Bioimage informatics

PS-Net: human perception-guided segmentation network for EM cell membrane

Ruohua Shi [1,2,3], Keyan Bi[4,5,6,7], Kai Du[8], Lei Ma[2,3,9], Fang Fang[4,5,6,7], Lingyu Duan[2,10], Tingting Jiang[1,2,*], Tiejun Huang[2,3]

[1]Advanced Institute of Information Technology, Peking University, Hangzhou, Zhejiang 310000, China
[2]National Engineering Research Center of Visual Technology, National Key Laboratory for Multimedia Information Processing, School of Computer Science, Peking University, Beijing 100871, China
[3]Beijing Academy of Artificial Intelligence, Beijing 100084, China
[4]School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China
[5]IDG/McGovern Institute for Brain Research, School of Psychological and Cognitive Sciences, Peking University, Beijing 100871, China
[6]Key Laboratory of Machine Perception (Ministry of Education), Peking University, Beijing 100871, China
[7]Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100084, China
[8]Institute for Artificial Intelligence, Peking University, Beijing 100871, China
[9]National Biomedical Imaging Center, College of Future Technology, Peking University, Beijing 100871, China
[10]Peng Cheng Laboratory, Shenzhen 518066, China

*Corresponding author. School of Computer Science, Peking University, Beijing 100871, China. E-mail: ttjiang@pku.edu.cn (T.J.)

Associate Editor: Hanchuan Peng

Abstract

Motivation: Cell membrane segmentation in electron microscopy (EM) images is a crucial step in EM image processing. However, while popular approaches have achieved performance comparable to that of humans on low-resolutionHanch.(n)-22atasets, they have shownlimited success when applied to high-resolution2atasets. The human system, onthe other hand, displays consistently excellent performance onboth low and high resolutions. To better understand this limitation, we conducted eye movement and perceptual consistency experiments. Our2ata showed that human observersare sensitive to the structure of the membrane while toleratingmisalignment, contrary to commonlyused evaluation criteria.ly, our results indicatedthat the human visual system processes images in both global–localand coarse-to-fine manners.

Results: Based on these observations, we propose a computational framework for membrane segmentation that incorporates these characteristics of human perception. This framework includes a novel evaluation metric, the perceptual Hausdorff distance (PHD), and an end-to-end network called the PHD-guided segmentation network (PS-Net) that is trained using adaptively tuned PHD loss functions and a multiscale architecture. Our subjective experiments showed that the metric is consistent with human perception than other criteria, and our

and identify the underlying causes for these differences in performance.

During our investigation into the differences in performance between humans and DL methods on cell membrane segmentation, we noticed that there is a discrepancy between human perception and commonly used evaluation criteria, such as the *F*1 score (Sasaki *et al.* 2007), IoU (Kosub 2019), and Betti number error (Betti) (Hu *et al.* 2021). For example, in Fig. 1, (b) is the ground truth of the cell membranes in (a), while (c) and (d) are two predictions by different algorithms. According to the *F*1, IoU, and Betti scores, prediction (d) is better than prediction (c). However, from a human perspective, the opposite is true because (d) lacks some important structures. To better understand this discrepancy, we conducted a subjective experiment in which subjects were shown three images: the ground truth and two different predictions. They were asked to indicate which prediction was more similar to the ground truth. We evaluated the consistency between the preferences of these criteria and humans. Surprisingly, results showed that these evaluation criteria are only 30%–40% consistent with human perception.

To better understand the mechanisms of the human visual system when comparing two images of cell membranes, we conducted an eye movement experiment to record subjects' saccades and fixations. In the experiment, subjects were shown two images of cell membranes side by side, such as the ground truth (b) and prediction (c) in Fig. 1e. Heatmaps and arrows were used to indicate fixations and saccades, respectively. Based on the data collected from eye movements, we found that humans focus primarily on the structure of membranes while using quick glances to compare other regions. For example, the red regions of the heatmaps correspond to junctions of cell membranes in (e) and missing edges in (f). This suggests that humans pay more attention to the skeleton of the cell membrane and missing edges, while ignoring thickness and misalignment errors. Additionally, we observed that humans use a global-local strategy and a coarse-to-fine approach to find differences. Specifically, according to the

proposed PS-Net outperforms existing methods on all evaluation criteria. We also demonstrate the versatility of our method by applying it to natural image segmentation datasets, where it also demonstrates state-of-the-art performance.

2 Related works

EM cell membrane segmentation, which can also be viewed as cell boundary detection, is a critical step in EM image analysis for neuron reconstruction. This task is more challenging than similar tasks on natural images, such as "delineation detection," due to the higher resolution, more complex structures, and more detailed information present in EM images. Since the release of the first annotated EM image dataset in the ISBI 2012 challenge (Arganda-Carreras *et al.* 2015), several extraordinary DL methods have been developed for this task. U-Net (Ronneberger *et al.* 2015) is a popular and successful DL model for biomedical image segmentation. Subsequent research efforts (Paszke *et al.* 2016, Chaurasia and Culurciello 2017, Shen *et al.* 2017, Yu *et al.* 2017, Hu *et al.* 2018, Khadangi *et al.* 2021) have sought to further improve EM segmentation performance using a U-shaped encoder–decoder architecture and effective feature extraction techniques, such as dual-channel blocks (Lou *et al.* 2021) and skip connections (Chaurasia and Culurciello 2017). These methods have achieved near-human performance on the ISBI 2012 dataset. However, as EM imaging techniques have advanced, the demand for the segmentation of ultra-high-resolution images has increased. For instance, the recently proposed U-RISC dataset (Shi *et al.* 2022) has a resolution of $120 \times 9958 \times 9959$. When applied to this dataset, the performance of these methods significantly decreased (from 98% on ISBI 2012 to 60% on U-RISC). This suggests that algorithms should not only focus on effectively extracting features from limited labeled images, but should also incorporate human-based strategies.

Evaluation for cell membrane segmentation. In the cell membrane segmentation task, both pixel accuracy and topographic accuracy are important. There are three main categories of evaluation criteria (Yeghiazaryan and Voiculescu 2018) that have been proposed for image segmentation: "pixel-wise" criteria, "topology-wise" criteria, and "point-wise" criteria. "Pixel-wise" criteria, such as the $F1$ and IoU, treat segmentation as a pixel-wise binary classification task and use statistics to evaluate the performance of models. These criteria are often used as optimization objectives, with popular loss functions including the cross-entropy loss and its variations (Chen *et al.* 2019, Khadangi *et al.* 2021), as well as the Dice loss (Dice 1945). "Topology-wise" criteria, like V-Rand and V-Info (Arganda-Carreras *et al.* 2015) consider both merge and split errors of membranes in their evaluation. Bs inesitstiombranes i04TD[(limited)-421.8m6180TD[(et)-7al.and use

provided insight into how humans visually compare images of cell membranes, leading to the development of a new evaluation criterion based on human perception, known as PHD.

In the design of the PHD, we consider the structural information of cell membranes and the human tolerance for slight misalignment. On the one hand, to capture the structural information, we represent membranes as point-sets and use the modified Hausdorff distance (Huttenlocher *et al.* 1993). As the results of eye movement experiments show, humans are more sensitive to changes in structure than to changes in thickness of membranes. Therefore, to alleviate the influence of the thickness change, the PHD extracts structural information (skeleton) from the segmentation results and represents it

the consistency with human perception initially increased before slowly decreasing to 0, indicating that humans do have tolerance for a certain level of offset. These results suggest that humans tend to tolerate small perturbations in cell membrane segmentation.

It is worth investigating whether using skeletonization can improve the performance of other evaluation metrics. The results in Fig. 3 show that using skeletonization can help some metrics, such as F1, IoU, and ASSD, to a certain extent. However, the consistency of $F1-SK$ is only 34.51%, and the consistency of IoU$-SK$ is 44.25%. These values are still far from the performance of PHD. This suggests that simply extracting the membrane skeleton is not sufficient to address the limitations of existing metrics.

5 PHD-guided segmentation network

Inspired by the PHD criterion and the global–local strategy with a coarse-to-fine approach observed in the eye movement experiment, we propose the PS-Net. This network includes a multiscale architecture with loss functions specifically designed to guide the segmentation process using PHD.

5.1 Overview of architecture

An overview of the network is depicted in Fig. 4. PS-Net consists of two branches for multiscale image segmentation: the "global branch" S_G, which uses the full image as input, and the local branch S_L, which uses N patches of the cropped original image with the same size as input. Both branches use the same u-shaped encoder–decoder architecture to make probability predictions, as well as a module for skeleton extraction. The global and local predictions are then combined to produce the final segmentation result.

5.1.1 Backbone

The U-Net (Ronneberger *et al.* 2015) is a convolutional neural network with a contracting path that captures contextual information and an expansive path that enables precise localization. It is often used as an encoder–decoder module in image segmentation tasks. In this work, the U-Net is utilized in the global and local branches of the PS-Net for probability prediction.

5.1.2 Skeleton extraction module

The structure extraction module of PS-Net uses the modified differentiable Zhang–Suendiffereinn.4(pa560.8lgJ-thmpa56ined.00041rg5

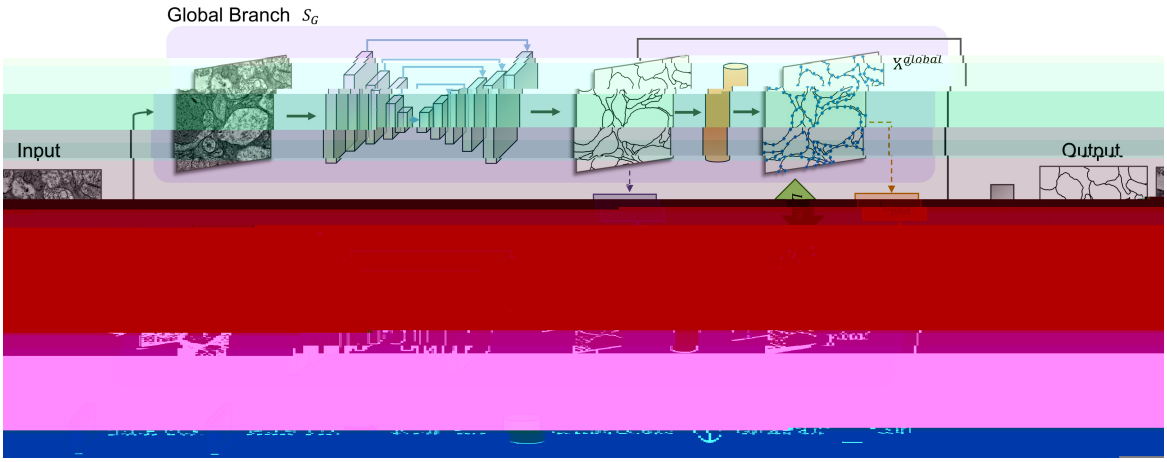

Figure 4. An overview of PS-Net. PS-Net has two branches to segment multiple scales of the input image. In the global branch, the u-shape segmentation module uses the original image as input and outputs its membrane probability map. In the local branch, the original image is cropped into N patches with the same size. Then, the patches are put into the segmentation module with N prediction maps as outputs. The two branches share weights during the training process. The structure extraction module is designed to compute the skeletons of the all the N predictions. Three loss functions: pixel-wise loss, PHD loss, and similarity loss are calculated during the training. PS-Net outputs the prediction from the results of two branches.

between the ground truth and predictions in both the global and local branches. As shown in Equation (4), the function compares the skeleton point-sets of the predictions, represented by X^{global} and X^{local}, with their respective ground truth, represented by Y^{global} and Y^{local}. In order to compute the loss for backpropagation, the soft-max function is applied to the likelihood map for binarization and the derivative of the binary image is shown in Supplementary Section S7.

$$L_{phd} = d_{PHD}(X^{global}, Y^{global}) + d_{PHD}(X^{local}, Y^{local}). \quad (4)$$

In addition, the similarity loss L_{sim} is used to measure the similarity between the global and local scales by calculating the PHD distance between the skeleton point-sets of the global prediction (X^{global} and the stitched local predictions $\hat{X}^{local}$). It is designed as $L_{sim} = d_{PHD}(X^{global}, \hat{X}^{local})$. The stitched local predictions $\hat{X}^{local}$ are obtained by stitching the skeleton point-sets X^{local} from the local branch, and have the same size as the global skeleton point-sets X^{global}. This loss helps to ensure that the prediction from the global branch and the stitched prediction from the local branch are consistent in terms of structure.

5.3 Coarse-to-fine training

During the training process, three loss functions are optimized with a coarse-to-fine strategy, which aims to assist the network focusing more on generating a coarse segmentation result, and then subsequently shifting to detailed information. Correspondingly, in our method, L_{pixel} measures the accuracy of each pixel in the image, which is the low-level (local) feature, while L_{phd} and L_{sim} measure the structural difference of membranes, which is the high-level (global) feature. In contrast to the two-stage refinement approach utilized by Chen *et al.* (2019), PS-Net employs pixel-wise loss for the first several epochs as to generate a coarse segmentation result. And then, to get a finer cell membrane structure, the weights of PHD loss and similarity loss are adaptively raised with the number of training epochs. Let λ_1 and λ_2 be the adaptive weights of L_{phd} and L_{sim}. The final loss function of PS-Net L is shown in Equation (5). The details of the parameter settings are shown in Supplementary Section S4.2.

$$L = L_{pixel} + \lambda_1 L_{phd} + \lambda_2 L_{sim}. \quad (5)$$

6 Segmentation experiments

The performance of PS-Net was evaluated on two EM image datasets. Results show that PS-Net outperforms existing methods. Then, ablation studies were performed to isolate the individual contributions of the main components and parameters of our approach. Furthermore, PS-Net was extended to two natural image segmentation datasets with SOTA performance.

6.1 Experiments on EM image datasets

We evaluated our method on two EM datasets: ISBI 2012 and U-RISC. We used a 3-fold cross-validation to tune hyperparameters for both our proposed method and eight baseline methods. The evaluation metrics included $F1$ score, IoU, V-Rand, V-Info, TPVF, TNVF, Hausdorff distance, and our proposed PHD-τ, where τ is the tolerance threshold. The baseline methods included U-Net (Ronneberger *et al.* 2015), CASENet (Yu *et al.* 2017), LinkNet (Chaurasia and Culurciello 2017), GLNet (Chen *et al.* 2019), SENet (Hu *et al.* 2018), U-Net++ (Zhou *et al.* 2018), Mosin. (Mosinska *et al.* 2018), and DMT (Hu *et al.* 2021). We report the mean and SD performance over the test set for all the methods. More details about the datasets, baseline models, and evaluation metrics are provided in Supplementary Section S4. * represents that the predicted results for evaluation are re-implemented by the official code.

For the ISBI 2012 dataset, our method achieves SOTA performance as reported in Table 1. We also summarized some leading quantitative results reported in original papers in Supplementary Sections S4.3 and S4.4. The results show that PS-Net obtained the best scores on all of these metrics (as shown in bold font). More visualizations of segmentation results are depicted in Fig. 5. Our method has fewer mistakes.

Table 1. Quantitative results of the methods on ISBI 2012 dataset.

Metrics	U-Net*	CASENet*	LinkNet*	GLNet*	SENet*	U-Net++*	Mosin.*	DMT*	PS-Net
*F*1 (%)	92.01±0.02	87.99±0.05	89.40±0.04	90.41±0.02	91.35±0.02	93.01±0.02	82.30±0.03	92.93±0.02	**93.98±0.02**
IoU (%)	92.31 ±0.01	89.61±0.01	91.02±0.01	81.89±0.02	84.24±0.01	89.56±0.02	90.88±0.01	92.19±0.01	**93.99±0.01**
V-Rand (%)	96.33±0.02	96.53±0.32	96.99±0.04	95.69±0.07	94.54±0.04	95.81±0.05	95.99±0.04	96.74±0.06	**98.37±0.02**
V-Info (%)	96.01±0.02	96.27±0.03	95.01±0.04	96.56±0.02	96.42±0.01	97.07±0.05	95.81±0.05	97.82±0.01	**98.75±0.02**
TNVF (%)	94.61±0.01	93.12±0.02	93.08±0.03	93.52±0.02	92.04±0.03	94.25±0.03	94.66±0.01	94.67±0.02	**94.68±0.01**
TPVF (%)	91.96±0.04	91.77±0.05	89.94±0.07	91.80±0.04	90.49±0.03	91.45±0.02	92.04±0.03	92.75±0.02	**93.00±0.04**
ASSD ↓	2.689±1.92	3.157±1.13	3.921±2.15	3.036±2.01	3.002±1.19	2.994±2.05	3.015±1.87	3.845±1.86	**2.041±1.98**
HD ↓	55.94±10.4	59.87±17.0	63.12±28.1	83.12±17.0	72.46±19.4	60.35±10.5	93.03±19.2	84.94±13.6	**54.62±13.8**
PHD-0 ↓	5.950±2.06	6.013±1.05	5.814±4.52	6.989±3.57	5.362±2.50	4.205±3.95	4.833±2.97	4.374±3.19	**3.954±1.04**
PHD-3 ↓	5.650±2.07	5.990±1.02	5.627±4.66	6.884±3.10	5.028±2.18	4.002±3.65	4.629±2.54	4.081±3.02	**3.661±1.25**
PHD-5 ↓	3.663±1.91	4.280±3.82	3.631±0.82	5.299±3.84	3.716±2.11	3.769±3.02	3.894±1.18	3.351±2.44	**3.042±1.53**
PHD-10 ↓	2.414±1.08	2.997±2.05	2.146±1.07	3.017±2.58	2.631±1.98	2.877±1.94	2.510±1.01	1.993±2.31	**1.045±0.99**
PHD-50 ↓	0.280±0.01	0.241±0.03	0.351±0.02	0.26±0.017	0.291±0.02	0.238±0.03	0.274±0.03	0.286±0.04	**0.244±0.01**

The boldface values indicate the best performance.

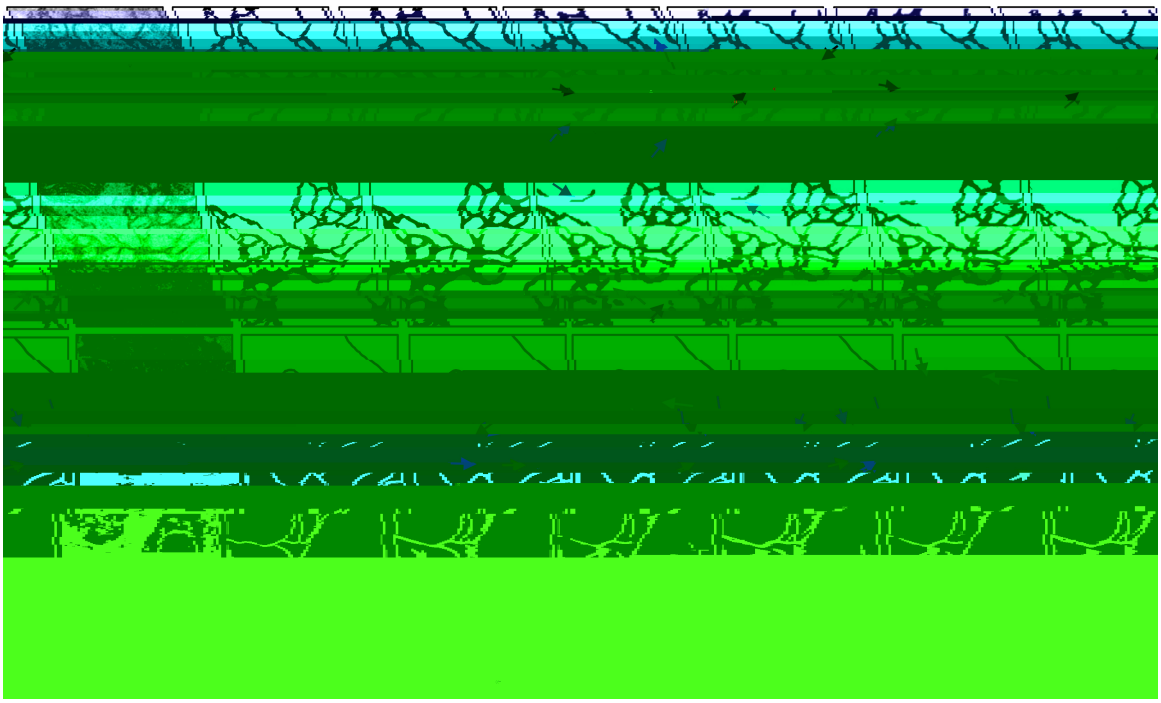

Figure 5. Segmentation results of ISBI 2012 (first two rows) and U-RISC (last two rows) datasets. Red arrow: false negative error. Blue arrow: false positive error.

More visualization results are shown in Supplementary Section S4.7.

For the U-RISC dataset, we first summarize the top four results reported in the leaderboard of the challenge (Supplementary Section S4.4). Our method has reached the best performance (promote approximately **11.5**% more than the winning team in the challenge). Similar to the experiments of ISBI 2012, to compare more results of other metrics, we train and test the six competitive methods, using the data division the same as the challenge. The scores and SD of eight evaluation metrics are reported on the testing images. The results in Table 2 show that PS-Net outperformed the other methods. In particular, it not only greatly improves the *F*1 score, but also performs well in other metrics. In addition, we observed an apparent decline of the PHD-τ scores at $\tau = 10$ and $\tau = 50$ for ISBI 2012 and U-RISC, respectively, which showed that the U-RISC was a more challenging dataset to gain a fine segmentation. Compared with other methods, our method is able to alleviate the missing structures and redundant predictions (as shown in Supplementary Section S4.8).

6.2 Ablation study on U-RISC

To evaluate the effectiveness of the proposed two strategies and three loss functions, we conducted several ablation experiments on the U-RISC dataset.

6.2.1 PHD-based loss functions

To evaluate the effectiveness of PHD-based loss functions, we trained the model using only the pixel-wise loss L_{pixel}, and then added the PHD loss L_{phd} and similarity loss L_{sim}. The results are presented in Table 3, where L_1, L_2, and L_3 represent L_{pixel}, L_{phd}, and L_{sim}, respectively. The results show that L_{phd} improved the performance of the three architectures. In particular, the *F*1 score increased by ∼6.63% and the PHD-0 score decreased by ∼1.83 when using L_{phd}. Additionally, the combination of L_{sim} with L_{phd} resulted in an ∼1.49% increase

Table 2. Quantitative results of the methods on U-RISC dataset.

Metrics	U-Net*	CASENet*	LinkNet*	GLNet*	SENet*	U-Net++*	Mosin.*	DMT*	PS-Net
*F*1 (%)	48.83±0.02	60.07±0.05	60.70±0.04	58.10±0.04	52.12±0.05	60.30±0.05	47.56±0.09	39.68±0.05	**67.69±0.02**
IoU (%)	32.33±0.02	43.07±0.05	**43.69±0.05**	41.05±0.04	35.41±0.05	43.29±0.04	40.29±0.08	37.98±0.06	43.63±0.03
V-Rand (%)	49.38±0.03	59.21±0.05	63.10±0.04	53.41±0.04	52.88±0.05	62.11±0.04	49.75±0.05	50.37±0.04	**68.93±0.02**
V-Info (%)	51.20±0.04	60.13±0.04	62.39±0.03	54.33±0.04	51.78±0.06	62.34±0.04	58.64±0.03	59.27±0.05	**65.32±0.03**
TNVF (%)	88.62±0.02	96.22±0.05	96.02±0.03	95.72±0.04	97.68±0.05	95.92±0.03	94.25±0.05	96.31±0.04	**97.82±0.02**
TPVF (%)	35.24±0.03	56.04±0.04	55.62±0.04	53.39±0.04	52.91 ±0.04	54.93±0.04	54.99±0.03	53.77±0.05	**56.17±0.03**
ASSD ↓	10.16±8.14	9.314±3.51	9.201±4.43	11.96±9.45	12.11±6.34	9.106±4.52	19.67±10.3	13.04±8.45	**7.808±4.15**
HD ↓	271.5±31.1	566.1±32.2	352.9±29.9	399.3±39.1	547.3±38.0	414.0±31.9	484.6±51.5	683.9±82.4	**252.8±30.2**
PHD-0 ↓	18.65±9.72	19.25±9.33	22.72±6.93	23.30±6.46	20.42±5.22	17.25±7.33	24.54±8.98	29.56±9.57	**15.29±5.79**
PHD-3 ↓	17.93±8.52	19.01±10.2	22.70±6.92	23.15±6.05	19.86±6.13	16.99±8.21	24.26±7.38	29.56±8.48	**15.01±6.29**
PHD-5 ↓	17.37±6.25	16.72±9.15	20.41±7.81	21.25±5.74	17.47±10.0	16.55±7.01	22.85±8.62	28.78±10.3	**13.52±5.03**
PHD-10 ↓	8.512±5.10	10.38±6.99	11.90±8.66	11.53±6.03	9.93±7.23	8.99±6.72	19.48±7.29	18.67±6.27	**6.979±6.67**
PHD-50 ↓	6.501±1.17	10.25±5.64	5.436±2.82	5.170±4.62	3.201±2.53	4.312±2.97	15.47±6.13	14.52±6.29	**1.594±2.06**

The boldface values indicate the best performance.

Table 3. Ablation study for the architectures and loss functions of PS-Net on U-RISC dataset.

Method	L_1	L_2	L_3	*F*1 (%)	V-Rand (%)	V-Info (%)	PHD-0↓	PHD-5↓	PHD-10↓	PHD-50↓
S_G	✓			51.57	53.01	53.92	21.61	19.59	10.42	7.227
S_L	✓			58.23	56.94	57.05	23.53	20.41	12.92	8.039
$S_G + S_L$	✓			59.57	58.71	59.80	20.91	17.04	9.367	6.294
S_G	✓	✓		53.81	54.78	54.79	17.58	16.11	8.829	3.142
S_L	✓	✓		61.98	63.62	61.03	17.14	16.32	8.994	3.035
$S_G + S_L$	✓	✓		66.20	67.24	65.00	16.07	15.21	7.878	2.770
$S_G + S_L$	✓	✓	✓	**67.69**	**68.93**	**65.32**	**15.29**	**13.52**	**6.969**	**1.594**

The boldface values indicate the best performance.

in the *F*1 score and a decrease of ∼0.78–1.176 in the PHD score. This indicates that the structure of the cell membrane plays an important role in its segmentation. Furthermore, for the selection of the tolerance, we conducted the ablation experiments summarized in Supplementary Table S5. The results show that PS-Net achieves the best performance when τ=2. Further, we also compared the PHD loss with another topology loss, clDice loss (Shit *et al.* 2022), in Supplementary Table S6, and the results verify the superiority of PHD loss.

6.2.2 Global–local strategy

To investigate the effectiveness of the global–local strategy, we conducted experiments using three architectures: S_G, S_L, and S_G+S_L. Results presented in Table 3 indicate that the combined approach of S_G+S_L outperforms either S_G or S_L alone. When using only pixel-wise loss, the *F*1 score of S_G+S_L is 59.57% compared to 51.57% for S_G and 58.23% for S_L. Similar improvements were observed on other evaluation criteria. These results suggest that the global–local strategy can be advantageous in segmentation, as it not only increases the local accuracy but also alleviates the global structure distance. Moreover, to provide further insight into the impact of PS-Net, we have illustrated the feature visualization in Supplementary Fig. S6 and conducted an attribution analysis for the global–local strategy in Supplementary Fig. S9. Our results indicate that the model trained with this strategy is able to capture more structural information, with a larger number of pixels contributing significantly to the prediction. These findings suggest that the global–local strategy enables the network to effectively utilize features of larger regions, thereby improving the segmentation performance. Due to space limitations, we have provided additional information in the Supplementary Materials.

6.2.3 Coarse-to-fine strategy

Additionally, the experiments were conducted to explore the effectiveness of the coarse-to-fine strategy by varying the parameters λ_1, λ_2, and k, as presented in Table 3 and Supplementary Table S5. The results indicate that gradually increasing the weights of the similarity loss and PHD loss resulted in improved segmentation performance. Notably, when the epoch is set to five, the introductions of the similarity loss and PHD loss yielded the best performance. These findings suggest that the coarse-to-fine strategy, with appropriate parameter tuning, can effectively improve the accuracy of segmentation tasks.

6.3 Experiments on natural image datasets

We further extend PS-Net to two natural image datasets: "Road" (Mnih 2013) and "CrackTree" (Zou *et al.* 2012). For evaluation, Pixel-wise accuracy, ARI, VOI, and Betti are chosen for comparison [reported by (Hu *et al.* 2021)]. The results in Table 4 also show that our work has SOTA performance. It is worth mentioning that PS-Net obtained a much better VOI score (**0.5117**) on the Road dataset.

7 Conclusions

In this study, we propose a novel criterion PHD and a PHD-based network for the task of cell membrane segmentation in EM images. The motivation for this approach arose from the discrepancy between commonly used metrics and human evaluations of segmentation results. To gain insight into the way humans analyze differences between segmentations, we conducted eye movement tracking experiments. These experiments revealed that humans utilize "global-local" and "coarse-to-fine" strategies in this process. Based on these observations, we incorporated these strategies into our model

through the use of separate global and local networks and the inclusion of PHD-based losses after initializing training with pixel-wise loss. Our proposed method was evaluated on several public EM and natural image datasets with consistently high performance.

Acknowledgements

We acknowledge High-Performance Computing Platform of Peking University for computational resources.

Supplementary data

Supplementary data are available at *Bioinformatics* online.

Conflict of interest

None declared.

Funding

This work was partially supported by the Natural Science Foundation of China [under contracts 62088102].

Data availability

The data underlying this article is available in https://github.com/EmmaSRH/PS-Net.

References

Arganda-Carreras I, Turaga SC, Berger DR *et al.* Crowdsourcing the creation of image segmentation algorithms for connectomics. *Front Neuroanat* 2015;**9**:142.

Beaucousin V, Simon G, Cassotti M *et al.* Global interference during early visual processing: ERP evidence from a rapid global/local selective task. *Front Psychol* 2013;**4**:539.

Shen W, Wang B, Jiang Y, *et al.* Multi-stage multi-recursive-input fully convolutional networks for neuronal boundary detection. In: *ICCV*. Venice, Italy, 22-29 October, IEEE/CVF 2017, 2391–400.

Shi R, Wang W, Li Z *et al.* U-RISC: an annotated ultra-high-resolution electron microscopy dataset challenging the existing deep learning algorithms. *Front Comput Neurosci* 2022;**16**:842760.

Shit S, Paetzold JC, Sekuboyina A, *et al.* clDice-a novel topology-preserving loss function for tubular structure segmentation. In: *CVPR*. Virtual, 19-25 June, IEEE/CVF 2021, 16560–9.

Spillmann L. From elements to perception: local and global processing in visual neurons. *Perception* 1999;**28**:1461–92.

Weng G, Dong B, Lei Y *et al.* A level set method based on additive bias correction for image segmentation. *Expert Syst Appl* 2021;**185**:115633.

Yan J, Chen H, Wang K, *et al.* Hierarchical attention guided framework for multi-resolution collaborative whole slide image segmentation. In: *MICCAI. Strasbourg, France, 27 September – 1 October, 2021, Proceedings, Part VIII 24*. Springer, 2021, 153–63.

Yeghiazaryan V, Voiculescu ID. Family of boundary overlap metrics for