Pseudo-Label Guided Image Synthesis for Semi-Supervised COVID-19 Pneumonia Infection Segmentation

Fei Lyu, Mang Ye, Jonathan Frederik Carlsen, Kenny Erleben, Sune Darkner, and Pong C. Yuen, *Senior Member, IEEE*

***Abstract*— Coronavirus disease 2019 (COVID-19) has become a severe global pandemic. Accurate pneumonia infection segmentation is important for assisting doctors in diagnosing COVID-19. Deep learning-based methods can be developed for automatic segmentation, but the lack of large-scale well-annotated COVID-19 training datasets may hinder their performance. Semi-supervised segmentation is a promising solution which explores large amounts of unlabeled data, while most existing methods focus on pseudo-label refinement. In this paper, we propose a new perspective on semi-supervised learning for COVID-19 pneumonia infection segmentation, namely pseudo-label guided image synthesis. The main idea is to keep the pseudo-labels and synthesize new images to match them. The synthetic image has the same COVID-19 infection regions as indicated in the pseudo-label, and the reference style extracted from the style code pool is added to make it more realistic. We introduce two representative methods by incorporating the synthetic images into model training, including single-stage Synthesis-Assisted Cross Pseudo Supervision (SA-CPS) and multi-stage Synthesis-Assisted Self-Training (SA-ST), which can work individually as well as cooperatively. Synthesis-assisted methods are featured in two aspects: 1) rectify the training bias caused by inaccurate pseudo-labels 2) expand the training data by using additional synthetic data. Extensive experiments on two COVID-19 CT datasets for segmenting the infection regions demonstrate our method is superior to existing schemes for semi-supervised segmentation, and achieves the state-of-the-art performance on both datasets.**

***Index Terms*— Semi-supervised learning, image synthesis, self-training, COVID-19 CT segmentation.**

I. Introduction

CORONAVIRUS disease 2019 (COVID-19) has caused a serious health crisis and generated unprecedented social disruptions globally [1]–[3]. By the end of 2021, more than 43% of the global population has been infected with COVID-19 at least once, and over 5.94 million people died worldwide because of the COVID-19 pandemic [4], [5]. Early detection of COVID-19 is helpful for prevent spreading and proper resource allocation during the pandemic [6]. Though reverse transcription polymerase chain reaction (RT-PCR) remains the gold standard for COVID-19 diagnosis, several studies suggest that the RT-PCR test has low sensitivity [7], [8]. Chest computed tomography (CT) is an important complement to the RT-PCR test, and has demonstrated its effectiveness in accurate diagnosis and follow-up assessment [9]. Recently, deep learning-based methods have been proposed to combat COVID-19 [10], [11], for example, the severity classification task helps distinguish those severe patients who need emergency medical care [11]. However, automatic segmentation of COVID-19 pneumonia regions still remains a challenging task [12]–[14]. It is impractical to collect large-scale well-labeled datasets due to the urgent nature of the pandemic. The labeling process is tedious and requires professional knowledge. On average, it takes around 400 minutes to annotate one CT volume with 250 slices [15]. Therefore, it is desirable to develop label-efficient deep learning models for automatic and accurate COVID-19 pneumonia infection segmentation.

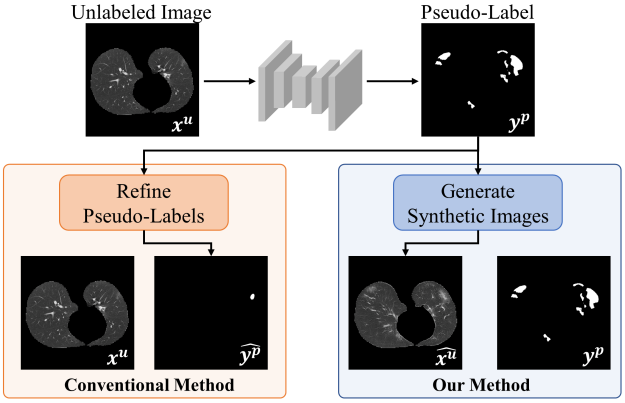

Fig. 1. Illustration of our motivation: Conventional semi-supervised learning methods focus on refining pseudo-labels of unlabeled images with elaborate designs. We consider this problem from another perspective, where we keep the pseudo-labels and synthesize new images to match them, then synthetic pairs are incorporated into model training.

This work was supported by the Health and Medical Research Fund Project under Grant 07180216. (Corresponding author: Pong C. Yuen.)

Fei Lyu and Pong C. Yuen are with the Department of Computer Science, Hong Kong Baptist University, Hong Kong. (e-mail: feilyu@comp.hkbu.edu.hk; pcyuen@comp.hkbu.edu.hk).

Mang Ye is with the School of Computer Science, Wuhan University, Wuhan, China. (e-mail: mangye16@gmail.com).

Jonathan Frederik Carlsen is with Department of Diagnostic Radiology, Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark. (e-mail: jonathan.carlsen@gmail.com).

Kenny Erleben and Sune Darkner are with Department of Computer Science, University of Copenhagen, Copenhagen, Denmark. (e-mail: kenny@di.ku.dk; darkner@di.ku.dk).

Recently, there is a surge of interest in semi-supervised segmentation, which explores large amounts of unlabeled data for better performance. They can be divided into two categories, one-stage and multi-stage. There are some successful one-stage methods, such as consistency regularization [16]–[19] and GAN-based methods [20], [21]. However, those single-stage methods are optimized with both labeled and unlabeled samples in a single stage, and the performance is highly affected by those inaccurate predictions in the early training stage. Self-training is a classical multi-stage method with three stages [22], [23]. It first trains a teacher model to generate pseudo-labels for unlabeled samples, and retrains a student model with the expanded training data. However, the quality of the generated pseudo-labels can not be guaranteed, and may bring marginal improvement with a poor teacher model. In this paper, we would like to explore the usage of pseudo-labels from unlabeled data more efficiently for semi-supervised COVID-19 pneumonia infection segmentation.

Inspired by recent advances in semantic image synthesis [24], [25], we propose to synthesize new CT images to match the pseudo-labels of unlabeled samples, rather than designing sophisticated mechanisms for assessing and refining those pseudo-labels. Recent techniques for semantic image synthesis allow controlling the style of each region individually in the segmentation mask, which yields high quality of the synthesized images. The main idea of pseudo-label guided image synthesis to control the layout of the generated image which has the same COVID-19 infection regions as indicated in the pseudo-labels, and add reference styles to make it more realistic. Considering the pseudo-label inevitably contains low-quality predictions and the extracted region-wise style codes from itself may be inaccurate, we propose to build a style code pool with the style codes extracted from all labeled samples which contain infection regions. The quality of the reference style codes sampled from the style code pool can be guaranteed, and leads to better synthesized results.

With the synthetic images, we introduce two novel methods by incorporating them into model training, *i.e.*, Synthesis-Assisted Cross Pseudo Supervision (SA-CPS) and Synthesis-Assisted Self-Training (SA-ST), which can work individually as well as cooperatively. SA-CPS is a single-stage method, which feeds the synthetic images together with the labeled and unlabeled images into two segmentation networks that have the same architecture but different initialization weights. Similar to self-training, SA-ST is a multi-stage method, which requires a teacher model to generate pseudo-labels for unlabeled data, and the re-trained student model can normally bring better performance than its teacher model. The teacher model can be a model trained only using a small amount of labeled data or a trained model from SA-CPS. In SA-ST, the synthetic images together with the labeled and unlabeled images are used to retrain a student model. The advantages of synthesis-assisted methods lie in two aspects. First, the contribution of the unlabeled data is averaged by the original pseudo pairs (*unlabeled image and its pseudo-label*) and the new synthetic pairs (*pseudo-label and its synthetic image*), such that the negative impacts caused by inaccurate pseudo-labels can be alleviated. Second, the synthetic data is used as additional supervision to train the model, which behaves like expanding the training data, and thus improving the model performance.

The main contributions are summarized as follows:

- We propose a new perspective on semi-supervised learning for COVID-19 pneumonia infection segmentation, where synthetic images are generated to match the pseudo-labels of unlabeled images and added to existing training data for improved performance.
- We introduce two representative methods by incorporating the synthetic images into model training, *i.e.*, single-stage Synthesis-Assisted Cross Pseudo Supervision and multi-stage Synthesis-Assisted Self-Training, which can work individually as well as cooperatively.
- We conduct extensive experiments on two COVID-19 CT datasets for segmenting the infection regions, and the results show that our methods outperform state-of-the-art methods when training with limited labeled data.

II. Related work

A. COVID-19 Pneumonia Infection Segmentation

Segmenting the pneumonia regions is a crucial task which can assist the quantification and diagnosis of COVID-19. It is a quite challenging task due to the high variation of infection appearances. Recently, deep learning methods have been employed to segment the COVID-19 infection regions automatically. Among various network architectures, U-Net and its variants are popular choices [12]–[14]. Moreover, some novel designs are incorporated into the models, such as attention mechanism [26], [27], joint learning [28], [29] and transfer learning [15], [30]. Despite their good performance, they could not realize full potential of deep learning models due to the small dataset size. Considering the challenge of data collection and annotation, some pioneer works attempt to train models with weak labels [31], [32] or noisy labels [33]. Nevertheless, the performance of existing label-efficient methods for COVID-19 pneumonia infection segmentation is obviously inferior to those fully supervised methods. In this work, we aim at developing a more advanced deep model trained with limited labeled data and achieve better segmentation performance.

B. Semi-Supervised Segmentation

Collecting pixel-level annotations for semantic segmentation is extremely costly and time-consuming. Therefore, semi-supervised learning that utilizes both labeled and unlabeled data is attracting more attention for training segmentation models. There are many successful works on semi-supervised semantic segmentation, such as consistency regularization [16], [17], GAN-based model [20], [21], self-training [22], [23], *etc*. The key idea of consistency regularization is to enforce the consistency of the predictions from augmented input images, perturbed features, or different networks. GAN-based models either generate additional training data or learn an additional discriminator to distinguish the prediction from the ground truth. In self-training, a teacher model is first trained with labeled data, then it is used to generate pseudo-labels on a large set of unlabeled data, finally a student model

is trained using both labeled and pseudo-labeled data. Different from other single-stage methods, self-training is multi-stage and can be iterated many times until reaching the satisfactory performance. Indeed, some works have investigated semi-supervised learning methods on COVID-19 pneumonia infection segmentation [18], [19], [34]–[36]. However, they focus more on developing advanced networks or refining the pseudo labels of unlabeled data. In comparison, we consider this problem from another perspective, where we retain the pseudo-labels and synthesize new images to match them, then the synthetic data is added to existing labeled and unlabeled data for model training. Our approach is more generic and simpler as we do not require specially designed loss functions or regularization techniques, and is less affected by low-quality pseudo-labels during training.

C. Semantic Image Synthesis

Semantic image synthesis refers to the task of converting a semantic segmentation mask to a realistic image, and it has been extensively studied in the computer vision community. The most representative work, Pix2Pix [37], introduces an encoder-decoder generator for semantic image synthesis. Pix2PixHD [38] is the follow-up work which improves Pix2Pix by proposing a coarse-to-fine training scheme. After this seminal work, many subsequent methods are proposed to improve the performance of semantic image synthesis, such as SPADE [24] and SEAN [25]. SPADE and SEAN propose novel normalization layers to synthesize images with high-quality, and edit the input image controlled by style image and segmentation masks. SEAN improves SPADE by allowing per-region style encoding, which leads to better synthesized results. In the medical domain, some recent works also explore the synthesis methods for different downstream tasks, such as data augmentation [39]–[41] and segmentation quality assessment [42]. Different from these works, we are interested in a new application of semantic image synthesis, where we synthesize images controlled by pseudo-labels from unlabeled data in the semi-supervised learning setting, and improve the segmentation performance with these synthetic images.

III. Method

In this paper, we propose a new perspective on semi-supervised learning for COVID-19 pneumonia infection segmentation with limited labeled data. Inspired by recent advances in semantic image synthesis [24], [25], we propose to synthesize new CT images to match the pseudo-labels of unlabeled samples, rather than designing sophisticated mechanisms for assessing and refining those pseudo-labels. The synthetic images and pseudo-labels are formulated as additional synthetic pairs, and are incorporated into model training for achieving better segmentation performance.

In the following section, we first clearly define the problem of semi-supervised COVID-19 pneumonia infection segmentation, then describe the details of pseudo-label guided image synthesis. With the synthetic images, we introduce the single-stage method Synthesis-Assisted Cross Pseudo Supervision (SA-CPS) and multi-stage method Synthesis-Assisted Self-Training (SA-ST) in detail.

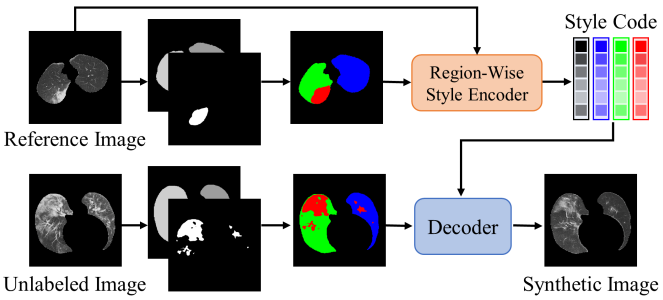

Fig. 2. Illustration of pseudo-label guided image synthesis. The pseudo-label of an unlabeled image is obtained from the model. The region-wise style encoder takes a reference image and outputs the style code. The decoder generates the synthetic image which matches the pseudo-label with the reference style.

A. Problem Definition

The task of semi-supervised COVID-19 pneumonia infection segmentation is formulated as follows. A dataset $\mathcal{D}$ is used for training the models, which contains a small amount of labeled samples and a large amount of unlabeled samples. $\mathcal{D}_L = \{x_i^l, y_i\}_{i=1}^N$ represent N labeled CT samples and $\mathcal{D}_U = \{x_i^u\}_{i=1}^M$ represent M unlabeled CT samples, where x_i^l and x_i^u denote CT images and y_i is the corresponding groundtruth segmentation mask. We aim at exploiting the large amount of unlabeled data ($M \gg N$) to obtain an improved segmentation model which can achieved the performance comparable to the model trained using fully labeled datasets.

B. Pseudo-Label Guided Image Synthesis

The pseudo labels generated by a model trained with a small amount of labeled samples inevitably contain inaccurate predictions. Unlike conventional methods, we synthesize new images to match these pseudo-labels rather than refining them, which is illustrated in Fig. 2.

The region-wise style encoder Enc extracts per region styles from the reference image, and the output **style code** $\mathcal{SC}$ is a $512 \times s$ dimensional matrix [25], where s is the number of semantic labels. To keep the layout of lung regions, we combine the masks of lung segmentation and COVID-19 pneumonia infection segmentation, to formulate a semantic mask of $s = 4$ component categories, *i.e.*, background, left lung, right lung and infection region. If a category does not exist in the input image, we set its corresponding column in $\mathcal{SC}$ to zero. With the reference code and the semantic mask, the decoder Dec generates a realistic CT image which contains infections in the segmented regions from the pseudo-label.

The next question is how to find a reference image to obtain the style code. One straightforward solution is to use the unlabeled image itself and its pseudo-label. However, the style code may be biased due to the inaccurate predictions in the pseudo-label. For example, an unlabeled image has no infections but its pseudo-label contains infection labels, such that the synthetic image still has no infections because no style information of the infection category is available. Another solution to borrow the style codes from the labeled samples, because their groundtruth segmentation masks are provided

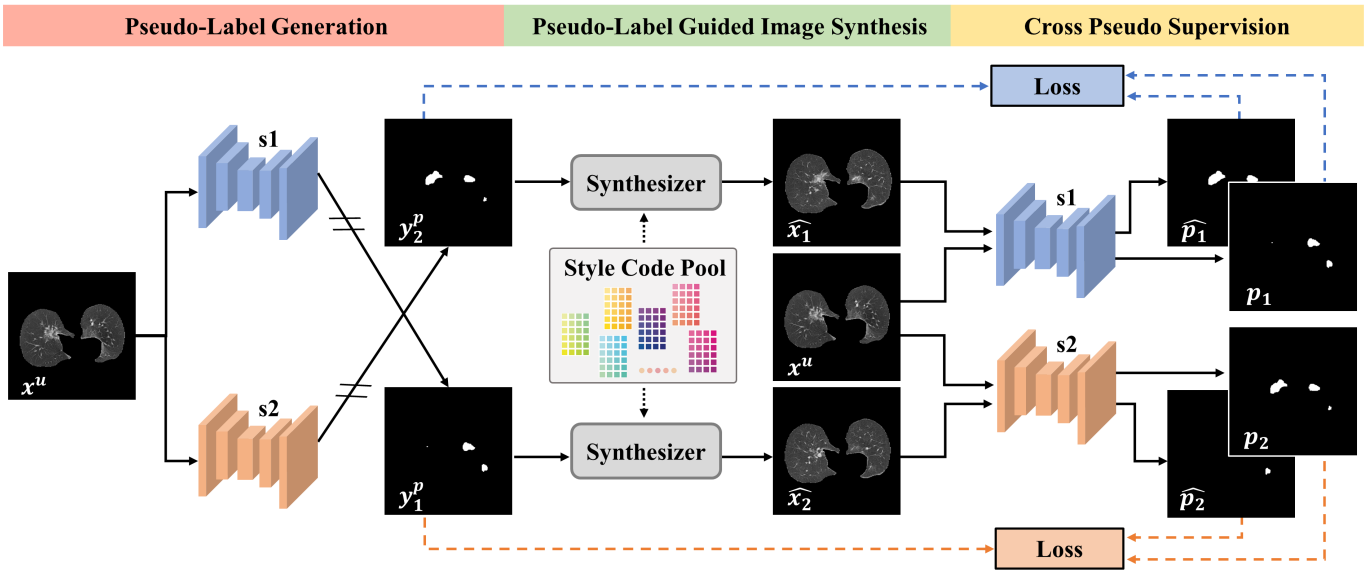

Fig. 3. Overview of synthesis-assisted cross pseudo supervision (SA-CPS) on unlabeled training data. The framework contains two parallel segmentation networks $s1$ and $s2$. Following the spirit of cross pseudo supervision, we use the pseudo-label y_1^p from $s1$ to supervise the other network $s2$, and use y_2^p from $s2$ to supervise network $s1$. ' // ' on ' $\rightarrow$' means stop gradient, which is used when generating pseudo-labels. Given the pseudo-label, we can synthesize a realistic CT image with the reference style code sampled from the style code pool. Each network is trained with three loss functions, including supervised loss, pseudo-supervised loss and synthetic-supervised loss.

and the extracted style codes are more accurate. Therefore, we build a **Style Code Pool** $\mathcal{P}$ with the style codes of labeled samples which contain infection regions:

$$\mathcal{P} = \{\, Enc(\, x_i^l,\, y_i)\,\}_{i=1}^{N^+}, \quad (x_i^l, y_i) \in \mathcal{D}_L, \tag{1}$$

where N^+ is the number of style codes from labeled training samples with infection regions, and $Enc(\cdot)$ denotes the output from the style encoder which is a 512×4 dimensional matrix.

To avoid large variations, we randomly sample K style codes in $\mathcal{P}$ and average them to obtain the reference style code $\mathcal{SC}_{ref}$ in each mini-batch iteration:

$$\mathcal{SC}_{ref} = \frac{1}{K}\sum_{i=1}^{K} \mathcal{SC}_i, \quad \mathcal{SC}_i \in \mathcal{P}, \tag{2}$$

where $\mathcal{SC}_i$ is the style code sampled from the style code pool.

With the reference style code $\mathcal{SC}_{ref}$, we can synthesize a realistic CT image $\hat{x}$ to match the pseudo-label y^p:

$$\hat{x} = Dec(\, y^p,\, \mathcal{SC}_{ref}), \tag{3}$$

where $Dec(\cdot)$ denotes the output from the decoder.

C. Synthesis-Assisted Cross Pseudo Supervision

In this subsection, we introduce the one-stage method Synthesis-Assisted Cross Pseudo Supervision (SA-CPS). Predictions from two peer networks can provide complementary information for each other, and the idea has been proven to be effective in semi-supervised learning [16], [19]. In SA-CPS, we adopt two parallel segmentation networks $s1$ and $s2$:

$$p_1 = f_{s1}(x), \quad p_2 = f_{s2}(x), \tag{4}$$

where $f_{s1}(\cdot)$ and $f_{s2}(\cdot)$ denote the predictions from $s1$ and $s2$. $s1$ and $s2$ have the same architecture but different initialization weights, to guarantee the diversity of two peer networks.

During training, each network takes a batch of labeled samples $\{x_i^l, y_i\}_{i=1}^B \in \mathcal{D}_L$ and unlabeled samples $\{x_i^u\}_{i=1}^B$, where B is the batch size. The training objective consists of a supervised loss $\mathcal{L}_s$ applied to labeled data and an unsupervised loss $\mathcal{L}_u$ applied to unlabeled data. Specifically, the supervised loss $\mathcal{L}_s$ is formulated using the combination of binary cross-entropy ($\mathcal{L}_{BCE}$) and dice loss functions ($\mathcal{L}_{DICE}$) :

$$\mathcal{L}_s = \mathcal{L}_{Seg}(f_{s1}(x_i^l),\, y_i) + \mathcal{L}_{Seg}(f_{s2}(x_i^l),\, y_i), \tag{5}$$

with

$$\mathcal{L}_{Seg}(p,\, y) = \mathcal{L}_{BCE}(p,\, y) + \mathcal{L}_{DICE}(p,\, y), \tag{6}$$

$$\mathcal{L}_{BCE}(p,\, y) = \sum(y \cdot \log(p) + (1-y) \cdot \log(1-p)), \tag{7}$$

$$\mathcal{L}_{DICE}(p,\, y) = 1 - \frac{2|p \cap y|}{|p| + |y| + \epsilon}, \tag{8}$$

where x_i is the input image and y_i is its pixel-level annotation, $(x_i^l, y_i) \in \mathcal{D}_L$. ϵ refers to the smooth parameter.

Next, we elaborate the unlabeled data training branch, as illustrated in Fig. 3. The unlabeled image has no groundtruth labels, and we first generate pseudo-labels from each network:

$$y_1^p = \mathbb{1}(f_{s1}(x_u) \geq \tau), \quad y_2^p = \mathbb{1}(f_{s2}(x_u) \geq \tau), \tag{9}$$

where $\mathbb{1}(.)$ is the indicator function, and τ is the threshold for binarization, $\tau \in (0, 1)$. Following the spirit of cross pseudo supervision, we use y_1^p from $s1$ to supervise the other network $s2$, and use y_2^p from $s2$ to supervise network $s1$. Given the pseudo-labels, we can synthesize CT images to match them:

$$\hat{x}_1 = Dec(\, y_2^p,\, \mathcal{SC}_{ref}), \quad \hat{x}_2 = Dec(\, y_1^p,\, \mathcal{SC}_{ref}), \tag{10}$$

where $\mathcal{SC}_{ref}$ is the reference style code sampled from the style code pool $\mathcal{P}$.

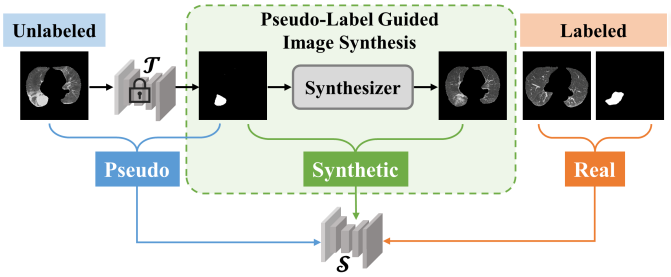

Fig. 4. Illustration of synthesis-assisted self-training (SA-ST). First, a teacher model ($\boldsymbol{T}$) is trained on a small amount of labeled data. Next, the fixed teacher model generates pseudo-labels for a large amount of unlabeled data. **Moreover**, we synthesize new images to match those pseudo-labels. Finally, a student model ($\boldsymbol{S}$) is trained using the combination of real, pseudo and synthetic training data.

The segmentation losses on the pseudo and synthetic pairs are written as:

$$\mathcal{L}_{pseudo} = \mathcal{L}_{Seg}(f_{s1}(x^u),\ y_2^p) + \mathcal{L}_{Seg}(f_{s2}(x^u),\ y_1^p), \quad (11)$$

$$\mathcal{L}_{synthetic} = \mathcal{L}_{Seg}(f_{s1}(\hat{x}_1),\ y_2^p) + \mathcal{L}_{Seg}(f_{s2}(\hat{x}_2),\ y_1^p). \quad (12)$$

The unsupervised loss $\mathcal{L}_u$ on the unlabeled data is the combination of the losses on both the pseudo and synthetic pairs, which is written as:

$$\mathcal{L}_u = \mathcal{L}_{pseudo} + \mathcal{L}_{synthetic}. \quad (13)$$

The whole training loss on both the labeled and unlabeled data is represented by:

$$\mathcal{L}_{SA-CPS} = \mathcal{L}_s + \lambda \cdot \mathcal{L}_u, \quad (14)$$

where λ is the balanced weight for the loss items.

D. Synthesis-Assisted Self-Training

In this subsection, we introduce the multi-stage method Synthesis-Assisted Self-Training (SA-ST). An illustration of SA-ST is shown in Fig. 4.

A teacher model is first trained, which can be a model trained only using a small amount of labeled data or a trained model from SA-CPS. Then, pseudo-labels can be obtained for unlabeled data, which yields pseudo-labeled data $\mathcal{D}_U^{pseudo} = \{x_i^u, y_i^p\}_{i=1}^N$. Given the pseudo-labels, we can synthesize images to match them, which brings the synthetic data $\mathcal{D}_U^{synthetic} = \{\hat{x}_i, y_i^p\}_{i=1}^N$.

During training, the student model takes a batch of labeled samples $\{x_i^l, y_i\}_{i=1}^B \in \mathcal{D}_L$, pseudo-labeled samples $\{x_i^u, y_i^p\}_{i=1}^B \in \mathcal{D}_U^{pseudo}$ and synthetic samples $\{\hat{x}_i, y_i^p\}_{i=1}^B \in \mathcal{D}_U^{synthetic}$, where B is the batch size. The training losses are represented by:

$$\mathcal{L}_s = \mathcal{L}_{Seg}(f_{\mathcal{S}}(x_i^l),\ y_i), \quad (15)$$

$$\mathcal{L}_{pseudo} = \mathcal{L}_{Seg}(f_{\mathcal{S}}(x_i^u),\ y_i^p), \quad (16)$$

$$\mathcal{L}_{synthetic} = \mathcal{L}_{Seg}(f_{\mathcal{S}}(\hat{x}_i),\ y_i^p). \quad (17)$$

The total loss is the weighted sum of these three losses, which is defined as:

$$\mathcal{L}_{SA-ST} = \mathcal{L}_s + \mu \cdot (\mathcal{L}_{pseudo} + \mathcal{L}_{synthetic}), \quad (18)$$

where μ is a trade-off coefficient.

The entire training procedure is illustrated in Algorithm 1.

Algorithm 1 Synthesis-Assisted Self-Training (SA-ST)

Input: $\mathcal{D}_L = \{x_i^l, y_i\}_{i=1}^N$: labeled training data,
$\mathcal{D}_U = \{x_i^u\}_{i=1}^M$: unlabeled training data,
Output: $\mathcal{S}$: trained segmentation model

1: Train a teacher model $\mathcal{T}$ using all the labeled samples from $\mathcal{D}_L$ or adopt a trained model from SA-CPS as $\mathcal{T}$
2: Generate pseudo-labels on unlabeled samples from $\mathcal{D}_U$, which yields pseudo-labeled data $\mathcal{D}_U^{pseudo}$
3: Synthesize images to match each pseudo-label, which results in synthetic data $\mathcal{D}_U^{synthetic}$
4: Train a student model $\mathcal{S}$ using all the training data, *i.e.*, $\mathcal{D}_{all} = \mathcal{D}_L \cup \mathcal{D}_U^{pseudo} \cup \mathcal{D}_U^{synthetic}$
5: **return** $\mathcal{S}$

IV. Experiments

A. Datasets and Experimental Settings

1) Datasets: We evaluate our method using two public datasets for COVID-19 pneumonia infection segmentation, and their statistics are summarized in Table I :

- COVID-19 Lung CT Lesion Segmentation Challenge-2020 (COVID-19-20) [43] creates the public platform to evaluate emerging methods for segmenting COVID-19 pneumonia regions from CT images. An open-source dataset is provided by the challenge, where 199 and 50 volumes from the dataset are used for training and validation, respectively. Since the challenge is over, we can not access the test data for evaluation. In our experiment, we randomly divide the 249 volumes into: (1) 180 volumes as training set; (2) 19 volumes as validation set; and (3) 50 volumes as testing set. To satisfy the setting of semi-supervised learning, we randomly sample **10%** (18 volumes), **20%** (36 volumes) and **30%** (54 volumes) of CT images in the original training set to construct the labeled training set, and the remaining is used as unlabeled training set.
- MosMedData [44] is a dataset of 1,110 chest CT volumes collected by municipal hospitals in Moscow, Russia. The dataset was originally used for triage, so as to prioritize those patients with severe COVID-19. Among 1,100 CT volumes, 50 CT volumes are annotated with binary masks depicting infection regions, which is a practical scenario suitable for semi-supervised learning. Later on, additional 32 CT volumes with carefully annotated lesions masks are publicly released in [11] to evaluate the performance of COVID-19 pneumonia infection segmentation. In our experiment, MosMedData is divided into: (1) 40 volumes as labeled training set; (2) 1,060 volumes as unlabeled training set; (3) 10 volumes as validation set; and (4) 32 volumes as testing set.

2) Evaluation metrics: Following the evaluation metrics used in the COVID-19-20 challenge, we employ Dice similarity coefficient (DSC) and Normalized surface Dice (NSD) to evaluate the infection segmentation performance. DSC is commonly used in evaluating segmentation performance which is defined as the overlap between the prediction results and

TABLE I
STATISTICS OF TWO DATASETS

Dataset	Training		Validation	Testing
	# Total Volume	# Labeled Volume	# Volume	# Volume
COVID-19-20	180	180	19	50
MosMedData	1,100	40	10	32

ground truths. NSD provides the normalized measure of agreement between the surface of the prediction and the surface of the ground truth at a specified tolerance, and $1\ mm$ is used by default. Python implementations of these two metrics are publicly available from the challenge website.

3) Pre-processing: During pre-processing, the image intensity values of all CT slices are first truncated to the range $[-1250, 250]$, and then normalized to $[0, 255]$. The lung region segmentation is an initial step which extracts left and right lung lobes from the CT slice, and we adopt an automatic model which can accurately segment the lung even under severe pathologies [45]. We ignore the regions outside the lung to reduce computational complexity. Before segmenting the COVID-19 pneumonia regions, we first train the synthesizer using the labeled data, closely following the implementation of SEAN [25].

4) Implementation details: The implementation is based on PyTorch 1.9.0, and all the experiments were conducted on an NVIDIA A100 GPU. The baseline segmentation network follows a 2D U-Net architecture, which is commonly used in medical image segmentation. The input images are resized to 512 × 512. We employ the SGD optimizer to train the models with an initial learning rate of $1e-2$ and a momentum of 0.9. A polynomial learning policy is used during training, where the initial learning rate is multiplied by $(1-\frac{epoch}{max\_epoch})^{power}$ with a power of 0.9. The number of training epochs is set to 20 for COVID-19-20 and 10 for MosMedData. The batch size is 16, consisting of 8 labeled images and 8 unlabeled images. Standard data augmentations such as random flipping and random rotation are adopted to avoid overfitting. The sampling number K is set to 10 for obtaining the reference style code. The coefficients for the loss functions are set to $\lambda = 0.5, \mu = 0.5$. The threshold for binarization when generating the pseudo-labels is set to $\tau = 0.5$ for all experiments.

B. Self Evaluation

In this subsection, we conduct self evaluation experiments on the COVID-19-20 dataset.

1) Effectiveness of SA-CPS and SA-ST: We first analyze the effectiveness of our proposed methods with 10% labeled CT volumes. A more detailed evaluation using different label ratios is shown in Table IV. Fig. 5 shows both our methods outperform the supervised baseline which is trained only using the labeled data. Specifically, the NSD score is increased from 56.17% to 61.52% the DSC score is increased from 60.26% to 65.46% when applying our single-stage Synthesis-Assisted Cross Pseudo Supervision (SA-CPS). As for our multi-stage Synthesis-Assisted Self-Training (SA-ST), the NSD score is increased by 2.74% and the DSC score is increased by 3.26%. SA-ST is a multi-stage method, where pseudo-labels are first

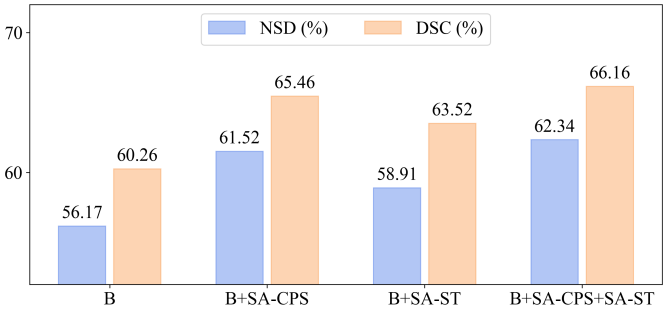

Fig. 5. Effectiveness of the proposed methods with 10% labeled CT volumes on COVID-19-20 test set. **B**: baseline methods which is trained only using the labeled data; **SA-CPS**: synthesis-assisted cross pseudo supervision; **SA-ST**: synthesis-assisted self-training.

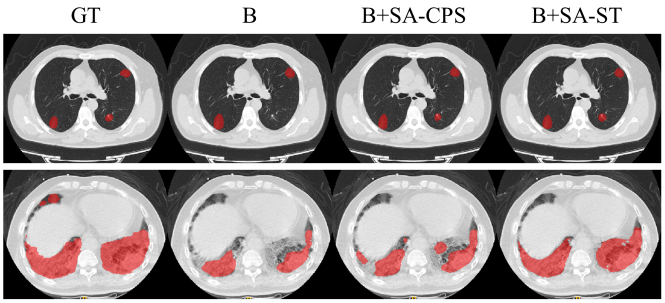

Fig. 6. Visualization results of methods trained with 10% labeled CT volumes on COVID-19-20 test set. **GT**: ground truth. **B**: baseline method trained only using the labeled data; **SA-CPS**: synthesis-assisted cross pseudo supervision; **SA-ST**: synthesis-assisted self-training. The red regions denote the segmented infections.

generated for unlabeled samples using a teacher model. If we use the trained model from SA-CPS as the teacher model for SA-ST, the NSD score is further improved to 62.34% and the DSC score is also improved to 66.16%. The results demonstrate that synthetic images are beneficial for training the models, and both synthesis-assisted methods SA-CPS and SA-ST can improve the segmentation performance.

Fig. 6 shows some visualization results of methods trained with 10% labeled CT volumes. The baseline method achieves less satisfactory performance on the test samples, but the segmentation results can be obviously improved after using both our synthesis-assisted methods. Moreover, the comprehensive evaluation results shown in Table IV show our methods consistently improve the baseline under different labeled ratios.

2) Analysis on Pseudo-Label Guided Image Synthesis : We synthesize new images to match the pseudo-labels of unlabeled samples, and study the performance of these synthetic images. The synthetic image is conditioned on the pseudo-label that describes the infection regions in the desired output image. The realistic image depends on the specified style code from an reference image, where the region-wise style code can be extracted using SEAN [25]. We compare the different methods for generating the reference style codes in Fig. 8. We can find that it achieves the best performance when sampling the reference style code from a style code pool with the style codes of labeled samples which contain infection regions. The main reason is that the pseudo-label inevitably

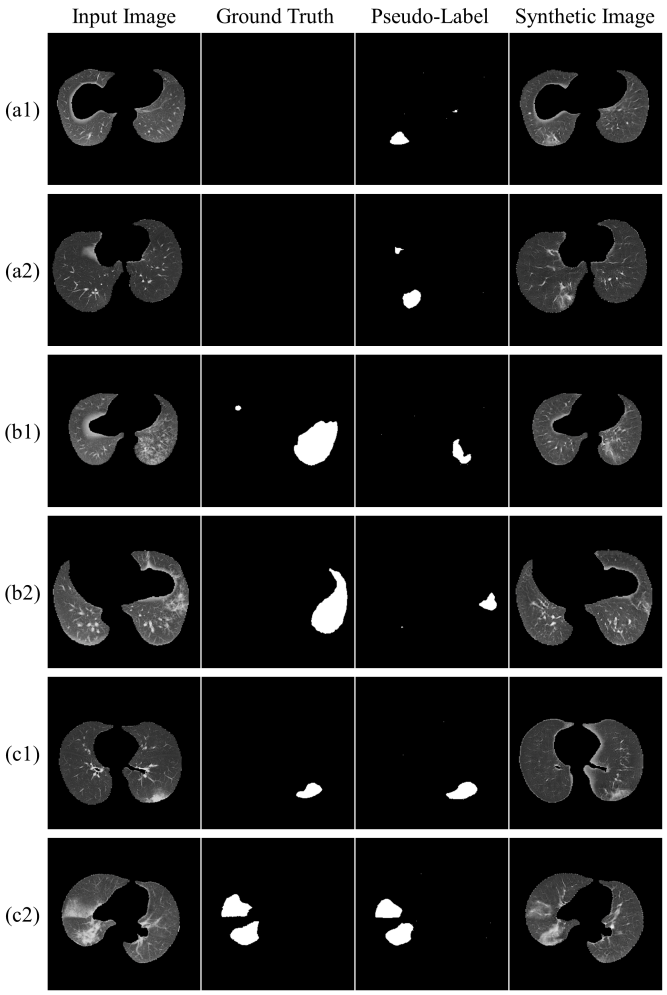

Fig. 7. Visualization results of pseudo-label guided image synthesis. Pseudo-labels are generated for unlabeled images but their quality is not guaranteed, which can be with: (a1-a2) high false positive, (b1-b2) high false negative, (c1-c2) high accuracy. Synthetic images are conditioned on the pseudo-labels that describe the infection regions in the desired output images.

contains low-quality predictions and the extracted region-wise style codes from itself may be inaccurate. Moreover, most of labeled images contain no infections and their style codes are uninformative for synthesizing infection regions. These two alternative options synthesize new images with inferior quality, and result in lower segmentation accuracy.

To better understand the motivation of our proposed pseudo-label guided image synthesis, we provide some visualization results in Fig. 7. The groundtruth labels are not provided for unlabeled samples during training. Pseudo-labels can be generated but their quality is not guaranteed and inevitably contain inaccurate predictions. For example, the input images have no infections but the pseudo-labels falsely predict some infections in Fig. 7(a1-a2), but the synthetic images contain the same COVID-19 infection regions as indicated in the pseudo-labels. In Fig. 7(b1-b2), the predicted infections regions only cover a small part of groundtruth infections regions, and the synthetic images are generated accordingly to match the pseudo-labels. When the predicted pseudo-labels are close to

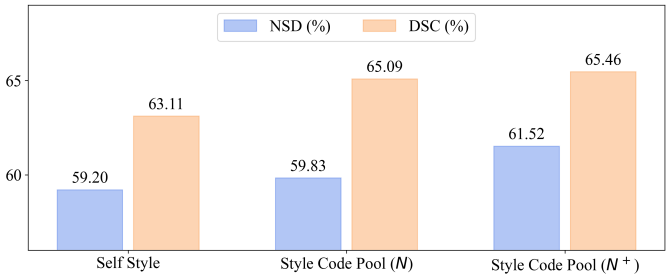

Fig. 8. Comparison of different methods for generating reference style codes. **Self Style**: the reference style code is obtained from the unlabeled image itself based on its pseudo-label; **Style Code Pool (N)**: the reference style code is sampled from a Style Code Pool with the style codes of all labeled samples; **Style Code Pool (N^+)**: the reference style code is sampled from a Style Code Pool with the style codes of labeled samples which contain infection regions.

TABLE II
COMPARISON OF THE SEGMENTATION RESULTS USING DIFFERENT LOSS COMBINATIONS IN SA-CPS.

L_s	L_u		NSD(%)	DSC(%)
	L_{pseudo}	$L_{synthetic}$		
✓	✓		56.83	62.59
✓		✓	58.73	63.59
✓	✓	✓	**61.52**	**65.46**

the ground truths as shown in Fig. 7(c1-c2), the synthetic images also satisfy the requirement.

To conclude, pseudo-label guided image synthesis could control the layout of the generated synthetic image which has the same COVID-19 infection regions as indicated in the pseudo-labels, and add reference styles sampled from a style code pool to make it more realistic. Synthetic images are incorporated into model training and contribute to the segmentation performance.

3) Analysis on SA-CPS: This part evaluates different loss combinations in SA-CPS. In the original CPS method, the pseudo-label from one network is used to supervise the other network for unlabeled data. In SA-CPS, additional synthetic images generated based on the pseudo-labels are used for training. Results shown in Table II demonstrate that both the pseudo and synthetic supervision on unlabeled data could improve the segmentation performance. When both the loss functions of pseudo and synthetic pairs from unlabeled data work together with supervision loss of labeled data, we can achieve the best performance, *i.e.*, the NSD score is 61.52% and the DSC score is 65.46%. The results demonstrate that the synthetic supervision could contribute the segmentation performance. Pseudo-labels inevitable contain low-quality predictions and the training bias may be brought due to the inaccurate supervision on unlabeled samples. However, the synthetic images are generated to match the pseudo-labels and the synthetic supervision is able to reduce the training bias. Moreover, the additional synthetic images behave like expanding the training data, and lead to better segmentation performance. We think these two reasons make our SA-CPS superior to original CPS.

4) Analysis on SA-ST: This part evaluates different input combinations in SA-ST. Self-training (ST) is a classical

TABLE III
COMPARISON OF THE SEGMENTATION RESULTS USING DIFFERENT INPUT COMBINATIONS IN SA-ST.

Real	Pseudo	Synthetic	NSD(%)	DSC(%)
✓	✓		57.50	63.25
✓		✓	58.71	61.99
✓	✓	✓	**58.91**	**63.52**

method for semi-supervised learning, and consists of multiple stages (train a teacher model using labeled data → generate pseudo-labels for unlabeled samples → train a student model using both labeled and unlabeled samples). Our proposed SA-ST adds synthetic images generated from pseudo-labels for training the student model, and the third stage becomes training a student model using labeled, unlabeled and synthetic samples. Results in Table III demonstrate that the additional synthetic images could bring extra gains to the original ST method, where the NSD score is increased to 58.91% and the DSC score is increased to 63.52%. Similar to SA-CPS, we think the synthetic images play two important roles in SA-ST, one is to rectify the training bias caused by inaccurate pseudo-labels and the other one is to expand the training data, thus leading to improved segmentation performance.

C. Comparison with State-of-the-Arts

We compare our methods with the following state-of-the-art methods for semi-supervised medical image segmentation:

- **Uncertainty-Aware Mean-Teacher(UA-MT) [46]**: follows the same spirit of mean teacher, and explores the uncertainty information to enable the student model learn from the meaningful and reliable targets.
- **Cross-Consistency Training(CCT) [17]**: is a cross-consistency based semi-supervised approach for semantic segmentation, where the predictions from the main decoder and auxiliary decoders are forced to be consistent.
- **Uncertainty-guided Dual-Consistency(UDC) [18]**: presents a dual-consistency learning scheme for semi-supervised COVID-19 lesion segmentation, which introduces image transformation equivalence and feature perturbation invariance for leveraging unlabeled data.
- **Self-Ensembling [19]**: presents a co-training framework for semi-supervised COVID-19 CT segmentation, and proposes a self-ensembling consistency regularization technique to alleviate the negative impacts caused by unreliable pseudo-labels from unlabeled samples.
- **Cross Pseudo Supervision(CPS) [16]**: is a consistency regularization approach, which enforces the consistency between the predictions from two segmentation networks perturbed with different initialization weights.
- **SemiInfNet [36]**: is a multi-stage method for semi-supervised COVID-19 infection segmentation, which progressively generates pseudo-labels for unlabeled data with a randomly selected propagation strategy.
- **Self-Training [23]**: is a multi-stage method, where a teacher model trained with labeled data is used to generate pseudo-labels for unlabeled data, then both labeled and unlabeled data are used to train a student model.

TABLE IV
COMPARISON OF COVID-19 PNEUMONIA INFECTION SEGMENTATION RESULTS ON COVID-19-20 TEST SET.

Ratio	Stages	Methods	NSD(%)	DSC(%)
10%	Single	Baseline	56.17	60.26
		UA-MT [46]	57.23	63.65
		CCT [17]	56.93	63.09
		UDC [18]	57.45	61.62
		Self-Ensembling [19]	56.65	61.36
		CPS [16]	56.83	62.59
		SA-CPS [Ours]	**61.52**	**65.46**
	Multiple	SemiInfNet [36]	57.46	63.16
		Self-Training [23]	57.50	63.25
		SA-ST [Ours]	**58.91**	**63.52**
		SA-ST(+SA-CPS) [Ours]	**62.34**	**66.16**
20%	Single	Baseline	62.78	68.67
		UA-MT [46]	63.26	69.07
		CCT [17]	61.84	68.19
		UDC [18]	62.74	68.48
		Self-Ensembling [19]	63.48	69.09
		CPS [16]	63.59	69.64
		SA-CPS [Ours]	**64.64**	**69.90**
	Multiple	SemiInfNet [36]	63.53	69.40
		Self-Training [23]	64.40	69.92
		SA-ST [Ours]	**65.12**	**70.40**
		SA-ST(+SA-CPS) [Ours]	**65.41**	**70.31**
30%	Single	Baseline	64.20	69.80
		UA-MT [46]	64.64	70.37
		CCT [17]	64.30	69.87
		UDC [18]	65.19	70.18
		Self-Ensembling [19]	64.74	70.09
		CPS [16]	65.17	70.39
		SA-CPS [Ours]	**66.92**	**71.31**
	Multiple	SemiInfNet [36]	64.81	69.98
		Self-Training [23]	65.16	70.55
		SA-ST [Ours]	**65.96**	**71.53**
		SA-ST(+SA-CPS) [Ours]	**67.34**	**71.61**
Full Supervision			69.48	73.56

Green and Blue represent the best results for single-stage and multi-stage methods, respectively. Red shows the result of combining our proposed SA-ST and SA-CPS.

Table IV shows the comparison results on COVID-19-20 test set, and we investigate the performance of these methods under different labeled ratios: 10%, 20% and 30%. For fair comparison, we adopt the same 2D U-Net architecture for all the methods. The baseline method is trained only using the labeled data, while 'Full Supervision' represents that all training samples are labeled and can be regarded as the upper bound. It can be observed that almost all the methods are superior to the baseline method by leveraging unlabeled data, which verifies the value of unlabeled data. Our methods achieve the best performance among all the competing methods, which demonstrates pseudo-label guided image synthesis is beneficial, no matter in the single-stage method or multi-stage method. The superiority is obvious, especially when only a small amount of images are labeled. If we take the trained model from SA-CPS as the teacher model in SA-ST, their combination could bring extra improvement. When only 10% CT volumes are labeled, the NSD score is increased by

6.17% and the DSC score is increased by 5.9%. When 30% CT volumes are labeled, the result is comparable to the fully supervised upper bound. Existing methods proposed novel techniques to avoid the negative impacts of imperfect pseudo-labels, which require sophisticated designs for assessing and refining pseudo-labels. Unlike these methods, our proposed pseudo-label guided image synthesis is a new perspective on semi-supervised learning for COVID-19 pneumonia infection segmentation, which is more generic and efficient.

D. Application to Large-Scale Unlabeled Data

The goal of semi-supervised learning is to achieve good performance by leveraging unlabeled data. In this section, we would like to study the performance of our method when the unlabeled data is abundant but annotation is limited. We use the MosMedData dataset due to its relatively large scale of unlabeled data, where around 1000 CT volumes are unlabeled.

The results on the MosMedData test set are shown in Table V, where state-of-the-art semi-supervised segmentation methods are included for comparison. It is observed that the segmentation performance is improved by leveraging the unlabeled data in all the semi-supervised learning methods. Our methods consistently achieve better performance compared to other competing methods. The combination of SA-CPS and SA-ST can bring obvious improvements over the baseline, where the NSD score is increased from 60.92% to 65.67%, and the DSC score is increased from 61.83% to 65.64%. The experiment results further demonstrate the superiority of our method under semi-supervised setting when encountering large-scale unlabeled data. Note that MosMedData test set was collected at inpatient clinics, but the MosMedData training set was collected from the Moscow out-patient clinics database created from two to six weeks later, the potential domain shift problem could bring challenges to the segmentation task on the test samples. We will try to combine some domain adaptation techniques in the future work.

We show some visualization results of COVID-19 pneumonia infection segmentation results on MosMedData test set in Fig. 9. We can observe that the baseline model only trained with a small amount of labeled data can bring low-quality prediction results, where many regions are mis-segmented. With the assistance from synthetic images, we can find that the segmentation results become more accurate.

TABLE V
COMPARISON OF COVID-19 PNEUMONIA INFECTION SEGMENTATION RESULTS ON MOSMEDDATA TEST SET.

Stages	Methods	NSD(%)	DSC(%)
Single	Baseline	60.92	61.83
	UA-MT [46]	63.62	62.39
	CCT [17]	63.09	62.42
	UDC [18]	63.97	63.06
	Self-Ensembling [19]	63.80	62.95
	CPS [16]	63.62	63.56
	SA-CPS [Ours]	**64.90**	**64.32**
Multiple	SemiInfNet [36]	62.10	62.33
	Self-Training [23]	62.99	62.91
	SA-ST [Ours]	**64.69**	**64.82**
	SA-ST(+SA-CPS) [Ours]	**65.67**	**65.64**

Green and Blue represent the best results for single-stage and multi-stage methods, respectively. Red shows the result of combining our proposed SA-ST and SA-CPS.

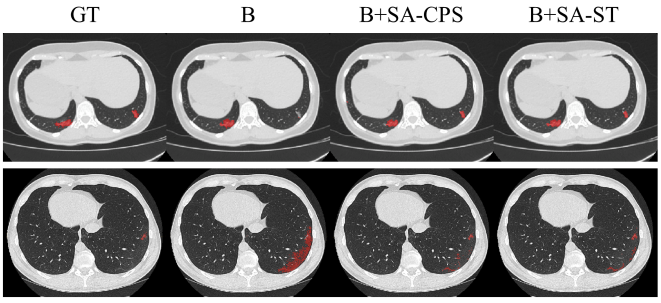

Fig. 9. Visualization results of COVID-19 pneumonia infection segmentation results on MosMedData test set. **GT**: ground truth. **B**: baseline method trained only using the labeled data; **SA-CPS**: synthesis-assisted cross pseudo supervision; **SA-ST**: synthesis-assisted self-training. The red regions denote the segmented infections.

V. CONCLUSION

In this paper, we propose a new perspective on semi-supervised learning for COVID-19 pneumonia infection segmentation, namely pseudo-label guided image synthesis, which aims to train a model using limited labeled data. Our approach keeps the pseudo-labels and synthesizes new images to match them. We introduce two novel representative methods by incorporating the synthetic images into model training, including Synthesis-Assisted Cross Pseudo Supervision and Synthesis-Assisted Self-Training. Extensive experiments on two public datasets have verified the effectiveness of our methods. In real-world situations, it is impractical to collect large-scale well-labeled datasets due to the urgent nature of pandemics. Our method is a generic semi-supervised learning method, and we believe it will enable new avenues of research into label-efficient learning in medical applications, such as fighting against new pandemics or rare diseases.

REFERENCES

[1] A. Clark, M. Jit, C. Warren-Gash, B. Guthrie, H. H. Wang, S. W. Mercer, C. Sanderson, M. McKee, C. Troeger, K. L. Ong *et al.*, "Global, regional, and national estimates of the population at increased risk of severe covid-19 due to underlying health conditions in 2020: a modelling study," *Lancet Glob. Heal.*, vol. 8, no. 8, pp. e1003–e1017, 2020.

[2] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu *et al.*, "A novel coronavirus from patients with pneumonia in china, 2019," *N. Engl. J. Med.*, 2020.

[3] S. K. Brooks, R. K. Webster, L. E. Smith, L. Woodland, S. Wessely, N. Greenberg, and G. J. Rubin, "The psychological impact of quarantine and how to reduce it: rapid review of the evidence," *The lancet*, vol. 395, no. 10227, pp. 912–920, 2020.

[4] R. M. Barber, R. J. Sorensen, D. M. Pigott, C. Bisignano, A. Carter, J. O. Amlag, J. K. Collins, C. Abbafati, C. Adolph, A. Allorant *et al.*, "Estimating global, regional, and national daily and cumulative infections with sars-cov-2 through nov 14, 2021: a statistical analysis," *The Lancet*, 2022.

[5] H. Wang, K. R. Paulson, S. A. Pease, S. Watson, H. Comfort, P. Zheng, A. Y. Aravkin, C. Bisignano, R. M. Barber, T. Alam *et al.*, "Estimating excess mortality due to the covid-19 pandemic: a systematic analysis of covid-19-related mortality, 2020–21," *The Lancet*, vol. 399, no. 10334, pp. 1513–1536, 2022.

[6] L. S. Canas, C. H. Sudre, J. C. Pujol, L. Polidori, B. Murray, E. Molteni, M. S. Graham, K. Klaser, M. Antonelli, S. Berry *et al.*, "Early detection of covid-19 in the uk using self-reported symptoms: a large-scale, prospective, epidemiological surveillance study," *Lancet Digit. Health*, vol. 3, no. 9, pp. e587–e598, 2021.

[7] 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 for coronavirus disease 2019 (covid-19) in china: a report of 1014 cases," *Radiology*, vol. 296, no. 2, pp. E32–E40, 2020.

[8] X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, and J. Liu, "Chest ct for typical coronavirus disease 2019 (covid-19) pneumonia: relationship to negative rt-pcr testing," *Radiology*, vol. 296, no. 2, pp. E41–E45, 2020.

[9] G. D. Rubin, C. J. Ryerson, L. B. Haramati, N. Sverzellati, J. P. Kanne, S. Raoof, N. W. Schluger, A. Volpi, J.-J. Yim, I. B. Martin *et al.*, "The role of chest imaging in patient management during the covid-19 pandemic: a multinational consensus statement from the fleischner society," *Radiology*, vol. 296, no. 1, pp. 172–180, 2020.

[10] 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*, vol. 181, no. 6, pp. 1423–1433, 2020.

[11] M. Goncharov, M. Pisov, A. Shevtsov, B. Shirokikh, A. Kurmukov, I. Blokhin, V. Chernina, A. Solovev, V. Gombolevskiy, S. Morozov *et al.*, "Ct-based covid-19 triage: Deep multitask learning improves joint identification and severity quantification," *Med. Image Anal.*, vol. 71, p. 102054, 2021.

[12] A. Amer, X. Ye, and F. Janan, "Residual dilated u-net for the segmentation of covid-19 infection from ct images," in *Proc. ICCV*, 2021, pp. 462–470.

[13] K. Gao, J. Su, Z. Jiang, L.-L. Zeng, Z. Feng, H. Shen, P. Rong, X. Xu, J. Qin, Y. Yang *et al.*, "Dual-branch combination network (dcn): Towards accurate diagnosis and lesion segmentation of covid-19 using ct images," *Med. Image Anal.*, vol. 67, p. 101836, 2021.

[14] L. Zhou, Z. Li, J. Zhou, H. Li, Y. Chen, Y. Huang, D. Xie, L. Zhao, M. Fan, S. Hashmi *et al.*, "A rapid, accurate and machine-agnostic segmentation and quantification method for ct-based covid-19 diagnosis," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2638–2652, 2020.

[15] J. Ma, Y. Wang, X. An, C. Ge, Z. Yu, J. Chen, Q. Zhu, G. Dong, J. He, Z. He *et al.*, "Toward data-efficient learning: A benchmark for covid-19 ct lung and infection segmentation," *Med. Phys.*, vol. 48, no. 3, pp. 1197–1210, 2021.

[16] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proc. CVPR*, 2021, pp. 2613–2622.

[17] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 12 674–12 684.

[18] Y. Li, L. Luo, H. Lin, H. Chen, and P.-A. Heng, "Dual-consistency semi-supervised learning with uncertainty quantification for covid-19 lesion segmentation from ct images," in *Proc. MICCAI*. Springer, 2021, pp. 199–209.

[19] C. Li, L. Dong, Q. Dou, F. Lin, K. Zhang, Z. Feng, W. Si, X. Deng, Z. Deng, and P.-A. Heng, "Self-ensembling co-training framework for semi-supervised covid-19 ct segmentation," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 11, pp. 4140–4151, 2021.

[20] N. Souly, C. Spampinato, and M. Shah, "Semi supervised semantic segmentation using generative adversarial network," in *Proc. ICCV*, 2017, pp. 5688–5696.

[21] Y. Zhang, L. Yang, J. Chen, M. Fredericksen, D. P. Hughes, and D. Z. Chen, "Deep adversarial networks for biomedical image segmentation utilizing unannotated images," in *Proc. MICCAI*. Springer, 2017, pp. 408–416.

[22] Y. Zhu, Z. Zhang, C. Wu, Z. Zhang, T. He, H. Zhang, R. Manmatha, M. Li, and A. J. Smola, "Improving semantic segmentation via efficient self-training," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021.

[23] B. Zoph, G. Ghiasi, T.-Y. Lin, Y. Cui, H. Liu, E. D. Cubuk, and Q. Le, "Rethinking pre-training and self-training," *Proc. NIPS*, vol. 33, pp. 3833–3845, 2020.

[24] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proc. CVPR*, 2019, pp. 2337–2346.

[25] P. Zhu, R. Abdal, Y. Qin, and P. Wonka, "Sean: Image synthesis with semantic region-adaptive normalization," in *Proc. CVPR*, 2020, pp. 5104–5113.

[26] T. Kitrungrotsakul, Q. Chen, H. Wu, Y. Iwamoto, H. Hu, W. Zhu, C. Chen, F. Xu, Y. Zhou, L. Lin *et al.*, "Attention-refnet: Interactive attention refinement network for infected area segmentation of covid-19," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 7, pp. 2363–2373, 2021.

[27] H. Hu, L. Shen, Q. Guan, X. Li, Q. Zhou, and S. Ruan, "Deep co-supervision and attention fusion strategy for automatic covid-19 lung infection segmentation on ct images," *Pattern Recognit.*, vol. 124, p. 108452, 2022.

[28] Y.-H. Wu, S.-H. Gao, J. Mei, J. Xu, D.-P. Fan, R.-G. Zhang, and M.-M. Cheng, "Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation," *IEEE Trans. Image Process.*, vol. 30, pp. 3113–3126, 2021.

[29] X. Wang, L. Jiang, L. Li, M. Xu, X. Deng, L. Dai, X. Xu, T. Li, Y. Guo, Z. Wang *et al.*, "Joint learning of 3d lesion segmentation and classification for explainable covid-19 diagnosis," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2463–2476, 2021.

[30] J. Liu, B. Dong, S. Wang, H. Cui, D.-P. Fan, J. Ma, and G. Chen, "Covid-19 lung infection segmentation with a novel two-stage cross-domain transfer learning framework," *Med. Image Anal.*, vol. 74, p. 102205, 2021.

[31] I. Laradji, P. Rodriguez, O. Manas, K. Lensink, M. Law, L. Kurzman, W. Parker, D. Vazquez, and D. Nowrouzezahrai, "A weakly supervised consistency-based learning method for covid-19 segmentation in ct images," in *Proc. WACV*, 2021, pp. 2453–2462.

[32] X. Liu, Q. Yuan, Y. Gao, K. He, S. Wang, X. Tang, J. Tang, and D. Shen, "Weakly supervised segmentation of covid19 infection with scribble annotation on ct images," *Pattern Recognit.*, vol. 122, p. 108341, 2022.

[33] G. Wang, X. Liu, C. Li, Z. Xu, J. Ruan, H. Zhu, T. Meng, K. Li, N. Huang, and S. Zhang, "A noise-robust framework for automatic segmentation of covid-19 pneumonia lesions from ct images," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2653–2663, 2020.

[34] X. Wang, Y. Yuan, D. Guo, X. Huang, Y. Cui, M. Xia, Z. Wang, C. Bai, and S. Chen, "Ssa-net: Spatial self-attention network for covid-19 pneumonia infection segmentation with semi-supervised few-shot learning," *Med. Image Anal.*, vol. 79, p. 102459, 2022.

[35] D. Yang, Z. Xu, W. Li, A. Myronenko, H. R. Roth, S. Harmon, S. Xu, B. Turkbey, E. Turkbey, X. Wang *et al.*, "Federated semi-supervised learning for covid region segmentation in chest ct using multi-national data from china, italy, japan," *Med. Image Anal.*, vol. 70, p. 101992, 2021.

[36] 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 Trans. Med. Imag.*, vol. 39, no. 8, pp. 2626–2637, 2020.

[37] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. CVPR*, 2017, pp. 1125–1134.

[38] 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 *Proc. CVPR*, 2018, pp. 8798–8807.

[39] Y. Jiang, H. Chen, M. Loew, and H. Ko, "Covid-19 ct image synthesis with a conditional generative adversarial network," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 2, pp. 441–452, 2020.

[40] S. Thermos, X. Liu, A. O'Neil, and S. A. Tsaftaris, "Controllable cardiac synthesis via disentangled anatomy arithmetic," in *Proc. MICCAI*. Springer, 2021, pp. 160–170.

[41] Q. Wang, X. Zhang, W. Zhang, M. Gao, S. Huang, J. Wang, J. Zhang, D. Yang, and C. Liu, "Realistic lung nodule synthesis with multi-target co-guided adversarial mechanism," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2343–2353, 2021.

[42] K. Li, L. Yu, and P.-A. Heng, "Towards reliable cardiac image segmentation: Assessing image-level and pixel-level segmentation quality via self-reflective references," *Med. Image Anal.*, vol. 78, p. 102426, 2022.

[43] H. Roth, Z. Xu, C. T. Diez, and et al., "Rapid artificial intelligence solutions in a pandemic - the covid-19-20 lung ct lesion segmentation challenge," *Research square*, p. rs.3.rs—571332, June 2021. [Online]. Available: https://europepmc.org/articles/PMC8183044

[44] S. P. Morozov, A. E. Andreychenko, I. A. Blokhin, P. B. Gelezhe, A. P. Gonchar, A. E. Nikolaev, N. A. Pavlov, V. Y. Chernina, and V. A. Gombolevskiy, "Mosmeddata: data set of 1110 chest ct scans performed during the covid-19 epidemic," *Digital Diagnostics*, vol. 1, no. 1, pp. 49–59, 2020.

[45] J. Hofmanninger, F. Prayer, J. Pan, S. Röhrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *European Radiology Experimental*, vol. 4, no. 1, pp. 1–13, 2020.

[46] L. Yu, S. Wang, X. Li, C.-W. Fu, and P.-A. Heng, "Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation," in *Proc. MICCAI*. Springer, 2019, pp. 605–613.