Generating ideal synthetic data for 3D reconstruction of FIB tomography data using generative adversarial networks

Anonymous Author(s)
Affiliation
Address
email

Abstract

Accurate 3D reconstruction of nanomaterials is essential for studying their physical properties. Focused Ion Beam (FIB) tomography is a preferred method for creating 3D image stacks of micrometer-sized material volumes at nanometer resolution. To achieve valid 3D reconstructions, it is crucial to segment these images using machine learning-based methods, as they help mitigate artifacts in the data. However, supervised machine learning requires a large amount of training data and ground truth, which is challenging because FIB tomography is a destructive technique. While training machine learning models on synthetic data and applying this to real data is possible, it is only partially accurate due to differences in data distributions. Moreover, generating synthetic training data is time-consuming, even with modern computing, because of the complex physical Monte Carlo modeling. This study proposes a machine learning pipeline that reduces the difference in FIB tomography data distribution using domain adaptation techniques and introduces a novel method for quickly generating synthetic data by considering physical effects without Monte Carlo simulations.

1 Introduction

Nanoporous materials have significant potential in fields like materials science and biochemistry due to their unique properties. To understand these properties, accurate 3D reconstruction of their structure is often required. When studying nanoporous materials, such as hierarchical nanoporous gold (HNPG), electron microscopy (EM) is one of the few imaging methods that can provide the necessary resolution, as pore sizes can be smaller than 20 nm. Focused ion beam (FIB) combined with a scanning electron microscope (SEM) allows for high-resolution volumetric data collection of such nanomaterials, with in-plane (xy) resolution of 1 nm and depth resolution of 10 nm or less. This is achieved by removing material and imaging the newly exposed cross-sections consecutively (1). However, because FIB tomography is a destructive technique, obtaining ground-truth values for the structure being studied is impossible.

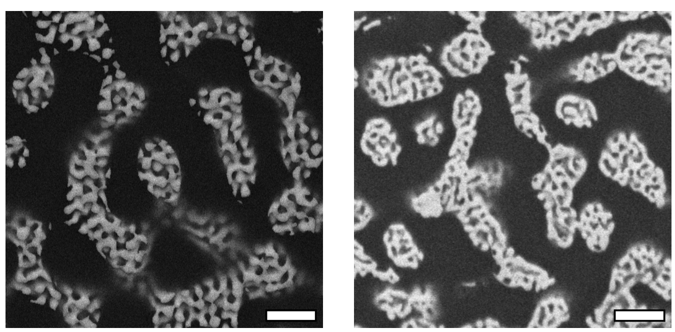
Additionally, these high-resolution image stacks often contain artifacts, such as the *shine-through effect* and intensity ambiguities (2). The *shine-through effect* occurs when structures from deeper layers become visible in the current milling plane, introducing extra information to the images. As a result, it is very challenging to uniquely map intensity and structural information in gray-scale FIB tomography images.

These challenges make it hard to semantically segment FIB tomography images using traditional methods like thresholding or k-means clustering, which rely mainly on intensity values. However, as demonstrated in (3), combining FIB tomography with machine learning can lead to accurate 3D reconstructions of nanomaterials. In their study, they trained machine learning models on synthetic data and then applied this knowledge to extract structural information from real FIB tomography data.

Training deep learning models with synthetic data is especially valuable in electron microscopy, where acquiring real data is difficult and expensive. In FIB tomography, the sample is destroyed, making it impossible to obtain ground-truth data. Synthetic data, however, can be easily reproduced and often include ground truth values, making them highly useful. Researchers have successfully used synthetic data in various fields, such as training machine learning models for autonomous vehicles to handle rare scenarios (4; 5). In electron microscopy, (6) generated simulated images of basic geometries for training purposes.

However, generating synthetic electron microscopy data presents challenges. It is time-consuming and resource-intensive, and creating realistic synthetic data requires comprehensive physical knowledge. The best synthetic data in electron microscopy are generated using Monte Carlo simulations, as suggested by (7). However, simulating 512 slices of 512 x 512 pixels using the Monte Carlo plugin in Dragonfly software (8) can take time in days on a high-performance CPU due to the need to calculate each electron trajectory.

Monte Carlo simulations are generally slow because they require iterative calculations for each event. Recently, generative adversarial networks (GANs) have shown promise in generating these events more efficiently. For instance, (9) demonstrated the use of GANs in particle physics, while (10) suggested a GAN-based approach for simulating electron-proton scattering events. Additionally, (11) proposed a 3D GAN based on StyleGAN2 (12) to generate realistic MR images. In electron microscopy, (13) utilized CycleGAN to create realistic scanning transmission electron images. However, to effectively replace Monte Carlo simulations, which incorporate physics-based knowledge, generative networks must also be provided with the necessary information to accurately simulate data using machine learning.



(A) (B)

Figure 1: Slice of a **(A)** synthetic hierarchical nanoporous gold (HNPG) generated using MCXray plugin and **(B)** real HNPG structure - scale bar: 200 nm

When generating synthetic data, it is not always guaranteed that it will match the distribution of real data. This mismatch often occurs because synthetic data typically follow strict physical formulas, making them too *ideal*. As a result, they lack the randomness and unique characteristics found in real-world data (see Figure 1). Due to these differences, models trained on synthetic data may underperform when applied to real data (14; 15).

In this study, we make two key contributions. First, we introduce a synthetic data pipeline that replaces the time-intensive Monte Carlo simulations with generative networks. This pipeline includes an additional artificial neural network (ANN) block designed to incorporate critical characteristics, such as realistic structural details and the physical relationships between variables like voltage, penetration depth, and atomic number. Second, we enhance the semantic segmentation of hierarchical nanoporous gold (HNPG) by applying domain adaptation techniques using generative networks. Our results demonstrate that our novel machine learning-based synthetic data pipeline performs on par with Monte Carlo-based simulation methods. Additionally, by applying domain adaptation to synthetic data before training, we achieve, on average, a 20% improvement in 3D reconstruction accuracy for semantic segmentation tasks.

2 Method

2.1 Acquiring Imaging Data

In this study, we analyzed real HNPG samples and generated synthetic samples using various methods.

2.1.1 Synthetic Samples

We began by generating binarized structures using the leveled wave method (LWM) (16), as described in (17). These binarized structures served as the first step in our synthetic FIB tomography data generation pipeline and provided the ground-truth values for our synthetic data.

Next, we used the Monte Carlo plugin in Dragonfly software (8) to generate *realistic* synthetic FIB tomography images, which incorporate *nearly* all relevant electron microscope physics (7). We created three synthetic datasets with different voltages (1kV, 2kV, and 4kV) using the same initial structure. These multi-voltage datasets (sMC-1kV, sMC-2kV, and sMC-4kV) are suitable for training multimodal machine learning models. Additionally, we generated three more synthetic datasets (sML-1kV, sML-2kV, and sML-4kV) using the same binarized structure but with the machine learning method outlined in Section 2.3 instead of Monte Carlo simulations. To evaluate the impact of domain adaptation, we also created domain-adapted versions of the Monte Carlo datasets (sMCDA-1kV, sMCDA-2kV, and sMCDA-4kV).

Furthermore, we prepared a dataset, sMC-BS, using Monte Carlo simulations to capture *realistic* material contrasts. We generated 50 virtual structures, each with an epoxy layer of varying thickness (0 to 190 nm in 10 nm increments) over a bulk material. These structures were simulated at different accelerating voltages and with various materials. By calculating the mean voxel intensities for each structure, we derived *realistic* backscattered coefficients. These coefficients, along with atomic number, accelerating voltage, and penetration depth, were used to train our artificial neural network (ANN) in a supervised manner.

2.1.2 Real Samples

An HNPG sample with a uniform random network structure and ligament sizes of 15 and 110 nm was prepared using the dealloying-coarsening-dealloying method (17). To enhance SEM imaging for solid-pore phase differentiation, the sample was infiltrated with epoxy resin (18). Following the approach in (19), multi-voltage FIB tomography was conducted using a Dual Beam FEI HeliosNanoLab G3 system with ASV4 software for automated control (20), which also monitored milling progress and compensated for drift (21).

To optimize HNPG tomography, two fiducial markers with intersecting trenches and a ruler system were used for drift compensation and precise slice thickness determination (22; 23). Three datasets at different accelerating voltages (1kV, 2kV, and 4kV) were prepared as described in (19) and are referred to as r-1kV, r-2kV, and r-4kV.

2.2 Machine Learning Architecture

We employed an encoder-decoder model based on cycle-consistent adversarial networks (CycleGAN) (24). This architecture utilizes two GANs with identical structures, each focusing on different tasks: the first maps data from the source domain to the target domain, while the second maps data from the target domain back to the source domain. This design enables the model to learn without requiring paired image data, making it suitable for unsupervised tasks.

For the encoder, we used a customized U-Net (25) with residual connections, and for the decoder, we adopted the architecture proposed in (24).

The cycle loss concept was key in addressing two primary challenges in this study: generating synthetic data and minimizing differences in data distribution.

2.3 Generating Synthetic Data Using Machine Learning

To efficiently generate Monte Carlo-like images, we employed a CycleGAN model combined with an artificial neural network (ANN) trained on physics-based data.

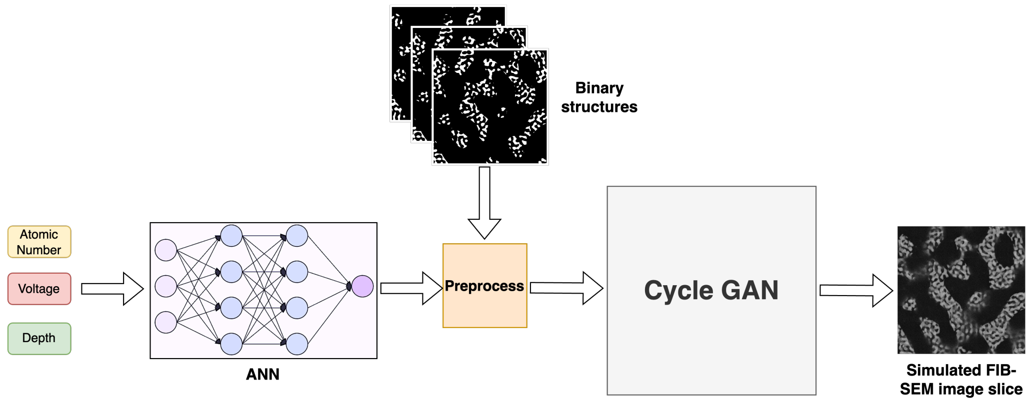


Figure 2: Block diagram of the experimental setup to obtain synthetic images using machine learning based on physical effects

In electron microscopy, the relationship between the backscattered electron coefficient (η), accelerating voltage (E), and atomic number (Z) is well-established (Equation 1) (26). However, a direct relation between η and penetration depth has not been explicitly defined in existing literature. To address this, we defined a basic ANN model, which had three input features, followed by four hidden layers containing 16, 32, 32, and 32 neurons, respectively, and a single output to map the backscattered electron coefficient with accelerating voltage, atomic number, and penetration depth.

$$\eta(Z, E) = E^{m(Z)*C(Z)} \tag{1}$$

Where

$$m(Z) = 0.1382 - \frac{0.9211}{Z^{0.5}} \tag{2}$$

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

Then, we trained this ANN model in a supervised manner using the sMC-BS dataset (see Section 2.1.1) to obtain the necessary backscattered coefficients to preprocess the leveled wave method (LWM) data before feeding it into the CycleGAN model. These backscattered coefficients were normalized across groups of atomic numbers and accelerating voltages and then applied as weights to generate weighted grayscale images from the binarized LWM data, typically using 10-15 slices to create the final preprocessed dataset.

These preprocessed gray-scale images served as the source domain, while the corresponding Monte Carlo simulation data acted as the target domain. The CycleGAN model was trained in an unsupervised manner, similar to domain adaptation, to enhance the robustness of our pipeline. The complete pipeline for this approach is illustrated in Figure 2.

2.4 Reducing Data Distribution Differences Using Machine Learning

To address the data distribution discrepancies between synthetic (source domain) and real FIB tomography data (target domain), we utilized a CycleGAN-based approach for unpaired image style transfer. This process, known as domain adaptation, was applied to synthetic datasets (sMC-1kV, sMC-2kV, sMC-4kV, sML-1kV, sML-2kV, and sML-4kV) to produce the final training datasets. The full pipeline for generating domain-adapted synthetic data and performing semantic segmentation on real HNPG data is illustrated in Figure 3.

2.5 Semantic Segmentation Using Machine Learning

After domain adaptation, we conducted semantic segmentation on all datasets following the approach outlined in (3). We trained three different models to evaluate performance: one using domain-adapted

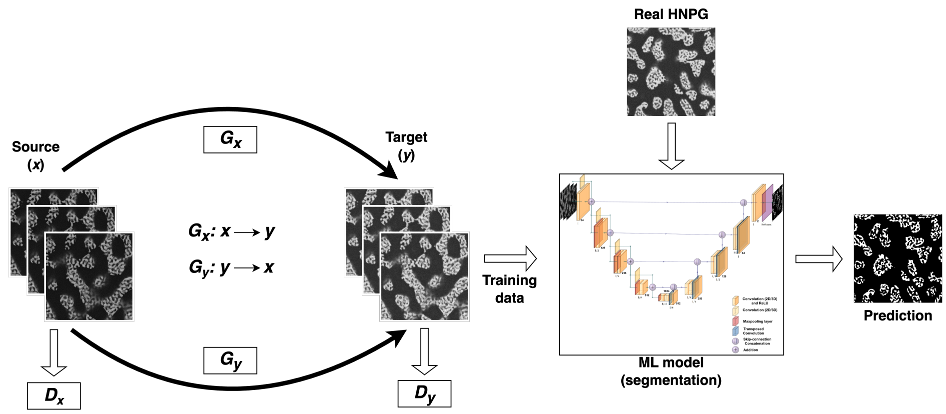


Figure 3: Block diagram of the experimental setup to reduce data distribution difference between synthetic and real HNPG images. G_x and G_y are CNN-based generators, and D_x and D_y are discriminators for the source domain (x) and the target domain (y) respectively

data, another using Monte Carlo simulation data, and a third using data generated with our proposed method. This comparative analysis allowed us to measure the improvement in segmentation accuracy due to domain adaptation and our innovative data generation technique.

2.6 Training Procedure

All machine learning models were trained on RTX 3090 GPUs. For CycleGAN, images were cropped into smaller patches (128 $\times$ 128) with a 64-pixel stride using a sliding window technique, and training was conducted on these 2D image patches. The loss functions employed were consistent with those proposed in (24), with an initial learning rate of 0.001, which decayed by a factor of 10 if no improvement in loss was observed for ten consecutive epochs.

For semantic segmentation, we adopted a structured training approach, presenting data as individual 2D slices, 3D volumetric stacks, or 2D slices combined with neighboring slices in smaller patches (64 $\times$ 64). The models were optimized using Dice loss in conjunction with the Adam optimizer, with an initial learning rate of 0.0001, reduced by a factor of 10 after 10 epochs without improvement.

The ANN model, used in the synthetic data generation pipeline, was trained using mean squared error (MSE) loss, with a learning rate decay scheme similar to that of the segmentation models. Detailed training parameters for all models are summarized in Table 1.

Table 1: Summary of parameters used for training ML models

Parameter	GANs	Semantic Segmentation	ANN
Patch size	128	64	-
Stride	0.5	0.5	-
Batch size	1	64	8
Epochs	100 with early stopping with patience=25		
Loss	CycleGAN Loss	Dice loss	MSE
Optimizer	Adam		
Learning rate	0.0001, adapted with a patience of 10 (reduction factor 10)		

2.7 Evaluation Criteria

We used two types of accuracy metrics to evaluate our methods: those based on ground truth values for synthetic datasets and anisotropy-based metrics for real data where ground truth is unavailable. These metrics are inspired by (3).

2.7.1 Synthetic Data

For synthetic datasets, where ground truth data is available, we used three metrics:

First, misplaced pixels (MP) measures the fraction of incorrectly classified pixels compared to the ground truth. It is calculated as:

$$MP = \left(1 - \frac{TP + TN}{TP + FP + FN + TN}\right) \times 100 \tag{4}$$

where TP, TN, FP, and FN are true positives, true negatives, false positives, and false negatives, respectively. Second, misplaced gold pixels (MGP) assesses the fraction of misclassified gold pixels. It is computed by:

$$MGP = \left(1 - \frac{TP}{TP + FN}\right) \times 100 \tag{5}$$

MGP is useful for evaluating imbalanced data but does not account for false positives. Third, mean Dice score (MDS) evaluates the overlap between predicted and ground truth regions (27). Calculated for each phase, the Dice score is:

$$DS = \frac{2TP}{2TP + FN + FP}. \tag{6}$$

The MDS averages the Dice scores for the solid and pore phases.

2.7.2 Real Data

We used anisotropy-based metrics for real data where ground truth is unavailable, assuming isotropy in the structure. This is a valid assumption for hierarchical nanoporous gold (28; 29). The metrics include:

First, the two-point correlation function (TPCF) error ($e_{L_2}^{TPCF}$) assesses anisotropy by comparing the TPCF values in different directions. It is calculated as:

$$e_{L_2}^{TPCF} = \frac{1}{2}\left(\frac{2 \times \sqrt{\sum_{i=1}^{n}(f_i^x - f_i^z)^2}}{\sqrt{\sum_{i=1}^{n}(f_i^x)^2} + \sqrt{\sum_{i=1}^{n}(f_i^z)^2}} + \frac{2 \times \sqrt{\sum_{i=1}^{n}(f_i^y - f_i^z)^2}}{\sqrt{\sum_{i=1}^{n}(f_i^y)^2} + \sqrt{\sum_{i=1}^{n}(f_i^z)^2}}\right) \tag{7}$$

where f_i^x, f_i^y, and f_i^z are the discretized functional values in the x-, y-, and z-directions, respectively. A value of zero indicates perfect isotropy; higher values suggest anisotropy.

Second, lineal path function (LPF) error ($e_{L_2}^{LPF}$) is computed analogous to $e_{L_2}^{TPCF}$ but based on the LPF. This metric evaluates anisotropy based on local correlations between points in the same phase.

Third, diameter error ($e_{L_2}^{D}$) compares the predicted ligament diameters in different directions. It is calculated by:

$$e_{L_2}^{D} = \frac{1}{2}\left(\sqrt{\frac{(D_{xz} - D_{xy})^2}{D_{xy}^2}} + \sqrt{\frac{(D_{yz} - D_{xy})^2}{D_{xy}^2}}\right) \tag{8}$$

where D_{ij} represents the average diameter of the ligaments in the ij-plane. A value of zero indicates a geometrically isotropic structure.

All metrics are normalized to a range of [0, 1], where 0 denotes perfect accuracy, and 1 indicates a complete mismatch.

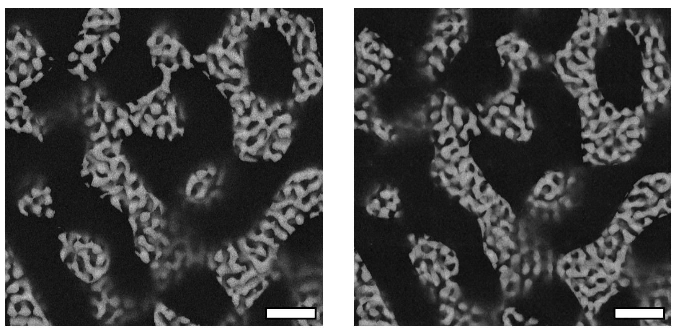
3 Results

3.1 Comparing different simulation techniques

To achieve optimal performance in downstream tasks for real datasets, it is crucial to train machine learning models using well-prepared data, including synthetic data. This study compares our novel machine learning-based simulation technique with the state-of-the-art Monte Carlo-based simulation method. We trained two machine learning models using two different datasets: sML-2kV (prepared using our ML-based method) and sMC-2kV (prepared using the Monte Carlo-based method). We then evaluated the models by calculating the overlapping regions of binary structures of r-2kV predicted by these segmentation models, using the mean Dice score (MDS) as described in Section 2.7.1. An MDS value of 0.83 indicates a high degree of overlap between the segmentations, demonstrating the effectiveness of our ML-based simulated data.

Table 2: Segmentation results of ML models trained on sML-1kV, sML-2kV and sML-4kV on respective test data prepared using our ML-based simulation method instead of Monte Carlo-based method

Data	MP $\downarrow$	MGP $\downarrow$	MDS $\uparrow$
sML-1kV	0.310	1.251	0.993
sML-2kV	0.539	2.415	0.987
sML-4kV	0.844	3.583	0.980



(A) (B)

Figure 4: A slice of a synthetic hierarchical nanoporous gold generated using **(A)** MCXray plugin and **(B)** our ML-based simulation method - scale bar: 200 nm

Additionally, Table 2 demonstrates that models trained on synthetic data prepared using our ML-based method exhibit comparable performance in terms of absolute errors. These results are directly comparable to those of ML models trained on synthetic data prepared using the Monte Carlo method (see Table 3 - only ML). The segmentation results are particularly promising for the real dataset r-2kV, with minimal anisotropy-based errors: $e_{L_2}^{TPCF}$ = 0.1526, $e_{L_2}^{LPF}$ = 0.0359, and $e_{L_2}^{D}$ = 0.0222. These values indicate the strong performance of segmentation models trained using our novel ML-based simulation method. Figure 6 in Appendix shows a visual comparison of segmentation performed using both ML models.

Another notable advantage is the significantly reduced preparation time for simulated data using our method, which takes seconds compared to days required for the computation-intensive Monte Carlo methods. Figure 4 provides a comparison of a single FIB tomography dataset slice simulated using both the Monte Carlo method and our ML-based method.

3.2 Comparing semantic segmentation after domain adaptation

We evaluated the impact of our domain adaptation technique on the segmentation performance of machine learning models. Specifically, we compared the segmentation performance on synthetic FIB tomography data and real HNPG datasets. The machine learning models were trained on synthetic training data both with and without domain adaptation, and the results were predicted on the same dataset.

3.2.1 Synthetic Data:

For synthetic datasets, which have ground truth values, we calculated absolute error-based metrics as described in Section 2.7.1. Table 3 presents the calculated MP, MGP, and MDS for synthetic test datasets (s-1kV, s-2kV, and s-4kV) predicted using machine learning models trained on sMC-1kV, sMC-2kV, and sMC-4kV data, and once with domain-adapted sMCDA-1kV, sMCDA-2kV, and sMCDA-4kV data. The comparable MP and MGP errors and MDS values for all corresponding test datasets indicate that the domain adaptation process does not significantly reduce performance. However, to demonstrate the clear advantage of domain adaptation for segmenting real HNPG datasets with substantial data distribution differences, we compared the performance of both machine learning models on the same real HNPG dataset in the next section. Figure 7 depicts a visual comparison of segmentation performed using different techniques.

Table 3: Comparison of segmentation results based on absolute errors with (ML+DA) and without (only ML) domain adaptation on test datasets

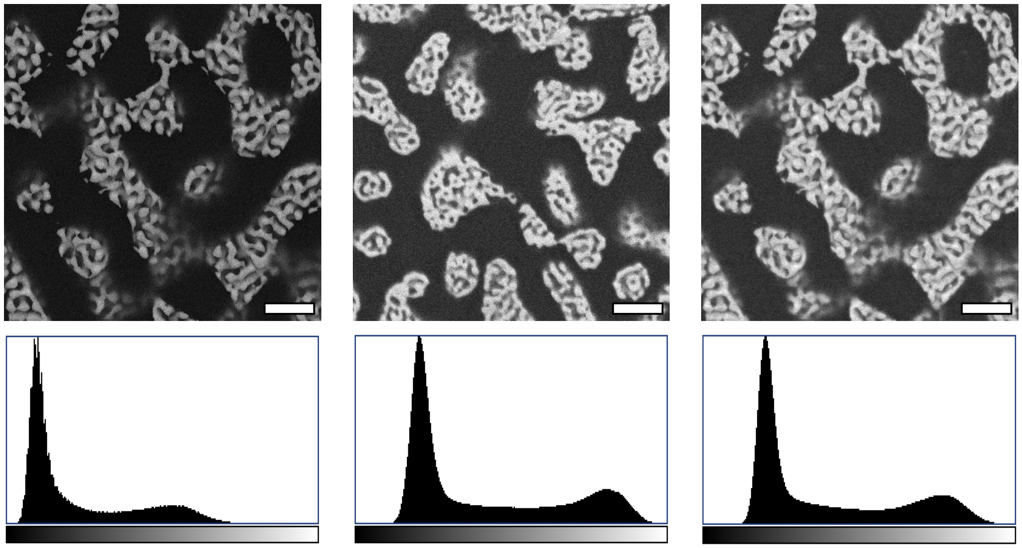
Method \ Measure		MP ↓	MGP ↓	MDS ↑
Original		0.000	0.000	1.000
ML + DA	s-1kV	0.401	1.654	0.991
	s-2kV	0.833	3.294	0.980
	s-4kV	1.809	7.428	0.958
only ML	s-1kV	0.245	0.963	0.994
	s-2kV	0.654	2.512	0.985
	s-4kV	1.407	6.466	0.967

3.2.2 Real Data:

Table 4: Comparison of segmentation results based on isotropy errors with (ML+DA) and without (only ML) domain adaptation on real datasets

Method \ Measure		$e_{L_2}^{TPCF} \downarrow$	$e_{L_2}^{LPF} \downarrow$	$e_{L_2}^{D} \downarrow$
ML + DA	r-1kV	0.132	0.047	0.029
	r-2kV	0.121	0.096	0.059
	r-4kV	0.140	0.038	0.017
only ML	r-1kV	0.145	0.059	0.030
	r-2kV	0.169	0.091	0.037
	r-4kV	0.271	0.181	0.138

It is crucial to assess the effect of training data prepared using domain adaptation on real HNPG datasets, as data distribution shifts are typically observed between synthetic and real FIB tomography datasets. Table 4 shows the superior performance of machine learning models trained using domain-adapted datasets. Notably, the very low errors for the r-4kV dataset for ML models with domain adaptation highlight the method's effectiveness, even for datasets with large artifacts, such as r-4kV. Since capturing microscopy images at lower voltages requires significant effort and expertise, our domain adaptation method offers a novel approach for microscopists to capture images at higher accelerating voltages while achieving similarly good segmentation results. Figure 5 provides a histogram comparison of sMC-2kV, sMCDA-2kV, and r-2kV, illustrating the qualitative improvement due to our domain adaptation technique. Figure 8 describes a visual comparison of segmentation of real HNPG microstructures performed using different techniques.



(A) (B) (C)

Figure 5: A slice of **(A)** a synthetic hierarchical nanoporous gold (HNPG) generated using MCXray plugin and **(B)** a real HNPG structure and **(C)** a synthetic HNPG structure after domain adaptation. The lower row represents histograms of the respective datasets - scale bar: 200 nm

4 Conclusion

In this study, we proposed a novel method for rapidly generating synthetic data by leveraging available physics knowledge, thereby bypassing the time-consuming Monte Carlo methods. This approach addresses the large data requirements for machine learning models, achieving performance comparable to traditional Monte Carlo based methods. Our method also lays the groundwork for more sophisticated techniques that can accommodate various materials and microscopy conditions. Furthermore, we demonstrated that reducing data shifts through domain adaptation techniques significantly improves reconstruction quality. This allows microscopy data acquired at higher voltages, which requires less effort, to be used for accurate 3D reconstruction when combined with our domain adaptation technique. Overall, our work provides a robust framework for enhancing 3D reconstruction accuracy in FIB tomography, making it a valuable tool for studying the physical properties of nanomaterials. However, it is important to note that generating synthetic data using our ML-based method has only been tested on the HNPG dataset. Future work should extend this approach to other materials, and consider replacing the current ANN with more sophisticated physics-based neural networks.

References

[1] 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.

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

[3] 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.

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

[5] Zhihang Song, Zimin He, Xingyu Li, Qiming Ma, Ruibo Ming, Zhiqi Mao, Huaxin Pei, Lihui Peng, Jianming Hu, Danya Yao, et al. Synthetic datasets for autonomous driving: A survey. *arXiv preprint arXiv:2304.12205*, 2023.

[6] 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.

[7] 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.

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

[9] Anja Butter, Sascha Diefenbacher, Gregor Kasieczka, Benjamin Nachman, and Tilman Plehn. Ganplifying event samples. *SciPost Physics*, 10(6):139, 2021.

[10] Y Alanazi, P Ambrozewicz, MP Kuchera, Y Li, T Liu, RE McClellan, W Melnitchouk, E Pritchard, M Robertson, N Sato, et al. Ai-based monte carlo event generator for electron-proton scattering. *arXiv preprint arXiv:2008.03151*, 2020.

[11] Sungmin Hong, Razvan Marinescu, Adrian V Dalca, Anna K Bonkhoff, Martin Bretzner, Natalia S Rost, and Polina Golland. 3d-stylegan: A style-based generative adversarial network for generative modeling of three-dimensional medical images. In *Deep Generative Models, and Data Augmentation, Labelling, and Imperfections: First Workshop, DGM4MICCAI 2021, and First Workshop, DALI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 1*, pages 24–34. Springer, 2021.

[12] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.

[13] Abid Khan, Chia-Hao Lee, Pinshane Y Huang, and Bryan K Clark. Leveraging generative adversarial networks to create realistic scanning transmission electron microscopy images. *npj Computational Materials*, 9(1):85, 2023.

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

[15] Liangchen Song, Yonghao Xu, Lefei Zhang, Bo Du, Qian Zhang, and Xinggang Wang. Learning from synthetic images via active pseudo-labeling. *IEEE Transactions on Image Processing*, 29:6452–6465, 2020.

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

[17] 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.

[18] 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.

[19] Trushal Sardhara, Alexander Shkurmanov, Yong Li, Lukas Riedel, Shan Shi, Christian J Cyron, Roland C Aydin, and Martin Ritter. Enhancing 3d reconstruction accuracy of fib tomography data using multi-voltage images and multimodal machine learning. *Nanomanufacturing and Metrology*, 7(1):4, 2024.

[20] Thermo Fisher Scientific Inc. Auto slice and view 4.0 [computer software]. 2017. Version: 4.1.0.1196.

[21] K Lepinay and F Lorut. Three-dimensional semiconductor device investigation using focused ion beam and scanning electron microscopy imaging (FIB/SEM tomography). *Microscopy and Microanalysis*, 19(1):85–92, 2013.

[22] 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.

[23] 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*, 5(2):112–118, 2022.

[24] 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.

[25] 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.

[26] 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.

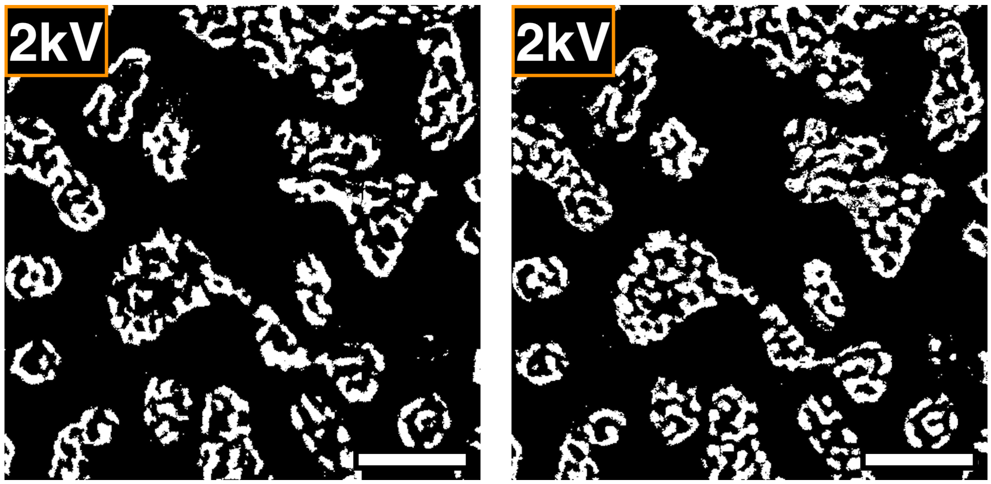
[27] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016.

[28] 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.

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

A Appendix

A.1 Visual segmentation results of comparing different simulation techniques

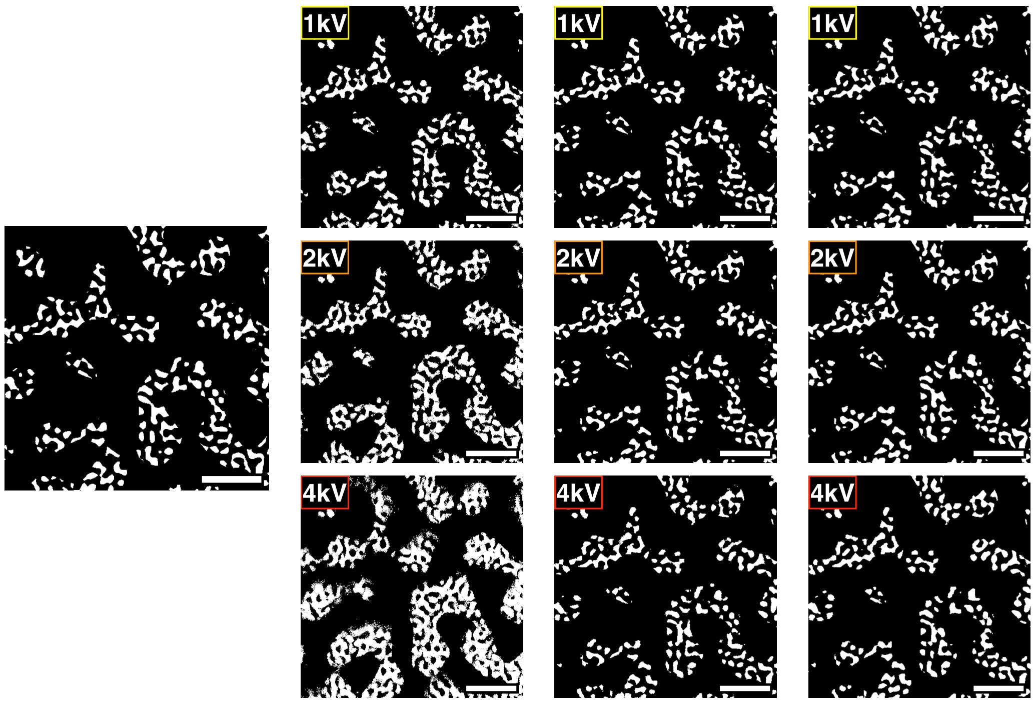


(A) (B)

Figure 6: Slice of a real HNPG microstructure segmented using the ML model trained on the synthetic data prepared using **(A)** MCXray plugin and **(B)** our ML-based simulation method. Note: x kV represents the original dataset imaged at an accelerating voltage x kV - scale bar: 300 nm

A.2 Visual segmentation results of comparing semantic segmentation after domain adaptation

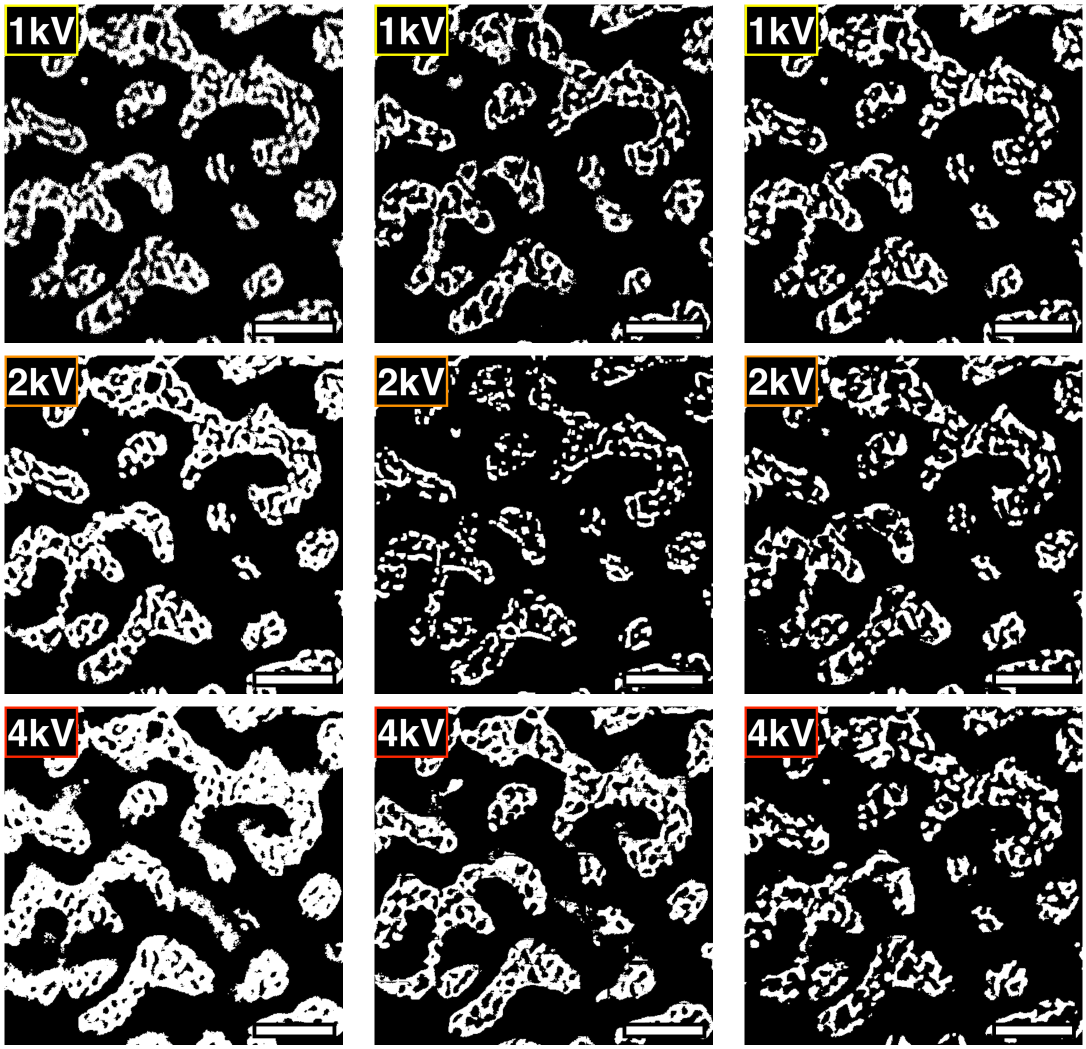
A.2.1 Synthetic data



(A) (B) (C) (D)

Figure 7: **(A)** Slice of a synthetic microstructure (ground truth) and segmentation results of Monte Carlo-simulated BSE images using **(B)** k-means clustering, **(C)** ML model trained on data without domain adaptation, and **(D)** ML model trained on data with domain adaptation. Note: *x* kV represents the original dataset imaged at an accelerating voltage *x* kV - scale bar: 300 nm

A.2.2 Real data



(A) (B) (C)

Figure 8: Slice of a real HNPG microstructure segmented using **(A)** k-means clustering, **(B)** ML model trained on data without domain adaptation, and **(C)** ML model trained on data with domain adaptation. Note: x kV represents the original dataset imaged at an accelerating voltage x kV - scale bar: 300 nm