POT-SAM3: Prompt-Only Tuning for One-Shot Electron Microscopy Segmentation

Xuncheng Gu[1] and Li Xiao[1]✉

Beijing University of Posts and Telecommunications, Beijing, China
andrewxiao@bupt.edu.cn

Abstract. Serial-section electron microscopy (EM) supports quantitative analysis of cellular ultrastructure, yet instance segmentation is hindered by costly expert annotations and strong domain shifts. Given the inter-slice continuity, a practical workflow is "few-section annotation + sequential propagation." SAM3 provides video-style detection and tracking-based segmentation that fits cross-section propagation, yet its concept recognition and segmentation generalization in the EM domain are unreliable and it still relies on per-frame prompts. We propose Prompt-Only Tuning for SAM3 (POT-SAM3), a parameter-efficient, low-interaction workflow for serial EM segmentation. POT-SAM3 freezes all SAM3 weights and learns only a small set of EM-target concept prompt tokens from a single annotated section, injecting EM target concepts into SAM3 without updating foundation weights. At inference, these tokens replace point/box/text prompts, enabling prompt-free detection and segmentation on each section, while the SAM3 tracker is used for cross-section mask propagation and candidate addition of newly appearing same-category targets. On three serial-section EM datasets (Lucchi, UroCell, and Guay) under one-shot evaluation, POT-SAM3 updates $< 0.1\%$ parameters and improves the average Dice from 0.464 (SAM3) to 0.770, substantially reducing sequence-level annotation and interaction cost. Project: https://github.com/AI4MyBUPT/POT-SAM3

Keywords: One-shot segmentation · Prompt tuning · Electron microscopy · Segment Anything Model

1 Introduction

Electron microscopy (EM) provides nanoscale visualization of intracellular ultrastructure and underpins studies in cell biology, pathology, and quantitative phenotyping [21]. Instance-level segmentation of targets in EM images (e.g., mitochondria, lysosomes, cells, and other organelles) is a prerequisite for downstream analyses such as counting, morphology measurement, spatial organization statistics, and interaction modeling [26, 9, 24, 15, 14]. In practice, EM segmentation is constrained by two factors: pixel-wise annotation is expensive and

✉ Corresponding author: andrewxiao@bupt.edu.cn

expert-dependent, and variations in sample preparation and imaging settings induce substantial domain shifts [4, 5], making it difficult for models to generalize reliably and remain transferable to new data.

Meanwhile, EM data are often acquired as serial sections, where adjacent slices provide continuous observations of the same 3D structures and thus admit stable cross-slice correspondences [10]. This makes "annotate a few sections and propagate to the entire sequence" a natural and scalable low-annotation workflow [25]. Compared with explicit 3D modeling or volumetric segmentation pipelines which often require reconstruction/registration and incur higher training and computational costs, tracking-based sequential propagation can exploit 3D continuity within a 2D pipeline to maintain cross-slice consistency at much lower engineering and annotation cost [22, 3]. However, under limited annotation, strong domain shift, and long-range propagation, existing methods still struggle to jointly satisfy accuracy, interaction cost, and cross-dataset usability: they either rely on more annotations and retraining, or require frequent per-slice human corrections during propagation, undermining the practical value of this paradigm.

From the perspective of building a universal, low-interaction sequence annotation system, foundation models offer a more scalable path [11, 23, 1, 20]. Built upon *concept prompt tokens*, SAM3 [1] supports an automated "first-frame supervision + sequential propagation" paradigm that aligns well with the continuity of serial-section EM [12], yet its direct application to EM remains unreliable. First, its concept recognition and segmentation generalization are insufficient in the EM domain; second, propagating from a single annotated section often fails to cover newly appearing same-category instances later in the sequence, and correcting them with per-slice point/box/mask prompts [11] would reintroduce substantial interaction cost. Therefore, enabling SAM3 for serial-section EM hinges on two key challenges: **(i)** how to inject EM target concepts into the model to obtain category-aware, "detectable and segmentable" behavior; and **(ii)** how to identify and incorporate candidate newly appearing same-category targets during propagation, thereby reducing reliance on per-slice human prompts.

To address these challenges, we propose POT-SAM3, an efficient automatic segmentation solution built on SAM3 via *Prompt-Only Tuning* for one-shot, prompt-free serial-section EM segmentation with tracking-based propagation. POT-SAM3 freezes all SAM3 parameters and learns only a small set of EM-target *concept prompt tokens* from a single annotated section, enabling one-shot semantic injection without modifying the backbone. During inference, these tokens are used as conditioning inputs to replace the point/box/text prompts required by conventional SAM-style interaction, allowing the SAM3 detector to detect and segment EM targets; the SAM3 tracker is then used for cross-section mask propagation and association, while detector-generated candidates are incorporated to handle newly appearing same-category targets in later sections.

We evaluate on three public serial-section EM datasets—Lucchi [17], UroCell [18], and Guay [6]—under a unified one-shot protocol: only the first section is annotated and no interactive prompts are provided thereafter. POT-SAM3

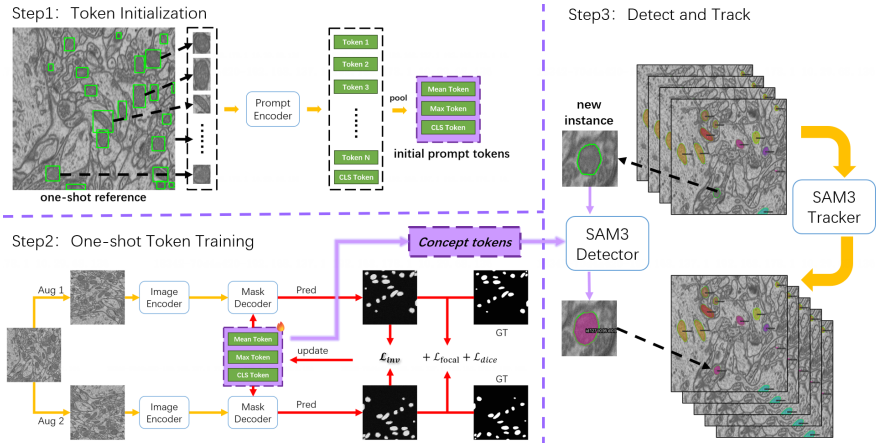


Fig. 1: **Overview of POT-SAM3.** (1) Token initialization—encode instance boxes from the one-shot annotated reference with the SAM3 prompt encoder and pool to obtain Mean/Max/CLS concept tokens; (2) One-shot token training—optimize only the concept tokens with segmentation supervision under two-view augmentations; (3) Detect and track—inject the learned concept tokens into the SAM3 detector for prompt-free instance discovery during tracking.

achieves the best or tied-best performance across all six tasks, improving the average Dice from 0.464 to 0.770 (+0.306) over vanilla SAM3 [1] and consistently outperforming PerSAM [27], SAM2 [23], and HSNet [19]. On Lucchi, POT-SAM3 even surpasses a 10-shot nnU-Net [7] baseline. Overall, by replacing human prompts with learnable concept prompt tokens, POT-SAM3 adapts SAM3's "detection + propagation" capability into a low-interaction annotation workflow for EM sequences.

2 Method

Problem setup. We study one-shot instance segmentation on an image sequence $\mathcal{V} = \{I_t\}_{t=1}^{T}$, where $I_t \in \mathbb{R}^{H\times W}$. Only a single labeled section is provided with instance masks $\mathcal{M}_1^* = \{M_{1,k}^*\}_{k=1}^{N_1}$, $M_{1,k}^* \in \{0,1\}^{H\times W}$, and no annotations are available for $t > 1$. Our goal is to predict a variable-length set of instance masks $\hat{\mathcal{M}}_t = \{\hat{M}_{t,i}\}_{i=1}^{\hat{N}_t}$ for each t, while (i) propagating labeled instances across sections and (ii) completing newly appearing same-category instances during propagation.

As illustrated in Fig. 1, POT-SAM3 learns a small set of *concept prompt tokens* from the single labeled section and injects them into SAM3 to enable automatic instance discovery during tracking [8], without requiring additional human prompts beyond the initial annotation. We denote the key SAM3 modules as an image encoder $f_E(\cdot)$, a prompt encoder $f_P(\cdot)$, an instance decoder/detector

$f_D(\cdot)$, and a propagation tracker $f_T(\cdot)$. Our pipeline contains three stages: (1) **token initialization** from instance boxes derived from the labeled masks (Sec. 2.1); (2) **one-shot token learning** with segmentation supervision and an inverse-warp consistency objective while freezing all SAM3 parameters (Sec. 2.2); and (3) **detection and tracking inference** by injecting the learned tokens into the SAM3 detector and using the SAM3 tracker for propagation and online completion (Sec. 2.3). We learn one token set per class and run the procedure independently for each class.

2.1 Concept Token Initialization

Given the labeled section I_1 and instance masks $\mathcal{M}_1^*$, we compute an axis-aligned bounding box for each instance, $b_k = \mathrm{Box}(M_{1,k}^*)$ (in `xywh` format), forming $\mathcal{B}_1 = \{b_k\}_{k=1}^{N_1}$. We feed $\mathcal{B}_1$ into the frozen prompt encoder to obtain an instance-aware prompt sequence:

$$P_1 = f_P(\mathcal{B}_1) \in \mathbb{R}^{(N_1+1)\times D}, \tag{1}$$

where the first N_1 vectors are box tokens and the last vector is a `cls` token. We denote them as $P_1^{box} \in \mathbb{R}^{N_1\times D}$ and $p_1^{cls} \in \mathbb{R}^D$. We construct three prototypes:

$$t^{cls} = p_1^{cls}, \quad t^{mean} = \mathrm{Mean}(P_1^{box}), \quad t^{max} = \mathrm{MaxPool}(P_1^{box}), \tag{2}$$

and stack them as the initial concept tokens:

$$T^{(0)} = \begin{bmatrix} (t^{cls})^\top \\ (t^{mean})^\top \\ (t^{max})^\top \end{bmatrix} \in \mathbb{R}^{3\times D}. \tag{3}$$

This initialization combines a class-level semantic anchor (`cls`) with instance-level statistics (mean/max), yielding a stable starting point for subsequent token-only learning.

2.2 One-shot Concept Token Learning

We freeze all SAM3 parameters and optimize only the concept tokens T. In each iteration, we sample two augmentations T_a and T_b of the labeled section to obtain two views:

$$I_a = T_a(I_1), \qquad I_b = T_b(I_1). \tag{4}$$

We encode each view with the image encoder, $F_a = f_E(I_a)$ and $F_b = f_E(I_b)$, and obtain token-conditioned prompt embeddings $P_T = f_P(T)$. The detector then produces Q candidate masks and scores:

$$\{(Z_i^a, \ell_i^a)\}_{i=1}^Q = f_D(F_a, P_T), \qquad \{(Z_i^b, \ell_i^b)\}_{i=1}^Q = f_D(F_b, P_T), \tag{5}$$

where Z_i are mask logits and ℓ_i are score logits. We apply sigmoid to obtain $\hat{M}_i = \sigma(Z_i)$ and $\hat{s}_i = \sigma(\ell_i)$. We match candidate masks to ground-truth instances

using Hungarian matching based on soft-IoU [2], and supervise the matched masks and their scores with focal [13] and Dice losses for masks and a focal binary loss for scores. We denote the instance-level supervision as

$$\mathcal{L}_{qry} = \mathcal{L}_{mask} + \mathcal{L}_{score}. \tag{6}$$

To stabilize one-shot learning and improve token reusability, we add auxiliary objectives and optimize:

$$\mathcal{L} = \mathcal{L}_{qry} + \lambda_{sem}\mathcal{L}_{sem} + \lambda_{iw}\mathcal{L}_{iw} + \lambda_{div}\mathcal{L}_{div}. \tag{7}$$

We apply dense semantic supervision $\mathcal{L}_{sem}$ using SAM3's semantic head. To encourage robustness under geometric perturbations, we enforce inverse-warp consistency between semantic predictions from the two augmented views:

$$\mathcal{L}_{iw} = \left\| T_a^{-1}(\hat{S}_a) - T_b^{-1}(\hat{S}_b) \right\|_1, \tag{8}$$

where $\hat{S}_a$ and $\hat{S}_b$ denote semantic predictions for I_a and I_b, and $T^{-1}(\cdot)$ warps predictions back to the original coordinate system. When multiple tokens are used for a class, we include a diversity regularizer $\mathcal{L}_{div}$ based on pairwise cosine similarity to avoid token collapse.

2.3 Inference: Detect and Track

At inference time, we inject the learned concept tokens T as conditioning inputs to SAM3 and perform automatic multi-instance segmentation on the serial sequence without additional human prompts beyond the initial annotation. For each section I_t, we extract visual features $F_t = f_E(I_t)$ and compute token-conditioned prompt embeddings $P_T = f_P(T)$. The SAM3 detector produces candidate instances:

$$\{(Z_{t,i}, \ell_{t,i})\}_{i=1}^{Q} = f_D(F_t, P_T), \tag{9}$$

yielding $\hat{M}_{t,i} = \sigma(Z_{t,i})$ and $\hat{s}_{t,i} = \sigma(\ell_{t,i})$. We filter candidates by confidence thresholding and optionally apply mask-level NMS to de-duplicate proposals, obtaining per-section high-confidence instances.

We maintain a set of active tracks $\mathcal{T}_t$ using the SAM3 tracker:

$$\mathcal{T}_t = f_T\Big(\mathcal{T}_{t-1}, \{(\hat{M}_{t,i}, \hat{s}_{t,i})\}_{i=1}^{Q}\Big), \tag{10}$$

which propagates masks and associates instances across sections to preserve cross-section consistency. To complete newly appearing instances, we treat a high-confidence proposal as a new instance if it has low overlap (measured by IoU) with all existing tracks, and we then add it to $\mathcal{T}_t$. To further recover missing masks in earlier sections, we adopt a bidirectional strategy: we first run forward tracking to discover and follow instances, then run reverse tracking initialized by the discovered instances, and finally fuse the forward and reverse masks to obtain per-section outputs $\hat{\mathcal{M}}_t$.

Table 1: One-shot serial-section EM segmentation results in Dice. Lucchi contains two independent mitochondria sequences. All one-shot methods use annotations from only the first section of each sequence; nnU-Net(10) is trained with 10 annotated sections. Abbreviations: Mito, mitochondria; Fv, fusiform vesicle; A.G., alpha granule.

Method	Lucchi		UroCell		Guay	
	Mito	Mito	Fv	Mito	Cell	A.G.
nnU-Net	0.795	0.751	0.331	0.540	0.895	0.669
nnU-Net(10)	0.900	0.895	**0.551**	**0.830**	**0.930**	**0.775**
HSNet	0.706	0.649	0.198	0.592	0.864	0.400
PerSAM	0.473	0.305	0.148	0.496	0.851	0.347
SAM2	0.717	0.721	0.319	0.473	0.823	0.529
SAM3	0.742	0.551	0.209	0.545	0.540	0.198
POT-SAM3	**0.913**	**0.918**	**0.473**	**0.726**	**0.907**	**0.684**

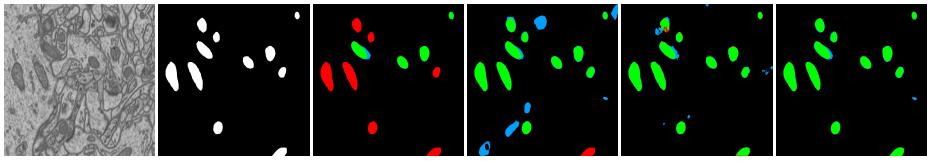
3 Experiments and Results

3.1 Experiments

We evaluate on three public serial-section EM datasets: Lucchi (mitochondria; two 165-section sequences), UroCell (fusiform vesicles and mitochondria; 100 sections), and Guay (cells and α granules; 50 sections). Under a unified one-shot protocol, we annotate only the first section of each sequence and provide no additional prompts thereafter. We also include nnU-Net with 10-shot supervision as a stronger baseline. All inputs are resized to 1004×1004.

Implementation. We implement POT-SAM3 in PyTorch and run all experiments on an NVIDIA A100 40GB GPU using the official SAM3 checkpoint, with bfloat16 AMP enabled. All SAM3 parameters are frozen; only class-specific concept prompt tokens are optimized. We train for 1000 epochs with batch size 2 using AdamW [16] (weight decay = 0), initial learning rate 0.1, and a warmup+cosine schedule (warmup 20 epochs, minimum LR ratio 0.001), with early stopping if the training loss does not improve for 100 epochs. Instance supervision uses Hungarian matching with IoU thresholds (≥ 0.8 positives, < 0.5 negatives) and losses of Focal ($\alpha = 0.8, \gamma = 2.0$)+Dice for masks and focal binary loss for scores, plus inverse-warp consistency ($\lambda_{iw} = 0.1$) and token diversity regularization ($\lambda_{div} = 0.1$). For serial inference with auto detection/tracking, we apply mask-NMS (IoU=0.1), add as new instances proposals with confidence > 0.5 whose IoU with all existing tracks is < 0.5 , and remove trajectories when the per-frame tracking score drops below 0.9.

Baselines. We include nnU-Net as a standard fully supervised baseline, reporting both one-shot and 10-shot results (nnU-Net and nnU-Net(10)). We use HSNet as a representative one-shot segmentation method, and PerSAM as a representative SAM-based one-shot method. We also compare against SAM2 and the unadapted SAM3 baseline to quantify the benefit of prompt-only token tuning.



(a) Frame 127 (b) GT (c) SAM2 (d) SAM3 (e) nnU-Net(10) (f) Ours

Fig. 2: Qualitative Comparison. Green: true positives; Red: false negatives; Blue: false positives.

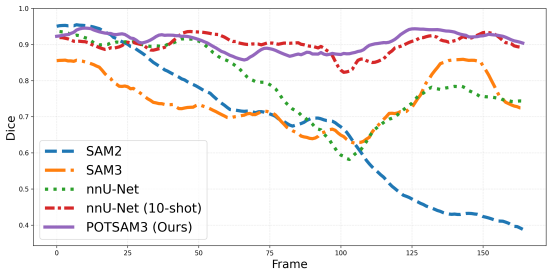


Fig. 3: Frame-wise Dice Comparison

3.2 Results

Table 1 summarizes the one-shot results (Dice) on three serial-section EM datasets. Under a unified protocol where only the first section is annotated and no interactive prompts are provided thereafter, POT-SAM3 achieves the best or tied-best performance across all six tasks. Notably, POT-SAM3 surpasses the 10-shot nnU-Net baseline on both Lucchi mitochondria sequences. These results demonstrate that one-shot concept-token learning can substantially improve SAM3's usability in EM without updating the foundation weights, effectively turning its "detection + propagation" capability into an automatic segmentation solution for serial sections.

We further analyze typical error modes in sequential inference by combining Fig. 2 and Fig. 3. The qualitative results in Fig. 2 show that SAM2 lacks per-frame detection capability and mainly propagates targets that already appear in the first frame, leading to many false negatives (FN) when new instances emerge in later sections. In contrast, SAM3 often produces more false positives (FP) due to unreliable concept recognition of EM targets. While nnU-Net(10) achieves strong quantitative scores, its visual results still exhibit fragmented predictions and coarse boundaries. By comparison, POT-SAM3 enhances the detectability and segmentability of EM targets via concept-token injection and uses tracking-based propagation to encourage cross-section continuity, yielding more stable and coherent segmentations. Consistently, the frame-wise Dice curves in Fig. 3 indicate that SAM2 performs well in early frames (before new instances appear) but degrades as the sequence progresses, whereas POT-SAM3 maintains more

Table 2: Ablation study of POT-SAM3

Tracker	Detector	Token	Aux. loss	Dice ↑
✓	✓			0.742
	✓	✓		$0.869^{\uparrow 0.127}$
✓	✓	✓		$0.889^{\uparrow 0.147}$
✓	✓	✓	✓	$\mathbf{0.913}^{\uparrow 0.171}$

stable performance throughout propagation, demonstrating stronger robustness to long-range propagation and newly appearing instances.

Ablation. Table 2 reports ablation results on the Lucchi dataset. Built on the SAM3 framework, we progressively add four key components: the tracker (Tracker), the detector (Detector), concept-token injection (Token), and auxiliary losses (Aux. loss). *Tracker+Detector* serves as the vanilla SAM3 baseline, which propagates and associates instances using box prompts extracted from the first-frame ground-truth masks, without any token learning or injection. *Detector+Token* disables the tracker and injects the learned concept tokens, using the detector for per-section instance discovery and segmentation to assess the gain from token injection. *Tracker+Detector+Token* further enables tracking-based propagation with token injection, while training tokens with only standard instance-level supervision (Focal+Dice) to evaluate the base "concept injection + sequential propagation" variant. *Tracker+Detector+Token+Aux. loss* is the full model, which incorporates auxiliary losses to improve the stability and reusability of one-shot token learning and achieves the best performance. Overall, Dice increases steadily as components are introduced, indicating that token injection, tracking/propagation, and auxiliary objectives contribute to concept alignment, cross-section consistency, and training stability, respectively.

4 Conclusion

We study one-shot serial-section electron microscopy segmentation with tracking-based propagation, where only a single section is annotated and no additional interactive prompts are available for the remaining sections. Across Lucchi, UroCell, and Guay, POT-SAM3 achieves the best or tied-best Dice on all six tasks. By freezing all foundation weights and learning only a small set of concept prompt tokens via Prompt-Only Tuning, POT-SAM3 mitigates EM concept mismatch and adapts SAM3's detection-and-propagation capability into a low-interaction "single-section supervision + sequential propagation" workflow. Our current evaluation focuses mainly on Dice-based segmentation quality; object-level and tracking-level assessment, including AP/PQ, split/merge analysis, identity consistency, and track fragmentation, remains an important direction for future work, together with more robust token learning and candidate completion/association under strong domain shift. Overall, POT-SAM3 provides a practical prompt-free workflow for serial-section EM, reducing sequence-level annotation and interaction burden while preserving strong Dice-based segmentation performance.

Acknowledgments. This research was supported by the Fundamental Research Funds for the Beijing University of Posts and Telecommunications under Grant 2025AI4S19, and the Beijing Municipal Natural Science Foundation under Grants L252199, L242134, and L242059.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Carion, N., Gustafson, L., Hu, Y.T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K.V., Khedr, H., Huang, A., et al.: Sam 3: Segment anything with concepts. arXiv preprint arXiv:2511.16719 (2025)
2. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European conference on computer vision. pp. 213–229. Springer (2020)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
4. Dorkenwald, S., Schneider-Mizell, C.M., Brittain, D., Halageri, A., Jordan, C., Kemnitz, N., Castro, M.A., Silversmith, W., Maitin-Shephard, J., Troidl, J., et al.: Cave: Connectome annotation versioning engine. Nature methods **22**(5), 1112–1120 (2025)
5. Guan, H., Liu, M.: Domain adaptation for medical image analysis: a survey. IEEE Transactions on Biomedical Engineering **69**(3), 1173–1185 (2021)
6. Guay, M.D., Emam, Z.A., Anderson, A.B., Aronova, M.A., Pokrovskaya, I.D., Storrie, B., Leapman, R.D.: Dense cellular segmentation for em using 2d–3d neural network ensembles. Scientific reports **11**(1), 2561 (2021)
7. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
8. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: European conference on computer vision. pp. 709–727. Springer (2022)
9. Kichuk, T., Dhamankar, S., Malani, S., Hofstadter, W.A., Wegner, S.A., Cristea, I.M., Avalos, J.L.: Using miter for 3d analysis of mitochondrial morphology and er contacts. Cell Reports Methods **4**(1) (2024)
10. Kievits, A.J., Lane, R., Carroll, E.C., Hoogenboom, J.P.: How innovations in methodology offer new prospects for volume electron microscopy. Journal of Microscopy **287**(3), 114–137 (2022)
11. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)
12. Li, X., Loy, C.C.: Video object segmentation with joint re-identification and attention-aware mask propagation. In: Computer Vision – ECCV 2018. Lecture Notes in Computer Science, vol. 11207, pp. 93–110. Springer (2018). https://doi.org/10.1007/978-3-030-01219-9_6

13. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE international conference on computer vision. pp. 2980–2988 (2017)
14. Liu, L., Wu, H., Yang, S., Yi, K., Hu, J., Xiao, L., Xu, T.: Using deepcontact with amira graphical user interface. STAR protocols **4**(4) (2023)
15. Liu, L., Yang, S., Liu, Y., Li, X., Hu, J., Xiao, L., Xu, T.: Deepcontact: High-throughput quantification of membrane contact sites based on electron microscopy imaging. Journal of Cell Biology **221**(9), e202106190 (2022)
16. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
17. Lucchi, A., Smith, K., Achanta, R., Knott, G., Fua, P.: Supervoxel-based segmentation of mitochondria in em image stacks with learned shape features. IEEE transactions on medical imaging **31**(2), 474–486 (2011)
18. Mekuč, M.Ž., Bohak, C., Hudoklin, S., Kim, B.H., Kim, M.Y., Marolt, M., et al.: Automatic segmentation of mitochondria and endolysosomes in volumetric electron microscopy data. Computers in biology and medicine **119**, 103693 (2020)
19. Min, J., Kang, D., Cho, M.: Hypercorrelation squeeze for few-shot segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6941–6952 (2021)
20. Noh, S., Lee, B.D.: A narrative review of foundation models for medical image segmentation: zero-shot performance evaluation on diverse modalities. Quantitative Imaging in Medicine and Surgery **15**(6), 5825–5858 (2025)
21. Peddie, C.J., Genoud, C., Kreshuk, A., Meechan, K., Micheva, K.D., Narayan, K., Pape, C., Parton, R.G., Schieber, N.L., Schwab, Y., et al.: Volume electron microscopy. Nature Reviews Methods Primers **2**(1), 51 (2022)
22. Popovych, S., Macrina, T., Kemnitz, N., Castro, M., Nehoran, B., Jia, Z., Bae, J.A., Mitchell, E., Mu, S., Trautman, E.T., et al.: Petascale pipeline for precise alignment of images from serial section electron microscopy. Nature communications **15**(1), 289 (2024)
23. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024)
24. Reagan, B.C., Kim, P.J.Y., Perry, P.D., Dunlap, J.R., Burch-Smith, T.M.: Spatial distribution of organelles in leaf cells and soybean root nodules revealed by focused ion beam-scanning electron microscopy. Functional Plant Biology **45**(2), 180–191 (2018)
25. Takaya, E., Takeichi, Y., Ozaki, M., Kurihara, S.: Sequential semi-supervised segmentation for serial electron microscopy image with small number of labels. Journal of Neuroscience Methods **351**, 109066 (2021)
26. Wei, D., Lin, Z., Franco-Barranco, D., Wendt, N., Liu, X., Yin, W., Huang, X., Gupta, A., Jang, W.D., Wang, X., et al.: Mitoem dataset: large-scale 3d mitochondria instance segmentation from em images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 66–76. Springer (2020)
27. Zhang, R., Jiang, Z., Guo, Z., Yan, S., Pan, J., Ma, X., Dong, H., Gao, P., Li, H.: Personalize segment anything model with one shot. arXiv preprint arXiv:2305.03048 (2023)