



Research article

Modified UNet++ with atrous spatial pyramid pooling for blood cell image segmentation

Kun Lan^{1,†}, Jianzhen Cheng^{2,*}, Jinyun Jiang¹, Xiaoliang Jiang¹ and Qile Zhang^{3,†,*}

¹ College of Mechanical Engineering, Quzhou University, Quzhou 324000, China

² Department of Rehabilitation, Quzhou Third Hospital, Quzhou 324000, China

³ Department of Rehabilitation, The Quzhou Affiliated Hospital of Wenzhou Medical University, Quzhou People's Hospital, Quzhou 324000, China

† The authors contributed equally to this work.

* **Correspondence:** Email: 906989116@qq.com, zjzhangqile@163.com.

Abstract: Blood cell image segmentation is an important part of the field of computer-aided diagnosis. However, due to the low contrast, large differences in cell morphology and the scarcity of labeled images, the segmentation performance of cells cannot meet the requirements of an actual diagnosis. To address the above limitations, we present a deep learning-based approach to study cell segmentation on pathological images. Specifically, the algorithm selects UNet++ as the backbone network to extract multi-scale features. Then, the skip connection is redesigned to improve the degradation problem and reduce the computational complexity. In addition, the atrous spatial pyramid pooling (ASSP) is introduced to obtain cell image information features from each layer through different receptive domains. Finally, the multi-sided output fusion (MSOF) strategy is utilized to fuse the features of different semantic levels, so as to improve the accuracy of target segmentation. Experimental results on blood cell images for segmentation and classification (BCISC) dataset show that the proposed method has significant improvement in Matthew's correlation coefficient (Mcc), Dice and Jaccard values, which are better than the classical semantic segmentation network.

Keywords: cell; UNet++; atrous spatial pyramid pooling; image segmentation; U-Net

1. Introduction

With the combination of clinical medicine and information science, medical image processing technology has gradually developed and been applied in various medical laboratories, which can reduce the workload of doctors and improve detection accuracy. In traditional medical examination, professional doctors mainly rely on the eyes to observe the diseased part or tissue of the patient, and judge the disease condition according to their own experience. This method is easily affected by the professional knowledge of doctors, and different doctors may draw different inspection conclusions for the same patient, so it lacks objectivity. When examining specimens containing cells, such as pathological sections, blood smears, etc., it will be a time-consuming task to examine or count them due to a large number of samples. It can provide an objective and reliable reference for doctors to judge diseases by automatically segmenting and discriminating the lesion location images through medical imaging technology. Therefore, it is of great practical significance and social value to carry out automatic blood smear cell image segmentation based on deep learning theory and has broad application prospects in the medical field.

At present, the convolutional neural network (CNN) [1–5] has been widely used in the field of medical image segmentation. The method uses a large amount of medical image data to train the network so that the desired features in the images can be extracted. The most popular deep learning network for nuclear segmentation is the U-Net model [6], which combines shallow, low-level, fine-grained features from the encoder, and deep, semantic, coarse-grained features from the decoder via long skip connections to effectively improve the accuracy of nuclear detection. With the deepening of the research, various extended frameworks have emerged [7–12]. Considering the complex background of white blood cell images and the appearance changes of tissues, Lu et al. [13] proposed a deep learning segmentation framework based on UNet++ [14] and ResNet [15]. Firstly, a contextual feature sensing module with residual function is designed to extract multi-scale information. Then, dense convolution blocks are introduced to obtain more features on multi-scale channels. Using U-Net as the backbone architecture, Chan et al. [16] proposed a deep architecture for image segmentation with better performance, showing strong competitiveness. Jumutc et al. [17] proposed an enhanced version of the U-Net approach, which introduces the concept of a single receptive field path to obtain better coarse, fine-grained information. To overcome the interference of light conditions and other external factors for the cell nucleus, Thi Le et al. [18] adopted fuzzy pooling operation to maintain the salient features of the image, thus overcoming the problem of noise generation. By choosing image enhancement algorithms such as Fourier transform and mean-shift clustering, Makem et al. [19] proposed a deep-based network framework for white blood cells, and the experimental results proved the effectiveness and robustness of this method.

Although various algorithms have been proposed, blood cell segmentation in pathological images still faces great challenges for several reasons: 1) The complex structure, diverse shape, irregular boundary and other factors cause high variability of blood cell image. 2) Accurate segmentation of blood cells from microscope images requires expert knowledge and can be labor-intensive, resulting in limited cell image databases for specific diseases. 3) The sample data of abnormal cells are relatively small, resulting in the imbalance of distribution among classes, which affects the segmentation effect of the deep learning model on rare data. Figure 1 shows different types of cell images from the BCISC dataset. It can be seen that the shapes, sizes and boundaries of cells in different images are different, which may lead to difficulties in nucleus segmentation.

Therefore, the existing methods can only deal with a specific image, and there is no general method that can automatically and effectively segment the cell images of all modes.

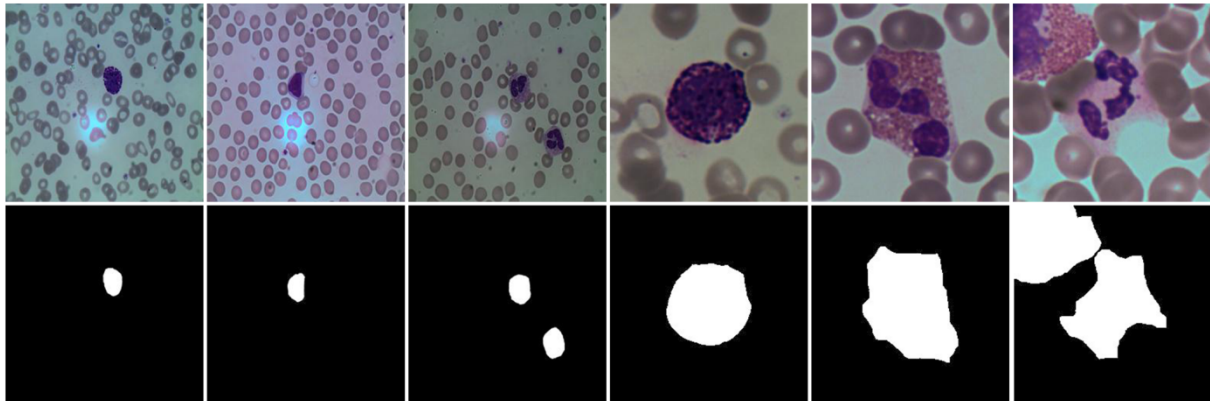


Figure 1. Different types of original blood cell images. The first row: original images. The second row: their respective masks.

Inspired by the UNet++ model, we apply the end-to-end deep learning method to blood cell image segmentation. By adding the ASSP module to UNet++ framework, the cell segmentation accuracy is improved without significantly increasing the number of parameters. Then the model structure of the jump connection is redesigned to ensure a stronger learning ability and feature extraction effect. On this basis, MSOF strategy is used to fuse the feature information from different semantic levels to further improve the accuracy of detection and recognition. Finally, the Dice loss function is utilized for training the network.

The remainder of this paper is divided as: Section 2 reviews related work and Section 3 gives the proposed method. The results and discussion of our algorithm on the BCISC dataset are illustrated in Section 4. Finally, the conclusion is discussed in Section 5.

2. Related works

2.1. U-Net structure

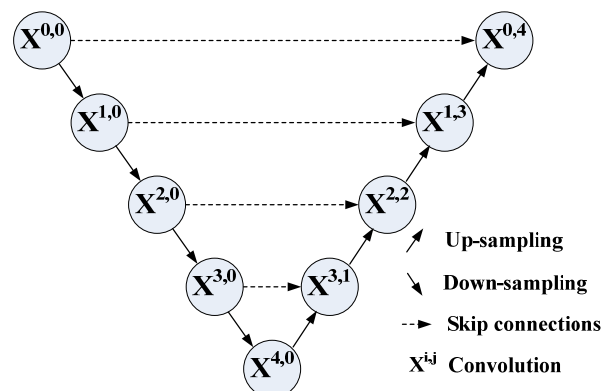


Figure 2. The network structure of U-Net.

U-Net [6] network is mainly designed to solve the problems of a small amount of data, unclear boundaries and large gray ranges in medical image analysis. Its structure is symmetric and similar to a U-shape, as shown in Figure 2. In the left part, 5 convolutional layers and 4 pooled down-sampling layers are used to decompose the image into features of different levels, which can capture the pixel information of the context. The right side is basically symmetric with the left side, and 5 convolution layers and 4 up-sampling layers are used to restore the input size of the image, which has the function of extracting the shallow, low-level and fine-grained features of the image. Another characteristic of U-Net is a skip connection, which completes the channel connection between the deep, semantic and coarse-grained feature maps from the decoder sub-network and the shallow, low-level and fine-grained feature maps from the encoder sub-network, so as to reduce the information loss in the process of feature extraction and achieve the purpose of accurate positioning.

The simple structure of U-Net is just suitable for the task of medical image segmentation, and it has an excellent performance in cell image segmentation. However, due to the fixed structure, it has two limitations: 1) The deeper the network structure, the larger the computational load, but the results could not be necessarily better. 2) Different data sets determine the different optimal depths of the network.

2.2. UNet++ structure

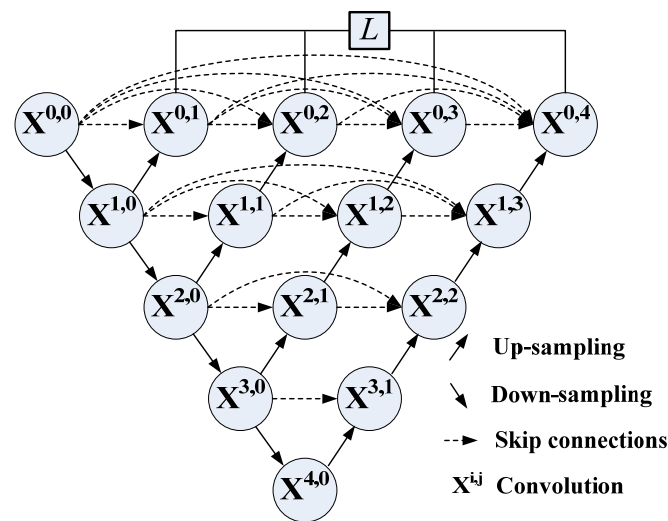


Figure 3. The network structure of UNet++.

Zhou et al. proposed a new neural network structure for semantic and instance segmentation, named UNet++ [14]. On the basis of maintaining the U-Net network structure, UNet++ improves the feature connection method between the encoder and the decoder, and its structure is shown in Figure 3. Take the node $X^{1,2}$ as an example, it receives all previous convolution units of the same level ($X^{1,0}$ and $X^{1,1}$) and the corresponding up-sampled output of $X^{2,1}$, and then gets it through a convolution and nonlinear correction unit. In this way, the semantic level of the encoder feature map is closer to the corresponding decoder part, which is more suitable for optimizer optimization. Assuming $x^{i,j}$ is the output of node $X^{i,j}$, the calculation of $x^{i,j}$ can be computed:

$$x^{i,j} = \begin{cases} H(x^{i-1,j}), j = 0 \\ H([\![x^{i,k}]_{k=0}^{j-1}\!] , u(x^{i+1,j-1})), j > 0 \end{cases} \quad (1)$$

where i represents the i th down-sampling layer in the encoder; j represents the j th convolution layer in skip connection, $H(\cdot)$ is the convolution operation with activation function, $u(\cdot)$ and $[\cdot]$ indicate the up-sample and concatenation operations, respectively.

The most obvious difference between UNet++ and U-Net is the new skip connection, which integrates different levels of semantic features of network output in network decoding. This method is very flexible, which can well remove unnecessary features and speed up the training. At the same time, UNet++ also introduces more parameters, and the memory occupied is greatly increased, which slows down the training speed and the convergence speed of the loss function.

3. Proposed methods

The improvement of UNet++ network mainly focuses on two parts: the feature extraction module and the feature output layer, as shown in Figure 4. Firstly, in order to better match blood cell features at different levels, the ASSP module is introduced to carry out the convolution of different cavities during feature extraction and connect feature map channels at different levels. Secondly, to make full use of the features of each level, the skip connection is redesigned and the multi-task learning module is introduced to optimize the segmentation results. The detailed technical and implementation of our modified Unet++ are explained below.

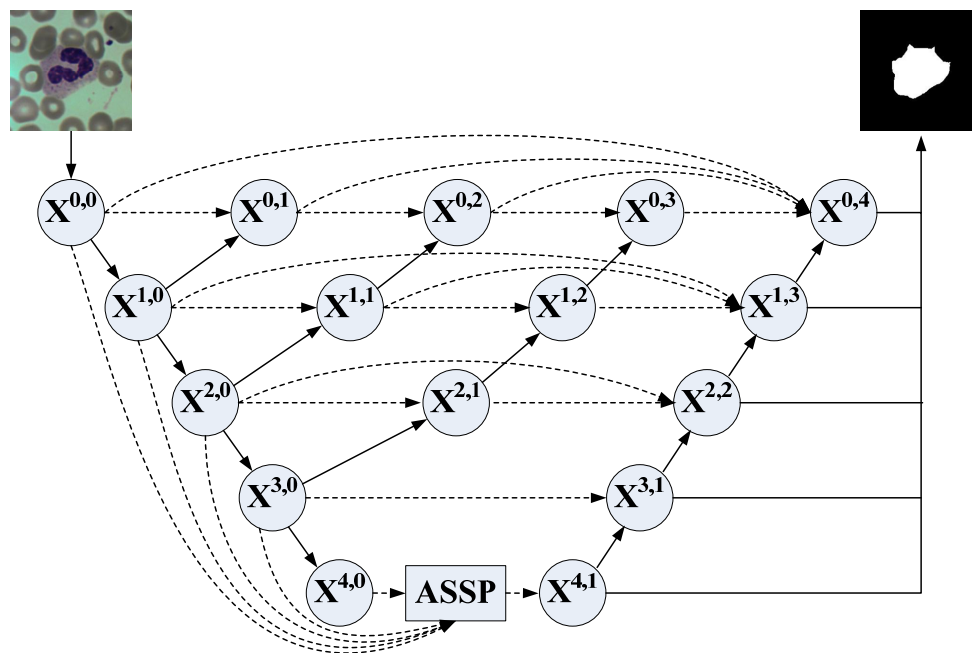


Figure 4. The overview of our proposed UNet++.

3.1. Redesigned skip connection

UNet++ and its variant [20–25] have a great advantage in obtaining multi-scale feature maps

because of their nested dense jump paths. However, since each node in the encoder and decoder is connected through the intermediate connection, this kind of dense connection leads to many model parameters and high computational complexity. Therefore, we present an improved UNet++ architecture, which only preserves the skip connections between the decoder and each node, the details are shown in Figure 4. The redesigned skip connection can ensure that the parameters are reduced without losing any information, and achieve better segmentation effect. The modified $x^{i,j}$ can be represented as:

$$x^{i,j} = \begin{cases} H(x^{i-1,j}), j = 0 \\ H([x^{i,j-1}, u(x^{i+1,j-1})]), j < 4 - i \\ H([x^{i,k}]_{k=0}^{j-1}, u(x^{i+1,j-1})), j = 4 - i \end{cases} \quad (2)$$

3.2. ASSP module

In deep learning, convolution is usually used for feature extraction. However, when the number of convolutions is too large, it is easy to lead to too many parameters and difficult weight optimization. By designing multiple parallel convolution kernels with different cavity rates, the size of the receptive field can be enlarged, and more spatial information can be retained while the number of parameters is unchanged. Figure 5 is the schematic diagram of dilated convolution, where (a) is a standard 3×3 convolution with the dilated factor of 1, and its receptive field size is 9; (b) is a 3×3 convolution with the dilated factor of 2, and its receptive field size is 49; (c) is a 3×3 convolution with the dilated factor of 3, and its receptive field size is 225. It can be seen that the applicability of the network to multi-scale objects can be increased by setting the inflation factor to obtain filters with different receptive fields.

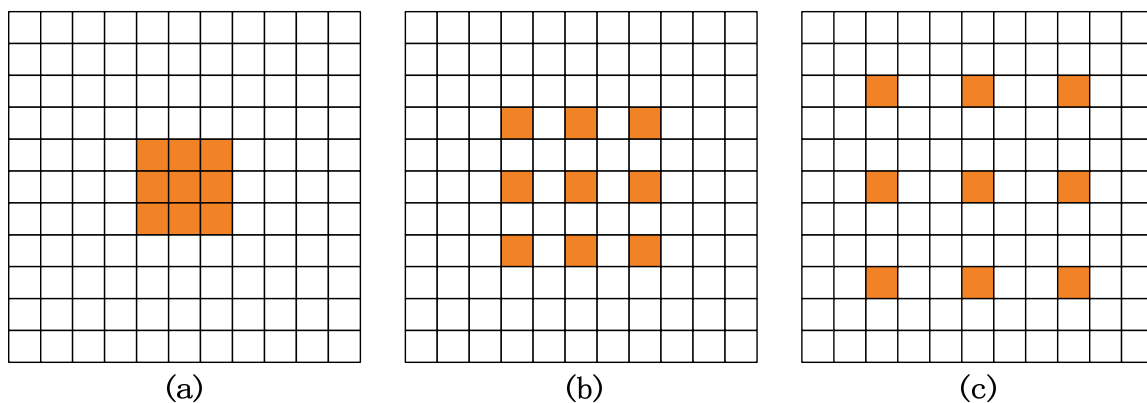


Figure 5. Dilated convolutions with different dilated rates.

The atrous spatial pyramid pooling module uses dilated convolution with different cavity rates for sampling, as shown in Figure 6. The ASPP module is mainly composed of eight convolution layers and one global average pooling (GAP) [26,27] layer in parallel. Specifically, a 3×3 convolution operation is performed on feature images $X^{0,0}$, $X^{1,0}$, $X^{2,0}$ and $X^{3,0}$ with stride of 16, 8, 4 and 2, respectively, while $X^{0,0}$ is subjected to one 1×1 convolution, pooling pyramid (three 3×3 dilated convolution) and

global average pooling layer (1×1 convolution). After that, the corresponding results are concatenated. Therefore, the network not only realizes the acquisition of feature maps of different scales and the integration of cross-channel information, but also expands the regional and contextual features of the received information.

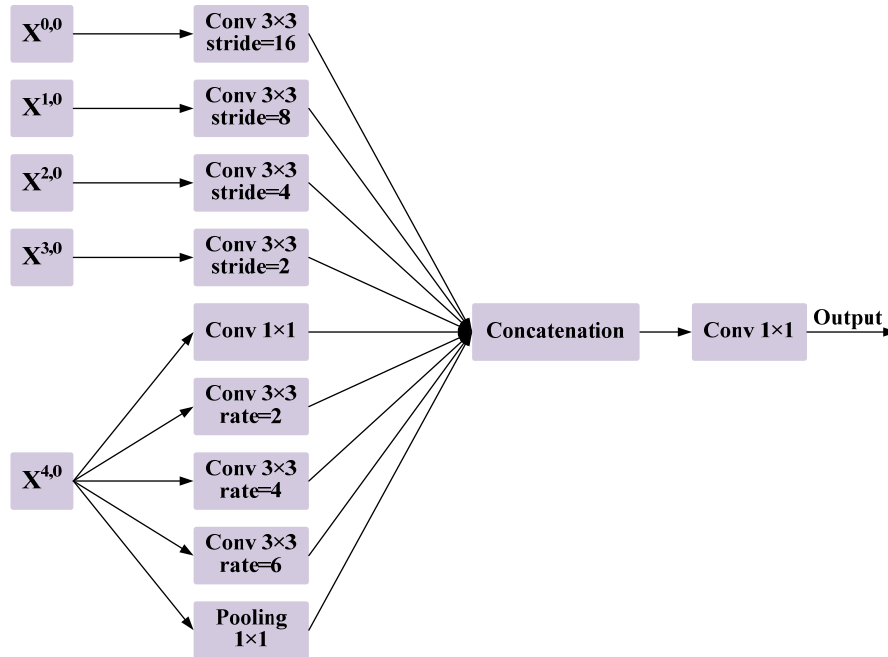


Figure 6. Backbone network structure with atrous spatial pyramid pooling module.

3.3. Multi-sided output fusion

For the UNet ++ backbone network structure, its depth increases from left to right, and the feature map corresponding to each output result becomes more and more refined from left to right. At the rightmost end of the network, the segmentation effect of the cell region is the best because of the increase of convolutional layers. When making a change prediction, the deepest network output is generally selected as the final result. However, there is useful information in relatively shallow levels of output, even if there are error areas in the deepest output predictions. Therefore, our model combines the shallow features with the deepest output features, which is beneficial to improve the prediction and segmentation accuracy of the overall model. As shown in Figure 4, the feature maps generated by the five convolution units $\{X^{0,4}, X^{1,3}, X^{2,2}, X^{3,1}, X^{4,1}\}$ are calculated by a 1×1 convolution and sigmoid function to get the final segmentation result.

3.4. Loss function

The Dice loss [28–30] is a region-based function, which can effectively solve the problem of imbalance between positive and negative samples because it only calculates a part of the region intersecting with the label without considering the background region. The loss function based on Dice coefficient can be expressed as:

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N y_i^2 \hat{y}_i^2}{\sum_{i=1}^N y_i^2 + \sum_{i=1}^N \hat{y}_i^2} \quad (3)$$

where L_{Dice} represents Dice loss, N denotes the total pixel, y_i represents the probability that the network predicts that voxel i belongs to the foreground target, and \hat{y}_i is the corresponding value of voxel i .

4. Experiments and results

In order to verify the high performance of the proposed method and its superiority for blood cell image segmentation, BCISC is used as the data set to test and discuss the experimental results. Next, we employ Mcc, Dice and Jaccard as evaluation indicators to compare with the current classical learning methods.

4.1. Experimental settings

4.1.1. Dataset description

The BCISC dataset [31] was provided by the Third People's Hospital of Fujian Province, which contains 400 training sets and 100 test sets. These images were from neutrophils, eosinophils, basophils, monocytes, and lymphocytes, to ensure the diversity of nuclear morphology. All images were taken by a physician for routine examination of the subject, and the ground truth was marked by a junior annotator and verified by an experienced radiologist. Taking into account the limitations of computer memory usage, we resized all images to 256×256 pixels.

4.1.2. Evaluation metrics

Considering the specificity of the nucleus segmentation task, the Mcc [32,33], Dice coefficient [34,35] and Jaccard index [36,37] are introduced as evaluation index. They are defined as:

$$Mcc = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FN)(TP+FP)(TN+FN)(TN+FP)}} \quad (4)$$

$$Dice = \frac{2TP}{2TP+FN+FP} \quad (5)$$

$$Jaccard = \frac{TP}{TP+FN+FP} \quad (6)$$

where TP, TN, FP, and FN denote numbers of true positives, true negatives, false positives, and false negatives on the pixels set, respectively.

4.1.3. Implementation details

All algorithms use the open-source deep learning framework Keras to write the network structure and realize the training of the network. The hardware and environment of the

experiment are Windows10 workstation configured with an Intel Xeon Gold 6248R CPU@3.00 GHz processor, 128 GB of 3200MHz DDR4 ECC RDIMM, NVIDIA Quadro RTX A6000 GPU memory of 48 GB. During model training, the original images and labels were sent to the network, and the Adam optimizer was used to optimize the network. The initial learning rate was set to 0.001, the number of training iterations was set to 400, and the batch size was set to 16.

4.2. Quantitative and qualitative results

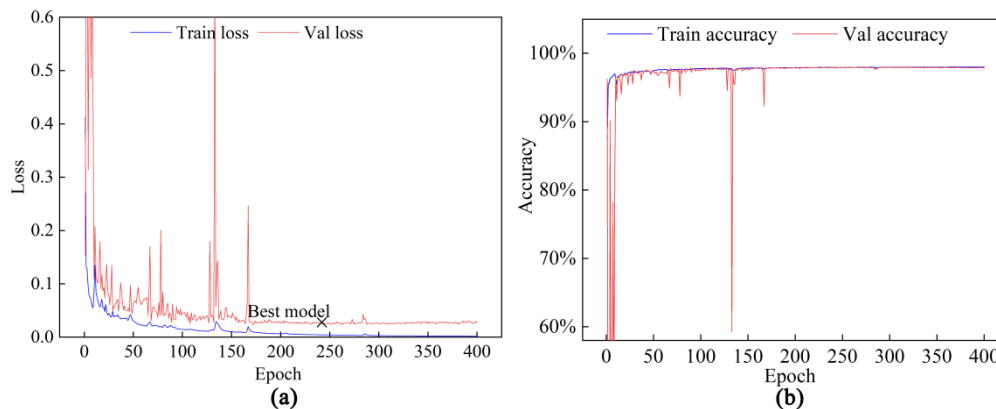


Figure 7. Visualization of loss and accuracy during training.

Figure 7 shows the convergence curve of the loss function and accuracy. It can be seen from Figure 7(a) that the number of trainings in this experiment is 400. During the first 20 training sessions, the gradient descends faster and the value of the loss function decreases faster. After that, the decline speed of the loss function becomes slower than before, which can avoid missing the optimal value due to the large gradient descent pace. When the loop reaches a certain level, the model optimization is no longer obvious. Figure 7(b) shows the changing trend of the accuracy index during the training process. As can be seen from the graph, the value of accuracy increases rapidly in the early stage of training, and then gradually becomes stable, with only a small range of fluctuation.

Table 1. Results of different algorithms on blood cell images.

Method	Mcc	Dice	Jaccard
DenseUnet [38]	0.9164	0.9189	0.8525
FCN [39]	0.8016	0.7987	0.6842
Segnet [40]	0.9221	0.9267	0.8656
U-Net+++ [41]	0.9278	0.9333	0.8757
UNet++ [14]	0.9370	0.9417	0.8902
U-Net [6]	0.9088	0.9140	0.8441
Our model	0.9418	0.9465	0.8989

To verify the performance superiority of our proposed approach on blood cell segmentation task, it is compared with DenseUnet [38], FCN [39], Segnet [40], U-Net+++ [41], UNet++ [14] and U-Net [6] methods. The main evaluation indicators included Mcc, Dice and Jaccard, and the results are listed in Table 1. It can be seen from the table that after multiple pooling, the feature

map size of the FCN method is very small. Although it contains more semantic information, there are serious false detections and missed detections. Although it contains a lot of semantic information, there is still serious false detection and missed detection phenomena, so all indicators are the lowest. The U-Net and Segnet models allow the decoder to learn the relevant feature information lost in the encoder pooling at each stage through the skip-connected architecture, so the performance of each evaluation index is better than FCN. The DenseUnet structure greatly expands the actual depth of the network and improves feature utilization. Therefore, it has great advantages over ordinary convolutional networks. UNet++ and UNet+++ have nested dense jump paths, making the semantic level of the encoder feature graph closer to the level of the corresponding encoder part, so it has been significantly improved on the basis of U-Net, and the phenomenon of false detection and missed detection is reduced. After the introduction of the ASSP module and multi-sided output fusion strategy, the Dice loss function is adopted to reduce the impact of sample imbalance. When compared with the U-Net, the Mcc, Dice and Jaccard values of our model are increased to 3.30%, 3.25% and 5.48%, respectively. The comprehensive evaluation index shows that our network has great advantages in cell localization and extraction.

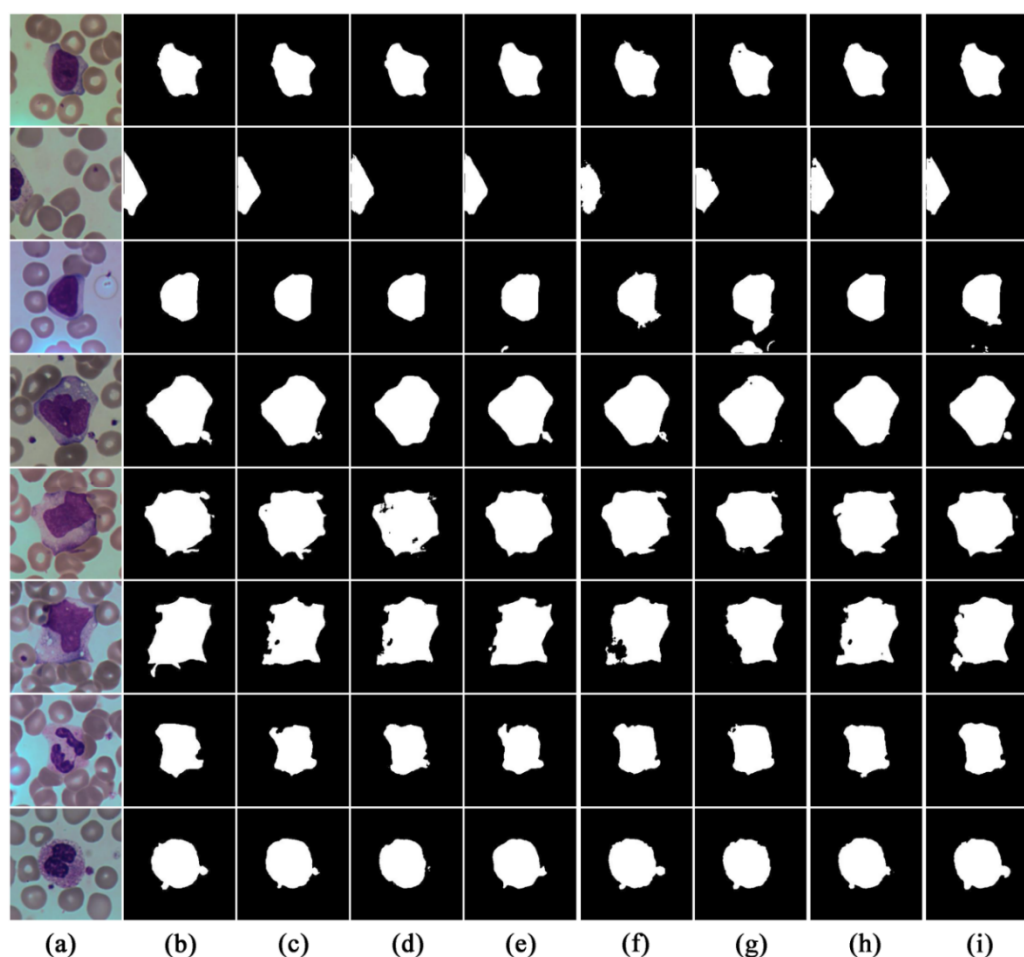


Figure 8. Result of blood cell images by different algorithms. From left to right: original images, their labels, DenseUnet, FCN, Segnet, U-Net+++, UNet++, U-Net and our method.

To further visually compare the differences in blood cell segmentation results by different network models, some results with obvious contrast were selected from the BCISC dataset for intuitive qualitative analysis and then compared with the segmentation results manually annotated by experts, as shown in Figure 8. According to the visualization results, the segmentation effect of FCN, Segnet and U-Net in the complex background is not very ideal, and the nucleus cannot be completely segmented. Through the multi-scale feature extraction techniques, U-Net+++ and UNet++ can improve the integrity of complex cell segmentation, but the edge details are obviously insufficient. However, the convolution kernel of the proposed model can extract information at different scales and suppress useless feature information, so it can better demonstrate the superiority of its segmentation performance.

5. Conclusions

In this study, based on the UNet++ architecture, we describe a new end-to-end method for blood cell segmentation on pathological images, which achieves good results in both localization and boundary segmentation. This is due to the introduction of the ASSP module, which can obtain blood cell image information features at each layer from different receiving fields. On the other hand, the multi-sided output fusion strategy can fuse feature information of different semantic levels to further improve the accuracy of detection and recognition. The quantitative and qualitative results on cell images indicate that, compared with other advanced deep learning-based methods, our model has great advantages in Mcc, Dice and Jaccard values. In the future, we will apply our technology to other cell images and medical images, and try to build an automated diagnostic system that can more accurately distinguish or predict benign and malignant lesions.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 62102227, 51805124, 62101206), Zhejiang Basic Public Welfare Research Project (Nos. LZY22E050001, LZY22D010001, LGG19E050013, LZY21E060001), Science and Technology Major Projects of Quzhou (2021K29, 2022K56).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. B. Dourthe, N. Shaikh, S. A. Pai, S. Fels, S. H. M. Brown, D. R. Wilson, et al., Automated segmentation of spinal muscles from upright open MRI using a multiscale pyramid 2D convolutional neural network, *Spring*, **47** (2022), 1179–1186. <https://doi.org/10.1097/BRS.0000000000004308>

2. K. Mariam, O. M. Afzal, W. Hussain, M. U. Javed, A. Kiyani, N. Rajpoot, et al., On smart gaze based annotation of histopathology images for training of deep convolutional neural networks, *IEEE J. Biomed. Health Inf.*, **26** (2022), 3025–3036. <https://doi.org/10.1109/JBHI.2022.3148944>
3. X. Y. Wei, Y. Y. Wang, L. Ge, B. Peng, Q. He, R. Wang, et al., Unsupervised convolutional neural network for motion estimation in ultrasound elastography, *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, **69** (2022), 2236–2247. <https://doi.org/10.1109/TUFFC.2022.3171676>
4. W. Ba, H. Wu, W. W. Chen, S. H. Wang, Z. Y. Zhang, X. J. Wei, et al., Convolutional neural network assistance significantly improves dermatologists' diagnosis of cutaneous tumours using clinical images, *Eur. J. Cancer*, **169** (2022), 156–165. <https://doi.org/10.1016/j.ejca.2022.04.015>
5. A. Iqbal, M. Sharif, M. A. Khan, W. Nisar, M. Alhaisoni, FF-UNet: a u-shaped deep convolutional neural network for multimodal biomedical image segmentation, *Cognit. Comput.*, **14** (2022), 1287–1302. <https://doi.org/10.1007/s12559-022-10038-y>
6. O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-assisted Intervention*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
7. Z. Li, H. Zhang, Z. Li, Z. Ren, Residual-attention UNet++: a nested residual-attention U-Net for medical image segmentation, *Appl. Sci.*, **12** (2022), 7149. <https://doi.org/10.3390/app12147149>
8. L. F. Yu, Z. Qin, Y. Ding, Z. G. Qin, MIA-UNet: multi-scale iterative aggregation U-Network for retinal vessel segmentation, *Comput. Model. Eng. Sci.*, **129** (2021), 805–828. <https://doi.org/10.32604/cmescs.2021.017332>
9. Y. J. He, J. S. Li, S. Shen, K. Liu, K. K. Wong, T. C. He, et al., Image-to-image translation of label-free molecular vibrational images for a histopathological review using the UNet plus/seg-cGAN model, *BioMed. Opt. Express*, **13** (2020), 1924–1938. <https://doi.org/10.1364/BOE.445319>
10. Y. Zhang, X. Liu, S. Wa, Y. Liu, J. Kang, C. Lv, GenU-Net++: an automatic intracranial brain tumors segmentation algorithm on 3D image series with high performance, *Symmetry*, **13** (2021), 2395. <https://doi.org/10.3390/sym13122395>
11. C. Wang, Z. Y. Zhao, Y. Yu, Fine retinal vessel segmentation by combining Nest U-net and patch-learning, *Soft Comput.*, **25** (2021), 5519–5532. <https://doi.org/10.1007/s00500-020-05552-w>
12. M. Lei, J. Li, M. Li, L. Zou, H. Yu, An improved UNet++ model for congestive heart failure diagnosis using short-term RR intervals, *Diagnostics*, **11** (2021), 534. <https://doi.org/10.3390/diagnostics11030534>
13. Y. Lu, X. J. Qin, H. Y. Fan, T. T. Lai, Z. Y. Li, WBC-Net: a white blood cell segmentation network based on UNet++ and ResNet, *Appl. Soft Comput.*, **101** (2021), 107006. <https://doi.org/10.1016/j.asoc.2020.107006>
14. Z. W. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, J. M. Liang, UNet++: a nested U-Net architecture for medical image segmentation, in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, (2018), 3–11. https://doi.org/10.1007/978-3-030-00889-5_1

15. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 770–778. <https://doi.org/10.1109/CVPR.2016.90>
16. S. X. Chan, C. Huang, C. Bai, W. L. Ding, S. Y. Chen, Res2-UNeXt: a novel deep learning framework for few-shot cell image segmentation, *Multimedia Tools Appl.*, **81** (2021), 13275–13288. <https://doi.org/10.1007/s11042-021-10536-5>
17. V. Jumutc, D. Bliznuks, A. Lihachev, Multi-Path U-Net architecture for cell and colony-forming unit image segmentation, *Sensors*, **22** (2022), 990. <https://doi.org/10.3390/s22030990>
18. P. Thi Le, T. Pham, Y. C. Hsu, J. C. Wang, Convolutional blur attention network for cell nuclei segmentation, *Sensors*, **22** (2022), 1586. <https://doi.org/10.3390/s22041586>
19. M. Makem, A. Tiedeu, G. Kom, Y. Nkandeu, A robust algorithm for white blood cell nuclei segmentation, *Multimedia Tools Appl.*, **81** (2022), 17849–17874. <https://doi.org/10.1007/s11042-022-12285-5>
20. B. Yang, M. X. Wu, W. Teizer, Modified UNet++ with attention gate for graphene identification by optical microscopy, *Carbon*, **195** (2022), 246–252. <https://doi.org/10.1016/j.carbon.2022.03.035>
21. S. Bhagat, M. Kokare, V. Haswani, P. Hambarde, R. Kamble, Eff-UNet++: a novel architecture for plant leaf segmentation and counting, *Ecol. Inf.*, **68** (2022), 101583. <https://doi.org/10.1016/j.ecoinf.2022.101583>
22. F. Hoorali, H. Khosravi, B. Moradib, Automatic Bacillus anthracis bacteria detection and segmentation in microscopic images using UNet++, *J. Microbiol. Methods*, **177** (2020), 106056. <https://doi.org/10.1016/j.mimet.2020.106056>
23. W. Zhao, Y. Zhao, L. Feng, J. Tang, Attention enhanced serial Unet++ network for removing unevenly distributed haze, *Electronics*, **10** (2021), 2868. <https://doi.org/10.3390/electronics10222868>
24. H. Zhao, H. Zhang, X. Zheng, A multiscale attention-guided UNet++ with edge constraint for building extraction from high spatial resolution imagery, *Appl. Sci.*, **12** (2022), 5960. <https://doi.org/10.3390/app12125960>
25. S. Safarov, T. K. Whangbo, A-DenseUNet: adaptive densely connected UNet for polyp segmentation in colonoscopy images with atrous convolution, *Sensors*, **21** (2021), 1441. <https://doi.org/10.3390/s21041441>
26. J. J. Li, Y. Han, M. Zhang, G. Li, B. H. Zhang, Multi-scale residual network model combined with global average pooling for action recognition, *Multimedia Tools Appl.*, **81** (2022), 1375–1393. <https://doi.org/10.1007/s11042-021-11435-5>
27. R. L. Kumar, J. Kakarla, B. V. Isunuri, M. Singh, Multi-class brain tumor classification using residual network and global average pooling, *Multimedia Tools Appl.*, **80** (2021), 13429–13438. <https://doi.org/10.1007/s11042-020-10335-4>
28. R. Arora, I. Saini, N. Sood, Multi-label segmentation and detection of COVID-19 abnormalities from chest radiographs using deep learning, *Optik*, **246** (2021), 167780. <https://doi.org/10.1016/j.ijleo.2021.167780>
29. X. M. Liu, S. C. Wang, Y. Zhang, D. Liu, W. Hu, Automatic fluid segmentation in retinal optical coherence tomography images using attention based deep learning, *Neurocomputing*, **452** (2021), 576–591. <https://doi.org/10.1016/j.neucom.2020.07.143>

30. T. D. T. Phan, S. H. Kim, H. J. Yang, G. S. Lee, Skin lesion segmentation by U-Net with adaptive skip connection and structural awareness, *Appl. Sci.*, **11** (2021), 4528. <https://doi.org/10.3390/app11104528>
31. *fpklipic*, Dataset, 2019. Available from: <https://github.com/fpklipic/BCISC>.
32. M. Jiang, F. Zhai, J. Kong, A novel deep learning model DDU-net using edge features to enhance brain tumor segmentation on MR images, *Artif. Intell. Med.*, **121** (2021), 102180. <https://doi.org/10.1016/j.artmed.2021.102180>
33. A. Oulefki, S. Agaian, T. Trongtirakul, A. K. Laouar, Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images, *Pattern Recognit.*, **114** (2021), 107747. <https://doi.org/10.1016/j.patcog.2020.107747>
34. Y. Y. Yang, C. Feng, R. F. Wang, Automatic segmentation model combining U-Net and level set method for medical images, *Expert Syst. Appl.*, **153** (2020), 113419. <https://doi.org/10.1016/j.eswa.2020.113419>
35. J. You, P. L. Yu, A. C. Tsang, E. L. Tsui, P. P. Woo, C. S. Lui, et al., 3D dissimilar-siamese-U-Net for hyperdense middle cerebral artery sign segmentation, *Comput. Med. Imaging Graphics*, **90** (2021), 101898. <https://doi.org/10.1016/j.compmedimag.2021.101898>
36. S. Mohajerani, P. Saeedi, Cloud and cloud shadow segmentation for remote sensing imagery via filtered Jaccard loss function and parametric augmentation, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **14** (2021), 4254–4266. <https://doi.org/10.1109/JSTARS.2021.3070786>
37. V. S. Bochkov, L. Y. Kataeva, wUUNet: Advanced fully convolutional neural network for multiclass fire segmentation, *Symmetry*, **13** (2021), 98. <https://doi.org/10.3390/sym13010098>
38. G. Huang, Z. Liu, V. Laurens, K. Q. Weinberger, Densely connected convolutional networks, in: *IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>
39. J. Long, E. Shelhamer, T Darrell, Fully convolutional networks for semantic segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2015), 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>
40. V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
41. H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, et al., UNet 3+: a full-scale connected UNet for medical image segmentation, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, (2020), 1055–1059. <https://doi.org/10.1109/ICASSP40776.2020.9053405>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)