Received: 14 May 2021 | Revised: 23 August 2021 | Accepted: 11 September 2021

DOI: 10.1002/mp.15231

RESEARCH ARTICLE

MEDICAL PHYSICS

CARes-UNet: Content-aware residual UNet for lesion segmentation of COVID-19 from chest CT images

Xinhua Xu[1] | **Yuhang Wen**[1] | **Lu Zhao**[2] | **Yi Zhang**[1] | **Youjun Zhao**[1] | **Zixuan Tang**[1] | **Ziduo Yang**[1] | **Calvin Yu-Chian Chen**[1,3,4,5]

[1] Artificial Intelligence Medical Center, School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, China

[2] Department of Clinical Laboratory, The Sixth Affiliated HospitalSun Yat-sen University, Guangzhou, China

[3] Department of Medical Research, China Medical University Hospital, Taichung, Taiwan

[4] Department of Bioinformatics and Medical Engineering, Asia University, Taichung, Taiwan

[5] Guangdong Provincial Key Laboratory of Fire Science and Technology, Guangzhou, China

Correspondence
Calvin Yu-Chian Chen, Artificial Intelligence Medical Center, School of Intelligent Systems Engineering, Sun Yat-sen University, Shenzhen, China.
Email: chenyuchian@mail.sysu.edu.cn

Xinhua Xu, Yuhang Wen, and Lu Zhao contributed equally to this study.

Funding information
National Natural Science Foundation of China, Grant/Award Number: 62176272; Guangzhou Science and Technology Fund, Grant/Award Number: 201803010072; Science, Technology&Innovation Commission of Shenzhen Municipality, Grant/Award Number: 20170818165305521; China Medical University Hospital, Grant/Award Numbers: DMR-107-067, DMR-108-132, DMR-110-097

Abstract

Purpose: Coronavirus disease 2019 (COVID-19) has caused a serious global health crisis. It has been proven that the deep learning method has great potential to assist doctors in diagnosing COVID-19 by automatically segmenting the lesions in computed tomography (CT) slices. However, there are still several challenges restricting the application of these methods, including high variation in lesion characteristics and low contrast between lesion areas and healthy tissues. Moreover, the lack of high-quality labeled samples and large number of patients lead to the urgency to develop a high accuracy model, which performs well not only under supervision but also with semi-supervised methods.

Methods: We propose a content-aware lung infection segmentation deep residual network (content-aware residual UNet (CARes-UNet)) to segment the lesion areas of COVID-19 from the chest CT slices. In our CARes-UNet, the residual connection was used in the convolutional block, which alleviated the degradation problem during the training. Then, the content-aware upsampling modules were introduced to improve the performance of the model while reducing the computation cost. Moreover, to achieve faster convergence, an advanced optimizer named Ranger was utilized to update the model's parameters during training. Finally, we employed a semi-supervised segmentation framework to deal with the problem of lacking pixel-level labeled data.

Results: We evaluated our approach using three public datasets with multiple metrics and compared its performance to several models. Our method outperforms other models in multiple indicators, for instance in terms of Dice coefficient on COVID-SemiSeg Dataset, CARes-UNet got the score 0.731, and semi-CARes-UNet further boosted it to 0.776. More ablation studies were done and validated the effectiveness of each key component of our proposed model.

Conclusions: Compared with the existing neural network methods applied to the COVID-19 lesion segmentation tasks, our CARes-UNet can gain more accurate segmentation results, and semi-CARes-UNet can further improve it using semi-supervised learning methods while presenting a possible way to solve the problem of lack of high-quality annotated samples. Our CARes-UNet and semi-CARes-UNet can be used in artificial intelligence-empowered computer-aided diagnosis system to improve diagnostic accuracy in this ongoing COVID-19 pandemic.

KEYWORDS
computed tomography (CT) image, content-aware residual UNet, coronavirus disease 2019 (COVID-19), deep learning, segmentation

1 | INTRODUCTION

Since 2019, the world has been suffering from the pandemic of coronavirus disease 2019 (COVID-19), which is caused by severe respiratory syndrome coronavirus 2 (SARS-CoV-2).[1,2] A common and effective method to diagnose COVID-19 is the reverse-transcriptase molecular polymerase chain reaction (RT-PCR) test by detecting the SARS-CoV-2 RNA.[3] However, data from several studies suggest that the RT-PCR test has low sensitivity[4] and accuracy due to little amounts of SARS-CoV-2 virus concentration in patient samples. Chest computed tomography (CT) radiography is another powerful tool for the detection of pulmonary infection, which plays a pivotal role in the COVID-19 diagnosis system for low cost and is highly sensitive. However, manually delineating the infected lung region of COVID-19 based on chest CT images by radiologists is a labor-intensive and highly subjective task. Besides, diagnosing patients over time can lead to a serious decline in the judgment of doctors because of the enormous and growing base number.

Artificial intelligence (AI) is now being developed rapidly to combine with CT to help radiologists and clinicians improve diagnostic accuracy and working efficiency. Recently Lessmann et al.[5] developed and validated the COVID-19 Reporting and Data System (CO-RADS) AI system, which correctly identified COVID-19 patients using chest CT scans and assigned CO-RADS and CT severity scores automatically. Zhang et al.[6] developed an AI system for novel coronavirus pneumonia diagnosis with a large CT database from 3777 patients, which was clinically applicable to assisting radiologists and physicians in performing a quick diagnosis.

Computer-aided diagnoses of COVID-19 empowered by AI techniques have been well-studied, which can be mainly divided into two categories: COVID-19 diagnosis and lesion segmentation.

The most common one is the automatic COVID-19 diagnostic based on CT slices or volumes.[7–10] For example, Bai et al.[9] proposed an AI system for differentiating COVID-19 and other pneumonia at chest CT based on EfficientNet[11] and concluded that AI assistance improves radiologists' performance in distinguishing different types of CT volume. Wang et al.[10] proposed a weakly supervised framework for COVID-19 classification. In their method, the lung region was first segmented using a weakly supervised method. Then the segmented 3D lung region was fed into a 3D deep neural network to predict the probability of COVID-19 infectious. Finally, the lesions were localized by a weakly supervised lesion localization method. Wang et al.[12] proposed a novel joint learning framework to improve COVID-19 diagnosis by effectively learning from heterogeneous datasets with distribution discrepancy. Owais[13] proposed a multilevel deep-aggregated boosted network to spot COVID-19 infection from heterogeneous data including CT images.

Another category is the COVID-19 lesion segmentation. Since our work is more about lesion segmentation in CT slices, we will discuss it more thoroughly. A large number of image segmentation algorithms have been developed in the literature,[14] among which deep learning methods for COVID-19 segmentation have achieved the most remarkable success. For example, traditional U-Net[15] and SegNet[16] continue to shine in COVID-19 lesion segmentation. Saood et al[17] compared U-Net with SegNet in COVID-19 lung CT image segmentation tasks. In their experiment, the Dice coefficient of SegNet is higher than that of U-Net. Thus, they concluded that SegNet shows a superior ability in segmentation tasks. Bizopoulos et al.[18] conducted extensive experiments to present the comparison of various COVID-19 lesion segmentation deep learning models. In their work, four deep learning architectures (UNet, LinkNet, Feature Pyramid Networks (FPN), Pyramid Scene Parsing Network (PSPNet)) were tested, combined with 25 randomly initialized and pre-trained encoders (variations of VGG, DenseNet, ResNet, ResNext, Dual Path Network (DPN), MobileNet, Xception, Inception-v4, EfficientNet), for a total of 200 tested models.

Some recently proposed new network architectures gain better segmentation results via elaborate architecture design. For example, Fan et al.[19] proposed Inf-Net and a semi-supervised segmentation framework, which was based on a randomly selected propagation strategy. On COVID-SemiSeg Dataset, Inf-Net gained the best performance, compared with other cutting-edge models owing to three specially designed components: edge attention module, parallel partial decoder, and reverse attention module. Wang et al.[20] proposed a noise-robust framework for COVID-19 lesion segmentation to tackle the inaccurate annotation caused by complex appearances of pneumonia lesions and high inter- and intra-observer variability. Zhou et al.[21] proposed a machine-agnostic segmentation and quantification method for CT-based COVID-19 diagnosis. Gao et al.[22] proposed a dual-branch combination network that was able to perform diagnosis and lesion segmentation simultaneously. Wu et al.[23] developed a joint classification and segmentation system to perform real-time and explainable COVID-19 chest CT segmentation and diagnosis. Yang et al.[24] presented a novel weakly supervised learning method based on a generative adversarial network for lesion segmentation with image-level labels only. Laradji et al.[25] proposed a weakly supervised consistency-based learning method for COVID-19 segmentation in CT slices and demonstrated that the weakly supervised learning method can also lead to competitive results compared with the fully supervised learning methods.

Although extensive research and advanced AI methods have been carried out focusing on accurate

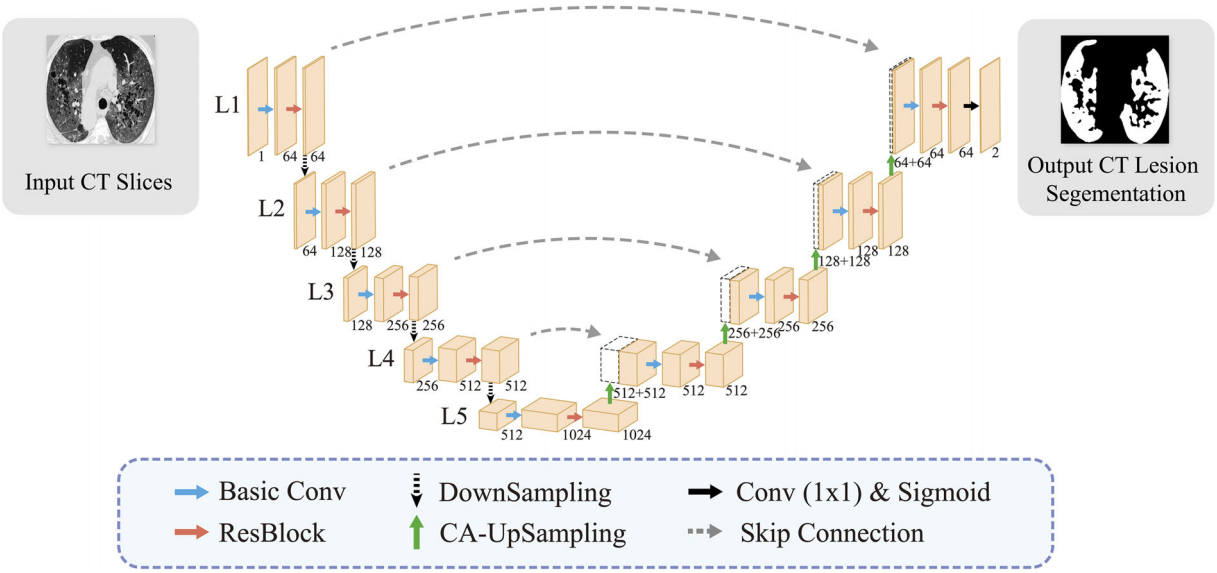

FIGURE 1 Network architecture of content-aware residual UNet (CARes-UNet)

prediction, COVID-19 infection detection in CT slices is still a challenging work for several reasons. Training a very deep neural network is hard due to the gradient vanishing and unstable optimizing. Besides, the complex boundary interactions, larger appearance variation, and low tissue contrast result in difficulty to identify infected regions. Moreover, because of the urgency of the COVID-19 pandemic, it is hard to collect an enormous amount of labeled data for deep learning. Finally, acquiring high-quality pixel-level annotations of lung infection in CT slices could be expensive and time-consuming.

In this study, we propose a content-aware residual UNet[15] (CARes-UNet) for lung infection segmentation from CT slices to deal with the issues mentioned above. To alleviate the lack of high-quality annotated samples, we applied a semi-supervised framework to train the model. We also conducted several comparisons and ablation studies to prove the effectiveness of the proposed CARes-UNet and semi-CARes-UNet.

2 | METHODS

2.1 | Overview of the network

The architecture of our CARes-UNet, which is built by redesigning the traditional U-Net, is shown in Figure 1. Our CARes-UNet comprises a downsampling path and an upsampling path. The downsampling path enlarged the receptive field while reducing computation cost. The upsampling path recovered the lost resolution in the downsampling path. The structure of our encoder block and downsampling block is shown in Figure 2. Every encoder block and downsampling block contained one basic convolution operation and one resblock, which was the same except that the basic convolution in the downsampling block did not change the channels number. The basic convolution operation and resblock at the same level in Figure 1 compose an encoder block together with a downsampling operation next to it, which are both explained in Figure 2. Batch normalization[26] and the rectified linear unit were added after each convolution operation. In L1 (Figure 1), the first basic convolution block was to change the channels from 1 to 64, and the resblock was used to extract features without changing the channel numbers.[27] After each downsampling block, the matrix resolution was down-sampled by applying a $3 \times 3 \times 3$ convolution operation with stride 2 between each pair convolution block to preserve information and reduce computation cost. Similarly, in other layers from L2 to L5, a convolution operation was used to transform the channels to 128, 256, 512, and finally to 1024. When the channel number reaches 1024, the feature map with minimal size was upsampled via a content-aware upsampling operation. The features in the downsampling path were concatenated with the features in the upsampling path by skip connection to provide the added information without the downsampling information abstraction.[28] Finally, the channel number was reduced to 64, and the final segmentation result comes out through a 1×1 convolution operator followed by a sigmoid layer to map the result into probability. Technical details of our network and loss function will be illustrated later.

2.2 | Residual structure

The residual connection was applied in our model to mitigate the degradation problem, which is shown in Figures 1 and 2 as resblock with detailed explanation in Figure 2. The degradation problem indicates that it

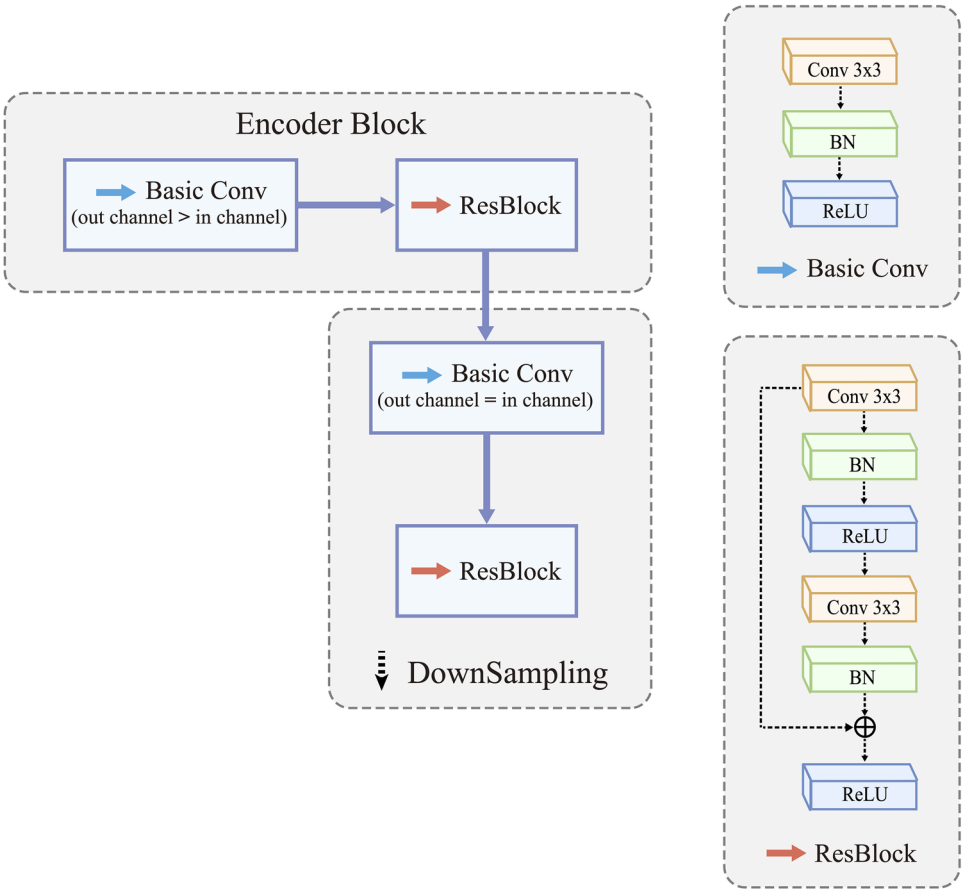

FIGURE 2 The structure of encoder block and downsampling block. Both of them consist of one basic convolution operation and one resblock. The meanings of colored arrows are the same as those in Figure 1

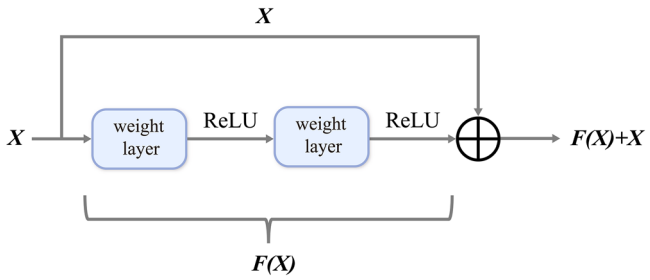

FIGURE 3 Residual connection

may be difficult for the solver to estimate the identity map through multiple nonlinear layers. With the re-expression of residual learning, the solver drives the weights of multiple nonlinear layers to zero to approximate the identity map if the identity map is optimal. However, this is usually impractical. A simple but effective solution is using the residual connection.

As shown in Figure 3, the original network fits $F(X)$ while the network fits $F(X) - X$ after introducing residual connection. When $F(X) = 0$, the output of the network $H(X) = X$ is an identity mapping, which is complex for ordinary multi-layer nonlinear networks to fit. After adding a residual connection, the fitting cost of identity mapping is greatly reduced. To sum up, the residual block can simplify the fitting of the identity mapping, solve the degradation problem, and greatly improve the learning ability of the deep model.

2.3 | Content-aware upsampling

Feature map upsampling is very important in segmentation tasks. An ideal method of feature upsampling should have at least three qualities. (1) Large receptive field: With a larger receptive field, the outcome of feature upsampling will make better usage of the information around. (2) Content-related: The upsampling filter should be related to the content of the feature map so that the method will provide us more appropriate information. (3) Lightweight: The upsampling filters in a network are important and, usually, very complicated, which will make the training and convergence of a model very difficult. So we expected our chosen upsampling filter to be as light as possible.

Inspired by Wang et al.,[29] we introduced a content-aware upsampling module in our network. As shown in Figure 4, reassembly kernels are generated by kernel prediction. First, we used a convolution layer to

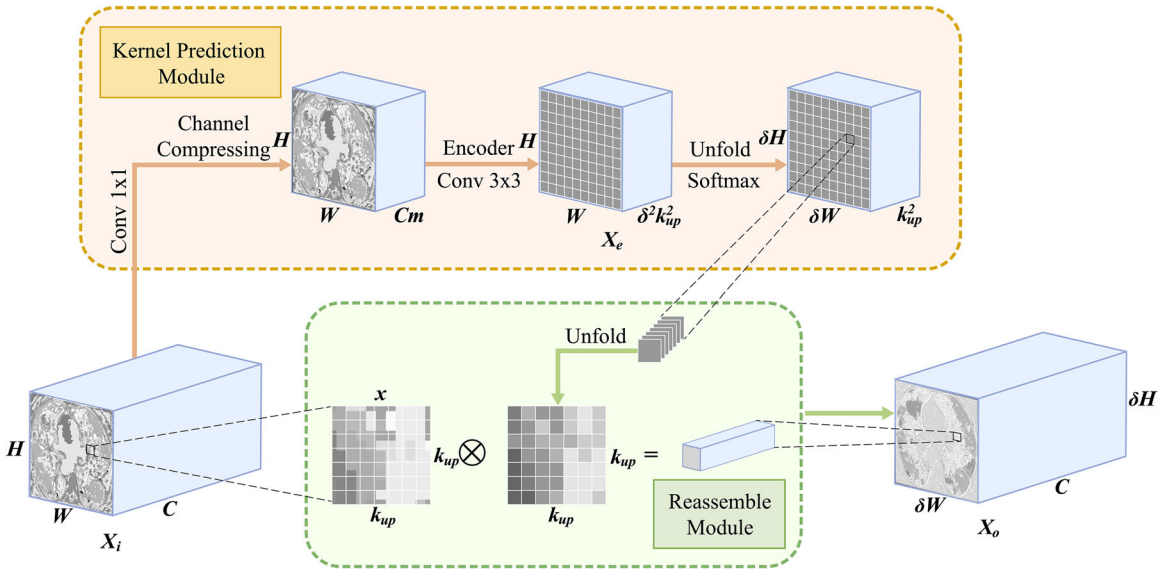

FIGURE 4 Content-aware upsampling. Kernel prediction module generates reassembly kernels. Reassemble module upsamples the feature map using predicted kernels. Unfold returns a view of the original tensor which contains all slices of size from tensor in the dimension and the cross-product sign represents the matrix product of two tensors

compress the channels of the input feature map to reduce computation. Next, we employed a 3 × 3 convolution layer to encode the feature map based on the content of input features. Then, before upsampling, we applied softmax to take a normalization step spatially to make sure that the sum of kernel values equals 1. After normalization, the operator would not rescale or change the mean values of the feature map. All the operations above are shown in Equations (1) and (2).

$$X_e = unfold\left(Conv_{3\times3}\left(Conv_{1\times1}\left(X_i\right)\right)\right), \tag{1}$$

$$Kernel\ (w,\ h,\ k) = \frac{e^{X_e(w,\ h,\ k)}}{\sum e^{X_e(w,\ h,k)}}, \tag{2}$$

where $Conv_{n\times n}$ is convolution with kernel sized $n \times n$, and X_e is the feature map after the encoding operation.

Finally, as shown in Equation (3)

$$X_0\left(w, h, k\right) = unfold\left(Kernel\left(w, h, :\right)\right) \otimes x, \tag{3}$$

the reassemble module used the predicted kernel generated in the kernel prediction module to upsample the feature map, where $X_o(w, h, k)$ is one pixel in the output feature map X_o with position (w, h, k);$Kernel(w, h, k)$, defined analogously, is one pixel in the kernel with position (w, h, k); operator $\otimes$ represents convolution; and x is a rectangular area with a size of $k_{up} \times k_{up}$ in X_i from $X_i(\frac{w}{\delta}, \frac{h}{\delta}, k)$ to $X_i(\frac{w}{\delta} + k_{up}, \frac{h}{\delta} + k_{up}, k)$ (δ represents the upsampling ratio).

With reassembly kernels, each pixel in the region contributed to the upsampled pixel differently according to the content of features, instead of distance. The content-aware reassemble module automatically assigned more weights to the important features, thus resulting in more accurate lesion localization.

2.4 | Ranger optimizer

Fast, effective, and stable optimizers are what all researchers in various fields are pursuing. We employed a Ranger optimizer for model training. Ranger optimizer is a combination of rectified Adam (RAdam)[30] and Lookahead.[31] RAdam is a new variant of Adam, which provides a dynamic heuristic to explicitly rectify the variance of adaptive learning rate based on derivations and avoid manual tuning warm-up during training. It can reduce variance during the early stage of training and compare superior with the heuristic warm-up. Lookahead can improve learning stability and reduce variance during training. Lookahead maintains a set of slow weights and fast weights, which get synced with the fast weights every k updates. It first updates k times fast weight using inner loop optimizer and then updates slow weight once in the direction of final fast weight.

RAdam is more robust in terms of learning rate variations, which provides a dynamic warm-up, while Lookahead lessens the need for extensive hyperparameter tuning. Therefore, it is intuitive that combining RAdam with Lookahead would result in a more powerful optimizer. By combining RAdam with Lookahead, the model can achieve higher performance and faster convergence with minimal computation cost.

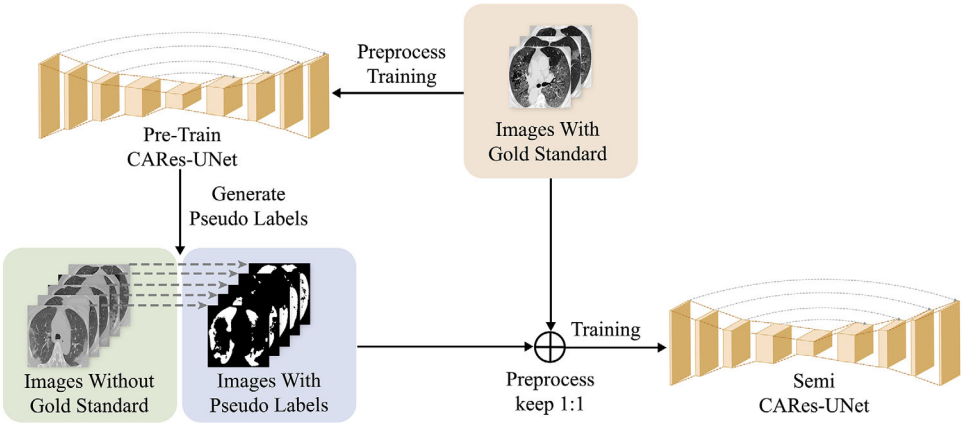

FIGURE 5 Overview of our semi-supervised framework

2.5 | Loss function

Motivated by Qiu et al.,[28] we defined the loss function L_{seg} as a combination of a weighted Dice loss L_{dice} and a weighted binary cross-entropy loss L_{BCE} for each segmentation supervision, that is,

$$L_{seg} = L_{dice} + \lambda L_{BCE}, \tag{4}$$

where λ is the weight, and we set it to 1 in our experiments. The L_{dice} is

$$L_{dice} = 1 - dice\left(\hat{x}, x^{gt}\right), \tag{5}$$

where $dice(a, b)$ is Dice coefficient between a and b, $\hat{x}$ and x^{gt} are the prediction result and the corresponding ground-truth, respectively. Our experimental results show that the combination of L_{dice} and L_{BCE} is sufficient to train our models.

2.6 | Semi-CARes-UNet

As we mentioned above, the high-quality labeled data is difficult to collect. In contrast, collecting a large amount of unlabeled data is feasible and less expensive. To harness the unlabeled samples in COVID-19 lung CT segmentation task, we combined our CARes-UNet with a semi-supervised learning method. An overview of our semi-supervised framework is shown in Figure 5. Specifically, we first trained CARes-UNet using 50 images with gold standard and used it to generate the pseudo labels of 1600 images without gold standard. Then, we utilized these pseudo labels as masks of the unlabeled data to train our semi-supervised model along with the labeled data. It is worth noting that the predictive pseudo labels may contain noise. The capacity of deep models is so high that they can memorize these noisy labels sooner or later during training.[32] To tackle this problem, we calculated the average entropy of each pseudo mask. We set the threshold of 0.6 manually to ensure that a sample with a generated pseudo mask would be selected only if its average entropy is lower than the threshold, which could reduce the noise in the pseudo labels. Similar threshold-based methods to filter noise are also used in other literature.[33,34] Also, we made sure that the labeled samples were as many as the samples with pseudo labels in every epoch during training.

Generally speaking, our semi-CARes-UNet has two advantages. First, our method is easy to implement. Using the trained supervised model, we can generate pseudo labels and treat them as the masks of the unlabeled CT slices. Second, the method can effectively enlarge our dataset and reduce over-fitting, which was confirmed by the recent study.[35]

3 | RESULTS AND DISCUSSION

3.1 | Datasets

Three datasets are used in our experiments, which are COVID-SemiSeg Dataset, Dataset of *COVID-19 Lung CT Lesion Segmentation Challenge–2020*, and UESTC-COVID-19 Dataset.

COVID-SemiSeg Dataset is a semi-supervised COVID-19 infection segmentation dataset built by Fan et al.[19] It consists of 1698 CT axial slices, in which 98 are labeled and 1600 are unlabeled. This COVID-19 infection segmentation dataset is open-source and can be accessed at https://github.com/DengPingFan/Inf-Net. Since our model was trained for lesion segmentation, we only made use of the lung infection dataset in COVID-SemiSeg Dataset and excluded its multi-class infection dataset. All the CT slices were collected by the Italian Society of Medical and Interventional Radiology. A radiologist segmented the CT images using different labels for identifying lung infections. Each sample in

the lung infection dataset is a grayscale image along with a corresponding ground truth mask. All the images have been preprocessed by the providers. In our experiments on COVID-SemiSeg Dataset, we took 50 labeled image slices for training and 48 labeled image slices for testing in supervised learning. The remaining 1600 unlabeled CT images were used for semi-supervised learning.

The Multi-national NIH Consortium for CT AI in COVID-19 provides an open-source dataset[36] for COVID-19 lung CT lesion segmentation via the NCI TCIA[37] public website to be used in *COVID-19 Lung CT Lesion Segmentation Challenge–2020*.[38] The training and validation part of this dataset consists of chest CTs from 199 and 50 patients, respectively, with positive RT-PCR for SARS-CoV-2 and ground truth annotations of COVID-19 lesions in the lung. Since the challenge was over, we only got the training and validation part of this dataset without the unseen testing samples. In our experiments on this dataset, we randomly selected training and testing samples with a ratio of approximately 3:1 from the received dataset. More precisely, 10 147 CT slices were used for training, and 3558 CT slices were used for testing. Before training, we cropped the images according to the bounding box of the lung region, resized to the same size, and normalized the images as the inputs of networks.

UESTC-COVID-19 Dataset[20] is used for lesion segmentation, which contains CT scans in the form of a 3D volume of 120 patients diagnosed with COVID-19. In our experiments on UESTC-COVID-19 Dataset, we also divided it into training and testing sets with a ratio of approximately 3:1. More precisely, 14 855 CT slices were used for training and 4056 CT slices were used for testing. All the images have been cropped by the dataset providers based on the bounding box of the lung region. In addition, the intensity has been normalized into a number ranging from 0 to 1 using window width/level of 1500/−650. Because the datasets of *COVID-19 Lung CT Lesion Segmentation Challenge–2020* and UESTC-COVID-19 Dataset are not designed for semi-supervised learning and do not include unlabeled samples needed in semi-supervised learning method, we only evaluate our proposed supervised method on these two datasets.

3.2 | Training details

All the models were implemented using Pytorch[39] and were trained on a PC equipped with GTX 2080Ti GPU. In each control experiment, all the networks shared the same training settings. Hyper-parameters were tuned either based on related researches and experiments[27,29] or by performing *k*-fold validation on the training set (two-fold validation on training set of COVID-SemiSeg Dataset, five-fold validation on training dataset of *COVID-19 Lung CT Lesion Segmentation Challenge–2020*, and five-fold validation on training set of UESTC-COVID-19 Dataset). In our experiments on COVID-SemiSeg Dataset, the initial learning rate was set to 5×10^{-4} and the training process finished after 1200 epochs. We used a weight decay of 0.0001 and a momentum of 0.9. In our experiments on the other two datasets, models were trained for 60 epochs. The initial learning rate was set to 5×10^{-4} too and was decayed by a factor of 0.1 at the 30th and the 50th epoch. Similarly, we also took a weight decay of 0.0001 and a momentum of 0.9. All the source code can be found at https://github.com/zylye123/CARes-UNet.

3.3 | Evaluation metrics

We used six different quantitative measurements, namely, Dice coefficient (*Dice*), sensitivity (*Sen*), specificity (*Spe*), structure measure (S_α), enhanced-alignment measure (E_ϕ^{mean}), and mean absolute error (MAE) to compare the segmentation results of the proposed method and baselines. We denoted S_p as final prediction after activated by a sigmoid function and G as ground truth. Then all the evaluation metrics adopted in our experiments could be formulated as follows:

1. *Dice*

$$Dice = \frac{2\left|S_p \cap G\right|}{\left|S_p\right| + |G|}. \tag{6}$$

2. *Sensitivity*

$$Sensitivity = \frac{TP}{TP + FN}. \tag{7}$$

3. *Specificity*

$$Specificity = \frac{TN}{TN + FP}. \tag{8}$$

4. *Structure Measure* (S_α)

$$S_\alpha = (1-\alpha) * S_o\left(S_p, G\right) + \alpha * S_r\left(S_p, G\right). \tag{9}$$

5. *Enhanced-alignment Measure* (E_ϕ^{mean})

$$E_\phi = \frac{1}{w \times h} \sum_x^w \sum_y^h \phi\left(S_p(x,y), G(x,y)\right). \tag{10}$$

6. *Mean Absolute Error (MAE)*

$$MAE = \frac{1}{w \times h} \sum_x^w \sum_y^h \left|S_p(x,y) - G(x,y)\right|, \tag{11}$$

TABLE 1 Model Comparison on coronavirus disease 2019 (COVID)-SemiSeg Dataset

Model	Dice ↑	Sen. ↑	Spec. ↑	S_α ↑	E_ϕ^{mean} ↑	MAE ↓
FCN[40]	0.622 (0.17)	0.508 (0.18)	0.952 (0.01)	0.676 (0.09)	0.783 (0.13)	0.0772 (0.05)
SegNet[16]	0.645 (0.16)	0.702 (0.19)	0.916 (0.09)	0.677 (0.11)	0.852 (0.13)	0.1031 (0.08)
UNet[15]	0.639 (0.17)	0.641 (0.18)	0.943 (0.04)	0.675 (0.10)	0.865 (0.09)	0.0890 (0.05)
LinkNet+xception[18]	0.689 (0.15)	0.704 (0.18)	0.942 (0.06)	0.710 (0.10)	0.891 (0.06)	0.0770 (0.03)
FPN+efficientnet-b1[18]	0.634 (0.15)	0.569 (0.19)	0.950 (0.05)	0.671 (0.08)	0.828 (0.07)	0.0830 (0.04)
PSPNet+vgg13[18]	0.685 (0.14)	0.661 (0.14)	0.954 (0.03)	0.710 (0.09)	0.889 (0.06)	0.0720 (0.05)
UNet+resnet34[18]	0.652 (0.12)	0.623 (0.15)	0.954 (0.07)	0.677 (0.10)	0.873 (0.08)	0.0860 (0.05)
Inf-Net[19]	0.682	0.692	0.943	0.781	0.838	0.082
Semi-Inf-Net[19]	0.739	0.725	0.960	0.800	0.894	0.064
CARes-UNet (ours)	0.731 (0.11)	0.738 (0.15)	0.955 (0.03)	0.746 (0.08)	0.913 (0.05)	0.0694 (0.05)
Semi-CARes-UNet (ours)	0.776 (0.10)	0.786 (0.13)	0.961 (0.02)	0.789 (0.07)	0.931 (0.04)	0.0561 (0.03)

Note: Average various metrics comparison of different models on COVID-SemiSeg Dataset. Red, blue, and cyan colors are used to indicate to first, second, third rank, respectively. ↑ indicates the metric is the higher the better while ↓ indicates the metric is the lower the better. Numbers in parentheses are standard deviations. Metrics of Inf-Net and semi-Inf-Net are directly cited from Fan et al.[19]

Abbreviations: CARes-UNet, content-aware residual UNet; E_ϕ^{mean}, enhanced-alignment measure; Dice, Dice coefficient; MAE mean absolute error; Sen, sensitivity; Spe, specificity, S_α, structure measure.

where *TP*, *TN*, *FP*, and *FN* represent true positive, true negative, false positive, and false negative between S_p and *G*.

3.4 | Baselines

The experiment baselines were set as follows. For a more comprehensive comparison, we compared our proposed CARes-UNet with U-Net,[15] SegNet,[16] FCN,[40] several models perform best in the method by Bizopoulos et al.,[18] Inf-Net,[19] and Semi-Inf-Net.[19] Among all, the Semi-Inf-Net achieved the best performance in the COVID-19 infection segmentation task in all evaluation metrics mentioned above.[19] In our experiments on COVID-SemiSeg Dataset, all the baseline models shared the same training and testing scenario, which is using 50 labeled image slices for training and 48 labeled image slices for testing. But for Semi-Inf-Net, which was trained in a semi-supervised way, 1600 unlabeled image slices were also used in training. Also, in our experiments on the other two datasets, all the baseline models shared the same training and testing scenario where the ratio of the training and testing samples is approximately 3:1.

3.5 | Experiment results

Quantitative results of baselines and our proposed CARes-UNet and semi-CARes-UNet on the COVID-SemiSeg Dataset are summarized in Table 1. As shown in Table 1, semi-CARes-UNet outperforms all the compared baselines in terms of Dice, sensitivity, specificity, E_ϕ^{mean}, MAE. Compared with UNet, Res-UNet, and UNet+resnet34, our CARes-UNet captured and utilized plentiful CT slice content information during upsampling and results in more accurate segmentation accuracy. Compared with Inf-Net, our model achieved better performance without the need of using complex network architecture.

Moreover, we observed from Table 1 that the model could gain from the semi-supervised learning method. For instance, semi-CARes-UNet obtained up to 4.5% improvement in terms of Dice, compared with CARes-UNet. These results reveal that semi-supervised learning can reduce over-fitting and have a regularising effect. However, only 4.5% improvement can be observed from adding 1600 unlabeled images with pseudo labels. This is because the pseudo labels may contain noise raised from the model's bias. Using the methods reported by Lee and colleagues,[41–43] dealing with the noisy labels may further increase the model's performance. In addition, the results also support the point raised by Castro et al.[44] that semi-supervised learning has limited benefits in causal tasks like segmentation.

Quantitative results on the dataset of *COVID-19 Lung CT Lesion Segmentation Challenge–2020* and UESTC-COVID-19 Dataset are presented in Tables 2 and 3, respectively. On these two large datasets for COVID-19 lesion segmentation, our CARes-UNet also gained the best results among all the models, especially in terms of Dice, sensitivity, S_α and E_ϕ^{mean}. These experimental results show that though expanding the small dataset to a relatively large dataset, our CARes-UNet could still attain high COVID-19 lesion segmentation performance.

Qualitative results are shown in Figure 6. Visually it reveals that CARes-UNet and semi-CARes-UNet detected the COVID-19 lesion areas more precisely than any other baseline model. Our proposed CARes-UNet and semi-CARes-UNet, compared with other baseline models, could reduce the area of the false negative regions while restricting the enlargement

TABLE 2 Model Comparison on COVID-19 Lung CT Lesion Segmentation Challenge – 2020 dataset

Model	Dice ↑	Sen. ↑	Spec. ↑	S_α ↑	E_ϕ^{mean} ↑	MAE ↓
FCN[40]	0.622 (0.17)	0.508 (0.18)	0.952 (0.01)	0.676 (0.09)	0.783 (0.13)	0.0772 (0.05)
SegNet[16]	0.645 (0.16)	0.702 (0.19)	0.916 (0.09)	0.677 (0.11)	0.852 (0.13)	0.1031 (0.08)
UNet[15]	0.639 (0.17)	0.641 (0.18)	0.943 (0.04)	0.675 (0.10)	0.865 (0.09)	0.0890 (0.05)
LinkNet+xception[18]	0.689 (0.15)	0.704 (0.18)	0.942 (0.06)	0.710 (0.10)	0.891 (0.06)	0.0770 (0.03)
FPN+efficientnet-b1[18]	0.634 (0.15)	0.569 (0.19)	0.950 (0.05)	0.671 (0.08)	0.828 (0.07)	0.0830 (0.04)
PSPNet+vgg13[18]	0.685 (0.14)	0.661 (0.14)	0.954 (0.03)	0.710 (0.09)	0.889 (0.06)	0.0720 (0.05)
UNet+resnet34[18]	0.652 (0.12)	0.623 (0.15)	0.954 (0.07)	0.677 (0.10)	0.873 (0.08)	0.0860 (0.05)
Inf-Net[19]	0.682	0.692	0.943	0.781	0.838	0.082
semi-Inf-Net[19]	0.739	0.725	0.960	0.800	0.894	0.064
CARes-UNet (ours)	0.731 (0.11)	0.738 (0.15)	0.955 (0.03)	0.746 (0.08)	0.913 (0.05)	0.0694 (0.05)
semi-CARes-UNet (ours)	0.776 (0.10)	0.786 (0.13)	0.961 (0.02)	0.789 (0.07)	0.931 (0.04)	0.0561 (0.03)

Average various metrics comparison of different models on dataset of *COVID-19 Lung CT Lesion Segmentation Challenge–2020*. Red, blue, and cyan colors are used to indicate top first, second, and third rank, respectively. ↑ indicates the metric is the higher the better while ↓ indicates the metric is the lower the better. Numbers in parentheses are standard deviation.

TABLE 3 Model Comparison on UESTC-COVID-19 Dataset

Model	Dice ↑	Sen. ↑	Spec. ↑	S_α ↑	E_ϕ^{mean} ↑	MAE ↓
FCN[40]	0.635 (0.48)	0.637 (0.48)	0.650 (0.48)	0.818 (0.24)	0.740 (0.35)	0.0141 (0.04)
SegNet[16]	0.668 (0.42)	0.682 (0.42)	0.773 (0.42)	0.852 (0.21)	0.821 (0.30)	0.0109 (0.03)
UNet[15]	0.652 (0.43)	0.658 (0.43)	0.724 (0.44)	0.899 (0.13)	0.900 (0.15)	0.0074 (0.01)
LinkNet+xception[18]	0.708 (0.39)	0.718 (0.40)	0.813 (0.39)	0.907 (0.16)	0.898 (0.16)	0.0075 (0.02)
FPN+efficientnet-b1[18]	0.698 (0.37)	0.705 (0.38)	0.820 (0.37)	0.901 (0.16)	0.910 (0.20)	0.0077 (0.02)
PSPNet+vgg13[18]	0.712 (0.38)	0.715 (0.39)	0.826 (0.38)	0.910 (0.15)	0.901 (0.17)	0.0074 (0.02)
UNet+resnet34[18]	0.710 (0.37)	0.721 (0.38)	0.831 (0.37)	0.901 (0.15)	0.901 (0.20)	0.0073 (0.02)
CARes-UNet (ours)	0.755 (0.37)	0.741 (0.38)	0.835 (0.36)	0.911 (0.15)	0.925 (0.18)	0.0080 (0.02)

Note: Average various metrics comparison of different models on UESTC-COVID-19 Dataset. Red, blue, and cyan colors are used to indicate top first, second, and third rank, respectively. ↑ indicates the metric is the higher the better while ↓ indicates the metric is the lower the better. Numbers in parentheses are standard deviation.

of the false positive regions. Especially the outcomes of semi-CARes-UNet are very close to the ground truth with the smallest area of false negative and false positive regions among all the results. We noticed that UNet, FCN, and SegNet gave unsatisfying segmentation outcomes where they segmented most of the small disconnected regions wrongly. This experimental result reveals that the plain UNet architecture was insufficient to capture the complex texture of chest CT images. Our CARes-UNet tackled these drawbacks by redesigning UNet architecture collaborated with an advanced optimizer. Also, our semi-CARes-UNet further boosted the performance using the semi-supervised learning method.

3.6 | Ablation study

In this subsection, we conducted several experiments on COVID-SemiSeg Dataset to validate the effectiveness of each key component of our proposed model, including residual connection, content-aware upsampling, and Ranger optimizer.

3.6.1 | Efficacy of residual connection

To investigate the effectiveness of residual connection, we designed two sets of controlled trials. We first set up a baseline network without the residual connections called CA-UNet. We then added residual connections to the network and denoted it as CARes-UNet. The results of two sets of controlled trials are summarized in Figure 7. It reveals that Res-UNet performs much better than plain UNet. Indeed, training a network with residual connections can be thought of as training an ensemble of different models on the dataset and getting the best possible accuracy.

3.6.2 | Efficacy of CARAFE

To verify the performance of CARAFE, we trained three models: U-Net upsampling with deconvolution layer, sub-pixel layer, and CARAFE. Quantitative results of the experiments are shown in Figure 8. We observed that U-Net with CARAFE outperforms U-Net with

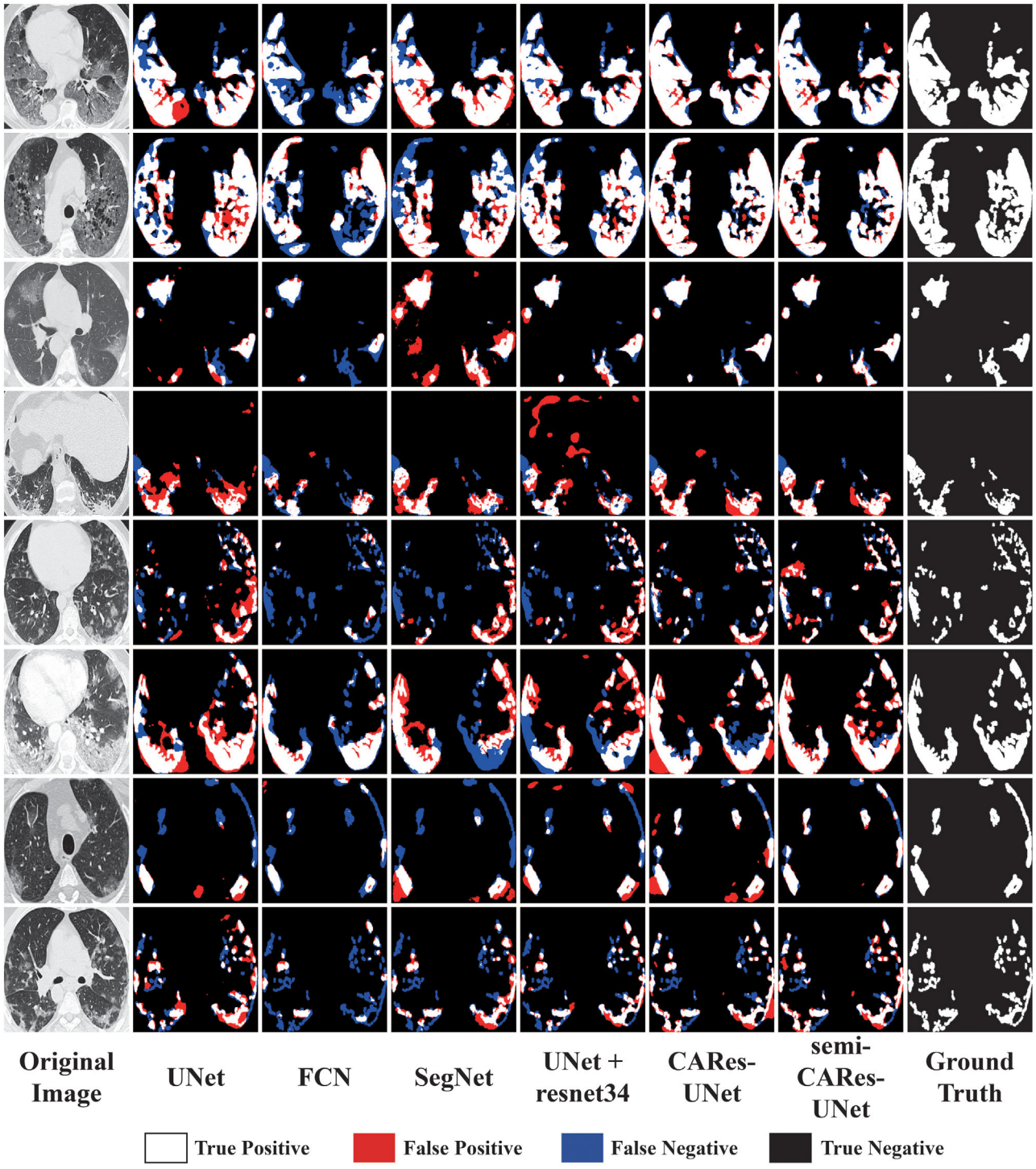

FIGURE 6 Qualitative analysis of different models. The first column on the left is the original computed tomography (CT) slices, and the first column on the right is their corresponding ground truth masks. Columns between them are images of lesion areas predicted by different models. White, red, blue, and black regions identify true positive, false positive, false negative, and true negative regions, respectively

deconvolution layer and U-Net with a sub-pixel layer in terms of various metrics. These results suggest that using CARAFE on UNet can improve the model's performance.

3.6.3 | Efficacy of Ranger optimizer

We also investigated the importance of the Ranger optimizer. As presented in Figure 9, though average Dice,

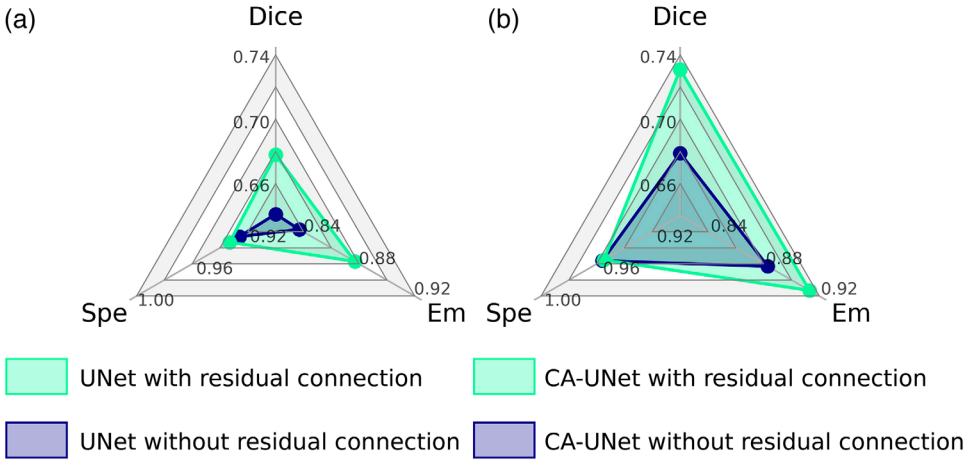

FIGURE 7 Comparison of models with and without residual connection. (a) UNet. (b) CA-UNet. All metrics are the higher the better. Models with residual connection are better than ones without residual connection

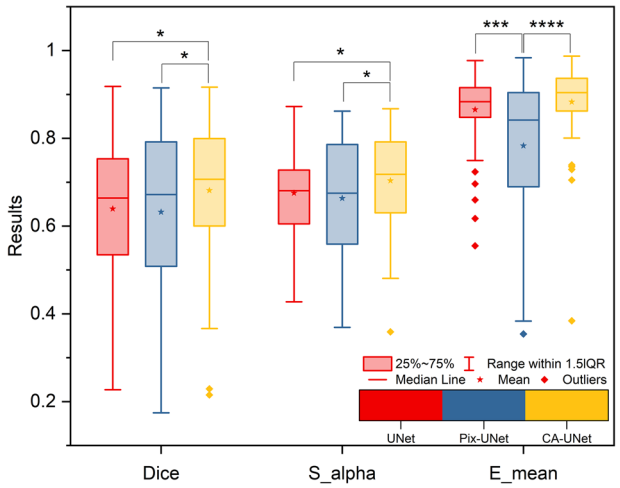

FIGURE 8 Comparison of UNet (without residual connection) with different upsampling blocks. Different colors represent different upsampling blocks and all metrics are the higher the better. The number of asterisks indicates the statistical significance calculated without outliers

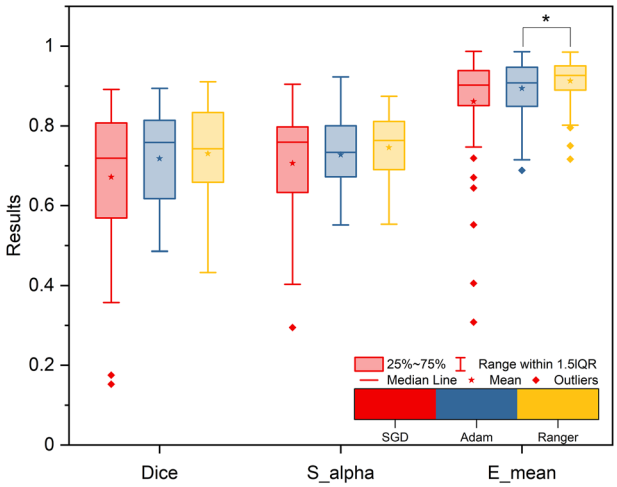

FIGURE 9 Comparison of CARes-UNet with different optimizers (learning rate is set to 5×10^{-4}). Different colors represent different optimizers, and all metrics are the higher the better. The number of asterisks indicates the statistical significance calculated without outliers. The test results were recorded after the models converged given enough training period

S_α and E_ϕ^{mean} on testing set of Ranger optimizer are higher than average Dice, S_α, and E_ϕ^{mean} of SGD and Adam optimizer, there will not be much significant difference if all the models are given enough training time to converge. However, if we compare the time consumed during the model convergence or performance on the testing set given a limited training period, the Ranger optimizer could perform better than the other optimizers. Training and testing loss versus epochs of three optimizers are depicted in Figure 10. It can be observed that the Ranger and Adam optimizer has a faster convergence than the SGD optimizer. The training loss of all the optimizers decreased fast before 100 epochs and then went up slowly, while the Ranger optimizer decreased faster than the others. Ranger optimizer showed less variance than Adam optimizer in testing as shown in Figure 10b, which can be inferred that the RAdam together with Lookahead strategy is able to improve learning stability.

3.7 | Discussion on parameters and computational efficiency

We compared different models about their parameters and computational efficiency. The results are shown in Table 4. multiply-adds (MAdd) and floating-point operations per second (FLOPs) represent the computational efficiency of a model. We observed that the total amount of parameters, MAdd and Flops increases

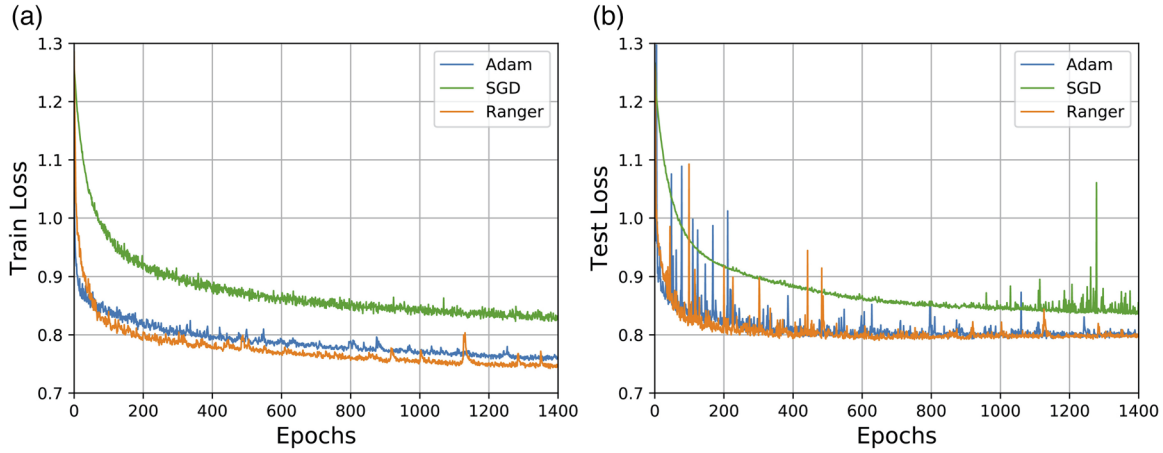

FIGURE 10 Comparison of convergence using different optimization algorithms (a) training curve of three optimizers. (b) Testing curve of three optimizers. Different colors represent different optimizers. Ranger has a faster convergence and less variance than SGD and Adam optimizer

TABLE 4 Comparison of models' attributes

Model	Total params	Total memory	Total multiply-adds (MAdd)	Total floating-point operations per second (FLOPs)	Total MemR+W
FCN	23.61 M	147.25 MB	10.87 G	5.45 G	401.99 MB
MCFCN	23.61 M	149.50 MB	10.88 G	5.45 G	406.49 MB
SegNet	29.44 M	224.50 MB	80.07 G	40.06 G	561.07 MB
UNet	18.81 M	221.50 MB	58.14 G	24.26 G	474.59 MB
Inf-Net	31.07 M	190.15 MB	14.72 G	7.37 G	524.44 MB
Pix-UNet	9.48 M	194.50 MB	24.31 G	12.17 G	394.92 MB
CA-UNet	16.73 M	250.11 MB	60.62 G	30.34 G	644.48 MB
PixRes-UNet	40.97 M	467.00 MB	92.05 G	46.06 G	1.04 GB
Res-UNet	59.63 M	579.00 MB	159.74 G	75.08 G	1.31 GB
CARes-UNet (ours)	54.42 M	592.61 MB	157.38 G	78.75 G	1.44 GB

Note: Comparison of models about total params, memory, MAdd, FLOPs and MemR+W measured by torchstat, a lightweight neural network analyzer based on Pytorch, in the same experimental setup. FLOPs and MAdd are units of measure for the computational performance.

inevitably when introducing residual structure into networks, leading to more total parameters and lower computational efficiency. But these can be accepted considering that the residual structure used in our proposed method could effectively improve the lesion segmentation results.

4 | CONCLUSION

In this study, we present a content-aware deep residual UNet for lung infection segmentation of COVID-19 from CT slices. We redesigned the UNet architecture by adding residual connection and replacing deconvolution layers with content-aware upsampling layers to deal with the within-class variation and between-class similarity. We also employed the Ranger optimizer to expedite convergence and improve learning stability. To tackle the problem that high-quality labeled data is difficult to collect as well as to utilize a large amount of existing unlabeled data, we applied the semi-supervised learning method to our CARes-UNet called semi-CARes-UNet. We conducted experiments and ablation studies to prove the effectiveness and robustness of our method. Results on three public datasets showed that our method outperforms several state-of-the-art methods.

ACKNOWLEDGMENTS

This work was supported by Guangzhou science and technology fund (Grant No. 201803010072), Science, Technology&Innovation Commission of Shenzhen

Municipality (JCYL 20170818165305521), and China Medical University Hospital (DMR-107-067, DMR-108-132, DMR-110-097). The annotation of the dataset was made possible through the joint work of Children's National Hospital, NVIDIA, and National Institutes of Health for the COVID-19-20 Lung CT Lesion Segmentation Grand Challenge.

This work was supported by National Natural Science Foundation of China (Grand No. 62176272), Guangzhou Science and Technology Fund (Grant No. 201803010072), Science, Technology&Innovation Commission of Shenzhen Municipality (JCYL 20170818165305521), and China Medical University Hospital (DMR-107-067, DMR-108-132, DMR-110-097).

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://github.com/DengPingFan/Inf-Net, reference number.[19]

The data that support the findings of this study are openly available at https://covid-segmentation.grand-challenge.org/Download/, reference number.[36–38]

The data that support the findings of this study are available from the University of Electronic Science and Technology of China. Restrictions apply to the availability of these data, which were used under license for this study. Data are available at https://faculty.uestc.edu.cn/HiLab/en/article/379152/content/3319.htm with the permission of the University of Electronic Science and Technology of China, reference number.[20]

REFERENCES

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet*. 2020;395(10223):470-473.
2. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395(10223):497-506.
3. Fang Y, Zhang H, Xie J, et al. Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology*. 2020;296(2):200432.
4. Ai T, Yang Z, Hou H, et al. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology*. 2020;296(2):E32-E40.
5. Lessmann N, Sánchez CI, Beenen L, et al. Automated assessment of COVID-19 reporting and data system and chest CT severity scores in patients suspected of having COVID-19 using artificial intelligence. *Radiology*. 2021;298(1):E18-E28.
6. Zhang K, Liu X, Shen J, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell*. 2020;181(6):1423-1433.
7. Harmon SA, Sanford TH, Xu S, et al. Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat Commun*. 2020;11(1):1-7.
8. Jin C, Chen W, Cao Y, et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. *Nat Commun*. 2020;11(1):1-14.
9. Bai HX, Wang R, Xiong Z, et al. AI augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other etiology on chest CT. *Radiology*. 2020;296(3):201491.
10. Wang X, Deng X, Fu Q, et al. A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT. *IEEE Trans Med Imaging*. 2020;39(8);2615-2625.
11. Tan M, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. 2019. arXiv Prepr arXiv190511946.
12. Wang Z, Liu Q, Dou Q. Contrastive cross-site learning with redesigned net for COVID-19 CT classification. *IEEE J Biomed Health Inform*. 2020;24(10):2806-2813.
13. Owais M, Lee YW, Mahmood T, Haider A, Sultan H, Park KR. Multilevel deep-aggregated boosted network to recognize COVID-19 infection from large-scale heterogeneous radiographic data. *IEEE J Biomed Health Inform*. 2021;25(6):1881-1891.
14. Minaee S, Boykov YY, Porikli F, Plaza AJ, Kehtarnavaz N, Terzopoulos D. Image segmentation using deep learning: a survey. *IEEE Trans Pattern Anal Mach Intell*. 2021.
15. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention, October 5-9, 2015; Munich, Germany.
16. Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(12):2481-2495.
17. Saood A, Hatem I. COVID-19 lung CT image segmentation using deep learning methods: u-Net versus SegNet. *BMC Med Imaging*. 2021;21(1):1-10.
18. Bizopoulos P, Vretos N, Daras P. Comprehensive comparison of deep learning models for lung and COVID-19 lesion segmentation in CT scans. 2020. arXiv Prepr arXiv200906412.
19. Fan D-P, Zhou T, Ji G-P, et al. Inf-Net: automatic COVID-19 lung infection segmentation from CT images. *IEEE Trans Med Imaging*. 2020;39(8):2626-2637.
20. Wang G, Liu X, Li C, et al. A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT Images. *IEEE Trans Med Imaging*. 2020;39(8):2653-2663.
21. Zhou L, Li Z, Zhou J, et al. A rapid, accurate and machine-agnostic segmentation and quantification method for CT-based COVID-19 diagnosis. *IEEE Trans Med Imaging*. 2020;39(8):2638-2652.
22. Gao K, Su J, Jiang Z, et al. Dual-branch combination network (DCN): towards accurate diagnosis and lesion segmentation of COVID-19 using CT images. *Med Image Anal*. 2020;67:101836.
23. Wu Y-H, Gao S-H, Mei J, et al. JCS: an explainable COVID-19 diagnosis system by joint classification and segmentation. *IEEE Trans Image Process*. 2021;30:3113-3126.
24. Yang Z, Zhao L, Wu S, Chen Y-C. Lung lesion localization of COVID-19 from chest CT image: a novel weakly supervised learning method. *IEEE J Biomed Health Inform*. 2021;25(6):1864-1872.
25. Laradji I, Rodriguez P, Manas O, et al. A weakly supervised consistency-based learning method for covid-19 segmentation in ct images. Paper presented at: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision; January 3-8, 2021; Waikoloa, HI.
26. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. 2015. *arXiv Prepr arXiv150203167*.
27. Han D, Kim J, Kim J. Deep pyramidal residual networks. Paper presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; July 21-26, 2017; Honolulu, HI.
28. Qiu Q, Yang Z, Wu S, et al. Automatic segmentation of hippocampus in hippocampal sparing whole brain radiotherapy: a multi-task edge-aware learning. *Med Phys*. 2021; 48(4):1771-1780.
29. Wang J, Chen K, Xu R, Liu Z, Loy CC, Lin D. Carafe: content-aware reassembly of features. Paper presented at: Proceedings

of the IEEE/CVF International Conference on Computer Vision; October 27-28, 2019; Seoul, South Korea.
30. Liu L, Jiang H, He P, et al. On the variance of the adaptive learning rate and beyond. 2019. *arXiv Prepr arXiv190803265*.
31. Zhang MR, Lucas J, Hinton G, Ba J. Lookahead optimizer: k steps forward, 1 step back. 2019. *arXiv preprint arXiv:1907.08610*.
32. Han B, Yao Q, Yu X, et al. Co-teaching: robust training of deep neural networks with extremely noisy labels. *Adv Neural Inf Process Syst*. 2018.
33. Wang K, Zhang D, Li Y, Zhang R, Lin L. Cost-effective active learning for deep image classification. *IEEE Trans Circuits Syst Video Technol*. 2017;27(12):2591-2600.
34. Xie Q, Luong M-T, Hovy E, Le QV. Self-training with noisy student improves imagenet classification. In: Boult T, Medioni G, Zabih R, Mortensen E, Masson M, Institute of Electrical and Electronics Engineers; Computer Vision Foundation, eds. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE; 2020:10684-10695. doi:10.1109/CVPR42600.2020.01070
35. Mittal S, Tatarchenko M, Çiçek Ö, Brox T. Parting with illusions about deep active learning. 2019. arXiv Prepr arXiv191205361.
36. An P, Xu S, Harmon SA, Turkbey EB, et al. CT images in Covid-19 [Data set]. *Cancer Imaging Arch*. 2020.
37. Clark K, Vendt B, Smith K, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26(6):1045-1057.
38. Roth H, Xu Z, Diez CT, et al. Rapid artificial intelligence solutions in a pandemic–The COVID-19-20 lung CT lesion segmentation challenge. *Res Sq*. 2021.
39. Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library. 2019. arXiv Prepr arXiv191201703.
40. Voulodimos A, Protopapadakis E, Katsamenis I, Doulamis A, Doulamis N. Deep learning models for COVID-19 infected area segmentation in CT images. 2020. *medRxiv*.
41. Lee H, Jeong W-K. Scribble2Label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency. In: Martel AL, Abolmaesumi P, Stoyanov D, Mateus D, Zuluaga MA, Zhou SK, Racoceanu D, Joskowicz L, eds. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2020:14-23.
42. Yu L, Wang S, Li X, Fu C-W, Heng P-A. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In: Shen D, Liu T, Staib LH, Essert C, Zhou S, Yap P-T, Khan A, eds. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2019:605-613.
43. Sedai S, Antony B, Rai R, et al. Uncertainty guided semi-supervised segmentation of retinal layers in OCT images. In: Shen D, Liu T, Staib LH, Essert C, Zhou S, Yap P-T, Khan A, eds. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2019:282-290.
44. Castro DC, Walker I, Glocker B. Causality matters in medical imaging. *Nat Commun*. 2020;11(1):3673.

How to cite this article: Xu X, Wen Y, Zhao L, et al. CARes-UNet: content-Aware residual UNet for lesion segmentation of COVID-19 from chest CT images. *Med. Phys.* 2021;48:7127–7140. https://doi.org/10.1002/mp.15231