CS^{2}C: Collaborative Spatial and Spectral Neural Clustering for Organelle Segmentation from Volumetric Electron Microscopy

Jimao Jiang[0000−0003−1463−4445] and Yuru Pei✉[0000−0001−8520−3509]

School of Intelligence Science and Technology, Key Laboratory of Machine Perception (MOE), State Key Laboratory of General Artificial Intelligence, Peking University, Beijing 100871, China
peiyuru@cis.pku.edu.cn

Abstract. Organelle segmentation is crucial for understanding the morphology of biological structures. Existing unsupervised methods leverage powerful feature extractors and clustering techniques to uncover organelle structures from volumetric electron microscopy images. However, these methods often struggle with noisy microscopy images and the computational complexity of numerical clustering. In this paper, we propose CS^{2}C, a novel collaborative spatial and spectral deep neural clustering framework, for multi-class organelle segmentation. The pillar of our approach is combining unsupervised deep spectral clustering and spatial clustering, which enhances a harmony of learned cluster assignments under the spatial and spectral superpixel-wise representation. Specifically, we adopt a masked autoencoder-based feature extractor to obtain powerful superpixel features, where spatial clustering is performed directly on these features. Beyond that, spectral clustering is applied in the spectral domain, naturally alleviating high-frequency perturbations in the image features. The entire framework is trained end-to-end using a combination of clustering loss and consistency regularization between spatial and spectral clustering. Extensive experiments demonstrate that our method outperforms state-of-the-art unsupervised methods on known benchmarks. Code is available at: https://github.com/JimaoJIANG/CS2C.

Keywords: Collaborative spatial and spectral clustering · Organelle segmentation · Volumetric electron microscopy.

1 Introduction

Nanometer-level organelle segmentation plays a crucial role in understanding the intricate morphology and organization of cellular structures [11, 16, 20, 22]. High-throughput imaging techniques, such as volumetric electron microscopy (VEM), provide high-resolution imaging of organelle structures with great efficiency. While manual segmentation remains the gold standard for a variety of downstream biological tasks, it is labor-intensive and relies heavily on expert knowledge. Transformer-based self-supervised learning techniques [8, 4] have

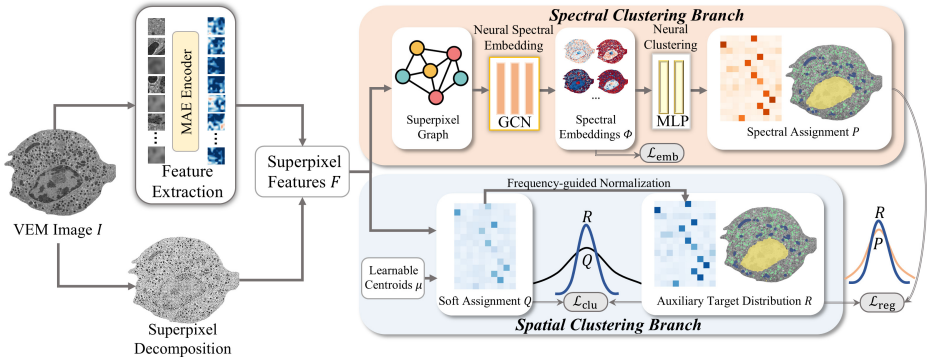

Fig. 1. Illustration of the proposed collaborative spatial and spectral neural clustering framework. When given a VEM slice image I, we conduct superpixel decomposition and adopts the a masked autoencoder (MAE) to extract superpixel features F. In the spatial clustering branch, we compute the soft cluster assignment probability Q and the axillary target distribution R. The cluster centroids are optimized by minimizing KL-divergence-based clustering loss $\mathcal{L}_{clu}$. In the spectral clustering branch, we feed the superpixel graph to the neural spectral embedding module for approximated spectral bases Φ, which are further used to infer the spectral clustering assignment matrix P via an MLP-based clustering module. We impose a consistency regularization $\mathcal{L}_{reg}$ of the spatial and spectral clustering. Our approach enables and end-to-end organelle segmentation with consistent cluster assignment across images.

shown great potential in capturing both local and global spatial correlations, making them well-suited for robust representations of subcellular structures. The deep neural network-based representation learning facilitates scalable and automated structural segmentation from microscopy images [7, 20, 10].

Clustering has long been a powerful unsupervised segmentation approach for data grouping based on feature similarity, and can be performed in either spatial or spectral domains. Spatial clustering methods typically operate directly on extracted image features by identifying local pixel similarities. However, these methods are sensitive to high-frequency noise in microscopy images, often leading to suboptimal segmentation results. On the other hand, spectral clustering operates in the spectral embedding space and is more robust to high-frequency perturbations. However, numerical spectral embedding and clustering can be computationally expensive when dealing with large-scale graphs [6]. Furthermore, additional synchronization operations [14] are required to maintain consistent cluster assignments across images.

In this paper, we exploit collaborative spatial and spectral clustering learning and propose a novel unsupervised two-branch deep neural clustering framework, CS^{2}C, for organelle segmentation, as shown in Fig. 1. Drawing inspiration from advances in self-supervised representation learning and deep embedding clustering [19], our framework leverages spatial clustering on rich semantic features extracted from the MAE model. In light of high frequency perturbations in mi-

croscopy image features, we exploit the deep spectral embedding and conduct clustering in the low-dimensional spectral embedding space, which is represented by the approximated spectral basis of the graph Laplacian matrix regarding the superpixel graph. We train the two-branch clustering model end-to-end by combining the clustering loss and the consistency regularization between spatial and spectral clustering assignments. Our approach promotes a unified solution to organelle segmentation in both spatial and spectral domains. Extensive experimental results demonstrate that our method outperforms state-of-the-art techniques on existing benchmarks. Our main contributions are as follows:

- We introduce an unsupervised collaborative neural spatial and spectral clustering framework for efficient organelle segmentation from VEM images.
- We present a simple yet effective two-branch design for unsupervised deep clustering, which mitigates high-frequency perturbations in microscopy image features and promotes consistent clustering in both spatial and spectral domains.
- We validate the effectiveness of our approach through comprehensive experiments, demonstrating superior performance in multi-class organelle segmentation compared to state-of-the-art methods.

2 Method

Given the input VEM images, the objective is to train a clustering model that takes a VEM slice I as input and outputs clustering assignments P for various organelle structures. As shown in Fig. 1, our framework follows the prevailing paradigm of feature extraction and clustering inference based on feature similarity. The core of our framework is harnessing a two-branch spatial-spectral clustering model for organelle segmentation. The spatial clustering branch performs clustering directly on the rich semantic superpixel features extracted by the MAE model, while the spectral clustering branch conducts neural spectral embedding and clustering in a low-dimensional space spanned by the approximated spectral basis of the graph Laplacian matrix. We introduce an end-to-end training strategy that combines a clustering loss with a consistency regularization between spatial and spectral clustering assignments. Our approach mitigates high-frequency perturbations in microscopy image features and encourages spatial-spectral consistency in clustering.

Superpixel Features. Given the fine granularity of organelle structures, the patches used in the vision transformer model must be small enough to capture detailed shapes and appearances. However, small patches increase computational complexity in feature extraction [8]. To address this, we convert the high-resolution microscopy image into a more compact representation via superpixel decomposition [1] instead of using regular patches. We employ MAE-based feature extraction [8], which has demonstrated effectiveness in handling fine-grained, repetitive subcellular structures [10]. We up-sample the feature embedding from the pre-trained MAE encoder to obtain l-channel features $F \in \mathbb{R}^{m \times n \times l}$

for a VEM image with a resolution of $m \times n$. The superpixel feature is then defined as the concatenation of mean and standard deviation of the pixels within each superpixel.

2.1 Collaborative Spatial-Spectral Neural Clustering

The core of our framework consists of two parallel clustering models: a spatial model and a spectral model. They predict clustering assignments from superpixel features, leveraging the spectral clustering's resilience to high-frequency perturbations and the alignment with centroid-based probability distributions in the original feature space.

Spectral Clustering. In the spectral branch, organelle segmentation is formulated as a cut on the superpixel graph within the spectral embedding space. Traditional independent per-graph spectral clustering methods often overlook cross-image relationships, leading to inconsistent cluster assignments. To address this, we propose a deep neural spectral clustering model that leverages its generalization capabilities to ensure consistent labeling across images. We generate a superpixel graph using MAE features, where the affinity matrix $A \in \mathbb{R}^{n_s \times n_s}$ is defined with the RBF kernel, and $a_{ij} = \exp \frac{-\|F_i - F_j\|_2^2}{\kappa}$ for $1 \leq i, j \leq n_s$. κ is the kernel's variance, and n_s is the number of superpixels. We further reduce the bandwidth and eliminate weakly correlated values as [2], and $A \leftarrow \max\left(A - \frac{\max A}{\alpha}, 0\right)$. where α controls the strength of the repulsion forces that diminish weak connections.

We utilize a Graph Convolutional Network (GCN)-based spectral embedding module to approximate the spectral basis $\Phi \in \mathbb{R}^{n_s \times u}$ of the graph Laplacian matrix L, avoiding eigenvector switching and sign flipping problems in the computationally expensive numerical eigendecomposition. u denotes the number of approximated spectral bases. The spectral embedding is optimized using two constraints: (1) the diagonalization of the symmetric Laplacian matrix L, and (2) the orthogonality of the approximated spectral basis Φ, as [18]. The spectral embedding loss:

$$\mathcal{L}_{emb} = \sum_{i \neq j} \left[(\phi_i^T L \phi_j)^2 + \phi_i^T \phi_j \right], \tag{1}$$

where ϕ_i and ϕ_j are the i-th and j-th column vectors of Φ. The eigendecomposition of the symmetric Laplacian L satisfies $\Lambda = \Phi^T L \Phi$, where Λ is a diagonal matrix containing the eigenvalues. The loss ensures that off-diagonal values of $\Lambda = \Phi^T L \Phi$ approach zero and enforces the orthogonality of Φ.

A neural clustering module f_θ is introduced to assign cluster labels based on the spectral embedding, eliminating the need for label synchronization like Hungarian matching [14, 2]. The cluster assignment probability matrix $P = \text{softmax}(f_\theta(\Phi))$ defines a cut on the superpixel graph using the perturbation-resistant spectral embedding.

Spatial Clustering. In the spatial branch, clustering is directly applied to the MAE features. Inspired by deep embedding clustering methods [19], we optimize learnable cluster centroids by minimizing the distance between a soft cluster assignment and an auxiliary target distribution. The soft cluster assignment matrix Q is computed using a Student's t-distribution kernel [12], and $q_{i,j} = \frac{(1+\|F_{s,i}-\mu_j\|/\tau)^{-(\tau+1)/2}}{\sum_{j'}(1+\|F_{s,i}-\mu_{j'}\|/\tau)^{-(\tau+1)/2}}$. $\mu \in \mathbb{R}^{k\times 2l}$ represents the learnable centroids, initialized using k-means, and τ controls the degrees of freedom. The auxiliary target distribution R is derived by frequency-guided normalization of Q, and

$$r_{i,j} = \frac{q_{i,j}^2 / \sum_k q_{k,j}}{\sum_{j'} \left(q_{i,j'}^2 / \sum_k q_{k,j'} \right)}. \tag{2}$$

We optimize the cluster centroids by minimizing the Kullback-Leibler (KL) divergence between the soft assignment Q and the target distribution R. The clustering loss is defined as:

$$\mathcal{L}_{clu} = \mathrm{KL}(R\|Q) = \sum_i \sum_j r_{i,j} \log \frac{r_{i,j}}{q_{i,j}}. \tag{3}$$

Minimizing the loss $\mathcal{L}_{clu}$ refines the centroids μ, improving cluster purity and focusing on high-confidence assignments.

Spatial-Spectral Consistency Regularization. Spectral clustering is known to be robust against high-frequency perturbations in image features, owing to the low-pass nature of the graph Laplacian's eigenbasis approximation. Spatial clustering, when applied to MAE features, is effective for iterative refinement of soft cluster assignments. To exploit the strengths of both clustering methods, we propose minimizing the distance between their respective probabilistic assignment matrices. Specifically, we use the KL divergence to measure the dissimilarity between the spectral clustering matrix P and the spatial clustering matrix R. The spatial and spectral consistency regularization loss $\mathcal{L}_{reg}$ is defined as:

$$\mathcal{L}_{reg} = \mathrm{KL}(P\|R) = \sum_i \sum_j p_{i,j} \log \frac{p_{i,j}}{r_{i,j}}. \tag{4}$$

Unlike cross-entropy loss, KL divergence updates the model more gently [3], helping to prevent abrupt shifts in data representations.

2.2 Training Loss

The overall training loss is a weighted combination of the neural spectral embedding loss $\mathcal{L}_{emb}$, the clustering loss $\mathcal{L}_{clu}$, and the spatial-spectral consistency regularization loss $\mathcal{L}_{reg}$.

$$\mathcal{L} = \mathcal{L}_{emb} + \gamma_1 \mathcal{L}_{clu} + \gamma_2 \mathcal{L}_{reg}, \tag{5}$$

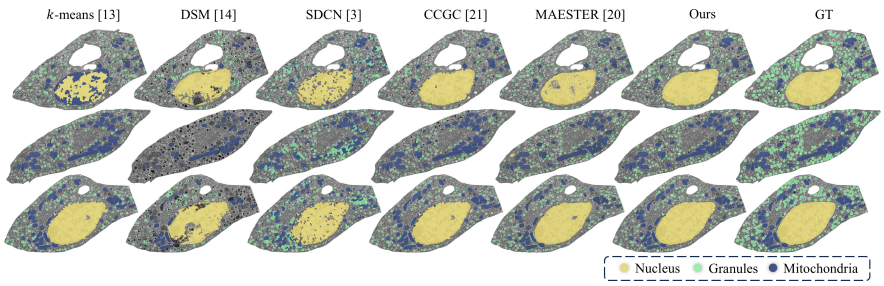

Fig. 2. Qualitative comparison of organelle segmentation by compared methods.

where γ_1 and γ_2 are hyperparameters that balance the regularized spatial and spectral clustering. Minimizing the combined loss function $\mathcal{L}$ optimizes both the neural spectral and spatial clustering modules simultaneously.

In the online inference process, given a VEM image I, first, the MAE-based feature extractor generates superpixel features, which are then used to construct the superpixel graph. Second, the spectral clustering module takes the weighted superpixel graph as an input and assigns each superpixel to a cluster, producing the clustering assignment matrix P and achieving the organelle segmentation.

3 Experiments

Datasets and Metrics. We evaluate the proposed CS^{2}C on the BetaSeg [9, 15], which consists of four preprocessed primary mouse pancreatic islet β cell volumes. The first three volumes are used for training, and the last one for testing as [20]. We divide the VEM slices into superpixels with a physical size of around 5 micrometers squared area using the SLIC [1] algorithm, with the compactness set to 0.2. During training, we use slices with a resolution of 560×560, containing approx. 3000 superpixels. We evaluate the performance using Dice Similarity Coefficient (DSC) and Intersection over Union (IoU), which measure the consistency of the predicted segmentation with the ground truth.

Implementation Details. We use the MAE based feature extractor [20], with the feature channel number l of 192. κ in RBF kernel, α in affinity matrix, and τ in Student's t-distribution kernel are set to 2, 4 and 1, respectively. We retain $u = 12$ spectral bases and set the cluster number k to 14 according to the number of organelle types empirically. The hyperparameters γ_1 and γ_2 are set to 1 in the loss function $\mathcal{L}$. The spectral embedding module includes three linear GCN layers with 384×96, 96×48, and 48×12 weight matrices. The MLP-based clustering module has two fully connected layers with dimensions of 14×14 and 14×8. We implement the proposed CS^{2}C using the PyTorch toolkit on a PC with an NVIDIA RTX 2080Ti GPU. We use the Adam optimizer with a momentum of 0.9 and 0.999. The learning rate is set to 0.01 for the first 3,000 iterations and then reduced to 0.0001 for 2,000 iterations. The mini-batch size is set to 1. The training takes approx. 6 hours. In the online testing process,

Table 1. Segmentation results regarding the DSC and IoU by compared methods and variants without spatial (`spa`) or spectral (`spe`) branches. We report the output clustering, CS^2C_{spe} and CS^2C_{spa}, from the spectral and spatial branches. (sp: superpixel)

	Feature Genre	Use Graph	Nucleus		Granules		Mitochondria		*Average*	
			DSC↑	IoU↑	DSC↑	IoU↑	DSC↑	IoU↑	DSC↑	IoU↑
k-means [13]	sp	No	0.927	0.864	0.442	0.284	0.754	0.605	0.708	0.584
Classical SC [17]	sp	Yes	0.903	0.823	0.365	0.224	0.769	0.625	0.679	0.557
DSM [14]	patch	Yes	0.848	0.735	0.347	0.210	0.743	0.591	0.646	0.512
DeepCUT [2]	patch	Yes	0.861	0.756	0.333	0.200	0.694	0.531	0.629	0.495
FastDGC [5]	sp	Yes	0.651	0.482	0.484	0.319	0.658	0.490	0.597	0.431
SDCN [3]	sp	Yes	0.902	0.821	0.442	0.284	0.665	0.498	0.669	0.534
CCGC [21]	sp	Yes	0.957	0.917	0.452	0.292	0.835	0.717	0.748	0.642
MAESTER [20]	pixel	No	0.943	0.892	0.556	0.385	0.778	0.636	0.759	0.638
w/o `spe`	sp	No	0.952	0.909	0.575	0.404	0.824	0.701	0.784	0.671
w/o `spa`	sp	Yes	0.954	0.913	0.573	0.401	0.861	0.756	0.796	0.690
CS^2C_{spa}	sp	No	0.958	0.919	0.601	0.430	0.838	0.722	0.799	0.690
CS^2C_{spe} (Ours)	sp	Yes	**0.966**	**0.935**	**0.603**	**0.431**	**0.872**	**0.772**	**0.814**	**0.713**

the feature extraction, spatial clustering, and spectral clustering of a 1082×545 image take 2.714 seconds, 0.005 seconds, and 0.002 seconds.

Experimental Results. We compare our approach to state-of-the-art deep clustering methods, including DSM [14], DeepCUT [2], FastDGC [5], SDCN [3], CCGC [21], and MAESTER [20], as well as methods that apply classical k-means [13] and spectral clustering [17] on MAE-based superpixel features. To ensure a fair comparison, we use the same superpixel features and graphs for compared deep clustering methods [5, 3, 21] as ours. Table 1 and Fig.2 present quantitative and qualitative results for organelle segmentation. Our method consistently outperforms all baselines. The compared methods rely on deep neural network-based feature extraction, such as DINOViT[14, 2] and MAE [20]. Clustering is performed either on individual pixels [20], patch graphs [14, 2], or superpixel graphs as in our approach. Pixel-level clustering [20] generally outperforms patch-based clustering [14, 2], though it incurs a higher computational cost during both training and testing. Notably, compared to MAESTER [20], which uses k-means on pixel-level MAE features, our method uses a compact superpixel graph that reduces computational time by more than a factor of 137, while achieving a DSC boost of 0.023 (nucleus), 0.047 (granules), and 0.094 (mitochondria). Fig.3(b) shows the effect of Gaussian noise levels from 1% to 8% on organelle segmentation. Across all noise levels, the performance drop of MAESTER[20] is approximately twice as large as that of our method. Superior performances of our approach can be attributed to spectral embedding and the collaborative learning strategy, mitigating negative impacts of noisy features.

Our method excels by harmonizing spatial and spectral clustering, offering performance gains over approaches using k-means or neural clustering networks

Table 2. Organelle segmentation accuracy when using different cluster number k.

k	Nucleus		Granules		Mitochondria		*Average*	
	DSC↑	IoU↑	DSC↑	IoU↑	DSC↑	IoU↑	DSC↑	IoU↑
8	0.940	0.887	0.243	0.139	0.645	0.476	0.609	0.500
10	0.948	0.900	0.181	0.099	0.666	0.495	0.597	0.498
12	**0.968**	**0.938**	0.568	0.397	0.869	0.768	0.802	0.701
14	0.966	0.935	**0.603**	**0.431**	**0.872**	**0.772**	**0.814**	**0.713**
16	0.964	0.931	0.584	0.412	0.870	0.770	0.806	0.704
18	0.965	0.933	0.513	0.345	0.871	0.771	0.783	0.683

directly on superpixel features [5, 3, 21], as well as deep spectral clustering methods [14, 17]. This makes our CS^{2}C SOTA among unsupervised baselines, and demonstrates the power by collaborative neural spatial and spectral clustering with generalization capacity and consistent clustering across images.

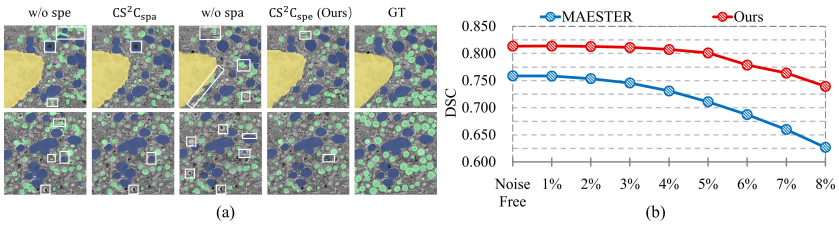

Fig. 3. (a) Organelle segmentation by variants of the proposed method without the spectral branch (`spe`) or the spatial branch (`spa`). We illustrate the output clustering, CS^{2}C$_{\texttt{spe}}$ and CS^{2}C$_{\texttt{spa}}$, from the spatial and spectral branches. Errors are white blocked. (b) Organelle segmentation accuracy when confronted with noisy VEM images.

Ablation Study. We conduct an ablation study to assess the contributions of the spatial and spectral clustering branches as shown in Table 1 and Fig.3(a). We compare our full framework with two variant models: (1) w/o `spe`, which removes the spectral branch and uses only $\mathcal{L}_{clu}$ for training. This setup combines DEC[19] with frozen MAE features. (2) w/o `spa`, which removes the spatial branch and relies on $\mathcal{L}_{emb}$ and k-means supervision for training. The results demonstrate that both the spatial and spectral branches contribute to final clustering. Specifically, the spatial and spectral branches yield IoU gains of 0.016 and 0.012, respectively. The spatial branch also improves structural segmentation along nucleus boundaries, helping to prevent label omissions (Fig. 3(a)). Overall, the proposed CS^{2}C, which uses both spatial and spectral branches in parallel, achieves consistent organelle segmentation aligned with the ground truth.

Parameter Analysis. Table 2 demonstrates the impact of cluster number k on subcellular segmentation performance. The results show that performance

is insensitive with moderate cluster numbers. However, both excessively small and large cluster numbers have limitations. With too few clusters, granules and mitochondria become difficult to distinguish, while too many clusters hinder the capture of fine-grained granules, resulting in reduced performance.

4 Conclusion

We presented a collaborative spatial-spectral neural clustering approach for organelle segmentation, which integrates neural clustering in both the spatial and spectral domains. Our method introduces a novel two-branch clustering framework that leverages the robustness of spectral clustering to high-frequency perturbations in image features, while also utilizing spatial clustering for iterative refinement of cluster assignments. We demonstrate the effectiveness of our approach with state-of-the-art accuracy in organelle segmentation.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China under Grant 62272011 and Beijing Natural Science Foundation 7232337.

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

References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**, 2274–2282 (2012)
2. Aflalo, A., Bagon, S., Kashti, T., Eldar, Y.C.: Deepcut: Unsupervised segmentation using graph neural networks clustering. 2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW) pp. 32–41 (2022), https://api.semanticscholar.org/CorpusID:254564022
3. Bo, D., Wang, X., Shi, C., Zhu, M., Lu, E., Cui, P.: Structural deep clustering network. In: Proceedings of the web conference 2020. pp. 1400–1410 (2020)
4. Caron, M., Touvron, H., Misra, I., J'egou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. 2021 IEEE/CVF International Conference on Computer Vision (ICCV) pp. 9630–9640 (2021), https://api.semanticscholar.org/CorpusID:233444273
5. Ding, S., Wu, B., Ding, L., Xu, X., Guo, L., Liao, H., Wu, X.: Towards faster deep graph clustering via efficient graph auto-encoder. ACM Transactions on Knowledge Discovery from Data **18**(8), 1–23 (2024)
6. Efroni, O., Ginzburg, D., Raviv, D.: Spectral teacher for a spatial student: Spectrum-aware real-time dense shape correspondence. In: 2022 International Conference on 3D Vision (3DV). pp. 1–10. IEEE (2022)
7. Han, H., Dmitrieva, M., Sauer, A., Tam, K.H., Rittscher, J.: Self-supervised voxel-level representation rediscovers subcellular structures in volume electron microscopy. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) pp. 1873–1882 (2022), https://api.semanticscholar.org/CorpusID:250980726

8. He, K., Chen, X., Xie, S., Li, Y., Doll'ar, P., Girshick, R.B.: Masked autoencoders are scalable vision learners. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 15979–15988 (2021), https://api.semanticscholar.org/CorpusID:243985980
9. Heinrich, L., Bennett, D., Ackerman, D., Park, W., Bogovic, J.A., Eckstein, N., Petruncio, A., Clements, J., Pang, S., Xu, S., Funke, J., Korff, W.L., Hess, H.F., Lippincott-Schwartz, J., Saalfeld, S., Weigel, A.V., Team, P., Ali, R., Arruda, R., Bahtra, R., Nguyen, D.: Whole-cell organelle segmentation in volume electron microscopy. Nature **599**, 141 – 146 (2021), https://api.semanticscholar.org/CorpusID:238421373
10. Kraus, O., Kenyon-Dean, K., Saberian, S., Fallah, M., McLean, P., Leung, J., Sharma, V., Khan, A., Balakrishnan, J., Celik, S., et al.: Masked autoencoders for microscopy are scalable learners of cellular biology. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11757–11768 (2024)
11. Lu, Z., Zuo, S., Shi, M., Fan, J., Xie, J., Xiao, G., Yu, L., Wu, J., Dai, Q.: Long-term intravital subcellular imaging with confocal scanning light-field microscopy. Nature Biotechnology pp. 1–12 (2024)
12. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)
13. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability/University of California Press (1967)
14. Melas-Kyriazi, L., Rupprecht, C., Laina, I., Vedaldi, A.: Deep spectral methods: A surprisingly strong baseline for unsupervised semantic segmentation and localization. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 8354–8365 (2022), https://api.semanticscholar.org/CorpusID:248811034
15. Müller, A., Schmidt, D., Xu, C.S., Pang, S., DCosta, J.V., Kretschmar, S., Münster, C., Kurth, T., Jug, F., Weigert, M., et al.: 3d fib-sem reconstruction of microtubule–organelle interaction in whole primary mouse β cells. Journal of Cell Biology **220**(2) (2021)
16. Seal, S., Trapotsi, M.A., Spjuth, O., Singh, S., Carreras-Puigvert, J., Greene, N., Bender, A., Carpenter, A.E.: Cell painting: a decade of discovery and innovation in cellular imaging. Nature methods pp. 1–15 (2024)
17. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Transactions on pattern analysis and machine intelligence **22**(8), 888–905 (2000)
18. Sun, D., Pei, Y., Zhang, Y., Xu, T., Wang, T., yan Zha, H.: Dense correspondence of deformable volumetric images via deep spectral embedding and descriptor learning. Medical image analysis **82**, 102604 (2022), https://api.semanticscholar.org/CorpusID:251947036
19. Xie, J., Girshick, R., Farhadi, A.: Unsupervised deep embedding for clustering analysis. In: International conference on machine learning. pp. 478–487. PMLR (2016)
20. Xie, R., Pang, K., Bader, G.D., Wang, B.: Maester: Masked autoencoder guided segmentation at pixel resolution for accurate, self-supervised subcellular structure recognition. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 3292–3301 (2023), https://api.semanticscholar.org/CorpusID:261080761

21. Yang, X., Liu, Y., Zhou, S., Wang, S., Tu, W., Zheng, Q., Liu, X., Fang, L., Zhu, E.: Cluster-guided contrastive graph clustering network. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 10834–10842 (2023)
22. Zhou, D., Gu, C., Xu, J., Liu, F., Wang, Q., Chen, G., Heng, P.A.: Repmode: learning to re-parameterize diverse experts for subcellular structure prediction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3312–3322 (2023)