

Research Article

Improved 3D U-Net for COVID-19 Chest CT Image Segmentation

Ruiyong Zheng , Yongguo Zheng , and Changlei Dong-Ye 

Collage of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao, Shandong 266590, China

Correspondence should be addressed to Yongguo Zheng; skd991317@sdust.edu.cn

Received 5 April 2021; Accepted 10 June 2021; Published 25 June 2021

Academic Editor: Shah Nazir

Copyright © 2021 Ruiyong Zheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Coronavirus disease 2019 (COVID-19) has spread rapidly worldwide. The rapid and accurate automatic segmentation of COVID-19 infected areas using chest computed tomography (CT) scans is critical for assessing disease progression. However, infected areas have irregular sizes and shapes. Furthermore, there are large differences between image features. We propose a convolutional neural network, named 3D CU-Net, to automatically identify COVID-19 infected areas from 3D chest CT images by extracting rich features and fusing multiscale global information. 3D CU-Net is based on the architecture of 3D U-Net. We propose an attention mechanism for 3D CU-Net to achieve local cross-channel information interaction in an encoder to enhance different levels of the feature representation. At the end of the encoder, we design a pyramid fusion module with expanded convolutions to fuse multiscale context information from high-level features. The Tversky loss is used to resolve the problems of the irregular size and uneven distribution of lesions. Experimental results show that 3D CU-Net achieves excellent segmentation performance, with Dice similarity coefficients of 96.3% and 77.8% in the lung and COVID-19 infected areas, respectively. 3D CU-Net has high potential to be used for diagnosing COVID-19.

1. Introduction

Coronavirus disease 2019 (COVID-19) has rapidly spread worldwide since its outbreak in December 2019 [1, 2]. In March 2020, the World Health Organization declared COVID-19 as a global pandemic [3].

The reverse transcription polymerase chain reaction (RT-PCR) test is the standard for COVID-19 detection. However, this test has a high false negative rate, and it cannot accurately detect the initial infection. Hence, infected patients cannot be diagnosed on time [4, 5]. Compared with the RT-PCR test, chest computed tomography (CT) provides higher sensitivity in the diagnosis of COVID-19; therefore, it can be used as one of the main clinical detection methods [6, 7].

The chest CT scans of patients with COVID-19 show characteristic imaging features, such as ground-glass opacity and occasional consolidation plaques in the lungs [8–10], which are considerably useful for diagnosing COVID-19 and evaluating the severity of a patient's condition. However, owing to a significant increase in the number of patients, it

has become quite challenging to use chest CT scans for COVID-19 detection because of the large workload and experience requirements for doctors.

Numerous deep learning methods have been used to segment and quantitatively analyse infected areas in chest CT scans [11–15]. Li et al. [11] proposed an automatic neural network architecture to detect COVID-19 from chest CT scans and distinguish it from other types of pneumonia and lung diseases. Fan et al. [12] used a Siamese convolutional neural network to assay COVID-19 and automatically evaluate the severity of lung diseases. Gao et al. [13] improved the detection capacity of the network for small lesions and improved the interpretability of the network using a lesion attention module with a dual-branch combination network and an attention mechanism. Paluru et al. [14] proposed Anam-Net to address anomalies in COVID-19 chest CT images. A lightweight convolutional neural network was embedded in Anam-Net, which contained significantly fewer parameters compared to U-Net. The use of Anam-Net in embedded systems demonstrated its applicability to point-of-care platforms. Yan et al. [15] proposed a

high-accuracy network (COVID-SegNet) to segment COVID-19 lesions from chest CT images. COVID-SegNet used multiscale feature fusion and enhanced features to segment lung and COVID-19 lesions accurately and automatically. Although these methods play an important role in the diagnosis and analysis of COVID-19, they are based on CT slices. These methods frequently neglect the correlation between continuous CT slices and cannot fully utilise the spatial information of CT scans.

It is challenging to automatically segment the lesions of COVID-19 pneumonia because of the complexity of CT spatial imaging, the difficulty of marking infected areas, and the difference between medical image characteristics. First, infections may have different characteristic appearances, such as ground-glass opacity and consolidation plaques. Second, lesions have irregular shapes and fuzzy boundaries, and a few lesions have a lower contrast compared to surrounding areas. Third, it is tedious and time consuming to artificially mark pulmonary infection, and it is frequently influenced by doctors' knowledge and clinical experience of lesions [9, 10, 12, 16].

We propose a deep learning method, named 3D CU-Net, to improve the segmentation performance of the neural network models for COVID-19. In addition, we propose a new feature encoding module (residual channel attention, Res_CA) for 3D CU-Net. In the feature extraction stage, the channel attention mechanism of local cross-channel information is used to recalibrate the feature weight of global information and enhance the performance of feature representation. We propose a pyramid fusion module with multiscale global information interaction in the bottom encoder, which enhances the performance of the network by fusing the feature information of different scales and improving the performance of the network for lesion area segmentation.

2. Related Work

2.1. U-Net Structure. U-Net [17] was proposed by Ronneberger et al. in 2015 for medical cell segmentation. It consists of a contraction path to obtain context information and an expansion path to recover a feature map. As high-level and low-level semantic information has the same importance in image segmentation, U-Net combines the high-definition features of an encoder with the advanced semantic features of a decoder stage to help restore the details of a target and obtain an accurate output.

2.2. Variants of U-Net. Numerous methods based on U-Net have achieved better results in different medical image segmentation tasks by integrating the new design concepts of networks. Oktay et al. [18] added an attention mechanism based on U-Net for targets with different shapes and sizes and used an attention gate to highlight the salient features of a skip connection. Xiao et al. proposed a model, named Res U-Net [19], with a weighted attentional mechanism to deal with extreme changes in the ocular vascular background. Feng et al. proposed

CPFNet [20], which improved the segmentation performance by utilising two pyramid modules to fuse multiscale context information.

Wang et al. proposed a new cross-channel information interaction method, named ECA-Net [21], to recalibrate features. They prevented the adverse effect of dimensionality reduction in SE-Net on channel attention. However, most U-shaped networks use only abstract features, neglect certain details, and cannot effectively use multiscale context information [20].

3. Proposed Method

3.1. Network Overview. We propose an automatic segmentation model, named 3D CU-Net, for COVID-19 lesions. The model is based on the 3D U-Net architecture, as shown in Figure 1. The network structure of 3D CU-Net is composed of a feature encoding module (Res_CA) with an attention mechanism, a pyramid dilated convolution module (PDS block) for extracting and fusing multiscale information at different resolutions, and a feature decoding module for segmentation. A fixed-size 3D slice extracted from a 3D CT image is used as the input of the network. The predicted segmentation result is obtained after a series of upsampling and downsampling operations for feature encoding and decoding. The model can ensure continuity between CT images and retain a certain amount of interlayer information. Thus, the 3D input contains more contextual information compared to a 2D image.

In the feature encoding part, an efficient channel attention mechanism [21] is used to reallocate feature weights under the guidance of global information, and residual networks are used to mitigate problems such as gradient vanishing. Global average pooling is used to obtain multiscale global information under different receptive fields to enhance the feature representation in the PDS module, thereby improving the segmentation performance of the network for the irregular shapes and sizes of lesions. Finally, segmentation results are obtained by a feature decoding module, which includes two consecutive $23 \times 3 \times 3$ convolutions and a residual connection with a $1 \times 1 \times 1$ convolution.

3.2. Feature Encoding Block. As shown in Figure 2, the feature encoding module mainly consists of the following two parts.

3.2.1. Feature Extraction. In each encoding module, except for the bottom encoder, two continuous $3 \times 3 \times 3$ convolutions are used to extract deeper feature information. This expands the receptive field, extracts more feature information, improves the complexity of the network, and reduces the amount of calculation and number of parameters. After each $3 \times 3 \times 3$ convolution, we add the ReLU activation function and batch normalisation to alleviate the problem of gradient disappearance and increase the speed of network learning.

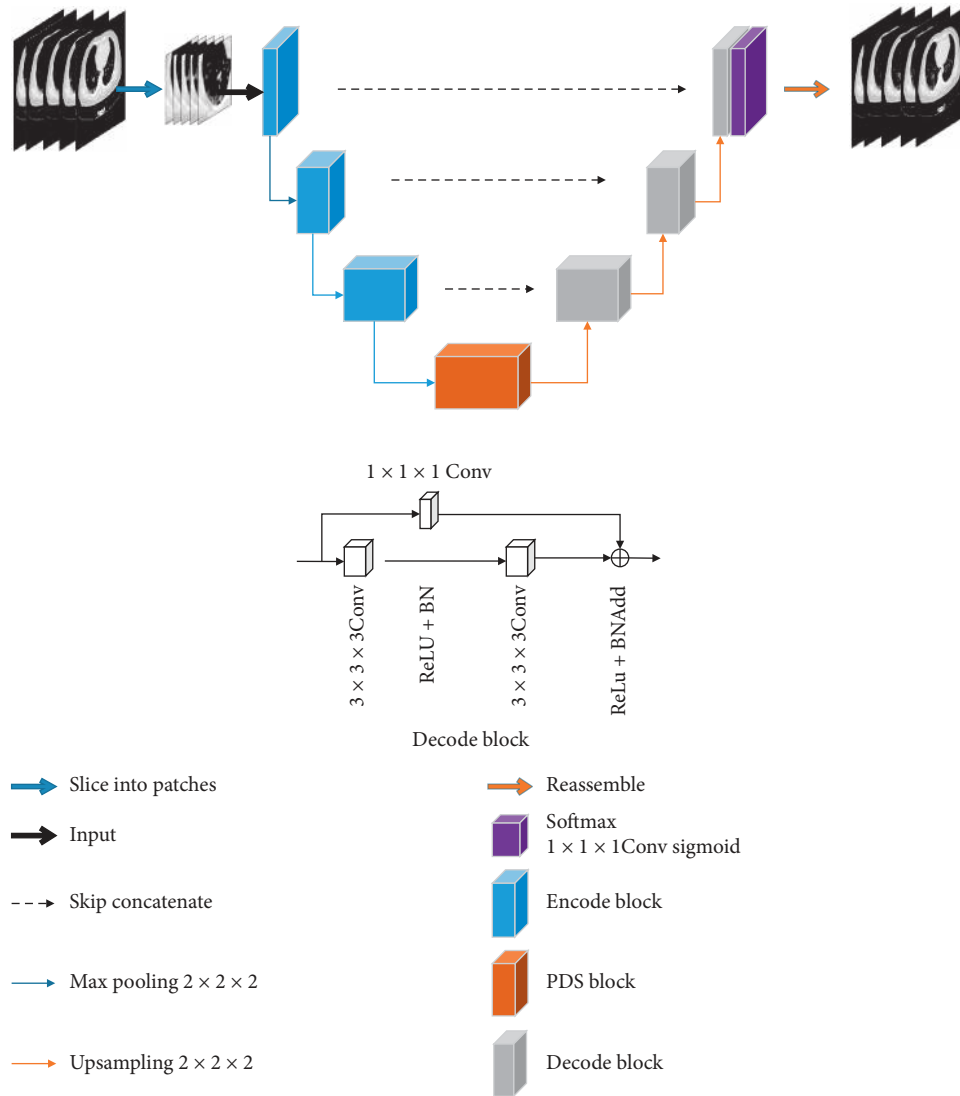


FIGURE 1: Overview of 3D CU-Net.

3.2.2. *Feature Calibration Block.* We introduce a channel attention mechanism to obtain representative features and highlight useful information. According to the correlation between adjacent channels, cross-channel interactive fusion methods are used to recalibrate the weights of the extracted features. Cross-channel information communication can effectively prevent the influence of the reduction in dimensions on channel attention and enhance the feature representation of lesion areas.

3.3. *Pyramid Fusion Module for Dilated Convolutional Global Information Interaction.* Multiscale context information helps improve the performance of semantic segmentation. Thus, we propose a pyramid fusion module that converts low-scale global information into high-scale features. As shown in Figure 3, a residual block is used to deepen the network and extract feature information. Then, a parallel expanded convolution with expanded sizes of 1, 2, and 4 is used to obtain the multiscale information of advanced features. Next, according to the correlation between feature

channel information at different scales, global average pooling is used to obtain the global channel features and their weights at different scales. Thus, the global information obtained in a small receptive field is used to enhance the feature expression ability of a large receptive field. Finally, the features at different scales are fused by stitching.

In the last part of this module, we connect the multiscale feature information that has been recalibrated with feature weights, normalise it using a $1 \times 1 \times 1$ convolution, and then fuse it with the original advanced features.

3.4. *Feature Decoding Block.* As shown in the decoding block in Figure 1, two $3 \times 3 \times 3$ convolutions and a residual connection with a $1 \times 1 \times 1$ convolution are applied to the feature map after the series connection, and a feature map is obtained with the same size as that of the original input image.

3.5. *Loss Function.* In the medical image segmentation task (lesion detection), the high imbalance of the training data leads

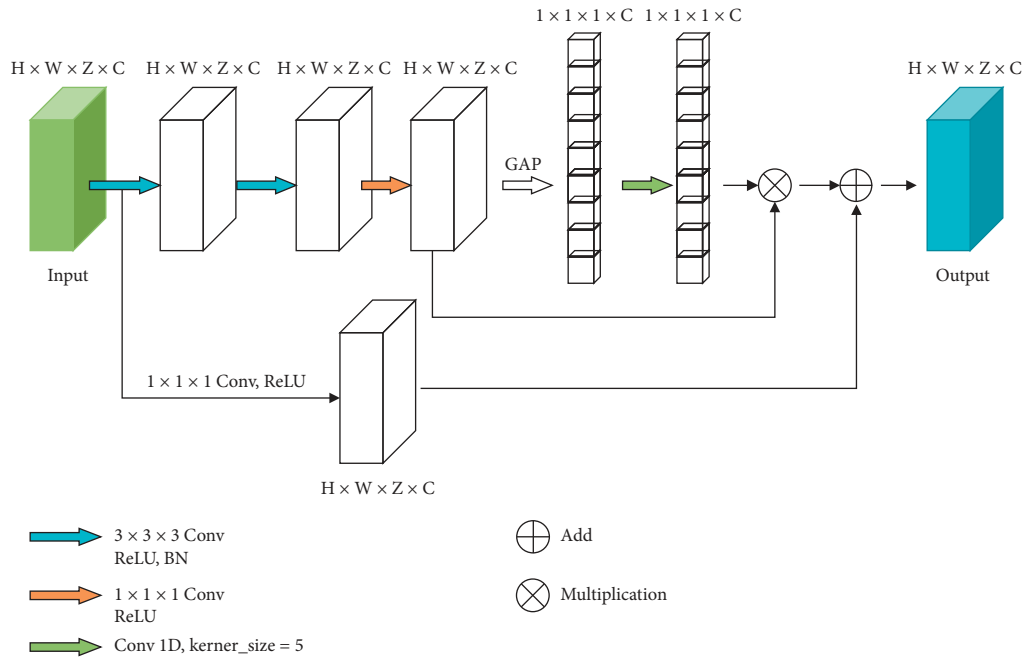


FIGURE 2: Network structure of the feature encoding block.

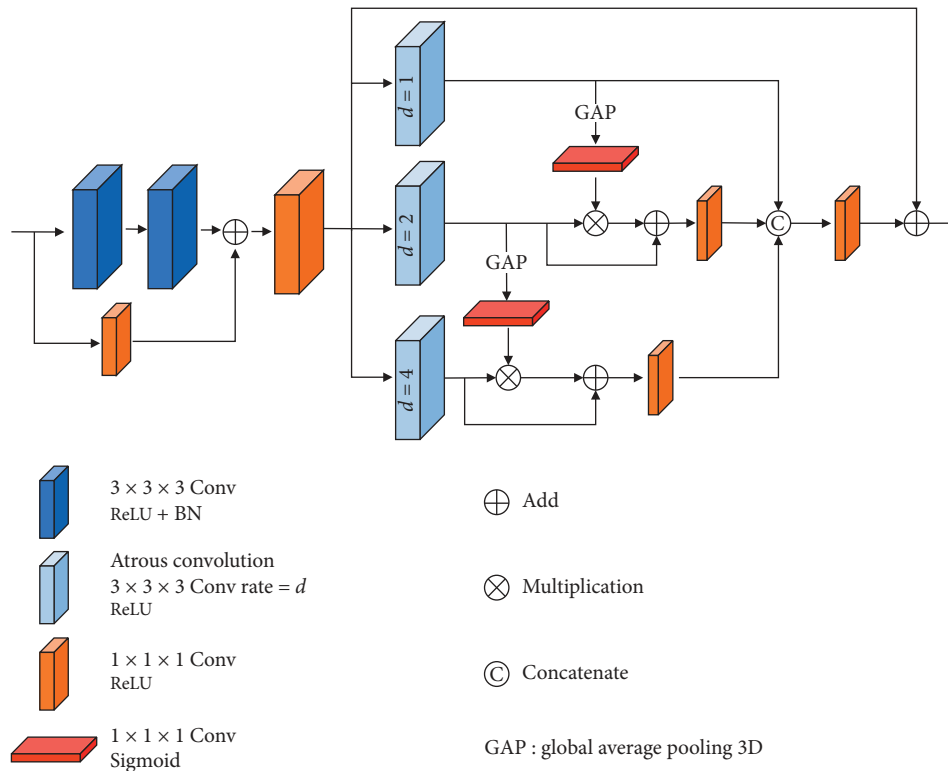


FIGURE 3: Convolutional global information interaction fusion network with pyramid expansion.

to high precision and low recall. Lesion areas account for only 1.12% of the total area in our data. We employ the combination of categorical cross-entropy loss and the Tversky loss as the loss

function of segmentation to resolve the problem of unbalanced data category distribution and improve the performance of network generalisation. The loss function is expressed as

TABLE 1: Comparison of results obtained using 3D CU-Net and 3D U-Net.

	Lung_left			Lung_right			Lesion		
	DSC	Sens	Spec	DSC	Sens	Spec	DSC	Sens	Spec
3D U-Net	0.925	0.943	0.997	0.948	0.959	0.997	0.705	0.707	0.999
3D CU-Net Ours_1: $\alpha = 0.5, \beta = 0.5$	0.966	0.968	0.998	0.964	0.975	0.997	0.778	0.738	0.999
3D CU-Net Ours_2: $\alpha = 0.3, \beta = 0.7$	0.960	0.969	0.998	0.963	0.966	0.998	0.771	0.837	0.998

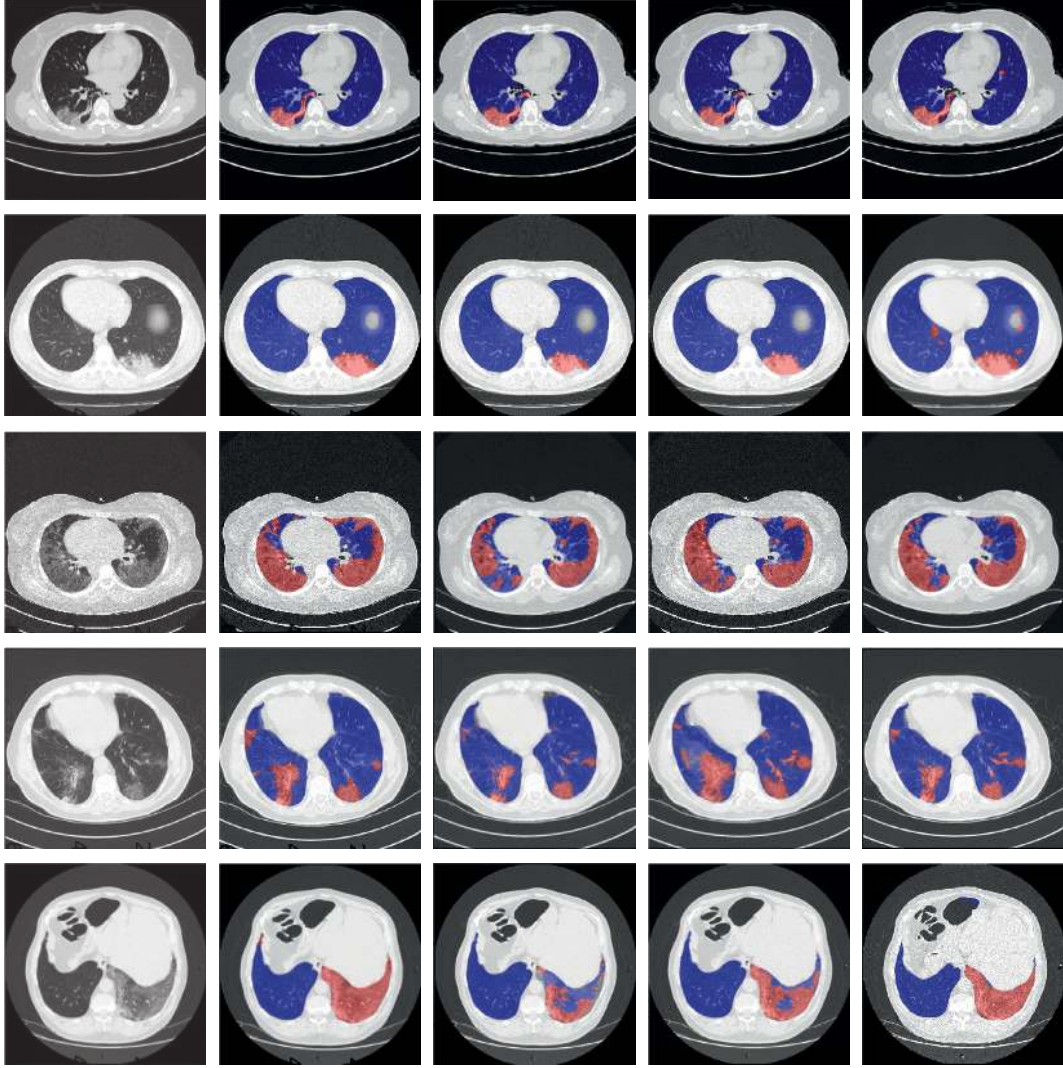


FIGURE 4: Visualisation of segmentation results.

$$\text{Loss}_{\text{Total}} = \text{Loss}_{\text{CE}} + \text{Loss}_{\text{Tversky}},$$

$$\text{Loss}_{\text{CE}} = -\frac{1}{N} \sum_{i=1}^N y_i \log f_i(x),$$

$$\text{Loss}_{\text{Tversky}} = N - \sum_{i=1}^N \frac{\text{TP}(c)}{\text{TP}(c) + \alpha \text{FN}(c) + \beta \text{FP}(c)},$$

(1)

where x is the input value, y_i is the true label corresponding to category i , and $f(x_i)$ is the model output value. TP, FN, and FP represent true positive, false negative, and false positive, respectively.

4. Experiments and Results

4.1. Experimental Data. We train and evaluate 3D CU-Net using the open COVID-19 CT dataset provided by Jun et al.

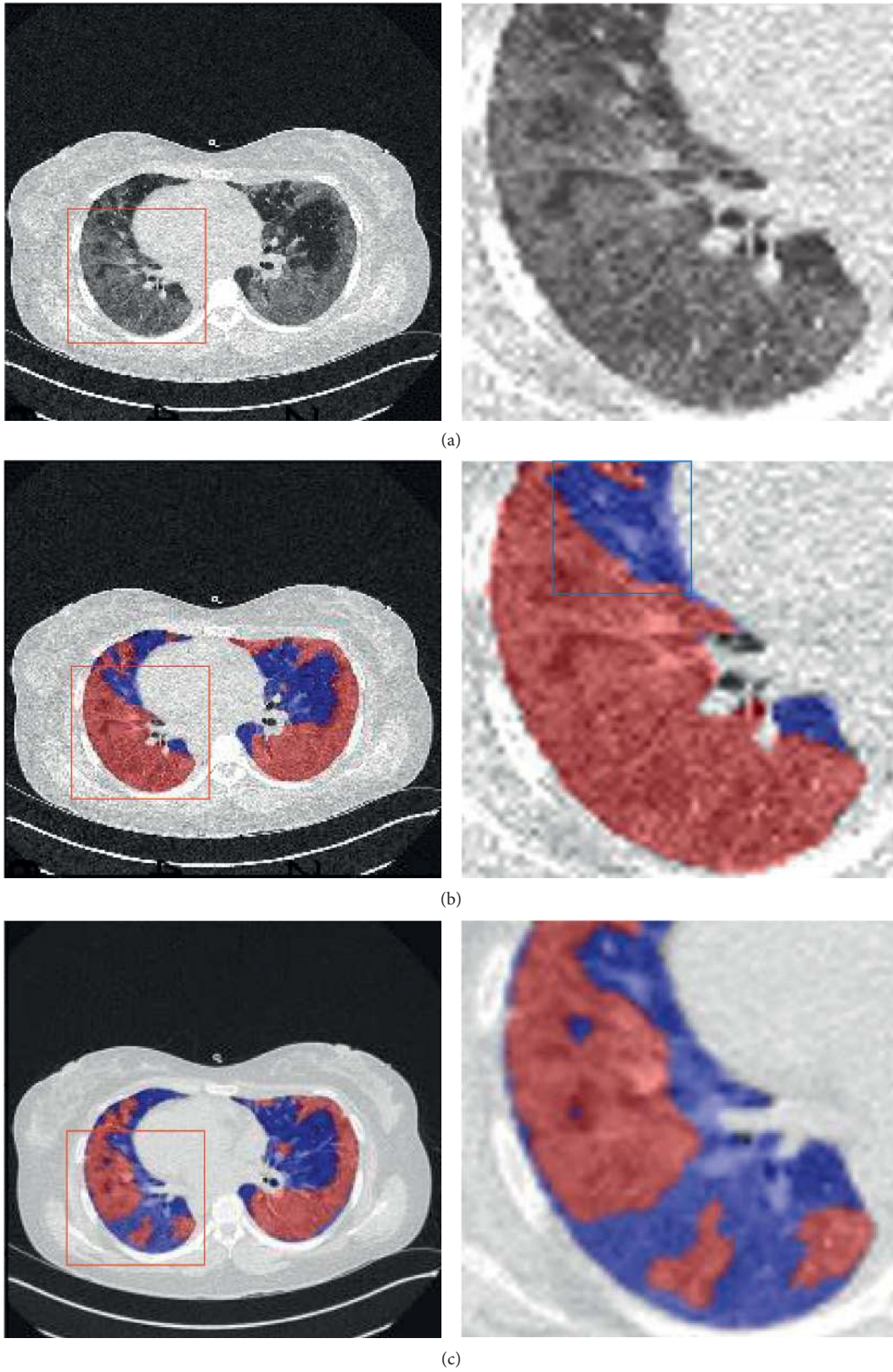


FIGURE 5: Continued.

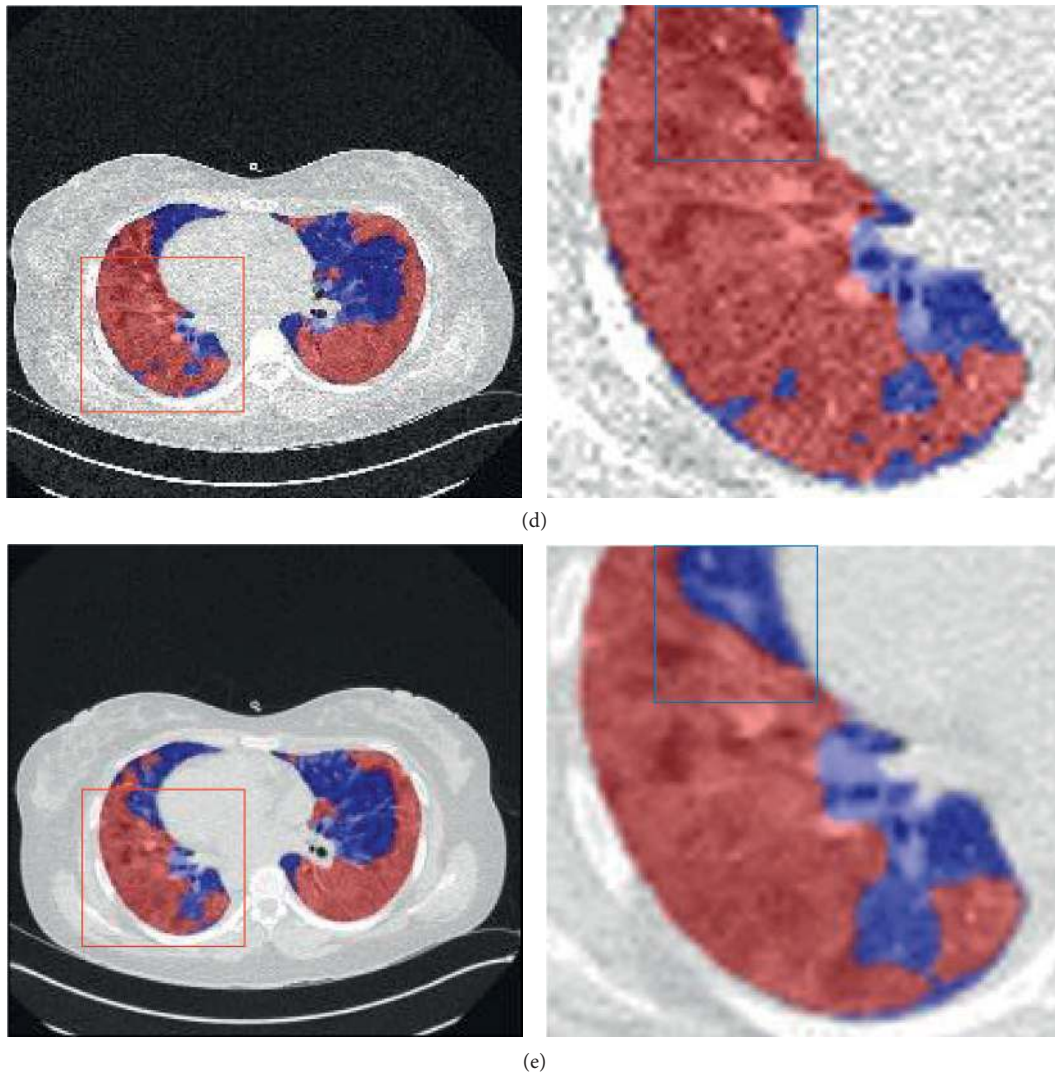


FIGURE 5: Local details of COVID-19 segmentation results.

[22]. In addition, MosMedData [23] provided by Forrest is used as an independent test dataset to further verify the performance of the model. The COVID-19 CT dataset consists of the chest CT scans of 20 COVID-19 patients validated with annotations by a senior radiologist, in which the left lung, right lung, and lesion areas are annotated. The dataset contains 3250 CT slices with sizes of 630×630 , 512×512 , and 401×630 . The lesion area accounts for only 2.12% of the CT slice area, and the slices with lesion markers account for 52.86% of the total slices. MosMedData was provided by Moscow Municipal Hospital. It consists of the chest CT scans of 50 confirmed COVID-19 patients, with lesion areas annotated by a few experts.

Ma et al. [24] proposed a COVID-19 infection segmentation benchmark based on 3D nnU-Net. The Dice coefficients of the COVID-19 CT dataset and MosMedData are 0.673 and 0.588, respectively.

We normalise the training data and pixel values between $[0, 1]$ by considering -250 and 1250 as thresholds. In addition, 3D CT images are resampled at a fixed isomorphic

resolution to normalise them into the same voxel spacing. We use random elastic deformation, random rotation, random scaling, Gaussian noise, and other common medical image data enhancement methods to enhance the training data and prevent the overfitting problem caused by a small amount of training data.

4.2. Experimental Details and Evaluation Metrics. 3D CU-Net is compared with standard 3D U-Net [25] in terms of segmentation results, and the performance of 3D CU-Net is further analysed using MosMedData.

We build an operating environment on a Linux server. The NVIDIA Tesla P100 GPU is used, and the TensorFlow 2.0 deep learning framework is adopted. The installation environment comprises cuda10.0, cudnn7.6.5, python3.6, opencv, gcc, etc. In the fitting process, we set the batch size as 2. We use Adam optimisation with an initial learning rate of 0.001 and a minimum learning rate of 0.00001. We reduce the learning rate by 0.1 times when loss does not decrease in

TABLE 2: Comparison of results on the basis of MosMedData.

	Lesion		
	DSC	Sens	Spec
3D U-Net	0.609	0.584	0.999
3D CU-Net Ours_1: $\alpha = 0.5, \beta = 0.5$	0.664	0.651	0.999
3D CU-Net Ours_2: $\alpha = 0.3, \beta = 0.7$	0.668	0.734	0.998

15 epochs. The training process ends when loss does not decrease in 50 epochs.

We employ 5-fold cross-validation for model fitting. Sixteen sets of CT images are used for fitting and the remaining four sets for validation. After each fitting process, the model is evaluated using the validation data.

In addition, after fitting, we utilise three widely used metrics in medical image analysis to evaluate the segmentation performance of the model for the left and right lungs and COVID-19 lesion areas. These metrics are the Dice similarity coefficient (DSC), sensitivity (Sens), and specificity (Spec).

4.3. Experimental Results. As shown in Table 1, the DSC is 77.8% and Sens is 73.8% for 3D CU-Net, compared with 3D U-Net, and the DSC and Sens increased by 7.3% and 3.1%, respectively. By adjusting the parameters of the Tversky loss ($\alpha = 0.3$ and $\beta = 0.7$), Sens for 3D CU-Net increases to 83.7% with few losses of DSC. In addition, the accuracy of overall lung segmentation improves.

Figure 4 shows the segmentation results obtained using 3D CU-Net and 3D U-Net for five different slices. The images from left to right are the original CT image, ground truth, segmentation results of 3D U-Net, results of 3D CU-Net, and results of 3D CU-Net with the Tversky loss parameters as $\alpha = 0.3$ and $\beta = 0.7$.

Figure 5 shows the local details of the CT image slice segmentation results shown in the third row of Figure 4. In the first column of Figure 5, rows (a)–(e) show the original CT image, ground truth, segmentation result of 3D U-Net, segmentation result of 3D CU-Net, and segmentation result of 3D CU-Net with the Tversky loss parameters as $\alpha = 0.3$ and $\beta = 0.7$, respectively. The second column shows the details of the area enclosed by the red box in the first column. 3D U-Net shows poor segmentation performance, and a large infected area is not identified, as shown in row (c). In contrast, 3D CU-Net provides better segmentation performance, and most infected areas are accurately identified, as shown in row (d). The area enclosed by the blue box in the second column of Figure 5 shows that setting $\alpha = 0.3$ and $\beta = 0.7$ effectively reduces the false positive rate of 3D CU-Net and improves the sensitivity of infection region segmentation.

Furthermore, we compare the performance of the model in terms of infected area segmentation on the basis of MosMedData. As shown in Table 2, the performance of 3D CU-Net is better than that of 3D U-Net, with a 5.9% improvement in the DSC and a 15% increase in Sens.

The experiments performed using the COVID-19 CT dataset and MosMedData show that 3D CU-Net provides excellent segmentation performance. For the left lung, right lung, and lesion areas, the DSC is 0.960, 0.963, and 0.771, Sens is 0.969, 0.966, and 0.837, and Spec is 0.998, 0.998, and 0.998, respectively.

It has great potential in evaluating COVID-19 infection. The above results suggest that the 3D CU-Net model has good performance in COVID-19 lesion segmentation.

5. Conclusion

We proposed a deep learning segmentation network (3D CU-Net) for detecting COVID-19 pulmonary infection. The proposed network was based on 3D U-Net. An attention mechanism was introduced for channel features in the encoding stage to enhance the representation ability of features. The full utilisation of the multiscale global information of high-level features extracted from the bottom encoder improved the accuracy of COVID-19 detection. The proposed network has high potential to be used for diagnosing COVID-19.

However, 3D CU-Net has certain limitations. Its accuracy must be improved for the irregular shapes and different sizes of lesions. In addition, the segmentation performance can be improved via further research and by utilising high-quality medical imaging data.

Data Availability

The labeled datasets used to support the findings of this study are publicly available.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors would like to thank Editage (<http://www.editage.com>) for English language editing. This work was supported by the 2016 Excellent Teaching Team Construction Plan of Shandong University of Science and Technology.

References

- [1] H. Lu, C. W. Stratton, and Y. W. Tang, "Outbreak of pneumonia of unknown etiology in Wuhan, China: the mystery and the miracle," *Journal of Medical Virology*, vol. 92, no. 4, pp. 401–402, 2020.
- [2] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health concern," *The Lancet*, vol. 395, no. 10223, pp. 470–473, 2020.
- [3] World Health Organization (WHO), *WHO Director-General's Opening Remarks at the media Briefing on COVID-19 11 March 2020*, WHO, Geneva, Switzerland, 2020, <https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020>.

- [4] Y. Fang, H. Zhang, J. Xie et al., "Sensitivity of chest CT for COVID-19: comparison to RT-PCR," *Radiology*, vol. 296, no. 2, pp. E115–E117, 2020.
- [5] T. Ai, Z. Yang, H. Hou et al., "Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases," *Radiology*, vol. 296, no. 2, pp. E32–E40, 2020.
- [6] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, and J. Liu, "Chest CT for typical coronavirus disease 2019 (COVID-19) pneumonia: relationship to negative RT-PCR testing," *Radiology*, vol. 296, no. 2, pp. E41–E45, 2020.
- [7] N. Zhu, D. Zhang, W. Wang et al., "A novel coronavirus from patients with pneumonia in China, 2019," *New England Journal of Medicine*, vol. 382, no. 8, pp. 727–733, 2020.
- [8] J. Lei, J. Li, X. Li, and X. Qi, "CT imaging of the 2019 novel coronavirus (2019-nCoV) pneumonia," *Radiology*, vol. 295, no. 1, p. 18, 2020.
- [9] M.-Y. Ng, E. Y. P. Lee, J. Yang et al., "Imaging profile of the COVID-19 infection: radiologic findings and literature review," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, Article ID e200034, 2020.
- [10] M. Chung, A. Bernheim, X. Mei et al., "CT imaging features of 2019 novel coronavirus (2019-nCoV)," *Radiology*, vol. 295, no. 1, pp. 202–207, 2020.
- [11] M. D. Li, N. T. Arun, M. Gidwani et al., "Automated assessment and tracking of COVID-19 pulmonary disease severity on chest radiographs using convolutional siamese neural networks," *Radiology: Artificial Intelligence*, vol. 2, no. 4, Article ID e200079, 2020.
- [12] D.-P. Fan, T. Zhou, G.-P. Ji et al., "Inf-net: automatic COVID-19 lung infection segmentation from CT images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2626–2637, 2020.
- [13] K. Gao, J. Su, Z. Jiang et al., "Dual-branch combination network (DCN): towards accurate diagnosis and lesion segmentation of COVID-19 using CT images," *Medical Image Analysis*, vol. 67, Article ID 101836, 2021.
- [14] N. Paluru, A. Dayal, H. B. Jenssen et al., "ANAM-net: anamorphic depth embedding-based lightweight CNN for segmentation of anomalies in COVID-19 chest CT images," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 3, pp. 932–946, 2021.
- [15] Q. Yan, B. Wang, D. Gong et al., "COVID-19 chest CT image segmentation network by multi-scale fusion and enhancement operations," *IEEE Transactions on Big Data*, vol. 7, no. 1, pp. 13–24, 2021.
- [16] L. Li, L. Qin, Z. Xu et al., "Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy," *Radiology*, vol. 296, no. 2, pp. E65–E71, 2020.
- [17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science*, pp. 234–241, Springer, Cham, Switzerland, 2015.
- [18] O. Oktay, J. Schlemper, L. L. Folgoc et al., "Attention u-net: Learning where to look for the pancreas," 2018, <https://arxiv.org/abs/1804.03999>.
- [19] X. Xiao, S. Lian, Z. Luo et al., "Weighted res-unet for high-quality retina vessel segmentation," in *Proceedings of the 9th International Conference on Information Technology in Medicine and Education (ITME)*, pp. 327–331, IEEE, Hangzhou, China, October 2018.
- [20] S. Feng, H. Zhao, F. Shi et al., "CPFNet: context pyramid fusion network for medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 10, pp. 3008–3018, 2020.
- [21] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: efficient channel attention for deep convolutional neural networks," in *Proceedings of the CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.
- [22] M. Jun, G. Cheng, W. Yixin et al., "Covid-19 ct lung and infection segmentation dataset [DB/OL]. Zenodo," 2020, <https://zenodo.org/record/3757476#.YLov8vkzaUk>.
- [23] S. P. Morozov, A. E. Andreychenko, I. A. Blokhin et al., "MosMedData: data set of 1110 chest CT scans performed during the COVID-19 epidemic," *Digital Diagnostics*, vol. 1, no. 1, pp. 49–59, 2020.
- [24] J. Ma, Y. Wang, X. An et al., "Towards efficient covid-19 ct annotation: a benchmark for lung and infection segmentation," 2020, <https://arxiv.org/abs/2004.12537>.
- [25] D. Müller, I. S. Rey, and F. Kramer, "Automated chest CT image segmentation of COVID-19 lung infection based on 3D U-Net," 2020, <https://arxiv.org/abs/2007.04774>.