# Joint EM Image Denoising and Segmentation with Instance-aware Interaction

Zhicheng Wang[1,2], Jiacheng Li[1], Yinda Chen[1,2], Jiateng Shou[1], Shiyu Deng[1], Wei Huang[1], and Zhiwei Xiong[1,2(✉)]

[1]MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China
[2]Anhui Province Key Laboratory of Biomedical Imaging and Intelligent Processing, Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
zhichengwang@mail.ustc.edu.cn zwxiong@ustc.edu.cn

**Abstract.** In large scale electron microscopy(EM), the demand for rapid imaging often results in significant amounts of imaging noise, which considerably compromise segmentation accuracy. While conventional approaches typically incorporate denoising as a preliminary stage, there is limited exploration into the potential synergies between denoising and segmentation processes. To bridge this gap, we propose an instance-aware interaction framework to tackle EM image denoising and segmentation simultaneously, aiming at mutual enhancement between the two tasks. Specifically, our framework comprises three components: a denoising network, a segmentation network, and a fusion network facilitating feature-level interaction. Firstly, the denoising network mitigates noise degradation. Subsequently, the segmentation network learns an instance-level affinity prior, encoding vital spatial structural information. Finally, in the fusion network, we propose a novel Instance-aware Embedding Module (IEM) to utilize vital spatial structure information from segmentation features for denoising. IEM enables interaction between the two tasks within a unified framework, which also facilitates implicit feedback from denoising for segmentation with a joint training mechanism. Through extensive experiments across multiple datasets, our framework demonstrates substantial performance improvements over existing solutions. Moreover, our framework exhibits strong generalization capabilities across different network architectures. Code is available at https://github.com/zhichengwang-tri/EM-DenoiSeg.

**Keywords:** Instance Segmentation · Electron Microscopy Images · Image denoising.

## 1 Introduction

Large scale EM imaging plays a pivotal role in neural circuit reconstruction, providing crucial insights for connectomics research [21]. However, the challenge of poor signal-to-noise ratio (SNR) significantly undermines the quality of subsequent segmentation tasks. While increasing dwell time and voltage can enhance SNR,

this approach contradicts the demands of high data throughput and low sample damage[5,17,8,22], as in connectomics studies[13].

Traditionally, image denoising and segmentation techniques have been developed and applied in isolation, with each task being approached separately [20,7,27,1,6]. Efforts to integrate denoising and segmentation can be broadly categorized into denoising-guided segmentation and segmentation-guided denoising methods. Denoising-guided segmentation methods focus on enhancing the robustness of segmentation models by incorporating noise resilience during training [4,28,19]. Conversely, segmentation-guided denoising methods utilize advanced segmentation prior to optimize the network's ability to reduce noise while preserving structural details [15,24]. However, these approaches typically yield unidirectional improvements. A recent work explores the synergy between semantic segmentation and image denoising via alternate boosting [26], yet which cannot handle instance segmentation. Despite progress in denoising [14,5,3] and instance segmentation [10,11,16] in the field of EM, there is a gap in research exploring their symbiotic relationship.

To fill this gap, we propose an instance-aware interaction framework to leverage the synergies of the two closely related tasks. Our framework consists of three components: a denoising network, a segmentation network, and a fusion network. We facilitate collaborative learning and promote task interaction at the feature level. Initially, we use a denoising network to process noisy images, improving segmentation performance by mitigating noise degradation. Subsequently, a segmentation network predicts pixel affinity map encoding crucial spatial structure information. In the fusion network, we introduce a novel Instance-aware Embedding Module (IEM) to fuse semantic and image features in a structure-aware manner, preserving cellular integrity during reconstruction. IEM computes similarity between semantic and image features, facilitating cross-modal interaction between heterogeneous representations. Lastly, our joint learning mechanism enables the fusion network to provide implicit yet effective feedback to the affinity learning process, thereby benefiting segmentation.

We conduct comprehensive experiments across multiple public benchmarks, demonstrating substantial performance improvements over existing solutions. For instance, we achieve an average reduction in VOI by 0.116 for segmentation and an increase in PSNR by 0.320dB for denoising on the AC4 dataset under two noise types. Furthermore, our framework exhibits robust generalization capabilities across various network architectures.

In summary, our contributions are threefold: 1) We present the first unified framework for joint EM image denoising and instance segmentation, leveraging synergies between the two tasks. 2) We introduce a novel instance-aware embedding module that integrates segmentation prior to enhance the performance of both denoising and segmentation through interaction design. 3) Extensive experiments validate the superiority of our framework over existing solutions in both denoising and segmentation performance across multiple datasets, demonstrating robust generalization capabilities.

## 2 Instance-aware Interaction Framework

### 2.1 Overview

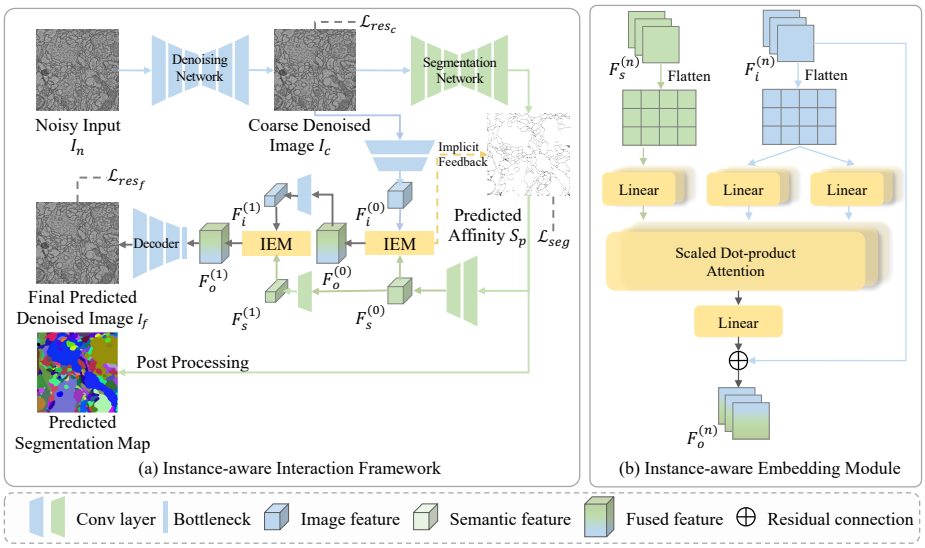


**Fig. 1**: The overview of the proposed instance-aware interaction framework. (a) Given a noisy input image $I_n$, denoising is first performed to mitigate noise degradation to obtain a coarse denoised image $I_c$. Then, a segmentation network predicts affinity map $S_p$ from this less noisy result. After segmentation, the affinity map $S_p$ is utilized as segmentation prior to guide the fusion process with the coarse denoised image to get the final denoised image $I_f$. With post processing, we obtain the final predicted segmentation map. (b)Detail of Instance-aware Embedding Module (IEM).

As illustrated in Fig. 1(a), our instance-aware interaction framework comprises three main components: a denoising network, a segmentation network, and a fusion network. For a noisy input image $I_n$, preliminary denoising is first performed via the denoising network to obtain a coarse denoised image $I_c$. Subsequently, the segmentation network takes the coarse denoised image as input to produce affinity map $S_p$. Finally, the affinity information predicted by the segmentation network is used to guide the reconstruction process in our fusion network to obtain the final denoising output $I_f$. At the same time, the affinity map predicted by the segmentation network can be converted to the final instance segmentation map by the Mutex [25] post-processing algorithm.

In our proposed framework, the denoising network and segmentation network can be various network combinations. We implement a dual Unet architecture in Table 1, 2, and 4. In Table 3, we expand the denoising network to DnCNN [27]

and RIDNet [1], and the segmentation network to TransUNet [7] to show the generalization capacity of our framework. Across all experiments, the comparison solutions for both denoising and segmentation networks are the same with ours.

### 2.2 Instance-aware Embedding Module

Incorporating high-level prior into low-level denoising necessitates detailed consideration of the gaps between the sources. To mitigate such discrepancies and utilize semantics effectively, we propose a Instance-aware Embedding Module (IEM) to enable the feature-level interaction, as shown in Fig. 1(b). IEM establishes connections between the segmentation network and the denoising network, thereby facilitating the integration of these two heterogeneous tasks. In our framework, we choose a Unet-like architecture [20] for the fusion network due to its exceptional performance. The network is further augmented with two IEMs that perform pixel-wise attention between image and semantic features to obtain the fused features. We integrate two IEMs into the second and third layers of the UNet encoder in order to enhance performance while conserving parameters. As illustrated in Fig. 1(a), the coarse denoised image $I_c$ is passed through a cascade of convolution layers to extract feature representations. We utilize the predicted affinity from the segmentation network as multi-scale segmentation prior. To be specific, we take two semantic/image features ($F_s^{(n)}, F_i^{(n)}, n = 0, 1$) with two spatial resolutions ($H/2^{2+n}, W/2^{2+n}$), with H and W denoting the height and width of the input image, which are then fused as refined output feature $F_o^{(n)}$ through IEM. Details of the fusion process are provided below.

To reconcile the discrepancies in channel dimensions and spatial resolutions for the computation of attention, the affinity maps $S_p$ undergo corresponding convolutional transformation. After the transformation, we get the image feature $F_i^{(n)}$ and semantic feature $F_s^{(n)}$ with the same shape. Next, we adopt the $MultiHeadAttention$ [23] mechanism to compute an attention map, which is then used to fabricate image feature $F_i^{(n)}$ to get the refined image feature $F^{(n)}$,

$$F^{(n)} = MultiHeadAttention(F_s^{(n)}, F_i^{(n)}, F_i^{(n)}). \tag{1}$$

Then, we apply $\texttt{ReLU}(\cdot)$ to the refined image feature and add it with the input image feature,

$$F_o^{(n)} = \texttt{ReLU}(F^{(n)}) + F_i^{(n)}, \tag{2}$$

and the final image feature $F_o^{(n)}$ is then sent to the next layer of the fusion network. Finally, a bottleneck layer, followed by a decoder network reconstructs the final denoised image $I_f$.

### 2.3 Joint Training Mechanism

We train our framework in an end-to-end manner. As shown in Fig. 1(a), for the coarse denoised image $I_c$, we employ a restoration loss,

$$\mathcal{L}_{res_c} = \|I_c - I_{gt}\|^2, \tag{3}$$

where $I_{gt}$ is the clean image. For the predicted affinity $S_p$, we utilize a weighted binary cross entropy loss for optimization,

$$\mathcal{L}_{seg} = -\frac{1}{N}\sum w_i \left[S_{gt}\log(S_p) + (1 - S_{gt})\log(1 - S_p)\right], \tag{4}$$

where $S_{gt}$ is the ground truth affinity. The final denoising result $I_f$ is also supervised by a restoration loss,

$$\mathcal{L}_{res_f} = \|I_f - I_{gt}\|^2. \tag{5}$$

Distinct from existing solutions [15,19], we employ a joint training approach, where the overall objection function is the combination of the above three losses, which terms

$$\mathcal{L}_{overall} = \mathcal{L}_{res_c} + \alpha\mathcal{L}_{seg} + \beta\mathcal{L}_{res_f}, \tag{6}$$

where $\alpha = 3$, $\beta = 50$ are the hyper-parameters.

## 3 Experiments

### 3.1 Experimental settings

**Datasets and metrics.** We assess our method on two representative biomedical datasets,including AC3/AC4 [12] and CREMI [9]. The AC3/AC4 dataset is imaged from the mouse somatosensory cortex, consisting of 256 and 100 consecutive EM images of size $1024 \times 1024$ pixels, respectively. In our experiments, we utilize 256 slices from the AC3 dataset for training, and 100 slices from the AC4 dataset for validation. The CREMI dataset, originating from the CREMI neuron segmentation challenge in EM volumes, includes three manually labeled subvolumes, of which CREMI-C contains the most challenging neuron types. Each subvolume contains 125 image slices, originally at a resolution of $1250 \times 1250$, which are then cropped to a standardized size of $1024 \times 1024$ pixels. We adopt the first 75 slices for training, and the subsequent 50 slices for evaluation across three subvolumns. We simulate two types of noise degradation. For film noise, we set the kernel size to 5 and the maximum intensity to 1.5. For Gaussian-Poisson mixture noise, the noise level of Gaussian noise is randomly set between 55 and 85, and the lambda parameter of the Poisson component is set to a random number between 0.6 and 0.8. The noisy images are pre-processed to ensure all experimental noisy images are the same. Across all experiments, we assume noise to be Gaussian-Poisson mixture noise by default. We employ PSNR and SSIM for denoising evaluation, and Variation of Information (VOI [18]) and Adapted Rand Error (ARAND [2]) for segmentation evaluation.

**Comparison solutions.** For each dataset, we train our framework and compare it to three competing baselines for evaluation of segmentation and denoising performance: a model trained purely for segmentation or denoising (referred to as SEG-Only/Dn-Only), and a sequential scheme [19] that first trains a denoiser and then the aforementioned segmentation network (referred to as Sequential), and another scheme [15] that first initializes the network for segmentation in the noiseless setting then trains the cascade of two networks in an end-to-end manner while fixing the weights of segmentation network (referred to as DMS).

**Table 1**: Quantitative comparison on CREMI dataset.

| Method | CREMI-A | | | | CREMI-B | | | | CREMI-C | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Denoising | | Segmentation | | Denoising | | Segmentation | | Denoising | | Segmentation | |
| | PSNR↑ | SSIM↑ | VOI↓ | ARAND↓ | PSNR↑ | SSIM↑ | VOI↓ | ARAND↓ | PSNR↑ | SSIM↑ | VOI↓ | ARAND↓ |
| DN-Only[20] | 28.389 | 0.781 | - | - | 28.137 | 0.759 | - | - | 27.348 | 0.756 | - | - |
| SEG-Only[20] | - | - | 1.208 | 0.251 | - | - | 1.587 | 0.246 | - | - | 1.665 | 0.235 |
| DMS[15] | 26.643 | 0.696 | 1.206 | 0.242 | 25.530 | 0.615 | 1.738 | 0.253 | 24.971 | 0.645 | 1.698 | 0.235 |
| Sequential[19] | 28.389 | 0.781 | 1.197 | 0.247 | 28.137 | 0.759 | 1.573 | 0.251 | 27.348 | 0.756 | 1.621 | 0.224 |
| Ours | **28.437** | **0.789** | **1.158** | **0.239** | **28.202** | **0.763** | **1.497** | **0.239** | **27.407** | **0.763** | **1.566** | **0.209** |

**Table 2**: Quantitative comparison on AC4 dataset. We additionally conduct experiments on film noise degradation.

| Method | Gaussian-Poission mixture | | | | Film | | | |
|---|---|---|---|---|---|---|---|---|
| | Denoising | | Segmentation | | Denoising | | Segmentation | |
| | PSNR↑ | SSIM↑ | VOI↓ | ARAND↓ | PSNR↑ | SSIM↑ | VOI↓ | ARAND↓ |
| DN-Only[20] | 23.568 | 0.631 | - | - | 25.101 | 0.725 | - | - |
| SEG-Only[20] | - | - | 1.723 | 0.333 | - | - | 1.562 | 0.303 |
| DMS[15] | 22.718 | 0.608 | 1.748 | 0.320 | 24.312 | 0.709 | 1.574 | 0.301 |
| Sequential[19] | 23.568 | 0.631 | 1.684 | 0.325 | 25.101 | 0.725 | 1.540 | 0.290 |
| Ours | **23.810** | **0.645** | **1.569** | **0.304** | **25.498** | **0.743** | **1.479** | **0.287** |

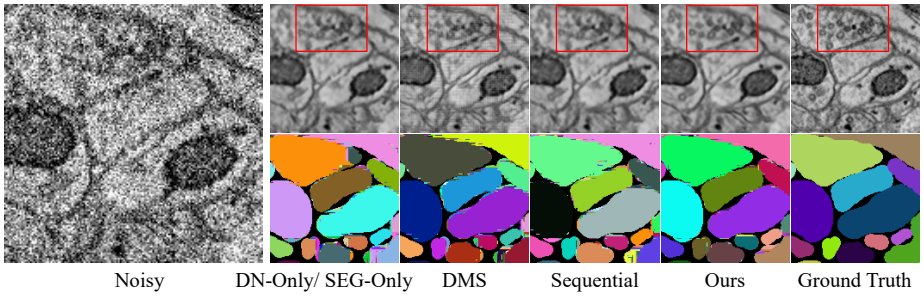


**Fig. 2**: Visual comparisons of different baselines on AC4 dataset. Our method excels in preserving cellular details, as demonstrated by the more complete preservation of vesicles. Moreover, our segmentation result exhibits greater structural precision.

### 3.2 Quantitative and Qualitative Evaluations

The quantitative results on CREMI and AC4 are presented in Table 1 and Table 2 respectively. The experimental results demonstrate that while the Sequential method yields certain segmentation improvement over the SEG-Only method, the gains are marginal and the denoising network does not improve in this approach due to its weights being frozen. The DMS method, while effective in

**Table 3**: Quantitative comparison on the CREMI-C dataset across three different network combinations.

| Method | Unet+TransUNet | | | | DnCNN+TransUNet | | | | RIDNet+TransUNet | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Denoising | | Segmentation | | Denoising | | Segmentation | | Denoising | | Segmentation | |
| | PSNR↑ | SSIM↑ | VOI↓ | ARAND↓ | PSNR↑ | SSIM↑ | VOI↓ | ARAND↓ | PSNR↑ | SSIM↑ | VOI↓ | ARAND↓ |
| DN-Only | 27.097 | 0.744 | - | - | 27.081 | 0.747 | - | - | 27.068 | 0.745 | - | - |
| SEG-Only[20] | - | - | 1.519 | 0.178 | - | - | 1.519 | 0.178 | - | - | 1.519 | 0.178 |
| DMS[15] | 24.358 | 0.566 | 1.617 | 0.205 | 24.205 | 0.548 | 1.675 | 0.222 | 23.708 | 0.492 | 1.649 | 0.214 |
| Sequential[19] | 27.097 | 0.744 | 1.509 | 0.180 | 27.081 | 0.747 | 1.511 | 0.174 | 27.068 | 0.745 | 1.516 | 0.175 |
| Ours | **27.240** | **0.754** | **1.488** | **0.174** | **27.242** | **0.758** | **1.461** | **0.169** | **27.183** | **0.753** | **1.461** | **0.172** |

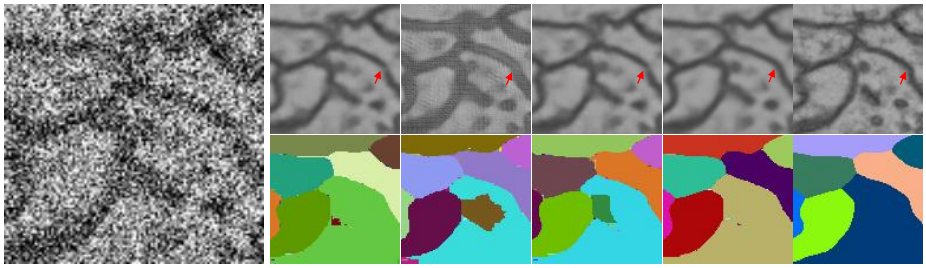


**Fig. 3**: Qualitative comparisons on the CREMI-C dataset with denoising-segmentation network combination being Unet-TransUNet. Our method excels in maintaining the continuity of cell boundaries and shows better segmentation accuracy.

certain natural image contexts, does not align with EM images with high-noise and complex texture, and fails to improve both denoising and segmentation performance. Our method consistently surpasses other methods in both low-level denoising and high-level segmentation with a substantial margin. Additionally, our method achieves significant improvements on CREMI-C and AC4 datasets, which present higher segmentation challenges due to greater size disparities compared to CREMI-A and CREMI-B. This indicates that our method is well-suited for more challenging scenarios. In Fig. 2, we show the visual denoising and segmentation results on AC4 dataset. Although the competing methods can satisfactorily remove the noise, the proposed framework can better preserve the subtle cellular structures such as the vesicles. Furthermore, the segmentation result of the our framework is more structural and precise.

### 3.3 Generalization Evaluation

**Generalization to different noise types.** We extend our experiments to add synthetic film noise on the AC4 dataset, as shown in Table 2, the results corroborate the versatility of our method across different noise conditions.

**Table 4**: The effectiveness of each component in our framework.

| Method | AC4 | | | | CREM-C | | | |
|---|---|---|---|---|---|---|---|---|
| | Denoising | | Segmentation | | Denoising | | Segmentation | |
| | PSNR↑ | SSIM ↑ | VOI↓ | ARAND↓ | PSNR↑ | SSIM↑ | VOI↓ | ARAND↓ |
| w/o DN to SEG | 23.785 | 0.642 | 1.617 | 0.308 | 27.249 | 0.754 | 1.653 | 0.242 |
| w/o SEG to DN | 23.268 | 0.611 | 1.647 | 0.333 | 26.030 | 0.708 | 1.632 | 0.235 |
| Joint DN & SEG (Ours) | **23.810** | **0.645** | **1.569** | **0.304** | **27.407** | **0.763** | **1.566** | **0.209** |

**Generalization to different network architectures.** In Table 3, we investigate a variety of network combinations for denoising and segmentation. To be specific, we utilize three distinct denoising networks: Unet [20], DnCNN [27], and RIDNet [1]. For the segmentation network, TransUNet [7] is chosen due to its exceptional segmentation capabilities. By comparing the results in Table 1 on CREMI-C with the first column of Table 3, where the denoising network remains constant but the segmentation network shifts from Unet to TransUNet, we note an improvement in segmentation performance. Specifically, TransUNet achieves a decrease in VOI from 1.665 to 1.519 (SEG-Only), and from 1.621 to 1.509 (Sequential). Our method further decreases this metric to 1.488, which suggests our method generalizes beyond CNN architectures, indicating its potential for generalization to more advanced architectural designs. Furthermore, as shown in Table 3, our method boosts the performance of three established denoising networks. In Fig. 3, we visualize the denoising and segmentation results for the CREMI-C dataset with denoise-segmentation network combination being Unet-TransUNet. Our method demonstrates superior preservation of cellular integrity and yields more precise segmentation results. Due to the high memory demands of transformer models, Table 3 adopts a block-testing strategy(256 $\times$256), resulting in a marginal decline in denoising performance compared to Table 1.

### 3.4 Ablation study

We conduct ablation studies on the AC4 dataset to prove the effectiveness of our method from various aspects.

**The effectiveness of denoising helping segmentation.** We conduct experiments without denoising network (*i.e.*, w/o DN to SEG). In this paradigm, we directly fuse the noisy image and predicted affinity in the fusion network. As shown in Table 4, without denoising network, both the denoising performance and segmentation performance slightly decrease. This demonstrates the rationality of our coarse-to-fine denoising paradigm.

**The effectiveness of segmenation helping denoising and implicit feedback with joint training.** We also conduct experiments to assess the effectiveness of the fusion network (*i.e.*, w/o SEG to DN). The results in Table 4 clearly

show that both denoising and segmentation quality significantly decrease without fusion network.

## 4 Conclusion

In this work, we present an instance-aware interaction framework for joint denoising and segmentation for EM images. Our method actively exploits the combined advantages inherent in two interdependent tasks, yielding a synergistic improvement that exceeds the outcomes achieved when addressing the tasks in isolation. A novel Instance-aware Embedding Module is proposed to integrate segmentation prior to assist low-level restoration. Further, two subtasks are mutually promoted through interaction. Extensive experiments verify the superiority of our method in both image denoising and segmentation.

## Disclosure of Interests

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

## Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grant 62021001.

## References

1. S. Anwar and N. Barnes. Real image denoising with feature attention. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3155–3164, 2019.
2. I. Arganda-Carreras, S. C. Turaga, D. R. Berger, D. Cireşan, A. Giusti, L. M. Gambardella, J. Schmidhuber, D. Laptev, S. Dwivedi, J. M. Buhmann, et al. Crowdsourcing the creation of image segmentation algorithms for connectomics. *Frontiers in neuroanatomy*, 9:142, 2015.
3. T.-O. Buchholz, M. Jordan, G. Pigino, and F. Jug. Cryo-care: content-aware image restoration for cryo-transmission electron microscopy data. In *International Symposium on Biomedical Imaging*, pages 502–506. IEEE, 2019.
4. T.-O. Buchholz, M. Prakash, D. Schmidt, A. Krull, and F. Jug. Denoiseg: joint denoising and segmentation. In *European Conference on Computer Vision*, pages 324–337. Springer, 2020.
5. S. Chang, L. Shen, L. Li, X. Chen, and H. Han. Denoising of scanning electron microscope images for biological ultrastructure enhancement. *Journal of Bioinformatics and Computational Biology*, 20(03):2250007, 2022.
6. C. Chen, Z. Xiong, X. Tian, Z.-J. Zha, and F. Wu. Real-world image denoising with deep boosting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(12):3071–3087, 2019.

7. J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.
8. S. Deng, W. Huang, C. Chen, X. Fu, and Z. Xiong. A unified deep learning framework for sstem image restoration. *IEEE Transactions on Medical Imaging*, 41(12):3734–3746, 2022.
9. J. Funke, S. Saalfeld, D. Bock, S. Turaga, and E. Perlman. Miccai challenge on circuit reconstruction from electron microscopy images, 2016.
10. J. Funke, F. Tschopp, W. Grisaitis, A. Sheridan, C. Singh, S. Saalfeld, and S. C. Turaga. Large scale image segmentation with structured loss based deep learning for connectome reconstruction. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 41(7):1669–1680, 2018.
11. W. Huang, S. Deng, C. Chen, X. Fu, and Z. Xiong. Learning to model pixel-embedded affinity for homogeneous instance segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 1007–1015, 2022.
12. N. Kasthuri, K. J. Hayworth, D. R. Berger, R. L. Schalek, J. A. Conchello, S. Knowles-Barley, D. Lee, A. Vázquez-Reina, V. Kaynig, T. R. Jones, et al. Saturated reconstruction of a volume of neocortex. *Cell*, 162(3):648–661, 2015.
13. N. Krasowski, T. Beier, G. Knott, U. Köthe, F. A. Hamprecht, and A. Kreshuk. Neuron segmentation with high-level biological priors. *IEEE transactions on medical imaging*, 37(4):829–839, 2017.
14. K. Lee and W.-K. Jeong. Iscl: Interdependent self-cooperative learning for unpaired image denoising. *IEEE Transactions on Medical Imaging*, 40(11):3238–3248, 2021.
15. D. Liu, B. Wen, X. Liu, Z. Wang, and T. S. Huang. When image denoising meets high-level vision tasks: A deep learning approach. In *IJCAI*, 2018.
16. X. Liu, B. Hu, W. Huang, Y. Zhang, and Z. Xiong. Efficient biomedical instance segmentation via knowledge distillation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 14–24. Springer, 2022.
17. T. Mullarkey, J. J. Peters, C. Downing, and L. Jones. Using your beam efficiently: Reducing electron dose in the stem via flyback compensation. *Microscopy and Microanalysis*, 28(4):1428–1436, 2022.
18. J. Nunez-Iglesias, R. Kennedy, T. Parag, J. Shi, and D. B. Chklovskii. Machine learning of hierarchical clustering to segment 2d and 3d images. *PloS one*, 8(8):e71715, 2013.
19. M. Prakash, T.-O. Buchholz, M. Lalit, P. Tomancak, F. Jug, and A. Krull. Leveraging self-supervised denoising for image segmentation. In *IEEE International Symposium on Biomedical Imaging*, pages 428–432, 2020.
20. O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention 18th International Conference*, pages 234–241. Springer, 2015.
21. A. Sheridan, T. M. Nguyen, D. Deb, W.-C. A. Lee, S. Saalfeld, S. C. Turaga, U. Manor, and J. Funke. Local shape descriptors for neuron segmentation. *Nature Methods*, 20(2):295–303, 2023.
22. J. Shou, Z. Xiao, S. Deng, W. Huang, P. Shi, R. Zhang, Z. Xiong, and F. Wu. Learning large-factor em image super-resolution with generative priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11313–11322, 2024.
23. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances In Neural Information Processing Systems*, 30, 2017.
24. S. Wang, B. Wen, J. Wu, D. Tao, and Z. Wang. Segmentation-aware image denoising without knowing true segmentation. *arXiv preprint arXiv:1905.08965*, 2019.

25. S. Wolf, A. Bailoni, C. Pape, N. Rahaman, A. Kreshuk, U. Köthe, and F. A. Hamprecht. The mutex watershed and its objective: Efficient, parameter-free graph partitioning. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 43(10):3724–3738, 2020.
26. S. Xu, K. Sun, D. Liu, Z. Xiong, and Z.-J. Zha. Synergy between semantic segmentation and image denoising via alternate boosting. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(2):1–23, 2023.
27. K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.
28. X. Zhang, S. Li, X. Li, P. Huang, J. Shan, and T. Chen. Destseg: Segmentation guided denoising student-teacher for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3914–3923, 2023.