RESEARCH ARTICLE

MEDICAL PHYSICS

Label-free coronavirus disease 2019 lesion segmentation based on synthetic healthy lung image subtraction

Chengyijue Fang[1,#] | **Yingao Liu**[1,#] | **Ying Liu**[2] | **Mengqiu Liu**[2] | **Xiaohui Qiu**[3] | **Yang Li**[4] | **Jie Wen**[2] | **Yidong Yang**[5,1]

[1]Department of Engineering and Applied Physics, University of Science and Technology of China, Hefei, China

[2]Department of Radiology, the First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China

[3]Department of Radiology, Bozhou People's Hospital, Bozhou, China

[4]Department of Radiology, the First Affiliated Hospital of Bengbu Medical College, Bengbu, China

[5]Department of Radiation Oncology, the First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China

Correspondence
Jie Wen, Department of Radiology, the First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China.
Email: jiewen@ustc.edu.cn
Yidong Yang, Department of Engineering and Applied Physics, School of Physical Sciences, University of Science and Technology of China, Hefei, China.
Email: ydyang@ustc.edu.cn

#These authors made equal contributions to this work.

Abstract

Purpose: Coronavirus disease 2019 (COVID-19) has become a global pandemic and is still posing a severe health risk to the public. Accurate and efficient segmentation of pneumonia lesions in computed tomography (CT) scans is vital for treatment decision-making. We proposed a novel unsupervised approach using a cycle consistent generative adversarial network (cycle-GAN) which automates and accelerates the process of lesion delineation.

Method: The workflow includes lung volume segmentation, healthy lung image synthesis, infected and healthy image subtraction, and binary lesion mask generation. The lung volume was first delineated using a pre-trained U-net and worked as the input for the following network. A cycle-GAN was trained to generate synthetic healthy lung CT images from infected lung images. After that, the pneumonia lesions were extracted by subtracting the synthetic healthy lung CT images from the infected lung CT images. A median filter and k-means clustering were then applied to contour the lesions. The auto segmentation approach was validated on three different datasets.

Results: The average Dice coefficient reached 0.666 ± 0.178 on the three datasets. Especially, the dice reached 0.748 ± 0.121 and 0.730 ± 0.095, respectively, on two public datasets Coronacases and Radiopedia. Meanwhile, the average precision and sensitivity for lesion segmentation on the three datasets were 0.679 ± 0.244 and 0.756 ± 0.162. The performance is comparable to existing supervised segmentation networks and outperforms unsupervised ones.

Conclusion: The proposed label-free segmentation method achieved high accuracy and efficiency in automatic COVID-19 lesion delineation. The segmentation result can serve as a baseline for further manual modification and a quality assurance tool for lesion diagnosis. Furthermore, due to its unsupervised nature, the result is not influenced by physicians' experience which otherwise is crucial for supervised methods.

KEYWORDS
COVID-19 pneumonia, label-free segmentation, lesion segmentation, unsupervised approach

1 | INTRODUCTION

The coronavirus disease 2019(COVID-19) has become a global public health problem and now still affects billions of people's life. According to the World Health Organization, the pandemic has caused over 2 million deaths by April 2021.[1] The typical symptom of COVID-19 includes cough, fever, and pneumonia after infection.[2] Clinically, computed tomography (CT) scans are commonly used to evaluate the progress and severity of pneumonia[3,4] due to their high resolution in three dimensions and broad availability compared to other imaging modalities. Accurate delineation of pneumonia lesions is vital for evaluating disease progression and assessing the severity of infection which is crucial for treatment decision making.[5] However, manual segmentation is time-consuming and labor-intensive. Therefore, automatic segmentation methods are highly demanded.

In the past decade, deep learning has shown its tremendous power and potential in various radiological applications, including image segmentation,[6,7] disease classification, and synthetic image generation.[8–11] Since the beginning of the COVID-19 pandemic, deep learning, incorporating various image modalities like CT, X-ray, and ultrasound,[4,12–15] has been applied in clinical diagnosis, predicting disease progression,[16] classifying pneumonia types and assessing the severity of infection.[4,17] However, existing methods are mostly based on supervised learning which requires substantial data labeling by radiologists as training references. For example, U-net networks have been used for the classification and segmentation of COVID-19 lesions in CT scans.[18–20] The results vary greatly among different studies, partially due to the inter- and intra-observer variations in the training lesion labeling by different radiologists.[18]

Compared to supervised learning, unsupervised learning does not require training labeling, and hence gets rid of the burden of manual lesion delineation and the inter- and intra-observer inconsistency. For example, Yao et al.[21] proposed a label-free pneumonia lesion segmentation method that employed an unsupervised statistical method to simulate infected lungs from healthy ones. Zhang et al.[22] developed an unsupervised method to augment lung images for better follow-up segmentation training using a conditional generative adversarial network (GAN).[22] However, most of these methods focused on data augmentation with unsupervised networks, and had to rely on supervised networks to train the lesion segmentation process.[6,21]

Cycle consistent GAN (cycle-GAN) is an unsupervised network that has been widely used in medical image analysis, such as synthetic CT generation[23–25] and image transformation between different MRI sequences.[9,24] Inspired by this, we propose a cycle-GAN-based unsupervised framework for COVID-19 lesion segmentation. The cycle-GAN is used to convert infected lung slices to healthy lung slices by transforming pneumonia lesions into normal lung tissues. Then, the lesion is retrieved by subtracting the simulated "healthy" lung from the original image. The network does not require any image pairing or manual training label, hence can improve efficiency and eliminate the inter- and intra-observer inconsistency otherwise presented in supervised networks.

TABLE 1 Dataset information

Database	Samples	Purpose	Description
USTC-1	77	Training	CT Images without labeling
USTC-2	14	Validation	CT Images with labeling
Coronacases	10	Test	CT Images with labeling
Radiopedia	8	Test	CT Images with labeling
UESTC	50	Test	CT Images with labeling

2 | MATERIALS AND METHODS

2.1 | Patient datasets

In this study, CT scans of 77 COVID-19 patients with positive reverse transcription-polymerase chain reaction results were collected between December 2019 and January 2020 in the First Affiliated Hospital of the University of Science and Technology of China (USTC-1 dataset). The data was anonymized before any analysis. The patient and CT scan information is listed in Table 1. All CT images were converted to $1\times1\times1$ mm^3 spatial resolution and cropped to 256×256 pixels per slice. The image window level was set to [-800, 100] HU and all images were normalized to [-1, 1] with a zero background before being fed into network training. We selected 1264 healthy CT slices and 1272 slices with pneumonia lesions from the 77 CT scans as the training dataset.

Lung volumes were extracted from CT images using a pre-trained 2D U-net lung segmentation network which was trained using 285 patients from a public dataset LUNA-16.[26,30] The average dice similarity of this lung segmentation approach reached over 0.98 in the former study.[26] The extracted lung images were then used as the input of cycle-GAN. Another dataset including 14 patients from the First Affiliated Hospital of the University of Science and Technology of China (USTC-2 dataset) was used as the validation set. Three more datasets, including two public databases (Coronacases and Radiopedia) and one obtained from a published study[21] (UESTC), were used as the test set. The details of these datasets are listed in Table 1.
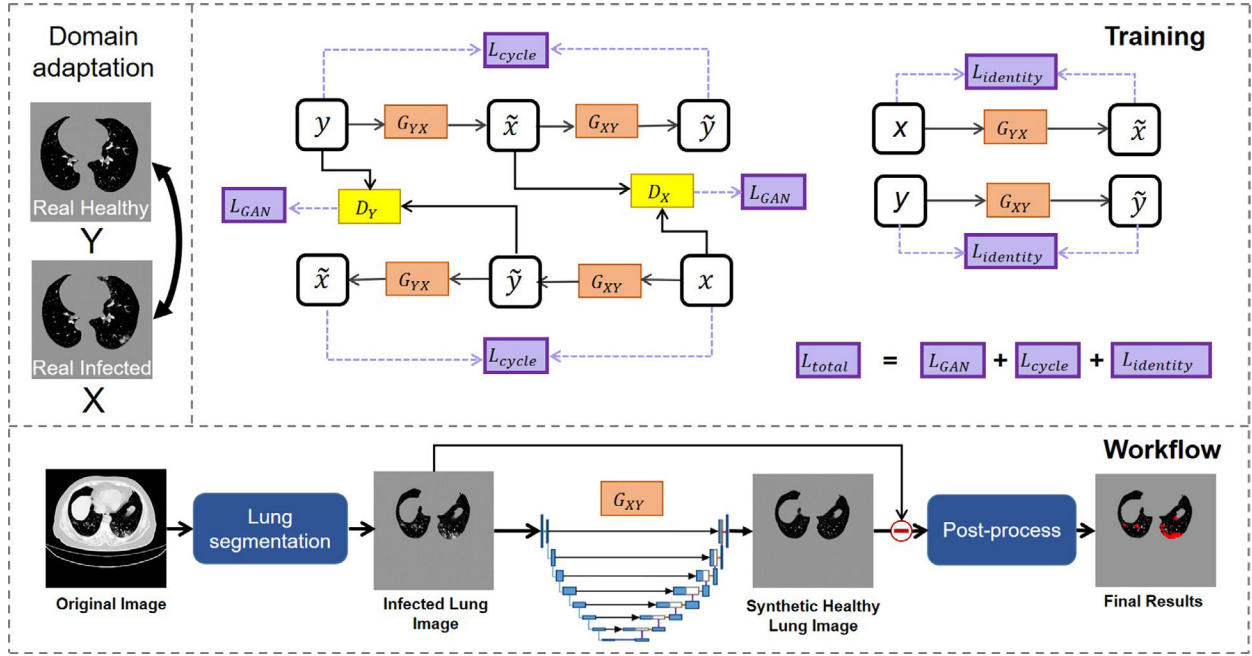
FIGURE 1 The scheme of the proposed method. G_{xy}, G_{yx} are generators for generating synthetic "Healthy" images and "Infected" images. $\ell_{GAN}, \ell_{cycle}, \ell_{idtentity}$ are three loss functions used for training. D_X, D_Y are two discriminators that distinguish real and synthetic images

2.2 | Cycle-GAN structure

In the training stage, we first used a cycle-GAN strategy to generate synthetic healthy lung CT slices. The network architecture is illustrated in Figure 1. We denote the infected lung CT slices by domain X and healthy lung CT slices by domain Y, the probability distribution for each domain is referred as P_x and P_y, respectively. The generator G_{XY} denotes the mapping process from domain X to domain Y, and G_{YX} denotes the mapping from domain Y to domain X. $\tilde{X}$, $\tilde{Y}$ are synthetic "infected" and "healthy" lung slices. Two adversarial discriminators D_X and D_Y are used to distinguish real input images and synthesized images. The architecture of the generator is a U-Net variant and consists of eight stages, as shown in Figure 2. Unlike the original U-net,[7] instance normalization, which can better preserve image details in the image generation process, is applied immediately after each convolutional layer except for the last one. All convolution filters in the generator have a size of 3×3 pixels. We set the channel number of the first block to 64. In the encoder part, the width and height of the feature map are halved using convolution with a stride of 2 instead of max pooling. In the first four stages, the channel number is doubled after the feature map passes each layer, while in the last four stages the channel number is fixed to 512. All the feature maps in the encoder part are concatenated with their counterparts in the decoder part. The encoder and decoder parts are symmetric.

The discriminator D_X and D_Y are implemented by a 70×70 Patch-GAN.[27] The architecture of the discrim-

inator is illustrated in Figure 3. The stride is set to 2 in the first three convolution layers and 1 in the last two layer, and the padding is 1 in all convolution layers. Leaky ReLu activation is applied with a slope of 0.2 after each convolution layer, except the last one. In the first convolution layer, a feature map with 64 channels is generated. After that, the channel number is doubled after the feature map passes each layer. In the last layer, the output is reduced to one channel.

The generators and discriminators are trained by solving a min-max problem: $\min\limits_{G_{XY}, G_{YX}} \max\limits_{D_X, D_Y} \ell_{total}(G_{XY}, G_{YX}, D_X, D_Y)$ where $\ell_{total}(G_{XY}, G_{YX}, D_X, D_Y)$ is defined as:

$$
\begin{aligned}
\ell_{total}\ (G_{XY}, G_{YX}, D_X, D_Y) &= \ell_{GAN}\ (G_{XY}, D_Y, X, Y) \\
&+ \ell_{GAN}(G_{YX}, D_X, Y, X) + \lambda_{cycle}\ell_{cycle}(G_{XY}, G_{YX}) \\
&+ \lambda_{identity}\ell_{identity}(G_{XY}, G_{YX}) \quad (1)
\end{aligned}
$$

where ℓ_{total} is the total loss aiming to learn the mapping function between the source and target domain. λ_{cycle} and $\lambda_{identity}$ are introduced in (1) mainly to weigh the importance of the three losses. After optimization, we set $\lambda_{cycle} = 10$ and $\lambda_{identity} = 5$.

ℓ_{GAN} is the loss function of the discriminator calculating the difference between synthetic "healthy" slices and real healthy slices. To maintain stability during the learning process, we here choose L2 loss in the LSGAN[28] as our loss function instead of the sigmoid cross-entropy in regular GANs.[29] The ℓ_{GAN} is defined
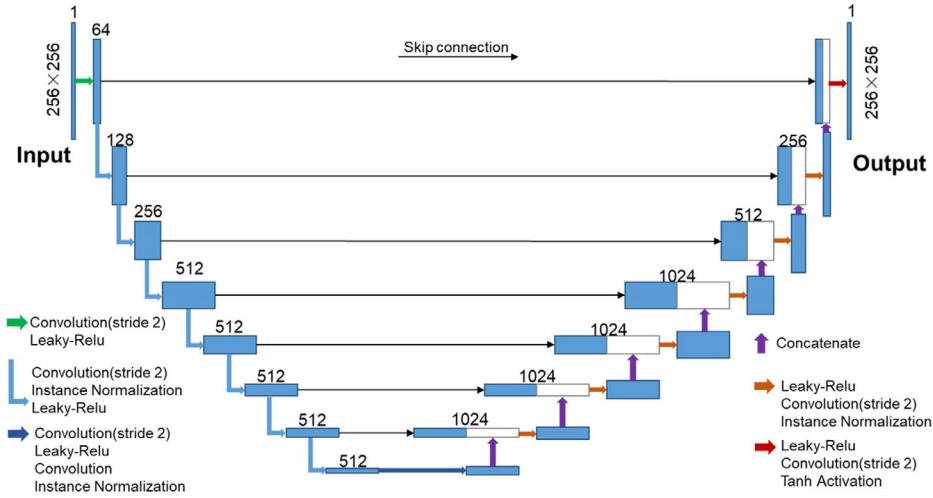
FIGURE 2 The structure of the U-Net generator. The size of the input and output image is 256 × 256. The encoder and decoder parts each have eight stages. Skip connection is applied to each stage
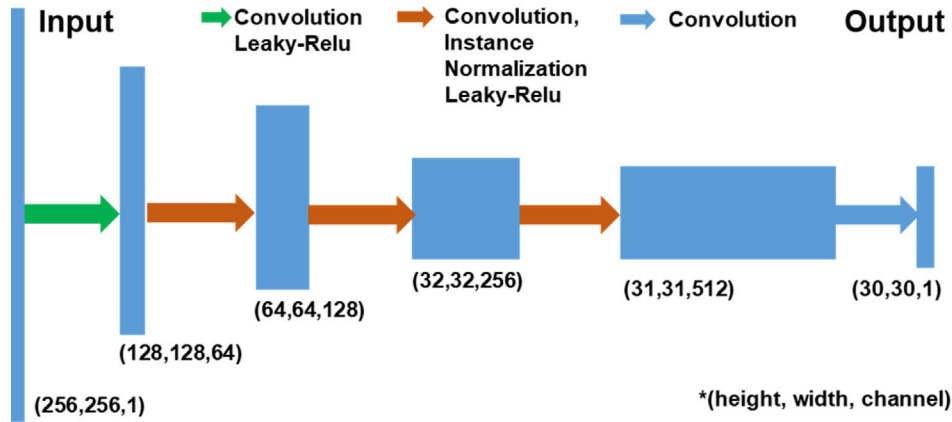
FIGURE 3 The structure of the discriminator. In the 1st layer, convolution and Leaky-ReLu are applied. In the 2nd–4th layers, convolution, instance normalization, and Leaky-Reu are implemented. Only one convolution layer is applied in the last layer

as:

$$\ell_{GAN} = \mathbb{E}_{x \sim P_x} \left[(D_Y (G_{YX} (x)) - 1)^2 \right] + \mathbb{E}_{y \sim P_y} \left[D_Y (y)^2 \right]$$
$$+ \mathbb{E}_{y \sim P_y} \left[(D_X (G_{XY} (y)) - 1)^2 \right]$$
$$+ \mathbb{E}_{x \sim P_x} \left[D_X (x)^2 \right] \qquad (2)$$

where $x \sim P_x$ denotes the learning process on domain X. ℓ_{cycle} is used to keep the consistency of the two generators G_{XY} and G_{YX}, and is defined as:

$$\ell_{cycle} (G_{XY}, G_{YX}) = \mathbb{E}_{x \sim Px} \left[|G_{YX} (G_{XY} (y)) - x|_1 \right]$$
$$+ \mathbb{E}_{y \sim Py} \left[|G_{XY} (G_{YX} (x)) - y|_1 \right] \qquad (3)$$

where $| \cdot |_1$ is the ℓ_1 norm. Since we only want to convert unhealthy lung CT slices into healthy ones, an identity loss $\ell_{identity}$ in (1) is designed to keep the image fea-

tures when a healthy slice is an input into the generator. Identity loss is defined as:

$$\ell_{identity} (G_{XY}, G_{YX}) = \mathbb{E}_{y \sim Py} \left[|G_{XY} (y) - y|_1 \right]$$
$$+ \mathbb{E}_{x \sim Px} \left[|G_{YX} (x) - x|_1 \right] \qquad (4)$$

We used the ADAM optimization method to train all the networks, with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. Kernels were initialized randomly with a Gaussian distribution. We updated the generators and the discriminators at each iteration. The input image slice was randomly cropped to patches of 256 × 256-pixel size. The number of mini-batches was 1, and the number of epochs was 100. The learning rate was initially set to 0.0002 and linearly decreased to 0 in the last 50 epochs. The training was stopped at the 85th epoch which rendered stabilized optimal dice similarity coefficient (DSC) results. All the hyperparameters were validated on the validation set (USTC-2). The training was conducted on a Linux
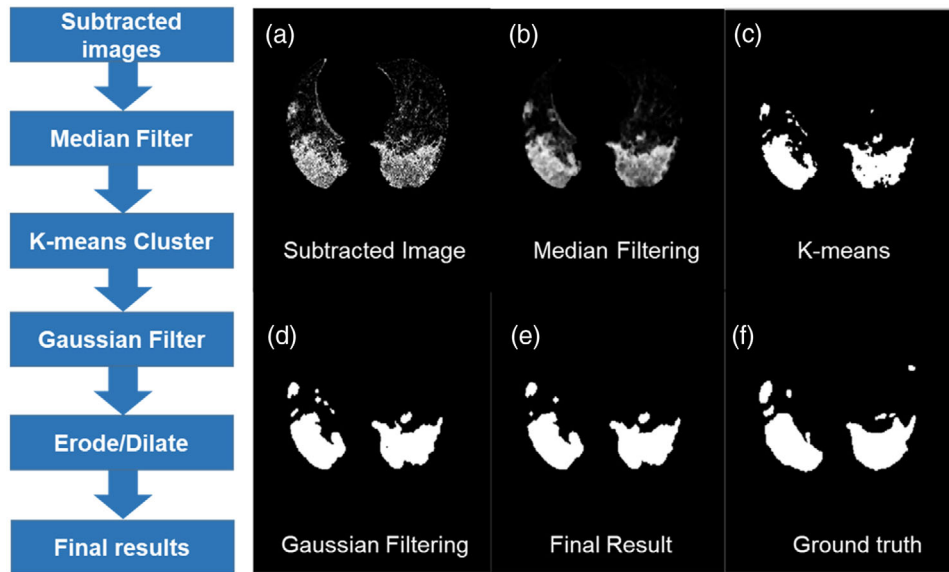
FIGURE 4 The framework of post-processing. (a–e) shows the image after subtraction of the synthetic "healthy" image from infected image (a), median filtering (b), k-means clustering with zero background (c), gaussian filtering (d), and hole filling, erosion, and dilation (e). (f) is the ground truth

machine with an NVIDIA RTX 2080Ti GPU and took 8 h.

2.3 | Image post-processing

The post-processing steps are illustrated in Figure 4. The synthetic healthy image was subtracted from its corresponding real infected image to obtain a different image. The lung CT slices from each patient were reshaped into a 3D volumetric image. Median filtering was applied to suppress noise and remove small islands. Then k-means clustering was used to segment the lesion from the low-intensity background. A 5 × 5 Gaussian kernel was employed, slice by slice, to smoothen the lesion edge. Finally, erosion and dilation with a radius of 1 pixel were performed to further remove small and isolated regions. All post-processing steps, except the Gaussian filtering, were implemented on the 3D volumetric image. The post-processing procedures were done in MATLAB 2018A and took less than 2 min for each patient. As a comparison, we also compared the k-means clustering method with Otsu thresholding which is commonly used for thresholding segmentation.

To summarize the lesion segmentation process, a U-net previously trained on a publicly available dataset LUNA-16[30] was applied to segment the lung volume first. Then a cycle-GAN network was trained and employed to generate the corresponding synthetic healthy lung images from the infected lung images in the USTC-1 dataset. Finally, the lesion was obtained after subtraction of the synthetic healthy lung images from the corresponding infected lung images. The subtraction images

containing lesions were combined into a 3D volumetric image based on which post-processing steps were performed to fine-tune the segmented lesion volume. The whole process is automated and does not need manual labeling of lung lesions as network training references.

2.4 | Segmentation evaluation

In this study, the DSC, volume precision (vPSC), and volume sensitivity (vSEN) defined in Equations (5)–(7) were used to evaluate the performance of the proposed segmentation method. The V_{pre} and V_{gt} in the equations represent the predicted and ground truth lesion volume.

$$\text{DSC } (\%) = \frac{2\left(V_{pre} \cap V_{gt}\right)}{V_{pre} + V_{gt}} \times 100\% \qquad (5)$$

$$\text{vPSC } (\%) = \frac{V_{pre} \cap V_{gt}}{V_{pre}} \times 100\% \qquad (6)$$

$$\text{vSEN } (\%) = \frac{V_{pre} \cap V_{gt}}{V_{gt}} \times 100\% \qquad (7)$$

When it comes to the diagnosis, doctors are more interested in whether there is a lesion and where it is rather than how accurate the delineation is. We proposed an evaluation approach to count whether there is any lesion in different subvolumes of the lung. To mimic this process, the lung volume was divided into 12 subvolumes as illustrated in Figure 5. Each subvolume was counted as positive if there was at least one lesion,
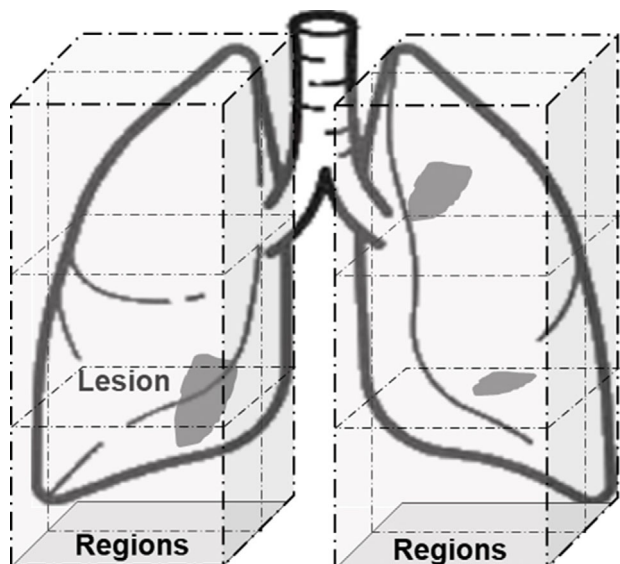
FIGURE 5 The 12 subvolumes in the lung volume division. The lung volume is divided into 12 regions equally

and negative otherwise. Three additional metrics were used for comparison between the proposed segmentation method and other methods. The detection accuracy (ACC), precision (PSC), and sensitivity (SEN) are defined respectively as:

$$ACC(\%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \qquad (8)$$

$$PSC(\%) = \frac{TP}{TP + FP} \times 100\% \qquad (9)$$

$$SEN(\%) = \frac{TP}{TP + FN} \times 100\% \qquad (10)$$

The paired *t*-test was used to statistically compare the proposed method with other methods, that is, nnUNet-2D, nnUNet-3D,[31] Cople-Net,[18] Inf-Net[32,] and NormNet.[21] As for those comparison methods, the nnUNet-2D/nnUNet-3D network was trained and validated on any two of the three datasets (Coronacases, Radiopedia, and UESTC) and tested on the remaining one dataset using a 5-fold cross-validation method; for all other networks, the network parameters were directly borrowed from the previous publications to build networks for the tests run on the three datasets.

3 | RESULTS

Table 2 shows the comparisons between the proposed method and several existing supervised and unsupervised methods. The results of our method reach a dice coefficient of 74.8 ± 12.1% and 73.0 ±

9.5% on the Coronacases and Radiopedia datasets, respectively. The volume precision and volume sensitivity are 81.3 ± 8.8%, 73.5 ± 20.5% on Coronacases and 77.3 ± 17.8%, 72.6 ± 11.1% on Radiopedia. As shown in Table 2, our approach is comparable with the supervised Cople-Net method and outperforms the semi-supervised Inf-Net method on these two datasets. The performance of our method on the UESTC database reaches 64.4 ± 19.2%, which is still higher than Inf-Net. We also compared our method with a state-of-the-art unsupervised label-free[21] method (denoted as "NormNet" in Table 2). Our method results in higher scores in most indices and is more robust than the "NormNet" method when implemented on different datasets. The results also indicate that k-means clustering performs better than Otsu thresholding in image post-processing.

Paired *t*-test results are shown in Table 3. It is observed that the proposed method is comparable with the supervised methods in most cases. For example, our method is comparable to nnUNet-2D, nnUNet-3D, and Cople-Net on the Coronacases dataset, and comparable to nnUNet-3D, Cople-Net, and Inf-Net on the Radiopedia dataset. In UESTC, however, our method renders an inferior outcome compared to the supervised methods, but still outperforms the unsupervised method NormNet.

The proposed method performs well even in small lesion segmentation. Figure 6a shows that a small lesion with only 2-mm width is correctly delineated. As shown in Figure 6b, our method can also readily separate lesions from the chest wall. Interestingly, our method catches some low contrast lesions which are skipped by radiologists during manual segmentation, as shown in Figure 6c.

Figure 7 demonstrates the performance of our method compared to existing supervised methods. These lesions are varied in shape, size, and position. As shown in Figure 7a, our method localized and delineated, while other methods missed, the lesion close to the trachea. Our method performs better than Inf-Net and Cople-Net on a large lesion segmentation as shown in Figure 7b, while is comparable to all existing supervised methods in segmenting small lesions isolated in the lung volume and nearby the chest wall as shown in Figure 7c,d.

As shown in Table 4, the average accuracy of our method on three datasets reaches 79.1%, which is comparable to the supervised methods. Our method outperforms the semi-supervised Inf-Net in detection accuracy and precision, except sensitivity. The results for the paired t-test on detection accuracy are shown in Table 5. As shown both in Tables 4 and 5, there is no significant difference between our method and nnUNet-2D, and a similar capability for lesion diagnosis can be observed between our method and all supervised methods when performed on database Coronacases and Radiopedia.
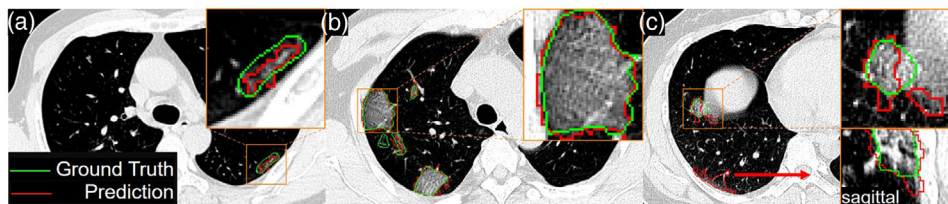
FIGURE 6 Segmentation results for lesions of various sizes. (a–c) are segmentation images from different patients. The upper right panel is the zoom-in of the boxed area in each image. The red outlines are from the proposed method, and the green is the ground truth. The red arrow in (c) points to a sagittal plane across the lesion
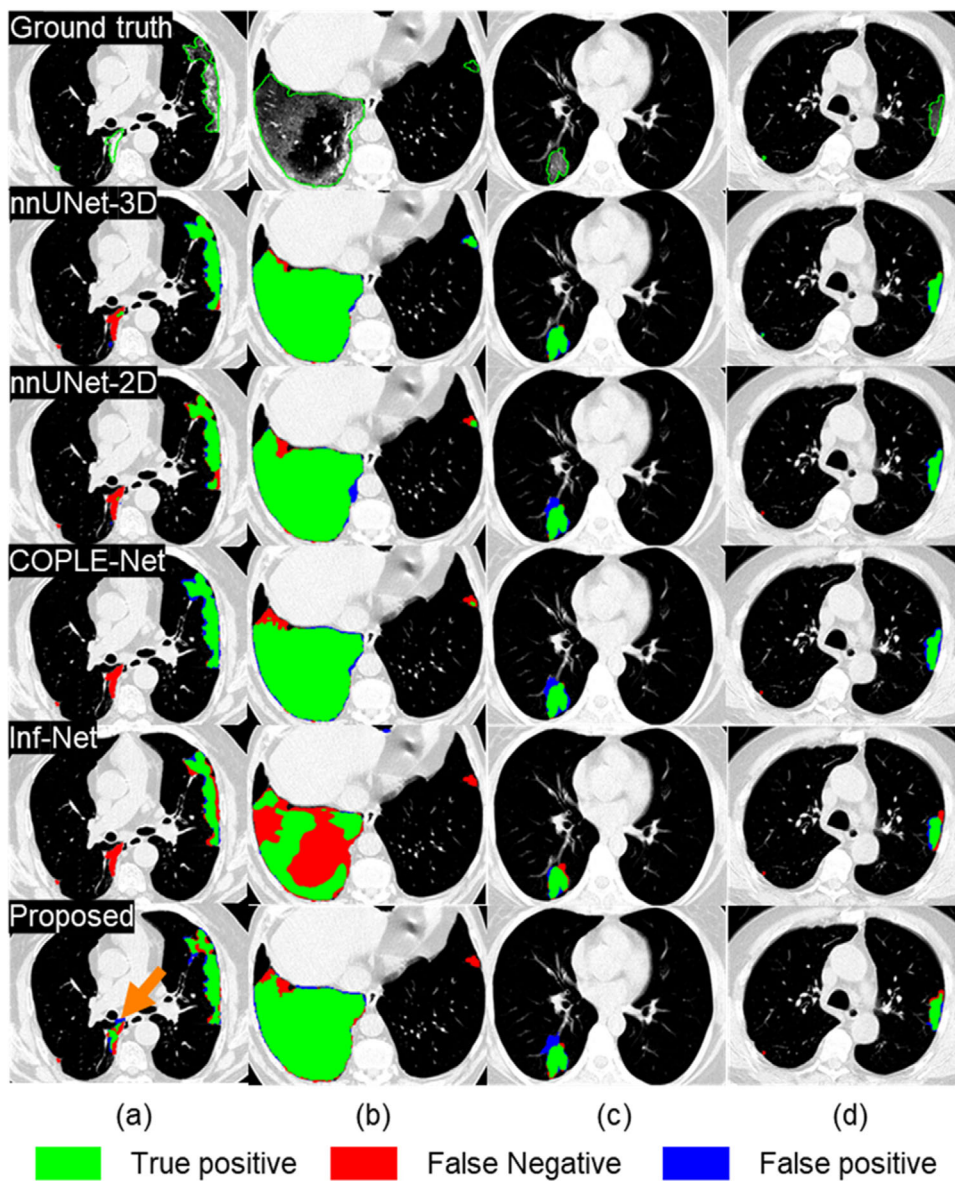
FIGURE 7 Results comparison between different segmentation methods. (a–d) show segmentation results from four different patients. The green, red, and blue areas are true positive, false negative, and false positive, respectively. The orange arrow in (a) points to a lesion that can only be detected by the proposed method

TABLE 2 The comparison between the proposed and existing methods on dice similarity coefficient (DSC), volume precision (vPSC), and volume sensitivity (vSEN)

	Method	Coronacases (10)			Radiopedia (8)		
		DSC (%)	vPSC (%)	vSEN (%)	DSC (%)	vPSC (%)	vSEN (%)
SUPERVISED METHOD	nnUNet-2D	72.2 ± 14.5	86.4 ± 16.1	64.2 ± 16.4	79.2 ± 0.5	75.5 ± 11.2	85.0 ± 6.4
	nnUNet-3D	77.0 ± 10.9	74.2 ± 14.7	82.2 ± 11.5	73.0 ± 9.6	66.5 ± 17.9	86.2 ± 9.2
	Cople-Net	76.7 ± 8.2	79.5 ± 14.9	75.9 ± 7.6	77.8 ± 6.7	83.1 ± 14.9	75.6 ± 10.1
	Inf-Net	59.5 ± 16.8	59.4 ± 19.2	65.2 ± 22.2	64.8 ± 12.0	72.6 ± 17.2	63.0 ± 18.2
UNSUPERVISED METHOD	NormNet	68.7 ± 15.8	85.1 ± 7.0	62.1 ± 22.8	59.4 ± 17.4	60.4 ± 19.7	61.8 ± 18.4
	Proposed (Otsu)	70.9 ± 14.3	68.2 ± 21.1	81.1 ± 9.1	65.3 ± 12.6	65.5 ± 15.6	70.6 ± 16.6
	Proposed (k-means)	74.8 ± 12.1	81.3 ± 8.8	73.5 ± 20.5	73.0 ± 9.5	77.3 ± 17.8	72.6 ± 11.1
		UESTC (50)			Average(68)		
		DSC (%)	vPSC (%)	vSEN (%)	DSC (%)	vPSC (%)	vSEN (%)
SUPERVISED METHOD	nnUNet-2D	74.8 ± 11.4	80.7 ± 15.6	71.5 ± 11.9	74.9 ± 11.4	80.9 ± 15.3	72.1 ± 13.2
	nnUNet-3D	69.5 ± 16.8	86.7 ± 13.1	61.1 ± 18.5	71.0 ± 15.5	82.5 ± 15.6	67.1 ± 19.5
	Cople-Net	84.0 ± 7.1	90.8 ± 5.6	79.1 ± 10.6	82.2 ± 7.8	87.4 ± 9.9	79.1 ± 10.8
	Inf-Net	41.8 ± 18.4	92.6 ± 3.6	36.6 ± 20.6	47.1 ± 19.6	43.1 ± 22.6	60.1 ± 18.2
UNSUPERISED METHOD	NormNet	61.4 ± 19.4	61.3 ± 26.1	77.6 ± 19.6	62.4 ± 18.6	64.7 ± 23.5	73.5 ± 20.0
	Proposed (Otsu)	62.4 ± 16.2	58.4 ± 19.6	75.0 ± 23.1	64.0 ± 14.8	59.0 ± 17.8	77.3 ± 21.6
	Proposed(k-means)	64.4 ± 19.2	74.2 ± 19.9	63.2 ± 24.7	66.6 ± 17.8	67.9 ± 24.4	75.6 ± 16.2

TABLE 3 The paired t-test p-values for dice similarity coefficient (DSC) comparison between the proposed and compared methods

	Coronacases	Radiopedia	UESTC	Average
nnUNet-2D	0.614	≤ 0.010	≤ 0.001	≤ 0.001
nnUNet-3D	0.302	0.618	0.017	0.008
Cople-Net	0.280	0.060	≤ 0.001	≤ 0.001
Inf-Net	0.005	0.110	≤ 0.001	≤ 0.001
NormNet	0.271	0.013	0.376	0.151

TABLE 5 The paired t-test p-values for detection accuracy comparison between the proposed and compared methods

| Accuracy	Coronacases	Radiopedia	UESTC	Total
nnUNet-2D	0.905	0.096	0.641	0.542
nnUNet-3D	0.953	0.247	0.004	0.036
Cople-Net	0.717	0.534	≤ 0.001	≤ 0.001
Inf-Net	0.297	0.083	≤ 0.001	≤ 0.001

4 | DISCUSSION

In this study, we propose a label-free method for delineating COVID-19 lesions using cycle-GAN. This unsupervised method shows great potential in lesion segmentation without employing labeled data and is validated on different databases. The proposed method is robust and less database-dependent and effectively eliminates the labeling inconsistency otherwise existing

TABLE 4 The accuracy (ACC), precision (PSC), and sensitivity (SEN) on three different datasets

Method	Coronacases (10)			Radiopedia (8)		
	ACC (%)	PSC (%)	SEN (%)	ACC (%)	PSC (%)	SEN (%)
nnUNet -2D	94.2 ± 8.0	93.4 ± 11.9	97.1 ± 6.7	88.3 ± 15.3	81.9 ± 26.0	99.0 ± 3.2
nnUNet -3D	83.3 ± 20.0	76.7 ± 26.2	98.0 ± 6.3	87.5 ± 20.5	81.7 ± 29.1	100.0 ± 0.0
Cople-Net	86.7 ± 15.8	79.5 ± 27.8	98.6 ± 4.5	91.7 ± 12.4	85.0 ± 25.2	95.8 ± 10.1
Inf-Net	70.0 ± 32.7	69.7 ± 32.7	100.0 ± 0.0	75.8 ± 32.0	74.6 ± 33.9	100.0 ± 0.0
Proposed (k-means)	90.0 ± 10.2	87.5 ± 20.4	96.4 ± 6.3	86.7 ± 12.6	95.0 ± 15.8	81.6 ± 18.8
	UESTC (50)			Average (68)		
	ACC (%)	PSC (%)	SEN (%)	ACC (%)	PSC (%)	SEN (%)
nnUNet -2D	76.5 ± 17.7	68.5 ± 24.1	99.3 ± 3.0	80.7 ± 14.5	74.0 ± 24.4	99.3 ± 3.0
nnUNet -3D	85.3 ± 12.7	78.8 ± 20.2	98.2 ± 5.2	85.4 ± 14.8	78.9 ± 22.1	98.2 ± 5.2
Cople-Net	90.8 ± 9.9	85.8 ± 17.8	97.5 ± 7.7	90.4 ± 11.1	84.8 ± 20.2	97.5 ± 7.7
Inf-Net	54.7 ± 24.5	54.6 ± 24.8	99.7 ± 1.8	59.9 ± 27.6	59.6 ± 28.0	99.7 ± 1.8
Proposed (k-means)	74.8 ± 22.7	73.8 ± 28.6	92.4 ± 12.9	79.1 ± 20.8	77.7 ± 26.8	92.4 ± 12.9
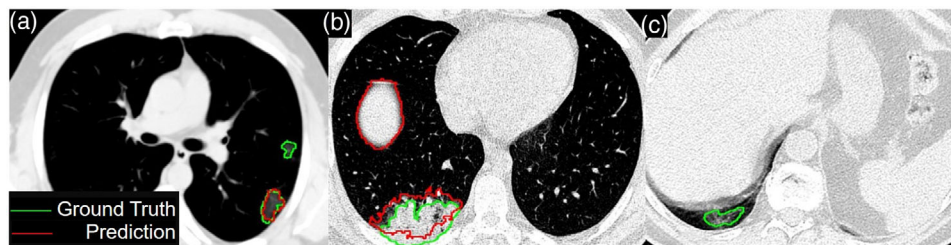
FIGURE 8 Segmentation limitations for the proposed method. The green contours are manual labels representing the ground truth and the red ones are automatic segmentation results by the proposed method

in supervised learning approaches. It can work as an efficient and independent automatic segmentation method or can provide a start point for physicians' follow-up refinement.

There are still some limitations to the proposed unsupervised method. First, as shown in Figure 8a, the method may miss some small and low contrast lesions. Second, the lung CT slices delineated by a U-net were used as the input, which may bear some false positive lesions misclassified from the top part of the liver. On three test sets, the top liver from three patients of a total of 68 (4.34%) was misclassified as a lesion. One example is shown in Figure 8b. Moreover, the sample number in this study might not be sufficiently large, particularly lacking patients with lesions in the top and bottom of the lung volume, potentially resulting in mis-delineation as shown in Figure 8c.

On the other hand, the network was trained with 2D images due to hardware limitations. This strategy didn't make full use of the three-dimensional property of CT images. Future studies using 3D image input may improve segmentation accuracy, particularly in terms of the contour continuity along the image thickness direction. Despite this, the proposed unsupervised method still achieved decent segmentation results, with a dice value of 74.8% and an accuracy of 90.0% on the Coronacases public database. In the future, the unsupervised method can be combined with more sophisticated postprocessing methods, such as texture analysis or even additional deep learning networks, to further improve segmentation results.

5 | CONCLUSIONS

In this work, we propose a label-free approach that can accurately and efficiently delineate the COVID-19 lesions automatically in CT scans. The training process of the unsupervised network does not rely on any labeled data. The automatic segmentation can provide a start point for further manual refinement and can work as a quality assurance tool for lesion diagnosis. Due to its unsupervised nature, the result is not influenced by physicians' experiences which otherwise is crucial for supervised methods.

CONFLICTS OF INTEREST
The authors declare no conflict of interest.

REFERENCES

1. WHO. Coronavirus (COVID-19) dashboard with vaccination data. https://covid19.who.int/. Accessed April 6, 2021.
2. Sheng WH. Coronavirus disease 2019 (COVID-19). *J Intern Med Taiwan*. 2020;31(2):61-66. https://doi.org/10.6314/JIMT.202004_31(2).01.
3. Liu KC, Xu P, Lv WF, et al. CT manifestations of coronavirus disease-2019: a retrospective analysis of 73 cases by disease severity. *Eur J Radiol*. 2020;126:108941. https://doi.org/10.1016/j.ejrad.2020.108941.
4. Yang R, Li X, Liu H, et al. Chest CT severity score: an imaging tool for assessing severe COVID-19. *Radiol Cardiothorac Imaging*. 2020;2(2):e200047. https://doi.org/10.1148/ryct.2020200047.
5. Abbasi-Oshaghi E, Mirzaei F, Farahani F, Khodadadi I, Tayebinia H. Diagnosis and treatment of coronavirus disease 2019 (COVID-19): laboratory, PCR, and chest CT imaging findings. *Int J Surg*. 2020;79:143-153. https://doi.org/10.1016/j.ijsu.2020.05.018.
6. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: learning dense volumetric segmentation from sparse annotation. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag; 2016. p. 424-432. https://doi.org/10.1007/978-3-319-46723-8_49.
7. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer Verlag; 2015. p. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28.

8. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016. http://image-net.org/challenges/LSVRC/2015/ Accessed April 6, 2021.

9. Lei Y, Harms J, Wang T, et al. MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks. *Med Phys*. 2019;46(8):3565-3581. https://doi.org/10.1002/mp.13617.

10. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. *Commun ACM*. 2020;63(11):139-144. https://doi.org/10.1145/3422622.

11. Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med Phys*. 2017;44(4):1408-1419. https://doi.org/10.1002/mp.12155.

12. Cai W, Yang J, Fan G, Xu L, Zhang B, Liu R. Chest CT findings of coronavirus disease 2019 (COVID-19). *J Coll Physicians Surg Pakistan*. 2020;30(1):S53-S55. 10.29271/jcpsp.2020.Supp1.S53.

13. Mahmud T, Rahman MA, Fattah SA. CovXNet: a multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. *Comput Biol Med*. 2020;122(June):103869. https://doi.org/10.1016/j.compbiomed.2020.103869.

14. Volpicelli G, Gargani L. Sonographic signs and patterns of COVID-19 pneumonia. *Ultrasound J*. 2020;12(1):22. https://doi.org/10.1186/s13089-020-00171-w.

15. Soldati G, Smargiassi A, Inchingolo R, et al. Is there a role for lung ultrasound during the COVID-19 pandemic? *J Ultrasound Med*. 2020;39(7):1459-1462. https://doi.org/10.1002/jum.15284.

16. Bernheim A, Mei X, Huang M, et al. Chest CT findings in coronavirus disease 2019 (COVID-19): relationship to duration of infection. *Radiology*. 2020;295(3):685-691. https://doi.org/10.1148/radiol.2020200463.

17. Tang Z, Zhao W, Xie X, et al. Severity assessment of COVID-19 using CT image features and laboratory indices. *Phys Med Biol*. 2021;66(3):35015. https://doi.org/10.1088/1361-6560/abbf9e.

18. Wang G, Liu X, Li C, et al. A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images. *IEEE Trans Med Imaging*. 2020;39(8):2653-2663. https://doi.org/10.1109/TMI.2020.3000314.

19. Shi F, Wang J, Shi J, et al. Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. *IEEE Rev Biomed Eng*. 2021;14:4-15. https://doi.org/10.1109/RBME.2020.2987975.

20. Hasanzadeh N, Paima SS, Bashirgonbadi A, Naghibi M, Soltanian-Zadeh H. Segmentation of COVID-19 infections on CT: comparison of four UNet-Based networks. In: *27th National and 5th International Iranian Conference of Biomedical Engineering, ICBME 2020*. Institute of Electrical and Electronics Engineers Inc.; 2020. p. 222-225. https://doi.org/10.1109/ICBME51989.2020.9319412

21. Yao Q, Xiao L, Liu P, Kevin Zhou S. Label-free segmentation of COVID-19 lesions in lung CT. *IEEE Trans Med Imaging*. 2021;40(10):2808-2819. https://doi.org/10.1109/TMI.2021.3066161.

22. Zhang P, Zhong Y, Deng Y, Tang X, Li X. CoSinGAN: learning COVID-19 infection segmentation from a single radiological image. *Diagnostics*. 2020;10(11):901. https://doi.org/10.3390/diagnostics10110901.

23. Harms J, Lei Y, Wang T, et al. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. *Med Phys*. 2019;46(9):3998-4009. https://doi.org/10.1002/mp.13656.

24. Liu Y, Lei Y, Wang T. Liver synthetic CT generation based on a dense-CycleGAN for MRI-only treatment planning. In: Landman BA, Išgum I, eds. *Medical Imaging 2020: Image Processing*. SPIE; 2020: 92. https://doi.org/10.1117/12.2549265. Vol.

25. Liu Y, Lei Y, Wang T, et al. CBCT-based synthetic CT generation using deep-attention cycleGAN for pancreatic adaptive radiotherapy. *Med Phys*. 2020; 47(6): 2472-2483. https://doi.org/10.1002/mp.14121.

26. Liu M, Liu Y, Yang Y, et al. Automated assessment of disease severity of COVID-19 using artificial intelligence with synthetic chest CT[J]. arXiv preprint arXiv:2112.05900, 2021.

27. Zhu J-Y, Park T, Isola P, Efros AA, Research BA. Unpaired image-to-image translation using cycle-consistent adversarial networks monet photos. 2017. https://github.com/junyanz/CycleGAN. Accessed April 6, 2021.

28. Mao X, Li Q, Xie H, Lau RYK, Wang Z, Smolley SP, Least Squares Generative Adversarial Networks. In: *Proceedings of the IEEE International Conference on Computer Vision*. Vol Oct 2017; 2017:2813-2821. https://doi.org/10.1109/ICCV.2017.304

29. Goodfellow IJ, Pouget-abadie J, Mirza M, Xu B, Warde-farley D. Generative adversarial nets. *Adv Neural Inf Process Syst*. 2014:2672-2680.

30. LUNA (LUng Nodule Analysis)16 - ISBI 2016 challenge - academic torrents. https://academictorrents.com/collection/luna-lung-nodule-analysis-16-isbi-2016-challenge Accessed December 27, 2021.

31. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203-211. https://doi.org/10.1038/s41592-020-01008-z.

32. Fan DP, Zhou T, Ji GP, et al. Inf-Net: automatic COVID-19 lung infection segmentation from CT images. *IEEE Trans Med Imaging*. 2020;39(8):2626-2637. https://doi.org/10.1109/TMI.2020.2996645.