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

C3D-UNET: A COMPREHENSIVE 3D UNET FOR COVID-19 SEGMENTATION WITH INTACT ENCODING AND LOCAL ATTENTION

Yiming Bao[1 #], *Hexiang Zeng*[2 #], *Chengfeng Zhou*[1], *Chen Liu*[3], *Lichi Zhang*[1],
Dahong Qian[1], *Jun Wang*[1*], *and Hongbing Lu*[2*]

[1] The school of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
[2] The college of Computer Science and Technology, Zhejiang University, China
[3] The Department of Radiology, First Affiliated Hospital to Amy Medical University, Chongqing, China

1

ABSTRACT

For COVID-19 prevention and treatment, it is essential to screen the pneumonia lesions in the lung region and analyze them in a qualitative and quantitative manner. Three-dimensional (3D) computed tomography (CT) volumes can provide sufficient information; however, extra boundaries of the lesions are also needed. The major challenge of automatic 3D segmentation of COVID-19 from CT volumes lies in the inadequacy of datasets and the wide variations of pneumonia lesions in their appearance, shape, and location. In this paper, we introduce a novel network called Comprehensive 3D UNet (C3D-UNet). Compared to 3D-UNet, an intact encoding (IE) strategy designed as residual dilated convolutional blocks with increased dilation rates is proposed to extract features from wider receptive fields. Moreover, a local attention (LA) mechanism is applied in skip connections for more robust and effective information fusion. We conduct five-fold cross-validation on a private dataset and independent offline evaluation on a public dataset. Experimental results demonstrate that our method outperforms other compared methods.

***Index Terms* —**3D COVID-19 segmentation, CT image analysis, deep learning

1. INTRODUCTION

The recent spread of COVID-19 throughout the entire world is cause for great concern. Although the real-time reverse transcriptase polymerase chain reaction (RT-PCR) is the gold standard for diagnosing COVID-19, computed tomography (CT) can provide valuable pneumonia lesion information in a faster manner. Further analysis of CT images can help establish computer-aided diagnosis (CAD) systems for quantitative and qualitative pneumonia evaluation [1].

The fine-detailed regions of the lesions are critical for exploiting CT volumes. On the other hand, manually outlining the lesions is time-consuming, even for experienced physicians. Generally, fast and automatic segmentation algorithms are expected to solve the above problems [2].

Recently, medical-image segmentation has developed rapidly, based on the extensive application of deep-learning models [3]. UNet has proven to be the most successful and efficient network architecture in many medical-image segmentation tasks [4, 5, 6]. Lately, researchers have conducted studies on COVID-19 segmentation, based on UNet architectures and other methods [7, 8]. Voulodimos et al. [7] evaluated the fully convolutional network (FCN) and UNet on a public COVID-19 segmentation dataset [9]. Yao et al. [8] built a framework to augment the training data with relevant knowledge from normal CT slices and trained a model to segment COVID-19 lesions, based on this framework. Wang et al. [14] proposed a 2D segmentation model named COPLE-Net and a noise-robust self-ensemble strategy. Zhou et al. [15] proposed a three-way segmentation network.

The above methods conduct COVID-19 segmentation in a 2D manner. Thus, the abundant 3D anatomic information in the lung region and pneumonia lesions has not been exploited. The main challenge of 3D COVID-19 segmentation is that the pneumonia lesions in the CT volumes exhibit wide variations in appearance, size, shape, and location in the lung region [1].

In this study, we develop a novel network called Comprehensive 3D-UNet (C3D-UNet) for robust and accurate 3D COVID-19 segmentation. Specifically, the proposed method makes two main contributions. First, we introduce an intact encoding (IE) strategy [12] in the down-sampling branch. A novel residual dilated convolution block with increased dilation rates is designed, enlarging the size of the receptive field in the shallow layers. Second, a local attention (LA) mechanism [10] is applied in the skip connection to fuse deep and shallow information, making feature extraction more efficient. Moreover, multi-layer outputs are all leveraged to predict the final lesion mask, which also enables the network to be trained using multi-level deep supervision.

[#] Equal Contribution: Y. Bao and H. Zeng.

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

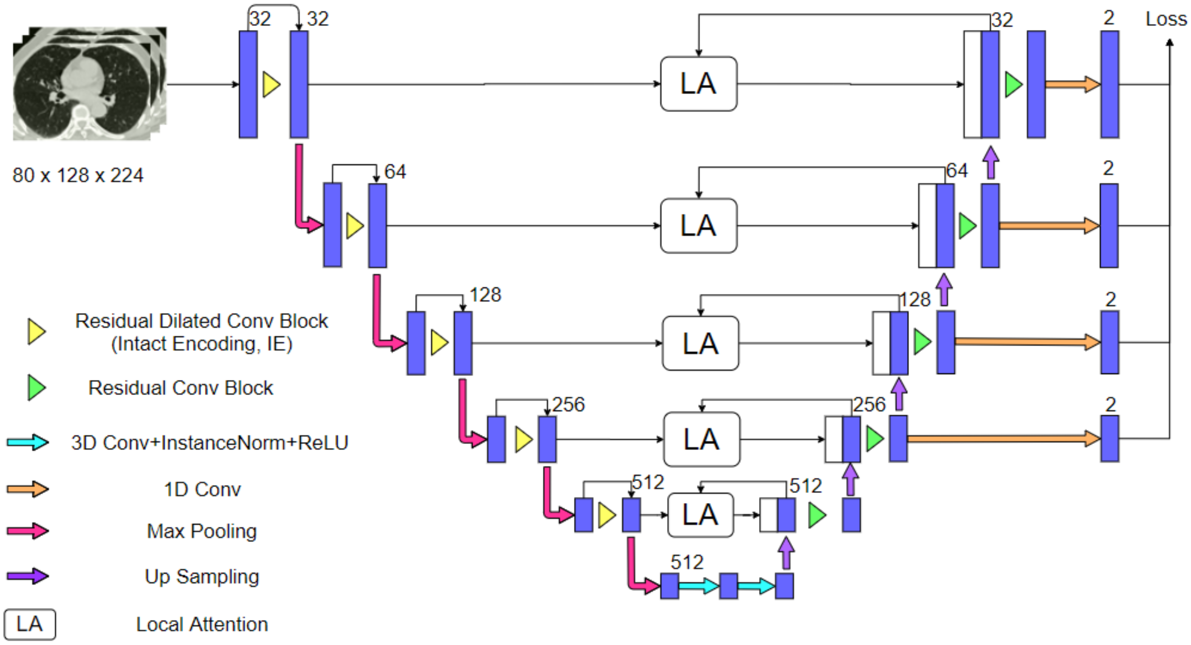

Fig. 2. Framework of the proposed C3D-UNet

We conducted extensive experiments to evaluate the effectiveness of the proposed method. The results of five-fold cross-validation on a private dataset and an independent offline evaluation on a public dataset show that our method achieves state-of-the-art performance.

2. METHOD

In this section, we will detail the C3D-UNet framework and elaborate on its mechanism for achieving more accurate and sophisticated 3D boundaries of pneumonia lesions.

2.1 C3D-UNet

The framework of the proposed C3D-UNet is illustrated in Fig. 1. The UNet series has been one of the most successful segmentation network structures in recent years. In this study, we designed a novel 3D segmentation neural network for pneumonia, especially COVID-19 segmentation, using 3D-UNet as the backbone.

As shown in Fig. 1, our network can be divided into a down-sampling branch, which encodes the fed CT volumes, V_{in}, to feature maps $M_{C\times D\times H\times W}$, and an up-sampling branch, which converts the feature maps back to 3D masks, V_{out} , with the same resolution as that of V_{in}. C denotes the channel number of the feature map, whereas D, H, and W denote the depth, width, and height of the shape, respectively.

The pneumonia lesions in CT volumes are dispersive and invasive, making it challenging to segment all of the lesion regions from an input patch. In the proposed framework, based on the residual convolution, we develop an IE strategy designed as residual dilated convolution blocks. In Fig. 1, each yellow arrow represents a residual dilated convolution block. Each block consists of three stacked dilated convolutional layers. Instead of stacking layers with the same dilation rate of 2 in each block, which will suffer from the gridding effect [11], we stack three layers with increasing dilation rates of 1, 2 and 3 in our blocks.

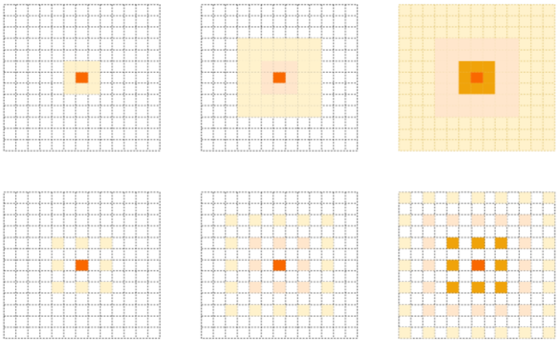

Fig. 2. Comparison of the receptive field and the used pixels. Upper: using different rates of 1, 2, and 3; lower: using the same dilation rate of 2. The deeper the color, the more times the pixel is used.

As illustrated in Fig. 2, our strategy achieves the same receptive field of 13 voxels and simultaneously utilizes information from all the voxels. Thus, the lower layers, rather than only the very deep layers in the network, can also receive sufficient and useful global information. This will be leveraged to help improve the segmentation performance by the skip connection and deep supervision, introduced next.

Although the resolution and contextual information of feature maps can be recovered by up-sampling, most of the texture information is still lost. Skip connections can fuse the texture information from the down-sampling branch with the contextual information from up-sampling, by concatenating the feature maps. However, the texture information in different regions should make different contributions to the lesion segmentation.

In the proposed method, the LA map generated from the up-sampling branch can highlight the local and important lesion regions in the features from the down-sampling branch

2.2 Loss Function

Loss functions can be built by calculating the difference between the predicted 3D masks V_{out} and the ground-truth mask V_{GT}. In this study, we minimize the binary cross entropy between V_{out} and V_{GT}.

Deep supervision is applied to avoid gradient exploration and make the model converge faster. The multi-level features in the up-sampling branch are utilized and aggregated to generate the final predicted mask in the training stage. The final loss function can be represented as follows:

$$L(V_{\text{out}}, V_{\text{GT}}) = \sum_{i=1}^{4} w_i \cdot L(V_{\text{out},i}, V_{\text{GT},i}),$$

where the weights $\{w_i | i = 1, 2, 3, 4\}$ equal 1 for the last layer output, and 0.5, 0.25, and 0.125 for the outputs of the previous layers,

$$L(V_{\text{out},i}, V_{\text{GT},i}) = \sum_{n=0}^{N_i} -p_n \, log(g_n) + (1 - p_n) \, log(1 - g_n),$$

where p_n and g_n are the prediction and ground truth for voxel n in layer i, respectively. The ground truth for layer i is generated by resizing the original mask to the same size as the prediction in layer i. N_i represents the number of voxels in the prediction or ground truth in layer i.

3. EXPERIMENTAL RESULT

3.1 Datasets

We collected one private COVID-19 dataset from several hospitals; the dataset contained 115 CT volumes corresponding to different patients infected by COVID-19. Pneumonia was confirmed in all patients via RT-PCR. All lesion boundaries were carefully drawn by experienced radiologists as ground-truth segmentation labels. This private dataset was used for five-fold cross-validation.

To further validate the proposed method, we conducted an external evaluation using a publicly available COVID-19 segmentation dataset [9]. This dataset consists of 10 CT volumes of confirmed COVID-19 patients. The lungs and areas of infection were labeled by two radiologists and verified by an experienced radiologist.

3.2 Data Preprocessing

Some data preprocessing strategies are applied to make the trained model more robust. First, lung-region coarse segmentation is executed via binarization, a morphological operation, and a region-growing algorithm in a step-by-step manner. Then, the resolution is normalized to make the data isotropic, with a voxel size of $1\,\text{mm} \times 1\,\text{mm} \times 1\,\text{mm}$. Finally, the voxels in each CT volume are normalized to a range of $(0, 1)$ to aid convergence.

After all of these preprocessing steps, $80 \times 128 \times 224$ patches are randomly cropped from the raw CT volumes, augmented by flipping and rotation, and fed into the network for training. In the validation stage, each patch was sampled in a step of half of its shape. The final prediction mask of the entire CT volume was produced by jointing the output patches and voting the overlapping voxels.

3.3 Implementation Details

To evaluate the performance and effectiveness of the proposed C3D-UNet, we conducted five-fold cross-validation on the private dataset for a reasonable comparison with two segmentation baselines, i.e. 3D-UNet and nnUNet.

All the models were implemented using Pytorch and trained on NVIDIA RTX 2080Ti GPUs. For fair comparison, the loss function and hyper parameters remained unchanged when training each model. The loss of each model was minimized by the stochastic gradient descent (SGD) optimizer under an initial learning rate of 0.001 and a decay rate of 10e-7. Each model was trained for 250 epochs.

3.4 Evaluation Metrics and Result

We used several common metrics for segmentation-result evaluation, including the Dice similarity coefficient (DSC), Jaccard similarity coefficient (JSC), and the Hausdorff distance (HD). Table 1 shows the evaluation results for both the private and public datasets. By observing the results, two main conclusions can be drawn:

Table 1. Evaluation results of ablation study

Model	DSC (%)	JSC (%)	HD (mm)
Results on private dataset			
3D-UNet	68.43 ± 4.12	54.85 ± 4.10	24.79 ± 7.99
nnUNet	70.59 ± 3.97	57.89 ± 4.10	22.69 ± 7.42
C3D-UNet (ours w/o IE)	70.52 ± 3.44	57.59 ± 3.74	22.41 ± 6.82
C3D-UNet (ours w/o LA)	71.80 ± 2.87	58.98 ± 2.88	20.92 ± 6.66
C3D-UNet (ours)	$\mathbf{73.41 \pm 2.99}$	$\mathbf{60.77 \pm 3.24}$	$\mathbf{18.73 \pm 7.21}$
Results on public dataset			
3D-UNet	79.84	67.21	11.78
nnUNet	82.37	70.52	6.67
C3D-UNet (ours)	$\mathbf{82.91}$	$\mathbf{71.25}$	$\mathbf{6.30}$

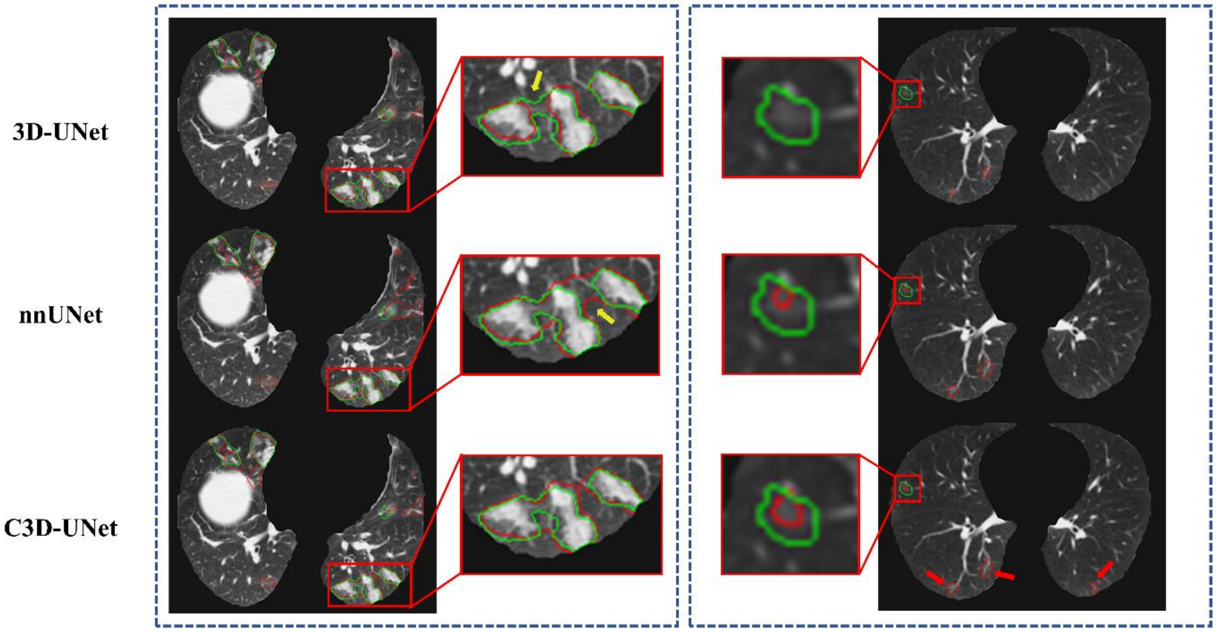

Fig. 3. Segmentation results of severely ill (left column) and moderately ill (right column) patients from private data. The obtained and ground-truth boundaries are outlined in red and green, respectively.

(1) The proposed C3D-UNet achieves both higher DSCs and JSCs and lower HDs than 3D-UNet and nnUNet.

(2) In the ablation experiments, the performances of C3D-UNet without IE and without LA are both inferior to those of the original C3D-UNet. This phenomenon reveals that IE and the LA mechanism are both essential and critical for our method.

Furthermore, we reviewed state-of-the-art methods for COVID-19 segmentation, as listed in Table 2. Except for Muller et al. [13] and our method, the other studies performed COVID-19 segmentation in a 2D manner.

Fig. 3 shows the segmentation results of two representative examples from the private dataset: a severely ill patient and a moderately ill patient, shown in the left and right columns, respectively. The obtained and the ground-truth lesion boundaries are outlined in red and green, respectively.

It can be seen that the proposed C3D-UNet can segment the lesion regions more accurately than both 3D-UNet and nnUNet. For example, both 3D-UNet and nnUNet have some mis-segmented regions, as indicated by the yellow arrows. In addition, the moderate patient contains only some slight lesions appearing as ground-glass nodules, most of which are difficult to annotate, even by experienced radiologists. However, all of the methods, especially C3D-UNet, could successfully identify these lesions (see regions indicated by the red arrows).

Table 2. State-of-the-arts in COVID-19 segmentation

Literature	Dataset	Method	Result
Yao et al. [8]	Coronacases [9]	VAE, nnUNet	DSC: 68.7%
Muller et al. [13]	Coronacases [9]	Data augmentation, 3D-UNet	DSC:76.1%
Wang et al. [14]	Private	2D COPLE-Net	DSC:80.29%
Zhou et al. [15]	Private	multi-view 2D-UNet	DSC:90.3%
Proposed	Private, Coronacases [9]	C3D-UNet	DSC:82.91%

4. CONCLUSIONS

In this study, we developed a novel C3D-UNet for COVID-19 segmentation in a 3D manner. Specifically, the designed framework had two main elements: an IE strategy, designed as a residual dilated convolution block with increased dilation rates for a wider receptive field, and a LA mechanism, which transferred information from the up-sampling branch to the down-sampling branch for effective information fusion.

Ablation studies have indicated that both of the above two elements are critical to the performance improvement of the proposed model. Moreover, experimental results of a five-fold cross-validation on a private dataset and an offline evaluation on a public dataset have demonstrated that our method outperforms the current state-of-the-art methods in 3D COVID-19 segmentation tasks.

5. COMPLIANCE WITH ETHICAL STANDARDS

For this study, ethical approval was obtained, and the informed consent requirement was waived (Approval Number: KY2020036).

6. ACKNOWLEDGMENTS

The authors certify that there is no conflict of interest with any individual/organization for the present work. The work described in this paper is supported in part by the Biomedical Engineering Interdisciplinary Research Fund in Shanghai Jiao Tong University under Grant YG2020YQ17, and by Chongqing Key Technology and Application Demonstration of Medical Depth Intelligent Diagnosis Platform under Grant cst2018jszx-cyztzxX0017.

7. REFERENCES

[1] Wang, J., Bao, Y., Wen, Y., Lu, H., Luo, H., Xiang, Y., ... & Qian, D. (2020). Prior-Attention Residual Learning for More Discriminative COVID-19 Screening in CT Images. IEEE Transactions on Medical Imaging.

[2] Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep learning techniques for medical image segmentation: Achievements and challenges. Journal of digital imaging, 32(4), 582-596.

[3] Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Medical Image Analysis, 101693.

[4] Li, X., Chen, H., Qi, X., Dou, Q., Fu, C. W., & Heng, P. A. (2018). H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. IEEE transactions on medical imaging, 37(12), 2663-2674.

[5] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016, October). 3D U-Net: learning dense volumetric segmentation from sparse annotation. In International conference on medical image computing and computer-assisted intervention (pp. 424-432). Springer, Cham.

[6] Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., ... & Maier-Hein, K. H. (2018). nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486.

[7] Voulodmos, A., Protopapadakis, E., Katsamenis, I., Doulamis, A., & Doulmis, N. (2020). Deep learning models for COVID-19 infected area segmentation in CT images. medRxiv.

[8] Yao, Q., Xiao, L., Liu, P., & Zhou, S. K. (2020). Label-Free Segmentation of COVID-19 Lesions in Lung CT. arXiv preprint arXiv:2009.06456.

[9] Ma, J., Wang, Y., An, X., Ge, C., Yu, Z., Chen, J., ... & Nie, Z. (2020). Towards Efficient COVID-19 CT Annotation: A Benchmark for Lung and Infection Segmentation. arXiv preprint arXiv:2004.12537.

[10] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7132-7141).

[11] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.

[12] Wang P, Chen P, Yuan Y, et al. Understanding convolution for semantic segmentation[C]//2018 IEEE winter conference on applications of computer vision (WACV). IEEE, 2018: 1451-1460.

[13] Müller D, Rey I S, Kramer F. Automated Chest CT Image Segmentation of COVID-19 Lung Infection based on 3D U-Net[J]. arXiv preprint arXiv:2007.04774, 2020.

[14] Wang G, Liu X, Li C, et al. A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images[J]. IEEE Transactions on Medical Imaging, 2020, 39(8): 2653-2663.

[15] Zhou L, Li Z, Zhou J, et al. A rapid, accurate and machine-agnostic segmentation and quantification method for CT-based COVID-19 diagnosis[J]. IEEE transactions on medical imaging, 2020, 39(8): 2638-2652.