Fine-Grained Classification of Neutrophils with Hybrid Loss

Qingtao Zhu[1], Danwei Lu[1], Tao Zhang[1], Junjun Yin[2], and Jian Yang[1(✉)]

[1] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
yangjian_ee@tsinghua.edu.cn

[2] School of Computer and Communication Engineering, University of Science and Technology, Beijing 100083, China

Abstract. Acute leukemia is a malignant clonal disease of hematopoietic stem cells, which is usually diagnosed by morphological examination of bone marrow cells. However, the morphological examination usually relies on the subjective inference of cell morphology experts and is labor-intensive. With the development of computer vision, automatic classification and counting of blood cells is increasingly popular, which greatly improves work efficiency. Within this context, we here propose a novel method for neutrophil classification, which is based on deep neural network. In brief, it first crops the single cells from the large images, and then makes use of the loss functions designed for face recognition and weakly-supervised fine-grained visual classification. With the hybrid loss, the trained network can focus on nucleus areas, extract features with inter-class differences and intra-class compactness. Experiments show that the proposed method can obtain higher overall accuracy. Data is available at https://github.com/stevenxmy/subAML-dataset.git.

Keywords: Acute leukemia · Fine-grained classification · Neural network

1 Introduction

Leukemia is a highly heterogeneous tumor with malignant hyperplasia of hematopoietic tissue or clonal hyperplasia of lymphoid tissue, which has high morbidity and fatality rate. Early haematological diagnosis of acute leukemia is crucial for the patients. Comprehensive haematological diagnostics are usually combined with complex inspection such as blood routine, bone marrow cell morphological examination and cytogenetic analysis. Particularly, morphological evaluation of leukocytes from peripheral blood or bone marrow samples is one of the most basic and important methods. During the hamatological diagnostics, trained human examiners observe bone marrow smears under microscopes, check whether there are abnormal cells and count different types of nucleated cells, the proportion of which depicts the pathological trend.

Y. Peng et al. (Eds.): ICIG 2021, LNCS 12888, pp. 102–113, 2021.
https://doi.org/10.1007/978-3-030-87355-4_9

The specific process of hamatological diagnostics is shown in Fig. 1. First, bone marrow smears are generated by bone marrow aspiration. Next, the cytochemical staining was performed by the Wright-Giemsa staining analysis technique. Then, morphological analysis and classification of bone marrow cells are conducted under oil immersion lens. Finally, preliminary diagnoses of the type of acute leukemia are made according to the French-American-British classification standard [1]. However, it takes quite a long time to cultivate an expert in cell morphology, and the cell examination is time-consuming and laborious.

Fig. 1. Diagram of the human examination of cell morphology

With the advancement of computer vision technologies and pattern recognition algorithms, emphasis is laid on automatic recognition of bone marrow cell types, which can reduce the burden on experts and improve work efficiency. A feasible solution is to make use of artificially designed feature extractions combined with feature selection and classifiers. Gurcan et al. [15] extracted chromaticity, morphology and texture features, and used SVM for classification. In recent years, deep learning has demonstrated outstanding performance in various fields of computer vision, such as image detection, segmentation, and recognition. Researchers have focused on the bone marrow cell classification based on deep neural networks. Christian et al. [11] used the ResNext network to classify 18,365 blood cell smear images of 200 patients collected by the Munich University Hospital Laboratory from 2014 to 2017. The blood cells are divided into 15 categories, while a few classes only contain dozens of images. It is not surprising that the precision and sensitivity of partial sub-categories obtained by such a generic classification network structure are very low. Hong et al. [6] made use of the deep neural network to classify the bone marrow cells based on 3000 marrow smear samples collected by Sir Run Run Shaw Hospital affiliated to Zhejiang University School of Medicine between 2016 and 2018. However, the sensitivity of certain types of cells such as myelocytes and promyelocytes are poor. In conclusion, although the classification of main blood cell types has been widely developed, the fine-grained classification of sub-categories, which is crucial for hamatological diagnostics, has not been well analyzed.

In recent years, face recognition and fine-grained visual classification(FGVC) have been increasingly popular. The common challenging problems of face recognition and FGVC are the large inter-class similarity and intra-class variability [12]. Researchers have focused on the loss functions to train deep neural networks for feature extraction, which greatly improve the performance of the networks. The morphological classification of blood cells is similar to the above tasks to a certain extent. However, there have been few research on white blood cell classification taking the algorithms of face recognition and FGVC into consideration, which may contribute to the precise classification of sub-categories of blood cells.

To address the gap above, we propose a pipeline similar to the face recognition which involves the steps of detection and classification. The feature pyramid network is applied to detect single cells. The backbone of EfficientNet [16] is adopted. A hybrid loss function containing the clustering component and diversity component is also designed to reinforce the inter-class difference and intra-class compactness, which borrows the idea of the loss functions designed for face recognition and weakly-supervised FGVC.

The contributions are summarized as follows:

- An half-automatic cell detection pipeline is designed to generate the actual single-cell images.
- The backbone of EfficientNet is adopted to enhance the feature extraction and save the memory.
- A hybrid loss for classification is designed by combining the center loss and diversity loss.
- Data cleaning and augmentation is done for part of the Munich AML Morphology Dataset [11] which is open source.

2 Related Work

2.1 Generic Object Detection

Generic object detection refers to the task of determining whether there are certain types of objects in the scene, and return the locations if the targets exist [9]. Traditional target detection algorithms mainly include preprocessing, window sliding, feature extraction, feature selection, feature classification and post-processing. The generic object detection algorithms based on deep learning are divided into two groups: two-stage methods with region proposal stage and object recognition stage, and single-stage methods without candidate boxes. Girshick et al. [14] proposed the classic dual-stage object detection method Faster R-CNN, which gained state-of-the-art results on PASCAL VOC 2007. Redmon et al. [13] implemented feature extraction, bounding box classification and regression in a branchless deep convolutional network, which effectively improved the low efficiency of Faster R-CNN. Recently, numerous studies on dual-stage and single-stage methods have been conducted, such as FPN [7] and YOLOv4 [2], the accuracy and speed of which are much better.

2.2 Fined-Grained Visual Classification

Fined-grained visual classification(FGVC) refers to the task of classifying subcategories of a certain category such as vehicles, birds and flowers, which draws considerable attention. Types of research approach on FGVC can be divided into two groups: strongly-supervised algorithms and weakly-supervised algorithms. Wei et al. [18] proposed an end-to-end model which localized parts and selected descriptors for FGVC of birds, which made use of strong supervision annotations. Chang et al. [3] proposed the mutual-channel loss with a discriminality component and a diversity component to force the channel maps of each class to be discriminative and mutually exclusive.

2.3 Face Recognition

Three basic steps are included in robust face recognition system: face detection, feature extraction and face recognition. With the development of deep learning, the second and third steps are usually merged in deep neural network. Different from generic object classification, there are numerous identities which may also change irregularly in total number. Therefore, feature embeddings and similarity measures are combined to recognize identities without fine-tuning the model frequently. Discriminative feature embeddings are usually extracted by training with particular losses such as center loss [19], CosFace [17] and ArcFace [4]. Cosine similarity and euclidean distance are usually adopted as the similarity measures.

3 Methods

3.1 Single Cell Detection

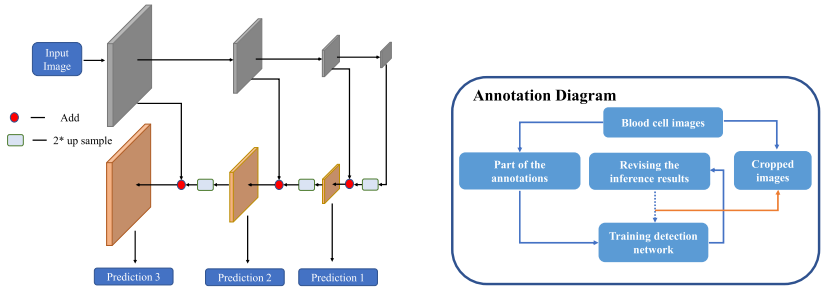
Although the common two-stage detection network Faster R-CNN has designed multiple sizes of anchor boxes and multiple aspect ratios, it has not solved the problem of low resolution of deep feature maps, which leads to the poor performance of small target detection. We adopt the idea of Feature Pyramid Network (FPN) [7] and combines multi-scale feature map information for detection. First, the proposal bounding boxes are generated through the region proposal network. After that, the coarse screening and preliminary anchor box regression are performed. Finally, the bounding boxes of different scales are finely selected while coordinates are revised by the fully connected network. The inference of FPN is depicted in Fig. 2(a). Specifically, due to the small scale variation range of blood cells, compared with generic object detetion, only three output scales are used as the final prediction to reduce the false alarm rate.

However, the locations of the bounding boxes may not have been annotated. The strategy of active learning is adopted. We first annotated part of the bounding boxes, then fine-tuned FPN for detection with the pretrained MS COCO [8] weights. After that, the inference results of the bounding boxes are manually revised, in case of the missed targets or the bad boxes. If the dataset is quite large, the revision work can be done in an iterative way, and the detection model can be updated for several times. Finally, the single-cell images are cropped from the original images. The diagram of the annotation work is depicted in Fig. 2(b).

3.2 EfficientNet Backbone

Tan et al. [16] made use of neural architecture search to design a baseline network EfficientNet-B0, and scale it up to build a series of models called EfficientNets.

The MBConv module in EfficientNet includes the squeeze-and-excitation [5] block to automatically recalibrates the channel-wise feature responses. In the squeeze-and-excitation block, the weights of the channels in feature maps are redistributed by global average pooling layer, fully connected layer and point-wise product layer. The channel attention blocks mentioned above contribute to the excellent feature extraction capabilities of the network.



(a) Inference of Feature Pyramid Network (b) Diagram of bounding box annotation

Fig. 2. FPN structure and half-automatic annotations

3.3 Loss Functions

Clustering Component. We adopt the center loss [19] as clustering component to improve the intra-class compactness.

The traditional softmax loss function is presented as follow.

$$\mathcal{L}_{softmax} = -\sum_{i=1}^{n} log \frac{e^{\boldsymbol{W}_{y_i}^T \boldsymbol{x}_i + \boldsymbol{b}_{y_i}}}{\sum_{j=1}^{m} e^{\boldsymbol{W}_j^T \boldsymbol{x}_i + \boldsymbol{b}_j}} \tag{1}$$

In Eq. 1, n denotes the batch size of input data, while m denotes the number of class. $\boldsymbol{x}_i \in \mathbb{R}^d$ denotes the ith extracted feature of the mini-batch. $\boldsymbol{W} \in \mathbb{R}^{d \times m}$ and $\boldsymbol{b} \in \mathbb{R}^d$denotes the weight and bias in the fully connected layer. Suppose the ith input belongs to class y_i, then the softmax loss is calculated as Eq. 1. However, the softmax loss mainly focuses on the inter-class differences. With the center loss [19] formulated in Eq. 2, the intra-class differences are minimized and the discriminative power of the convolutional neural network is improved.

$$\mathcal{L}_{center} = \frac{1}{2} \sum_{i=1}^{m} ||\boldsymbol{x}_i - \boldsymbol{c}_{y_i}||^2 \tag{2}$$

On each mini-batch, the center of features are computed once and the parameters of the center loss are updated. Scalar α is used to avoid perturbations caused by bad annotations. The gradients of $\mathcal{L}_{center}$ and the updating step is depicted as follow. $\boldsymbol{c}_{y_i}$ denotes the center of the feature embeddings extracted by the network. The centers of the embedding features are only updated when the predictions are correct.

$$\frac{\partial \mathcal{L}_{center}}{\partial \boldsymbol{x}_i} = \boldsymbol{x}_i - \boldsymbol{c}_{y_i} \tag{3}$$

$$\Delta \boldsymbol{c}_j = \frac{\sum_{i=1}^{m} \delta(y_i = j) \cdot (\boldsymbol{c}_j - \boldsymbol{x}_i)}{1 + \sum_{i=1}^{m} \delta(y_j = j)} \tag{4}$$

Diversity Component. The diversity component is presented as follow. W and H denote the width and height of the extracted feature map $\boldsymbol{F}$, which has ξ channels. Different from the diversity loss in mutual-channel loss [3], we take all feature channels into consideration instead of manually splitting the channels into m groups, where m denotes the number of classes. Therefore, the number of hyperparameters is reduced.

$$\mathcal{L}_{div} = -\sum_{k=1}^{WH} \underset{j=1,2,\ldots,\xi}{max} \left[\frac{e^{\boldsymbol{F}_{j,k}}}{\sum_{k'=1}^{WH} e^{\boldsymbol{F}_{j,k'}}} \right] \tag{5}$$

The diversity loss aims to reinforce the attention on different areas of the images, which may be crucial for the fine-grained classification of cells.

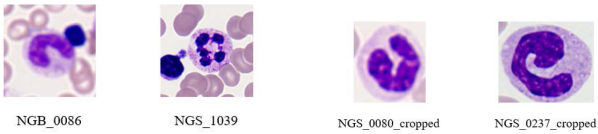
Hybrid Loss. The hybrid loss adopted in our approach is presented in Eq. 6. Considering the problem of convergence and the bad samples, the center loss and diversity loss are combined with the softmax loss. α and β are the hyperparameters which controls the proportion of $\mathcal{L}_{softmax}$, $\mathcal{L}_{center}$ and $\mathcal{L}_{div}$.

$$\mathcal{L}_{hybrid} = \mathcal{L}_{softmax} + \alpha \cdot \mathcal{L}_{center} + \beta \cdot \mathcal{L}_{div} \tag{6}$$

4 Experiments

4.1 Implementation Details

Datasets. As given in Table 1, we employ the subset of the Munich AML Morphology Dataset called Sub-AML Dataset which contains myelocytes in different stages as our training data and conduct comparison with other approaches. Additionally, promyelocyte (PMO) and myelocyte (MYB) cells are combined into the PMOMYB class, since many cells in these two classes are much alike. Although it is claimed that the Munich AML Morphology Dataset only contains single-cell images, there are still images containing more than one cell, as depicted in Fig. 3(a). Please note that the metamyelocyte class is removed because of the low proportion. There are 8484 segmented neutrophils in the original dataset, while some should be considered as band neutrophils, as depicted in Fig. 3(b). To best of our knowledge, data imbalance is unfriendly to network training. We apply Gaussian blur, average blur, random jittering in HSV space, rotation and flipping to alleviate data imbalance. To avoid data leaking, single-cell cropped images augmented from the same images can not be included in both training and validation folds. Finally, the Sub-AML Dataset contains 11655 PMOMYB cells, 13760 band neutrophils (NGB) and 11976 segmented neutrophils (NGS). The precision and sensitivity of different categories gained by the method in [11] are also depicted in Table 1.



(a) Images containing more than one cell (b) Suspected annotations

Fig. 3. Multi-cell images and suspected annotations

Table 1. The subset of Munich AML Morphology Dataset

Class	Precision	Sensitivity	Images	Chosen	No.	Final No.
Erythroblast	0.75	0.87	78			
Lymphocyte (typical)	0.96	0.95	3937			
Lymphocyte (atypical)	0.20	0.07	11			
Monoblast	0.52	0.58	26			
Monocyte	0.90	0.90	1789			
Myeloblast	0.94	0.94	3268			
Promyelocyte (bilobled)	0.45	0.41	18			
Smudge cell	0.53	0.77	15			
Eosinophil	0.95	0.95	424			
Basophil	0.48	0.82	79			
Promyelocyte	**0.63**	**0.54**	**70**	√	111	11655
Myelocyte	**0.46**	**0.43**	**42**	√		
Metamyelocyte	0.07	0.13	15			
Neutrophil (band)	**0.25**	**0.59**	**109**	√	344	13760
Neutrophil (segmented)	**0.99**	**0.96**	**8484**	√	1996	11976

Experimental Settings. As mentioned in Sect. 3.1, single cells are detected by FPN and cropped from the 400 × 400 images in the Munich AML Morphology Dataset with a half-automatic pipeline. Single-cell images are then resized to 224 × 224. We set the batch size to 32 and train models on one NVIDIA RTX 2080TI GPU. The training process is finished at 60 epochs in Pytorch 1.4.0. Momentum is set to 0.9 and weight decay is set to 5e−4. The learning rate starts from 0.0005 and is divided by 10 at 30, 48 epochs. Additionally, 5-fold cross-validation is used to evaluate the methods. The accuracy of fold-k is denoted as 'acc$k$'.

4.2 Ablation Study

Performance Metrics. We calculate accuracy as the comparison metric, which represents the ratio of correct predictions among the total samples.

Center Loss. We explore the proportion of center loss on the Sub-AML Dataset with the backbone of EfficientNet-B0. As depicted in Table 2, the center loss is

effective for improving classification accuracy. The best value of α observed in our experiments is 0.5. We also compare center loss with classic margin based losses such as ArcFace and CosFace, and the selection of hyperparameters refers to [4] and [17], in which scaling parameter is set to 64 and margin parameters are set to 0.5 and 0.35. However, the performance of the ArcFace and CosFace is bad, which may due to the small number of classes.

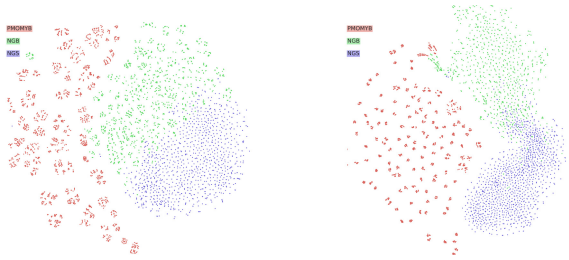
Table 2. Verification results of different α

Backbone	Loss Function	α	acc1	acc2	acc3	acc4	acc5	Average acc
Efficient-B0	Center	0.0	96.27%	95.24%	95.87%	95.04%	96.29%	95.74%
Efficient-B0	Center	0.1	96.50%	**96.04%**	96.10%	95.05%	96.27%	96.00%
Efficient-B0	Center	0.3	**96.77%**	94.96%	96.73%	**95.88%**	95.97%	96.06%
Efficient-B0	Center	0.5	96.14%	95.66%	96.73%	95.37%	96.65%	**96.11%**
Efficient-B0	Center	0.7	96.52%	95.00%	96.69%	94.24%	96.90%	95.87%
Efficient-B0	Center	0.9	96.38%	95.07%	**96.85%**	94.87%	**96.93%**	96.01%
Efficient-B0	ArcFace	/	94.22%	92.31%	94.10%	93.90%	94.24%	93.75%
Efficient-B0	CosFace	/	93.99%	92.75%	93.02%	93.07%	93.15%	93.20%

Since the dimension of the feature embedding is high, we adopt the t-SNE visualization tool [10], which can retain the local structure of the extracted data and reveal important global structure.

As depicted in Fig. 4, points with three colors which refer to the three classes show local clustering. Additionally, the similarity between NGB and NGS cells leads to the overlaps on the borders. To put it in a nutshell, the center loss successfully enhances the intra-class compactness of the testing data as well as the inter-class discrimination.

Diversity Loss. We explore the proportion of diversity loss on the Sub-AML Dataset with the backbone of EfficientNet-B0. As depicted in Table 3, the best value of β varies in each fold, and the diversity loss does work under certain hyperparameters. On average, the best value of β observed in our experiments is 0.05, while the weight of center loss is set to 0. The overall accuracy is not improved with only diversity loss. However, there is much difference on the attention maps of the extracted features. In Fig. 5, more emphasis is laid on nucleus areas. It is obvious that the attention areas of the features extracted by the neural network trained with diversity loss are sometimes more abundant.



(a) Without center loss (b) With center loss

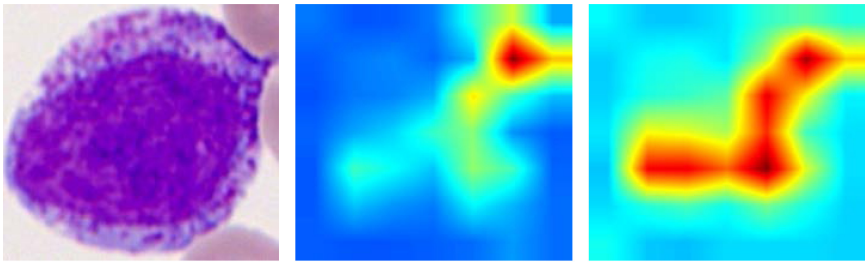
Fig. 4. T-SNE visualization of the feature embeddings

Table 3. Vertification results of different β

Backbone	β	acc1	acc2	acc3	acc4	acc5	Average acc
Efficient-B0	0.0	96.27%	95.24%	95.87%	95.04%	96.29%	95.74%
Efficient-B0	0.00005	96.29%	94.83%	95.58%	**95.21%**	96.29%	95.64%
Efficient-B0	0.0005	**96.47%**	**95.36%**	95.53%	94.20%	96.24%	95.56%
Efficient-B0	0.005	95.83%	95.10%	95.51%	95.15%	**96.33%**	95.58%
Efficient-B0	0.05	96.06%	95.29%	95.77%	95.03%	96.14%	95.66%
Efficient-B0	0.5	96.14%	95.19%	**96.26%**	94.75%	95.98%	95.66%

4.3 Evaluation Results

We train EfficientNet-B0 model with hybrid loss of different hyperparameters. Since the datasets of blood cell morphology are very few, we only compare our work with other CNN-based methods on Sub-AML Dataset, such as ResNext-101, VGG-16, ResNet-18, ResNet-50 and DenseNet-121, as depicted in Table 4. The performance of mutual-channel loss is also evaluated. Our method achieves the best accuracy among all the methods listed, while the backbone of Efficient-B0 has the smallest FLOPs [16]. Take fold-1 for instance, the best accuracy on test set is achieved at epoch 14. The curves of loss and accuracy are depicted in Fig. 6(a) and 6(b), while the confusion matrix of our method is depicted in Fig. 6(c). The convergence of our method is fast. With the data cleaning and augmentation work, the accuracy of the sub-category, especially NGB, have been significantly improved, compared with the results in [11]. Additionally, even though the diversity loss is not effective separately, the hybrid loss obviously contributes to the improvement of overall accuracy.

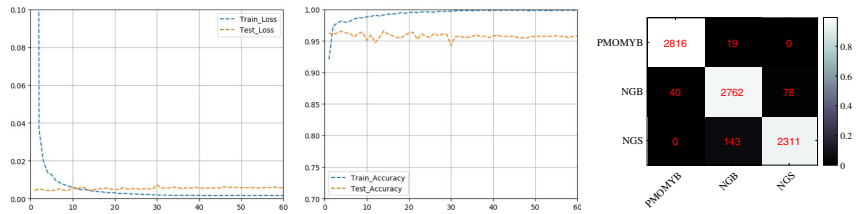


(a) Original image (b) Without diversity loss (c) With diversity loss

Fig. 5. Original single-cell image and the attention map of the extracted features

Table 4. Vertification performance of different methods on Sub-AML Dataset

Method	(α, β)	acc1	acc2	acc3	acc4	acc5	Average acc
Proposed	(0.5, 0.0005)	**96.57%**	**96.10%**	**96.71%**	**95.97%**	**96.86%**	**96.44%**
Mutual-channel loss	/	95.96%	95.92%	95.88%	95.14 %	96.18%	95.82%
Efficient-B0	/	96.27%	95.24%	95.87%	95.04%	96.29%	95.74%
ResNet-18	/	93.62%	91.93%	93.54%	94.49%	94.20%	93.56%
ResNet-50	/	94.97%	93.75%	94.77%	94.97%	94.62%	94.83%
ResNext-101	/	96.38%	95.49%	95.97%	94.77%	96.29%	95.78%
VGG-16	/	94.26%	93.29%	93.96%	93.98%	95.44%	94.19%
DenseNet-121	/	95.84%	95.27%	95.95%	94.96%	96.11%	95.63%



(a) Training and testing loss (b) Training and testing accuracy (c) Confusion matrix

Fig. 6. The curves and confusion matrix of fold-1

5 Conclusions

Automatic morphology examination of blood cells is a promising field, since it can save much labor. There have already exist a few studies which focus on rough classification of five types of blood cells. However, little research lays emphasis on fine-grained classification of the sub-categories, which is more difficult.

In this paper, we borrow the idea from the pipeline of face recognition, and propose half-automatic bounding box annotation method and resize the images which contain only one cell to a fixed size. Hybrid loss function is designed to gain better performance on the fine-grained classification of the neutrophils in

adjacent stages. Clustering component reinforces the compactive distribution of the feature embeddings of the blood cells belonging to the same class, which also contributes to the inter-class discrimination. Diversity component reinforces the broad receptive field of the network, which helps to discover important local features. From the comprehensive experiments done on the Sub-AML Dataset, we demonstrate that the method proposed outperforms the popular deep neural network based methods. With the increasing blood smear data, combined with rough classification techniques, our fine-grained classification of sub-categories may greatly contribute to the improvement of the overall accuracy of automatic blood cell fine classification in the future, which is crucial for the diagnosis of leukemia.

Acknowledgements. We thank Christian Matek and Antje for the morphologcial dataset of leukocytes.

References

1. Bennett, J.M., et al.: Proposals for the classification of the acute leukaemias French-American-British (fab) co-operative group. Br. J. Haematol. **33**(4), 451–458 (1976)
2. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: YOLOv4: optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)
3. Chang, D., et al.: The devil is in the channels: mutual-channel loss for fine-grained image classification. IEEE Trans. Image Process. **29**, 4683–4695 (2020)
4. Deng, J., Guo, J., Xue, N., Zafeiriou, S.: ArcFace: additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4690–4699 (2019)
5. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
6. Jin, H., et al.: Developing and preliminary validating an automatic cell classification system for bone marrow smears: a pilot study. J. Med. Syst. **44**(10), 1–10 (2020)
7. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017)
8. Lin, T.Y., et al.: Microsoft COCO: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 740–755. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_48
9. Liu, L., et al.: Deep learning for generic object detection: a survey. Int. J. Comput. Vis. **128**(2), 261–318 (2020)
10. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(11) (2008)
11. Matek, C., Schwarz, S., Spiekermann, K., Marr, C.: Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. Nat. Mach. Intell. **1**(11), 538–544 (2019)
12. Qiu, C., Zhou, W.: A survey of recent advances in CNN-based fine-grained visual categorization. In: 2020 IEEE 20th International Conference on Communication Technology (ICCT), pp. 1377–1384. IEEE (2020)

13. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
14. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks (2016)
15. Sarrafzadeh, O., Rabbani, H., Talebi, A., Banaem, H.U.: Selection of the best features for leukocytes classification in blood smear microscopic images. In: Medical Imaging 2014: Digital Pathology, vol. 9041, p. 90410P. International Society for Optics and Photonics (2014)
16. Tan, M., Le, Q.: EfficientNet: rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning, pp. 6105–6114. PMLR (2019)
17. Wang, H., et al.: CosFace: large margin cosine loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5265–5274 (2018)
18. Wei, X.S., Xie, C.W., Wu, J., Shen, C.: Mask-CNN: localizing parts and selecting descriptors for fine-grained bird species categorization. Pattern Recogn. **76**, 704–714 (2018)
19. Wen, Y., Zhang, K., Li, Z., Qiao, Yu.: A discriminative feature learning approach for deep face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 499–515. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_31