Data Augmentation for Semantic Segmentation in the Context of Carbon Fiber Defect Detection using Adversarial Learning

Silvan Mertes[1], Andreas Margraf[2][a], Christoph Kommer[1], Steffen Geinitz[2] and Elisabeth André[1]

[1]*University of Augsburg, Universitätsstraße 1, 86159 Augsburg, Germany*
[2]*Fraunhofer IGCV, Am Technologiezentrum 2, 86159 Augsburg, Germany*

Keywords: Image-to-Image Translation, Carbon Fiber, Data Augmentation, Computer Vision, Industrial Monitoring.

Abstract: Computer vision systems are popular tools for monitoring tasks in highly specialized production environments. The training and configuration, however, still represents a time-consuming task in process automation. Convolutional neural networks have helped to improve the ability to detect even complex anomalies withouth exactly modeling image filters and segmentation strategies for a wide range of application scenarios. In recent publications, image-to-image translation using generative adversarial networks was introduced as a promising strategy to apply patterns to other domains without prior explicit mapping. We propose a new approach for generating augmented data to enable the training of convolutional neural networks for semantic segmentation with a minimum of real labeled data. We present qualitative results and demonstrate the application of our system on textile images of carbon fibers with structural anomalies. This paper compares the potential of image-to-image translation networks with common data augmentation strategies such as image scaling, rotation or mirroring. We train and test on image data acquired from a high resolution camera within an industrial monitoring use case. The experiments show that our system is comparable to common data augmentation approaches. Our approach extends the toolbox of semantic segmentation since it allows for generating more problem-specific training data from sparse input.

1 INTRODUCTION

1.1 Motivation

Optical sensors such as high resolution cameras are widely used in the field of online process monitoring (OPM) to detect changes and anomalies in production environments. The image acquisition task set aside, the potential of large image data can only be explored with well-selected and problem-tailored segmentation and classification models. Machine learning algorithms have outperformed classic image processing approaches in several studies over the last decade (Cavigelli et al., 2017; McCann et al., 2017). Especially convolutional neural networks (CNN) proved effective and applicable to many problem domains due to their ability to generalize well on large training sets (Simonyan and Zisserman, 2014; He et al., 2016). In highly specialized industrial environments, however, image data is usu-

ally sparse due to the time-consuming and therefore expensive collection and labeling procedure. This results in a lack of labeled data necessary to train various machine learning algorithms for defect detection. In this paper we present scenarios for carbon fiber textiles which exhibit unique surface structures and heterogeneous anomalies. Since regular image processing with common filters such as edge, contour, thresholding or fourier transform would not cover the variety of anomalies, we propose a CNN based semantic segmentation method that relies on an adverserial network approach for data augmentation (DA). Hereby, DA is based on *pix2pix* image-to-image translation and used to increase detection reliability and reduce training effort for industrial machine vision.

1.2 Toward Better Data Augmentation

A major benefit of CNNs is their ability to classify images, i. e. to predict which category the images or single pixles belong to based on the "euclidean-distance" or some kind of similarity score. However, state-of-

[a] https://orcid.org/0000-0002-2144-0262

the art CNNs need large amounts of training data to produce reliable results. Since it is time-consuming and expensive to manually create annotations, DA is a popular tool to artificially enlarge the training set. Common DA methods rely on rotation, scaling, shifting and similar algorithms. However, these simple algorithms do not explicitly generate new patterns, but rather alter existing data within very narrow limits. GANs are designed to create output images as realistic as possible so that they become *indistinguishable* from real-world photographs. Modified GANs for image-to-image translation, such as *pix2pix*, enhance the capabilities of original GANs to transfer images from one domain to another. We make use of this property as we teach GANs to transfer randomly generated pixelwise label data to image data specific to our problem domain. Thus, we create label and image pairs that are truly new and can be used as training data for CNNs in a semantic segmentation context.

1.3 Structure

For a better understanding of the topics covered, the remainder of this paper is divided into four sections: First of all, we provide an overview on related work and existing technology in section 2. We then describe the proposed approach in section 3 and present the experimental setup in section 4. In addition, we discuss the results and compare them to related concepts in section 5. Finally, we draw conclusions from our findings and give an outlook on future research in section 6.

2 RELATED WORK

The following section will give an overview of previous work published in related fields of research, i. e. machine learning (ML), artifical neural networks (ANN), computer vision (CV), OPM and Organic Computing (OC). Identifying anomalies, e. g. misaligned carbon fibers on textile surfaces is a challenging task, especially from an engineer's perspective. On the one hand, camera sensors need to provide an outstanding image quality with a high resolution, on the other hand heterogenous surface patterns serve as tough barriers for the design and training of machine learning models. Geinitz et al. therefore designed (Geinitz et al., 2016) a line scan sensor with an adapted image processing filter pipeline in order to handle the observed variability in surface images.

Margraf et al. proposed an evolutionary learning approach (Margraf et al., 2017) for the automated design of image processing solutions for carbon fiber

fault detection which was extended by Stein et al. (Stein et al., 2018) with an architecture for the automated generation of processing pipelines. A big leap in classifying large image sets by means of CNN has been achieved by AlexNet (Krizhevsky et al., 2012), GoogleNet (Szegedy et al., 2015) and VGGNet (Simonyan and Zisserman, 2014). CNNs have been used in the context of steel defect classification by Masci et al. (Masci et al., 2012) and for photometric stereo images by Soukup et al. (Soukup and Huber-Mörk, 2014). Ren et al. proposed a region proposal network for real-time object detection (Ren et al., 2015). Ferguson et al. used CNNs and transfer learning for defect detection in X-ray images (Ferguson et al., 2018). Staar et al. examined CNNs for industrial surface inspection (Staar et al., 2019). Pixel-based segmentation was first introduced by Long et al. (Long et al., 2015). Schlegl et al. proposed unsupervised anomaly detection with GANs for marker discovery (Schlegl et al., 2017). Also, Di Mattia et al. presented a survey on GANs for anomaly detection (Di Mattia et al., 2019). Furthermore, transfering colors from a given photographic context into another has been discussed in related publications (Zhang et al., 2016; Xie and Tu, 2015). The concept of image-to-image translation was first introduced by (Isola et al., 2017) who presented the *pix2pix* architecture to project various image domains such as edge objects or label images to colored photographs. Images with a large gap between a small foreground and a comparably large background pose a challenge to pixel-based segmentation which is reflected in the case of carbon fiber images. In this context, networks with a multi-channel feature map seem to perform better which was confirmed using the U-Net architecture (Ronneberger et al., 2015). It should be noted that this architecture makes heavy use of data augmentation while taking only few input data. However, its potential is limited when using very small input data sets. For the specific purpose of data augmentation, (Frid-Adar et al., 2018) and (Mariani et al., 2018) successfully applied GANs to classification tasks. Also, (Choi et al., 2019) presented an approach to use image-to-image translation networks for semantic segmentation tasks. Therefor, they transformed labeled data to related image domains so that the original label still fits to the newly created image. (Huang et al., 2018) utilized image-to-image networks to be able to use multiple image domains to train shared segmentation subtasks. However, to the best of our knowledge, there exists no work using GANs to generate completely new image and label pairs for enhanced training datasets in the context of semantic image segmentation. The authors of this paper are aware that metaheuristics for hy-
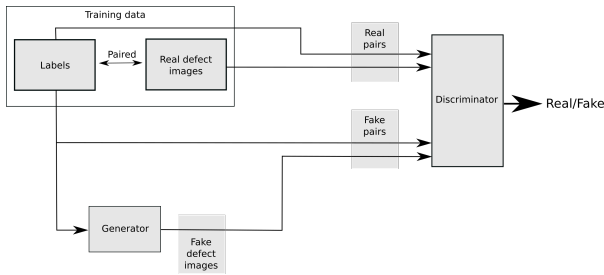
Figure 1: Training of a *pix2pix* network to perform image-to-image translation between labels and defect images (Step 1).

perparameter optimization, e. g. swarm intelligence (Strumberger et al., 2019) exist. However, this field of research is not subject to the presented approach.

3 APPROACH

In this paper, we take a first step toward a *pix2pix* based approach in the context of carbon fiber defect detection. We show how *pix2pix*, an image-to-image translation approach, can be adapted and utilized to enlarge training datasets for semantic segmentation. In the specific context of quality monitoring, we demonstrate that our approach is able to improve semantic segmentation if only small datasets are available. We also compare our approach to conventional DA which will be discussed in section 5. Thereby, it can be used to create training data for a problem domain that typically lacks of labeled data. While common strategies for DA rely on straightforward transformations such as flipping or rotating the training images, our concept is based on a on a randomized label generator combined with a *pix2pix* architecture to create completely new synthetic training data.

Our approach is three-folded: in a first step, we train a *pix2pix* network to perform an image-to-image translation from labels of defect images to their corresponding image data, i.e. we teach the *pix2pix* network to generate new defect images that correlate with given label data. To train the network, we use a set of existing real data pairs of images and labels. Fig. 1 shows the basic principle of this step. The network architecture and training procedure can be infered from (Isola et al., 2017). The only change that we made to the original architecture was to adapt the size of the input layer to our problem domain.

The structure of carbon fiber faults usually appears as mostly straight or curved lines of varying thickness. Under keen observation the basic structure of the single fibers can be regarded as combined

Table 1: Parameters for the fake label generator.

Parameter	Lower bound	Upper bound
a_1	15	30
a_2	0.02	0.03
a_3	1	50
a_4	-0.5	0.5
a_5	-0.5	0.5
a_6	-0.5	0.5
a_7	0.005	0.0095

graphs with different rotations. In stage 2, a composition of well-tuned stochastic functions destined to create 'fake' labels is applied to the data. The authors are aware of the fact that this method is unique to the domain of carbon fibers and that the experimental design of such stochastic functions can even take more time than applying common data augmentation techniques. However, there are more related problem domains in the context of anomaly detection in which surface anomalies can be imitated with ordinary stochastic trigonometric functions. For example, (Haselmann and Gruber, 2017) showed that anomalies in images of plastic parts manufactured by a foil-insert-molding process can be represented as trigonometric functions.

Several experiments showed that the following function allows to generate graphs that show a similar structure as carbon fiber defects. The trigonometric function $f(x)$ is denoted as follows:

$$f(x) = a_1 \cdot sin(a_2 \cdot x) + a_3 \cdot sin(x) + \\ + a_4 \cdot cos(a_5 \cdot x) + a_6 \cdot x + a_7 \cdot x^2$$

where the parameters a_n are chosen randomly within certain defined intervals. We found appropriate intervals through experimentation and visual inspection. By studying the real defect images it could be noted that the curved structures of the defects show big similarities to trigonometric functions. To model different shapings of the curvings, we chose parameter intervals for trigonometric functions so that the resulting graphs would cover a wide range of structures. For example, we tuned one sine function to use big amplitudes and therefore form the global structure of the label, whereas another sine function only uses small amplitudes to cover curvings that occur on a rather microperspective level. Polynomic functions were also included to model additional types of curvings that remind of aperiodic curvings. The intervals are listed in table 1. For every fake label we randomly set the variables and plotted the resulting graph for $x \in [0, w]$ where w represents the width of the sample images. After creating those plots they were rotated randomly. Multiple graphs were overlapped and the thickness of the resulting lines was varied to cre-
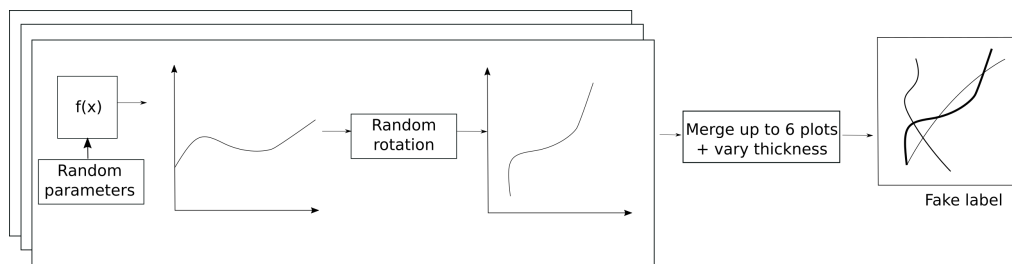
Figure 2: Heuristic to generate fake labels using the label generator (Step 2).

ate images with realistic fiber-like graphs so they can barely be distinguished from real labels. Fig. 2 illustrates this procedure. It should be emphasized that the whole process was performed automatically, i. e. no human involvement was necessary to create 'fake' labels. Fig. 5 depicts a random selection of fake labels created with this heuristic.

Finally, the previously generated label data is fed to the trained *pix2pix* model. By pairing the synthetic labels of the function generator with the synthetic image data of *pix2pix*, we get new image/label pairs to train a CNN for semantic segmentation of defects as shown in Fig. 3. We chose U-Net for this purpose since this architecture could achieve promising results in related fields, e. g. biomedical image segmentation (Ronneberger et al., 2015).

4 EXPERIMENTS AND DISCUSSION

4.1 Domain of Carbon Fiber Defect Images

We chose the domain of carbon fiber defect monitoring to test and evaluate the proposed approach. During production, the surface of carbon fibers apears similiar to common textiles. Ideally, the single fibers are aligned in parallel and form a carpet of straight lines. From time to time, single fibers come undone due to the mechanical impact executed by spools in the transportation system. Misaligned fibers are generally considered a potential defect. Since the position, shape and size of loose or cracked fibers vary heavily, their is no golden sample for single defects. The goal is to identify the defects on a carbon fiber carpet by training a U-Net architecture to perform a binary segmentation of the pixels that contain defects. Fig. 4 shows four examples of defect images with corresponding binary labels.

4.2 Experimental Setup

To evaluate our concept, we ran several experiments to compare the system with conventional data augmentation methods. For data augmentation the following image transformations were applied:

- Randomised crop of squares of different size (*RandomSizedCrop*)
- Horizontal and vertical flip
- Rotation (for 180 degrees)
- Elastic transformation
- Grid distortion

We arranged the image data in four different sets and performed multiple training procedures of a U-Net architecture. Afterwards, we compared the quality of the models trained on the different sets. Every training pair for the U-Net architecture consists of a real or fake defect image and a real or fake binary label image. The following listing shows the composition of the various datasets. The pairs of real data each consist of a real defect image and a corresponding manually labeled image.

Dataset 1 contains 300 pairs of real defect data and corresponding binary label images. For this dataset, only original data was selected and DA was *not* applied.

Dataset 2 contains the same 300 pairs of defect data and corresponding labels as *dataset 1*. During training, however, we applied an *online* form of data augmentation. For each image some of the aforementioned transformation operations were applied with a given probability.

Dataset 3 contains 3000 pairs of defect data and corresponding labels. 2700 of the 3000 data pairs were generated by applying the *pix2pix* based data augmentation approach on *dataset 1* while 300 data pairs were taken from *dataset 1*.

Dataset 4 contains the same 3000 pairs of defect data and corresponding labels as *dataset 3*. We also applied stochastic data augmentation as for *dataset*
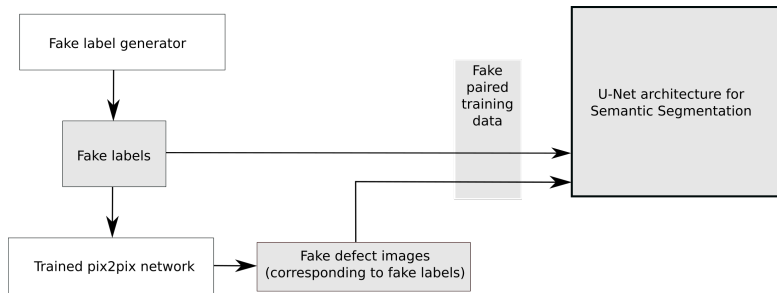
Figure 3: The generation and preparation of training data for U-Net using a trained *pix2pix* model and the fake label generator to create fake training pairs. (Step 3).
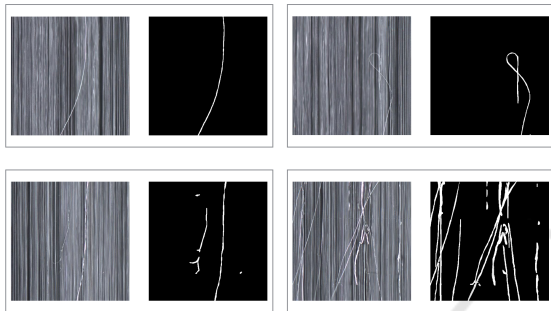

Figure 4: Examples of real image data pairs labelled by experts. The misalgined fibers are visible on top of the fiber carpet.

2, i.e. each image was transformed with a given probability during training. Thus, *dataset 4* combines common data augmentation with our approach.

As mentioned before, the datasets were used to train a U-Net architecture for semantic segmentation. For testing and evaluation we used a separate dataset that contained real defect data and corresponding annotations. This test set was created with the help of domain experts.

4.3 *pix2pix* Configuration

The *pix2pix* GAN configuration is given in table 2. We stopped the training after 3200 epochs, as we could not observe any further improvement of the generated images by that time. Fig. 5 shows a selection of pairs of random generated labels and images generated through application of the *pix2pix* model.

4.4 U-Net Configuration

During training, one setting was applied to the U-Net architecture. Also, DA with conventional image transforms was applied to 2 out of 4 datasets.

Table 2: *pix2pix* Configuration.

Parameter	Value
Learning rate	0.0005
Batch Size	1
Epochs	3200
Loss Function	Mean Squared/Absolute Error

All data augmentation methods are based on the library published by (Rizki et al., 2002). In our experiments, a stochastic component was added to image transformations, i.e. all operations were performed with a given probability. Thus, the randomized crop was given the probability $p = 0.25$ and a window size interval of $[400, 512]$ pixels. Furthermore, the probability for flipping, rotation, elastic transform and grid distortion was set to $p = 0.5$. In the latter case, only one operation, i.e. either elastic transform or grid distortion was allowed (*OneOf*). The rotation was set to exactly 180 degrees. *Elastic Transform* was performed with the parameters $\alpha = 10$, $\sigma = 10$, $alpha\_affine = 512 \cdot 0.05$ and $border\_mode = 4$. *Grid Distortion* was given the parameters $num\_steps = 2$ and $distort\_limit = 0.4$. The operations were applied using the given parameters on every incoming original image. The U-Net model itself was slightly adapted from (Yakubovskiy, 2019) to fit the dataset. The default size of the training images was 512x512, yet the default U-Net setting accepts 28x28. As an encoder, the U-Net uses a ResNet-18 model. The architecture was adapted to fit the input size before applying the model. The training was then performed using the parameters as presented in table 3.

Table 3: U-Net Configuration.

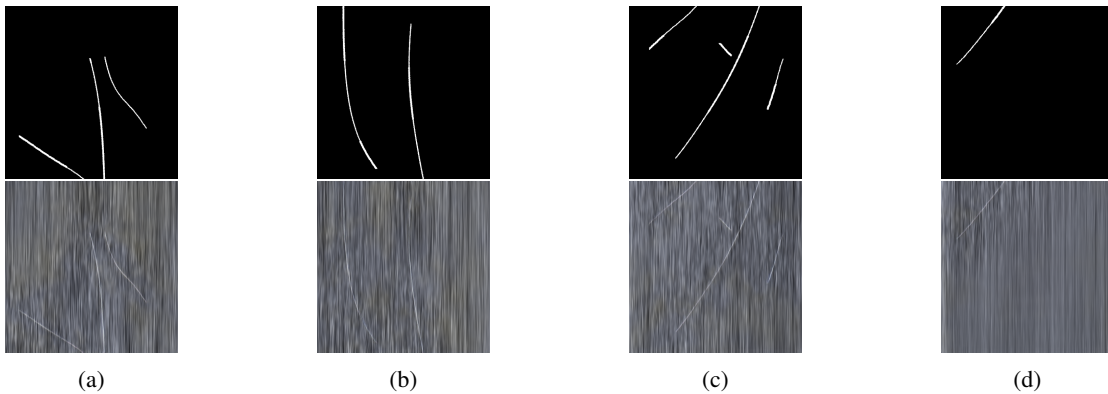| Parameter	Value
Learning rate	0.0001
Batch Size	10
Epochs	200
Loss Function	Binary Cross Entropy / Dice Loss

Figure 5: Samples of synthetic labels (top row) and corresponding *pix2pix* outputs (bottom row) imitating misaligned fibers.

5 EVALUATION

It should be noted that *accuracy*, *MCC* and $F_\beta - Score$ are less dependable for an objective evaluation of the classification model. The proportion between background pixels (i.e. the non-defect pixels) and foreground pixels (i.e. the defect pixels) per image is thoroughly unbalanced. While *accuracy* returns the proportion of true results among all data points examined, *MCC* and $F_\beta - Score$ aim to balance out true and false positives and negatives of the binary classification result. In contrast, the *Jaccard index* or *Intersection over Union (IoU)* is used to measure the similarity of two sets, i. e. the similarity of the ground truth and the prediction. It is defined as the area of overlap over the area of union or formally:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

In the following section, we will focus on the *IoU* metric for an objective and problem related evaluation. For the sake of completeness, all relevant statistical scores are published for each experiment.

5.1 Discussion of Results

For all datasets the training was aborted after 200 epochs since convergence was clearly visible. As Fig. 7 suggests, both training accuracy and loss converge from epoch 100 onwards for all 4 datasets. Training on dataset 1 was stopped at a loss rate of ~ 0.7, while for both datasets 3 and 4 the training ended at a loss rate of ~ 0.4. For dataset 2, model training reached an IoU of ~ 0.6 and ~ 0.5 for validation when the process was aborted. At the same time, the training loss ended at ~ 0.4 and reached a value of ~ 0.2 for validation. Furthermore, training on dataset 1 reached an IoU score of ~ 0.7 while dataset 3 and 4 achieved

an IoU value of ~ 0.8 after 200 epochs. All training results for the four datasets are shown in table 4. The trained models were consecutively applied on the test dataset. The model trained on dataset 1 reached an *accuracy* of 0.985 and *IoU* of 0.391 on the test set. Likewise, the model trained with dataset 2 reached an *accuracy* of 0.992 and an *IoU* of 0.593. As can be seen, the *IoU* for the model based on dataset 3 reached an IoU of 0.579 and an *accuracy* of 0.991, while training with dataset 4 achieved a value of 0.575 for the *IoU* and 0.991 for the *accuracy*. All metrics for the test runs were acquired from prediction on 25 randomly selected sample images as presented in table 4.

Fig. 6 shows a random selection of defect images taken from the test set with red overlays representing the ROIs predicted by the U-Net model. It should be noted that training without DA leads to more false positives which remainds of noise in the overlays as can be seen in Fig. 6b and 6c.

Figure 7 shows that the loss rate for *dataset 1* drops heavily for 50 epochs and converges around a value of 0.5 for the test set and just over 0.0 for the training set. At the same time, the *IoU* value increases heavily for 50 epochs before it slows down and converges after 125 epochs around an *IoU* value of 0.85 for the training set and around 0.4 for the test set.

For *dataset 2* the loss rate drops considerably during the first 5 epochs, then decreases constantly but slightly until it converges around 0.2 after 125 epochs during training. The loss on the test set develops the same way except it converges around a value of 0.4. Again, the *IoU* value increases clearly within less then 5 epochs during training, then only slightly continues to rise before converging around 0.6 after epoch 125. During testing, the *IoU* value increases constantly between epoch 0 and 75, then converges around 0.4 as illustrated in the second row of Fig. 7.

Table 4: U-Net results from test runs on the datasets 1 through 4 for batch size 5.

	PPV	TPR	IoU	ACC	MCC	F1	F2
Dataset 1	0.539169	0.586753	0.390778	0.985035	0.55487	0.561956	0.576576
Dataset 2	0.772803	0.718101	0.592925	0.991935	0.740872	0.744448	0.728413
Dataset 3	0.745926	0.721067	0.578888	0.991419	0.729034	0.733286	0.725905
Dataset 4	0.756767	0.705387	0.57502	0.991471	0.72631	0.730175	0.715098

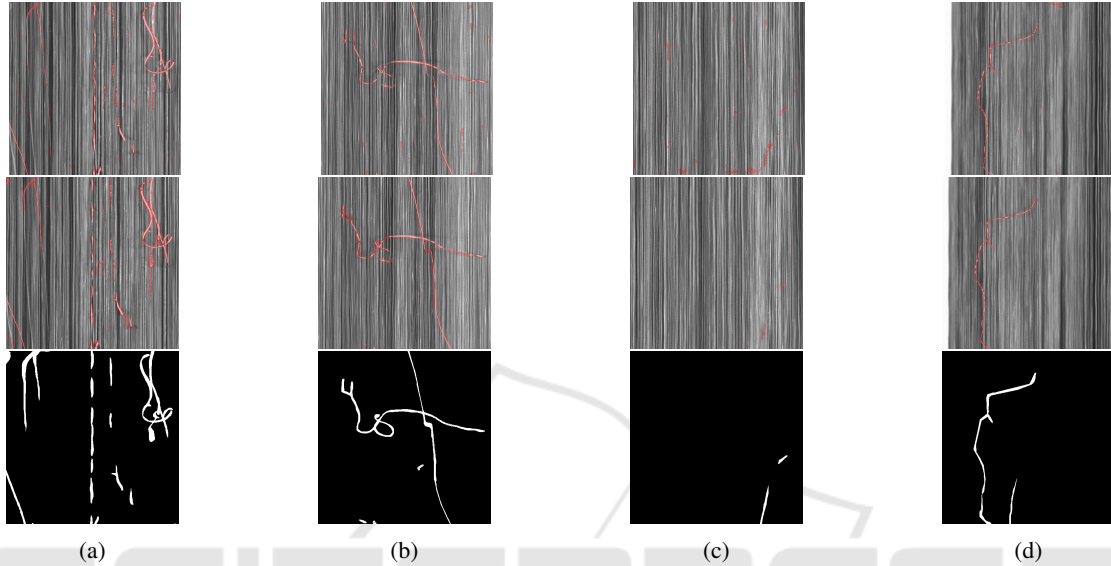   (a)   (b)   (c)   (d)

Figure 6: Real carbon fiber defects from the test set with red overlay from U-Net segmentation for *dataset 1* (top row), *dataset 3* (center row) and the *ground truth* (bottom row).
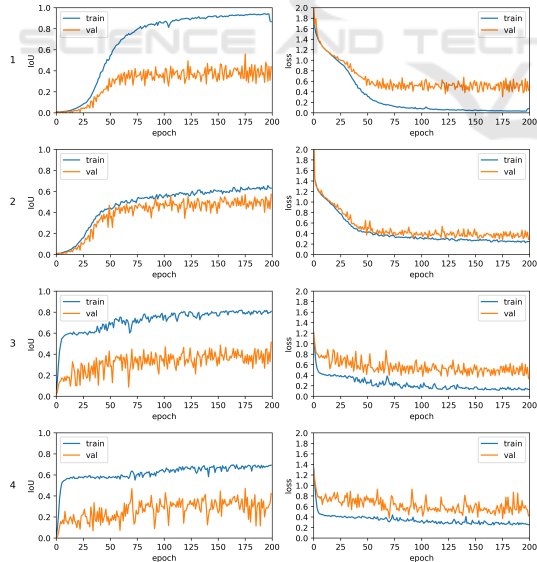

Figure 7: IoU Score and Loss during U-Net training for dataset 1 through 4 from top to bottom row.

As can be seen, the U-Net model trained on dataset 3 significantly outperformed the model trained on dataset 1. This shows that our approach could substantially improve the quality and diversity of the raw training set. When comparing the results of dataset 2 and dataset 3, it becomes apparent that the proposed approach is slightly worse, however not significantly diverts from conventional data augmentation techniques which were applied on dataset 2. The difference comprises within less then 0.02 for the *IoU*.

The combination of synthetic data generated using a GAN with subsequent data augmentation as for dataset 4 did not lead to any improvement. The model trained on dataset 4 outperforms dataset 1, but leads to a slighty lower *IoU*, *accuracy* and *MCC* than dataset 2 and 3. However, the degradation ranges within less then 0.02 for the *IoU* and is therefore not significant under the given circumstances. As the results from table 4 and the samples depicted in figure 6 suggest, the pairs of synthetic images and labels of carbon fiber defects were successfully used to replace traditional data augmentation for semantic segmentation network training. With an *IoU* of 0.579, the presented approach performs comparably to U-Net training with regular data augmentation. Since the absolute difference between dataset 2 and 3 results in a value of 0.01, it appears negligible. The results show that synthesized training data helps to improve the detection quality of a U-Net segmentation model to a great ex-

tent. Moreover, the augmented dataset could be created based on few samples of only 300 images with an image size of 512*x*512 pixels in which the *ROI* on average only covers 1 % of an image frame.

We consider our *pix2pix* based image generation a more realistic and application-oriented form of DA. The experiments were conducted with and without traditional DA in order to evaluate the effectiveness of our approach. Excessive use of traditional DA might superimpose the 'real' data within the training set due to its low level form of manipulation, reproduction and reuse which raises the risk of overfitting during model training. GAN based data generation is also less prone to repetitive patterns since it tries to project the variation found in the original data to the synthetic data.

In summary, the proposed approach shows great potential for semantic segmentation on sparse data. At this point, we cannot evaluate the whole extent of GAN based DA, but we encourage the research community to examine the application of our approach to other fields of research and related use cases. We expect benefits especially in the field of deep learning, industrial monitoring and neuroevolution.

6 CONCLUSION AND OUTLOOK

In this paper, we proposed an image-to-image translation approach for the detection of misaligned fibers and fuzzballs on carbon fiber surfaces. We discussed related GAN approaches and designed a novel concept for generating synthetic defects based on sparse labeled data using a *pix2pix* model.

Within our experiments on four different datasets we showed that the *pix2pix* based approach could substantially improve the pixel-based classification quality of U-Net models. The synthetic defects helped to augment the dataset so that segmentation quality improves significantly on sparse data. However, the approach did not outperform regular DA techniques but still achieves similiar quality scores. Furthermore, the approach can be used to train neural networks for semantic segmentation on comparably sparse data since the GAN manages to generate realistic, yet artifical labels from few samples. At this time, further experimentation is necessary to evaluate the whole potential. It should however be noted, that by using the label generator with mathematical models, training data can always be created for a specific use case as for carbon fiber images. Conventional DA only applied very general image transformations without any reference to specific requirements in the application scenario. We demonstrated that the approach of-

fers great potential for further semi-supervised training and exhibits high relevance for anomaly detection in industrial applications. We also showed that in combination with traditional DA the approach did not improve the pixel-based classification quality further. Under the given conditions the assumption can be made that GAN based augmentation already provides a well-balanced and diverse dataset so that conventional image transformation methods do not add any additional value. As current research activities still focus on adapting semantic segmentation and object detection networks to different domains of industrial image data, our research work will investigate the potential for an extended use of unsupervised learning using GANs. For this purpose, our efforts are concerned with network designs that manage to train models with even less or no input data in order to efficiently solve industrial monitoring tasks. In this spirit, we want to shift the attention from defect generation by designing problem-specific algorithms to a problem-independent approach. The field of neuroevolution, e.g. still remains unexplored for industrial applications. Future work will strive for a closer look to hyperparameter optimization in the context of deep learning. Furthermore we are working on a more generic approach that allows to use GANs not only for the generation of image data, but also for label data that is currently generated by our random label generator. Thus, we try to apply concepts borrowed from the fields of *Evolutionary* and *Organic Computing* to equip our approach with self-configuring and self-learning properties. The application of evolutionary computation, i.e. genetic algorithms and co-evolution, constitutes another topic of our research agenda.

ACKNOWLEDGEMENTS

REFERENCES

Cavigelli, L., Hager, P., and Benini, L. (2017). CAS-CNN: A deep convolutional neural network for image compression artifact suppression. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 752–759.

Choi, J., Kim, T., and Kim, C. (2019). Self-ensembling with gan-based data augmentation for domain adaptation in semantic segmentation. In *Proceedings of the IEEE*

International Conference on Computer Vision, pages 6830–6840.

Di Mattia, F., Galeone, P., De Simoni, M., and Ghelfi, E. (2019). A survey on gans for anomaly detection. *arXiv preprint arXiv:1906.11632*.

Ferguson, M. K., Ronay, A., Lee, Y.-T. T., and Law, K. H. (2018). Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning. *Smart and sustainable manufacturing systems*, 2.

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. (2018). Synthetic data augmentation using gan for improved liver lesion classification. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 289–293. IEEE.

Geinitz, S., Margraf, A., Wedel, A., Witthus, S., and Drechsler, K. (2016). Detection of filament misalignment in carbon fiber production using a stereovision line scan camera system. In *Proc. of 19th World Conference on Non-Destructive Testing*.

Haselmann, M. and Gruber, D. (2017). Supervised machine learning based surface inspection by synthetizing artificial defects. In *2017 16th IEEE international conference on machine learning and applications (ICMLA)*, pages 390–395. IEEE.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Huang, S.-W., Lin, C.-T., Chen, S.-P., Wu, Y.-Y., Hsu, P.-H., and Lai, S.-H. (2018). Auggan: Cross domain adaptation with gan-based data augmentation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 718–731.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

Margraf, A., Stein, A., Engstler, L., Geinitz, S., and Hähner, J. (2017). An evolutionary learning approach to self-configuring image pipelines in the context of carbon fiber fault detection. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE.

Mariani, G., Scheidegger, F., Istrate, R., Bekas, C., and Malossi, C. (2018). Bagan: Data augmentation with balancing gan. *arXiv preprint arXiv:1803.09655*.

Masci, J., Meier, U., Ciresan, D., Schmidhuber, J., and Fricout, G. (2012). Steel defect classification with max-pooling convolutional neural networks. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pages 1–6. IEEE.

McCann, M. T., Jin, K. H., and Unser, M. (2017). Convolutional neural networks for inverse problems in imaging: A review. *IEEE Signal Processing Magazine*, 34(6):85–95.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems 28*, pages 91–99. Curran Associates, Inc.

Rizki, M. M., Zmuda, M. A., and Tamurino, L. A. (2002). Evolving pattern recognition systems. In *IEEE Transactions on Evolutionary Computation*, volume 6, pages 594–609.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., and Langs, G. (2017). Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, pages 146–157. Springer.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Soukup, D. and Huber-Mörk, R. (2014). Convolutional neural networks for steel surface defect detection from photometric stereo images. In *International Symposium on Visual Computing*, pages 668–677. Springer.

Staar, B., Lütjen, M., and Freitag, M. (2019). Anomaly detection with convolutional neural networks for industrial surface inspection. *Procedia CIRP*, 79:484–489.

Stein, A., Margraf, A., Moroskow, J., Geinitz, S., and Haehner, J. (2018). *Toward an Organic Computing Approach to Automated Design of Processing Pipelines*. ARCS Workshop 2018; 31th International Conference on Architecture of Computing Systems. VDE.

Strumberger, I., Tuba, E., Bacanin, N., Jovanovic, R., and Tuba, M. (2019). Convolutional neural network architecture design by the tree growth algorithm framework. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9.

Xie, S. and Tu, Z. (2015). Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1395–1403.

Yakubovskiy, P. (2019). Segmentation models. `https://github.com/qubvel/segmentation_models`.

Zhang, R., Isola, P., and Efros, A. A. (2016). Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer.