

Prompt-DAS: Annotation-Efficient Prompt Learning for Domain Adaptive Semantic Segmentation of Electron Microscopy Images

Jiabao Chen¹, Shan Xiong¹, and Jialin Peng^{1(✉)}

College of Computer Science and Technology, Huaqiao University, China
2004pj1@163.com

Abstract. Domain adaptive segmentation (DAS) of numerous organelle instances from large-scale electron microscopy (EM) is a promising way to enable annotation-efficient learning. Inspired by SAM, we propose a promptable multitask framework, namely Prompt-DAS, which is flexible enough to utilize any number of point prompts during the adaptation training stage and testing stage. Thus, with varying prompt configurations, Prompt-DAS can perform unsupervised domain adaptation (UDA) and weakly supervised domain adaptation (WDA), as well as interactive segmentation during testing. Unlike the foundation model SAM, which necessitates a prompt for each individual object instance, Prompt-DAS is only trained on a small dataset and can utilize full points on all instances, sparse points on partial instances, or even no points at all, facilitated by the incorporation of an auxiliary center-point detection task. Moreover, a novel prompt-guided contrastive learning is proposed to enhance discriminative feature learning. Comprehensive experiments conducted on challenging benchmarks demonstrate the effectiveness of the proposed approach over existing UDA, WDA, and SAM-based approaches.

Keywords: Domain adaptive segmentation · Weak supervision · Electron microscopy · Mitochondria · Promptable learning.

1 Introduction

Accurate semantic segmentation of subcellular organelles, e.g., mitochondria, from various types of large-scale electron microscopy (EM) sequences is essential for cancer research and biology study [9]. Although deep neural networks including convolutional neural networks [13] and vision transformers (ViTs) [3] have revolutionized the field of semantic segmentation for nearly all applications, including EM image segmentation [4,10,12,17,19], the existing deep neural network models necessitate extensive pixel-wise annotations, involving expensive annotation budgets by experts. Furthermore, they typically show significant performance deterioration when applied directly to image datasets exhibiting different distributions. This challenge is particularly relevant in the context of EM images, which experience significant domain shifts attributable to variations in mi-

croscopy techniques and tissue types. Manually annotating numerous organelle instances from large-scale EM images is time-consuming and labor-intensive.

To reduce the burden of annotating each domain, we explore domain adaptation, which aims to reuse a well-trained model on a given source domain and adapt it to a target domain with a different distribution. Although the unsupervised domain adaptation (UDA) completely assumes no annotation on the target domain, UDA methods still show relatively low performance on complicated tasks, which prohibits their practical usage. To alleviate this issue, we leverage sparse points as [11] on the target domain as cheap weak labels to boost the segmentation performance with minimal annotation effort. In other words, we consider weakly supervised domain adaptation (WDA) with the same setting as WDA-Net [11]. Compared to full point annotation for all object instances and pixel-wise annotation, the partial points demand substantially less time and expert knowledge [11]. Thus, annotating sparse points on a small number of object instances in EM images can be easily completed by non-experts.

Recently, prompt-driven foundation models have shown remarkably strong generalization ability without training on specific targets, intriguing a trend toward more flexible segmentation paradigms. Notably, SAM [6], which is pre-trained on billion-scale datasets of natural images, has demonstrated impressive performance on various segmentation tasks with points, boxes, or masks as user-generated prompts. These promptable segmentation models also pave the way for longstanding interactive segmentation, which can be responsive to user intention or progressively refine the segmentation, guided by the user input.

However, SAM still has several limitations. First, SAM still struggles with domain shifts and usually shows low performance on medical image tasks, especially with point prompts, due to the lack of medical knowledge, ambiguous boundaries, and complex shapes. To enhance the performance, several studies [2,16,20], have proposed modifying or fine-tuning SAM using medical data, such as SAM-Med2D [2], Med-SA [16]. Second, SAM lacks the functionality to segment all object instances of the same class without prompts on all instances, making it particularly challenging to segment numerous organelle instances from EM images. Third, SAM exhibits lower performance when using points as prompts, especially for medical images, as several studies [18] have also shown.

Inspired by SAM, we introduce Prompt-DAS, a promptable transformer model for domain-adaptive segmentation of EM images. Our model is flexible enough to utilize prompts during both the training and testing stages, offering advantages over previous UDA and WDA methods. To achieve a minimal annotation burden, we use sparse points as cost-effective prompts for the semantic segmentation of all object instances within the EM images. Unlike SAM-based models, which require training on billion-scale datasets, we adapt a model trained on a source domain from scratch to a new target domain that already has sparse points available for the target training data and does not assume the availability of prompts during the testing stage. However, when point prompts are available during the testing stage, the output of our method can further align with user intent. Moreover, our model conducts one-pass segmentation of

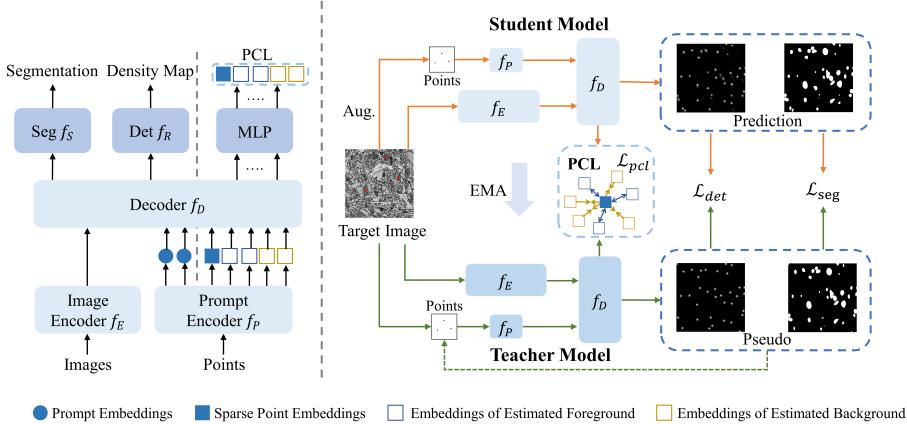


Fig. 1: Overview of our Prompt-DAS model for domain adaptive segmentation.

many object instances with any point prompts, presenting an advantage over SAM. To enhance discriminative feature learning, we introduce a novel prompt-guided contrastive learning. Comprehensive experiments conducted on challenging benchmarks demonstrate the effectiveness of the proposed approach.

2 Method

Problem. Suppose a source domain $\mathcal{D}^s = \{(x^s, y^s)\}$ with full pixel-wise labels y^s , and a target domain $\mathcal{D}^t = \{(x^t, \bar{c}^t)\}$ with point labels on centers of a few object instances. The binary dot label map \bar{c}^t takes 1 only at the annotated sparse points. Additionally, the full dot label of a pixel-wise label y , denoted as c , has a corresponding density map d that is obtained through convolution with a Gaussian kernel k_σ , expressed as $d = k_\sigma * c$. Our objective is to develop a model that is flexible enough to perform UDA, WDA, as well as interactive versions of UDA and WDA. The model should also be flexible enough to be capable of effectively utilizing both training and testing prompts when provided.

Overview. Figure 1 illustrates our Prompt-DAS, which encompasses an image encoder f_E , a point prompt encoder f_P that processes $M \geq 0$ points at once as inputs, a multitask decoder f_D followed by a semantic segmentation head f_S , and a regression-based center-point detection head f_R . In scenarios where $M = 0$ points are provided on target training data, our model operates as UDA, referred to as Prompt-DAS (0%). In scenarios where $M > 0$ points are given on the target training data, our model conducts WDA learning. By default, we use 15% sparse points, and our model is designated as Prompt-DAS (15%). When prompt points are provided during the testing phase, our model is capable of executing interactive UDA/WDA segmentation, denoted as Prompt-DAS+.

To tackle the issue of label scarcity on the target domain during domain adaptation learning, we conduct pseudo-label learning for both the segmenta-

tion and detection tasks under the mean-teacher framework [14]. The output of the detection head f_R is used to provide prompts for the segmentation task. Furthermore, the segmentation head f_S is guided by a prompt-based contrastive loss, enhancing the discriminability of prompt embeddings.

Promptable Detection. The proposed Prompt-DAS utilizes an auxiliary detection task to enhance the segmentation learning. The center point detection task is relatively easier than the dense segmentation task, particularly given sparse points as training prompts and partial supervision. While the joint learning of multiple tasks can implicitly boost the segmentation performance, confident detection outputs are further employed to augment the ground-truth point prompts for the segmentation task. Following the teacher-student framework [14], pseudo-labels for the unlabeled regions are generated by selecting most highest local maxima points with a threshold from the predicted density map by the teacher model, which is updated by the exponential moving average of the student network. Note that local maxima points can be identified through Non-Maxima Suppression. For the target data, student network training is supervised by both the ground-truth sparse points and pseudo labels, and the M sparse points are also used as training prompts. In the scenario of UDA, where there are no point annotations on the target domain, we use the estimated confident points from the prediction of the source model as the pseudo sparse points. For the source data, ground truth center points are used as the supervision, and randomly sampled n_s center points are used as training prompts.

$$\mathcal{L}_{det} = \frac{1}{|\mathcal{D}^s|} \sum_{x^s} MSE(F_R(x^s), d^s) + \frac{1}{|\mathcal{D}^t|} \sum_{x^t} MSE(F_R(x^t), \hat{d}^t) \quad (1)$$

where $F_R = f_R \circ f_D \circ f_E$, MSE represents mean square error loss, and \hat{d}^t represents the density map generated by the target pseudo labels and ground truth points.

Promptable Segmentation. To alleviate label scarcity, we leverage pseudo-labeling in the teacher-student framework. Thus, both ground truth source labels and target pseudo-labels are used to supervise the model training.

$$\mathcal{L}_{seg} = \frac{1}{|\mathcal{D}^s|} \sum_{x^s} CE(F_S(x^s), y^s) + \frac{1}{|\mathcal{D}^t|} \sum_{x^t} CE(F_S(x^t), \hat{y}^t) \quad (2)$$

where $F_S = f_S \circ f_D \circ f_E$, CE represents the standard cross-entropy loss, \hat{y}^t represents the pseudo labels generated by the teacher model on the target domain.

Similar to the detection head, we also use points as the segmentation training prompts. For the source domain, we use the n_s points sampled for the detection task as the training prompts. Note that n_s is a random number during training. Since the target data only has a few points as the annotation, we propose to use both the estimated points from the detection output and ground-truth sparse points as training prompts to assist the segmentation training. The target point prompts are generated by selecting most highest local maxima points with a threshold from the predicted density map by the detection head.

Prompt-guided Contrastive Learning (PCL). To learn more discriminative embeddings during pseudo-label learning, we further introduce contrastive

learning with the guidance of prompts, which can provide representative features to distinguish mitochondria instances from the background organelle. As shown in the left figure of Fig. 1, our contrastive learning aims to pull the feature embeddings of the estimated foreground points closer to that of the ground-truth sparse points while simultaneously pushing away from the foreground embeddings from the background embeddings. An MLP layer ϕ is utilized before conducting contrastive learning. Let $z^t = \phi(f_D(f_P(p^t)))$ denote the embedding derived from the target domain point p^t . We employ an attention mask mechanism following DN-DETR [8] to prevent information leakage from PCL.

Queries are generated from pixels identified as foreground exhibiting a sufficiently high confidence. Utilizing the pseudo-labels produced by the teacher model, we select three points from each instance with a confidence greater than δ_f , resulting in N^q foreground prompt embeddings $\{z_i^t\}_{i=1}^{N^q}$. Concurrently, we identify N^n points with a confidence level below δ_b , resulting in N^n background prompt embeddings $\{\mu_k^b\}_{k=1}^{N^n}$. Since mitochondrial instances display high similarity, we employ the average embedding of sparse point prompts as the sparse prompt embedding μ^f . The prompt-guided contrastive loss is defined as follows:

$$\mathcal{L}_{pcl} = - \sum_{i=1}^{N^q} \log \left[\frac{\exp(\mu^f \cdot z_i^t / \tau))}{\exp(\mu^f \cdot z_i^t / \tau)) + \sum_{k=1}^{N^n} \exp(\mu_k^b \cdot z_i^t / \tau)} \right] \quad (3)$$

In our experiments, we set $N^n = 256$, $\delta_f = 0.9$, and $\delta_b = 0.1$.

3 Experiments

Benchmark and Metrics. We evaluate the proposed method using the MitoEM dataset [15], which is 3600 times larger than previous datasets and presents a greater challenge due to the wide diversity of mitochondria in terms of shape and density. This dataset comprises two volumes of $30 \times 30 \times 30 \text{ } \mu\text{m}^3$ derived from the temporal lobe of an adult human and the primary visual cortex of an adult rat. The two datasets are named MitoEM-Human and MitoEM-Rat. The MitoEM-Human dataset contains significantly more mitochondria instances and a higher number of small mitochondria instances compared to the MitoEM-Rat dataset. Both datasets consist of 500 images of size 4096×4096 , where 400 images are allocated for training and 100 images for testing on each dataset. We consider the cross-domain segmentation between MitoEM-Human and MitoEM-Rat. We evaluate our method using the semantic-level Dice similarity coefficient (Dice) and the instance-level Aggregated Jaccard Index (AJI) [7], as well as Panoptic Quality (PQ) [5].

Implementation Details. We use the pre-trained ViT-S/8 with DINO [1] as our image encoder. Our decoder f_D is similar to that of SAM but with mask attention and cross-attention to prevent information leakage. In contrast to SAM, our prompt encoder is a standard positional embedding. Our MLP is the same as the MLP of SAM and other standard transformers. More details can be found in our released code <https://github.com/JiabaoChen1/Prompt-DAS>. The model

Table 1: Comparison results. We compare our Prompt-DAS under different settings with UDA, WDA, and SAM-based approaches. For WDA, sparse points on target training data are used as training prompts to achieve minimal annotation efforts. For interactive segmentation, indicated by a "+", all center points of the testing data are used as testing prompts to fulfill the requirements of SAM.

Methods	Prompts		Type	Human → Rat			Rat → Human		
	Training	Testing		Dice	AJI	PQ	Dice	AJI	PQ
SAM [6]	NoAdapt			32.0	14.3	30.0	20.8	11.4	18.7
SAM-Med2D [2]				15.9	-	-	22.5	-	-
Med-SA [16] [†]				75.5	56.6	27.5	72.4	55.0	33.4
Our Source Model				88.6	76.7	68.7	78.1	62.8	55.5
SAM+ [6]	(Interact.)	✓		40.6	1.2	26.2	40.3	4.6	26.6
SAM-Med2D+ [2]		✓	NoAdapt	72.6	55.6	39.7	78.1	61.2	42.2
Med-SA+ [16]		✓	(Interact.)	86.2	70.2	59.9	83.8	68.1	59.0
Our Source Model+		✓		89.8	78.8	74.1	87.6	76.0	70.6
DAMT-Net [10]	UDA			88.7	76.3	61.8	85.4	72.3	63.7
UALR [17]				86.3	71.6	53.7	83.8	69.7	60.0
DA-ISC [4] (2.5D)				88.6	75.7	65.8	85.6	72.7	63.8
CAFA [19] (2.5D)				89.2	-	-	86.6	-	-
WDA-Net (0%) [11]				88.2	74.5	59.0	85.5	72.3	60.6
Prompt-DAS (0%)				92.4	82.2	74.3	88.0	76.6	68.1
WeSAM (15%) [20] [‡]	WDA	✓		7.5	0.1	2.2	3.6	1.3	1.6
WDA-Net (15%) [11]		✓		91.7	80.7	74.0	88.7	77.6	67.8
Prompt-DAS (15%)		✓		93.3	83.6	74.5	89.2	78.6	69.1
WeSAM (15%)+ [20] [‡]	(Interact.)	✓	✓	89.9	79.6	73.9	82.3	66.0	65.5
Prompt-DAS(15%)+		✓	✓	93.5	84.4	74.2	90.8	81.5	72.3
Supervised model			Oracle	94.6	86.4	79.2	92.6	84.6	75.8

[†] Fine-tuning using the source data

[‡] Fine-tuning using the source data and target data with 15% sparse point labels

is trained for 16k iterations with a batch size of 2, using the AdamW optimizer with an initial learning rate of 1×10^{-5} . The input images are subjected to random cropping, resulting in a size of 384×384 pixels. We use a polynomial decay of power 0.9 to control the learning rate decay. For a fair comparison, the same data augmentations as those in WDA-Net are used. During the inference phase, we apply a sliding window with the same resolution as used during training. The implementation is conducted using PyTorch, and our model is trained for 6 hours on one RTX 4090 GPU with 24 GB of memory.

Quantitative Evaluations. In Table 1, we compare our Prompt-DAS model with six different types of models. 1) Methods without domain adaptation (NoAdapt): SAM-based models without testing prompts, including SAM [6], SAM-Med2D [2], and Med-SA[†] [16], where [†] means training using our source EM

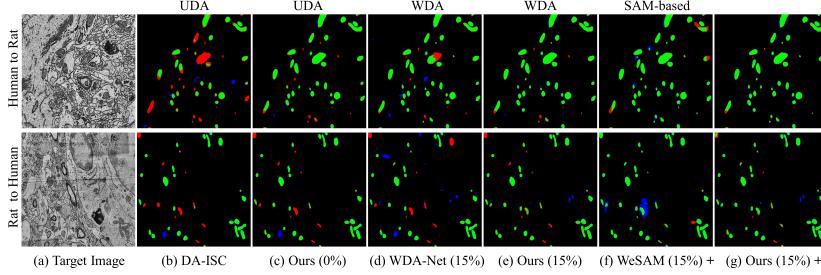


Fig. 2: Qualitative comparison results on two adaptation tasks. Green: true positives; Red: false negatives; Blue: false positives.

data; 2) NoAdapt with testing prompts for interactive segmentation: SAM+, SAM-Med2D+, and Med-SA+; 3) SOTA 2D and 2.5D UDA methods, including DAMT-Net [10], UALR [17], DA-ISC [4], CAFA [19], and WDA-Net (0%) [11], and WeSAM [20] with fine-tuning on the source and target data; 4) SOTA WDA methods, including WDA-Net (15%) [11], and WeSAM(15%)[‡] [20], which use 15% sparse points on the target training set and trained/fine-tuned under the same setting as our Prompt-DAS(15%); 5) Interactive version the WDA methods: WeSAM(15%)[‡], and Prompt-DAS(15%)⁺; 6) The upperbound, which is the model fully supervised trained on the target domain. Additionally, five settings of our model are in comparison, including our source model, our source model for interactive segmentation, our Prompt-DAS (0%) for UDA segmentation, our Prompt-DAS (15%) for WDA segmentation, and our Prompt-DAS (15%)⁺ for interactive WDA segmentation. It is noteworthy that, SAM, SAM-Med2D, Med-SA, and WeSAM are foundation models trained on billion-scale datasets, with or without adaptation using large-scale medical data. In contrast, our Prompt-DAS model is trained from scratch. In the context of interactive segmentation, all center points are used as testing prompts, as required by SAM; however, this setting is impractical for clinical usage. Conversely, during the testing stage, our model can effectively utilize sparse point prompts on partial object instances, enhancing its usability in real-world scenarios.

Table 1 presents quantitative results on two domain adaptation tasks. It is worth noting that the SAM and its medical versions show severely degraded performance both with and without testing prompts when directly applied to EM images. With fine-tuning using the EM data, the WeSAM demonstrates greatly improved performance over other SAM-based models. Our Prompt-DAS with 15% sparse points as training prompts achieves the highest performance on both cross-domain segmentation tasks compared to all UDA, WDA, and SAM-based methods. Notably, our Prompt-DAS (0%), the UDA version of our model, nearly outperforms all UDA, WDA, and SAM-based methods in comparison except the WDA-Net (15%) on MitoEM-Rat → MitoEM-Human. With the inclusion of testing prompts, our Prompt-DAS (15%)⁺ achieves further performance gains, specifically a 1.6% increase in Dice over the Prompt-DAS (15%) for MitoEM-Rat

Table 2: Ablation study on Human → Rat.

	Pseudo-labeling		Training	PCL	Dice (%)	PQ (%)
	Detection	Segmentation	Prompts			
I					88.6	68.7
II	✓				89.2	70.4
III		✓			89.5	70.0
IV	✓	✓			90.4	71.8
V	✓		✓		89.8	71.7
VI	✓	✓	✓		92.7	74.1
Full	✓	✓	✓	✓	93.3	74.5

Table 3: The impact of testing prompt amount on interactive segmentation.

Testing Prompts (Points)	Human → Rat		Rat → Human	
	Dice (%)	PQ (%)	Dice (%)	PQ (%)
0	93.3	74.5	89.2	69.1
15%	93.5	74.2	90.0	69.8
50%	93.5	74.2	90.4	70.9
100%	93.5	74.2	90.8	72.3

→ MitoEM-Human. Compared to WDA-Net (15%), our method demonstrates greater flexibility in alignment with human intention. Moreover, our Prompt-DAS model exhibits performance that closely approaches the supervised upper bound, with only a minimal performance gap. Visual comparison results in Fig. 2 further confirm the advantage of our method.

Ablation study. In Table 2, we evaluate the contributions of the key components of our approach: 1) Detection Pseudo-labeling; 2) Segmentation Pseudo-labeling; 3) Using sparse points as Training Prompts; 4) PCL: prompt-based contrastive learning. As shown in Table 2, the base Model I, our source model, can be improved by adding pseudo-labeling-based detection or segmentation. By conducting multitask learning, Model IV obtains a performance gain of 1.8% in Dice over Model I. With training prompts, we further gain an improvement of 2.3% in Dice over Model IV and obtain Model VI. With additional PCL, our full model further gains an improvement of 0.6% in Dice and 0.4% in PQ.

Influence of testing prompts. Compared to SAM, our model is flexible enough to utilize partial center points as testing prompts. Table 3 presents the influence of testing prompts for interactive segmentation. For Rat → Human, our model can gain improved performance with more point prompts. However, for Human → Rat, adding point prompts does not improve the performance. The main reason is that our model’s performance is already very close to the supervised upper bound, and there is a minimal number of false negatives. Moreover, our model can achieve similar performance with only 15% partial points, taking less than 1/5 of the annotation time of full points.

4 Conclusion

In this study, we develop a promptable transformer model for domain adaptive segmentation of EM images, which can conduct UDA, WDA, and interactive segmentation with various prompting configurations, including sparse points. Our model augments the segmentation task with a detection task, which can significantly alleviate label scarcity and generate pseudo-prompts for the segmentation. Furthermore, the segmentation is guided by a prompt-based contrastive loss, enhancing the discriminability of prompt embeddings. Comprehensive experiments conducted on challenging benchmarks demonstrate the SOTA performance of our approach. The limitation of our model is its requirement for source data and labels for training. In future studies, we will consider a source-free setting.

Acknowledgments. This work was partially supported by the NSFC (No. 1247011276) and Xiamen Natural Science Foundation (No. 3502Z202373042).

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

References

1. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9650–9660 (2021)
2. Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., et al.: Sam-med2d. arXiv preprint arXiv:2308.16184 (2023)
3. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (2021)
4. Huang, W., Liu, X., Cheng, Z., Zhang, Y., Xiong, Z.: Domain adaptive mitochondria segmentation via enforcing inter-section consistency. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 89–98 (2022)
5. Kirillov, A., He, K., Girshick, R., Rother, C., Dollar, P.: Panoptic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9404–9413 (2019)
6. 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)
7. Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A.: A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging* **36**(7), 1550–1560 (2017)
8. Li, F., Zhang, H., Liu, S., Guo, J., Ni, L.M., Zhang, L.: Dn-detr: Accelerate detr training by introducing query denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13619–13627 (2022)

9. Neikirk, K., Lopez, E.G., Marshall, A.G., Alghanem, A., Krystofiak, E., Kula, B., Smith, N., Shao, J., Katti, P., Hinton Jr, A.: Call to action to properly utilize electron microscopy to measure organelles to monitor disease. *European Journal of Cell Biology* **102**(4), 151365 (2023)
10. Peng, J., Yi, J., Yuan, Z.: Unsupervised mitochondria segmentation in em images via domain adaptive multi-task learning. *IEEE Journal of Selected Topics in Signal Processing* **14**(6), 1199–1209 (2020)
11. Qiu, D., Xiong, S., Yi, J., Peng, J.: Weakly-supervised cross-domain segmentation of electron microscopy with sparse point annotation. *IEEE Transactions on Big Data* **11**(2), 359–371 (2025)
12. Qiu, D., Yi, J., Peng, J.: Wda-net: Weakly-supervised domain adaptive segmentation of electron microscopy. In: *IEEE International Conference on Bioinformatics and Biomedicine*. pp. 1132–1137 (2022)
13. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 234–241 (2015)
14. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: *Advances in Neural Information Processing Systems*. vol. 30 (2017)
15. 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 (2020)
16. Wu, J., Ji, W., Liu, Y., Fu, H., Xu, M., Xu, Y., Jin, Y.: Medical sam adapter: Adapting segment anything model for medical image segmentation. *arXiv preprint arXiv:2304.12620* (2023)
17. Wu, S., Chen, C., Xiong, Z., Chen, X., Sun, X.: Uncertainty-aware label rectification for domain adaptive mitochondria segmentation. In: *International Conference on Medical Image Computing and Computer Assisted Intervention*. pp. 191–200 (2021)
18. Xie, B., Tang, H., Cai, D., Yan, Y., Agam, G.: Self-prompt sam: Medical image segmentation via automatic prompt sam adaptation. *arXiv preprint arXiv:2502.00630* (2025)
19. Yin, D., Huang, W., Xiong, Z., Chen, X.: Class-aware feature alignment for domain adaptative mitochondria segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 238–248 (2023)
20. Zhang, H., Su, Y., Xu, X., Jia, K.: Improving the generalization of segmentation foundation model under distribution shift via weakly supervised adaptation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 23385–23395 (2024)