



## WBC-Net: A white blood cell segmentation network based on UNet++ and ResNet

Yan Lu<sup>a,b</sup>, Xuejun Qin<sup>c</sup>, Haoyi Fan<sup>d</sup>, Taotao Lai<sup>a</sup>, Zuoyong Li<sup>a,\*</sup>

<sup>a</sup> Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, College of Computer and Control Engineering, Minjiang University, Fuzhou, 350121, China

<sup>b</sup> College of Mathematics and Computer Science, Fuzhou University, Fuzhou, 350108, China

<sup>c</sup> Department of Laboratory Medicine, The Affiliated People's Hospital of Fujian University of Traditional Chinese Medicine, Fuzhou, 350001, China

<sup>d</sup> School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, 150080, China



### ARTICLE INFO

#### Article history:

Received 30 June 2020

Received in revised form 5 November 2020

Accepted 7 December 2020

Available online xxxx

#### Keywords:

White blood cell

Image segmentation

Convolutional neural network

### ABSTRACT

The counting and identification of white blood cells (WBCs, i.e., leukocytes) in blood smear images play a crucial role in the diagnosis of certain diseases, including leukemia, infections, and COVID-19 (corona virus disease 2019). WBC image segmentation lays a firm foundation for automatic WBC counting and identification. However, automated WBC image segmentation is challenging due to factors such as background complexity and variations in appearance caused by histological staining conditions. To improve WBC image segmentation accuracy, we propose a deep learning network called WBC-Net, which is based on UNet++ and ResNet. Specifically, WBC-Net designs a context-aware feature encoder with residual blocks to extract multi-scale features, and introduces mixed skip pathways on dense convolutional blocks to obtain and fuse image features at different scales. Moreover, WBC-Net uses a decoder incorporating convolution and deconvolution to refine the WBC segmentation mask. Furthermore, WBC-Net defines a loss function based on cross-entropy and the Tversky index to train the network. Experiments on four image datasets show that the proposed WBC-Net achieves better WBC segmentation performance than several state-of-the-art methods.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

Manual counting and identification of white blood cells (WBCs) play a vital role in the initial diagnosis of human blood diseases. To do this, pathologists typically inspect blood smears visually through a microscope. However, this manual technique is time- and labor-intensive. Automatic WBC segmentation is the key to developing an automated WBC analysis technique. Difficulties in WBC segmentation are caused by dyeing impurities, cytoplasm with low image contrast, and variable appearances under different staining conditions. Existing WBC segmentation methods comprise unsupervised [1–6] and supervised methods [7–19].

Unsupervised methods achieve WBC segmentation based on characteristics such as color, shape, and brightness, and include image thresholding [1,2], image clustering [3,4], and edge detection [5,6]. Image thresholding-based methods [1,2] usually use image features to determine a threshold, which is used to divide image pixels into object pixels and background pixels. They usually achieve satisfactory segmentation of WBC nuclei

with high contrast, but fail to achieve satisfactory segmentation of WBC cytoplasm with low contrast. Image clustering-based methods [3,4] usually use image pixel features to group similar pixels, but they are unstable due to the difficulty in selecting suitable features. Edge detection-based methods [5,6] usually extract object boundaries based on their sharp characteristics. However, it is difficult to detect entire weak object boundaries with low contrast.

Supervised methods regard a segmentation problem as a classification problem. Traditional classifiers include the K-nearest neighbor (KNN) [7], and support vector machine (SVM) [8]. These methods usually first extract image features, and then use them to classify image pixels. KNN [7] usually classifies image pixels by measuring the similarity based on image pixel features. KNN is simple, but its classification accuracy with rare classes is usually low when the number of samples is unbalanced. SVM [8] achieves image pixel classification by finding the optimal segmentation hyperplane. However, SVM usually fails to obtain satisfactory results on nonlinear classification problems.

Traditional classifiers have difficulty in obtaining high classification accuracy on large-scale image datasets due to their weak image representation capability. During recent years, deep learning methods have attracted increasing attentions in the fields of

\* Corresponding author.

E-mail address: [fzulzytdq@126.com](mailto:fzulzytdq@126.com) (Z. Li).

computer vision and medical image analysis due to their strong image representation capability. For example, Mask R-CNN [9,10] evolved from Faster R-CNN [11] by adding a parallel branch (for object mask prediction) to the original branch (for bounding box recognition). Mask R-CNN achieves object detection and image segmentation concurrently, but is not good at segmenting small-scale image datasets.

Motivated by two famous deep learning networks, UNet++ [15] and ResNet [16], we propose a new WBC segmentation network called WBC-Net. WBC-Net first develops a context-aware feature encoder to extract multi-scale image features, then introduces mixed skip pathways to fuse them. Then, it reconstructs feature maps using a decoder with convolutional blocks and deconvolutional blocks. Finally, it defines a new loss function based on cross-entropy and the Tversky index for training the network.

The rest of this paper is organized as follows. Section 2 reviews related work, then Section 3 describes the proposed method's theory and implementation. Section 4 describes our experimental results and analysis, then Section 5 draws conclusions.

## 2. Related work

Due to the challenges of WBC segmentation, traditional image segmentation methods usually fail to obtain satisfactory segmentation results. Recently, convolutional neural network (CNN)-based methods have been widely applied in the field of computer vision for tasks such as image classification [20,21], object detection [22], image denoising [23], and semantic segmentation [13, 14]. Among them, a fully convolutional network (FCN) [13] can obtain good performance in image segmentation. FCN removes the fully-connected layers of a traditional CNN, and uses an up-sampling operation to achieve end-to-end (i.e., pixels-to-pixels) image segmentation. However, FCN's segmentation performance is limited by its simple structure. U-Net [14] is an updated FCN with an encoder-decoder structure. An encoder module achieves feature extraction, while a decoder module restores feature maps with the original image resolution. U-Net usually obtains good segmentation results on small-scale image datasets, and is suitable for medical image segmentation due to its encoder-decoder structure. U-Net uses skip connections to combine shallow feature maps from the encoder module with deep feature maps from the decoder module to recover fine-grained object details. However, U-Net may irrationality fuse semantically dissimilar feature maps. To resolve this issue, UNet++ [15] revises the skip connections of U-Net to reduce the semantic gap between feature maps generated by the encoder module and those generated by the decoder module. To overcome the performance bottleneck in deep learning caused by vanishing gradients, He et al. [16] proposed the famous Residual Network (ResNet), for which classification accuracy improves with network depth.

Inspired by the successful applications of UNet++ and ResNet in the field of computer vision, we propose a new deep learning method called WBC-Net for improving the segmentation accuracy of WBCs in blood smear images. The main contributions of this paper are as follows:

- (1) Motivated by UNet++ and ResNet, we propose a new deep learning network called WBC-Net that uses an improved encoder-decoder structure for segmentation.
- (2) The encoder module of WBC-Net better extracts and fuses multi-scale features by combining convolutional blocks and residual blocks, as well as mixed skip pathways.
- (3) Inspired by deep supervision, we designed a loss function based on cross-entropy and the Tversky index, and calculated the loss function value of each layer for full training of the prior features.
- (4) Experiments with four datasets comprising 873 images show that WBC-Net provides better segmentation performance than several state-of-the-art methods.

## 3. Proposed method

To use the strong image representation capability of deep learning for improving WBC segmentation accuracy, we propose a novel WBC segmentation network inspired by UNet++ and ResNet that we call WBC-Net. The architecture of WBC-Net includes three main stages: feature extraction, feature reconstruction, and feature fusion (Fig. 1). The feature extraction stage uses a context-aware feature encoder with convolutional blocks and residual blocks to extract multi-scale feature maps. The feature reconstruction stage uses a decoder with convolution and deconvolution to adjust the sizes of the feature maps to achieve end-to-end WBC segmentation. The feature fusion stage introduces mixed skip pathways to reduce the semantic gap between the feature maps generated by the encoder module and those generated by the decoder module. In short, WBC-Net uses the encoder module to extract image features at different scales, and uses the decoder module to reconstruct the feature maps. Moreover, we calculate the loss function value of each layer for full training of the prior features.

### 3.1. Context-aware feature network

WBC-Net uses residual blocks to better extract multi-scale features in the encoder module. Generally speaking, in deep learning, a deeper neural network usually generates better processing results. However, a deeper network brings more challenges in terms of optimizing network weights and being prone to vanishing gradients and network performance degradation. Inspired by the famous residual network ResNet [16], WBC-Net designs a direct connection channel between the input and output feature, and learns the residuals rather than the original features. In a traditional CNN, the relationship between the input  $x_i$  and output  $x_{i+1}$  is mapped as:

$$x_{i+1} = F(x_i, w_i), \quad (1)$$

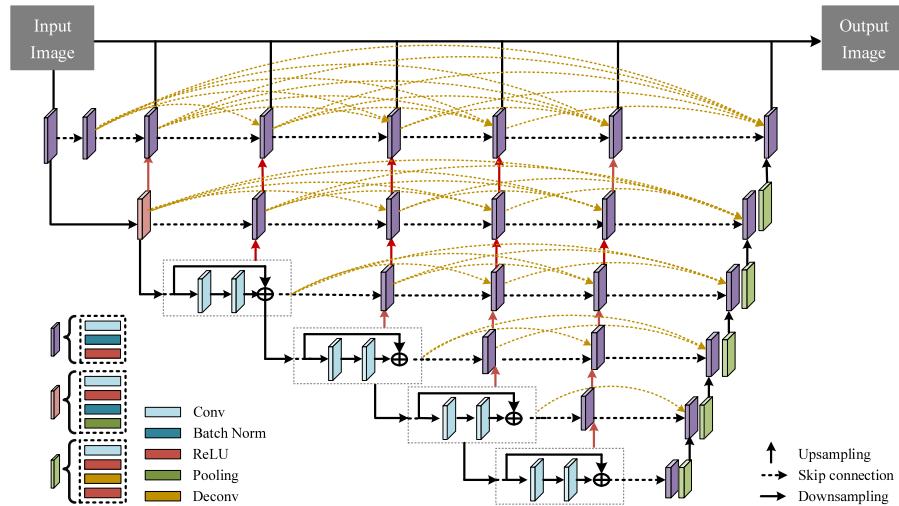
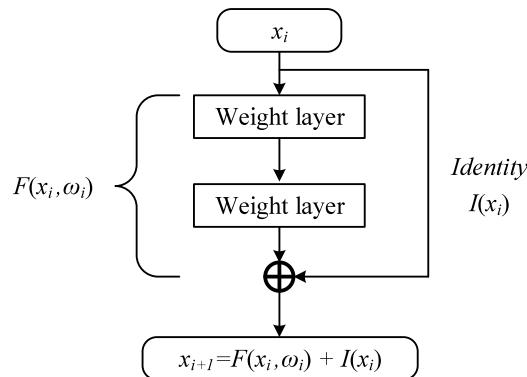
where  $w_i$  is the weight of the  $i$ th layer, and  $F(\cdot)$  is a non-linear transformation [10]. Unlike a traditional CNN, the relationship between the input  $x_i$  and output  $x_{i+1}$  in the building blocks of ResNet [16] is mapped as:

$$x_{i+1} = F(x_i, w_i) + I(x_i), \quad (2)$$

where  $I(\cdot)$  is the fast identity connection. As shown in Fig. 2, ResNet directly introduces a residual mapping (i.e., an identity mapping between the input and output), and utilizes residual mapping instead of the original feature mapping. Motivated by the superiority of ResNet in avoiding network performance degradation with increasing network depth, we embedded the reshaped ResNet34 into multiple shallow sub-networks in the encoder module of WBC-Net. As shown in Fig. 1, the encoder module of WBC-Net combines convolutional blocks with residual blocks, where each residual block contains two  $3 \times 3$  convolutional blocks.

### 3.2. Mixed skip pathways

In deep learning-based feature extraction, shallow features are simple and specific, while deep features are complex and abstract [15]. Both types of image features have their own advantages, and both are usually fused by skip connection. FCN [13] and U-Net [14] implement skip connection by using addition and concatenation, respectively. They use skip connection to combine shallow feature maps originating from the encoder module with deep feature maps originating from the decoder module. However, both FCN and U-Net have difficulty in fusing shallow and deep feature maps. UNet++ [15] uses skip pathways to reduce

**Fig. 1.** Architecture of WBC-Net.**Fig. 2.** Architecture of building block in ResNet [16].

the semantic gap between feature maps generated by the encoder module and those generated by the decoder module. Inspired by UNet++, the WBC-Net uses skip pathways on dense convolutional blocks to bridge the semantic gap between shallow and deep feature maps. Fig. 3 illustrates the skip pathways of WBC-Net, where rectangles indicate convolutional blocks. The solid lines are originally used in U-Net, and the dashed lines indicate skip connections on dense convolutional blocks. U-Net concatenates  $X^{0,0}$  with  $X^{1,2}$  after upsampling to obtain  $X^{0,3}$ , and there is a semantic gap between  $X^{0,0}$  and  $X^{1,2}$ . To bridge the gap, WBC-Net adds skip connections between  $X^{0,0}$  and  $X^{0,2}$ , and between  $X^{0,1}$  and  $X^{0,3}$ .

### 3.3. Deep supervision structure

Fig. 4 shows that WBC-Net performs skip connections on dense convolutional blocks. When only using the final convolutional block to calculate the loss of our network, the middle and final convolutional blocks do not connect during the backpropagation process, and the middle convolutional blocks cannot receive the gradient from the backpropagation. Inspired by deep supervision [24], we treat each layer as a sub-network and calculate each layer's loss to ensure full training of the entire network. As shown in Fig. 4, we take the outputs under different scales as the sub-network, and the outputs under five scales (i.e.,  $X^{0,0}$ ,  $X^{0,1}$ ,  $X^{0,2}$ ,  $X^{0,3}$ ,  $X^{0,4}$ ) are used to calculate the WBC-Net loss. Accordingly, the segmentation outputs of five branches are averaged.

### 3.4. Loss function

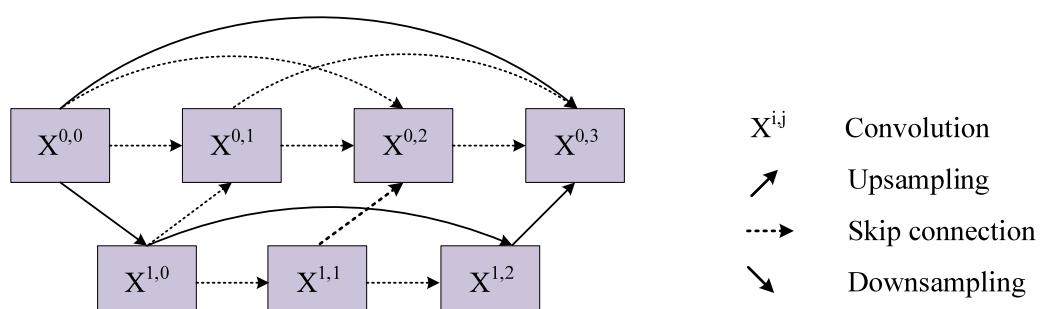
WBC-Net defines the loss function as:

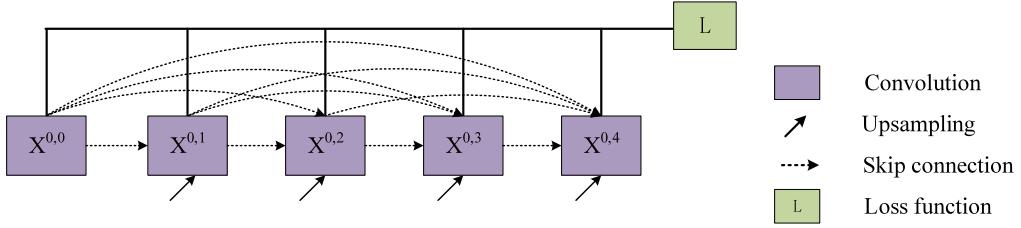
$$L = L_{BCE} + L_{Tversky}, \quad (3)$$

where  $L_{BCE}$  indicates the binary cross-entropy loss [25] used to check the accuracy of each pixel's predicted label. The definition of  $L_{BCE}$  is,

$$L_{BCE} = - \sum_{i=1}^N [g_{ic} \log p_{ic} + (1 - g_{ic}) \log(1 - p_{ic})], \quad (4)$$

where  $N$  indicates the total number of image pixels,  $c$  indicates the object class  $c$  (i.e., WBC), and  $p_{ic}$  indicates the predicted probability of pixel  $i$  belonging to object class  $c$ . Additionally,

**Fig. 3.** Skip pathways of WBC-Net.



**Fig. 4.** Deep supervision of WBC-Net.

$g_{ic}$  denotes the training label of pixel  $i$  belonging to object class  $c$  in a ground truth image. The training labels (i.e.,  $g_{ic}$ ) 1 or 0 denote whether pixel  $i$  belongs to the object class (i.e., WBC) or background class (i.e., non-WBC), respectively. However, when object pixels are far less than background pixels, segmentation performance with unbalanced samples will be degraded. To alleviate this issue, we added a second item  $L_{Tversky}$  (i.e., the Tversky loss [26]), and which is defined as:

$$L_{Tversky} = \sum_c (1 - T_c), \quad (5)$$

$$T_c = \frac{\sum_{i=1}^N p_{ic} g_{ic} + \epsilon}{\sum_{i=1}^N p_{ic} g_{ic} + \alpha \sum_{i=1}^N p_{ic} g_{\bar{c}} + \beta \sum_{i=1}^N p_{\bar{c}} g_{ic} + \epsilon}, \quad (6)$$

where  $p_{ic}$  and  $g_{ic}$  are the same as those in Eq. (4). The symbols  $p_{ic}$  and  $g_{\bar{c}}$  indicate the predicted probability and training label of pixel  $i$  belonging to the background class  $\bar{c}$  (i.e., non-WBC), respectively. The parameters  $\alpha$  and  $\beta$  are used to trade-off false-positives and false-negatives, and both of them are empirically set to 0.3 and 0.7, respectively. The symbol  $\epsilon$  is a very small constant used to prevent dividing by zero.

#### 4. Experimental results

To validate the effectiveness of WBC-Net in WBC segmentation, we compared it with five state-of-the-art methods: AHT [4], SVM [8], FCN [13], U-Net [14], and U-ResNet34 [17]. This section first introduces the image datasets used and the segmentation performance evaluation measures. Then, the implementation of the experiments is described. Finally, the WBC segmentation results obtained by different methods are qualitatively and quantitatively compared and analyzed.

##### 4.1. Datasets and evaluation methods

We used four image datasets in our experiments, denoted as Dataset1, Dataset2, Dataset3, and Dataset4. Dataset1 [8] was collected by Jiangxi Tekang Technology Company, China, and contains 300 individual WBC images (176 neutrophils, 22 eosinophils, 1 basophil, 48 monocytes, and 53 lymphocytes) that are  $120 \times 120$  pixels in size. Dataset2 [8] contains 100 individual WBC images (30 neutrophils, 12 eosinophils, 3 basophils, 18 monocytes, and 37 lymphocytes) that are  $300 \times 300$  pixels in size. Dataset3 is also called LISC [27] and contains 242 WBC images (50 neutrophils, 39 eosinophils, 53 basophils, 48 monocytes, and 52 lymphocytes) that are  $712 \times 568$  pixels in size. Dataset4 was collected from the Third People's Hospital of Fujian Province, China. It contains 231 WBC images (49 neutrophils, 49 eosinophils, 44 basophils, 43 monocytes, and 46 lymphocytes) that are  $2048 \times 1536$  pixels in size. The images in Dataset1 were captured under rapid staining conditions, while those of the other three image datasets were captured under standard staining conditions. The WBCs in each image of Dataset1 and Dataset2 cover a high proportion of the image area, while those of Dataset3 and Dataset4 cover low proportion.

To quantitatively evaluate the accuracy of WBC segmentation, we used six common measures: Precision [28], Dice [29], mIoU [30], Misclassification Error (ME) [31], False Positive Rate (FPR) [32], and False Negative Rate (FNR) [33]. These measures are defined as:

$$\text{Precision} = \frac{|G_f \cap P_f|}{|P_f|}, \quad (7)$$

$$\text{Dice} = \frac{2 |G_f \cap P_f|}{|G_f| + |P_f|}, \quad (8)$$

$$\text{mIoU} = \frac{1}{2} \left( \frac{|G_b \cap P_b|}{|G_b \cup P_b|} + \frac{|G_f \cap P_f|}{|G_f \cup P_f|} \right), \quad (9)$$

$$\text{ME} = 1 - \frac{|G_b \cap P_b| + |G_f \cap P_f|}{|G_f| + |G_b|}, \quad (10)$$

$$\text{FPR} = \frac{|G_b \cap P_f|}{|G_b|}, \quad (11)$$

$$\text{FNR} = \frac{|G_f \cap P_b|}{|G_f|}, \quad (12)$$

where  $P_f$  and  $P_b$  indicate the WBC and non-WBC regions in the predicted segmentation results, respectively, and  $G_f$  and  $G_b$  denote the WBC and non-WBC regions in the ground truths, respectively. Lower values of ME, FPR, and FNR indicate better segmentation accuracy, while higher values of Precision, Dice, and mIoU indicate better segmentation accuracy.

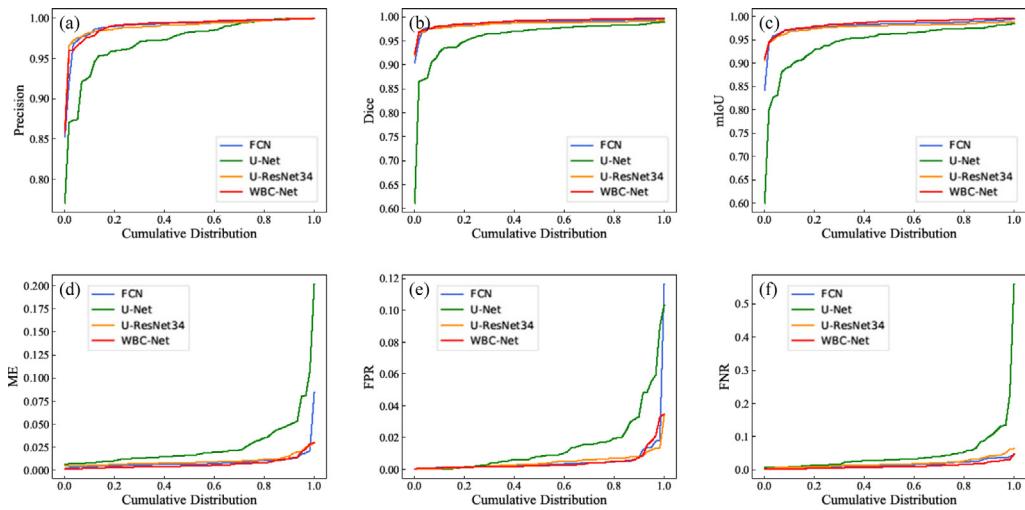
##### 4.2. Implementation details

The experiments were run on a server with the Linux operating system, 2.30 GHz processor, 128 GB RAM, and NVIDIA TESLA P100 GPU. When evaluating the segmentation performance of the above six methods on the four image datasets, 80% and 20% of the images of each dataset were used to construct the training set and testing set, respectively. To expand the scale of training samples, we performed a series of data augmentation operations, including rotation, zooming, and flipping horizontally from left to right.

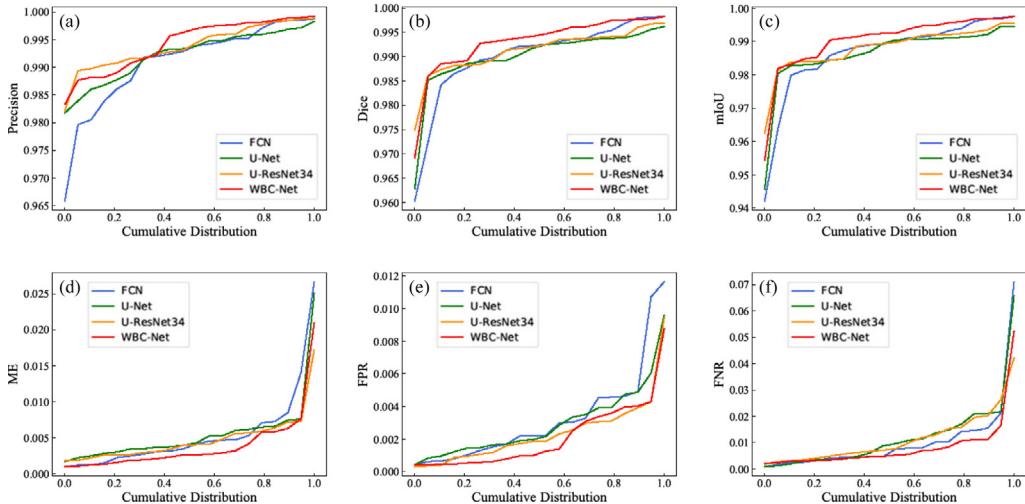
##### 4.3. Results and analysis

###### 4.3.1. Quantitative results

To quantitatively compare the effectiveness of the different methods in WBC segmentation, we plotted cumulative distribution curves of the six accuracy measure obtained by applying the four deep learning-based methods to the four datasets (Figs. 5–8). These figures do not display the cumulative distributions from the classic non-deep learning-based methods (AHT [4], SVM [8]). This is because their curves were dramatically lower than the others,



**Fig. 5.** Cumulative distributions of six measures of segmentation accuracy: (a) Precision, (b) Dice, (c) mIoU, (d) ME, (e) FPR, and (f) FNR resulting from applying four segmentation methods to Dataset1: FCN [13], U-Net [14], U-ResNet34 [17], and WBC-Net.



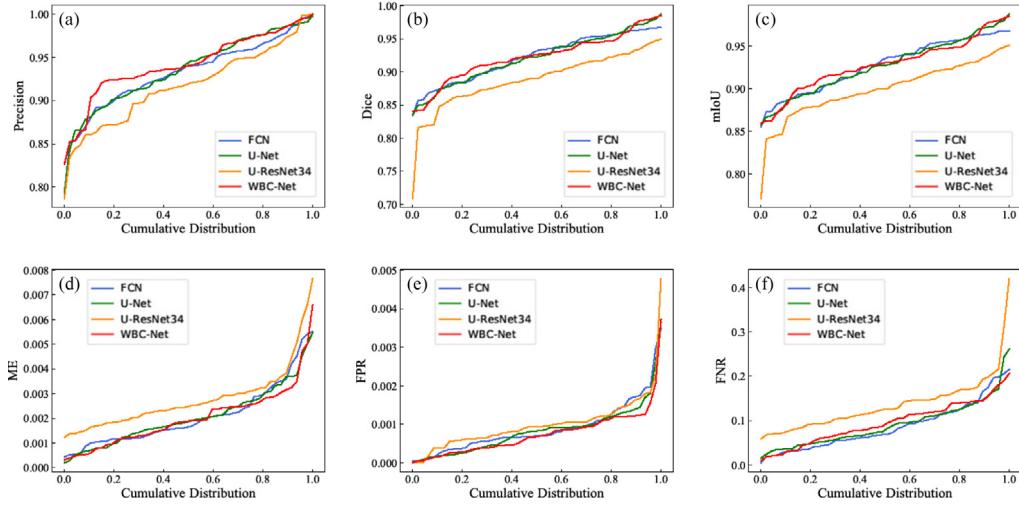
**Fig. 6.** Cumulative distributions of six measures of segmentation accuracy: (a) Precision, (b) Dice, (c) mIoU, (d) ME, (e) FPR, and (f) FNR resulting from applying four segmentation methods to Dataset2: FCN [13], U-Net [14], U-ResNet34 [17], and WBC-Net.

such that including them would complicate visual comparison of the figures. When drawing each cumulative distribution curve for a certain measure, we first sorted the measured values of all images in ascending order. For a point  $(x, y)$  on a cumulative distribution curve,  $x$  and  $y$  indicate the normalized image index and its corresponding measured value. Cumulative distribution curves [34] are usually used to display the trend and overall segmentation performance, where better segmentation accuracy is indicated by higher values of Precision, Dice, and mIoU and lower values of ME, FPR, and FNR.

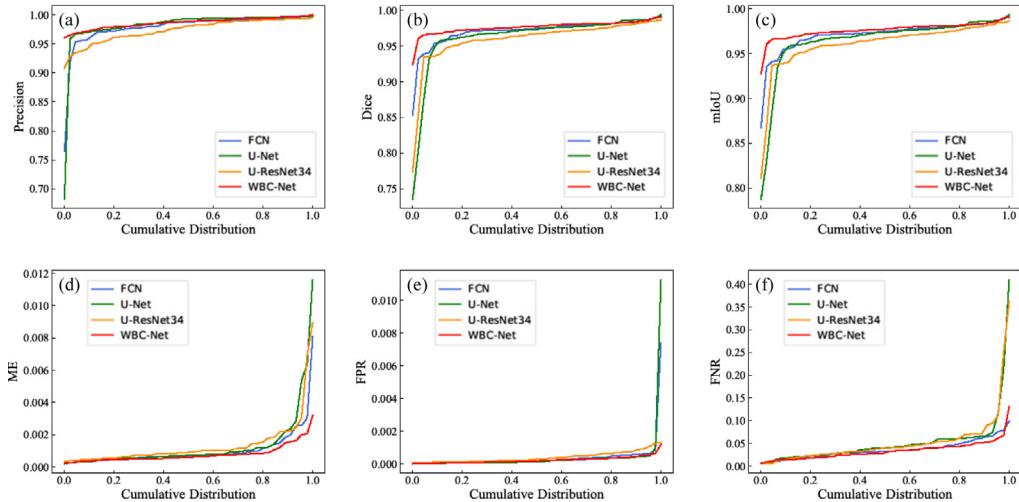
From Fig. 5, one can observe that U-Net obtained the worst segmentation accuracy, having the lowest curves of Precision, Dice, and mIoU and the highest curves of ME, FPR, and FNR. The other three methods obtained similar curves for the six measures. However, the robustness of FCN was lower than those of U-ResNet34 and WBC-Net, as FCN had the starting fragment of the mIoU curve with a large span, as well as the ending fragment of the ME and FPR curves with large spans. In short, WBC-Net had the best segmentation accuracy with Dataset1, as it had the highest curves of Precision, Dice, and mIoU and the lowest curves of ME, FPR, and FNR. From Fig. 6, one can observe that FCN and U-Net usually had the lowest accuracy and robustness, having the

lowest curves of Precision, Dice, and mIoU and the highest curves of ME, FPR, and FNR. U-ResNet34 obtained the intermediate segmentation accuracy, having the middle curves in each subfigure. WBC-Net obtained the best segmentation accuracy with Dataset2, having the highest curves for the first three measures and the lowest curves for the last three. Fig. 7 shows that U-ResNet34 obtained the worst segmentation accuracy and robustness, as it had the lowest curves for the first three measures, and the highest curves for the last three. Fig. 8 shows that WBC-Net had the best segmentation accuracy, having the highest curves for the first three measures and the lowest curves for the last three. Fig. 8 also shows that U-ResNet34 had the worst segmentation accuracy, having the lowest curves for the first three measures and the highest curves for the last three.

Table 1 lists the averages of the six segmentation measures obtained by applying each of the six methods to the four image datasets. Values indicating the best performance are marked in bold. From Table 1, one can observe that WBC-Net obtained the best performance on Dataset1, Dataset2, and Dataset4. For Dataset3, WBC-Net had the best values of the first five measures and the third-best FNR value. Overall, WBC-Net had the best segmentation performance on all four image datasets. For FCN,



**Fig. 7.** Cumulative distributions of six measures of segmentation accuracy: (a) Precision, (b) Dice, (c) mIoU, (d) ME, (e) FPR, and (f) FNR resulting from applying four segmentation methods to Dataset3: FCN [13], U-Net [14], U-ResNet34 [17], and WBC-Net.



**Fig. 8.** Cumulative distributions of six measures of segmentation accuracy: (a) Precision, (b) Dice, (c) mIoU, (d) ME, (e) FPR, and (f) FNR resulting from applying four segmentation methods to Dataset4: FCN [13], U-Net [14], U-ResNet34 [17], and WBC-Net.

it produced the second-best results on Dataset1 and Dataset4. U-Net had the second-best results for Dataset3. U-ResNet34 was second-best on Dataset2, and performed worst on Dataset3 and Dataset4. For the traditional methods (AHT and SVM), the segmentation results on the two datasets with small WBCs (Dataset3 and Dataset4) were worse than those on the two datasets with large WBCs (Dataset1 and Dataset2). Additionally, both AHT and SVM were worse than the deep learning-based methods on all image datasets.

#### 4.3.2. Qualitative results

To further qualitatively compare the effectiveness of the different methods in WBC segmentation, Figs. 9–10 show the segmentation results of six methods (AHT [4], SVM [8], FCN [13], U-Net [14], U-ResNet34 [17], and WBC-Net) applied to representative images from four image datasets (Dataset1, Dataset2, Dataset3, and Dataset4). In detail, Fig. 9(a) and (b) show visual segmentation results for the best two and worst two performances of the six methods on Dataset1 and Dataset2, respectively. Similarly, Fig. 10(a) and (b) show this for Dataset3 and Dataset4, respectively. The blue dashed line and red solid line in

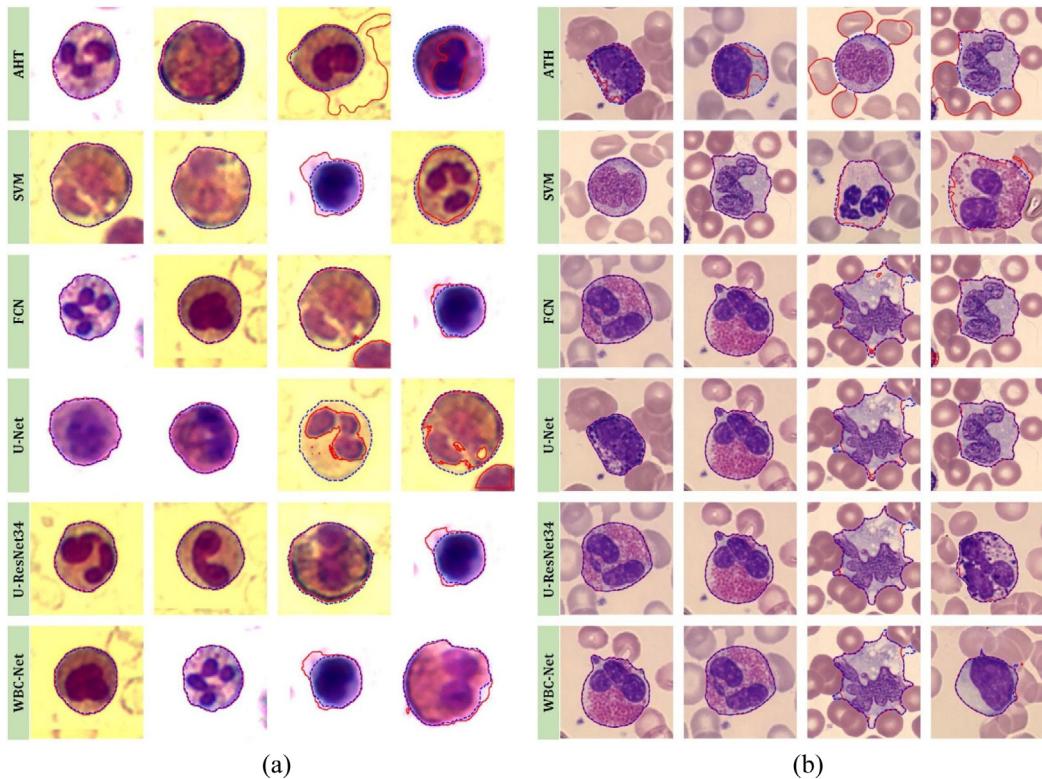
each subfigure indicate the ideal segmentation result (manually-determined ground truth) and the automated segmentation results obtained by various methods.

Fig. 9(a) shows that the four original WBC images were captured under rapid staining conditions, and had different background colors (i.e., white and yellow) due to different staining reagents. From Fig. 9(a), one can conclude that (1) different methods have different best and worst results, since they have their own characteristics. (2) The two best results of all methods achieved satisfactory segmentation; the blue dashed lines (representing ground truth) and the red solid lines (automatic segmentation result) basically coincide. (3) The two worst segmentation results of the different methods have some flaws. In detail, AHT exhibited serious over- and under-segmentation. SVM exhibited obvious over-segmentation and a little under-segmentation. FCN exhibited over-segmentation of different degrees. U-Net displayed serious under- and over-segmentation. U-ResNet34 exhibited a little under-segmentation and obvious over-segmentation. WBC-Net showed various degrees of over-segmentation. Fig. 9(b) shows that in the four images captured under standard staining conditions, segmentation difficulty was mainly caused by the adhesion of WBCs to neighboring red blood cells. One can conclude that (1) the two best results obtained by

**Table 1**

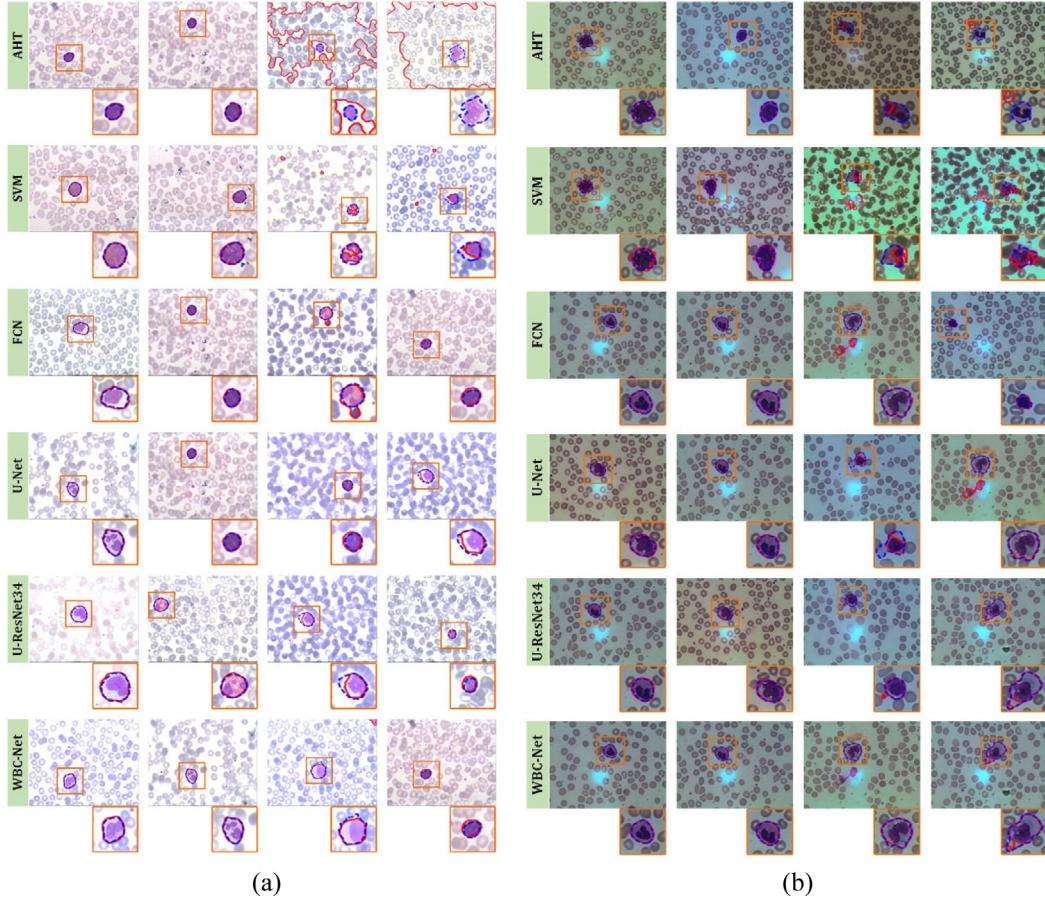
Average values of precision, Dice, mIoU, ME, FPR, and FNR on four image datasets. The best results are marked in bold.

Dataset	Method	Precision (%)	Dice (%)	mIoU (%)	ME (%)	FNR (%)	FPR (%)
Dataset1	AHT [4]	87.22	88.50	85.10	7.37	8.09	6.63
	SVM [8]	95.19	96.85	95.55	1.94	2.18	1.94
	FCN [13]	98.90	98.62	97.97	0.89	0.57	1.64
	U-Net [14]	97.00	95.77	94.23	2.65	1.48	4.91
	U-ResNet34 [17]	98.88	98.43	97.72	0.99	0.47	1.99
	WBC-Net	<b>98.97</b>	<b>98.92</b>	<b>98.44</b>	<b>0.67</b>	<b>0.44</b>	<b>1.12</b>
Dataset2	AHT [4]	89.82	73.77	72.24	12.46	5.67	30.57
	SVM [8]	98.6	97.83	97.06	1.18	0.52	3.00
	FCN [13]	99.08	99.00	98.64	0.54	0.33	1.07
	U-Net [14]	99.23	98.99	98.62	0.55	0.029	1.23
	U-ResNet34 [17]	99.38	99.11	98.79	0.48	0.24	1.14
	WBC-Net	<b>99.42</b>	<b>99.28</b>	<b>99.02</b>	<b>0.39</b>	<b>0.21</b>	<b>0.85</b>
Dataset3	AHT [4]	32.40	37.35	53.65	22.70	22.76	19.30
	SVM [8]	90.19	78.82	82.92	0.51	0.10	27.50
	FCN [13]	93.19	92.16	92.72	0.21	0.09	<b>8.60</b>
	U-Net [14]	93.53	92.08	92.67	0.21	<b>0.08</b>	9.12
	U-ResNet34 [17]	91.72	88.78	89.90	0.28	0.10	13.64
	WBC-Net	<b>94.18</b>	<b>92.24</b>	<b>92.80</b>	<b>0.20</b>	<b>0.08</b>	9.49
Dataset4	AHT [4]	84.33	76.40	83.25	0.62	0.15	25.71
	SVM [8]	60.35	56.15	70.03	1.32	0.61	43.95
	FCN [13]	97.76	97.00	97.08	0.10	0.04	3.65
	U-Net [14]	98.07	96.19	96.42	0.13	0.04	5.26
	U-ResNet34 [17]	97.33	95.88	96.07	0.13	0.04	5.27
	WBC-Net	<b>98.62</b>	<b>97.60</b>	<b>97.63</b>	<b>0.08</b>	<b>0.02</b>	<b>3.37</b>

**Fig. 9.** The two best and two worst results for (a) Dateset1, and (b) Dateset2 produced by the six methods (shown as one per row). Red lines indicate predicted results, and blue lines indicates ground truths.

all methods excluding AHT achieved satisfactory segmentation. AHT exhibited various degrees of under-segmentation. (2) The two worst segmentation results of the different methods show some flaws. In detail, AHT exhibited serious over-segmentation as it erroneously segmented several neighboring red blood cells into the WBC region. SVM and FCN exhibited some under- and over-segmentation. U-Net, U-ResNet34, and WBC-Net exhibited various degrees of under-segmentation. WBC-Net had the least under-segmentation.

**Fig. 10** shows eight microscopic images from the two image datasets captured under standard staining conditions (Dataset3 and Dataset4) at low magnification. The segmentation of WBCs with small sizes is more difficult. The two best and two worst results are shown in **Fig. 10(a)** and (b). An enlarged image of the WBC location is shown in the bottom-right corner of each subfigure for convenient observation of WBC segmentation details. **Fig. 10(a)** shows that (1) the two best segmentation results of all methods are essentially satisfactory. (2) The two



**Fig. 10.** The two best and two worst results for (a) Dateset3, and (b) Dateset4 produced by the six methods (shown as one per row). Red lines indicate predicted results, and blue lines indicates ground truths.

worst segmentation results of all methods excluding WBC-Net have obvious flaws. In detail, AHT erroneously segments many red blood cells into the WBC region. SVM exhibits simultaneous under-segmentation and over-segmentation. FCN exhibits over-segmentation. Both U-Net and U-ResNet34 exhibit under-segmentation. WBC-Net shows slight under- and over-segmentation. Fig. 10(b) shows that (1) The two best segmentation results are essentially satisfactory. (2) One of the two worst segmentation results obtained by AHT, SVM, FCN, and U-Net contains serious errors; some segmented object regions are far away from the WBCs. WBC-Net shows slight under-segmentation.

The two best and two worst segmentation results of the different methods shown in Figs. 9–10 may correspond to different WBC images. For convenient comparison, we show segmentation results for four randomly selected images from each dataset in Figs. 11 and 12. Fig. 11(a) and (b) show that AHT obtains the worst segmentation performance. SVM, FCN, and U-Net obtain intermediate performance. U-ResNet34 and WBC-Net both obtain the best performance, with WBC-Net being superior. Fig. 12(a) and (b) show that the four deep learning-based methods obtain better segmentation performance than AHT and SVM, of these, WBC-Net performs best.

## 5. Discussions

In this section, several groups of ablation experiments are described, which verify the impacts of several key design elements and parameters of WBC-Net on WBC segmentation accuracy. Each group of experiments used 786 WBC images to construct a training set, and 87 WBC images to construct a testing set.

### 5.1. Impact of building blocks

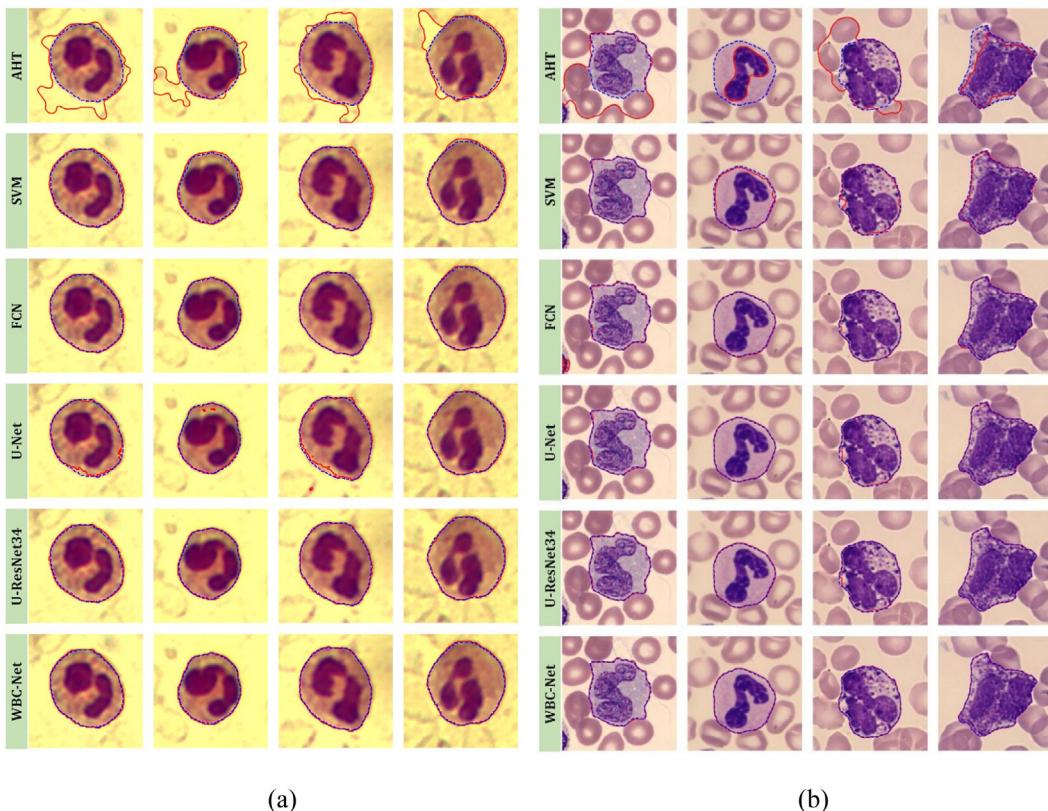
WBC-Net embeds several building blocks of ResNet34 into multiple sub-networks in the encoder module to resolve the issue of the vanishing gradient. To verify the impact of building blocks on WBC-Net, we conducted it with and without them. The experimental results in Table 2 show that WBC-Net obtained higher segmentation accuracy with building blocks than without them. This demonstrates the necessity of embedding several building blocks into the encoder module.

### 5.2. Impact of mixed skip pathways

WBC-Net uses mixed skip pathways to fuse shallow features from the encoder module with deep features from the decoder module. We conducted WBC-Net with and without mixed skip pathways to verify their impact. The experimental results in Table 3 show that WBC-Net obtained higher segmentation accuracy with mixed skip pathways than without them.

### 5.3. Impact of deep supervision structure

WBC-Net uses a deep supervision structure to calculate the loss function value in the training process. We conducted WBC-Net with and without a deep supervision structure to verify its impact. Table 4 shows that the higher segmentation accuracy was obtained with a deep supervision structure.



**Fig. 11.** Example segmentation results of the different methods on (a) Dateset1 and (b) Dateset2. Red lines indicate predicted results, and blue lines indicate ground truths.

**Table 2**  
Impact on segmentation performance of incorporating building blocks into WBC-Net.

Dataset	Method	Precision (%)	Dice (%)	mIoU (%)	ME (%)	FNR (%)	FPR (%)
Dataset1	Without building blocks	97.99	98.11	97.40	1.07	0.78	<b>1.72</b>
	With building blocks	<b>98.48</b>	<b>98.21</b>	<b>97.53</b>	<b>1.01</b>	<b>0.57</b>	2.01
Dataset2	Without building blocks	98.89	98.15	97.50	1.05	0.38	2.45
	With building blocks	<b>98.90</b>	<b>98.33</b>	<b>97.74</b>	<b>0.93</b>	<b>0.35</b>	<b>2.15</b>
Dataset3	Without building blocks	95.70	93.82	94.19	0.18	0.06	7.60
	With building blocks	<b>96.20</b>	<b>94.20</b>	<b>94.60</b>	<b>0.16</b>	<b>0.05</b>	<b>7.19</b>
Dataset4	Without building blocks	93.04	93.86	94.26	0.24	0.13	<b>4.88</b>
	With building blocks	<b>94.03</b>	<b>94.35</b>	<b>94.66</b>	<b>0.20</b>	<b>0.10</b>	5.01

**Table 3**  
Impact on segmentation performance of using mixed skip pathways in WBC-Net.

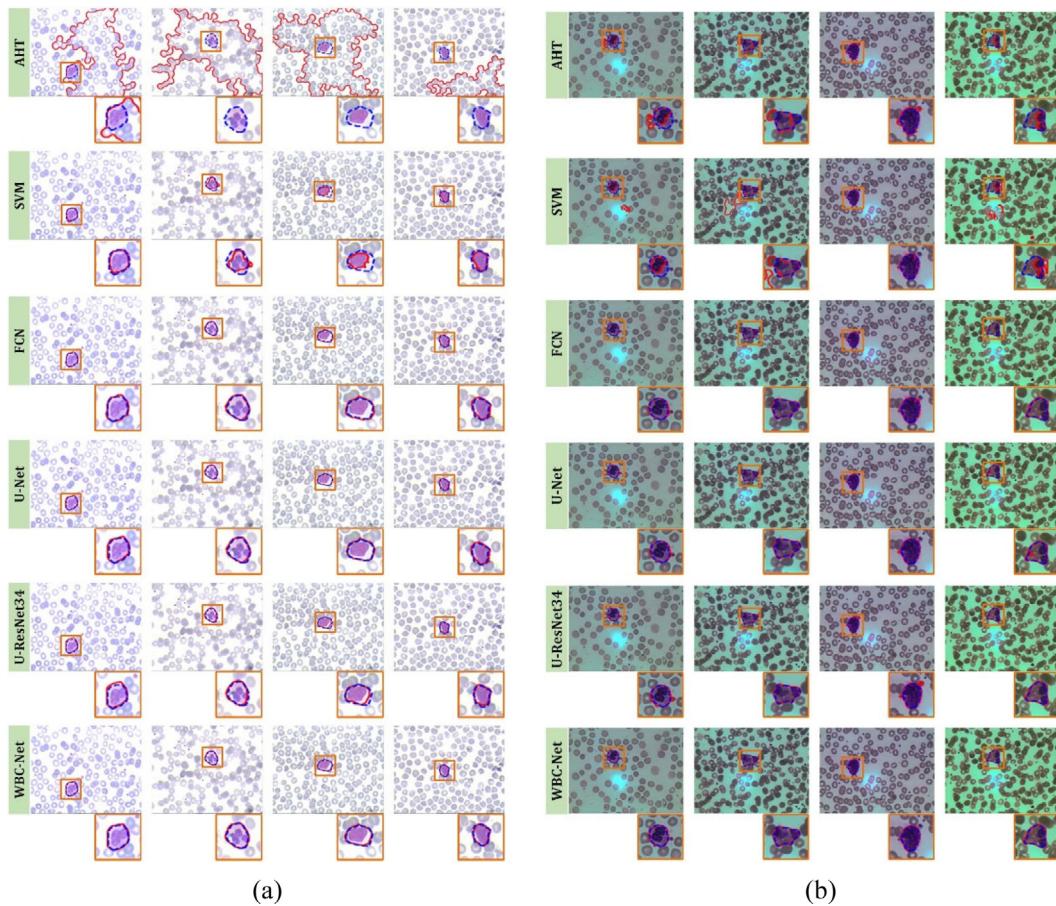
Dataset	Method	Precision (%)	Dice (%)	mIoU (%)	ME (%)	FNR (%)	FPR (%)
Dataset1	Without skip pathways	98.32	98.07	97.35	<b>1.00</b>	0.58	2.11
	With skip pathways	<b>98.48</b>	<b>98.21</b>	<b>97.53</b>	1.01	<b>0.57</b>	<b>2.01</b>
Dataset2	Without skip pathways	98.01	98.15	97.32	1.12	0.37	2.71
	With skip pathways	<b>98.90</b>	<b>98.33</b>	<b>97.74</b>	<b>0.93</b>	<b>0.35</b>	<b>2.15</b>
Dataset3	Without skip pathways	94.87	93.24	93.74	0.19	0.07	7.80
	With skip pathways	<b>96.20</b>	<b>94.20</b>	<b>94.60</b>	<b>0.16</b>	<b>0.05</b>	<b>7.19</b>
Dataset4	Without skip pathways	93.71	91.99	92.78	0.29	<b>0.10</b>	8.70
	With skip pathways	<b>94.03</b>	<b>94.35</b>	<b>94.66</b>	<b>0.20</b>	<b>0.10</b>	<b>5.01</b>

#### 5.4. Impact of loss function

The loss function in WBC-Net includes two items:  $L_{BCE}$  and  $L_{Tversky}$ . The first item is based on cross-entropy, while the second item is based on the Tversky index. To verify the impact of these two loss items on WBC-Net, we performed segmentation with  $L_{BCE}$ , or  $L_{Tversky}$ , or both of them. Table 5 shows that the highest segmentation accuracy was obtained with both loss items.

#### 6. Conclusions

In this paper, we proposed a novel WBC segmentation method based on UNet++ and ResNet, namely, WBC-Net. Specifically, WBC-Net first designs a context-aware feature encoder with convolutional blocks and residual blocks to extract multi-scale image features. Then, WBC-Net introduces mixed skip pathways to fuse semantically similar features to reduce the semantic gap between shallow and deep feature maps. Furthermore, WBC-Net uses a



**Fig. 12.** Example segmentation results of the different methods on (a) Dateset3 and (b) Dateset4. Red lines indicate predicted results, and blue lines indicate ground truths.

**Table 4**  
Impact on segmentation performance of incorporating a deep supervision structure into WBC-Net.

Dataset	Method	Precision (%)	Dice (%)	mIoU (%)	ME (%)	FNR (%)	FPR (%)
Dataset1	Without deep supervision	<b>98.49</b>	97.96	97.22	1.12	0.59	2.52
	With deep supervision	98.48	<b>98.21</b>	<b>97.53</b>	<b>1.01</b>	<b>0.57</b>	<b>2.01</b>
Dataset2	Without deep supervision	98.82	98.23	97.60	0.99	0.36	2.29
	With deep supervision	<b>98.90</b>	<b>98.33</b>	<b>97.74</b>	<b>0.93</b>	<b>0.35</b>	<b>2.15</b>
Dataset3	Without deep supervision	95.25	93.33	93.77	0.19	0.07	8.05
	With deep supervision	<b>96.20</b>	<b>94.20</b>	<b>94.60</b>	<b>0.16</b>	<b>0.05</b>	<b>7.19</b>
Dataset4	Without deep supervision	<b>94.22</b>	93.99	94.40	0.23	0.12	5.71
	With deep supervision	94.03	<b>94.35</b>	<b>94.66</b>	<b>0.20</b>	<b>0.10</b>	<b>5.01</b>

**Table 5**  
Impact on segmentation performance of using a loss function in WBC-Net.

Dataset	Method	Precision (%)	Dice (%)	mIoU (%)	ME (%)	FNR (%)	FPR (%)
Dataset1	$L = L_{BCE}$	98.44	98.17	97.39	1.07	0.58	2.16
	$L = L_{Tversky}$	96.35	97.84	97.08	1.18	1.41	<b>0.54</b>
	$L = L_{BCE} + L_{Tversky}$	<b>98.48</b>	<b>98.21</b>	<b>97.53</b>	<b>1.01</b>	<b>0.57</b>	2.01
Dataset2	$L = L_{BCE}$	98.40	98.30	97.69	0.95	0.57	1.72
	$L = L_{Tversky}$	97.76	98.15	97.51	1.01	0.74	<b>1.40</b>
	$L = L_{BCE} + L_{Tversky}$	<b>98.90</b>	<b>98.33</b>	<b>97.74</b>	<b>0.93</b>	<b>0.35</b>	2.15
Dataset3	$L = L_{BCE}$	96.42	93.92	94.37	0.17	0.05	7.83
	$L = L_{Tversky}$	<b>96.90</b>	91.12	92.04	0.25	<b>0.04</b>	13.16
	$L = L_{BCE} + L_{Tversky}$	96.20	<b>94.20</b>	<b>94.60</b>	<b>0.16</b>	0.05	<b>7.19</b>
Dataset4	$L = L_{BCE}$	93.33	92.83	93.37	0.26	0.11	7.12
	$L = L_{Tversky}$	93.99	88.74	90.23	0.40	<b>0.10</b>	14.53
	$L = L_{BCE} + L_{Tversky}$	<b>94.03</b>	<b>94.35</b>	<b>94.66</b>	<b>0.20</b>	<b>0.10</b>	<b>5.01</b>

loss function combining cross-entropy with the Tversky index to train the network. Experiments on four image datasets obtained

under rapid and standard staining conditions demonstrate that

WBC-Net has better segmentation accuracy and robustness compared with several state-of-the-art methods. Future work will explore shape-aware loss for different types of WBCs, and will extend WBC-Net to a semi-supervised segmentation architecture via adversarial learning.

### CRediT authorship contribution statement

**Yan Lu:** Conceptualization, Methodology, Software. **Xuejun Qin:** Data collection, Data interpretation. **Haoyi Fan:** Software, Validation. **Taotao Lai:** Visualization, Investigation. **Zuoyong Li:** Supervision, Writing - reviewing & editing.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (61972187, 61772254, 61702101), the Fuzhou Science and Technology Project (2020-RC-186), the Fujian Provincial Leading Project (2019H0025), the Government Guiding Regional Science and Technology Development (2019L3009), and the Natural Science Foundation of Fujian Province (2020J02024, 2020J01825, 2019J01756).

### References

- [1] A. GençTav, S. Aksoy, S. ÖNder, Unsupervised segmentation and classification of cervical cell images, *Pattern Recognit.* 45 (12) (2012) 4151–4168.
- [2] X. Li, J. Liu, Z. Liu, X. He, Automatic detection of leukocytes for cytometry with color decomposition, *Optik* 127 (24) (2016) 11901–11910.
- [3] N.M. Salem, Segmentation of white blood cells from microscopic images using kmeans clustering, in: Proceedings of the 31st IEEE National Radio Science Conference (NRSC), 2014, pp. 371–376.
- [4] X. Zhou, C. Wang, Z. Li, F. Zhang, Adaptive histogram thresholding-based leukocyte image segmentation, in: Proceedings of the Advances in Intelligent Information Hiding and Multimedia Signal Processing, 2020, pp. 451–459.
- [5] Y.K. Chan, M.H. Tsai, D.C. Huang, Z.H. Zheng, Leukocyte nucleus segmentation and nucleus lobe counting, *BMC Bioinformatics* 11 (1) (2010) 558.
- [6] P. Viswanathan, Fuzzy c means detection of leukemia based on morphological contour segmentation, *Procedia Comput. Sci.* 58 (2015) 84–90.
- [7] C. Di Ruberto, A. Loddo, L. Putzu, A multiple classifier learning by sampling system for white blood cells segmentation, in: Proceedings of the International Conference on Computer Analysis of Images and Patterns, 2015, pp. 415–425.
- [8] X. Zheng, Y. Wang, G. Wang, J. Liu, Fast and robust segmentation of white blood cell images by self-supervised learning, *Micron* 107 (2018) 55–71.
- [9] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969.
- [10] H. Fan, F. Zhang, L. Xi, Z. Li, G. Liu, Leukocytmask: An automated localization and segmentation method for leukocyte in blood smear images using deep neural networks, *J. Biophotonics* 12 (7) (2019) e201800488.
- [11] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2015, pp. 91–99.
- [12] Y. Wang, L. Zhou, B. Yu, L. Wang, 3d auto-context-based locality adaptive multimodality gans for pet synthesis, *IEEE Trans. Med. Imaging* 38 (6) (2018) 1328–1339.
- [13] 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, pp. 3431–3440.
- [14] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234–241.
- [15] Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, J. Liang, UNet++: A nested u-net architecture for medical image segmentation, in: Proceedings of the Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2018, pp. 3–11.
- [16] 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, pp. 770–778.
- [17] V. Iglovikov, A. Shvets, Ternausnet: U-Net with vgg11 encoder pre-trained on imagenet for image segmentation, 2018, arXiv preprint [arXiv:1801.05746](https://arxiv.org/abs/1801.05746).
- [18] J. Wen, Y. Xu, Z. Li, Z. Ma, Y. Xu, Inter-class sparsity based discriminative least square regression, *Neural Netw.* 102 (2018) 36–47.
- [19] Y. Wang, B. Yu, L. Wang, C. Zu, 3d conditional generative adversarial networks for high-quality pet image estimation at low dose, *Neuroimage* 174 (2018) 550–562.
- [20] J. Li, B. Zhang, D. Zhang, Shared autoencoder gaussian process latent variable model for visual classification, *IEEE Trans. Neural Netw. Learn. Syst.* 29 (9) (2018) 4272–4286.
- [21] J. Li, B. Zhang, G. Lu, D. Zhang, Generative multi-view and multi-feature learning for classification, *Inf. Fusion* 45 (2019) 215–226.
- [22] Z.Q. Zhao, P. Zheng, S. Xu, X. Wu, Object detection with deep learning: A review, *IEEE Trans. Neural Netw. Learn. Syst.* 30 (11) (2019) 3212–3232.
- [23] C. Tian, Y. Xu, Z. Li, W. Zuo, L. Fei, H. Liu, Attention-guided cnn for image denoising, *Neural Netw.* 124 (2020) 117–129.
- [24] C.Y. Lee, S. Xie, P. Gallagher, Z. Zhang, Deeply-supervised nets, in: Proceedings of the International Conference on Artificial Intelligence and Statistics, 2015, pp. 562–570.
- [25] P.T. De Boer, D.P. Kroese, S. Mannor, R.Y. Rubinstein, A tutorial on the cross-entropy method, *Ann. Oper. Res.* 134 (1) (2005) 19–67.
- [26] A. Tversky, Features of similarity, *Psychol. Rev.* 84 (4) (1977) 327.
- [27] S.H. Rezatoghi, H. Soltanian-Zadeh, Automatic recognition of five types of white blood cells in peripheral blood, *Comput. Med. Imaging Graph.* 35 (4) (2011) 333–343.
- [28] H. Fan, F. Zhang, R. Wang, L. Xi, Z. Li, Correlation-aware deep generative model for unsupervised anomaly detection, in: Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2020, pp. 688–700.
- [29] V.K. Singh, S. Romani, H.A. Rashwan, Akram, Conditional generative adversarial and convolutional networks for x-ray breast mass segmentation and shape classification, in: Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, 2018, pp. 833–840.
- [30] C. Zhou, H. Fan, Z. Li, Tonguenet: Accurate localization and segmentation for tongue images using deep neural networks, *IEEE Access* 7 (2019) 148779–148789.
- [31] W.A. Yasno, J.K. Mui, J.W. Bacus, Error measures for scene segmentation, *Pattern Recognit.* 9 (4) (1977) 217–231.
- [32] X. Zhou, Z. Li, H. Xie, T. Feng, Y. Lu, Leukocyte image segmentation based on adaptive histogram thresholding and contour detection, *Curr. Bioinform.* 15 (3) (2020) 187–195.
- [33] Z. Zhong, T. Wang, K. Zeng, X. Zhou, Z. Li, White blood cell segmentation via sparsity and geometry constraints, *IEEE Access* 7 (2019) 167593–167604.
- [34] Z. Zhang, L. Liu, F. Shen, H.T. Shen, Binary multi-view clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* 41 (7) (2019) 1774–1782.