2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Full Scale Attention for Automated COVID-19 Diagnosis from CT Images

Zheng Cao[1], Cailin Mu[2], Haochao Ying[3] and Jian Wu[4]

Abstract— **The wide spread of coronavirus pneumonia (COVID-19) has been a severe threat to global health since 2019. Apart from the nucleic acid detection, medical imaging examination is a vital diagnostic modality to confirm and treat the disease. Thus, implementing the automatic diagnosis of the COVID-19 bears particular significance. However, the limitations of data quality and size strongly hinder the classification and segmentation performance and it also result in high misdiagnosis rate. To this end, we propose a novel full scale attention mechanism (FUSA) to capture more contextual dependencies of features, which enables the model easier to classify positive cases and improve the sensitivity. Specifically, FUSA parallelly extracts the information of channel domain and spatial domain, and fuses them together. The experimental study shows FUSA can significantly improve the COVID-19 automated diagnosis performance and eliminate false negative cases compared with other state-of-the-art ones.**

I. Introduction

The large outbreak of Coronavirus Disease 2019 (COVID-19) has caused a global health emergency. This disease causes respiratory illness from mild fever, cough, nasal congestion, fatigue to progressive pneumonia, and even severe respiratory failure and death [1]. Chest computed tomography (CT) image examination is a rapid and accurate method to diagnosis COVID-19. Thus, automatically and efficiently detecting COVID-19 infections area from CT image are meaningful for the proper management and treatment of patients.

Nowadays, deep learning based algorithms have achieved great success in auxiliary medical image diagnosis, such as the lesion recognition in chest CT image, which even perform better than radiologists [2]. Several deep learning models also have been proposed for COVID-19 diagnosis, which can be broadly classified into two categories based on different tasks. One is the infection positive levels classification from suspected patients, while the other is the lesion area segmentation of lung. Hasan et al. [3] proposed a feature extraction method of Q-deformed entropy to classify COVID-19 positive image. Hussain et al. [4] proposed a 22-layer convolutional neural network CoroDet and reported its performance in a set of COVID-19 classification tasks. Pathak et al. [5] utilized transfer learning to overcome the sample unbalance problem. Singh et al. [6] established a CNN based on multi-objective differential evolution to predict whether CT is coronet positive or not. Wang et al. [7] proposed COVID-net with projection-expansion-projection-extension model to enhance the network performance and reduce the computational complexity. Due to the information loss in convolution, directly classification methods still have high false negative rate. As for the segmentation task, Fan et al. [8] designed a parallel partial decoder to aggregate global semantic information and proposed a semi-supervised architecture to enhance the segmentation result. He et al. [9] introduced a self-trans method for self-supervised learning to achieve better feature representation and reduce the risk of overfitting. Gozes et al. [10] combined 2D and 3D segmentation to build a robust segmentation model. Chen et al. [11] applied residual connection neural network in U-Net to enable the model to learn more robust features. Zheng et al. [12] proposed a weak supervision learning architecture to reduce dependence of accurate label. Saeedizadeh et al. [13] improved U-Net segmentation result through adding extra regularization in loss function. However, current segmentation methods cannot completely aggregate the features within the CT images, which also leads to high false negative results.

This paper aims to propose a novel full scale attention (FUSA) mechanism to capture enough contextual dependencies of characteristic from chest CT image. Firstly, we design the FUSA module by fusing the enhanced information of the channel domain and the spatial domain. Next, for COVID-19 classification and segmentation tasks, we propose two deep learning architecture FUSA-ResNet and FUSA-ResUnet respectively. Finally, both the comparative study and visualization results show our method can dramatically improve the COVID-19 diagnosis performance. In particular, our method can suppress the false negative results and has a higher recall.

II. Methods

A. Full Scale Attention Mechanism

In this section, we want to build a module to capture the contextual information as much as possible, so that the model evaluates the global features and obtain the areas that the model should pay attention to at the global level. In most studies, only average pooling is used for the preliminary pooling. In order to enable the module

*This research was partially supported by the National Research and Development Program of China under grant No. 2019YFB1404802.

[1]Zheng Cao is with College of Computer Science and Technology, Zhejiang University, Hangzhou 310044, China z.cao@zju.edu.cn

[2]Cailin Mu is with School of Software Technology, Zhejiang University, Ningbo 315048, China mu_charlie@163.com

[3]Haochao Ying is with School of Public Health, Zhejiang University, Hangzhou 310058, China haochaoying@zju.edu.cn

[4]Jian Wu is with The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310044, China wujian2000@zju.edu.cn

to integrate more layers of characteristic information, we consider average pooling and maximum pooling at the same time. The extracted channel domain features are mapped to the same dimension with the number of channels as the original feature image again through the full connection layer, while the spatial domain features are mapped through the convolution layer of 1×1 convolution kernel to obtain the single-channel feature with the same size as the spatial feature. The attention mechanism of channel domain and spatial domain are not simply superposition. It is considered to join the features of channel domain and spatial domain, and then let the network learn how to integrate the features of channel domain and spatial domain through the convolutional layer. However, the dimensions of channel domain features and spatial domain features that can be used at this time are different. In this paper, channel domain features and spatial domain features are expanded to the shape of the input feature map and then stitched. The channel domain feature uses the value of the feature at the corresponding channel location to fill all the values of the same channel. The spatial domain feature uses the value of the feature in the corresponding spatial position to fill the value of the corresponding position of all channels. After splicing, the new feature map is reduced to the size of the original feature map through a 1×1 size convolution kernel to get the result of feature fusion of channel domain and spatial domain, in which the feature fusion strategy is completely left to the network adaptive learning. In order to get a smoother attention feature map, we utilize a sigmoid layer after fusion layer. It eventually outputs an attention feature map that represents the importance of the information on a global scale.

The schematic diagram of full scale attention mechanism (FUSA) is shown in Fig. 1, assuming that the shape of the original feature map is $H \times W \times C$. For the channel domain branch, the features extracted by maximum pooling and average pooling are stitched together to obtain a feature with a length of $2C$. The full connection layer is used to re-map it to a length of C, and then it is expanded to the shape of the original feature map. For spatial domain branches, the features extracted by maximum pooling and average pooling are spliced together to obtain a feature map with the shape of $H \times W \times 2$. The convolution layer with the convolution size of 1×1 is used to remap the feature map with the size of $H \times W \times 1$, and then it is expanded to the shape of the original feature map. The channel domain is connected to the expanded feature map of the spatial domain, and the size of the feature map after splicing is $H \times W \times 2C$. Since it needs to be transformed into the same feature as the input shape, the convolution and Sigmoid operation are used to make the shape of the convolution $H \times W \times C$, and Sigmoid smoothing is done to finally get the attention feature map.

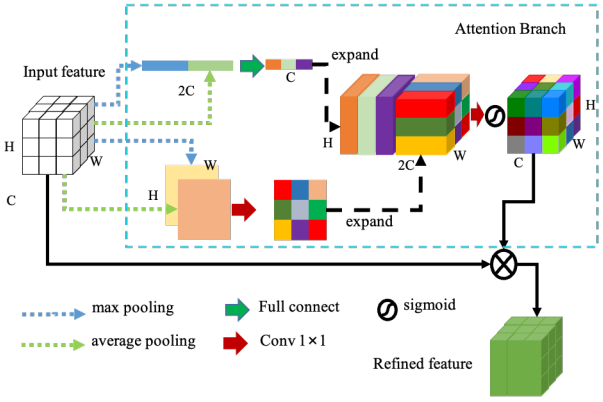

Fig. 1. An overview of full scale attention (FUSA) module.

B. Classification Model Design

For the COVID-19 classification task, we combine FUSA and classic deep learning model ResNet [14], by replacing the residual block with the FUSA-Residual block. As shown in Fig. 2, each residual block contains a FUSA module in the end, to adjust the input feature map globally. Then sum up the feature map adjusted by the FUSA module and the feature map after the residual connection.

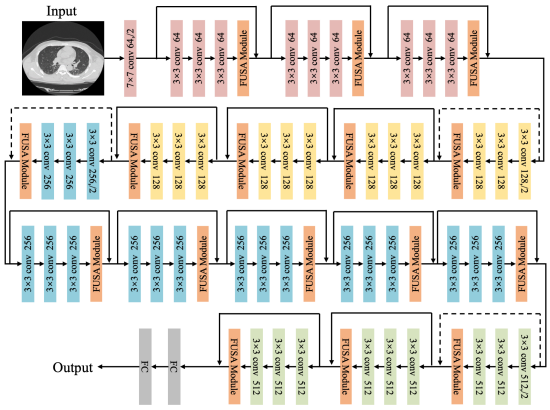

Fig. 2. Diagram of FUSA-ResNet-50 architecture for COVID-19 classification.

C. Segmentation Model Design

For the COVID-19 segmentation task, we designs the FUSA-ResUnet model based on the FUSA module and the U-Net model [15]. As shown in Fig. 3, the basic module of each layer is replaced by the global attention residual module for the ordinary U-Net model. The network is mainly composed of the encoding part of feature extraction and the decoding part of parsing and upsampling. The encoder module is composed of four global attention residuals, each global attention residual module includes two groups of convolution, batch normalization layer, ReLU activation function, and output the sum result of the FUSA module. Moreover, each global attention residuals module is followed by a maximum pooling to subsample the feature.

After the image enters the network, it will first pass through a residual module to obtain the feature with a channel number of 64. After that, the number of channels of the feature map output by the four global attention residual modules is 128,256,512,1024 in turn. The decoder module is basically the same as the encoder module. First, the feature is sampled to the feature size of the corresponding layer of the encoder, and then the multi-scale context information of different network levels is integrated through feature splicing. Then, the intermediate result of the decoding sub-module is obtained through the global attention residual module.

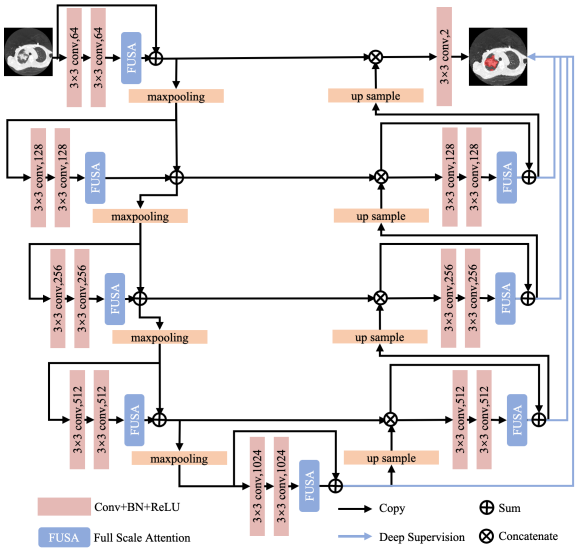

Fig. 3. Diagram of FUSA-ResUnet architecture for COVID-19 Segmentation.

III. Experiments and Results

A. Dataset

The dataset in this paper is provided by the National Institutes of Health (NIH) in the United States for the purpose of research on artificial intelligence for COVID-19 CT image [16]. And it is annotated by the National Children's Hospital in collaboration with Nvidia, which presents CT data from 199 patients with accurate delineation. In this paper, 2D CT slice images are used. Each slice is resized to a square 512×512 image. The dataset includes a total of 13705 CT image slices, in which 4981 slices contains positive COVID-19 lesions. The positive sample accounts for 36.34% in the dataset, and the area containing focal areas accounts for 1.14% of the total area.

B. Data pre-processing and Data augmentation

Due to there are large areas of background in the original data, the random clipping strategy is based on foreground region for the segmentation task. Considering the data imbalance, the up-sampling method is adopted to try to balance the number of samples of positive and negative lesions. At the same time, we set random cutting at different scales to enable the model to focus more on the lesions in some small areas. The clipping method we used is to select 1/2 size, 1/4 size and 1/8 size of the original picture respectively to clip the original picture, and fill the clipped picture back to the size of the original image.

C. Classification Task

In order to analyze the effectiveness of the proposed model, we compared the performance of FUSA-ResNet-50 with VGG-net, ResNet-50, DenseNet and MobileNet in the classification task. Moreover, we also set the experiments of FUSA-ResNet-18 as a comparison to ResNet-18. Tab. I lists the comparative algorithms performance on COVID-19 CT classification task, which reports the accuracy, precision, recall, F1 score and AUC respectively. Among them, the proposed FUSA-ResNet50 achieved the highest accuracy of 88.17%, F1 socre of 84.47% and AUC of 0.94. FUSA-ResNet18 model reaches the best recall of 80.13%. Comparing the FUSA-ResNet-50 with ResNet-50 model, it can be found that the FUSA module improves the accuracy rate of the model by 1.42% and the recall rate by 4.7%. In COVID-19 diagnosis, we need eliminate the false negative as much as we can to prevent the infection, so the recall rate needs to be as high as possible. The increase of recall rate means that the addition of FUSA model can bring better results for the model. Comparing the proposed FUSA-ResNet-18 with the ResNet-18, it can be found that the FUSA module in this paper improves the accuracy rate of the ResNet-18 model by 0.7% and the recall rate by 3.89%. Through the ablation experiments with ResNet-50 and ResNet-18, the FUSA module proposed in this paper can improve the accuracy of the model, and it can significantly improve the recall rate. It shows that the proposed FUSA module can effectively reduce the rate of missed diagnosis of the model, which is of great significance for the clinical application.

TABLE I
Comparative Study of Classification Results.

Model	Acc	Precision	Recall	F1	AUC
VGG-Net	81.80%	89.16%	62.42%	73.43%	0.90
DenseNet	84.78%	89.30%	70.19%	78.80%	0.92
MobileNet	86.75%	86.97%	78.95%	82.77%	0.93
ResNet-18	86.89%	89.69%	76.24%	82.42%	0.93
FUSAResNet18	**87.59%**	**88.00%**	**80.13%**	**83.88%**	**0.94**
ResNet-50	86.75%	**90.34%**	75.16%	82.05%	0.93
FUSAResNet50	**88.17%**	**89.66%**	**79.86%**	**84.47%**	**0.94**

D. Segmentation Task

In COVID-19 CT image segmentation task, we set the experiments to compare the performance of our proposed FUSA-ResUnet aganist LinkNet [17], ResUNet [15], Deeplabv3 [18] and PSPNET [19], as shown in Tab. II. The dice coefficient, precision and recall of each algorithm are reported. With full scale attention module, FUSA-ResUnet reaches 89.76% dice, 88.82% precision and 77.38% recall respectively. In terms of the comparison with ResUnet, FUSA-ResUnet has a 1.93% increasing in dice coefficient, 4.55% higher precision and a significant 9.72% recall improvement.

TABLE II

EVALUATION OF SEGMENTATION RESULTS.

Model	Dice	Precision	Recall
LinkNet	87.41%	88.01%	63.95%
ResUNet	87.83%	84.27%	67.66%
PSPNet	88.67%	86.47%	70.25%
DeepLabV3	89.29%	83.11%	75.08%
FUSA-ResUnet	**89.76%**	**88.82%**	**77.38%**

E. Result Visualization and Discussion

The visualization was carried out through the CT labeling software ITK-SNAP. The comparison of the segmentation result of different model is shown in Fig. 4. The blue lesions are correct prediction (true positive), the green lesions are incorrect prediction (false positive), and the red lesions are misdiagnosed prediction (false negative). Basically, the FUSA-ResUnet can have a robust segmentation result, which eliminate most false negative area in ResUnet. The model could achieve a better segmentation result on some large lesions in comparison of other methods. However, in terms of lesion in early symptoms, it still have misdiagnosis area. In other words, there are fewer green regions and more red regions in the segmentation results. This is consistent with the fact that the precision of the model is relatively high and the recall is relatively low in the indicators shown in Tab. II.

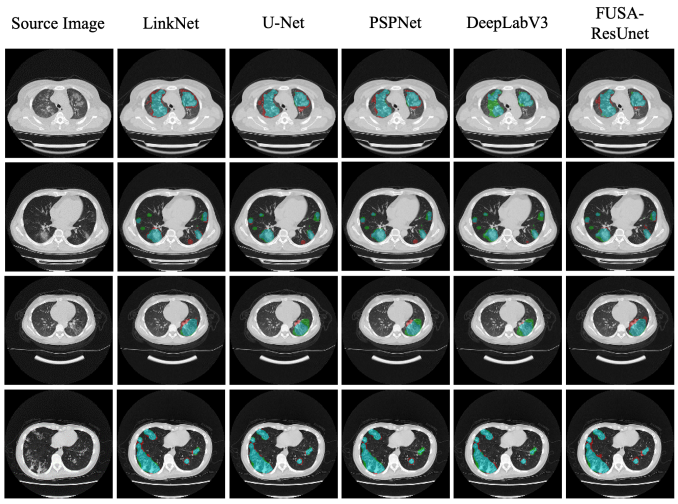

Fig. 4. Visualization of segmentation results.

IV. CONCLUSION

In conclusion, we proposed a full scale attention mechanism FUSA to enhance the deep learning performance in COVID-19 diagnosis from CT images. For classification and segmentation tasks, our improved model FUSA-ResNet and FUSA-ResUnet were tested on a public COVID-19 CT dataset respectively. The experimental results showed promising results that attention mechanism can significantly improve the model performance. Moreover, FUSA based methods can effectively reduce the false negative rate, which was also vital for epidemic control by finding as much as suspected cases. FUSA-ResNet-50 achieved 88.17% testing accuracy for COVID-19 classification and 89.76% dice for segmentation of infection lesions. In this work, we only considered 2D CT slice image for the tasks in this paper. With the introduction of 3D contextual attention information, the model may be able to perform better.

REFERENCES

[1] T. P. Velavan and C. G. Meyer, "The covid-19 epidemic," *Tropical medicine & international health*, vol. 25, no. 3, p. 278, 2020.

[2] Z. Cao, B. Yu, B. Lei, H. Ying, X. Zhang, D. Z. Chen, and J. Wu, "Cascaded se-resunet for segmentation of thoracic organs at risk," *Neurocomputing*, 2021.

[3] A. M. Hasan, M. M. Al-Jawad, H. A. Jalab, H. Shaiba, R. W. Ibrahim, and A. R. AL-Shamasneh, "Classification of covid-19 coronavirus, pneumonia and healthy lungs in ct scans using q-deformed entropy and deep learning features," *Entropy*, vol. 22, no. 5, p. 517, 2020.

[4] E. Hussain, M. Hasan, M. A. Rahman, I. Lee, T. Tamanna, and M. Z. Parvez, "Corodet: A deep learning based classification for covid-19 detection using chest x-ray images," *Chaos, Solitons & Fractals*, vol. 142, p. 110495, 2021.

[5] Y. Pathak, P. K. Shukla, A. Tiwari, S. Stalin, and S. Singh, "Deep transfer learning based classification model for covid-19 disease," *Irbm*, 2020.

[6] D. Singh, V. Kumar, M. Kaur *et al.*, "Classification of covid-19 patients from chest ct images using multi-objective differential evolution–based convolutional neural networks," *European Journal of Clinical Microbiology & Infectious Diseases*, vol. 39, no. 7, pp. 1379–1389, 2020.

[7] Z. Wang, Q. Liu, and Q. Dou, "Contrastive cross-site learning with redesigned net for covid-19 ct classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 10, pp. 2806–2813, 2020.

[8] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Inf-net: Automatic covid-19 lung infection segmentation from ct images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2626–2637, 2020.

[9] X. He, X. Yang, S. Zhang, J. Zhao, Y. Zhang, E. Xing, and P. Xie, "Sample-efficient deep learning for covid-19 diagnosis based on ct scans," *MedRxiv*, 2020.

[10] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, and E. Siegel, "Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis," *arXiv preprint arXiv:2003.05037*, 2020.

[11] X. Chen, L. Yao, and Y. Zhang, "Residual attention u-net for automated multi-class segmentation of covid-19 chest ct images," *arXiv preprint arXiv:2004.05645*, 2020.

[12] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and X. Wang, "Deep learning-based detection for covid-19 from chest ct using weak label," *MedRxiv*, 2020.

[13] N. Saeedizadeh, S. Minaee, R. Kafieh, S. Yazdani, and M. Sonka, "Covid tv-unet: Segmenting covid-19 chest ct images using connectivity imposed u-net," *arXiv preprint arXiv:2007.12303*, 2020.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[15] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[16] P. An, S. Xu, S. Harmon, E. Turkbey, T. Sanford, A. Amalou, M. Kassin, N. Varble, M. Blain, V. Anderson, F. Patella, G. Carrafiello, B. Turkbey, and W. BJ, "Ct images in covid-19 [data set]," *The Cancer Imaging Archive*, 2020.

[17] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE Visual Communications and Image Processing (VCIP)*. IEEE, 2017, pp. 1–4.

[18] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv preprint arXiv:1706.05587*, 2017.

[19] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2881–2890.