Defect detectability analysis via probability of defect detection between traditional and deep learning methods in numerical simulations

Miroslav Yosifov[1,2], Patrick Weinberger[1], Michael Reiter[3], Fröhler Bernhard[1], Jan De Beenhouwer[2], Jan Sijbers[2], Johann Kastner[1], Christoph Heinzl[4,5]

[1]University of Applied Sciences Upper Austria, Campus Wels, Stelzhamerstrasse 23, 4600 Wels, Austria
e-mail: miroslav.yosifov@fh-wels.at

[2]imec-Vision Lab, Dept. of Physics, University of Antwerp, Universiteitsplein 1, 2610 Antwerpen, Belgium

[3]Nemak Linz GmbH, Linz, Austria

[4]University of Passau, Innstraße 43, Passau, Germany

[5]Fraunhofer Institute for Integrated Circuits IIS, Division Development Center X-ray Technology,
Flugplatzstraße 75, Fürth, Germany

Abstract

X-ray computed tomography (XCT) is one of the most powerful imaging techniques in non-destructive testing (NDT) for detecting, analysing and visualising defects such as pores, fibres, cracks etc. in industrial specimens. Detecting defects in X-ray images, however, is still a challenging problem, as it strongly depends on the quality of the XCT images. Numerical XCT simulation proved to be valuable in order to increase both image quality and detection performance. In this work, we thus analyse the differences between traditional segmentation techniques (i.e., k-means, watershed, Otsu thresholding) and deep learning-based methods (i.e., U-Net, V-Net, modified 3D U-Net) in terms of their defect detection capacity using virtual XCT images. For this purpose, we apply the probability of defect detection (POD) approach on simulated X-ray computed tomography data from aluminium cylinder heads. The XCT simulation tool SimCT was used to generate X-ray radiographs and respective reconstructions from a specimen series which features different well-defined defects with varying sizes, shapes and locations. To generate POD curves and to specify detection limits, the segmentation algorithms are used in predefined regions for defect detection via a hit/miss approach. A comparison and visualisation of six different types of defects is illustrated in 2D and 3D images, together with their POD curves and detection limits.

Keywords: probability of detection, X-ray computed tomography, numerical simulation, deep learning, U-Net, V-Net

1 Introduction

Probability of detection (POD) [1–3] is increasingly used in the field of X-ray based digital radiography and X-ray computed tomography (XCT). Especially, the strong trend in the industry towards fast in-line and at-line inspections demands for POD analyses in terms of process qualification. Typically, in-line and at-line inspections reach very short cycle times to perform the data acquisition as well as the decision process between "OK" and "NOT-OK" parts. Less projections and consequently shorter scanning times cause more noise in XCT images and therefore strongly demand the use of POD methods to optimise scanning parameters and characterise the reliability of the inspection processes (e.g., detection of pores, inclusions, cracks). XCT imaging and visual analysis play an important role to segment or extract features in 2D and 3D, especially comparing the simulations and real scans [4].

Up until recently, mostly traditional segmentation methods such as k-means, watershed and Otsu thresholding have been used for this purpose, which are still highly popular in material science applications [5]. To improve segmentation accuracy upon the traditional methods, several deep learning algorithms such as U-Net [6], V-Net [7] and 3D-U-Net [8] were introduced in the biomedical field. An updated version of 3D U-Net (M-3DUnet) aimed at improving the detection in material science was presented in a recent work [9], extending open_iA [10] to use pre-trained networks of deep learning methods.

In our previous work, we focused on applying the Zeiss automated defect detection (ZADD) [11] for global defect detection for the generation of respective POD curves [12]. In this paper, we investigate various traditional and deep learning methods for the segmentation of pores and defects locally (using sub-volumes for segmentation). For this purpose, we trained various deep learning techniques, splitting datasets into sub-volumes, followed by the application of the above mentioned neural network models in terms of global and local detection. We compare the used methods and also show how to improve detection limits by optimising hyper-parameters in V-Net, U-Net and M-3DUnet deep learning methods.

In contrast to previous work, we focus here on the POD comparison for different kinds of defect detection algorithms applied on artificial XCT data of industrial specimens as generated by numerical XCT simulations. Numerical simulations allow the definition and generation of specific pores or defects in specific locations for artificial XCT data. This allows us to define different scenarios in the experiments and to create a simulation series with several options. Further related work was presented by Rao et al. on the creation of artificial pores in real specimens [13]. In their work, two artificial defects were created by drilling

holes with diameters 0.2 mm and 0.3 mm in non-porous samples then investigated porosity in two artificial defects and compared on different segmentation algorithms. However, this process is time-consuming and high uncertainty may occur in the drilling process. Therefore, simulating artificial XCT data is a more efficient process and it allows for avoiding the high uncertainty level as well as time-consuming.

2   Related Work

In order to put our work into context, we focus on POD using XCT simulation. We also review convolutional neural networks used for analysing XCT data.

2.1   Probability of Detection and Reliability

Nowadays, reliability is one of the most important methods in NDT for quality and safety certifications, especially in the aeronautics [3]. Determining which size of defect or pore can still be detected within the specimen is highly important for quality control. For this reason, POD [1–3] was introduced, and it is now increasingly applied in various NDT scenarios for evaluating the capability of an inspection system or process in terms of flaw detection. The POD method is frequently used in the aeronautics and the automotive industry, as well as several other application areas to calculate quantitative results regarding detection limits for an inspection. Especially in the aerospace industry, simulating XCT and ultrasonic images with POD determination may help to reduce high non quality costs [14, 15].

POD approach was applied on several research areas such as eddy current testing [16] or ultrasound inspection [17] for artificial cracks. POD in essence is a statistical measure to determine and introduce detection limits on task specific data. In XCT imaging, the characteristics and quality of XCT images play a critical role in the creation of effective POD curves to determine the detectability of defects with regards to size, shape or location. In our case, size will be our measurement aspect to create respective POD curves, i.e., we use the diameter for spheres, the height for cylinders and the volume for irregular shapes. To figure out the minimum detected defect size is important. However, to find out the maximum size of potentially undetected defects is more critical for the industry. Therefore, we will also focus on the maximum size of undetected defects in our analysis. Reliability determination and POD are necessary applications for uncertainty determination via repeating the experimental process [18]. However, in this work, the main focus will be only on determining POD curves and detectability limits without uncertainty analysis.

2.2   XCT Simulation and Modelling in NDT

Simulation, modelling and data processing is necessary to improve, understand and analyse XCT systems in NDT, most importantly to generate better projection images and optimise parameters for industrial samples. Therefore, many kinds of XCT simulation tools were developed over the last years. Specifically, XCT simulation tools can be classified into two types: numerical simulation which uses ray-tracing algorithms and Monte Carlo simulations as, e.g., used for biomedical and materials science [19]. Examples for numerical simulation tools are Scorpius XLab [20], arTist [21] and the ASTRA toolbox [22], which were implemented to simulate radiographic images by ray-tracing algorithms. In our experiments, SimCT [23] was used to generate radiographic images which also integrates CERA [24] for the reconstruction of XCT data.

Monte Carlo simulation is another powerful method for the simulation of X-rays and thus the generation of respective X-ray images. The GATE [25] simulation tool is used to assist in new medical and XCT devices, to develop reconstruction algorithms, for modelling positron emission tomography [26] and single photon emission computer tomography [27] as well as for the optimisation of imaging parameters in the field of nuclear medicine [28]. The performance of Monte Carlo simulation, more specifically, the number of the simulated X-rays, plays an important role in order to improve image quality. However, this process is computationally expensive and requires powerful devices.

2.3   Convolutional Neural Networks

In recent years, deep learning has shown impressive results in several domains. In the field of image processing mostly convolutional neuronal networks (CNN) are used [8]. Especially U-Net and V-Net have shown impressive result in the biomedical field [6, 7]. U-Net was developed at the university of Freiburg by Ronneberger et al. [6]. V-Net share a lot of similarities with U-Net (skip connections and U shape), the difference between them is that V-Net uses instead of max-pooling layer, also convolutional layers and residual blocks [7]. Also in the field of NDT several publication have shown promising results using CNNs for the inspection of XCT datasets [9, 29, 30]. Firstly, the modification of the 3D U-Net architecture to segment both pores and fibres in carbon and glass fibre-reinforced material was presented in 2020 by Yosifov [9]. Secondly, a modification of the 3D U-Net architecture was applied to segment and distinguish fibre bundles by Weinberger et al. [29]. Another publication applied U-Net to the analysis of simulated XCT scans of aluminium parts [30].

Table 1: XCT Scanning and device parameters.

Parameters	Unit	Value	Parameters	Unit	Value
Acceleration voltage	kV	200	Source-object-distance	mm	1186.36
Anode current	μA	7200	Source-detector-distance	mm	1518.35
Integration time	ms	33	Magnification	–	1.2766
Detector pixels	–	2048x2048	Voxel size	μm	313.32
Binning	–	2x2	Focal spot size	mm	1
Number of projections	–	400	Total cycle time	min	1

Table 2: Artificial defect properties.

	Shape	Material type	Class	Minimum range a	Maximum range b
Defect type DT1	cylinder	Al	broken water jacket	31 μm	2310 μm
Defect type DT2	sphere	air	porosity	31 μm	1320 μm
Defect type DT3	sphere	air	damage	31 μm	3140 μm
Defect type DT4	sphere	SiO_2	residual sand	31 μm	4960 μm
Defect type DT5	irregular/extracted	air	porosity	0.34 mm^3	84.37 mm^3
Defect type DT6	irregular/extracted	air	porosity	0.22 mm^3	55.94 mm^3

3 Methods

In the following sections, we describe our methodology in terms of the used defect detection and segmentation methods. We further discuss the generation of virtual XCT data as well as the used simulation setup. Finally, we introduce the applied POD concept.

3.1 Defect Detection and Segmentation

2D as well as 3D segmentation techniques are required for the extraction and detection of features such as pores, cracks or other defects. We can categorise typical segmentation techniques into four classes: region-, pixel-, edge- and model-based techniques [31]. In this work, we focus on k-means [32] as a representative of model-based segmentation techniques, on watershed[33] segmentation as a region-based method and on Otsu thresholding [5] as a pixel-based method in the area of traditional segmentation. In this experiment, edge-based techniques are not in our focus, because of their limitations in terms of the sensitivity to noise and the required manual intervention [31]. Instead of edge detection, manual segmentation methods were considered to generate proper segmentation masks. Each method features disadvantages and advantages depending on the data type. Threshold methods are simple without the need to apply pre-processing steps, yet highly depending on peaks in the histograms of respective data to be segmented. Watershed segmentation is based on a topological interpretation of the data and typically requires a pre-processing step (e.g., the computation of gradient magnitudes as input to the watershed filter). A significant disadvantage of watershed segmentation is over-segmentation of regions (i.e., the separation into more different regions than necessary), which requires an additional post-processing step in most cases. To avoid over-segmentation, more advanced versions of watershed segmentation were presented, e.g., the marker-controlled watershed segmentation by Wang et al. [34]. This approach introduces additional pre-processing steps to transform and clean the input images, before they are passed on to the watershed segmentation. More recently, also deep learning based techniques have been successfully applied for image and volume segmentation, e.g., in biomedical, material science and computer vision applications via U-Net, V-Net and related approaches, both in 2D and 3D. The U-Net architecture is one of the convolutional neural networks that are state of art in medical image segmentation. The V-Net [7] architecture is very similar to U-Net on which it builds. As mentioned in subsection 2.3, the main difference is the usage of the residual blocks. Recently a modified version of U-Net working in 3D (M-3DUnet) was introduced by Yosifov [9], employing overlapping tiles to improve the performance of the network. This modified 3D U-Net (M-3DUnet) version features additional convolution and ReLU layers in each decoder part coupled with a sigmoid function. In this paper, we consider three traditional methods and three different deep learning architectures for the POD comparisons.

3.2 Generation of Virtual XCT Data using Simulation

For our experiment, virtual XCT datasets were generated via the simulation tool SimCT. To generate realistic data using XCT simulation, optimised parameters were applied, which are listed in Table 1. The detailed XCT data generation process and analysis setup were presented in our previous work [12].

For the generation of simulated XCT data, four artificial defect types (DT1, 2, 3 and 4) were defined with different size, shape, location, density and material type. We created 100 different sizes (in terms of radius, height and volume) and automatically integrated these defects at varying locations via a uniform distribution function. DT5 and 6 were extracted from a real specimen and 100 STL datasets were created from these defects with varying size. The defect characteristics are presented in Table 2.
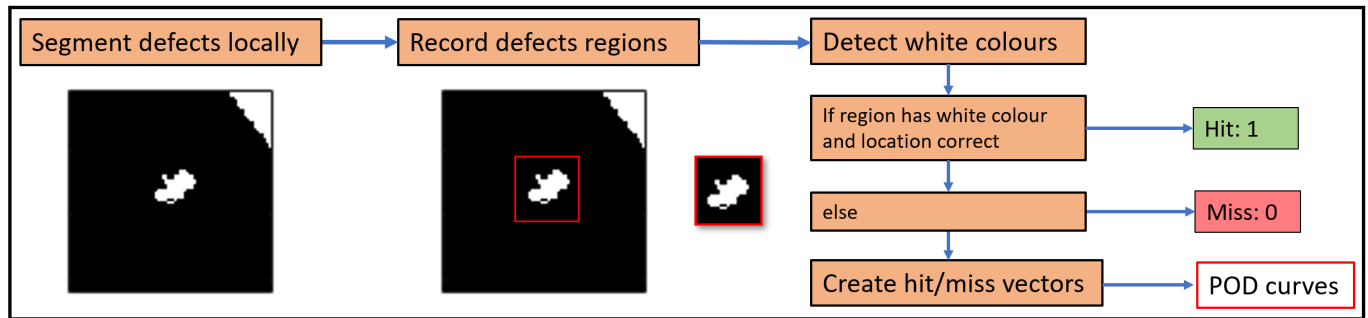
Figure 1: Hit/miss approach

The minimum and maximum range of each defect is considered respectively as height for cylinders, diameter for spheres and as volume for irregular shapes.

The specimen including the defects as explained were simulated via SimCT considering the following physical effects: focal spot blur, a modulation transfer function, image noise and scattering [23]. A more detailed discussion on the used parameters is published in our previous work [12].

3.3 Probability of Detection Concept and Hit/Miss Model

POD is a statistical measure to determine the performance of an inspection process regarding its detection of features of interest. Specifically in our case, it measures the detection capacity of pores and other defects. a_{90} is described as the size for which $POD(a_{90}) = 0.90$, and $a_{90/95}$ (detection limit) is the lower 95% confidence level of a_{90}. In this work, XCT based inspection scenarios were evaluated in terms of the detection properties of artificial defects via several detection methods based on artificial XCT data. For this purpose, we generated virtual XCT datasets using the SimCT simulation tool [23], which are including different types of artificial defects. We then computed their POD curves, the detection limits, as well as a POD curve comparison. The hit/miss approach[1] generates a binary signal of detection which returns "defect present" or "defect free". In our case, the binary response will be generated by traditional segmentation methods as well as deep learning methods in the region of the object of interest (see Figure 1). In some cases, the segmentation algorithm segments the defect as two separate defects. In this case, if the spatial tiling of the volume "cuts through" a pore, effectively making it two "open pores", one extra step is applied to detection to create the hit/miss vector. The results of the segmentation techniques were also analysed visually: If the defect shape is still stable, we count it as detected. Otherwise if there is reasonable change in the main defect shape, we count it as miss.

Generalised linear models (GLM) [35] are continuous link functions with model parameters providing a matrix of probabilities for the binary response of success/failure scenarios. The Probit linking function [12] is used for binary regression in order to create POD curves in this experiment.

3.4 Data Pre-processing and Training

We cut out the 100 defect versions with different sizes from the simulated specimens. The input sub-volume size was $64 \times 64 \times 64$ voxels. To train networks efficiently, the input images were normalised in the range between 0 and 1. To create a ground truth for the training, we applied manual segmentation for defects 5 and 6. The main reason for using manual segmentation was found in the traditional segmentation methods failing in terms of the grey value separation. For all other defects, Otsu threshold based segmentation was used to create ground truth.

Regarding deep learning techniques, we implemented U-Net, V-Net and M-3DUnet both in Matlab and Python. The MONAI[1] framework is applied to implement our U-Net, V-Net and M-3DUnet architectures for the training process using Python. We used a U-Net based architecture which is part of the framework MONAI. MONAI is a framework for deep learning mainly in the medical field; it offers different models, pre-processing and training libraries for images and volume data.

For the training of the networks an Nvidia RTX6000 GPU was used. We used a batch size of 8 and a learning rate of 5e-4, with the Adam optimiser, and 50 epochs. As metric for the segmentation accuracy, the validation accuracy and for the loss function to follow training process for over-fitting or under-fitting, the Sørensen–Dice coefficient [36] was used.

For the training process, 80 out of 100 sub-volumes were used for defect types DT1, 2, 3 and 4. The remaining 20 sub-volumes were used for validation. For DT5 and 6, we used 60 sub-volumes for the training and 20 sub-volumes for validation. We were unable to properly segment the small defects for defect types 5 and 6 regarding the grey values separation due to a lack of contrast. Therefore, only 80 samples were generated as a ground truth. In the testing part, 100 sub-volume versions of each defect were tested.

---

[1]MONAI is a freely available, community-supported, PyTorch-based framework for deep learning in healthcare imaging. ; MONAI.io

4 Results and Discussion

Here we present a visualisation of defects and segmentation results in 2D and 3D views together with their POD curves, for the comparison of different segmentation algorithms. Additionally, the statistics in terms of detection as well as a comparison of the smallest detected defect, largest undetected defect, as well as a_{90} and $a_{90/95}$ are investigated for each defect type.
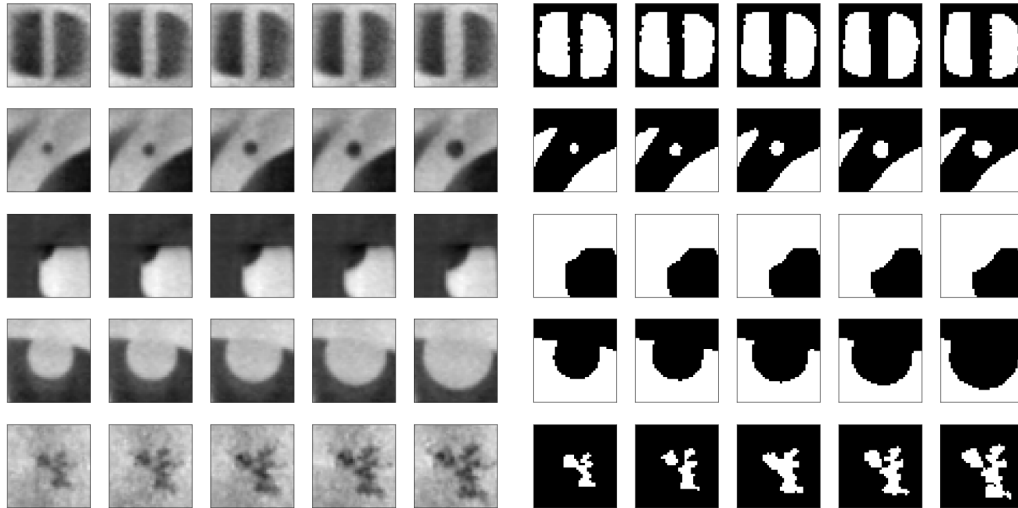

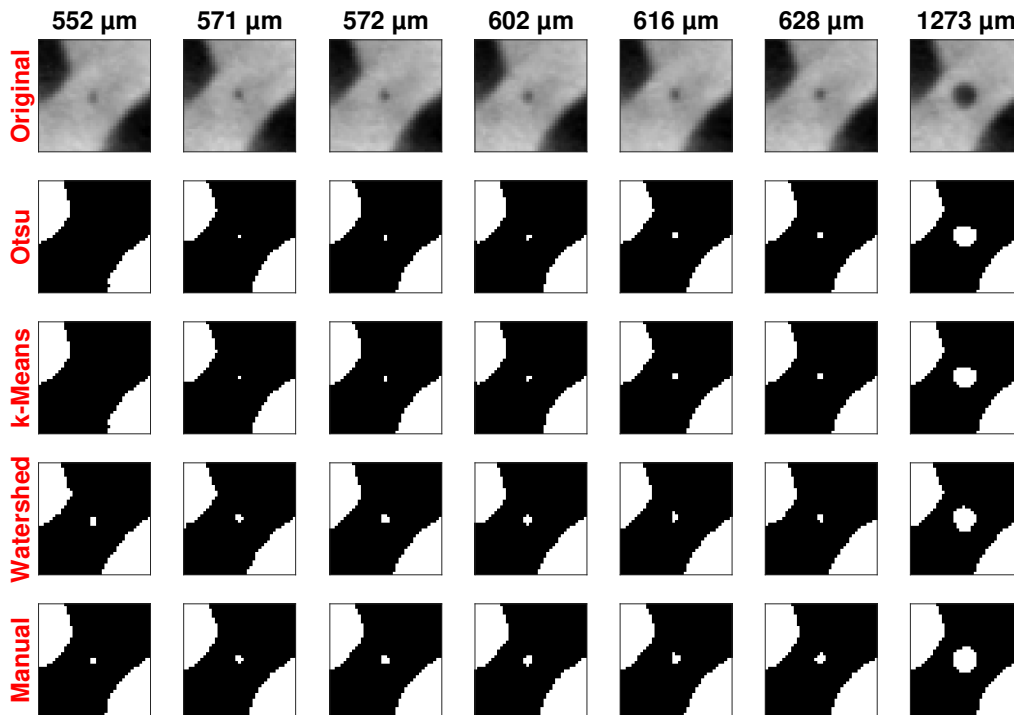Figure 2: Defect presentation and example segmentation results in 2D.

Figure 3: Original slices and segmentation results of defect type DT2.

4.1 Visualisation and Comparison of Defect Segmentation Results in 2D

In this section, five different defect types with their respective different versions in terms of changing radius size for spheres, height for cylinders and volume for irregular shapes are presented in Figure 2. Segmentation results for DT1 to 4 were generated using Otsu thresholding, while manual segmentation was used for DT5. For the comparison of the segmentation results of DT2, which are investigated using traditional methods, please see Figure 3.

We tried to segment defect types DT5 and 6 also with traditional methods. However, all of the methods failed to return proper segmentation results. Therefore, we have applied a data transformation process (cleaning data with opening-closing by recon-
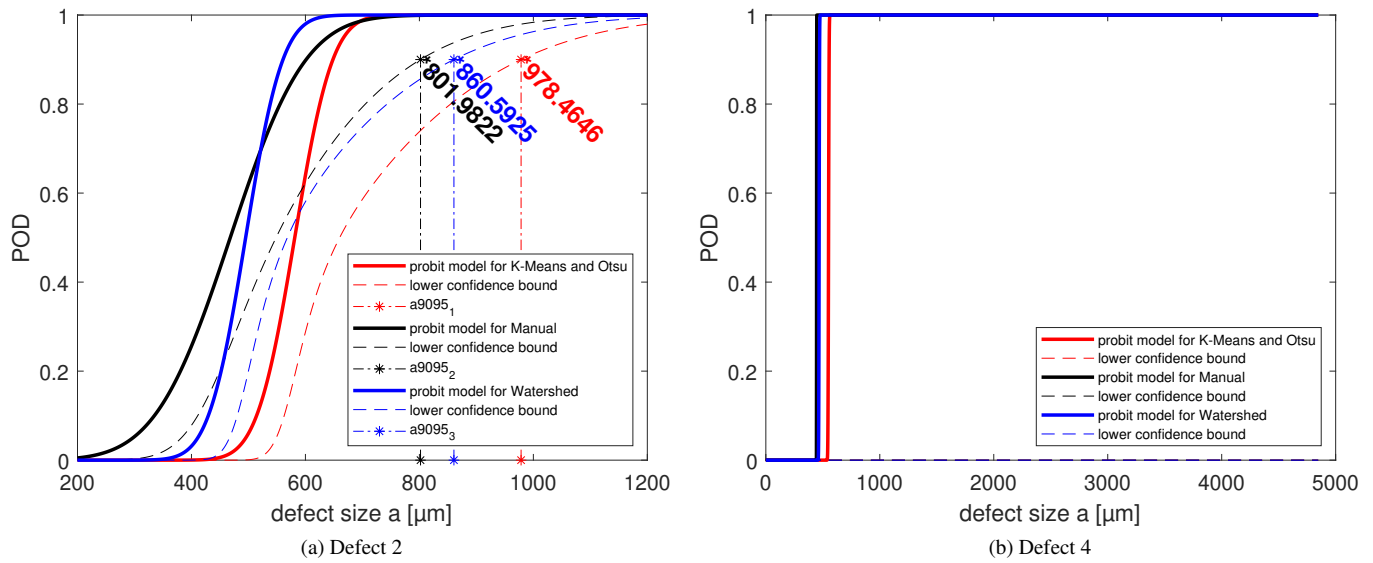
(a) Defect 2

(b) Defect 4

Figure 4: Comparison of POD curves for defect types DT2 and 4 in the traditional methods

struction [34]) followed by a manual segmentation to generate a reasonable ground truth for the training. We should also mention here that the segmented version of the defects sizes may differ with small uncertainties regarding the location changes and selected slice in 2D experiments. We only check here if a defect is detected or not. For our future work, we also plan to carry out an uncertainty analysis on the generated segmentation results.

4.2 POD Comparisons and Defect Statistics in 2D

In our investigation, for defect type DT2, 58 defects were detected out of 100 by Otsu thresholding and K-means segmentation, and slightly more (63) for watershed and finally 67 were detected by manual segmentation. The improvement for watershed results from its tendency to over-segment. For detectability this is an advantage, even though it also leads to the segmentation of extra voxels when compared to manual segmentation, as visible in the left upper corner in slices with size 552 and 571 μm in Figure 3. Minimum detected defect size was 184 μm with manual segmentation.

For defect type DT4, the algorithms vary only slightly in terms of detection numbers. For all of the methods, the detection numbers are over 85 and there is a perfect separation between detected and undetected defects (see Figure 4). The POD result of the investigation in terms of the hit/miss experiment consisting of 100 virtual XCT scans performed for one specimen with several defect types with varying defect size using SimCT. POD curves for defect types 2 and 4 demonstrated with their detection limits (confidence levels at 95 percent) with traditional segmentation methods. The detection limits ($a_{90/95}$) were calculated at 802 μm for manual segmentation, 860 μm for watershed segmentation, and 978 μm for Otsu thresholding (see Figure 4 (a)). There was also a perfect separation observable for defect types DT1 and 3, and detection results are presented in Table 3. Results are similar for each type of method for defect type DT3.

4.3 Visualisation and Comparison of Defect Segmentation Results in 3D

For each type of defect, a 3D view of six different defects in different sizes, which have been selected from the total of 100 XCT simulations, are presented in Figure 5.

Table 3: Maximum undetected (max failure) and minimum detected defect sizes (min success).

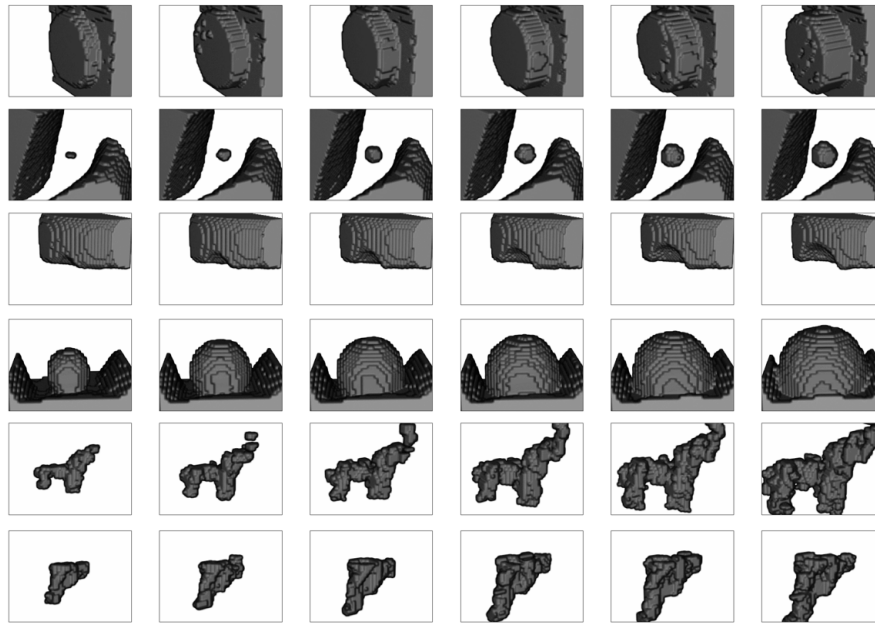|	Defect type DT1	Defect type DT2	Defect type DT3	Defect type DT4
	max failure /min success	max failure /min success	max failure /min success	max failure /min success
Otsu Threshold	449 /552 (μm)	730 /546 (μm)	681 /849 (μm)	523 /576 (μm)
K-means	449 /552 (μm)	730 /546 (μm)	681 /849 (μm)	523 /576 (μm)
Watershed	352 /393 (μm)	572 /352 (μm)	681 /849 (μm)	448 /486 (μm)
Manual	171 /184 (μm)	552 /310 (μm)	681 /849 (μm)	442 /448 (μm)

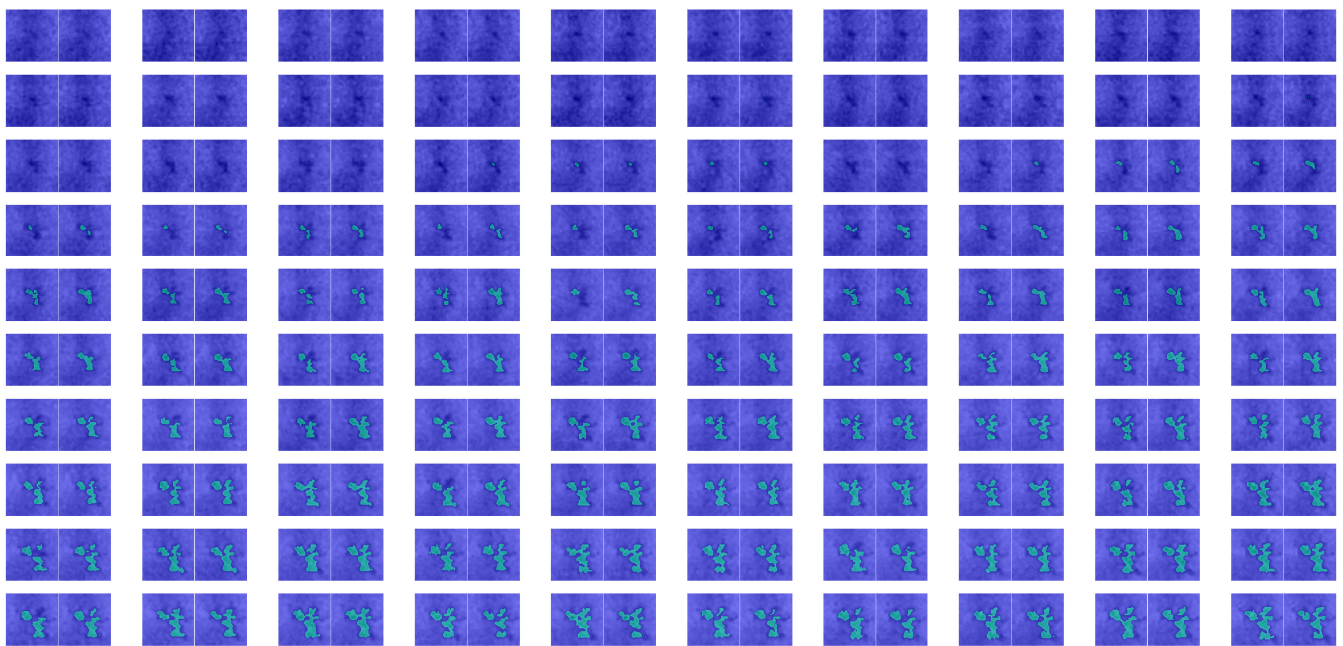Figure 5: Defect renderings in 3D sorted in six size versions.

Figure 6: Slice images for the segmentation results of ground truth (left sub-image) and M-3DUnet estimation (right sub-image) for defect type DT5. All images are sorted according to defect volume (mm^3). The white lines indicate the separation between ground truth and M-3DUnet estimation.

Predictions from M-3DUnet and ground truth for the defect type DT5 (irregular shape) with its variation regarding volume size in a sorted range are presented in Figure 6.

4.4 POD Comparisons and Defect Statistics in 3D

For defect type DT2, the detection limits for sphere diameter ($a_{90/95}$) were calculated respectively with 821 μm, 812 μm, and 772 μm for V-Net, U-Net and M-3DUnet (see Figure 7 (a)). $a_{90/95}$ for volume was respectively determined as 0.9, 0.8 and 0.78 mm^3. For defect type DT5, the detection limits for volume($a_{90/95}$) were calculated respectively with 7.2 mm^3, 10.83 mm^3, and 15.98 mm^3 for V-Net, U-Net and M-3DUnet (see Figure 7 (B)). For other types of defects, we were not able to create POD curves
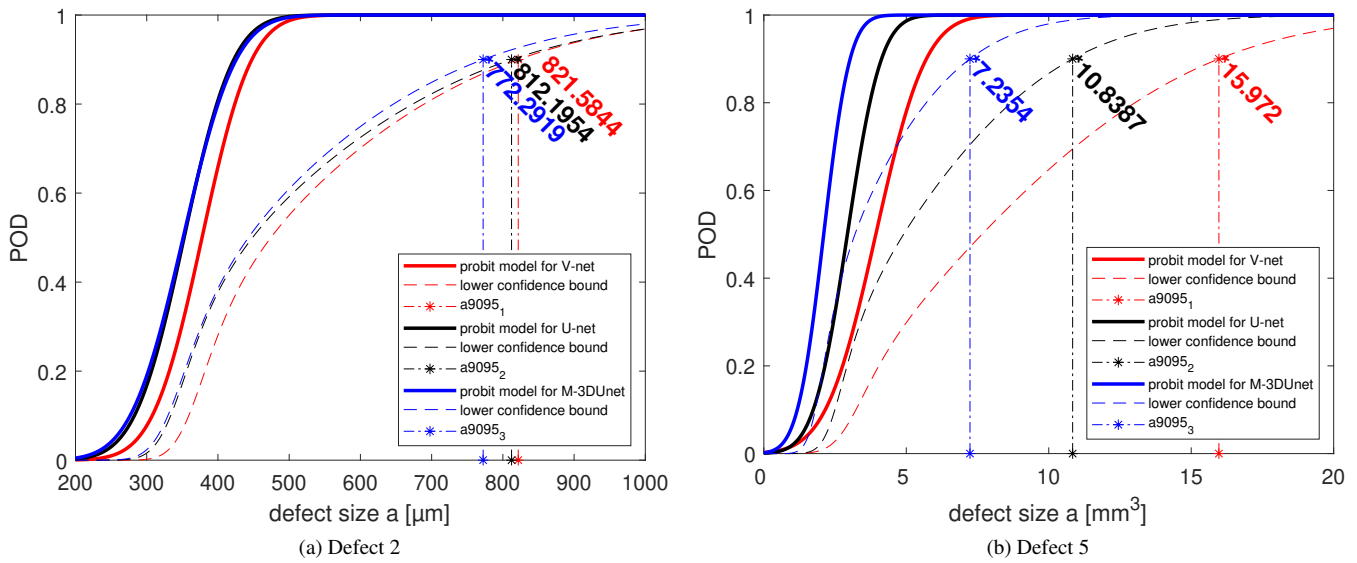
(a) Defect 2

(b) Defect 5

Figure 7: Comparison of POD curves for defect types DT2 and DT5.

Table 4: Maximum undetected (max failure) and minimum detected defect sizes (min success).

	Defect type DT1	Defect type DT2	Defect type DT4	Defect type DT5
	max failure /min success	max failure /min success	max failure /min success	max failure /min success
GT	352 /393 (μm)	459 /456 (μm)	486 /523 (μm)	4.45 /4.38 (mm^3)
V-Net	352 /393 (μm)	389 /310 (μm)	442 /446 (μm)	3.43 /2.42 (mm^3)
U-Net	352 /393 (μm)	374 /310 (μm)	442 /446 (μm)	3.43 /2.20 (mm^3)
M-3DUnet	352 /393 (μm)	374 /287 (μm)	442 /446 (μm)	2.62 /1.55 (mm^3)

regarding the perfect separation between detected and undetected defects.

All of the defect versions were segmented for defect type DT6, with U-Net, V-Net and M-3DUnet. Segmentation results can be discussed on the small type of defects via shape analysis for DT6. U-Net and M-3DUnet segmentation results were more accurate than the V-Net segmentation results regarding the segmented shape of defects. In V-Net segmentation, there were reasonable changes in the defect shape, especially for small pores. For defect type DT1, DT2, DT4 and DT5, the maximum size for undetected defects, and the minimum size for detected defects are presented in Table 4 as well as detection numbers for each type of defects (see Table 5). Defect types DT1, DT3, DT4 and DT6 had similar results for each type of method.

5 Conclusion

In this work, 2D and 3D defect detection and segmentation results using traditional and deep learning methods were presented for six different type of defects. Furthermore, we have determined POD curves and respective statistics (minimum detected defect, maximum undetected defect sizes and detection numbers) for each type of defect to investigate its detectability. In the 2D experiment, we focused on traditional segmentation methods and found that the detectability using watershed segmentation is higher than the other tested traditional methods. However, watershed segmentation tends to over-segment the defects present in our samples. K-means and Otsu thresholding segmentation are behaving similarly in terms of the detection in most of cases with slight differences regarding the shape of the segmented defect. Another finding was that we were capable of increasing the detection capability via manual segmentation. However, when the defect size is small (approximately 400 μm), a proper segmentation becomes very challenging on the noisy data. Still, the detectability can be increased if a manual segmentation is applied on each of the defects in series. However, this is a time consuming process. In the 3D investigations, two POD curves were successfully generated and detection limits were calculated for 3D segmentation results via the CNN methods. One of our most interesting findings was that we were able to increase detection numbers at least a little for all defect types, and even considerably for the irregular shapes via CNN methods.

For future work, we are planning to create a better ground truth by converting the STL models of defects to volumes. Moreover, we will focus on the task-specific uncertainties which are related to physical effects in the following steps: uncertainty in the radiographs (projections), uncertainty in the reconstructed images, training uncertainty in deep learning as well as segmentation uncertainty. While all of the investigations progressed on local detection in this study, it is also planned to apply global detection.

Table 5: Detected defects number from 100 sample for each defined defect and minimum detected defect.

	Defect 1	Defect 2	Defect 3	Defect 4	Defect 5	Defect 6
GT	81	64	89	88	74	82
V-Net	81	72	91	91	77	100
U-Net	81	74	91	91	81	100
M-3DUnet	81	74	91	91	86	100

References

[1] A. P. Berens, NDE reliability data analysis, in: ASM Handbook, Volume 17 - Nondestructive Evaluation and Quality Control, Vol. 17, AMS International, 1989, pp. 689–701.

[2] S. Amrhein, M. Rauer, M. Kaloudis, Characterization of computer tomography scanners using the probability of detection method, Journal of Nondestructive Evaluation 33 (2014) 643–650. `doi:10.1007/s10921-014-0258-4`.

[3] U.S. Department Of Defense, Handbook Non-destructive evaluation system reliability assessment, Vol. MIL-HDBK-, U.S. Department Of Defense, 1999.

[4] C. Heinzl, S. Stappen, STAR: Visual computing in materials science, Computer Graphics Forum 36 (3) (2017) 647–666. `doi:10.1111/cgf.13214`.

[5] N. Otsu, A threshold selection method from gray-level histograms, IEEE Transactions on Systems, Man, and Cybernetics 9 (1) (1979) 62–66. `doi:10.1109/TSMC.1979.4310076`.

[6] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, Medical Image Computing and Computer-Assisted Intervention – MICCAI. Lecture Notes in Computer Science 1 (2015) 234–241. `doi:10.1007/978-3-319-24574-4_28`.

[7] F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in: International Conference on 3D Vision (3DV), 2016, pp. 565–571. `doi:10.1109/3DV.2016.79`.

[8] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: Learning dense volumetric segmentation from sparse annotation, Medical Image Computing and Computer-Assisted Intervention – MICCAI. Lecture Notes in Computer Science 9901 (2016) 424–432. `doi:10.1007/978-3-319-46723-8_49`.

[9] M. Yosifov, Extraction and quantification of features in XCT datasets of fibre reinforced polymers using machine learning techniques, Master's thesis, Umeå University, Department of Computing Science (2020). URL `http://urn.kb.se/resolve?urn=urn:nbn:se:umu:diva-174684`

[10] B. Fröhler, J. Weissenböck, M. Schiwarth, J. Kastner, C. Heinzl, open_iA: A tool for processing and visual analysis of industrial computed tomography datasets, The Journal of Open Source Software 4 (2019) 1185. `doi:10.21105/joss.01185`.

[11] M. Reiter, C. Gusenbauer, R. Huemer, J. Kastner, At-line X-ray computed tomography of serial parts optimized by numerical simulations, in: Digital Industrial Radiology and Computed Tomography(DIR 2019), 2019.

[12] M. Yosifov, M. Reiter, S. Heupl, C. Gusenbauer, B. Fröhler, R. Fernández-Gutiérrez, J. D. Beenhouwer, J. Sijbers, J. Kastner, C. Heinzl, Probability of detection applied to x-ray inspection using numerical simulations, Nondestructive Testing and Evaluation 37 (5) (2022) 536–551. `doi:10.1080/10589759.2022.2071892`.

[13] G. Rao, B. Plank, J. Kastner, Comparison of different segmentation methods for porosity evaluation in CFRP-reference samples with real porosity samples, in: 6th Conference on Industrial Computed Tomography (iCT), 2016, pp. 1–8. URL `https://www.ndt.net/article/ctc2016/papers/ICT2016_paper_id43.pdf`

[14] M. Reseco Bato, A. Hor, A. Rautureau, C. Bes, Experimental and numerical methodology to obtain the probability of detection in eddy current ndt method, NDT & E International 114 (2020) 102300. `doi:10.1016/j.ndteint.2020.102300`.

[15] N. Dominguez, F. Reverdy, F. Jenson, POD evaluation using simulation: A phased array ut case on a complex geometry part, in: 40th Annual Review of Progress in Quantitative Nondestructive Evaluation: Incorporating the 10th International Conference on Barkhausen Noise and Micromagnetic Testing, Vol. 1581 of American Institute of Physics Conference Series, 2014, pp. 2031–2038. `doi:10.1063/1.4865073`.

[16] J. C. Aldrin, H. A. Sabbagh, R. K. Murphy, E. H. Sabbagh, J. S. Knopp, E. A. Lindgren, M. R. Cherry, Demonstration of model-assisted probability of detection evaluation methodology for eddy current nondestructive evaluation, AIP Conference Proceedings 1430 (1) (2012) 1733–1740. `doi:10.1063/1.4716421`.

[17] N. Dominguez, V. Feuillard, F. Jenson, P. Willaume, Simulation assisted pod of a phased array ultrasonic inspection in manufacturing, AIP Conference Proceedings 1430 (1) (2012) 1765–1772. `doi:10.1063/1.4716425`.

[18] W. Dewulf, H. Bosse, S. Carmignato, R. Leach, Advances in the metrological traceability and performance of x-ray computed tomography, CIRP Annals 71 (2) (2022) 693–716. `doi:10.1016/j.cirp.2022.05.001`.

[19] O. González, S. Rodríguez, R. Pérez-Jiménez, B. R. Mendoza, A. Ayala, Comparison of monte carlo ray-tracing and photon-tracing methods for calculation of the impulse response on indoor wireless optical channels, Opt. Express 19 (3) (2011) 1997–2005. `doi:10.1364/OE.19.001997`.

[20] S. Reisinger, S. Kasperl, M. Franz, J. Hiller, U. Schmid, Simulation-based planning of optimal conditions for industrial computed tomography, in: International Symposium on Digital Industrial Radiology and Computed Tomography, 2011.

[21] C. Bellon, G. Jaenisch, aRTist–analytical RT inspection simulation tool, in: 17th World Conference on Nondestructive Testing, 2007.

[22] W. Van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg, J. Sijbers, Fast and flexible X-ray tomography using the ASTRA toolbox, Opt. Express 24 (22) (2016) 25129–25147. `doi:10.1364/OE.24.025129`.

[23] M. Reiter, M. Erler, C. Kuhn, C. Gusenbauer, J. Kastner, SimCT: a simulation tool for X-ray imaging, in: 6th Conference on Industrial Computed Tomography (iCT), 2016.

[24] CERA User Manual Version 6.0.0 (2019).
URL `https://www.cera-systeem.nl/en/downloads`

[25] S. Jan, G. Santin, D. Strul, S. Staelens, K. Assié, D. Autret, S. Avner, R. Barbier, M. Bardiès, P. Bloomfield, D. Brasse, V. Breton, P. Bruyndonckx, I. Buvat, A. Chatziioannou, Y. Choi, Y. Chung, C. Comtat, D. Donnarieix, C. Morel, GATE: a simulation toolkit for PET and SPECT, Physics in medicine and biology 49 (2004) 4543–61. `doi:10.1088/0031-9155/49/19/007`.

[26] I. Buvat, D. Lazaro, Monte carlo simulations in emission tomography and GATE: An overview, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 569 (2) (2006) 323–329, proceedings of the 3rd International Conference on Imaging Technologies in Biomedical Sciences. `doi:10.1016/j.nima.2006.08.039`.

[27] R. J. Jaszczak, R. E. Coleman, C. B. Lim, Spect: Single photon emission computed tomography, IEEE Transactions on Nuclear Science 27 (3) (1980) 1137–1153. `doi:10.1109/TNS.1980.4330986`.

[28] S. Staelens, K. Vunckx, J. De Beenhouwer, F. Beekman, Y. D'Asseler, J. Nuyts, I. Lemahieu, Gate simulations for optimization of pinhole imaging, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 569 (2) (2006) 359–363, proceedings of the 3rd International Conference on Imaging Technologies in Biomedical Sciences. `doi:10.1016/j.nima.2006.08.071`.

[29] P. Weinberger, B. Plank, B. Fröhler, J. Kastner, C. Heinzl, Segmentation of multiple features in carbon fiber reinforced polymers using a convolutional neural network, in: Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XVI, International Society for Optics and Photonics, 2022. `doi:10.1117/12.2612895`.

[30] P. Fuchs, T. Kröger, C. S. Garbe, Defect detection in ct scans of cast aluminum parts: A machine vision perspective, Neurocomputing 453 (2021) 85–96. `doi:10.1016/j.neucom.2021.04.094`.

[31] R. Sarma, Y. K. Gupta, A comparative study of new and existing segmentation techniques, IOP Conference Series: Materials Science and Engineering 1022 (1) (2021) 12–27. `doi:10.1088/1757-899x/1022/1/012027`.

[32] J. A. Hartigan, M. A. Wong, Algorithm AS 136: A K-Means clustering algorithm, Applied Statistics 28 (1) (1979) 100–108. `doi:10.2307/2346830`.

[33] L. Belaid, W. Mourou, Image segmentation: a watershed transformation algorithm, Image Analysis & Stereology 28 (2) (2009) 93–102. `doi:10.5566/ias.v28.p93-102`.

[34] D. Wang, P. Vallotton, Improved marker-controlled watershed segmentation with local boundary priors, 2010 25th International Conference of Image and Vision Computing New Zealand (2010) 1–6.

[35] P. McCullagh, J. Nelder, Generalized Linear Models, 2nd Edition, Chapman and Hall, 1989.

[36] T. Sørensen, A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons, Kongelige Danske Videnskabernes Selskab, Biologiske Skrifter (1948) 1–34.