



Research article

Faster R-CNN with improved anchor box for cell recognition

Tingxi Wen^{1,2,3}, Hanxiao Wu¹, Yu Du¹ and Chuanbo Huang^{1,*}

¹ College of Engineering, Huaqiao University, Quanzhou 362021, China

² Postdoctoral Workstation of Linewell Software Company Limited, Quanzhou 362000, China

³ Nlinks Technology Group Co., LTD, China

* **Correspondence:** Email: huangchuanbo@hqu.edu.cn.

Abstract: As the basic units of the human body structure and function, cells have a considerable influence on maintaining the normal work of the human body. In medical diagnosis, cell examination is an important part of understanding the human function. Incorporating cell examination into medical diagnosis would greatly improve the efficiency of pathological research and patient treatment. In addition, cell segmentation and identification technology can be used to quantitatively analyze and study cellular components at the molecular level. It is conducive to the study of the pathogenesis of diseases and to the formulation of highly effective disease treatment programs. However, because cells are of diverse types, their numbers are huge, and they exist in the order of micrometers, detecting and identifying cells without using a deep learning-based computer program are extremely difficult. Therefore, the use of computers to study and analyze cells has a certain practical value. In this work, target detection theory using deep learning is applied to cell detection. A target recognition network model is built based on the faster region-based convolutional neural network (R-CNN) algorithm, and the anchor box is designed in accordance with the characteristics of the data set. Different design methods influence cell detection results. Using the object detection method based on our novel faster R-CNN framework to detect the cell image can help improve the speed and accuracy of cell detection. The method has considerable advantages in dealing with the identification of flowing cells.

Keywords: faster R-CNN; anchor box; convolutional neural network; object detection; cell recognition

1. Introduction

With the rapid development of science in recent years, computer technology has achieved many breakthroughs in image processing, which is being used increasingly in the biomedical field; such usage greatly promotes the development of medical image analysis [1]. At present, the prevention, detection, and treatment of many diseases are inseparable from biomedical image analysis. Biomedical image processing technology refers to the use of computer hardware to conduct digital processing of biomedical images and subsequently perform segmentation, recognition, analysis, interpretation, and other operations on digital medical images. Afterward, we can obtain the required feature information, which is then used for subsequent analysis and transmission. Cell segmentation and recognition can be used to conduct quantitative analyses of and research on cell components at the molecular level; they are conducive to the study of the pathogenesis of diseases and development of effective treatment programs for diseases [2].

Cell research is crucial in biomedicine. Cell image analysis can quickly obtain cell state information, which is important for the study and diagnosis of pathology. However, due to the diversity in the types of cells, cells are affected by cell surface tension [3], cytoplasm viscosity [4], cell membrane stiffness [5], and other aspects. Different cell scales, including spherical, rod-shaped, doughnut-shaped, and cubical, vary greatly because of the adaptation of cells to their functions. Cell-cell adhesion often occurs between cells [6–10] and result in cell stacking. In addition, the grayscale, clarity, blur, and brightness of cell images are affected by the lighting environment.

On the basis of these cell characteristics, cells have been used for medical image recognition and problem analysis. With an object detection algorithm, we can achieve the automatic identification, classification, and positioning of medical images of cells. Hence, detection precision and medical diagnosis can be improved, detection tasks can be completed reliably, and the development of medical technology can be promoted.

Currently, cell classification mainly utilizes methods based on nuclear features [11] and image clipping [12] and can be achieved with the convolutional neural network (CNN) [13].

Several segmentation algorithms have been developed in recent years to solve the aforementioned problem. Kang et al. [14] proposed a cell image segmentation method that combines image segmentation, multi-resolution, and Babbitt distance technologies. With a multi-scale framework, the low contrast problem of bright-field cell images was solved by determining the difference in cell and background distribution intensities. Chen et al. [15] presented a non-negative constrained principal component tracking model, combined the model with spectral clustering, and used standardized cutting as a measure of segmentation quality to realize the effective segmentation of bright-field cell images.

Cell segmentation based on morphology mainly adopts the geometry or gray characteristics of cells to achieve cell segmentation. This method does not require cell labeling for cell images and is suitable for the segmentation of small data sets. However, this method relies on the consistency of cell images; its segmentation efficiency decreases in the presence of noise.

Cell image segmentation based on deep learning has been introduced gradually through the rapid development of artificial intelligence. Ronneberger et al. [16] proposed the U-Net network, which can be used to solve the biomedical image segmentation problem of small data sets. Naylor et al. [17] utilized and compared three network structures, namely, FCN, U-Net, and mask R-CNN, to segment histological sections stained with H&E. They described the segmentation problem as a regression

task to predict the distance transformation of binarization images. Kromp et al. [18] compared and evaluated the segmentation performance of U-Net, U-Net with ResNet34 skeleton, Deepcell, and mask R-CNN. They trained and tested fluorescence core images of different samples, sample preparation types, marking quality, image scale, and segmentation complexity. They found that three of these deep learning frameworks can be used to segment fluorescence nuclear images on most sample preparation types and tissue sources.

Cell segmentation based on deep learning relies on numerous annotated data sets. However, compared with cell segmentation based on morphology, it has higher accuracy and segmentation efficiency.

In 2019, a study by Google showed that using CNN can realize the detection of breast cancer metastasis in lymph nodes and demonstrates ideal precision. In particular, Google introduced the augmented reality microscope [19] platform. This platform adopts the deep learning model; it was established based on TensorFlow and uses various machine learning algorithms. Hence, it can be utilized to solve problems in the classification, identification, and quantification of tumor cells and exerts a considerable influence on medical diagnosis [20].

On the basis of these studies, we implement the framework pertaining to Python Keras by using faster R-CNN [21] in blood cell recognition. The deep learning method is proven to be feasible in cell image classification and feature extraction of the original faster R-CNN. Network adjustment based on the proposed network regions of the anchor box allows the original faster R-CNN to be improved.

This experiment studies the use of the object detection algorithm based on deep learning and aims to integrate feature extraction [22], identification of regions of interest (ROIs), classification, and positioning into a network model [23]. We then adjust the network in accordance with the characteristics of cell images in the data set to reduce human intervention in the training process.

As shown in Figure 1, the cell that is captured in the image can be located and labeled (the rectangular box) using end-to-end target detection network model prediction. The category of each label can also be predicted.

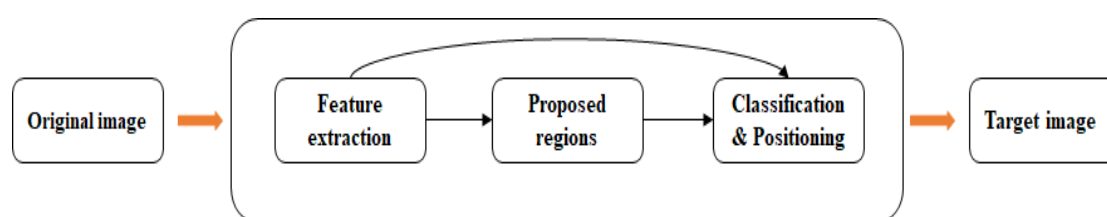


Figure 1. End-to-end cell detection framework.

2. Methods

2.1. Experiment methods

Faster R-CNN improves the extraction of candidate regions [24]. It uses CNN to obtain the region box and realizes end-to-end model training. Therefore, the faster R-CNN network model is used as the basic framework in this experiment to construct an end-to-end cell detection model.

The original network structure and parameters are adjusted in accordance with the characteristics of cell images to realize end-to-end training of the network model by using cell data set images. High detection precision is expected on the basis of the test set.

2.1.1. Feature extraction network

In literature [20], the faster R-CNN model utilizes the VGG16 [25] network as the backbone for feature extraction. The VGG16 network contains 13 convolutional, 13 activation, 4 pooling, and 3 full-connection layers. In the faster R-CNN model, only the convolutional, activation, and pooling layers are used for feature extraction.

However, in accordance with the characteristics of cells, we know that cells are of various types, and the morphological differences between cells are variable. Therefore, a detailed classification is required for cell detection. In addition, cell detection is often used in biomedical image analysis, which requires high precision.

A deep network is advantageous in learning and expressing strong semantic information, and acquiring robustness to changes in the shape and position of objects is easy. However, as the number of network layers increases, convergence difficulties arise, and the network performance deteriorates with the deepening of the network. We know that ResNet [8] can solve such a problem. ResNet uses a residual learning module, as shown in Figure 2, to help the network achieve identity mapping. When the internal features of a certain layer of the network have been optimized, the network after that layer does not change the features.

In the residual learning module [24], the output of shallow network f_1 is superimposed with the output of deep network f_2 as the input of deeper network f_3 . When shallow network f_1 is optimal, the output of f_2 approaches 0 to ensure that the loss does not increase and realize constant mapping of the network.

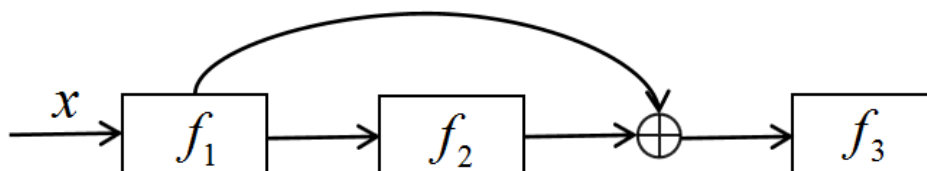


Figure 2. Residual learning module.

Therefore, in accordance with literature [8], this experiment changed the feature extraction network to ResNet50 and processed the data and images by using a deeper CNN to improve the model effect and enhance the precision of cell detection.

2.1.2. Anchor box design

The feature can be regarded as a 256-channel image with a scale of 51×39 . For each position of the image, 9 possible candidate windows are considered; the three areas are 128×128 , 256×256 , and 512×512 , and each area is divided into 3 aspect ratios, namely, 2:1, 1:2, and 1:1. These

candidate windows are called anchors. The author introduced the multi-scale methods commonly used in detection (detecting targets of various sizes) through these anchors.

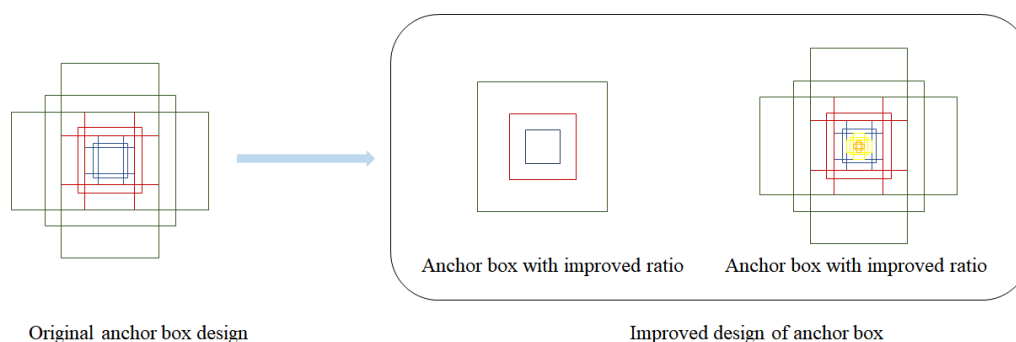


Figure 3. Comparison of the anchor box before and after adjustment.

For these 51×39 positions and $51 \times 39 \times 9$ anchors, the following shows the calculation steps for each position:

- (a) Let k be the number of anchors corresponding to a single position. At $k = 9$, the area suggestion function is completed by adding a 3×3 sliding window operation and two convolutional layers.
- (b) The first convolutional layer encodes each sliding window position of the feature map into a feature vector, and the second convolutional layer outputs k area scores for each sliding window position, indicating the probability that the anchor at that position is an object. The total output length of this part is $2 \times k$ (one anchor corresponds to two outputs: the probability of being an object + the probability of not being an object) and k regression recommendations (frame regression). One anchor corresponds to four frame regression parameters. Therefore, the total output length of the box regression part is $4 \times k$, and performing non-maximum suppression on the score area or the output score Top-N (300 in the text) area tells the detection network which areas should be paid attention to. Essentially, it implements the selective functions of Search, EdgeBoxes, and other methods.

In the training stage of the RPN network, the feature images are scanned by sliding windows, and different anchor boxes of a specified scale and proportion are obtained in the center of each sliding window. This anchor box participates in the regression operation of boundary boxes to obtain candidate regions. Therefore, in the training stage of the RPN network, increasing the anchor box size is beneficial for detecting objects with large scale differences. By observing the shape of the detected target, the anchor box length and width can be adjusted to an appropriate proportion, which is conducive to improving the final detection precision. Increasing or decreasing the number of anchors has a certain effect on time performance.

In the study that proposed the faster R-CNN method, the author describes the anchor box as follows: “The anchor point is located at the center of the sliding window in question and is related to a scale and aspect ratio. By default, we use 3 scales and 3 aspect ratios, and $k = 9$ anchor points are generated at each sliding position.” That is, the value is selected according to the data set. On the basis of the morphology of biological blood cells, we know that red blood cells are double-concave and disc-shaped, white blood cells are colorless and nucleated spherical, platelets are very small (with a diameter of 2–4 μm) and biconvex flat discs, and eosinophils are spherical. That is, most blood cells are spherical. An important feature of a sphere or a circle is that the distance from any

point on the circumference to the center of the circle is a radius r , that is, during the sliding process of the window, the length-to-width ratio of the blood cell is approximately 1:1. Moreover, Figure 3 indicates that the area of the candidate frame determines the area of the sliding window to be detected, and the total output length of the frame regression part is $4 \times k$. Thus, when the scale of the candidate frame changes, the area of the sliding window also changes; when the ratio and scale of the candidate frame change, the total output length of the frame regression part also changes. We reduced the area of the preset candidate frame in accordance with the characteristics of the small platelet volume.

Organisms have cells of different sizes and shapes, but the size differences of cells of the same kind are usually small. In this experiment, we adjusted the number, scale, and aspect ratio of the anchor box in accordance with different cell sizes. The design of the anchor box is expected to improve the performance of network detection of large and small objects.

3. Experiment

3.1. Data set

A data-enhanced public blood cell data set was used in this study. The original data set contains 364 blood cell pictures, and the pixel size of each picture is 640×480 . The data set has three categories, namely, red blood cells (RBCs), white blood cells (WBCs), and platelets. Several database samples are shown in Figure 4.

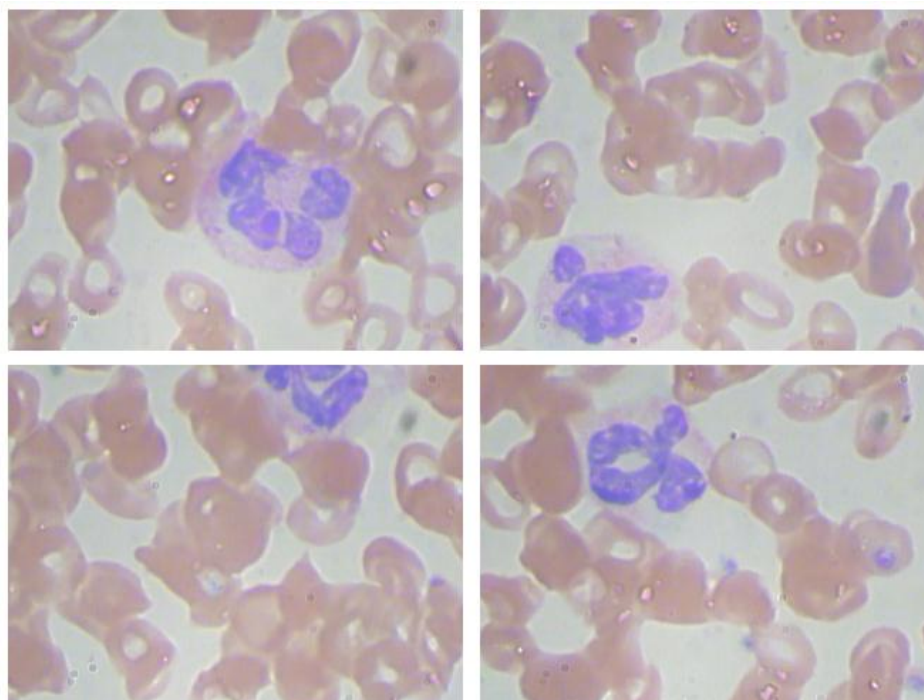


Figure 4. Database samples.

Through careful collation of the data set, we found that the data set has the following characteristics.

(a) In the actual composition of human blood, RBCs outnumber other blood cells. Therefore, in the data image, RBSs are attached to one another and tend to overlap, whereas WBCs and platelets usually exist discretely.

(b) Given that different cells often have different shapes, the data images show that large-scale color differences exist among the three types of cells, but the differences between cells of the same type are small. The platelet size is the smallest, and the WBC size is the largest.

(c) The images in this data set are biomedical images collected in the laboratory by using professional medical image acquisition equipment. Therefore, the brightness and darkness of each image are similar, minimally affected by the environment, and have a high degree of similarity. The image background is consistent and relatively simple.

(d) Incomplete cell images exist at the edge of the image due to the limitation in the field of vision or image clipping.

(e) The number of images in the data set is small, and the number of samples is limited.

In addition to the 364 blood cell images in the original image data file, each image is also matched with a mark file to mark the position and category information in the corresponding blood cell image.

To prevent the model from overfitting and increase the amount of data in the data set, we enhanced the data set by randomly cropping the image data, flipping left and right, color dithering, rotating, and zooming. We enhanced the data set to 10,000 and matched the corresponding marked files.

3.2. Experimental environment

This experiment was based on the Keras deep learning framework, and we used the Ubuntu 16.04.6 system in our program. The GPU model was based on NVIDIA RTX2080 Ti. The experiment also used the softmax optimizer. The error was divided into RPN and fast RCNN losses, and the final loss was the sum of the two.

3.3. Experimental strategy

The faster R-CNN model with ResNet50 as the backbone was used for training. The MAP of the model was tested on the test set. The time performance was analyzed, and the advantages and disadvantages of the model were identified.

3.3.1. Experiments based on static pictures

Specific experimental strategies were implemented as follows:

(a) In the training stage, the model was iterated 1000 times on the training set, with a learning rate of 0.0001. In the RPN network, the anchor box size had default values of 128, 256, and 512, and the length and width ratios were 1:1, 1:2, and 2:1, respectively. In addition, the IOU intersection ratio between the regression box and the real box was set as the background when it was less than 0.3 and

set as the foreground when it was greater than 0.7. The maximum and minimum thresholds of the regression box classification score were set to 0.5 and 0.1, respectively.

(b) By observing the image characteristics of the data set, we found that the cell morphology was mostly round. Therefore, on the basis of (a), the length/width ratio of the anchor box was modified to 1:1 to reduce the number of anchor boxes.

(c) On the basis of (a), the anchor box size was increased and adjusted to [16,32,64,128,256].

The results of the three groups of experiments were compared to determine the influence of different anchor box adjustment strategies on the cell detection results.

3.3.2. Experiments based on dynamic video

Using weight files trained in the experiments based on static pictures, we inputted blood cell video at a rate of 5 frames per second, performed cell recognition operations on the video, and took screenshots of the recognition image results every second to determine the effect of video detection of cell flow.

3.4. Experimental results and analysis

3.4.1. Mean average precision

Mean average precision (MAP), the average accuracy of the main set, is the sum of the average accuracies of all subjects. MAP is a single-value indicator that reflects the performance of the system on all related documents. The higher the number of related documents retrieved by the system is (the higher the rank is), the higher MAP is. If the system does not return relevant documents, then the accuracy rate defaults to 0.

MAP is an indicator commonly used in object detection to evaluate the quality of a model. MAP is the average precision value of multiple verification set individuals, and it is used as an indicator to measure the accuracy of object detection. The following briefly introduces the calculation method of this indicator. Two calculation methods can be applied to MAP. First, the interval [0,1][0,1][0,1] is divided, where *Recall* is divided into 11 equal parts, namely, $Recall = 0, 0.1, 0.2, \dots, 0.9, 1$. Each value of r_i is taken as the lower limit corresponding to a *Recall* value greater than or equal to threshold r_i in the statistical table. *Precision*'s maximum value is p_i , so we have 11 *Precision* values, namely, $\{p_0, p_1, \dots, p_{10}\}$.

Then, $AP = \frac{1}{11} \sum_{i=0}^{11} p_i$, and MAP is the average of the APs of all categories. The mathematical representation of this process is as follows:

$$mAP = \frac{1}{|Class|} \sum_{i=0}^{|Class|} AP_k, \quad (1)$$

$$AP_k = \frac{1}{11} \sum_{i=0}^{11} p_i, \quad (2)$$

$$p_i = \max_{r_j^i \in \{r'_0, r'_1, \dots, r'_{N-1}\}} \{p_j \mid r'_j \geq r_i\}. \quad (3)$$

Second, the area under the PR curve is directly calculated as AP, and MAP is the average of the AP values. This study uses the second method to calculate MAP.

3.4.2. Precision

To comprehensively evaluate the impact of the proposed improvement strategies on detection accuracy, this experiment used ResNet50 as the main engine, and multiple comparative tests were conducted to verify the effectiveness of the improvement strategies and evaluate the model. The experimental results are shown in Table 1.

Table 1. Comparative experiment on different improvement strategies.

	Anchor box number	Anchor box dimension	Anchor box ratio	MAP
No. 1	9	[128,256,512]	1:1,1:2,2:1	93.94%
No. 2	3	[128,256,512]	1:1	94.25%
No. 3	15	[16,32,64,128,256]	1:1,1:2,2:1	90.98%

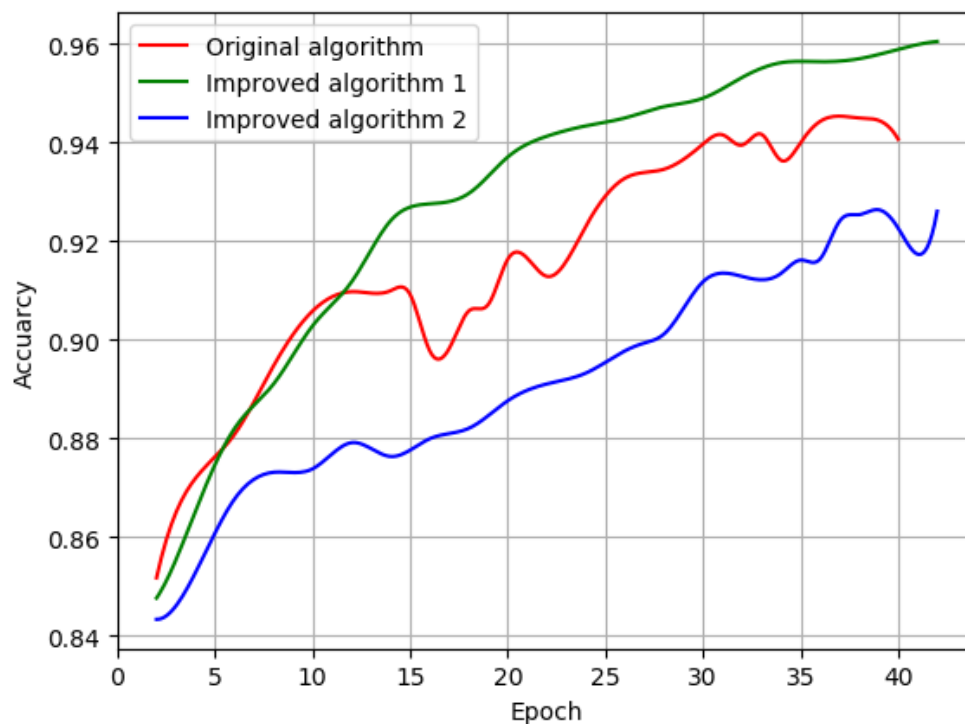


Figure 5. Curves of detection accuracy variation.

As indicated in Table 1, the ResNet50-based faster R-CNN model was adopted without adjusting the anchor box, and its MAP was 93.94%. When the number of anchor boxes was modified, the ratio of the adjusted anchor box was 1:1, and MAP was 94.25%, which was 0.31% higher than that of the original model. When the anchor box size was adjusted, MAP decreased to 90.98% in comparison with that of the original model.

We marked the algorithm after adjusting the proportion of the anchor box as the improved Algorithm 1 and the algorithm after adjusting the anchor box scale as the improved Algorithm 2. Figure 5 shows the curves of detection accuracy variation under the different experimental strategies. Before and after adjusting the anchor box, in the case of epoch, the accuracy of the three had a certain difference in the rate of increase. When the ratio of the anchor box was adjusted, the rate of increase in accuracy was the fastest, but it was the slowest when the scale of the anchor box was adjusted. Adjusting the anchor box ratio improved the time performance and recognition accuracy of the algorithm. Combined with the recognition accuracy of a single type, after adjusting the scale of the anchor box, the accuracy increased slowly, and the recognition of small objects showed higher accuracy.

3.4.3. Time performance

To compare the time performance of different methods, we compared the detection time required to complete the detection of 70 blood cells on the test set by using the ResNet50-based faster R-CNN model with a modified anchor box scale and proportion. The results are shown in Table 2.

Table 2. Time performance analysis.

Method	Mean test time(s)	Total test time(s)
Faster R-CNN (ResNet50)	0.852	59.681
Faster R-CNN (ResNet50+Reduce the anchor box)	0.747	52.312
Faster R-CNN (ResNet50+Add anchor box)	2.507	175.473

As shown in Table 2, the cell data set was detected using the ResNet50-based faster R-CNN model. When the anchor box was not adjusted, the average detection time of each image was 0.852 s, and the detection cost of all 70 test sets was 59.681 s. When the anchor box was adjusted and the number of the anchor box was reduced, the detection time of each image was 0.747 s on the average, which was 0.105 s less than that of the original model. The detection time of all data sets was 52.312 s, which was 7.369 s less than that of the original model. The model test speed decreased significantly by adjusting the scale and increasing the number of the anchor box. The average detection time for each picture was 2.507 s, and the test cost of all test sets was 175.473 s.

3.4.4. Detection effect diagram

Figure 6 shows the detection effect of the faster R-CNN model on the same cell image under the strategies of unchanged anchor box, adjusted proportion of anchor box, and adjusted scale of anchor

box. Comparison of the detection effects indicated that the ratio of the adjusted anchor box was 1:1. The detection of stacked cells was not obvious, but the discrete cells were easy to detect. With an increased anchor box size [16,32,64,128,256], small target platelets and stacked cells could be easily detected.

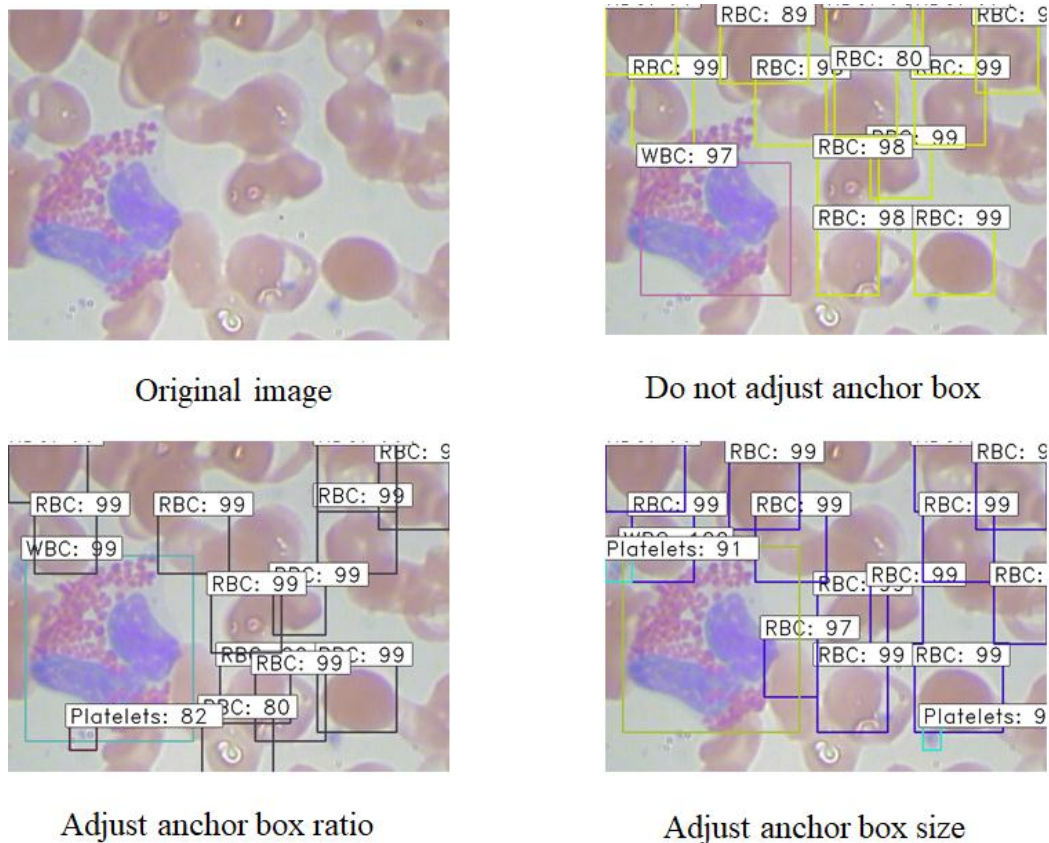


Figure 6. Inspection effect before and after anchor box adjustment.

Figure 7 shows the results of video stream detection. The method also had a strong effect on cell video stream detection, which was almost similar to static picture detection.

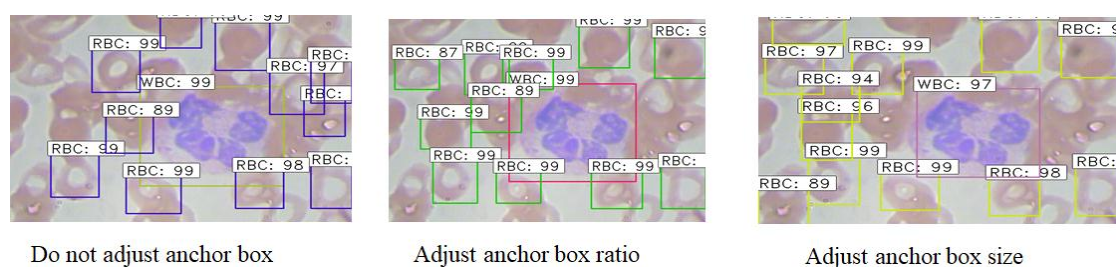


Figure 7. Cell flow video detection results.

3.4.5. Loss function comparison

Figure 8 shows the loss function curves of the different experimental strategies. Before and after adjusting the anchor box, in the case of epoch, the drop rate of the three did not differ much.

However, the loss when the anchor box ratio was adjusted decreased relative to when the ratio was not adjusted. After the scale, the time required to reach the corresponding level increased slightly.

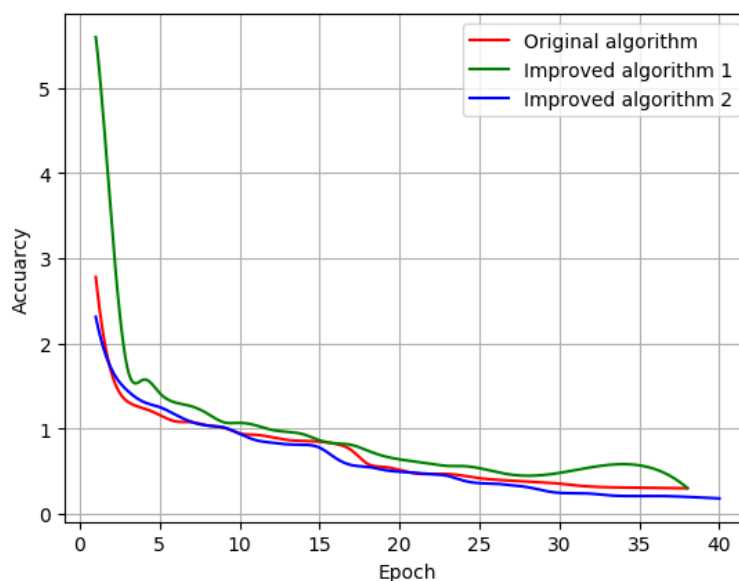


Figure 8. Loss function curves of the different experimental strategies.

3.4.6. Experimental analysis

The experimental results of the three experimental strategies of not adjusting the anchor box, adjusting the scale of the anchor box, and adjusting the proportion of the anchor box revealed the following points.

(a) Adjusting the anchor box ratio of 1:1 in accordance with the cell morphology ratio features greatly improved the detection speed and accuracy while reducing the number of generated anchor boxes. However, when the detection effect image was observed, the performance in detecting stacked cells was poor when the anchor box ratio was adjusted to 1:1. Morphology-based analytical methods, such as the watershed algorithm, can be considered in subsequent studies to improve the detection performance in stacked cells.

(b) The detection effect image showed that modification of the anchor box scale resulted in the detection of numerous small targets, such as platelets, and exhibited good performance in detecting stacked cells. However, due to the increase in the number of anchor boxes, the detection took a long time. In addition, the labels of cell data sets are often obtained through manual labeling, and the precision of labeling depends on the professionalism of workers who are responsible for the marking work. Moreover, small targets are often difficult to observe, resulting in a low precision of platelet labeling in cell data sets. Therefore, modification of the anchor box scale does not perform well in terms of accuracy.

4. Discussions

When the anchor box ratio is adjusted from the default 1:1, 1:2, and 2:1 to 1:1 in accordance with the morphological characteristics of cells, the cell shape becomes relatively fixed. Hence, after

the anchor box ratio is adjusted to the data used in the experiment, the anchor box only needs to search one scale when traversing the image, and this feature increases the search speed. Given that the focus is on the part of the search that is in line with the target graphics features, the method's recognition of cells is improved in the same round in terms of accuracy, thereby improving MAP. The time performance experiment shows that the improved anchor box ratio can enhance MAP and time performance to a certain extent. The ratio of the anchor box is adjusted to [16,32,64,128,256] to detect small target platelets that are difficult to detect. The number of anchor boxes increases accordingly due to the increase in the ratio. The reduced computing power decreases MAP and time performance. At the same time, due to the addition of small scales, the resolution of small objects and cell stacks is clear, making the detection of small target platelets and stacked RBCs easy.

This study proposes a method to improve the efficiency of blood cell and incomplete cell stack detection in medical images. The method optimizes the network structure to extract cell features effectively and appropriately reduces the time performance of model detection.

5. Conclusions

This study proposes an improved model of faster R-CNN for specific problems in blood cell detection. Experimental results show that the improved anchor box ratio can enhance MAP and time performance. After adjusting the size of the anchor box, the performance in MAP and time is reduced to a certain extent, but small targets and accumulated RBCs can be easily detected. Detection in a cell flow video can also achieve relatively good MAP and time performance.

Owing to the limitation in the data set sample and time, the experimental model still has a large room for improvement in terms of test precision and time performance. In view of these problems, on the basis of the training model for large data sets, labels can be added in the future to improve the test accuracy and obtain a more applicable training model. By using a large number of data sets, anchor box design, multi-scale feature fusion, watershed algorithm, and other methods could also be adopted to improve the detection precision for cells of different scales and stacked cells.

With the introduction of new algorithms and the acquisition of massive data sets in the future, this method has the potential to become an important part of medical image analysis in cell segmentation. In addition to cell recognition, the framework could be extended to other areas related to image object recognition.

Acknowledgments

This work was supported by Natural Science Foundation of Fujian Province (No.2020J01086), Science and Technology Program of Quanzhou (No.2019C108) and Fujian Provincial Big Data Research Institute of Intelligent Manufacturing.

Conflict of interest

No conflict of interest exists in the submission of this manuscript, and manuscript is approved by all authors for publication. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for

publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

References

1. S. D. Olabariaga, J. G. Snel, C. P. Botha, R. G. Belleman, Integrated support for medical image analysis methods: from development to clinical application, *IEEE Trans. Inf. Technol. Biomed.*, **11** (2007), 47–57.
2. F. Yang, M. A. Mackey, F. Ianzini, G. M. Gallardo, M. Sonka, Segmentation and quantitative analysis of the living tumor cells using Large Scale Digital Cell Analysis System, *Med. Imaging 2004: Image Proc.*, **5370** (2004), 1755–1763.
3. A. M. Bilek, K. C. Dee, D. P. Gaver III, Mechanisms of surface-tension-induced epithelial cell damage in a model of pulmonary airway reopening, *J. Appl. Physiol.*, **94** (2003), 770–783.
4. K. Luby-Phelps, Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area, *Int. Rev. Cytol.*, **192** (2000), 189–221.
5. A. Csiszár, B. Hoffmann, R. Merkel, Double-shell giant vesicles mimicking gram-negative cell wall behavior during dehydration, *Langmuir*, **25** (2009), 5753–5761.
6. K. K. L. Wong, Three-dimensional discrete element method for the prediction of protoplasmic seepage through membrane in a biological cell, *J. Biomech.*, **65** (2017), 115–124.
7. K. K. L. Wong, G. Fortino, D. Abbott, Deep learning-based cardiovascular image diagnosis: a promising challenge, *Future Gener. Comput. Syst.*, **110** (2020), 802–811.
8. K. K. L. Wong, J. Wu, G. Liu, W. Huang, D. N. Ghista, Coronary arteries hemodynamics: effect of arterial geometry on hemodynamic parameters causing atherosclerosis, *Med. Biol. Eng. Comput.*, 2020.
9. B. S., Gardiner, K. K. L. Wong, G. R. Joldes, A. J. Rich, C. W. Tan, A. W. Burgess, et al., Discrete element framework for modelling extracellular matrix, deformable cells and subcellular components, *PLoS Comput. Biol.*, **11** (2015), e1004544.
10. R. G. Joldes, K. K. L. Wong, D. W. Smith, C. W. Tan, B. S. Gardiner, Controlling seepage in discrete particle simulations of biological systems, *Comput. Methods Biomech. Biomed. Eng.*, **19** (2016), 1160–1170.
11. K. Metze, R. C. Ferreira, R. L. Adam, Classification of thyroid follicular lesions based on nuclear texture features—Lesion size matters, *Cytometry*, **77** (2010), 1101–1102.
12. J. Cui, J. X. Chen, Image-based clipping, *U. S. Patent*, 2007.
13. J. Liu, B. Xu, L. Shen, J. Garibaldi, G. Qiu, HEp-2 cell classification based on a deep autoencoding-classification convolutional neural network, *IEEE 14th Int. Symp. Biomed. Imaging (ISBL 2017)*, (2017), 1019–1023.
14. S. M. Kang, J. W. L. Wan, A multiscale graph cut approach to bright-field multiple cell image segmentation using a Bhattacharyya measure, *Med. Imaging 2013: Image Proc.*, **8669** (2013).
15. Y. Chen, J. W. L. Wan, Bright-field cell image segmentation by principal component pursuit with an Ncut penalization, *Med. Imaging 2015: Image Proc.*, **9413** (2015), 94133F.
16. O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, Cham, (2015), 234–241.

17. P. Naylor, M. La é F. Reyal, T. Walter, Segmentation of nuclei in histopathology images by deep regression of the distance map, *IEEE Trans. Med. Imaging*, **38** (2018), 448–459.
18. F. Kromp, L. Fischer, E. Bozsaky, I. Ambros, W. Doerr, Taschner-Mandl S, Ambros P, Hanbury A. Deep Learning architectures for generalized immunofluorescence based nuclear image segmentation, preprint, arXiv: 1907.12975.
19. Y. H. Huang, T. C. Yu, P. H. Tsai, Y. X. Wang, W. L. Yang, M. Ouhyoung, Scope+: a stereoscopic video see-through augmented reality microscope, in *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, (2015), 33–34.
20. P. H. C. Chen, K. Gadepalli, R. MacDonald, Y. Liu, S. Kadowaki, K. Nagpal, et al., An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis, *Nat. Med.*, **25** (2019), 1453–1457.
21. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, in *Advances in Neural Information Processing Systems*, (2015), 91–99.
22. I. Guyon, M. Nikraves, S. Gunn, L. Zadeh, *Feature Extraction*, Springer Berlin Heidelberg, 2006.
23. U. Orhan, M. Hekim, M. Ozer, EEG signals classification using the K-means clustering and a multilayer perceptron neural network model, *Expert Syst. Appl.*, **38** (2011), 13475–13481.
24. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2016), 770–778.
25. X. Xie, X. Han, Q. Liao, G. Shi, Visualization and pruning of SSD with the base network VGG16, in *International Conference on Deep Learning Technologies*, (2017), 90–94.



AIMS Press

©2020 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)