Title: Breaking Free from the Acquisition Dogma for Volume Electron Microscopy

Bohao Chen[1,2], Fangfang Wang[3], Haoyu Wang[3], Yanchao Zhang[2,4], Zhuangzhuang Zhao[2], Haoran Chen[2,5], Hua Han[2,4]*, Xi Chen[2]*, Yunfeng Hua[3]*

[1]School of Advanced Interdisciplinary Sciences, University of Chinese Academy of Science, Beijing, China
[2]State Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[3]Shanghai Institute of Precision Medicine, Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[4]School of Future Technology, University of Chinese Academy of Sciences, Beijing, China
[5]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China
E-mail: hua.han@ia.ac.cn; xi.chen@ia.ac.cn; yunfeng.hua@shsmu.edu.cn

Abstract

Volume electron microscopy (VEM) enables three-dimensional (3D) visualization of thick specimens through serial sectioning and nanometer-resolution imaging. In an effort to overcome daunting challenges in subsequent data analysis, people tend to employ excessively slow imaging and high resolution, both of which significantly prolong the acquisition. Here, using authentic and synthetic VEM datasets through image stack re-registration, we demonstrate the implications of 3D oversampling in maintaining structural information fidelity against high background noise. This provides a fact-based argument for a higher priority of improving the spatial sampling rate, particularly in the most difficult axial direction, than suppressing the image noise during a VEM acquisition. To accelerate VEM acquisition by leveraging the

oversampled 3D context, we explore distinct machine learning-based methods for restoring serial images that contain either low-contrast snapshots or skipped sections. On the datasets of similar acquisition consumptions, 3D context-based (volumetric) denoising outperforms state-of-the-art interpolation and 2D denoising approaches. Furthermore, the volumetric denoising models can proceed in a self-supervised manner, thereby no longer relying on specific training datasets. This work is not only instructive for planning efficient large-scale acquisition on commercial setups but also benchmarks the methodology of optimizing VEM acquisition to facilitate automated image processing.

Introduction

Volume electron microscopy (VEM) is a series of rapidly evolving technologies that enable the automated acquisition of a large image stack either from consecutive ultra-thin sections or iteratively exposed block faces of a resin-embedded sample.[1,2] Although both serial section and block face-types of VEM can provide a lateral resolution of about 5 nm, the minimal cutting thickness that can be maintained over hundreds of cycles differs from each other due to distinct mechanics.[2-5] Typically, a commercial ultra-microtome can produce high-quality serial sections by cutting with the diamond blade at a step size of 30-40 nm, below which increased occurrences of varying cutting quality and section damage are observed.[6] In contrast, serial-block-face (SBF) scanning electron microscope (SEM) can homogeneously remove the sample surface at every 25 nm with an in-chamber ultramicrotome, as the resin block is much more robust against mechanical deformation than ultra-thin sections.[7] Among all the VEM, focused ion beam (FIB) SEM provides the highest axial resolution owing to its capability of sample surface milling with Ga ions at a precision of up to 5 nm.[5,8] Considering limited setup availability and reasonable running costs, it is therefore unsurprising that most VEM volumes (except for very few produced by FIB-SEM) have been registered with anisotropic voxel sizes to date. Although the importance of improved axial resolution is stressed for the sake of not only an error-free structure annotation (utmost important for neurites) but also the development of automated segmentation algorithms, the implication of anisotropic resolution on the preservation of structural details and segmentation difficulty has not yet been studied comprehensively.[3]

With the ambition to map tissue volumes at an ever-increasing scale, both spatial and temporal sampling rates need to be taken into account. Collective efforts have been made to accelerate VEM image acquisition, including sample contrast methods with saturated heavy metal deposition[9-11], fast scopes empowered by a camera array[12] or multiple electron beams[13], automated and parallelized imaging pipelines[14], etc. Nevertheless, a trade-off among pixel size, acquisition duration, and signal-to-noise ratio (SNR) of images is often made by users of commercial VEM for a feasible or affordable imaging plan. Moreover, for SBF and FIB-SEM, improved image quality by prolonging pixel dwell time is frequently accompanied by compromised axial resolution, as heightened electron dose may lead to severe radiation damage to the sample block face that hinders the next round of ultra-thin surface removal.[3,4] This raises a key question for almost every VEM acquisition, namely, should a higher priority be given to either fine imaging or thin sectioning?

For the above dilemma, recent developments in artificial intelligence (AI)-based image restoration may offer a promising solution without compromise. By learning features from a variety of published VEM datasets, state-of-the-art methods have enabled image denoising by diffusion[15] or general[16] models, as well as generating datasets of isotropic voxel size through nonlinear interpolation[17-20]. However, the accuracy of model inference was predominantly assessed by comparison between the ground truth and synthetic data recovered from artificial degradations or among different model outcomes, leaving the general application of those specifically trained models poorly tested. Moreover, image denoising and pixel interpolation are

independent tasks, for which AI models demonstrate differential performance, leading us to seek a machine-friendly VEM acquisition strategy with regard to maximum recovery of structural information using the AI model of choice.

Here, we utilized a public FIB-SEM dataset and simulated the influence of various acquisition conditions on dataset quality. Next, we compared specific AI performances on improving the dataset quality of test stacks acquired using a hybrid slow and fast imaging routine on an SBF-SEM. Last, we demonstrated a use-dependent optimization of SBF-SEM acquisition for a self-supervised AI denoising.

Results

Structural information fidelity under different acquisition settings

To explore quantitatively the implication of section thickness and image background noise on the VEM dataset quality, we took a publicly accessible high-quality FIB-SEM dataset of the mouse brain tissue[21] to generate a series of synthetic datasets through noise masking and spatial down-sampling (Fig. 1a). First, the original image stack with an isotropic voxel size of 5 nm was down-sampled in both lateral (XY) and axial (Z) dimensions to yield new stacks at 10-nm and 15-nm voxel sizes. Next, to simulate low-SNR datasets acquired from the same sample due to shortened pixel dwell times, Poisson-Gaussian noise of anticipated intensity was added to the raw images. Meanwhile, Z-extraction of the stack was carried out by excluding every two, three, and four sections, respectively. To enable an intuitive comparison of dataset quality among combinations of thicker cutting and faster imaging, we artificially equalized the scales of Z-extraction and noise level by normalizing to the estimated acquisition time based on reported reference values[22] (Supplementary Fig. 1, also see Materials and Methods).

Quantitative quality assessments of these synthetic datasets were carried out by computing their effective spatial resolution (Fig. 1b and c) and organelle segmentation accuracy (Fig. 1d and e). Notably, the obtained effective resolution of data with a high Z-sampling rate was robust to varying image SNRs (Fig. 1c), resulting in a “comfort zone” for relatively stable data quality upon different imaging speeds estimated based on the SNR. Further, we compared the performance of automated mitochondrial

segmentation among the synthetic datasets. Accurate segmentations (low 1-IoU values) were achievable from the datasets of high Z-sampling rate despite a modest SNR (Fig. 1e), which typically featured high effective resolutions (Supplementary Fig. 2). These observations were found to be consistent across all the simulated scenarios of further data down-sampling (Fig. 1c and e). Besides modeling with synthetic datasets, this spatial sampling-dependent influence of SNR on the VEM dataset quality was validated further by physically acquired FIB-SEM datasets from a different brain sample, but repeatedly at two different scanning rates (Supplementary Fig. 3).

Deep learning-based restoration of serial images

Considering that anisotropic VEM voxel sizes are primarily caused by mechanical challenges in reliably cutting thinner slices or creating uniform block faces at smaller step sizes, the above results substantiate a yet underestimated role of section thickness in counteracting noise and maintaining VEM dataset quality for automated image processing. We speculate that spatial oversampling contributes to robust digital preservation of biological structures, which are continuous across consecutive sections. To test whether such 3D structural constraints could be utilized for AI-based restoration of serial images, we designed a hybrid imaging routine that allowed the acquisition of arbitrary interlaced high-SNR reference and low-SNR snapshot slices by alternating imaging speeds. In practice, a mouse brain sample was cut consecutively at 50 nm and imaged at 2 μs as well as 0.5 μs pixel dwell time using a commercial SBF-SEM (Fig. 2a). The obtained images were arranged in a use-dependent fashion to generate distinct

types of test datasets (Fig. 2b), including a reference stack (R-stack, all scanned at 2 μs/pixel) and a snapshot stack (S-stack, all scanned at 0.5 μs/pixel) as well as a subset of references by skipping every other slices (½R-stack) and a hybrid dataset with a fixed combination of ¾ snapshots and ¼ references (S/r-stack).

These datasets provided a unique opportunity to compare AI performances on tasks including interpolation, 2D denoising, and newly established volumetric denoising (see Supplementary Fig. 4, Materials and Methods for the network architecture and more details) based on a shared ground truth (Fig. 2c). First, as one would anticipate, significantly fewer artifacts and blurring regions were observed in the denoised images of S/r- and S-stacks using additional information from 3D structural context. Second, it was unexpected that the outcomes of volumetric denoising models were comparable between using input datasets with and without high-SNR images (S/r- and S-stack), implying a limited contribution of doping reference frames in the input stack to the inference accuracy of denoising models. Third, we noticed better structural preservation in terms of less aliasing and blurring in the volumetric denoised images (S/r- and S-stack) than the interpolated ½R-stack, again stressing the pivotal role of axial sampling beyond the image SNR. These observations were further confirmed by quantifying the effective resolution of the AI output stacks (Fig. 2d). In fact, volumetric denoising achieved even a slightly worse score of data quality using S/r- than S-stack as the input (Fig. 2e), because the denoised outcome of S/r-stack still kept those reference frames containing low residual noise (Supplementary Fig. 5).

Optimization of VEM acquisition for volumetric denoising

As for SBF-SEM, the axial resolution is mainly limited by the mechanical cutting of the resin block. Prolonged pixel dwell time heightened the radiation damage, which alters the mechanical properties of the block face and makes it more difficult to cut. Effective volumetric denoising of a dataset containing exclusively fast-scanned images creates a margin for cutting at smaller step sizes by SBF-SEM, and in turn, the AI performance may benefit from an increased axial sampling rate. To test this idea, we acquired three image stacks from a sample using 0.5 μs dwell time but at different cutting thicknesses (50 nm, 25 nm, and 20 nm). Due to the absence of high-SNR reference frames, we adopted a completely self-supervised strategy for the model training. Based on visual inspection, the AI delivered reasonable denoising with input stacks acquired at all tested cutting thicknesses (Fig. 3a-c, Supplementary Video 1 and 2), but further quantification revealed an improvement of the effective resolution in the XZ plane from 850.3 ± 361.8 nm to 209.5 ± 6.6 nm (50 nm cutting), 306.5 ± 98.9 nm to 104.1 ± 3.3 nm (25 nm cutting), and 263.6 ± 97.0 nm to 83.6 ± 2.0 nm (20 nm cutting) after blind volumetric denoising by AI (Fig. 3d). Note that the achieved effective XZ resolution by denoising were found to decrease linearly with the physical reduction of cutting thickness (Fig. 3e), serving as a good sign for minimal residual noise in the model outcomes. By comparing the cutting-thickness-independent XY component of effective resolution, we further controlled that such improvements in data quality were not due to inter-sample variability (Fig. 3f and g).

Discussion

In principle, VEM imaging of biological specimens is a 3D-digitalization process, in which the sampling rate and sampling error correspond to voxel size and its SNR. For most subcellular structures, the lateral resolution of modern electron microscopy is more than sufficient, thereby leading to oversampled structural details in XY. In the third dimension, however, the increase of sampling rate is substantially limited by technical challenges or time constraints, hindering isotropic voxel sizes of VEM. In the present work, we have comprehensively studied the trends of VEM dataset quality changes induced by different tradeoffs between Z-sampling and image SNR (Fig. 1). By doing so, we characterized the so-called "comfort zone" for VEM acquisition at adequate Z-sampling rates, within which the improvement of effective resolution and eventually also segmentation accuracy is saturated upon the SNR enhancement. Although the approach may be tissue type-specific, it is particularly instructive for choosing the parameters to maximize the acquisition efficiency of VEM in the future. Especially, emerging techniques like gas-cluster ion beam[23] and parallel ion beam source[24] are capable of pushing the thickness of large-scale surface etching down to 10-nm precision, paving the way for industrial-scale tissue mapping at desired resolutions.

Moreover, our quantification implies that sufficient spatial sampling appears to be the determinant for structural information fidelity with an add-on effect of the background noise (Fig. 1). In line with this notion, our fair comparison between AI performance on different image restoration tasks (Fig. 2) suggests that volumetric denoising outperforms 2D denoising and interpolation in a realistic experiment setting

enabled by the hybrid acquisition routine. This finding is significant in two aspects. First, it provides proof-of-principle evidence for learning structural features with spatial continuity by AI to accurately classify weak signals and random noise in the VEM stacks. Second, as computer devices are much more powerful in memorizing and comparing large image sequences with a wide contrast range than human beings, this stimulates new thinking of suitable VEM sampling strategies primarily for AI-based image processing rather than visual assessment. Indeed, those high-SNR reference images that are favorable to visual inspection turn out to be not essential for accurate volumetric denoising by AI (Fig. 2). Nevertheless, our attempt with such a hybrid acquisition may serve as a human-in-the-loop paradigm for future method development of AI-based image processing beyond denoising, for instance, super-resolution or defocus correction.

Lastly, note that self-supervised volumetric denoising no longer requires the reference frame (Fig. 3), making this approach not dataset-specific and thereby promising for a general application on VEM volumes of different tissue types and VEM setups. More importantly, it provides a new perspective for serial block-face imaging with improved Z-resolution or more robust surface removal, which are negatively influenced by radiation damage in the practice of both SBF and FIB-SEM. To the best of our knowledge, the actual cutting thickness of SBF-SEM is still limited by 25 nm[25-27], while our AI-denoising enables an acquisition with shortened dwell time and improved axial resolution at least to 20 nm using a commercial setup (Fig. 3). As inter-section differences become smaller upon reduction in the cutting thickness, we expect

an improved Z-alignment of serial images and that may further contribute to the performance of volumetric denoising.

In summary, the present work argues for a pivotal role of axial resolution in maintaining VEM data quality by numerical scoring. In line with this notion, volumetric denoising appears to be a more efficient and effective task for image stack restoration than generative interpolation using machine learning. This result triggers new thinking of adapting our current VEM sampling strategies in favor of AI-based processing, which can substantially differ from our current gold standard - visual assessment by human annotators, which is practical primarily on single images.

Methods

Animal experiment

Adult C57BL/6 mice (male) were purchased from Shanghai Jihui Experimental Animal Feeding Co., Ltd. (Shanghai, China). All procedures were approved by the ethical committee of Shanghai Ninth People's Hospital (No. SH9H-2019-A567-1) and conducted at Shanghai Institute of Precision Medicine.

The mouse brain samples were prepared according to the previously published procedure[9] with minor modifications. In brief, animals were anesthetized with isoflurane inhalation and transcardially perfused with 30 mL 1.5 M cacodylate buffer (CB, pH 7.4), followed by an ice-cold fixative mixture containing 2% paraformaldehyde and 2.5% glutaraldehyde in 0.08 M CB. The dissected brain tissues were post-fixed at 4°C in the same fixative for 48 hours and then sliced 500-μm thick using a vibratome (Leica VT 1200S) in 0.15 M CB. The *en bloc* staining was done by multiple rounds of incubation with heavy metals, including 2% OsO_4 (in 0.15 M CB), 2.5% ferrocyanide (in 0.15 M CB), 2% OsO_4 (in 0.15 M CB) at room temperature (RT) for 1.5, 1.5, and 1.0 hours, respectively. After being washed with 0.15 M CB and water for 30 min each, the samples were incubated at RT in thiocarbonhydrazide (saturated aqueous solution) for 1 hour at RT, 2% OsO_4 aqueous solution for 1.5 hours at RT, and 1% uranyl acetate (aqueous solution) for 2 hours at 50°C with intermediate washing step (twice in water for 30 min each). The sample dehydration was carried out through a graded acetone series (50%, 75%, 90%, 30 min each, all cooled at 4°C) and in pure acetone (100%, three times, 30 min each). Finally, the samples were infiltrated through a 1:1 mixture of acetone and Spurr's resin monomer (4.1g ERL 4221, 0.95g DER 736, 5.9g NSA, and 1% DMAE) overnight on a rotator, then embedded in pure resin for 8 hours before being baked in a 70°C oven for 72 hours.

Data acquisition of SBF-SEM

The sample was mounted on an aluminum metal rivet using epoxy adhesive for 12 hours and trimmed to a rectangular shape (0.8 × 0.8 mm^2) using a sample trimmer

(Leica, TRIM2). Next, the sample surface was prepared by a diamond knife using an ultramicrotome (Leica, UC7). Sputter coating of a thin layer of gold (Leica, EM ACE600) was applied to the sample block, which was then imaged using a field-emission SEM (Zeiss, Gemini300) equipped with an in-chamber ultramicrotome (Gatan, 3View2XP) and back-scattered electron detector (Gatan, OnPoint).

The SEM imaging was carried out with an incident beam energy of 2 keV and a pixel size of 15 nm. For Fig. 2, serial images (4000 × 4000 pixels, 365 slices) were acquired at a dwell time of 0.5 μs with a cutting thickness of 50 nm. For each imaging and cutting cycle, repeated scans with a dwell time of 2.0 μs were manually triggered to capture redundant images of improved quality. These serial images were assigned in use-dependent orders to compose the reference stack (all frames acquired with 2.0-μs dwell time), the snapshot stack (all frames acquired with 0.5-μs dwell time), and the fast-slow-scan dataset (a combination of one 2.0-μs-dwell-time frame and three 0.5-μs-dwell-time frames). For Fig. 3, the snapshot stack was reused for the 50-nm-cutting experiment. And two independent datasets were acquired using the same imaging parameters, except that the cutting thickness was reduced to 25 nm (902 slices) and 20 nm (626 slices). All SBF-SEM datasets contained at least 301 slices for model training and testing after Z-alignment using a self-written MATLAB script based on cross-correlation maximum between consecutive slices before analysis[28].

Data simulation on a public FIB-SEM volume

To simulate dataset quality under different image settings (Fig. 1), a publicly available dataset was used, which was acquired from the mouse CA1 hippocampus region using FIB-SEM at an isotropic voxel size of 5 nm (https://www.epfl.ch/labs/cvlab/data/data-em/)[21]. For the XY down-sampling, binning of image pixels at a factor of 2 and 3 was used, yielding a final XY pixel size of 10 and 15 nm. To perform down-sampling along the Z-direction at a factor of k, the new image stack kept every k slice from the original stack. For a given volume, this process resulted in a k-fold increase in the slice thickness, a k-fold reduction in the number of imaged slices, and an equivalent 1/k fraction of the original acquisition time.

To simulate the effect of a reduced pixel dwell time, we manipulated the image signal-to-noise ratio by introducing Poisson-Gaussian noise of different levels to the original images. The outcomes were scored using the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE)[22] for a quantitative calibration. In detail, the variance of the noise used for masking was adjusted so that the obtained BRISQUE score of the masked image matched closely the documented value for images acquired with a specific dwell time[15]. By doing so, we were able to generate a series of datasets with comparable noise levels to images scanned with different dwell times of 0.5 μs, 1.0 μs, 1.5 μs, 2.0 μs, 4.0 μs, and 16.0 μs, respectively (Supplementary Fig. 1).

Computing effective spatial resolution of EM volumes

Sectional Fourier Shell Correlation (SFSC) analysis was used to estimate the effective spatial resolution of the cubed VEM stacks[29]. This method determined directional resolution by calculating the normalized cross-correlation within wedge-shaped sectors of individual Fourier shells. For analysis of an EM volume, the input data was first split into two pairs of sub-volumes based on pixel coordinate on the XY plane:

Pair 1: One sub-volume containing all pixels at (even, even) coordinates and the other containing pixels at (odd, odd) coordinates.

Pair 2: A second pair formed from pixels at (even, odd) and (odd, even) coordinates. The SFSC for each pair was then calculated using the normalized cross-correlation between their Fourier transforms:

$$SFSC_{\mathrm{pair}_i}(r_i, \alpha_i) = \frac{\left(\sum_{r \in r_i, \alpha \in \alpha_i} F_1(r, \alpha, \phi) \cdot F_2(r, \alpha, \phi)\right)}{\sqrt{\sum_{r \in r_i, \alpha \in \alpha_i} F_1^2(r, \alpha, \phi) \cdot \sum_{r \in r_i, \alpha \in \alpha_i} F_2^2(r, \alpha, \phi)}},$$

where $F_1(r, \alpha, \phi)$ and $F_2(r, \alpha, \phi)$ denote the voxels in two Fourier-transformed sub-volumes within pair_i that are located at a given distance r_i from the origin and within an orientation sector, defined by α and ϕ. Finally, the SFSC curves on each pair were averaged to yield the final SFSC measurement for the input volume:

$$SFSC = \frac{SFSC_{\mathrm{pair}_1} + SFSC_{\mathrm{pair}_2}}{2}$$

To ensure computational efficiency while obtaining a robust measurement for the entire volume, our analysis was performed on 50 sub-volumes of a uniform shape (128 × 128 × 128) randomly sampled from input volume. The final effective spatial resolution reported for a given volume is the average of these individual SFSC measurements.

Python package for the SFSC analysis was downloaded from an open-source Github repository (sakoho81/miplib: A Python software library with a variety of functions for (optical) microscopy image restoration, reconstruction and analysis.) and implemented for this study. The resolution cut-off was determined by a default criterion (SNR-0.5) recommended from official examples, and the obtained directional resolution profiles were visualized as a polar plot. For comparison among EM volumes, we used the resolution cut-offs on the 90° axis (corresponding to the XZ plane), on which this value was expected to be equally sensitive to image quality and slice thickness.

Quantification of automated mitochondrial segmentation

The data quality was assessed based on the mitochondrial segmentation accuracy of a published 3D AI segmentation model combining a U-Net with Swin-Transformer blocks (STU-Net, uni-medical/STU-Net: The largest pre-trained medical image segmentation model (1.4B parameters) based on the largest public dataset (>100k annotations), up until April 2023., arXiv:2304.06716). Models were trained independently for down-sampled image stacks of isotropic voxel sizes (5, 10, and 15 nm). The training data and its corresponding ground truth were provided by the aforementioned publicly available FIB-SEM dataset. The training was conducted for 200 epochs using a Stochastic Gradient Descent (SGD) optimizer with momentum of 0.99 and a fixed learning rate of 1×10^{-3}. The model from the final epoch was used for inference. Each model was applied to its corresponding set of testing data with the same pixel size, and the segmentation performance was quantified on a slice-by-slice basis by calculating the Intersection-over-Union (IoU) against the human ground truth, which

was provided by the same public dataset and processed with the corresponding down-sampling factors. We plotted the segmentation error as 1-IoU (Fig. 1e) to ease visual comparison with the effect resolution of the corresponding dataset (Fig. 1d).

AI-based Z-interpolation of serial images

The input image stack was acquired at a 2-μs dwell time and spaced at 100 nm in Z (Fig. 2b). For the interpolation, we fed the pretrained Real-Time Intermediate Flow Estimation (RIFE) model[30] (previously shown to have outstanding performance on the FIB-SEM data[31]) with the input stack. By a 2-fold interpolation, the RIFE artificially improved the z-spacing of the original dataset from 100 nm to 50 nm.

Volumetric denoising network architecture and model training

For volume denoising, we adopted a cascaded U-Net-based deep denoising model named GShiftNet[32] as the core architecture (Supplementary Fig. 4). The proposed model contained an initial feature extraction block enhanced by channel attention, followed by an encoder for fusing multi-frame features by channel-wise circular shifts and 8-direction spatial shifts. The processed features were then passed through three cascaded U-Net sub-networks for deep representation learning. Throughout the network, Channel Attention Blocks (CABs) refined features by adaptively reweighting channel responses via global average pooling, making the model focus on informative signals while maintaining stability through residual connections. The final reconstruction module aggregated multi-stage features and combined them with the input via a global residual path, followed by a convolutional layer to produce the denoised output.

All volumetric denoising training was performed using 3D patches of a fixed size (128 × 128 × 5). While the initial patches were all randomly sampled from the data volumes, the precise method of generating the training pairs differed for each workflow, as detailed in the following subsections. Standard data augmentation techniques, including random flips and rotations, were applied to each patch prior to the network

feeding. Each epoch consisted of 540 training pairs and 32 validation pairs. All models were trained for a fixed 100 epochs, and the model checkpoint from the final epoch was used for all subsequent inference. During training, the model was optimized using a weighted sum of the L1 and structural similarity loss (defined as 1–SSIM). The loss function L_{total} used in training is defined as:

$$L_{total} = L_1 + (1 - SSIM)$$

Training was performed with Adam (initial LR 2 × 10-4, $\beta_1 = 0.9$, $\beta_2 = 0.99$) and a plateau-based learning rate scheduler. All experiments were run on a Linux server equipped with a single Nvidia RTX 4090 GPU.

Supervised training of volumetric denoising with serial images

For the supervised denoising experiments (Fig. 2), two distinct data preparation pipelines were implemented to generate source and target pairs to train our denoising model:

1. For the fast and slow imaging of S/r-stack, whose input sequence consisted of five consecutive frames (the 1st and 5th frames were captured by slow-scan; the 2nd to 4th frames were from fast scan), an approach was employed to prevent the model from learning a trivial identity map for the slow-scanned frames. To construct paired training samples, a larger patch of size (256 × 256 × 5) was initially extracted from both the hybrid stack and the corresponding high-SNR ground truth. From these, the input and supervision patches (128 × 128 × 5) were derived using a deterministic checkerboard subsampling scheme: the input patch was composed of pixels located at even rows and even columns (top-left), while the supervision patch was constructed from the diagonally opposite subset (bottom-right). This design enabled robust end-to-end denoising of hybrid patches.
2. For the fast-scan mode, training pairs were generated by directly extracting a 3D patch (128 × 128 × 5) from the snapshot stack and pairing it with the corresponding patch from the ground truth at the same spatial location. This pairing strategy enabled the model to learn a direct mapping from noisy inputs to their clean

counterparts. For the comparable 2D supervised denoising model (Fig. 2), the training process was identical, except that the input patch depth was reduced to one (128 × 128 × 1). This minimal modification ensured a fair comparison between the 2D and 3D approaches by keeping the model architecture, number of training samples, and optimization strategy consistent.

Self-supervised training of volumetric denoising with serial images

For the self-supervised denoising experiments lacking paired low-noise references (Fig. 3), training pairs were generated directly from the raw images. The pairing strategy followed the same principle as the supervised model training scheme described above. Specifically, a randomly cropped patch (256 × 256 × 5) was divided into two sub-patches using a deterministic checkerboard pattern, which were then assigned as the source and target, respectively. This blind-spot design encouraged the network to learn underlying structural representations rather than simply fitting noise patterns.

Volumetric denoising inference of serial images

For inference on large volumes, a patch-based approach was implemented using an overlapping sliding window strategy. To prevent block artifacts in the denoised stacks, adjacent patches (128 × 128 × 5) were processed with a 12.5% overlap in the XY plane (16 × 16), and 20% (one layer) overlap in the Z-axis. The final denoised pixels in these overlapping regions were the mean of the denoised outcomes from the adjacent patches. While this inference pipeline was applied to all denoising strategies, the hybrid fast-slow mode included an additional post-processing step to ensure a fair comparison with the interpolation method. After the initial seamless volume was generated, the slices corresponding to the positions of the original slow-scan frames (e.g., the 1st and 5th for every patch) were replaced with the original, unprocessed high-SNR slow-scan images. This step guaranteed that both volumetric denoising methods and the RIFE interpolation method share identical data at the ground truth layers.

Statistical analysis

All statistical tests were performed using GraphPad Prism 10.1.2 built-in functions. One-sample t tests for Fig. 2d (½SO, ****$p < 0.0001$, $n = 50$, $t = 49.78$, $df = 49$; S/r-stack, ****$p < 0.0001$, $n = 50$, $t = 17.07$, $df = 49$, two-tailed; S-stack, ****$p < 0.0001$, $n = 50$, $t = 13.29$, $df = 49$, all two-tailed), 3d (effective XZ resolution: 50-nm, ****$p < 0.0001$, $n = 50$, $t = 12.58$, $df = 49$; 25-nm, ****$p < 0.0001$, $n = 50$, $t = 14.72$, $df = 49$; 20-nm, ****$p < 0.0001$, $n = 50$, $t = 13.28$, $df = 49$, all two-tailed), and 3f (effective XY resolution: 50-nm, ****$p < 0.0001$, $n = 50$, $t = 65.92$, $df = 49$; 25-nm, ****$p < 0.0001$, $n = 50$, $t = 53.37$, $df = 49$; 20-nm, ****$p < 0.0001$, $n = 50$, $t = 54.64$, $df = 49$, all two-tailed). Two-sample t tests for Fig. 2e (interpolated vs. S/r-denoised, ****$p < 0.0001$, $n = 50$, $t = 12.15$, $df = 98$; interpolated vs. S-denoised, ****$p < 0.0001$, $n = 50$, $t = 15.44$, $df = 98$; S/r-denoised vs. S-denoised, ****$p < 0.0001$, $n = 50$, $t = 18.38$, $df = 98$), Fig. 3d inset (effective XZ resolution: 50-nm vs. 25-nm, ****$p < 0.0001$, $n = 50$, $t = 101.2$, $df = 98$; 50-nm vs. 20-nm, ****$p < 0.0001$, $n = 50$, $t = 129.6$, $df = 98$; 25-nm vs. 20-nm, ****$p < 0.0001$, $n = 50$, $t = 37.23$, $df = 98$), Fig. 3f inset (effective XY resolution: 50-nm vs. 25-nm, $p = 0.3391$, $n = 50$, $t = 0.9607$, $df = 98$; 50-nm vs. 20-nm, $p = 0.2149$, $n = 50$, $t = 1.248$, $df = 98$; and 25-nm vs. 20-nm, *$p = 0.0339$, $n = 50$, $t = 2.151$, $df = 98$).

Data availability

The FIB-SEM mouse CA1 hippocampus dataset, including the raw image stacks and the corresponding mitochondrial segmentation annotations used in this study, is available from the website (https://www.epfl.ch/labs/cvlab/data/data-em/). Both the fast-scan and slow-scan SBF-SEM image stacks are available from the corresponding author on reasonable request.

Code availability

The source codes used in this paper are available at https://github.com/hmzawz2/OptimizedAcquisition4vEM.

Acknowledgements

This work was supported by the Scientific Research Instrument and Equipment Development Project of Chinese Academy of Sciences (PTYQ2025TD0002 to H.H.), the National Natural Science Foundation of China (32171461 to H.H., 82171133 to Y.H.), and the Innovative Research Team of High-level Local Universities in Shanghai (SHSMU-ZLCX20211700).

Author contributions

Y.H. and X.C. conceived the project. Y.H., X.C., and H.H. supervised this study. F.W. prepared and imaged the SBF-SEM mouse brain samples. B.C. conducted the image analysis, quantifications, and validation. B.C., Y.Z., and H.C. conducted all the model training and testing. Y.Z. and H.C. performed segmentation of the FIB-SEM data. B.C., H.W., Y.H., and X.C. analyzed all the data and results. Y.H. and B.C. drafted the first manuscript with edits and approval of all authors.

Competing interests

The authors declare no competing interests.

References:

1 Collinson, L. M. *et al.* Volume EM: a quiet revolution takes shape. *Nat Methods* **20**, 777-782, doi:10.1038/s41592-023-01861-8 (2023).

2 Peddie, C. J. *et al.* Volume electron microscopy. *Nat Rev Methods Primers* **2**, 51, doi:10.1038/s43586-022-00131-9 (2022).

3 Briggman, K. L. & Bock, D. D. Volume electron microscopy for neuronal circuit reconstruction. *Curr Opin Neurobiol* **22**, 154-161, doi:10.1016/j.conb.2011.10.022 (2012).

4 Wanner, A. A., Kirschmann, M. A. & Genoud, C. Challenges of microtome-based serial block-face scanning electron microscopy in neuroscience. *Journal of Microscopy* **259**, 137-142, doi:10.1111/jmi.12244 (2015).

5 Xu, C. S. *et al.* Enhanced FIB-SEM systems for large-volume 3D imaging. *Elife* **6**, doi:10.7554/eLife.25916 (2017).

6 Schalek, R. *et al.* Development of High-Throughput, High-Resolution 3D Reconstruction of Large-Volume Biological Tissue Using Automated Tape Collection Ultramicrotomy and Scanning Electron Microscopy. *Microscopy and Microanalysis* **17**, 966-967, doi:10.1017/s1431927611005708 (2011).

7 Denk, W. & Horstmann, H. Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol* **2**, e329, doi:10.1371/journal.pbio.0020329 (2004).

8 Knott, G., Marchman, H., Wall, D. & Lich, B. Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. *J Neurosci* **28**, 2959-2964, doi:10.1523/JNEUROSCI.3189-07.2008 (2008).

9 Hua, Y., Laserstein, P. & Helmstaedter, M. Large-volume en-bloc staining for electron microscopy-based connectomics. *Nat Commun* **6**, 7923, doi:10.1038/ncomms8923 (2015).

10 Mikula, S. & Denk, W. High-resolution whole-brain staining for electron microscopic circuit reconstruction. *Nat Methods* **12**, 541-546, doi:10.1038/nmeth.3361 (2015).

11 Song, K., Feng, Z. & Helmstaedter, M. High-contrast en bloc staining of mouse whole-brain and human brain samples for EM-based connectomics. *Nat Methods* **20**, 836-840, doi:10.1038/s41592-023-01866-3 (2023).

12 Bock, D. D. *et al.* Network anatomy and in vivo physiology of visual cortical neurons. *Nature* **471**, 177-182, doi:10.1038/nature09802 (2011).

13 Eberle, A. L. *et al.* High-resolution, high-throughput imaging with a multibeam scanning electron microscope. *Journal of Microscopy* **259**, 114-120, doi:10.1111/jmi.12224 (2015).

14 Hayworth, K. J. *et al.* Ultrastructurally smooth thick partitioning and volume stitching for large-scale connectomics. *Nat Methods* **12**, 319-322, doi:10.1038/nmeth.3292 (2015).

15 Lu, C. *et al.* Diffusion-based deep learning method for augmenting ultrastructural imaging and volume electron microscopy. *Nat Commun* **15**, 4677, doi:10.1038/s41467-024-49125-z (2024).

16 Sun, H. *et al.*, doi:10.21203/rs.3.rs-6602548/v1 (2025).

17 Ferede, F. A., Khalighifar, A., John, J., Venkataraman, K. & Khairy, K. in *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)* 1-5 (2025).

18 Joshi, S. *et al.* InterpolAI: deep learning-based optical flow interpolation and restoration of biomedical images for improved 3D tissue mapping. *Nature Methods* **22**, 1556-1567,

doi:10.1038/s41592-025-02712-4 (2025).

19 Wang, Z. *et al.* STDIN: Spatio-temporal distilled interpolation for electron microscope images. *Neurocomputing* **505**, 188-202, doi:10.1016/j.neucom.2022.07.037 (2022).

20 Heinrich, L., Bogovic, J. A. & Saalfeld, S. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017 Lecture Notes in Computer Science* Ch. Chapter 16, 135-143 (2017).

21 Lucchi, A., Li, Y. & Fua, P. in *2013 IEEE Conference on Computer Vision and Pattern Recognition* 1987-1994 (2013).

22 Mittal, A., Moorthy, A. K. & Bovik, A. C. in *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)* 723-727 (2011).

23 Hayworth, K. J. *et al.* Gas cluster ion beam SEM for imaging of large tissue samples with 10 nm isotropic resolution. *Nat Methods* **17**, 68-71, doi:10.1038/s41592-019-0641-2 (2020).

24 Ma, H. *et al.*, doi:10.1101/2025.04.25.650569 (2025).

25 Helmstaedter, M. *et al.* Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500**, 168-174, doi:10.1038/nature12346 (2013).

26 Svara, F. *et al.* Automated synapse-level reconstruction of neural circuits in the larval zebrafish brain. *Nat Methods* **19**, 1357-1366, doi:10.1038/s41592-022-01621-0 (2022).

27 Briggman, K. L., Helmstaedter, M. & Denk, W. Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471**, 183-188, doi:10.1038/nature09818 (2011).

28 Jiang, Y. *et al.* Efficient cell-wide mapping of mitochondria in electron microscopic volumes using webKnossos. *Cell Rep Methods* **5**, 100989, doi:10.1016/j.crmeth.2025.100989 (2025).

29 Koho, S. *et al.* Fourier ring correlation simplifies image restoration in fluorescence microscopy. *Nat Commun* **10**, 3103, doi:10.1038/s41467-019-11024-z (2019).

30 B, P., S, N. & Sabaresan, V. in *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* 1-5 (2024).

31 Gambini, L. *et al.* Video frame interpolation neural network for 3D tomography across different length scales. *Nature Communications* **15**, doi:10.1038/s41467-024-52260-2 (2024).

32 Li, D. *et al.* in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 9822-9832 (2023).

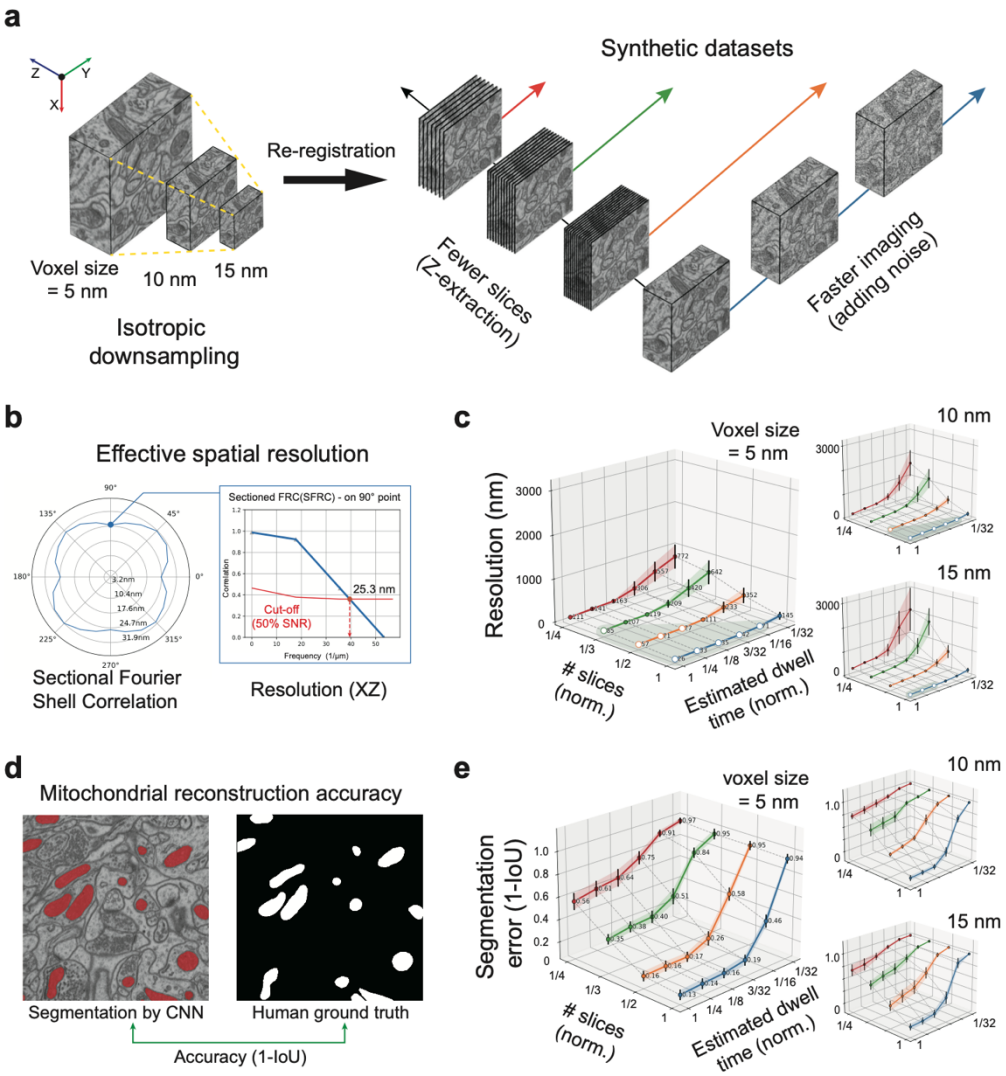

Fig. 1: Fidelity of structural information in different VEM samplings. **a,** Data simulation procedure of a public FIB-SEM volume. Down-sampling yields three synthetic datasets with isotropic voxel sizes. The obtained datasets are progressively masked with noise of varying levels and extracted in the axial direction (Z). **b,** Quantitative assessment of data quality using the effective spatial resolution, for which a sectional Fourier Shell Correlation (SFSC) is computed for each volume across different angles (as a polar plot). The resolution cut-off in the XZ plane (90° direction) is used to score changes in response to Z-extraction and noise masking. **c,** Effective XZ resolutions of datasets under different Z-extraction and noise levels. Z-sampling rate (# slices) and noise mask (estimated dwell time) are normalized to the original imaging consumption and scaled approximately to the same fractions. Distinct rising phases of colored lines suggest that resolution worsening in datasets

with low slice numbers is significantly accelerated upon heightened noise levels associated with fast scanning. Boundary conditions for the so-called “comfort zone” of acquisition (grey area), which is defined by an effective resolution below 100 nm (open dots). Inset, a similar tendency is observed in modeling with datasets of down-sampled voxel sizes. **d,** Data quality assessment by segmentation accuracy (Intersection-over-Union, IoU) of mitochondria by scoring the convolutional neural network (CNN)-based model inference with human annotation. **e,** Segmentation error (as 1 - IoU) on synthetic datasets reveals declined model performance upon reduced Z-sampling and shortened dwell time.

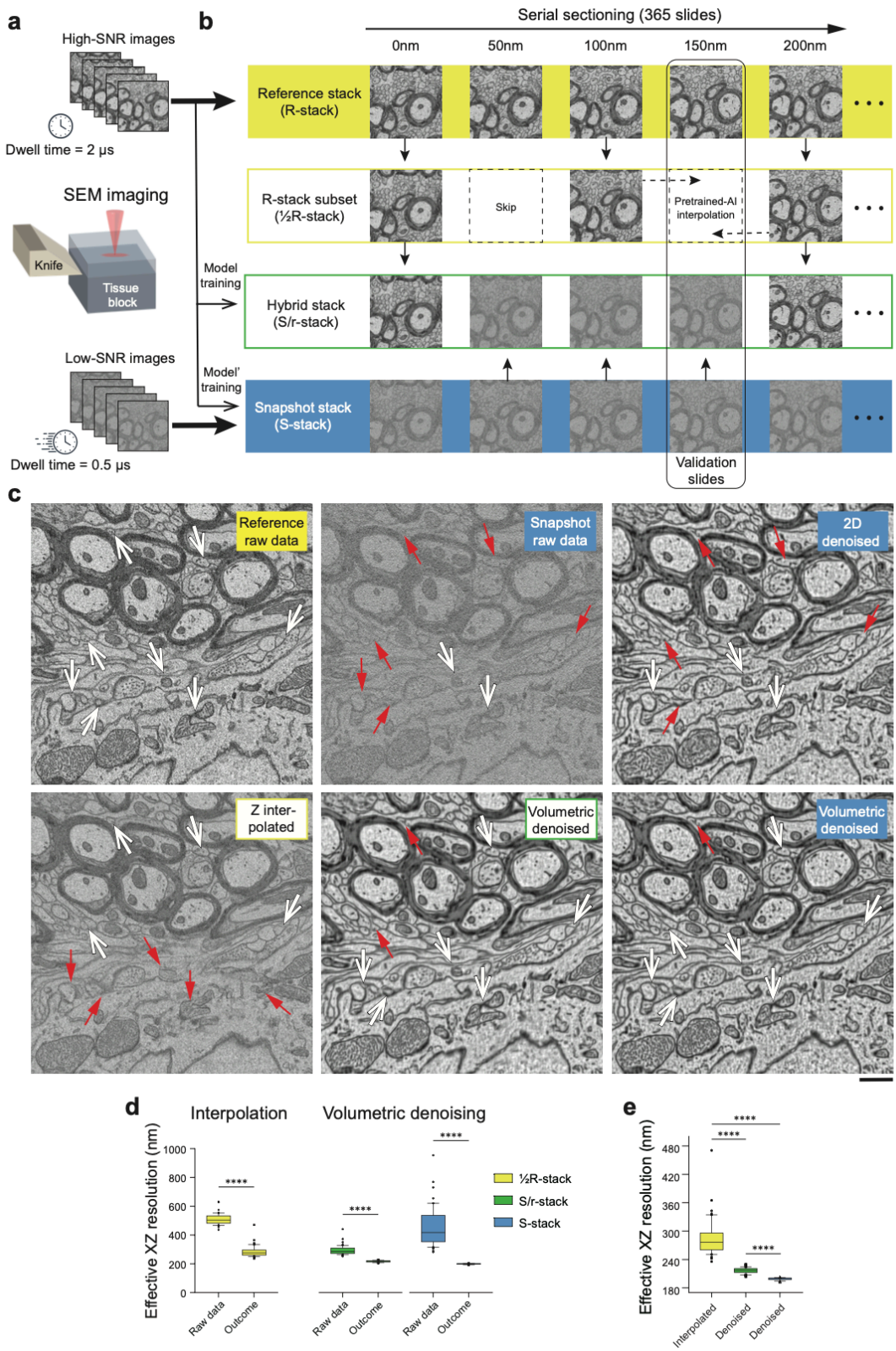

Figure 2: Denoising of image stack using 3D structural context.

a, Schematic illustration of a hybrid image acquisition using SBF-SEM. Repeat block-face scanning produces image pairs of low and high signal-to-noise-ratio (SNR). **b,** Full stack of high-SNR images (R-stack, yellow solid) serves as reference, whereas the

snapshot dataset containing exclusively low-SNR images (S-stack, blue solid) is for the method development. Two synthetic image stacks: For the ½R-stack (yellow box), every other slow-scan-only slice is replaced with an interpolated virtual slice. S/r-stack (green box) is made up of fixed combinations of one reference and three snapshot slices. **c,** Representative image of the same slice in the reference stack (up left) and snapshot dataset (up middle), as well as AI-denoised S-stack only using 2D information (up right). Similarly, the same slice was generated by 3D methods, including AI-based interpolation (bottom left) as well as supervised volumetric (V-) denoising on the S/r-stack (bottom middle) and the S-stack (bottom right). Arrows indicate faithfully preserved structures (white) and artifacts (red). Scale bar, 1000 nm. **d,** Improvement of effective resolution by AI-based image restoration. (Left) By interpolation, the effective resolution for the ½R-stack drops from 510.0 ± 38.1 nm to 284.6 ± 39.2 nm. The V-denoising improves the effective resolution for the S/r-stack (middle) from 293.9 ± 32.6 nm to 216.5 ± 6.2 nm, as well as for the S-stack (right) from 458.0 ± 138.3 nm to 198.9 ± 2.7 nm. **e,** Comparison of the achieved effective resolutions by different 3D approaches from d.

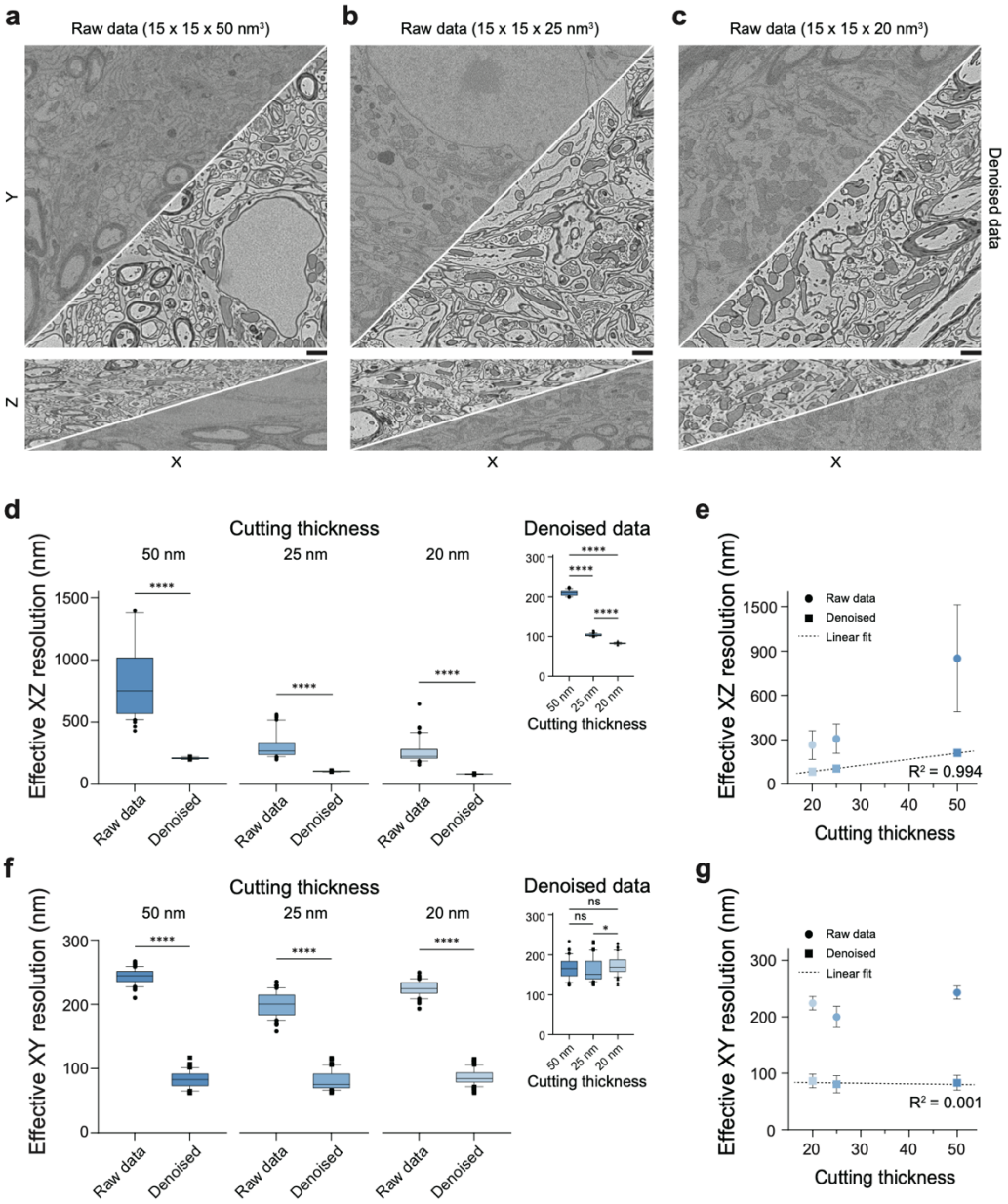

Fig. 3: Improved Z-sampling allowed by AI-based volumetric denoising

a-c, Three SBF-SEM datasets acquired in the snapshot mode (0.5 µs/pixel) but at different cutting thicknesses (a: 50 nm, b: 25 nm, c: 20 nm) for the same sample. Representative raw image and volumetric denoised image by a self-supervised model are shown in both XY (top) and XZ (bottom) views. Scale bar, 1 µm. **d,** Differential improvements in the effective XZ resolution after denoising are observed for datasets of different cutting thicknesses (50-nm: from 850.3 ± 361.8 nm to 209.5 ± 6.6 nm; 25-nm: from 306.5 ± 98.9 nm to 104.1 ± 3.3 nm; 20-nm from 263.6 ± 97.0 nm to 83.6 ± 2.0 nm). Inset: Plot of the achieved XZ resolution from d. against physical cutting

thickness. **e,** Plot of the effective XZ resolution measured from raw datasets (dots) and denoised datasets (squares) as a function of the corresponding cutting thickness. Linear fit: $R^2 = 0.994$. **f,** Similar to d, comparison of the effective resolution in the XY direction (50-nm: from 243.0 ± 11.9 nm to 83.5 ± 13.2 nm; 25-nm: from 200.2 ± 18.8 nm to 80.7 ± 15.1 nm; 20-nm: from 224.5 ± 11.9 nm to 86.6 ± 12.1 nm. Inset: Comparable effective XY resolution after denoising. **g,** Similar to e, plot of the obtained effective XY resolution against the cutting thickness of the datasets.