David Barton*, Felix Hess, Patrick Männle, Sven Odendahl, Marc Stautner, and Jürgen Fleischer

Image segmentation and robust edge detection for collision avoidance in machine tools

Bildsegmentierung und robuste Kantenerkennung zur Kollisionsvermeidung in Werkzeugmaschinen

Abstract: Collisions are a major cause of unplanned downtime in small series manufacturing with machine tools. Existing solutions based on geometric simulation do not cover collisions due to setup errors. Therefore a solution is developed to compare camera images of the setup with the simulation, thus detecting discrepancies. The comparison focuses on the product being manufactured (workpiece) and the fixture holding the workpiece, thus the first step consists in segmenting the corresponding region of interest in the image. Subsequently edge detection is applied to the image to extract the relevant contours. Additional processing steps in the spatial and frequency domain are used to alleviate effects of the harsh conditions in the machine, including swarf, fluids and sub-optimal illumination. The comparison of the processed images with the simulation will be presented in a future publication.

Keywords: Manufacturing, collision avoidance, frequency domain, semantic segmentation, deep learning.

Zusammenfassung: Kollisionen sind ein wesentlicher Grund für ungeplante Ausfallzeiten in der Kleinserienfertigung mit Werkzeugmaschinen. Bestehende Lösungen, die auf einer geometrischen Simulation basieren, können Kollisionen aufgrund von Rüstfehlern nicht vermeiden. Deshalb wird eine Lösung entwickelt, um Kamerabilder der Aufspannsituation mit der Simulation zu vergleichen. Für den Vergleich werden das zu bearbeitende Material (Werkstück) und die Vorrichtung zur Spannung des Werkstücks betrachtet. Daher ist der erste Schritt die Segmentierung des entsprechenden Bereichs im Bild. Danach werden die relevanten Konturen mithilfe einer Kantenerkennung extrahiert. Zusätzliche Bearbeitungsschritte im Frequenz- und Bildbereich werden eingesetzt, um die Auswirkungen der rauen Umgebungsbedingungen in der Werkzeugmaschine zu kompensieren, insbesondere Späne, Medien und ungünstige Beleuchtung. Der Abgleich der verarbeiteten Bildern mit der Simulation wird in einer künftigen Veröffentlichung beschrieben.

Schlagwörter: Fertigung, Kollisionsvermeidung, Frequenzbereich, Semantische Segmentierung, Deep Learning.

*Corresponding author: David Barton, Karlsruhe Institute of Technology, wbk Institute of Production Science, Kaiserstraße 12, 76131 Karlsruhe, Germany, e-mail: david.barton@kit.edu, ORCID: https://orcid.org/0000-0003-0287-3326
Felix Hess, Patrick Männle, Jürgen Fleischer, Karlsruhe Institute of Technology, wbk Institute of Production Science, Kaiserstraße 12, 76131 Karlsruhe, Germany
Sven Odendahl, Marc Stautner, Moduleworks GmbH, Henricistraße 50, 52072 Aachen, Germany

1 Introduction

In order to remain competitive in a global market, manufacturing is under pressure to continuously improve quality, costs and flexibility. There is a trend towards more variety in the final products, leading in turn to smaller batch sizes in production, including single-part production and mass customisation [1]. To be able to produce such parts in an economic way, it is necessary to optimise different stages of the production process. During the preparation phase, the planning and setup efforts need to be minimised, which relies heavily on experience: skilled workers know how to set up production for a new batch and experienced engineers are needed to safely plan the manufacturing process. The scarcity of these skills and the cost of training increase the need for support from digital solutions. New CAM (computer-aided manufacturing) software concepts help to detect problems during process planning, but show deficits when used in an Industry 4.0 environment [2]. The concept of a Digital Twin connects digital process planning with information retrieved via simulation or sensor measurements of the real process [3].

Another approach is to optimise the process on the machine level, for example by reducing downtime through condition-based maintenance and process monitoring [4, 5]. One major cause of downtime are collisions between

machine parts and production equipment, especially in small series manufacturing [5]. When correctly used, a Digital Twin allows to use advanced CAM algorithms to already avoid some collisions during the planning phase [6]. Other types of collisions cannot be detected in advance due to the potential for human error during the frequent and highly manual operation of setting up fixtures and the raw material for the part being manufactured (workpiece) [7]. This can be prevented by using a collision avoidance system. To apply such a system, all geometric features need to be modelled correctly by hand or via importing machine geometries and additional elements through given process-planning data. The positioning of these elements must be precise to ensure that collision checking and avoidance algorithms work correctly. This is especially the case for the workpiece, fixtures, and other supporting elements, as their geometry or position within the machine can change after the planning stage.

To overcome these problems, the authors propose a concept for collision avoidance consisting of a combination of geometric simulation and sensor-based inspection, thus avoiding collisions caused by discrepancies between simulation and real contents of the work area. The present contribution gives an overview of the state of the art in the fields of collision avoidance and image segmentation, then presents the authors' overall approach followed by a more detailed description of the pre-processing steps applied to images of the work area. These steps comprise image segmentation and edge detection, the result is a pre-processed edge image ready for comparison with the simulation.

2 State of the art

2.1 Collision avoidance

Existing solutions for collision avoidance in machine tools can be divided into the following categories: collision check during process planning, simulation-based dynamic collision avoidance, camera-based monitoring, and monitoring based on distance measurement. Collision check during process planning is a widespread and commercially available approach, in which a geometric model of machine tool, fixture, workpiece, and cutting tool is used to simulate and verify the planned machining steps [8]. In dynamic collision avoidance, a similar geometric model is used to check for collisions (Figure 2), however it is integrated in an in-time simulation running during machine operation based on real-time and look-ahead data from the machine control unit [9]. These sys-
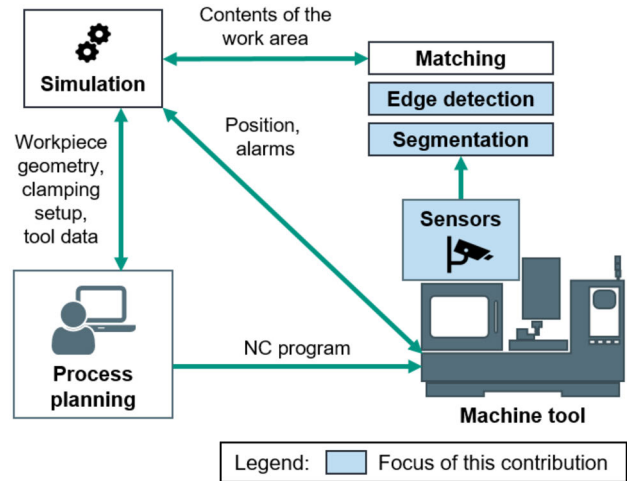

Figure 1: System architecture.

tems come in two flavours. They either check only for collisions between moving but constant geometries, or they also consider the changing geometry of the workpiece by simulating the material removal in real time as well.

Camera-based monitoring approaches aim to detect discrepancies between the real contents of the machine's work area and a reference geometry (either the geometry used to check the program during process planning, or the situation when previously manufacturing identical parts). Existing solutions for camera-based monitoring either overlay images from the geometric simulation and the real situation in the machine, and rely on a visual check by the operator [10], or rely on reference images from previous parts of the same type [11]. Monitoring based on distance measurement relies on laser triangulation, ultrasound, or inductive sensors to check the distance between moving parts of the machine (e. g. the main spindle) and obstacles such as the fixture and workpiece [12], however the position and number of sensors is limited due to high costs and limited mounting space.

Another approach to reducing costs due to collisions is collision detection. Acceleration, force or motor current signals are used to detect impacts and unexpectedly high loads, following which the movement of the feed axes is stopped as quickly as possible. This can limit the resulting damage to the machine, thus reducing repair costs and downtime [13]. The approaches described above either require a visual check by the operator, don't cover errors in setting up fixture and workpiece, require significant effort for sensor integration, require reference images from previous manufacturing of identical parts, or aren't able to entirely avoid collisions.
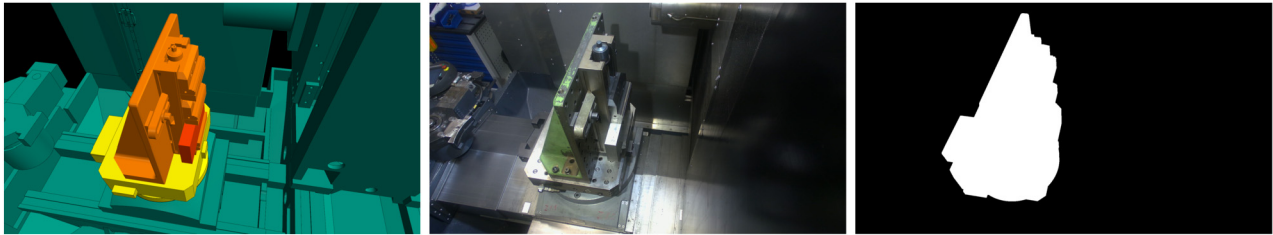
Figure 2: Simulation image (left), undistorted camera image (middle) and binary mask (right).

2.2 Semantic segmentation with deep learning

In recent years deep learning models for semantic segmentation have outperformed classic image segmentation approaches with the help of increasing computational resources [14]. The performance is measured with segmentation challenges on publicly available data sets such as PASCAL VOC 2012 [15] or MS COCO [16]. These data sets generally do not contain image scenes from industrial applications. The models most commonly used belong to the family of deep convolutional neural networks (DCNN). The advantages of deep learning models in computer vision are that these models can learn more complex features and often require less expert knowledge in comparison to traditional computer vision techniques [17]. U-Net is a model architecture with an encoder-decoder structure that has a high performance on segmentation tasks [18]. The U-Net model architecture with minor modifications is for example used to segment spatter in a laser welding process [19] and to segment an electrical motor and its components [20].

3 General approach

The present approach aims to combine the advantages of simulation and sensor-based approaches in a cost-effective solution for collision avoidance focussing on small-series and single-part manufacturing. In this context, it is especially relevant to ensure the first produced part is a good part, with minimal effort for setting up, running-in and human supervision. The combined system aims to detect mistakes in setting up or in the geometry model as well as discrepancies occurring during manufacturing (e. g. different workpiece shape due to a broken or wrong tool in a previous step, displaced workpiece due to inadequate clamping). The geometry of machine, fixture, workpiece and tool are modelled in the simulation-based collision avoidance system ModuleWorks CAS, and

the model is updated during machine operation based on data from the machine control unit and a material removal simulation [9]. The data obtained from the control unit comprises all the information necessary to simulate the current and the future state of the machine within a certain time span. Besides pure axis data, this also includes information on states in which the tool should not be allowed to cut material (e. g. during rapid movement or jog movements). If a future collision is detected based on the look-ahead data in the simulation component during automated movement, an alarm is sent to the machine to enable the feed axes to be stopped in time. During manual jog movement, the feed is controlled in such a way that the machine slowly approaches a future collision situation and finally stops before the contact occurs.

In order for CAS to work properly, the setup in the machine needs to be correct at all times. The workpiece in particular must be placed with a high accuracy to ensure safe process conditions. The level of accuracy required depends on the machining process, ranging from below one mm to orders of magnitude smaller. The lower end of this range cannot be checked with contactless sensor data alone within a cost-effective solution. For the placement of fixtures and the workpiece, the system therefore provides the possibility to position objects to a work offset measured by a probing process, which is usually also required to set up the machining process itself. However, the probing process is also prone to collisions because the initial position of the objects still has to be entered manually or based on information from the CAM project. At this stage, but also during the machining process itself, CAS is enhanced by a continuous sensor-based validation of the modelled situation. To accomplish this, the simulation component of the collision avoidance system periodically transmits an image of the current geometric model to a separate software system, which is tasked with matching the geometry from the simulation with sensor data acquired in the machine's work area (Figure 2). If a discrepancy is detected by the matching algorithm, an alarm is sent to the machine
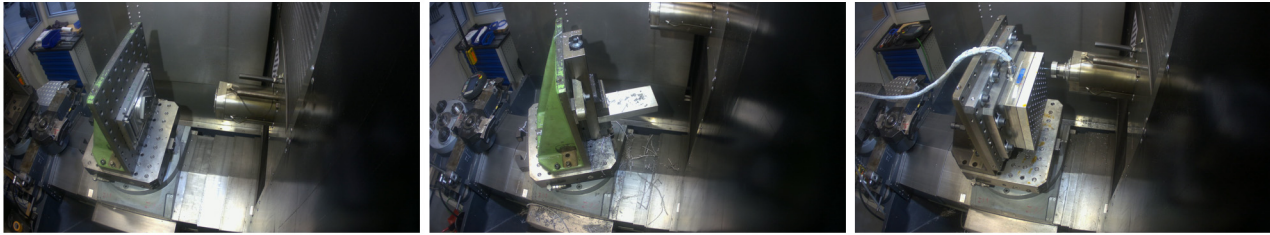
Figure 3: Three different combinations of workpiece and fixture.

control unit. An overview of the resulting system architecture is shown in Figure 1. The approach is tested in the machining centre DMC 60H, though care is taken to develop a solution that is applicable to a wide range of machines. The following section is dedicated to the image processing steps used to prepare sensor data for the matching algorithm.

4 Image processing

In the first prototypical implementation of the concept, a single camera[1] with a resolution of 1920x1080 pixels is used to observe the machine setup. Additionally a time-of-flight camera[2] with a resolution of 352x264 pixels is installed next to the other camera. Simulation-based collision avoidance typically allows for a safety clearance of 3 mm between bodies in the geometric model. In order to detect all critical discrepancies, the measurement and matching in this approach aims to detect deviations of 1 mm or more from the simulated geometry. If required due to the manufacturing process, smaller deviations could then be handled by probing. Damage during probing can be avoided thanks to the previous matching based on camera images.

The image processing consists of two steps, the following sections each describe one of these steps. First the region of interest (ROI) in the image is identified and segmented from the original image. Then the contours within this region are extracted, thus delivering the input required for a matching algorithm. The matching algorithm, aiming to compare the contours of workpiece and fixture with the geometric simulation, is still under development.

4.1 Object segmentation

The goal of semantic segmentation in general is to assign a class label to each pixel of an image. Here in particular the goal is to assign the class label 1 to the ROI and the class label 0 to the background. The difficulty here is to achieve good segmentation results for various combinations of workpiece and fixture, including new unseen combinations in the future (see Figure 3). Furthermore, the segmentation needs to be robust against disturbing factors occurring in the working area, and to be applicable to other machine tools with little integration effort.

We propose a deep learning model based on the model architecture mentioned in section 2.2. Images for training are captured with the camera setup described above and automatically labelled using the existing geometric simulation of the working area. The performance of the deep learning model will be compared to a baseline method using an additional time-of-flight camera and traditional computer vision techniques.

In the baseline method, the ROI is detected in the low-resolution depth image from the time-of-flight camera. The detection algorithm consists in applying a threshold to the distance between the current depth image and a background depth map containing the maximum distance captured at this spatial location. The image points of the largest interconnected region are then mapped to the corresponding image points in the high-resolution camera creating a sparse mask. The geometric relationship between the views is estimated with a stereo calibration process, the resulting sparse mask is densified with morphological operations. A general problem with this approach is the spindle occluding the ROI. We use the information about the spindle location and appearance from the geometric simulation. With this information, the image points corresponding to the spindle are subtracted from the ROI.

---

1 Datasheet available at https://rotoclear.com/c2.html
2 Datasheet available at https://www.ifm.com/de/en/product/O3D303

4.1.1 Data generation

For the pixelwise binary classification problem at hand a binary mask containing the ROI needs to be created. Typically these masks are created manually by drawing a polygon around the ROI. Algorithm-assisted labelling tools exist too, but these are still time consuming. Therefore the goal is to generate the binary masks automatically, requiring the human operator only to check the result.

To use the image provided by the geometric simulation three conditions need to be fulfilled:

1. The machine setup needs to be correctly modelled in the simulation.
2. The image from the geometric simulation needs to be generated from the camera's location and view direction. The camera's field of view also needs to be taken into account.
3. The image produced by the camera needs to contain as little distortion as possible.

Condition 1 requires a human operator to check the setup and supervise the machine during data generation. To fulfil the conditions 2 and 3, the extrinsic and intrinsic matrix and the distortion parameters of the camera need to be estimated. This is achieved by calibrating the camera based on multiple images of a chessboard pattern in different locations of the working area and different orientations relative to the camera, using an algorithm based on [21].

The location and view direction of the camera in the geometric simulation are derived from the extrinsic matrix of a calibration image with a known position in the machine coordinate system. The resulting simulation image is segmented via thresholding in the HSV colour space after assigning distinct colours to the ROI and the background, thus generating a mask for the fixture and workpiece. Additionally, the distortion parameters and intrinsic matrix are used to estimate an undistorted camera image. The simulation image, the undistorted camera image and the resulting mask are shown in Figure 2. The accuracy of the calibration was verified by overlaying corresponding camera and simulation images.

4.1.2 Segmentation model

The segmentation model is based on the U-Net architecture mentioned in section 2.2. The input feature map x is the undistorted camera image and the output is a feature map $\hat{y}$ of the same size predicting the class label for each pixel $x_{i,j}$. The binary mask containing the ROI is created by applying a threshold (e. g. $\hat{y}_{i,j} > 0.5$) to the output feature map.

The classic U-Net consists of an encoder-decoder structure with five stages. The encoder stages downsample the feature maps with successive convolution and max pooling layers. The decoder stages upsample the feature
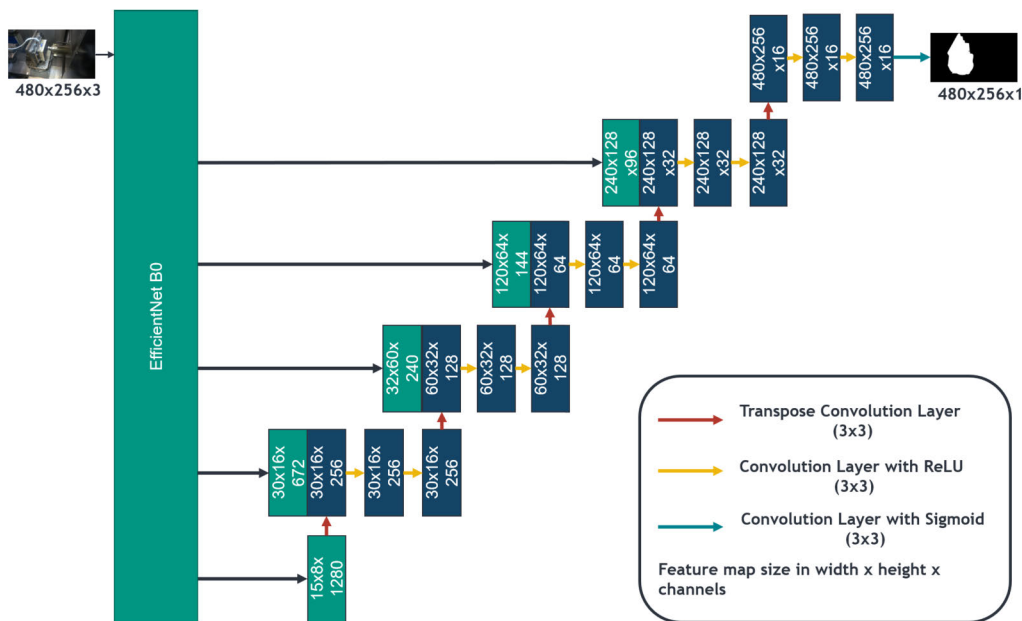

Figure 4: Model architecture of best performing model with EfficientNet backbone: Rectangles represent the feature maps, arrows the connections and layers between the feature maps. For regularisation L2 weight regularisation was used.

maps with the help of transpose convolution and normal convolution layers. Furthermore, skip connections concatenating feature maps from each encoder stage to the corresponding decoder stage feature maps are used. We adapt the classic U-Net to increase its model efficiency, thus reducing the training time and the resources needed for inference. The adaption consists in replacing the encoder stages with the image classification network EfficientNet [22]. EfficientNet is designed to produce good classification results with a low number of model parameters. The compound scaling method of this model family enables the model capacity of the encoder to be increased easily if the model is underfitting, without becoming less efficient. Moreover, pretrained weights for EfficientNet on ImageNet [23] are available. These weights can be used as a feature extractor or can be fine-tuned in combination with the decoder structure. Furthermore, we test the use of a shallower model containing less encoder and decoder stages since the problem at hand is only a binary classification problem. A weighted focal loss function is used [24]. The focal loss introduces an additional modulating factor $(1 - p_t)^\gamma$ in comparison to the standard binary cross entropy loss function. The modulating factor decreases the loss function for easily classified pixels and consequently increases the contribution of hard examples to the overall loss. The focal loss for the predicted probability p is defined as follows. The ground truth is denoted as y.

$$L_{fl}(p, y) = \begin{cases} -\alpha \cdot (1 - p)^\gamma \cdot \log(p), & \text{if } y = 1 \\ -(1 - \alpha) \cdot p^\gamma \cdot \log(1 - p), & \text{otherwise} \end{cases} \quad (1)$$

To evaluate the increased model efficiency of the adapted model architecture a comparison with the classic U-Net architecture was performed. The implementation of Efficient U-Net was based on [25]. Both architectures were trained from scratch with the same data set. The effect of using pretrained encoder weights was assessed for the Efficient U-Net. To find appropriate hyperparameters a random search with 15 trials and 35 epochs was used. The best model from the random search was trained for 25 additional epochs.

Furthermore, the effect of pretraining the decoder on a publicly available data set was explored. Here we used a subset of Pascal VOC 2012 excluding the object categories with living creatures. The model was trained on Pascal VOC for 100 epochs and appropriate hyperparameters were found with a random search. The pre-trained weights were fine-tuned on our data set in two phases. First, the top decoder stage was unfrozen and trained to convergence with a learning rate of 0.01. Second, the whole decoder was unfrozen and trained for 10 epochs with a reduced learning rate of 10^{-5}.

Our data set comprises 866 data points (camera image and mask). The mean ratio of ROI to background image is 16,1 %. Around half of the data points are manually labelled. The data set includes images with different lighting conditions, workpieces and fixtures. In a subset of the images, disturbing factors such as swarf, oil and cooling liquid are present. The data set is split into subsets for training (567 data points), validation (189 data points) and testing (110 data points). The images and masks are
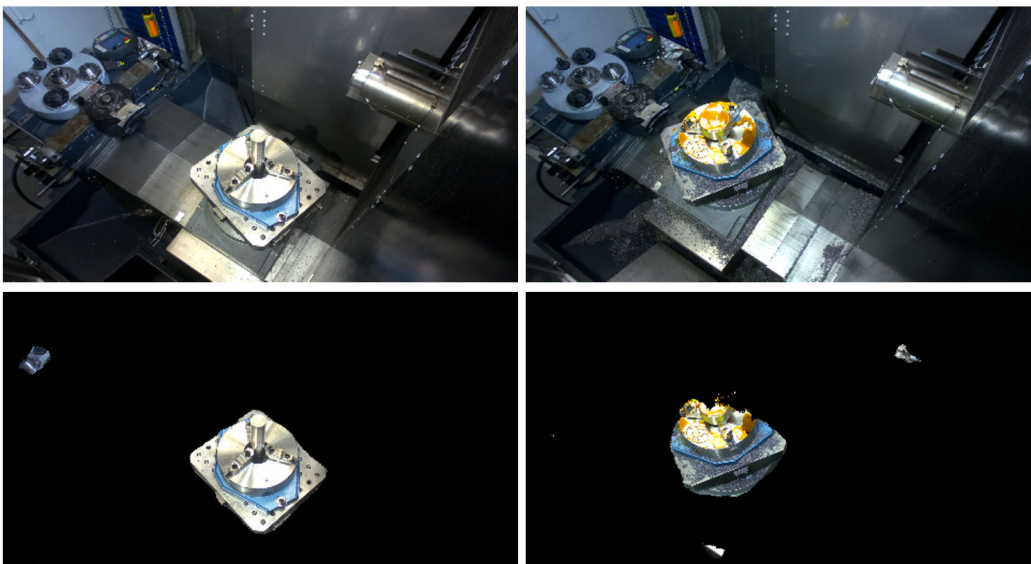

Figure 5: Segmentation results (bottom) on test data (top) with a new fixture: Examples with a good segmentation result (left) and a larger error (right).

downscaled to 480x256 pixels. The following augmentation methods are applied randomly to the training data before each epoch:

- Horizontal flip
- Affine transformation with shifting, scaling and rotation
- Perspective transformation
- Blur image, add Gaussian noise, sharpen image or improve contrast of the image
- Change contrast, brightness and hue, saturation and value
- Gamma correction

The performance of the segmentation model on the test data set can be seen in Table 1a. In addition, new data was recorded with a previously unseen fixture, new workpieces and an adapted camera position (see Table 1b). Figure 5 includes examples of new images segmented by the best performing model, the architecture of this model is outlined in Figure 4. The results on the new data suggest that the model is not yet able to generalise sufficiently to entirely new types of fixtures, which may be due to the limited variety of fixtures in the training data. Pretrained weights in the encoder lead to a significant improvement in performance on the second data set. Additionally pretraining the decoder weights further improves the performance.

Table 1: Mean Intersection over Union (IoU) and mean F1-Score of the model architectures on the different test data sets.

(a) Test data with known fixtures.

	Model Size	IoU	F1-Score
U-Net	34513864	0.9320	0.9646
Efficient U-Net	5223880	0.9358	0.9666
Eff. U-Net, encoder pretrained on ImageNet	10898692	0.9328	0.9650
Eff. U-Net fully pretrained on Pascal VOC	10898692	0.9324	0.9648

(b) Test data with a previously unseen fixture.

	Model Size	IoU	F1-Score
U-Net	34513864	0.3405	0.5053
Efficient U-Net	5223880	0.5452	0.6950
Eff. U-Net, encoder pretrained on ImageNet	10898692	0.7700	0.8674
Eff. U-Net fully pretrained on Pascal VOC	10898692	0.7957	0.8831

Finally the performance of the deep learning model is compared to the baseline method (segmentation based on thresholds in a time-of-flight image). The segmentation results of the deep learning model are post-processed with morphological operations and largest blob detection. The results (Table 2) indicate a slightly better performance with the baseline method than with the deep learning method. In future work the deep learning model will be trained with additional data. The generalisation of the segmentation model will be tested on additional combinations of workpiece and fixture as well as other machine tools. For deployment in an industrial environment, the model was implemented on the embedded system Jetson AGX Xavier.

Table 2: Mean Intersection over Union (IoU) and mean F1-Score of the deep learning model and the baseline method.

	IoU	F1-Score
Deep Learning Model	0.8009	0.8865
Baseline Method	0.8343	0.9101

4.2 Contour extraction

The contours of fixture, workpiece and other obstacles must be detected in a sufficient quality for a subsequent comparison with data from the geometric simulation. We use Canny edge detection as the core algorithm to detect the relevant contours. To minimise the detection of false edges due to shadows and reflections on metallic surfaces, images captured with different lighting conditions can be combined [26].

The conditions in machine tools lead to challenges due to obstruction by swarf (metal chips resulting from the cutting process, ranging from small particles to long tendrils), fluids (oil and coolant), and sub-optimal lighting conditions. Figure 3 includes an example of a workpiece partially covered by swarf, while the image on the right in Figure 5 shows a situation with oil and swarf. For each of these challenges, suitable image processing methods are evaluated using images acquired in the machining centre DMC 60H.

4.2.1 Spatial domain

The present approach uses processing in the spatial domain to detect the contours of fixtures and workpieces through Canny edge detection, and to compensate for the influence of lighting and fluids. As no object segmentation model has been finally implemented for this application, the ROI was selected manually when developing these methods.
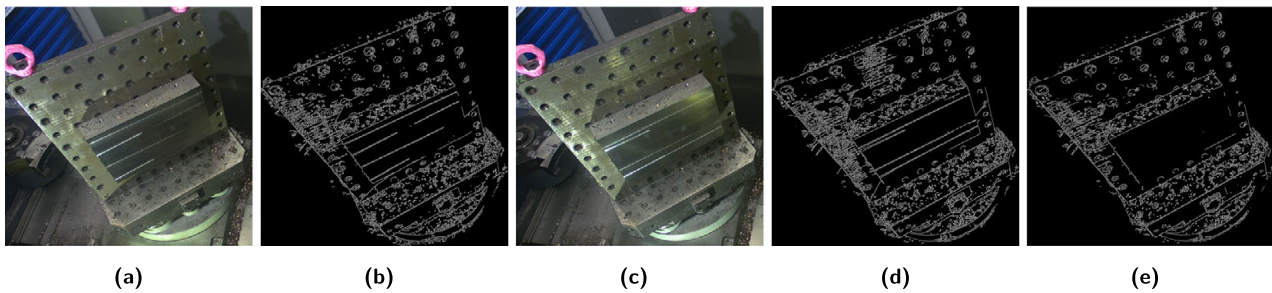
Figure 6: Removal of additional edges due to lighting. Images taken while varying the lighting (a, c) display additional edges due to the lighting conditions (b, d). These are identified and removed by comparing the images, thus leading to the improved image (e).
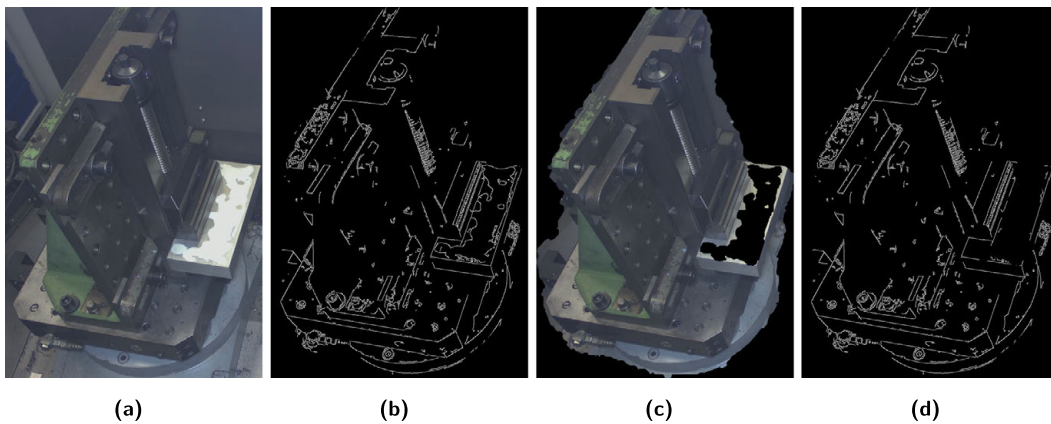

Figure 7: Removal of additional edges due to coolant. (a) Original image with coolant; (b) Edges detected in original image; (c) Coolant identified and marked in black; (d) Image after removal of edges due to coolant.

The conditions for image acquisition in machine tools can be improved by adding light sources, however the structure of machine tools and the presence of reflecting metallic surfaces mean undesired artefacts due to reflection and shadows remain frequent. Two light sources are used to successively illuminate the scene from different angles. In the resulting images, artefacts linked to the illumination appear in different positions. This effect is used by removing edges that do not appear in the same position in both images (within a tolerance of one pixel). The result is shown in Figure 6.

Coolant and cutting oil are frequently used to lubricate and cool machining processes. These may cover patches of the workpiece or fixture, thus causing additional edges in the captured images and hampering the detection of contours. The present approach uses the following steps to identify such additional edges:

– Bilateral filter
– Segmentation based on thresholding of pixel colour to identify coolant
– Adding similarly coloured neighbouring pixels to the segment
– Dilation of the identified segment

The original image is subjected to Canny edge detection, then all edges within the identified segment are removed. An example for this procedure is shown in Figure 7.

4.2.2 Frequency domain

Additional image processing is performed in the frequency domain, with the aim of removing edges due to swarf and other causes such as scratches, chipped paint, and corrosion. These undesirable features are linked to randomly oriented edges and high spatial frequencies (Figure 8).

After applying the 2-dimensional discrete Fourier transform (2D DFT) to the original image, the logarithmically scaled amplitude spectrum is subjected to a filter mask. After inverse 2D DFT, Canny edge detection is performed on the filtered image. The filter mask aims to select the dominant directions in an image and eliminate high frequencies, it is generated automatically for each image.

The dominant directions in the image appear as lines in the amplitude spectrum. The spectrum is binarised based on a threshold k, then the number of white pixels is counted for each line passing through the centre of the
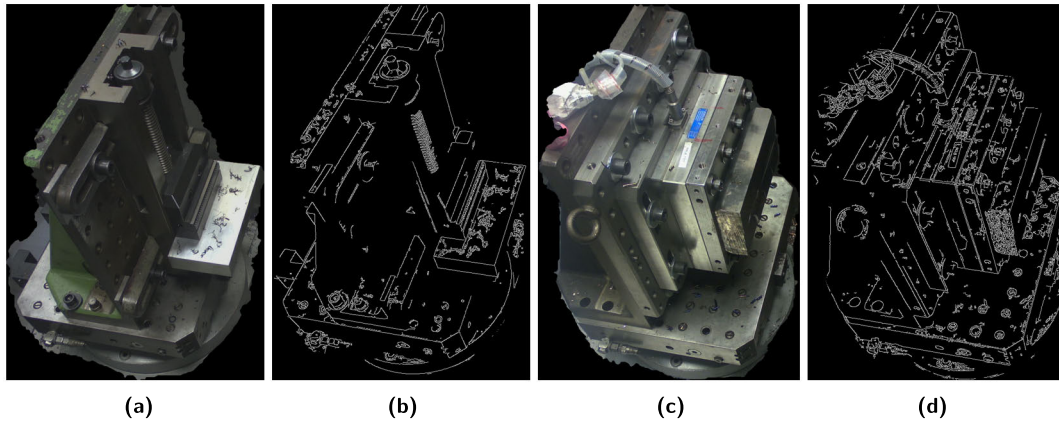
Figure 8: Examples of original images containing swarf and edge images without filtering.
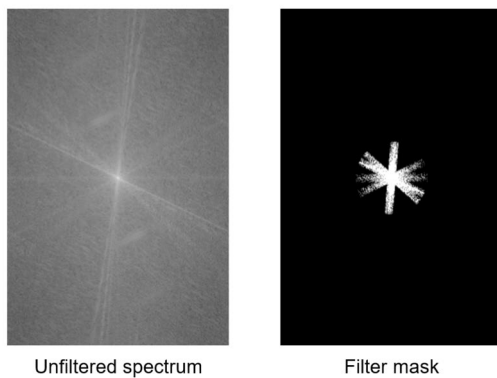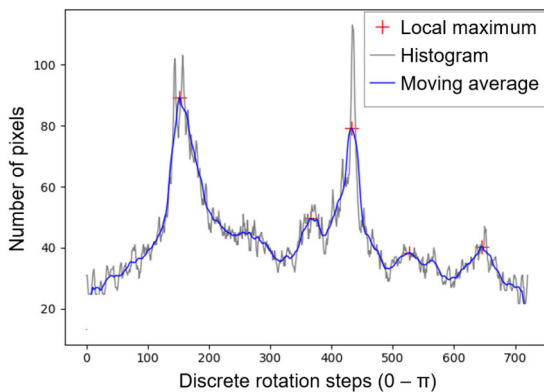

Figure 9: Selection of dominant directions in the amplitude spectrum, applied to Figure 8a.

image, thus creating a histogram of directions. This histogram is smoothed by applying a moving average, then local maxima with a prominence of at least p are determined (Figure 9). The filter mask for dominant directions is the union of the following:

- Stripes with a width of b around each of the identified dominant directions,
- A disc with a radius of r_1 in the centre of the image.

The complete filter mask is the intersection of the above with a low pass filter (with a radius of r_2). Figure 10 shows the resulting images after filtering, inverse DFT and Canny edge detection for the examples from Figure 8.

The parameters k, p, $r1$, $r2$, b and the parameters for Canny edge detection are determined manually based on a representative selection of images, whereas the automatically generated filter mask adapts to scenes with different orientations.

5 Summary and future work

A concept was developed for a collision avoidance system covering a larger range of collision causes than existing solutions and especially well-suited to small series and single part manufacturing. The proposed system runs during the operation of a machine tool and combines a state-of-the-art geometric simulation with a camera-based inspection of the work area. The encouraging initial results presented in this contribution concern the processing of images acquired in the harsh conditions of a machine tool's work area. First, a method to detect the object based on a deep learning model using pre-trained weights was presented. The deep learning model was compared to a method requiring an additional depth camera. The method requiring depth data performed better, the authors aim to improve the deep learning model in future work with a larger set of training data. Second, a method to detect the contours in the segmented image region was introduced, combining Canny edge detection with additional filter techniques in the spatial and frequency domains. Further work is needed to develop a matching algorithm to compare the detected contours resulting from these pro-
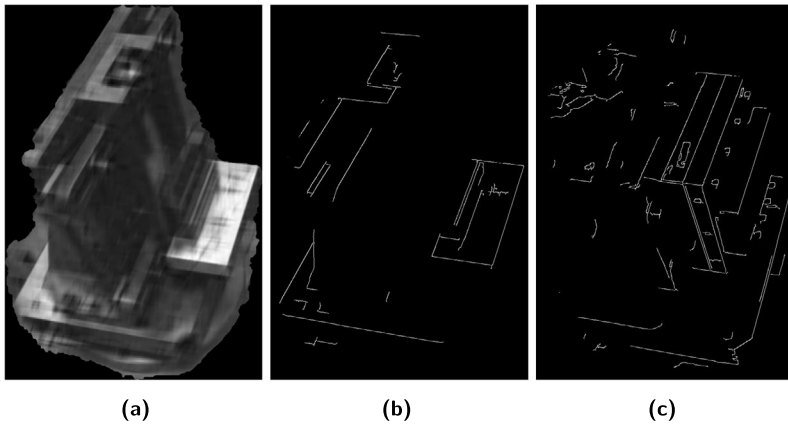
Figure 10: Results after filtering. (a) Fig. 8a after filtering in frequency domain and inverse transformation; (b) Edges detected in Fig. 10a; (c) Edges detected after filtering of Fig. 8c.

cessing steps with the simulation. The authors also plan to extend the concept in order to adjust the simulation model and tool path to the measured reality of the working area.

References

1. F. Piller and A. Kumar, Mass customization: Providing custom products and services with mass production efficiency, *Journal of Financial Transformation*, vol. 18, no. 3, pp. 125–131, 2006.

2. E.-M. Jakobs, C. Digmayer, S. Vogelsang and M. Servos, Not ready for industry 4.0: Usability of CAx systems, in: *Advances in Usability and User Experience*, 2018, pp. 51–62.

3. B. Schleich, M.-A. Dittrich, T. Clausmeyer, R. Damgrave, J. A. Erkoyuncu, B. Haefner, J. de Lange, D. Plakhotnik, W. Scheidel and T. Wuest, Shifting value stream patterns along the product lifecycle with digital twins, *Procedia CIRP*, vol. 86, pp. 3–11, 2019.

4. D. Barton, P. Gönnheimer, F. Schade, C. Ehrmann, J. Becker and J. Fleischer, Modular smart controller for Industry 4.0 functions in machine tools, *Procedia CIRP*, vol. 81, pp. 1331–1336, 2019.

5. D. Barton, R. Stamm, S. Mergler, C. Bardenhagen and J. Fleischer, Industrie 4.0 Nachrüstkit für Werkzeugmaschinen: Modulare Lösung für zustandsorientierte Instandhaltung und Prozessüberwachung, *WT Werkstattstechnik*, vol. 110, no. 7-8, 2020.

6. D. Plakhotnik, L. Glasmacher, T. Vaneker, Y. Smetanin, M. Stautner, Y. Murtezaoglu and F. van Houten, CAM planning for multi-axis laser additive manufacturing considering collisions, *CIRP Annals*, vol. 68, no. 1, pp. 447–450, 2019, doi: https://doi.org/10.1016/j.cirp.2019.04.007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0007850619300332.

7. E. Abele, C. Brecher, S. C. Gsell, A. Hassis and D. Korff, Steps towards a protection system for machine tool main spindles against crash-caused damages, *Production Engineering*, vol. 6 no. 6, pp. 631–642, 2012, issn: 0944-6524. doi: 10.1007/s11740-012-0422-6.

8. P. Hehenberger, Computerunterstützte Fertigung, 2011, doi: 10.1007/978-3-642-13475-3.

9. M. Armendia, T. Fuertjes, D. Plakhotnik, J. Sossenheimer and D. Flum, Cyber-physical system to improve machining process performance, in *Twin-Control*, M. Armendia, M. Ghassempouri, E. Ozturk and F. Peysson, Eds., vol. 58, Cham: Springer International Publishing, 2019, pp. 197–208. isbn: 978-3-030-02202-0. doi: 10.1007/978-3-030-02203-7_12.

10. P. Sommer, ed., *iWindow - Intelligentes Maschinenfenster: Abschlussbericht Verbundforschungsprojekt*, Fortschritt-Berichte VDI Reihe 2, Fertigungstechnik vol. 697, Düsseldorf: VDI Verlag, 2018, isbn: 978-3-18-369702-1.

11. Heidenhain GmbH, *Optionen und Zubehör für TNC-Steuerungen*, 2019.

12. T. Li, Y. Wang, K. Liu, H. Liu, J. Zhang, X. Sheng and D. Guo, Virtual grid and BPNN based collision avoidance control of automatic fixture system, *The International Journal of Advanced Manufacturing Technology*, vol. 95, no. 5-8, pp. 2843–2853, 2018, issn: 0268-3768. doi: 10.1007/s00170-017-1432-5.

13. T. Rudolf, C. Brecher and F. Possel-Dölken, Contact-based collision detection—a new approach to avoid hard collisions in machine tools, in: *International Conference on Smart Machining Systems*, 2007.

14. A. Garcia-Garcia, S. Orts-Escolano, S O. Oprea, V. Villena-Martinez and J. Garcia-Rodriguez, A Review on Deep Learning Techniques Applied to Semantic Segmentation, *CoRR*, 2017. arXiv:1704.06857v1.

15. M. Everingham, S. M Eslami, L. Van Gool, C. K. Williams, J. Winn and A. Zisserman, The Pascal Visual Object Classes Challenge: A Retrospective, *International Journal of Computer Vision*, vol. 111, no. 1, pp. 98–136, 2014, issn: 15731405. doi: 10.1007/s11263-014-0733-5.

16. T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár and C. L. Zitnick, Microsoft COCO: Common objects in context, PART 5, vol. 8693 LNCS, 2014, pp. 740–755. doi: 10.1007/978-3-319-10602-1_48. [Online]. Available: https://cocodataset.org/#home.

17. N. O' Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan and J. Walsh, Deep learning vs. traditional computer vision, *CoRR*, 2019.

18. O. Ronneberger, P. Fischer and T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, Springer Verlag, 2015, pp. 234–241, isbn: 9783319245737. doi: 10.1007/978-3-319-24574-4_28.

19. J. Hartung, A. Jahn, M. Stambke, O. Wehner, R. Thieringer and M. Heizmann, Camera-based spatter dection in laser welding with deep learning approach, in: *Forum Bildverarbeitung 2020*, T. Längle and M. Heizmann, Eds., KIT Scientific Publishing, 2020, pp. 317–328, isbn: 978-3-7315-1053-6. doi: 10.5445/KSP/1000124383.

20. N. Mitsche and M. Heizmann, Semantische Segmentierung von Ankerkomponenten von Elektromotoren, German, in: *Forum Bildverarbeitung 2020*, T. Längle and M. Heizmann, Eds., KIT Scientific Publishing, 2020, pp. 328–340, isbn: 978-3-7315-1053-6. doi: 10.5445/KSP/1000124383.

21. Z. Zhang, A flexible new technique for camera calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22 no. 11, pp. 1330–1334, 2000, issn: 01628828. doi: 10.1109/34.888718.

22. M. Tan and Q. V. Le, EfficientNet: Rethinking model scaling for convolutional neural networks, in: *36th International Conference on Machine Learning, ICML 2019*, vol. 2019-June, 2019, pp. 10691–10700, isbn: 9781510886988. arXiv:1905.11946.

23. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg and L. Fei-Fei, ImageNet Large Scale Visual Recognition Challenge, *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015, issn: 15731405. doi: 10.1007/s11263-015-0816-y. arXiv:1409.0575.

24. T.-Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, Focal Loss for Dense Object Detection, (2017). arXiv:1708.02002. [Online]. Available: https://github.com/facebookresearch/Detectron. http://arxiv.org/abs/1708.02002.

25. P. Yakubovskiy, *Segmentation Models*, https://github.com/qubvel/segmentation_models, 2019.

26. T. Pfeifer and L. Wiegers, Reliable tool wear monitoring by optimized image and illumination control in machine vision, *Measurement: Journal of the International Measurement Confederation*, vol. 28, no. 3, pp. 209–218, 2000, issn: 02632241. doi: 10.1016/S0263-2241(00)00014-2.

Bionotes

David Barton
Karlsruhe Institute of Technology, wbk
Institute of Production Science,
Kaiserstraße 12, 76131 Karlsruhe, Germany
david.barton@kit.edu

David Barton is a research associate in the Machine Tools and Mechatronics group since 2016. His research interests include vibrations in machining systems and retrofitting monitoring solutions to existing machines.

Felix Hess
Karlsruhe Institute of Technology, wbk Institute of Production Science, Kaiserstraße 12, 76131 Karlsruhe, Germany

Felix Hess is pursuing a master's degree in mechatronics at KIT. Currently he is writing his master thesis at wbk on automatic object segmentation for image-based collision avoidance in machine tools.

Patrick Männle
Karlsruhe Institute of Technology, wbk Institute of Production Science, Kaiserstraße 12, 76131 Karlsruhe, Germany

Patrick Männle is pursuing a master's degree in mechatronics at KIT. He wrote his bachelor thesis at wbk on the compensation of confounding factors for image-based collision avoidance in machine tools.

Sven Odendahl
Moduleworks GmbH, Henricistraße 50,
52072 Aachen, Germany
sven@moduleworks.com

Dr.-Ing. Sven Odendahl is development team leader of the CNC Collision Avoidance Team at ModuleWorks GmbH. His current research focuses on algorithms for real-time collision avoidance systems and topics facilitating the setup and usability of Digital-Twin-related solutions for machine builders and end customers in real-world applications.

Marc Stautner
Moduleworks GmbH, Henricistraße 50,
52072 Aachen, Germany
marc@moduleworks.com

Jürgen Fleischer
Karlsruhe Institute of Technology, wbk
Institute of Production Science,
Kaiserstraße 12, 76131 Karlsruhe, Germany
juergen.fleischer@kit.edu

Dr.-Ing. Marc Stautner is Research Director of ModuleWorks and Managing Director of AixPath GmbH, his research topics are heading to the digitalization of manufacturing process chains. The integration of digital planning, of toolpaths for subtractive and additive manufacturing and of simulations for manufacturing processes, into solutions for digital twins is his main development goal.

Prof. Dr.-Ing. Jürgen Fleischer is head of the wbk Institute of Production Sciences at KIT since 2003 and visiting professor at the Tongji University in Shanghai since 2012. His current research focuses on components for machine tools and handling systems, automated manufacturing systems for lightweight structures as well as the production of energy storing systems and traction drives for electric mobility.