Fast 3D Electron Microscopy Segmentation via Super-Resolution and Targeted Re-Imaging

Humza Iqbal, SickKids Nanoscale Biomedical Imaging Facility, University of Toronto

Abstract—Three-dimensional electron microscopy (3D EM) enables quantitative analysis of subcellular structure by stacking thousands of ultrathin sections into volumetric reconstructions. While modern segmentation models can map organelles at scale, their accuracy often depends on acquiring high-resolution images, which can be time-consuming and expensive to collect over large tissue areas. This work investigates a hybrid strategy to reduce acquisition burden while preserving downstream segmentation quality. Starting from low-resolution scans, we combine image restoration with learned error localization to selectively re-image only the regions most likely to induce segmentation failures. We evaluate five pipelines spanning classical interpolation and denoising baselines, learned super-resolution, error-guided selective re-imaging, and a combined super-resolution plus re-imaging approach. We train models on mouse liver volumes and transfer the segmentation pipeline to human liver, focusing on mitochondria as an initial organelle target. The resulting framework aims to produce segmentation maps comparable to those obtained from full high-resolution acquisition, while requiring only a fraction of the high-resolution imaging, resulting in order-of-magnitude time saves.

Index Terms—Electron Microscopy, Super-Resolution, Semantic Segmentation, Error Prediction

1 Introduction

Three-dimensional electron microscopy (3D EM) is increasingly used to study cellular structure in intact tissue by reconstructing volumetric regions at nanometer resolution. In a typical serial-section workflow, tissue is sectioned into thousands of ultrathin slices ($35 - 50\,\mathrm{nm}$), each section is imaged with an electron microscope, and the resulting image stack is aligned and segmented into organelle maps. These segmentations enable downstream quantitative analyses that are difficult to perform in 2D, including organelle morphology, proximity measures between structures, and volumetric statistics such as surface area and volume of organelles.

A practical bottleneck in this pipeline is imaging time. For segmentation models to reliably delineate fine boundaries, images often must be acquired at high spatial resolution. In the automated tape-collecting ultramicrotome (ATUM), where sections are organized serially on wafers to be imaged, this can imply scanning thousands of large tissue areas at small pixel sizes, which materially increases both time and cost. As a result, collecting full high-resolution datasets for large fields of view can be prohibitive, particularly when the objective is to quantify specific organelles within large tissue volumes rather than to reconstruct every structure at the highest possible fidelity. Even in the case where a researcher may be interested in a smaller region, there can still be a substantial cost to get all the data at high resolution, over the thousands of sections.

This paper explores an acquisition-efficient alternative: instead of imaging the full field of view at high resolution, we begin with low-resolution scans, apply image processing and learned super resolution, and then selectively re-image only the regions that are most error-prone for the segmentation model. Intuitively, if segmentation failures are spatially localized, then high-resolution acquisition can be concentrated where it matters most. We further hypothesize that super-resolution can reduce the burden on the error detection network by making the low-resolution imagery more similar to the high-resolution domain, enabling selective re-imaging at even larger scale gaps.

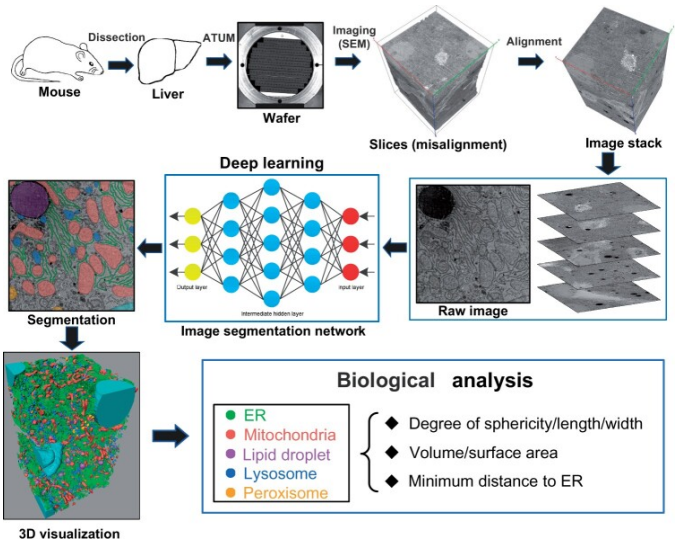

Fig. 1. Overview of the ATUM-SEM to 3D volume pipeline: sectioning, imaging, segmentation, and volumetric reconstruction. [1]

Our contributions are threefold. First, we adapt selective re-imaging ideas to an EM setting where the low-quality versus high-quality distinction is a change in pixel size (for example, $32\,\mathrm{nm}$ versus $8\,\mathrm{nm}$) rather than a change in dwell time at fixed grid resolution, which is a common framing in prior work. Second, we introduce a hybrid pipeline that combines super-resolution with error-guided selective re-imaging, producing a composite image that uses learned restoration globally and true high-resolution acquisition only where needed. Third, we study domain transfer by training on mouse liver and applying the resulting mito-

chondria segmentation pipeline to human liver. In this initial study we focus on mitochondria segmentation in human liver, and we discuss extensions to additional organelles and acquisition policies in Section 5.

2 Related Work

2.1 Machine-learning-guided selective re-imaging in EM

The closest conceptual predecessor to our work is SmartEM, which accelerates electron microscopy by using learning to decide which regions should be rescanned more slowly to improve segmentation quality [2]. SmartEM introduces an error prediction component (ERRNET) that operates on intermediate representations produced from fast-scan imagery, and it uses this predicted error to prioritize slow rescans. A central ingredient is a topology-aware segmentation discrepancy measure based on the variation of information (VI). Given two segmentations S_1 and S_2, SmartEM decomposes VI into merge and split terms,

$$\mathrm{VI}_{\mathrm{merge}} = H(S_1 \times S_2) - H(S_1), \tag{1}$$

$$\mathrm{VI}_{\mathrm{split}} = H(S_1 \times S_2) - H(S_2), \tag{2}$$

$$\mathrm{VI}(S_1, S_2) = \mathrm{VI}_{\mathrm{merge}} + \mathrm{VI}_{\mathrm{split}}, \tag{3}$$

and leverages the additivity of entropy to attribute error contributions to individual objects in the segmentations [2]. In practice, SmartEM uses image processing to isolate the specific border discrepancies responsible for topological differences and trains ERRNET to detect these discrepancies directly from membrane probability maps produced by a separate network [2].

The key difference in our setting is the source of the low-quality versus high-quality image pair. In SmartEM, image pairs differ primarily by dwell time while scanning at the same grid resolution, whereas our goal is to trade image scale itself (for example, $32\,\mathrm{nm}$ pixel size versus $8\,\mathrm{nm}$ pixel size). This changes both the nature of the errors and the operational interface, because the high-resolution image is not merely a cleaner version of the same pixel grid but corresponds to a different sampling lattice that must be fused back into a common coordinate system. In addition, SmartEM was designed primarily around connectomics-style settings that prioritize extremely precise boundary tracing in relatively constrained anatomical contexts, while our target is larger tissue-scale organelle segmentation in liver, where errors may be driven by different visual cues and where selective re-imaging is valuable over broader fields of view.

2.2 Unsupervised pretraining for EM segmentation

Our segmentation backbone is built on RETINA, a reconstruction-pretrained TransUNet variant designed for electron microscopy segmentation on the CEM500K dataset [3]. RETINA leverages a hybrid CNN-transformer encoder and performs self-supervised pretraining by reconstructing augmented versions of unlabeled images. 2D images undergo geometric and photometric augmentations, are encoded, and then decoded back into image space, with reconstruction measured by mean squared error between the reconstructed output and the original image after consistent geometric transforms [3]. This reconstruction-based pretraining aims to encourage robust feature representations that transfer to downstream segmentation tasks when fine-tuned with labeled data [3]. In our study, we fine-tune this backbone on mouse liver patches for mitochondria segmentation and use the resulting models within several downstream acquisition and fusion pipelines.

2.3 Single-image super-resolution

Single-image super-resolution (SR) learns to reconstruct a high-resolution (HR) image from a single low-resolution (LR) observation. In our setting, the LR scan at $32\,\mathrm{nm}$ is mapped to an HR prediction on the $8\,\mathrm{nm}$ grid. We use an EDSR-style model [4], which is a deep residual CNN designed for accurate $\times 4$ upsampling. Given an LR input I^{32}, the network predicts an HR image $\hat{I}^{8} = f_\theta(I^{32})$. Training minimizes an image reconstruction loss against the paired HR target I^{8}; EDSR reports that an ℓ_1 objective is effective for SR and tends to yield better quantitative fidelity than ℓ_2 in their ablations [4].

Architecturally, EDSR uses a stack of residual blocks to progressively refine feature representations while keeping optimization stable. At a high level, each block adds a learned correction to its input features, and the final feature map is upsampled to the target resolution using a learned upsampling module (pixel-shuffle) rather than fixed interpolation [4]. EDSR also removes batch normalization layers, which the authors motivate as beneficial for SR because it reduces memory overhead and avoids constraining feature ranges, allowing the model to better preserve intensity and texture statistics relevant to high-frequency detail [4]. In our pipeline, SR serves as a learned global enhancement step that increases apparent detail in $32\,\mathrm{nm}$ imagery before segmentation, and it is also used as the base canvas for fusion in the hybrid SR + re-imaging approach.

2.4 Uncertainty-driven adaptive acquisition

A complementary thread of work drives adaptive acquisition using uncertainty estimates from a learned model. Ye *et al.* propose uncertainty-based adaptive imaging, where a denoising network predicts both a reconstruction and a pixel-wise uncertainty, and subsequent scans acquire additional measurements only at the most uncertain locations [5]. This policy tightly couples acquisition to model behavior and can yield substantial savings by repeatedly refining the signal in ambiguous regions. Although we do not implement this uncertainty-driven acquisition in our current experiments, we include it here because it provides a principled alternative for selecting re-imaging regions, which we discuss as a potential extension in Section 5.

2.5 Datasets

We train low-resolution, high-resolution, and error localization components using a publicly released mouse liver FIB-SEM volume that is part of the CellMap 2024 Segmentation Challenge [6]. The dataset provides an $8\,\mathrm{nm} \times 8\,\mathrm{nm} \times 8\,\mathrm{nm}$ volume spanning approximately $102.0\,\mu\mathrm{m} \times 101.8\,\mu\mathrm{m} \times$

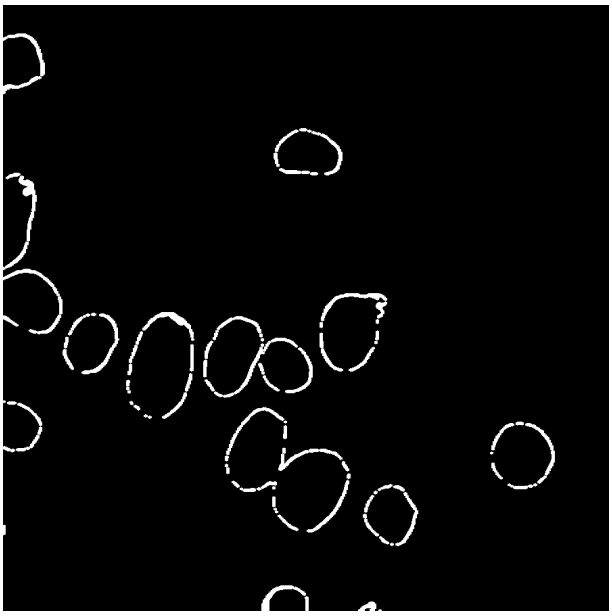

Fig. 2. Example ERRNET output. Error heatmap produced by the error predictor, based on a predicted segmentation map.

$71.5\,\mu\mathrm{m}$, enabling supervised evaluation of organelle segmentations and controlled generation of lower-resolution imagery via downsampling. In addition, we collect paired human liver images at $32\,\mathrm{nm}$ and $8\,\mathrm{nm}$ for training the super-resolution model and for evaluating the transfer of mouse mitochondria segmentation to human tissue.

3 Proposed Method

3.1 Problem setup and notation

Let I^8 denote an image acquired at $8\,\mathrm{nm}$ pixel size (high resolution) and let I^{32} denote an image acquired at $32\,\mathrm{nm}$ pixel size (low resolution). The scale factor between these regimes is $s = 4$ in each spatial dimension. Our goal is to obtain a mitochondria segmentation that matches the quality of a segmentation produced from a fully acquired I^8 image, while reducing the amount of high-resolution acquisition by operating primarily on I^{32} and selectively re-imaging only a subset of pixels at $8\,\mathrm{nm}$.

We denote a segmentation network by $g(\cdot)$, which outputs a mitochondria probability map or label map, and we denote an error prediction network by $e(\cdot)$, which produces a spatial error score intended to localize where the segmentation is most likely to deviate from the high-resolution reference. We denote a super-resolution network by $f(\cdot)$ that maps the low-resolution image to an $8\,\mathrm{nm}$-aligned reconstruction.

3.2 Learned components and training data

We fine-tune the RETINA segmentation backbone [3] separately in two regimes. The first model is trained for low-resolution inference on $32\,\mathrm{nm}$ images, and the second model is trained for high-resolution inference on $8\,\mathrm{nm}$ images. Mouse liver training data are formed by extracting 10 depth chunks in z, sampling 10 slices per chunk (100 slices total), and then generating aligned pairs of $8\,\mathrm{nm}$ and downsampled $32\,\mathrm{nm}$ imagery. Patches are extracted at 244×244 pixels so they fit the segmentation model input. Under this patching, a single $32\,\mathrm{nm}$ patch corresponds to a 4×4 grid (16 patches) at $8\,\mathrm{nm}$ covering the same field of view. In total, we use 11,974 low-resolution patches and 138,219 high-resolution patches for supervised fine-tuning.

We train an EDSR-style super-resolution network [4] using paired human liver imagery acquired at $32\,\mathrm{nm}$ and $8\,\mathrm{nm}$ from the same section. We extract paired patches from these two images to create approximately 10,000 training pairs. The super-resolution model operates directly on the low-resolution input and learns to predict an $8\,\mathrm{nm}$-aligned reconstruction without requiring pre-upsampling.

We train two ERRNET-style error predictors in the mouse domain, one associated with the $32\,\mathrm{nm}$ segmentation regime and one associated with the $8\,\mathrm{nm}$ regime. For each regime, the input to the error predictor is the segmentation output produced by the corresponding segmentation model on imagery at that resolution. Targets are derived from the ground-truth label maps of the mouse liver dataset [6], using an error signal that reflects segmentation disagreement. Inspired by SmartEM [2], error localization is intended to highlight spatial regions that drive topological or object-level discrepancies, providing a map that can be thresholded or ranked to select regions for re-imaging. Empirically, these heatmaps concentrate around mitochondria boundaries, behaving similarly to edge-detector for mitochondria 2.

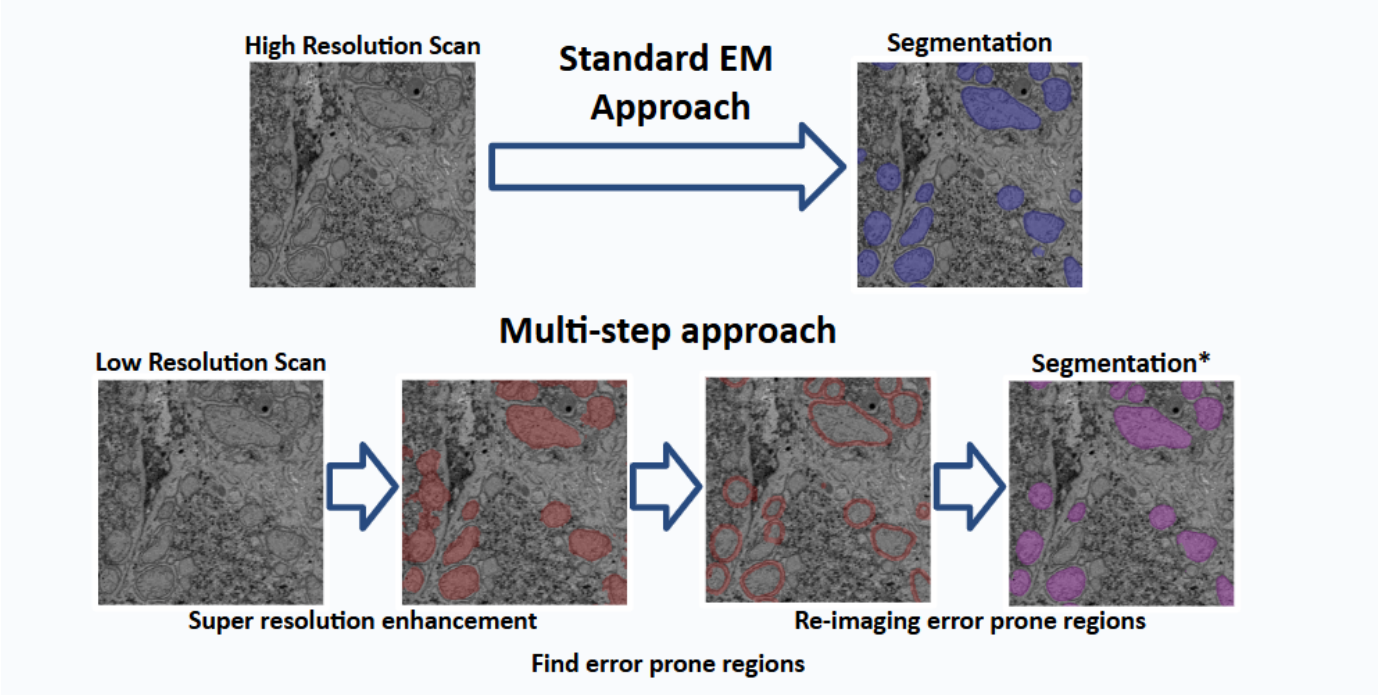

Fig. 3. Standard approach vs our approach: low resolution ($32\,\mathrm{nm}$) scan, super-resolution prediction, error prediction, selective $8\,\mathrm{nm}$ re-imaging, fusion, and final segmentation.

3.3 Selective re-imaging and image fusion

Selective re-imaging proceeds as follows. Starting from a low-resolution scan I^{32}, we compute a segmentation $g_{32}(I^{32})$ and then compute an error heatmap $E^{32} = e_{32}(g_{32}(I^{32}))$. We convert the heatmap into a binary mask $\mathcal{M}$ of pixels to re-image, and we apply morphological dilation with a radius of 10 pixels to include contextual boundaries around predicted errors. The microscope then acquires a high-resolution image $I^8_{\mathcal{M}}$ restricted to $\mathcal{M}$ (with coordinates expressed in the $8\,\mathrm{nm}$ grid). Finally, we overlay this high-resolution data back into an $8\,\mathrm{nm}$-aligned canvas so that a downstream segmentation model can operate on a single composite image.

There are two experiments we did with re-imaging, in one case we run the error network on a segmentation map generated from a low-resolution image. In this case, we bicubic-upsample the low-resolution image to the $8\,\mathrm{nm}$ grid and replace pixels in the upsampled image with the re-imaged high-resolution pixels in $\mathcal{M}$. The other expeirment already upsamples with the super resolution network, so we directly replace pixels in $\hat{I}^8$ with the re-imaged observations $I^8_{\mathcal{M}}$, producing a hybrid image that integrates learned restoration with true high-resolution measurement.

3.4 Experimental pipelines

We evaluate five pipelines that progressively add modeling components.

The bicubic baseline upsamples I^{32} to the $8\,\mathrm{nm}$ using bicubic interpolation and then applies the $8\,\mathrm{nm}$ segmentation model to the upsampled image. This tests whether a purely geometric upsampling step is sufficient to close the gap between acquisition regimes.

The denoise-plus-interpolate baseline applies non-local means denoising to I^{32} to suppress low-resolution noise and texture artifacts, then upsamples by bicubic interpolation, and finally applies the $8\,\mathrm{nm}$ segmentation model. This serves as a classical improvement over interpolation alone, with some noise reduction. EM images are notoriously noisy, so such a simple method for denoising may be destructive.

The super-resolution baseline replaces interpolation with the learned super-resolution network. We compute $\hat{I}^8 = f(I^{32})$ and then segment with the $8\,\mathrm{nm}$ model. This tests whether learned restoration can recover fine mitochondria boundary cues that are lost at $32\,\mathrm{nm}$.

The selective re-imaging pipeline uses I^{32} to find segmentation failures via the $32\,\mathrm{nm}$ segmentation model and the $32\,\mathrm{nm}$ error predictor, re-images those error-prone regions at $8\,\mathrm{nm}$, and overlays these patches onto the $8\,\mathrm{nm}$ image (we upsample the low-resolution image with bicubic interpolation). The final segmentation is performed using the $8\,\mathrm{nm}$ segmentation model.

The hybrid super-resolution plus selective re-imaging pipeline first computes a super-resolved image $\hat{I}^8$, then predicts errors using an $8\,\mathrm{nm}$ segmentation model and an $8\,\mathrm{nm}$ error predictor applied to $\hat{I}^8$, re-images selected regions at $8\,\mathrm{nm}$, and overlays true high-resolution measurements on top of the super-resolved image. The final segmentation is performed using the $8\,\mathrm{nm}$ segmentation model. This pipeline is designed to reduce both the number of re-imaged pixels and the error prediction difficulty by presenting the error model with inputs that better match the high-resolution appearance while still allowing regions of ambiguity to be resolved with true acquisition.

TABLE 1
Quantitative comparison across reconstruction and segmentation metrics. Time saved is reported relative to full high-resolution acquisition; values can be interpreted as approximate speedups of $1/(1 - \text{Time Saved})$.

Method	PSNR	SSIM	IoU	VI	Time Saved (%)
Bicubic	27.12	0.70	0.61	0.32	**94**
NLM + Linear	26.92	0.69	0.47	0.34	**94**
Super Resolution	27.76	0.73	**0.70**	0.22	**94**
Err Reimaging	**28.13**	**0.76**	0.6641	0.27	75
Super Resolution + Err Reimaging	28.059	0.75	0.65	**0.21**	88

4 Results

Table 1 summarizes quantitative performance across five pipelines. We report PSNR and SSIM between each method's reconstructed image and the high-resolution reference, along with mitochondria segmentation agreement measured by IoU and variation of information (VI). The super-resolution baseline achieves the highest IoU (0.70), indicating that learned restoration alone can substantially narrow the resolution gap for mitochondria segmentation. The hybrid super-resolution plus error-guided re-imaging pipeline attains the best VI (0.21), suggesting improved object-level consistency even when overlap-based scores are not maximal. Error-guided re-imaging also improves reconstruction metrics, achieving the best PSNR (28.13) and SSIM (0.76), consistent with directly inserting true high-resolution measurements in selected regions. I also hypothesize that there is a design flaw in the study design, which artificially reduces performance for the re-imaging methods, I expand on this in the discussion.

We also have qualitative examples in Fig. 4, which compares reconstructed images and predicted mitochondria segmentations for two samples. The bicubic baseline often produces visibly over-smoothed reconstructions, and this translates into segmentation artifacts near mitochondria boundaries. A similar boundary-weakening effect is observed in the error-guided re-imaging pipeline, since large portions of the field still rely on upsampled low-resolution context outside the re-imaged mask. Comparing the two re-imaging variants, the SR + error-guided re-imaging pipeline tends to produce smoother, more visually coherent segmentations, but it can also be conservative and occasionally misses complete mitochondria.

5 Discussion

The results in Table 1 show that super-resolution provides the strongest overall improvement in segmentation overlap, achieving the highest IoU while retaining the largest time savings. In contrast, the error-guided re-imaging pipelines do not consistently outperform super-resolution in segmentation quality, despite improving reconstruction fidelity (PSNR and SSIM). This outcome is unexpected because selective re-imaging inserts true high-resolution measurements, which should reduce ambiguity in difficult regions.

A likely explanation is a study-design mismatch between how the error models are trained and how the methods are evaluated. On mouse liver, the error networks are trained using ground-truth labels from the dataset, which we treat as correct and complete for supervision. During evaluation, however, the human liver reference segmentations are produced by the high-resolution segmentation model itself rather than by manual labels. Visual inspection suggests that these high-resolution predictions sometimes include false positives where non-mitochondrial structures are segmented as mitochondria. In such regions, the error network is not incentivized to request re-imaging because, under the evaluation protocol, those regions do not appear erroneous. When re-imaging overlays high-resolution evidence that does not support these false positives, these regions will not be segmented, which can reduce agreement with the model-generated reference and disproportionately penalize re-imaging methods. This mismatch can inflate the apparent performance of super-resolution-only pipelines and reduce measured performance for any pipeline that selectively replaces content, even when the replacement is biologically more plausible.

A direct path to resolving this mismatch is to strengthen the segmentation model so that the reference segmentations used in evaluation are closer to true ground truth. This can be pursued by expanding fine-tuning with additional mouse liver data and other available liver datasets, and by introducing a small curated set of human liver labels to calibrate the model's human-domain behavior. A mixed mouse plus human training strategy could reduce systematic false positives in human tissue and improve evaluation fidelity. The main limitation is reproducibility across new clients and tissue types, because obtaining manual labels requires domain expert effort and may need to be repeated for each tissue and organelle set.

Another improvement is to reduce domain mismatch for the error networks by collecting additional paired human liver data and re-training the error prediction. Here we can train the err-net on human data because of our improved segmentation model. Since the re-imaging decision depends on how the segmenter fails, training the error model on the same domain used at test time should improve mask quality and therefore improve the effectiveness of selective re-imaging. This is feasible given that the human data collection pipeline is now establishedd.

There is also significant headroom in the super-resolution component. The current model is trained on limited paired human data, and improved training diversity, more sections, and modern super-resolution objectives could further improve restoration quality and robustness. Better super-resolution should make error localization easier and reduce the amount of re-imaging required, which is consistent with the hybrid pipeline achieving the best VI in Table 1.

In terms of efficiency, the reported time savings correspond to large nominal speedups. A 94% time savings implies roughly a 16.7$\times$ reduction in acquisition time under

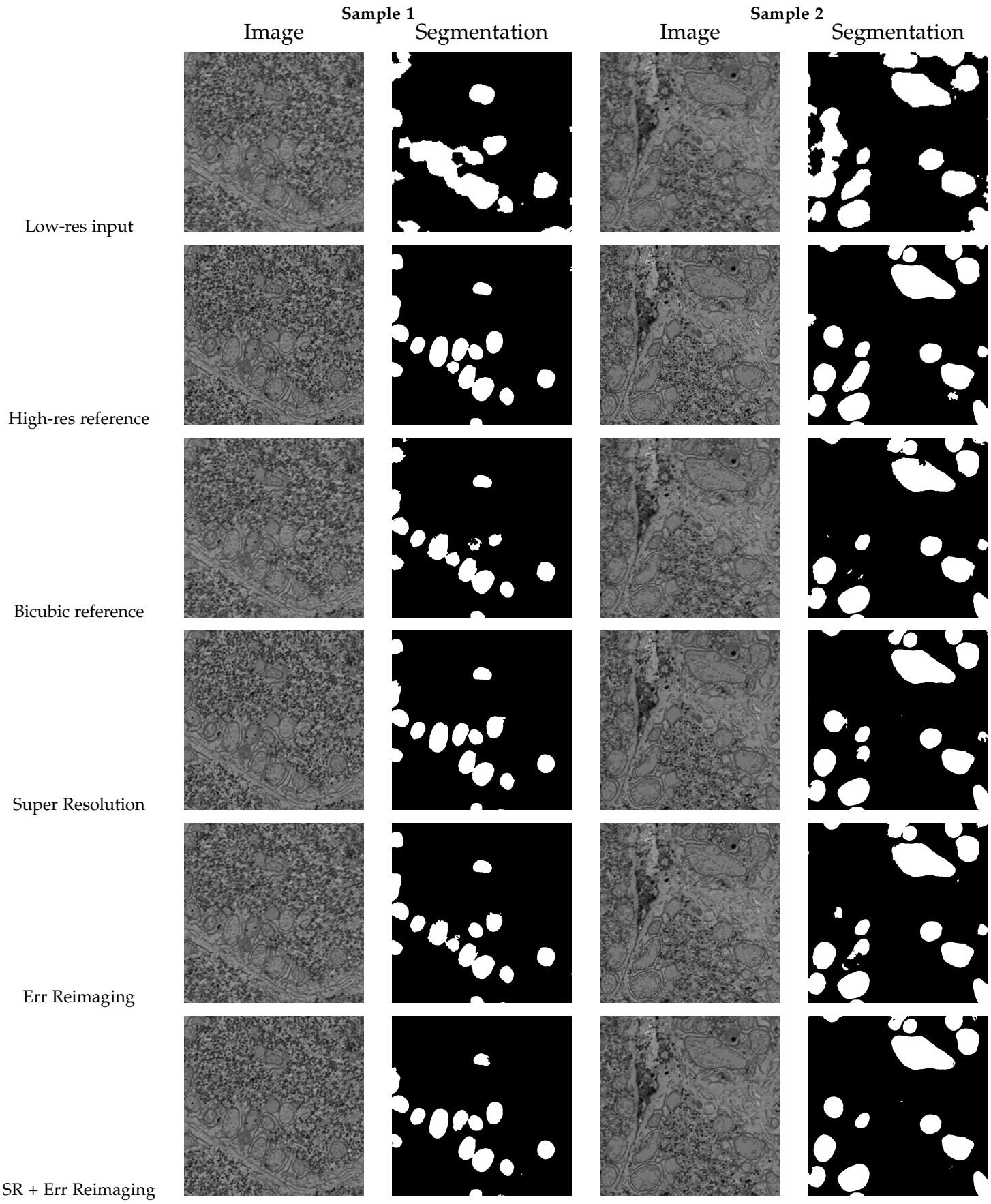

Fig. 4. Qualitative comparison of reconstructed images and predicted mitochondria segmentation maps for two representative samples. Each row shows the method output image and its corresponding segmentation.

an idealized model, while 88% and 75% correspond to approximately 8.3× and 4×, respectively. In practice, a full-section workflow will incur overhead from tiling, stage motion, focusing, and microscope control latency, particularly when re-imaging many small regions. These factors will reduce realized speedups, and measuring end-to-end wall-clock throughput on complete sections and even long series of sections, is also another step to imporvement.

Finally, after addressing these technical details, the natural extension is to generalize from mitochondria to multiple organelle classes, since mouse liver labels are available for a broader set of structures. A subsequent phase is to generalize the pipeline to new tissue types. In a client-facing version of the workflow, a small number of sections from the new tissue could be collected to adapt the super-resolution and error prediction components, assuming an existing segmentation model is available for the target organelles, and then the selective acquisition policy could be applied at scale.

REFERENCES

[1] Y. Jiang, L. Li, X. Chen, J. Liu, J. Yuan, Q. Xie, and H. Han, "Three-dimensional atum-sem reconstruction and analysis of hepatic endoplasmic reticulum–organelle interactions," *Journal of Molecular Cell Biology*, vol. 13, no. 9, pp. 636–645, 2021.

[2] Y. Meirovitch, J. Barthel, N. Yayon, H. Peled, O. Biber, E. Shalev, M. Kirmiz, J. Ullmann, A. Levy, N. Kashtan, J. W. Lichtman *et al.*, "SmartEM: Machine-learning-guided electron microscopy," *bioRxiv*, 2023.

[3] C. Xing, R. Xie, and G. D. Bader, "RETINA: Reconstruction-based pre-trained enhanced TransUNet for electron microscopy segmentation on the CEM500K dataset," *PLOS Computational Biology*, vol. 21, no. 5, p. e1013115, 2025.

[4] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," *arXiv preprint arXiv:1707.02921*, 2017.

[5] C.-T. Ye, Y. Song, Z. He, Y. Xu, W. Yao, Q. Liu, H. Zhang, and G. Chen, "Learned, uncertainty-driven adaptive acquisition for photon-efficient scanning microscopy," *Optics Express*, vol. 33, no. 6, pp. 12 269–12 285, 2025.

[6] FIB-SEM Technology Group, CellMap Project Team, D. Bennett, W.-P. Li, S. Pang, C. S. Xu *et al.*, "jrc_mus-liver: Mouse liver fib-sem volume (cellmap 2024 segmentation challenge)," Janelia Research Campus data release, 2024.