



FAM: focal attention module for lesion segmentation of COVID-19 CT images

Xiaoxin Wu¹ · Zhihao Zhang² · Lingling Guo³ · Hui Chen¹ · Qiaojie Luo⁴ · Bei Jin⁵ · Weiyan Gu⁵ · Fangfang Lu² · Jingjing Chen⁶

Received: 15 April 2022 / Accepted: 12 August 2022 / Published online: 4 September 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

The novel coronavirus pneumonia (COVID-19) is the world's most serious public health crisis, posing a serious threat to public health. In clinical practice, automatic segmentation of the lesion from computed tomography (CT) images using deep learning methods provides a promising tool for identifying and diagnosing COVID-19. To improve the accuracy of image segmentation, an attention mechanism is adopted to highlight important features. However, existing attention methods are of weak performance or negative impact to the accuracy of convolutional neural networks (CNNs) due to various reasons (e.g. low contrast of the boundary between the lesion and the surrounding, the image noise). To address this issue, we propose a novel focal attention module (FAM) for lesion segmentation of CT images. FAM contains a channel attention module and a spatial attention module. In the spatial attention module, it first generates rough spatial attention, a shape prior of the lesion region obtained from the CT image using median filtering and distance transformation. The rough spatial attention is then input into two 7×7 convolution layers for correction, achieving refined spatial attention on the lesion region. FAM is individually integrated with six state-of-the-art segmentation networks (e.g. UNet, DeepLabV3+, etc.), and then we validated these six combinations on the public dataset including COVID-19 CT images. The results show that FAM improve the Dice Similarity Coefficient (DSC) of CNNs by 2%, and reduced the number of false negatives (FN) and false positives (FP) up to 17.6%, which are significantly higher than that using other attention modules such as CBAM and SENet. Furthermore, FAM significantly improve the convergence speed of the model training and achieve better real-time performance. The codes are available at GitHub (<https://github.com/RobotvisionLab/FAM.git>).

Keywords COVID-19 · Lesion segmentation · Attention mechanism · Machine learning · Public health

1 Introduction

According to the report from the Center for Systems Science and Engineering (CSSE) of Johns Hopkins University, until May 25, 2022, COVID-19 has resulted in

526,824,747 infections, of which, 6,280,794 deaths. Rapid detection of the infection is essential to prompt isolation and treatment of the patients. At present, reverse transcription-polymerase chain reaction (RT-PCR) is the most widely adopted method for COVID-19 diagnosis. However, RT-PCR suffers from some drawbacks such as time

Xiaoxin Wu and Zhihao Zhang have contributed equally to this work and share first authorship.

✉ Fangfang Lu
lufangfang@shiep.edu.cn

✉ Jingjing Chen
joyjchan@gmail.com

¹ State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

² College of Computer Science and Technology, Shanghai University of Electric Power, Shanghai, China

³ College of Chemical Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang, China

⁴ School of Stomatology, Stomatology Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China

⁵ Department of Oral and Maxillofacial Surgery, Taizhou Hospital, Wenzhou Medical University, Taizhou, Zhejiang, China

⁶ Zhejiang University City College, Hangzhou, Zhejiang, China

consuming, false negative caused by the sampling quality [1]. The chest computed tomography (CT) images captured from COVID-19 patients frequently include patchy bilateral shadows or ground-glass opacity in the lung [2], hence chest CT is adopted as an dominant method for the diagnosis of COVID-19 [1, 3]. Compared with RT-PCR, chest CT image is easy to obtain in the clinical practice, therefore it can be used for the severity classification of COVID-19 patients, for which, contouring the lesion is an essential procedure. The traditional manual contouring is tedious, time-consuming, and heavily depending on doctor's clinical experience, therefore, there is an urgent need for an automated lesion segmentation method specially designed for COVID-19 CT images.

Nowadays, convolution neural network (CNN), a typical deep learning method, is becoming an essential for the segmentation of COVID-19 CT image. Widely used CNNs include FCN [4], SegNet [5], UNet series (UNet [6], UNet++ [7], UNet3+ [8], etc.), and DeepLab series [9–11]. These methods usually rely on a large-scale dataset with high-quality pixel-level annotation of COVID-19 lesions. The need for large-scale data collection and data labeling before the model training prevents it from wide adoption in the context of public health.

Attention mechanism is a technology widely used in the fields such as natural language processing (NLP), statistical learning, image processing, voice recognition. It stems from the particularly selective attention of human vision. Attention mechanism focuses on important information with high weights, ignores unrelated information with low weights, and continuously adjusts weights so that important information can be selected in different situations, thus making it expandable and robust. With the consideration of application contexts, attention can be grouped into spatial, channel, layer, mixed, and temporal domains. Spatial and channel domains are most widely used in the tasks of image processing. Many excellent attention modules such as SENet [12], CBAM [13], SKNet [14] are proposed. To improve the accuracy of deep learning models for COVID-19 lesion segmentation, attention modules (or their variations) have been integrated into the state-of-the-art segmentation networks. However, existing attention modules always cannot fully utilize the characteristics of CT images. Moreover, they always disrupt the original feature distribution of the input data, resulting in low segmentation accuracy and the inefficiency of the network training convergence.

To address the abovementioned issue, we propose a novel design, in which the lesion in CT image is treated as rough spatial attention and then combined with a channel attention module to achieve a novel plug-and-play attention module, (named focal attention module (FAM)) for lesion segmentation of COVID-19 CT images. The main contributions of this study are as follows:

- 1 A novel spatial attention module is proposed. It introduces the shape prior information of the lesion region to improve the feature analysis and weight redistribution of the attention module and accelerates the convergence of the network training.
- 2 We sequentially combine the spatial attention module in the form of the residual block with the channel attention module, constructing a novel Focal Attention Module (FAM) for lesion segmentation of COVID-19 CT images.
- 3 FAM is integrated into six state-of-the-art networks and is validated on the public COVID-19 CT image dataset.

The rest of the paper is organized as follows: Sect. 2 describes the work related to the proposed method. Section 3 details the design and implementation of the proposed method. The experiment and discussion is described in Sect. 4. Finally, we conclude the study in Section.

2 Related work

In this section, we first discuss these existing deep learning-based methods for lesion segmentation of COVID-19 CT images, followed by related work on the attention mechanism, in the end we introduce the applications of shape priors in image segmentation.

2.1 Lesion segmentation of COVID-19 CT images

The data annotation is usually with labor cost and time-consuming, large-scale segmentation datasets of COVID-19 lesions are rarely available. Meanwhile, training networks on a small-scale dataset suffers from the issues such as overfitting and poor generalization performance. Existing deep learning methods are proposed to attenuate these models' reliance on a large-scale dataset. The attention mechanism is used to enhance the capability of feature extraction of the network. For example, Fan et al. [15] combined a semi-supervised learning model and FCN8s network with implicit reverse attention and explicit edge attention mechanism to achieve a novel model. It achieves a sensitivity of 72.5% and an accuracy of 96.0%. Chen et al. [16] proposed a residual attention UNet and applied a soft attention mechanism to enhance the capability of feature learning of the model. The proposed model achieves a performance with a segmentation accuracy of 89%. Zhao et al. [17] integrated their proposed spatial-wise and channel-wise attention modules on UNet++ [7]. The Dice Similarity Coefficients (DSC) of the model is 88.99%. A number of novel loss functions and special network modules are also proposed. For example, Wang et al. [18] proposed noise-robust dice loss to solve the problem of poor training results caused by low-quality labels, and the DSC of the model is 80.72%. Inspired by

contrast enhancement methods and Atrous Spatial Pyramid Pooling (ASPP) [10], Yan et al. [19] proposed a novel Progressive Atrous Spatial Pyramid Pooling (PASPP) module to progressively aggregate information and obtain more useful contextual features, and the DSC of the model is 72.60%. Elharrouss et al. [20] proposed a multi-class segmentation network based on an encoder-decoder structure, and the multi-input stream of the network allows the model to learn more features. It achieves a sensitivity of 71.1%. In addition, multi-scale features fusion [21], multipoint supervised training [22], and conditional generation model [23] are promising for improving the segmentation accuracy of COVID-19 lesions.

2.2 Attention mechanism

Attention is an essential and complex cognitive function in the human brain [24]. With attention, people can work methodically while receiving a large amount of information through vision, hearing, touch, etc. The human brain can select small portions of interested information from these large amounts of input information to focus on, meanwhile ignoring other portions.

In the context of computer vision, attentions can be divided into soft attention and hard attention [25]. For soft attention, by calculating the attention weight, all data is included in the attention range, and no filter condition for the data feature is set. Hard attention sets the filtering condition after calculating the attention weight and forms a part of the attention weight value that does not meet the condition to 0. Contrarily, soft attention is probabilistic, end-to-end differentiable, and utilizes back-propagation and forward-propagation to learn the attention weight without the posterior sampling. There are a number of studies regarding the soft attention. Inspired by translation and rotation without deformation of the pooling mechanism, Jaderberg et al. [26] proposed a spatial transformation module that could learn the transformation from the network. It was widely used for Optical Character Recognition (OCR). Hu et al. [12] proposed a channel attention model (SENet), but SENet cannot capture spatial contextual information. Woo et al. [13] expanded the SENet and proposed an attention module (CBAM) to constrain and enhance the input feature map from the channel and spatial dimensions. But, the spatial attention module of CBAM fails to capture information at different scales and is not able to establish a long-range dependency. Inspired by the classical non-local means method [27] for image processing, Wang et al. [28] proposed an attention module (non-local neural networks) for capturing long-range dependencies. Fu et al. [29] amalgamated the advantages of CBAM and Non-local Neural Networks to propose the DANet, an attention module widely used in

semantic segmentation. Drawing on the idea of residual networks, Wang et al. [30] proposed a novel solution to solve the problem of information reduction caused by stacked attention modules. A Criss-Cross Attention is proposed by Huang et al. [31], to reduce the calculations of Non-local Neural Networks. Gao et al. [32] proposed a Spatially Modulated Co-Attention (SMCA) mechanism to accelerate training convergence, but it suffers from the increased time of computation and inference. A particular channel attention module [33] was proposed to distinguish the esophagus and surrounding tissues from esophageal cancer. However, there are limited literatures regarding the hard attention, and studies [34–36] argued that reinforcement learning is required for training in hard attention due to its non-differentiability.

Although there are a number of studies [16, 17, 37, 38] introducing the attention mechanism for lesion segmentation on COVID-19 CT images, the improvement in the performance and accuracy of these models is still urgently expected in academia and industry.

2.3 Shape priors in image segmentation

Traditional segmentation methods (e.g., thresholding, watershed, and region growing) usually suffer from the lack of robustness and poor segmentation accuracy due to the noise, low contrast, and complexity of objects in medical images. Recently, the rapid development of deep learning methods promoted the adoption of deep learning-based image segmentation algorithms in medical image segmentation. Studies [39–43] have shown that integrating prior knowledge of objects into rigorous segmentation formulas can improve the segmentation accuracy of a specific target. The prior knowledge has been utilized in various forms, e.g., user interaction, object shape and appearance [44].

The shape is one of the most important geometric attributes of anatomical objects, and shape priors can reduce the search space of the potential segmentation outputs for deep learning models [45]. Ravishankar et al. [46] incorporated the shape model explicitly in FCN through a novel loss function that penalizes the deviation of the predicted segmentation mask from a learned shape model. Avanti et al. [47] used stacked automatic encoders to infer the target shape, then the inferred shape is incorporated into deformable models to improve the accuracy and robustness. In addition, Ngo et al. [48] and Cremers et al. [40] combined level set and genetic algorithms with deep learning to improve the training effect of the model on small datasets. Zhao et al. [49] obtained the shape prior of the lung region through threshold segmentation to optimize the segmentation of the lung.

3 Proposed method

3.1 Design rationale of focal attention module

Given an intermediate feature map $\mathbf{F} \in \mathbb{R}^{C \times H \times W}$, and an input CT image $\mathbf{I} \in \mathbb{R}^{1 \times H \times W}$ is defined as the input. FAM sequentially infers a 1D channel attention map $\mathbf{M}_c \in \mathbb{R}^{C \times 1 \times 1}$ and a 2D spatial attention map $\mathbf{M}_s \in \mathbb{R}^{1 \times H \times W}$ as illustrated in Fig. 1. The attention module is formularized as:

$$\mathbf{F}' = \mathbf{M}_c(\mathbf{F}) \otimes \mathbf{F}, \tag{1}$$

$$\mathbf{F}'' = (\mathbf{M}_s(\mathbf{I} + 1) \otimes \mathbf{F}', \tag{2}$$

where \otimes denotes element-wise multiplication. During multiplication, the attention values are broadcasted accordingly: channel attention values are broadcasted along the channel dimension, and vice versa. \mathbf{F}'' is the final refined output. Different from naive stacking attention modules (e.g., CBAM, SENet), the feature map refined by the spatial attention module (as depicted in Fig. 2) is combined as a residual branch into the feature map refined by the channel attention module due to the following analysis:

1. When the input \mathbf{I} is a negative sample, spatial attention obtained from its lung image \mathbf{L} by distance transformation contains less feature information. In this case, stacking will degrade the value of features in deep layers.

2. Residual branch works as feature selectors, which enhance good features and suppress noise from trunk features.
3. Inspired by Residual Attention Network [38], attention residual learning not only keeps good properties of original features but also allows to be refined by the spatial attention module.

3.2 Channel attention module

The channel attention module focuses on “what” is meaningful given the feature maps. To compute the channel attention efficiently, spatial information of a feature map is first aggregated by average-pooling and max-pooling operations, respectively, thus two different spatial context descriptors (i.e. \mathbf{F}_{avg}^c and \mathbf{F}_{max}^c) are obtained. Both descriptors are then forwarded to a multi-layer perceptron (MLP) with one hidden layer, achieving two output feature maps. Finally, the output feature maps are merged using element-wise summation. To reduce the number of parameters, the hidden layer size is set to $\mathbb{R}^{C/r \times 1 \times 1}$, where r is the reduction ratio. The channel attention is formularized as:

$$\begin{aligned} \mathbf{M}_c(\mathbf{F}) &= \sigma(\text{MLP}(\text{AvgPool}(\mathbf{F})) + \text{MLP}(\text{MaxPool}(\mathbf{F}))) \\ &= \sigma\left(\mathbf{W}_1\left(\mathbf{W}_0\left(\mathbf{F}_{avg}^c\right)\right) + \mathbf{W}_1\left(\mathbf{W}_0\left(\mathbf{F}_{max}^c\right)\right)\right), \end{aligned} \tag{3}$$

where σ represents the sigmoid function, $\mathbf{W}_0 \in \mathbb{R}^{C/r \times C}$ and $\mathbf{W}_1 \in \mathbb{R}^{C \times C/r}$. \mathbf{W}_0 and \mathbf{W}_1 are the weights of the MLP and are shared for both inputs. The ReLU activation function is followed by \mathbf{W}_0 .

Fig. 1 The overview of focal attention module, which consists of channel attention module and spatial attention module

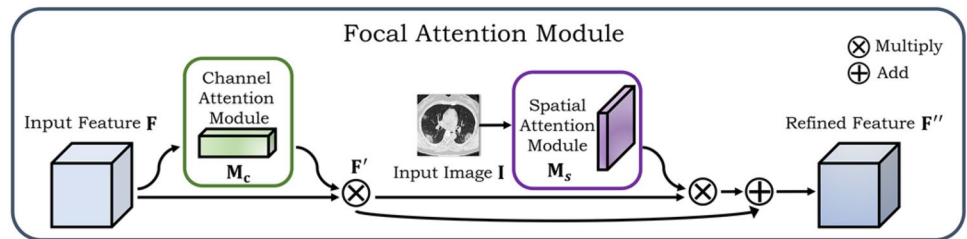
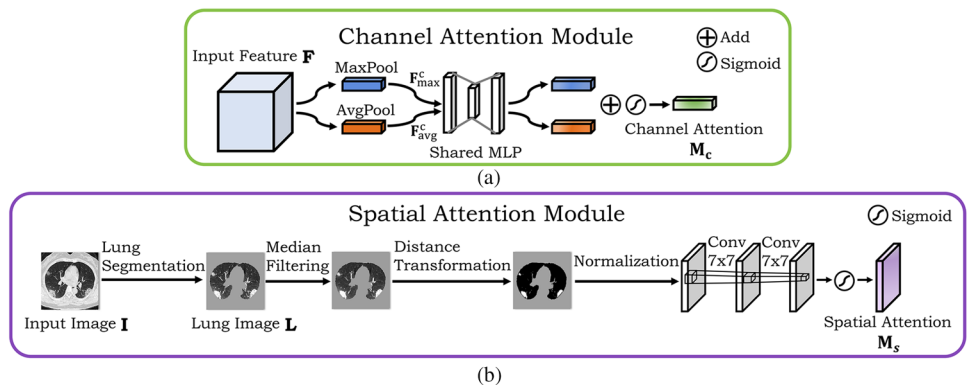


Fig. 2 a The channel attention module: max-pooling, average-pooling outputs, and a multi-layer perceptron; b the spatial attention module obtains a rough shape prior of the lesion region by median filtering and distance transformation



3.3 Spatial attention module

The spatial attention focuses on “where” is the interested in given feature maps. The spatial attention in CBAM is learned from a 7×7 convolution layer, and it has two shortcomings when dealing with the specific task of lesion segmentation for COVID-19 diagnosis: (1) low efficiency on learning process; (2) the change of spatial attention to the feature space easily cause the problematic convergence of training and poor network generalization performance, especially when the dataset is small or the parameters of the backbone network are few. Two mechanisms are introduced to address these issues: (1) adopting the residual structure and refined the feature maps with spatial attention while preserving trunk network features (as shown in Fig. 1); (2) utilizing the shape prior of the COVID-19 lesion region to reduce the search space of the spatial attention module. The main steps (as shown in Fig. 2b) while computing the spatial attention include lung segmentation, median filtering, and distance transformation.

3.3.1 Lung segmentation

To efficiently obtain the shape prior of the lesion region, the lung needs to be segmented from CT images. Currently, many excellent methods of lung segmentation have been proposed and widely used. These methods are mainly divided into three types: traditional image processing-based algorithms, deep learning-based algorithms and the combination of the two former methods. Because segmentation of the lung is not the focus of this paper, the lung region is segmented with a simple mask operation from labels in the dataset.

3.3.2 Median filtering

Median filtering is introduced to eliminate partial noise pixels consisting of the pulmonary trachea and pulmonary vessels from the lung image. Median filtering, a nonlinear method that can preserve the details of the edges of an image

while eliminating noise, has been widely used in fields such as image enhancement and image recovery. As shown in Fig. 3a, a few noise pixels in the lung region, such as regions of the tiny pulmonary trachea and pulmonary vessels, interfere significantly with the accurate segmentation of lesions. As shown in Fig. 3b, the median filtering eliminates most of the small pulmonary trachea and pulmonary vessels. For the large pulmonary trachea and pulmonary vessels, the median filtering also reduces their pixel region. Meanwhile, median filtering retains the nature of the lesions (i.e. ground-glass opacity) with little reduction in the area of the lesions due to its large pixel region.

3.3.3 Distance transformation

Distance transformation (DT) is to convert a digital binary image that consists of object and non-object pixels into another image in which each object pixel owns a value corresponding to the minimum distance from the non-object by a distance function [50, 51]. Distance transformation is widely used for target thinning, object skeleton extraction. Euclidean distance, city block distance, and chessboard distance are widely used measures for distance transformation. The full workflow of distance transformation is introduced as follows:

Given an image J , it's binarized to get an image J_b . In J_b , 1 is associated with object pixel and 0 with the background pixel. Hence, we have a pixel set \mathcal{O} represented by all the object pixels and \mathcal{O}^c represented by all the background pixels.

$$\mathcal{O} = \{t | J_b(t) = 1\}, \quad (4)$$

$$\mathcal{O}^c = \{b | J_b(b) = 0\}, \quad (5)$$

where t and b represent the pixel of objects and background respectively. The distance transformation (DT) generates a map D , in which the value of each pixel in \mathcal{O} is the smallest distance from this pixel to \mathcal{O}^c :

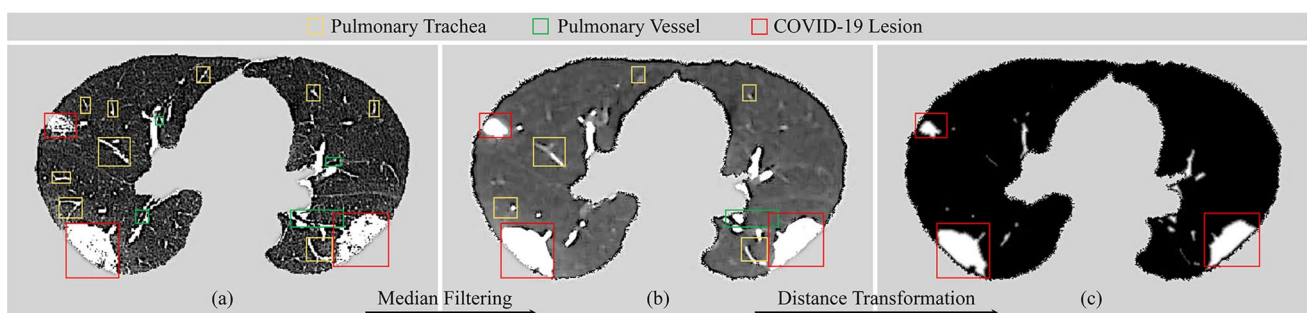


Fig. 3 The process of eliminating noise pixels in the lung region of CT image step by step. **a** A lot of noise pixels (i.e. pulmonary trachea and pulmonary vessels inside the lung region). **b** Applying median

filtering to partially eliminate the noise pixels. **c** Applying distance transformation to further eliminate the noise pixels and extract the main lesion region

$$D(t) = \min \{d(t, b) | t \in \mathcal{O}, b \in \mathcal{O}^c\}, \tag{6}$$

where the image D is called the distance map of J . It is assumed that \mathcal{O} contains at least one pixel. Otherwise, the output of the DT is undefined, i.e., the outliers will be ignored in the distance transformation. Moreover, $d(t, b)$ represents Euclidean distance, is formularized as:

$$d(t, b) = \sqrt{(t_x - b_x)^2 + (t_y - b_y)^2} \tag{7}$$

$$0 \leq x < W, \quad 0 \leq y < H,$$

where H and W represent the height and width of the image J respectively.

As shown in Fig. 3c, distance transformation is used to eliminate noise pixels (i.e., the pulmonary trachea and pulmonary vessels) and extract the main lesion region. By applying

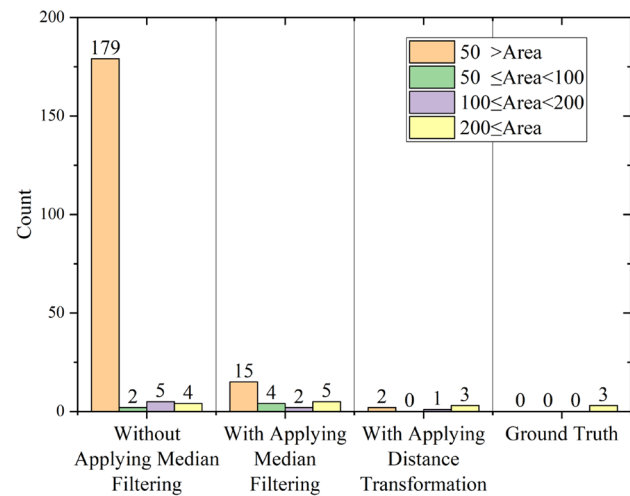


Fig. 4 The distributions of connected regions in a lung image without/with applying sequential median filtering, with applying distance transformation

sequential median filtering and distance transformation, the distribution of connected regions in a lung image is shown in Fig. 4. The connected regions containing more than 200 pixels represent the lesion region, and those with a small area represent the region of the pulmonary trachea and pulmonary vessels. By applying distance transformation, the distribution of connected regions in the lung image is close to the ground truth. Distance maps of several lung images are shown in Fig. 5. By comparing the distance map with the corresponding lesion label, the main lesion region is extracted.

Normalization serves as an activation function to obtain the shape prior of the lesion region. As shown in Fig. 6b, distance transformation strengthens the weight of the main connected regions and weakens the weight of the edge parts (as shown in Fig. 6a). Furthermore, as shown in Fig. 6c, normalization removes both the edge and connected regions with a small area. The normalization function $Norm$ is formularized as:

$$Norm(X) = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \tag{8}$$

where X represents an image matrix, X_{\min} and X_{\max} represent the minimum and maximum values in X respectively. Next, as illustrated in Fig. 2b, two 7×7 convolution layers are utilized to learn the attention weight of edge pixels which has a low boundary contrast to surroundings, as well as adaptively tuning up attention weights to other regions. Finally, refined spatial attention is obtained. Spatial attention is formularized as:

$$\mathbf{M}_s(\mathbf{I}) = \sigma(f_{7 \times 7}^{\text{cov}}(f_{7 \times 7}^{\text{cov}}(Norm(DT(f^{\text{med}}(f^{\text{seg}}(\mathbf{I}))))))), \tag{9}$$

where \mathbf{I} is the input CT image, σ represents the sigmoid function, f^{seg} and f^{med} represent the lung segmentation network and median filtering, respectively. $f_{7 \times 7}^{\text{cov}}$ is the 7×7 convolution layer. DT represents the distance transformation.

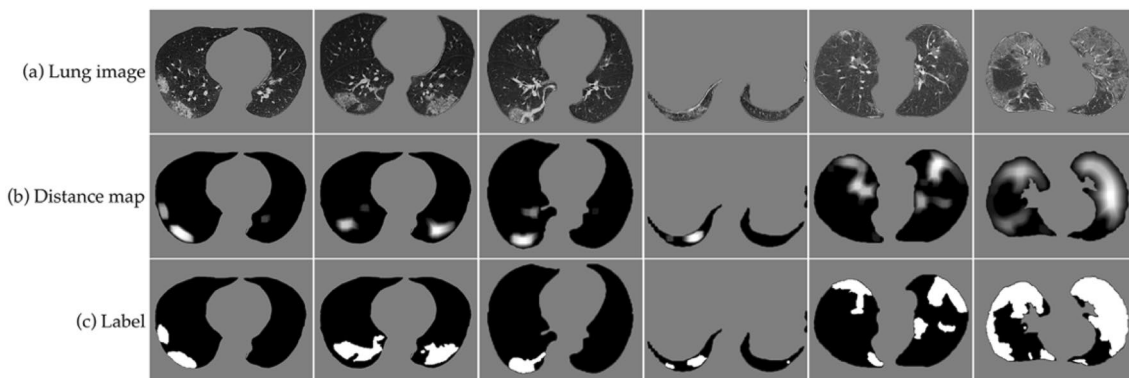


Fig. 5 Distance maps of several lung images: **a** lung region is segmented from CT images in the dataset. **b** Distance maps of lung images obtained by distance transformation. **c** By comparing the distance map with the corresponding lesion label, the main lesion region is extracted

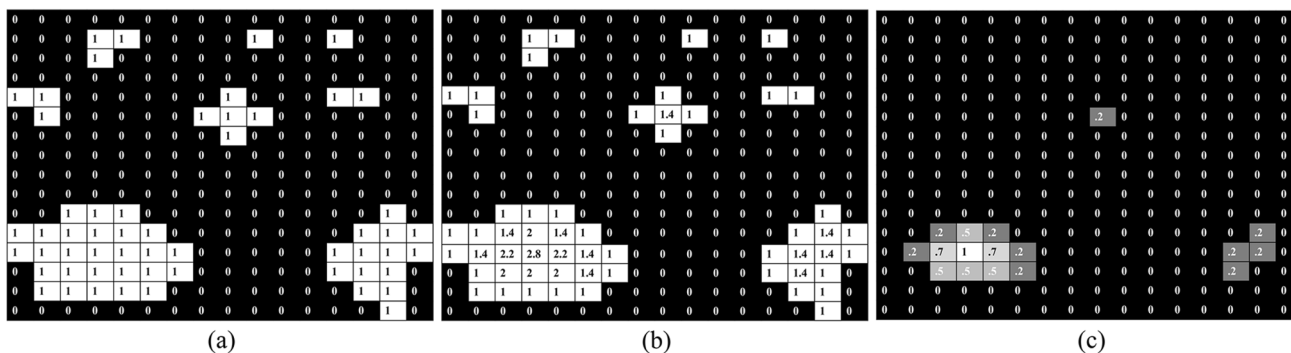
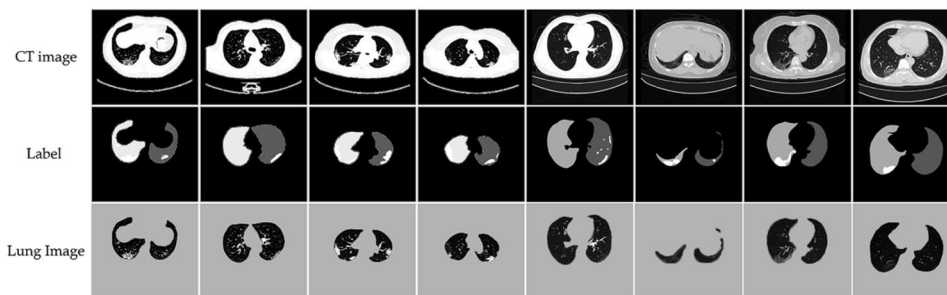


Fig. 6 A numerical example of distance transformation: **a** a binary image containing several connected regions; **b** the distance map of (a); **c** is the normalization of (b)

Fig. 7 The dataset contains segmentation labels for the left lung, right lung and COVID-19 lesions. The lung region is segmented based on the lung label



4 Experiment

4.1 Dataset and metrics

The dataset contains 20 groups of labeled CT scans of COVID-19 patients. Two radiologists manually labeled the regions of left lung, right lung, and COVID-19 lesions on the images and an extra verification of the labeled regions was carried out by another radiologist. CT slices are exported from the CT scans as 2D images (752×670 pixels), and a group of processed images are listed in Fig. 7. The dataset contains a total of 3520 CT images and is divided into a training set and a testing set at a ratio of 4:1 (as detailed in Table 1). It is noted that “positive” means CT images with the lesion, and “negative” means healthy CT images.

The Dice Similarity Coefficients (DSC), false negatives (FN), false positives (FP) and inference time of the networks are adopted as evaluation metrics. DSC is a standard metric for comparing the pixel-wise results between the ground truth and predicted segmentation. It is formularized as follows:

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|}, \tag{10}$$

Table 1 Details of the dataset

	Positive	Negative	Total
Train	1474	1341	2815
Test	369	336	705
Total	1843	1677	3520

where A is the lesion label, and B denotes the segmented lesion image.

4.2 Training method

An Adam optimizer with a learning rate of 0.001 is employed to minimize the binary cross-entropy (BCE) loss. The total number of training epochs and batch size is set to 100 and 1, respectively. The network weights are initialized with Kaiming initialization, and the network biases are initialized by 0. Moreover, the positive and negative samples are trained alternately, and the dataset is shuffled in each iteration. The formula of the BCE loss is as follows:

$$Loss_{BCE} = -[G \log(P) + (1 - G) \log(1 - P)], \tag{11}$$

where G is the label and P is the output of the network.

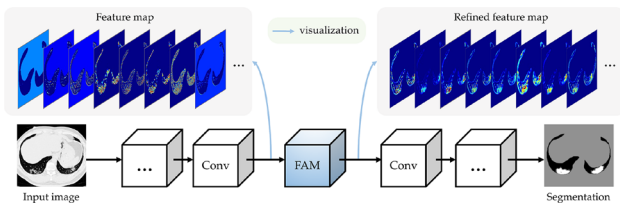


Fig. 8 The structure of the integration of FAM with the network

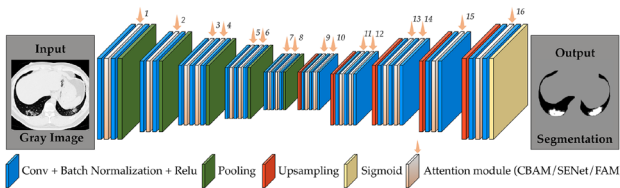


Fig. 9 The workflow of attention modules integrated with SegNet

4.3 Ablation analysis

FAM can be conveniently integrated into any CNNs as a lightweight and plug-and-play attention module. As shown in Fig. 8, FAM is integrated into convolution layers to refine the intermediate feature map. FAM and two state-of-the-art attention modules (i.e. SENet [19] and CBAM [20]) are respectively inserted into every two convolution layers of six state-of-the-art segmentation networks [6–11, 16, 21] for ablation experiments. All the related networks and modules are reproduced in the framework PyTorch, trained and inferred on a single NVIDIA GeForce RTX 2080Ti GPU with CUDA v10.2 and cuDNN v7.6.5. The main concerns in this ablation analysis contains: shape prior, time complexity,

the performance of the network and convergence rate of the network training.

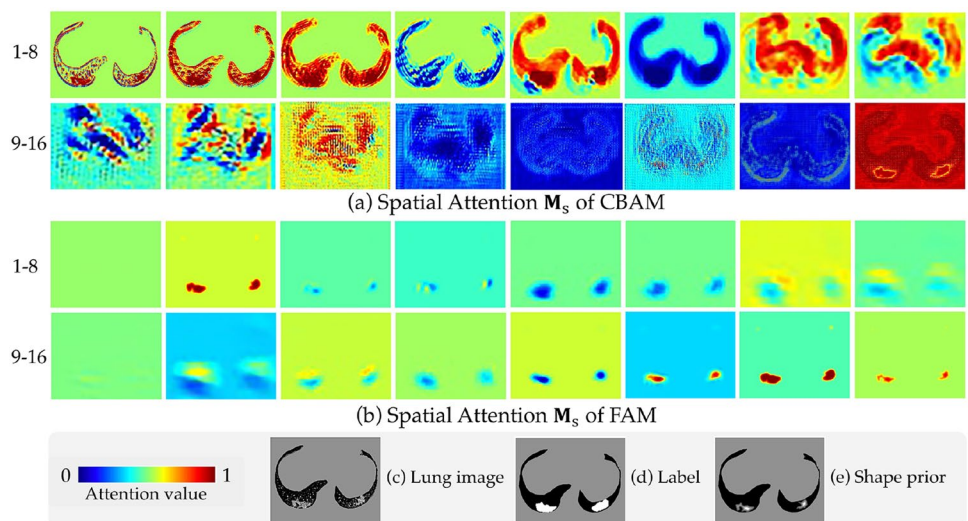
4.3.1 Shape prior

FAM is constructed based on the shape prior and the spatial attention module of CBAM. In the case that CBAM and FAM are integrated respectively between each convolution layer in SegNet (as shown in Fig. 9). A visualization of the spatial attention is built in Fig. 10 to illustrate the effect of shape prior. Since that SegNet is a typical encode-decode structure, in which, low-level features and high-level semantic features are extracted accordingly, the feature map contained in the middle layers is abstract. CBAM adopts such a feature extraction workflow because of the connection between its spatial attention module and the network, i.e., the input of the spatial attention module comes from inside the network rather than outside. Thus, the spatial attention of CBAM integrated into the middle layers is abstract, and that integrated into other layers highlights only the lung region rather than the lesion region, resulting in the inaccuracy of the spatial attention learning of CBAM. The shape prior information of the lesion region is introduced in the spatial attention module to reduce the search space. In summary, FAM focuses on the lesion region without being disturbed by the no-lesion region during the spatial attention learning process.

4.3.2 Time complexity

Time complexity determines the training and inference time of the network. The network with high time complexity suffer from poor real-time performance. Floating Point

Fig. 10 Spatial attention comparison between CBAM and FAM in SegNet. We visualized the spatial attention for sixteen attention modules of SegNet when using CBAM and FAM. **c** is the lung image segmented from the input CT image. **d** and **e** are corresponding lesion label and shape prior



Operations (FLOPs) is a classical metric of the time complexity. The time complexity of a network is formularized as follows:

$$T = O\left(\sum_{i=1}^l f_{FLOPs}(\text{layer})\right), \tag{12}$$

where l denotes the number of layers in the network. f_{FLOPs} represents the function that calculates the FLOPs of a layer. FLOPs for each type of layer is formularized as:

$$f_{FLOPs}^{conv} = 2C_{in}C_{out}K_wK_hQ_wQ_h, \tag{13}$$

$$f_{FLOPs}^{linear} = 2N_{in}N_{out}, \tag{14}$$

$$f_{FLOPs}^{pooling} = C_{in}Q_wQ_h, \tag{15}$$

$$f_{FLOPs}^{relu} = C_{in}Q_wQ_h, \tag{16}$$

$$f_{FLOPs}^{sigmoid} = 4C_{in}Q_wQ_h, \tag{17}$$

where f_{FLOPs}^{conv} , f_{FLOPs}^{linear} , $f_{FLOPs}^{pooling}$, f_{FLOPs}^{relu} and $f_{FLOPs}^{sigmoid}$ represent the function that calculate the FLOPs of convolution layer, full connect layer, global pooling layer, Relu layer and sigmoid layer respectively. These five types of layers are used in FAM, SENet and CBAM. C_{in} and C_{out} denote the channel number of the input and output feature map respectively. K_w and K_h denote the width and height of the convolution kernel respectively. Q_w and Q_h denote the width and height of the input feature map respectively. N_{in} and N_{out} denote the number of input and output neurons. We set the channel number of the input feature map to 16 and set the reduction ratio of the channel attention module to 16. The size of the input feature map is 652×752 . FLOPs of FAM, SENet and CBAM are listed in Table 2. SENet achieves the smallest FLOPs because it lacks a spatial attention module. FAM and CBAM share a similar structure, but FLOPs of the latter is less than the former by about 15 million. It is because FAM owns fewer pooling layers than CBAM. As shown in Table. 3, FLOPs of six state-of-the-art segmentation networks increase very little when FAM is integrated between every two convolution layers. Moreover, the inference time of FAM is 4 ms faster than that for CBAM on average. In conclusion, FAM is able to greatly improve the performance of the network with very little time complexity increasement.

Table 2 FLOPs of FAM, SENet and CBAM

	Conv	Linear	Pooling	Relu	Sigmoid	Total
SENet	0	64	7,844,864	1	64	7,844,993
CBAM	99,047,552	0	31,379,456	2	34,363,392	164,790,402
FAM	99,047,552	0	15,689,728	505,346	34,363,392	149,606,018

Bold represents values of our proposed method

4.3.3 Performance analysis of the networks with attention modules

The average DSC, FN and FP of all images in the test set are adopted to analyze the performance of the integration of attention modules to the networks (as detailed in Table 3). FAM adds almost no extra parameters to the network, as it only has one more 7×7 convolution than CBAM. The integration of FAM improves the DSC of all these networks, with an improvement of 2% for SegNet. In addition, the integration of FAM reduces the FN and FP, with a reduction of 17.6% for PSPNet.

SENet improves the performance of UNet, PSPNet and UNet++, meanwhile it degrades the performance of FCN, SegNet and deeplabV3+. CBAM improves the performance of SegNet and PSPNet, meanwhile it degrades the performance of FCN, UNet, Deeplabv3 and UNet++. This finding shows that the effectiveness of the integration of the attention module depends on the structure of the network.

Figure 11. shows the segmentation result of SegNet integrated with different attention modules on the dataset. As shown in Fig. 11, SegNet without the integration of the attention module suffers from false detection in lesions of both left and right lobes to some extent. The integration of SENet alleviates it but CBAM exacerbates it. Although the SegNet integrated with FAM has many false detections, overall it has the highest segmentation accuracy.

4.3.4 Convergence analysis of the network training

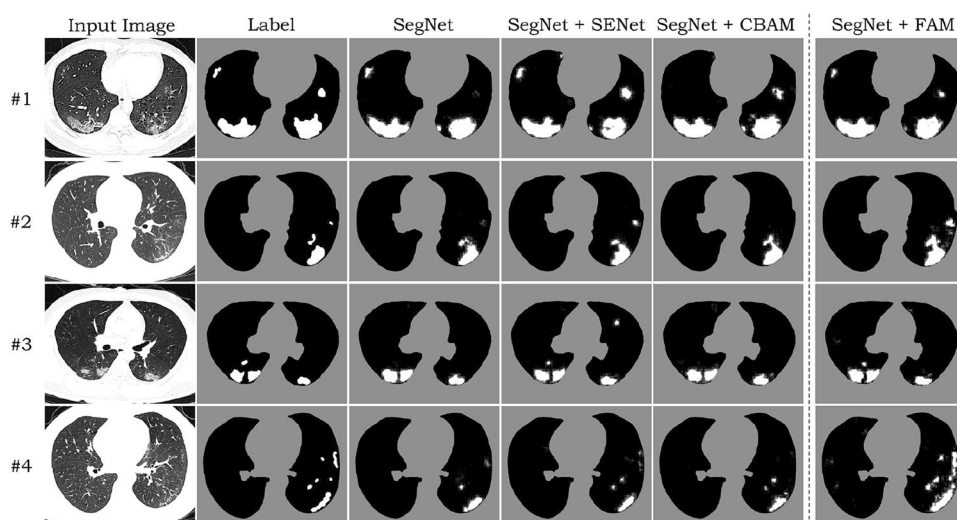
Six networks with attention modules are trained with the same experimental settings and dataset. As shown in Fig. 12, all networks converge within 60 iterations. FAM accelerates model training better than SegNet or CBAM does. For SegNet, UNet++ and DeepLabV3+, SENet does not significantly accelerate the model training. For UNet and UNet++, CBAM decelerates the convergence of the model training. However, FAM accelerates the training of FCN, UNet, SegNet, PSPNet and DeepLabV3+, as well as minimizing the loss.

For UNet++, FAM performs better than the other two attention modules. For UNet and PSPNet, FAM makes the training process more stable. The result shows that FAM achieves better convergence rate and less converged loss

Table 3 Performance comparison of FAM, SENet and CBAM on state-of-the-art segmentation network

Network	Attention module	#Param (k)	FLOPs (G)	DSC%	FN + FP	Inference time (ms)
FCN	–	19,169.03	201.18	85.80	824	29
	SENet	19,319.04	201.25	85.59	830	33
	CBAM	19,319.83	201.32	85.65	823	56
	FAM	19,320.04	201.30	86.41	762	53
UNet	–	31,042.37	421.05	86.03	770	69
	SENet	31,260.48	421.44	86.13	781	72
	CBAM	31,261.37	422.12	85.22	818	114
	FAM	31,262.25	421.57	86.56	767	110
SegNet	–	29,442.43	308.98	85.13	839	56
	SENet	29,742.47	309.79	85.03	837	60
	CBAM	29,744.05	309.93	85.80	796	106
	FAM	29,744.83	309.84	87.13	720	101
DeepLabV3+	–	59,233.51	170.07	87.42	712	42
	SENet	59,856.02	170.75	86.85	746	45
	CBAM	59,858.71	170.79	86.51	766	91
	FAM	59,860.01	170.76	88.54	601	85
PSPNet	–	27,509.47	313.67	86.67	769	56
	SENet	27,677.10	313.74	86.98	763	60
	CBAM	27,677.45	313.81	86.93	772	92
	FAM	27,677.83	313.80	88.12	634	90
UNet++	–	47,170.32	1539.32	85.67	820	187
	SENet	47,415.21	1539.58	85.85	800	191
	CBAM	47,417.20	1539.89	85.70	794	240
	FAM	47,418.06	1539.62	87.06	723	232

Bold represents values of our proposed method

Fig. 11 The comparison of the segmentation results on lesions of COVID-19 CT images by applying various combinations of SegNet and attention modules

value of the model training among these six networks than CBAM does. In addition, although SENet only applies its attention in the channel dimension, it achieves better performance than CBAM in specific networks such as FCN, UNet and PSPNet.

5 Conclusion

In this study, a lightweight and plug-and-play attention module, is proposed to improve the lesion segmentation

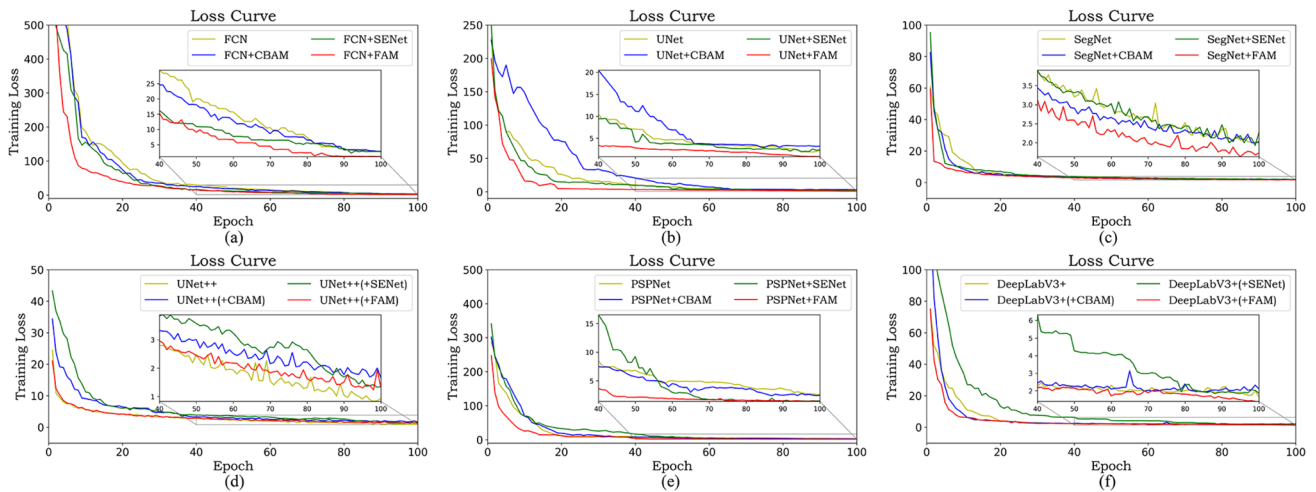


Fig. 12 Diagram of training networks integrated with various attention modules. **a–c** and **d–f** show that FAM accelerates the network training as well as minimizes the loss

performance of CNNs for COVID-19 CT images. FAM refines the input feature map from channel and space dimensions to maximize the network representation. In the spatial attention of FAM, shape prior of the lesion region is used to reduce the search space for attention learning. In addition, the feature map refined by spatial attention is added to the network as a residual branch. A set of experiments proved that: (1) FAM could improve the segmentation performance on a small-scale public COVID-19 CT image dataset; (2) FAM could accelerate the convergence speed of the model training; (3) FAM is capable of being stacked in a deep segmentation network without performance loss. (4) FAM could achieve better real-time performance.

FAM is promising for practical use in public health. In future, we will work towards improving the generated shape prior to enhance the generalization performance of FAM based on the up-to-date COVID-19 CT image datasets.

Acknowledgements This study was supported by Natural Science Foundation of Zhejiang Province (No. LQ21H190004), China Postdoctoral Science Foundation (No. 2020T130102ZX), Postdoctoral Science Foundation of Zhejiang Province (No. ZJ2020031), the Educational Commission of Zhejiang Province of China (No. Y202147553).

Data availability The datasets generated during and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Ai, T., Yang, Z., Hou, H., et al.: Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology* **296**(2), E32–E40 (2020). <https://doi.org/10.1148/radiol.20200642>
2. Adams, H.J., Kwee, T.C., Yakar, D., et al.: Chest CT imaging signature of coronavirus disease 2019 infection: in pursuit of the scientific evidence. *Chest* **158**(5), 1885–1895 (2020). <https://doi.org/10.1016/j.chest.2020.06.025>
3. Xu, X., Tian, H., Zhang, X., Qi, L., He, Q., Dou, W.: DisCOV: distributed COVID-19 detection on X-ray images with edge-cloud collaboration. *IEEE Trans. Serv. Comput.* (2022). <https://doi.org/10.1109/TSC.2022.3142265>
4. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(4), 640–651 (2016). <https://doi.org/10.1109/TPAMI.2016.2572683>
5. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017). <https://doi.org/10.1109/TPAMI.2016.2644615>
6. Ronneberger, O., Fischer, P., Brox, T.: “U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp 234–241 (2015). https://doi.org/10.1007/978-3-319-24574-4_28
7. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., et al.: Unet++: a nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp. 3–11. Springer (2018). https://doi.org/10.1007/978-3-030-00889-5_1
8. Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., Wu, J.: Unet 3+: A full-scale connected UNET for medical image segmentation. In: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, pp. 1055–1059. (2020) <https://doi.org/10.1109/ICASSP40776.2020.9053405>

9. Chen, L.-C., Papandreou, G., Kokkinos, I. et al.: Semantic image segmentation with deep convolutional nets and fully connected CRFS. (2014) [Online]. <https://arxiv.org/abs/1412.7062>
10. Chen, L.-C., Papandreou, G., Kokkinos, I., et al.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 834–848 (2017). <https://doi.org/10.1109/TPAMI.2017.2699184>
11. Florian, L.-C., Adam, S. H.: Rethinking atrous convolution for semantic image segmentation. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE/CVF, (2017) [Online]. <https://arxiv.org/abs/1706.05587>
12. Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E.: Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**(8), 2011–2023 (2020). <https://doi.org/10.1109/TPAMI.2019.2913372>
13. Woo, S., Park, J., Lee, J.-Y., Kweon, I. S.: Cbam: Convolutional block attention module. In: *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19 (2018) https://doi.org/10.1007/978-3-030-01234-2_1
14. Wu, W., Zhang, Y., Wang, D., Lei, Y.: Sk-net: deep learning on point cloud via end-to-end discovery of spatial keypoints. *Proc. AAAI Conf. Artif. Intell.* **34**(04), 6422–6429 (2020). <https://doi.org/10.1609/aaai.v34i04.6113>
15. Fan, D.P., Zhou, T., Ji, G.P., et al.: Inf-Net: automatic COVID-19 lung infection segmentation from CT images. *IEEE Trans. Med. Imaging* **PP**(99), 1–1 (2020). <https://doi.org/10.1109/TMI.2020.2996645>
16. Chen, X., Yao, L., Zhang, Y.: Residual attention u-net for automated multi-class segmentation of covid-19 chest CT images. (2020) [Online]. Available: <https://arxiv.org/abs/2004.05645>
17. Zhao, S., Li, Z., Chen, Y., et al.: SCOAT-Net: aa novel network for segmenting COVID-19 lung opacification from CT images. *Pattern Recogn.* **119**, 108109 (2021). <https://doi.org/10.1016/j.patcog.2021.108109>
18. Wang, G., Liu, X., Li, C., et al.: A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images. *IEEE Trans. Med. Imaging* **39**(8), 2653–2663 (2020). <https://doi.org/10.1109/TMI.2020.3000314>
19. Yan, Q., Wang, B., Gong, D. et al.; COVID-19 chest CT image segmentation—a deep convolutional neural network solution. (2020) [Online]. Available: <https://arxiv.org/abs/2004.10987>
20. Elharrouss, O., Subramanian, N., Al-Maadeed, S.: An encoder-decoder-based method for COVID-19 lung infection segmentation. (2020) [Online]. Available: <https://arxiv.org/abs/2007.00861>
21. Qiu, Y., Liu, Y., Li, S., Xu, J.: Miniseg: an extremely minimum network for efficient COVID-19 segmentation. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 6, (2021) 4846–4854. <https://ojs.aaai.org/index.php/AAAI/article/view/16617>
22. Pei, H.-Y., Yang, D., Liu, G.-R., et al.: MPS-net: multi-point supervised network for CT image segmentation of covid-19. *IEEE Access* **9**, 47144–47153 (2021). <https://doi.org/10.1109/ACCESS.2021.3067047>
23. Zhang, P., Zhong, Y., Deng, Y., et al.: CoSinGAN: learning COVID-19 infection segmentation from a single radiological image. *Diagnostics* **10**(11), 901 (2020). <https://doi.org/10.3390/diagnostics10110901>
24. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(11), 1254–1259 (1998). <https://doi.org/10.1109/34.730558>
25. Wang, F., Tax, D. M.: Survey on the attention based RNN model and its applications in computer vision. (2016) [Online]. Available: <https://arxiv.org/abs/1601.06823>
26. Jaderberg, M., Simonyan, K., Zisserman, A.: Spatial transformer networks. *Adv. Neural Inform. Process. Syst.* **28** (2015). <https://dl.acm.org/doi/abs/https://doi.org/10.5555/2969442.2969465>
27. Buades, A., Coll, B., Morel, J.-M.: A non-local algorithm for image denoising. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2. IEEE, pp. 60–65 (2005). <https://doi.org/10.1109/CVPR.2005.38>
28. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 7794–7803 (2018). <https://doi.org/10.1109/CVPR.2018.00813>
29. Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H.: Dual attention network for scene segmentation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 3146–3154 (2019). <https://doi.org/10.1109/CVPR.2019.00326>
30. Wang, F., Jiang, M., Qian, C., Yang, S., Li, C., Zhang, H., Wang, X., Tang, X.: Residual attention network for image classification. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 3156–3164 (2017). <https://doi.org/10.1109/CVPR.2017.683>
31. Huang, Z., Wang, X., Huang, L., Huang, C., Wei, Y., Liu, W.: Ccnet: Criss-cross attention for semantic segmentation. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 603–612 (2019). <https://doi.org/10.1109/TPAMI.2020.3007032>
32. Gao, P., Zheng, M., Wang, X., Dai, J., Li, H.: Fast convergence of detr with spatially modulated co-attention (2021) [Online]. <https://doi.org/10.48550/arXiv.2108.02404>
33. Huang, G., Zhu, J., Li, J., Wang, Z., Cheng, L., Liu, L., Li, H., Zhou, J.: Channel-attention U-Net: channel attention mechanism for semantic segmentation of esophagus and esophageal cancer. *IEEE Access* **8**, 122798–122810 (2020). <https://doi.org/10.1109/ACCESS.2020.3007719>
34. Zhao, B., Wu, X., Feng, J., et al.: Diversified visual attention networks for fine-grained object classification. *IEEE Trans. Multimed.* **19**(6), 1245–1256 (2017). <https://doi.org/10.1109/TMM.2017.2648498>
35. Mnih, V., Heess, N., Graves, A.: Recurrent models of visual attention. *Adv. Neural Inform. processing Syst.* **27** (2014). <https://dl.acm.org/doi/abs/https://doi.org/10.5555/2969033.2969073>
36. Liu, X., Xia, T., Wang, J. et al.: Fully convolutional attention networks for fine-grained recognition. (2016) [Online]. <https://arxiv.org/abs/1603.06765>
37. Zhao, X., Zhang, P., Song, F. et al.: D2a u-net: automatic segmentation of COVID-19 lesions from CT slices with dilated convolution and dual attention mechanism. (2021) [Online]. <https://arxiv.org/abs/2102.05210>
38. Zhou, T., Canu, S., Ruan, S.: Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism. *Int. J. Imaging Syst. Technol.* **31**(1), 16–27 (2021). <https://doi.org/10.1002/ima.22527>
39. Zhou, X., Xu, X., Liang, W., Zeng, Z., Yan, Z.: Deep-learning-enhanced multitarget detection for end-edge-cloud surveillance in smart IoT. *IEEE Internet Things J.* **8**(16), 12588–12596 (2021). <https://doi.org/10.1109/JIOT.2021.3077449>
40. Cremers, D., Osher, S.J., Soatto, S.: Kernel density estimation and intrinsic alignment for shape priors in level set segmentation. *Int. J. Comput. Vis.* **69**(3), 335–351 (2006). <https://doi.org/10.1007/s11263-006-7533-5>
41. Li, K., Tao, W.: Adaptive optimal shape prior for easy interactive object segmentation. *IEEE Trans. Multimed.* **17**(7), 994–1005 (2015). <https://doi.org/10.1109/TMM.2015.2433795>

42. Wang, H., Zhang, H.: Adaptive shape prior in graph cut segmentation. In: 2010 IEEE International Conference on Image Processing. IEEE, pp 3029–3032 (2010). <https://doi.org/10.1109/ICIP.2010.5653335>
43. Veksler, O.: Star shape prior for graph-cut image segmentation. In: European Conference on Computer Vision. Springer, pp 454–467 (2008). https://doi.org/10.1007/978-3-540-88690-7_34
44. Nosrati, M. S., Hamarneh, G.: Incorporating prior knowledge in medical image segmentation: a survey (2021) [Online]. Available: <https://arxiv.org/abs/1607.01092>
45. Lee, M.C.H., Petersen, K., Pawlowski, N., Glocker, B., Schaap, M.: TeTrIS: template transformer networks for image segmentation with shape priors. *IEEE Trans. Med. Imaging* **38**(11), 2596–2606 (2019). <https://doi.org/10.1109/TMI.2019.2905990>
46. Ravishankar, H., Venkataramani, R., Thiruvenkadam, S., Sudhakar, P., Vaidya, V.: Learning and incorporating shape models for semantic segmentation. In: International conference on medical image computing and computer-assisted intervention. Springer, pp 203–211 (2017). https://doi.org/10.1007/978-3-319-66182-7_24
47. Avendi, M.R., Kheradvar, A., Jafarkhani, H.: A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI. *Med. Image Anal.* **30**, 108–119 (2016). <https://doi.org/10.1016/j.media.2016.01.005>
48. Ngo, T.A., Lu, Z., Carneiro, G.: Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance. *Med. Image Anal.* **35**, 159–171 (2017). <https://doi.org/10.1016/j.media.2016.05.009>
49. Zhao, C., Xu, Y., He, Z., Tang, J., Zhang, Y., Han, J., Shi, Y., Zhou, W.: Lung segmentation and automatic detection of COVID-19 using radiomic features from chest CT images. *Pattern Recogn.* **119**, 108071 (2021). <https://doi.org/10.1016/j.patcog.2021.108071>
50. Rosenfeld, A., Pfaltz, J.L.: Sequential operations in digital picture processing. *J. ACM (JACM)* **13**(4), 471–494 (1966). <https://doi.org/10.1145/321356.321357>
51. Shih, F.Y., Wu, Y.-T.: Fast Euclidean distance transformation in two scans using a 3x3 neighborhood. *Comput. Vis. Image Underst.* **93**(2), 195–205 (2004). <https://doi.org/10.1016/j.cviu.2003.09.004>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Xiaoxin Wu received the B.E. degree from ChongQing Medical University, China, in 2013, the Doctoral degree in Zhejiang University, China, in 2019. Currently he is an attending physician in the First Affiliated Hospital, Zhejiang University School of Medicine. His current research focuses on the diagnosis and treatment of emerging infectious diseases.



Zhihao Zhang received the B.E. degree in Electrical Information Science and Technology from Nanjing XiaoZhuang University, Nanjing, China, in 2018. He is currently working toward the M.S. degree in Artificial Intelligence and Big Data from Shanghai University of Electric Power. His current research focuses on image processing, deep learning and defect detection.



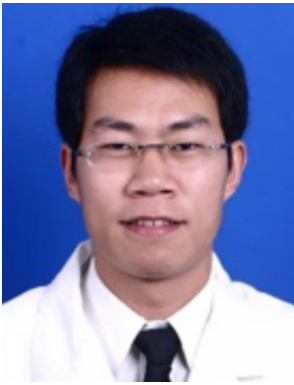
Lingling Guo received the B.E. degree from Zhejiang University, China, in 2011, the Ph.D. degree in Chemical Engineering from Alabama University, USA, in 2015. Currently she is an assistant professor in Zhejiang University of Technology. Her current research focuses on the industrial information and automation.



Hui Chen received the B.E. degree from Tianjin University of Traditional Chinese Medicine, China, in 2012. Currently he is a supervisor nurse in the First Affiliated Hospital, Zhejiang University School of Medicine. His current research focuses on the treatment of emerging infectious diseases.



Qiaojie Luo received the Ph.D. degree in Stomatology from Zhejiang University in 2015. Currently she is a dentist and science researcher in the Affiliated Stomatology Hospital of Zhejiang University. Her current research focuses on implantology and adhesive dentistry.



Bei Jin received the B.S. degree in Stomatology from Wenzhou Medical University, Zhejiang, China, in 2005. He is currently an attending physician in Taizhou Hospital of Zhejiang Province. His current research focuses on Oral and Maxillofacial Surgery.



Fangfang Lu received the Ph.D. degree in control theory and control engineering from Shanghai Jiaotong University, Shanghai, China, in 2013. She is currently an assistant professor in Shanghai University of Electric Power. Her current research focuses on medical image processing, machine learning, pattern recognition and image quality assessment.



Weiyang Gu received the M.S. degree in Stomatology from Wenzhou Medical University, Zhejiang, China, in 2017. She is currently an attending physician in Taizhou Hospital of Zhejiang Province. Her current research focuses on Prosthodontics and Oral Implantology.



Jingjing Chen received the Ph.D. in Computer Science from Hong Kong Baptist University in 2016. He was a research fellow in Blockchain with the Sunyard System Engineering Co. Ltd. during 2017–2019, and research fellow with the School of Traditional Chinese Medicine of Hong Kong Baptist University during 2019–2020. Currently he is a research fellow with the school of Economics of Fudan University. He also serves as research fellow with Fudan-Stanford China Institute for Financial Technology and Security. His research interests include: Blockchain, e-Government and Enterprise Information System.