

Neural Field-Based 3D Surface Reconstruction of Microstructures from Multi-Detector Signals in Scanning Electron Microscopy

Supplementary Material

1. Multi-View and Multi-Detector SEM Data Acquisition

All SEM data are acquired using a ZEISS Gemini 560 SEM system (Fig. 1) equipped with a field-emission gun and controlled through the ZEISS SmartSEM software. Microscale specimens are mounted on a 5-axis motorized stage and sequentially rotated under SmartSEM control to enable multi-view imaging. For each microstructure, the stage is tilted along two orthogonal axes in 5° increments from -45° to $+45^\circ$, yielding 37 distinct viewing directions. At each viewpoint, an SE image is captured using the Everhart-Thornley detector located at the side of the SEM chamber. In addition, a pneumatically retractable 4Q-BSE detector positioned beneath the objective lens captures four BSE images from its four outer quadrants. All images are recorded at a resolution of 1024×768 pixels. To ensure stable imaging and prevent mechanical interference during stage tilting, the accelerating voltage is set between 5–7 kV and the working distance is maintained between 7–12 mm. Before data acquisition, SEM parameters are optimized for each microstructure to improve image clarity and minimize charging artifacts. After optimization, all SEM parameters, including accelerating voltage, working distance, magnification, brightness, and contrast, are kept fixed throughout the entire imaging sequence.

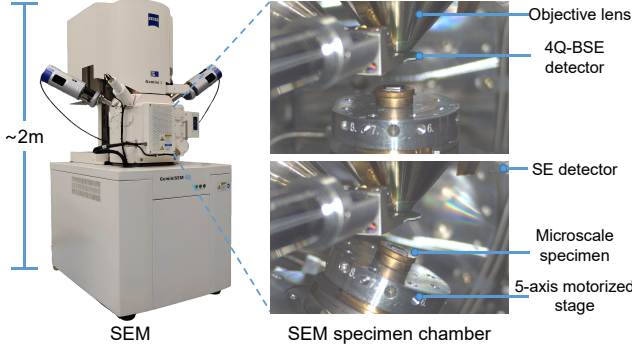


Figure 1. **Overview of the SEM imaging setup for multi-view and multi-detector scanning.** External photograph of the SEM system used in our experiments and internal chamber views with the motorized specimen stage in flat and tilted positions.

To improve the efficiency of multi-view and multi-detector data acquisition, we utilize the compucentric function in SmartSEM, enabling the SEM to maintain focus on the microstructure as the stage tilts or rotates. To enable this

function, the rotation center and its distance from the sample surface are determined in accordance with SmartSEM operational guidelines. With this information, the software automatically computes stage offsets after each stage motion, ensuring the microstructure remains centered in the field of view. Furthermore, we packaged all data acquisition processes into a unified automated macro script, utilizing the SmartSEM macro command interface to realize multi-detector imaging and stage movements. Therefore, operator intervention is minimal to occasional manual corrections for accumulated stage drift, resulting in a low manual workload for the entire data acquisition process.

2. Neural Field Implementation Details

The learnable parameters Θ of our neural field consist of both the multi-resolution hash table parameters and the MLP weights. At each training iteration, we randomly sample 256 rays, with 1024 points along each ray. Each 3D point is first encoded using multi-resolution hashing [9], a technique that significantly accelerates training and inference of neural fields. We employ a 16-level hash table with a 2-dimensional feature per level, producing a 32-dimensional concatenated feature vector. This hash-encoded feature is then concatenated with the raw 3D coordinate and fed into a shallow MLP consisting of a single hidden layer with 64 units and ReLU activation. The MLP maps the 3D coordinate to its corresponding SDF value. This neural field representation is built upon an open-source PyTorch [12] implementation [5] of NeUS [17]. The SDF is subsequently converted into a density value using an unbiased and occlusion-aware weighting function [17]. Finally, volume rendering is performed along each camera ray to compute surface intersection depths \hat{z} and normals \hat{n} . For loss weighting, we set $\lambda_1 = 0.5$ for TPL samples and 0.1 for pollen and SiC samples. The remaining weights λ_2 , λ_3 , and λ_4 are fixed to 0.1, 1.0, and 1.0 for all samples. The parameter α is set to 0.25 for TPL microstructures and 0.5 for pollen and SiC. Following prior findings on the validity range of the BSE signal model [11, 14], the BSE loss \mathcal{L}_{BSE} is applied to regions with incidence angle θ less than 60 degrees. All experiments are conducted on a workstation equipped with an AMD 5950X CPU, NVIDIA RTX 4090 GPU, and 128 GB RAM. After training, we discretize the space into a $512 \times 512 \times 512$ voxel grid and infer the SDF value at each vertex. These SDF values are then converted into meshes using the Marching Cubes algorithm [8] to produce 3D models shown in our experiments.

3. Sample Preparation

3.1. TPL Microstructure Fabrication

Microstructures were fabricated using the commercial photoresist IP-S (Nanoscribe GmbH), which was dispensed onto an indium tin oxide-coated glass substrate (25 mm × 25 mm × 0.7 mm, Nanoscribe GmbH). A commercial direct laser writing system (Photonic Professional GT2, Nanoscribe GmbH), equipped with a 780 nm femtosecond laser (80 MHz repetition rate, 80–100 fs pulse duration) and a 25× objective lens with a numerical aperture (NA) of 0.8, was employed for the 3D printing of microstructures. CAD models were processed in Describe v2.7 (Nanoscribe GmbH) to generate executable job files, with the slicing and hatching distances set to 500 nm and 400 nm, respectively. We set printing parameters with a laser power of 50 mW and a scanning speed of 50,000 $\mu\text{m/s}$. These job files were then transferred to NanoWrite v1.8 (Nanoscribe GmbH) to initiate the writing process. After printing, the microstructures were developed in propylene glycol methyl ether acetate (PGMEA) for 20 minutes, followed by rinsing in isopropyl alcohol (IPA) for 5 minutes at room temperature to remove unpolymerized photoresist. Prior to SEM imaging, the printed microstructures on the substrate were sputter-coated with a thin platinum layer using an ion sputter coater (GVC-2000, Gewei Instrument) at 15 mA for 60 seconds.

3.2. Peach Pollen and SiC Sample Preparation

The Okubo peach pollen was sourced from Dangshan County Orchard Farm (China). The micro-scale SiC powder was purchased from Beesley New Materials (Suzhou) Co., Ltd. (China). It was produced via a high-temperature reaction between silica and petroleum coke, followed by crushing and classification to obtain particles with an average diameter of approximately 45 μm . Both samples were prepared using the same procedure. They were mounted on 9.5 mm aluminum pin stubs using carbon conductive adhesive tabs and coated with gold using an E-1010 ion sputter coater (Hitachi) at a 15 mA current for 60 seconds, effectively preventing charging artifacts during SEM imaging.

4. Challenges in Acquiring Ground-Truth 3D Geometry for SEM Surface Reconstruction

Quantitative evaluation of SEM 3D reconstruction remains fundamentally challenging due to the difficulty of obtaining micro-nanoscale ground-truth surface geometry. Here, we provide a brief discussion of two primary factors.

4.1. Limitations of Micro-Nanoscale 3D Metrology Techniques

Although a variety of high-resolution 3D metrology techniques exist, including atomic force microscopy (AFM),

white-light interferometry (WLI), confocal microscopy, X-ray micro-CT, and Focused Ion Beam Scanning Electron Microscope (FIB-SEM) tomography, it remains difficult for these techniques to provide surface measurements that are both non-destructive and matched to the resolution and depth-of-field of SEM. Each technique has intrinsic limitations: AFM offers a limited field of view and suffers from tip-sample convolution [4]; WLI [7] and confocal microscopy [2] lack SEM-level lateral resolution and are incompatible with steep or high-aspect-ratio microstructures; micro-CT provides insufficient surface fidelity [13]; and FIB-SEM damages the sample surface during scanning [18]. Consequently, obtaining a reference 3D surface that faithfully matches SEM-observed geometry is highly challenging.

4.2. Why CAD Models Are Not Valid Ground Truth for TPL Microstructures

Although we have CAD models used for TPL fabrication, these models cannot serve as ground truth for evaluation. The reason is that the final printed geometry inevitably deviates from the design model due to the physics of the TPL process. As discussed in Supplementary Sec. 3.1, factors such as laser power, scanning trajectory, slicing/hatching distances, voxel elongation, and development conditions all strongly influence the polymerization process and introduce fabrication-dependent variations [16]. These deviations are significant in practice. For instance, the layered patterns observed on the forehead of the Lucy sample (Fig. 4 in the main paper) are absent from the CAD model. They originate from the layer-by-layer characteristics of the TPL printing process rather than the intended geometry. Consequently, using the CAD model as ground truth would incorrectly penalize reconstructions for accurately recovering real, physically printed surface features that do not appear in the ideal design model.

5. Construction of Simulated Dataset

We construct a simulated dataset for quantitative evaluation based on the CAD models used for TPL fabrication. To simulate the multi-view SEM imaging setup, cameras are placed in Blender v4.1 at the same tilt and rotation angles as in our real experiments. Surface blur and noise are applied to the CAD models using Blender modifiers to generate coarse geometric initializations. To approximate the appearance of real 4Q-BSE images, we simulate each component of the BSE formation process separately. Shadow-free BSE intensities are computed using our forward model $\mathcal{F}(\bar{n}; \bar{\Phi})$, where \bar{n} denotes the surface normals of the CAD geometry and $\bar{\Phi}$ is the set of BSE parameters estimated from NFH-SEM reconstructions of real TPL samples. A key element of BSE imaging is the soft shadowing caused by self-occlusion. Since BSE shadow formation resembles a ren-

dering configuration in which the four detector quadrants behave as area light sources and the electron beam serves as the view direction, we model shadows in Blender by placing quadrant-shaped area lights in fixed poses relative to the camera. The resulting shadow intensity maps $\bar{\psi}$ are rendered using the Blender Cycles engine. To match the noise characteristics of real BSE images, we estimate the pixel-wise standard deviation ($\sigma = 0.9142$ in our case) from a flat region of a TPL microstructure and use it to synthesize additive Gaussian noise $\mathcal{N}(0, \sigma^2)$. The final simulated BSE images are given by:

$$b' = \mathcal{F}(\bar{n}; \bar{\Phi}) - \bar{\psi} + \mathcal{N}(0, \sigma^2). \quad (1)$$

Finally, by combining the shadow-free intensity maps, the rendered soft shadow maps, and the synthesized noise, we obtain simulated 4Q-BSE images b' with high consistency with real-world images (Fig. 2).

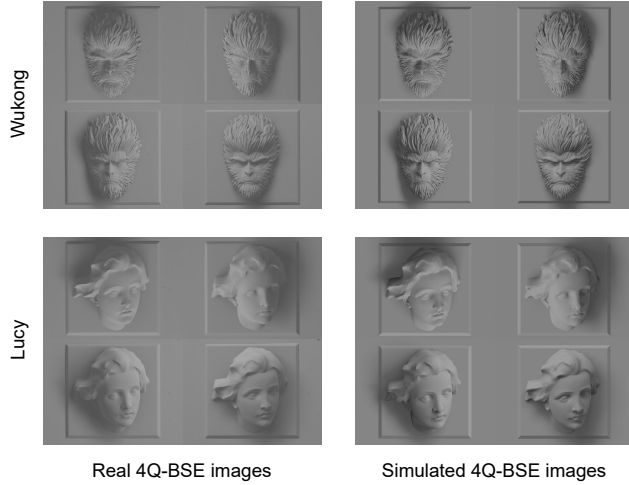


Figure 2. **Comparison between real and simulated 4Q-BSE images.** The simulated quadrant responses closely match the real detector signals, validating both the accuracy of our learned BSE forward model and the correctness of the BSE signal formulation used in our simulation pipeline.

6. Evaluation Metrics

We use the Chamfer distance and the average angular difference of surface normals between the reconstructed geometry and the ground truth to evaluate geometric accuracy. In addition to these standard metrics, we define two evaluation metrics for the BSE forward model and the estimated 4Q-BSE shadow maps produced by NFH-SEM.

We evaluate the accuracy of the BSE forward model $\mathcal{F}^{\hat{\Phi}}$ with learned parameters $\hat{\Phi}$, by comparing it to the ground-

truth model $\mathcal{F}^{\bar{\Phi}}$ used to generate the simulated dataset:

$$\mathcal{E}_{\text{BSE}} = \frac{1}{4T} \sum_i^{A,B,C,D} \sum_q^T \left| \mathcal{F}_i(\theta_q, \varphi_i; \bar{\Phi}) - \mathcal{F}_i(\theta_q, \varphi_i; \hat{\Phi}) \right|. \quad (2)$$

By fixing the input φ to the quadrant directions φ_i , each quadrant’s forward model depends solely on θ . We uniformly sample T angles θ_q within the supervision range of \mathcal{L}_{BSE} , and compute the intensity difference between $\mathcal{F}^{\hat{\Phi}}$ and $\mathcal{F}^{\bar{\Phi}}$. The mean error \mathcal{E}_{BSE} measures how accurately the BSE forward model is learned.

Here, we aim to define a metric to demonstrate the percentage of BSE shadows successfully separated by NFH-SEM. BSE shadows are inherently soft and do not exhibit clear, binary boundaries, making it unsuitable to evaluate shadow estimation accuracy using hard shadow masks. Moreover, stronger shadows result in a larger BSE loss and therefore have a more significant negative impact on the learning of both geometry and the BSE forward model. For these reasons, the shadow evaluation metric must consider not only the spatial distribution of shadows but also their intensity. After training, shadow-free 4Q-BSE images are computed using the estimated surface normals \hat{n} and the learned model $\mathcal{F}^{\hat{\Phi}}$. Subtracting the measured BSE images from these predictions yields the estimated shadow maps, $\hat{\psi} = \left| \mathcal{F}(\hat{n}; \hat{\Phi}) - b' \right|$. These shadows are not absorbed into the neural field geometry or the forward model, demonstrating NFH-SEM’s ability to disentangle shadows in BSE images. Given the ground-truth shadow intensities $\bar{\psi}$, we define a 4Q-BSE soft shadow accuracy metric:

$$\mathcal{S}_{\text{shadow}} = 100 \times \left(1 - \frac{1}{4P} \sum_i^{A,B,C,D} \sum_u^P \frac{\sum_v^Q |\bar{\psi}_{iuv} - \hat{\psi}_{iuv}|}{\sum_v^Q (\bar{\psi}_{iuv} + \hat{\psi}_{iuv})} \right), \quad (3)$$

where P is the number of viewpoints with index u , and Q is the number of pixels per image with index v . This percentage score is computed by measuring the pixel-wise intensity difference between the estimated and ground-truth shadow maps and normalizing it by their total intensity. Higher values of $\mathcal{S}_{\text{shadow}}$ indicate more accurate shadow estimation.

7. Additional Experiments

7.1. Results on Learning-Based MVS Methods

We evaluate two representative learning-based MVS methods, GeoMVSNet [19] and MVSFormer++ [1], by applying their DTU [6] pretrained models to our NFH-SEM dataset. As shown in Fig. 3, both methods generalize poorly to the microscopic SEM domain. Their reconstructed surfaces are highly noisy, contain large structural inconsistencies, and often fail to produce complete geometry. These observations further highlight the domain gap between conventional RGB-based MVS training data and SEM imaging.

7.2. Robustness to Geometric Initialization

NFH-SEM relies on a coarse initial geometry to guide neural field optimization. To evaluate the robustness of our method to the quality of this initialization, we replace the initialization used in the main paper with the much worse surface reconstructions generated by GeoMVSNet [19] and MVFormer++ [1]. As shown in Fig. 3, NFH-SEM still converges to clean, complete, and high-fidelity surfaces even when initialized with geometry that contains severe noise, large missing regions, and surface holes. This experiment highlights the strong robustness of NFH-SEM to noisy or incomplete initial geometry and demonstrates the effectiveness of our reconstruction pipeline.

7.3. Ablations on TPL Microstructures

We provide additional qualitative ablation studies on real TPL microstructures using the same ablation settings and component definitions as in the simulation experiments presented in the main paper. As shown in Fig. 5, different strategies for gradient extraction from 4Q-BSE signals lead to notably different reconstruction results. Simplified BSE forward models or gradients derived from single-view approximations produce degraded surface geometry. As illustrated in Fig. 6, the shadow separation strategy further improves reconstruction quality, especially around structural edges affected by strong shadowing. The qualitative trends observed on real TPL data are consistent with the quantitative results from simulations, reinforcing the importance and effectiveness of each module in NFH-SEM.

7.4. Experiments on Simulated 3Q-BSE Data

Most modern SEM systems are equipped with multi-quadrant BSE detectors. The most common configuration is a four-quadrant BSE (4Q-BSE) detector with 90° spacing between adjacent quadrants, as in the ZEISS Gemini 560 SEM used in our experiments. In contrast, SEM systems from Thermo Fisher Scientific (formerly FEI) often employ a three-quadrant BSE (3Q-BSE) detector [10], which uses 120° spacing between quadrants. While prior methods are typically tailored to a specific detector configuration, most commonly the widely used 4Q-BSE detector [3, 14, 15], our method generalizes across different detector setups and can be readily adapted to a wide range of SEM systems. To demonstrate the flexibility of NFH-SEM, we constructed a simulated 3Q-BSE dataset, as shown in Fig. 7. All other experimental conditions remain identical to those used in our 4Q-BSE simulations, except for the detector configuration. As shown in Fig. 8, NFH-SEM reliably reconstructs accurate surface geometry under the 3Q-BSE configuration, demonstrating its flexibility and practical applicability across different SEM systems.

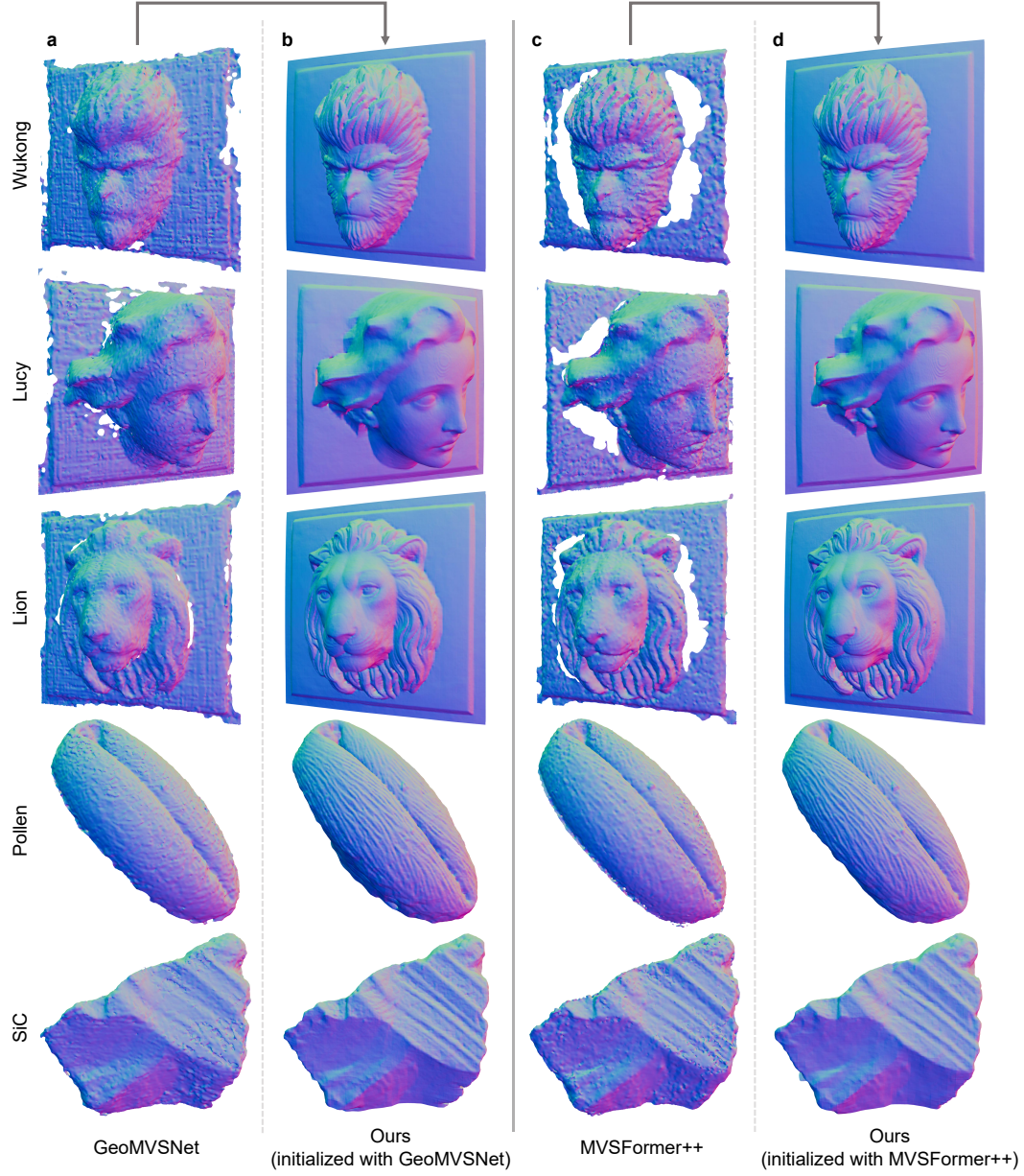


Figure 3. **Evaluation of learning-based MVS methods and robustness of NFH-SEM to different initializations.** (a, c) Surface reconstructions of GeoMVSNet [19] and MVSFormer++ [1]. (b, d) NFH-SEM reconstructions initialized with (a) and (c), respectively.

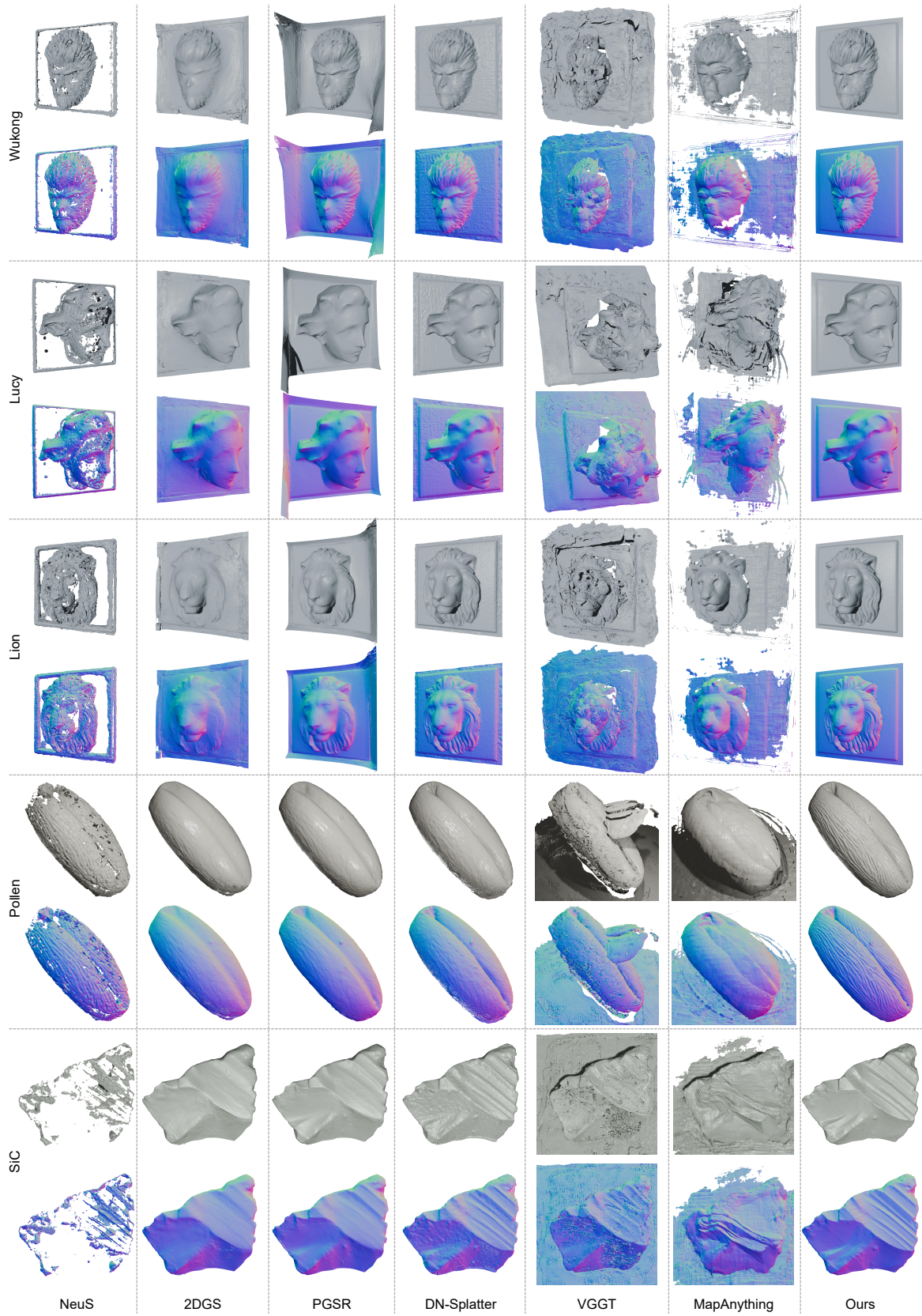


Figure 4. Results of learning-based surface reconstruction methods on the NFH-SEM dataset.

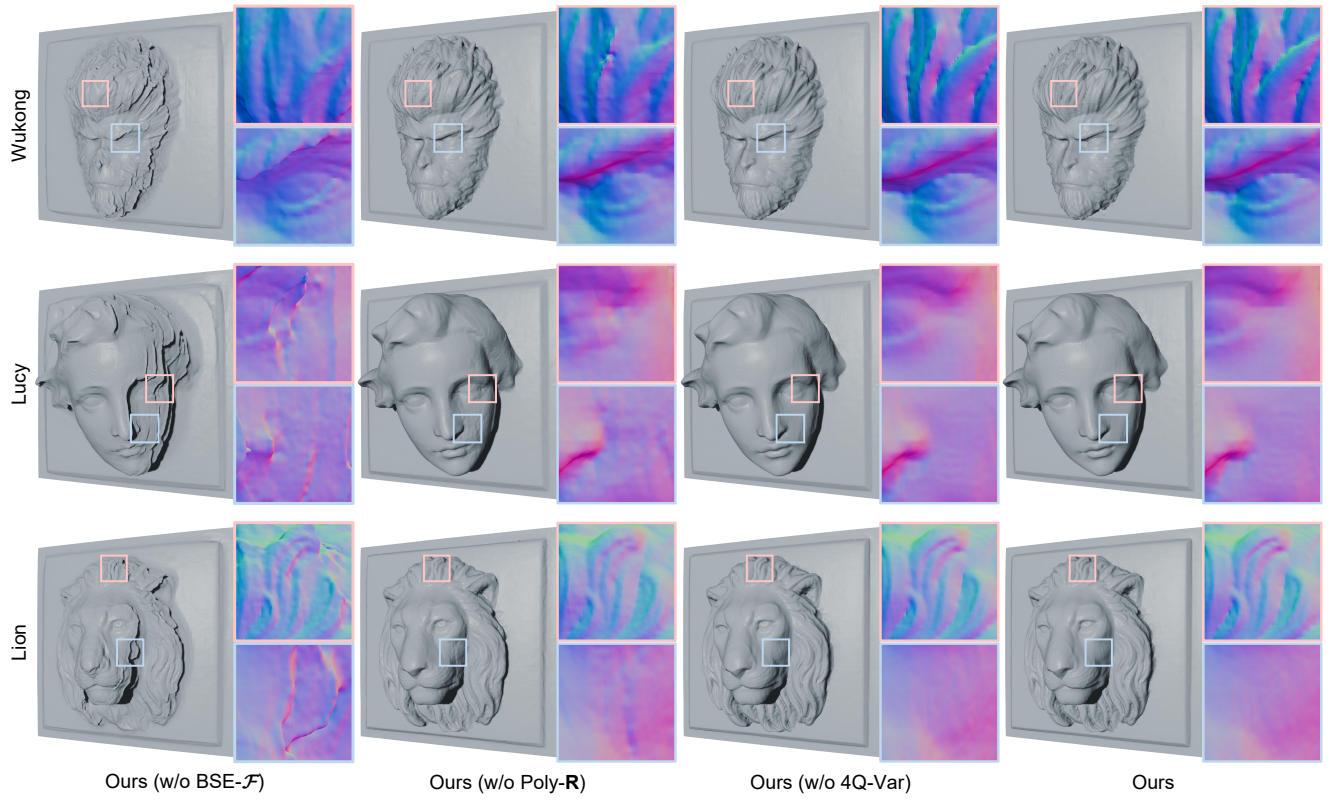


Figure 5. Comparison of BSE gradient extraction strategies for TPL microstructure reconstruction.

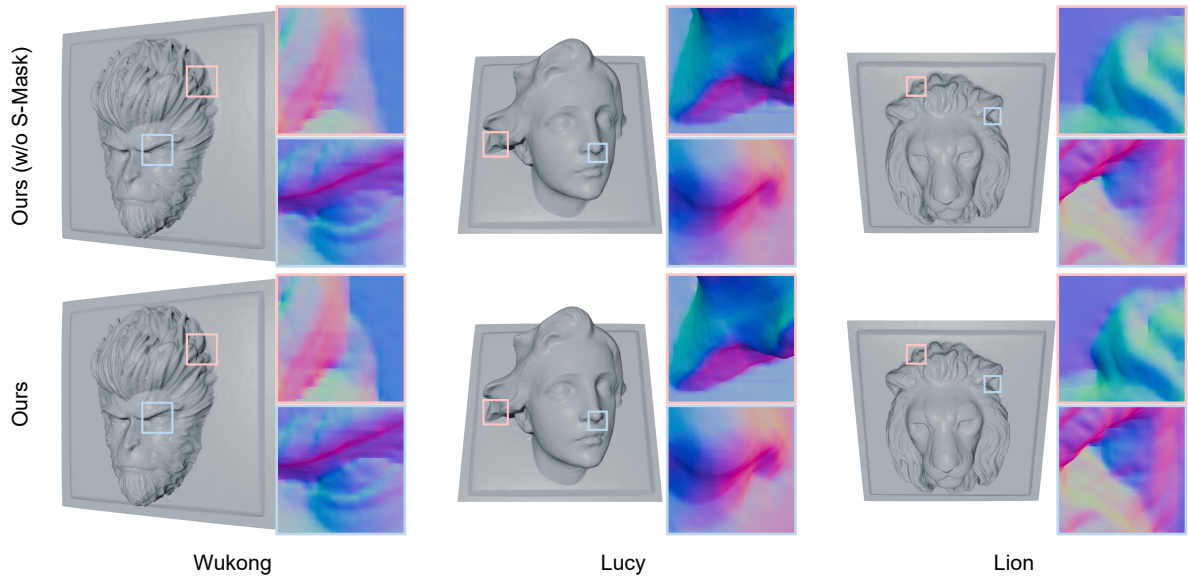


Figure 6. Effect of the shadow separation strategy on TPL microstructure reconstruction.

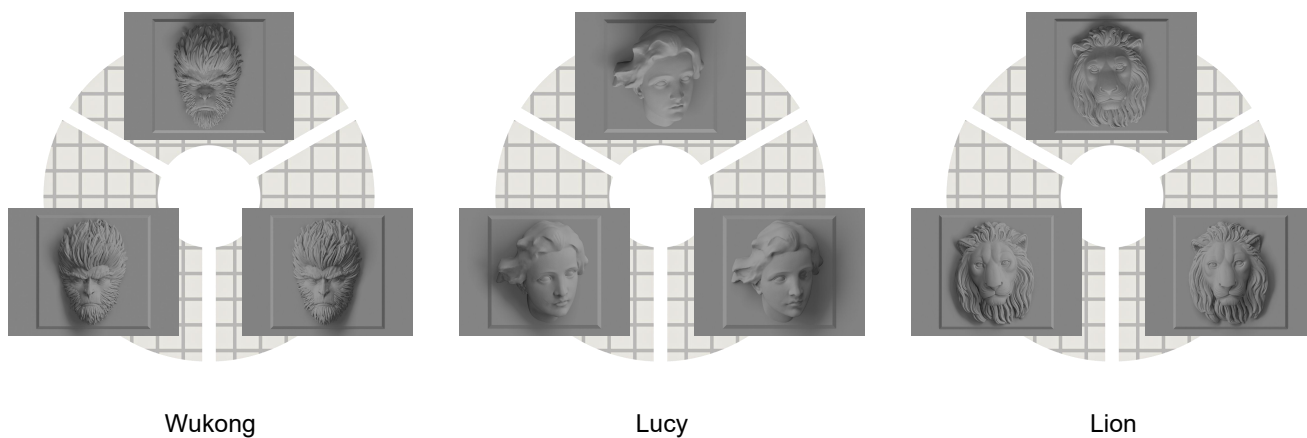


Figure 7. **Simulated BSE images acquired using a three-quadrant BSE detector.**

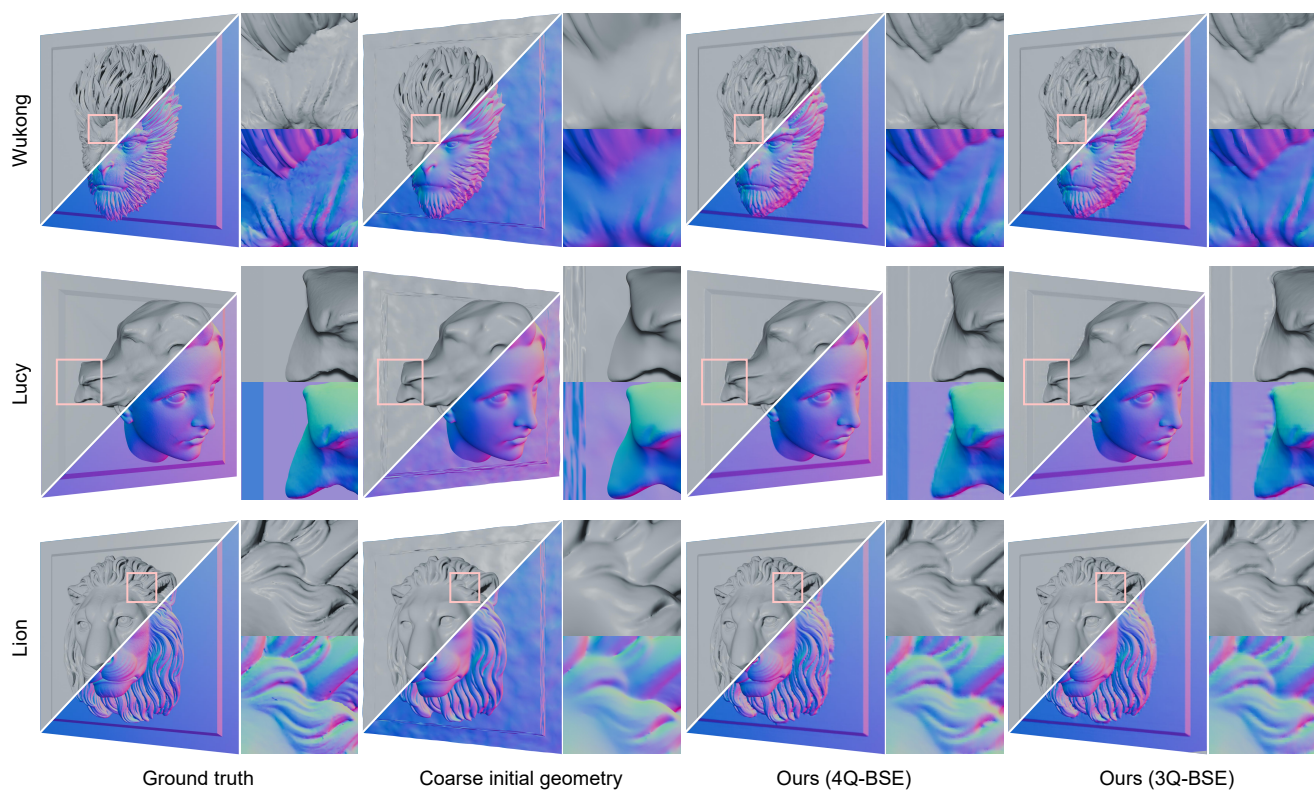


Figure 8. **NFH-SEM reconstruction results on simulated data under different BSE detector configurations.**

References

- [1] Chenjie Cao, Xinlin Ren, and Yanwei Fu. MVSFormer++: Revealing the devil in transformer's details for multi-view stereo. In *International Conference on Learning Representations (ICLR)*, 2024. 3, 4, 5
- [2] Amicia D Elliott. Confocal microscopy: principles and modern practices. *Current protocols in cytometry*, 92(1):e68, 2020. 2
- [3] Matteo Giardino, Devanarayanan Meena Narayana Menon, and Davide Luca Janner. Regularization techniques for 3D surface reconstruction from four quadrant backscattered electron detector images. *Ultramicroscopy*, 250:113746, 2023. 4
- [4] F. Golek, P. Mazur, Z. Ryszka, and S. Zuber. AFM image artifacts. *Applied Surface Science*, 304:11–19, 2014. 2
- [5] Yuan-Chen Guo. Instant neural surface reconstruction, 2022. <https://github.com/bennyguo/instant-nsr-pl>. 1
- [6] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engin Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 406–413, 2014. 3
- [7] Peter Lehmann, Stanislav Tereschenko, and Weichang Xie. Fundamental aspects of resolution and precision in vertical scanning white-light interferometry. *Surface Topography: Metrology and Properties*, 4(2):024004, 2016. 2
- [8] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3D surface construction algorithm. In *Seminal graphics: pioneering efforts that shaped the field*, pages 347–353. 1998. 1
- [9] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022. 1
- [10] Jan Neggers, Eva Hériprié, Marc Bonnet, Denis Boivin, Alexandre Tanguy, Simon Hallais, Fabrice Gaslain, Elodie Rouesne, and Stéphane Roux. Principal image decomposition for multi-detector backscatter electron topography reconstruction. *Ultramicroscopy*, 227:113200, 2021. 4
- [11] J Paluszynski and W Słowko. Surface reconstruction with the photometric method in SEM. *Vacuum*, 78(2-4):533–537, 2005. 1
- [12] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. PyTorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32:8024–8035, 2019. 1
- [13] Grzegorz Pyka, Greet Kerckhofs, Jan Schrooten, and Martine Wevers. The effect of spatial micro-CT image resolution and surface complexity on the morphological 3D analysis of open porous structures. *Materials Characterization*, 87:104–115, 2014. 2
- [14] Witold Słowko. Specific features of the miniature ionisation BSE multi-detector unit for 3D imaging in environmental conditions. *Micron*, 126:102752, 2019. 1, 4
- [15] W Słowko and M Krysztof. Detector system for three-dimensional imaging in the variable pressure/environmental SEM. *Acta Physica Polonica A*, 123(5):877–879, 2013. 4
- [16] Hao Wang, Wang Zhang, Dimitra Ladika, Haoyi Yu, Darius Gailevičius, Hongtao Wang, Cheng-Feng Pan, Parvathi Nair Suseela Nair, Yujie Ke, Tomohiro Mori, et al. Two-photon polymerization lithography for optics and photonics: fundamentals, materials, technologies, and applications. *Advanced Functional Materials*, 33(39):2214211, 2023. 2
- [17] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *Advances in Neural Information Processing Systems*, 34:27171–27183, 2021. 1
- [18] C Shan Xu, Kenneth J Hayworth, Zhiyuan Lu, Patricia Grob, Ahmed M Hassan, José G García-Cerdán, Krishna K Niyogi, Eva Nogales, Richard J Weinberg, and Harald F Hess. Enhanced FIB-SEM systems for large-volume 3D imaging. *elife*, 6:e25916, 2017. 2
- [19] Zhe Zhang, Rui Peng, Yuxi Hu, and Ronggang Wang. GeoMVSNet: Learning multi-view stereo with geometry perception. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 21508–21518, 2023. 3, 4, 5