International Journal of Computer Science and Information Technology
ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 8, Number 3, Year 2026
DOI: https://doi.org/10.62051/ijcsit.v8n3.05
Journal homepage: https://wepub.org/index.php/IJCSIT/index

DilNet: A Dilation Privileged Information for Accurate Langerhans Cell Segmentation

Ankai Dong, Pan Su

School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China

ABSTRACT

Accurate segmentation of Langerhans cell morphological characteristics is of great significance for the diagnosis of corneal diseases and the assessment of activation degree. However, the precise segmentation of corneal Langerhans cells remains unexplored. Manual dense annotation of Langerhans cells is a time-consuming and labor-intensive task. In order to achieve automated and accurate segmentation of Langerhans cells, this paper proposes a prior denoising framework, through boundary-aware refinement module which consists of a random mask dilation strategy to force the model to locate the target by understanding the target features rather than relying on the environment and a shrinking values generation strategy to gradually capture the target contour and improve the robustness of the refinement model. Experiments on one dataset and twelve types of neural networks have proven that our method significantly improves the accuracy of Langerhans cell segmentation and has strong versatility.

KEYWORDS

Medical Segmentation; Supervised learning; Privileged Information; Langerhans cell Dilation

1. INTRODUCTION

Langerhans cells, a type of dendritic cells (DCs), play a vital role in the human immune system by capturing and presenting antigens. In the field of ophthalmology, Langerhans cells present in the cornea serve as a crucial defense mechanism, safeguarding the cornea against infections and damage. Currently, clinical analysis focuses on examining the number, distribution, and condition of Langerhans cells, providing physicians with valuable insights for diagnosis and treatment planning. Therefore, leveraging deep learning networks for automated Langerhans cell extraction holds immense potential for enhancing quantitative analysis and improving the diagnostic accuracy of corneal diseases.

However, existing automated segmentation methods for Langerhans cells still have limitations, which hinders their clinical application. Most of the previous studies on automatic segmentation of Langerhans cells have mainly relied on traditional computer vision methods. For example, Braiki et al. [1] proposed an automatic segmentation method based on automatic thresholding and mathematical morphology. In their subsequent study [2], a combination of K-means clustering and mathematical morphology was used, followed by segmentation using the region-based Chan-Vese active contour model, which improved the accuracy of Langerhans cell segmentation. However, such methods tend to misclassify artifacts and noise as Langerhans cells when encountering such interference that is highly similar to the cellular saliency, and thus fail to accurately extract Langerhans cells. Furthermore, to mitigate the occlusion effect and deal with halo regions in images,

Suberi et al. [3] proposed a hybrid of low and high sigma in Gaussian kernel filtering with Local Adaptive Threshold (H-GLAT) through logical operator AND.

Recently, deep learning methods have been widely used in medical image tasks due to their high accuracy, robustness and generalization. However, deep learning methods have rarely been applied to dendritic cell segmentation. Existing work [4] directly adopted Mask R-CNN architectures to segment dendritic cells for clinical analysis. The above methods cannot learn the deep semantic information of DCs and are hard to distinguish noise which is similar to DCs. In addition, the above methods are easily affected by imaging quality, leading to inaccurate segmentation of DCs contours.

To address the above problems, we propose a novel framework, which consists of a privileged information generation module, a boundary-aware refinement module, and a denoising refinement module. The privileged information generation module is used to generate dilation masks containing a small amount of noise. Then dilation masks are fed into boundary-aware refinement module which consists of a random mask dilation strategy to force the model to learn the target features and a shrinking values generation strategy to gradually capture the target contour and improve the robustness of the refinement model.

(1) To the best of our knowledge, this paper is the first work to accurately segment Langerhans cells, providing a technical basis for further analysis of Langerhans activation levels and corresponding functions in future.

(2) Our paper proposes a prior denoising framework (a novel prior denoising framework based on privileged information and boundary-aware refinement) to obtain the local target features with less noise response first. Then boundary-aware refinement module is proposed to force the model to locate the target by understanding the target features rather than relying on the environment and a shrinking values generation strategy to gradually capture the target contour and improve the robustness of the refinement model.

2. METHODOLOGY

The proposed DilNet, based on Unet, comprises the Privileged Information Generation Module, Boundary-Aware Refinement Module, and Refinement Module. The network structure is shown in Fig. 1. The input data traverse through the encoder and then sequentially through the three modules to yield the final segmentation map. The Privileged Information Generation Module and Boundary-Aware Refinement Module are devised based on a random mask generation strategy and a contraction value generation strategy to enhance feature representations and bolster model robustness. In DilNet, we employ a combined loss function comprising cross-entropy loss and dice coefficient loss as the segmentation loss to train the framework.

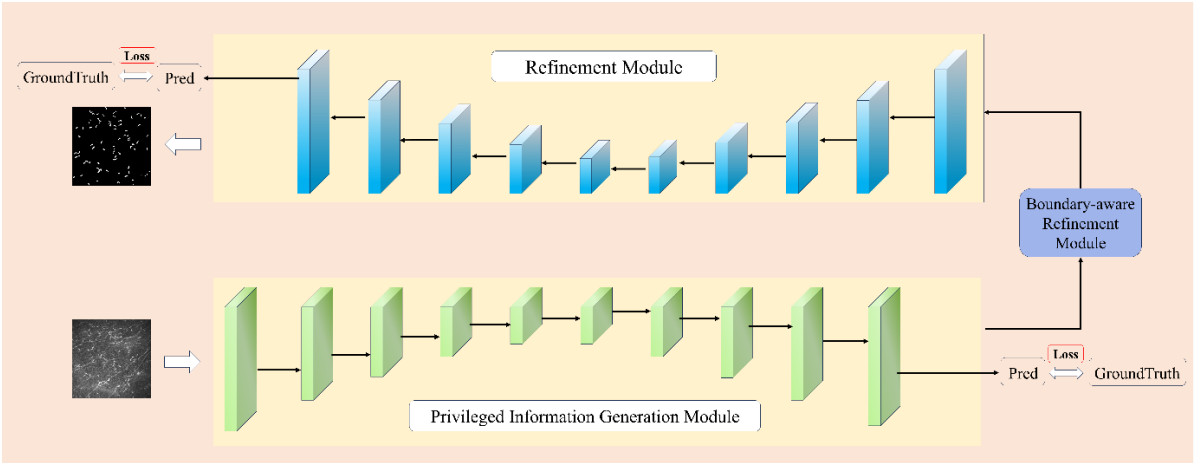

Figure 1. The architecture of framework

2.1. Formulation.

Given N training samples $X = \{(x_i, y_i)\}_{i=1}^{N}$, where x_i denotes the input image and y_i denotes the corresponding label. Let Frepresent the privileged information generation model. The segmentation map generated by the model can be expressed as

$$M_i = F(x_i) \tag{1}$$

Where M_i denotes the segmentation map of sample x_i generated by model F. During training, two strategies are applied to the generated segmentation map $M_i$to construct privileged information for training the refined model. Specifically, the random mask generation strategy is applied at the T-th epoch, while the shrinking values generation strategy is applied at the $(T+1)$-th epoch.

2.2. Boundary-aware Refinement Module.

To obtain better feature representations and improve model robustness, two strategies are introduced: random mask generation strategy and shrinking values generation strategy.

Random Mask Generation Strategy: We argue that the model may rely on the surrounding context when locating the target rather than truly understanding the target features. Based on this observation, a random mask dilation strategy is proposed to continuously transform the background region, forcing the model to focus on learning the intrinsic characteristics of the target.

In order to ensure that the dilation mask contains as many target regions as possible and achieve effective mask transformation, the dilation mask is generated through an ensemble learning mechanism. Specifically, a weight pool is defined as

$$S = \{W_1, W_2, \dots, W_K\} \tag{2}$$

Where K denotes the number of weights randomly selected from the privileged information generator.

Let $D_p(\cdot)$denote the operation that dilates the mask by $p$pixels. The process can be formulated as follows:

$$M_i^* = \sum_{k=1}^{K} F_{W_k}(x_i) \tag{3}$$

$$x_i^* = D_p(M_i^*) \odot x_i \tag{4}$$

Where x_i^* represents the transformed image and $\odot$ denotes element-wise multiplication.

Shrinking Values Generation Strategy: It is further assumed that focusing on the regions near the target boundary can effectively improve model accuracy and robustness, since these regions often contain fuzzy boundaries and similarity noise.

We argue that focusing on the region near the target can improve model accuracy and robustness because these regions often contain fuzzy boundaries and similar noise patterns. To validate this assumption, we introduce a shrinking values generation strategy to progressively shrink the effective target region during training. This mechanism allows the model to gradually capture the target contour and improves the robustness of the refinement process.

The above process can be expressed as

$$x_i^* = S(M_i) \odot x_i \tag{5}$$

$$s_t = 1 - \frac{t}{T} \tag{6}$$

Where t denotes the current training epoch and T represents the total number of training epochs. The coefficient S_t decreases linearly during training, which gradually shrinks the effective region around the target and forces the model to focus more on boundary information.

3. EXPERIMENTS

3.1. Dataset and Implementation Details

The proposed framework was evaluated on a dataset consisting of 560 corneal confocal microscopy images acquired using Heidelberg Retina Tomography equipment with a Rostock Cornea Module (HRT-III) microscope. The images were collected at a depth of 40–60 μm below the anterior surface of the central cornea. Each image has a resolution of 384 × 384 pixels, corresponding to a physical area of 400 × 400 μm. The annotations were produced by experienced ophthalmologists using the open-source software ITK-SNAP.

All models were implemented using the PyTorch framework with Python 3.8.0. The experiments were conducted on a workstation equipped with an NVIDIA GeForce RTX 3090 GPU. The dataset was divided into training, validation, and testing subsets with a ratio of 6:2:2. Model optimization was performed using the Adam optimizer with its default parameters. The batch size was set to 4, and the learning rate followed the configuration provided in the released code of the corresponding networks. Each model was trained for 300 epochs, and early stopping was adopted to prevent overfitting.

3.2. Comparison with State-of-the-Art Methods

To evaluate the effectiveness of the proposed framework, it was compared with several representative medical image segmentation methods, including ResUNet [5], CSNet [6], AttUNet [7], CENet [8], CPFNet [9], UNet++ [10], MedNeXt [11], DeepLabV3+ [12], nnUNet [13], TransUNet [14], SegFormer [15], UTNet [16], SwinUNet [17], ResUTNet [18], ENet [19], and UNet [20]. The evaluation was conducted using several widely adopted segmentation metrics, including mean Intersection over Union (mIoU), Dice coefficient (Dice), Modified Hausdorff Distance (MHD) [21], sensitivity (SE), and area under the ROC curve (AUC). Higher values of Dice, mIoU, SE, and AUC indicate better segmentation performance, whereas a lower MHD value reflects more accurate boundary prediction.

3.3. Quantitative Results

The quantitative comparison between the proposed method and the competing models is presented in Table 1. The proposed framework achieves the best overall performance among the compared methods. It produces the highest Dice and mIoU scores while obtaining the lowest MHD value, which indicates improved segmentation accuracy and boundary consistency.

Table 1. Comparison of quantitative results between Proposed and SOTA methods

Model	SE (↑)	Dice (↑)	AUC (↑)	mIoU (↑)	MHD (↓)
ResUNet	0.872	0.859	0.936	0.763	3.020
AttUNet	0.872	0.878	0.936	0.794	2.21
CSNet	0.857	0.855	0.928	0.756	3.230
CENet	0.828	0.806	0.913	0.686	3.660
CPFNet	0.837	0.814	0.918	0.697	4.250
UNet++	0.859	0.877	0.929	0.794	2.470
MedNeXtT	0.842	0.801	0.920	0.679	4.570
MedNeXtB	0.881	0.852	0.940	0.751	3.010
UNet	0.851	0.867	0.925	0.774	2.621
nnUNet	0.867	0.875	0.933	0.782	2.870
TransUNet	0.835	0.785	0.917	0.658	6.060
UTNet	0.890	0.873	0.944	0.783	2.530
SwinUNet	0.738	0.739	0.869	0.605	2.700
ResUTNet	0.860	0.822	0.929	0.709	3.860
Proposed	0.904	0.905	0.952	0.831	1.509

Compared with nnUNet, which shows strong performance among the baseline methods, the proposed method provides improvements in Dice score and sensitivity while also reducing the MHD value. These improvements suggest that the proposed framework enhances both the accuracy of cell region prediction and the precision of boundary localization.

The results also indicate that convolutional neural network based segmentation models tend to perform more effectively than transformer based approaches for the Langerhans cell segmentation task. The proposed method can also be integrated with different backbone architectures, and consistent improvements can be observed after applying the proposed strategies.

3.4. Qualitative Results

The qualitative segmentation results are illustrated in Fig. 2. In the visualization results, white regions correspond to correctly segmented areas, red regions indicate over-segmentation, and green regions represent under-segmentation.

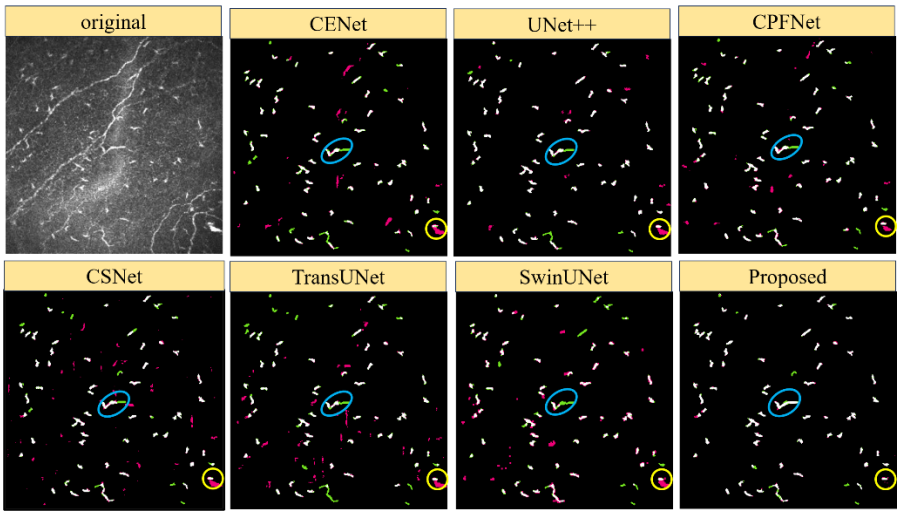

Figure 2. Performance comparisons of different methods. Under-segmentation is shown in red, and over-segmentation is shown in green

Compared with the baseline models, the proposed framework demonstrates a stronger ability to suppress noise regions. In several cases, background noise that is incorrectly segmented by other models is successfully ignored by the proposed method. The segmentation boundaries of Langerhans cells are also captured more accurately, resulting in clearer cell morphology. In addition, the proposed method shows improved capability in distinguishing activated Langerhans cells that exhibit visual similarity to nerve fiber branches.

3.5. Ablation Study

An ablation study was conducted to investigate the contribution of each component in the proposed framework. The evaluated configurations include Baseline, Baseline + NSF, Baseline + NSF + SVGS, Baseline + NSF + RMGS, and Baseline + All. In these configurations, NSF denotes the noise-shielding framework, RMGS represents the random mask generation strategy, and SVGS refers to the shrinking values generation strategy.

The experimental results summarized in Table 2 show that incorporating additional strategies improves segmentation performance compared with the baseline model. The introduction of the noise-shielding framework reduces the influence of background noise and leads to more stable predictions. When boundary refinement strategies are included, the segmentation boundaries become more accurate and the MHD value decreases accordingly. When all components are applied simultaneously, the model achieves the best overall performance while maintaining robust noise suppression.

Table 2. Ablation results for dendritic cells segmentation.

Model	SE (↑)	Dice (↑)	AUC (↑)	mIoU (↑)	MHD (↓)
Baseline	0.851	0.867	0.925	0.774	2.621
Baseline + NSF	0.936	0.879	0.967	0.791	2.341
Baseline + NSF + SVGS	0.882	0.887	0.941	0.801	2.003
Baseline + NSF + RMGS	0.930	0.891	0.964	0.816	1.674
Baseline + All	0.904	0.905	0.952	0.831	1.509

4. DISCUSSION

4.1. Effectiveness Analysis of Similarity Constraints and Relocation Constraints

To investigate whether the proposed method reduces the reliance on background information when locating the target, experiments were conducted using dilated images with different dilation values. During training, the models were trained with images containing a relatively small dilation range, while during testing the dilation range was extended. This setting evaluates the robustness of the model when the background distribution changes.

Fig. 3 presents the Dice and MHD results obtained by different models under varying dilation values during the testing stage. In the figure, proposed represents the model trained with the proposed strategies, whereas plain represents the model trained without these strategies.

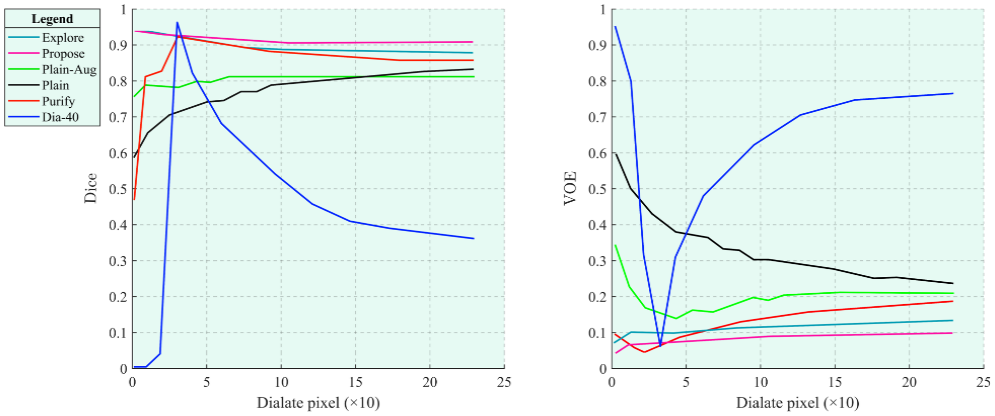

Figure 3. Experimental results of different models under different dilation images. The x-axis represents dilated values while y-axis represents metric. proposed and plain respectively represent the model with/without proposed strategy, and plain-aug represents the model with data augmentation

The results show that the proposed model maintains stable performance across a wide range of dilation values. In contrast, the plain model performs well only when the dilated images are close to the original image. This phenomenon indicates that the plain model is sensitive to background variations. The proposed method demonstrates stronger robustness when background conditions change. The observation suggests that the proposed strategy encourages the model to learn intrinsic target representations instead of relying on surrounding contextual information for localization.

The model trained with conventional data augmentation (plain-aug) performs better than the plain model under most dilation conditions. This improvement can be explained by the fact that data augmentation introduces variations in the background, which partially encourages the model to focus on the target region. However, the proposed strategy still achieves better results than the traditional augmentation approach. The results indicate that the proposed relocation and similarity constraints provide a more effective mechanism for improving feature learning.

4.2. Robustness Evaluation under Data Corruption

To further examine the effectiveness of the similarity constraint, several types of data corruption were introduced to evaluate model robustness. Five corruption levels were applied to the test images. The augmentation types include random rotation within the range of −45°to 45°, random horizontal and vertical flipping, RandomBrightnessContrast (RBC), Gaussian blur (GB), and Gaussian noise (GN).

Fig. 4 presents the performance of three models under different augmentation conditions. In the figure, Dia-40 denotes a model trained using dilated images with a dilation value of 40 without applying the similarity constraint. Plain refers to a model trained using original images without similarity constraints. Proposed denotes the model trained using dilated images with the similarity constraint.

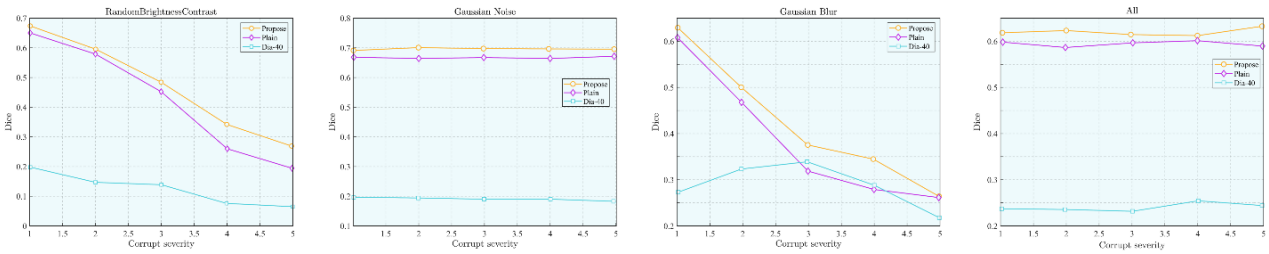

Figure 4. Verify the robustness of the three models under different severity levels of different data enhancements. Dia-40 represents a model trained without SC and dilated images. Plain represents a model trained without SC and non-dilated images. Proposed represents a model trained with SC and dilated images

The results show that the Dia-40 model exhibits unstable performance under various corruption settings. This behavior may be caused by the reduction of background information in dilated images, which makes the model more sensitive to noise. The performance of the plain model gradually declines as the corruption severity increases. In comparison, the proposed model maintains relatively stable performance across different corruption levels. These results indicate that the similarity constraint improves the robustness of the segmentation model. The findings also highlight the necessity of the relocation constraint when dilated images are used during training.

4.3. Application in Universal Networks

To evaluate the general applicability of the proposed strategy, it was integrated into both CNN-based and Transformer-based segmentation networks. The experimental results are summarized in Table 3.

Table 3. Quantitative results of proposed strategies applied to CNN and Transformer models. Red indicates performance improvement compared with the original model.

Model	SE (↑)	Dice (↑)	AUC (↑)	mIoU (↑)	MHD (↓)
Proposed (ResUNet)	0.880 (+0.80)	0.882 (+2.30)	0.940 (+0.004)	0.796 (+3.30)	1.669 (+1.351)
Proposed (CENet)	0.852 (+2.40)	0.841 (+3.50)	0.925 (+0.012)	0.733 (+4.70)	1.793 (+1.867)
Proposed (CSNet)	0.874 (+1.70)	0.884 (+2.90)	0.937 (+0.009)	0.799 (+4.30)	1.724 (+1.506)
Proposed (CPFNet)	0.873 (+3.60)	0.861 (+4.70)	0.936 (+0.018)	0.762 (+6.50)	1.631 (+2.619)
Proposed (AttUNet)	0.904 (+3.20)	0.900 (+2.20)	0.952 (+0.016)	0.824 (+3.00)	1.397 (+0.813)
Proposed (UNet++)	0.887 (+2.80)	0.895 (+1.80)	0.943 (+0.014)	0.816 (+2.20)	1.277 (+1.193)
Proposed (DeepLabV3)	0.798 (+3.70)	0.775 (+1.50)	0.898 (+0.018)	0.640 (+1.80)	4.003 (+1.115)
Proposed (MedNeXtT)	0.884 (+4.20)	0.859 (+5.80)	0.941 (+0.021)	0.761 (+8.20)	2.002 (+2.568)
Proposed (MedNeXtB)	0.883 (+0.20)	0.876 (+2.40)	0.943 (+0.003)	0.787 (+3.60)	1.664 (+1.346)
Proposed (TransUNet)	0.919 (+8.40)	0.838 (+5.30)	0.959 (+0.042)	0.658 (+7.00)	6.060 (+3.167)
Proposed (SegFormer)	0.830 (+14.7)	0.643 (+20.4)	0.912 (+0.074)	0.477 (+19.2)	2.958 (+1.882)
Proposed (UTNet)	0.902 (+1.20)	0.889 (+1.60)	0.951 (+0.007)	0.806 (+2.30)	1.725 (+0.805)
Proposed (SwinUNet)	0.744 (+0.60)	0.755 (+1.60)	0.872 (+0.003)	0.623 (+1.80)	1.899 (+0.801)
Proposed (ResUTNet)	0.889 (+2.90)	0.866 (+4.40)	0.944 (+0.015)	0.770 (+6.10)	1.860 (+2.00)

The results show that performance improvements are observed across all evaluated models after applying the proposed strategy. The improvements are reflected in several evaluation metrics, including Dice, sensitivity, mIoU, and MHD. These consistent improvements demonstrate that the proposed strategy can enhance the segmentation performance of different network architectures.

The results suggest that the proposed framework can be integrated into a wide range of segmentation models without modifying their original network structures. This property indicates that the proposed method has strong generalization capability and practical applicability for Langerhans cell morphology segmentation.

5. CONCLUSION

This paper proposes a priori denoising framework consisting of a Privileged Information Generation Module and a Refinement Module. The proposed framework aims to first obtain local target features with reduced noise responses. Based on the generated privileged information, a boundary-aware refinement module is introduced to guide the model to locate the target by learning intrinsic target features rather than relying on surrounding contextual information.

In addition, a shrinking values generation strategy is designed to gradually capture the target contour and improve the robustness of the refinement model. Experimental results on the dataset and twelve types of neural networks demonstrate that the proposed method achieves consistent performance improvements and shows good generalization capability across different network architectures.

REFERENCES

[1] Braiki M., Benzinou A., Nasreddine K., et al. Segmentation of dendritic cells from microscopic images using mathematical morphology. 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP). Monastir, Tunisia, 2016, p. 282-287.

[2] Braiki M., Benzinou A., Nasreddine K., et al. Automatic human dendritic cells segmentation using k-means clustering and Chan-Vese active contour model. Computer Methods and Programs in Biomedicine. Vol. 195 (2020) No. 1, p. 105520.

[3] Suberi A.A.M., Zakaria W.N.W., Tomari R., et al. Optimization of overlapping dendritic cell segmentation in phase contrast microscopy images. IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES). Kuala Lumpur, Malaysia, 2016, p. 246-250.

[4] Setu M.A.K., Schmidt S., Musial G., et al. Segmentation and evaluation of corneal nerves and dendritic cells from in vivo confocal microscopy images using deep learning. Translational Vision Science & Technology. Vol. 11 (2022) No. 6, p. 24-24.

[5] Xiao X., Lian S., Luo Z., et al. Weighted Res-UNet for high-quality retina vessel segmentation. International Conference on Information Technology in Medicine and Education (ITME). Hangzhou, China, 2018, p. 327-331.

[6] Mou L., Zhao Y., Chen L., et al. CS-Net: Channel and spatial attention network for curvilinear structure segmentation. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). Shenzhen, China, 2019, p. 721-730.

[7] Oktay O., Schlemper J., Folgoc L., et al. Attention U-Net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999. 2018.

[8] Gu Z., Cheng J., Fu H., et al. CE-Net: Context encoder network for 2D medical image segmentation. IEEE Transactions on Medical Imaging. Vol. 38 (2019) No. 10, p. 2281-2292.

[9] Feng S., Zhao H., Shi F., et al. CPFNet: Context pyramid fusion network for medical image segmentation. IEEE Transactions on Medical Imaging. Vol. 39 (2020) No. 10, p. 3008-3018.

[10] Zhou Z., Siddiquee M.M.R., Tajbakhsh N., et al. UNet++: A nested U-Net architecture for medical image segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support (DLMIA). Granada, Spain, 2018, p. 3-11.

[11] Roy S., Koehler G., Ulrich C., et al. MedNeXt: Transformer-driven scaling of ConvNets for medical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Vancouver, Canada, 2023, p. 405-415.

[12] Chen L.C., Zhu Y., Papandreou G., et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. European Conference on Computer Vision (ECCV). Munich, Germany, 2018, p. 801-818.

[13] Isensee F., Jaeger P.F., Kohl S.A., et al. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods. Vol. 18 (2021) No. 2, p. 203-211.

[14] Chen J., Lu Y., Yu Q., et al. TransUNet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306. 2021.

[15] Xie E., Wang W., Yu Z., et al. SegFormer: Simple and efficient design for semantic segmentation with transformers. Advances in Neural Information Processing Systems. Vol. 34 (2021), p. 12077-12090.

[16] Gao Y., Zhou M., Metaxas D.N. UTNet: A hybrid transformer architecture for medical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Strasbourg, France, 2021, p. 61-71.

[17] Cao H., Wang Y., Chen J., et al. Swin-UNet: Unet-like pure transformer for medical image segmentation. European Conference on Computer Vision (ECCV). Tel Aviv, Israel, 2022, p. 205-218.

[18] Diakogiannis F.I., Waldner F., Caccetta P., et al. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS Journal of Photogrammetry and Remote Sensing. Vol. 162 (2020) No. 1, p. 94-114.

[19] Paszke A., Chaurasia A., Kim S., et al. ENet: A deep neural network architecture for real-time semantic segmentation. arXiv preprint arXiv:1606.02147. 2016.

[20] Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation. International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Munich, Germany, 2015, p. 234-241.

[21] Dubuisson M.P., Jain A.K. A modified Hausdorff distance for object matching. International Conference on Pattern Recognition (ICPR). Jerusalem, Israel, 1994, p. 566-568.