Synthetic image data augmentation for fibre layup inspection processes: Techniques to enhance the data set

Sebastian Meister[1] · Nantwin Möller[1] · Jan Stüve[1] · Roger M. Groves[2]

Abstract

In the aerospace industry, the *Automated Fiber Placement* process is an established method for producing composite parts. Nowadays the required visual inspection, subsequent to this process, typically takes up to 50% of the total manufacturing time and the inspection quality strongly depends on the inspector. A Deep Learning based classification of manufacturing defects is a possibility to improve the process efficiency and accuracy. However, these techniques require several hundreds or thousands of training data samples. Acquiring this huge amount of data is difficult and time consuming in a real world manufacturing process. Thus, an approach for augmenting a smaller number of defect images for the training of a neural network classifier is presented. Five traditional methods and eight deep learning approaches are theoretically assessed according to the literature. The selected *conditional Deep Convolutional Generative Adversarial Network* and *Geometrical Transformation* techniques are investigated in detail, with regard to the diversity and realism of the synthetic images. Between 22 and 166 laser line scan sensor images per defect class from six common fiber placement inspection cases are utilised for tests. The GAN-Train GAN-Test method was applied for the validation. The studies demonstrated that a conditional *Deep Convolutional Generative Adversarial Network* combined with a previous Geometrical Transformation is well suited to generate a large realistic data set from less than 50 actual input images. The presented network architecture and the associated training weights can serve as a basis for applying the demonstrated approach to other fibre layup inspection images.

Keywords Image data augmentation · Automated fiber placement · Inline inspection · Generative adversarial networks · Laser line scan sensor

Introduction

Lightweight structures are now commonly used in aerospace manufacturing. The Airbus A350 XWB and the wing and fuselage production of the Boeing 787 are examples of an increasing demand for these lightweight components (Marsh 2010; McIlhagger et al. 2020). Compared to metallic materials, *Carbon Fiber Reinforced Plastic* (CFRP) offers superior stiffness and strength properties. Thus, lightweight structures are often made from CFRP. The manufacturing of these mostly complex lightweight structures is typically quite expensive. In order to make production economical, fast and efficient production techniques are essential. To meet the high safety requirements of the aerospace industry a visual inspection follows the fibre layup process.

Typically today, this manual inspection takes between 32% (Rudberg et al. 2014) and 50% (Eitzinger 2019) of the total production time. Moreover, due to the manual inspection process, it is sometimes impossible to fulfil the required inspection accuracy. This aspect offers great potential for improvements in quality and speed.

A crucial stage of the automated inline inspection is the reliable classification of manufacturing defects within a sensor image. Machine learning methods are very well suited for this purpose (Schmidt et al. 2019). Unfortunately, especially the very common approaches based on *Artificial Neural Network* (ANN), often require very large training data sets, which are quite difficult to produce in a reliable production process or during the development of a classification system (Huang et al. 2019; Tan and Le 2019; Zambal et al.

✉ Sebastian Meister
  sebastian.meister@dlr.de

1 Center for Lightweight Production Technology (ZLP), German Aerospace Center (DLR), Ottenbecker Damm 12, Stade, Germany

2 Aerospace Non-destructive Testing Laboratory, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, The Netherlands
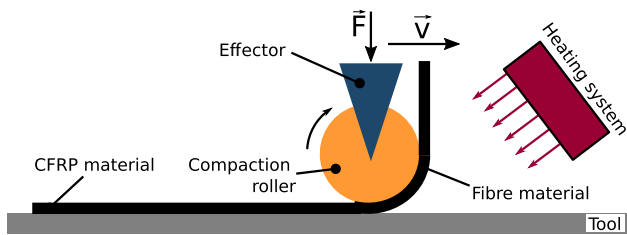
Fig. 1 AFP manufacturing process using a heating system to apply temperature and a compaction roller to apply pressure to the laid up fibre material. **F** is the compaction force and **v** the effector velocity

Fig. 2 Schema of five common AFP process defects as well as a proper material lay up.

2019b) For this reason we investigate methods for generating a large training data set based on a few previously acquired real defect images for the *Automated Fiber Placement* (AFP) inspection process, in this paper.

The AFP technology is relatively novel, but is increasingly applied in industry. Thus, we have chosen this technique for further investigations in this paper, aiming a proper transferability of our research results (Cemenska et al. 2015; Weimer et al. 2016; Black 2018) Since a *Laser Line Scan Sensor* (LLSS) is frequently used in research and development for the inline inspection of AFP processes, in this paper we will focus on greyscale depth images from such a sensor (Cemenska et al. 2015; Weimer et al. 2016; Ucan et al. 2019; Meister et al. 2020) A LLSS is based on the principle of triangulation to obtain topology data from a laser beam that is projected onto a surface and reflected to a camera sensor at an angle to the laser illumination. The research question of this publication is:

Which methods can be used to generate synthetic image data of fiber placement defects from the AFP process, using a small data set with mostly less than 100 images?

The methodology of this paper is to design and assess a *Deep Convolutional Generative Adversarial Network* (DCGAN) to generate a large dataset with several thousand images from less than 50 input images per class. A visual image evaluation combined with the GAN-Train GAN-Test method will then be applied to assess the synthetic data generated this way.

Related research

Manufacturing process

Several fiber placement technologies are available on the market today. Popular methods are the *Automated Fiber Placement* (AFP) (Lengsfeld et al. 2014; Maass 2012), *Dry Fiber Placement* (DFP) (Lengsfeld et al. 2014; Maass 2012), *Automated Tape Laying* (ATL) (Lengsfeld et al. 2014) and *Direct Roving Placement* (DRP) (Grohmann et al. 2016). These methods apply CFRP material in layers onto a mould.
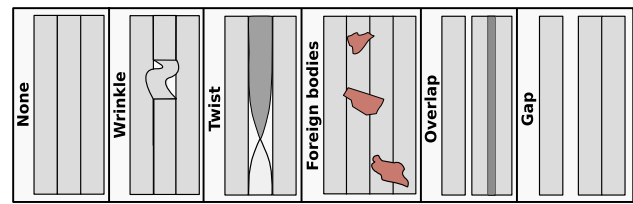
This process has been described by Campbell (Campbell 2004) and is illustrated in Fig. 1. The AFP technology is preferably utilised to manufacture complex composite structures (Rudberg et al. 2014; Campbell 2004). In the AFP process several narrow pre impregnated material strips, so-called tows, are deposited along a previously programmed path (Oromiehie et al. 2019). Therefore, composite material e.g. carbon prepreg material is transferred to an effector. This effector carries the material to the mould's surface. Afterward, the material is heated to increase its tag properties and pressed onto the mould (Lengsfeld et al. 2014). Each structural component consists of many CFRP prepreg layers (Campbell 2004) Different part geometries can be manufactured using the AFP process. Moreover, Rudberg (2019) expects an increasing use of the AFP technology in future applications.

Several different defects can result from the fiber placement process. These defects are often directly related to the fibre layup (Oromiehie et al. 2019). Harik et al. (2018) have investigated the relationship between AFP defects and process planning, layup strategies and processing. Potter (2009) studied the factors that causes deviations in the AFP production. According to Harik et al. (2018) and Potter (2009), all defects that can occur during the fiber layup result in geometric changes and deviations from an exact layup surface. Thus, common AFP defect types from the literature are wrinkles, twists, foreign bodies, overlaps and gaps. These defects, together with a reference sample with no defect, are illustrated in the Fig. 2. The associated geometric defect measures and their characteristics are summarised in the Table 1. Wrinkle and twist have different but distinct shapes. These defect types protrude from the materials surface. This leads to greater changes in height and result in clear edges of the defect. In the longitudinal direction wrinkles causes a single edge. In contrast, twists have a very small growth in altitude over their distance. Gap and overlap defects have very similar geometrical properties. Both are quite flat and show only minor changes in topology. Gaps have two small edges at their beginning and their end, perpendicular to the fibre orientation. Overlaps on the other hand show three small edges transverse to the fibre direction. This is due to the fact that these defects are a combination of a gap and an overlapping tow, in most cases. Also gaps and overlaps have nearly

Table 1 The table summarises the geometrical dimensions of the fiber placement defects from Fig. 2

	Wrinke	Twist	Gap	Overlap	For. Mat.
Typical ratio: (l/w)	0.5 to 2	5 to 10	≤ course length	≤ course length	unk.
Thickness deviation (+/−)	≥ 3x CPT (+)	≥ 2x CPT (+)	≤ 1x CPT (−)	≤ 1x CPT (+)	unk.
Source	Oromiehie et al. (2019) and Harik et al. (2018)	Oromiehie et al. (2019), Harik et al. (2018) and Heinecke and Willberg (2019)	Oromiehie et al. (2019), Harik et al. (2018) and Heinecke and Willberg (2019)	Oromiehie et al. (2019), Harik et al. (2018) and Heinecke and Willberg (2019)	Harik et al. (2018) and Heinecke and Willberg (2019)

The value range of the length-to-width (l/w) ratio is presented. Due to the large variance in defects geometry no absolute values are given. CPT for the investigations is about 0.125 mm. For the thickness measure + means an increase in thickness and − indicates a thickness decrease

no edges apparent along the tows. Their similarities makes the distinction between these two classes mostly very difficult. The inconspicuous form of these defects enables the possibility to analyse algorithms for this use case. Furthermore, these previously mentioned defect types are commonly applied as example defects in the related research (Oromiehie et al. 2019; Harik et al. 2018; Heinecke and Willberg 2019) Additionally, foils as typical foreign bodies in manufacturing processes are considered. They show quite different reflection properties in comparison with layed up fibre material (Potter 2009; Miesen et al. 2015).

A manual, visual inspection of each ply is very time consuming and mostly does not fulfil the actual quality requirements of this inspection process. Therefore, the common LLSS technology for the recording of the corresponding defect image data in the production process is described below.

Sensor based inspection and data processing

Inline inspection for AFP processes is of great interest in research and industry today. Electroimpact (Cemenska et al. 2015; Black 2018), InFactory Solutions (Weimer et al. 2016), Danobat Composites (Black 2018) and Profactor (Gardiner 2018) used LLSS systems for the inline *Quality Assurance* (QA) of AFP processes. This technology allows the acquisition of 3D topology information of the materials surface, which may have contributed to its success (Weimer et al. 2016) Schmitt et al. (2008) and Schmitt et al. (2007) started investigating LLSS based methods for contour scanning of fabrics and preforms. They demonstrated that a LLSS is a suitable system for fabric and preform inspection. Miesen et al. (2015) proposed a method for detecting defects with a point laser displacement system. They discussed factors influencing deviations in their research and analysed the accuracy of such systems. They also presented different types of defects and their corresponding geometric characteristics.

Sacco et al. (2018) investigated the defect segmentation for LLSS depth images of AFP fiber placement defects using a *Convolutional Neural Network* (CNN). They considered 15 different types of fiber placement defects and attempt to segment and classify them correctly. For training their fully connected CNN they used 800 x 800 pixel LLSS depth images. They also suggested the use of a *Generative Adversarial Network* (GAN) for a more stable segmentation of their defect types and mentioned the need of a large database for the training of an ANN. Furthermore they add the fact that a GAN generates artificial data sets as part of its operating principle. Zambal et al. (2019a; 2019b) introduced an end-to-end deep learning defect detection and segmentation approach for the AFP process inspection considering synthetically generated training data. Therefore, they applied a U-Net CNN structure, which Ronneberger et al. (2015) have introduced in 2015. Additionally, they used realistic depth maps of a LLSS for validation. Their results also indicate difficulties in differentiating between gaps, missing tows and overlaps. Beyond that, they mentioned difficulties in recording a large amount of real training data in a real world scenario. Therefore a data synthesis is indispensable. Furthermore Tabernik et al. (2019) explained deep-learning methods which are suitable for analysing surface anomalies. They demonstrated their application for the detection and classification of cracks in surfaces within one shot. Therefore they connected a segmenting CNN with five convolutional layers and a classifying CNN with six convolutional layers. The designed ANN architecture allows the model to be trained with only 25-30 data sets. The sufficiency of such a small data set is a key requirement for practical applications, in their opinion. In contrast to Zambal et al. (2019b; 2019a) they stated that the U-Net architecture from Ronneberger et al. (2015) performs much worse for defect segmentation. Luo et al. (2020) investigated various GAN based methods to generate synthetic training data especially for unbalanced or very small training data sets for deep learning fault diagnosis systems for produc-

tion machines. In particular, they evaluated the performance and trainability of GAN, *Conditional Generative Adversarial Network* (CGAN) and *Conditional Deep Convolutional Generative Adver- sarial Network* (CDCGAN) architectures. They demonstrated their approach using two diagnostic data sets for a bearing and a gearbox.

Meister et al. (2020) described in their paper a technique for the smoothing of LLSS scan images of AFP laying defects. Their approach is based on the *Contrast Limited Adaptive Histogram Equalization* (CLAHE) algorithm and is very well suited for LLSS scan images with low information density. This technique is also used within this paper for image pre-processing. Furthermore, the defect segmentation methods examined in the study of Meister et al. provide a promising way of extracting appropriate individual defect images from the overall scan images in a real fibre placement inspection process. These individual defect images provide a sensible input for a classifier.

Subsequently, the principles of ANN training and approaches for artificially augmenting a database are introduced.

Image data augmentation techniques

Within this section we present various methods for the synthesis of depth image data from fiber placement inspection, in order to use them for the training of neural networks. For this purpose, promising training data characteristics are introduced at first. Subsequently, suitable techniques for image data augmentation are discussed.

Review on training data sets from related research

Deep learning techniques require very large data sets to train these ANN, compared to e.g. a *Support Vector Machine* (SVM). However, the minimum amount of training data needed depends strongly on the architecture and trainable parameters of the ANN. This in turn is influenced from the application case and the characteristics of the data to be used. In order to determine a reasonable amount of data to be synthesised and applied for the subsequent training of an ANN, similar use cases from the literature are considered.

Wan et al. (2013) examined the classification of handwritten numbers from zero to nine from the *Modified National Institute of Standards and Technology* (MNIST) data set. They concluded that a data set of 7000 grayscale images of size 28x28 is well suited for the classification of the 10 classes. Huang et al. (2019) compared the classification accuracy of different classifying ANN on the three public data sets *Canadian Institute For Advanced Research* (CIFAR)-10, Stanford Cars and Oxford Pets. For these three classification tasks with the best accuracies from 94.8 to 99.0% they used between 3680 and 50,000 training images. Tan and Le (2019) compared different training data which consisted of

2040 to 75750 data samples of various types for training their transfer learning approach. Wu et al. (2019) used a GAN based approach for contrast adjustment of *Magnetic Resonance Imaging* (MRI) data. Therefore they trained their GAN with 2000 original images. Jain et al. (2020) evaluated different GAN based techniques for augmenting an image dataset for training a CNN classifier for the detection of defects on metallic surfaces. They first applied a Geometrical Transformation to generate a set of 9000 images. Subsequently, 5400 of these images were randomly selected and used for training the GAN. Finally, the GAN processes generated 3600 images for training the CNN classifier in order to examine the performance improvement in the detection of surface defects. Schmidt et al. (2019) used image based inspection data from a thermographic camera for the inspection of an AFP process. In their work they compared the application of a pre-trained ResNet-101 ANN with a custom developed ANN structure for the classification of different fiber placement defect types. For this purpose, they performed various experiments with differently sized training data sets with between 1000 and 3000 training images. In their investigations the classification results from their self developed ANN are more accurate than those of the pre-trained ANN. Within the previously mentioned work from Zambal et al. (2019b) and Zambal et al. (2019a), they trained their CNN with 5000 synthetic defect samples. Joshi et al. (2018) pointed out the disadvantages of individual classifiers such as SVM or ANN for the part inspection. As a solution in their paper they proposed a hybrid approach of using different individual classifiers. In order to demonstrate the performance of their approach they carried out three different classification tasks which could be applied similarly for the inspection of components. For training of their algorithms they captured 2000 real part images but from only 25 different components. In order to get feasibliy large data sets for this research, subsequently techniques for the augmentation of small data sets are presented.

Image augmentation techniques from literature

In this section various techniques for image data augmentation from related research are presented and subsequently compared in "Methodology" section. On this basis feasible methods for data synthesis in this paper are selected.

Shorten and Khoshgoftaar (2019) summarised various deep learning and basic image manipulation techniques for data augmentation with the aim of avoid an overfitting in training processes. Their focus was especially on GAN based methods. Furthermore they discussed different types of image data biases such as lighting, occlusion or image scale and their influences on a machine learning algorithm. Cubuk et al. (2019) explained the properties of basic image manipulation approaches such as kernel filtering, Geometrical Transformation, random erasing, color space transformation

and mixing images. With regard to these techniques they investigated rules for the efficient automated composition of these different methods. Their aim was to automatically find the best augmentation policy and improve the performance of a classifying ANN. To evaluate their approach they additionally applied a GAN based augmentation method and carried out validation experiments on common image data sets. They stated that their traditional approach leads to slightly better classification results than the GAN method. Perez and Wang (2017) shared their perspective. They additionally summarised various deep learning methods for the artificial augmentation of a data set. They proposed a combined usage of GAN methods and traditional procedures for efficient augmentation of a data set. Moreover, they point out the major issue of overfitting when the applied training data sets are not sufficiently representative and diverse. In order to obtain a reasonable trade-off between computational effort and synthesis result they proposed a combination of traditional techniques with ANN based augmentation methods. Mikolajczyk and Grochowski (2018) took a closer look at the ANN based generation of artificial image data. As a further supplement they proposed neural style transfer methods.

From the references given above we can also conclude that GAN and *Autoencoder* (AE) techniques are often stated to be very suitable for this application. A GAN should produce qualitatively better image augmentation results than the AE, with the drawback that the GAN behave sometimes unstably for particular use cases. According to the literature, a GAN consists of two forward connected ANN, a so-called generator and a discriminator. They face each other as competitors. In case the balance between these two components is not preserved and thus the Nash equilibrium is fulfilled the GAN becomes unstable.

With the aim of reducing this issue, various enhancements of the basic GAN were developed. Radford et al. (2016) introduced the DCGAN and Goodfellow et al. (2014) and Goodfellow (2017) explained some details on the working principle of this technique. Furthermore, they confirmed the novelty and the promising usage of GAN based methods in future applications. However, they also indicate that some research is still needed especially with regard to the better understanding of network stability. Arjovsky et al. (2017) mentioned the probably more stable *Wasserstein Generative Adversarial Network* (WGAN). This WGAN uses a Wasserstein loss function which performs similar to the DCGAN but is less likely to become unstable at its limits. Karras et al. (2018) gives a detailed description of the *Progressive Growing Generative Ad- versarial Network* (PGGAN). Referring to the progressive growing training principle different resolutions of a training image are considered. The level of image detail increases with the training of deeper layers in the GAN. This procedure is designed to minimise the computational effort and improve the stability of the training process.

On this basis, the Table 2 presents a detailed comparison of different established GAN and AE methods of Goodfellow (2017), Shorten and Khoshgoftaar (2019) and Creswell et al. (2018). These algorithms are assessed on the basis of criteria from the literature. The impact of individual criteria is considered in a weighted manner. Therefore, an expected value w_e is specified on the basis of the use case and the presented literature. In order to handle the subjective specification of w_e and to ensure the robustness of the performed evaluation, weighting intervals $[w_e - 0.5, w_e + 0.5]$ are specified for each criterion. The presented range of results is determined by 25 runs with randomly selected weights according to the Monte Carlo method.

The assessment Table 2 shows that the DCGAN provides the best rating followed closely by the WGAN. The AE techniques tend to yield worse evaluation results than the GAN approaches.

On the basis of these results, the DCGAN will be examined in more detail in this paper. We considered only the DCGAN for further investigations in this paper despite the assessment result close to the WGAN. This DCGAN algorithm is more commonly used than the WGAN. Thus there is more information available in the literature which can be used to improve the synthesis results. Furthermore, the WGAN is basically a modified DCGAN which applies the Wasserstein loss function to avoid instabilities during the training. Arjovsky et al. (2017) However, if the algorithms stability does not cause any issues the WGAN should generate very similar results as the DCGAN.

In order to find a suitable configuration for the DCGAN applied here, Table 3 compares different, reasonable DCGAN settings from the literature. Therefore, the parameters from Radford et al. (2016) are the basis for the subsequent improvements of Perarnau et al. (2016), Neff (2018), Salimans et al. (2016) and Brownlee (2019) for a DCGAN. Additionally, the table shows the mutual intersections of the parameters. Furthermore, Odena et al. (2017) presented an auxiliary classifier GAN configuration which potentially provides useful guidance for the GAN parametrisation and selection of test parameters.

In order to clarify this, for the following investigations in this paper it is necessary to implement and configure a technique which is able to generate synthetic depth image data of fibre layup defects. This needs to be done in such a way that the algorithm runs in a stable way and the generated image data looks as realistic as possible, although it is different from the real input data. The DCGAN data augmentation method seemed to be very promising from the assessments in the Table 2 and was therefore selected for investigations in this paper. This DCGAN method first extracts the image features and then reproduces a totally new image from these abstract representations.

Table 2 Comparison and weighted evaluation of the different commonly used GAN techniques GAN, DCGAN, WGAN, PGGAN and AE methods GMMNAE, VAR, SMCAE, AAE from the literature

Criteria	Weights range	GAN				AE			
		GAN Goodfellow et al. (2014)	DCGAN Radford et al. (2016)	WGAN Arjovsky et al. (2017)	PGGAN Karras et al. (2018)	GMMN-AE Li et al. (2015)	VAE Jorge et al. (2018)	SMCAE Zhang et al. (2015)	AAE Makhzani et al. (2016)
Match with real images	[4.5, 5.5]	3	4	4	4	3	3	3	4
Diversity of generated images	[3.5, 4.5]	3	3	4	4	3	3	3	3
Resolution/sharpness	[3.5, 4.5]	3	4	4	5	3	3	3	3
SNR	[3.5, 4.5]	3	5	5	5	5	5	5	5
Architectures complexity	[1.5, 2.5]	4	4	2	1	–	4	4	2
Stability of technique	[2.5, 3.5]	3	4	5	4	–	5	5	5
Calculation effort	[1.5, 2.5]	4	3	3	4	–	3	3	3
Available theoretical background information	[1.5, 2.5]	5	5	3	1	–	5	3	2
Weighted average		[3.27, 3.35]	[3.97, 4.06]	[3.92, 4.00]	[3.77, 3.93]	[2.18, 2.44]	[3.73, 3.81]	[3.59, 3.66]	[3.51, 3.63]

The assessment criteria are also available in the literature. For the weighting of the individual criteria according to their importance for the assessment, an interval $[w_e - 0.5, w_e + 0.5]$ is specified to mind the robustness and influence of individual weights. The expected value range w_e of the weights varies between unimportant (1) and very important (5). The rating v_a ranges from 0 (no match) to 5 (absolutely correct). "–" indicates: Insufficient information available

Table 3 Various suitable GAN settings available in the literature are summarised and the quantity of values is presented

Parameters	Radford et al. (2016)	Perarnau et al. (2016)	Neff (2018)	Salimans et al. (2016)	Brownlee (2019)	Quantity of values
Image size	$64 \times 64 \times 3$	$64 \times 64 \times 3$	$256 \times 256 \times 2$	$128 \times 128 \times 3$	–	{ $64 \times 64 \times 3$, 128×128 x 3, $256 \times 256 \times 2$ }
Batch size	128	64	16	64	–	{ 16, 64, 128 }
Noise vector size	100	100	128	100	100	{ 100, 128 }
Optimiser	Adam	Adam	Adam	Adam	Adam	{ Adam }
Learning rate	0.0002	0.0002	0.0001	–	0.0002	{ 0.0001, 0.0002 }
β_1	0.5	0.5	0.5	0.5	0.5	{ 0.5 }
β_2	0.999	0.999	0.9	–	0.999	{ 0.9, 0.999 }
ϵ	0.00000001	0.00000001	0.000000001	–	0.00000001	{ 0.00000001 }
Activation func. Generator	ReLu	ReLu	ReLu	ReLu	ReLu/ LeakyReLu	{ ReLu, (LeakyReLu) }
Activation func. Discriminator	LeakyReLu	LeakyReLu	ReLu	LeakyReLu	LeakyReLu	{ ReLu, LeakyReLu }
α size	0.2	0.2	–	0.2	0.2	{ 0.2 }
Kernel size	5×5	4×4	5×5	5×5	–	{ 4×4, 5 x 5 }
Initial weights	0 ± 0.02	–	He Ini.	–	0 ± 0.02	{ 0 ± 0.02, He Ini. }
Batch normalisation	Generator + Discriminator	Generator + Discriminator	Generator + Discriminator	Generator + Discriminator	Generator + Discriminator	{ Generator + Discriminator }
Dropout	0	0	0	Yes (Y), value unknown	–	{ Y, 0 }
Dimension expansion	Upsampling to 2×2	Upsampling to 2×2	Upsampling to 2×2	–	Upsampling to 2×2	{ Upsampling to 2 x 2 }
Dimension reduction	Convolution with stride 2	Convolution with stride 2	Convolution with stride 2	Convolution with stride 2	Convolution with stride 2	{ Convolution with stride 2 }
Convolution layers	4	5	6	6	–	{ 4, 5, 6 }
Number Kernels	512 to 128	512 to 64	1024 to 64	512 to 64	–	{ 512 to 64, 1024 to 64, 512 to 128 }

The other deep learning augmentation methods from the "Image data augmentation techniques" section were not considered since the focus of this paper was the investigation of the usability of such data enhancement techniques rather than the detailed validation of many different methods. Subsequently, techniques to evaluate the quality of synthetic fibre placement defect images are discussed. Such an analysis is essential to evaluate the quality of the artificially generated depth images.

Performance assessment of GAN based synthesised data

In order to evaluate the performance of a GAN for the generation of synthetic image data of fiber placement defects suitable assessment methods have to be selected regarding this application. Therefore, Borji (2019) summarised several methods for assessing the performance of GAN techniques. Besides the manual, visual assessment of an image, Borji (2019) outlined the further sensible GAN-Train GAN-Test method. Shmelkov et al. (2018) suggested and developed this technique with the aim to evaluate the variety and quality of the generated images. This method is based on a two-step approach using the real input data and the artificially generated images. For the GAN-Train step a classifying ANN is trained with the generated images from a GAN. The performance is measured by classifying the real images with the previously mentioned ANN. For the GAN-Test step the classifier ANN is trained with real data. The generated images from the GAN are used for the automated assessment of the ANN classification results.

The GAN-Train GAN-Test method enables an evaluation of the diversity and realism of the generated, artificial defect image data without the need for an unavailable reference data set or another pre-trained ANN. The GAN-Train method primarily provides information about the diversity but also about the realism of the generated images. In contrast, the GAN-Test method focuses on investigating only the realism of the synthetic images. However, the two observations cannot be sharply separated. This means that the results should always be interpreted jointly.

For the investigations in this paper, a CNN was applied to classify the images during every GAN-Train GAN-Test assessment. Such a CNN can significantly reduce the number of weights needed to train an ANN since it uses kernels, which examine individual parts of the input data incrementally. The number of weights required depends on the number and size of the applied kernels. Therefore, fewer parameters have to be trained than in an ANN without kernels. This approach improves the efficiency of the classifier (Khan et al. 2020; Vasilev et al. 2019).

With the aim to focus on the actual image quality analysis, only a rudimentary CNN is applied for the GAN-Train

GAN-Test evaluation in this paper. Therefore, Khan et al. (2020) presented feasible CNN architectures and parametrisations to use in combination with the GAN-Train GAN-Test method. In addition, Chen et al. (2018) explained an approach particularly suitable for the AFP inspection.

The following section explaines the procedures for testing and evaluation.

Methodology

This section gives details on the experimental setup as well as the test procedure and evaluation. For the studies in this paper appropriate defect types were chosen. According to those introduced in "Manufacturing process" section, no defect regions, wrinkles, twists, foils representing foreign bodies, gaps and overlaps were selected for the following studies. Figure 2 schematically illustrates these defect types. Accordingly, Fig. 3 displays six randomly selected and smoothed real defect images per class, which were used as inputs for examinations in this paper. They have been acquired using the experimental setup described below. The individual defect images were manually labeled in the overall LLSS scan image using the tool LabelImg (Tzutalin 2015). Based on these labels, the individual defect images were extracted and used individually for the experiments. For the investigations considered, we used a different number of defect images per defect class. This is because defect types like gaps and overlaps can be considered as the combination of many individual partial areas and thus several real and independent defect images can be extracted from a single defect. In certain cases defects are located quite close to the edge of the overall defect image in our database. Thus, they become useless as training samples due to pre-processing steps and filter effects at the edge of the image. All origin input images were previously resized to a reasonable size of 128×128 px. This image size was chosen because the essential characteristics of a defect are still represented here, but the amount of data has been significantly reduced. Larger images may require additional layers in the ANN which in turn increases the training effort. The actual amount of data considered per defect type and the corresponding rounded half amounts are presented in Table 4. These "rounded half amounts" are needed for data compilation at a later stage and hence they are mentioned here.

In order to perform reliable investigations, representative original data must be acquired. These fibre layup defect images must be generated in a reproducible and representative way with respect to the actual fiber placement process. For this reason, a feasible experimental setup was applied, as shown in Fig. 4. This assembly is independent of disturbing influences from the manufacturing process such as contamination, thermal radiation or tilting of the layup effector. This test setup consisted of a KUKA jointed-arm robot,

Table 4 The number of available images per defect type are listed

Defect type	No defect	Wrinkle	Twist	Foreign body	Overlap	Gap
Numb. of defects	86	49	53	22	166	93
Half amount of defect	43	25	27	11	83	47

These are the maximum numbers of usable data sets per class. Additionally, the corresponding rounded half amounts of images are listed, as these are needed for the data compilation in Table 8
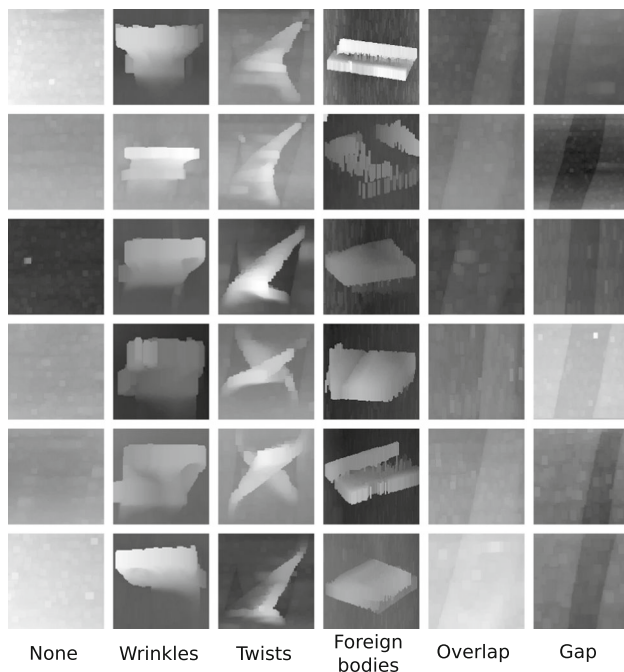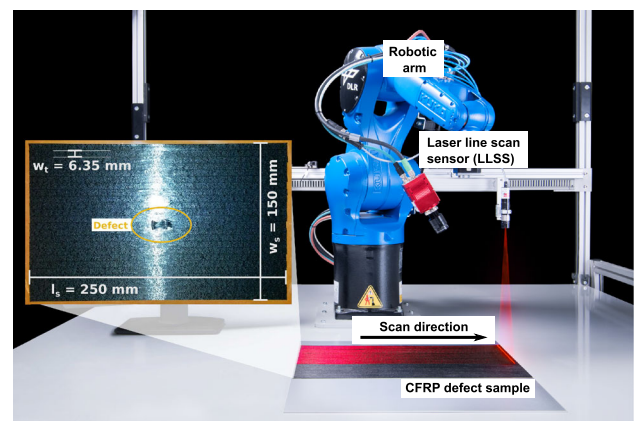

Fig. 3 Six randomly selected and smoothed real defect grayscale depth images per class were applied as inputs in this paper. These have the image dimension of 128×128 px each and were captures by the LLSS presented in Fig. 4

Fig. 4 The experimental setup for image data acquisition is illustrated. A KUKA robot with attached C5 LLSS is used. This machine carried out a linear motion parallel to the material surface

the *Automation Technology GmbH* (AT) C5-4090 LLSS (Automation Technology GmbH 2019) and a CFRP prepreg material sample.

Image data acquisition and processing

The previously mentioned AT C5 sensor captured 16-bit grayscale depth images of dimensions 4096 (W) x 500 (H) px representing the topology of a 250 x 150 mm fiber layup sample. The width of the measurement image results from the maximum resolution in the width direction of the installed AMS CMV12000 sensor chip (ams AG 2020). The height resolution is determined by the exposure time per pixel line and the time between the acquisition of individual height profile lines. Accordingly, the image resolution decreases with increasing exposure time for the same sample size and equivalent scanning velocity. A laser voltage of 5V was applied to determine precise topological information using the FIR-PEAK laser line detection algorithm (Automation

Technology GmbH 2014). The FIR-PEAK method involves a derivative filter that detects the zero crossing point of the first derivative of the laser intensity image. The recorded image data was transmitted via an Ethernet connection using the GenICam protocol (European Machine Vision Association 2009).

The scanning of the fiber placement defect samples was performed by moving the robot arm linearly along the entire sample at a velocity of 200 mm/s.

All calculations in this paper were performed on a computer with an Intel Xeon Gold 5122 @ 3.60 GHz CPU, 48 GB RAM and a NVIDIA Quadro P6000 GPU. Furthermore, OpenCV 3.4.1 (Bradski 2000), Keras 2.2.4 and Tensorflow 1.13.1 were used in conjunction with Python 3.7.5. The training of all ANN investigated in this paper were carried out on the GPU.

Data augmentation methods

The artificial augmentation of the data set under consideration was carried out by means of Geometrical Transformation as a traditional technique and the conditional DCGAN as a deep learning based method. The variation of the individual parameters of both methods was performed on the basis of the application case and the literature. The quality of the input images as well as an appropriate parameterisation of the synthesis methods depends on a large number of different factors. Within the scope of this paper, a suitable image

pre-processing was applied to adjust brightness and contrast. Furthermore, influential parameters of the image synthesis methods have been varied according to the literature. The investigations in this paper serve to give a basic overview of a reasonable configuration. However, the varied settings represent only a subset of all variations and serve as rough guidance values.

The Geometrical Transformation is presented in the "Image data augmentation techniques" section as an efficient but simple to use method for data enhancement. In order to carry out a meaningful image data augmentation according to the application, certain value ranges were assigned to the applied Geometrical Transformations. These are presented in Table 5. For the plausibility of these value ranges, we have considered that the LLSS applied for data acquisition determines linewise height profiles of the surface immediately after the fibre deposition, during the placement process. This means that the orientation of the laid up tows can only rotate slightly. This implies that most defect types must be aligned along the moving direction of the effector. Only defects which are not directly related to the fibre material can freely vary their orientation. The image size of the individual defects within the measurement image is also limited since the distance of the sensor to the surface varies only marginally during the production process. The disadvantage of this method is, that the images are only modified geometrically, but the diversity of the images is not changed.

A convenient alternative is the DCGAN approach. The DCGAN architecture applied in this paper was developed on the basis of designs from the literature. Taking these into account, parameters with an anticipated large influence on the result and the corresponding reasonable value ranges were determined (Radford et al. 2016; Brownlee 2019). The configurations from the literature are presented in the Table 3. The derived test parameters are listed in the Table 6. A basic configuration of the DCGAN from Radford et al. (2016) is presented, which was varied for certain key parameters. These basic parameters were considered to be suitable parameters for a stable DCGAN by both Radford et al. (2016) and Brownlee (2019). Thus, this parameter set has the best maturity level of the presented literature. In order to determine reasonable settings for the batch size, layer structures and DCGAN parameters, three preliminary tests were performed. The different combinations of parameters were applied in tests and the generated defect images were compared visually in order to find a feasible configuration for answering the research question. Inspired from the data sets of the case studies mentioned in "Image data augmentation techniques" section, 5000 images were used for the training of the DCGAN. In this paper the associated class labels are attached to the first layer using the Keras *Concatenate* function. As discussed above, we cannot give an exact value for the necessary amount of training data. However, after reviewing the very

different examples from the literature, this specified number of training samples seemed reasonable for the use case considered. In order to clarify this once more, our approach was focused on demonstrating the feasibility of enhancing an depth image inspection database for this application case and finding a suitable setting. However, we only rudimentary investigated the performances of different parameter settings.

Validation methods

For the analysis of the synthetic defect images an appropriate assessment method was required. However, it must only use the origin input images themselves or the data generated during the process. Here it is noteworthy that the images must be evaluated with regard to their diversity, realism and defect orientation. Furthermore the applicability of these generated images for machine learning methods is of interest. For this purpose, a plain cross-validation or a data separation into a validation and a test data set is only possible to a limited extent. As mentioned above, a subset of those real input images is shown in the Fig. 3 and was intended to serve the traceability of the manual, visual assessments in this paper. On the basis of the aspects mentioned in "Performance assessment of GAN based synthesised data" section, the GAN-Train GAN-Test method appeared as a promising and easy to use technique in addition to the manual, visual assessment. Within this Paper the GAN-Train GAN-Tests results are presented using confusion matrices with the actual class displayed on the ordinate and the predicted class on the abscissa.

A CNN classifier was applied for the validation with the GAN-Train GAN-Test method in this paper. Due to the previously explained efficient operation principle, CNN are particularly well suited for the classification of the image data for the GAN-Train GAN-Test evaluation in this paper. The utilised architecture of the CNN classifier was build up on the conceptual ideas and the architecture of Chen et al. (2018), which have already successfully applied and validated their approach for image-based inspection in the AFP process. For this reason, their network architecture was considered to be appropriate for the application under consideration in this paper.

We did not make use of pre-trained ANN for the experiments presented here since the literature discussed in "Image data augmentation techniques" section indicates that these pre-trained networks perform rather poorly or similarly for the considered AFP inspection application. However, for the considered AFP inspection case pre-trained networks presumably do not provide a significant performance advantage over self-trained ANN. This referred to both, the classifier used for validation and the GAN applied for data synthesis.

Table 5 For the application case of the AFP fibre material layup of several narrow tows in parallel, defect-type dependent, reasonable parameters for the Geometrical Transformation are given here

Performed operation	Wrinkle	Foreign body	Twists, Gaps, Overl., No Def.	Step size
Vertical shift	0 ± 10 px	0 ± 10 px	0 ± 10 px	1 px
Horizontal shift	0 ± 10 px	0 ± 10 px	0 ± 10 px	1 px
Vertical mirroring	No	Yes	Yes	1
Horizontal mirroring	Yes	Yes	Yes	1
Rotation	$0 \pm 10°$	$0 \pm 45°$	$0 \pm 10°$	$1°$
Scaling	$0 \pm 4\%$	$0 \pm 4\%$	$0 \pm 4\%$	1 %
Brightness adjustment	$0 \pm 30\%$	$0 \pm 30\%$	$0 \pm 30\%$	1 %
Possible variations	10168578	88127676	20337156	–

Table 6 Different feasible architectures and setting are available in the literature and thus given here for application in this paper

Parameter	Base config.	Variations	References
Image size	128×128	–	Radford et al. (2016)
Batch size	128	64	Radford et al. (2016)
Noise vector size	100	–	Radford et al. (2016)
Optimiser	Adam	–	Radford et al. (2016)
Learning rate	0.0002	0.0001	Radford et al. (2016) and Neff (2018)
β_1	0.5	–	Radford et al. (2016)
β_2	0.999	0.9	Radford et al. (2016) and Neff (2018)
ϵ	0.00000001	–	Radford et al. (2016)
Activation func. Generator	ReLU	–	Radford et al. (2016)
Activation func. Discriminator	Leaky ReLU	–	Radford et al. (2016)
α	0.2	–	Radford et al. (2016)
Kernel size	5×5	–	Radford et al. (2016)
Initial weights	0 ± 0.02	–	Radford et al. (2016)
Batch normalisation	Generator, Discriminator	–	Radford et al. (2016)
Dropout	0	0.25, 0.5	Radford et al. (2016) and Odena et al. (2017)
Dimension expansion	Upsampling to 2x2	–	Radford et al. (2016)
Dimension reduction	Convolution with stride 2	–	Radford et al. (2016)
Convolution layers (Gener./Discr.)	5/5	6/6, 5/6, 6/5	Radford et al. (2016), Neff (2018), Salimans et al. (2016), Odena et al. (2017) and Gulrajani et al. (2017)
Number Kernels	256 to 16	256 to 8	Radford et al. (2016)

Experiments

As mentioned at the beginning of this section, three preliminary experiments were carried out to sequentially estimate reasonable parameters. Afterwards, two validation experiments were performed. The applied settings correspond either to the basic configuration mentioned in the Table 6 or to the optimised value, if the corresponding parameter has already been investigated. This approach also gives an impression of the sensitivity of an algorithm regarding the considered parameters.

Preliminary tests

We used the method of visual image quality assessment for all three preliminary tests for making the decisions. In order to obtain a trustworthy result all preliminary tests were repeated three times redundantly for each parameter examined and the generated results were analysed. For the first two preliminary tests 25,000 epochs were run for the DCGAN training. The generated images were observed after 250 epochs each. In the first two preliminary experiments the highest quality image result was usually achieved after 16,000 to 20,000 epochs. Thus, from the third preliminary experiment onwards only

20,000 epochs were run for each training. The utilised visual assessment approach was described above in the "Performance assessment of GAN based synthesised data" section.

The first experiment aimed at the selection of a suitable batch size. On the basis of the literature we only had to choose between batch size 64 and 128. Furthermore, it is noteworthy that a larger batch size results in an increased training effort of the ANN when considering an equivalent number of epochs passed.

The second preliminary experiment served to estimate a suitable structure of the convolution layers for the generator and discriminator of the DCGAN. After comparing the literature from Table 3, the necessary tests were limited to the four reasonable configurations with five and six layers for each of the two GAN components.

Subsequently, in the third preliminary test all combinations of the actual DCGAN parameters learning rate, β_2 and dropout factor were investigated using the parameters mentioned in Table 3. The best performing parameter combination was determined according to the needs for stability and quality of the generated images for the application case considered.

Subsequently, with the aim to answer the research question, two validation tests were performed under consideration of the previously determined test parameters.

Validation experiments

The aim of this first validation experiment was the investigation of the quality and diversity of images synthetically generated by the conditional DCGAN. For this purpose, a manual, visual assessment as well as a GAN-Train GAN-Test evaluation was carried out and the results were discussed. In order to check the robustness of the results three different synthetic data sets with 5000 defect images per class were utilised. The individual runs and the corresponding data sets are presented in the test matrix from Table 7. The CNN classifier mentioned above was applied for the automated image classification within the GAN-Train GAN-Test approach. The previously determined, best suited parameters for the DCGAN as well as the training weights of the GAN from the previous experiments were applied. Thus, three synthetic defect image data sets AUG_DCGAN_<N> are generated. Regarding the GAN-Train runs 1.x the artificial DCGAN images were used as training data sets for the CNN classifier. Thus, the Geometrical Transformation enhanced data set AUG_GT_All were applied for validation. For the GAN-Test runs 2.x the data sets were used vise versa. The traditionally augmented data set was applied for training the CNN classifier and the images generated by the DCGAN were used as validation data for the classifier.

In order to answer the research question, the second validation experiment investigated the applicability of the

Table 7 The data compilation for the first validation test is listed

	Run	DS training	DS test
GAN-Train	1.1	AUG_DCGAN_1	AUG_GT_All
	1.2	AUG_DCGAN_2	AUG_GT_All
	1.3	AUG_DCGAN_3	AUG_GT_All
GAN-Test	2.1	AUG_GT_All	AUG_DCGAN_1
	2.2	AUG_GT_All	AUG_DCGAN_2
	2.3	AUG_GT_All	AUG_DCGAN_3

DS: Data set; AUG_GT_All: Data set containing 5000 images which have been generated by geometrical image transformation including the original input data; AUG_DCGAN_<N> consists of 5000 images which have been generated in different runs <N> using the DCGAN with given weights

considered methods for the synthesis of differently sized and diverse composed image data sets. For this purpose, the classification performance of a CNN classifier was evaluated for different training and validation data. Therefore, the real input defect images, the traditionally enhanced data and the image data generated by the DCGAN were again analysed using the GAN-Train GAN-Test method. The performed experiments and the corresponding data sets are listed in Table 8.

The data set AUG_GT_X were created by Geometrical Transformation with the settings from Table 5. This Geomatrical Transformation was based on a certain amount X of randomly selected real input images. AUG_GT_10 was therefore based on ten origin input images per class, AUG_GT_Half on half of all available input images per class and AUG_GT_ALL on all existing real input images per defect class. Except for the AUG_GT_All data set, the other individual data sets were based on different original data than were used as test data sets in the experiment. This means that the original test data was never part of the actual training database.

The previous test has shown the high quality, realism and diversity of the images from the data set AUG_DCGAN_2. Furthermore, the prior tests have indicated only marginal differences between the individual data sets generated by the DCGAN. Therefore, for comparability in this experiment just the data set AUG_DCGAN_2 was applied for the GAN-Train GAN-Test procedure. All the corresponding results are presented in the following section.

Results

In the following, the results of the three preliminary tests and the final two validation experiments are presented and analysed.

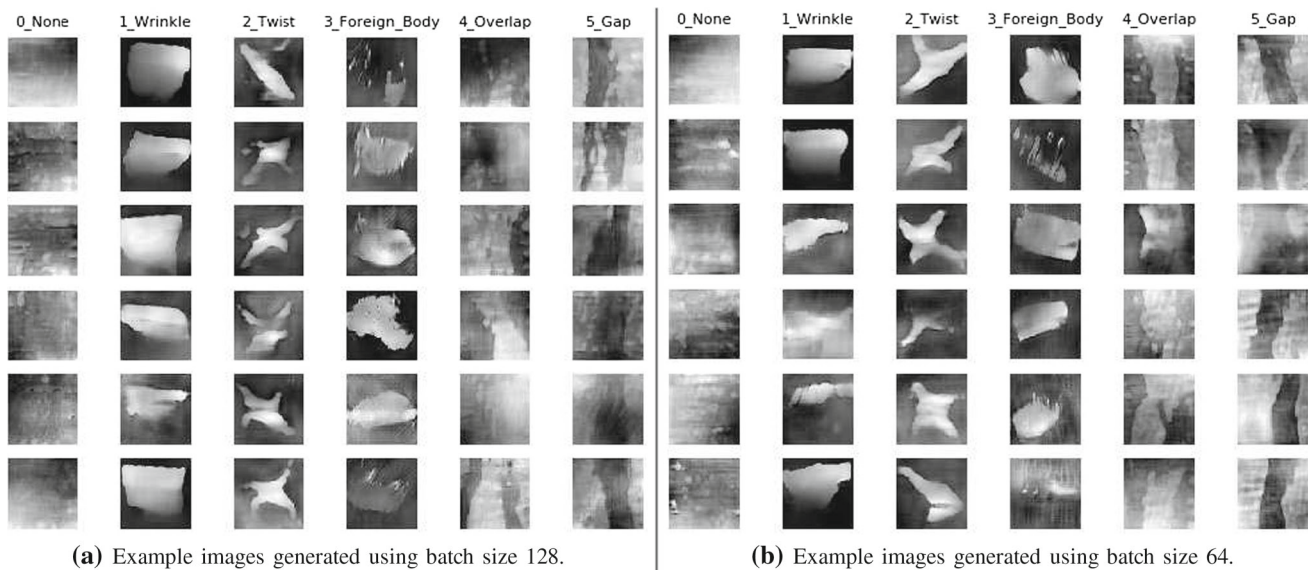**Table 8** The data compilation for the second validation test is listed

	Run	DS training	DS test
Geometrical Transform. Augm.	1.1	AUG_GT_10	RE_All−10
	1.2	AUG_GT_Half	RE_Half
	1.3	AUG_GT_All	RE_All
DCGAN Augm.	2.1	AUG_DCGAN_2	RE_All−10
	2.2	AUG_DCGAN_2	RE_Half
	2.3	AUG_DCGAN_2	RE_All

DS: Data set; AUG_GT_<N> contains <N> randomly chosen images which have been generated by geometrical image transformation excluding the input data. AUG_DCGAN_2 is the best performing data set from the Table 7. RE_<X> represent the collection of selected original data <X> for tests

Table 9 The visual evaluation results considering the batch sizes 64 and 128 are presented

	None	Wrinkle	Twist	Foreign body	Overlap	Gap
Batch size 128	+	+	+	+	○	○
Batch size 64	+	+	+	+	+↓	+↓

+: Good; ○: Medium; −: Bad; ↑: Tends to be better; ↓: Tends to be worse

(a) Example images generated using batch size 128.

(b) Example images generated using batch size 64.

Fig. 5 Synthetically generated images considering two different batch sizes for DCGAN image augmentation

Preliminary tests

Table 9 presents the results of the manual visual image assessment for the generated conditional DCGAN data with batch sizes 64 and 128. Both perform very similarly, as you can see from the Fig. 5.

Only for gap and overlap defects the DCGAN with batch size 64 generates slightly superior quality images. Since a manual and therefore uncertain evaluation method was used here, we can assume that there is no significant difference in the quality and variance of the images. However, it should be noted that for a comparable number of epochs the training time of an ANN increases with rising batch size, as already described in "Methodology" section. Thus, it makes sense to choose a batch size that is as small as possible but of sufficient

quality. For this reason we have selected batch size 64 for this use case and for the following experiments.

In the second preliminary test the generated images from different previously defined DCGAN generator (G) and discriminator (D) layer structures (G/D) are examined. The results of the manual visual image assessment are evaluated qualitatively and compared in Table 10.

These results indicate again only a perceptible difference for gaps and overlaps. We can see clearly that the DCGAN architectures (6/5) and (6/6) generate the best quality synthetic images. The structure (5/5) produces only mediocre artificial gap and overlap defect images. The composition (5/6) synthesises especially very poor gap defect images. Since variant 3 with the architecture (6/5) presumably yield

Table 10 Visual evaluation results for different DCGAN Generator (G) and Diskriminator (D) layer structures are displayed

Var. x (G/D) [Kernel]	None	Wrinkle	Twist	Foreign body	Over-lap	Gap
Var. 1 (5/5) [256..16]	+	+	+	+	○	○
Var. 2 (5/6) [256..8]	+	+	+	+	+↓	−
Var. 3 (6/5) [256..8]	+↑	+↑	+↑	+↑	+↑	+↑
Var. 4 (6/6) [256..8]	+	+	+	+	+	+

+: Good; ○: Medium; −: Bad; ↑: Tends to be better; ↓: Tends to be worse

the best defect images for all classes, this configuration is applied for further investigations in this paper. It should be noted, that the difference to the architecture (6/6) is maginal and therefore this configuration provides a reasonable alternative. However, since the focus in this paper is on the investigation of the general feasibility of an artificial image data augmentation of this particular AFP defect images, only the presumably best configuration (6/5) was considered for the subsequent experiments.

The parameter configurations presented in Table 11 were compared in the third preliminary experiment. The resulting quality of the generated synthetic images was evaluated visually.

The performance of the individual settings is color coded. The parameter sets 1, 7 and 10 generate qualitatively good synthetic images. The other settings create rather poor or unsuitable images. Except for the configurations 4 and 11, a dropout factor > 0 seems to have a major negative impact on the quality and variety of the synthetically generated images. For setting 11 this deterioration in quality is also evident but considerably less than in the other configurations with an equal dropout factor of 0.25. With setting 4 the image quality is poor and the variety between the images is smaller, despite a dropout factor of 0. This setting combines a learning rate of 0.0001 with a β_2 of 0.9. It is possible that the combination of both parameters has a significant influence on the quality of the synthesised images for the considered depth map data set of the AFP fibre layup defects. However, a generally valid conclusion cannot be derived from this since the settings of the DCGAN and the resulting synthetic images are highly depend on the input data set. Based on the visual assessment, configuration 1 generates the highest quality synthetic defect images. Furthermore, this configuration contains a presumably beneficial learning rate and β_2 parameter combination. These are the reasons for choosing this parameter set 1 for the subsequent validation experiments. Figure 6 illustrates a visual representation of six randomly selected images per class which were generated synthetically with the DCGAN using parameter set 1. Consequently, the following network architecture from Table 12 is applied for the conditional DCGAN for the following validation experiments. The corresponding specific layer structure for the Generator is presented in Table 13 and for the Discriminator

Table 11 Results from the visual image quality assessment of artificial defect images considering various DCGAN parameter sets are listed and color-coded according to their performances

Parameter sets:		**1**	2	3	*4*	5	6	**7**	8	9	**10**	11	12
Learn.rate	0.0001	x	x	x	*x*	x	x						
	0.0002							x	x	x	**x**	x	x
β_2	0.999	x	x	x				x	x	x			
	0.9				*x*	x	x				x	x	x
Drop-out	0.0	**x**			*x*			**x**			**x**		
	0.25		*x*			*x*			*x*			*x*	
	<u>0.5</u>			<u>x</u>			<u>x</u>			<u>x</u>			<u>x</u>

Good result: Bold; Medium result: Italic; Medium (Bad) result: Bold Italic; Bad result: Underlined
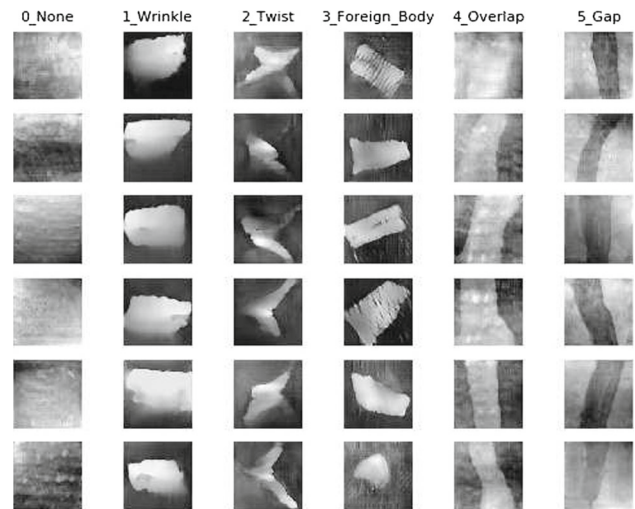
Fig. 6 Six randomly selected, synthetic images per class using parameter set 1 from Table 11 and the conditional DCGAN after 20000 epochs of training

in Table 14. Accordingly, for the Generator 99.98% and for the Discriminator 99.91% of all the parameters are trainable.

The training of the DCGAN takes about 77 min in this scenario. For this the settings and computing hardware specified above were used for passing through the 20000 epochs. The generation of 5000 training images for each of the six classes, making a total of 30000 images, takes < 3 min in this experiment.

Table 12 The list give the key parameters for the DCGAN architecture and its configuration applied for validation tests in this study

Parameter	Value
Image size	$128 \times 128 \times 1$
Batch size	64
Noise vector size	100
Optimiser	Adam
Learning rate	0.0001
β_1	0.5
β_2	0.999
ϵ	0.00000001
Activation func. Generator	ReLU
Activation func. Discriminator	Leaky ReLU
α	0.2
Kernel size	5 x 5
Initial weights	0 ± 0.02
Batch normalisation	Generator, Discriminator
Dropout	0
Dimension expansion	Upsampling to 2×2
Dimension reduction	Convolution with stride 2
Convolution layers (Gener./Discr.)	6/5
Number Kernels	256 to 8 (descending with powers of 2)

Validation experiments

Within this first validation experiment the quality and diversity of the generated images are investigated. Furthermore, the manual visual assessment is compared with the results from the GAN-Train GAN-Test evaluation. To review the realism of the synthetically generated images, they are compared with the illustrated example images from Fig. 3. For this purpose, the conditional DCGAN with the previously defined parameter set was applied. Figure 7 gives the mean value and the standard deviation of the three similar performed runs using the data sets from Table 7. The GAN-Train results are presented on the left and the GAN-Test results on the right hand side of the figure.

When looking at the GAN-Train confusion matrix we notice values > 88% along the diagonal for all class assignments, except the non defective assignment. This has a classification rate of only 78.07%. These results indicate that the diversity of the defect patterns are fairly high. Beyond that this also means that the non defective test patterns look very similar to each other. The large standard deviation of $\sigma = 23.36\%$ for the three generated data sets further indicates that the CNN classifier probably has difficulties in deriving suitable features from the non defect images. Obviously, this finding is plausible as an accurate and reliable fiber placement process is designed to achieve a consistently good fibre placement quality. This results in a very smooth LLSS depth image. The diversity of the images without defects thus should be slightly less than for the images with defects. Actual defect images are therefore subject to

very strong variations in the appearance due to the characteristic defect shape. Furthermore, we notice that non defect images are classified as overlap defects with a mean value of 17.77%. This is likely due to the fact that the overlap defects also have less distinctive geometric attributes present in a LLSS scan image. Additionally this could indicate that the classifier applied for this validation needs to be properly configured to correctly distinguish between these two types of defects. However, it is also conceivable that the DCGAN generates an insufficient representation of these overlap and non-defect images. The comparatively high standard deviation of $\sigma = 21.72\%$ is another indicator for potential deficits in the synthetic data generation using DCGAN or in the GAN-Train GAN-Test evaluation with the CNN classifier. However, since the synthetic non-defect images and the overlap images from the Fig. 5 and the Fig. 6 are visually very well distinguishable, we assume here that the deviating classification results rather indicate an insufficient configuration of the CNN classifier regarding these particular defect types. For the geometrically more complex defect types wrinkle and twist even mean values of > 96% are yielded. This is due to the very characteristic shape which can simply be varied from an image generator. Furthermore, these defect geometries can be mapped easily to the feature maps of a CNN classifier. The result for the GAN-Test investigations differs slightly with regard to the value range of all mean values and regarding the weakest classification result. In this observation all mean classification results are > 94%, except for overlap defects with 87.36% mean classification rate. Overlap fiber

(a) GAN-Train results (mean and standard deviation σ) considering three individual runs and given in %. The classifying CNN was therefore trained with the DCGAN augmented data and tested with the Geometrical Transformation data.

(b) GAN-Test results (mean and standard deviation σ) considering three individual runs and given in %. The classifying CNN was therefore trained with the the Geometrical Transformation data and tested with DCGAN augmented data.

Fig. 7 Results from GAN-Train GAN-Test evaluation, averaged over three individual runs and using the data sets previously described in the Table 7. An individually trained CNN classifier is applied to generate the results

Table 13 The table illustrates the layer architecture and the amount of parameters of the DCGAN generator (G) with six convolution layers, which is applied for validation experiments

Layer type	Output shape
Noise vector + labels	*(None, 106, 1, 1)*
Dense	(None, 2048)
Reshape	(None, 2, 2, 512)
UpSampling2D	(None, 4, 4, 512)
Conv2D	(None, 4, 4, 256)
Batch normalization	(None, 4, 4, 256)
ReLU activation	(None, 4, 4, 256)
UpSampling2D	(None, 8, 8, 256)
Conv2D	(None, 8, 8, 128)
Batch normalization	(None, 8, 8, 128)
ReLU activation	(None, 8, 8, 128)
UpSampling2D	(None, 16, 16, 128)
Conv2D	(None, 16, 16, 64)
Batch normalization	(None, 16, 16, 64)
ReLU activation	(None, 16, 16, 64)
UpSampling2D	(None, 32, 32, 64)
Conv2D	(None, 32, 32, 32)
Batch normalization	(None, 32, 32, 32)
ReLU activation	(None, 32, 32, 32)
UpSampling2D	(None, 64, 64, 32)
Conv2D	(None, 64, 64, 16)
Batch normalization	(None, 64, 64, 16)
ReLU activation	(None, 64, 64, 16)
UpSampling2D	(None, 128, 128, 16)
Conv2D	(None, 128, 128, 1)
ReLU activation	(None, 128, 128, 1)

Total parameters: 4586817 | Trainable parameters: 4585825 | Non-trainable parameters: 992

Table 14 The table illustrates the layer architecture and the amount of parameters of the DCGAN discriminator (D) with five convolution layers, which is applied for validation experiments

Layer type	Output shape
Image + labels	*(None, 128, 128, 7)*
Conv2D	(None, 64, 64, 16)
Batch normalization	(None, 64, 64, 16)
LeakyReLU	(None, 64, 64, 16)
Conv2D	(None, 32, 32, 32)
Batch normalization	(None, 32, 32, 32)
LeakyReLU	(None, 32, 32, 32)
Conv2D	(None, 16, 16, 64)
Batch normalization	(None, 16, 16, 64)
LeakyReLU	(None, 16, 16, 64)
Conv2D	(None, 8, 8, 128)
Batch normalization	(None, 8, 8, 128)
LeakyReLU	(None, 8, 8, 128)
Conv2D	(None, 4, 4, 256)
Batch normalization	(None, 4, 4, 256)
LeakyReLU	(None, 4, 4, 256)
Flatten	(None, 4096)
Dense	(None, 1)

Total parameters: 1097377 | Trainable parameters: 1096385 | Non-trainable parameters: 992

placement defects appear very similar to gap defects or non defect images. Therefore the classifier is more likely to identify these types as gap defects or non defect images. This is plausible for the mix-up of gaps as well as overlaps and basically matches the visual impression when viewing a LLSS scan image. Nevertheless, this result is different from the correct classification, which leads to a decrease of this value. The mixing up of overlaps and non-defect images probably

has a similar origin as discussed above for the GAN-Train results.

In addition, Table 15 presents the results for a comparison of the false positive and false negative results. The false positives values refer to the amount of no defects which are categorised as defect. False negatives are the number of defects which are recognised as no defect. This false negative rate has a special meaning here, because it describes the amount of defects that are missed. This is particularly problematic in a manufacturing process.

The numbers in the table indicate the influence of gaps and overlaps on the individual false values. Without considering gaps and overlaps the false values are < 2 %. Taking all defect types into account these false values vary between 2.65 % and 21.93 %. In this case the GAN-Train values are significantly larger than for the GAN-Test scenario. This relationship changes substantially for the observations without gaps and overlaps. In particular it should be noted that the false negative rate for the GAN-Train evaluation with 0.24 % is significantly lower than for the GAN-Test procedure. This indicates that the CNN classifier has a significantly better performance with respect to critical defect misses if trained with the synthetic DCGAN data. However, this is only valid if the gap and overlap defects are excluded.

The following second validation experiment serves as a comparison of the Geometrical Transformation and the conditional DCGAN for differently sized initial input data sets. The used data sets are introduced in "Methodology" section in Table 8, with the aim of performing a slightly modified GAN-Train GAN-Test evaluation. The corresponding results are presented in Fig. 8. For the results from the Fig. 8a, c, e ten images, half of the total or all of the available real training data are enlarged with the Geometrical Transformation to 5000 images for training the CNN classifier. To generate the results in the Fig. 8b, d, f the previously introduced data set AUG_DCGAN_2 is applied for the CNN training. The test data sets each consist of the remaining available original data, which has not been previously utilised for the data augmentation.

The image data generated with the DCGAN provide classification results with a total mean classification rate of 90.17% for all the different data sets. Thus, this assessment appears to be relatively independent of the size of the test data set. We recognise once more a slightly increasing misclassification between no defect images, gaps and overlaps as previously discussed. Particularly noticeable is the increasing number of twists being classified as overlaps. This unexpected behaviour is especially noticeable when comparing the results for ten and all real training images per class, with miss classification rates of > 20%. However, this tendency is also clearly apparent when considering the half amount of available test images from the Fig. 8d, having a misclassification value of 7.69%.

For the training images generated via Geometrical Transformation a distinctly heterogeneous behaviour appears from the evaluation of the different data sets. When applying ten initial images for training and the remaining available images for tests obvious classification deficits are evident for the defect types twist, foreign body and overlap. This is displayed in the Fig. 8a. Furthermore, we notice that foreign bodies are often recognised as wrinkles or twists. However, the classification results for no defects, wrinkles and gaps are unexpectedly high compared to the classification rate of foreign bodies in this experiment. Compared to the findings from Fig. 8 of the DCGAN generated images using all-10 test samples it is clearly evident that ten input images are not enough to model a sufficient diversity of defects and to train a CNN, even after a geometric augmentation. In contrast, the classification of only a few test samples can lead to a similar diversity problem. This makes it difficult to assess their realism. Nevertheless a robustly trained classifier is capable of properly classifying such defects. The results in Fig. 8a appear unrepresentative in comparison to the remaining results of this study. They seemed affected from a beneficial or non-beneficial aggregation of the randomly composed training data set. Considering the results from Fig. 8, applying half of the available data set for the geometric augmentation leads to a significant increase in the classification rate compared to the usage of just ten initial training images per class. Noteworthy here is the increase in the classification rate for foreign bodies. Due to the small number of defect images available of this type only one more initial training image was additionally applied. This fact strengthens the previous assumption of the low representativeness of ten randomly selected initial training images. Except for foreign bodies and overlaps, the CNN classifier trained with the part data set listed in the Table 4 yields classification rates of > 95%. This indicates a sufficiently good CNN classification rate for the remaining defect types when trained with only 25 to 47 initial defect images, depending on the class. The relatively low classification rate for overlap defect images is quite surprising, since the applied part data set with 83 images contains the largest number of training images of all classes. Thus, the previous findings of this paper are strengthened that especially the characteristics of overlap defects are difficult to abstract appropriately using image features. When comparing the results from the Fig. 8e, f we realise that the classification rate for no defects, wrinkles, twists, and foreign bodies is 100%. For the difficult to characterise gaps and overlaps we observe classification rates of > 95%. This results in the very great mean classification rate of 98.89% with a standard deviation of $\sigma = 1.66\%$. Compared to these results, the CNN classifier trained with the DCGAN enhanced data set only yields a mean classification rate of 90.3% with a standard deviation of $\sigma = 51.81\%$. These results illustrate the limitations of the traditional data augmentation regarding the

Table 15 The estimated false positive (No defect → Defect) and false negative (Defects → No defects) values corresponding to the experiment in Fig. 7 are presented

	False positives (%)	False negatives (%)
GAN-Train	21.93	10.82
GAN-Train (no gap/overlap)	1.55	0.24
GAN-Test	2.65	8.88
GAN-Test (no gap/overlap)	0.13	1.48

AUG_GT_All: Data set containing 5000 images which have been generated by geometrical image transformation including the original input data; AUG_DCGAN_<N> consists of 5000 images which have been generated in different runs <N> using the DCGAN with given weights
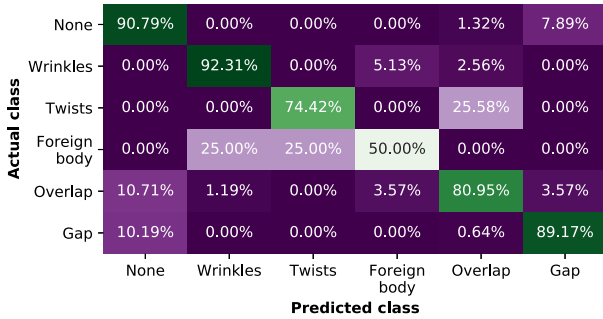
(a) DS training: AUG_GT_10; DS test: RE_All−10 from Table 8 run 1.1
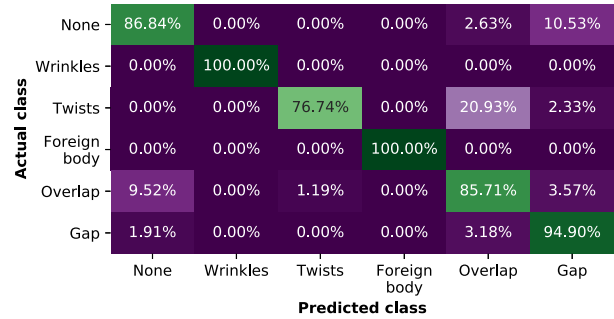
(b) DS training: AUG_DCGAN_2; DS test: RE_All−10 from Table 8 run 2.1
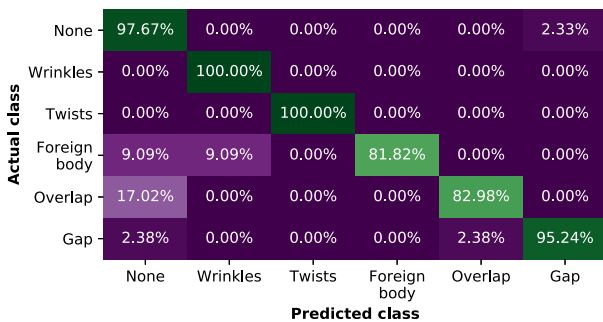
(c) DS training: AUG_GT_Half; DS test: RE_Half from Table 8 run 1.2
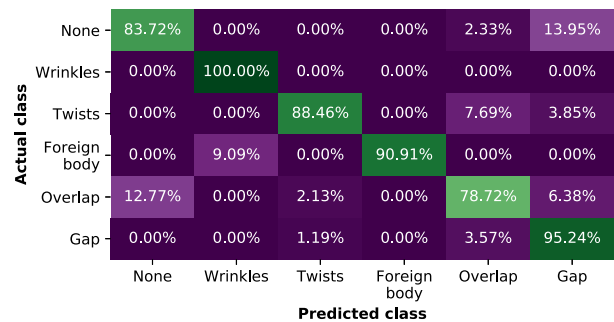
(d) DS training: AUG_DCGAN_2; DS test: RE_Half from Table 8 run 2.2
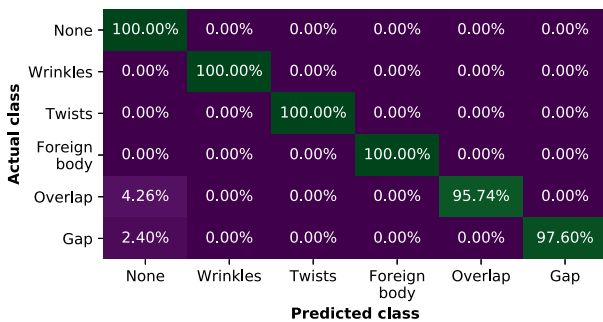
(e) DS training: AUG_GT_All; DS test: RE_All from Table 8 run 1.3
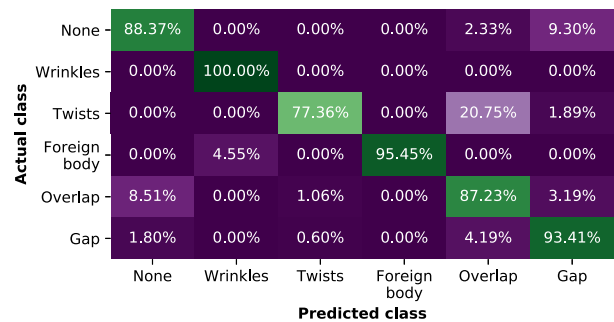
(f) DS training: AUG_DCGAN_2; DS test: RE_All from Table 8 run 2.3

Fig. 8 Results from the GAN-Train GAN-Test evaluation considering different amounts of comparison and test data, corresponding to data sets (DS) from the Table 8

diversity of the generated images. In this case, the images initially applied for the traditional data augmentation are only geometrically transformed, while the actual image content remains identical. Hence, obviously very good classification results are achieved when in this case a classifier is trained with these images which are basically the same for their subsequent classification. In contrast, the data generated with DCGAN obviously must have a different appearance than the original data. For this reason, the classification rate is usually < 100% with a significant standard deviation, despite good results have been achieved. In summary, we would like to mention again that the considered GAN-Train GAN-Test method serves to investigate the diversity and realism of different defect types for the stated image synthesis techniques. The GAN-Test results in the Fig. 8d, f partly show lower classification rates than the corresponding GAN-Train results in the Fig. 8c, e. This only indirectly characterises the performance of the DCGAN. As already mentioned above, a great advantage of data synthesis with DCGAN is the different but realistic appearance of the resulting artificial images compared to the real recorded defect images. As a result, the GAN-Test analysis often achieves classification rates of < 100%. As this indicates varying synthetic data, these results are desirable. The extremely high classification rates of often 100% in the GAN-Train analysis reveal an insufficient diversity for the image data augmentation with the Geometrical Transformation. Thus, this can quickly lead to an overfitting of a classifier and would not offer any added value in real applications. We also want to clarify that the original real input images provide the basis for the data augmentation for all the results presented above. However, they are not part of the actual CNN training data set. For the investigations of all-10 test images and half of the data set, the remaining original images are used for tests. This data has not previously been considered for data enhancement. For the tests on all real data these images were initially considered for data augmentation, but they are not part of the actual CNN training data set.

Additionally, Table 16 summarises the calculated false positive and false negative values corresponding to the results in Fig. 8. Similar to the investigations in Table 15 the false positives values describe the number of no defects predicted as defects and the false negatives consider the inverse case.

In addition to the results from Table 15 we see that the false negative rates for training with DCGAN synthesised training data are very similar for all runs. For training with Geometrical Transformation augmented data these false negative rates and the corresponding false positive values decreases continuously with increasing amount of applied input data. In particular for the runs x.1 and x.2 the false negatives rates using the DCGAN training data are much less than for training with the corresponding Geometrical Transformation training data. These results indicate that the data augmenta-

tion for fibre layup defects is very beneficial especially for very small data sets. Therefore it is crucial to apply a valid and high quality data augmentation model. Below the presented results are discussed and compared with related studies.

Discussion

The research discussed above of Radford et al. (2016), Perarnau et al. (2016), Neff (2018), Salimans et al. (2016) and Brownlee (2019) illustrate fundamental approaches for designing a DCGAN architecture. Unfortunately, these investigations are based on everyday pictures and medical image data. This data is rather different from AFP inspection images considered in this study. Thus, the investigations in this paper complement the ideas of Sacco et al. (2018). As they have suggested, we applied a GAN in this paper for artificial data augmentation, which they had briefly mentioned in their publication. Similar to Zambal et al. (2019b; 2019a), we demonstrated the possibility of generating synthetic AFP inspection topology data. Furthermore, we theoretically evaluated different data augmentation techniques and investigated selected approaches in detail. However, none of the previous studies cover the detailed investigation of different DCGAN configurations for the inspection in fiber composite manufacturing. These necessary investigations are initiated in this paper and different feasible methods are compared.

On the basis of a detailed literature analysis the traditional enhancement method Geometrical Transformation and the deep learning technique conditional DCGAN are determined as suitable methods for an artificial data enhancement. Their performances as well as their advantages and disadvantages are conclusively demonstrated in two validation experiments. We also demonstrated that for the classes no defect, wrinkles, twists, foreign bodies and gaps probably 25 to 47 representative origin defect images are already sufficient to achieve a good CNN classification result after a proper data augmentation to 5000 training images per class. This fits quite well with the results of Tabernik et al. (2019) which used 25 to 30 initial data sets for training their ANN classifier for the surface cracks detection.

Overlaps behave very unevenly in our investigations. Thus, for the use of overlaps for algorithm training or data augmentation probably considerably more data is required than was available for this work. However, it is also possible that the input data needs to be more representative.

In this regard, we need to mention that training the DCGAN with all possible variations of defect patterns is practically impossible. This is also not necessary, since the DCGAN abstracts the defect pattern and learns their appearance. Thus training the DCGAN with a various realistic defect images is more important than training with as many as possible images. In this study, fibre layup defects with

Table 16 The estimated false positive (No defect → Defect) and false negative (Defects → No defects) values corresponding to the experiment in Fig. 8 are presented

	False positives (%)	False negatives (%)
Run 1.1 (a) AUG_GT_10 training data	9.21	20.90
Run 2.1 (b) AUG_DCGAN_2 training data	13.16	11.43
Run 1.2 (c) AUG_GT_Half training data	2.33	26.11
Run 2.2 (d) AUG_DCGAN_2 training data	16.28	12.77
Run 1.3 (e) AUG_GT_All training data	0.00	6.66
Run 2.3 (f) AUG_DCGAN_2 training data	11.63	10.31

DS: Data set; AUG_GT_All: Data set containing 5000 images which have been generated by geometrical image transformation including the original input data; AUG_DCGAN_<N> consists of 5000 images which have been generated in different runs <N> using the DCGAN with given weights

different characteristics were recorded and then transformed geometrically. Nevertheless, it is important to note that this is a replicated process. For a reliable comparison of the synthetic defects with industrially arising process defects, an extensive empirical analysis in a real world production environment would be essential. However, this was not part of this study.

In order to provide a basic assessment of the diversity and realism of the synthetic defect images, the GAN-Train GAN-Test method was used successfully in this paper. With this GAN-Train GAN-Test method we have demonstrated that the DCGAN generates diverse but realistic training data, similar to the expectations from the literature review. According to its operating principle the Geometrical Transformation generates very realistic synthetic data, but the variety of these artificial images is very limited. This can easily lead to an overfitting of a classifier. Beside that, Cubuk et al. (2019) have already demonstrated for another applications, that a significant improvement in classification performance can be achieved with a reasonably configured Geometrical Transformation.

Considering the results mentioned above, the research question is answered. The Geometrical Transformation method and the DCGAN based technique can be used to generate synthetic image data of fiber placement defects from the AFP process, using less than 50 initial representative data samples. Only overlap defects shall be an exception in our studies. Since the data augmentation and training with overlap defects behave quite unexpectedly in our investigations, further tests should therefore be examined. For these investigations also the gap and no defect classes need to be taken into account. In further investigations, it seams to be reasonable to examine the robustness of the proposed methods for other sensor settings and different materials with deviating optical material properties. In this context, it might also be beneficial to investigate further promising methods from the literature summarised in the Table 2 for application in this adapted scenario.

In summary, the aim of this study was not to increase the classification rate of a CNN classifier but to compare

and evaluate different data synthesis methods for the AFP inspection application case. In order to apply these techniques to a real world scenario, representative defect image data from the corresponding inspection process need to be recorded and utilised for data synthesis. Therefore an important aspect is to ensure the diversity of the initial training images. Thus these images should consist of many potential defect characteristics. The challenge is to extract these striking defect images from a real process. In most cases, typical defects occur, which each contain very similar characteristics. Obviously, this can lead to the fact that in real applications considerably more defect images have to be recorded than the actually required minimum quantity of representative training images. Moreover, the enhancement of the very simple CNN classifier used here is certainly advisable to achieve a very high and robust classification result.

Conclusion

The investigations in this paper demonstrate that the *Geometrical Transformation* and a reasonably configured conditional *Deep Convolutional Generative Adversarial Network* are well suited for the synthetic data generation from less than 50 representative origin images per class. The GAN-Train GAN-Test method proves to be a suitable tool for the independent evaluation of artificially generated image data. However, this method has the inherent property that it links the diversity and realism of the images always to another defined comparison data set, which has to be taken into account appropriately.

The data synthesis techniques investigated in this paper offer major advantages for the fibre composite industry. Firstly, this paper provides the necessary information for the utilisation of suitable data synthesis techniques for inspection image data from fibre composite production. Thus the results of our studies emphasise the application of such methods. Secondly, the presented approaches enable the abstraction and synthetic generation of potentially confidential manufacturing data. Thus a sufficient amount of realistic data can be

provided easily to developers of inspection systems, without spreading potentially confidential manufacturing information. Furthermore, these artificially generated depth images can be applied for a simulative optimisation of an AFP inspection algorithms.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

ams AG: Datasheet ds000603 - cmv12000 - cmos image sensor. techreport 3.0, ams AG, Tobelbader Strasse 30, 8141 Premstaetten, Austria (2020). https://ams.com/documents/20143/36005/CMV12000_DS000603_3-00.pdf/d27f4643-e11b-86f9-4e09-ec055cb4c8e1. Datasheet DS000603 v3-00.

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks. In D. Precup, & Y. W. Teh (Eds.), *Proceedings of the 34th international conference on machine learning, proceedings of Machine Learning Research* (Vol. 70, pp. 214–223). PMLR, International Convention Centre, Sydney, Australia. http://proceedings.mlr.press/v70/arjovsky17a.html.

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 (2019). https://www.automationtechnology.de/cms/wp-content/uploads/2019/03/C5-Series_specifications_web.pdf. Rev 1.2.

Automation Technology GmbH: The FIR filter. techreport 1.0, Automation Technology GmbH, Hermann-Bössow-Straße 6-8, 23843 Bad Oldesloe, Germany (2014). https://www.stemmer-imaging.com/media/uploads/cameras/12/122195-Automation_Technology_AppNote_FIR_Filter.pdf. Rev. 1.0.

Black, S. (2018). Improving composites processing with automated inspection, part II. Compositesworld. https://www.compositesworld.com/articles/improving-composites-processing-with-automated-inspection-part-ii.

Black, S. (2018). Improving composites processing with automated inspection. compositesworld. https://www.compositesworld.com/articles/improving-composites-processing-with-automated-inspection.

Borji, A. (2019). Pros and cons of GAN evaluation measures. *Computer Vision and Image Understanding*, 179, 41–65. https://doi.org/10.1016/j.cviu.2018.10.009.

Bradski, G. (2000). The OpenCV Library. Dr. Dobb's Journal of Software Tools. https://opencv.org/.

Brownlee, J. (2019). Generative adversarial networks with python: deep learning generative models for image synthesis and image translation. Machine Learning Mastery. https://books.google.de/books?id=YBimDwAAQBAJ.

Campbell, F. (2004). *Manufacturing processes for advanced composites*. Amsterdam: Elsevier.

Cemenska, J., Rudberg, T., & Henscheid, M. (2015). Automated in-process inspection system for AFP machines. *SAE International Journal of Aerospace*, 8(2), 303–309. https://doi.org/10.4271/2015-01-2608.

Chen, M., Jiang, M., Liu, X., & Wu, B. (2018). Intelligent inspection system based on infrared vision for automated fiber placement. In *2018 IEEE international conference on mechatronics and automation (ICMA)*. IEEE. https://doi.org/10.1109/icma.2018.8484646.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1), 53–65. https://doi.org/10.1109/msp.2017.2765202.

Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2019). AutoAugment: Learning augmentation strategies from data. In *2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. IEEE. https://doi.org/10.1109/cvpr.2019.00020.

Eitzinger, C. (2019). Inline inspection helps accelerate production by up to 50%. Lightweight Design worldwide

European Machine Vision Association: Emva genicam standard. techreport 2.0, European Machine Vision Association (EMVA) (2009). https://www.emva.org/. Release 2.0.

Gardiner, G. (2018). Zero-defect manufacturing of composite parts. CompositesWorld. https://www.compositesworld.com/blog/post/zero-defect-manufacturing-of-composite-parts.

Goodfellow, I. J. (2017). Generative adversarial networks. In *NIPS 2016 Tutorial* (Vol. abs/1701.00160). arXiv:1701.00160.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 27, pp. 2672–2680). Red Hook: Curran Associates Inc.

Grohmann, Y., Stoffers, N., Kühn, A., & Mahrholz, T. (2016). Development of the direct roving placement technology (DRP). In *ECCM17—17th European conference on composite materials*. https://elib.dlr.de/107706/.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. (2017). Improved training of wasserstein gans. In *Proceedings of the 31st international conference on neural information processing systems, NIPS'17* (pp. 5769–5779). Curran Associates Inc., Red Hook, NY, USA. https://doi.org/10.5555/3295222.3295327.

Harik, R., Saidy, C., J. Williams, S., Gurdal, Z., & Grimsley, B. (2018). Automated fiber placement defect identity cards: Cause, anticipation, existence, significance, and progression. In *SAMPE 18*. https://www.researchgate.net/publication/326464139_Automated_fiber_placement_defect_identity_cards_cause_anticipation_existence_significance_and_progression.

Heinecke, F., & Willberg, C. (2019). Manufacturing-induced imperfections in composite parts manufactured via automated fiber placement. *Journal of Composites Science*, 3(2), 56. https://doi.org/10.3390/jcs3020056.

Huang, Y., Cheng, Y., Bapna, A., Firat, O., Chen, D., Chen, M., et al. (2019). Gpipe: Efficient training of giant neural networks using pipeline parallelism. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 32, pp. 103–112). Red Hook: Curran Associates Inc.

Jain, S., Seth, G., Paruthi, A., Soni, U., & Kumar, G. (2020). Synthetic data augmentation for surface defect detection and classification using deep learning. *Journal of Intelligent Manufacturing*,. https://doi.org/10.1007/s10845-020-01710-x.

Jorge, J., Vieco, J., Paredes, R., Sánchez, J. A., & Miguel Benedí, J. (2018). Empirical evaluation of variational autoencoders for data augmentation. In *Proceedings of the 13th international joint conference on computer vision, imaging and computer graphics theory and applications* (pp. 96–104). SCITEPRESS—Science and Technology Publications. https://doi.org/10.5220/0006618600960104.

Joshi, K. D., Chauhan, V., & Surgenor, B. (2018). A flexible machine vision system for small part inspection based on a hybrid SVM/ANN approach. *Journal of Intelligent Manufacturing*, *31*(1), 103–125. https://doi.org/10.1007/s10845-018-1438-3.

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive growing of gans for improved quality, stability, and variation. In *ICLR 2018* (Vol. abs/1710.10196). arXiv:1710.10196.

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. (2020). A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*,. https://doi.org/10.1007/s10462-020-09825-6.

Lengsfeld, H., Fabris, F. W., Krämer, J., Lacalle, J., & Altstädt, V. (2014). Faserverbundwerkstoffe. Hanser Fachbuchverlag. https://www.ebook.de/de/product/22746074/hauke_lengsfeld_felipe_wolff_fabris_johannes_kraemer_javier_lacalle_volker_altstaedt_faserverbundwerkstoffe.html.

Li, Y., Swersky, K., & Zemel, R. (2015). Generative moment matching networks. In *ICML'15* (pp. 1718–1727). JMLR

Luo, J., Huang, J., & Li, H. (2020). A case study of conditional deep convolutional generative adversarial networks in machine fault diagnosis. *Journal of Intelligent Manufacturing*,. https://doi.org/10.1007/s10845-020-01579-w.

Maass, D. (2012). Automated dry fiber placement for aerospace composites. In *Composites manufacturing 2012*. Danobat.

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., & Frey, B. (2016). Adversarial autoencoders. EPrint arXiv:1511.05644v2 (Unpublished)

Marsh, G. (2010). Airbus A350 XWB update. *Reinforced Plastics*, *54*(6), 20–24. https://doi.org/10.1016/s0034-3617(10)70212-5.

McIlhagger, A., Archer, E., & McIlhagger, R. (2020). Manufacturing processes for composite materials and components for aerospace applications. In P. Irving, & C. Soutis (Eds.), *Polymer composites in the aerospace industry* (pp. 59–81). Elsevier. https://doi.org/10.1016/b978-0-08-102679-3.00003-4.

Meister, S., Wermes, M.A.M., Stueve, J., & Groves, R.M. (2020). Algorithm assessment for layup defect segmentation from laser line scan sensor based image data. In D. Zonta, & H. Huang (Eds.), Sensors and smart structures technologies for civil, mechanical, and aerospace systems 2020. SPIE. https://doi.org/10.1117/12.2558434.

Miesen, N., Sinke, J., Groves, R. M., & Benedictus, R. (2015). Simulation and detection of flaws in pre-cured CFRP using laser displacement sensing. *The International Journal of Advanced Manufacturing Technology*, *82*(1–4), 341–349. https://doi.org/10.1007/s00170-015-7305-x.

Mikolajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. In *2018 International interdisciplinary PhD workshop (IIPhDW)*. IEEE. https://doi.org/10.1109/iiphdw.2018.8388338.

Neff, T. (2018). Data augmentation in deep learning using generative adversarial networks. Master's thesis, Graz University of Technology. https://www.tugraz.at/fileadmin/user_upload/Institute/ICG/Images/team_bischof/mib/paper_pdfs/StudentsMasterTheses/2018_03_DA_neff.pdf.

Odena, A., Olah, C., & Shlens, J. (2017). Conditional image synthesis with auxiliary classifier gans. In *Proceedings of the 34th international conference on machine learning, ICML'17* (Vol. 70, pp. 2642–2651). JMLR.org. https://doi.org/10.5555/3305890.3305954.

Oromiehie, E., Prusty, B. G., Compston, P., & Rajan, G. (2019). Automated fibre placement based composite structures: Review on the defects, impacts and inspections techniques. *Composite Structures*, *224*, 110987. https://doi.org/10.1016/j.compstruct.2019.110987.

Perarnau, G., van de Weijer, J., Raducanu, B., & Álvarez, J. M. (2016). Invertible conditional gans for image editing. In *NIPS 2016 workshop on adversarial training*.

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv:1712.04621

Potter, K. (2009). Understanding the origins of defects and variability in composites manufacture. In *ICCM international conferences on composite materials*. http://iccm-central.org/Proceedings/ICCM17proceedings/Themes/Plenaries/P1.5%20Potter.pdf.

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. ICLR. https://arxiv.org/abs/1511.06434.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Lecture notes in computer science* (pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28.

Rudberg, T. (2019). *Webinar: Building AFP system to yield extreme availability*. CompositesWorld. Video.

Rudberg, T., Nielson, J., Henscheid, M., & Cemenska, J. (2014). Improving AFP cell performance. *SAE International Journal of Aerospace*, *7*(2), 317–321. https://doi.org/10.4271/2014-01-2272.

Sacco, C., Radwan, A. B., Harik, R., & Tooren, M. V. (2018). Automated fiber placement defects: Automated inspection and characterization. In *SAMPE 18—Long Beach* (p. 13). McNAIR Center for Aerospace Innovation and Research, Department of Mechanical Engineering, College of Engineering and Computing, University of South Carolina 1000 Catawba St., Columbia, SC, 29201, USA. https://www.nasampe.org/store/ViewProduct.aspx?ID=11833782.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training gans. In *Proceedings of the 30th international conference on neural information processing systems, NIPS'16* (pp. 2234–2242). Curran Associates Inc., Red Hook, NY, USA. https://doi.org/10.5555/3157096.3157346.

Schmidt, C., Hocke, T., & Denkena, B. (2019). Artificial intelligence for non-destructive testing of CFRP prepreg materials. *Production Engineering*,. https://doi.org/10.1007/s11740-019-00913-3.

Schmitt, R., Niggemann, C., & Mersmann, C. (2008). Contour scanning of textile preforms using a light-section sensor for the automated manufacturing of fibre-reinforced plastics. In F. Berghmans, A. G. Mignani, A. Cutolo, P. P. Meyrueis, & T. P. Pearsall (Eds.), *Optical sensors 2008* (Vol. 7003, pp. 436–447). SPIE. https://doi.org/10.1117/12.779005.

Schmitt, R., Orth, A., & Niggemann, C. (2007). A method for edge detection of textile preforms using a light-section sensor for the automated manufacturing of fibre-reinforced plastics. In W. Osten, C. Gorecki, & E. L. Novak (Eds.), *Optical measurement systems for industrial inspection V*. SPIE. https://doi.org/10.1117/12.726177

Shmelkov, K., Schmid, C., & Alahari, K. (2018). How good is my gan? In *The European conference on computer vision (ECCV)*.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. *Journal of Big Data*,. https://doi.org/10.1186/s40537-019-0197-0.

Tabernik, D., Šela, S., Skvarč, J., & Skočaj, D. (2019). Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, *31*(3), 759–776. https://doi.org/10.1007/s10845-019-01476-x.

Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In K. Chaudhuri, & R. Salakhutdinov (Eds.), *Proceedings of the 36th international conference on machine learning, proceedings of machine learning research* (Vol. 97, pp. 6105–6114). PMLR, Long Beach, California, USA. http://proceedings.mlr.press/v97/tan19a.html.

Tzutalin. (2015). LabelImg. https://github.com/tzutalin/labelImg.

Ucan, H., Scheller, S., Nguyen, D.C., Nieberl, D., Beumler, T., Haschenburger, A., Meister, S., Kappel, E., Prussak, R., Deden, D., Mayer, M., Pantelelis, N., Zapp, P., Hauschild, B., & Menke, N. (2019). Automated, quality assured and high volume oriented production of fiber metal laminates (FML) for the next generation of passenger aircraft fuselage shells. In *The fourth international symposium on automated composites manufacturing*. https://elib.dlr.de/127353/.

Vasilev, I., Slater, D., & Spacagna, G. (2019). *Python deep learning* (2nd ed.). Birmingham: Packt Publishing.

Wan, L., Zeiler, M., Zhang, S., LeCun, Y., & Fergus, R. (2013). Regularization of neural networks using dropconnect. In *Proceedings of the 30th international conference on international conference on machine learning, ICML'13* (Vol. 28, pp. III-1058–III-1066). JMLR.org. http://yann.lecun.com/exdb/publis/pdf/wan-icml-13.pdf.

Weimer, C., Friedberger, A., Helwig, A., Heckner, S., Buchmann, C., & Engel, F. (2016). 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*. https://www.researchgate.net/profile/Christian_Weimer/publication/308778487_INCREASING_THE_PRODUCTIVITY_OF_CFRP_PRODUCTION_PROCESSES_BY_ROBUSTNESS_AND_RELIABILITY_ENHANCEMENT/links/57efa78208ae886b8975147a.pdf.

Wu, K., Qiang, Y., Song, K., Ren, X., Yang, W., Zhang, W., et al. (2019). Image synthesis in contrast MRI based on super resolution reconstruction with multi-refinement cycle-consistent generative adversarial networks. *Journal of Intelligent Manufacturing*, *31*(5), 1215–1228. https://doi.org/10.1007/s10845-019-01507-7.

Zambal, S., Heindl, C., & Eitzinger, C. (2019a). Machine learning for CFRP quality control. In *SAMPE 19*. Zenodo. https://doi.org/10.5281/ZENODO.3381930.

Zambal, S., Heindl, C., Eitzinger, C., & Scharinger, J. (2019b). End-to-end defect detection in automated fiber placement based on artificially generated data. In C. Cudel, S. Bazeille, & N. Verrier (Eds.) *Fourteenth international conference on quality control by artificial vision*. SPIE. https://doi.org/10.1117/12.2521739.

Zhang, X., Fu, Y., Jiang, S., Sigal, L., & Agam, G. (2015). Learning from synthetic data using a stacked multichannel autoencoder. In *IEEE 14th international conference on machine learning and applications* (Vol. abs/1509.05463). IEEE. https://doi.org/10.1109/icmla.2015.199. arXiv:1509.05463.