COVID-19 CT Image Synthesis with a Conditional Generative Adversarial Network

Yifan Jiang, *Member, IEEE,* Han Chen, Murray Loew, *Fellow, IEEE* and Hanseok Ko, *Senior Member, IEEE*

***Abstract*—Coronavirus disease 2019 (COVID-19) is an ongoing global pandemic that has spread rapidly since December 2019. Real-time reverse transcription polymerase chain reaction (rRT-PCR) and chest computed tomography (CT) imaging both play an important role in COVID-19 diagnosis. Chest CT imaging offers the benefits of quick reporting, a low cost, and high sensitivity for the detection of pulmonary infection. Recently, deep-learning-based computer vision methods have demonstrated great promise for use in medical imaging applications, including X-rays, magnetic resonance imaging, and CT imaging. However, training a deep-learning model requires large volumes of data, and medical staff faces a high risk when collecting COVID-19 CT data due to the high infectivity of the disease. Another issue is the lack of experts available for data labeling. In order to meet the data requirements for COVID-19 CT imaging, we propose a CT image synthesis approach based on a conditional generative adversarial network that can effectively generate high-quality and realistic COVID-19 CT images for use in deep-learning-based medical imaging tasks. Experimental results show that the proposed method outperforms other state-of-the-art image synthesis methods with the generated COVID-19 CT images and indicates promising for various machine learning applications including semantic segmentation and classification.**

***Index Terms*—COVID-19, computed topography, image synthesis, conditional generative adversarial network**

I. Introduction

CORONAVIRUS disease 2019 (COVID-19) [1], which was first identified in Wuhan, China, in December 2019, was declared a pandemic in March 2020 by the World Health Organization (WHO). As of 21 July, there had been more than 14 million confirmed cases and 609,198 deaths across 188 countries and territories [2]. COVID-19 is the result of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and its most common symptoms include fever, dry cough, a loss of appetite, and fatigue, with common complications including pneumonia, liver injury, and septic shock [3], [4].

There are two main diagnostic approaches for COVID-19: rRT-PCR and chest computed tomography (CT) imaging [4]. In rRT-PCR, an RNA template is first converted by reverse transcriptase into complementary DNA (cDNA), which is then used as a template for exponential amplification using polymerase chain reaction (PCR). However, the sensitivity of rRT-PCR is relative low for COVID-19 testing [5], [6]. As an alternative, chest CT scans can be used to take tomographic images from the chest area at different angles with post-computed X-ray measurements. This approach has a higher sensitivity to COVID-19 and is less resource-intensive than traditional rRT-PCR [5], [6].

Over time, artificial intelligence (AI) has come to play an important role in medical imaging tasks, including CT imaging [7], [8], magnetic resonance imaging (MRI) [9] and X-ray imaging [10]. Deep learning is a particularly powerful AI approach that has been successfully employed in a wide range of medical imaging tasks due to the massive volumes of data that are now available. These large datasets allow deep-learning networks to be well-trained, extending their generalizability for use in various applications. However, the collection of COVID-19 data for use in deep-learning models is far more difficult than normal data collection. Because COVID-19 is highly contagious [4], medical staff require full-length protection for CT scans, and the CT scanner and other equipment need to be carefully disinfected after an operation. In addition, certain tasks, such as CT image segmentation, require well-labeled data, which is labor-intensive. These problems mean that the COVID-19 CT data collection process can be difficult and time-consuming.

In order to speed up the COVID-19 CT data collection process for deep-learning-based CT imaging and to protect medical personnel from possible infection when coming into contact with COVID-19 patients, we propose a novel image synthesis method based on a conditional generative adversarial network (cGAN) for deep-learning chest CT imaging. The fundamental principle of the proposed method is to employ CT images and corresponding well-labeled semantic segmentation maps to train a cGAN model. Figure 1 presents example chest CT images from several COVID-19 patients. The proposed model takes lung segmentation maps as input and employs them to generate synthesized CT images during the training stage. In the testing stage, we prepare a data augmented segmentation map and then use the pre-trained model to generate realistic synthesized lung CT images. The main contributions of the proposed method are as follows:

(1) A safe COVID-19 chest CT data collection method based on image synthesis is presented. It is designed to significantly reduce the infection risk and the workload of medical staff. To the best of our knowledge, the proposed method represents the first use of image synthesis technology for COVID-19 chest CT imaging.
(2) The proposed approach can generate the COVID-19 chest CT images with ground-glass opacity and consolidation in a single model.

This research work is supported by a National Research Foundation (NRF) grant funded by the MSIP of Korea (number 2019R1A2C2009480). (Corresponding author: Hanseok Ko.) Y. Jiang, H. Chen, and H. Ko are with the School of Electrical Engineering, Korea University, Seoul 02841, South Korea (e-mail: yfjiang@ispl.korea.ac.kr;hanchen@ispl.korea.ac.kr; hsko@korea.ac.kr). M. Loew is with School of Engineering & Applied Science, George Washington University, Washington, DC 20052, United States (e-mail: loew@gwu.edu).

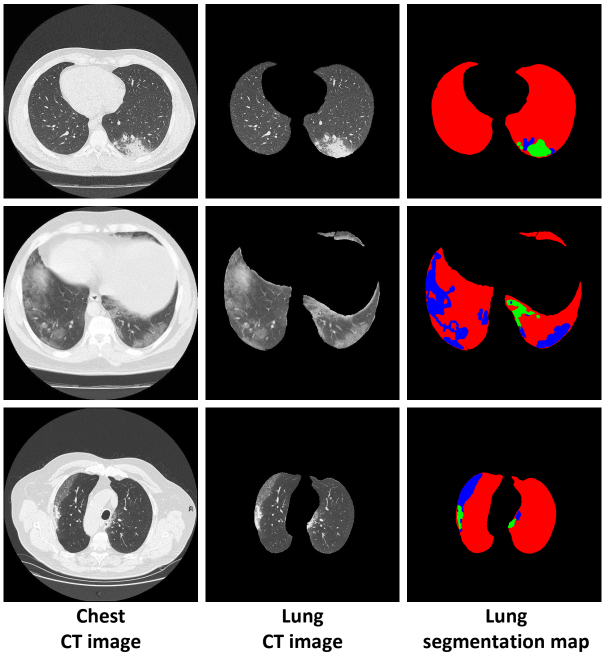

Fig. 1. Example CT images from three COVID-19 patients. The first column shows CT images of the entire chest, the second column contains CT images of the lungs only, and the third column shows the corresponding segmentation map, with the lung region colored red, ground-glass opacity colored blue, and areas of consolidation colored green.

(3) The proposed method outperforms other state-of-the-art image synthesizers in several image-quality metrics and demonstrates its potential for use in image synthesis for computer vision tasks such as semantic segmentation for COVID-19 chest CT imaging.

(4) Benefiting from the novel restoration performance and flexible segmentation mapping, the proposed approach holds significant promise for use in practical applications, such as the rapid diagnosis of COVID-19 or other diseases, using a simple process that starts with a data augmented segmentation map of the target image in order to reconstruct medical simulation images.

II. Related works

Generative adversarial networks. Generative adversarial networks (GANs) were first reported in 2014 [11], and they have since been widely applied to many practical applications, including image synthesis [12], [13], [14], [15], image enhancement [16], [17], human pose estimation [18], [19], and video generation [20], [21]. A GAN structure generally consists of a generator and a discriminator, where the goal of the generator is to fool the discriminator by generating a synthetic sample that cannot be distinguished from real samples. A common GAN extension is the conditional generative adversarial network (cGAN) [22], which generates images that are conditional on class labels. cGAN always produces more realistic results than traditional GANs due to the extra information from these conditional labels.

Conditional image-to-image translation. Conditional image-to-image translation (synthesis) methods can be mainly divided into three categories by the input conditions. Class-conditional methods take class-wise labels as input to synthesize image [22], [23], [24], [25]. More recently, some text-conditional methods have been introduced [26], [27]. The conditional GANs based methods [12], [13], [14], [15], [28], [29], [30], [31], [27], [26], [32] have been widely used on various image-to-image translation methods, for instance unsupervised image-to-image translation [30], high quality image-to-image translation [13], multi-modal image-to-image translation [28], [14], [15], semantic layout conditional image-to-image translation. [12], [13], [14], [15]. In the case of semantic layout conditional methods, the main idea is to synthesize realistic images under the navigation of semantic layout, so that they are easier to control particular part in the image.

Conditional image-to-image translation. Conditional image-to-image translation methods can be divided into three categories based on the input conditions. Class-conditional methods take class-wise labels as input to synthesize images [22], [23], [24], [25] while, more recently, text-conditional methods have been introduced [26], [27]. cGAN-based methods [12], [13], [14], [15], [28], [29], [30], [31], [27], [26], [32] have been widely used for various image-to-image translation methods, including unsupervised [30], high-quality [13], multi-modal [28], [14], [15], and semantic layout conditional image-to-image translation [12], [13], [14], [15]. In semantic layout conditional methods, realistic images are synthesized under the navigation of the semantic layout, meaning that it is easier to control a particular region of the image.

AI-based diagnosis using COVID-19 CT imaging. Since the outbreak of COVID-19, many researchers have turned to CT imaging technology in order to diagnose and investigate this disease. COVID-19 diagnosis methods based on chest CT imaging have been introduced in order to improve test efficiency [33], [34], [35], [36]. Rather than using CT imaging for rapid COVID-19 diagnosis, semantic segmentation approaches have been employed to clearly label the focus position in order to make it easier for medical personnel to identify infected regions in a CT image [37], [38], [39], [40], [41]. As an alternative to working at the pixel-level, high-level classification or detection approaches have been proposed [42], [43], [44], which can allow medical imaging experts to rapidly locate areas of infection, thus speeding up the diagnosis process. Though two CT image synthesis methods have been previously reported [45], [46], they did not focus on COVID-19 or lung CT imaging; to the best of our knowledge, our proposed method is the first designed specifically for COVID-19 CT image synthesis.

III. COVID-19 CT Image Synthesis with a Conditional Generative Adversarial Network

In this paper, we propose a cGAN-based COVID-19 CT image synthesis method. Here, COVID-19 CT image synthesis is formulated as a semantic-layout-conditional image-to-image translation task. The proposed method is inspired by Pix2pixHD [13], with the structure consisting of two main components: a global-local generator and a multi-resolution discriminator. During the training stage, the semantic segmentation map of a corresponding CT image is passed to

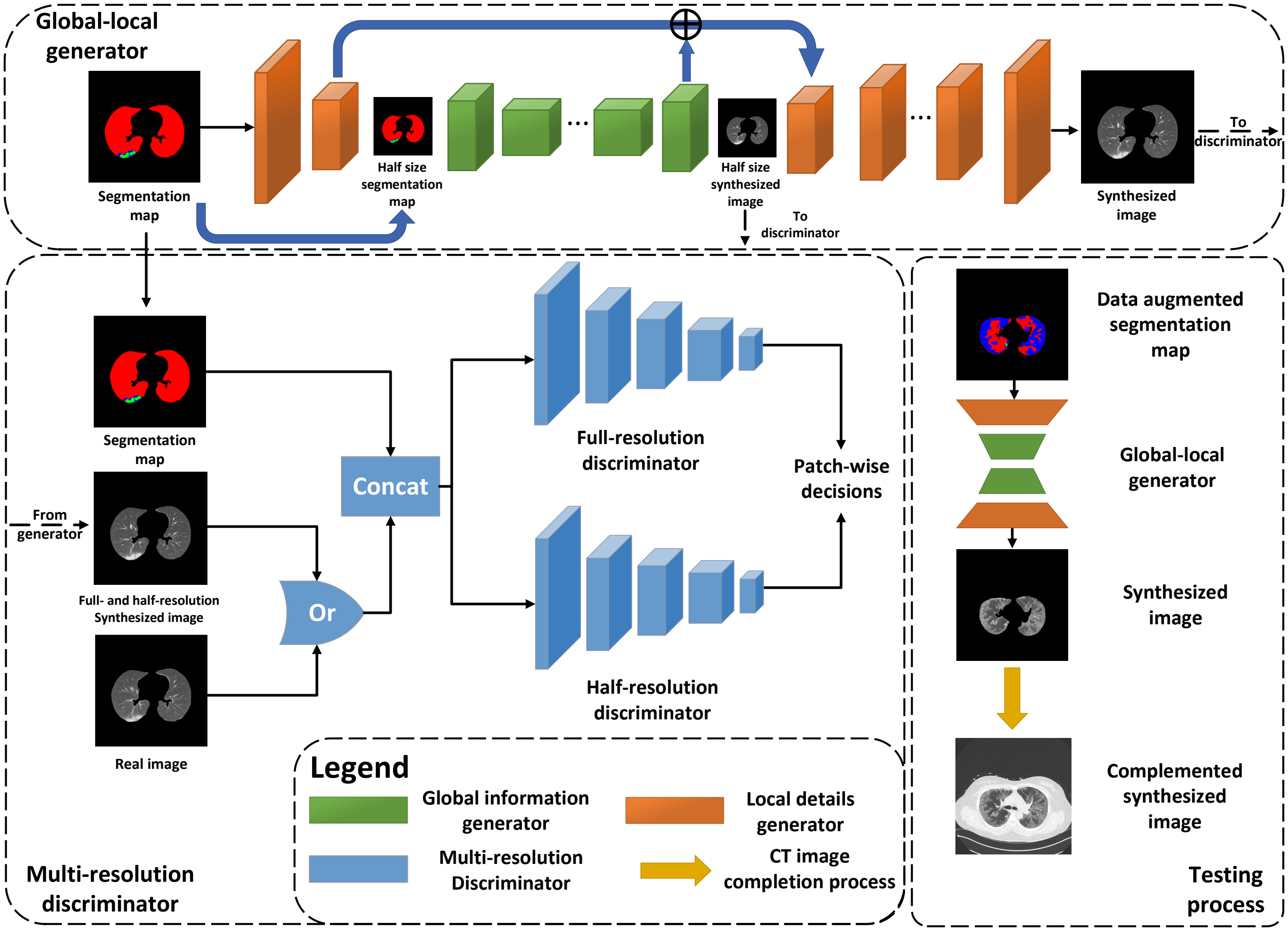
Global-local generator
Segmentation map
Half size segmentation map
Half size synthesized image
To discriminator
Synthesized image
To discriminator
Segmentation map
From generator
Full- and half-resolution Synthesized image
Real image
Concat
Or
Full-resolution discriminator
Half-resolution discriminator
Patch-wise decisions
Multi-resolution discriminator
Legend
Global information generator
Local details generator
Multi-resolution Discriminator
CT image completion process
Data augmented segmentation map
Global-local generator
Synthesized image
Complemented synthesized image
Testing process

Fig. 2. Overview of the proposed method. The upper section containing global-local generator blocks and multi-resolution discriminator blocks represents the training process, while the lower right section shows the testing process. Within the global-local generator blocks, two types of generator are present: a global information generator and local detail generator. The multi-resolution discriminator is depicted in gray. The synthesized images are transferred from the generator to the discriminator, and this process is shown as the dashed arrow. The yellow arrow shows the completion step for the process in which the non-lung region for the synthesized lung image is added.

the global-local generator, where the label information from the segmentation map is extracted via down-sampling and re-rendered to generate a synthesized image via up-sampling. The segmentation map is then concatenated with the corresponding CT image or synthesized CT image to form the input for the multi-resolution discriminator, which is used to distinguish the input as either real or synthesized. The decisions from the discriminator are used to calculate the loss and update the parameters for both the generator and discriminator. During the testing stage, only the generator is involved. A data augmented segmentation map is used as input for the generator, from which a realistic synthesized image can be obtained after extraction and re-rendering. This synthesized lung CT image is then combined with the non-lung area to form a completely synthesized CT image as the final result. Figure 2 presents an overview of the proposed method.

A. Global-local generator

The global-local generator G here mainly consists of two parts: global information generator G_1 and local details generator G_2. These two generators work together in a coarse-to-fine way, where G_1 takes charge of learning and re-rendering global information, the global information always contains high-level knowledge (semantic segmentation labels, image structure information). While G_2 does details enhancement task in order to generate detailed information (image texture, tiny structures).

The global-local generator G has two sub-components: global-information generator G_1 and local-detail generator G_2. These generators work together by moving in a coarse-to-fine direction. G_1 takes charge of learning and re-rendering global information, which always contains high-level knowledge (e.g., semantic segmentation labels and image structure information). G_2 is then used for detail enhancement (e.g., image texture and fine structures).

We train the global-local generator using a three-step process:

1) Individual training for the global information generator: The training process for G starts with the training of the global information generator G_1. The design of G_1 follows

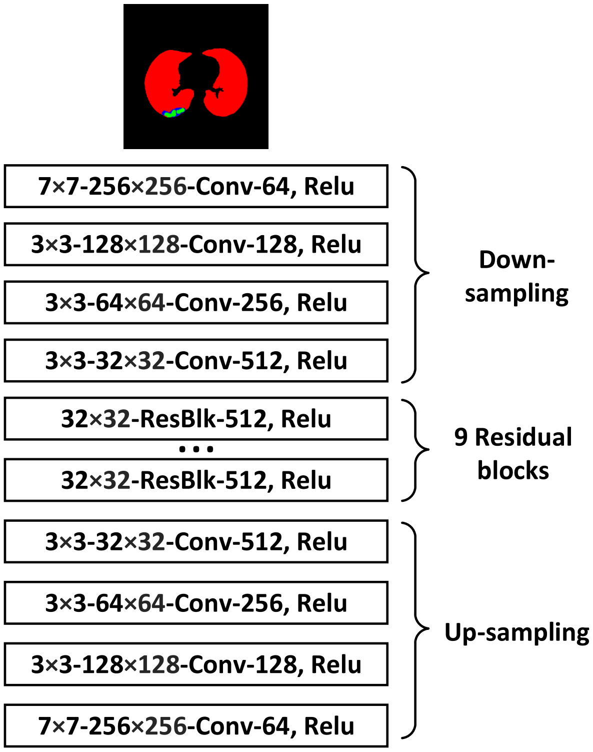

Fig. 3. The network structure of the global information generator G_1. The parameters of each layer are separated by the notation '-', e.g. for the first layer, 7×7 means the kernel size is 7, 256×256 is the size of the feature, Conv denotes the category of the layer, 64 is the channel number, and Relu is the activation function.

PatchGAN [47]. As shown in Figure 3, G_1 takes a half-resolution (256×256) segmentation map as input, which is then sent for down-sampling to reduce the feature dimensions to 32×32. Nine residual blocks that maintain the dimensions at 32×32 are used to reduce the computational complexity and generate a large reception field [47]. Finally, the features are up-sampled and reconstructed back into a half-resolution (256×256) synthesized image.

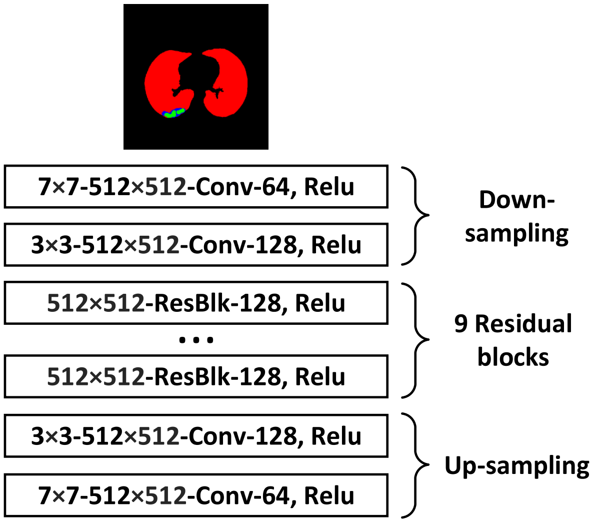
7×7-512×512-Conv-64, Relu
3×3-512×512-Conv-128, Relu
Down-sampling
512×512-ResBlk-128, Relu
• • •
512×512-ResBlk-128, Relu
9 Residual blocks
3×3-512×512-Conv-128, Relu
7×7-512×512-Conv-64, Relu
Up-sampling

Fig. 4. Network structure of local details generator G_2.

2) *Individual training for the local details generator:* The structure of the local detail generator G_2, which is similar to the structure of G_1, is shown in Figure 4. Rather than taking a low-resolution segmentation map as input, the local detail generator begins the synthesis process with a full-resolution segmentation map (512×512) and maintains this size throughout. That allows the local detail generator to fully learn the fine texture and structure and focus on low-level information within the input image. G_2 has a similar encoding-decoding training procedure as G1, though the output synthesized image is 512×512.

3) *Joint training for the global-local generator:* After training G_1 and G_2 separately, a joint training process is conducted. This is shown in the global-local generator region of Figure 2. In the joint training stage, both G_1 and G_2 take the same input but with different resolutions (half- and full-resolution, respectively). The two networks run a forward process that differs from the individual training stage in which the up-sampling process in G_2 takes the element-wise sum from the output feature maps from the up-sampling process in G_1 and the output feature maps from down-sampling in G_2, meaning that G_2 receives both global and local information to reconstruct the output.

This training strategy enables the global-local generator G to effectively learn both global information and local details while also stabilizing the training process by simplifying it into three relatively simple procedures.

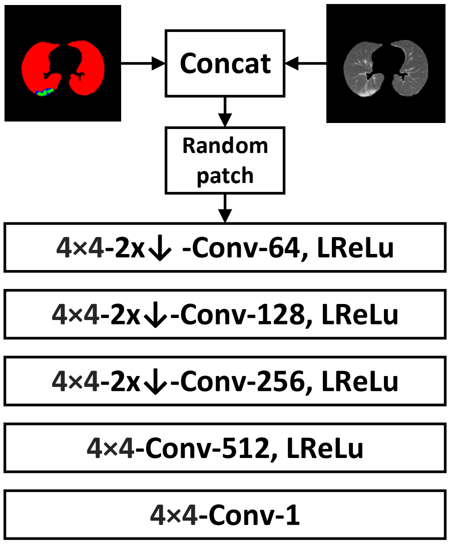

Fig. 5. Network structure of the multi-resolution discriminator D. $2\times \downarrow$ denotes down-sampling with a factor of 2.

B. Multi-resolution discriminator

A multi-resolution discriminator D is proposed in this paper. D consists of two sub-components: the full-resolution discriminator D_1 and the half-resolution discriminator D_2. The design of these discriminators follows SPADE [14]; rather than making a decision for the whole image, we utilize a patch-wise discriminator. This means that the proposed discriminator can perceive both the global information and the details of the image. As shown in Figure 5, we first down-sample the segmentation map, synthesized image, and real image into half-resolution form, then the synthesized and real images are randomly chosen to be concatenated with the segmentation map to form two inputs (full- and half-resolution) for D. Before sending the input to the discriminator, it should be randomly sampled as a series of 70×70 patches. After passing these patches through the discriminator, two decision matrices are obtained, which represent the patch-wise decisions for the two inputs.

The patch-wise sampling method enables the multi-resolution discriminator D to effectively learn local details, which can significantly improve the quality of the synthesized image. By assigning global and local discrimination to individual discriminators D1 and D2, the global structure can be maintained while also enhancing the details of the synthesized images.

C. Learning objective

The overall learning objective of proposed approach can be represented by equation (1):

$$\min_{G} \left[\max_{D_1,D_2} \sum_{i=1}^{2} \mathcal{L}_{cGAN}\{G(m), D_i(m,x), D_i(m,G(m))\} + \lambda \sum_{i=1}^{2} \mathcal{L}_{FM}\{G(m), D_i(m,x), D_i(m,G(m))\} \right] \tag{1}$$

There are two main loss terms in the overall learning objective function (1): the loss for the cGAN $\mathcal{L}_{cGAN}$ and the loss for feature matching $\mathcal{L}_{FM}$. The variable x is the real input image and m is the corresponding segmentation map. G represents global-local generator while D_i represents the full-resolution discriminator D_1 or half-resolution discriminator D_2. $G(m)$ denotes the synthesized image produced by generator G with input segmentation map s, $D_i(m,x)$ and $D_i(m,G(m))$ are the patch-wise decisions made by multi-resolution discriminator D with the real image or synthesized image as input, respectively. λ is the weight factor of feature matching loss term.

We designed the cGAN loss function based on pix2pix [12], as shown in (2)

$$\mathcal{L}_{cGAN} = \mathbb{E}_{m,x}[logD(m,x)] + \mathbb{E}_{m,x}[log(1 - D(m,G(m)))] \tag{2}$$

This loss term allows cGAN to generate a realistic synthesized image that can fool discriminator under the condition of the input segmentation map.

In order to train multi-resolution discriminator D, feature matching loss is employed (Eq. (3)), which is inspired by ref [14].

$$\mathcal{L}_{FM} = \mathbb{E}_{m,x} \sum_{i=1}^{3} \frac{1}{N_i} [||D^i(m,x) - D^i(m,G(m))||_1] \tag{3}$$

where i represents the i^{th} layer of D and N_i is the total number of elements in the i^{th} layer. This loss manages the differences of intermediate features from both full- and half-resolution discriminators in order to stabilize the training process and allow D_1 and D_2 to synchronously learn the details from the inputs with different resolutions.

D. Testing process

Rather than using both global-local generator G and multi-resolution discriminator D as in the training stage, we only utilize the pre-trained G in the testing process. The input for G in this stage is a data augmented segmentation map that can be obtained using standard image editing software. After passing it through G, a synthesized CT image of the lung area is generated. The final step in the process combines the synthesized lung image with the corresponding non-lung area from the real image to produce a complete synthesized image.

IV. Experiments

A. Experimental settings

Dataset. In order to evaluate the proposed method and compare its performance to other state-of-the-art methods, we use 829 lung CT slices from nine COVID-19 patients, which were made public on 13 April 2020, by Radiopaedia [48]. This dataset includes the original CT images, lung masks, and COVID-19 infection masks. The infection masks contain ground-glass opacity and consolidation labels, which are the two most common characteristics used for COVID-19 diagnosis in lung CT imaging [49]. In this experiment, we select 373 slices that contained clear areas of infection. We divide the selected dataset into training and test sets consisting of 300 and 73 images, respectively. To fully train the deep-learning-based model, data augmentation pre-processing is applied. The 300 original images from the training set are augmented to produce 12,000 images, while the 73 images from the test set are augmented to produce 10,220 images (Table I).

TABLE I
ORGANIZATION OF THE COVID-19 CT IMAGE DATASET

Dataset	Original count	After data augmentation
Training set	300	12,000
Test set	73	10,220
Overall	373	22,220
Before selection	829	-

The data augmentation methods include random resizing and cropping, random rotation, Gaussian noise, and elastic transform.

Evaluation metrics. To accurately assess model performance, we utilize both image quality metrics and medical imaging semantic segmentation metrics:

Four image quality metrics are considered in this study: Frchet inception distance (FID) [50], peak-signal-to-noise ratio (PSNR) [51], structural similarity index measure (SSIM) [51], and root mean square error (RMSE) [15]. FID measures the similarity of the distributions of real and synthesized images using a deep-learning model. PSNR and SSIM are the most widely used metrics when evaluating the performance of image restoration and reconstruction methods. The former represents

TABLE II
IMAGE QUALITY EVALUATION RESULTS OF SYNTHETIC CT IMAGES
(THE BEST EVALUATION SCORE IS MARKED IN BOLD. ↑ MEANS HIGHER NUMBER IS BETTER, AND ↓ INDICATES LOWER NUMBER IS BETTER.)

Categories	Complemented images				Lung only images			
Metrics	FID (↓)	PSNR (↑)	SSIM (↑)	RMSE (↓)	FID (↓)	PSNR (↑)	SSIM (↑)	RMSE (↓)
OURS	**0.0327**	**26.89**	**0.8936**	**0.0813**	0.3641	**28.17**	**0.8959**	**0.2747**
SEAN [15]	0.0341	26.69	0.8922	0.0837	**0.3575**	28.02	0.8952	0.2795
SPADE [14]	0.0389	26.50	0.8903	0.0854	0.4812	27.79	0.8928	0.2864
Pix2pixHD [13]	0.0430	26.63	0.8893	0.0840	0.4283	27.82	0.8910	0.2856
Pix2pix [12]	0.0611	26.56	0.8870	0.0913	8.4077	26.56	0.8855	0.3301

the ratio between the maximum possible intensity of a signal and the intensity of corrupting noise, while the latter reflects the structural similarity between two images.

Three semantic segmentation metrics for medical imaging are used in this experiment: the dice score (Dice), sensitivity (Sen), and specificity (Spec) [52], [53]. The dice score evaluates the area of overlap between a prediction and the ground truth, while sensitivity and specificity are two statistical metrics for the performance of binary medical image segmentation tasks. The former measures the percentage of actual positive pixels that are correctly predicted to be positive, while the latter measures the proportion of actual negative pixels that are correctly predicted to be negative. These three metrics are employed for semantic segmentation based on the assumption that, if the quality of the synthesized images is high enough, excellent segmentation performance can be achieved when using the synthesized images as input.

Implementation details. We transform all of the CT slices into gray-scale images on a Hounsfield unit (HU) scale [-600,1500]. The sizes of the images and segmentation maps are then rescaled from 630×630 to 512×512. All of the image synthesis methods are trained with 20 epochs, with a learning rate that is maintained at 0.0002 for the first 10 epochs before linearly decaying to zero over the following ten epochs. Global-local generator G and multi-resolution discriminator D are trained using an Adam optimizer with parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The feature matching loss weight λ is set at 10. The batch size used to train the proposed method is 16. All of the experiments are run in an Ubuntu 18.04 environment using an Intel i7 9700k CPU and two GeForce RTX Titan graphics cards (48 GB VRAM).

B. *Quantitative results*

The performance of the proposed method was assessed according to both image quality and medical imaging semantic segmentation.

1) Image quality evaluation: In this study, common image quality metrics are employed to assess the synthesis performance of the proposed method and four other state-of-the-art image synthesis methods: SEAN [15], SPADE [14], Pix2pixHD [13], and Pix2pix [12]. We evaluate image quality for two synthetic image categories: complete and lung-only images. The complete images are those CT images generated by merging a synthesized lung CT image with its corresponding non-lung CT image. The evaluation results are presented in Table II.

The proposed method outperforms other state-of-the-art methods based on the four image quality metrics for both the complete and lung-only images. Due to the design of the global-local generator and multi-resolution discriminator, the proposed model can generate realistic lung CT images for COVID-19 with a complete global structure and fine local details and maintain a relatively high signal-to-noise ratio. Thus, the proposed method can achieve state-of-the-art image synthesis results based on image quality.

2) Medical imaging semantic segmentation evaluation: To evaluate the reconstruction capability of the proposed method, we utilize Unet, a common medical imaging semantic segmentation approach [54]. We first train the Unet model on a mix of synthetic and real CT images.

This evaluation consists of two independent experiments: (1) keeping the total number of images the same while replacing the real data with synthesized data from a proportion of 0% to 50% in steps of 10% and (2) keeping the number of real images the same and adding a certain proportion of synthetic images from 0% to 50% in steps of 10%. The first experiment evaluates how similar the synthetic and real data are and the second evaluates the image synthesis potential of the synthetic data. We consider three categories in the assessment: ground-glass opacity, consolidation, and infection (which considers both ground-glass opacity and consolidation). The evaluation results for the two experiments are presented in Table III and Table IV, respectively. The pre-trained Unet model is then tested with a fixed real CT image dataset. 10,220 images from the test set are divide equally into 10 folds, the evaluation results are reported with the format as *MEAN* ± *95% CONFIDENCE INTERVAL* among above folds.

In Table III, we describe the experimental results of different replacing ratios of synthetic data. We can obtain the best performance when using pure real data as a training set. By replacing the real data with a ratio of synthetic data, the semantic segmentation performance of Unet does not decrease and stay at a stable level. By replacing real data with 30% synthetic data, the Unet obtains the best performance on the Spec metric for ground-glass opacity focus, also it gets the best performance on Dice and Sen metrics for consolidation focus. The experimental results from Table III show that synthetic CT images are similar to real CT images. They are realistic enough even replacing the real data with a large ratio of synthetic data, the semantic segmentation performance of Unet still seems promising.

Table III presents the experimental results for different replacement ratios for the synthetic data. We obtain the best

TABLE III
EXPERIMENTAL RESULTS FOR CT IMAGES USING SEMANTIC SEGMENTATION METHODS
(REPLACING REAL DATA WITH DIFFERENT PROPORTIONS OF SYNTHETIC DATA)
(THE BEST EVALUATION SCORE IS MARKED IN BOLD. ↑ MEANS HIGHER NUMBER IS BETTER, AND ↓ INDICATES LOWER NUMBER IS BETTER. RATIO MEANS REPLACING CERTAIN PROPORTION OF SYNTHETIC DATA. ε REPRESENTS A SMALL POSITIVE QUANTITY WHICH IS SMALLER THAN $1e^{-5}$.)

Focus	Ground-glass opacity			Consolidation			Infection		
Ratio	Dice (%, ↑)	Sen (%, ↑)	Spec (%, ↑)	Dice (%, ↑)	Sen (%, ↑)	Spec (%, ↑)	Dice (%, ↑)	Sen (%, ↑)	Spec (%, ↑)
0%	**87.55±0.20**	**86.84±0.31**	99.82±0.01	84.88±0.33	82.80±0.51	99.96±ε	**89.57±0.18**	**88.58±0.23**	99.82±0.01
10%	87.34±0.27	85.08±0.45	**99.85±0.01**	85.91±0.44	84.23±0.55	99.96±ε	89.35±0.32	87.14±0.39	**99.85±0.01**
20%	84.22±0.36	83.38±0.41	99.77±0.01	84.30±0.19	83.60±0.33	99.95±ε	86.67±0.36	85.83±0.27	99.76±0.01
30%	87.43±0.35	85.73±0.41	99.84±0.01	**86.01±0.21**	**87.14±0.21**	99.95±ε	89.32±0.22	88.11±0.30	99.82±0.01
40%	87.07±0.27	86.70±0.29	99.80±0.01	85.81±0.24	81.94±0.36	**99.97**±ε	88.92±0.20	87.90±0.30	99.81±0.01
50%	86.98±0.35	86.46±0.40	99.80±0.01	85.58±0.23	82.18±0.36	**99.97**±ε	89.19±0.24	88.12±0.37	99.82±0.01

TABLE IV
EXPERIMENTAL RESULTS FOR CT IMAGES USING SEMANTIC SEGMENTATION METHODS
(ADDING SYNTHETIC DATA WITH DIFFERENT PROPORTIONS)
(THE BEST EVALUATION SCORE IS MARKED IN BOLD. ↑ MEANS HIGHER NUMBER IS BETTER, AND ↓ INDICATES LOWER NUMBER IS BETTER. RATIO MEANS ADDING CERTAIN PROPORTION OF SYNTHETIC DATA. ε REPRESENTS A SMALL POSITIVE QUANTITY WHICH IS SMALLER THAN $1e^{-5}$.)

Focus	Ground-glass opacity			Consolidation			Infection		
Ratio	Dice (%, ↑)	Sen (%, ↑)	Spec (%, ↑)	Dice (%, ↑)	Sen (%, ↑)	Spec (%, ↑)	Dice (%, ↑)	Sen (%, ↑)	Spec (%, ↑)
0%	87.55±0.20	86.84±0.31	99.82±0.01	84.88±0.33	82.80±0.51	**99.96**±ε	89.57±0.18	88.58±0.23	99.82±0.01
10%	87.65±0.40	85.82±0.32	**99.84±0.01**	86.12±0.33	86.02±0.63	99.95±ε	89.67±0.31	88.11±0.32	**99.84±0.01**
20%	87.87±0.34	87.67±0.28	99.81±0.01	85.52±0.34	84.16±0.53	**99.96**±ε	89.87±0.12	89.44±0.20	99.81±0.01
30%	87.99±0.36	87.25±0.36	99.82±0.01	86.33±0.30	86.38±0.51	99.95±ε	89.78±0.22	89.17±0.27	99.82±0.01
40%	**88.33±0.22**	**88.71±0.32**	99.81±0.01	**87.25±0.28**	86.30±0.38	**99.96**±ε	**90.19±0.17**	**90.34±0.31**	99.81±ε
50%	88.16±0.30	86.88±0.01	**99.84±0.01**	87.09±0.34	**86.54±0.47**	**99.96**±ε	90.06±0.30	88.88±0.18	99.83±0.01

performance when using pure real data as the training set. By replacing the real data with a proportion of synthetic data, the semantic segmentation performance of Unet does not decrease, but rather remains stable. By replacing real data with 30% synthetic data, Unet obtains the best performance for the Spec metric for ground-glass opacity and for the Dice and Sen metrics for consolidation. The experimental results thus indicate that the synthetic CT images are similar to real CT images. They are sufficiently realistic for semantic segmentation with Unet to be successful even when real data is replaced with a large proportion of synthetic data.

Table IV presents the semantic segmentation results when a certain proportion of extra synthetic data is added to the real data. The best performance is obtained when adding 40% synthetic data. Overall, the results indicate that the synthetic CT images are sufficiently diverse and realistic, meaning that they have the potential to be utilized in image synthesis to improve the dataset quality for deep-learning-based COVID-19 diagnosis.

C. Qualitative results

To intuitively demonstrate synthetic results and easily compare them with the results from other state-of-the-art image synthesis methods, we show the synthetic examples in both Figure 6 and Figure 7 in this subsection.

The synthetic images from three individual cases are compared in Figure 6. The first case shows that a consolidation infection area locates on the lower left of CT image. By comparing the synthetic results from the proposed method, SEAN [15] and SPADE [14], the infection area remains the original structure and texture in the result which is generated by the proposed method, however, we found that in the results of SEAN and SPADE, there some unnatural artifacts (holes) are generated in the position that yellow arrow points out. For the second case, a large area of ground-glass infection is detected, the results of SPADE and SEAN ignore some small lung area in the middle of the infection area, but the proposed method can still reflect above small lung area correctly. In the final case, it contains both two categories of infection area: consolidation and ground-glass opacity, and the ground-glass opacity are surrounded by the consolidation area. If we focus on the surrounded area, we can found out that the boundary of two infection area is not clear in the synthetic image of SEAN, and the ground-glass area are mistakenly generated as lung area in the synthesized image of SPADE. The result of the proposed method in case 3 shows that it has the ability to handle this complex situation and produce realistic synthetic CT images with high image quality.

We present some synthetic examples that are generated by the proposed method in Figure 7. We select one example for each patient (8 samples from 9 patients; patient #3 is skipped because the segmentation maps were miss-labeled). For Patient #0, the consolidation area is located at the bottom of the lung area; the synthetic image shows a sharp and high-contrast consolidation area that can be easily distinguished from the surrounding non-lung region. The slides for Patients #1 and #4 have a similarity in that the lung area contains widespread ground-glass opacity. Consolidation is sporadically located within this ground-glass opacity. The small consolidation area can be easily identified due to the clear boundary between the two infection areas. Patient #6 shows ground-glass opacity and consolidation that are distant from each other. The results thus illustrate that the proposed method can handle the two types of infection areas together in a single lung CT image. The CT slides of Patients #5, #7, and #8 show the simplest cases, with only a single category of infection (ground- glass opacity). The experimental results thus indicate that realistic ground-glass opacity can be obtained using the proposed method.

V. Conclusion and future study

In this paper, we proposed a cGAN-based COVID-19 CT image synthesis method that can generate realistic CT images that include the two main infection types, ground-glass opacity, and consolidation. The proposed method takes the semantic segmentation map of a corresponding lung CT image, and the cGAN structure learns the characteristics and information of the CT image. A global-local generator and a multi-resolution discriminator are employed to effectively balance global information with local details in the CT image. The experimental results show that the proposed method is able to generate realistic synthetic CT images and achieve state-of-the-art performance in terms of image quality when compared with common image synthesis approaches. In addition, the evaluation results for semantic segmentation performance show that the high image quality and fidelity of the synthetic CT images enable their use in image synthesis for COVID-19 diagnosis using AI models. For future research, the authors plan to fully utilize high-quality synthetic COVID-19 CT images to improve specific computer vision approaches that can help in the fight against COVID-19, such as lung CT image semantic segmentation and rapid lung CT image-based COVID-19 diagnosis.

References

[1] "Coronavirus disease (covid-19) pandemic," https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

[2] "Johns hopkins coronavirus resource center," https://coronavirus.jhu.edu/map.html.

[3] W.-j. Guan, Z.-y. Ni, Y. Hu, W.-h. Liang, C.-q. Ou, J.-x. He, L. Liu, H. Shan, C.-l. Lei, D. S. Hui *et al.*, "Clinical characteristics of coronavirus disease 2019 in china," *New England journal of medicine*, vol. 382, no. 18, pp. 1708–1720, 2020.

[4] "Interim clinical guidance for management of patients with confirmed coronavirus disease (covid-19)," https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html.

[5] T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, and L. Xia, "Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases," *Radiology*, p. 200642, 2020.

[6] Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, and W. Ji, "Sensitivity of chest ct for covid-19: comparison to rt-pcr," *Radiology*, p. 200432, 2020.

[7] K. Yan, Y. Peng, V. Sandfort, M. Bagheri, Z. Lu, and R. M. Summers, "Holistic and comprehensive annotation of clinically significant findings on diverse ct images: learning from radiology reports and label ontology," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8523–8532.

[8] Z. Cui, C. Li, and W. Wang, "Toothnet: automatic tooth instance segmentation and identification from cone beam ct images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6368–6377.

[9] Z. Zhang, A. Romero, M. J. Muckley, P. Vincent, L. Yang, and M. Drozdzal, "Reducing uncertainty in undersampled mri reconstruction with active acquisition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2049–2058.

[10] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, and Y. Zheng, "X2ct-gan: reconstructing ct from biplanar x-rays with generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 10 619–10 628.

[11] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[12] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[13] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8798–8807.

[14] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2337–2346.

[15] P. Zhu, R. Abdal, Y. Qin, and P. Wonka, "Sean: Image synthesis with semantic region-adaptive normalization," 2019.

[16] Y.-S. Chen, Y.-C. Wang, M.-H. Kao, and Y.-Y. Chuang, "Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6306–6314.

[17] Z. Wan, B. Zhang, D. Chen, P. Zhang, D. Chen, J. Liao, and F. Wen, "Bringing old photos back to life," *arXiv preprint arXiv:2004.09484*, 2020.

[18] W. Yang, W. Ouyang, X. Wang, J. Ren, H. Li, and X. Wang, "3d human pose estimation in the wild by adversarial learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5255–5264.

[19] L. Ma, X. Jia, Q. Sun, B. Schiele, T. Tuytelaars, and L. Van Gool, "Pose guided person image generation," in *Advances in Neural Information Processing Systems*, 2017, pp. 406–416.

[20] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, G. Liu, A. Tao, J. Kautz, and B. Catanzaro, "Video-to-video synthesis," *arXiv preprint arXiv:1808.06601*, 2018.

[21] T.-C. Wang, M.-Y. Liu, A. Tao, G. Liu, J. Kautz, and B. Catanzaro, "Few-shot video-to-video synthesis," *arXiv preprint arXiv:1910.12713*, 2019.

[22] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[23] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," in *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017, pp. 2642–2651.

[24] H. Caesar, J. Uijlings, and V. Ferrari, "Coco-stuff: Thing and stuff classes in context," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1209–1218.

[25] L. Mescheder, A. Geiger, and S. Nowozin, "Which training methods for gans do actually converge?" *arXiv preprint arXiv:1801.04406*, 2018.

[26] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, and X. He, "Attngan: Fine-grained text to image generation with attentional generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1316–1324.

[27] S. Hong, D. Yang, J. Choi, and H. Lee, "Inferring semantic layout for hierarchical text-to-image synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7986–7994.

[28] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[29] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman, "Toward multimodal image-to-image translation," in *Advances in neural information processing systems*, 2017, pp. 465–476.

[30] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Advances in neural information processing systems*, 2017, pp. 700–708.

[31] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 172–189.

[32] B. Zhao, L. Meng, W. Yin, and L. Sigal, "Image generation from layout," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8584–8593.

[33] H. Kang, L. Xia, F. Yan, Z. Wan, F. Shi, H. Yuan, H. Jiang, D. Wu, H. Sui, C. Zhang *et al.*, "Diagnosis of coronavirus disease 2019 (covid-19) with structured latent multi-view representation learning," *IEEE transactions on medical imaging*, 2020.

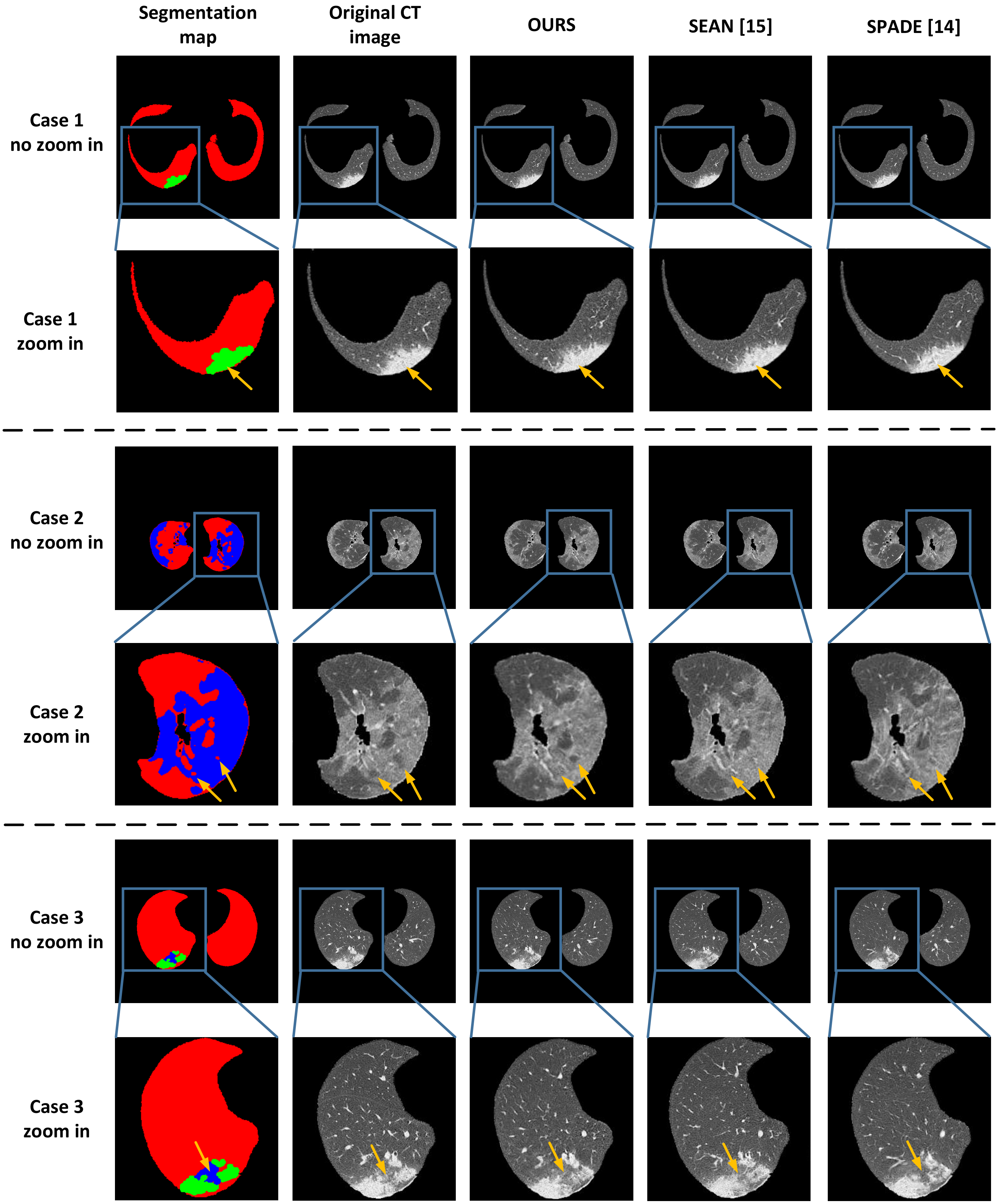

Fig. 6. Synthetic lung CT images generated by the proposed method and the other two competitive state-of-the-art image synthesis approaches. The first column shows the segmentation map including the lung (red), ground-glass opacity (blue), and consolidation (green) areas. The second column shows the original CT image. The third, fourth, fifth columns show the synthetic samples which are generated by the proposed method, SEAN [15] and SPADE [14] in order. Each case is presented with zoom in order to show more details, and the yellow arrows point out the special area which is described in the main text.

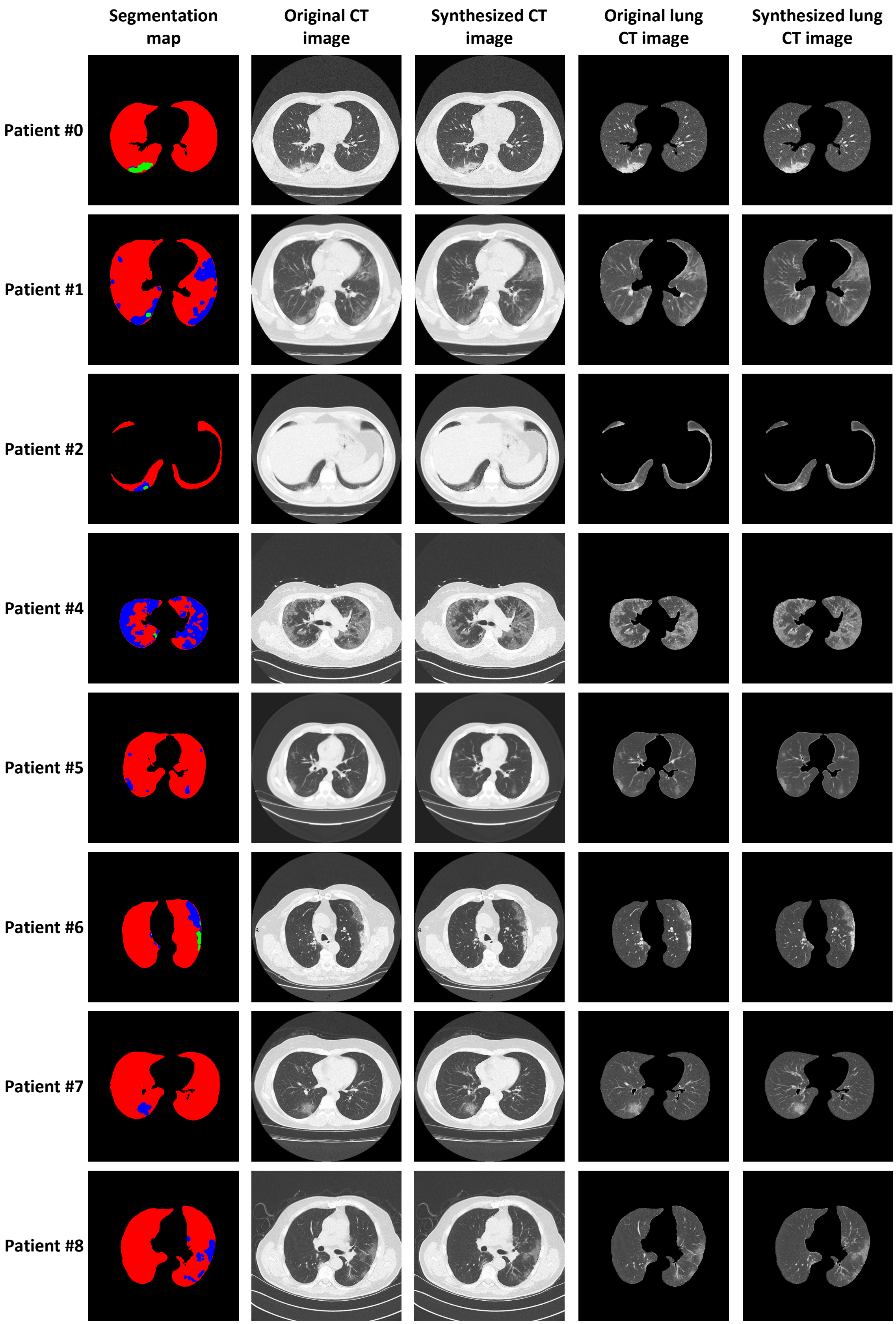

Fig. 7. Synthetic lung CT images generated by the proposed method. Eight samples are selected, each from an individual patient. The first column shows the segmentation map including the lung (red), ground-glass opacity (blue), and consolidation (green) areas. The second and third columns show the original and synthetic CT images, respectively. The synthetic CT images here merge the synthetic lung CT image and the corresponding real non-lung area. The fourth and fifth columns depict CT images for the original lung and synthesized CT images, respectively.

[34] K. Zhang, X. Liu, J. Shen, Z. Li, Y. Sang, X. Wu, Y. Zha, W. Liang, C. Wang, K. Wang *et al.*, "Clinically applicable ai system for accurate diagnosis, quantitative measurements, and prognosis of covid-19 pneumonia using computed tomography," *Cell*, 2020.

[35] A. A. Ardakani, A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi, "Application of deep learning technique to manage covid-19 in routine clinical practice using ct images: Results of 10 convolutional neural networks," *Computers in Biology and Medicine*, p. 103795, 2020.

[36] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song *et al.*, "Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct," *Radiology*, p. 200905, 2020.

[37] T. Zhou, S. Canu, and S. Ruan, "An automatic covid-19 ct segmentation based on u-net with attention mechanism," *arXiv preprint arXiv:2004.06673*, 2020.

[38] W. Xie, C. Jacobs, J.-P. Charbonnier, and B. van Ginneken, "Relational modeling for robust and efficient pulmonary lobe segmentation in ct scans," *IEEE Transactions on Medical Imaging*, 2020.

[39] A. Voulodimos, E. Protopapadakis, I. Katsamenis, A. Doulamis, and N. Doulamis, "Deep learning models for covid-19 infected area segmentation in ct images," *medRxiv*, 2020.

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

[41] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Inf-net: Automatic covid-19 lung infection segmentation from ct images," *IEEE Transactions on Medical Imaging*, 2020.

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

[43] O. Gozes, M. Frid-Adar, N. Sagie, H. Zhang, W. Ji, and H. Greenspan, "Coronavirus detection and analysis on chest ct with deep learning," *arXiv preprint arXiv:2004.02640*, 2020.

[44] S. Hu, Y. Gao, Z. Niu, Y. Jiang, L. Li, X. Xiao, M. Wang, E. F. Fang, W. Menpes-Smith, J. Xia *et al.*, "Weakly supervised deep learning for covid-19 infection detection and classification from ct images," *arXiv preprint arXiv:2004.06689*, 2020.

[45] A. D. Lauritzen, X. Papademetris, S. Turovets, and J. A. Onofrey, "Evaluation of ct image synthesis methods: From atlas-based registration to deep learning," *arXiv preprint arXiv:1906.04467*, 2019.

[46] Y.-W. Chen, H.-Y. Fang, Y.-C. Wang, S.-L. Peng, and C.-T. Shih, "A novel computed tomography image synthesis method for correcting the spectrum dependence of ct numbers," *Physics in Medicine & Biology*, vol. 65, no. 2, p. 025013, 2020.

[47] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European conference on computer vision*. Springer, 2016, pp. 694–711.

[48] "Covid-19 ct segmentation dataset," https://medicalsegmentation.com/covid19/.

[49] M. Chung, A. Bernheim, X. Mei, N. Zhang, M. Huang, X. Zeng, J. Cui, W. Xu, Y. Yang, Z. A. Fayad *et al.*, "Ct imaging features of 2019 novel coronavirus (2019-ncov)," *Radiology*, vol. 295, no. 1, pp. 202–207, 2020.

[50] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems*, 2017, pp. 6626–6637.

[51] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th International Conference on Pattern Recognition*. IEEE, 2010, pp. 2366–2369.

[52] A. Fenster and B. Chiu, "Evaluation of segmentation algorithms for medical imaging," in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*. IEEE, 2006, pp. 7186–7189.

[53] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.

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