DOI: 10.1002/jbio.201800488

FULL ARTICLE

LeukocyteMask: An automated localization and segmentation method for leukocyte in blood smear images using deep neural networks



Haoyi Fan¹ | Fengbin Zhang^{1*} | Liang Xi¹ | Zuoyong Li^{2,3*} | Guanghai Liu⁴ | Yong Xu⁵

¹School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, China

²Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, Fuzhou, China

³Fujian Province Collaborative Innovation Center of Traditional Chinese Medicine Health Management 2011, Fujian University of Traditional Chinese Medicine, Fuzhou, China

⁴College of Computer Science and Information Technology, Guangxi Normal University, Guilin, China

⁵Bio-Computing Research Center, Harbin Institute of Technology (Shenzhen), Shenzhen, China

*Correspondence

Fengbin Zhang, School of Computer Science and Technology, Harbin University of Science and Technology, Harbin, 150080, China.

Email: zhangfengbin@hrbust.edu.cn Zuoyong Li, Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, Fuzhou, 350121, China. Email: fzulzytdq@126.com

Funding information

Fujian Provincial Leading Project, Grant/ Award Number: 2017H0030; Key Project of College Youth Natural Science Foundation of Fujian Province, Grant/Award Number: JZ160467; Project for Innovative Talents in University of Heilongjiang Province of China, Grant/Award Number: UNPYSCT-2015048; Natural Science Foundation of Fujian Province; Natural Science Foundation of Heilongjiang Province, Grant/Award Digital pathology and microscope image analysis is widely used in comprehensive studies of cell morphology. Identification and analysis of leukocytes in blood smear images, acquired from bright field microscope, are vital for diagnosing many diseases such as hepatitis, leukaemia and acquired immune deficiency syndrome (AIDS).



The major challenge for robust and accurate identification and segmentation of leukocyte in blood smear images lays in the large variations of cell appearance such as size, colour and shape of cells, the adhesion between leukocytes (white blood cells, WBCs) and erythrocytes (red blood cells, RBCs), and the emergence of substantial dyeing impurities in blood smear images. In this paper, an end-to-end leukocyte localization and segmentation method is proposed, named LeukocyteMask, in which pixel-level prior information is utilized for supervisor training of a deep convolutional neural network, which is then employed to locate the region of interests (ROI) of leukocyte, and finally segmentation mask of leukocyte is obtained based on the extracted ROI by forward propagation of the network. Experimental results validate the effectiveness of the propose method and both the quantitative and qualitative comparisons with existing methods indicate that LeukocyteMask achieves a state-of-the-art performance for the segmentation of leukocyte in terms of robustness and accuracy.

KEYWORDS

bright field microscope, cell segmentation, deep neural networks, white blood cells

BIOPHOTONICS

Number: F2018019; National Natural Science Foundation of China, Grant/Award Numbers: 61172168, 61772254

1 | INTRODUCTION

As the principal components of immune cells, white blood samples (WBCs) play a significant role in disease diagnosis such as leukaemia, hepatitis and acquired immune deficiency syndrome (AIDS). The diagnosis of those diseases by pathologists mainly relies on the visual inspection of WBCs in blood smear images captured by the bright field microscopy, due to it's the low-cost and widespread acceptance. During the inspection of WBCs, it is necessary to segment the cells firstly from the original images to analyze the WBCs' properties. A typical blood smear image consists of RBCs, WBCs, platelets and background materials, and the goal of cell segmentation is to extract WBCs from such a complex scene for subsequent diagnosis. However, owing to the limitations of various properties of cells in blood smear images such as size, colour and shape of cells, the adhesion between WBCs and RBCs, and the emergence of substantial dyeing impurities in blood smear images, as shown in Figure 1, it is impractical and time-consuming for pathologists to produce manual segmentation for the whole slide image.

Fortunately, in the past few decades, advances in computer-aided methods have led to a faster and more reproducible medical image analysis than manual analysis [1, 2]. Varied automated segmentation tools have been developed and roughly split into 2 groups, unsupervised and supervised methods.

The unsupervised approaches include clustering-based methods [3-5], thresholding-based methods [6, 7] and shape-based methods [8, 9]. The method proposed in Ref. [8] utilizes colour and shape prior for cell segmentation by defining 2 transformations and introduces an efficient use of these transformations in a marker-controlled watershed algorithm. Inspired by the successful applications of saliency detection [10–12] in image processing, a nucleus saliency model based on average absolute difference is built in Ref. [9] to remove dyeing impurities and erythrocyte fragments for a precise localization of each leukocyte. Although abovementioned methods are effective when the colour of the WBCs is distinct from surrounding materials such as RBCs, platelets and background materials in the staining image, those traditional unsupervised methods do not perform well when region of interests (ROIs) of cells have a large of variations in colour, size and shape and lots of parameters need to be adjusted to avoid over/under-segmentation. Therefore, owing to the similar colour and the adhesion between WBCs and RBCs, and the emergence of substantial dyeing impurities in blood smear images, satisfying results from those unsupervised methods leave much to be desired, and moreover, all those unsupervised methods assume some structure in the data that may not fit every case.

The supervised approaches model the problem of segmentation as a multi-class classification task. The traditional approaches vary and include neural network [13], K-nearest neighbor classifier [14], support vector machine (SVM) [15-19], Bayesian classifier [20, 21], random forest [22], etc. Different from unsupervised ones, which assume certain structure in the data to fit the model for segmentation, supervised methods do not assume any structure rather aim to learn it from the data. Such methods generally require 2 independent steps: feature extraction and classification. In Ref. [21], features are extracted from L*a*b colour space to distinguish the cell ROI and non-cell ROI; For the same purpose, scale-invariant feature transform (SIFT) is applied in Ref. [15], to construct more reliable discriminating features; texture information is also integrated for pixel classification in Ref. [23]. More recently, a self-supervised learning method Ref. [19] combining the K-means clustering and SVM classifier to perform WBC segmentation, in [19], a supervised initial segmentation module is construct firstly and supervised refinement of the initial segmentation based on the generated coarse WBC region is employed for final segmentation. Although each of those traditional classifiers had its fair share of success, lots of limitations during feature extraction and classification are still existed among them, for example, one common limitation existed in those methods is that the features are manually designed based on certain prior knowledge which may not robust enough to all situations.

Recently, approaches based on deep convolution neural network (CNN) have achieved remarkable success in the field of computer vision and image processing, for example, image classification [24], object detection [25], image retrieval [26], face recognition [27] and semantic segmentation [28]. In medical image segmentation, lots of CNNbased methods are also widely used benefitting from its powerful ability of feature learning and representation. Among these methods, the fully convolutional network (FCN) [28] has shown the state-of-the-art performance in cell and organ segmentation problems [29-32]. In Ref. [29], U-Net is developed based on FCN and takes skip connection between encoder and decoder into consideration, which localizes objects better by extending the symmetricautoencoder design to combine high-resolution features from the encoding path with upsampled outputs in the decoding

BIOPHOTONICS

3 of 17



FIGURE 1 Examples of blood smear images which containing leukocytes, erythrocytes, platelet and substantial dyeing impurities. Columns 1 to 5 are basophils, eosinophils, lymphocytes, monocytes and neutrophils, respectively; Row 1 are rapidly-stained images; Rows 2 to 4 are traditional wright-stained images

path. In Ref. [30], FCN is first trained to learn a coarse model for the pixel-level prediction of nuclei segmentation, then, sub-regions concluding nucleus are cropped from both the coarse prediction and the original image, and the final refined segmentation is obtained by employing a graphbased approach on the 2 sub-images cropped from the prediction and the original image, respectively. Actually, with such a graph-based segmentation stage, the model proposed in Ref. [30] is no more an end-to-end trainable architecture. In Ref. [31], U-Net-based network is employed to identify and segment the heart region of Drosophila at different developmental stages, which achieves a high intersection over union (IoU) rate on custom optical coherence microscopy images. In Ref. [32], a custom designed convolutional neural network operating on focus stack of images is used to build a focus stacking-based approach for automated quantitative detection of Plasmodium falciparum

malaria from blood smear. However, all above-mentioned CNN-based methods segment cells or organ directly on the whole image, which are vulnerable to the complex background such as substantial dyeing impurities in blood smear images, and rely solely on strong supervision via high-quality, but high-cost, dense segmentations.

In this paper, a new end-to-end leukocytes (WBCs) localization and segmentation method, called LeukocyteMask, which segments WBCs in a pixel-to-pixel manner automatically based on the deep convolution neural network, is proposed to solve previous issues as mentioned above. LeukocyteMask employs multi-scale feature maps from an improved feature pyramid network (FPN) [33] to acquire stronger semantic features for the localization of WBC ROI, and then generates a precise segmentation mask based on the proposed ROI, which does not require any preprocessing and is robust to variant cells of different appearance and complex background under different imaging and staining conditions.

The main contributions of this paper are as follows:

1. The proposed LeukocyteMask is an end-to-end method which segments leukocytes in a pixel-to-pixel manner. Different from previous CNN-based methods [23–26] which perform segmentation directly on the original image with complex background. In Leukocyt-eMask, candidate region for each WBC is located first, on which segmentation is then carried out to obtained the final precise segmentation results of WBC.

2. Different from the original Mask-RCNN model [34], a modified feature pyramid network (FPN) based on ResNet [24], which is designed in terms of the characters of WBCs in blood smear images, is employed as the backbone of LeukocyteMask to extract the discriminative features of WBCs.

3. By observing the training samples, a new data augmentation method for blood smear images is proposed. We postulate that the distribution of cells in blood smear images has some invariance with respect to not only affine transformations, but also elastic deformations caused by the adhesion between cells and the growth of cell itself, therefore, an elastic transformation [35] previously used for document analysis is applied for WBC segmentation to enhance the generalizability and robustness of the model.

4. Quantitative and qualitative comparisons among our proposed method and the current state-of-the-art methods [8, 28, 29] are conduct on 4 different commonly used dataset, and the results show that LeukocyteMask outperforms other methods significantly and achieves an outstanding result in terms of both accuracy and robustness.

The rest of the paper is organized as follows: First, in Section 2, the proposed method for leukocyte localization and segmentation are described in detail. Then, experiment results on 4 different datasets compared to the other state-of-the-art methods are presented in Section 3. Finally, this paper is concluded in Section 4.

2 | PROPOSED METHOD

For the segmentation of leukocytes in blood smear images, our goal is to automatically segment the leukocytes from the erythrocytes and complex background, without any manual intervention and preprocessing. To achieve this, we model the problem as a binary dense labelling task: Given a blood smear image captured by the bright filed microscope, which contains WBCs, RBCs and other noise background regions such as dyeing impurities, we aim to predict either "WBC" or "non-WBC" (including RBCs and noise background) labels for each pixel in the blood smear image. To make full use of prior knowledge such as shape, colour and texture of cells, and spatial information of images, an improved Mask-RCNN together with a WBC-oriented FPN layer is employed in the proposed LeukocyteMask architecture to solve the modelled binary labelling problem. The architecture of LeukocyteMask is illustrated in Figure 2, in which, 3 stages for leukocyte localization and segmentation is constructed: feature extraction, region proposal, and prediction. The details of the proposed method will be described in the following sections.

2.1 | Towards WBC-oriented feature extraction

As proved in many previous researches [36, 37], a discriminating feature extraction network should be deep enough with many convolution layers such that multi-level features can be sufficiently learned. However, along with the increase of network's depth, it would be more difficult to optimize the weights of network since it may produce the vanishing or exploding gradients problem. Inspired by ResNet [24],



FIGURE 2 Architecture of LeukocyteMask



Residual Network (ResNet50)

FIGURE 3 Residual network without fully connected layer head. (conv stands for convolutional layers)

we address this issue by fitting a residual mapping instead of the original mapping, and by adding several connections between layers of deep convolutional neural network. As shown in Figure 3, a deep ResNet is a modularized architecture that is constructed from multiple ResNet building blocks. Each block has a shortcut connection in parallel with traditional convolutional layers, which connects the input feature directly to its output.

In the traditional convolutional layers (Figure 4(A)), a mapping between input and output of each layer is learned as following:

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

where x_i and x_{i+1} are the input and output of the *i*th layer, respectively, w_i represents a set of weights and biases associated with the *i*th layer, $F(\cdot)$ is a nonlinear transformation which consists of convolutions, batch normalization (BN) [38], and rectified linear units (ReLU) [39].

Different from traditional CNNs, a ResNet building block (Figure 4(B)) performs mapping as following:

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

where x_i and x_{i+1} are the input and output of the *i*th ResNet building block, respectively, w_i is a set of weights and biases associated with the *i*th block and $I(\cdot)$ is the identity function.

Considering the powerful ability of feature learning and network optimization, an improved ResNet architecture is utilized in our proposed LeukocyteMask with a WBC-oriented modification to extract more reliable and distinctive features for WBC identification and localization. Due to the fine texture



FIGURE 4 Learning flow of (A): a traditional CNN, (B): a ResNet building block

features appeared in the WBCs which are different from images in the nature scene, the conv1 building block of the original ResNet50 is replaced with 2 convolutional layers in which the filer size is 3, to extract more finer basic features. Moreover, owing to the different cell body colour caused by different imaging and staining conditions, the number of building blocks in conv3_x, conv4_x is reduced to 2 and 3, respectively to prevent overfitting. The configuration details of the modified Res Net are shown in the bottom left of Figure 5.

Although the features extracted through solely ResNet capable of representing relatively discriminative characteristics of WBCs, it has been proven that better performance can be boosted by employing pyramid representations for multi-scale feature maps [33]. Therefore, a FPN based on the modified WBC-oriented ResNet is used as the backbone of feature extraction in LeukocyteMask, to extract more



FIGURE 5 WBC-oriented FPN architecture

6 of 17

representative and reliable multi-scale features. As shown in Figure 5, the FPN architecture used in LeukocyteMask consists of 3 parts: the Bottom-up Pathway, the Top-down Pathway, and the Lateral Connection (LC) between 2 pathways.

The Bottom-up Pathway is comprised of the modified ResNet which consists of 5 convolution modules and in each module, multiple building blocks are utilized to extract features of WBCs. The output of the last building block in each convolution module is employed to build corresponding feature maps in different pyramid levels, as shown in Figure 6, the extracted feature maps in different levels represent discriminatively for WBC compared to the RBCs and other substances.

As going to the top-down pathway, feature maps extracted from bottom-up pathway are upsampled by a factor of 2 using nearest neighbours up-sampling, after that, the upsampled feature maps are merged with the corresponding bottom-up feature maps in the Lateral Connection module by going through a 1×1 convolution layer with an element-wise addition operator and then a 3×3 convolution is applied to all merged maps to obtain the final pyramid feature maps in the Top-down Pathway. The pyramid feature maps are denoted as P2, P3, P4, P5, respectively in Figure 5, and P6 is just a subsampling of P5 with a factor of 2, which is different from the original FPN, to cover a larger perceptual field for WBC to increase classification accuracy. The 3×3 convolution filter used here is to weaken the aliasing effect due to up-sampling [33]. Finally, the extracted pyramid feature maps in different levels are utilized in the next region proposal stage, as described in Section 2.2, to locate WBC ROIs at various scales.



FIGURE 6 Examples of feature maps extracted from WBC-oriented FPN. Columns 1 to 11 are the original image, output of conv1, conv2_x, conv3_x, conv4_x, conv5_x, P2, P3, P4, P5 and P6, respectively; Rows 1 to 4 are 4 different types of WBC samples. (The feature maps at different levels are randomly selected for visualization)

2.2 | Architecture of LeukocyteMask

Inspired by the successful use of Mask-RCNN [34] in instance segmentation tasks, LeukocyteMask is proposed based on an improved Mask-RCNN architecture, in which, a WBC-oriented FPN network is designed for leukocytes localization and segmentation in blood smear images captured by bright field microscope. The proposed LeukocyteMask consists of 3 stages for leukocytes localization and segmentation: feature extraction, region proposal and prediction.

In the feature extraction stage, an improved FPN network, as shown in Figure 5, is employed as a WBC-oriented feature extraction module, to extract more discriminating and reliable features for the next region proposal stage. The details of the modified FPN are described in Section 2.1.

In the region proposal stage, WBC ROIs are localized for the final segmentation of leukocyte. In this stage, Region Proposal Network (RPN) [40] is utilized, which consists of 2 convolution layers, as shown in Figure 8, to locate the regions that might contain WBC objects in the feature maps generated from FPN in the last stage. Then, different from the original RPN, which followed by a RoI pooling layer [41] to crop and resize the feature maps, a RoIAlign layer [34] is applied in LeukocyteMask which uses bilinear interpolation to resolve the misalignment issue encountered in the RoI pooling layer. As illustrated in Figure 8, in RPN, with feature maps generated from FPN as input, a slidingwindow is slid over them and then mapped to a 2048-dimensional vector by a convolutional layer with a filter size of 3×3 . Follow that, the vector is fed into 2 sibling branches with 2 1 \times 1 convolutional layers, 1 for ROI box classification and the other for ROI box regression. For each sliding-window, both branches simultaneously predict k region proposals, in which 3 different anchor aspect ratios {2:1, 1:1, 1:2} and 5 scales (the scales are determined by the 5 feature maps P2, P3, P4, P5 and P6 generated from FPN) are utilized in LeukocyteMask, yielding k = 15 anchors in this case at each sliding position. The anchor here means the proposal centred at the sliding window, depicted as a red box in Figure 8 for visualization. Finally, 4 *k* outputs which encoding the coordinates of *k* boxes (*x*, *y*, width, height of box) and 2 *k* scores which measure the probability of WBC/non-WBC for each proposal, are obtained in box classification branch and box regression branch, respectively.

For training of RPN, a binary classification problem is modelled, in which the anchors that have an Intersectionover-Union (IoU) higher than 0.7 (a threshold set in this paper) with ground-truth box are assigned positive labels and those anchors whose IoU lower than 0.3 are labelled negative instances, the anchors whose IoU between 0.3 and 0.7 do not contribute to the training objective. The objective function used here is defined in Eq. (3), which is optimized by back-propagation, stochastic gradient descent and momentum:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cla}} \sum_{i} L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_{i} p_i^* L_{reg}(t_i, t_i^*)$$
(3)

where p_i is the probability that the *i*th candidate anchor contains a WBC, the ground truth p_i^* masks whether an anchor truly contains a WBC, where 1 is a positive mark and 0 if negative. t_i is the parameterized coordinate of the bounding box predicted by the RPN, and t_i^* is the coordinate of the ground truth corresponding to the positive anchor. L_{cls} is the log loss of the binary classification (WBC/non-WBC), and L_{reg} is the smooth L1 loss of the predicted bounding box and the ground truth box. The total loss of the RPN is normalized by N_{cla} , N_{reg} and a balancing weight λ . Finally, a series



FIGURE 7 Region Proposal Network (RPN) used in LeukocyteMask for WBC localization

Anchor boxes of 3 types aspect ratios

8 of 17 Journal of BIOPHOTONICS

of WBC ROI proposals are obtained by optimizing this loss function via network training.

In the prediction stage, a RoIAlign layer is utilized firstly, by using bilinear interpolation to rectify the misaligned WBC proposals [34], to convert each of proposals generated from RPN layer into a fixed size feature maps. After alignment, 2 following parallel branches, called Localization Branch and Mask Branch, respectively, perform the final WBC localization and segmentation tasks as illustrated in Figure 2.

In the Localization Branch, each fix sized feature map is fed into 2 concatenated fully connected layers (FCs), after which, 2 sibling FCs are forked, as a classifier and a regressor, respectively, to perform box-refinement and boxclassification.

In the Mask Branch, the feature maps of positive proposals selected by RPN are fed into a tiny FCN, which consists of a few stacked convolutional layers and enables a pixel-to-pixel semantic segmentation, and then a soft mask, represented by float numbers between 0 and 1, is obtained, as shown in Figure 7. Different from the traditional binary mask, the soft mask used here holds more details for the final segmentation.

For training of Mask Branch, an average binary crossentropy loss function is used, as defined in Eq. (4):

$$L_{\text{mask}} = -\frac{1}{m^2} \sum_{1 \le i,j \le m} \left[y_{ij} \log \hat{y}_{ij}^k + (1 - y_{ij}) \log \left(1 - \hat{y}_{ij}^k \right) \right] \quad (4)$$

where y_{ij} is the label of a pixel (i, j) in the ground truth of size $m \times m$; \hat{y}_{ij}^k is the predicted value of the same pixel in the ground truth for class *k*. Here, *k* is 1 for WBC class and 0 for non-WBC class in blood smear images.

Finally, LeukocyteMask is trained by using a multi-task loss, similar with the original Mask-RCNN [34], which combines the losses of box-classification, box-refinement of Localization Branch, and the loss of mask segmentation of Mask Branch, as defined in Eq. (5): FAN ET AL.

where box-classification loss L_{cls} and box-refinement loss L_{box} are same as those defined in [40]. By using such a multi-task loss function, the mutual promotion between the mask prediction and region proposal leads to a more precise localization of the WBC boundary.

3 | EXPERIMENTAL RESULTS AND DISCUSSION

In this section, experimental results are reported to validate the proposed model for WBC localization and segmentation. First, datasets and their evaluation criteria are presented. Then, a new WBC-specific data augmentation technique used for generating more training data and the setup details of implementation are described. Finally, experimental results compared with several latest methods are shown, which proves that the proposed LeukocyteMask not only achieves state-of-the-art segmentation accuracy on the commonly used dataset, but also performs stable on the new collected complex dataset.

3.1 | Dataset and evaluation methods

To evaluate the proposed method in terms of both accuracy and robustness, different from previous works [8, 9, 19], in which at most 2 different datasets are used for evaluation, in this paper, 4 datasets called Dataset 1 [19], Dataset 2 [19], BCISC¹ and LISC [42], respectively, captured under various imaging and staining conditions by different medical microscopy institutes, are evaluated in the experiment.

Dataset1 was obtained from Jiangxi Tecom Science Corporation, China, which contains totally 300 sub-images of single WBC with size of 120×120 (176 neutrophils, 22 eosinophils, 1 basophil, 48 monocytes, and 53 lymphocytes); Dataset 2 consists of 100 colour images with size of 300×300 (30 neutrophils, 12 eosinophils, 3 basophils, 18 monocytes and 37 lymphocytes) which is published in

FIGURE 8 Soft mask (right) of a WBC image (left) predicted by 0.9 LeukocyteMask. For visualization, color 0.8 map is used here in which the highest 0.7 value 1 is mapped as red, the middle 0.6 range is yellow-green, and the lowest value 0 is blue. The values in the soft 0.5 mask are represented as the confidence of 0.4 a pixel classified as WBC. It can be seen 0.3 that the predictions on the WBC boundary are soft 0.2



CellaVision²; The third dataset, called BCISC, was collected by us with the help of the Third People's Hospital of Fujian Province which consists of 268 sub-images of single WBC with size of 256×256 (51 neutrophils, 54 eosinophils, 56 basophil, 54 monocytes, and 53 lymphocytes) and labelled by pathologists; And the last dataset, LISC [42], includes the haematological images taken from peripheral blood of healthy subjects, in which 257 sub-images of single WBC with size of 256×256 (56 neutrophils, 39 eosinophils, 55 basophil, 48 monocytes and 59 lymphocytes), are manually segmented by experts. Data examples of 4 datasets are shown in Figure 1.

For evaluation metrics, 6 measures are used in the experiment, which consist of 3 commonly used measure scores in deep learning-based methods [30, 31, 43], namely Precision, Dice coefficient (Dice) and mean Intersection over Union (mIoU), respectively, and other 3 commonly used metrics for traditional segmentation methods [9, 17, 19], namely false positive rate (FPR), false negative rate (FNR) and misclassification error (ME), respectively. All 6 metrics are defined as in Eqs. (6)–(11):

$$Presion = \frac{\left|F_g \cap F_p\right|}{\left|F_p\right|} \tag{6}$$

$$\text{Dice} = \frac{2|F_g \cap F_p|}{|F_g| + |F_p|} \tag{7}$$

$$mIoU = \frac{1}{2} \left(\frac{|F_g \cap F_p|}{|F_g \cup F_p|} + \frac{|B_g \cap B_p|}{|B_g \cup B_p|} \right)$$
(8)

$$FPR = \frac{\left|B_g \cap F_p\right|}{\left|B_g\right|} \tag{9}$$

$$FNR = \frac{\left|F_g \cap B_p\right|}{\left|F_g\right|} \tag{10}$$

$$ME = 1 - \frac{\left|B_g \cap B_p\right| + \left|F_g \cap F_p\right|}{\left|F_g\right| + \left|B_g\right|} \tag{11}$$

where F_p and B_p are WBC region (foreground) and non-WBC region (background) of the prediction of model, respectively; F_g and B_g are WBC region (foreground) and non-WBC region (background) of the ground truth, respectively; $|\cdot|$ is the cardinality of a set.

3.2 | Data augmentation

Deep learning-based models require a large set of training samples to achieve a good generalization capability.

However, as is typical for many deep learning-based medical image analysis tasks, the existing datasets with annotation are quite small by computer vision standards and collecting new large datasets with annotation is time-consuming and sometimes completely impossible. To overcome these limitations, data augmentation is used in many works during the training of network.

Data augmentation is a technique that artificially increasing the volume of the training set by applying several distortions such as brightness changing, zooming, rotation to the original images, in this case, the blood smear images captured by microscope. Utilizing the data augmentation technique instead of training model directly on the tiny original dataset is essential for such circumstance of the absence of large dataset and controlling the overfitting of the model training. However, the distortions applied to original images should not alter spatial pattern and inner characteristic of target objects.

Different from traditional data augment techniques as used in Refs. [28, 31], in which only affine transformation was utilized which functions weakly in this case, in this paper, by observing the samples in the collected datasets, we postulate that the distribution of data has some invariance with respect to not only affine transformations, but also elastic deformations caused by the adhesion between cells and the growth of cell itself. So, in the experiment, not only affine transformations such as flip from horizontal (FlipH), flip vertical (FlipV), Rotation (range from -180° to 180°), are employed, but also a kind of elastic transformation (ET) used in Refs. [35] for document analysis is applied to the original images to increase the diversity of data samples. Furthermore, Gaussian blur (GB), and gamma transformation (GT) which is a nonlinear operation for simulation of variant imaging illuminance, are also used for augmentation. The samples of augmented images are shown in Figure 9.

In the experiment, 3 types of data augmentation configurations: no augmentation (noAug), augmentation without elastic transformation (Aug-noET) and augmentation with elastic transformation (Aug-ET), respectively, are applied for model training. The learning curves of LeukocyteMask under different configurations for 4 datasets are depicted in Figure 10. As is shown, by adding elastic transformation to data augmentation process, although the training of network with Aug-ET is slightly slower than 2 other configurations (noAug, AugnoET), the validation losses developed during the training are dropped obviously, which improves the segmentation performance consequently as described in the Section 3.4.

3.3 | Implement Details

The model is implemented by using Keras³ library and trained on Ubuntu 16.04 OS with 2.5GHz Intel Core i7

10 of 17 Journal of BIOPHOTONICS

CPU, 16GB RAM, and NVIDIA GTX 1080Ti graphic card with 11GB memory. Stochastic gradient descent is used for training, with momentum 0.9, batch size 2, initial learning rate 10^{-3} , weight decay 2×10^{-4} . The training set,

validation set and test set are produced by randomly splitting 60%, 20% and 20% of each dataset, respectively, for training and testing. It should be noted that the training data used for FCN [28] and U-Net [29] are same with the data used during



FAN ET AL.

FIGURE 9 Data augmentation examples of: Dataset1 (Column 1), Dataset2 (Column 2), BCISC (Column 3), and LISC (Column 4). Row 1 to 7 are the original image, and corresponding FlipH, FlipV, Rotation, ET, GB, and GT results, respectively



FIGURE 10 Network convergence of LeukocyteMask during training under different augmentation techniques (Aug, Aug-noET and Aug-ET) on 4 datasets



FIGURE 11 Box-and-whisker plots of different metrics for performance comparison among LeukocyteMask with different augmentation configurations (noAug, Aug-noET, Aug-ET) and other 2 deep learning-based methods (FCN [28], U-Net [29])

the training of LeukocyteMask (Aug-ET) model. In the experiment, 5-fold cross validation are utilized for each model to evaluate the performance.

3.4 | Results and analysis

The segmentation performances of the proposed LeukocyteMask with different data augmentation techniques are compared with one of the most recent traditional methods: watershed-based method [8], and 2 deep learningbased methods: FCN [28] and U-Net [29], respectively. However, different from above mentioned works, in which the mean values of different metrics are measured for performance evaluation, in this paper, the distributions of segmentation results are analyzed by using box-and-whisker plots, to conduct comprehensive analysis on segmentation performance of various methods under different setups. It should be noted that a soft mask threshold is set as 0.5 to get the final binary segmentation masks for all the soft masks generated by LeukocyteMask in the experiment.

12 of 17	Journal of				
	BIOPHOTONICS				

TABLE 1 Average value of Precision, Dice coefficient (Dice), mean Intersect of Union (mIoU), false positive rate (FPR), false negative rate (FNR), and misclassification error (ME) from watershed, FCN, U-Net, and the proposed LeukocyteMask with noAug, Aug-noET, Aug-ET, respectively on Dataset 1, Dataset 2, BCISC and LISC

Dataset	Method	Precision	Dice	mIoU	FPR	FNR	ME
Dataset 1	Watershed	0.99268	0.61732	0.59585	0.00166	0.55657	0.18049
	FCN	0.96001	0.97342	0.96221	0.01833	0.01237	0.01621
	U-Net	0.93319	0.96163	0.94548	0.03231	0.00727	0.02390
	LeukocyteMask(noAug)	0.98982	0.97311	0.96214	0.00255	0.03055	0.01666
	LeukocyteMask(Aug-noET)	0.98918	0.97317	0.96373	0.00286	0.02660	0.01528
	LeukocyteMask(Aug-ET)	0.99544	0.98196	0.96975	0.00198	0.02277	0.01447
Dataset 2	Watershed	0.96203	0.66295	0.64202	0.00941	0.50111	0.13915
	FCN	0.94608	0.93881	0.92634	0.01747	0.06107	0.02770
	U-Net	0.96083	0.95589	0.94678	0.01254	0.04238	0.02015
	LeukocyteMask(noAug)	0.97785	0.96151	0.94365	0.00639	0.06352	0.02569
	LeukocyteMask(Aug-noET)	0.98929	0.96953	0.96073	0.00209	0.04160	0.01725
	LeukocyteMask(Aug-ET)	0.99432	0.98242	0.96671	0.00182	0.03645	0.01648
BCISC	Watershed	0.97407	0.83183	0.82264	0.00406	0.26130	0.05050
	FCN	0.94755	0.95826	0.95200	0.01075	0.02896	0.01377
	U-Net	0.93005	0.95082	0.94427	0.01472	0.02423	0.01664
	LeukocyteMask(noAug)	0.98064	0.96730	0.95537	0.00283	0.05201	0.01419
	LeukocyteMask(Aug-noET)	0.98248	0.96757	0.95846	0.00263	0.04191	0.01279
	LeukocyteMask(Aug-ET)	0.98947	0.97835	0.96471	0.00192	0.03540	0.01103
LISC	Watershed	0.81088	0.79061	0.74763	0.03738	0.21102	0.05426
	FCN	0.93018	0.91828	0.91827	0.00698	0.08280	0.01331
	U-Net	0.91661	0.93671	0.93557	0.00756	0.03848	0.01028
	LeukocyteMask(noAug)	0.98044	0.95956	0.95064	0.00137	0.06741	0.00869
	LeukocyteMask(Aug-noET)	0.98276	0.96562	0.95607	0.00121	0.05813	0.00791
	LeukocyteMask(Aug-ET)	0.98443	0.97428	0.95952	0.00117	0.05160	0.00784

Best results are marked in bold.

3.4.1 | Quantitative results

As shown in Figure 11 and Table 1, the proposed LeukocyteMask with Aug-ET provides the best results of almost all computed metrics on all 4 datasets, with very remarkable Precision, Dice, and mIoU, which are obviously higher than all other 5 methods. On the Dataset1, although, Watershed and FCN yield the better results than LeukocyteMask (Aug-ET) in terms of FPR and FNR (-0.032%, and -1.04%, respectively), the true cause for this is because the predictions conducted by those 2 methods are not closed enough to the real boundary of white blood cells, as shown in Figure 12(A) for a qualitative inspection. Watershed performs poor on segmenting cytoplasm of cells and in most cases only nucleuses are masked out. For FCN, there exists a relatively large gap between the prediction and the ground truth near the boundary of cell, in this case, a lower FNR is easier to obtained while with a high FPR. For BCISC and LISC, although a lower FNR is achieved by U-Net, the segmentation results are not stable enough, as shown in Figure 11 and Figure 12(B-C), for the highest k FNR (k = 3 in Figure 12) results, U-Net performs worse than LeukocyteMask in terms of robustness and compactness of segmentation for variant cells. As demonstrated in Figure 11, except above 2 mentioned special cases, it can be seen that, compared with the previous methods, LeukocyteMask improves the segmentation performance by a large margin, not only with a higher segmentation accuracy, but also a more stable performance on the different datasets.

3.4.2 | Qualitative results

Examples of ground truth and segmentation results of different methods on 3 datasets are shown in Figure 13. In Figure 13, the top 3 performance results and the bottom 3 performance results (measured by the average value of the

BIOPHOTONICS -



FIGURE 12 The highest 3 FPR and highest 3 FNR results for different methods. (A): Results on Dataset1, Row 1: LeukocyteMask (Aug-ET); Row 2: Watershed; Row 3: FCN. (B): Results on BCISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. (C): Results on LISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. (C): Results on LISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. (C): Results on LISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. (C): Results on LISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. (C): Results on LISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. (C): Results on LISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. (C): Results on LISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. (C): Results on LISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. (C): Results on LISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. (C): Results on LISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. (C): Results on LISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. (C): Results on LISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. (C): Results on LISC, Row 1: LeukocyteMask (Aug-ET); Row 2: U-Net. Columns 1 to 3 are the corresponding highest 3 FPR results and Columns 4 to 6 highest 3 FNR results. (Red Solid line indicates the predicted result and Blue Dashed line means the ground truth)



FIGURE 13 Segmentation results of different methods on 4 datasets. (A): Results on Dataset 1; (B): Results on Dataset 2; (C): Results on BCISC; (D): Results on LISC. For (A)-(D): Rows 1 to 4 are the predictions of Watershed, FCN, U-Net and the proposed LeukocyteMask(Aug-ET); Columns 1 to 3: Predictions of top 3 performance, Columns 4 to 6: Predictions of bottom 3 performance. (Red Solid line indicates the predicted result and Blue Dashed line means the ground truth)

BIOPHOTONICS

15 of 17



FIGURE 13 Continued

6 metrics) are displayed for segmentation accuracy and robustness comparison of different models. For watershed, marker-controlled watershed algorithm is not robust enough to mark the cytoplasm of the WBCs in different blood smear images and thus in many cases only nucleus regions are masked out in the final segmentation. For FCN and U-Net, the algorithms employ segmentation directly on the whole image, which is easily misled by the complex background such as RBCs and substantial dyeing impurities in the blood smear images, as shown in Figure 13, the predictions are interfered by RBCs and some extra dyeing impurities, which reduces the accuracy of the final segmentation results. Different from FCN and U-Net, the proposed LeukocyteMask carries out segmentation only on the ROIs localized by the RoIAlign layer, as demonstrated in Figure 2, which narrows the scope of segmentation, to improve the accuracy of final segmentation. Combining the results illustrated in Table 1 and Figure 13, it is obviously that the proposed LeukocyteMask leads to the most accurate and compact segmentation masks compared with above mentioned methods.

4 | CONCLUSION

In this paper, an end-to-end leukocyte (white blood cell, WBC) localization and segmentation method in blood smear images, called LeukocyteMask, is proposed, which segments WBC in a pixel-to-pixel manner automatically. Quantitative and qualitative results show that our proposed approach achieves significantly improvement to the state-of-the-art models, not only with a higher segmentation accuracy, but also a more stable performance. Future work would collect more images data and extend the current framework to evaluate it on more datasets which are captured under various imaging conditions. Furthermore, we would apply the proposed method to the automated cell analysis system for cell identification and counting.

ACKNOWLEDGMENTS

This work is partially supported by National Natural Science Foundation of China (61172168, 61772254 and 61866005), Natural Science Foundation of Heilongjiang Province (F2018019), Project for Innovative Talents in University of Heilongjiang Province of China (UNPYSCT-2015048), Guangxi Natural Science Foundation of China (2018GXNSFAA138017), Key Project of College Youth Natural Science Foundation of Fujian Province (JZ160467), Fujian Provincial Leading Project (2017H0030).

AUTHOR BIOGRAPHIES

Please see Supporting Information online.

ENDNOTES

¹https://github.com/fpklipic/BCISC.

²http://blog.cellavision.com/.

³https://github.com/keras-team/keras.

REFERENCES

- [1] F. Xing, L. Yang, IEEE Rev Biomed Eng 2016, 9, 234.
- [2] J. Wen, Y. Xu, Z. Li, Z. Ma, Y. Xu, Neu Networks. 2018, 102, 36.
- [3] N. Theera-Umpon, White blood cell segmentation and classification in microscopic bone marrow images. in *International Conference on Fuzzy Systems and Knowledge Discovery*, Springer, Berlin, Heidelberg 2005, p. 787.
- [4] C. Zhang, X. Xiao, X. Li, Y. J. Chen, W. Zhen, J. Chang, C. Zheng, Z. Liu, *Sensors.* 2014, 14, 16128.
- [5] L. Zhi, L. Jing, X. Xiao, Y. Hui, X. Li, J. Chang, C. Zheng, Sensors. 2015, 15, 22561.
- [6] A. Gençtav, S. Aksoy, S. Önder, Pattern Recogn 2012, 45, 4151.
- [7] D. C. Huang, K. D. Hung, Y. K. Chan, J Syst Software. 2012, 85, 2104.
- [8] S. Arslan, E. Ozyurek, C. Gunduzdemir, Cytometry 2014, 85, 480.
- [9] X. Zheng, Y. Wang, G. Wang, Z. Chen, Micron 2014, 56, 17.
- [10] G. Liu, J. Yang, IEEE Trans Image Process 2019, 28, 6.
- [11] G.-H. Liu, J.-Y. Yang, Z. Li, Pattern Recogn 2015, 48, 2554.
- [12] Z. Li, G. Liu, D. Zhang, Y. Xu, Pattern Recogn 2016, 52, 317.
- [13] F. Yi, Z. Chongxun, P. Chen, L. Li, *Proc IEEE Conf Eng Med Biol*, IEEE Computer Society, Shanghai 2005, p. 6476.
- [14] H. Kong, M. Gurcan, K. Belkacem-Boussaid, *IEEE Trans. Med. Imaging* 2011, 30, 1661.
- [15] Y. Song, F. Weidong, D. Dagan, H. Huang, Y. Wang, BMC Bioinformatics 2013, 14, 173.
- [16] C. D. Ruberto, A. Loddo, L. Putzu, Mach Vis Appl. 2016, 27, 1.
- [17] C. Pan, S. P. Dong, S. Yoon, J. C. Yang, Expert Syst Appl. 2011, 38, 7479.
- [18] C. Pan, H. Lu, F. Cao, Segmentation of Blood and Bone Marrow Cell Images via Learning by Sampling, Springer, Berlin, Heidelberg 2009, p. 336.
- [19] X. Zheng, Y. Wang, G. Wang, J. Liu, Micron 2018, 107, 55.
- [20] J. Prinyakupt, C. Pluempitiwiriyawej, *Biomedl Eng* 2014, 20 (1), 1.
- [21] O. Sertel, J. Kong, U. V. Catalyurek, G. Lozanski, J. H. Saltz, M. N. Gurcan, J Signal Processing Systems. 2008, 55, 169.
- [22] M. Saidi, M. E. A. Bechar, N. Settouti, M. A. Chikh, *MedPRAI*, ACM, Tebessa 2016, p. 31.
- [23] M. Farhan, P. Ruusuvuori, O. Yli-Harja, M. Emmenlauer, P. Raemoe, *BMC Bioinformatics*. 2013, 14, 1.
- [24] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, CVPR, Las Vegas, NV 2016, p. 770.
- [25] T.-Y. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, S. J. Belongie, *Feature pyramid networks for object detection*, *CVPR*, IEEE Computer Society, Hawaii 2017, p. 936.
- [26] F. Shen, Y. Xu, L. Liu, Y. Yang, Z. Huang, H. T. Shen, *IEEE Trans. Pattern Anal. Mach. Intell.* 2018, 40, 3034.
- [27] W. Yang, J. Li, H. Zheng, R. Y. Da Xu, *IEEE Access.* 2018, 6, 7445.

BIOPHOTONICS 17 of 17

- [28] E. Shelhamer, J. Long, T. Darrell, *IEEE Trans. Pattern Anal.* Mach. Intell. 2017, 39, 640.
- [29] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, MICCAI, Springer International Publishing, Munich, 2015, p. 234.
- [30] L. Zhang, M. Sonka, L. Lu, R. M. Summers, J. Yao, *ISBI*, IEEE Computer Society, Melbourne, 2017, p. 234.
- [31] L. Duan, X. Qin, Y. He, X. Sang, J. Pan, T. Xu, J. Men, R. E. Tanzi, A. Li, Y. Ma, J. Biophotonics 2018, 11, e201800146.
- [32] G. P. Gopakumar, M. Swetha, G. S. Siva, G. R. K. S. Subrahmanyam, *J. Biophotonics* 2017, *11*, e201700003.
- [33] T. Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature Pyramid Networks for Object Detection, *CVPR*, IEEE Computer Society, Hawaii 2017, p. 2117.
- [34] K. He, G. Gkioxari, P. Dollár, R. Girshick, *Mask R-CNN, ICCV*, IEEE Computer Society, Venice 2017, p. 2980.
- [35] P. Y. Simard, D. Steinkraus, J. C. Platt, Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis, ICDAR, IEEE Computer Society, Edinburgh, 2003, p. 958.
- [36] G. E. HintonA. Krizhevsky, I. Sutskever, Imagenet Classification with Deep Convolutional Neural Networks NIPS, Curran Associates, Nevada 2012, p. 1097.
- [37] K. Simonyan, A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv: 1409.1556. 2014.
- [38] S. Ioffe, C. Szegedy, Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, ICML, JMLR. org, Lille 2015, p. 448.

- [39] V. Nair, G. E. Hinton, Rectified Linear Units Improve Restricted Boltzmann Machines, ICML. Omnipress, Haifa 2010, p. 807.
- [40] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks NIPS, Curran Associates, Quebec 2015, p. 91.
- [41] R. Girshick, Fast R-CNN, ICCV, IEEE Computer Society Santiago 2015, p. 1440.
- [42] S. H. Rezatofighi, H. Soltanian-Zadeh, Comput. Med. Imaging Graph. 2011, 35, 333.
- [43] V. K. Singh, S. Romani, H. A. Rashwan, F. Akram, N. Pandey, M. M. K. Sarker, S. Abdulwahab, J. Torrents-Barrena, A. Saleh, M. Arquez, M. Arenas, D. Puig, Conditional Generative Adversarial and Convolutional Networks for X-ray Breast Mass Segmentation and Shape Classification, MICCAI, Springer International Publishing, Granada, 2018, p. 833.

How to cite this article: Fan H, Zhang F, Xi L, Li Z, Liu G, Xu Y. LeukocyteMask: An automated localization and segmentation method for leukocyte in blood smear images using deep neural networks. *J. Biophotonics*. 2019;e201800488. <u>https://doi.org/10.1002/jbio.201800488</u>