Lung lesions detection from CT images based on the modified Faster R-CNN

Linlin Xu
School of Management
Heifei University of Technology
Hefei, China
2019110767@mail.hfut.edu.cn

Xuemin Mao
School of Management
Heifei University of Technology
Hefei, China
xueminmao@hfut.end.cn

Minmin Sun
School of Management
Heifei University of Technology
Hefei, China
2019110813@mail.hfut.edu.cn

Wentao Liu
Institute of Intelligent Machines, Hefei Institutes of Physical Science,Chinese Academy of Science
Hefei, China
517581496@qq.com

Yifan Wang
Institute of Intelligent Machines, Hefei Institutes of Physical Science,Chinese Academy of Science
Hefei, China
yifanwang0623@163.com

Yuyang Tang
Institute of Intelligent Machines, Hefei Institutes of Physical Science,Chinese Academy of Science
Hefei, China
1455510277@qq.com

***Abstract*—At the end of 2019, the outbreak of the COVID-19 epidemic brought huge economic losses all around the world, and also seriously affect human work, life and study. A large number of infected patients brought huge workload to doctors, but deep learning methods can effectively assist their diagnosis. This paper is based on Faster R-CNN, an end-to-end target detection model, to realize the detection of lesions in CT images of the novel coronavirus, which contributes to track the later condition of the confirmed patients and conduct timely treatment. Firstly, Kmeans++ was carried out to cluster the dimensions of the bounding box of the ground truth in annotated CT images, and appropriate anchors sizes and ratios were selected. Then, the performance of the Faster R-CNN model based on VGG-16 and ResNet-50 on the original datasets and the augment datasets is compared. Finally, the results show that, in the enhanced dataset, the Faster R-CNN model based on VGG-16 achieved a better performance, the Recall and Precision of which on the overall test set reached 68.12% and 65.58% respectively, and the missed detection rate(MR) was 31.88%.**

Keywords—COVID-19, Faster R-CNN, Kmeans++, lesion detection.

I. INTRODUCTION

The COVID-19 broke out at the end of December 2019 has had a huge impact on people's work, life and study. On February 11, the World Health Organization (WHO) officially named it COVID-19 (Corona Virus Disease 2019). On March 11, the epidemic was officially upgraded to a global pandemic. As of July 28, 2020, the cumulative number of confirmed cases of the COVID-19 worldwide has exceeded 16 million, and the cumulative number of deaths has exceeded 600,000. The epidemic continues to spread.

After the SARS virus that broke out in 2003, COVID-19 is another serious lung infectious disease which has always been one of the most dangerous diseases that endanger human life and health[1]. At present, reverse transcription polymerase chain reaction (RT-PCR) is the golden standard for the detection of new coronaviruses and CT images with the characteristics of high resolution and painlessness are also one of the most important materials for medical diagnosis. According to a research, nucleic acid detection has some limitations including poor timeliness and low sensitivity; in contrast, CT images are highly sensitive to COVID-19[2]. However, with the outbreak of the epidemic, in the face of a large number of CT images, radiologists are prone to misdiagnosis due to some subjective factors such as fatigue.

Computer-aided diagnosis can effectively solve the above problems, the ability of which is constantly improving with the application of deep learning. Since the outbreak of the COVID-19, many scholars have studied how to diagnose COVID-19 efficiently and accurately using deep learning algorithms based on CT images. Ali et al.[3] tested the COVID-19 recognition performance of ten basic networks (AlexNet, VGG-16, VGG-19, SqueezeNet, GoogleNet, MobileNet-V2, ResNet-18, ResNet-50, ResNet-101, Xception).All ten networks can achieve good results, among which ResNet101 and Xception networks achieve the best performance with AUC reaching 0.994. Bin et al.[4] proposed a novel lesion-attention deep neural network (LA-DNN).By adding the annotations of the five major lesions of COVID-19 in the dataset, the attention of the model is focused on the lesion area to improve model performance. Yongchao Xu et al.[5] proposed a joint learning framework based on 3D-Desenet, adding a federal learning mechanism to train dataset, which improved the recognition effect of COVID-19, and the sensitivity increased to 98%. Anmol et al.[6] proposed an interpretable deep learning system for CT analysis based on YOLO v3, which can distinguish GGO of COVID-19 from GGO of other pulmonary diseases, indicating that GGO has a certain heterogeneity. Ophir et al.[7] integrated the U-net,ResNet-50, Grad-CAM and Corona score algorithms to detect the lesions and quantify the severity of COVID-19 manifestations from chest CT

This study has received funding by the National Natural Science Foundation of China (No.41871302), the National Key Research and Development Program of China (No. 2018YFB1004600).

Correspondence: Xuemin Mao

scans. Xinggang Wang et al. [8] used a weakly supervised structure to diagnose COVID-19, the main body of which is a new 3D deep convolutional neural network called DeCoVNet. Deng-Ping Fan et al.[9] proposed a new COVID-19 lung CT infection segmentation network called Inf-Net, which uses implicit reverse attention and explicit edge attention to improve the ability to identify the infected areas (Sensitivity 0.870, specificity 0.974).

From the above analysis, it can be concluded that the current research on the COVID-19 diagnosis mainly focuses on two aspects: COVID-19 recognition and lung lesion segmentation. Among them, the former is limited to the distinction COVID-19 and non COVID-19, and the latter pays more attention to lung infection. However, the lung lesions caused by COVID-19 have varieties of forms, including the ground-glass opacities (GGO), consolidation, crazy paving appearance, air bronchograms, and interlobular septal thickening etc [4]. The current research focuses more on the location of the lesion and fewer on classifications of lesions. Moreover, the steps of many methods for positioning the lesions are complicated. This paper is based on Faster R-CNN, an end-to-end target detection model, to realize the lesions detection in CT images of COVID-19, which helps to track the later condition of confirmed patients and conduct timely treatment.

II. METHODS

A. *Faster R-CNN*

The two-stage algorithm is one of the most commonly used detection algorithms for lung lesion detection[10,11], mainly including R-CNN, SPP-Net, Fast R-CNN and Faster R-CNN. Among them, Faster R-CNN[12] is a further improvement of the previous three networks and has the best detection performance. Its core contribution is to build a region proposal network (RPN) instead of selective search and share convolution layers with the feature extraction network layers, achieving end-to-end training. The structure is shown in Fig. 1, divided into two stages:

(1) The first stage: Input CT images and then extract feature maps as the input of RPN to obtain anchors

(2) The second stage: Input the anchors and feature maps into the ROI pooling layer to complete target classification and positioning.

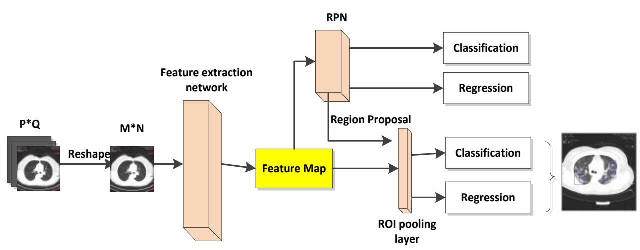

Fig.1. The structure of Faster R-CNN

Thereinto, RPN is the core of Faster R-CNN, which is a fully convolutional network. The network is to take a feature map of any size as input , and then generate a set of anchor boxes through a 3×3 sliding window, and obtain the class and regression gradient of the anchor boxes at last, as shown in Fig.2 .The loss function of RPN is divided into classification loss $L_{cls}$and bounding box regression loss L_{reg}. The overall loss function is as follows:

$$L(\{p_i\}+\{t_i\})=\frac{1}{N_{cls}}\sum_i L_{cls}\ p_i,p_i^* +\frac{1}{N_{reg}}\ {}_i\, p_i^* L_{reg}\ t_i,t_i^* \qquad (1)$$

where i represents the i-th anchor box, p_i is the probability of predicting the i-th box as a positive sample, p_i^* refers to whether the i-th predicting box is a true positive sample, t_i is the predicted coordinates of the i-th box, t_i^* is the coordinate of the labeled box corresponding to the i-th candidate box. N_{cls} refers to the number of candidate frames participating in training, N_{reg} refers to the number of feature map pixels, and the λ respects balance weight. Thereinto, the formula of $L_{cls}$and L_{reg} are as follows:

$$L_{cls}\ p_i,p_i^* = -\log[p_i p_i^*+(1-p_i)(1-p_i^*)] \qquad (2)$$

$$L_{reg}\ t_i,t_i^* = \begin{cases} 0.5(t_i-t_i^*)^2 & t_i-t_i^* <1 \\ t_i-t_i^*|-0.5 & \text{else} \end{cases} \qquad (3)$$

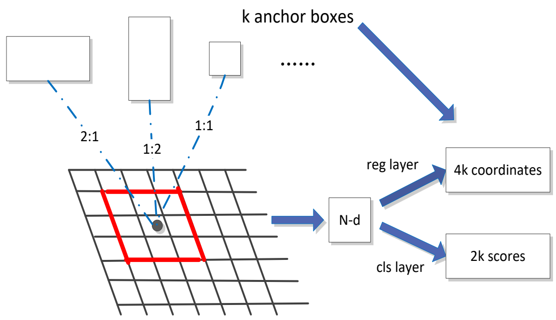

Fig. 2. PRN introduction

B. *K-means++*

The original RPN has three different anchor sizes, which are 128, 256 and 512 pixels. Each size has three ratios of length and width, which are 1:1, 1:2, and 2:1. A total of nine sizes of anchors are generated. Due to differences in datasets, the size and aspect ratio of the original anchors cannot fully meet the requirements of the CT image dataset. In this paper, the K-means++ algorithm is used to analyze the size and aspect ratio of the ground truth. The number of cluster centers is 9 and the cluster diagram is shown in Fig. 3. The size is mainly concentrated between 10-120 pixels, and the 9 cluster centers and the aspect ratio are shown in Table I . In order to meet the requirements of the sample size in the CT dataset, this paper determines that the anchor sizes are 7.5, 15, 30, and 60, and the aspect ratios are 1:1, 1:2 and 2:1, so there are a total of 12 kinds of anchors.

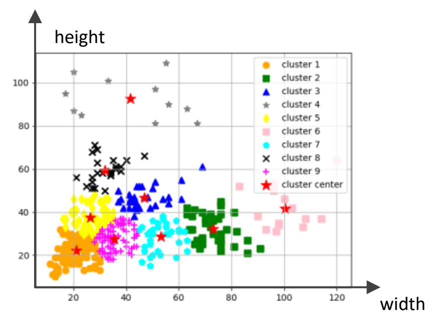

Fig. 3. K-means++ cluster diagram

TABLE I K-MEANS++ CLUSTER CENTERS

cluster	1	2	3	4	5	6	7	8	9
width	21.06	73.06	46.97	41.45	26.45	100.21	53.11	31.92	35.43
height	22.12	32.25	46.73	92.64	37.6	41.71	28.54	59.33	27.51
ratio	0.95	2.27	1.01	0.45	0.70	2.40	1.86	0.54	1.29

III. EXPERIMENT AND RESULTS

A. Datasets preprocessing

This study contains a total of 263 CT images annotated by professional radiologists with the size of 512×512.There are two kinds of lesions: GGO and consolidation(Con).In the imaging process of CT images, it is inevitable that some salt and pepper noises will be generated ,which can be effectively removed by the median filter algorithm [13,14], so the median filter with 3×3 convolution is used to denoise the images. The deep learning model requires a large amount of data, so the dataset is expanded to 1238 pictures through rotation, cropping, translation, mirroring and cutout. As shown in Fig. 4, the left side is the denoised CT image, and the right side is the original CT image.

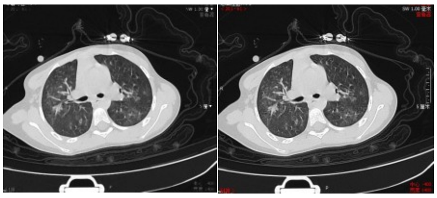

Fig. 4. Comparison of the original image and denoising image

B. Model training and evaluation

The experiment uses ResNet-50 and VGG-16 as the basic feature extraction network of Faster R-CNN. The anchor sizes of the RPN network are 7.5, 15, 30 and 60, and the aspect ratio is 1:1, 1:2 and 2:1. The size of the original dataset is 263, and the size of the augmented dataset is 1238. Both are divided into training set, validation set and test set in a ratio of 7:2:1. The training environment is based on windows10 with 128GB memory, tensorflow-gpu 1.9.0 and Python3.7. The graphics card is Nvidia TITAN X (Pascal).

In order to verify the effectiveness of the model, considering the limitations of AP indicators that cannot accurately describe the accuracy of numbers [15], this paper uses Precision, Recall and missing detection rate(MR) to evaluate the performance of the model. They are also universal evaluation indexes. The calculation formulas are as follows:

$$\text{Precision} = \frac{\text{True detection}}{\text{True detection+False detection}} \tag{4}$$

$$\text{Recall}=\frac{\text{True detection}}{\text{True detection+missing}} \tag{5}$$

$$\text{MR}=\frac{\text{missing}}{\text{missing+Ture detection}} \tag{6}$$

where True detection (Td) represents the number of prediction boxes in the test set corresponding to ground truths, False detection (Fd) represents the number of prediction boxes in the test set not corresponding to ground truths, and missing refers to the number of lesions in the test set that are not in the prediction box but in ground truth.

In addition, considering the rigor and scientific nature of medical diagnosis, the predict boxes with probability higher than 0.6 will be counted and the prediction box can be positive as long as it has overlaps with the object.

C. Analysis of experimental results

This paper conducted 4 comparison experiments based on the original dataset with 263 CT images and the augmented dataset with 1238 CT images, and compared the performance of Faster R-CNN under the two basic feature extraction networks of VGG-16 and ResNet-50. The loss curve and the average detection accuracy curve of the training process are shown in Fig. 5 and Fig. 6. Analyzing the loss curve, it can be found that the loss curves of three experiments all converged to one point with no overfitting but the loss of the ResNet-50 trained in the enhanced dataset increased in the last few trainings, resulting in an overfitting phenomenon. For the three normal experiments, in the original data set, the convergence speed of ResNet-50 is significantly higher than that of VGG-16. After increasing the data set, the reduction of the convergence speed of VGG-16 is reasonable. In addition, their average detection accuracy was ultimately higher than 99%.

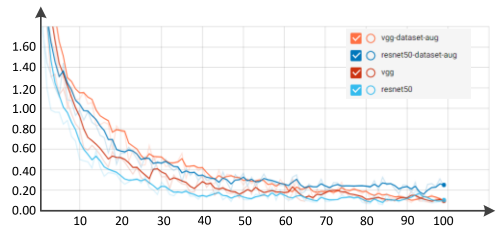

Fig. 5. Loss curve

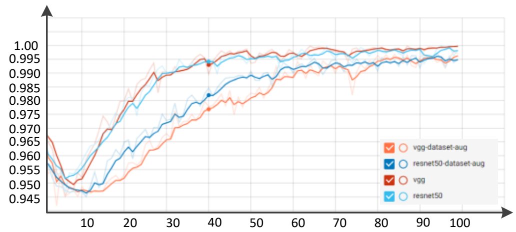

Fig. 6. Average detection accuracy curve

The experimental results are shown in Table II. In the original dataset, the performance of Faster R-CNN based on the VGG-16 model is very poor. The Recall and Precision on the entire test set are only 54.54% and 38.71% respectively and the MR is as high as 45.45%, which means a lot of lesions will not be found. Thereinto, the Recall, Precision and MR of GGO are 60.71%, 33.33% and 39.29% respectively, and the three indicators of the consolidation(Con) are 43.75%, 63.64% and 56.25% respectively. Similarly, the performance of Faster R-CNN on the basis of ResNet-50 is also poor, but its overall performance is better than the model based on VGG-16 . Its Recall, Precision and MR on the entire test set

TABLE II EVALUATION RESULTS OF CT LESION DETECTION PERFORMANCE

Datasets	Based network	Lesion	Td	Fd	Missing	MR	Recall	Precision
Datasets -augmented GGO:129 Consolida-tion:78 TOTAL: 207	VGG -16	GGO	92	56	37	28.68%	71.32%	62.16%
		Con	49	18	29	37.18%	62.82%	73.13%
		TOTAL	141	74	66	31.88%	68.12%	65.58%
	ResNet-50（overfitting）	GGO	86	80	43	33.33%	66.67%	51.81%
		Con	39	37	39	50.00%	50.00%	51.32%
		TOTAL	125	117	82	39.61%	60.39%	51.65%
Datasets-original GGO:28 Consolida-tion:16 TOTAL: 44	VGG -16	GGO	17	34	11	39.29%	60.71%	33.33%
		Con	7	4	9	56.25%	43.75%	63.64%
		TOTAL	24	38	20	45.45%	54.54%	38.71%
	ResNet-50	GGO	18	11	10	35.71%	64.29%	62.07%
		Con	9	5	7	43.75%	56.25%	64.29%
		TOTAL	27	16	17	38.64%	61.36%	62.79%

are 61.36%, 62.79% and 38.64% respectively. Correspondingly ,the Recall, Precision and MR of GGO are 64.29%, 62.07% and 35.71% and the three indexes of consolidation(Con) are 56.25% ,64.29% and 43.75% respectively.

Therefore, this article improves the performance of the model by appropriately augment the CT dataset. It is obvious that the performance of Faster R-CNN based on VGG-16 is significantly improved in the augmented dataset, with Recall and Precision being the highest, and MR being the lowest. The three indicators on the whole test set reach 68.12%, 65.58% and 31.88% respectively. For GGO, the Recall, Precision and MR are 71.32%, 62.16% and 28.68%, and for consolidation(Con), they are 62.82%, 73.13% and 37.18%. A sample of the test results is show in the Fig. 7. The left is the annotated image, and the right is the prediction image.

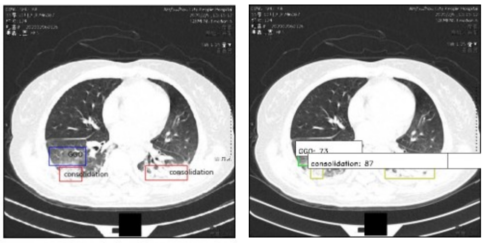

Fig. 7. A sample of detection results

In summary, Faster R-CNN based on VGG-16 achieved the best detection performance in the enhanced dataset. The model based on ResNet-50 can also achieve similar results on the original dataset, but when it was trained on the augmented dataset, over-fitting phenomenon is occurred, which led to the decline of model performance. We will also continue to study the reasons for overfitting to better improve the performance of the Faster R-CNN model.

IV. CONCLUSION

This paper proposes to detect COVID-19 lung lesion from CT images based on modified Faster R-CNN. The original Faster R-CNN is modified by the K-means++ method , which is used to change the size and ratio of the anchors to suit the requirements of the COVID-19 CT dataset . And two basic networks, VGG-16 and ResNet-50, are used as the feature extracted layers. After data enhancement, the VGG-16-based model achieved better detection performance, with Recall and Precision reaching 68.12% and 65.58%, respectively, and MR dropping to 31.88%. Therefore, Faster R-CNN can realize the classification and positioning of lesions in COVID-19 CT images.

At present, the accuracy of the model still needs to be improved, but this study can contribute to track the later condition of the confirmed patients and conduct timely treatment. At the same time, the Faster R-CNN model learns limited features on small data sets, and increasing data sets can effectively improve its performance. Therefore, we will continue to collect data and further research will be conducted to improve the detection accuracy of COVID-19 lesions by adding CT images and improving network structure. In addition, in the enhanced dataset, the overfitting phenomenon of the model based on ResNet-50 also needs further analysis.

REFERENCES

[1] CHEN W, ZHENG R, BAADE P D, et al. Cancer statistics in China,2015［J］. CA Cancer J Clin, 2016, 66(2): 115-132.

[2] Yicheng, F., et al., Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. Radiology, 2020.

[3] Ardakani, A.A., et al., Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks. Computers in Biology and Medicine, 2020. 121: p. 103795.

[4] Liu, B., et al., Online COVID-19 diagnosis with chest CT images: Lesion-attention deep neural networks. medRxiv, 2020: p. 2020.05.11.20097907.

[5] Xu, Y., et al., A collaborative online AI engine for CT-based COVID-19 diagnosis. 2020.

[6] Warman, A., et al., Interpretable Artificial Intelligence for COVID-19 Diagnosis from Chest CT Reveals Specificity of Ground-Glass Opacities. medRxiv, 2020: p. 2020.05.16.20103408.

[7] Gozes, O., et al. Coronavirus Detection and Analysis on Chest CT with Deep Learning. 2020. arXiv:2004.02640

[8] Wang, X., et al., A Weakly-supervised Framework for COVID-19 Classification and Lesion Localization from Chest CT. IEEE Transactions on Medical Imaging, 2020: p. 1-1

[9] Fan, D.-P., et al. Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images. 2020. arXiv:2004.14133

[10] Xie H,Yang D, Sun N, et al. Automated pulmonary nodule detection in CT images using deep convolutional neural networks[J].Pattern Recognit，2019，85：109-119

[11] DING J, LI A, HU Z, et al. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks[C]//Proceedings of the 2017 International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2017: 559-567

[12] Ren, S., et al., Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017. 39(6): p. 1137-114

[13] Gao Ting, et al., 3-D adaptive template matching for detecting lung nodules. Journal of Image and Graphics, 2014. 19(09): p. 1384-1391.(In Chinese)

[14] Han Guanghui , Liu Xiabi, and Zheng Guangyuan, Detection method of lesion area in lung CT image. Acta Automatica Sinica, 2017. 43(12): p. 2071-2090. (In Chinese)

[15] Dou Yong , Qiao Peng and Jin Ruochun, Exploring the shortcomings of the average accuracy AP index and its impact. Science China: Information Science , 2019. 49(10): p. 1369-1382(In Chinese)