COFI - Coarse-Semantic to Fine-Instance Unsupervised Mitochondria Segmentation in EM

Anusha Aswath[1](✉), Ahmad Alsahaf[2], B. Daan Westenbrink[3], Ben N. G. Giepmans[2], and George Azzopardi[1]

[1] Bernoulli Institute, University Groningen, Groningen, The Netherlands
a.aswath@rug.nl

[2] Department of Biomedical Sciences of Cells and Systems, University Medical Center Groningen, Groningen, The Netherlands

[3] Department of Cardiology, University Medical Center Groningen, Groningen, The Netherlands

Abstract. Instance segmentation is crucial for insightful analysis in the increasing use of large-scale electron microscopy (EM) to gain a better understanding of disease causes or progression. Instance segmentation is a more granular version of semantic segmentation, as it identifies and distinguishes individual object instances, whereas semantic segmentation only identifies object classes. In this study, we introduce a two-stage unsupervised approach called COFI, which stands for Coarse-Semantic to Fine-Instance segmentation, for the application of mitochondria segmentation in large-scale 2D EM images. In its first stage, it produces a rough region mask by clustering image patches and prompting a user to select the regions of interest. This is followed by a boundary delineation method based on the brain-inspired COSFIRE filter which is augmented by an inhibition component that makes it robust to image texture and noise. The effectiveness of the proposed COFI approach is evaluated on an EM dataset of the heart muscle of a mouse tissue, which consisted of four tiles of 16384 × 16384 pixels, containing a total of 2287 instances of mitochondria among other subcellular structures. It consistently achieved panoptic quality measures that are substantially superior to competing supervised methodologies. Besides its elevated effectiveness, the proposed COFI approach is conceptually simple and sufficiently versatile as the structure of interest is not intrinsic to the method.

Keywords: Instance segmentation · unsupervised · mitochondria

1 Introduction

Segmentation is an important step in the analysis of electron microscopy (EM) images in biology. Through segmentation, sub-cellular structures can be identified and labeled, which improves the biological understanding of the analyzed

N. Tsapatsoulis et al. (Eds.): CAIP 2023, LNCS 14185, pp. 87–97, 2023.
https://doi.org/10.1007/978-3-031-44240-7_9

samples. EM is increasingly being employed in large-scale biological initiatives, whether for volume imaging (3D EM) or large-area mapping (2D EM). In both methods, the goal to resolve nanoscale features (2–10 nm/pixel) is linked with the desire to set these findings in a larger context, which could be a large area or a 3D volume. High-throughput large-scale EM imaging is now possible due to enhanced automation that generates petabytes of image data [1,2]. Hence, there is a need for developing automatic tools for EM segmentation.

Large-scale 2D EM or nanotomy[1] provides an unbiased analysis of structures in EM images with the right cellular context [1]. We propose a new methodology for instance segmentation of mitochondria in 2D EM. Instance segmentation involves assigning each pixel to the correct class – mitochondria in this case – and identifying each component of that class as a separate instance. Figure 1 shows an example of a cropped region from an EM image with multiple mitochondria and corresponding ground truth maps of semantic, contour, and instance segmentation. Mitochondria are the primary energy providers for cell activities, thus essential for metabolism. Results of instance segmentation can be used to quantify morphological properties of mitochondria, which is not only crucial to basic research, but also informative to the clinical studies of several diseases.

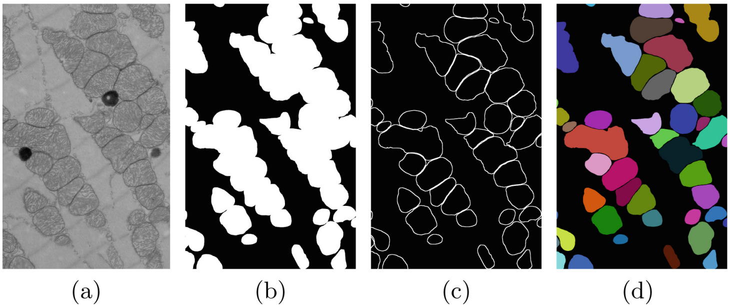

Fig. 1. Example of expected segmentation. (a) A region with apposing mitochondria, and the ground truth (b) semantic, (c) contour and (d) instance segmentation maps.

We propose a two-stage unsupervised pipeline. The first stage entails unsupervised semantic segmentation through clustering of overlapping patches using their feature embeddings encoded by a pre-trained network and prompting a user to select regions of interest among the resulting clusters. The second stage involves the COSFIRE filter approach with surround inhibition for edge delineation. It is inspired by simple cells of the mammalian visual cortex, and is robust to delineating edges and lines in the presence of texture [3].

[1] www.nanotomy.org.

2 Related Work

Previous methods for mitochondria segmentation have primarily used handcrafted features [4] or those derived using supervised learning to encode images [5,6]. The success of encoder-decoder architectures such as FCN, U-Net, and DeepLabv3+ for semantic segmentation, has enabled pixel-wise classification of EM images. Relevant image regions can also be obtained using prior knowledge of an object's shape or texture through fragment matching. Due to its adaptability to noise and local variations, such methods are, however, more effective for image denoising and texture synthesis than pixel-based techniques. The work in [7] investigated a patch processing approach based on region homogeneity, utilizing CNNs as feature extractors and performing boundary refinement using watersheds. Boundary-based segmentation is a preferred technique for instance segmentation due to its ability to provide fine-grained results in combination with other techniques such as object proposals or region-based segmentation to improve performance. Instance segmentation of mitochondria was preferred with semantic region mapping and boundary prediction, in comparison with top-down approaches, as variability in their appearances, shape, and the presence of overlapping instances makes the use of object proposal networks impractical [8].

Manually marking ground truth in EM images is tedious, which makes supervised methods challenging. This may be addressed by transfer learning, which takes a supervised model that was pre-trained on a large dataset and fine-tunes it on a different dataset. Self-supervised learning has emerged as a label-free alternative to pre-training, utilizing a contrastive loss function to learn meaningful representations. It can achieve high accuracy in various downstream tasks through fine-tuning with a simple linear classification or an MLP head [9]. Pre-trained models for unlabeled EM data have become possible with the release of CEM1.5M, a large and diverse dataset that provides ample cellular context [10].

The brain-inspired COSFIRE filter approach that we use here has proven to be effective for unsupervised delineation of curvilinear structures in complex and noisy backgrounds. It achieves orientation selectivity by aggregating the collective responses of a set of difference-of-Gaussian functions that are linearly aligned in their areas of support [11]. This approach has demonstrated success in various applications, such as delineating blood vessels in retinal fundus images, roads and rivers from aerial images [12,13]. The COSFIRE model has been extended with push-pull inhibition [14] and surround suppression [3]. The push-pull inhibition is effective in suppressing high-frequency noise, while surround suppression inhibits responses in the neighbourhoods of dominant contours.

3 Method

The proposed COFI method comprises two components. First, it uses a pre-trained network to generate a rough object location map by clustering embeddings and selecting regions of interest. Then, the instance-level fine delineation is performed by the inhibition-augmented COSFIRE filter approach.

3.1 Dataset Description and Annotation

The proposed pipeline is evaluated on a nanotomy dataset of the heart muscle of a mouse tissue, which consisted of the four tiles shown in Fig. 2. Manual annotation of individual instances was a laborious task due to various factors such as high-resolution noise, image artifacts, surrounding structures with similar textures, and side-by-side mitochondria. Manual delineating all 2287 instances of mitochondria, took approximately 8 hours per tile, totaling four working days. The instance segmentation ground truth masks were obtained using the polygon tool of ImageJ [15] and were further proofread by biomedical experts.

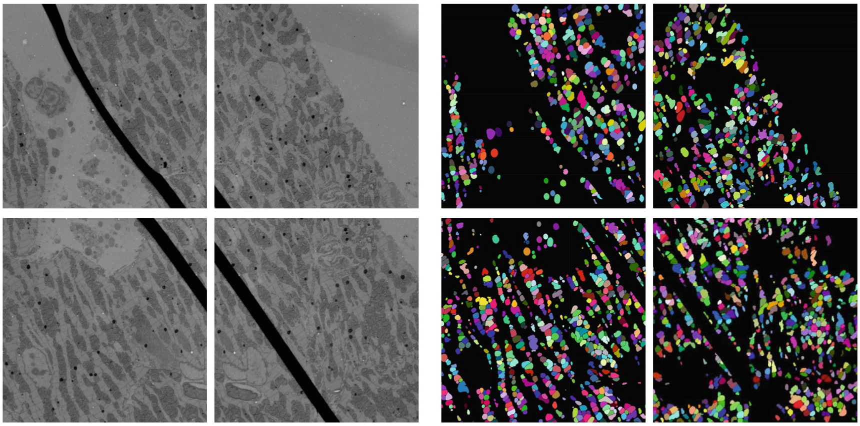

Fig. 2. EM data set used here. Left: Set of four 2D EM tiles of 16384 × 16384 pixels each at a resolution of 2.5 nm/pixel. Right: Corresponding ground truth instance maps.

3.2 Coarse Semantic Segmentation

The first stage utilizes feature embeddings of image patches from networks pre-trained using unsupervised contrastive learning. The contrastive loss function L compares pairs of image representations to separate representations from different images and brings together those from different views of the same image:

$$L = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j=1}^{N}[y(i,j)\cdot d(f_i,f_j) + (1-y(i,j))\cdot \max(margin-d(f_i,f_j),0)] \quad (1)$$

where N is the number of training samples, $y(i,j)$ is a flag indicating whether the pair of features (f_i, f_j) is from the same image $(y(i,j) = 1)$ or different images $(y(i,j) = 0)$, $d(f_i, f_j)$ is the distance between the features of images i and j (e.g., Euclidean distance), and *margin* is a hyperparameter that controls the distance between features from different images.

We use 128 × 128 pixel-sized patches with 50% overlap to partition a given 2D EM image. These values are chosen as they provide a good tradeoff between information content and region homogeneity. The embeddings contain inherent distances that distinguish similar input image patches from dissimilar ones, which are then clustered using K-means into relevant regions. By using a graphical user interface, a biologist then manually selects the clusters that correspond to the regions of interest, i.e., those containing mitochondria. The output of this first component in our pipeline is a binary map that is produced by merging all patches that belong to the selected clusters[2], Fig. 3.

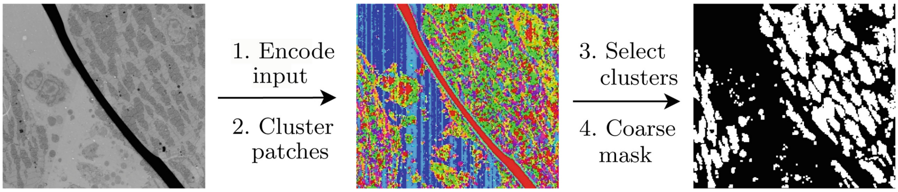

Fig. 3. Coarse semantic segmentation. An encoder extracts features from input image patches followed by clustering and selection of clusters to produce the coarse mask.

3.3 Fine Instance Segmentation

Fine instance segmentation is achieved by simultaneously processing each connected component in the binary coarse semantic map. This part of our pipeline consists of the following steps: a) membrane delineation with the inhibition-augmented COSFIRE filter, b) watershed segmentation, and c) object selection.

A. Inhibition-Augmented COSFIRE Filter. A COSFIRE filter can be configured to be selective for any given pattern of interest. For this application, where the goal is to delineate boundaries, we configure a COSFIRE filter to be selective for lines. It takes input from a linearly aligned set of responses of a difference-of-Gaussians (DoG) filter. We denote by B a line-selective COSFIRE filter, which is defined as a set of 3-tuples:

$$B = \{(\sigma_i, \rho_i, \phi_i) \mid i = 1, \ldots, n\} \tag{2}$$

where each tuple i indicates the distance ρ_i and the polar angle ϕ_i of the response of a DoG filter whose outer standard deviation is σ_i. The inner standard deviation of the DoG function is set to $0.5\sigma_i$. The COSFIRE filter's response $r_B(x, y)$ in a given (x, y) location is the geometric mean of the n DoG responses at the polar coordinates defined in B, with respect to (x, y). For a more in-depth

[2] Effectively, user selection of clusters can be assisted by cluster validity indices, in that the user gets automatic suggestions of which other clusters are mostly similar to the already selected ones.

explanation of the technical details and how COSFIRE filters achieve rotation-invariance, we refer the reader to [16].

COSFIRE filters can be augmented with surround suppression in the same way as originally proposed in [17]. This is needed here to accentuate the membranes while ignoring the inner cristae for the delineation of mitochondria. The surround inhibition term is computed for every (x, y) location by convolving a normalized center-off DoG function I_γ (γ denotes the standard deviation of the inner Gaussian function) with the COSFIRE response map r_B. Further to [3] the standard deviation of the outer Gaussian function is set to 4γ. Normalization of this DoG kernel consists of first applying the Heaviside step function, which maps all negative values to zero and all positive values to 1. Then all values of one are L_1-normalized such that their sum equals to 1. The final COSFIRE response map R is then achieved by the linear function:

$$R = r_B - \alpha r_{I_\gamma} \tag{3}$$

where α denotes the inhibition strength. Figure 4 shows examples of COSFIRE response maps for different α values. The inhibition term suppresses responses to spurious strokes (i.e. cristae) in the surrounding of mitochondria walls.

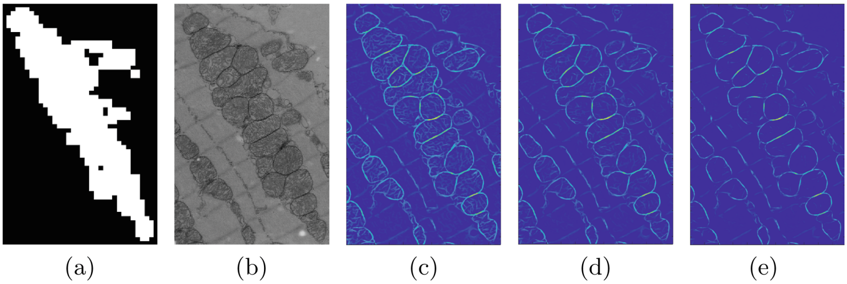

Fig. 4. Examples of boundary delineations with a COSFIRE filter. (a) A connected component from the coarse segmentation map, (b) the corresponding EM region, and COSFIRE response maps for (c) $\alpha = 0$, (d) $\alpha = 1$ and (e) $\alpha = 2$.

The response map R is transformed to a binary contour map by first thinning R with non-maximum suppression to obtain the ridges and then by applying hysteresis thresholding, which is characterized by the high t_h and low t_l threshold values. We keep t_h as a hyperparameter and set $t_l = 0.5t_h$.

B. Watershed Segmentation. First, the Euclidean distance map is computed from the thresholded COSFIRE binary map obtained above and all values below the mean distance are set to zero. The resulting thresholded distance map is used to generate the first watershed output (Fig. 5b). In the second stage, the ridges of the watershed output of the first stage are superimposed on the thresholded distance map (Fig. 5c) and used to generate the final watershed output (Fig. 5d).

C. Object Selection. First, the objects that fall outside the coarse semantic mask are removed. For the remaining components, we compute the contrast from the gray-level co-occurrence matrix (GLCM) determined from the corresponding intensity pixels of the input image and keep all objects with a contrast less than λ standard deviations from the mean.

4 Experiments and Results

We evaluate the performance of the proposed method in three different setups. The first two, which we denote by UG and US, use (U)nsupervised semantic segmentation with networks that are pre-trained on the (G)eneral ImageNet dataset [9] and on the (S)pecific CEM1.5M dataset [10] of EM images with many instances of mitochondria, respectively. For the third approach, denoted by SS, we replace the unsupervised stage with the state-of-the-art MitoNet, which is a (S)upervised ConvNet trained for (S)emantic segmentation of mitochondria. Finally, we compare the results of these three methods with the (S)supervised (I)nstance segmentation variant of MitoNet [18], denoted by SI.

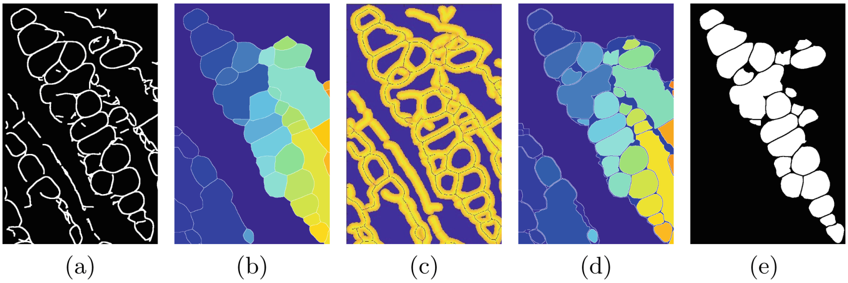

Fig. 5. Example of fine instance segmentation from a COSFIRE contour map. (a) COSFIRE binary map, which is used as input to the (b–d) watershed algorithm followed by (e) object selection to achieve the final instance segmentation.

Performance Measures. We measure two performance indicators, namely the global similarity measure Intersection-Over-Union (IoU) and the Panoptic Quality (PQ), which is a more detailed measure suitable for instance segmentation. IoU is the intersection between the predicted (PR) and ground truth (GT) masks divided by the union of the two masks, across all pixels in a given image. PQ unifies both segmentation and detection, making it a useful metric for cellular EM segmentation [19]. They are defined as:

$$IoU = \frac{PR \cap GT}{PR \cup GT} \tag{4}$$

$$PQ = \underbrace{\frac{\Sigma_{j \in TP} IoU(GT^j, PR^{j*})}{|TP|}}_{\text{Segmentation Quality (SQ)}} \times \underbrace{\frac{|TP|}{|TP| + \frac{1}{2}|FP| + \frac{1}{2}|FN|}}_{\text{Detection Quality (DQ)}} \tag{5}$$

where TP, FP, and FN stand for the number of true positive, false positive, and false negative objects, respectively. Following the mitochondria instance segmentation work in [8], we consider a mitochondrium as TP if it has at least 30% IoU overlap with a GT object. The FP and FN objects are the unmatched segments in PR and GT, respectively. $PR^{j,*}$ denotes the object in PR that is matched with the largest overlapping region (in IoU) with GT^j.

Experiments. Pre-trained encoders (ResNet50) of UG and US were applied to all 128 × 128 sized patches of the four tiles, which represented each of them with a 2048-element feature vector obtained from the last layer of the encoder. The vectors were then min-max normalized. Next, we applied truncated SVD to reduce the dimensions from 2048 to 1000, in order to enhance clustering effectiveness by eliminating noise and irrelevant features. These lower-dimensional vectors were then clustered using K-Means with (K =) 10 clusters. Finally, three and four clusters, respectively, were selected for the UG and US methods by visually inspecting the clustering results.

For the second stage, we applied a grid search to fine-tune three parameters of the COSFIRE filters, namely σ, α, and t_H, which are related to the contour thickness, inhibition strength, and hysteresis thresholding, respectively, along with the parameter λ which we used in the object selection step. The fine-tuning was done on the single component shown in Fig. 4 and Fig. 5, which was randomly selected from the coarse semantic segmentation in the first stage. The random selection was constrained to pick a component with 10 to 20 mitochondria. The determined parameters are: $\sigma = 4,\ \alpha = 2,\ t_H = 0.6$, and $\lambda = 2.5$.

Table 1. Comparison of the coarse semantic segmentation outputs using IoU.

Method	Tile 1	Tile 2	Tile 3	Tile 4
UG	0.64	0.69	0.67	0.64
US	0.66	0.69	0.69	0.68
SS	0.81	0.84	0.81	0.83

Results. We report two sets of results. Table 1 presents the IoUs of the UG, US, and SS that measure the quality of the coarse semantic segmentation for each of the four tiles with respect to GT. The second set of results is illustrated in Fig. 6 shows PQ – the product of the segmentation quality (SQ) and detection quality (DQ) – that measures the quality of the final instance segmentation. The consistently high SQ of the UG, US, and SS methods is attributable to the precise delineation by the COSFIRE filter, which yields fine instance segmentation masks. The DQ metric indicates the effectiveness of detecting the right components. While our UG and US unsupervised variants achieve modest IoUs in the first stage due to under-segmentation, their final detection quality outperforms that of the supervised counterparts. Among them, the US approach achieves the best performance, which can be attributed to the fact that the underlying encoder was pre-trained on the dataset CEM1.5 of EM images.

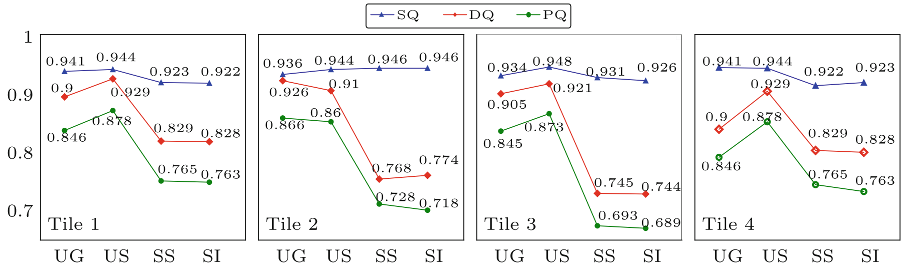

Fig. 6. Line plots of SQ, DQ, and PQ for all four tiles. The two unsupervised variants UG and US show consistent superiority over their supervised counterparts SS and SI.

5 Discussion and Conclusion

The results of the instance segmentation indicate that the proposed unsupervised variants, UG and US, of the COFI approach perform substantially better than the supervised approach despite having very coarse segmentation maps. This improvement is attributable to the COSFIRE operator, whose inhibition component makes it particularly effective in delineating the walls of apposing mitochondria in challenging backgrounds. The initial stage of the COFI method has the greatest influence on the detection quality (DQ). Any missing components from the first stage cannot be recovered by the COSFIRE filter in the second stage. It is also remarkable that for our images although the UG method uses an encoder that was pre-trained on ImageNet, it still yields very high results that come very close to the best results achieved with an encoder that was pre-trained on the more specific CEM1.5 dataset of EM images (US). To gain more insight, we augment the COSFIRE delineation operator with a supervised semantic segmentation approach (SS) based on MitoNet. The results show that the COSFIRE operator performs equally well as the supervised instance segmentation (SI) on MitoNet-based semantic maps.

The proposed COFI approach is unsupervised and versatile, in that the structure of interest (mitochondria here) is not an intrinsic component. The patch-based classification of high-resolution EM images provides the necessary redundancy to capture semantically important textured regions, which is then fine-tuned in the second stage by the COSFIRE filter. The COSFIRE filter with inhibition turned out to be very robust in delineating the mitochondria walls from the cristae within them. In future work, we will evaluate the proposed COFI approach on bigger datasets, other cellular tissues and different sub-cellular structures that are important for the study of biological processes.

Acknowlegements. This project has received funding from the Centre for Data Science and Systems Complexity at the University of Groningen (www.rug.nl/research/fse/themes/dssc/). Part of the work has been performed in the UMCG Microscopy and Imaging Center (UMIC), sponsored by ZonMW grant 91111.006 and the Netherlands Electron Microscopy Infrastructure (NEMI), NWO National Roadmap for Large-Scale

Research Infrastructure of the Dutch Research Council (NWO 184.034.014). Dr. Westenbrink is supported by the Netherlands Organisation for Scientific Research (NWO VENI, grant 016.176.147) and the Netherlands Heart Foundation Senior Clinical Scientist Grant (2019T064). Thanks also to Anouk Wolters for technical assistance.

References

1. de Boer, P., et al.: Large-scale electron microscopy database for human type 1 diabetes. Nat. Commun. **11**(1), 1–9 (2020)
2. Titze, B., Genoud, C.: Volume scanning electron microscopy for imaging biological ultrastructure. Biol. Cell **108**(11), 307–323 (2016)
3. Melotti, D., Heimbach, K., Rodríguez-Sánchez, A., Strisciuglio, N., Azzopardi, G.: A robust contour detection operator with combined push-pull inhibition and surround suppression. Inf. Sci. **524**, 229–240 (2020)
4. Lucchi, A., Li, Y., Fua, P.: Learning for structured prediction using approximate subgradient descent with working sets. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1987–1994 (2013)
5. Luo, Z., Wang, Y., Liu, S., Peng, J.: Hierarchical encoder-decoder with soft label-decomposition for mitochondria segmentation in EM images. Front. Neurosci. **15**, 687832 (2021)
6. Yuan, Z., Ma, X., Yi, J., Luo, Z., Peng, J.: HIVE-Net: centerline-aware hierarchical view-ensemble convolutional network for mitochondria segmentation in EM images. Comput. Methods Programs Biomed. **200**, 105925 (2021)
7. Oztel, I., Yolcu, G., Ersoy, I., White, T., Bunyak, F.: Mitochondria segmentation in electron microscopy volumes using deep convolutional neural network. In: 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1195–1200. IEEE (2017)
8. Wei, D., et al.: MitoEM dataset: large-scale 3D mitochondria instance segmentation from EM images. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12265, pp. 66–76. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59722-1_7
9. Chen, T., Kornblith, S., Swersky, K., Norouzi, M., Hinton, G.E.: Big self-supervised models are strong semi-supervised learners. Adv. Neural. Inf. Process. Syst. **33**, 22243–22255 (2020)
10. Conrad, R., Narayan, K.: CEM500K, a large-scale heterogeneous unlabeled cellular electron microscopy image dataset for deep learning. Elife **10**, e65894 (2021)
11. Azzopardi, G., Petkov, N.: A CORF computational model of a simple cell that relies on LGN input outperforms the gabor function model. Biol. Cybern. **106**, 177–189 (2012)
12. Azzopardi, G., Strisciuglio, N., Vento, M., Petkov, N.: Trainable COSFIRE filters for vessel delineation with application to retinal images. Med. Image Anal. **19**(1), 46–57 (2015)
13. Strisciuglio, N., Petkov, N.: Delineation of line patterns in images using B-COSFIRE filters. In: 2017 International Conference and Workshop on Bioinspired Intelligence (IWOBI), pp. 1–6. IEEE (2017)
14. Strisciuglio, N., Azzopardi, G., Petkov, N.: Robust inhibition-augmented operator for delineation of curvilinear structures. IEEE Trans. Image Process. **28**(12), 5852–5866 (2019)
15. Schneider, C.A., Rasband, W.S., Eliceiri, K.W.: NIH image to ImageJ: 25 years of image analysis. Nat. Methods **9**(7), 671–675 (2012)

16. Azzopardi, G., Rodríguez-Sánchez, A., Piater, J., Petkov, N.: A push-pull CORF model of a simple cell with antiphase inhibition improves snr and contour detection. PLoS ONE **9**(7), e98424 (2014)
17. Grigorescu, C., Petkov, N., Westenberg, M.: Contour detection based on nonclassical receptive field inhibition. IEEE Trans. Image Process. **12**(7), 729–739 (2003)
18. Conrad, R., Narayan, K.: Instance segmentation of mitochondria in electron microscopy images with a generalist deep learning model trained on a diverse dataset. Cell Syst. **14**(1), 58–71 (2023)
19. Kirillov, A., He, K., Girshick, R., Rother, C., Dollár, P.: Panoptic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9404–9413 (2019)