DTU Library

Automated generation of labeled synthetic training data for machine learning based segmentation of 3D-woven composites

Friemann, Johan; Mikkelsen, Lars P.; Oddy, Carolyn; Fagerström, Martin

Published in:
Proceedings of the 21st European Conference on Composite Materials

Link to article, DOI:
10.60691/yj56-np80

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

AUTOMATED GENERATION OF LABELED SYNTHETIC TRAINING DATA FOR MACHINE LEARNING BASED SEGMENTATION OF 3D-WOVEN COMPOSITES

Johan Friemann[1*], Lars P. Mikkelsen[2], Carolyn Oddy[1†,3] and Martin Fagerström[1‡]

[1]Division of Material and Computational Mechanics, Department of Industrial and Materials Science, Chalmers University of Technology, Gothenburg, Sweden
Email: *johan.friemann@chalmers.se, †carolyn.oddy@chalmers.se, ‡martin.fagerstrom@chalmers.se, Web Page: https://www.chalmers.se/en/

[2]Wind Energy Materials and Components Division, Department of Wind and Energy Systems, Technical University of Denmark, Roskilde, Denmark
Email: lapm@dtu.dk, Web Page: https://www.dtu.dk/english/

[3]GKN Aerospace Sweden, Trolhättan, Sweden
Email: carolyn.oddy@gknaerospace.com, Web Page: https://www.gknaerospace.com/

Keywords: x-ray computed tomography, woven composites, segmentation, synthetic training data

Abstract

A novel pipeline for the generation of synthetic tomograms of woven composite materials, to be used for training of machine learning based segmentation algorithms is presented. The pipeline is completely based on open source software and heavily utilizes the graphical processing unit for fast data generation. The proposed method generates a surface mesh of the woven geometry, scans it, reconstructs the scan, and generates a voxel labeling of the generated tomogram. It is demonstrated that the method can generate images that show good agreement with experimentally produced x-ray computed tomography images of a 3D-woven carbon fiber reinforced polymer composite.

1. Introduction

3D-woven carbon fiber reinforced polymer (CFRP) composites have garnered great interest recently due to their potentially good specific mechanical properties. When compared with the more common 2D-woven, or laminated, composites they show improved out of plane stiffness and strength. Moreover, the through layer reinforcement offers protection against delamination. In order to efficiently utilize them in industrial applications, accurate computational models are required. Typically, the modeling workflow starts with the analysis of representative volume elements. The utilization of x-ray computed tomography (XRCT) to generate RVE geometries is a popular method. However, the segmentation of CT images, and their transformation into computational meshes can be challenging.

Owing to the low contrast to noise ratio in scans of CFRP, classical segmentation algorithms such as filtering or thresholding, often fail. This is especially true for large field of view scans performed with the intention of providing geometries for meso-scale models. When a scan aims to resolve the entire weave structure of an often quite large unit cell, the diffuse boundaries between the matrix infused yarns and the surrounding pure matrix make separation of phases challenging. As an alternative, machine learning based methods show great promise. One pressing issue of machine learning methods is their need for large amounts of labeled data for training. High quality scans take several hours (or require access to synchrotron facilities), and the labeling must often be performed painstakingly by hand. It is therefore of interest to investigate the prospects of sim to real transfer learning, where the segmentation algorithm is trained on synthetic data but utilized on real data.

Ali et al. [1] successfully demonstrated how to use synthetic XRCT images to train a deep convolutional neural network to perform semantic segmentation of yarns and matrix regions of a woven composite. However, the generated 3D-models were directly transformed to gray-scale and noise was added as post processing. As such, the method does not account for any physical effects, or the typical artifacts introduced when performing tomographic reconstruction. It thereby reduces the image's realism. Mendoza et al. [2] used an alternative approach for generating training data. They trained a neural network to produce artificial XRCT images that were paired with segmentations derived from voxelized finite element simulation results. This method has the benefit of creating images with the type of noise and localized artifacts one would expect to see in a real tomogram. However, it fails to capture reconstruction artifacts. Furthermore, the model was limited to segmenting images from the same sample it was trained on. An alternative to both mentioned approaches would be to simulate the XRCT process and reconstruct images in the same way as real scans. This results in training sets that enable algorithms to account for the type of artifacts encountered in real XRCT reconstructions. The datasets can also be varied enough to enable algorithms to segment a wider range of samples.

The aim of this work is to present a novel pipeline for the generation of synthetic XRCT images of 3D-woven composites, paired with an automatically labeled segmentation for meso-scale models. The pipeline enables the automated generation of representative training data for machine learning based segmentation algorithms. Furthermore, the physically-based images (including reconstruction artifacts) also enable the evaluation of a scan setup before any real CT scan has been performed, potentially saving time and resources.

2. Method

To demonstrate the potential of synthetic XRCT images of woven composites a digital twin of a layer to layer angle interlock sample is created. A virtual scan of the digital twin is compared to an actual scan of the original sample. Furthermore, a method for automatic segmentation and labeling of the virtual scan is demonstrated.

2.1. Generation of synthetic tomogram-segmentation pairs

To begin with, representative surface meshes of the 3D-woven composite are generated. Following, a simulated XRCT is performed, and subsequently reconstructed. The reconstructed volume is segmented into its constituent phases by voxelizing the meshes making up the composite material. Everything from weave-pattern and material parameters to X-ray scan geometry and spectrum can be customized. This enables a user to generate a tailor-made training data set with domain randomization.

2.1.1. Generation of 3D-woven geometries

The fiber yarn bundles are considered as homogeneous volumes, i.e. individual fibers are not resolved. The open source software TexGen [3] is used to generate the geometry. TexGen models yarns using splines, where the interpolation nodes are assigned the associated cross-sections. This enables easy generation of surface meshes. In the current pipeline the user inputs the specifications of a unit cell of a layer to layer angle interlock composite, and can thereafter generate a sample consisting of several repetitions of this unit cell. Each unit cell can have different random distortions, enabling a realistic representation of irregularities in a real sample. TexGen does not guarantee that yarns do not penetrate each other. Therefore, a Boolean difference operation is performed on the meshes wherever the warp material cuts the weft material. For the current material under study this seemingly arbitrary step is motivated by the weft material being less dense, and should in practice give way to the warp material.

2.1.2. XRCT simulation and reconstruction

The open source software gVirtualXRay [4, 5] is used to generate synthetic X-ray tomographic projections. The software allows for a physically motivated X-ray attenuation through closed surface meshes following Beer-Lamberts law. Computations exploit the graphical processing unit (GPU) for real time simulation of X-ray projections. The simulated attenuation coefficients of different materials take density, chemical composition, and incident X-ray energy into account, all of which can be conveniently controlled from a python interface. The X-ray spectrum is derived with the open source software SpekPy [6]. The spectrum can be configured for different target materials, tube voltage and currents, and filter materials. For each projection, independently sampled photonic noise is added. Photonic noise is Poisson distributed, and is dependent on the incident photon flux. This is easily computed from gVirtualXRay and SpekPy based on the chosen configuration parameters. Once all the projections of the virtual scan has been simulated, tomographic reconstruction is performed with the open source software ASTRA [7]. The virtual projections are reconstructed with the FDK algorithm on the GPU.

2.1.3. Labeling

The CT segmentation problem consists of assigning each voxel in the reconstructed volume to its corresponding constituent phase. For a woven composite, there are three classes, namely weft, warp, and matrix material. To begin with, the meshes that encompass the virtual sample are rotated and translated such that the samples position and orientation matches the one in the tomography simulation. The sample is thereafter transformed into the same coordinate system as the reconstruction volume. This is achieved by scaling by the reconstruction voxel size, and shifting by half the number of voxels per side, resulting in the coordinate (0, 0, 0) in scanner coordinates to be mapped on to center of the voxel grid. The geometry is voxelized on the GPU in parallel over the mesh triangles, utilizing an algorithm by Schwartz and Seidel [8]. Each voxel is assigned an integer value according to the material phase it occupies. As the voxelization is performed in the same frame of reference as the sample inside the X-ray simulation, the resulting voxel grid becomes a 1 to 1 segmentation of the reconstructed volume.

2.2. XRCT scan

The layer to layer angle interlock sample was scanned on an actual XRCT machine. The physical sample was scanned on a *ZEISS XRadia 410 Versa* CT scanner. A tube voltage of 40 kV and power of 10 W was utilized. The proprietary “LE1” filter was utilized to filter the incident X-rays, and a detector exposure of 5 s was used. The sample was mounted 60 mm away from the X-ray source, while the detector was placed 80 mm behind the sample. A 0.4 times magnification optical system was used to focus the scintillator signal. All this together with a detector binning of 2 resulted in an effective voxel size of approximately 28.9 µm. A scan of 360 degrees with 3201 projections was carried out. To compare with the simulated images which are described in the next paragraph, an additional binning 2 was performed on the physical scan projections (total binning 4). Furthermore, every 4 projection was used, resulting in a total of 801 projections. The physical scan was reconstructed with the ASTRA implementation of FDK in the same way as the virtual scan.

A digital twin of the described sample was created, scanned and reconstructed with the proposed pipeline. The virtual scan was performed with the same optic, sensor resolution, and X-ray source parameters as the physical scan. The source filter, of unknown composition, was heuristically approximated with a 2 mm thick aluminum filter. The weft material was heuristically approximated by pure carbon with a density of 1.72 g/cm^3, the warp material was similarly approximated by pure carbon of density 1.82 g/cm^3. The matrix material was approximated by an arbitrary epoxy-like composition of 40.4% carbon, 48.1% hydrogen, 1.9% chlorine, and 9.6% oxygen [9]. The density of the matrix material was chosen as 1.08 g/cm^3. The material densities (and the fact that the yarns are 100% carbon) are unrealistic. They were chosen to best match the relative contrasts of the experimental scan, given that their actual composition was unknown. A 360 degree scan was performed with 801 projections and binning 4.

3. Results

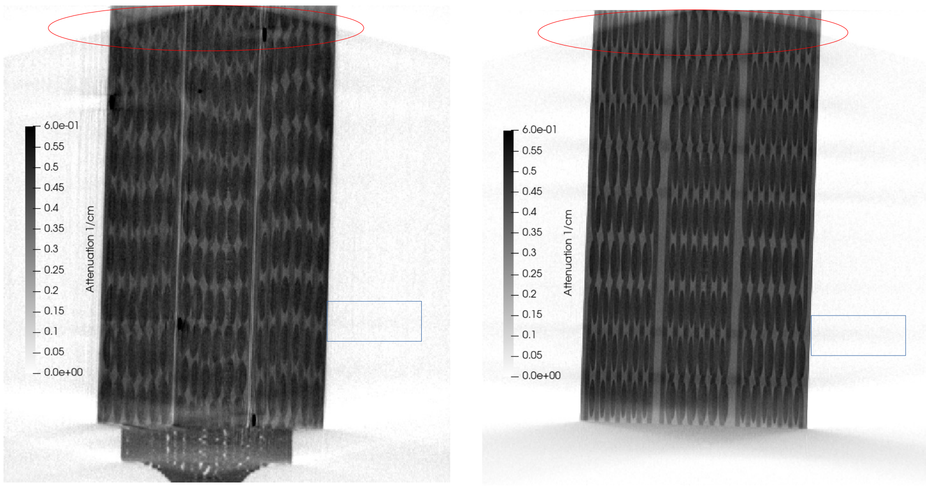

Figure 1: The real sample (left) is shown next to the virtual scan of its digital twin (right). The warp yarns are facing the viewer. An example of a cone beam artifact is highlighted with a red oval, while an example of a streak artifact is shown with a blue rectangle.

A slice from the tomographic reconstruction of the scan of the sample, together with its virtually scanned digital twin, is shown in Figure 1. The displayed images show qualitative agreement. Note how the virtual scan displays the expected cone beam reconstruction artifact at the bottom and top. Furthermore, streak artifacts (that arise due to the limited number of projections) from the gaps between the relatively denser warp yarns can be seen in both images. It is worth noting that the discrepancy between the images at the bottom is due to the lack of a plastic sample mount in the virtual scan. The real scan shows slightly more noise. This can partly be explained by the very absorbent glass fiber tracer yarns (seen as black patches) present in the real sample. However, it can also be explained by the lack of sensor and electronic noise among others in the simulated image.

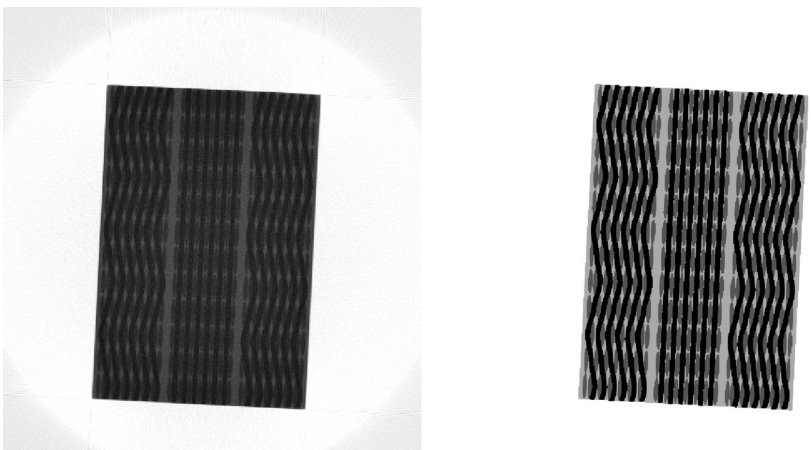

Figure 2: A slice of a virtual XRCT scan (left) is shown together with its automatically generated segmentation (right) is displayed. The weft yarns are facing the viewer.

Figure 2 displays a virtual scan next to its automatically generated labeled segmentation. The weft (dark gray), warp (black), and matrix (light gray) material can be seen as distinct phases with perfect contrast. A machine learning algorithm could for example take the left image as input and use the right image as the ground truth.

3.1 Quantitative comparison

A qualitative similarity between the simulated XRCT scan and the real scan can be seen in the previous section. In order to properly train an algorithm with synthetic data, the images need to be realistic enough for the model to be able to generalize. A few quantitative comparison metrics are shown. The first quantitative comparison is the histogram of attenuation values. This can be seen in Figure 3. The peak for the weft and warp yarns around 0.5 cm^{-1} agrees well between the two scans. The matrix peak around 0.42 cm^{-1} also agrees quite well in location. It is however much more prominent in the simulated image, which can be explained by the matrix rich regions between the layers (can be seen in Figure 1) not present in the real sample.

The histogram does not give any information on structure, as shuffling the volume would leave the histograms intact. The cosine similarity index can be used to infer about structural similarity between images. It is defined as

$$c = \boldsymbol{a} \cdot \boldsymbol{b} \,/\, (||\boldsymbol{a}||\ ||\boldsymbol{b}||), \tag{1}$$

where $\boldsymbol{a}$ and $\boldsymbol{b}$ are vectors that consist of the flattened XRCT images, $\cdot$ is the scalar product and $||\boldsymbol{x}||$ is the euclidean norm of $\boldsymbol{x}$. The cosine similarity index measures the cosine of the "angle" between two images, where the index of an image and its scaled version should result in the value 1.

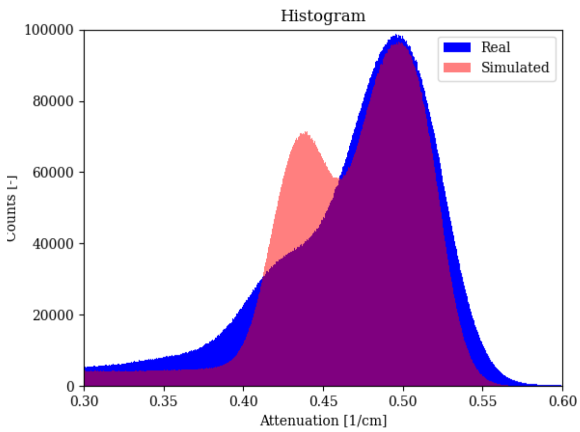

Figure 3: Attenuation value histograms for the simulated and real XRCT scans are displayed.

Computing the cosine similarity index between the XRCT tomograms whose slices are displayed in Figure 1, a value of 0.93 is received. This demonstrates that the images structure are closely related.

4. Conclusions

In this paper a novel pipeline for the generation realistic synthetic, automatically labeled training data for machine learning assisted segmentation of 3D-woven composites has been presented. The method enables the generation of a training data set for a given segmentation problem that is tailored for the material and available equipment. It is demonstrated that the proposed method can generate synthetic tomograms that agree with experimentally retrieved images both qualitatively and quantitatively. This indicates their potential in being used for sim to real transfer training of a segmentation algorithm.

Acknowledgments

The authors acknowledge funding from Horizon Erope through the MSCA Doctoral Network RELIANCE, grant no. 101073040.

References

[1] M. Ali and Q. Guan and R Umer and W. Cantwell and T. Zhang. Deep learning based semantic segmentation of µCT images for creating digital material twins of fibrous reinforcements. *Composites Part A: Applied Science and Manufacturing*, 139(106131), 2020.

[2] A. Mendoza and R. Trullo and Y. Wielhorski. Descriptive modeling of textiles using FE simulations and deep learning. *Composites Science and Technology,* 213(108897), 2021.

[3] L. Brown and A. Long. Modelling the geometry of textile reinforcements for composites: TexGen. *Composite Reinforcements for Optimum Performance (Second Edition).* Woodhead Publishing, Cambridge, 2021.

[4] F. Vidal and M. Garnier and N. Freud and J. Létang and N. John. Simulation of X-ray attenuation on the GPU. In *Proceedings of Theory and Practice of Computer Graphics 2009*, pages 25-32, 2009.

[5] F. Vidal and P. Villard. Development and validation of real-time simulation of X-ray imaging with respiratory motion. *Computerized Medical Imaging and Graphics*, 49:1-15, 2016.

[6] G. Poludniowski and A. Omar and R. Bujila and P. Andreo. Technical Note: SpekPy v2.0—a software toolkit for modeling x-ray tube spectra. *Medical Physics,* volume 48:3630-3637, 2021.

[7] W. van Aarle and W. Palenstijn and J. De Beenhouwer and T. Altantzis and S. Bals and K. Batenburg and J. Sijbers. The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography, *Ultramicroscopy*, 157:35–47, 2015.

[8] M. Schwarz and H. Seidel. Fast parallel surface and solid voxelization on GPUs. *ACM Transactions on Graphics,* 29(179), 2010.

[9] National Center for Biotechnology Information. PubChem Compound Summary for CID 169944, Epoxy resin. https://pubchem.ncbi.nlm.nih.gov/compound/Epoxy-resin. Accessed May 14, 2024.