2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)
April 13-16, 2021, Nice, France

LCOV-NET: A LIGHTWEIGHT NEURAL NETWORK FOR COVID-19 PNEUMONIA LESION SEGMENTATION FROM 3D CT IMAGES

Qianfei Zhao, Huan Wang, Guotai Wang

School of Mechanical and Electrical Engineering
University of Electronic Science and Technology of China, Chengdu, China

ABSTRACT

The wide spread of coronavirus disease 2019 (COVID-19) has become a global concern and millions of people have been infected. Chest Computed Tomography (CT) imaging is important for screening and diagnosis of this disease, where segmentation of the lung infections plays a critical role for quantitative assessment of the disease progression. Currently, 3D Convolutional Neural Networks (CNNs) have achieved state-of-the-art performance for automatic medical image segmentation tasks. However, most 3D segmentation CNNs have a large set of parameters and huge floating point operations (FLOPs), causing high command for equipments. In this work, we propose LCOV-Net, a lightweight 3D CNN for accurate segmentation of COVID-19 pneumonia lesions from CT volumes. The core component of LCOV-Net is a lightweight attention-based convolutional block (LACB), which consists of a spatiotemporal separable convolution branch to reduce parameters and a lightweight feature calibration branch to improve the learning ability. We combined our LACB module with 3D U-Net as LCOV-Net, and tested our method on a dataset of CT scans of 130 COVID-19 patients for the infection lesion segmentation. Experimental results show that: (1) our LCOV-Net outperforms existing lightweight networks for 3D segmentation and (2) compared with the widely used 3D U-Net, our LCOV-Net improved the Dice score by around 20.36% and reduced the parameter number by 90.16%, leading to 27.93% speedup. Models and code are available at https://github.com/afeizqf/LCOVNet.

Index Terms— COVID-19, 3D CNN, Efficient model

1. INTRODUCTION

COVID-19 poses a huge threat to almost all countries, resulting nearly a million death up to September, 2020. Segmentation of COVID-19 pneumonia lesions from Computed Tomography (CT) images is important for accurate diagnosis and treatment decision. An automatic segmentation system is highly desirable as manual segmentation is time-consuming and depends on experience of the annotator. Recently, CNNs [1, 2, 3] have achieved impressive performance in medical image segmentation and of great value to the COVID-19 pneumonia lesion segmentation task. However, how to balance the accuracy versus efficiency for the semantic segmentation model remains challenging. It is desirable to develop lightweight CNNs to obtain more efficient 3D segmentations with reduced computational cost. Recently, for 2D image classification, many lightweight networks has been proposed. MobileNet [4] reduces parameters and FLOPs with pointwise convolutions and groupwise convolutions. MobileNetV2 [5] used inverted residuals and linear bottlenecks to become lightweight. In ShuffleNet [6], channel shuffle is proposed to alleviate the reduced performance caused by groupwise convolution, and ShuffleNetV2 [7] takes both memory access cost and FLOPs into account to get a trade-off between speed and accuracy. However, few attempts have been made to design efficient segmentation network, especially in 3D medical segmentation tasks. For example, S3D-Unet [8] uses $1 \times 1 \times 3$ convolutions and $3 \times 3 \times 1$ convolutions to replace 3D convolutions. DMFNet [9] combines pointwise convolutions, groupwise convolutions and dilated convolutions for efficient 3D segmentation. Despite their smaller parameter number, they have a reduced segmentation performance compared with typical 3D segmentation models such as the V-Net [2]. Therefore, it remains challenging to keep high segmentation performance with a lightweight model structure.

In this paper, we propose a lightweight CNN (LCOV-Net) for accurate segmentation of COVID-19 pneumonia lesions from 3D CT volumes. We first propose a Lightweight Attention-based Convolutional Block (LACB) to replace standard 3D convolution operations, where the LACB consists of one branch with spatiotemporal separable 3D convolution [10] to reduce the parameters, and another branch with attention-based feature calibration to enhance the block's learning ability. We then combine LACB with a encoder-decoder structure to form our LCOV-Net. Experimental results with 3D CT volumes of 130 COVID-19 patients showed that our LCOV-Net outperformed existing lightweight 3D segmentation networks, and compared with the well recognized 3D U-Net [3], LCOV-Net is around 10 times smaller, and it led to 20.36% increase in terms of Dice score [2], and 27.93% improvement in terms of inference speed.

Corresponding author: Guotai Wang (guotai.wang@uestc.edu.cn)

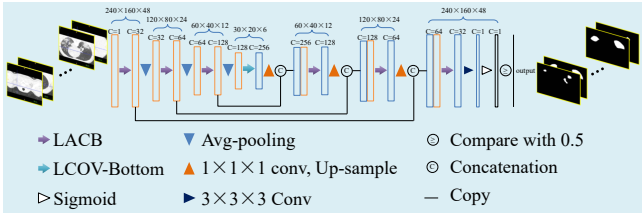


Fig. 1. The proposed LCOV-Net for COVID-19 pneumonia lesion segmentation from 3D CT images. To keep the network lightweight, we replace standard 3D convolution blocks [3] with our Lightweight Attention-based Convolutional Blocks (LACBs) and LCOV-Bottom block, as detailed in **Fig.2**.

2. METHOD

Our proposed LCOV-Net is shown in **Fig. 1**, and it mainly uses LACB and LCOV-Bottom to achieve a lightweight structure. The details of LACB and LCOV-Bottom blocks are illustrated in **Fig. 2(a)** and **Fig. 2(b)**, respectively.

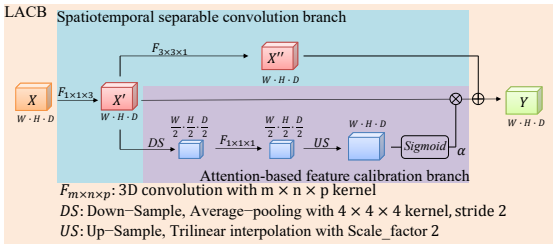
2.1. Lightweight Attention-based Convolutional Block and LCOV-Bottom block

Let X and Y denote the input and output feature map of LACB respectively. As shown in **Fig. 2(a)**, our LACB consists of two branches. The first is a spatiotemporal separable convolution branch that leads to a reduced number of parameters, and the second is an attention-based feature calibration branch that improves the model's performance.

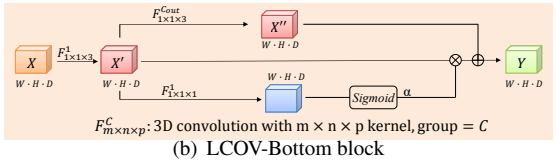
Standard 3D convolutions require a large computational cost. We propose to replace 3D convolutions with spatiotemporal separable 3D convolutions [10] to decrease the parameter number for more efficient computation. This operation utilizes one $1 \times 1 \times 3$ convolution to learn inter-slice (a.k.a., temporal [8]) features and one $3 \times 3 \times 1$ convolution to learn intra-slice features. Each convolution operation is followed by a batch norm layer and a ReLU. Unlike S3D-UNet [8] that divides a 3D convolution into three parallel branches that keep a large number of parameters, we only use one spatiotemporal separable convolution to keep the model lightweight. We denote the output of the inter-slice convolution and intra-slice convolution as X' and X'', respectively.

Using a single spatiotemporal separable convolution will lead to limited feature learning ability despite the reduction of parameter number. To address this problem, we design an attention-based feature calibration branch without introducing many extra parameters. This branch works parallelly with the spatiotemparal separable convolution branch and learns a high-level feature to provide more context information with a larger receptive field.

The feature calibration branch takes X' as input and it consists of four layers. First, a down-sampling layer reduces the resolution of X' by half, then a pointwise convolution is used for feature mapping with inter-channel interaction,



(a) LACB block



(b) LCOV-Bottom block

Fig. 2. Our proposed LACB block and LCOV-Bottom block.

which is followed by an upsampling layer to recover the resolution, and the output is sent into a Sigmoid activation function to obtain an attention coefficient α . We implement the down-sampling by $4 \times 4 \times 4$ average pooling with a stride of 2 and the upsampling by trilinear interpolation. The use of down-sampling followed by upsampling helps to reduce the computational cost for the pointwise convolution and enlarge the receptive field for context learning at the same time. We use α to calibrate X' and the result is added to X''. Thus, the entire LACB block can be summarized as:

$$Y = \alpha X' + X'' \tag{1}$$

Let C_{in} and C_{out} denote the channel numbers in the input and output features of LACB respectively. LACB is used to replace a standard convolutional block with two 3D convolutional layers in 3D U-Net [3]. The parameter number for the standard block is $3^3 \times (C_{in} \cdot C_{out} + C_{out}^2)$, and that for our LACB is $(1 \times 1 \times 3 \times C_{in} \cdot C_{out} + 3 \times 3 \times 1 \times C_{out}^2 + 1 \times 1 \times 1 \cdot C_{out}^2)$, which is $(3 \times C_{in} \cdot C_{out} + 10 \times C_{out}^2)$. Take $C_{in} = 16$ and $C_{out} = 32$ for example, the parameter number is reduced by 3.52 times.

At the bottleneck of LCOV-Net (**Fig. 1**), as the feature map has a small resolution and a large channel number, we use another module named as LCOV-Bottom to reduce parameters. LCOV-Bottom is a variant of LACB, but it additionally uses groupwise convolution (group number = channel number) [4] for the $3 \times 3 \times 1$ convolution in the first branch. And in the feature calibration branch, we remove the down-sampling and up-sampling operations due to the low spatial resolution and only keep the pointwise convolution, as shown in **Fig. 2(b)**.

2.2. LCOV-Net: Lightweight Network for COVID-19 Pneumonia Lesion Segmentation

Theoretically, our LACB and LCOV-Bottom do not depend on a specific network backbone. To demonstrate their ef-

fectiveness, we combine them with the widely recognized encoder-decoder backbone [3] to design our LCOV-Net, as shown in **Fig. 1**. The backbone consists of an encoding path for multi-scale feature learning, and a decoding path to recover fine-grained label for each voxel. A skip connection between the encoder and the decoder is used at each resolution level for better use of low- and high-level features. Differently from U-Net [1], we use our LACB instead of two 3D convolutions at each resolution level of the encoder and decoder, and the bottleneck is replaced with our LCOV-Bottom. To further reduce the number of parameters, we use average-pooling for down-sampling in the encoder, and replace deconvolution in the decoder with pointwise convolution followed by trilinear interpolation.

3. EXPERIMENTS AND ANALYSIS

3.1. Data and Evaluation Metrics

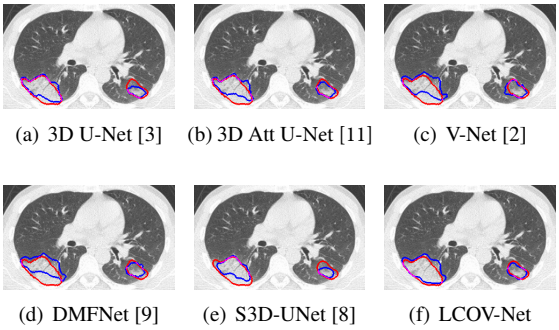
We used clinical chest CT scans of 130 pneumonia patients with COVID-19 from 10 different hospitals. The original image size was 512×512, with pixel size 0.59 to 0.93 mm^2 and inter-slice spacing 1.0 to 5.0 mm. Manual segmentation was used as the ground truth. For preprocessing, we cropped the images by the lung region and normalized the intensity to [0, 1] by a window/level of 1500/-600. For images with a slice number smaller than 48, we pad them along the z-axis to 48 slices by zero. Each slice was then scaled and zero-padded in 2D so that the output size was 320×224. We randomly split the 3D volumes into 80, 20 and 30 for training, validation and testing, respectively.

The segmentation accuracy was measured by Dice score [2] and Average Symmetric Surface Distance (ASSD). The model complexity and efficiency were measured in terms of parameter number and inference time for a 3D volume.

We trained LCOV-Net on a Ubuntu desktop with PyTorch and an NVIDIA GeForce RTX 2080Ti GPU for 400 epochs. For training, we used the Dice loss function [2] and adopted the SGD optimizer with learning rate 0.0025, weight decay 3×10^{-4}. We chose cosine annealing as our learning rate scheduler. During training, images were randomly cropped to a patch of $240 \times 160 \times 48$ voxels. In testing phase, the predict maps were generated by sliding windows of $320 \times 224 \times 48$ for inference and the results were post-processed by morphological opening and closing to reduce noise.

3.2. Comparison with Existing Networks

We compared our LCOV-Net with three widely recognized and high-performance 3D networks: 3D U-Net [3], V-Net [2] and 3D attention U-Net [11]. In addition, it was compared with two existing lightweight structures for medical image segmentation: S3D-Net [3] that combines spatiotemporal convolution with residual inception structure and DMF-Net [9] that uses multi-fiber unit with group convolution.



(a) 3D U-Net [3] (b) 3D Att U-Net [11] (c) V-Net [2]

(d) DMFNet [9] (e) S3D-UNet [8] (f) LCOV-Net

Fig. 3. Visual comparison of different networks for COVID-19 pneumonia lesion segmentation. Red: ground truth. Blue: prediction.

These networks were trained in the same way as LCOV-Net.

Table 1 shows the quantitative comparison between these networks. The average Dice and ASSD achieved by LCOV-Net was 78.67% and 5.78 mm, respectively. Compared with the widely used 3D U-Net, LCOV-Net improved the Dice score by around 20.36% and reduced the parameter number by 90.16%, leading to 27.93% speedup. Our model also achieved higher accuracy than V-Net, and it is more than 3 times faster than V-Net. The parameter number of our LCOV-Net was only 0.8M, and it takes 0.8s in average for segmenting a 3D volume. Compared with existing lightweight 3D networks S3D-UNet and DMFNet, LCOV-Net is more efficient, and its segmentation is more accurate.

Fig. 3 shows a visual comparison of the results obtained by different networks. It demonstrates that segmentation obtained by LCOV-Net is very close to the ground truth in shape, size and location. In contrast, a lot of under-segmentation and over-segmentation are obtained by the other networks, which demonstrates that LCOV-Net has an excellent performance in dealing with the COVID-19 pneumonia lesions.

3.3. Ablation Study

For ablation study, we compared LCOV-Net with three variants: LCOV-Net-A represents only the spatiotemporal separable convolution branch is used in our LACB. LCOV-Net-B represents the the spatiotemporal separable convolution branch is used with skip connection in LACB, without using the feature calibration branch. LCOV-Net-C denotes that the bottleneck uses LACB instead of the LCOV-Bottom.

Quantitative comparison in **Table 2** shows that all these variants outperform 3D U-Net [3], which indicates that the spatiotemporal separable 3D convolutions may be more suitable to our dataset with a large range of inter-slice spacing. LCOV-Net surpasses the other three models in terms of Dice score and ASSD, suggesting the effectiveness of feature calibration in the LACB module. It also proves that it is better to use LACB-Bottom at the bottleneck of LCOV-Net.

Table 1. Quantitative comparison of different networks for COVID-19 pneumonia lesion segmentation.

Model	Dice (%)	ASSD (mm)	Params (M)	Runtime (sec/ volume)
3D U-Net [3]	65.36±20.40	6.88±4.03	8.01	1.11
V-Net [2]	75.49±14.07	7.10±4.06	45.60	3.01
3D Att U-Net [11]	71.41±18.51	6.24±3.48	8.06	1.30
S3D-Unet [8]	52.46±21.06	8.68±7.43	1.89	1.05
DMFNet [9]	75.62±12.53	7.27±6.50	1.15	0.82
LCOV-Net	**78.67±13.11**	**5.78±3.12**	**0.79**	**0.80**

Table 2. Comparison of different variants of LCOV-Net.

Model	Dice (%)	ASSD (mm)	Params (M)	Runtime (sec/volume)
LCOV-Net-A	77.34±13.98	6.48±3.39	**0.74**	**0.64**
LCOV-Net-B	77.59±14.35	6.31±3.93	**0.74**	0.65
LCOV-Net-C	77.47±13.33	7.30±7.34	0.79	0.80
LCOV-Net	**78.67±13.11**	**5.78±3.12**	0.79	0.80

4. CONCLUSION

In this paper, we propose a lightweight 3D CNN (LCOV-Net) for accurate and efficient segmentation of COVID-19 pneumonia lesions from CT volumes. We mainly introduce a novel lightweight attention-based convolutional block (LACB), which consists of one branch with spatiotemporal separable convolution to reduce parameters and another branch using attention for feature calibration. The proposed LCOV-Net is based on LACB and an encoder-decoder structure, and the bottleneck is implemented by our LCOV-Bottom structure. Compared with the widely used 3D U-Net, our LCOV-Net improved the Dice score by around 20.36% and reduced the parameter number by 90.16%, leading to 27.93% speedup. In the future, it is of interest to apply our LACB and LCOV-Net to other network backbones and more medical image segmentation datasets.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research was conducted retrospectively using CT scans of COVID-19 patients. Ethical approval was granted by the local Institutional Review Board.

6. ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant 61901084, and by key research and development project (No. 20ZDYF2817) of Sichuan province, China.

7. REFERENCES

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.

[2] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in *3DV*. IEEE, 2016, pp. 565–571.

[3] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger, "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation," in *MICCAI*. Springer, 2016, pp. 424–432.

[4] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *arXiv preprint arXiv:1704.04861*, 2017.

[5] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen, "MobileNetv2: Inverted Residuals and Linear Bottlenecks," in *CVPR*, 2018, pp. 4510–4520.

[6] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," in *CVPR*, 2018, pp. 6848–6856.

[7] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun, "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design," in *ECCV*, 2018, pp. 116–131.

[8] Wei Chen, Boqiang Liu, Suting Peng, Jiawei Sun, and Xu Qiao, "S3D-UNet: Separable 3D U-Net for Brain Tumor Segmentation," in *International MICCAI Brainlesion Workshop*. Springer, 2018, pp. 358–368.

[9] Chen Chen, Xiaopeng Liu, Meng Ding, Junfeng Zheng, and Jiangyun Li, "3D Dilated Multi-fiber Network for Real-time Brain Tumor Segmentation in MRI," in *MICCAI*. Springer, 2019, pp. 184–192.

[10] Saining Xie, Chen Sun, Jonathan Huang, Zhuowen Tu, and Kevin Murphy, "Rethinking Spatiotemporal Feature Learning for Video Understanding," *arXiv preprint arXiv:1712.04851*, vol. 1, no. 2, pp. 5, 2017.

[11] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al., "Attention U-Net: Learning Where to Look For The Pancreas," *arXiv preprint arXiv:1804.03999*, 2018.