Electronic Engineering and Informatics
G. Izat Rashed (Ed.)

doi:10.3233/ATDE240093

COVID-19 Detection and Localization: A Novel Fused Attention Mechanism Network Based on YOLO V5

Wenyi YANG[a,c1], Ruihu WANG[b2], Jianxun ZHANG[a3], Chang LIU[d4], Li ZENG[c5], Xiaoyu SHEN[c6]
[a]*Chongqing University of Technology, College of Computer Science and Engineering, Chongqing, China*
[b]*Chongqing University of Arts and Sciences, School of Tourism & Event Management, Chongqing, China*
[c]*Chongqing Vocational and Technical University of Mechatronics, Information Engineering Institute, Chongqing, China*
[d]*Industrial and Commercial Bank of China, Shapingba Sub-branch, Chongqing, China*

Abstract. COVID-19 is a very contagious viral disease that has shown characteristics of being insidious and spreading rapidly since its emergence at the end of 2019, leading to many suspected cases and increased difficulty in protection. However, COVID-19 has become a huge clinical challenge in terms of rapid identification and precise treatment due to the similarity between its pathogen and the common pneumonia virus. In order to help physicians to diagnose patients quickly and accurately, we proposed a powerful target detection virus model. This innovative model based on the deep network of the YOLO V5 model structure, is used to detect lesion areas of COVID-19, combined with attention mechanism thinking. Meanwhile, to improve the category imbalance problem in the data, we proposed an innovative loss function based on Focal Loss. This approach helps our network to be more focused on the learning of complex cases and effectively improves the model accuracy. Finally, we compared it with other experiments. The proposed improved YOLO V5 model detected novel coronavirus infections with an mAP of 62.23%. Obviously, the effective YOLO V5 we propose has better results and judgment ability compared to existing models and can be effectively used as an aid to detect novel coronavirus infections.

Keywords. YOLO V5, Attention Mechanism, Category Imbalance, Focal Loss

---

[1] Wenyi YANG, Chongqing University of Technology, College of Computer Science and Engineering; Chongqing Vocational and Technical University of Mechatronics, Information Engineering Institute; e-mail: 773496652@qq.com

[2] Corresponding author: Ruihu WANG, Chongqing University of Arts and Sciences, School of Tourism & Event Management, e-mail: 50432096@qq.com

[3] Corresponding author: Jianxun ZHANG, Chongqing University of Technology, College of Computer Science and Engineering, e-mail: 466908695@qq.com

[4] Chang LIU, Industrial and Commercial Bank of China, Shapingba Sub-branch, e-mail: 461373535@qq.com

[5] Li ZENG, Chongqing Vocational and Technical University of Mechatronics, Information Engineering Institute, e-mail: 876971844@qq.com

[6] Xiaoyu SHEN, Chongqing Vocational and Technical University of Mechatronics, Information Engineering Institute, e-mail: 601566245@qq.com

1. Introduction

COVID-19 (Coronavirus disease), refers to pneumonia caused by the novel coronavirus infection in 2019. Novel coronavirus infections are highly infectious, the impact on different patients does not vary and from current case data, most develop mild to moderate disease and recover without hospitalization. However, a small number of severe patients can cause many complications, such as heart disease, septic shock, etc. When the patient's condition becomes critical, over 60% of patients lose their lives [1]. The rapid spread of novel coronavirus infections and their global prevalence poses a serious threat to human health and safety, as well as to economic and social development, public health, trade, and other areas. Rapid diagnosis and accurate treatment of novel coronavirus infections are therefore of great importance.

Current techniques that can accurately detect novel coronavirus infections include nucleic acid testing, antigen testing, and imaging. Firstly, nucleic acid testing results have a high false negative rate, the test needs to be repeated several times, which is very time-consuming and costly in terms of human resources, making the test very inefficient. Secondly, the antigen test is performed using a new coronavirus antigen test kit. This test is not as accurate as the new coronavirus nucleic acid test, but it is a low-cost, short, simple, and convenient test. Third, imaging refers to taking advantage of CT and chest X-rays that examine the patient's medical image and assist the physician in making a rapid diagnosis. Chest CT of novel coronavirus infections has shown certain characteristics and its results are more accurate than those of nucleic acid detection. However, due to the high cost of CT scans, chest X-rays are cheaper and more readily available. Therefore, this study gets chest X-ray lung images on patients with novel coronavirus infections.

As medical technology continues to evolve, medical images are important for disease diagnosis. According to statistics, more than 90% of medical data comes from imaging examinations [2], and there are many problems with relying on radiologists alone to analyze the data. Therefore, it is important to apply AI methods to the detection of novel coronavirus infections in medical images.

2. Related Work

To increase the precision and recognition of interpreting chest radiographs by medical aid systems and to save the effort and time of overworked physicians [3], researchers have done much work on this. For example, Misra et al. [4] came up with the multi-channel ResNet network method and confirmed that the integrated network was more accurate than the single network model, with an accuracy of 94% and a recall of 100%. Wang et al. [5] has proposed the COVID-Net, thus enabling better and timely treatment of patients. Pathak et al. [6] used deep transfer learning (DTL) techniques to build the classification network for detecting COVID-19 infected cases while using 10-fold cross-validation to prevent overfitting problems, and the model achieved 96.2% training accuracy and 93.02% test accuracy, respectively. Song et al. [7] proposed a new coronavirus pneumonia diagnosis system based on deep learning, which used ResNet-50 to extract features. The model results with AUC and recall (sensitivity) scored 0.99 and 0.93 with other patients, respectively. Asnaoui et al. [8] compared the performance of deep learning models (VGG16, VGG19, and so on) for detection and classification using confusion matrices. The models were evaluated and the results showed that Inception_ResNet_V2 and DenseNet201 had good accuracy rates of 92.18% and 88.09%

respectively.

Currently, detection is divided into two types, the first is two-stage. Its object detection algorithms are those that manufacture pre-selected boxes and then classify the objects through CNN. Such as R-CNN and Fast R-CNN etc. Another type is a one-stage object detection algorithm, directly learning extracted features. One-stage object detection algorithms contain YOLO [9] series, and SSD [10]. YOLO v5 as a single-stage algorithm performance is extremely excellent and directly processes common image data files with high speed and accuracy. YOLO v5 is widely used in real-life applications but has not yet been applied to detect novel coronavirus infections.

3. Method

3.1 Introduction of the YOLO v5 algorithm

YOLO v5, one of the popular one-stage target detection algorithms, has many merits, for instance, fast detection speed, high precision, and low model weight, etc. The input, the backbone, the neck, and the output form its network structure. YOLO v5 uses Mosaic for data enhancement on the input side, The optimal anchor frame is calculated adaptively according to the data set, and pre-processes the image using adaptive image scaling operations; the backbone network consists mainly of Conv, C3, and SPPF modules; The PANet structure acts as part of the neck network; the output side mainly detects objects of different sizes. To better improve the detection of the model, we have improved the model. One is to use Group Normalization instead of bulk normalization in networks, as the training data provided are very small and cannot be trained and learned from large-scale data, using Group Normalization can have a better performance effect. Secondly, we improve the C3 module by adding an attention mechanism model and propose the C3CABM (CBAM [11], Convolutional Block Attention Module) module, The improved C3CBAM module network is shown in Figure 1.

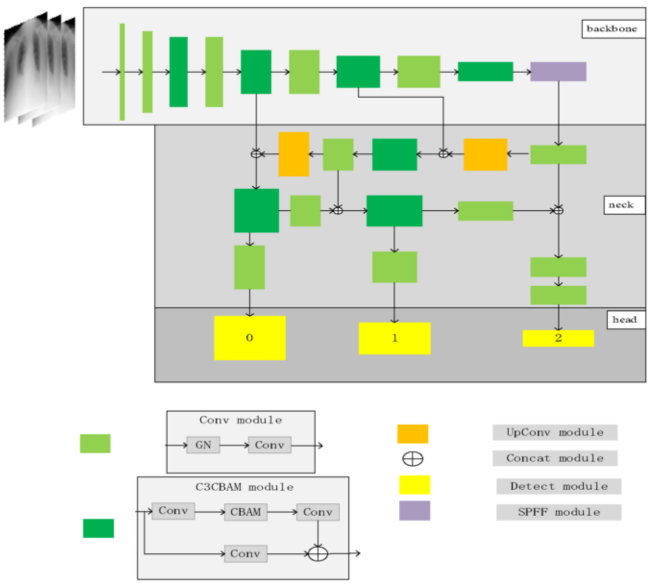

Figure 1. Modified YOLOV5

We use a convolutional function with a kernel size of 3*3 instead of the Bottleneck module and finally output it by splicing it with the characteristic map of the other channel. After adding the CBAM module, this network model allows the model to further expand the perceptual field and depth of the structure, improve the focus on lung features, extract more detailed features through the residual connectivity, and ultimately ameliorate the detection performance of the network model. This module can learn data features better compared to the previous one and focus on learning the target sample data. Thirdly, using multi-scale feature fusion, using PANet network idea, combined with ResNet residual network connection, using dilated convolution instead of an upsampling operation, we add more scale fusion feature information, so the proposed the modified structure can combine advanced information on semantic features and information on low-level detail features for multi-level learning and comparison.

3.2 Loss Function

YOLO v5's loss consists of Classes loss, Objectness loss, and Location loss [12]. Yolov5 uses a BCE loss function by default to calculate classification loss. However, the BCE loss function has problems in dealing with highly unbalanced multi-label classification problems, resulting in models that perform very poorly on a few categories. To address this issue, we propose a Focal loss function based on which can effectively optimize highly unbalanced multi-label classification problems. Focal loss was first proposed by Kai-Ming He, in order to address the model performance issues caused by data imbalance, it was originally used in the field of graphics. The definition of the focal loss function is Eq. (1), and unify into a single expression such that Eq. (2).

$$L_{fl} = \begin{cases} -(1-p)^{\gamma} \log(p) \ if \ \mathrm{y}=1 \\ -p^{\gamma} \log(1-p) \ if \ \mathrm{y}=0 \end{cases} \tag{1}$$

$$L_{fl} = -(1-p_t)^{\gamma} \log(p_t), \text{ then } \ p_t = \begin{cases} p \ if \ \mathrm{y}=1 \\ 1-p \ otherwise \end{cases} \tag{2}$$

Compared to cross-entropy loss, focal loss does not change for samples that are classified incorrectly, and it decreases for samples that are classified correctly. Overall, Focal loss improves the performance of the model by setting higher weights for a few categories of incorrect samples, focusing attention on the categories with high error rates, and the model is more focused on learning and detecting difficult figures.

4. Experiments

4.1 Parameters and Datasets

The following configuration is used in our article: Windows 10 operating system, Intel Core i7-10700F with a main frequency of 2.9 GHz, 16 G of RAM, and NVIDIA GeForce RTX 3080 (16 GB) as the graphics card model. The experiment was performed on the PyTorch 1.10 framework on the Pycharm development tool, and the experimental hyperparameters were in Table 1.

Table 1. Hyperparameters

Hyperparameter Name	Value	Unit
Learning Rate	0.001	%
Number of iterations	300	Per count
Batch Size	1	Per count
Momentum	0.937	%
Weight decay	0.0005	%

The dataset for this experiment was derived from data published by the Radiological Society of North America in conjunction with the Kaggle Medical Imaging Pneumonia Detection Challenge [13], and consisted of 6,334 chest X-rays. These radiographs were marked by a group of experienced radiologists to classify COVID-19 as 'Negative for Pneumonia', 'Typical Appearance', 'Indeterminate Appearance', and 'Atypical Appearance'. The DICOM format is not conducive to direct observation of lung images, and the processed image data format consists of a loss-free compressed PNG (Portable Network Graphics) format at a size of 512 pixels by 512 pixels. The data set was expanded to 19,002 images by flipping, cropping, and panning, and randomly divided the dataset into training, validation, and testing sets in a 6:2:2 ratio before training.

4.2 Evaluation criteria

In this paper, Precision, Recall, Average Precision, and mean Average Precision are used as evaluation indicators. Precision indicates the proportion of all targets predicted by the model that are correctly predicted; Recall is used to predict the ratio of correct samples to all correct samples; AP@ indicates the average of the algorithm's precision across different intersection and merge ratios (IOUs); mAP is used to indicate the average precision of the mean of each species in the dataset. A higher mAP means that the model is better at detecting the average for each category.

4.3 Ablation experiments

To confirm the effectiveness of the innovative yolov5 model for novel coronavirus infection detection performance, ablation experiments will be conducted for each module to analyze the performance of each module. We can see the experimental results in Table 2. Firstly, the improved C3 module is added to the network structure. Secondly, the loss function is modified to a focal loss function. Finally, the two are combined to generate the final improved model, and the experimental comparison with the original method is performed.

Table 2. Comparative analysis of the results of ablation experiments

Algorithm	C3CBAM	Focal Loss	P /%	R /%	mAP@.5 (%)	mAP@.5:.95 (%)
Yolov5			80.13	75.75	59.16	26.36
	√		81.45	76.84	60.14	27.02
		√	82.26	77.19	61.41	28.16
	√	√	82.71	78.01	62.23	28.97

From the data in Table 2, we can see that the addition of the CBAM module has increased Precision by 1.65%, Recall by 1.44%, mAP@0.5 by 2.5%, and mAP@0.5:0.95 by 2.5%. The effectiveness of the CBAM module in enhancing the expression of pneumonia features was demonstrated. The introduction of the focal loss function improved Precision by 2.66%, Recall by 1.9%, mAP@0.5 by 3.8%, and mAP@0.5:0.95 by 6.83%, demonstrating that the Focal Loss function improves the model's target detection accuracy by adding both modules to the base model. mAP@0.5 improves by 5.19%. These results suggest that the improved method proposed can enhance the pneumonia features and can detect the location of pneumonia more accurately.

4.4 Comparison experiments

Based on the dataset in this article, compare the improved YOLOv5 model with the most advanced object detection algorithms. Figure 2(a) gives the annotated data, and Figures 2(b)-2(e) gives the detection results of SSD, Faster-RCNN, EfficientDet D7, and the improved yolov5, respectively. The SSD algorithm, as a one-stage algorithm, does not require a complex network structure and multiple stages of processing, but the SSD algorithm has the problem of positive and negative sample imbalance during training, so the detection effect is poor. The Faster-RCNN algorithm has better detection results than SSD because it uses RPN to select balanced true and false cases. The improved yolov5 algorithm also has higher localization accuracy due to the attention mechanism module used in this algorithm, which can capture more sensory field information and extract lung features in the features. Meanwhile, the Focal Loss function can resolve the model performance problem, so our modified yolov5 algorithm has better results.

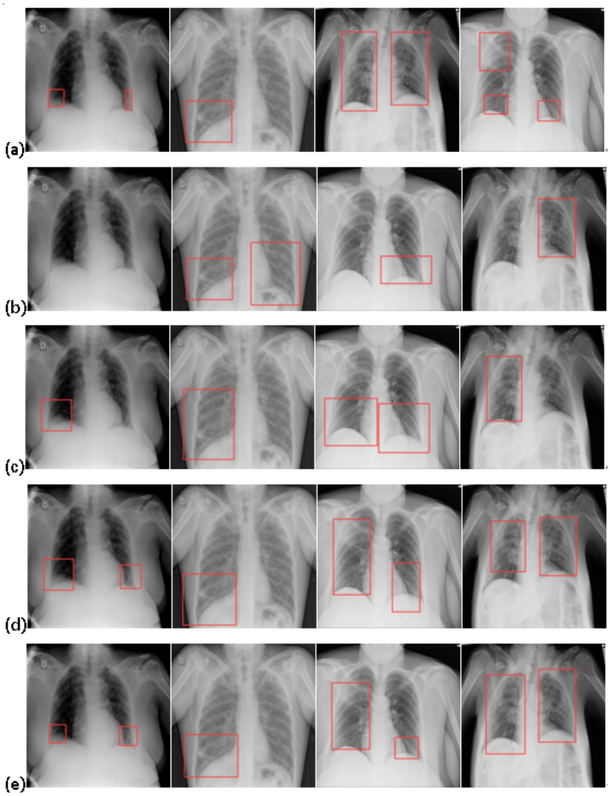

Figure 2. Effect of different algorithms

5. Conclusion

The comparison experiments show that the proposed improved YOLOV5 network model has a better representation. By learning and analyzing the input image features and using CAM and SAM, the novel network structure we propose can learn the typical features of COVID-19 more effectively. At the same time, we incorporate the Focal Loss function, which allows the network model to reduce the degree of learning of simple samples and focus on the learning of difficult and indistinguishable samples, improving the influence of class imbalance on the network model. The proposed innovative model offers better detection compared to common detection models, better detection compared to SSD algorithms, and faster speed compared to Faster-RCNN and EfficientDet D7, while the detection accuracy of this model can be guaranteed.

References

[1] China CDC Wkly. The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19) - China, 2020. 2020 Feb 21;2(8):113-122. PMID: 34594836; PMCID: PMC8392929.

[2] ZY. Jin. (2018) Prospects and Challenges: when Medical Imaging Meets Artificial Intelligence [J]. Medical Journal of Peking Union Medical College Hospital, 2018, 9(1): 2-4. DOI: 10.3969/j.issn.1674-9081.2018.01.001.

[3] M. Malik, J. Peirce, M. Wert, et al. (2021) Psychological First Aid Well-Being Support Rounds for Frontline Healthcare Workers During COVID-19. Front Psychiatry. 2021 May 28; 12:669009. doi: 10.3389/fpsyt.2021.669009.

[4] S. Misra, S. Jeon, S. Lee, et al. (2020) “Multi-Channel Transfer Learning of Chest X-ray Images for Screening of COVID-19” Electronics 9, no. 9: 1388. DOI:10.3390/electronics9091388.

[5] Wang L, Lin ZQ, Wong A. (2020) COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. Sci Rep. 2020 Nov 11;10(1):19549. DOI: 10.1038/s41598-020-76550-z.

[6] Y Pathak, P K, Shukla, A Tiwari, etc. (2020) Deep Transfer Learning Based Classification Model for COVID-19 Disease. Ing Rech Biomed. 2022 Apr; 43(2):87-92. Epub 2020 May 20. DOI: 10.1016/j.irbm.2020.05.003.

[7] Yang. S Son. Z, Li. L., etc. (2021) Deep Learning Enables Accurate Diagnosis of Novel Coronavirus (COVID-19) With CT Images. IEEE/ACM Trans Comput Biol Bioinform. 2021 Nov-Dec; 18(6):2775-2780. DOI: 10.1109/TCBB.2021.3065361. Epub 2021 Dec 8.

[8] El Asnaoui K, Chawki Y. (2020) Using X-ray images and deep learning for automated detection of coronavirus disease. J Biomol Struct Dyn. 2021 Jul;39(10):3615-3626. Epub 2020 May 22. DOI: 10.1080/07391102.2020.1767212.

[9] Redmon, Joseph, Divvala, Santosh, Girshick, Ross. (2015) You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. DOI: 10.48550/arXiv.1506.02640.

[10] Liu, Wei, Anguelov, Dragomir, Erhan, Dumitru. (2015) Ssd: Single shot multibox detector. DOI: 10.1007/978-3-319-46448-0_2.

[11] Woo, Sanghyun, Park, Jongchan, Lee, Joon-Young. (2018) CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. DOI: 10.48550/arXiv.1807.06521.

[12] Lin T Y, Goyal, Priya, Girshick, Ross, He. (2017) Focal Loss for Dense Object Detection. DOI: 10.48550/arXiv.1708.02002.

[13] P. Lakhani, J. Mongan, C. Singha, et al. (2021) The 2021 SIIM-FISABIO-RSNA Machine Learning COVID-19 Challenge: Annotation and Standard Exam Classification of COVID-19 Chest Radiographs. doi.org/10.1007/s10278-022-00706-8.