ISSN: 2710-0588

RefineDet with Improved Attention Block for Chest X-ray Image Lesion Detection

Xiang Peng[1, *], Chuncheng Chen[2]

[1] School of Mathematics and Computer Science, Guangdong Ocean University, Zhanjiang, Guangdong, 524000, China

[2] School of Electronics and Information Engineering, Guangdong Ocean University, Zhanjiang, Guangdong, 524000, China

* Corresponding Author: xi3@stu.gdou.edu.cn

Abstract

In recent years, artificial intelligence technology has been widely used to assist radiologists in diagnosing and analyzing medical images. The use of artificial intelligence technology can well assist doctors in the localization of lesions. However, the mainstream target detection models at this stage are difficult to be practically applied in medical systems because of factors such as the use of large backbone networks and high input resolution, which leads to low model accuracy and high consumption of computational resources. In this paper, we propose a fast detection speed and high accuracy of the lung lesion detection network IAB- RefineDet. By improving the channel and spatial attention mechanisms and introducing the improved attention module into RefineDet, the lesion detection accuracy is dramatically improved without significantly increasing the number of parameters. We conduct extensive experiments on VinDr-CXR, the world's largest publicly available chest radiograph detection dataset, and comparative experiments with existing mainstream target detection models. The experimental results show that IAB-RefineDet achieves a mAP of 16.23%, and the lesion detection performance is significantly better than mainstream deep learning models.

Keywords

Lesion Detection; Deep Learning; Attention Mechanism; Chest X-ray.

1. Introduction

As global industrialization accelerates, the issue of human lung health presents a major challenge. According to a report by the World Health Organization, five of the top ten causes of death worldwide are related to the lungs [1]. Therefore, early screening for lung diseases is important to reduce their fatality. Due to the large difference in the ratio of radiologists to patients, radiologists need to examine a large number of CXR (Chest X-ray) images every day, and the diagnostic process must be completed in a short period of time, which may lead to some misdiagnosis due to the influence of doctor's experience as well as personal subjective factors. In order to reduce the work pressure of radiologists and to improve the diagnostic accuracy of chest x-ray images, computer-aided diagnosis using artificial intelligence techniques has become increasingly popular.

Thanks to the rapid development in the field of artificial intelligence, combining deep learning to assist doctors in diagnosis has become a new trend. In terms of lung disease screening, the target detection model can substantially improve the efficiency of radiologists' disease screening by providing doctors with areas where lung lesions may occur. Medical studies have shown that computer-aided diagnosis can play an important role in disease screening when the

ISSN: 2710-0588

false-positive rate of the computer-aided diagnosis system is reduced and the sensitivity is higher than 80% [2].

Deep learning-based target detection in the medical image field mainly follows the model of generalized target detection tasks, such as YOLO [3], Faster RCNN [4], etc. Among them, the one-stage target detection methods can achieve faster detection speed, and the two-stage target detection methods can achieve higher detection accuracy. Although the one-stage target detection algorithms dominate in speed and have smaller models, which are more likely to be applied in the auxiliary detection system of hospitals, there is still a large gap in their detection accuracy compared to the two-stage target detection algorithms. And the second-stage target detection model is difficult to be deployed in the hospital's detection system due to the problems of larger model and higher occupation of computational resources.

Zhang et al. proposed a one-stage target detection network, RefineDet [5], by adding two modules, ARM (Anchor Refinement Module) and ODM (Object Detection Module), to the network for initial filtering and further screening of anchor frames, respectively. And the TCB (Transfer Connection Block) module is used to fuse the features between ARM and ODM, so that the one-stage target detection network can have the accuracy of the two-stage target detection network while maintaining a faster detection speed. This provides the possibility of model landing. In this paper, RefineDet is used as a benchmark network, inheriting the advantages of its architecture and making improvements for its insufficient feature extraction capability and low lesion detection accuracy. It is committed to designing a target detection algorithm with high accuracy and fast detection speed, which can more effectively assist doctors in the judgment of lung diseases.

We validate the performance of IAB-RefineDet by conducting experiments on VinDr-CXR, the world's largest publicly available chest X-ray detection dataset. We compare the experiments with mainstream deep learning models in terms of mAP and number of parameters, and the experimental results are shown in Table1, which shows that our designed IAB-RefineDet achieves 16.23% mAP while keeping a moderate amount of parameters. The IAB-RefineDet has a performance improvement of 6.95% compared to the benchmark network, and the detection accuracy is significantly better than that of the mainstream deep learning models.

2. Method

2.1. The Overall Structure of proposed IAB-RefineDet

The network structure of IAB-RefineDet is shown in Fig. 1. The ARM is used to perform preliminary screening of the frames, remove the frames of non-target classes, and roughly adjust the position of the anchor to provide a better initial anchor frame for the subsequent target regression. The role of the ODM module is to further precise the position of the anchor frame and predict the class information of the anchor frame. The TCB module exists between the ARM and ODM modules, which integrates contextual information to a greater extent by transferring the features of ARM module in different output layers to the corresponding ODM module, and improves the detection capability of ODM module. However, in chest X-ray image lesion detection, RefineDet has the problem of underutilization of features, in order to solve this problem, we designed the IAB module by combining the attention mechanism and added it before the TCB module, so that the features extracted by the backbone network are processed by the IAB first. By adding the designed IAB module before TCB, the lesion detection performance of the model is significantly improved.

ISSN: 2710-0588

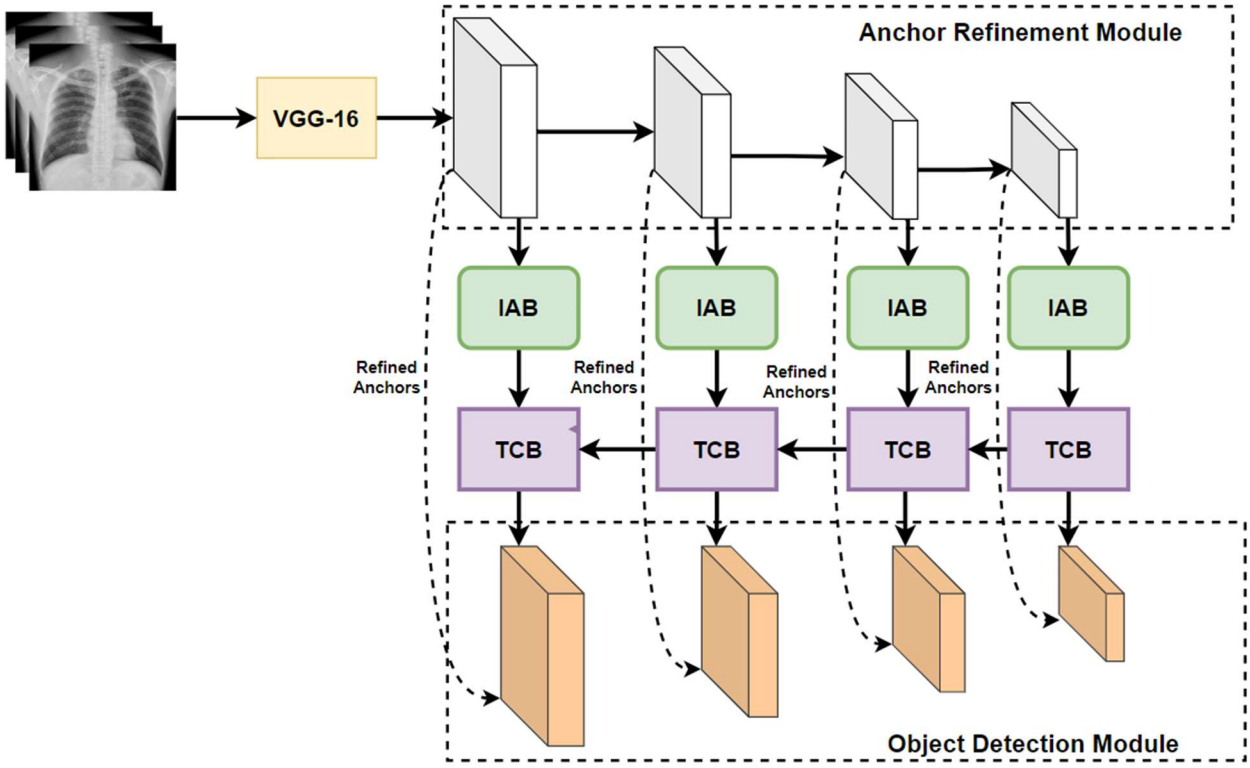

Fig. 1 The overall structure of IAB-RefineDet

2.2. The Convolutional Block Attention Module

The attention mechanism in deep learning is a processing mechanism that can learn autonomously and selectively focus on important features. The attention model enables the network to allocate more computational resources to more important feature information. The Convolutional Block Attention Module (CBAM) [6] is a simple and effective attention module proposed by Woo et al. to extract features in both channel and spatial dimensions. The channel attention module and the spatial attention module are connected in series between the input and output of the CBAM structure. Both modules employ global maximum pooling and global average pooling to extract richer global and local semantic information. For the feature maps generated by the convolutional network, CBAM first computes the channel attention templates based on the channel dimensions and then recalibrates the channel weights of the original feature images by multiplying the channel attention templates with the original feature images. Subsequently, the feature maps with extracted channel attention are input into the spatial attention module and the channel attention module.

2.3. The Improved Attention Block

Inspired by CBAM, we design the IAB as shown in Fig. 2. The IAB consists of a channel attention mechanism and a spatial attention mechanism. The feature map is first processed by the channel attention mechanism, and the resulting feature map is multiplied with the original feature map to readjust the channel dimension weights of the original feature image. Then the multiplied feature map is processed by the spatial attention mechanism, and the obtained feature map is multiplied with the output feature map of the channel attention mechanism. So that the network adaptively learns the importance of different pixel positions in the same channel, and finally get the filtered salient features. Finally it is fused with the original feature map by shortcut.

ISSN: 2710-0588

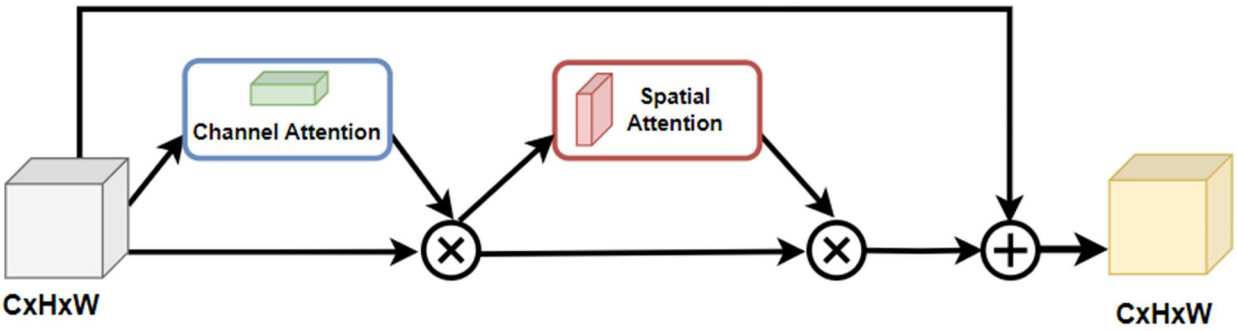

Fig. 2 The improved attention block includes both spatial and channel attention components

2.4. Channel Attention Block

To adjust the weights of different channels in the feature graph more effectively. For the channel attention mechanism, we also use global maximum pooling and global average pooling for semantic information extraction from the feature map. Maximum pooling considers only the elements with maximum values in the pooled region and ignores the other elements to extract the salient features of the target and maximize the retention of texture, structure, contour and other information in the image. Average pooling calculates the average value of all elements in the pooled region to retain more information about the background of the image. Using two pools simultaneously both removes redundant information and ensures richer high-level feature extraction. Two feature vectors of size HxWx1 are then obtained. Then we use a convolutional layer of size 1x1 to capture the more complex channel attention features in the obtained two feature vectors, and finally the feature vectors of the two paths are fused and processed by a sigmoid function.

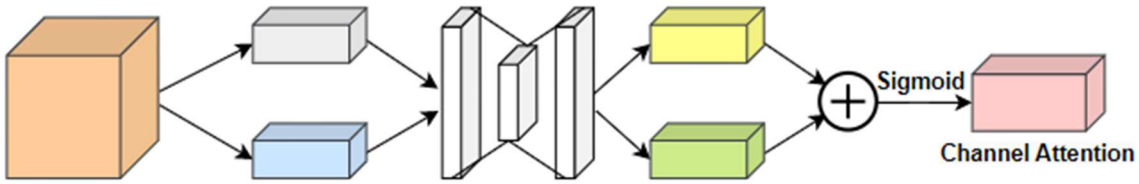

Fig.3 The structure of channel attention block

2.5. Spitial Attention Block

The spatial attention block is connected to the channel attention block for extracting key information from different locations in the same feature map and generating a spatial attention map using the spatial relationship of the features. For the spatial attention module, the input to this module is the recalibrated feature map from the channel attention module. Semantic information is first extracted from the HxWxC feature map along the spatial dimension using global average pooling and global maximum pooling to obtain the HxWx2 feature map, and then multiscale information is extracted from the obtained feature map after three convolutional layers of different sizes. Finally, the multiscale information extracted from each branch is fused and processed by sigmoid function.

ISSN: 2710-0588

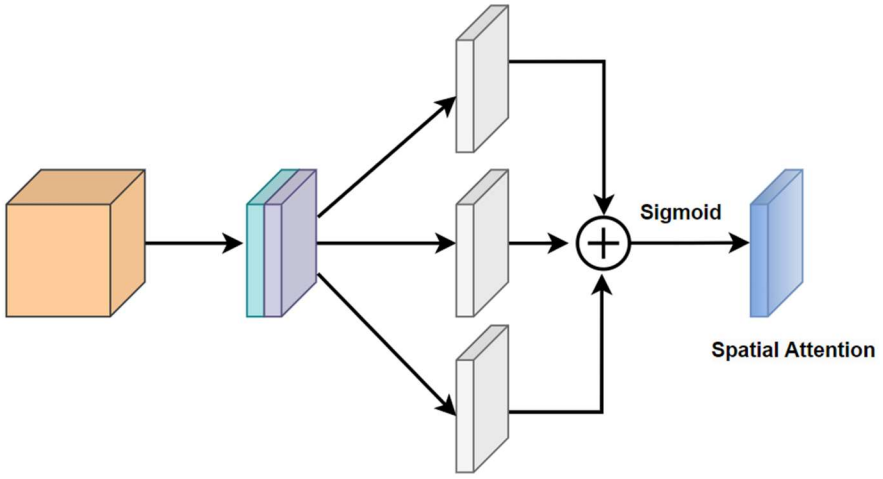

Fig.4 The structure of spatial attention block

3. Comparison with Mainstream Target Detection Algorithms

In order to prove the superiority of the model, we conduct comparative experiments on mainstream target detection models. The two performance metrics of mAP and number of parameters of the models are compared, and the experimental results are shown in Table 1. In terms of detection accuracy, the benchmark model RefineDet has a mAP of 9.28%, and the mAP of SSD [7] and RetinaNet [8] is also below 10%. Yolov3 [9] achieves a detection accuracy of 10.84%. The Faster R-CNN has a slightly higher detection accuracy due to its two-stage detection network, but it is still lower than our model IAB-RefineDet, and the number of Faster R-CNN parameters is larger compared to our model. The experimental results show that IAB-RefineDet achieves a mAP of 16.23%, which is significantly better than the mainstream deep learning models while keeping the number of parameters moderate. This shows that IAB-RefineDet can maintain high detection speed while keeping high detection accuracy.

Table 1. Performance comparison with other models

Methods	Backbone	mAP(%)	Params(M)
RetinaNet	ResNet-50	9.34	38.66
SSD	VGG-16	6.21	36.58
Faster R-CNN	VGG-16	11.71	42.32
	ResNet-50	12.30	45.59
Yolov3	DarkNet-53	10.84	40.73
RefineDet	VGG-16	9.28	36.77
IAB-RefineDet	VGG-16	16.23	40.23

4. Conclusion

In this paper, we propose a highly efficient and accurate lung lesion detection network IAB-RefineDet. by designing an attention processing method including a channel attention module and a spatial attention module. The lesion detection performance of the network is dramatically improved. The meanAP of our IAB-RefineDet is 6.95% higher than that of the benchmark network, and the meanAPs are all significantly better than those of the current mainstream deep learning models. This well meets the needs of helping radiologists in the clinic for lung disease diagnosis and effectively alleviates the current problems of low detection accuracy and high resource consumption in lung lesion detection.

ISSN: 2710-0588

References

[1] Li Y, Luo L, Lin H, et al. Dual-consistency semi-supervised learning with uncertainty quantification for COVID-19 lesion segmentation from CT images[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24. Springer International Publishing, 2021: 199-209.

[2] Yu L, Wang S, Li X, et al. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation[C]//Medical image computing and computer assisted intervention–MICCAI 2019: 22nd international conference, Shenzhen, China, October 13–17, 2019, proceedings, part II 22. Springer International Publishing, 2019: 605-613.

[3] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[4] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(6): 1137-1149.

[5] Zhang S, Wen L, Bian X, et al. Single-shot refinement neural network for object detection[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4203-4212.

[6] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.

[7] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[8] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[9] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arxiv preprint arxiv:1804.02767, 2018.