High Resolution Mask R-CNN-based Damage Detection on Titanium Nitride Coated Milling Tools for Condition Monitoring by using a New Illumination Technique

Mühenad Bilal[1][a], Sunil Kancharana[1][b], Christian Mayer[1], Daniel Pfaller[1], Leonid Koval[1][c],
Markus Bregulla[1], Rafal Cupek[2][d] and Adam Ziębiński[2][e]

[1] *Technische Hochschule Ingolstadt, Esplanade 10, Ingolstadt 85057, Germany*

[2] *Silesian University of Technology, Institute of Informatics, Gliwice, Poland*

Keywords: Predictive Maintenance, Machine Learning, Damage Detection, Illumination Source, Mask R-CNN.

Abstract: The implementation of intelligent software in the manufacturing industry is a technology of growing importance and has highlighted the need for improvement in automatization, production, inspection, and quality assurance. An automated inspection system based on deep learning methods can help to enhance inspection and provide a consistent overview of the production line. Camera-based imaging systems are among the most widely used tools, replacing manual industrial quality control tasks. Moreover, an automatized damage detection system on milling tools can be employed in quality control during the coating process and to simplify measuring tool life. Deep Convolutional Neural Networks (DCNNs) are state-of-the-art methods used to extract visual features and classify objects. Hence, there is great interest in applying DCNN in damage detection and classification. However, training a DCNN model on Titanium-Nitride coated (TiN) milling tools is extremely challenging. Due to the coating, the optical properties such as reflection and light scattering on the milling tool surface make image capturing for computer vision tasks quite challenging. In addition to the reflection and scattering, the helical-shaped surface of the cutting tools creates shadows, preventing the neural network from efficient training and damage detection. Here, in the context of applying an automatized deep learning-based method to detect damages on coated milling tools for quality control, the light has been shed on a novel illumination technique that allows capturing high-quality images which makes efficient damage detection for condition monitoring and quality control reliable. The method is outlined along with results obtained in training a ResNet 50 and ResNet 101 model reaching an overall accuracy of 83% from a dataset containing bounding box annotated damages. For instance and semantic segmentation, the state-of-the-art framework Mask R-CNN is employed.

1 INTRODUCTION

Machining Process Monitoring(MPM) (Liang et al., 2004) plays an important role in reducing cost, ensuring greater product variability, and improving manufacturing productivity and reliability (Caggiano, 2018). Monitoring of the production process (Cupek et al., 2015), production variants (Cupek et al., 2018; Yli-Ojanperä et al., 2019) and other parameters such as current supply (Grzechca et al., 2017; Yingjie, 2014) and even speed of the engines are of growing importance to provide real-time data for manufacturers. Moreover, there is a vital demand for Tool Condition Monitoring (TCM) (Short and Twiddle, 2019), especially when it comes to evaluating the milling process regarding tool wear and the resultant surface roughness.

- Break-in
- Normal wear
- Abnormal wear

Coating can increase the durability of cutting tools by 10-12 times (Spišák and Majernikova, 2017). In

[a] https://orcid.org/0000-0003-4065-8467
[b] https://orcid.org/0000-0002-0718-6480
[c] https://orcid.org/0000-0003-4845-6579
[d] https://orcid.org/0000-0001-8479-5725
[e] https://orcid.org/0000-0003-4554-6667

Bilal, M., Kancharana, S., Mayer, C., Pfaller, D., Koval, L., Bregulla, M., Cupek, R. and Ziębiński, A.
High Resolution Mask R-CNN-based Damage Detection on Titanium Nitride Coated Milling Tools for Condition Monitoring by using a New Illumination Technique.
DOI: 10.5220/0010781800003124
In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 5: VISAPP*, pages 305-314
ISBN: 978-989-758-555-5; ISSN: 2184-4321

almost all micro-size industries, digital image processing techniques are used as a measurement of quality assurance (Chen and Lee, 2010). Using an Imaging System(IS) for damage detection on coated milling tools is accompanied by many difficulties such as high damage density, low contrast intensity, in-homogeneity, and damage shape variations. Also, weak boundaries and strong gradient on the tool contours that overlap with the damages can decrease the detection accuracy. Additionally, due to the complex geometry of the cutting tools, a deficient illumination uniformity results in large intensity variation of different image regions, making training a DCNN model insufficient for inspection applications. To overcome these difficulties, a new illumination technique to ensure uniform illumination for capturing high quality images for computer visions tasks such as object detection and semantic segmentation was developed. Additionally, using these images can improve and accelerate training DCNN models, increasing damage detection accuracy. To our best knowledge this is the first work, in which an object with optical critical properties is inserted into a Cylindrical Shaped Enclosure (CSE) to capture high quality images for object detection, instance segmentation and pixelwise damage detection tasks.

The object detection algorithms have been continuously improved by the computer vision community. Parts of this advanced technique have been driven by popular object detection algorithms like SSD (Liu et al., 2016), R-CNN (Girshick et al., 2014), Fast R-CNN (Girshick, 2015), Faster R-CNN (Ren et al., 2015) and YOLO (Redmon et al., 2016). For the automatic damage recognition and localization, the state-of-the-art target detection framework Mask-R-CNN was employed, which extends Faster R-CNN by adding a branch for predicting segmentation masks on each Region of Interest (RoI) and a branch for classification and bounding box simultaneously. The mask is a fully convolutional network that takes an image of arbitrary size as input and produces sized output with efficient inference and learning for each RoI, predicting a segmentation mask in a pixel-to-pixel manner by adding only a small fraction of computational overhead to Faster R-CNN. This enables a fast system and rapid experimentation.

In the first stage, Mask R-CNN uses a Region Proposal Net (RPN) network (Girshick et al., 2014) to generate a sparse set of rectangle proposals (Faster, 2015). Each proposal represents a RoI on the feature maps indicating whether there is a target or not. Using RoI-Pooling in the next step, the feature extraction of each proposal from a CNN feature map is performed. Finally, the two processing branches mentioned above classify the object and predict the masks. The mask prediction indicates whether the pixels lies in the predicted bounding box of the objects or not.

Additionally, the Faster R-CNN includes an Feature Pyramid Networks (FPN), which combines low-resolution, semantically strong features with high-resolution, semantically weak features via a top-down architecture with lateral connection to build an in-network feature pyramid from single scale input. This results in excellent gains in both accuracy and speed (Tsung-Yi Lin et al., 2017). Moreover, the FPN can enhance small damage detection below 30 μ by just using a standard commercial camera system.

This paper is focused on coated cutting tools, which are widely used in the milling industry. Due to the increased demand for all kinds of high precision and high accuracy cutting tools determining the wear or damages of cutting edges is of great importance (Schulz and Moriwaki, 1992). For this purpose, a DCNN based tool measuring and inspection system for determining the wear condition of the cutting edges and coating homogeneity will definitely support tool manufactures as well as machining process.

The main steps and aim of this paper can be summarized as follows:

1. Use of Cylindrical Shaped Enclosure (CSE) for capturing high-quality images to avoid unwanted reflection and ensure homogeneous illumination on the optical-critical components.
2. Preprocessing the data by cropping each image into 36 small image fractions to improve the performance of the DCNN model. In addition to this, the cropped images will support the FPN to detect small damages.
3. Due to the high image quality, few annotated images are used to train the model and perform high-accuracy damage detection.
4. Predicting damages on the cropped images and merging them to reconstruct the original image with the corresponding damages.
5. Fine tune the model by modifying the hyper parameter such as the area of anchor boxes with various backbones.

The rest of the paper is structured as follows: The measurement apparatus is described in section 2. Section 3 explains the method using Mask R-CNN algorithm followed by section 4 which includes experimental results and analysis. The paper is ended with conclusion as section 5.

2 IMAGE ACQUISITION OF MILLING TOOLS WITH CYLINDRICAL SHAPED ENCLOSURE MEASUREMENT SETUP

The proposed measurement setup (see Figure 1) has been filed for an (EU) European patent. It consists of a cylindrical shaped enclosure (CSE) whose inner walls are coated with Barium Sulfate (BaSO4) to enhance multi-light scattering. This idea is inspired by a conventional integrating sphere, which is used as a light source with a uniform luminance field at the exit port and also as a uniform illumination field at various distances for photo metric and radiometric applications (Liu et al., 2015). For uniform distribution of light 14 multi-spectral Light Emitting Diodes (LED) are distributed uniformly around the circumference of CSE. Diffusion disks in front of the LED are mounted. The measurement setup also consists of a camera system which has a commercial camera along with a slider. The slider helps in adjusting the focal length of the lens according to the tool length. This unique and innovative light source can be used for various computer vision tasks such as object detection and semantic segmentation. A rotation plate is located below CSE to ensure that images are captured in a sequence of 15° so that, the entire 360° view of the tool is obtained.

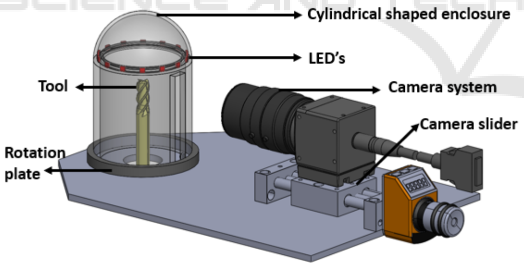

Figure 1: Measurement setup for image acquisition of components with high reflection co-efficient and complex helical shaped structures.

The proposed measurement allows to capture high-quality images without any reflections. To prove the mentioned point, a TiN coated milling tool was captured using the proposed measurement setup and normal illumination conditions without any controlled environment. The difference can be observed in the Figure 2

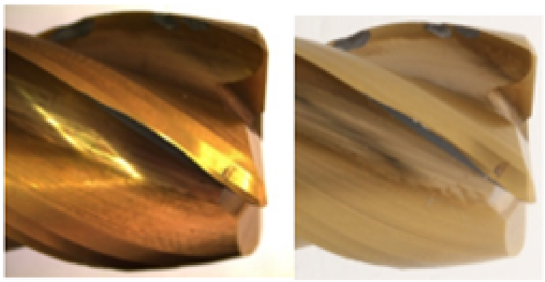

Figure 2: Comparing the images of the same tool captured using normal illumination conditions(left) and the proposed measurement setup.

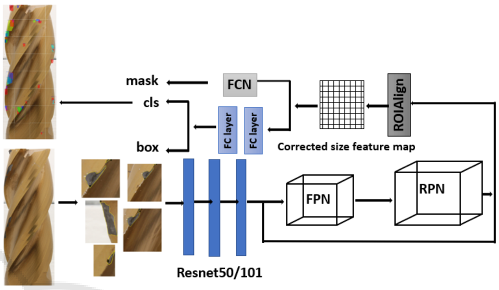

Figure 3: The Mask R-CNN framework for instance segmentation used for high resolution damage detection on milling tools.

3 METHOD USING MASK R-CNN ALGORITHM

The objective of this work is to develop a method to capture high quality images of milling tools with TiN coating to detect and segment damages on these tools by using the state-of-the-art segmentation and object detection framework Mask R-CNN.

Figure 3 illustrates the architecture of Mask R-CNN. Mask R-CNN has three outputs, a class label, bounding box, and object mask. It consists of a backbone network for generating multi-scale features maps, FPN to enhance extracting semantic and abstract information from the feature maps, RPN module for generating a plenty of region proposals for refining bounding boxes and a mask head for generating binary masks of the objects in out cases the damages occur on the drilling tools.

The working principle of our proposed Mask R-CNN based algorithm can be described step-wise as follows:

1. Capture high quality images by using our measurement setup, described in Figure 1.
2. Generate a data set by cropping each image to small fraction of 36 images and assign each image to an element of an 9x4 matrix. This step pre-

vents losing of spatial information and enhances training feature extraction by FPN and reduces the training and test time.

3. Feed the cropped image to a residual network ResNet101 or ResNet50 with FPN for enhancing feature extraction to generate feature maps.
4. The feature map is then scanned by the RPN network with a sliding window, looking for the potential candidate for generating proposals with different sizes and aspect ratios.
5. Now the feature maps obtained from the RPN possess large number of framed candidates as proposals. The next step is to use softmax classifier, frame regression and non-maximum suppression to discard inaccurate proposals and remain only top-scoring predictions as RoI's for the next step.
6. The remaining Region of Interest (RoI) on the feature maps are then sent to the Region of Interest Alignment layer (RoIAlign layer) to perform pooling and quantization on RoI thereby a fixed size of feature map for each proposal is generated.
7. The new feature map undergoes the two branches as mentioned above. The first one is a fully connected layer for object classification and frame regression and the second branch is a fully convolutional network for pixel segmentation and mask prediction.
8. Finally, the damages on each cropped image were obtained. Since the cropped images have been assigned to a 4×9 matrix the cropped images can be merged together to depict the damages marked with bounding boxes, scores and masks.

3.1 Related Work

In this section, an introduction of the DCNN with special emphasis on the Region Based Object Detection (RBOD) and Semantic Segmentation (SS) methods is provided. Several research studies have been undertaken to develop an DCNN for locating class-specific and class agnostic bounding boxes (Szegedy et al., 2013; Szegedy et al., 2014; Erhan et al., 2014). Fully-Connected layer (FC) has been used to train a model for predicting a box with special coordinates to localize single objects and for detecting multiple class-specific tasks (Sermanet et al., 2014). These techniques have been employed for the region-based CNN (R-CNN) object detection approach (Girshick et al., 2014). Using R-CNN Girshick et al were able to present a simple and scalable detection algorithm that improves mAP on PASCAL VOC 2012 dataset by more than 30%. R-CNN lacks of computation sharing, resulting in slow convolutional operations performance in forward pass for each object, resulting in a high training time and test time as well. R-CNN combined with spatial pyramid pooling networks (SPP-nets) can speed up R-CNN by sharing computation power up to 100 times at test time and 3 times at training time (He et al., 2015). SPPnets has still some drawbacks. During fine-tuning the SPPnets cannot update the convolutional layers that proceed the spatial pyramid pooling and decrease the accuracy of DCNN. To overcome the disadvantages of R-CNN and SPPnet Girshick et al. introduced Fast R-CNN (Girshick, 2015) as an extension of R-CNN, which then extended by Faster R-CNN in 2017 (Ren et al., 2017). Fast R-CNN is faster than R-CNN and precedes training on VOC07 dataset 9 times faster than R-CNN (Girshick, 2015). Faster R-CNN is flexible and robust two-stage system and considered to be the leading frame work in several benchmarks (Tsung-Yi Lin et al., 2017; Shrivastava et al., 2016). The common idea behind Faster R-CNN is to use convolutional feature map generated by a DCNN (e.g. Resnet) to determine region proposals with different anchor sizes by using sliding windows for feature extraction, whereas Fast R-CNN takes an input as an entire image with a set of object proposals, which are extracted by a region of interest pooling layers.

Applying instance segmentation and object detection tasks simultaneously proves to be challenging, because it requires correct detection of all objects in the image and segmenting each instance of the object consecutively. The computer vision community has improved beside object detection semantic segmentation tasks separately. In large part, this have been driven by powerful baseline systems, which are based on segment proposals methods (Girshick et al., 2014; Hariharan et al., 2014; Hariharan et al., 2017). Jonathan Long et al. defined a fully convolutional network (FCN) for segmentation. FCN combines layers of the feature hierarchy and refines the spatial precision of the output at the same time (Shrivastava et al., 2016). Deep Mask (DM) model with two branches has been introduced by (Pinheiro et al., 2015). For high quality object segmentation the masks use only the upper-layer to extract CNN features and predict the likelihood of that segmented object. To improve the object segmentation masks, and increase the pixel segmentation accuracy a deep learning approaches based on augmentation feedforward networks with top-down refinement, called SharpMask, has been proposed in 2016 (Pinheiro et al., 2016).

Mask- R-CNN has been introduced by (He et al., 2017; Nur Ömeroğlu et al., 2019) to extend Faster R-

CNN by adding a second branch for predicting object mask beside the exiting branch for bounding box regression, adding only a small overhead to Faster R-CNN. Since that time, Mask-R-CNN based methods for object detection and classification task have been widely used to determine the category and localization of multiclass objects, e.g., to identify and segment polyps in the colonoscopy images (Kang and Gwak, 2019), detect ships on high resolution sensing images (You et al., 2019), quantification of blueberries in the wilds (Gonzalez et al., 2019) and for a variety of practical damage detection application (Zhang et al., 2020).

3.2 Different Backbones of Mask R-CNN

In this section the ResNet Backbone used in Mask R-CNN is discussed. The residual backbone networks ResNet101 and ResNet50 have been used. While ResNet101 consists of 101 layers, ResNet50 consists of 50 layers. Both networks combine FPN to get feature maps of four levels P2, P3, P4, P5 corresponding to last residual block for the conv2, conv3, conv4 and conv5 outputs (Tsung-Yi Lin et al., 2017). Thus the proposed backbone enhances in extracting damages with different scale.

3.3 The Improvement of Detection Accuracy by Adjusting RPN

The region-based detector RPN has been used in Fast R-CNN, Faster R-CNN and Mask R-CNN to generate initial regions proposals at various scales and aspect ratios. This is done by using appropriate multiple anchor boxes as shown in figure 4. The RPN takes different size of feature maps generated by the FPN module and provides outputs of object region boundary and their associated object scores. The scores specify the likelihood of each proposed region containing an RoI to determine the level of the feature pyramid in which the sliding (red window) is performed. The regions scanned by the sliding window are called anchors. Anchors are boxes centered at the sliding window and are associated with different sizes and aspects ratios distributed over the whole feature map. The vector undergoes two 1×1 convolutional layer for box regression and box classification. At each sliding-window location multiple region proposals are predicted with maximal proposals k referred to the anchor boxes. The box regression layer outputs 4k coordinates and the classification layer outputs 2k proposals with probability score to estimate whether an object exist or not. If an anchor box has an Intersection over Union ratio (IoU) with ground truth greater than 0.8, it is considered as positive label, otherwise it considered as negative label. Therefore, the scale and the size of the anchor boxes were tuned and adjusted to improve the damage detection accuracy. To reduce redundancy, Non-Maximum Suppression (NMS) was applied to suppress low scored proposals.

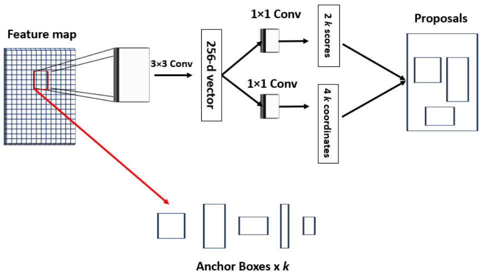

Figure 4: The RPN architecture. The red box represents the sliding window.

3.4 Region of Interest Alignment Layer (RoIAlign)

As mentioned above, image segmentation at pixel level is applied by the mask branch to determine whether a given pixel is a part of the target (here the damage) or not. During the convolutional and polling operations accompanied by quantization, the image sizes changes and causes a positional offset on the RoI. This process affects the accuracy of the small targets. Therefore, (Region of interest alignment layer) RoIAlign is applied, in which the sampling points are increased to calculate each sampling point by a bilinear interpolation to derive the value of the entire RoI with less offset and error. RoIAlign improve the average precision highly (He et al., 2020).

3.5 Training and Loss Function

During the training process, optimizing the loss function plays an import role for both object detection and semantic segmentation. The training process contains forward propagation and backward propagation. Forward propagation starts with extracting the feature map and has three branches for calculating the general loss: The mask loss, the classification loss and the location regression loss, respectively. The Back-propagation updates the parameters of each layer in the network and minimizes the loss function by momentum optimization algorithm (Sutskever et al., 2013; Rumelhart et al., 1986). The RPN module is trained by object/non-object binary classification to each anchor. A positive label is assigned for the anchor with the highest IoU overlapped with the ground-truth box or higher than 0.8 overlapped with the any ground-truth box. Following the multi-task loss in Fast R-CNN (Girshick, 2015) the loss function of the first branch for the classification and regression is given by:

$$L(\{p_i\},\{t_i\}) =$$
$$\frac{1}{N_{cls}}\sum_i L_{cls}(p_i,\ p_i^*) + \lambda\frac{1}{N_{reg}}\sum_i p_i^* L_{reg}(t_i,t_i^*) \quad (1)$$

Where i is the index of an anchor in a mini-batch and p_i is the predicted probability of anchor being an object or not. The ground-truth label p_i^* is 1 if the anchor is positive and 0 if the anchor is negative. The vector t_i contains 4 coordinates of the ground-truth boxes and t_i^* assigned represents the coordinates of the predicted bounding box.

The first term of equation 1 L_{cls} is the log loss over the binary classification and assigned to the two classes, namely damage or no damage. The location regression L_{reg} is the smooth $L1$ (Faster, 2015) loss between the vector t_i^* and t_i. The loss t_i is only activated, when the anchor is positive and is balanced by λ (Pinheiro et al., 2016). The outputs are marked with bounding boxes assigned to the localization of the damage with a probability of being there a damage or not. Since in this work the object detection and the semantic segmentation are combined to classify each pixel assigned to the damages, the mask branch outputs a Km^2 dimensional matrix for each RoI corresponds to a K binary masks of $m \times m$ dimension for each of the K classes. Similar to (He et al., 2020) L_{mask} is defined for the k_{th} mask of the RoI associated with the ground truth class k as the average binary cross entropy loss:

$$L_{mask} =$$
$$\frac{1}{m^2}\sum_{1\le i,j\le m}[y_{ij}log(\hat{y}_{ij}^k + (1-y_{ij}))log(1-\hat{y}_{ij}^k)] \quad (2)$$

Where y_{ij} is the label of a cell $(i,\ j)$ in the true mask for the region of size $m \times m$ and $\hat{y}_{ij}^k$ is the predicted value of the same cell in the mask learned for the ground-truth class k. The multitasking loss function of Mask R-CNN is therefore given by:

$$L = L(\{p_i\},\{t_i\}) + L_{mask} \quad (3)$$

4 EXPERIMENTAL RESULTS AND ANALYSIS

The dataset of cutting tools captured from the measurement system mentioned in Figure 1 was used. The measurement system generates homogeneous illuminated images of the TiN coated cutting tools with high contrast and low noise. The high-quality images allow us to achieve reliable results using small data sets. The network of (Waleed Abdulla, 2017) was modified and implemented in this study.

4.1 Dataset

There is a lack of datasets of TiN coated milling tool, especially for damage detection applications. Capturing valuable data of optical critical objects is challenged by a lot of difficulties. One of them is avoiding reflection and shadow in the images. The exposure of helical shaped cutting tool makes it even more challenging since the complex shape of the cutting tool tend to have a variation of brightness, contrast and artifacts. To overcome these challenges, a new illumination technique to capture high quality images was developed, which was allowing to use only a few images as a training data set and getting reliable results with less than 25 training epochs. For full inspection 24 images from different angles were captured by rotating the milling tool. Each image was cropped to 36 small fragments of a fixed size 512×512 pixels, making a total of 864 images. Around 144 damages were annotated by experts.

4.1.1 Data Augmentation

The goal was to achieve high performance with only a few manually annotated images. Therefore, the following data augmentation technique were applied to increase the training data set from 518 to 5180 images.

1. The images were randomly flipped (horizontally and vertically).
2. The images were randomly rotated in a degree range between -90 to 90.
3. The images have been scaled from 50% to 150% of their original size.

4.1.2 Cropping Images

The Mask R-CNN and other current instance segmentation methods are designed for supervised learning. Typically a large amount of labeled data for training are required to obtain good results. In this work it has been shown that only a few images in combination with transfer learning and appropriate data augmentation can generate high resolution damage detection, reaching an Average Precision (AP) of higher than 0.83. For the experiment, 24 high quality images are used, where each image is cropped into 36 small fragments, resulting in total 864 images to fully utilize the spatial information. For damage detection database ground truth annotations of 144 images have been done manually by drawing a bounding box over the damages.

4.2 Implementation Details

The algorithm was implemented in Python and all of the experiments were performed using NVIDIA Tesla K80 24GB, Linux operating system, 2 virtual CPUs with capacity of 2GHz and 7680MiB system memory. The Mask-R-CNN was trained using ResNet101 and ResNet50 as a backbone architecture for 25 epochs using a learning momentum of 0.9, a learning rate of 0.001, weights decayed by 0.0001, batch-size of 4 images per GPU. ResNet101 took 4 hours 15 minutes for training whereas ResNet50 took only 3 hours 58 minutes.

4.3 Used Evaluation Metrics

The anchors with IoU higher than 0.7 for all of the Ground Truth (GT) boxes are assigned to positive labels, whereas anchors with IoU less than 0.3 for all of the GT boxes are assigned to negative labels. For evaluating the performance of prepared models, the standard metrics Intersection-over-Union (overall IoU) and the precision haven been used. The average precision AP over different IoU thresholds has been considered from 0.5 to 0.95 at an interval step of 0.05 (0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95) (Ren et al., 2017). The precision metric is indicated as: $(AP, AP_{50}, AP_{55}, AP_{60}, AP_{65}, \ldots AP_{95})$, here the AP_{50} indicates the Average Precision at IoU threshold of 0.5 and AP_{55} indicates the Average Precision at IoU threshold of 0.55 and so on.

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

Precision represents the exactness as the ratio between the number of correctly detected pixels and the total number of detected pixels.

- True Positives (TP): The number of pixels correctly identified as a mask (white pixels).
- True Negatives (TN): The number of pixels correctly identified as not part of a mask (black pixels).
- False Positives (FP): The number of pixels incorrectly identified as a mask.
- False Negatives (FN): The number of pixels incorrectly identified as not part of a mask.

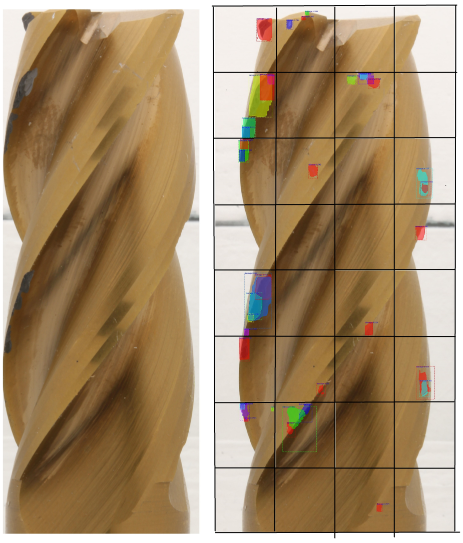

Figure 5: Results of a high-resolution damage detection and instance segmentation of an TiN coated milling tool. On the right the damages are marked with the bounding boxes and the prediction probability.

4.4 Damage Detection Result

The result presented in the current paper is based on 864 images. 60% of the whole dataset was used for training, 20% used for validation and the remaining 20% was used for testing the model.

4.4.1 High Resolution Milling Tool Damage Detection Result

Figure 5. displays an example of an TiN milling tool image captured by the proposed novel measurement setup (left side). On right side the result after applying the Mask-R-CNN algorithm to detected damages can be observed. The damages are depicted and marked with a bounding boxes and prediction probability. It can be clearly seen in Figure 5 that almost all damages have been detected. Semantic segmentation can be used to determine the size, the shape and localization of the damages.

4.4.2 ResNet101 vs ResNet50

Generally, deep learning requires a huge amount of data and in most cases, it is difficult to find the data sets especially for optical critical components such as drilling or milling tools. So, due to this reason the training data was augmented.

At first the detection stage was performed using ResNet101 and it was followed by ResNet50 backbone. Several tests have been done by using different training parameters such as:

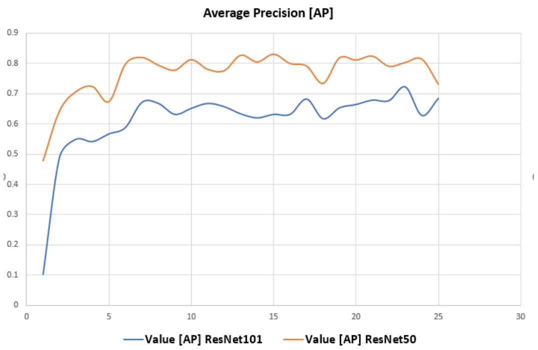

Figure 6: Comparison of average precision by using the backbones ResNet50 and ResNet101.The AP as a function of epochs of both models with different backbone architecture (ResNet101 and ResNet50).

- Number of epochs
- RPN anchor scales
- AP at diffrent IoU (0.50, 0.55, 0.60, . . . , 0.95)

Firstly, the average precision was compared by using the backbones ResNet50 and ResNet101 as seen in Figure 6. To compare the detection and learning performance the Mask-R-CNN was trained using the ResNet101 and ResNet50 architecture for 25 epochs by simultaneously calculating the AP as a function of epochs. Figure 6 shows the AP@50 for both models as a function of epochs. An AP of 0.83, was achieved with the ResNet50 backbone architecture at epoche 21 whereas with the ResNet101 backbone architecture an AP of only 0.71 at epoche 23 was achievable. The ResNet50 has a smaller number of layers, which helps avoiding overfitting. The model file size of ResNet50 is about 180 MB compared with 250 MB of the ResNet101, making ResNet50 more effective for a variety of applications with less computational complexity. In both cases the risk of overfitting has been increased after the 23 epochs. Although the damages differ in shape and size, the necessary complexity and depth of an appropriate neural network cannot be easily determined. Due to the small data size and only two classes (damage or not a damage) ResNet 50 seems to perform better than ResNet 101. In the context of instance detection and instance segmentation, the damage identification and its position location must be done. The Intersection over Union (IoU) measures the overlap between the predicted boundary and the ground of truth boundary. Thus, the average precision for different IoU was calculated. An appropriate scales of anchor boxes can improve the efficiency and accuracy of the region proposal generation and hence improve the overall object detection accuracy. Therefore, the AP for different IoU and different scales of anchor boxes was evaluated as shown in Figure 7. It was found, that the model with the ResNet50 backbones architecture performs better than ResNet101, especially by adopting the anchor boxes scales $\{16^2, 32^2, 64^2, 128^2, 256^2\}$ with the aspect ratio $\{1:1, 1:2, 2:1\}$.

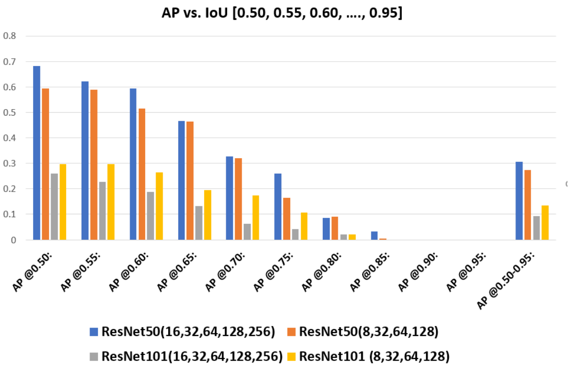

Figure 7: AP at diffrent IoU (0.50, 0.55, 0.60, . . . , 0.95). The ResNet50 architecture performs better than the ResNet101 architecture.

5 CONCLUSION

The results show that high-quality and good-resolution images that are captured using the proposed measurement setup are capable of achieving superior results with the help of deep convolution neural networks. For training the network, each image has been divided into 36 fragments to ensure high resolution damage detection by utilizing the highest capability of the FPN. Both in instance and semantic based image segmentation promising results have been achieved using few images combined with data augmentation, which pave ways for new opportunities in inspection applications. To identify the damages, Mask R-CNN which consists of feature extraction of the images followed by other convolutional layers was implemented. ResNet 50 and ResNet 101 architectures were fine tuned for feature extraction. The segmentation using ResNet 50 has achieved better results with less computational time when compared to ResNet 101.

As the future scope for the current work, a bigger dataset will be generated that includes different cutting tools with different coating and variants. Another area of research would focus on surface roughness estimation of the tools using the images from the developed measurement setup.

REFERENCES

Caggiano, A. (2018). Cloud-based manufacturing process monitoring for smart diagnosis services. *International Journal of Computer Integrated Manufacturing*, 31(7):612–623.

Chen, J.-Y. and Lee, B.-Y. (2010). Development of a simplified machine for measuring geometric parameters of end mills.

Cupek, R., Erdogan, H., Huczala, L., Wozar, U., and Ziebinski, A. (2015). Agent based quality management in lean manufacturing. In Núñez, M., Nguyen, N. T., Camacho, D., and Trawiński, B., editors, *Computational Collective Intelligence*, volume 9329 of *Lecture Notes in Computer Science*, pages 89–100. Springer International Publishing, Cham.

Cupek, R., Ziębiński, A., Drewniak, M., and Fojcik, M. (2018). Improving kpi based performance analysis in discrete, multi-variant production. In Nguyen, N. T., Hoang, D. H., Hong, T.-P., Pham, H., and Trawiński, B., editors, *Intelligent Information and Database Systems*, volume 10752 of *Lecture Notes in Computer Science*, pages 661–673. Springer International Publishing, Cham.

Erhan, D., Szegedy, C., Toshev, A., and Anguelov, D. (2014). Scalable object detection using deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2147–2154.

Faster, R. (2015). Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, page 9199.

Girshick, R. (2015). Fast r-cnn. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448. IEEE.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587.

Gonzalez, S., Arellano, C., and Tapia, J. E. (2019). Deepblueberry: Quantification of blueberries in the wild using instance segmentation. *IEEE Access*, 7:105776–105788.

Grzechca, D., Ziębiński, A., and Rybka, P. (2017). Enhanced reliability of adas sensors based on the observation of the power supply current and neural network application. In Nguyen, N. T., Papadopoulos, G. A., Jędrzejowicz, P., Trawiński, B., and Vossen, G., editors, *Computational Collective Intelligence*, volume 10449 of *Lecture Notes in Computer Science*, pages 215–226. Springer International Publishing, Cham.

Hariharan, B., Arbeláez, P., Girshick, R., and Malik, J. (2014). Simultaneous detection and segmentation. In *European Conference on Computer Vision*, pages 297–312.

Hariharan, B., Arbelaez, P., Girshick, R., and Malik, J. (2017). Object instance segmentation and fine-grained localization using hypercolumns. *IEEE transactions on pattern analysis and machine intelligence*, 39(4):627–639.

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969.

He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2020). Mask r-cnn. *IEEE transactions on pattern analysis and machine intelligence*, 42(2):386–397.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9):1904–1916.

Kang, J. and Gwak, J. (2019). Ensemble of instance segmentation models for polyp segmentation in colonoscopy images. *IEEE Access*, 7:26440–26447.

Liang, S. Y., Hecker, R. L., and Landers, R. G. (2004). Machining process monitoring and control: the state-of-the-art. *J. Manuf. Sci. Eng.*, 126(2):297–310.

Liu, L., Zheng, F., Zhu, L., Li, Y., Huan, K., Shi, X., and Liu, G. (2015). Luminance uniformity of integrating sphere light source. In *2015 International Conference on Optoelectronics and Microelectronics (ICOM)*, pages 265–268.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., and Berg, A. C. (2016). Ssd: Single shot multibox detector. 9905:21–37.

Nur Ömeroğlu, A., Kumbasar, N., Argun Oral, E., and Ozbek, I. Y. (2019). Mask r-cnn algoritması ile hangar tespiti hangar detection with mask r-cnn algorithm. *27th Signal Processing and Communications Applications Conference (SIU)*, 31(7):1–4.

Pinheiro, P. O., Collobert, R., and Dollar, P. (2015). Learning to segment object candidates.

Pinheiro, P. O., Lin, T.-Y., Collobert, R., and Dollàr, P. (2016). Learning to refine object segments.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99.

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1137–1149.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533–536.

Schulz, H. and Moriwaki, T. (1992). High-speed machining. *CIRP annals*, 41(2):637–643.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., and LeCun, Y. (2014). Overfeat: Integrated recognition, localization and detection using convolutional networks. 2nd international conference on learning representations, iclr 2014. In *2nd International Conference on Learning Representations, ICLR 2014*.

Short, M. and Twiddle, J. (2019). An industrial digitalization platform for condition monitoring and predictive maintenance of pumping equipment. *Sensors (Basel, Switzerland)*, 19(17).

Shrivastava, A., Gupta, A., and Girshick, R. (2016). Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 761–769.

Spišák, E. and Majernikova, J. (2017). Increasing of durability of cutting tools. *Advances in Science and Technology Research Journal*, 11:141–146.

Sutskever, I., Martens, J., Dahl, G., and Hinton, G. (2013). On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147.

Szegedy, C., Reed, S., Erhan, D., Anguelov, D., and Ioffe, S. (2014). Scalable, high-quality object detection. *arXiv preprint arXiv:1412.1441 [Titel anhand dieser ArXiv-ID in Citavi-Projekt übernehmen]*.

Szegedy, C., Toshev, A., and Erhan, D. (2013). Deep neural networks for object detection.

Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie (2017). Feature pyramid networks for object detection.

Waleed Abdulla (2017). Mask r-cnn for object detection and instance segmentation on keras and tensorflow.

Yingjie, Z. (2014). Energy efficiency techniques in machining process: a review. *The International Journal of Advanced Manufacturing Technology*, 71(5-8):1123–1132.

Yli-Ojanperä, M., Sierla, S., Papakonstantinou, N., and Vyatkin, V. (2019). Adapting an agile manufacturing concept to the reference architecture model industry 4.0: A survey and case study. *Journal of Industrial Information Integration*, 15(5):147–160.

You, Y., Cao, J., Zhang, Y., Liu, F., and Zhou, W. (2019). Nearshore ship detection on high-resolution remote sensing image via scene-mask r-cnn. *IEEE Access*, 7:128431–128444.

Zhang, Q., Chang, X., and Bian, S. B. (2020). Vehicle-damage-detection segmentation algorithm based on improved mask rcnn. *IEEE Access*, 8:6997–7004.