Journal of Network Intelligence ©2021 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **6**, Number **3**, August **2021**

Applying Deep Learning Algorithm to Cell Identification

Chien-Chih Lin

Department of Electrical Engineering
National Changhua University of Education
u358554@taipower.com.tw

Jin-Xiang Liao

Department of Electrical Engineering
National Changhua University of Education
have88888888@gmail.com

Ming-Sung Chiu

Department of Electrical Engineering
National Changhua University of Education
Hugojely@gmail.com

Ming-Tsung Yeh

Department of Electrical Engineering
National Changhua University of Education
jack.yeh@tw.unisys.com

Yi-Nung Chung*

College of Engineering
National Changhua University of Education
ynchung@cc.ncue.edu.tw

Chao-Hsing Hsu

College of Engineering
Chienkuo Technology University
hsu@ctu.edu.tw

Correspond Author: ynchung@cc.ncue.edu.tw

Received May 2021; revised June 2021

Abstract. *There are many factors may affect human health, such as diseases and parasite infections. When scientists are fighting viruses, bacteria and parasites, they need to observe the influence on cells. By using the traditional method to identify cells, the test results may have errors because of human factors. One image processing technology together with deep learning algorithm is proposed in this paper, which can identify cells accurately and does not affect by environment and noise. This algorithm is proposed to segment and identify the cells. Based on the experimental results, the Mask R-CNN has better performance than other architectures for target recognition. Therefore, the Mask R-CNN architecture is proposed to identify the cells in this paper.*

Keywords: Neural network technology, Deep learning, Mask R-CNN

---

1. **Introduction.** In order to find the source of diseases, scientists need to observe the influence on cells. The traditional recognition methods are easily interfered by noise and human factors. In recent years, the neural network algorithm is applied to image processing [1-3], which can achieve much better results for image recognition. There are some methods for target detection and identification in the reference articles, which usually applied the neural network or the convolutional neural network (CNN) [4-5]. For example, the R-CNN algorithm was presented in [6], the faster R-CNN algorithm was presented in [7], the YOLO v1 was presented in [8], the YOLO v2 was presented in [9], and etc.

The convolutional neural network (CNN) is a fundamental algorithm of deep learning which is an extended architecture of neural networks [4-5]. There are three layers in a CNN algorithm which are convolution layer, pooling layer and fully connected layer. The CNN algorithm can make a good advantage for image detection and recognition. Some other extended research algorithms such as Alex Net, VGG16, ResNet50, and ResNet101 are studied by researchers and have published in articles [10-12]. However, the neural network algorithm which is applied for image processing usually needs pre-processing for the input data. Following the pre-processing, the training and identification processes are necessary. In the pre-processing, the system needs to segment the object one by one and then resizes them to the same size. The efficiency of this kind of algorithm is not quite well. In this paper, one better algorithm is proposed for image detection and recognition, which is Mask R-CNN method [10-11]. It does not need additional processing of the image in this algorithm. This algorithm can identify the target on original image and it has better accuracy also. Therefore, the Mask R-CNN method is a good choice to identify the cells in image processing algorithm.

Plasmodium is a parasite whose hosts are humans. This kind of disease infects about two hundred million people every year and four hundred thousand people will die. It affects human health very much. The traditional discrimination method for plasmodium is observed by using the microscope after staining cells. If the scientists use the traditional method to identify the plasmodium, it may cause errors because of the human factors. The image processing together with deep learning algorithm is a better method, which can identify objects accurately without being affected by the noise. In this paper, the Mask R-CNN algorithm is proposed to segment and identify the plasmodium. The cells infected by plasmodium can be selected and identified by using the proposed algorithm. And then, the infected person can be further treated in hospital. The scientists can observe the changes of the infected cells during the treatment also.

There are two problems to identify plasmodium in human cells. The first problem is that the identification rate of cells infected by plasmodium is low. The second problem is that if the rate of overlap cells is too high or the background is too complexed, the recognition rate will be low also. In order to solve these two problems, much more sampling data of cells infected by plasmodium is necessary. However, the sampling data of cells infected by plasmodium is very hard to be obtained. One easier method is to generate the data by computer, but the data generation is not easy also. Because we need to cut the cells which infected by the plasmodium first, and then we need to synthesize the cut cells one by one among the thousands of input data. Different backgrounds will be involved in the process of image synthesis also. After the synthesis process is completed, the marked data needs to be enhanced. If the overlap rate of cells is too high, the recognition rate will be affected. This problem can be solved by increasing the coverage region of the data, which means to increase the amount of database for high overlap images. The traditional image processing is hard to solve this kind of problems. It needs an extended image processing algorithm such as neural networks together with deep learning algorithm. In this paper, the Mask R-CNN method is proposed for image detection and recognition. Because it does not need additional processing of the image in this algorithm. This algorithm can identify the target on original image and it has high accuracy also.

The rest of this paper is arranged as follows. The system algorithm proposed in this paper is described in the second section. The third section is the experimental test. In this section, the experimental results by using the proposed algorithm are presented. The conclusion is presented in the last section.

2. **System Algorithm.** In this paper, the Mask R-CNN method is proposed for image detection and recognition, which does not need additional processing to deal with the images in this algorithm. There are three processes in this algorithm. At first, the ResNet101 is used in feature pyramid network (FPN) architecture as the backbone. It uses residual learning to reduce the vanishing of gradient caused by layers in the network [12-13]. Secondly, the region proposal networks (RPN) is proposed to classify the background and modify the bounding box. At last, the ROI-Align architecture which is a region extraction method is proposed to identify the proposal target and refine the bounding box again. Based on the proposed method, the cells infected with plasmodium can be selected.

The architecture of Mask R-CNN network is shown in Figure 1 and its flowchart is shown in Figure 2. This algorithm can be divided into three parts which are FPN, RPN and ROI-Align architecture. At first, the input image is resized as a fixed size, and then the preliminary feature image is trained in this algorithm. In this process, the FPN network architecture is conducted. The preliminary feature maps are added to the following layers and the anchor frames can be found from these feature maps. After performing the preliminary RPN network process, the background image is removed. The rest image samples are performed as a preliminary boundary regression. After ROI-Align process, the system conducts a process to remove the background one more time and to perform the boundary regression again.

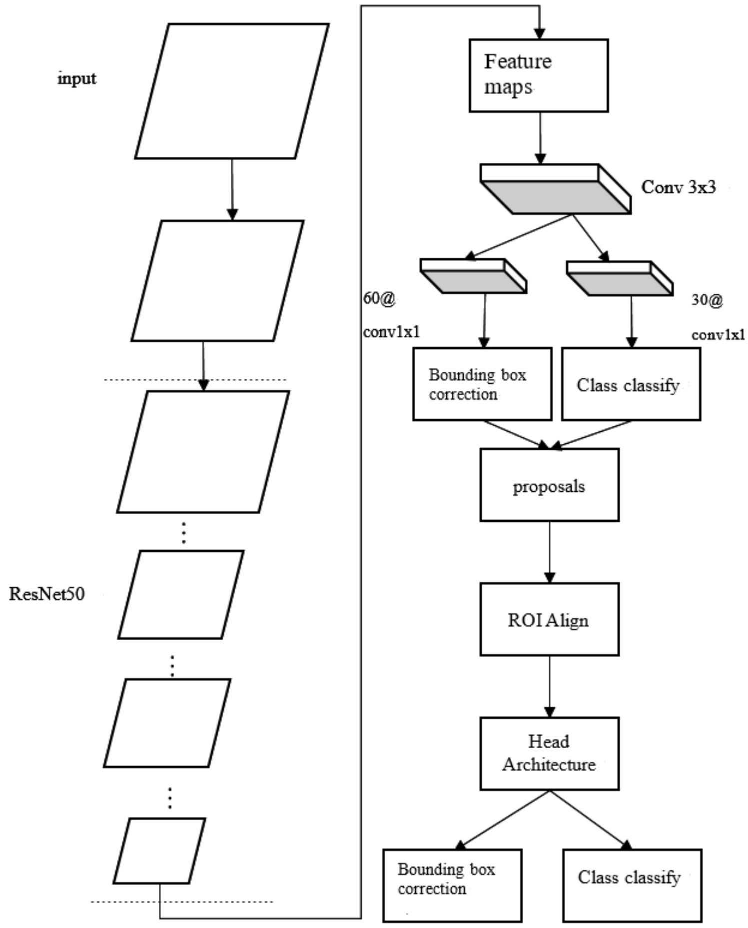


Figure 1. The architecture of Mask R-CNN network

In this algorithm, the cell image is resized to 1600x1600 (pixels), and then let the image rotate 90 degrees, 180 degrees and 270 degrees respectively. And then the image processing algorithm is conducted for image enhancement and put it into the backbone for training. In this algorithm, the FPN architecture applies ResNet101 as the backbone. The system uses residual learning to reduce the vanishing of gradient caused by layers in the network. The RPN proposes anchor bounding box and refines the bounding box also. Finally, the head architecture re-identifies the proposal targets and refines the bounding box again. The more detail processes are presented as follows.

FPN Network: The preliminary feature is trained in FPN network and it will produce four different sizes of feature maps which are P2, P3, P4 and P5. The overall architecture diagram of the FPN network is shown in Figure 3. C1 is the resized input data. After computing the convolution with C1 in different layers, it will produce different feature maps which are C2, C3, C4 and C5. P5 is obtained by computing 1x1 convolution with C5. The feature maps of P2, P3 and P4 can be obtained by computing the convolution following the architecture of FPN network.

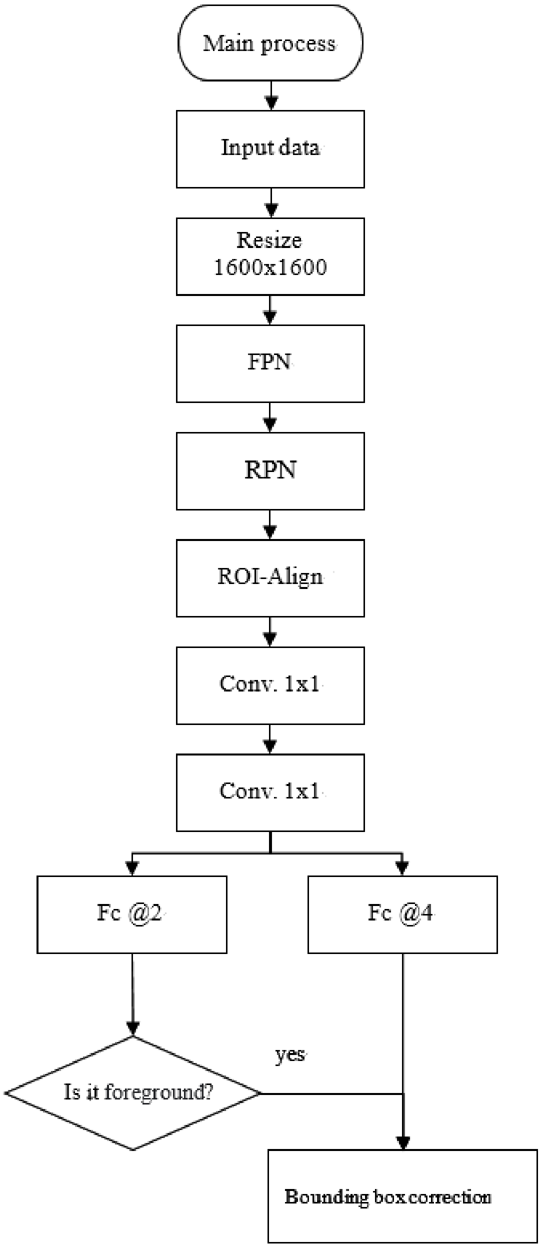


FIGURE 2. The flowchart of Mask R-CNN network

Target frame extraction process: There are many different objects in a feature map, therefore the target must be selected before training process. The purpose of anchor is to select the targets by using different frame on the feature map. Therefore, a feature map will contain a large number of anchor frames. The system will select 6000 higher score's anchor frames first. And then, the system will set a threshold to select 128 anchor frames. The anchor frames are divided into three types which are positive anchor, negative anchor, and neutral anchor. The effective anchor frames are positive anchor and negative anchor. In the training process, the ratio of positive anchor and negative anchor is 1:2 for training. The selected positive anchor contains the trained targets which is shown in Figure 4. The negative anchor is the background part which is shown in Figure 5.

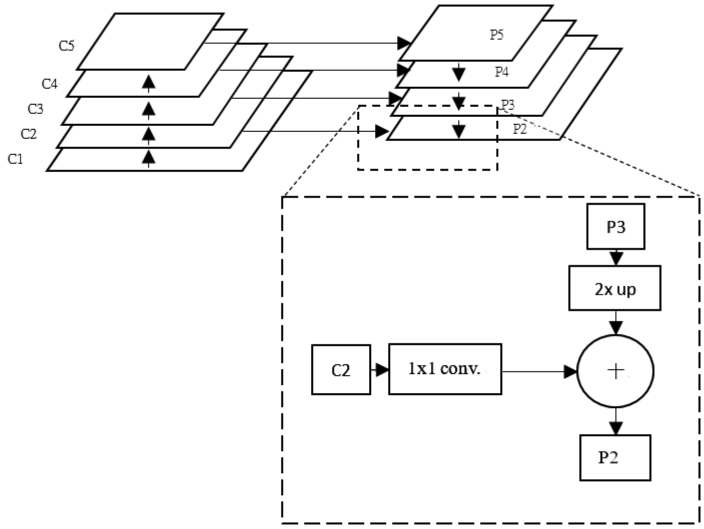


Figure 3. The architecture of FPN network

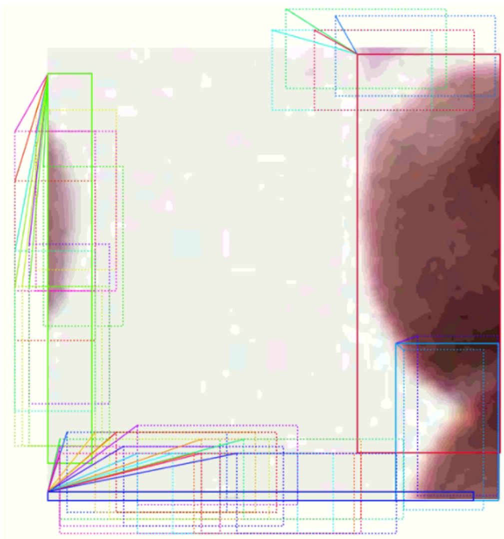


Figure 4. Positive anchor frames

The architecture of RPN network: The flowchart of region proposal network (RPN) is shown in Figure 6. The feature maps are the output of FPN network, which are P2, P3, P4 and P5. At first, the RPN network computes a 3x3 convolution for the input data. And then, it divided into two 1x1 convolution computation branches. The left-hand side branch has 30 channels and the right-hand side branch has 60 channels. Following the computation algorithm shown as Figure 6, the proposed region can be selected. The frames in RPN layer will be made the first time classification and frame correction. The classification is divided into foreground and background. If the frame that is judged to be the background, it will be discarded. The proposed bounding box of RPN network will be the input of ROI-Align network. The size of the frame will be 7X7 (pixels) in ROI-Align network. In ROI-Align network, the system will perform the second time classification and frame correction after two fully connected layers. The classification is divided into

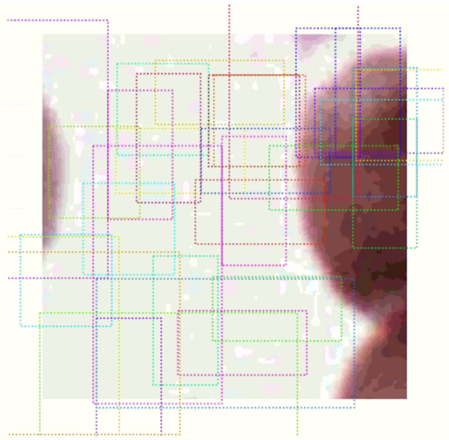


Figure 5. Negative anchor frames

foreground and background. If the frame that is judged to be the background, it will be discarded also. It is quite closed to ground truth after the second time frame correction.

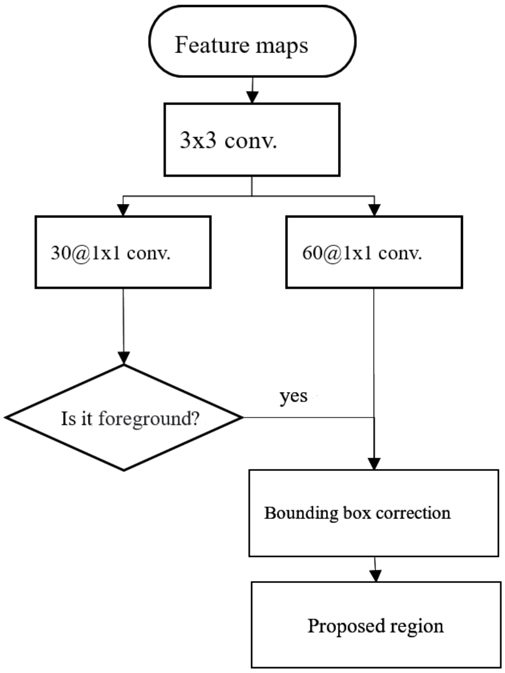
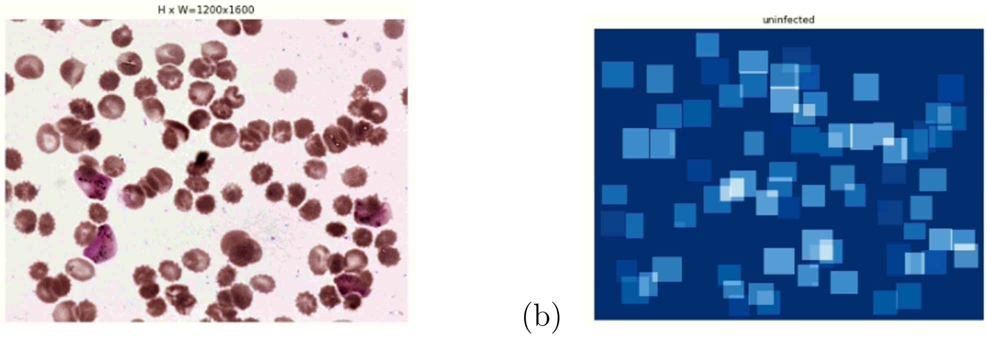


Figure 6. The flowchart of region proposal network

3. **Experimental results.** The experimental results are presented in this section. The database used in this experiment is Malaria Bounding Boxes, which is an open source malaria cell on Kaggle [14]. There are two types of classification targets which are infected cells and normal cells. The input initial image is shown in Figure 7(a), and every infected or normal cells need to be made a mask. A schematic diagram of the mask is shown in Figure 7(b).

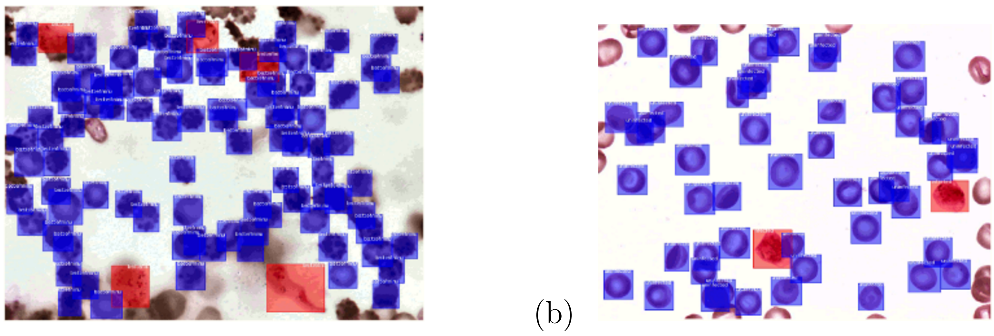
The mask on the image is shown in Figure 8. The system puts the data into the network for following process. After the FPN network layer, five kinds of feature maps with different sizes will be formed. And put them into the RPN layer. The purpose of



(a) (b)

FIGURE 7. Input image

the RPN layer is to make the target extraction, which is carried out under different size feature maps. The anchor frame is shown in Figure 9. This frame can be used to detect the targets. In the RPN layer, the frame is divided into foreground and background. If the frame is judged to be a background, then it will be discarded. After entering the ROI Align layer, the size will be adjusted to the same size. And then, the system makes the second time classification and frame correction. The classification is still divided into foreground and background. If it is a background and it will be discarded also. After the second frame correction, the result is quite close to ground truth. The training loos curve is shown in Figure 10.



(a) (b)

FIGURE 8. The mask on the image

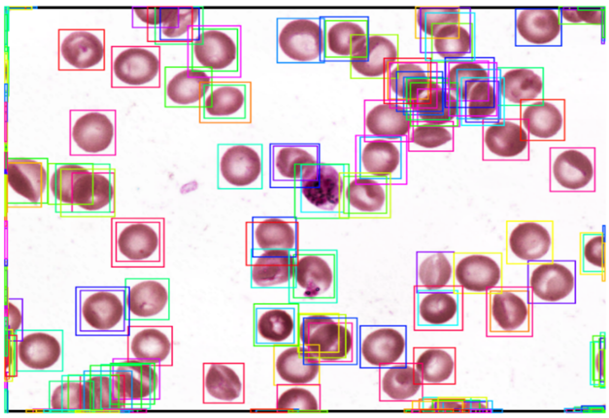


FIGURE 9. Anchor frame

The test and prediction result is shown in Figure 11. In the Figure, the red box is the prediction box, and the green box is the ground truth. Based on the experimental

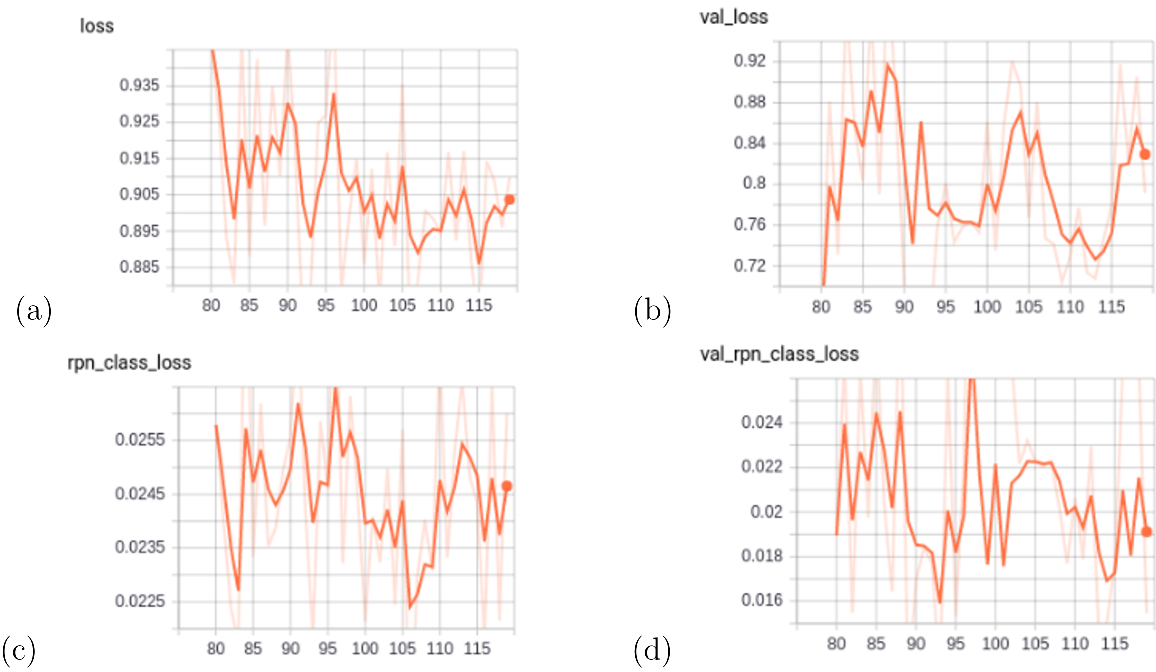


Figure 10. The training loos curve

results, the accuracy of cell prediction is quite well. But if the cell overlap rate is high, the difficulty of predicting will be increased. The problem can be solved by using more sample data which is in different environments and it needs to do data enhancement also. The accuracy of test results is shown in Table 1. Based on the experimental results, the test accuracy of using ResNet101in FPN network is better than that of using ResNet50.

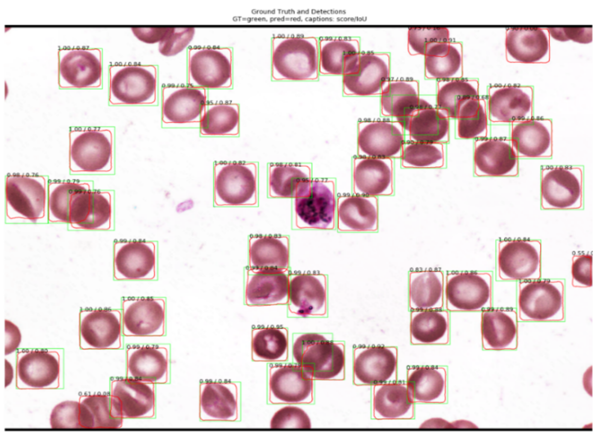


Figure 11. The test and prediction result

4. **Conclusion.** In this paper, the Mask R-CNN algorithm is proposed to cell detection and recognition. Based on the experimental results, the proposed method can identify objects accurately without being affected by environment noises. The cells infected by Plasmodium can be selected out. After identifying the plasmodium, the infected person can be further treated in the hospital. The system chooses Res-Net as the basic structure of backbone in the network. However, if the Res-Net network has too many layers, it may cause the training gradient to disappear and may cause the training results to diverge. Therefore, the residual learning structure is applied to the structure to converge

Table 1. The test accuracy

	backbone	AP [val.]	AP [test]
Mask R-CNN	ResNet50	0.49	0.51
Mask R-CNN	ResNet101	0.54	0.53

the network. Based on the experimental result, the accuracy of choosing ResNet101 as a backbone will be better than that of ResNet50 for FPN network. Based on the experimental results, the cells can be effectively identified and the bounding box can be effectively predicted.

Acknowledgment. This work was supported by the Ministry of Science and Technology under Grant MOST 106-2221-E-018-028- and Department of National Defense project in 2020.

REFERENCES

[1] F. Zhang, T. Y. Wu, Y. Wang, R. Xiong, G. Ding, P. Mei, L. Liu, Application of quantum genetic optimization of LVQ neural network in smart city traffic network prediction, *IEEE Access*, vol. 8, pp. 104555-104564, 2020

[2] K. Wang, C. M. Chen, M. S. Hossain, G. Muhammad, S. Kumar, S. Kumari, Transfer reinforcement learning-based road object detection in next generation IoT domain, *Computer Networks*, vol. 193, 108078, 2021

[3] K. K. Tseng, J. Lin, C. M. Chen, M. M. Hassan, A fast instance segmentation with one-stage multi-task deep neural network for autonomous driving, *Computers & Electrical Engineering*, vol. 93, 107194, 2021

[4] J. M. -T. Wu, Z. Li, G. Srivastava, J. Frnda, V. G. Diaz J. C. -W. Lin, A CNN-based Stock Price Trend Prediction with Futures and Historical Price, *2020 International Conference on Pervasive Artificial Intelligence (ICPAI)*, pp. 134-139, 2020

[5] J. M. -T. Wu, Z. Li, N. Herencsar, B. Vo, J. C. –W. Lin, A graph-based CNN-LSTM stock price prediction algorithm with leading indicators, *Multimedia Systems*, 10.1007/s00530-021-00758-w, 2021

[6] R. Girshick, J. Donahue, T. Darrell, J., Malik, UC Berkeley, Rich feature hierarchies for accurate object detection and semantic segmentation, *arXiv*:1311.2524v5 [cs.CV] 22 Oct 2014

[7] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *arXiv*:1506.01497v3 [cs.CV] 6 Jan 2016.

[8] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection, *arXiv*:1506.02640v5 [cs.CV] 9 May 2016.

[9] J. Redmon, A. Farhadi, YOLO9000: Better, Faster, Stronger, *arXiv*:1612.08242v1 [cs.CV] 25 Dec 2016

[10] R. Anantharaman, M. Velazquez, Y. Lee, Utilizing Mask R-CNN for Detection and Segmentation of Oral Diseases, *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2197-2204, 2018.

[11] A. O. Vuola, S. U. Akram, J. Kannala, Mask-RCNN and U-Net Ensemble for Nuclei Segmentation, *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 208- 212, 2019.

[12] E. K. Wang, C. M. Chen, M. M. Hassan, A. Almogren, A deep learning based medical image segmentation technique in Internet-of-Medical-Things domain, *Future Generation Computer Systems 108*, vol. 135-144, 2020

[13] K. He, X. Zhang, S. Ren, J. Sun, M. Research, Deep Residual Learning for Image Recognition, *arXiv*:1512.03385v1 [cs.CV], 10 Dec., 2015

[14] Ljosa et al. P. vivax (malaria) infected human blood smears, *Broad Bioimage Benchmark Collection*, https://bbbc.broadinstitute.org BBBC041, 2012.