Attention-RefNet: Interactive Attention Refinement Network for Infected Area Segmentation of COVID-19

Titinunt Kitrungrotsakul, Qingqing Chen, Huitao Wu, Yutaro Iwamoto, *Member, IEEE*, Hongjie Hu, Wenchao Zhu, Chao Chen, Fangyi Xu, Yong Zhou, Lanfen Lin, Ruofeng Tong, Jingsong Li, and Yen-Wei Chen, *Member, IEEE*

***Abstract*—COVID-19 pneumonia is a disease that causes an existential health crisis in many people by directly affecting and damaging lung cells. The segmentation of infected areas from computed tomography (CT) images can be used to assist and provide useful information for COVID-19 diagnosis. Although several deep learning-based segmentation methods have been proposed for COVID-19 segmentation and have achieved state-of-the-art results, the segmentation accuracy is still not high enough (approximately 85%) due to the variations of COVID-19 infected areas (such as shape and size variations) and the similarities between COVID-19 and non-COVID-infected areas. To improve the segmentation accuracy of COVID-19 infected areas, we propose an interactive attention refinement network (Attention RefNet). The interactive attention refinement network can be connected with any segmentation network and trained with the segmentation network in an end-to-end fashion. We propose a skip connection attention module to improve the important features in both segmentation and refinement networks and a seed point module to enhance the important seeds (positions) for interactive refinement. The effectiveness of the proposed method was demonstrated on public datasets (COVID-19CTSeg and MICCAI) and our private multicenter dataset. The segmentation accuracy was improved to more than 90%. We also confirmed the generalizability of the proposed network on our multicenter dataset. The proposed method can still achieve high segmentation accuracy.**

***Index Terms*—COVID-19, CT scan, deep learning, segmentation, interactive, attention, seed point, refinement.**

Manuscript received December 24, 2020; revised March 19, 2021 and May 10, 2021; accepted May 18, 2021. Date of publication May 25, 2021; date of current version July 20, 2021. This work was supported in part by China Postdoctoral Science Foundation under Grant 2020M671826, in part by Zhejiang Lab Program under Grants 2020ND8AD01, 2020ND0AA01 and 2018DG0ZX01, and in part by the Zhejiang University special scientific research fund for COVID-19 prevention and control under Grant 2020XGZX051. *(Titinunt Kitrungrotsakul and Qingqing Chen are co-first authors.) (Corresponding authors: Hongjie Hu; Lanfen Lin; and Yen-Wei Chen.)*

Titinunt Kitrungrotsakul is with the Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou 311121, China (e-mail: titinunt121@hotmail.com).

Qingqing Chen is with the Department of Radiology, Sir Run Run Shaw Hospital, Hangzhou, Zhejiang 310016, China (e-mail: qingqingchen@zju.edu.cn).

Huitao Wu is with Intelligent Healthcare Team, Baidu Inc., Beijing 100085, China (e-mail: wuhuitao@zhejianglab.com).

Yutaro Iwamoto is with the Graduate School of Information Science and Engineering, Ritsumeikan University, Shiga 525-0058, Japan (e-mail: yiwamoto@fc.ritsumei.ac.jp).

Hongjie Hu, Wenchao Zhu, Chao Chen, Fangyi Xu, and Yong Zhou are with the Department of Radiology, Sir Run Run Shaw Hospital, Hangzhou, Zhejiang 310016, China (e-mail: hongjiehu@zju.edu.cn; zhuwenchao503@163.com; Y215180005@zju.edu.cn; 21718390182090@zju.edu.cn; zhyong@zju.edu.cn).

Lanfen Lin is with the College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China (e-mail: llf@zju.edu.cn).

Ruofeng Tong is with the Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou 311121, China, and also with the College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China (e-mail: trf@zju.edu.cn).

Jingsong Li is with the Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou 311121, China, and also with the College of Biomedical Engineering & Instrument Science, Zhejiang University, Hangzhou 310027, China (e-mail: ljs@zju.edu.cn).

Yen-Wei Chen is with the Graduate School of Information Science and Engineering, Ritsumeikan University, Shiga 525-0058, Japan, and with the Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou 311121, China, and also with the College of Computer Science and Technology, Zhejiang University, Hangzhou 310027, China (e-mail: chen@is.ritsumei.ac.jp).

Digital Object Identifier 10.1109/JBHI.2021.3082527

I. Introduction

SARS-COV-2 pneumonia, or COVID-19 pneumonia, has spread around the world since December 2019. By January 2020, the World Health Organization (WHO) declared COVID-19 pneumonia a Public Health Emergency of International Concern. According to a report containing 72,314 cases from the Chinese Center for Disease Control and Prevention, approximately 81% of patients with COVID-19 were classified as mild; however, 14% and 5% were recognized as severe and critical, respectively. In total, the average mortality rate was 2.3%, and the mortality rate of critical cases was approximately 49% [1]. Quantitative evaluation of COVID-19 severity is important for adequate medical treatment. The CT involvement score (the proportional volume of the infection area in the lung) which measures the infection progress in individual lung lobes, is an essential measure of the severity of COVID-19 [2]. CT is extensively used as a key complement for diagnosing COVID-19; moreover, high sensitivity was reported in patients infected by COVID-19 [3]–[7]. Some typical CT images of an

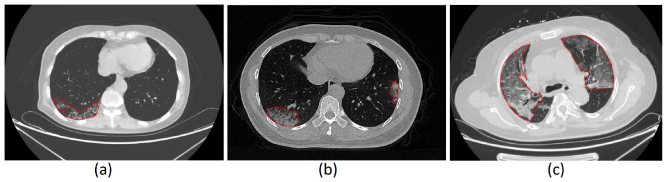

Fig. 1. Sample slices of infected lungs obtained from CT images. The red mask indicates the infected areas. a) and b) the infected lungs damaged by COVID-19 in the early stage. c) the infected lung of a severe case.

infected lung are shown in Fig. 1. The red masks indicate the infected areas. Figs. 1(a) and 1(b) show the infected lungs in the early stage, and Fig. 1(c) shows the infected lung of a severe case. Infected area segmentation of COVID-19 in CT images is an essential preprocessing step for quantitative evaluation of COVID-19 severity. Due to the large volume of CT images, COVID-19 pneumonia diagnosis has greatly increased the time and workload requirements of physicians and radiologists [8]. It takes experienced radiologists several hours to complete the localization and accurate masking of lung lesions in a single CT volume. The analysis and diagnosis of CT volume have been limited by finding and locating lung lesions for qualitative or semiquantitative evaluation [9]–[12].

Recently, deep learning-based segmentation methods such as U-Net and its variations [13]–[15], DeepLab [16], [17], and the fully convolutional network (FCN) [51] have achieved state-of-the-art performance. Deep learning-based segmentation methods (such as U-Net, U-Net++, and V-Net) have also been utilized for the segmentation of COVID-19 lesions [40]–[42]. An Inf-Net for COVID-19 lung CT infection segmentation using implicit reverse attention and explicit edge attention to improve the infected regions was proposed by Fan *et al.* [43]. Although these automatic methods are useful, the segmentation accuracy is not high enough (approximately 85%) due to variations among COVID-19 infected areas (such as shape and size variations) and similarities of COVID-19 and non-COVID-19 areas.

Additionally, various U-Net methods [44], [45] have been proposed as an improved version of skip connection attention. The propagation of local information from lower features to high-level features is used to retrieve spatial information. These methods pass high-level features to the attention gate and combine the same features with the feature attention output. With those processes, the gradient will be very small, similar to the problem that was mentioned in the ResNet paper [48].

We propose an interactive attention refinement network (attention RefNet) to improve the segmentation accuracy of COVID-19 infected areas. The network is combined with a backbone segmentation network (e.g., U-Net) to refine the initial segmentation result through the backbone segmentation network. The proposed technique is motivated by our preliminary research [49]. The main contributions and advantages of the proposed method are as follows.

1) We proposed an interactive attention refinement network (Attention RefNet), which can relate to any segmentation network. We also presented an automatic seed point generation technique for the training process leading to the end-to-end training of the refinement network and segmentation network. Via the attention RefNet, we can enhance the segmentation accuracy and generalizability of the model (i.e., its applicability to multicenter data).
2) We proposed a skip connection attention (SCA) module generated and combined from various level features to improve the important features in both segmentation and refinement networks for initial segmentation and refinement. Using channel and spatial attention, the SCA is formed. The U-Net with the proposed SCA can be considered an enhanced version of U-Net where the skip connection is replaced by the SCA. The SCA map is also utilized in the refinement network to capture the important features.
3) We proposed a seed point module (SPM), which is attached to the refinement network. The SPM is used to transform the seed points inputted by users to obtain useful information.
4) We conduct extensive experiments on thee COVID-19 datasets, including a private dataset from Chinese hospitals, a public dataset (COVID-19-CT-Seg) [35] and another public dataset (MICCAI COVID-19) [36], to verify the effectiveness of our proposed method. The private dataset is a multicenter dataset collected from multiple hospitals in China and was used to verify the generalizability of the proposed method. We compared our attention RefNet against various state-of-the-art approaches for COVID-19 segmentation. We demonstrate that the proposed technique achieves higher accuracy than the other state-of-the-art methods. We also confirm the generalizability of the proposed network using our multicenter dataset. The presented method can still achieve high segmentation accuracy, even when applied to other datasets collected in various hospitals not included in the training dataset.

II. Related Work

In this section, we present the research related to the present work, including COVID-19 infected area segmentation and interactive segmentation.

A. COVID-19 Infected Area Segmentation

Medical image segmentation is a popular technique for computer-aided diagnosis (CAD) systems. The approaches include learning-based and nonlearning-based methods. Since learning-based methods, especially deep learning-based methods, have achieved state-of-the-art segmentation performance, we focused on deep learning-based methods. Several deep learning-based methods, such as U-Net++ [18], 3D U-Net [53] and U-Net3+ [19], have been proposed for image segmentation. In U-Net, which is extensively utilized for biomedical image segmentation, the concept of an encoder and a decoder with a skip connection is used. The present methods (i.e., U-Net,

U-Net++, and V-Net) were also used for COVID-19 segmentation [38]–[40]. Fan proposed Inf-Net, a COVID-19 lung CT infection segmentation network [43]. This network uses implicit reverse attention and explicit edge attention to improve the infected regions. Though these automatic methods are useful, the segmentation accuracy is not high enough (approximately 85%) due to variations of the COVID-19 infected area (such as shape and size variations) and the similarity of COVID-19 and non-COVID-19 areas. Interactive refinements are required to improve the segmentation accuracy.

B. Interactive Segmentation

Considering the problem of the accuracy of automatic segmentation methods, interactive segmentation methods were proposed. Among the conventional interactive segmentation methods, graph cuts [28] are common in both natural and medical imaging fields. Graph cuts represent an image as a graph with nodes and edges representing the pixels and pixel relationships of the image. The objective of graph cuts is to divide a graph into two parts by minimizing the cutting costs. A large memory usage problem is mainly faced by graph cuts when applied to medical data.

Along with conventional methods, deep learning techniques have been used for solving the problem of segmentation with interactive segmentation techniques. Deep Extreme Cut (DEXTR) [54] is a recently developed deep learning technique using the concept of interactive image segmentation. To perform segmentation using DEXTR, the user should provide the four most extreme points of the object. The Gaussian distribution technique is employed to transform the seed points before sending them to the deep learning network.

Another deep learning method is DeeplGeoS [55], which uses the concept of interactive segmentation. DeeplGeoS generates an initial segmentation from the backbone segmentation network. Seed points from the user are utilized as refinement criteria in their network. The refinement network has an architecture similar to the backbone segmentation network; however, it has three extra input channels, including two geodesic distances of foreground and background seed points and initial segmentation from the first network. The refinement is performed using seed points as the extra channel with the initial segmentation result. Via this refinement technique, low-level information of the input image will be overlooked.

Regarding seed point generators, most existing seed point generators are designed for nondeep learning methods such as region growth. In [46] and [47], Juan Shan *et al.* designed a seed point generator for region growth. Their method automatically generated a seed point for initial segmentation by calculating the image histogram. The threshold is used to binarize the image into background and foreground. The performance is dependent on the threshold.

III. Proposed Method

In this research, we present an interactive refinement network with skip connection attention and a seed point map for the infected area segmentation of COVID-19, and the method is called "Attention RefNet". The network is utilized to refine the results of the segmentation backbone. In our Attention-RefNet, the seed point map and skip connection attention are employed to identify the infected area.

A. Overview

Fig. 2 presents the architecture of the proposed "Attention RefNet," which includes two networks. The first network is the modified segmentation backbone network, which is utilized to extract the initial segmentation features. The second network is the interactive refinement network, which is employed to refine the segmentation features and perform the segmentation on the refined features. We also add seed points as interactive inputs of the refinement network to further improve the segmentation results. The output (feature maps) of the segmentation backbone is utilized as the input of the refinement network. The two networks are trained end-to-end and are explained along with our main contributions in detail in the following subsections.

Regarding the backbone architecture, we utilized a typical encoder-decoder network such as U-Net or an FCN. In the encoder part, we keep the layer structure the same as in the original work. However, in the decoder part, we employ and replace the skip connection with Skip Connection Attention (SCA), which is designed to emphasize the important and sensitive areas of infected cells. The SCA features were generated from high-level and low-level network features.

Regarding the refinement architecture (RefineNet), the segmentation results of a test image are refined based on the backbone features. Refinement attention, which results from the backbone feature, is utilized in this network to emphasize the important parts of the infected areas. A multiscale feature technique is used for accurate segmentation of the infected areas owing to the multiple sizes of infected areas.

The characteristics of COVID-19-infected areas are scattered throughout the lung region, making the segmentation process here different from liver, lung, or organ segmentation. In this network, interactive seed point (line or dot) refinement is used to locate the incorrect segmented region from the network and present more information to our refinement network.

The details of skip connection attention (SCA) for the backbone network, RefineNet, and interactive seed points are presented in the next subsections.

B. Segmentation Network (U-Net) With Skip Connection Attention

To enhance the information of the infected area, the concept of spatial and channel attention modules is applied. We propose skip connection attention to combine information from low-level features and high-level features. The skip connection attention module is utilized to replace an original skip connection. The component of our SCA module is represented in Fig. 3.

Fig. 3 shows the process of the SCA module containing two parts. In the upper part, SCA performs global pooling ($GlobalPooling(\cdot)$) on low-level features (f_l) to transform the

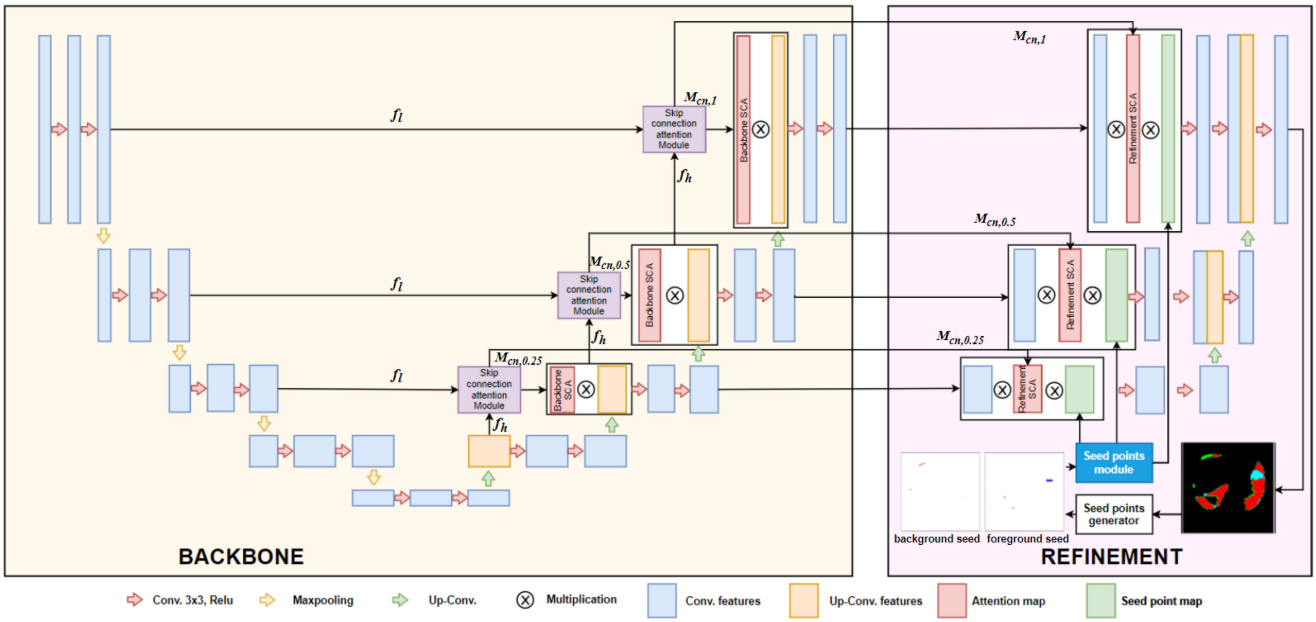

Fig. 2. The architecture overview of our Attention-RefNet. The yellow part represents the backbone network with modifying the skip connection attention (SCA) and the pink part shows the refinement network with the SCA and seed point module (SPM).

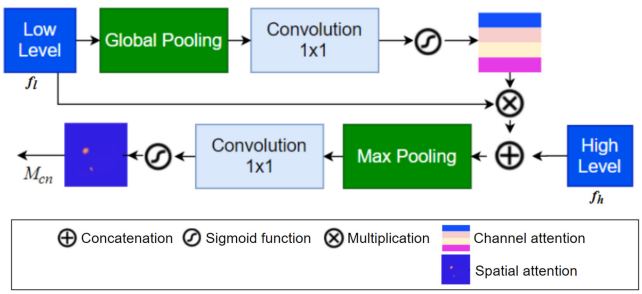

Fig. 3. Diagram of the proposed skip connection attention (SCA).

feature image from $h \times w \times c$ to $1 \times 1 \times c$. After global pooling, a 1×1 convolution ($Conv_{1\times1}(\cdot)$) is applied. The sigmoid function (σ) is employed to determine the corresponding matrix affecting the label. The output M_c of the upper part of the SCA is calculated as follows:

$$M_c(f_l) = \sigma(Conv_{1\times1\times1}(GlobalPooling(f_l))) \quad (1)$$

The other components in SCA, except for the upper part in Fig. 3(lower), are utilized to calculate a spatial attention map through the relationship of features to comprehend the network and find the informative area. The output M_c is utilized for elementwise multiplication with low-level features and then concatenated with high-level features (f_h). Max pooling ($MaxPooling(\cdot)$) and 1×1 convolutional ($Conv_{1\times1}(\cdot)$) are used. The ultimate output of skip connection attention map M_{cn} is created utilizing the sigmoid function as follows:

$$M_{cn}(f_l, f_h) = \sigma(Conv_{1\times1\times1}(MaxPooling([M_c(f_l) \otimes f_l, f_h]))) \quad (2)$$

In our proposed network, we apply and replace conventional skip connections with SCA un the backbone network. The original skip connection is utilized to gather low-level features. However, in our SCA, we employ the result of SCA and perform elementwise multiplication with a high-level feature since the new output features f_b are calculated as:

$$f_b = M_{cn}(f_l, f_h) \otimes f_h \quad (3)$$

C. Interactive Refinement Network

The right part in Fig. 2 represents the proposed Attention RefNet architecture, which is utilized to refine the feature map results from the backbone segmentation network (the left part in Fig. 2). The feature map results output from the backbone segmentation network are inserted into the refinement network along with seed points for final segmentation (refinement).

1) Multi-Scale: To address small and sensitive regions, the concept of a pyramid network [50] is applied to increase the segmentation performance for a small infected area. Adding the 1×1 convolution, initial feature maps are generated before the last three convolutional layers, where the size of the feature map results of the last three layers are $1\times$, $0.5\times$, and $0.25\times$ of the original input image size, respectively. Each feature map result is merged up at each scale together through the concatenation method and convolution (Up-Conv.). Hence, the size of the final segmentation result of the second refinement network is the same as the size of the input image.

2) Attention Refinement: To ensure that important features from the backbone are not missed during the refinement process in the refinement network, a skip connection attention (SCA) map (M_{cn}) is used for attention refinement. Hence, the performance of the refinement network is enhanced by performing elementwise multiplication on the output feature f_o^r of the

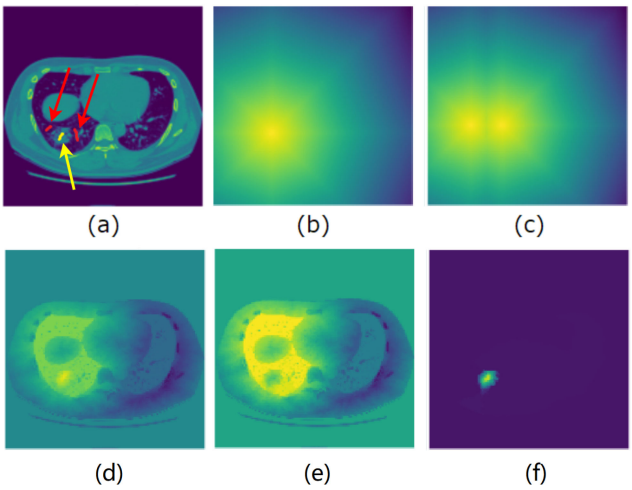

Fig. 4. (a) The seed points in the original image, where red indicates the background seeds and yellow indicates foreground seeds; (b) the foreground Euclidian distance; (c) the background Euclidian distance; (d) the foreground geodesic distance; (e) the background geodesic distance; and (f) the seed point map generated from Eq. 5.

backbone network at each scale of the network, where r is a scale factor ($1\times$, $0.5\times$, and $0.25\times$).

$$\hat{f}_o^r = M_{cn}(f_l^r, f_h^r) \otimes f_o^r \tag{4}$$

A 3×3 convolution is performed after elementwise multiplication with SCA. Concatenation with a lower scale factor feature is applied to merge multiscale features, and another 3×3 convolution is used to the new features $\hat{f}_o^r$ to create the final segmentation result.

3) Seed Point: We use seed points (both foreground and background seed points) given by users for further refinement of the segmentation result. Using the geodesic distance, we transformed the seed points to seed point maps. In our preliminary interactive segmentation method [49], we utilized the Euclidian distance to convert the seed points to seed point maps where the texture information of the image is lost, resulting in false refinement or segmentation. Examples of Euclidean distance maps reflecting only the distance to the seed point (not any image) are presented in Figs. 4(b) and 4(c). However, we utilize the geodesic distance [55] to generate the seed point map $f_d(x)$ considering the image information, as shown in Eqs. 5–7

$$f_d(x) = G_{\mathcal{F}}(x) - G_{\mathcal{B}}(x) \tag{5}$$

$$G_p(x) = \min_{j\in\mathcal{P}} \mathcal{D}_j(x) \tag{6}$$

$$\mathcal{D}_j(x) = \int_C |\nabla I(s) \cdot V(s)|\, ds \tag{7}$$

where $G_{\mathcal{F}}$ and $G_{\mathcal{B}}$ denote the geodesic distance maps of the foreground and background, respectively. $D_j(x)$ is the geodesic distance between pixel x and the j-th seed point. C denotes the minimum path between pixel x and the j-th seed point, and s is a pixel on path C. $V(s)$ is a unit vector that is tangent to the direction of the path. Examples of geodesic distance maps are presented in Figs. 4(d) and 4(e). The figures show that the geodesic distance map reflects the image structure.

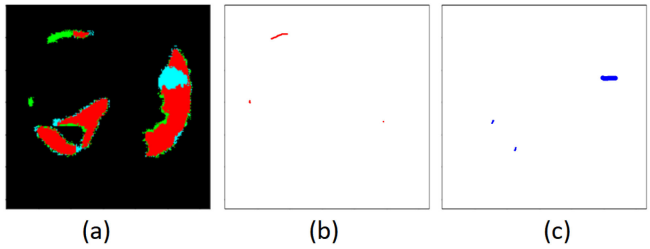

Fig. 5. Automatic seed point generation for the training model. The green color mask is oversegmented, and the blue color mask is undersegmented. (b) The background seed points (red line) generated from an oversegmentation mask (a). (c) The foreground seed points (blue line) generated from the undersegmentation mask (a).

The seed point map (f_d^r) for scale r is applied to Eq. 4 by performing elementwise multiplication, where r denotes a scale factor ($1\times$, $0.5\times$, and $0.25\times$).

$$\hat{f}_o^r = M_{cn}(f_l^r, f_h^r) \otimes f_o^r \otimes f_d^r \tag{8}$$

4) Automatic Seed Point Generation for Training: To reduce the user interactions (provision of seed points), we propose an automatic seed point generation technique for the training phase. We used both the initial segmentation result (f_y) and the ground truth (f_x) obtained by the backbone segmentation network to generate the foreground/background points automatically. The subtraction mask (δf) is generated by the subtraction of both images, such that each pixel is either -1, 0, or $+1$. The $+1$ mask is referred to as the oversegmentation mask; conversely, the -1 mask is considered the undersegmentation mask.

$$\delta f = f_x - f_y \begin{cases} \delta f > 0, & \mathcal{F} \\ \delta f < 0, & \mathcal{B} \end{cases} \tag{9}$$

To generate seed points from those masks, we used a skeleton and utilized the skeleton lines as foreground seed points (undersegmentation mask) and background seed points (oversegmentation mask), as shown in Fig. 5.

Compared to existing methods [46], [47], we directly use the difference between the ground truth mask (f_x) and the initial segmentation result (f_y) from the segmentation network to generate seed points in the training phase. The points of $(f_x - f_y) > 0$ are used as object seed points, and the points of $(f_x - f_y) < 0$ are used as background seed points. The advantage of the proposed seed point generation method is that we can focus on the oversegmentation and undersegmentation areas and refine them.

IV. Experiments and Results

A. COVID-19 Segmentation Datasets

In this work, we assessed the proposed method on two COVID-19 segmentation datasets: a private multicenter dataset and a public dataset.

The first dataset is a COVID-19 private multicenter dataset containing 210 CT volumes of COVID-19 patients with ground

TABLE I
THE DETAIL OF OUR DATASET WITH DIFFERENT MANUFACTURE

Center	CT volume	Manufacture
Hospital1	8	SIEMENS
Hospital2	6	GE
Hospital3	184	GE
Hospital4	12	UIH
Total volume	210	

TABLE II
THE NUMBER OF TRAINING, VALIDATION AND TEST SAMPLE

Center	Training	validation	Test
Hospital1	4	2	2
Hospital2	4	1	1
Hospital3	134	24	26
Hospital4	7	2	3
Total volume	149	29	32

truth (GT) segmentation masks. The CT volumes were collected from 4 hospitals in China, including Yifu (Hospital 1), Zhejiang (Hospital 2), Taizhou (Hospital 4), and Jingmen (Hospital 3), as shown in Table I. Each center uses a different scanner; however, the sizes of the CT images are the same ($512 \times 512 \times 200$). The segmentation masks were labeled by the radiologists of each hospital. From our dataset, we extracted 14,000 2D CT axial slices from 210 CT volumes. We randomly selected 29 CT volumes as validation data, 149 CT volumes as training data, and 32 CT volumes as test data, as shown in Table II, for our experiments in section IV-C and IV-D1. Furthermore, in section IV-E, the training, validation, and test data of the experiments are presented based on the hospital to assess the generalizability of the models on the segmentation task. The ethics committee of Sir Run Run Shaw Hospital approved the study, and informed consent was obtained from all patients.

The second dataset used is a public dataset (COVID-19-CT-Seg) [35]. The COVID-19-CT-Seg dataset contains 20 patients, each with only one CT volume and corresponding lung and infected area masks. The masks were labeled by two radiologists and verified by an experienced radiologist. We extracted approximately 40 2D CT axial slices from each 3D volume. The five-fold cross-validation method employed in our experiment in section IV-D2.

The last dataset is another public dataset from MICCAI (COVID-19 Lung CT Lesion Segmentation Challenge) [36]. The MICCAI dataset contains 199 CT volumes with segmentation masks. We randomly divided the 199 volumes into five groups and employ a five-fold cross-validation method to evaluate the segmentation methods. The results are shown in section IV-D3.

B. Evaluation Methods

In this research, the overall performance of each method was assessed utilizing three extensively adopted metrics. The overlap index (Dice coefficient), Sensitivity (SEN), and Positive Predictive Value (PPV) were used as the evaluation measures, and they are calculated using Eqs. 10–12, respectively.

$$DICE = \frac{2|A \cap B|}{|A| + |B|} \tag{10}$$

$$SEN = \left(\frac{A \cap B}{B}\right) \tag{11}$$

$$PPV = \left(\frac{A \cap B}{A}\right) \tag{12}$$

where A denotes the segmented infected area mask and B denotes the ground truth infected area mask. These three evaluation metrics are utilized to measure the segmentation performance.

C. Ablation Study

The ablation experiment results on the private dataset and public dataset are summarized in Tables III and IV, respectively. U-Net is the baseline network. IM1 represents the enhanced model of U-Net by replacing the skip connection with the proposed SCA but without a refinement network. IM2, IM3, and IM4 denote the improved models with refinement networks but without seed points. IM5 and IM6 are enhanced models with refinement networks and seed points.

Effectiveness of refinement: According to both Tables III and IV, a higher segmentation accuracy was obtained by the models with refinement networks (IM2-6) than by the models without refinement networks (U-Net and IM1). In comparison to the backbone segmentation network (U-Net), the segmentation accuracy (Dice) was improved by the proposed method (IM6) from 0.8226 to 0.9275 on our private dataset and from 0.8332 to 0.9137 on the public dataset.

Effectiveness of SCA: We assessed the effectiveness of the proposed SCA on the backbone network and refinement network. By comparing IM1 (with SCA) with U-Net (without SCA), we can validate the effectiveness of SCA on the backbone network. It is observed that the backbone SCA enhanced the segmentation performance for both datasets. We also conduct statistical analysis ($t-test$) between IM1 and U-Net. The p-value result was 0.0002, as shown in Table V. The results of the analysis confirmed the statistical significance of the difference between models ($<$0.05). Moreover, we can validate the effectiveness of SCA on the refinement network by comparing IM3 with IM2. The refinement of SCA also enhanced the segmentation performance for both datasets, with the $p-value$ of the difference between IM2 and IM3 being less than 0.05. Additionally, the performance comparison between IM0 and IM6 shows the improvement of the network from SCA. The segmentation accuracy (Dice) was improved from 0.8852 to 0.9275 for our private dataset and from 0.8792 to 0.9137 for the public dataset. Moreover, we implement and replace our attention part in the Convolutional Block Attention Module (CBAM) and called it IM CBAM. We compare IM CBAM with IM 6, which is our proposed attention module. The results in both tables show that our attention module achieves a Dice value of approximately 0.01.

TABLE III
ABLATION STUDIES OF OUR ATTENTION-REFNET ON OUR PRIVATE DATASET

Network	Refinement	Attention			Seed point Map		Multi-scale	Evaluation matrix		
		CBAM	Skip Connection Attention		Euclidian	Geodesic		DICE	SEN	PPV
			Backbone Attention	Refinement Attention						
U-Net	–	–	–	–	–	–	–	0.8226	0.8496	0.8115
IM 0	◯	–	–	–	–	◯	◯	0.8852	0.8929	0.8830
IM 1	–	–	◯	–	–	–	–	0.8269	0.8547	0.8101
IM 2	◯	–	◯	–	–	–	–	0.8273	0.8550	0.8029
IM 3	◯	–	◯	◯	–	–	–	0.8298	0.8531	0.8107
IM 4	◯	–	◯	◯	–	–	◯	0.8436	0.8767	0.8003
IM 5	◯	–	◯	◯	◯	–	◯	0.8761	0.8998	0.8533
IM CBAM	◯	◯	–	–	–	◯	◯	0.9174	0.9343	0.9217
IM 6	◯	–	◯	◯	–	◯	◯	0.9275	0.9388	0.9359

TABLE IV
ABLATION STUDIES OF OUR ATTENTION-REFNET ON THE COVID-19-CT-SEG DATASET

Network	Refinement	Attention			Seed point Map		Multi-scale	Evaluation matrix		
		CBAM	Skip Connection Attention		Euclidian	Geodesic		DICE	SEN	PPV
			Backbone Attention	Refinement Attention						
U-Net	–	–	–	–	–	–	–	0.8332	0.8450	0.8141
IM 0	◯	–	–	–	–	◯	◯	0.8792	0.8911	0.8808
IM 1	–	–	◯	–	–	–	–	0.8394	0.8468	0.8209
IM 2	◯	–	◯	–	–	–	–	0.8429	0.8539	0.8153
IM 3	◯	–	◯	◯	–	–	–	0.8461	0.8548	0.8337
IM 4	◯	–	◯	◯	–	–	◯	0.8601	0.8921	0.8220
IM 5	◯	–	◯	◯	◯	–	◯	0.8751	0.9008	0.8572
IM CBAM	◯	◯	–	–	–	◯	◯	0.9062	0.9133	0.9077
IM 6	◯	–	◯	◯	–	◯	◯	0.9137	0.9154	0.9190

TABLE V
THE STATISTICAL ANALYSIS ($t-Test$) RESULTS OF THE MODELS

Network	U-Net	IM1	IM2	IM3	IM4	IM5	IM CBAM	IM6
IM 0	<0.0001	-	-	-	-	-	-	-
IM 1	0.00027	-	-	-	-	-	-	-
IM 2	<0.0001	0.01086	-	-	-	-	-	-
IM 3	<0.0001	<0.0001	0.00123	-	-	-	-	-
IM 4	<0.0001	<0.0001	<0.0001	<0.0001	-	-	-	-
IM 5	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	-	-	-
IM CBAM	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	-	-
IM 6	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	-

Effectiveness of Multiscale Fusion: We investigated the effectiveness of multiscale fusion by comparing IM4 (multi-scale fusion) with IM3 (single-scale fusion). It was found that multiscale fusion can enhance the segmentation accuracy (Dice) by approximately 0.015 for both datasets.

Effectiveness of Seed Points: We assessed the effectiveness of seed points by comparing IM5 and IM6 (with seed points) with IM2, IM3, and IM4 (without seed points). According to Tables III and IV, adding seed points can considerably improve the segmentation accuracy. The proposed geodesic distance (IM6) can further enhance the segmentation accuracy for both datasets compared to the conventional Euclidean distance (IM5).

D. Segmentation Benchmark Evaluation

1) Benchmark Evaluation on Our Private Dataset: We compared our Attention-RefNet with automatic segmentation methods, including FCN [51], V-Net [52], U-Net [13], U-Net++ [18], 3D U-Net [53] and Inf-Net [43]; and interactive segmentation methods, including graph cuts [28], DEXTR [54] and DeeplGeoS [55]. All experimental results are presented in Table VI.

TABLE VI
THE QUANTITATIVE RESULTS OF INFECTED AREA SEGMENTATION ON OUR PRIVATE DATASET WITH 149 CT VOLUMES AS THE TRAINING SAMPLE, 29 CT VOLUMES AS THE VALIDATION SAMPLE, AND 32 CT VOLUMES AS THE TEST SAMPLE

Methods	Approach	DICE	SEN	PPV
FCN [51]	Automatic	0.7661	0.8023	0.7555
V-Net [52]	Automatic	0.8117	0.8234	0.8021
U-Net [13]	Automatic	0.8226	0.8496	0.8115
3D U-Net [53]	Automatic	0.8412	0.8584	0.8315
U-Net++ [18]	Automatic	0.8430	0.8590	0.8317
Inf-Net [43]	Automatic	0.8676	0.8787	0.8600
Graph Cut [28]	Interactive	0.8213	0.7851	0.8584
DEXTR [54]	Interactive	0.7811	0.8073	0.7839
DeeplGeoS [55]	Interactive	0.8996	**0.9389**	0.8750
Attention-RefNet	Interactive	**0.9275**	0.9388	**0.9359**

The performances of the automatic segmentation methods are approximately 0.82, except for the FCN, which achieves a Dice coefficient of only 0.77 since it was originally proposed for natural image segmentation. Inf-Net is the automatic segmentation that obtains the highest performance in the automatic segmentation section in Table VI. Since Inf-Net is well suited for COVID-19 segmentation and easy to adapt to other COVID-19 datasets, Inf-Net achieves a Dice coefficient of 0.87 on our dataset.

TABLE VII
THE QUANTITATIVE RESULTS OF INFECTED AREA SEGMENTATION ON THE COVID-19-CT-SEG DATASET USING 5-FOLDS TRAINING

Method	DICE	SEN	PPV
U-Net [13]	0.8422	0.8252	0.8567
U-Net++ [18]	0.8628	0.8613	0.8658
Inf-Net [43]	0.8917	0.8903	0.9021
GraphCut [28]	0.8519	0.8431	0.8622
DEXTR [54]	0.7823	0.7771	0.8041
DeeplGeoS [55]	0.8966	**0.9203**	0.8837
RefineNet [49]	0.9043	0.9001	0.9112
Attention-RefNet	**0.9137**	0.9154	**0.9190**

In our experiments, the seed points were used only one time for all interactive segmentation methods. The seed points are given manually by the user based on their observation. In our method and the DeeplGeoS method, the average seed points for each region are approximately 4-6 px, as shown in Fig. 5. For graph cuts, we manually provide seed points similar to the skeletonization method. In addition, for DEXTR, we give the seed points based on the most extreme points. Graph cuts can obtain a Dice coefficient of 0.82, although DEXTR cannot obtain good performance and yields a Dice coefficient of approximately 0.78. However, DeeplGeoS can yield a Dice coefficient of approximately 0.9. The reason for not obtaining Dice coefficient greater than 0.8 by DEXTR is that DEXTR needs the user to give the four extreme points on the boundary of every segmentation target. Compared to our Attention RefNet, our method can obtain a Dice coefficient of 0.93, which is approximately 0.1 better than that of U-Net. Additionally, it is worth mentioning that all interactive segmentations in our experiment are provided as single time seed points.

In the statistical analysis ($t-test$) of the accuracy comparisons of the proposed method and other segmentation methods, the p-values for FCN, V-Net, U-Net, 3D U-Net, U-Net++, InfNet, graph cuts, DEXTR, and DeeplGeoS were all less than 0.0001. As all p-values were less than 0.05, we statistically confirmed the superior performance of the proposed method to that of other segmentation methods.

2) Benchmark Evaluation on COVID-19-CT-Seg: In this section, we assess our proposed network in comparison with other state-of-the-art methods on COVID-19-CT-Seg using a five-fold training method (3 folds for training, 1 fold for validation and 1 fold for testing). DEXTR, DeeplGeoS, and graph cuts were the interactive segmentation methods and U-Net, U-Net++, and Inf-Net were the automatic segmentation approaches that we used for the comparisons in this section. Table VII presents the performance comparison of our technique and other state-of-the-art methods.

High Dice coefficients of at least 0.84 are obtained by the automatic deep segmentation methods U-Net, U-Net++, and Inf-Net. Compared to U-Net, U-Net++ obtains a Dice coefficient of 0.86, and Inf-Net obtains a Dice coefficient of 0.89. Nevertheless, interactive segmentation methods such as DeeplGeoS can achieve better segmentation than automatic segmentation. DeeplGeoS allows users to provide seed points to refine segmentation results on unclear boundaries. The method can obtain a Dice coefficient of approximately 0.9. Based on the previous section, the DEXTR method using four extreme points did not work well, and the Dice performance was 0.78. Comparing our Attention-RefNet with other methods, it is found that AttentionRefNet can obtain a Dice coefficient of 0.91.

TABLE VIII
THE QUANTITATIVE RESULTS OF INFECTED AREA SEGMENTATION ON THE MICCAI COVID-19 DATASET

Method	DICE	SEN	PPV
U-Net [13]	0.6139	0.6108	0.6227
U-Net++ [18]	0.6158	0.6131	0.6175
Inf-Net [43]	0.6593	0.6504	0.6618
UNet-MONAI [56]	0.6780	0.6751	0.6803
MICCAI's 1st rank	0.7352	0.7458	0.7243
MICCAI's 4th rank	0.7322	0.7414	0.7285
MICCAI's 6th rank	0.6912	0.6887	0.6953
DeeplGeoS [55]	0.7787	0.7913	0.7585
RefineNet [49]	0.7899	0.7904	0.7891
Attention-RefNet	**0.8122**	**0.8142**	**0.8101**

To evaluate the performance of our proposed method over other methods, we conduct additional statistical analysis ($t-test$). The p-values for U-Net, U-Net++, Inf-Net, graph cuts, DEXTR, DeeplGeoS, and RefineNet compared to our proposed method were all less than 0.0001. The results of the analysis confirmed the statistically significant (p-value < 0.05) superior performance of our proposed method compared to other methods.

3) Benchmark Evaluation on MICCAI Dataset: In this section, we assess our proposed network in comparison with other state-of-the-art methods on the MICCAI COVID-19 dataset (199 CT volumes with segmentation masks) [36] using a five-fold cross-validation method (3 folds for training, 1 fold for validation, and 1 fold for testing).as well as the experiments using COVID-19-CT-Seg dataset in the previous subsection. In addition to UNet, U-Net++, Inf-Net, DeeplGeos, RefineNet, we also compared the proposed method with U-Net-MONAI (the baseline method of MICCAI challenge dataset) [56] and methods of the 1st, 4nd and 6th ranks of the challenge. The results are summarized in Table VIII. The first seven methods are automatic methods and the last three methods are interactive methods. As shown in Table VIII, the method of the 1st rank in MICCAI Challenge, which uses U-Net [57] with reverse attention [58], outperformed other automatic methods, while the interactive methods outperformed the automatic methods. The proposed Attention RefNet achieved the best results in the automatic methods.

In the statistical analysis, $(t-test)$ of the accuracy comparisons of the proposed method and other segmentation methods, the $p-values$ for other methods were all less than 0.0001. We statistically confirmed the superior performance of the proposed method over other segmentation methods.

E. Evaluation of Generalizability

To validate the proposed models generalizability, we conduct two experiments on our private multicenter dataset. First, we train the model on multicenter data and test it on the data from another data center (target center) that is not included in the training data. Then, we train the model with one data point and test it on other different data points.

1) N-1 Centers Used for Training With Another Center Used for Testing: We selected one data center as the target center for the test; the other centers were used to train the model. The

TABLE IX
EVALUATION OF THE SEGMENTATION RESULT ON MULTI CENTERS AS TRAINING FOR THE TEST ON SELECTED DATA CENTER(S)

Target center(s)	Number of CT	U-Net			DeeplGeoS			Attention-RefNet		
		DICE	SEN	PPV	DICE	SEN	PPV	DICE	SEN	PPV
Hospital1	8	0.6117	0.6734	0.5481	0.8507	0.8719	0.8498	0.8555	0.8649	0.8659
Hospital2	6	0.8434	0.8505	0.8298	0.9020	0.9177	0.9095	0.9089	0.9209	0.9123
Hospital3	184	0.5313	0.6720	0.4987	0.8001	0.8497	0.7751	0.8272	0.8452	0.8015
Hospital4	12	0.6395	0.6711	0.5918	0.8725	0.8988	0.8651	0.8769	0.8873	0.8773

TABLE X
EVALUATION OF THE SEGMENTATION RESULT TRAINING MODEL ON THE HOSPITAL3 DATA

Target center(s)	Number of CT	U-Net			DeeplGeoS			Attention-RefNet		
		DICE	SEN	PPV	DICE	SEN	PPV	DICE	SEN	PPV
Hospital1	8	0.5987	0.6789	0.5356	0.8484	0.8690	0.8327	0.8578	0.8626	0.8597
Hospital2	6	0.8571	0.8739	0.8502	0.9038	0.9319	0.8941	0.9118	0.9333	0.9061
Hospital4	12	0.6232	0.6715	0.5881	0.8527	0.8709	0.8477	0.8711	0.8854	0.8619

segmentation performance in this section is calculated based on a slice-by-slice segmentation of U-Net and AttentionRefNet (IM6). According to Table IX, although we used multicenter data for training, the segmentation performance of the conventional segmentation model (U-Net) will be reduced significantly if the target center is not included in the training centers. Note that hospital 2 achieved better segmentation performance with U-Net, DeeplGeoS and Attention-RefNet, where the Dice performances were 0.87 and 0.88 for DeeplGeoS and Attention-RefNet, respectively. The main reason is that more than 90% of the training samples are from hospital 3. Hospital 2 and hospital 3 use scanners produced by the same manufacturer, and the quality and specificity of the CT volumes are close to each other.

In contrast to the conventional U-Net, the proposed AttentionRefNet and DeeplGeos can significantly enhance the segmentation performance since the target center is not included among the training centers. It should be noted that when we used hospital 3 as the testing sample, AttentionRefNet obtained a lower Dice coefficient, only 0.53, than those obtained in other experiments. The main reason is that the number of training samples was lower in this experiment, and the training sample included scans taken with equipment from multiple manufacturers.

2) Single-Center as Training for Other Centers: We also conducted another experiment to assess the effectiveness of the refinement network on addressing the generalizability problem. We used only data from hospital 3 (single center), which is a dominant center in our dataset (Table I), for training and evaluated the network performance using data of other centers (hospital 1, hospital 2, and hospital 4).

Table X shows high performance on data from hospital 2 for U-Net, DeeplGeos and Attention-RefNet, with a minimum Dice score of approximately 0.85. The hospital 1 data are generated from a Siemens scanner, and the hospital 4 data are obtained from a UIH scanner, which cannot obtain good results in the conventional U-Net model. However, both interactive methods can enhance the performance of data from both centers to a Dice coefficient of 0.85 for hospital 1 and a Dice coefficient of up to 0.87 for hospital 4 when using our proposed method.

As for the experimental results from both Table Table IX and Table X, our technique can be utilized to solve the problem of low performance at various centers through interactive refinement.

V. DISCUSSION AND CONCLUSION

According to Tables VI and VII, automatic segmentation methods such as U-Net, U-Net++, and Inf-Net can yield average performance for the COVID-19 segmentation task. The performance of those approaches is based on the quality of the training sample masks and network designs. Fig. 6 presents the visualization of segmentation results from U-Net, U-Net++ and Attention-RefNet on our private dataset. Over- and undersegmentation occur in U-Net. In Fig. 6(c), the first, third, and fifth rows show oversegmentation, and the second row shows undersegmentation. The U-Net results in the fourth row include both over- and undersegmentation. Comparing U-Net++ with U-Net, the results show that the problem of over- and undersegmentation can be solved. Nevertheless, U-Net++ still cannot present a good segmentation result on the fourth row image input since the fragment-like infected area causes segmentation by the U-Net++ model to fail, similar to the first row in some areas. The seed points are used in our Attention-RefNet as a segmentation constraint. We can separately place the foreground seed points on the infected area on each fragmented area and present background seed points to separate fragment areas. The segmentation results of Attention-RefNet are presented in Fig. 6(e), indicating the improved segmentation result from U-Net.

Furthermore, the performance of interactive segmentation methods is dependent on training sample masks, network designs, and the seed points. For example, DEXTR is an interactive segmentation technique with performance lower than those of automatic segmentation methods for the COVID-19 dataset. The main reason is that when using the DEXTR method, the user should provide the four extreme points on the boundary. Nonetheless, the boundary between infected and noninfected areas is unclear. The DEXTR method requires the four extreme points, includes some parts of the noninfected areas and ignores some parts of the infected areas. In addition to the difficulty of seed point interaction usage, DEXTR becomes more difficult and complex to use on the COVID19 dataset due to the fragmentation of COVID-19-infected areas. Users should provide four extreme points on every fragment.

In this work, we propose an interactive segmentation network termed Attention-RefNet. The main objective is to refine the

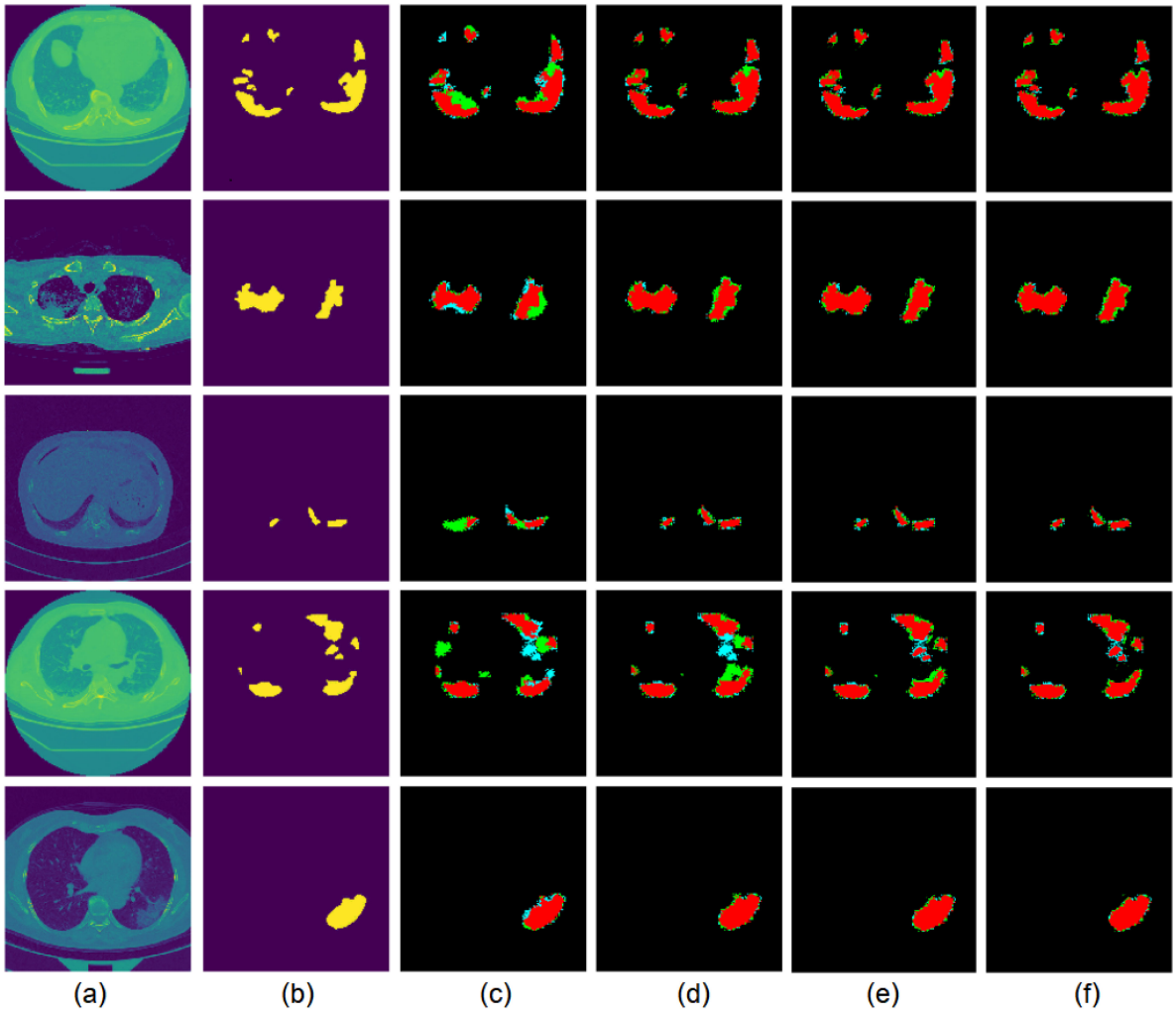

Fig. 6. The visualization of the segmentation methods on COVID-19 data (private data). (a) input lung image with infected area, (b) manual segmentation, (c) segmentation results from U-Net, (d) segmentation results from U-Net++, (e) segmentation results from DeeplGeoS and (f) segmentation from Attention-RefNet. The red pixels are correctly segmented, the green pixels are oversegmented, and the blue pixels are undersegmented.

segmentation results generated from automatic segmentation backbone networks. The automatic segmentation backbone network is the first part of our proposed network that is utilized to generate the initial segmentation. The refinement network is the second part used to refine segmentation results by obtaining the information from user interaction. The network can utilize feature information from the image and seed points from the user to refine the segmentation result through combining skip connection attention (SCA) and a seed point module (SPM). The network can generate seed points by itself over the training phase while requiring the seed points from the user only in the test phase. High segmentation accuracy as measured by the Dice coefficient can be obtained according to the experimental results on our private COVID-19 dataset comprising data for approximately 200 CT images and on a public dataset (COVID-19-CT-Seg). The segmentation results on both COVID-19 datasets revealed that the proposed technique outperforms current state-of-the-art methods.

REFERENCES

[1] W. J. Guan *et al.*, "Clinical characteristics of coronavirus disease 2019 in China," *New Engl. J. Med*, vol. 382, pp. 1708–1720, 2020.
[2] J. Zhang *et al.*, "Relationship of chest CT score with clinical characteristics of 108 patients hospitalized with COVID-19 in wuhan, china," *Respir. Res.*, vol. 21, p. 180, Jul. 2020, doi: 10.1186/s12931-020-01440-x.
[3] C. Hari *et al.*, "COVID-19 pneumonia: A review of typical CT findings and differential diagnosis," *Diagn. Interventional Imag.*, vol. 101, pp. 263–268, 2020.
[4] X. Peng, X. Xu, Y. Li, L. Cheng, X. Zhou, and B. Ren, "Transmission routes of 2019-ncov and controls in dental practice," *Int. J. Oral Sci.*, vol. 12, no. 1, pp. 1–6, 2020.
[5] S. Tian, W. Hu, L. Niu, H. Liu, H. Xu, and S.-Y. Xiao, "Pulmonary pathology of early phase 2019 novel coronavirus (COVID-19) pneumonia in two patients with lung cancer," *J. Thoracic Oncol.*, vol. 15, no. 5, pp. 700–704, 2020
[6] C. Sohrabi *et al.*, "World health organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *Int. J. Surg.*, vol. 76, pp. 71–76, 2020.
[7] S. Mohagheghi, M. Alizadeh, S. M. Safavi, A. H. Foruzan, and Y.-W. Chen, "Integration of CNN, CBMIR, and visualization techniques for diagnosis and quantification of COVID-19 disease," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 6, pp. 1873–1880, 2021.

[8] R. Alizadehsani *et al.*, "Risk factors prediction, clinical outcomes, and mortality of COVID-19 patients," *medRxiv*, vol. 93, no. 4, pp. 2307–2320, Apr. 2021.

[9] D. Colombi *et al.*, "Well-aerated lung on admitting chest CT to predict adverse outcome in COVID-19 pneumonia," *Radiology*, vol. 296, no. 2, pp. E86–E96, 2020.

[10] K. Li *et al.*, "CT image visual quantitative evaluation and clinical classification of coronavirus disease (COVID-19)," *Eur. Radiol.*, vol. 30, no. 8, pp. 4407–4416 2020.

[11] F. Pan *et al.*, "Time course of lung changes on chest CT during recovery from 2019 novel coronavirus (COVID-19) pneumonia," *Radiology*, vol. 295, no. 3, pp. 715–721, Jun. 2020.

[12] M. Prokop *et al.*, "CO-RADSCa categorical CT assessment scheme for patients with suspected COVID-19: Definition and evaluation," *Radiology*, vol. 296, no. 2, Apr. 2020.

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

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

[15] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention U-net for lesion segmentation," *IEEE 16th Int. Symp. Biomed. Imaging (ISBI 2019)*, 2019, pp. 683–687, *arXiv:1810.07842*.

[16] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected CRFs," in *Proc. Int. Conf. Learn. Representations*, pp. 1–14, 2015.

[17] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–8484, Apr. 2018, *arXiv:1606.00915*.

[18] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++ : A nested U-net architecture for medical image segmentation," *Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support*, Sep. 2018, pp. 3–11, Art. no. 11045.

[19] H. Huang *et al.*, "UNet 3+ : A. full-scale connected UNet for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 1055–1059

[20] H. R. Roth *et al.*, "Hierarchical 3D fully convolutional networks for multi-organ segmentation," Apr. 2017, *arXiv:1704.06382*.

[21] K. He, X. Cao, Y. Shi, D. Nie, Y. Gao, and D. Shen, "Pelvic organ segmentation using distinctive curve guided fully convolutional networks," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 585–595, Feb. 2019.

[22] X. Liu *et al.*, "Automatic organ segmentation for CT scans based on super-pixel and convolutional neural networks," *J. Digit. Imag.*, vol. 31, 2018, Art. no. 748C760.

[23] M. Larsson, Y. Zhang, and F. Kahl, "Robust abdominal organ segmentation using regional convolutional neural networks," *Appl. Softw. Comput.* vol. 70, pp. 465–471, 2018.

[24] Y. Tang, "Deep learning using linear support vector machines," Feb. 2015, *arXiv:1306.0239v4*.

[25] R. Adams and L. Bischof, "Seeded region growing," *Pattern Anal. Mach. Intell.* vol. 16, no. 6, pp. 641–647, 1994.

[26] Z. G. Pan and J. F. Lu, "A bayes-based region-growing algorithm for medical image segmentation," *J. Comput. Sci. Eng.*, vol. 9, no. 4, pp. 32–38, 2004.

[27] S. Zhou, J. Wang, S. Zhang, Y. Liang, and Y. Gong, "Active contour model based on local and global intensity information for medical image segmentation," *Neurocomputing*, vol. 186, pp. 107–118, 2016.

[28] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *Int. J. Comput. Vis.*, vol. 70, no. 2, pp. 109–131, 2006.

[29] L. Grady, T. Schiwietz, S. Aharon, and R. Westermann, "randomwalks for interactive organ segmentation in two and three dimensions: Implementation and validation," *MICCAI*, vol. 8, no. 2, pp. 773–780, Oct. 2005.

[30] Y. Yuan *et al.*, "Hybrid method combining superpixel, random walk and active contour model for fast and accurate liver segmentation," *Computerized Med. Imag. Graph.*, vol. 70, pp. 119–134, 2018.

[31] T. Okada *et al.*, "Automated segmentation of the liver from 3D CT images using probabilistic atlas and multi-level statistical shape model," *J. Academic Radiol.*, vol. 63, pp. 1390–1403, 2008.

[32] C. Dong *et al.*, "Segmentation of liver and spleen based on computational anatomy models," *Comput. Biol. Med.*, vol. 67, pp. 146–160, 2015.

[33] 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, 2021.

[34] 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.

[35] M. Jun *et al.*, "COVID-19 CT lung and infection segmentation dataset," *Zenodo*, 2020.

[36] COVID-19 lung CT lesion segmentation challenge - 2020 (COVID-19–20) 2020. [Online]. Available: https://covid-segmentation.grand-challenge.org/COVID-19-20/

[37] F. Shan *et al.*, "Lung infection quantification of COVID-19 in CT images with deep learning," 2020, *arXiv:2003.04655*.

[38] S. Wang *et al.*, "A deep learning algorithm using CT images to screen for corona virus disease (COVID-19)," *medRxiv*, Eur Radiol. Feb. 2021, pp. 1–9.

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

[40] T. Zhou, S. Canu, and S. Ruan, "Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism," *Int. J. Imag. Sys. Techn.*, Nov. 2020, *arXiv:2004.06673*, doi: 10.1002/ima.22527.

[41] 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*.

[42] S. Rajaraman, J. Siegelman, P. O. Alderson, L. S. Folio, L. R. Folio, and S. K. Antani, "Iteratively pruned deep learning ensembles for COVID-19 detection in chest X-rays," *IEEE Access*, vol. 8, pp. 115041–115050, 2020.

[43] D. P. Fan *et al.*, "INF-Net: Automatic COVID-19 lung infection segmentation from CT images," *IEEE Trans. Med. Imag.*, vol. 39, no. 8, pp. 2626–2637, Aug. 2020.

[44] T. L. B. Khanh *et al.*, "Enhancing U-Net with spatial-channel attention gate for abnormal tissue segmentation in medical imaging," *Appl. Sci.*, vol. 10, no. 17, 2020, Art. no. 5729.

[45] O. Oktay *et al.*, "Attention U-Net: Learning where to look for the pancreas," 2020, *arXiv:1804.03999*.

[46] J. Shan, H. D. Cheng, and Y. Wang, "A novel automatic seed point selection algorithm for breast ultrasound images," *Int. Conf. Pattern Recognit.*, Jan. 2009, pp. 1–4, doi: 10.1109/ICPR.2008.476133.

[47] M. A. Mohammed, M. K. A. Ghani, R. I. Hamed, M. K. Abdullah, and D. A. Ibrahim, "Automatic segmentation and automatic seed point selection of nasopharyngeal carcinoma from microscopy images using region growing based approach," *J. Comput. Sci.*, vol. 20, pp. 61–69, 2017.

[48] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *IEEE Conf. Comput. Vis. Pattern Recognition (CVPR)*, 2016, pp. 770–778, *arXiv:1512.03385*.

[49] T. Kitrungrotsakul, I. Yutaro, L. Lin, R. Tong, J. Li, and Y. W. Chen, "Interactive deep refinement network for medical image segmentation," 2020, *arXiv:2006.15320*.

[50] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6230–6239.

[51] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[52] 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.*, 2016, pp. 565–571.

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

[54] K. K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool, "Deep extreme cut: From extreme points to object segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 616–625.

[55] G. Wang *et al.*, "DeepIGeoS: A deep interactive geodesic framework for medical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1559–1572, Jul. 2019

[56] MONAI, "A U-Net model for lung lesion segmentation from CT images," 2020. [Online]. Available: https://github.com/Project-MONAI/tutorials/tree/master/3d_segmentation/challenge_baseline

[57] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation," *Nature Methods*, vol. 18, 203–211, 2021.

[58] S. Chen, X. Tan, B. Wang, and X. Hu, "Reverse attention for salient object detection," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 234–250.