Probabilistic Modelling combined with a CNN for boundary detection of carbon fiber fabrics

1st Sebastian Zambal
Machine Vision
PROFACTOR GmbH
4407 Steyr, Austria
sebastian.zambal@profactor.at

2nd Christoph Heindl
Visual Computing
PROFACTOR GmbH
4407 Steyr, Austria
christoph.heindl@profactor.at

3rd Christian Eitzinger
Machine Vision
PROFACTOR GmbH
4407 Steyr, Austria
christian.eitzinger@profactor.at

***Abstract*—For many industrial machine vision applications it is difficult to acquire good training data to deploy deep learning techniques. In this paper we propose a method based on probabilistic modelling and rendering to generate artificial images of carbon fiber fabrics. We deploy a convolutional neural network (CNN) to learn detection of fabric contours from artificially generated images. Our network largely follows the recently proposed U-Net architecture. We provide results for a set of real images taken under controlled lighting conditions. The method can easily be adapted to similar problems in quality control for composite parts.**

***Index Terms*—Deep learning, probablistic modelling, U-Net**

I. Introduction

In this work we adress the problem of automatically detecting fabric boundaries. We do this based on images that are acquired with a photometric stereo vision system.

For the production of parts made of carbon fiber reinforced plastics (CFRP) there exist different production technologies. Most of these include a processing step in which carbon fiber material is placed, layer by layer. During this lay-up, it is important to monitor fabric boundaries in order to detect defects due to misalignment of individual fabric patches.

Inspection of surfaces in production by means of computer vision poses multiple challenges. In many cases, the exact surface properties and the relevant defects, deviations, or features to be detected are very domain-specific. Expert knowledge might be required for correct interpretation of images. Furthermore, it is often difficult to acquire a sufficiently large amount of representative data at the development stage of a vision-based monitoring system. Other challenges are related to the fact that physical accessibility (a test installation) might be difficult. In addition, machine time can be very costly and as a result data for feasibility studies can only be acquired to a limited extent. In general, it might be necessary to keep disturbance of running production at a minimum. Furthermore, when data is acquired for evaluation of a surface inspection system, specific defects might only occur under certain conditions. When acquiring data for system validation, it is often difficult to set up conditions such that the recorded data covers all possible interesting cases. The above issues make it difficult to transfer powerful deep learning methods to the specific problem at hand, simply because of lack of good data. Furthermore, it is often not clear how existing expert knowledge and knowledge about the production process can be included into the design of a vision system.

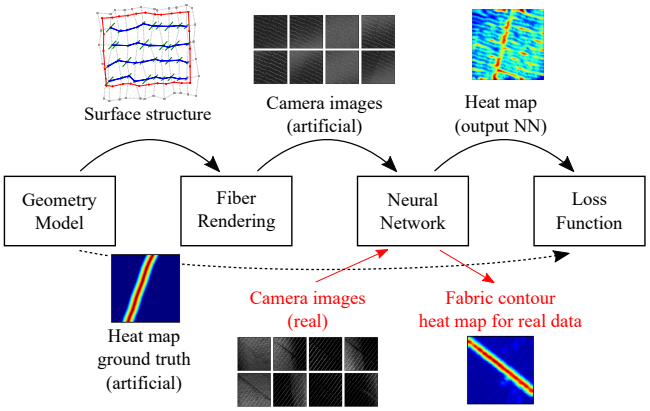

Fig. 1. Overview of our approach: A probabilistic model generates artificial surface geometry. A rendering stage converts this to artificial images under different simulated lighting directions. A neural network is trained on this artificial data. The loss is calculated for heat maps as output by the neural network and ground truth heat maps from the initial modelling stage. For the final application, the neural network is applied to real surface images (red arrows).

We are not aware of any prior methods for the very specific task of boundary detection of carbon fiber fabrics. For solving similar computer vision problems, the conventional approach is to select a set of image features and define simple thresholds to detect interesting regions in the image. However, this typically is very error-prone and does not generalize well. Machine learning methods have shown impressive performance on various computer vision tasks in the recent past. However, the lack of large amounts of data is a real bottleneck for many special applications.

Fig. 1 provides an overview about our method. We propose a probabilistic geometry model to generate surface structure of non-crimp carbon fiber fabrics. A fiber rendering process converts generated geometries into simulated photometric stereo images. Simulated images are fed into a deep neural network. The neural network is trained to learn heat maps of fabric boundaries via a loss function that takes into account artificial ground truth heat maps. The performance of the

neural network for fabric boundary detection is evaluated on real photometric stereo images.

II. Related Work

For optical analysis of carbon fiber surfaces two different methods have been applied in the past. The first builds upon polarization of light reflected at carbon fibers [1]. The second is based on a photometric stereo approach for carbon fibers [2], [3]. In the present work we build upon the second method for image acquisition. However, we do not use reflection properties to calculate fiber orientations or other features. Instead, we deploy the reflection model of carbon fibers to generate artificial images. These are then used to train a neural network to perform fabric boundary detection. In section III-D we explain in more detail the relationship of our method to the photometric stereo approach.

Generative models are currently extensively studied by a large community of researchers. We see two main concepts in this context: data-driven generative models and hand-crafted modelling. Typical approaches that follow the data-driven concept are variational auto-encoders (VAE) [4] and generative adversarial networks (GAN) [5]. Interesting extensions of these approaches are Cycle GANs [6], and Balancing GANs [7]. In the end all of these models are highly data-driven. Hand-crafted modelling on the other hand is successfully used in very domain-specific applications where data acquisition is difficult. Examples are computer-generated images for machine learning in autonomous driving [8], [9], object localization [10], [11], or tracking of honeybees [12]. In hand-crafted modelling the focus is more on manually specifying the data generating process.

There is no sharp border between data-driven and hand-crafted modelling. An example for the mixture of data-driven and hand-crafted concepts is an application for detection of human hands [13]. Hand-crafted elements of the model include a 3d mesh of a hand and rendering of synthetic images. For conversion of synthetic images to pseudo-real images an approach inspired by Cycle GANs [6] is used. The latter requires a large amount of hand images for training. However, a special loss needs to be added to avoid geometric deviations. Data augmentation is also a conecpt that relates to generative models. It can be understood as a hand-crafted modification (e.g. random geometric transformations, smooth deformations, etc.) of plain data.

The method outlined in this work focuses on a hand-crafted model. Hence, our method is conceptually very similar to domain randomization [9], [11]. The advantage compared to data-driven methods is that little or no data is required. This is very interesting for applications like ours, where data acquisition and labeling is very costly. Furthermore, domain knowledge can be included in the generative model. This is difficult in strongly data-driven approaches.

An important aspect of our method is related to rendering of carbon fibers. Carbon fibers reflect light in a very special anisotropic way. Therefore, our model needs to take this into account for image synthesis. Modeling of reflection properties was also done in work on surface material characterization [14]. This method makes use of a database of bi-directional texture functions (BTF) to synthesize images of different materials. Machine learning methods are used to infer materials based on artificial images. In a similar way we make use of a carbon fiber light reflection model to synthesize images. Our method relies on a deep neural network to infer hidden variables (boundary contours) of our model.

In order to learn inference of hidden parameters in our model (fabric boundaries), we deploy a deep neural network. For our purpose, so called image-to-image networks, such as DeepLab [15] or U-Net [16], are suitable. In this work, we use a slightly modified version of the recently proposed U-Net architecture [16].

III. A model for non-crimp fabrics

Our method aims at creation of artificial images that mimic real images of a visual inspection system for carbon fiber surfaces. Our model consists of a first part which defines a probabilistic model of carbon fiber surface structure. Once the structure is sampled, a set of artificial images with different virtual light source positions are generated.

A. Fabric structure modeling

Here, we focus on a specific type of so called non-crimp fabrics (NCF). In this type of fabric, carbon fibers are aligned in a single direction. In order to keep the fibers in place, they are stitched together with a narrow sewing yarn. Compared to woven material, carbon fibers in NCF undergo less distortion. This has a positive effect on the composite part's mechanical stability. Example photometric stereo images of NCF material with sewing yarns running at 45° relative to the carbon fibers are shown on the left side of Fig. 9. In the following, we stick to this type of material, although the general approach can easily be extended to other fabric types.

We understad the generation of the geometric alignment of carbon fibers and sewing yarn as a sampling process from a bayesian network. Sampling from the model involves a sequence of draws from random variables which are conditioned on their parent variables. This approach to probabilistic modelling helps to break down the probability distribution of a complex process by specifying its constituent "local" conditional probability distributions. In order to define conditional probabilities we use standard probability distributions, e.g. uniform or normal distribution. Parameters for these distributions are defined by a domain expert.

Due to the directed structure of bayesian networks, a top-down approach for modelling is targeted. On the top level we model a piece of fabric as a grid of points on a plane. These points define the locations where the sewing yarn penetrates the fabric. Grid points are initially evenly spaced along x- and y-direction of the grid. In a second step points are displaced randomly. This accounts for imperfections in real fabrics. The outer contour of the fabric is defined by additional points interpolated from the outer two rows of grid points. Finally,

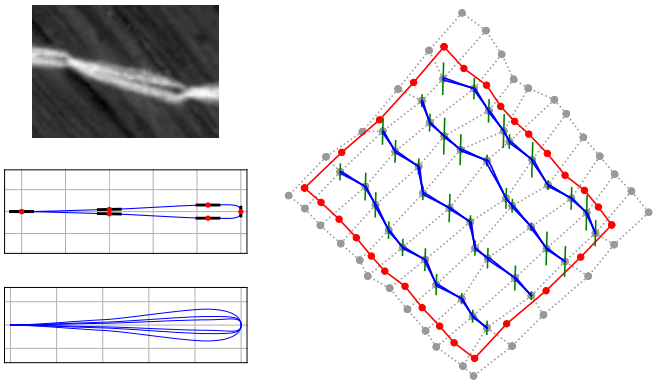

Fig. 2. Left: photo of real stitching yarn loop (top), control points of bezier curve (center), and three examples for different shapes of loops (bottom). Right: Example of generated structure of carbon fiber fabrics with grid structure (gray), fabric boundary (red), sewing yarns (blue), and small gaps where sewing yarn penetrates the top layer (green).

the complete grid is rotated by an angle that is uniformly distributed in the range [-π, +π].

Sewing yarns form small elongated loops between the locations where they penetrate the fabric. We model these loops as simple bezier curves defined by start- and end-point together with 4 additional points on both sides of the line that connects start- and end-point. Fig. 2 (left) illustrates the geometric model for these loops.

At the locations where sewing yarns penetrate the surface, carbon fibers are slightly pushed to the side. Therefore, carbon fibers of the layer underneath become visible along an elongated region around the grid point. The main direction of these regions is along the fiber direction of the top layer. Fig. 2 (right) shows an example for a single sample of the complete geometry of a small fabric patch. Individual structures (sewing yarns, sewing yarns, small gaps) are each represented by polygonal contours.

B. Fibrous material reflection model

In this sub-section we discuss how the above mentioned geometry is converted to images that are (ideally) close to real images of real surfaces acquired with a real camera system. An important aspect of this rendering process is to use a reasonable reflection model of fibrous material. Typically, carbon or glass fiber surfaces reflect light in a highly anisotropic specular way. Two angles in the geometric configuration of surface point, optical center of the camera, and light source determine how much light is reflected into the camera:

1) Angle α between the vector from surface point to the camera's optical center and the vector pointing in the direction of the fiber on the surface.
2) Angle β between the vector from surface point to light source and the vector pointing in the direction of the fiber on the surface.

When α and β are equal, the maximum amount of light is reflected towards the camera. When α and β are different, the amount of light reflected into the camera decreases quickly. We model this relationship with a simple exponential function. The brightness b of a pixel that shows fiber material is proportional to

$$b \propto e^{-|\alpha-\beta|\cdot k}, \tag{1}$$

where k describes the decay of reflection intensity with increasing difference of α and β (in radian). We use an empirical value of $k = 10$. Fig. 3 illustrates this relationship. For this reflection model the maximum reflection intensity is not only reached for a single configuration. In fact, α and β are equal for any configuration where the light source is located on the surface of a cone. This cone is made up of all lines that run through the observed surface point and form an angle equal to α with the carbon fiber. The carbon fiber is the symmetry axis of that cone.

The described model is only valid if the diameter of fibers is less than the size of the region covered by a single camera pixel. Furthermore, the above model explains only the fibrous specul reflection. An interesting extension of our method could involve learning of BRDFs [17] for more realistic rendering.

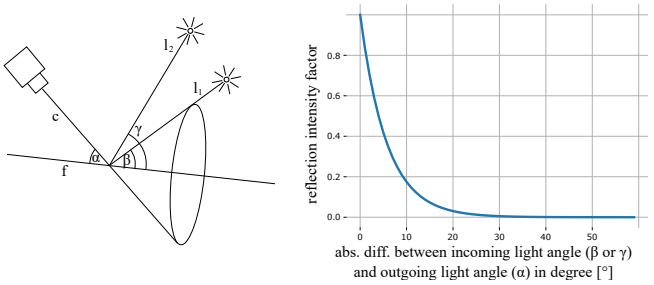

Fig. 3. The reflection model for rendering specular fibrous materials. Intensity of the reflection depends on the absolute difference of angles between vectors from surface point to camera (c) and surface point to light source (l_1, l_2). The relationship between difference angle and reflection intensity is approximated by a simple exponential function.

C. Rendering of fibrous material

Once the arrangement of carbon fibers and sewing yarn on the fabric surface is fixed, this geometric representation is converted to artificial images. The simulated surface is described by three feature maps: azimuthal angle, diffuse reflectivity, and specular reflectivity. Rasterization of the simulated geometry contours is done to fill dense feature maps. For example, points along the sewing yarn are assigned high values for diffuse reflectivity, low values for specular reflectivity, and an azimuth angle that corresponds to the main orientation of the sewing yarn. After rasterization of the geometric description, a set of surface property maps or textures is available: diffuse map, specular map, fiber orientation map.

In principle, additional maps could be generated to enhance surface description. One example is height deviation. Also, isotropic specularity could be useful to be added for modelling foreign objects (e.g. small metallic pieces).

Based on surface property maps, artificial photometric stereo images are rendered. The artificial photometric stereo images corresponding to the single fabric illustrated in Fig. 2

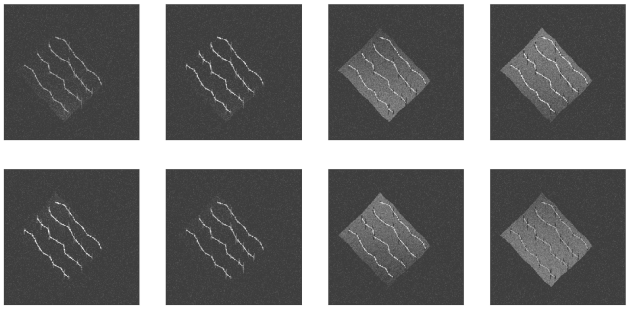

Fig. 4. Artificial photometric stereo images of a single carbon fibre fabric patch. Each image corresponds to a different light source. The images correspond to the geometry illustrated in Fig. 2.

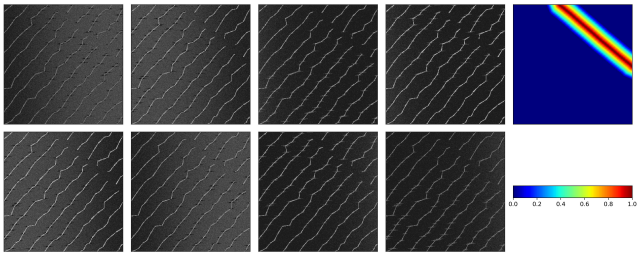

Fig. 5. Example for complete training sample: Artificial photometric stereo raw images (left) with contour heat map (right).

(right) are shown in Fig. 4. A complete sample from our model is shown in Fig. 5. For training data we render a bottom fabric patch that fills the complete image. On top of this we render one or two additional smaller patches. Therefore, all synthetic images contain borders at the transitions from top layers to bottom layers.

In addition to the rendered images, it is straight forward to calculate also a heat map for fabric contours. Fabric contours present in the corresponding geometry sample are rasterized into a heat-map image. In combination with a distance transform, the sharp contours are converted to smooth heat maps with the maximum value of 1 at the very contour and decreasing intensity at points with increasing distance to the contour. The heat map is zero for all points that have more than a threshold distance to the contour. Fig. 5 (right) shows an example of such a heat map. When training a neural network, rendered images are used as input and heat maps are forwarded to the loss function. They are used as ground truth for comparison with heat maps generated by the neural network.

D. Relationship to Photometric Stereo for Carbon Fibers

Before we proceed with describing the use of a neural network for detection of fabric gaps, we briefly describe the photometric stereo vision system. The real data which we use for evaluation comes from such a system. We also outline the concept of photometric stereo for fiber reflection analysis (FRA) [2]. In some sense, the method proposed in this paper is inverse to FRA. In FRA the surface properties (fiber orientation, diffuse reflectivity, etc.) are calculated from a set of differently illuminated images. In our approach differently illuminated images are synthesized based on a probabilistic model of the surface.

FRA is based on a very similar reflection model as outlined above. Surface properties (fiber orientation, diffuse reflectivity, etc.) are calculated from given raw images that are illuminated from different directions. Eight raw images are acquired by the photometric stereo system. The number of eight images represents a good trade off between speed (few images) and information content (many images).

An image processing pipeline based on FRA typically attempts to detect relevant findings on the surface via thresholding of features of the output of FRA (fiber orientation, diffuse reflectivity, etc., etc.). However, it is not straight forward to set up criteria for identification of some findings based on FRA output. Fabric boundaries are one example for that. The output of FRA for a surface region with fabric boundaries is shown in Fig. 6. While the fabric boundary is visible in all feature images, it is not easy to manually define criteria for automatic detection.

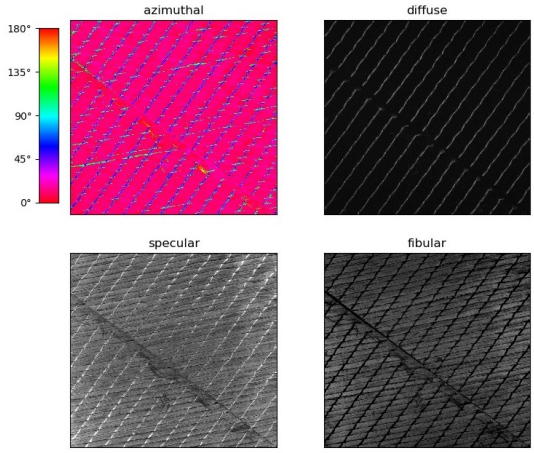

Fig. 6. Image modalities calculated based on a special reflection model. Although these modalities show specific properties of the surface (e.g. azimuth angle), the problem of specifying critical defects in terms of these features remains a challenge.

Deep learning methods have proven very powerful for object detection. End-to-end learning removes the need to manually define thresholds for features. We suggest to use these methods for the above photometric stereo system. In principle, it would be possible to consider FRA features as input for a deep neural network. However, we argue that these features do not contain more information than the original raw photometric stereo images. Therefore, we completely skip FRA features. Instead, we use only raw photometric stereo images as input to our deep neural network.

IV. Neural Network for Segmentation

We identify the carbon fabric boundary detection as image segmentation problem. Therefore, we adapt an existing deep neural network architecture, the U-Net [16], for our application. The U-Net network architecture implements the

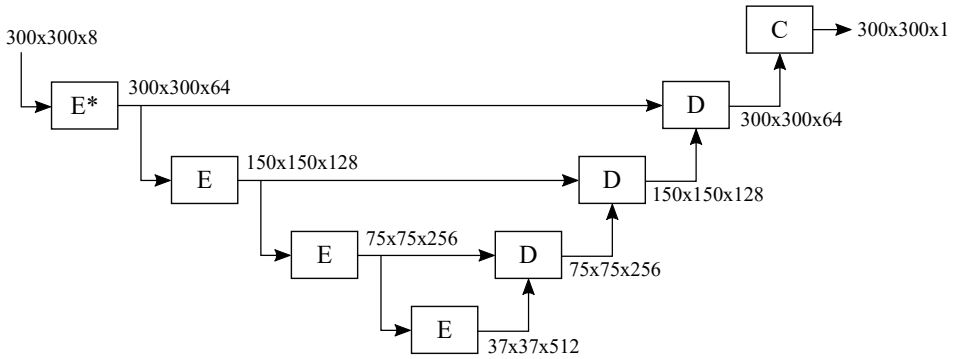

Fig. 7. Neural network architecture for segmentation. The numbers specify input/output of building blocks as: height x width x channels. The numbers given in this figure are only examples. The given network architecture supports input images of any size (except very small images).

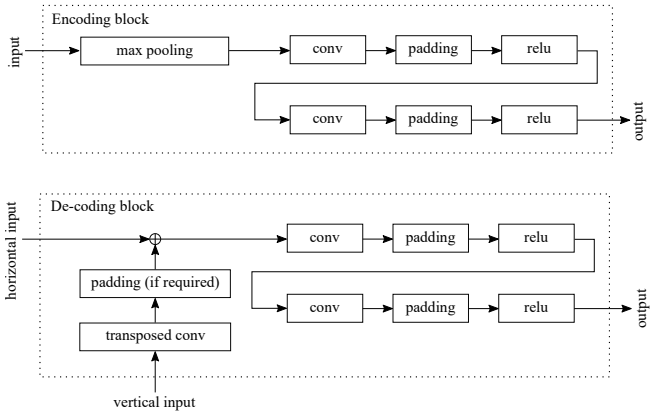

Fig. 8. Internal structure of encoding and de-coding blocks of the neural network.

idea of using features at multiple resolutions: small details on high resolutions and context information on coarse resolutions. The flow of information on the one hand follows a down-sampling/upsampling path. On the other hand, information is forwarded across stages with equal resolution. Therefore, details and context are both forwarded to the final output of the neural network.

Fig. 7 shows an overview of our network architecture. In the original U-Net design, the output image size is smaller than the input image size. In our implementation we add padding layers to prevent this image shrinkage. Therefore, the output images of our network are equal in size to the input images. Blocks denoted E represent encoding blocks. Blocks marked with D are de-coding blocks. The final convolution to convert the output to the required set of channels is marked with C. An encoding layer consists of the following sequence of operations: max pooling, convolution, padding, relu, convolution, padding, relu. At the initial encoding layer E^*, the max pooling operation is skipped.

Details of encoding and de-coding layers are shown in Fig. 8. De-coding layers have two inputs: an input from the encoding layer of the same level of image resolution (horizontal input) and an input with smaller spatial size from the de-coding layer from one level below (vertical input). A transposed convolution is applied to the low-resolution input to make it the same size as the high-resolution input. In some cases (odd image size), it is necessary to add a padding layer to make sure that size of vertical input and output of transposed convolution are equal in width and height. The remaining steps in the de-convolution block are: concatenation of horizontal and up-scaled vertical inputs, convolution, padding, relu, convolution, padding, and relu.

Nice properties of the network architecture described above are that (1) the image input size is equal to the output size and (2) the network is fully convolutional. The latter property makes it possible to apply the network to images of any size (except very small images).

The input to our network are photometric stereo raw images as shown in Fig. 9 (gray-scale images on the left). We are using 8 photometric images as individual channels for the network input. The complete dimensions of the input therefore are: $B \times C \times H \times W$ with batch size B, raw image channels $C = 8$, and image size $H \times W$

The output of the network comes from a final convolutional layer. Depending on the output size of this convolutional layer, a varying number of output feature maps may be calculated. In this work we use only a single output channel as heat map for fabric contours. Training data contains ground truth heat maps with values between 0 and 1, where a value of 1 is present at the very edge of the fabric. Ground truth and network output are compared in the loss function. For each pixel, the loss is calculated as the squared difference between ground truth and network output. The total loss is the average over individual pixel losses.

V. Results

We evaluate our method on two fabric patches of size 300x300mm. Regions that show edge boundaries on these surfaces were randomly chosen and captured with a photometric stereo vision system. The complete data set is publicly available online[1].

[1]https://zenodo.org/record/3237980

We provide a qualitative evaluation of results obtained with the described method. Fig. 9 shows four selected examples of results achieved with the above method. For the first two examples shown, the fabric contour was detected without any false positives in the image. In the third example our method failed to detect the contour boundary running to the left of the image. Also, some false positive high values in the heat map occur. In the bottom image the contour region was detected quite well, but again there is a small spot of false positive boundary.

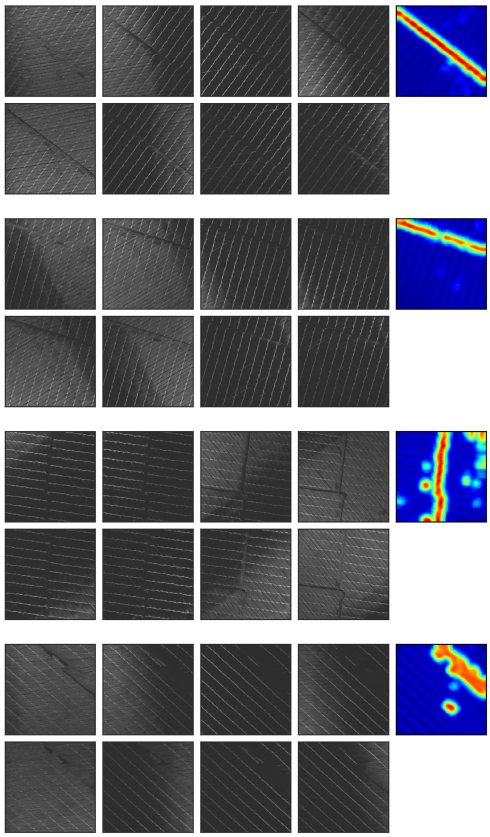

Fig. 9. Segmentation results for four selected examples. Raw input images are shown on the left. Output (heat map for contours) of the neural network (trained exclusively on artificial data) is shown on the right.

VI. Conclusions and Future Work

The method presented in this paper uses a man-made generative model in combination with a deep neural network. The study presented in this paper shows that such an approach can work in principle. We strongly believe that modelling of training data can open the door for many highly specialized machine vision application in industry. In future work we plan to extend this method in two main directions: (1) extending and improving the modelling process and (2) including additional features and surface defects to be detected by the method.

Acknowledgment

Work presented in this paper has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 721362 (project "ZAero") and by the European Union in cooperation with the State of Upper Austria within the project "Investition in Wachstum und Beschäftigung" (IWB).

References

[1] M. Schöberl, K. Kasnakli, and A. Nowak, "Measuring strand orientation in carbon fiber reinforced plastics (cfrp) with polarization," in *19th World Conference on Non-Destructive Testing 2016*, 2016.

[2] W. Palfinger, S. Thumfart, and C. Eitzinger, "Photometric stereo on carbon fiber surfaces," in *35th Workshop of the Austrian Association for Pattern Recognition*, 2011.

[3] S. Zambal, W. Palfinger, M. Stöger, and C. Eitzinger, "Accurate fibre orientation measurement for carbon fibre surfaces," *Pattern Recognition*, vol. 48, no. 11, pp. 3324–3332, 2015.

[4] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.

[5] I. J. Goodfellow, J. Pouget-abadie, M. Mirza, B. Xu, D. Warde-farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *In NIPS*, 2014.

[6] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *IEEE International Conference on Computer Vision (ICCV)*, 2017.

[7] G. Mariani, F. Scheidegger, R. Istrate, C. Bekas, and A. C. I. Malossi, "BAGAN: data augmentation with balancing GAN," *CoRR*, vol. abs/1803.09655, 2018.

[8] M. Johnson-Roberson, C. Barto, R. Mehta, S. N. Sridhar, K. Rosaen, and R. Vasudevan, "Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks?," in *2017 IEEE International Conference on Robotics and Automation, ICRA 2017, Singapore, Singapore, May 29 - June 3, 2017*, pp. 746–753, 2017.

[9] A. Prakash, S. Boochoon, M. Brophy, D. Acuna, E. Cameracci, G. State, O. Shapira, and S. Birchfield, "Structured domain randomization: Bridging the reality gap by context-aware synthetic data," in *International Conference on Robotics and Automation (ICRA)*, 2019.

[10] A. Krull, E. Brachmann, F. Michel, M. Ying Yang, S. Gumhold, and C. Rother, "Learning analysis-by-synthesis for 6d pose estimation in rgb-d images," in *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[11] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*, pp. 23–30, IEEE, 2017.

[12] L. Sixt, B. Wild, and T. Landgraf, "Rendergan: Generating realistic labeled data," *Frontiers in Robotics and AI*, vol. 5, p. 66, 2018.

[13] F. Mueller, F. Bernard, O. Sotnychenko, D. Mehta, S. Sridhar, D. Casas, and C. Theobalt, "Ganerated hands for real-time 3D hand tracking from monocular RGB," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 49–59, 2018.

[14] M. Weinmann, J. Gall, and R. Klein, "Material classification based on training data synthesized using a BTF database," in *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III*, pp. 156–171, 2014.

[15] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2018.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015 - 18th International Conference Munich, Germany, October 5 - 9, 2015, Proceedings, Part III*, pp. 234–241, 2015.

[17] R. Vidaurre, D. Casas, E. Garces, and J. Lopez-Moreno, "BRDF estimation of complex materials with nested learning," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1347–1356, 2019.