Pneumonia Detection Algorithm Based on Improved YOLOv3

Hailong Liu, Jinrong Cui(✉), and Chaoda Peng

College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China
tweety1028@163.com

Abstract. Pneumonia is a kind of disease caused by bacteria, viruses and other pathogens, which can seriously endanger human health, and has strong infectivity. Timely and accurate detection of pneumonia symptoms can not only make patients receive timely treatment, but also prevent the disease from spreading to others. This paper proposes an improved object detection algorithm YOLOv3-P which was based on YOLOv3. Using the idea of Path Aggregation Network (PANet) for reference, after feature fusion, the location information is enhanced through a bottom-up path, which makes full use of the feature information of each layer. And the better backbone network CSPDarkNet53 was used to replace DarkNet53 of YOLOv3, to better extract features from the pneumonia images. Experiments on the lung X-ray image data set provided by the North American Society of Radiology show that the average precision of the algorithm reaches 50.43%, which was improved compared with the YOLOv3 algorithm, and has good performance compared with other common object detection algorithms. YOLOv3-P can help doctors judge the location of pneumonia tissue faster and more accurately.

Keywords: Object detection · Pneumonia · Deep learning · YOLOv3

1 Introduction

Pneumonia is a respiratory disease caused by bacteria, viruses and other pathogens. It can lead to fever, cough, headache and other symptoms, and has a strong infectious. The number of people infected with pneumonia was increasing every year in the world. Pneumonia is a serious threat to human health. The diagnosis of pneumonia based on X-ray images is one of the most important methods for the diagnosis of pneumonia. But because the X-ray image is a black-and-white image, it is difficult for doctors to distinguish the diseased part from the normal part due to the lack of important information such as color and texture.

In recent years, machine learning has developed rapidly, and it has a wide range of applications in the field of computer vision [1–3]. In 2012, AlexNet [4] was proposed by Krizhevsky. AlexNet won the championship in the Imagenet image classification competition of the year. Convolutional neural network can learn deep feature information

S. Wang et al. (Eds.): IoTCare 2021, LNICST 415, pp. 313–320, 2022.
https://doi.org/10.1007/978-3-030-94182-6_22

from images, and it can be applied to object detection fields to improve the detection effect of the algorithm. R. Girshick et al. Proposed the first object detection algorithm R-CNN [5] based on convolutional neural network in 2014.

R. Joseph. proposed YOLOv1 [6] object detection algorithm in 2016. The detection accuracy of R-CNN series algorithm is very high, but its detection speed is slow. In order to solve the problem of real-time detection in object detection tasks, YOLOv1 was born. YOLOv1 algorithm directly inputs the image into the network and directly regresses the size and position of the target. However, compared with the two-stage object detection algorithm, its positioning accuracy is lower, especially for the very close targets and small-scale targets.

In 2017, R. Joseph improved based on YOLOv1 and proposed YOLOv2 [7] object detection algorithm. YOLOv2 absorbs the idea of Faster R-CNN algorithm and also adopts the method of the prior box. It uses the clustering method of K-means [8] to obtain the width and height of the prior box, to find a more suitable prior box.

In 2018, R. Joseph made some improvements on the basis of YOLOv2 and proposed YOLOv3 algorithm [9]. YOLOv3 uses three different proportions of feature maps to get the prediction box, which improves the detection effect of small targets, and proposes DarkNet53 feature extraction network.

With the development of the neural network, it is possible to use deep learning to detect and classify medical images. Setio [10] used multiple Convolutional Neural Networks (CNN) to identify pulmonary nodules, and fused the final results, and achieved good results. In reference [11], the de-noising self-coding method is used to extract the depth features of pulmonary nodules for classification, and its performance in fine-grained classification is better than that of traditional morphological and texture bottom feature learning methods.

In view of the current pneumonia detection algorithm prone to misdiagnosis and missed diagnosis, this paper improves the YOLOv3 detector, greatly improves its performance, reduces the phenomenon of misdiagnosis and missed diagnosis, so as to help doctors diagnose pneumonia faster and more accurately. X-ray images lack color and texture information, and it is difficult to distinguish the diseased part from the normal part. To solve this problem, this paper uses the idea of PANet [12] for reference, and further improves the Feature Pyramid Networks (FPN) [13] structure of YOLOv3. After the feature fusion of FPN, the positioning information is enhanced through the bottom-up path, making full use of the feature information of each layer. And use better backbone network CSPDarkNet53 [14] to replace the backbone network DarkNet53 of YOLOv3, so as to better extract features from the image.

The experimental results on the lung X-ray image data set provided by the Radiological Society of North America show that the detection accuracy of YOLOv3-P is greatly improved compared with the original algorithm YOLOv3, and it has better performance than several other commonly used target detection algorithms, which can help doctors judge the location of pneumonia tissue faster and more accurately.

2 Related Work

2.1 Two Stage Detectors

The first two stage detector based on deep learning is R-CNN, which uses CNN to extract features from images, but the network is complex and the training speed is slow. SPP-Net [15] can input images of any size, only need to do a convolution feature extraction to get the feature image, but it cannot achieve end-to-end detection. Fast R-CNN [16] uses RPN to generate suggestion boxes, which greatly speeds up the generation of suggestion boxes. Compared with other algorithms mentioned above, the speed of the algorithm is fast, but it has not achieved real-time detection yet.

2.2 Single Stage Detectors

Different from two stage detectors, single stage detectors do not need to generate suggestion boxes, only need to directly regress the type and location of the target, thus greatly speeding up the detection speed. YOLOv1 divides the input image into grids, and each grid cell is responsible for detecting the falling objects. SSD [14] significantly increases the detection effect by detecting targets of different sizes on feature maps of different scales.

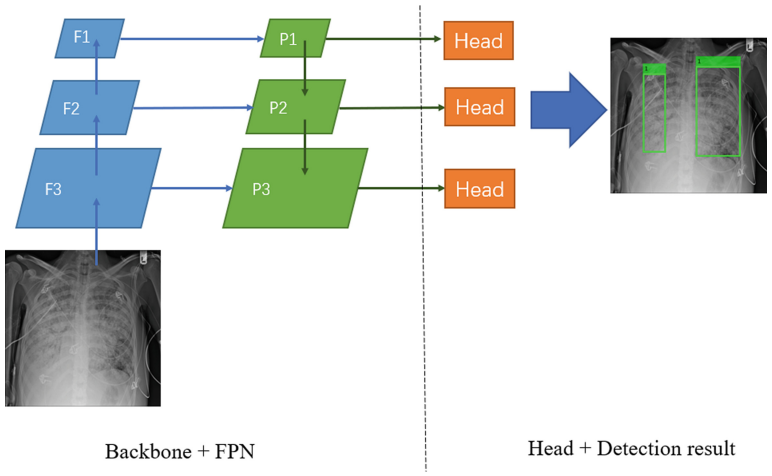

Fig. 1. Network structure of YOLOv3.

2.3 YOLOv3

YOLOv3 is an improved version of YOLOv2, and its backbone network is DarkNet53. YOLOv3 uses the idea of FPN for reference and fuses the feature images of the same size in the up-sampling stage. YOLOv3 network outputs three different scale feature maps, which are 13×13, 26×26, 52×52. Different size feature maps are used to detect different size objects. Figure 1 shows the network structure of YOLOv3.

3 Our Approach

In this paper, through the improvement of YOLOv3, YOLOv3-P is proposed. The algorithm first extracts features through the CSPDarkNet53 backbone network, and then obtains three feature maps with different scales through the PAN. Finally, each feature map is classified and regressed to get the final result. The network structure of the YOLOv3-P algorithm is shown in Fig. 2.

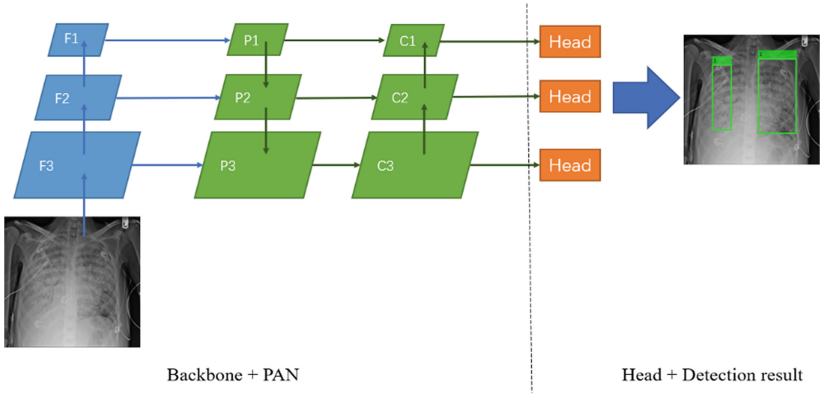

Fig. 2. Network structure of YOLOv3-P.

The training process of YOLOv3-P includes three stages: feature extraction, target location and loss calculation. The main network used in feature extraction is CSPDarkNet53 network. Object location is divided into two steps: classification and location regression. The loss calculation includes classification loss and location loss. After the lung images are input into the YOLOv3-P network, three feature maps F1, F2 and F3 with different sizes are output through the backbone network. Among them, F1 is 52 $\times$ 52, F2 is 26 $\times$ 26, F3 is 13 $\times$ 13. F3 gets P3 through 1 $\times$ 1 convolution reduction channel, P3 is up sampled to get a feature map of size 26 $\times$ 26 and fused with F2 to get P2, P2 is up sampled to get a feature map of size 52 $\times$ 52 and fused with F1 to get P1. After obtaining different sizes of feature maps P1, P2 and P3 through FPN, P1 passes through 1 $\times$ 1 convolution reduction channel to obtain feature map C1 with size of 52 $\times$ 52, C1 is down sampled to obtain feature map with size of 26 $\times$ 26, which is fused with P2 to obtain C2, C2 is down sampled to obtain feature map with size of 13 $\times$ 13, which is fused with P3 to obtain C3. At the same time, after each feature fusion, the fused feature image will be convoluted by 3 without changing the size and channel. After obtaining different size feature maps C1, C2 and C3, different size targets are predicted on each feature map.

3.1 PANet

The PANet was proposed in reference [17], which shortens the information path between low-level and high-level features through a bottom-up path. The specific method is, after

extracting the image features through the backbone network, first through the top-down path, the high-level features are fused by up-sampling and low-level features, and then through the bottom-up path, the whole feature layer is enhanced by using more accurate positioning information of low-level features, to get a better feature map for subsequent classification and regression.

3.2 CSPDarkNet53

CSPDarkNet53 is a backbone network based on DarkNet53, the backbone network of YOLOv3, and learning from the experience of CSPNet [18]. CSPNet solves the problem that network reasoning needs a lot of computation from the perspective of network architecture. Replacing DarkNet53 with CSPDarkNet53 can enhance CNN's learning ability, make the network lightweight while maintaining accuracy, and reduce the amount of computation and memory costs.

4 Experiments

4.1 Dataset

The experimental data set comes from the X-ray images of the lung provided by the Radiological Society of North America, including 4000 Gy-scale images with the size of 1024 pixels × 1024 pixels, 75% of which are used for training and the rest for testing. The input size of the network is 416 pixels × 416 pixels, and the gray image into RGB (Red, Green, Blue) image.

4.2 Training Details

In the experiment, the pre training model of DarkNet53 and CSPDarkNet53 on COCO data set is used as the feature extraction network, and Adam [19] is used to optimize the model. The experimental environment is: the deep learning framework is pytorch1.2 + cuda10.0, the GTX Tian x graphics card with 12 GB video memory, and the operating system is Ubuntu 18.04.

4.3 Metrics

In order to select the appropriate algorithm, this paper uses the average precision (AP) as the evaluation index. P-R curve is based on the two variables of precision (P) and recall (R). AP is obtained by calculating the area under P-R curve. And the Intersection over Union (IoU) is 0.5. The evaluation index of detection speed is frame per second (FPS).

4.4 Ablation Study

In order to verify the influence of different improved methods on the algorithm, this paper sets up three groups of experiments, which are: 1) YOLOv3; 2) YOLOv3 + PAN; 3) YOLOv3 + CSPDarkNet53 + PAN. The experimental results are shown in Table 1.

Table 1. Influence of different improvement methods on the performance of YOLOv3.

Method	AP/%	FPS
YOLOv3	44.21	40.96
YOLOv3 + PAN	46.47	39.12
YOLOv3 + PAN + CSPDarkNet53	50.43	34.12

As can be seen from Table 1, after replacing FPN in YOLOv3 with PAN, the AP of the algorithm is improved by 2.26%. On this basis, after replacing DarkNet53 with CSP-DarkNet53, the AP can continue to improve by 3.96%. This shows that the improvement of this algorithm to YOLOv3 can improve the AP of the algorithm.

4.5 Compared with Other Detectors

In order to verify the detection performance of YOLOv3-P, this paper compares YOLOv3-P with other object detection algorithms, and the experimental results are shown in Table 2.

Table 2. Performance comparison of different algorithms.

Method	AP/%	FPS
SSD	41.78	70.35
Faster R-CNN	46.12	9.47
YOLOv3	44.21	40.96
YOLOv3-P	50.43	34.12

As can be seen from Table 2, in the lung X-ray image data set, the AP of our algorithm YOLOv3-P is higher than other algorithms. The AP of SSD is lower than that of YOLOv3, and the AP of Faster R-CNN is higher than that of YOLOv3. The AP of YOLOv3-P is 8.65% and 4.31% higher than SSD and Faster R-CNN, respectively.

4.6 Detection Effect of Different Algorithms

Figure 3 shows the detection effect of different algorithms on four groups of pneumonia images, in which the blue box represents the real coordinate box, the green box represents the detection box, and the red box represents the detection error box.

As can be seen from Fig. 3, SSD algorithm has poor detection effect, large positioning error, and false detection and missing detection problems. The reason is that SSD algorithm does not fuse feature maps of different sizes, and positive and negative samples are unbalanced in the training process. The detection effect of YOLOv3 is better than that of SSD, because YOLOv3 algorithm combines different scale feature maps and

makes better use of the information of pneumonia image. Faster R-CNN algorithm uses region proposal network (RPN) to generate suggestion box, which makes the detection effect better and the problems of missed detection and false detection less. YOLOv3-P has the least false detection and missing detection problems, and the positioning accuracy is higher, because this algorithm replaces FPN with PAN, to make better use of the information of different sizes of feature maps. At the same time, it uses a more excellent backbone network CSPDarkNet53, so that the network can extract better feature information.

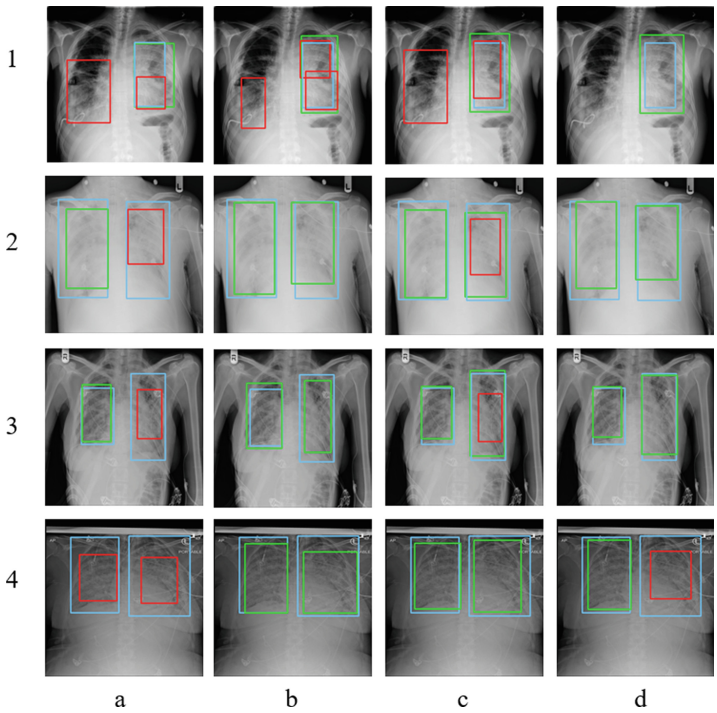

Fig. 3. Detection effect of different algorithms. Blue box: real coordinate box; Green box: detection box; Red box: detection error box. (a) SSD; (b) Faster R-CNN; (c) YOLOv3;(d) YOLOv3-P (Color figure online)

5 Conclusions

In view of the lack of color and texture information in the X-ray of the diseased part of pneumonia, the feature is not obvious and so on, this paper proposes the YOLOv3-P algorithm based on the improvement of the YOLOv3 algorithm. The algorithm uses a better feature extraction network CSPDarkNet53 to extract features better, and uses pan structure to make better use of feature information. The experimental results show that YOLOv3-P improves the detection accuracy significantly compared with the original algorithm, and has good performance compared with other classical target detection

algorithms. YOLOv3-P can help doctors judge the location of pneumonia tissue faster and more accurately.

Acknowledgments. This work supported by the Opening Project of Guangdong Province Key Laboratory of Computational Science at the Sun Yat-Sen University 2021011.

References

1. Gao, G., Yang, J., Jing, X.-Y., Shen, F., Yang, W., Yue, D.: Learning robust and discriminative low-rank representations for face recognition with occlusion. Pattern Recogn. **66**, 129–143 (2017)
2. Zhang, Z., et al.: Inductive structure consistent hashing via flexible semantic calibration. IEEE Trans. Neural Netw. Learn. Syst. **32**, 4514 (2020)
3. Zhang, Z., Lai, Z., Xu, Y., Shao, L., Wu, J., Xie, G.: Discriminative elastic-net regularized linear regression. IEEE Trans. Image Process. (TIP) **26**(3), 1466–1481 (2017)
4. Wu, Q., Shen, C., Wang, P., et al.: Image captioning and visual question answering based on attributes and external knowledge. IEEE Trans. Pattern Anal. Mach. Intell. 1 (2017)
5. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014). https://doi.org/10.1109/CVPR.2014.81
6. Redmon, J., Divvala, S., Girshick, R., et al.: You only look once: unified, real-time object detection. Comput. Vis. Pattern Recogn., 779 (2016)
7. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: IEEE Conference on Computer Vision & Pattern Recognition, pp. 6517–6525 (2017)
8. Hartigan, J.A., Wong, M.A.: A K-means clustering algorithm. Appl. Statist. **28**(1), 100 (1979)
9. Redmon, J., Farhadi, A.: YOLOv3: an incremental improvement. arXiv e-prints (2018)
10. Setio, A., Ciompi, F., Litjens, G., et al.: Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. IEEE Trans. Med. Imaging **35**(5), 1160–1169 (2016)
11. Cheng, J.Z., Ni, D., Chou, Y.H., et al.: Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in CT scans. Sci. Rep. **6**, 24454 (2016)
12. Liu, S., Qi, L., Qin, H., et al.: Path aggregation network for instance segmentation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2018)
13. Lin, T.Y., Dollar, P., Girshick, R., et al.: Feature pyramid networks for object detection. IEEE Comput. Soc. (2017)
14. Bochkovskiy, A., Wang, C.Y., Liao, H.: YOLOv4: optimal speed and accuracy of object detection (2020)
15. He, K., Zhang, X., Ren, S., et al.: Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1904–1916 (2014)
16. Ren, S., He, K., Girshick, R., et al.: Faster R-CNN: towards real-time object detection with region proposal networks (2017)
17. Liu, W., Anguelov, D., Erhan, D., et al.: SSD: single shot MultiBox detector. In: European Conference on Computer Vision. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46448-0_2
18. Wang, C.Y., Liao, H., Wu, Y.H., et al.: CSPNet: a new backbone that can enhance learning capability of CNN. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE (2020)
19. Kingma, D., Ba, J.: Adam: a method for stochastic optimization. Comput. Sci. (2014)