3D reconstruction of FIB tomography data using machine learning

Dissertation (cumulative) approved by the

Doctoral Degree Committee of
Hamburg University of Technology

in pursuit of the academic degree of

Doctor of Engineering (Dr.-Ing.)

written by
Trushal Sardhara

from
Ahmedabad, India

2024

Chair of
Examination Board: Prof. Dr. Roland C. Aydin

Reviewers: Prof. Dr.-Ing Christian J. Cyron
Dr.rer.nat. Gunnar Schaan

Date of Oral Examination: 06.08.2024

DOI: https://doi.org/10.15480/882.13216

To my family and friends, whose love and support have been my greatest strength.

Acknowledgements

This journey has been an incredible, challenging, and enriching experience, and I could not have completed it without the support of many amazing people.

First and foremost, I extend my deepest gratitude to my advisors, Prof. Dr.-Ing. Christian J. Cyron, Dr.-Ing. Martin Ritter, and Prof. Dr. Roland C. Aydin. Your guidance, patience, and unwavering encouragement have been invaluable. I have learned so much from each of you, and I am profoundly grateful for your mentorship. I also sincerely thank Dr. Nicolas Piché for his continuous support with the Dragonfly software. A special thank you to Dr.-Ing. Martin Ritter for inspiring me to pursue my PhD and for making electron microscopy so fascinating.

I am deeply thankful to my colleagues and friends at the Institute for Continuum and Material Mechanics, TUHH, and the electron microscopy unit (BeEM), TUHH. My gratitude extends to Dr. Alexander Shkurmanov, Dr. Tobias Krekeler, and Dr. Gunnar Schaan for the countless brainstorming sessions and shared laughs along the way. You made this journey not only possible but truly enjoyable.

To my family and friends, words cannot express my gratitude. To my parents, thank you for your unwavering belief in me and for supporting me in every way imaginable. To my friends, thank you for always being there with a laugh when I needed it most; your support made this journey so much more manageable and enjoyable.

Finally, thank you to everyone who has been part of this journey, whether through a word of encouragement, a helping hand, or simply your presence. This work is as much yours as it is mine.

Abstract

FIB-SEM tomography, which combines scanning electron microscope (SEM) images of consecutive sample cross-sections created with a focused ion beam (FIB), is a powerful tool for acquiring high-resolution images of nanomaterials. These images enable accurate 3D reconstruction of nanomaterials from 2D image slice stacks. However, the quality of these images, particularly those of hierarchical nanoporous materials, can be compromised by issues such as the *shine-through effect* and ambiguity in image intensities. The *shine-through effect* causes structures from posterior regions to become visible in the current plane, complicating the semantic segmentation of FIB-SEM images using traditional methods like thresholding or k-means clustering. These methods rely solely on intensity value numbers, which can be misleading due to the nonunique mapping between actual structure and image intensities. Furthermore, inconsistencies in the measured thicknesses of FIB-SEM image slices compared to the target thickness are challenging for accurate volumetric reconstruction. Machine learning-based solutions can help overcome these issues. In the research presented in this thesis, the *shine-through effect* is not treated as a problem but a potential solution, providing valuable additional information for an over-deterministic system. However, machine learning methods often require extensive training datasets and ground-truth values, which are difficult to obtain in electron microscopy. To address this, a data pipeline is implemented to generate synthetic data, including back scattered electron (BSE) grayscale and binarized paired images, using Monte Carlo simulations. This study demonstrates that the presented machine learning-based approach significantly improves the 3D reconstruction accuracy of micro- and nanostructures. Furthermore, the generated data pipeline can efficiently train additional machine learning models for various applications.

Table of contents

List of figures **xi**

Nomenclature **xiii**

1 Introduction **1**
1.1 FIB-SEM tomography 1
1.2 Objectives 5

2 Theory **7**
2.1 Basics of electron microscopy 7
2.2 Acquiring samples 9
2.2.1 Real Samples 9
2.2.2 Synthetic Samples 10
2.3 Supervised machine learning 11
2.4 Multimodal machine learning 13
2.5 Evaluation strategy 15
2.6 Evaluation metrics 18
2.6.1 Absolute error metrics 18
2.6.2 Isotropy based error metrics 19

3 First steps towards accurate 3D reconstruction using machine learning **21**
3.1 Background of the study 21
3.2 Discussion 23
3.3 Contribution to the field 24
3.4 Problems yet to be addressed 24
3.5 Article I - Training deep neural networks to reconstruct nanoporous structures from FIB tomography images using synthetic training data . 26

4 Role of slice thickness quantification in the 3D reconstruction 47
4.1 Background of the study . . . 47
4.2 Discussion . . . 49
4.3 Contribution to the field . . . 51
4.4 Problems yet to be addressed . . . 51
4.5 Article II - Role of slice thickness quantification in the 3D reconstruction of FIB tomography data of nanoporous materials . . . 52

5 Improving 3D reconstruction further using multimodal machine learning 67
5.1 Background of the study . . . 67
5.2 Discussion . . . 68
5.3 Contribution to the field . . . 70
5.4 Problems yet to be addressed . . . 71
5.5 Article III - Enhancing 3D reconstruction accuracy of multi-voltage FIB tomography images using multimodal machine learning . . . 72

6 Future work 87
6.1 Domain adaptation using machine learning . . . 87
6.2 BSE simulations using machine learning . . . 88

7 Conclusion 91

References 93

List of figures

1.1 A single slice of real HNPG showing **(A)** the *shine-through effect*, where the additional visible structure is shown using '?' (scale bar: 100 nm) and **(B)** simulated BSE image (2 kV) of gold cuboids generated using MCXray plugin to show ambiguity in intensities. 1) shows simulated intensities in xy-plane (I1-I6). (scale bar: 50 nm) 2) represents the arrangement of gold cuboids beneath the imaging plane (d1-d6). This experiment is inspired by [1]. 3

1.2 3D reconstruction solution pipeline for FIB-tomography data, including important steps such as generating virtual data, semantic segmentation method, and predicted 3D volumetric structure. 6

2.1 Trajectories of electrons inside bulk gold sample at accelerating voltage of 2 kV. Trajectories are simulated using Monte Carlo simulator software Casino [2]. 8

2.2 A single slice of back-scattered electron scanning microscope image of epoxy-infiltrated hierarchical nanoporous gold (HNPG) imaged at an accelerating voltage of 2kV. (scale bar: 200 nm) 9

2.3 A single image slice of synthetic FIB-tomography dataset was prepared using (A) virtual binary structure (LWM) and the corresponding (B) simulated BSE image. (scale bar: 200 nm) 10

2.4 Basic concept of convolution operation used commonly in convolutional neural networks (CNNs) 12

2.5 Basic block diagram representing unimodal machine learning model containing U-Net for semantic segmentation of FIB tomography data . 13

2.6 Basic block diagram setup representing multimodal machine learning model containing U-Net for semantic segmentation of FIB tomography data 13

2.7 Different multimodal ML architectures: **(A)** early fusion, **(B)** intermediate fusion, and **(C)** late fusion. Note: Each CNN model block (purple color) contains deep CNNs for semantic segmentation. 14
2.8 Basic concept of the two-point correlation function. Here, P1 and P2 are two points on the two different ligaments, and the functional value is 1 for this setup. 15
2.9 Basic concept of the lineal path function. Here, P1 and P2 are two points on the same ligaments, and the functional value is 1 for this setup. 16
2.10 3D view of virtual anisotropic structure, mimicking the *shine-through effect* in z-direction, used to test efficacy of anisotropic functions 16
2.11 Functional values of calculated anisotropy functions for virtual anisotropic structure in x-, y- and z-directions: **(A)** two-point correlation function, **(B)** lineal path function in yz-plane. 17

6.1 **(A)** Slice of a synthetic example microstructure generated using MCXray plugin and **(B)** a slice of real HNPG structure and **(C)** a slice of synthetic dataset after domain adaptation - scale bar: 200 nm 88
6.2 (A) A single image slice of digital microstructure generated using leveled wave method and corresponding simulated BSE image using (B) Monte Carlo method and (C) Machine learning method. (scale bar: 200 nm) . 89

Nomenclature

Acronyms / Abbreviations

2D	Two Dimensional
3D	Three Dimensional
AI	Artificial Intelligence
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DS	Dice Score
FIB	Focused Ion Beam
GAN	Generative Adversarial Network
HNPG	Hierarchical Nanoporous Gold
ICD	In-Column Detector
LMIS	Liquid Metal Ion Source
LPF	Lineal Path Function
LWM	Leveled Wave Method
MC	Monte Carlo
MGP	A fraction of total Misplaced Gold Pixels
ML	Machine Learning
MP	A fraction of total Misplaced Pixels

NPG Nanoporous Gold

SEM Scanning Electron Microscope

TLD Therough-Lens Detector

TPCF Two-point Correlation Function

Chapter 1

Introduction

Nanoporous materials are of significant interest across various fields, including biology, materials science, and chemistry, due to their potential for diverse applications. To fully harness their potential, a thorough understanding of their physical properties is crucial. Among the various investigative approaches, accurate 3D reconstruction is a powerful tool for elucidating key characteristics such as surface area, solid fraction, pore volume, ligament size, and connectivity. Hierarchical nanoporous materials, with ligament sizes ranging from 15 to 110 nm, necessitate high-resolution imaging techniques for detailed analysis and characterization. Consequently, advanced imaging techniques capable of capturing fine details at the nanoscale are indispensable for uncovering the complex structures and properties of these materials.

1.1 FIB-SEM tomography

Scanning electron microscopy (SEM) and focused ion beam (FIB) methods are combined in FIB-SEM tomography to obtain high-resolution cross-section images. SEM captures high-resolution images with resolutions of up to 1 nm in the xy-plane. FIB is used to mill away material layers within a range of 5 to 10 nm. This process generates a layer-by-layer image stack, with each image featuring voxel sizes of up to 1 nm in the SEM plane (xy) and 10 nm in the z-plane [3, 4]. This approach allows for acquiring detailed 3D information of micrometer-sized volumes with exceptional spatial resolution. FIB tomography is used in various fields, including materials science, nano-mechanics, biology, and geology, for imaging [5–7], and in combination with other analytical techniques for tasks such as 3D crystal orientation determination [8]. Several methodologies have been employed to reconstruct the three-dimensional microstructure of hierarchical nanoporous gold (HNPG) from FIB tomography data [9–11].

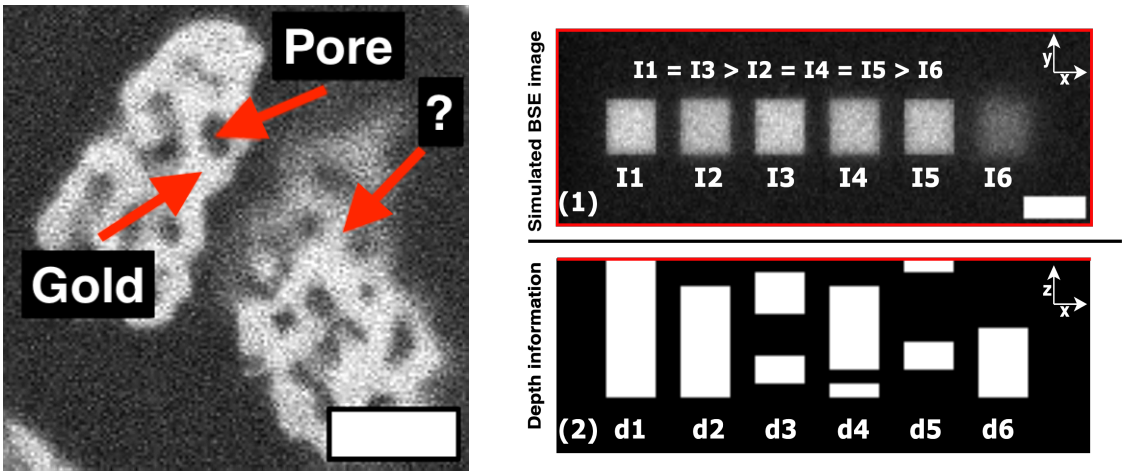
However, despite their high resolution, FIB-SEM images present several challenges:

- ***Shine-through effect*:** The *shine-through effect* is a phenomenon where structures from deeper layers become visible in the currently milled plane, introducing additional information that complicates the analysis of FIB-SEM images [12]. This effect can mask the actual features of interest, as illustrated in Figure 1.1 (A).
- **Ambiguity in image intensities:** Another challenge is the ambiguity in image intensities, which can result in a non-unique mapping between the actual structure and the intensity values in the image. As shown in Figure 1.1 (B), multiple voxels may have the same image intensity despite having different underlying structures at varying depths. This ambiguity presents a significant hurdle for analysis methods that rely solely on intensity values, underscoring the need to consider more complex image features [13].
- **Inconsistency in slice thicknesses:** Machine limitations often lead to discrepancies between the measured and target thickness in the z-plane [14]. Moreover, the measured thicknesses may vary between milled slices, making it challenging to assume a constant slice thickness or a zero difference between measured and target values. These inconsistencies can significantly affect the accuracy of 3D reconstruction efforts, highlighting the need for careful consideration of these factors in image analysis and processing.

The 3D reconstruction of nanoporous materials requires the classification of each voxel in an image as either material or pore, a process known as semantic segmentation. Traditional methods such as thresholding [15] or k-means clustering [16] are effective for bulk materials [17]; however, they struggle with nanomaterials due to the above mentioned *shine-through effect* and ambiguity in image intensities [13]. These methods often overlook complex image features, leading to inaccurate voxel predictions.

The *shine-through effect* is particularly problematic in hierarchical nanoporous materials. However, materials like nanoporous gold (NPG), which exhibit minimal *shine-through effect*, can be segmented using the k-means method choosing an appropriate number of clusters (k), as shown by [18].

To overcome the limitations of traditional segmentation methods, machine learning-based approaches, specifically convolutional neural networks (CNNs) [19], have been proposed as promising alternatives [20, 13]. CNN-based networks have gained prominence in various image analysis tasks within the machine learning domain. These



(A) (B)

Fig. 1.1 A single slice of real HNPG showing **(A)** the *shine-through effect*, where the additional visible structure is shown using '?' (scale bar: 100 nm) and **(B)** simulated BSE image (2 kV) of gold cuboids generated using MCXray plugin to show ambiguity in intensities. 1) shows simulated intensities in xy-plane (I1-I6). (scale bar: 50 nm) 2) represents the arrangement of gold cuboids beneath the imaging plane (d1-d6). This experiment is inspired by [1].

methods exploit the ability of CNNs to learn complex features from images, thereby improving the accuracy of semantic segmentation in challenging imaging scenarios. In the context of segmentation challenges in hierarchical nanoporous materials, the use of advanced machine learning-based methods, particularly CNNs, is crucial. Researchers have successfully used CNN-based methods to accurately segment hierarchical nanoporous gold structures, as demonstrated by [13].

However, using machine learning-based segmentation methods presents its challenges [21–23]. A primary hurdle is the scarcity of ground-truth values and insufficient training data [24], which are vital for training and validating machine learning models. In electron microscopy, acquiring adequate imaging data is time-consuming, costly, and requires expertise in sample preparation and microscope operation [25]. Furthermore, obtaining ground-truth annotations for electron microscopy images is particularly challenging due to the destructive nature of the FIB-SEM process, which limits the feasibility of traditional annotation methods.

To address the challenge of insufficient data, generating synthetic data is a viable solution [26]. Synthetic data, which consists of computer-generated samples, provides the benefit of scalability. This allows users, even those without expert knowledge, to generate unlimited samples, limited only by the computational resources at their

disposal. In electron microscopy, particularly in FIB-SEM imaging, Monte Carlo-based methods can *accurately* simulate backscattered electrons [27]. The MCXray plugin is used for this purpose in this study. This plugin is developed by [28] and integrated into the Dragonfly software [29] to generate synthetic FIB-SEM samples. Programs for simulating very simple geometries were developed over two decades ago, marking the inception of such simulations. Since then, these programs have undergone continuous refinement and enhancement to accommodate evolving needs and advancements [30–33, 2, 34]. Further details regarding this synthetic data generation pipeline can be found in Section 2.2.2. This pipeline yields a pair of binary and corresponding realistic simulated FIB-SEM images, which serve as valuable training data for machine learning algorithms. Subsequently, a technique known as transfer learning can be used to segment real FIB-SEM images using the optimized machine learning models [35, 36].

However, the simulated datasets often present a level of perfection that is not typically observed in electron microscopy images. The real electron microscopy images are known to exhibit two types of noise, as highlighted in a previous study [37]. The first type, Poisson noise, results from five different electron sources such as primary and secondary emissions, scintillator, photocathode, and photomultiplier [38], while the second type, Gaussian noise, originates from secondary emission and electronics [37]. Among these sources, the noise introduced by detection systems is often considered negligible [39, 40]. Due to this discrepancy, models trained on synthetic data may often perform poorly on real data [41]. When simulating BSE images using MCXray [28], Gaussian noise is naturally included [2].

Two methods could be adopted to overcome this difference in data distribution. The first method is extracting difference vectors manually using traditional image processing methods and correcting the simulations [13]. In this method, the remaining Poisson noise and additional Gaussian noise are added to effectively assess the impact of noise in synthetic FIB tomography data. The second method uses machine learning-based generative adversarial networks (GANs) [42]. Since paired simulated data (source domain) and real FIB tomography data (target domain) are not available, techniques like cycle consistency presented in [43] may be helpful.

A novel approach based on measuring the anisotropy of the predicted 3D volumetric structure is adopted to address the issue of needing many ground-truth annotations. Prior studies have shown that the structure of nanoporous gold exhibits self-similarity and, therefore, isotropy [44, 10]. This implies that if the predicted structure demonstrates isotropic behavior, the predicted 3D volume can be considered accurate. Specifically, two-point correlation [45] and lineal-path functions [46] are

calculated to quantify the isotropy of the predicted binary structures. This statistical analysis provides a quantitative assessment of the performance of the machine learning models. Further details regarding this method are explained in Section 2.6.2.

This study proposes a comprehensive solution pipeline based on these strategies, as illustrated in Figure 1.2. The efforts of this study are primarily focused on improving three key components: synthetic data generation, machine learning methods, and validation techniques. By refining these elements, it is aimed to achieve the most accurate 3D reconstruction of hierarchical nanoporous materials.

1.2 Objectives

In this thesis, the primary focus lies in acquiring accurate semantic segmentation of FIB tomography data. Accurate segmentation is a critical step towards obtaining reliable 3D reconstructions and, subsequently, conducting comprehensive analyses of the physical properties of nanostructures. To this end, each component depicted in Figure 1.2 is systematically implemented and improved through a series of iterative steps.

The first step toward enhancing the pipeline involves generating synthetic data that closely mimics real FIB tomography images. This step is essential, as realistic synthetic data is the foundation for training robust machine learning models and refining segmentation algorithms. Efforts of this study have yielded significant contributions in this domain, as evidenced by the work documented in Article I and the ongoing work presented briefly in Section 6.1.

Subsequently, this study delves into developing machine learning models tailored for the semantic segmentation of FIB tomography images. This pivotal step entails creating and optimizing convolutional neural network architectures capable of accurately segmenting material and pore regions within the images. Through rigorous experimentation and refinement, documented in Article I, Article III, and the ongoing work (Section 6.1), substantial progress has been made in advancing the state-of-the-art semantic segmentation methodologies for FIB tomography data.

Equally essential to the success of the proposed pipeline is the establishment of robust evaluation strategies to quantitatively and qualitatively assess the performance of proposed segmentation models. By rigorously evaluating these methods against ground-truth data and implementing comprehensive metrics, the reliability and accuracy of the segmentation results are ensured. Main contributions in this domain are documented in Article I and Article II.

Overall, this iterative approach to pipeline improvement has yielded promising outcomes, paving the way for more accurate and reliable FIB tomography data semantic segmentation. The solutions and methodologies developed in this study can be readily adapted to other microstructures and even extended to diverse domains, showcasing the versatility and applicability of our approach.

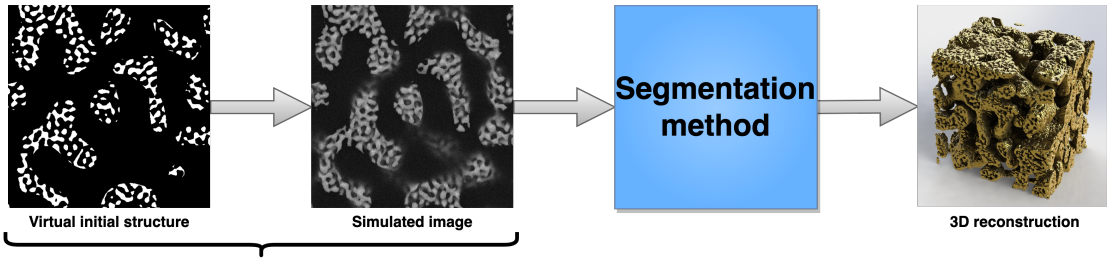


Fig. 1.2 3D reconstruction solution pipeline for FIB-tomography data, including important steps such as generating virtual data, semantic segmentation method, and predicted 3D volumetric structure.

Chapter 2

Theory

2.1 Basics of electron microscopy

The scanning electron microscope traces back to the pioneering work of [47, 48] and subsequent advancements by [49] and Prof. Sir Charles Oatley and his team at the University of Cambridge [50]. Operating on the principle of emitting a finely focused beam of energetic electrons, typically ranging from 0.1 to 30 keV, the SEM allows the visualization of specimens at the microscopic scale [40]. This electron beam traverses various optical components, scanning the specimen in a raster pattern. Upon interaction with the specimen, the electron beam generates, among others, backscattered electrons (BSEs) and secondary electrons (SEs), which are then captured by specialized detectors.

This study uses FEI Helios Nanolab G3, which combines an SEM and a focused ion beam (FIB) column. In such a combined system, sometimes called "dual-beam", the FIB column is mounted at an angle of 52° with the vertical electron (SEM) column. The electron and ion beams coincide at a working distance of 4 mm from the SEM pole piece [51]. The FIB column contains a liquid metal ion source (LMIS), which can provide ions in the diameter range of around 5 nm. Gallium (Ga) is used as LMIS for this FIB system. Gallium exhibits useful properties, such as low melting point, low volatility, low surface free energy, low vapor pressure, and excellent electrical, mechanical, and vacuum properties [52]. Extracted Ga ions from LMIS are accelerated using a potential down the ion column. When these accelerated Ga+ ions come in contact with the surface of the sample, sputtered atoms, secondary electrons, and secondary ions are generated. Such a gallium beam is used to create trenches and cross-sections by milling sample material away. Consecutive milling of cross-sections before imaging is the principle of the FIB tomography process. It is a destructive method. However, various artifacts are observed while performing FIB milling, including charging effects, material

redeposition, and curtaining [53]. In order to minimize the effect of the materials' redeposition, trenches around the region of interest should be milled. In the case of FIB tomography, curtaining was typically observed for an initial 10-30 slices, which were removed before further sample analysis.

For imaging, the focus of this study predominantly lies in using the potential of backscattered electrons. These electrons, having undergone scattering within the specimen, are directed back toward the surface from which they initially entered. Retaining a substantial portion of their incident energy, BSEs serve as an essential imaging signal in SEM, providing deep insights into various specimen properties such as composition, topography, mass thickness, and crystallography.

A fundamental parameter for quantifying BSEs is the *backscattered electron coefficient*, denoted as η. This coefficient represents the ratio of the number of backscattered electrons (NBSE) to the total number of beam electrons that penetrate the specimen (NB) [40]; Mathematically expressed as:

$$\eta = \frac{\text{NBSE}}{\text{NB}} \tag{2.1}$$

Figure 2.1 shows the trajectory of electrons generated using Monte Carlo simulations of the bulk gold sample within an SEM setup. Trajectories shown in red are backscattering electrons (NBSE). Trajectories shown in blue are primary beam electrons that did not escape the sample (NB-NBSE).

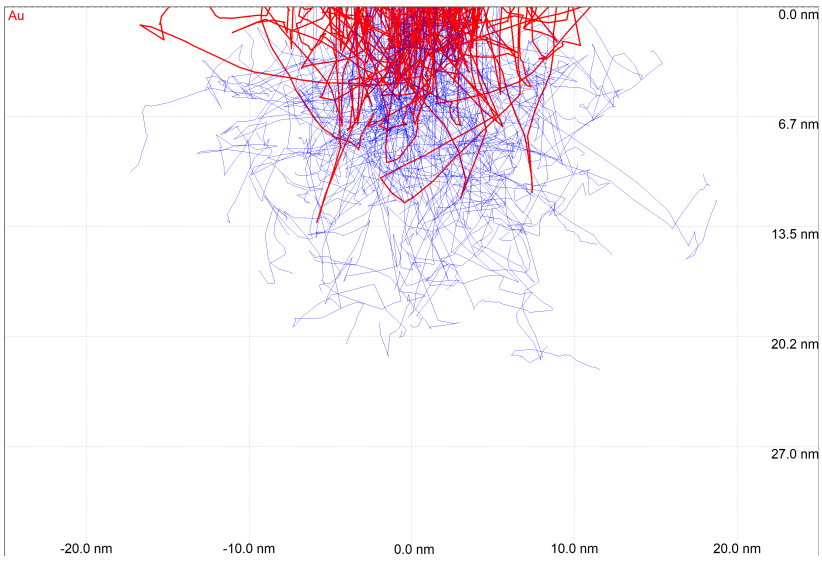


Fig. 2.1 Trajectories of electrons inside bulk gold sample at accelerating voltage of 2 kV. Trajectories are simulated using Monte Carlo simulator software Casino [2].

2.2 Acquiring samples

In this research, hierarchical nanoporous gold (HNPG) emerges as a pivotal material due to its controllable pore morphology and suitability for investigating mechanical and surface properties [54]. HNPG has received significant attention in the scientific community [55], underscoring its relevance for diverse applications. Real HNPG FIB tomography images, alongside computer-generated samples, constitute the primary datasets used to train different machine learning models. Detailed procedures for acquiring both real and synthetic samples are elaborated on in the articles referenced within this thesis.

2.2.1 Real Samples

HNPG is synthesized using a dealloying-coarsening-dealloying method, as outlined in a previous work [56]. This synthesis process yields HNPG with a uniform random network structure and distinct ligament sizes (15 nm and 110 nm). For the investigations in this study, HNPG samples are infiltrated with epoxy resin to mitigate the *shine-through effect* during FIB-SEM tomography [57]. A representative slice from a real HNPG sample FIB tomography is depicted in Figure 2.2. It is essential to ensure that the slice thickness remains significantly smaller than the pore size ($<$ 10 nm) for accurate characterization. Specific experimental parameters and procedures may vary between articles and should be referenced accordingly.

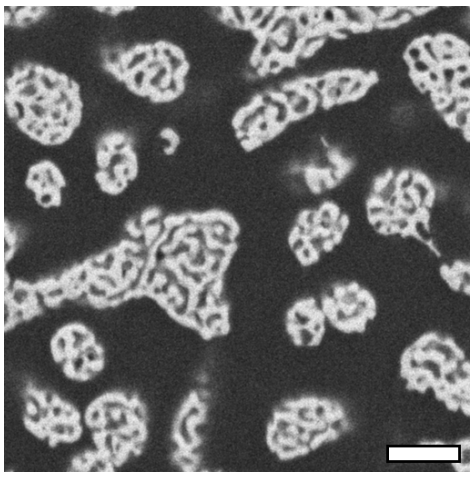
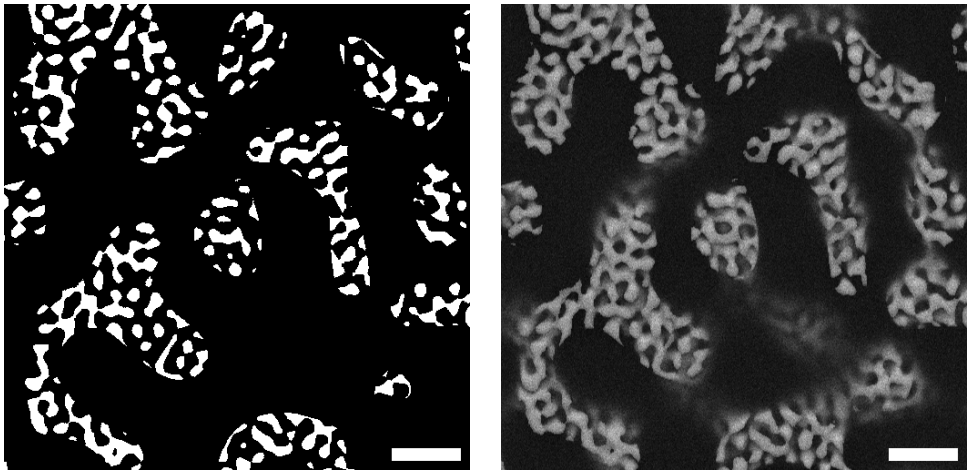


Fig. 2.2 A single slice of back-scattered electron scanning microscope image of epoxy-infiltrated hierarchical nanoporous gold (HNPG) imaged at an accelerating voltage of 2kV. (scale bar: 200 nm)

2.2.2 Synthetic Samples

Synthetic samples are prepared in two stages. Initially, binary structures are generated using the leveled wave method proposed in the previous study [56]. The leveled-wave model [58] forms the foundation for each level within the virtual HNPG hierarchy. It begins by superimposing waves of uniform wavelength but with random wave vector directions to create a concentration field. These waves, all with equal magnitude, determine the ligament size in the virtual nanoporous network. Next, a level cut is chosen to binarize the structure into pore or solid phases, generating virtual nanoporous microstructures by eliminating the pore phase. Different level cuts can be utilized to attain the desired solid fraction. These binary structures serve as ground-truth values for machine learning algorithms and as initial structures for Monte Carlo simulations. To simulate realistic FIB tomography data, the binary structures are processed using the MCXray plugin [28], integrated into the Dragonfly software [29].



(A) (B)

Fig. 2.3 A single image slice of synthetic FIB-tomography dataset was prepared using (A) virtual binary structure (LWM) and the corresponding (B) simulated BSE image. (scale bar: 200 nm)

Backscattered electron signals are mainly used due to their direct correlation with atomic number (Z) and accelerating voltages (E), as dictated by established formulas [59].

$$\eta(Z,E) = E^{m(z)*C(Z)} \tag{2.2}$$

Where

$$m(z) = 0.1382 - \frac{0.9211}{Z^{0.5}} \tag{2.3}$$

$$C(Z) = 0.1904 - 0.2236 \cdot ln(Z) + 0.1292 \cdot (ln(Z))^2 - 0.01491 \cdot (ln(Z))^3 \tag{2.4}$$

A comparison between a single slice of a binary structure and its corresponding simulated slice is presented in Figure 2.3. Additional parameters such as accelerating voltages, sample stage tilt angle, and working distance are calibrated per individual study requirements. This synthetic data pipeline, coupled with paired ground-truth values in the form of binary structures, facilitates the seamless adoption of supervised machine learning methodologies for the 3D reconstruction of microstructures.

2.3 Supervised machine learning

In machine learning, training a neural network with paired input and output images, known as ground truth, forms the basis of many exciting models. Neural networks, inspired by the structure of the human brain, comprise input, output, and hidden layers [60, 61]. These layers consist of multiple interconnected neurons, with each connection governed by a learnable parameter known as a weight. When a neural network contains many neurons and hidden layers, it is categorized as a deep learning network [62]. During supervised training, the network aims to approximate a function:

$$\hat{y} = f(x; \theta) \tag{2.5}$$

Here, θ represents the learnable parameters, and x denotes the input values. A loss function is computed by comparing the predicted output $\hat{y}$ with the ground-truth values. The objective is to minimize this loss by updating all network parameters. This supervised learning strategy [63] is commonly employed in tasks such as classification or regression.

In contrast, unsupervised learning methods [64] come into play when paired data is unavailable. This approach is commonly used to uncover latent features within the data. However, the focus of this work primarily revolves around supervised learning.

Given that images contain a large feature space, conventional feed-forward networks may not yield optimal results. For image-related tasks, specialized neural networks known as convolutional neural networks (CNNs) prove to be highly effective [19]. These networks use a mathematical operation called convolution, expressed as:

$$s(t) = (x * w)(t) \tag{2.6}$$

Here, x represents the input, w is the learnable parameters or kernel, and s denotes the output feature map. Figure 2.4 shows a basic convolution operation. For a more comprehensive understanding of CNNs and their various components, refer to [65].

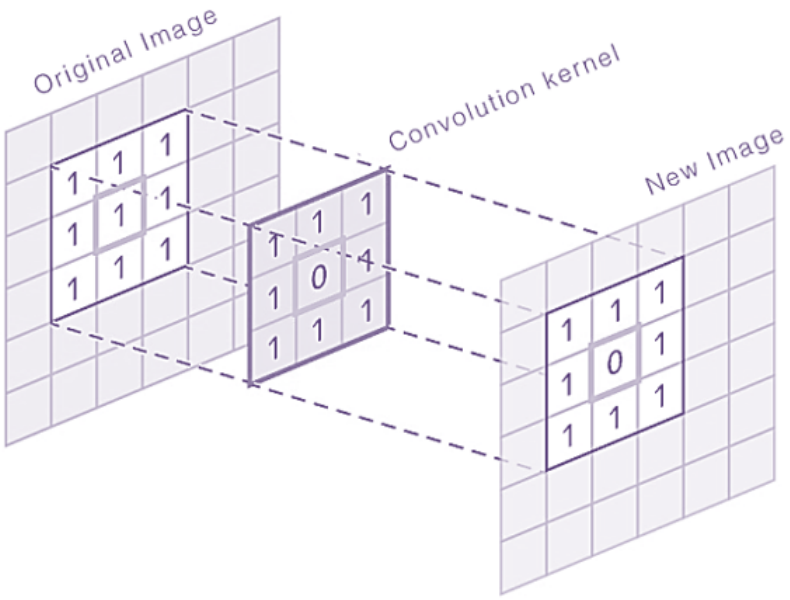


Fig. 2.4 Basic concept of convolution operation used commonly in convolutional neural networks (CNNs)

This research employs CNNs primarily for two tasks: 1) Image semantic segmentation and 2) Image inpainting.

Image segmentation is a classification task wherein each image voxel is categorized as either solid or pore phase. Consequently, CNNs produce a segmentation map as an output, with values corresponding to the available categories for classification. For this purpose, this work uses an encoder-decoder architecture, such as the U-Net, a CNN-based model [66, 67], with custom modifications including varying depths of the encoder-decoder, integration of residual connections, and padding strategies.

On the other hand, image inpainting involves a generative approach aimed at filling in missing pixels within images. This task also uses CNN-based networks, often employing a pair of networks engaged in adversarial competition. Known as generative adversarial networks (GANs) [42], these architectures are prevalent in various generative tasks, including image inpainting [68], domain adaptation [69], and the generation of entirely novel data points.

2.4 Multimodal machine learning

In machine learning, *modality* refers to the various types of data derived from different sensors or sources humans can perceive or interact with. Examples of modalities encompass images, language, audio signals, and more. When machine learning models use data from a single modality, they are termed unimodal machine learning models. An illustration of such a model is a single U-net trained exclusively on synthetic samples simulated at an accelerating voltage of 1kV, aimed at segmenting images (see Figure 2.5).

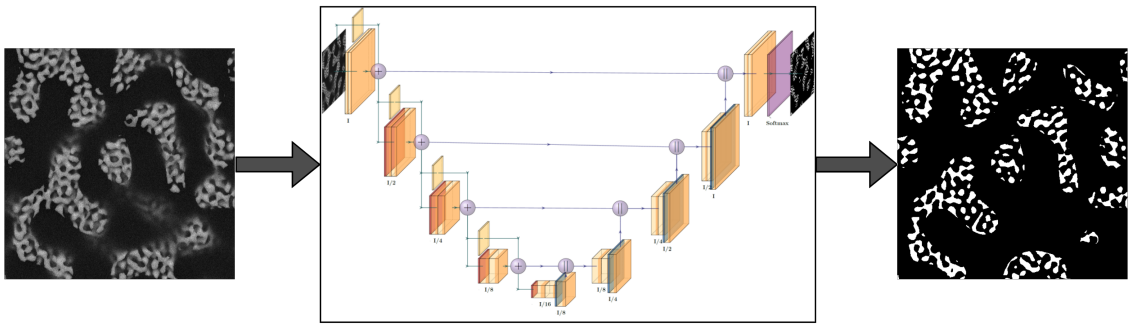


Fig. 2.5 Basic block diagram representing unimodal machine learning model containing U-Net for semantic segmentation of FIB tomography data

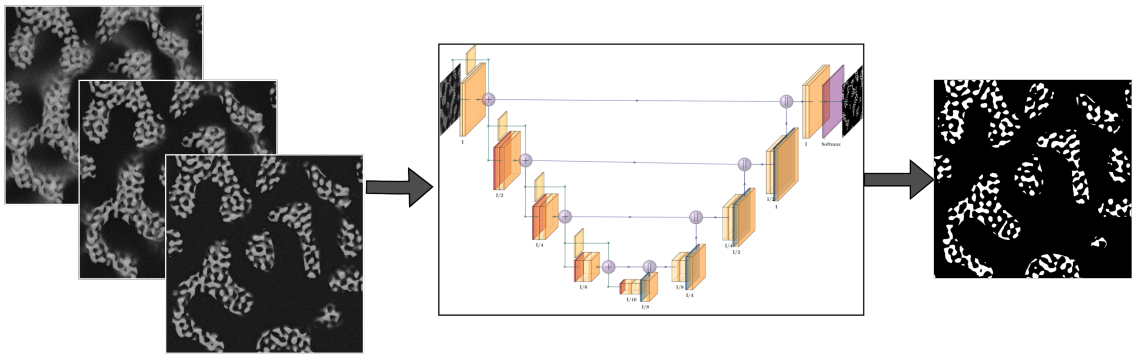


Fig. 2.6 Basic block diagram setup representing multimodal machine learning model containing U-Net for semantic segmentation of FIB tomography data

While unimodal models are effective, supplementing them with additional information from diverse sources can often yield superior results. Models integrating data from multiple sources are called multimodal machine learning models [70]. The decision to incorporate extra information is typically determined through various assessment

methods summarized in [71]. Moreover, several alignment steps are necessary to synchronize data from different modalities before training multimodal machine learning models.

This research categorizes segmentation models trained on synthetic data simulated at accelerating voltages of 1kV, 2kV, and 4kV concurrently as multimodal segmentation models (see Figure 2.6). Figure 2.7 illustrates various types of multimodal segmentation models, with detailed explanations in Article III.

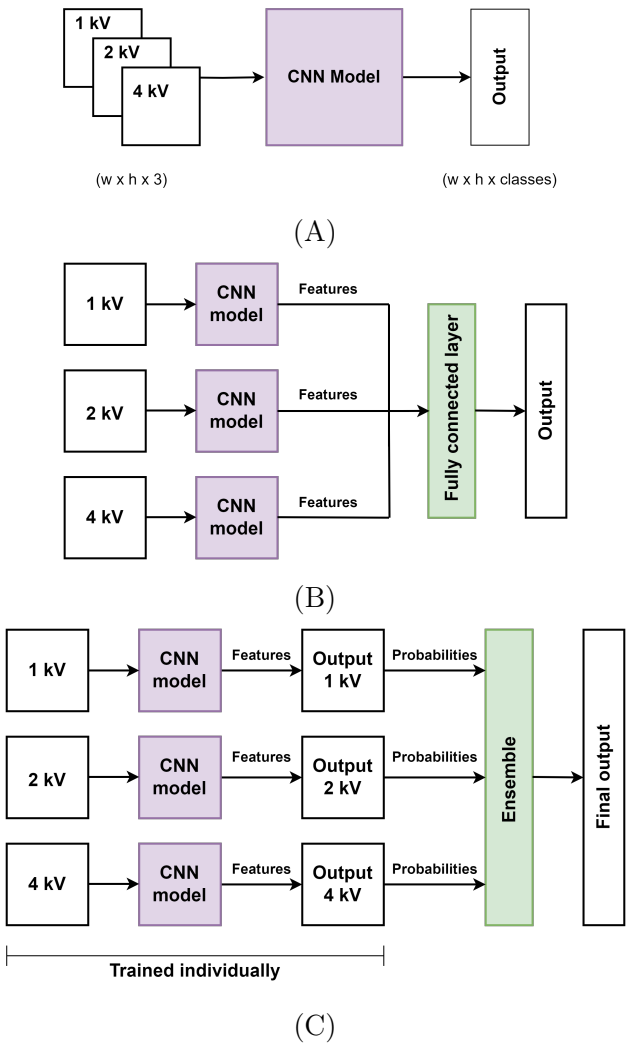


Fig. 2.7 Different multimodal ML architectures: **(A)** early fusion, **(B)** intermediate fusion, and **(C)** late fusion. Note: Each CNN model block (purple color) contains deep CNNs for semantic segmentation.

2.5 Evaluation strategy

In this study, two distinct evaluation strategies are used to assess the performance of our segmentation models.

The first strategy contains absolute error metrics when ground-truth values are available for the dataset. This approach is often more accurate than other metrics discussed within this work.

The second strategy uses the concept of isotropy and is used when ground-truth data is unavailable. This strategy is tailored for isotropic volumes, particularly relevant to HNPG structures. It is well-established that the volumetric structure of HNPG exhibits self-similarity, indicative of isotropy [10]. Two functions are calculated to quantify isotropy: the two-point correlation function and the lineal-path function.

The two-point correlation function measures the probability of encountering two points within the same material phase, regardless of whether they reside on the same ligament (see Figure 2.8). This function not only captures local correlations but also evaluates long-range correlations. The formula for calculating the two-point correlation function is as follows [45]:

$$S_2(k) = \frac{1}{2k+1} + \Sigma_{l=0}^{2k}[\frac{1}{N_2} \sum_{\substack{1\leq i\leq n \\ 1\leq j\leq n}} f_{ij} \cdot f_{i+k} \cdot f_{j+\frac{\pi l}{4k}}] \tag{2.7}$$

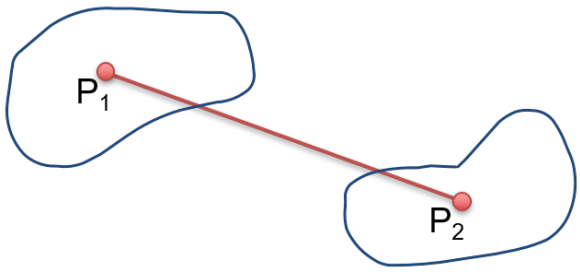


Fig. 2.8 Basic concept of the two-point correlation function. Here, P1 and P2 are two points on the two different ligaments, and the functional value is 1 for this setup.

The second function used for evaluating isotropy is the lineal path function, which accurately measures the probability of encountering two points on the same ligament within the same material phase (see Figure 2.9). Unlike the two-point correlation function, the lineal path function focuses solely on local correlations, particularly for short-range points.

The lineal path function is calculated using the following formula [46]:

$$L(k) = \frac{\phi_i \int_0^\infty (y-k)\, p^i(y)\, \Theta(y-k) dy}{\int_0^\infty y\, p^i(y)\, dy} \tag{2.8}$$

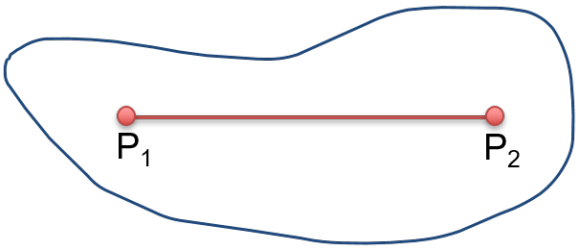


Fig. 2.9 Basic concept of the lineal path function. Here, P1 and P2 are two points on the same ligaments, and the functional value is 1 for this setup.

If both of these functions do not change over all the directions for the structure, then the structure can be assumed to be isotropic. However, it has been noted that relying on only one function is insufficient to determine a structure's isotropy accurately [45]. To this end, short- and long-range correlations should be examined before drawing any conclusions about the isotropy of the structures.

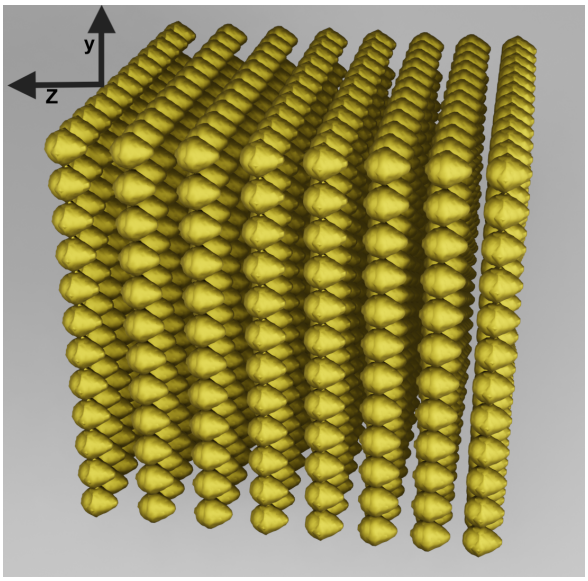
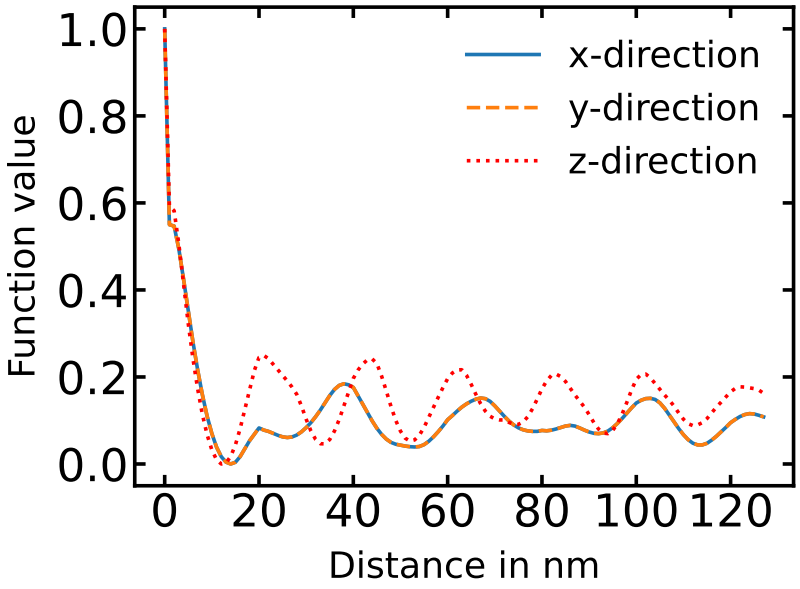


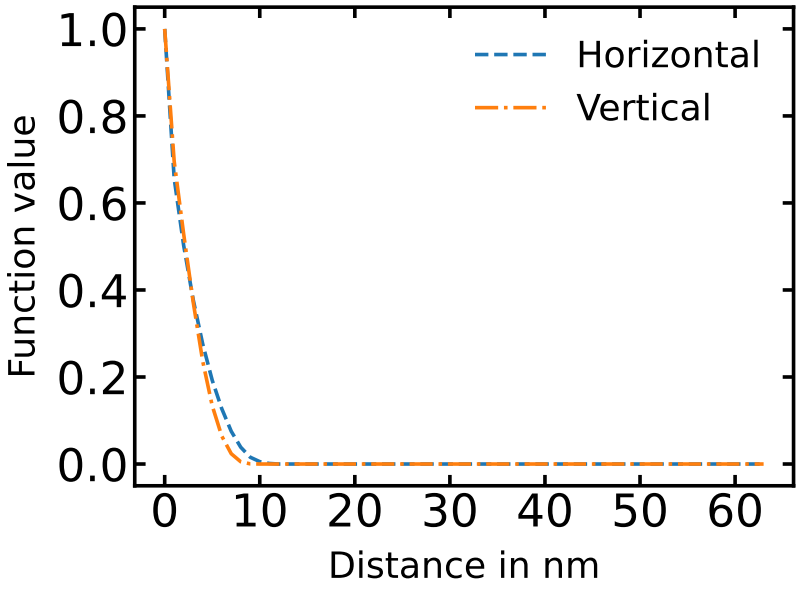
Fig. 2.10 3D view of virtual anisotropic structure, mimicking the *shine-through effect* in z-direction, used to test efficacy of anisotropic functions

This study tested these functions on virtually generated non-isotropic structures 2.10. To achieve this, virtual structures were created using non-isotropic spheres

aligned in the z-direction to emulate the *shine-through effect*. These non-isotropic spheres were positioned on a grid within a fully solid volume, with uniform intervals in the xy-plane but incrementally spaced intervals along the z-direction to introduce additional anisotropy. The resulting structure is depicted in Figure 2.10.



(A)



(B)

Fig. 2.11 Functional values of calculated anisotropy functions for virtual anisotropic structure in x-, y- and z-directions: **(A)** two-point correlation function, **(B)** lineal path function in yz-plane.

The main objective of creating such a virtual structure was to demonstrate that the functions remain consistent along the x- and y-directions, indicating isotropy in the xy-plane while displaying significant differences in the z-direction where intentional anisotropy was introduced. To achieve this, both of these functions were calculated in the x-, y-, and z-directions, as illustrated in Figure 2.11.

The plotted data shows that both functions exhibit identical values in the x- and y-directions, confirming isotropy in these planes. However, the functions in the z-direction display notable discrepancies, reflecting the intentional introduction of anisotropy along this axis. This behavior highlights the capability of these functions to detect anisotropy in a specified direction accurately.

2.6 Evaluation metrics

Based on the availability of ground-truth data, the following evaluation metrics are used in this work.

2.6.1 Absolute error metrics

All the metrics calculated by comparing predictions to the actual values are absolute, and the performance of the machine learning models can be measured very accurately using these metrics. A total of three absolute error metrics are calculated when ground-truth values are available. First, the fraction of total misplaced pixels (MP) is calculated using the formula:

$$MP = \left(1 - \frac{TP + TN}{TP + FP + FN + TN}\right) \times 100 \tag{2.9}$$

where TP denotes the number of true positives, TN denotes the number of true negatives, FP denotes the number of false positives, and FN denotes the number of false negatives. MP is the same as the absolute accuracy of the prediction.

The second metric is called the fraction of total misplaced gold pixels (MGP), is calculated using the formula:

$$MGP = \left(1 - \frac{TP}{TP + FN}\right) \times 100 \tag{2.10}$$

In other words, this metric is the same as the sensitivity, which means a probability of predicting positive samples as positive. However, this metric should be used with special care since there is no punishment in the term for predictions of false positives.

The third metric, mean Dice score (MDS) [72], is based on calculating overlapping regions and is probably most suitable for imbalanced datasets like in our case. The Dice score for a single category is calculated using the formula:

$$DS = \frac{2\,TP}{2\,TP + FN + FP} \tag{2.11}$$

This score takes values between 0 and 1, 1 being the best possible score indicating completely overlapping regions of a category between ground truth and predicted images. Dice scores for each category are measured using this formula and then averaged to get a final mean Dice score. This final metric also calculates values between 0 and 1.

2.6.2 Isotropy based error metrics

In the absence of ground-truth values, measuring the performance of machine learning models becomes challenging for nanomaterials, such as HNPG. Nevertheless, evaluating model performance on real data is essential as it indicates the effectiveness of knowledge transfer from synthetic to real data and vice versa. This study addresses this challenge by implementing evaluation metrics considering the inherent isotropy assumptions for HNPG and factor in anisotropy caused by the *shine-through effect* in the z-direction.

As discussed earlier, evaluating isotropy or, conversely, anisotropy is crucial. A two-point correlation function is calculated in three spatial directions (x, y, and z) to quantify the anisotropy. Denoted as f^x, f^y, and f^z, these functions are discretized uniformly to obtain function values f_i^x, f_i^y, and f_i^z at n data points (i ranging from 1 to n). The final error metric based on TPCF involves calculating the L_2 differences between the TPCF in the x- and z-directions and the y- and z-directions. These differences are then averaged to derive the overall anisotropy metric:

$$e_{L_2}^{TPCF} = \frac{1}{2}\left(\frac{2 \times \sqrt{\sum_{i=1}^{n}(f_i^x - f_i^z)^2}}{\sqrt{\sum_{i=1}^{n}(f_i^x)^2} + \sqrt{\sum_{i=1}^{n}(f_i^z)^2}} + \frac{2 \times \sqrt{\sum_{i=1}^{n}(f_i^y - f_i^z)^2}}{\sqrt{\sum_{i=1}^{n}(f_i^y)^2} + \sqrt{\sum_{i=1}^{n}(f_i^z)^2}}\right) \tag{2.12}$$

This metric provides a quantitative measure of the deviation from isotropy, with higher values indicating greater anisotropy in the structure. Similarly, the second metric is calculated based on the lineal path function. This local range-based function provides additional information about the anisotropy of the microstructure. The final anisotropy metric $e_{L_2}^{LPF}$ is calculated analogously to $e_{L_2}^{TPCF}$.

The third metric, $e^D_{L_2}$, is similar and based on the ligament diameters. These average ligament diameters are derived from the underlying area of probability distributions of LPF functions in all directions. The final metric then calculates the average difference of such ligament diameters in the x- and z-directions and y- and z-directions:

$$e^D_{L_2} = \frac{1}{2}\left(\sqrt{\frac{(D_{xz} - D_{xy})^2}{D^2_{xy}}} + \sqrt{\frac{(D_{yz} - D_{xy})^2}{D^2_{xy}}}\right) \tag{2.13}$$

where D_{ij} represents the calculated average diameter of ligaments in the ij plane, with $i, j \in \{x, y, z\}$. These errors are zero for a completely isotropic structure. Thus, zero errors indicate the best 3D reconstruction performance of the machine learning models on the real microstructures. On the other hand, higher errors in these functions indicate the presence of anisotropy in real structures, suggesting possible errors in the segmentation.

Chapter 3

First steps towards accurate 3D reconstruction using machine learning

3.1 Background of the study

In our study, detailed in Article I, the primary objective was to compare various segmentation methodologies. Specifically, we sought to evaluate the efficacy of classical approaches, such as Otsu's thresholding method and k-means clustering, alongside a machine learning-based approach in segmenting real hierarchical nanoporous gold.

Otsu's thresholding method calculates final predictions using a threshold value derived from the intensity histogram of the images. In contrast, k-means clustering requires the assumption of a predefined value for k, typically set to 3 in our case. The additional third cluster is designated for artifacts resulting from the *shine-through effect*, distinct from the two clusters representing material and pore phases, respectively. Notably, these traditional methods offer the advantage of being deployable without the need for extensive training data. They calculate predictions using mathematical formulas, making them interpretable and readily reproducible.

However, a fundamental limitation of such classical methods lies in their reliance solely on intensity values for prediction. They neglect more complex features that are important in image-based applications. Relying solely on intensity values requires accurate and unique mappings between various categories to ensure accurate segmentation. The presence of the *shine-through effect* and inherent ambiguity in intensity values often prevent traditional methods from making accurate predictions. Consequently, exploring additional complex features beyond intensity values presents a way to achieve more accurate segmentation predictions.

Given these considerations, our study highlights the need for methods that use diverse image features beyond intensity values alone. By incorporating such features, segmentation algorithms can potentially enhance accuracy and robustness, particularly in scenarios where traditional methods fall short. This approach advances segmentation methods in image processing, especially in complex materials like HNPG in electron microscopy.

In such conditions, machine learning-based methods emerge as viable solutions capable of extracting complex features from images and accurately learning the mapping between grayscale images and segmentation maps. However, implementing CNNs presents several challenges. Major among these challenges is the requirement for extensive training data. Moreover, if the CNNs are to be trained in a supervised manner, ground-truth values (true segmentation values) are also required. However, obtaining such data in electron microscopy poses considerable difficulties due to cost, time, and expertise constraints. Furthermore, without ground-truth values, evaluating the performance of segmentation methods becomes an entirely different challenge. Therefore, in addition to comparing machine learning-based methods with traditional segmentation methods, we also proposed a strategy to generate synthetic data with ground-truth values and a method to quantitatively evaluate isotropic materials without any ground-truth values.

We proposed using synthetic data to address the scarcity of paired training data. All machine learning networks were trained on these synthetic training data, and the optimized networks were used to calculate 3D reconstructions of real NPG and HNPG microstructures. The initial step in synthetic data generation involved generating virtual initial structures that closely resemble real microstructures. To achieve this, we implemented three methods for generating binarized virtual initial structures and simulated them using the MCXray plugin incorporated in Dragonfly software [29]. Subsequently, machine learning networks trained using each dataset were evaluated on the same test data. Finally, we determined that virtual initial structures generated using the leveled wave method, closely resembling real HNPG microstructures, performed best as virtual initial structures for synthetic data.

Moreover, in Article I, we proposed a method to evaluate the performance of machine learning models for isotropic data in the absence of ground-truth values. Our method is not limited to HNPG but extends to all isotropic structures. As mentioned in Section 2.6.2, we computed all three error metrics based on the two-point correlation and lineal path functions in the x-, y-, and z-directions and calculated the respective L_2 errors.

3.2 Discussion

We thoroughly validated our trained models, assessing their performance across various datasets. This process included segmenting one synthetic LWM dataset and four real FIB-SEM images (refer to Table 5 from Article I). The results comparing different segmentation methods are presented in Tables 6, 7, and 8 from Article I.

A key finding from these tables is the significantly reduced $e_{L_2}^{TPCF}$, $e_{L_2}^{LPF}$ and $e_{L_2}^{D}$, achieved by our newly introduced deep learning-based segmentation methods. In contrast, traditional methods like Otsu's or k-means clustering (with k = 3) exhibit higher L_2 errors. This indicates that our novel deep learning approaches effectively mitigate the *shine-through effect*, leading to geometries that closely resemble isotropic microstructures.

We examined the prediction outcomes for datasets captured using a through-lens detector (TLD) and again using the more surface-sensitive in-column detector (ICD). Comparing the segmentation results between ICD and TLD, we observed a higher solid fraction even for ICD with traditional methods, contrasting with our machine learning-based methods for the TLD dataset. This disparity underscores the challenge of traditional methods in effectively mitigating artifacts. Moreover, our deep learning-based segmentation methods consistently yielded lower MP values than Otsu's and k-means clustering, reflecting their superior accuracy in classifying pixels across TLD and ICD imaging data.

In line with the principles of sensor fusion [73], this performance disparity further underscores the superiority of our deep learning-based segmentation methods. A detailed analysis reveals a notable occurrence of misplaced pixels for both real datasets when employing k-means clustering. This discrepancy can be attributed to the selection of k = 3 for k-means clustering, designed to accommodate the presence of three distinct clusters—solid phase, pore phase, and artifacts (i.e., pixels attributed to the *shine-through effect*). While this assumption holds merit for TLD images due to their susceptibility to the *shine-through effect*, ICD images exhibit reduced sensitivity to this issue. Consequently, using k = 3 may no longer be suitable, resulting in suboptimal segmentation outcomes for ICD datasets.

Moreover, our study compared various CNN-based segmentation methods, including the 2D CNN with adjacent slices and the 3D CNN, against the simplest 2D CNN. Notably, both the 2D CNN with adjacent slices and the 3D CNN demonstrated superior performance compared to the standard 2D CNN. This enhanced performance suggests their potential for greater generalization across different datasets, offering possibilities for robust segmentation methods.

Table 8 from Article I reveals a notable disparity in the solid volume fraction for the 2D CNN segmentation, particularly evident in the npg_TLD dataset. This discrepancy suggests limitations in the standard 2D CNN's ability to accurately classify solid pixels within this dataset. This issue may be because 2D CNNs do not use all the available spatial features of such volumetric datasets.

Additionally, we explored the impact of window size, i.e., the number of adjacent slices, on segmentation performance. Our analysis, as shown in Figure 8 from Article I, highlights the superiority of 2D CNN with adjacent slices, particularly for window sizes 9 and 5 for the hnpg_epoxy_TLD and npg_TLD datasets, respectively.

These findings highlight the importance of methodological nuances, such as incorporating adjacent slices and optimizing window sizes, in improving segmentation accuracy and generalizability across different datasets. Such insights help design machine learning-based segmentation methods and advance the understanding of complex materials like HNPG within image processing.

3.3 Contribution to the field

- A novel method to generate synthetic FIB tomography images can help prepare a large number of training images with ground-truth values for training ML models or other applications. The same data pipeline can be used for HNPG and other complex structures and materials with different parameters.

- Our novel evaluation strategy based on isotropy can help the community evaluate the performance of their models for isotropic structures in materials science. This strategy can pave a new road toward measuring the performance of real nanoporous structures without ground-truth values.

- An extensive comparison between traditional and machine learning-based segmentation methods can help researchers select an effective method that mitigates the *shine-through effect* and achieves accurate 3D reconstruction.

3.4 Problems yet to be addressed

1. Measured thicknesses of each slice of the FIB tomography dataset differ from their target thicknesses. Additionally, these measured thicknesses are not constant. This irregularity may introduce additional inaccuracy in the 3D reconstruction of microstructures after semantic segmentation.

2. Machine learning-based semantic segmentation methods can still be improved because they can not be considered optimum. Machine learning strategies like multimodal machine learning may help achieve more accurate 3D reconstruction than the models trained on only a single modality dataset.

3. Data distribution of synthetic data and real data is significantly different. If this data distribution difference can be lowered using machine learning techniques such as domain adaptation, then the accuracy of semantic segmentation may improve and become more robust.

3.5 Article I - Training deep neural networks to reconstruct nanoporous structures from FIB tomography images using synthetic training data

T. Sardhara, R.C. Aydin, Y. Li, N. Piché, R. Gauvin, C.J. Cyron, M. Ritter: **Training deep neural networks to reconstruct nanoporous structures from FIB tomography images using synthetic training data.** Frontiers in Materials 9 (2022) 837006 (DOI: 10.3389/fmats.2022.837006)*

frontiers in Materials

ORIGINAL RESEARCH
published: 28 February 2022
doi: 10.3389/fmats.2022.837006

Training Deep Neural Networks to Reconstruct Nanoporous Structures From FIB Tomography Images Using Synthetic Training Data

Trushal Sardhara [1], Roland C. Aydin [2], Yong Li [3], Nicolas Piché [4], Raynald Gauvin [5], Christian J. Cyron [1,2] and Martin Ritter [6]*

[1]Institute for Continuum and Material Mechanics, Hamburg University of Technology, Hamburg, Germany, [2]Institute of Material Systems Modeling, Helmholtz-Zentrum Hereon, Geesthacht, Germany, [3]Institute of Materials Physics and Technology, Hamburg University of Technology, Hamburg, Germany, [4]Object Research Systems, Montreal, QC, Canada, [5]Department of Mining and Materials Engineering, McGill University, Montreal, QC, Canada, [6]Electron Microscopy Unit, Hamburg University of Technology, Hamburg, Germany

Focused ion beam (FIB) tomography is a destructive technique used to collect three-dimensional (3D) structural information at a resolution of a few nanometers. For FIB tomography, a material sample is degraded by layer-wise milling. After each layer, the current surface is imaged by a scanning electron microscope (SEM), providing a consecutive series of cross-sections of the three-dimensional material sample. Especially for nanoporous materials, the reconstruction of the 3D microstructure of the material, from the information collected during FIB tomography, is impaired by the so-called *shine-through effect*. This effect prevents a unique mapping between voxel intensity values and material phase (e.g., solid or void). It often substantially reduces the accuracy of conventional methods for image segmentation. Here we demonstrate how machine learning can be used to tackle this problem. A bottleneck in doing so is the availability of sufficient training data. To overcome this problem, we present a novel approach to generate synthetic training data in the form of FIB-SEM images generated by Monte Carlo simulations. Based on this approach, we compare the performance of different machine learning architectures for segmenting FIB tomography data of nanoporous materials. We demonstrate that two-dimensional (2D) convolutional neural network (CNN) architectures processing a group of adjacent slices as input data as well as 3D CNN perform best and can enhance the segmentation performance significantly.

Keywords: electron microscopy, synthetic training data, 3D reconstruction, semantic segmentation, SEM simulation, 3D CNN, 2D CNN with adjacent slices, machine learning

OPEN ACCESS

Edited by:
Surya R. Kalidindi, Georgia Institute of Technology, United States

Reviewed by:
Stefan G. Stanciu, Politehnica University of Bucharest, Romania
David Montes De Oca Zapiain, Sandia National Laboratories, United States

****Correspondence:***
Trushal Sardhara
trushal.sardhara@tuhh.de

Specialty section:
This article was submitted to Computational Materials Science, a section of the journal Frontiers in Materials

Received: *16 December 2021*
Accepted: *27 January 2022*
Published: *28 February 2022*

Citation:
Sardhara T, Aydin RC, Li Y, Piché N, Gauvin R, Cyron CJ and Ritter M (2022) Training Deep Neural Networks to Reconstruct Nanoporous Structures From FIB Tomography Images Using Synthetic Training Data. Front. Mater. 9:837006. doi: 10.3389/fmats.2022.837006

1 INTRODUCTION

Nanoporous materials bear great potential in microtechnology, chemical engineering, biomedical engineering, energy technology and electronics and communication technology. So-called FIB tomography combines the sequential removal of material layers by FIB with SEM imaging. It is a powerful technique for 3D imaging of nanoporous materials with a resolution of approx. 1 nm in the SEM plane and 10 nm in the out-of-plane direction (Knott et al., 2008).

However, accurate 3D reconstruction of nanoporous structures remains a challenge because of the so-called *shine-through effect* in FIB tomography data (Prill et al., 2013). Due to this effect, the

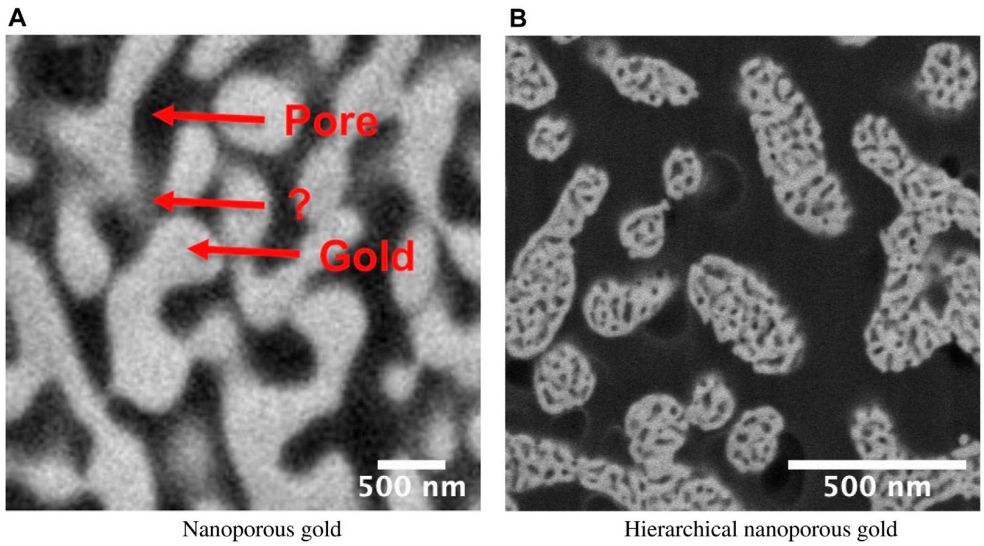


FIGURE 1 | Back-scattered electron scanning microscope images of epoxy-infiltrated nanoporous gold (npg) and hierarchical nanoporous gold (hnpg). **(A)** nanoporous gold structures below the cross-section plane shine through the epoxy-filled pores so that for some pixels it is unclear whether they belong to the solid (gold) or the pore (epoxy) phase (arrow with the question mark). **(B)** The influence of the shine-through effect is increased in hierarchical nanoporous gold due to the small pore sizes within the ligaments.

intensity of pixels in the SEM images generally depends not only on the material at the respective position in the plane currently imaged but also on structures in deeper layers. This effect occurs because these structures may shine through the nanopores up to the surface currently imaged by SEM, in case of back-scattered electron (BSE) imaging even in infiltrated nanoporous materials. Hence, there is no unique mapping between the intensity of a voxel in FIB tomography data and the material composition at exactly the position of this voxel. This ambiguity makes segmentation of FIB tomography data of nanoporous materials highly non-trivial (**Figure 1A**).

Classical methods like thresholding work best for standard materials without nanopores (Salzer et al., 2015). However, they fail for nanoporous materials with strong *shine-through effects* because of the ambiguity in the local voxel intensity. Machine learning algorithms like random forests or the k-means algorithm can help classify material and pores (Rogge and Ritter, 2018; Fager et al., 2020). However, deep learning-based (DL) methods, especially convolutional neural networks (CNN), bear the potential to outperform such methods when processing images. Over the last years, CNNs have more and more outperformed such classical methods across all disciplines (Krizhevsky et al., 2012; Girshick et al., 2014). For example, CNNs were used for the semantic segmentation of electron microscopy images of neuronal membranes (Ciresan et al., 2012). For the segmentation of FIB tomography images of porous membranes, the deep learning architecture ResUNet was applied, using initial training data generated by a random forest algorithm (Tracey et al., 2019). It is thus consequential to apply convolutional neural networks like U-Net, which was originally developed for biomedical images (Ronneberger et al., 2015), with some modifications also to FIB tomography data (Fend et al., 2021).

Due to *shine-through effects* in FIB tomography datasets, structures are visible through several subsequent SEM slices. Taking this information into account is an important step towards accurate segmentation of FIB tomography data of nanoporous materials. The machine learning architecture called CNN 2.5D has recently been reported to be particularly powerful (Vu et al., 2020) to incorporate such partial spatial information in a specific direction. CNN 2.5D feeds several adjacent slices into channels of a 2D CNN architecture. A similar approach, but with 3D kernels, is pursued by two other recently proposed machine learning architectures, 3D U-Net (Çiçek et al., 2016), and VNet (Milletari et al., 2016). The latter seeks to prevent information loss when the network grows deeper (Milletari et al., 2016).

Deep learning requires, however, a large amount of training data. In the context of the semantic segmentation of FIB tomography data, a sufficient amount of images is required whose pixels are labelled as belonging to specific categories (e.g., the solid or the pore phase in a nanoporous material). Obtaining sufficient training data from experiments can be expensive and time-consuming. To overcome this problem, synthetic training data is frequently used (Nikolenko, 2019).

In electron microscopy, two steps are required to generate synthetic training images for deep learning segmentation methods. The first step is the generation of a realistic geometric structure. The second step is the computer simulation of the FIB tomography of this structure, i.e., synthetic back-scattered electron (BSE) imaging data.

For the first step (Fend et al., 2021), previously used geometric primitives like spheres and cylinders. However, these do not adequately resemble the microstructure of the nanoporous materials of our interest. Therefore, these are suitable only to a limited extent for generating synthetic training data for the case we are interested in.

For the second step, Monte Carlo simulations of electron microscopy imaging can be performed. These simulate the trajectories of numerous electrons, thereby providing realistic information on the appearance of SEM images of specific structures. In order to perform such simulations on very simple geometries, first programs were developed more than two decades ago and have seen continuous improvements (Lowney J, 1994; Lowney, 1995; Karabekov et al., 2003; Zhang et al., 2012; Hovington et al., 1997; Gauvin et al., 2006).

In this paper, we compare different deep learning architectures for accurately segmenting FIB tomography data of nanoporous structures despite *shine-through effects*. We present a novel approach to generate synthetic FIB-SEM images using Monte Carlo (MC) simulations to overcome the lack of training data for deep learning methods. To obtain as realistic synthetic training data as possible, these simulations are not performed on geometries consisting of simple geometric primitives. Instead, we compare three different ways to generate largely realistic microstructures and use the most promising of them, the so-called levelled-wave algorithm (Li et al., 2020) as a basis for our study. Using the in silico training data generated this way, we demonstrate that 2D CNN with a group of adjacent slices as input data and 3D CNN can surpass the segmentation performance of classical methods by more than 100%. In the absence of ground-truth data, we measure the segmentation performance with a novel approach, which exploits specific geometrical properties of nanoporous gold and hierarchical nanoporous gold, such as isotropy.

2 MATERIALS AND METHODS

2.1 Generation of Synthetic Training Data

Synthetic FIB tomography data can be generated in two steps. The first step is the generation of virtual microstructures, and the second step is the generation of virtual SEM images of these microstructures using MC simulations.

2.1.1 Generation of Virtual Microstructures

To generate artificial microstructures that closely resemble the ones of nanoporous materials, we compared three different methods: the levelled wave method (LWM), self-similarity method (SSM) and random pore generation method (RPGM).

2.1.1.1 Levelled Wave Method

Nanoporous materials are often produced by dealloying. Theoretical analysis reveals (Li et al., 2020) that this leads to a microstructure whose geometry can be described by a superimposition of several wave vectors with an identical wavelength but different random orientations (Li et al., 2020). Subsequently, the Gaussian random field generated this way is subjected to a thresholding algorithm to divide it into a solid and a pore phase, resulting in microstructures as illustrated in **Figure 2A**.

2.1.1.2 Self-Similarity Method

A structure is called self-similar if it resembles exactly or partly itself. In this method, a hierarchical microstructure is generated using the thresholded images of a real nanoporous gold structure, hence the name "self-similarity method (SSM)". In the first step, FIB-SEM images of nanoporous gold are segmented to get binary images identifying the solid phase (intensity 255) and pore phase (intensity 0) using the k-means algorithm. Then, these binary images are resized using bilinear interpolation (Press et al., 1992) according to required voxel dimensions. In the next step, a mask, smaller in size than the binary images from the previous step, is prepared by resizing binary slices and rotating them at a random angle. Final output images are then calculated by performing arithmetic AND operations on binary images with masks using convolution (**Supplementary Figure S13**). These AND operations with masks generated from the original binary structure make the final structure self-similar. One slice from SSM is shown in **Figure 2B**.

2.1.1.3 Random Pore Generation Method

RPGM is a relatively simple method. In the first step, a volume that is fully solid (intensity 255) is chosen. Subsequently, void spheres are introduced at random locations and with a radius drawn from a Gaussian random distribution using a masking operation (**Supplementary Figure S14**). The advantage of RPGM is that it is a straightforward method. Its disadvantage is that it produces microstructures that exhibit considerable differences compared to actual nanoporous materials, limiting their value for training accurate machine learning-based segmentation algorithms in our case. The final sample geometry generated by this method is shown in **Figure 2C**.

2.1.2 Monte Carlo Simulation of Electron Microscopy

The virtual microstructures generated in the above-described ways are used in Monte Carlo simulations to generate synthetic FIB tomography data. The simulation of the BSE images is performed using the software MCXray. This software is an extension of the Monte Carlo simulation tools Casino (Hovington et al., 1997) and Win X-Ray (Gauvin et al., 2006). It was developed by (Gauvin and Michaud, 2009) and then incorporated in the Dragonfly Software [ORS, Montreal, Canada] (Object Research Systems, 2018). MCXray allows simulations of complex microstructures even consisting of different materials. Simulated BSE images of virtual microstructures generated by the above described three methods are presented in **Figure 3**. In all the Monte Carlo simulations for this paper, we assume beam energy of 2 kV, a beam current of 5×10^{-11} A, FIB sample stage tilt angle of 0 *degrees*, and detector to sample distance of 104 *mm*. We performed these simulations for nanoporous gold and hierarchical nanoporous gold, and considered pores in vacuum (**Figure 3**) as well as infiltrated with epoxy resin (see for example **Figures 4–B**). These simulations have a high *shine-through effect*

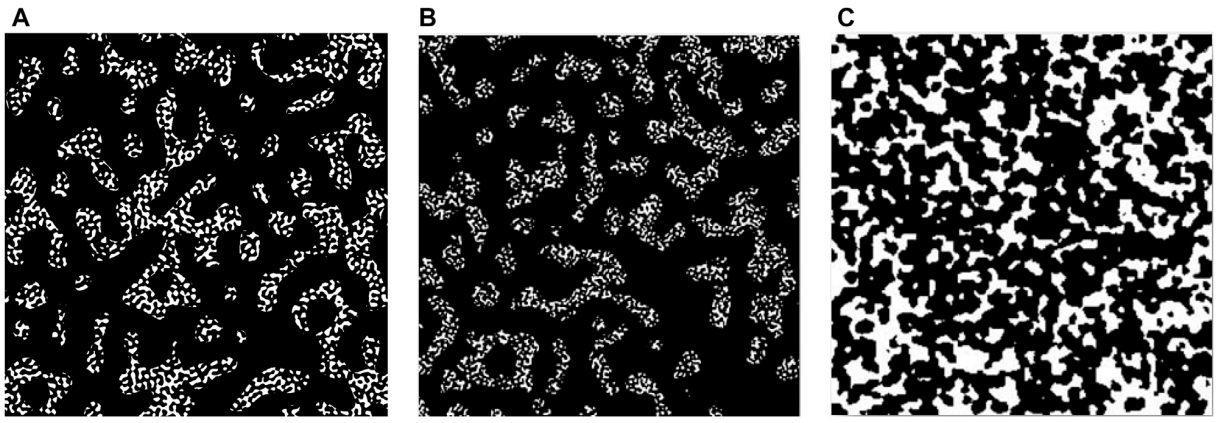


FIGURE 2 | Virtual microstructures generated by **(A)** levelled wave method **(B)** self-similarity method **(C)** random pore generation method.

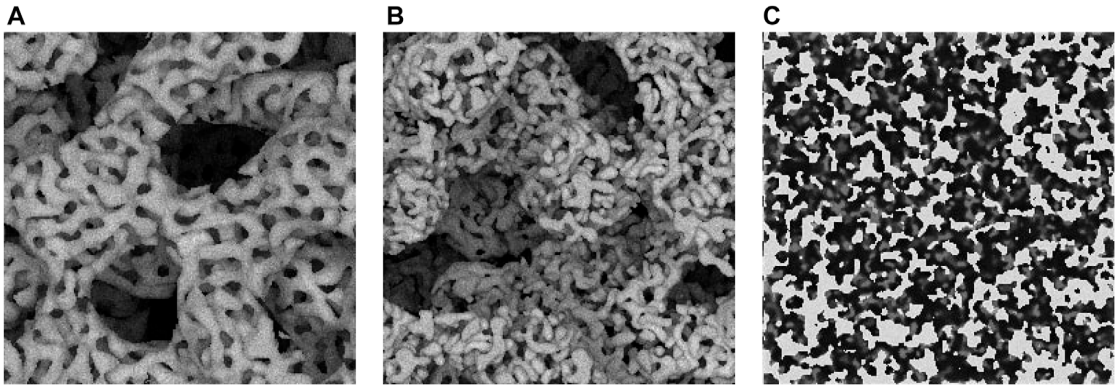


FIGURE 3 | Simulated BSE images using Monte Carlo simulation method and data generated using **(A)** LWM as initial virtual structure **(B)** SSM as initial virtual structure **(C)** RPGM as initial virtual structure.

and mimic images from a less surface sensitive through-the-lens (TLD) back-scattered electron detector.

2.2 Preprocessing of Training Data

In our study, we applied noise and blur filters to make the simulated images more similar to actual FIB tomography data. Subsequently, we applied the online data augmentation technique to increase the size of the training dataset.

2.2.1 Adding Noise

Electron microscopy images typically exhibit two types of noise (Cizmar et al., 2008). Primary electrons can generate Poisson noise, and the rest of the electrons from five noise sources can generate Gaussian noise (Timischl et al., 2012). Not all five sources are equally important, and noise added by detection systems is often assumed to be negligible (Sim et al., 2004; Goldstein, 2003). MCXray (Gauvin and Michaud, 2009) simulations of BSE images naturally include Gaussian noise (Hovington et al., 1997). We added the remaining Poisson noise and some additional Gaussian noise to understand the effect of noise in synthetic SEM images. To this end, we used the Scikit image library[1] in this project. First, Poisson noise was added to the image, and then Gaussian noise, to get a realistic noisy simulated BSE image. For the Gaussian noise, we heuristically chose a zero mean value and a variance of 0.001. After adding the noise, the intensities of the image were renormalized to a range from -1 to 1, converting the noisy image thereby to a standard unsigned 8-bit image. As training data for our study, we used these resulting noisy images.

2.2.2 Blurring of Edges

In the images generated by MCXray simulations[2], the edges were observed to be unrealistically sharp. Simple solutions like applying Gaussian filtering to the whole image may not work

[1]https://scikit-image.org

[2]Dragonfly software version 2021.1.0.118 13. Blurring of edges may not require in future versions of the software

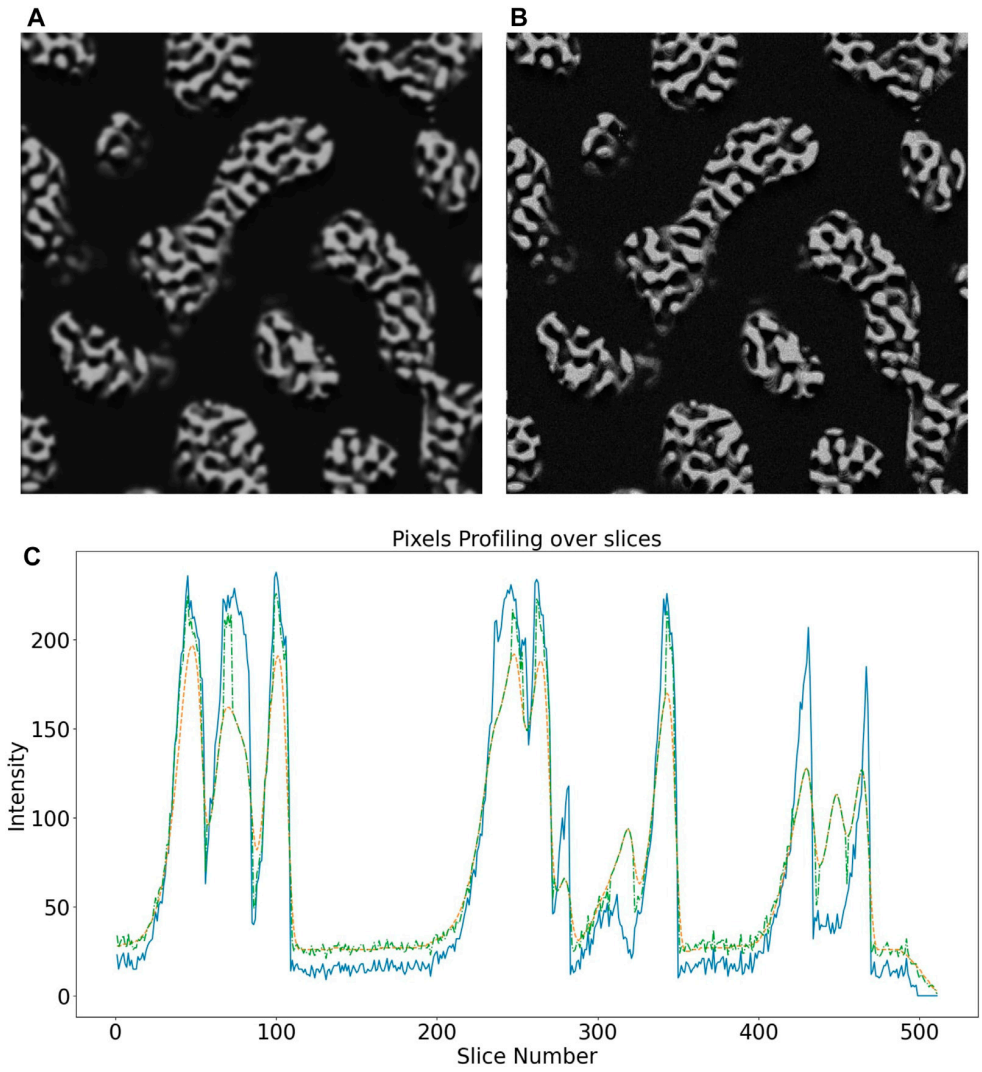


FIGURE 4 | Synthetic SEM image after applying **(A)** standard Gaussian blur filter and **(B)** the 4-step blurring procedure introduced herein **(C)** intensities of pixel located at a random location in the xy-plane across slices before blurring (blue), after applying Gaussian blur filter (orange) and 4-step blurring procedure (green).

as it may remove necessary Gaussian randomness of intensities from the solid ligaments in the simulated image. Therefore, it is very important to blur only the edges of the ligaments. This can be achieved by the following steps of masking and blurring.

1. Find edges of the solid ligaments
2. Generate mask with maximum weights at the edges from step (1)
3. Blur whole original image
4. Blend source image and blurred image from step (3) by mask from step (2)

The difference between this procedure and standard Gaussian blurring is illustrated in **Figure 4**. The image generated by the above 4-step procedure looks more realistic.

2.2.3 Data Augmentation

Data augmentation is a very powerful technique in deep learning when there is not enough training data available (Wang et al., 2017). Herein, we used online data upsampling; during the training process itself, the training data was augmented by applying random flips, rotations, brightness changes, and stretch transforms.

2.3 Machine Learning Architectures for Segmentation

In our synthetic FIB tomography data - as in real data - *shine-through effects* occur. Hence, it can be expected that accurate segmentation is not possible by processing image data layer by layer but rather in the group of the layers. Herein, we tested two machine learning architectures for segmentation that address this need.

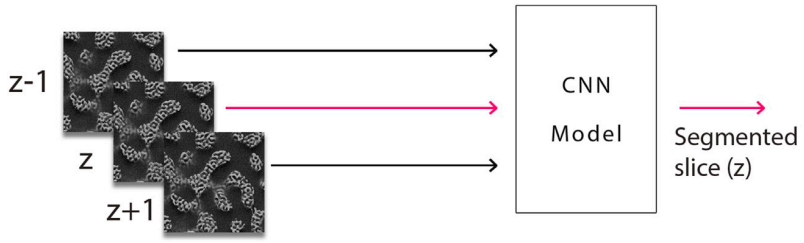


FIGURE 5 | Basic block diagram of 2D CNN receiving as input for the segmentation of slice *z* also the slices *z* – 1 (above) and *z* + 1 (below).

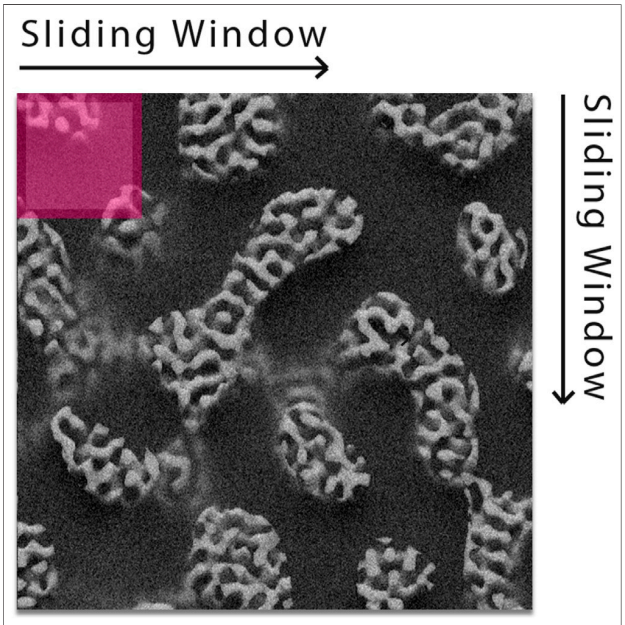


FIGURE 6 | Sliding window method.

2.3.1 2D CNN With Adjacent Slices

Shine-through effects establish a correlation between adjacent slices. To account for this in semantic segmentation, one has to segment different slices not separately but rather process information from several adjacent slides in each segmentation operation. As a way to avoid computationally costly 3D convolutions, one can process information from several adjacent layers simultaneously by 2D CNN (**Figure 5**). To this end, we used the 2D CNN architecture U-Net (Ronneberger et al., 2015). A general limitation of this approach is that the number of slices that can be included at a time is naturally limited by the number of slices available in the 3D image stack that forms the FIB tomography data. Indeed this limits the depth to which segmentation can be performed.

2.3.2 3D CNN

While 2D CNNs are computationally cheaper than 3D CNN, they may always be prone to miss out on recognizing some spatial features. To overcome this problem, we also tested 3D CNN. These use full 3D convolutions. We compared the 3D CNN architectures 3D U-Net (Çiçek et al., 2016), VNet (Milletari et al., 2016) and ResUNet3D with minor modifications. In the U-Net architecture (Ronneberger et al., 2015) we used padding in the convolutional blocks to retain the original image size. Moreover, we also added residual connections in one of our 3D U-Net models. We considered the number of encoding blocks as a hyperparameter and tuned it to improve performance.

TABLE 1 | Parameters used for training 2D CNN with adjacent slices.

Parameter	Value
Patch size	64
Stride	0.5
Batch size	64
Epochs	100 with early stopping with patience = 10
Loss	Dice loss
Optimizer	Adam
Learning rate	0.00001 with reducing it by factor of 0.10 with patience of 10

TABLE 2 | Parameters used for training 3D CNN.

Parameter	Value
Patch size	64
Stride	0.5
Batch size	1
Epochs	100 with early stopping with patience = 10
Loss	Dice Loss
Optimizer	Adam
Learning rate	0.0001 with reducing it by factor of 0.10 with patience of 10

2.4 Training of Neural Networks

All machine learning architectures were implemented using PyTorch; data loaders were written in *Python*, and models were trained on Tesla K80 GPUs.

2.4.1 2D CNN With Adjacent Slices

Input to the deep neural network was provided using the sliding window technique (**Figure 6**), where small patches (64 × 64 pixels) were generated from the original training dataset and used as the final training set for the network. We used 3, 5, 7 and 9 slices as the number of adjacent slices, making the input

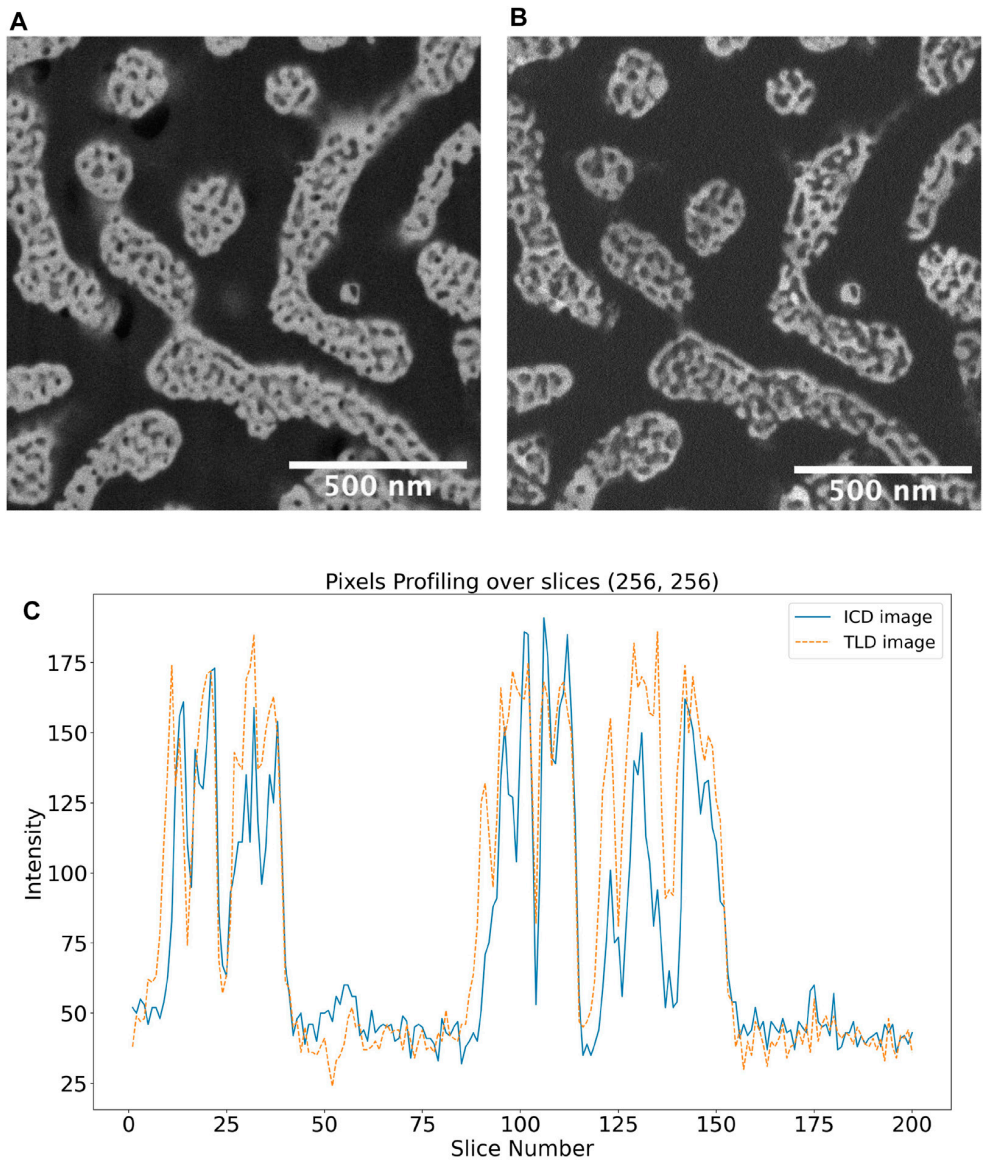


FIGURE 7 | SEM images of hierarchical nanoporous gold simultaneously recorded with **(A)** TLD and **(B)** ICD detector **(C)** intensities of pixel located at a random location in the xy-plane in SEM TLD (orange) and ICD (orange) images.

window size 64 × 64 × 3, 5, 7 and 9, to understand the effect of the number of adjacent slices on the segmentation performance. Additional parameters used for the training process are specified in **Table 1**.

2.4.2 3D CNN

We trained and compared a total of three 3D CNN models, namely U-Net3D, ResUNet and VNet. We used the same sliding window technique for all of them. However, unlike for the 2D CNN, the window was moving in all three spatial directions with a given stride. The training parameters are summarized in **Table 2**. Inspired by (Milletari et al., 2016), we used a squared Dice loss layer with the necessary smoothness value to avoid zero division in all three 3D CNN architectures. We used the mean Dice loss as the final loss value to account for possible data imbalance (Milletari et al., 2016).

2.5 Evaluation Criteria

Due to the unavailability of labelled datasets, the CNN models were trained based on synthetic training and validation data. Moreover, to evaluate the performance of various segmentation methods for real BSE images, we used the concept of anisotropy (**Supplementary Section S1**). The *shine-through effect* makes the images anisotropic in the *z*-direction (though the underlying material microstructure is in a statistical sense isotropic, that is,

TABLE 3 | Performance of different segmentation methods applied to dataset npg40.

	Target data	DL method (SSM)	DL method (LWM)	DL method (RPGM)	Otsu's algorithm	k-means clustering (k = 3)
Solid volume fraction (ϕ)	0.34	0.34	0.34	0.36	0.55	0.49
Relative error of solid volume fraction (%)	0.00	−1.99	0.50	4.33	33.69	17.94
Dice coefficient	1.00	0.98	0.99	0.89	0.75	0.78

TABLE 4 | Impact of preprocessing on segmentation performance for real hnpg_epoxy dataset.

DL model	Preprocessing	ϕ (TLD)	ϕ (ICD)	MP	e_{L2} TPCF	e_{L2} LPF	$e_{D_{L2}}$
2D CNN with adjacent slices	No	0.11	0.08	0.08	0.12	0.07	0.02
2D CNN with adjacent slices	Yes	0.16	0.10	0.10	0.07	0.008	0.004
3D CNN	No	0.19	0.11	0.12	0.15	0.18	0.12
3D CNN	Yes	0.18	0.13	0.10	0.13	0.12	0.08

lacking any preferred direction). Therefore, if the segmented images are isotropic, then it can be concluded that the segmentation method has (likely) been able to filter the *shine-through effect* and performs well. To measure anisotropy of the segmented images, we used two-point correlation functions (TPCF), **Supplementary Figures S7, S8**, and lineal path functions (LPF) for segmented ICD images (**Supplementary Figures S10, S11**) using our method and Otsu's thresholding method. This set of real imaging data, referred to henceforth as hnpg_epoxy dataset, has a voxel size of $2.6 \times 2.6 \times 10\ nm^3$, which was interpolated to $5.2 \times 5.2 \times 5.2\ \mathrm{nm}^3$ using bicubic interpolation. To quantify anisotropy of the segmented images, we calculated the TPCF and LPF in the x-, y- and *z*-direction. Subsequently, we computed the relative L2 differences for the TPCF and LPF in *z*-direction and compared them to the relative L2 differences for the TPCF and LPF in x- and *y*-direction, respectively. Generally, the relative L2 difference between two functions f and g can be computed in a discrete setting with n given data points as

$$e_{L2}(f,g) = \frac{2 \times \sqrt{\sum_{i=1}^{n} (f_i - g_i)^2}}{\left(\sqrt{\sum_{i=1}^{n} (f_i)^2} + \sqrt{\sum_{i=1}^{n} (g_i)^2}\right)} \tag{1}$$

where the f_i, g_i with $i = 1, \ldots, n$ are the given data points. Finally, we averaged the L2 differences of the *z*-direction compared to the x- and *y*-direction. This average is referred to henceforth (e.g., in **Table 4**) as TPCF or LPF e_{L2} difference. The larger both are, the higher the anisotropy of the segmented images, which can be considered a hint that the associated segmentation method has not been able to filter *shine-through effects*.

As an additional measure of anisotropy, we calculated the average diameters of ligaments in xy-, yz- and xz-planes using lineal path functions (D_{xy}, D_{yz}, D_{xz}). Then, we computed the averaged relative difference $e_{D_{L2}}$ of the diameters in the xz- and yz-planes compared to the one in the xy-plane as

$$e_{D_{L2}} = \frac{1}{2}\left(\sqrt{\frac{\left(D_{xz} - D_{xy}\right)^2}{D_{xy}^2}} + \sqrt{\frac{\left(D_{yz} - D_{xy}\right)^2}{D_{xy}^2}}\right) \tag{2}$$

TABLE 5 | Specifications of real FIB-SEM gold datasets used in this study. Note: All dataset discussed in the table have epoxy material as pore filling.

Dataset name	Gold structure	Pixel size [nm^3]	Detector type
hnpg epoxy ICD	hnpg	2.6_2.6_10	ICD
hnpg epoxy TLD	hnpg	2.6_2.6_10	TLD
npg ICD	npg	3.4_3.4_7	ICD
npg TLD	npg	3.4_3.4_7	TLD

Generally, the average diameter of ligaments in the *z*-direction is expected to be larger than in the x- and *y*-direction in the presence of non-filtered *shine-through effects*. Therefore, pronounced *shine-through effects* can be expected to result in large $e_{D_{L2}}$.

In the absence of ground truth, we also checked the congruency between images from two different sensors. We segmented FIB-SEM images from a low-loss in-column back-scattered electron detector (ICD) and compared the results with segmented images from the TLD. The ICD detector is situated at an upper position in the column, ensuring that only the most elastically scattered electrons (i.e., back-scattered electrons) are collected. The signal is highly sensitive to Z-contrast with almost no topographical information. Also, low energy loss increases the probability of near-surface interaction and therefore near-surface information (Ritter, 2019) (**Figures 7–B**). Therefore, one can expect the ICD images to have relatively small *shine-through effects* anyway so that they can provide at least some hint at the (not exactly known) ground truth. We calculated the fraction of misplaced pixels by

$$MP = \frac{(TP - IP)}{TP} \times 100 \tag{3}$$

where TP is the total number of pixels, IP the number of pixels identically segmented for TLD and ICD images, and MP is the fraction of pixels where the segmentation of TLD and associated ICD images disagreed.

TABLE 6 | Performance of different segmentation methods applied to synthetic LWM dataset. MP is here computed not using ICD images as reference but exact synthetic microstructure generated by the LWM.

Model name	ϕ (TLD)	ϕ (LWM)	MP	e_{L2} TPCF	e_{L2} LPF	$e_{D_{L2}}$
k-means clustering	0.33	0.12	0.26	0.83	0.75	0.95
Otsu's algorithm	0.54	0.12	0.43	0.96	0.65	0.71
2D CNN	0.12	0.12	0.04	0.18	0.03	0.02
2D CNN with adjacent slices	0.13	0.12	0.03	0.18	0.02	0.02
3D CNN	0.12	0.12	0.03	0.17	0.01	0.003

TABLE 7 | Performance of different segmentation methods applied to real hnpg_epoxy_TLD dataset.

Model name	ϕ (TLD)	ϕ (ICD)	MP	e_{L2} TPCF	e_{L2} LPF	$e_{D_{L2}}$
k-means clustering	0.39	0.13	0.39	0.20	0.41	0.37
Otsu's algorithm	0.38	0.29	0.12	0.21	0.29	0.24
2D CNN	0.20	0.10	0.16	0.14	0.05	0.01
2D CNN with adjacent slices	0.16	0.10	0.10	0.07	0.008	0.004
3D CNN	0.18	0.13	0.10	0.13	0.12	0.08

TABLE 8 | Performance of different segmentation methods applied to real npg_TLD dataset.

Model name	ϕ (TLD)	ϕ (ICD)	MP	e_{L2} TPCF	e_{L2} LPF	$e_{D_{L2}}$
k-means clustering	0.33	0.24	0.57	0.46	0.28	0.24
Otsu's algorithm	0.54	0.40	0.19	0.40	0.18	0.12
2D CNN	0.24	0.21	0.18	0.38	0.20	0.10
2D CNN with adjacent slices	0.40	0.36	0.17	0.16	0.07	0.05
3D CNN	0.44	0.31	0.20	0.12	0.06	0.04

3 RESULTS AND DISCUSSION

3.1 Selection of Best Synthetic Training Data

The first question we sought to address was which of the three above described methods of generating synthetic training data (LWM, SSM, RPGM) is most suitable for training deep neural networks for semantic segmentation. We performed a study using the CNN architecture U-Net to answer this question. We trained the U-Net on the above three types of synthetic data, splitting them into training data (60%), validation data (20%), and test data (20%). Subsequently, we compared the performance of the trained U-Nets to two classical segmentation methods, Otsu's thresholding algorithm (Otsu, 1979) and k-means clustering (Lloyd, 1982) with k = 3.

To evaluate the performance of these segmentation methods, we generated a synthetic dataset called npg40, which was prepared using MC simulations on synthetic microstructures generated by k-means segmentation of real nanoporous gold SEM images. The performance of the models was then evaluated using metrics as the relative error of solid volume fraction and the Dice coefficients. As shown in **Table 3**, the deep learning model trained on the LWM data exhibits the lowest solid volume fraction error and highest Dice coefficient. Therefore, we trained all our deep learning models, shown below, using synthetic LWM data only after this preliminary test.

3.2 Role of Preprocessing

It is instructive to compare the performance of segmentation methods trained on synthetic training data with and without preprocessing ("blur the edges" and subsequent addition of noise). To this end, we trained 3D CNN and 2D CNN processing adjacent slices (on synthetic LWM data) and subsequently compared their segmentation performance on the (real) hnpg_epoxy dataset. **Table 4** reveals that preprocessing increases segmentation performance for both 2D CNN processing adjacent slices and 3D CNN, underlining that preprocessing indeed helps to generate more realistic synthetic training data.

3.3 Semantic Segmentation

We validated our trained models and their performance both by the segmentation of one synthetic LWM dataset and four sets of real FIB-SEM images (see **Table 5**). Results are presented in **Tables 6**, **7**, and **8**. In these tables, it is apparent that the L2 differences for the TPCF and LPF and also the ligament diameter anisotropy is much lower for the novel segmentation methods based on deep learning introduced in this paper compared to classical methods like Otsu's method or k-means clustering (with k = 3). This suggests that our novel deep learning methods filter *shine-through effects* much better than classical methods and thus produce a geometry much closer to the real one. The excessive solid volume fraction of the segmented ICD data compared to the associated TLD data suggests that, in particular, Otsu's thresholding method in many

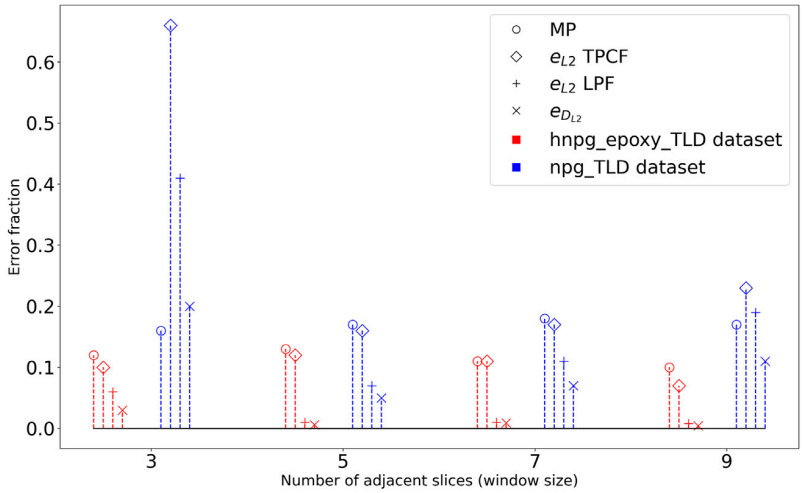


FIGURE 8 | Effect of window size when using 2D CNNs with adjacent slices for segmentation for the real datasets from **Table 5**.

cases fails to classify pixels visible due to the *shine-through effect*. Our deep learning-based segmentation methods exhibit much lower MP values compared to Otsu's method and k-means clustering. That is, they succeed in more cases to classify pixels in TLD and ICD imaging data identically. Following the concept of sensor fusion, also this is indicative of a better performance of our deep learning-based segmentation methods. It is interesting to discuss why there is a particularly large number of misplaced pixels for both real data sets when using k-means clustering. To explain this, it is important to note that we selected k = 3 for k-means clustering to account for the presence of a total of three different clusters, namely a solid phase, a pore phase and artifacts (i.e. pixels due to the shine-through effect). This assumption is certainly reasonable for TLD images because of their sensitivity to shine-through effects. By contrast, ICD images are much less prone to this problem, so k = 3 may no longer be a good assumption, resulting in poor performance of the associated segmentation.

Among the CNN-based segmentation methods, the 2D CNN with adjacent slices and the 3D CNN performed better than the standard 2D CNN. They probably can generalize better to the different datasets. For example, the very low solid volume fraction in **Table 8** for the 2D CNN segmentation suggests that this method fails to classify solid pixels for the npg_TLD dataset.

We also studied the effects of window size (number of adjacent slices) on segmentation performance. It is evident from **Figure 8** that 2D CNN processing adjacent slices provided the best performance for windows of sizes 9 and 5 for the hnpg_epoxy_TLD and the npg_TLD datasets, respectively.

4 CONCLUSION

Deep learning can play an important role in the segmentation of FIB tomography data. A potential caveat is the limited availability of training data. We demonstrated that the lack of training data can be overcome by generating virtual microstructures and simulating them using the MCXray method, providing ample synthetic FIB tomography training data. We compared three different methods to generate realistic synthetic nanoporous geometries. Our results reveal that the training data generated by the levelled wave method (LWM) is most effective for deep learning of image segmentation. A major problem in this context is the typically missing availability of ground truth. That is, for real materials, most often, there is no information available that would characterize their microstructure more accurately and reliably than segmented FIB tomography data, which makes it naturally difficult to evaluate the performance of a specific segmentation method. Herein, we introduced a novel approach to measure the extent to which *shine-through effects* - which can be expected to be the dominant source of errors in the semantic segmentation of images of nanoporous materials - are filtered out by different segmentation methods. This method did not require any full ground truth but rather exploited that nanoporous materials typically exhibit an isotropic microstructure. This way, the degree of anisotropy of the segmented images could be used as a proxy of the segmentation error. We tested different deep learning architectures for the segmentation of FIB tomography data and identified 2D CNNs with adjacent slices as image channels and 3D CNNs as the best architectures of the ones tested herein. Generally, 3D CNNs were found to be computationally more expensive to train than 2D CNN with adjacent slices.

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed for this study can be found in TUHH Open Research repository (https://tore.tuhh.de/handle/11420/11060).

AUTHOR CONTRIBUTIONS

TS, RA, CC, and MR contributed to conception and design of the study. YL generated the LWM database. NP and RG helped

with the formation of Monte Carlo simulations. TS wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 986 – Project number 192 346 071.

ACKNOWLEDGMENTS

We thank Stefan Bartels for his contribution by providing us the necessary computation power to generate the simulated images in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmats.2022.837006/full#supplementary-material

REFERENCES

Ciresan, D., Giusti, A., Gambardella, L., and Schmidhuber, J. (2012). Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images. *Adv. Neural Inf. Process. Syst.* 25, 2843–2851.

Cizmar, P., Vladár, A. E., Ming, B., and Postek, M. T. (2008). For the National Institute of Standards and TechnologySimulated Sem Images for Resolution Measurement. *Scanning* 30, 381–391. doi:10.1002/sca.20120

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). "3d U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in International conference on medical image computing and computer-assisted intervention (Berlin, Germany: Springer), 424–432. doi:10.1007/978-3-319-46723-8_49

Fager, C., Röding, M., Olsson, A., Lorén, N., von Corswant, C., Särkkä, A., et al. (2020). Optimization of FIB-SEM Tomography and Reconstruction for Soft, Porous, and Poorly Conducting Materials. *Microsc. Microanal* 26, 837–845. doi:10.1017/s1431927620001592

Fend, C., Moghiseh, A., Redenbach, C., and Schladitz, K. (2021). Reconstruction of Highly Porous Structures from FIB-SEM Using a Deep Neural Network Trained on Synthetic Images. *J. Microsc.* 281, 16–27. doi:10.1111/jmi.12944

Gauvin, R., Lifshin, E., Demers, H., Horny, P., and Campbell, H. (2006). Win x-ray: A New Monte Carlo Program that Computes X-ray Spectra Obtained with a Scanning Electron Microscope. *Microsc. Microanal.* 12, 49–64. doi:10.1017/s1431927606060089

Gauvin, R., and Michaud, P. (2009). Mc x-ray, a New Monte Carlo Program for Quantitative X-ray Microanalysis of Real Materials. *Microsc. Microanal* 15, 488–489. doi:10.1017/s1431927609092423

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 580–587. doi:10.1109/cvpr.2014.81

Goldstein, J. (2003). *Scanning Electron Microscopy and X-Ray Microanalysis*. Berlin, Germany: Springer US.

Hovington, P., Drouin, D., and Gauvin, R. (1997). Casino: A New Monte Carlo Code in C Language for Electron Beam Interaction—Part I: Description of the Program. *Scanning* 19, 1–14. doi:10.1002/sca.4950190104

Karabekov, A., Zoran, O., Rosenberg, Z., and Eytan, G. (2003). Using Monte Carlo Simulation for Accurate Critical Dimension Metrology of Super Small Isolated Poly-Lines. *Scanning: J. Scanning Microscopies* 25, 291–296.

Knott, G., Marchman, H., Wall, D., and Lich, B. (2008). Serial Section Scanning Electron Microscopy of Adult Brain Tissue Using Focused Ion Beam Milling. *J. Neurosci.* 28, 2959–2964. doi:10.1523/jneurosci.3189-07.2008

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet Classification with Deep Convolutional Neural Networks. *Commun. ACM* 60, 84–90. doi:10.1145/3065386

Li, Y., Dinh Ngô, B.-N., Markmann, J., and Weissmüller, J. (2020). Datasets for the Microstructure of Nanoscale Metal Network Structures and for its Evolution during Coarsening. *Data in brief* 29, 105030. doi:10.1016/j.dib.2019.105030

Lloyd, S. (1982). Least Squares Quantization in Pcm. *IEEE Trans. Inform. Theor.* 28, 129–137. doi:10.1109/tit.1982.1056489

Lowney J, M. E. (1994). *User's Manual for the Program Monsel-1: Monte Carlo Simulation of Sem Signals for Linewidth Metrology*. Gaithersburg, Maryland, U.S.: NIST special publication.

Lowney, J. (1995). *Monsel-ii Monte Carlo Simulation of Sem Signals for Linewidth Metrology*. Gaithersburg, Maryland, U.S.: NIST special publication.

Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). "V-net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in 2016 fourth international conference on 3D vision (3DV) (IEEE), 565–571. doi:10.1109/3dv.2016.79

Nikolenko, S. (2019). "Synthetic Data in Deep Learning," in School-conference "Approximation and Data Analysis 2019", 21.

Object Research Systems (2018). *Object Research Systems (ORS) Inc*. Montreal, Canada: Object Research Systems (ORS) Inc. Dragonfly 3.6 [computer software].

Otsu, N. (1979). A Threshold Selection Method from gray-level Histograms. *IEEE Trans. Syst. Man. Cybern.* 9, 62–66. doi:10.1109/tsmc.1979.4310076

Press, W. H., Teukolsky, S. A., Flannery, B. P., and Vetterling, W. T. (1992). *Numerical Recipes in Fortran 77: Volume 1, Volume 1 of Fortran Numerical Recipes: The Art of Scientific Computing*. Cambridge: Cambridge University Press.

Prill, T., Schladitz, K., Jeulin, D., Faessel, M., and Wieser, C. (2013). Morphological Segmentation of Fib-Sem Data of Highly Porous media. *J. Microsc.* 250, 77–87. doi:10.1111/jmi.12021

Ritter, M. (2019). Comparison of 3d-Fib Tomography Volume Reconstruction with Various In-Column Detectors on an Epoxy-Infiltrated Nonoporous Metal Test Sample," in EMAS 2019—16th European Workshop on Modern Developments and Applications in Microbeam Analysis.

Rogge, F., and Ritter, M. (2018). Cluster Analysis for Fib Tomography of Nanoporous Materials," in EMAS 2019—19th International Microscopy Congress (IMC19).

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional Networks for Biomedical Image Segmentation," in International Conference on Medical image computing and computer-assisted intervention (Berlin, Germany: Springer), 234–241. doi:10.1007/978-3-319-24574-4_28

Salzer, M., Prill, T., Spettl, A., Jeulin, D., Schladitz, K., and Schmidt, V. (2015). Quantitative Comparison of Segmentation Algorithms for Fib-Sem Images of Porous media. *J. Microsc.* 257, 23–30. doi:10.1111/jmi.12182

Sim, K. S., Thong, J. T. L., and Phang, J. C. H. (2004). Effect of Shot Noise and Secondary Emission Noise in Scanning Electron Microscope Images. *Scanning* 26, 36–40. doi:10.1002/sca.4950260106

Timischl, F., Date, M., and Nemoto, S. (2012). A Statistical Model of Signal-Noise in Scanning Electron Microscopy. *Scanning* 34, 137–144. doi:10.1002/sca.20282

Tracey, J., Lin, S. S., Jankovic, J., Zhu, A., and Zhang, S. (2019). Iterative Machine Learning Method for Pore-Back Artifact Mitigation in High Porosity Membrane Fib-Sem Image Segmentation. *Microsc. Microanal* 25, 186–187. doi:10.1017/s1431927619001661

Vu, M. H., Grimbergen, G., Nyholm, T., and Löfstedt, T. (2020). Evaluation of Multislice Inputs to Convolutional Neural Networks for Medical Image Segmentation. *Med. Phys.* 47, 6216–6231. doi:10.1002/mp.14391

Wang, J., and Perez, L. (2017). The Effectiveness of Data Augmentation in Image Classification Using Deep Learning. *Convolutional Neural Networks Vis. Recognit* 11, 1–8.

Zhang, P., Wang, H. Y., Li, Y. G., Mao, S. F., and Ding, Z. J. (2012). Monte Carlo Simulation of Secondary Electron Images for Real Sample Structures in Scanning Electron Microscopy. *Scanning* 34, 145–150. doi:10.1002/sca.20288

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary Material

1 CONCEPT OF ANISOTROPY

Validation of the machine learning model in our case is a difficult process as we do not have ground-truth data available. As nanoporous gold is self-similar, we should see isotropy in all directions. We used two functions, namely the two-point correlation function and the lineal path function, to measure anisotropy in the microstructures. The two-point correlation function is the autocorrelation function showing the probability of finding another point of the same phase in the neighbourhood. Similarly, the lineal path function shows the probability of having two points in the same cluster. Both functions in all directions should remain very similar for the isotropic structures.

To test how well both functions work to detect anisotropy, we prepared one virtual structure with controlled anisotropy in the z-direction. This structure contained spheres stretched in the z-direction, making them elliptical in the +z direction. The intensity of the stretched part increases by going deeper in the z-direction to mimic the shine-through effect. Distances between each sphere layer in z-direction also change (Figure S1). These changes make the virtual structure isotropic in x and y-direction while anisotropic in the z-direction.

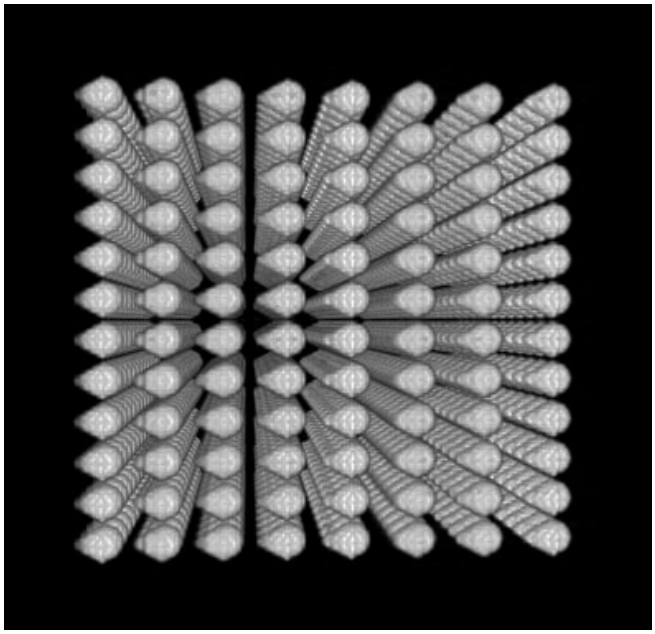


Figure S1. 3D view of virtual anisotropic structure

1.1 Test of isotropy – Two point correlation function

We calculated two-point correlation functions (Berryman, 1985) for prepared virtual structure in x, y and z-directions. As expected, we get the same two-point correlation functions in x and y-directions (yz and xz planes), but we get significantly different functions in the z-direction (xy plane) (Figure S2).

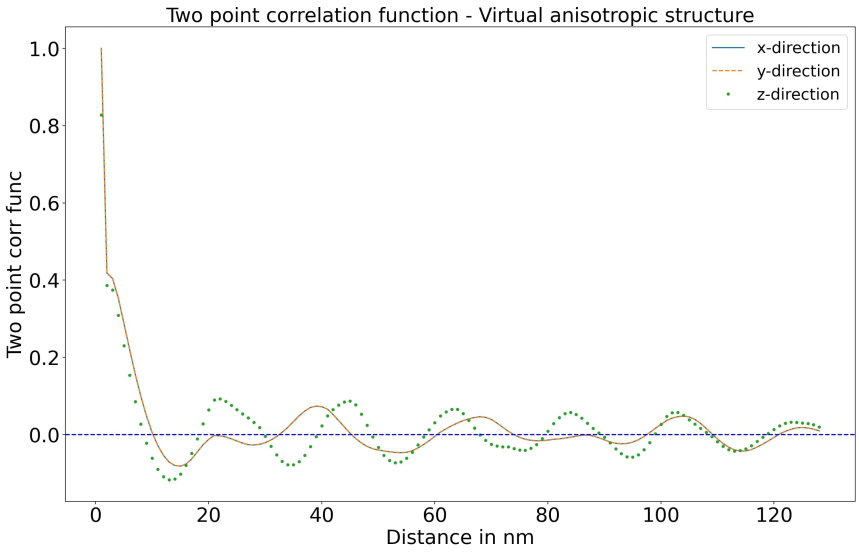


Figure S2. Two-point correlation function plots of virtual anisotropic structure in z-dir (xy plane), y-dir (xz plane) and x-dir (yz plane)

1.2 Test of isotropy – Lineal path function

(Torquato and Haslach Jr, 2002) proposed inserting lines of arbitrary length and orientation in the image and then counting the fraction of these lines that lie wholly within a single phase to calculate the Lineal path function. This function can also be useful to calculate average ligament size.

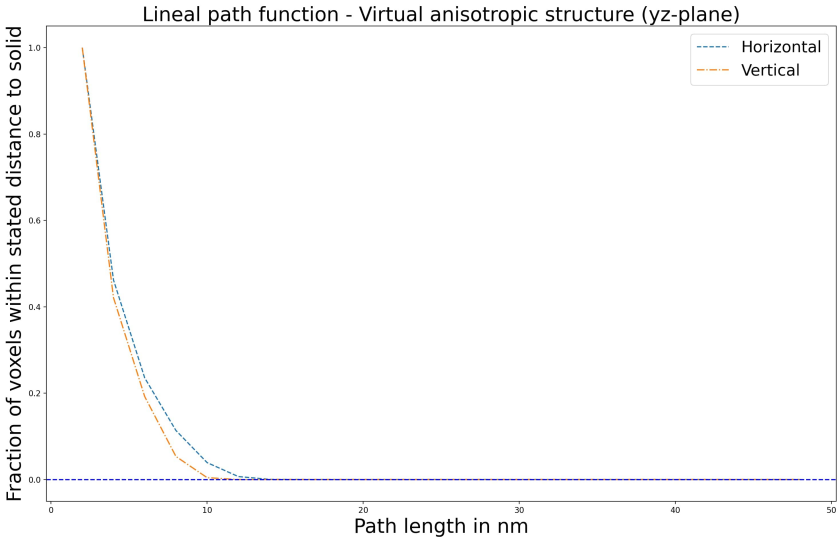


Figure S3. Lineal path function plots of virtual anisotropic structure in yz-plane

We have calculated a total of two lineal path functions whose directions are perpendicular to each other. Different lengths and errors between two functions suggest anisotropy in the structure. We repeated this procedure for all three planes.

As we know, there is some anisotropy in the z-direction; we can see length differences in the function in the yz-plane (Figure S3).

2 TRAINING PLOTS

In figures S4-S6, training/validation loss curves for different segmentation methods used in this research are shown. The figures show mean Dice loss over multiple epochs, with training scores (blue) and validation scores (orange). We used online data augmentation only during the training phase. Therefore, the training dataset was comparatively more complex than the validation dataset. Due to this complex training dataset, values of the Dice loss during the validation cycle are lower than values of the Dice loss during the training cycle in all three training/validation loss curves.

3 SUPPLEMENTARY TABLES AND FIGURES

REFERENCES

Berryman, J. G. (1985). Measurement of spatial correlation functions using image processing techniques. *Journal of Applied Physics* 57, 2374–2384

Torquato, S. and Haslach Jr, H. (2002). Random heterogeneous materials: microstructure and macroscopic properties. *Appl. Mech. Rev.* 55, B62–B63

3.1 Figures

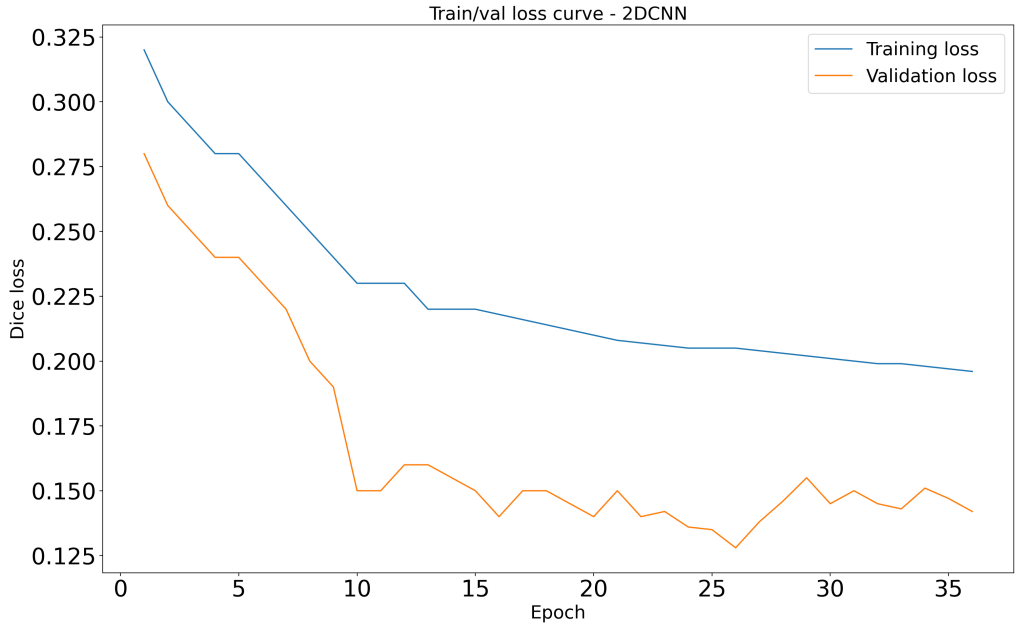


Figure S4. Training/validation loss curve for 2D CNN segmentation method

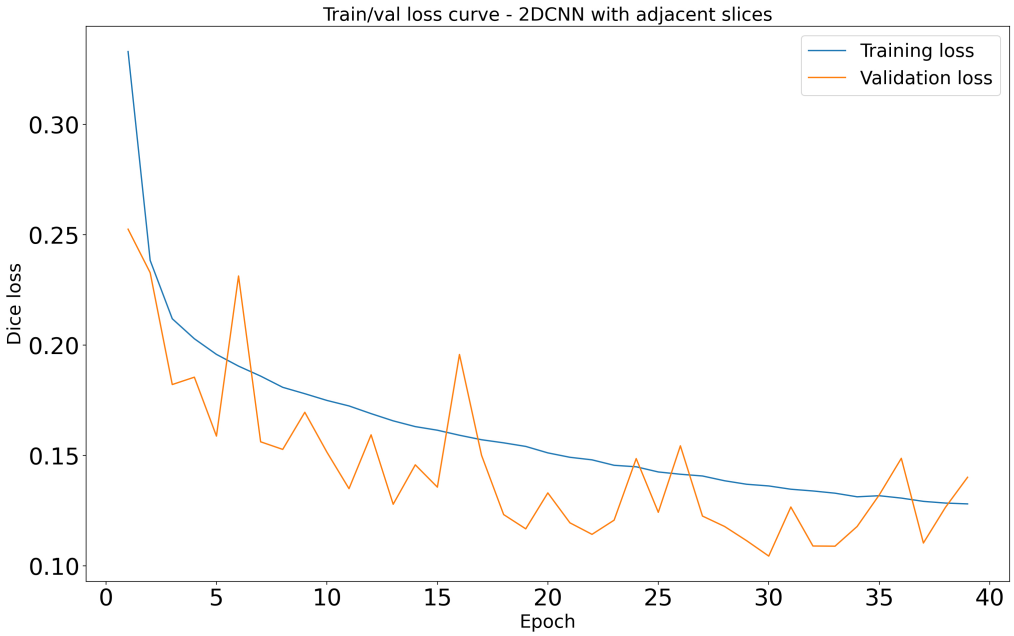


Figure S5. Training/validation loss curve for 2D CNN with adjacent slices segmentation method

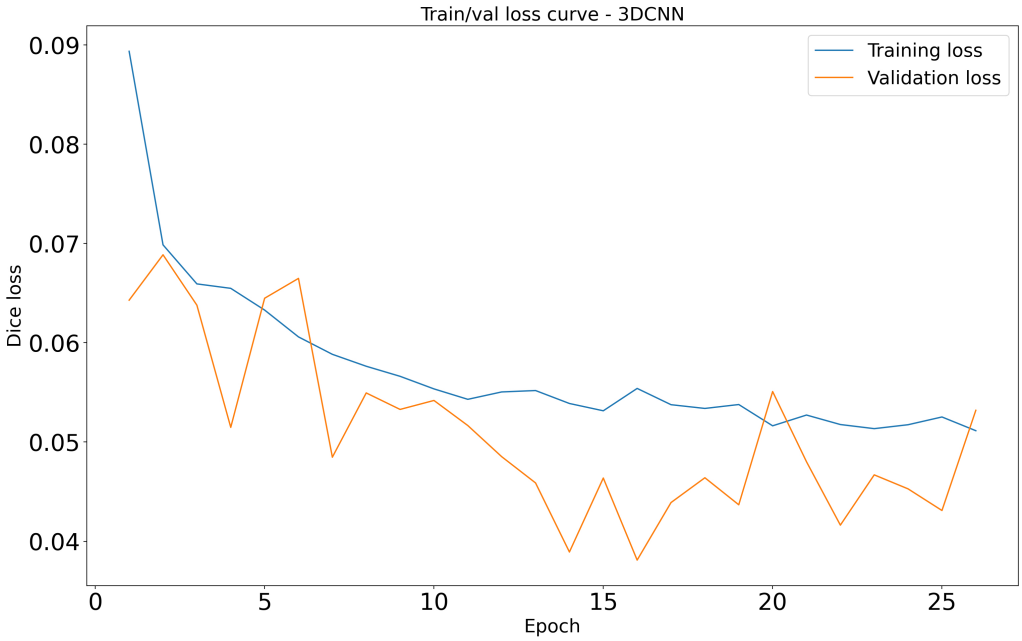


Figure S6. Training/validation loss curve for 3D CNN segmentation method

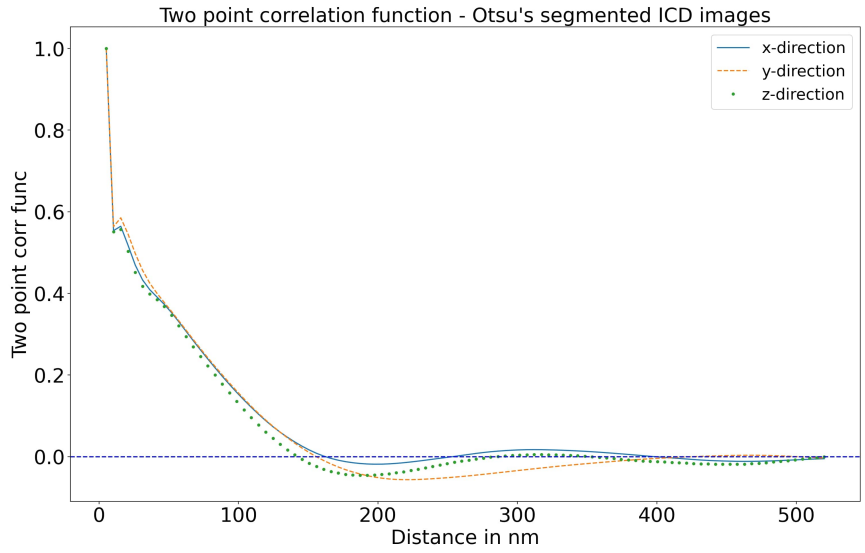


Figure S7. Two-point correlation function plots of segmented real hnpg_epoxy ICD dataset (using Otsu's method) in z-dir (xy plane), y-dir (xz plane) and x-dir (yz plane)

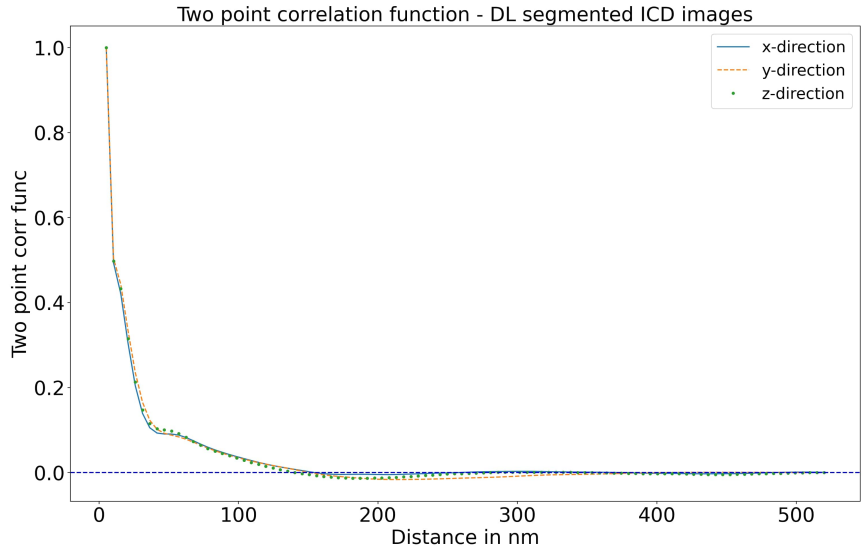


Figure S8. Two-point correlation function plots of segmented real hnpg_epoxy ICD dataset (using our deep learning method) in z-dir (xy plane), y-dir (xz plane) and x-dir (yz plane)

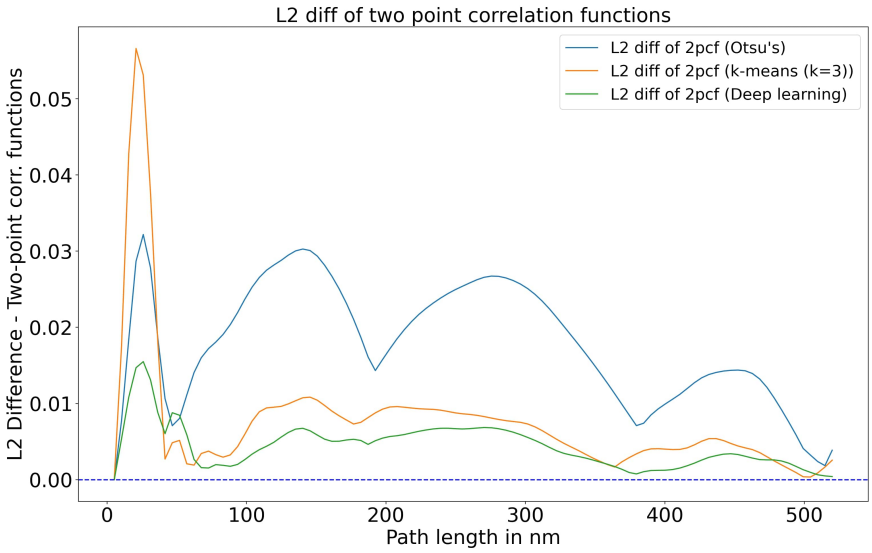


Figure S9. L2 differences of two-point correlation functions of segmented structures using Otsu's, k-means and our deep learning method

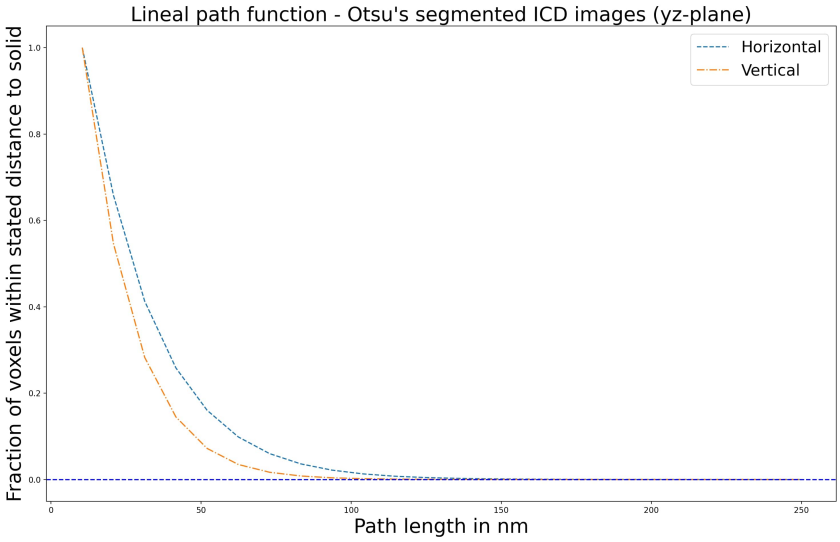


Figure S10. Lineal path function plots of segmented real hnpg_epoxy TLD dataset (using Otsu's method) in yz-plane

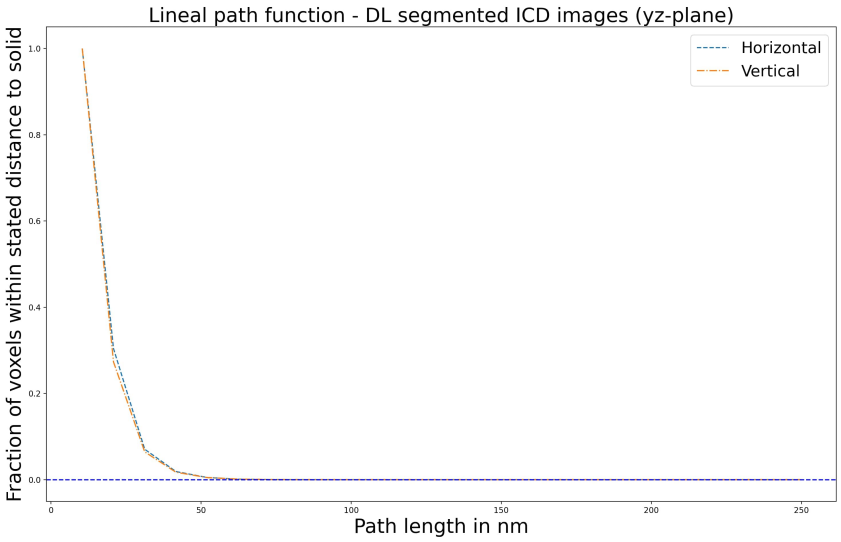


Figure S11. Lineal path function plots of segmented real hnpg_epoxy ICD dataset (using our deep learning method) in yz-plane

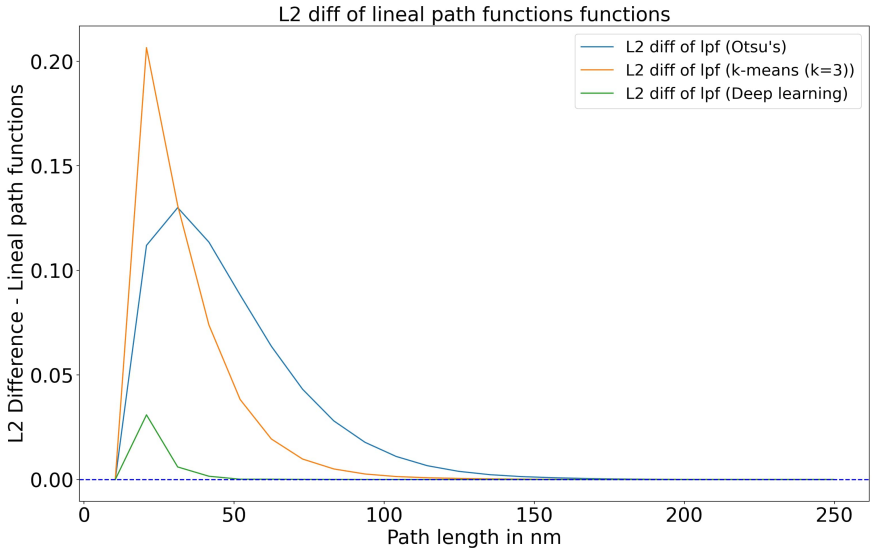


Figure S12. L2 differences of lineal path functions of segmented structures using Otsu's, k-means and our deep learning method

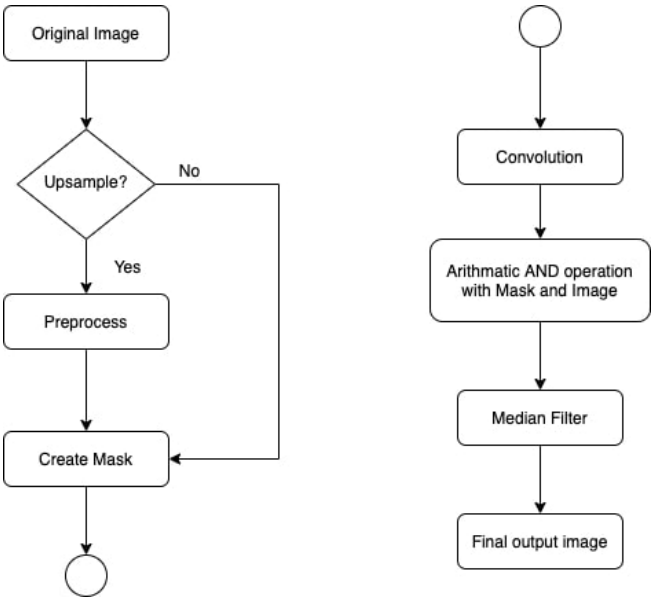


Figure S13. Methodology of SSM

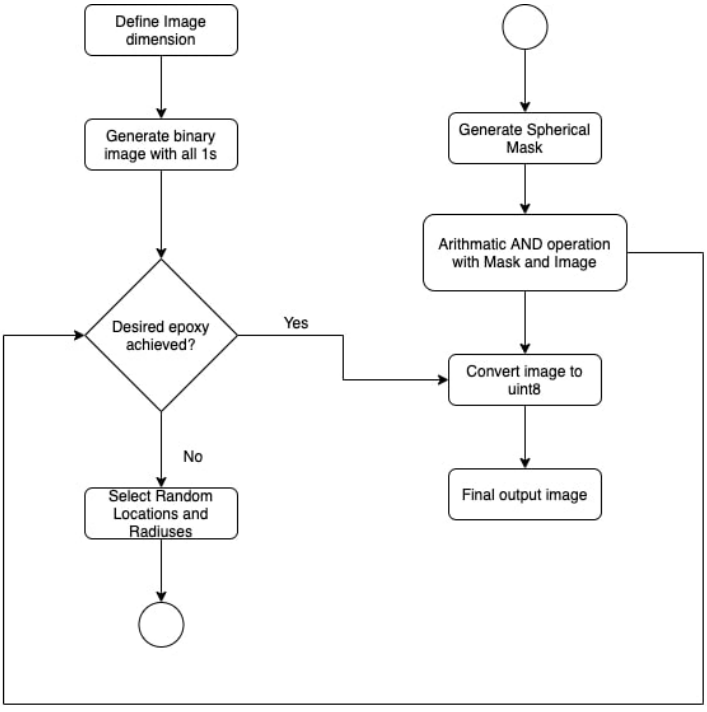


Figure S14. Methodology of RPGM

Chapter 4

Role of slice thickness quantification in the 3D reconstruction

4.1 Background of the study

While Article I focused on segmentation methods and the evaluation strategy of trained models, Article II focused on addressing errors post-semantic segmentation. These errors often arise due to inherent limitations of the FIB system, including factors such as beam shape [74], charging effects [75], high surface roughness [76], and sample drift [77]. These factors lead to considerable variations in the thickness of milled slices [78, 79], resulting in significant deviations from the intended target thickness. Consequently, these measurement errors in slice thickness lead to inaccuracies in the 3D reconstruction of microstructures. Therefore, accurately quantifying the actual thickness of each slice is important for accurate reconstruction of the microstructure volume. Furthermore, correct thickness measurements are essential for accurately calculating isotropy functions, particularly considering the lower resolution in the z-direction (10 nm) compared to the higher resolution in the x- and y-directions (2.6 nm). Before calculating these isotropy functions, it is imperative to standardize the resolutions across all three directions. Any inaccuracies in the measurement of slice thicknesses can introduce errors when increasing the resolution in the z-direction to match that in the x- and y-directions.

To ensure accurate and precise measurement of thicknesses, we implemented a novel ruler-based approach inspired by the study of [80]. Instead of fiducial markers, ruler markings were utilized as an alternative for estimating the thickness of milled slices. These markings facilitate slice-to-slice alignment in the plane during reconstruction

and provide accurate measurements of slice thickness in the direction perpendicular to the plane, with minimal error [79].

In our study, we implemented the ruler-based approach and the auto slice & view (ASV) method using fiducial markers to measure slice thicknesses. Subsequently, we compared the thicknesses measured by both methods to analyze which approach yielded accurate results.

Once accurate thickness measurements were obtained, we proposed a novel slice repositioning method to make resolutions in all directions the same. This method involved repositioning each slice according to new positions calculated on a smaller grid size equal to the desired resolution (in our case, equal to the resolution in the xy-plane). During this process, some slices could have the same position on the grid. In such cases, overlapping slices were distributed to the nearest free positions to ensure no real data information was lost.

Following the repositioning of slices, gaps were created between each slice due to the lower resolution in the z-direction. These gaps need to be filled with meaningful information to increase the accuracy of the resolution of the microstructure. We approached this problem as an image inpainting task rather than an image super-resolution task, as it was evident that these gaps were not uniformly distributed. In image inpainting, missing parts in images are reconstructed using either classical interpolation methods, such as nearest neighbor [81] or cubic splines [82], or machine learning-based methods.

We proposed a novel machine learning-based image inpainting method to calculate missing information in images. This method was compared with the classical cubic splines method to demonstrate its superior performance. This comparison aimed to highlight the efficacy and accuracy of our proposed machine learning-based approach in filling the gaps and generating missing information for the microstructure images.

To evaluate the efficacy of our slice repositioning method, we generated various synthetic datasets with an isometric resolution of $2.5 \times 2.5 \times 2.5\,nm^3$, derived from a dataset designed to mimic real data with variable thicknesses. This mimicking dataset contained varying slice thicknesses based on a realistic gamma distribution.

Subsequently, we prepared different datasets by repositioning slices according to their measured thicknesses and then generating missing gaps using one of three methods: nearest-neighbor, cubic splines, or our proposed machine learning method. Detailed information about the different datasets is provided in Table 1 of Article II.

This detailed evaluation allowed us to measure the performance and accuracy of our slice repositioning method and the effectiveness of different interpolation methods in reconstructing missing information in the microstructure images.

4.2 Discussion

This study compared classical interpolation methods, including cubic spline and nearest neighbor interpolation, and our novel machine learning-based interpolation approach. We generated 250 mimicking datasets from 25 distinct LWM datasets, each having ten different thickness distributions. Each of these mimicking datasets underwent interpolation using the three aforementioned methods.

Notably, we observed that the cubic spline method occasionally created gaps within ligaments. To address this, we applied image processing techniques to solve this issue, ensuring the best possible outcome with cubic spline interpolation (further details provided in the supplementary section of Article II). It is important to emphasize that the data used for interpolation were entirely different from the data used to train the machine learning model.

The error metrics introduced in Section 2.4.1 of Article II were calculated for each interpolated dataset, with results reported in Table 2 from Article II. Our machine learning-based interpolation method demonstrated superior performance compared to classical interpolation methods. Additionally, error measures from Section 2.4.2 of Article II were calculated for each dataset, with mean and standard deviation values presented in Table S3 in the supplementary section of Article II, confirming the superiority of our machine learning approach.

Furthermore, in Section 3.1 of Article II, we compared the three interpolation methods, assuming each dataset underwent the same slice repositioning procedure introduced in Article II. However, it should be noted that this is an ideal scenario and is not typically encountered in practice.

To measure the achievable improvement of our machine learning-based interpolation and slice repositioning method compared to classical interpolation methods, we analyzed three different cases: DNN-s, SRCS-s, and RML-s datasets (see Table 1 of Article II).

Table 3 from Article II illustrates that while the current state-of-the-art (SRCS-s) outperforms no sophisticated post-processing (DNN-s), it falls short of the combined repositioning and machine learning-based interpolation method introduced in this study (RML-s). These results, based on mean values and standard deviations across 250 datasets, highlight the significant advantages of our combined approach. Additionally,

Table S4 in the supplementary material of Article II reaffirms the superior performance of our method through anisotropy-based error measures.

Figure 7 from Article II visually depicts the predictions obtained in the three cases above, providing further insight into the comparative performance of the different interpolation approaches. However, it is imperative to note that the results presented thus far were confined to synthetic imaging data. This was primarily due to the availability of ground truth data, enabling an accurate performance comparison.

In the subsequent phase, our focus shifts towards measuring the performance of our combined repositioning and ML-based interpolation method using a real HNPG dataset. To this end, we compared five imaging datasets derived from real HNPG data: DNN-r, SRCS-r (ASV), SRCS-r (ruler), RML-r (ASV), and RML-r (ruler) datasets. Like Section 3.2 from Article II, DNN-r represents a scenario where FIB tomography data was generated without sophisticated post-processing for correcting slice thicknesses. At the same time, SRCS-r (ASV) and SRCS-r (ruler) represent state-of-the-art post-processing methods involving slice repositioning on a grid with subsequent cubic splines interpolation. Similarly, RML-r (ASV) and RML-r (ruler) represent cases where our novel combination of ML-based interpolation and slice repositioning is employed, with slice thickness information derived from either the ASV or the ruler-based method, respectively.

The comparison results, detailed in Table 4 and Figure 8 from Article II, are based on indirect quantification using anisotropy measures defined in Section 2.4.2, as ground truth data is unavailable for the real HNPG dataset. As expected, the results reveal that state-of-the-art post-processing (SRCS-r (ruler)) yields superior performance compared to no post-processing (DNN-r). Notably, the superior performance of the ruler-based method, evidenced by the SRCS-r (ruler) and RML-r (ruler) datasets, highlights its capability to measure slice thicknesses with minimal errors.

Moreover, when using slice thickness information from the same source, our novel combination of ML-based interpolation and slice repositioning (RML-r (ruler)) outperforms the state-of-the-art (SRCS-r (ruler)), as shown in Table 4 and Figure 8 from Article II.

Interestingly, all considered L_2 errors are higher for the RML-r (ASV) and SRCS-r (ASV) datasets than any other dataset, including the DNN-r dataset, which uses a simplistic post-processing approach. Upon careful investigation, we discovered that the ASV method measured thicknesses equal to or less than 1 nm for a notably large number of slices (Figure 3C from Article II), suggesting substantial inaccuracies in the

ASV thickness calculation method. Hence, caution should be exercised when using ASV to measure thicknesses.

4.3 Contribution to the field

- A novel slice repositioning-based method to move FIB tomography data slices to their actual correct position helps to build a first step towards making voxel resolution equal in the x-, y-, and z-directions.
- Generating missing information using the concept of image inpainting instead of image super-resolution helps adapt even for a dataset with information missing at inconsistent intervals.
- A comparison between different thickness measurement techniques and an extensive comparison of different interpolation methods such as nearest-neighbor, cubic splines, and our machine learning method can help researchers select required procedures for accurate 3D reconstruction after semantic segmentation.

4.4 Problems yet to be addressed

1. After mitigating inaccuracies in slice thicknesses, more focus can be given to making semantic segmentation methods accurate. Machine learning techniques using additional information from different sensors can help achieve better segmentation performance.
2. The Difference between data distribution is still a challenge and can be solved using machine learning-based domain adaptation techniques.

4.5 Article II - Role of slice thickness quantification in the 3D reconstruction of FIB tomography data of nanoporous materials

T. Sardhara, A. Shkurmanov, Y. Li, S. Shi, C.J. Cyron, R.C. Aydin, M. Ritter: **Role of slice thickness quantification in the 3D reconstruction of FIB tomography data of nanoporous materials.** Ultramicroscopy 256 (2024) 113878 (DOI: 10.1016/j.ultramic.2023.113878)*

Ultramicroscopy 256 (2024) 113878

Contents lists available at ScienceDirect

Ultramicroscopy

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

Role of slice thickness quantification in the 3D reconstruction of FIB tomography data of nanoporous materials

Trushal Sardhara [a,1], Alexander Shkurmanov [b,1], Yong Li [c,1], Shan Shi [d,e,1], Christian J. Cyron [a,f,1], Roland C. Aydin [f,1,2], Martin Ritter [b,1,2,*]

[a] Institute for Continuum and Material Mechanics, Hamburg University of Technology, Hamburg, Germany
[b] Electron Microscopy Unit, Hamburg University of Technology, Hamburg, Germany
[c] Materials Physics and Technology, Hamburg University of Technology, Hamburg, Germany
[d] Research Group of Integrated Metallic Nanomaterials Systems, Hamburg University of Technology, Hamburg, Germany
[e] Institute of Materials Mechanics, Helmholtz-Zentrum Hereon, Geesthacht, Germany
[f] Institute of Material Systems Modeling, Helmholtz-Zentrum Hereon, Geesthacht, Germany

ARTICLE INFO

Dataset link: TUHH Open Research repository

Keywords:
FIB
Slice thickness determination
Slice repositioning
Accurate reconstruction
Image inpainting

ABSTRACT

In focused ion beam (FIB) tomography, a combination of FIB with a scanning electron microscope (SEM) is used for collecting a series of planar images of the microstructure of nanoporous materials. These planar images serve as the basis for reconstructing the three-dimensional microstructure through segmentation algorithms. However, the assumption of a constant distance between consecutively imaged sections is generally invalid due to random variations in the FIB milling process. This variation complicates the accurate reconstruction of the three-dimensional microstructure. Using synthetic FIB tomography data, we present an algorithm that repositions slices according to their actual thickness and interpolates the results using machine learning-based methods. We applied our algorithm to real datasets, comparing two standard approaches of microstructure reconstruction: *on-the-fly* via image processing and *ruler-based* via sample structuring. Our findings indicate that the *ruler-based* method, combined with our novel slice repositioning and interpolation algorithm, exhibits superior performance in reconstructing the microstructure.

1. Introduction

Nanoporous materials promise to have a transformative effect on various fields of science and technology, such as drug development, chemical catalysis, and sensor and actuator technology. To study their physical properties, it is essential to reconstruct their microstructure accurately. FIB (focused ion beam) tomography can be used to collect high-resolution SEM (scanning electron microscope) images of consecutive layers of the material. Thereby, a specific layer is imaged, which is subsequently removed with the FIB, and the next layer (in the z-direction) is imaged [1]. By repeating this procedure, a stack of planar images is acquired, which can be used to reconstruct the three-dimensional microstructure. The in-plane resolution of this process is approximately 1 nm, and the resolution in depth is around 10 nm (slice thickness) [2]. FIB tomography is used in many disciplines such as materials science, nano-mechanics, biology, and geology for imaging [3–5] or in combination with different analytical information such as 3D crystal orientation [6]. Different techniques have been applied to reconstruct the three-dimensional microstructure of HNPG from FIB tomography data [7–9]. However, reconstructing the three-dimensional microstructure of materials from FIB tomography data faces two problems, particularly in nanoporous materials.

The first one is the *shine-through effect*. When specific layers of nanoporous materials are imaged, material from deeper layers often shines through the pores and contaminates the images taken from layers on top of them. Therefore, there is no unique relation between the recorded image intensities and the actual structure for any specific layer [10]. This ambiguity complicates the accurate reconstruction of the three-dimensional structure of the material from layer-by-layer imaging. As a potential solution, convolutional neural networks (CNNs) were proposed by Sardhara et al. [11] to accurately segment and reconstruct epoxy-infiltrated hierarchical nanoporous gold (HNPG).

The second major source of uncertainties results from the FIB milling process. For example, due to the beam shape [12], charging

* Corresponding author.
E-mail address: ritter@tuhh.de (M. Ritter).
[1] All authors contributed to the manuscript.
[2] Aydin and Ritter contributed equally to the manuscript.

https://doi.org/10.1016/j.ultramic.2023.113878
Received 30 June 2023; Received in revised form 13 October 2023; Accepted 21 October 2023
Available online 25 October 2023

effects [13], high surface roughness [14] and sample drift [15] of the FIB, there are substantial variations in the thicknesses of the milled slices [16,17] and significant deviations from the predefined target thickness. For an accurate reconstruction of the three-dimensional structure of the material, it is vital to quantify the thickness of each slice rather than simply assume it to be equal to the predefined target thickness as the accuracy of the reconstruction process is strongly influenced by variations in the thickness of the milled slices [9,18].

Mangipudi et al. [7] studied the effect of varying slice thicknesses on the accuracy of reconstruction for nanoporous gold (NPG) based on microstructural descriptors such as connectivity of ligaments, pore size distribution and pair correlation function. Their study examined two different parts of the same nanoporous gold sample with different microscopes to obtain data for comparison. Tricubic interpolation [19] was used to achieve isometric resolutions.

In this study, our first objective is to investigate the impact of varying slice thicknesses on the accuracy of reconstructing hierarchical nanoporous gold using synthetic FIB tomography images. We generate these synthetic FIB tomography images following the procedure outlined in [11]. In the subsequent step, we show how the imaging data can be corrected by repositioning the imaged slices in the z-direction using existing thickness information. We generate a total of two datasets of the same hierarchical nanoporous gold structure using two different slice thickness determination methods, specifically the Auto Slice&View (ASV) *on-the-fly* method [20] and the *ruler-based* method by Shkurmanov et al. [21]. Repositioning of slices creates gaps in the three-dimensional stack of slices in the z-direction. To close these gaps, we use machine learning-based methods for inpainting (i.e., generating missing parts in the images). We validate our approach on the synthetic FIB tomography dataset, which mimics the real FIB tomography dataset in terms of structure and varying thicknesses. Ultimately, we calculate the isotropy of all generated nanostructures and demonstrate that our combination of methods improves overall reconstruction performance by more than 25% on average compared to approaches where the problem of variable slice thickness is either ignored or addressed using the established ASV *on-the-fly* method.

2. Material and methods

2.1. Acquisition of imaging data

We used both imaging data from real hierarchical nanoporous gold (HNPG) samples as well as computer-generated synthetic imaging data mimicking the real HNPG imaging data to augment our data basis for machine learning.

2.1.1. Real samples

Hierarchical nanoporous gold characterized with a uniform random network and two well defined ligament sizes (15 nm and 110 nm) was prepared by a dealloying-coarsening-dealloying method, as described in [22]. For FIB tomography, HNPG samples were infiltrated with epoxy resin to provide additional support during FIB milling [23] and to reduce the *shine-through effect*. Fig. 1 shows the SEM image of a single slice from a FIB tomography dataset of an HNPG sample. Generally, for HNPG, the slice thickness should be considerably smaller than the pore size, i.e., it should not exceed 10 nm in our case. Therefore, we chose the following FIB tomography parameters for the ion-beam image. We used an in-plane pixel size of 8.4 nm corresponding to a resolution of 3072×2048 pixels in the field of view with a horizontal length of 25.9 μ m. The dwell time was chosen as 750 ns per pixel. Along the depth direction (denoted as the z-direction), approximately 200 slices were milled with a target thickness of 10 nm.

In recent years, researchers have explored various approaches for estimating slice thickness in FIB tomography after the milling process. Among these approaches, fiducial markers and ruler markings have emerged as widely adopted methods. Furthermore, image-based

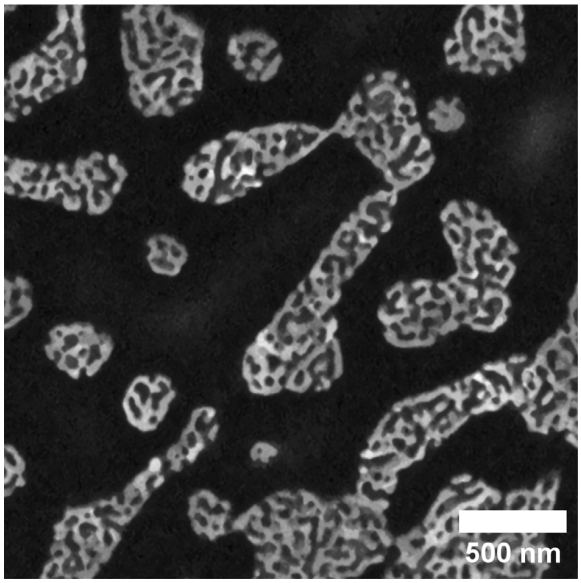


Fig. 1. SEM backscattered electron image of a FIB cross-section of infiltrated hierarchical nanoporous gold (HNPG) using an accelerating voltage of 2 kV.

thickness correction techniques have been proposed, including the use of Gaussian processes for estimating thicknesses in serial sectioning electron microscopy [24,25]. In this study, we focus on fiducial markers and ruler markings as state-of-the-art methods, considering their extensive usage in the field. It is worth noting that we employed these methods to analyse our imaging data and remove outliers.

ASV using fiducial markers: The utilization of fiducial markers, which are placed on the samples and tracked along the milled slices [15,26], enables *on-the-fly* correction of drift between consecutive milling operations (Fig. 2 C). Various electron microscope manufacturers have integrated this approach in their FIB tomography software, e.g., in our case Thermo Fisher Scientific with Auto Slice&View (ASV) [20].

For the HNPG sample studied herein, the fiducial marker was formed by two crossed trenches milled on a platinum (Pt) deposited area separated from the region of interest (ROI). FIB tomography was performed using an FEI HeliosNanoLab G3 FIB-SEM. The auto slice-and-view software ASV4 [20], providing automated control of the tomography process, was integrated into this system. This software employs the ion-beam image (top view image) to estimate the positioning of the beam for milling the subsequent slice. It utilizes this information to provide thickness estimates for each slice, as depicted in Fig. 3 B and C, for the real HNPG sample analysed in this article.

***Ruler-based* method:** As an alternative to fiducial markers, ruler markings in or around the sample can also be used to estimate the thickness of the milled slices. These ruler markings offer a reliable method for ensuring slice-to-slice alignment in the xy-plane during reconstruction while providing slice thickness measurements in the z-direction with minimal error [17]. Over the past decade, this approach has successfully been applied to various material systems ranging from silicon wafers to hierarchical nanoporous gold (HNPG) [21,27,28]. While recent studies have demonstrated the use of wedge-shaped sample geometry as an implicit ruler, eliminating the need for explicit ruler markings [7], we focus on the classical approach of employing explicit ruler markings in this study. In this approach, the ruler can be formed by two parallel milled trenches and two additional trenches inclined at a certain angle with respect to the parallel ones. These trenches are filled with a platinum/carbon (Pt/C) deposition layer (Fig. 2 A, B). This makes their outline clearly visible in the cross-sectional images

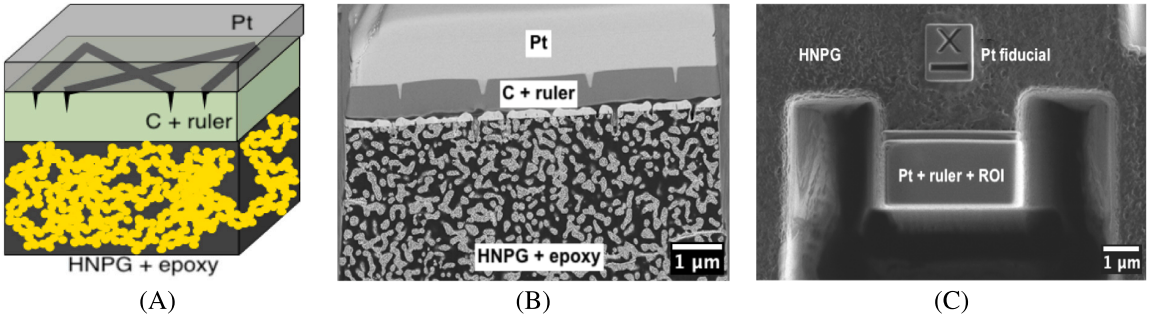


Fig. 2. Ruler in **(A)** schematic illustration and **(B)** real hierarchical nanoporous gold sample (implemented using carbon and platinum); **(C)** top view of the hierarchical nanoporous gold structure with fiducial marker used by ASV.

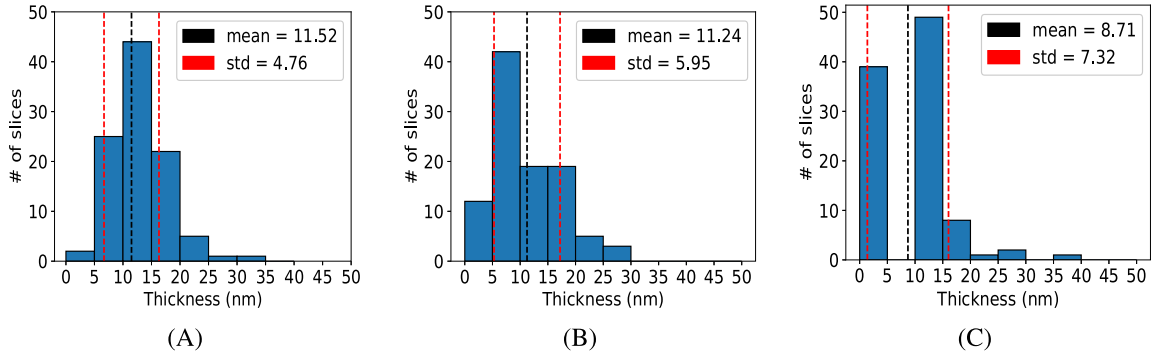


Fig. 3. Histograms of slice thicknesses variations of a **(A)** synthetically made HNPG structure using Gamma distribution, **(B)** real HNPG sample as measured by the *ruler-based* method and **(C)** real HNPG sample as measured by the ASV4 method.

collected during FIB tomography. In consecutive slices, the markers resulting from the two parallel trenches always remain at the same position, whereas the two markers from the inclined trenches alter their in-plane position. The actual thickness of the milled slices can be computed by comparing the in-plane position of these markers in consecutive slices.

2.1.2. Synthetic samples

To augment the data basis for machine learning, we generated, in addition to the real samples, also synthetic sample data. To this end, we used the levelled wave method (LWM) described by Li et al. [29] to generate synthetic microstructures similar to real HNPG [22]. This dataset, henceforth referred to as the *original* dataset, had the same voxel size (2.5 nm) in all three directions. We used it to prepare an additional dataset that mimicked real FIB tomography data in terms of varying slice thicknesses. This *mimicking* dataset had a target slice thickness in the z-direction of 10 nm with a realistic gamma distribution of the actual thickness. The shape (α) and scale (β) parameters of that gamma distribution were calculated by fitting the gamma distribution to the average of the thickness distributions measured by the *ruler-based* and the ASV4 method. The *mimicking* dataset was generated by repeating the following two steps for each of its slices: first, a slice thickness $N_s \times 2.5$ nm with $N_s \in \mathbb{N}$ was drawn from a gamma distribution discretized into bins of width 2.5 nm; second, the slice was assigned the geometry of the slice in the *original* dataset starting at the same z-coordinate. Fig. 3 A shows the final slice thickness distribution of the *mimicking* dataset.

2.2. Processing of imaging data

The imaging data from the real samples were subjected to all processing steps described in the following, while the synthetic samples were only subjected to the final two steps.

2.2.1. Semantic segmentation

To binarize the greyscale images into a solid and a void phase, we used semantic segmentation. In an earlier study [11], various segmentation methods were compared, including machine learning-based approaches specifically designed to suppress the *shine-through effect*. The results of the study indicated that machine learning-based methods were highly effective in mitigating this effect. Therefore, we adopted the machine learning-based method for the semantic segmentation of our imaging data. The subsequent steps relied on these binarized data only.

2.2.2. Slice repositioning

In the binarized (segmented) imaging data, we repositioned every slice in z-direction to its correct location according to its measured thickness (Fig. 4), rounding to a multiple of some discrete spacing parameter (in our case 2.5 nm). In certain cases, such rounding may place multiple slices in the same position. Where this was the case, overlapping slices were distributed to the nearest free positions to prevent a complete loss of the associated microstructural information.

2.2.3. Interpolation for isometric resolution

The varying slice thickness leads to data where the resolution in z-direction is generally lower than that in the SEM plane and also varies in z-direction. However, to accurately evaluate quantities such as two-point correlation and lineal path functions, it is necessary to project FIB tomography data on a grid with uniform spacing equal in the three spatial directions [11]. To this end, we use interpolation in z-direction, yielding a grid spacing in z-direction comparable to the resolution in the SEM plane. Various interpolation techniques have been proposed for the post-processing of FIB tomography, including different classical and machine learning-based techniques [30,31].

For image interpolation, cubic splines [32] are frequently used. That is, cubic splines are fitted to the available imaging data in each zx-plane or alternatively in each zy-plane (which typically yields, in the

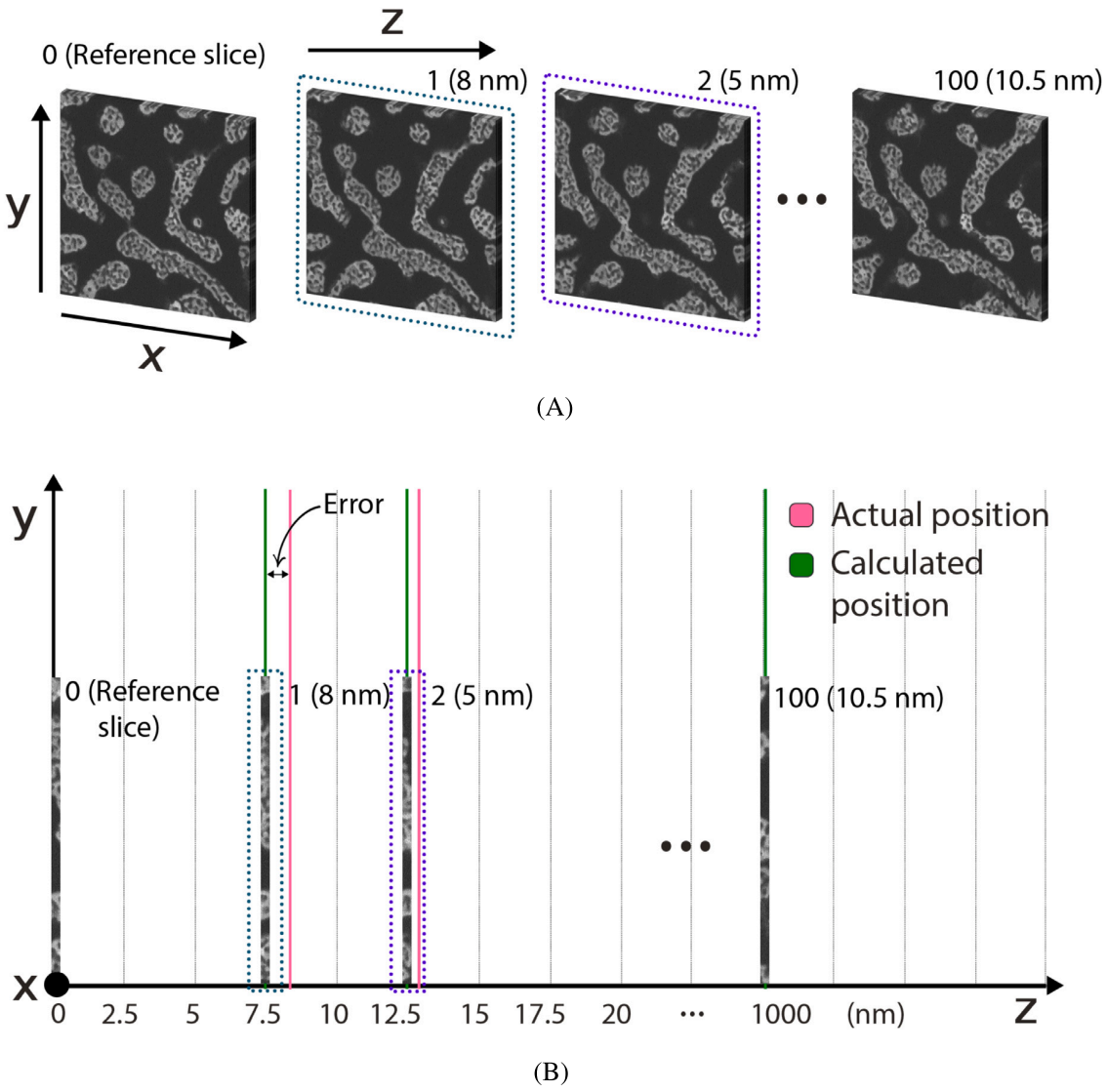


Fig. 4. **(A)** Slices with associated measured thicknesses and **(B)** repositioned slices for a grid spacing of 2.5 nm in the *yz*-plane. 0th slice is considered as a reference slice, and the next slices are repositioned to calculated positions represented using vertical green lines from their actual positions (pink line).

end, nearly identical results). Subsequently, the fitted splines are used to calculate the intensities of the missing voxels between the available data.

When using machine learning for interpolation, different approaches can be taken. The underlying problem is often interpreted as an image super-resolution problem [30]. However, machine learning can address not only super-resolution but also image-inpainting problems quite well [33,34]. Here, we decided to interpret the problem of approximating missing values due to variable slice thickness in *z*-direction as an image-inpainting (impainting) problem. That is, after segmenting the imaging data (Section 2.2.1) and repositioning slices (Section 2.2.2), the voxels between the repositioned slices are interpreted as missing image data (red bars in Fig. 5 B). To fill these gaps in the image, an impainting problem is solved (Fig. 5 C). To do so, we trained a machine learning architecture to fill the gaps in the image. The training data was formed by the set of all *zx*- and *zy*-planes that could be extracted from the synthetic imaging data (for which ground truth was available so that supervised machine learning was possible). Details of the architecture and training are provided in the supplementary material.

2.3. Image datasets for method validation and testing

Synthetic datasets for comparing interpolation methods: To compare different interpolation methods, we generated synthetic datasets following the steps described in Sections 2.1.2, 2.2.1 and 2.2.2. Subsequently, three interpolation methods were employed to generate the missing slices (see discussion in Section 2.2.3). The synthetic *repositioned nearest-neighbour interpolation* (RNN-*s*) dataset was generated using a basic nearest-neighbour interpolation approach, while the synthetic *repositioned cubic spline interpolation* (RCS-*s*) dataset was generated using cubic spline interpolation in the *zx*-plane. The synthetic *repositioned machine learning-based interpolation* (RML-*s*) dataset, on the other hand, was generated using our novel machine learning-based interpolation technique.

Synthetic datasets for evaluating slice repositioning: To evaluate slice repositioning, we generated two additional types of synthetic datasets with an isometric resolution ($2.5 \times 2.5 \times 2.5$ nm^3) from the *mimicking* dataset. This *mimicking* dataset contained varying slice thicknesses based on a realistic gamma distribution (Section 2.1.2). The first

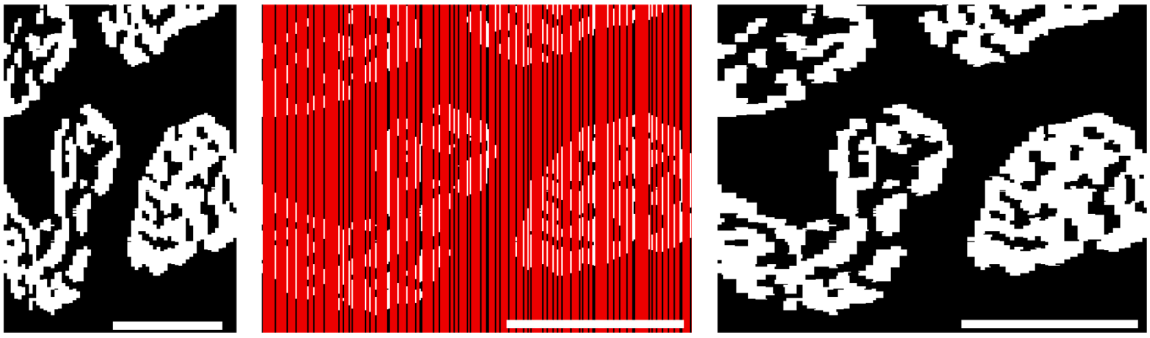


Fig. 5. Segmented imaging data (zx-plane) **(A)** before repositioning of slices and **(B)** after repositioning of slices with missing regions in red; **(C)** image where in the missing regions the image was completed by the machine learning method developed herein — scale bar: 200 nm.

dataset is referred to as the synthetic *direct NN interpolation* (DNN-*s*) dataset. It was generated by interpolating values in z-direction using the nearest-neighbour interpolation method without any prior slice repositioning. That is, the slices with the available data were all assumed to exhibit an exact spacing of 10 nm in z-direction, which is, in theory, the intention of the FIB milling process but, in practice, not achieved as discussed above. The second dataset is referred to as the synthetic *standard repositioned cubic spline interpolation* (SRCS-*s*) dataset. To generate it, first, slices were repositioned on a grid with a spacing of 10 nm. Subsequently, cubic spline interpolation with a resolution of 2.5 nm in z-direction was applied for missing slice generation.

Real datasets: From real HNPG FIB images, we generated five binarized datasets with an isometric resolution of $2.5 \times 2.5 \times 2.5$ nm^3 to test our method consisting of repositioning and interpolation. The first three datasets are referred to as real *direct NN interpolation* (DNN-*r*), real *standard repositioned cubic spline interpolation (ASV)* (SRCS-*r (ASV)*) and real *standard repositioned cubic spline interpolation (ruler)* (SRCS-*r (ruler)*) datasets. They were generated analogously to the DNN-*s* and SRCS-*s* datasets, considering thickness calculation using ASV *on-the-fly* method and *ruler-based* method for the last two, respectively. The only difference was that the generation process did not start from the synthetic mimicking dataset but rather from real HNPG FIB imaging data. The other two datasets are referred to as real *repositioned ML interpolation* (RML-*r (ASV)*) and real *repositioned ML interpolation* (RML-*r (ruler)*) datasets. They were generated using slice repositioning, with thickness calculation using the ASV *on-the-fly* method and *ruler-based* method, respectively. Then the missing slices were generated with our ML-based interpolation method. Table 1 summarizes all the generated datasets.

2.4. Evaluation criteria for 3D image reconstruction

2.4.1. Metrics based on ground truth values

In this study, exact microstructure (ground truth) data are available for all synthetic datasets, allowing us to use absolute error metrics to evaluate the accuracy of reconstructed 3D structures. We calculated three metrics, the fraction of total misplaced pixels (MP), the fraction of total misplaced gold pixels (MGP) and the mean Dice score (MDS), to evaluate the reconstruction. The first metric is calculated as

$$MP = \left(1 - \frac{TP + TN}{TP + FP + FN + TN}\right) \times 100 \tag{1}$$

where TP is the number of true positive pixels, FN the number of false negative pixels, TN the number of true negatives, and FP the number of false positive pixels. MP measures the percentage of pixels where the prediction of synthetic datasets and associated ground truth images disagree. The second metric is calculated as

$$MGP = \left(1 - \frac{TP}{TP + FN}\right) \times 100 \tag{2}$$

MGP measures the percentage of falsely predicted gold pixels compared to ground truth images. MGP is especially useful when the data is imbalanced, like in hierarchical nanoporous gold, where the solid phase covers significantly less volume than the pore phase.

The Dice score (DS) measures the similarity between two sets of (pixel or voxel) data, in this case, the reconstructed and ground truth datasets. It ranges from 0 to 1, with 1 indicating perfect overlap between the two datasets. We calculated the Dice score (DS) for both phases (solid and pore) separately by

$$DS = \frac{2TP}{2TP + FN + FP}. \tag{3}$$

Subsequently, we averaged the Dice scores of the solid and pore phase to obtain the mean Dice score (MDS) as an overall error measure.

2.4.2. Metrics based on anisotropy in the absence of ground truth values

In many cases, including the real HNPG datasets, we do not know the exact microstructure (ground truth). Therefore, to assess the quality of image reconstruction for such microstructures, we use the concept of anisotropy as explained in [11]. That is, we know that the real microstructure can be assumed to be isotropic. Therefore, deviations of the reconstructed image from isotropy can be considered a measure of the reconstruction error. To measure such deviations, we rely on the datasets with isometric resolution and calculate three metrics. The first one relies on the two-point correlation function (TPCF). Let this function – evaluated in the three spatial directions – be denoted by f^x, f^y, and f^z. Discretizing these functions uniformly with n data points yields function values f_i^x, f_i^y, and f_i^z with $i = 1, \dots, n$. Now we compute the normalized L_2-difference between the TPCF in z-direction and x-direction on the one hand and in z-direction and y-direction on the other hand, and average both. This yields:

$$e_{L_2}^{TPCF} = \frac{1}{2}\left(\frac{2 \times \sqrt{\sum_{i=1}^n (f_i^x - f_i^z)^2}}{\sqrt{\sum_{i=1}^n (f_i^x)^2} + \sqrt{\sum_{i=1}^n (f_i^z)^2}} + \frac{2 \times \sqrt{\sum_{i=1}^n (f_i^y - f_i^z)^2}}{\sqrt{\sum_{i=1}^n (f_i^y)^2} + \sqrt{\sum_{i=1}^n (f_i^z)^2}}\right) \tag{4}$$

The above function equals zero if applied to an isotropic dataset where the TPCF is equal in all three spatial directions. By contrast, higher values of $e_{L_2}^{TPCF}$ generally indicate a higher anisotropy in the underlying geometric data. The normalization in Eq. (4) ensures that $e_{L_2}^{TPCF} \in [0; 1]$. An analogous measure of anisotropy can also be defined based on the lineal path function (LPF). We refer to that measure as $e_{L_2}^{LPF}$. One can compute it analogously to Eq. (4) by interpreting the function values f_i^x, f_i^y, and f_i^z as values of a (discretized) lineal path function in the three spatial directions. As a third metric to assess anisotropy, we compute

$$e_{L_2}^{D} = \frac{1}{2}\left(\sqrt{\frac{(D_{xz} - D_{xy})^2}{D_{xy}^2}} + \sqrt{\frac{(D_{yz} - D_{xy})^2}{D_{xy}^2}}\right) \tag{5}$$

Table 1
Description of real HNPG and synthetic datasets with isometric resolution ($2.5 \times 2.5 \times 2.5$ nm^3) obtained from real HNPG and synthetic *mimicking* datasets ($2.5 \times 2.5 \times 10$ nm^3) respectively.

Dataset name	Thickness information	Reposition grid (nm)	Repositioning (Section 2.2.2)	Interpolation method
RNN-*s*	Gamma distribution	2.5	✓	Nearest-neighbour
RCS-*s*	Gamma distribution	2.5	✓	Cubic splines
RML-*s*	Gamma distribution	2.5	✓	Our ML method
DNN-*s*	-	-	-	Nearest-neighbour
SRCS-*s*	Gamma distribution	10	✓	Cubic splines
DNN-*r*	-	-	-	Nearest-neighbour
SRCS-*r (ruler)*	Ruler	10	✓	Cubic splines
SRCS-*r (ASV)*	ASV	10	✓	Cubic splines
RML-*r (ASV)*	ASV	2.5	✓	Our ML method
RML-*r (ruler)*	Ruler	2.5	✓	Our ML method

Note: "–" means "not applicable" or "not applied"; "Our ML method" is explained in Section 2.2.3.

Here, D_{ij} is the average diameter of the ligaments in the ij-plane with $i, j \in \{x, y, z\}$. Also, $e^D_{L_2}$ equals zero for isotropic geometric data, and increasing values indicate an increasing anisotropy. Again, the normalization ensures $e^D_{L_2} \in [0; 1]$. We calculated TPCF, LPF and D_{ij} for image patches of 256×256 pixels as a convergence study revealed that this is sufficient to ensure a negligible error from the finite-size effects.

3. Results and discussion

3.1. Comparison of interpolation methods

In this study, we compared the performance of classical interpolation methods (such as cubic spline and nearest neighbour interpolation) with our machine learning-based interpolation method. To do so, we generated 250 different *mimicking* datasets from 25 different LWM datasets and 10 different thickness distributions per LWM dataset. Then we applied three interpolation methods to each of these *mimicking* datasets. We found that the cubic spline method sometimes created holes in the ligaments, which we then fixed using image processing techniques to obtain the best result possible with cubic spline interpolation by a reasonable effort (see supplementary section). We underline that the data we used for the interpolation was completely separate from the data used to train the machine learning model. The error measures introduced in Section 2.4.1 calculated for each interpolated dataset are reported in Table 2. The ML-based interpolation method introduced herein apparently outperforms both classical interpolation methods. Additionally, we also calculated the error measures introduced in Section 2.4.2 for each interpolated dataset. The means and standard deviations of these errors are presented in Table S3 in the supplementary section. They confirm the superiority of the ML-based interpolation introduced herein. Fig. 6 illustrates the results of the different interpolation methods for two regions of an example microstructure.

Table 2
Mean errors (and standard deviations) of images reconstructed from synthetic data using nearest neighbour (NN), cubic spline (CS), and machine learning based (ML) interpolation. The errors are measured relative to the ground truth (original dataset). For the sake of clarity, the second line of the table reports the considered error measures for the hypothetical case of an exact reconstruction of the original dataset.

Dataset	MP ↓	MGP ↓	MDS ↑
Original	0	0	1
RNN-*s*	4.89 ± 0.53	19.98 ± 2.07	$0.89 \pm 0.12 \times 10^{-1}$
RCS-*s*	3.12 ± 0.11	13.52 ± 0.46	$0.93 \pm 0.23 \times 10^{-2}$
RML-*s*	2.35 ± 0.11	10.04 ± 0.42	$0.95 \pm 0.25 \times 10^{-2}$

3.2. Comparison of combined repositioning and interpolation (synthetic data)

In Section 3.1, we compared three different interpolation methods under the assumption that they are all combined with the particularly accurate method for slice repositioning introduced in this paper. In reality, however, this is typically not the case.

To get an impression of how much improvement can be obtained by applying both the novel ML-based interpolation and particularly accurate slice repositioning introduced in this paper compared to standard approaches widely used so far in practice, we, therefore, compare the following three cases: The first case is represented by DNN-*s* data. It resembles a situation where no slice repositioning at all is applied, and data between available slices are added in the most simplistic manner, that is, by nearest neighbour interpolation. This case can represent a situation where no sophisticated algorithmic post-processing is applied to available FIB tomography data. The second case is represented by the SRCS-*s* dataset. It represents a situation where state-of-the-art algorithms are used for post-processing, that is, standard repositioning with target thickness (10 nm in our case) spacing together with cubic spline interpolation. This resembles the post-processing quality that is currently often achieved in practice. The third case is represented by the RML-*s* dataset. It illustrates the quality that can be achieved in post-processing by combining both the accurate repositioning and ML-based interpolation methods introduced in this paper.

Table 3 reveals that the current state-of-the-art (second case) performs better than no sophisticated post-processing at all (first case) but worse than the combined repositioning and ML-based interpolation method introduced in this paper (third case). The numbers reported in Table 3 represent mean values and standard deviations for 250 different datasets (created from 25 different LWM microstructures with 10 different slice thickness distributions each). Table S4 in the supplementary material demonstrates that the anisotropy-based error measures introduced in Section 2.4.2 also confirm the superior performance of the

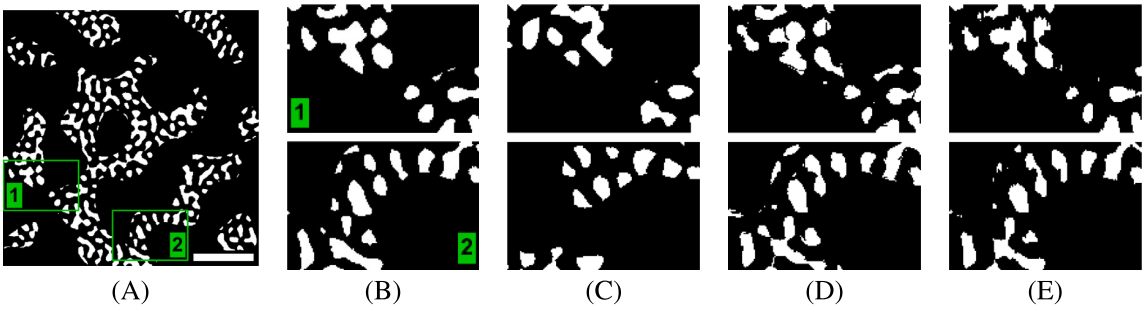


Fig. 6. **(A)** Slice of a synthetic example microstructure with two regions highlighted and **(B)** zoomed in; reconstruction of this ground truth using **(C)** nearest neighbour (NN), **(D)** cubic spline (CS), and **(E)** machine learning-based (ML) interpolation — scale bar: 300 nm.

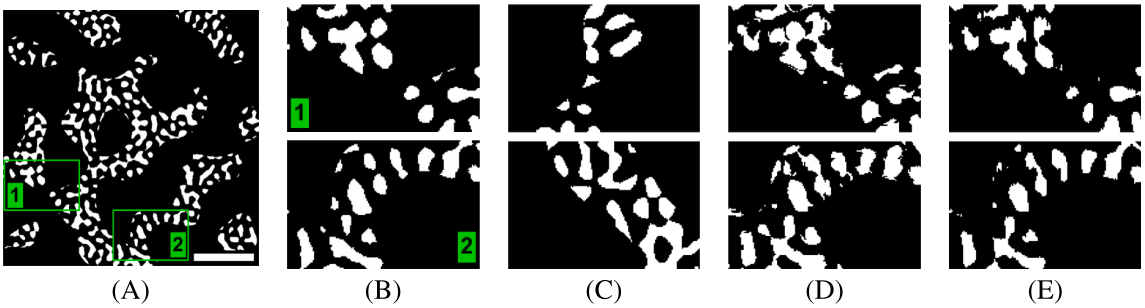


Fig. 7. **(A)** Slice of a synthetic example microstructure with two regions highlighted and **(B)** zoomed in; reconstruction of this ground truth using **(C)** no repositioning and nearest neighbour interpolation (DNN-*s* dataset), **(D)** standard repositioning and cubic spline interpolation (SRCS-*s* dataset), and **(E)** our repositioning method combined with machine learning based interpolation (RML-*s* dataset) - scale bar: 300 nm.

Table 3
Summarized results of an experiment investigating the effect of varying slice thickness on 250 different synthetically interpolated datasets, with and without the repositioning slices approach.

Dataset	MP ↓	MGP ↓	MDS ↑
Original	0	0	1
DNN-*s*	19.97 ± 0.91	81.62 ± 3.18	0.53 ± 0.02 $\times 10^{-1}$
SRCS-*s*	5.29 ± 0.37	22.38 ± 1.34	0.88 ± 0.79 $\times 10^{-2}$
RML-*s*	2.35 ± 0.11	10.04 ± 0.42	0.95 ± 0.25 $\times 10^{-2}$

combined accurate repositioning and ML-based interpolation methods introduced herein. Fig. 7 illustrates the results obtained in the three cases studied in this subsection.

3.3. Comparison of performance for real HNPG data

In the previous two subsections, we have shown that our novel ML-based interpolation method is superior to established nearest neighbour or cubic spline interpolation. We have also shown that combining our novel ML-based interpolation method with the particularly accurate repositioning method introduced in this paper brings considerable advantages compared to state-of-the-art processing. However, the results shown in the previous two subsections are for synthetic imaging data only. The reason for that was the availability of the ground truth for such data; therefore, a particularly reliable comparison of the performance is possible.

In the next step, we examine the performance of our novel combined repositioning and ML-based interpolation method for real-world data. To this end, we compare five imaging datasets based on real-world data: the DNN-*r*, SRCS-*r (ASV)*, SRCS-*r (ruler)*, RML-*r (ASV)* and RML-*r (ruler)* dataset. Similarly to Section 3.2, DNN-*r* represents a situation where FIB tomography data is used without any sophisticated post-processing in the most direct manner. SRCS-*r (ASV)* and SRCS-*r (ruler)* represent a real dataset post-processed in a state-of-the-art manner (repositioning of slices on 10 nm grid and cubic-spline interpolation). By contrast, RML-*r (ASV)* and RML-*r (ruler)* represent a case where we apply the accurate slice repositioning and novel ML-based interpolation introduced in this paper in a manner where the slice thickness information is either obtained from the ASV method or the ruler method. The results are compared in Table 4 and Fig. 8. Since no ground truth is available in this case, the quality of the results is indirectly quantified using the anisotropy measures defined in Section 2.4.2. As expected, state-of-the-art post-processing (SRCS-*r (ruler)*) yields better results than no advanced post-processing at all (DNN-*r*). The superior performance for SRCS-*r (ruler)* and RML-*r (ruler)* datasets show that the *ruler-based* method can measure slice thicknesses with minimal errors. Moreover, provided slice thickness information from the same source, our novel combination of ML-interpolation and enhanced slice repositioning (RML-*r (ruler)*) outperforms the state-of-the-art (SRCS-*r (ruler)*).

Interestingly, all considered errors are higher for the RML-*r (ASV)* and SRCS-*r (ASV)* datasets than for any other dataset, even the DNN-*r* dataset that is based on a very simple procedure without advanced post-processing. When examining the reasons for this phenomenon, we found that the ASV method yielded thicknesses less than or equal to 1 nm for an unrealistically large number of slices (Fig. 3 C), suggesting substantial inaccuracies of the ASV thickness calculation method. The ASV-based data should thus be taken with caution.

4. Conclusion

During FIB tomography, the thickness of the slices sequentially ablated by milling is typically larger than the image resolution achieved in the plane of these slices. Moreover, the slice thickness is typically not constant and may substantially deviate from the desired

Fig. 8. Image reconstruction of an example region as given by the datasets **(A)** DNN-*r*, **(B)** SRCS-*r (ASV)*, **(C)** SRCS-*r (ruler)*, **(D)** RML-*r (ASV)* and **(E)** RML-*r (ruler)* - scale bar: 300 nm.

Table 4
Error measures based on anisotropy of reconstructed microstructure (originally 64 slices) as achieved by four different microstructure reconstruction methods.

Dataset name	$e_{L_2}^{TPCF}\downarrow$	$e_{L_2}^{LPF}\downarrow$	$e_{L_2}^{D}\downarrow$
DNN-*r*	0.24	0.15	0.09
SRCS-*r (ruler)*	0.22	0.07	0.04
SRCS-*r (ASV)*	0.27	0.16	0.10
RML-*r (ASV)*	0.26	0.17	0.10
RML-*r (ruler)*	0.21	0.06	0.02

target thickness. Both introduce considerable inaccuracies in the reconstruction of the microstructure, particularly, if a reconstruction with a similar effective resolution in all spatial directions is desired. To address this issue, FIB tomography data are typically post-processed during the reconstruction process. The state-of-the-art post-processing often relies on positioning the recorded slices in space based on slice thickness information obtained either from the *ruler-based* or the ASV method. Subsequently, the microstructure between the recorded slices is interpolated by standard methods such as cubic spline interpolation.

In this paper, we introduced a new post-processing method. It relies on positioning the recorded slices in space using a finer grid than usual so far. Moreover, interpolation between these slices is not performed in a classical fashion (e.g., using cubic splines) but rather by a machine learning architecture specifically trained for this purpose and treating the interpolation as a sort of impainting problem.

Using both synthetic imaging data obtained from levelled wave method as well as real FIB tomography data, we demonstrated that the novel post-processing method we suggest can outperform the state-of-the-art. It is important to note that additional performance studies should be conducted for various microstructure types to validate and corroborate these findings. If future investigations confirm the trends observed in our research, the post-processing pipeline presented in this paper holds great potential for effectively reconstructing complex microstructures from FIB tomography data.

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Germany – SFB 986 – Project number 192346071.

CRediT authorship contribution statement

Trushal Sardhara: Conception and design of the study, Wrote the first draft of the manuscript. **Alexander Shkurmanov:** Acquired FIB-SEM images of HNPG and calculated slice thicknesses. **Yong Li:** Generated the LWM database. **Shan Shi:** Prepared the HNPG sample. **Christian J. Cyron:** Conception and design of the study. **Roland C. Aydin:** Conception and design of the study. **Martin Ritter:** Conception and design of the study.

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.

Data availability

The datasets generated and analyzed for this study can be found in TUHH Open Research repository.

Appendix A. Supplementary material

Supplementary material associated with this article can be found, in the online version, at TUHH Open Research repository.

References

[1] B. Inkson, M. Mulvihill, G. Möbus, 3D determination of grain shape in a FeAl-based nanocomposite by 3D FIB tomography, Scr. Mater. 45 (7) (2001) 753–758.
[2] G. Knott, H. Marchman, D. Wall, B. Lich, Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling, J. Neurosci. 28 (12) (2008) 2959–2964.
[3] C. Villinger, H. Gregorius, C. Kranz, K. Höhn, C. Münzberg, G. von Wichert, et al., FIB/SEM tomography with TEM-like resolution for 3D imaging of high-pressure frozen cells, Histochem. Cell Biol. 138 (4) (2012) 549–556.
[4] Y. Liu, H.E. King, M.A. Van Huis, M.R. Drury, O. Plümper, Nano-tomography of porous geological materials using focused ion beam-scanning electron microscopy, Minerals 6 (4) (2016) 104.
[5] N. Welch, F. Gray, A. Butcher, E. Boek, J. Crawshaw, High-resolution 3D FIB-SEM image analysis and validation of numerical simulations of nanometre-scale porous ceramic with comparisons to experimental results, Transp. Porous Media 118 (3) (2017) 373–392.
[6] S. Korte, M. Ritter, C. Jiao, P. Midgley, W. Clegg, Three-dimensional electron backscattered diffraction analysis of deformation in MgO micropillars, Acta Mater. 59 (19) (2011) 7241–7254.
[7] K.R. Mangipudi, V. Radisch, L. Holzer, C.A. Volkert, A FIB-nanotomography method for accurate 3D reconstruction of open nanoporous structures, Ultramicroscopy 163 (2016) 38–47.
[8] K. Hu, M. Ziehmer, K. Wang, E.T. Lilleodden, Nanoporous gold: 3D structural analyses of representative volumes and their implications on scaling relations of mechanical behaviour, Phil. Mag. 96 (32–34) (2016) 3322–3335.
[9] Y. Fam, T.L. Sheppard, A. Diaz, T. Scherer, M. Holler, W. Wang, et al., Correlative multiscale 3D imaging of a hierarchical nanoporous gold catalyst by electron, ion and X-ray nanotomography, ChemCatChem 10 (13) (2018) 2858–2867.
[10] C. Fager, M. Röding, A. Olsson, N. Lorén, C. von Corswant, A. Särkkä, et al., Optimization of FIB–SEM tomography and reconstruction for soft, porous, and poorly conducting materials, Microsc. Microanal. 26 (4) (2020) 837–845.
[11] T. Sardhara, R.C. Aydin, Y. Li, N. Piché, R. Gauvin, C.J. Cyron, M. Ritter, Training deep neural networks to reconstruct nanoporous structures from FIB tomography images using synthetic training data, Front. Mater. 9 (2022) http://dx.doi.org/10.3389/fmats.2022.837006, URL https://www.frontiersin.org/articles/10.3389/fmats.2022.837006.
[12] A.A. Tseng, Recent developments in micromilling using focused ion beam technology, J. Micromech. Microeng. 14 (4) (2004) R15.
[13] L. Holzer, F. Indutnyi, P. Gasser, B. Münch, M. Wegmann, Three-dimensional analysis of porous BaTiO3 ceramics using FIB nanotomography, J. Microsc. 216 (1) (2004) 84–95.

[14] B. Winiarski, A. Gholinia, K. Mingard, M. Gee, G. Thompson, P. Withers, Broad ion beam serial section tomography, Ultramicroscopy 172 (2017) 52–64.
[15] M. Schaffer, J. Wagner, B. Schaffer, M. Schmied, H. Mulders, Automated three-dimensional X-ray analysis using a dual-beam FIB, Ultramicroscopy 107 (8) (2007) 587–597.
[16] M. Cantoni, L. Holzer, Advances in 3D focused ion beam tomography, Mrs Bull. 39 (4) (2014) 354–360.
[17] H. Jones, K. Mingard, D. Cox, Investigation of slice thickness and shape milled by a focused ion beam for three-dimensional reconstruction of microstructures, Ultramicroscopy 139 (2014) 20–28.
[18] H.-B. Kim, G. Hobler, A. Steiger, A. Lugstein, E. Bertagnolli, Simulation-based approach for the accurate fabrication of blazed grating structures by FIB, Opt. Express 15 (15) (2007) 9444–9449.
[19] F. Lekien, J. Marsden, Tricubic interpolation in three dimensions, Internat. J. Numer. Methods Engrg. 63 (3) (2005) 455–471.
[20] Thermo Fisher Scientific Inc, Auto slice and view 4.0 [computer software], 2017, URL https://www.thermofisher.com/de/de/home/electron-microscopy/products/software-em-3d-vis/auto-slice-view-4-software.html Version: 4.1.0.1196.
[21] A. Shkurmanov, T. Krekeler, M. Ritter, Slice thickness optimization for the focused ion beam-scanning electron microscopy 3D tomography of hierarchical nanoporous gold, Nanomanuf. Metrol. (2022) 1–7.
[22] S. Shi, Y. Li, B.-N. Ngo-Dinh, J. Markmann, J. Weissmüller, Scaling behavior of stiffness and strength of hierarchical network nanomaterials, Science 371 (6533) (2021) 1026–1033.
[23] B. Peña, G.R. Owen, K. Dettelbach, C. Berlinguette, Spin-coated epoxy resin embedding technique enables facile SEM/FIB thickness determination of porous metal oxide ultra-thin films, J. Microsc. 270 (3) (2018) 302–308.
[24] P. Hanslovsky, J.A. Bogovic, S. Saalfeld, Post-acquisition image based compensation for thickness variation in microscopy section series, in: 2015 IEEE 12th International Symposium on Biomedical Imaging, ISBI, IEEE, 2015, pp. 507–511.
[25] T.D. Ambegoda, J.N. Martel, J. Adamcik, M. Cook, R.H. Hahnloser, Estimation of z-thickness and xy-anisotropy of electron microscopy images using gaussian processes, 2020, arXiv preprint arXiv:2002.00228.
[26] M.D. Uchic, M.A. Groeber, D.M. Dimiduk, J. Simmons, 3D microstructural characterization of nickel superalloys via serial-sectioning using a dual beam FIB-SEM, Scr. Mater. 55 (1) (2006) 23–28.
[27] B. Van Leer, R. Kelley, B. Winiarski, Investigation of slice thickness for FIB tomography in a plasma focused ion beam system, Microsc. Microanal. 24 (S1) (2018) 858–859.
[28] J.A. Taillon, C. Pellegrinelli, Y.-L. Huang, E.D. Wachsman, L.G. Salamanca-Riba, Improving microstructural quantification in FIB/SEM nanotomography, Ultramicroscopy 184 (2018) 24–38.
[29] Y. Li, B.-N.D. Ngô, J. Markmann, J. Weissmüller, Datasets for the microstructure of nanoscale metal network structures and for its evolution during coarsening, Data Brief 29 (2020) 105030.
[30] K. Hagita, T. Higuchi, H. Jinnai, Super-resolution for asymmetric resolution of FIB-SEM 3D imaging using AI with deep learning, Sci. Rep. 8 (1) (2018) 1–8.
[31] K. de Haan, Z.S. Ballard, Y. Rivenson, Y. Wu, A. Ozcan, Resolution enhancement in scanning electron microscopy using deep learning, Sci. Rep. 9 (1) (2019) 1–7.
[32] H. Hou, H. Andrews, Cubic splines for image interpolation and digital filtering, IEEE Trans. Acoust. Speech Signal Process. 26 (6) (1978) 508–517.
[33] J. Yamanaka, S. Kuwashima, T. Kurita, Fast and accurate image super resolution by deep CNN with skip connection and network in network, in: International Conference on Neural Information Processing, Springer, 2017, pp. 217–225.
[34] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, T.S. Huang, Generative image inpainting with contextual attention, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 5505–5514.

Supplementary Material

After repositioning slices, we arrived at imaging data with gaps between the available imaging data layers. To solve this impainting problem, we used convolutional neural networks (CNN) to interpolate between the available layers. In the following details of data preparation, network architecture and training are summarized.

1 DATA PREPARATION

We trained and validated the machine learning model on the segmented (binarized) synthetic imaging data (resolution of 2.5 nm in all directions). To this end, we extracted all available zx- and zy-planes from the imaging data and split these planes into training data (ca. 80%) and validation data (ca. 20%). Masks were generated for all missing regions in the dataset. For training, we used data from our synthetic samples, for which complete images without missing regions could be prepared as ground truth (because, for these samples, the exact microstructure was known). We used an online data augmentation technique to increase the training data. That is, we applied random rotations and horizontal/vertical flips to the training data. We did not apply any transformations to the validation dataset.

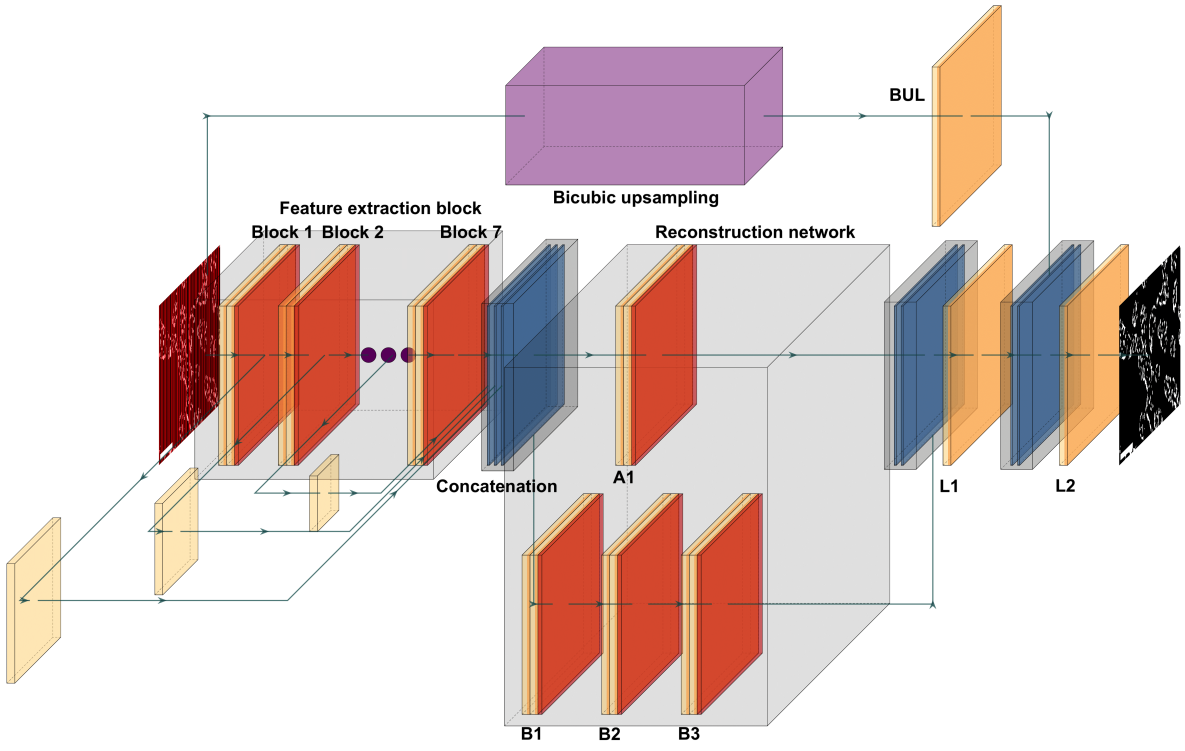


Figure S1. Machine learning architecture used in this paper to interpolate voxelated data points (i.e., to solve the inpainting problem of interest)

2 ARCHITECTURE

Our machine-learning model architecture was implemented using PyTorch and was inspired by Yamanaka et al. (2017). Their ML model was developed to solve problems of image super-resolution; by contrast, we used our modified ML model to solve an image inpainting problem. In this ML model, encoded features from the input image are extracted by seven convolutional blocks in the feature extraction network. These encoded features are then concatenated and passed through a reconstruction network, which reconstructs an output array of the same size as the input image. We use a different number of filters in the feature extraction and the reconstruction network blocks than Yamanaka et al. (2017). The number of filters in both network blocks is summarized in Table S1. In our modified ML model, we concatenated the output of the reconstruction network and a separate image dataset prepared from the input of the overall architecture using bicubic interpolation (Figure S1). This was in contrast to adding them together as mentioned by Yamanaka et al. (2017). Our approach provided good results even after a short training period of approximately 14 hours and 30 minutes. However, our machine learning-based method is computationally expensive compared to the classical method, cubic splines interpolation. Our method with a batch size of 128 takes approximately 44 seconds compared to 7 seconds for cubic splines interpolation for 128 images of size 512×512 pixels. It should be noted that the time taken to calculate results using our method depends on several parameters, including batch size and GPU memory capacity.

Table S1. Number of filters used in each block of our machine learning architecture

Feature extraction								Reconstruction				Concatenate		Bicubic layer
								A1	B1	B2	B3	L1	L2	BUL
96	81	70	60	50	41	32	16	96	64	48	32	1		1

3 TRAINING

We trained our machine learning architecture on Tesla K80 GPUs. We provided small image patches (64×64 pixels) with a stride of 32 pixels as input to train the ML model using a sliding window technique. We used a mean squared error (MSE) loss function and used the Adam optimization algorithm with an initial learning rate of 0.001. We reduced the learning rate by a factor of 10 if the loss did not decrease for ten consecutive training epochs. Additional parameters used for the training process are specified in Table S2. Figure S2 illustrates a typical training data and validation data MSE loss curve.

4 IMAGE POST-PROCESSING

We noticed clusters of missing pixels, *holes*, in some of the ligaments after performing cubic-splines interpolation (Figure S3). These *holes* were filled using image processing techniques to improve the quality of these interpolated structures. Specifically, we used the Scikit-image package [1] to detect *holes* which were entirely covered by the gold structures. We only considered *holes* having an area smaller than 20 pixels to ensure that any secondary hierarchy of the ligaments was not classified as artefacts. *Holes* detected in this manner were then filled with a gold intensity to enhance the overall structure of the ligaments.

[1] https://scikit-image.org/docs/stable/api/skimage.morphology.html

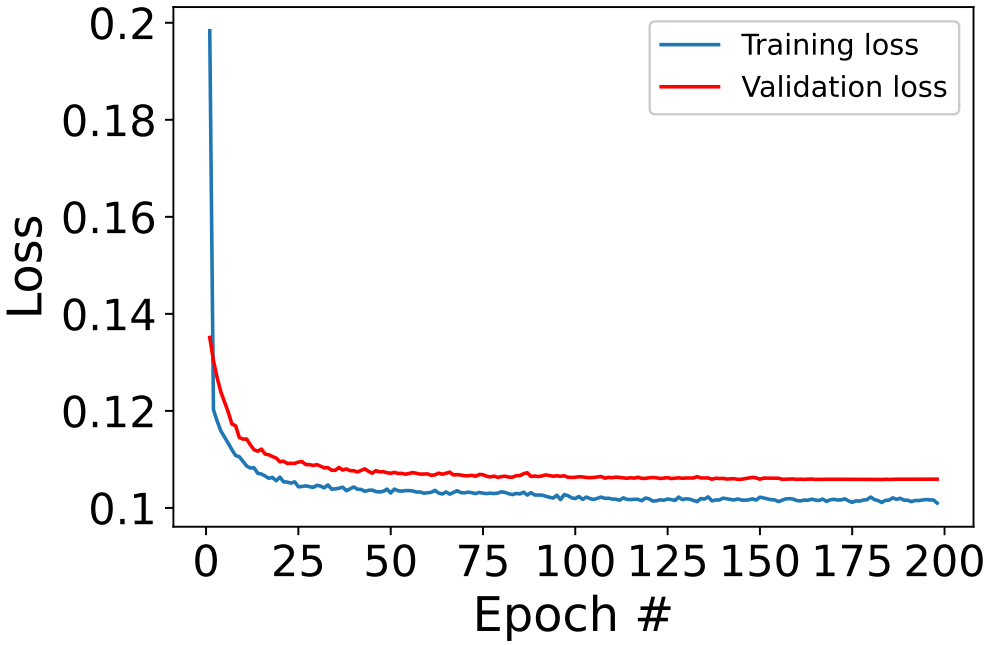


Figure S2. Training data and validation data MSE loss function over 200 epochs

Table S2. Parameters used for training machine learning model (interpolation)

Parameter	Value
Patch size	64
Stride	0.5
Batch size	64
Epochs	500 with early stopping with patience=25
Loss	MSE loss
Optimizer	Adam
Learning rate	0.001 with reducing it by factor of 0.10 with patience of 10

(A) (B)

Figure S3. Interpolated ligament of hierarchical nanoporous gold using cubic-splines interpolation **(A)** without post-processing and **(B)** after post-processing

5 ADDITIONAL RESULTS USING ANISOTROPY-BASED ERROR MEASURES FOR SYNTHETIC DATA

Table S3. Performance comparison of different interpolation techniques on 250 different synthetic datasets.

Dataset name	$e_{L_2}^{TPCF} \downarrow$	$e_{L_2}^{LPF} \downarrow$	$e_{L_2}^{D} \downarrow$
Original - [Target values]	0.1038 ± 0.0283	0.0188 ± 0.0131	0.0138 ± 0.0096
RNN-*s*	0.1149 ± 0.0247	0.0479 ± 0.0222	0.0306 ± 0.0185
RCS-*s*	0.1091 ± 0.0278	0.0482 ± 0.0188	0.0299 ± 0.0173
RML-*s*	0.1086 ± 0.0274	0.0341 ± 0.0131	0.0211 ± 0.0114

Table S4. Performance comparison of different interpolation techniques on 250 different synthetic datasets.

Dataset name	$e_{L_2}^{TPCF} \downarrow$	$e_{L_2}^{LPF} \downarrow$	$e_{L_2}^{D} \downarrow$
Original - [Target values]	0.1038 ± 0.0283	0.0188 ± 0.0131	0.0138 ± 0.0096
DNN-*s*	0.2104 ± 0.0467	0.0633 ± 0.0276	0.0386 ± 0.0191
SRCS-*s*	0.1153 ± 0.0267	0.0716 ± 0.0222	0.0455 ± 0.0198
RML-*s*	0.1086 ± 0.0274	0.0341 ± 0.0131	0.0211 ± 0.0114

REFERENCES

Yamanaka, J., Kuwashima, S., and Kurita, T. (2017). Fast and accurate image super resolution by deep cnn with skip connection and network in network. In *International Conference on Neural Information Processing* (Springer), 217–225

Chapter 5

Improving 3D reconstruction further using multimodal machine learning

5.1 Background of the study

Article III focuses on improving machine learning-based segmentation methods. While Article I extensively discussed how the *shine-through effect* could impair accurate semantic segmentation and strategies to mitigate this issue using machine learning, Article III sheds light on the potential of leveraging the *shine-through effect* as a solution for accurate segmentation. Contrary to considering the *shine-through effect* solely as a challenge, we demonstrate its usage as a valuable source of additional information for accurate segmentation.

As explained previously, the *shine-through effect* results in the visibility of structures in the current plane from posterior regions. Moreover, the visibility of these additional structures increases with increasing accelerating voltage (see Figure 1 of Article III). We try to use this property as an opportunity to incorporate additional information that aids in accurate segmentation. Essentially, the *shine-through effect* offers a preview of structures from subsequent slices, providing *future* insight into the microstructure. Machine learning models can obtain meaningful insights by leveraging this extra information across multiple images captured at different accelerating voltages.

Thus, we proposed developing an over-deterministic system using these multi-voltage images. These images were obtained following a similar methodology described in previous articles, albeit with a modification: instead of acquiring a single image using SEM followed by milling using FIB, three images of the same plane were captured at different accelerating voltages before material removal. Similarly, synthetic datasets for

training machine learning models were generated by simulating the same slice multiple times at various accelerating voltages.

We treated each image acquired at a single accelerating voltage as a distinct modality to harness the additional information provided by multi-voltage images. To this end, we implemented a multimodal system by combining all images of the same region obtained at different accelerating voltages. In machine learning, a specialized technique known as multimodal machine learning exists to handle such multimodal datasets. This approach integrates information from different sensors or sources to obtain robust and accurate predictions. The architecture and techniques used in multimodal machine learning are explained comprehensively in Article III.

Multimodal machine learning models were trained in a manner similar to single-modality machine learning models. However, in two types of multimodal machine learning (Type 1 and Type 2 in Article III), input data were not fed individually but rather combined at the input level or features learned from different modalities at the classifier level.

In this study, we compared machine learning models trained on images obtained at single accelerating voltages (ML-singleV) with our proposed multimodal machine learning approach. Furthermore, we conducted comparative analyses among different multimodal machine learning model architectures. Additionally, we tried to answer whether our multimodal machine learning model could only learn from images captured at the lowest accelerating voltage, making multimodal completely unnecessary. This concern is relevant since images obtained at the lowest accelerating voltage exhibit minimal *shine-through effect*, thereby providing the most accurate signal from FIB-SEM. To answer this, we trained and optimized three different machine learning models on synthetic data simulated at 1kV, 2kV, and 4kV. Subsequently, to make the process interpretable, we implemented a voting-based technique to derive the optimal segmentation result. This study contributes to explaining the contributions of each modality in multi-voltage images.

5.2 Discussion

Our study compared three multimodal machine learning architectures to determine the most effective multimodal network for semantic segmentation. We trained all three architectures mentioned in Section 2.2 of Article III using synthetic datasets (s-1kV, s-2kV, and s-4kV). Subsequently, after fine-tuning these models, we rigorously

evaluated their performance using the metrics outlined in Section 2.5.1 of Article III, leveraging the availability of ground truth data specifically for the synthetic dataset.

It is essential to highlight that the synthetic dataset used for calculating metrics remained entirely distinct from that used during the training process, ensuring unbiased evaluation. Our findings, summarized in Table 2 from Article III, revealed that the late fusion architecture emerged as the top-performing option among the three architectures compared.

Interestingly, the intermediate fusion architecture exhibited notably poorer metrics than the late fusion architecture. This decline in performance can be attributed to the large size of the ML model, posing optimization challenges due to limited training data availability. Metrics such as MP, MGP, and MDS (see Section 2.6.1) notably favored the late fusion architecture over the more straightforward early fusion method.

Furthermore, we analyzed the behavior of the late fusion network regarding the *shine-through effect* by substituting the ensemble layer with an interpretable voting mechanism. This experiment revealed insightful weight distributions of 0.5, 0.3, and 0.2 for CNN models individually trained on the s-1kV, s-2kV, and s-4kV datasets, respectively. These findings suggest that the model effectively gains valuable insights from datasets with varying degrees of *shine-through effect*.

Based on this detailed comparative study, we determined the late fusion architecture as the preferred choice for all subsequent experiments of this study, ensuring robustness and reliability in our segmentation approach.

Following our initial analysis, we conducted a comprehensive comparative study, wherein we evaluated the performance of our multimodal machine learning method (ML-multiV) against the cluster-based k-means clustering algorithm and individual ML models trained on single kV datasets, denoted as ML-singleV. In particular, ML-singleV (s-1kV), ML (s-2kV), and ML (s-4kV) represent scenarios where the ML model was trained and tested using individual single kV datasets, namely s-1kV, s-2kV, and s-4kV, respectively. The outcomes of this comparison are presented in Table 3 from Article III, which includes key metrics such as MP, MGP, and MDS for the reconstructed synthetic data.

An important finding of our study is the superior performance of our multimodal ML method, ML-multiV, which surpassed all alternative techniques mentioned in this study. All ML models, including ML-singleV variants - even at 4kv -, outperformed the cluster-based k-means clustering method when applied to their respective individual datasets. Notably, the most promising results were observed for the multimodal ML approach, particularly when leveraging information from all three datasets for

segmentation. This highlights the significant advantage of incorporating additional information from images captured at various accelerating voltages.

To provide a deeper insight into the performance of each segmentation method for synthetic data, Supplementary Table 7 from Article III presents a detailed exploration of anisotropy-based metrics, further validating the trends observed in Table 3 from Article III. Additionally, Figure 7 from Article III offers a visual representation of the segmentation results, providing further evidence of the effectiveness of our multimodal ML technique.

Our study extends beyond synthetic datasets by validating the efficacy of our multimodal machine learning (ML) approach on real data obtained from actual HNPG FIB tomography data (r-1kV, r-2kV, and r-4kV). Using anisotropy-based metrics outlined in Section 3.5.2 of Article III, we evaluated the segmentation performance of our ML-multiV method on the real HNPG data. The results, presented in Table 4 from Article III, reaffirm the superiority of our approach over alternative segmentation methods when applied to real FIB tomography data.

A visual illustration of these findings is provided in Figure 8 from Article III, showcasing a single slice from a segmented real HNPG dataset using different segmentation methods. This illustration demonstrates the enhanced capability of all ML-based techniques in effectively mitigating the challenging *shine-through effect* compared to classical segmentation methods.

Notably, our ML-multiV approach exhibits an impressive average improvement of more than 50 % in performance based on anisotropy measures compared to k-means clustering. Furthermore, the improvement over k-means clustering (r-1kV) in anisotropy measures exceeds 30%, underscoring the substantial benefits of our multimodal ML approach in addressing the complexities inherent in real FIB tomography data.

5.3 Contribution to the field

- We proposed a novel method to consider artifacts due to the *shine-through effect* as additional information instead of a problem like in conventional methods. Implementation of the interpretable voting-based method and results of weights show that additional information due to the *shine-through effect* helps improve semantic segmentation accuracy.

- An extensive comparison of different multimodal architectures for segmenting FIB tomography data can help researchers select the architecture according to the availability of the datasets and computation power.

- Our multimodal machine learning method shows superior performance over single modality machine learning models, highlighting the potential of the additional information.

5.4 Problems yet to be addressed

1. It is very difficult to acquire multi-voltage images due to machine limitations and expertise needed for the acquisition. Additionally, the acquisition and post-acquisition processing of the dataset is very time-consuming. Therefore, it would be a tradeoff between better accuracy and additional effort required to acquire the dataset.

2. Data distribution difference should be reduced using domain adaptation techniques, and machine learning-based methods have the potential to replace slow and compute-intensive Monte Carlo-based simulation methods.

5.5 Article III - Enhancing 3D reconstruction accuracy of multi-voltage FIB tomography images using multimodal machine learning

T. Sardhara, A. Shkurmanov, Y. Li, L. Riedel, S. Shi, C.J. Cyron, R.C. Aydin, M. Ritter: **Enhancing 3D reconstruction accuracy of multi-voltage FIB tomography images using multimodal machine learning.** Nanomanufacturing and Metrology 7, 4 (2024) (DOI: 10.1007/s41871-024-00223-y) *

Nanomanufacturing and Metrology (2024) 7:4
https://doi.org/10.1007/s41871-024-00223-y

ORIGINAL ARTICLE

Enhancing 3D Reconstruction Accuracy of FIB Tomography Data Using Multi-voltage Images and Multimodal Machine Learning

Trushal Sardhara[1] · Alexander Shkurmanov[2] · Yong Li[3] · Lukas Riedel[4] · Shan Shi[4,5] · Christian J. Cyron[1,6] · Roland C. Aydin[1,6] · Martin Ritter[2]

Received: 6 November 2023 / Revised: 10 January 2024 / Accepted: 10 January 2024

Abstract

FIB-SEM tomography is a powerful technique that integrates a focused ion beam (FIB) and a scanning electron microscope (SEM) to capture high-resolution imaging data of nanostructures. This approach involves collecting in-plane SEM images and using FIB to remove material layers for imaging subsequent planes, thereby producing image stacks. However, these image stacks in FIB-SEM tomography are subject to the *shine-through effect*, which makes structures visible from the posterior regions of the current plane. This artifact introduces an ambiguity between image intensity and structures in the current plane, making conventional segmentation methods such as thresholding or the k-means algorithm insufficient. In this study, we propose a multimodal machine learning approach that combines intensity information obtained at different electron beam accelerating voltages to improve the three-dimensional (3D) reconstruction of nanostructures. By treating the increased *shine-through effect* at higher accelerating voltages as a form of *additional* information, the proposed method significantly improves segmentation accuracy and leads to more precise 3D reconstructions for real FIB tomography data.

Highlights

1. FIB-SEM tomography can be used to acquire high-resolution image stacks of nanostructures.
2. Machine learning reduces artifacts and ambiguities introduced in FIB-SEM tomography because of the ***shine-through effect***.
3. Our approach treats the ***shine-through effect*** as a valuable ***additional*** information for precise reconstruction.
4. Using multi-voltage images, our multimodal ML approach improves 3D reconstruction accuracy.

Keywords Multimodal machine learning · Multi-voltage images · FIB-SEM · Overdeterministic systems · 3D reconstruction · FIB tomography

✉ Martin Ritter
ritter@tuhh.de

[1] Institute for Continuum and Material Mechanics, Hamburg University of Technology, Hamburg, Germany

[2] Electron Microscopy Unit, Hamburg University of Technology, Hamburg, Germany

[3] Institute of Materials Physics and Technology, Hamburg University of Technology, Hamburg, Germany

[4] Institute of Materials Mechanics, Helmholtz-Zentrum Hereon, Geesthacht, Germany

[5] Research Group of Integrated Metallic Nanomaterials Systems, Hamburg University of Technology, Hamburg, Germany

[6] Institute of Material Systems Modeling, Helmholtz-Zentrum Hereon, Geesthacht, Germany

Published online: 27 February 2024

1 Introduction

Structural analysis is essential in materials science, enabling the precise characterization and modeling of complex materials. This knowledge facilitates the implementation of these materials in the mass production industry and everyday life [1]. However, high-resolution imaging techniques are typically required to investigate materials with nanoscale features, which limit the range of suitable methods. For instance, when examining hierarchical nanoporous materials such as those in our case with ligament sizes larger than 100 nm and nanopores smaller than 20 nm, conventional techniques, such as X-ray nanotomography, may lack the necessary resolution [2].

Electron microscopy, particularly focused ion beam (FIB) tomography in conjunction with scanning electron microscopy (SEM), offers the essential resolution for analyzing conductive hierarchical nanoporous materials. FIB removes nanometer-thin layers of material after each SEM imaging step, building a stack of images for subsequent three-dimensional (3D) reconstruction. Although FIB-SEM tomography has been applied to the study of nanoporous gold and hierarchical nanoporous gold (HNPG) in various studies [3–5], several challenges persist.

First, the variation in slice thickness in FIB tomography images can introduce inaccuracies in 3D reconstructions. A slice repositioning method was proposed to address this issue, and a machine learning (ML) method was adopted to solve this problem from the perspective of an image inpainting problem [6].

The *shine-through effect* is another challenge, where structures from posterior regions become visible through pores in the currently milled plane [7]. This effect complicates the mapping between actual structure and image intensity [8, 9]. While classical machine learning methods like random forests or k-means clustering are effective in cases where the *shine-through effect* is minimal [8, 10], the utility of more advanced machine learning techniques for suppressing this effect in cases where it is non-negligible was demonstrated [11, 12]. They used simulated data to train machine learning models and successfully applied these optimized models to real nanostructures. This technique, known as transfer learning, has exhibited notable performance in materials science applications [13].

The extent of the *shine-through effect* is directly related to the beam energy (Fig. 1), allowing its use as an auxiliary method to establish an overdetermined system and collect a comprehensive SEM dataset. In other words, by analyzing images acquired at different voltages, it is possible to separate the *shine-through effect* signal from the surface area signal, ultimately enabling the reconstruction of hierarchical nanoporous structures with significantly improved accuracy.

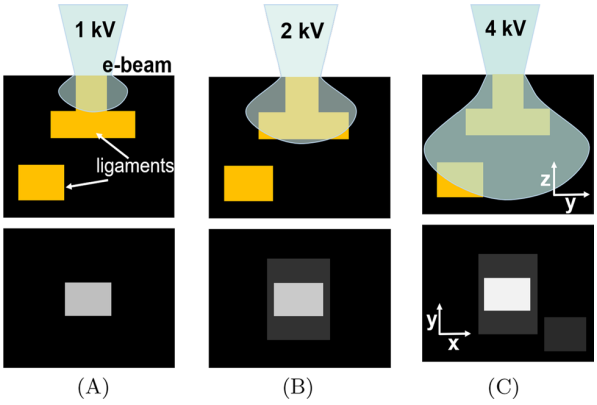


Fig. 1 Schematic overview of multi-voltage image acquisition: penetration of electron beam (e-beam) into microstructure from the top (top row) for acceleration voltages of 1 kV (**A**), 2 kV (**B**), and 4 kV (**C**) and resulting planar SEM image (bottom row)

We use advanced ML methods to solve the overdetermined system obtained from images obtained at different voltages.

In general, multimodal ML leverages different data types, such as speech and images, to improve task-specific performance [14]. In medical imaging, multimodal data fusion, such as combining positron emission tomography with computed tomography and/or magnetic resonance imaging, enhances lesion quantification [15–18]. Shared image features from unregistered views enhance classification [19]. A comprehensive summary of multimodal ML developments in medical imaging can be found in [20]. In electron microscopy, combining signals from high-angle annular dark-field and energy-dispersive X-ray spectroscopy facilitates the 3D reconstruction of nanostructures [21], and combining transmission X-ray microscopy with FIB-SEM improves the image quality for shale analysis [22].

In this study, we present a novel ML-based method for a multi-voltage (multiV) FIB tomography dataset of HNPG. To develop this method, we created synthetic multiV images using a Monte Carlo-based method [12]. These simulated multiV data are used to train ML models. We compared different segmentation methods and demonstrated significant segmentation improvements with multiple modalities. In particular, our multimodal ML method outperforms single-modality ML methods in mitigating the *shine-through effect*.

2 Materials and Methods

2.1 Acquisition of Imaging Data

In this study, we used imaging data obtained from both real HNPG samples and computer-generated synthetic imaging

data, closely resembling actual HNPG imaging data, to construct datasets for ML.

2.1.1 Real Samples

An HNPG sample featuring a uniform random network structure with two well-defined ligament sizes of 15 and 110 nm was prepared using the dealloying-coarsening-dealloying method [23]. To improve SEM imaging and distinguish solid and pore phases, epoxy resin infiltration was employed because it provides a dark, homogeneous background during SEM imaging [24]. MultiV FIB tomography of the HNPG sample was performed using a Dual Beam FEI Helios NanoLab G3 system, integrating ASV4 Thermo Fisher Scientific Inc (2017) software [25] for automated tomography control. This software acquires electron- and ion-beam images to monitor milling progress and compensate for image drift resulting from various sources, including charge effects, mechanical stage drift, and thermal-induced drift [26].

To perform drift compensation during FIB tomography, two fiducial markers were prepared and positioned on the side of the region of interest (ROI). Each fiducial marker comprised two perpendicularly intersecting trenches milled on a platinum-deposited area, separate from the ROI. In addition, a ruler system was implemented on top of the ROI to ensure precise slice thickness determination, following a methodology similar to that presented in [27] and optimized for HNPG tomography [3]. This approach involved sputtering a 1-µm carbon layer atop the ROI to smoothen the surface and render the ruler visible (see Fig. 2).

During multiV tomography, each slice was imaged three times with accelerating voltages of 1, 2, and 4 kV at a constant current of 50 pA. This process generated three real FIB tomography image datasets referred to as r-1 kV, r-2 kV, and r-4 kV (where the letter "r" refers to the fact that the data are based on real rather than synthetic, computer-generated microstructures). The SEM image parameters included a resolution of 3072 × 2048 and a horizontal field of view spanning 10 µm, resulting in a pixel size of 3.26 nm. Low-noise imaging was achieved with a dwell time of 30 µs. A through-the-lens detector (TLD) was used to detect backscattered electrons (BSEs).

In total, 316 slices were milled using an ion beam with an aperture of 80 pA and an accelerating voltage of 30 kV. This dataset yielded a mean slice thickness of $< d > = 10.16$ nm,

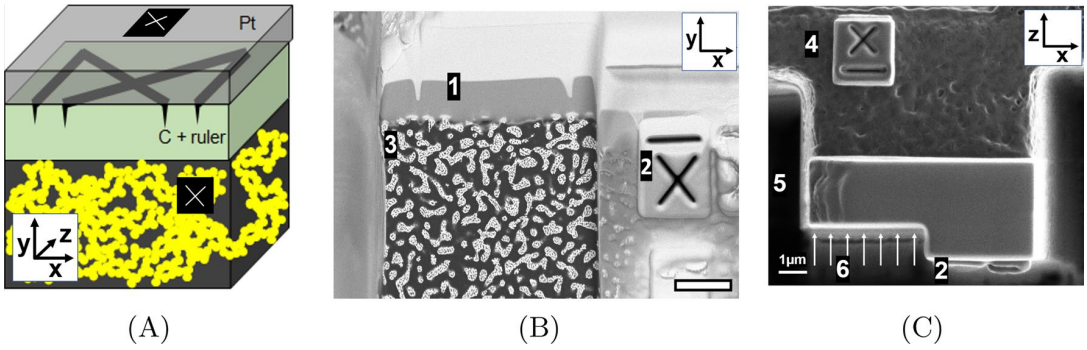


Fig. 2 Experimental dataset. **A** Schematic overview. Dark areas with white crosses represent fiducial structures for drift compensation. **B** Imaged block face-electron-beam view (BSE, TLD, 2 kV, 50 pA, 52° tilt) and **C** milling view (ETD, SE, 30 kV, 80 pA, 0° tilt) with (1) ruler structure, (2) e-beam fiducial structure, (3) HNPG infiltrated with epoxy, (4) ion-beam fiducial structure, (5) 30-µm deep trenches, and (6) direction of ion mill (scale bar: 1 µm)

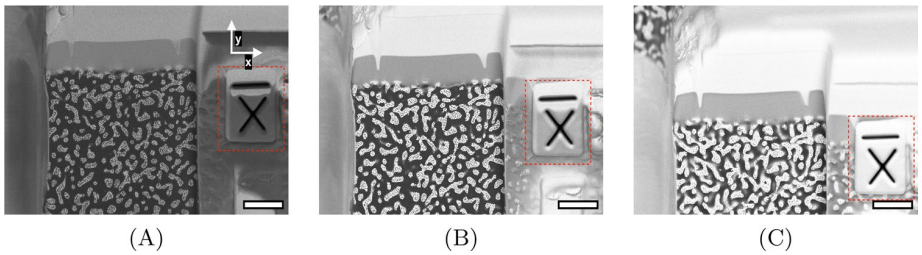


Fig. 3 Single raw FIB-SEM slice images of real HNPG at **A** 1 kV, **B** 2 kV, and **C** 4 kV, each featuring a fiducial marker in the *xy*-plane. Misalignment is evident from the marked rectangles (red), showcasing variations in *x*- and *y*-directions between images acquired at different accelerating voltages. Scale bar: 1000 nm

with a standard deviation of $\sigma = 3.12$ nm, aligning closely with the target thickness of 10 nm. Notably, 9% of the slices, characterized by thickness outliers (z-scores of 1.5), were excluded from the thickness calculation analysis.

Owing to instrument constraints, misalignments frequently occur in images captured at different accelerating voltages within and across stacks (see Fig. 3). Consequently, image registration is essential to ensure that the same region is consistently exposed across all voxel groups for images obtained at different accelerating voltages. This registration process comprises two key steps:

Intra-stack image registration

The first step involves aligning the images within each image stack acquired at a specific accelerating voltage (1, 2, and 4 kV). The initial image in each stack serves as a reference. Subsequent images are translated in the (in-plane) *x*- and *y*-directions to maximize the correlation between consecutive slices. We used the Fiji plugin [28] 'Register virtual stack slices' to perform intra-stack registration.

Inter-stack image registration

After intra-stack registration, images across stacks acquired with different voltages (1, 2, and 4 kV) were aligned considering the image stack acquired at 2-kV accelerating voltage as the reference. Calculating correlations between corresponding slices in 1- and 2-kV stacks enabled translational shifts for optimal alignment. A supplementary fiducial marker, milled on the right side of the sample (Fig. 2), enhanced the inter-stack correlation as no *shine-through effect* is noticed for the region across different image stacks. Aligning the 4- and 2-kV stacks followed the same procedure. Figure 4 shows an identical slice from each dataset after the registration procedure.

2.1.2 Synthetic Samples

Supervised ML on electron microscopy data, particularly for FIB tomography of HNPG, can be challenging because of the high cost and time involved in acquiring substantial training datasets. Because ground truth images are unavailable for the FIB tomography data of HNPG, we generated synthetic FIB tomography images, as described in [12]. In the first step, virtual initial structures resembling real HNPG structures were generated using the method described in [23]. The leveled-wave model [29] provides a basis for each virtual HNPG hierarchy level. This model starts with a concentration field formed by the superimposition of waves with

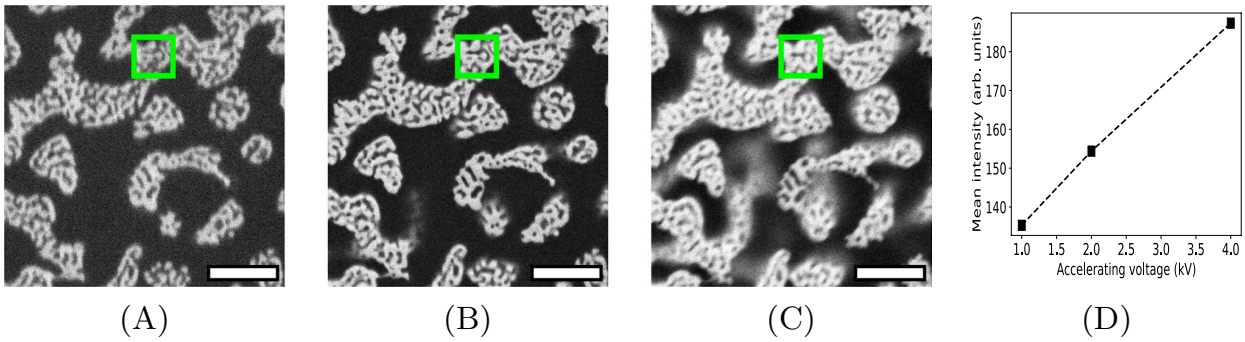


Fig. 4 Aligned single identical slice of real FIB-SEM image of HNPG imaged at accelerating voltages of **A** 1 kV, **B** 2 kV, and **C** 4 kV; **D** mean intensity plot of the highlighted area (green square) for 1-, 2-, and 4-kV images. Scale bar: 300 nm

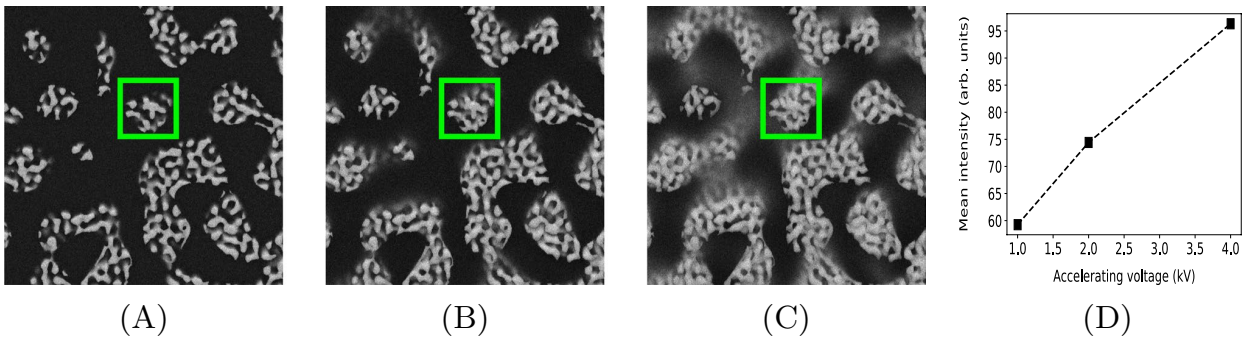


Fig. 5 Same slice of simulated FIB tomography data with accelerating voltages of **A** 1 kV, **B** 2 kV, and **C** 4 kV; **D** mean intensity plot of the highlighted area (green square) for 1-, 2-, and 4-kV images. Scale bar: 300 nm

the same wavelength but with random wave vector directions. The wave vectors have the same magnitude, which dominates the ligament size of the virtual nanoporous network. Subsequently, a level cut was selected for binarization into a pore or solid phase, and virtual nanoporous microstructures were generated after removing the pore phase. Various level cuts can be applied to achieve the desired solid fraction. In the next step, the process of FIB-SEM tomography was simulated on these binary structures by Monte Carlo simulations using the MCXray plugin of Dragonfly software [30]. We generated three datasets using the same virtual initial structure but at different accelerating voltages (1, 2, and 4 kV), similar to the real HNPG data. As shown in Fig. 5, these datasets, named s-1 kV, s-2 kV, and s-4 kV (where "s" refers to "synthetic"), resemble the real HNPG data acquired at different accelerating voltages.

2.2 Machine Learning Architectures for Semantic Segmentation

We used multimodal ML to segment the FIB tomography data semantically. Inspired by [20], we developed three architectures for 3D nanostructure reconstruction, adapted from [12], employing different data fusion techniques:

2.2.1 Early Fusion

In the early fusion architecture (Fig. 6(A)), we fused multimodal images at the input level by channel-wise concatenating single two-dimensional (2D) slices from the s-1 kV, s-2 kV, and s-4 kV datasets. These slices form the input tensor, passing through custom 2D and 3D U-Net models to produce the final segmentation output.

2.2.2 Intermediate Fusion

The intermediate fusion architecture (Fig. 6(B)) contains separate branches for each imaging modality, focusing on images at different accelerating voltages. These images pass through modality-specific subnetworks to extract low-level features, which are then concatenated and classified using a fully connected neural network to yield the final segmentation. This architecture posed challenges because of its size during the optimization process.

2.2.3 Late Fusion

The late fusion architecture (Fig. 6(C)) resembles intermediate fusion, with each branch dedicated to a different modality, initially trained and optimized. These trained ML models estimate the probabilities of each pixel belonging to either the material or pore phase. In the second step, the output probabilities from these optimized branches are collected and concatenated. These concatenated probabilities then train a final neural network layer (ensemble), which, once optimized, produces the ultimate segmentation output.

2.3 ML Model Training Process

We trained all ML architectures on Tesla K80 GPUs. Images were initially cropped into smaller patches (64×64) with a 32-pixel stride using a sliding window technique. Our networks are trained using a structured approach where data are presented as individual 2D slices, a 3D volumetric stack, or a 2D slice coupled with its neighboring slices for a 2D convolutional neural network (CNN), 2D CNN with adjacent slices, and 3D CNN. We used Dice loss in conjunction with the Adam optimizer, starting with an initial learning rate of 0.0001. If the learning rate did not decrease for ten consecutive training epochs, the learning rate was reduced by a factor of 10. Table 1 shows details of the training parameters, and supplementary (Appendix) Fig. 10 illustrates a typical training and validation data Dice loss curve.

2.4 Data Augmentation

ML requires a large set of training data. When it is difficult to collect sufficient training data, data augmentation can be a powerful method for increasing the data size for ML [31]. Differences in data distributions between synthetic and real datasets and variations stemming from different microscopes emphasize the need for extensive data augmentation in electron microscopy. We applied online data upsampling, where the data size was increased during training by applying different image processing operations such as random flips, image rotations, random croppings, and changes in brightness.

2.5 Evaluation Criteria for 3D Image Reconstruction

2.5.1 Metrics Based on Ground Truth Values

In cases where ground truth data are available as a binary virtual initial structure, e.g., for synthetic datasets, we calculated the following three absolute error metrics:

The first metric, the fraction of misplaced pixels (MP), measures the percentage of pixels whose predicted and ground truth values do not match. This metric was calculated using the following formula:

$$\mathrm{MP} = \left(1 - \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}}\right) \times 100 \quad (1)$$

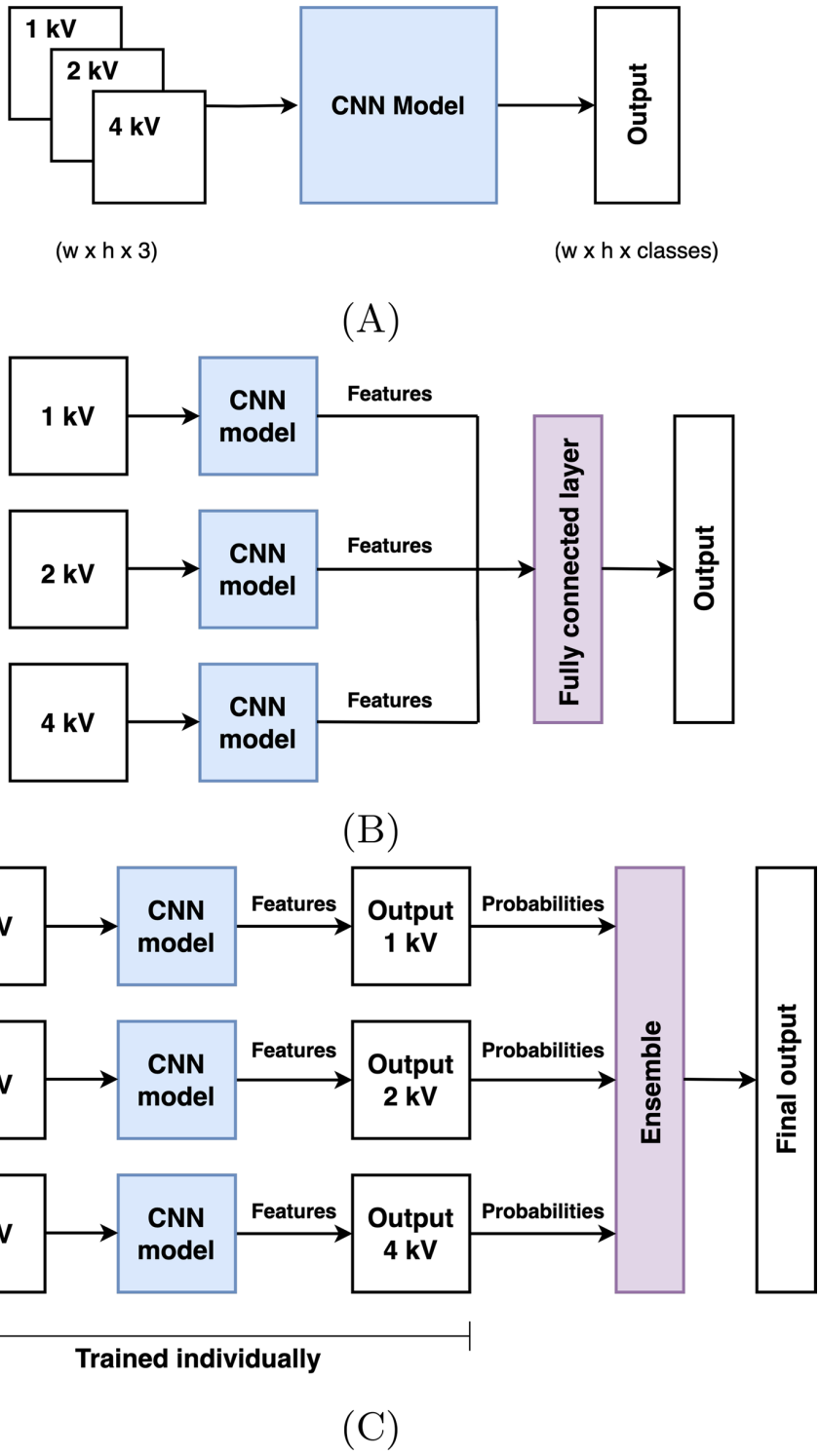


Fig. 6 Different multimodal ML architectures: **A** early fusion, **B** intermediate fusion, and **C** late fusion. Note: Each CNN model block (blue color) contains deep CNNs for semantic segmentation (see Appendix 1)

Table 1 Summary of parameters used for training ML models

Parameter	Value
Patch size	64
Stride	0.5
Batch size	64
Epochs	100 with early stopping with patience=25
Loss	Dice loss
Optimizer	Adam
Learning rate	0.0001, adapted with a patience of 10 (reduction factor 10)

where TP denotes the number of true positives, TN denotes the number of true negatives, FP denotes the number of false positives, and FN denotes the number of false negatives.

The second metric, the percentage of misplaced gold pixels (MGP), calculates falsely predicted gold pixels using ground truth values. One should be careful while using this metric alone, as FP is not considered in this metric. MGP is calculated as follows:

$$\mathrm{MGP} = \left(1 - \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}\right) \times 100 \tag{2}$$

The third metric, the Dice score (DS), measures the overlapping regions between predicted and ground truth binarized images. It provides a valuable measure when data imbalances exist. DS takes values from 0 to 1, 1 being the best performance indicator. DS is calculated as follows:

$$\mathrm{DS} = \frac{2\,\mathrm{TP}}{2\,\mathrm{TP} + \mathrm{FN} + \mathrm{FP}}. \tag{3}$$

We computed the DS for both the material and pore phases and then averaged them to obtain the mean Dice score (MDS).

2.5.2 Metrics Based on Anisotropy in the Absence of Ground Truth Values

When ground truth data are unavailable, e.g., for real HNPG FIB tomography, anisotropy-based metrics are used to assess ML model performance under the assumption of isotropy of the actual nanostructure [12]. These metrics account for the *shine-through effect* observed in the z-direction. Three metrics are calculated in the x-, y-, and z-directions and compared to evaluate anisotropy based on the three functions. The first metric is based on the two-point correlation function (TPCF), which is the probability of having two points in the same material phase at a given distance from each other in x-, y-, and z-directions. Let this function calculated in the three spatial directions be denoted as f^x, f^y, and f^z. When these functions are uniformly discretized with n data points, we obtain the function values f_i^x, f_i^y, and f_i^z, where i ranges from 1 to n. TPCFs statistically characterize the microstructure in different spatial directions. For isotropic microstructures, identical TPCFs are expected in all spatial directions. Conversely, the differences between the TPCFs in different directions are a measure of the anisotropy of the microstructure. To quantify these differences, we calculated the L_2 differences between the TPCF in x- and z-directions and y- and z-directions. These L_2 differences were averaged to obtain the final anisotropy metric as follows:

$$e_{L_2}^{\mathrm{TPCF}} = \frac{1}{2}\left(\frac{2\times\sqrt{\sum_{i=1}^{n}(f_i^x - f_i^z)^2}}{\sqrt{\sum_{i=1}^{n}(f_i^x)^2} + \sqrt{\sum_{i=1}^{n}(f_i^z)^2}} + \frac{2\times\sqrt{\sum_{i=1}^{n}(f_i^y - f_i^z)^2}}{\sqrt{\sum_{i=1}^{n}(f_i^y)^2} + \sqrt{\sum_{i=1}^{n}(f_i^z)^2}}\right) \tag{4}$$

A similar anisotropy metric can be defined using the lineal path function (LPF) instead of the TPCF. The LPF measures the probability that two points at a certain distance in a certain direction can be connected by a line fully located in the same phase. The resulting anisotropy metric $e_{L_2}^{\mathrm{LPF}}$ is calculated analogously to $e_{L_2}^{\mathrm{TPCF}}$. The third metric, $e_{L_2}^{D}$, is a similar metric and calculates the average difference of ligament diameters in x- and z-directions and y- and z-directions:

$$e_{L_2}^{D} = \frac{1}{2}\left(\sqrt{\frac{(D_{xz} - D_{xy})^2}{D_{xy}^2}} + \sqrt{\frac{(D_{yz} - D_{xy})^2}{D_{xy}^2}}\right) \tag{5}$$

Here, D_{ij} denotes the calculated average diameter of ligaments in the ij plane with $i, j \in \{x, y, z\}$. When calculated for a completely isotropic structure, zero errors indicate the best reconstruction performance. However, higher error values for these functions suggest anisotropy in the reconstructed structure. For an underlying actual isotropic microstructure, this indicates a segmentation error.

3 Results and Discussion

3.1 Comparison of Different Multimodal ML Architectures

First, we performed a comparative analysis of the three multimodal ML architectures to identify the most effective segmentation model. We trained all three architectures,

as described in Sect. 2.2, using synthetic datasets (s-1 kV, s-2 kV, and s-4 kV). After fine-tuning these models, we evaluated their performance using the metrics outlined in Sect. 2.5.1, leveraging the availability of ground truth data for the synthetic dataset. It is essential to emphasize that the synthetic dataset used for metrics calculation remained entirely isolated from that used during the training process. Our results, presented in Table 2, indicate that the late fusion architecture excelled among the three options.

Notably, the intermediate fusion architecture exhibited significantly poorer metrics than the late fusion architecture. This decline in performance can be attributed to the large size of the ML model, which posed challenges for optimization given the limited training data. MP, MGP, and MDS were significantly less favorable for the more straightforward multimodal architecture, early fusion. In addition, we examined the behavior of the late fusion network on the *shine-through effect* by substituting the ensemble layer with a voting mechanism. This adaptation revealed weights of 0.5, 0.3, and 0.2 for CNN models trained individually on the s-1 kV, s-2 kV, and s-4 kV datasets, respectively, signifying that the model learns valuable insights from the 2- and 4-kV datasets despite their *shine-through effects*. On the basis of these comparative findings, we decided to proceed with the late fusion architecture for all subsequent parts of this study.

Table 2 Performance comparison of the different multimodal architectures on the synthetic dataset

Architecture	MP ↓	MGP ↓	MDS ↑
Original	0	0	1
Early fusion	2.8	20.33	0.93
Intermediate fusion	2.58	6.92	0.94
Late fusion	0.23	0.94	0.99

The first row in the table represents the target metric values for the original dataset (ground truth)

3.2 Comparison of Segmentation Methods (Synthetic Data)

Subsequently, we performed a comparative analysis, pitting our multimodal ML method (ML-multiV) against the cluster-based k-means clustering algorithm and ML models trained using individual single kV datasets, denoted as ML-singleV. The results are tabulated in Table 3, featuring metrics such as MP, MGP, and MDS for the reconstructed simulated structure. ML-singleV (s-1 kV), ML-singleV (s-2 kV), and ML-singleV (s-4 kV) represent cases where the

Table 3 Quantitative evaluation of different segmentation methods using synthetic datasets

Method		k-means (k=3)			ML-singleV			ML-
Measure	Original	s-1 kV	s-2 kV	s-4 kV	s-1 kV	s-2 kV	s-4 kV	multiV
MP ↓	0.000	4.455	9.490	18.626	0.245	0.654	1.407	0.225
MGP ↓	0.000	8.776	15.573	20.904	0.963	2.512	6.466	0.944
MDS ↑	1.000	0.904	0.815	0.698	0.994	0.985	0.967	0.995

The last column shows the results of our new multimodal ML method trained on a synthetic multiV dataset

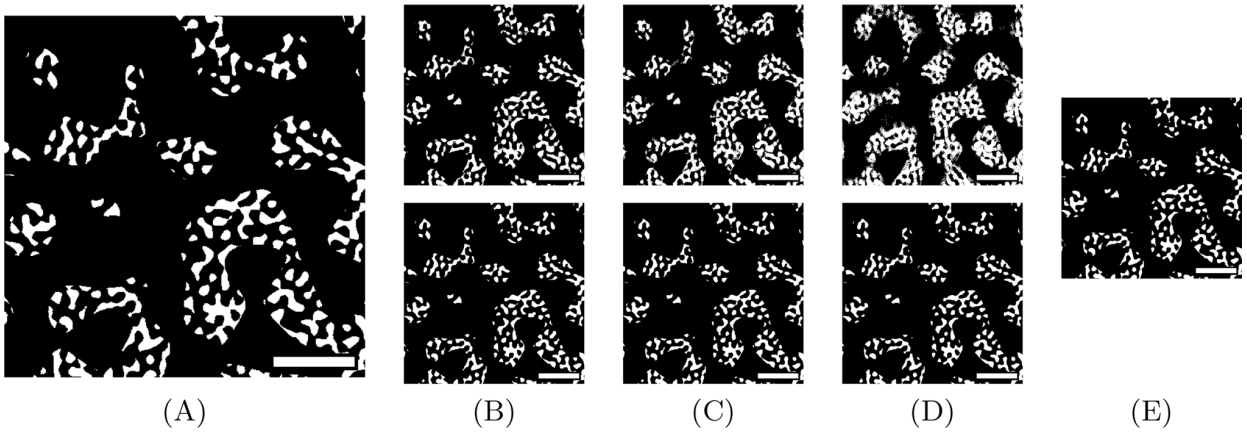


Fig. 7 **A** Slice of a synthetic microstructure (ground truth) and segmentation results of Monte Carlo-simulated BSE images using k-means clustering (top row) and the ML-singleV method (bottom row) of **B** s-1 kV, **C** s-2 kV, **D** s-4 kV datasets, and **E** our multimodal ML-multiV method. Scale bar: 300 nm

ML model was trained and tested using only a single kV dataset, s-1 kV, s-2 kV, and s-4 kV, respectively.

Notably, our multimodal ML method, ML-multiV, outperformed all alternative techniques assessed. All ML models surpassed the cluster-based k-means clustering method when applied to their respective individual datasets. The best results were observed for the multimodal ML approach, particularly when considering all three datasets for segmentation. This underscores the distinct advantage gained from the *additional* information acquired through images captured at various accelerating voltages.

Supplementary (Appendix) Table 7 provides a detailed exploration of anisotropy-based metrics for each segmentation method, reaffirming the trends observed in Table 3. In addition, Fig. 7 offers a visual representation of the segmentation outputs, further supporting the efficacy of our multimodal ML technique.

3.3 Comparison of Performance for Real HNPG Data

We have demonstrated the superior performance of our multimodal ML approach on synthetic datasets created using Monte Carlo simulations of the FIB tomography process. To validate these results on real data, we assessed the segmentation performance of our ML-multiV method on real HNPG FIB tomography data (r-1 kV, r-2 kV, and r-4 kV) using anisotropy-based metrics (see Sect. 2.5.2). The outcomes presented in Table 4 confirm the advantages of our method, ML-multiV, over alternatives when applied to real FIB tomography data. Figure 8 shows a single slice from a segmented real HNPG dataset using different segmentation methods. This figure illustrates the enhanced ability of all ML-based techniques in mitigating the *shine-through effect* when compared with classical segmentation methods. Note that ML-multiV, on average, exhibits more than 50% improvement in performance based on anisotropy measures compared with k-means clustering. Furthermore, the enhancement over k-means clustering (r-1 kV) in anisotropy measures exceeds 30%.

4 Conclusion

FIB-SEM tomography data are affected by artifacts such as the *shine-through effect* and the resulting ambiguity in image intensities. These artifacts make it difficult to use cluster-based segmentation methods, such as k-means clustering. More advanced ML-based methods can efficiently suppress the *shine-through effect*, even when trained only on a single set of synthetic FIB tomography images. The reconstruction performance can be further improved using more than one set of FIB tomography data for the same system. Exploiting this idea, we developed an overdetermined system for FIB-SEM tomography of nanostructured materials such as HNPG. We accomplished this by combining FIB tomography data collected for the same system at different acceleration voltages. The additional data collected in this way can help decide with greater certainty whether certain high-intensity pixels of the SEM image belong to the surface area or are rather shine-through artifacts from deeper layers that should ideally be neglected.

Table 4 Error measures based on anisotropy of reconstructed microstructure achieved using four different microstructure reconstruction methods

Method	k-means (k=3)			ML-singleV			ML-
Measure	r-1 kV	r-2 kV	r-4 kV	r-1 kV	r-2 kV	r-4 kV	multiV
$e_{L_2}^{\mathrm{TPCF}}\downarrow$	0.1645	0.1899	0.2742	0.1454	0.1686	0.2708	0.1206
$e_{L_2}^{\mathrm{LPF}}\downarrow$	0.0531	0.2494	0.2369	0.0591	0.0911	0.1810	0.0489
$e_{L_2}^{D}\downarrow$	0.0312	0.1943	0.1858	0.0304	0.0370	0.1383	0.0119

The best possible error value is 0 for all measures

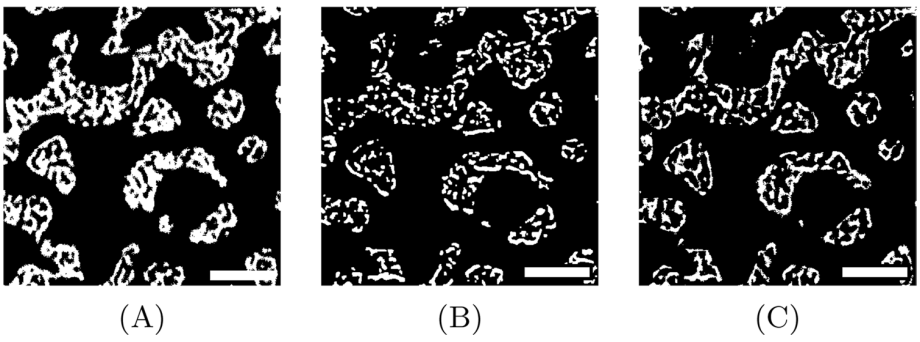


Fig. 8 Image segmentation results of an example region of a real HNPG dataset using **A** k-means clustering (r-1 kV), **B** ML-singleV (r-2 kV), and **C** our multimodal ML-multiV. Scale bar: 300 nm

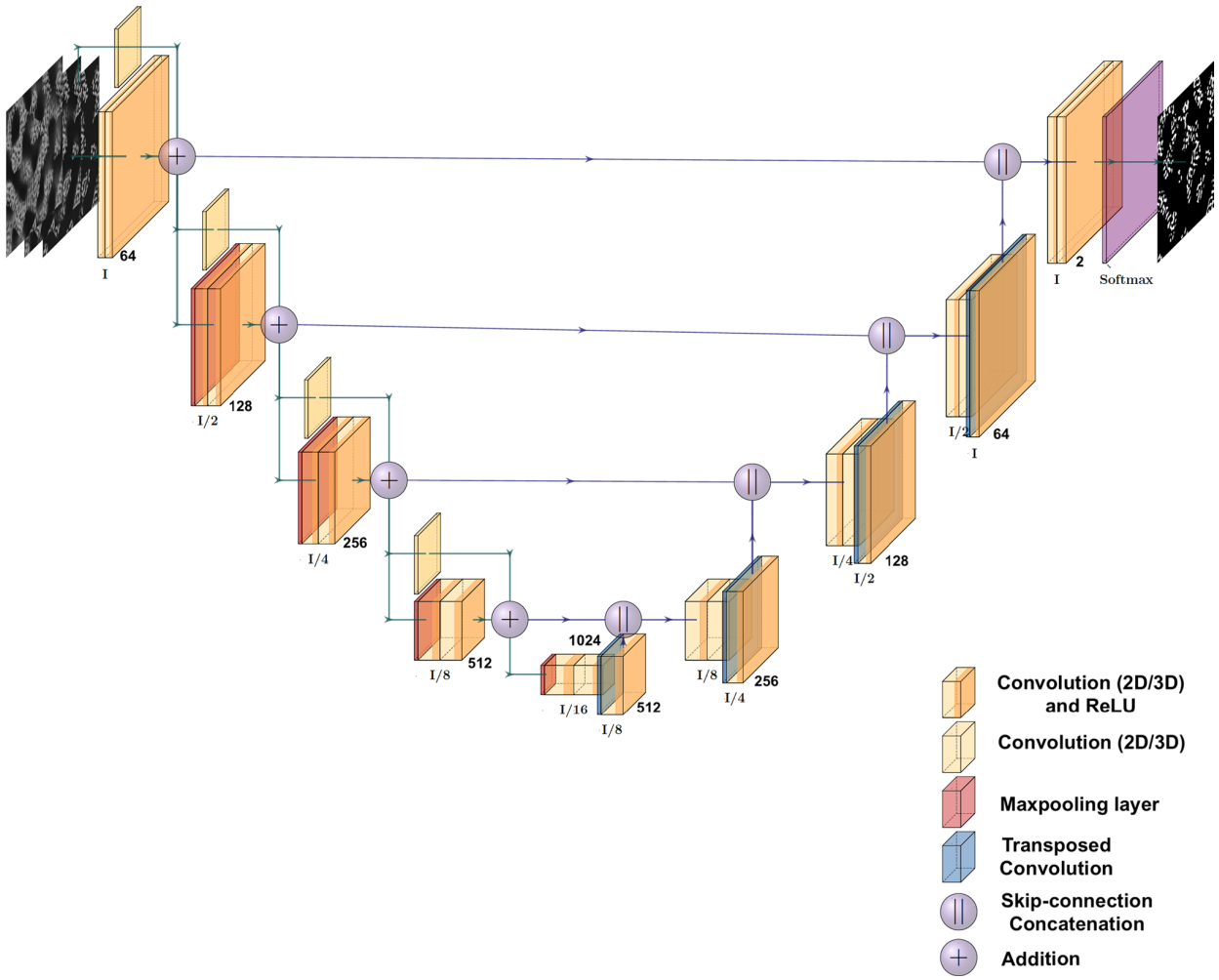


Fig. 9 Schematic representation of U-Net architecture with residual connections

This study demonstrated that a multimodal ML method using imaging data collected with various acceleration voltages can outperform classical segmentation methods using only one dataset (obtained with a single acceleration voltage). We demonstrated this specifically using a multimodal ML architecture that combined imaging information obtained at acceleration voltages of 1, 2, and 4 kV.

Appendix 1: Machine Learning Architecture

We designed custom U-Net-based architectures for multimodal ML models (Fig. 6). These architectures were tailored for each fusion method and differed in their approach to processing image data. In all U-Net architectures [32, 33], we incorporated customizations such as residual connections, padding, and the number of encoding/decoding blocks (Fig. 9). After evaluating the performance of various architectures for different fusion techniques, we selected the following:

In the early fusion approach, we opted for a 3D U-Net model comprising four encoding blocks with residual connections.

In the case of intermediate fusion, we narrowed our focus to a 2D CNN and a 2D CNN with adjacent slices because of the significant size of the multimodal ML model. To our surprise, the 2D CNN proved to be the most effective, probably because of the constraints of the limited dataset size and the overall large model size after intermediate fusion. Consequently, we adopted a 2D U-Net model that features two encoding blocks with residual connections. The fully connected layer block incorporated two convolutional layers: The first expanded output channels to 64, followed by a second output classifier layer with a kernel size of 1.

For the late fusion approach, we trained and compared the 2D CNN, 2D CNN with adjacent slices, and 3D CNN architectures, with the 2D CNN featuring adjacent slices outperforming the others. This approach employed seven adjacent slices, four encoding blocks, and residual connections and included an ensemble block with a convolutional layer and a kernel size of 1.

Appendix 2: Training Curves

Figure 10 depicts training and validation Dice loss curves. These curves display the mean Dice loss over epochs, with the blue line representing training scores and the orange line representing validation scores. We used data augmentation during training, which made the training dataset more complex. Consequently, validation Dice loss values were consistently lower than training Dice loss values in all curves.

Appendix 3: Effect of 3D Reconstruction on Material Properties (Solid Fraction)

The solid fraction (ϕ), which is a critical property in materials science, quantifies the volume of the solid phase within a structure. It significantly influences mechanical and optical properties, underscoring the importance of its accurate measurement, which relies on precise 3D reconstruction. The solid fraction can be calculated from the binary structure as follows:

$$\phi = \frac{N_m}{N_{\text{total}}} \tag{6}$$

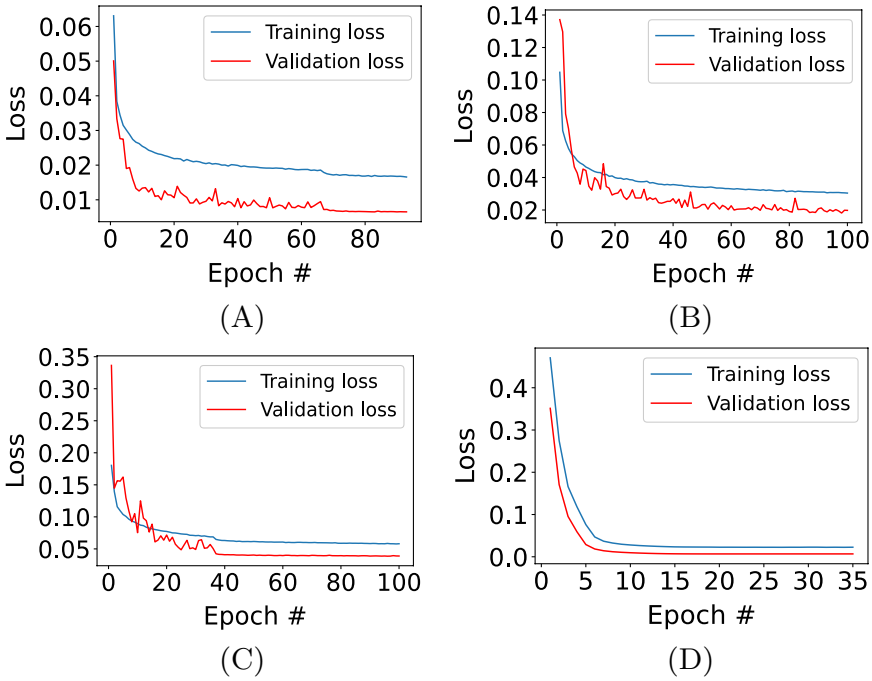


Fig. 10 Training and validation data Dice loss function for ML method singleV trained on **A** s-1 kV, **B** s-2 kV, **C** s-4 kV, and **D** our multimodal ML-multiV method

Table 5 Quantitative evaluation of different segmentation methods on material properties (solid fraction) of synthetic datasets

Method		k-means (k=3)			ML-singleV			ML-
Measure	Original	s-1 kV	s-2 kV	s-4 kV	s-1 kV	s-2 kV	s-4 kV	multiV
Solid Fraction	0.123	0.147	0.180	0.258	0.123	0.123	0.121	0.123

Table 6 Quantitative evaluation of different segmentation methods on material properties (solid fraction) in real datasets

Method	k-means (k=3)			ML-singleV			ML-
Measure	r-1 kV	r-2 kV	r-4 kV	r-1 kV	r-2 kV	r-4 kV	multiV
Solid Fraction	0.214	0.253	0.355	0.164	0.123	0.247	0.130

Table 7 Performance comparison of different segmentation methods based on anisotropy for synthetic datasets

Method		k-means (k=3)			ML-singleV			ML-
Measure	Original	s-1 kV	s-2 kV	s-4 kV	s-1 kV	s-2 kV	s-4 kV	multiV
$e_{L_2}^{\mathrm{TPCF}}\downarrow$	0.0965	0.1609	0.2313	0.2317	0.0963	0.0976	0.0936	0.0961
$e_{L_2}^{\mathrm{LPF}}\downarrow$	0.0087	0.0734	0.1442	0.2336	0.0100	0.0102	0.0143	0.0100
$e_{L_2}^{D}\downarrow$	0.0120	0.0583	0.1087	0.1805	0.0111	0.0091	0.0078	0.0106

where N_m represents voxels within phase m and N_{total} denotes the total voxel count. Synthetic data with ground truth allow us to observe the impact of reconstruction on the solid fraction. Table 5 shows that the ML-based methods yield similar solid fractions, while with k-means clustering, the solid fraction increases with increasing voltage, indicating its struggle with the *shine-through effect*.

Real FIB-SEM image data (r-1 kV, r-2 kV, and r-4 kV) lack ground truth, limiting direct comparison. Nevertheless, Table 6 shows that ML-singleV (r-1 kV), ML-singleV (r-2 kV), and ML-multiV exhibit solid fractions indicative of *realistic* hierarchical nanoporous materials.

Appendix 4: Additional Results Using Anisotropy-Based Error Measures for Synthetic Data

In this section, we computed anisotropy-based errors as detailed in Sect. 2.5.2 to compare the performance of different segmentation methods. Table 7 presents the comprehensive metrics for the synthetic datasets including for the original dataset, serving as reference values for evaluating the outcomes obtained from different methods.

Author Contributions All authors read and approved the final manuscript. Sardhara, Aydin, Cyron, and Ritter contributed to the conception and design of the study. Shkurmanov acquired FIB-SEM images of HNPG and calculated slice thicknesses. Li generated the LWM database. Shi and Riedel prepared the HNPG sample. Sardhara wrote the first draft of the manuscript. All authors contributed to the manuscript.

Funding This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - SFB 986 - Project number 192346071.

Availability of Data and Material The datasets generated and analyzed for this study can be found in TUHH Open Research repository. DOI: https://doi.org/10.15480/882.8927

Declarations

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Liu Y, Zhao T, Ju W, Shi S (2017) Materials discovery and design using machine learning. J Materiomics 3(3):159–177
2. Fam Y, Sheppard TL, Diaz A, Scherer T, Holler M, Wang W, Wang D, Brenner P, Wittstock A, Grunwaldt J-D (2018) Correlative multiscale 3D imaging of a hierarchical nanoporous gold catalyst by electron, ion and X-ray nanotomography. ChemCatChem 10(13):2858–2867
3. Shkurmanov A, Krekeler T, Ritter M (2022) Slice thickness optimization for the focused ion beam-scanning electron microscopy 3D tomography of hierarchical nanoporous gold. Nanomanuf Metrol 5(2):112–118
4. Richert C, Huber N (2018) Skeletonization, geometrical analysis, and finite element modeling of nanoporous gold based on 3D tomography data. Metals 8(4):282
5. Hu K, Ziehmer M, Wang K, Lilleodden ET (2016) Nanoporous gold: 3D structural analyses of representative volumes and their implications on scaling relations of mechanical behaviour. Phil Mag 96(32–34):3322–3335
6. Sardhara T, Shkurmanov A, Li Y, Shi S, Cyron CJ, Aydin RC, Ritter M (2023) Role of slice thickness quantification in the 3D reconstruction of FIB tomography data of nanoporous materials. Ultramicroscopy. https://doi.org/10.1016/j.ultramic.2023.113878
7. Prill T, Schladitz K, Jeulin D, Faessel M, Wieser C (2013) Morphological segmentation of FIB-SEM data of highly porous media. J Microsc 250(2):77–87
8. Fager C, Röding M, Olsson A, Lorén N, Corswant C, Särkkä A, Olsson E (2020) Optimization of FIB-SEM tomography and reconstruction for soft, porous, and poorly conducting materials. Microsc Microanal 26(4):837–845
9. Sardhara T, Shkurmanov A, Aydin R, Cyron CJ, Ritter M (2023) Towards an accurate 3D reconstruction of nano-porous structures using fib tomography and monte carlo simulations with machine learning. Microsc Microanal 29(Supplement_1):545–546
10. Rogge F, Ritter M (2019) Cluster analysis for FIB tomography of nanoporous materials. In: Conference: IMC19, Sydney
11. Fend C, Moghiseh A, Redenbach C, Schladitz K (2021) Reconstruction of highly porous structures from fib-sem using a

deep neural network trained on synthetic images. J Microsc 281(1):16–27
12. Sardhara T, Aydin RC, Li Y, Piché N, Gauvin R, Cyron CJ, Ritter M (2022) Training deep neural networks to reconstruct nanoporous structures from FIB tomography images using synthetic training data. Front Mater 9:837006
13. Liu Y, Yang Z, Zou X, Ma S, Liu D, Avdeev M, Shi S (2023) Data quantity governance for machine learning in materials science. Natl Sci Rev 125
14. Yuhas BP, Goldstein MH, Sejnowski TJ (1989) Integration of acoustic and visual speech signals using neural networks. IEEE Commun Mag 27(11):65–71
15. Beyer T, Townsend DW, Brun T, Kinahan PE, Charron M, Roddy R, Jerin J, Young J, Byars L, Nutt R (2000) A combined PET/CT scanner for clinical oncology. J Nucl Med 41(8):1369–1379
16. Bagci U, Udupa JK, Mendhiratta N, Foster B, Xu Z, Yao J, Chen X, Mollura DJ (2013) Joint segmentation of anatomical and functional images: Applications in quantification of lesions from PET, PET-CT, MRI-PET, and MRI-PET-CT images. Med Image Anal 17(8):929–945
17. Lian C, Ruan S, Denœux T, Li H, Vera P (2018) Joint tumor segmentation in PET-CT images using co-clustering and fusion based on belief functions. IEEE Trans Image Process 28(2):755–766
18. Suk H-I, Lee S-W, Shen D, Initiative ADN et al (2014) Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis. Neuroimage 101:569–582
19. Carneiro G, Nascimento J, Bradley AP (2015) Unregistered multiview mammogram analysis with pre-trained deep learning models. In: conference on medical image computing and computer-assisted intervention, pp 652–660. Springer
20. Guo Z, Li X, Huang H, Guo N, Li Q (2019) Deep learning-based image segmentation on multimodal medical imaging. IEEE Trans Radiat Plasma Med Sci 3(2):162–169
21. Huber R, Haberfehlner G, Holler M, Kothleitner G, Bredies K (2019) Total generalized variation regularization for multi-modal electron tomography. Nanoscale 11(12):5617–5632
22. Anderson TI, Vega B, Kovscek AR (2020) Multimodal imaging and machine learning to enhance microscope images of shale. Comput Geosci 145:104593
23. Shi S, Li Y, Ngo-Dinh B-N, Markmann J, Weissmüller J (2021) Scaling behavior of stiffness and strength of hierarchical network nanomaterials. Science 371(6533):1026–1033
24. Peña B, Owen GR, Dettelbach K, Berlinguette C (2018) Spin-coated epoxy resin embedding technique enables facile SEM/FIB thickness determination of porous metal oxide ultra-thin films. J Microsc 270(3):302–308
25. Thermo Fisher Scientific Inc: Auto slice and view 4.0 [computer software] (2017). Version: 4.1.0.1196. Accessed 2022-07-13
26. Lepinay K, Lorut F (2013) Three-dimensional semiconductor device investigation using focused ion beam and scanning electron microscopy imaging (FIB/SEM tomography). Microsc Microanal 19(1):85–92
27. Jones H, Mingard K, Cox D (2014) Investigation of slice thickness and shape milled by a focused ion beam for three-dimensional reconstruction of microstructures. Ultramicroscopy 139:20–28
28. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B et al (2012) Fiji: an open-source platform for biological-image analysis. Nat Methods 9(7):676–682
29. Soyarslan C, Bargmann S, Pradas M, Weissmüller J (2018) 3D stochastic bicontinuous microstructures: generation, topology and elasticity. Acta Mater 149:326–340
30. Object Research Systems (ORS) Inc, C. Montreal: Dragonfly 3.6 [computer software] (2018)
31. Perez L, Wang J (2017) The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621
32. Ronneberger O, Fischer P, Brox T (2015) U-Net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, October 5–9, 2015, Proceedings, Part III 18, pp 234–241. Springer
33. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Medical image computing and computer-assisted intervention–MICCAI 2016: 19th international conference, Athens, Greece, October 17–21, 2016, Proceedings, Part II 19, pp. 424–432. Springer

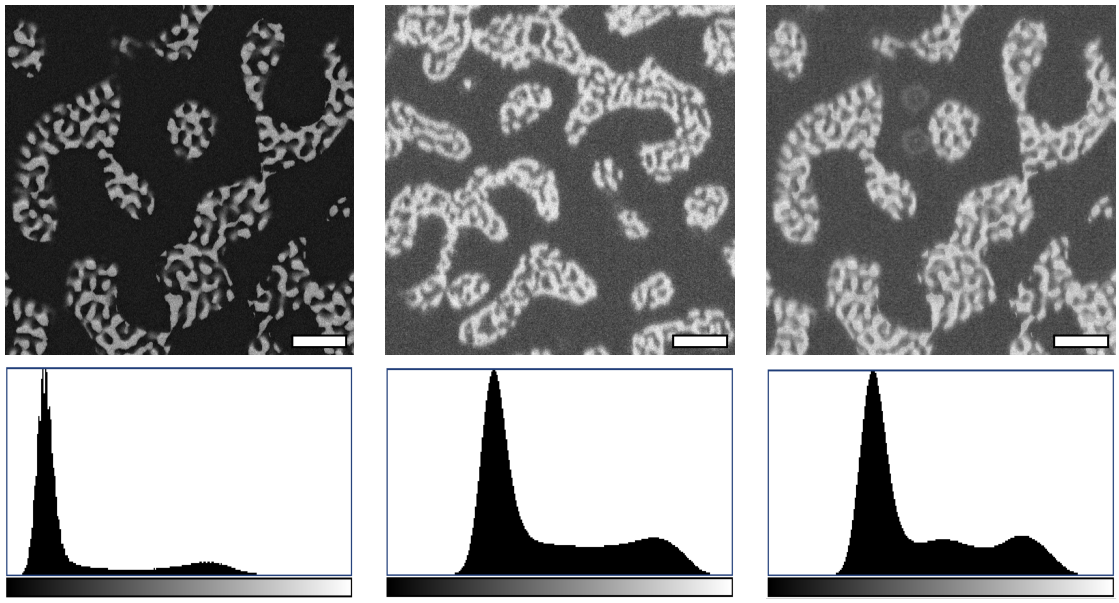
Chapter 6

Working towards unleashing full potential of machine learning

6.1 Domain adaptation using machine learning

The divergence between the data distributions of simulated and real datasets poses a significant challenge for machine learning-based segmentation methods (refer to Figure 6.1). Addressing this issue, [83] provides a comprehensive review of solutions, particularly in medical image analysis. [84] applied such a methodology for 3D surface reconstruction of fin-like structures using a scanning electron microscope. Simulated datasets tend to lack inherent imperfections in real-world data, such as random noise and irregular slice thickness introduced by nature and machine limitations. Moreover, replicating the exact BSE image across different electron microscopes proves challenging even with the same parameters [85, 86]. However, generative machine learning models offer a promising approach to mitigate these challenges through domain adaptation techniques, which seek to reduce the discrepancy between data distributions [41].

In electron microscopy, obtaining paired image data poses a significant challenge, as is common in many domains. This constraint necessitates training machine learning models in an unsupervised manner, as recommended by [43]. Initial results demonstrate the potential of machine learning-based domain adaptation methods in reducing the disparity in data distributions (see Figure 6.1).

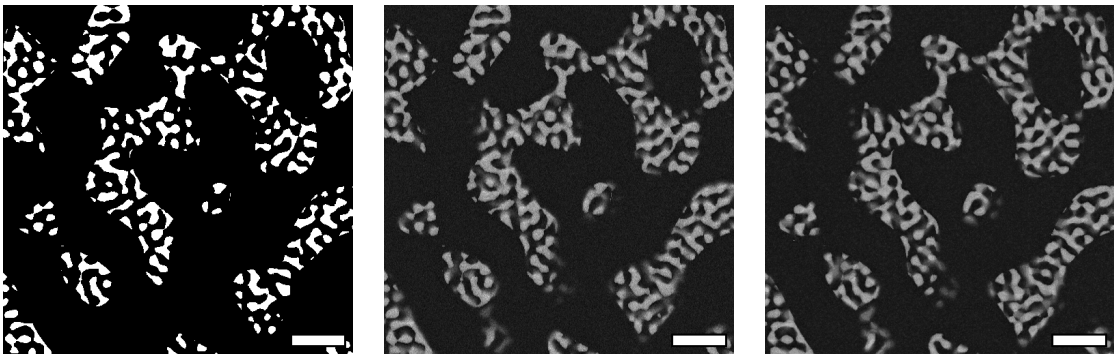


(A) (B) (C)

Fig. 6.1 **(A)** Slice of a synthetic example microstructure generated using MCXray plugin and **(B)** a slice of real HNPG structure and **(C)** a slice of synthetic dataset after domain adaptation - scale bar: 200 nm

6.2 BSE simulations using machine learning

Machine learning offers the potential to replace computationally intensive Monte Carlo simulations by integrating generative methods with physics-informed machine learning approaches. The latter holds the potential for harmonizing data with mathematical, physical models [87]. In this context, physics-informed machine learning can serve as a preliminary step to compute the backscattered electron coefficient and establish its correlation with slice depth, accurately emulating the *shine-through effect*. These preprocessed images are then fed into a generative machine learning model, resembling domain adaptation, to yield the final outcomes. The initial results of this pipeline, as shown in Figure 6.2, demonstrate the potential to simulate BSE images for materials beyond hierarchical nanoporous gold. The simulations obtained using our machine learning-based method take only 20 seconds once the model is optimized compared to some days every time, depending on available computation power.



(A) (B) (C)

Fig. 6.2 (A) A single image slice of digital microstructure generated using leveled wave method and corresponding simulated BSE image using (B) Monte Carlo method and (C) Machine learning method. (scale bar: 200 nm)

Chapter 7

Conclusion

FIB-SEM tomography is a powerful technique for obtaining high-resolution images of relatively large volumes, enabling their precise 3D reconstruction of micro and nanostructures. However, these images are often impaired by the *shine-through effect* and intensity ambiguities, complicating the semantic segmentation of the microstructures. This issue is particularly noticed when using traditional methods such as Otsu's thresholding or k-means clustering for hierarchical nanoporous materials.

Machine learning-based methods offer a promising solution for accurately segmenting these imperfect images. However, the lack of training data presents a significant hurdle to effectively applying these methods. To address this, a novel approach involving the generation of synthetic binarized microstructures is introduced. These binarized structures are then simulated using the Monte Carlo method. Among the various methods tested for generating such binarized microstructures, the leveled wave method proved to be the most effective for generating synthetic training data as expected.

This study identifies machine learning networks such as 2D CNNs with adjacent slices and 3D CNNs as the optimal architectures for semantic segmentation of hierarchical nanoporous gold. However, it is important to note that 3D CNNs demand significantly higher computational resources compared to their 2D counterparts. These machine learning-based models outperform traditional segmentation methods like Otsu's thresholding and k-means clustering.

FIB tomography is a destructive technique, making obtaining ground-truth values for real NPG and HNPG datasets challenging. This absence of ground truth is a significant obstacle in evaluating the performance of segmentation methods on real microstructures. To circumvent this issue, an innovative approach is proposed using the information that self-similar networks such as nanoporous and hierarchical nanoporous gold structures are isotropic structures. This isotropy or, subsequently, anisotropy

information can serve as a proxy for segmentation error when no ground-truth values are available.

Functions like two-point correlation and lineal path functions can be used to calculate the isotropy of any microstructures. However, in the post-processing of FIB tomography data, inaccuracies often arise due to variable slice thicknesses and disparities in resolution. To address these issues, a novel method that employs a slice repositioning algorithm combined with machine learning-based interpolation based on image inpainting is proposed. This method outperforms conventional interpolation techniques like nearest-neighbor or cubic splines and offers more accurate interpolation results. These findings suggest significant potential for interpolating complex microstructures to make their voxel size identical in all directions. However, it is important to note that further studies should be conducted on other microstructures to validate and corroborate these findings.

In an attempt to enhance segmentation performance further, an overdetermined system by combining FIB tomography data from different acceleration voltages is introduced. This system was formed on the principle that additional information due to the *shine-through effect* increases by increasing accelerating voltage. This multivoltage image approach significantly improved the segmentation results, highlighting the superiority of multimodal machine learning methods compared to classical and machine learning-based methods that rely on single datasets.

However, data distribution differences can be a major hurdle while unleashing the full performance of the trained machine learning models on real nanoporous materials. The initial results of this study show that this problem may be solved using machine learning-based domain adaptation techniques. Similarly, computation-intensive Monte Carlo-based simulation methods for generating synthetic BSE images may also be generated using physics-informed machine learning combined with generative machine learning models.

In summary, this study introduces various methods, including virtual FIB tomography data generation, isotropy-based validation, slice repositioning, image inpainting, and a multimodal machine learning approach based on multivoltage images. Incorporating these methods in the segmentation pipeline can be the new beginning of more accurate 3D reconstruction results of FIB tomography data, particularly for nanoporous materials.

References

[1] Trushal Sardhara, Alexander Shkurmanov, Roland Aydin, Christian J Cyron, and Martin Ritter. Towards an accurate 3D reconstruction of nano-porous structures using fib tomography and monte carlo simulations with machine learning. 2023.

[2] Pierre Hovington, Dominique Drouin, and Raynald Gauvin. CASINO: A new monte carlo code in C language for electron beam interaction—part i: Description of the program. *Scanning*, 19(1):1–14, 1997.

[3] Graham Knott, Herschel Marchman, David Wall, and Ben Lich. Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. *Journal of Neuroscience*, 28(12):2959–2964, 2008.

[4] BJ Inkson, M Mulvihill, and G Möbus. 3D determination of grain shape in a FeAl-based nanocomposite by 3D FIB tomography. *Scripta materialia*, 45(7):753–758, 2001.

[5] Clarissa Villinger, Heiko Gregorius, Christine Kranz, Katharina Höhn, Christin Münzberg, Götz von Wichert, Boris Mizaikoff, Gerhard Wanner, and Paul Walther. FIB/SEM tomography with TEM-like resolution for 3D imaging of high-pressure frozen cells. *Histochemistry and cell biology*, 138(4):549–556, 2012.

[6] Yang Liu, Helen E King, Marijn A Van Huis, Martyn R Drury, and Oliver Plümper. Nano-tomography of porous geological materials using focused ion beam-scanning electron microscopy. *Minerals*, 6(4):104, 2016.

[7] NJ Welch, F Gray, AR Butcher, ES Boek, and JP Crawshaw. High-resolution 3D FIB-SEM image analysis and validation of numerical simulations of nanometre-scale porous ceramic with comparisons to experimental results. *Transport in Porous Media*, 118(3):373–392, 2017.

[8] S Korte, Martin Ritter, Chengge Jiao, PA Midgley, and WJ Clegg. Three-dimensional electron backscattered diffraction analysis of deformation in mgo micropillars. *Acta Materialia*, 59(19):7241–7254, 2011.

[9] Kodanda Ram Mangipudi, Volker Radisch, Lorenz Holzer, and Cynthia A Volkert. A FIB-nanotomography method for accurate 3d reconstruction of open nanoporous structures. *Ultramicroscopy*, 163:38–47, 2016.

[10] Kaixiong Hu, Markus Ziehmer, Ke Wang, and Erica T Lilleodden. Nanoporous gold: 3D structural analyses of representative volumes and their implications on

scaling relations of mechanical behaviour. *Philosophical Magazine*, 96(32-34):3322–3335, 2016.

[11] Yakub Fam, Thomas L Sheppard, Ana Diaz, Torsten Scherer, Mirko Holler, Wu Wang, Di Wang, Patrice Brenner, Arne Wittstock, and Jan-Dierk Grunwaldt. Correlative multiscale 3D imaging of a hierarchical nanoporous gold catalyst by electron, ion and X-ray nanotomography. *ChemCatChem*, 10(13):2858–2867, 2018.

[12] Cecilia Fager, Magnus Röding, Anna Olsson, Niklas Lorén, Christian von Corswant, Aila Särkkä, and Eva Olsson. Optimization of FIB-SEM tomography and reconstruction for soft, porous, and poorly conducting materials. *Microscopy and Microanalysis*, 26(4):837–845, 2020.

[13] Trushal Sardhara, Roland C Aydin, Yong Li, Nicolas Piché, Raynald Gauvin, Christian J Cyron, and Martin Ritter. Training deep neural networks to reconstruct nanoporous structures from FIB tomography images using synthetic training data. *Frontiers in Materials*, 9:837006, 2022.

[14] Trushal Sardhara, Alexander Shkurmanov, Yong Li, Shan Shi, Christian J. Cyron, Roland C. Aydin, and Martin Ritter. Role of slice thickness quantification in the 3D reconstruction of FIB tomography data of nanoporous materials. *Ultramicroscopy*, page 113878, 2023.

[15] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66, 1979.

[16] Stuart Lloyd. Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[17] M Salzer, Torben Prill, A Spettl, D Jeulin, K Schladitz, and V Schmidt. Quantitative comparison of segmentation algorithms for FIB-SEM images of porous media. *Journal of microscopy*, 257(1):23–30, 2015.

[18] F Rogge and M Ritter. Cluster analysis for FIB tomography of nanoporous materials". *Conference: IMC19 Sydney*, 2019.

[19] Keiron O'Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[20] C Fend, A Moghiseh, C Redenbach, and K Schladitz. Reconstruction of highly porous structures from fib-sem using a deep neural network trained on synthetic images. *Journal of Microscopy*, 281(1):16–27, 2021.

[21] Léon Bottou. From machine learning to machine reasoning: An essay. *Machine learning*, 94:133–149, 2014.

[22] Nadia Burkart and Marco F Huber. A survey on the explainability of supervised machine learning. *Journal of Artificial Intelligence Research*, 70:245–317, 2021.

[23] Sotiris B Kotsiantis, Ioannis Zaharakis, P Pintelas, et al. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1):3–24, 2007.

[24] Carla E Brodley, Umaa Rebbapragada, Kevin Small, and Byron Wallace. Challenges and opportunities in applied machine learning. *Ai Magazine*, 33(1):11–24, 2012.

[25] JORGE Perdigao, P Lambrechts, and G Vanherle. Microscopy investigations: techniques, results, limitations. *Am. J. Dent*, 13:3D18D, 2000.

[26] Sergey I Nikolenko. *Synthetic data for deep learning*, volume 174. Springer, 2021.

[27] David C Joy. *Monte Carlo modeling for electron microscopy and microanalysis*, volume 9. Oxford University Press, 1995.

[28] Raynald Gauvin and Pierre Michaud. MC X-ray, a new Monte Carlo program for quantitative x-ray microanalysis of real materials. *Microscopy and Microanalysis*, 15(S2):488–489, 2009.

[29] Canada Object Research Systems (ORS) Inc, Montreal. Dragonfly 3.6 [computer software]. 2018.

[30] J R Lowney and Egon Marx. User's manual for the program MONEL-I: Monte Carlo simulation of SEM signals for linewidth metrology. 1995.

[31] J R Lowney. MONSEL-II Monte Carlo simulation of sem signals for linewidth metrology. 1995.

[32] Albert Karabekov, Oren Zoran, Zvika Rosenberg, and Guy Eytan. Using Monte Carlo simulation for accurate critical dimension metrology of super small isolated poly-lines. *Scanning: The Journal of Scanning Microscopies*, 25(6):291–296, 2003.

[33] P Zhang, HY Wang, YG Li, SF Mao, and ZJ Ding. Monte Carlo simulation of secondary electron images for real sample structures in scanning electron microscopy. *Scanning*, 34(3):145–150, 2012.

[34] Raynald Gauvin, Eric Lifshin, Hendrix Demers, Paula Horny, and Helen Campbell. Win X-ray: A new monte carlo program that computes X-ray spectra obtained with a scanning electron microscope. *Microscopy and Microanalysis*, 12(1):49–64, 2006.

[35] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. A survey on deep transfer learning. In *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27*, pages 270–279. Springer, 2018.

[36] Yue Liu, Zhengwei Yang, Xinxin Zou, Shuchang Ma, Dahui Liu, Maxim Avdeev, and Siqi Shi. Data quantity governance for machine learning in materials science. *National Science Review*, page nwad125, 2023.

[37] Petr Cizmar, András E Vladár, Bin Ming, Michael T Postek, National Institute of Standards, and Technology. Simulated sem images for resolution measurement. *Scanning*, 30(5):381–391, 2008.

[38] F Timischl, M Date, and S Nemoto. A statistical model of signal–noise in scanning electron microscopy. *Scanning*, 34(3):137–144, 2012.

[39] KS Sim, JTL Thong, and JCH Phang. Effect of shot noise and secondary emission noise in scanning electron microscope images. *Scanning: The Journal of Scanning Microscopies*, 26(1):36–40, 2004.

[40] Joseph I Goldstein, Dale E Newbury, Joseph R Michael, Nicholas WM Ritchie, John Henry J Scott, and David C Joy. *Scanning electron microscopy and X-ray microanalysis*. springer, 2017.

[41] Wouter M Kouw and Marco Loog. An introduction to domain adaptation and transfer learning. *arXiv preprint arXiv:1812.11806*, 2018.

[42] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[43] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[44] Zhen Qi and Jorg Weissmuller. Hierarchical nested-network nanostructure by dealloying. *Acs Nano*, 7(7):5948–5954, 2013.

[45] Salvatore Torquato and HW Haslach Jr. Random heterogeneous materials: microstructure and macroscopic properties. *Appl. Mech. Rev.*, 55(4):B62–B63, 2002.

[46] James G Berryman. Measurement of spatial correlation functions using image processing techniques. *Journal of Applied Physics*, 57(7):2374–2384, 1985.

[47] KNOLL Max. Static potential and secondary emission of bodies under electron irradiation. *Z Tech Physik*, 16:467, 1935.

[48] M Knoll and R Theile. Scanning electron microscope for determining the topography of surfaces and thin layers. *Z Phys*, 113:260, 1939.

[49] Von Ardenne. The scanning electron microscope: theoretical fundamentals (in german). *Z Physk*, 109:553, 1938.

[50] Sir Charles William Oatley. *The Scanning Electron Microscope: Pt. 1-Oatley, CW the instrument*. Cambridge University Press, 1972.

[51] Richard J Young and Mary V Moore. Dual-beam (FIB-SEM) systems: techniques and automated applications. *Introduction to focused ion beams: instrumentation, theory, techniques and practice*, pages 247–268, 2005.

[52] Fred A Stevie, LA Giannuzzi, and BI Prenitzer. The focused ion beam instrument. *Introduction to focused ion beams: instrumentation, theory, techniques and practice*, pages 1–12, 2005.

[53] Ron Anderson and Stanley J Klepeis. Practical aspects of FIB TEM specimen preparation: With emphasis on semiconductor applications. *Introduction to Focused Ion Beams: Instrumentation, Theory, Techniques and Practice*, pages 173–200, 2005.

[54] Jörg Weissmüller, Roger C Newman, Hai-Jun Jin, Andrea M Hodge, and Jeffrey W Kysar. Nanoporous metals by alloy corrosion: formation and mechanical properties. *Mrs Bulletin*, 34(8):577–586, 2009.

[55] Erkin Seker, Michael L Reed, and Matthew R Begley. Nanoporous gold: Fabrication, characterization, and applications. *Materials*, 2(4):2188–2215, 2009.

[56] Shan Shi, Yong Li, Bao-Nam Ngo-Dinh, Jürgen Markmann, and Jörg Weissmüller. Scaling behavior of stiffness and strength of hierarchical network nanomaterials. *Science*, 371(6533):1026–1033, 2021.

[57] B Peña, G Rh Owen, KE Dettelbach, and CP Berlinguette. Spin-coated epoxy resin embedding technique enables facile SEM/FIB thickness determination of porous metal oxide ultra-thin films. *Journal of Microscopy*, 270(3):302–308, 2018.

[58] Celal Soyarslan, Swantje Bargmann, Marc Pradas, and Jörg Weissmüller. 3D stochastic bicontinuous microstructures: Generation, topology and elasticity. *Acta materialia*, 149:326–340, 2018.

[59] H-J Hunger and L Küchler. Measurements of the electron backscattering coefficient for quantitative EPMA in the energy range of 4 to 40 kev. *physica status solidi (a)*, 56(1):K45–K48, 1979.

[60] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133, 1943.

[61] Halbert White. Learning in artificial neural networks: A statistical perspective. *Neural computation*, 1(4):425–464, 1989.

[62] Li Deng, Dong Yu, et al. Deep learning: methods and applications. *Foundations and trends® in signal processing*, 7(3–4):197–387, 2014.

[63] Pádraig Cunningham, Matthieu Cord, and Sarah Jane Delany. Supervised learning. In *Machine learning techniques for multimedia: case studies on organization and retrieval*, pages 21–49. Springer, 2008.

[64] Zoubin Ghahramani. Unsupervised learning. In *Summer school on machine learning*, pages 72–112. Springer, 2003.

[65] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[66] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[67] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.

[68] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424, 2000.

[69] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. *Advances in neural information processing systems*, 19, 2006.

[70] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 689–696, 2011.

[71] Paul Pu Liang, Amir Zadeh, and Louis-Philippe Morency. Foundations and recent trends in multimodal machine learning: Principles, challenges, and open questions. *arXiv preprint arXiv:2209.03430*, 2022.

[72] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016.

[73] Martin Liggins II, David Hall, and James Llinas. *Handbook of multisensor data fusion: theory and practice*. CRC press, 2017.

[74] Ampere A Tseng. Recent developments in micromilling using focused ion beam technology. *Journal of micromechanics and microengineering*, 14(4):R15, 2004.

[75] Lorenz Holzer, Fedir Indutnyi, PH Gasser, Beat Münch, and Markus Wegmann. Three-dimensional analysis of porous batio3 ceramics using FIB nanotomography. *Journal of microscopy*, 216(1):84–95, 2004.

[76] Bart Winiarski, Ali Gholinia, K Mingard, M Gee, GE Thompson, and PJ Withers. Broad ion beam serial section tomography. *Ultramicroscopy*, 172:52–64, 2017.

[77] Miroslava Schaffer, Julian Wagner, Bernhard Schaffer, Mario Schmied, and Hans Mulders. Automated three-dimensional X-ray analysis using a dual-beam FIB. *Ultramicroscopy*, 107(8):587–597, 2007.

[78] Marco Cantoni and Lorenz Holzer. Advances in 3D focused ion beam tomography. *Mrs Bulletin*, 39(4):354–360, 2014.

[79] HG Jones, KP Mingard, and DC Cox. Investigation of slice thickness and shape milled by a focused ion beam for three-dimensional reconstruction of microstructures. *Ultramicroscopy*, 139:20–28, 2014.

[80] Alexander Shkurmanov, Tobias Krekeler, and Martin Ritter. Slice thickness optimization for the focused ion beam-scanning electron microscopy 3D tomography of hierarchical nanoporous gold. *Nanomanufacturing and Metrology*, pages 1–7, 2022.

[81] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[82] Hsieh Hou and H Andrews. Cubic splines for image interpolation and digital filtering. *IEEE Transactions on acoustics, speech, and signal processing*, 26(6):508–517, 1978.

[83] Hao Guan and Mingxia Liu. Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185, 2021.

[84] Tim Houben, Thomas Huisman, Maxim Pisarenco, Fons van der Sommen, and Peter de With. Training procedure for scanning electron microscope 3D surface reconstruction using unsupervised domain adaptation with simulated data. *Journal of Micro/Nanopatterning, Materials, and Metrology*, 22(3):031208–031208, 2023.

[85] Carlos Becker, C Mario Christoudias, and Pascal Fua. Domain adaptation for microscopy imaging. *IEEE Transactions on Medical Imaging*, 34(5):1125–1139, 2014.

[86] Róger Bermúdez-Chacón, Carlos Becker, Mathieu Salzmann, and Pascal Fua. Scalable unsupervised domain adaptation for electron microscopy. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 326–334. Springer, 2016.

[87] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.