SAUNet++: An automatic segmentation model of COVID-19 lesion from CT slices

Hanguang XIAO

Chongqing University of Technology

Zhiqiang RAN (✉ zqran@2019.cqut.edu.cn)

Chongqing University of Technology

Shingo MABU

Yamaguchi University Graduate School of Sciences and Technology for Innovation: Yamaguchi Daigaku
Daigakuin Sosei Kagaku Kenkyuka

Banglin ZHANG

Chongqing University of Technology

Bolong ZHANG

Chongqing University of Technology

Chang LIU

Chongqing University of Technology

Research Article

Keywords: coronavirus disease 2019 (COVID-19), image segmentation, computed tomography (CT), squeeze excitation residual (SER), atrous spatial pyramid pooling (ASPP), generalized dice loss (GDL)

Posted Date: August 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-812932/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at The Visual Computer on February 8th, 2022. See the published version at <https://doi.org/10.1007/s00371-022-02414-4>.

SAUNet++: An automatic segmentation model of COVID-19 lesion from CT slices

Hanguang Xiao¹ · Zhiqiang Ran^{1,2} · Shingo Mabu² ·
Banglin Zhang¹ · Bolong Zhang¹ · Chang Liu¹

Received: date / Accepted: date

Abstract The coronavirus disease 2019 (COVID-19) epidemic has spread worldwide and the healthcare system is in crisis. Accurate, automated and rapid segmentation of COVID-19 lesion in computed tomography (CT) images can help doctors diagnose and provide prognostic information. However, the variety of lesions and small regions of early lesion complicate their segmentation. To solve these problems, we propose a new SAUNet++ model with squeeze excitation residual (SER) module and atrous spatial pyramid pooling (ASPP) module. The SER module can assign more weights to more important channels and mitigate the problem of gradient disappearance, the ASPP module can obtain context information by atrous convolution using various sampling rates. In addition, the generalized dice loss (GDL) can reduce the correlation between lesion size and dice loss, and is introduced to solve the problem of small regions segmentation. We collected multinational CT scan data from China, Italy and Russia and conducted extensive experiments. In the experiments, SAUNet++ and GDL were compared to advanced segmentation models and popular loss functions, respectively. The experimental results demonstrated that our methods can effectively improve the accuracy of COVID-19 lesion segmentation on the dice similarity coefficient (our: 87.38% VS U-Net++: 86.08%), sensitivity (our: 93.28% VS U-Net++: 89.85%) and hausdorff distance (our: 19.99mm VS U-Net++: 27.69mm), respectively.

Keywords coronavirus disease 2019 (COVID-19) · image segmentation, computed tomography (CT) · squeeze excitation residual (SER) · atrous spatial pyramid pooling (ASPP) · generalized dice loss (GDL)

1 Introduction

COVID-19 caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was highly contagious, and world health organization (WHO) announced that the COVID-19 outbreak had entered a “global pandemic”. Up to August 10, 2021, there have been 203,295,170 confirmed cases of COVID-19, including 4,303,515 deaths, have been reported to the WHO [1]. Mutated strains have emerged in many countries at present, and even if a variety of vaccines have been developed, there is no guarantee that they will be safe and effective against all viruses. Rapid and accurate detection of COVID-19 is essential to prevent the spread of the outbreak.

Reverse transcription polymerase chain reaction (RT-PCR) is the gold standard of COVID-19 detection. Nasopharyngeal and oropharyngeal swabs are frequently used as screening samples for nucleic acid in clinical settings. But in practice, it is found that the detection of nucleic acid by swabbing the pharynx is easy to lead to false negatives because of the quality of samples and the low viral load of the pharynx [2]. For patients with clinical symptoms, it is easy to delay treatment and increase the risk of infection among medical staff. For asymptomatic infection, it can easily lead to misdiagnosis and increased risk of transmission. An important addition to RT-PCR, CT offers the advantages of non-invasiveness, high resolution and low noise [3]. It can clearly describe the image characteristics related to COVID-19, such as ground-glass opacity (GGO) and lung consolidation as shown in Fig. 1. In the early stage of the disease, there were mainly multiple small patches in the extrapulmonary zone. After the development of the disease, GGO appeared in many lung lobes, which at the same time caused the thickening of interlobular septum and the appearance of ‘crazy-paving sign’. After entering the severe stage, both lungs showed diffuse lesions, mainly lung consolidation, accompanied by distortion of lung structure, bronchiectasis, pleural effusion and so on [4]. Accurate segmentation of COVID-19 lesion is useful in evaluating the degree of lung infection and disease progression, and provides a foundation for follow-up therapy of patients. However, manual labelling of

Zhiqiang Ran
E-mail: zqran@2019.cqut.edu.cn

¹ School of Artificial Intelligence, Chongqing University of Technology, Chongqing 401135, China

² Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Yamaguchi 755-8611, Japan

Fig. 1 COVID-19 CT performance: (a) the lung contains multiple GGO, (b) red covers the GGO lesion, (c) there are a large number of lung consolidation in both lungs, and (d) the red covers the consolidation and GGO lesion.

lesion is a time-consuming process, and precision is largely affected by the subjective knowledge and experience of doctors.

In recent years, the medical image segmentation model based on deep learning has shown satisfactory results in assisting the diagnosis of human diseases. These models can automatically extract important image features that are difficult for doctors to detect, improving the effectiveness of disease diagnosis [5] [6]. Ronneberger et al. [7] proposed U-Net with symmetric encoder-decoder structure, which is the most classic biomedical image segmentation network. U-Net++ [8] combines four structures of different depths and uses dense skip connections of different lengths to enhance U-Net. Attention U-Net [9] introduces the attention mechanism to focus on the target and suppress the background features. 3D U-Net [10] converts all convolution operations in U-Net to 3D in order to restore 3D information. At present, many improved methods based on U-Net are used to segment organs or lesions. However, the following issues remain in the segmentation of COVID-19 lesion [11]. 1) The diversity of lesions: There is a large difference in texture, position, shape and other characteristics between lesions, which increases the difficulty of segmentation. 2) The small region of early lesion: Early detection of COVID-19 patients is beneficial for subsequent treatments, but early GGO is often difficult to be recognized because of its small regions, low contrast, and blurred appearance. 3) Lack of datasets with professional labels: The COVID-19 urgency makes it challenging to collect enough data to train segmentation models, and it is expensive and time-consuming to obtain high-quality pixel-level masks.

In this study, we propose a deep learning model called (squeeze excitation residual atrous spatial pyramid pooling U-Net++, SAUNet++) to segment the COVID-19 lesion from CT images. The contributions made by this work are threefold:

- On the basis of U-Net++ model, squeeze excitation residual (SER) [12] module and atrous spatial pyramid pooling (ASPP) [13] module are introduced, which has not been tried in the relevant studies. The SER module can assign more weight to more important channels and mitigate the problem of gradient disappearance, while the ASPP module can obtain context information by atrous convolution using various sampling rates.
- The generalized dice loss (GDL) is introduced to solve the problem of small regions segmentation. Compared to the common dice loss, it can balance the contribution of each region to the loss and make the training more stable.
- We collected multinational CT scans from four different datasets. A total of 11359 slices with high-quality pixel-level masks were used for the experiment.

2 Related Work

2.1 Lung Parenchyma Segmentation

Computer aided diagnosis (CAD) system based on CT images aims to provide valuable ‘second opinions’ to doctors in order to reduce their workload and improve the accuracy of lung disease diagnosis [14] [15]. CAD systems are usually divided into four stages: preprocessing, region of interest (ROI) detection, feature extraction and result evaluation. Lung parenchyma segmentation is one of the important steps of ROI detection, because most of the lesions only exist in the lung parenchyma. Therefore, it is important to accurately segment the lung parenchyma region, minimize the searching area, and maximize the sensitivity of lesion detection [16]. Kumar et al. [17] firstly used iterative threshold method to roughly segment lung parenchyma region, then used improved bidirectional chain code method to obtain the judgment points of lung boundary in vertical and horizontal directions. Finally, on the basis of support vector machine (SVM) analysis of position, distance, concavity ratio and other information, the lung boundary is modified by midpoint method. Nithila et al. [18] proposed an active contour model based on signed pressure force function. Firstly, the level set function is regularized through the Gaussian filter to get the initial contour, then contour constant is added to SPF function, and the pressure sign is adjusted to make the outline of the object shrink or expand. Finally, edge stop function is used to pull the contour to the lung parenchyma boundary. However, the traditional machine learning methods based on threshold and active contours have the shortcomings of over-segmentation and modeling difficulties. In contrast, deep learning methods are easier to achieve full automation. At present, a variety of deep learning networks have been applied to lung parenchyma segmentation. Geng et al. [19] proposed a lung parenchyma segmentation method based on VGG-16 and atrous convolution. In order to avoid gradient disappearance or redundant calculation during model training, both Khanna et al. [20] and Zhang et al. [21] introduced residual structure into U-Net. Tan et al. [22] proposed lung generative adversarial networks (GAN), U-Net is used to generate the predicted lung parenchyma mask, and CNN is used to discriminate the authenticity of the mask. The classification results of CNN help U-Net to generate more accurate masks, and these images are helpful to improve the classification accuracy. As a result, the performance of both networks is continuously improved in the iterative process. To test the noise suppression capability of SAUNet++, our method did not segment the lung parenchyma before segmenting the COVID-19 lesion.

2.2 COVID-19 Lesion Segmentation

In order to quickly screen COVID-19, many deep learning methods for segmentation of COVID-19 lesion have been proposed, which mainly focus on improving the network architecture. Zhao et al. [23] proposed D2A U-Net for segmenting COVID-19 lesions in CT slices based on atrous convolution and dual attention mechanism. Zheng et al. [4] proposed a multi-scale discriminative network, which includes channel attention module, pyramid convolution module and residual refinement module. Joseph Raj et al. [24] proposed ADID-UNET, in which convolution and pooling layers were replaced by dense network blocks to speed up information transmission, attention gate is used to suppress the background, and improved atrous convolution is used to obtain more edge features. Amyar et al. [25] proposed a deep learning model that combines segmentation, classification and reconstruction tasks. The model can improve segmentation performance by using relevant information in different tasks. Some studies have improved the loss function. Wang et al. [26] proposed noise-robust dice loss function, where COPLE-Net and adaptive self-ensembling method can effectively segment COVID-19 lesion from the noise-containing slices. Zhou et al. [27] also used the attention mechanism in the channel and spatial dimensions, and focal tversky loss (FTL) is introduced to focus on small lesion segmentation by adjusting the coefficients. However, due to the limited annotated data, it is difficult to train the model with strong generalization. In order to get more data, Zhang et al. [28] proposed a GAN network with pyramid structure, in which each GAN can generate CT slices with different sizes, and two-stage generators can strengthen constraints and enrich image details. Abdel-Basset et al. [29] proposed an semi-supervised few-shot segmentation method with dual-path architecture. The conditioner path learns the relevant knowledge representation, and then share it to the segmentation path to achieve new image

segmentation. The framework can effectively segment COVID-19 lesion with a limited number of data. In our study, we proposed SER module and ASPP module to improve U-Net++ instead of U-Net, introduced GDL to compare with FTL, and collected more images to train the model.

3 Methodology

In this section, we describe the details of SAUNet++ in terms of network architecture, core modules and loss function.

3.1 SAUNet++

The SAUNet++ architecture is shown in Fig. 2, which is improved based on U-Net and U-Net++. The original U-Net utilizes four downsamplings to obtain high-dimensional information, and then uses four upsamplings to restore the feature map to the original size, and four skip connections are used to merge the feature maps [7]. The improved methods based on U-Net do not always use four-layer structures, however, the combination of downsampling, upsampling and skip connection remains unchanged. Different features can be extracted from the structures of different depths, which are all important for segmentation, so U-Net++ combines four structures of different depths and uses skip connections of different lengths [8]. Because U-Net++ uses unpadded convolution, the boundary pixels of the feature map in the same layer are missing after two convolutions, so the feature map in the contracting path needs to be cropped first and then connected with the feature map in the expanding path. To enhance the structure of U-Net++, we use the SER modules to replace the original two convolutions, and add the ASPP module at the end of downsampling.

Fig. 2 The architecture of our proposed SAUNet++ model, which mainly consists of SER modules, an ASPP module, downsampling, upsampling and skip connections. The network receives an input of CT slice and directly outputs the mask with the same size as the slice, where the white region represents the COVID-19 lesion. Downsampling is achieved by maxpooling. The numbers in the circle represent the numbers of channels in the feature maps.

3.2 SER

Our proposed channel attention mechanism called SER module is built upon the squeeze and excitation block and residual block, as shown in Fig. 3. Because the skip connection is to superimpose the feature maps on the channels, we first use C 1×1 convolution kernels to change the number of channels. The feature map with size $W \times H \times C$ is

used to obtain a deeper feature map through two 3×3 convolutions. Dropout and batch normalization (BN) were used after each 3×3 convolution. In order to solve the problem of channel dependence, we first consider to output the signal of each channel in the feature map through the squeeze operation [12]. The squeeze operation is realized by global average pooling, and C 1×1 feature data are obtained.

$$z = F_{sq}(u) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u(i, j) \quad (1)$$

where z is the data after the channel squeeze, F_{sq} is the squeeze function, u is the input of the squeeze and excitation block, W is the width of the feature map, and H is the height of the feature map [30].

The excitation operation aims to completely capture the dependence of channels through two fully connected (FC) layers. In order to limit the complexity of the model, the first FC layer reduces the dimension of the channel in the ratio of $r = 6$. After the activation function is applied, the second FC layer restores the channel to the original dimension.

$$s = F_{ex}(z, W) = \sigma(W_2 \delta(W_1 z)) \quad (2)$$

where s is the data after the channel excitation, F_{ex} is the excitation function, δ is the ReLU function, σ is the sigmoid function, $W_1 \in R^{\frac{C}{r} \times C}$ and $W_2 \in R^{C \times \frac{C}{r}}$.

The squeeze and excitation block readjust the weights of each channel in the feature map, enhance the weights of task-related pixels, and suppress the weights of background and noises. The introduction of the residual block is to prevent the problem that the model cannot be converged when the depth of the network is deepened [31]. The SER module can significantly increase the accuracy of the segmentation while slightly increasing the model complexity and computing time, and it is easier to integrate than other attention mechanisms [32]. Finally, the function of the SER module is represented by the following equation.

$$y = x + u \times s \quad (3)$$

where x is the input of SER module, y is the output of SER module.

Fig. 3 SER module. Our proposed SER module mainly consists of convolution, squeeze and excitation block and residual block. SER module pays more attention to the more important feature channels and strengthens the transmission ability of features. W , H and C represent the width, height and channel of the feature map, respectively.

3.3 ASPP

The combination of convolutional layer and pooling layer can expand the receptive field of the image without adding additional parameters, but pooling operation will compress the size of the image, which loses the details of the image. Nevertheless, upsampling cannot recover all the spatial information. Atrous convolution covers the image with a larger receptive field, and only some pixels in the receptive field are used for convolution, which is equivalent to inserting holes in the convolution kernel [33]. Therefore, atrous convolution can expand the receptive field and control the resolution of feature maps without losing details [34].

ASPP integrates atrous convolutions into the spatial pyramid pooling structure, which can obtain context information by atrous convolution using various sampling rates [35]. However, with the increase of the sampling rate, the weight of the filter will become smaller. When the size of the atrous convolution kernel is close to the feature map, the filter will degenerate from 3×3 to 1×1 , which only focuses on the local features and ignores the global information [36]. To address the challenge and incorporate global context information into the model, our method adopts image-level features, as shown in Fig. 4.

Fig. 4 ASPP module. Our proposed ASPP module is used to deal with deep features after downsampling, so the size of the feature map is $16 \times 16 \times 512$, and we use four 128 atrous convolutions with sampling rate of [1, 2, 4, 6], respectively. We concatenate the channels, and send the feature map into a 1×1 convolution, batch normalization and LeakyReLU.

3.4 Loss Function

Dice loss is widely used to guide the training of image segmentation models, but the segmentation performance is poor when facing small regions. In the case of an extreme imbalance between the number of pixels in the lesion and the number of pixels in the background, once a small number of pixels are predicted wrong, it will lead to dramatic changes in dice loss, resulting in unstable training.

Abraham et al. [37] proposed focal tversky loss (FTL) to deal with class imbalance:

$$FTL = \left(1 - \frac{\sum_{i=1}^N p_{ic}g_{ic}}{\sum_{i=1}^N p_{ic}g_{ic} + \alpha \sum_{i=1}^N p_{i\bar{c}}g_{ic} + \beta \sum_{i=1}^N p_{ic}g_{i\bar{c}}}\right)^{\frac{1}{\gamma}}, \quad (4)$$

where p_{ic} and $p_{i\bar{c}}$ are probabilities that pixel i is of the lesion class c and non-lesion class $\bar{c}$, respectively. g_{ic} and $g_{i\bar{c}}$ are probabilities for another class. The N is the total number of pixels in a slice. In our work, $\alpha = 0.7$, $\beta = 0.3$, $\gamma = \frac{4}{3}$.

Sudre et al. [38] proposed GDL that uses weights inversely proportional to lesion, in order to better segment small regions:

$$GDL = 1 - 2 \frac{\sum_{l=1}^2 w_l \sum_n r_{ln} p_{ln}}{\sum_{l=1}^2 w_l \sum_n r_{ln} + p_{ln}}, \quad (5)$$

$$w_l = \frac{1}{(\sum_{n=1}^N r_{ln})^2 + \varepsilon}, \quad (6)$$

where $r_{ln} \in \{0, 1\}$ and $p_{ln} \in [0, 1]$ represent the true voxel values and the probability predicted as class l , respectively. N is the total number of pixels in a slice. ε is set at 10^{-5} to prevent the loss function from the issue of dividing by 0. The contribution of each lesion is corrected by the inverse of its volume, thus reducing the correlation between lesion size and dice loss.

4 Experimental Results

In this section, we introduce the datasets, evaluation metrics and training details in the experiments. Then we conduct extensive comparative experiments for different models and loss functions and finally analyze the experimental results.

4.1 Datasets

The CT slices used in the experiment are derived from four datasets: COVID-19-CT-Seg dataset (China) [39], MosMed dataset (Russia) [40], UESTC-COVID-19 dataset (China) [26], and MedSeg dataset (Italy) [11]. The COVID-19-CT-Seg dataset contains 20 complete COVID-19 CT scans from Wenzhou Medical University. The lung parenchyma and lesions were labeled by three experienced radiologists. The MosMed dataset is provided and annotated by Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of the Moscow Health Care Department. The dataset contains CT scans of 1110 patients, and is divided into 5 categories according to the degree of infection, but only 50 cases in the category of several GGO and involvement of lung

parenchyma less than 25% are labeled. The UESTC-COVID-19 dataset contains CT scans of 120 patients, some of which contain noise. The slices have been cropped based on the lung boundary area and the grayscale has been normalized. The MedSeg dataset is published by the Italian Society of Medical and Interventional Radiology. Radiologists labeled slices from 60 patients with GGO, lung consolidation and pleural effusion.

As shown in Fig. 5, we extracted 11,359 slices containing the lesion from a total of 250 CT scans of COVID-19 patients with varying degrees of infection, where 8,000 were used for training models, 2,000 for validation, and 1,359 for testing. These slices have been resized to 256×256 and the lung window has been adjusted to improve the display of detailed lung tissues, that is, window width of 1200 HU with a window level of -500 HU.

Fig. 5 Number distribution of CT slices. COVID-19-CT-Seg dataset and UESTC-COVID-19 dataset contain complete CT scans. MosMed dataset and MedSeg dataset contain only a few slices for each patient.

4.2 Evaluation Metrics

We evaluated the performance of COVID-19 lesion segmentation models with the following metrics: dice similarity coefficient (DSC), intersection over union (IoU), sensitivity (Sen), specificity (Spe), F1-score (F1) and hausdorff distance (HD). P value is used to evaluate the significant difference between other methods and ours.

DSC and IoU are designed to evaluate the overlap rate of GT and a segmentation result (SR). GT is the gold standard for segmentation given by experienced doctors for labelling protocols. SR is the lesion predicted by a segmentation model.

$$DSC = \frac{2|GT \cap SR|}{|GT| + |SR|} \quad (7)$$

$$IoU = \frac{|GT \cap SR|}{|GT \cup SR|} = \frac{DSC}{2 - DSC} \quad (8)$$

Sen indicates the proportion of all pixels within the lesion that are properly classified, which measures the ability of the model to distinguish positive examples. On the contrary, Spe measures the ability of the model to distinguish negative examples. F1 is the harmonic average of accuracy and recall. In order to define these metrics, we also use the values of true positive (TP), true negative (TN), false positive (FP) and false negative (FN).

$$Sen = \frac{TP}{TP + FN} \quad (9)$$

$$Spe = \frac{TN}{TN + FP} \quad (10)$$

$$F1 = \frac{2 \times |TP|}{2 \times |TP| + |FN| + |FP|} \quad (11)$$

HD is defined as the maximum surface distance between the objects. x and y are the set of boundary pixels of GT and SR. The $d(x, y)$ is the minimum of the Euclidean distances between a voxel x and voxels in a set of y . The better predicted mask has a smaller HD value.

$$d(x, y) = \max_{x \in GT} \min_{y \in SR} \|x - y\| \quad (12)$$

$$HD = \max\{d(x, y), d(y, x)\} \quad (13)$$

4.3 Implementation Details

The proposed method and the methods for comparison are implemented in Keras with a single Nvidia GeForce RTX 2080 SUPER 8GB. We train models by Adam optimizer and GDL. The learning rate = 10^{-5} , batchsize = 6, epochs = 200. If the validation loss does not decrease in 10 consecutive epochs, the model will automatically stop training to prevent overfitting. The validation set and the test set are divided randomly.

4.4 Comparison of Models

We compared our proposed SAUNet++ with five state-of-the-art medical image segmentation models by using the same datasets and GDL. U-Net [7] is the most classic biomedical image segmentation model. U-Net++ [8] combines four encoder-decoder structures of different depths and adds a series of dense skip connections. RA-Unet [41] adds the attention residual mechanism. AttSEResUnet adds residual block and both the spatial-level and channel-level attention gate. COPLE-Net [26] mainly adds squeeze and excitation module and bridge layers on the basis of U-Net. The performance of the six models on the test set is shown in Table 1. The SAUNet++ achieved good performance on DSC ($87.38 \pm 11.23\%$), IoU ($79.04 \pm 14.81\%$), Sen ($93.28 \pm 10.09\%$), Spe ($99.55 \pm 0.70\%$), F1 ($87.54 \pm 11.22\%$) and HD ($19.99 \pm 35.06\text{mm}$). It is observed that our model obtains 1.3%, 1.65%, 3.43%, 4.35%, 7.7mm improvements on DSC, IoU, Sen, F1, HD, respectively due to adding SER modules and an ASPP module. In addition, we also used the t-test method to calculate the p-value to evaluate the statistical significance of our method. The results prove that SAUNet++ has a statistically significant improvement in DSC (all p-values < 0.01). The visual comparison of the output of different models is shown in Fig. 6. U-Net can only get rough segmentation masks, resulting in missed segmentation when the lesion is small, and over-segmentation when the lesion is more complex. With the change of the model structure and the introduction of SER and ASPP modules, the improved models can obtain more accurate boundaries. The output masks obtained by or proposed SAUNet++ model are closest to the gold standard.

Table 1 Model Comparison for COVID-19 Lesion Segmentation. The GDL was Used for Training.

Models	DSC (%)	IoU (%)	Sen (%)	Spe (%)	F1 (%)	HD (mm)	p-value (DSC)
U-Net [7]	83.12 $\pm$ 14.71	73.26 $\pm$ 17.39	83.51 $\pm$ 15.95	99.62$\pm$0.68	83.19 $\pm$ 14.54	31.90 $\pm$ 45.33	3.54×10^{-17}
U-Net++ [8]	86.08 $\pm$ 12.99	77.39 $\pm$ 16.30	89.85 $\pm$ 13.30	99.56 $\pm$ 0.76	86.08 $\pm$ 12.99	27.69 $\pm$ 41.47	5.44×10^{-3}
RA-Unet [41]	85.17 $\pm$ 14.65	76.37 $\pm$ 17.54	91.10 $\pm$ 12.55	99.53 $\pm$ 0.75	85.17 $\pm$ 14.65	28.58 $\pm$ 41.58	1.13×10^{-5}
AttSEResUnet	85.31 $\pm$ 13.45	76.31 $\pm$ 16.69	93.03 $\pm$ 11.70	99.49 $\pm$ 0.73	85.31 $\pm$ 13.45	22.06 $\pm$ 35.71	1.44×10^{-5}
COPLE-Net [26]	84.83 $\pm$ 14.73	75.86 $\pm$ 17.59	92.23 $\pm$ 13.20	99.51 $\pm$ 0.69	84.90 $\pm$ 14.55	20.89 $\pm$ 35.01	4.33×10^{-7}
SAUNet++	87.38$\pm$11.23	79.04$\pm$14.81	93.28$\pm$10.09	99.55 $\pm$ 0.70	87.54$\pm$11.22	19.99$\pm$35.06	-

Table 2 Loss Function Comparison for COVID-19 Lesion Segmentation. The SAUNet++ was Used for Training.

Models	DSC (%)	IoU (%)	Sen (%)	Spe (%)	F1 (%)	HD (mm)	p-value (DSC)
Dice loss	86.30 $\pm$ 11.98	77.51 $\pm$ 15.49	93.04 $\pm$ 10.66	99.51 $\pm$ 0.73	86.47 $\pm$ 11.98	21.45 $\pm$ 34.80	1.56×10^{-2}
FTL [37]	86.60 $\pm$ 11.71	77.92 $\pm$ 15.25	88.25 $\pm$ 12.03	99.69$\pm$0.52	87.24 $\pm$ 11.42	20.94 $\pm$ 34.97	7.67×10^{-2}
GDL	87.38$\pm$11.23	79.04$\pm$14.81	93.28$\pm$10.09	99.55 $\pm$ 0.70	87.54$\pm$11.23	19.99$\pm$35.06	-

Fig. 6 The visual comparison of the output of different models. The first line is the original CT slices in the test set. The corresponding segmentation output from the U-Net, U-Net++, RA-Unet, AttSEResUnet, COPL-Net, SAUNet++ are shown on the second, third, fourth, fifth, sixth, seventh line respectively. The GT mask corresponding to each slice is shown on the last line.

4.5 Comparison of Loss Functions

In order to better illustrate the ability of GDL to segment small regions, we use dice loss, FTL and GDL to train the SAUNet++ model. The performance of the three methods on the test set is shown in Table 2. Compared with dice loss, FTL has better performance on DSC and HD, but the ability to find positive pixels is decreased. The

GDL obtains 0.78%, 1.12%, 0.24%, 0.3%, 0.95mm improvements on DSC, IoU, Sen, F1, HD, respectively, which is the best results among the three methods in each evaluation metrics. The results prove that GDL has a statistically significant improvement in DSC (all p-values < 0.1). The visual comparison of the outputs of SAUNet++ using different loss functions is shown in Fig. 7. We can see that GDL can effectively guide the model to segment the micro-infected regions which cannot be identified by other methods. In addition, the segmentation error rate is low when facing the non-infected regions which are highly similar to lesions.

Fig. 7 The visual comparison of the output of SAUNet++ using different loss function. The first line is the original CT slices in the test set, and we selected slices with less infection to better observe the effect of loss function on small region segmentation. The corresponding segmentation output from dice loss, FTL and GDL are shown on the second, third and fourth line respectively. The GT mask corresponding to each slice is shown on the last line.

5 Discussion

DSC is the most popular evaluation metric for image segmentation because it can express the degree of overlap at the pixel level. As shown in Table 1, when using a multi-layer encoder-decoder structure such as U-Net++ to segment the lesion, DSC is improved more obviously compared with U-Net baseline. Therefore, it is proved that, compared with the small changes of attention mechanism and residual connection based on U-Net, it is more effective to improve the backbone network structure directly. In view of the excellent performance of U-Net++, we added SER and ASPP modules. Although the embedded modules of squeezing and excitation and atrous convolution are common, but there is no relevant research to introduce them into U-Net++. Comparing SAUNet++ with U-Net++, the performance of the model is obviously improved because the SER module adjusts the channel weights and the ASPP module increases the receptive field. In particular, Sen increased by 3.43%, which is very important for doctors to find the lesions as much as possible when treating the diseases. And SAUNet++

is better than state-of-the-art medical image segmentation models in most of the evaluation metrics. By analyzing the segmentation results, the excellent performance of SAUNet++ is mainly attributed to the improvement on the segmentation of: (1) GGO with high transparency; (2) lesions located near blood vessels and trachea; (3) lesions located on the chest wall.

The reason for directly using GDL as the loss function is that the real goal of segmentation is to maximize the DSC metric. When a slice contains multiple lesions, a dice index is obtained for each region, while the generalized dice index integrates the dice indices of multiple regions and uses one index to quantify the segmentation results [42]. It makes the GDL training model more stable and easier to identify small infected regions. As shown in Table 2, when using the GDL to train SAUNet++ model, DSC is 1.08% higher than the dice loss baseline. Generally speaking, the improvement of the model is more effective than the improvement of the loss function. Even the SAUNet++ model using the simplest loss function can reach more than 86% DSC, while most models trained with GDL have less than 86% DSC.

There are slight differences among COVID-19 patients infected in different countries, and multinational datasets make the model not limited to local use. And there are more pictures in our dataset than in most studies of COVID-19 lesion segmentation. Because our data come from different datasets and doctors have different standards, the labeling of some regions is controversial, which is inevitable.

Some studies suggest that the lung parenchyma is segmented to reduce the search region before segmenting the lesion. However, as shown in Fig. 5 and Fig. 6, even if the lung parenchyma is not segmented in advance, a well-trained model will hardly judge the tissues outside the lung as COVID-19 lesion. If a lung parenchyma segmentation model needs to be added, it will make the operation more difficult for doctors.

However, there are some problems need to be solved in future study. Due to the lack of multi-category labeling of lesions, the output of our model is a single category segmentation image. The proposed SAUNet++ is a supervised learning method, so each image needs a corresponding mask. However, more unlabeled images are not used effectively. A semi-supervised learning method will have better prospects for segmentation of COVID-19 lesion.

6 Conclusion

In this study, we proposed a novel SAUNet++ model to segment the lesion in COVID-19 CT images. The SAUNet++ adds the channel attention mechanism of SER module and the ASPP module with different sampling rates on the basis of U-Net++. In addition, we used the improved loss function to deal with the problem of small region segmentation. In order to prove the performance of our method, we use the datasets from China, Italy and Russia to train and test different models. The experimental results showed that GDL is more effective than FTL in small region segmentation. The SAUNet++ achieved good performance on DSC ($87.38 \pm 11.23\%$), IoU ($79.04 \pm 14.81\%$), Sen ($93.28 \pm 10.09\%$), Spe ($99.55 \pm 0.70\%$), F1 ($87.54 \pm 11.22\%$) and HD ($19.99 \pm 35.06\text{mm}$), which surpassed the performance of state-of-the-art medical image segmentation models. Our model can also be used to segment other lung diseases, such as common pneumonia. In the future work, we will divide the lesion into different classes to evaluate the proposed method in multi-class segmentation.

Acknowledgements This work was supported in part by the National Natural Science Foundation of China (Grant No. 61971078), Science and Technology Foundation of Chongqing Education Commission (Grant No. CQUT20181124), Graduate Student Innovation Program of Chongqing (Grant No. CYS20351), and Undergraduate Student Innovation and Entrepreneurship Training Program of Chongqing (Grant No. S202111660048).

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. World Health Organization.: Weekly epidemiological update on COVID-19 - 10 August 2021. (2021)
2. Tavaré, A.N., Braddy, A., Brill, S., Jarvis, H., Sivaramakrishnan, A., Barnett, J., Creer, D.D., Hare, S.S.: Managing high clinical suspicion COVID-19 inpatients with negative RT-PCR: a pragmatic and limited role for thoracic CT. *Thorax*. 75(7), 537-538(2020)
3. Fang, Y., Zhang, H., Xie, J., Lin, M., Ying, L., Pang P., Ji, W.: Sensitivity of chest CT for COVID-19: comparison to RT-PCR. *Radiology*. 296(2), E115-E117(2020)
4. Zheng, B., Liu, Y., Zhu, Y., Yu, F., Jiang, T., Yang, D., Xu, T.: MSD-Net: Multi-Scale Discriminative Network for COVID-19 Lung Infection Segmentation on CT. *IEEE Access*. 8, 185786-185795(2020)
5. Kaushal, C., Singla, A.: Automated segmentation technique with self-driven post-processing for histopathological breast cancer images. *CAAI Transactions on Intelligence Technology*. 5(4), 294-300(2020)

6. Alhasson, H.F., Willcocks, C.G., Alharbi, S.S., Kasim, A., Obara, B.: The relationship between curvilinear structure enhancement and ridge detection methods. *Visual Computer*. 3, 1-21(2020)
7. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: The 18th International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI), Munich. 234-241(2015)
8. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, Granada. 3-11(2018)
9. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999(2018)
10. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Conference on Medical image computing and computer-assisted intervention, Athens. 424-432(2016)
11. Fan, D., Zhou, T., Ji, G., Zhou, Y., Chen, G., Fu, H., Shen, J., Shao, L.: Inf-net: Automatic covid-19 lung infection segmentation from ct images. *IEEE Transactions on Medical Imaging*. 39(8), 2626-2637(2020)
12. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, Salt Lake City. 7132-7141(2018)
13. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587(2017)
14. Xiao, H., Ran, Z., Huang, J., Ren, H., Liu, C., Zhang, B., Zhang, B., Dang, J.: Research progress in lung parenchyma segmentation based on computed tomography. *Journal of Biomedical Engineering*. 38(2), 379-386(2021)
15. Shi, L., Ma, H., Zhang, J.: Automatic detection of pulmonary nodules in CT images based on 3D Res-I network. *Visual Computer*. 37, 1343-1356(2021)
16. Ziyad, S.R., Radha, V., Vayyapuri, T.: Overview of computer aided detection and computer aided diagnosis systems for lung nodule detection in computed tomography. *Current Medical Imaging*. 16(1), 16-26(2020)
17. Kumar, S.P., Latte, M.V.: Lung parenchyma segmentation: fully automated and accurate approach for thoracic CT scan images. *IETE Journal of Research*. 66(3), 370-383(2020)
18. Nithila, E.E., Kumar, S.: Segmentation of lung from CT using various active contour models. *Biomedical Signal Processing Control*. 47, 57-62(2019)
19. Geng, L., Zhang, S., Tong, J., Xiao, Z.: Lung segmentation method with dilated convolution based on VGG-16 network. *Computer Assisted Surgery*. 24(sup2), 27-33(2019)
20. Khanna, A., Londhe, N.D., Gupta, S., Semwal, A.: A deep Residual U-Net convolutional neural network for automated lung segmentation in computed tomography images. *Biocybernetics Biomedical Engineering*. 40(3), 1314-1327(2020)
21. Zhang, Z., Wu, C., Coleman, S., Kerr, D.: DENSE-INception U-net for medical image segmentation. *Computer methods programs in biomedicine*. 192, 105395(2020)
22. Tan, J., Jing, L., Huo, Y., Li, L., Akin, O., Tian, Y.: Lgan: Lung segmentation in ct scans using generative adversarial network. *Computerized Medical Imaging Graphics*. 87, 101817(2021)
23. Zhao, X., Zhang, P., Song, F., Fan, G., Sun, Y., Wang, Y., Tian, Z., Zhang, L., Zhang, G.: D2A U-Net: Automatic Segmentation of COVID-19 Lesions from CT Slices with Dilated Convolution and Dual Attention Mechanism. arXiv preprint arXiv:2102.05210(2021)
24. Joseph Raj, A.N., Zhu, H., Khan, A., Zhuang, Z., Yang, Z., Mahesh, V.G.V., Karthik, G.: ADID-UNET—a segmentation model for COVID-19 infection from lung CT scans. *PeerJ Computer Science*. 7(2021)
25. Amyar, A., Modzelewski, R., Li, H., Ruan, S.: Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation. *Computers in Biology Medicine*. 126, 104037(2020)
26. Wang, G., Liu, X., Li, C., Xu, Z., Ruan, J., Zhu, H., Meng, T., Li, K., Huang, N., Zhang, S.: A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images. *IEEE Transactions on Medical Imaging*. 39(8), 2653-2663(2020)
27. Zhou, T., Canu, S., Ruan, S.: Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism. *International Journal of Imaging Systems and Technology*. 31(1), 16-27(2021)
28. Zhang, P., Zhong, Y., Deng, Y., Tang, X., Li, X.: CoSinGAN: Learning COVID-19 Infection Segmentation from a Single Radiological Image. *Diagnostics (Basel)*. 10(11), 901(2020)
29. Abdel-Basset, M., Chang, V., Hawash, H., Chakraborty, R.K., Ryan, M.: FSS-2019-nCov: A deep learning architecture for semi-supervised few-shot segmentation of COVID-19 infection. *Knowledge-Based Systems*. 212, 106647(2021)
30. Gu, J., Sun, X., Zhang, Y., Fu, K., Wang, L.: Deep residual squeeze and excitation network for remote sensing image super-resolution. *Remote Sensing*. 11(15), 1817(2019)
31. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, LAS VEGAS. 770-778(2016)
32. Roy, A.G., Siddiqui, S., Pölsterl, S., Navab, N., Wachinger, C.: 'Squeeze & excite' guided few-shot segmentation of volumetric images. *Medical image analysis*. 59, 101587(2020)
33. Wang, P., Chen, P., Yuan, Y., Liu, D., Huang, Z., Hou, X., Cottrell, G.: Understanding convolution for semantic segmentation. In: 2018 IEEE winter conference on applications of computer vision (WACV), Nevada. 1451-1460(2018)
34. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122(2015)
35. Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*. 40(4), 834-848(2017)
36. Ma, J., Dai, Y., Tan, Y.-P.: Atrous convolutions spatial pyramid network for crowd counting and density estimation. *Neuro-computing*. 350, 91-101(2019)
37. Abraham, N., Khan, N.M.: A novel focal tversky loss function with improved attention u-net for lesion segmentation. In: 2019 IEEE 16th International Symposium on Biomedical Imaging, Venice. 683-687(2019)
38. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep learning in medical image analysis and multimodal learning for clinical decision support(DLMIA), Québec City. 240-248(2017)
39. Ma, J., Wang, Y., An, X., Ge, C., Yu, Z., Chen, J., Zhu, Q., Dong, G., He, J., He, Z.: Toward data-efficient learning: A benchmark for COVID-19 CT lung and infection segmentation. *Medical physics*. 48(3), 1197-1210(2021)
40. Morozov, S., Andreychenko, A., Pavlov, N., Vladzimirsky, A., Ledikhova, N., Gombolevskiy, V., Blokhin, I.A., Gelezhe, P., Gonchar, A., Chernina, V.Y.: Mosmeddata: Chest ct scans with covid-19 related findings dataset. arXiv preprint arXiv:1804.03999(2020)
41. Jin, Q., Meng, Z., Sun, C., Cui, H., Su, R.: RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans. *Frontiers in Bioengineering and Biotechnology*. 8, 1471(2020)
42. Crum, W.R., Camara, O., Hill, D.L.: Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE transactions on medical imaging*. 25(11), 1451-1461(2006)