

## Article

# Automatic Recognition of Blood Cell Images with Dense Distributions Based on a Faster Region-Based Convolutional Neural Network

Yun Liu <sup>1,2,\*</sup> , Yumeng Liu <sup>1</sup>, Menglu Chen <sup>1</sup>, Haoxing Xue <sup>1</sup>, Xiaoqiang Wu <sup>1</sup>, Linqi Shui <sup>1</sup>, Junhong Xing <sup>1</sup>, Xian Wang <sup>1</sup>, Hequn Li <sup>1</sup> and Mingxing Jiao <sup>1</sup> 

<sup>1</sup> School of Mechanical and Precision Instrument Engineering, Xi'an University of Technology, Xi'an 710048, China; mylbling@163.com (Y.L.); menglu\_chen@outlook.com (M.C.); xhx5215@163.com (H.X.); qiangxiaoqiang2253@163.com (X.W.); shuilinqi@xaut.edu.cn (L.S.); xjh3729@xaut.edu.cn (J.X.); wangxian@xaut.edu.cn (X.W.); lhq@xaut.edu.cn (H.L.); jiaomx@xaut.edu.cn (M.J.)

<sup>2</sup> State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China

\* Correspondence: lyun@xaut.edu.cn

**Abstract:** In modern clinical medicine, the important information of red blood cells, such as shape and number, is applied to detect blood diseases. However, the automatic recognition problem of single cells and adherent cells always exists in a densely distributed medical scene, which is difficult to solve for both the traditional detection algorithms with lower recognition rates and the conventional networks with weaker feature extraction capabilities. In this paper, an automatic recognition method of adherent blood cells with dense distribution is proposed. Based on the Faster R-CNN, the balanced feature pyramid structure, deformable convolution network, and efficient pyramid split attention mechanism are adopted to automatically recognize the blood cells under the conditions of dense distribution, extrusion deformation, adhesion and overlap. In addition, the Align algorithm for region of interest also contributes to improving the accuracy of recognition results. The experimental results show that the mean average precision of cell detection is 0.895, which is 24.5% higher than that of the original network model. Compared with the one-stage mainstream networks, the presented network has a stronger feature extraction capability. The proposed method is suitable for identifying single cells and adherent cells with dense distribution in the actual medical scene.

**Keywords:** automatic recognition; densely distributed cell image; deep learning; Faster R-CNN; feature extraction



**Citation:** Liu, Y.; Liu, Y.; Chen, M.; Xue, H.; Wu, X.; Shui, L.; Xing, J.; Wang, X.; Li, H.; Jiao, M. Automatic Recognition of Blood Cell Images with Dense Distributions Based on a Faster Region-Based Convolutional Neural Network. *Appl. Sci.* **2023**, *13*, 12412. <https://doi.org/10.3390/app132212412>

Academic Editor: Andrea Prati

Received: 28 September 2023

Revised: 1 November 2023

Accepted: 8 November 2023

Published: 16 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Cell detection techniques have played an increasingly important role in modern clinical medicine and pathological diagnosis, because the shape and number of red blood cells contribute to detecting blood diseases. The manual cell detection method with lower efficiency is easily affected by subjective factors, while automatic instruments, such as cell analyzers, are not only expensive but also interfered with by impurities like white blood cells, resulting in a lower detection accuracy. Therefore, more and more researchers have been focusing on the automatic detection of cells.

The traditional cell detection methods may be divided into three categories. The first kind is cell detection based on traditional segmentation algorithms [1], commonly including watershed segmentation [2,3], global threshold segmentation [4] and color segmentation [5,6]. These methods make use of the gray or color features of the image to realize cell segmentation and recognition. The second type are recognition algorithms based on edge detection, such as the canny operator [7,8] and the Sobel operator [9,10]. These algorithms utilize the object edge information as the detection result. The third category are the traditional machine learning algorithms, for example, support vector machine

(SVM) [11] and neural networks [12]. They are used to classify the obtained features and further perform border regression by the target category. However, the traditional detection methods above need to select a suitable model for the automatic recognition of samples with different features; therefore, they are poor in generalization and robustness.

As deep learning is increasingly integrated into the medical field, much research on cell detection methods of deep learning has been carried out. The typical target detection network is composed of a faster region-based convolutional neural network (Faster R-CNN) [13], you only look once (YOLO) [14] and a single shot multi-box detector (SSD) [15], which mainly focus on cell segmentation or single-cell recognition. In 2018, Han et al. [16] proposed a method based on a generative adversarial network to realize the detection of cancer cells in pathological sections of breast cancer, in which the U-Net network was a generative network. In 2020, Banik et al. [17] used the K-means algorithm and a CNN to classify and locate white blood cells, respectively, thus separating them from whole blood smear images. In the same year, Chen et al. [18], by referring to a LeNet-5 network, constructed a model using a traditional texture feature classifier and shallow CNN to effectively identify the type of milk somatic cells, which provided a new idea for subsequent research on classification methods. In 2021, Lavitt et al. [19] combined the xResNet architecture of a CNN with transfer learning to realize cell counting, in which the cell counting problem was transformed into a regression task. In 2022, based on the design criteria of optimizing the accuracy and speed of detection with minimal resources, Anand et al. [20] developed a customized deep-learning architecture (YOLO-mp), and obtained the number of pathogens in microscope images of thick blood smears. Although these methods realize cell recognition and classification, most of them complete single-cell recognition under the premise of sparse distribution. However, dense cell distribution is inevitable in the actual medicine scenario, in which the target sizes caused by adhesion and overlap are different and the cell morphological changes caused by squeezing are irregular. There, the problem of unrecognizable adherent cells appears, resulting in an increased false detection rate. The current algorithm research on these cases is usually limited to general target detection or specific scene detection. Few studies involve the extraction of adherent cells' features, which makes feature extraction inadequate and the recognition rate low.

In this paper, a deep-learning target recognition network based on Faster R-CNN is proposed in order to recognize both single cells and adherent cells in a densely distributed cell scene. On the basis of the Faster R-CNN, a balanced feature pyramid (BFP) [21] structure, deformable convolution network (DCN) [22] and efficient pyramid split attention (EPSA) mechanism [23] are adopted to form a method named BDE\_RCNN. This method can effectively extract multi-scale features of cells and thus realizes the automatic recognition of binary classification for single cells and adherent cells. The proposed method has better feature extraction capability in comparison to other networks in terms of cell recognition. This will provide an effective solution for the automatic recognition of densely distributed cells in an actual medical scene.

## 2. Methods

### 2.1. Network Structure

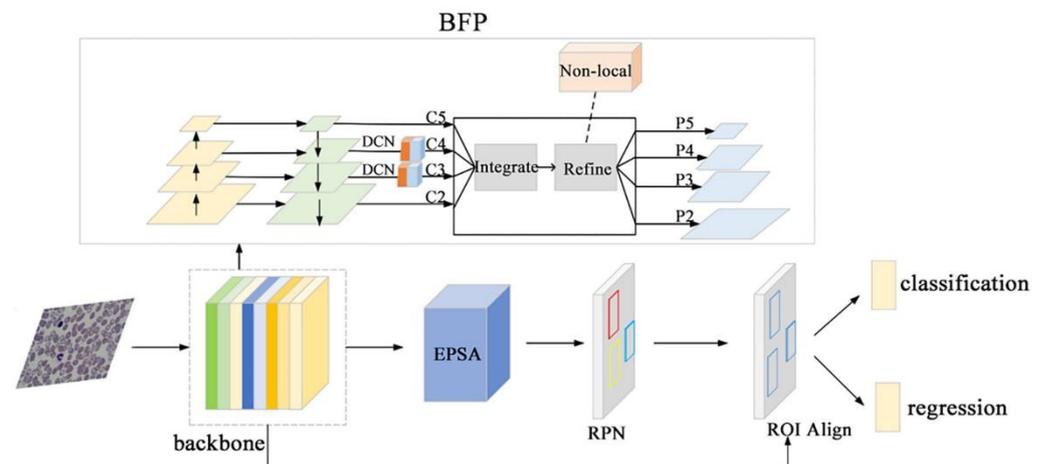
Faster R-CNN is a two-stage target detection model for end-to-end training. It is mainly composed of a backbone network, a region proposal network (RPN), region of interest pooling (ROI Pooling), classification and regression. The backbone network is a basic CNN, in which the input image is processed by convolution, rectified linear unit (ReLU) activation, pooling, etc. Its output image with high-dimensional features is imported into the RPN and processed by a  $3 \times 3$  convolutional layer to generate the anchors with multiple size proportions. The two outputs of the classification layer and regression layer are obtained for every anchor.

For the medical scene with densely distributed cells, there are many characteristics, such as a large number of cell samples, dense and disorderly distributions and larger

morphological difference between adherent cells and single cells. Thus, the recognition problems of the conventional Faster R-CNN are as follows:

1. The labeling results inevitably contain some non-target information due to the dense distributions of cells in the complex environment, which leads to an increased false detection rate.
2. The squeezing of cells may happen, resulting in irregular morphology. This increases the difficulty of feature extraction.
3. There are single red cells, agglomerated red cells, white blood cells or other impurities in an actual medical scene. These cause the problem of multi-scale recognition and reduce the range of receptive field.

In the BDE\_RCNN method, ResNet50 is selected to replace Vgg16 as the backbone network in the Faster R-CNN. Firstly, a feature pyramid network (FPN) [24] is added in the feature extraction network to make full use of the visual features extracted. Considering the dense distributions of cells, the FPN is integrated with a non-local module [25] to produce a BFP structure, in which the feature extraction capability is enhanced by correctly identifying multi-scale targets. Secondly, due to the diversity of irregular changes in cell morphology, it is difficult for the conventional convolutional networks to learn the multi-posture features caused by cell squeezing and deformation. The DCN is adopted to reform the C3 and C4 layers in the backbone feature extraction network, thus improving the feature extraction capability of the multi-posture features. Finally, the EPSA module is introduced to extract the multi-scale features of spatial information in each channel feature map by channel segmentations. It solves the loss problem of feature details caused by the different sizes of cells and the overlap of adherent cells. In addition, the resolution of the feature map decreases with the deepening of network depth in target detection of conventional networks, so the top-level feature map will lose the cell details. By using ROI Align in the improved network, instead of ROI Pooling, the problem of region mismatch caused by two quantization errors can be avoided in the pooling process. Figure 1 shows the structure of BDE\_RCNN.



**Figure 1.** BDE\_RCNN structure.

## 2.2. BFP Module

The BFP is a feature integration method. It combines the respective advantages of FPN and non-local to solve the unbalanced feature level of FPN in feature fusion. The BFP uses the feature map information of multiple levels to enhance the expression ability of each level in the feature map. The feature extraction network proposed is a ResNet50. Its last four convolutional layers are selected as the bottom-up network of the feature pyramid, defined as C2, C3, C4 and C5. After the C5 layer, a transverse connection and top-down network of new feature layers are established, namely P5, P4, P3 and P2. The complementary forms of

high- and low-level information can improve the detection performance of the network and produce richer semantic features.

In order to integrate the features of multiple layers while preserving their respective semantic layers, the BFP unifies the feature maps of different levels to the C4 layer by adaptive maximum pooling or the interpolation method. The pooling processing is adopted for a small feature map, while the bilinear interpolation is carried out for a large-scale feature map. By average operation, the balanced semantic feature is expressed as:

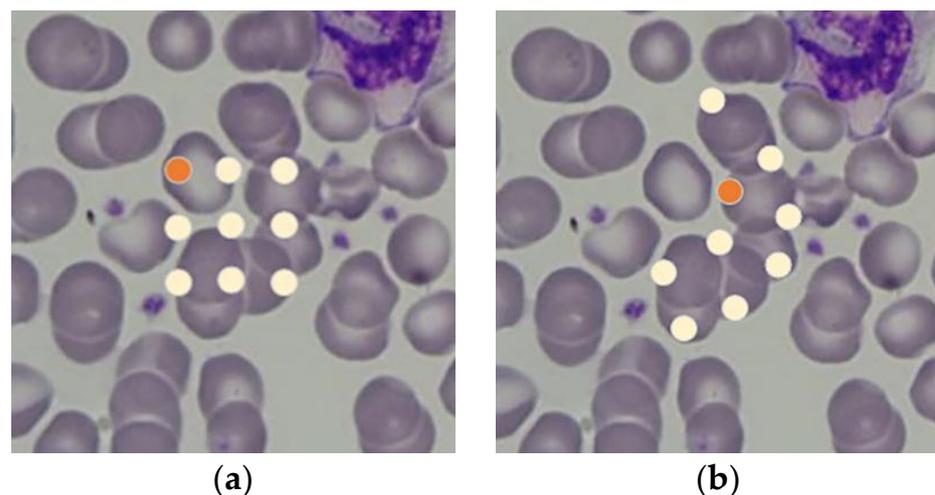
$$C = \frac{1}{L} \sum_{l=l_{\min}}^{l=l_{\max}} C_l, \quad (1)$$

where  $C_l$  is the prediction feature layer,  $l$  is the hierarchy,  $L$  is the total number of predicted feature layers,  $l_{\max}$  and  $l_{\min}$  are the highest level and lowest level, respectively.

The balanced semantic features are refined by an embedded non-local, which will integrate the global information and further re-scale it to enhance the original features. This way of feature integration allows each scale to utilize richer details for recognition of dense cells at multiple scales.

### 2.3. DCN Module

Cases of cell adhesion, squeezing and morphological change exist in a cell dataset, resulting in irregular deformations of cell morphology. To solve this problem, the DCN module is introduced in the C3 and C4 layers of the backbone, respectively. On the basis of traditional convolution, the DCN adds the direction vectors of adjusting the convolution kernel to obtain more areas of interest according to sample shapes. This allows the kernel to closely align with characteristic objects, facilitating the learning of more complex transformations and reducing the background information in the receptive field. Figure 2 shows the sampling location diagrams of standard convolution and DCN, respectively. The differently colored balls denote the sampling location results, in which the orange ball is the initial position in each diagram. It can be seen that the DCN enables the sampling points of convolution kernel to shift in the input feature diagram, which makes the convolutional module extract more accurate target features.



**Figure 2.** Sampling location diagrams of standard convolution and DCN. (a) Standard convolution; (b) DCN.

### 2.4. EPSA Module

In order to effectively acquire the characteristic information of different sizes for single cells and adherent cells, the lightweight EPSA module is adopted, which replaces a  $3 \times 3$  convolution with PSA in the residual network bottleneck.

First of all, the spatial pyramid convolutional (SPC) module divides the input channels and extracts the multi-scale features according to the spatial information in the feature map of each channel, which improves the multi-scale representation capability at a finer level. Secondly, the squeeze-and-excitation weight (SEWeight) module is used to extract the channel attention from the feature maps with different scales to obtain the attention vector of each channel. Then, the softmax algorithm recalibrates the attention vector features of multi-scale channels to obtain the attention weights. Finally, the dot product is processed between recalibrated weights and corresponding feature maps by elements. The EPSA module not only increases the ranges of trunk feature extraction and receptive field, but also significantly separates the important context information features. So, EPSA is superior to the existing attention modules in multi-scale cell recognition.

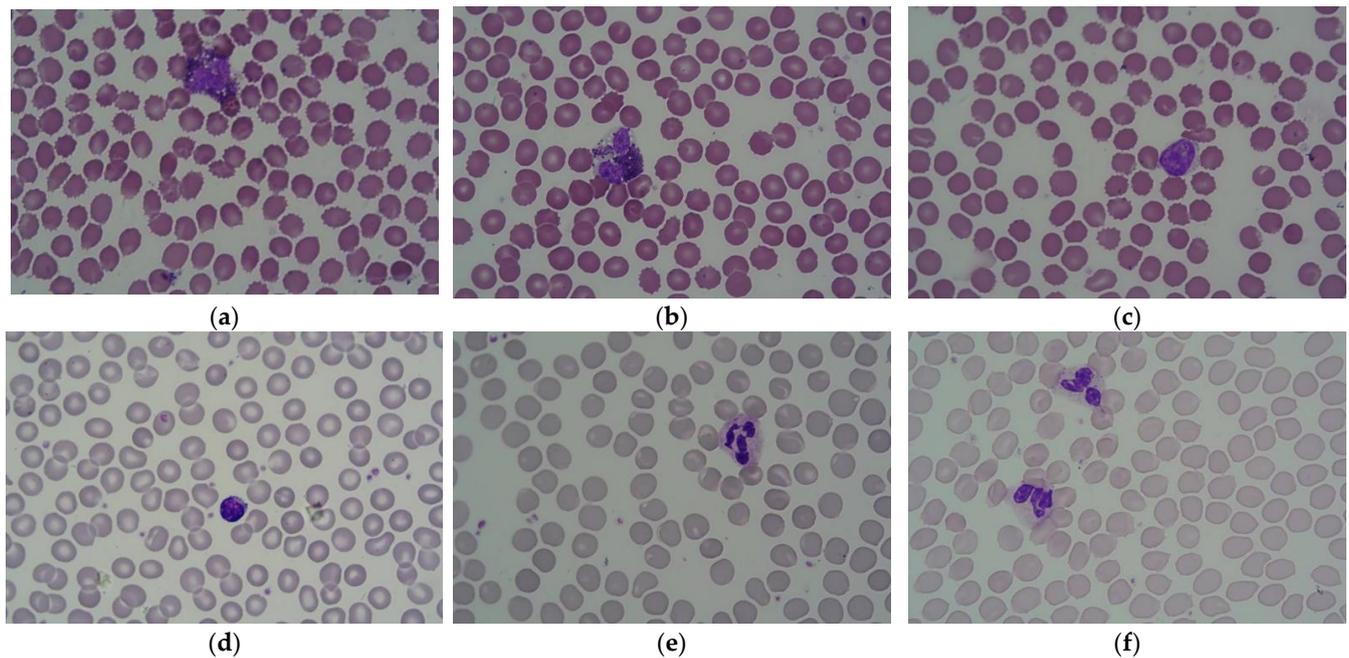
### 2.5. ROI Align Module

The input of ROI Pooling is the candidate region coordinate obtained by RPN calculation in the Faster R-CNN. Because there is only one regression process, the coordinates, mapped from the candidate region in the original map, are the floating point numbers in the feature map. However, ROI Pooling requires quantizing coordinates to integers, and the quantization errors produced twice cause the position deviation of the candidate region in the original feature map. Therefore, this process affects the accuracy of detection results. In order to resolve the problem, ROI Align [26] is chosen to replace ROI Pooling. The principle is to cancel the quantization operation and obtain the floating point coordinates by the bilinear interpolation so as to complete the whole continuous operation. When the feature information is entered into the pyramid, the last four layers of data are entered into the ROI Align layer in the prediction of output features, thereby increasing the feature extraction capability of the network.

## 3. Experiments and Analyses

### 3.1. Dataset

The dataset is derived from the Infahan Medical Image and Signal Processing (MISP) dataset. The blood smear images were taken by a Nikon ECLIPSE 50i microscope with a magnification of 100 times, in which a large number of single cells and adherent cells appear in dense distributions. These images reflect the diversity of cell morphology and the environmental complexity of actual medical scenes, affecting the accuracy of cell recognition. They are neither labeled nor preprocessed, so it is necessary to draw the boundary boxes on cell images by labelme 5.0.2 software. The processes of annotation and labeling can accurately mark and segment single cells and adherent cells to train and evaluate the model. The sample examples of original dataset are shown in Figure 3. These images were obtained in different conditions, such as different shooting angles, lighting conditions, equipment settings and other factors. In order to deal with various images in the actual scene, the original dataset is expanded to 260 images by randomly rotating, scaling, cropping and adjusting contrast and brightness. The images are  $775 \times 519$  pixels. The expanded dataset is randomly divided into a test set, validation set and training set according to the ratio of 3:3:20. Among them, the numbers of labeled single cells and adherent cells are 11,350 and 9414, respectively.



**Figure 3.** Sample dataset. (a) Sample 1; (b) Sample 2; (c) Sample 3; (d) Sample 4; (e) Sample 5; (f) Sample 6.

### 3.2. Evaluation Indexes

In order to comprehensively assess the performance of the proposed model, four evaluation indexes of recall, precision, mean average precision (mAP) and F1 score are adopted in here. All these evaluations are calculated based on the confusion matrix [27].

In target detection, the intersection over union (IoU) is the overlap rate between the candidate bounding box and ground truth box. DR indicates the candidate bounding box and GT indicates the ground truth box, and the IoU is given as:

$$IoU = \frac{S_{GT} \cap S_{DR}}{S_{GT} + S_{DR} - S_{GT} \cap S_{DR}}, \quad (2)$$

where  $S_{GT}$  represents the area of ground truth box,  $S_{DR}$  represents the area of candidate bounding box,  $S_{GT} \cap S_{DR}$  represents the intersection area of the two.

Recall is the number ratio of correctly detected targets to all real labeled targets. It is expressed as:

$$R = \frac{T_P}{T_P + F_N}, \quad (3)$$

where  $T_P$  refers to the number of true positive samples, that is, the positive samples are correctly identified as positive samples,  $F_N$  refers to the number of false negative samples, i.e., the positive samples are incorrectly identified as negative samples.

Precision is the number ratio of correctly detected targets to all predicted targets. It is given as:

$$P = \frac{T_P}{T_P + F_P}, \quad (4)$$

where  $F_P$  refers to the number of false positive samples, that is, the negative samples are incorrectly identified as positive samples.

Average precision (AP) is used to comprehensively measure the model quality by considering recall and precision. AP is described as:

$$A_t^c = \frac{1}{N} \sum_{R \in \{0, 0.01, \dots, 1\}} \max_{\tilde{R} \geq R} \left\{ P(\tilde{R}) \right\}, \quad (5)$$

where  $t$  is the threshold of IoU,  $c$  is the given category number,  $N$  is the number of confidence thresholds selected,  $\tilde{R}$  is the next recall value. Because there are only two categories, namely single type and adherent type, thus  $c = 2$ .

The mAP is the average of AP of all categories. It is calculated as:

$$m_{AP} = \frac{\sum_{i=1}^c A_i}{c}. \quad (6)$$

For cell target detection, the larger the mAP value, the better the detection performance of the model, thus the higher the recognition rate.

F1 score also combines precision and recall and is their harmonic mean. It avoids the single maximum value of precision or recall to comprehensively reflect the overall index. F1 is expressed as:

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (7)$$

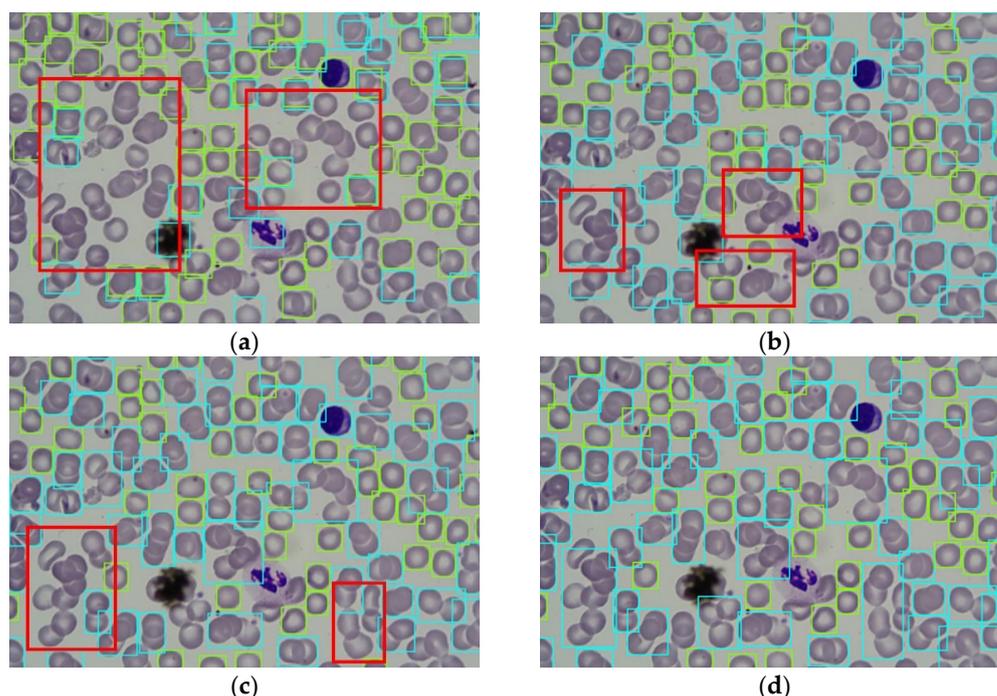
### 3.3. Result Analyses

#### 3.3.1. Results of Ablation Experiments

The ablation experiments are carried out among the original network and three improved network models, respectively. The network parameters are as follows: the momentum gradient is 0.9, the initial learning rate is 0.005, the weight attenuation is 0.0005, the batch size of the input image is 4 during training and each experiment is trained with 40 epochs. These networks all use the same dataset and training parameters. The original network selects Vgg16 as the backbone, while the three improved networks use ResNet50 as the backbone. The improved model 1 adds EPSA and ROI Align, the improved model 2 adds BFP on the basis of model 1 and the improved model 3, namely BDE\_RCNN, adds DCN on the basis of model 2. It is decided that when the IoU between the candidate bounding box and ground truth box is larger than 0.7, the sample is considered as a positive sample. Otherwise, when it is smaller than 0.3, it is considered as a negative sample. The positive and negative samples are used to train the classification function of RPN, and then the suggested areas are accurately output for the subsequent full connection layer. Figure 4 shows the detection results of four network models, in which the green marks and blue marks denote the detection results of single red blood cells and adherent red blood cells, respectively. It is seen from the detection results of the original network that there are missed cells, as shown in the red boxes in Figure 4a. This is mainly because the extraction step of the original model is larger, resulting in inadequate learning of small features in densely distributed scenes. Compared to the original network model, the three improved models significantly improve the cell detection effects. Model 1 and model 2 still have a few missing cells, as shown by the red boxes marked in Figure 4b,c. But BDE\_RCNN correctly identifies almost all single cells and adherent cells, as shown in Figure 4d. The recognition results show that the proposed BDE\_RCNN has the best effect for cell detection.

In order to more directly compare the recognition effect of each algorithm, the evaluation indexes of four models are listed in Table 1. A check indicates a selected model and a cross indicates an unselected model. The experimental datasets include three types: one is the dataset of single cells (SRBC), one is the dataset of adherent cells (ARBC) and the other is the mixture dataset of single cells and adherent cells (MIX). It is shown from recall, precision, mAP and F1 that BDE\_RCNN is better than other models in SRBC, ARBC and MIX. In SRBC, the mAP of BDE\_RCNN increases by 25.7%, 15.9% and 1.5% and F1

increases by 27.1%, 16.1% and 1.5%, respectively, compared to the original model, model 1 and model 2. In ARBC, the mAP of BDE\_RCNN increases by 54.4%, 18.6% and 2.3% and F1 increases by 54.5%, 18.8% and 2.2%, respectively, compared to the other three models. In MIX, the mAP of BDE\_RCNN also increases by 24.5%, 16.4% and 2.4% and F1 correspondingly increases by 26.6%, 17.2% and 3.2%, respectively. The experimental results indicate that the BDE\_RCNN method performs best in multiple scenes with different datasets when it adopts BFP, DCN, EPSA and ROI Align modules, especially for adherent cells with dense distribution. The proposed method effectively solves the automatic recognition problem of red blood cells under the conditions of dense distribution, extrusion deformation, adhesion and overlap.



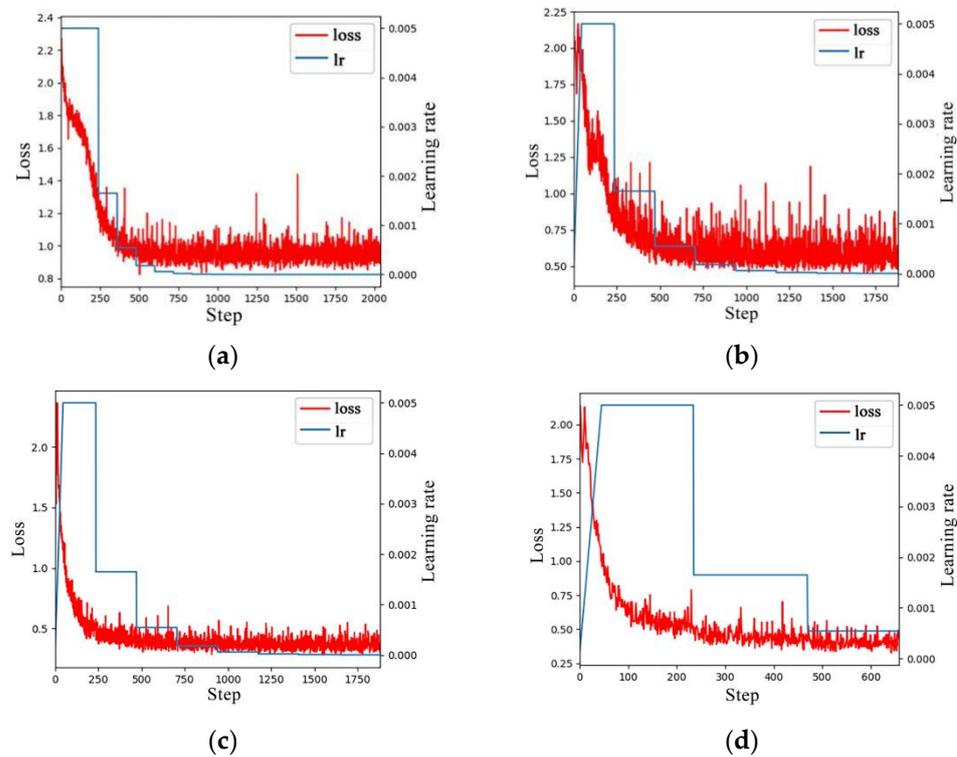
**Figure 4.** Detection results of four models. (a) Original model; (b) model 1; (c) model 2; (d) BDE\_RCNN.

**Table 1.** Index comparison results of four models.

Model	BFP	DCN	EPSA	ROI Align	Type	Recall	Precision	mAP	F1
Original model	×	×	×	×	SRBC	0.686	0.761	0.731	0.722
					ARBC	0.587	0.505	0.545	0.543
					MIX	0.664	0.754	0.719	0.706
Model 1	×	×	✓	✓	SRBC	0.772	0.811	0.793	0.791
					ARBC	0.666	0.752	0.710	0.706
					MIX	0.736	0.791	0.769	0.763
Model 2	✓	×	✓	✓	SRBC	0.886	0.923	0.905	0.904
					ARBC	0.782	0.865	0.823	0.821
					MIX	0.838	0.897	0.874	0.866
BDE_RCNN	✓	✓	✓	✓	SRBC	0.895	0.942	0.919	0.918
					ARBC	0.796	0.887	0.842	0.839
					MIX	0.868	0.921	0.895	0.894

The convergences of training loss function and the changes in model learning rate (lr) for four models are shown in Figure 5. It is shown that the first three models tend

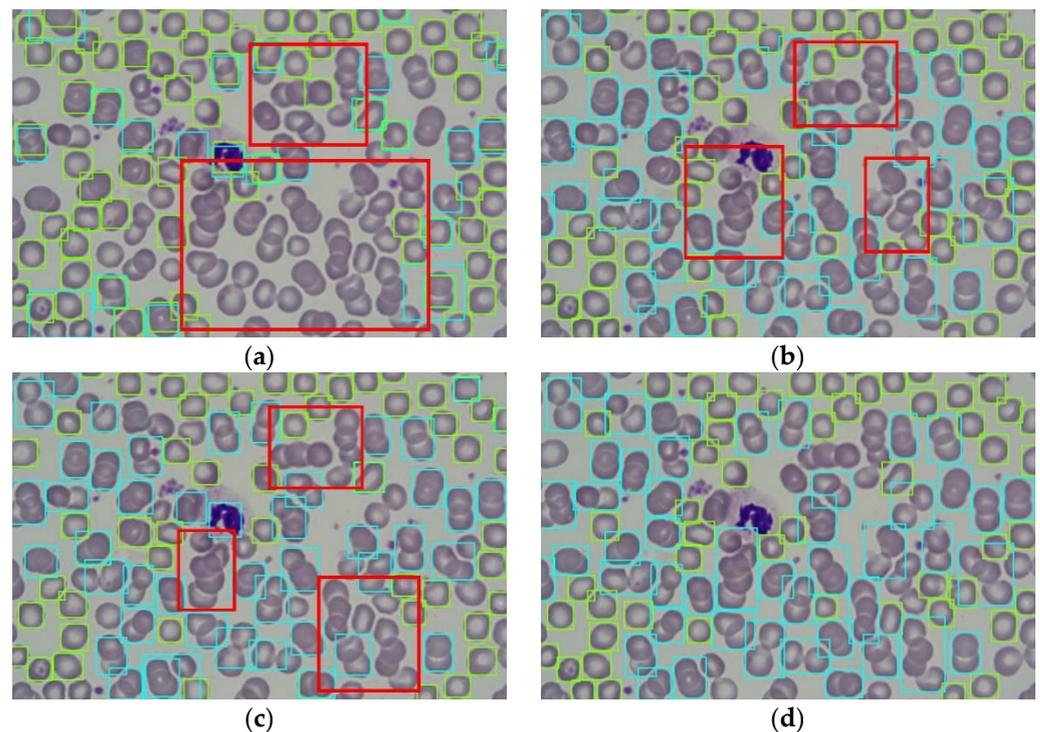
to converge at 1750 iterations, as shown in Figure 5a–c, while BDE\_RCNN converges at 600 iterations, as shown in Figure 5d. The BDE\_RCNN method is superior to other models in terms of convergence times. Thus, with the reduction of loss curve, a higher level of training model can be obtained at fewer iterations. In addition, the learning rate curve of BDE\_RCNN begins to decline when the iteration numbers reach about 250, as shown in Figure 5d. Compared with other models, BDE\_RCNN more easily approaches the optimal model solution of training parameters in the training process.



**Figure 5.** Loss/Lr curves of four models. (a) Original model; (b) model 1; (c) model 2; (d) BDE\_RCNN.

### 3.3.2. Results of Comparison Experiments

In order to further verify the effectiveness of the proposed algorithm, several single-stage mainstream models such as SSD, YOLOv5 and RetinaNet [28] are compared to BDE\_RCNN, as shown in Figure 6. Among them, SSD network is based on a convolutional network of forward propagation to generate the boundary frame with fixed size and target scores of the sample. Its final detection result is obtained by non-maximum suppression. YOLOv5 is one of the typical single-stage algorithms. By dividing the feature graph into multiple lattices and detecting the targets in each lattice, the position and category of targets within the lattices can be predicted at once. Using the FPN structure and focal loss function, the RetinaNet network solves the imbalance of positive and negative samples in the target detection network. Figure 6 shows that the three single-stage models fail to detect the whole cells and miss some cells for the same dataset, as shown by the red boxes marked in Figure 6a–c, respectively. Meanwhile, BDE\_RCNN can correctly recognize the single cells and adherent cells, as shown in Figure 6d. This is attributed to lacking the step of candidate region extraction in the single-stage network, and only the first-stage network completes the two tasks of classification and regression.



**Figure 6.** Detection results of four different models. (a) SSD; (b) YOLOv5; (c) RetinaNet; (d) BDE\_RCNN.

Table 2 shows the comparison results of SSD, YOLOv5, RetinaNet and BDE\_RCNN in the evaluation indexes. It is seen that the overall detection accuracy of the improved BDE\_RCNN is better than that of single-stage networks. For mAP and F1 indexes, BDE\_RCNN is 46.5% and 44% better than SSD, respectively. Compared to YOLOv5, they increase by 17.1% and 17.8%, respectively. Similarly, the two indexes increase by 6.9% and 7.5% compared to RetinaNet, respectively. These results indicate that BDE\_RCNN has stronger feature extraction capability compared to single-stage mainstream networks in the actual scene, thus effectively improving the automatic recognition effect for cells.

**Table 2.** Index comparison results of mainstream models and BDE\_RCNN.

Model	Recall	Precision	mAP	F1
SSD	0.586	0.659	0.611	0.620
YOLOv5	0.717	0.804	0.764	0.758
RetinaNet	0.808	0.856	0.837	0.831
BDE_RCNN	0.868	0.921	0.895	0.893

#### 4. Conclusions

In this paper, an automatic recognition method, BDE\_RCNN, of adherent cells in densely distributed scene is proposed based on Faster R-CNN. To improve the network performance, BFP, DCN and EPSA are applied to automatically recognize the cells under the conditions of dense distribution, extrusion deformation, adhesion and overlap. In addition, the region mismatch caused by quantization error is avoided by the ROI Align module. The experimental results show that BDE\_RCNN can obtain a higher level of training model with fewer iterations and its mAP and F1 values are 24.5% and 26.6% higher than that of the original Faster R-CNN, respectively. Compared with several single-stage mainstream models, BDE\_RCNN has stronger feature extraction capability, thus effectively improving the cell recognition effect. The proposed method can maximize the network recognition rate and realize the automatic recognition of single cells and adherent cells in an

actual scene. It may be applied in holographic imaging such as particle field measurement and biological cell morphology observation.

**Author Contributions:** Conceptualization, Y.L. (Yumeng Liu) and Y.L. (Yun Liu); methodology, Y.L. (Yumeng Liu); software, M.C. and X.W. (Xiaoqiang Wu); validation, H.X.; formal analysis, Y.L. (Yumeng Liu); investigation, L.S. and J.X.; writing—original draft preparation, Y.L. (Yumeng Liu); writing—review and editing, Y.L. (Yun Liu), X.W. (Xian Wang) and H.L.; supervision, M.J.; project administration, Y.L. (Yun Liu). All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China (61805195, 51875455, 52205067, 52004213), Key Research and Development Plan Project in Shaanxi Province (2023-YBGY-400), Natural Science Basic Research Program of Shaanxi (2020JQ-625), Xi'an Science and Technology Plan Project (22GXFW0089), Seed Fund for Creativity and Innovation of Postgraduates of Xi'an University of Technology (252082206), 2023 Training Program of Innovation and Entrepreneurship for Undergraduates (202310700045).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** A publicly available dataset was analyzed in this study. These data can be found here: <http://www.coder100.com/index/index/content/id/1569951> (accessed on 27 November 2021).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Vala, M.H.J.; Baxi, A. A review on otsu image segmentation algorithm. *Int. J. Adv. Res. Comput. Eng. Technol.* **2013**, *2*, 387–389.
2. Haris, K.; Efstratiadis, S.N.; Maglaveras, N.; Katsaggelos, A.K. Hybrid image segmentation using watersheds and fast region merging. *IEEE Trans. Image Process.* **1998**, *7*, 1684–1699. [[CrossRef](#)] [[PubMed](#)]
3. Yi, F.; Moon, I.; Javidi, B. Cell morphology-based classification of red blood cells using holographic imaging informatics. *Biomed. Opt. Express.* **2016**, *7*, 2385–2399. [[CrossRef](#)] [[PubMed](#)]
4. Gao, C.; Zhou, D.; Guo, Y. An iterative thresholding segmentation model using a modified pulse coupled neural network. *Neural Process. Lett.* **2014**, *39*, 81–95. [[CrossRef](#)]
5. Sundara, S.M.; Aarthi, R. Segmentation and evaluation of white blood cells using segmentation algorithms. In Proceedings of the 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 23–25 April 2019; pp. 1143–1146.
6. Zhang, Y.C.; Xu, N.; Chen, H.M.; Lam, W.H.; Zhang, X.; Qiu, T. A robust and high-performance white blood cell segmentation algorithm. In Proceedings of the IEEE 2020 5th International Conference on Computer and Communication Systems (ICCCS), Shanghai, China, 15–18 May 2020; pp. 347–351.
7. Harris, C.G.; Stephens, M.J. A combined corner and edge detector. In Proceedings of the 4th Alvey Vision Conference, Manchester, UK, 31 August–2 September 1988; pp. 147–151.
8. Tang, Z.J.; Huang, L.Y.; Zhang, X.Q.; Lao, H. Robust image hashing based on color vector angle and Canny operator. *Int. J. Electron. Commun.* **2016**, *70*, 833–841. [[CrossRef](#)]
9. Zhang, Y.; Han, X.; Zhang, H.; Zhao, L.M. Edge detection algorithm of image fusion based on improved sobel operator. In Proceedings of the 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC 2017), Chongqing, China, 3–5 October 2017; pp. 457–461.
10. Ma, Y.L.; Ma, H.Y.; Chu, P.C. Demonstration of quantum image edge extraction enhancement through improved sobel operator. *IEEE Access* **2020**, *8*, 210277–210285. [[CrossRef](#)]
11. Tai, W.L.; Hu, R.M.; Hsiao, H.C.W.; Chen, R.M.; Tsai, J.J.P. Blood cell image classification based on hierarchical SVM. In Proceedings of the 2011 IEEE International Symposium on Multimedia, Dana Point, CA, USA, 5–7 December 2011; pp. 129–136.
12. Rezatofighi, S.H.; Soltanian-Zadeh, H. Automatic recognition of five types of white blood cells in peripheral blood. *Comput. Med. Imag. Grap.* **2011**, *35*, 333–343. [[CrossRef](#)] [[PubMed](#)]
13. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
14. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
15. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single shot multibox detector. *arXiv* **2015**, arXiv:1512.02325.

16. Han, H.Z.; Wei, B.; Sui, D.; Li, S. A U-Net-based method for detection of cancer cells in pathological sections of breast cancer. *J. Precis. Med.* **2018**, *33*, 471–473.
17. Banik, P.P.; Saha, R.; Kim, K.D. An automatic nucleus segmentation and CNN model based classification method of white blood cell. *Expert Syst. Appl.* **2020**, *149*, 113211. [[CrossRef](#)]
18. Chen, L.; Xue, H.R.; Gao, X.J.; Zhang, X.L.; Bai, J. Milk somatic cells recognition based on dichotomy method and BP neural network. *J. Inner Mong. Agric. Univ. (Nat. Sci. Ed.)* **2020**, *41*, 69–74.
19. Lavitt, F.; Rijlaarsdam, D.J.; van der Linden, D.; Weglarz-Tomczak, E.; Tomczak, J.M. Deep learning and transfer learning for automatic cell counting in microscope images of human cancer cell lines. *Appl. Sci.* **2021**, *11*, 4912. [[CrossRef](#)]
20. Anand, K.; Meena, J.; Srinivas, B.; Animesh, M.; Girija, C.; Praveen, K.S.; Sanjib, M.; Timir, K.P.; Jyoti, M.; Ajat, H. Deep learning for real-time malaria parasite detection and counting using YOLO-mp. *IEEE Access* **2022**, *10*, 102157–102172.
21. Pang, J.M.; Chen, K.; Shi, J.P.; Feng, H.J.; Ouyang, W.L.; Lin, D.H. Libra R-CNN: Towards balanced learning for object detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 821–830.
22. Dai, J.F.; Qi, H.Z.; Xiong, Y.W.; Li, Y.; Zhang, G.D.; Hu, H.; Wei, Y.C. Deformable convolutional networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 764–773.
23. Zhang, H.; Zu, K.; Lu, J.; Zou, Y.; Meng, D.Y. EPSANet: An efficient pyramid squeeze attention block on convolutional neural network. *arXiv* **2021**, arXiv:2015.14447.
24. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.M.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 936–944.
25. Wang, X.; Girshick, R.; Gupta, A.; He, K.M. Non-local neural networks. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
26. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. [[CrossRef](#)] [[PubMed](#)]
27. Yue, W.; Liu, S.; Li, Y. Eff-PCNet: An efficient pure CNN network for medical image classification. *Appl. Sci.* **2023**, *13*, 9226. [[CrossRef](#)]
28. Tsung, Y.L.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.