



Contents lists available at ScienceDirect

Composites Science and Technology

journal homepage: www.elsevier.com/locate/compscitech

Deep learning segmentation of wood fiber bundles in fiberboards

Pierre Kibleur^{a,d,*}, Jan Aelterman^{b,c,d}, Matthieu N. Boone^{c,d}, Jan Van den Bulcke^{a,d},
Joris Van Acker^{a,d}

^a Department of Environment, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000, Ghent, Belgium

^b Department of Telecommunications and information Processing - imec, Faculty of Engineering and Architecture, Ghent University, Sint-Pietersnieuwstraat 41, 9000, Ghent, Belgium

^c Department of Physics and Astronomy, Faculty of Sciences, Ghent University, Proeftuinstraat 86, 9000, Ghent, Belgium

^d Ghent University Center for X-ray Tomography (UGCT), Proeftuinstraat 86, 9000, Ghent, Belgium

ARTICLE INFO

Keywords:

Bio composites
Natural fibers
Material modeling
X-ray computed tomography

ABSTRACT

Natural fiber composites and fiberboards are essential components of a sustainable economy, making use of bio-sourced, and also recycled materials. These composites' structure is often complex, and their mechanical behavior is not yet fully understood. A major barrier in comprehending them is the ability to identify the fibers in situ, i.e. embedded in complex fibrous networks such as medium-density fiberboards (MDF). To that end, the first step is to separate individual wood fibers from fiber bundles. Modern material studies on real world, dense fibrous materials using X-ray microtomography and 3D image analysis were always limited in accuracy. However, recent machine learning techniques and particularly deep learning may help to overcome this challenge. In this work, we compare existing segmentation algorithms with the performance of convolutional neural networks (CNNs). We explain the need for network complexity, and demonstrate that our best algorithm, based on the UNet3D architecture, reaches unprecedented accuracy. Moreover, it achieves the first segmentation sufficiently qualitative to extract morphometric measurements of the fiber bundles and accurately estimate their density. Among other applications, the proposed method thus enables the design of more realistic material models of MDF, and is a milestone towards the understanding and improvement of this wood-based product.

1. Introduction

The past decades, fiber composites have alleviated several material limitations that bulk materials encounter, and have been used in many innovative applications. With environmental considerations in mind, natural fiber composites have further advantages: using the fibers from flax, hemp, jute, bamboo, wood, etc., not only ensures renewability and biodegradability, but also contributes to the consumption (or storage) of CO₂ from the atmosphere [1]. In this regard, wood fibers are particularly interesting. Wood accounts for up to 80% of the biomass on Earth, and roughly half of the wood mass is carbon. Naturally, wood is already thought of as a paramount element to fight climate change [2], but as a bulk material wood is limited: the resource used to make sawn timber requires a form of homogeneity that is not necessarily found in nature. By breaking wood into smaller pieces, this limitation can be overcome and novel materials can be created, allowing to exploit a larger proportion of the available, naturally-grown biomass.

Yet, such re-engineering of wood has its limitations. Some wood-

based panels (WBPs) exhibit exacerbated weaknesses compared to solid wood, and these necessitate modern material investigations. Making use of the analysis possibilities offered by X-ray microtomography (μ CT), several studies already pursued cutting-edge 3D material characterizations at the micro-scale [3]. However, these studies are limited by the material density, because the advanced image processing methods that can be applied at low densities [4,5] fail at material densities above $\sim 500 \text{ kg m}^{-3}$ [6]. And while no current method works at high densities (above $\sim 1000 \text{ kg m}^{-3}$), even the simpler methods do not work well at medium-high density [7,8].

Medium-density fiberboard (MDF), a wood fiber composite with a density of approx. 750 kg m^{-3} , embodies these different considerations. It is made of mechanically defibrated wood fibers, it is dense, and it exhibits a relatively high weakness to water absorption. This creates a real need for the investigation and modeling of MDF. Furthermore, it is also the second most produced WBP in the world, with applications everywhere. To produce it, wood chips are first transformed into individual wood fibers by disk refining. The process is fast: around 30 tons of

* Corresponding author. Department of Environment, Faculty of Bioscience Engineering, Ghent University, Coupure Links 653, 9000, Ghent, Belgium.

E-mail address: pierre.kibleur@ugent.be (P. Kibleur).

<https://doi.org/10.1016/j.compscitech.2022.109287>

Received 7 July 2021; Received in revised form 15 January 2022; Accepted 21 January 2022

Available online 29 January 2022

0266-3538/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

fibers are produced per hour with a thermo-mechanical defibrillator (or refiner). The process is similar to the pulp production in the pulp & paper sector. However, not all fibers are perfectly defibrated. Specifically, fiber bundles can remain, which preserve some structure and properties of solid wood. Eventually embedded in the material, their presence is expected to affect the panel's properties differently than single fibers. Their identification is thus pertinent and critical to the study of MDF.

While advanced traditional image processing methods such as the tracking of a single fiber's cross-section [3] are limited by density in practice, deep learning has alleviated several technical barriers in different areas of image processing [9]. Applied to the study of MDF, deep learning could settle new standards in the segmentation accuracy of fiber bundles. These features are indeed recognizable by the trained human eye. And although there could be some ambiguity in recognizing them, supervised learning is therefore promising.

Moreover, precisely segmenting the fiber bundles embedded in MDF is capital for future studies. Particularly for modeling, it could greatly impact the accuracy and thus relevance of mechanical material models of MDF [10]. In dynamic experiments, the presence of fiber bundles could guide the analysis of strain fields to understand how they affect the properties of MDF. In analyses about the additives of MDF (e.g. glue), it could permit to assert the performance of the adhesive by comparing the glue adherence and penetration in bundles and single fibers. Finally, it could even allow single fiber analyses at high densities. Since a fiber bundle contains fibers with similar orientations, this information could be used to isolate each wood fiber in a fiber bundle - something that cannot be achieved outside of a bundle at higher material densities - and thus retrieve individual fiber measurements. Although these measurements would not necessarily be sufficient to fully characterize the individual fibers outside of the fiber bundles, they could nevertheless bring valuable and so far unique information about the wood fibers.

In this work, we exploit the potential of deep learning for fiber bundle segmentation. 2D, pseudo-3D and full 3D network architectures were considered, and each method was qualitatively and quantitatively evaluated against one another. Eventually, the segmentation results achieved enabled a first advanced characterization of fiber bundles in a commercial MDF panel.

2. Materials and methods

2.1. Dataset

A 9 mm thick, commercially-available MDF panel (seen on Fig. 1), with $\sim 775 \text{ kg/m}^3$ density was provided by the manufacturer. We extracted a sample of size $2 \times 2 \times 9 \text{ mm}^3$ with a table saw. This was the



Fig. 1. Left: MDF panel samples of different thicknesses. These are standard line-production panels cut for promotional purposes. Out of such panels, millimetric samples were extracted for X-ray tomography. Right: 2 mm-wide, 9 mm thick MDF sample mounted on a carbon stick for X-ray CT scanning.

smallest sample we could extract without visible damage to the otherwise brittle fibrous structure. MDF is a wood fiber composite, with fibers sourced from different local wood species. Here, we expect a majority ($\sim 80\%$) of fibers to be from Norway spruce (*Picea abies*). MDF has a symmetry with respect to its middle in the thickness direction [6]. Therefore only the bottom half of the sample was scanned with μCT .

The sample was scanned with the in-house built CT system Nano-wood [11] of the Ghent University Centre for X-ray Tomography (UGCT, Ghent, Belgium). 2001 projections were acquired at a tube voltage of 50 kV, power of 5 W, 1 s exposure, and no filter. The source-detector distance was about 1000 mm, and the source-object distance was 20 mm, resulting in a magnification of ~ 50 . 16-bit volume reconstructions were obtained with Octopus Reconstruction [12]. The Feldkamp algorithm, based on the filtered-back projection (FBP), was used to reconstruct the images. In these images, the approximate voxel pitch was $2.5 \mu\text{m}$. With a detector panel ~ 1800 pixels wide, the system's horizontal field of view was about 4.5 mm wide.

2.1.1. Labeled set

From the reconstructed volume measuring $1088 \times 1686 \times 772$ voxels³, two distinct regions of interest (ROI₁ and ROI₂, of dimensions $401 \times 400 \times 267$ voxels³ each, centered around a couple of noticeable fiber bundles) were extracted. All fiber bundles inside these ROIs were manually and individually identified as shown in Fig. 2, to produce so-called groundtruths (GND₁ and GND₂) that could be compared with a software prediction of the fiber bundles. As shown in Fig. 2, only (a part of) ROI₁ would then be used to supervise the training of CNNs. ROI₂ was only used after the networks have been trained, and allows unbiased estimations of the segmentation accuracy.

Visually, fiber bundles are recognizable from their cross-section. In MDF, fiber orientation is mainly horizontal, with random in-plane orientations. This reduced the problem to identifying fiber bundle cross-sections in the two orthogonal vertical cross-sections. In Dragonfly 2020.1 (Object Research Systems, Montreal, Canada - free of charge for non-commercial use), we identified the contours of a fiber bundle on several 2D slices throughout the depth dimension (1/50 slices). Then, we interpolated these sparse 2D cross-sections along the depth direction, to retrieve the full 3D segmentation of one single fiber bundle. The result of the interpolation was manually refined/cleaned when necessary. The operation was repeated for all bundles that could be identified in the ROIs. Finally, we combined these manually annotated bundles into a single image with a logical "OR", to retrieve a given groundtruth. A view of the GND₁ is shown in Fig. 3.

2.2. Structure tensor approach

A method to segment fiber bundles had been proposed before, based on morphological properties derived from the structure tensor T [8]. The eigenvalues and corresponding eigenvectors of the structure tensor provide information about the local gradient directions. Specifically, we may derive in every voxel a strength parameter $\xi := \log(\lambda_{\max}/\lambda_{\min})$, from the extremal eigenvalues λ_{\max} and λ_{\min} of T [10]. Previous studies have shown that ξ is high at points of strong directionality [13]. We might therefore use it to automatically detect the presence of fiber bundles, because fibers in a fiber bundle share the same orientation. Consequently in sufficiently large neighborhoods, the average strength parameter should highlight the presence of fiber bundles. Thus, after locally smoothing ξ in cubes of side $60 \mu\text{m}$ (24 voxels in our dataset - or about 3-4 contiguous full wood fibers assuming a diameter of $15-20 \mu\text{m}$ [14]), we manually adapted a threshold to obtain a plausible binary segmentation of fiber bundles in the entire volume. This approach was implemented in Python, primarily using the scikit-image package [15].

2.3. Deep learning

Deep learning segmentation was carried out in Dragonfly 2020.1. A

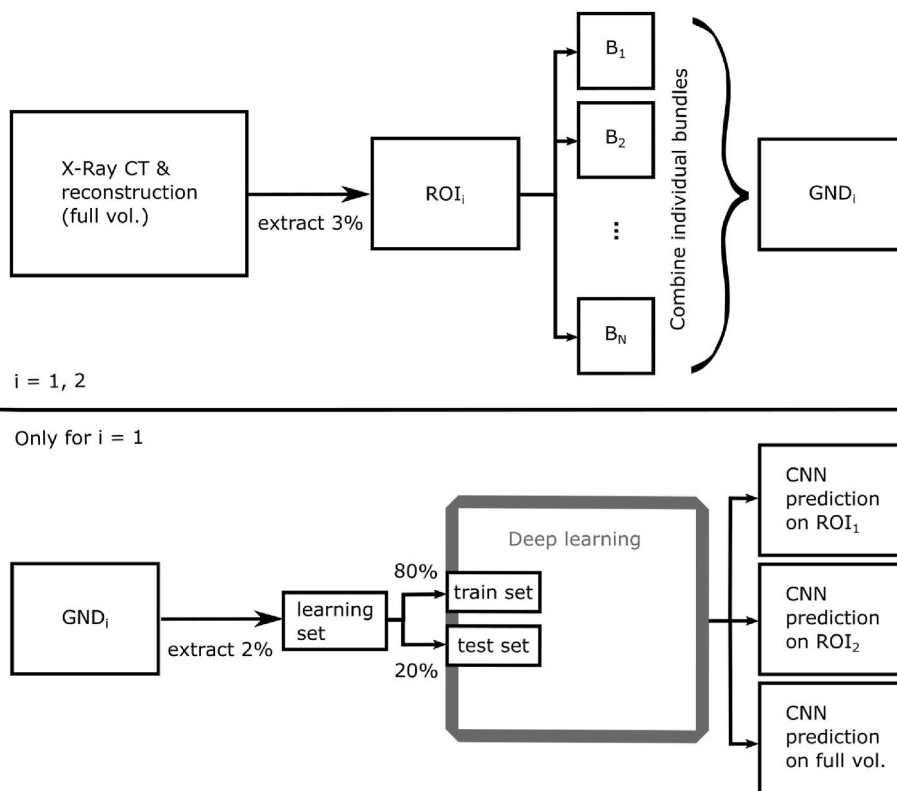


Fig. 2. Diagrams of the experiment. Top: creation of two groundtruths from the full volume (vol.). Bottom: supervised training of a convolutional neural network (CNN) from a subset of GND_1 , and its possible predictions.

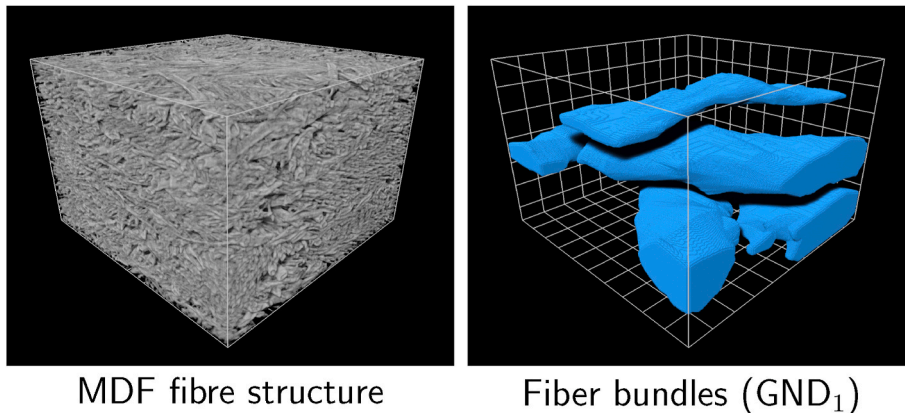


Fig. 3. Raw MDF structure (left) and manually annotated fiber bundles (right) in ROI_1 . One grid square is $100 \mu m^2$. The vertical direction corresponds to the panel thickness direction: fibers bundles and fibers thus lay in the horizontal plane.

few corresponding slices from ROI_1 and GND_1 were used as a learning set to train a convolutional neural network (CNN). Specifically, every 50th slice in the depth direction of ROI_1 was extracted. Hence, the resulting learning set had dimensions $9 \times 400 \times 267$ voxels, corresponding to 2% of the ROI in volume. This unique learning set was used to train all network architectures. The three deep learning approaches are schematized in Fig. 4, and described below.

2.3.1. UNet2D

At first, a standard UNet architecture for image segmentation was considered (later referred to as UNet2D). It was parametrized with 32 layers, and totaled $2 \cdot 10^7$ nodes. Input to this network, were 2D patches of 400×400 pixels, organized in batches of a single patch each epoch. These extremal settings followed two considerations. First, the fiber

bundles are rather large features that should be markedly different from their surroundings. Second, it is advised by the developers of the UNet architecture to use maximal (as GPU memory permits it) patch sizes, at the cost of minimizing the batch size [16]. The network was trained in 250 epochs, but had stabilized in a little more than 100.

2.3.2. Bilateral UNet2D

Because of the variability of the 2D network’s predictions throughout the volume’s depth (see section 4.3.1), a second strategy was tried. The network -that was already trained-would be applied twice, along two orthogonal axes in the horizontal plane. This strategy is later referred to as Bilateral UNet2D. It is simply an extension of the UNet2D strategy, by complementing its result with a segmentation of orthogonal slices from the same 3D dataset. The two segmentations were merged with a logical

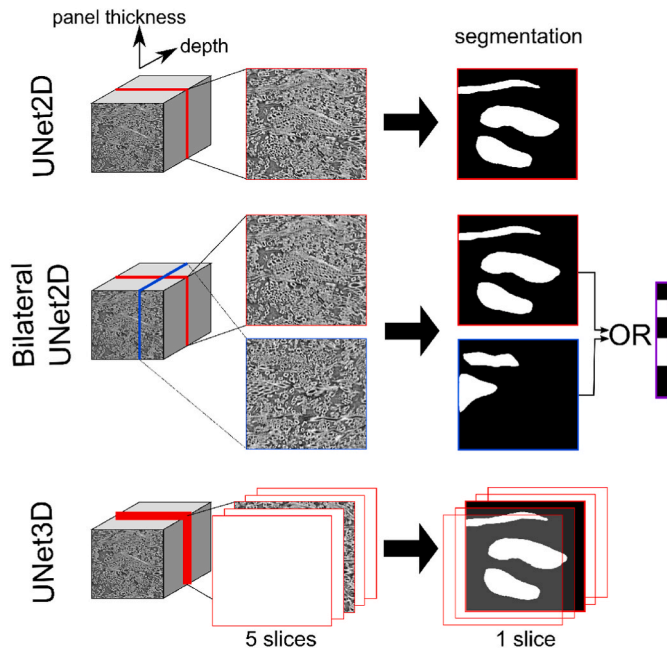


Fig. 4. Visualization of the different machine learning strategies.

“OR”.

2.3.3. UNet3D

Finally, a more complex UNet3D architecture [17] was tested. The first layer counted 32 filters, and the network was organized in 5 levels, which resulted in a network weighing $3 \cdot 10^8$ nodes. Patches were $128 \times 128 \times 5$ voxels³, where the smallest dimension was taken along the ROI's depth direction. The batch size was once again 1. This network was trained in 200 epochs (completed in half a day on a NVIDIA Quadro P5000), but had stabilized in just 75.

2.4. Post processing

2.4.1. Binary cleaning

First, binary “holes” in the fiber bundle phase smaller than 10^7 voxels were removed. Second, it was smoothed by binary opening with a spheroidal structuring element. We implemented this structuring element after identifying that given the fiber bundles aspect, we were expecting to retrieve “flat” shapes. This suggested the use of a prolate structuring element. A grid search on the vertical and horizontal radii nevertheless resulted in choosing these as 11 and 13 voxels, respectively. Finally, the fiber bundle phase was cleaned of objects smaller than 10^6 voxels, as a fiber bundle should always be large.

2.4.2. Measuring performance

Having a 3D groundtruth, we could measure the performance of a given method with Matthew's correlation coefficient (MCC) [18]. The MCC, which can be used even if the cardinality of each segmented class varies, ranges from -1 (inverse prediction) to 1 (perfect prediction). 0 indicates an average random prediction.

2.4.3. Labeling

Given fiber bundles are disconnected in the labeled set, we could directly proceed to labeling. However, in case of contiguous bundles in the CNN prediction, we chose to perform a watershed transform. This made the separation of different fiber bundles more robust. An Euclidean distance map was computed, and mean-thresholded, to produce the watershed transform markers.

2.4.4. Shape descriptive metrics

After labeling, it becomes possible to count and measure the fiber bundles. In that endeavor, we defined the (normalized) surface to volume ratio as $SA:V := \frac{S}{V} r_{eq}$, where r_{eq} is a bundle's equivalent radius, S its surface (adapted from Ref. [19]), and V its volume. The surface to volume ratio, which would be 3 for a sphere and 6 for a cube, illustrates a general shape of the identified fiber bundles. Additionally, the aspect ratio was defined as the ratio of a bundle's longest semi-axis length over its smallest semi-axis length, both calculated from the extremal eigenvalues of the inertia tensor, with the principal axis theorem. This second ratio suggests a form factor, which could intuitively be interpreted as the elongation of the bundle.

2.4.5. Horizontal orientation

Finally, labeled 3D objects could be analyzed to retrieve their horizontal orientation. For each object, its 3D inertia tensor was computed, and reduced to its non-vertical components. The resulting 2×2 tensor was diagonalized, and the eigenvector corresponding to its largest eigenvalue indicated its horizontal orientation. This value, in the range $]0, 180]$ degrees, was then mapped onto the binary fiber bundle segmentation for visualization. As other post-processing steps, computing the horizontal orientation for each fiber bundle in the volume was not particularly computationally demanding (the full volume of $1088 \times 1686 \times 772$ voxels³ was processed in ~ 5 min on an Intel Xeon W-2133 CPU @ 3.6 GHz).

3. Results

3.1. Segmentation

The performance of each algorithm is presented in Table 1. While it shows that the prediction quality improved with the CNN complexity, it also highlights that the simpler CNNs already vastly outperformed the morphometric methods on either test ROI. Two significant jumps in prediction accuracy were performed; the first from ST3D to UNet2D; the second from UNet2D to UNet3D.

The prediction of selected algorithms (ST3D, UNet2D and UNet3D) are overlaid on the groundtruth GND_1 on 3 of its slices, in Fig. 5. For comparison purposes, these 3 slices were selected to correspond to the maximal, median, and minimal MCC score/slice of the UNet3D segmentation.

In short, the ST3D segmentation lacks homogeneity and completeness; UNet2D is plausible, but seems to miss part of the fiber bundles along the depth direction, and may completely fail to recognize fiber bundles orthogonal to it; UNet3D, while not perfect either, largely alleviates UNet2D's shortcomings: even in the worst MCC slice, the bundle phase is close to the groundtruth.

Once a visually convincing fiber bundle segmentation was reached with UNet3D, i.e. when the algorithm was able to recognize the same bundles that a human operator had, with shapes resembling fiber bundles, case-specific image processing routines could be applied to smooth the segmentation. A comparison of the bundle phase before and after post-processing is shown in Fig. 6. Clean bundles can furthermore be color-coded with their properties (e.g. volume, orientation, length, etc.). In Fig. 6, the fiber bundles in GND_1 are colored based on the angle associated to their horizontal orientation.

Table 1
Measured performance of different fiber bundle segmentation methods.

Segmentation	$MCC _{ROI_1}$	$MCC _{ROI_2}$
ST3D	0.435	0.506
UNet2D	0.589	0.554
Bilateral UNet2D	0.625	0.577
UNet3D	0.723	0.602
Post-processed UNet3D	0.748	0.705

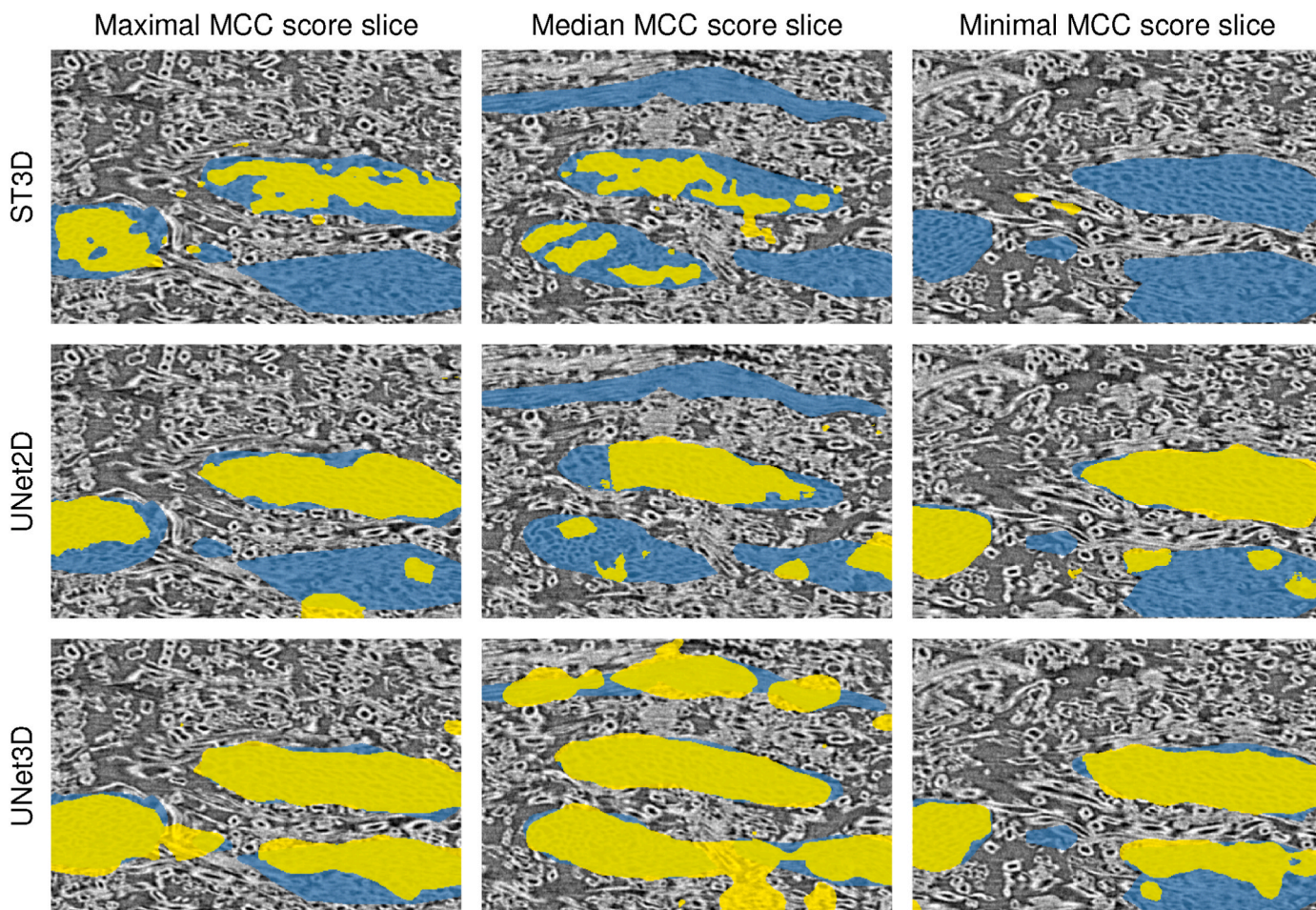


Fig. 5. Selected slices for different ROI₁ segmentations: ST3D, UNet2D and UNet3D. In blue, the groundtruth; in yellow, an automatic segmentation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

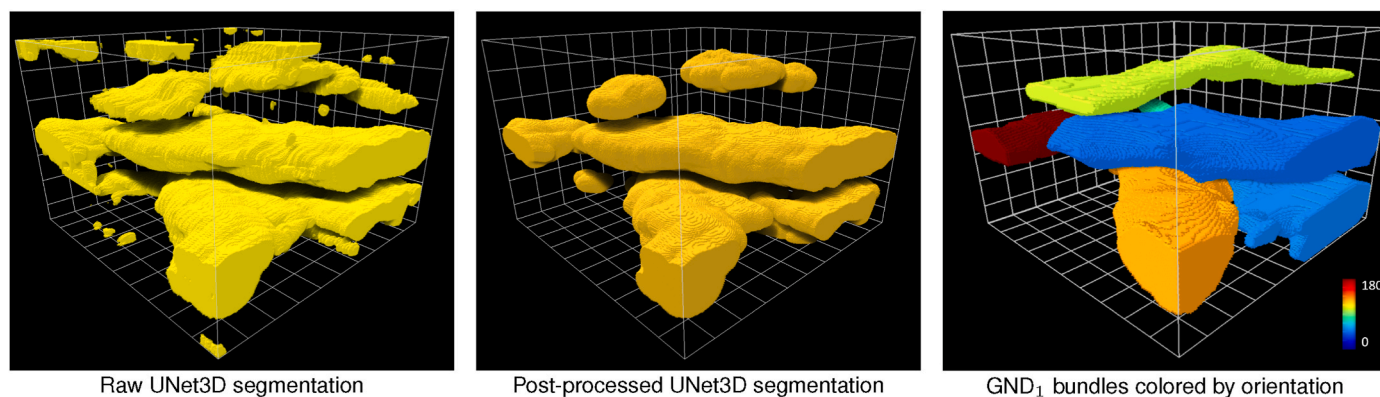


Fig. 6. Post-processing the UNet3D segmentation results in a smoothed and rounder phase, without the small irregularities that could not constitute a fiber bundle. After labeling, horizontal orientations can be determined for each individual bundle. One grid square corresponds to 100 μm^2 .

3.2. Fiber bundle analysis

Overall, 3 segmentations (GND₁, the post-processed UNet3D segmentation of ROI₁, and the post-processed UNet3D segmentation of the entire scanned volume) were labeled and analyzed. Several properties of the labeled fiber bundles were computed for each of these 3 segmentations, and presented in Table 2.

Comparing the metrics obtained from the groundtruth with those obtained from the UNet3D segmentation of the same ROI gives an

insight into the final, unverifiable results of the full volume segmentation. The automatic segmentation of ROI₁ yielded an appropriate estimate of the amount and density of fiber bundles, but tended to underestimate their volume, possibly by failing to capture them along their entire length (as the median aspect ratio is lower in automatically segmented fiber bundles), while overestimating their width (since the bundle density is correctly estimated).

The labeled fiber bundles in the entire MDF volume, color-coded for their horizontal orientation, are shown in Fig. 7. Alongside, the binary

Table 2
Measured properties of the segmented fiber bundles.

Property	GND ₁	ROI ₁ (UNet3D)	Full volume (UNet3D)
Bundle count	8	7	141
Bundle density [%]	21	20	12
Median bundle volume [μm^3]	$1.2 \cdot 10^7$	$5.6 \cdot 10^6$	$3.8 \cdot 10^6$
Median surface to volume ratio	9.8	8.6	8.7
Median aspect ratio	2.7	1.6	1.8

prediction capabilities of the network are displayed by comparing them with details of the input image, selected through the sample.

4. Discussion

4.1. Evaluating performance

To quantitatively evaluate different algorithms, we must first address the quality of the groundtruths. Due to the material's nature, it is intricate to identify fiber bundles manually. Since the defibrating process (during which solid wood chips are separated into fibers) is not perfect, "bundles" with 2–100 fibers can remain. However, there is no unambiguous geometrical definition of a processed, natural and hollow fiber in such images as ours. Natural fibers, with different sources, come in different shapes. For MDF production, they are furthermore heavily processed, which alters the shapes that can be observed, various examples of which can be found in Ref. [20]. Therefore, we chose to focus on the larger bundles, whose geometrical shape is less determinant than their size, and which are expected to have the largest impact on the mechanical properties of the material. A minimal bundle size - based on surface area - had already been an identification criterion before [6]. Yet, it is possible that our algorithm would be capable of picking up the smaller bundles like it does the bigger ones, without the discerning power to reject the firsts. This would create a surge in observed false

positives. Besides, an imperfect groundtruth could confuse a CNN's training. Although this is out of the current scope, we therefore recommend a more elaborate manual segmentation performed by multiple independent operators, as is done in other fields [21]. Following this methodology may enable extremely detailed studies on the fiber bundles, as the segmentation accuracy increases.

4.2. Structure tensor approach

The structure tensor method aims to find homogeneities in the orientation field, and corroborate these with the presence of fiber bundles. This sound approach was validated in various materials, e.g. Ref. [8] or [22]. However, since the source code was not available, we had to re-implement it. Fixing parameters was done to the best of our ability with the given information, to produce the highest quality segmentation possible. Nonetheless, it is possible that our implementation does not reproduce the optimal implementation of this algorithm. In the worst case scenario, the ST3D accuracy could be underestimated due to our implementation. Notwithstanding, we believe that the improvements achieved with CNNs render any ST3D implementation obsolete.

4.3. Deep learning

4.3.1. Weakness in the depth direction

Segmenting fiber bundles requires 3D algorithms: even the best 2D predictions would indeed be varying in the third dimension, from one slice to the next. This is shown in Fig. 8, which presents a 2D slice normal to the panel thickness. The obvious inconsistencies in the depth direction lead to a sharp decrease in prediction accuracy.

For this reason, we first considered using pseudo-3D methods, such as combining UNet2D predictions in two orthogonal directions. However, our bilateral UNet2D method had noticeable shortcomings; the variability of UNet2D in the depth direction was arguably too large to satisfyingly compensate. Instead of exploring smarter ways of combining several non-coplanar segmentations than with an "OR", we chose to use a fully 3D method, based on UNet3D. As a drawback, a 3D

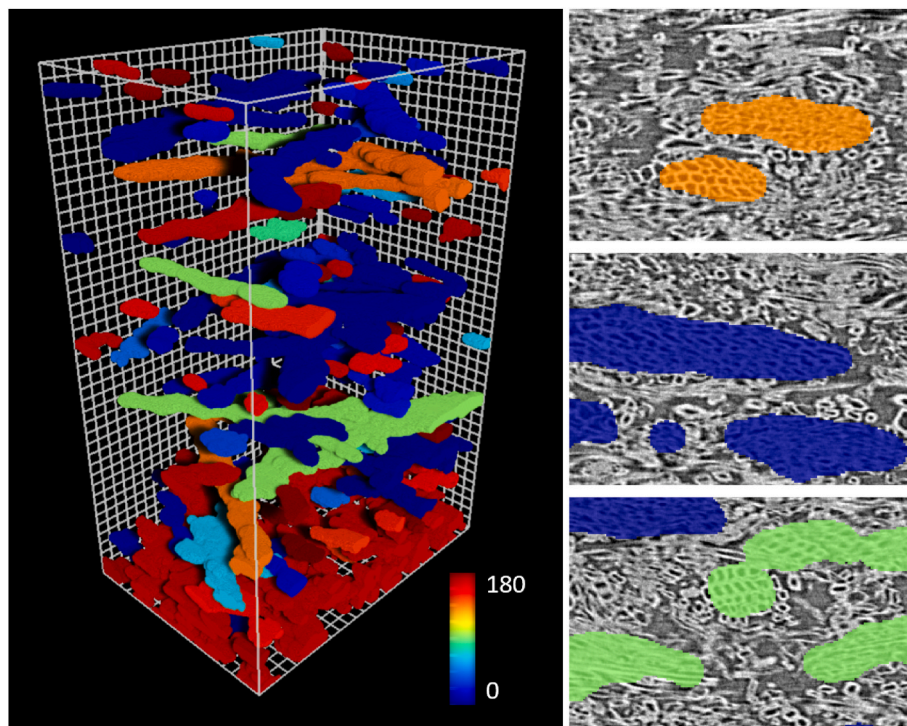


Fig. 7. Left panel: full volume segmentation by UNet3D, post-processed, and colored by horizontal orientation. Right panel: details of the post-processed segmentation overlaid on the MDF structure.

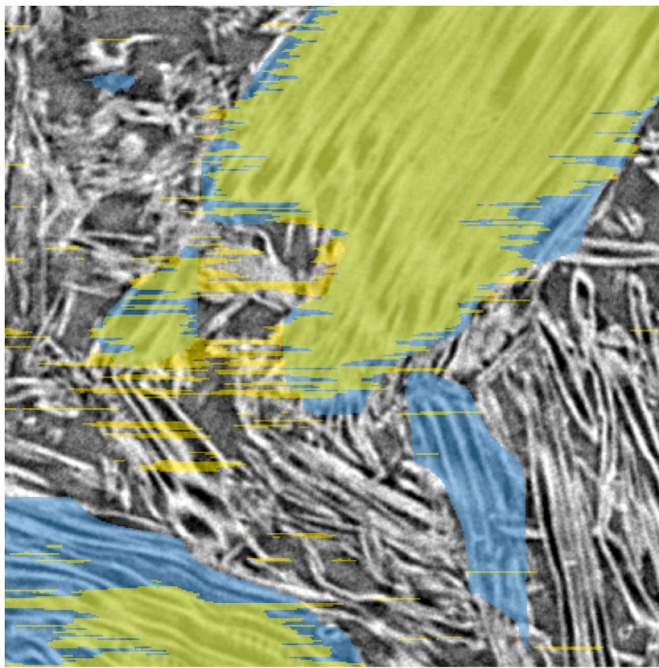


Fig. 8. While plausible on 2D slices (Fig. 5), a volume segmentation with UNet2D shows depth (bottom to top) variability too important to satisfactorily post-process. The groundtruth (GND₁) is shown in blue, while the UNet2D segmentation is shown in yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

algorithm is sooner bottlenecked by computational resources than a 2D one. In our case, this meant limiting the patch size to $128 \times 128 \times 5$ voxels³. We could thus question the performance of UNet3D using deeper patches, e.g. $128 \times 128 \times 128$ voxels³. Although aside from computational resources, this generates another limitation in the amount of training data that can be considered by the network before overfitting. Specifically in each of our manually-annotated ROIs, we could extract over 500 non-overlapping $128 \times 128 \times 5$ voxels³ patches, but only 20 non-overlapping $128 \times 128 \times 128$ voxels³ patches. Furthermore, if overfitting was a worry for UNet2D, it would be easy to generate more learning data: one would simply have to label extra 2D slices. In 3D, the same augmentation is much more demanding.

4.3.2. Learning process

In machine learning, overfitting is a risk. In our case, using a learning set to train CNNs that is included in ROI₁, which we partly used for validation, could raise concerns. The learning set is a small subset of the ROI₁ (~2% of it). Yet, these 2% are equally distributed within ROI₁, and we could argue that they fully represent its variability, as such not leading to overfitting. During preliminary investigations, a sensitivity analysis on porosity (i.e. analyzing the evolution of segmented air in a cubic ROI of side L, as a function of L) showed that a ROI's variability would stabilize beyond $\sim 100 \mu\text{m}^3$. ROI₁ is much larger (its smallest side is larger than $500 \mu\text{m}$), while although discontinuous, the learning set is much smaller in equivalent volume. However, we were concerned that the segmentation accuracy could be slightly overestimated on a ROI that was not completely novel to the network's weights. Therefore, we applied the same methodology to a second ROI of the same size as ROI₁, located elsewhere in the scanned volume. The results shown in Table 1 revealed that UNet3D's accuracy was lower on ROI₂ than on ROI₁. However, the prediction accuracy of morphological methods also differed by about the same amount between the two ROIs. Furthermore, the CNN accuracies remained significantly higher also in ROI₂. Therefore, we were able to prove that the high prediction accuracy achieved on ROI₁ does not correspond to an overfitting artifact, but to the real

prediction accuracy of the network. Differences inherent to the fiber bundles inside each ROI would explain the variability of the algorithms' prediction accuracies.

Besides, densifying a sparsely annotated dataset (as we have done in ROI₁) is common practice, and happens to be the first suggested use of UNet3D [17]. The same authors expected that by training on just two distinct datasets, the network could be generalized to a third. We do not claim generalization yet, as we currently focus on dynamic experiments and models involving only one sample. Generalization could be explored in future, although a non-generalizable network will also be a valuable asset for distinct future research with transfer learning.

Finally, the learning set was extracted along the single depth direction. Because fibers are oriented randomly in the horizontal plane, we assumed that the information present in vertical cross-sections would be qualitatively independent of rotation.

4.4. Fiber bundle morphometry

Only our best segmentation was post-processed, because specifics can widely vary from one image to another. Although standard, this process reflected on the subsequent morphometric analysis. Particularly, we notice in Table 2 that compared to the manually-annotated bundles, automatically annotated and post-processed bundles are less elongated. This is likely aggravated by the binary opening with a large, almost spherical structuring element. In Fig. 6, we can indeed see that a thinner bundle at the top of the ROI is processed into 3 smaller and rounder ones. Despite such undesirable effects, post-processing still positively impacts the bundle segmentation overall, as shown in Table 1. Furthermore, it is not yet expected to have perfectly segmented fiber bundles: in Fig. 7, we notice that parts of the bundles - particularly edges - are missed. Yet for all applications previously mentioned (modeling, dynamic experiments, etc.), having a first reliable estimate of the bundles location and density is already a major achievement. The current segmentation achieves this reliability for the first time at high density. Where previous segmentations methods at such densities resulted in segmented fiber bundles with incomplete shapes [8], the bundle shapes presented in this work resemble the realistic shapes traditional methods could segment at low density [4].

Moreover, we remark that while the raw UNet3D segmentation contains false positives (Fig. 6) and false-negatives alike, post-processing mainly removes false positives, by grinding away the bundle phase. The bundle density reported in Table 2 is therefore a slight underestimate of the real bundle density. Before post-processing, the bundle density in the entire volume (Fig. 7) was 16%, which was reduced to 12%. We expect the true value in between 12 and 16%, yet closer to the lower value. False positives are indeed noticeable, while the post-processed bundle density in the ROI was close to the groundtruth's (20 vs 21%). These proportions are in line with those previously reported by manual segmentation in different MDF panels [8].

Finally, the discrepancy in fiber bundle density between ROI₁ and the full volume can be logically explained: ROI₁ was selected to contain fiber bundles. It is therefore no surprise that it contains slightly more, and slightly larger fiber bundles than the entire volume would.

5. Conclusion

We investigated morphological and deep learning algorithms to automatically segment fiber bundles in MDF using μCT scanning. Using Matthew's correlation coefficient with two manually-annotated groundtruths, we showed that these algorithms had varying degrees of accuracy, the best one being UNet3D. We demonstrated that the complexity of this architecture is necessary. We then analyzed the morphometry of the identified fiber bundles. Results obtained on the groundtruths and on automatically segmented data showed satisfactory correspondence, although this could be improved depending on the needs of subsequent applications. Because of its unprecedented

performance, our method opens several paths for future. Moreover, because it was developed on an extremely challenging material at higher density than is generally investigated, we believe that this method could be applied to several other natural fiber composites with similar or better results.

Data availability

The data, code, and trained networks are available to download from 10.5281/zenodo.4898996.

Author statement

Pierre Kibleur: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization.

Jan Aelterman: Conceptualization, Methodology, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

Matthieu N. Boone: Conceptualization, Methodology, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Jan Van den Bulcke: Conceptualization, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Joris Van Acker: Conceptualization, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was funded by the Research Foundation - Flanders in the Strategic Basic Research Program MoCCha-CT (S003418N). The Special Research Fund of Ghent University is acknowledged for the support to the UGCT Centre of Expertise (BOF.EXP.2017.0007). The authors would like to thank Stijn Willen for sample preparation.

References

- [1] G. Churkina, A. Organschi, C.P.O. Reyer, A. Ruff, K. Vinke, Z. Liu, B.K. Reck, T. E. Graedel, H.J. Schellnhuber, Buildings as a global carbon sink, *Nat. Sustain.* 3 (4) (2020) 269–276.

- [2] J. Hildebrandt, N. Hagemann, D. Thrän, The contribution of wood-based construction materials for leveraging a low carbon building sector in Europe, *SCS* 34 (June) (2017) 405–418.
- [3] A. Miettinen, A. Ojala, L. Wikström, R. Joffe, B. Madsen, K. Nättinen, M. Kataja, Non-destructive automatic determination of aspect ratio and cross-sectional properties of fibres, *Comp. Part A* 77 (2015) 188–194.
- [4] J. Vigié, P. Latil, L. Orgéas, P. Dumont, S. Rolland du Roscoat, J.-F. Bloch, C. Marulier, O. Guiraud, Finding fibres and their contacts within 3D images of disordered fibrous media, *Compos. Sci. Technol.* 89 (2013) 202–210.
- [5] A. Madra, J. Adrien, P. Bretkopf, E. Maire, F. Trochu, A clustering method for analysis of morphology of short natural fibers in composites based on X-ray microtomography, *Comp. Part A* 102 (2017) 184–195.
- [6] T. Walther, H. Thoenen, Synchrotron X-ray microtomography and 3D image analysis of medium density fiberboard (MDF), *Holzforschung* 63 (5) (2009) 581–587.
- [7] T. Walther, K. Terzic, T. Donath, H. Meine, F. Beckmann, H. Thoenen, Microstructural analysis of lignocellulosic fiber networks, in: *Optics & Photonics*, vol. 6318, 2006, p. 631812.
- [8] J. Sliseris, H. Andrä, M. Kabel, O. Wirjadi, B. Dix, B. Plinke, Estimation of fiber orientation and fiber bundles of MDF, *Mater. Struct.* 49 (10) (2016) 4003–4012.
- [9] Y. Sinchuk, P. Kibleur, J. Aelterman, M.N. Boone, W. Van Paepegem, Variational and deep learning segmentation of very-low-contrast X-ray computed tomography images of carbon/epoxy woven composites, *Materials* 13 (4) (2020) 936.
- [10] J. Sliseris, H. Andrä, M. Kabel, B. Dix, B. Plinke, Virtual characterization of MDF fiber network, *Euro. J. of Wood and Wood Products* 75 (3) (2017) 397–407.
- [11] M. Dierick, D. Van Loo, B. Masschaele, J. Van den Bulcke, J. Van Acker, V. Cnudde, L. Van Hoorebeke, Recent micro-CT scanner developments at UGCT, *Nucl. Instrum. Methods Phys. Res. B* 324 (2014) 35–40.
- [12] M. Dierick, B. Masschaele, L. van Hoorebeke, Octopus, a fast and user-friendly tomographic reconstruction package developed in LabView®, *Meas. Sci. Technol.* 15 (7) (2004) 1366–1370.
- [13] O. Wirjadi, M. Godehardt, K. Schladitz, B. Wagner, A. Rack, M. Gurka, S. Nissle, A. Noll, Characterization of multilayer structures in fiber reinforced polymer employing synchrotron and laboratory X-ray CT, *Int. J. Mater. Res.* 105 (7) (2014) 645–654.
- [14] G. Standfest, S. Kranzer, A. Petutschnigg, M. Dunky, Determination of the microstructure of an adhesive-bonded medium density fiberboard (MDF) using 3-D sub-micrometer computer tomography, *J. Adhes. Sci. Technol.* 24 (8–10) (2010) 1501–1514.
- [15] S. van der Walt, J.L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J.D. Warner, N. Yager, E. Gouillart, T. Yu, scikit-image: image processing in Python, *PeerJ* 2 (July) (2014) e453.
- [16] O. Ronneberger, P. Fischer, T. Brox, U-Net, Convolutional networks for biomedical image segmentation, in: *Lecture Notes in Computer Science*, vol. 9351, 2015, pp. 234–241.
- [17] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: learning dense volumetric segmentation from sparse annotation, in: *Lecture Notes in Computer Science*, vol. 9901, 2016, pp. 424–432.
- [18] B. Matthews, Comparison of the predicted and observed secondary structure of T4 phage lysozyme, *BBA - Protein Structure* 405 (2) (1975) 442–451.
- [19] A. Wang, X. Yan, Z. Wei, ImagePy: an open-source, Python-based and platform-independent software package for bioimage analysis, *Bioinformatics* 34 (18) (2018) 3238–3240.
- [20] J. Bache-Wijig, P. Henden, Individual Fiber Segmentation of Three-Dimensional Microtomograms of Paper and Fiber-Reinforced Composite Materials (July).
- [21] A. Stojkovic, J. Aelterman, H. Luong, H. Van Parys, W. Philips, Highlights analysis system (HAnS) for low dynamic range to high dynamic range conversion of cinematic low dynamic range content, *IEEE Access* 9 (2021) 43938–43969.
- [22] I. Straumit, S.V. Lomov, M. Wevers, Quantification of the internal structure and automatic generation of voxel models of textile composites from X-ray computed tomography data, *Comp. Part A* 69 (2015) 150–158.