# Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Scans

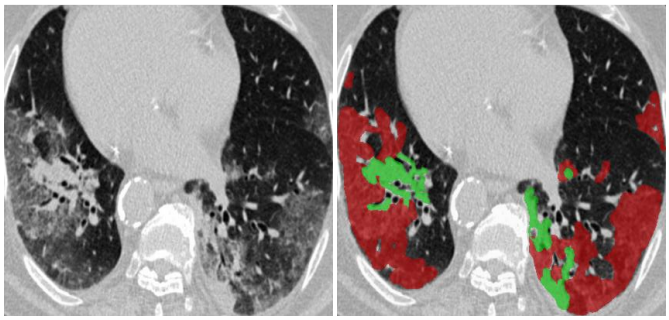
Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao

***Abstract*—Coronavirus Disease 2019 (COVID-19) spread globally in early 2020, causing the world to face an existential health crisis. Automated detection of lung infections from computed tomography (CT) images offers a great potential to augment the traditional healthcare strategy for tackling COVID-19. However, segmenting infected regions from CT scans faces several challenges, including high variation in infection characteristics, and low intensity contrast between infections and normal tissues. Further, collecting a large amount of data is impractical within a short time period, inhibiting the training of a deep model. To address these challenges, a novel COVID-19 Lung Infection Segmentation Deep Network (*Inf-Net*) is proposed to automatically identify infected regions from chest CT scans. In our *Inf-Net*, a parallel partial decoder is used to aggregate the high-level features and generate a global map. Then, the implicit reverse attention and explicit edge-attention are utilized to model the boundaries and enhance the representations. Moreover, to alleviate the shortage of labeled data, we present a semi-supervised segmentation framework based on a randomly selected propagation strategy, which only requires a few labeled images and leverages primarily unlabeled data. Our semi-supervised framework can improve the learning ability and achieve a higher performance. Extensive experiments on a COVID-19 infection dataset demonstrate that the proposed *Inf-Net* outperforms most cutting-edge segmentation models and advances the state-of-the-art performance.**

***Index Terms*—COVID-19, CT scan, infection segmentation, semi-supervised learning.**

## I. Introduction

SINCE December 2019, the world has been facing a global health crisis: the pandemic of a novel Coronavirus Disease (COVID-19) [1], [2]. According to the global case count from the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) [3] (updated April 12th, 2020), 1,777,666 identified cases of COVID-19 have been reported so far, including 108,867 deaths and impacting more than 180 countries/regions. For COVID-19 screening, the reverse-transcription polymerase chain reaction (RT-PCR) has been considered the gold standard. However, the shortage of equipment and strict requirements for testing environments limit the rapid and accurate screening of suspected subjects. Further, RT-PCR testing is also reported to suffer from high false negative rates [4]. As an important complement to RT-PCR tests, the radiological imaging techniques, *e.g.*, X-rays and computed tomography (CT), have also demonstrated effectiveness in both current diagnosis, including follow-up assessment and evaluation of disease evolution [5], [6]. Moreover, a clinical study with 1014 patients in Wuhan China, has shown that chest CT analysis can achieve 0.97 of sensitivity, 0.25 of specificity, and 0.68 of accuracy for the detection of COVID-19, with RT-PCR results for reference [4]. Similar observations were also reported in other studies [7], [8], suggesting that radiological imaging may be helpful in supporting early screening of COVID-19.



(A) (B)

Fig. 1. Example of COVID-19 infected regions (B) in CT axial slice (A), where the red and green masks denote the GGO and consolidation, respectively. The images are collected from [9].

Compared to X-rays, CT screening is widely preferred due to its merit and three-dimensional view of the lung. In recent studies [4], [10], the typical signs of infection could be observed from CT scans, *e.g.*, ground-glass opacity (GGO) in the early stage, and pulmonary consolidation in the late stage, as shown in Fig. 1. The qualitative evaluation of infection and longitudinal changes in CT scans could thus provide useful and important information in fighting against COVID-19. However, the manual delineation of lung infections is tedious and time-consuming work. In addition, infection annotation by radiologists is a highly subjective task, often influenced by individual bias and clinical experiences.

Recently, deep learning systems have been proposed to detect patients infected with COVID-19 via radiological imaging [6], [15]. For example, a COVID-Net was proposed to detect COVID-19 cases from chest radiography images [16]. An anomaly detection model was designed to assist radiologists in analyzing the vast amounts of chest X-ray images [17]. For CT imaging, a location-attention oriented model was employed in [18] to calculate the infection probability of COVID-19. A weakly-supervised deep learning-based software system was developed in [19] using 3D CT volumes to detect COVID-

Corresponding authors: *Geng Chen, Huazhu Fu, and Jianbing Shen.*

D.-P. Fan is with Inception Institute of Artificial Intelligence, Abu Dhabi, UAE, and also with College of Computer Science, Nankai University, Tianjin, China. (e-mail: dengpfan@gmail.com)

G.-P. Ji is with School of Computer Science, Wuhan University, Hubei, China. (e-mail:gepengai.ji@gmail.com)

T. Zhou, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao are with the Inception Institute of Artificial Intelligence, Abu Dhabi, UAE. (e-mails: {tao.zhou, yi.zhou, geng.chen, huazhu.fu, jianbing.shen, ling.shao}@inceptioniai.org)

TABLE I
A SUMMARY OF PUBLIC COVID-19 IMAGING DATASETS. #COV AND #NON-COV DENOTE THE NUMBERS OF COVID-19 AND NON-COVID-19 CASES. † DENOTES THE NUMBER IS PRESENTED FROM CORRESPONDING GITHUB LINK.

| Dataset | Modality | #Cov/#Non-COV | Task |
|---|---|---|---|
| COVID-19 X-ray Collection [11] | X-rays | 229† / 0 | Diagnosis |
| COVID-19 CT Collection [11] | CT volume | 20 / 0 | Diagnosis |
| COVID-CT-Dataset [12] | CT image | 288 / 1000 | Diagnosis |
| COVID-19 Patients Lungs [13] | X-rays | 70 / 28 | Diagnosis |
| COVID-19 Radiography [14] | X-rays | 219 / 2,686 | Diagnosis |
| COVID-19 CT Segmentation [9] | CT image | 110 / 0 | Segmentation |

19. A paper list for COVID19 imaging-based AI works could be found in [20]. Although plenty of AI systems have been proposed to provide assistance in diagnosing COVID-19 in clinical practice, there are only a few works related infection segmentation in CT scans [21], [22]. COVID-19 infection detection in CT scans is still a challenging task, for several issues: *1) The high variation in texture, size and position of infections in CT scans is challenging for detection.* For example, consolidations are tiny/small, which easily results in the false-negative detection from a whole CT scans. *2) The inter-class variance is small.* For example, GGO boundaries often have low contrast and blurred appearances, making them difficult to identify. *3) Due to the emergency of COVID-19, it is difficult to collect sufficient labeled data within a short time for training deep model.* Further, acquiring high-quality pixel-level annotation of lung infections in CT scans is expensive and time-consuming. Table I reports a list of the public COVID-19 imaging datasets, most of which focus on diagnosis, with only one dataset providing segmentation labels.

To address above issues, we propose a novel COVID-19 Lung Infection Segmentation Deep Network (*Inf-Net*) for CT Scans. Our motivation stems from the fact that, during lung infection detection, clinicians first roughly locate an infected region and then accurately extract its contour according to the local appearances. We therefore argue that the area and boundary are two key characteristics that distinguish normal tissues and infection. Thus, our *Inf-Net* first predicts the coarse areas and then *implicitly* models the boundaries by means of reverse attention and edge constraint guidance to *explicitly* enhance the boundary identification. Moreover, to alleviate the shortage of labeled data, we also provide a semi-supervised segmentation system, which only requires a few labeled COVID-19 infection images and then enables the model to leverage unlabeled data. Specifically, our semi-supervised system utilizes a randomly selected propagation of unlabeled data to improve the learning capability and obtain a higher performance than some cutting edge models. In a nutshell, our contributions in this paper are threefold:

- We present a novel COVID-19 Lung Infection Segmentation Deep Network (*Inf-Net*) for CT Scans. By aggregating features from high-level layers using a parallel partial decoder (PPD), the combined feature takes contextual information and generates a global map as the initial guidance areas for the subsequent steps. To further mine the boundary cues, we leverage a set of implicitly recurrent reverse attention (RA) modules and explicit edge-attention guidance to establish the relationship between areas and boundary cues.
- A semi-supervised segmentation system for COVID-19 infection segmentation is introduced to alleviate the shortage of labeled data. Based on a randomly selected propagation, our semi-supervised system has better learning ability (see § IV).
- We also build a semi-supervised COVID-19 infection segmentation (*COVID-SemiSeg*) dataset, with 100 labeled CT scans from the COVID-19 CT Segmentation dataset [9] and 1600 unlabeled images from the COVID-19 CT Collection dataset [11]. Extensive experiments on this dataset demonstrate that the proposed *Inf-Net* and *Semi-Inf-Net* outperform most cutting-edge segmentation models and advances the state-of-the-art performance. The code and dataset will be released at: https://github.com/DengPingFan/Inf-Net

## II. RELATED WORKS

In this section, we discuss three types of works that are most related to our work, including: segmentation in chest CT, semi-supervised learning, and artificial intelligence for COVID-19.

### A. Segmentation in Chest CT

CT imaging is a popular technique for the diagnosis of lung diseases [23], [24]. In practice, segmenting different organs and lesions from chest CT scans can provide crucial information for doctors to diagnose and quantify lung diseases [25], [26]. Recently, many works have been provided and obtained promising performances. These algorithms often employ a classifier with extracted features for nodule segmentation in chest CT. For example, Keshani *et al.* [27] utilized the support vector machine (SVM) classifier to detect the lung nodule from CT scans. Shen *et al.* [28] presented an automated lung segmentation system based on bidirectional chain code to improve the performance. However, the similar visual appearances of nodules and background makes it difficult for extracting the nodule regions. To overcome this issue, several deep learning algorithms have been proposed to learn a powerful visual representations [29]–[31]. For instance, Wang *et al.* [29] developed a central focused convolutional neural network to segment lung nodules from heterogeneous CT scans. Jin *et al.* [30] utilized GAN-synthesized data to improve the training of a discriminative model for pathological lung segmentation. Jiang *et al.* [31] designed two deep networks to segment lung tumors from CT scans by adding multiple residual streams of varying resolutions.

### B. Semi-supervised Learning

Due to the challenge in obtaining full labels for large-scale data, more attention has been drawn to semi-supervised learning (SSL), and the main goal of SSL is to improve model performance using a limited number of labeled data

and a large amount of unlabeled data [32]. Currently, there is increasing focus on training deep neural network using the SSL strategy [33]. These methods often optimize a supervised loss on labeled data along with an unsupervised loss imposed on either unlabeled data [34] or both the labeled and unlabeled data [35], [36]. Lee *et al.* [34] provided to utilize a cross-entropy loss by computing on the pseudo labels of unlabeled data, which is considered as an additional supervision loss. In summary, existing deep SSL algorithms regularize the network by enforcing smooth and consistent classification boundaries, which are robust to a random perturbation [36], and other approaches enrich the supervision signals by exploring the knowledge learned, *e.g.*, based on the temporally ensembled prediction [35] and pseudo label [34].

### C. Artificial Intelligence for COVID-19

Artificial intelligence technologies have been employed in a large number of applications against COVID-19 [6], [15]. Joseph *et al.* [15] categorized these applications into three scales, including patient scale (*e.g.*, medical imaging for diagnosis [37], [38]), molecular scale (*e.g.*, protein structure prediction [39]), and societal scale (*e.g.*, epidemiology [40]). In this work, we focus on patient scale applications [18], [22], [37], [38], [41]–[44], especially those based on CT scans. For instance, Wang *et al.* [37] proposed a modified inception neural network [45] for classifying COVID-19 patients and normal controls. Instead of directly training on complete CT images, they trained the network on the regions of interest, which are identified by two radiologists based on the features of pneumonia. Chen *et al.* [38] collected 46,096 CT image slices from COVID-19 patients and control patients of other disease. The CT images collected were utilized to train a U-Net++ [46] for identifying COVID-19 patients. Their experimental results suggest that the trained model performs comparably with expert radiologists in terms of COVID-19 diagnosis. In addition, other network architectures have also been considered in developing AI-assisted COVID-19 diagnosis systems. Typical examples include ResNet, used in [18], [41], and U-Net [47], used in [42]. Finally, deep learning has been employed to segment the infection regions in lung CT scans so that the resulting quantitative features can be utilized for severity assessment [43], large-scale screening [44], and lung infection quantification [15], [21], [22] of COVID-19.

## III. Proposed Method

In this section, we first provide details of our *Inf-Net* in terms of network architecture, core network components, and loss function. We then present the semi-supervised version of *Inf-Net* and clarify how to use a semi-supervised learning framework to enlarge the limited number of training samples for improving the segmentation accuracy. We also show an extension of our framework for the multi-class labeling of different types of lung infections. Finally, we provide the implementation details.

### A. Lung Infection Segmentation Network (Inf-Net)

**Overview of Network:** The architecture of our *Inf-Net* is shown in Fig. 2. As can be observed, CT images are first fed to two convolutional layers to extract high-resolution, semantically weak (*i.e.*, low-level) features. Herein, we add an edge attention module to *explicitly* improve the representation of objective region boundaries. Then, the low-level features $f_2$ obtained are fed to three convolutional layers for extracting the high-level features, which are used for two purposes. First, we utilize a *parallel partial decoder* (PPD) to aggregate these features and generate a global map $S_g$ for the coarse localization of lung infections. Second, these features combined with $f_2$ are fed to multiple *reverse attention* (RA) modules under the guidance of the $S_g$. It is worth noting that the RA modules are organized in a cascaded fashion. For instance, as shown in Fig. 2, $R_4$ relies on the output of another RA $R_5$. Finally, the output of the last RA, *i.e.*, $S_3$, is fed to a *sigmoid* activation function for the final prediction of lung infection regions. We now detail the key components of *Inf-Net* and our loss function.

**Edge Attention Module**: Several works have shown that edge information can provide useful constraints to guide feature extraction for segmentation [48]–[50]. Thus, considering that the low-level features (*e.g.*, $f_2$ in our model) preserve some sufficient edge information, we feed the low-level feature $f_2$ with moderate resolution to the proposed ***edge attention*** (EA) module to explicitly learn an edge-attention representation. Specifically, the feature $f_2$ is fed to a convolutional layer with one filter to produce the edge map. Then, we can measure the dissimilarity of the EA module between the produced edge map and the edge map $G_e$ derived from the ground-truth (GT), which is constrained by the standard Binary Cross Entropy (BCE) loss function:

$$\mathcal{L}_{edge} = -\sum_{x=1}^{w}\sum_{y=1}^{h}[G_e log(S_e) + (1-G_e) log(1-S_e)], \quad (1)$$

where $(x, y)$ are the coordinates of each pixel in the predicted edge map $S_e$ and edge ground-truth map $G_e$. The $G_e$ is calculated using the gradient of the ground-truth map $G_s$. Additionally, $w$ and $h$ denote the width and height of corresponding map, respectively.

**Parallel Partial Decoder:** Most existing medical image segmentation networks, *e.g.*, U-Net [47], U-Net++ [46], *etc*, segment interested organs/lesions using all high- and low-level features in the encoder branch. However, Wu *et al.* [51] pointed out that, compared with high-level features, low-level features demand more computational resources due to larger spatial resolutions, but contribute less to the performance. Inspired by this observation, we propose to only aggregate high-level features with a ***parallel partial decoder*** component. Specifically, for an input CT image $I$, we first extract two sets of low-level features $\{f_i, i = 1, 2\}$ and three sets of high-level features $\{f_i, i = 3, 4, 5\}$ using the first five convolutional blocks of Res2Net [52]. We then utilize the partial decoder $p_d(\cdot)$ [51], a novel decoder component, to aggregate the high-level features with a paralleled connection. The partial decoder yields a coarse global map $S_g = p_d(f_3, f_4, f_5)$, which then serves as global guidance in our RA modules.

**Reverse Attention Module:** In clinical practice, clinicians usually segment lung infection regions via a two-step pro-

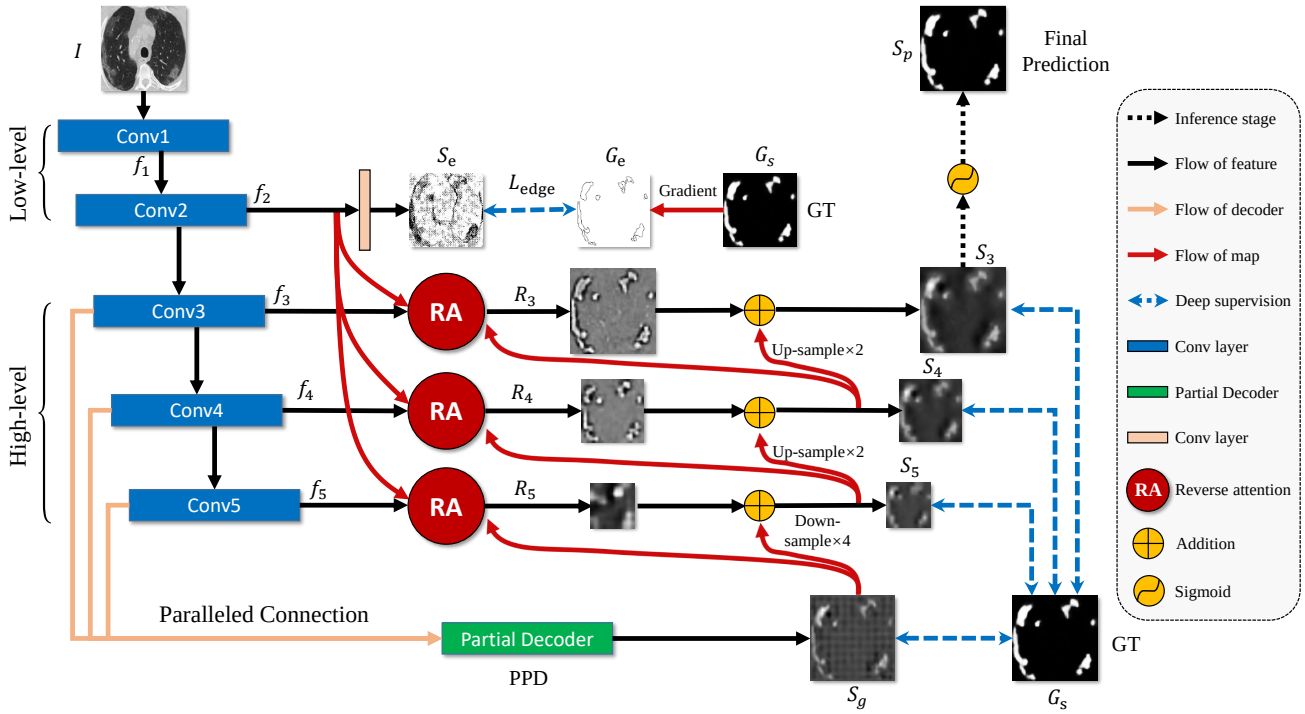



Fig. 2. The architecture of our proposed *Inf-Net* model, which consists of three reverse attention (RA) modules connected to the paralleled partial decoder (PPD). See § III-A for details.

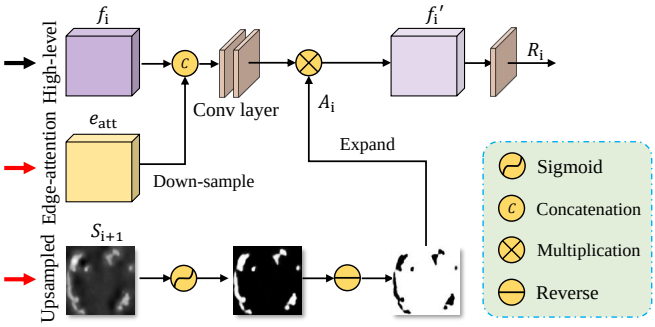



Fig. 3. Reverse Attention Module is utilized to implicitly learning edge features.

cedure, by roughly localizing the infection regions and then accurately labeling these regions by inspecting the local tissue structures. Inspired by this procedure, we design *Inf-Net* using two different network components that act as a rough locator and a fine labeler, respectively. First, the PPD acts as the rough locator and yields a global map $S_g$, which provides the rough location of lung infection regions, without structural details (see Fig. 2). Second, we propose a progressive framework, acting as the fine labeler, to mine discriminative infection regions in an erasing manner [53], [54]. Specifically, instead of simply aggregating features from all levels [54], we propose to adaptively learn the ***reverse attention*** in three parallel high-level features. Our architecture can sequentially exploit complementary regions and details by erasing the estimated infection regions from high-level side-output features, where the existing estimation is up-sampled from the deeper layer.

We obtain the output RA features $R_i$ by multiplying (element-wise $\odot$) the fusion of high-level side-output features $\{f_i, i = 3, 4, 5\}$ and edge attention features $e_{att} = f_2$ with RA weights $A_i$, *i.e.*,

$$R_i = \mathcal{C}(f_i,\ Dow(e_{att})) \odot A_i, \tag{2}$$

where Dow$(\cdot)$denotes the down-sampling operation, $\mathcal{C}(\cdot)$ denotes the concatenation operation follow by two 2-D convolutional layers with 64 filters.

The RA weight $A_i$ is de-facto for salient object detection in the computer vision community [54], and it is defined as:

$$A_i = \mathcal{E}(\ominus(\sigma(\mathcal{P}(S_{i+1})))), \tag{3}$$

where $\mathcal{P}(\cdot)$ denotes an up-sampling operation, $\sigma(\cdot)$ is a *sigmoid* activation function, and $\ominus(\cdot)$ is a reverse operation subtracting the input from matrix $E$, in which all the elements are 1. Symbol $\mathcal{E}$ denotes expanding a single channel feature to 64 repeated tensors, which involves reversing each channel of the candidate tensor in Eq. (2). Details of this procedure are shown in Fig. 3. It is worth noting that the erasing strategy driven by RA can eventually refine the imprecise and coarse estimation into an accurate and complete prediction map.

**Loss Function:** As mentioned above in Eq. (1), we propose the loss function $\mathcal{L}_{edge}$ for edge supervision. Here, we define our loss function $\mathcal{L}_{seg}$ as a combination of a weighted IoU loss $\mathcal{L}^w_{IoU}$ and a weighted binary cross entropy (BCE) loss $\mathcal{L}^w_{BCE}$ for each segmentation supervision, *i.e.*,

$$\mathcal{L}_{seg} = \mathcal{L}^w_{IoU} + \mathcal{L}^w_{BCE}. \tag{4}$$

The two parts of $\mathcal{L}_{seg}$ provide effective global (image-level) and local (pixel-level) supervision for accurate segmentation.

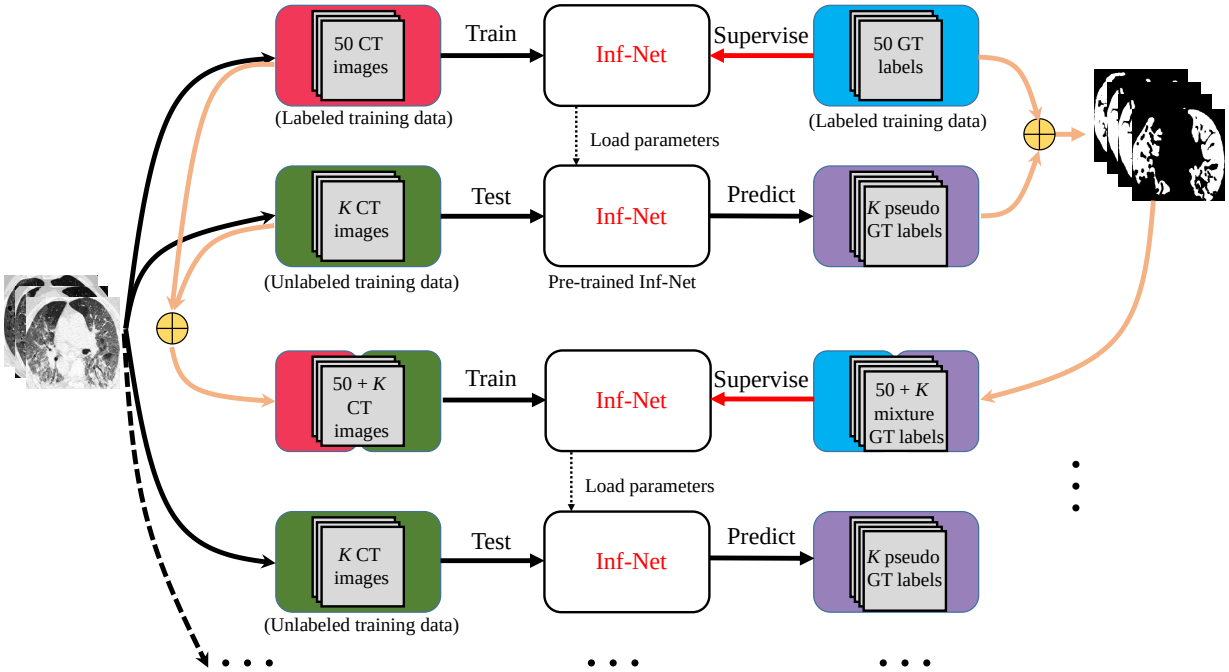



Fig. 4. Overview of the proposed *Semi-supervised Inf-Net* framework. Please refer to § III-B for more details.

Unlike the standard IoU loss, which has been widely adopted in segmentation tasks, the weighted IoU loss increases the weights of hard pixels to highlight their importance. In addition, compared with the standard BCE loss, $\mathcal{L}_{BCE}^{w}$ puts more emphasis on hard pixels rather than assigning all pixels equal weights. The definitions of these losses are the same as in [55], [56] and their effectiveness has been validated in the field of salient object detection.

Finally, we adopt deep supervision for the three side-outputs (*i.e.*, $S_3$, $S_4$, and $S_5$) and the global map $S_g$. Each map is up-sampled (*e.g.*, $S_3^{up}$) to the same size as the object-level segmentation ground-truth map $G_s$. Thus, the total loss in Eq. (4) is extended to

$$\mathcal{L}_{total} = \mathcal{L}_{seg}(G_s, S_g^{up}) + \mathcal{L}_{edge} + \sum_{i=3}^{i=5} \mathcal{L}_{seg}(G_s, S_i^{up}). \quad (5)$$

### B. Semi-Supervised Inf-Net

Currently, there is very limited number of CT images with segmentation annotations, since manually segmenting lung infection regions are difficult and time-consuming, and the disease is at an early stage of outbreak. To resolve this issue, we improve *Inf-Net* using a semi-supervised learning strategy, which leverages a large number of unlabeled CT images to effectively augment the training dataset. An overview of our semi-supervised learning framework is shown in Fig. 4. Our framework is mainly inspired by the work in [57], which is based on a random sampling strategy for progressively enlarging the training dataset with unlabeled data. Specifically, we generate the pseudo labels for unlabeled CT images using the procedure described in Algorithm 1. The resulting CT images with pseudo labels are then utilized to train our model using a two-step strategy detailed in Section III-D.

**Algorithm 1** Semi-Supervised *Inf-Net*

**Input:** Labeled training data $\mathcal{D}_{\text{Labeled}}$ and unlabeled training data $\mathcal{D}_{\text{Unlabeled}}$
**Output:** Trained *Inf-Net* $\mathcal{M}$

```
Construct a training dataset D_Training using all the labeled CT images from D_Labeled
Train our model M using D_Training
repeat
    Perform testing using the trained model M and K CT images randomly selected from D_Unlabeled, which yields network-labeled data D_Net-labeled, consisting of K CT images with pseudo labels
    Enlarge the training dataset using D_Net-labeled, i.e., D_Training = D_Training ∪ D_Net-labeled
    Remove the K testing CT images from D_Unlabeled
    Fine-tune M using D_Training
until D_Unlabeled is empty
return Trained model M
```

The advantages of our framework, called *Semi-Inf-Net*, lie in two aspects. First, the training and selection strategy is simple and easy to implement. It does not require measures to assess the predicted label, and it is also threshold-free. Second, this strategy can provide more robust performance than other semi-supervised learning methods and prevent over-fitting. This conclusion is confirmed by recently released studies [57].

### C. Extension to Multi-Class Infection Labeling

Our *Semi-Inf-Net* is a powerful tool that can provide crucial information for evaluating overall lung infections. However, we are aware that, in a clinical setting, in addition to the overall evaluation, clinicians might also be interested in the quantitative evaluation of different kinds of lung infections, *e.g.*, GGO and consolidation. Therefore, we extend *Semi-Inf-Net* to a multi-class lung infection labeling framework so that it can provide richer information for the further diagnosis and treatment of COVID-19. The extension of *Semi-Inf-Net* is based on an infection region guided multi-class labeling framework, which is illustrated in Fig. 5. Specifically, we utilize the infection segmentation results provided by *Semi-Inf-Net* to guide the multi-class labeling of different types of lung infections. For this purpose, we feed both the infection segmentation results and the corresponding CT images to a multi-class labeling network (FCN8s [58] in our case, which is popular in medical image segmentation tasks [59], [60]). This framework can take full advantage of the infection segmentation results provided by *Semi-Inf-Net* and effectively improve the performance of multi-class infection labeling.

### D. Implementation Details

Our model is implemented in PyTorch, and is accelerated by an NVIDIA TITAN RTX GPU. We describe the implementation details as follows.

**Pseudo label generation:** We generate pseudo labels for unlabeled CT images using the protocol described in Algorithm 1. The number of randomly selected CT images is set to 5, *i.e.*, $K = 5$. For 1600 unlabeled images, we need to perform 320 iterations with a batch size of 16. The entire procedure takes about 50 hours to complete.

**Semi-supervised *Inf-Net*:** Before training, we uniformly resize all the inputs to $352 \times 352$. We train *Inf-Net* using a multi-scale strategy with different scale ratios $\{0.75, 1, 1.25\}$. The Adam optimizer is employed for training and the learning rate is set to $1e-4$. Our training phase consists of two steps: (i) Pre-training on 1600 CT images with pseudo labels, which takes about 180 minutes to converge over 100 epochs with a batch size of 24. (ii) Fine-tuning on 50 CT images with the ground-truth labels, which takes about 15 minutes to converge over 100 epochs with a batch size of 16. For a fair comparison, the training procedure of ***Inf-Net*** follows the same setting described in the second step.

***Semi-Inf-Net*+FCN8s:** Follow the default settings of FCN8s [58], we resize all the inputs to $512 \times 512$ before training. The network is initialized by a uniform Xavier, and trained using an SGD optimizer with a learning rate of $1e-10$, weight decay of $5e-4$, and momentum of $0.99$. The entire training procedure takes about 45 minutes to complete.

## IV. Experiments

### A. COVID-19 Segmentation Dataset

As shown in Table I, there is only one segmentation dataset for CT data, *i.e.*, the COVID-19 CT Segmentation dataset [9][1], which consists of 100 axial CT images from different COVID-19 patients. All the CT images were collected by the Italian Society of Medical and Interventional Radiology, and are available at here[2]. A radiologist segmented the CT images using different labels for identifying lung infections. Although this is the first open-access COVID-19 dataset for lung infection segmentation, it suffers from a small sample size, *i.e.*, only 100 labeled images are available.

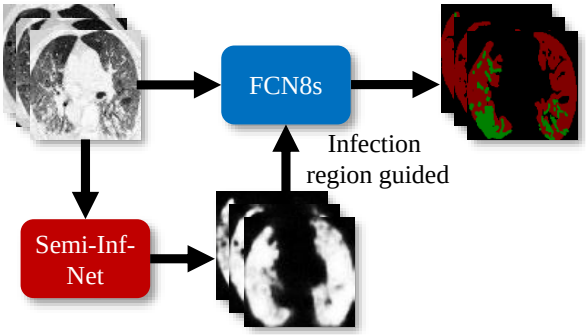



Fig. 5. Illustration of infection region guided FCN8s [58] for multi-class labeling task. We feed both the infection segmentation results provided by *Inf-Net* and the CT images into FCN8s for improving the accuracy of multi-class infection labeling.

In this work, we collected a semi-supervised COVID-19 infection segmentation dataset (*COVID-SemiSeg*), to leverage large-scale unlabeled CT images for augmenting the training dataset. We employ COVID-19 CT Segmentation [9] as the labeled data $\mathcal{D}_{\text{Labeled}}$, which consists of 45 CT images randomly selected as training samples, 5 CT images for validation, and the remaining 50 images for testing. The unlabeled CT images are extracted from the COVID-19 CT Collection [11] dataset, which consists of 20 CT volumes from different COVID-19 patients. We extracted 1,600 2D CT axial slices from the 3D volumes, removed non-lung regions, and constructed an unlabeled training dataset $\mathcal{D}_{\text{Unlabeled}}$ for effective semi-supervised segmentation.

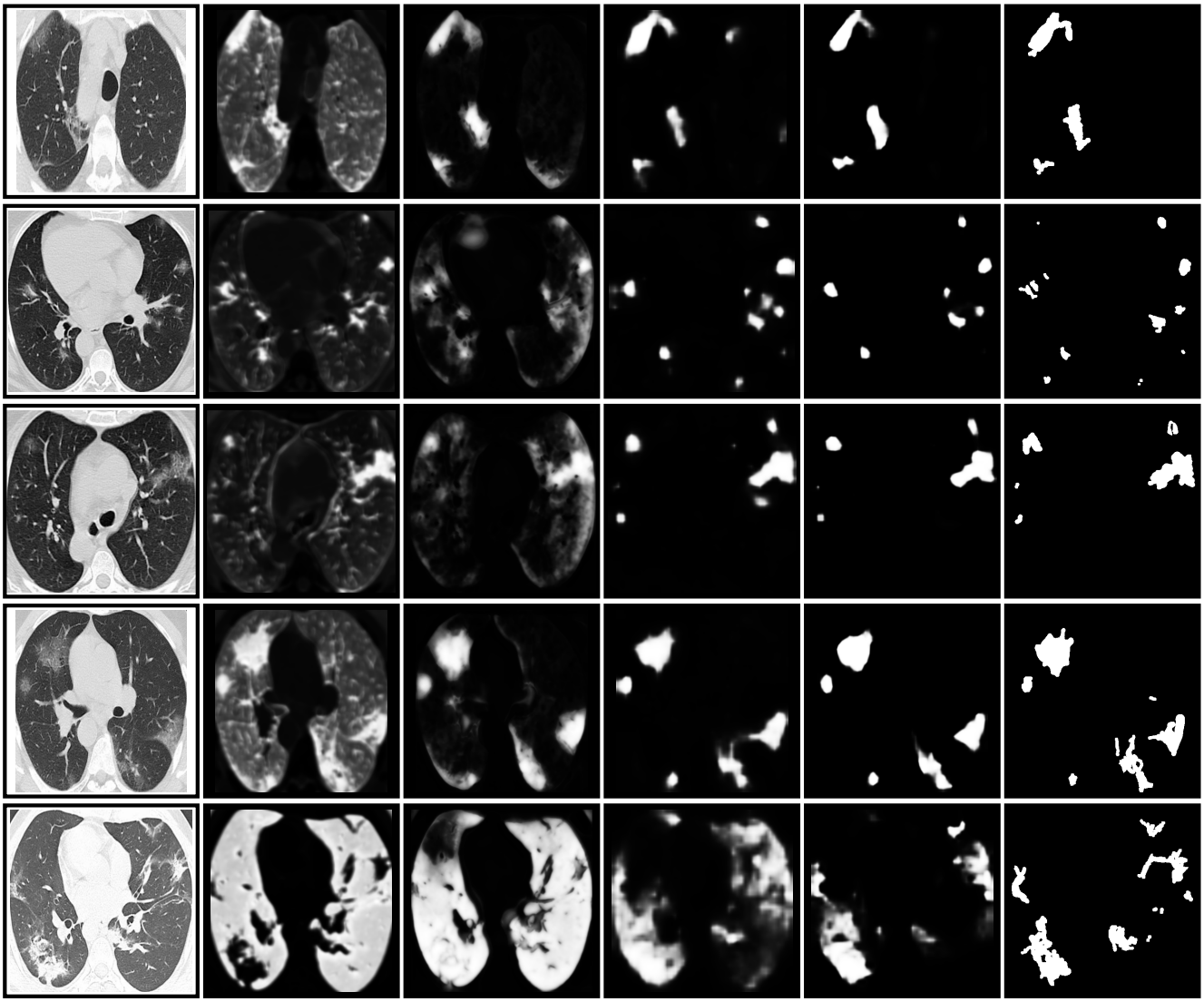
### B. Experimental Settings

**Baselines.** For the infection region experiments, we compare the proposed *Inf-Net* and *Semi-Inf-Net* with two classical segmentation models in the medical domain, *i.e.*, U-Net [47] and U-Net++ [46]. For the multi-class labeling experiments, we compare our model with two cutting-edge models from the computer vision community: DeepLabV3+ [61] and FCN8s [58].

**Evaluation Metrics.** Following [22], [44], we use three widely adopted metrics, *i.e.*, the Dice similarity coefficient, Sensitivity (Sen.), and Specificity (Spec.). We also introduce three golden metrics from the object detection field, *i.e.*, Structure Measure [62], Enhance-alignment Measure [63], and Mean Absolute Error. In our evaluation, we choose $S_3$ with $sigmoid$ function as the final prediction $S_p$. Thus, we measure the similarity/dissimilarity between final the prediction map and object-level segmentation ground-truth $G$, which can be formulated as follows:

[1] http://medicalsegmentation.com/covid19/

[2] https://www.sirm.org/category/senza-categoria/covid-19



CT Image | U-Net (MICCAI'15) [47] | U-Net++ (TMI'19) [46] | ***Inf-Net* (Ours)** | ***Semi-Inf-Net* (Ours)** | Ground Truth

Fig. 6. Visual comparison of lung infection segmentation results.

*1) Structure Measure ($S_\alpha$):* This was proposed to measure the structural similarity between a prediction map and ground-truth mask, which is more consistent with the human visual system:

$$S_\alpha = (1-\alpha) * S_o(S_p, G) + \alpha * S_r(S_p, G), \tag{6}$$

where $\alpha$ is a balance factor between object-aware similarity $S_o$ and region-aware similarity $S_r$. We report $S_\alpha$ using the default setting ($\alpha = 0.5$) suggested in the original paper [62].

*2) Enhanced-alignment Measure ($E_\phi^{mean}$):* This is a recently proposed metric for evaluating both local and global similarity between two binary maps. The formulation is as follows:

$$E_\phi = \frac{1}{w \times h} \sum_x^w \sum_y^h \phi(S_p(x,y), G(x,y)), \tag{7}$$

where $w$ and $h$ are the width and height of ground-truth $G$, and $(x, y)$ denotes the coordinate of each pixel in $G$. Symbol $\phi$ is the enhanced alignment matrix. We obtain a set of $E_\phi$ by converting the prediction $S_p$ into a binary mask with a threshold from 0 to 255. In our experiments, we report the mean of $E_\xi$ computed from all the thresholds.

*3) Mean Absolute Error (MAE):* This measures the pixel-wise error between $S_p$ and $G$, which is defined as:

$$MAE = \frac{1}{w \times h} \sum_x^w \sum_y^h |S_p(x,y) - G(x,y)|. \tag{8}$$

### C. Segmentation Results

*1) Quantitative Results:* To compare the infection segmentation performance, we consider the two state-of-the-art models U-Net and U-Net++. Quantitative results are shown in Table II. As can be seen, the proposed *Inf-Net* outperforms U-Net and U-Net++ in terms of Dice, $S_\alpha$, $E_\phi^{mean}$, and MAE by a large margin. We attribute this improvement to our implicit reverse attention and explicit edge-attention modeling, which provide robust feature representations. In addition, by introducing the semi-supervised learning strategy into our

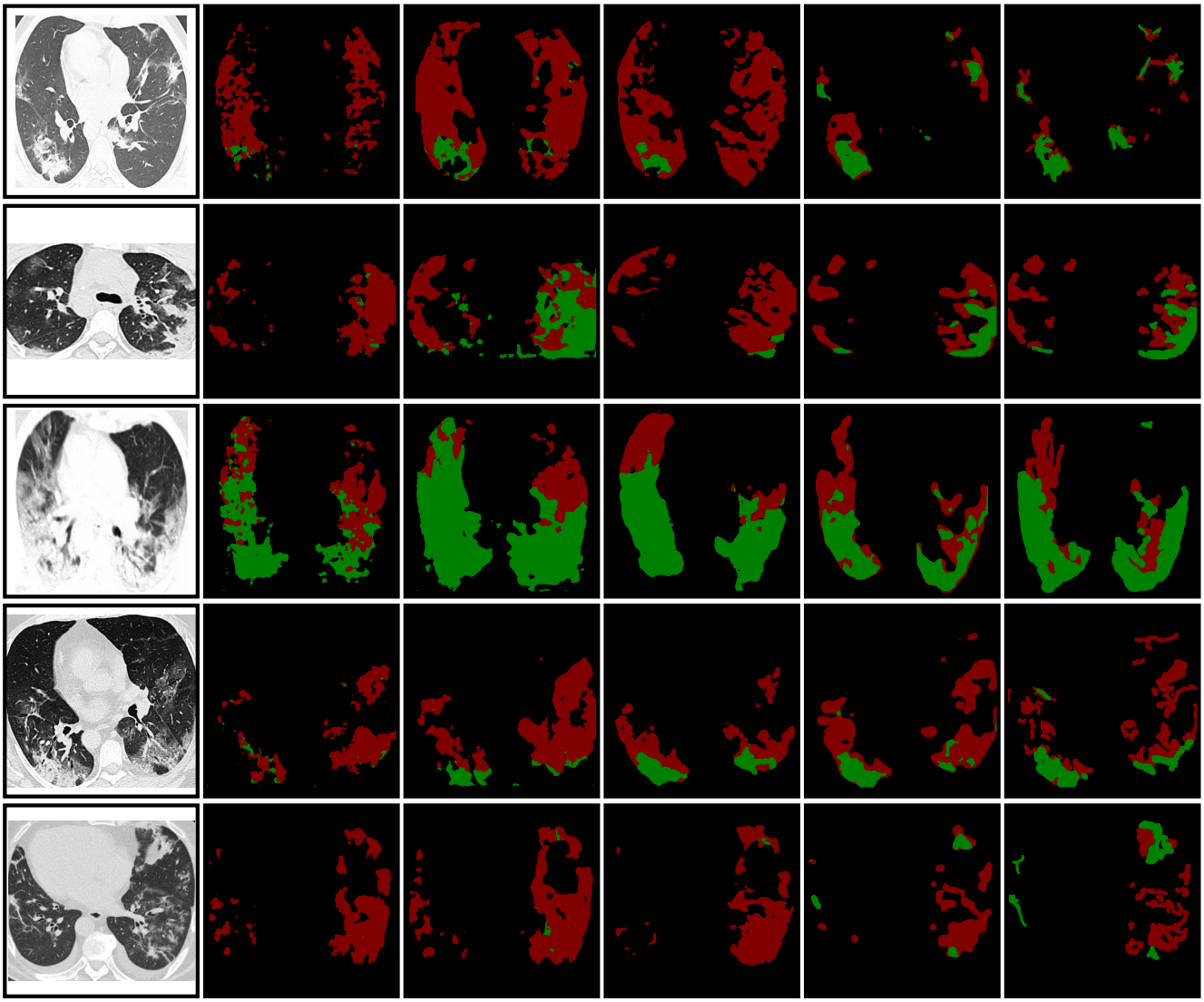



Fig. 7. Visual comparison of multi-class lung infection segmentation results, where the red and green labels indicate the GGO and consolidation, respectively.

TABLE II
QUANTITATIVE RESULTS OF INFECTION REGIONS ON OUR *COVID-SemiSeg* DATASET. THE BEST TWO RESULTS ARE SHOWN IN **RED** AND BLUE FONTS.

| Methods | Dice | Sen. | Spec. | $S_\alpha$ | $E_\phi^{mean}$ | MAE |
|---|---|---|---|---|---|---|
| U-Net (MICCAI'15) [47] | 0.439 | 0.534 | 0.858 | 0.622 | 0.625 | 0.186 |
| U-Net++ (TMI'19) [46] | 0.518 | 0.672 | 0.902 | 0.722 | 0.720 | 0.120 |
| ***Inf-Net* (Ours)** | **0.682** | **0.692** | **0.943** | **0.781** | **0.838** | **0.082** |
| ***Semi-Inf-Net* (Ours)** | **0.739** | **0.725** | **0.960** | **0.800** | **0.894** | **0.064** |

framework, we can further boost the performance with a 5.7% improvement in terms of Dice.

As an assistant diagnostic tool, the model is expected to provide more detailed information regarding the infected areas. Therefore, we extent to our model to the multi-class (*i.e.*, GGO and consolidation segmentation) labeling. Table III shows the quantitative evaluation on our *COVID-SemiSeg* dataset, which indicate that our Semi-Inf-Net&FCN8s pipeline achieves the state-of-the-art performance on GGO segmentation in most evaluation metrics. For more challenging consolidation segmentation, the proposed pipeline also achieves competitive results. For instance, in terms of Dice, our method outperforms the cutting-edge model, FCN8s, by 1.7%. Overall, the proposed pipeline performs better than existing state-of-the-art models on multi-class labeling in terms of Dice, $S_\alpha$, $E_\phi^{mean}$, and MAE. Our method also performs very well in the specificity metric. However, for the sensitivity metric, our model still has room for improvement.

*2) Qualitative Results:* The lung infection segmentation results, shown in Fig. 6, indicate that our *Semi-Inf-Net* and *Inf-Net* outperform the baseline methods remarkably. Specifically, they yield segmentation results that are close to the ground truth with much less mis-segmented tissue. In contrast, U-Net gives unsatisfactory results, where a large number of mis-segmented tissues exist. U-Net++ improves the results, but the performance is still not promising. The success of

TABLE III
QUANTITATIVE RESULTS OF GROUND-GLASS OPACITIES AND CONSOLIDATION ON OUR *COVID-SemiSeg* DATASET. THE BEST TWO RESULTS ARE SHOWN IN **RED** AND **BLUE** FONTS.

| Methods | Dice | Sen. | Spec. | $S_\alpha$ | $E_\phi^{mean}$ | MAE |
|---|---|---|---|---|---|---|
| Ground-glass opacities | | | | | | |
| DeepLabV3+ (stride = 8) [61] | 0.375 | 0.478 | 0.863 | 0.544 | 0.675 | 0.123 |
| DeepLabV3+ (stride = 16) [61] | 0.443 | **0.713** | 0.823 | 0.548 | 0.655 | 0.156 |
| FCN8s [58] | **0.471** | 0.537 | **0.905** | **0.582** | **0.774** | **0.101** |
| ***Semi-Inf-Net*+FCN8s (Ours)** | **0.646** | **0.720** | **0.941** | **0.711** | **0.882** | **0.071** |
| Consolidation | | | | | | |
| DeepLabV3+ (stride = 8) [61] | 0.117 | 0.120 | **0.584** | 0.603 | 0.621 | 0.052 |
| DeepLabV3+ (stride = 16) [61] | 0.188 | **0.245** | 0.560 | 0.617 | **0.688** | 0.061 |
| FCN8s [58] | **0.221** | **0.212** | 0.567 | **0.652** | 0.651 | **0.039** |
| ***Semi-Inf-Net*+FCN8s (Ours)** | **0.238** | 0.186 | **0.639** | **0.661** | **0.660** | **0.035** |

*Inf-Net* is owed to our coarse-to-fine segmentation strategy, where a parallel partial decoder first roughly locates lung infection regions and then multiple edge attention modules are employed for fine segmentation. This strategy mimics how real clinicians segment lung infection regions from CT scans, and therefore achieves promising performance. In addition, the advantage of our semi-supervised learning strategy is also confirmed by Fig. 6. As can be observed, compared with *Inf-Net*, *Semi-Inf-Net* yields segmentation results with more accurate boundaries. In contrast, *Inf-Net* gives relatively fuzzy boundaries, especially in the subtle infection regions.

We also show the multi-class infection labeling results in Fig. 7. As can be observed, our model, *Semi-Inf-Net* & FCN8s, consistently performs the best among all methods. It is worth noting that both GGO and consolidation infections are accurately segmented by *Semi-Inf-Net* & FCN8s, which further demonstrates the advantage of our model. In contrast, the baseline methods, DeepLabV3+ with different strides and FCNs, all obtain unsatisfactory results, where neither GGO and consolidation infections can be accurately segmented.

### D. Ablation Study

In this subsection, we conduct several experiments to validate the performance of each key component of our *Semi-Inf-Net*, including the PPD, RA, and EA modules.

*1) Effectiveness of PPD:* To explore the contribution of the parallel partial decoder, we derive two baselines: No.1 (backbone only) & No.2 (backbone+PPD) in Table IV. The results clearly show that PPD is necessary for boosting performance.

*2) Effectiveness of RA:* We investigate the importance of the RA module. From Table IV, we observe that No.3 (backbone + RA) increases the backbone performance (No.1) in terms of major metrics, *e.g.*, Dice, Sensitivity, MAE, *etc*. This suggests that introducing the RA component can enable our model to accurately distinguish true infected areas.

*3) Effectiveness of PPD & RA:* We also investigate the importance of the combination of the PPD and RA component (No.4). As shown in Table IV, No.4 performs better than other settings (*i.e.*, No.1∼No.3) in most metrics. These improvements demonstrate that the reverse attention together with the parallel partial decoder are the two central components responsible for the good performance of *Inf-Net*.

TABLE IV
ABLATION STUDIES OF OUR *Semi-Inf-Net*. THE BEST TWO RESULTS ARE SHOWN IN **RED** AND **BLUE** FONTS.

| Methods | Dice | Sen. | Spec. | $S_\alpha$ | $E_\phi^{mean}$ | MAE |
|---|---|---|---|---|---|---|
| (No.1) Backbone | 0.442 | 0.570 | 0.825 | 0.651 | 0.569 | 0.207 |
| (No.2) Backbone+PPD | **0.669** | **0.744** | 0.880 | 0.720 | 0.810 | 0.125 |
| (No.3) Backbone+RA | 0.625 | **0.826** | 0.809 | 0.668 | 0.736 | 0.177 |
| (No.4) Backbone+PPD+RA | 0.655 | 0.690 | **0.927** | **0.761** | **0.812** | **0.098** |
| (No.5) Backbone+PPD+RA+EA | **0.739** | 0.725 | **0.960** | **0.800** | **0.894** | **0.064** |

*4) Effectiveness of EA:* Finally, we investigate the importance of the EA module. As shown in Table IV, No.5 performances better than the setting (*i.e.*, No.4) in most of metrics. This improvement validates that EA module further improves the segmentation performance in our *Inf-Net*.

## V. CONCLUSION

In this paper, we have proposed a novel COVID-19 lung CT infection segmentation network, named *Inf-Net*, which utilizes an implicit reverse attention and explicit edge-attention to improve the identification of infected regions. Moreover, we have also provided a semi-supervised solution, *Semi-Inf-Net*, to alleviate the shortage of high quality labeled data. Extensive experiments on our *COVID-SemiSeg* dataset have demonstrated that the proposed *Inf-Net* and *Semi-Inf-Net* outperform the cutting-edge segmentation models and advance the state-of-the-art performances. Our system has great potential to be applied in assessing the diagnosis of COVID-19, *e.g.*, quantifying the infected regions, monitoring the longitudinal disease changes, and mass screening processing.

## REFERENCES


[1] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health concern," *The Lancet*, vol. 395, no. 10223, pp. 470–473, feb 2020.

[2] C. Huang, Y. Wang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *The Lancet*, vol. 395, no. 10223, pp. 497–506, feb 2020.

[3] "Coronavirus COVID-19 global cases by the center for systems science and engineering at johns hopkins university," https://coronavirus.jhu.edu/map.html, accessed: 2020-04-02.

[4] T. Ai, Z. Yang *et al.*, "Correlation of chest CT and rt-pcr testing in coronavirus disease 2019 (COVID-19) in China: A report of 1014 cases," *Radiology*, vol. 2019, p. 200642, feb 2020.

[5] G. D. Rubin, L. B. Haramati *et al.*, "The role of chest imaging in patient management during the COVID-19 pandemic: A multinational consensus statement from the fleischner society," *Radiology*, p. 201365, apr 2020.

[6] F. Shi, J. Wang *et al.*, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for COVID-19," *arXiv*, 2020.

[7] Y. Fang, H. Zhang *et al.*, "Sensitivity of chest CT for COVID-19: Comparison to RT-PCR," *Radiology*, p. 200432, 2020.

[8] M.-Y. Ng, E. Y. Lee, and otehrs, "Imaging profile of the COVID-19 infection: Radiologic findings and literature review," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, p. e200034, 2020.

[9] "COVID-19 CT segmentation dataset," https://medicalsegmentation.com/covid19/, accessed: 2020-04-11.

[10] Z. Ye, Y. Zhang, Y. Wang, Z. Huang, and B. Song, "Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review," *European Radiology*, vol. 2019, no. 37, pp. 1–9, mar 2020.

[11] J. P. Cohen, P. Morrison, and L. Dao, "COVID-19 image data collection," *arXiv*, 2020.

[12] J. Zhao, Y. Zhang, X. He, and P. Xie, "COVID-CT-Dataset: a CT scan dataset about COVID-19," *arXiv*, 2020.

[13] "COVID-19 Patients Lungs X Ray Images 10000," https://www.kaggle.com/nabeelsajid917/covid-19-x-ray-10000-images, accessed: 2020-04-11.

[14] M. E. H. Chowdhury, T. Rahman *et al.*, "Can AI help in screening Viral and COVID-19 pneumonia?" *arXiv*, 2020.

[15] V. Rajinikanth, N. Dey *et al.*, "Harmony-search and otsu based system for coronavirus disease (COVID-19) detection using lung CT scan images," *arXiv preprint arXiv:004.03431*, 2020.

[16] L. Wang and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images," *arXiv*, mar 2020.

[17] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, "COVID-19 Screening on Chest X-ray Images Using Deep Learning based Anomaly Detection," *arXiv*, mar 2020.

[18] X. Xu, X. Jiang *et al.*, "Deep learning system to screen coronavirus disease 2019 pneumonia," *arXiv*, 2020.

[19] C. Zheng, X. Deng *et al.*, "Deep learning-based detection for COVID-19 from chest CT using weak label," *medRxiv*, 2020.

[20] H. Fu, D.-P. Fan, G. Chen, and T. Zhou, "COVID-19 Imaging-based AI Research Collection," https://github.com/HzFu/COVID19_imaging_AI_paper_list.

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

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

[23] I. Sluimer, A. Schilham, M. Prokop, and B. Van Ginneken, "Computer analysis of computed tomography scans of the lung: a survey," *IEEE TMI*, vol. 25, no. 4, pp. 385–405, 2006.

[24] B. Kamble, S. P. Sahu, and R. Doriya, "A review on lung and nodule segmentation techniques," in *Advances in Data and Information Sciences*. Springer, 2020, pp. 555–565.

[25] 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," *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018.

[26] 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," *International journal of biomedical imaging*, vol. 2012, 2012.

[27] 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," *Computers in Biology and Medicine*, vol. 43, no. 4, pp. 287–300, 2013.

[28] S. Shen, A. A. Bui, J. Cong, and W. Hsu, "An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy," *Computers in Biology and Medicine*, vol. 57, pp. 139–149, 2015.

[29] S. Wang, M. Zhou *et al.*, "Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation," *Medical image analysis*, vol. 40, pp. 172–183, 2017.

[30] 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 *MICCAI*. Springer, 2018, pp. 732–740.

[31] J. Jiang, Y.-C. Hu *et al.*, "Multiple resolution residually connected feature streams for automatic lung tumor segmentation from CT images," *IEEE TMI*, vol. 38, no. 1, pp. 134–144, 2018.

[32] X. J. Zhu, "Semi-supervised learning literature survey," University of Wisconsin-Madison Department of Computer Sciences, Tech. Rep., 2005.

[33] J. E. van Engelen and H. H. Hoos, "A survey on semi-supervised learning," *Machine Learning*, vol. 109, no. 2, pp. 373–440, feb 2020.

[34] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Workshop on challenges in representation learning, ICML*, vol. 3, 2013, p. 2.

[35] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," *ICLR*, 2017.

[36] A. Rasmus, M. Berglund, M. Honkala, H. Valpola, and T. Raiko, "Semi-supervised learning with ladder networks," in *NIPS*, 2015, pp. 3546–3554.

[37] S. Wang, B. Kang *et al.*, "A deep learning algorithm using CT images to screen for corona virus disease (COVID-19)," *medRxiv*, 2020.

[38] J. Chen, L. Wu *et al.*, "Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study," *medRxiv*, 2020.

[39] A. Senior, J. Jumper, and D. Hassabis, "AlphaFold: Using AI for scientific discovery," *DeepMind. Recuperado de: https://deepmind.com/blog/alphafold*, 2018.

[40] Z. Hu, Q. Ge, L. Jin, and M. Xiong, "Artificial intelligence forecasting of COVID-19 in China," *arXiv*, 2020.

[41] Y. Song, S. Zheng *et al.*, "Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images," *medRxiv*, 2020.

[42] O. Gozes, M. Frid-Adar *et al.*, "Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis," *arXiv*, 2020.

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

[44] F. Shi, L. Xia *et al.*, "Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification," *arXiv*, 2020.

[45] C. Szegedy, W. Liu *et al.*, "Going deeper with convolutions," in *CVPR*, 2015, pp. 1–9.

[46] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," *IEEE TMI*, pp. 3–11, 2019.

[47] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015, pp. 234–241.

[48] J.-X. Zhao, J.-J. Liu *et al.*, "Egnet: Edge guidance network for salient object detection," in *ICCV*, 2019, pp. 8779–8788.

[49] Z. Wu, L. Su, and Q. Huang, "Stacked cross refinement network for edge-aware salient object detection," in *ICCV*, 2019, pp. 7264–7273.

[50] Z. Zhang, H. Fu *et al.*, "ET-Net: A generic edge-attention guidance network for medical image segmentation," in *MICCAI*, 2019, pp. 442–450.

[51] Z. Wu, L. Su, and Q. Huang, "Cascaded partial decoder for fast and accurate salient object detection," in *CVPR*, 2019, pp. 3907–3916.

[52] S. Gao, M.-M. Cheng *et al.*, "Res2Net: A new multi-scale backbone architecture," *IEEE TPAMI*, 2019.

[53] Y. Wei, J. Feng, X. Liang, M.-M. Cheng, Y. Zhao, and S. Yan, "Object region mining with adversarial erasing: A simple classification to semantic segmentation approach," in *CVPR*, 2017, pp. 1568–1576.

[54] S. Chen, X. Tan, B. Wang, and X. Hu, "Reverse attention for salient object detection," in *ECCV*, 2018, pp. 234–250.

[55] X. Qin, Z. Zhang *et al.*, "BASNet: Boundary-aware salient object detection," in *CVPR*, 2019, pp. 7479–7489.

[56] J. Wei, S. Wang, and Q. Huang, "F3Net: Fusion, feedback and focus for salient object detection," in *AAAI*, 2020.

[57] S. Mittal, M. Tatarchenko, Ö. Çiçek, and T. Brox, "Parting with illusions about deep active learning," *arXiv preprint arXiv:1912.05361*, 2019.

[58] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015, pp. 3431–3440.

[59] D. Nie, L. Wang, Y. Gao, and D. Shen, "Fully convolutional networks for multi-modality isointense infant brain image segmentation," in *ISBI*. IEEE, 2016, pp. 1342–1345.

[60] M. Pekala, N. Joshi, T. A. Liu, N. M. Bressler, D. C. DeBuc, and P. Burlina, "Deep learning based retinal OCT segmentation," *Computers in Biology and Medicine*, vol. 114, p. 103445, 2019.

[61] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *ECCV*, 2018, pp. 801–818.

[62] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji, "Structure-measure: A new way to evaluate foreground maps," in *ICCV*, 2017, pp. 4548–4557.

[63] D.-P. Fan, C. Gong *et al.*, "Enhanced-alignment measure for binary foreground map evaluation," *IJCAI*, pp. 698–704, 2018.