Explainability of deep learning classifier decisions for optical detection of manufacturing defects in the Automated Fiber Placement process

Meister, Sebastian; Wermes, Mahdieu A.M.; Stuve, Jan; Groves, Roger M.

DOI
10.1117/12.2592584

Publication date
2021

Document Version
Final published version

Published in
Automated Visual Inspection and Machine Vision IV

Explainability of deep learning classifier decisions for optical detection of manufacturing defects in the automated fiber placement process

Meister, Sebastian, Wermes, Mahdieu A., Stüve, Jan, Groves, Roger

Sebastian Meister, Mahdieu A. M. Wermes, Jan Stüve, Roger M. Groves, "Explainability of deep learning classifier decisions for optical detection of manufacturing defects in the automated fiber placement process," Proc. SPIE 11787, Automated Visual Inspection and Machine Vision IV, 1178705 (20 June 2021); doi: 10.1117/12.2592584

SPIE.

Event: SPIE Optical Metrology, 2021, Online Only

Explainability of deep learning classifier decisions for optical detection of manufacturing defects in the Automated Fiber Placement process

Sebastian Meister[a], Mahdieu A. M. Wermes[a], Jan Stüve[a], and Roger M. Groves[b]

[a] Center for Lightweight Production Technology, German Aerospace Center, Ottenbecker Damm 12, Stade, Germany
[b] Aerospace Non-Destructive Testing Laboratory, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands

ABSTRACT

Automated fibre layup techniques are commonly used composite manufacturing processes in the aviation sector and require a manual visual inspection. Neural Network classification of defects has the potential to automate this visual inspection, however, the machine decision-making processes are hard to verify. Thus, we present an approach for visualising *Convolutional Neural Network* (CNN) based classifications of manufacturing defects and quantifying its robustness. Our investigations have shown that especially *Smoothed Integrated Gradients* and *DeepSHAP* are particularly well suited for the visualisation of CNN classifications. The *Smoothed Integrated Gradients* technique also reveals advantages in robustness when evaluating degraded input images.

Keywords: Defect classifications, CNN, Inline Inspection, xAI, Computer Vision, Composite Manufacturing, Laser Line Scan Sensor

1. INTRODUCTION

With the production of the Airbus A350 XWB and the Boeing 787, lightweight components are widespread in the aviation sector.[1,2] Such structures are usually made of *Carbon Fiber Reinforced Plastic* (CFRP). The production of these often sophisticated lightweight components is normally rather expensive. Hence, efficient production processes are necessary for economical manufacturing. To satisfy the high safety standards in the aviation industry, the fibre placement process involves a visual inspection. This visual check takes usually up to 50 %[3] of the manufacturing duration. This provides huge opportunities for significant improvements in terms of speed and quality through the automation of this step.

For automated inspection, a reliable machine-based classification of manufacturing defects in an image is necessary.[4,5] Methods from the field of machine learning are highly suitable for such a classification of fibre layup defects.[6,7] However, these techniques often face the disadvantage that the models' decisions are hard to comprehend. This applies especially to *Artificial Neural Network* (ANN) or deep learning methods in general.

In order to be able to carry out a comprehensive analysis for this, the importance of individual pixels or small image areas for the classification decision must be examined initially. For this purpose, we choose three suitable techniques for the investigations in this paper. The analysis in this paper provide the basis for subsequent studies on correlating the image areas that are important for a classification decision with the actual characteristics of a given fibre layup defect.

For the investigations in this paper we consider height profile scans of layup defects from the *Automated Fiber Placement* (AFP) manufacturing process.[8–10] Therefore, we applied a *Laser Line Scan Sensor* (LLSS) for data recording as this device is often used in industry and in research, for this particular application case.[6,8,9,11,12]

Further author information: (Send correspondence to Sebastian Meister)
Sebastian Meister: E-mail: sebastian.meister@dlr.de, Telephone: +49 531 295 3710
Mahdieu A. M. Wermes: E-mail: mahdieu.wermes@dlr.de, Telephone: +49 531 295 3784
Jan Stüve: E-mail: jan.stueve@dlr.de, Telephone: +49 531 295 3700
Roger M. Groves: E-mail: R.M.Groves@tudelft.nl, Telephone: +31 15 27 88230

Automated Visual Inspection and Machine Vision IV, edited by Jürgen Beyerer, Michael Heizmann, Proc. of SPIE Vol. 11787, 1178705 · © 2021 SPIE
CCC code: 0277-786X/21/$21 · doi: 10.1117/12.2592584

The methodology of this paper involves an analysis and visualisation of the importance of individual image areas for the *Convolutional Neural Network* classification of typical fibre layup defects. For this evaluation, the *Explainable Artificial Intelligence* techniques *Smoothed Integrated Gradients*, *Guided Gradient Class Activation Mapping* as well as *Deep Learning Important Features with Shapley Additive Explanations* are applied and the results are examined.

2. STATE OF THE ART

2.1 Fibre placement inspection

The AFP process is often used for the production of complex composite structures.[13,14] Various defects can arise in the AFP process.[15] Common fibre placement defects are *wrinkles*, *twists*, *foreign bodies*, *overlaps* and *gaps*.[6,11,15–17] For the aim of automated inspection, LLSS based technique are suitable for capturing defect topology data during AFP manufacturing. This field is currently receiving much attention from industry and research.[8–10,18,19] In our recently published studies from Meister et al.[6,11] on defect detection, classification and data synthesis we already used a LLSS to acquire the investigated test data.

2.2 Explainable Artificial Intelligence techniques

This section briefly introduces different *Explainable Artificial Intelligence* (xAI) methods for analysing the importance of individual image areas with respect the decision of an ANN. Comparatively much research has been done in the field of Gradient and Decomposition based approaches. Due to the amount of available knowledge, these methods are probably quite promising. Because of their diverse functional concepts and their performance characteristics as stated in the study of Yeh et al.,[20] the *Smooth Integrated Gradients* (*Smooth IG*),[21,22] *Deep Learning Important Features (LIFT) with Shapley Additive Explanations (SHAP)* (*DeepSHAP*)[23] and *Guided Gradient Class Activation Mapping* (*Guided Grad-CAM*)[24] approaches are highly appropriate for examining an unfamiliar scenario. These techniques evaluate several types of information obtained from the neural activation of an ANN in order to reveal the significance of certain parts of the image for a machine decision.

3. METHODOLOGY

3.1 Experimental setup

As suitable defect types, *none*, *wrinkles*, *twists*, *foreign bodies*, *gaps* and *overlaps* were selected for the experiments. The corresponding input defect images were captured with the following test setup. Each defect image was cropped manually from an entire LLSS image. The individual defect images were downsized to a sensible dimension of 128×128 px.

For recording of representative original images the test setup from Figure 1 was used. This setup was invariant to disruptive effects from the production procedure like pollution, radiation of the heater or effector rotation. The setup involved an articulated-arm robot, the *Automation Technology GmbH* (AuTech) C5-4090 LLSS[25] as well as a CFRP specimen. An AuTech C5-4090 LLSS recorded 16-bit single channel depth images with the size of 4096 (w) $\times$ 500 (h) px. This measurement image contained a CFRP specimen with the dimension 250×150 mm. Moreover, the material specimen was scanned at a speed of 200 mm/s.

In addition, the well suited *Convolutional Neural Network* (CNN) classifier with 16 hidden layers outlined in Meister et al.[6] was used as the basis for the investigations in this study.

3.2 Selection of xAI techniques

In this scenario a rather unknown classification case was considered. Thus, for the basic analysis in this paper, relatively novel xAI methods with different operating principles were selected.

Guided Grad-CAM was chosen because it does not only consider the importance of individual pixels, but describes the importance of small image areas.

The *Smooth IG* method has the advantage to use the same reference for all defect classes. Accordingly, the calculation results are comparable across the defect samples of the same class and thus the results from individual examples can be interpreted in a more general way. Additionally, the inherent integration and smoothing

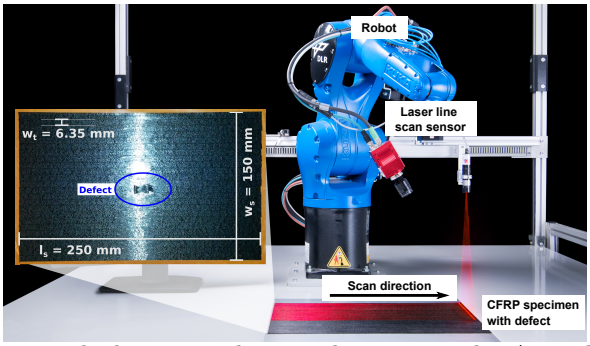


Figure 1: Data recording setup in which a jointed-arm robot moves the AuTech C5-4090 sensor parallel to a CFRP specimen and captures height profile data of this specimen.

approach reduces noise in the outcomes.
According to the research of Yeh et al.[20] as well as Lundberg and Lee,[23] in their study the *DeepSHAP* method yields the most faithful and robust results. Hence we also selected this approach for our investigations.

3.3 Investigation of the xAI findings

Initially, for the previously introduced CNN[6] and the considered defect categories, a visual xAI explanation was calculated for an example layup defect images. These visual explanations were presented as greyscale images, where the pixel intensity represents their overall influence on the CNN decision.
Secondly the correlation between the xAI explanation result and the classification decision of the CNN were examined. For each considered input defect image, the degree of activation at the output neurons of the CNN for the associated defect class were observed. These served as the reference activations. Then, based on the calculated xAI importance for each pixel, the n most important pixels in the original input image were set to zero. For this manipulated input image, the previously described analysis of neuron activations was then performed again. The influence of the neural activations was presented here as the degree of variation in neural activation in relation to the previously defined reference activation for the unmodified, original input image. This experimental design was inspired by the research of Srinivas and Fleuret.[26]

4. RESULTS

4.1 Visual xAI evaluation

Figure 2 presents the visualised xAI results for the six classes considered in this paper. In each case, the original input image which serves as input for the CNN is shown on the far left. To the right, for each of the three examined xAI methods, the absolute importance values are given as a greyscale image. The *Smooth IG* method yields the most homogeneous result for the example images. But this method primarily assigns greater importance to the brighter pixels in the defect image. This results from the fact that through the multiplication with the corresponding gradients, the intensities of the pixels in the input image have a strong influence on the result of this xAI method. Thus, the contours of the defects are rather similarly visible in the input image and in the xAI calculation.
The *DeepSHAP* results show a consistent visual similarity of the input image to the xAI explanation. Hence, *gaps* and *overlaps* are difficult to recognise and the results appear quite similar to *none*. The importance of a defect region also seems to depend on the difference in brightness in the input image. However, these result images are clearly subject to a kind of statistical noise. In the xAI output images a kind of skewed pattern of the important pixels is partly visible.
The greyscale result images from the *Guided Grad-CAM* calculations reveal important defect regions, which clearly distinguish themselves from the background. For *gaps*, *overlaps* and *foreign bodies* the respective xAI outcomes match quite well with the defect region. For all the other classes, they only partially correspond to the actual defect region in the original input image. Especially for *none* defect images this behaviour is quite conspicuous and might be attributed to pre-processing artefacts within the image.

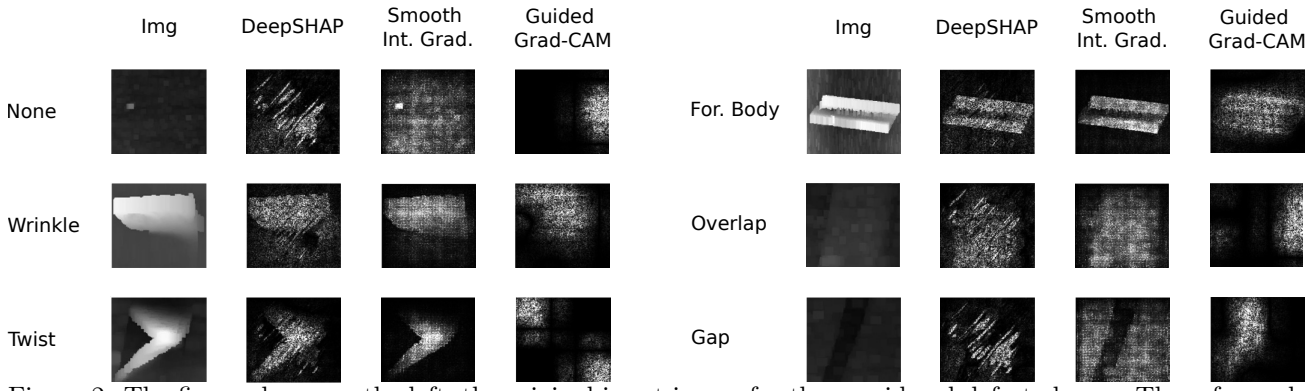
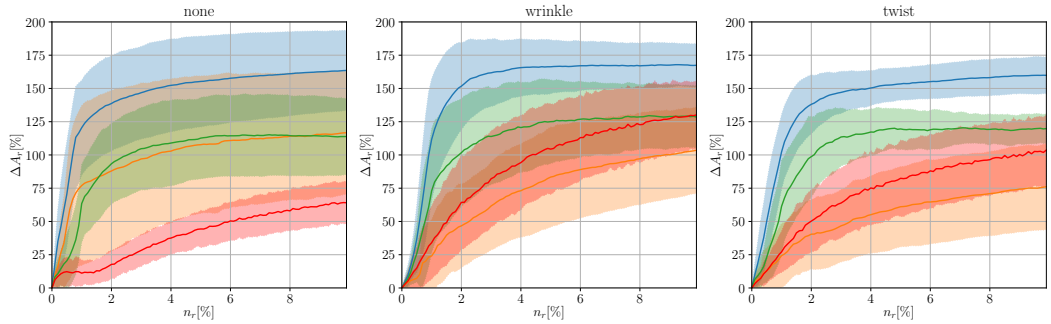


Figure 2: The figure shows on the left, the original input image for the considered defect classes. Then, for each of the three xAI methods *DeepSHAP*, *Smooth IG*, *Guided Grad-CAM*, the greyscale image of the magnitude of the importance values are displayed.



(a) Results for class *none*. (b) Results for class *wrinkle*. (c) Results for class *twist*.

(d) Results for class *foreign body*. (e) Results for class *overlap*. (f) Results for class *gap*.

Figure 3: Class dependant change in neural activations ΔA_r over the percentage of removed pixels n_r. Methods: *DeepSHAP* (blue), *Smooth IG* (green), *Guided Grad-CAM* (orange), reference (red). The associated standard deviations are presented as a coloured tubes.

4.2 Variance in neural activation for modified data

In Figure 3 the change in relative neural activation ΔA_r is plotted over the relative percentage of removed pixels n_r for the respective examined classes. The graphs represent the mean values over all input data sets of a certain defect class. The coloured tubes indicate the associated standard deviations. The red curve serves as a reference, as introduced above. For all curves, an initially very steep rise is evident up to 1-2 % removed pixels. Then the curves flatten out considerably. In most cases, an almost constant value of the change in neural activation ΔA_r is reached for n_r between 2 % and 10 %. According to the *DeepSHAP* calculation, the modification of the input image leads to the smallest change in neural activation of $\Delta A_r = 40$ % for *foreign bodies*. For all the

other classes $\Delta A_r > 150$ % for $n_r \geq 8$ % is reached. The *Smooth IG* results yield a similar curve with a much lower end value of $\Delta A_r = 22$ % for *foreign bodies* and about $\Delta A_r = 100$ % for the other classes at $n_r \geq 8$ %. The *Guided Grad-CAM* method achieves lower ΔA_r values for $n_r \geq 8$ % compared to the reference curve which considers randomly removed pixels in the initial image. For the *wrinkle*, *twist* and *overlap* classes, the ΔA_r values for $n_r \geq 8$ % are very similar for the *Guided Grad-CAM* and *Smooth IG* outcomes. The strong increase of ΔA_r for the *Guided Grad-CAM* calculation up to about $n_r = 1$ %, nevertheless, indicates a strong identification of the most important image pixels for almost all classes except for *foreign bodies*. Across all xAI techniques, the *foreign body* class is conspicuous. For this, the ΔA_r values for $n_r \geq 8$ % are significantly lower than for the other classes. Also, the standard deviations of the ΔA_r values are much smaller. The comparatively small amount of input images can of course be an issue for the significance of these findings. Based on the often almost constant ΔA_r values for $n_r \geq 8$ % and the gradient of the initial steep curve rise we can estimate the performance of a xAI method. Further research is needed for more detailed conclusions on this.

5. DISCUSSION

The previously selected approaches *Smooth IG*,[21,22] *DeepSHAP*[23,27] and *Guided Grad-CAM*[24] for the visualisation of image areas that are particularly important for the classification decision provide a sound possibility to represent the behaviour of a CNN. However, the different operating principles of the individual methods must be taken into account. In some cases, these can have a great influence on the results. The analysis of the neural activations of the CNN for a defined input image provide extensive insights about the actual behaviour of the CNN as well as potential uncertainties in the classification process. Further research will be conducted to analyse the individual xAI procedures in more detail.
Concluding, we can state that for the considered use case especially the xAI methods *Smooth IG* and *DeepSHAP* are particularly well suited for the representation of important image regions for the decision-making process of the CNN. In the following section, the major findings of this paper are summarised and the added value for the community is highlighted.

6. CONCLUSION

The findings from this paper have revealed that the importance of individual image pixels for the decision-making process of a *Convolutional Neural Network* classifier can be visualised and assessed. According to the research in this study, the *Explainable Artificial Intelligence* methods *Smooth Integrated Gradients* and *Deep Learning Important Features with Shapley Additive Explanations* are particularly well suited for this purpose.
The results from this paper support the developers of camera-based inspection systems for the lightweight composite industry in the design and integration of reliable complex machine learning applications. Furthermore, the presented findings provide guidance for respective certification procedures for such machine learning algorithms.

ACKNOWLEDGMENTS

This research is part of the project DHiiP-AIR and was financially supported by the Federal Ministry for Economic Affairs and Energy. This project has received funding from the Federal Ministry for Economic Affairs and Energy under the funding code No. 20W1911F.

REFERENCES

1. G. Marsh, "Airbus A350 XWB update," *Reinforced Plastics* **54**, pp. 20–24, Nov. 2010.
2. A. McIlhagger, E. Archer, and R. McIlhagger, "Manufacturing processes for composite materials and components for aerospace applications," in *Polymer Composites in the Aerospace Industry*, P. Irving and C. Soutis, eds., pp. 59–81, Elsevier, 2020.
3. C. Eitzinger, "Inline inspection helps accelerate production by up to 50 %," *Lightweight Design worldwide* , Mar. 2019.
4. European Union Aviation Safety Agency, "Human intelligence roadmap - a human-centric approachto ai in aviation," techreport Vers. 1.0, European Union Aviation Safety Agency, Feb. 2020. Vers. 1.0.

5. EASA AI Task Force and Daedalean AG, "Concepts of design assurancefor neural networks (codann)," techreport Vers. 1.0, European Union Aviation Safety Agency and Daedalean AG, Mar. 2020. Vers. 1.0.
6. S. Meister, N. Möller, J. Stüve, and R. M. Groves, "Synthetic image data augmentation for fibre layup inspection processes: Techniques to enhance the data set," *Journal of Intelligent Manufacturing* , Feb. 2021.
7. C. Schmidt, T. Hocke, and B. Denkena, "Artificial intelligence for non-destructive testing of CFRP prepreg materials," *Production Engineering* , July 2019.
8. J. Cemenska, T. Rudberg, and M. Henscheid, "Automated in-process inspection system for AFP machines," *SAE International Journal of Aerospace* **8**, pp. 303–309, Sept. 2015.
9. C. Weimer, A. Friedberger, A. Helwig, S. Heckner, C. Buchmann, and F. Engel, "Increasing the productivity of CFRP production processes by robustness and reliability enhancement," in *CAMX 2016 - The Composites and Advanced Materials Expo and Conference*, (Airbus Group Innovations, 81663 Munich, Germany; AirbusInfactory Solutions GmbH, 81663 Munich, Germany), Sept. 2016.
10. S. Black, "Improving composites processing with automated inspection." compositesworld, Jan. 2018. https://www.compositesworld.com/articles/improving-composites-processing-with-automated-inspection.
11. S. Meister, M. A. M. Wermes, J. Stueve, and R. 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*, D. Zonta and H. Huang, eds., SPIE, Apr. 2020.
12. S. Meister, M. A. M. Wermes, J. Stüve, and R. M. Groves, "Review of image segmentation techniques for layup defect detection in the automated fiber placement process," *Journal of Intelligent Manufacturing* , May 2021.
13. T. Rudberg, J. Nielson, M. Henscheid, and J. Cemenska, "Improving AFP cell performance," *SAE International Journal of Aerospace* **7**, pp. 317–321, Sept. 2014.
14. F. Campbell, *Manufacturing Processes for Advanced Composites*, Elsevier Science & Technology, 2004.
15. E. Oromiehie, B. G. Prusty, P. Compston, and G. Rajan, "Automated fibre placement based composite structures: Review on the defects, impacts and inspections techniques," *Composite Structures* **224**, p. 110987, sep 2019.
16. R. Harik, C. Saidy, S. J. Williams, Z. Gurdal, and B. Grimsley, "Automated fiber placement defect identity cards: cause, anticipation, existence, significance, and progression," in *SAMPE 18*, 01 2018.
17. F. Heinecke and C. Willberg, "Manufacturing-induced imperfections in composite parts manufactured via automated fiber placement," *Journal of Composites Science* **3**, p. 56, June 2019.
18. G. Gardiner, "Zero-defect manufacturing of composite parts." CompositesWorld, Nov. 2018. accessed: 2019-06-18.
19. S. Black, "Improving composites processing with automated inspection, part II." compositesworld, June 2018. accessed: 2019-06-19.
20. C.-K. Yeh, C.-Y. Hsieh, and A. S. Suggala, "On the (in)fidelity and sensitivity of explanations," *NeurIPS* , 2019.
21. M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proceedings of the 34 th International Conference on Machine Learning*, 2017.
22. D. Smilkov, N. Thorat, B. Kim, and F. V. ´and Martin Wattenberg, "Smoothgrad: removing noise by adding noise." arXiv:1706.03825, 2017.
23. S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 2017.
24. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localizatio," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017.
25. Automation Technology GmbH, "C5 series - user manual for high speed 3d sensors," techreport 1.2, Automation Technology GmbH, Hermann-Bössow-Straße 6-8, 23843 Bad Oldesloe, Germany, Mar. 2019. Rev 1.2.
26. S. Srinivas and F. Fleuret, "Full-gradient representation for neural network visualization," in *2019 Conference on Neural Information Processing Systems*, 2019.

27. A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," in *Proceedings of the 34th International Conference on Machine Learning*, D. Precup and Y. W. Teh, eds., *Proceedings of Machine Learning Research* **70**, pp. 3145–3153, PMLR, (International Convention Centre, Sydney, Australia), Aug. 2017.