



Research article

UNet segmentation network of COVID-19 CT images with multi-scale attention

Mingju Chen^{1,2}, Sihang Yi^{1,2,*}, Mei Yang³, Zhiwen Yang^{1,2} and Xingyue Zhang^{1,2}

¹ School of Automation and Information Engineering, Sichuan University of Science & Engineering, Yibin 644002, China

² Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science & Engineering, Yibin 644002, China

³ Zigong Third People's Hospital, Zigong 643000, China

* **Correspondence:** Email: 321085404110@stu.suse.edu.cn.

Abstract: In recent years, the global outbreak of COVID-19 has posed an extremely serious life-safety risk to humans, and in order to maximize the diagnostic efficiency of physicians, it is extremely valuable to investigate the methods of lesion segmentation in images of COVID-19. Aiming at the problems of existing deep learning models, such as low segmentation accuracy, poor model generalization performance, large model parameters and difficult deployment, we propose an UNet segmentation network integrating multi-scale attention for COVID-19 CT images. Specifically, the UNet network model is utilized as the base network, and the structure of multi-scale convolutional attention is proposed in the encoder stage to enhance the network's ability to capture multi-scale information. Second, a local channel attention module is proposed to extract spatial information by modeling local relationships to generate channel domain weights, to supplement detailed information about the target region to reduce information redundancy and to enhance important information. Moreover, the network model encoder segment uses the Meta-ACON activation function to avoid the overfitting phenomenon of the model and to improve the model's representational ability. A large number of experimental results on publicly available mixed data sets show that compared with the current mainstream image segmentation algorithms, the pro-posed method can more effectively improve the accuracy and generalization performance of COVID-19 lesions segmentation and provide help for medical diagnosis and analysis.

Keywords: COVID-19; lesion segmentation; multi-scale attention; UNet; local channel attention;

1. Introduction

Since December 2019, the COVID-19 epidemic has caused a huge impact on the world economy, and also seriously affected people's normal clothing, food, housing and transportation [1,2]. As of March 23, 2022, data from the World Health Organization (WHO) shows that about 470 million people have been diagnosed with COVID-19 worldwide, and about 6.09 million people have died due to the temporary lack of effective treatment for COVID-19. Radiologic techniques, including chest X-ray and computed tomography (CT) scans, provide critical imaging tools for detecting pneumonia infections associated with COVID-19. X-ray radiography offers the advantages of portable equipment for on-site execution, reduced patient movement and lower radiation dose. However, chest CT is more effective than X-ray radiography in detecting COVID-19 due to its high spatial resolution. This is because CT can be used to detect small lesions in the lungs that show signs of infection with greater sensitivity than X-ray radiography [3]. Typical signs of pneumonia lesions, such as ground-glass shadows, can be observed on CT images. These characteristics provide effective information for the quantitative assessment of COVID-19 conditions and become an important means to combat COVID-19.

CT images of COVID-19 usually have relatively complex textures, edges and other image features. In actual diagnosis, image segmentation can be used to divide areas of concern to reduce interference from other areas. After the segmentation of the COVID-19 lesion area, CT quantitative analysis technology can be used to analyze the CT value of the lesion area and the lesion proportion to evaluate the COVID-19 disease. Studies have shown that the use of lung CT images for screening has higher sensitivity than RT-PCR technology and can be used as an auxiliary screening means to improve the reliability of screening [4]. On the other hand, the use of quantitative CT analysis is effective in assessing the degree of infection based on the performance of the patient's lung CT images, which can be used as an important tool for disease tracking and monitoring [5,6]. The use of CT imaging technology to diagnose people can separate doctors from patients, control the spread of the epidemic within a certain range and is safer than RT-PCR technology in the detection of COVID-19.

Traditional segmentation algorithms, such as segmentation methods based on region [7], edge detection [8] and thresholding [7,9], as well as wavelet transforms [10], have limited segmentation capabilities and are unable to effectively segment CT images of diseased lungs. Conventional Computer Aided Diagnosis (CAD) systems usually use traditional digital image processing techniques [11] in the analysis of infected areas of lung CT, or use machine learning techniques that require manual feature extraction [12], both of which are complex and have large errors.

In recent years, deep learning methods based on convolutional neural networks (CNNs) have achieved remarkable results in the field of medical image segmentation. Currently, the commonly used deep learning networks for image segmentation are FCN [13], SegNet [14] and U-Net [15]. U-Net and its improved U-Net (U-Net++ [16] and KUB-UNet [17], which performs excellent in X-ray images) have been widely used in medical image segmentation. However, when the above method is used to segment CT images of the novel coronavirus pneumonia, the gradient is easy to disappear and the features are not fully utilized, resulting in low segmentation performance.

For the existing deep learning model of COVID-19 CT image segmentation, there are still some problems, such as low segmentation accuracy, poor model generalization performance, large model

parameters and difficult deployment. This paper proposes a novel coronavirus CT image segmentation model integrating multi-scale attention and adaptive activation function of UNet, we will call it LMSA-Net. LMSA-Net focuses on lightweight network to facilitate model training and deployment. Proposing a UNet COVID-19 segmentation network based on improved multiscale convolutional attention for lung parenchyma as well as COVID-19 lesions, respectively. Experimental results show that this method has better performance and competitiveness than the existing classical methods.

In summary, the major contributions of the work in this paper are as follows:

- In this paper, an improved multi-scale convolutional attention (MSCA) encoder module is proposed to be applied to the encoder part of the U-shaped structure on the network structure of UNet. By replacing the channel global information coding with local information coding, and modeling only the K-nearest-neighbor element relationship of the channel domain with correlation, the redundancy information is effectively reduced and the segmentation performance is improved.
- An attention module using local channel (LCA) domain interaction is proposed, which replaces the channel attention computation in the original MSCA structure to make up for the lack of spatial context information modeling in the original MSCA structure and to enhance the localization ability of the model.
- The Meta-ACON function is used as the activation function of the network, which has an adaptive first-order derivative size and an adaptive activation mode. The network locally uses the adaptive activation function Meta-ACON to avoid overfitting and to improve the network representation, which leads to better convergence and generalization of the network model.

The rest of the paper is organized as follows. Section 2 describes the related work; Section 3 presents the improved methodology; Section 4 analyzes the experimental results; Section 5 discusses the limitations of the methodology of this paper; and Section 6 summarizes the whole paper and draws conclusions. In addition, we list some important acronyms from this paper in Table 1.

Table 1. Summary of acronyms in this article.

Full name of the noun	Abbreviations
World Health Organization	WHO
computed tomography	CT
Computer Aided Diagnosis	CAD
convolutional neural networks	CNNs
multi-scale convolutional attention	MSCA
Generative adversarial networks	GAN
Spatial Feature and Resolution Maximization	SFRM
chest X-ray	CXR
local channel attention	LCA
Intersection over Union	IoU
Mean Intersection over Union	MIoU
eXplainable Artificial Intelligence	XAI

2. Related works

CT image segmentation of COVID-19 pneumonia lesions currently provides quantitative features [18]

for mass screening [19] and quantitative analysis of lung infections [20] to accurately assess disease severity. However, there are fewer studies on improving the generalization performance of COVID-19 pneumonia lesion segmentation.

Recently, many scholars have proposed the novel coronavirus lesion segmentation algorithm based on traditional methods [7,9,20]. For example, Shen et al. [7] performed lesion segmentation based on thresholding and region growth. Oulefki et al. [9] proposed an image contrast enhancement algorithm and multilevel image thresholding method for pneumonia lesion segmentation.

Moreover, a large number of methods based on deep learning have been proposed. Cao et al. [21] and Huang et al. [22] utilized Unet to segment the region of pneumonia lesions for quantitative analysis. Shan et al. [23] applied VB-net for segmentation. Chaganti et al. [24] trained two networks to obtain lung regions and foci respectively, and used lung region filtering to obtain foci for automatic segmentation and quantization. Yan et al. [25] proposed a CNN based segmentation network for lesion segmentation. Fan [26] and others proposed Inf-net, which uses a parallel encoder and multi-attention mechanism and a semi-supervised learning approach to cope with the small number of samples in the dataset. Jiang et al. [27] proposed to use Generative adversarial networks (GAN) to generate data for training to supplement the problem of insufficient data. Kumar et al. [28] proposed the RFA module by combining convolutional layers with discrete wavelet transform, which is a model that ensures the completeness of the information on the basis of further improvement of the sensory field thereby improving the segmentation accuracy of the COVID-19. He et al. [29] proposed a label optimization-based segmentation network model for COVID-19. He et al. [30] developed a novel evolvable adversarial framework for COVID-19 infection segmentation. Mu et al. [3] proposed a progressive global sensing and local polishing network for automatic segmentation of pneumonia infections caused by COVID-19 in CT images. Song et al. [31] proposed a novel self-supervised deep learning method for automatic segmentation of COVID-19 infected lesions and assessment of infection severity. On the other hand, X-ray images have achieved excellent results in the field of medical image segmentation techniques. Ran et al. designed an automated, accurate, reliable, robust and intelligent adjunct system for mass screening of COVID-19, non-COVID-19 viral pneumonia and bacterial pneumonia from healthy chest radiographs. In the same year, Ran's team developed the Spatial Feature and Resolution Maximization (SFRM) GAN, which effectively reduces the visibility of bones in a chest X-ray (CXR) while ensuring maximum retention of critical information. It is worth mentioning that the images examined above are usually subject to uncertainty in their actual implementation, which affects the quality of performance. Literature [32,33] use fuzzy logic-based tools for preprocessing the studied images, which provides ideas in preprocessing the segmented images.

UNet is a prominent model in the field of medical image segmentation and its network structure is depicted in Figure 1. It utilizes an encoder-decoder architecture for feature extraction and high-resolution image reconstruction. The encoder part employs four maximum pooling operations for downsampling and two 3×3 convolution operations for feature extraction before each downsampling step. The decoder section restores the image resolution through 4 upsampling operations, using either transposed convolution or linear interpolation. Following each upsampling, two 3×3 convolution operations are applied to adjust the image dimensions, and the feature maps from both the encoder and decoder are concatenated along the channel dimension to minimize information loss. Finally, the output is passed through a 1×1 convolutional layer.

The above methods have achieved some performance in COVID-19 lesion segmentation, but the evaluation of the models is more difficult due to the large amount of work involved in the production

of the dataset and the lack of uniform standards. In addition, some of the models are designed without considering lightness, and the long training time and large number of parameters are not conducive to local deployment. It is easy to overfitting when the data set has fewer samples. Aiming at the above problems, we improve the network on the basis of the structure of UNet and propose a fused multiscale attention COVID-19 CT image segmentation model based on UNet, which can effectively solve and alleviate the above problems.

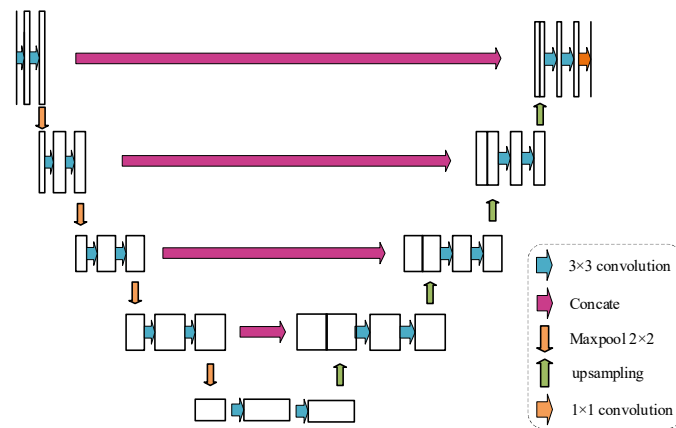


Figure 1. UNet network structure.

3. Approach

This section describes the proposed LMSA-Net. First, we outline the architecture of the whole network and then describe the structure of each module in detail.

3.1. Network architecture

Figure 2 illustrates the overall model design of LMSA-Net, and we used the UNet network with multi-scale feature fusion to effectively segment complex CT images of COVID-19. Specifically, in this paper, a multi-scale convolutional encoder (MSCA) module is designed in the encoder part of the U-shaped structure to replace the 3×3 convolutional feature extraction of UNet, which consists of three sets of depth grouped convolutions with different sizes of strips and a local channel attention (LCA) mechanism to improve the network's ability of capturing multi-scale information, as well as feature extraction. LCA improves the channel global information coding to local information coding, modeling only the channel domain K nearest neighbor element relationships with correlation, reducing redundant information to improve segmentation performance.

We employed the Meta-ACON function as the activation function of the network. This function adapts its first-order partial derivative size and activation modes, including linear and nonlinear modes. This effectively mitigates network overfitting, leading to better convergence and generalization of the network. To avoid introducing additional parameters that could slow down network training, the Meta-ACON function is only used as the activation function in the encoding section to enhance the network's representation capacity. The decoding section still utilizes the ReLU activation function. Furthermore, to address information loss caused by downsampling in UNet using max-pooling, we employed a 2D

convolution with a kernel size of 3 and a stride of 2 for downsampling the feature map. This optimization aims to improve the segmentation performance of the network. Finally, the output of LMSA-Net is obtained through a 1×1 convolution.

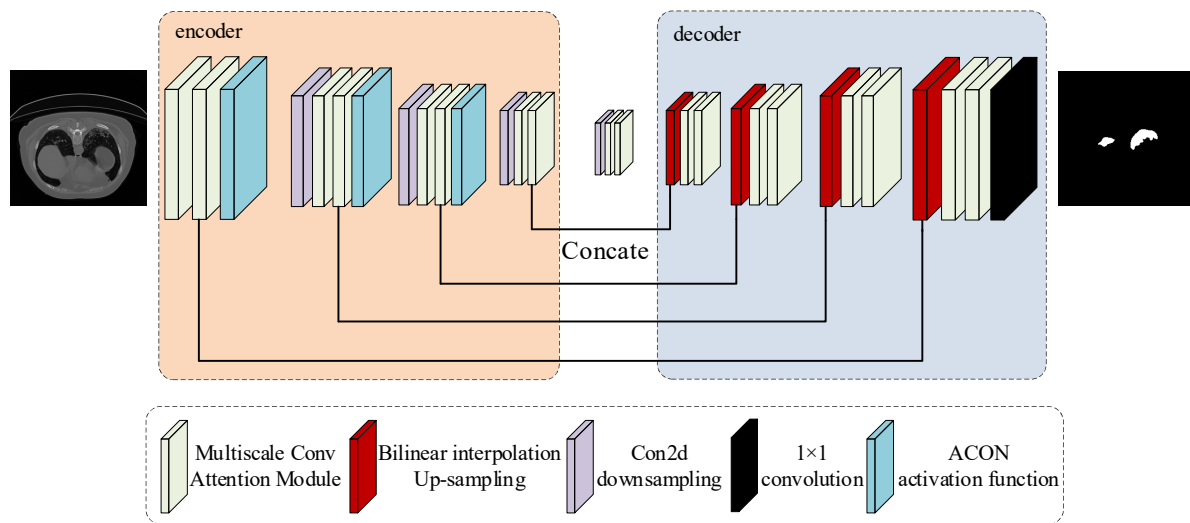


Figure 2. Overview of the proposed LMSA-Net framework. Input is lung image and output is segmented image.

3.2. Architecture of the multiscale convolutional attention module

Using a single-sized 3×3 convolution for each layer of feature extraction for UNet, the network as a whole has limited ability to capture multi-scale contextual information. In this paper, the multi-scale convolutional attention (MSCA) module is proposed to re-place the 3×3 convolution of UNet, so as to obtain a more comprehensive receptor field when extracting context information. For the task of COVID-19 lesion region segmentation, the network fusion of multi-scale information can have the lesion region feature information of different sizes and locations, thus improving the network's modeling ability of global features.

The MSCA module is used as an encoder to solve the problem of insufficient modeling of multi-scale information of the input image during the training process of the UNet network. MSCA is a parallel multiscale convolutional attention structure capable of fusing multiscale information and reconstructing feature maps using channel domain attention. As shown in Figure 3, the MSCA structure consists of two branching paths, which are multi-scale information extraction and fusion and attention weight reconstruction. Specifically, the feature map F obtained from downsampled feature extraction is first subjected to 5×5 deep convolution, and then multi-scale contextual information is extracted by three sets of bar deep convolution of different sizes and feature fusion is performed in the form of element summation. Each group uses 1×7 and 7×1 , 1×11 and 11×1 and 1×21 and 21×1 to replace the depth grouping convolution of 7×7 , 11×11 and 21×21 sizes, respectively. In addition, we model the channel dimensions using 1×1 convolution to generate an attention weight map to reconstruct the input image. Finally, the fused features are fed into the decoder section for scaling, which is restored to the original image size by bilinear interpolation upsampling and skip connections. In multi-scale information extraction and fusion, deep convolution and deep strip convolution are used

to reduce the number of parameters and speed up training. This process is described as:

$$\begin{aligned} Att &= Conv_{1 \times 1} \left(\sum_{i=0}^3 Scale_i(DW - Conv(F)) \right) \\ F' &= Att \otimes F \end{aligned} \quad (1)$$

Where, F is the input feature map, F' is the output feature map, $Scale_i$ denotes the modeling of multi-scale information and DW denotes Depthwise Convolution.

As can be seen from Eq (1), the MSCA structure has fewer parameters as well as excellent multi-scale feature capture capabilities and channel domain attention mechanisms. When segmenting the COVID-19 lesion region, the network can effectively spatially localize the infected region and give more attention to the complex texture and edges of the infected region to increase the accuracy of the model segmentation. In addition, the MSCA structure uses a large number of strip convolutions, which is more favorable for the extraction of slender types of lesions. However, MSCA uses 1×1 convolution for modeling within the channel range while ignoring the spatial dimension when performing attentional weight generation.

To address the above problems, we propose a local channel attention (LCA) module to generate feature weights based on the MSCA structure. LCA synthesizes information from both the channel domain as well as the spatial domain to generate attention weights, which can be more effective in giving more accurate attention to the complex morphology of the lesion region, and the new weight generation module is described as:

$$MSCA = LCA \left(\sum_{i=0}^3 Scale_i(DW - Conv(F)) \right) \quad (2)$$

where LCA is the designed local channel attention module, $Scale_i$ denotes the modeling of multi-scale information and DW denotes Depthwise Convolution.

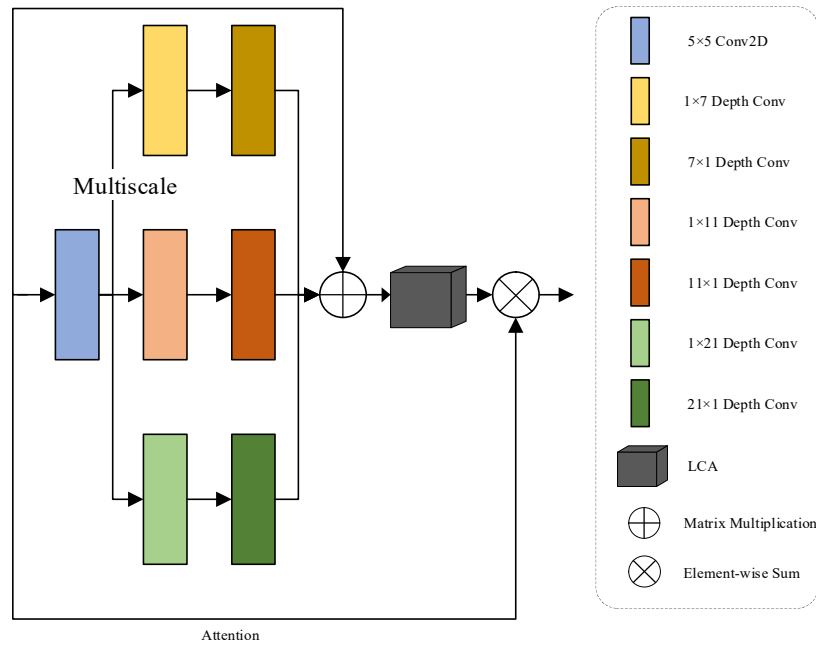


Figure 3. Multi scale convolutional attention module.

3.3. Architecture of the local channel attention module

The LCA module is a network component for attention mechanisms whose main purpose is to enhance feature representation in the channel dimension. Prior to this, SE-Net [34] enabled the network to adaptively learn the correlation and importance between features by introducing a local attention module. However, SE-Net increases the number of parameters and computational complexity of the network. ECA-Net [35] by introducing an efficient local attention module, the network is able to adaptively adjust the weights between channels to better capture important features and suppress irrelevant ones. However, ECA-Net ignores the spatial dimensionality information and is poorly adapted to different scales. Inspired by SE-Net and ECA-Net, in order to enable the network to adaptively learn correlations between channels while fusing spatial dimensional features, we design a lightweight local channel attention (LCA) module. LCA can synthesize spatial as well as channel information to generate attentional weights, which are adjusted by spatial location as well as channel layer when reconstructing the input image. The LCA module has a smaller number of covariates and a larger improvement in segmentation performance, allowing the network to give more precise attention to the COVID-19 foci region and perform more biased computations. We integrate the LCA module into the structure of MSCA. Specifically, for input $x \in \mathbb{R}^{H \times W \times C}$, two 1D pooling kernels are used to perform an average pooling operation along the horizontal and vertical directions, respectively. For channel C at height h the output is [34]:

$$z_h^c(h) = \frac{1}{W} \sum_{0 \leq j \leq W} x_c(j, w) \quad (3)$$

Where W is the width of the input channel. Similarly for channel C the output at width W is:

$$z_w^c(h) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w) \quad (4)$$

Where H is the height of the input channel. After the feature information is embedded, the outputs z_w^c and z_h^c are spliced along the spatial dimension and activated using the ReLU activation function after a linear transformation. The process can be represented as:

$$f = \delta\left(T_1 \left[z^h, z^w \right]_{\text{concat}}\right) \quad (5)$$

where $[\dots]_{\text{concat}}$ is the spatial dimension of the splicing operation and T_1 is a linear transformation of convolution size 1×1 . The output $f \in \mathbb{R}^{C/r \times l \times (H+W)}$ is the inter-mediate feature map generated by information embedding. The attention weight generation part decomposes f into two feature vectors $f^h \in \mathbb{R}^{C/r \times h \times l}$ and $f^w \in \mathbb{R}^{C/r \times l \times w}$. Each feature vector after decomposition is then linearly transformed by one-dimensional convolution, and finally the attention weight vectors g^w and g^h are generated by the sigmoid function, and the process can be expressed as [35]:

$$\begin{aligned} g^w &= \sigma(T_w f^w) \\ g^h &= \sigma(T_h f^h) \end{aligned} \quad (6)$$

From Eqs (5) and (6), LCA is modeled using a linear transformation when calculating the channel dimensional attention weights using the channel dimensions to capture the global information. Due to the difference in the dimension of the weight map vectors computed in LCA, we use 2D convolution to capture the channel local information and then use the sigmoid function to generate weights G^w and G^h on the channel dimensions. The process can be expressed as [34,35]:

$$\begin{aligned} [f^h, f^w] &= P_2 T_k P_1 [z^h, z^w] \\ k &= \left\lfloor \frac{\log_2(C)}{\lambda} + \frac{b}{\lambda_{\text{odd}}} \right\rfloor \\ G^w &= \sigma(f^w) \\ G^h &= \sigma(f^h) \end{aligned} \quad (7)$$

where T_k denotes a two-dimensional convolution operation with convolution kernel size k , and λ and b are two hyperparameters to control the convolution kernel size. P_1 and P_2 represent permute operations, C adjusts the sequence of vector dimensions to $1 \times (H + W) \times C$ and P_2

adjusts the sequence of vector dimensions to $C \times (H + W) \times 1$. The number of parameters required for the original global channel linear transformation is $2C^2 / r$. The number of parameters required for the improvement in this paper is k^2 . The calculation of the original CA module compared to the designed LCA is shown in Figure 4.

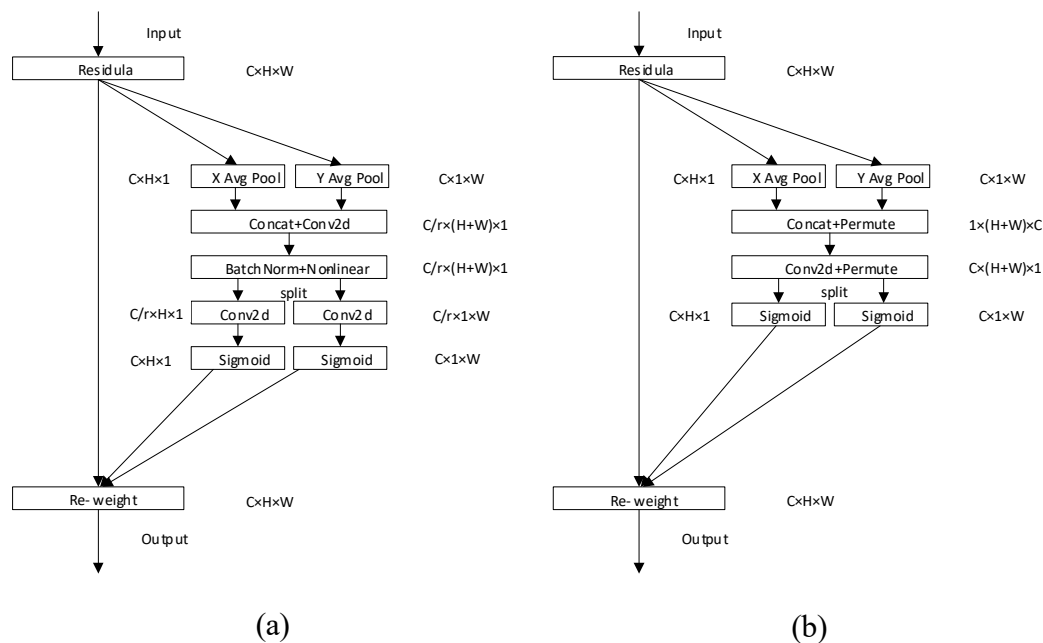


Figure 4. Calculation diagrams before and after the improvement of the LCA module. (a) represents before improvement, (b) represents after improvement.

3.4. Construction of Meta-ACON adaptive activation functions

In order to prevent the problem of model overfitting in the study of COVID-19, the Meta-ACON activation function was used as the activation function for some of the locations of the network. The Meta-ACON activation function is an extension of the ACON family of activation functions that dynamically selects either a nonlinear mapping or a linear mapping, improving the generalization of the model. Compared to the Swish [36] function, the Meta-ACON activation function has learnable upper and lower bounds on the first-order derivatives, which is the key to its performance improvement. In the New Crown Pneumonia study, the use of the Meta-ACON activation function prevented overfitting and improved the generalization of the network model due to the small dataset and the high correlation of the lung CT image sequences.

The ACON family of functions is derived from a smooth approximation of the max function [37], which is described as a smooth approximation of a differentiable function:

$$S_{\beta}(x_1, x_2, \dots, x_n) = \frac{\sum_{i=1}^n x_i e^{\beta x_i}}{\sum_{i=1}^n e^{\beta x_i}} \quad (8)$$

where β is the switching factor and S_β is the smoothed maximum value function. When $\beta = 0$, the output of S_β is the summation average of x_i . When $\beta = \infty$, S_β is equivalent to the max function. When $n = 2$, the S_β function [37] can be represented by a sigmoid function:

$$S_\beta(\eta_a(x), \eta_b(x)) = (\eta_a(x) - \eta_b(x)) \times \sigma[\beta(\eta_a(x) - \eta_b(x))] + \eta_b(x) \quad (9)$$

where σ is the sigmoid function, $\eta_a(x)$ and $\eta_b(x)$ correspond to different ACON functions when they are different expressions [36]: ACON-A, ACON-B, ACON-C. From the literature [37] they correspond as follows:

$$\begin{cases} x \times \sigma(\beta x) & \eta_a(x) = x, \eta_b(x) = 0 \\ (1-p)x \times \sigma[\beta(1-p)x] + px & \eta_a(x) = x, \eta_b(x) = px \\ (p_1 - p_2)x \times \sigma[\beta(p_1 - p_2)x] + p_2x & \eta_a(x) = p_1x, \eta_b(x) = p_2x \end{cases} \quad (10)$$

where p_1 , p_2 and β are learnable parameters, and $p_1 \neq p_2$, $\beta = G(X)$, $X \in \mathbb{R}^{H \times W \times C}$ and $G(X)$ are called the generating functions of β . Unlike the Swish function with fixed upper and lower bounds (1.0998, -0.0998) for the first order derivatives, the ACON-C function has upper and lower bounds for the first order derivatives [37]:

$$\begin{aligned} \max\left(\frac{d}{dx}[f_{ACON-C}(x)]\right) &\approx 1.0998p_1 - 0.0998p_2 \\ \min\left(\frac{d}{dx}[f_{ACON-C}(x)]\right) &\approx 1.0998p_2 - 0.0998p_1 \end{aligned} \quad (11)$$

where $f_{ACON-C}(x)$ is the maximum/minimum value of the upper/lower bound output. Meta-ACON designs the generating function $G(X)$ of β on the basis of the ACON-C function, where the values of $X \in \mathbb{R}^{C \times H \times W}$. β are differentiated by the channel dimension and its value is determined by the feature map of each channel. Its functional relationship expression is as follows [37]:

$$\beta_c = \sigma W_1 W_2 \sum_{h=1}^H \sum_{w=1}^W x_{c,h,w} \quad (12)$$

where β_c is the structure of the output channel, C is the number of channels, h is the height of the channel, W is the width of the channel, W_1 and W_2 are 1×1 convolution operations and W_2 will

adjust the channel dimension to C/r and w_1 will adjust the channel dimension to C . It is worth mentioning that since the switching factor β determines the nonlinearity in the activation, we generate β values for each sample at different levels, which can be pixel level, channel level and layer level. In this paper, Meta-ACON provides a function space $G(x)$ at the nonlinearity level by customizing the nonlinearity of each layer. Specifically, the learning switching factor β depends explicitly on the input sample $X \in \mathbb{R}^{C \times H \times W}$, $\beta = G(x)$ and the function space $G(x)$ uses the channel structure and ACON-C as described above.

3.5. Loss function

Dice coefficient is a region-based metric that is often used in image processing to assess region overlap and can be used to evaluate the performance of model segmentation. Among them, the Dice Loss weighted by the Laplace smoothing coefficient, referred to as L_D in this paper, can reduce the overfitting during the training process, and the formula of L_D is as follows [38]:

$$L_D = 1 - \frac{2(A \cap B) + s}{|A| + |B| + s} \quad (13)$$

where A represents the region with positive segmentation results, B represents the region with positive labeling and s is the Laplace smoothing factor. Using L_D as the model loss function, when the segmented foreground region is small, smaller segmentation errors may cause L_D to change substantially, leading to an unstable training process. The commonly used cross-entropy loss function L_{CE} is formulated as [39]:

$$L_{CE} = A \log B + (1-A) \log(1-B) \quad (14)$$

In this paper, the joint function L of L_D and L_{CE} is used as the loss function of the model:

$$L = \lambda L_{CE} + (1-\lambda) L_D \quad (15)$$

where λ represents the ratio of L_D and L_{CE} . In this paper, we take $\lambda = 0.5$. The joint loss function

L both accelerates the convergence of the network and ensures the equalization of the samples, which improves the effectiveness of the segmentation experiment.

4. Experiments

The experiments in this paper consisted of two tasks: lung parenchyma segmentation and COVID-19 lesion segmentation. Since the lesion area has the characteristics of being finer, more complex edges and more similar gray values to the torso, its difficulty in the image segmentation task is much greater than the segmentation of the lung parenchyma portion. Therefore, we conduct the experiment mainly to evaluate the segmentation of the lesion area, and the segmentation of the lung parenchyma will only be conducted for the comparison experiments with the other models.

4.1. DATASETS

The dataset used in this paper was merged from two COVID-19 datasets containing CT scan data of 20 COVID-19 cases [40] with annotations and 50 COVID-19 cases [41] totaling 70 cases, respectively. The former was named COVID-Dataset-1 and the latter was named COVID-Dataset-2. The ratio of infected area to lung area in COVID-19 patients' lungs in COVID-Dataset-1 spanned from 0.01 to 59%, and the patients' right and left lung areas as well as the infected areas were annotated by two radiologists with 5–10 years of experience. These annotations were then validated and refined by senior radiologists with more than 10 years of experience. The dataset included more than 1800 slices. COVID-Dataset-2 was derived from a publicly available dataset from the Moscow City Hospital in Russia. This dataset was annotated by experts from the Moscow healthcare sector on CT images of the lungs of 50 individuals. The patients' infection grades ranged from no infection, minor infection, moderate infection, severe to critical infection. The grades of infection were classified as 0%, less than 25%, less than 50%, 50–75% and more than 75% of lung involvement. Severe and critical infections presented as a coexistence of gross glass shadows and solid lung lesions on CT images, while the others presented as gross glass shadows.

The dataset in this paper was divided into two categories: the lung parenchyma dataset and the lesion segmentation dataset. The lung parenchyma dataset was derived from the lung parenchyma mask and its samples in COVID-Dataset-1. The lesion segmentation dataset synthesized the lesion area mask and its samples from COVID-Dataset-1 and COVID-Dataset-2.

4.2. Experimental environment and implementation details

The experimental environment was accelerated for training on the CUDA platform with CUDA version 9.0. The GPU model was NVIDIA TITAN XP with a memory size of 12 GB. The network model was built based on the PyTorch 1.3.1 platform. The training used a learning rate of 0.001, a batch size of 10, an RMSProp function for the optimizer, a weight decay coefficient of $1E-5$, a momentum parameter of 0.9 and a binary cross-entropy function for the loss function. The reduction factor r was set to 8. The model tended to stabilize after 70 epochs of training, where the loss was close in both the training set and the validation set. The model reached the optimal state. The loss function curve is shown in Figure 5.

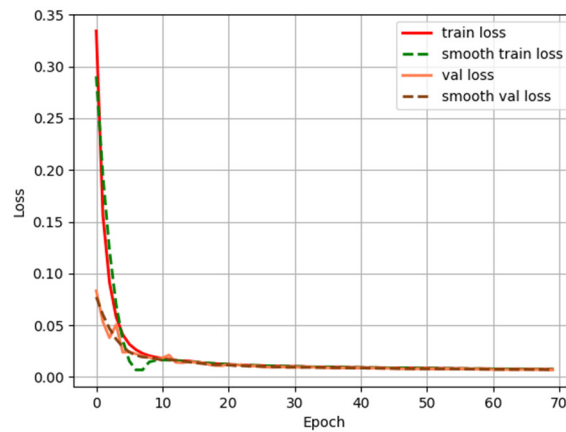


Figure 5. Loss function curve for lesion segmentation task.

4.3. Experimental results

Image segmentation under deep learning is a pixel-level classification task and the evaluation metrics are similar to those of classification tasks. We are concerned with the categorization of infected areas in COVID-19 CT images, which vary in size from lesion to lesion. Therefore, the evaluation metrics included Precision, Recall, Dice coefficient, Sensitivity (Sen), Specificity (Spec) and Intersection over Union (IoU) for the classification of infected lesions, as well as F1 scores and Mean Intersection over Union (MIoU) for all classifications. In this paper, the weight generation part of coordinate attention is designed to capture the channel local information by convolution operation, which improves the performance and the number of parameters. Therefore, the experiment compares the number of parameters with the above classification performance evaluation indexes.

In this case, precision and recall and intersection ratios are used in image segmentation for a particular classification at the pixel level, and precision is the ratio of the number of correct classifications to the number of all those predicted to be in that class, with a higher precision indicating fewer misdetections. Recall is the ratio of the number of correctly categorized to all the number whose true value is in that category; the higher the recall, the fewer the missed tests. IoU refers to the ratio of the intersection and concatenation of the network's predicted classification results and the true values, with the following formula:

$$IoU = \frac{TP}{TP + FP + FN} \quad (16)$$

where TP is true positive, FP is false positive and FN is false negative, and they all represent the number of pixels between the predicted result and the true label. MIoU is an evaluation metric that combines the area of infection (foreground) and background with the following formula:

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{FN + FP + TP} \quad (17)$$

where TP is true positive, FP is false positive and FN is false negative, and they all represent

the number of pixels between the predicted result and the true label, and k represents the number of classifications after removing the background, and the value of k in this paper is 1.

4.3.1. Quantitative analysis

In order to verify the state-of-the-art of the network designed in this paper, we chose to compare the network proposed in this paper with other improved models based on UNet networks. We conducted experiments on the dataset used in this paper using a total of six models, UNet, Attention-UNet [42], UNet++ [16], Res-UNet [43], Inf-Net [26] and MiniSeg [44]. Various segmentation indexes are shown in Table 2, it can be seen that the method in this paper is slightly lower than Attention-UNet in terms of recall, the number of parameters is only second to UNet++ and UNet model, and all other indexes are higher than the same type of network. Taken together, the proposed LMSA-Net performs optimally.

Table 2. Results of quantitative comparison of lesion segmentation.

Network Models	IoU (%)	Accuracy (%)	Recall Rate (%)	MioU (%)	F1 Score (%)	Dice	Sen	Spec	Params (M)
UNet	69.78	85.29	78.03	84.14	81.50	0.311	0.664	0.836	14.79
Attention-UNet	73.70	83.92	86.83	86.07	85.35	0.470	0.701	0.930	34.87
UNet++	72.19	82.23	85.91	86.74	84.03	0.444	0.853	0.948	9.046
Res-UNet	72.99	85.39	81.97	85.28	83.64	0.485	0.665	0.945	13.04
Inf-Net	74.11	85.99	82.33	85.96	84.46	0.669	0.766	0.965	23.32
MiniSeg	74.63	87.01	82.64	86.06	85.08	0.696	0.794	0.971	16.44
LMSA-Net	76.27	88.91	84.49	87.98	86.64	0.722	0.866	0.983	15.44

4.3.2. Qualitative analysis

In order to visualize the gap between this paper and other UNet-based networks in terms of segmentation details of the lesion region, the segmentation maps output from our network and the above model were visually compared, and the results obtained are shown in Figure 6, in which the first column is Ground Truth of the test map mask, and the subsequent columns are the mask outputs of the test network. As can be seen in the first two rows, where the lesion area is small and the shape edges are not yet complex, the gap between the models is small under naked eye observation, and in the last row, where the edges of the infected area are more complex, the outputs of the models are distorted to varying degrees. Among them, the segmentation effect of UNet within the red box labeling is the worst, which is related to its feature extraction ability and the lack of localization ability, the LMSA-Net proposed in this paper has the smallest void region compared to other models, and the output of this paper's algorithm in the region labeled by the red box is the most similar to the GT diagram with the best performance.

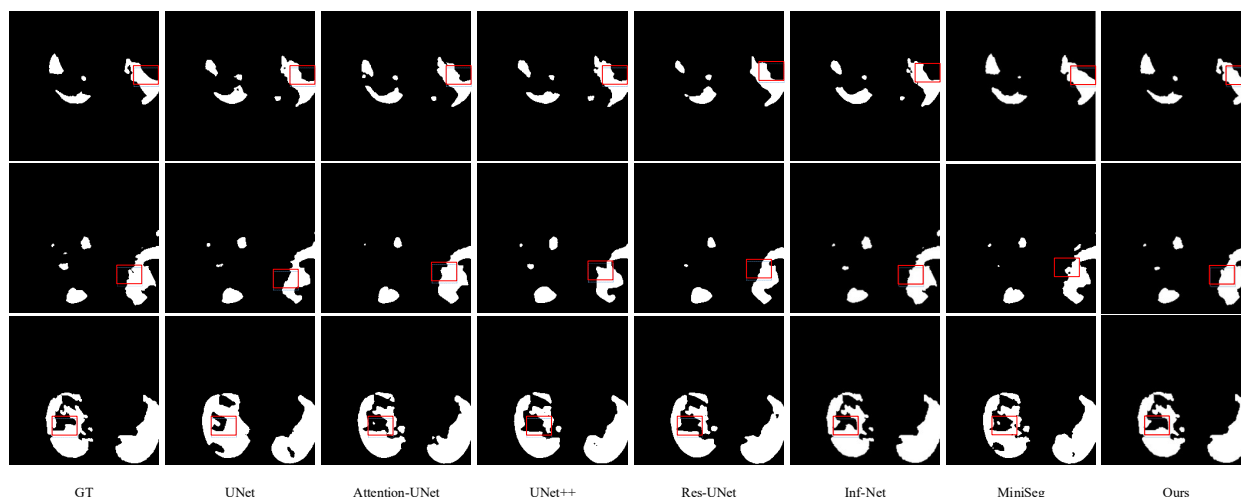


Figure 6. Qualitative comparison of lesion segmentation (zoomed in for a better view). Represented from left to right: the GT image, UNet, Attention-UNet, UNet++, Res-UNet and ours.

4.4. Lung parenchyma segmentation test

In order to validate the performance of the model on lung parenchyma segmentation, we conducted a comparative lung parenchyma segmentation experiment, in which the original 3D data were sliced through the lung window adjustment to obtain a sample size of 3520 slices, and the sequence of slices was randomly disrupted and then assigned to the training set, the validation set and the test set in an 8:1:1 manner and scaled. The final information of the dataset is obtained as shown in Table 3 below.

The model training parameters differed from the lesion segmentation experiments only in the number of epochs during training; after a total of 40 epochs of training for the lung parenchyma segmentation the loss function no longer decreased and the model converged. The lung parenchyma segmentation loss function curve is shown in Figure 7.

The model of this paper will carry out the lung parenchyma segmentation comparison experiment. The experiment is shown in Figure 8 can be seen in the lung parenchyma segmentation performance. As can be seen from Table 4, all models can achieve better results in the segmentation of lung parenchyma, which is caused by the smooth edges of the lungs, the large difference in gray values between the lung parenchyma region and the trunk, and the obvious image features. Nevertheless, the model proposed in this paper can still outperform the other models in the comparative experiments in all metrics. As can be seen from Figure 8, the lung parenchyma segmented by the model in this paper outperforms the other models in terms of continuity as well as edge detail.

Table 3. Pulmonary parenchymal segmentation experimental dataset information.

Names	Quantities	Sizes	Grayscale Map Size
Training set	2816	256×256	8 bit
Validation set	352	256×256	8 bit
Test set	352	256×256	8 bit

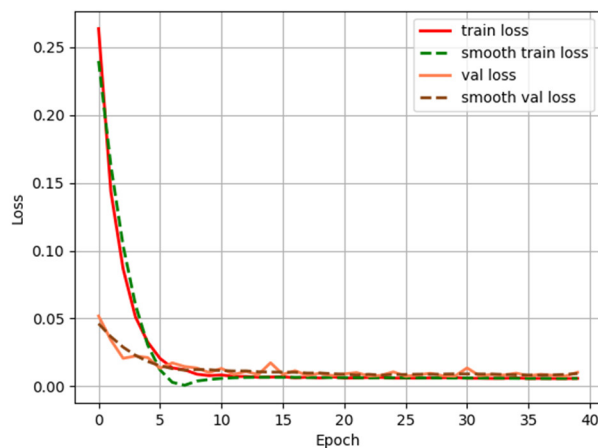


Figure 7. Lung parenchyma segmentation loss function curves.

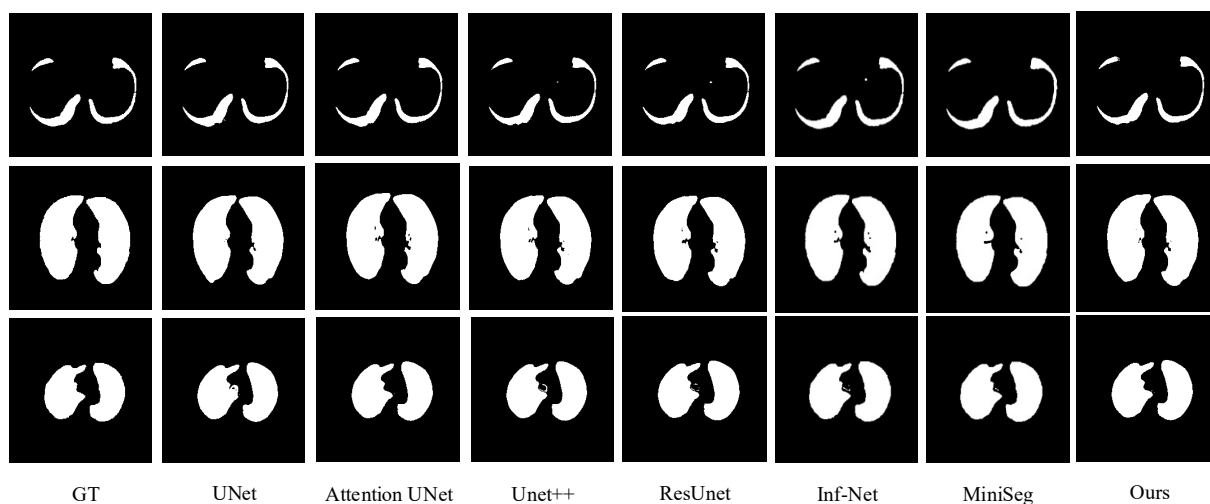


Figure 8. Qualitative comparison of lung parenchyma segmentation (clearer when magnified). From left to right, GT images, UNet, Attention-UNet, UNet++, Res-UNet and our images.

Table 4. Comparative quantitative results of lung parenchymal segmentation.

Network	IoU(%)	Accuracy (%)	Recall Rate (%)	MioU (%)	Dice	Sen	Spec	F1 Score (%)
UNet	91.49	92.84	97.34	95.56	0.577	0.656	0.892	95.03
Attention-UNet	94.51	96.53	97.82	96.97	0.641	0.699	0.865	97.17
UNet++	92.01	94.22	97.67	97.25	0.689	0.745	0.932	95.91
ResUNet	93.65	95.91	97.54	96.49	0.645	0.754	0.912	96.72
Inf-Net	94.32	96.14	97.39	96.99	0.706	0.811	0.942	97.10
MiniSeg	94.69	96.66	97.34	97.01	0.712	0.821	0.961	96.98
LMSA-Net	95.21	97.77	97.31	97.35	0.822	0.865	0.987	97.55

4.5. Ablation studies

In order to validate the effectiveness of the designed modules in improving the performance of network lesion segmentation, we train and test the computational analysis of different modules on mixed datasets for evaluation metrics. We analyze the influence of each module on the final performance of the network from four aspects: the design and improvement of LCA module, MSCA module, loss function and activation function. For the fairness of the experiments, we conducted the ablation experiments using the UNet with the added LCA module as the BASELINE. The baseline network uses the ReLU activation function and ensures the consistency of each parameter when adding modules. Among them, the results of the benchmark network modeling and ablation experiments are shown in Table 5.

Table 5. Performance of each module on the lesion segmentation dataset.

Baseline	MSCA	Meta-ACON	IoU (%)	Accuracy (%)	Recall Rate (%)
√			74.13	87.36	82.23
√	√		75.02	88.23	83.15
√	√	√	76.27	88.91	84.49

4.5.1. Ablation of the LCA modules

In order to verify the performance of the improved LCA module proposed in this paper compared with that of the pre-improved lesion segmentation, we counted the overall number of parameters and test set segmentation indexes of the LCA module network before and after the improvement by adding only the LCA module based on the UNet model with the rest of the network structure, as well as the hyperparameters remaining unchanged. As can be seen in Table 6, the improved LCA module has a slight decrease in recall, but is better than before the improvement in the rest of the metrics: the intersection and merge ratio is improved by 0.65 percent, the precision rate is improved by 0.45 percent, the average intersection and merge ratio is improved by 0.42 percent and the number of references is reduced by 0.075 M. The improved LCA module is also better than the improved LCA module. It can be seen that the designed LCA module is able to reduce the number of parameters of the model while maintaining the performance.

Table 6. Comparison of the performance of the CA and LCA modules of the lesion segmentation experiment.

Names	Params (M)	Accuracy (%)	IoU (%)	Recall Rate (%)	MioU (%)	F1 Score (%)
CA module	29.017	86.91	73.48	82.25	86.77	84.52
LCA module	28.942	87.36	74.13	82.23	87.19	84.72

4.5.2. Ablation of the MSCA module

In order to verify the performance of this paper for the improved MSCA module compared with the pre-improved lesion segmentation, we counted the test set segmentation indexes of the MSCA module network before and after the improvement by adding only the MSCA module on the basis of

the UNet model, while the rest of the network structure as well as the hyper-parameters remain unchanged. As can be seen from Table 7, after adding the LCA module, the model is more accurate in the segmentation task, and all the indexes have been improved compared with those before the improvement.

Table 7. Comparison of MSCA module performance before and after lesion segmentation experiments.

Names	Accuracy (%)	IoU (%)	Recall Rate (%)	MioU (%)	F1 Score (%)
MSCA	87.30	73.66	82.19	86.91	84.67
MSCA+LCA	88.23	74.13	83.15	87.19	85.61

4.5.3. Loss function comparison

In this section, the joint loss function L used is compared with Dice Loss, a commonly used loss function for medical image segmentation. Dice Loss is a region-dependent loss function that will work better when the dataset is severely unbalanced in terms of positive and negative samples. In order to verify the performance of both on the lesion segmentation dataset used in this paper, comparative experiments were done with the network structure as well as various other parameters unchanged. As can be seen from Table 8, the joint loss function L is much better in the segmentation task, and all the indexes are improved compared to Dice Loss.

Table 8. Comparison of loss functions for lesion segmentation experiments.

Loss function	IoU (%)	Accuracy (%)	Recall Rate (%)	MIoU (%)
Dice Loss	75.11	88.24	83.78	87.35
L Loss	76.27	88.91	84.49	87.98

4.5.4. Activation function comparison

In order to verify that the Meta-ACON activation function used in this paper improves the performance of lesion segmentation, we conducted ablation experiments on common activation functions in medical image segmentation and Meta-ACON activation function respectively with the network structure as well as the hyperparameters remaining unchanged and counted the segmentation metrics of the test set. From Table 9, it can be seen that the Meta-ACON activation function outperforms the other activation functions in each metric. It can be seen that the Meta-ACON activation function is effective in avoiding overfitting as well as improving network characterization.

Table 9. Comparison of activation functions for lesion segmentation experiments.

Activation function	IoU (%)	Accuracy (%)	Recall Rate (%)	MIoU (%)
ReLU	74.14	85.36	82.95	85.46
Sigmoid	74.65	85.91	83.15	86.02
Leaky ReLU	75.03	86.11	83.66	86.33
Meta-ACON	76.27	88.91	84.49	87.98

5. Discussion

To date, variants produced by COVID-19 continue to affect human health. In this paper, although our approach achieves good results, it is worth noting that it has some limitations. With respect to the dataset, COVID-19 lesions have similar imaging characteristics to other types of virus-induced pneumonia. Due to the lack of laboratory confirmation of the etiology of these cases, the vast majority of the available datasets do not compare other viral pneumonias detected, which leads to the small size of the dataset used and affects the generalization ability of the model to some extent. In addition, a deep learning-based approach may introduce certain biases. This is due to data imbalance, data labeling errors, data quality and diversity, model complexity and capacity. In recent years, how to address and improve deep learning-based methods has become a focus of research. The medical community is currently exploring the systematic adoption of argumentation frameworks for eXplainable Artificial Intelligence (XAI) in medical informatics. Proposals in [45] investigate the benefits of using logic methods for interpretable AI by demonstrating how the natural characteristics of interpretability and expressiveness of logic methods can contribute to the design of intelligent systems that are ethical, interpretable and rational. On the other hand, some approaches solve the problem through artificial intelligence based on simple segmentation techniques, ensuring an understandable diagnosis for both the expert and the patient. The idea of using a multi-instance learning paradigm to classify pneumonia X-ray images has been proposed in [46] and appears to be very fast in practice, especially considering that no preprocessing techniques are used. In any case, some of the above solutions can provide a fast, efficient and reliable alternative to the problems related to the possible biases introduced by the COVID-19 deep learning-based approach. This approach complements traditional medical diagnostic strategies and accelerates research in image analysis while reducing the burden on physicians.

6. Conclusions

In this paper, we propose a new Unet segmentation model for Crown pneumonia based on a multi-scale attention mechanism as well as an adaptive activation function. Specifically, the MSCA structure for multi-scale information fusion is introduced into the network structure of UNet to enhance the network's ability to capture multi-scale information and to enhance the network's ability to localize a small target such as a new coronary pneumonia. Additionally, a local channel-domain interaction attention (LCA) module is used to replace the channel-domain attention weight calculation in the original MSCA, which makes the module synthesize the spatial-domain attention information as well as the channel-domain attention information, and cuts down the number of parameters under the premise of guaranteeing the performance, so as to make the network tend to be lightweight. In addition, the network has changed the activation function and used an adaptive activation function, Meta-ACON, to enhance the network's characterization. The experimental results show that the model proposed in this paper achieves better objective evaluation metrics and subjective visualization on both the hybrid datasets COVID-Dataset-1 and COVID-Dataset-2 compared to the current mainstream lung parenchyma segmentation and COVID-19 lesion segmentation algorithms.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported in part by the Project of The Opening Fund of Artificial Intelligence Key Laboratory of Sichuan Province (2021RYY04); Natural Science Foundation of Sichuan, China (2023NSFSC1987); Sichuan University of Science & Engineering Postgraduate Innovation Fund Project (Y2022118).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. C. Wang, P. W. Horby, F. G. Hayden, G. F. Gao, A novel coronavirus outbreak of global health concern, *Lancet*, **395** (2020), 470–473. [https://doi.org/10.1016/S0140-6736\(20\)30185-9](https://doi.org/10.1016/S0140-6736(20)30185-9)
2. C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet*, **395** (2020), 497–506. [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5)
3. N. Mu, H. Wang, Y. Zhang, J. Jiang, J. Tang, Progressive global perception and local polishing network for lung infection segmentation of COVID-19 CT images, *Pattern Recognit.*, **120** (2021), 108168. <https://doi.org/10.1016/j.patcog.2021.108168>
4. Y. Fang, H. Zhang, J. Xie, M. Lin, L. Ying, P. Pang, et al., Sensitivity of chest CT for COVID-19: comparison to RT-PCR, *Radiology*, **296** (2020), E115–E117. <https://doi.org/10.1148/radiol.2020200432>
5. S. Inui, A. Fujikawa, M. Jitsu, N. Kunishima, S. Watanabe, Y. Suzuki, et al., Chest CT findings in cases from the cruise ship diamond princess with coronavirus disease (COVID-19), *Radiol. Cardiothorac. Imaging*, **2** (2020), e200110. <https://doi.org/10.1148/ryct.2020200110>
6. C. Butt, J. Gill, D. Chun, B. A. Babu, Deep learning system to screen coronavirus disease 2019 pneumonia (Retracted Article), *Appl. Intell.*, **53** (2023), 4874–4874. <https://doi.org/10.1007/s10489-020-01714-3>
7. P. Arbelaez, M. Maire, C. Fowlkes, J. Malik, Contour detection and hierarchical image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **33** (2011), 898–916. <https://doi.org/10.1109/TPAMI.2010.161>
8. C. Shen, N. Yu, S. Cai, J. Zhou, J. Sheng, K. Liu, et al., Quantitative computed tomography analysis for stratifying the severity of Coronavirus Disease 2019, *J. Pharm. Anal.*, **10** (2020), 123–129. <https://doi.org/10.1016/j.jpha.2020.03.004>
9. A. Oulefki, S. Agaian, T. Trongtirakul, A. K. Laouar, Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images, *Pattern Recognit.*, **114** (2021), 107747. <https://doi.org/10.1016/j.patcog.2020.107747>
10. J. Wen, S. Xuan, Y. Li, Q. Gao, Q. Peng, Image-segmentation algorithm based on wavelet and data-driven neutrosophic fuzzy clustering, *Imaging Sci. J.*, **67** (2019), 63–75. <https://doi.org/10.1080/13682199.2018.1549694>

11. R. Pohle, K. D. Toennies, Segmentation of medical images using adaptive region growing, in *Medical Imaging 2001: Image Processing*, (2001), 1337–1346. <https://doi.org/10.1117/12.431013>
12. Z. Tu, S. C. Zhu, Image segmentation by data-driven Markov chain Monte Carlo, *IEEE Trans. Pattern Anal. Mach. Intell.*, **24** (2002), 657–673. <https://doi.org/10.1109/34.1000239>
13. E. Shelhamer, J. Long, T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 640–651. <https://doi.org/10.1109/TPAMI.2016.2572683>
14. V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: a deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
15. O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
16. Z. Zhou, M. R. Siddiquee, N. Tajbakhsh, J. Liang, UNet++: a nested U-Net architecture for medical image segmentation, in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, (2018), 3–11. https://doi.org/10.1007/978-3-030-00889-5_1
17. G. Rani, P. Thakkar, A. Verma, V. Mehta, R. Chavan, V. S. Dhaka, et al., KUB-UNet: segmentation of organs of urinary system from a KUB X-ray image, *Comput. Methods Programs Biomed.*, **224** (2022), 107031. <https://doi.org/10.1016/j.cmpb.2022.107031>
18. Z. Tang, W. Zhao, X. Xie, Z. Zhong, F. Shi, T. Ma, et al., Severity assessment of COVID-19 using CT image features and laboratory indices, *Phys. Med. Biol.*, **66** (2021), 035015. <https://doi.org/10.1088/1361-6560/abfb9e>
19. X. Wang, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, et al., A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT, *IEEE Trans. Med. Imaging*, **39** (2020), 2615–2625. <https://doi.org/10.1109/TMI.2020.2995965>
20. Y. Akbari, H. Hassen, S. Al-Maadeed, S. M. Zughair, COVID-19 lesion segmentation using lung CT scan images: comparative study based on active contour models, *Appl. Sci.*, **11** (2021), 8039. <https://doi.org/10.3390/app11178039>
21. Y. Cao, Z. Xu, J. Feng, C. Jin, X. Han, H. Wu, et al., Longitudinal assessment of COVID-19 using a deep learning-based quantitative CT pipeline: illustration of two cases, *Radiol. Cardiothorac. Imaging*, **2** (2020), e200082. <https://doi.org/10.1148/ryct.2020200082>
22. L. Huang, R. Han, T. Ai, P. Yu, H. Kang, Q. Tao, et al., Serial quantitative chest CT assessment of COVID-19: a deep learning approach, *Radiol. Cardiothorac. Imaging*, **2** (2020), e200075. <https://doi.org/10.1148/ryct.2020200075>
23. F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, et al., Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction, *Med. Phys.*, **48** (2021), 1633–1645. <https://doi.org/10.1002/mp.14609>
24. S. Chaganti, P. Grenier, A. Balachandran, G. Chabin, S. Cohen, T. Flohr, et al., Automated quantification of CT patterns associated with COVID-19 from chest CT, *Radiol. Artif. Intell.*, **2** (2020), e200048. <https://doi.org/10.1148/ryai.2020200048>
25. Q. Yan, B. Wang, D. Gong, C. Luo, W. Zhao, J. Shen, et al., COVID-19 chest CT image segmentation – a deep convolutional neural network solution, preprint, arXiv:2004.10987.

26. D. P. Fan, T. Zhou, G. Ji, Y. Zhou, G. Chen, H. Fu, et al., Inf-Net: automatic COVID-19 lung infection segmentation from CT images, *IEEE Trans. Med. Imaging.*, **39** (2020), 2626–2637. <https://doi.org/10.1109/TMI.2020.2996645>
27. Y. Jiang, H. Chen, M. Loew, H. Ko, COVID-19 CT image synthesis with a conditional generative adversarial network, *IEEE J. Biomed. Health. Inf.*, **25** (2020), 441–452. <https://doi.org/10.1109/JBHI.2020.3042523>
28. V. K. Singh, M. Abdel-Nasser, N. Pandey, D. Puig, LungINFseg: segmenting COVID-19 infected regions in lung CT images based on a receptive-field-aware deep learning framework, *Diagnostics*, **11** (2021), 158. <https://doi.org/10.3390/diagnostics11020158>
29. T. He, H. Liu, Z. Zhang, C. Li, Y. Zhou, Research on the application of artificial intelligence in public health management: leveraging artificial intelligence to improve COVID-19 CT image diagnosis, *Int. J. Environ. Res. Public Health*, **20** (2023), 1158. <https://doi.org/10.3390/ijerph20021158>
30. J. He, Q. Zhu, K. Zhang, P. Yu, J. Tang, An evolvable adversarial network with gradient penalty for COVID-19 infection segmentation, *Appl. Soft Comput.*, **113** (2021), 107947. <https://doi.org/10.1016/j.asoc.2021.107947>
31. Y. Song, J. Liu, X. Liu, J. Tang, COVID-19 infection segmentation and severity assessment using a self-supervised learning approach, *Diagnostics*, **12** (2022), 1805. <https://doi.org/10.3390/diagnostics12081805>
32. M. Versaci, G. Angiulli, P. di Barba, F. C. Morabito, Joint use of eddy current imaging and fuzzy similarities to assess the integrity of steel plates, *Open Phys.*, **18** (2020), 230–240. <https://doi.org/10.1515/phys-2020-0159>
33. P. R. G. Kurka, A. A. D. Salazar, Applications of image processing in robotics and instrumentation, *Mech. Syst. Signal Process.*, **124** (2019), 142–169. <https://doi.org/10.1016/j.ymsp.2019.01.015>
34. J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-excitation networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **42** (2020), 2011–2023. <https://doi.org/10.1109/TPAMI.2019.2913372>
35. Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, Q. Hu, ECA-Net: efficient channel attention for deep convolutional neural networks, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2020), 11531–11539. <https://doi.org/10.1109/CVPR42600.2020.01155>
36. P. Ramachandran, B. Zoph, Q. V. Le, Searching for activation functions, preprint, arXiv:1710.05941.
37. N. Ma, X. Zhang, M. Liu, J. Sun, Activate or not: learning customized activation, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2021), 8028–8038.
38. R. Zhao, B. Qian, X. Zhang, Y. Li, R. Wei, Y. Liu, et al., Rethinking dice loss for medical image segmentation, in *2020 IEEE International Conference on Data Mining (ICDM)*, (2020), 851–860. <https://doi.org/10.1109/ICDM50108.2020.00094>
39. F. Milletari, N. Navab, S. Ahmadi, V-Net: fully convolutional neural networks for volumetric medical image segmentation, in *2016 Fourth International Conference on 3D Vision (3DV)*, (2016), 565–571. <https://doi.org/10.1109/3DV.2016.79>
40. J. Ma, Y. Wang, X. An, C. Ge, Z. Yu, J. Chen, et al., Towards efficient COVID-19 CT annotation: a benchmark for lung and infection segmentation, preprint, arXiv:2004.12537.

41. S. P. Morozov, A. E. Andreychenko, N. A. Pavlov, A. V. Vladzimirskyy, N. V. Ledikhova, V. A. Gombolevskiy, Mosmeddata: chest ct scans with COVID-19 related findings dataset, 2020. <https://doi.org/10.1101/2020.05.20.20100362>
42. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, et al., Attention U-Net: learning where to look for the pancreas, preprint, arXiv:1804.03999.
43. F. I. Diakogiannis, F. Waldner, P. Caccetta, C. Wu, ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data, *ISPRS J. Photogramm. Remote Sens.*, **162** (2020), 94–114. <https://doi.org/10.1016/j.isprsjprs.2020.01.013>
44. Y. Qiu, Y. Liu, S. Li, J. Xu, MiniSeg: an extremely minimum network for efficient COVID-19 segmentation, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **35** (2021), 4846–4854. <https://doi.org/10.1609/aaai.v35i6.16617>
45. L. Caroprese, E. Vocaturo, E. Zumpano, Argumentation approaches for explainable ai in medical informatics, **16** (2022), 200109. <https://doi.org/10.1016/j.iswa.2022.200109>
46. E. Zumpano, A. Fuduli, E. Vocaturo, M. Avolio, Viral pneumonia images classification by Multiple Instance Learning: preliminary results, in *Proceedings of the 25th International Database Engineering & Applications Symposium*, (2021), 292–296. <https://doi.org/10.1145/3472163.3472170>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)