.



**Abb. 3:** TLRK mit codiertem 2x2-Kreismuster und 3D-Referenzmerkmalen



**Abb. 4:** Testaufbau mit TLRK und A4-Testfeld. Alle Marker haben eine eindeutige Codierung durch die ternären Lyndon-Kreismuster.

Zur 3D-Digitalisierung des Aufnahmesettings wird ein 3D-Scan mittels Artec LEO erstellt, auf den dann die Bildaufnahmen registriert werden sollen. Damit erreicht die 3D-Oberfläche eine Punktgenauigkeit von unter 0.1mm und eine Ortsauflösung von ca. 0.2mm (laut Herstellerangaben [3]). Zusätzlich werden Aufnahmen des Messbereiches unter verschiedenen Beleuchtungs- und Filterkonfigurationen angefertigt (IR-Infrarot, LA-Laser, VIS-Normallicht, UV-Ultraviolett). Für diese Bildaufnahmen wird eine Nikon Z7II mit einem Nikkor Z

50mmF1.2S Objektiv verwendet (siehe Tabelle 1). Diese wurden mit den Experten für kriminalistische Fotografie des Landeskriminalamtes Sachsen abgestimmt und beruhen auf deren Erfahrungen und entsprechen einem realen Aufnahmeszenario. Zur Auswertung und Analyse der multispektralen Bildaufnahmen wurde der Testaufbau vorab mit einem Zeiss ATOSQ, MF100x100, exakt eingemessen, sodass die Soll-Lage der Gray-Code-Referenzmarken und damit die Objektkontrollpunkte zu den jeweiligen TLRK im 3D-Raum bekannt sind.

	Spektrum [nm]	Strahlenquelle	Filter
IR	960	Dedolight Mini-DLOB	IR-Hochpass(700 nm)
LA	447	Crime-lite® blau	Bandpass (530 nm)
VIS	380 ... 789	Licht	Neutral
UV	365	Dedolight Mini-DLOB	ohne Filter

**Tab. 1:** Gegenüberstellung der Konfigurationen des verwendeten Aufnahmesystems mit Nikon Z7II und Nikkor Z50mm F1.2S. In absteigender Reihenfolge von der am Sensor ankommenden minimalen Wellenlänge.

Für die Registrierung der TLRK im 3D-Scan wird deren Geometrie mithilfe eines Iterative-Closest-Point-Best-Fit-Verfahrens ausgerichtet. Dabei muss zuvor die Dreiecksregion im 3D-Scan ausgewählt werden, die zur Anpassung der Rotations- und Translationsparameter verwendet werden soll. Die ermittelte Transformation wird anschließend auf die kodierten und unkodierten Objektpunktkoordinaten der Referenzmarken und Passmarken angewendet, sodass eine gemeinsame Messszene entsteht.

Nach den multispektralen Bildaufnahmen werden die TLRK detektiert, sodass in Verbindung mit den Umgebungspasspunkten das Messbild ausgerichtet, simultan kalibriert und orientiert wird, wodurch IOR und EOR bereitstehen. Grundsätzlich wird dafür nur ein TLRK benötigt, um diese Simultankalibrierung durchzuführen.

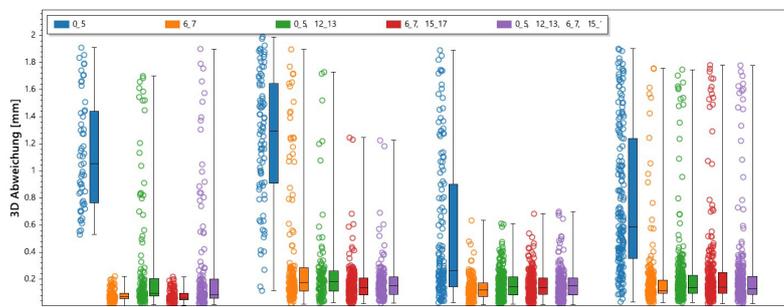
Im Gegensatz zu einer Testfeldkalibrierung wird das Aufnahmesystem bei der Simultankalibrierung nicht über das gesamte Messvolumen des Schärfentiefebereichs des Objektivs ausgeglichen. Stattdessen findet die IOR-Approximation nur für den begrenzten Bildbereich statt, in dem die TLRK abgebildet sind. Dadurch kann es potenziell zu größeren Abbildungsfehlern der Textur des 3D-Scans in den Bildbereichen des Spurmerkmals kommen, die nicht von Korrespondenzpunkten abgedeckt sind. Zur Untersuchung der Simultankalibrierung werden nun verschiedene Zusammenstellungen der TLRK im Testaufbau betrachtet, um die Texturierungsgenauigkeit bzw. die Projektion der Bildpunkte auf den 3D-Scan zu untersuchen: Es werden alle 4 TLRK ([0\_5], [12\_13], [6\_7], [15\_17]), die gegenüberliegenden TLRK ([0\_5]&[12\_13]

oder [6\_7]&[15\_17]) sowie einzelne TLRK ([0\_5] oder [6\_7]) genutzt. Die Genauigkeitsbeurteilung der 2D- und 3D-Abweichungen erfolgt anhand der Distanz zum nächstgelegenen Objektkontrollpunkt. Dabei wird die 2D-Ist-Koordinate, also die detektierten und unverzerrten Pixelkoordinaten (Bildpunkte) des Testfelds in der Messaufnahme, auf den 3D-Scan projiziert. Durch das Schneiden der Sehstrahlen mit dem Dreiecksnetz ergibt sich die 3D-Ist-Koordinate der Objektkontrollpunkte und die 3D-Genauigkeit in Millimetern. Ebenfalls werden die eingemessenen 3D-Soll-Koordinaten der Objektkontrollpunkte in das Bild projiziert, sodass die 2D-Genauigkeit in Pixeln bereitsteht.

### 3 ERGEBNISSE UND DISKUSSION

Grundlage der Ergebnisanalyse ist somit die Gesamtbeurteilung der Simultankalibrierung mit der Referenzmarkendetektion und den verwendeten Ellipsenfitting-Algorithmen. Messbilder mit unscharfen Teilbereichen werden nicht gesondert aussortiert, sodass eine generelle Aussage über den bereitstehenden Anwendungsprozess getroffen werden kann. Angeglichene IOR-Parameter sind die Brennweite und der Bildhauptpunkt, da die Einbeziehung weiterer Verzeichnungsparameter keine Verbesserung der Ergebnisse zeigte. Es werden mindestens fünf identische Messpositionen im Bereich von 90° bis 45° zum Testfeld pro Wellenbereich (VIS, UV, Laser und IR) verwendet und zur Gesamtgenauigkeit zusammengefasst. Beim Vergleich der TLRK-Zusammenstellungen (siehe Abb. 5) mit den jeweiligen Wellenspektren liegt die erreichbare RMSE (Root Mean Squared Error) zwischen 0.275mm und 0.522mm für alle vier TLRK, zwischen 0.231mm und 0.526mm für gegenüberliegende TLRK und zwischen 0.176mm und 0.859mm für einzelne TLRK. Die maximale Abweichung von 0.859mm wird dabei im

Bereich der forensischen Spurendokumentation als ausreichend gering angesehen, muss aber in jedem Strafverfahren bzgl. der gerichtlichen Verwendbarkeit einzeln beurteilt werden. Wie zu erwarten, unterscheiden sich die RMSE-Werte der TLRK-Zusammenstellungen dahingehend, dass mehr TLRK und eine regelmäßige Verteilung um das Messobjekt bessere Genauigkeiten sicherstellen (Vgl. ([0\_5] zu [6\_7]). Hervorzuheben ist bei einzelnen TLRK, dass die Abweichungsstreuung mit zunehmendem Abstand zum TLRK wesentlich zunimmt und nicht nur einzelne Ausreißer >1mm zu beobachten sind (siehe Abb. 6). Um eine ausreichende Genauigkeit im Messvolumen sicherzustellen, sollten daher mindestens zwei TLRK an den gegenüberliegenden Seiten des Messobjekts platziert werden. Die Wellenfrequenzen weisen zu einander über die Kombinationen ähnliche Abweichungscharakteristika auf, wobei die geringe Abweichung von IR mit einem maximalen RMSE von 0.275mm bei der Verwendung von mindestens zwei TLRK eine sehr geringe Abweichungsstreuung besitzt. Eine Abweichungs- zu Wellenlängenkorrelation kann jedoch nicht eindeutig festgestellt werden. Letztendlich ist das zentralperspektivische Abbildungsmodell für jedes Aufnahmesystem und damit in jedem verwendeten Wellenspektrum nutzbar.



**Abb. 6:** Streuungs- und Lagemaße der 3D-Genauigkeit der jeweiligen Wellenfrequenz über TL-Referenzmarkerkombinationen.

Ternary Lyndon Marker Kombinationen

	[0,5]	[6,7]	[0,5], [12,13]	[6,7], [15,17]	[0,5], [12,13], [6,7], [15,17]
VIS	RMSE 3d:0.742 [mm] 2d:11.067 [pix]	RMSE 3d:0.386 [mm] 2d:5.182 [pix]	RMSE 3d:0.451 [mm] 2d:6.405 [pix]	RMSE 3d:0.474 [mm] 2d:6.533 [pix]	RMSE 3d:0.46 [mm] 2d:6.442 [pix]
IR	RMSE 3d:0.576 [mm] 2d:8.395 [pix]	RMSE 3d:0.179 [mm] 2d:1.997 [pix]	RMSE 3d:0.269 [mm] 2d:2.686 [pix]	RMSE 3d:0.246 [mm] 2d:2.524 [pix]	RMSE 3d:0.275 [mm] 2d:2.751 [pix]
UV	RMSE 3d:0.859 [mm] 2d:11.655 [pix]	RMSE 3d:0.529 [mm] 2d:6.251 [pix]	RMSE 3d:0.415 [mm] 2d:5.863 [pix]	RMSE 3d:0.347 [mm] 2d:4.335 [pix]	RMSE 3d:0.367 [mm] 2d:4.543 [pix]
LA	RMSE 3d:0.873 [mm] 2d:12.419 [pix]	RMSE 3d:0.234 [mm] 2d:3.226 [pix]	RMSE 3d:0.526 [mm] 2d:5.751 [pix]	RMSE 3d:0.231 [mm] 2d:3.18 [pix]	RMSE 3d:0.522 [mm] 2d:5.721 [pix]

**Abb. 5:** 3D- und 2D-Genauigkeit der jeweiligen elektromagnetischen Wellenfrequenz über TL-Referenzmarkerkombinationen.

#### 4 ZUSAMMENFASSUNG UND AUSBLICK

Die vorliegende Arbeit untersucht die 3D-Integration multispektraler Bilddaten in der forensischen Spurensicherung. Die forensische Fotografie nutzt verschiedene Wellenlängenbereiche, um latente Spuren, wie Hämatome, sichtbar zu machen, die im sichtbaren Spektrum nicht erkennbar sind. Um die Vorteile multispektraler 2D-Aufnahmen mit der Präzision und Messbarkeit von 3D-Scans zu kombinieren, wurde ein Ansatz entwickelt, der photogrammetrische Referenzkörper nutzt. Mittels eines Texturierungsprozesses werden die multispektralen 2D-Aufnahmen auf das 3D-Modell projiziert, wodurch eine detaillierte und messtechnisch auswertbare 3D-Darstellung des Spurenträgers entsteht.

Die Genauigkeit des Verfahrens wurde anhand von Tests mit verschiedenen TLRK-Anordnungen und Beleuchtungsszenarien evaluiert. Die Ergebnisse zeigen, dass bei Verwendung von mindestens zwei TLRK eine maximale Abweichung von unter 1mm erreicht wird, was im Bereich der forensischen Spurendokumentation als ausreichend präzise angesehen werden kann. Damit leistet der Beitrag einen wichtigen Anteil zur Entwicklung standardisierter Verfahren für die 3D-Integration multispektraler Bilddaten in der forensischen Praxis. Die vorgestellte Methode bietet ein vielversprechendes Werkzeug für die forensische Spurensicherung und -dokumentation und hat das Potenzial, die Beweisführung in Gerichtsverfahren zu stärken.

Die Arbeit des Forschungsprojektes MultiForensic3D wird im Zuge der Bekanntmachung „Anwender – Innovativ: Forschung für die zivile Sicherheit II“ des Bundesministeriums für Bildung und Forschung (BMBF) im Rahmen des Programms „Forschung für die zivile Sicherheit“ der Bundesregierung gefördert.

#### LITERATUR

1. Amit Raj, Cusuh Ham, Connelly Barnes, Vladimir G. Kim, Jingwan Lu, James Hays: Learning to generate textures on 3d meshes. In: CVPR Workshops (2019)
2. Arnd Voßenkaul: Virtuelle Realität - Polizeiarbeit 2.0. In: Der Kriminalist, vol. 3/2018 (2018)
3. Artec3D: Studie über gerichtsmedizinische Verfahren vergleicht Tatortdokumentation durch 3d-scans mit fotografie (2021), <https://www.artec3d.com/de/cases/>
4. Baker, H.C., Marsh, N., Quinones, I.: Photography of faded or concealed bruises on human skin. *Journal of Forensic Identification* (2013)
5. Black, H.I., Coupaud, S., Daéid, N.N., Riches, P.E.: On the relationships between applied force, photography technique, and the quantification of bruise appearance. *Forensic science international* (2019)
6. Brown, D.C.: Close-range camera calibration. *Photogrammetric Engineering* (37 (8)), 855–866 (1971)
7. Catmull, E.E.: A Subdivision Algorithm for Computer Display of Curved Surfaces. Dissertation, University of Utah (1974)
8. Elias, M., Eltner, A., Liebold, F., Maas, H.G.: Assessing the influence of temperature changes on the geometric stability of smartphone- and raspberry pi cameras. *Sensors* (Basel, Switzerland) (2020)
9. Gajinov, Z., Matić, M., Prčić, S., Đuran, V.: Optical properties of the human skin. *Serbian Journal of Dermatology and Venerology* (2010)
10. Hough, P.V.C.: Method and means for recognizing complex patterns (US3069654A) (25031960)
11. Lee, W.C., Khoo, B.E.: Forensic light sources for detection of biological evidences in crime scene investigation: A review. *Malaysian Journ. of Forensic Sciences* (2010)
12. Limmen, R.M., Ceelen, M., Reijnders, U.J.L., Joris Stomp, S., de Keijzer, K.C., Das, K.: Enhancing the visibility of injuries with narrow-banded beams of light within the visible light spectrum. *Journal of forensic sciences* (2013)
13. Loipführer, M.: Efficient and Robust Circle Grids for Fiducial Detection. Masterarbeit, Technische Universität München, München (2022)
14. Luhmann, T.: Nahbereichsphotogrammetrie - Grundlagen, Methoden und Anwendungen. Wichmann Herbert, Rostock, 3. auflage edn. (2010)
15. Norberti, N., Tonelli, P., Giaconi, C., Nardi, C., Focardi, M., Nesi, G., Miele, V., Colagrande, S.: State of the art in post-mortem computed tomography: a review of current literature. *Virchows Archiv* (2019)
16. Pierre Duval, J.: Factorizing words over an ordered alphabet. *Journal of Algorithms* 4(4), 363–381 (1983). [https://doi.org/10.1016/0196-6774\(83\)90017-2](https://doi.org/10.1016/0196-6774(83)90017-2)
17. Ralf Breker: Virtuelle Realität: Aufbruch in eine neue Wirklichkeit. In: *Kriminalistik* 2019, vol. 43 (2019)
18. Scafide, K.N., Sharma, S., Tripp, N.E., Hayat, M.J.: Bruise detection and visibility under alternate light during the first three days post-trauma. *Journal of forensic and legal medicine* (2020)
19. Shen, I.C., Wang, Y.H., Chen, Y.M., Chen, B.Y.: Texturing and deforming meshes with casual images, <https://arxiv.org/pdf/1809.03144>
20. Zhou, G.: Accurate determination of ellipse centers in digital imagery. In: *ASPRS Annual Convention*, pp. 256–264 (1986)

# Automated Re-Photography: Localization of Historical Photographs for Urban Environments using Structure-Feature Based Camera Pose Estimation

Oliver Tworkowski<sup>1</sup>

<sup>1</sup>University of Applied Sciences (HTW) Berlin, Treskowallee 8, 10318 Berlin, Germany

eMail: Oliver.Tworkowski@Student.Htw-Berlin.de

**Abstract.** In the presented paper, the author will examine and evaluate the latest 6-DoF camera localization techniques, with a particular emphasis on their suitability for use with historical photographs in urban environments. The objective is to identify a robust and adaptable pipeline that can be used on any historical photograph, even in the absence of known information about the photograph. The paper will commence with an overview of select papers from recent years on the general and historical localization of cameras. In order to gain insight into the requirements for historical photo localization, it is first necessary to analyze the inherent challenges of such a scenario. One promising approach is structure-feature-based hierarchical localization with local learning tasks for the extraction and matching of robust and accurate features. Initially, a synthetic dataset will be employed to assess the viability of this approach, with the objective of eventually utilizing the findings for the implementation of a practical application for automated re-photography. Subsequently, the localization results from the proof of concept application will be evaluated using authentic historical photographs. The findings demonstrate that this structure-feature-based methodology can accurately establish correspondences in challenging scenarios. However, it is deficient in identifying structures in image pairs with minimal co-visibility. While this foundation is encouraging for automated localization of historical photographs, numerous enhancements could facilitate greater efficiency and adaptability to diverse scenarios.

**Keywords.** re-photography · pose estimation · camera localization · structure-from-motion

## 1 INTRODUCTION

In recent years, the intersection of technology and history has opened new ways for experiencing and understanding our past. As we navigate the digital age, the concept of interactive history has emerged as a compelling approach to bring historical narratives closer to the public through Mixed Reality (XR). The combination of recent advancements in deep learning for computer vision (CV) tasks and the continual improvement of mobile device hardware has the potential to provide users with highly immersive experiences.

One common subject of analog historical exploration is the process of localizing a place from a historical photograph and attempting to reproduce a contemporary photograph from the exact same position and orientation. This practice is referred to as repeat photography (re-photography) [1] and results in two images that can be compared side-by-side or even overlapped, thus immersing the viewer in the context. The practice of re-photography is ubiquitous across a range of media, industries, and motifs [2].

The presented paper demonstrates an attempt at fully automating the process through the use of a mobile app with a sophisticated structure-feature-based localization pipeline. To facilitate comprehension of the operational principles of 6-DoF camera localization, a concise overview of the fundamental concepts will be provided, followed by an examination of the prevailing approaches in the context of the challenges associated with the use of historical photographs in that process.

An early form of photogrammetry has been introduced as early as the 1880s by Albrecht Meydenbauer [3]. Erwin Kruppa significantly contributed to the field with his foundational ground work for determining the epipolar geometry between two calibrated cameras with no more than five common correspondences in the early 1900s [4]. Throughout the 20th century, the computational capabilities of computers facilitated the digitization of data, enabling the creation of sophisticated algorithms that, by the turn of the 21st century, constituted a comprehensive photogrammetry pipeline for the purposes of localization and mapping. [5]. Classic structure-from-motion uses handcrafted feature detectors and descriptors like SIFT [CITE LOWE 2004] or ORB. The 2010s opened an exciting window of opportunity for alternative deep learning-based approaches.

With regard to the subject of re-photography, and more particularly the utilization of historical images, a few digitization initiatives have been undertaken. However, there is as of yet no fully automated pipeline incorporating this. This may be due to a number of reasons, one of which being that this specific task is inherently difficult to generically implement. Another being, that arguably the main concept of re-photography is the freedom the user has in manually localizing a historical photo. In an early effort in 2010 Bae et al. present such a semi-automated re-photography app that provides step-by-step guidance to the user in achieving the correct pose. The authors employ a combination of SIFT features and Approximate Nearest Neighbor in conjunction with a low ratio test threshold for robust feature matching and RANSAC for outlier detection [6]. In a more recent attempt, students from the university of Osnabrück in Germany developed an Interactive Web Application for the Creation, Organization, and Visualization of Repeat Photographs [1], [7]. The platform offers a method for interactive image registration and matching, while leaving the actual capturing of the re-photography up to the user. In 2019 developed an app, that gives the user automated feedback about where to move next and how to rotate the camera to achieve the goal. As the authors describe efficiency and time cost as one of their requirements, they use binary ORB features which are faster to process. The use case of the app however, is to solve a typical problem of travel photography, where a photo with a motive ought to be retaken from the exact perspective, like family photos [8]. In a comprehensive study from 2022, the reader is given a detailed overview of the applications and techniques employed in re-photography. The authors conclude that further research is necessary to advance the field of automated re-photography, with particular attention to the potential of deep local features. In the development

of such applications, future research should also prioritize the human-computer interaction component [2].

## 2 STATE-OF-THE-ART 6-DoF IMAGE-BASED CAMERA POSE LOCALIZATION

Localizing a camera pose with 6-DoF consists of finding all degrees of freedom in a 3D Cartesian coordinate system, which comprises of the position and orientation along the three axis respectively. Such pose estimators can generally be split into two main categories: structure-based and regression-based [5]. Structure-based methods depend on a virtual 3D structure of the scene for localization, whereas regression-based methods predict a pose solely based on learning and optimization using deep neural networks.

Usually virtual 3D structure for localization is represented as a point cloud, which has been generated by using feature-based correspondences. The correspondences can either be matched directly from 2D features to 3D points or entirely through 3D points, or through hierarchical matching using image retrieval with “global” features or local features. Learning-based features have shown very promising improvements in accuracy and robustness over the years compared to handcrafted features [CITE]. On the side of regression-based pose estimators there are three major categories: absolute camera pose regression, relative – and scene coordinate regression.

Absolute pose regression is an end-to-end deep learning approach for predicting the position and orientation of a camera pose within a given world coordinate system. Such pose regressors have proven to handle monocular images, image sequence auxiliary and even whole videos as input for localization. On the other hand, relative pose regression predicts the relative pose first, based on explicit retrieval or through a CNN using a geo-referenced dataset for example. Due to being trained on general multi-unseen scenes, such regressors are more nscalable [5]. Scene coordinate regression predicts 3D coordinates directly from pixels in the query image. This way a dense correspondence mapping is calculated. However the authors state that SCR's have “not yet proven their capacity to be as effective in large-scale scenes” [5, p. 17].

In another paper from 2019, Shavit and Ferens present a similar categorization, but instead only distinct between end-to-end and hybrid deep pose estimation [9].

## 2.1 CHALLENGES

Using historical photos for image-based computer vision tasks like localization introduces new challenges in addition to already common challenges of contemporary photos. Four major categories encapsulate all such challenges well: methodology, metadata, visual quality, environment.

**Methodology** The availability of comprehensive historical datasets with consistent metadata is sparse. Many historical photos are yet to be digitized and labeled correctly which is an effort. Digitized historical photos in public archives often lack the option to be filtered correctly and private or paid archives have tedious licensing conditions to fulfill to be used for further processing. Nonetheless, there have been initiatives to improve access to such datasets, like the I-Media- Cities [10] archive. In another effort an archive has been used together with content-based image retrieval with LEA and DELF to create a selected subset of parameterized valuable historical photos [11], [12].

**Metadata** Metadata for historical images, particularly intrinsic and extrinsic camera parameters, is often lost or not transferred, which poses challenges for computer vision tasks requiring calibrated cameras. During digitization, photo quality depends on the operator's skills and equipment, leading to issues such as cropping, damage, or low resolution. Cropped images can shift the principal point, complicating automatic localization. Many historical images were captured with non-central principal points, and when camera parameters are unknown, estimation is necessary. Although parameters like skew and lens distortion are often negligible, they can enhance extrinsic parameter estimation. Selfcalibration methods, relying on vanishing points, offer limited accuracy with few images. Recent deep learning models, such as DEEPPFOCAL, DeepCalib, and MLFocalLengths, predict focal lengths but are trained on contemporary datasets, highlighting the need for further research in historical contexts. Accurate focal length estimation is crucial for 6-DoF localization, as different focal lengths can yield similar images at different positions. [CITE]

**Environment and Visual Quality** The environment can have a significant impact on the resulting visual quality of photos, especially when analog film has been treated improperly or the photographer lacked the skill to capture a qualitative photo. Furthermore, transient objects, repetitive patterns and textureless surfaces, illumination changes and other influences generally lead to an increased difficulty when trying to localize such historical photos. [CITE]

## 2.2 DISCUSSION

While state-of-the-art camera pose regression for localization has not been researched yet, structure-feature-based localization has already been achieved using learning-based features in an adapted structure-from-motion (SfM) pipeline [11] with success. Learning-based features have already shown their capabilities of being applied to multi-temporal photos [13] for detection, description and matching, where handcrafted features usually fail. State-of-the-art structure-feature based localization techniques may even outperform end-to-end regression-based localization [5]. This advantage may be due to the scene geometry being retained, which is not the case with fully predicted poses [9]. One implementation that stands out in challenging datasets, delivering robust and accurate results efficiently is HF-Net, a hierarchical localization pipeline using the coarse-to-fine strategy combined in a joint network [14]. This localization strategy can also be observed when using global and local features in order sequentially, which is what the authors used as an inspiration. When compared to such a pipeline using NetVlad+SuperPoint+SuperGlue (global -, local features and matching) HF-Net showed significant efficiency improvements, but also a slight decrease in accuracy. Although this approach relies on features, which generally are prone to errors when dealing with little co-visibility, SuperPoint features have already shown their capabilities of handling such scenarios well [CITE]. Furthermore, hierarchical localization employs 2D-3D matching, meaning features on a historical photo could be matched against an existing point cloud extracted from contemporary images.

## 3 MOBILE PROOF OF CONCEPT APPLICATION FOR HISTORICAL PHOTOS

In order to show compelling real world historical photo localization examples, a proof of concept mobile AR application will be developed. The app will encompass an automated hierarchical localization pipeline consisting of four steps: (1) acquiring of auxiliary contemporary photos and their respective real world coordinates  $t$  acting as ground truth, (2) generating a reference frame containing the estimated camera poses of both the historical and auxiliary photos with position  $t$  and orientation  $q$ , (3) importing the transformed estimated camera poses and ground truth camera poses into the real world coordinate system, (4) applying Helmert transformation to align both estimated and ground truth camera poses to receive the real world camera pose of the historical photo (see Figure 1). The Helmert transformation is a specific case of a similarity transformation mentioned commonly used in geodesy for transforming geodetic coordinate systems[11], [12].

### 3.1 SYNTHETIC DATASET PRE-EVALUATION

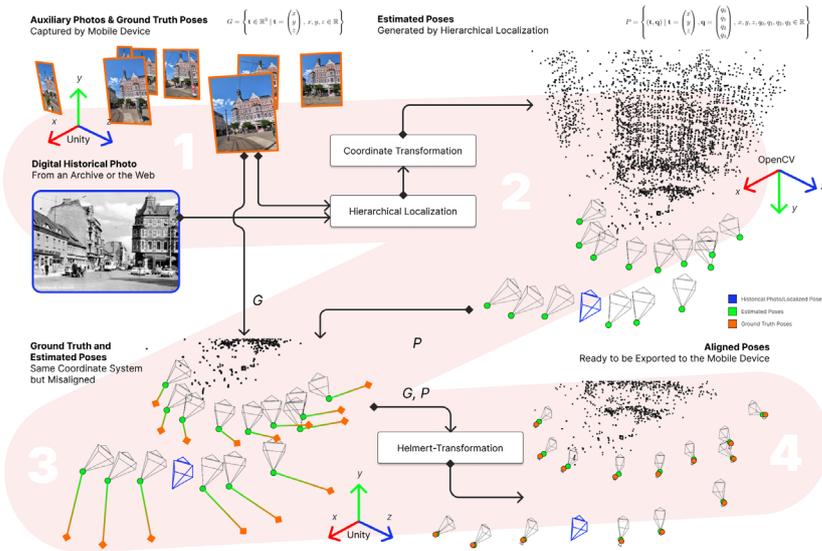
In order to choose the best feature-based matching using hierarchical localization, a preliminary evaluation has been conducted using a custom synthetic dataset [CITE dieses SfM paper?]. The comparative analysis will use two different base-localization approaches: an extended SfM pipeline using learning-based features (similar to [11]) and the full hierarchical localization using 2D-3D matching.

Both methods will be run with the same set of feature pipelines with the most promising results for historical photos based on previous benchmarks and analysis [CITE]. Figure 2a shows these results as a weighted error sum of the rotation error  $E_r$  and translation error  $E_t$ . The translation error is weighted less due to it being less impactful on the finally matched images and possibly inaccurately selected focal lengths (see Equation 1). Figure 2b shows the visual result of the best feature pipeline for both localization approaches.

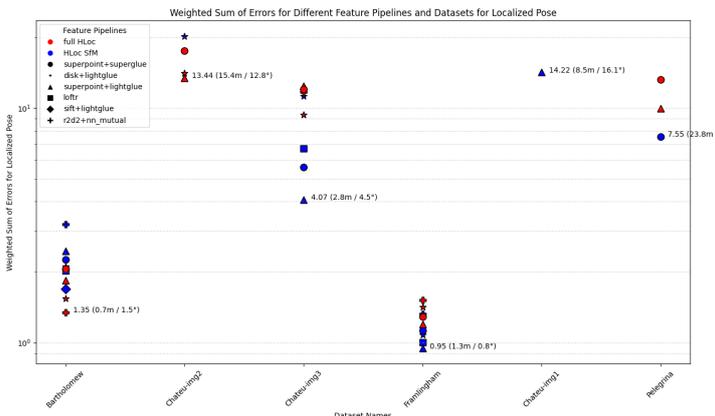
$$E = 0.25 \cdot E_t + 0.75 \cdot E_r \quad (1)$$

Against expectations, the SfM pipeline performs similar or better than the hierarchical localization (hloc) pipeline on the synthetic dataset of historical photos. Neither method is able to successfully localize the Chateau1 subset due to the challenging matching scenario. In the other two Chateau subsets, hloc seems to be estimate a more stable focal length leading to the assumption of also delivering more accurate localizations due to the historical photos in that subset likely being captured by the same camera. The visual analysis suggests all other localizations may be correct. However, some estimated pose positions are beneath the ground plane, suggesting the cause to be an inaccurately estimated focal length.

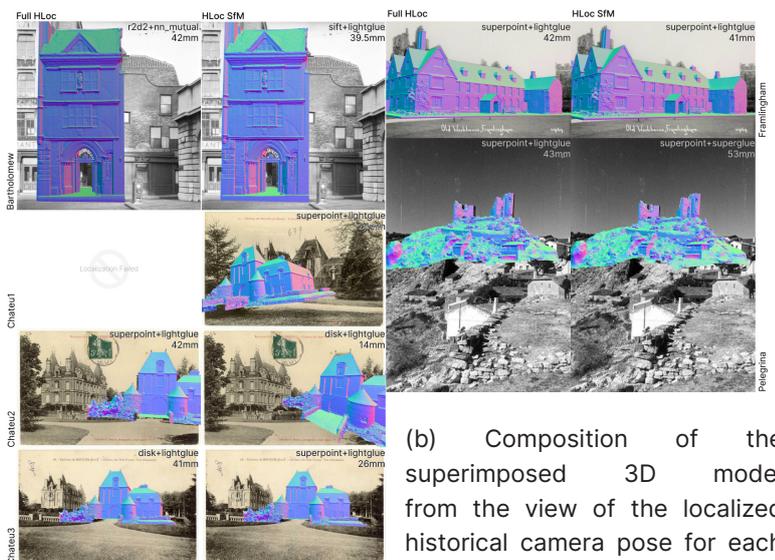
*SuperPoint+LightGlue* prevails as the feature pipeline delivering the most accurate results and will thus be used in the app.



**Fig. 1:** Proposed automated hierarchical localization pipeline for proof of concept app



(a) Comparison of feature pipelines for full Hierarchical Localization (red) and only SfM Hierarchical Localization (blue) on synthetic photogrammetry localization datasets. Each symbol displays a weighted error sum from a different feature pipeline for the localized historical photo camera pose. The errors are calculated as a weighted sum from the translation and rotation error. The lower the error, the better the pipeline localization pose. Only inliers with a summed error of below 50 are shown.



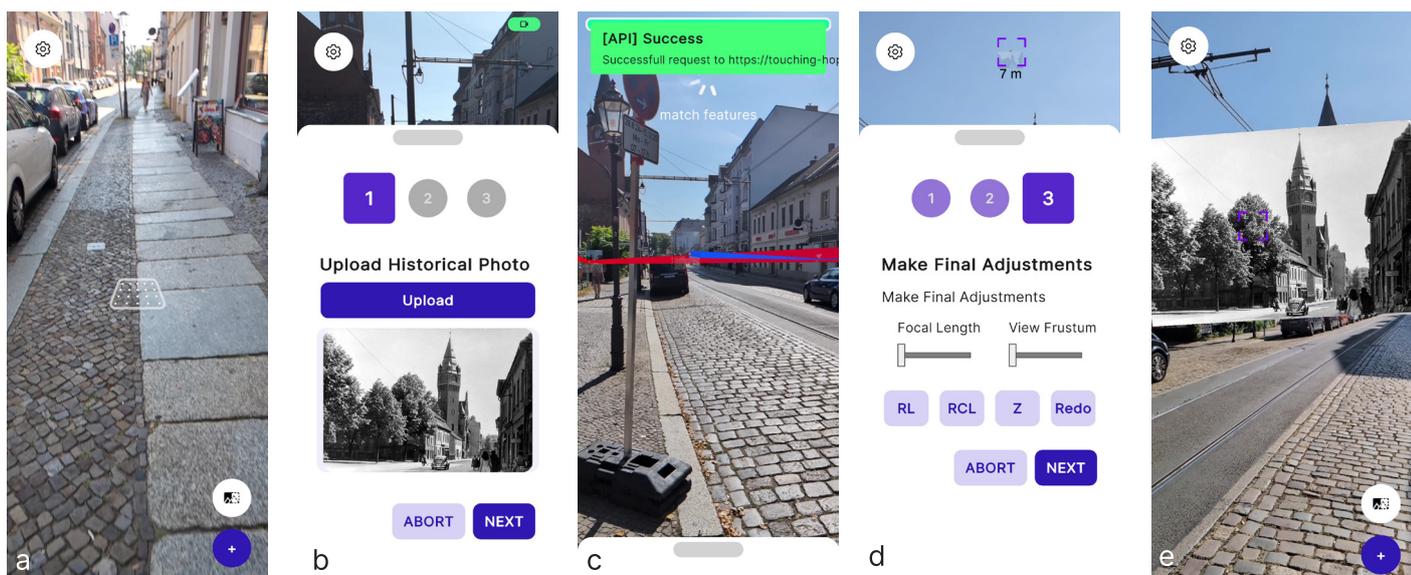
**Fig. 2:** Results of preliminary evaluation

(b) Composition of the superimposed 3D model from the view of the localized historical camera pose for each best performing pipeline for full hierarchical localization and SfM localization together with the estimated focal length in millimeters.

### 3.2 WORKFLOW AND APP NAVIGATION

The user starts by establishing a tracking context through the AR Foundation framework. The user then proceeds to create a localization context and uploading a historical photo to be localized in the first step. Afterwards, the ground truth poses are captured by walking around and manually pressing a trigger. The localization can then be initialized and started. Once the pose has been localized in the backend, the bottom drawer extends, showing additional adjustments for the pose. Such options include the adjustment of the focal length, the view frustum size of the spatial 3D photo and an option to level out the pose to eye level, reset it, or repeat the localization with the existing context. In the last step, the user can persist the pose with all necessary information. Currently the poses are only persisted throughout the session, but in the future AR anchor-based persistence could be used.

higher focal length. Localization B initially failed due to insufficient image correspondences but succeeded on a subsequent attempt with height adjustments. The Helmert transformation achieved an average alignment error of 14.38 cm, though Exhibit F had a 40 cm deviation, which was minimal given the large-scale scene. Feature matching using SuperPoint+LightGlue was effective under strong illumination but struggled with single-plane correspondences, impacting accuracy in some cases. Automated re-photography showed potential by accurately localizing photos in overgrown areas where manual matching was impossible.



### 3.3 EVALUATION AND DISCUSSION

The evaluation with real world examples uses seven exhibits (numbered A-G), ranging from 1906 to 1987. All query photos use a single grayscale channel. Localization typically required approximately 6.28 minutes, utilizing an average of 15 auxiliary images and one historical photo per attempt. The effort and duration increased with the number of input images. Notably, Localization run C was faster by reusing the previous tracking context, which enhanced both performance and precision.

Most localization attempts were successful except for runs A and B. Localization A matched the historical photo but had significant positional errors due to the large scene scale, limited camera spread, and

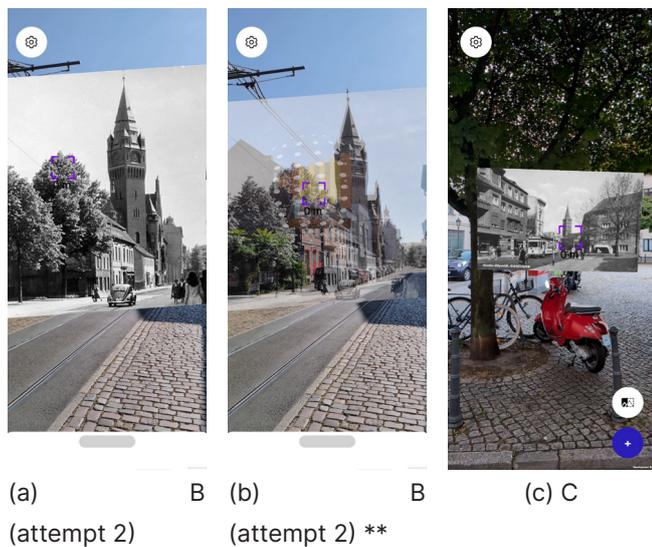
- (a) establishing tracking context from scanning environment
- (b) start localization by uploading historical photo to be localized
- (c) started localization with feedback about progress
- (d) localized camera pose settings in an extended drawer and the marker with distance to the localized pose
- (e) successfully localized camera pose without transparency

**Fig. 3.** Proof of concept app localization process

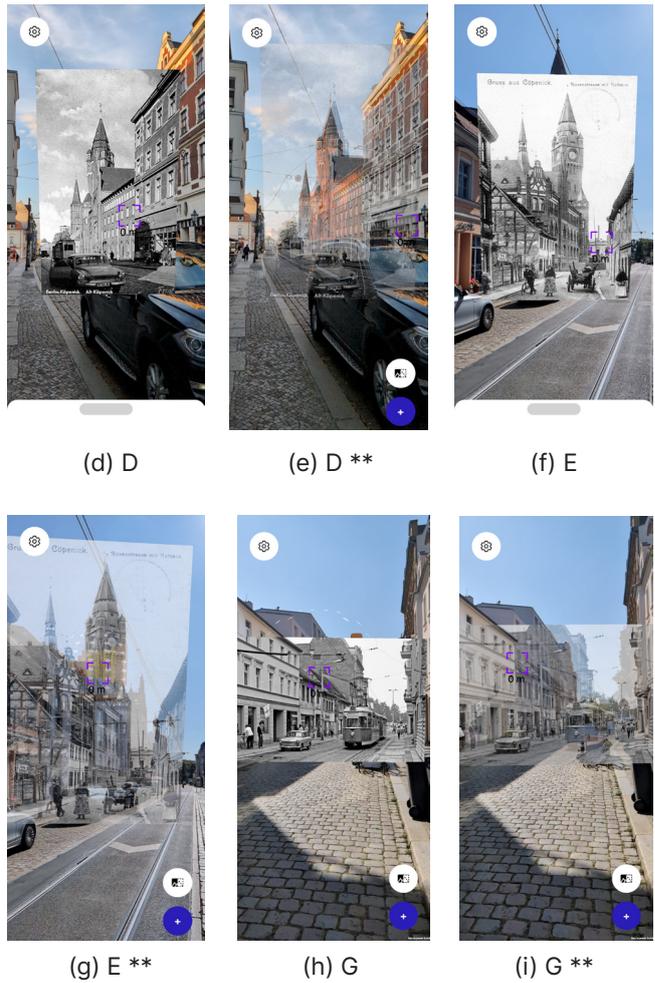
The results demonstrate that hierarchical localization combined with Helmert transformation effectively localizes historical photographs with reasonable accuracy. The success of Localization run C highlights the benefits of reusing tracking contexts, which improves both speed and precision by maintaining consistency in tracking data. However, challenges remain, particularly with feature correspondences limited to single planes, leading to misalignments in localized poses. This issue is evident when matching features from facades without sufficient overlap in other regions, resulting in peripheral misalignments.

Exhibit	Auxiliary Images	Duration (min)	Inliers Avg.	Transform Offset (cm)
A	13	06:10	508/541	14.06
B	17	07:20	13/21	11.38
B (attempt 2)	16	06:50	110/307	16.85
C	14	03:50	292/520	9.49
D	11	07:13	573/788	12.22
E	12	04:40	108/197	5.54
F	20	08:12	825/881	39.94
G	18	07:36	122/659	5.59

**Tab. 1:** Summary of all localization runs in the PoC app for historical photos from the test dataset with their exhibit. This table shows the number of auxiliary images used for that localization, the duration from app start to localized pose, the number of inliers in relation to the total number of feature matches across all images and the average position offset between ground truth and estimated poses after the transformation has been applied.



**Fig. 4:** All successful localization runs displayed as spatial photos in the proof of concept app with\*\* and without transparency



The Helmert transformation proved reliable for pose alignment, maintaining low average errors even in large-scale scenes. Nonetheless, certain cases like Exhibit F indicate that deviations can occur, although they remain acceptable within the context of expansive environments. The feature matching pipeline demonstrated robustness under strong lighting conditions but was less effective when image pairs lacked sufficient co-visibility, emphasizing the need for diverse and overlapping features in both historical and auxiliary images.

Automated re-photography emerged as a promising approach for relocalizing historical photos in challenging environments, where manual matching is infeasible. This automation leverages auxiliary images with clear views of the objects of interest, enhancing the app's ability to accurately determine original photograph positions. However, scalability remains a concern due to the timeconsuming nStructure from Motion (SfM) component and the substantial storage required for auxiliary and ground truth data.

Future improvements should focus on reducing pipeline redundancies and enhancing feature matching confidence to handle a broader range of historical photos. Such improvement may include the usage of 3D tracking data directly and incorporating it into the 2D-3D matching process and running localization entirely on the phone with HF-Net directly. Exploring feature-less matching methods could also address limitations related to single-plane correspondences. Future work should also explore the capabilities of pose regressors for historical photos. Additionally, optimizing storage solutions and enabling simultaneous localization of multiple photos would enhance the system's scalability and practicality for real-world applications. Addressing these technical challenges is essential for advancing the accuracy and efficiency of historical photo localization in diverse and large-scale environments.

#### 4 CONCLUSIONS

**Acknowledgments.** The original paper has been written as a master thesis at the University of Applied Sciences (HTW) Berlin.

#### REFERENCES

- [1] Repeat photography - institute of cognitive science, Uni-osnabrueck.de, 2021. [Online]. Available: [https://www.ikw.uni-osnabrueck.de/research\\_groups/computer\\_vision/research/repeat\\_photography.html](https://www.ikw.uni-osnabrueck.de/research_groups/computer_vision/research/repeat_photography.html) (visited on 07/19/2024).
- [2] A. Schaffland and G. Heidemann, "Heritage and repeat photography: Techniques, management, applications, and publications," *Heritage*, vol. 5, pp. 4267–4305, Dec. 2022. doi: 10.3390/heritage5040220. [Online]. Available: <https://www.mdpi.com/2571-9408/5/4/220> (visited on 07/22/2024).
- [3] J. Albertz, *Museum theulegium: Albrecht meydenbauer pionier der photogrammetrischen dokumentation des kulturerbes*, Theulegium.de, Aug. 2010. [Online]. Available: <http://www.theulegium.de/index.php?id=46> (visited on 09/02/2024).
- [4] G. Gallego, E. Mueggler, and P. Sturm, Translation of „zur ermittlung eines objektes aus zwei perspektiven mit innerer orientierung“ by erwin kruppa (1913), arXiv.org, 2017. [Online]. Available: <https://arxiv.org/abs/1801.01454> (visited on 05/04/2024).
- [5] M. Xu, Y. Wang, B. Xu, et al., "A critical analysis of image-based camera pose estimation techniques," *Neurocomputing*, vol. 570, pp. 127 125– 127 125, Feb. 2024. doi: 10.1016/j.neucom.2023.127125. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231223012481> (visited on 04/14/2024).
- [6] S. Bae, A. Agarwala, and F. Durand, "Computational rephotography," *ACM transactions on graphics*, vol. 29, pp. 1–15, Jun. 2010. doi: 10.1145/1805964.1805968. [Online]. Available: <https://dl.acm.org/doi/epdf/10.1145/1805964.1805968> (visited on 05/24/2024).
- [7] An interactive web application for the creation, organization, and visualization of repeat photographs | proceedings of the 1st workshop on structuring and understanding of multimedia heritage contents, ACM Conferences, 2019. [Online]. Available: <https://dl.acm.org/doi/10.1145/3347317.3357247> (visited on 07/19/2024).
- [8] Instant rephotography | human-computer interaction interact 2019, Guide Proceedings, 2019. [Online]. Available: [https://dl.acm.org/doi/abs/10.1007/978-3-030-29387-1\\_12](https://dl.acm.org/doi/abs/10.1007/978-3-030-29387-1_12) (visited on 07/19/2024).
- [9] Y. Shavit and R. Ferens, Introduction to camera pose estimation with deep learning, arXiv.org, 2019. [Online]. Available: <https://arxiv.org/abs/1907.05272> (visited on 04/14/2024).
- [10] L. B. I. for Digital History, I-media-cities lbi for digital history, I-Media-Cities, 2020. [Online]. Available: <https://www.geschichte.lbg.ac.at/forschungsprojekte/i-media-cities/> (visited on 08/14/2024).
- [11] F. Maiwald, C. U. Lehmann, and T. Lazariv, "Fully automated pose estimation of historical images in the context of 4d geographic information systems utilizing machine learning methods," *ISPRS international journal of geo-information*, vol. 10, pp. 748–748, Nov. 2021. doi: 10.3390/ijgi10110748. [Online]. Available: <https://www.mdpi.com/2220-9964/10/11/748> (visited on 11/26/2023).
- [12] F. Maiwald, J. Brusckke, D. Schneider, M. Wacker, and F. Niebling, "Giving historical photographs a new perspective: Introducing camera orientation parameters as new metadata in a large-scale 4d application," *Remote sensing*, vol. 15, pp. 1879–1879, Mar. 2023. doi: 10.3390/rs15071879. [Online]. Available: <https://www.mdpi.com/2072-4292/15/7/1879> (visited on 04/15/2024).
- [13] L. Morelli, F. Bellavia, F. Menna, and F. Remondino, "Photogrammetry now and then from hand-crafted to deep-learning tie points," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-2/W1-2022, pp. 163–170, Dec. 2022. doi: 10.5194/isprs-archives-xlVIII-2-w1-2022-163-2022. (visited on 04/10/2023).
- [14] P.-E. Sarlin, C. Cadena, R. Siegwart, and M. Dymczyk, From coarse to fine: Robust hierarchical localization at large scale, Apr. 2019. [Online]. Available: <https://arxiv.org/pdf/1812.03506> (visited on 08/03/2024).

# Modulares Target und Ansatz für die Geometrische Kalibrierung multimodaler Bildgebungssysteme

Evgeny Degtyarev, David Reese, and Rico Nestler

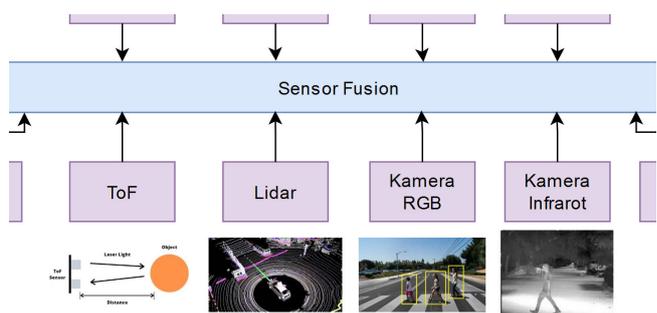
Zentrum für Bild- und Signalverarbeitung e. V.,  
Werner-von-Siemens-Straße 12, 98693 Ilmenau

evgeny.degtyarev@zbs-ilmenau.de, david.reese@zbs-ilmenau.de, rico.nestler@zbs-ilmenau.de  
www.zbs-ilmenau.de

**Zusammenfassung:** Bei der targetbasierten multimodalen 2D- und 3DSensorkalibrierung werden meist ebene, oft modenspezifische Kalibrierobjekte (Targets) verwendet, die vor einer Bildsensoranordnung orthogonal positioniert werden müssen. Diese Vorgehensweise stellt im Fall eines gemeinsam genutzten Kalibrierobjekts jedoch eine Herausforderung dar, da die einzelnen Bildgebungen oft individuell deutlich zum Target geneigt sind. Der vorliegende Beitrag widmet sich der Konzeption und Umsetzung eines modularen, multimodal nutzbaren Targets, mit dem dieses Problem überwunden, die allgemeine Kalibrierungsgenauigkeit verbessert und der multimodale Kalibrierprozess vereinfacht werden soll. Das vorgeschlagene Target integriert mehrere Funktionen, um eine Nutzung durch verschiedene Bildgebungsverfahren, z.B. RGB, TOF, Thermografie, sowohl für die intrinsische als auch die extrinsische Kalibrierung zu ermöglichen. Es umfasst planare Muster, geometrische Elemente und modulare 3D-Primitive, die am Target angebracht oder von ihm entfernt werden können, um ausreichende, robuste und betrachterunabhängige geometrische Merkmale für verschiedene Arten der Bildgebung zu bieten. Die Targetmodularität ermöglicht die gleichzeitige Kalibrierung von 2D und 3D-Sensorsystemen verschiedener Bildmodalitäten und löst damit die üblichen Herausforderungen bei der multimodalen Datenfusion und Sensorausrichtung. Der Beitrag beschreibt die Konstruktionsprinzipien eines multimodalen Targets und demonstriert die Anwendung des Targets innerhalb einer automatisierten Kalibrierungspipeline für die extrinsische Kalibrierung von multimodalen Kamera-Setups.

## 1 MOTIVATION

Die zunehmende Integration verschiedener bildgebender Sensortechnologien in modernen technischen Systemen erfordert präzise und effiziente Kalibrierungsmethoden, um geometrische Bezüge zwischen den erfassten Daten herstellen und nutzen zu können. Traditionelle, targetbasierte Ansätze sind allgemeiner Stand der Technik, aber für spezifische Bildgebungen optimiert. Aktuelle Entwicklungen in der targetlosen geometrischen Kalibrierung, insbesondere auch KI-basierte Ansätze, bieten vielversprechende neue Möglichkeiten der Nutzung auch für unterschiedliche Bildprinzipien, können häufig aber nur der Nachführung einer klassischen initialen Kalibrierung dienen.



**Abb. 1:** Prinzip der multimodalen Sensordatenfusion durch extrinsische geometrische Kalibrierung von bildgebenden RGB-, Infrarot- und 3D-Sensoren.

Vor dem Hintergrund eines steigenden Bedarfs an universellen, multimodal nutzbaren Kalibrierlösungen widmet sich dieser Beitrag der Entwicklung und Implementierung eines Kalibrierverfahrens für die gleichzeitige extrinsische Kalibrierung von unterschiedlichen 2D- und 3D-Bildgebungsprinzipien. Eine präzise extrinsische Kalibrierung ist erforderlich, um die relative Lage (Pose) der Sensoren in der Welt und zueinander korrekt zu bestimmen. Hierzu ist ein spezielles Target zu entwerfen, das für alle Bildgebungen geeignet ist. Die Eignung beinhaltet die Bereitstellung von nutzbaren Merkmalen in Targetabbildern in einer hinreichenden Zahl, um eine präzise Kalibrierung zu ermöglichen. Von diesen Anforderungen werden die Gestaltungsmöglichkeiten des Targets und die Auswahl von und Verbindung mit nutzbaren Algorithmen zur Bildauswertung und Kalibrierung bestimmt. Die besondere Herausforderung liegt in einem Targetentwurf, das für verschiedene Sensormodalitäten, d.h. für optische (VIS, NIR, MWIR, LWIR - aktiv und passiv) und geometrische Szenenmerkmale, z.B. RGB-Kameras, Time-of-Flight-Sensoren, einsetzbar ist. Jeder Sensortyp stellt hinsichtlich Materialwahl, geometrischer Gestaltung und Oberflächenbeschaffenheit ganz spezifische Anforderungen an das Target. Insbesondere die Kalibrierung von Sensoren, die die thermische Emission von Objekten bildgebend erfassen, erfordert spezielle Lösungen, da konventionelle optische Muster in Wärmebildern nicht erkennbar sind.

Der Beitrag ist gegliedert in Überlegungen zur Gestaltung und Konstruktion eines multimodalen Targets und die Diskussion der softwareseitigen Umsetzung eines Kalibrierverfahrens unter Berücksichtigung spezieller Aspekte der verwendeten Geometrien der Targetelemente. Er schließt mit einer Zusammenfassung und Erläuterungen zur Umsetzbarkeit in konkreten Anwendungen.

## 2 TARGETGESTALTUNG

Targetbasierte geometrische Kalibrierverfahren dominieren nach wie vor in Anwendungen, die hohe Genauigkeiten der aus den Bildgebungen abzuleitenden geometrischen Szeneninformationen erfordern. Typischerweise werden für optisch remissionsbasierende Bildgebungsprinzipien planare Targets mit Schachbrett- oder ChArUco-Mustern eingesetzt. Diese Methoden sind zwar etabliert und mathematisch fundiert, stoßen jedoch bei der Integration weiterer Sensormodalitäten an ihre Grenzen. Insbesondere die Einbeziehung der für eine thermische Kalibrierung häufig notwendigen beheizten Kalibrierkörper erhöhen die Komplexität

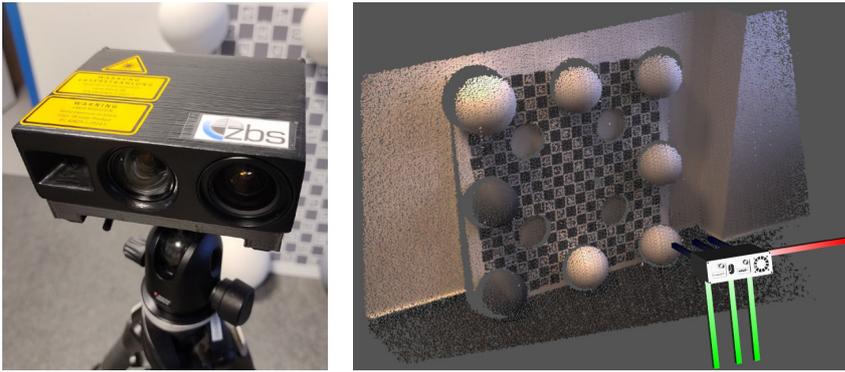
und Kosten des Kalibriervorgangs erheblich. Das in diesem Beitrag vorgestellte Konzept eines modularen, multimodalen Targets soll die Nutzbarkeit unterschiedlicher Targeteigenschaften flexibel und kostengünstig in sich vereinen. Die Grundlage bildet eine Basisplattform aus einem Aluminium- Verbundwerkstoff, die ein integriertes 2D-ChArUco-Muster für die grundlegende optische Kalibrierung enthält. Weiterhin sind 2D-geometrische Elemente integriert. Die Plattform verfügt über definierte Aufnahmepunkte, an denen weitere 3D-Primitive montiert werden können.

Die geometrischen Primitive, hauptsächlich Kugeln und Zylinder, können anwendungsspezifisch bedarfsgerecht (Formen, Größen und Materialien) mittels additiver Fertigung hergestellt werden. Diese Fertigungsmethode ermöglicht eine kostengünstige und präzise Produktion bei gleichzeitig hoher Flexibilität im Design. Als Primärelemente dienen Kugeln, die durch ihre Geometrie eine weitestgehend winkelunabhängige Erkennbarkeit gewährleisten. Die Kugelzentren können geometrisch durch drei Parameter beschrieben und validiert werden: den Radius, den Abstand zur Grundebene sowie die Abstände zu benachbarten Kugeln. Für die extrinsische Kalibrierung mit Thermographiekameras im Sensorverbund können einfache Heizelemente in die Kugeln integriert werden. Das Material und die Oberflächeneigenschaften lassen sich kooperativ anpassen. Ergänzend aufgebrachte Zylindergeometrien als Sekundärelemente erlauben eine präzise Achsdefinition und können zur Validierung der Kugelzentren dienen.

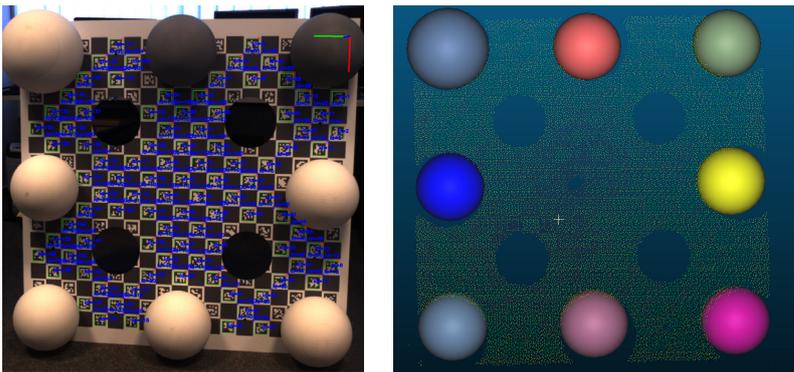
## 3 ALGORITHMISCHE UMSETZUNG DER MULTIMODALEN KALIBRIERPIPELINE

### 3.1 TARGETAUSWERTUNG UND EXTRINSISCHE KALIBRIERPIPELINE FÜR OPTISCHE VIS-MWIR-SENSOREN

Für den visuellen Spektralbereich wird eine Posenerkennung per planarem Target mit ChArUco-Mustern verwendet. Automatisch erkannte Marker mit einem bei ZBS umgesetzten Verfahren sind in Abbildung 3 dargestellt. Diese Muster können auch zur gleichzeitigen intrinsischen Kalibrierung der Bildgebung verwendet werden. Hierfür müssen mehrere Posen aufgenommen werden.



**Abb. 2:** Prototypischer Aufbau eines multimodalen Single-Shot-3DSensorsystems des ZBS e.V. (links) mit Beispielaufnahme einer RGBattributierten 3D-Punktwolke nach geometrisch extrinsischer Kalibrierung (rechts). Die 3D-Rekonstruktion erfolgt hier aktiv mit flächiger Musterprojektion und einer Kamerabeobachtung (Single-Shot). Die RGB-Kamera zur multimodalen Ergänzung der 3D-Daten ist in der Mitte installiert.



**Abb. 3:** Multimodales Target mit acht Kugelementen unterschiedlicher Größe und Farbe (links). Die Kugeln wurden aus mattem PLA im 3D-Druckverfahren hergestellt. Ergebnis der Sphärendetektion aus einer 3D-Punktwolke eines prototypischen Single-Shot-3D-Sensorsystems des ZBS e.V. (rechts).

### 3.2 TARGETAUSWERTUNG UND EXTRINSISCHE KALIBRIERPIPELINE FÜR THERMOGRAPHISCHE SENSOREN

Für den thermographischen Spektralbereich wurde ein spezieller Ansatz entwickelt, bei dem die Kugeln mit elektronischen Heizelementen ausgestattet sind. Eine individuelle Temperierung jeder Kugel ermöglicht die zuverlässige Detektion und insbesondere eindeutige Identifikation bei unterschiedlichen Targetposen im Wärmebild (siehe Abbildung 4). Zusätzliche geometrische Elemente auf dem planaren Target (Bohrungen) ermöglichen optional eine Genauigkeitssteigerung der Auswertungen im Fall nahezu senkrechter

Ausrichtungen der Kamera zum Target. Die Bohrungssegmente können aufgrund der natürlichen Temperaturunterschiede des Targets zum Umfeld leicht detektiert werden.

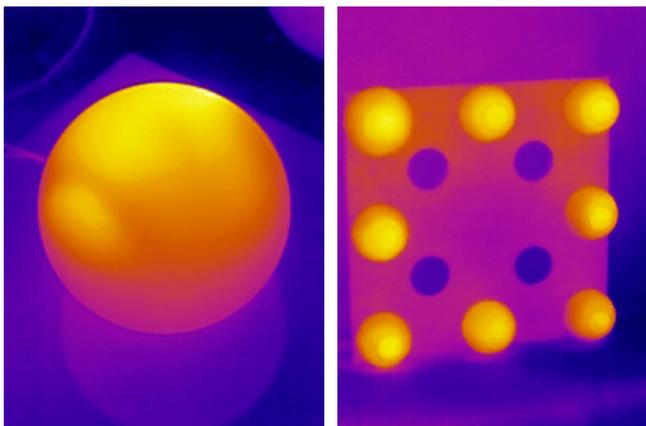
Ein fundamentales Problem bei der Verwendung von Kugeln als Kalibrierobjekt ist die perspektivische Projektion, da eine Kugel in der Bildebene stets als Ellipse abgebildet wird. Der typischerweise zur Kalibrierung verwendete Schwerpunkt der Kugelprojektion (Ellipsensegment), der wegen seines integralen Charakters zur genauen Strukturortbestimmung genutzt werden kann, entspricht im Allgemeinen nicht der Projektion des Kugelmittelpunkts. Der ansichtsabhängige systematische Fehler, der entsteht, weil der Sehstrahl durch den Ellipsenschwerpunkt nicht durch den tatsächlichen Kugelmittelpunkt verläuft, muss für die Nutzung des Targetabbildes für Kalibrierzwecke korrigiert werden [3]. Die Größe dieser Abweichung hängt vom Blickwinkel und der radialen Position im Bild ab. Auf dieser Grundlage wurde ein eigenes Korrekturverfahren entwickelt und in die Verarbeitung der Targetdaten einbezogen.

Die Posen-Erkennung für die Thermographiekamera erfolgt in mehreren Schritten:

- 1. Ellipsen-Korrektur:** Für die perspektivische Abbildung von Kugeln wird eine systematische Korrektur der projizierten Zentren vorgenommen. Der entwickelte Algorithmus modelliert und korrigiert die geometrische Beziehung zwischen der tatsächlichen Kugelposition und ihrer elliptischen Projektion in der Bildebene. Diese Korrektur ist besonders wichtig bei größeren Blickwinkeln und bei Kugelpositionen am Bildrand, wo die perspektivischen Einflüsse am stärksten ausgeprägt sind.
- 2. Blob-Detektion:** Die Abbilder der erwärmten Kugeln werden im geometrisch korrigierten Wärmebild mittels OpenCV-Blob-Detector [2] lokalisiert. Durch die unterschiedlichen Temperaturniveaus der einzelnen Kugeln wird eine eindeutige Zuordnung und Unterscheidung im Targetabbild ermöglicht.
- 3. Pose-Estimation:** Mit den korrigierten Bildpunkten und ermittelten geometrischen Parametern der geometrischen Primitive wird mittels solvePNPAlgorithmus [2] die finale Pose des Targets bestimmt.

### 3.3 TARGETAUSWERTUNG UND EXTRINSISCHE KALIBRIERPIPELINE FÜR 3D-SENSOREN

Die Verarbeitung von erfassten 3D-Punktwolken Daten eines intrinsisch kalibrierten 3D-Sensoren erfolgt in mehreren aufeinanderfolgenden Schritten, die eine robuste Erkennung und geometrische Positionsbestimmung der Kalibrierelemente ermöglichen:



**Abb. 4:** Targetaufnahme mit Thermographiekamera TOPDON TS001: Abbild einer aktiv erwärmten Kugel auf einer Aluminiumplatte (links) und Abbild des Targets mit Bohrungen und erwärmten Kugeln (rechts)

- 1. Initiale Segmentierung:** Mittels DBScan-Algorithmus [1] wird die Punktwolke in einzelne zusammenhängende Objekte segmentiert. Dieser dichtebasierte Ansatz ermöglicht eine effektive Trennung der geometrischen Primitive voneinander und vom Hintergrund.
- 2. Geometrische Filterung:** Die segmentierten Objekte werden anhand ihrer Größenordnung und geometrischen Eigenschaften gefiltert. Insbesondere wird für potenzielle Kugelsegmente überprüft, ob ihre Punktverteilung mit dem bekannten Radien des Targets kompatibel sind.
- 3. Parameterextraktion:** Für die gefilterten Segmente erfolgt eine Parameterbestimmung mittels RANSAC-Verfahren und anschließender Least-Squares-Optimierung. Dabei werden geometriespezifisch folgende Eigenschaften ermittelt, z.B.:
  - für Kugeln: Mittelpunktkoordinaten und Radius
  - für Zylinder: Achsenrichtung, Radius und Länge
  - ergänzend: Parameter der Basisebene, falls diese im Sichtfeld liegt

- 4. Modellvalidierung:** Die extrahierten Parameter werden anhand der bekannten Target-Geometrien im Targetaufbau validiert:
  - bei Kugeln: Abstände zwischen Kugelmittelpunkten, Radien der Kugeln, Abstände zur Basisebene (falls sichtbar)
  - bei Zylindern: Durchmesser und Länge
  - bei Ebenen: Ausrichtung und Ebenheit

- 5. Iterative Optimierung:** Mittels RANSAC werden die Parameter solange optimiert, bis entweder die Abweichungen unter einen definierten Schwellwert fallen oder eine maximale Iterationszahl erreicht ist. Dies gewährleistet eine robuste, schnelle und genaue Anpassung, auch bei verrauschten Daten und vermeidet Overfitting.

- 6. Pose-Estimation:** Die finale Transformation wird durch Least-Squares-Registrierung nach Umeyama [4] bestimmt. Dabei werden z.B. gemessene Kugelmittelpunkte mit dem Target-Modell in Übereinstimmung gebracht. Die eindeutige Zuordnung der Korrespondenzen wird durch Referenzkugel ermöglicht.

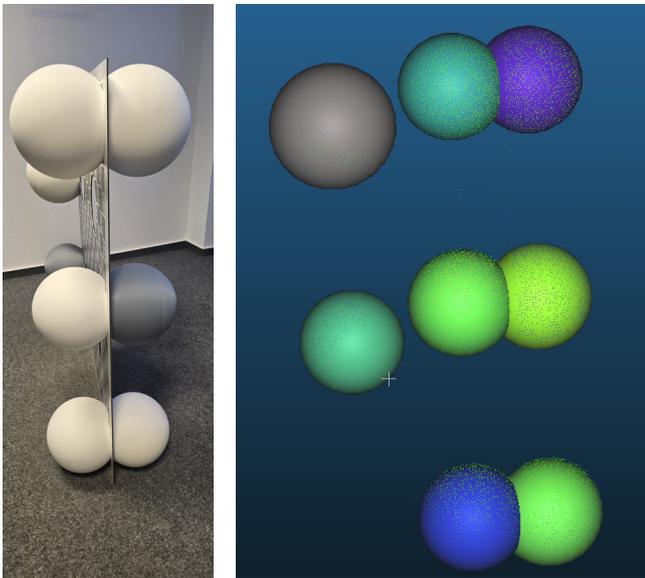
Die dargestellte Pipeline ermöglicht die robuste Erkennung der Target-Pose auch unter schwierigen Bedingungen des praktischen Einsatzes, wie partieller Verdeckung oder unterschiedlichen Aufnahmewinkeln unterschiedlicher Targetposen zum Kamerasetup.

Die Kalibrierung und Posen-Erkennung von bildgebenden 3D-Sensoren bietet bei senkrechter Ansicht zum Target, wie in Abbildung 3 dargestellt, im Vergleich zu einem konventionellen planaren Target zunächst keine signifikanten Vorteile. Im Gegenteil: Die Kugelelemente können sogar die Erfassung der planaren Geometrie erschweren, da sie relevante Bereiche verdecken. Die besonderen Stärken des entwickelten Targets zeigen sich jedoch bei der Posen-Erkennung auch unter extremen Winkeln. Abbildung 5 demonstriert dies anhand einer normalerweise ungebräuchlichen Seitenansicht auf das Target. Durch die beidseitige Anordnung der Kugeln am Target kann der ursprüngliche Modellparameter „Abstand zur Ebene“ durch den „Abstand zwischen den Kugelreihen“ ersetzt werden. Selbst die im Hintergrund befindlichen Kugeln, die aufgrund der perspektivischen Ansicht deutlich weniger Messpunkte aufweisen als die sechs Kugeln im Vordergrund, können von noch zuverlässig erkannt und in die Modellberechnung einbezogen werden. Diese Robustheit bei extremen Aufnahmewinkeln stellt einen entscheidenden Vorteil dieses Targets gegenüber herkömmlichen planaren Targets dar.

#### 4 ZUSAMMENFASSUNG UND AUSBLICK

Das vorgestellte modulare multimodale Target demonstriert mehrere entscheidende Vorteile für die praktische Anwendung in der geometrischen Kalibrierung von multimodalen Kamerasetups. Die Einbeziehung der Bedürfnisse verschiedener Sensormodalitäten in die Targetgestaltung resultiert in einem vielseitig nutzbaren, flexiblen Design, das sowohl für RGB-Kameras, Thermographiekameras als auch 3D-Sensoren geeignet ist. Das ChArUco-Muster erlaubt die präzise Kalibrierung optischer Sensoren, während die ebenen geometrischen Elemente und die optional beheizbaren Kugelemente die thermographische Bildgebung unterstützen. Gleichzeitig dienen die geometrischen 3D-Primitive der robusten Posen-Erkennung mit 3D-Modalitäten.

Ein besonderer Vorteil des Targetkonzeptes liegt in seiner modularen Bauweise, die ein schnelles Um- und Nachrüsten ermöglicht. Die einzelnen Komponenten können je nach Anwendungsfall flexibel montiert oder demontiert werden, was die Anpassung an verschiedene, wechselnde Kalibrierszenarien vereinfacht. Dabei gewährleistet die standardisierte Befestigung eine präzise und reproduzierbare Positionierung der Elemente.



**Abb. 5:** Multimodales 2D-/3D-Target mit 8 Kugeln in Seitenansicht senkrecht zur Ebenennormalen (links). Rechts dargestellt ist die 3D-rekonstruierte Szene aus Sicht der 3D-Modalität.

Durch Nutzung additiver Fertigungsverfahren für die 3D-Primitive können geometrische Elemente für spezifische Anforderungen aus Bildmodalität und Erfassungsraum kostengünstig und individuell realisiert

werden. So können beispielsweise Strukturen für die Integration von Heizelementen oder spezielle Oberflächeneigenschaften direkt im Druckprozess realisiert werden. Diese Flexibilität trägt wesentlich zur Praktikabilität und Weiterentwicklungsfähigkeit des Targets bei. Das beschriebene Verfahren und die algorithmische Umsetzung einer multimodalen, automatisierten Kalibrierpipeline berücksichtigt modenspezifische Besonderheiten des Targets und Targetabbildes. Im Beitrag wurden Vorschläge zur Gewährleistung hoher Genauigkeiten der Targetauswertung gemacht. Eine quantitative Validierung der Kalibriergenauigkeit mit realen Sensorsystemen stellt eine besondere Herausforderung dar, da ein direkter Vergleich mit konventionellen Targets und Verfahren durch die fehlende Ground-Truth-Referenz erschwert wird. Dies betrifft insbesondere 3D-Sensoren, bei denen die Schaffung reproduzierbarer Vergleichsbedingungen aufgrund systeminhärenter Messunsicherheiten und unterschiedlicher Messprinzipien nur eingeschränkt möglich ist. Um dennoch eine objektive Bewertung verschiedener Kalibrierverfahren zu ermöglichen, wird aktuell an einer Simulationsumgebung gearbeitet, die sowohl idealisierte als auch realistische Sensoreigenschaften verschiedener Modalitäten abbilden soll. Diese ermöglicht zukünftig systematische Vergleichsstudien unter kontrollierten Bedingungen.

Mit dem Ergebnis der Arbeiten stehen potentiellen Anwendern multimodaler Kamerasetups Hardware- und Software-Werkzeuge des ZBS e.V. (3D-EasyCalib[5]) zur Verfügung, welche die Erzeugung korrekt fusionierter multimodaler Bilddaten für weiterführende Bildanalysen ermöglichen.

#### REFERENZEN

- [1] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96, page 226–231. AAAI Press, 1996.
- [2] OpenCV. The OpenCV Reference Manual, 2.4.13.7 edition, April 2014.
- [3] Rudi Penne, Bart Ribbens, and Pedro Roios. An exact robust method to localize a known sphere by means of one image. International Journal of Computer Vision, 127, 08 2019.
- [4] Shinji Umeyama. Least-squares estimation of transformation parameters between two point patterns. IEEE Transactions on Pattern Analysis & Machine Intelligence, 13(04):376–380, 1991.
- [5] Darko Vehar, Rico Nestler, and Karl-Heinz Franke. 3d-easycalib™ - toolkit zur geometrischen kalibrierung von kameras und robotern. In 22. Anwendungsbezogener Workshop zur Erfassung, Modellierung, Verarbeitung und Auswertung von 3D-Daten, 3D-NordOst. GFal Gesellschaft zur Förderung angewandter Informatik e. V., December 2019.

# Depth Completion by Colorization for Solid-State LiDAR Sensors

Ying Yin Li, Jan Herschel, Roman Abayev, Dietrich Paulus and Anselm von Gladiss

Arbeitsgruppe Aktives Sehen, Institut für Computervisualistik,  
Universität Koblenz, Universitätsstraße 1, 56070 Koblenz

eMail: [liyinyinsimon@uni-koblenz.de](mailto:liyinyinsimon@uni-koblenz.de) / [www.uni-koblenz.de](http://www.uni-koblenz.de)

**Abstract.** This paper presents a depth completion method inspired by colorization, utilizing sparse depth data captured by solid-state LiDAR sensors, namely the Cepton Vista-P60 and the Livox Mid-100. Existing state-of-the-art methods for depth completion have primarily focused on Velodyne datasets featuring a nearly homogeneous sampling pattern in the horizontal direction. Our paper aims to achieve depth completion for inhomogeneous sampling patterns, making it applicable to a broader range of scenarios. To facilitate research in this domain, we introduce a comprehensive dataset consisting of consecutive sparse data measurements. The dataset is split into training, validation, and testing data. Ground truth data is generated by integrating multiple time frames. Our dataset allows for the evaluation of algorithms specifically tailored for solid-state LiDAR sensors. Both the implementation and the dataset are publicly available. The implementation is available at

<https://github.com/lyysl/Depth-Completion-through-Colorization>.

Dataset is available at

<https://www.kaggle.com/datasets/liyinyinsimon/solid-state-lidar>.

## 1 INTRODUCTION

LiDAR technology offers significant advantages for various applications due to its ability to measure distances accurately using emitted light. It may operate effectively at obscure vision conditions and provides rapid data acquisition. However, LiDAR data is often sparse and does not fully sample the field of view (FOV). This sparsity poses challenges because conventional computer vision algorithms require dense data to function effectively. Therefore, depth completion [1, 2], which estimates depth information for all pixels in a sparse depth image, is essential for leveraging the full potential of LiDAR in computer vision tasks. Among the various types of LiDAR [3, 4] sensors, solid-state LiDAR [5, 6] technology is rapidly gaining traction as a promising solution. These sensors offer a potentially higher update frequency and angular resolution when compared to traditional mechanical LiDAR sensors. Moreover, solid-state LiDARs have already demonstrated their value as cost-effective and lightweight alternatives to the traditional mechanical LiDAR sensors [7]. Furthermore, solid-state LiDAR technology enhances resistance to vibrations as it eliminates the need for rotating mechanical structures.

The process of depth completion usually involves projecting a sparse 3D point cloud onto a 2D space. Within these 2D sparse depth scenes, missing depth values need to be accurately estimated. The core objective of depth completion is to predict a dense depth image from the provided sparse depth data. For this study, we exclusively utilize sparse depth data as input.

To quantitatively assess the efficacy of depth completion results, ground truth data becomes essential. However, manually annotating ground truth data can be a laborious and challenging task. To address this, we obtain ground truth data for our solid-state LiDAR dataset by integrating multiple scans from the same scene, assuming the

scenes remain static during the data acquisition process. This approach ensures a reliable and comprehensive evaluation of our depth completion method.

We outline the main contributions of this paper as follows:

- We present a novel dataset comprising Cepton Vista-P60 and Livox Mid-100 LiDAR data, along with corresponding ground truth information. This dataset serves as a valuable resource for training and evaluating methods designed to process sensor data from the Cepton Vista-P60 and the Livox Mid-100 LiDAR sensor.
- We propose a depth completion method by solving a sparse matrix equation using only sparse depth data. Also, we evaluate the depth completion results using different time integrations.
- We evaluate various state-of-the-art approaches for depth completion on this dataset. We assess the proposed algorithm to gauge its potential for utilization with the Cepton Vista-P60 and Livox Mid-100 LiDAR sensors.

This work is structured as follows. Section 2 addresses related work on depth completion and solid-state LiDAR. Section 3 proposes a depth completion method using sparse depth images. Section 4 describes a new solid-state LiDAR dataset and presents experimental evaluations using different algorithms. Finally, Section 5 provides a summary of the paper.

## 2 RELATED WORK

In this section, we present an overview of the state-of-the-art methods for depth completion, focusing on their utilization of sparse depth data as inputs. With the growing prominence of LiDAR sensors in the field of autonomous driving [8], these algorithms have gained popularity and widespread attention. The evaluation of such algorithms has conventionally relied on benchmark datasets like KITTI [9, 10], which incorporate the use of the Velodyne HDL-64E LiDAR sensor. Our paper introduces a diverse set of algorithms, including both neural network-based and non-neural network-based approaches. These algorithms are employed to assess and compare their performance using the dataset we have provided. Additionally, the integration of solid-state LiDAR technology offers new opportunities for enhancing depth completion methods, which we explore in this study.

Compressed Sensing (CS) [11] addresses the reconstruction of sparse signals from limited

measurements, formulated as an optimization problem with an L1 norm constraint. The Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [12] effectively solves for the compressed sensing problem. Sparsity Invariant CNNs (SI-CNN) [10] utilize sparse convolution to process sparse depth maps, enabling the network to produce dense depth maps as outputs. The sparse convolution layer consists of three main operations: elementwise multiplication with a binary mask, convolution with the same padding, and normalization.

Fast Depth Completion on the CPU (IP-Basic) [13] is a non-neural networkbased approach. Classical image processing algorithms are employed for fast depth completion on the CPU. Techniques including depth inversion, dilation, hole closure, and blur operations achieve impressive speed and performance, outperforming some neural network-based methods.

Deep Convolutional Compressed Sensing (DCCS) [14] combines compressed sensing and deep learning for depth completion using sparse depth measurements. Built upon the Alternating Direction Neural Network (ADNN) [15], this framework leverages compressed sensing techniques for reconstructing sparse signals, applicable to depth map reconstruction from sparse data.

A probabilistic normalized convolutional neural network (pNCNN) has been proposed in [16]. The normalized convolution learns to operate only on confident points to avoid calculations from missing and noisy measurements. Additionally, the normalized convolutional network, within the proposed probabilistic framework, also provides a meaningful output uncertainty measure.

A non-learning depth completion method has been proposed in [17] with the underlying assumption that unobserved pixels share a common surface with the nearest observed value. The pipeline consists of two primary technical components: an outlier removal module and a surface geometry model.

ScaffNet [18] integrates Spatial Pyramid Pooling (SPP) [19] with an encoderdecoder architecture to enhance the densification of sparse point cloud inputs. By combining pooling outcomes with small and large kernel sizes, an equilibrium between data density and salient details is achieved.

A new simulated and real-life dataset for non-repetitive circular scanning (NRCS) LiDAR has been introduced in [20]. Additionally, they propose a neural network, ST-DepthNet, featuring a spatio-temporally extended U-Net architecture.

### 3 METHODS

Colorization using optimization [21] is a method that adds colors to grayscale images through optimization techniques. The core principle of the method involves minimizing the difference between the weighted averages of neighboring pixel values. These weighting functions are computed based on the squared difference and normalized correlation between two pixel values. By optimizing a cost function, the algorithm effectively converts grayscale images into their corresponding colored versions. In this contribution, the optimization techniques from colorization are repurposed to compute depth information.

#### 3.1 PROBLEM FORMULATION

The proposed method for depth completion, namely colorization, is modified as follows:

$$\min_x \left( \sum_x \left( U(x) - \sum_{s \in N(x)} w_{xs} U(s) \right) \right), \quad \text{s.t.} \quad U(x_{\text{mask}}) = d. \quad (2)$$

and

$$U(x_{\text{mask}}) = Gx = d. \quad (3)$$

The weight  $w_{xs}$  is the average over the total number of nearest neighbor pixels. By combining the optimization equation and the constraint, we obtain:

$$(B + G)x = Ax = d. \quad (4)$$

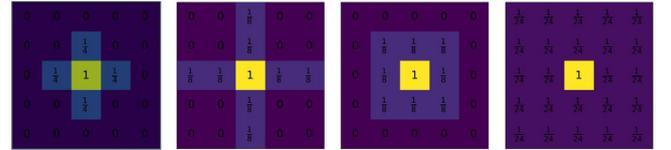
Here,  $A$  is a sparse matrix, and we aim to solve the sparse matrix system:

$$(B + G)x = Ax = d. \quad (5)$$

The matrix  $A$  is sparse because it includes only the nearest neighbors, making it efficient in terms of storage and computation. The sparsity depends on the number of nearest neighbors chosen, and the matrix size is determined by the dimensions of the sparse depth input image, which can be substantial. The matrix  $A$  is nearly-symmetric due to the nature of the weights, which represent a weighted average over the nearest neighbors. This nearly-symmetric property allows for efficient solving using algorithms [22, 23] designed for sparse matrix equations. The constraint of the equation ensures that the dense depth output at the original sparse input pixel locations does not change, preserving the integrity of the sparse input data. The objective function and the constraint ensure that the constructed matrix is weakly chained diagonally dominant (WCDD). Consequently, the matrix is non-singular [24], guaranteeing a unique solution. Unlike optimization or neural network-based methods, this approach does not require parameter tuning or training and consistently converges to the same

solution. Additionally, the proposed algorithm tackles linear algebra problems, which can be optimized for speed and efficiency by leveraging GPUs [25], enhancing its overall performance.

The proposed method relies on neighboring weights. We experimented with various sizes and shapes of nearest neighbors [13] to assess the impact of different weights on the results, which are depicted in Fig. 1.



**Fig. 1:** Various radius sizes and shapes of nearest neighbors. Radius from left to right: 1, 2, 1, 2. Shape from left to right: cross, cross, full, full. The sum of the weights of the neighbors equals the weight of the center pixel.

#### 3.3 GROUND TRUTH GENERATION

We approximate ground truth data using a process called „binning“. This involves combining consecutive sparse measurements taken over 50 frames of a static scene. The single frames differ due to the non-repetitive sampling trajectories of the sensors. For each pixel position, we take the median of the sampled values that correspond to non-zero entries in the images.

#### 3.4 EVALUATION METRICS

To assess the performance of the depth completion algorithms, we employ two widely utilized error metrics: mean absolute error (MAE) and root mean square error (RMSE), both reported in meters. Notably, the evaluation focuses solely on foreground objects, excluding distant elements, the maximum depth is set to 100 meters.

The MAE metric is computed as the average absolute difference between the ground truth and predicted values, as expressed by the formula:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (6)$$

where  $y_i$  denotes the ground truth pixel value at the  $i$ -th pixel position,  $\hat{y}_i$  corresponds to the pixel value derived from the outcome of different algorithms at the same pixel position, and  $n$  denotes the total count of valid pixels.

Similarly, the RMSE metric is calculated as the square root of the average squared difference between the ground truth and predicted values, employing the formula:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (7)$$

## 4 EXPERIMENTS

### 4.1 DATASET

The collected dataset serves the purpose of depth completion utilizing solid-state LiDAR sensor data, specifically data from the Cepton Vista-P60 and Livox Mid-100. Acquired in an outdoor setting, the dataset is in the form of rosbag files, with point cloud data transformed into 2D images saved as .png files. The image dimensions are standardized at 1250×375, and any depth values exceeding 100 meters are clipped. The dataset used in this study was collected in a campus environment. All scenes within the dataset are static, ensuring consistency when aggregating sparse data from consecutive time frames. The campus setting includes various outdoor scenarios typical of an urban environment, such as pathways, buildings, greenery, and other common campus features.

For the Cepton dataset, there are a total of 672 samples, distributed with 168 samples for each binning. Similarly, the Livox dataset comprises 620 samples, with 155 samples corresponding to each binning. The dataset is further categorized into training, validation, and testing sets. The Cepton dataset includes 125 training, 33 validation, and 10 testing samples for each binning. Meanwhile, the Livox dataset has 114 training, 31 validation, and 10 testing samples for each binning. Notably, the testing samples include RGB images as reference data.

An important aspect to highlight is that both the Cepton Vista-P60 and the Livox Mid-100 LiDAR devices follow non-repetitive trajectories [5, 26], allowing for integration over time, provided that the scenes are static.

### 4.2 EXPERIMENTAL RESULTS

In Table 1, the proposed method is evaluated with different neighbor sizes and shapes using Cepton and Livox datasets with a binning of 5. The results indicate that the configuration with a neighbor radius of 1 and a cross shape yields the best metric. This configuration will subsequently be used for later experiments.

**Cepton**

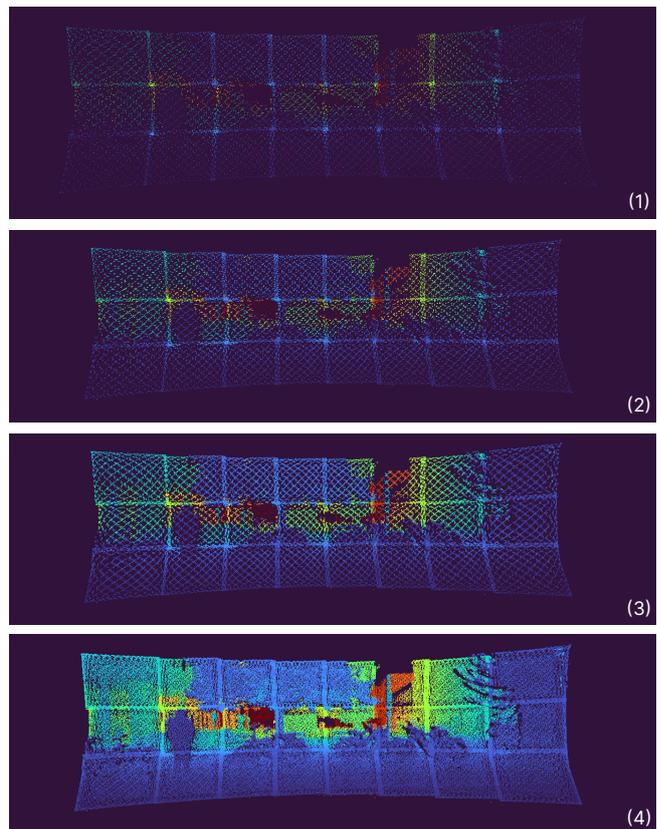
Radius	Shape	MAE (m)	RMSE (m)
1	Cross	<b>1.07</b>	<b>4.05</b>
1	Full	1.13	4.07
2	Cross	1.22	4.15
2	Full	1.32	4.24

**Livox**

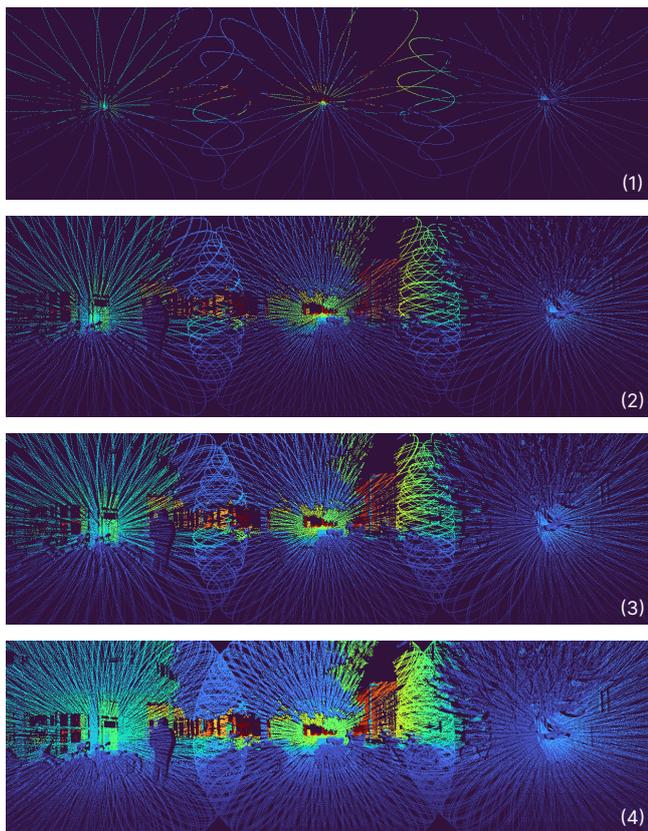
Radius	Shape	MAE (m)	RMSE (m)
1	Cross	<b>0.85</b>	<b>2.82</b>
1	Full	0.89	2.83
2	Cross	0.95	2.87
2	Full	1.03	2.96

**Tab. 1:** Effect of different radius sizes and shapes of the neighbors on the performance of Colorization on the Cepton and Livox datasets with a binning of 5 (binning time of 0:5 s and 168:7 ms, respectively).

To investigate how different sparsity levels in the input affect the depth completion results, we generate various binnings of sparse inputs, namely 1, 5, 10, and 25. The corresponding binning times for Cepton are approximately 0:1 s, 0:5 s, 1 s, and 2:5 s. For Livox, the binning times are approximately 33:7 ms, 168:7 ms, 337:4 ms, and 843:5 ms. The corresponding sparse inputs are shown in Fig. 2.



(a) Cepton Vista-P60



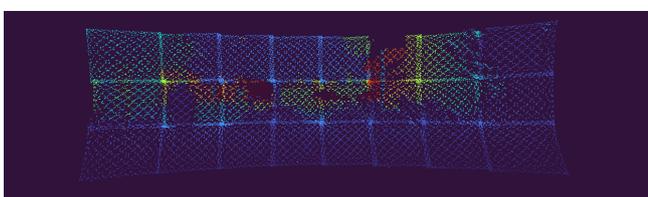
(b) Livox Mid-100

**Fig. 2:** The images display sparse depth inputs with varying degrees of sparsity for (a) Cepton Vista-P60 and (b) Livox Mid-100. In both (a) and (b), the images are arranged as follows: (1) shows binning value 1, (2) shows binning value 5, (3) shows binning value 10, and (4) shows binning value 25.

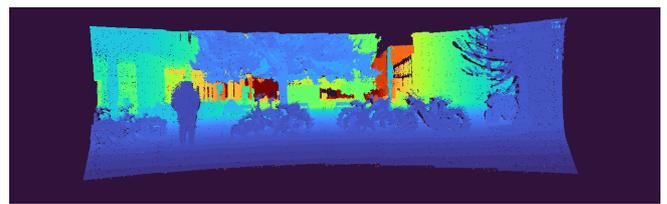
Fig. 3 and 4 display the reference RGB image, the sparse depth input image, and the ground truth for Cepton and Livox datasets. Additionally, the figures show sample predictions obtained using various algorithms with a binning of 5.



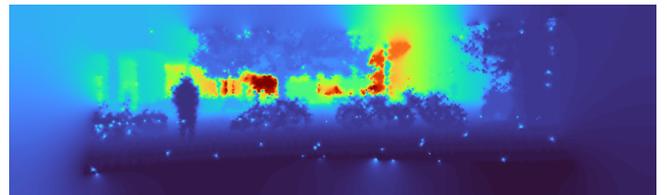
(a) Reference RGB image



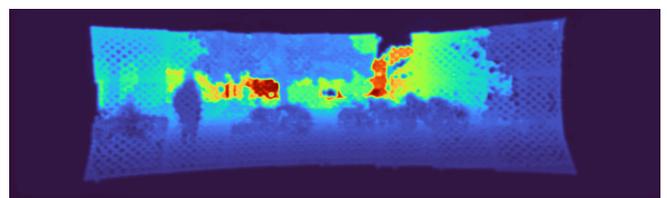
(b) Sparse depth input



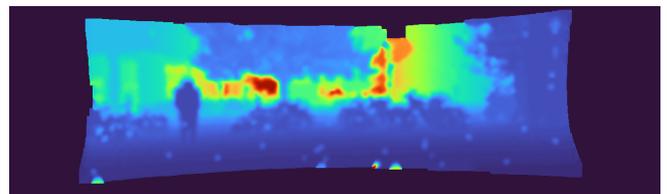
(c) Ground truth



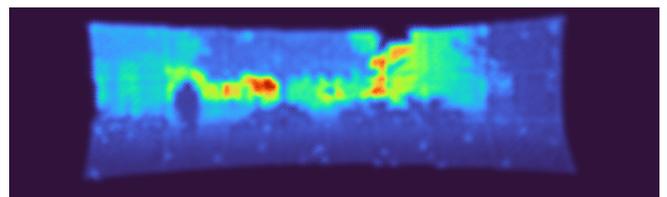
(d) Colorization



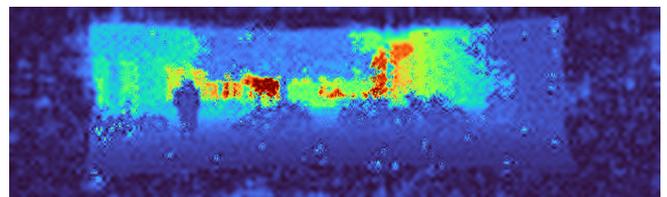
(e) ScaffNet [18]



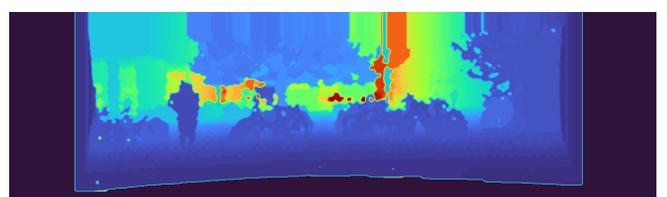
(f) SI-CNN [10]



(g) DCCS [14]



(h) CS [11]

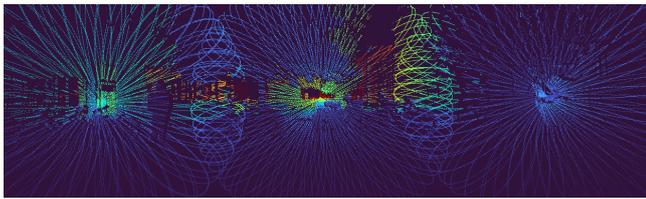


(i) IP-Basic [13]

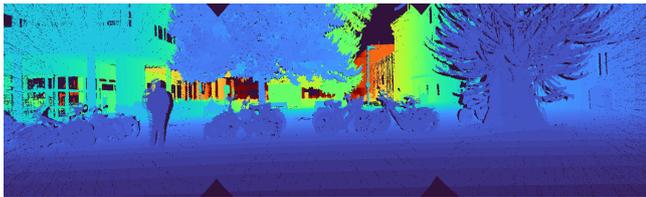
**Fig. 3:** Depth completion results for the Cepton dataset with a binning of 5 (binning time of 0:5 s). (a) Reference RGB image, (b) Sparse depth input, (c) Ground truth, and (d) to (i) Predictions from different methods.



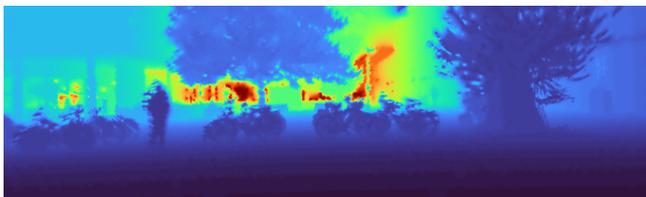
(a) Reference RGB image



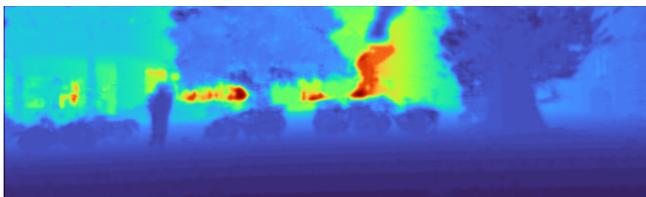
(b) Sparse depth input



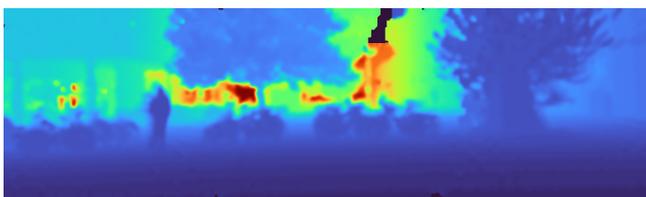
(c) Ground truth



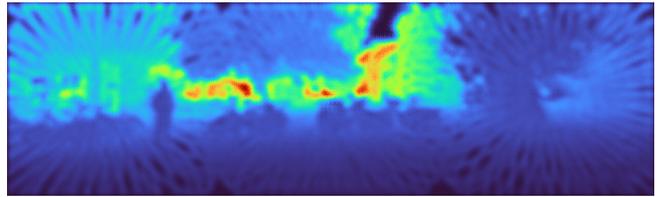
(d) Colorization



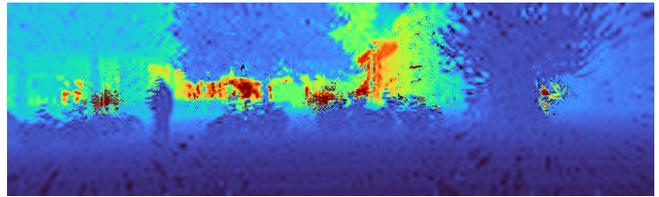
(e) ScaffNet [18]



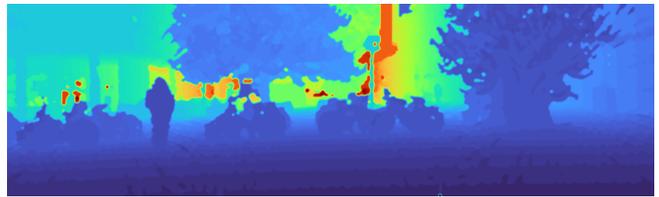
(f) SI-CNN [10]



(g) DCCS [14]



(h) CS [11]



(i) IP-Basic [13]

**Fig. 4:** Depth completion results for the Livox dataset with a binning of 5 (binningtime of 168:7 ms). (a) Reference RGB image, (b) Sparse depth input, (c) Ground truth, and (d) to (i) Predictions from different methods.

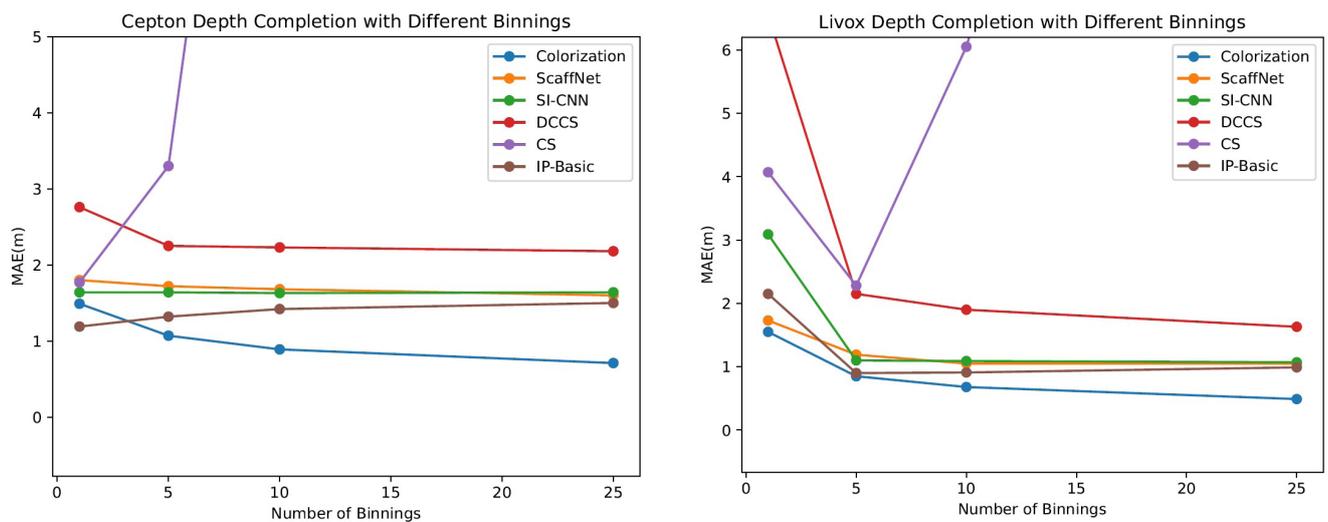
Table 2 presents the results of a comparison between different algorithms on the Cepton and Livox datasets, showing both mean absolute error (MAE) and root mean square error (RMSE). Additionally, the table shows the impact of different binnings in the sparse input. Notably, our proposed method exhibits the best performance in most cases compared to other algorithms.

	Cepton				Livox			
Binning time (ms)	100	500	1,000	2,500	33.6	168.7	337.4	843.5
Frames integrated	1	5	10	25	1	5	10	25
Method	MAE (m)				MAE (m)			
Colorization	1.49	<b>1.07</b>	<b>0.89</b>	<b>0.71</b>	<b>1.55</b>	<b>0.85</b>	<b>0.68</b>	<b>0.49</b>
ScaffNet [18]	1.80	1.72	1.68	1.60	1.73	1.19	1.05	1.05
SI-CNN [10]	1.64	1.64	1.63	1.64	3.09	1.10	1.09	1.07
DCCS [14]	2.76	2.25	2.23	2.18	6.56	2.15	1.90	1.63
CS [11]	1.77	3.30	13.97	30.40	4.07	2.28	6.05	23.55
IP-Basic [13]	<b>1.19</b>	1.32	1.42	1.50	2.15	0.90	0.91	0.99
Method	MAE (m)				MAE (m)			
Colorization	<b>4.64</b>	<b>4.05</b>	<b>3.72</b>	<b>3.16</b>	<b>3.97</b>	<b>2.82</b>	<b>2.50</b>	<b>2.02</b>
ScaffNet [18]	6.11	5.88	5.69	5.46	4.57	3.82	3.54	3.52
SI-CNN [10]	4.82	4.66	4.62	4.56	8.18	3.14	3.04	2.96
DCCS [14]	6.25	5.40	5.27	5.09	11.44	4.65	4.22	3.74
CS [11]	4.98	8.32	24.40	40.33	7.34	7.11	15.58	35.31
IP-Basic [13]	5.12	5.33	5.48	5.63	6.13	3.63	3.66	3.82

**Tab. 2:** Evaluation results on different methods on Cepton and Livox datasets.

The results obtained from Cepton and Livox datasets reveal distinct behaviors as the number of binnings in the input increases. In the case of the Cepton dataset, Colorization, ScaffNet and DCCS exhibit a decreasing trend in MAE as the number of binnings increases, as shown in Fig. 5. Conversely, the remaining methods demonstrate a general increase in MAE with higher

binnings. For the Livox dataset, a similar trend emerges where all methods experience a decrease in MAE as the binning value ascends from 1 to 5. Nevertheless, it is worth noting that some algorithms display a subsequent increase in MAE when confronted with larger numbers of binnings.



**Fig. 5:** Depth completion results with different binnings.

Further observations reveal that, across all algorithms for Livox depth completion, the most significant decrease in MAE occurs between binnings of 1 and 5. Beyond a binning of 5, the improvement in MAE becomes less pronounced or even results in increased MAE. In the case of Cepton depth completion, certain algorithms exhibit a decrease in MAE, while others experience an increase in MAE from a binning of 1 to 5. The depth completion results using the solidstate LiDAR dataset indicate that having a higher number of valid data points in the sparse input does not necessarily guarantee a more accurate prediction result.

A limitation of the proposed method is its processing time, as it requires solving a large matrix system. Table 3 shows that after optimization with GPU utilization, the running time is reduced by about 4 times. When the input image resolution is halved, the running time is further reduced by about 5 times.

Colorization	Input Image Shape	Time in s
Without GPU	1250×375	~6
With GPU	1250×375	~1.5
With GPU	625×188	~0.3

**Tab. 3:** Time optimization for processing a single sample for Colorization.

Another limitation of the method is its reliance on nearest neighbors with equal weights, which does not account for shape edges across different surfaces. A potential future development is to incorporate RGB images as guidance to dynamically adjust the weights.

## 5 CONCLUSION

In this paper, we propose a depth completion method that addresses the problem through colorization. Additionally, we provide a new benchmark for solid-state LiDAR sensors, specifically the Cepton Vista-P60 and the Livox Mid-100. We observe the potential of utilizing solid-state LiDAR sensors for depth completion tasks and other computer vision applications.

## REFERENCES

[1] Junjie Hu, Chenyu Bao, Mete Ozay, Chenyou Fan, Qing Gao, Honghai Liu, and Tin Lun Lam. Deep depth completion from extremely sparse data: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[2] Muhammad Ahmed Ullah Khan, Danish Nazir, Alain Pagani, Hamam Mokayed, Marcus Liwicki, Didier Stricker, and Muhammad Zeshan Afzal. A comprehensive survey of depth completion approaches. *Sensors*, 22(18):6969, 2022.

[3] Jordan Britt, Christopher Rose, and David M Bevil. A comparative study of lidar and camera-based lane departure warning systems. In *Proceedings of the 24th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2011)*, pages 324–332, 2011.

[4] Dinh Van Nam and Kim Gon-Woo. Solid-state lidar based-slam: A concise review and application. In *2021 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 302–305. IEEE, 2021.

[5] Roman Abayev, Anselm von Gladiss, and Dietrich Paulus. Eigenschaften verschiedener solid-state-lidar-systeme. In St. Rothstock, B. Hohnhäuser, D. Krueger, and M. Pochanke, editors, *3D-NordOst 2021*, pages 19–27, Berlin, Germany, 2021. Gesellschaft zur Förderung angewandter Informatik e.V. (GFal). ISBN 978-3-942709-27-9.

[6] Nanxi Li, Chong Pei Ho, Jin Xue, Leh Woon Lim, Guanyu Chen, Yuan Hsing Fu, and Lennon Yao Ting Lee. A progress review on solid-state lidar and nanophotonics-based lidar sensors. *Laser & Photonics Reviews*, 16(11):2100511, 2022.

[7] Han Wang, Chen Wang, and Lihua Xie. Lightweight 3-d localization and mapping for solid-state lidar. *IEEE Robotics and Automation Letters*, 6(2): 1801–1807, 2021.

[8] Michael Gschwandtner. Support framework for obstacle detection on autonomous trains. na, 2012.

[9] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.

[10] Jonas Uhrig, Nick Schneider, Lukas Schneider, Uwe Franke, Thomas Brox, and Andreas Geiger. Sparsity invariant cnns. In *2017 international conference on 3D Vision (3DV)*, pages 11–20. IEEE, 2017.

[11] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[12] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2 (1):183–202, 2009.

[13] Jason Ku, Ali Harakeh, and Steven L Waslander. In defense of classical image processing: Fast depth completion on the cpu. In *2018 15th Conference on Computer and Robot Vision (CRV)*, pages 16–22. IEEE, 2018.

[14] Nathaniel Chodosh, ChaoyangWang, and Simon Lucey. Deep convolutional compressed sensing for lidar depth completion. In *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision*, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part I 14, pages 499–513. Springer, 2019.

[15] Calvin Murdock, MingFang Chang, and Simon Lucey. Deep component analysis via alternating direction neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 820–836, 2018.

- [16] Abdelrahman Eldesokey, Michael Felsberg, Karl Holmquist, and Michael Persson. Uncertainty-aware cnns for depth completion: Uncertainty from beginning to end. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12014–12023, 2020.
- [17] Yiming Zhao, Lin Bai, Ziming Zhang, and Xinming Huang. A surface geometry model for lidar depth completion. IEEE Robotics and Automation Letters, 6(3):4457–4464, 2021.
- [18] Alex Wong, Safa Cicek, and Stefano Soatto. Learning topology from synthetic data for unsupervised depth completion. IEEE Robotics and Automation Letters, 6(2):1495–1502, 2021.
- [19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(9):1904–1916, 2015.
- [20] Örkény Zováthi, Balázs Pálffy, Zsolt Jankó, and Csaba Benedek. Stdepthnet: A spatio-temporal deep network for depth completion using a single non-repetitive circular scanning lidar. IEEE Robotics and Automation Letters, 2023.
- [21] Anat Levin, Dani Lischinski, and Yair Weiss. Colorization using optimization. In ACM SIGGRAPH 2004 Papers, pages 689–694. 2004.
- [22] Timothy A Davis. Algorithm 832: Umfpack v4. 3—an unsymmetricpattern multifrontal method. ACM Transactions on Mathematical Software (TOMS), 30(2):196–199, 2004.
- [23] Xiaoye S Li. An overview of superlu: Algorithms, implementation, and user interface. ACM Transactions on Mathematical Software (TOMS), 31(3): 302–325, 2005.
- [24] PN Shivakumar and Kim Ho Chew. A sufficient condition for nonvanishing of determinants. Proceedings of the American mathematical society, pages 63–66, 1974.
- [25] Siu Kwan Lam, Antoine Pitrou, and Stanley Seibert. Numba: A llvm-based python jit compiler. In Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, pages 1–6, 2015.
- [26] Mark O Mints, Roman Abayev, Nick Theisen, Dietrich Paulus, and Anselm von Gladiss. Online calibration of extrinsic parameters for solid-state lidar systems. Sensors, 24(7):2155, 2024.

# Digitalisierung des Aufmaßprozesses von Treppenläufen und -podesten

Erik Magerl<sup>1</sup>, Tina Ambrosat<sup>1</sup>, Dominik Hack<sup>1</sup>, Michael Geist<sup>1</sup>, Konrad Jagusch<sup>1</sup>, Jan Sender<sup>1</sup>, Julian Jakubzyk<sup>2</sup>

<sup>1</sup>Fraunhofer-Institut für Großstrukturen in der Produktionstechnik IGP,  
Albert-Einstein-Straße 30, D-18059 Rostock  
E-Mail: vorname.name@igp.fraunhofer.de  
www.igp.fraunhofer.de

<sup>2</sup>Metallbau Ott GmbH, Eikboomstraße 12, 18209 Bad Doberan  
E-Mail: jakubzyk@metallbau-ott.de  
www.metallbau-ott.de

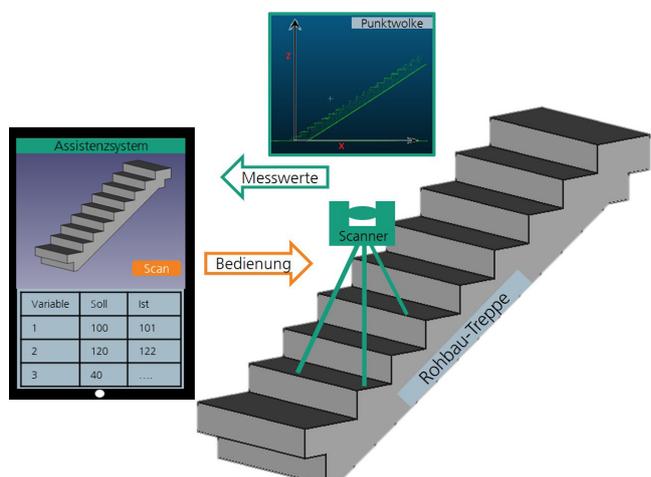
**Zusammenfassung:** Die Erfassung von Aufmaßen von Treppenläufen und podesten ist ein notwendiger Schritt in der Vorfertigung von metallischen Treppengeländern. Derzeit erfolgt dies in einem manuellen Prozess mit einfachen Messmitteln oder terrestrischem Laserscanner, ohne unmittelbare Datenauswertung und -kontrolle. Fehler werden somit nicht erkannt und führen zu hohen Kosten und Aufwänden. Um dem entgegenzuwirken, wird in diesem Beitrag ein Assistenzsystem in Kombination mit Messsystematik zur Ermittlung der Aufmaße und Erstellung der 3D-Modelle vorgestellt. Dieses beinhaltet eine angeleitete Vermessung mit einem terrestrischen Laserscanner und einer Vor-Ort Auswertung, sodass Fehler im Rohbauobjekt erkannt und korrigiert werden können. Insgesamt wird eine digitale Datenhaltung ermöglicht und die Aufmaßermittlung erfolgt objektiv und reproduzierbar.

**Keywords:** Treppenaufmaß, TLS, Assistenzsystem, Digitalisierung

## 1 EINFÜHRUNG

Im Bauwesen sind aufgrund von hohen Rohbautoleranzen präzise Aufmaße der tatsächlichen Geometrie des Rohbaus unverzichtbar, um eine präzise Vorfertigung der Bauteile sicherzustellen [1]. Treppengeländer müssen sowohl ästhetischen als auch sicherheitstechnischen Anforderungen gerecht werden [2] und unterliegen in Deutschland strengen Regularien, die eine exakte Anpassung an die baulichen Gegebenheiten erfordern [3]. So ist beispielsweise die Mindesthöhe von Geländern, abhängig von der Gebäudetypologie, fest vorgeschrieben. Entsprechend gilt es, die Höhe des Fertigfußbodens im Rohbau zu erfassen und das Geländer daran zu planen.

Die Einpassung von Treppengeländern, die einen hohen Vorfertigungsgrad bei Fertigungstoleranzen im Millimeterbereich haben, gestaltet sich herausfordernd, da bereits geringe Abweichungen signifikante Nacharbeiten nach sich ziehen können.



**Abb. 1:** Schema der Verknüpfung des Messsystems und der parametrischen Konstruktion mit dem Assistenzsystem

Neben der ursprünglichen Methode zur Geometrieerfassung mit einfachen Messmitteln wie Gliedermetermessstab und Rotationslaser hat sich der Einsatz von terrestrischen Laserscannern etabliert. Mit diesen lassen sich die Strukturen im Rohbau schnell und mit hoher Genauigkeit in Form von 3D-Punktwolken erfassen [4]. Allerdings ist die Auswertung dieser Daten sehr zeitaufwändig und nur von qualifiziertem Personal durchführbar. Nicht-erfasste Maße, z.B. durch Abschattungen, müssen nachträglich aufwändig ergänzt werden.

In diesem Beitrag wird ein Assistenzsystem vorgestellt, das das Personal durch den Prozess der Aufmaßermittlung mittels terrestrischem Laserscanner führt (siehe Abbildung 1). Dabei wird die Vermessung angeleitet und die Aufmaße werden direkt automatisiert aus den Messdaten abgeleitet, sodass die Ergebnisse vor Ort vorliegen. Nicht-erfasste Maße können mit einfachen Messmitteln erfasst und ergänzt werden. Somit wird der Prozess robuster, erheblich beschleunigt und der Personalaufwand wird reduziert.

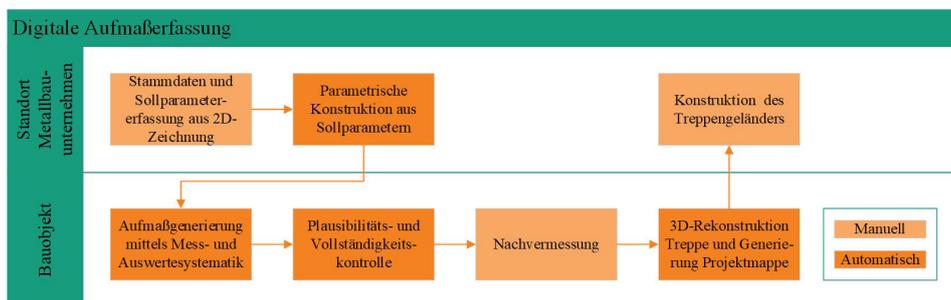
## 2 PROZESS ZUR DIGITALEN AUFMASSERFASSUNG

Der digitale Prozess umfasst sowohl die Vorbereitungsphase am Unternehmenssitz als auch die Erfassung und Auswertung vor Ort auf der Baustelle (siehe Abbildung 2). Das Assistenzsystem führt die Anwendenden durch den Prozess und erleichtert Projektierungs- und Dokumentationsaufgaben. Die projektspezifischen Stammdaten zu den Treppenhäusern, Etagen sowie zugehörigen Parametern zu den Treppenläufen und -podesten auf Basis von 2D-Rohbau-Zeichnungen werden zu Beginn festgelegt. Anhand dieser wird mittels parametrisierter Konstruktion ein Soll-Modell des Treppenhauses erstellt.

Für die Aufmaßgenerierung leitet das Assistenzsystem systematisch durch den Messprozess und ermöglicht eine automatisierte Ableitung der notwendigen Aufmaße. Der eingesetzte terrestrische Laserscanner erfasst eine Punktwolke der Rohbaugeometrie, anhand der die tatsächlichen Aufmaße abgeleitet werden. Mit dem automatischen Abgleich der Aufmaße mit den Solldaten wird unmittelbar vor Ort eine Plausibilitäts- und Vollständigkeitsprüfung durchgeführt. Dies erfolgt in einem mehrstufigen Validierungsprozess, bei dem die

Daten in drei Kategorien eingeteilt werden: plausible, nicht-plausible und fehlende Aufmaße. Aufmaße innerhalb der Bautoleranz gelten als plausibel, wohingegen Messergebnisse außerhalb als nicht-plausibel angesehen werden. Die Vollständigkeitsüberprüfung identifiziert die fehlenden Aufmaße, die durch das Messsystem und die nachgelagerte Auswertung nicht erfasst wurden. Im Assistenzsystem werden die drei Kategorien visuell hervorgehoben, sodass die Mitarbeitenden alle Daten zur Korrektur und Nachprüfung effizient identifizieren können. Die manuelle Nachvermessung erfolgt mit traditionelle Messmitteln, wie dem Gliedermetermaßstab, und die Ergebnisse werden ebenfalls einer Plausibilitätsprüfung unterzogen. Ziel ist es, sicherzustellen, dass die erfassten Aufmaße eine exakte Grundlage für die präzise und unmittelbare Erstellung der 3D-Modelle des tatsächlichen Treppenhauses bilden, anhand denen das Treppengeländer geplant werden kann.

Im Folgenden werden das Vorgehen zur Vermessung mit einem terrestrischen Laserscanner sowie die automatisierte Aufmaßgenerierung aus den Punktwolken beschrieben.

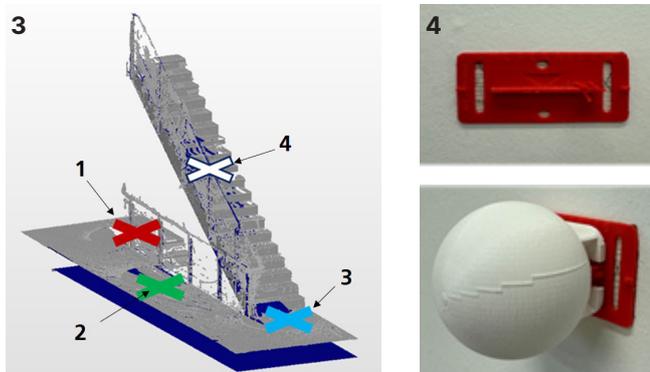


**Abb. 2:** Innovierter digitaler Prozesse nach Verwendungsstandorten des Assistenzsystems

### 2.1 MESSSYSTEMATIK

Terrestrische Laserscanner wurden für viele verschiedene Anwendungen konzipiert und verfügen entsprechend über eine Vielfalt an Messeinstellungen. Für den Anwendungsfall Treppengeometrie im Rohbau hat sich gezeigt, dass eine niedrige Auflösung bei mittleren Qualitätseinstellungen einen geeigneten Kompromiss zwischen Geometrieerfassung, Scandauer und Datenmenge darstellt. Für eine effiziente Vermessung mit einem terrestrischen Laserscanner gilt es, jeden Standpunkt relativ zur Treppe so zu wählen, dass die relevante Geometrie ausreichend aufgenommen werden kann. Zeitgleich müssen sich die Punktwolken der Einzelscans genügend überlappen, sodass eine Registrierung dieser zueinander möglich ist. Jedoch ist

die Gesamtzahl der Scannerstandpunkte aus Zeit- und Speicherkapazitätsgründen gering zu halten.



**Abb. 3:** Scanner-Standpunktplanung zur Treppemessung mit vier Standpunkten (als Kreuz dargestellt)

**Abb. 4:** Standard-Markierung eines Meterriss (oben) mit 3D gedruckter Kugeladapter (unten)

In Abbildung 3 ist die Standortplanung für die Vermessung einer Etage mit vier Standpunkten dargestellt. Die Einzelscanregistrierung erfolgt automatisiert, flächenbasiert über die Wand- Decken- und Bodenoberflächen.

Meterrissmarkierungen (siehe Abbildung 4) sind essenziell zur etagenweise Höhenreferenz, jedoch sind diese in terrestrischen Scandaten anhand des groben Einzelpunktrasters schwer zu erkennen. Zur Erfassung dient hier ein Ansteckadapter mit einer Regelgeometrie als Ziel. Diese Sphären sind eindeutig von anderen Strukturen in Treppenhäusern zu unterscheiden und liefern nach Erkennung einen berechneten Kugelmittelpunkt, wodurch die Höhe des Meterrissmarker ermittelt werden kann. Die erkannten Höhenmarkierungen dienen weiterhin zur etagenweisen Ausrichtung der Scandaten zur Referenzhöhe.

## 2.2 AUSWERTESTRATEGIE

Anhand der via terrestrischem Laserscanner erfassten Punktwolken des Treppenhauses gilt es, die Aufmaße für die Vorfertigung des Treppengeländers abzuleiten. Dazu wird die Information, welche Aufmaße gesucht sind, benötigt. Diese wird aus dem dreidimensionalen Sollmodell der Treppen und -podeste abgeleitet.

Zur Generierung eines digitalen Sollmodells eines Treppenhauses wurden parametrische Beschreibungen für Treppenarten (siehe Abbildung 5), -auflager und -podeste entwickelt. Diese enthalten, im Falle der Treppen, Parameter wie Stufenanzahl, Steigung (Stufenhöhe), Auftritt (Stufentiefe) und Treppenbreite. Zur

Geländerplanung werden neben der Treppengeometrie die der Auflager (Übergang zwischen Treppe und Etage) und dem Treppenauge benötigt. Da eine Etage in der Regel eben ist, umfasst deren Parametrisierung neben der Deckenstärke die aneinandergereihten Strecken und Winkel, die das Treppenaug geometrisch beschreiben. Die jeweiligen Parameter des Treppenhauses sind den Architekturzeichnungen zu entnehmen, sodass das Sollmodell automatisiert generiert werden kann. Zeitgleich werden die gesuchten Aufmaße hinterlegt.



**Abb. 5:** Parametrisierte Treppenarten, v.l.n.r.: gerade Treppe ohne bzw. mit Podest, Podesttreppe mit zwei 90°-Wendungen, Podesttreppe mit 180°-Wendung [5]

Die Aufmaße werden treppen- und etagenweise aus den Punktwolken abgeleitet. Es gilt es zunächst, die Treppe aus der Punktwolke zu segmentieren und an der Wange (offene Seite der Treppe) auszurichten. Danach sind die Elemente wie Stufen und Podeste zu identifizieren und klassifizieren. Beide Schritte basieren auf einer Ebenensegmentierung mittels RANSAC. Die Klassifizierung erfolgt mithilfe der vorhandenen Sollwerte wissensbasiert anhand der Richtung des Normalenvektors und der Lage des Schwerpunkts im Raum.

Die gesuchten Aufmaße werden indirekt abgeleitet, indem jeweils zwei benachbarte Stufenebenen mit der Ebene, die die Treppenwange repräsentiert, zu einem Punkt verschnitten werden. Die 3D-Abstände der Schnittpunkte bilden die Aufmaße.

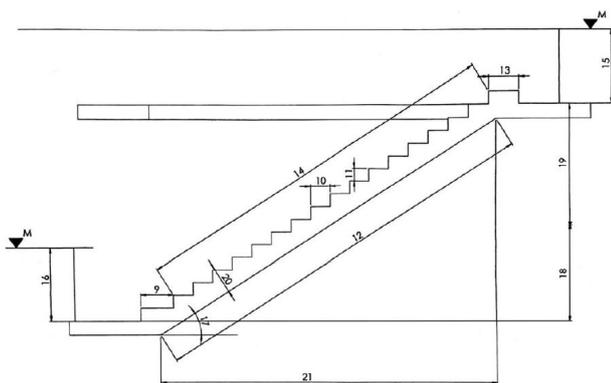
## 3 ANALYSE DES PROZESSES

Die Aufmaßermittlung von Treppenhäusern im Rohbau mit einem Assistenzsystem sieht den Einsatz von einem terrestrischen Laserscanner mit automatisierter Datenauswertung vor. Im Folgenden werden die experimentell erreichte Ergebnislänge mit dem Messsystem sowie die Vorteile der Verwendung des Assistenzsystems betrachtet.

### 3.1 ERGEBNISGÜTE DES LASERSCANNERS IN DER AUFMASSERFASSUNG

Am Beispiel einer geraden Treppe im Rohbau (siehe Abbildung 6) wurden die Aufmaße (ohne die

Meterrisshöhe in der nächsten Etage, Maß 15) mit der traditionellen Messmethode (Gliedermetermessstab) direkt sowie einem terrestrischen Laserscanner (Z+F Imager 5010) indirekt ermittelt. Als Referenz wurde die Treppengeometrie zusätzlich mit einem höhergenauen Messsystem (Leica Absolute Tracker AT960-LR) aufgenommen und die Aufmaße wurden ebenfalls indirekt als 3D-Abstände der Schnittpunkte von Treppenebenen ermittelt.



**Abb. 6:** Maßübersicht für Treppenaufmaß

Jedem Aufmaß wurde eine Maßanforderung basierend auf den Ansprüchen der Vorfertigung zugeordnet und jedes Messergebnis entsprechend des Referenzwertes und der Maßanforderung bewertet. Die Ergebnisse sind in der Tabelle 1 farbcodiert dargestellt. Es zeigt sich, dass die Aufmaßermittlungen mit dem terrestrischen Laserscanner im Gegensatz zum Gliedermetermessstab den Maßanforderungen entsprechen. Die hohe Abweichung im Maß 18+19 ist nicht auf das Messmittel, sondern auf einen groben Fehler in der Dokumentation zurückzuführen.

Maße	Lasertracker	Laserscanner	Gliedermetermaßstab	Maßanforderung
9	460,9	461,4	459	± 2 mm
10	280,1	280,8	280	± 2 mm
11	176,8	177,1	177	± 2 mm
12	5627,1	5623,7	5620	± 5 mm
13	420,6	422,6	420	± 2 mm
14	5294,7	5295,1	5300	± 5 mm
16	1203,6	1202,9	1205	± 5 mm
18 + 19	3019,2	3019,7	3201	± 5 mm
20	368,8	369,6	370	± 2 mm
21	4760,4	4756,4	4746	± 5 mm

**Tab. 1:** Vergleich ausgewählter erfasster Aufmaße (Maße in mm)

### 3.2 VORTEILE DES ASSISTENZSYSTEMS

Das zentrale Ergebnis der Aufmaßermittlung mit dem Assistenzsystem ist ein 3D-Ist-Modell der Treppengeometrie, das direkt im Anschluss an die Vermessung generiert wird. Dieses ist aufgrund der Qualität der angeleitete erfassten Daten und der integrierten Datenüberprüfung konsistent und bildet eine valide Datengrundlage für die Vorfertigung von Treppengeländern. Da nur ein Mitarbeitender für den innovierten Prozess und kein manuelles Eingreifen nach Abschluss der Vermessung notwendig ist, wird mit dem Assistenzsystem signifikant Personal- und Zeitaufwand eingespart.

Das Assistenzsystem führt den Anwendenden durch den gesamten Prozess der Vermessung und übernimmt die Aufmaßgenerierung. Eine messtechnische Vorqualifikation anwenderseitig ist somit nicht notwendig. Die Generierung der Aufmaße vor Ort auf der Baustelle gewährleisten die vollständige Datenaufnahme und verhindern somit erneute Anfahrts- und Vermessungskosten. Zudem bietet das Assistenzsystem eine umfangreiche digitale Datenhaltung. So werden Rohdaten, Zwischenergebnisse sowie Notizen und Fotos dokumentiert, sodass die projektweite Bearbeitung transparent ist und unmittelbar zur Verfügung steht.

Im Vergleich zur traditionellen Messmethodik werden die ergonomischen Bedingungen für die Mitarbeitenden verbessert, da die Notwendigkeit für manuelle und körperlich belastende Tätigkeiten reduziert wird.

### 4 FAZIT UND AUSBLICK

Die Aufmaßermittlung mit terrestrischen Laserscanner und dem vorgestellten Assistenzsystem, das durch den ganzen Prozess von der Projektierung bis zur Ableitung der Aufmaße führt, liefert ein konsistentes 3D-Modell des erfassten Treppenhauses für die anschließende Geländerkonstruktion. Der Prozess ist robuster, reproduzierbar und effizienter im Vergleich zur Vermessung ohne Anleitung und automatisierte Auswertung.

Das Assistenzsystem befindet sich noch in der Entwicklung. Insofern stehen die Analyse der Eignung in der Praxis und die Integration des Systems in den Arbeitsalltag noch aus.

Mit der Vermessung des Rohbaus mittels terrestrischem Laserscanner werden neben der für die Aufmaßermittlung von Treppenläufen notwendigen auch redundante Geometrien erfasst. Durch Bereitstellung dieser Daten

für den Bauherren und anderer Gewerke kann hier eine Wertschöpfung erreicht werden.

Derzeit beschränkt sich das System auf gerade Treppen mit und ohne Podest sowie mit 90°- und 180°-Winkel. Gewendelte Treppen werden bisher nicht berücksichtigt, sodass das System nach nachgewiesener Eignung um diese erweitert werden sollte. Die Aufnahmemöglichkeit meterrissbezogener Maße sollte ebenfalls weiterentwickelt werden, indem ein weiterer, höhenverstellbarer Adapter konstruiert wird. Dieser soll auf Baustellen nicht nur bei den genormten Meterrissmarkern, sondern auch bei angezeichneten Meterstrichen angewandt werden können. Ebenfalls ist eine Skalierung der Anwendung des Aufmaßprozesses auf andere Bereiche wie Außentreppengeländer, Balkonumrandungen und Fassaden denkbar.

## FÖRDERHINWEIS

Das Verbundvorhaben „DAVoM: Digitalisierung des Aufmaßprozesses für die Vorfertigung im Metallbau“ wird kofinanziert von der Europäischen Union aus dem Europäischen Fonds für regionale Entwicklung in der Förderperiode 2021 bis 2027.

## LITERATUR

[1] R. Ertl, Toleranzen im Hochbau: Kommentar zur DIN 18202 Maßgerechtes Bauen mit Abweichungen : mit 629 Abbildungen und 123 Tabellen, 4. Aufl. Köln: RM Rudolf Müller, 2021.

[2] TÜV SÜD AG, Geländer - Normen, Maße, Vorschriften: Gesetzliche Vorschriften für Treppengeländer. [Online]. Verfügbar unter: <https://www.tuvsud.com/de-de/indust-re/bautechnik-info/allgemeine-bautechnik/treppen-bruestungen-gelaender/gelaender>.

[3] Gebäudetreppen: Begriffe, Messregeln, Hauptmaße, DIN 18065:2015-03, DIN Deutsches Institut für Normung e. V - DIN-Normenausschuss Bauwesen (NABau), Berlin, Mrz. 2015.

[4] C. Wu, Y. Yuan, Y. Tang und B. Tian, „Application of Terrestrial Laser Scanning (TLS) in the Architecture, Engineering and Construction (AEC) Industry“ (eng), Sensors, Jg. 22, Nr. 1, S. 265, 2021, doi: 10.3390/s22010265.

[5] D. Teixeira, Grundriss Butler: Treppenarten. [Online]. Verfügbar unter: <https://www.grundriss-butler.de/die-verschiedenen-treppenarten> (Zugriff am: 25. Oktober 2024).

# Lidar-based Missing Plant Detection in Steep Vineyards

Nick Theisen, Philipp Räder, Daniel Müller and Dietrich Paulus

Active Vision Group, Universität Koblenz, Universitätsstraße 1, 56070 Koblenz

{nicktheisen,philraed,muellerd,paulus}@uni-koblenz.de

**Abstract.** In this paper we propose a method for missing and dead plant estimation given a 3D pointcloud of vineyards. Following existing approaches we estimate the plant canopy volume, i.e. the volume of the plant foliage, and perform classification through thresholding. We propose two additional thresholding methods and show their effectiveness by comparing them to an average volume thresholding baseline. Furthermore, we demonstrate our approach on pointclouds acquired with different sensors (lidar, solid-state lidar and camera) and platforms (UAV and UGV).

**Keywords.** Viticulture · Leaf-volume estimation · Missing plant detection.

## 1 INTRODUCTION

The cultural landscape of the German Moselle valley is characterized by viticulture in steep-slope vineyards. Due to the challenging topographic and environmental conditions inspection is still mostly done manually. As manual work is expensive and the topographic and environmental conditions in vineyards pose a high risk for workers, reducing the amount of time people spend in vineyards needs to be reduced.

Research towards this goal focuses on robotic platforms for automation, e. g. [4]. Another way to reduce this time is automatic sensor-based monitoring. However, without according extraction of high-level information the benefit to wine makers is limited. Therefore we propose a method for missing plant detection given a 3D pointcloud of a vineyard and a list of plant coordinates. Manual inspection and documentation takes multiple person hours per vineyard and is dangerous. In contrast, our approach requires only few minutes excluding the time of data collection. An overview over the approach is shown in Fig. 1.

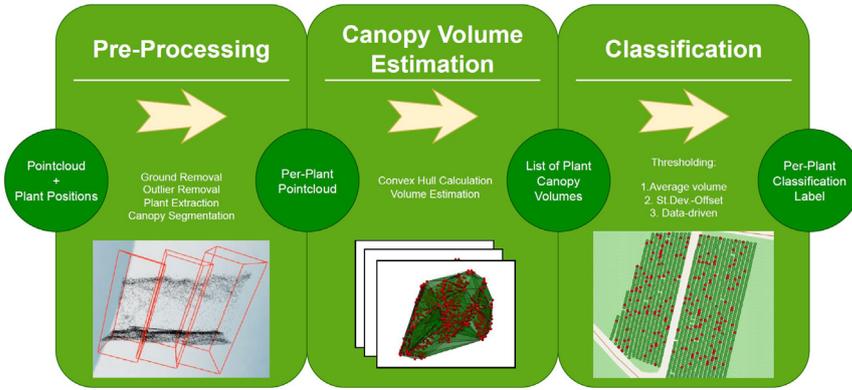
Analysis of pointclouds in agri- and viticultural settings is an active research field. Jurado et al. [3] detect grapevine trunks and derive missing plant locations. Di Gennaro and Matese [1] use alpha-shapes to estimate plant canopy volume, i.e. the volume of plant foliage, and perform thresholding to detect missing plants. Both use UAV-based 3D pointclouds, which is not optimal for capturing the plant geometries. Pagliari et al. [6] showed that pointclouds collected from a ground-vehicle are better suited for analysis of plant geometry.

Our approach is similar to [1], but instead of alpha-shapes we use the convex hull of canopy points to estimate its volume. This is faster and does not require tuning of the alpha parameter. Further, we compare three different

thresholding strategies for classification. We evaluate our approach on multiple datasets acquired with different sensors and platforms (UAV, UGV).

## 2 MISSING PLANT DETECTION

The proposed approach consists of three consecutive steps. First, the input pointcloud is pre-processed to remove outliers and segment the canopy points of each individual plant. Second, we estimate the canopy volume by fitting a convex hull. Third, we perform thresholding on the estimated volumes to identify missing plants. An overview is shown in Fig. 1.



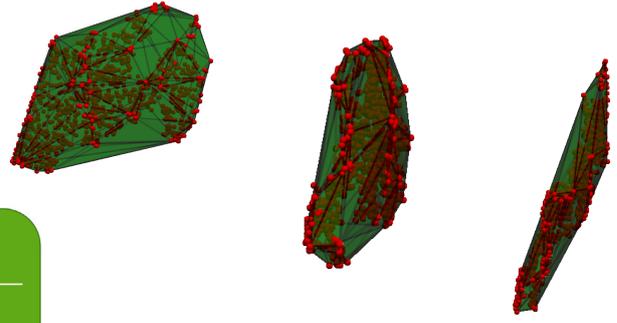
**Fig. 1:** Overview over missing plant detection approach. Input are geo-referenced 3D maps of a vineyard and the geographical coordinates of the wine plants. Processing steps have a dark green background, while (intermediate) results have a brighter green background.

**Pre-processing** The input to our approach is a 3D pointcloud showing the whole vineyard and a list of plant coordinates. After pre-processing we return individual pointclouds showing each plant's canopy.

During pre-processing we extract the points that show canopy. Therefore, we identify and remove the ground points using RANSAC to fit a ground plane into pointcloud segments. This is done in a piecewise-fashion to better approximate steep-slope vineyards complex topology. Segments of size 10m x 10m were empirically identified as effective trade-off between accuracy and robustness. Then duplicate points and outliers are removed using functions provided by the the Open3D-Library [8]. Using the plant coordinates we extract bounding boxes of size 1,2m x 1,2m around each plant. The trunk points are removed by estimating the height of the bottom of canopy and deleting all points below. We calculate a histogram of the number of points over the up vector for each plant. Only a minor proportion of points falls onto the trunk, so beginning from the ground we define the bottom canopy height as the first bin that

contains > 1.5% of all points.

**Volume Estimation** We use the convex hull to estimate plant canopy volume. We considered alpha shapes, such as proposed by [1], but found that the impact on classification performance is marginal and convex hull is faster and parameterfree.



**Fig. 2:** Example of a convex hull (green) fit to the pointcloud (red) of the extracted foliage of a single plant in 3D shown from different perspectives. Viewpoint orthogonal to plant-rows (left), parallel to plant-rows (middle) and top view (right).

**Thresholding** Given the plant volume  $v_i$  of the  $i$ -th plant, we perform binary classification using a volume threshold  $\tau$ . If  $f_{\tau}(v_i) = v_i < \tau$  we assign the positive class, i.e. plant missing plant, otherwise negative. In [1], plants are grouped in groups of three and the average volume over all groups is used as threshold for classification. We follow this approach for single plants, denoted as  $\tau_{\text{avg}}$ . This is likely to produce many false positives. Hence, we propose two additional approaches. We offset the average by the standard deviation  $\tau_{\text{std}}$  and we use labeled samples to adjust the threshold such that the F1-Score over the train set  $V$  is maximized. The thresholds are calculated as follows:

$$\tau_{\text{avg}} = \bar{v} = \frac{1}{N} \sum_i^N v_i;$$

$$\tau_{\text{std}} = \sqrt{\frac{1}{N} \sum_i^N (v_i - \bar{v})^2};$$

$$\tau_{\text{opt}} = \arg \max_{\tau} (F1(f_{\tau}(V))).$$

### 3 EXPERIMENTS

**Datasets** The examined vineyards is located in Bernkastel-Kues in Germany and belongs to DLR Mosel. More information about the vineyard can be found in Hermes [2]. 3D pointclouds of the vineyards were collected with different sensors and platforms (see Tab. 1). L1-Lidar and L1-opt where acquired from a DJI Matrice 300 RTK drone at an altitude of 35m and Hesai-XT from a robotic platform on ground-level. More information on the acquisition of the latter dataset can be found in [5]. Additionally, we use a list of plant coordinates<sup>2</sup>.

Name	Sensor	Method/Software	Points	Acquisition Date
L1-Lidar	Zenmuse L1 Lidar	DJI Terra	66,793,953	03.08.2022
L1-opt	Zenmuse L1 Camera	Photogrammetry <sup>1</sup>	43,365,338	03.08.2022
Hesai-XT	Hesai-XT-32	SLAM	78,118,515	08.08.2022

**Tab. 1:** Overview over acquired pointcloud datasets.

Similar to [3] this list could also be automatically derived from the pointcloud. Finally, for evaluation we derived our ground truth labels for missing plants from a plant growth rating file<sup>3</sup>. It contains 1712 manual per-plant annotations about the plants' growth stage, ranging from 0 to 9. We follow the rating scheme proposed by Porten [7] and define 0 as positive cases, otherwise negative.

**Experimental Setup** We focus our evaluation on quantifying the difference in classification performance given different thresholding strategies and acquisition methods and platforms. In all experiments we use the same parameters for pre-processing and volume estimation. We use 5-fold cross-validation to get an objective estimate of the models performance. Based on predictions we calculate accuracy, precision, recall and F1-score and calculate mean and standard deviation. Accuracy, precision and recall are not very expressive individually. Hence, we provide the results for completeness, but focus on F1-Score.

Dataset	Thresholding	Accuracy	Precision	Recall	F1-Score
L1-opt	$\tau_{avg}$	0.474 ± 0.130	0.176 ± 0.037	<b>0.901</b> ± 0.034	0.293 ± 0.054
	$\tau_{std}$	<b>0.884</b> ± 0.015	0.000 ± 0.000	0.000 ± 0.000	0.000 ± 0.000
	$\tau_{opt}$	0.831 ± 0.073	<b>0.449</b> ± 0.172	0.558 ± 0.153	<b>0.445</b> ± 0.059
L1-Lidar	$\tau_{avg}$	0.469 ± 0.200	0.150 ± 0.041	<b>0.662</b> ± 0.140	0.236 ± 0.045
	$\tau_{std}$	0.836 ± 0.081	0.557 ± 0.273	0.339 ± 0.237	0.290 ± 0.123
	$\tau_{opt}$	<b>0.843</b> ± 0.092	<b>0.616</b> ± 0.273	0.343 ± 0.233	<b>0.313</b> ± 0.138
Hesai-XT	$\tau_{avg}$	0.525 ± 0.136	0.180 ± 0.040	<b>0.801</b> ± 0.075	0.291 ± 0.057
	$\tau_{std}$	<b>0.877</b> ± 0.033	<b>0.527</b> ± 0.123	0.446 ± 0.154	0.447 ± 0.089
	$\tau_{opt}$	0.864 ± 0.039	0.495 ± 0.147	0.510 ± 0.144	<b>0.462</b> ± 0.061

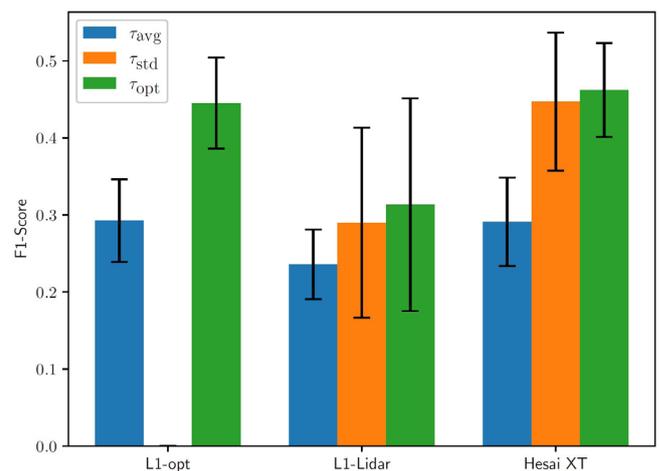
**Tab. 2:** Results of our missing plant detection experiments. We used 5-fold crossvalidation and report the mean and standard deviation per metric. The best results per dataset are highlighted.

L1-opt and Hesai-XT result in similar classification performance, indicating that ground-based acquisition is not better in general. However, L1-opt's average volume (ca. 0.25m<sup>3</sup>) is lower than Hesai-XT (ca. 0.55 m<sup>3</sup>) and L1-Lidar (ca. 0.44 m<sup>3</sup>) and noise seems higher leading to a negative threshold and hence no detection with  $\tau_{std}$ . The L1-Lidar pointcloud has the least amount of points which results in lower density and hence in lower classification performance.

<sup>1</sup> Agisoft Metashape Professional

<sup>2</sup> provided by AeroDCS GmbH, Koblenz, Germany

<sup>3</sup> provided by DLR Mosel, Bernkastel-Kues, Germany



**Fig. 3:** Comparison of different thresholds on all three pointcloud datasets with respect to F1-Score.

A comparison of the thresholding methods is shown in Fig. 3. In general the best performance is achieved by  $\tau_{opt}$ . The baseline  $\tau_{avg}$  has many false positives and lowest performance. Surprisingly,  $\tau_{std}$  performed almost as well as the optimized threshold without requiring annotated samples.

## 5 CONCLUSION

We presented a simple approach for missing plant detection based on volume estimation of 3D canopy pointclouds. We compared different acquisition techniques and showed that ground-based acquisition is not better in general and that photogrammetric pointclouds might lead to underestimation of plant volume. We showed that the most effective thresholding method is by fitting it on annotated train data. However, in real scenarios this might not be available as annotation is expensive. Luckily, a threshold based on standard-deviation-adjusted average achieves comparable results without relying on labeled samples. Both presented approaches achieve ca. 7-10% improvement over the average-thresholding baseline. In conclusion, we demonstrated that volume-estimation-based thresholding for missing-plant detection can be effective in steep-slope vineyards. However, acquisition technique and thresholding strategy must be chosen with care. Even though manual annotation of plant coordinates must only be done once, it would still be useful to combine our approach with plant localization to omit manual annotation completely.

## ACKNOWLEDGMENTS

This work took place in the context of the project „Smarter Weinberg 5G“, funded by the German Federal Ministry for Digital and Transport. Thanks to our project partners Vision & Robotics GmbH, DLR Mosel and aeroDCS GmbH for data provision.

## REFERENCES

- [1] S.F. Di Gennaro and A. Matese. Evaluation of novel precision viticulture tool for canopy biomass estimation and missing plant detection based on 2.5D and 3D approaches using RGB images acquired by UAV platform. *Plant Methods*, 16(1):91, 2020. ISSN 1746-4811. <https://doi.org/10.1186/s13007-020-00632-2>.
- [2] Christopher Hermes. Einsatz berührungsloser messtechniken zur bestimmung pflanzenphysiologischer parameter bei reben (vitis vinifera l.). Master's thesis, Justus-Liebig-Universität Gießen, 2020.
- [3] Juan M. Jurado, Luís Pádua, Francisco R. Feito, and Joaquim J. Sousa. Automatic grapevine trunk detection on uav-based point cloud. *RemoteSensing*, 12(18), 2020. ISSN 2072-4292. <https://doi.org/10.3390/rs12183043>. URL <https://www.mdpi.com/2072-4292/12/18/3043>.
- [4] Christian Kötting, Eike Gassen, and Karsten Berns. A robot platform for steep slope vineyards. In Andrey Ronzhin, Karsten Berns, and Alexander Kostyaev, editors, *Agriculture Digitalization and Organic Production*, pages 221–233, Singapore, 2022. Springer Nature Singapore. ISBN 978-981-16-3349-2.
- [5] Mark O. Mints, Nick Theisen, Peer Neubert, and Dietrich Paulus. Point-cloud-based change detection for steep slope vineyard agriculture. In *2023 IEEE SENSORS*, pages 1–4, 2023. <https://doi.org/10.1109/SENSORS56945.2023.10325099>.
- [6] Andrea Pagliai, Marco Ammoniaci, Daniele Sarri, Riccardo Lisci, Rita Perria, Marco Vieri, Mauro Eugenio Maria D'Arcangelo, Paolo Storch, and Simon-Paolo Kartsiotis. Comparison of aerial and ground 3d point clouds for canopy size assessment in precision viticulture. *Remote Sensing*, 14(5), 2022. ISSN 2072-4292. <https://doi.org/10.3390/rs14051145>. URL <https://www.mdpi.com/2072-4292/14/5/1145>.
- [7] Matthias Porten. Validierung\* des vegetativen und generativen Wuchszustands von Reben in variierenden Bewirtschaftungs- und Stressbelastungssystemen. doctoralthesis, Universität Trier, 2020.
- [8] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. arXiv:1801.09847, 2018.

# Alpha Shape For 3D Point Cloud Analysis

Burak Hafizoglu, Leonardo Maben, Stephan Brodkorb, Benjamin Hohnhäuser

Gesellschaft zur Förderung angewandter Informatik e.V., Volmerstraße 3, 12489 Berlin, Germany

www.gfai.de

**Abstract.** This paper presents a framework for calculating overspray in industrial coating applications by leveraging 3D point cloud data and alpha shape analysis. The Möller-Trumbore ray-triangle intersection algorithm, combined with alpha shapes, enables effective identification of coated surfaces and overspray regions by assessing intersection points between the robot's spray nozzle and target object. Overspray quantities are estimated by calculating the distribution values of intersected and non-intersected points, based on their relative positions to the spray nozzle and object's surface. Experimental results show that alpha shapes are effective in reconstructing 3D boundaries, minimizing coating material waste and optimizing coating precision. This methodology offers economic and environmental benefits by reducing overspray and improving the efficiency of coating applications.

**Keywords.** 3D Point Clouds, Alpha Shapes, Overspray Calculation

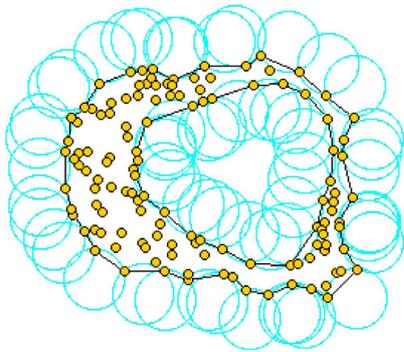
## 1 INTRODUCTION

With an increasing demand for precise representation of complex surfaces, 3D point cloud data has become essential in fields ranging from industrial robotics to geospatial analysis. Numerous individual points, each of which represents a position in three dimensions, constitute 3D point cloud data. These points, which are often represented by  $x$ ,  $y$ , and  $z$  coordinates, capture the surface features and spatial organization of objects or surroundings. Dealing with 3D point clouds can be challenging in terms of noise, incompleteness, irregular spacing, and complicating the extraction of geometric information. With the detailed spatial information captured in point clouds, these datasets serve as the foundation for reconstructing 3D surfaces through mesh creation. This process enables the visualization and modelling of complex shapes, facilitating applications like surface reconstruction and terrain modelling. The 3D Delaunay triangulation is particularly well-suited for this purpose, offering a robust geometrical approach and solid theoretical foundation which represents the meshes formed through triangles [1].

Triangle meshes provide an accurate and structured representation of the surfaces a coating robot encounters. By using triangular meshes, realistic coating scenarios can be simulated, facilitating precise coating applications and optimized path planning in robotic systems. Overspray, the portion of coating material that misses the target surface, increases waste and leads to higher costs. To minimize the economic and environmental impact of overspray, the triangular mesh data of the object can be used for effective boundary recognition of the target area. This enables adjustments to the spray gun's angle, direction, and distance to reduce overspray. Accurately calculating the shape of target surfaces is crucial for effective coating application. To identify the exact surface affected by the robot's spray gun, the Möller-Trumbore ray-triangle intersection algorithm and the alpha shape method are utilized.

## 2 METHODS

The Möller-Trumbore ray-triangle intersection algorithm efficiently determines whether a ray intersects a triangle in 3D space [2]. This method is particularly valuable in coating simulations, as it allows for the rapid identification of specific triangles covered by the robot's nozzle, enabling precise tracking of coated areas. The alpha shape algorithm is employed as a primary tool to achieve highly accurate reconstruction of the geometric structure of 3D point clouds. Alpha shapes belong to a family of closed polygons in 3D space that represent the shape of a finite set of points (or point cloud) in the Euclidean plane.



**Fig. 1:** Finding Alpha Value

The algorithm begins with a finite set of points and uses Delaunay triangulation as a preprocessing step to compute the circumradius  $\sigma$  for each point in the set. The alpha shape is then formed by introducing a parameter  $\alpha$ , which defines a spherical region around each point in the dataset. The  $\alpha$ -complex of the given point set, for any real number  $\alpha$ , is the simplicial complex generated by the circumcircles of each simplex in the triangulation whose radii are at most  $1/\sqrt{\alpha}$ . The radius of the circumcircle for each simplex is denoted as  $r(\sigma)$ , and the alpha shape algorithm determines the threshold for retaining simplexes using the equation shown below [4].

$$r(\sigma) \leq \frac{1}{\sqrt{\alpha}} \quad (1)$$

The alpha shape is constructed by connecting points that fall within the defined threshold value, while excluding any connections that are obstructed by other points. The result is a boundary that consists of arcs, caps, and points, which approximates the shape formed by the point set as it can be seen in Fig. 1 as black irregular line, which resulted a concave shape. Adjusting the alpha value allows for capturing different levels of detail in shape reconstruction. Higher alpha values yield smoother,

simpler shapes, while lower values capture finer details, including concavities and small-scale features. Selecting an appropriate alpha value is essential for accurate shape representation. The alpha value can either be adjusted manually to emphasize/remove desired details or chosen optimally which will represent the natural structure of the point set. This flexibility ensures that the reconstructed shape aligns closely with the intended level of detail.

## 3 RESULTS

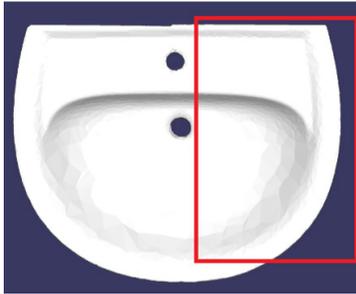
The coated area is influenced by factors such as nozzle angle, spray pattern, nozzle type, the paint gun's position, the object's surface slope, and the details of the triangular mesh. For instance, different nozzles produce varying spray patterns; a nozzle emitting an elliptical spray on a flat surface may create a conical pattern in 3D space. To accurately determine the coated area, points are generated along the base ellipse of the spray pattern, and rays are projected toward these points to check for intersections with the triangular mesh using the Möller-Trumbore ray-triangle intersection algorithm. Rays that miss triangles result in overspray, forming non-intersecting points. These non-intersecting points are used to create a concave polygon with the alpha shape algorithm, representing the overspray area.

Calculating the area of this polygon provides an estimation of overspray, assuming an equal distribution of glaze material. However, in reality, most glaze material tends to accumulate at the center after the spraying process is completed for a given position. The alpha shape method is also essential for accurately defining the exact shape of the intersecting points. This is crucial for determining whether specific triangles are contained within the polygon formed by the intersecting points. By identifying and adding these triangles to a list, the precision of the simulation is enhanced. This approach not only enables clear and precise visualization of the coating on the object but also significantly improves the accuracy of the simulation.

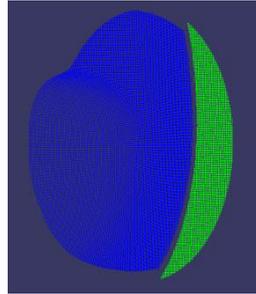
Fig. 2-5 illustrates how the alpha shape procedure is beneficial in the coating process. Fig. 2 shows a coated sink, with the target surface to be coated represented by the red rectangle. In this case, a certain amount of material has been sprayed onto the sink, but due to the angle and position, some of the material misses the object and ends up being oversprayed. To calculate the overspray, the amount of glaze material that lands on the object must first be determined. Fig. 3 shows the intersecting (blue) and non-intersecting (green) points after applying the Möller-Trumbore Intersection Algorithm, generated using the alpha shape algorithm. In Fig. 4,

the intersecting points highlight the irregular shape of the coating surface, while the non-intersecting points indicate the overspray area. Fig. 5 displays the coated sink according to the shape of the intersecting points. The overspray ratio can be estimated by analyzing the non-intersecting area using the elliptical pattern.

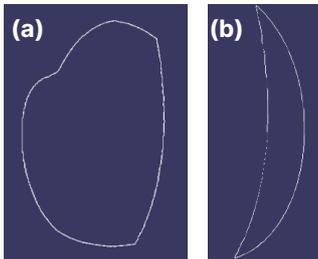
If the spray is evenly distributed, this position will result in an overspray of 18.4636% of the glaze material.



**Fig. 2:** Target area of coating process (Red rectangle)



**Fig. 3:** Intersecting Points (Blue) / non-intersecting Points (Green)



**Fig. 4:** a) Alpha shape of intersecting points  
b) Alpha shape of non-intersecting points



**Fig. 5:** Partially coated sink using intersecting points alpha shape

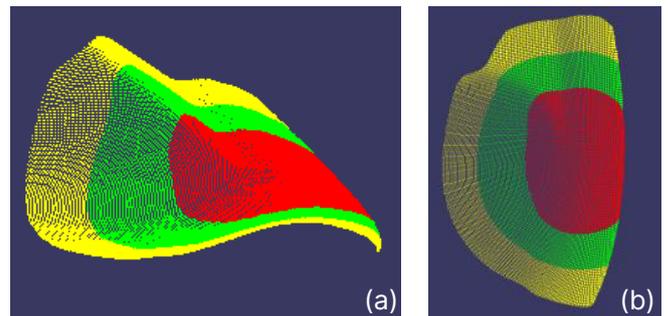
To estimate the potential overspray in grams, a 3D distribution function is developed. This function accounts for both intersecting and non-intersecting points, as well as their respective distances from the spray gun position and the center of the ellipse representing the target surface for each position. The distances are weighted, with 30% assigned to the distances between the points and the spray gun position, and 70% assigned to the distances between the points and the ellipse center. These weights can be adjusted based on real-world tests to achieve more accurate results.

Fig. 6 presents the distribution of glaze material is calculated using 3D distribution function and visualized with different color levels describing the distribution density: yellow represent the least material, green

indicating medium distribution, and red illustrating the highest accumulation of material. Fig. 7 indicates the distribution of glaze material onto sink surface. The distribution value of the points can be related to their corresponding triangles using barycentric coordinates, allowing for the estimation of their approximate gram values. The visualization and calculation of the gram values demonstrate the user to identify which parts are thicker or thinner, thereby improving the efficiency of material usage considering the right position and angle of nozzle, and robot trajectories.

Fig. 6 showcases the distribution of glaze material, calculated using the 3D distribution function and visualized with different color levels representing the distribution density: yellow represents the least material, green indicates medium distribution, and red illustrates the highest accumulation of material. Fig. 6 c) shows the distribution of glaze material on the sink surface. The distribution values of the points can be linked to their corresponding triangles using barycentric coordinates, enabling the estimation of their approximate gram values. The visualization and calculation of these gram values help users identify areas with thicker or thinner coatings, thereby improving material usage efficiency by optimizing the nozzle position, angle, and robot trajectories.

The total glaze amount for this position is calculated to be 12.61 grams, determined from the spray rate of the nozzle and the process duration. Following the application of the 3D distribution, the overspray is found to be 2.03 grams, resulting in an overspray ratio of 16.12%.



**Fig. 6:** a, b) Distribution of glaze material  
c) Distribution of glaze material onto sink surface

Alpha shapes are useful not only for overspray calculations but also for making covered regions easier to visualize in coating simulations. The experiments demonstrated that alpha shapes effectively reconstruct the boundaries of 3D objects. The appropriate alpha value was determined to capture fine details of the object and achieve highly accurate results. Alpha shapes generated precise models that were both topologically and geometrically accurate, as demonstrated through various case studies, including applications in industrial robotics and geomatics surveys.

#### 4 DISCUSSION

The findings highlight the adaptability and reliability of the alpha shape approach when working with point cloud data. The process of calculating and visualizing overspray in grams can be applied to every spray gun position. Depending on the specific application, the choice of the alpha value requires careful consideration. While the method is highly effective for shape reconstruction, it can be computationally intensive for large datasets. In such cases, optimizations or parallel processing techniques may be necessary. Additionally, the distribution formula must be adjusted based on real-world tests, as robots may have different nozzles, glaze parameters, and other variables.

#### 5 CONCLUSION

Alpha shapes provide a powerful tool for 3D data analysis, capturing fine details of point cloud data and playing a crucial role in object reconstruction. The method is highly adaptive, enabling precise identification of the boundaries of the coated surface, and facilitates the calculation of distributed and oversprayed gram values. This calculation can significantly improve the efficiency of the coating process, benefiting both the economy and the environment. Future studies may focus on optimizing computational time to determine the optimal alpha value for large-scale datasets in robotics, geomatics, and other fields.

#### REFERENCES

- [1] Tianyun Su, Wen Wang, Haixing Liu, Zhendong Liu, Xinfang Li, Zhen Jia, Lin Zhou, Zhuanling Song, Ming Ding, and Aiju Cui. An adaptive and rapid 3d delaunay triangulation for randomly distributed point cloud data. *The Visual Computer*, 1–25, 2022.
- [2] Tomas Moller and Ben Trumbore. Fast, minimum storage ray/triangle intersection. In *ACM SIGGRAPH 2005 Courses*, 7–es. 2005.
- [3] Herbert Edelsbrunner and Ernst P Mücke. Three-dimensional alpha shapes. *ACM Transactions On Graphics (TOG)*, 13(1):43–72, 1994.
- [4] Erik Carlsson and John Carlsson. Computing the alpha complex using dual active set methods. *arXiv preprint arXiv:2310.00536*, 2023.

# Automatisierung von Lackierprozessen bei kleinen Losgrößen: Entwicklung eines Multikamerasystems zur Echtzeit-Erfassung manueller Sprühbahnen

Habib Ben Khedher<sup>1</sup>, Martin Wolff<sup>2</sup>, Nurradeen Beshar<sup>3</sup>, Benjamin Hohnhäuser<sup>4</sup>

Gesellschaft zur Förderung angewandter Informatik e.V. Volmerstraße 3, Berlin, Germany

<sup>1</sup>ben\_khedher@gfai.de, <sup>2</sup>wolff@gfai.de, <sup>3</sup>beshar@gfai.de, <sup>4</sup>hohnhaeuser@gfai.de

**Abstract.** Die Automatisierung von Beschichtungsprozessen ist in der industriellen Fertigung besonders bei kleinen Losgrößen eine Herausforderung. Im Rahmen des FuE-Kooperationsprojekts „**AI4Tech - AuLaTeach**“ wird ein System entwickelt, das Lackierroboter durch Vormachen programmieren kann. Ziel ist die Aufzeichnung von Sprühbahnen (6D) über die Zeit, inklusive aller technologischen Prozessparametern, die in ein ausführbares Roboterprogramm überführt werden. Die Erfassung der Trajektorien erfolgt mittels eines Multikamerasystems, das aktive Marker auf der Sprühlackierpistole verfolgt. Ein 3D-Kalibrierkörper und der Levenberg-Marquardt-Algorithmus sorgen für eine präzise räumliche Kalibrierung der Kameras. Der Flood-Fill-Algorithmus wird zur Marker-Erkennung eingesetzt, während ein paarweiser Frame-Vergleich die Trajektorienverfolgung unterstützt. Zur Überwindung von Sichtverdeckungen wird zusätzliche Sensorik an der Pistole installiert. Die so erfassten Daten werden direkt in ein gemeinsames Koordinatensystem transformiert, was eine präzise Reproduktion des Beschichtungsprozesses ermöglicht.

## 1 EINLEITUNG

In vielen Bereichen der industriellen Fertigung nimmt der Automatisierungsgrad stetig zu. Hierbei verringert sich der anzutreffende Automatisierungsgrad kontinuierlich mit der Abnahme der produzierten Losgröße. Dazu bestimmt die angewandte Technologie und der Prozess selbst ebenfalls das Maß an Automation, welches sinnvollerweise umsetzbar ist. Für mittlere und kleine Losgrößen ist eine Automatisierung von Beschichtungsprozessen unter wirtschaftlichen Gesichtspunkten häufig schwer oder nicht darstellbar. Allgemein nimmt die Automatisierungstiefe für Beschichtungsprozesse jedoch deutlich mit sinkender Losgröße ab. Werkstücke in kleinen Losgrößen werden in den meisten Fällen von Hand lackiert, dabei ist das Einsparpotenzial in Bezug auf Material, Zeit und Energie durch eine Automatisierung hier gerade am größten.

Um diese Herausforderung anzugehen, wird im Rahmen des FuE-Kooperationsprojektes „**AI4Tech - AuLaTeach**“ (**16KN117724**) bei der Gesellschaft zur Förderung angewandter Informatik e.V.

im Forschungsbereich Bildverarbeitung / Industrielle Anwendungen und die Firma „THÜMA – Maschinen und Service GmbH“ zusammen ein System entwickelt, zur Programmierung von Lackierrobotern für kleine Losgrößen durch Vormachen.

Ziel dieses FuE-Vorhabens ist die Entwicklung eines Systems zur Aufzeichnung von Sprühbahnen (6D) über die Zeit, inklusive aller relevanten technologischen Prozessparameter (Drücke, Ventilöffnung etc.), als Trajektorie in einem Weltkoordinatensystem (stationär), die von einem Werker beim fachgerechten, manuellen Beschichten eines Werkstücks ausgeführt wird. Diese Trajektorie wird automatisch in eine strikte Werkzeug-Werkstück-Relation überführt und gespeichert.

Diese aufgezeichneten Trajektorien können dann direkt in das Roboterkoordinatensystem transformiert und

zusammen mit den Prozessparametern in ein ausführbares Roboterprogramm überführt werden, so dass der Roboter in der Lage ist, das exakt gleiche Beschichtungsprogramm an einem identischen Werkstück auszuführen. Die Qualität der Beschichtung soll dabei der fachgerechten, manuellen Beschichtung gleichwertig sein.

Darüber hinaus soll das Beschichtungsergebnis anhand einer bestehenden Auftrags- und Schichtdickensimulation berechnet und visualisiert werden, um dem Benutzer eine Möglichkeit zu geben, die Qualität des erzeugten roboterbasierten Beschichtungsprogramms auch in der Simulation bewerten zu können. Ein solches direktes Feedback ist insbesondere notwendig, wenn Auftragsbahnen modifiziert wurden.

Zur Anpassung der Auftragsbahnen sollen zwei unterschiedliche Varianten durch das System unterstützt werden. Die erste klassische ist eine Benutzungsschnittstelle in einem integrierten Offline-Programmiersystem. Es soll möglich sein, Stützstellen der Bewegungsbahn zu verschieben und in ihrer Parametrierung anzupassen und direkt die Wirkung der Änderung in der Auftragssimulation als Ergebnis visualisiert zu bekommen. Hierdurch kann die direkte Wirkung von Änderungen an der Auftragsbahn bewertet werden.

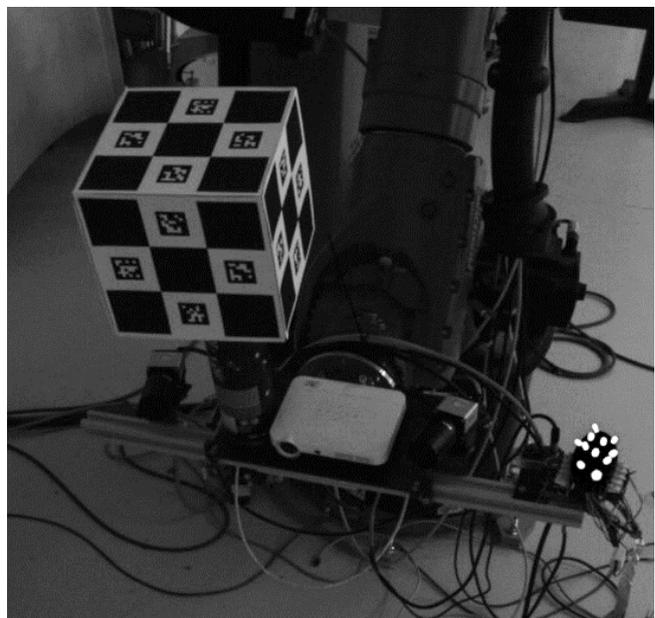
Die zweite Möglichkeit Auftragsbahnen anzupassen soll in einer VR-Umgebung (virtual reality) umgesetzt werden. Hier sollen prinzipiell dieselben Anpassungen wie im ersten Fall möglich sein, nur kann hier zusätzlich mit der Lackierpistole gearbeitet werden, allerdings mit virtueller Farbe, so dass der haptische Eindruck dem ursprünglichen Lackierprozess sehr ähnlich ist. Der Materialauftrag wird mit Hilfe der oben beschriebenen Schichtdickensimulation dem Benutzer direkt angezeigt. Prinzipiell ist hierdurch auch eine vollkommen virtuelle Vorgabe von Lackierbahnen mit Ergebnisvorhersage möglich, die dann in ein ausführbares Roboterprogramm überführt werden kann.

## TECHNISCHE ANSÄTZE

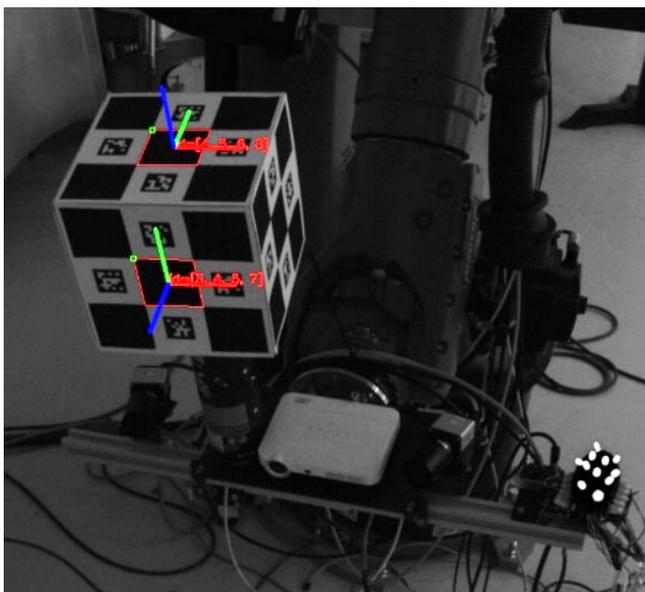
Für die Erfassung der Trajektorien wird ein Multikamerasystem eingesetzt, das einen festgelegten Arbeitsraum überwacht und aktiv Marker (siehe Abbildung 1) auf der Sprühlackierpistole erfasst und verfolgt. Um eine stabile Erkennung der Marker im Bild zu gewährleisten, wird ein geeigneter Bandpassfilter verwendet, der sicherstellt, dass im aufgenommenen Bild nur die relevanten Marker sichtbar sind. Jede Kamera führt dabei eine individuelle Erkennung und Verfolgung

der Marker durch (siehe Abbildung 2).

Die räumliche Kalibrierung der Kameras erfolgt durch die Verwendung eines präzise gestalteten 3D-Kalibrierkörpers (siehe Abbildung 1), bestehend aus einem Würfel, dessen Flächen jeweils mit Diamond-Charuco-Boards versehen sind. Im Rahmen des Kalibrierungsprozesses extrahiert jede Kamera einen Translations- und Rotationsvektor, die ihre jeweilige räumliche Positionierung und Orientierung relativ zum Kalibrierkörper im dreidimensionalen Raum beschreiben. Unter Berücksichtigung des bekannten Sichtfeldes jeder Kamera auf die spezifischen Charuco-Boards wird in einem weiteren Schritt der Levenberg-Marquardt-Algorithmus angewandt. Dieser Algorithmus dient der nichtlinearen Optimierung der erfassten Translations- und Rotationsvektoren, indem er die Kalibrierdaten durch eine iterative Anpassung mit statischen Referenzaufnahmen der Szene verfeinert. Dieser Prozess stellt eine maximale Präzision in der Bestimmung der Position und Ausrichtung der Kameras sicher und minimiert potenzielle Kalibrierungsfehler durch eine konsistente, auf empirischen Daten basierende Optimierung.



**Abb. 1:** Aufnahme aus eine Kameraposition



**Abb. 2:** Ermittelte Pose des Kalibrierkörpers

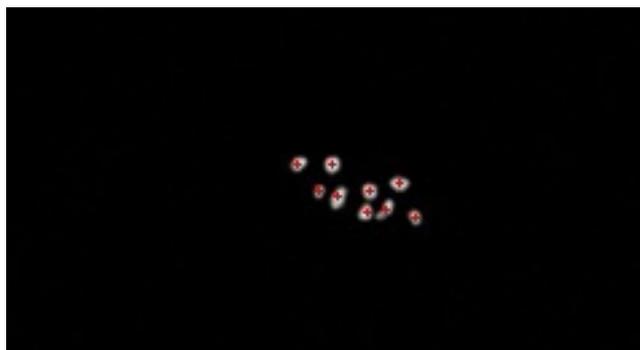
Eine zentrale Recheneinheit sorgt dafür, dass die Kameras synchron arbeiten, und mit den Messdaten wie Druck, Ventilstellung und Durchfluss aus den Sensoren der Sprühpistole zusammengeführt werden. So werden die Kamerabilder optimal abgestimmt und wichtige Prozessdaten aufgenommen.

### ERKENNUNG DER SPRÜHBAHNEN

Für die präzise Erkennung der aktiven Marker wird der Flood-Fill-Algorithmus verwendet, welcher das Bild systematisch Pixel für Pixel durchläuft, um zusammenhängende Bereiche (Blobs) zu identifizieren. Der Algorithmus überprüft dabei für jedes Pixel, ob es bereits markiert wurde und ob es zu einem zusammenhängenden Bereich oder Innenbereich gehört. Sobald ein zusammenhängender Bereich erkannt wird, werden die entsprechenden Pixel einem dominanten Blob zugeordnet, sofern bestimmte Kriterien erfüllt sind. Dieser Ansatz ermöglicht eine zuverlässige Segmentierung von Bildbereichen basierend auf Pixelverbindungen und unterstützt die präzise Erfassung von Objekten oder Strukturen im Bild. [1][2]



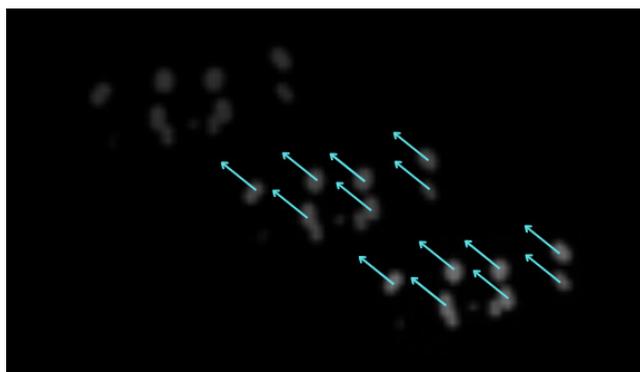
**Abb. 3:** Bild vor der Erkennung



**Abb. 4:** Ermittelte Pose des Kalibrierkörpers

### VERFOLGUNG DES TARGET MIT AKTIV MAKERN

Die Verfolgung basiert auf einem paarweisen Frame-Vergleich, bei dem für jeweils zwei aufeinanderfolgende Bilder einer Kamera der Richtungsvektor (2D) für jedes Target über die Berechnung des Schwerpunkts ermittelt wird (siehe Abbildung 5). Diese Methode ermöglicht eine präzise Bestimmung der Geschwindigkeit des Targets und unterstützt bei temporären Obstruktionen die eindeutige Zuordnung von Blobs. Zudem erlaubt die Bewegungsrichtung eine prädiktive Schätzung des Target-Standorts im folgenden Frame, wodurch Tracking-Verluste minimiert werden. Durch die Kombination dieser Bewegungsdaten mit den Positionsinformationen aus mehreren Kameraperspektiven wird die Zuverlässigkeit der Trajektorienverfolgung selbst bei teilweiser Verdeckung der Targets erheblich gesteigert.[3]



**Abb. 5:** Optische Fluss aus 3 sukzessive zusammengeführten Frame

### OKLUSIONSPROBLEM

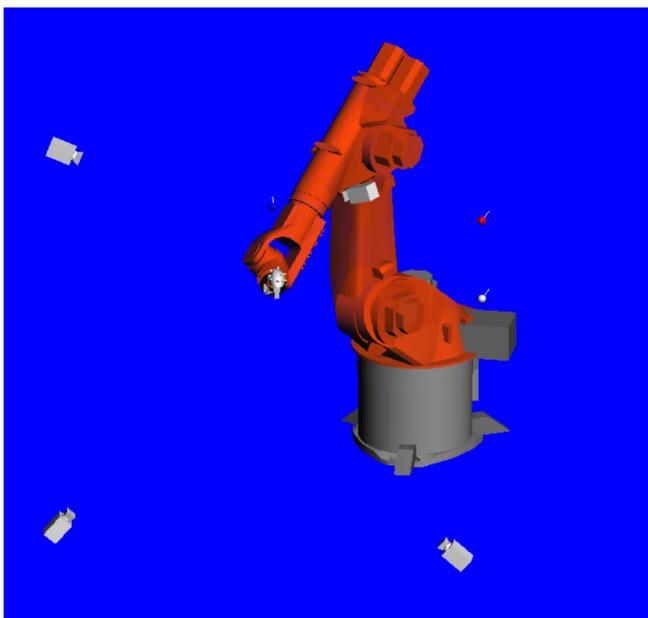
In bestimmten Blickwinkeln der Szene kann das Target teilweise verdeckt sein, was zu Unterbrechungen in der Trajektorienaufzeichnung führt. Ein Ausfall mehrerer Ansichten kann ebenfalls zu Berechnungsungenauigkeiten beitragen. Um diese Problematik zu lösen, wird an der Sprühpistole zusätzliche Sensorik installiert, die ab einem definierten Schwellenwert die Posenbestimmung

unterstützt. Sobald das Target wieder in das Sichtfeld gelangt, wird es mittels einer Transformation in das gemeinsame euklidische Koordinatensystem erneut erkannt und korrekt seinen ursprünglichen IDs zugeordnet.

## AUFBAU UND DURCHFÜHRUNG VON ERKENNUNG MITTELS MESSEXPERIMENT

### VERSUCHSAUFBAU

Zur Validierung des Algorithmus wurde ein kontrolliertes Testszenario mit einem KR60-KUKA-Knickarmroboter (siehe Abbildung 6) eingerichtet, der vorab definierte Bahnen wiederholt abfährt. Diese experimentelle Anordnung erlaubt die Erfassung von Ground-Truth-Daten zur Validierung der berechneten Pose und ermöglicht eine detaillierte Analyse potenzieller Kamerapositionen. Der Versuchsaufbau umfasste den 3D-Kalibrierkörper sowie das Target mit aktiven Markern, das üblicherweise an der Sprühpistole angebracht ist. Eine Kamera wurde systematisch in verschiedenen Positionen um den Arbeitsbereich des Roboters platziert, um die Bewegungen unter gleichbleibenden Bedingungen zu erfassen. Der 3D-Kalibrierkörper und das Target mit aktiven Markern waren fest am Endeffektor des Roboterarms montiert, wodurch eine exakte Übereinstimmung der aufgenommenen Bewegungen mit den Referenzdaten gewährleistet und eine präzise Evaluierung des Algorithmus ermöglicht wurde.



**Abb. 6:** 3D-Darstellung des Messexperimentes

### VERSUCHDURCHFÜHRUNG

Das Messexperiment wurde in einem strukturierten Ablauf durchgeführt, um präzise Ground-Truth-Daten zu gewährleisten.

Im ersten Schritt wurde eine Referenzaufnahme von vom Standpunkt der Kamera erstellt, um die extrinsische Kalibrierung mittels 3D-Kalibrierkörper zu bestimmen. Anschließend wurde ein Bandpassfilter vor die Kamera gesetzt, um die LEDs am Target als eindeutige „Blobs“ sichtbar zu machen und Bildstörungen durch unerwünschte Objekte zu minimieren.

Im zweiten Schritt wurden Bilddaten während einer simulierten Bewegungsfahrt des Roboterarms aufgenommen. Der Roboterarm folgte dabei vorab definierten Trajektorien, die in einzelne Segmente unterteilt waren. Die ersten drei Segmente umfassten Bewegungen entlang der X-, Y- und Z-Achsen des Roboterkoordinatensystems. An fest definierten Abständen entlang dieser Achsen hielt der Roboterarm in einer definierten position an, und die Kameras erfassten statische Aufnahmen des Targets.

Die abschließende Fahrt beinhaltete eine komplexe Trajektorie, die Bewegungen entlang aller drei Achsen umfasste und dynamische Punkte im Roboterkoordinatensystem abdeckte. Diese Abschlussfahrt ermöglichte die Erfassung zusätzlicher Datensätze zur Analyse von Bewegungen in mehrdimensionalen Raumachsen und diente der Validierung des Algorithmus bei komplexeren Bewegungsmustern.

### AUSBLICK

Die ersten Messversuche haben gezeigt, dass Modifikationen am Target, bestehend aus der LED-Marker-Konstellation und der Trägerkonstruktion, erforderlich sind, um die Präzision der Poseberechnung weiter zu steigern. Mit dem aktuellen Target sind folgende Anpassungen geplant: Die LEDs werden weiter auseinander positioniert, um eine Überlagerung der Markerbilder in den Kameras zu vermeiden. Dazu werden die Abmessungen des Targets in Länge und Breite vergrößert, um eine präzisere Erfassung zu ermöglichen.

Zusätzlich werden LEDs mit einem breiteren Abstrahlwinkel und reduzierter Intensität verwendet, um die Sichtbarkeit und Differenzierbarkeit der Marker in verschiedenen Kameraperspektiven zu verbessern.

Zur Optimierung der Testfahrten wird der Aufnahmeabstand

verringert, da sich der bisherige Abstand als zu groß erwiesen hat. Durch kleinere Schrittweiten bei den Messungen soll die Trajektorienverfolgung präziser getestet werden. Diese Anpassung ermöglicht eine detailliertere Erfassung der Bewegungen, wodurch die Genauigkeit und Zuverlässigkeit der Trajektorienanalyse weiter verbessert wird.

Anschließend ist die Entwicklung einer Datenfusion geplant, bei der die eingebaute Sensorik der Sprühpistole mit der internen Sensorik (Roboterpfad) sowie der externen Sensorik (Drehgeber des Drehtisches) kombiniert wird. Ziel dieser Datenfusion ist es, eine präzisere Synchronisation und Steuerung des gesamten Systems zu ermöglichen.

## FÖRDERHINWEIS

Dieses Forschungsprojekt wird durch das Zentrale Innovationsprogramm Mittelstand (ZIM) des Bundesministeriums für Wirtschaft und Klimaschutz (BMWK) unter der Projektnummer **16KN117724** gefördert. Wir danken dem BMWK für die finanzielle Unterstützung und die Möglichkeit, innovative Forschung und Entwicklung voranzutreiben.

Gefördert durch:



aufgrund eines Beschlusses  
des Deutschen Bundestages



## REFERENCES

- [1] S. Guha, Computer Graphics Through OpenGL® From Theory to Experiments, 15. Dez. 2022, EBook, CRC Press, pp. 473–474, ISBN: 9781000618389
- [2] V. Kshirsagar, „Flood Fill Algorithm - Explained with Examples,“ FavTutor, <https://favtutor.com/blogs/flood-fill-algorithm> (accessed Nov. 6, 2024)
- [3] OpenCV Documentation, „Optical Flow,“ OpenCV, [https://docs.opencv.org/3.4/d4/dee/tutorial\\_optical\\_flow.html](https://docs.opencv.org/3.4/d4/dee/tutorial_optical_flow.html) (accessed Nov. 6, 2024).

# Additively Manufactured Forging Dies with Contour-Optimized Cooling

Robin Roj, Jens Kneifel, Hans-Jürgen Gittel, Christian Pelshenke, Ralf Theiß and Peter Dültgen

Forschungsgemeinschaft Werkzeuge und Werkstoffe e. V., Papenberger Str. 49, 42859 Remscheid, Germany  
roj@fgw.de

**Abstract.** This scientific paper investigates strategies to enhance the economic viability and competitiveness of hot forging dies. A specific case study involving a benchmark hot forging die geometry, produced using additive manufacturing in a hybrid design, is presented. Stellite, known for its durability and suitability for laser powder bed fusion, was chosen as the die material. With a strong focus on cost-effectiveness, the innovative hybrid design combines an additively manufactured stellite engraving with a conventionally manufactured nickel bronze carrier. Key features, such as inlet and outlet ports for the cooling medium have been incorporated into the carrier, and its coefficient of thermal expansion has been adapted for the application-specific temperature conditions. Material properties, including compressive strength and offset yield strength, were assessed through uniaxial compression tests, transferred into a simulation-driven die design process. Demonstrators, produced via laser powder bed fusion, exhibited reduced abrasive wear when equipped with cooling systems. The results indicate extended die service life and economic benefits, highlighting the potential of the hybrid design and additive manufacturing approach for enhancing competitiveness in hot forging die production and utilization. This research offers valuable insights for the forging industry's pursuit of sustainability and efficiency.

**Keywords.** Cobalt-based Alloys, Forging, Integrated Cooling Channels, Laser Powder Bed Fusion, Stellite.

## 1 INTRODUCTION

The tool life and setup costs have a significant influence on the unit cost of a forging. Forging dies wear on the surface mainly due to high cyclic thermo-mechanical loads. To protect these from wear, there have been efforts in recent years to keep dies cooler for forging and hot forming by means of channels close to the surface. Due to the 3D contour, these can usually only be produced by laser powder bed fusion (LPBF). However, the materials available for LPBF usually have little resistance to the abrasive wear that also occurs.

Buchmayr shows that 70% of die damage is due to abrasive wear caused by high thermal stresses and consequently this shortens the service life [1]. Furthermore, 10-30% of the cost of the final product is due to the cost of the forging die. However, in addition to die costs, setup costs and times must also be considered. Furthermore, the study deals with damage to forging dies as well as dies and their repair options. Various repair methods are presented and discussed. It is emphasized that material selection, heat treatment, and coating play a crucial role in repair.

A paper by Foster et al. describes the use of laser metal deposition (LMD) with powder and the hardfacing alloy Stellite 21® to restore hot forging tools and dies [2]. It is successfully demonstrated that LMD is a promising process for manufacturing repair components for forging tools, as it can reduce tool wear and extend the life of forging tools. Foster et al. found that damaged or worn forging dies can be repaired using LMD, extending their lifetime and saving costs in comparison to conventional repair methods. The study identified Stellite 21® as a promising material because it offers good wear resistance, machinability, forgeability, as well as thermal conductivity, and improves the performance of the dies.

Topology optimization and lattice structures implemented during additive manufacturing (AM) can significantly increase the cooling efficiency of forging dies, resulting in

a significant improvement in tool life and wear resistance. Chantzis et al. developed a method for designing forging dies and dies with improved cooling performance and reduced wear under cyclic loading by AM [3]. For this purpose, the design for AM of hot stamping dies with improved cooling performance under cyclic loading conditions was investigated. The researchers used topology optimization and lattice structures to improve the cooling efficiency of the tool. The results show a significant improvement in tool life and wear resistance, as the optimized design can significantly reduce the temperature gradient and load on the tool. This leads to an increase in fatigue life and a reduction in the risk of cracks.

Komodromos, Kolpak, and Tekkaya also developed a method for designing forging tools and dies with improved cooling performance and reduced wear under cyclic loading by AM [4]. They dealt with the fabrication of forging tools and dies with integrated cooling channels using directed energy deposition (DED). The results show improvement in surface finish, reduction in wear, and significant improvement in tool life and wear resistance. The study investigated the design for AM of hot stamping dies with improved cooling performance under cyclic loading conditions. The researchers used topology optimization and lattice structure design to create cooling channels inside the tool and ball burnishing to improve surface quality. The study found that the optimized design can significantly reduce the temperature gradient and stress on the tool and increase fatigue life.

A publication by Tang et al. highlights the advantages and disadvantages of AM, such as design flexibility in terms of material properties and shape, and shorter lead times [5]. However, AM is usually associated with high manufacturing costs. Shorter lead times enable a faster response to customer requirements as well as market trends, and the flexibility of the design makes it possible to meet these requirements precisely. This feature is especially important for small and medium-sized enterprises (SME), as they usually produce one-offs or small batches.

## 2 MATERIALS AND METHODS

The present work uses the knowledge from the above sources to make the production of and with forging dies more economically attractive and to expand the competitiveness of SMEs. To test the scientific approaches, a forging die with a round engraving and elevation in the center was chosen. Fig. 1a reveals the geometry and shows an additively manufactured

specimen immediately after the printing process. To achieve the overall goal of increasing efficiency in the forging process, the following solution steps were implemented:

1. The stellite type Celsit 21 was selected as the material to be additively processed. This is a cobalt-based alloy with a carbon content of 0.25 m%. A large number of parameter studies were carried out, in which test specimens were printed in the form of cuboids and then analyzed metallographically. These studies were used to find parameters as well as to obtain high quality results. Fig. 1b shows one of these parameter studies, in which twelve cuboids are still attached to the build platform directly after LPBF.



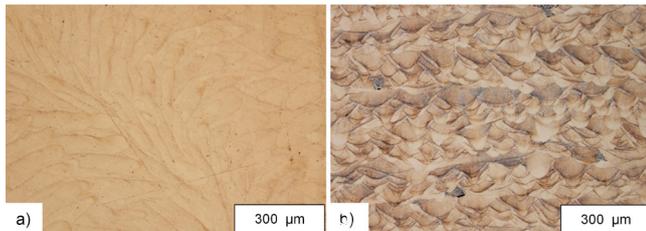
**Fig. 1:** a) Additively manufactured forging die without post-processing; b) Test print of cuboids for parameter determination.

2. The LPBF printer AconityMIDI from the manufacturer Aconity was used for processing. In this, argon was applied as an inert gas to achieve a low-oxygen atmosphere. The machine is also equipped with a build platform heater for temperatures up to 800 °C. As part of the parameter studies for the qualification of Celsit 21, the machining parameters and the temperature of the build platform heating were varied. Fig. 2 shows two examples of microstructural images taken during the metallographic examination of the cuboids, which on the one hand show good quality without inclusions, voids or impurities, and on the other hand show the solidification of the weld beads during AM.

3. The layout of the hybrid forging die was designed in such a way that the coolant connections are attached to a nickel bronze carrier, and thus the coolant is passed through it as well as the additively manufactured engraving. The carrier was designed to allow fabrication by casting or by milling. A nickel bronze alloy called Hovadur K220 was chosen as the carrier material.

4. The economic aspects of production and operation were already taken into account during the design phase

of the hybrid tool. During the development of the carrier and engraving, care was taken to ensure that both were designed for the thermo-mechanical loads in the forging process and that the volume of the component to be manufactured additively was as small as possible. Thus, the printing time was reduced as much as possible and the corresponding costs for AM were kept as low as possible.



**Fig. 2:** Microstructure images at 100x magnification; a) Cube 7; b) Cube 2.

5. Even before the hybrid tools were produced, virtual simulation techniques based on finite element analysis were used to simulate the effect of the forces acting in the forging process on the one hand, and to determine the thermally induced expansions of the material on the other. In the design phase, various options were considered for the connection of the additively manufactured engraving made of Celsit 21 and the milled nickel bronze carrier, such as bolting, soldering or shrink-fitting.

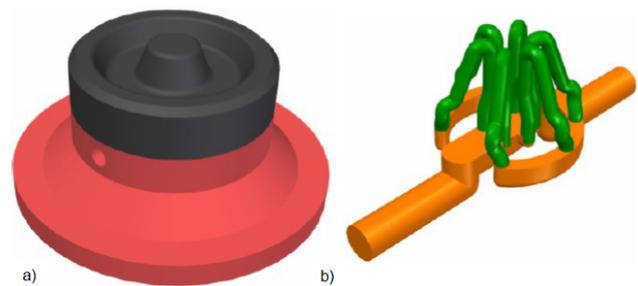
6. After investigating the idealized force ratios in one cycle by means of a thermo-mechanical simulation, it was determined that the connection could be made by soldering the two components together. Furthermore, investigations were carried out to determine the coefficient of thermal expansion (CTE) of both materials. These showed that although the alloys selected for the coating material Celsit 21 and the base material Hovadur K220 have a different CTE, the heat-related expansions during brazing as well as forging do not lead to any significant stresses.

7. After the forging tools were manufactured, serial forging tests were carried out, both, with and without coolant. In addition to varying the coolant flow, the influence of the semi-finished product temperature on the wear behavior of the printed forging tools was also investigated.

## 2.1 COOLING LAYOUT

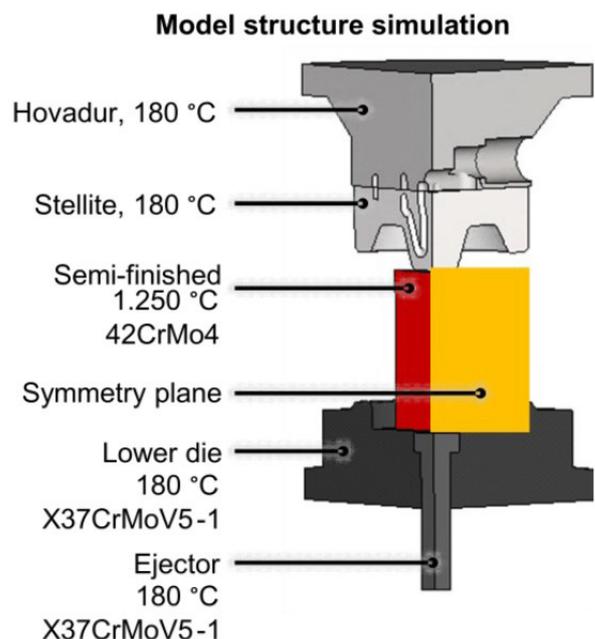
During development, it was decided to implement a hybrid design of the forging die. This consists of a carrier and engraving, including a layout for a symmetrical channel design with the aim of achieving homogeneous, wear-reducing mold cooling.

According to the described considerations for the joining technology, different concepts for the design were developed. Fig. 3a shows the conceptual design. Both, the supply and return lines are integrated into the carrier, together with the associated connections for the cooling medium. The base of the carrier has a larger diameter than the engraving and can thus be used for mounting in the machine. The joining surface is flat, with both components having the same diameter. Fig. 3b shows the concept for the design of the internal cooling channel network. While the flow and return have a comparatively large diameter, the system splits into six thinner arms in the mandrel area.



**Fig. 3:** a) Hybrid-built forging tool consisting of a carrier and engraving; b) Internal channel system with a cooling channel mesh for contour-adjacent tempering.

Fig. 4 shows the installation situation in the machine as well as the materials and temperatures to be expected.



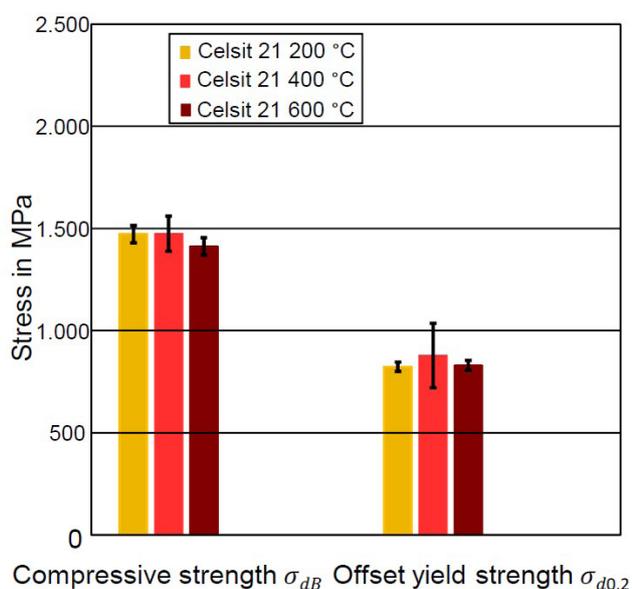
**Fig. 4:** Overview of the cut simulation model.

It can be seen that the division into the six branches is already done at the lower end. This makes the structure appear more complex and accurate. In accordance with the parameter studies, the cooling channels are designed with a diameter of 2 mm at the thinnest section. To ensure effective heat dissipation in the thermally highly stressed area, the cooling is concentrated on the mandrel area and not on the annular outer structure. This ensures homogeneous heat dissipation in the wear-critical area. The channels run 2 mm below, along the flanks of the engraving and each have a diameter of 2 mm. They are supplied by the forward and return flow, each with a diameter of 6.2 mm.

### 3 RESULTS

#### 3.1 UNIAXIAL CYLINDER COMPRESSION TESTS

During the forming process, the main loads act in the mandrel area of the die. The investigation was carried out at three different temperatures (200 °C, 400 °C, and 600 °C) and the results are shown in Fig. 5. Here,  $\sigma_{dB}$  (compressive strength) as well as  $\sigma_{d0.2}$  (offset yield strength) were determined from each temperature and the mean values were calculated. No relevant difference can be seen between the different temperatures for  $\sigma_{dB}$  and the crush limit  $\sigma_{d0.2}$ .  $\sigma_{dB}$  reaches the maximum value of 1,466 MPa at 200 °C. For  $\sigma_{d0.2}$ , a maximum value of about 870 MPa is reached at 400 °C. At the temperatures studied, the standard deviation, for both  $\sigma_{dB}$  and  $\sigma_{d0.2}$ , was below 5% (40 MPa), with the  $\sigma_{d0.2}$  upset limit having a higher standard deviation of 17% (150 MPa) at a temperature of 400 °C.



**Fig. 5:** Results of the compression tests with respect to  $\sigma_{dB}$  and  $\sigma_{d0.2}$  as a function of temperature.

Some standard deviation in the uniaxial cylinder compression tests can be explained due to the manufacturing variant with LBP of the specimens. This can be concluded by the slightly increased porosity compared to solid material, which cannot be completely avoided even depending on the selected process parameters.

#### 3.2 THERMAL SIMULATION WITH COOLING

For the thermal cooling simulation, a surrogate body was introduced into the cooling channel geometry to represent the cooling medium. The simulation results show that the highest stresses are present in the breakdown of the cooling channels, which are located directly in the center just below the mandrel surface.

Within the first cycle, a maximum temperature of 850 °C occurs in the body made of Celsit 21 in the area near the surface. This maximum temperature is limited to the mandrel radius. The remaining area of the surface has a temperature between 300 °C and 600 °C, depending on the contact time and contact pressure. The temperature inside the stellite tool remains constant at 180 °C.

#### 3.3 CONNECTION OF CARRIER AND ENGRAVING

With regard to the connection technology of the additively manufactured engraving, carrier and installation of the hybrid tool in the machine, different concepts were developed. In the final version, four jaws were provided for fixing, which are used for bolting to the rest of the machine structure. For this purpose, a step is to be incorporated both in the jaws and in the engraving for force transmission.

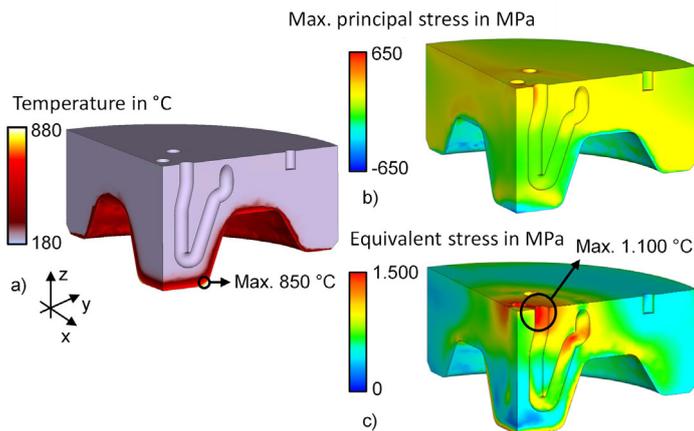
In determining the final concept, both the effort required during production and as-assembly as well as cost-effectiveness and quality were taken into account. Vacuum brazing was identified as a suitable process. A flat joining surface was preferred due to simplified production and soldering. This determination had a corresponding effect on the design of the layout of the cooling channels.

Hovadur K220 was selected as the substrate material. This is characterized by its high thermal conductivity coefficient and good brazing properties with constant corrosion and abrasion resistance. Hovadur K220 is a nickel bronze that meets all technical requirements. High-vacuum brazing was selected as the joining technique, using silver and copper solders with germanium and cobalt content as the brazing alloy. It was also investigated whether the solder should be used as a foil, paste, or in combination.

### 3.4 STRESS AND THERMAL SIMULATION WITHOUT COOLING

The results of the thermal simulation are shown in Fig. 6. To estimate possible cracking, the calculated positive maximum principal stress is compared with the tensile strength at the present temperature. By comparing with the calculated results (max. principal stress about 600 MPa), it can be concluded that the risk of cracking in the first forging cycle is unlikely.

A maximum of 1,100 MPa was calculated for the equivalent stress according to von Mises. This value is compared to the reported yield strength for the material at 180 °C, which ranges from 650 MPa (cobalt-based alloy number 22) to 1,200 MPa (cobalt-based alloy number 60) according to the literature, depending on the alloy [6]. The equivalent stress can be used to determine when the material-specific yield point is reached and is thus used to estimate tool failure due to plastic deformation.



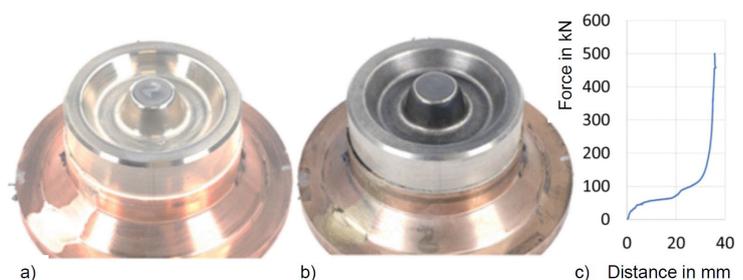
**Fig. 6:** Thermal simulation result without cooling effect; a) Overview of the entire simulation model; b) Temperature distribution; c) Equivalent stress according to von Mises.

If all results are compared accordingly, it can be concluded that plastic deformation in the area of the cooling channel partitioning can already be assumed to be very probable in the first cycle. In this context, all comparable alloys considered tend to fail in the critical range with the material characteristic values from the literature. The measured material characteristic value from the uniaxial cylinder compression tests is below the calculated equivalent stress according to von Mises. Accordingly, there is a high probability that plastic deformation will occur locally within the first forming cycle during the complete forming travel. This is also in line with the conclusion reached in the comparison with other grades of stellite.

In another numerical investigation, the cooling effect in one cycle was analyzed. For this purpose, a thermal simulation was carried out in which the components of the upper tool were modeled as rigid with thermal conduction and a substitute body for the cooling medium was introduced. The same maximum temperatures were calculated in the surface region of the mandrel as in the thermo-mechanically coupled simulation without cooling effect. Due to the cooling effect in the cooling channels, the depth of the heat input could be reduced. The area with the set initial temperature could be increased in the direction of the mandrel surface from the lowest point of the cooling channels by 0.2 mm (6.2% of the total distance). This clearly shows that the use of a cooling medium has a positive effect, even with a short exposure time of 0.14 s (process time) and can dissipate the introduced heat. In the experimental forging cycles, a longer exposure time, which is applied both during forming and during the transfer time of the semi-finished product, is expected to have a greater effect.

### 3.5 LOAD AND FORGING EXPERIMENTS

It became clear from the numerical investigations that further experiments should be carried out to evaluate the tool load before the serial forging tests. For example, in particular in run-in tests, the occurring tool load should be checked by a gradual increase of the stroke in order to avoid damage to the tool (cf. Fig. 7). In addition, the following should be considered when classifying the simulation results: Although there is a risk of plastic deformation in the first cycle, the highest stresses are located in a localized area inside the die. Due to the locally limited critical areas, it can be assumed that the plastic material behavior will not occur due to the lack of flow possibilities. Further-more, a positive effect due to internal cooling on the die temperature has been calculated. This effect is expected to be particularly effective over the course of several forg-ing cycles.

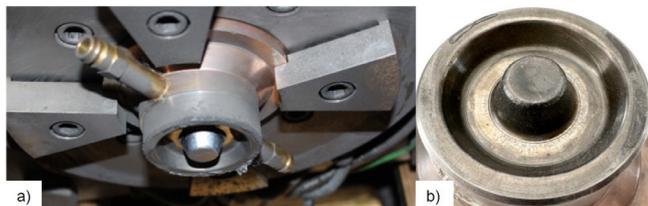


**Fig. 7:** Forging tool 1 made of Celsit 21; a) Before the load tests; b) After the load tests; c) Forcedisplacement curve at minimum flash height.

A total of three tools were tested. Tools 1 and 2 were the hybrid variants, and tool 3 was a conventional specimen made from quenched and tempered X38CrMoV5-1 with a hardness of  $48 \pm 2$  HRC. For all tests, the forming temperature was  $1,200 \text{ }^\circ\text{C}$ , the ram force 500 kN, and the burr path height 3.4 mm. Forging with 500 cycles was performed with all three variants, and only tool 2 was operated with cooling. Fig. 8 shows tool 2 as installed in the machine and following the tests after 500 strokes.

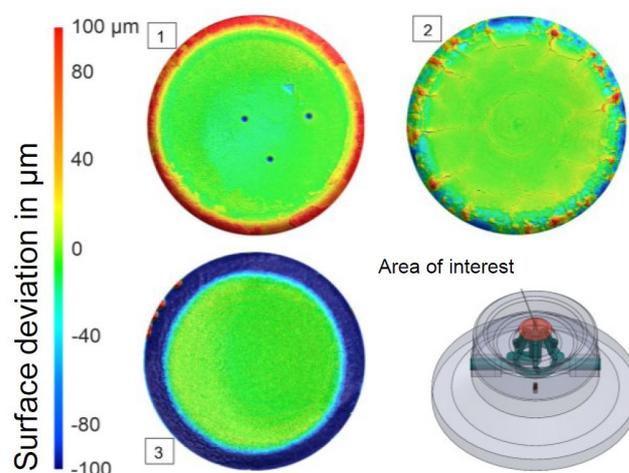
The use of tool 1 made of Celsit 21 allowed higher forming of the cylindrical semi-finished products, without the onset of failure. The maximum load applied in this test occurred at a measured ram force of 500 kN. The tool showed no crack initiation and thus suitability for forging applications.

The surfaces of the tools were measured before and after the tests on the VR-3200 3D profilometer from Keyence (cf. Fig. 9). The deviations in the range of 0.01 mm can be attributed to a remaining measurement inaccuracy and positioning deviation. Tool 1 shows no impairment of the surfaces.



**Fig. 8:** Forging tool 2; a) In the installed state in the machine; b) After 500 strokes.

The planned tempering and compression tests were also carried out with tool 1. Here, a medium temperature of  $30 \text{ }^\circ\text{C}$  was started and successively increased by  $10 \text{ }^\circ\text{C}$  until  $180 \text{ }^\circ\text{C}$  and 18 bar medium pressure were reached. Distilled water was circulated as the medium within the cooling system over a period of 1 h, and no leakage occurred. Nevertheless, the first tests were run without flow of the cooling medium to prevent uncontrolled leakage. The tool was then loaded in the forging process with two press strokes on a Eumuco press. During the subsequent application of media pressure, leakage occurred due to water escaping along the brazed seam.



**Fig. 9:** Load diagram; a) Tool 1, additively manufactured without cooling; b) Tool 2, additively manufactured with cooling; c) Tool 3, conventionally manufactured without cooling; d) Investigation area.

The die stability was still intact, and no breaks were visible on the engraving. The die did not detach from the carrier, so tool 1 continued to be used. Subsequent forging tests continued without internal cooling due to the leakage. A total tool life of 500 strokes was achieved with tool 1. Lateral misalignment was observed after forging. Fixation was provided only by the dowel pins. It can be assumed that the misalignment caused by transverse forces during forging led to failure of the joint after the first few strokes and to leakage during cooling.

The die withstood the applied forging loads and did not break in the area of the flash track. This indicates sufficient strength of the material and the suitability of the composite material. The maximum depth of removal for the additively manufactured dies is significantly less than for the hot-work tool steel, which favors the insertion of the channel structure near the surface.

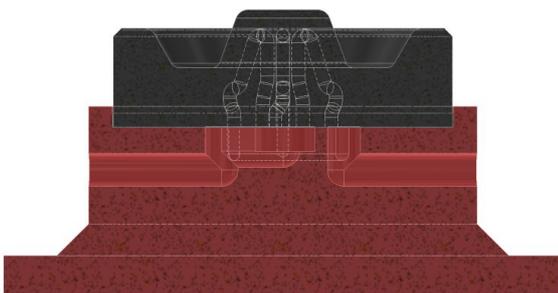
Tool 1 shows a pure increase in the profile height in the radius area, without any negative wear into the interior of the die being evident. The course of the surface profiles at tool 3, however, clearly indicates a dominant erosion within the stressed area. At tool 2, as already seen in the three-dimensional view, both directions of action are apparent. However, this shows the least impairment of the die surface in the mandrel radius. The cooling, which was applied to this tool continuously over the entire period of use, shows a positive effect, both in preventing the widening of the crack network and structural volume and through reduced wear depth.

#### 4 DISCUSSION AND CONCLUSIONS

The channels run below the stressed crown areas and have a rough surface. No particle deposits or deposits of the cooling medium can be observed in the channel passages. They do not show any visual signs of the propagation of a crack network below the surface, which could indicate a leakage cause. The wear phenomena are therefore far away from the areas of the cooling structures that cannot be detected with the naked eye. Based on this, it can be assumed that the cooling structures can continue to operate intact, even with multiple forging cycles. Reducing the distance between the cooling channels and the tool surface can therefore be considered useful in terms of improved tempering effect.

In the new concept, the engraving is embedded in a setting in the carrier (cf. Fig. 10). While soldering is still intended for joining, and the solder paste can also be added at the edges, this design allows for compensation of possible lateral forces. In the case of the layout implemented in the project, these forces can potentially lead to shearing of the engraving from the carrier. The manufacturing effort is only minimally increased in the implementation of this concept during the production of the carrier. In terms of cooling performance, reducing the channel spacing to the die surface layer, and increasing the channel surface area can serve the further optimization of the cooling effect and can be adapted to the individual cycle times of the industry and the corresponding heat input.

The properties of Celsit 21 demonstrate excellent suitability for forging applications. In addition to its resistance to the tempering effect typical of hot work steels, effective hardening occurs in the edge area of the highly stressed dies due to material compaction. Due to manufacturing reasons, the Celsit 21 dies have pores, which are closed near the surface during forging. Cracks in the crown area reduce the surface quality and can lead to breakouts with further use. The crack depth correlates with the applied mechanical load in the examined variants.



**Fig. 10:** Concept for further improvement to better absorb lateral forces.

It has also been shown that die tempering and forging temperature have a significant influence on tool wear behavior. To prevent critical crack formation, the use of additively manufactured dies with integrated cooling channels offers a new technological approach. Economic implementation could be achieved by reducing the volume of additively manufactured Celsit 21. In comparison to tool 3 made of hot work steel, all examined dies show a lower wear depth and thus suitability for die forging.

In summary, the results show that the use of stellite as die material in LPBF represents a promising opportunity to increase the tool life of forging dies and thereby reduce the piece costs of forged parts. The hybrid structure of the tool, consisting of an additively manufactured engraving and conventionally manufactured carrier, enables customized supply and removal of the cooling medium and can contribute to a more economical production of forged parts. The conducted experiments have shown that the technology can also be successfully applied under industry-like conditions. For a perfect transfer to industry, it would be necessary to produce larger tools that are also designed in a hybrid shape. While the conventional production of the carriers using milling techniques is trivial, the challenge will be the production of additively manufactured components including cooling channels on a large scale. The challenges in detail will be the size of the printing platform, the temperature management in order to generate crack- and pore-free parts, and the duration for the printing job. While the technical issues might be solved by appropriate machines, the cost-effectiveness is another topic to consider. Nevertheless, the findings of this work can be utilized not only by the forging industry but also by other medium-sized industries such as tool manufacturers, foundries, and 3D printing companies. Overall, the findings represent an important step towards more cost-effective manufacturing.

#### ACKNOWLEDGMENTS

The research project IGF 20773 N „Additiv gefertigte Kühlung von Schmiedegesesenken“ from the Research Association Forschungsgemeinschaft Werkzeuge und Werkstoffe e.V. (FGW), Papenberger Straße 49, 42859 Remscheid, Germany, is supported by the Federal Ministry of Economic Affairs and Climate Action the German Federation of Industrial Research Associations (AiF) as part of the program for promoting industrial cooperative research (IGF) based on a decision by the German Bundestag.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## REFERENCES

- [1] Buchmayr, B.: Damage, Lifetime, and Repair of Forging Dies. BHM Berg- und Hüttenmännische Monatshefte 162, 88–93 (2017)
- [2] Foster, J, Cullen, C., Fitzpatrick, S., Payne, G., Hall, L., Marashi, J.: Remanufacture of hot forging tools and dies using laser metal deposition with powder and a hard-facing alloy Stellite 21®. Journal of Remanufacturing 9, 189–203 (2019)
- [3] Chantzis, D., Xiaochuan, L., Politis, D. J., Shi, Z., Wang, L.: Design for additive manufacturing (DfAM) of hot stamping dies with improved cooling performance under cyclic loading conditions. Additive Manufacturing 37, 101720 (2021)
- [4] Komodromos, A., Kolpak, F., Tekkaya, A. E.: Manufacturing of Integrated Cooling Channels by Directed Energy Deposition for Hot Stamping Tools with Ball Burnished Surfaces. BHM Berg- und Hüttenmännische Monatshefte 167, 428–434 (2022)
- [5] Tang, S.-Y., Yang, L., Fan, Z.-T., Jiang, W.: A review of additive manufacturing technology and its application to foundry in China. China Foundry 18(4), 249–264 (2021)
- [6] Knotek, O., Lugscheider, E., Eschnauer, H.: Hartlegierungen zum Verschleiss-Schutz: Aufbau, Eigenschaften, Verarbeitung, Anwendung. Verlag Stahleisen, Köln (1975)

# Towards a robust automated surface inspection method for CT-scanned cannulas

Lennart Siefke<sup>1</sup> and Anna Linden<sup>2</sup>

<sup>1</sup>Humboldt-Universität zu Berlin, Rudower Chaussee 25, 12489 Berlin-Adlershof, lennart.siefke@hu-berlin.de

<sup>2</sup>Bundesanstalt für Materialforschung und -prüfung, Unter den Eichen 87, 12205 Berlin, anna.linden@bam.de

**Abstract.** For certain cardiovascular diseases, cannulas are implanted into the blood circuit. To match the patients individual anatomy of the heart, there is research for cannulas to be custom-designed and manufactured aided by 3D printing. However, cannulas have to hold very high standards with regard to the smoothness of their surfaces, as rough patches can lead to formation of blood clots. Therefore, this work uses computer vision to detect such patches as part of quality assurance. First, the produced cannula is scanned using a precise CT scanner and transformed into a 3D mesh object. Rough patches in an otherwise smooth but curved surface are detected by using cosine similarity between neighboring faces and a statistical evaluation. In the end, this method is able to raise a warning when curved surfaces are not smooth enough and visualizes the problematic patches. However, there is just limited access to test data currently and the scanner used needs to be upgraded.

## 1 INTRODUCTION

Additive manufacturing represents a significant advancement in the production of personalized medical devices. In the context of 3D-printed injection molds, the surface quality of the mold directly impacts the surface roughness of the resulting injection-molded components like the cannula. Inserting a cannula into the circulatory system is potentially dangerous, as a rough inner surface may result in the development of blood clots. This happens when components of the blood get stuck in indentations in the wall of the cannula and is further supported by slow blood flow. To present numbers for this geometric problem, the smallest components of blood are thrombocytes with a diameter between 1.5  $\mu\text{m}$  and 3  $\mu\text{m}$ . Blood clots can become life-threatening when plugging up blood vessels. Our use case is bound to analyzing cannulas which are tubular objects. However, there are cannulas with different radiuses and bend angles, which need to be supported by our approach.

To tackle these problems, the goals of this work are

1. Use computer vision to quantify the smoothness of the surface of a cannula with an unknown radius.
2. Visualize where harmful indentations in the surface are located.

The short-term application is to validate the surfaces of prototypes and samples. However, in the long term, we would like to validate the surfaces of every printed cannula so a non-destructing method is preferred.

## 2 RELATED WORK

Lavoué [Lav09] segments meshes by roughness to support applications like compression or robust watermarking. The approach includes local smoothing and comparing the original to the smoothed mesh as a rough surface is not very similar to its smoothed surface. Because of the smoothing, the approach is more suitable to detect rough areas of a larger size and is therefore not ideal for our use case.

In [JPO+17], the authors use a 3D camera to scan the surface on the wings of an Airbus 320 to detect defects. The authors estimate normals and curvature values on point clouds and segment defects by a threshold for the curvatures, defined by an expert guess. Whereas the approach seems to be very promising, it only works for plane-like surfaces.

Pauly et al. [PMG04] analyze surfaces of point clouds by how likely it is for a point to be located on the surface of a point-sampled surface. The likelihood of a point to belong to the surface between discrete points is estimated by least-square fitting of interpolated surfaces. Generally, it is possible to use this approach to detect outliers. However, the detection of outliers would be influenced by a non-trivial relation between the resolution, the interpolation method, the fitting, and the shape of the object and its defects.

Rabbani et al. [RHV06] segment point clouds by smoothness. The key idea is to specify an allowed threshold for the difference between angles of neighboring surface normals and perform region growing. Defects can be segmented using their approach, but setting the parameters is difficult for our use case as the threshold for the angle between neighboring normals depends on the scan resolution and radius of the cannula.

### 3 MATERIAL UNDER TEST

The mold for the test cannula design was fabricated using a selective laser sintering (SLS) process with the *EOS Formiga P100*, employing *EOS PA2200* material. The current layer height is 100  $\mu\text{m}$  with the possibility of reducing it to a minimum of 50  $\mu\text{m}$  in the future. To eliminate excess powder, the surface underwent a cleaning process utilizing the Powershot system from DyeMansion. Subsequently, the three-piece mold was assembled and filled with the heat-curable liquid silicone rubber (LSR) *NuSil MED 4960*. The LSR was cured by placing the mold in an oven at 90 C for 12 hours. Following demolding, the silicone cannula underwent post-curing at 165 C for 10 minutes. To achieve a smoother surface finish, the cannula was partially coated with an RTV silicone dispersion through a dip-coating process. Fig. 1 presents this cannula. Additionally, there is an earlier prototype made from the same CAD-model. The earlier prototype is called C1, the non-coated part of the current prototype is called C2 and the coated part is called C2C. For a fair comparison, every cannula is sliced into just a small tube so C1, C2 and C2C have the same shape.

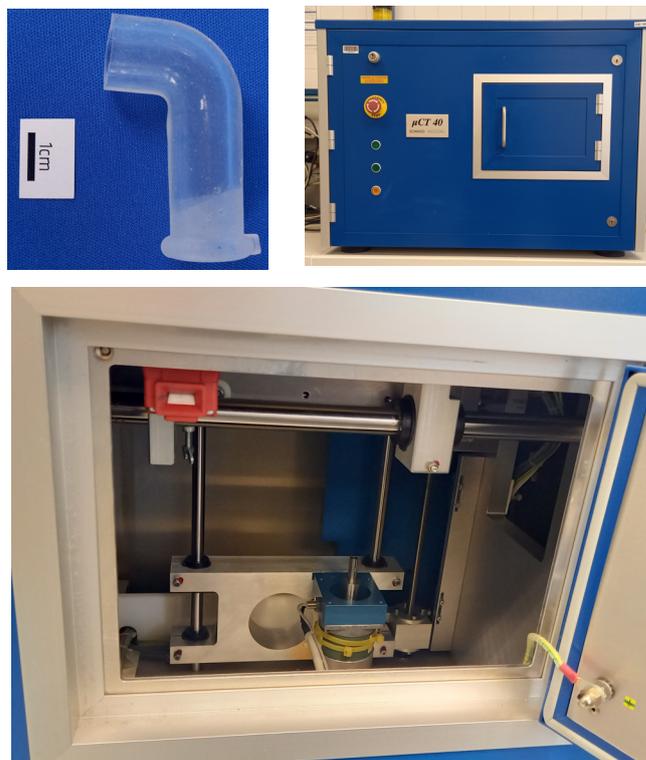
## 4 APPROACH

There are various methods to scan surfaces of objects and analyze their smoothness, with different pros and cons. The method for analyzing the surface depends on the technique to scan the cannula.

### 4.1 RELEVANT 3D-SCANNING TECHNIQUES

An obvious choice to scan surfaces is to use a profilometer. However, optical methods like structured light scanning or interference microscopy require a non-transparent material or proximity to the surface. To access the inner surface of the cannula, it would be necessary to cut the cannula open. Thus, such methods are unsuitable for this application. Computed tomography (CT) scanning is an advanced imaging technique that employs X-rays to visualize internal structures, primarily for medical applications. Micro-computed tomography ( $\mu\text{CT}$ ) is a specialized variant of CT that offers higher resolution imaging, characterized by voxel sizes in the micrometer range. Due to its typically smaller maximum specimen size compared to conventional CT systems,  $\mu\text{CT}$  is primarily employed for material specimens and small biological samples.  $\mu\text{CT}$  offers high precision with non-destructive scanning and is therefore the chosen technique for our application.

### 4.2 $\mu\text{CT}$ -SCAN



**Fig. 1:** Cannula made with liquid silicone rubber and the Scanco  $\mu\text{CT}$ -Scanner to scan it. The transparent part of the cannula is coated to smooth the surface.

The cannula was scanned using the Scanco  $\mu$ CT 45 (see Fig. 1) at 70 kV and 114  $\mu$ A with a resulting voxel size of 18  $\mu$ m. Although the voxel size already yields high precision, even higher precision would be necessary to detect notches relevant for thrombocytes to get stuck. Nonetheless, the voxel size is sufficient to detect undesired artefacts generated by additive manufacturing with a layer height of 100  $\mu$ m. To detect smaller structures, a higher resolution is needed. The quality of a 3D scan is always a trade-off between the resolution and the field of view (FOV) due to geometrybased magnification. A larger FOV comes with a smaller resolution and vice versa. To maximize the resolution, the object-to-source distance needs to be minimized. This distance is limited by the size and geometry of the object. For the cannula it would be possible to have an overview scan of the whole specimen with a reasonable resolution to analyze the surface structure and then scan regions of interest with a maximum resolution. This functionality is supported by PXR CT-COMPACT nano for example. With these information, both a general and a few specific statements about the surface quality of the cannula could be generated. Using the precisely scanned ROIs, a conclusion can be drawn to the whole surface. However, the used scanner is not capable of this function. Thus, this work focuses on the data analysis approach and algorithm.

For further evaluation, a surface mesh is desired. Compared to a point cloud, a surface mesh has the benefit of information about neighborhood relationships. The CT provides DICOMs, which are stacks of 2D-images. Silicon and air result in very different pixel values and are easy to separate. Using 3D-Slicer [KPV14,BWH24], a surface mesh is exported.

### 4.3 ROUGHNESS OF ADJACENT FACES

Common definitions for roughness like the roughness average  $R_A$  require an expected value for each measured point, which is usually defined by a mean line through the measurements. The surfaces of cannulas are curved and may be bent. Whereas a plane fitting could yield the expected value for each point, it would introduce some error and supporting a variety of shapes would be difficult. Therefore, we focus on other methods. To quantify smoothness of a surface, the first step is to quantify smoothness in adjacent faces in the surface mesh. A face is the smallest surface component of a mesh, usually a triangle, and is flat by definition. Roughness can only exist between adjacent faces. Therefore, we compare the normal vectors of adjacent faces as they tell in which direction a face is pointing. The cosine similarity  $S_C$  is

a metric which tells how much two vectors point into the same direction. It is calculated similar to the angle between two vectors with the difference of not applying the inverse cosine; the dot product of the vectors is divided by the product of the magnitude of the vectors:

$$S_C(\mathbf{n}_1, \mathbf{n}_2) = \frac{\mathbf{n}_1 \cdot \mathbf{n}_2}{\|\mathbf{n}_1\| \cdot \|\mathbf{n}_2\|} \quad (1)$$

The domain of the cosine similarity is  $[-1; 1]$ . A value of 1 tells that two vectors are coplanar and pointing into the same direction, a value of 0 indicates a difference of  $90^\circ$  and a value of -1 tells that two vectors are coplanar but pointing into opposite directions. The optimal cosine similarity between normals of faces on a curved surface depends on the radius and the resolution of the scanner. Theoretically, an infinitely small voxel size results in almost coplanar adjacent faces. The cosine similarity of their normals would approach 1 without ever reaching 1. As a rule of thumb, a high resolution scan of a tube with a small diameter results in cosine similarities close to 1. As it is ambiguous for two normals to have a cosine similarity of -1 or 1, the absolute is applied to  $S_C$  to define the roughness  $r$ .

To reason about the roughness of a face, the face has to be compared to its 1 to  $n$  adjacent faces. Therefore, the roughness  $r_i$  of a face  $i$  is modeled by the minimum of the absolute cosine similarity of its normal to all the normals of its adjacent faces  $A_i$ :

$$r_i = \min_{\forall j \in A_i} |S_C(\mathbf{n}_i, \mathbf{n}_j)| \quad (2)$$

### 4.4 DETERMINING ROUGHNESS OF SURFACES BY STATISTICAL EVALUATION

The surface of a cannula is expected to be mostly smooth enough, even if defects are present. This assumption allows using a statistical approach to analyze the cosine similarities of neighboring face normals and to separate outliers. The roughnesses  $R$  of each face are expected to be left skewed, as they are capped to 1 by definition of the cosine similarity. For skewed distributions, it is common to call a sample an outlier when it has a distance to the boundary of the interquartile range (IQR) of 1:5 IQR [Dek05]. The IQR itself is defined as the difference between the 0:75-quantile and the 0:25-quantile. Building on this, the faces of the mesh are separated into outliers  $O$  and inliers  $I$ :

$$\begin{aligned} O &= \{r \in R \mid r < q_n(0.25) - 1.5 \cdot IQR \vee r > q_n(0.75) + 1.5 \cdot IQR\} \\ I &= R \setminus O \end{aligned} \quad (3)$$

Ideally, the outliers  $O$  build an empty set. This separation of in- and outliers adapts to the input mesh and allows finding outliers without much knowledge of the resolution or shape of the object. As the method does not require comparing the surface with another object, the approach can be categorized as a no-reference method.

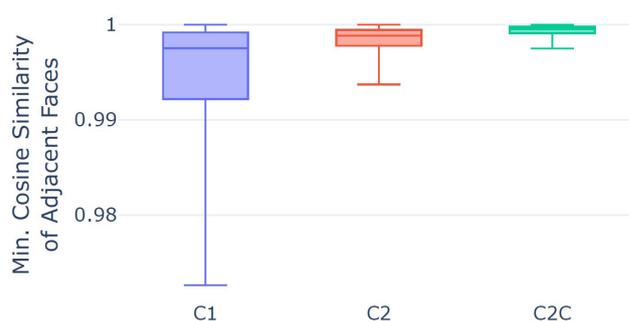
## 5 RESULTS

Performing the statistical evaluation on the three cannulas validates the expectations made in 3: C2C is smoother than C2, which is smoother than C1. This can be seen in the average, median and standard deviation presented in Table 1. Comparing average and medians is difficult, however, as it is neither known which minimal value describes a rough tube nor which value describes a perfect tube given the scan resolution and radius. On the other side, comparing the spread of the inliers is meaningful. The standard deviation of C2 is 2.5 times higher than the standard deviation of C2C and the standard deviation of C1 is even 11.2 times higher than the standard deviation of C2. Reference metrics for each cannula may be gained by manufacturing and evaluating multiple cannulas of the same type which will be investigated in further work. Fig. 2 presents the distribution of roughness in the smooth part of the surfaces of the cannulas.

Inliers of	In-/Outlier Threshold	Number of Outliers	Average	Median	$\sigma$	$\frac{\sigma}{\sigma_{C2C}}$
C1	0.9726	11.2%	0.9947	0.9975	0.0062	11.2
C2	0.9937	10.2%	0.9984	0.9988	0.0014	2.5
C2C	0.9975	9.9%	0.9994	0.9995	0.0005	1

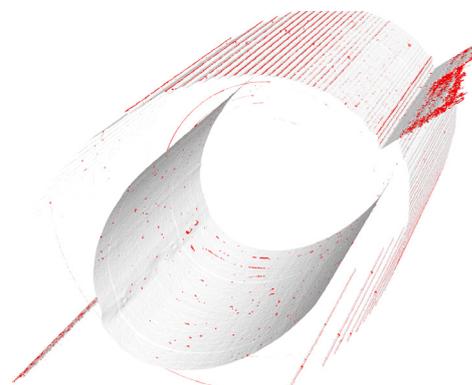
**Tab. 1:** Statistical evaluation performed on the three cannulas. Note that the division  $\frac{\sigma}{\sigma_{C2C}}$  is performed on the non-rounded values.

Fig. 3 presents the visualization of the outliers on the cannulas and a close up view on the backside of the inner wall of C2C to grab a better view on the outliers. The outer surface of C1 is covered by stair-stepping which is detected properly. C2 is more smooth but has a few defects which are also detected. The detection of rough patches on C2C however is a little problematic as lots of minor rough patches are detected on a very smooth surface. A close look on these patches reveals the inner surface to be covered by those patches very uniformly. This is a hint for a bad signal-to-noise ratio, especially as the outer surface with the smaller curvature is less effected but in a similar pattern. The surface seems to be smooth enough to be influenced by the noise of the  $\mu$ CT or the processing chain. These patches also explain why the relative number of outliers of C2C is the lowest by just a small margin. Interestingly, the statistical evaluation is robust enough to tell that C2C is the smoothest cannula.

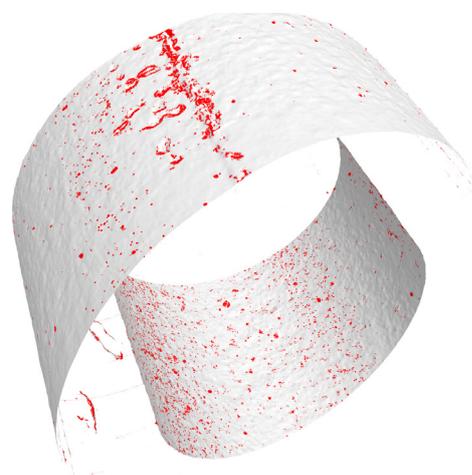


**Fig. 2:** With the tubular shape and a high scan resolution, the distribution of the roughness of inliers is left skewed and close to 1. The coated cannula has by far the least spread of its smooth faces.

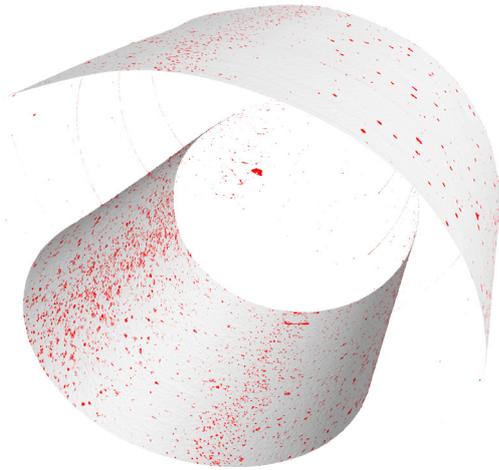
Sadly, the precision of the scanner or the processing chain is not sufficient for a proper detection of minor defects on the coated surface. The pipeline from raw CT-data to the surface mesh relies on external software and algorithms. As further processing depends strongly on the angle between adjacent faces, more control over the export to the surface mesh could be beneficial to handle noise.



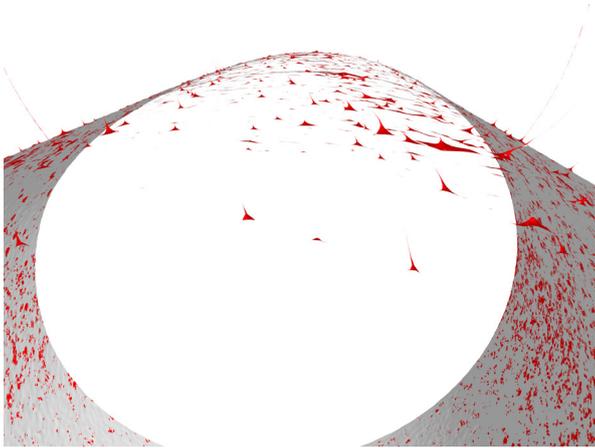
(a) Cannula C1



(b) Cannula C2



(c) Cannula C2C



(d) Cannula C2C

**Fig. 3:** Visualization of rough patches in red on a segment of the scanned cannulas. (d) shows evenly distributed minor outliers on the inner wall which are a hint for a bad signal-to-noise ratio.

## 6 CONCLUSION

In this paper, we presented an approach to quantify smoothness and detect rough patches in a mesh which requires low knowledge about the shape. The only assumption about the surface is that it is mostly smooth. To be able to detect defects on inner surfaces without destroying the object,  $\mu$ CT was used to scan the object. The processing chain from raw  $\mu$ CT-data to the mesh needs more investigation though and a more precise scanner would be beneficial.

## ACKNOWLEDGEMENTS

This work was supported by the IBB AMBER-ProFIT program and co-financed by the European Regional Development Fund (EFRE).

## REFERENCES

- BWH24.** BWH and 3D Slicer contributors. 3D Slicer image computing platform. <https://slicer.org/>, 2024.
- Dek05.** Michel Dekking, editor. *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Springer Texts in Statistics. Springer, London, 2005.
- JPO+17.** Igor Jovancevic, Huy-Hieu Pham, Jean-José Orteu, Rémi Gilblas, Jacques Harvent, Xavier Maurice, and Ludovic Brèthes. 3D Point Cloud Analysis for Detection and Characterization of Defects on Airplane Exterior Surface. *Journal of Nondestructive Evaluation*, 36(4):74, October 2017.
- KPV14.** Ron Kikinis, Steve D. Pieper, and Kirby G. Vosburgh. 3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support. In Ferenc A. Jolesz, editor, *Intraoperative Imaging and Image-Guided Therapy*, pages 277–289. Springer, New York, NY, 2014.
- Lav09.** Guillaume Lavoué. A local roughness measure for 3D meshes and its application to visual masking. *ACM Transactions on Applied Perception*, 5(4):1–23, January 2009.
- PMG04.** Mark Pauly, Niloy J. Mitra, and Leonidas J. Guibas. Uncertainty and variability in point cloud surface data. In *Eurographics Symposium on Point-Based Graphics*, volume 84, 2004.
- RHV06.** T. Rabbani, F. V. Heuvel, and G. Vosselman. SEGMENTATION OF POINT CLOUDS USING SMOOTHNESS CONSTRAINT. 2006.

# Schwingform-Projektion zur Berechnung von Ähnlichkeitsmaßen

Kai Henning

Gesellschaft zur Förderung angewandter Informatik, Abteilung Strukturmechanik/Mustererkennung, Volmerstraße 3, 12489 Berlin, Deutschland

**Abstract.** In der Strukturmechanik gewinnen bildbasierte Verfahren zur Schwingformmessung an Bedeutung, da sie kontaktlos Schwingformen aus Videodaten erfassen können. Ein Vergleich mit numerischen Vorhersagen aus Finite-Elemente-Berechnungen (FEM) ist notwendig, stellt jedoch eine Herausforderung dar, da die Datensätze im gleichen lokalen Koordinatensystem vorliegen müssen. Diese Arbeit präsentiert einen Ansatz zur Transformation von 3D-FEM-Schwingformdaten, um sie mit experimentellen Daten zu vergleichen. Die Methodik wurde an mehreren Datensätzen getestet und zeigte vielversprechende Ergebnisse bei Ähnlichkeitsmaßen wie dem Modal Assurance Criterion (MAC) und der Normalized Cross-Correlation (NCC).

**Keywords.** Projektion, Schwingform, Modalanalyse, GPGPU

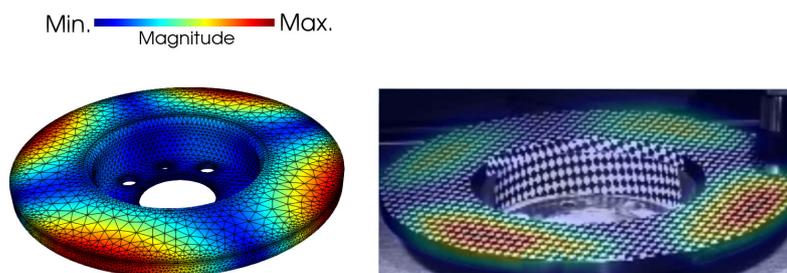
## 1 EINFÜHRUNG

Für mechanische Systeme können Schwingformen sowohl experimentell als auch numerisch erfasst werden. Durch die Kombination experimenteller Messungen und numerischer Simulationen ist eine umfassende Untersuchung der dynamischen Eigenschaften solcher Systeme möglich. Für diese Kombination ist es jedoch erforderlich, die gemessenen Daten mit den Simulationsergebnissen zu vergleichen. Damit dieser Vergleich mit traditionellen Ähnlichkeitsmaßen durchgeführt werden kann, muss eine Kompatibilität der Daten gewährleistet sein. Im Allgemeinen ist eine Schwingform wie folgt definiert:

**Definition 1.** Eine Schwingform, auch Eigenmode oder Modalfunktion genannt, ist eine vektorwertige Funktion  $\mathbf{u} : \Omega \subset \mathbb{R}^3 \rightarrow \mathbb{R}^3$ , die die räumliche Verteilung der Schwingungen eines dynamischen Systems beschreibt. Hierbei ist  $\Omega$  der betrachtete Körper, und  $\mathbf{u}(p)$  gibt die Verschiebung in jedem Punkt  $p = (x_p, y_p, z_p) \in \Omega$  an. Eine Schwingform ist mit ihrer Eigenfrequenz  $\omega$  verbunden, die die Frequenz beschreibt, bei der  $\Omega$  in dieser Form schwingt.

## 2 DATENSTRUKTUREN

Im folgenden beschreiben wir die mathematische Struktur der beiden Datensätze, die es zu vergleichen gilt.



**Abb. 1:** Links: Visualisierung einer Schwingform, die für eine Autobremsscheibe mit der FEM berechnet wurde. Rechts: Entsprechende Schwingform der bildbasierten Messung der Autobremsscheibe.

## 2.1 FEM - SCHWINGFORMBERECHNUNG

Die Schwingformberechnung mit der Finite-Elemente-Methode (FEM) ermöglicht die Analyse der dynamischen Eigenschaften eines Körpers  $\Omega \subset \mathbb{R}^3$  auf Grundlage eines detaillierten CAD-Modells. Die Schwingformberechnung erfolgt auf einer diskreten Menge  $\mathbf{V} \subset \Omega$ , die aus dem CAD-Modell generiert wird. Die Schwingform wird dann durch die Menge

$$\mathbf{U} = \{\mathbf{u}_i = \mathbf{u}(p_i) \mid p_i \in \mathbf{V}\} \quad (1)$$

ausgedrückt. Zusätzlich wird die Menge  $\mathbf{V}$  durch eine Tetraederisierung aus Tetraedern  $T_j$  strukturiert, die die Geometrie des Körpers approximieren:  $\Omega \approx \bigcup_{j \in J} T_j$ . Jeder Tetraeder  $T_j$  wird durch seine vier Eckpunkte definiert und hat vier Flächen, die durch die Triangulierungen  $\{S_{j1}, S_{j2}, S_{j3}, S_{j4}\}$  gegeben sind, wobei jede Fläche  $S_{jk}$  ein Dreieck ist, das durch drei der vier Eckpunkte des Tetraeders gebildet wird. Die Oberfläche  $\partial\Omega$  lässt sich dann durch eine Triangulierung aus Dreiecken  $S_k$  darstellen, die eine Teilmenge der Tetraederisierung ist:

$$\partial\Omega \approx \mathbf{S} = \bigcup_{k \in K} S_k \subset \bigcup_{j \in J} T_j. \quad (2)$$

## 2.2 BILDBASIERTE SCHWINGFORMMESSUNG

Schwingformmessungen die auf Videodaten basieren, messen die Schwingformen in der Bildebene  $E \subset \mathbb{R}^2$  der Kamera und können diese daher nur in den lokalen Koordinaten  $(x, y) \in E$  ausdrücken. Nach der Analyse eines Videos mit der Auflösung  $w \times h$  erhalten wir einen Tensor  $\mathbf{I}$  der Dimension  $w \times h \times 3$ . Dieser Tensor repräsentiert an seinen Einträgen  $(c, r)$ , die in die Bildebene projizierte Schwingform. Die Beziehung zwischen den Pixelindizes und den Schwingformwerten wird durch eine Gleichung der Form

$$\mathbf{I}_{\iota(p)} = \mathcal{M}\mathbf{u}(p) \quad (3)$$

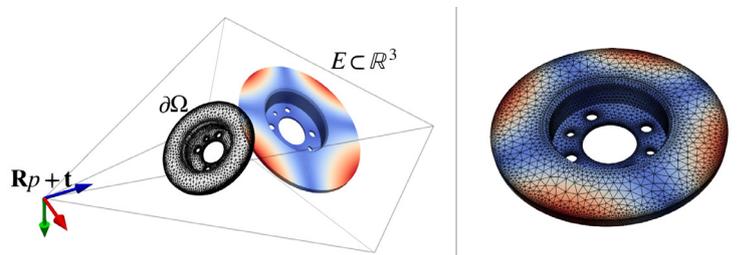
beschrieben. Hierbei ist  $\iota : p \rightarrow (c, r)$  eine projektive Abbildung die Punkte  $p \in \mathbb{R}^3$  in Pixelindizes  $(c, r) \in w \times h$  abbildet und  $\mathcal{M}$  eine lineare Transformation die  $\mathbf{u}$  aus dem lokalen Koordinatensystem auf der Oberfläche von  $\Omega$  in die Bildebene  $E$  transformiert. Konkrete Setups und Messungen mit dieser Methode beschreiben wir in [3].

## 3 SCHWINGFORMPROJEKTION

Das Ziel besteht darin, die Darstellung aus Kapitel 2.1 in die Form (3) zu überführen. Dafür müssen die Geometriedaten aus der FEM mit den Kameraparametern der bildbasierten Schwingformmessung in Beziehung gesetzt werden.

Dazu wird die Kamera als ane Raum  $\mathbb{A}^3$  im  $\mathbb{R}^3$  aufgefasst. Der Ursprung wird durch den Translationsvektor  $\mathbf{t} \in \mathbb{R}^3$  festgelegt, während die Orientierung durch die Rotationsmatrix  $\mathbf{R} \in \mathbb{R}^{3 \times 3}$  definiert ist. Die Größen  $\mathbf{t}$  und  $\mathbf{R}$  repräsentieren die extrinsischen Parameter der Kamera und beziehen sich relativ zu  $\partial\Omega$ . Diese Parameter beeinflussen nicht direkt die Projektion, bestimmen jedoch, welche Punkte  $p \in \partial\Omega$  in  $E$  abgebildet werden.

Zur Bestimmung von  $\mathbf{t}$  und  $\mathbf{R}$  kann eine Homographie zwischen  $E$  und  $\partial\Omega$  berechnet werden. Hierfür sind vier Punktepaare  $(p_i, q_i) \in E \times \partial\Omega$  für  $i = \{1, 2, 3, 4\}$  erforderlich, aus denen eine lineare Abbildung berechnet wird [5, tutorial\_homography]. Alternativ kann der Benutzer die Werte für  $\mathbf{t}$  und  $\mathbf{R}$  interaktiv eingeben, indem er die 3D-Geometrie aus der FEM-Berechnung manuell so ausrichtet und verschiebt, dass die Konguration wie in Abb. 2 erreicht wird.



**Abb. 2:** Links: Kameraframe,  $\partial\Omega$  und die Bildebene  $E$  so angeordnet im  $\mathbb{R}^3$ , dass die Abbildung rechts durch Projektion berechnet werden kann.

Die projektive Abbildung der Szene auf die Bildebene wird dann durch die intrinsischen Parameter der Kamera festgelegt. Um die Abbildung der 3DPunktkoordinaten in die Pixelindizes zu bestimmen, betrachten wir den projektiven Raum  $\mathbb{P}^2$ , dessen Elemente alle Geraden durch den Ursprung in  $\mathbb{R}^3$  repräsentieren, formal definiert man  $\mathbb{P}^2 := (\mathbb{R}^3 \setminus \{0\}) / \sim$ , wobei  $(x_1, y_1, z_1) \sim (x_2, y_2, z_2)$  gilt, wenn  $\exists \lambda \in \mathbb{R} \setminus \{0\}$  so dass  $(x_2, y_2, z_2) = \lambda \cdot (x_1, y_1, z_1)$  erfüllt ist. Die Projektion der Objektoberfläche  $\partial\Omega$  auf  $\mathbb{P}^2$  ist dann gegeben als:

$$\text{proj}(\partial\Omega) = \{\ell \in \mathbb{P}^2 \mid \exists (x_1, y_1, z_1) \in \partial\Omega : \ell = \{\lambda \cdot (x_1, y_1, z_1) \mid \lambda \in \mathbb{R} \setminus \{0\}\}\}.$$

Da diese Projektion nicht injektiv ist, denieren wir die sichtbare Projektion durch  $V(\text{proj}(\partial\Omega)) = \{(x, y, z) \in \partial\Omega \mid \exists \ell \in \text{proj}(\partial\Omega) : (x, y, z) = d(\ell)\}$ , wobei  $d(\ell) = \arg \min_{(x, y, z) \in \ell} \|(x, y, z)\|$ . Da in unserer Anwendung von rektifizierten Bildern ausgegangen wird, muss die Linse und deren Verzerrung nicht modelliert werden. Das Lochkamera-Modell  $P : p \rightarrow p/z_p D, p \in \mathbb{R}^3$  mit  $z_p \neq 0$  und  $D > 0$  ist daher ausreichend, um die Projektion auf die Bildebene anhand von Sichtstrahlen zu modellieren. Die Brennweite wird dabei über den Parameter  $D$  und die  $z$  Achse als Sichtachse festgelegt. Da der Bildsensor aufgrund seiner Dimensionierung, d.h. Breite  $w$  oder Höhe  $h$ , begrenzt ist wird das Sichtfeld durch die Brennweite und Dimensionierung eingeschränkt. Das horizontale Sichtfeld lässt sich mit  $\alpha = 2 \tan^{-1}(\frac{w}{2D})$  bestimmen.[1] Aufgrund der Symmetrie des Sichtkegels ergeben sich die Grenzen für das Sichtfeld zu  $x_{\min} = -x_{\max}$  und  $y_{\min} = -y_{\max}$ . Die reellen Koordinaten  $(x_{\text{proj}}, y_{\text{proj}}) := (\frac{x_p}{z_p} D, \frac{y_p}{z_p} D) \in (x_{\min}, x_{\max}) \times (y_{\min}, y_{\max})$  werden auf ganzzahlige Pixelkoordinaten gerundet :

$$c = \left\lfloor \frac{x_{\text{proj}} + x_{\max}}{2x_{\max}} \cdot (w - 1) \right\rfloor, r = \left\lfloor \frac{y_{\text{proj}} + y_{\max}}{2y_{\max}} \cdot (h - 1) \right\rfloor$$

Damit erhalten wir eine bijektive Abbildung :

$$\iota : \Lambda(\text{proj}(\partial\Omega)) \mapsto w \times h, \\ p \rightarrow (c, r)$$

mit der wir Simulationsergebnisse an den Knoten  $p$  den Pixeln  $c, r$  zuordnen können. Da  $\mathbf{u}(p)|_{\partial\Omega} \in N_{\partial\Omega} \oplus T_{\partial\Omega}$  für  $p \in \partial\Omega$  und  $\mathbf{R}$  eine orthonormale Matrix ist, vereinfacht sich die Transformation  $\mathcal{M} : T_{\partial\Omega} \oplus N_{\partial\Omega} \mapsto E, \mathbf{u}_p \rightarrow I_{i(p)}$  zu

$$\mathbf{M}\mathbf{u}(p) := \mathbf{R}\mathbf{u}(p) \quad (4)$$

[1, Kapitel 10.4.2 Surface Normals].

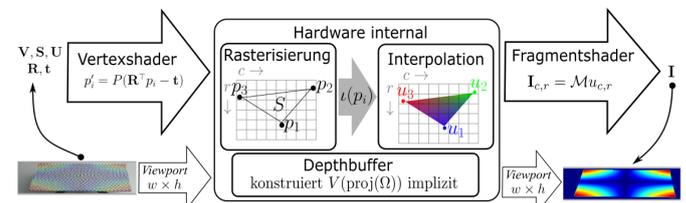
#### 4 IMPLEMENTIERUNG

Für die Implementierung können die Vorteile der programmierbaren Render-Pipeline genutzt werden. Sie ist Bestandteil jeder modernen Grakkarte und kann über verschiedene APIs oder Frameworks angesprochen werden, z.B. mit OpenGL, Direct3D oder VTK. Sie ist äußerst flexibel und kongurierbar und wird vereinfacht in zwei Phasen unterteilt: Vertexshading und Fragmentshading.[1,10.10 The Realtime Graphics Pipeline] Alle  $p_i \in \mathbf{V}$  passieren diese Phasen, und zwar so, dass die Kombinatorik von  $\mathbf{S}$  beachtet wird wenn wir die Renderpipeline für Dreiecke kongurieren.[4]

Im Vertexshader werden Kameratransformationen und Projektionen<sup>1</sup> auf jedes  $p_i \in \mathbf{V}$  angewendet, um die Positionen in der Bildebene  $E$  festzulegen.

<sup>1</sup>  $P$  muss für spezifische APIs spezifische Bedingungen erfüllen die in der jeweiligen Dokumentation spezifiziert sind, z.B. in [4, GluPerspective\_code] oder [1].

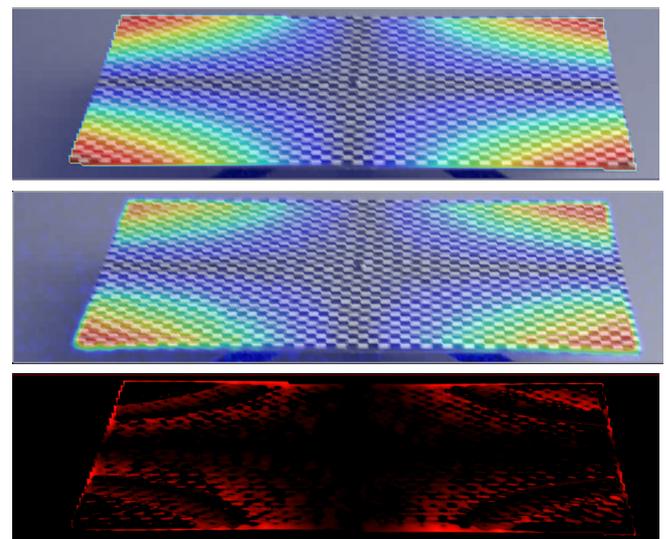
Zwischen den Positionen in  $E$  werden alle Attribute, die mit  $p_i$  assoziiert sind nach dem kongurierten primitiv interpoliert bevor die Attribute an den Fragmentshader übergeben werden. Im Fragmentshading, erhalten die Pixel ihre Farben, indem Farbwerte aus den übergebenen Attributen generiert und an die im Vertexshader bestimmten Positionen in den Framebuffer geschrieben werden.[1,10.10] Der Framebuffer sollte so konguriert werden, dass anstelle von Integer Float Werte gespeichert werden können und  $V(\text{proj}(\partial\Omega))$  implizit durch die Verwendung eines Tiefenbuffers konstruiert wird.[2] Nachdem durchlaufen der Renderpipeline für alle Dreiecke in  $\mathbf{S}$  kann das Ergebnis aus dem Framebuffer ausgelesen werden.



**Abb. 3:** Übersicht der Schwingformprojektion, als Eingabe dienen die 3D Geometriedaten sowie die assoziierte Schwingform  $U$  aus Kapitel 2.1, die Kameradaten  $R$  und  $t$  als auch die Dimensionierung des Bildsensors aus Kapitel 2.2 die als *Viewport* die Dimension des Framebuffers vorgibt.

#### 5 ERGEBNISSE

Wir evaluieren unsere Ergebnisse mit drei verschiedenen Ähnlichkeitsmaßen: Modal Assurance Criterion, Normalized Cross Correlation und einem Dierenzbild.



**Abb. 4:** Der Datensatz mit dem wir unsere Ergebnisse evaluieren. Oben: Projizierte numerisch berechnete Schwingform und mittig: aus Videodaten extrahierte Schwingform, jeweils eingefärbt nach Betrag. Unten, das Dierenzbild der beiden Datensätze.

Das **Modal Assurance Criterion (MAC)** kann verwendet werden, um reellwertige Ergebnisse aus der Simulation mit komplexwertigen Ergebnissen aus einer bildbasierten Schwingformmessung zu vergleichen. Es bildet zwei diskrete Schwingformen  $\mathbf{U}_A$  und  $\mathbf{U}_B$  auf eine reelle Zahl ab,  $\text{MAC} : \mathbb{R}^3 \times \mathbb{C}^3 \mapsto [0, 1]$ , wobei

$$\text{MAC}(\mathbf{U}_A, \mathbf{U}_B) = \frac{|\mathbf{U}_A^T \mathbf{U}_B|^2}{(\mathbf{U}_A^T \mathbf{U}_A)(\mathbf{U}_B^H \mathbf{U}_B)}$$

. Für die in Abb. 4 abgebildete Schwingform erhalten wir nach dem MAC einen Wert von 95%. Die Normalized Cross Correlation (NCC) ist im Wertebereich  $[-1, 1]$  definiert, deswegen verwenden wir für unsere Evaluierung eine normalisierte Form,  $\text{NCC}_{\text{norm}} : \mathbb{R}^3 \times \mathbb{C}^3 \mapsto [0, 1]$  die negative Korrelationen ignoriert

$$\text{NCC}_{\text{norm}}(\mathbf{U}_A, \mathbf{U}_B) = \max \left( 0, \frac{\sum_i (U_{A,i} - \bar{U}_A) \cdot (U_{B,i}^* - \bar{U}_B^*)}{\sqrt{\sum_i |U_{A,i} - \bar{U}_A|^2} \cdot \sqrt{\sum_i |U_{B,i} - \bar{U}_B|^2}} \right)$$

. Für die in Abb. 4 abgebildete Schwingform erhalten wir nach dem NCC einen Wert von 99%. Ein Differenzbild (vgl. Abb. 4)  $\mathbf{D} = \mathbf{U}_A - \mathbf{U}_B$  ist ein Punkt-zu-Punkt-Vergleich und kann auf den Bereich  $[0, 1]$  normalisiert werden:  $\mathbf{D}_{\text{norm}} = |\mathbf{D}| / \max|\mathbf{D}|$ . Während das Differenzbild spezifische Abweichungen in der Bildebene lokalisiert, liefern MAC und NCC als globale Ähnlichkeitsmaße eine quantitative Bewertung der Übereinstimmung zwischen den Datensätzen.

## 6 ZUSAMMENFASSUNG

Wir haben in dieser Arbeit ein Verfahren konstruiert, das es ermöglicht die Ergebnisse numerischer 3D Schwingformsimulationen mit den Ergebnissen bildbasierter Schwingformmessungen über traditionelle Metriken zu vergleichen. Mit diesem Werkzeug ist es nun möglich weitere traditionelle Verfahren, die strukturdynamische Messungen und Simulation verbinden zu nutzen.

## LITERATUR

[1] Dunn, F., Parberry, I.: 3D Math Primer for Graphics and Game Development. Jones & Bartlett Learning (2002)

[2] Göttsche, D.: Gpgpu basic math tutorial (2005), <http://www.mathematik.uni-dortmund.de/goeddeke>, e-Mail: dominik.goeddeke@math.uni-dortmund.de

[3] Herfert, D., Krause, J., Gollnick, M., Henning, K.: Innovative Werkzeuge zur Eigenfrequenzmessung und Eigenformanalyse von Bremscheiben. GFal e.V. (2023)

[4] KhronosGroup: OpenGL wiki. <https://www.khronos.org/opengl/wiki/> (2023)

[5] OpenCV: Homography, <https://docs.opencv.org/4.x/d9/dab/>

# Online Optimization of Stereo Camera Calibration Accuracy

Kai Cordes , Lin Chen , Sönke Südbeck, and Hellward Broszio

VISCODA GmbH  
 {cordes,chen,suedbeck,broszio}@viscoda.com  
 www.viscoda.com

**Abstract.** Depth estimation from stereo cameras is a fundamental Computer Vision task with many applications in automated driving, robotics, scene reconstruction, and medical diagnosis. Depth measurements are based on calibration parameters which are computed in a special calibration procedure. In practice, camera orientations may change slightly over time due to mechanical or thermal effects. Then, calibration parameters no longer match the current stereo camera. This results in inaccuracies in the depth measurement and even complete malfunctions.

We propose a method for the evaluation and optimization of the calibration accuracy during use of a stereo camera. It is based on automatically selected image regions and the rectified stereo configuration. The evaluation monitors the actual accuracy of the calibration. The optimization adjusts the calibration parameters. The time-consuming recalibration using calibration patterns is avoided.

**Keywords.** Accuracy · Calibration · Stereo · Camera · Optimization · Online

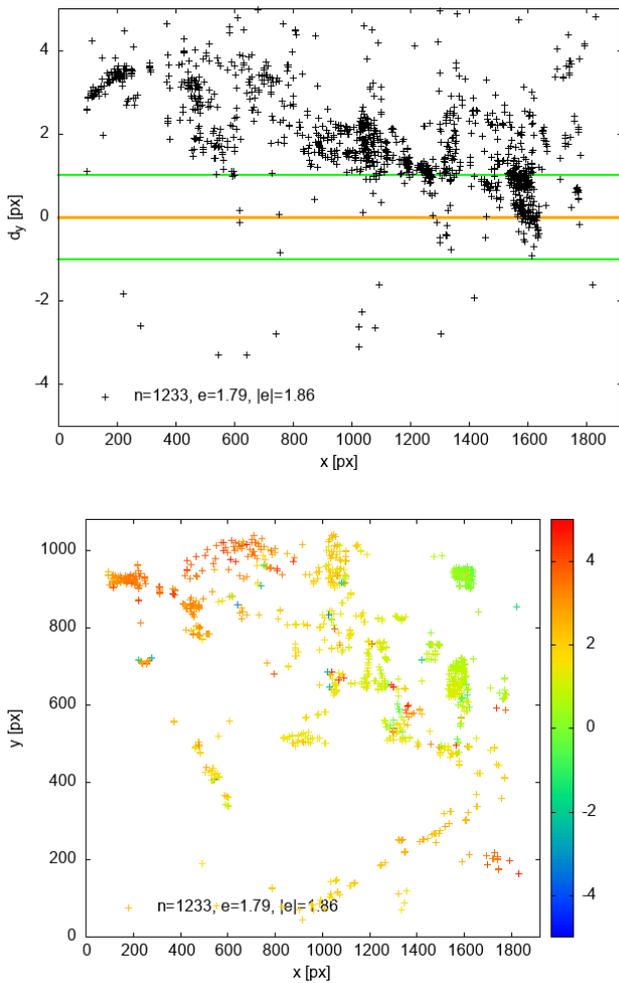
## 1 INTRODUCTION

Stereo cameras are used in a large variety of applications to measure the depth of a 3D scene. For highly accurate measurements, the imaging properties of the stereo camera must be known. These properties are determined in a separate procedure known as calibration using special calibration patterns. Camera calibration is a time-consuming, complicated, and semi-automatic task and there are several approaches for the optimization of the calibration parameters [7,13,6]. The accuracy of the calibrations depends on the respective calibration procedure. However, it has been shown that the relative alignment of the cameras change during use due to vibration and mechanical or thermal effects [3]. Thus, accurate stereo calibration cannot be guaranteed over long periods [14,9]. Even well-established datasets have significant systematic errors resulting from sub-optimal calibration parameters for different recording days. In most cases, the camera orientations are subject to changes during data acquisition and the camera parameters are not adapted.

Small errors in the camera parameters lead to significantly decreased reconstruction accuracy [5,15,9]. For given disparity error  $\Delta d$ , the range uncertainty increases quadratically with distance [2]. If vertical misalignment in the rectified stereo image occurs, corresponding pixels do not have the same y-coordinate (same scanline) and traditional stereo matching algorithms such as SGM [8] provide suboptimal results. Hence, vertical misalignment should be avoided.



(a) Rectified stereo image



(b) Vertical disparities  $d_y$  for automatically selected corresponding image points in the rectified image pair in Fig. 1a. We show  $d_y$  with respect to the  $x$ -coordinate (left) and color-coded with  $(x/y)$  position in the left stereo image.

**Fig. 1:** Visualization of vertical disparities  $d_y$  in pixels for the stereo camera in use. For rectified images computed from accurate stereo calibration parameters, the vertical disparity should be zero for all correct point correspondences. They have green color in the right visualization of Fig. 1b. In this example, significant errors with  $d_y > 0$  (red color) occur.

To quantify the calibration accuracy of stereo camera systems, we evaluate the vertical misalignment using corresponding image points for rectified stereo image pairs. To demonstrate, vertical disparities  $d_y$  of corresponding image points are visualized in Fig. 1 for a rectified stereo image pair captured several days after calibrating the cameras. We show  $d_y$  for corresponding points with respect to the  $x$ -coordinate (Fig. 1b, left) and color-coded with respect to the  $(x/y)$  position (Fig.

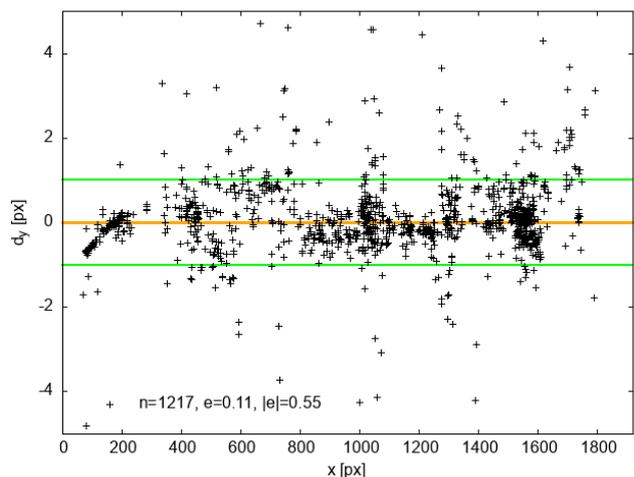
1b, right) in the left stereo image. The example shows a large systematic error with  $d_y > 0$  in the left part of the image. Thus, the calibration parameters no longer match the current camera configuration. The proposed measure for stereo calibration accuracy is used to optimize the calibration parameters. The optimization procedure minimizes the vertical misalignment in the rectified images. We demonstrate that the resulting images have significantly reduced stereo calibration error.

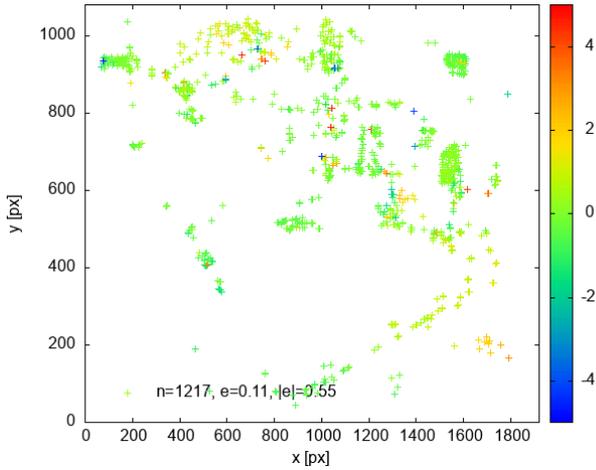
## 2 Accuracy of Stereo Calibration

Usually, stereo images and cameras are transformed in rectified stereo configuration using a preprocessing step [12]. The rectification computes new extrinsic camera parameters for which both cameras share the same rotation angles and image target plane. It follows, that corresponding image points in left and right image have the same  $y$ -coordinate which eases the analysis significantly. Then, the depth  $z$  is calculated from the horizontal disparity  $d_x$ , the baseline  $b$ , and the focallength  $f$  as  $z = f \cdot b/d_x$ . For a reasonable rectification output, accurately calibrated cameras are required. Otherwise, a vertical offset  $d_y$  is encountered when comparing corresponding points. Inversely, this offset can be used to quantify the accuracy of the original calibration.



(a) Rectified stereo image





(b) Vertical disparities  $d_y$  for corresponding image points in the rectified image pair in Fig. 1a with  $d_y$  with respect to the x-coordinate and the  $(x/y)$  position in the left stereo image.

**Fig. 2:** Visualization of the vertical disparities  $d_y$  with optimized parameters. Most correspondences have a  $d_y$  near zero and have green color in the right visualization of Fig. 2b. The mean vertical disparity is  $\epsilon = 0.08$  px. The significant error shown in Fig. 1b is eliminated.

For the stereo calibration accuracy measure, we make use of scale invariant keypoint detection and descriptor computation. For keypoint detection, classical approaches still provide higher subpixel accuracy compared to machine learning approaches. As shown in [4], A-KAZE [1] keypoints provide dominant subpixel localization accuracy. To ease the analysis, stereo images and cameras are rectified. We assume a small, but nonzero vertical offset  $d_y$  and limit the correspondence analysis to a small search space. Thus, the probability of outliers, i.e., wrongly established correspondences should be small. The disparity  $\Delta \mathbf{d}$  for each corresponding feature point pair  $\mathbf{p}_l, \mathbf{p}_r$  is defined as  $\Delta \mathbf{d} = \mathbf{p}_r - \mathbf{p}_l = (d_x, d_y)^t$ . The mean of vertical disparities for all corresponding  $n$  features points is [3]:

$$\epsilon = \frac{1}{n} \sum_{i=1}^n d_y^{(i)} \quad (1)$$

The mean of the absolute vertical disparities for all corresponding  $n$  features points is [3]:

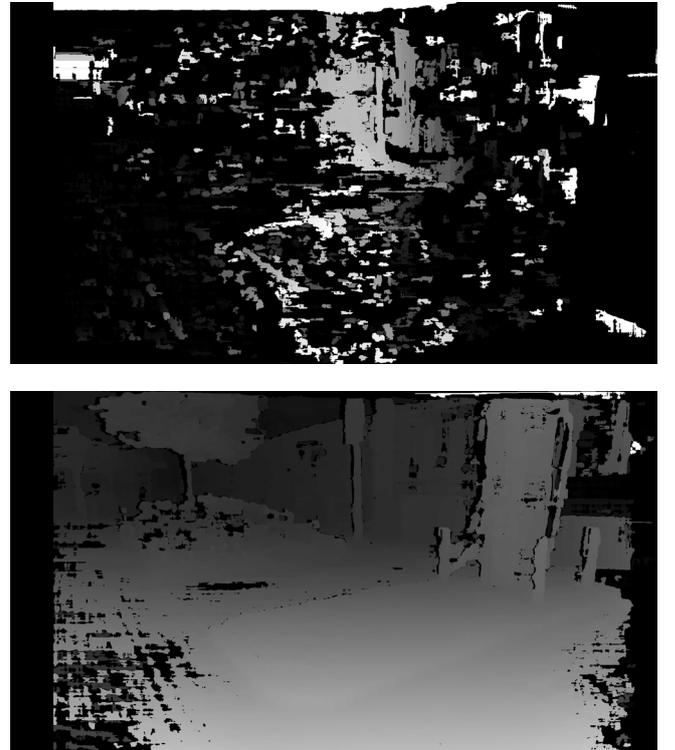
$$|\epsilon| = \frac{1}{n} \sum_{i=1}^n |d_y^{(i)}| \quad (2)$$

The mean of the vertical disparities  $\epsilon$  in equation (1) shows the systematic error resulting from inaccurate calibration parameters and is independent from the feature localization error. The absolute vertical disparities

$|\epsilon|$  in equation (2) provides the magnitude of the error. For the example in Fig. 1,  $n = 1233$  corresponding keypoints are established. We obtain  $\epsilon = 1.79$  px and  $|\epsilon| = 1.86$  px.

### 3 OPTIMIZATION OF STEREO CALIBRATION

For the optimization of camera parameters, the mean vertical disparity (equation (2)) is minimized [3]. For each iteration, images and cameras are rectified and correspondences are established to obtain the cost function value  $\epsilon$  as described in Sect. 2. We optimize 6 parameters (relative position and angles of the right camera).



**Fig. 3:** Depth maps computed by SGM (Semi-Global-Matching) for the rectified stereo images from Fig. 1 (above) and the optimized calibration as shown in Fig. 2 (below).

The resulting relative position  $C_r$  and angles for the stereo camera in the example experiment are as follows:

$$\begin{aligned} C_r &= (7.91 \text{ mm}, -0.24 \text{ mm}, -3.67 \text{ mm}) \\ (pan, tilt, roll)_r &= (0.09796^\circ, -0.09354^\circ, -0.07998^\circ) \end{aligned} \quad (3)$$

For the new camera parameters,  $n = 1217$  keypoints are obtained resulting in  $\epsilon = 0.11$  px and  $|\epsilon| = 0.55$  px. The corresponding rectified images are shown in Fig. 2a. For validation, the vertical disparities are shown in Fig. 2b. Compared to the original versions (Fig. 1), a significant decrease of the vertical disparities is visible. The systematic error diminishes. In Fig. 3 we

compare depth maps computed by SGM (Semi-Global-Matching) [8] as implemented in OpenCV. Similar to the experimental results in [10], a significant improvement is achieved for the camera parameters optimized with our approach (below) compared to the original stereo calibration (above). Currently, it is unclear, how much these calibration errors affect machine learning models for stereo depth estimation. Initial tests with RAFT-Stereo [11] indicate that these approaches appear less sensitive to small calibration errors. But, an accuracy analysis requires much more effort since the performance is dependent on both train and test datasets and their calibrations. In [3], it is shown that stereo vision datasets have various calibration error structures.

For classical depth estimation methods, cf. Fig. 3, our experiments show the practicability of the parameter optimization. Since no calibration patterns are needed, this procedure can be applied to adjust the camera parameters during acquisition as *Onlinecalibration*.

#### 4 CONCLUSIONS

Stereo cameras in use are subject to thermal and mechanical stresses that lead to fluctuations in the relative alignment of the two cameras. With the developed method, a consistently high accuracy of depth estimation can be ensured during use. Therefore, a keypoints correspondence analysis with high localization accuracy is employed. From the keypoints, vertical disparities are computed. The accuracy measure provides the possibility for error control and indicates the need for a recalibration.

The method is applied to self-recorded stereo images of a test vehicle. The calibration errors occurred under heavy mechanical stress caused by vibrations, e.g. from driving on cobblestones, are corrected. The proposed methodology enables Online Calibration since calibration patterns are not needed.

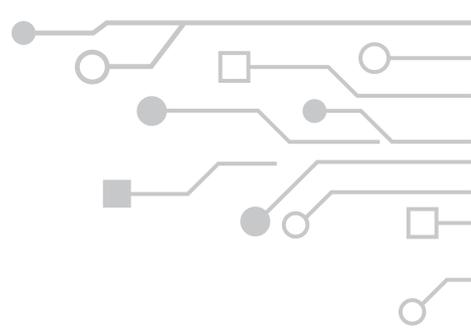
*This work was partially supported by the German Federal Ministry of the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (GreenAutoML4FAS project no. 67KI32007A).*

#### REFERENCES

- [1] Alcantarilla, P.F., Solutions, T.: Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell* 34(7), 1281–1298 (2011) 3
- [2] Bansal, M., Jain, A., Camus, T., Das, A.: Towards a practical stereo vision sensor. In: *Conference on Computer Vision and Pattern Recognition (CVPR) -Workshops*. pp. 63–63. IEEE (2005) 1
- [3] Cordes, K., Broszio, H.: Accuracy evaluation and improvement of the calibration of stereo vision datasets. In: *European Conference on Computer Vision (ECCV) - VCAD workshop (2024)* 1, 3, 4
- [4] Cordes, K., Grundmann, L., Ostermann, J.: Feature evaluation with high-resolution images. In: Azzopardi, G., Petkov, N. (eds.) *Computer Analysis of Images and Patterns*. pp. 374–386. Springer International Publishing (2015) 3
- [5] Dang, T., Hoffmann, C., Stiller, C.: Continuous stereo self-calibration by camera parameter tracking. *IEEE Transactions on image processing* 18(7), 1536–1550 (2009) 1
- [6] Furgale, P., Rehder, J., Siegwart, R.: Unified temporal and spatial calibration for multi-sensor systems. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. pp. 1280–1286. IEEE (2013) 1
- [7] Geiger, A., Moosmann, F., Car, O., Schuster, B.: A toolbox for automatic calibration of range and camera sensors using a single shot. In: *International Conference on Robotics and Automation (ICRA)* (2012) 1
- [8] Hirschmüller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. vol. 2, pp. 807–814. IEEE (2005) 2, 4
- [9] Kumar, A., Mannan, F., Jafari, O.H., Li, S., Heide, F.: Flow-guided online stereo rectification for wide baseline stereo. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 15375–15385 (2024) 1
- [10] Ling, Y., Shen, S.: High-precision online markerless stereo extrinsic calibration. In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. pp. 1771–1778. IEEE (2016) 4
- [11] Lipson, L., Teed, Z., Deng, J.: Raft-stereo: Multilevel recurrent field transforms for stereo matching. In: *International Conference on 3D Vision (3DV)*. pp. 218–227. IEEE (2021) 4
- [12] Loop, C., Zhang, Z.: Computing rectifying homographies for stereo vision. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. vol. 1, pp. 125–131. IEEE (1999) 2
- [13] Scharstein, D., Hirschmüller, H., Kitajima, Y., Krathwohl, G., Nešić, N., Wang, X., Westling, P.: High-resolution stereo datasets with subpixel-accurate ground truth. In: *German Conference on Pattern Recognition (GCPR)*. pp. 31–42. Springer (2014) 1
- [14] Warren, M., McKinnon, D., Upcroft, B.: Online calibration of stereo rigs for long-term autonomy. In: *IEEE International Conference on Robotics and Automation*. pp. 3692–3698 (2013) 1
- [15] Xu, D., Zeng, Q., Zhao, H., Guo, C., Kidono, K., Kojima, Y.: Online stereovision calibration using on-road markings. In: *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. pp. 245–252. IEEE (2014) 1







## AI4EA Workshop (Berlin Workshop on Artificial Intelligence for Engineering Applications)



Mitveranstalter:

AI4Tech (ZIM-Innovationsnetzwerk)

NET4AI (GRW-Netzwerk)

Berlin  
27.11./ 28.11.2024



# Testing ML-based Systems Using Probabilistically Extended Ontologies

Hans-Werner Wiesbrock and Nicolas Grube

ITPower Solutions GmbH, Kolonnenstraße 26, 10829 Berlin, Germany  
 {hans-werner.wiesbrock, nicolas.grube}@itpower.de

**Abstract.** Systems based on Machine Learning (ML-based systems) are currently employed in numerous applications, including the recognition of access authorizations, vehicles, and individuals in traffic. The fundamental technology underlying these examples is deep neural feedforward nets. A substantial proportion of these applications are safety-critical, and the accuracy and precision of their classifications are of paramount importance. The question thus arises as to how the new systems can be demonstrated to fulfill these requirements.

This paper presents a novel systematic approach developed in the research project KI-LOK. The approach is based on the concept of probabilistically extended ontologies (PEON). It considers the statistical nature of these ML-based systems and permits the derivation of test end criteria. Moreover, this approach is consistent, whereby if developers and independent testers have worked in accordance with the prescribed methodology, they will receive comparable quality values.

This paper employs an illustrative example to demonstrate that the statistical nature of ML-based systems requires the development of novel testing concepts. One potential solution to this issue is the probabilistic extension of ontologies, which is subsequently described and illustrated by a case study from the KI-LOK project.

**Keywords.** Test and Statistic for ML Based Systems · Verification of accuracy / precision of neural nets · Probabilistically Extended Ontology (PEON) · Sampling Abstract Test Data

## 1 INTRODUCTION

The application of machine learning (ML) systems has proliferated across a diverse range of domains, including those that are safety-critical. Furthermore, an increasing number of experts are issuing warnings about the potential uncertainties and risks associated with the uncontrolled and accelerated development of artificial intelligent (AI) systems [2]. This raises the question of how such systems can be qualitatively assessed, even independently of development.

In the research project KI-LOK we developed a black box test for object recognition systems based on the systematic software test. This test addresses some of the outstanding questions such as:

- How and to what extent should test data be selected for the quality assurance of an AI system?
- How should the executed tests be evaluated?
- What statement about the quality of the test object can be derived from the executed tests?
- When should one stop testing (Completion Criteria)?

Furthermore, this approach can be extended to a multitude of other, more general systems, see [15,8,16].

This article begins with an overview of the rationale behind the use of statistics in this context. It then introduces the concept of probabilistically extended ontologies, as outlined in the referenced literature, [16]. This method extends the category partition or classification tree method, replacing the combinatorial criteria with a statistical approach. This methodology allows for the systematic resolution of the aforementioned questions, thereby providing a development-independent test methodology. The aforementioned activities will be integrated into an autonomous test process. By means of

illustrative examples drawn from the railroad industry, we demonstrate how this procedure can be implemented in practice. In addition, elements of the methodology can be used to delineate and evaluate balanced, comprehensive training data. This latter point will be discussed in future work.

## 2 ML-BASED SYSTEMS AND STATISTICS

In contrast to conventional software, machine learning (ML) software is not developed on the basis of rules; rather, it learns inductively from a large number of examples. This has a number of consequences. For example, fixing bugs in these systems is not a straightforward process, nor is it possible in most cases! The question itself is also worth considering: What is an error in an object recognition system? A closer look reveals the problems.

**Example 1** *It is evident that a classification program designed to differentiate between dogs and cats based on visual cues would yield erroneous results if it identified a Siamese cat as a German shepherd. Nevertheless, if the system provides the correct response with greater frequency than a human, it will be employed regardless. The quality of such a system is not determined by the individual case, but rather by statistical analysis. In order to ascertain the classification of the input image as either cat or dog, Bayesian conditional expectation is employed.*

Accuracy and precision are calculated as the averaged values over a large number of test results. In addition, the learning process uses statistical quantities such as maximum entropy resp. likelihood to facilitate optimization. ML-based systems function statistically, thus necessitating statistical testing. A straightforward Gedankenexperiment can be employed to illustrate this concept.

**Example 2** *Look at the simple classification system outlined in Example 1, which can distinguish between cats and dogs in animal images. Suppose we received ground truth data stating that dogs are correctly identified 95 % of the time, while cats are correctly identified 85 % of the time. What, then, is the accuracy of the system, expressed as the mean value of correctly recognized images?*

*A tester tests the system using 900 dog and 100 cat images and arrives at the conclusion that the accuracy of the system is  $(900 \cdot 0.95 + 100 \cdot 0.85) / 1000 = 94\%$ .*

*A second tester employs 900 images of cats and 100 images of dogs, yielding a result of  $(100 \cdot 0.95 + 900 \cdot 0.85) / 1000 = 86\%$  see Table 1*

Tester	Dog	Cat	accuracy
Tester 1	900	100	94%
Tester 2	100	900	86%

**Tab. 1:** Gedankenexperiment

*What is the correct response?*

In [8], we utilize a straightforward ML-based model to illustrate the extent to which the test outcomes for the system's accuracy are contingent upon the distribution of the test data. In particular, these results indicate that the conventional combinatorial testing approach is no longer a viable method for evaluating the accuracy of ML-based systems.

From a mathematical perspective, the precision and accuracy of ML-based systems and other conventional statistical quality indicators are ultimately determined by the outcomes of measurements and test runs. The validity and significance of such an approach is contingent upon the specific set of samples under consideration. Furthermore, the fundamental tenets of the central limit theorem, which pervades nearly all such methodologies, dictate that samples must be drawn from the base distribution. When deterministic algorithms are used, which do not adhere to the base distribution, it is impossible to verify the validity of any statistical quality criteria e.g., the precision and accuracy of a ML-based system. Without knowledge of the distribution, a mean cannot be determined without sampling.

In conclusion, these observations imply that in order to measure reliable quality indicators for ML-based systems, it is necessary to have direct access to the base distribution. It is only through the utilization of a probability model for the input data that a systematic and reasonable approach to sampling test data can be achieved. The concept of a *probabilistically extended ontology* enables us to address this issue.

## 3 PEON, THE CONCEPT OF PROBABILISTICALLY EXTENDED ONTOLOGIES

Systematic software testing is a methodical, tool-supported, and well-proven procedure in which the input space of the test object is adequately covered by tests. This is achieved through the use of established techniques, such as the category partition method, limit value analysis, and the classification tree method, see [7]. This method is a black box approach that permits quality assessment to be conducted independently of the development process.

In the case of ML-based systems, their target domain, the operational design domain (ODD), may include for example, images of people, road traffic, train routes, video sequences, and audio recordings. . . The environment is too complex to be represented by a tree-like structure. Ontologies are becoming an increasingly popular means of formalizing the description of highly complex environments within the context of an ML-based system, see [1,6,11,9,5,3,4,12,17,13]. They comprise the entities with their properties and logical dependencies as they can occur in the target environments. Nevertheless, to date, ontologies have been employed primarily for the purpose of identifying edge cases. We will use ontologies more pervasive in a wider range of applications.

Let us consider a given ontology with a set of defined entities, for instance person. These entities may exhibit a variety of characteristics, including size, thickness, age, etc. In some images, a person may be depicted. The person may be of varying sizes, with varying degrees of thickness or thinness, and may belong to any age group. Images can be grouped accordingly. This results in a partitioning of the ODD as specified by the ontology. An example of which, is pictured on the left in figure 1. Such structures serve as a replacement for category or classification trees.

Nevertheless, the data set is insufficient for a statistical evaluation, which is a prerequisite for the assessment of ML-based systems. As demonstrated by the experiments presented in Chapter 2, or for further details [8], it is crucial to define the probability of occurrence for the various partitions. In the case of the simple dog/cat example 2, it is necessary to specify the distribution of images to be classified between these two categories. In Germany, there are approximately 12 million dogs and 16 million cats. Consequently, the correct answer for Germany would be approximately 92.6 %.

As previously stated, an ontology defines a partition of the ODD. This must now be augmented with occurrence probabilities for each partition, as illustrated in Figure 1 .

**Definition 1** A probabilistically extended ontology is defined as an ontology augmented with a probability distribution on the related partition reflecting the occurrence of the partition.

To rephrase, a probabilistically extended ontology provides us with a probabilistic model of the ODD. Depending on the partition the failure rate of the test object will be different. A large person in yellow clothing will be easier to see than a small person in camouflage.

**uniformity hypothesis for a PEON:** In light of this concept, a criterion is proposed for the entities and refined properties that should be modelled in the ontology. It is modelled on the uniformity hypothesis of the classification tree analysis, [7] and depends on the test object. The uniformity hypothesis for a PEON states that the refinement of a (ontology) partition should continue until the failure rate of the ML-based system on exemplars of the partition is likely to be equal. This hypothesis allows us to derive a statistical model of test performance from the PEON, thereby facilitating a more precise evaluation of the test, see [15,16].

**sampling, sample size and test completion criteria:** If we want to achieve an accuracy of specified significance by testing, the probability model helps us to sample an appropriate set of abstract test cases. We can also estimate the size of the sample set needed for this by using the central limit theorem.

**test evaluation:** When the sampled test cases are executed on the test object, we can compute the various statistics for their results against the probability model defined by the *probabilistically extended ontology*.

In this way we can answer the questions 1 in chapter 1. At first sight, modelling such a distribution of probabilities of occurrence seems impossible. However, we have developed a number of techniques, language concepts and algorithms to achieve this goal. In the KI-LOK project,

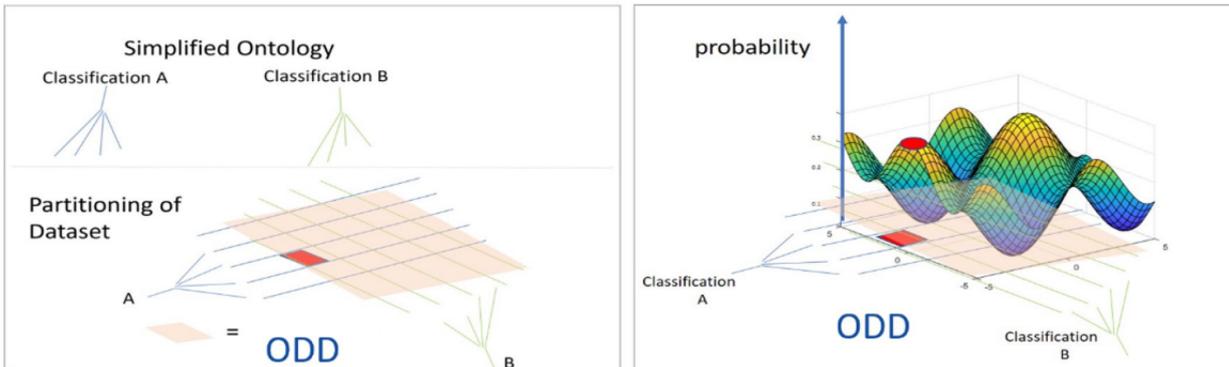


Fig. 1: The partitioning is conducted in accordance with an established ontology, with the occurrence probabilities represented accordingly.



of sections, which can be delineated by their length, curvature and orientation, see Figure 3. Again it is reasonable to assume that we get statistical data for the distribution of curvatures and lengths from some railway institute. Given the symmetry inherent in riding back and forth, we set the orientation to be uniformly distributed. This defines the marginal distributions for track parts.

There are various categories of railway lines, including urban, regional and highspeed lines. These are typically situated in different environments and their distribution is presumed to be well-known.

High-speed lines usually have very long sections with only very slight curves, whereas urban lines have relatively short sections with sharper curves. We model this dependency as follows. The different types have a natural order urban, regional,... highspeed, the curvatures also highly curved,... almost straight. We index the different possibilities and can then describe the dependency informally: The higher the index of the type, the more likely we expect a higher index of curvature.

High-speed lines typically comprise lengthy sections with minimal curvature, whereas urban lines are characterised by comparatively shorter sections with sharper curves. The dependency in question is modelled as follows: the various categories exhibit a natural order, progressing from urban to regional and then to high-speed. The curvatures also vary considerably, from highly curved to almost straight. The different possibilities are indexed, allowing for an informal description of the dependency: the higher the index of the category, the more probable it is that the index of curvature will be higher. We model this dependency linearly, see Figure 3.

In this way we obtain a PEON description of the various types of railway tracks and subsequently sample abstract track descriptions. In the second step, the aforementioned descriptions can be readily sampled for concrete data regarding length, curvature, and alignment, thereby enabling the sampling of specific railway tracks.

### 5.2 SIGNALS ALONG THE TRACK

A case study of the KI-LOK project dealt with the recognition of signals on the railroad line. There are different types such as dwarf signal, Andreas cross etc., closely related to them different statuses

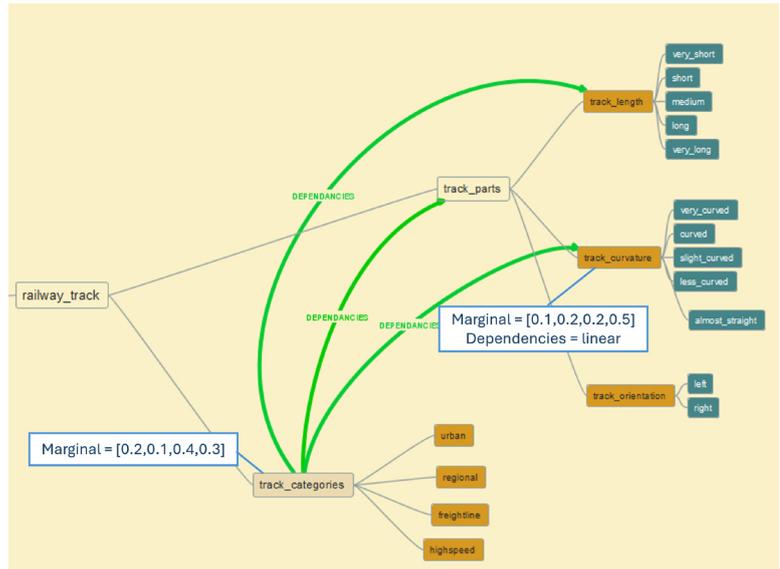


Fig. 3: Section of a probabilistically extended ontology for railway tracks

of the signals and their localization to the tracks or to the ego train, see Figure 4. COUNT = [1,4] at the node signals describes that between one and four signals are to be sampled. Depending on the route category, there may be more or fewer signals, but this is not modeled in this example. The localisation needs to be specified for many objects along the route, not just for signals. Instead of noting this required property separately under each signal, we can bundle these properties into separate subtrees, as is common in ontologies, see Figure 4. In general, the position of the signals along the track is constantly changing due to the movement of the train, so their longitudinal spacing can be assumed to be uniformly distributed.

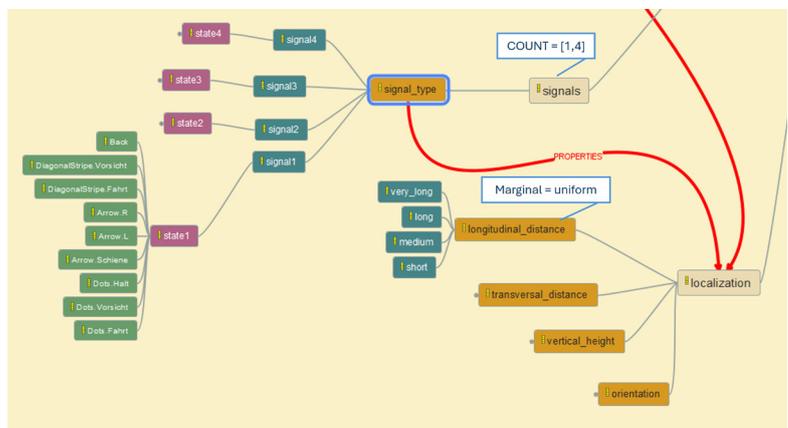
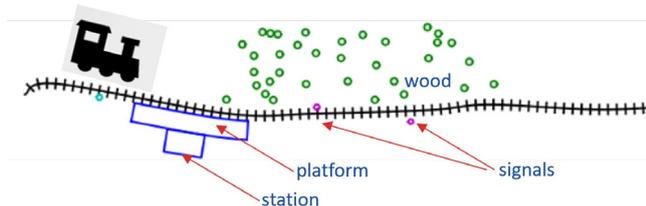


Fig. 4: PEON for trackside signals

Again we can sample from this abstract settings for signals and in a second step concrete data for appearance, state and localisation of signals, see Fig. 5.



**Fig. 5:** PEON generated conceptual picture of a railway track

### 5.3 TRANSFER TO OTHER DOMAINS

As we have seen in the previous sections, a PEON is built according to the ODDs context and extends the concept of ontologies. There are various methodologies for designing a PEON. Our preferred approach to general scenarios is a top-down methodology, whereby the design process begins with a comprehensive overview of the system, encompassing the environment, track, and structures, and then proceeds to the finer details, such as the signals, their location and types.

It is similarly important to consider the reliability of the system in special conditions, for example, in inclement weather. In such circumstances, the system should continue to perform satisfactorily, even in conditions of heavy fog. In order to address this issue, the development of appropriate PEONs can be undertaken to describe these conditions, which are also relevant in other areas, see figure 2.

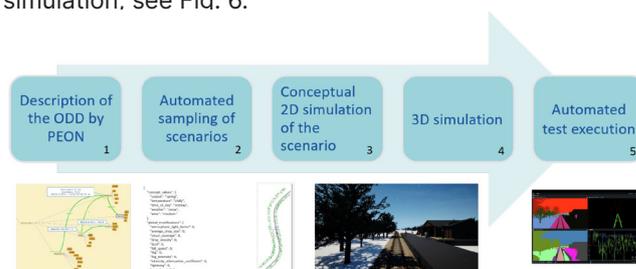
The concepts of ontologies and PEONs entail that they can be iteratively refined and supplemented. Therefore this general approach can be directly transferred to other domains. When the automotive domain is targeted, the environment and structures would be similar, while roads take the part of railroad tracks and the signals in part map to different assets (stop signals to stop signs) or are of distinctly different types (switch signals have no equivalent in the automotive domain, where as no stopping signs have no equivalent in the train domain). But one should take care. While some entities and their properties in a PEON may be identical across the train and automotive domains, they may also be subject to domain-specific prior distributions for re-evaluation. Failure to do so could result in statistical representativeness being skewed, thereby compromising the desired results.

The ability to transfer parts of a PEON between domains has implications for the scalability of this approach.

The construction of a new ontology for each domain is unnecessary, as prior distributions can be adapted in parts of the ontology, thereby facilitating transferability and scalability.

### 6 PEON, A BASIC BUILDING BLOCK OF A PROOF-OF-CONCEPT (POC) IMPLEMENTATION

Following the analysis of [10] and their proposed safety evaluation process, which is another result of the KI-LOK project, we want to propose an implementation of the process in a toolchain (figure 6) to accelerate the acceptance of the approach in the industry. We are also focusing on a black box test to enable adoption without the need for a specialised test setup. Together with our project partners, we developed a proof-of-concept implementation that can be used to validate an object detection system for railways. The basis and statistical reference point is a probabilistically extended ontology developed for several use cases. From this we generate abstract test cases, i.e. specified partitions from which the input data should be selected. The abstract data are then sampled to concrete descriptions, i.e. what detailed path the track should follow, where signals should be placed, etc., as sketched in Chap. 5.1,5.2. We get what is called a conceptual picture of the test, a 2-dimensional description of the scene. The result is then further enhanced by the addition of concrete assets and is expanded into a complete 3-dimensional visual simulation, see Fig. 6.



**Fig. 6:** Concept of the POC toolchain and demonstrator

The tool chain allows for the automatic determination of the quality of an object recognition system. The requisite number of test cases is calculated based on the quality and significance required for a scene, using a statistical model derived from the PEON. A corresponding number of abstract test cases are then sampled and enriched to create complete visual simulations. This data is fed into the test object, the object recognition system, and the system's ability to recognise the various objects is measured. The statistical evaluation of these results then provides transparent, objective and comprehensible criteria for the acceptance or rejection of the test object.

## 7 CONCLUSION AND FUTURE WORK

It is imperative that ML-based systems be evaluated statistically, given that they are based on statistical models. This in turn requires the development of new testing concepts. The novel concept of probabilistically extended ontologies (PEON) facilitates a statistical characterization of the target environment, the ODD, thereby serving as a reference for statistical test outcomes. In preliminary trials, we demonstrated that this methodology is reliable, whereby if developers and testers have followed the prescribed procedures, they will produce comparable assessments of the system's quality, see [8]. In this paper we have shown that, in general, combinatorial testing is not able to reproduce the quality of the systems claimed by the developers. To date, preliminary PEONs have been developed for railroad lines with forests, platforms, various track signals, and weather conditions (rain, fog, etc.). However, further modeling, such as that of road traffic and people, remains to be done. With these additions, the potential for further development and expansion of suitable modeling concepts is significant.

**Acknowledgments.** This work has been funded by <https://ki-lok.itpower.de/> (German Federal Ministry for Economic Affairs and Climate Action and managed by TÜV Rheinland (project no.: 19121007A)) and <https://iml4e.org> (German Federal Ministry for Education and Research (project no.: 01IS21021C))

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## REFERENCES

- [1] ASAM e.V.: Asam openscenario: Version 2.0.0 concepts <https://www.asam.net/standards/detail/openscenario/>
- [2] Bengio, Y., Hinton, G., et. al.: Pause giant ai experiments: An open letter (22032023), <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>
- [3] Bogdoll, D., Guneshka, S., Zöllner, J.M.: One ontology to rule them all: Corner case scenarios for autonomous driving, <http://arxiv.org/pdf/2209.00342v2>
- [4] Breitenstein, J., Termöhlen, J.A., Lipinski, D., Fingscheidt, T.: Corner cases for visual perception in automated driving: Some guidance on detection approaches, <http://arxiv.org/pdf/2102.05897v1>
- [5] Breitenstein, J., Termöhlen, J.A., Lipinski, D., Fingscheidt, T.: Systematization of corner cases for visual perception in automated driving. In: 2020 IEEE Intelligent Vehicles Symposium (IV). pp. 1257–1264. IEEE (2020). <https://doi.org/10.1109/IV47402.2020.93047896>.
- [6] Fremont, D.J., Kim, E., Dreossi, T., Ghosh, S., Yue, X., Sangiovanni-Vincentelli, A.L., Seshia, S.A.: Scenic: A language for scenario specification and data generation, <http://arxiv.org/pdf/2010.06580.pdf>
- [7] Grimm, K., Grochtmann, M.: Systematischer software-test mit der klassifikationsbaummethode. In: Reichel, H. (ed.) Informatik - Wirtschaft - Gesellschaft, pp. 253–259. Springer eBook Collection Computer Science and Engineering, Springer Berlin Heidelberg, Berlin, Heidelberg (1993). [https://doi.org/10.1007/978-3-642-78486-6\\_40](https://doi.org/10.1007/978-3-642-78486-6_40), [https://link.springer.com/chapter/10.1007/978-3-642-78486-6\\_40](https://link.springer.com/chapter/10.1007/978-3-642-78486-6_40)
- [8] Grube, N., Massah, M., Tebbe, M., Wancura, P., Wiesbrock, H.W., Grossmann, J., Kharm, S.: On a systematic test of ml-based systems: Experiments on test statistics. In: 2024 IEEE International Conference on Artificial Intelligence Testing (AITest). pp. 11–20. IEEE (7/15/2024 - 7/18/2024). <https://doi.org/10.1109/AITest62860.2024.00010>
- [9] Lihua Zhao, Ryutaro Ichise, Seiichi Mita, Yutaka Sasaki: Core ontologies for safe autonomous driving (2015), [https://www.researchgate.net/publication/286313863\\_Core\\_Ontologies\\_for\\_Safe\\_Autonomous\\_Driving](https://www.researchgate.net/publication/286313863_Core_Ontologies_for_Safe_Autonomous_Driving)
- [10] Roßbach, J., De Candido, O., Hammam, A., Leuschel, M.: Evaluating AI-Based Components in Autonomous Railway Systems: A Methodology. In: Hotho, A., Rudolph, S. (eds.) KI 2024: Advances in Artificial Intelligence, vol. 14992, pp. 190–203. Springer Nature Switzerland, Cham (2024). [https://doi.org/10.1007/978-3-031-70893-0\\_14](https://doi.org/10.1007/978-3-031-70893-0_14), [https://link.springer.com/10.1007/978-3-031-70893-0\\_14](https://link.springer.com/10.1007/978-3-031-70893-0_14), series Title: Lecture Notes in Computer Science
- [11] Schuldt, F.: Ein Beitrag für den methodischen Test von automatisierten Fahrfunktionen mit Hilfe von virtuellen Umgebungen (2017). <https://doi.org/10.24355/dbbs.084-2017042421210>, [https://leopard.tu-braunschweig.de/receive/dbbs\\_mods\\_00064747](https://leopard.tu-braunschweig.de/receive/dbbs_mods_00064747)
- [12] Stefani Guneshka: Ontology-based corner case scenario simulation for autonomous driving. Bachelor thesis, FZI Research Center for Information Technology Karlsruhe
- [13] Ul-Haq, S., Qamar, U.: Ontology based test case generation for black box testing. In: Proceedings of the 2019 8th International Conference on Educational and Information Technology. pp.236–241. ACM, New York, NY, USA (2019). <https://doi.org/10.1145/3318396.3318442>
- [14] Villani, C.: Optimal transport: Old and new, Grundlehren der mathematischen Wissenschaften, vol. 338. Springer, Berlin (2009), <http://swbplus.bsz-bw.de/bsz28657621xcov.htm>
- [15] Wiesbrock, H.W., Grossmann, J.: Probabilistically extended ontologies: A basis for systematic testing of ml-based systems. <https://doi.org/10.4271/2024-01-3002>, <https://www.sae.org/publications/technical-papers/content/2024-01-3002/>
- [16] Wiesbrock, H.W., Grossmann, J.: Outline of an independent systematic blackbox test for mlbased systems (7/15/2024 - 7/18/2024). <https://doi.org/10.1109/AITest62860.2024.00009>
- [17] Zipfl, M., Koch, N., Zöllner, J.M.: A comprehensive review on ontologies for scenariobased testing in the context of autonomous driving, <https://arxiv.org/pdf/2304.10837.pdf>

# Finding Predictive Features for Energy Consumption of CNC Machines

Hafez Kader<sup>1</sup>, Robin Ströbel<sup>2</sup>, Alexander Puchta<sup>2</sup>, Jürgen Fleischer<sup>2</sup>, Benjamin Noack<sup>1\*</sup> and Myra Spiliopoulou<sup>3\*</sup>

\*These authors contributed equally to this work.

<sup>1</sup>Autonomous Multisensor Systems Group, Institute for Intelligent Cooperating Systems, Otto von Guericke University Magdeburg, Germany.

<sup>2</sup>Institute of Production Science, Karlsruhe Institute of Technology (KIT), Germany.

<sup>3</sup>Knowledge Management and Discovery Lab, Faculty of Computer Science, Otto von Guericke University Magdeburg, Germany.\*

**Abstract.** With rising energy costs and a growing emphasis on sustainable and efficient production, predicting the energy consumption of CNC machines has become increasingly important. Accurate predictions can lead to significant energy savings, better planning, more informed decision-making, and alignment with smart manufacturing and Industry 4.0 initiatives. Extensive research has been conducted in this area, utilizing both physical and analytical models, as well as expert knowledge from experiments. More recently, machine learning models have also been employed using a wide range of input features. In this paper, we examine the energy consumption of CNC machines by analyzing various features explored in different studies. We propose a method that ranks these features based on their predictive power, then groups the rankings to highlight a small subset of the most predictive features. Furthermore, we assess the stability of the predictive power of these features over time, allowing us to not only rank them by their predictive strength but also evaluate their long-term stability. Our findings indicate that only a few features are highly predictive, and their predictive power remains consistent over time.

**Keywords.** Time series prediction, CNC machine, feature relevance.

## 1 INTRODUCTION

Computer numerical control (CNC) machining is crucial for part manufacturing, and is extensively used across industries to produce components and finished products [1]. CNC-controlled tools, such as three-axis milling machines with the spindle (SP), are common but energy-intensive [2]. Rising energy costs and climate concerns have led to increased efforts to optimize energy consumption in manufacturing [3], particularly in machine tools [4]. Nearly one-third of the world's energy consumption and 36% of CO<sub>2</sub> emissions are attributed to the manufacturing sector [5, 6], highlighting the need for energy-efficient production.

Manufacturers are under pressure to reduce energy use and CO<sub>2</sub> emissions due to environmental regulations and consumer demand for sustainable practices [7]. Efforts to minimize energy consumption in CNC machining not only reduce costs but also improve environmental impact, making energy-efficient CNC tools essential for meeting environmental goals. However, achieving energy savings remains a challenge [8].

Predicting energy consumption in CNC machining is critical for improving efficiency [9]. Companies are interested in simplifying the prediction process [10], and machine learning models are increasingly used to forecast energy use. However, different features are used as inputs in the models across various studies. Understanding the significance of these features is crucial for enhancing model transparency, explainability, and improving prediction accuracy.

In our paper, we present a method for ranking features based on their predictive power, grouping this ranking into categories according to their importance. The approach allows for the prioritization of features, helping to decide

which should be emphasized or deprioritized in the future. Additionally, we quantify the stability of these features, examining whether they maintain their categories over time or shift between categories. Therefore, the research question addressed in our paper is:

How can we rank features according to their predictive power in importance categories, and how can we determine the stability of these categories over time?

The structure of the paper is organized as follows: Section 2 presents a comprehensive review of prior research on machine energy consumption. Section 3 outlines the methodology, detailing how features are ranked into importance categories and how their stability is determined. In Section 4, the dataset used in this study is introduced, followed by a description of the experimental setup, including the implementation and evaluation metrics. The results are presented in Section 5, along with a discussion of the findings. Finally, Section 6 provides the conclusion and offers recommendations for future research.

## 2 RELATED WORK

The industrial sector is increasingly focusing on the development of sustainable products with minimal environmental impact [11]. A key aspect of this is reducing the energy consumption of machine tools, which is critical for improving the sustainability of manufacturing processes [5]. The energy consumption of these machines can be categorized into two main components: primary units, responsible for kinematic movements like axis control and the main spindle operation, and secondary units, which include auxiliary functions such as coolant supply and hydraulic systems. Several studies have aimed to optimize machining parameters to reduce energy consumption [12]. However, modeling energy usage remains challenging due to uncontrollable factors such as random errors and missing data in observed samples [9]. The complexity of these variables makes precise energy prediction difficult. This is particularly true when the focus is on understanding the process itself rather than on the auxiliary components of the machine, as modeling the process is inherently more complex and challenging. In recent years, machine learning methods have been increasingly employed to predict energy consumption. A major advantage of machine learning is its ability to process large datasets and uncover deeper insights into machining processes. These techniques allow for the extraction of valuable patterns and correlations that were previously difficult to detect. One of the pioneering efforts in this field was by Dietmair et al., who proposed a model for predicting and optimizing

energy efficiency in machining operations [13]. Similarly, Cao et al. et al. developed a model that categorized CNC code, enabling predictions of energy consumption for each grouped state [14]. Quintana et al. used a neural network model with seven input parameters, including feed rate, to enhance prediction accuracy [15]. Kante et al. further refined energy prediction using a neural network model with variable input parameters, improving model performance [16]. Borgia et al. applied machine learning techniques using features such as axis speed, acceleration, spindle speed, and torque to predict energy consumption [17]. Duc and Trinh developed an approach that incorporated tool wear as a factor in energy consumption prediction [18]. Ströbel et al. created a machine learning model that used CNC-derived features like speed and acceleration for each axis to improve energy prediction accuracy [19]. Camposeco-Negrete et al. optimized energy consumption using an ANOVA model [20], while Liu et al. studied how process parameters influenced both tool wear and the energy consumption of machine tools. [21]. Additionally, Draganescu et al. [22] aimed to correlate energy consumption at the tool tip with the efficiency of the main spindle drive under varying cutting conditions. Kolar et al. proposed a model for predicting CNC machining energy consumption, considering factors like material removal rate, spindle speed, coolant pumps, tool changers, and fan motors [23].

As highlighted in the literature, energy consumption predictions in manufacturing tools are often based on analytical or experimental approaches. These approaches vary significantly and are often derived from simulations, which can lack precision. A thorough evaluation of the importance of these features and their influence on prediction accuracy is often absent. In our earlier work [24] we employed a long short-term memory (LSTM) model to rank features based on their predictive power. However, an analysis of the categorisation of these features, as well as their stability in improving energy consumption predictions, is still lacking. In summary, while progress has been made in optimizing energy consumption in machining processes, a deeper understanding of how specific features contribute to the model and their relationship to energy consumption is still needed to enhance prediction accuracy.

## 3 METHODOLOGY

In our study, we do not consider factors that are more straightforward to model, such as auxiliary components like cooling conditions, due to the extensive research already conducted in those areas. Instead, we focus on the energy consumption of the spindle, which

constitutes the largest portion of the machine's total energy consumption, represents the process-dependent aspect of the machine, and poses greater challenges for prediction. This is the primary target of our work. Due to the strong correlation between the higher-resolution current signal and the power signal [19, 24], it is possible to establish predictable relationships between features and energy consumption. Furthermore, the accessibility of the current signal in brownfield machines supports its selection as the target  $y$  in our paper. Following the approach of our earlier work [24], this study develops a model  $M$  for each feature  $f$  to assess its influence on the prediction. An LSTM is also applied here to predict the spindle's current signal as accurately as possible. The LSTM is a powerful recurrent neural network that detects complex patterns in time series and captures long-term dependencies [25]. It outperforms traditional models like ARIMA [26], is robust against noisy data, and effectively handles issues such as exploding or vanishing gradients [27]. At each time step  $t$ , the performance of each Model  $M$  is quantified using an evaluation metric  $E$ . Based on this, features are ranked at each time point, with the best feature receiving a rank of 1, and the worst feature receiving a rank equal to the total number of features  $m$ . Individual feature rankings can change rapidly with slight shifts in prediction values, whereas features within ranking categories are less likely to fluctuate.

Unlike in our earlier work [24], this study groups these ranks into „bins“ (ranking categories  $\mathbf{R}$ ) to categorize the importance of the features. The advantage of grouping rankings into bins is that it allows features to be categorized based on their importance to the prediction. Additionally, it provides a more reliable way to quantify the stability of feature predictions. The number of bins used depends on the number of features  $m$  and a parameter  $k$ . The number of bins is calculated as follows:

$$\#bins = \frac{m}{k}, \quad (1)$$

where  $k \leq m$ . The larger the value of the parameter  $k$ , the more the  $m$  features are grouped into the same bin. The features are then ordered by importance. With  $k = m$ , each feature stands alone, with no grouping into bins at all. As  $k$  decreases, the number of bins increases. The ranks are transformed into bins in a linear fashion, that can be described as:

$$bin(r) = \lceil r * k / m \rceil. \quad (2)$$

This function ensures that each feature ranking is assigned a distinct bin. For instance, if there are 4 features and 2 bins, the first two features are assigned

to the first bin, and the last two features are assigned to the second bin. This allocation follows the calculation  $bin(1) = 1 * 2/4 = 0.5$ , which rounds up to bin 1, and so on.

This approach categorizes feature importance, with particular attention given to features in the top bins. Features considered less important are grouped into the last bin. Each  $\mathbf{f}$  is assigned a unique rank based on its performance at each time point  $t$ , ensuring no two features share the same rank. However, if two features have identical error values at a specific time point  $t$ , they are assigned the average of their respective ranks. It is recommended to use a natural number for the number of bins. This can be accomplished by determining the parameter  $k$ .

After transforming ranks into bins  $\mathbf{R} \in \mathbb{N}^{m,t}$  for each feature  $\mathbf{f}$  at each time point  $t$  with the ComputeBins function, for every  $\mathbf{f}$  the matrix  $\mathbf{B}$  is divided into partitioned windows  $n$  over the entire time  $t$  to obtain a unique bin assignment for each  $\mathbf{f}$  and later to analyze the stability of the predictive power of each  $\mathbf{f}$ . Within each window (with window size  $l \in \{1, \dots, t\}$ ), the most frequent bin for each  $\mathbf{f}$  is identified, and this bin is considered the „winner“ for that feature in the respective window. If two or more bins occur with equal frequency, the average of these bins is used as the bin winner. A  $\mathbf{f}$  may retain the same bin across different windows. If fewer data points are available in the final window than the window size, we still consider the most frequent bin. We denote the result as the matrix  $\mathbf{B} \in \mathbb{N}^{m,n}$ .

To transform all bins over time into one final bin, the most frequent bin across all windows is identified for each feature, assigning it to one of the predefined categories. If multiple bins are equally frequent, the average these bins is taken as the final bin. A pseudocode for the method is supplied in Algorithm 1.

---

**Algorithm 1** Feature binning
 

---

**Input:** Feature matrix  $\mathbf{F} \in \mathbb{R}^{m,t}$ , Target  $\hat{y}$ , Model  $M$ , ComputeBins function ComputeBins(), window size  $l$ , evaluation function  $E()$ , rank function  $rank()$

**Output:** Feature bin vector  $\mathbf{b}$

```

1: for feature  $\mathbf{f}$  in  $\{1, \dots, m\}$  do
2:    $M' \leftarrow train\_model(M, \mathbf{f})$ 
3:   for time  $\theta$  in  $\{1, \dots, t\}$  do
4:      $\hat{y} \leftarrow evaluate\_model(M', \mathbf{f}, \theta)$ 
5:      $\mathbf{E}_{\mathbf{f}, \theta} \leftarrow (y - \hat{y})^2$ 
6:   end for
7: end for
8: for time  $\theta$  in  $\{1, \dots, t\}$  do
9:    $\mathbf{R}_{:, \theta} = ComputeBins(rank(\mathbf{E}_{:, \theta}))$ 
10: end for
11: for feature  $\mathbf{f}$  in  $\{1, \dots, m\}$  do
12:   for window  $i$  in  $n$  do
13:      $\mathbf{B}_{\mathbf{f}, n} = most\_frequent(\mathbf{R}_{\mathbf{f}, 1+(i-1)*l:i*l})$ 
14:   end for
15:    $\mathbf{b}_{\mathbf{f}} = most\_frequent(\mathbf{B}_{\mathbf{f}, :})$ 
16: end for

```

---

Based on these bins and their frequency in the windows  $n$ , the stability of each feature's predictiveness can be determined. To quantify this, for each feature, starting with the first window, the bin in which the feature resides is compared to subsequent windows.

A feature's stability is measured by how often its bin changes across all windows. A feature is more stable if it changes categories less frequently, and it is considered most stable if it remains in the same bin for all windows. For example, a feature is stably predictive if it consistently stays in bin 1 across all windows, and stably non-predictive if it consistently remains in the last bin across all windows. Stability can be expressed as a percentage of windows in which a feature stays in the same bin. A feature is considered relatively stable if it stays in the same bin for a high percentage of windows (e.g., 80% or more).

In contrast, predictiveness refers to the ability of a feature to be useful for making predictions. A feature is considered stably predictive if it predominantly resides in bin 1 (indicating high predictive power), while a feature is stably nonpredictive if it is predominantly in the last bin (indicating low or no predictive power). The predictiveness of a feature can also be assessed by the frequency of its occurrence in higher bins (e.g., bin 1 or bin 2).

Thus, a feature's stability and predictiveness are not binary, but rather represent a spectrum. The more frequently a feature remains in the same bin or in higher bins, the more stable or predictive it is considered.

By analyzing the bins and their changes or repetitions across windows, one can gain insights into when a feature is predictive within the time series. The more bins and windows used, the more detailed the insights, though the computational cost increases as well.

The advantage of the method lies in its model independence, meaning it is not tailored to any specific model and can be universally applied. Furthermore, the evaluation metric for binning and stability can be adapted to the specific requirements of the task. Additionally, this method does not require data preprocessing or expert knowledge, as it is purely data-driven.

## 4 MATERIALS AND EXPERIMENTAL SETUP

This section offers an overview of the datasets used to apply the method outlined in our paper. It also details the parameters for the method and the machine learning models. Lastly, the evaluation metrics used to measure

the effectiveness of the proposed method are presented.

### 4.1 DATASETS

The datasets were collected using two DMG machines: a CMX 600 V and a retrofitted DMC 60H milling machine, both equipped with a Siemens Industrial Edge system operating at a sampling rate of 500 Hz. The experiments involved machining aluminum (ALU) and steel (S), with varying feed rates and spindle speeds to create comprehensive datasets. The time series data for these processes were published in references [28] and [29]. The CMX is a newly manufactured machine, while the DMC is an older retrofitted model. The datasets were chosen for their high frequency and axis-specific data, aligning with the objective of this work to predict and understand spindle consumption. The machines captured a total of 52 features across the X, Y, Z axes and spindle parameters (SP), including actual axis position, commanded axis position, commanded speed, contour deviation, control difference, feed rate, load, torque, and torque feed forward. Additionally, the spindle drive current value is used as the target output in this study. For more information on these features, please refer to [30]. Then datasets are labeled as CMX-ALU, CMX-S, DMC-ALU, and DMC-S, indicating the machine and material used.

### 4.2 IMPLEMENTATION

The parameters  $k$ , representing the number of bins, and  $l$ , indicating the size of the partitioned windows, are determined based on the specific use case and the characteristics of the existing dataset. In this study, we focus on the most predictive features, assigning all rankings up to 10 to the first bin, while rankings below 10, totaling 42, are grouped into the second bin. The parameter  $l$  is set to 500 to identify the most frequent bins within this window and to assess the stability of these bins across other windows. The LSTM model architecture comprises three layers with 64, 32, and 25 neurons, respectively. It employs a dropout rate of 0.5, uses the ReLU activation function, and is optimized with the Adam optimizer (learning rate of 0.01). The loss function utilized is the mean squared error (MSE). To facilitate the use of raw data from different machines, the data was normalized using a standard scaler. The implementation was carried out in Python, utilizing libraries such as, scikit-learn, and TensorFlow.

### 4.3 EVALUATION METRIC

The squared error (SE) metric is utilized to assess the performance of the applied models. It is computed as the

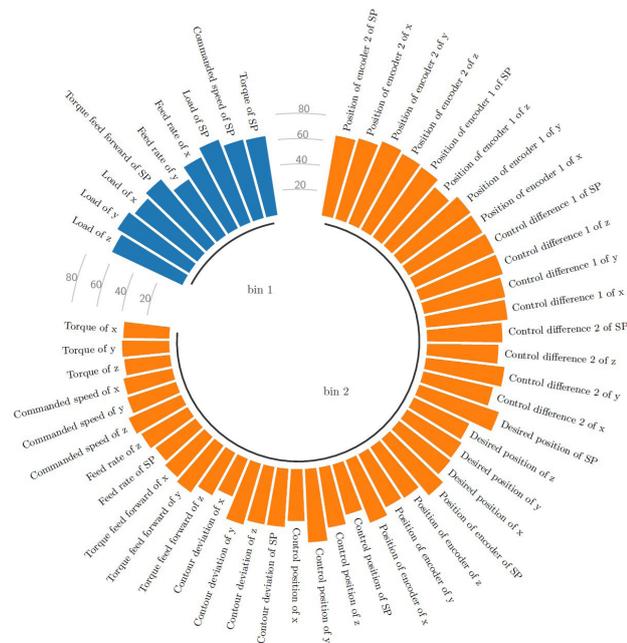
square root of the average of the squared differences between the predicted values  $\hat{y}$  and the actual values  $y$ . This is mathematically defined as follows [31]:

$$SE(y, \hat{y}) = (y - \hat{y})^2 \tag{3}$$

It is important to note that larger errors have a disproportionately higher impact on the SE value compared to smaller errors, due to the nature of squaring the differences. The optimal SE value is 0, with the worst-case reaching  $+\infty$ .

### 5 RESULT AND DISCUSSION

Our paper presents the features of the four datasets to which the applied method was used, categorized according to their bin assignment. Figure 1 illustrates the stability of all features across all windows. The length of the bars represents the frequency of bin occurrences for each feature, while the color indicates the final bin assignment. The most frequent value within each window and across all windows was chosen as the final bin. This result pertains to dataset CMX-ALU, with comparable outcomes observed across the other datasets as well. Table 1 highlights all features that consistently appeared in the first bin across all four datasets. In contrast to Table 1, Table 2 displays five of the features that consistently appeared in the last bin across all four datasets after the method applied in this paper was used.



**Fig. 1:** The bin assignment for each feature and its stability across all windows. The length of the bars represents the frequency of occurrence of the bin values for each feature, while the color indicates the final bin assignment. This result relates to the dataset CMX-ALU.

It was found that certain features, such as spindle load, spindle torque, X and Y axis feed rate, spindle commanded speed, and spindle torque feed forward, x, y, and z axis load, consistently ranked among the best across all four datasets and remained stable, demonstrating strong predictive power.

Features	bin Assignment
Torque of spindle	1
Commanded speed of spindle	1
Load of spindle	1
Feed rate of x-axis	1
Feed rate of y-axis	1
Torque feed forward of spindle	1

**Tab. 1:** The final bin assignment of the best features, which are consistent across all four datasets where the method is applied, shows that the top binning occurred most frequently across all windows.

Features	bin Assignment
Positions of encoder 1	2
Position of encoder 2	2
Actual and commanded position of the axis	2
Control difference 1	2
Control difference 2	2

**Tab. 2:** Five of the least predictive features and their final bin assignment are based on the most frequent value within the window. This applies to all four datasets after the method was applied.

In contrast, features like the positions of encoder 1 and 2, actual and target axis positions, and control differences 1 and 2 were consistently among the worst, maintaining their status as stable but non-predictive. Other features, such as torque of x, y and z-axis, x, y, and z commanded speed, were unstable, with their predictive power fluctuating over time.

These results were partly expected, as energy consumption is influenced by the speed of the axes required to remove material during machining. Additionally, spindle load and torque are critical factors, given their correlation with the energy required for material removal. Conversely, positional data is not a reliable indicator of energy consumption, as it provides limited insight into the machine's overall energy use.

The findings highlight the importance of assessing both the predictive power of features and their stability over time. Further research is needed to explore the interaction

between features and their collective predictive power over time. Incorporating physical knowledge into data collection, feature selection, and dataset composition are important to provide enhanced insights. Additionally, integrating analytical knowledge may provide deeper understanding. Expanding this analysis to a broader range of datasets, materials, machines, and features will be crucial to further improving our understanding of the role features play in predicting energy consumption.

## 6 CONCLUSION

In our work, we developed a method to categorize features based on their importance in predicting energy consumption. Additionally, the stability of these categories was quantified to determine if the features consistently remained in the same bins over time. The raw machine signal data was fed directly into the machine learning model without pre-processing, relying on technical knowledge of the processes. The method was applied to four datasets from two machines using two different materials, enabling a comprehensive analysis. It offers deeper insights into the relationship between features and energy consumption, helping to understand how features contribute to predictions over time. As a result, certain features can be identified for closer scrutiny in the future, while others may be omitted, enhancing transparency and revealing process interdependencies. These transparent results allow for accurate energy consumption predictions, promoting sustainable production and cost reduction.

Building on the work of [19] and our earlier work [24], future research will focus on gathering more extensive datasets to enhance our understanding of machine tool energy consumption. One key challenge is determining the importance of features over time. Although our current method is model-agnostic, it does not consider interactions between features. In future work, we plan to extend this approach by incorporating the interactions between features, as these relationships may provide valuable insights into the predictive power and stability of the features. Additionally, the model relies solely on raw signal data, without incorporating relevant domain knowledge. To overcome these limitations, we intend to develop more sophisticated techniques for assessing feature importance, integrate process-specific expertise into the model, and further explore the stability of feature importance over time to achieve a more comprehensive understanding.

## ACKNOWLEDGMENT

The IGF project 22849 BG of the research association Gesellschaft zur Förderung angewandter Informatik e.V. (GFaI) was supported via the AiF in a program to promote the Industrial Community Research and Development (IGF) by the Federal Ministry for Economic Affairs and Climate Action on the basis of a resolution of the German Bundestag.

## REFERENCES

- [1] M. Soori, B. Arezoo, and R. Dastres, "Machine learning and artificial intelligence in CNC machine tools, a review," *Sustainable Manufacturing and Service Economics*, vol. 2, 2023, 100009.
- [2] W. Liu, "Content architecture and future trends of energy efficiency research on machining systems," *Journal of Mechanical Engineering*, vol. 49, no. 19, pp. 87–94, 2013.
- [3] T. Gutowski, J. Dahmus, and A. Thiriez, "Electrical energy requirements for manufacturing processes," in 13th CIRP international conference on life cycle engineering, vol. 31. Leuven, Belgium, 2006, pp. 623–638.
- [4] W. H. Choi, J. Kim, and J. Y. Lee, "Development of fault diagnosis models based on predicting energy consumption of a machine tool spindle," *Procedia Manufacturing*, vol. 51, pp. 353–358, 2020.
- [5] P. S. Bilga, S. Singh, and R. Kumar, "Optimization of energy consumption response parameters for turning operation using Taguchi method," *Journal of Cleaner Production*, vol. 137, pp. 1406–1417, 2016.
- [6] E. Efficiency, "Tracking industrial energy efficiency and CO2 emissions," *International Energy Agency*, vol. 34, no. 2, pp. 1–12, 2007.
- [7] M. Bornschlegl, S. Kreitlein, M. Bregulla, and J. Franke, "A method for forecasting the running costs of manufacturing technologies in automotive production during the early planning phase," *Procedia CIRP*, vol. 26, pp. 412–417, 2015.
- [8] W. Cai, F. Liu, O. Dinolov, J. Xie, P. Liu, and J. Tuo, "Energy benchmarking rules in machining systems," *Energy*, vol. 142, pp. 258–263, 2018.
- [9] J. Pan, C. Li, Y. Tang, W. Li, and X. Li, "Energy consumption prediction of a CNC machining process with incomplete data," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 5, pp. 987–1000, 2021.
- [10] R. Ströbel, A. Bott, L. Hutt, S. Groß, and J. Fleischer, "Untersuchung des Umgangs produzierender Unternehmen mit Energieverbrauchsmonitoring und -vorhersage," *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, vol. 119, no. 1-2, pp.80–84, 2024. [Online]. Available: <https://doi.org/10.1515/zwf-2024-1009>
- [11] S. A. Bagaber and A. R. Yusoff, "Multi-objective optimization of cutting parameters to minimize power consumption in dry turning of stainless steel 316," *Journal of Cleaner Production*, vol. 157, pp. 30–46, 2017.
- [12] Z. M. Bi and L. Wang, "Energy modeling of machine tools for optimization of machine setups," *IEEE Transactions on Automation Science and Engineering*, vol. 9, no. 3, pp. 607–613, 2012.

- [13] A. Dietmair and A. Verl, "Energy consumption modeling and optimization for production machines," in 2008 IEEE International Conference on Sustainable Energy Technologies. IEEE, 2008, pp. 574–579.
- [14] J. Cao, X. Xia, L. Wang, Z. Zhang, and X. Liu, "A novel CNC milling energy consumption prediction method based on program parsing and parallel neural network," *Sustainability*, vol. 13, no. 24, p. 13918, 2021.
- [15] G. Quintana, J. Ciurana, and J. Ribatallada, "Modelling power consumption in ball-end milling operations," *Materials and Manufacturing Processes*, vol. 26, no. 5, pp. 746–756, 2011.
- [16] G. Kant and K. S. Sangwan, "Predictive modelling for energy consumption in machining using artificial neural network," *Procedia CIRP*, vol. 37, pp. 205–210, 2015.
- [17] S. Borgia, S. Pellegrinelli, G. Bianchi, and M. Leonesio, "A reduced model for energy consumption analysis in milling," *Procedia CIRP*, vol. 17, pp. 529–534, 2014.
- [18] L. A. Duc and N. D. Trinh, "A new study for prediction and optimisation of energy consumption during high-speed milling," *International Journal of Computer Integrated Manufacturing*, vol. 35, no. 12, pp. 1352–1377, 2022.
- [19] R. Ströbel, Y. Probst, S. Deucker, and J. Fleischer, "Time series prediction for energy consumption of computer numerical control axes using hybrid machine learning models," *Machines*, vol. 11, no. 11, p. 1015, 2023.
- [20] C. Camposeco-Negrete, J. de Dios Calderón Nájera, and J. C. Miranda-Valenzuela, "Optimization of cutting parameters to minimize energy consumption during turning of AISI 1018 steel at constant material removal rate using robust design," *The International Journal of Advanced Manufacturing Technology*, vol. 83, no. 5, pp. 1341–1347, 2016.
- [21] Z. Liu, Y. Guo, M. Sealy, and Z. Liu, "Energy consumption and process sustainability of hard milling with tool wear progression," *Journal of Materials Processing Technology*, vol. 229, pp. 305–312, 2016.
- [22] F. Draganescu, M. Gheorghe, and C. Doicin, "Models of machine tool efficiency and specific consumed energy," *Journal of Materials Processing Technology*, vol. 141, no. 1, pp. 9–15, 2003.
- [23] M. Kolar, J. Vyroubal, and J. Smolik, "Analytical approach to establishment of predictive models of power consumption of machine tools' auxiliary units," *Journal of Cleaner Production*, vol. 137, pp. 361–369, 2016.
- [24] H. Kader, R. Ströbel, A. Puchta, J. Fleischer, B. Noack, and M. Spiliopoulou, "Feature ranking for the prediction of energy consumption on CNC machining processes," in 2024 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), 2024, pp. 1–7.
- [25] K. Smagulova and A. P. James, "A survey on LSTM memristive neural network architectures and applications," *The European Physical Journal Special Topics*, vol. 228, no. 10, pp. 2313–2324, 2019.
- [26] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A comparison of ARIMA and LSTM in forecasting time series," in 2018 17th IEEE international conference on machine learning and applications (ICMLA). IEEE, 2018, pp. 1394–1401.
- [27] R. Fu, Z. Zhang, and L. Li, "Using LSTM and GRU neural network methods for traffic flow prediction," in 2016 31st Youth academic annual conference of Chinese association of automation (YAC). IEEE, 2016, pp. 324–328.
- [28] R. Ströbel, Y. Probst, and J. Fleischer, "Training and validation dataset of milling processes for time series prediction," Karlsruhe Institute of Technology. doi: 10.5445/IR/1000157789, 2023.
- [29] R. Ströbel, M. Mau, S. Deucker, and J. Fleischer, "Training and validation dataset 2 of milling processes for time series prediction," Karlsruhe Institute of Technology. doi: 10.35097/1738, 2023.
- [30] P. Gönzheimer, R. Ströbel, R. Dörfinger, M. Mattes, and J. Fleischer, "Interoperable system for automated extraction and identification of machine control data in brownfield production," *Manufacturing Letters*, vol. 35, pp. 915–925, 2023.
- [31] T. O. Hodson, T. M. Over, and S. S. Foks, "Mean squared error, deconstructed," *Journal of Advances in Modeling Earth Systems*, vol. 13, no. 12, e2021MS002681, 2021.

# Feedback-driven object detection and iterative model improvement for accurate annotations

Sönke Tenckhoff\*, Mario Koddenbrock\*, and Erik Rodner

\*These authors contributed equally to this work.

KI-Werkstatt/Fachbereich 2, University of Applied Sciences Berlin,  
Wilhelminenhofstr. 75A, 12459 Berlin, Germany  
firstname.lastname@htw-berlin.de

**Abstract.** Automated object detection has become increasingly valuable across diverse applications, yet efficient, high-quality annotation remains a persistent challenge. In this paper, we present the development and evaluation of a platform designed to interactively improve object detection models. The platform allows uploading and annotating images as well as fine-tuning object detection models. Users can then manually review and refine annotations, further creating improved snapshots that are used for automatic object detection on subsequent image uploads—a process we refer to as semi-automatic annotation resulting in a significant gain in annotation efficiency.

Whereas iterative refinement of model results to speed up annotation has become common practice, we are the first to quantitatively evaluate its benefits with respect to time, effort, and interaction savings. Our experimental results show clear evidence for a significant time reduction of up to 53% for semi-automatic compared to manual annotation. Importantly, these efficiency gains did not compromise annotation quality, while matching or occasionally even exceeding the accuracy of manual annotations. These findings demonstrate the potential of our lightweight annotation platform for creating high-quality object detection datasets and provide best practices to guide future development of annotation platforms.

The platform is open-source, with the frontend and backend repositories available on GitHub<sup>1</sup>. To support the understanding of our labeling process, we have created an explanatory video demonstrating the methodology using microscopy images of *E. coli* bacteria as an example. The video is available on YouTube<sup>2</sup>.

**Keywords.** object detection · global average times · semi automatic annotation · bounding boxes · active learning

<sup>1</sup> <https://github.com/ml-lab-htw/iterative-annotate>

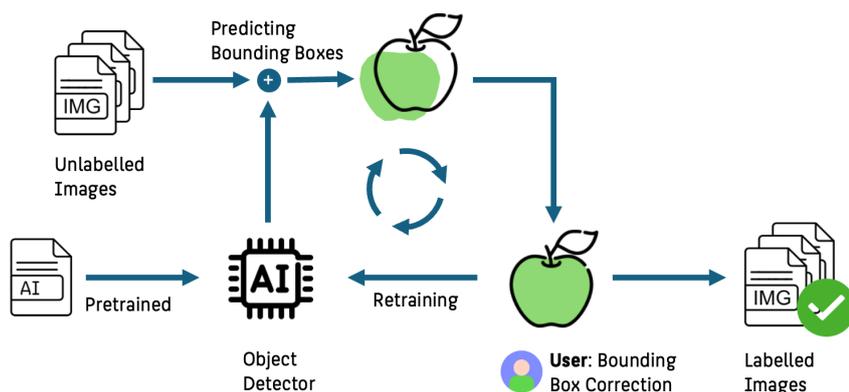
<sup>2</sup> <https://www.youtube.com/watch?v=CM9uhE8NN5E>

## 1 INTRODUCTION

Object detection has become a critical component in various computer vision applications, including but not limited to autonomous driving [19], surveillance [7], robotics [9], microscopy [13], and manufacturing [3]. These applications rely on high-quality labeled datasets to train and fine-tune models. Traditionally, creating such datasets requires manually tagging images—a process that is both labor-intensive and prone to human error. As object detection models continue to evolve, so do the strategies for efficiently generating these annotated datasets, which has become a key research focus [21,8].

While accurate, manual annotation is not scalable for large datasets. Consequently, there has been growing interest in methods that integrate machine learning techniques to assist with image labeling, particularly in semi-automatic annotation workflows [5,4]. Semi-automatic annotation bridges the gap between fully manual and fully automated annotation by combining the strengths of machine learning and human expertise. Fully automated systems often struggle with rare or ambiguous cases, while semi-automatic methods allow human annotators to correct these mistakes, resulting in higher accuracy and efficiency. This approach promises to significantly improve efficiency without compromising the quality of annotations.

In this paper, we introduce a platform designed to enhance the object detection labeling process through an interactive, feedback-driven loop that iteratively refines model-generated annotations. Our platform allows users to create and manage annotation projects, upload image bundles, and apply a pre-trained Single Shot Detector (SSD) [12] to perform initial object detection and labeling. Users can then manually adjust the generated labels, which are used to incrementally fine-tune the model's performance, creating snapshots for future predictions. This iterative process, which we refer to as semi-automatic annotation, aims to reduce annotation time while maintaining high accuracy. The workflow of the labeling process is illustrated in Fig. 1.



**Fig. 1:** Illustration of the iterative annotation workflow in our platform. The process begins with a pre-trained object detection model predicting bounding boxes on unlabeled images. Users then correct these predictions, and the refined annotations are fed back into the model for incremental improvement. This feedback loop progressively enhances model accuracy, reducing manual annotation effort over time.

We evaluate the platform by conducting quantitative experiments to assess the efficiency of semi-automatic annotations versus fully manual annotation workflows. Our results demonstrate a significant time-saving of up to 53% without compromising annotation quality. The findings presented in this paper provides further quantitative insights and best practices with respect to semiautomatic annotation.

## 2 RELATED WORK

Efficient image annotation is essential for training high-quality object detection models. Manual annotation tools such as LabelMe [16] and VGG Image Annotator [6] laid the groundwork by providing interfaces for human annotators to label images. While accurate, manual annotation is time-consuming and not scalable for large datasets.

Iterative refinement or semi-automatic annotation is indeed a common strategy (see [1,17,2,14] just for a few examples). However, a extensive quantitative evaluation of the resulting speed-up as well as suitable best practices are missing in the literature. Semi-automatic annotation systems, such as Amazon Sage- Maker Ground Truth [4], combine machine learning with human-in-the-loop corrections, speeding up the annotation process without sacrificing quality. Similarly, Papadopoulos et al. [15] demonstrated how interactive tools could reduce the time required by allowing users to refine pre-labeled bounding boxes generated by object detectors.

The concept of iterative refinement has also been extensively explored from an active learning perspective, where algorithms select suitable yet unlabeled examples to be annotated or reviewed [18]. Tools like LabelStudio [20] apply this feedback loop in object detection, enabling the model to learn from progressively refined annotations.

Our platform builds on these advancements by combining the strengths of semi-automatic annotation with an iterative refinement mechanism. This approach reduces the manual annotation workload while continuously improving model accuracy. Our approach can be of course combined with active learning techniques, however, we decided to focus on semi-automatic annotation only.

## 3 PLATFORM DESIGN AND IMPLEMENTATION

Building on these previous advancements, our work focuses on the integration of semi-automatic annotation with iterative model refinement to address the limitations observed in manual workflows. The platform leverages semi-automatic annotation, where users correct initial model predictions, and these corrections are used to incrementally improve the model. This process is shown in Fig. 1. It involves:

1. Generating predictions using the SSD model.
2. Allowing users to refine the annotations.
3. Using the refined annotations to fine-tune the model incrementally.

### 3.1 OBJECT DETECTION MODEL

At the core of the platform's object detection capabilities is a standard SSD model[12]. We employ an SSD300 model pre-trained on the COCO dataset [10], chosen for its balance between speed and accuracy. SSD predicts bounding boxes and object classes in a single forward pass, making it ideal for real-time annotation tasks.

### 3.2 USER INTERFACE

The platform's user interface (UI) is intuitive and user-friendly, facilitating efficient annotation of large image datasets. It consists of several components guiding users through project creation, annotation, and model fine-tuning.

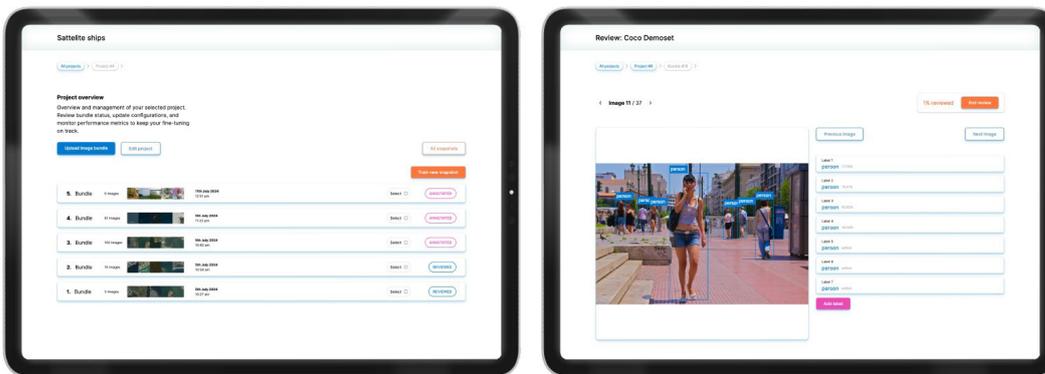
**Project Management Interface** The project management interface allows users to create, manage, and monitor annotation projects. Fig. 2 (left) shows a summary of active projects, with key statistics like uploaded image bundles and fine-tuning status. Users can start new projects, upload images, and review progress through detailed metrics.

**Annotation Editor** The annotation editor enables users to review and refine bounding boxes generated by the model. Fig. 2 (right) shows the interface for adjusting, resizing, or deleting boxes. Users can also add new boxes for missed objects, improving annotation quality through real-time feedback.

**Fine-Tuning Setup** After refining annotations, users initiate model fine-tuning. The user interface allows selection of image bundles, model snapshots, and training configurations. Performance metrics, such as accuracy and loss values, help users track progress and make decisions on further training.

**Snapshot Management** Snapshot management ensures version control during fine-tuning. Users can view different versions, compare performance, and roll back when needed, aiding experimentation.

**User Interaction and Shortcuts** The UI supports keyboard shortcuts for actions like navigating images and adjusting labels, improving efficiency. The platform is also optimized for minimal interaction, ensuring that users can complete the annotation and fine-tuning processes with as few clicks as possible.



**Fig. 2:** The user interface (UI) of the annotation platform. The left pane shows the project management interface, where users can manage projects, view uploaded image bundles, and monitor the status of model fine-tuning. The right pane displays the annotation editor, where users can review and adjust predicted bounding boxes, creating high-quality annotations through a semi-automatic workflow.

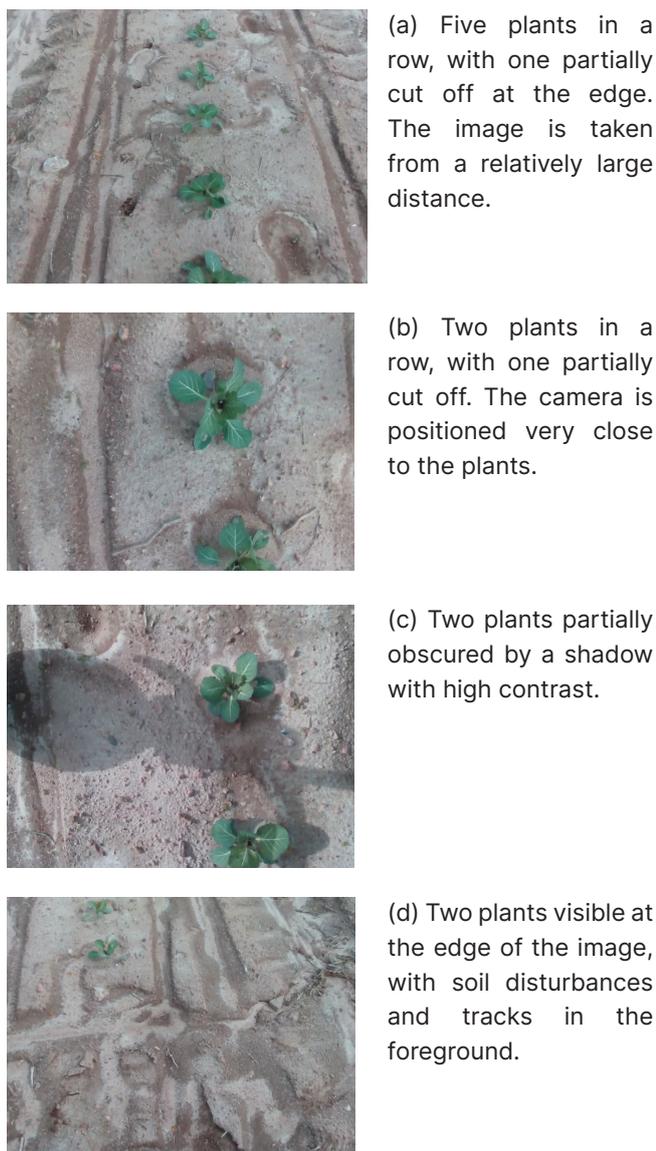
## 4 QUANTITATIVE EVALUATION

### 4.1 DATASETS

The platform was evaluated using the agricultural dataset of [11], selected for its challenging conditions that closely simulate real-world agricultural environments. This dataset, consisting of 200 images captured from an autonomous agricultural robot, includes annotated rows of crops from multiple angles and under varying lighting conditions. For our experiments, the dataset was divided into 20 bundles, each containing 10 images, allowing systematic testing of the platform's adaptability across diverse scenarios. As shown in Fig. 3, these variations—including dense foliage, occlusions, and inconsistent lighting—created a complex environment for the object detection model. Such characteristics enabled us to rigorously assess the platform's performance in handling domainspecific object detection tasks in realistic agricultural monitoring scenarios.

### 4.2 PERFORMANCE MEASURES

To evaluate the performance of the platform, we employed several key metrics commonly used in object detection tasks: a) mean intersection over union (IoU), for measuring bounding box overlap between detection and ground-truth. We also calculated the b) F1 score, which is the harmonic mean of precision and recall. Precision and recall were calculated using an IoU threshold of 0.5 to determine true positives for single detections. The F1 score was computed per image bundle. For the semi-automatic approach, the F1 score was calculated after the automatic labeling was completed, to quantify how much manual intervention was still required to improve accuracy.



**Fig. 3:** Example images from the agricultural dataset [11], captured by an autonomous agricultural robot. These images illustrate some of the challenging conditions encountered during the annotation process, including varying camera distances, shadows, and partial occlusion of plants.

Another critical aspect of the platform’s evaluation was annotation efficiency. We measured the time required for manual annotation and compared it to the time spent using the semi-automatic annotation process. In addition, we tracked specific user interactions, including the number of bounding boxes created, adjusted, or deleted during the process, allowing for a detailed analysis of the effort saved by using the platform. All measurements were performed on the dataset described in section 4.1.

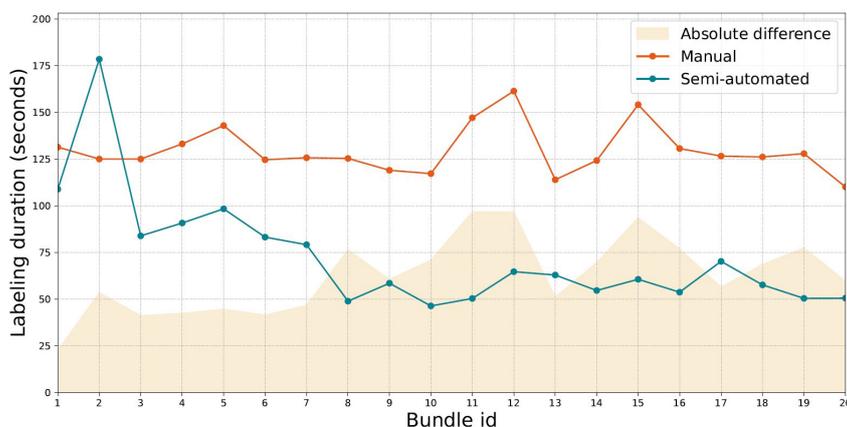
## 5 RESULTS

The experiments conducted in this study demonstrate the effectiveness of the proposed semi-automatic annotation platform. The key findings are summarized as follows:

1. The semi-automatic annotation process reduced annotation time by up to 53.82% compared to manual labeling, especially in later iterations as the model’s predictions improved.
2. User interaction effort, significantly decreased in the semi-automatic process, with nearly no manual box creation needed by the 20th bundle.
3. F1 scores for semi-automatic annotations matched or exceeded the quality of manual annotations, ensuring that the efficiency gains did not compromise accuracy.

### 5.1 EFFICIENCY GAINS

A key objective of this work was to reduce the time required for object annotation. Our experiments, show significant time savings using the semi-automatic annotation platform compared to manual labeling. Fig. 4 provides a comparison of the annotation duration across different bundles. It shows, that the semiautomatic process reduces the time spent on annotation by up to 53.82%, with the most significant reductions observed in later bundles, where the model’s accuracy improved. So as the model becomes more refined, user intervention decreases.



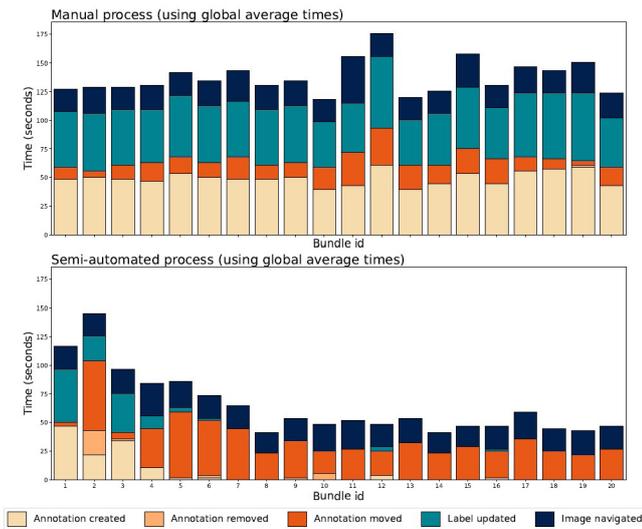
**Fig. 4:** Comparison of the average annotation time per bundle (in seconds) between manual and semi-automatic processes. The results show a significant time reduction using the semi-automatic approach, particularly in later bundles as the model’s accuracy improves, reducing the need for manual intervention.

### 5.2 INTERACTION TIME REDUCTION

In addition to overall time savings, we measured the time required for individual user interactions, such as creating, adjusting, or removing bounding boxes. Fig. 5 illustrates the time taken for each interaction in both manual and semi-automatic workflows. The results show that the semi-automatic process required less time on almost every bundle, with up to 67% less time needed for the creation of bounding boxes compared to the manual approach.

### 5.3 MODEL PERFORMANCE

As shown in Fig. 6 (left), F1 score improved consistently over the course of multiple feedback loops, with the semi-automatic approach achieving comparable performance to the manual approach after just 5 iterations. This indicates that the semi-automatic method significantly reduces the need for manual corrections and can function as a fully automated process after sufficient iterations.



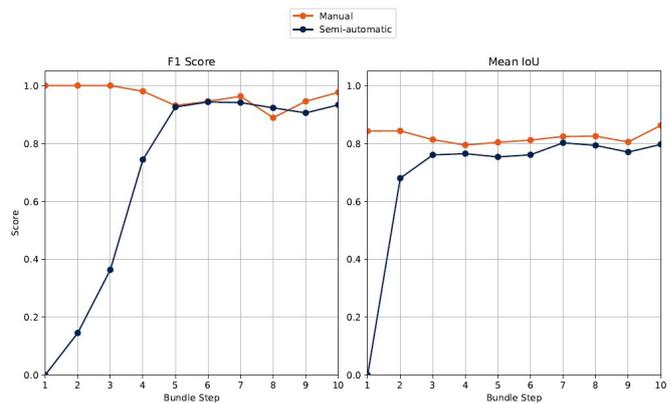
**Fig. 5:** User interaction time metrics comparing the time required for creating, adjusting, or removing bounding boxes in manual and semi-automatic workflows.

### 5.4 ANNOTATION QUALITY

In contrast to detection performance, annotation quality was assessed using Intersection over Union (IoU), a widely recognized metric for object detection tasks. IoU measures the overlap between the predicted bounding box and the ground truth bounding box, providing a value between 0 and 1. A higher IoU indicates better alignment between the predicted bounding boxes and the actual objects in the image, thus reflecting higher annotation accuracy.

In this study, the mean IoU was calculated per image bundle, allowing us to track the improvements in annotation quality over the iterative annotation steps. For the semi-automatic approach, mean IoU was computed after the automatic labeling, to assess how accurately the model was able to label objects without requiring further manual adjustments.

The results indicate a steady increase in mean IoU from 0.68 in the first bundle to 0.83 in the final bundle, as illustrated in Fig. 6 (right). The comparison shows that automatically generated bounding boxes quickly catch up to the accuracy of manual annotations, with the final iterations producing IoU scores that are on par with fully manual approaches. Eventually, the semi-automatic approach transitions into a fully automated process as the model becomes capable of producing high-quality labels independently.



**Fig. 6:** F1 score (left) and mean Intersection over Union (IoU) (right) over multiple iterations of annotation. The analysis is shown up to bundle step 10, because the semi-automatic approach achieved comparable performance to the manual approach by bundle 5, with no further significant divergence in subsequent bundles.

## 6 CONCLUSION

In this work, we presented a feedback-driven, semi-automatic annotation platform designed to iteratively improve object detection models. The primary goal was to evaluate the efficiency of such a platform in reducing the time and effort required for creating high-quality labeled datasets compared to traditional manual annotation methods. Through the integration of user feedback and iterative model fine-tuning, we demonstrated that the proposed system achieves significant gains in annotation efficiency without compromising accuracy.

Our experimental results show that the semi-automatic annotation process can reduce annotation time by up to 53.82% compared to manual labeling, while maintaining or even improving the quality of the annotations. Intersection over Union (IoU) showed consistent improvements across multiple iterations, indicating that the platform effectively enhances the model's performance over time. These results validate the effectiveness of the semi-automatic approach, especially for large-scale datasets, where manual annotation becomes impractical.

The ability to provide feedback to the model in a dynamic, iterative manner highlights the importance of a human-in-the-loop approach, where machine learning techniques complement human expertise to accelerate the dataset creation process.

## FUNDING ACKNOWLEDGMENT

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number 528483508.

## REFERENCES

- [1] Adhikari, B., Huttunen, H.: Iterative bounding box annotation for object detection. In: 2020 25th International Conference on Pattern Recognition (ICPR). pp. 4040–4046. IEEE (2021)
- [2] Aghdam, H.H., Gonzalez-Garcia, A., Weijer, J.v.d., López, A.M.: Active learning for deep detection neural networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3672–3680 (2019)
- [3] Ahmad, H.M., Rahimi, A.: Deep learning methods for object detection in smart manufacturing: A survey. *Journal of Manufacturing Systems* 64, 181–196 (2022)
- [4] Das, P., Ivkin, N., Bansal, T., Rouesnel, L., Gautier, P., Karnin, Z., Dirac, L., Ramakrishnan, L., Perunicic, A., Shcherbatyi, I., et al.: Amazon sagemaker autopilot: a white box automl solution at scale. In: Proceedings of the fourth international workshop on data management for end-to-end machine learning. pp. 1–7 (2020)
- [5] De Gregorio, D., Tonioni, A., Palli, G., Di Stefano, L.: Semiautomatic labeling for deep learning in robotics. *IEEE Transactions on Automation Science and Engineering* 17(2), 611–620 (2019)
- [6] Dutta, A., Zisserman, A.: The via annotation software for images, audio and video. In: Proceedings of the 27th ACM International Conference on Multimedia. pp. 2276–2279 (2019)
- [7] Fedorov, A., Nikolskaia, K., Ivanov, S., Shepelev, V., Minbaleev, A.: Traffic flow estimation with data from a video surveillance camera. *Journal of Big Data* 6, 1–15 (2019)
- [8] García-Aguilar, I., García-González, J., Luque-Baena, R.M., López-Rubio, E.: Automated labeling of training data for improved object detection in traffic videos by fine-tuned deep convolutional neural networks. *Pattern Recognition Letters* 167, 45–52 (2023)
- [9] Griffin, B.: Mobile robot manipulation using pure object detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp.561–571 (2023)
- [10] Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part V 13. pp. 740–755. Springer (2014)
- [11] Liu, D., Li, Z., Wu, Z., Li, C.: Dt/mars-cyclegan: Improved object detection for mars phenotyping robot (2023), <https://arxiv.org/abs/2310.12787>
- [12] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: Ssd: Single shot multibox detector. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. pp. 21–37. Springer (2016)
- [13] Midtvedt, B., Pineda, J., Skärberg, F., Olsén, E., Bachimanchi, H., Wesén, E., Esbjörner, E.K., Selander, E., Höök, F., Midtvedt, D., et al.: Single-shot self-supervised object detection in microscopy. *Nature communications* 13(1), 7492 (2022)
- [14] Munro, R.: Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI. Manning (2021)
- [15] Papadopoulos, D.P., Uijlings, J.R.R., Keller, F., Ferrari, V.: Extreme clicking for efficient object annotation. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 4940–4949 (2017)
- [16] Russell, B.C., Torralba, A., Murphy, K.P., Freeman, W.T.: Labelme: a database and web-based tool for image annotation. *International Journal of Computer Vision* 77(1-3), 157–173 (2008)
- [17] Schmidt, S., Rao, Q., Tatsch, J., Knoll, A.: Advanced active learning strategies for object detection. In: 2020 IEEE intelligent vehicles symposium (IV). pp. 871–876. IEEE (2020)
- [18] Settles, B.: Active learning literature survey (2009)
- [19] Takumi, K., Watanabe, K., Ha, Q., Tejero-De-Pablos, A., Ushiku, Y., Harada, T.: Multispectral object detection for autonomous vehicles. In: Proceedings of the on Thematic Workshops of ACM Multimedia 2017. pp. 35–43 (2017)
- [20] Tkachenko, M., Malyuk, M., Holmanyuk, A., Liubimov, N.: Label Studio: Data labeling software (2020–2024), <https://github.com/HumanSignal/label-studio>, open source software available from <https://github.com/HumanSignal/label-studio>
- [21] Wu, J.Z., Zhang, D.J., Hsu, W., Zhang, M., Shou, M.Z.: Label-efficient online continual object detection in streaming video. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 19246–19255 (2023)

# Enhancing Investment Casting Quality Control: AI-Driven Defect Detection on Ceramic Shells in Investment Casting

Wenchao Tang, Ali Zafar, Alexander Küll, and Hendrik Holling

Access e.V., Intzestraße 5, 52072 Aachen, Germany

**Abstract.** Investment casting is one of the main production processes for aerospace applications. The melt is poured into ceramic shell molds which allow the casting of very fine geometrical features for turbine blades and vane rings. The ceramic shell is therefore crucial for the casting quality. One of the most frequent occurring issues in the ceramic shell production are flow lines. Traditionally, manual inspection is conducted for quality control. Over the years, Access e.V. has accumulated an extensive archive of handmade pictures of ceramic shells, exhibiting great potential for AI applications. This study aimed to evaluate the feasibility of applying two advanced computer vision algorithms, Mask R-CNN and nnU-Net, to detect the flow lines on ceramic shells. A multi-stage approach was applied to expand the training dataset of 17 images to 127 images. Each image was annotated together with ceramic experts. Both Mask R-CNN and nnU-Net algorithms were then tested on the same validation dataset for comparison. The results showed that both demonstrated very good capabilities in detecting the defect, with the nnU-Net model showing superior performance. It was also observed that the model of nnU-Net trained on larger datasets didn't necessarily improve the detection results due to the extra noise introduced, whereas the Mask R-CNN model experienced significant improvement. The results, however, underline the potential of AI-driven automated defect detection. Future work will focus on expanding the models' capability to detect multiple defects and integrating the developed AI models with an automated camera system for real-time quality monitoring. This will streamline the quality control process and reduce significantly the time required by manual inspections for ceramic shell production.

**Keywords.** Investment casting · ceramic shells · defect detection · artificial intelligence · computer vision · Mask R-CNN · nnU-Net.

## 1 INTRODUCTION

Investment casting, also known as lost-wax casting, is a precision manufacturing technique that has been used for centuries to produce intricate components with complex geometries [1, 2]. At the heart of this process are the ceramic shells, which function as molds that define the final shape, accuracy, and quality of the castings [3,4]. These ceramic shells must exhibit high precision, surface integrity and stability to meet the stringent requirements of industries like aerospace, automotive, and biomedical sectors [5].

The ceramic shell is created by first forming a wax pattern of the desired part, which is then repeatedly dipped into a ceramic slurry and coated with fine refractories to build up layers—a process called shelling [6]. After sufficient buildup and drying, the wax is melted away, and the shell undergoes debinding and sintering to achieve the necessary strength, resulting in a hollow mold for casting. Throughout this process, it is challenging to maintain consistent quality, as various problems such as surface cracks, inclusions, shell thickness irregularities, and flow lines can occur [4, 7, 8]. Among these, flow lines occur particularly frequently.

Flow lines are caused by unwanted accumulations of the ceramic slurry and sand mixture on the shell surface [7,9]. This issue often arises due to insufficient draining time, improper shell positioning, undercuts in geometries or high slurry viscosity [7–9]. Flow lines are especially common during the early stages of shell production when process parameters—such as slurry composition, draining time, and handling techniques—are still being optimized. These defects lead to uneven layer application, which changes the local thermal conditions and could result in defects in the castings [8]. Traditionally, ceramic shell quality has been assessed through visual inspection by skilled operators, which is time-consuming and prone to subjectivity and human error [10].

Leveraging advanced deep learning techniques could provide a viable solution for automating the defect detection process, thereby addressing the limitations of human inspection [11–13]. Bhatt et al. [14] showcased how image-based defect detection improved inspection accuracy in aerospace industry while reducing reliance on manual inspection. Birlutiu et al. [15] applied deep learning techniques to detect surface anomalies, where even subtle defects were detected which could even be missed by human inspection. In the casting industry, Ferguson et al. [16] also demonstrated the effectiveness of CNN-based defect detection, using transfer learning to identify small-sized defects of castings with limited datasets.

Despite advances in artificial intelligence, its application in investment casting—specifically for detecting ceramic shell defects remains limited, highlighting a significant gap in the field. Flow lines are particularly challenging to detect due to their thin, irregular appearance and tendency to blend seamlessly with surrounding surface textures. Even deep learning models like convolutional neural networks struggle to identify these defects because they often merge indistinguishably with the material's texture [17]. While AI methods have improved the detection of more apparent surface defects such as cracks and porosity in casting industries [17], flow lines remain understudied. To address the challenge of detecting flow lines in ceramic shells, we studied two advanced deep learning models: Mask R-CNN for instance segmentation and nnU-Net for semantic segmentation.

Mask R-CNN builds upon the Faster R-CNN framework by adding a branch that predicts segmentation masks for each detected region of interest, making it suitable for irregular and varying-sized defects [18, 19]. It is capable of detecting individual defects and precisely localize them, which makes it ideal for isolated or distinct flow lines [18]. Mask R-CNN has proven effective in industries like motor part inspections, where small, subtle defects are critical. Techniques such as data augmentation and transfer learning have enhanced its performance, especially with limited data [20]. Its adaptability to different types of defects makes it a promising choice for ceramic defect detection.

The nnU-Net model, based on the U-Net architecture, features a symmetrical encoder-decoder structure with skip connections [21]. This design retains fine-grained spatial information essential for precise segmentation tasks. Its automated optimization makes it highly adaptable to different datasets, which is crucial for identifying fine-grained continuous patterns that may

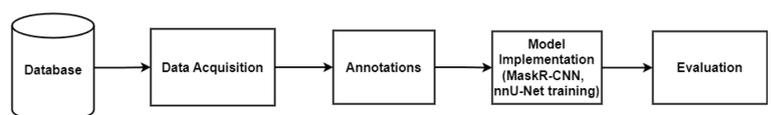
not be easily distinguishable as individual objects [22, 23]. nnU-Net has shown success in identifying subtle cracks due to its ability to automatically adapt to different datasets, enhancing robust pixel-level segmentation [22]. This capability allows nnU-Net to detect fine defects that traditional methods might overlook, consistently achieving high accuracy across various scenarios.

However, each model has limitations. Mask R-CNN may struggle with defects that cannot be neatly separated into distinct instances, while nnU-Net might find it difficult to differentiate overlapping defects or those with ambiguous boundaries. This study aims to test the feasibility of both algorithms in detecting flow lines on ceramic shells in investment casting.

## 2 METHODOLOGY

### 2.1 THE TRAINING PIPELINE

Figure 1 illustrates the general pipeline used in this study to detect flow lines in ceramic shells using Mask R-CNN and nnU-Net. The pipeline involves several steps. First, high-resolution images of ceramic surfaces capturing potential defects are acquired. These images are then annotated by ceramic experts, marking the flow lines. Next, the annotated dataset is used to train Mask R-CNN for instance segmentation and nnU-Net for pixel-level semantic segmentation. The predictions made by these models are then evaluated using metrics introduced in section 2.3.



**Fig. 1:** Illustration of general training pipeline

This process is iterative and is repeated several times as the dataset grows. The first training was conducted with Mask R-CNN using a dataset consisting of only 25 images exhibiting flow lines. These images were hand-picked by ceramic experts and captured under various lighting conditions during real-world shelling processes. The trained Mask R-CNN model was then used to filter a large database of over 2000 handmade ceramic images, pre-select images exhibiting flow lines to expand the dataset. These images were then manually examined to finalize which images to add to the dataset. In all these images, the flow lines were manually annotated using the VGG Image Annotator (VIA) tool under the supervision of experts, to ensure accurate representations of defects and to facilitate effective model training. The size of the

dataset grew iteratively in this manner and eventually resulted in two differently sized training datasets consisting of 41 and 127 images for the comparative study, with another 8 images as a validation dataset serving as the common ground for evaluation of both algorithms. Initially the smaller dataset was tested before introducing the larger one to assess the impact of data size on generalization and learning capacity.

## 2.2 IMPLEMENTATION OF MASK R-CNN AND NNUNET ALGORITHMS

With the key parameters summarized in table 1, the details for both Mask RCNN and nnU-Net that are implemented in this study are introduced.

**Mask R-CNN Implementation** Mask R-CNN allows images to be input in their original formats without extensive preprocessing, preserving the authentic characteristics of the ceramic surfaces. A custom configuration class was implemented, utilizing the ResNet-101 architecture for its deep layers that enhance the extraction of complex features critical for detecting subtle defects [24]. Key parameters such as minimum detection confidence, RPN anchor scales, learning rate, optimizer, and augmentation techniques were carefully selected based on empirical evaluations to optimize performance [25]. The model was trained for approximately 500 epochs, with all layers fine-tuned to leverage comprehensive feature representations. Data augmentation techniques like rotations, flipping, and scaling were applied to simulate diverse defect appearances and prevent overfitting.

**nnU-Net Implementation** The nnU-Net model was implemented following the official guidelines [23], which automatically adapt the architecture and parameters based on the dataset's characteristics. Intensity normalization and binary mask encoding were performed during dataset preparation to ensure consistency [23, 26]. The dataset was structured as per nnU-Net's requirements, including organizing images and labels in a standardized format to ensure smooth model training [23]. The 2D U-Net architecture was employed due to the availability of 2D images. The framework automatically adjusted key parameters such as input patch size, batch size, and network depth for optimal performance [27]. The loss function combined cross-entropy with Dice loss to enhance segmentation accuracy [28]. The model was trained using stochastic gradient descent with Nesterov momentum [29] over 1000 epochs, with a dynamically reduced learning rate based on performance plateaus [30]. Data augmentation techniques, including random rotations, flips, scaling, and intensity shifts, are applied to improve robustness and generalizability [31].

Parameter	Mask R-CNN	nnU-Net
Architecture	Mask R-CNN (ResNet-101)	2D U-Net
Depth	Not specified	9 stages
Feature maps per stage	Not specified	32, 64, 128, 256, 512, 512, 512, 512, 512
Convolution kernel	3 × 3	3 × 3
Patch size	Not applicable	1024 × 1536 pixels
Batch size	2	2
Learning rate	0.001	0.01 - 0.0001
Optimizer	Adam	SGD with Nesterov momentum
Normalization	Not specified	Instance Normalization
Activation function	ReLU	Leaky ReLU
Min. detection confidence	0.9	Not applicable
RPN anchor scales	[32, 64, 128, 256, 512]	Not applicable
Augmentation	Rotation, flipping, scaling	Rotation, flipping, scaling, intensity shifts
Epochs	500	1000
Deep supervision	Not applicable	Enabled
Loss function	Cross-Entropy	Dice + Cross-Entropy
Input channels	3 (RGB)	3 (RGB)
Channel normalization	Not specified	Z-score

**Tab. 1:** Key parameters of the implemented Mask R-CNN and nnU-Net

## 2.3 EVALUATION METRICS

To comprehensively evaluate the performance of the Mask R-CNN [18] and nnUNet [23] algorithms in detecting flow lines on ceramic shells, a multi-faceted evaluation approach was employed, integrating qualitative visual assessments, quantitative metrics, and the customized precision score tailored to the detection objectives.

**Visual Evaluation** The initial phase of the evaluation involved visual inspection. Side-by-side comparisons of original images, ground truth annotations, and model predictions were conducted, providing a direct qualitative assessment of model performance [32]. The following aspects were focused on during the visual inspection:

- **Prediction Results Evaluation:** Determining whether all ground truth defects were successfully identified by both models.
- **False Positive Analysis:** Systematically analyzing instances where models produced false positives,

examining their context and potential causes.

– **Model Comparison:** Comparing the performance of Mask R-CNN and nnU-Net, as well as assessing the impact of dataset size expansion within each model.

Due to the different nature of instance segmentation (Mask R-CNN) and semantic segmentation (nnU-Net), Mask R-CNN predictions were converted to binary masks by merging all detected instances into a single output [33]. This step ensured comparability given the focus on detecting a single defect type. Following the visual evaluation, the models were evaluated quantitatively using established metrics.

**Quantitative Metrics** Several standard quantitative metrics were employed to rigorously evaluate model performance [34, 35]. These metrics are calculated at the pixel level, rather than in the traditional classification sense. Pixels aligning with ground truth mask pixels are considered true positives, providing a detailed assessment of model accuracy in segmenting defect regions. The details are summarized in table 2.

**Calibrated Precision Score** A calibrated precision score was developed as a post-processing strategy to address the limitations of strict pixel-level calculations described above. In practical applications, accurately identifying the presence of a defect is more important than achieving pixel-perfect predictions. However, the standard precision score does not fully capture this capability. For instance, models may be penalized with lower precision scores even when defects are correctly identified, due to slight misalignments in the predicted masks or predictions with very small areas. To better reveal and compare model performance in the practical context, following calibration of standard precision score was implemented.

The first step involved removing noises, specifically small predictions likely to be false positives. This was done by filtering out noise using a threshold derived from the smallest annotation in the ground truth. Specifically, the threshold was set at 20 percent of the pixel count of the smallest ground truth annotation. If any predicted masks contained fewer pixels than the threshold, they were removed, thus enhanced the quality of the predicted masks. The precision scores were then calculated using the standard formula, providing a more meaningful comparison of the models and emphasizing their practical capability to detect defects in real-world scenarios.

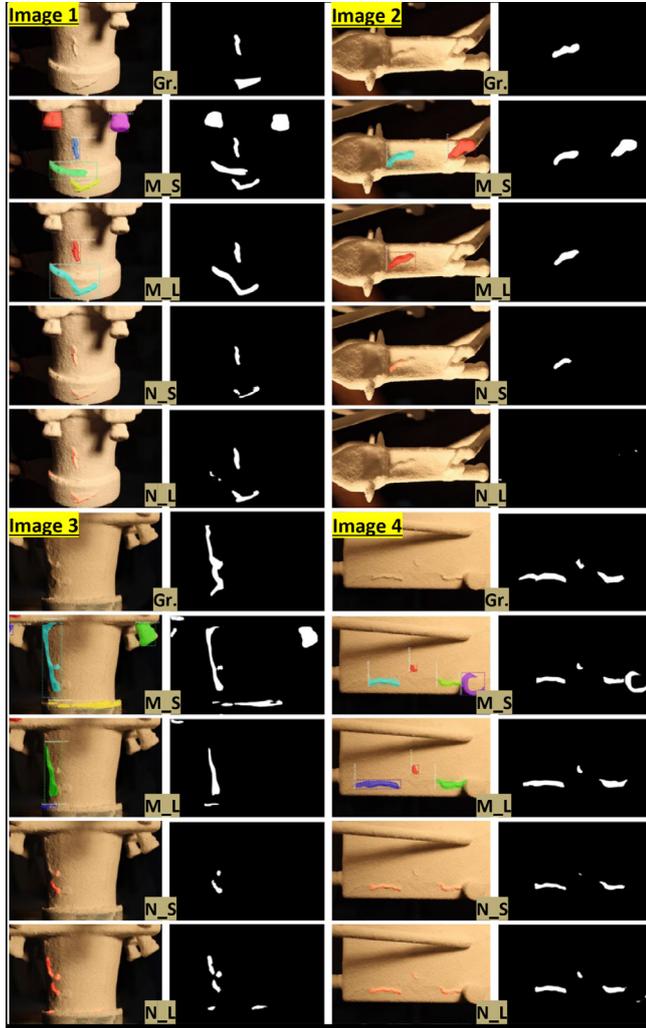
Metric	Description	Formula <sup>1</sup>
Accuracy	Measures the ratio of correctly predicted pixels (both defect and non-defect) to the total number of pixels [34].	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	Ratio of correctly identified defect pixels to all pixels predicted as defects [36].	$\frac{TP}{TP+FP}$
Recall	Ratio of correctly identified defect pixels to all actual defect pixels [36].	$\frac{TP}{TP+FN}$
IoU	Measures the overlap between predicted and ground truth segmentation masks [34].	$\frac{TP}{TP+FP+FN}$
Dice Coefficient	Similar to IoU, assesses the overlap between predicted and ground truth masks [28].	$\frac{2 \times TP}{2 \times TP + FP + FN}$
Hausdorff Distance	Measures the maximum distance between the predicted segmentation and the ground truth [37].	<i>No specific formula</i>

**Tab. 2:** Quantitative metrics for model evaluation

<sup>1</sup>Abbreviations used in the formulas: TP: True Positives; TN: True Negatives; FP: False Positives; FN: False Negatives.

### 3 RESULTS

#### 3.1 QUALITATIVE RESULTS



**Fig. 2:** Qualitative comparison of predicted masks from Mask R-CNN and nnU-Net for 4 validation images. Abbreviations are as follows: Gr.: Ground truth; M\_S: Mask R-CNN trained on the small dataset; M\_L: Mask R-CNN trained on the large dataset; N\_S: nnU-Net trained on the small dataset; N\_L: nnU-Net trained on the large dataset.

Figure 2 presents 4 representative examples of defect detection results from both algorithms, nnU-Net and Mask R-CNN, on the validation dataset containing flow lines. For each algorithm, two models trained on the small (denoted with „S“) and large datasets (denoted with „L“), respectively, are included: Mask R-CNN\_S, Mask R-CNN\_L, nnU-Net\_S, and nnU-Net\_L. To ensure comparability between the semantic segmentation outputs of nnU-Net and the instance segmentation outputs of Mask R-CNN, all detected instances from Mask R-CNN were merged into binary masks.

The visual inspection showed that both models successfully detected most of the ground truth defects across various test images. Notably, nnU-Net\_S produced more precise predictions, closely following the contours of the defects. For example, for image 3 in figure 2, nnU-Net\_S's prediction aligns accurately with the contour of the ground truth annotation, whereas both Mask R-CNN\_S and Mask R-CNN\_L show slight deviations to that. Mask R-CNN also exhibited a higher tendency to produce false positives, especially when trained on the smaller dataset (Mask R-CNN\_S). In the same image, Mask R-CNN\_S incorrectly identified geometrical features as defects, possibly due to the complex surface textures. In contrast, nnU-Net\_S demonstrated better control over false positives in similar areas, indicating a more robust performance against such confounding factors. Comparing the performance between models and within models due to dataset size expansion revealed distinct trends. For Mask R-CNN, increasing the training dataset size to create Mask R-CNN\_L reduced clearly the number of false positives and improved the model's ability to distinguish defects from geometrical features, as seen in all images in figure 2. However, for nnUNet, increasing the dataset size to create nnU-Net\_L did not notably enhance performance. In some cases, nnU-Net\_L even missed defects that nnU-Net\_S detected, resulting in false negatives—for instance, in figure 2, Image 2.

The qualitative results suggest that nnU-Net\_S outperformed the other models, offering higher precision with fewer false positives. The expansion of the dataset size improved Mask R-CNN's performance but had a limited impact on nnU-Net. These findings show the importance of model selection and dataset size in defect detection tasks. To confirm these results and provide further assessment of the models' performance, evaluations using quantitative metrics is conducted and shown below.

#### 3.2 QUANTITATIVE RESULTS

The quantitative metrics introduced in the methodology section are applied to all four models. The confusion matrix on pixel level and the Hausdorff distance for each model is shown in table 3. Based on that, the result of calculated quantitative metrics is shown graphically in figure 3.

Metric	Mask R-CNN_S	Mask R-CNN_L	nnU-Net_S	nnU-Net_L
<b>Confusion Matrix Elements</b>				
True Negatives	60,364,894	61,134,283	<b>62,062,874</b>	62,095,151
False Positives	1,979,562	1,210,173	<b>281,582</b>	249,305
False Negatives	608,253	<b>446,099</b>	641,719	691,751
True Positives	748,283	<b>910,437</b>	714,817	664,785
Hausdorff Distance	28.1286	21.7297	<b>20.3734</b>	21.9567

Note: Best values for each metric are highlighted in bold.

Tab. 3: Confusion matrix on pixel-level & Hausdorff distance

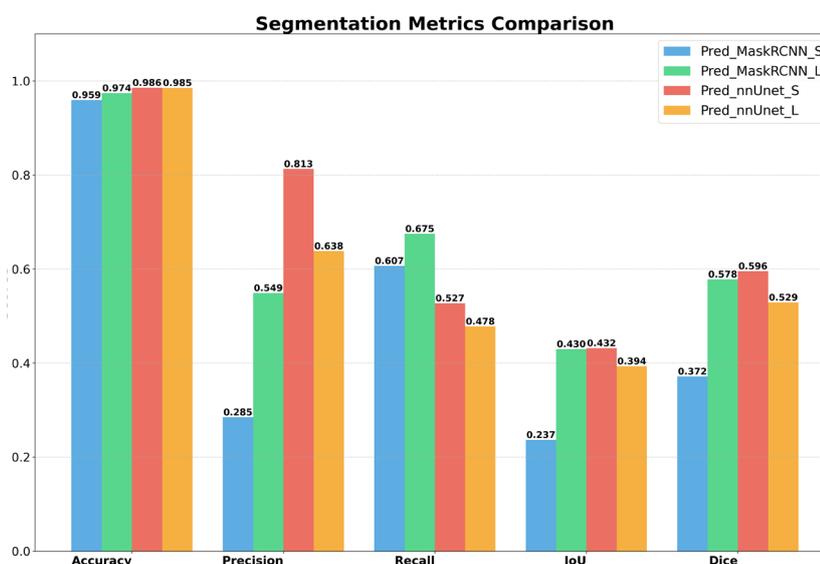


Fig. 3: Comparison of segmentation results with quantitative metrics

**Model performance** All models achieved high overall accuracy, ranging from 95.94% to 98.55%, with nnU-Net\_S performing best at 98.55%. However, due to the class imbalance (with over 60 million true negatives), accuracy alone is not a sufficient indicator of performance. A key distinction emerged between the models in terms of precision and recall, where nnU-Net\_S demonstrated the highest precision at 81.29%, making it more conservative and effective at reducing false positives, though it exhibited lower recall (52.71%), meaning it missed more defects. Conversely, Mask R-CNN\_L achieved the highest recall at 67.50%, indicating its sensitivity in detecting defects but with a lower precision (54.90%), leading to more false positives. The IoU and Dice Coefficient metrics confirmed these trends, with nnU-Net\_S slightly outperforming Mask R-CNN\_L in both IoU (43.19% vs. 42.97%) and Dice Coefficient (59.55% vs. 57.79%). This highlights that while both models offered a balanced performance, nnU-Net\_S exhibited a more cautious approach in its predictions, whereas Mask R-CNN\_L showed a more aggressive approach in capturing potential defects.

**Impact of Dataset Size** The dataset size had a notable impact on the performance of Mask R-CNN models, but less so for nnU-Net models. Transitioning from Mask R-CNN\_S to Mask R-CNN\_L resulted in significant improvements across metrics: precision nearly doubled (from 28.48% to 54.90%) and recall also increased (from 60.69% to 67.50%). This suggests that the Mask R-CNN architecture benefited substantially from the larger dataset, which enhanced its generalization capabilities and reduced false positives. In contrast, the nnU-Net models did not show similar gains. Increasing the dataset size from nnU-Net\_S to nnU-Net\_L actually led to a decline in performance, with precision dropping from 81.29% to 63.80% and recall from 52.71% to 47.80%. This unexpected result suggests that nnU-Net may already perform optimally with a smaller dataset for this specific task, or that additional data introduced complexity that the model struggled to leverage effectively.

**Hausdorff Distance** The Hausdorff Distance, which measures the maximum deviation between predicted and ground truth boundaries, further illustrates the models' performance and dataset's impact. nnU-Net\_S exhibited the lowest Hausdorff Distance (20.37), demonstrating its superior ability to precisely capture the boundaries of defects, which is essential for accurate defect characterization. In contrast, Mask R-CNN\_S had the highest Hausdorff Distance (28.13), indicating less precise boundary delineation. Mask R-CNN\_L showed significant improvement (21.73), confirmed the positive impact of dataset size on this model.

These quantitative results indicate that both algorithms have distinct strengths and trade-offs. For applications requiring high precision and minimal false positives, the model nnU-Net\_S is the preferred choice due to its conservative nature and more accurate boundary

predictions. On the other hand, Mask R-CNN\_L is better suited for applications where detecting as many defects as possible (high recall) is prioritized, even if it means accepting more false positives. The choice of model should depend on the specific requirements of the defect detection task. If missing a defect is more detrimental than dealing with false positives, Mask R-CNN\_L is preferable. Conversely, for tasks where false positives carry a higher cost, nnU-Net\_S is more appropriate.

### 3.3 CALIBRATED PRECISION SCORE RESULTS

Using the custom evaluation metrics described in the methodology, all four models based on their predictions on 8 unseen test images were processed. A 20% threshold was applied to remove noise from the original predictions, and the calibrated precision scores were calculated. Figure 4 illustrates the effect of this post-processing.

Figure 5 presents the maximum, minimum, and average precision scores of the 8 test images and their distributions, both before and after noise removal. It can be observed that Mask R-CNN\_L showed the most improvement after noise removal, with its mean precision increasing from 0.55 to 0.58 and enhanced consistency. nnU-Net\_S had initially the highest mean precision at 0.81 before noise removal but experienced a slight decline to

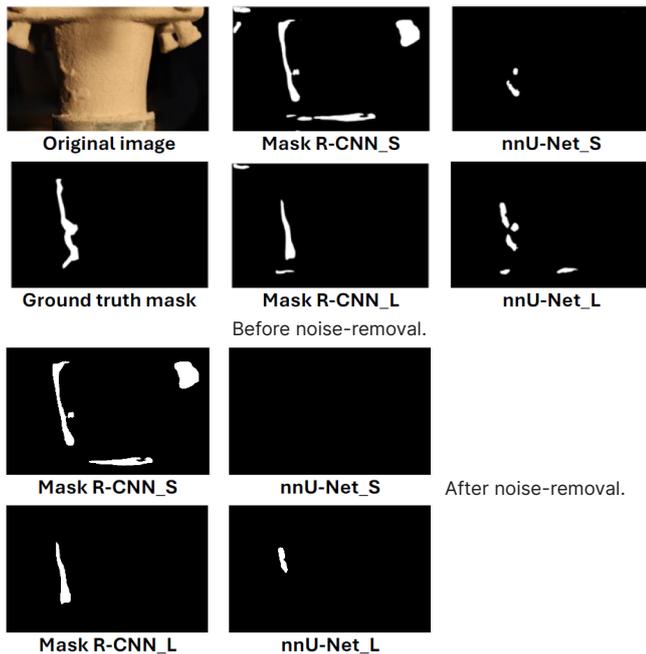


Fig. 4: Example of predicted masks from Mask R-CNN and nnU-Net, before and after noise-removal with 20 percent threshold applied.

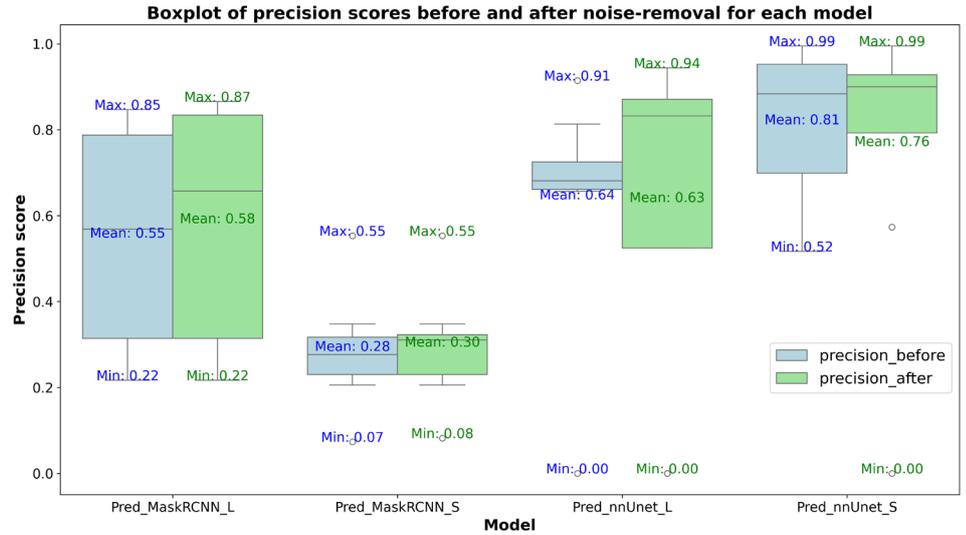


Fig. 5: Comparison of segmentation results before and after noise-removal.

0.76 afterward, along with increased variability. nnU-Net\_L had a minimal drop in mean precision from 0.64 to 0.63 but improved its maximum precision to 0.94. Mask R-CNN\_S exhibited marginal improvement, with its mean precision increasing slightly from 0.28 to 0.30, but remained the weakest model.

Before noise removal, nnU-Net\_S exhibited the highest mean precision and most consistent performance, with nnU-Net\_L also performing well but showing slightly more variability. Mask R-CNN\_L showed moderate performance, while Mask R-CNN\_S struggled with low precision scores, making it the weakest model at this stage. After applying noise removal, the effects varied across the models, where Mask R-CNN\_L benefited the most, with increases in both mean precision and consistency, indicating that noise removal significantly reduced variability and improved performance. In contrast, Mask R-CNN\_S experienced only marginal gains, suggesting limited benefit from this process. Interestingly, nnU-Net\_L had a slight drop in mean precision but improved in top-end performance, while nnU-Net\_S, previously the top performer, saw a minor decline in mean precision and increased variability, particularly in lower-performing predictions. This suggests that noise removal introduced more inconsistency in the nnU-Net models.

Overall, Mask R-CNN\_L benefited most from the noise removal, which could be introduced into the evaluation pipeline to boost model performance. For the nnU-Net models, however, noise removal had a negative effect and is thus not recommended.

## 4 CONCLUSION

The comparative analysis of nnU-Net and Mask R-CNN models for detecting flow lines in ceramic shell images reveals distinct performance characteristics of each model suited to specific application requirements. The key findings are summarized as follows:

- **nnU-Net\_S Model:** Exhibited superior precision and boundary accuracy, making it the ideal choice in scenarios where minimizing false positives is critical.
- **Mask R-CNN\_L Model:** Demonstrated higher recall, successfully identifying more defects, but with an increase in false positives. This makes it suitable in cases where avoiding missed defects is a priority.
- **Impact of Dataset Size:** Mask R-CNN showed significant improvements with a larger dataset, leading to better generalization and fewer false positives. In contrast, nnU-Net's performance plateaued and even slightly declined with increased data size.
- **Post-processing Effects:** Noise removal post-processing improved the performance of Mask R-CNN\_L but had a detrimental effect on nnU-Net models.

These findings demonstrate the importance of selecting appropriate models and post-processing strategies based on specific operational requirements. Despite these differences, the study successfully verified the feasibility of both algorithms for detecting flow lines in ceramic shells.

## 5 FUTURE WORK

Future work will focus on further improving the performance of the models. For nnU-Net, this could involve optimizing its ability to leverage larger datasets, potentially through further data augmentation techniques or architectural modifications to prevent the decline in performance observed with increased data. Additionally, exploring alternative post-processing methods specifically tailored for nnU-Net may help mitigate the negative effects seen with the noise removal.

For Mask R-CNN, strategies to enhance precision without sacrificing recall, such as employing more sophisticated loss functions or adjusting class weight balancing, could further reduce the false positives.

Furthermore, automating the data acquisition process for ceramic shells by implementing a automated camera

system, rather than relying on manual photo capture, could significantly streamline the automated quality control process for ceramic shells and accelerate the data acquisition process, which enables further expansion of the model's capabilities on detecting multiple defects. Implementing further shell defects into the system is also planned.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support provided by the Federal Ministry for Economic Affairs and Climate Action (BMWK) for the LUFO FAST Project 'Konzeption und Entwicklung von integrierten, innovativen und digital vernetzten Prozessketten mit extrem schnellen Fertigungszyklen als Basis einer agilen Produktion für zukünftige hybridelektrische Luftfahrtantriebe' (20L2105C). The authors also thank Dr. Jürgen Jakumeit at Access e.V. for the fruitful discussions that enriched our work.

## REFERENCES

- [1] Campbell, J. (2015). *Complete Casting Handbook: Metal Casting Processes, Metallurgy, Techniques and Design*. Elsevier.
- [2] Beeley, P. R. (2001). *Foundry Technology*. Butterworth-Heinemann.
- [3] Kanyo, J. E., Chlup, Z., Galusek, D., & Galusek, D. (2020). An overview of ceramic molds for investment casting of nickel superalloys. *Journal of the European Ceramic Society*, 40(15), 4955–4973.
- [4] Jones, S., & Yuan, C. (2003). Advances in shell moulding for investment casting. *Journal of Materials Processing Technology*, 135(2–3), 258–265.
- [5] Bünck, M., et al. (2017). Titanium aluminide casting technology development. *JOM*, 69, 2565–2570.
- [6] Bünck, M., Salber, R., & Stoyanov, T. (2024). Resource-Efficient Manufacturing Technology for Titanium Aluminide Aerospace Components. *Transactions of the Indian National Academy of Engineering*, 9(1), 141–154.
- [7] Sidhu, B. S., Kumar, P., & Mishra, B. K. (2008). Effect of slurry composition on plate weight in ceramic shell investment casting. *Journal of Materials Engineering and Performance*, 17(4), 489–498.
- [8] Dong, R.-Z., Wang, Y.-Y., Li, J.-G., & Chen, Q. (2024). An investigation of ceramic shell thickness uniformity and its impact on precision in turbine blade investment casting. *Journal of Manufacturing Processes*, 131, 507–522.
- [9] Chakrabarti, B. K. (2002). Drying conditions and their effect on ceramic shell investment casting process. *Materials Science and Technology*, 18(8), 935–940.
- [10] Kujawińska, A., & Vogt, K. (2015). Human factors in visual quality control. *Management and Production Engineering Review*, 6(4), 65–77.

- [11] Kuo, J.-K., Chiou, Y.-C., Tseng, T.-C., & Hung, C.-H. (2022). Inspection of sandblasting defect in investment castings by deep convolutional neural network. *The International Journal of Advanced Manufacturing Technology*, 120(3), 2457–2468.
- [12] Yousef, N., & Sata, A. (2024). Implementing deep learning-based intelligent inspection for investment castings. *Arabian Journal for Science and Engineering*, 49(2), 2519–2530.
- [13] Cumbajin, E., Sihuay, K. G., Toma, R. S., & Loayza, F. A. (2023). A real-time automated defect detection system for ceramic pieces manufacturing process based on computer vision with deep learning. *Sensors*, 24(1), 232.
- [14] Bhatt, P. M., Malhan, R. K., Rajendran, P., Shah, B. C., Thakar, S., Yoon, Y. J., & Gupta, S. K. (2021). Image-Based Surface Defect Detection Using Deep Learning: A Review. *Journal of Computing and Information Science in Engineering*, 21(4), 040801. <https://doi.org/10.1115/1.4049535>
- [15] Birlutiu, A., Burlacu, A., Kadar, M., & Onita, D. (2017). Defect Detection in Porcelain Industry Based on Deep Learning Techniques. 19th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 263–270. <https://doi.org/10.1109/SYNASC.2017.00049>
- [16] Ferguson, M., Ak, R., Lee, Y.-T. T., & Law, K. H. (2017). Automatic Localization of Casting Defects with Convolutional Neural Networks. *IEEE International Conference on Big Data (Big Data)*, 1726–1735. <https://doi.org/10.1109/BigData.2017.8258115>
- [17] Dučić, N., Manasijević, S., Jovičić, A., Čojbašić, Ž., & Radiša, R. (2022). Casting Process Improvement by the Application of Artificial Intelligence. *Applied Sciences*, 12(7), 3264. <https://doi.org/10.3390/app12073264>
- [18] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 2961–2969).
- [19] Zhou, X., Nie, Y., Wang, Y., Cao, P., Ye, M., Tang, Y., & Wang, Z. (2020). A Realtime and High-efficiency Surface Defect Detection Method for Metal Sheets Based on Compact CNN. 13th International Symposium on Computational Intelligence and Design (ISCID), 259–264. <https://doi.org/10.1109/ISCID51228.2020.00064>
- [20] Zhu, H., Wang, Y., & Fan, J. (2022). IA-Mask R-CNN: Improved Anchor Design Mask R-CNN for Surface Defect Detection of Automotive Engine Parts. *Applied Sciences*, 12(13), 6633. <https://doi.org/10.3390/app12136633>
- [21] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS (Vol. 9351, pp. 234–241).
- [22] Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., & Maier-Hein, K. H. (2018). nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. *arXiv*. <https://arxiv.org/abs/1809.10486v1>
- [23] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2), 203–211.
- [24] Tahir, H., Iftikhar, A., & Mumraiz, M. (2021). Forecasting COVID-19 via Registration Slips of Patients using ResNet-101 and Performance Analysis and Comparison of Prediction for COVID-19 using Faster R-CNN, Mask R-CNN, and ResNet-50. *International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, 1–6. <https://doi.org/10.1109/ICAECT49130.2021.9392487>
- [25] Ocer, N. E., Kaplan, G., Erdem, F., Kucuk Matci, D., & Avdan, U. (2020). Tree Extraction from Multi-scale UAV Images using Mask R-CNN with FPN. *Remote Sensing Letters*, 11(9), 847–856. <https://doi.org/10.1080/150704X.2020.1784491>
- [26] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Lecture Notes in Computer Science (Vol. 11045, pp. 3–11).
- [27] Heller, N., Isensee, F., Maier-Hein, K. H., et al. (2021). The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 Challenge. *Medical Image Analysis*, 67, 101821.
- [28] Milletari, F., Navab, N., & Ahmadi, S. A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *Fourth International Conference on 3D Vision (3DV)* (pp. 565–571).
- [29] Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013). On the importance of initialization and momentum in deep learning. In *International Conference on Machine Learning (ICML)* (pp. 1139–1147).
- [30] Smith, L. N. (2017). Cyclical learning rates for training neural networks. In *IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 464–472).
- [31] Perez, L., & Wang, J. (2017). The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *arXiv preprint arXiv:1712.04621*.
- [32] Cahill, N. D., Hwang, J. Y., & Prakash, B. A. (2010). Quantitative methods for comparing segmented images. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)* (pp. 4665–4668).
- [33] Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature Pyramid Networks for Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2117–2125).
- [34] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2), 303–338.
- [35] Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Medical Imaging*, 15(1), 29.
- [36] Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.
- [37] Huttenlocher, D. P., Klanderman, G. A., & Rucklidge, W. J. (1993). Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9), 850–863.

# Humanity-Centered Production – a Spotlight on Future Productivity

Jörg Reiff-Stephan

TH Wildau, Wildau 15745, Germany  
jrs@th-wildau.de

**Abstract.** The effects of unexpected global events, such as the COVID-19 pandemic, the Russian-Ukrainian war and the energy crisis, on industrial production processes are immanent and demand a transformation of behavior. In industrial development, the change from Industry 4.0 to a new stage is taking place. The transformation known as Industry 5.0 emphasizes close collaboration between humans and machines and pursues a human-centric, sustainable and resilient approach. The focus is on the concept of Humanity-Centered Production (HCP), which takes into account not only individual human needs, but also the entire ecosystem and long-term effects. In contrast to the human-centered approach, HCP expands the focus to the whole of humanity and the environment. The principles of HCP include a systemic view of production systems, a longterm perspective and community involvement. HCP also aims to achieve five of the 17 UN Sustainable Development Goals, particularly in the areas of education, decent work, responsible consumption and innovation. The integration of cognitive, ergonomic and social aspects into production systems is intended to promote the competitiveness and sustainability of future production landscapes.

**Keywords:** Humanity-centered Production; Digital Transformation; Sustainable Ecosystem.

## 1 INTRODUCTION

Unexpected events have challenged production processes in recent years, such as the COVID-19 pandemic, the Russian-Ukrainian war and the energy crisis. In addition to the social impact, these events have also disrupted the traditional industrial production system [1]. Industry 4.0 requires adaptability to changing conditions. The digital and sustainable transformation is leading to a new approach: Industry 5.0. This new phase emphasizes greater collaboration between humans and machines and takes a more coordinated approach than Industry 4.0. With a human-centric strategy, Industry 5.0 aims for greater sustainability and resilience [3, 4]. While Industry 4.0 focuses on networking through cyber-physical systems, Industry 5.0 emphasizes the relationship between humans and machines (see Fig. 1).

In Industry 5.0, where complex industrial processes are susceptible to disruption due to the use of modern technologies such as AI, big data analytics and IoT, resilience is essential. The term goes beyond simply enduring difficulties and also emphasizes increased performance and flexibility in the face of setbacks. The need for resilience has been highlighted by the unexpected events described, which implies that organizations must develop systems that can withstand disruption and recover quickly. Resilience is mainly attributed to flexibility and inherent redundancy that enable systems to overcome disruptions or failures.

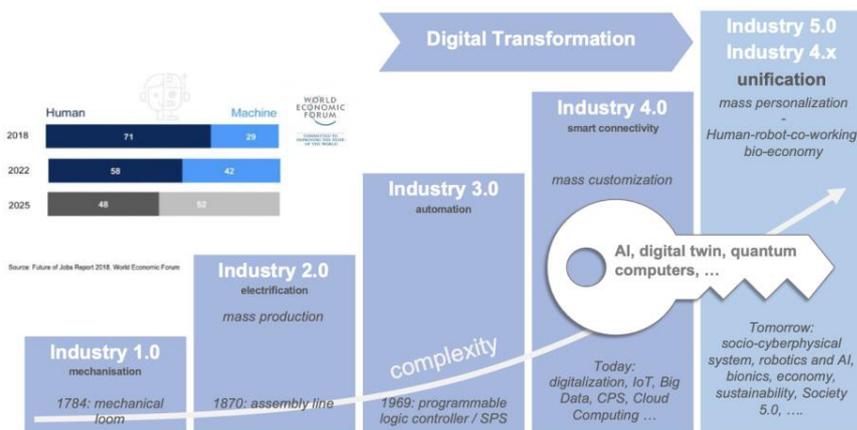


Fig. 1: The path to Industry 5.0 [3]

To prevent and successfully respond to disruptions in the Industry 5.0 scenario, organizations must proactively strengthen their resilience through techniques such as modular production systems, flexible manufacturing facilities and risk management procedures, including cybersecurity measures as well as social interactions to consider the workforce and society. The emphasis on resilience and sustainability is not just a buzzword in Industry 5.0, but a key design principle. The awareness of the essential role that people play in this technological environment is what characterizes Industry 5.0. A special synergy is created when people and machines work together. Humans are adaptable, can solve problems and make subtle decisions. This human-machine collaboration promotes sustainable operations by reducing the need for ongoing maintenance and ensuring consistent production. Because human workers can quickly adapt to changing circumstances and deal with unforeseen problems, Industry 5.0 places a strong emphasis on the human element as a means of developing resilience. In Industry 5.0, a holistic strategy that utilizes the capabilities of humans and robots proves essential to achieving sustainability and resilience [9].

This new paradigm shift favors less technology through the use of social content. It will increase productivity as part of a global concept known as Humanity-Centered Production (HCP).

## 2 DIFFERENTIATION OF TERMS

*How does the term „humanity-centered“ differ from „human-centered“? Don't the terms „human“ and „humanity“ have similar meanings?*

The meaning of these terms cannot be derived solely from the words themselves; the context must be taken into account. The first use of the term „human-centered“ appeared in the early 1980s and at that time focused primarily on the individual person, for whom the optimization of production processes was intended [8].

This approach has many merits and is still the prevailing paradigm today. Three decades later, however, we have developed a heightened sensitivity to bias against social groups and are increasingly concerned about the impact that humans have on the environment. The term „Humanity-Centered“ emphasizes the rights of all humanity and addresses the entire ecosystem (the term ecosystem encompasses all living things and the Earth's environment).

The HCP focuses on the question of how production systems can be tailored to people's needs, skills and

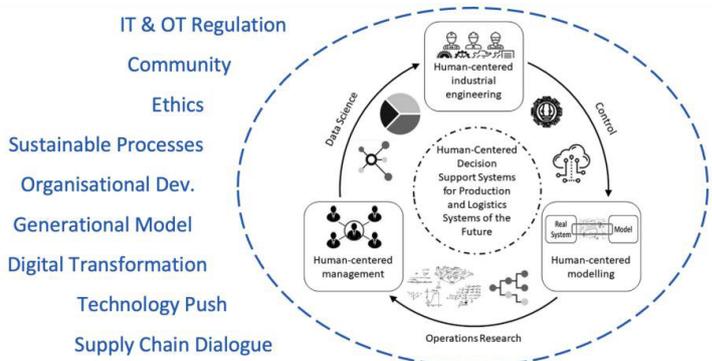
experiences. This also includes the question of what social participation could look like in new working environments. It is also about how people can be actively involved in production processes through new concepts and supported by technical and data-driven tools.

## 3 BASIC APPROACH

*„If we produce for humanity, we must not stop at the individual. We must consider our entire global environment in a sustainable way: all living beings, the quality of soil, water and air. The loss of species. The changes in the climate. We are an integral part of the ‚earth‘ system, where changes in one component can affect all other components.“*

The HCP embraces the fundamentals of current holistic, lean production strategies and technological capabilities, but expands them to explicitly consider all living things, the ecosystem and the long-term impacts in the future.

In the dynamically developing field of technology, people's tendency to accept technological advances is not solely motivated by the benefits they offer [7, 2]. Acceptance is often determined by social use and therefore implies looking beyond the limits of use. Technology must be used under the conditions of the general public without harming it.



**Fig. 2:** Humanity-centered production approach

Based on Norman [5], five principles for dealing with industrial processes are defined:

1. Tackle problems at the root, not just the current problem (which is often a symptom rather than a cause).
2. Focus on the entire ecosystem of humans, all living beings and the physical environment.
3. Take a long-term, systemic perspective, recognizing that most complications arise from the interdependencies of different parts and that many of the most damaging effects on society and the

eco-structure only become apparent years or even decades later.

4. Continuously evaluate the use of technologies and organizational practices to ensure that they truly meet the needs of the people for whom they are intended.
5. Plan together with the community and support the needs of the community as much as possible.

Taking these five principles into account, collaboration in the production environment and an understanding of the use of high technologies such as artificial intelligence will lead to a better culture [6]. With clear, culturally anchored application, this will also lead to greater motivation in the working environment, which in turn is expected to lead to higher productivity.

Implementing the HCP approach fundamentally addresses 5 of the 17 United Nations Sustainable Development Goals (SDGs):

- **Quality education (4)** - Through partnerships with educational institutions, the aim is to transfer knowledge and train the next generation of professionals, which contributes to the promotion of high-quality education.
- **Decent work and economic growth (8)** - By creating a working environment that promotes innovation and efficiency, the vision contributes to the creation of decent work and economic growth.
- **Industry, innovation and infrastructure (9)** - The vision aims to research and develop innovative technologies and production infrastructures to strengthen the industry and increase efficiency.
- **Responsible consumption and production (12)** - The vision aims to optimize production processes through the use of resource-efficient technologies and the promotion of ecological and social sustainability.
- **Partnerships for the goals (17)** - The cooperation with industrial companies and educational institutions shows that we are committed to working in partnership to jointly achieve the Sustainable Development Goals.

#### 4 IMPLICATION

Based on the principles and their relevance to sustainability goals, the transformation process in industry will make the automation of production processes increasingly urgent, particularly in light of demographic change. In addition, from an ecological perspective, manufacturing companies worldwide are increasingly dependent on the collection of energy data and its processing in efficient process chains in order to meet the challenges of future value creation networks. Digitalization is an important key to sustainability, resilience, flexibility and long-term prosperity. However, the digital penetration of many production areas is only taking place with a delay. The introduction of digital technologies depends not only on the technical and economic added value or their organizational implementation, but also on whether this added value is perceived and accepted.

Adoption of digital technologies and actual usage behavior is based on the perceived usefulness in combination with the perceived ease of use. Depending on value profiles and the purpose in the context of the social task, the productive factors must be adapted to the respective circumstances and constantly scrutinized.

Optimizing the interaction between man and machine is therefore more urgent than ever in order to advance the development of dynamic socio-cyber-physical production systems in harmony with the respective cultural framework conditions. Until now, digital production processes have only mapped the data-technical connection of the means of production to virtual representations through data models in the context of the duality of virtual (cyber) and physical production systems. In future, a socio-technical component will be added to these cyber-physical production systems in order to support the people using them. The integration of socio-cognitive aspects in automation processes must therefore be taken into account in the context of product development, factory planning, entrepreneurship and innovation management in order to shape the future together responsibly.

#### 5 CONCLUSION AND OUTLOOK

Over the past decade, advances in technology and in the organizational structures of production systems have led to profound changes. These changes have led to a human-centered approach that emphasizes the importance of human factors in the design and operation of production systems. Future production landscapes as well as the competitiveness of companies and societal frameworks will require continuous adaptation to dynamic changes.

It is crucial to develop a comprehensive understanding of the diverse impacts of our activities.

Humanity-Centered Production (HCP) provides a strategic framework in this context and offers guidelines that will have a significant impact on global practice. By integrating ergonomic, cognitive and social considerations into production processes, human-centered production aims to improve productivity, safety and overall well-being, thereby promoting sustainable and competitive production environments.

## REFERENCES

- [1] Alojaiman, B.: Technological Modernizations in the Industry 5.0 Era: A Descriptive Analysis and Future Research Directions. *Processes* 11(1318) (2023)
- [2] Abdalla, R.: Examining awareness, social influence, and perceived enjoyment in the TAM framework as determinants of ChatGPT. *Journal of Open Innovation: Technology, Market, and Complexity* 3(19) (2024)
- [3] Günther, N., Prell, B., Reiff-Stephan, J.: Industry 5.0 - From the vision of the human-centered approach to sociocyberphysical production systems for practice. In: *Proceedings of the 18th AALE (2022)*, Leipzig: Open Access, pp. 257-266 (2022)
- [4] Yitmen, I., Almusaed, A.: Industry 4.0 Transformation Towards Industry 5.0 Paradigm - Challenges, Opportunities and Practices. DOI: 10.5772/intechopen.1005583 (2024)
- [5] Norman, D. A.: *Design for a Better World: Meaningful, Sustainable, Humanity Centered*. MIT Press (2024)
- [6] Pedreschi, D., Pappalardo, L., Ferragina, E., Baeza-Yates, R., Barabasi, A-L., Dignum, F., Dignum, V., Eliassi-Rad, T., Giannotti, F., Kertesz, J., Knott, A., Ioannidis, Y., Lukowicz, P., Passarella, A., Pentland, A. S., Shawe-Taylor, J., Vespignani, A.: Human-AI coevolution. arXiv:2306.13723 [cs.AI] (2024)
- [7] Prell, B., Reinhardt, H., Mauersberger, A., Rauscher, A., Reiff-Stephan, J., Ihlenfeldt, S.: A survey on sustainability approaches in manufacturing. In: *Procedia CIRP*, 116 (2023)
- [8] Rosenbrock, H. H.: *Designing human-centered technology: a cross-disciplinary project in computer-aided manufacturing*. Springer, London (1989)
- [9] Hassan, M. A., Zardari, S., Farooq, M. U., Alansari, M. M., Nagro, S. A.: Systematic Analysis of Risks in Industry 5.0 Architecture. *Appl. Sci.* 14(1466) (2024)

# Optimized test case selection in a Continuous Integration process using Reinforcement Learning

Torben Dreyer<sup>1</sup>, Michael Thess<sup>2</sup>, Hans-Werner Wiesbrock<sup>1</sup>, Kai-Uwe Basener<sup>3</sup>, Jens Bielefeldt<sup>4</sup>

<sup>1</sup> ITPower Solutions GmbH, Kolonnenstrasse 26, 10829 Berlin, Germany  
{torben.dreyer,hans-werner.wiesbrock}@itpower.de

<sup>2</sup> IT Entwicklung, Judith-Auer-Str. 4A, 10369 Berlin, Germany  
thess@signal-cruncher.com

<sup>3</sup> KOSTAL Industrie Elektrik GmbH & Co. KG, Lange Eck 11, 58099 Hagen, Germany  
k.basener@kostal.com

<sup>4</sup> Compleo Charging Solutions GmbH & Co. KG, Ezzestraße 8, 44379 Dortmund, Germany  
jens.bielefeldt@compleo-cs.com

**Abstract.** Continuous Integration (CI) is an important concept in modern software development to maintain a high quality in projects with increasing complexity yet decreasing time for development cycles. It uses a continuous and automated test of the software on a daily basis. An important challenge in CI processes is an optimal selection of test cases with high potential to find faults in the software given the changes in the respective CI cycle. This selection is needed since typically not all test cases can be executed in each cycle because of limited time and hardware resources. Reinforcement Learning (RL) is a method in Artificial Intelligence (AI) where an agent learns by interaction with a system. It is well suited to cope with dynamic environments like a CI process.

This article describes an implementation of a tabular RL agent to perform an optimal selection of test cases in a CI process. A real CI process was analysed within the scope of these studies, resulting in a rule tree to improve process. The RL agent is tested in a dynamical simulation of a real CI process. The RL agent shows a better performance compared to a random selection. A model-dependent RL approach shows the best performance in a reasonable number of CI cycles as needed for a future application in a real CI process.

**Keywords:** Continuous Integration Process, Reinforcement Learning, Test Case Selection.

## 1 INTRODUCTION

Continuous Integration (CI) is a concept in software development processes that gained importance over the last years to maintain high quality standards for software in an environment with increasing complexity and decreasing time of software development cycles. The goal of a CI process is to find faults in the software early during the development to save additional costs and time that would be needed to fix these faults at a later stage. A build of the current state of the software is created and tested daily. With this build it is possible to run automated tests on all test stages such as component tests in software, integration tests and system tests on hardware (hardware in the loop), which need special test environments. This process allows finding potential faults that can only be observed in a complete build of the software because they rely on an interplay of the different components of the software.

High quality standards demand for a high number of test cases that need to be created and maintained. A main challenge of CI is to select a set of test cases with a high potential to find possible faults given the changes in the software in a specific development cycle. The necessity of a selection process is derived from the inherent limitations of time and hardware resources, which preclude the possibility of running all test cases in each cycle. The selection of test cases is a complex task that is usually done by experienced software testers.

This article describes the development of an AI agent using Reinforcement Learning to perform the task of test selection that is usually done by humans. Reinforcement Learning is a special concept of machine learning in which the AI agent learns by interaction with its environment. This makes RL well suited for dynamic and changing systems.

Previous applications of an RL agent in a CI process have been studied in references [1,2]. In the referenced studies, the RL agent is trained and tested with static historical data. The studies presented in this article aim to go one step further and to train and apply such an RL agent in a dynamic CI environment.

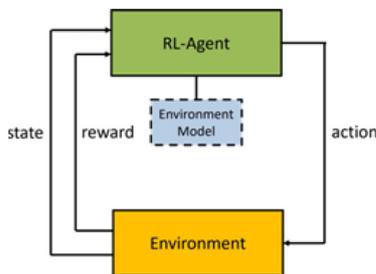
Within the scope of these studies, a real CI process at KOSTAL Industrie Elektrik GmbH & Co. KG and Compleo Charging Solutions GmbH & Co. KG has been examined in Section 3 resulting in a rule tree to improve the process.

The tabular agent used in this article is described in Section 4. It uses a model-dependent approach to get early estimates for states previously unknown to the agent and to obtain a good performance within a reasonable number of cycles. The agent is tested in Section 5 in a dynamic simulation of a real CI process. The average reward represented by the average number of failing test cases selected by the agent is compared to a random selection of testcases as a function of learning steps or simulated CI cycles.

## 2 BACKGROUND

### 2.1 REINFORCEMENT LEARNING

Reinforcement learning is an interdisciplinary area of machine learning and optimal control, concerned with how an intelligent agent should take actions in a dynamic environment in order to maximize a reward signal [4]. Central to this – as usual in AI – is the agent. The environment is typically stated in the form of a Markov decision process (MDP), as many reinforcement learning algorithms use Dynamic Programming (DP) techniques [5]. Fig. 2 shows the interaction between agent and environment in RL.



**Fig. 1:** Schematic representation of a model-based Reinforcement Learning process.

In each step the agent passes into a new state  $s$ , for which it receives a reward  $r$  from the environment, whereupon it decides on a new action  $a$  from the admissible action set  $A(s)$  for  $s$ , by which in most cases it learns, and the environment responds in turn to this

action. We differentiate between episodic tasks, which come to an end, and continuing tasks without any end state. The goal of the agent is to select the action in each state so as to maximize the sum of rewards over all future interactions. The selection of the actions by the agent is referred to as its policy  $\pi$ , and that policy which results in maximizing the sum of all rewards is referred to as the optimal policy  $\pi^*$ . In model-free RL algorithms, it is assumed that an agent neither has any prior knowledge of the environment, nor attempts to learn the environment dynamics. In model-based RL-algorithms [6], the agent uses a model of the environment, either known a priori or learned by the agent from the environment dynamics, see Fig. 2. Using this environment model, the agent learns through simulation of the environment (planning). The advantage of model-based RL over model-free RL is that it requires less real interaction steps to learn a stable policy. The disadvantage is the additional bias from the environment model, that can lead to systematic errors. Both, model-free and model-based RL approaches can be combined.

The state and actions spaces can consist of continuous or discrete attributes. In our work, we model the RL problem as continuing task over discrete state and action spaces using a model-based approach.

### 2.2 RELATED WORK

Over recent years, different RL approaches for test case prioritization of the CI process have been proposed. Spieker et al. prioritize test cases based on execution time and previous execution and failure history [1]. They use a pointwise ranking model and Q learning as RL algorithm. Bagherzadeh et al. extended the approach by including code-based features into the states, additional listwise and pairwise ranking, and different RL learning methods [2]. Bertolino et. al. included other machine learning methods and compared them with the RL approaches [3].

However, all these approaches are based on simulations of historic datasets. This means failure rates are independent of their previous executions, dynamically selected by the agent. In this work, we will address this shortcoming by introducing an environment model, where failure rates depend on previous test case executions.

## 3 IMPROVEMENTS FOR A CONTINUOUS INTEGRATION PROCESS

Within the scope of the studies presented in this paper, case studies of CI processes from the development projects of KOSTAL Industrie Elektrik GmbH & Co.

KG and Compleo Charging Solutions GmbH & Co. KG have been examined. The Open Charge Point Protocol (OCPP) is an open communication protocol developed for the interaction between charging stations for electric vehicles and a central management system. In the OCPP case study, the results of the specification tests of the OCPP client integrated in the CI system were examined.

The automatic scheduling in the CI system triggers a pipeline execution for all existing branches when changes are made in the source code management (SCM) system. The results were archived during the research project.

The archived results have been analyzed and it was found that the relation between failed tests and changes in the software is ambiguous, since the CI process was often executed over several commits. In order to obtain more data for the research project, a tool was developed in Python that simulates the CI process and executes the OCPP tests for each individual commit. The obtained data was analyzed manually, resulting in a rule tree shown in Fig. 3.

The rule tree uses a pattern consisting of a trigger source, trigger and trigger event. The trigger source acts as a trigger for the database and essentially includes changes in the source code and the content of the commit message. The trigger describes the condition that must be met at the trigger source. A successful trigger results in a trigger event, i.e. the follow-up action.

The data was first sorted according to the respective trigger sources. It was further checked which events (triggers) could have a direct influence on the tests (trigger event) and finally, possible rules for test selection were defined.

Based on the experience of the project participants, there were already general assumptions about rules that could be confirmed. These were also taken into account in the diagram. For example, it was expected that a code change in a test would lead to changed test conditions and thus this test would have to be considered in any case.

In addition, a lot of new insights could be gained from the commit messages.

The information contained, such as action words (e.g. removed, documented), branch names, requirements markers and buzz words (e.g. TMP, WIP, Todo), provides a crucial indication of possible effects (trigger events), which can be used to adjust the test selection.

It is also important not only to examine the content, but also to take the meta-information into account. If the commit is not written by the module owner, the probability of errors in the tests increases and a broader contextual view must be taken when selecting tests.

The buzzwords can be used to restrict the test selection, as this clearly describes the change. For example, it makes no sense to run the tests if the commit is labelled WIP (work in progress) and the change is therefore not yet complete.

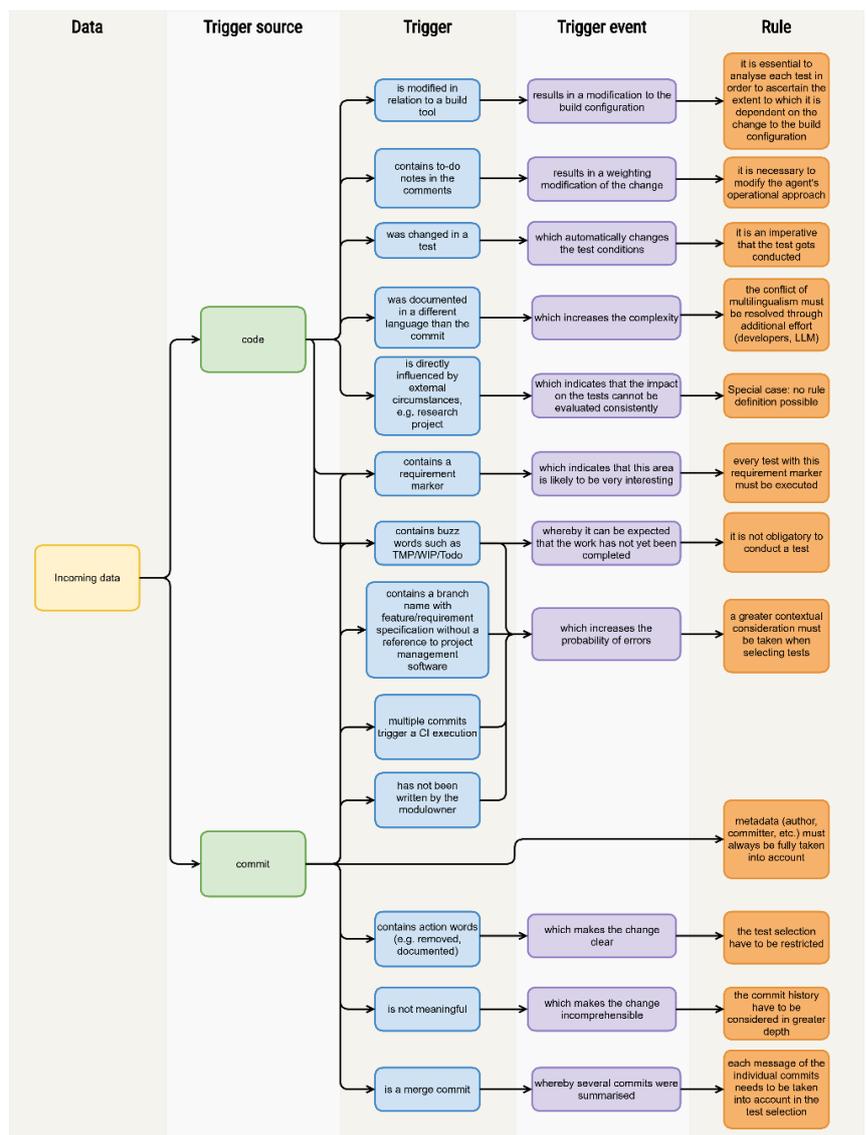


Fig. 3: Rule tree developed for the optimization of a specific CI process

#### 4 BUILDING AN RL-AGENT

There exists a large number of RL methods in order to determine the optimal policy  $\pi^*$ . They are mostly rooted in DP. In the current work, we focus on Q-learning. Like many RL methods it calculates an action-value function  $Q:S \times A \rightarrow \mathbb{R}$  which for each pair of state and actions assigns the expected sum of future rewards, the expected return.

The core of the Q-learning algorithm is a Bellman equation from DP as a simple value iteration update, using the weighted average of the current Q value and the quadruple  $(s_t, a_t, r_{t+1}, s_{t+1})$  from the new time step  $t$ . Based on the action-value function, a corresponding policy can be calculated.

##### 4.1 TABULAR AGENT

For discrete state and actions spaces of sufficiently low dimensions, the state-action function  $Q$  can be directly stored as a table. This is the simplest type of representations of  $Q$  and does not require function approximations techniques. The resulting agent is called tabular agent. We use Q-learning with a tabular agent.

##### 4.2 AGENTS ENVIRONMENT MODEL

To expedite the agent’s learning process, we employ a model-based approach. This entails informing the agent of the existence of relationships between processed features and test failures. As a result, the agent is able to construct a histogram of failed tests from a given run and previously processed features from the sprint. This enables the agent to utilize this data when it is required to act in previously unobserved states.

#### 5 EXPERIMENTS

Based on the experience gained from numerous projects, a simulation environment was developed with which the various hyperparameters of the agent, as well as an optimal reward and the action and state spaces, could be developed. This was done in order to facilitate the investigation of the aforementioned parameters.

##### 5.1 ENVIRONMENT SIMULATION

Based on careful studies of the behaviour of the CI process (Section 3), we designed an environment to simulate the failure rates of test cases in each new CI

run based on previous executions and properties of the test cases.

#### DATA MODEL

As shown in Fig. 4 in a development process, various features or functionalities are implemented and modified in development cycles, which are described here in terms of their relevance to safety and assigned a unique identifier in the *Functionality* table of the accompanying data model. Concurrently, tests are designed to ascertain the correct implementation of the various features within the system, see table *TestCase*. The relationship between the evaluated features and test cases is designated as *Test\_Func\_Rel*, wherein the weight represents the depth of the test case utilized for the assessment of the feature. Versions are recorded periodically (*SW\_Version* table) and the functionalities that have been modified since the last version are noted (the *Version\_Func\_Rel* table).

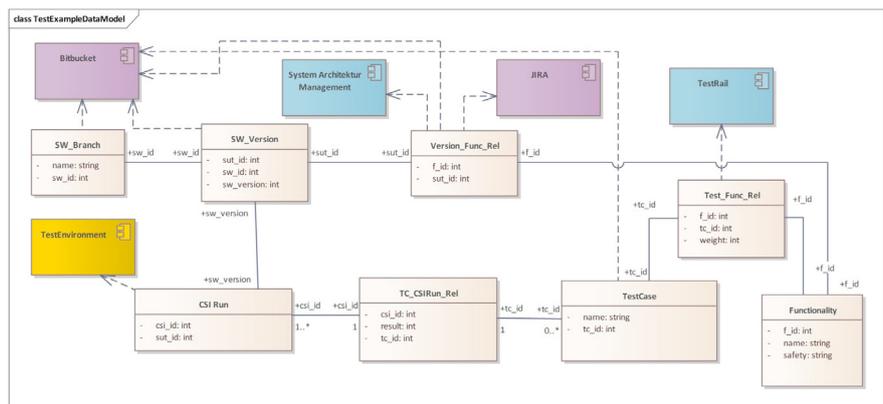


Fig. 4: Data model for the simulation of a CI Process

A crucial element of a CI process is that an automatic test run is always triggered after versioning (the *CSI Run* table). This executes a subset of test cases (specified in the table *TC\_CSIrun\_Rel*) and stores the test results (in column *result* of the aforementioned table).

#### SIMULATION

A typical CI process was modelled dynamically as follows. In each simulation step, a new software version is created, resulting in a new row in the table *SW\_Version*. Additionally, features that have been edited since the previous version are selected. This selection is project-specific and depends on factors such as the average time required for developing new functionalities or resolving bugs. The subsequent phase is the execution of a CI test run. To this end, the agent will select a number of test cases to be run on a test engine. The result of each test

run will be modelled stochastically. It is more likely that a test case associated with a feature that was processed in the previous sprint will fail. Similarly, if a test case failed in the previous CI test run and no bug fixing has taken place, it is more likely to fail again.

### STATE, ACTION, REWARD

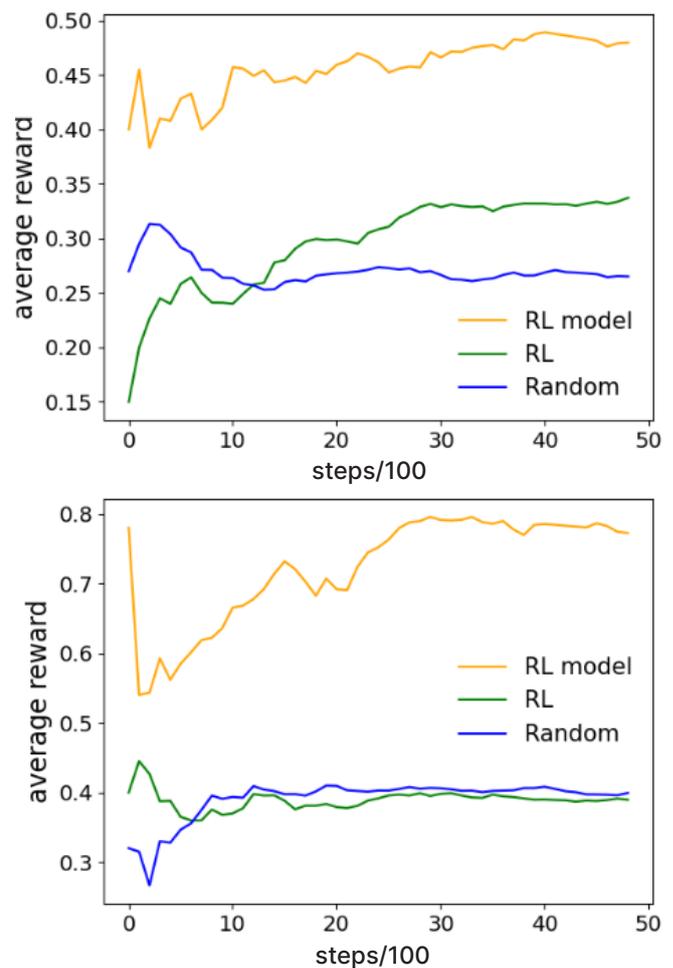
In order to select appropriate test cases for the CI run, the agent requires information regarding the system's current state. In this simulation, the failed test cases from the previous run and the features that were processed in the last sprint are utilized to define the state. The agent selects a subset of test cases to be executed in the next CI run, which defines the agent's action. As a reward, the number of previously selected test cases that have failed is received.

### 6 RESULTS

The RL agent is applied to the simulated environment described above. Figure 5 shows the average reward for the dynamic test case selection using the RL agent with the environment model in yellow compared to a RL agent without the environment model in green and a random selection of test cases in blue. The training of the agents runs through 5000 steps. The reward represents the number of failing test cases found by the respective agent or by the random selection. A higher reward is therefore equal to a better performance of the agent.

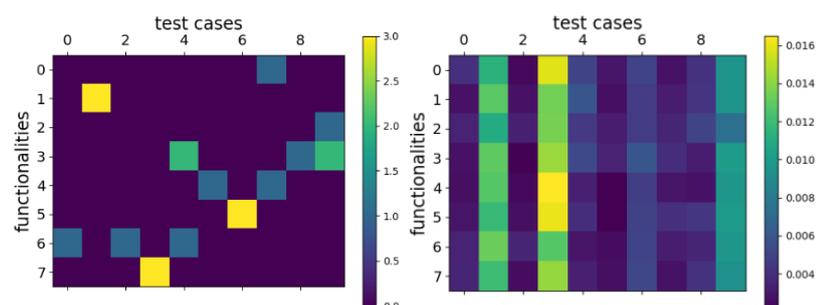
The left figure shows the average reward in a simulation with 10 test cases and 8 functionalities. Up to three functionalities can be modified in each cycle. A better performance is observed for both agents compared to the random selection after about 1000 steps. We observe a better performance for the agent with the environment model. Especially in the first few hundred steps the model provides an estimate for previously unknown steps leading to a better performance in the early cycles. This is even more important in a more complex simulation with 25 test cases for 20 functionalities shown in the right figure. The average reward does not improve significantly in the first 5000 cycles using the agent without the environment model. Using the model-dependent approach however, leads to a better performance in a reasonable number of cycles.

The plots shown in Fig. 5 are just one example. The statistical nature of the problem leads to differences between different runs, especially for the model-dependent approach in the first cycles. The model in the agent is randomly initialized. A better performance of the model-dependent approach at the end of the run was observed in multiple studied examples.



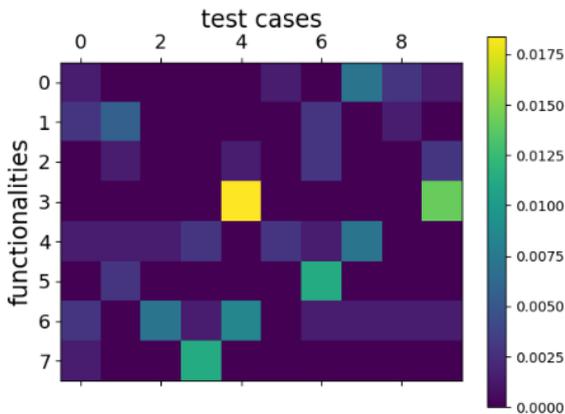
**Fig. 5:** Average reward as a function of the training steps for an RL agent with environment model (yellow), an RL agent without model (green) and a random test case selection (blue) in simulated CI processes with 10 test cases and 8 functionalities (left) and with 25 test cases and 20 functionalities (right).

The better performance with the environment model is important for a possible future application in a real CI process because in a real application you can usually not afford to be ineffective for a few hundred cycles. In a process with daily cycles, this would mean ineffectiveness for the first years. The environment model in the agent could further be initialized with information from a test management tool or from historical data to get an even better and more stable performance in the early cycles.



**Fig. 6:** Test case to functionality relation used in the simulation (left) and learned by the environment model in the agent (right).

Fig. 6 shows the relation between test cases and functionalities used in the simulation of the CI process on the left. Test cases are related to functionalities by weights ranging from 0 to 3. The higher the weighting, the deeper the functionality is tested by the test case and the higher the probability that it will fail, if the functionality is changed. The figure on the right shows the test case to functionality relation in the environment model of the agent. This relation is constantly updated. The reason for the stripe structure in this figure is that several functionalities can be modified in the same cycle, resulting in ambiguities relating the functionalities to the test cases. Therefore, stripes are visible for the dominant failing test cases. This can be avoided by updating the model only in states where one functionality is modified and no testcases have failed in the previous run. The resulting model is shown in Fig. 7. This, however, leads to a slower learning process because less events are selected to update the model. It could be beneficial to use this variant in case of an initialization by historical data, but this would be part of future work.



**Fig. 7:** Test case to functionality relation learned by the environment model in the agent. The model is only updated in cycles with only one modified functionality and no failing test cases.

## 7 CONCLUSION

A successful application of an RL agent for a test case selection in a simulated CI process has been shown. The RL agent was studied with and without an environment model used to obtain an estimate of the optimal selection in states previously unknown to the agent. Its performance is compared to a random test case selection. A better performance is observed for both RL agents compared to the random selection in a simulated process. The model-dependent approach shows a much better performance, especially in the first cycles of the CI process. This advantage in the performance grows with

increasing complexity of the process (more test cases and functionalities). The model-dependent approach is interesting for an application in a real CI process, where it is important to reach a reasonable performance in a limited number of cycles. It also allows to include prior knowledge into the agent by a respective initialization of the environment model. Additionally, improvements of an existing CI process have been studied resulting in a rule tree in Fig. 3. In a next step, the tabular action-value function should be replaced by function approximation.

**Acknowledgments.** This study was funded by the Bundesministeriums für Bildung und Forschung under the grant number 01IS22035A.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## REFERENCES

- [1] H. Spieker, A. Gotlieb, D. Marijan und M. Morten, "Reinforcement learning for automatic test case prioritization and selection in continuous integration", Proceedings of 26th International Symposium on Software Testing and Analysis (ISSTA'17), p. 12–22, 07 2017.
- [2] M. Bagherzadeh, N. Kahani und L. Briand, "Reinforcement Learning for Test Case Prioritization", IEEE Transactions on Software Engineering, Bd. PP, 04 2021.
- [3] A. Bertolino, A. Guerriero, B. Miranda, R. Pietrantuono, and S. Russo, "Learning-to-rank vs ranking-to-learn: Strategies for regression testing in continuous integration", in 42nd International Conference on Software Engineering (ICSE), 2020.
- [4] R. S. Sutton and A. G. Barto. "Reinforcement Learning: An Introduction", Second Edition (see here for the first edition) MIT Press, Cambridge, MA, 2018.
- [5] D.P. Bertsekas. "Dynamic Programming and Optimal Control", Athena Sci., 1995.
- [6] T.M. Moerland, J. Broekens, A. Plaat, C.M. Jonker, "Model-based Reinforcement Learning: A Survey", Foundations and Trends in Machine Learning, 16(1), 2023.

# Low-Cost Smart Sensor IIoT Platform for Real-Time Monitoring and Predictive Maintenance

Jens Kneifel, Hala Alfaris, Robin Roj, Ralf Theiß and Peter Dültgen

Forschungsgemeinschaft Werkzeuge und Werkstoffe e. V., Papenberger Str. 49, 42859 Remscheid, Germany  
roj@fgw.de

**Abstract.** The industrial internet of things offers a transformative opportunity to revolutionize manufacturing through real-time monitoring and predictive maintenance. However, the high costs of traditional measurement equipment have limited its adoption, especially in resource-constrained environments. This paper presents a smart sensor system based on the ESP32 microcontroller, designed to make industrial internet of things applications accessible to small and medium-sized enterprises. The system includes multiple sensor nodes deployed across various machines, transmitting data to a hub gateway device powered by a Raspberry Pi. To enhance functionality, specific libraries for the low-cost sensors were developed and power-efficient features were implemented, including a sleep mode to extend battery life. The collected data is analyzed using artificial intelligence models for regression, clustering, classification, and forecasting, providing valuable insights into machine behavior. This enables predictive maintenance, quality monitoring, wear detection, and total tool life analysis, ultimately leading to improved productivity, reduced downtime, and increased overall efficiency. By leveraging low-cost sensors and optimizing their performance, the proposed system offers a cost-effective and efficient solution for Industry 4.0 applications.

**Keywords:** Artificial Intelligence, Industrial Internet of Things, Industry 4.0, Predictive Maintenance, Sensor Nodes.

## 1 SCIENTIFIC BACKGROUND AND MOTIVATION

The high costs of traditional measurement equipment has been a major barrier to the widespread adoption of advanced sensors with software agents in the field of industrial internet of things (IIoT), particularly for small and medium-sized enterprises (SME) in resource-constrained environments. Previous studies, such as the work by Rossi et al. have demonstrated the potential of low-cost micro-electromechanical systems (MEMS) accelerometers for vibration monitoring in rotating structures, providing a reliable and cost-effective alternative to traditional optical systems [1].

Building on this foundation, this paper proposes a smart sensor system that leverages low-cost sensors and advanced data analytics to provide an affordable yet effective solution for IIoT applications. The system is based on the ESP32 microcontroller, known for its versatility and cost-effectiveness. By deploying multiple sensor nodes across various machines and transmitting data to a central hub powered by a Raspberry Pi, a comprehensive monitoring network was created. The integration of artificial intelligence (AI) models for data analysis further enhances the system's capabilities. This approach not only reduces reliance on expensive equipment but also leverages machine learning (ML) and deep learning algorithms to enhance data analysis.

This research extends the application of MEMS accelerometers to new domains, specifically adapting the experimental setup to a rotating fan and exploring their potential for classification and error detection in drilling machines. This advancement confirms the versatility of MEMS accelerometers and opens new possibilities for their use in industrial applications.

## 2 MATERIALS AND METHODS

### 2.1 SENSOR SELECTION PROCESS

The smart sensor system relies on a combination of affordable open-source sensors, a microcontroller, a Raspberry Pi-based AI hub and AI models to enable monitoring, realtime data collection, processing, and analysis. The hardware and network configuration play a crucial role in ensuring seamless communication between the various components and efficient data transmission. Table 1 shows a comparison of commercially available sensor boards.



**Fig. 1:** A Test stand with two fans operating at the same frequency and two sensors: a MEMS sensor (a) and an IEPE sensor (b) mounted on the fan casing with a magnet. An intentional imbalance was introduced on the left side.

	ADXL337	ADXL326	LIS3DH	ADXL345	ISM330 DHCX	ADXL335
Manufacturer	Sparkfun	Adafruit	Adafruit	Sparkfun	Adafruit	Sparkfun
Max. data rate in Hz	500	550	5,300	3,200	6,660	550
Measuring range in m/s <sup>2</sup>	± 29.42	± 156.91	± 19.61 to ± 156.91	± 19.61 to ± 156.91	± 19.61 to ± 156.91	± 29.42
Sensitivity (LSB = least significant bit)	300 V/g	57 V/g	1-12 mg/LSB	4 mg/LSB	0.06 - 0.47 mg/LSB	300 V/g
Output	analog	analog	digital	digital	digital	analog
Approx. costs in €	17.30	6.90	5.90	24.40	23.75	15.50

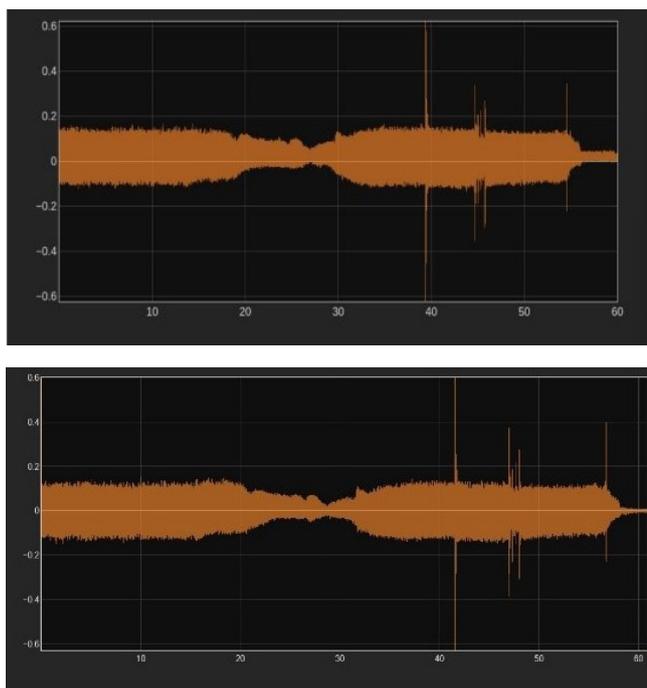
**Tab. 1:** Comparison of six MEMS-based sensor boards.

The data collection process in the smart sensor system is designed to balance accuracy and power efficiency. Low-budget sensors are configured for real-time analysis by collecting data at specific intervals over a predetermined duration. To minimize power consumption, the sensor is set to low-power configurations between data collection intervals. The objective of the sensor selection process was to identify a cost-effective accelerometer sensor that meets specific performance criteria. The key metrics used in the selection process included precision, data rate, sensitivity, reproducibility, and overall usability.

To ensure a standardized comparison, a test setup containing two fans was set up (cf. Fig. 1). An intentional imbalance was introduced allowing the sensors to capture rotational frequencies. Both digital and analog MEMS sensors mounted on the fan casing were evaluated in this setup, providing a comprehensive basis for comparing their performance under identical operating conditions.

Measurements were taken over a 10 s period for each test, with simultaneous data collection from both the integrated electronics piezo-electric sensor (IEPE), measuring the fan's casing, and the MEMS sensor, which was linked to a Raspberry Pi. The MEMS accelerometer sensor was connected to the I<sup>2</sup>C pins of an ESP32 microcontroller for data acquisition, a common protocol supported by all the sensors in comparison. To ensure consistency, only the corresponding z-axis data from the MEMS sensor was compared with the IEPE sensor. The measurement range was set to ±19.61 m/s<sup>2</sup> where possible and the data output rate was targeted at approximately 400 Hz which is supported by all the sensors to be analyzed, though actual rates varied.

The ISM330DHCX sensor was selected based on its performance in comparison to the sensors in Table 1, especially for its higher output data rates (ODR), making it an effective alternative to IEPE sensors operating at a sample rate up to 6.7 kHz. A comparison was made between the selected ISM330DHCX sensor and the industrial piezo vibration sensor 601A02 (sensitivity 554 mV/g) from PCB connected to a Raspberry Pi with an MCC172 IEPE measurement hardware attached on top (HAT) to ensure it as a viable substitution (cf. Fig. 2).



**Fig. 2:** The signal plots from the ISM330DHCX MEMS sensor (above) and the IEPE sensor (below) highlight comparable performance in detecting fan anomalies. Although the IEPE sensor provided a slightly cleaner signal, the ISM330DHCX was able to capture similar vibration patterns with sufficient accuracy. Given the lower costs and ease of integration of the ISM330DHCX, it was selected as the preferred sensor for the validation experiment, as it offers a cost-effective alternative.

## 2.2 SENSOR CONFIGURATION

After selecting the ISM330DHCX sensor, it was configured to use SPI communication instead of I<sup>2</sup>C. This change was necessary to accommodate the higher ODR supported by the sensor and ensure faster data transfer. The I<sup>2</sup>C protocol was only used in the sensor selection process with an ODR of 400 Hz to make the sensors comparable. To standardize the format of the sensor data for transmission, a JSON string template was created. This template includes key parameters such as metadata and the corresponding sensor data for acceleration and gyroscope readings across the x, y, and z axes. The ISM330DHCX sensor allows for a wide range of ODR from 1.6 Hz to 6,667 Hz. For the validation experiment, the sensor was configured to have a sampling rate of 6,667 Hz to capture fast changing and time sensitive events. A higher ODR contributes significantly to improving overall sensing accuracy and system responsiveness, as well as being useful for calculating frequency domain features with limited timeframes for measurement. The ISM330DHCX sensor has a dynamic, user-selectable full-scale acceleration range allowing flexibility depending on the application's needs. The choice of range directly affects the sensor's sensitivity. For the sensor configuration, a range of

$\pm 78.48 \text{ m/s}^2$  was chosen. Smaller ranges allow a more precise detection of subtle accelerations. The signal-to-noise ratio (SNR) is influenced by this range and the system's noise bandwidth, with lower noise enabling the detection of smaller acceleration values. Therefore, careful selection of the range was crucial for balancing sensitivity and noise performance.

## 2.3 SENSOR OPERATION MODES

The ISM330DHCX sensor operates in two modes: continuous mode and interval sleep mode. In continuous mode, the sensor collects data while connected to a power source, primarily for AI training or continuous monitoring. The first in - first out buffer (FIFO) is employed in both continuous and sleep modes to ensure consistent data buffering and an equivalent time interval between the samples. For continuous mode, the FIFO is configured to provide continuous updates, where older data is discarded as new data arrives. The data sample collection loop runs while the number of samples is less than the data buffer size.

In sleep mode, the ISM330DHCX sensor is set to the lowest possible ODR to minimize power consumption while the controller is in a deep sleep state. This configuration allows for efficient energy usage, extending the system's operational lifespan when running on battery. The controller wakes up periodically to collect and transmit data, then returns to sleep, enabling the sensor to possibly operate for a few months in low-power mode. For the FIFO buffer, FIFO mode is utilized because the measurement duration is small enough to fit within one or two full FIFO buffers, enhancing efficiency. The FIFO can store up to 9 KB of data with compression enabled, and its depth can be resized by setting the WTM bits in FIFO\_CTRL1 and FIFO\_CTRL2. The buffer is cleared between readings and no FIFO related triggers are used in this application. The loop for data sample collection continuously polls the FIFO watermark flag to determine if sufficient samples are available for processing. When the flag is set, it retrieves the number of samples and enters a loop to read data from the FIFO buffer, collecting different sample types based on their respective tag. The valid data is then buffered as a JSON string for transmission to the AI hub. Finally, the system prepares to enter sleep mode, efficiently managing power consumption. This logical flow enables structured data handling and energy-efficient operation.

In the validation experiment, acceleration and gyroscope data are collected to test the buffering capabilities for multiple sample types. Different parameters, such as

temperature, could also be buffered in the FIFO. The ISM330DHCX can act as a sensor hub for additional sensors, like a power sensor. An example code for configuring an external sensor to the hub, specifically the LIS2MDL external magnetometer sensor, is referenced in the ISM330IS datasheet which also provides an example for the sensor hub within the sensor board [2].

## 2.4 MICROCONTROLLER

The ESP32 serves as the central device for data processing and communication between the sensors and the AI hub. It buffers the sensor data, adds relevant metadata, and manages data communication with the server. To optimize power consumption, a dynamic measurement strategy was implemented that adjusts the required measurement time and interval. The microcontroller was then configured to operate in a low-power mode between measurements, allowing it to run on a battery for extended periods without replacement. Additionally, the FIFO buffer was utilized within the sensor to ensure consistent data collection rates and synchronize the various parameters measured (gyro, acceleration, temperature), thereby increasing the overall quality of the collected data. For the connection, the `setupMQTT` function is responsible for initializing and configuring the MQTT client on the ESP32. It sets up timers for reconnecting to the MQTT broker and Wi-Fi and defines callback functions for various MQTT events. The ESP32 establishes a Wi-Fi connection with a designated router for connecting the smart sensors with the Raspberry Pi. Next for buffering, the sensor setup is initialized with the configurations specified in the sensor section. A crucial part of this initialization is allocating space on the external PSRAM of the ESP32 to buffer the samples. This buffering mechanism is essential for ensuring smooth data acquisition and processing. By utilizing the external PSRAM, the ESP32 can temporarily store a larger number of samples, allowing for more efficient data handling and transmission. This approach helps prevent data loss and enables the microcontroller to manage high-frequency sampling rates effectively.

In the main loop, the ESP32 continuously checks the number of samples in the FIFO. If the number of samples exceeds the threshold, it reads the data from the ISM330DHCX sensor and publishes the data to the MQTT topic. The ESP32 enters deep sleep mode

after transmitting the data. During this time, power consumption is significantly reduced, allowing the device to run on battery power for extended periods. The power consumption of the ESP32 in deep-sleep and active modes is critical. In deep-sleep mode, the ESP32 consumes approximately 10  $\mu$ A, while Wi-Fi transmission in active mode draws 240 mA. To evaluate the impact of different active durations on battery life, a scenario is considered where the ESP32 sleeps for one hour and then transmits data over Wi-Fi for varying durations. Table 2 shows the average current consumption and estimated battery life using a 1,000 mAh LiPo battery. As active time increases, overall power consumption rises significantly, reducing battery longevity.

Sleep time (per cycle)	Active time (per cycle)	ESP32 average current consumption (in mA)	ISM330DHCX average current consumption (in mA)	Total average current consumption (in mA)	Estimated battery life (1000 mAh LiPo)
1 h	1 s	0.077	0.006	0.083	~12,112.34 h (~504.68 days)
1 h	10 s	0.675	0.010	0.684	~1,461.07 h (~60.88 days)
1 h	1 min	3.944	0.03	3.974	~251.62 h (~10.48 days)
10 min	1 s	0.409	0.008	0.417	~2,396.33 h (~99.85 days)
10 min	10 s	3.944	0.03	3.974	~251.62 h (~10.48 days)
10 min	1 min	21.827	0.141	21.968	~45.52 h (~1.90 days)

**Tab. 2:** Estimated battery life with different settings of measurement intervals.

When the wake-up timer triggers, the ESP32 wakes up, performs the data collection and transmission tasks, and then returns to deep sleep mode. This cycle repeats. The configuration of the ESP32 was challenging due to the need to extend library functionalities. Functions that access all the FIFO registers were necessary to get the status of the FIFO and check registers. To dynamically allocate memory on the PSRAM, a debug file was included. This involved significant efforts in extending the library functionalities to support the specific requirements of the project. This setup ensures that the ESP32 can handle high-frequency data sampling and efficient data transmission using MQTT. The `board_build.psrाम` option in the Platform IO project configuration enables the use of external PSRAM, which is crucial for buffering large amounts of data. The `build_flags` include necessary configurations to fix known issues with the ESP32 PSRAM cache and to enable dynamic memory allocation using PSRAM.

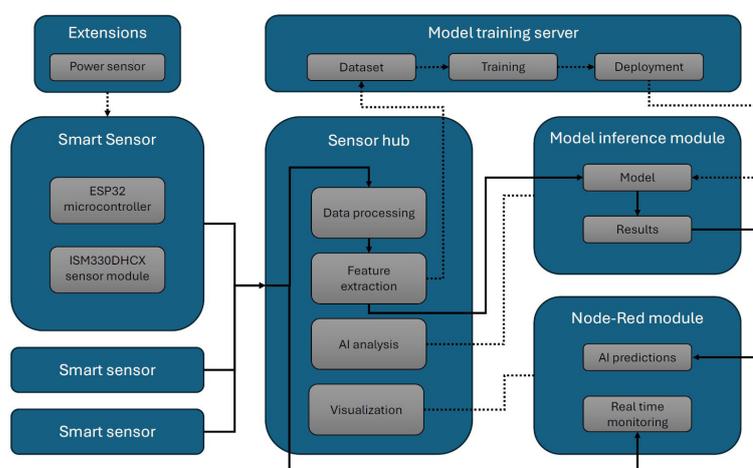
## 2.5 AI HUB

The sensor hub, built with a Raspberry Pi, collects data via Wi-Fi from multiple sources and processes it. To improve data handling, an asynchronous MQTT server manages client connections, allowing multiple communications at once. Each smart sensor sends its data (3-axis acceleration, 3-axis gyro data, optional temperature readings) to the hub. The hub then formats this data for analysis.

The server script assigns the data to the correct sensor topic and processes it further. The MQTT protocol facilitates data exchange between the sensors and the hub. The data pipeline extracts features and analyzes the data using AI models, acting as an inference server that turns sensor signals into AI predictions.

A diagram of the data flow is shown in Fig. 3. The server's backend runs on Python. For feature extraction, the TSFEL library and SciPy are used to efficiently extract relevant features from the data. For ML, PyCaret is used for initial data analysis, and XGBoost is the standard classification model.

The AI model's classifications are published as MQTT messages on a dedicated topic and displayed on a user interface created with NodeRed. This open-source software is popular in the Internet of Things (IoT) field and allows programming of data flow and linking signals with display or control elements. The interface is accessible via a web browser on any device within the same network, including mobile devices.



**Fig. 3:** The data flow of the smart sensor system, showcasing how raw sensor data is processed, analyzed, and visualized in real-time.

During initial testing in live detection mode, it was observed that, from data acquisition to feature extraction, processing in an AI model, transmission, and display on the graphical user interface, it takes on average less than 1 s processing time. This enables live monitoring, but the main application focuses on long-term monitoring, which is why the sleep function with battery operation is considered a crucial component of the smart sensor setup. The server can be operated in two modes: initially, a pure data acquisition mode is available, where incoming data is only collected and assigned to recognized classes. This data acquisition part serves to create a dataset with which the AI model will be trained. If an AI model is present, the inference operation mode is switched on, and incoming data is immediately further processed in the data pipeline, and the result of the classification is displayed. For the first version of the server, each sensor is treated as its own data collector and incoming data is individually processed and evaluated. Preparations have been made on the server side to enable data combinations, such as sensor fusion, where data from two sensors are combined to process as if it were a single sensor with 12 axes, or as axis extension, where the axes of multiple sensors located at different positions on the subject are considered together, allowing, e.g., three sensors to extract one axis each and combine it into a 3-axis dataset.

## 2.6 AI

To develop an effective predictive maintenance system, a standard model is provided with the XGBoost algorithm and additionally an automated ML framework (AutoML), that gives access to a multitude of industry standard algorithms, has been used. The AutoML mode trains multiple AI models and selects the most suitable based on performance-accuracy trade-offs. The level of accuracy required depends on the specific application and the quantity and quality of the collected data. An automated training process is used to compare various ML algorithms and select the best-fitted one. The model trains in the frameworks Pycaret for ML and Pytorch for deep learning. This approach enables the optimization of the model accuracy while minimizing required computational complexity after deployment. The automated ML process is intended for more complex classification or regression problems where the provided standard model does not achieve sufficient accuracy. The collected dataset is sent to a model-training server, which can be constantly optimized to provide updates to the automated model training process without the need to run updates on the edge devices. The standard model running the XGBoost algorithm can be trained on the edge device,

the AI hub, itself and therefore provides a complete offline alternative. The standard model has been shown to be sufficient in most standard maintenance tasks such as fault and machine state detection and estimating the remaining useful life (RUL) of tools. In the standard configuration, 19 features from the vibration analysis range are calculated for each axis, providing a total of 114 features for each collected sample. Features are extracted from both the time and frequency domains to capture the underlying characteristics of the data input. Time-domain features include statistical metrics such as the minimum, maximum, mean, variance, standard deviation, and more specialized features like root mean square (RMS), peak-to-peak amplitude (P2P), and crest factor. These features provide insights into the signal's amplitude, variability, and distributional shape, incorporating descriptors like skewness and kurtosis to quantify asymmetry and the tail behavior of the data. Additionally, the frequency domain features are extracted using the fast Fourier transform (FFT), yielding metrics such as the maximum frequency component, sum of spectral components, spectral mean, variance, and their respective skewness and kurtosis. This method, as demonstrated and validated by Pedotti, Zago, and Fruett [3], was employed in the analysis. These features collectively enable the model to leverage both time-dependent and frequency-dependent patterns, enhancing its ability to detect and classify complex behaviors in the data.

## 2.7 VISUALIZATION

Node-RED is a key component of the monitoring and visualization stack, offering realtime insights and alerts. It displays crucial information extracted from processed data, such as which machines are currently operating and their status, enabling operators to address potential issues proactively. The user interface also shows prediction results, including anomaly alerts and wear-out predictions, with customizable threshold levels tailored to the specific needs of each industrial application. To optimize system efficiency, Node-RED is leveraged for data orchestration or offloading tasks like data storage and visualization. This frees the sensor hub's backend to concentrate on critical tasks such as client management. Data is logged and stored offline for AI model tuning, supporting the continuous improvement of the predictive maintenance system.

## 3 VALIDATION EXPERIMENT

### 3.1 SETUP DESCRIPTION

The focus of the validation experiment is to prove the smart sensor system's capabilities as a predictive maintenance solution with multi state classification and rapid prototyping capabilities. The subject is an industrial drill, specifically realizing wear detection for the tools used (cf. Fig. 4). The experiment is designed with a single sensor setup to validate the system in its minimalist form, so 3-axis of acceleration and 3-axis of gyro data were collected. To test its fast-learning potential, a relatively small dataset was purposely collected and a small frame of data for each classification was used to ensure fast response times in live mode.



**Fig. 4:** Industrial drill setup with a smart sensor mounted on the clamping jaw using a neodymium magnet (left) and (a) a new drill and (b) a worn-out, defective drill with visible tip abrasions and cracks (right).

The ISM330DHCX sensor was securely mounted on the drilling machine's stationary table using a magnetic attachment. Two 8 mm drills were employed on a block made of aluminum: a standard drill and a defective, worn-out drill. The Raspberry Pi and router were positioned nearby for efficient data transfer. Data was collected in four distinct states: machine inactive, idle, drilling with a drill in good condition, and drilling with a worn-out drill. A total of 6,667 data points were gathered per class, each stored in a separate directory in CSV format for subsequent model training.

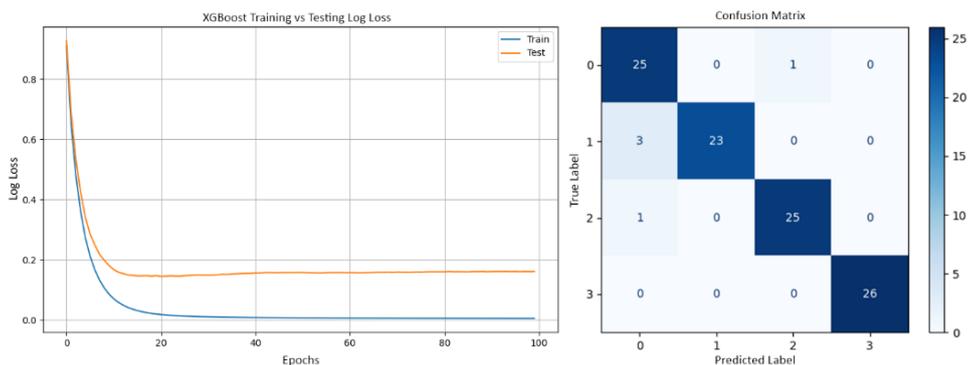
### 3.2 FEATURE EXTRACTION

With the collected dataset containing four distinct classes, 19 features were extracted on each axis, so that each data point stemming from 50 samples of raw data contributed a total of 119 features. Because of the split of 50 samples for a row of data, a total of 519 states could be generated in the dataset for model training. Each class will be represented by about 129 rows of data. The size of 50 samples per row was purposely designed to ensure quick response time and test the system with small batches of data and therefore its rapid prototyping capabilities.

### 3.3 MODEL TRAINING

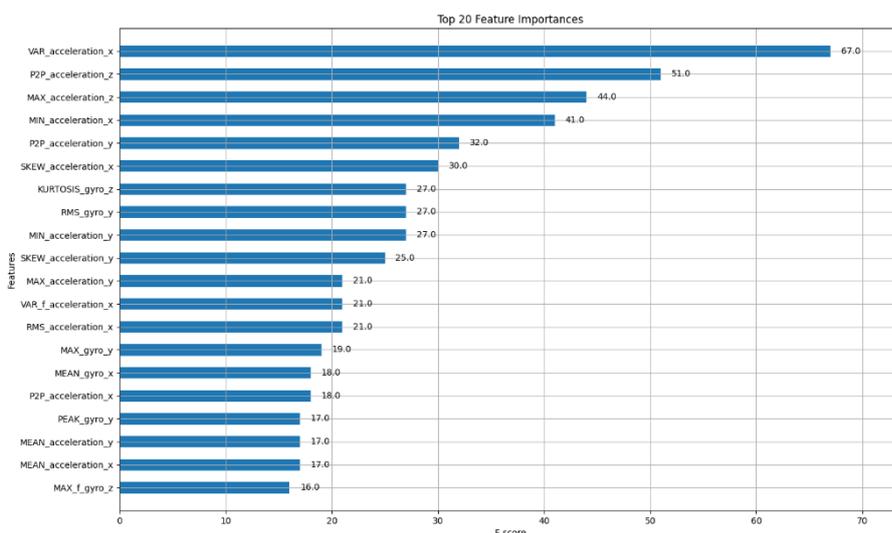
The model training process employed the standard model using the XGBoost algorithm. The dataset was fed into the model over 100 epochs, with the algorithm optimizing the decision boundaries with each pass. After training, the model achieved an accuracy of 95.19%, demonstrating its effectiveness in learning from the data. To better understand the model's behavior and performance, several key visualizations were generated: a training curve, a confusion matrix, and a feature importance plot. The training curve plots the model's accuracy over the 100 epochs, showing a rapid improvement in performance during the early epochs. Specifically, the curve rises sharply at the beginning but starts to flatten around 30 epochs, suggesting that the model quickly converges and reaches its optimal learning state early on (cf. Fig. 5). This indicates that further training beyond this point yields diminishing returns, though the model continues to refine itself slightly across subsequent epochs.

The confusion matrix provides insight into the classification performance across different classes. It reveals that the majority of predictions made by the model are correct, as shown by the high number of true positives and true negatives. Misclassifications are relatively rare, indicating that the model has learned to distinguish well between the various classes.



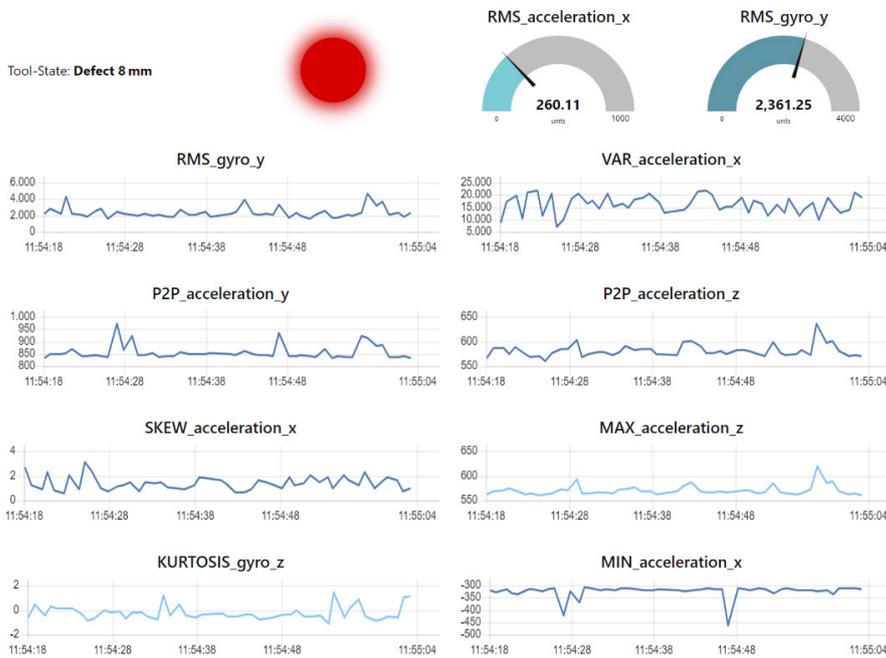
**Fig. 5:** Training and testing log loss curves for the XGBoost model, illustrating convergence and performance stability over 100 epochs (left). The training log loss decreases consistently, while the testing log loss stabilizes, indicating effective generalization. Confusion matrix representing the model's predictions, achieving an overall accuracy of 95.19% (right). The matrix highlights true positives, true negatives, false positives, and false negatives, demonstrating the model's classification effectiveness across different classes.

The feature importance plot further sheds light on the inner workings of the XGBoost model by ranking the features based on their contribution to the final predictions (cf. Fig. 6). The plot reveals a smooth and gradual decrease in feature importance, indicating that many of the features in the dataset play a meaningful role in the model's decision-making process, contributing to its robustness and generalizability.



**Fig. 6:** Bar plot displaying the top 20 features ranked by their importance in the XGBoost model. Features are sorted by their contribution to the model's predictive power, with higher values indicating greater importance in influencing the target variable. This visualization helps identify key predictors driving the model's decisions.

Following the initial data collection phase, the server can be manually switched to live mode. In this mode, smaller batches of 50 data points are collected and processed in near real-time. Extracted features from these batches are used to generate predictions using the trained XGBoost ML model. The reduced data size enables faster prediction times without compromising accuracy. To facilitate visualization, the server sends the extracted features and corresponding predictions via MQTT, each on its dedicated topic, ensuring synchronized and accurate representation of the drill's status. Node-RED then receives these messages and processes them for visualization purposes (cf. Fig. 7).



**Fig. 7:** The live monitoring user interface built with Node-RED, displaying the top eight features and the model's classification results. A red-light indicator shows that a defective drill has been detected.

The system updates the user interface every second, demonstrating the model's ability to detect changes in the drill's condition quickly and reliably. To enhance explainability, the top eight features were plotted in real-time, allowing users to track variations over time as drills with different conditions are used. The interface can be displayed both on a local screen and accessed remotely via a mobile device's web browser, indicating that the system's processing power is sufficient for real-time monitoring.

#### 4 RESULTS

The ISM330DHCX sensor-based system demonstrated its ability to monitor milling machine performance by collecting data across four distinct machine states: inactive, idle, drilling with a standard drill, and drilling

with a worn-out drill, making it an effective alternative or addition to traditional data acquisition methods. By capturing and analyzing this data, the system effectively detected anomalies related to machine operation and tool wear. This early identification of states and defects helps avoiding costly downtimes and improves product quality.

Additionally, the sensor's use in detecting wear on machine components presents a cost-effective solution compared to more expensive vibration analysis tools. The system supports two modes of operation: one for more extensive AI tasks that require advanced anomaly detection and tool life estimation via a dedicated training server, and a more compact, edge-based mode using the XGBoost algorithm. The latter can be trained directly on the edge device, such as the AI hub, providing an offline, self-sufficient alternative that is ideal for simpler maintenance tasks like fault detection and machine state monitoring.

Future improvements could include incorporating a standard for advanced tool life and wear estimation features, enabling more accurate predictions of tool longevity and optimized maintenance schedules. The system's integration with Node-RED for realtime visualizations of predictions further enhances its utility for live monitoring, providing continuous insight into machine health. The data flow can also be adjusted to work with external systems such as enterprise resource planning or cloud-based monitoring platforms.

#### 5 DISCUSSION AND CONCLUSIONS

In this paper, a cost-effective smart sensor system for industrial monitoring and predictive maintenance was presented. This system leverages affordable, low-cost components and open-source based software to enable real-time data collection, processing, and analysis. By optimizing power consumption and implementing AI algorithms, the system provides valuable insights into machine behavior, enabling predictive maintenance, quality monitoring, wear detection, and tool life analysis. The results demonstrate the effectiveness of the approach to improve productivity, reducing downtime, and increasing overall efficiency. As Industry 4.0 continues to evolve, the smart sensor system proposed in this paper could play a significant role in transforming industrial manufacturing processes by lowering the barriers of entry.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## REFERENCES

- [1] Rossi, A.; Bocchetta, G.; Botta, F.; Scorza, A.: Accuracy Characterization of a MEMS Accelerometer for Vibration Monitoring in a Rotating Framework. *Applied Sciences* 13(8), 5070 (2023)
- [2] STMicroelectronics, AN5850: ISM330IS - Always-on 3-axis Accelerometer and 3-axis Gyroscope with ISPU, [https://www.st.com/resource/en/application\\_note/an5850-ism330is-always-on-3axis-accelerometer-and-3axis-gyroscope-with-ispu--intelligent-sensor-processing-unit-stmicroelectronics.pdf](https://www.st.com/resource/en/application_note/an5850-ism330is-always-on-3axis-accelerometer-and-3axis-gyroscope-with-ispu--intelligent-sensor-processing-unit-stmicroelectronics.pdf), last accessed 2024/10/22
- [3] Romansini, M.; de Aguirre, P. C. C.; Compassi-Severo, L.; Girardi, A. G.: A Review on Vibration Monitoring Techniques for Predictive Maintenance of Rotating Machinery. *Eng* 4(3), 1797–1817 (2023)
- [4] Pedotti, L. A. D. S.; Zago, R. M.; Fruett, F.: Fault diagnostics in rotary machines through spectral vibration analysis using low-cost MEMS devices. *IEEE Instrumentation & Measurement Magazine* 20(6), 39–44 (2017)

# AI-based Real-Time Analysis of Sawing Processes for the Identification of Materials

Christoph Dünn<sup>1</sup>, Jens Kneifel<sup>1</sup>, Robin Roj<sup>1</sup>, André Jaquemod<sup>2</sup>, Kamil Güzel<sup>2</sup>, Christian Pelshenke<sup>1</sup>, Ralf Theiß<sup>1</sup>, Hans-Christian Möhring<sup>2</sup> and Peter Dültgen<sup>1</sup>

<sup>1</sup>Forschungsgemeinschaft Werkzeuge und Werkstoffe e. V., Papenberger Str. 49, 42859 Remscheid, Germany, roj@fgw.de

<sup>2</sup>Universität Stuttgart, Institut für Werkzeugmaschinen, Holzgartenstraße 17, 70174 Stuttgart, Germany

**Abstract.** In industrial machining processes, guaranteeing process stability is of great economic relevance for ensuring workpiece quality. If deviations are not detected instantly, this leads to production faults and the associated reworking or new production. This paper presents exemplarily some results of a research project that was focused on self-learning monitoring methods to evaluate the tool and workpiece condition and to report critical situations. The project involved validating machine learning methods and developing a guideline. An important sub-goal is to help small and medium-sized enterprises to implement such a process analysis based on artificial intelligence in their production environments. In detail, this paper presents the results of a classification of different materials that are automatically detected on a band saw. For that reason, the machine was equipped with retrofit sensor technology and several test specimens made of different materials were cut. The collected data was analyzed by numerous feature-based models in order to identify the optimal artificial intelligence method. It could be shown that automatic classification can identify the different materials by their vibration characteristics. Technical maturity and transferability to other machining processes and machines was ensured through the use of open-source software for the in-house development of the entire infrastructure for data acquisition, processing, and analysis. This includes a range of functions from the areas of anomaly detection, predictive quality, predictive maintenance, and other specific approaches on the way to Industry 4.0.

**Keywords:** Artificial Intelligence, Predictive Maintenance, Retrofit, Sawing Processes, Sensor Data Analysis.

## 1 INTRODUCTION

The use of machine learning (ML) methods in industrial production offers significant economic benefits for machine operators by optimizing processes, improving product quality, and minimizing downtime. ML algorithms analyze large amounts of production data, identify patterns as well as anomalies and can be used to optimize machine parameters in real time. This leads to a significant reduction in operating costs and improved resource utilization. ML algorithms detect and correct production errors at an early stage, reduce the reject rate and increase product quality. The investment in ML-based technologies quickly pays for itself through savings in operating costs, quality improvements and a reduction in downtime. Optimized production processes enable higher machine utilization and better scalability. The use of ML can significantly reduce operating costs. However, implementation requires technical specialist know-how and a solid database. In general, the results prove that ML has a wide range of possible applications and supports quality assurance. Primarily, the handling of large amounts of data, data processing, and feature extraction were considered. A standardized plug-and-play solution for implementation in small and medium-sized enterprises (SME) would be desirable but cannot be classified as feasible. The feasibility and potential in manufacturing SMEs and the means by which these technological solutions are implemented in production were also considered.

ML methods are attracting particular attention in various application areas. Starting with methods of fuzzy logic and the use of neural networks (NN), methods from the field of artificial intelligence (AI) and ML have been researched for many years with regard to monitoring tasks in machining production technology. E.g., [1-3] provide an overview of the application of NN in machining processes. Current research approaches

usually pursue the goal of drawing conclusions about the process using a large number of different sensors on the tool, workpiece, or machining center as well as complex analysis measures. AI and ML methods have been used for some time now. In 2018, Dimla provided an approach for automated tool monitoring in which NNs are also applied [4].

AI methods offer great potential in the processing and evaluation of sensor-based component, system, and process status information in production and for the implementation of monitoring and process control strategies [5]. ML makes it possible to profitably utilize an increasing amount of available data in manufacturing companies. This data can represent information about production processes and their process parameters or component quality. By applying ML approaches, this data can be analyzed and evaluated in order to be used to reduce maintenance costs, detect defective parts, or for intelligent process control. With the help of statistical procedures, heuristics, AI methods, and swarm intelligence, fuzzy correlations or systems that contain unknown parameters and parameter dependencies can be approximated, mapped, and described.

Using AI algorithms, self-optimizing machine systems in the manufacturing environment are the focus of several research projects. In particular, the Decision Tree, Support Vector Machine (SVM), Random Forest, k-Nearest Neighbor (KNN) and Bayesian Networks methods have been researched in recent years and are increasingly being investigated with regard to their suitability for use in the manufacturing environment [6].

In contrast to a wide range of exemplary work on the use of individual AI or ML methods, there are only isolated approaches in which different algorithms are examined comparatively. Ambhore et al. summarize the current state of knowledge on tool monitoring [7]. Among other things, they present current approaches for evaluating the tool condition based on cutting force, structure-borne sound, acceleration, power, temperature, and workpiece surface measurements. It is concluded that the use of modern AI models and classification algorithms will enable more reliable and robust monitoring systems in the future. Nevertheless, the use of AI and ML methods continues to pose considerable challenges, especially for SMEs, as their implementation requires a certain amount of specialist knowledge on the one hand and a not inconsiderable amount of effort is required for successful application due to the possible variety of data on the other.

Wear-resistant materials such as Inconel or Stellite are indispensable in high-performance applications, particularly in the aerospace and power generation industries. Although their outstanding strength, hardness, and heat resistance are necessary, they make machining considerably more difficult. They are therefore classified as difficult to machine. Inconel, a nickel-chromium alloy, is often used in extreme temperature environments, such as in turbines or aircraft engines. These alloys retain their strength even at high temperatures and offer high resistance to oxidation and corrosion. However, machining leads to rapid tool wear and high temperatures, which makes it considerably more difficult. Stellite, another high-strength alloy based on cobalt, has similar machining difficulties, particularly due to its hardness and wear resistance.

There is often no alternative to using such materials as they are required in critical applications where conventional materials would fail. Numerous approaches have been developed to improve the machining of these difficult-to-cut materials, such as optimized cutting tools made of special carbides or ceramics that withstand the high probability of wear. In addition, innovative cooling methods and lubricants are used to reduce the heat generated during machining [8].

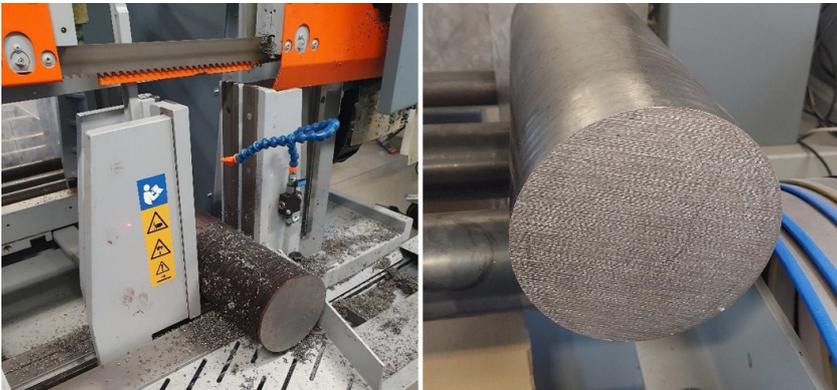
Acoustic sensors measure the airborne noise generated by the machine and can be used to detect abnormal operating noises that indicate malfunctions or wear. They are an important tool for early fault detection and help to improve machine availability [9]. Edge computing-capable sensors can process data directly on site, which is particularly advantageous in dynamic processes. The integration of this sensor data with ML makes it possible to better understand complex relationships in machining processes and to make predictions about machine conditions or process deviations. These predictions are particularly important for predictive maintenance and process optimization [10]. The greatest challenge in sensor technology lies in integration and data processing. The large amount of data generated must be processed efficiently and converted into useful information. Another limitation is sensor placement, as not every location in a machine tool is accessible or suitable to enable precise measurements. In addition, environmental conditions such as high temperatures, humidity, or mechanical disturbances can affect the accuracy and reliability of the sensors.

In this paper, some of the results from a research project aimed at providing SMEs with easy access to AI technology and process analysis are presented. This

objective was considered at every stage of development. In particular, this paper focuses on the AI-based automatic detection of different materials processed by a sawing machine. Similar to a previous publication [11], this paper aims to demonstrate how quickly and efficiently in-house developed AI systems, utilizing low-cost hardware and open-source software, can solve machine-specific problems and support the specialized requirements of SMEs.

## 2 MATERIALS AND METHODS

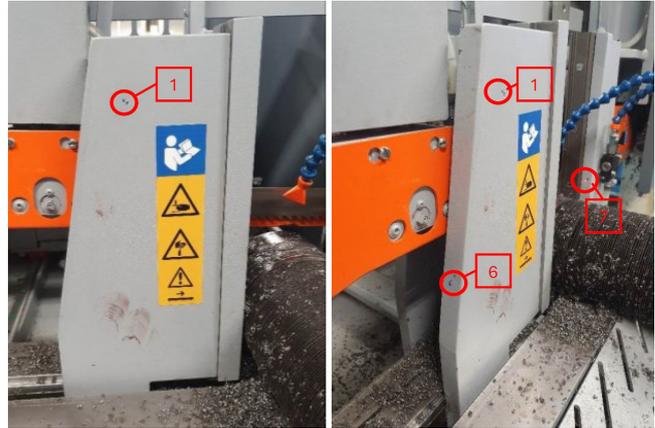
At the beginning of the machining and data collection tests, it was decided to first machine standard materials in different shapes and use them to test data acquisition and further processing. Accordingly, cylindrical round material made of 1.0503 (C45+C/+SH) according to DIN EN 10277:2018-09 with a diameter of 150 mm was purchased [12]. Fig. 1 shows an example of this raw material in the form of 1 m long pieces on the MEBA saw. Subsequently, it was decided to focus on material with a diameter of 40 mm. Thus, other materials including cast rods made of Celsit V and round material made of Inconel 718 were processed and analyzed [13,14].



**Fig. 1:** Cylindrical C45 with a diameter of 150 mm.

In the course of the test setup and the first test measurements on the MEBA e-cut 400A bandsaw machine, it was first necessary to equip it with appropriate sensor technology. In order to best capture the vibrations that occur within the bandsaw during the sawing processes, tests were carried out to determine the optimum positioning of the retrofit hardware. A laser Doppler vibrometer was used for these tests, with which the vibrations on the band saw were recorded at different positions. Due to natural vibrations, certain housing components as well as the guard plates of the band saw were not taken into consideration. This was to avoid falsified recordings of the sensors. Another aspect in the selection of the test positions for possible sensor

arrangements was, among other things, perfect and risk-minimized cable routing for the sensors. The cables had to be adapted to the full range of movement of the respective components of the band saw and possible damage caused by the saw blade and sawdust had to be avoided. The test positions are shown in Fig. 2.



**Fig. 2:** Definition of several potential sensor positions at the bandsaw MEBA e-cut 400A.

Test position 1 is on the clamping jaw facing the front, whereby position 6 is on the side of the clamping jaw. The left one is moved adaptively towards the component to be sawn and is therefore a moving component. The aim of the measurements is to directly record the vibrations in the component during the sawing process, whereby a moving component is only considered secondarily. Test positions 2, 3, 4, 5, and 7 are located on the right-hand clamping jaw, which is permanently integrated and connected to the band saw. This means that the vibrations that occur within the entire band saw are transmitted to this clamping jaw. It should be noted that test position 7 is located to the side of the right clamping jaw and therefore

close to the object to be sawn. The danger here, however, is that during operation of the band saw, components with a large edge length are also processed, which could come into contact with a possible sensor at position 7 in an emergency. For this reason, a possible positioning in this area was initially only considered secondarily. The results obtained using the vibrometer show that points 3, 4, and 5 have little noise in the entire frequency range, making them suitable for positioning sensors. Since the recordings of the frequencies in points 4 and 5 show less noise than point 3, these two points were shortlisted for the positioning of the required sensors. The following diagram in Fig. 3 depicts the results of the measurements taken with the laser Doppler vibrometer. The X-axis shows the time in ms. The Y-axis shows the frequency in Hertz. The intensity of the discoloration represents the frequency. The vertical lines can be ignored as they are overload signals. A direct comparison of the plots of the different points shows that the horizontal lines are relatively uniform. In order to avoid or minimize possible contamination, point 4 was ultimately selected as the position for the sensor, as it is not positioned near the workpiece to be processed.

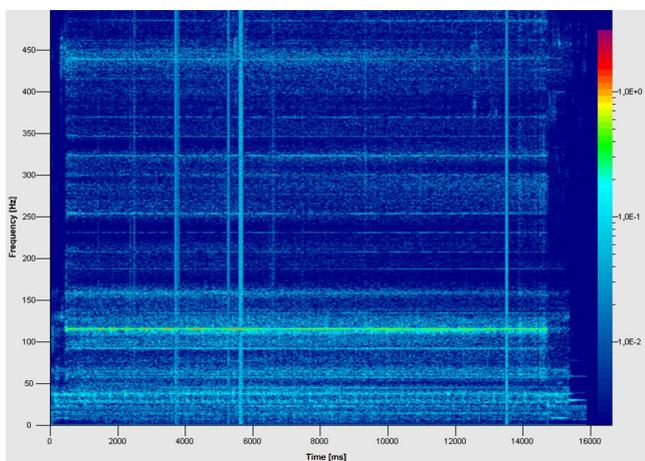


Fig. 3: Frequency in point 4 recorded by a laser Doppler vibrometer.

For further preparation of the sawing tests, the hardware was installed and used as an edge server for data processing. The setup was designed to enable live monitoring of the sawing test, live display of the NN, and live image transmission for observing the sawing test. Thanks to the structure of the data processing system, the recorded vibrations from the sawing tests could be processed and transmitted without any problems. The data was properly stored, processed, and made available to the NN. This allowed the data recorded by the sensors to be evaluated in near real time during the tests and all data control steps to be carried out via remote access.

The carbide band saw blade used for the tests with the dimensions 6220 mm x 41 mm x 1.3 mm and a tothing of 3/4 K was heavily stressed in the first tests. There are no fixed criteria as to when a band is considered worn, as wear increases continuously. Quality defects on the workpiece do not necessarily occur, but the probability of damage increases. If the parameters remain constant, wear is particularly evident in the vibration pattern. The tests before the belt change were mostly carried out in a worn condition. After changing to a new band saw blade with the same settings, some tests could be repeated and a direct comparison carried out. The measurement procedure was identical for all tests, although different sensors and measurement parameters were initially experimented with. For a better overview, these preliminary tests will not be discussed in detail. First, the saw was prepared, the parameters set, and the material clamped. The sensor settings (sensitivity and measuring range) were made via the aforementioned user interface. The same interface was also used to name the measurement data and document the tests. The system was programmed in such a way that the measurements were started automatically from the *OPC UA* signal via the actual pressure and the measured values were recorded with the self-made hardware called *NodeOne* [15]. The automated recording meant that large amounts of data could be collected quickly as long as the same material was used.

Due to the high sampling rates, numerous data points were generated per second, which contain a lot of information but cannot be read out efficiently. Data preparation was therefore required for feature-based learning. Various feature extractors are available for this purpose. A frequently used open-source Python library is the *Time Series Feature Extraction Library* (TSFEL) [16], which provides a large number of spectral, statistical and temporal features. The data preparation process is shown in the flowchart in Fig. 4.



Fig. 4: Flow chart for the visualization of principles for data preparation.

To ensure a fast response time of the system, the raw data was divided into sections with a length of 1 s each and then combined into several features. The overlap of the data blocks was set at 50%. The first models were trained on this basis. The open-source library PyCaret was used

to compare various algorithms [17]. One example of this is the Light Gradient Boosting Machine (LGBM) framework [18], which is based on decision tree algorithms and is used for tasks such as ranking, classification, and other ML tasks. The focus of the framework is on performance and scalability. For further evaluation, the relevance of each feature in the selected model was examined so that the number of features could be reduced in subsequent iterations until an efficient ratio between model accuracy and computational effort existed.

First, the data preparation for the automated detection of different materials was carried out. The measurement data of the different materials was annotated for this classification. No distinction was made between the wear of the band saw blade. The setup data set with all features and the classification was combined and mixed. This gave the model a uniform training that was independent of the order of the data.

The dataset was split into two parts: 80% of the data was used for training, again 80% of this training set was used for the training process, and 20% was used to validate the training to avoid overfitting. The remaining 20% of the total data points were used for the final validation. This has the advantage that this data was not included in the training and thus a more realistic estimation of the model performance is based on new, previously unknown data. The training data set was loaded, and the training was performed with the open-source Python library PyCaret. This makes it possible to test the following different model categories:

#### 1. Tree-based models

- *Extra Trees Classifier*
- *Random Forest Classifier*
- *Gradient Boosting Classifier*
- *LGBM*
- *Decision Tree Classifier*

#### 2. Ensemble methods

- *Random Forest Classifier (also tree-based)*
- *Extra Trees Classifier (also tree-based)*
- *Gradient Boosting Classifier (also tree-based)*
- *LGBM (also tree-based)*

#### 3. Linear models

- *Logistic Regression*
- *Ridge Classifier*
- *SVM - Linear Kernel*
- *Linear Discriminant Analysis (LDA)*

#### 4. Non-linear models

- *SVM - Linear kernel (also linear)*
- *KNN Classifier*

#### 5. Bayes models

- *Naive Bayes*
- *Quadratic Discriminant Analysis*
- *LDA (also linear)*

#### 6. Baseline models

- *Dummy Classifier*

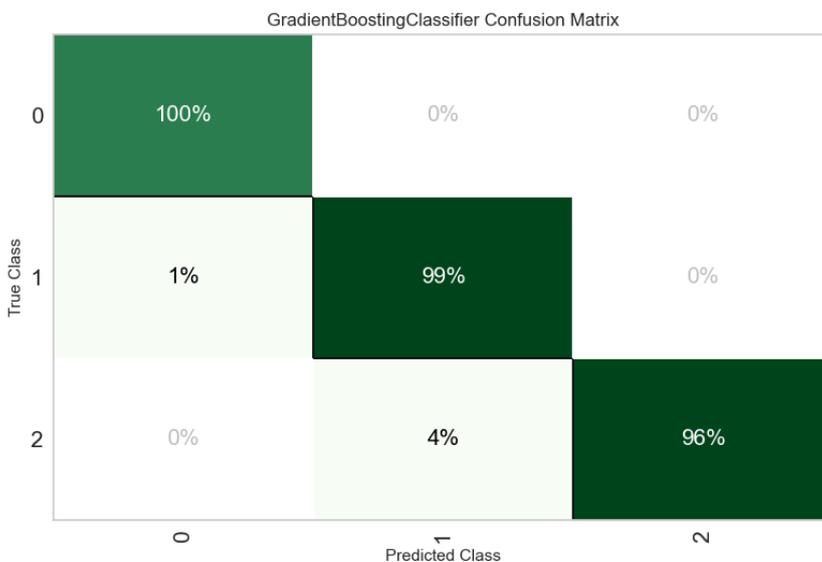
Table 1 shows the F1 score and the accuracies. These metrics offer a comparison of the performance of the models and help to select the optimum for the respective task.

Model	F1 Score	Accuracy
Extra Trees Classifier	0.9954	0.9954
LGBM	0.9945	0.9946
Gradient Boosting Classifier	0.9942	0.9942
Random Forest Classifier	0.9934	0.9934
KNN Classifier	0.9931	0.9931
LDA	0.9903	0.9904
Logistic Regression	0.9891	0.9892
Ridge Classifier	0.9891	0.9892
SVM - Linear Kernel	0.9855	0.9857
Decision Tree Classifier	0.9803	0.9803
Ada Boost Classifier	0.9292	0.9359
Naive Bayes	0.8758	0.8649
Quadratic Discriminant Analysis	0.5288	0.5368
Dummy Classifier	0.3426	0.5083

**Tab. 1:** Overview of the F1 scores and accuracies of all models for classification.

### 3 RESULTS

The results show that most of the models tested are highly accurate and enable reliable classification. The LGBM was selected for further evaluation. This model is based on decision tree algorithms and is designed for tasks such as ranking, classification, and other ML tasks. The focus in the development of LGBM is on performance and scalability, which makes it a suitable choice for large and complex datasets. The confusion matrix in Fig. 5 illustrates the performance of LGBM in classifying the three materials C45, Celsit V, and Inconel.



**Fig. 5:** Confusion matrix for classification of different materials.

**C45:** The model achieves a perfect result for this class with an accuracy of 100%. All samples of C45 were correctly detected without any confusion with other classes.

**Celsit V:** Here too, the model performs almost perfectly with an accuracy of 99%. Only 1% of the samples from Celsit V were incorrectly classified as C45.

**Inconel:** The Inconel class was classified correctly 96% of the time. However, 4% of the Inconel samples were recognized as Celsit V, which indicates a low difficulty of the model to completely separate these two classes.

Overall, the model shows a very high classification accuracy, especially for the classes C45 and Celsit V. Most confusions occur with the Inconel class, which was incorrectly classified as Celsit V in some cases. These confusions could be an indication of a structural similarity between the two classes in the feature space and represent a potential starting point for further model optimization. In addition, the model can be extended to include other materials and the accuracy for similar materials can be further investigated. The accuracy was checked using the previously separated validation data. The results of this validation are shown in Table 2. These confirm that the model retains a high level of accuracy even with the new data not used in the training.

Accuracy on test data	0.99
Precision on test data	0.99
Recall on test data	0.99
<i>F1 score</i> on test data	0.99

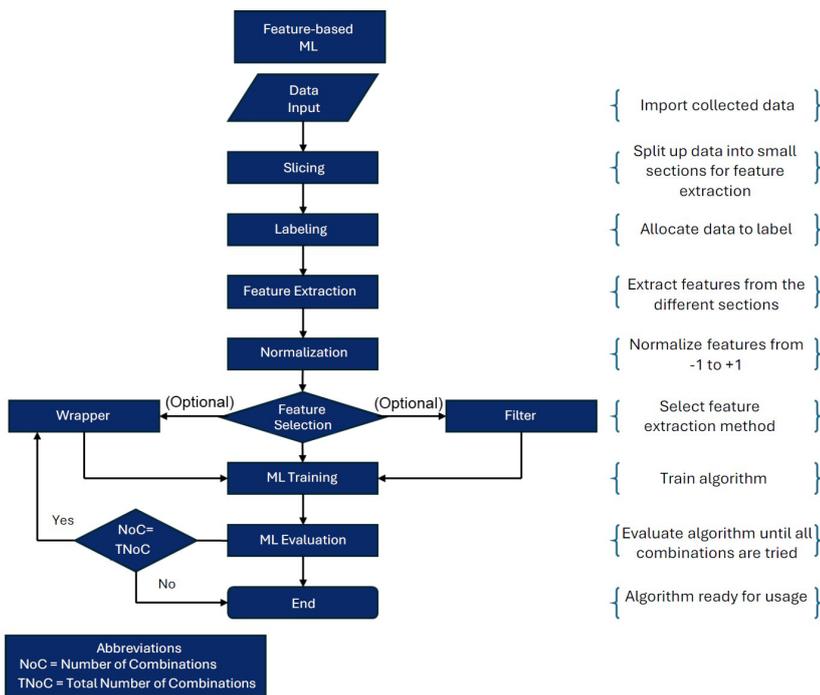
**Tab. 2:** Results of the classification of different materials.

## 4 DISCUSSION AND CONCLUSIONS

The flowchart in Fig. 6 visualizes the procedure and serves as an initial orientation for SMEs. It provides an overview of the key steps and requirements that should be taken into account when introducing ML to a company. It also visualizes basic information and typical processes, but is not necessarily universally valid, as every company has specific challenges and framework conditions that need to be adapted individually.

First, the company should formulate clear goals for the use of ML. It is important to understand which problems are to be solved and which benefits can be achieved through ML. This could include the optimization of production processes or the prediction of maintenance requirements. As ML models require high quality data, it is important to take stock of the existing data sources and ensure that the data is complete, accurate, and relevant. In many cases, it is necessary to cleanse and transform data before modeling to make it usable for ML algorithms. In the case of decomposition, this means that the data must be recorded. To do this, sensors must be placed in the selected process.

The sensors are used to collect the relevant data, which is followed by the selection of the right ML tools and platforms. For SMEs, it makes sense to implement the use of existing ML solutions and not to carry out costly in-house development. After data preparation and the selection of suitable ML methods, model development begins. In this phase, various ML models are created and trained on the basis of the collected data. The model is optimized through repeated training processes and validations in order to make accurate predictions or decisions. The ML model is then integrated into the processes of the company. This can include integration into existing software systems, the automation of certain tasks, or the support of decision-making processes. The integration should be carried out in such a way that it does not disrupt ongoing operations and fits seamlessly into the existing infrastructure. At this point, the need for early planning and strategic integration into company structures becomes clear and the research results show the complexity of the entire topic.



**Fig. 6:** Overview of necessary steps to integrate ML into the processes of a company.

The main objective was to provide SMEs in the field of machining with an easy introduction to the use of AI for analyzing their manufacturing processes. The focus was on the machining of difficult-to-cut and inhomogeneous materials in order to increase the efficiency of machine and tool use and to reduce wear.

Several sub-goals were pursued, all of which were geared towards the overarching goal. Particular emphasis was placed on showing companies various options for gradually integrating AI technology into their existing infrastructure. This should enable them to achieve initial successes and recognize the potential of AI-based analyses for optimizing their manufacturing processes.

The analysis of the manufacturing processes focused in particular on difficult-to-machine and inhomogeneous materials using the example of sawing. By using ML, data from the ongoing processes could be evaluated in (near) real time, enabling companies to identify weak points in their processes and take targeted optimization measures (predictive quality, predictive maintenance). It was shown that it is possible to improve the efficiency of machine and tool utilization with low-threshold software and hardware solutions. In the laboratory environment, it was possible to reduce tool wear. Nevertheless, it should be noted that the mentioned experiments were conducted on a laboratory level. The reported high accuracies for the models may be unrealistic or overestimated without proper real-world validation, raising concerns

about potential overfitting or biases in the dataset. For future work, an industrial transfer should be considered, which will split up those clear laboratory results into variant real-world insights. Then, it will be much more important to provide a robust model that is able to compensate possible interferences. This demonstrates the potential to achieve not only a longer tool life but also higher production quality and speed in an SME production environment.

During the course of this project, it became evident that the use of AI in manufacturing offers numerous advantages, particularly for SMEs, which often have limited opportunities to optimize their processes based on data due to a lack of expertise and potentially a shortage of personnel. The findings and methods developed as part of the project now offer these companies a valuable starting point for

securing their long-term competitiveness and asserting themselves in an increasingly technology-driven market environment.

**Acknowledgments.** The research project IGF 22150 N „Analyse der Anwendung von Machine Learning Methoden für die Echtzeitüberwachung der Zerspanung inhomogener Werkstoffe zur Beurteilung von Werkzeugzustand und Werkstückqualität in kleinen und mittelständischen Unternehmen“ from the Research Association Forschungsgemeinschaft Werkzeuge und Werkstoffe e.V. (FGW), Papenberger Straße 49, 42859 Remscheid, Germany, is supported by the Federal Ministry of Economic Affairs and Climate Action the German Federation of Industrial Research Associations (AiF) as part of the program for promoting industrial cooperative research (IGF) based on a decision by the German Bundestag.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## REFERENCES

- [1] Cuentas, S.; Peñabazena-Niebles, R.; Garcia, E.: Support vector machine in statistical process monitoring: a methodological and analytical review. *The International Journal of Advanced Manufacturing Technology* 91(4–5), 485–500 (2017)
- [2] Chandrasekaran, M.; Muralidhar, M.; Krishna, C. M.; Dixit, U. S.: Application of soft computing techniques in machining performance prediction and optimization: a literature review. *The International Journal of Advanced Manufacturing Technology* 46(5), 445–464 (2010)
- [3] Sick, B.: On-Line and Indirect Tool Wear Monitoring in Turning with Artificial Neural Networks: A Review of more than a Decade of Research. *Mechanical Systems and Signal Processing* 16(4), 487–546 (2002)
- [4] Dimla, E.: Development of an Innovative Tool Wear Monitoring System for Zero-Defect Manufacturing. *International Journal of Mechanical Engineering and Robotics Research* 7(3), 305–312 (2018)
- [5] Balazinski, M.; Czogala, E.; Jemielniak, K.; Leski, J.: Tool condition monitoring using artificial intelligence methods. *Engineering Applications of Artificial Intelligence* 15(1), 73–80 (2002)
- [6] Fahle, S.; Prinz, C.; Kuhlentötter, B.: Systematic review on machine learning (ML) methods for manufacturing processes – Identifying artificial intelligence (AI) methods for field application. *Procedia CIRP* 93, 413–418 (2020)
- [7] Ambhore, N.; Kamble, D.; Chinchankar, S.; Wayal, V.: Tool Condition Monitoring System: A Review. *Materials Today: Proceedings* 2(4–5), 3419–3428 (2015)
- [8] Pedroso, A. F. V.; Sousa, V. F. C.; Sebbe, N. P. V.; Silva, F. J. G.; Campilho, R. D. S. G.; Sales-Contini, R. C. M.; Jesus, A. M. P.: A Comprehensive Review on the Conventional and Non-Conventional Machining and Tool-Wear Mechanisms of INCONEL®. *Metals* 13(3), 585 (2023)
- [9] Carrino, S.; Guerne, J.; Dreyer, J.; Ghorbel, H.; Schorderet, A.; Montavon, R.: Machining Quality Prediction Using Acoustic Sensors and Machine Learning. *Proceedings* 63(1), 31 (2020)
- [10] Lee, W. J.; Wu, H.; Yun, H.; Kim, H.; Jun, M. B. G.; Sutherland, J. W.: Predictive Maintenance of Machine Tool Systems Using Artificial Intelligence Techniques Applied to Machine Condition Data. *Procedia CIRP* 80, 506–511 (2019)
- [11] Kneifel, J.; Böttinger, C.; Roj, R.; Pelshenke, C.; Theiß, R.; Dültgen, P.: Vibration Analysis of Sawing Processes aided by AI. In: 2nd Berlin Workshop on Artificial Intelligence for Engineering Applications, pp. 6–12. GFal - Gesellschaft zur Förderung angewandter Informatik e.V., Berlin (2023)
- [12] DIN EN 10277:2018-09, Bright steel products - Technical delivery conditions. Beuth Verlag GmbH, Berlin (2018)
- [13] Deutsche Edelstahlwerke Specialty Steel GmbH & Co. KG, Metallpulver und Stranggussstäbe, [https://swisssteel-group.com/content-media/documents/Brochures/Special-Material/050\\_DEW\\_Metallpulver\\_D.pdf](https://swisssteel-group.com/content-media/documents/Brochures/Special-Material/050_DEW_Metallpulver_D.pdf), last accessed 2024/10/18
- [14] VDM Metals International GmbH, VDM® Alloy 718, [https://www.vdm-metals.com/filead-min/user\\_upload/Downloads/Data\\_Sheets/Datenblatt\\_VDM\\_Alloy\\_718.pdf](https://www.vdm-metals.com/filead-min/user_upload/Downloads/Data_Sheets/Datenblatt_VDM_Alloy_718.pdf), last accessed 2024/10/18
- [15] Kneifel, J.; Roj, R.; Woyand, H.-B.; Theiß, R.; Dültgen, P.: An IIoT-Device for Acquisition and Analysis of High-Frequency Data Processed by Artificial Intelligence. *IoT* 4(3), 244–264 (2023)
- [16] Fraunhofer AICOS, Time Series Feature Extraction Library (TSFEL), <https://tsfel.readthedocs.io/en/latest/>, last accessed 2024/10/18
- [17] Moez Ali, PyCaret: An open source, low-code machine learning library in Python, <https://pycaret.org/>, last accessed 2024/10/18
- [18] Microsoft Corporation, Light Gradient Boosting Machine, <https://lightgbm.readthedocs.io/en/stable/>, last accessed 2024/10/18

# An Object-Based Approach for Classifying Flake Graphite Cast Iron Micrographs through Semantic Segmentation

Marc Goedecke and Martin Pfaff

Gesellschaft zur Förderung angewandter Informatik e.V., Volmerstraße 3, D-12489 Berlin, Germany  
www.gfai.de

**Abstract.** Classifying microstructures in flake graphite cast iron (GJL) micrographs can be challenging due to the frequent occurrence of mixed microstructures and the smooth transitions between their microstructure types. In this paper, we propose a method to segmentate the microstructures of flake graphite cast iron micrographs based on the DIN EN ISO 945-1 standard. This method uses an object-based approach, storing separate micrograph regions corresponding to specific microstructure types as individual binary images (binary objects) in a dataset. We can then create synthetic micrographs by randomly arranging these objects and converting the result into a grayscale image using a U-Net model. Since we assign each object to a specific microstructure type, we can also create the corresponding semantic segmentation masks for the synthetically generated micrographs. With the resulting dataset of the synthetically generated micrographs and their respective masks, we trained a U-Net model to enable an objective, reproducible, and accurate evaluation of the micrographs. Tested on real-world micrographs, the trained model showed difficulties in generalization and using global texture features and patterns. Possible solutions to these challenges and approaches to improve the results are given in an outlook.

**Keywords:** flake graphite cast iron · GJL · DIN EN ISO 945-1 · semantic segmentation · classification

Gefördert durch:



aufgrund eines Beschlusses  
des Deutschen Bundestages



## 1 INTRODUCTION

Technological advancements increase the demands placed on products during their development and evolution. This raises expectations for materials like cast iron, necessitating a meaningful quantitative description of their microstructure to determine and control the material properties.

Flake graphite cast iron (GJL) is a versatile material used in various machines and systems. Its unique properties make it suitable for complex geometries such as cylinder crankcases, cylinder blocks, compressors, steam turbines, pumps, and valves. [1]

For qualitative characterization of the microstructure of GJL, the samples are ground and polished so that the lamellar structure is visible under the microscope. The classification is then performed according to industry standards (DIN EN ISO 945-1 [2], ASTM A247 [3]), which provides reference images for the different classes.

However, using the current standard shows difficulties in evaluating the microstructures, as the structures represent only a simplified two-dimensional representation of the three-dimensional GJL objects. This problem was previously highlighted by the results of the DIAgraph II (IGF No. 20650 N) project ring test, where participants produced varying segmentation masks for a single GJL micrograph [4]. Additionally, in the current DIAgraph ML (IGF No. 22783 N) project, we demonstrated that even a single person's assessment can vary over time when classifying the same GJL micrograph at different times. Therefore, there is an actual demand for a deterministic and objective evaluation of GJL microstructures.

Methods such as focused ion beam tomography (FIB) [5] or computer tomography (CT) [6] can also offer initial approaches for characterizing the 3-dimensional structure of the GJL. Moreover, these approaches can

also be used to get correlations between 2-dimensional cross-sections and 3-dimensional structures of GJL objects.

In this paper, we present an approach for semantic segmentation that classifies GJL microstructures into their respective arrangement classes. To achieve this, we train a U-Net model [7] to generate a segmentation mask for a given input micrograph. The segmentations are based on the DIN EN ISO 945-1 standard.

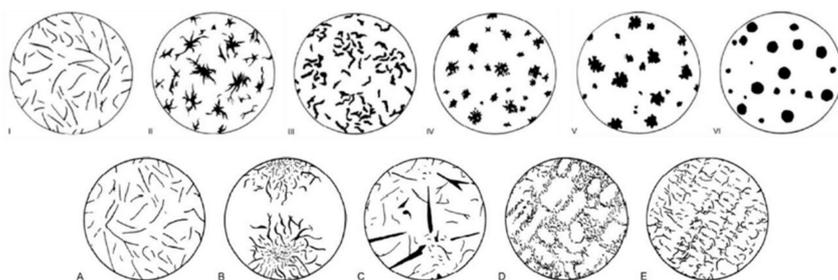
Given the time-consuming nature of creating segmentation masks and the variability in classifying complete microstructures, we focus on creating objects using crops of the microstructures. Each object is explicitly assigned to a specific arrangement class. By randomly arranging these objects, we can generate synthetic micrographs and their corresponding semantic segmentation masks. These synthetic micrographs are then used to train the U-Net model, and the model's performance is evaluated using common metrics such as Intersection-over-Union (IoU) and Dice Score [8].

Afterward, we will discuss the results based on real-world micrographs and address adjustments we plan to make for the future progress of this project.

## 2 PRELIMINARIES

### 2.1 GRAPHITE STANDARD: DIN EN ISO 945-1

The DIN EN ISO 945-1 (Fig. 1) [2] standard classifies cast iron into six different graphite form types. These forms are typically referred to as flake graphite (form I), vermicular graphite (form III), temper carbon (form V), and spheroidal graphite (form VI). Furthermore, for cast iron with flake graphite (form I), there is an additional subdivision into five different graphite arrangement classes, which are known as A-graphite, B-graphite (rosette graphite), C-graphite (primary graphite), D-graphite (interdendritic graphite/undercooled graphite), and E-graphite (degenerate interdendritic graphite). The arrangement classes represent an ideal state; however, there are many intermediate stages between these classes, which complicates the qualitative characterization of the microstructure. [2] [9]



**Fig. 1:** Drawings visualizing the six types of graphite forms in cast iron (top row), along with the five arrangement classes of flake graphite cast iron (bottom row). The images represent micrographs at 100× magnification. [2]

### 2.2 U-NET

U-Net is a convolutional neural network (CNN) first introduced by Ronneberger et al. [7]. The model has an encoder-decoder structure with lateral skip connections. These skip connections allow the model to pass information from the encoder layer to the corresponding decoder layer to produce a more exact output. It was named after its architectural shape since the encoder and decoder are constructed symmetrically (contraction step in encoder:  $3 \times 3$  Conv, ReLU,  $3 \times 3$  Conv, ReLU,  $2 \times 2$  max-pool, expansion step in decoder:  $2 \times 2$  ConvTranspose,  $3 \times 3$  Conv, ReLU,  $3 \times 3$  Conv, ReLU). Although the number of channels doubles for each max-pooling operation in the contraction step, it is halved during each transposed convolution of the expansion step. Since the output of the transposed convolution is concatenated with the skip connection, the first  $3 \times 3$  convolution also halves the number of channels.

Since U-Net relies on operations like convolution and max-pooling, the architecture can handle images of arbitrary size as long as the GPU memory is sufficient and the image dimensions allow for contraction. This ensures that the resulting segmentation is seamlessly connected. If GPU memory is limited, we can split the images into overlapping patches, process them independently, and stitch them together to create a seamless segmentation (overlap-tile strategy).

Our implementation differs from the original paper when talking from U-Net in this paper. We used  $3 \times 3$  convolutional layers with padding and added batch normalization [10] after ReLU.

## 3 OBJECT-BASED APPROACH

This chapter will present the object-based approach for the semantic segmentation of GJL micrographs based on the DIN EN ISO 945-1 standard.

### 3.1 OBJECT DATASET

In this part, we will create binary objects for the different arrangement classes. Why binary objects instead of grayscale? Using binary objects, we can isolate the objects from other artifacts in the crop. Furthermore, a separation of the background is desirable due to differences in saturation, brightness, and contrast across the micrographs.

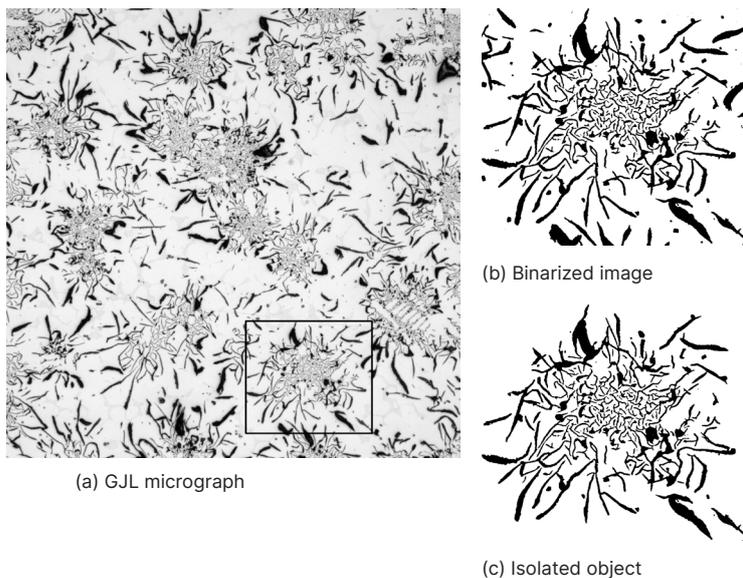
To separate an object, we ensure that the micrograph is given in grayscale format. We cropped the corresponding object area from

the micrograph and select an appropriate threshold for binarization. Afterward, we isolate the object by removing artifacts or other elements (Fig. 2). For specific objects, such as B-graphite, we apply a sharpening method (unsharp mask) before the binarization to better visualize the fine details of the lamellar structure. For A-graphite, we filtered the contours in the images to automatically detect and extract the objects. After reviewing and sorting out the degenerated objects, we created 1411 A objects of 14 micrographs. For the other types of graphite, we had to make the crops ourselves. We ended up with 200 B, 51 C, 76 D, and 84 E objects (81, 11, 9, and 24 micrographs).

Note: To get a better workflow, we trained a U-Net to transform grayscale image crops into binary crops. Equivalent to the U-Net, which transforms binary objects back into grayscale objects. (Section 3.2).

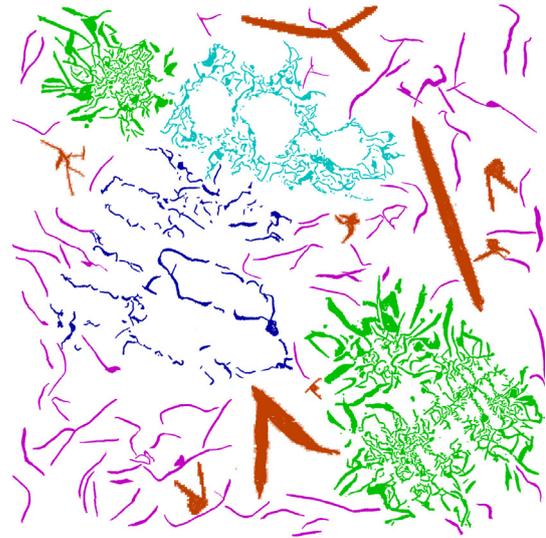
### 3.2 SYNTHETIC MICROGRAPHS

Using the object dataset, we create a binary mask by randomly placing objects in a  $2048 \times 2048$  area. Before positioning, each object is randomly mirrored (horizontal/vertical) and rotated in 45-degree increments. An object is only placed if its contour (white pixels Fig. 2c) does not overlap with the contours of already placed objects. To avoid early placement conflicts, each object is given multiple attempts to find a non-overlapping position. Since each object is assigned to a specific arrangement class, we generate a corresponding semantic mask by applying color coding, as shown in Fig. 3a.



**Fig. 2:** Object extraction in the GJL micrograph: (a) the original micrograph with the target object outlined by a bounding rectangle; (b) a binarized version of the cropped image; and (c) the isolated object after removing artifacts or other objects.

To transform the binary mask into a synthetic micrograph, we trained a U-Net model using  $256 \times 256$  images with binary inputs and grayscale outputs. As images, we used 120 micrographs of a single B graphite sample divided into 13310 training patches and 1210 validation patches. Moreover, this B-graphite sample offers a spectrum of finer and coarser lamellar. Further, we used the binary cross-entropy with logits loss function and the Adam optimizer. After training the model, we perform inference on the binary mask to obtain the synthetic micrograph (see Fig. 3b).



(a) Mask with color-coded classes



(b) Grayscale synthetic image

**Fig. 3:** Fig. Synthetic image generation: (a) random placement of predefined objects with color-coded classes (A: purple, B: green, C: orange, D: cyan, E: blue); and (b) the corresponding grayscale synthetic image.

### 3.3 SEMANTIC SEGMENTATION

With the ability to generate synthetic images along with their corresponding semantic segmentation masks, we created a dataset of 1000 training, 100 validation, and 50 test pairs to train a U-Net model to segment GJL micrographs. During training, we apply random data augmentation, such as changing the brightness or contrast, flipping, rotating, and blurring the input image. The output consists of a binary mask for each GJL class. However, we train the model like a multi-label segmentation even though the binary classes have no overlapping. We wanted to predict the model's confidence for each class independently, so we used a sigmoid activation with a binary cross-entropy loss function over a softmax with categorical cross-entropy.

Fig. 4 shows the good performance of our model based on the IoU and Dice score metrics (visualized using the Seaborn library [11] and Matplotlib [12]). However, due to the small object dataset, the synthetic test images were generated using the same objects as the training and validation data, which may have influenced the results.

### 3.4 EVALUATING ON REAL-WORLD MICROGRAPHS

To test the performance of our trained segmentation model on real-world data, we selected three images that several experts had already classified (Fig. 5). All micrographs showed problems with correctly classifying graphite A. As soon as the lamellar appears to become thicker, they are classified as C. In addition, many A graphite objects have also been classified as E graphite, implying that the existing A objects in the object dataset need to be more diverse.

Furthermore, the B object in Fig. 5c is classified as D. This outcome was expected, as the objects in our object database differ significantly from this one. However, in Fig. 5b, the B objects were correctly identified. Although the objects in the microstructure are not part of the object database, they have a significantly higher similarity because they come from the same sample.

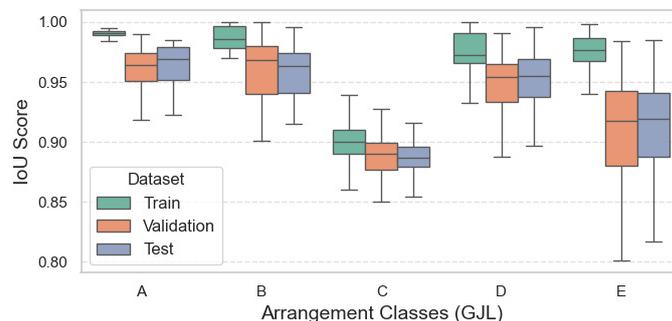
Finally, the classification of D graphite in Fig. 5a was largely successful. This was very positive since we had to divide the D graphite structures (and E) multiple times to get an object of feasible size for the random placement process in Section 3.2.

## 4 OUTLOOK

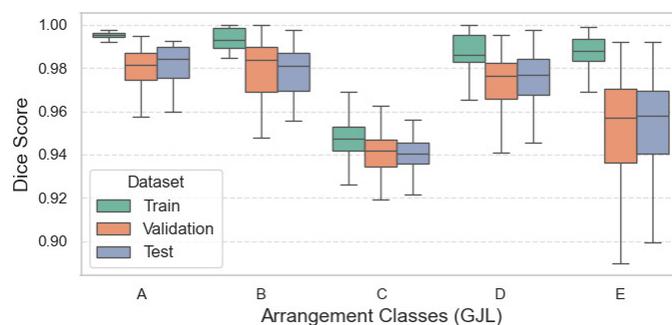
The evaluation of the trained U-Net revealed that it struggles with generalizing to real-world micrographs and

does not recognize global texture features and patterns. The analysis of the micrographs is further complicated by varying the preparation methods, camera qualities, and scaling factors, leading to quality inconsistencies.

To improve future results, we will expand the object dataset by including as many diverse objects as possible. Further, we will add input layers to train the U-Net, covering texture features like the homogeneous distribution of the objects displayed in the microstructure and extracting characteristics specific to the arrangement classes. The object dataset allows us to test whether we can target these features. This means that we could use separate sets of objects to create synthetic micrographs for the training and validation sets. Additionally, we could focus on creating mixed microstructures that occur in reality, such as arrangements like AB, AC, AD, ADE, AE, or DE, and consider the scaling factor (micrometer/pixel) in our representation of the synthetic images and texture layers. Instead of using binary cross-entropy as a loss function, we plan to experiment with categorical cross entropy and focal loss. [13] [14]

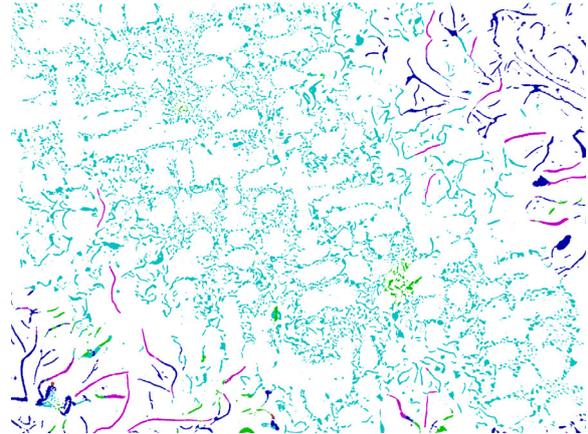
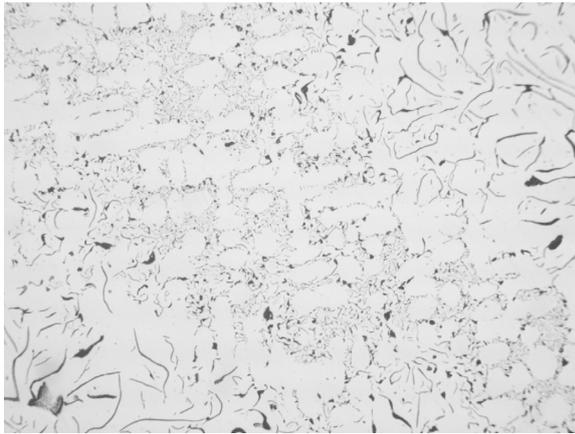


(a) Intersection-over-Union (IoU): Ranges from 0 to 1 with 1 indicating a perfect performance.

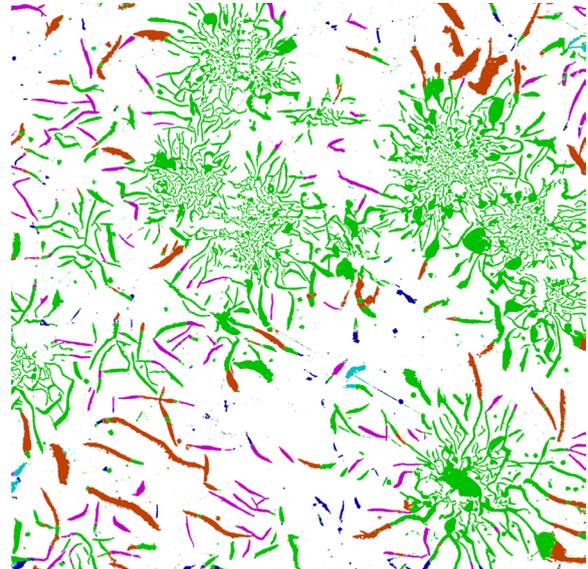
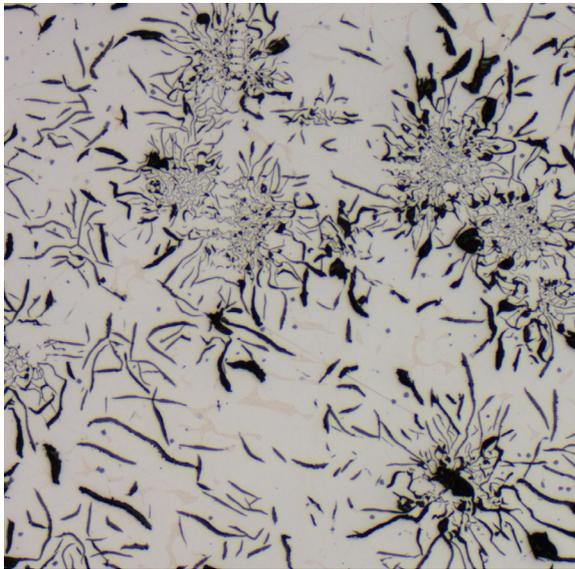


(b) Dice Score: Ranges from 0 to 1 with 1 indicating a perfect performance.

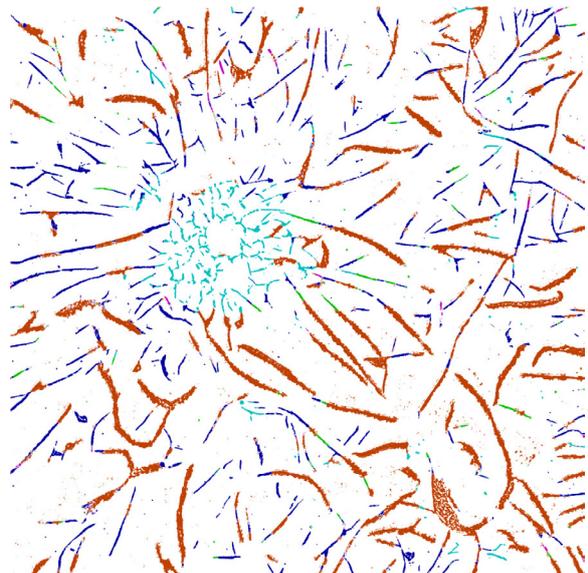
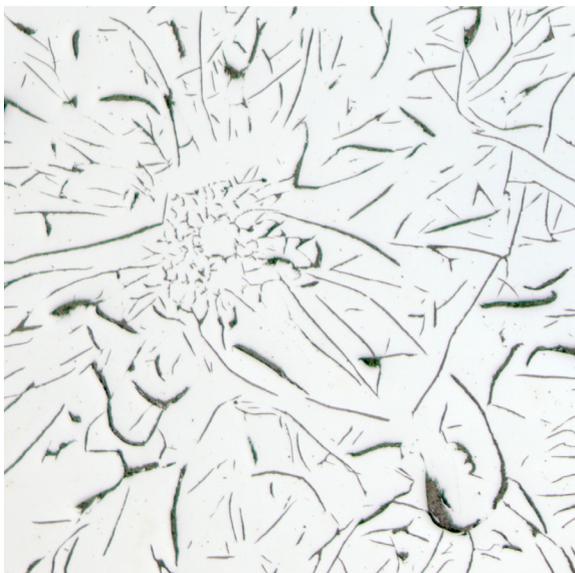
**Fig. 4:** Boxplot diagrams of the performance metrics: (a) Intersection-over-Union and (b) Dice Score, computed for each class using the training, validation, and test datasets from the synthetically generated GJL micrographs.



(a) AD graphite structure (left) and corresponding U-Net segmentation mask (right).



(b) AB graphite structure (left) and corresponding U-Net segmentation mask (right).



(c) AB graphite structure (left) and corresponding U-Net segmentation mask (right).

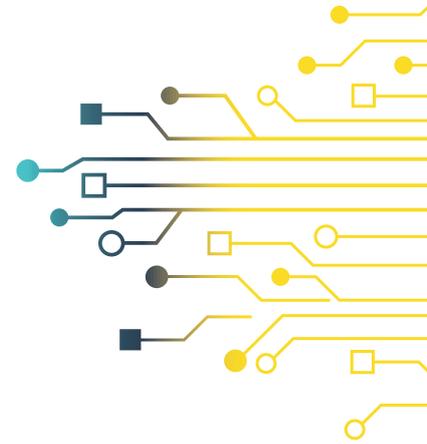
**Fig. 5:** Illustration of three GJL graphite structures (left) and their corresponding segmentation masks (right) generated using U-Net. The segmentation uses the following color code: A: purple, B: green, C: orange, D: cyan, E: blue. Each structure has been qualitatively characterised by multiple experts, ensuring reliable identification of graphite features.

## 5 CONCLUSION

The qualitative characterization of GJL is a complex challenge. The evaluation of the different micrographs often varies from expert to expert. We have presented an approach that should enable an objective and deterministic evaluation by using objects or structures from clearly classified areas of these micrographs. It became clear how challenging it is to create an object data set representing the numerous facets of the lamellar structure. There are ways to diversify the object dataset further; however, this approach will reach its limits the closer we approach the intermediate forms of two classes. Using three-dimensional objects could offer a solution in the future, as two-dimensional objects can be derived from them by looking at different cross-sections.

## REFERENCES

- [1] Forschungsvereinigung Gießereitechnik e.V. Schlussbericht zum IGF-Forschungsvorhaben 20650 N: Bereichsbezogene Klassifikation von Graphitausbildungen in Gusseisen mit Lamellen- bzw. Vermikulargraphit sowie von nicht den Grundformen zuzuordnenden Graphitstrukturen mittels digitaler Bildverarbeitung (DIAgraph II). Technical report, RWTH Aachen, Lehrstuhl für das Gesamte Gießereiwesen und Gießerei-Institut (GI) und Gesellschaft zur Förderung angewandter Informatik e.V. (GFal), Germany, 2022.
- [2] Deutsches Institut für Normung e.V. DIN EN ISO 945-1: Mikrostruktur von Gusseisen – Teil 1: Graphitklassifizierung durch visuelle Auswertung. Beuth Verlag, Berlin, Germany, 2018.
- [3] ASTM A247-17. Standard test method for evaluating the microstructure of graphite in iron castings, 2017.
- [4] G. Hansen. IGF-Forschungsvorhaben 20650 N „DIAgraph II“ - Bericht zum Ringversuch 1: Subjektive Zuordnung für eine Vielzahl von einzelnen lamellaren, vermikularen und entarteten Graphitpartikeln und Bereichen. Technical Report 20650 N, IGF (Industrielle Gemeinschaftsforschung), Germany, 2020.
- [5] Alexandra Velichko. Quantitative 3d characterization of graphite morphologies in cast iron using fib microstructure tomography. 2008.
- [6] Simon N Lekakh, Xueliang Zhang, Wesley Tucker, Hyoung K Lee, Tara Selly, and James D Schiffbauer. Micro-ct quantitative evaluation of graphite nodules in sgi. *International Journal of Metalcasting*, 14:318–327, 2020.
- [7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.
- [8] Dominik Müller, Iñaki Soto-Rey, and Frank Kramer. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1):210, 2022.
- [9] Ernst Brunhuber. *Giesserei Lexikon*, Ausgabe 1988. Fachverlag Schiele & Schön, 1988.
- [10] Sergey Ioffe. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [11] Michael L. Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021.
- [12] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.
- [13] T-YLPG Ross and GKHP Dollár. Focal loss for dense object detection. In *proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2980–2988, 2017.
- [14] Raúl Gómez Bruballa. Understanding categorical cross-entropy loss, binary cross-entropy loss, softmax loss, logistic loss, focal loss and all those confusing names. [https://gombru.github.io/2018/05/23/cross\\_entropy\\_loss/](https://gombru.github.io/2018/05/23/cross_entropy_loss/), 2018. Accessed: 2024-11-27.



## Über uns:

Die Gesellschaft zur Förderung angewandter Informatik e. V. (GFai) ist eine industriennahe Forschungseinrichtung mit über 30 Jahren Erfahrung in der angewandten Informatik. Wir entwickeln praxisorientierte IT-Lösungen für Unternehmen und Institutionen, insbesondere in den Bereichen optische Mess- und Prüftechnik, Schall und Schwingungen, 3D-Datenverarbeitung, Robotik, Energie- und Umwelttechnik, Netzwerkinfrastruktur sowie wissensbasierte Dialogsysteme.

Neben der Forschung engagiert sich die GFai in der Ausbildung und Qualifizierung von Fachkräften. In Zusammenarbeit mit Hochschulen fördern wir den Wissenstransfer und begleiten Studierende sowie Nachwuchstalente auf ihrem Weg in die Praxis.

Zusammen mit ihren Tochterunternehmen gfai tech GmbH und Adalogic GmbH unterstützt die GFai den Transfer von Innovationen in den Mittelstand Deutschlands und setzt Impulse für digitale und nachhaltige Entwicklungen.

Weitere Informationen finden Sie unter [www.gfai.de](http://www.gfai.de).

## Kontaktdaten

### AI4EA:

E-Mail: [AI4EA@gfai.de](mailto:AI4EA@gfai.de)

Internet: [www.gfai.de/aktuelles/veranstaltungen/ai4ea-workshop/ai4ea](http://www.gfai.de/aktuelles/veranstaltungen/ai4ea-workshop/ai4ea)

### 3D-ISA:

E-Mail: [3D-ISA@gfai.de](mailto:3D-ISA@gfai.de)

Internet: [www.gfai.de/aktuelles/veranstaltungen/workshop-3d-nordost/workshop](http://www.gfai.de/aktuelles/veranstaltungen/workshop-3d-nordost/workshop)

## IMPRESSUM

### Herausgeber / Organisation:

F. Böhm, B. Hohnhäuser, M. Bauer

Gesellschaft zur Förderung angewandter Informatik e. V. (GFai)  
Volmerstraße 3, 12489 Berlin

Telefon: +49 30 814563-300

Telefax: +49 30 814563-302

eMail: [sekretariat@gfai.de](mailto:sekretariat@gfai.de)

Internet: [www.gfai.de](http://www.gfai.de)

### ISBN-Nummer:

978-3-942709-34-7