Ye Wang^{1†}, Rui Ma^{2,4†}, Xiaoqing Ma¹, Honghua Cui³, Yubin Xiao¹, Xuan Wu¹ and You Zhou^{1*}

^{1*}School of Computer Science and Technology, Jilin University, Changchun, 130012, Jilin Province, China.

²School of Artificial Intelligence, Jilin University, Changchun, 130012, Jilin Province, China.

³Second Hospital of Jilin University, Changchun, 130041, Jilin Province, China.

⁴Engineering Research Center of Knowledge-Driven Human-Machine Intelligence, MOE, Changchun, 130012, Jilin Province, China.

*Corresponding author(s). E-mail(s): zyou@jlu.edu.cn; Contributing authors: wangyejlu1996@gmail.com; ruim@jlu.edu.cn; maxq20@mails.jlu.edu.cn; chh1214@126.com; xiaoyb@smail.xtu.edu.cn; wuuu20@mails.jlu.edu.cn; †These authors contributed equally to this work.

Abstract

Fine-grained classification and counting of bone marrow erythroid cells are vital for evaluating the health status and formulating therapeutic schedules for leukemia or hematopathy. Due to the subtle visual differences between different types of erythroid cells, it is challenging to apply existing image-based deep learning models for fine-grained erythroid cell classification. Moreover, there is no large open-source datasets on erythroid cells to support the model training. In this paper, we introduce BMEC (Bone Morrow Erythroid Cells), the first large finegrained image dataset of erythroid cells, to facilitate more deep learning research on erythroid cells. BMEC contains 5,666 images of individual erythroid cells, each of which is extracted from the bone marrow erythroid cell smears and professionally annotated to one of the four types of erythroid cells. To distinguish the erythroid cells, one key indicator

is the cell *shape* which is closely related to the cell growth and maturation. Therefore, we design a novel shape-aware image classification network for fine-grained erythroid cell classification. The shape feature is extracted from the shape mask image and aggregated to the raw image feature with a shape attention module. With the shape-attended image feature, our network achieved superior classification performance (81.12% top-1 accuracy) on the BMEC dataset comparing to the baseline methods. Ablation studies also demonstrate the effectiveness of incorporating the shape information for the fine-grained cell classification. To further verify the generalizability of our method, we tested our network on two additional public white blood cells (WBC) datasets and the results show our shape-aware method can generally outperform recent state-of-the-art works on classifying the WBC. The code and BMEC dataset can be found on https://github.com/wangye8899/BMEC.

Keywords: Bone Marrow Erythroid Cell, Fine-Grained Cell Classification, Shape Attention, Feature Fusion

1 Introduction

The erythrocytes or red blood cells are one of the most critical cells in the body. There are four types of erythroid cells (Figure 1) based on their growth and maturation. In a healthy body, they maintain a relatively balanced state. However, in many blood disorders such as anemia, the number of these erythroid cells will change and become unbalanced. Therefore, the *fine-grained* classification and counting of different types of erythroid cells are vital in diagnosing and preventing the related diseases [1].

One typical way for cell classification and counting is to manually observe the morphological or shape differences of the cells through the microscope. Such process relies heavily on the hematologists' experience and skills, and it is easy to make mistakes and introduce subjectivity to the results. Furthermore, it may take several years to train a competent hematologist to distinguish the specific type of erythroid cells. In recent years, several image processing and machine learning techniques for classifying and counting erythroid cells have been proposed. Most existing methods [2–4] mainly follow the traditional machine learning pipeline, including data prep-processing, feature extraction, feature selection, and classification steps. Since these methods rely on manually designed features, the generalizability of the classification model may be limited. In addition, the morphological and shape differences are not fully and deeply exploited in the model. Even though satisfied results have been achieved for classifying white blood cells or certain abnormal types of erythroid cells whose shapes are visibly distinctive, traditional methods may not perform well for fine-grained classification of erythroid cells with only subtle shape differences.

On the other hand, deep learning techniques have shown superior performance than traditional machine learning methods in computer vision tasks such as image classification [5], object detection [6] and segmentation [7]. Deep neural networks can automatically extract deep image features which are more robust and more generalizable than hand-crafted features. However, most pre-trained network models [5, 8–11] cannot directly learn the subtle shape differences which are crucial for fine-grained classification of erythroid cells. To enable fine-grained erythroid cell classification, a more customized and shapeaware network needs to be designed and trained on erythroid cell images with finer-level category annotation. Unfortunately, unlike the leukocytes or white blood cells (WBC), for which many open-source datasets are available (e.g., LISC [12], BCCD [13] and Raabin-WBC [14]), there is no existing open-source fine-grained datasets to support the deep learning on erythroid cells.

In this paper, we introduce BMEC (Bone Morrow Erythroid Cells), a large fine-grained image dataset of erythroid cells, to facilitate the study for erythroid cells in real clinical scenarios. BMEC dataset contains 5,666 images of individual erythroid cells. Each image contains one erythroid cell extracted from the bone marrow erythroid cell smears and is annotated to one of the four types of erythroid cells by the hematologists (Figure 1). To classify the erythroid cells into fine-grained categories, we propose a novel *shape-aware* image classification model which explicitly encode the shape information into the network (Figure 2). Specifically, we first extract the cell shape mask image from the input erythroid cell image. Then, the shape mask feature and the input image feature are extracted using the existing backbone image feature extraction models. A shape attention module is further employed to compute a shape-attended image feature, which is aggregated to the shape feature and the raw image feature to form a fused feature for final classification.

We conducted extensive quantitative evaluations of different backbone models for our shape-aware network on BMEC and our final model with Swin-Transformer [5] and VGG19 [8] achieved 81.12% top-1 accuracy which is consistently superior than other backbone model combinations. We also performed ablation studies to verify the effectiveness of explicitly using shape information for fine-grained erythroid cell classification. Furthermore, we tested our shape-aware classification network on two public leukocytes or WBC datasets and the results show our method can outperform recent state-ofthe-art work on classifying the WBC. In summary, our contributions are as follows:

• We introduce BMEC, the first large dataset of erythroid cell images with fine-grained category annotation, to support more deep learning research on erythroid cells. This dataset contains 5,666 professionally annotated images of four types of erythroid cells, while there are only subtle shape differences between these fine-grained types, making our dataset more chanllenging than existing WBC datasets.

Dataset	Quantity	Cell Type	Classes	Label	Microscope
LISC [12] BCCD [13] Raabin-WBC[14] CBC [15] erythrocytesIDB [16] BMEC (ours)	$250 \\ 349 \\ 17,965 \\ 360 \\ 629 \\ 5,666$	WBC WBC WBC Erythroid cells Erythroid cells Erythroid cells	$5 \\ 5 \\ 3 \\ 3 \\ 4$	one expert one expert two experts two experts one expert three experts	Axioskope40 Regular light micro Olympus Cx18 ECLIPSE 50i micro Leika micro Olympus BX43

Table 1 Comparison of related WBC and erythroid cell datasets with our BMEC.

- We design a novel shape-aware network which employs a shape attention module to compute a shape-attended image feature and uses the shapeaggregated feature to improve the fine-grained classification accuracy of erythroid cells. To our best knowledge, we are the first to explicitly incorporate the shape information to the deep learning framework for this task.
- We perform extensive evaluations of our network with different backbone models and conduct ablation studies to verify the effectiveness of shape information. The results on BMEC and two other WBC datasets show the superiority of the proposed shape-aware network and models in classifying the blood cell images.

2 Related Work

In this section, we first discuss several datasets on blood cells, including both leukocytes and erythrocytes, with the emphasis on the data characteristics and how they are collected; see Table 1. We then provide a brief review of the traditional and deep learning techniques related to blood cell classification and counting.

2.1 Blood Cell Datasets

WBC datasets. There are many available white blood cell image datasets for investigating the WBC classification and counting tasks [12–14, 17–19]. Here, we only describe three representative WBC datasets and highlight their difference with our BMEC; see Table 1. The LISC [12] is an early WBC dataset that includes the hematological images taken from peripheral blood of healthy subjects. The smears were stained by through the Gismo-right technique and a Sony Model No.SSCDC50AP camera was used to capture the observation of a Axioskope 40 microscope at 100X magnification. Then, 250 WBC images were extracted and labelled by one blood expert into five fine-grained categories. Similarly, BCCD [13] contains 349 WBC images that have been taken from peripheral blood and annotated by one expert. The smears were also stained by the Gismo-right technique and the observations of a regular light microscope at 100X were captured using a CCD color camera. Each WBC image categories. Raabin-WBC [14] is a recent, large and public dataset of white blood cell images extracted from the normal peripheral blood samples. The peripheral blood smears were stained by the Giemsa technique and the images were taken using a Samsung Galaxy S5 camera to capture the observation of an Olympus CX18 microscope at 100X magnification. In total, 17,965 WBCs images are cropped from the smear images and labelled into five categories by two blood experts. Comparing to the white blood cells, different types of erythroid cells look more similar and the shape difference between the types is subtle. Therefore, it is more challenging for fine-grained erythroid cell classification, e.g., over 99% accuracy can be achieved on Raabin-WBC while only around 80% accuracy can be obtained on our BMEC. Due to the difficulty on distinguishing the erythroid cell types, we asked three hematologists to jointly label the BMEC images in a more rigorous manner.

Erythroid cell datasets. Compared to the WBC datasets, there are only a few open-source datasets containing the erythrocytes, while the data quality, quantity and variety are limited comparing to our BMEC. As shown in Table 1. CBC [15] is a hybrid dataset which contains three types of blood cells: red blood cells, white blood cells, and platelets. The images in CBC were taken using a Nikon V1 camera to capture the observation of an ECLIPSE 50i microscope at 100X magnification, and then annotated by two experts. It focuses more on investigating the differences between the three types of blood cells rather than classifying red blood cells into fine-grained types. In addition, CBC only has 360 images which are difficult for training the deep neural networks. One most related dataset on erythroid cell classification is the erythrocytesIDB [16] which contains 629 images of individual ervthroid cells extracted from 196 full filed peripheral blood images. A Kodak EasyShare V803 camera was used to capture the observation of a Leika microscope at 100X magnification. The erythrocytesIDB images are annotated by one expert into three categories based on the cell shape: circular, elongated and other shape. Comparing to BMEC, the shape differences for erythroid cells in erythrocytesIDB are more significant since they focus on the study in cell deformation caused by certain blood diseases. Hence, it's much easier to achieve high accuracy (98% in [16])on erythrocytesIDB. In contrast, our BMEC provides a much larger number of images with minor shape differences, so that it can be used to train a more generalizable deep learning model for fine-grained erythroid cell classification.

2.2 Cell Classification Techniques

Along with the release of more open-source blood cell datasets, automatic cell classification and counting techniques have been proposed, including the traditional image processing and machine learning based methods, as well as the deep learning based frameworks.

Traditional methods. Generally, cell classification using traditional image processing and machine learning techniques involves image segmentation, feature extraction, classification, and counting steps. For example,

Alomari et al. [2] employed the Circular Hough Transform (CHT) in an iterative circle detection framework to segment, classify and count WBCs and RBCs. Such hough transform based shape detection methods are usually timeconsuming and their detection and classification accuracy may be significantly reduced when the shape of the blood cells becomes more complex. Meanwhile, machine learning based methods aim to classify the blood cells based on some features (e.g., shape and color) extracted from the cell images. Lippeveld et al. [3] applied the traditional machine learning techniques such as random forest [20] and gradient boosting classifier [21] to classify and count the stain-free WBCs images based on the size, location, texture, and signal strength feature. Tavakoli et al. [22] proposed a new segmentation and feature extraction algorithm which computes three shape features and four novel color features for classifying WBC images. Although above traditional methods have achieved appealing results for WBC or RBC classification on certain datasets such as Raabin-WBC and LISC, they heavily rely on hand-crafted features which are hardly generalizable to more challenging tasks, e.g., fine-grained classification on our BMEC.

Deep learning based cell classification. Deep learning can automatically extract features by learning from a large amount of data to compute tasks such as classification [23, 24] and detection [25, 26]. To apply deep learning models for the cell classification, one way is to use models pre-trained on general large image datasets for feature extraction and perform particular traditional feature selection and classification methods on the smaller-scale WBC or RBC dataset. In [27], features extracted from pre-trained AlexNet [28], GoogLeNet [29] and ResNet50 [9] models are filtered using the Maximal Information Coefficient and Ridge feature selection methods, and quadratic discriminant analysis was used as a classifier on a dataset of WBCs. Sahlol et al. [30] employed a statistically enhanced Salp Swarm Algorithm[31] to filter the features extracted from WBC images using the VGG [8] model pre-trained on ImageNet and performed the classification using a traditional decision tree.

The other way to apply deep learning for cell classification is using the transfer learning technique which starts from a pre-trained CNN model and performs model fine-tuning on a specific WBC or RBC dataset. For example, Alzubaidi et al. [32] first trained a CNN model on a combination of three WBC datasets and then employed transfer learning to classify RBC images in erythrocytesIDB [16]. Pasupa et al. [33] exploited pre-trained ResNet [9] and DenseNet [10] and performed fine-tuning on their RBC images. In addition, they have handled the class imbalance problem by incorporating the focal loss [34] . Comparing to traditional machine learning techniques, deep learning based classification methods can achieve superior performance without manual feature design. However, it is difficult to either train a traditional classifier or perform transfer learning for the fine-grained classification task, especially when there is no sufficient finely annotated training data. We propose a novel



Fig. 1 Full-field erythroid cell image and four different types of erythroid cells in BMEC cropped by manually annotated ROI: proerythroblast (Pro), basophilic erythroblast (Bas), polychromatophilic erythroblast (Pol), and Orthochromic erythroblast (Ort).

shape-aware transfer learning framework that utilizes different types of pretrained backbone models and fine-tunes the network on our large-scale BMEC for fine-grained erythroid cell classification.

Shape-aware cell classification. The cell's morphological or shape characteristics such as the circularity, convexity and size have been investigated in traditional methods [2, 3, 22] for cell classification. For deep learning based methods, given only the original cell images, the networks can automatically learn some shape features [35] which are generally more robust than the handcrafted ones. However, the implicit learning of deep shape features requires a large amount of training data. Even when transfer learning is used, it may be still difficult for the network to learn the fine-grained shape variations. To explicitly incorporate the shape information in the network, Tavakoli et al. [36] passed two binary shape images along with the RGB images of WBCs to a customized CNN with five channels (three for RGB and two for shape masks) and showed the generalizability of the model can be improved when the shape masks are used. Similarly, we also explicitly integrate the shape information into the network. Instead of using the nucleus and ROC (Representative Of the Convex hull = convex hull - nucleus) images as in [36], we choose convex hull as our shape mask image since we find the contour is more suitable to represent the shape of erythroid cells. Moreover, we design a novel shape attention module which computes a shape-attended image feature, to improve the learning of fine-grained shape-aware image features.

3 BMEC Dataset

Our BMEC dataset was built based on the data acquired at the Second Hospital of Jilin University, Changchun, China. It contains 5,666 erythroid cell images which are annotated into four fine-grained types by three hematologists. The shape differences between the types are subtle, which makes the classification on BMEC more challenging than existing WBC datasets or RBC datasets with abnormal erythroid cells.

Cells	$\mathbf{Percent}^1$	Diameter	Shape
Pro	$\begin{array}{c} 0.5\%\text{-}4\%\\ 1\%\text{-}5\%\\ 12\%\text{-}20\%\\ 6\%\text{-}12\%\end{array}$	15um-20um	elliptical, protrusions
Bas		10um-18um	round, protrusions
Pol		8um-15um	round, rough edges
Ort		7um-10um	round, smooth edges

 Table 2
 The characteristics of different erythroid cells

 summarized in [37].

¹The Percent column indicates the percentage of the corresponding cell type in all types of bone marrow cells, including both WBC and erythroid.

Table 3 The statistics of the BMEC dataset in terms of distributions of different cells.

Dataset	Quantity	Percent	Pro^{1}	Bas^1	Pol^1	Ort^1
Training Set.	3,865	68.2%	146(2.6%)	428(7.6%)	1,702(30.0%)	1,589(28.0%)
Validation Set	655	11.6%	28(0.5%)	69(1.2%)	312(5.5%)	246(4.4%)
Test Set	1,146	20.2%	48(0.8%)	115(2.0%)	526(9.3%)	457(8.1%)
Total	$5,\!666$	100%	222(3.9%)	612(10.8%)	2,540(44.8%)	2,292(40.5%)

¹proerythroblast cells (Pro), basophilic erythroblast cells (Bas), polychromatophilic erythroblast cells (Pol), orthochromic erythrobla cells (Ort).

Smears Imaging and Cell Extraction. The original data contains 239 sets of bone marrow erythroid cell smears collected from 128 patients using the BEION V4.90 system between 2019 and 2021. To meet the requirements of clinical evaluation and diagnosis, each smear contains more than 200 blood cells. In the collection process, the hematologist first stained all smears using the Richter stain method. Then, the observation of an Olympus BX43 microscope at 100X magnification was captured into images of 2592 \times 1944 resolution.

To extract individual erythroid cells from the smear images, two experienced hematologists worked together to select a ROI (region of interest) for each cell and crop the selections into cell images (Figure 1). When extracting the cells, the hematologists mainly follow two principles: 1) not introduce too much background and noise; 2) preserve the whole cell shape to the maximum extent. When one hematologist completed the cropping, the other hematologist examined the results to ensure the quality of the extracted images, e.g., no other cells rather except the central erythroid cell appear in the image. In the end, 5,666 erythroid cell images are extracted from the blood smears. Each cell image is resized to 224×224 resolution.

Data Annotation. Based on the growth and maturation, the erythroid cells are annotated into four types (Figure 1 right): proerythroblast (Pro), basophilic erythroblast (Bas), polychromatophilic erythroblast (Pol), and orthochromic erythrobla (Ort). Note that we only focus on the erythroid cells which have normal shapes and do not consider the abnormal erythroid cells with elongated shapes such as in erythrocytesIDB [16]. For reference, in Table

2, we show the characteristics (e.g., percentage, size and shape) summarized in [37], for the four types of erythroid cells.

As the shape differences of the collected erythroid cells are indeed very subtle, to reduce the subjectiveness in annotation, we invited three hematologists and designed a rigorous annotation process to obtain more consistent labels. First, two hematologists independently annotated all the cell images without any interference. The annotation results of the two hematologists were compared and the cell images with different assigned labels were re-annotated by the two hematologists for the second pass. If the re-annotation results were still different, the third hematologist would provide another annotation on these images, and the final cell category would be determined by voting the annotations from the three hematologists.

Statistics. As shown in Table 3, we divide the BMEC dataset into a training set, a validation set and a test set with an approximate ratio of 7:1:2. It can be observed that for each set, the percentage of the four cell types follows the relative cell distribution mentioned in Table 2, e.g., the number of Pol is about four times of Bas.

In the following, we summarize the key characteristics that discriminate our BMEC dataset from other blood cell datasets:

- The first fine-grained erythroid cell dataset. BMEC dataset is the first dataset that contains four fine-grained types of normal erythroid cells: Pro, Bas, Pol, and Ort. It is more challenging than existing WBC or erythroid cell datasets since the shape differences between the types are subtle.
- Largest number of erythroid cells. As shown in Table 1, BMEC is currently the largest dataset for erythroid cells, which is expected to enable more deep learning research on erythroid cells.
- **Professional and rigorous annotation**. The images in the BMEC dataset were professionally annotated by three hematologists following a rigorous labeling process, which reduces the subjectiveness and ensures the quality of the annotation.
- Free public access. We make the BMEC dataset freely available for all communities, with the hope that it can inspire more follow-ups research on erythroid cells.

4 Method

4.1 Overview

Our pipeline consists of four modules as shown in Figure 2: a cell shape extraction module, a dual-branch network, a shape attention module and a cell classification network. The shape extraction module transforms the RGB erythroid cell image into the binary shape mask. The RGB image and shape mask are separately passed to the dual-branch network to generate the raw and shape features. The shape attention module calculates the shape attention weights and generates a shape-attended image feature from the raw image



Fig. 2 The overview of our approach. The novel shape attention module calculates a shapeattended image feature using the raw and shape features (\otimes represents the broadcasted element-wise matrix multiplication operation). The fused shape-aware feature is used for the final fine-grained classification.

Algorithm 1 The cell segmentation pipeline following [22]

- 1: Input RGB erythroid cell image
- 2: Convert the input RGB image to CMYK color space
- 3: Calculate $KM = (K_component) (M_component)$
- 4: Convert the RGB input image to HLS color space and calculate MS = Min(M_component, S_component)
- 5: Calculate V = MS KM
- 6: Exploit Otsu's thresholding algorithm [38] to segment the cell and obtain the cell nucleus ${\cal N}$
- 7: Calculate the convex hull of the cell to acquire the shape mask S based on N
- s: Calculate the cell cytoplasm C = S N
- 9: Output the convex hull S, nucleus N, cytoplasm C as the binary segmentation masks

feature. Then, the raw image feature, shape-attended image feature and the shape feature are aggregated to a fused shape-aware feature, which is passed to the lightweight classification network for prediction.

4.2 Cell Shape Extraction

From Figure 1 and Table 2, it can be observed the shape information is crucial to distinguish different fine-grained types of erythroid cells, especially when the appearances of the cells are similar. Therefore, we explicitly extract the cell shape information as a binary shape mask and treat it as a prior knowledge (or inductive bias) for the classification model. We employ the cell image segmentation method proposed in [22] to extract the shape mask for each BMEC cell image. The main steps of cell shape extraction are shown in Algorithm 1, while Figure 3 shows the corresponding intermediate and segmentation results. Three types of binary mask images are obtained: the convex hull, nucleus and cytoplasm. Figure 4 shows the image segmentation results of different types of erythroid cells.





Fig. 4 This figure shows the RGB images and other segmentation results (Nucleus mask, Shape mask, and Cytoplasm mask) of four types of erythroid cells.

Unlike [36] which utilized the nucleus and ROC (convex hull - nucleus) images as the shape images, we choose the convex hull image as our shape image. The main reason is [36] mainly focuses on the WBCs whose nucleuses are in apparently different shapes. However, for erythroid cells, their nucleuses are more intact and look quite similar comparing to WBCs (see Figure 4). Also, the cytoplasm (i.e., the ROC in [36]) images of the erythroid cells are noisy and not suitable to represent the cell shapes. In contrast, convex hull contains the key shape characteristics and it is more robust to noise comparing to the nucleus. We pass the binary image of convex hull to our network and also compute the shape attention to enhance the original image feature.

4.3 Dual-Branch Network

To explicitly integrate the shape information into the deep learning framework, we design a dual-branch network which encodes the RGB image and the binary shape mask (convex hull image) in two separate branches.

As shown in Figure 2, the top branch network takes the RGB image as input and extracts the raw feature using Swin-Transformer [5], the state-of-the-art backbone model for image feature extraction. Although the RGB image branch can learn the pixel intensity variations between different cells, it does not focus on the cell's shape or structure characteristics. Therefore, we introduce a separate bottom branch network which employs a VGG19 [9] model to explicitly compute the shape feature from the binary shape mask image obtained from the cell shape extraction module. The raw image feature and the shape feature from two branches are integrated in the shape attention module (Section 4.4) and fused in the later stage. In Section 5.4, we show the shape branch is an effective complement to the RGB branch for fine-grained cell classification. In addition, our dual-branch network is a general design such that the backbone models in each branch can be replaced by other deep image feature

extraction models. In current our experiments, we find using the combination of Swin-Transformer and VGG19 for the two branches can achieve the best performance (Table 4 and 5).

4.4 Shape Attention

To further integrate the shape information to the extracted image feature, we propose a novel shape attention module which takes the raw image feature and shape feature as input and outputs a shape-attended image feature (see Figure 2 middle). First, the shape attention weights are computed from the binary shape mask image following the similar attention mechanism used in SENet [39]. Then, the raw image feature is multiplied by the shape attention weights to get the shape-attended feature. The mathematical formulation is as follows:

$$F_R = Swin - Transformer(X_R),\tag{1}$$

$$F_S = VGG19(X_S),\tag{2}$$

$$W_{SA} = Sigmoid(W_2(ReLU(W_1(F_S))))), \qquad (3)$$

$$F_{SA} = W_{SA} \otimes F_R. \tag{4}$$

Here, $X_R \in R^{224 \times 224 \times 3}$ and $X_S \in R^{224 \times 224 \times 1}$ are the input RGB image and binary shape mask, respectively; F_R and F_S are the corresponding features extracted using Swin-Transformer [5] and VGG19 [8]. As shown in Equation 3 and Figure 2, the shape attention weights W_{SA} are calculated by passing F_S through two fully connected layers (W_1 and W_2) which are followed by a ReLU and a Sigmoid layer, respectively. Finally, the shape-attended feature F_{SA} is obtained by performing element-wise multiplication of the shape attention weights W_{SA} and the raw image feature F_R . In summary, the shape attention module is a lightweight network that can efficiently computes the correlation among the shape feature and enhances the raw image feature with self-learned cell shape information. Moreover, the shape attention module can be combined with any backbone models to integrate the shape information to the image feature.

4.5 Cell Classification Network

The last module of our pipeline is a conventional classification network which consists of two sets of fully connected, ReLu and Dropout layers (Figure 2 right). The raw image feature F_R , shape feature F_S and shape-attended image feature F_{SA} are concatenated to form an fused cell feature $F_C = \{F_R, F_{SA}, F_S\}$ for final classification. With the shape feature and shape attention explicitly considered, the shape-aware network can achieve superior performance comparing to only using the RGB feature for the fine-grained erythroid cell classification (see Table 8).

5 Experiments

In this section, we first provide the implementation details of our method. Then, we qualitatively evaluate the performance of different backbones in our dual-branch network and show the combination of Swin-Transformer (for RGB image) and VGG19 (for shape mask) can work best on our fine-grained BMEC dataset. To further verify the effectiveness and generalizability of our shape-aware network, we compare the results on two WBC datasets and show our method can consistently achieve the best accuracy on different datasets. In the end, we perform an ablation study to show the contribution of the shape attention module for improving the classification accuracy. The code and BMEC dataset can be found on https://github.com/wangye8899/BMEC.

5.1 Implementation Details

The network architecture in Figure 2 is implemented with the Timm [40] and PyTorch [41]. Both branch networks (Swin-Transformer, VGG19) are initialized with ImageNet pre-trained weights. Furthermore, the shape attention module and cell classification network are initialized with Kaiming initialization [42]. We adopt the SGD optimizer with a momentum of 0.9, weight decay of 2e-5 and a cosine lr schedule with the warm-up strategy [43]. We set the warm-up learning rate to 1e-4, the maximum lr to 1e-3, and the minimum lr to 1e-4. Four data augmentation techniques are used: random crop, random horizontal flip with probability of 0.5, random vertical flip with probability of 0.5, and color jitter with factor of 0.4. The network is trained for 300 epochs with batch size of 32 on a cloud-based server with a 4.90GHz CPU and three NVIDIA RTX 3090 24G GPUs and the total training takes 3 hours. During inference, the computation time is instantaneous which is similar to other deep learning based classification methods.

5.2 Classification Results on BMEC

Our dual-branch network allows flexible choices of different backbone models for the cell classification. To quantitatively evaluate the performance of different backbones for the two branches, we conducted classification experiments with various classic and SOTA image feature extraction models on our BMEC dataset and compared the results in Table 4 and 5. Since the RGB image in the top branch contains more color variations than the binary shape mask in the bottom branch, it is expected that a more complex backbone model is needed for the top branch, while a relative simple model may be sufficient for the bottom branch. Therefore, in Table 4, we fixed the bottom branch with a VGG19 which is a classic CNN based model and tested different backbones for the top branch. It is shown the SOTA Swin-Transformer model can produce the best accuracy and F1 score comparing to other models such as ResNet [9], DenseNet [10] or ResNest [11] for extracting features from the RGB images. Similarly, we fixed the top branch network as Swin-Transformer and evaluated different models for the bottom branch. The results in Table 5 shows the

	D	· .		5	5	
Тор	Bottom	Inputs	Acc	Pre	Rec	F.1
ResNet18 [9]		RGB+Shape	79.51%	74.27%	73.53%	72.42%
ResNet50 [9]		RGB+Shape	79.58%	74.33%	73.67%	72.51%
DenseNet121 [10]	VGG19 [8]	RGB+Shape	79.65%	73.42%	73.15%	72.20%
ResNest [11]		RGB+Shape	80.12%	76.47%	74.64%	73.04%
Swin-T $[5]^1$		RGB+Shape	81.12%	76.28%	75.73%	74.00%

Table 4 Comparison of different top branch backbone networks for classification onBMEC. The bottom branch network is fixed as VGG19.

¹Swin-Transformer (Swin-T)

Table 5 Comparison of different bottom branch backbone networks for classification onBMEC. The top branch network is fixed as Swin-Transformer.

Тор	Bottom	Inputs	Acc	Pre	Rec	F1
Swin-T [5] ¹	ResNet18 [9]	RGB+Shape	80.62%	78.68%	74.93%	73.23%
	ResNet50 [9]	RGB+Shape	80.45%	75.44%	76.21%	72.68%
	Swin-T [5] ¹	RGB+Shape	80.36%	77.45%	75.81%	73.33%
	VGG19 [8]	RGB+Shape	81.12%	76.28%	75.73%	74.00%

¹Swin-Transformer (Swin-T)

VGG19 model is most suitable for the bottom branch, while one possible reason is that other models may underfit or overfit the binary shape mask images. Based on above experiments, we finally choose Swin-Transformer and VGG19 for the top and bottom branches, and our final model achieved 81.12% accuracy on BMEC. Note that the backbones can always be easily evaluated and replaced when newer and more powerful image models are proposed.

5.3 Classification Results on WBCs Datasets

To further demonstrate the effectiveness of our shape-aware network, we conducted classification experiments on two other public WBCs datasets, LISC [12] and Raabin-WBC [14]. We retrained our model using their training data and compared the testing results with the reported numbers in [12] and [14]. As shown in Table 6 and 7, we achieved the best classification accuracy on both datasets, showing the generalizability of our method. In addition, the high accuracy (over 98.5%) on both WBC datasets indicates the classification of WBCs is relatively easier than the erythroid cells since the shape differences of WBCs are more prominent. We hope our BMEC dataset can inspire and facilitate more research on fine-grained classification and other learning tasks for erythroid cells.

5.4 Ablation Study on Shape information

To verify the effectiveness of the shape information, we performed an ablation study which compared our full dual-branch network with the RGB-only single branch network (see Table 8). Specifically, we disabled the bottom branch and

Table 6 Comparison of different methods for classification on LISC dataset [12]. Due to the data imbalance problem in LISC, other methods directly discarded the cell categories with a low number of images. In contrast, we trained and tested using all categories and still obtained the best accuracy.

Study	Method	Category Number	Accuracy
Rezatofighi et al. (2011) [12]	SVM	5	96.00%
Jung et al. (2019) [44]	CNN	5	97.00%
Baydilli et al. (2020) [45]	Capsule Networks	5	96.86%
Harshanand et al. (2020) [46]	CNN	5	97.64%
Tavakoli et al. (2021) [22]	Segmentation	5	92.21%
Muhammad et al. (2021) [47]	Feature Selection and ELM	4	96.60%
Ours	DB and SA^1	6	98.51%

¹Dual Branch and Shape Attention (DB and SA)

Tał	ole 7	Comparison of different	methods for	classification	on Raabin-	WBC dataset $[14]$.
	-					

Study	Method	Category Number	Accuracy
Kouzeh et al. (2021) [14]	VGG16	6	98.09%
Kouzeh et al. (2021) [14]	MnasNet1	6	98.59%
Kouzeh et al. (2021) [14]	DenseNet121	6	98.87%
Kouzeh et al. (2021) [14]	ShuffleNet-V2	6	99.03%
Tavakoli et al. (2021) [22]	Segmentation	6	94.65%
Ours	$\overset{{}_{{}}}{\operatorname{DB}}$ and SA^1	6	$\boldsymbol{99.17\%}$

¹Dual Branch and Shape Attention (DB and SA)

Тор	Bottom	Inputs	Acc	Pre	Rec	F1
BecNet18 [9]	—	RGB	78.79%	71.91%	73.09%	71.77%
	VGG19 [<mark>8</mark>]	RGB+Shape	79.51%	74.27%	73.54%	72.42%
D N -+ 50 [0]	—	RGB	79.31%	72.16%	73.18%	71.85%
ResNet50 [9]	VGG19 [8]	RGB+Shape	79.58%	74.33%	73.67%	72.51%
DemoeNet191 [10]		RGB	79.30%	71.45%	72.23%	70.91%
Denselvet121 [10]	VGG19 [8]	RGB+Shape	79.65%	73.42%	73.15%	72.20%
ResNest [11]		RGB	80.09%	74.49%	73.95%	72.87%
	VGG19 [8]	RGB+Shape	80.12%	76.47%	74.64%	73.04%
Carrier Tr [E]]		RGB	80.62%	75.81%	74.79%	72.85%
Swiii-1 [0]-	VGG19 [8]	RGB+Shape	81.12%	76.28%	75.73%	74.00%

Table 8 Ablation study on the shape information using our BMEC dataset.

¹Swin-Transformer (Swin-T)

only used the raw image feature for classification. In this case, the network becomes a conventional RGB image based classification model. When two branches are enabled, both the shape feature extracted from the binary shape mask and shape-attended image feature computed from the shape attention module are fused to the raw image feature for classification. For the single branch network, different backbone models are tested on the BMEC RGB

images. For our dual-branch network, the VGG19 is used in each configuration so that we can compare the performance improvement for each case. From the results in Table 8, the dual-branch network can always outperform the singlebranch, confirming the effectiveness of the shape information for improving the fine-grained cell classification.

6 Conclusion, Limitation and Future Work

We introduce BMEC, the first and largest professionally annotated image dataset for fine-grained classification of erythroid cells. The BMEC is public available for all communities and we hope it can encourage more deep learning research on erythroid cells. We propose a novel shape-aware network which explicitly utilizes the shape information to improve the classification accuracy in a dual-branch deep learning framework. We conduct extensive evaluations of different backbones and perform experiments on BMEC and two other WBC datasets. The results show the superiority and generalizability of our shapeaware dual-branch network. Ablation studies also verify the effectiveness of the shape information for the fine-grained cell classification.

Although our method has achieved superior performance on cell classification, it still has some limitations which can inspire more future work. Our current cell shape extraction still involves several image processing steps which may be affected by the image noise and staining quality. It will be interesting to automatically learn the cell shape segmentation in the network and then apply the learned shape mask in our dual-branch network. Meanwhile, our classification works on the cropped cell images which are prepared in the BMEC dataset. Developing a cell detection and classification network which can directly work on the full-field cell images is also a promising future work. Due to the privacy issue, we cannot release the full-field cell images used in our BMEC, but it is still possible to train a cell classification network using the cropped BMEC cell images.

Declarations

- Funding: This research is supported by the National Natural Science Foundation of China (Grants No. 61772227, 61972174, 61972175, 62202199), Science and Technology Development Foundation of Jilin Province (No. 20180201045GX, 20200201300JC, 20200401083GX, 20200201163JC), the Jilin Development and Reform Commission Fund (No. 2020C020-2).
- Conflict of interest/Competing interests: The authors declare no conflict of interest.
- Availability of data and materials: The code and dataset have been published on github (https://github.com/wangye8899/BMEC).
- Code availability: https://github.com/wangye8899/BMEC
- Authors' contributions: Conceptualization, Y.W. and R.M.; methodology, Y.W., R.M. and Y.Z.; validation, X.M., Y.X. and X.W.; formal analysis, Y.W. and R.M.; investigation, H.C.; resources, Y.Z.; data curation, H.C.

and X.M.; writing—original draft preparation, Y.W.; writing—review and editing, R.M.; visualization, Y.W. and R.M.; supervision, Y.Z.; project administration, Y.Z.; All authors have read and agreed to the published version of the manuscript.

References

- Chitra, P., Jebarani, M., Kavipriya, P., Srilatha, K., Sumathi, M., Lakshmi, S.: Detection of aml in blood microscopic images using local binary pattern and supervised classifier. Research Journal of Pharmacy and Technology 12(4), 1717–1720 (2019)
- [2] Alomari, Y.M., Sheikh Abdullah, S.N.H., Zaharatul Azma, R., Omar, K.: Automatic detection and quantification of wbcs and rbcs using iterative structured circle detection algorithm. Computational and mathematical methods in medicine (2014)
- [3] Lippeveld, M., Knill, C., Ladlow, E., Fuller, A., Michaelis, L.J., Saeys, Y., Filby, A., Peralta, D.: Classification of human white blood cells using machine learning for stain-free imaging flow cytometry. Cytometry Part A 97(3), 308–319 (2020)
- [4] Petrović, N., Moyà-Alcover, G., Jaume-i-Capó, A., González-Hidalgo, M.: Sickle-cell disease diagnosis support selecting the most appropriate machine learning method: Towards a general and interpretable approach for cell morphology analysis from microscopy images. Computers in Biology and Medicine **126**, 104027 (2020)
- [5] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. arXiv preprint arXiv:2103.14030 (2021)
- [6] Xu, M., Zhang, Z., Hu, H., Wang, J., Wang, L., Wei, F., Bai, X., Liu, Z.: End-to-end semi-supervised object detection with soft teacher. Proc. ICCV (2021)
- [7] Strudel, R., Garcia, R., Laptev, I., Schmid, C.: Segmenter: Transformer for semantic segmentation. (2021)
- [8] Simonyan, K., Zisserman, A.: Very deep convolutional networks for largescale image recognition. arXiv preprint arXiv:1409.1556 (2014)
- [9] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: IEEE CVPR, pp. 770–778 (2016)
- [10] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: IEEE CVPR, pp. 4700–4708 (2017)

- 18 Shape-Aware Fine-Grained Classification of Erythroid Cells
- [11] Zhang, H., Wu, C., Zhang, Z., Zhu, Y., Zhang, Z., Lin, H., Sun, Y., He, T., Mueller, J., Manmatha, R., Li, M., Smola, A.J.: Resnest: Split-attention networks. arXiv: Computer Vision and Pattern Recognition (2020)
- [12] Rezatofighi, S.H., Soltanian-Zadeh, H.: Automatic recognition of five types of white blood cells in peripheral blood. Computerized Medical Imaging and Graphics 35(4), 333–343 (2011)
- [13] https://github.com/Shenggan/BCCD_Dataset
- [14] Kouzehkanan, Z.M., Saghari, S., Tavakoli, E., Rostami, P., Abaszadeh, M., Mirzadeh, F., Satlsar, E.S., Gheidishahran, M., Gorgi, F., Mohammadi, S., et al.: Raabin-wbc: a large free access dataset of white blood cells from normal peripheral blood. bioRxiv (2021)
- [15] Alam, M.M., Islam, M.T.: Machine learning approach of automatic identification and counting of blood cells. Healthcare Technology Letters 6(4), 103–108 (2019)
- [16] Gonzalez-Hidalgo, M., Guerrero-Pena, F., Herold-García, S., Jaume-i-Capó, A., Marrero-Fernández, P.D.: Red blood cell cluster separation from digital images for use in sickle cell disease. IEEE journal of biomedical and health informatics 19(4), 1514–1525 (2014)
- [17] Labati, R.D., Piuri, V., Scotti, F.: All-idb: The acute lymphoblastic leukemia image database for image processing. In: 2011 18th IEEE International Conference on Image Processing, pp. 2045–2048 (2011). IEEE
- [18] Zheng, X., Wang, Y., Wang, G., Liu, J.: Fast and robust segmentation of white blood cell images by self-supervised learning. Micron 107, 55–71 (2018)
- [19] Naruenatthanaset, K., Chalidabhongse, T.H., Palasuwan, D., Anantrasirichai, N., Palasuwan, A.: Red blood cell segmentation with overlapping cell separation and classification on imbalanced dataset. arXiv preprint arXiv:2012.01321 (2020)
- [20] Breiman, L.: Random forests. Machine learning 45(1), 5–32 (2001)
- [21] Friedman, J.H.: Stochastic gradient boosting. Computational statistics & data analysis 38(4), 367–378 (2002)
- [22] Tavakoli, E., Ghaffari, A., Kouzehkanan, Z.M., Hosseini, R.: New segmentation and feature extraction algorithm for classification of white blood cells in peripheral smear images. bioRxiv (2021)

- [23] Liu, C., Huang, L., Wei, Z., Zhang, W.: Subtler mixed attention network on fine-grained image classification. Applied Intelligence 51(11), 7903– 7916 (2021)
- [24] Wang, L., He, K., Feng, X., Ma, X.: Multilayer feature fusion with parallel convolutional block for fine-grained image classification. Applied Intelligence 52(3), 2872–2883 (2022)
- [25] Sun, C., Ai, Y., Wang, S., Zhang, W.: Mask-guided ssd for small-object detection. Applied Intelligence 51(6), 3311–3322 (2021)
- [26] Zhou, Y., Wang, Y., Wu, J., Hassan, M., Pang, W., Lv, L., Wang, L., Cui, H.: Erythroidcounter: an automatic pipeline for erythroid cell detection, identification and counting based on deep learning. Multimedia Tools and Applications, 1–16 (2022)
- [27] Toğaçar, M., Ergen, B., Cömert, Z.: Classification of white blood cells using deep features obtained from convolutional neural network models based on the combination of feature selection methods. Applied Soft Computing 97, 106810 (2020)
- [28] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems 25, 1097–1105 (2012)
- [29] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: IEEE CVPR, pp. 1–9 (2015)
- [30] Sahlol, A.T., Kollmannsberger, P., Ewees, A.A.: Efficient classification of white blood cell leukemia with improved swarm optimization of deep features. Scientific reports 10(1), 1–11 (2020)
- [31] Mirjalili, S., Gandomi, A.H., Mirjalili, S.Z., Saremi, S., Faris, H., Mirjalili, S.M.: Salp swarm algorithm: A bio-inspired optimizer for engineering design problems. Advances in Engineering Software 114, 163–191 (2017)
- [32] Alzubaidi, L., Fadhel, M.A., Al-Shamma, O., Zhang, J., Duan, Y.: Deep learning models for classification of red blood cells in microscopy images to aid in sickle cell anemia diagnosis. Electronics 9(3), 427 (2020)
- [33] Pasupa, K., Vatathanavaro, S., Tungjitnob, S.: Convolutional neural networks based focal loss for class imbalance problem: A case study of canine red blood cells morphology classification. Journal of Ambient Intelligence and Humanized Computing, 1–17 (2020)
- [34] Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense

object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)

- [35] Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: ECCV (2014)
- [36] Tavakoli, S., Ghaffari, A., Kouzehkanan, Z.M.: Generalizability in white blood cells' classification problem. bioRxiv (2021)
- [37] Gregory, T.R.: The bigger the c-value, the larger the cell: genome size and red blood cell size in vertebrates. Blood Cells, Molecules, and Diseases 27(5), 830–843 (2001)
- [38] Otsu, N.: A threshold selection method from gray level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9, 62–66 (1979)
- [39] Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E.: Squeeze-and-excitation networks. In: Computer Vision and Pattern Recognition (2018)
- [40] Wightman, R.: PyTorch Image Models. GitHub (2019). https://doi.org/ 10.5281/zenodo.4414861
- [41] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch. In: NIPS-W (2017)
- [42] He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: International Conference on Computer Vision (2015)
- [43] Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016)
- [44] Jung, C., Abuhamad, M., Alikhanov, J., Mohaisen, A., Han, K., Nyang, D.: W-net: a cnn-based architecture for white blood cells image classification. arXiv preprint arXiv:1910.01091 (2019)
- [45] Baydilli, Y.Y., Atila, Ü.: Classification of white blood cells using capsule networks. Computerized Medical Imaging and Graphics 80, 101699 (2020)
- [46] Harshanand, B., Sangaiah, A.K.: Comprehensive analysis of deep learning methodology in classification of leukocytes and enhancement using swish activation units. Mobile networks and applications 25(6), 2302–2320 (2020)
- [47] Khan, M.A., Qasim, M., Lodhi, H.M.J., Nazir, M., Javed, K., Rubab, S., Din, A., Habib, U.: Automated design for recognition of blood cells

diseases from hematopathology using classical features selection and elm. Microscopy Research and Technique 84(2), 202–216 (2021)