Dual Multiscale Mean Teacher Network for Semi-Supervised Infection Segmentation in Chest CT Volume for COVID-19

Liansheng Wang, *Member, IEEE*, Jiacheng Wang, *Student Member, IEEE*, Lei Zhu, Huazhu Fu, *Senior Member, IEEE*, Ping Li, *Member, IEEE*, Gary Cheng, Zhipeng Feng, Shuo Li, *Senior Member, IEEE*, and Pheng-Ann Heng, *Senior Member, IEEE*

***Abstract*—Automated detecting lung infections from computed tomography (CT) data plays an important role for combating coronavirus 2019 (COVID-19). However, there are still some challenges for developing AI system: 1) most current COVID-19 infection segmentation methods mainly relied on 2-D CT images, which lack 3-D sequential constraint; 2) existing 3-D CT segmentation methods focus on single-scale representations, which do not achieve the multiple level receptive field sizes on 3-D volume; and 3) the emergent breaking out of COVID-19 makes it hard to annotate sufficient CT volumes for training deep model. To address these issues, we first build a multiple dimensional-attention convolutional neural network (MDA-CNN) to aggregate multiscale information along different dimension of input feature maps and impose supervision on multiple predictions from different convolutional neural networks (CNNs) layers. Second, we assign this MDA-CNN as a basic network into a novel dual multiscale mean teacher network (DM^{2}T-Net) for semi-supervised COVID-19 lung infection segmentation on CT volumes by leveraging unlabeled data and exploring the multiscale information. Our DM^{2}T-Net encourages multiple predictions at different CNN layers from the student and teacher networks to be consistent for computing a multiscale consistency loss on unlabeled data, which is then added to the supervised loss on the labeled data from multiple predictions of MDA-CNN. Third, we collect two COVID-19 segmentation datasets to evaluate our method. The experimental results show that our network consistently outperforms the compared state-of-the-art methods.**

***Index Terms*—Chest computed tomography (CT), coronavirus 2019 (COVID-19), infection segmentation, semi-supervised learning.**

I. Introduction

As an ongoing pandemic, novel coronavirus 2019 (COVID-19) has infected about 6 898 613 cases and incurred 399 832 deaths in the world by 7 June 2020. Reverse transcription-polymerase chain reaction (RT-PCR) test is considered as the gold standard of screening COVID-19. However, RT-PCR testing is time-consuming and requires repeated testing for accurate confirmation of a COVID-19 case due to its low sensitivity, thereby resulting in the ineffectiveness of timely confirming CVOID-19 patients. By working as a complement to RT-PCR, easily accessible imaging equipments [e.g., chest X-ray and computed tomography (CT)] have provided huge assistance to clinicians in both current diagnosis and follow-up assessment of disease evolution [1], [2], [3]. Further, quantitative CT information (e.g., lung burden, percentage of high opacity, and lung severity score) are used widely to monitor disease progression and understand the course of COVID-19 [4], [5], [6]. In clinical practice, CT screening is usually more preferred since typical infection signs can be observed from CT data, covering ground-glass opacity (GGO) in the early stage to pulmonary consolidation in the late stage. Moreover, the qualitative evaluation of infection and longitudinal changes in CT images could thus provide useful and important information in combating COVID-19.

Accurate segmentation of the COVID-19 infected region plays a crucial role in achieving a reliable quantification of infection in chest CT images. However, the manual delineation of lung infections is tedious, labor-consuming, and expensive for radiologists. Also, annotating the lung infections in CT is a challenging task due to highly variant textures, sizes and positions of infected regions, as well as low contrast and blurred GGO boundaries [7]. Recently, convolutional neural networks (CNNs) have been developed to automatically

Manuscript received 17 December 2021; revised 26 October 2022; accepted 10 November 2022. Date of publication 14 December 2022; date of current version 16 September 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFE0113900; in part by the National Natural Science Foundation of China under Project 61902275; and in part by the AI Singapore Tech Challenge (Open-Theme) Funding under Grant AISG2-TC-2021-003. This article was recommended by Associate Editor P. Shi. *(Corresponding author: Lei Zhu.)*

Liansheng Wang and Jiacheng Wang are with the Department of Computer Science, School of Informatics, Xiamen University, Xiamen 361005, China.

Lei Zhu is with the ROAS Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511400, Guangdong, China, and also with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong, SAR, China (e-mail: leizhu@ust.hk).

Huazhu Fu is with the Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore 138632 (e-mail: hzfu@ieee.org).

Ping Li is with the Department of Computing and the School of Design, The Hong Kong Polytechnic University, Hong Kong (e-mail: p.li@polyu.edu.hk).

Gary Cheng is with the Department of Mathematics and Information Technology, The Education University of Hong Kong, Hong Kong.

Zhipeng Feng is with the Zhongshan Hospital, Xiamen University, Xiamen 361005, China.

Shuo Li is with the Department of Medical Biophysics, Western University, London, ON N6A 4V2, Canada.

Pheng-Ann Heng is with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2022.3223528.

Digital Object Identifier 10.1109/TCYB.2022.3223528

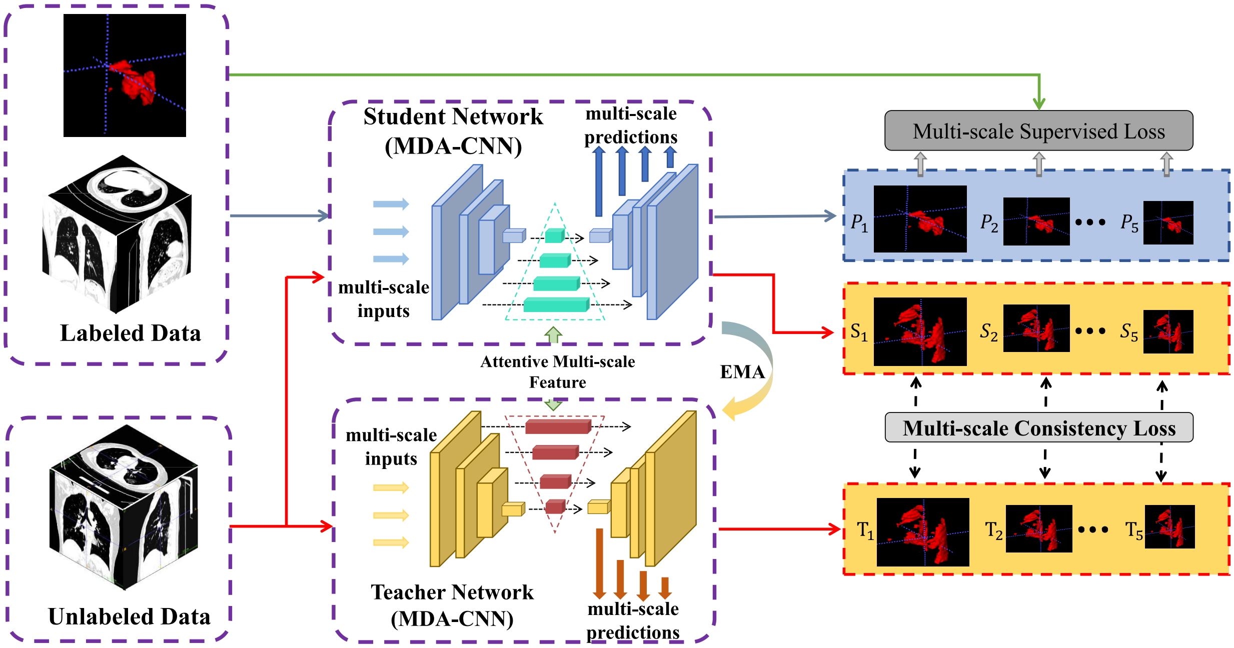

Fig. 1. Schematic illustration of the developed DM^{2}T-Net. We first develop an MDA-CNN (see Fig. 2) to detect infected regions. The MDA-CNN produces five segmentation results from different CNN layers. After that, we compute a multiscale supervised loss on the five segmentation results (i.e., P_1 to P_5) of unlabeled data. For unlabeled data, we compute a multiscale supervised loss on two pairs of five results from the student network (i.e., S_1 to S_5) and the teacher network (i.e., T_1 to T_5). Finally, we combine the supervised loss and consistency loss to train our 3-D COVID-19 infection segmentation network.

segment the lung infections from CT images [3], [7], [8], [9], [10], [11], [12]. Such automatic segmentation tools contribute to the quantification of lung infections and can eliminate the possibility of subjective impact. Moreover, for saving medical resources and accelerating daily diagnosis of the overburdened hospitals under the COVID-19's large-scale outbreak, building artificial intelligence (AI) system will be greatly helpful. However, there are still some challenges for developing AI system.

1) Most current COVID-19 infection segmentation methods mainly relied on 2-D CT images, resulting in a lower segmentation accuracy due to lack of inter-relations among different 2-D images of raw clinical CT volume. Moreover, infected regions of a 3-D volume only exist in a few 2-D images and using 2-D images to train a segmentation network tends to produce many false positives in these images without any lung infection. By contrast, the 3-D context makes a meaningful difference in exploring inter-relation between continuous slices and determining the infection regions from features of more dimensions so as to increase the segmentation accuracy.
2) The 3-D volume segmentation methods [13], [14] only exploited the single scale of the input 3-D CT volume, which do not consider the multiple level receptive field sizes on 3-D volume.
3) It is difficult to collect sufficient high-quality labeled 3-D CT data within a short time for training a deep model, which limits the developing of deep 3-D segmentation models.

To address above issues, we propose a dual multiscale mean teacher network (DM^{2}T-Net) for boosting the 3-D COVID-19 lung infection segmentation performance. As shown in Fig. 1, our DM^{2}T-Net utilizes two kinds of multiscale structures. One is the multiple dimensional-attention for learning a hierarchical representations of 3-D volume, while another is multiscale consistency loss for constraining the semi-supervised learning. Specifically, a novel multiple dimensional-attention CNN (MDA-CNN) is designed as the basic network for both teacher and student networks of DM^{2}T-Net. The proposed MDA-CNN attentively integrates multiple scales of the input 3-D volumes along different dimensions for segmenting lung infections, simultaneously, produces multiple side-outputs at different CNN layers. Two kinds of loss are employed to constrain our DM^{2}T-Net. One is multiscale supervised loss in the student network for labeled data to integrate the deeply supervised side-outputs. The second is multiscale consistency loss for the unlabeled data to encourage multiple predictions at different CNN layers from the student and teacher networks to be consistent. Overall, the main contributions are summarized as follows.

1) We develop an MDA-CNN for 3-D lung infection segmentation, which attentively aggregated CNN features extracted from multiple dimensional-scale information of the input 3-D data and generates multiple predictions at different CNN layers.
2) We propose a DM^{2}T-Net for leveraging the unlabeled data. A multiscale consistency loss is devised on the side-output predictions to encourage the predictions consistent on both intermodel and intramodel. As a semi-supervised learning model, our framework has the potential to be used for other 3-D segmentation tasks.
3) Moreover, we collect two COVID-19 segmentation datasets to evaluate our method. The experiments show that our proposed network outperforms state-of-the-art

methods on both supervised and semi-supervised manners. We have released the code, trained models, and collected unlabeled 3-D CT data at https://github.com/jcwang123/DM2TNet.

II. Related Work

This section reviews three kinds of works that are most related to our method, including segmentation in chest CT, semi-supervised learning, and data-driven methods for COVID-19.

A. Segmentation in Chest CT

Segmenting organs and tumors from chest CT images provides crucial information for clinicians to diagnose and quantify lung diseases [15], [16]. Early data-driven algorithms converted the lung segmentation task into voxel classification and then applied different classification models with manually designed features for segmenting target regions. Wu et al. [17] designed a set of texture and shape features to represent voxels and then trained conditional random field (CRF) model to classify voxels for lung segmentation. Keshani et al. [18] segmented lung nodules by the support vector machine (SVM) classifier with 2-D stochastic and 3-D anatomical features. However, relying on these hand-crafted features is difficult to segment nodules due to similar appearances of nodules and background details. Motivated by the outstanding performance of CNNs in medical image analysis [19], [20], [21], [22], [23], [24], [25], deep learning-based methods have been introduced to learn discriminative representation for lung nodule detection from CT images. Wang et al. [26] formulated a central focused CNN to capture both 2-D and 3-D lung nodule features for identifying lung nodules from heterogeneous CT images. Jin et al. [27] developed a conditional GAN model to generate CT-realistic high-quality 3-D lung nodules and utilized these synthesized data to enhance the pathological lung segmentation model [28]. Jiang et al [29] simultaneously leveraged features across multiple image resolution and CNN feature levels via residual connections to identify the lung tumors.

B. Semi-Supervised Learning

Annotations in large-scale medical data are tedious, time-consuming, and difficult to obtain. More and more researchers have shifted their attentions from supervised fashion to the semi-supervised learning, which improves the model performance by combining limited labeled data and sufficient unlabeled data [30]. From a high-level view, these semi-supervised learning methods devised an objective function, which consists of supervised loss on labeled data and unsupervised learning on unlabeled data or both the labeled data and unlabeled data. Lee [31] picked up the class which has the maximum predicted probability and used this class as the pseudo-labels for unlabeled data. Bai et al. [32] alternately updated the network parameters and the segmentation on unlabeled data in a semi-supervised learning framework. Based on an adversarial learning-based semi-supervised fashion [33], [34], Zhang et al. [35] encouraged the segmentation of unlabeled images to be similar to those of the labeled ones in a deep adversarial network. Yu et al. [36] estimated an uncertainty information as a guidance to eliminate unreliable predictions and maintain only the reliable ones (low uncertainty) when devising the consistency loss of student and teacher network predictions for labeled and unlabeled data.

C. AI Techniques for COVID-19

AI methods, especially deep learning techniques, have been employed widely in medical imaging applications against COVID-19 [1], [37]. Tang et al. [38] calculated quantitative features from chest CT images and then passed these features into to train a random forest model for COVID-19 severity assessment. Wang et al. [39] modified an inception transfer-learning model for the identification of viral pneumonia images. Chen et al. [41] first collected 46 096 image slices from 106 admitted COVID-19 patients, and then trained U-Net++ [40] to extract valid areas and detect suspicious lesions in CT images. Song et al. [42] formulated a detail relation extraction neural network (DRE-Net) to extract top-K details and obtain the image-level predictions for patient-level diagnoses. Xie et al. [11] presented a relational two-stage U-Net to segment pulmonary lobes in CT images by introducing a nonlocal neural network module to model the global structured relationships. Chen et al. [7] exploited both the residual network and attention mechanism to improve the efficacy of the U-Net for the lung CT image segmentation. Wu et al. [43] created a COVID-19 dataset with 3855 labeled CT images and performed a joint explainable classification and accurate lesion segmentation. Qiu et al. [10] presented a lightweight deep learning model for efficient COVID-19 image segmentation. Observing that the boundary of the infected region can be enhanced by adjusting the global intensity, Qiu et al. [10] introduced a deep CNN with feature variation block, which adaptively adjusted the global properties of the features for segmenting COVID-19 infection in 2-D images. Zhou et al. [8] incorporated a spatial and channel attentions to U-Net model for capturing richer contextual relationships for COVID-19 CT segmentation. He et al. [44] first used a set of 2-D image patches to represent each CT image and then developed a multitask multi-instance to assess the severity of COVID-19 patients and segment the lung lobe simultaneously. Fan et al. [3] devised an Inf-Net to utilize an implicit reverse attention and explicit edge-attention for the identification of infected regions and then further enhance the segmentation performance by embedding the Inf-Net into a semi-supervised learning strategy. Ding et al. [45] proposed the MT-nCov-Net to formulate 2-D lesion segmentation as a multitask shape regression problem that enables the feature fusion between various tasks. Pang et al. [46] proposed a novel group equivariant segmentation framework by encoding the 2-D inherent symmetries, that is, rotations and reflections, for learning more precise representations. Methods above almost relied on 2-D images to train CNNs for the classification or segmentation. Ma et al. [13] created a COVID-19 3-D CT dataset with 20 cases and explored the U-Net [47] for 3-D lung and infection segmentation.

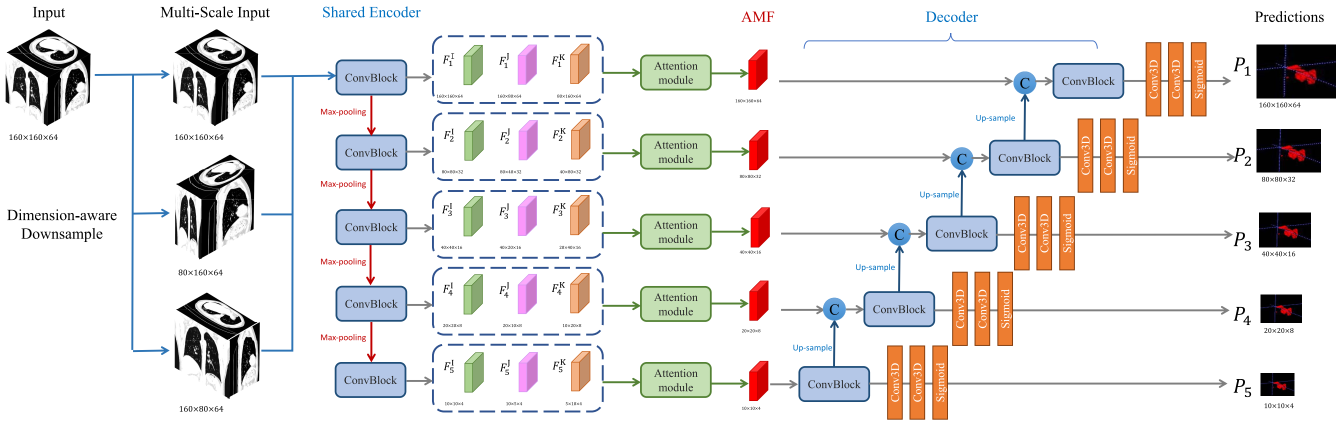

Fig. 2. Schematic illustration of our MDA-CNN. Given an input 3-D CT volume, MDA-CNN first down-samples it along two spatial dimensions and then passes the three 3-D volumes to obtain a set of CNN features (F_i^J) with different spatial resolutions. Then, the attention modules are introduced after each CNN layer to attentively aggregate features (AMF_i) from three volumes and these aggregated features are then adjacently merged. Finally, multiple side-outputs (P_i) are produced from different decoder layers. Note that we empirically take P_1 as the output segmentation result of MDA-CNN.

III. Proposed Method

Fig. 1 shows the schematic illustration of the proposed DM^{2}T-Net which integrates dual multiscale details of labeled data and unlabeled data for COVID-19 lung infection segmentation. In our DM^{2}T-Net, an MDA-CNN is developed to attentively aggregate CNN features extracted from multiple dimensional-scale information and produce multiple predictions from different CNN layers. This MDA-CNN is then integrated into DM^{2}T-Net as the basic network for both the student and the teacher networks. In the training stage, the labeled data is fed into the student network, and a multiscale supervised loss is calculated to constrain the consistency of multiple side-outputs on intramodel. Then, the unlabeled data is inputted into the both student and teacher networks, respectively. Meanwhile, a multiscale consistency loss is devised on the two groups of side-outputs to encourage the predictions consistent on intermodel.

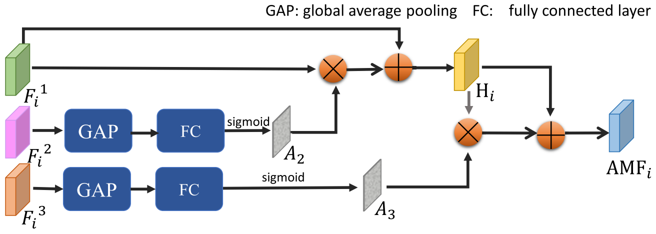

Fig. 3. Schematic illustration of the developed attention module of Fig. 2.

A. Multiple Dimensional-Attention CNN

Although achieving remarkable results, existing 3-D segmentation networks produce unsatisfactory results when detecting 3-D lung infected regions, since only single-scale information of input volume is considered. To address this issue, we argue that exploring multiscale information is helpful to boost lung infection segmentation in 3-D CT volume. In this article, we propose an MDA-CNN to model and fuse the complementary of multiple dimensional-scale details within a single network, as shown in Fig. 2. Given a 3-D lung CT volume ($\mathcal{I}$), we first generate another two auxiliary volumes (denoted as $\mathcal{J}$ and $\mathcal{K}$) by downsampling $\mathcal{I}$ along the two spatial dimensions. Then, we pass the three volumes into several convolutional blocks and obtain multiple feature maps with five different spatial resolutions for each volume ($\mathcal{I}$, $\mathcal{J}$, and $\mathcal{K}$). We use ${F_i}^{\mathcal{I}}$, ${F_i}^{\mathcal{J}}$, and ${F_i}^{\mathcal{K}}$ to denote the three features at the ith CNN layer (see Fig. 2). After that, we develop an attention module in each CNN layer to learn attention maps for aggregating the features with multiple dimension-aware scale information. By doing so, the dimension-aware multiscale representations of the input volume are well modeled and fused together for segmenting COVID-19 infected lung areas.

Fig. 3 shows the schematic illustration of the attention module at the ith CNN layer. It takes three feature maps (${F_i}^{\mathcal{I}}$, ${F_i}^{\mathcal{J}}$, and ${F_i}^{\mathcal{K}}$) from $\mathcal{I}$, $\mathcal{J}$, and $\mathcal{K}$ at the ith CNN layer and outputs a new feature map (AMF_i) to attentively aggregate the input features. The attention module starts by using one average pooling operation (GAP) on ${F_i}^{\mathcal{J}}$, one fully connected (FC) layer, and one Sigmoid activation function to obtain an attention map (A_2). Then, we multiply the attention map A_2 with ${F_i}^{\mathcal{I}}$, and the resultant features are then elementwisely added with ${F_i}^{\mathcal{I}}$ to obtain a refinement of ${F_i}^{\mathcal{I}}$, which is denoted as H_i. Similarly, we further employ an average pooling operation on ${F_i}^{\mathcal{K}}$ to obtain another attention map (A_3), which is then multiplied with H_i. We then elementwisely add the resultant feature map to H_i to obtain attentive multiscale features (denoted as AMF_i), which is taken as the output of the developed attention module. After the encoder layers and attention modules, we obtain five attentive feature maps (denoted as AMF_i) with different scales $i = 1, \ldots, 5$, which are then adjacently merged for decoding them. Then, we use three 3×3 CNN layers, one 1×1 CNN layer, and one Sigmoid activation layer on each decoder output to produce five segmentation predictions, and then apply the deep supervision mechanism [48] to impose the supervision on each segmentation result.

B. Multiscale Supervised Loss on Labeled Data

Given a labeled data, we can have a pair of input 3-D CT data and the corresponding annotated lung infection mask. It is natural that we take the annotated infection mask as the ground truth of the COVID-19 infected region segmentation. As shown in Fig. 2, the proposed MDA-CNN predicts multiple segmentation results at different CNN layers and these results have different spatial resolutions; see P_1 to P_5 of the last column in Fig. 2. Hence, we downsample the original segmentation mask into the same resolution of the prediction result at each CNN layer as its ground truth. After obtaining the ground truths for the segmentation result at each CNN layer, we devise a multiscale supervised loss (denoted as $\mathcal{L}^s$) for a labeled image (x_n) by adding the supervised losses of all the CNN layers

$$\mathcal{L}^s(x_n) = \sum_{k=1}^{5} \Phi_{\text{dice}}(P_k, G_k) \tag{1}$$

where P_k denotes the predicted lung infection detection map at the kth CNN layer. G_k is the down-sampled ground truth at kth CNN layer and it has the same resolution of P_k. Here, we empirically use the dice loss to compute the difference between P_k and G_k.

C. Multiscale Consistency Loss on Unlabeled Data

The unlabeled 3-D CT data was fed into the student network and teacher network to obtain their infection segmentation results, which are the five predictions at different CNN layers. We then devise a multiscale consistency loss ($\mathcal{L}^c$) to enforce the five predictions from the student network and teacher network to be consistent. The definition of $\mathcal{L}^c$ is given by

$$\mathcal{L}^c(y_m) = \sum_{k=1}^{5} \Phi_{\text{MSE}}(S_k, T_k) \tag{2}$$

where y_m is the input unlabeled image. S_k and T_k are the segmentation results of the student network and the teacher network at the kth CNN layer. Here, we empirically use the mean square error (MSE) loss to compute the difference of S_k and T_k.

Like the original mean teacher framework [49], the multiscale consistency loss on unlabeled data is designed to improve the segmentation performance. The reason why the consistency loss on unlabeled data can increase the segmentation accuracy is summarized as: our semi-supervised network first generates an initial pseudo-label for unlabeled data by training the segmentation network on the labeled data, and then computes the consistency loss on the predictions of a teacher network and a student network to progressively refine the pseudo-labels for unlabeled data. By doing so, we can generate more reliable predictions on unlabeled data, and these reliable unlabeled data are combined with the labeled data into the training process, thereby boosting the overall segmentation performance.

D. Our Loss Function

The total loss ($\mathcal{L}_{\text{total}}$) of our network is computed as

$$\mathcal{L}_{\text{total}} = \sum_{n=1}^{N} \mathcal{L}^s(x_n) + \lambda \sum_{m=1}^{M} \mathcal{L}^c(y_m) \tag{3}$$

where N is the number of labeled CT scans of the training set. M is the number of unlabeled CT scans of our training set. $\mathcal{L}^s(x_n)$ denotes the multiscale supervised loss (1) for the nth labeled image (x_n) while $\mathcal{L}^c(y_m)$ is the multiscale consistency loss (2) for the mth unlabeled image (y_m). The weight λ balances the multiscale supervised loss of labeled data and the multiscale consistency loss of unlabeled data. As suggested in [36] and [49], we compute λ via a time-dependent Gaussian warming up function: $\lambda(i) = \lambda_{\max} e^{(-5(1-i/i_{\max})^2)}$, where t denotes the current training iteration and $i_{\max}$ is the maximum training iteration. We empirically set $\lambda_{\max} = 5$ in our experiments.

Note that averaging model weights over CNN training steps tends to produce a more accurate model than using the final weights directly. Based on this, the mean teacher [49] computes the weights of a network (called teacher network) as an exponential moving average (EMA) weights of a network (called student network) to generate a better target model, which the student network learns from. Hence, one network learns from another network during the training process of the mean teacher framework, and the former network is named the student network, and the latter network is named the teacher network. In the t training iteration, the teacher network parameters Ω_t' is computed by

$$\Omega_t' = \beta \Omega_{t-1}' + (1 - \beta)\Omega_t \tag{4}$$

where Ω_t is the student network parameter at the t training iteration. We set EMA decay $\beta = 0.99$ as same as in [36] and [49].

IV. Results and Discussion

A. Evaluation Dataset and Metric

Evaluation Dataset: For evaluation, we build a new CT volume segmentation dataset (named COVID-19-P20) with 11 unlabeled data and 20 labeled data. The labeled data is collected from COVID-19 3-D CT dataset [13], which provides 20 COVID-19 CT volume data with pixel-level lung infection masks. The infections are first delineated by junior annotators with 1–5 years experience, then refined by two radiologists with 5–10 years experience and, finally, all the annotations were verified and refined by a senior radiologist with more than 10 years experience in chest radiology. According to [13], the last ten scans have been adjusted to the lung window [−1250, 250], and then normalized to [0, 255]. Meanwhile, we adjust the first ten scans to the lung window [−1000, 400], and the intensity values are normalized to [0,1]. Then, we also collect 11 3-D lung CT scans from 11 confirmed COVID-19 cases (8 female and 3 male), as the unlabeled data. These unlabeled data is captured from Philips in Zhongshan Hospital, affiliated to Xiamen University. Similar to the labeled dataset [13], we also adjust these 11 scans to the lung window [−1000, 400], and then normalized to [0, 1] for training.

Here, we conduct a two-fold evaluation on 20 labeled data. Specifically, COVID-19 3-D CT dataset [13] contains ten CT volumes from Coronacases and ten volumes from Radiopaedia. We randomly select five volumes from Coronacases and randomly select five volumes from

Radiopaedia to form the first fold, and the remaining ten scans are for the second fold. Then, one fold combined with the unlabeled data is used as the training set, while the rest fold is utilized as the test set. Finally, we compute the mean±variance of two-fold segmentation results of each method for comparisons.

We extensively build the experiment on a large public dataset, MosMedData [53], which has 50 labeled CT volumes and 806 unlabeled CT volumes. The infections are annotated by the experts of Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department. During the annotation for every given image ground-glass opacifications and regions of consolidation are selected as positive (white) pixels on the corresponding binary pixel mask. Same preprocessing as COVID-19-P20 is adopted here and five-fold evaluation is conducted on the labeled data.

Evaluation Metric: We employ five widely used metrics to quantitatively evaluate the COVID-19 lung infection segmentation performances, including *Dice* coefficient, *Jaccard* coefficient, normalized surface dice (NSD), average distance of boundaries (ADBs), and Hausdorff distance of boundaries (95th percentile; HD95). In general, a better segmentation performance shall have higher *Dice*, *Jaccard*, and NSD scores, as well as lower ADB and HD95 scores. It is noteworthy that all the metrics in this article are calculated in volumewise, which is more meaningful for clinics and convincing for the assessment of 3-D segmentation.

Dice and *Jaccard* coefficients compute the region-based similarity of the predicted segmentation result P and the ground truth G

$$\text{Dice} = \frac{2 \cdot |P \cap G|}{|P| + |G|}, \quad \text{Jaccard} = \frac{|P \cap G|}{|P \cup G|} \tag{5}$$

where $|P \cap G|$ denotes the number of voxels in the intersection area of P and G; $|P \cup G|$ is the number of voxels in the union area of P and G; $|P|$ and $|G|$ are the number of voxels in the region P and the region G, respectively.

NSD evaluates how close the segmentation and ground-truth surfaces are to each other at a specified tolerance, defined as

$$\text{NSD} = \frac{\left|\partial G \cap B^{(\tau)}_{\partial P}\right| + \left|\partial P \cap B^{(\tau)}_{\partial G}\right|}{|\partial G| + |\partial P|} \tag{6}$$

where $B^{(\tau)}_{\partial P}$ and $B^{(\tau)}_{\partial G}$ denote the border region of segmentation surface and ground truth, and they are: $B^{(\tau)}_{\partial P} = \{u \in R^3 | \exists \widetilde{u} \in \partial P, \|u - \widetilde{u}\| \le \tau\}$, and $B^{(\tau)}_{\partial G} = \{v \in R^3 | \exists \widetilde{v} \in \partial G, \|v - \widetilde{v}\| \le \tau\}$. And tolerance τ is empirically set as 1 mm and 3 mm for lung segmentation and infection segmentation, respectively.

ADB and HD estimate the surface distance between the predicted segmentation result and the manual ground truth

$$\begin{aligned} \text{ADB} &= \frac{1}{2}\left\{\frac{\sum_{v_i \in \Phi_P} h(v_i, \Phi_G)}{|G|} + \frac{\sum_{v_j \in \Phi_G} h\big(v_j, \Phi_P\big)}{|P|}\right\} \\ \text{HD} &= \max\left(\max_{v_i \in \Phi_P} h(v_i, \Phi_G), \max_{v_j \in \Phi_G} h\big(v_j, \Phi_P\big)\right) \\ h(v_i, \Phi_G) &= \min_{v_j \in \Phi_G} \text{dist}\big(v_i, v_j\big) \\ h\big(v_j, \Phi_P\big) &= \min_{v_i \in \Phi_P} \text{dist}\big(v_j, v_i\big) \end{aligned} \tag{7}$$

where Φ_P and Φ_G denote the surface of the prediction segmentation and ground truth, respectively. v_i is a vertex of Φ_P and v_j is a vertex of Φ_G. $\text{dist}(v_i, v_j)$ is the Euclidean distance between the vertex v_i and the vertex v_j. Apparently, ADB counts the average surface distance of the predicted segmentation and the ground-truth surfaces. HD computes the maximum distance between two segmentation surfaces, and HD95 is a modified HD by using the 95% percentile instead of the maximum distance (100% percentile) in HD in order to eliminate the impact of a small subset of the outliers; see [36] for more details.

B. Implementation

Training Parameters: All parameters of our network are initialized from scratch, without requiring any pretrained weight. We augment the training set using a random flipping in all directions and adding Gaussian noise with the noise intensity $\sigma = 10$. Adam is employed to optimize the whole network with an initial learning rate of 0.0003 and 5000 iterations. We randomly sample 3-D blocks with a size of 160 × 160 × 64 for COVID-19-P20 or 160 × 160 × 32 for MosMedData from each training CT volume for training our network on a single TITAN RTX. The mini-batch size is 4, which consists of two labeled images and two unlabeled images. The model size of our network is 18.79 Mb (megabyte), and the training time is 23 h.

Inference: In the testing stage, we adopt the sliding window with a 50% overlapping rate to continually crop a set of volumes with a size of 160 × 160 × 64 for COVID-19-P20 or 160 × 160 × 32 for MosMedData. Moreover, we feed these cropped volumes into the student network of developed DM^{2}T-Net to generate multiple segmentation masks. Finally, we obtain the final segmentation of our network by stitching these small segmentation masks according to their crop positions. The average inference time (including preprocessing time) is 2.12 s for one volume.

C. Comparison With the State-of-the-Art Methods

We compare our method against seven state-of-the-art segmentation methods, including 2-D U-Net [14], U-Net++ [40], DLA [50], 3-D U-Net [47], V-Net [51], nn-UNet [52], and UA-MT [36]. Among them, the first three segmentation methods are based on 2-D images while the other four methods directly perform the segmentation on 3-D volumes. And the last one (UA-MT) is a state-of-the-art 3-D semi-supervised segmentation method, which presented an uncertainty-aware self-ensembling model. To make the comparisons fair, we adopt the released code of compared methods and fine-tune the parameters to obtain their best segmentation results.

Table I reports the quantitative results of different methods on COVID-19-P20. Apparently, the 3-D deep-learning-based methods [47], [51], [52] have superior performance of five metrics scores than 2-D-image-based CNNs (i.e., 2-D U-Net [14] and 2-D U-Net++ [40]), since these 3-D segmentation methods can learn more interslice relations among 3-D volume and reduce the false predictions on 2-D slice without any COVID lung infection, which usually happen in the segmentation results of 2-D U-Net and 2-D U-Net++. Moreover, due to the additional unlabeled data in the training set, the

TABLE I
RESULTS (MEAN ± VARIANCE) OF DIFFERENT SEGMENTATION METHODS ON **COVID-19-P20**. WE USE THE BOLD FONTS TO HIGHLIGHT THE BEST PERFORMANCE

Method	data type	Dice ↑	Jaccard ↑	NSD ↑	ADB ↓	HD95 ↓
2D U-Net ([14])	2D	63.98±19.72	49.63±18.20	72.61±23.01	7.27±12.92	18.77±30.71
DLA ([50])	2D	64.82±18.66	50.23±16.74	73.29±21.21	6.29±11.80	20.44±32.10
U-Net++ ([40])	2D	66.01±19.12	51.70±17.35	72.23±21.77	7.49±11.94	23.93±33.81
3D U-Net ([47])	3D	65.91±22.55	52.60±20.82	73.83±25.67	10.42±22.53	27.00±46.60
V-Net ([51])	3D	65.89±20.10	51.89±18.71	70.69±24.82	9.32±15.74	26.00±35.97
nn-UNet ([52])	3D	67.89±20.56	54.38±19.51	73.00±25.15	9.61±17.24	25.30±35.10
UA-MT ([36])	3D	69.32±19.07	55.60±17.85	74.73±23.20	8.22±16.52	24.60±36.69
Our method (DM^{2}T-Net)	3D	**72.59±18.55**	**59.42±17.09**	**80.44±20.19**	**4.45±8.06**	**16.34±24.76**

TABLE II
RESULTS (MEAN ± VARIANCE) OF DIFFERENT SEGMENTATION METHODS ON **MOSMEDDATA**. WE USE THE BOLD FONTS TO HIGHLIGHT THE BEST PERFORMANCE

Method	data type	Dice ↑	Jaccard ↑	NSD ↑	ADB ↓	HD95 ↓
2D U-Net ([14])	2D	53.47±22.39	39.47±19.75	77.94±23.52	6.45±15.51	17.61±22.40
DLA ([50])	2D	55.71±18.90	40.91±17.75	80.34±18.20	5.08±7.75	22.01±28.64
U-Net++ ([40])	2D	56.03±22.46	42.05±20.39	77.71±22.71	5.51±6.94	22.05±28.17
3D U-Net ([47])	3D	54.42±23.51	40.69±20.70	75.31±27.09	8.21±14.43	21.35±27.39
V-Net ([51])	3D	48.95±19.76	34.61±17.03	71.88±20.39	5.16±8.19	21.20±18.23
nn-UNet ([52])	3D	56.30±23.55	42.62±21.31	76.45±27.16	9.26±18.25	20.59±29.97
UA-MT ([36])	3D	57.31±20.53	42.87±19.05	78.55±21.42	6.89±13.90	20.68±26.53
Our method (DM^{2}T-Net)	3D	**60.19±19.22**	**45.56±18.44**	**80.95±20.99**	**6.55±13.79**	**21.11±27.74**

semi-supervise method, UA-MT [36], further outperforms all these 3-D supervised techniques in terms of all the five metrics. Compared to the best-performing existing method (UA-MT), our method has 4.72% improvement on Dice, 6.87% improvement on Jaccard, 7.64% improvement on Jaccard, and 45.86% reduction on ADB, 33.58% reduction on HD95, respectively. It indicates that our network can more accurately detect COVID-19 infected lung regions than state-of-the-art methods from 3-D CT scans. We extensively evaluate the effectiveness of our method on MosMedData. The results in Table II show that our network has larger Dice, Jaccard, and NSD scores, as well as smaller ADB and HD95 scores than state-of-the-art methods. It further indicates that our network can more accurately segment COVID-19 infected regions from CT scans. Moreover, compared to COVID-19-P20, we can find that our network and state-of-the-art methods suffer from a degraded performance on MosMedData for all five metrics. The main reason is that the data in MosMedData is more challenging and the infected regions in MosMedData are smaller than COVID-19-P20, thereby increasing the segmentation difficulties. On the other hand, we argue that there are two main reasons why 2-D U-Net performs poorly on two datasets in our network. First, existing works computed the DSC value of 2-D U-Net on 2-D slices, while our work computes all five metrics (including DSC metric) on the whole 3-D volume. Second, existing works have not tested the 2-D U-Net model on slices without COVID-19 infections, and thus a large amount of false positives will not be involved for computing the DSC score.

Figs. 4 and 5 visually compare the COVID-19 lung infection segmentation results produced by our network and compared methods. Apparently, compared methods tend to include many noninfection regions or neglect parts of infection regions in their segmentation results, while our network predicts more accurate infection segmentation results. For these challenging inputs with multiple infection regions and different infection region sizes in Fig. 5, our network can still better segment these infected regions than all the compared methods. It further verifies the effectiveness of the developed dual multiscale mean teacher framework in our work. From the perspective of clinical importance, our method brings obvious improvement on those small and challenging infections, which are even hard for junior radiologists to determine. Although these infections are too small to make a great difference on the statistics, successful segmentation of them has much larger significance in practice.

D. Ablation Analysis

Here, we provide several experiments to validate the effectiveness of main components of our network, including the multiple dimensional-scale mechanism, multiscale supervised loss, and multiscale consistency loss.

Effectiveness of Multiple Dimensional-Scale Downsampling: First, we construct a basic model (denoted as "basic") by removing the teacher network from our method, the dimensional-scale downsampling operations of the input

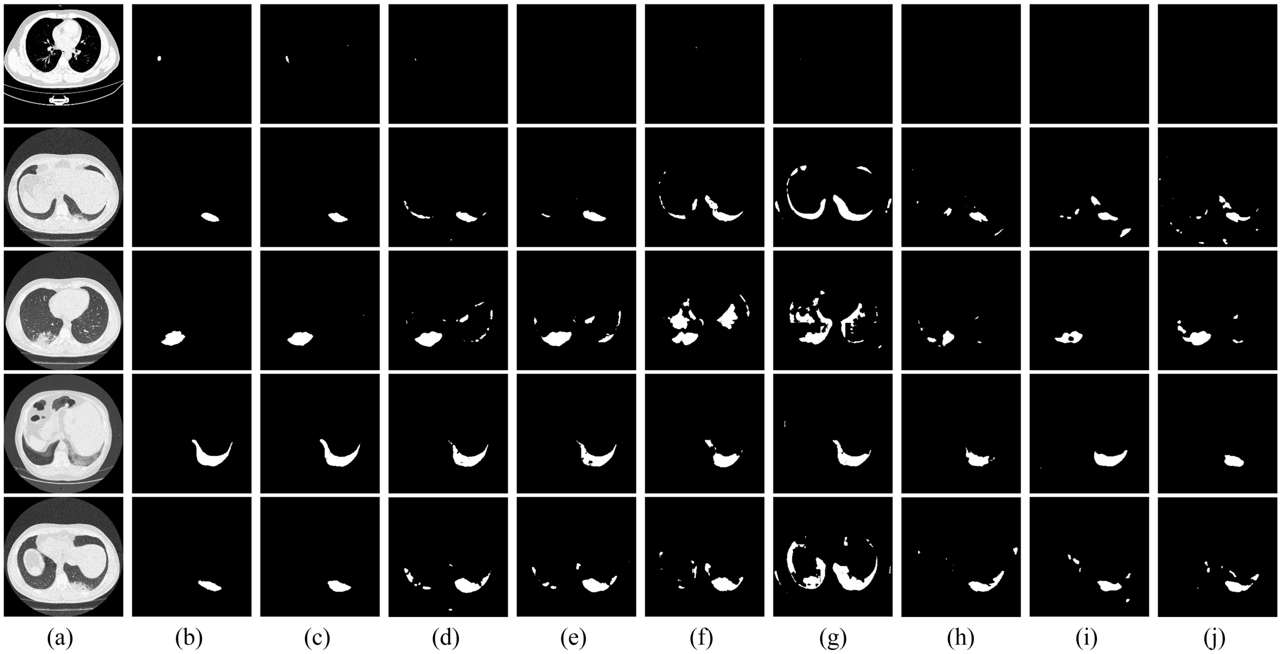

Fig. 4. Visual comparison of segmentation results produced by different methods on the COVID-19-P20 dataset. (a) Input images; (b) ground truths (denoted as GT); (c)–(h) segmentation results predicted by our method, UA-MT [36], nn-UNet [52], 3-D U-Net [47], DLA [50], U-Net++ [40], and 2-D U-Net [14]. Apparently, our network can more accurately identify COVID-19 lung infected regions than other methods.

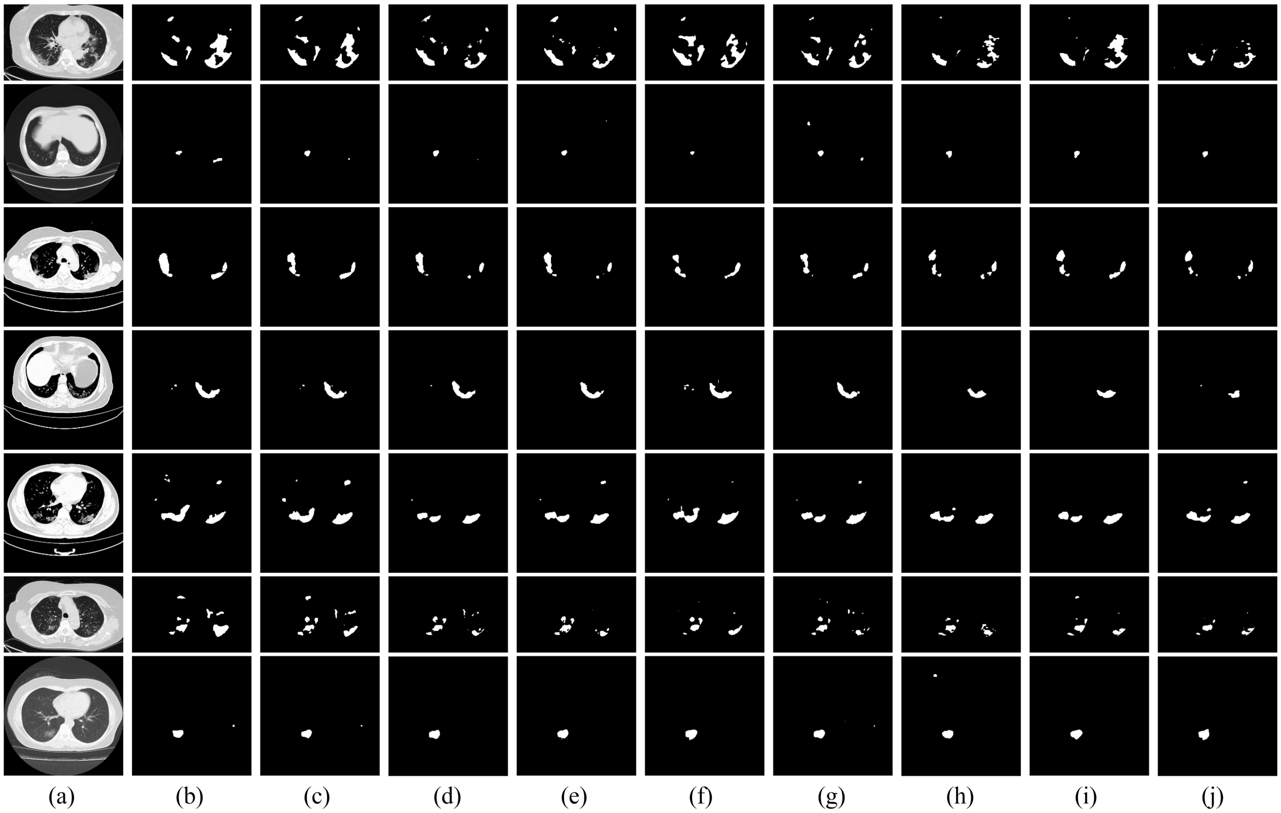

Fig. 5. Visual comparison of segmentation results produced by different methods on the COVID-19-P20 dataset (continued from Fig. 4). (a) Input images with multiple infected regions; (b) ground truths (denoted as GT); (c)–(h) segmentation results predicted by our method, UA-MT [36], nn-UNet [52], 3-D U-Net [47], DLA [50], U-Net++ [40], and 2-D U-Net [14]. Apparently, our network can more accurately identify COVID-19 lung infected regions than other methods.

volume, and the segmentation predictions P_2, P_3, P_4, and P_5 (see Fig. 1). In this way, the basic model almost becomes the classical 3-D U-Net model. After that, we add the multiple dimensional-scale mechanism to basic by fusing features from multiple down-sampled volumes for building another model (denoted as "basic + multidimensional-scale") to evaluate the

TABLE III
RESULTS (MEAN ± VARIANCE) OF DIFFERENT ABLATION STUDY EXPERIMENTS ON THE COVID-19-P20 DATASET. WE USE THE BOLD FONTS TO HIGHLIGHT THE BEST PERFORMANCE

Method	multiple-dimension-scale	dual multi-scale	semi-supervised	Dice ↑	Jaccard ↑	NSD ↑	ADB ↓	HD95 ↓
basic	×	×	×	67.89±20.56	54.38±19.51	73.00±25.15	9.61±17.24	25.30±35.10
basic+multi-dimensional-scale	√	×	×	70.86±18.79	57.38±17.55	77.19±21.69	7.09±12.83	22.35±34.12
basic+dual-multiscale	√	√	×	71.53±18.85	58.21±17.55	78.29±21.51	6.04±10.49	19.89±27.70
semi-basic	×	×	√	69.89±19.48	56.45±18.67	75.84±23.51	7.60±16.81	19.92±33.89
semi-multi-dimensional-scale	√	×	√	71.20±18.70	57.76±17.35	77.15±21.29	6.41±14.07	20.26±36.86
Our method (DM^{2}T-Net)	√	√	√	**72.59±18.55**	**59.42±17.09**	**80.44±20.19**	**4.45±8.06**	**16.34±24.76**

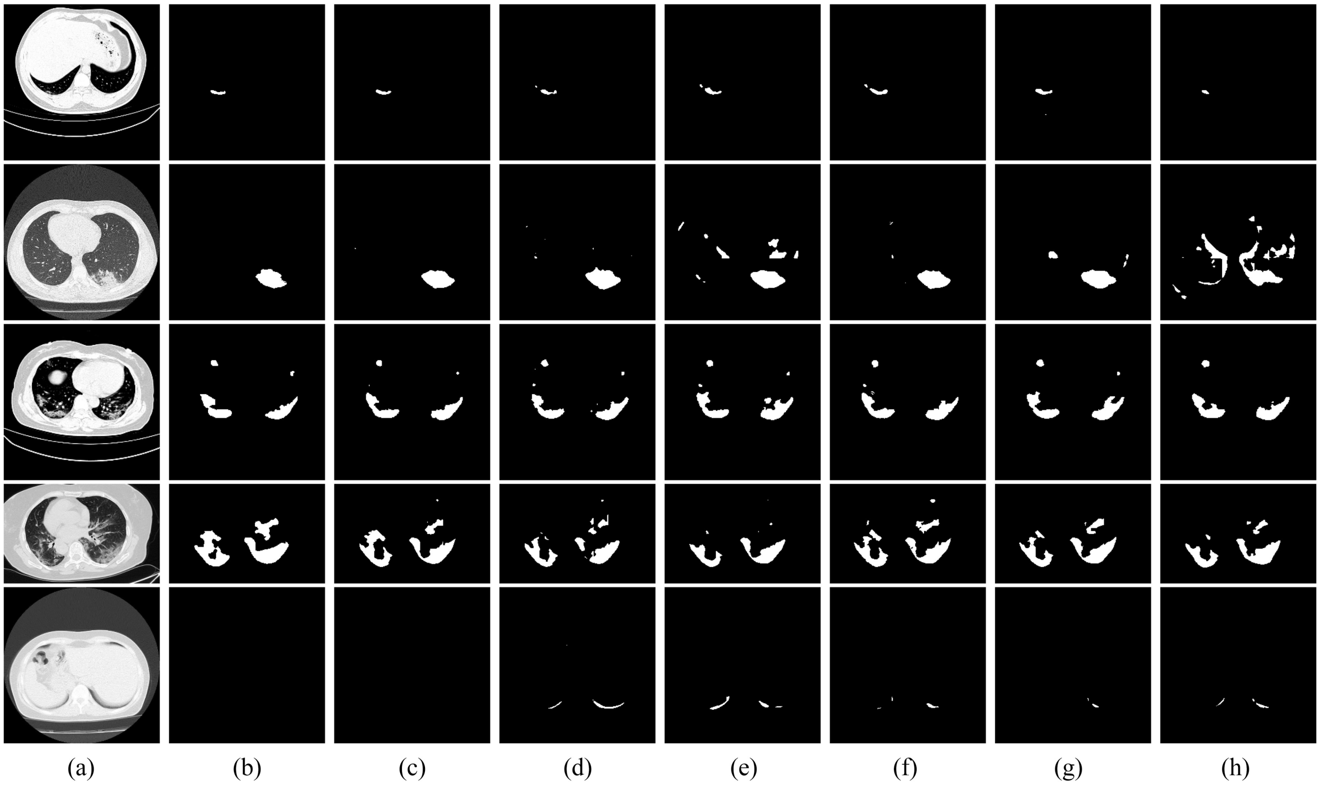

Fig. 6. Visual comparisons on COVID-19 infection segmentation results produced by different networks of the ablation study experiment on the COVID-19-P20 dataset. Note that "mul-dimsca" "dual-mulsca," and "semi-mdsca" denote "basic + multidimensional-scale," "basic + dual-multiscale," and "semi-multidimensional-scale," respectively. (a) Inputs. (b) GT. (c) Our method. (d) semi-mdsca. (e) semi-basic. (f) dual-mulsca. (g) mul-dimsca. (h) Basic.

contribution of our multiple dimensional-scale mechanism. As shown in Table III, basic + multidimensional-scale has a superior performance of five metrics over basic. It shows that fusing features from multiple down-sampled volumes enables our method to accurately identify COVID-19 lung infected regions.

Effectiveness of Multiscale Supervised Loss: We then investigate the importance of the multiscale supervised loss. To do so, we build a model (denoted as "basic + dual-multiscale") by removing the multiscale consistency loss from our network. Compared to basic + multidimensional-scale, we predict additional four segmentation results from P_2, P_3, P_4, and P_5 in basic + dual-multiscale, and thus formulate the multiscale supervised loss [see (1)]. From the results shown in Table III, basic + dual-multiscale performs better than basic + multidimensional-scale. It demonstrates that aggregating the supervised losses from different CNN layers via a multiscale supervised loss helps our method to better identify COVID-19 infected regions in our method.

Effectiveness of Multiscale Consistency Loss: We finally investigate the importance of the multiscale consistency loss by constructing another two models with unlabeled data. The first one ("semi-basic") adds the unlabeled data and encourages the segmentation results of basic from the student network and the teacher network to be consistent. The second model ("semi-multidimensional-scale") is to produce the infection segmentation results via basic + multidimensional-scale and regularize the segmentation results from the student and teacher network to be consistent.

Table III reports the quantitative results of "semi-basic," semi-multidimensional-scale, and our method. Apparently, "semi-basic" can more accurately segment COVID-19 infected lung regions than basic due to its superior performance of all the five metrics. It indicates that the additional consistency loss from the unlabeled data incurs a superior infection segmentation performance. Then, as shown in Table III, semi-multidimensional-scale outperforms semi-basic in terms of all five metrics, demonstrating that aggregating CNN features

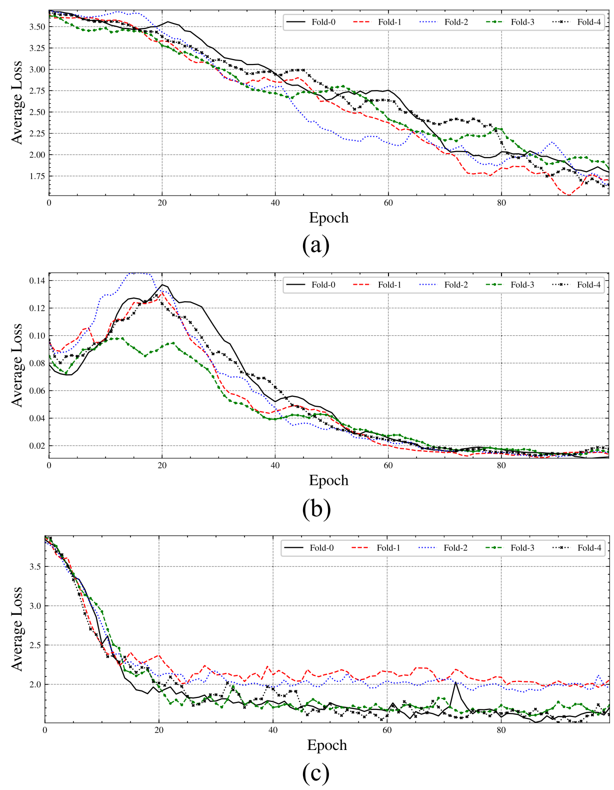

Fig. 7. Loss analysis during training our DM^{2}T-Net on the COVID-19-P20 dataset. (a) Training multiscale supervised loss. (b) Training multiscale consistency loss. (c) Testing multiscale supervised loss.

from multiple dimensional-scaled inputs enables a more accurate supervised loss and a more accurate consistency loss, thereby improving the lung infection segmentation accuracy. Moreover, compared to semi-multidimensional-scale, our method has higher Dice, Jaccard, and NSD scores, as well as lower ADB and HD95 scores, which shows that computing the consistency loss from multiscale predictions at CNN layers further boosts the segmentation performance of our method.

Visual Comparisons: Fig. 6 visually compares the segmentation results produced by our method and the five constructed baseline networks of the ablation study experiment (see Table III). By observing these segmentation results, we can easily find that our DM^{2}T-Net better segments the COVID-19 infections than all the five baseline networks. It further proves the effectiveness of considering unlabeled data and dual multiscale information within an end-to-end network in our work.

E. Discussion

In this part, we provide the detailed analysis about: 1) how the multiscale consistency loss makes contribution to the DM^{2}T-Net, including the loss curves and visual consistency between teacher and student networks and 2) generalization analysis of the infection segmentation performance.

How Does Multiscale Consistency Loss Work? Fig. 7 presents the loss curves during semi-supervised learning process. It includes the multiscale supervised segmentation loss [Fig. 7(a)] for labeled data and multiscale consistency loss [Fig. 7(b)] for unlabeled data during training, as well as the testing multiscale supervised loss [Fig. 7(c)]. It is observed that the consistency loss progressively increases at the first 20 epochs and then converges to a lower value after training 60 epochs. Since the teacher network's parameters are an average of consecutive student networks [49], the difference at early epochs will be increasing and the averaging weight over large training steps tends to produce a more accurate supervision to the student network. Moreover, the supervised loss on the training set and the testing set decreases and then reaches stable values as the epoch number increases.

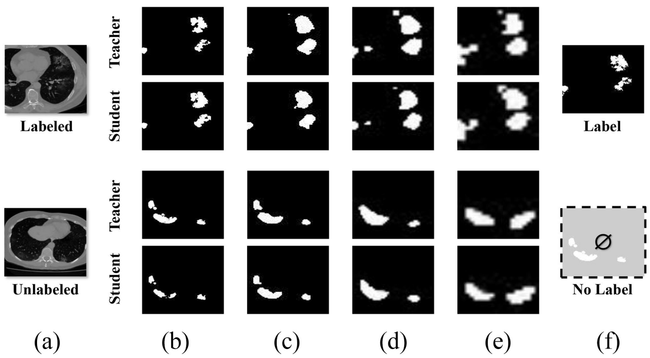

Fig. 8. Segmentation results from the teacher and student networks at different scales on the COVID-19-P20 dataset. (a) Inputs. (b) Scale 1. (c) Scale 2. (d) Scale 3. (e) Scale 4. (f) GT.

We additionally visualize all layers outputs (scales) from the teacher and student networks in Fig. 8. First, considering the teacher network's parameters are the averaged result of the student network's, we can find that the teacher network's predictions are naturally more accurate than that of the student network, when compared to the ground truth of the labeled data. More importantly, combining the multiscale labeled data loss and the multiscale unlabeled data enables our network to produce good predictions for labeled data and unlabeled data. Hence, integrating the unlabeled data into the network training improves the COVID-19 infected region segmentation performance.

Generalization Analysis: We make the discussion about model generalization in Table IV, where we use the model trained on one dataset to segment the infections of the other one dataset and assess the performance. From the quantitative results, we have the following observations.

1) Due to the imaging variance and infection difference on two datasets, the segmentation performance of both 3-D-based UA-MT and 2-D-based U-Net++ decrease, but the 3-D-based segmentation method has a better generalization capability than the 2-D-based U-Net++ on both cross-dataset evaluation settings. The underlying reason is that the 3-D-based method is able to capture more high-level information of lung infections among multiple image slices than the 2-D-based segmentation performance.
2) More importantly, compared to the best existing 3-D-based method (UA-MT), our method still outperforms it in terms of Dice and HD95, which demonstrates the generalized advancement of our network. We have added this experiment into Section IV-E of the revised manuscript.

TABLE IV
GENERALIZATION ANALYSIS, WHERE WE USE ONE DATASET TO TRAIN THE NETWORK AND EVALUATE IT ON THE OTHER ONE DATASET. LET C AND M DENOTE THE COVID-19-P20 AND MOSMEDDATA DATASETS, SO THAT C → M MEANS TRAINING THE NETWORK ON C AND EVALUATING IT ON M

	C → M		M → C	
	Dice ↑	HD95 ↓	Dice ↑	HD95 ↓
U-Net++ ([40])	11.09±11.48	51.24±24.78	23.61±17.31	66.82±26.13
UA-MT ([36])	43.95±24.69	44.95±34.12	50.24±17.70	51.98±38.58
Ours	**48.81±23.51**	**35.34±30.57**	**51.36±17.80**	**41.55±32.19**

V. CONCLUSION

This work has presented a novel COVID-19 lung infection segmentation network from 3-D CT volumes by developing a DM^{2}T-Net. Our key idea is to first develop an MDA-CNN to explore multiple dimensional-scale details of the input 3-D CT scan. Moreover, we also employ a semi-supervised system to leverage additional unlabeled data and dual multiscale information for further boosting the COVID-19 infected lung region segmentation. Two datasets for COVID-19 segmentation are collected to evaluate the effectiveness, where our model finally achieves the Dice score of 72.59% on the COVID-19-P20 dataset and achieved the Dice score of 60.19% on the MosMedData dataset. The results show that our DM^{2}T-Net performs better than the state-of-the-art methods by a large margin.

REFERENCES

[1] F. Shi et al., "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 4–15, 2020.

[2] G. D. Rubin et al., "The role of chest imaging in patient management during the COVID-19 pandemic: A multinational consensus statement from the Fleischner society," *Chest*, vol. 296, no. 1, pp. 1–8, 2020.

[3] D.-P. Fan et al., "Inf-net: Automatic COVID-19 lung infection segmentation from CT scans," 2020, *arXiv:2004.14133*.

[4] G. Chassagnon et al., "AI-driven CT-based quantification, staging and short-term outcome prediction of COVID-19 pneumonia," 2020, *arXiv:2004.12852*.

[5] S. Chaganti et al., "Quantification of tomographic patterns associated with COVID-19 from chest CT," 2020, *arXiv:2004.01279*.

[6] N. Zheng et al., "Predicting COVID-19 in China using hybrid AI model," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 2891–2904, Jul. 2020.

[7] X. Chen, L. Yao, and Y. Zhang, "Residual attention U-Net for automated multi-class segmentation of COVID-19 chest CT images," 2020, *arXiv:2004.05645*.

[8] T. Zhou, S. Canu, and S. Ruan, "An automatic COVID-19 CT segmentation based on U-Net with attention mechanism," 2020, *arXiv:2004.06673*.

[9] Q. Yan et al., "COVID-19 chest CT image segmentation—A deep convolutional neural network solution," 2020, *arXiv:2004.10987*.

[10] Y. Qiu, Y. Liu, S. Li, and J. Xu, "MiniSeg: An extremely minimum network for efficient COVID-19 segmentation," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 4846–4854.

[11] W. Xie, C. Jacobs, J.-P. Charbonnier, and B. van Ginneken, "Relational modeling for robust and efficient pulmonary lobe segmentation in CT scans," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2664–2675, Aug. 2020.

[12] H. Kang et al., "Diagnosis of Coronavirus disease 2019 (COVID-19) with structured latent multi-view representation learning," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2606–2614, Aug. 2020.

[13] J. Ma et al., "Towards efficient COVID-19 CT annotation: A benchmark for lung and infection segmentation," 2020, *arXiv:2004.12537*.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervent.*, 2015, pp. 234–241.

[15] P. M. Gordaliza, A. Muñoz-Barrutia, M. Abella, M. Desco, S. Sharpe, and J. J. Vaquero, "Unsupervised CT lung image segmentation of a mycobacterium tuberculosis infection model," *Sci. Rep.*, vol. 8, no. 1, pp. 1–10, 2018.

[16] A. Munoz-Barrutia, M. Ceresa, X. Artaechevarria, L. M. Montuenga, and C. Ortiz-de Solorzano, "Quantification of lung damage in an elastase-induced mouse model of emphysema," *Int. J. Biomed. Imag.*, vol. 2012, Nov. 2012, Art. no. 734734.

[17] D. Wu et al., "Stratified learning of local anatomical context for lung nodules in CT images," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2791–2798.

[18] M. Keshani, Z. Azimifar, F. Tajeripour, and R. Boostani, "Lung nodule segmentation and recognition using SVM classifier and active contour modeling: A complete intelligent system," *Comput. Biol. Med.*, vol. 43, no. 4, pp. 287–300, 2013.

[19] B. Lei, P. Yang, T. Wang, S. Chen, and D. Ni, "Relational-regularized discriminative sparse learning for Alzheimer's disease diagnosis," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1102–1113, Apr. 2017.

[20] L. Wu, J.-Z. Cheng, S. Li, B. Lei, T. Wang, and D. Ni, "FUIQA: Fetal ultrasound image quality assessment with deep convolutional networks," *IEEE Trans. Cybern.*, vol. 47, no. 5, pp. 1336–1349, May 2017.

[21] H. Chen et al., "Ultrasound standard plane detection using a composite neural network framework," *IEEE Trans. Cybern.*, vol. 47, no. 6, pp. 1576–1586, Jun. 2017.

[22] B. Sheng et al., "Retinal vessel segmentation using minimum spanning Superpixel tree detector," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2707–2719, Jul. 2019.

[23] H. Wu, X. Chen, P. Li, and Z. Wen, "Automatic symmetry detection from brain MRI based on a 2-channel convolutional neural network," *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4464–4475, Sep. 2021.

[24] H. Fu et al., "Angle-closure detection in anterior segment OCT based on multilevel deep network," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3358–3366, Jul. 2020.

[25] M. Liu, J. Zhang, C. Lian, and D. Shen, "Weakly supervised deep learning for brain disease prognosis using MRI and incomplete clinical scores," *IEEE Trans. Cybern.*, vol. 50, no. 7, pp. 3381–3392, Jul. 2020.

[26] S. Wang et al., "Central focused convolutional neursal networks: Developing a data-driven model for lung nodule segmentation," *Med. Image Anal.*, vol. 40, pp. 172–183, Aug. 2017.

[27] D. Jin, Z. Xu, Y. Tang, A. P. Harrison, and D. J. Mollura, "CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervent.*, 2018, pp. 732–740.

[28] A. P. Harrison, Z. Xu, K. George, L. Lu, R. M. Summers, and D. J. Mollura, "Progressive and multi-path holistically nested neural networks for pathological lung segmentation from CT images," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervent.*, 2017, pp. 621–629.

[29] J. Jiang et al., "Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 134–144, Jan. 2019.

[30] V. Cheplygina, M. de Bruijne, and J. P. Pluim, "Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Med. Image Anal.*, vol. 54, pp. 280–296, May 2019.

[31] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Workshop Challenges Rep. Learn. (ICML)*, vol. 3, 2013, p. 2.

[32] W. Bai et al., "Semi-supervised learning for network-based cardiac MR image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervent.*, 2017, pp. 253–260.

[33] N. Dong, M. Kampffmeyer, X. Liang, Z. Wang, W. Dai, and E. Xing, "Unsupervised domain adaptation for automatic estimation of cardiothoracic ratio," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervent.*, 2018, pp. 544–552.

[34] D. Nie, Y. Gao, L. Wang, and D. Shen, "ASDNet: Attention based semi-supervised deep networks for medical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervent.*, 2018, pp. 370–378.

[35] Y. Zhang et al., "Deep adversarial networks for biomedical image segmentation utilizing unannotated images," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervent.*, 2017, pp. 408–416.

[36] 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. Int. Conf. Med. Image Comput. Comput. Assisted Intervent.*, 2019, pp. 605–613.

[37] D. Dong et al., "The role of imaging in the detection and management of COVID-19: A review," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 16–29, 2021.

[38] Z. Tang et al., "Severity assessment of coronavirus disease 2019 (COVID-19) using quantitative features from chest CT images," 2020, *arXiv:2003.11988*.

[39] S. Wang et al. "A deep learning algorithm using CT images to screen for corona virus disease (COVID-19)." 2020. [Online]. Available: https://www.medrxiv.org/content/10.1101/2020.02.14.20023028v5

[40] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2018, pp. 3–11.

[41] J. Chen et al., "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: A prospective study," *Sci. Rep.*, vol. 10, Nov. 2020, Art. no. 19196.

[42] Y. Song et al., "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 18, no. 6, pp. 2775–2780, Nov./Dec. 2021.

[43] Y.-H. Wu et al., "JCS: An explainable COVID-19 diagnosis system by joint classification and segmentation," 2020, *arXiv:2004.07054*.

[44] K. He et al., "Synergistic learning of lung lobe segmentation and hierarchical multi-instance classification for automated severity assessment of COVID-19 in CT images," 2020, *arXiv:2005.03832*.

[45] W. Ding, M. Abdel-Basset, H. Hawash, and O. M. Elkomy, "MT-NCOV-Net: A multitask deep-learning framework for efficient diagnosis of COVID-19 using tomography scans," *IEEE Trans. Cybern.*, early access, Nov. 8, 2021, doi: 10.1109/TCYB.2021.3123173.

[46] S. Pang et al., "Beyond CNNs: Exploiting further inherent symmetries in medical image segmentation," *IEEE Trans. Cybern.*, early access, Aug. 31, 2022, doi: 10.1109/TCYB.2022.3195447.

[47] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assist. Intervent.*, 2016, pp. 424–432.

[48] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1395–1403.

[49] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Proc. NIPS*, 2017, pp. 1195–1204.

[50] F. Yu, D. Wang, E. Shelhamer, and T. Darrell, "Deep layer aggregation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2403–2412.

[51] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, 2016, pp. 565–571.

[52] F. Isensee, J. Petersen, S. A. Kohl, P. F. Jäger, and K. H. Maier-Hein, "nnU-net: Breaking the spell on successful medical image segmentation," 2019, *arXiv:1904.08128*.

[53] S. Morozov et al., "MosMedData: Chest CT scans with COVID-19 related findings dataset," 2020, *arXiv:2005.06465*.

Liansheng Wang (Member, IEEE) received the Ph.D. degree in computer science from The Chinese University of Hong Kong, Hong Kong, in 2012.

He is currently an Associate Professor with the Department of Computer Science, Xiamen University, Xiamen, China. His research interests include medical image processing and analysis.

Jiacheng Wang (Student Member, IEEE) received the B.S. degree from Xiamen University, Xiamen, China, in 2018, where he is currently pursuing the master's degree with the Department of Computer Science.

His main research interests include medical image processing and machine learning.

Lei Zhu received the Ph.D. degree from the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, in 2017.

He is currently an Assistant Professor with the Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China. Before that, he worked as a Postdoctoral Researcher with The Hong Kong Polytechnic University, Hong Kong, The Chinese University of Hong Kong, and the University of Cambridge, Cambridge, U.K. His research interests include computer vision, computer graphics, image and video processing, AI-based healthcare, and deep learning.

Huazhu Fu (Senior Member, IEEE) received the Ph.D. degree from Tianjin University, Tianjin, China, in 2013.

He is a Senior Scientist with the Institute of High Performance Computing, A*STAR, Singapore. Previously, he was a Research Fellow with Nanyang Technological University, Singapore, from 2013 to 2015; a Research Scientist with I2R, A*STAR, from 2015 to 2018; and a Senior Scientist with the Inception Institute of Artificial Intelligence, Abu Dhabi, UAE, from 2018 to 2021. His research interests include computer vision, AI in healthcare, and trustworthy AI.

Dr. Fu received the Best Paper Award of ICME 2021 and the Best Paper Award of OMIA Workshop in MICCAI 2022. He has served as an Associate Editor for IEEE TRANSACTIONS ON MEDICAL IMAGING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS. He is also a member of the IEEE BISP TC.

Ping Li (Member, IEEE) received the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Hong Kong, in 2013.

He is currently an Assistant Professor with the Department of Computing and an Assistant Professor with the School of Design, The Hong Kong Polytechnic University, Hong Kong. He has an excellent research project reported by the *ACM TechNews*, which only reports the top breakthrough news in computer science worldwide. More importantly, however, many of his research outcomes have strong impacts to research fields, addressing societal needs and contributed tremendously to the people concerned. His current research interests include image/video stylization, colorization, artistic rendering and synthesis, realism in nonphotorealistic rendering, computational art, and creative media. He has published many scholarly research articles, pioneered several new research directions, and made a series of landmark contributions in his research areas.

Gary Cheng received the Ph.D. degree in computing from The Hong Kong Polytechnic University, Hong Kong, in 2001.

He is currently the Acting Head and an Associate Professor with the Department of Mathematics and Information Technology, The Education University of Hong Kong (EdUHK), Hong Kong. His academic background is in computer science, but he has specialized in teaching Information Technology in education for over a decade. With substantial years of work experience in Hong Kong academia, he has built a wealth of knowledge and a network of support to unleash the potential of technology for teacher education. He has a proven track record of exploring and evaluating the use of emerging technologies to enhance teaching and learning. Over the years, he has been involved in research projects funded by EdUHK and the Research Grant Council of Hong Kong on a range of topics mainly related to technology-enhanced learning. He has also devoted himself to organizing events and activities to promote coding, STEM, and AI in education, such as Inter-Primary School Mobile Apps Design Competition in 2017, STEAM Education: 3D Chinese Cultural Architectural Design Competition in 2018, STEM Competition in Smart Product Design in 2019, Workshop and Seminar Series on Big Data/AI with Applications to Education and Beyond in 2020, and International Conference on Education and Artificial Intelligence in 2020.

Zhipeng Feng holds the bachelor of medicine degree from Nanchang University, Nanchang, China, in 2000.

He is currently a Radiologist with the Department of Medical Imaging, Zhongshan Hospital, Xiamen University, Xiamen, China. His research interests include medical image processing and analysis.

Shuo Li (Senior Member, IEEE) received the Ph.D. degree from Concordia University, Montreal, QC, Canada, in 2006.

He is currently a Professor with Western University, London, ON, Canada. He is also a member of the MICCAI Society Board, a Research Fellow with the Lawson Institute of Health, and a member of the IEEE Engineering in Medicine and Biology Society's Translational Engineering and Healthcare Innovation Technology Committee. He has published more than 100 papers in top international journals and conferences, such as IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, *Medical Image Analysis*, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON MEDICAL IMAGING, and CVPR. His research interests include cardiac electrophysiology and cardiac image analysis.

Prof. Li is the Deputy Editor of *Medical Image Analysis* and *Computerized Medical Imaging and Graphics*. He is also a Guest Editor of the IEEE TRANSACTIONS ON MEDICAL IMAGING, *Computerized Medical Imaging and Graphics*, and *Computer Vision and Image Understanding*.

Pheng-Ann Heng (Senior Member, IEEE) received the B.Sc. degree from the National University of Singapore, Singapore, in 1985, and the M.Sc. degree in computer science, the M.Art degree in applied math, and the Ph.D. degree in computer science from Indiana University, Bloomington, IN, USA, in 1987, 1988, and 1992, respectively.

He is a Professor with the Department of Computer Science and Engineering, The Chinese University of Hong Kong (CUHK), Hong Kong, where he has been serving as the Director of Virtual Reality, Visualization and Imaging Research Center since 1999 and the Director of Center for Human–Computer Interaction, Shenzhen Institute of Advanced Integration Technology, Chinese Academy of Science/CUHK, Shenzhen, China, since 2006. He has been appointed as a Visiting Professor with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, as well as a Cheung Kong Scholar Chair Professor with the Ministry of Education and University of Electronic Science and Technology of China, Chengdu, China, since 2007. His research interests include AI and VR for medical applications, surgical simulation, visualization, graphics, and human–computer interaction.