SmartEM: machine-learning guided electron microscopy

Yaron Meirovitch[1,2,3,*,✉], Core Francisco Park[1,2,*], Lu Mi[4,5,*], Pavel Potocek[6,7,*], Shashata Sawmya[4], Yicong Li[8], Ishaan Singh Chandok[1,2], Thomas L. Athey[4], Neha Karlupia[1,3], Yuelong Wu[1,3], Daniel R. Berger[1,3], Richard Schalek[1,3], Hanspeter Pfister[1,8], Remco Schoenmakers[6], Maurice Peemen[6], Jeff W. Lichtman[1,3,✉], Aravinthan D.T. Samuel[1,2,✉], and Nir Shavit[4,✉]

[1]Center for Brain Science, Harvard University, Cambridge, MA 02138, USA
[2]Department of Physics, Harvard University, Cambridge, MA 02138, USA
[3]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA
[4]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[5]Allen Institute for Brain Science, Seattle, WA, USA
[6]Thermo Fisher Scientific, Eindhoven, the Netherlands
[7]Saarland University, 66123, Saarbrücken, Germany
[8]Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Boston, MA 02134, USA
[*]Equal contribution

Correspondence: *yaron.mr@gmail.com (YM); jeff@mcb.harvard.edu (JL); samuel@physics.harvard.edu (AS) shanir@csail.mit.edu (NS)*

Summary

Connectomics provides essential nanometer-resolution, synapse-level maps of neural circuits to understand brain activity and behavior. However, few researchers have access to the high-throughput electron microscopes necessary to generate enough data for whole circuit or brain reconstruction. To date, machine-learning methods have been used after the collection of images by electron microscopy (EM) to accelerate and improve neuronal segmentation, synapse reconstruction and other data analysis. With the computational improvements in processing EM images, acquiring EM images has now become the rate-limiting step. Here, in order to speed up EM imaging, we integrate machine-learning into real-time image acquisition in a single-beam scanning electron microscope. This SmartEM approach allows an electron microscope to perform intelligent, data-aware imaging of specimens. SmartEM allocates the proper imaging time for each region of interest – scanning all pixels equally rapidly, then re-scanning small subareas more slowly where a higher quality signal is required to achieve accurate segmentability, in significantly less time. We demonstrate that this pipeline achieves a 7-fold acceleration of image acquisition time for connectomics using a commercial single-beam SEM. We apply SmartEM to reconstruct a portion of mouse cortex with the same accuracy as traditional microscopy but in less time.

electron microscopy | connectomics | machine-learning | adaptive scanning

Introduction

Serial-section Electron Microscopy (ssEM) is widely used to reconstruct circuit wiring diagrams in entire brains of small animals like *C. elegans*, *Drosophila*, and zebrafish (White et al., 1986; Witvliet et al., 2021; Xu et al., 2020; Hildebrand et al., 2017) and brain regions in mammals (Morgan et al., 2016; Kasthuri et al., 2015; Bock et al., 2011; Abbott et al., 2020; Lu et al., 2023; Song et al., 2023). Comparing the growing numbers of connectomes of animals with different genetic backgrounds, life experiences, and diseases will illuminate the anatomical nature of learning, memory, and developmental plasticity, the nature of brain evolution, as well as the nature of anatomical abnormalities that cause neuropathology and disease (Kornfeld et al., 2020; Shapson-Coe et al., 2024; Loomba et al., 2022; Karlupia et al., 2023; Bidel et al., 2023). To achieve wide-scale deployment for comparative connectomics, data acquisition and analysis pipelines need to become more widely available (Swanson and Lichtman, 2016). At present, connectome datasets are mostly acquired by the few laboratories and institutions equipped with specialized and expensive high-throughput electron microscopes such as the TEMCA (Transmission Electron Microscopy Camera Array) or the Zeiss 61- or 91-beam scanning electron microscope (SEM) (Bock et al., 2011; Shapson-Coe et al., 2024). Until recently, dataset acquisition had not been a limiting factor in connectomics (Lichtman et al., 2014). A more significant bottleneck had been data analysis – segmenting serial-section electron micrographs to reconstruct the shape and distribution of nerve fibers, identify synapses, and map circuit connectivity. However, recent improvements in machine-learning and image analysis (Beier et al., 2017; Januszewski et al., 2018; Meirovitch et al., 2019; Sheridan et al., 2023) have dramatically sped data analysis, creating a now urgent need for faster image acquisition. The field needs more electron microscopes to deliver datasets as fast as they can now be analyzed. One way to meet this need is to enable widely-available electron microscopes, like more affordable point-scanning SEMs, to collect connectomic datasets.

When using a point-scanning SEM for connectomics, the time budget for image acquisition time is mostly dictated by the dwell time that the electron beam spends on each pixel. In practice, SEM imaging of well-prepared, high-contrast, electron-dense tissue for connectomics usually uses dwell times $\geqslant\sim 1000$ ns/pixel. By comparison, the time spent moving the beam between pixels is negligible; modern SEMs use electrostatic scan generators that deflect the electron beam to any pixel in an image (Mohammed and Abdullah, 2018; Anderson et al., 2013). To accelerate a SEM for connectomics, one must reduce the total dwell time over all pixels, but without losing information needed to accurately determine the wiring diagram.

For connectomics, the salient measure of image accuracy is neuronal segmentation – being able to correctly identify each neuron's membrane boundary and to correctly identify each synapse. In standard SEM, image acquisition is done by specifying a fixed homogeneous dwell time for all pixels. The longer the dwell time, the higher the signal-to-noise per pixel, and the more accurate the segmentation. Thus, there is a fundamental trade-off between SEM imaging time and segmentation accuracy. Previous approaches to improving segmentation accuracy with rapidly acquired images have involved post-acquisition image processing such as de-noising (Minnen et al., 2021) or "super-resolution"/upsampling (Fang et al.,

2021). However, image processing that works entirely after the completion of image acquisition is limited by the amount of original information acquired. No technique unambiguously "creates" information that was not acquired in the first place. Fast, lossy image acquisition can miss critical information that precludes accuracy in any subsequent segmentation.

Our solution to the problem of missing information in an initial rapidly acquired image is to immediately recover information during real-time microscope performance. To do this, we developed a "smart" SEM pipeline that rapidly identifies error-prone regions as well as high-salience regions (such as synapses) in every rapidly acquired image, and immediately and slowly re-scans only these regions. We define error-prone regions as only those that would confer full segmentation accuracy to a composite image, which is built from the initial rapidly acquired image (adequate wherever segmentation is accurate at short dwell time) and fused with long dwell time re-scans (necessary wherever segmentation is error-prone). When error-prone and high-salience regions are relatively few in number and small in size, re-scanning adds little to the total image acquisition time budget while fully restoring segmentation accuracy. We sought an image acquisition pipeline that achieves the accuracy of uniform long dwell time acquisition with nearly the speed of uniform short dwell time acquisition.

We implemented smartness in the pipeline with machine-learning algorithms running within SEM computer hardware. This pipeline, called SmartEM, can be applied in any context where images exhibit high spatial heterogeneity in segmentation accuracy as a function of imaging time – a fundamental characteristic of brain images where nerve fibers and synapses can vary in size and density from region to region. Unavoidable spatial heterogeneity in any specimen is why a smart selection of which regions to collect at short dwell times and which regions to re-scan at long dwell times can achieve full segmentation accuracy but with much less total dwell time. Applied to connectomics, the SmartEM pipeline yields a substantial 7-fold speedup for a widely available point-scanning SEM, allowing the microscope to be used for connectomics solely by implementing our machine-learning algorithms in the GPU-equipped SEM support computer. Spatial heterogeneity characterizes numerous SEM applications, and SmartEM can thus be applied to speed reconstruction of other specimens in biology, in material sciences and in electronic circuit fabrication.

Results

Suitability of adaptive dwell times for connectomics

To establish the rationale for our connectomics pipeline by point-scanning SEM – automatically applying short dwell times to most brain regions that are "easy" to segment and long dwell times to fewer brain regions that are "hard" to segment – we quantitatively tested how spatial heterogeneity in representative mammalian brain images affects segmentation accuracy with different dwell times. To perform these tests, we used a recent high-quality sample comprising 94 sections of mouse visual cortex (Karlupia et al., 2023). We re-imaged these 94 sections at 4 nm pixel resolution using a point-scanning Verios G4 UC SEM from Thermo Fisher Scientific and a range of fixed dwell times from 25 to 1200 ns/pixel.

We note that when these images were originally acquired in a previous study using standard point-scanning SEM, the dwell time was 800 ns/pixel (Karlupia et al., 2023). This dwell time was determined by a "rule-of-thumb" and is close to the 800 – 1000 ns/pixel needed for maximal segmentation accuracy for this dataset (**Figure 1A, 1B**).

Our segmentation algorithm – mapping EM images to membrane predictions (EM2MB) followed by a standard watershed transform – provided an objective assessment of segmentation accuracy of images collected with different dwell times. We adapted EM2MB to SEM images taken with different dwell-times (see **Supplement**). We automatically segmented 256 randomly selected 2000×2000 pixel regions taken from the 94-section sample with different dwell times. Automatic segmentation with ultrafast dwell times (25 ns/pixel) produced frequent merge and split errors compared to automatic segmentation of the same regions with overly slow dwell times (1200 ns/pixel) (**Figure 1A**). As dwell times increased, segmentation errors gradually disappeared.

To quantify segmentation accuracy, we calculated the Variation of Information (VI; Meila (2003)) between each automatically segmented region at each faster dwell time and the segmentation obtained at the slowest dwell time (**Figure 1B**). Segmentation accuracy increased with slower dwell times, and saturated at 800-1000 ns/pixel, consistent with the rule-of-thumb practice in choosing the dwell times for connectomics. With >1000 ns dwell times, segmentation accuracy using EM2MB is saturated. At 25 ns/pixel, acquisition speed is 40× faster than at 1000 ns/pixel, but with lower accuracy.

Brain tissue is typically heterogeneous in the difficulty of segmentation across image regions (**Figure 1C, 1D**). Thus, segmentation accuracy varied substantially from region to region. For slow dwell times (1000 ns), segmentation accuracy was narrowly distributed around small VI, indicating less segmentation errors. For ultrafast dwell times (25 ns), segmentation accuracy was broadly distributed. Some regions exhibited the same low VI with both ultrafast and slow dwell times ("easy" to segment regions). Some regions exhibited drastically higher VI for ultrafast dwell times than slow dwell times ("hard" to segment regions) (**Figure 1C**). For each region, we determined the minimum dwell time to reach the same segmentation accuracy as the slowest dwell time (see **Supplement: Determination of maximal segmentation quality**). We observed a broad distribution of minimum dwell times across pixel regions. Most 2000×2000 pixel regions are accurately segmented with dwell times <150 ns, but a small number (∼25%) required longer dwell times. Minimum dwell times exhibited a broad-tailed distribution from 50 – 1200 ns/pixel (**Figure 1D**).

Challenges in smart microscopy

We sought a SmartEM pipeline to identify and adapt to spatial heterogeneity in the segmentation accuracy of brain tissue for connectomics when imaged at different dwell times. To implement this pipeline with a point-scanning SEM, we needed to solve several challenges. The SEM needs to automatically identify error-prone locations in an initial rapidly acquired brain

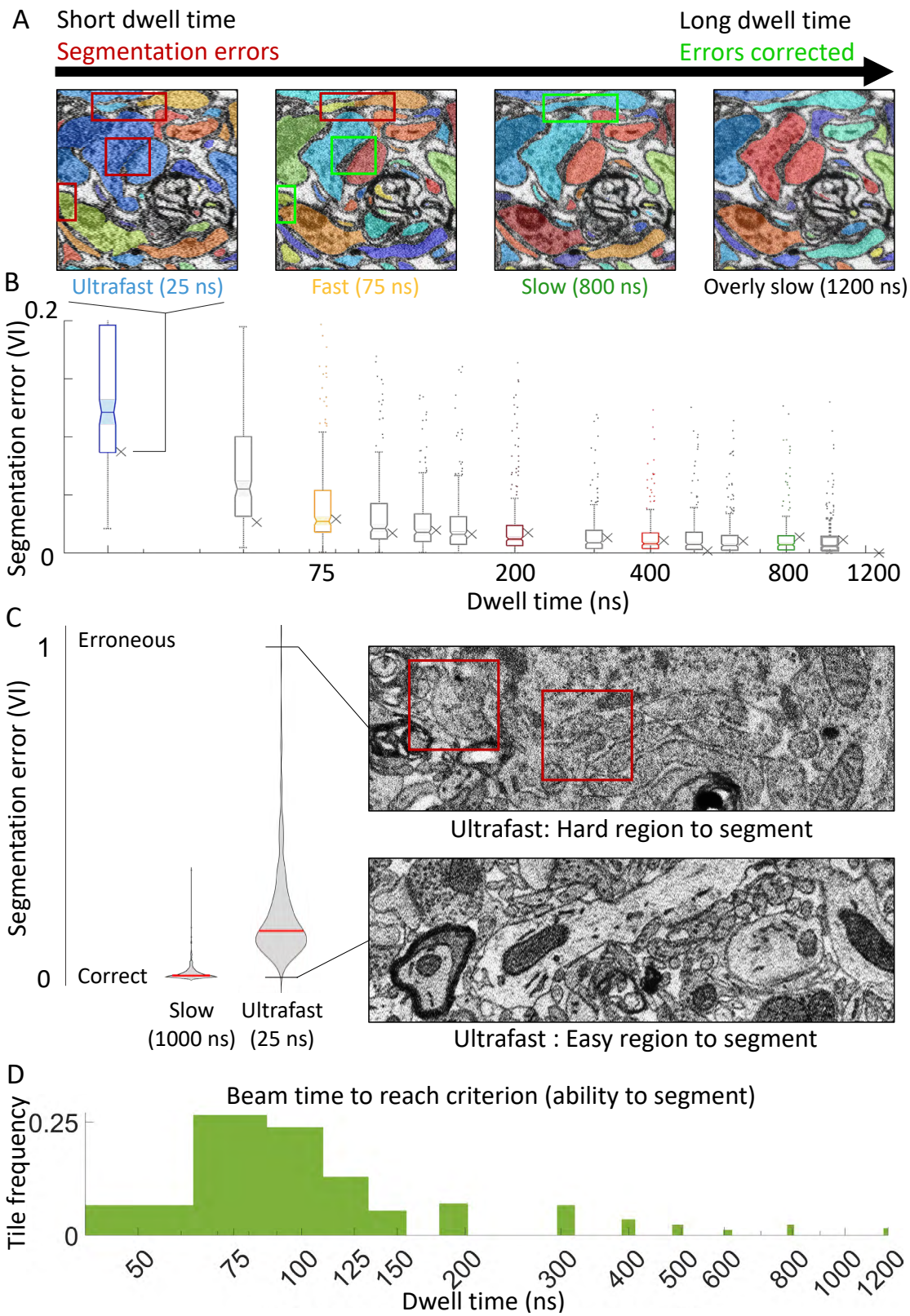

Figure 1. The effect of the beam's dwell time on the ability to segment the EM into neuronal elements. **A.** Scanning the same EM tile with different dwell times. Short dwell time scans result in segmentation errors (red squares) that are resolved by longer scans (green squares). Increasing the dwell time improves the segmentation accuracy of short dwell time images (25 ns/pixel and 75 ns/pixel) but does not improve the segmentation accuracy of sufficiently long dwell time images (800 ns/pixel). **B.** The segmentation quality of the same images used in (A) are represented by x markers, alongside the distribution of segmentation qualities of 256 images (scatter and boxes) for 13 dwell times, from 25 ns to 1000 ns, calculated relative to a reference image taken at 1200 ns/pixel. Segmentation error is quantified by variation of information (y-axis). VI drops rapidly with increased dwell times, saturating with dwell times near 800 ns. Wide distributions indicated by whiskers at each dwell time indicate that some image tiles can be segmented at any dwell time. **C.** Segmentation of neuronal tissue has varying quality due to tissue heterogeneity: taking an image at 25 ns could lead to an image that can be segmented at high quality (bottom image) or low quality (top image), compared to taking the images slowly (at 1000 ns). **D.** The majority of image regions (greens areas add up to 1.00) can be segmented at faster dwell times (75 ns to 125 ns), while some regions require longer dwell times (between 400 ns to 800 ns) to reach the segmentation quality criterion. Thus, adapting dwell time for different regions would save imaging time without reducing segmentation quality.

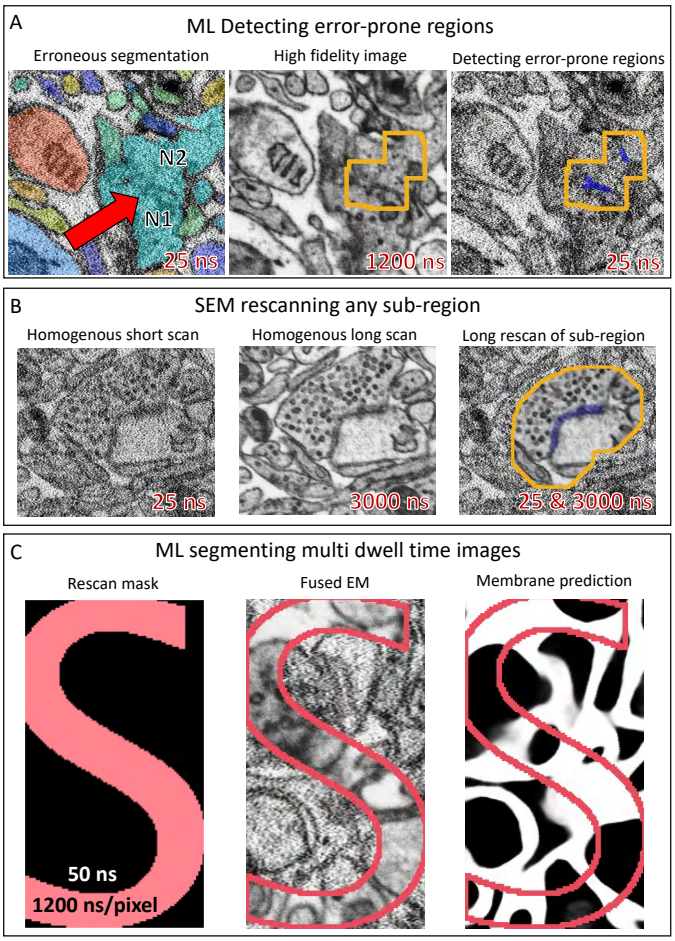

Figure 2. Smart microscope challenges. **A.** An erroneous segmentation of a rapidly acquired image (25 ns/pixel) with a red arrow indicating the location of a merge error between two neurons (N1, N2). Slowly acquiring the same image at 1200 ns/pixel captures the neuronal boundary (middle). The output of the ERRNET neural network that was trained to predict segmentation errors from EM is shown on the right. Blue indicates the location where the network predicts a possible merge or a split error. The yellow outline is a window around the predicted error to provide further context needed for downstream correction. **B.** The SEM readily captures any part of an image at different dwell times, homogeneously at short dwell times (left), homogeneously at long dwell times (middle), or homogeneously at short dwell times with a sub-region taken at long dwell times (right). Here, the yellow outline for the long dwell time sub-region contains a synaptic cleft. **C.** Predicting neuronal borders from fused EM images using FUSED2MB.

image. The SEM needs to immediately re-scan pixel neighborhoods around error-causing locations to guarantee accurate segmentation. After image acquisition, the pipeline needs to accurately segment composite images built from the initial rapidly acquired images fused with re-scanned error-prone regions. We review solutions to these challenges that we incorporate in the smart microscopy pipeline described below.

Detecting error-prone regions by an SEM. To identify error-prone regions in initial rapidly acquired images, we developed a machine learning (ML) algorithm to run on the microscope's support computer. **Figure 2A** shows a rapidly acquired image tile and its segmentation containing a merge error (red arrow in left panel); the slowly acquired image tile that would not produce an error (middle panel); the prediction of a neural network (ERRNET, see below) that identifies error-causing locations in the rapidly acquired image (corresponding to cell membranes associated with the merge error, highlighted in blue in right panel); and the specification of error-prone region to be re-scanned that would remedy segmentation errors in post-processing (yellow outline in middle and right panels). ERRNET operates in real-time within SEM computer hardware that is equipped with a high-performance GPU, and is much faster than initial image acquisition – per pixel processing for a single commodity GPU is <100 ns/pixel; N GPUs operating in parallel require $< 100/N$ ns/pixel. A related idea where EM acquisition is guided based on uncertainty measures estimated by neural network models was described in Shavit et al. (2021).

SEM re-scanning any sub-region. To use the prediction of error-prone regions during real-time SEM operation, we modified the scanning procedure of the microscope to re-scan error-prone regions at slow dwell times right after the fast scan. In addition to re-scanning error-prone regions, neural networks can be trained for data-aware re-scan of additional regions of interest like synaptic clefts for applications in connectomics. **Figure 2B** depicts data-aware re-scan where the microscope is guided to re-take regions around synaptic clefts that are predicted from an initial fast scan image of a section of mammalian cortex. SEM microscopes with electrostatic scan generators are able to conduct efficient and rapid re-scan without wasted time in moving the electron beam (Mohammed and Abdullah, 2018; Anderson et al., 2013). When ERRNET and re-scan software are seamlessly integrated within SEM computer hardware, the total time spent acquiring an image is the total number of pixels $\times$ the short initial dwell time plus the total number of *re-scanned* pixels $\times$ their long dwell time.

Segmentation of multi-dwell time images. After image acquisition, a smart microscopy pipeline generates a complete rapidly acquired image and set of slowly re-scanned regions of each sample. When pixels from the re-scanned regions are substituted into corresponding locations in initially rapidly acquired images, composite images are produced with pixels of multiple dwell times. Previous segmentation algorithms for connectomics have dealt with a single pre-fixed dwell time (Januszewski et al., 2018; Meirovitch et al., 2019; Sheridan et al., 2023) – these algorithms generalize poorly to homogeneous images taken at different dwell times or to heterogeneous images composed of regions taken at different dwell times. The smart microscopy pipeline demands new algorithms to accurately segment composite images where different regions are obtained at different dwell times. We developed a data augmentation training procedure technique for a neural network with a U-Net (Ronneberger et al., 2015) architecture (FUSED2MB) to accurately detect membranes in an image with heterogeneous dwell times as well as if the image was taken with a single uniformly applied dwell time (see **Supplement**). **Figure 2C** shows an example of an image that has multiple dwell times (slow scanning arbitrarily within an S-shaped region surrounded by fast scanning). The predicted membranes by FUSED2MB are unperturbed when crossing between regions taken with different dwell times.

Thus, the challenges in building a smart microscopy pipeline are met by extensively using machine learning in both

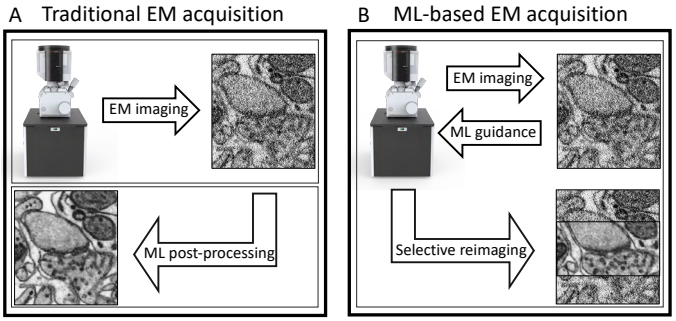

Figure 3. Traditional versus ML-based image acquisition. **A.** Traditional microscopes acquire images without real-time adaptation to the data itself. Acquired images are often processed off-line using machine learning algorithms that might enhance the existing information content (e.g. using super resolution, de-noising and in-painting ML techniques). **B.** The SmartEM pipeline uses dataset-specific machine learning algorithms to guide image acquisition in real-time to increase information content.

guiding image acquisition and image analysis. As illustrated in **Figure 3**, our pipeline differs from previous applications of machine learning in electron microscopy, where image analysis using neural networks was only used to improve image appearance after image acquisition.

The smart microscopy pipeline

We built an integrated smart pipeline that meets the above challenges, **Figure 4A** shows an example of our smart microscopy pipeline run on a small tile from the mouse cortex dataset (Karlupia et al., 2023). The components of the SmartEM are outlined in **Figure 4B** and their design and implementation are described below in detail.

Determining the standard dwell time needed for high accuracy segmentation. The goal of the SmartEM pipeline is to reach the segmentation accuracy of standard SEM with uniform slow scanning, but in much less time. To fairly assess the improvement of SmartEM over standard SEM, we needed first to determine the shortest dwell time for standard imaging that leads to accurate segmentation (800 – 1000 ns/pixel in the example in **Figure 1B**).

We also needed an objective metric for assessing accurate segmentation. In the example shown in **Figure 1B**, the microscope automatically estimates 800 ns as the minimal dwell time needed for accurate segmentation. This estimate was based on the output of automatic membrane prediction by a neural network (Pavarino et al., 2023).

To accomplish this, we trained a neural network called SLOWEM2MB to perform automatic membrane prediction using long dwell time images. We acquired a small and diverse subset of long dwell time images from random locations in a specimen, typically twenty 5×5 μm EM tiles, and performed manual segmentation by an expert to create training data for SLOWEM2MB.

Next, we used SLOWEM2MB to train a separate neural network called EM2MB that was capable of predicting membranes with long or short dwell time images. The single-beam SEM allowed for re-imaging the same regions at different dwell times. The single-beam SEM allowed repeatedly re-imaging the same region at different dwell times. In this way it was possible to guide the microscope to collect a large sample of EM images from different random locations in the specimen, using different dwell times ranging from 25 to 2500 ns/pixel as shown in **Figure 4**. SLOWEM2MB was applied to the long dwell time image at each location to automatically create segmentations that we could use as "ground truth" to train EM2MB to predict segmentations with long or short dwell time images. Both SLOWEM2MB and EM2MB were implemented using a U-net architecture.

SLOWEM2MB and EM2MB calculated the trade-off between pixel dwell time and segmentation accuracy. EM2MB was used to automatically segment all dwell time images (e.g. from 25 to 1000 ns/pixel for the mouse cortex dataset) and compare them to a reference automatic segmentation corresponding to the longest dwell time image (e.g. 1200 ns/pixel image). Thus, it was possible to identify the shortest dwell time for which mean accuracy across tiles was not further improved by longer dwell time imaging. This minimum dwell time was defined by SmartEM as the required dwell time to achieve agreement with ultra-slow dwell time segmentation.

Learning to detect error-causing locations in short dwell time images. To further reduce imaging time we adjusted pixel dwell time based on segmentation accuracy. Most image regions can be segmented with full accuracy after scanning with a short dwell time. Additional dwell time is only selected for those regions that require longer imaging to segment properly. This selection was accomplished via a neural network (ERRNET) that learned what regions required longer dwell time after scanning whole images with short dwell time. ERRNET learns the features of error-causing locations in raw short dwell time images that produce segmentation differences – erroneous merges or splits – in comparison to long dwell time images.

To assemble "ground-truth" to train ERRNET, the microscope first takes a large set of images from random locations in the specimen at multiple dwell times (e.g. from 25 to 1200 ns/pixel). These images are segmented to distinctly label every contiguous neuron cross section. Automatic labeling can be done using membrane probabilities, a seeding procedure, and a standard region-growing algorithm such as watershed (Vincent and Soille, 1991). Segmented images at all dwell times are compared to reference segmented images taken with the longest dwell time (1200 ns/pixel for the mouse cortex dataset in **Figures 1A, 1B**, longer than needed for fully accurate segmentation with SLOWEM2MB). To automatically learn segmentation discrepancies between short and long dwell time images, we developed a method to produce a binary error mask that defines the morphological differences between two segmented images based on the variation of information (VI) clustering metric (Meila, 2003) (See **Supplement** for details). We trained ERRNET to predict error-causing regions in short dwell time image as shown in **Figure 4**. We used the VI metric to detect objects that are morphologically different between segmentations of short and long dwell time images, and then mapped the borders that differ for these objects (described in **Supplement**) (Meila, 2003). We noted that all segmentation errors in short dwell time images can be repaired (i.e. leading to identical segmentation as long dwell time images) by selectively replacing only regions surrounding discrepancy-causing locations in short

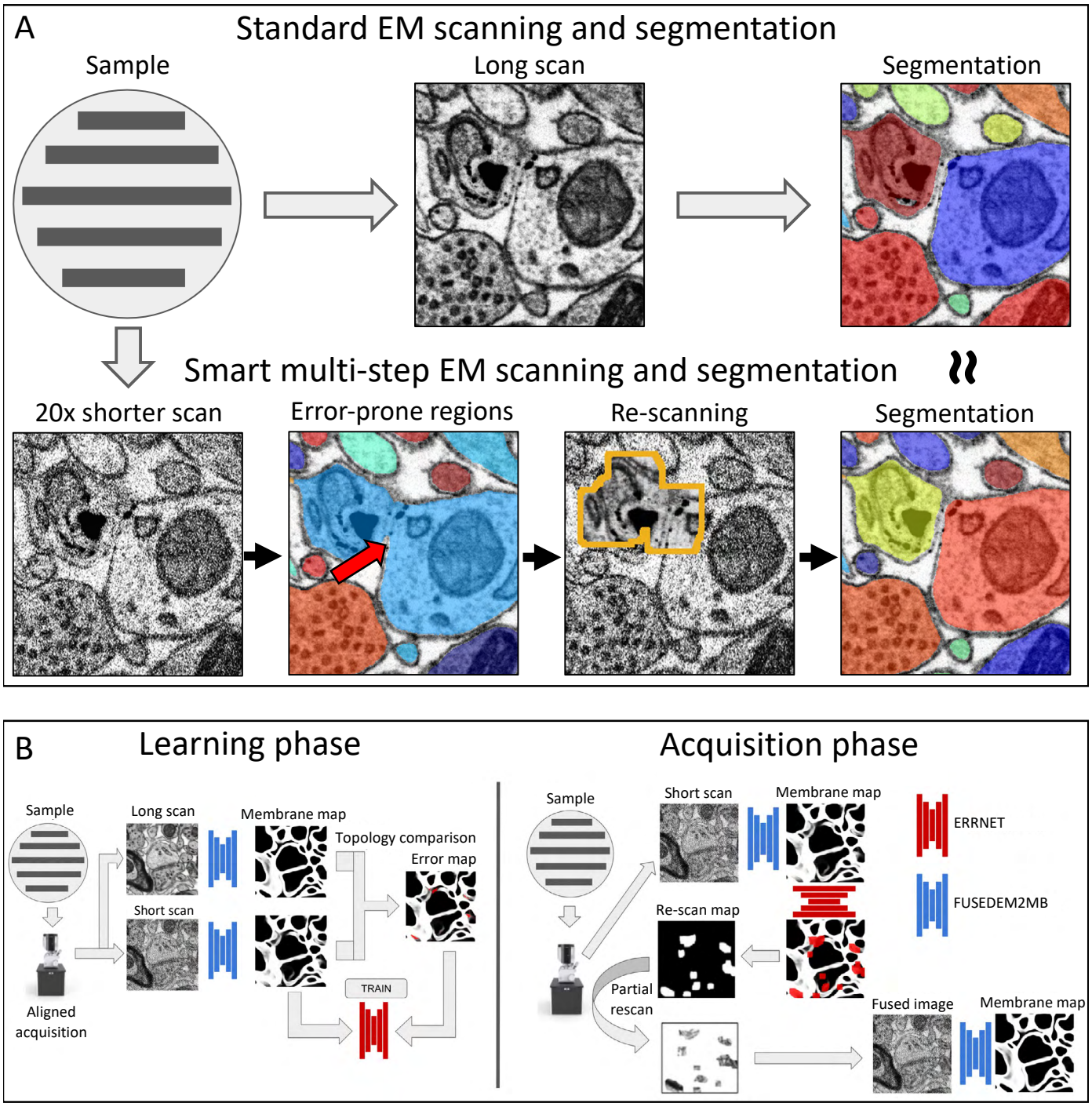

Figure 4. SmartEM pipeline. **A.** The smart multi-step imaging compared to standard imaging. In standard EM, the sample is first scanned with a long dwell time and then segmented (top). In the SmartEM pipeline, the sample is first scanned at a short dwell time, error-prone regions are detected and re-scanned and then segmented. **B.** The learning and acquisition phases of SmartEM. Left: For training, SmartEM requires aligned stacks of high-quality (long scan) images and low-quality (short scan) images. A membrane detector, FUSEDEM2MB (blue), is trained on this dataset to re-produce the high quality results of a membrane detector that runs only on the long scan images. Once FUSEDEM2MB is trained, the membrane predictions between the short and long dwell times is compared (topology comparison) and a binary error map featuring the differences between the two predictions is produced. A second network, ERRNET (red) is trained to predict this error map from the membrane predictions of the short dwell time images. Right: The two trained networks FUSEDEM2MB and ERRNET are used for smart acquisition. First a short scan is performed and the membrane prediction is generated from FUSEDEM2MB. This prediction serves as an input to ERRNET to generate an error map. The error map is processed and used to guide a long dwell time re-scan. For verification, the composite image of the two dwell times is segmented.

dwell time images with corresponding regions taken from long dwell time images.

Detecting error-prone regions in short dwell time images, re-scanning, and producing fused images. In real-time operation, the SEM microscope must take an initial rapidly acquired image, execute ERRNET to detect error-prone locations, define a re-scan mask by padding error-prone locations to capture enough context to improve segmentation accuracy, and then immediately re-scan all error-prone regions using slow dwell times.

Unifying and enhancing images. The final output of the pipeline are images where some pixels are captured with slower dwell times than others. Although the raw appearance of rapidly captured regions (high pixel noise) and slowly captured regions (low pixel noise) does not degrade segmentation accuracy, it does create visually unappealing contrasts (**Figure 3B**). To standardize the SmartEM image for human interpretation, we also built an algorithm that translates the style of the SmartEM images to look like standard EM images with homogeneous dwell times. A similar technique was described in Shavit et al. (2021, 2023). This stylized output does not supplant, but is saved in addition to, the raw composite SmartEM images. We note that stylized images often retain the correct details of the ultrastructure seen in homogeneous long dwell time images (**Figure S8**).

Technique Evaluation

We developed our SmartEM pipeline to expedite connectomic reconstruction on our widely available point-scanning SEM, the Verios G4 UC from Thermo Fisher Scientific. Here, we quantitatively estimate the practical improvement in quality and speed offered by this pipeline for connectomics.

Improving accuracy. One premise of the smart microscopy pipeline is that automatically detecting error-prone regions and replacing them with slower dwell time pixels will reduce segmentation errors. To attempt to validate this premise, we compared the accuracy of a segmentation pipeline trained to deal with short dwell time images (FASTEM2MB at 100 ns/pixel) to a SmartEM pipeline trained to deal with composite images made from short and long dwell times (FUSEDEM2MB at 100 ns/pixel and 2500 ns/pixel). The performance of these networks was compared to the standard segmentation pipeline with slow image acquisition (SLOWEM2MB at 2500 ns/pixel). For fair comparison, we used the same long dwell time for the re-scanning in the smartEM pipeline and for the uniform scan in the standard pipeline. We found that using these dwell times, SmartEM pipeline is $\sim$5$\times$ faster than the standard segmentation pipeline with slow image acquisition and $\sim$2-3$\times$ more accurate (based on VI) than the standard pipeline operating quickly (100 ns/pixel) (**Figure S2**). Thus, fusing long dwell time pixels into a rapidly acquired image can improve segmentation accuracy.

Another premise of the SmartEM pipeline is that given the additional time spent in re-scanning part of an image, the improvement in segmentation accuracy is superior to the improvement that would be obtained by giving the same amount of extra time to a standard pipeline that somewhat more slowly acquires all pixels at the same dwell time. To attempt to validate this premise, we used a FastEM pipeline by choosing competitively fast settings for the standard pipeline, with pixels taken homogeneously at 75 ns. We compared the performance of FastEM with a SmartEM pipeline tuned to take the same average time when combining both the initial scan and the smart re-scan. The initial SmartEM scan dwell time was set to 25 ns, the re-scan dwell time to 200 ns, and a portion of the 12.5% most "error susceptible" regions were adaptively selected per tile for re-scan, so as to provide an exact average of 75 ns/pixel. We compared the variation of information of 64 segmented 2048$\times$2048 pixel image tiles of fastEM and SmartEM to a reference slowEM and found that the SmartEM had less error (signed-rank Wilcoxon test; p<0.05, and p<0.025 for N=38 tiles devoid of cell bodies).

Estimating speed-up. We considered two scenarios for the large-scale collection of a connectome dataset. The first involves a fixed imaging time budget to acquire a selected data volume at the selected pixel resolution. Here, the task is to intelligently allocate the imaging time to optimize segmentation accuracy. We note that this optimization is not feasible with a standard EM pipeline that would fix the homogeneous dwell time to fill the time budget. The second scenario involves setting the pipeline quality according to the quality of a standard EM imaging pipeline. Here, the task is to determine SmartEM parameters that maintain this quality while minimizing the required imaging time per volume. Below we analyze both scenarios.

Scenario 1: Optimized accuracy with fixed imaging time budget We fix the total imaging time budget for a given specimen. From this requirement the pixel dwell time is determined after subtracting overhead factors (such as image focusing, astigmatism correction, and mechanical stage movement) from the total budget. For example, the user might need to image a given specimen – 100$\times$100$\times$100 μm tissue, cut in 30 nm thick sections, imaged at 4 nm spatial resolution – within 5 days of continuous EM operation. These constraints determine the average dwell time per pixel

$$\frac{(5 \cdot 24 \cdot 3600 \text{ sec})(4^2 \cdot 30 \text{ nm}^3)}{(100\mu\text{m})^3} = 207.36 \text{ ns}.$$

For a standard EM pipeline, 207.36 ns becomes the homogeneous pixel dwell time. For the SmartEM pipeline, the initial scan and re-scan of all error-prone regions should sum to an average of 207.36 ns/pixel. This average dwell time, which we call *effective dwell time*, can be achieved with different combinations of initial fast dwell time, re-scan slow dwell time, and percentage of re-scanned pixels:

$$T_{\text{effective}} = T_{\text{initial}} + \alpha \cdot T_{\text{re-scan}}$$

where T represents dwell times.

For example, an effective average dwell time of 207.6 ns is achieved with an initial dwell time of $T_{\text{initial}} = 100$ ns, re-scan rate of $\alpha = 5\%$, and re-scan dwell time of $T_{\text{initial}} = (207.36 - 100)/0.05 = 2147.2$ ns. These parameter settings correspond to a specific segmentation accuracy (VI) relative to the reference homogeneous long scan image. SmartEM considers a grid of

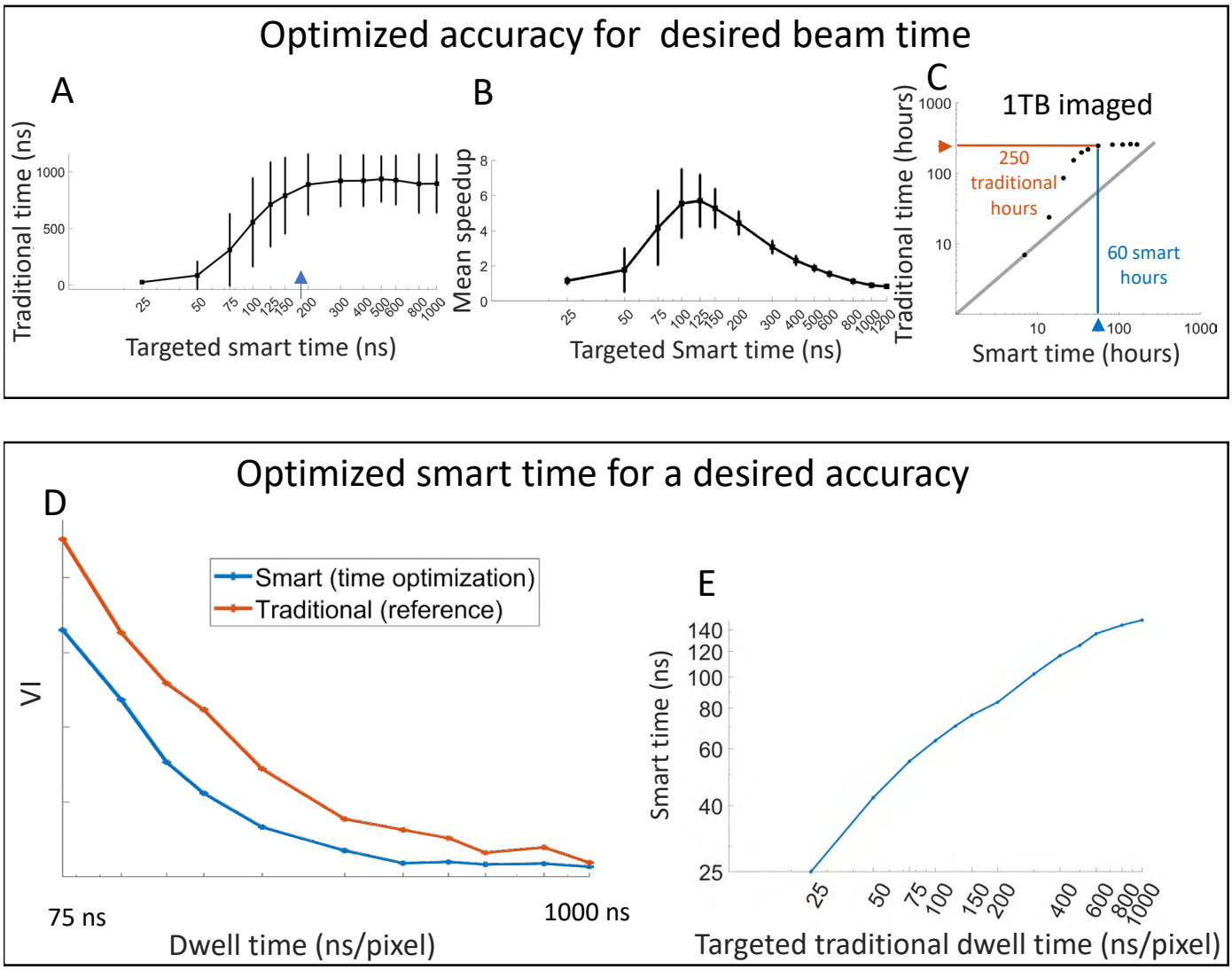

Figure 5. SmartEM acquisition time. In the first imaging scenario **A-C**, the imaging time is constrained by a time budget, which for a fixed volume and pixel resolution, provides the average dwell time (beam time). The task is to intelligently determine the SmartEM parameters that optimize segmentation accuracy. **A** For every targeted smart time (effective dwell time) the smartEM parameters that optimize segmentation accuracy are determined, including the standard (homogeneous) dwell time achieving that segmentation accuracy (y-axis). The longer the smart time is, the longer homogenous time is needed to achieve the same accuracy, with an asymptote around smart dwell time of 200 ns/pixel, equivalent to a homogeneous dwell time of about 800 ns/pixel. Error bars represent 1 s.d. **B** The speed up of the curve in A. The maximal speed up is achieved around the inflection point in A, around 125 ns/pixel - longer smart imaging up to 200 ns/pixel will still improve segmentation accuracy but with smaller speed up. **C** The data in A and B is shown for a fixed volume of 1 TB at 4nm per pixel and a slice thickness of 30 nm. **D** The variation of information of SmartEM compared to slowEM is calculated for each effective dwell time by optimizing the pipeline's parameters and the average VI across tiles is depicted. This allow calculating the two dwell times in the SmartEM (blue) versus standard (red) settings that produce the same accuracy (on average and per tile). **E** In the second imaging scenario, the quality of the EM is set in advance in terms of a desired dwell time of a standard pipeline, and the task is to find smart EM parameters that would provide that quality in a minimal amount of imaging time per volume. Maximal information for segmentation is achieved with around 140 ns/pixel for SmartEM and with around 800-1000 ns/pixel for standard EM.

parameter settings and calculates the T_{initial}, $T_{\text{re-scan}}$ and α settings that produce maximal accuracy (minimal VI) compared to the segmentation of reference tiles, while guaranteeing the effective dwell time (see **Supplement**).

Figure 5A presents the results of parameter optimization for different effective dwell times (smart imaging time) and image tiles. This optimization links any effective dwell time (achieved by optimizing the VI for different T_{initial}, $T_{\text{re-scan}}$) to an accuracy-equivalent standard homogeneous dwell time. For example, an effective dwell time of 200 ns (blue arrow) already attains the maximal quality using a specific set of initial, re-scan dwell times, and re-scan rates that are determined per tile. This quality is comparable to standard homogeneous scan at 800 ns/pixel.

Figure 5B depicts the time saved by SmartEM compared to standard microscopy. For the mouse cortex dataset, the maximal saving compared to standard EM is achieved when smart EM is used at an effective dwell time of $\sim$125 ns/pixel, which corresponds to the inflection point in **Figure 5A**, leads to an accuracy akin to $\sim$725 ns/pixel by the standard pipeline. This effective dwell time produces images with nearly maximal possible segmentation accuracy (**Figure 1**). **Figure 5C** estimates the time to replicate the accuracy of SmartEM using standard microscopy using 1TB of mouse cortex (where 1 Byte corresponds to 1 pixel at $4\times4\times30$ nm^3). The SmartEM microscope running for 60 hours of continuous imaging achieves the same

quality as a standard pipeline running for 250 hours.

Scenario 2: Minimizing imaging time with fixed image quality In the second scenario a certain volume needs to be segmented, and SmartEM is asked to minimize imaging cost. Imaging time is not determined in advance, but the quality of the smart EM has to meet a quality standard. SmartEM needs to acquire the volume in a way that leads to comparable segmentation to standard EM but in significantly less time.

First, the operator determines the dwell time that is needed to achieve a specific quality with standard homogeneous scanning. This dwell time can be obtained from the SmartEM pipeline estimate of a minimum homogeneous dwell time (**Figure 1**). Once the image quality is effectively determined by selecting a reference dwell time for uniform scanning, SmartEM needs to acquire the volume in a way that leads to comparable segmentation accuracy, but in considerably shorter time.

To analyze the expected imaging time of SmartEM in the mouse cortex dataset, we first segmented images taken at homogeneous dwell times from 25 to 1200 ns uniform dwell times from the same areas. We did the same for each image and each effective dwell time, where each effective dwell time is derived from the maximally accurate parameter set of initial and re-scan dwell times and re-scan rate (**Figure 5D**). In all experiments we used the same error detector (ERRNET) and the same neural network model to predict membrane from composite images (FUSEDEM2MB). To match each standard homogeneous dwell time to an effective smart dwell time, we calculated the shortest smart dwell time that produces segmentation that is statistically indistinguishable from the standard dwell time across tiles (see **Supplement**). **Figure 5E** depicts the relation between the targeted standard dwell time and the smart time that yields the same accuracy. The highest possible quality of standard EM at 1000 ns/pixel (see **Figure 1**) is attained by a smart effective dwell time of $\sim$140 ns/pixel. This $\sim$7.1$\times$ speed-up from standard to SmartEM is achieved by selecting the percentage of re-scanned pixels in each image tile, and letting ERRNET determine re-scan locations.

In **Figure S9** we tested the speed-up achieved by SmartEM when re-scan rates are fixed in advance and only one pair of initial and re-scan dwell times are used for imaging. These constraints allow direct comparison of the initial and re-scan dwell times that optimize imaging time. As effective dwell time increases, the time spent on re-scan also increases – the longest dwell time for re-scan, with the equivalent segmentation accuracy as uniform dwell time (at 800 ns), is 1000 ns (initial dwell time at 200 ns and effective dwell time at 300 ns). Efficiently correcting errors in comparison to standard imaging with slower dwell times also requires slower re-scan rates. Maximally slow re-scan dwell times can be needed even when producing sub-optimal segmentation (i.e. faster than 800-1000 ns of homogenous dwell time). A re-scan dwell time of 2-3$\times$ longer than the homogeneous dwell time can yield optimal speed up. As the effective dwell time is allowed to exceed 500 ns and come closer to the homogeneous dwell time, SmartEM no longer requires a specific value for the initial scan dwell time (blue curve in **Figure S9**) which becomes an arbitrary choice while the re-scan rate is maximal and only the percentage of re-scan is increased to correct remaining errors.

Image acquisition with widely available point-scanning SEM is now a limiting factor in connectomics. This evaluation indicates that the SmartEM pipeline can yield >7$\times$ speed up compared to standard image acquisition with a point-scanning SEM without compromising quality and, at standard fast acquisition (75ns-200ns), smartEM offers better quality.

Imaging mouse cortex with SmartEM

Figure 6 shows the outcome of SmartEM. A volume of size $60\times68\times3$ μm^3 (**Figure 6A**) and a section size 205×180 μm^2 was imaged at 4 nm pixel resolution. For volume acquisition, we used an initial dwell time of 75 ns/pixel, re-scan of 800 ns/pixel, and re-scan rate of 10% providing an effective dwell time of

$$\text{Effective dwell time} = 75 + 0.1 \cdot 800 = 155\,\text{ns/pixel}.$$

This average dwell time for SmartEM corresponds to a standard dwell time of $\sim$1000 ns for traditional microscopy. To test the pipeline on larger sections, we acquired a 205×180 μm^2 composed of 30×30 individual tiles with the same pixel resolution. For the SmartEM parameters, we used an initial dwell time of 75 ns/pixel and a rescan of 600 ns/pixel and a rescan rate of 10% providing an effective dwell time of

$$\text{Effective dwell time} = 75 + 0.1 \cdot 600 = 135\,\text{ns/pixel}.$$

As mentioned above, this effective dwell time corresponds to the maximal possible speed up of SmartEM for this dataset, producing images with segmentation quality akin to standard EM at $\sim$1000 ns/pixel. We depict the segmentation of pipeline outputs in **Figure 6B, 6C, 6D** using segmentation code that was deployed on the microscope's support computer using existing tools (Pavarino et al., 2023). This 2-dimensional segmentation can be used as input to a 3D-dimensional agglomeration algorithm (Karlupia et al., 2023) to produce high quality 3-dimensional neuron reconstruction.

We also assessed the ability to detect synapses on short dwell time images (from 25 ns to 1000 ns) and applied this detection to the above initial scan of 75 ns/pixel with excellent results that are comparable to slow scan imaging as shown in **Figure 6E, 6F, S7**. In **Figure 6G, 6H** we show the ability of SmartEM to detect and exclude regions of no interest, where cytoplasm far from membrane is detected from initial scan, allowing SmartEM to force the skipping of the long dwell time scanning from these regions. In **Figure 6I, 6J, S8** we demonstrate the ability to translate the fused images to a uniform looking EM tiles with quality akin to long dwell time imaging.

Neuronal reconstruction of mouse cortex using SmartEM

We tested SmartEM in application to connectomics. Connectomics requires accurate agglomeration of 2D cross-sections

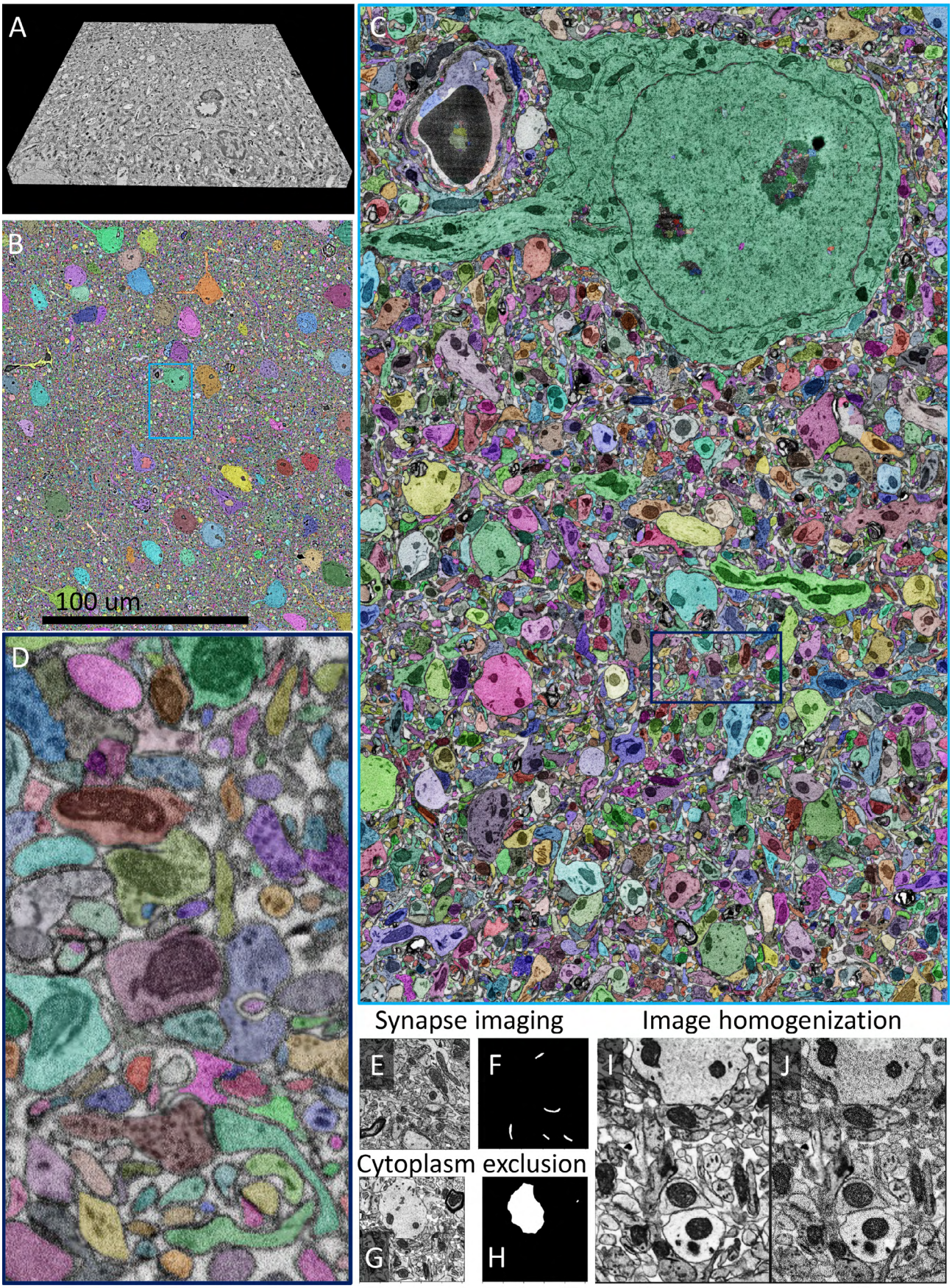

Figure 6. Segmentation of a mouse cortex volume using SmartEM. **A.** Stitched and aligned SmartEM volume of size $60 \times 68 \times 3\ \mu m^3$. **B.** Segmentation of SmartEM volume using FUSEDEM2MB and watershed transform. **C.** Location of the highlighted region in B with respect to the total volume. **D** Detailed depiction of segmentation in the boxed region in B (rotated). **E,F.** Automatic detection of synapses from short dwell time images. **G,H.** Automatic detection of regions to be excluded from short dwell time images. **I,J.** An Image (I) stylized from a composite dwell times image (J) to appear akin to homogeneous dwell times.

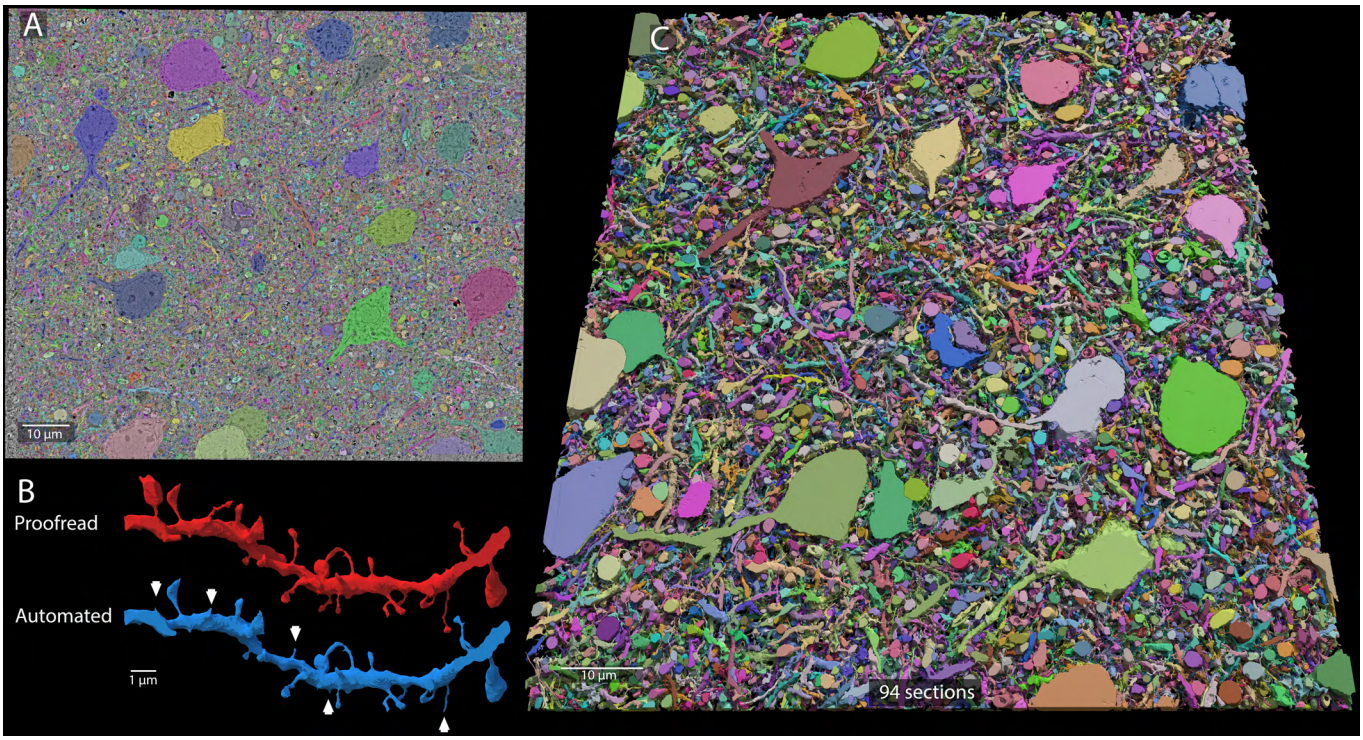

Figure 7. Neuronal reconstruction of a mouse cortex SmartEM volume taken at an average time of 99 ns/pixel. **A.** A section showing overlay of fused EM and an automated neuronal reconstruction, long and short dwell time pixels at 800 ns/pixel ($T_{\text{re-scan}}$) and 75 ns/pixel (T_{initial}), respectively, and a rescan rate of 3% (α). **B.** A dendrite reconstruction proofread by an expert (red) achieved by manually itemizing and reconstructing all dendritic spines from the fusedEM image stack. An automated reconstruction (blue) achieves a high reconstruction rate of the dendritic spines. Arrowheads indicate split errors. **C.** A rendering of the automated 3D reconstruction of all sections in the dataset. The high quality of automated reconstruction has sparse merge errors common to current segmentation algorithms.

(see **Figure 7A**) into 3D reconstruction of neuron volumes and synapses (see **Figure 7C**). We asked whether the quality of aligned SmartEM fused images supports automated reconstruction and proofreading with comparable performance to traditional imaging. We used competitive SmartEM parameters to acquire a mouse cortex volume with average dwell time of 99 ns/pixels (initial long dwell time: 75 ns/pixel; rescan long dwell time: 800 ns/pixel; rescan rate 3%) and a lightweight 3D segmentation algorithm (see **S**upplement: Segmentation and neuronal reconstruction). We assessed the quality of resulting SmartEM image volume with automated reconstruction of fine processes and expert manual annotation (**Figure 7B**).

Reconstruction of dendritic spines. Connectomes can contain "split" errors (fragmenting the volume of one cell) or "merge" errors (joining the volume of two cells). To benchmark SmartEM performance, we studied a challenging and significant problem: split errors in the 3D reconstruction of dendritic spines. Spines are the fine processes that protrude from dendrites and contain synapses. We used automated 3D agglomeration to reconstruct all neurons inside the cortex volume (**Figure 7C**, see **Supplement**). We randomly selected three dendrites (see **Figure 7B**). We counted spines that were fully automatically reconstructed without split errors and spines with split errors. Expert human annotators verified every correct reconstruction and verified that every split error was correctable with proofreading. The percentage of correct spines was approximately 58%, 53% and 75% in the three dendrites. The combined percentage of correct spines was 65%, comparable to the rate of correct spine capture in recent automated reconstruction of human cortex (67%) (Shapson-Coe et al., 2024).

Discussion

The future and flexibility of SmartEM

Data analysis for connectomics is rapidly becoming faster, easier, and cheaper thanks to rapid improvements in machine-learning and the broadening availability of cloud-based tools and computational power. Data acquisition speed is now becoming a bottleneck, rate-limited by the availability and speed of microscope hardware. High-throughput electron microscopes, like the Zeiss multibeam SEM, are not commonly available. This SmartEM pipeline – because it is entirely implemented in accessory computer hardware – can make existing, widely available point-scanning SEMs usable for connectomics with modest cost and modification.

The implementation of the three tools of the SmartEM pipeline are designed so they can be altered depending on use case from user to user or preparation to preparation.

Tool 1 allows an SEM to identify error-prone regions in any rapidly acquired image, but this concept can be implemented with different underlying component algorithms. As described above, **Tool 1** is built by training the ERRNET neural network to detect error-prone regions on the basis of segmentation differences that arise with fast and slow dwell times. Training the ERRNET network allows a choice about what segmentation algorithm to use to train the network. We used our recently devel-

oped two-dimensional segmentation algorithm (Pavarino et al., 2023; Karlupia et al., 2023), but different laboratories will likely have their own preferred segmentation algorithms. ERRNET can be trained with any reliable segmentation algorithm. Because ERRNET is trained before image acquisition, the speed or performance of the segmentation algorithm used to train ERRNET has no effect on pipeline performance. **Tool 2** allows an SEM to perform the slow re-scan of any region within an initially rapidly acquired image in real-time during microscope operation. This slow scan can be done with any point-scanning SEM with electrostatic scan generators that deflect the electron beam to any pixel in an image much faster than the fastest dwell time per pixel (>25 ns) (Mohammed and Abdullah, 2018; Anderson et al., 2013). Electrostatic scan generators are common to modern SEM systems built by most manufacturers. **Tool 3** that performs segmentation of multi-dwell time images is used off-line after image acquisition. The method that we implemented to train Fused2MB can be extended to other segmentation algorithms that work with fixed dwell times. Users could adapt their own segmentation algorithms to work with the multi-dwell time images that emerge from the SmartEM pipeline.

Diverse use cases for SmartEM

The underlying concept of SmartEM with a point-scanning SEM can improve the efficiency and accuracy of image acquisition in any context where it makes sense to adapt the time spent on different regions, much like the human eye, which rapidly captures most of a visual scene with low-resolution (non-foveal) imaging and dwells on selected parts of the visual image to remove ambiguity with high-resolution (foveal) imaging (Thorpe et al., 1996). Point-scanning SEM is used in materials science and manufacturing to assess samples that vary in the spatial density of information content. Any application where regions of high information content can be predicted (but not accurately reconstructed) with an initial rapidly acquired image can benefit from immediate re-scan of those regions, guided by our SmartEM approach. Imaging approaches that take advantage of electron beam sensitive materials, such as cryo-EM would benefit from the selective re-scanning of SmartEM. The objects of interest that are sparsely distributed in the specimen, such as a specific mixture of molecules, will be identified from the rapid initial scan and slowly re-scanned.

Here, we focused on re-scanning for connectomics to capture information in error-prone regions with respect to neuronal segmentation. But re-scanning could also be used to capture information that is salient in other ways. As we showed, we can also perform re-scanning to selectively capture high-quality images of every chemical synapse in a connectome, thereby providing high-quality morphological reconstructions of salient structures in an image volume in addition to resolving the problem of error-prone regions, while still providing substantial pipeline speedup. SmartEM can be adapted to other applications in cell biology or pathology by recognizing and re-scanning other sparse cellular structures of interest (e.g., mitochondria and other organelles).

The SmartEM pipeline can not only be "taught" to capture the most salient features of an image, but can also be used to neglect regions without interest. In most connectomics of larger brains, nearly all objects in the field of view will be neural structures. But in small invertebrates, neural tissue might constitute only a small part of the field of view. The *C. elegans* nerve ring (brain) is <10% of the total volume of the body, and wraps around the pharynx. Any two-dimensional brain section of the *C. elegans* nervous system will also include substantial non-neural tissue. To date, connectomic datasets have been acquired by carefully designating the region-of-interest for each image. The SmartEM pipeline may simplify and speed image acquisition by allowing the microscope to spend the time budget for each image section on neurons instead of non-neuronal tissue without needing the user to laboriously specify each region of interest.

Adaptability of SmartEM for other microscopes and other applications

Tape-based serial-section sample collection, where specimens are stored permanently and can be re-imaged at any time, is suited to SmartEM because any information that is lost during imaging can be recovered. When specimens are imaged for the purpose of connectomics, the SmartEM pipeline might gloss over features that might eventually be of interest to other scientists for other applications (e.g., cell biology). Because serial-sections stored on tape can be safely archived for years, they can be revisited at any time.

Instead of collecting serial sections on tape, one can use block face imaging with serial tissue removal. One block face approach, Focused Ion Beam SEM (FIB-SEM), has distinct advantages over tape-based serial-section sample collection, including thinner tissue layers (4-8 nm) and better preservation of image alignment (Knott et al., 2008). The principal disadvantage of FIB-SEM has been the slow pace of traditional point-scanning SEM with >1000 ns dwell times. This can be problematic when the microscope is used to collect extremely large specimens, and must be continuously operational for days or weeks without technical glitch. However, a FIB-SEM that implements the SmartEM pipeline would be able to operate much faster, increasing the likelihood of capturing an entire specimen in single long runs. SmartEM is expected to provide greater speed up on block face imaging because the imaging component is a larger part of the entire acquisition pipeline compared to serial-section SEM. Similar benefits will be obtained with other block face imaging approaches such as Serial Block Face SEM (SBF-SEM) where a diamond knife slices the specimen (Denk and Horstmann, 2004). The downside of block face approaches, whether with traditional imaging or the SmartEM pipeline, is that each section is destroyed by ablation after being imaged, forbidding revisiting the sample to capture any information that was inadvertently lost.

Improvements for SmartEM

The performance of this software pipeline that runs in real-time during microscope operation should improve further as machine learning algorithms perform segmentation of rapidly acquired images more accurately, a trend that can be expected as more imaging is performed to generate training data for neu-

ral networks and as neural networks themselves improve over time. We expect gradual improvement in how rapid the initial rapid image acquisition can be, and gradual improvement in how many error-prone regions need to be re-scanned. A further order-of-magnitude improvement in the SmartEM pipeline may make point-scanning SEM systems comparably fast as more expensive multibeam systems. We note that current multibeam SEM systems cannot be sped up with this SmartEM strategy, because their multiple beams are coordinated and cannot be independently controlled, a fundamental requirement of this approach.

Summary

All components needed to implement the SmartEM pipeline on the ThermoFisher Verios G4 UC will be provided as open source software. The basic conceptual workflow o f the SmartEM pipeline is adaptable to other microscope platforms.

Code Availability

Machine learning software and all models will be made available upon publication on a public repository and are currently available on request.

Acknowledgements

Research reported in this paper was supported by the NIH BRAIN Initiative under award number 1U01NS132158 and by NIH grants 5U24NS109102 and U01 NS108637. L.M.'s work was supported in part by a fellowship from MathWorks. T.A. is supported by the MIT-Novo Nordisk Artificial Intelligence Postdoctoral Fellows Program.

Declaration of Interests

P.P., M.P. and R.S. are employees of Thermo Fisher Scientific.

References

Abbott, L.F., Bock, D.D., Callaway, E.M., Denk, W., Dulac, C., Fairhall, A.L., Fiete, I., Harris, K.M., Helmstaedter, M., Jain, V., et al. (2020). The mind of a mouse. Cell *182*, 1372–1376.

Anderson, H.S., Ilic-Helms, J., Rohrer, B., Wheeler, J., and Larson, K. (2013). Sparse imaging for fast electron microscopy. In Computational Imaging XI (International Society for Optics and Photonics), vol. 8657, p. 86570C.

Bailoni, A., Pape, C., Hütsch, N., Wolf, S., Beier, T., Kreshuk, A., and Hamprecht, F.A. (2022). Gasp, a generalized framework for agglomerative clustering of signed graphs and its application to instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11645–11655.

Beier, T., Pape, C., Rahaman, N., Prange, T., Berg, S., Bock, D.D., Cardona, A., Knott, G.W., Plaza, S.M., Scheffer, L.K., et al. (2017). Multicut brings automated neurite segmentation closer to human performance. Nature methods *14*, 101–102.

Bidel, F., Meirovitch, Y., Schalek, R.L., Lu, X., Pavarino, E.C., Yang, F., Peleg, A., Wu, Y., Shomrat, T., Berger, D.R., et al. (2023). Connectomics of the *Octopus vulgaris* vertical lobe provides insight into conserved and novel principles of a memory acquisition network. eLife *12*.

Bock, D.D., Lee, W.C.A., Kerlin, A.M., Andermann, M.L., Hood, G., Wetzel, A.W., Yurgenson, S., Soucy, E.R., Kim, H.S., and Reid, R.C. (2011). Network anatomy and in vivo physiology of visual cortical neurons. Nature *471*, 177–182.

Denk, W., and Horstmann, H. (2004). Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. PLoS biology *2*, e329.

Fang, L., Monroe, F., Novak, S.W., Kirk, L., Schiavon, C.R., Yu, S.B., Zhang, T., Wu, M., Kastner, K., Latif, A.A., et al. (2021). Deep learning-based point-scanning super-resolution imaging. Nature Methods *18*, 406–416.

Hayworth, K.J., Morgan, J.L., Schalek, R., Berger, D.R., Hildebrand, D.G.C., and Lichtman, J.W. (2014). Imaging atum ultrathin section libraries with wafermapper: a multi-scale approach to em reconstruction of neural circuits. Frontiers in Neural Circuits *8*, 68–68.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778.

Hildebrand, D.G.C., Cicconet, M., Torres, R.M., Choi, W., Quan, T.M., Moon, J., Wetzel, A.W., Scott Champion, A., Graham, B.J., Randlett, O., et al. (2017). Whole-brain serial-section electron microscopy in larval zebrafish. Nature *545*, 345–349.

Ioffe, S., and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning, F. Bach, and D. Blei, eds. (Lille, France: PMLR), *Proceedings of Machine Learning Research*, vol. 37, pp. 448–456. https://proceedings.mlr.press/v37/ioffe15.html.

Isola, P., Zhu, J.Y., Zhou, T., and Efros, A.A. (2016). Image-to-image translation with conditional adversarial networks .

Januszewski, M., Kornfeld, J., Li, P.H., Pope, A., Blakely, T., Lindsey, L., Maitin-Shepard, J., Tyka, M., Denk, W., and Jain, V. (2018). High-precision automated reconstruction of neurons with flood-filling networks. Nature methods *15*, 605–610.

Karlupia, N., Schalek, R.L., Wu, Y., Meirovitch, Y., Wei, D., Charney, A.W., Kopell, B.H., and Lichtman, J.W. (2023). Immersion fixation and staining of multicubic millimeter volumes for electron microscopy–based connectomics of human brain biopsies. Biological Psychiatry .

Kasthuri, N., Hayworth, K., Berger, D., Schalek, R., Conchello, J., Knowles-Barley, S., Lee, D., Vázquez-Reina, A., Kaynig, V., Jones, T., et al. (2015). Saturated reconstruction of a volume of neocortex. Cell *162*, 648–661.

Kingma, D.P., and Ba, J. (2014). Adam: A method for stochastic optimization .

Knott, G., Marchman, H., Wall, D., and Lich, B. (2008). Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. The Journal of Neuroscience *28*, 2959–2964.

Kornfeld, J., Januszewski, M., Schubert, P., Jain, V., Denk, W., and Fee, M.S. (2020). An anatomical substrate of credit assignment in reinforcement learning. BioRxiv pp. 2020–02.

Li, Y., Meirovitch, Y., Kuan, A.T., Phelps, J.S., Pacureanu, A., Lee, W.C.A., Shavit, N., and Mi, L. (2023). X-ray2em: Uncertainty-aware cross-modality image reconstruction from x-ray to electron microscopy in connectomics. ArXiv.org .

Lichtman, J.W., Pfister, H., and Shavit, N. (2014). The big data challenges of connectomics. Nature neuroscience *17*, 1448–1454.

Loomba, S., Straehle, J., Gangadharan, V., Heike, N., Khalifa, A., Motta, A., Meyer, H.S., Helmstädter, M., and Sievers, M. (2022). Connectomic comparison of mouse and human cortex. Science *377*, 171–+.

Lu, X., Wu, Y., Schalek, R.L., Meirovitch, Y., Berger, D.R., and Lichtman, J.W. (2023). A scalable staining strategy for whole-brain connectomics. bioRxiv pp. 2023–09.

Meila, M. (2003). Comparing clusterings by the variation of information. In Learning Theory and Kernel Machines (Berlin, Heidelberg: Springer Berlin Heidelberg), *Lecture Notes in Computer Science*, vol. 2777, pp. 173–187.

Meirovitch, Y., Mi, L., Saribekyan, H., Matveev, A., Rolnick, D., and Shavit, N. (2019). Cross-classification clustering: An efficient multi-object tracking technique for 3-d instance segmentation in connectomics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8425–8435.

Mi, L., Wang, H., Meirovitch, Y., Schalek, R., Turaga, S.C., Lichtman, J.W., Samuel, A.D.T., and Shavit, N. (2021). Learning guided electron microscopy with active acquisition. Medical Image Computing and Computer Assisted Intervention pp. 77–87.

Minnen, D., Januszewski, M., Blakely, T., Shapson-Coe, A., Schalek, R.L., Ballé, J., Lichtman, J.W., and Jain, V. (2021). Denoising-based image compression for connectomics. bioRxiv .

Mirza, M., and Osindero, S. (2014). Conditional generative adversarial nets .

Mohammed, A., and Abdullah, A. (2018). Scanning electron microscopy (sem): A review. In Proceedings of the 2018 International Conference on Hydraulics and Pneumatics—HERVEX, Băile Govora, Romania. vol. 2018, pp. 7–9.

Morgan, J.L., Berger, D.R., Wetzel, A.W., and Lichtman, J.W. (2016). The fuzzy logic of network connectivity in mouse visual thalamus. Cell *165*, 192–206.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library .

Pavarino, E.C., Yang, E., Dhanyasi, N., Wang, M.D., Bidel, F., Lu, X., Yang, F., Francisco Park, C., Bangalore Renuka, M., Drescher, B., et al. (2023). membrain: an interactive deep learning matlab tool for connectomic segmentation on commodity desktops. Frontiers in Neural Circuits *17*.

Pizer, S., Johnston, R., Ericksen, J., Yankaskas, B., and Muller, K. (1990). Contrast-limited adaptive histogram equalization: speed and effectiveness. In [1990] Proceedings of the First Conference on Visualization in Biomedical Computing (IEEE Comput. Soc. Press), pp. 337–345.

Potocek, P. (2021). Adaptive specimen image acquisition using an artificial neural network. US Patent 10,928,335.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (Springer), pp. 234–241.

Shapson-Coe, A., Januszewski, M., Berger, D.R., Pope, A., Wu, Y., Blakely, T., Schalek, R.L., Li, P.H., Wang, S., Maitin-Shepard, J., et al. (2024). A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution. Science *384*, eadk4858.

Shavit, N., Samuel, A., Lichtman, J., and Mi, L. (2021). System and method for learning-guided electron microscopy. US Patent 11,164,721.

Shavit, N., Samuel, A., Lichtman, J., and Mi, L. (2023). System and method for learning-guided electron microscopy. US Patent 11,557,459.

Sheridan, A., Nguyen, T.M., Deb, D., Lee, W.C.A., Saalfeld, S., Turaga, S.C., Manor, U., and Funke, J. (2023). Local shape descriptors for neuron segmentation. Nature Methods *20*, 295–303.

Song, K., Feng, Z., and Helmstaedter, M. (2023). High-contrast en bloc staining of mouse whole-brain and human brain samples for em-based connectomics. Nature Methods *20*, 836–840.

Swanson, L.W., and Lichtman, J.W. (2016). From cajal to connectome and beyond. Annual Review of Neuroscience *39*, 197–216.

Thermo Fisher Scientific (2018). Software SEM and FIB SEM software for Python-based scripting control of your instrument. Available at https://www.thermofisher.com/cz/en/home/electron-microscopy/products/software-em-3d-vis/autoscript-4-software.html, AutoScript 4.

Thermo Fisher Scientific (2020). Verios 5 XHR SEM. Scanning electron microscopy characterization of nanomaterials with sub-nanometer resolution and high material contrast. Available at https://www.thermofisher.com/us/en/home/electron-microscopy/products/scanning-electron-microscopes/verios-xhr-sem.html.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. nature *381*, 520.

Vincent, L., and Soille, P. (1991). Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis & Machine Intelligence *13*, 583–598.

Wang, H., Rivenson, Y., Jin, Y., Wei, Z., Gao, R., Gunaydin, H., Bentolila, L.A., Kural, C., and Ozcan, A. (2019). Deep learning enables cross-modality super-resolution in fluorescence microscopy. Nature Methods *16*, 103–110.

Weigert, M., Schmidt, U., Boothe, T., Mueller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., et al. (2018). Content-aware image restoration: pushing the limits of fluorescence microscopy. Nature Methods *15*, 1090–1097.

White, J.G., Southgate, E., Thomson, J.N., and Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. Philosophical Transactions of the Royal Society B: Biological Sciences *314*, 1–340.

Witvliet, D., Mulcahy, B., Mitchell, J.K., Meirovitch, Y., Berger, D.R., Wu, Y., Liu, Y., Koh, W.X., Parvathala, R., Holmyard, D., et al. (2021). Connectomes across development reveal principles of brain maturation. Nature *596*, 257–261.

Xu, C.S., Januszewski, M., Lu, Z., Takemura, S.y., Hayworth, K.J., Huang, G., Shinomiya, K., Maitin-Shepard, J., Ackerman, D., Berg, S., et al. (2020). A Connectome of the Adult Drosophila Central Brain. bioRxiv .

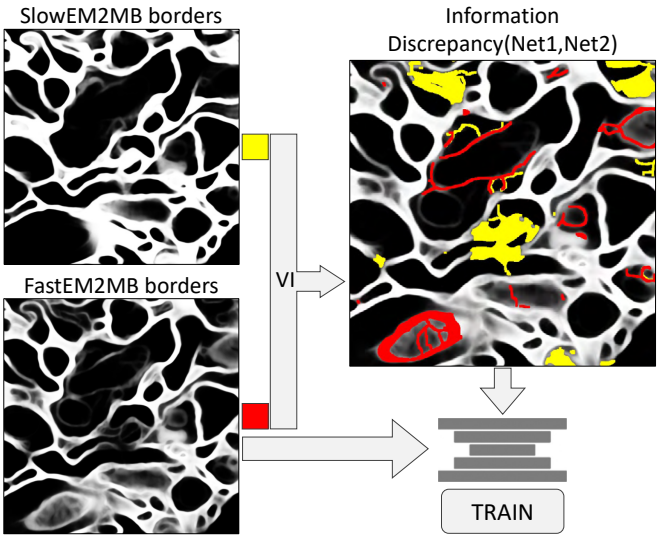

Figure S1. The discrepancy between segmentation with long dwell time (using SLOWEM2MB) and short dwell time (using FASTEM2MB) is defined based on VI. VI is the sum of individual error terms contributed by each object in the two segmented images. The most variable objects are flagged. Image processing is used to delineate specific borders that appear in only one segmented image. Yellow represents segmented objects that are uniquely predicted in the long dwell time image. Red represents segmented objects that are uniquely predicted in the short dwell time image. A neural network (ERRNET) is trained to predict all red and yellow discrepancies only using short dwell time images. This is possible because variation occurs where membrane predictors are uncertain and often with typical, at times biologically implausible, membrane prediction.

Supplemental Information

Segmenting composite images

The smart microscope should be able to analyze images composed from multiple dwell times (see **Figures 1C, 2B, 2C, 4A, 6A-6D**). We tested whether replacing error-prone regions in a short dwell time image with regions taken from long dwell time images improves segmentation outcomes. **Figure S2** depicts the segmentation outcome of a short dwell time image taken at 100 ns/pixel segmented with a dedicated 100 ns network FASTEM2MB (**S2A,S2E**), and by FUSEDEM2MB (**S2B,S2F**). The segmentation quality of these networks are similar (top panel; VI=0.025 and VI=0.022). In most scenarios, the network trained to deal with fused EM (FUSEDEM2MB) produces better results than networks trained to handle a fixed dwell time, even if the input to the two networks consists of a single homogeneous dwell time. **Figures S2C, S2G** depict the segmentation of an image where the error-prone regions were detected by an error detector and replaced with long dwell time pixels (2500 ns). The error level is typically and substantially cut by $\sim$ 3-4 $\times$. The 2500 ns reference image and its segmentation are shown in **Figures S2D, S2H**. All error estimates based on VI shown in **Figure S2** are presented as the sum of the merge error term and split error term.

Imaging procedure

The SEM is automated to acquire acquire images of individual tiles of every specimen section that are eventually stitched and aligned to form a total image volume **Figure 4**. The microscope navigates through multiple specimen sections held on tape and defines every specimen region of interest (S-ROI). Each S-ROI is captured at high spatial resolution by multi-tile acquisition. To identify the S-ROI and automate stage position and rotation control, we used SEM Navigator, a custom interface akin to earlier WaferMapper software (Hayworth et al., 2014). The list of S-ROIs is exported into a text file, which is subsequently processed by the SmartEM pipeline (coded in Python/Matlab) using the Thermo Fisher Scientific Autoscript (Thermo Fisher Scientific, 2018) package. The SmartEM pipeline controls the Verios (Thermo Fisher Scientific, 2020) microscope, moves to S-ROI and individual tile positions, controlling the entire acquisition sequence.

For all image acquisitions, we used the Verios UHR (Ultra High Resolution) imaging mode with 4nm/pixel spatial resolution and $\sim$ 4 mm working distance. Image contrast was obtained using a back-scattered electron detector with 2000 V stage bias. The initial short dwell time scan was obtained using the full frame acquisition Autoscript interface. The subsequent long dwell time re-scan utilized the standard interface of Autoscript patterning

To optimize image quality and tuning time for both short movements between neighboring tiles and long movements neighboring sections, we customized sequences of various autofunctions. These autofunctions included autocontrast/brightness (ACB), auto-focus (AF), auto-stigmation (AS), auto-focus/stigmation (AFS), and auto-lens (AL) alignment.

Because we used different interfaces for the initial short dwell time scan and long dwell time re-scan, an additional alignment procedure was necessary to achieve pixel-resolution precision in the re-scan. The basic system configuration for the re-scan acquisition is described in Potocek (2021).

When the re-scan long dwell time was shorter than $\sim$ 500 ns/pixel, an unavoidable artifact due to limited system response of the electron deflection system occurred at the edge of re-scan regions. We excised this artifact by omitting a 1-pixel boundary from every re-scan region.

Segmentation quality metric

To compare the segmentation quality of different samples we used a variation of information (VI) metric (Meila, 2003). In principle all comparisons that we made in this study can be accomplished with other metrics of segmentation quality as long as they can be applied to 2-dimensional images. We expect the choice of segmentation metric to have little effect as long as any metric assesses similar topological attributes as VI (i.e., whether objects are split or merged). Our implementation of the VI running on CPU/GPU is available at https://pypi.org/project/python-voi/.

Using VI to build ERRNET. To train the error detectors we needed to locate the specific regions that contribute to the largest segmentation differences between image pairs, which is not provided by the VI metric. VI combines split and merge errors. The two error measures are defined by comparing the entropy of three segmented images (Meila, 2003), $S_1 \in L_1^N$, $S_2 \in L_2^N$ and $S_1 \times S_2 \in L_1^N \times L_2^N$ for two N-pixel labeling (instance segmentation) S_1 and S_2 that needs to be compared, where the Ls represents the sets of pixel labels. The segmented

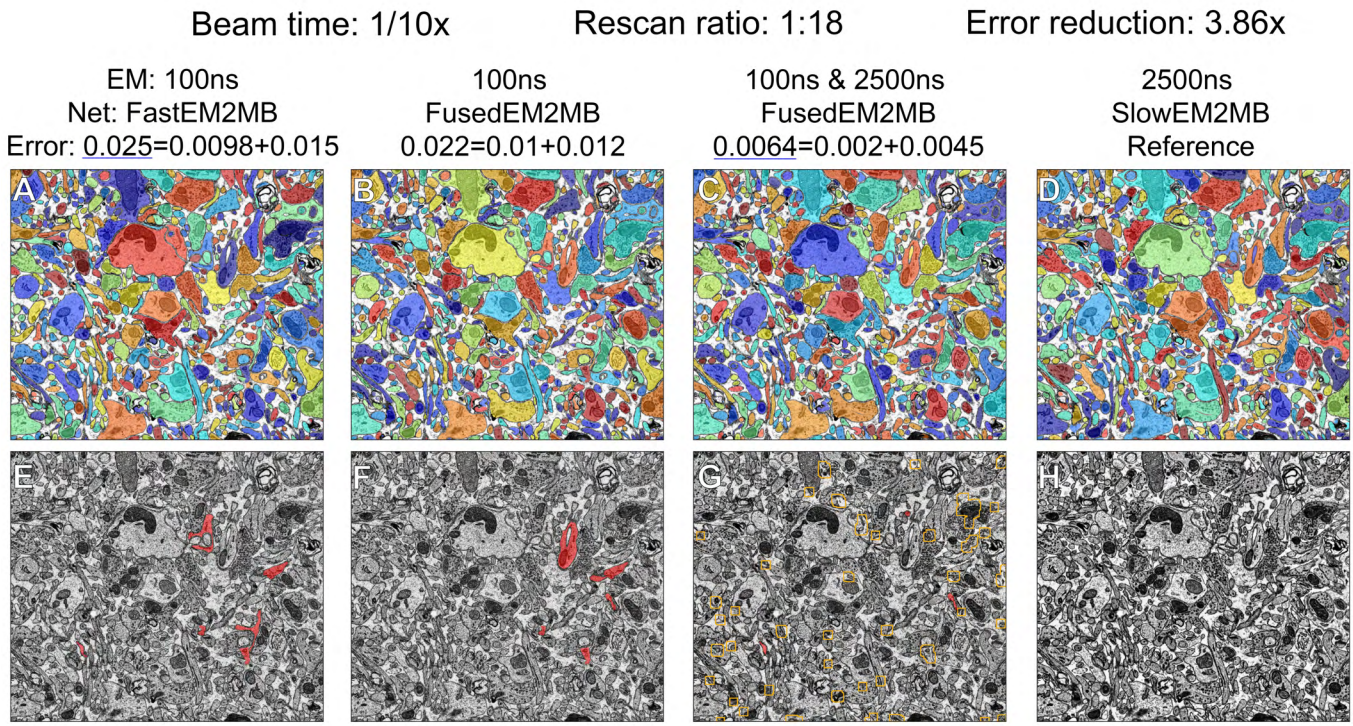

Figure S2. Composite EM images fusing short and a long dwell time regions are better segmented compared to short dwell time images. We tested whether replacing error-prone regions harms the ability to segment. Composite images tend to be segmented with dramatically higher accuracy. Error of the instance segmentation is assessed in terms of the Variation of Information (VI) compared to the segmented reference image, where VI is composed of a merge and split error terms as in Equation 1.

image $S_1 \times S_2$ is labeled by concatenating the labels from S_1 and S_2 for each pixel. The VI is then the sum of two error terms $\mathrm{VI}_{\mathrm{merge}}$ and $\mathrm{VI}_{\mathrm{split}}$

$$\begin{aligned} \mathrm{VI}_{\mathrm{merge}} &= H(S_1 \times S_2) - H(S_1), \\ \mathrm{VI}_{\mathrm{split}} &= H(S_1 \times S_2) - H(S_2), \\ \mathrm{VI} &= \mathrm{VI}_{\mathrm{merge}} + \mathrm{VI}_{\mathrm{split}}. \end{aligned} \tag{1}$$

Due to the additivity of the entropy measure (Meila, 2003), $\mathrm{VI}_{\mathrm{merge}}$ and $\mathrm{VI}_{\mathrm{split}}$ can be broken into individual constituents, representing the amount of error contributed by each individual label in each segmentation. We could thus rank objects in each segmentation according to the amount of variation they contribute to overall VI (**Figures S5**). The error contributed by the set of pixels that are both in segment $s_1 \in S_1$ and $s_2 \in S_2$ (i.e. the error contributed by a segment in $S_1 \times S_2$) is

$$W(s_1 \cap s_2) - W(s_1)$$

and

$$W(s_1 \cap s_2) - W(s_2),$$

for the split and merge errors, respectively, where $W(A) = -\frac{|A|}{N} \cdot \log \frac{|A|}{N}$, $|A|$ is the number of pixels in A and N is the number of pixels in the image.

Once the significantly incompatible objects are detected in each segmentation, we used image processing to delineate the borders that are responsible for the topological differences between the two segmented images (**Figure S1**). We then produced binary masks from these errors and trained neural networks (ERRNET) to detect them directly from membrane probability maps, themselves produced by another neural network (FASTEM2MB). Detecting borders allows our technique to disregard small "cosmetic" variations between two segmentations that do not cause meaningful topological differences.

Determination of maximal segmentation quality. We developed an unbiased estimate for the minimal dwell needed for 2D segmentation. We compared segmentations from N images for each pair of dwell times $d_1 < d_2$ and an overly slow dwell time d_{ref}. We asked whether the VI of the d_2 images was significantly smaller ($p<0.05$; Wilcoxon signed rank test) than d_1 images compared to d_{ref} images. When two dwell times were not sufficiently different, we call these dwell times equivalent. We defined the minimum dwell time with near maximal segmentation ability as that dwell time beyond which VI does not improve.

Forcing fast scan imaging of desired regions

The acceleration of SmartEM depends on the quantity of rescanned pixels. Since the re-scanning mask is learned rather than calculated through a fixed process, regions irrelevant to the connectomics task may contain error-prone regions and appear in the re-scan map, potentially reducing speedup. To exclude irrelevant regions from slow re-scan, we built another neural network module (EMEXCLUDE) to calculate what regions should be excluded from any re-scan, even if they might be flagged as error-prone by ERRNET. Developing a separate EMEXCLUDE module (rather than adding this capability to ERRNET) conferred additional flexibility to the SmartEM pipeline by allow-

ing us to adaptively choose what regions should be excluded from re-scan without retraining ERRNET. Bypassing irrelevant pixels (e.g., cell nuclei, blood vessels) during re-scan boosts efficiency by conserving time and computational resources.

Here, we implement EMEXCLUDE to exclude regions that are sufficiently far from any cellular membranes. To do this, we utilize the Euclidean distance transform on input binary membranes. This transform calculates the shortest Euclidean distance from each zero pixel (background) to any non-zero (foreground) pixel in the image. To train EMEXCLUDE, we binarize the distance transform with a fixed threshold (**Figure S13**). The features of irrelevant regions we learned as a semantic segmentation task using paired EM images and their binary masks (see **Neural network models**). The SmartEM pipeline applies EMEXCLUDE in real-time on short dwell time images and precludes re-scanning irrelevant regions that might have been predicted by ERRNET. To assess the performance of different modules in the SmartEM pipeline, we exclude EMEXCLUDE from speedup tests shown in **Figure 5**. For the cytoplasm exclusion described above, the average exclusion proportion is about 23% as shown in **Figure S13**. The speedup tests shown in **Figure 5** would improve with the implementation of EMEXCLUDE.

Identifying additional high-interest regions for slow re-scan

ERRNET identifies regions susceptible to segmentation errors and re-scans them at a higher quality to improve segmentation accuracy. The same strategy can be re-formulated, not only to identify error-prone regions, but to identify additional image-specific regions of special interest, such as synapses or any sub-cellular component of biological interest. Here, we built an additional neural network module (EMINCLUDE) to re-scan regions identified as synapses, because of their high relevance to connectomics. Mouse cortex typically contains $\sim$1-1.5 synapses per μm^3 (Kasthuri et al., 2015), or $\sim$2-3 synapses per field of view when image tiles are $\sim 8\times 8$ μm^2. Because of synapse sparsity, the re-scan time does not substantially increase. We trained EMINCLUDE with a set of manually-annotated long dwell time SEM images.

To train EMINCLUDE, we first trained a neural network to detect synapses using manual annotations of long dwell time images (EMINCLUDE). The high performance of EMINCLUDE is shown in **Figure S7**. We paired short dwell time images with the binary masks for synapse locations predicted by EMINCLUDE (which had used long dwell time images to make the predictions). This procedure created ground truth to train EMINCLUDE. A snapshot of the synapse detection and re-scan mask generation pipeline is shown in **Figure S11**. The hyper-parameters and training details of EMINCLUDE are similar to EMEXCLUDE.

Optional image homogenization

The SmartEM pipeline produces composite image with pixels acquired at different dwell times. A human observer will note contrast differences at interfaces between pixels with different dwell times. To increase human image interpretability, we built

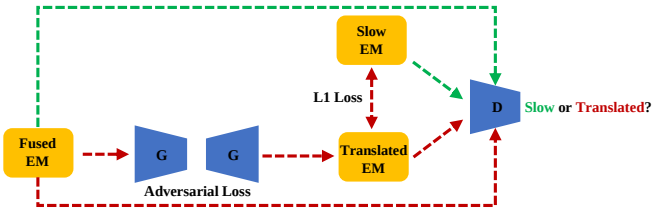

Figure S3. Image Translation Model. G: *generator.* D: *discriminator.* The generator G takes a fused EM as input and produces a translated EM (i.e., fake slow EM) that looks similar to slow EM (i.e., taken by the microscope). The discriminator D takes as input a concatenation of a fused EM and another image that is either slow EM (green arrows) or a translated EM (red arrows). The aim of the discriminator is to classify whether the second image is slow EM or translated EM. The model is trained with a combination of adversarial loss and L1 loss.

an image translator component that homogenizes SmartEM images to look like standard EM images with uniform dwell times. **Figure S8** shows a specific example, a fused EM image that is a mosaic of sub-images with different dwell times. To mitigate dwell time contrasts and produce a visually coherent image, we applied a conditional generative adversarial network (IMAGEHOMOGENIZER, cGANs) (Mirza and Osindero, 2014). Previous studies used deep learning to improve the quality of microscopy images (Fang et al., 2021; Wang et al., 2019; Weigert et al., 2018; Mi et al., 2021), de-noise EM images (Minnen et al., 2021), and perform image reconstruction across different modalities (Li et al., 2023). IMAGEHOMOGENIZER contains two convolutional neural networks (CNN): a *generator* and a *discriminator* (Isola et al., 2016). Training data are a composite image and a uniformly long dwell time image, where the composite image is generated by randomly combining pixels from short dwell time and long dwell time images in different proportions (**Figures 6B,6C,6D** where the composite images consist of 75 ns and 600 ns pixel dwell times). As shown in **Figure S3**, during the training process, the *generator* translates the simulated composite images to resemble long dwell time images, and the *discriminator* attempts to distinguish the translated images from real long dwell time images. The training process utilizes L1 loss and adversarial loss. After image homogenization by the *generator*, the fused EM images are more suitable for human inspection and retain the visual details of fine ultrastructure **Figure S8**.

Neural network architectures

For all neural network models, we strove for simple architectures that would allow straightforward reproducibility of results. A U-Net like architecture (Ronneberger et al., 2015) was used to train membrane detection of homogeneous dwell time EMs (SLOWEM2MB, FASTEM2MB), any dwell-time EM (EM2MB), and composite EM where each image fuses more than one dwell time (FUSEDEM2MB). We found that FUSEDEM2MB, once trained, could be used for all membrane prediction tasks without compromising quality. The same U-net architecture was also used to train ERRNET, EMINCLUDE, and EMEXCLUDE. We tried the U-net architecture for image homogenization, but achieved better results with conditional GANs.

Architecture for FUSEDEM2MB and ERRNET. The selected architecture, similar to the UNET(Ronneberger et al., 2015), shown in **Figure S14** has 3 sets of 2D-Convolution, Batch-Normalization(Ioffe and Szegedy, 2015), ReLU in each layer. We use residual connections(He et al., 2016) adding the output of the first convolution to the last one in each layer. This architecture showed the highest segmentation accuracy when varying the number of CBR (Conv-BatchNorm-ReLU) in each layer (2∼4), the usage of residual connections, and the type of residual connections (concatenation or addition).

U-Net architecture for EMEXCLUDE. We trained a fully convolutional UNET model over 200 epochs, employing a learning rate of 0.01. The model was configured with five layers of depth and filter sizes progressively sequenced as 32, 64, 128, 256, and 512. To introduce non-linearity and manage potential negative inputs, we incorporated a leakyReLU activation function.

Image Normalization and Augmentation. To train the FUSEDEM2MB network, we used the CLAHE (Pizer et al., 1990) normalization with *clipLimit=3* to bring all images to a common color space. We used on the fly rotation, flip, translation to augment the images in the training set. Although images are naturally 2048×2048, we sub-sampled 256×256 squares to train the network. To allow the network to deal with images with multiple dwell times, we randomly replace patches at random locations with different dwell times. Specifically, each sample was generated by choosing a baseline image at a single dwell time and replacing up to 30 patches with a maximum size of 11×11 pixels with the corresponding pixels of an image with longer dwell time.

To train ERRNET, we normalized membrane probabilities to [0,1] as an input to the network. We used the same procedure for on the fly translation and rotation but did not replace patches.

Training Procedure. We used the Pytorch framework (Paszke et al., 2019) to implement and optimize the network. The Adam optimizer (Kingma and Ba, 2014) with learning rate 0.001 was used to update the network parameters. We used a batch size of 16 images. We trained the FUSEDEM2MB network for 50000 gradient steps. We evaluated validation loss every 1000 steps over 100 batches. The network converged after ∼35000 gradient steps. The same procedure was used to train ERRNET. ERRNET converged after ∼8000 gradient steps.

Image Translation Networks. IMAGEHOMOGENIZER uses a conditional GAN called pix2pix (Isola et al., 2016), consisting of a generator CNN and discriminator CNN. The generator includes an encoder and decoder that downsamples and then upsamples the input image. The discriminator tries to discriminate between slow EM and translated EM. At the training stage, we use a batch size of 1 and randomly crop 128 × 128 image tiles from a larger composite EM image. The model is first trained with a constant learning rate of 0.0002 for 100 epochs and then for another 100 epochs, during which the learning rate decays to zero. At the inference stage, the whole composite EM image is passed to the model without cropping.

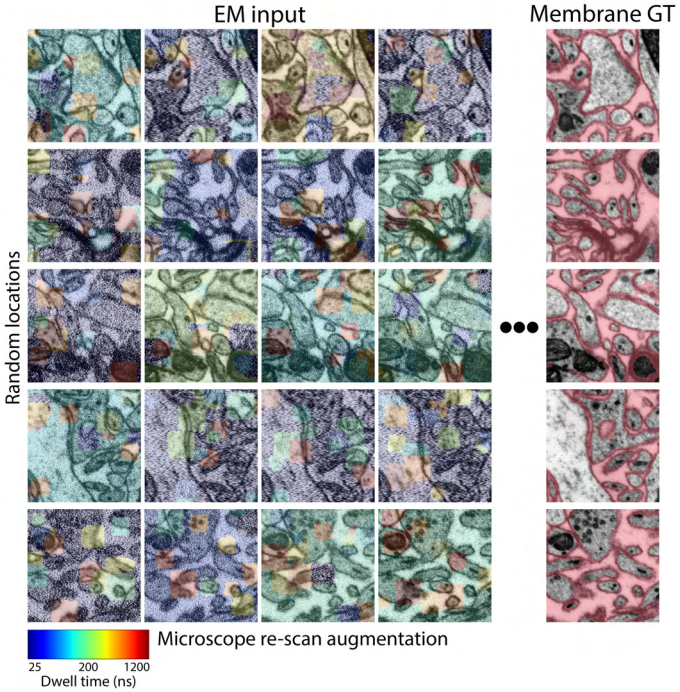

Figure S4. Dwell-time re-scan data augmentation. Rows 1-5 show different locations in the EM sample. Columns 1-4 show different augmented composite images that were taken at different dwell times; short dwell time pixels in blue, representing 25 ns scans; long dwell time pixels in red, representing 1200 ns pixels. Column 5 shows the groundtruth classes for each region that were obtained from the long dwell time neural network (SLOW2EM). The aim of FUSEDEM2MB is to classify membrane pixels. Additional augmentations such as translation, rotation, and flip are used during training.

Image stitching and alignment

The stitching and alignment of the sample volume was performed on composite dwell time images. After applying a band-pass filter to raw images, we used conventional block matching technique to obtain matching points between neighboring images, from which elastic transformations mapping the raw data to the aligned volume were computed by mesh relaxation. Code for stitching and alignment is available at Stitching and alignment code. We applied the same stitching and alignment transformations to the fast, composite, and homogenized images to produce three sets of final volumes.

Segmentation and neuronal reconstruction

Neuron reconstruction technique. To reconstruct neurons in 3D, we applied a lightweight segmentation method that we previously used to reconstruct neurons from the same sample imaged by a multi-beam SEM (Karlupia et al., 2023) and tissue prepared using a whole mouse brain staining technique (Lu et al., 2023). First, pixels straddling intra-cellular spaces were predicted by a CNN, based on the pre-trained FUSEDEM2MB network. To improve the network accuracy, we fine-tuned FUSEDEM2MB using thirty-six 1024×1024 SmartEM tiles obtained from random locations in the target volume and annotated by an expert. Predictions from FUSEDEM2MB were used as a starting point for the annotation process of the training set. All sections were segmented in 2D using the fine-tuned network and watersheds (Pavarino et al., 2023). Second, a CNN was trained to predict from the EM the medial axis of all objects in 2D. This process required no

additional human annotation. Third, 2D object segments were agglomerated across sections based on shape alignment and similarity. In addition, 2D segments were agglomerated if their medial axes were well-aligned using a fixed threshold determining large overlaps. Fourth, agglomerated objects containing a large number of adjacent 2D segments were flagged as objects with possible merge errors. This was done by building a Regional Adjacency Graph whose nodes are 2D segments and edges represent spatial adjacency. Then these objects were re-agglomerated iteratively from the original 2D object segments until the merge-error criterion was attained using an iterative clustering technique (Bailoni et al., 2022). Fifth, orphans were detected and connected to other orphans or non-orphan objects based on their best estimate from the agglomeration graph, i.e., connecting them to objects that did not pass the agglomeration threshold in the first iteration. The results of the reconstruction are shown in **Figure 7C**.

Criterion for filtering dendritic spines. Three dendrites were randomly selected for quantitative analysis. We defined correctly segmented spines as spines whose segmentation includes their synapse-containing regions. Incorrectly segmented spines were split errors that occurred before the synaptic region. To avoid confusing spines with dendritic filopodia, we excluded putative spines from analysis if no potential synapse was contained in the image volume. There observed three types of error: *Type 1* errors occur when the spine is prematurely truncated by a split error that occurs before the spine's corresponding synapse that was not due to an obvious image artifact (e.g., tissue preparation, folds in the section). *Type 2* errors occur when the spine is not tracked at all due to a split error at its base on the dendrite that was not due to an obvious image artifact. *Type 3* errors occur when the spine is lost due to an obvious artifact. We observed such errors caused by local aberrations in tissue preparation in sections 56, 65, 66, 77 and 88. The distribution of incorrect spines and their corresponding error type is shown in Table S1. To characterize only errors that might be associated with the SmartEM technique, we exclude the rate of Type 3 errors from consideration.

Statistical tests

All statistical tests were done using the Wilcoxon signed-rank test for paired samples. The test was used to assess the cases where two dwell times produce similar segmentation quality by comparing the variation of information of individual samples to a single reference taken at a longer dwell time.

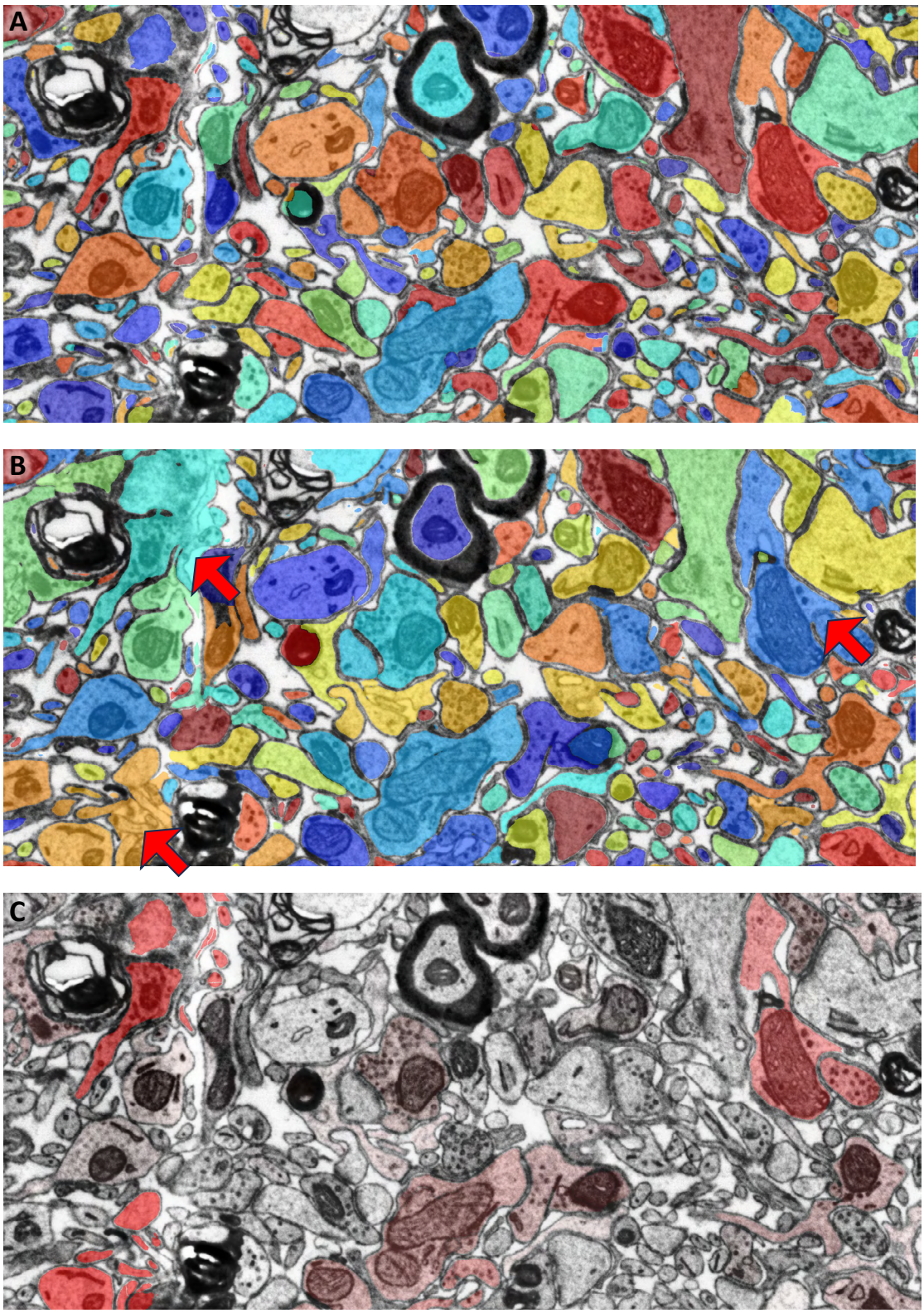

Figure S5. Ranking objects of two segmented images based on contribution to variation of information. **A.** Segmentation of long dwell time image at 1000 ns. **B.** Segmentation of short dwell time image at $\sim$ 100 ns overlaid on 1000 ns EM. Some large errors are indicated with red arrows. **C.** Objects that vary between the two segmented images. Red heatmap indicates contribution to variation of information (Meila, 2003) where variable objects come from either of the two segmented images. The largest variation is captured by the three objects indicated by red arrows.

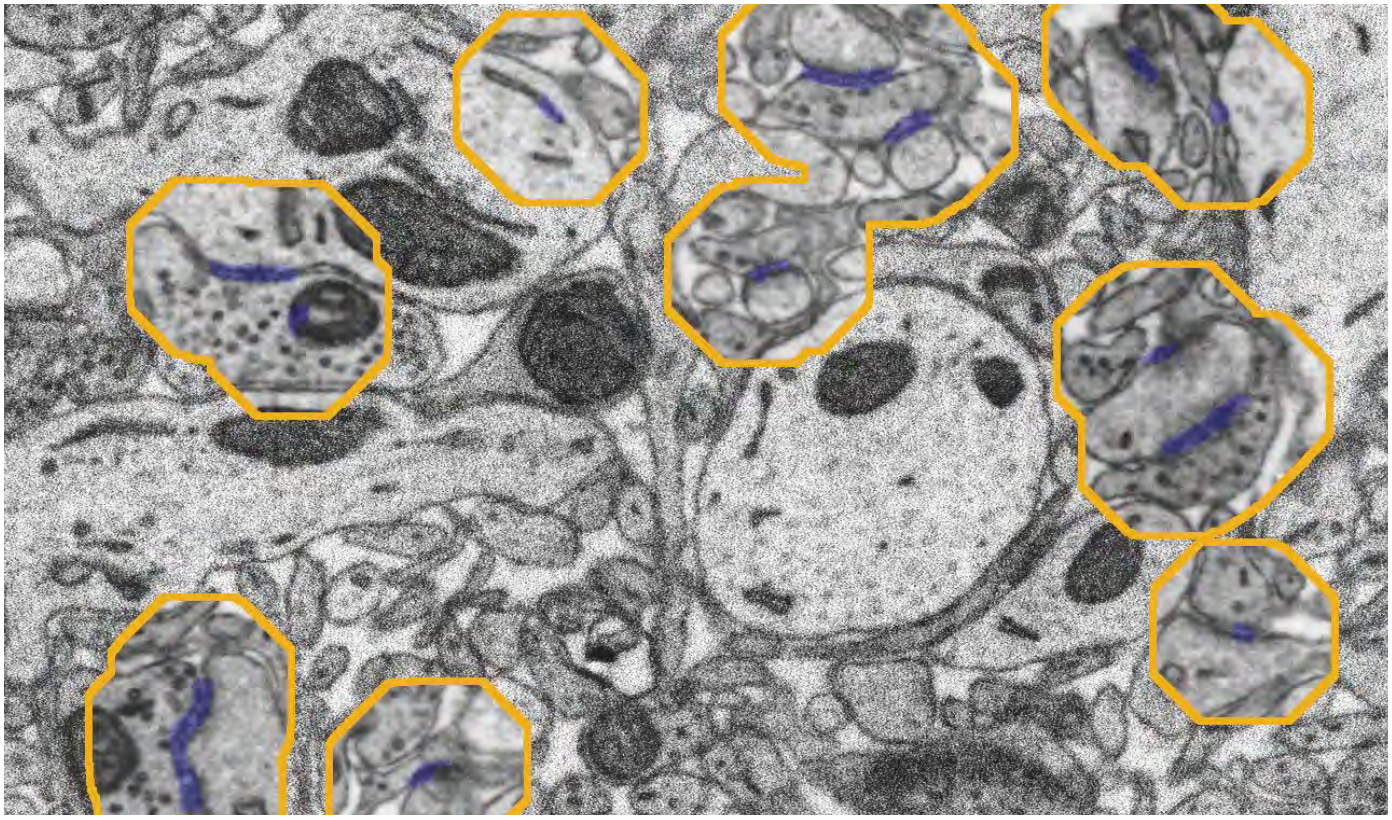

Figure S6. Data-aware imaging of synapses at long dwell time. SmartEM takes a short dwell time image (50 ns/pixel), predicts locations that contain synapses, and re-scans these regions at long dwell time (1200 ns/pixel). The blue overlay presents synapse predictions by EMINCLUDE. Yellow outlines represent locations for re-scan based on dilation of EMINCLUDE predictions.

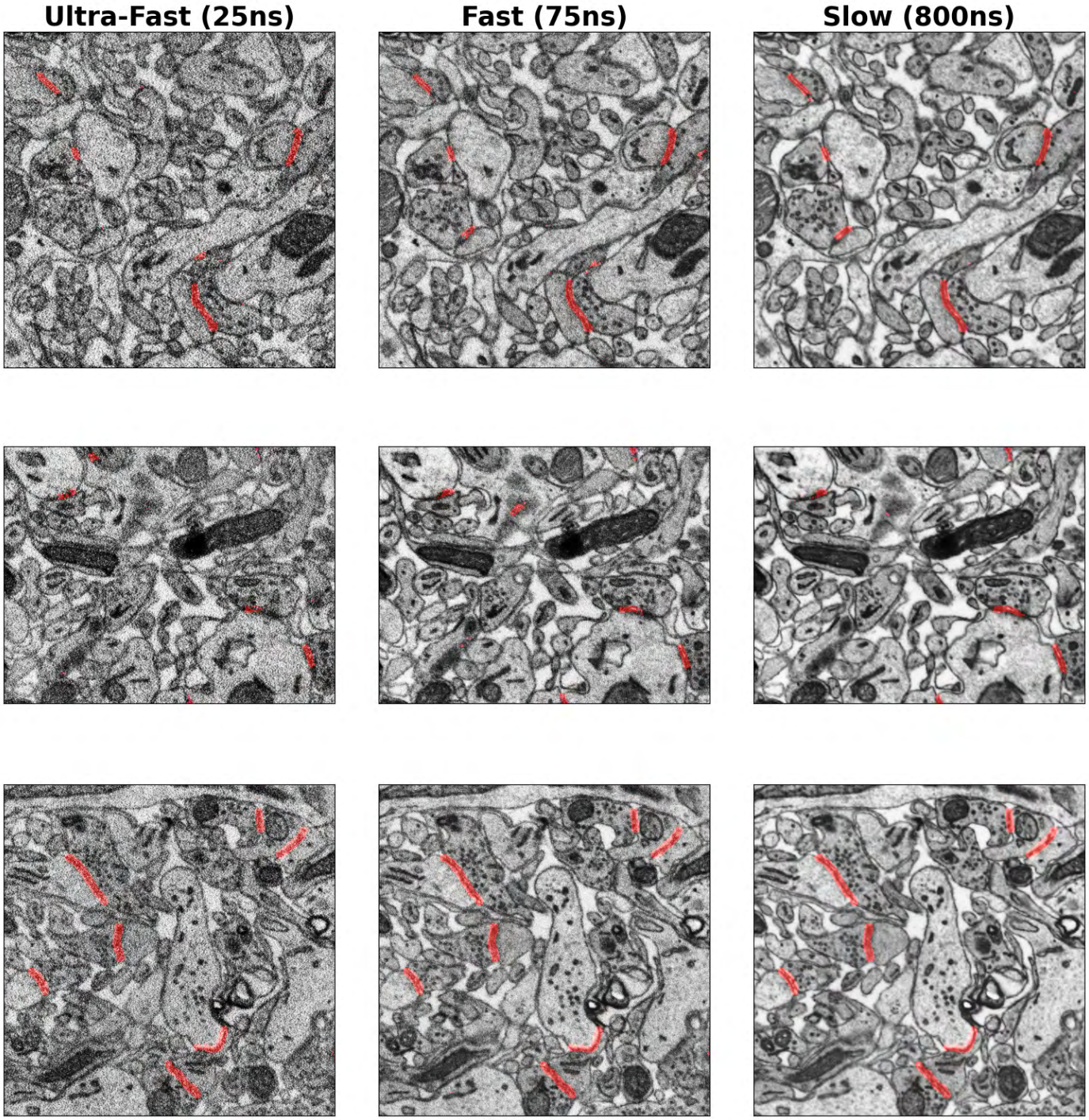

Figure S7. Synapse detection in ultrafast (25 ns), fast (75 ns) and slow (800 ns) dwell time. EMINCLUDE works at multiple dwell times.

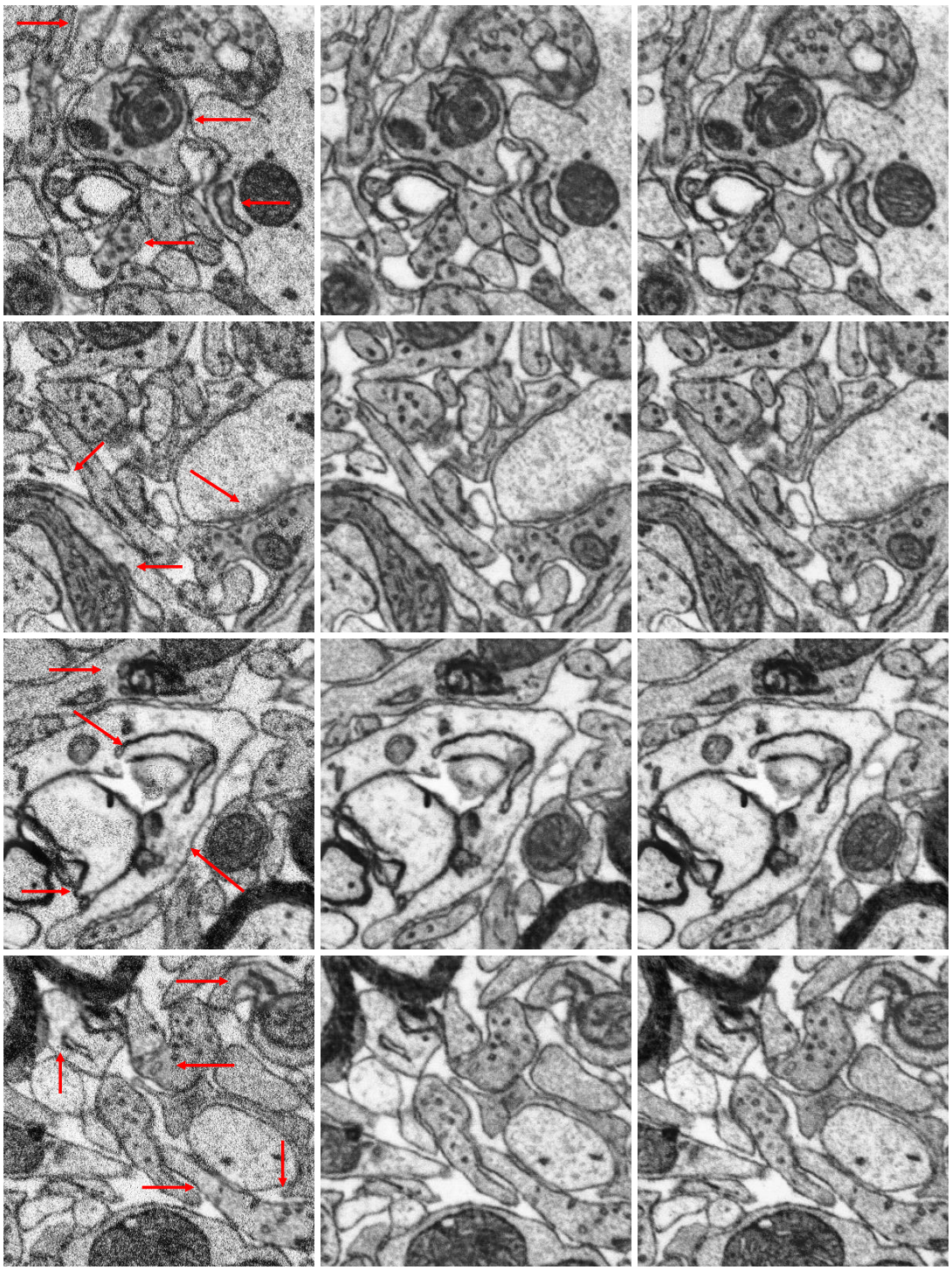

Figure S8. Examples of image homogenization by IMAGEHOMOGENIZER. Left column: composite EM with two dwell times (75 ns/pixel and 600 ns/pixel). Middle column: homogenized EM from composite EM, exhibiting similar visual coherence compared to slow EM. Right column: slow EM (600 ns/pixel). Red arrows indicate the locations with slow dwell time of 600 ns/pixel in composite EM.

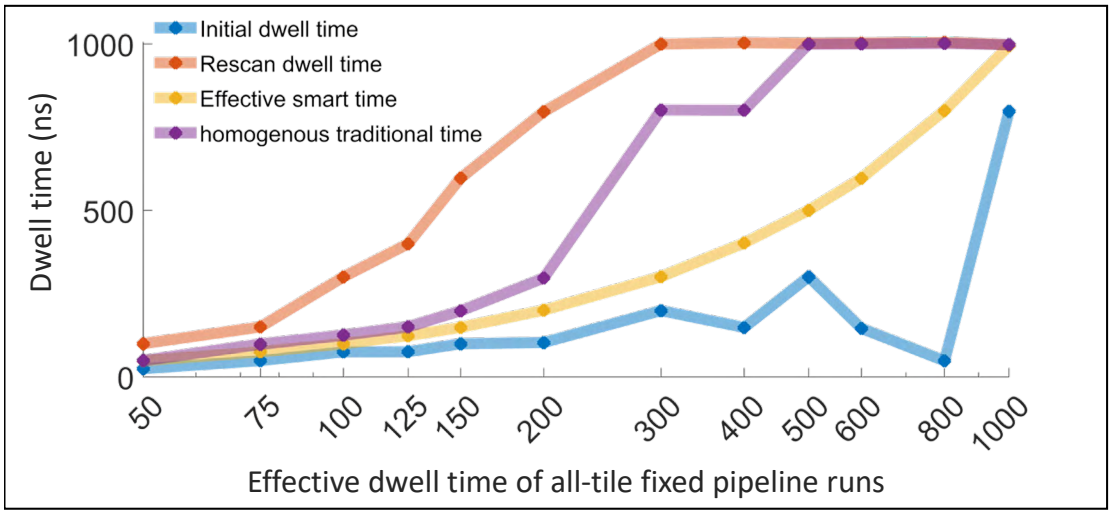

Figure S9. Globally fixed SmartEM parameters and their respective speedup compared to traditional EM.

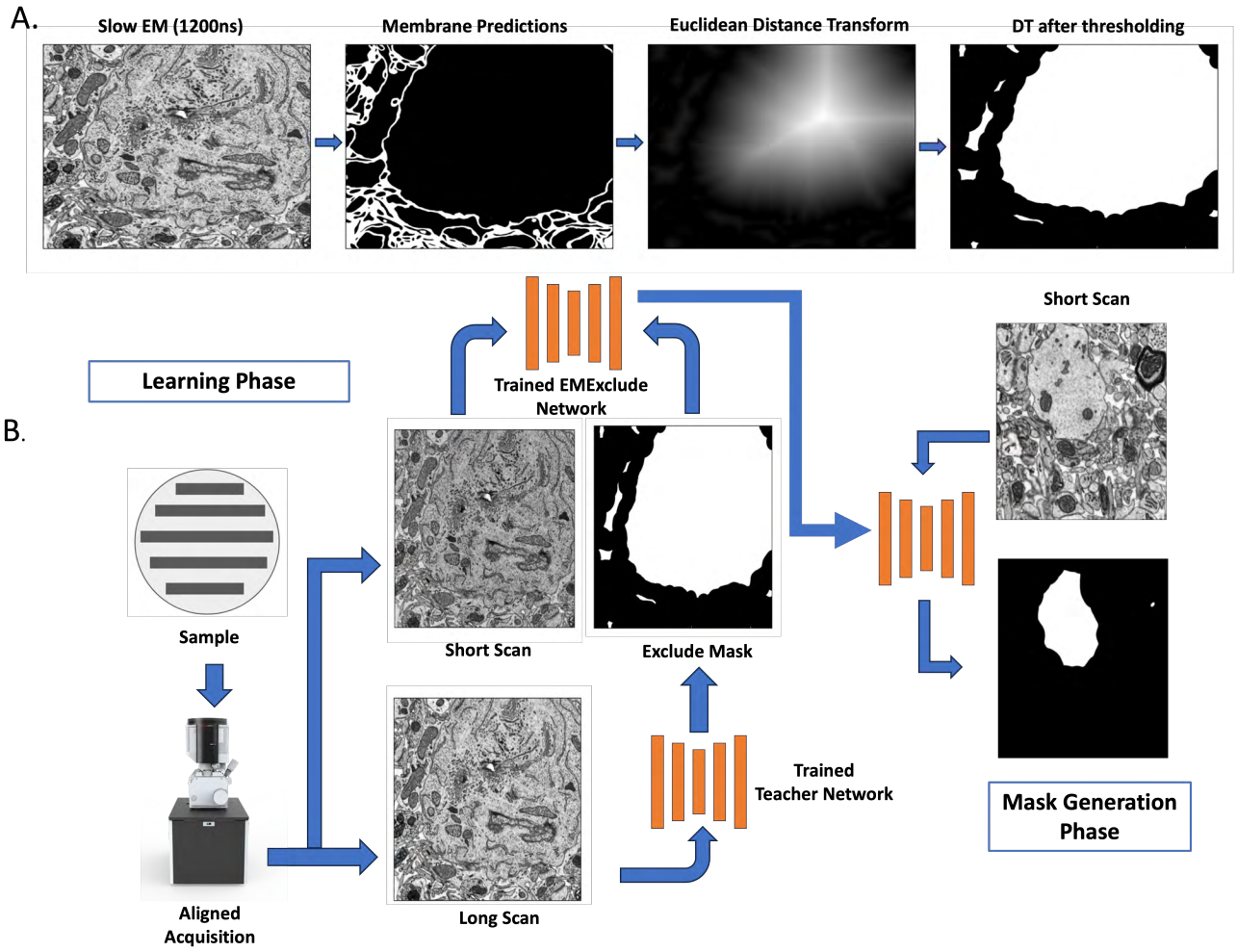

Figure S10. A. The process of generating the EMEXCLUDE ground truth: EM images taken at slow dwell times are processed to determine regions that should be excluded in subsequent scans. The sequence begins with the raw EM image, proceeds to membrane predictions highlighting essential structures, and then applies a Euclidean distance transform to emphasize key features. The final output is a binary differentiation after thresholding, which identifies areas of minimal interest, establishing the EMEXCLUDE ground truth. B. The EMEXCLUDE ground truth is paired with fast EM images to train a neural network, enabling it to recognize and exclude similar non-essential regions in new scans. Once trained, the network processes new EM images in real-time, generating EMEXCLUDE masks.

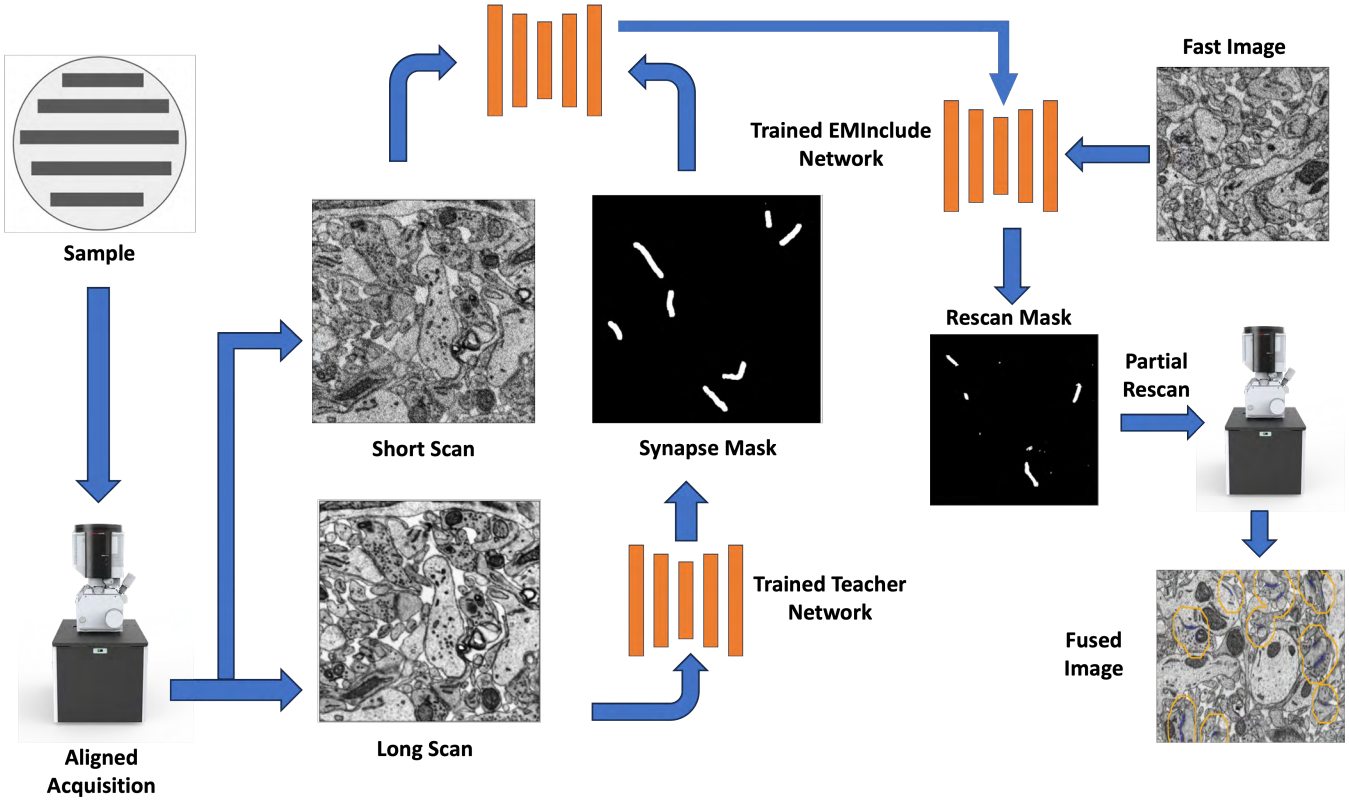

Figure S11. Synapse detection and rescan mask generation pipeline: aligned acquisition provides electron microscopy (EM) images at varying dwell times. A teacher network is trained to identify synapses from slow dwell time images, and these identified labels train a student network, EMInclude, for synapse detection on faster dwell time images. This student network predicts synapse locations to generate a rescan mask, directing the microscope for targeted slow point scans of selected synapses. The outcome is a fused EM image that integrates different dwell times, optimizing scanning speed and detail in areas of interest.

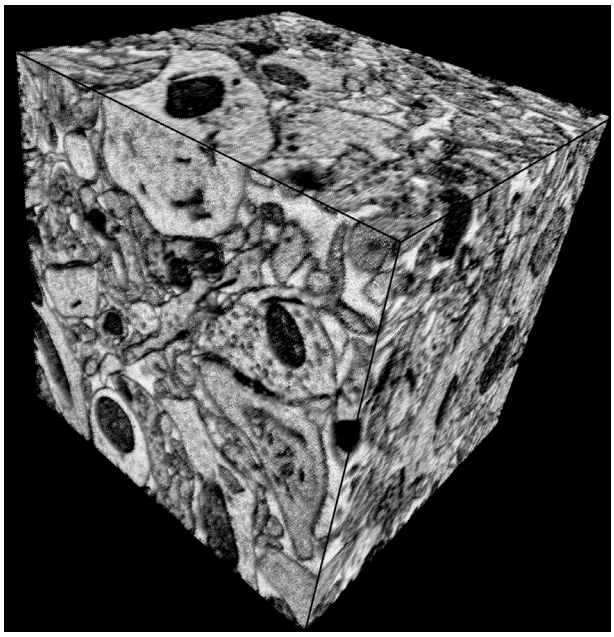

Figure S12. A cubical 3 μm portion of an aligned smart EM output from 94 serial sections.

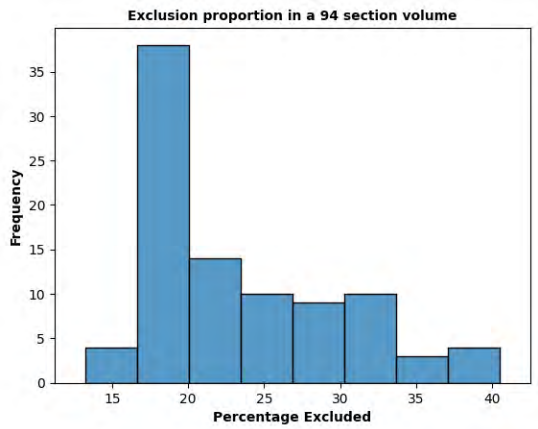

Figure S13. Percentage of EM that can be excluded in a $60 \times 68 \times 3\mu m^3$ section. On average, around 23% of the volume can be excluded from rescanning.

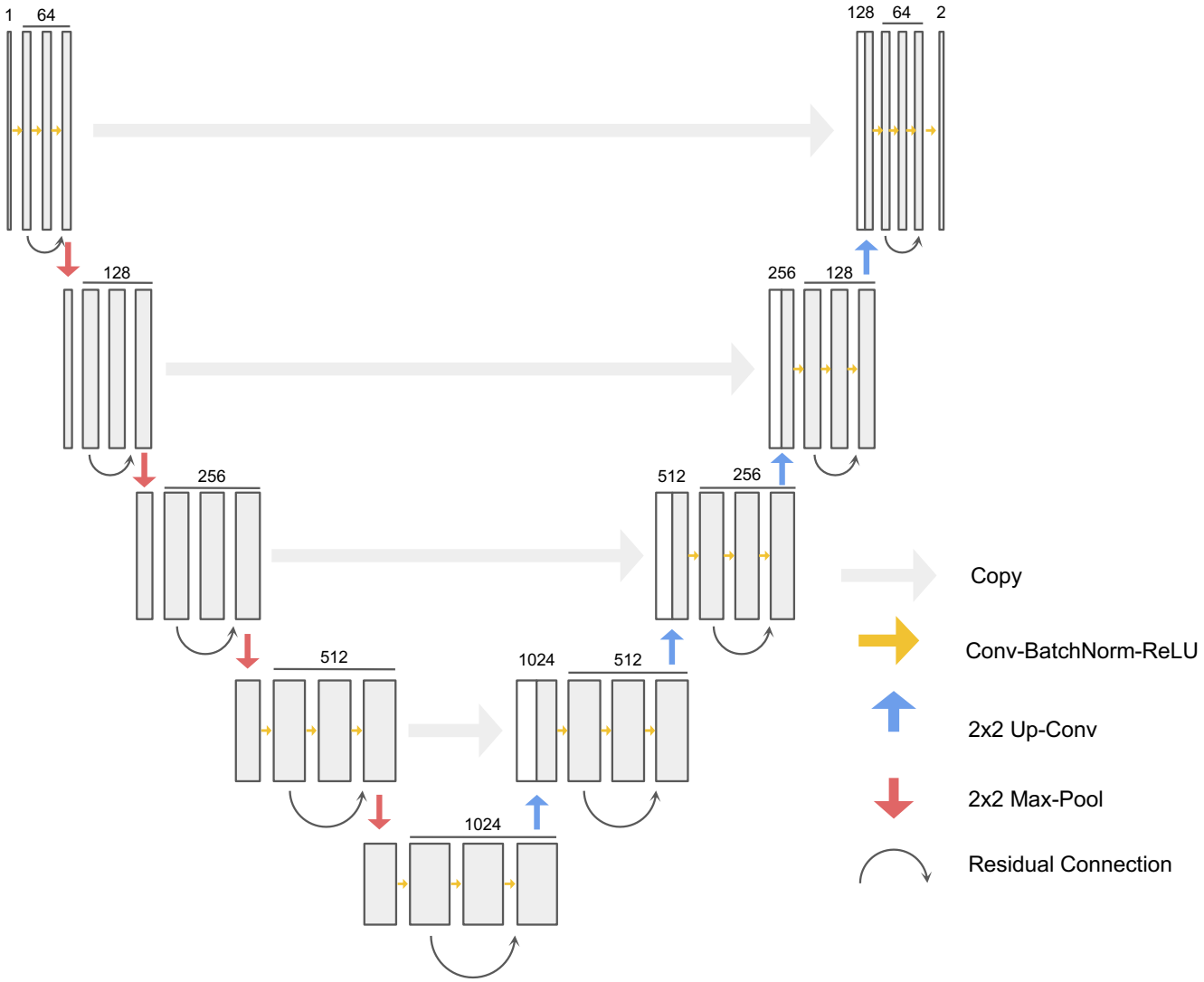

Figure S14. CNN architecture used for the FUSEDEM2MB and ERRNET. The architecture is similar to U-Net (Ronneberger et al., 2015), but has 3 layers of (Convolution, Batch–Normalization, ReLU) in each layer and has additional residual connections (He et al. (2016)). The architecture is fully convolutional and for both FUSEDEM2MB and ERRNET the input dimension is 1, respectively for the grayscale image and the membrane probability. In both cases the output dimension is 2, respectively for 0:not-membrane,1:membrane and 0:no-error,1:error

Dendrite	# Correct Spines	# Type 1 Errors	# Type 2 Errors	# Type 3 Errors	% Correct Spines (excluding Type 3)
1	7	3	2	2	58%
2	9	5	3	10	53%
3	21	6	1	5	75%
Combined	37	14	6	17	65%

Table S1. Distribution of correctly and incorrectly segmented dendritic spines by automated reconstruction.