PAPER • OPEN ACCESS

Generative adversarial networks for creating realistic training data for machine learning-based segmentation of FIB tomography data

To cite this article: Trushal Sardhara *et al* 2025 *Mach. Learn.: Sci. Technol.* **6** 025023

IOP Publishing Mach. Learn.: Sci. Technol. 6 (2025) 025023 https://doi.org/10.1088/2632-2153/adc870

MACHINE LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
23 January 2025

REVISED
14 March 2025

ACCEPTED FOR PUBLICATION
1 April 2025

PUBLISHED
22 April 2025

PAPER

Generative adversarial networks for creating realistic training data for machine learning-based segmentation of FIB tomography data

Trushal Sardhara[1], **Christian J Cyron**[1,2], **Martin Ritter**[3] **and Roland C Aydin**[1,2,*]

[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] Electron Microscopy Unit, Hamburg University of Technology, Hamburg, Germany
[*] Author to whom any correspondence should be addressed.

E-mail: roland.aydin@tuhh.de

Keywords: domain adaptation, fast simulation, synthetic data, FIB-SEM tomography, 3D reconstruction

Abstract

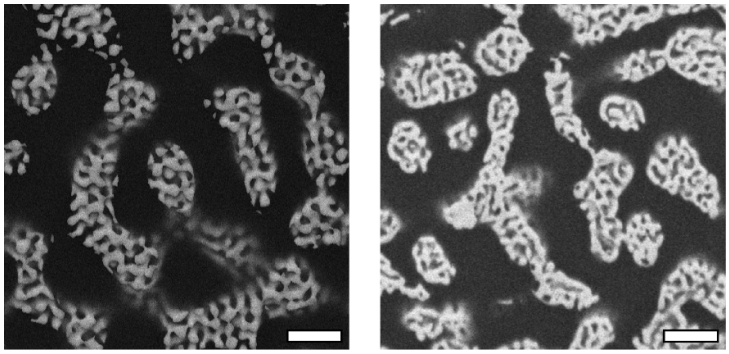
Accurate 3D reconstruction of the structure of nanomaterials is essential for studying their physical properties. Focused ion beam (FIB) tomography is a preferred method for creating 3D image stacks of micrometer-sized material volumes at nanometer resolution. To achieve valid 3D reconstructions from FIB tomography data, semantic segmentation of these images using machine learning-based methods is often beneficial. However, supervised machine learning requires a large amount of training data and ground truth, which is challenging because FIB tomography is a destructive technique. This motivates the use of synthetic training data generated with Monte Carlo simulations of the FIB tomography process. However, these simulations are computationally expensive, and the resulting synthetic imaging data still differs from real FIB tomography data in terms of the statistical distribution of various features. In this study, we propose a novel approach to overcome both problems, i.e. requiring high computation time and difference in data distribution, using generative adversarial networks.

1. Introduction

Nanoporous materials have significant potential in fields like materials science and biochemistry due to their unique properties [1]. To understand these properties, accurate 3D reconstruction of their structure is often required. When studying nanoporous materials, such as hierarchical nanoporous gold (HNPG) [2], 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. A 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 (z) resolution of 10 nm or less. This is achieved by consecutively removing material and imaging the newly exposed cross-sections [3]. However, since FIB tomography is inherently destructive, the material removed during the process cannot be re-imaged or analyzed using other techniques. This irreversible nature makes it challenging to validate structural features with alternative methods after FIB tomography.

Additionally, these high-resolution image stacks often contain artifacts, such as the *shine-through effect* and intensity ambiguities [4]. The *shine-through effect* occurs when structures from deeper layers become visible in the current milling plane, introducing extra information to the images. Ambiguities occur when their effects result in equal intensities despite completely different compositions in the z-direction [5]. 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 [6], combining FIB tomography with machine learning can lead to accurate 3D



(A) (B)

Figure 1. Comparison of hierarchical nanoporous gold (HNPG) structures: (A) synthetic HNPG generated through leveled wave method of computer simulations and subsequently using the MCXray plugin, and (B) real HNPG obtained from FIB tomography (scale bar: 200 nm). Notable differences are observed between the synthetic and real imaging data, arising from subtle discrepancies between the simulated imaging process and actual experimental conditions.

reconstructions of nanomaterials. In that study, we 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 EM, where acquiring real data is difficult and expensive. The sample is destroyed in FIB tomography, 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 [7, 8]. In EM, [9] generated simulated images of basic geometries for training purposes.

However, generating synthetic EM data presents challenges. It is time-consuming and resource-intensive, and creating realistic synthetic data requires comprehensive physical knowledge. The best synthetic data in EM are generated using Monte Carlo simulations, as suggested by [10]. However, simulating 512 slices of 512×512 pixels using the Monte Carlo plugin, MCXray, in Dragonfly software [11] can take approximately seven 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, [12] demonstrated the use of GANs in particle physics, while [13] suggested a GAN-based approach for simulating electron–proton scattering events. Additionally, [14] proposed a 3D GAN based on StyleGAN2 [15] to generate realistic magnetic resonance images. In EM, [16] 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 i.e. a relation of backscattered coefficient with atomic number, accelerating voltage, and penetration depth to accurately simulate data using machine learning.

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 [17, 18].

In this study, we are trying to 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 incorporating 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 HNPG by applying domain adaptation (DA) 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 DA to synthetic data before training, we achieve an average of 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 the following methods.

2.1.1. Synthetic samples

We began by generating binarized structures using the leveled wave method (LWM) [19], as described in [2]. 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 [11] to generate *realistic* synthetic FIB tomography images, which incorporate relevant electron microscope physics [10]. Using the same initial structure, we created three synthetic datasets with different voltages (1 kV, 2 kV, and 4 kV). 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 DA, we also created domain-adapted versions of the Monte Carlo datasets (sMCDA-1kV, sMCDA-2kV, and sMCDA-4kV). All datasets were split into training (80%) and test data (20%).

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 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 [2]. To enhance SEM imaging for solid-pore phase differentiation, the sample was infiltrated with epoxy resin [20]. Following the approach in [21], multi-voltage FIB tomography was conducted using a Dual Beam FEI HeliosNanoLab G3 system with ASV4 software for automated control [22], which also monitored milling progress and compensated for drift [23]. Imaging parameters included a 3072 $\times$ 2048 pixel resolution, a 10 μm horizontal field of view (3.26 nm pixel size), and a 30 μs dwell time for low-noise imaging. Backscattered electrons were detected using a through-the-lens detector.

To optimize HNPG tomography, two fiducial markers with intersecting trenches and a ruler system were used for drift compensation and precise slice thickness determination [24, 25]. Each slice was imaged at 1 kV, 2 kV, and 4 kV with a constant beam current of 50 pA as described in [21], generating three FIB-SEM datasets: r-1kV, r-2kV, and r-4kV. These datasets will be used to assess machine learning model performance on real microstructures.

2.2. Machine learning architecture

We employed an encoder-decoder model based on cycle-consistent adversarial networks (CycleGAN) [26]. 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 [27] with residual connections, and for the decoder, we adopted the architecture proposed in [26].

The cycle loss concept was key in addressing two primary challenges in this study: generating synthetic data and minimizing differences in data distribution. The cycle loss was chosen because we wanted to train our models on unpaired image datasets, as we do not have ground-truth datasets for simulation and DA problems.

2.3. Generating synthetic data using machine learning

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

In EM, the relationship between the backscattered electron coefficient (η), accelerating voltage (E), and atomic number (Z) is established (equation (1)). This relation $\eta = f(E, Z)$ was obtained analytically [28]. However, in this study, we try to find a direct relation between η and penetration depth (D). This

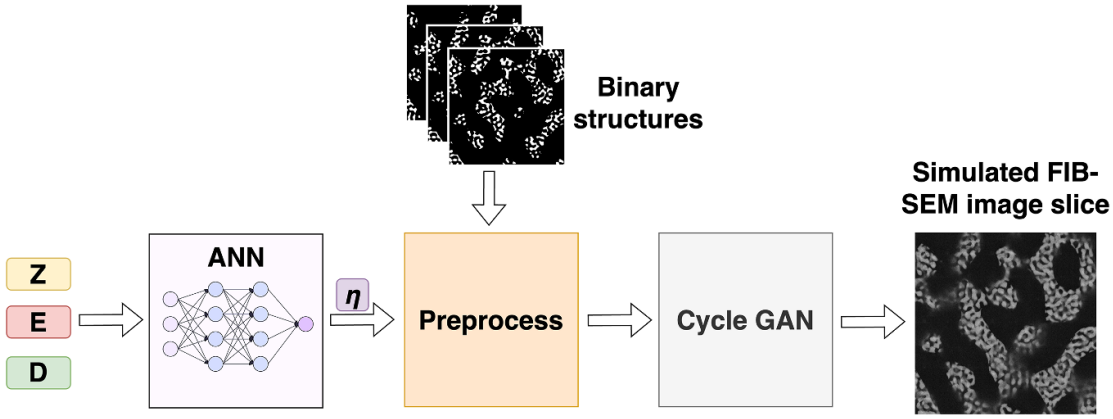


Figure 2. Block diagram of the experimental setup for generating synthetic images using machine learning based on physical principles. In the first step, an artificial neural network (ANN) is trained to predict the backscattered coefficient (η) using inputs such as atomic number (Z), accelerating voltage (E), and penetration depth (D). The second step involves training a CycleGAN to simulate synthetic FIB-SEM image slices using preprocessed inputs derived from η and a binary LWM structure.

relationship is essential for constructing weighted structures that integrate information from the planes beneath the currently sliced plane. By leveraging these structures, the simulation process becomes more efficient, as it enables problem-solving through image translation techniques. This approach preserves the original structural details and physical relationships while eliminating the need for complex 3D generative methods, which are computationally expensive and challenging to train with limited data. 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 LWM data before feeding it into the CycleGAN model. Predictions of backscattered electron coefficients for Au at 2 kV against respective simulated values are shown in figure C1(B). 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 grayscale 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 DA, 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 DA, 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 DA, we conducted semantic segmentation on all datasets following the approach outlined in [6]. We trained three different models to evaluate performance: one using domain-adapted data, another using Monte Carlo simulation data, and a third using data generated with our proposed method. This comparative

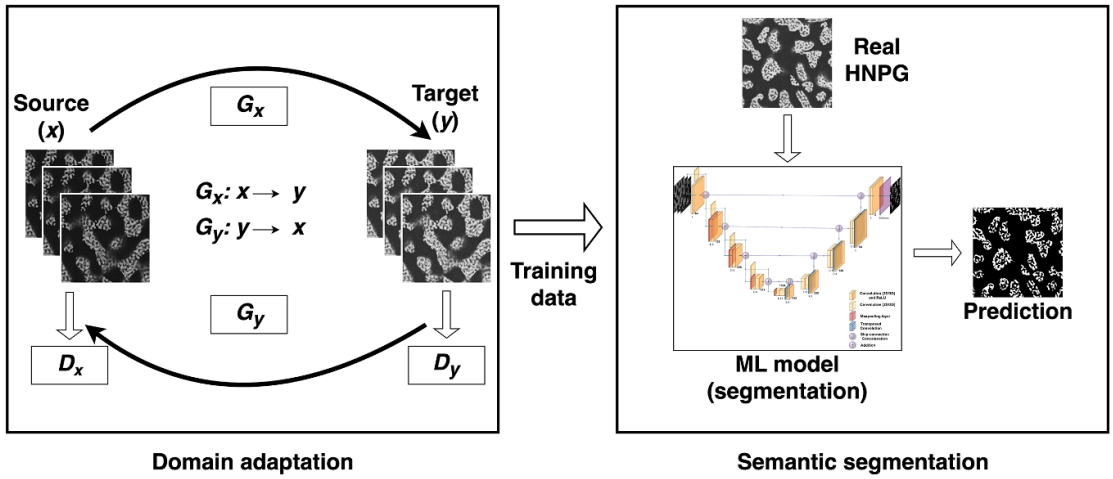


Figure 3. Block diagram of the experimental setup designed to reduce data distribution differences between synthetic and real HNPG images. The first step, domain adaptation, employs CNN-based generators (G_x and G_y) and discriminators (D_x and D_y) for the source domain (x) and target domain (y), respectively. The second step involves performing a downstream task—semantic segmentation—using the domain-adapted synthetic FIB-SEM image training data.

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)		

analysis allowed us to measure the improvement in segmentation accuracy due to DA 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 × 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 [26], 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.

We adopted a structured training approach for semantic segmentation, presenting data as individual 2D slices, 3D volumetric stacks, or 2D slices combined with neighboring slices in smaller patches (64 × 64). The models were optimized using Dice loss in conjunction with the Adam optimizer, with an initial learning rate of 0.0001, which was 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. It should be noted that, before finalizing the training parameters, we performed hyperparameter tuning on a small subset of the dataset, optimizing key parameters such as learning rate, batch size, and network architecture. Training plots for all GANs and the ANN model are provided in the appendix C.

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 [6].

2.7.1. Synthetic data

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

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

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

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

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

MGP is useful for evaluating imbalanced data but does not account for false positives. Third, the mean Dice score (MDS) evaluates the overlap between predicted and ground truth regions [29]. Calculated for each phase, the Dice score is:

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

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 HNPG [30, 31]. The metrics include:

First, the two-point correlation function (TPCF) error ($e_{L_2}^{\mathrm{TPCF}}$) assesses anisotropy by comparing the TPCF values in different directions. The TPCF quantifies how likely two points separated by a given distance belong to the same material phase, reflecting spatial structure and phase distribution. In an isotropic structure, this probability should be direction-independent. It is calculated as:

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

where $f^x{}_i$, $f^y{}_i$, and $f^z{}_i$ 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 ($e_{L_2}^{\mathrm{LPF}}$) is computed analogously to $e_{L_2}^{\mathrm{TPCF}}$ but based on the LPF, which measures the probability that an entire line segment lies within a single phase. Unlike TPCF, which captures phase distribution at a given length scale, the LPF is more sensitive to phase connectivity and continuity.

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

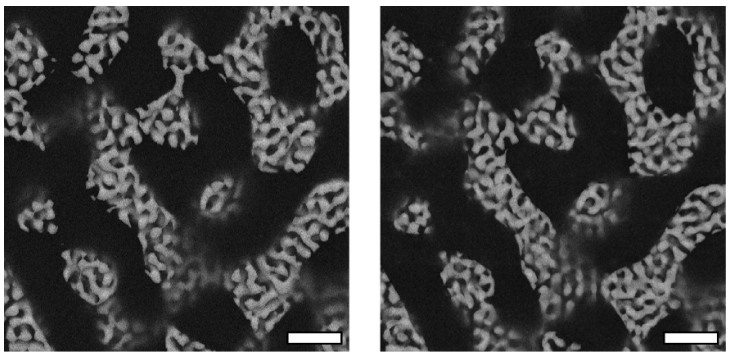
$$e_{L_2}^{D} = \frac{1}{2}\left(\sqrt{\frac{\left(D_{xz} - D_{xy}\right)^2}{D_{xy}^2}} + \sqrt{\frac{\left(D_{yz} - D_{xy}\right)^2}{D_{xy}^2}}\right) \tag{8}$$

where D_{ij} represents the average diameter of the ligaments in the ij-plane. This metric directly evaluates the geometric anisotropy of the structure, as variations in ligament thickness across different planes indicate directional dependencies in material formation. A value of zero suggests 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.

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 (%) ↓	MGP (%) ↓	MDS ↑
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. Slices of synthetic imaging data for hierarchical nanoporous gold generated using (A) the MCXray plugin and (B) our new machine learning-based simulation method leveraging generative adversarial networks (scale bar: 200 nm). The ML-based simulation method produces synthetic FIB-tomography data that closely resembles the computationally expensive yet accurate Monte Carlo-based methods.

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 MDS as described in section 2.7.1. An MDS value of 0.83 (1 represents full overlap) indicates a high degree of overlap between the segmentations, demonstrating the effectiveness of our ML-based simulated data.

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: $e_{L_2}^{\mathrm{TPCF}} = 0.1526$, $e_{L_2}^{\mathrm{LPF}} = 0.0359$, 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 A1 in appendix visually compares 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 the 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 DA

We evaluated the impact of our DA 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 DA, and the results were predicted on the test dataset withheld from the same dataset.

Table 3. Comparison of machine learning-based segmentation results based on absolute errors with (ML+DA) and without (only ML) domain adaptation on multi-voltage synthetic test datasets.

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

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

Method \ Measure		$e_{L_2}^{\mathrm{TPCF}}$ ↓	$e_{L_2}^{\mathrm{LPF}}$ ↓	$e_{L_2}^{D}$ ↓
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

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 (see only ML), and once with domain-adapted sMCDA-1kV, sMCDA-2kV, and sMCDA-4kV data (see ML + DA). Both methods achieve low absolute MP and MGP errors, though predictions using models with DA (ML + DA) show relative increases. For sMCDA-1kV, MP rises by 60% (from 0.245% to 0.401%) and MGP by 72% (from 0.963% to 1.654%) with DA. Similar trends are observed for sMCDA-2kV and sMCDA-4kV. MDS values remain consistently high (>0.95), indicating that segmentation quality is well preserved. While DA (ML + DA) yields slightly lower MDS than only ML, the differences are minimal. These results suggest that while DA introduces small relative variations in synthetic datasets, segmentation accuracy remains high. However, to demonstrate the clear advantage of DA 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 B1 depicts a visual comparison of segmentation performed using different techniques.

3.2.2. Real data

It is crucial to assess the effect of training data prepared using DA 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 DA (ML+DA) 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 DA 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 DA technique. Figure B2 describes a visual comparison of the segmentation of real HNPG microstructures performed using different techniques.

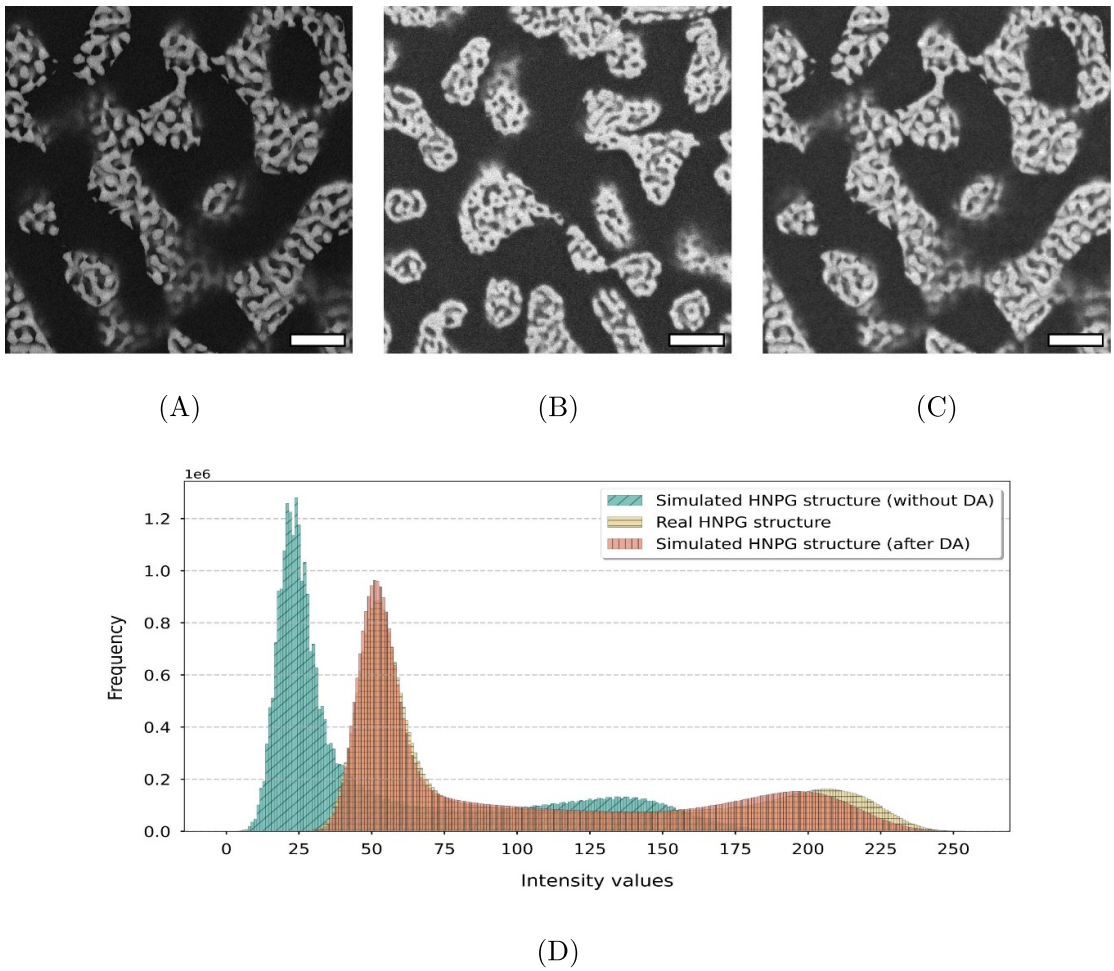


Figure 5. Slices of hierarchical nanoporous gold (HNPG) structures: (A) synthetic imaging data generated through computer simulations using the MCXray plugin, (B) real HNPG structure, and (C) synthetic imaging data after domain adaptation using our new approach leveraging generative adversarial networks. (D) Histograms corresponding to each dataset (scale bar: 200 nm). While Monte Carlo simulations alone cannot produce fully realistic synthetic imaging data, our approach illustrated in (C) achieves a much closer resemblance to real data.

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 DA techniques significantly improves reconstruction quality for real FIB tomography data. This allows microscopy data acquired at higher voltages, which requires less effort, to be used for accurate 3D reconstruction when combined with our DA 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.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.15480/882.14344.

Author contributions

TS, RA, CC, and MR contributed to the conception and design of the study. TS wrote the first draft of the manuscript. All authors contributed to the manuscript revision, and read, and approved the submitted version.

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - SFB 986 - Project number 19 234 6071.

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.

Appendix A. Visual segmentation results of comparing different simulation techniques

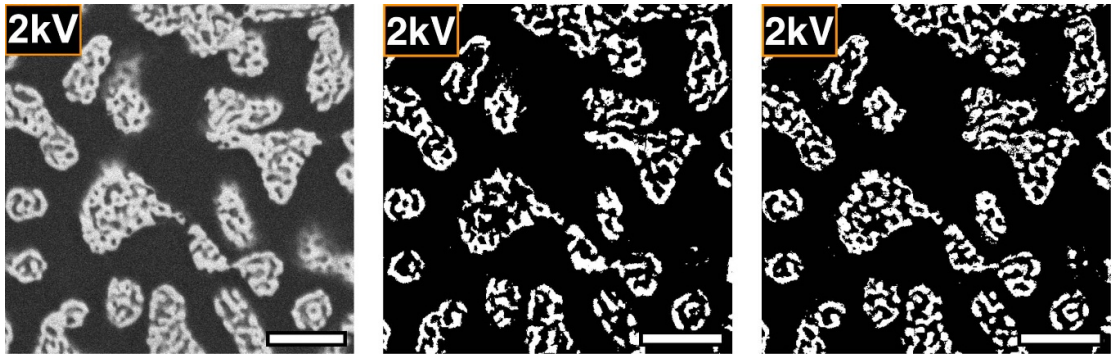


Figure A1. (A) Slice of a real HNPG microstructure and its segmentation using the ML model trained on synthetic data generated via (B) the MCXray plugin and (C) our ML-based simulation method. Note: *x* kV denotes the original dataset imaged at an accelerating voltage of *x* kV—scale bar: 300 nm.

Appendix B. Visual segmentation results of comparing semantic segmentation after domain adaptation

B.1. Synthetic data

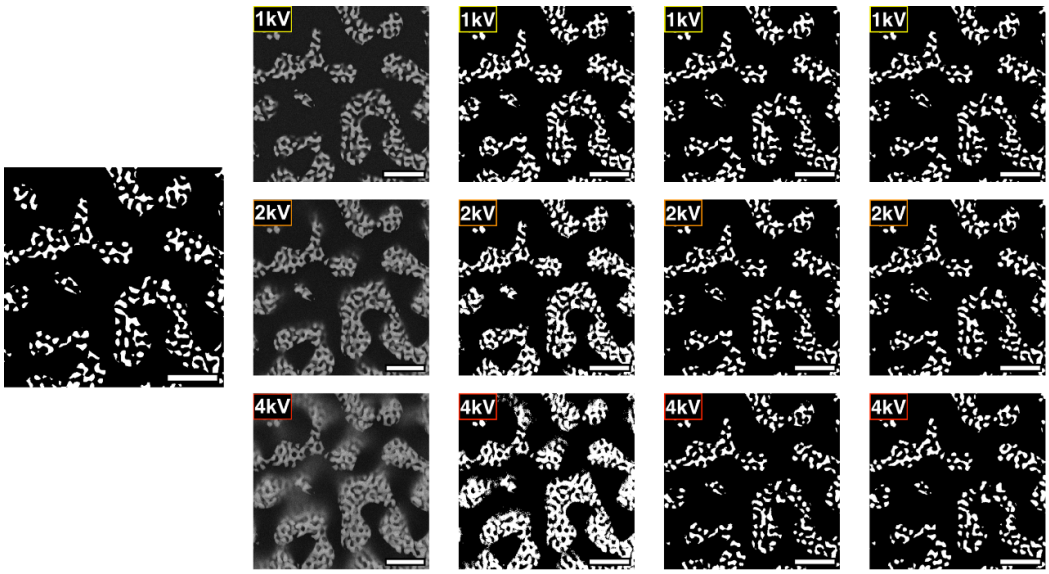


Figure B1. (A) Slice of a synthetic binary microstructure (ground truth), (B) Monte Carlo-simulated BSE images and their segmentation results obtained using (C) *k*-means clustering, (D) an ML model trained on data without domain adaptation, and (E) an ML model trained on data with domain adaptation.
Note: x kV indicates the original dataset imaged at an accelerating voltage of *x* kV—scale bar: 300 nm.

B.2. Real data

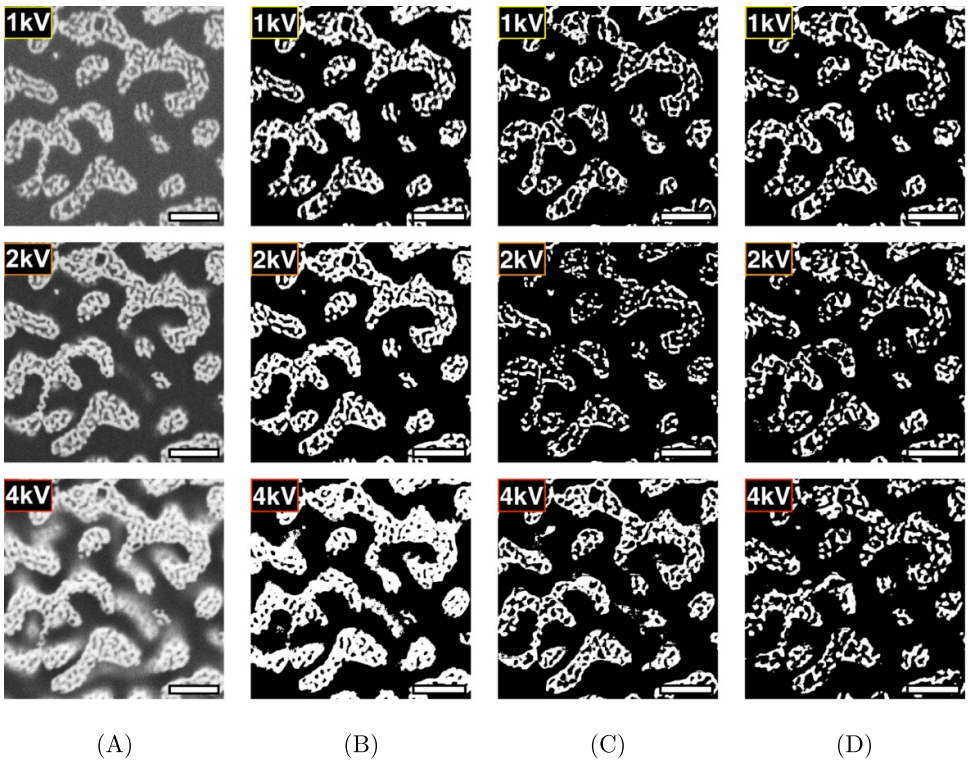


Figure B2. (A) Slice of a real HNPG microstructure and its segmentations using (B) *k*-means clustering, (C) ML model trained on data without domain adaptation, and (D) ML model trained on data with domain adaptation.
Note: x kV represents the original dataset imaged at an accelerating voltage *x* kV—scale bar: 300 nm.

Appendix C. Training plots

Figures C1–C3 present the training plots for ANN, CycleGANs for ML-based simulation, and CycleGANs for domain adaptation. The ANN optimization process took approximately one minute on a CPU. In contrast, training a single epoch of our CycleGAN models required about 25 min, completing a total of 25 088 steps per epoch. On the same hardware, inference for both domain adaptation and ML-based simulations was significantly faster, taking approximately 20 s. Notably, our results indicate that domain adaptation models achieve optimal performance within just one epoch for our dataset.

C.1. Backscattered electron coefficients prediction using ANN

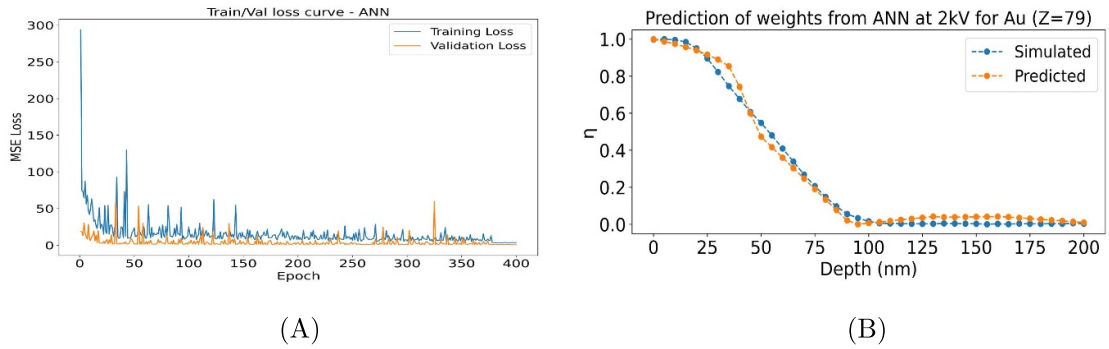


Figure C1. (A) Training and validation loss curves of ANN for predicting backscattered electron coefficients, and (B) prediction of backscattered electron coefficients for Au ($Z = 79$) at 2 kV.

C.2. ML-based simulation networks

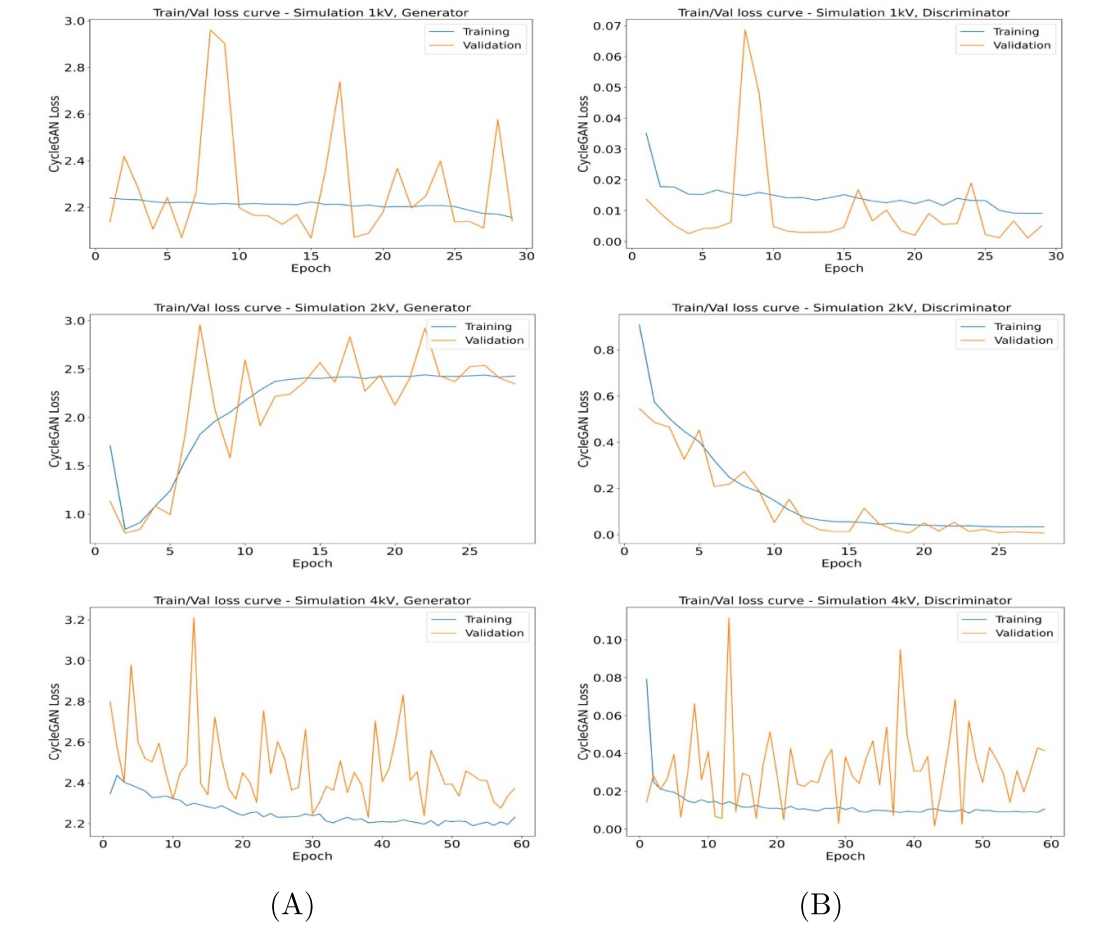


Figure C2. Training and validation loss curves of GANs for simulating FIB-SEM tomography data at accelerating voltages of 1 kV, 2 kV, and 4 kV. (A) Loss curves for the generators, and (B) loss curves for the discriminators.

C.3. Domain adaptation networks

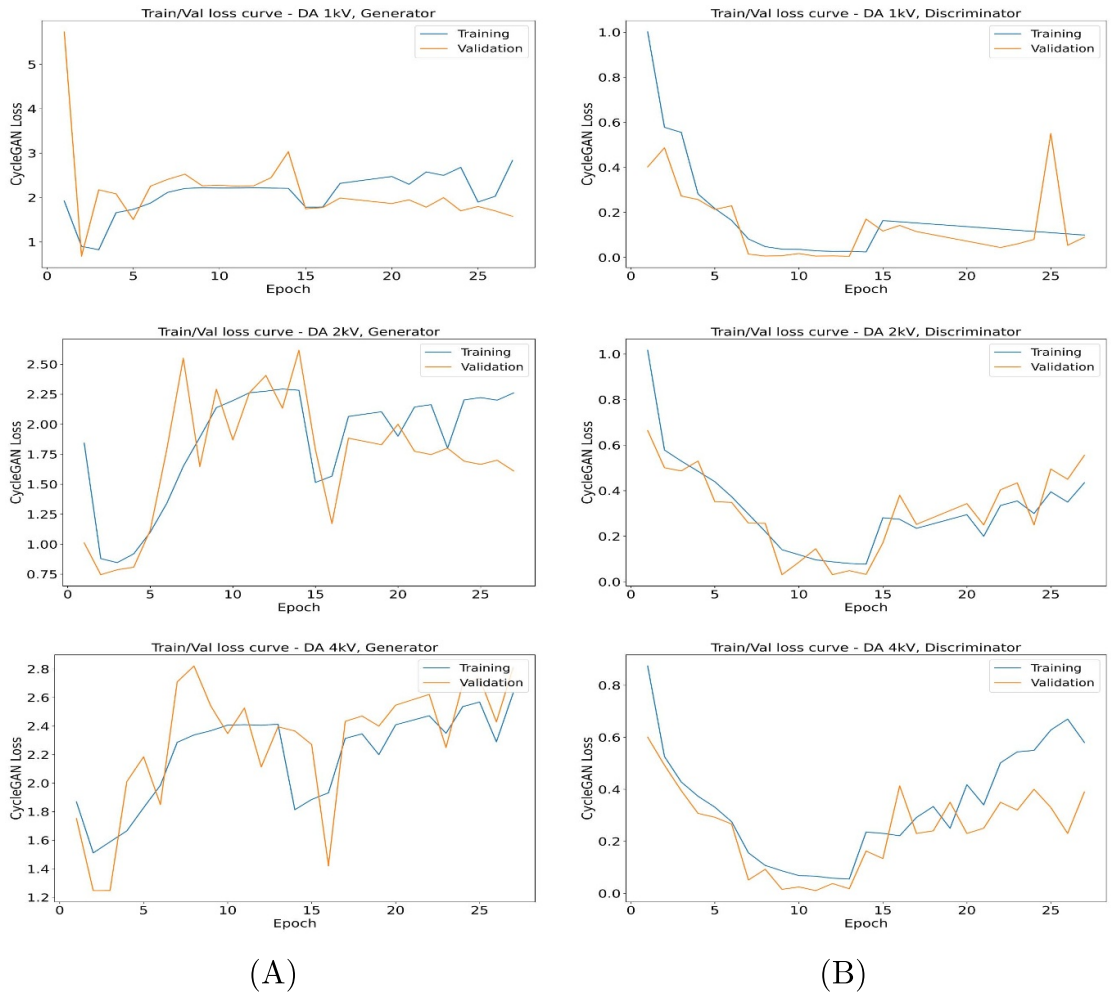


Figure C3. Training and validation loss curves of GANs for domain adaptation for FIB-SEM tomography data at accelerating voltages of 1 kV, 2 kV, and 4 kV. (A) Loss curves for the generators, and (B) loss curves for the discriminators.

ORCID iDs

Trushal Sardhara https://orcid.org/0000-0001-9915-3039
Christian J Cyron https://orcid.org/0000-0001-8264-0885

References

[1] Polarz S and Smarsly B 2002 Nanoporous materials *J. Nanosci. Nanotechnol.* **2** 581–612
[2] Shi S, Li Y, Ngo-Dinh B-N, Markmann J and Weissmüller J 2021 Scaling behavior of stiffness and strength of hierarchical network nanomaterials *Science* **371** 1026–33
[3] 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** 2959–64
[4] Rogge F and Ritter M 2019 Cluster analysis for FIB tomography of nanoporous materials *Conf.: IMC19 Sydney*
[5] Sardhara T, Shkurmanov A, Aydin R, Cyron C J and Ritter M 2023 Towards an accurate 3D reconstruction of nano-porous structures using FIB tomography and Monte Carlo simulations with machine learning
[6] Sardhara T, Aydin R C, Li Y, Piché N, Gauvin R, Cyron C J and Ritter M 2022 Training deep neural networks to reconstruct nanoporous structures from FIB tomography images using synthetic training data *Frontiers in Materials* **9** 837006
[7] Nikolenko S I 2021 *Synthetic Data for Deep Learning* vol 174 (Springer)
[8] Song Z, He Z, Li X, Ma Q, Ming R, Mao Z, Pei H, Peng L, Hu J, Yao D 2023 Synthetic datasets for autonomous driving: a survey (arXiv:2304.12205)
[9] 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
[10] 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** 488–9
[11] 2018 Canada Object Research Systems (ORS) Inc, Montreal. Dragonfly 3.6 [computer software]
[12] Butter A, Diefenbacher S, Kasieczka G, Nachman B and Plehn T 2021 GANplifying event samples *SciPost Phys.* **10** 139

[13] Alanazi Y, Ambrozewicz P, Kuchera M P, Li Y, Liu T, McClellan R E, Melnitchouk W, Pritchard E, Robertson M, Sato N 2020 AI-based Monte Carlo event generator for electron-proton scattering (arXiv:2008.03151)
[14] Hong S, Marinescu R, Dalca A V, Bonkhoff A K, Bretzner M, Rost N S and Golland P 2021 3D-StyleGAN: a style-based generative adversarial network for generative modeling of three-dimensional medical images *Deep Generative Models and Data Augmentation, Labelling and Imperfections: First Workshop, DGM4MICCAI 2021 and First Workshop, DALI 2021, Held in Conjunction With MICCAI 2021Proc. (Strasbourg, France, 1 October 2021)* vol 1 (Springer) pp 24–34
[15] Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J and Aila T 2020 Analyzing and improving the image quality of StyleGAN *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition* pp 8110–9
[16] Khan A, Lee C-H, Huang P Y and Clark B K 2023 Leveraging generative adversarial networks to create realistic scanning transmission electron microscopy images *npj Comput. Mater.* **9** 85
[17] Kouw W M and Loog M 2018 M. An introduction to domain adaptation and transfer learning (arXiv:1812.11806)
[18] Song L, Xu Y, Zhang L, Du B, Zhang Q and Wang X 2020 Learning from synthetic images via active pseudo-labeling *IEEE Trans. Image Process.* **29** 6452–65
[19] Soyarslan C, Bargmann S, Pradas M and Weissmüller J 2018 3D stochastic bicontinuous microstructures: generation, topology and elasticity *Acta Mater.* **149** 326–40
[20] Peña B, Owen G R, Dettelbach K E and Berlinguette C P 2018 Spin-coated epoxy resin embedding technique enables facile SEM/FIB thickness determination of porous metal oxide ultra-thin films *J. Microsc.* **270** 302–8
[21] Sardhara T, Shkurmanov A, Li Y, Riedel L, Shi S, Cyron C J, Aydin R C and Ritter M 2024 Enhancing 3D reconstruction accuracy of FIB tomography data using multi-voltage images and multimodal machine learning *Nanomanuf. Metrol.* **7** 4
[22] Thermo Fisher Scientific Inc 2017 Auto slice and view 4.0 [computer software] Version: 4.1.0.1196
[23] Lepinay K and Lorut F 2013 Three-dimensional semiconductor device investigation using focused ion beam and scanning electron microscopy imaging (FIB/SEM tomography) *Microsc. Microanal.* **19** 85–92
[24] Jones H G, Mingard K P and Cox D C 2014 Investigation of slice thickness and shape milled by a focused ion beam for three-dimensional reconstruction of microstructures *Ultramicroscopy* **139** 20–28
[25] Shkurmanov A, Krekeler T and Ritter M 2022 Slice thickness optimization for the focused ion beam-scanning electron microscopy 3D tomography of hierarchical nanoporous gold *Nanomanuf. Metrol.* **5** 112–8
[26] Zhu J-Y, Park T, Isola P and Efros A A 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks *Proc. IEEE Int. Conf. on Computer Vision* pp 2223–32
[27] Ronneberger O, Fischer P and Brox T 2015 U-Net: convolutional networks for biomedical image segmentation *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th Int. Conf.Proc., Part III (Munich, Germany, 5–9 October 2015)* vol 18 (Springer) pp 234–41
[28] Hunger H-J and Küchler L 1979 Measurements of the electron backscattering coefficient for quantitative EPMA in the energy range of 4 to 40 keV *Phys. Status Solidi* a **56** K45–K48
[29] Milletari F, Navab N and Ahmadi S-A 2016 V-Net: fully convolutional neural networks for volumetric medical image segmentation *2016 4th Int. Conf. on 3D Vision (3DV)* (IEEE) pp 565–71
[30] Hu K, Ziehmer M, Wang K and Lilleodden E T 2016 Nanoporous gold: 3D structural analyses of representative volumes and their implications on scaling relations of mechanical behaviour *Phil. Mag.* **96** 3322–35
[31] Qi Z and Weissmuller J 2013 Hierarchical nested-network nanostructure by dealloying *ACS Nano* **7** 5948–54