Segmentation Aware Attention Mechanism for Defect Classification of both Virgin and Recycled Carbon Fiber Fabric

Abhiram Kolli, Denis Krajnc, Christian Eitzinger
Machine Vision
A-4407 Steyr-Gleink, Austria — Im Stadtgut D1
{akolli, dkrajn, ceitzi}@profactor.at

Abstract

Using neural networks for photometric stereo based surface inspection is fast emerging, especially for carbon fiber matrix (before binding with resins) / fabric. However, the non-rigid nature and high reflectivity of the carbon fiber material makes them very difficult to perform robust digital image based quality inspection. To overcome this issue, 1). a dataset was collected with photometric stereo feature modalities and open sourced, 2). studied the classification performance on the dataset and 3). propose a segmentation supervised multi-head defect classification model. Several standard vision classifier models are retrained on the the created dataset to benchmark the performance of our model. This paper demonstrates that a segmentation supervised multi-head neural network outperforms the benchmark. The realized classifier model achieves over 30% *higher accuracy compared to standard CNNs and transformer models. The dataset for classification and its parser is available at https://zenodo.org/records/11203952.*

1. Introduction

Carbon fiber composites are widely being adopted by various industries such as aerospace, wind turbine, automotive, construction, sports, and leisure [20]. Each industry that uses carbon fiber matrices has its specific stringent requirements to be fulfilled. This emphasizes the need for better industry-scale automated quality inspection methods of the carbon fiber matrix.

Estimating the surface properties of carbon fiber materials without using any intrusive methods is popularly used for quality inspection in the industrial manufacturing sector. Photometric stereo is one such method, where several images of the same material of interest are captured from either different viewing angles or from different light sources. Desired features are computed from these multi-view images to identify the material surface properties such as carbon fiber angles, texture, and fiber reflectivity.

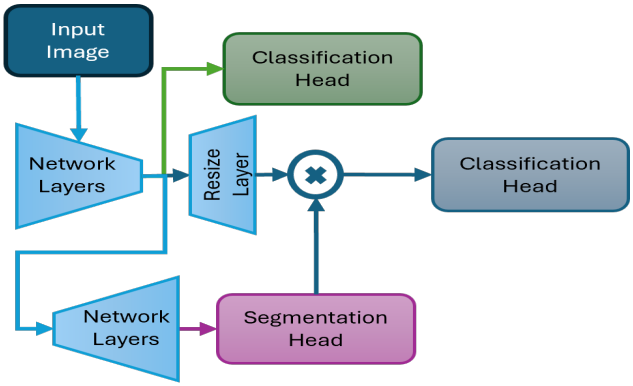


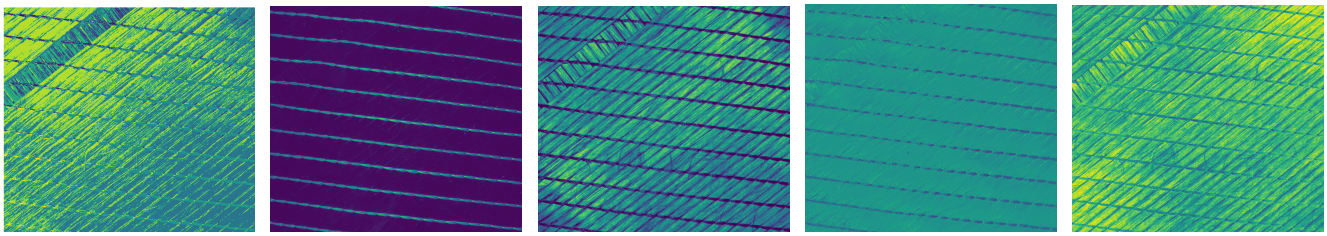
Figure 1. An overview of the proposed segmentation supervised multi-head defect classification model.

However, such features alone are not robust enough when provided as an input to a neural network to classify the carbon fiber defects. One predominant reason identified is the non-rigid nature of the carbon fiber matrices. Due to this, the neural networks either require a lot of training samples, higher graphic card memory, and/or have relatively lower throughput, which (due to higher model complexity) is not suitable for the industrial scale.

To overcome these issues, several models both in transformers and convolutional neural networks (CNNs) were developed. Neither of these models satisfactorily achieve higher accuracy for classifying defects in photometric stereo images. Hence, this paper demonstrates a neural network that has a relatively smaller number of parameters and better accuracy (see Fig. 1).

2. Related work

There are several proposed techniques to use machine learning for defect classification and detection. However, these mechanisms are not viable to implement for an industrial setup and might also not achieve near real-time speed



(a) azimuthal (b) Diffusion (c) Fiber reflectivity (d) Polar (e) Specular

Figure 2. Illustration of all the 5 input feature modalities pre-computed using photometric stereo.

for model inference.

This study aims to streamline the entire process of real-time feature extraction (see Fig. 2) from an onboard computer of the sensor (see Fig. 3) and to perform training and testing of neural networks for carbon fiber defect detection.

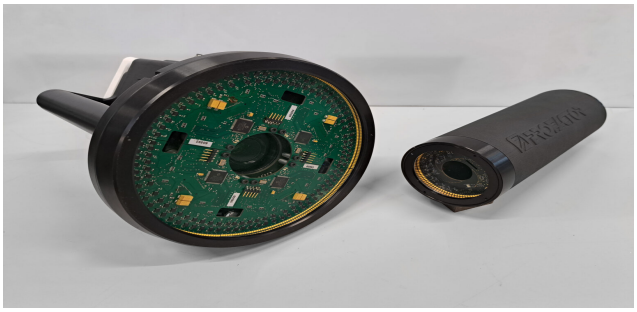


Figure 3. An illustration of in house developed photometric stereo based sensors. Big sensor on the left was used for dataset creation.

2.1. Supervised learning-based defect identification

Supervised defect identification is a traditional machine learning (ML) approach where the neural network is trained with defect samples labeled according to the classes to which they belong. In this approach, the defect samples are collected to be as diverse as possible. The collected sample data are further augmented to obtain better performance on the testing dataset. Some approaches also employ regularization techniques to improve the classifier's performance.

There are several methods to perform defect identification, where the input image can either be directly classified or image segmentation can be performed. In [12], the defects are initially extracted using segmentation through a ResNet-based fully convolutional network. Afterwards, the obtained region is classified through a machine learning process. Such a multi-stage process might lead to higher computation time. Similarly, an image segmentation technique based on the U-Net architecture was used in [18]. However, to overcome the issue of small data size, synthetic sample generation was employed. Meister et al. in [5] proposed best practices to improve synthetic datasets for the fiber layup inspection process. Furthermore, Meister et al. reviewed segmentation-based defect detection in multiple works [9, 8].

Unlike segmentation-based carbon fiber defect detection, in [7], the authors implemented explainable AI by combining CNN and support vector machine (SVM) training for the classification of carbon fiber defects. In [6], the authors compared the classification of defects across various network architectures. Silenzi et al. evaluated ten pretrained classifiers to determine if the samples were defect-free, had recoverable defects, or had non-recoverable defects [13]. Among these classifiers, DenseNet121 achieved the best results with data augmentation. Wen et al. used an object detection mechanism to identify defects in carbon fibers, employing YOLOv5 for defect detection [16].

However, these supervised approaches are heavily reliant on the diversity of training samples, which is not practical to obtain in real-world scenarios. This is because it requires extensive knowledge of the types of defects that might occur in the future and the ability to generate those defects. Otherwise, one must retrain the neural network to include new defect patterns.

2.2. Unsupervised learning-based defect identification

In contrast to supervised approaches, there are no defect samples present in the training set for unsupervised defect identification. This approach is used to overcome the inability to: (1) model defects mathematically, (2) create a neural network that can generalize defect identification, and (3) collect samples of all possible defect patterns that might arise during the layup process in the future.

Szarski et al. proposed vision transformers (ViT) combined with a normalizing flow-based model to classify and detect anomalies in dry carbon fiber textiles [14]. The authors also used interpretability techniques to visualize ViT intermediate layers for human inspection of the results. Brysch et al. in [1] used two Siamese networks for defect segmentation—one network generated a mask while the other performed detection. This process is resource-intensive, and the virtual mask generation influences the detection outcome. Yang et al. used the reconstruction error of autoencoders as an anomaly detection mechanism dur-

ing the layup process [3]. The error threshold value was obtained from the receiver operating characteristic (ROC) curve. This region was then passed through difference of Gaussians (DoG) filters to identify local maxima representing anomaly regions.

These unsupervised approaches offer advantages over supervised methods. Specifically, these models (1) achieve relatively higher accuracy and (2) require fewer training samples. However, models based on this approach tend to generate relatively higher false positives compared to supervised approaches [11].

3. Methodology

3.1. Dataset creation

The dataset was created using an in-house developed sensor (Fig. 3) [19] based on the photometric stereo principle [17]. The sensor captures images at a resolution of $40\,\mu\text{m} - 60\,\mu\text{m}$ with a field of view of $50 \times 50\,\text{mm}$.

An onboard computer computes five feature modalities (azimuthal, diffuse, fiber refractivity, polar, and specular) from the raw images (see Fig. 2). These features are used to estimate carbon fiber orientation and defect properties (e.g., contour, dimension, and area).

The dataset includes non-crimp fabric (NCF) from virgin, recycled, and woven carbon fiber materials. Industry experts identified three defect types: gaps, missing stitches, and fuzzballs. A total of 190 samples were collected: 99 gaps, 37 missing stitches, and 54 fuzzballs. Each sample is represented as a $1080 \times 1200 \times 5$ tensor.

Two datasets were prepared: one for classification and one for segmentation. For classification, each image was labeled with a single defect type to prevent test-time confusion. To reduce class imbalance bias, a manually annotated binary segmentation mask of the predominant defect was created, where pixels are labeled as 1 for defect and 0 for clean regions.

3.2. Machine Learning

Architecture. The proposed model uses a single input and multi-head output design. The input tensor has shape $32 \times 512 \times 512 \times 5$ (batch size, height, width, channels). It is processed through two convolutional branches: one splits the tensor into patches (light blue in Fig. 4), while the other retains the original size (dark blue). The patches are reshaped to $32 \times 64 \times 64 \times 320$, as defined in Eq. (1):

$$N_c = \frac{Height_i}{Height_p} \times \frac{Width_i}{Width_p} \times Channels_i \tag{1}$$

The segmentation output is a weighted sum (alpha blend) of classifier embeddings and the segmentation mask:

$$\text{output}_l = \alpha \cdot \text{o}_r + (1 - \alpha) \cdot \text{o}_s \tag{2}$$

where $\alpha = 0.5 \in [0, 1]$ is the blending factor.

Loss Functions. The model includes two classification heads (cross-entropy loss) and one segmentation head. The segmentation is trained using a hybrid loss (Eq. (3)) as proposed by [10]:

$$L_T = L_{\text{bce}} + L_{\text{ssim}} + L_{\text{iou}} \tag{3}$$

The binary cross-entropy (BCE) loss [2] is:

$$L_{\text{bce}} = -\sum_{(r,c)} \Big[G_{(r,c)} \log(S_{(r,c)}) + (1 - G_{(r,c)}) \log(1 - S_{(r,c)}) \Big] \tag{4}$$

The structural similarity index (SSIM) loss [15] is:

$$L_{\text{ssim}} = 1 - \frac{2(\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{5}$$

The intersection over union (IoU) loss [4] is:

$$L_{\text{IoU}} = 1 - \frac{\sum_{r=1}^{H} \sum_{c=1}^{W} S_{(r,c)} G_{(r,c)}}{\sum_{r=1}^{H} \sum_{c=1}^{W} \left[S_{(r,c)} + G_{(r,c)} - S_{(r,c)} G_{(r,c)} \right]} \tag{6}$$

Performance Evaluation. Due to limited data, an 80-20% train-validation split was used with 100-fold cross-validation. Top-K accuracy was evaluated to identify misclassification severity. Segmentation was evaluated via mean squared error. All experiments were conducted using TensorFlow 2.15 and Keras, on an NVIDIA V100 GPU (16 GB VRAM), with Adam optimizer, learning rate 0.001, batch size 16, 200 epochs, and early stopping (patience: 20).

4. Results

The proposed model was benchmarked against standard architectures (Xception, InceptionV3, MobileNetV2, and CCT) using single-channel input due to GPU memory limitations. Our model significantly outperformed others, as shown in Tab. 1.

4.1. Classification only

Since the standard models take only 3 input channels, all the models in Tab. 1 were trained and validated using 3 features despite our models having 5 channels as input capability. The performance of the classifier-only model outperformed the standard models and reached 77.59% (see Tab. 1). This low performance can be attributed to either 1) the pre-computed photometric stereo features being non-discriminative, 2) hyperparameters might not be tuned for the task, or 3) noisy labels. To test these, several ablation studies were performed.

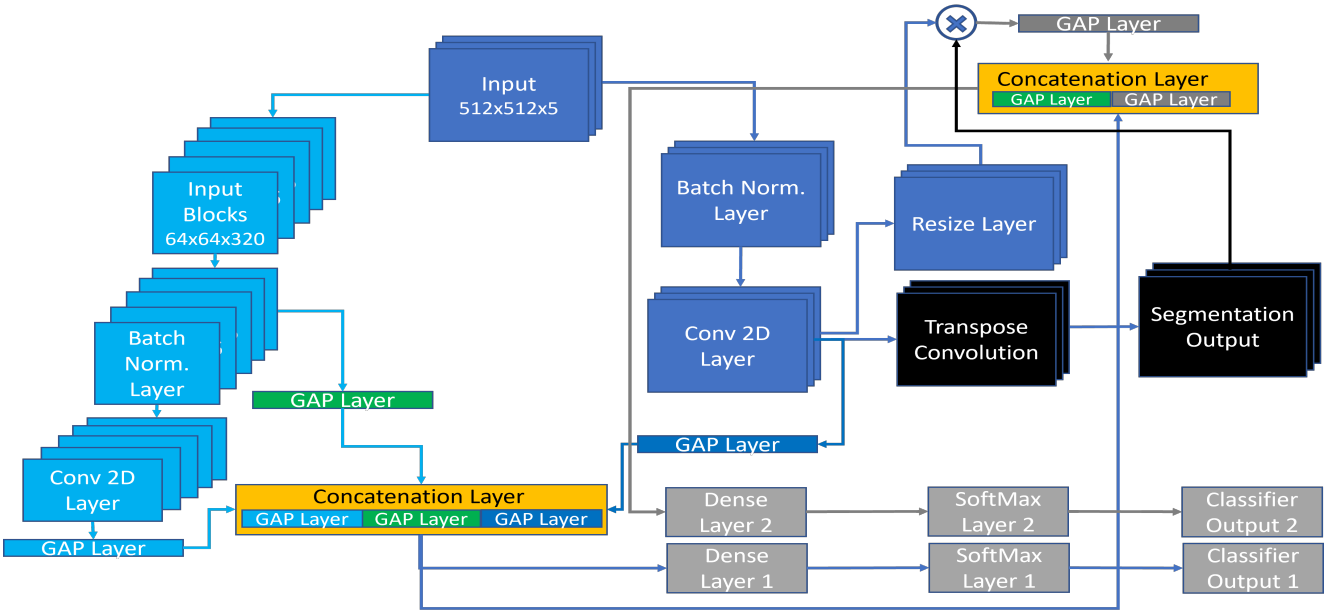


Figure 4. Proposed neural network architecture where the light blue blocks are the patch based convolution blocks, dark blue blocks are the full convolutional filter and the black blocks are segmentation related blocks. Each convolutional blocks output are given to global average pooling and the yeloow blocks are the concatenation layers. This enables the feature diversity. The gray blocks are fully connected neural layers that belong to 2 classifier heads.

Model	Parameters	Accuracy (%)
Xception	22.9M	40.00
InceptionV3	23.9M	60.00
MobileNetV2	3.5M	40.00
CCT	2.68M	42.00
CCN (ours)	36,771	**77.59**

Table 1. Benchmark results on the defect classification task. The highest accuracy is shown in bold.

No. of input channels	Accuracy (%)
3 (pre-computed features)	77.79
4 (pre-computed features)	80.39
5 (pre-computed features)	**82.39**
5 (raw input images only)	70.80
5 + 5 (raw input images)	81.79

Table 2. Accuracy based on the number of feature channels for the network's input layer. The value in bold is the highest.

No. of Channels	Activations	Accuracy (%)
3	ReLU	77.79
4	ReLU	80.39
5	Swish	77.60
5	PReLU	72.60
5	ReLU	**82.39**

Table 3. Performance comparison based on different configurations. Value in bold is the highest.

Tab. 2 shows how the accuracy is influenced by the number of input channels for compact convolutional network (CCN) model. It can be inferred that with the increase in pre-computed features the accuracy of the model increased however inclusion of raw input images aprat from the pre-computed features did not linearly imporove the accuracy of the proposed model.

Tab. 3 studied the effect of activation functions and the number of channels over the accuracy of the designed model. It is evident that the rectified linear unit (ReLU) outperformed the Swish and PReLU activation functions with 5 input channels.

The effect of block size (see, light blue color blocks in Fig. 4) and the dilation rate for all the convolutional blocks were also studied. Since these aspects influence the receptive field of the convolutional filters, it is essential to study

them. From Tab. 4, it is evident that the block size of 64×64 with a dilation rate of $(2, 2)$ and $(2, 5)$ achieve the highest accuracy. Any further increase in the dilation rate did not influence the accuracy of the neural network.

Block size	Dilation rate	Accuracy (%)
32 x 32	(1, 1)	66.80
64 x 64	(1, 1)	80.79
64 x 64	(2, 2)	**81.99**
64 x 64	(2, 5)	**81.99**

Table 4. Classifier performance with varying patch size and convolutional layer's dilation rate. The value in bold is the highest.

Since the global average pooling was used to reduce the number of model parameters, it is essential to determine the robustness of the GAP layer and its impact on the number of dense neurons. It is evident from Tab. 5 that additive white Gaussian noise (AWGN) of 0.01 with 4096 dense neurons has better accuracy over the others.

Channels	Hyper parameters	Accuracy (%)
5	No (128)	81.99
5	0.01 (128)	80.19
5	0.02 (128)	79.99
5	0.01 (4096)	**82.79**
5	0.02 (4096)	79.2

Table 5. Performance comparison based on different configurations. The values in the hyper parameters column are AWGN and within the braces are number of neurons present in the dense layer. The value in bold is the highest.

From the above prior results, it is evident that the accuracy of the neural network has plateaued. To further improve the performance, the network was trained using label smoothing. The label smoothing reduced the classifier's performance significantly (see Tab. 6) even with a smaller smoothing factor of 0.2.

Reduced accuracy with label smoothing indicates that the inter-class distribution of the embedding space is similar. To confirm this, the neural network was trained with top-K accuracy, where $K = 2$. The output of this resulted in the classifier's performance of 96.31% after 100-fold cross-validation (see Tab. 7). Upon careful curation of the dataset, it is evident that only the predominant defect is expert-labeled as the corresponding defect, despite having multiple defects in an image sample.

Channels	Smoothing (α)	Accuracy (%)
4	-	80.39
5	-	**82.39**
4	0.2	65.99
5	0.2	68.80

Table 6. Accuracy based on the number of channels and label smoothing. The value in bold is the highest. Smoothing (α) is the smoothing factor.

4.2. Classification and segmentation

Scenario	Accuracy (%)
top-K Classifier Only	**96.31**
No top-K Classifier Only	82.19
Seg. + Classifier	92.36
Seg. + Classifier 2 w/o ab	95.02
Seg. + Classifier 2 w ab	<u>95.52</u>

Table 7. Mean accuracy for 100-fold cross-validation across different scenarios. The seg. is used as short form for the segmentation included. The value in bold is highest and the underlined value is second highest. w/o ab indicates without alpha blending and w abindicates with alpha blending

Three scenarios for the same architecture (see Fig. 4) are experimented for the segmentation and classification training. For the segmentation task, a segmentation mask label dataset is used. In the first scenario, segmentation and classifier models were trained at the same time as a multi-task self-supervised network with 100-fold cross-validation (see red violin plot in Fig. 6). In the second scenario, the segmented output mask is used as the region of interest for the classifier (see violet violin plot in Fig. 5). In the third scenario, the segmented mask and the network embeddings to the classifier are taken as the weighted sum (see yellow violin plot in Fig. 5). The mean accuracies for 100-fold cross-validation of all the scenarios in Fig. 6 are shown in Table 5. The results of the top-K classifier scenario are not practically useful in identifying all the defects but were rather used for analysis purposes only. Therefore, this work considers the combination of the classifier and segmentation network trained together, which has an accuracy of 92.36% as the true performance metric. The mean square error for the segmentation task is 0.098. Fig. 6 illustrates all three types of defects on different carbon fiber matrices, their corresponding binary masks obtained as the network output,

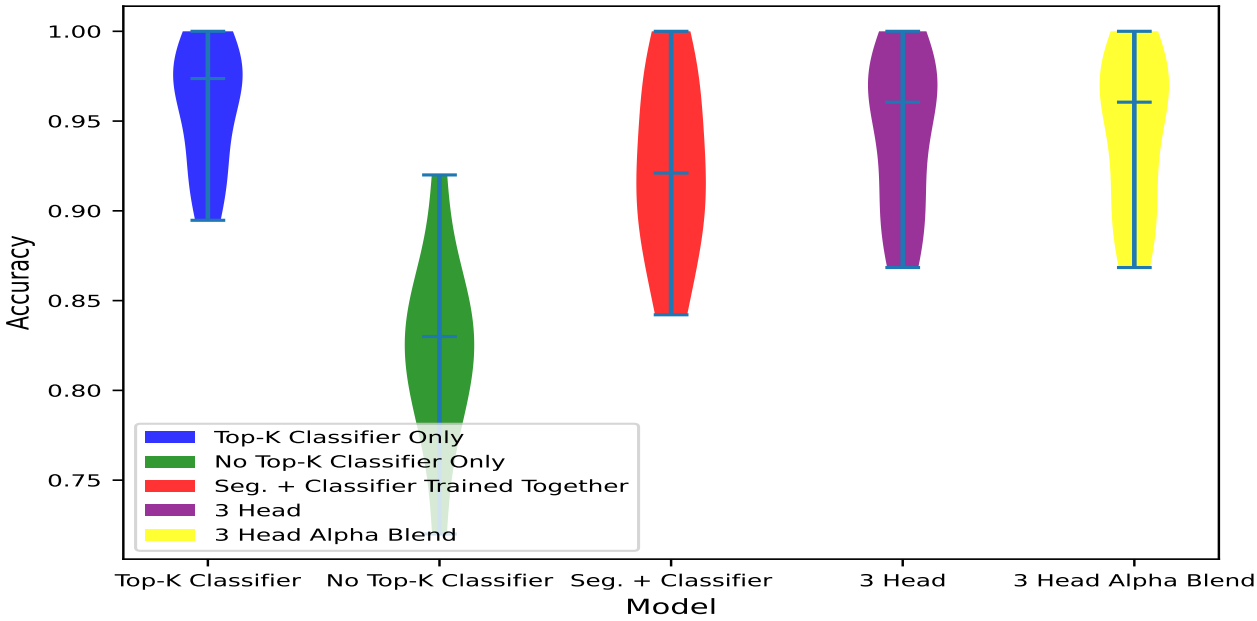
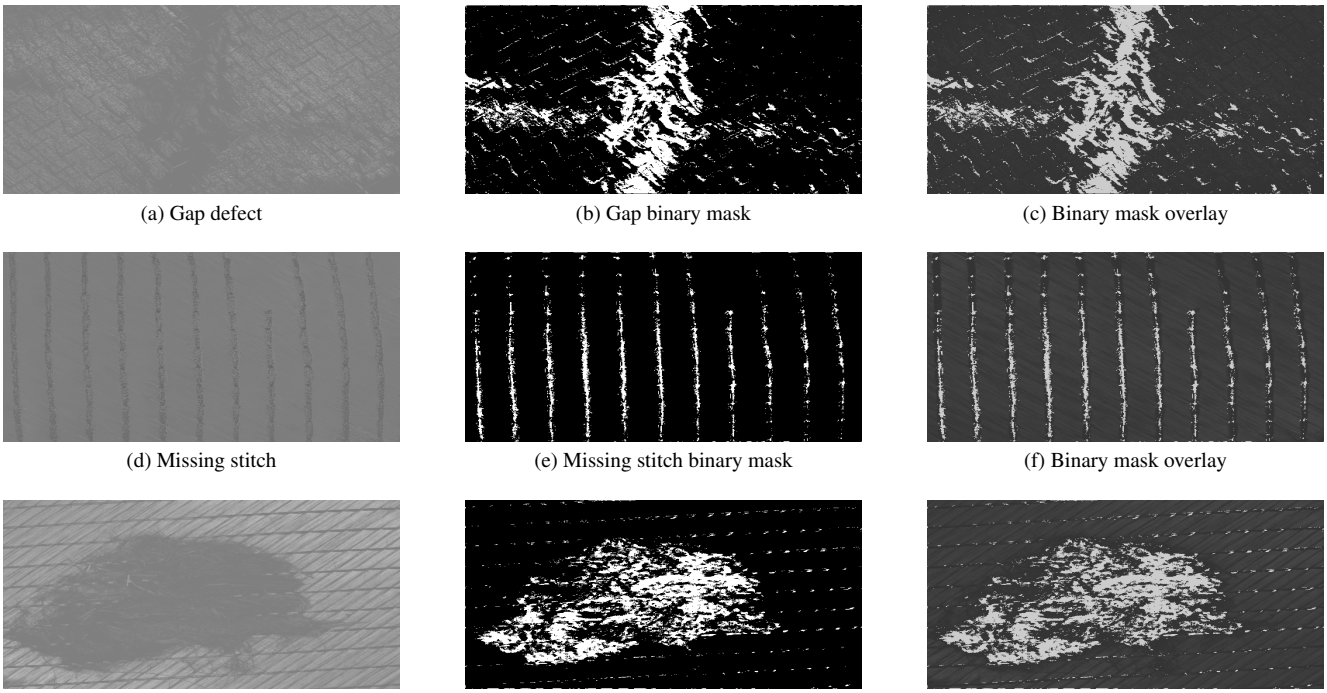


Figure 5. Illustration of violin plots indicating the accuracies with respect to each fold of training and validation. This indicates the characteristics of the network's accuracy for a given dataset.



(a) Gap defect (b) Gap binary mask (c) Binary mask overlay

(d) Missing stitch (e) Missing stitch binary mask (f) Binary mask overlay

(g) Fuzz ball (h) Fuzz ball binary mask (i) Binary mask overlay

Figure 6. Illustration of defect segmentation output, alpha blended with specular feature modality as one of the input channel.

and alpha blending of the input with that of the obtained binary mask. The alpha blend ratios and the input images can further be fine-tuned to obtain better visualization. This visualization, once refined, has the following advantages: 1) helps to visualize the correspondence between the raw input and the defect blob of the network, 2) helps to verify the network's performance, and 3) helps in localizing the defect on the original carbon fiber matrix rather than just viewing an out-of-the-box blob.

4.3. Discussion

This paper has trained lite weight neural networks that are able to classify the defects without relying on data augmentation techniques. This is due to the varying range of values per channel obtained from the in house developed sensor (see, Fig. 3). An optimized network architecture is developed to balance the higher performance and improved speed and lack of data. Such approach improved the performance of our network compared to standard network architectures (CNNs and transformer) by 37%. Ablation studies improved the understanding of the data distribution and labeling errors that are common in the real-world scenarios. Nevertheless, by additionally including the segmentation task, the model's accuracy improved from 77.79% to 92.36% (see Fig. 4, Classifier Output block). Thus segmented output region alone is further classified (see Fig. 4, Classifier Output 2 block) there by achieving 95.52%

5. Conclusion

This work identified the inherent issues with the data labeling method and data mislabeling on this open sourced dataset. Furthermore, these issues were solved systematically through constructive hypothesis and proved through experimentation such as benchmarking with known models (CNNs, transformer), using random search based hyper parameter tuning, label smoothing, distribution of embeddings using top-K, introducing multi-task mechanism such as segmentation and attention mechanism that classifies the segmented embedding. These efforts improved the performance of the network by 10% (82.39% vs 95.52%), emphasizing the importance of data quality and the necessary strategies.

6. Acknowledgment

This research was conducted within the framework of MC4 and COMPASS projects.

The work presented in this paper is related to the project "MC4" and has received funding from the European Union's Horizon Europe research and innovation program under grant agreement No 101057394. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Co-funded by the European Union

COMPASS Project is co-funded by the European Commission within the Horizon Europe Programme Views, grant number 101136940, and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] Marco Brysch, Mohammad Bahar, Hans Christoph Hohensee, and Michael Sinapius. Single system for online monitoring and inspection of automated fiber placement with object segmentation by artificial neural networks. *Journal of Intelligent Manufacturing*, 33(7):2013–2025, 2022.

[2] Pieter-Tjerk De Boer, Dirk P Kroese, Shie Mannor, and Reuven Y Rubinstein. A tutorial on the cross-entropy method. *Annals of operations research*, 134:19–67, 2005.

[3] Assef Ghamisi, Todd Charter, Li Ji, Maxime Rivard, Gil Lund, and Homayoun Najjaran. Anomaly detection in automated fibre placement: Learning with data limitations. *Frontiers in Manufacturing Technology*, 4:1277152, 2024.

[4] Gellért Máttyus, Wenjie Luo, and Raquel Urtasun. Deeproadmapper: Extracting road topology from aerial images. In *Proceedings of the IEEE international conference on computer vision*, pages 3438–3446, 2017.

[5] Sebastian Meister, Nantwin Möller, Jan Stüve, and Roger M Groves. Synthetic image data augmentation for fibre layup inspection processes: Techniques to enhance the data set. *Journal of Intelligent Manufacturing*, 32(6):1767–1789, 2021.

[6] Sebastian Meister and Mahdieu Wermes. Performance evaluation of cnn and r-cnn based line by line analysis algorithms for fibre placement defect classification. *Production Engineering*, 17(3):391–406, 2023.

[7] Sebastian Meister, Mahdieu Wermes, Jan Stueve, and Roger M Groves. Cross-evaluation of a parallel operating svm–cnn classifier for reliable internal decision-

making processes in composite inspection. *Journal of Manufacturing Systems*, 60:620–639, 2021.

[8] Sebastian Meister, Mahdieu AM Wermes, Jan Stüve, and Roger M Groves. Review of image segmentation techniques for layup defect detection in the automated fiber placement process: A comprehensive study to improve afp inspection. *Journal of Intelligent Manufacturing*, 32(8):2099–2119, 2021.

[9] Sebastian Meister, Mahdieu Amin Mahdieu Wermes, Jan Stüve, and Roger M Groves. Algorithm assessment for layup defect segmentation from laser line scan sensor based image data. In *Sensors and smart structures technologies for civil, mechanical, and aerospace systems 2020*, volume 11379, pages 139–153. SPIE, 2020.

[10] Xuebin Qin, Deng-Ping Fan, Chenyang Huang, Cyril Diagne, Zichen Zhang, Adrià Cabeza Sant'Anna, Albert Suarez, Martin Jagersand, and Ling Shao. Boundary-aware segmentation network for mobile and web applications. *arXiv preprint arXiv:2101.04704*, 2021.

[11] Ji Qiu, Hongmei Shi, Yuhen Hu, and Zujun Yu. Unraveling false positives in unsupervised defect detection models: A study on anomaly-free training datasets. *Sensors*, 23(23):9360, 2023.

[12] Christopher Sacco, Anis Baz Radwan, Andrew Anderson, Ramy Harik, and Elizabeth Gregory. Machine learning in composites manufacturing: A case study of automated fiber placement inspection. *Composite Structures*, 250:112514, 2020.

[13] Andrea Silenzi, Vincenzo Castorani, Selene Tomassini, Nicola Falcionelli, Paolo Contardo, Andrea Bonci, Aldo Franco Dragoni, and Paolo Sernani. Quality control of carbon look components via surface defect classification with deep neural networks. *Sensors*, 23(17):7607, 2023.

[14] Martin Szarski and Sunita Chauhan. An unsupervised defect detection model for a dry carbon fiber textile. *Journal of Intelligent Manufacturing*, 33(7):2075–2092, 2022.

[15] Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. Ieee, 2003.

[16] Liwei Wen, Shihao Li, and Jiajun Ren. Surface defect detection for automated tape laying and winding based on improved yolov5. *Materials*, 16(15):5291, 2023.

[17] Robert J Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *Image understanding systems and industrial applications I*, volume 155, pages 136–143. SPIE, 1979.

[18] Sebastian Zambal, Christoph Heindl, Christian Eitzinger, and Josef Scharinger. End-to-end defect detection in automated fiber placement based on artificially generated data. In *Fourteenth international conference on quality control by artificial vision*, volume 11172, pages 371–378. SPIE, 2019.

[19] Sebastian Zambal, Werner Palfinger, Matthias Stöger, and Christian Eitzinger. Accurate fibre orientation measurement for carbon fibre surfaces. *Pattern Recognition*, 48(11):3324–3332, 2015.

[20] Jin Zhang, Gang Lin, Uday Vaidya, and Hao Wang. Past, present and future prospective of global carbon fibre composite developments and applications. *Composites Part B: Engineering*, 250:110463, 2023.