

# SAM4EM: efficient memory-based two stage prompt-free segment anything model adapter for complex 3D neuroscience electron microscopy stacks

Uzair Shah  
CSE, HBKU  
Doha, Qatar

Marco Agus  
CSE, HBKU  
Doha, Qatar

magus@hbku.edu.qa

Daniya Boges  
CEMSE, KAUST  
Thuwal, Saudi Arabia

Vanessa Chappini  
University of Turin  
Turin, Italy

Mahmood Alzubaidi  
CSE, HBKU  
Doha, Qatar

Jens Schneider  
CSE, HBKU  
Doha, Qatar

Markus Hadwiger  
CEMSE, KAUST  
Thuwal, Saudi Arabia

Pierre J. Magistretti  
BESE, KAUST  
Thuwal, Saudi Arabia

Mowafa Househ  
CSE, HBKU  
Doha, Qatar

Corrado Calì  
University of Turin  
Turin, Italy

corrado.cali@unito.it

## Abstract

We present SAM4EM, a novel approach for 3D segmentation of complex neural structures in electron microscopy (EM) data by leveraging the Segment Anything Model (SAM) alongside advanced fine-tuning strategies. Our contributions include the development of a prompt-free adapter for SAM using two stage mask decoding to automatically generate prompt embeddings, a dual-stage fine-tuning method based on Low-Rank Adaptation (LoRA) for enhancing segmentation with limited annotated data, and a 3D memory attention mechanism to ensure segmentation consistency across 3D stacks. We further release a unique benchmark dataset for the segmentation of astrocytic processes and synapses. We evaluated our method on challenging neuroscience segmentation benchmarks, specifically targeting mitochondria, glia, and synapses, with significant accuracy improvements over state-of-the-art (SOTA) methods, including recent SAM-based adapters developed for the medical domain and other vision transformer-based approaches. Experimental results indicate that our approach outperforms existing solutions in the segmentation of complex processes like glia and post-synaptic densities.

## 1. Introduction

In the following we provide further information about the process we carried out for annotating the data (Sec. 2), and additional benchmarks against the most recent SOTA

transformer-based methods (Sec. 3), including a qualitative analysis on noisy microscope data performed by domain scientists participating to the project (Sec. 4).

## 2. Data Curation

We constructed a dataset from high-resolution 3D reconstructions of neuropil ultrastructure in the somatosensory cortex of mice. The dataset incorporates dense reconstructions obtained from serial electron microscopy (EM) images, focusing on axons, dendrites, synapses, and mitochondria, as detailed in [1]. The source data was curated from the publicly available Dryad repository [1].

The dataset comprises three volumetric samples of dimensions  $5\mu m \times 5\mu m \times 5\mu m$ , extracted from layer 1 of the somatosensory cortex of mice. These samples, corresponding to three experimental groups (Mouse 1, Mouse 3, and Mouse 4), were selected based on their morphological diversity and relevance to our analysis. Each sample includes voxelized binary masks representing glia, mitochondria, and post-synaptic densities, enabling comprehensive morphological and connectivity analyses.

To produce the segmentation masks, the high-resolution 3D reconstructions were processed using a combination of semi-automated segmentation tools (e.g., Ilastik and TrakEM2), followed by manual refinement for accuracy. These reconstructions were further voxelized to binary masks using custom scripts developed in Python, ensuring compatibility with downstream analyses. Each voxelized mask retains precise structural details, enabling quantitative



Figure 1. **Mice glia datasets:** full 3D reconstruction of the glia cells from the 3 stacks considered in this project [1]. From left to right: Mouse 1, Mouse 3, and Mouse 4.

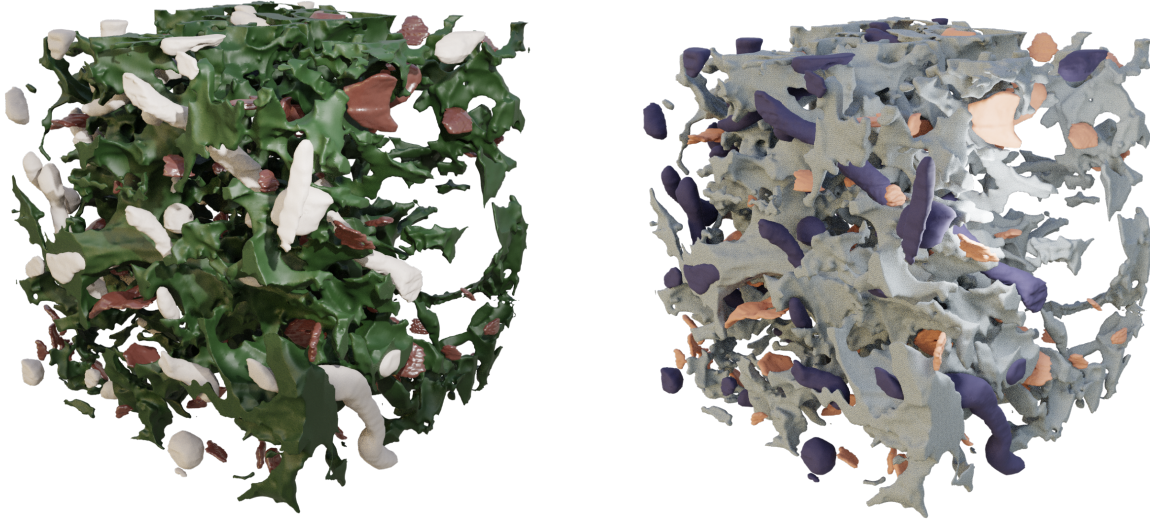


Figure 2. **Mouse 3 rendering:** Full 3D reconstruction of the data in the stack rendered using Blender software. Left: glia (green), mitochondria (white), synapses (pink). Right: glia (light gray), mitochondria (purple), and post-synaptic densities (orange).

analyses of cellular and synaptic elements.

We visualized and validated the dataset using Blender’s advanced rendering capabilities to ensure the fidelity of the segmentation and the representational quality of the data. Fig. 1 shows the full 3D reconstruction of glial cells from the three samples, highlighting the fractal complexity. Fig. 2 illustrates the reconstructed glia, mitochondria, and post-synaptic densities from Mouse 3, highlighting the spatial organization and structural relationships within the sample.

The finalized dataset is publicly available via Dropbox (<https://bit.ly/42k3B1c>) and upon request for reproducibility and future research applications.

### 3. Additional comparisons

To comprehensively evaluate our method’s performance, we conducted additional benchmarking against recent state-

of-the-art transformer architectures in medical image segmentation. While several notable architectures such as AT-Former [6], DualRel [5], and FragViT [4] show promising results in literature, their implementations were not publicly available for direct comparison. Therefore, we focused our comparison on two widely-adopted transformer-based architectures: TransUNet [3] and SwinUNet [2], which have demonstrated comparable performance to the aforementioned methods.

Given the architectural differences and convergence characteristics, we adapted the training protocols accordingly. SwinUNet, being a lightweight architecture, required extended training periods to achieve optimal performance. Specifically, we trained SwinUNet for 350 epochs on the Lucchi dataset and 150 epochs on other datasets, while TransUNet training was capped at 100 epochs across all datasets.

Table 1 presents the quantitative results of this compar-

ative analysis. TransUNet exhibited competitive performance across all datasets, achieving mIoU scores of 56.5%, 70.5%, 37.3%, and 84.8% on Mice-Glia, Mice-Mito, Mice-Synapses, and Lucchi datasets respectively, which are comparable to our proposed SAM4EM. SwinUNet, while being a lightweight architecture, showed relatively lower performance with mIoU scores of 48.4%, 64.1%, 32.9%, and 78.1% across the same datasets. These results demonstrate that while transformer-based architectures can achieve competitive performance in medical image segmentation, our proposed method maintains its advantages in terms of efficiency and consistency across diverse anatomical structures.

## 4. Qualitative analysis

To further challenge our method, we have tested SAM4EM on another EM dataset, acquired using different conditions. Specifically, this dataset was acquired using Serial block-face Electron Microscopy (SBF-EM), whose z-resolution is lower (50 nm-thick sections) compared to the training dataset, acquired with FIB-SEM (12 to 15nm thickness), but the system allows for larger fields of view. In particular we selected a stack from brain parenchyma, containing a cell body, of 50 micrometers width, at very high resolution (10 nm pixel size). The field of view of the training dataset was an order of magnitude smaller (5 micrometers, at 5 nanometers per pixel). Because of the different acquisition methods, images obtained using SBF-EM are qualitatively different, with a slightly higher noise. A domain expert performed a qualitative assessment of the inference on this stack with our proposed SAM4EM against the transformer-based TransUNet [3]. Fig. 3 shows two details extracted from few slices: in red, the glia, in green the mitochondria, and in blue the post-synaptic densities. The domain expert reported the following:

- The classifications obtained with both nets are rather poor, with a lot of objects missing per each category. This might be due, as previously mentioned, by the technical differences of the acquisition methods used to acquire the training datasets.
- Nevertheless, we can say that SAM4EM performed better than TransUNet. In fact, despite the paucity of objects detected, SAM4EM was still able to correctly detect mitochondria, although not all, and the red profile correspond to astrocytic processes. While TransUNet misclassified many of the processes detected, notably with mitochondria, which is basically non-existing, and the detection of glial processes included many neuronal processes.
- Finally both models performed extremely poorly on classification of synaptic densities. Most likely, by training SAM4EM with another, more similar dataset, it is likely that the segmentation might improve significantly, while TransUNet is visibly too faulty.

## 5. Additional Discussions and Clarifications

### 5.1. Memory Efficiency and Scalability

SAM4EM is designed to handle large-scale EM data efficiently during both training and inference. Our model achieves memory efficiency through several key mechanisms:

- **LoRA adaptation:** Reduces trainable parameters by approximately 85% compared to full fine-tuning
- **Efficient memory attention:** Utilizes an 8-slot attention mechanism instead of full self-attention
- **Sliding window processing:** Enables handling of large volumes during inference

In inference mode, SAM4EM successfully processes volumes up to 50m width (approximately 10x larger than training volumes), as demonstrated in our qualitative analysis in Section 4. For context, the model requires approximately 4GB GPU memory for processing 512x512 resolution images, with peak usage not exceeding 6GB.

The method scales linearly with volume size through sliding window processing, maintaining consistent segmentation quality even on larger datasets. This efficiency enables deployment on moderate GPU hardware while retaining high segmentation accuracy on complex neural structures.

### 5.2. Architectural Design Choices

The architectural components of SAM4EM were specifically designed to address the challenges of segmenting complex fractal structures in EM data:

- **Multi-scale feature enhancement:** Processes features at 1/4, 1/8, and 1/16 resolutions to effectively capture both fine details and broader contextual information critical for tracing intricate cellular processes
- **3D memory-based attention:** Ensures structural consistency across consecutive slices, particularly important for maintaining coherence in branching cellular structures
- **Bi-directional refinement mechanism:** Specifically targets irregular boundaries characteristic of glia cells and synaptic junctions

The effectiveness of these design choices is validated by our performance improvements on challenging structures, particularly glia cells (70.5% vs. H-SAM's 68.7% Dice) and synaptic junctions (53.8% vs. H-SAM's 42.3% Dice).

### 5.3. Comparison Methodology

Our evaluation focused on comparing SAM4EM with recent foundation model adaptations (H-SAM, SAMed, UN-SAM) rather than methods requiring full training from scratch. This decision was motivated by our research objective: investigating the potential of fine-tuning foundation models for complex 3D segmentation tasks with limited annotated data.

Table 1. Quantitative comparison across datasets using Dice coefficient (Dice $\uparrow$ ) and mean Intersection over Union (mIoU $\uparrow$ ) metrics. Higher values indicate better performance. Comparing the performance with recent transformer architectures designed for medical image segmentation.

Model	Mice-Glia		Mice-Mito		Mice-Synapses		Lucchi	
	Dice $\uparrow$	mIoU $\uparrow$	Dice $\uparrow$	mIoU $\uparrow$	Dice $\uparrow$	mIoU $\uparrow$	Dice $\uparrow$	mIoU $\uparrow$
TransUNet [3]	71.8	56.5	81.7	70.5	53.6	37.3	91.7	84.8
SwinUNet [2]	63.7	48.4	76.8	64.1	46.3	32.9	85.7	78.1

Traditional segmentation methods would require extensive labeled data for training, whereas our approach leverages the pre-trained knowledge of the SAM foundation model, achieving superior performance through efficient adaptation. This comparison framework better reflects the practical scenario where annotation resources are limited but high segmentation accuracy is required for complex neural structures.

#### 5.4. Future Work

While this work focused on demonstrating SAM4EM’s effectiveness on fully annotated datasets of moderate size, we recognize the potential for further reducing annotation requirements. Preliminary investigations suggest that our approach could significantly reduce the annotation burden through semi-supervised learning and active learning strategies.

Future work will explore:

- Quantifying annotation time savings through interactive segmentation approaches
- Extending the method to handle dense segmentation tasks with multiple cell types and organelles
- Developing specialized data augmentation techniques for electron microscopy data to further reduce annotation requirements

**Acknowledgments.** This publication was funded by the PPM-7th Cycle grant (PPM 07-0409-240041, AMAL-For-Qatar) from the Qatar National Research Fund, a member of the Qatar Foundation. The findings herein reflect the work and are solely the responsibility, of the authors.



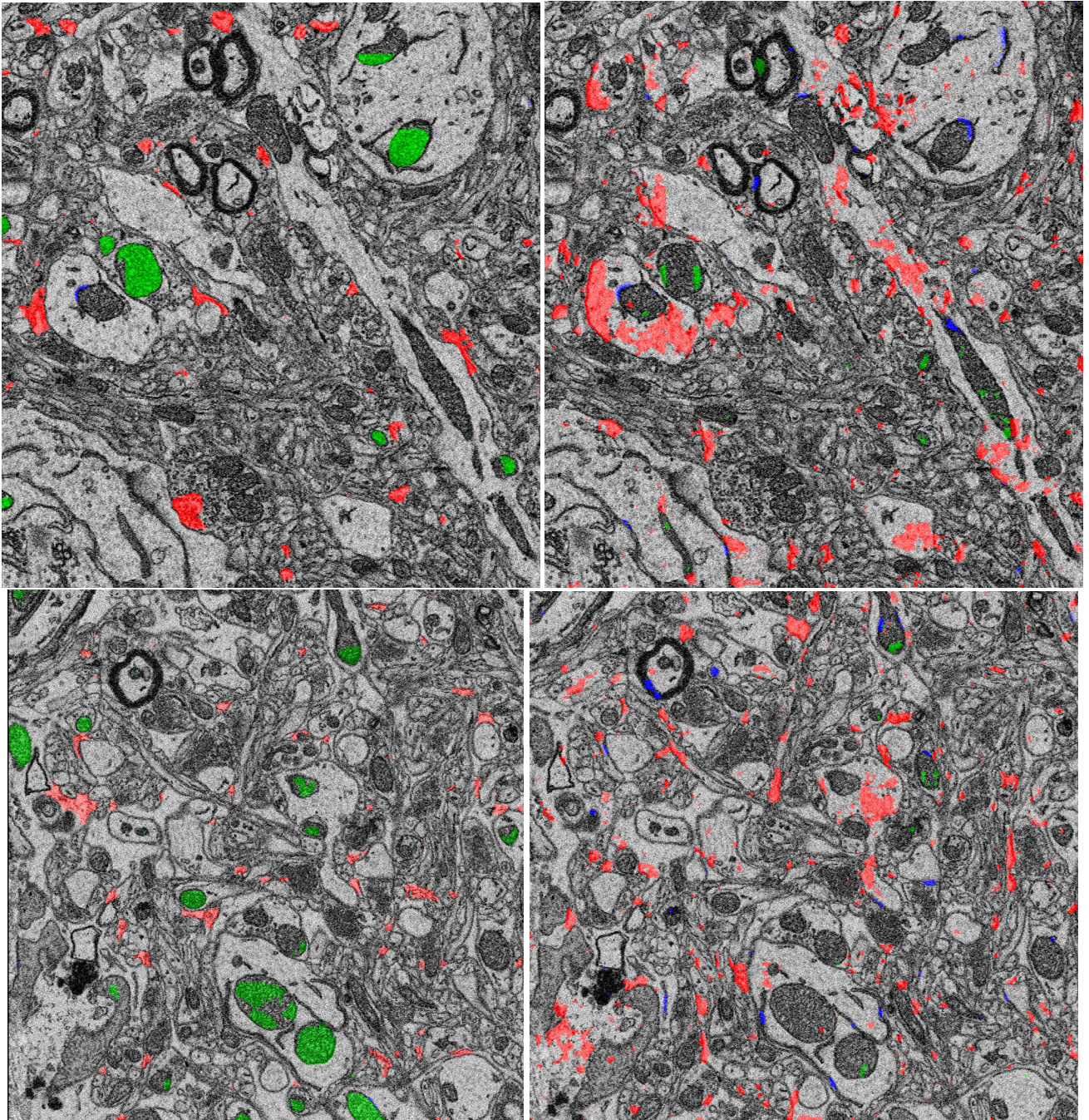


Figure 3. **Qualitative comparison:** a domain expert benchmarked the prpoosed SAM4EM and TransUnet [3] architecture on a stack representing brain parenchyma. From top to botto, two details are shown: on the left, the inference obtained with SAM4EM, on the right the output from TransUnet.



## References

- [1] Corrado Cali, Marta Wawrzyniak, Carlos Becker, Bohumil Maco, Marco Cantoni, Anne Jorstad, Biagio Nigro, Federico Grillo, Vincenzo De Paola, Pascal Fua, et al. The effects of aging on neuropil structure in mouse somatosensory cortex—a 3d electron microscopy analysis of layer 1. *PloS one*, 13(7):e0198131, 2018. [1](#), [2](#)
- [2] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *Proceedings of the European Conference on Computer Vision Workshops(ECCVW)*, 2022. [2](#), [4](#)
- [3] Jieneng Chen, Jieru Mei, Xianhang Li, Yongyi Lu, Qihang Yu, Qingyue Wei, Xiangde Luo, Yutong Xie, Ehsan Adeli, Yan Wang, et al. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. *Medical Image Analysis*, page 103280, 2024. [2](#), [3](#), [4](#), [5](#)
- [4] Naisong Luo, Rui Sun, Yuwen Pan, Tianzhu Zhang, and Feng Wu. Electron microscopy images as set of fragments for mitochondrial segmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(4):3981–3989, Mar. 2024. [2](#)
- [5] Huayu Mai, Rui Sun, Tianzhu Zhang, Zhiwei Xiong, and Feng Wu. Dualrel: Semi-supervised mitochondria segmentation from a prototype perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19617–19626, June 2023. [2](#)
- [6] Yuwen Pan, Naisong Luo, Rui Sun, Meng Meng, Tianzhu Zhang, Zhiwei Xiong, and Yongdong Zhang. Adaptive template transformer for mitochondria segmentation in electron microscopy images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 21474–21484, October 2023. [2](#)