



Research article

Dmbg-Net: Dilated multiresidual boundary guidance network for COVID-19 infection segmentation

Zhenwu Xiang, Qi Mao*, Jintao Wang, Yi Tian, Yan Zhang and Wenfeng Wang

School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

* **Correspondence:** Email: asdenglish@126.com.

Abstract: Accurate segmentation of infected regions in lung computed tomography (CT) images is essential for the detection and diagnosis of coronavirus disease 2019 (COVID-19). However, lung lesion segmentation has some challenges, such as obscure boundaries, low contrast and scattered infection areas. In this paper, the dilated multiresidual boundary guidance network (Dmbg-Net) is proposed for COVID-19 infection segmentation in CT images of the lungs. This method focuses on semantic relationship modelling and boundary detail guidance. First, to effectively minimize the loss of significant features, a dilated residual block is substituted for a convolutional operation, and dilated convolutions are employed to expand the receptive field of the convolution kernel. Second, an edge-attention guidance preservation block is designed to incorporate boundary guidance of low-level features into feature integration, which is conducive to extracting the boundaries of the region of interest. Third, the various depths of features are used to generate the final prediction, and the utilization of a progressive multi-scale supervision strategy facilitates enhanced representations and highly accurate saliency maps. The proposed method is used to analyze COVID-19 datasets, and the experimental results reveal that the proposed method has a Dice similarity coefficient of 85.6% and a sensitivity of 84.2%. Extensive experimental results and ablation studies have shown the effectiveness of Dmbg-Net. Therefore, the proposed method has a potential application in the detection, labeling and segmentation of other lesion areas.

Keywords: computer-aided detection; dilated convolution; multi-scale supervision; COVID-19 segmentation

1. Introduction

The ongoing pandemic poses significant challenges to public health systems and societies at large. The global tally of confirmed coronavirus disease 2019 (COVID-19) cases has reached a staggering 770,875,433 as of September 27, 2023 [1]. This figure represents the cumulative number of individuals who have tested positive for the virus across various countries and regions worldwide. The necessity to develop machine-based tools for detecting COVID-19 signs in diagnostic imagery has arisen due to the worldwide spread of the pandemic [2,3]. The COVID-19 outbreak has caused a global public health crisis with unprecedented disruption and long-lasting effects [4]. Moreover, nonlaboratory evaluations, such as computer-aided analysis of chest radiographs (X-ray) or computed tomography (CT) scans, have been implemented to scrutinize the lungs for signs of COVID-19 [5]. Compared to X-ray, CT screening is widely preferred due to its advantages and three-dimensional view of the lungs [6,7]. CT enhances the efficiency of medical image analysis conducted by physicians, strengthens their clinical image perception skills and improves patient treatment rates.

Although there have been advancements in the development of intelligent diagnostic systems and treatment methods for COVID-19, numerous challenges still need to be addressed [8]. First, the presence of diverse morphological variances and the varied positioning of infected areas in lung CT images significantly impact feature analysis and information extraction, especially object boundaries and small targets. Second, infected areas have a wide range of infection characteristics, and the images have low contrast between normal tissue and infected lesions, which causes difficulty in segmenting unclear structural boundaries.

Benefiting from the rapid advancement of deep learning techniques [9,10]. Rani et al. [11] successfully addressed a potential obstacle to feature extraction in chest X-rays by implementing bone suppression and lung segmentation preprocessing techniques. These methods not only overcome the obstacle but also ensure the preservation of the highest possible spatial information and resolution. Fan et al. [12] developed the Inf-Net network by employing reverse attention and implementing semi-supervised learning. Additionally, several works have tried to model the global context from the perspectives of boundary information using an encoder-decoder architecture [13,14]. However, these methods fail to offer the flexibility needed for adjusting the receptive fields to match the diverse scales of the intended objects. The fixed receptive fields of the feature map hinder their ability to dynamically accommodate the varying scales of the targets. In contrast, a few studies have shown that learning the appropriate receptive fields and achieving an optimal configuration can enhance the perception of semantic cues associated with various objects [15,16].

Compared with most computer-aided detection/diagnosis methods, several studies have been conducted on methods to reduce a high false-positive rate [17–19]. More recently, object segmentation methods have leveraged global appearance models to accurately identify and delineate target regions, encompassing both foreground and background. Yan et al. [20] designed a multiscale cascading deep belief network via calculating the Fourier spectrum to capture multiple scale characteristics. The diagnosis network acquires richer and distinctive feature information from the original signal [21,22]. Bougourzi et al. [23] proposed a hybrid loss function to address the challenge of segmenting COVID-19 infection pixels with blurry boundaries. Furthermore, most prior research has concentrated on precision in specific areas while disregarding the boundaries [24–26]. Fan et al. [27] integrated boundary features with the multi-layer output features to generate the ultimate output. Additionally, when considering the specific tasks associated with the segmentation of medical images, gaining a

deeper comprehension of the intrinsic relationships that exist among the individual pixels becomes imperative [28,29].

The infected areas in lung CT images are usually scattered with complex backgrounds [30,31]. The edge detection method uses local gradient representation to identify object boundaries and then separates the closed-loop region into objects [32]. This idea is used in network design. However, directly transmitting the coarse low-level features may lead to redundancy interference. Several research studies have provided evidence that edge information can serve as a valuable constraint in guiding the extraction of features [33,34]. Most of the previous studies focused on regional accuracy but neglected boundary quality. In fact, clear regions and boundaries hold significant importance in the segmentation of COVID-19 infection. This study incorporates explicit modelling of edge information within the network to effectively utilize edge cues.

The main contributions of this study are as follows: 1) A novel model of the dilated multiresidual boundary guidance network (Dmbg-Net) is proposed, which can learn more boundary details of lung lesions. In this method, a dilated residual block is used to expand the respective field. To fully leverage the contextual information, the method allows for exponential enlargement of the receptive field while maintaining resolution and intact coverage. To effectively acquire the low-contrast and boundary areas, edge information is generated by an edge-attention guidance preservation (EGP) block, which provides detailed structured information. 2) A loss function is designed for amplifying the flow of gradients that effectively amplifies the saliency of the pertinent regions and suppresses that of the irrelevant regions. In the encoding path, the Dmbg-Net encoder-decoder framework is designed for ResNet-50, which causes the network to converge quickly. The results of ablation studies verified the effectiveness and accuracy of the proposed model. After comparative experiments, the proposed method outperforms the state-of-the-art convolutional neural networks (CNNs), including U-Net, UNet++ and GFNet.

The remainder of this paper is organized as follows. Section 2 provides an overview and detailed component description of the proposed approach. Section 3 illustrates the experimental validation and evaluation of COVID-19 infection segmentation tasks using the proposed model. Section 4 discusses the shortcomings between the proposed model and the existing segmentation network. Finally, the conclusion is presented in Section 5.

2. The proposed method

In this section, we begin by providing a detailed overview of our Dmbg-Net, focusing on its network architecture. Subsequently, we cover various network components, including the dilated residual block, edge-attention guidance preservation block and edge aggregation pile block. Furthermore, we clarify how we utilize a receptive field strategy to enhance segmentation accuracy and employ a progressive multi-scale supervision strategy to effectively segment fine-grained lung lesions. Finally, we introduce the loss function utilized by our model.

2.1. Structure of Dmbg-Net

The structure and details of the proposed Dmbg-Net are shown in Figure 1. The framework is based on an encoder-decoder architecture. It comprises five primary components: A feature encoder, dilated residual block, edge-attention guidance preservation (EGP) block, feature decoder and edge

aggregation pile (EP) block. We have provided comprehensive details of the dilated residual block in Figure 1 for clarity. The model uses ResNet as a backbone network. To further enhance the features learned from ResNet, the dilated residual block aims to enhance and adapt the receptive fields to capture contextual information at the most suitable scale. Additionally, the EGP block is specifically designed to effectively handle shape edges, particularly semantic boundaries. Finally, the EP module has been developed to progressively guide the integration of multilevel feature maps, enabling the capture of richer semantic information.

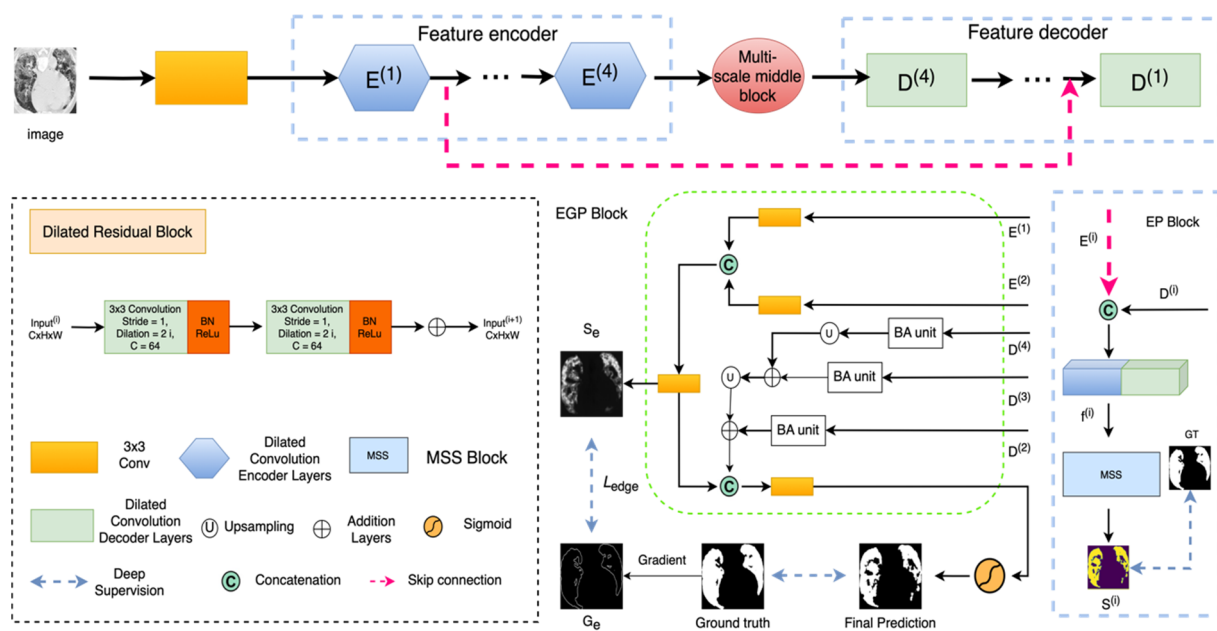


Figure 1. Architecture of Dmbg-Net, which includes the dilated residual block, edge-attention guidance preservation (EGP) block and edge aggregation pile (EP) block.

CT images are monochromatic, which means that the pixel values in all three channels are identical when the image is uploaded. Depending on the nature of the image, a monochromatic input may possess three channels (RGB image). We employ a 3×3 convolutional module with a stride of 1 and padding of 1 to process the input image. Subsequently, batch normalization is applied and activated using the rectified linear unit (ReLU). This approach can effectively process monochromatic CT images while preserving high accuracy in medical image segmentation tasks.

The feature extraction stream of our Dmbg-Net architecture consists of a stack of three convolutional layers: A 1×1 convolutional layer, a 3×3 convolutional layer and another 1×1 convolutional layer. On the one hand, we use a skip connection to allow information to flow directly from one layer to another without passing through intermediate layers. The model has the capability to generate high-level features that are specific to each class. On the other hand, adding the Maxing layer reduces model overfitting and improves its ability to capture global features in the input image. The Dmbg-Net architecture is divided into five layers, during which a pooling layer is linked between each of these stages. The first 2 layers consist of three cascaded convolutional layers, each with a different function. The first layer comprises a standard convolutional layer with a kernel size of 3×3 . This layer employs a series of filters to extract low-level features. The second layer is a depthwise

separable convolutional layer that applies separate filters to each channel of the low-level feature map. Each feature decoder includes a 1×1 convolution, a 3×3 transposed convolution and a 1×1 convolution. Therefore, this model combines hierarchical features from all the convolutional layers into a holistic framework. This holistic framework allows for automatic learning of all parameters, eliminating the need for additional overhead or increased complexity.

Given the significance of semantic relations and boundary constraints in the segmentation task, we present a comprehensive model for COVID-19 infection segmentation. The proposed model emphasizes the decoding process by incorporating global semantics and boundary constraints. In each feature encoder, we jointly consider the dilated convolutions, residual learning and edge guidance module to offer ample multiscale context for feature decoding. The traditional approach for segmenting GGOs relies on either grey information or the Euclidean distance, often leading to erroneous GGO segmentation. We employ explicit supervision to compare the generated boundary map with the boundary ground truth acquired through the boundary extractor to ensure the efficacy of learning. Specifically, it tackles the challenge of scattered lesion positioning and irregular issues. By leveraging the attributes derived from the feature decoder, we effectively transfer the representations of edge attention to the layers at a higher level to improve the final results.

2.2. Dilated residual block

To ensure accurate and complete regions, we propose a dilated residual block, which is made from dilated convolution and one residual path. Dilated convolutions, sometimes called atrous convolutions or convolutions with holes, have gained substantial popularity in deep learning. The convolutional kernel weight is set to zero at corresponding locations, which is given by Eq (1):

$$f(x_i) = \sum \theta_{k,r}^c * x_i^c \quad (1)$$

where $*$ denotes the convolution operation and $f(x_i)$ convolves the input feature map x_i . A dilated convolution $\theta_{k,r}$ with a kernel size of k and dilation rate r determines the pattern for these holes. By employing this approach, the receptive field is effectively enlarged without the need to introduce any additional network parameters that require learning.

In state-of-the-art medical image algorithms, dilated convolutions are used to improve the receptive field while maintaining spatial information. Different from the existing methods, we utilize different dilated convolution rates to learn information about the contents of various high-level features. The dilated residual block comprises two sets of 3×3 convolutions, as in Eq (2):

$$f(x_i) = \Phi^C(\sum \theta_{k,2 \times i}^c * x_i^c) + x_i \quad (2)$$

where $*$ denotes the convolution operation and $f(x_i)$ convolves the input feature map x_i . A dilated convolution $\theta_{k,2 \times i}$ with a kernel size of k and dilation rate of $2 \times i$ (i denotes the model feature decoder number). Φ^C denotes a 3×3 convolution with C channels, followed by the BatchNorm layer and a ReLU activation function.

Specifically, we employ an encoder architecture that consists of three convolution layers with dilation rates of 2, 4 and 8. As shown in Figure 1, this design results in an effective kernel size of 5, 9 and 17 for each convolution operation at their respective levels. The feature maps of the dilated residual block are combined through a residual connection. This aggregation process ensures seamless

integration of the feature maps, enhancing the overall performance of the model. The use of dilated convolutions in the encoder enables the network to capture contextual information from a larger receptive field while preserving spatial resolution, which is important for contextual information processing. After each encoder layer, a max pooling operation with a stride of 2 and a 2×2 kernel is applied. This operation decreases the spatial dimensions of the feature maps by half while preserving their essential features. The feature maps processed by these dilated residual blocks have sizes of 88×88 , 176×176 and 352×352 .

The bottleneck convolution blocks are depicted in Figure 1. These blocks execute the combined residual transformation. The bottleneck is employed to compel the model to learn a compressed representation of the input data. This compressed representation should exclusively encompass the vital and valuable information needed for the decoder to restore the input. In the initial U-Net architecture, the bottleneck comprises two convolutional layers with a size of 3×3 , both activated by the rectified linear unit (ReLU) function. As shown in Figure 2, the bottleneck of the multi-scale middle block includes avgpooling, convolutional units and maxpooling. Each convolution includes a BatchNorm layer and a ReLU activation function. The bottleneck employs a dilation rate of 8 while maintaining a stride of 1 and padding of 8.



Figure 2. The bottleneck of the multi-scale middle block.

This architecture facilitates efficient feature extraction from medical images of multiple scales while preserving spatial resolution. By utilizing dilated convolutions in the encoder, we can expand the receptive field of each layer without compromising the spatial resolution. By incorporating feature encoder layers with distinct receptive field sizes, Dmbg-Net facilitates the acquisition of multiscale information at the low-level and object level. This unique attribute of our network enhances its ability to capture more boundary information.

2.3. Edge-attention guidance preservation block

The foundation of our framework takes significant inspiration from the research of [27]. The authors assumed that the utilization of edge detection and region segmentation approaches could be beneficial. To optimize the feature extraction process in Dmbg-Net, we design an edge-attention guidance preservation (EGP) block that fuses the features of each layer by leveraging the concept of edge detection. These representations are used to guide our proposed approach and improve the precision of the final outcomes. The EGP block consists of five inputs, two outputs and a boundary aggregation (BA) unit component. The encoder layers E1 and E2 enhance the boundary information of the high-level decoder features D2, D3 and D4. Specifically, the focus of this block is acquiring boundary context, conserving the distinctive features of local edges within the E1 and E2 layers, and designing BA units to aggregate multiscale side outputs from the decoding layer.

To emphasize the significant boundaries in feature decoding, modifications are made to the U-Net network by incorporating edge attention representations. We devise the BA module to incorporate

the guidance of boundaries for low-level features in the integration of features. Details of this module are shown in Figure 3. In the BA module, the initial step involves utilizing global average pooling to consolidate the overall contextual details of the inputs. Subsequently, two 1×1 convolutional layers, each employing distinct nonlinear activation functions, are employed to assess the significance of each layer and produce the weights across the channel dimension. By utilizing sigmoid functions, we can generate more distinctive features in the output, thereby enhancing the representativeness of the results.

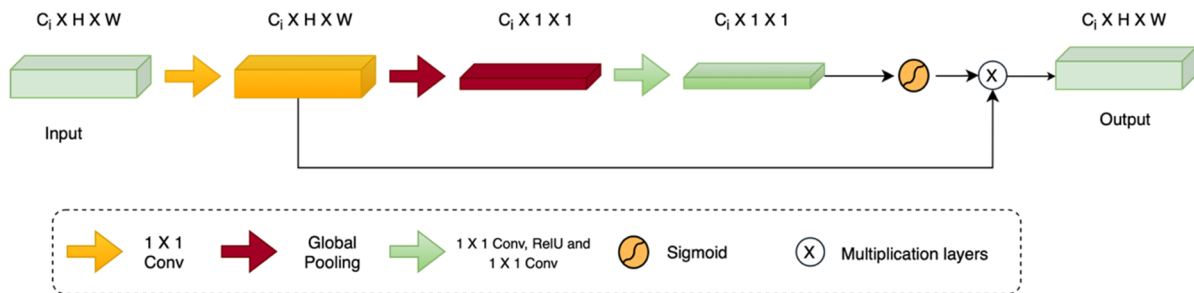


Figure 3. The structure of the boundary aggregation unit.

Several studies have demonstrated that utilizing edge data can offer valuable limitations to direct the extraction of features for segmentation [12,13]. Consequently, considering that the proposed model incorporates low-level features, effectively retaining ample edge data, these low-level features are incorporated into our analysis. The resolution of the output feature from the BA unit is matched by upsampling the outputs of the EGP block, feeding them to the 1×1 and 3×3 convolutional layers and concatenating them. Next, we evaluate the dissimilarity between the generated edge map and the ground truth (GT) edge map. This evaluation is performed using the standard binary cross-entropy (BCE) loss function. The BCE loss of edge map can be written as Eq (3):

$$\mathcal{L}_{edge} = - \sum_{x=1}^w \sum_{y=1}^h [G_e \log(S_e) + (1 - G_e) \log(1 - S_e)] \quad (3)$$

where (x, y) are the coordinates of each pixel in the predicted edge map S_e and edge ground-truth map G_e . The w and h parameters represent the dimensions of the respective maps. While obtaining the feature map, this study calculates the gradient of the ground truth of the input image to obtain the edge ground truth of the boundary. The edge-attention guidance preservation block serves as a mechanism for transferring edge information from early encoding layers to high-level layers, where it can be combined with other features to guide segmentation. Using this approach, the model can leverage the strengths of edge detection and attention mechanisms to enhance the efficiency of the image segmentation network.

2.4. Edge aggregation pile block

We design a progressive multiscale supervision strategy for edge aggregation pile (EP) blocks. Details of this module are shown in Figure 1 (right). Incorporating skip connections and leveraging the concatenation of feature decoders, the EP block consists of two inputs and one output and a component of a multiscale supervision (MSS) block, which are described in detail below. As illustrated

in Figure 4, the EP blocks consist of two convolutional units with a size of 3×3 . These units have a dilation rate of $2 \times i$, where i represents the number of feature decoders in the model. The stride is set to 1, and the padding is adjusted to $2 \times i$. Output $C \times H \times W$ aims to create segmentation outcomes by utilizing the feature encoder and feature extractor. By incorporating skip connections, the feature decoder can acquire additional information from the encoder, compensating for the information loss caused by pooling and convolutional operations. These techniques enable efficient feature extraction at multiple scales while preserving spatial information and reducing overfitting. Next, a 3×3 convolutional layer and a bilinear upsampling module are employed, and then a 1×1 convolutional layer is applied, followed by a sigmoid activation function. The proposed model benefits from deep multiscale supervision, as it incorporates channels with various sizes of 176, 88 and 44. These modules allow Dmbg-Net to incorporate edge information at multiple scales and improve its ability to segment complex objects with irregular shapes.

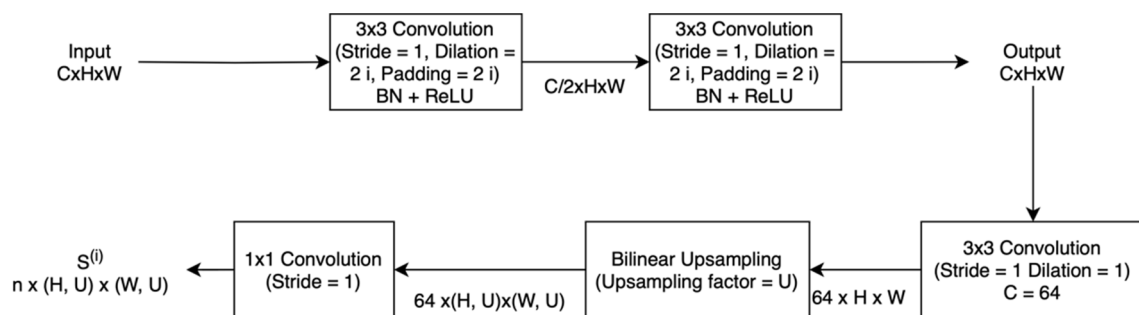


Figure 4. The structure of the edge aggregation pile block.

With the constraint of deep supervision, we can acquire enhanced feature mapping and generate final predictions. By simultaneously considering the information pertaining to the three high-level attributes D4, D3 and D2, a unified output feature resolution is achieved. Effective guidance helps the network learn missing components and intricate aspects of the perimeter, resulting in more comprehensive and accurate predictions. Hence, optimizing the gradient flow throughout the various layers of the model during the backpropagation process makes it possible to achieve faster convergence.

2.5. Loss function

The loss function L_{seg} is defined as a combination of the weighted IoU loss function L_{IOU}^w and the weighted binary cross entropy (BCE) loss function L_{BCE}^w . The loss is calculated as follows:

$$L_{\text{seg}} = L_{\text{IOU}}^w + \lambda L_{\text{BCE}}^w \quad (4)$$

where λ represents the weight (assigned a value of 1 in this study). The two components of L_{seg} offer efficient global (image-level) and local (pixel-level) guidance. This approach ensures that the obtained results are reliable. Deep supervision involves computing the loss for both the hidden layers and the overall model, subsequently refining the model by leveraging the aggregated loss value. This study

implements deep supervision for four outputs ($L_{seg}^i, i = 1, 2, 3, 4$) and the boundary loss \mathcal{L}_{edge} . The corresponding total loss is given as follows:

$$\mathcal{L}_{total} = \sum L_{seg}^i + \mathcal{L}_{edge} \quad (5)$$

3. Experiments and result analysis

3.1. Image database and evaluation indicators

The labelled CT images are taken from the COVID-19 CT segmentation dataset, composed of 100 axial CT images collected by the Italian Society of Medical and Interventional Radiology [35]. With only 98 images available, this dataset is the first open-access resource for segmenting lung infections caused by COVID-19. We extracted 920 high-quality CT images from the COVID-19 CT collection dataset [36], which comprises twenty 3D CT volumes obtained from different COVID-19 patients. To better train the model and obtain a relatively sufficient training sample, two common datasets were combined to obtain 1018 high-quality CT images, which were further divided into 718 training images and 300 test images.

We utilize the Dmbg-Net architecture described in this study for the infected region experiment, incorporating dilated convolutions. We use an optimal receptive field strategy to improve Dmbg-Net, primarily based on an encoder-decoder network. This strategy is contrasted with two well-known segmentation models, U-Net and U-Net++, and the most recent model GFNet. The source code of the proposed method is available at <https://github.com/pure-sky/Dmbg-Net>.

Based on the studies by Fan et al. [12] and Fan et al. [27], we use six widely adopted metrics, i.e., the Dice similarity coefficient (DSC), sensitivity (Sen.), structure measure (S_α), enhance-alignment measure (E_ϕ), mean absolute error (MAE) and precision (Prec.). The formula provided below represents a novel metric that evaluates the local and global similarity of two binary graphs.

$$E_\phi = \frac{1}{w \times h} \sum_x^w \sum_y^h \phi(S_p(x, y), G(x, y)) \quad (6)$$

where w and h represent the width and height of the ground-truth map G , respectively, and (x, y) represents the coordinates of each pixel in G . The symbol ϕ is an enhanced alignment matrix.

$$S_\alpha = (1 - \alpha) * S_o(S_p, G) + \alpha * S_r(S_p, G) \quad (7)$$

where α is the balance factor used to control object-aware similarity S_o and region-aware similarity S_r . In this study, we employ the identical metric value as the original text by utilizing the default setting ($\alpha = 0.5$). The primary purpose of the structure measure S_α is to assess the degree of structural similarity between the prediction map and the ground-truth mask. The MAE metric measures the pixel-level error between S_p and G , which is written as:

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

where w and h represent the width and height of the ground-truth map G , respectively, and (x, y) represents the coordinates of each pixel in G . Among these indicators, Sen. and S_α can reflect the segmentation integrity, and the DSC, Prec., E_ϕ and MAE can evaluate the overall performance. To maximize performance, except for the MAE metric, the numerical value of these measurements is higher.

3.2. Implementation details and experimental results

The proposed model is configured with the PyTorch toolbox and trained on a single Quadro RTX 6000 (24 GB) GPU. To ensure fairness in the training process, we apply a consistent resizing technique to all input data, resulting in a standardized size of 352×352 . Our training approach for Dmbg-Net incorporates a multi-scale strategy [12,34]. We begin by resampling the training images with varying scaling ratios, such as 0.75, 1 and 1.25. Subsequently, we employ the resampled images to train Dmbg-Net, thereby enhancing the overall generalization capabilities of the proposed model. The batch size is set to 4, and the Epoch is adjusted to 200. The training utilizes the Adam optimizer, with a learning rate of $3e-4$.



Figure 5. Segmented results by different methods. (a) Input images. (b) Ground truth. (c)–(l) Results of the proposed method, U-Net, UNet++, Attention-UNet, FCN, Inf-Net, GFNet, BCS-Net and BS-Net, respectively.

To assess the segmentation capabilities of the Dmbg-Net model proposed in this study, we employ it for COVID-19 infection segmentation. We conduct a comparative analysis against the classical

algorithms currently in use. Figure 5 illustrates the visual comparison among the proposed model and other state-of-the-art methods. The original CT images are presented in the first column, while the second column displays the evaluation standard, which represents the manual marking performed by radiologists. From left to right are the results of the proposed method, U-Net [9], UNet++ [37] Attention-UNet [38], FCN [10], Inf-Net [12], GFNet [27], BCS-Net [13] and BS-Net [14]. The proposed method demonstrates superior advantages of accuracy, completeness and sharpness compared to other techniques. For instance, in the first image, classical medical image segmentation networks, including U-Net, UNet++ and Attention-UNet, frequently fail to effectively mitigate the interference caused by background regions, such as the areas between the left lung and the background. The proposed method successfully overcomes this challenge, ensuring precise and comprehensive segmentation results.

The COVID-19 segmentation network exhibits superior performance in terms of recognition. However, the existing approaches, including Inf-Net, GFNet and BCS-Net, are unable to completely mitigate these interferences. For instance, the suppression of the area above the right lung proves to be inadequate. The proposed Dmbg-Net preserves the structural boundary of the desired area, even in cases where the structural boundary of the image is unclear or exhibits textural variations. Figure 6 is the box plot comparing Dmbg-Net with other state-of-the-art techniques. The training loss curve shows that the proposed method achieves fast convergence. The findings demonstrate that our Dmbg-Net outperforms all existing methods and exhibits superior stability and robustness.

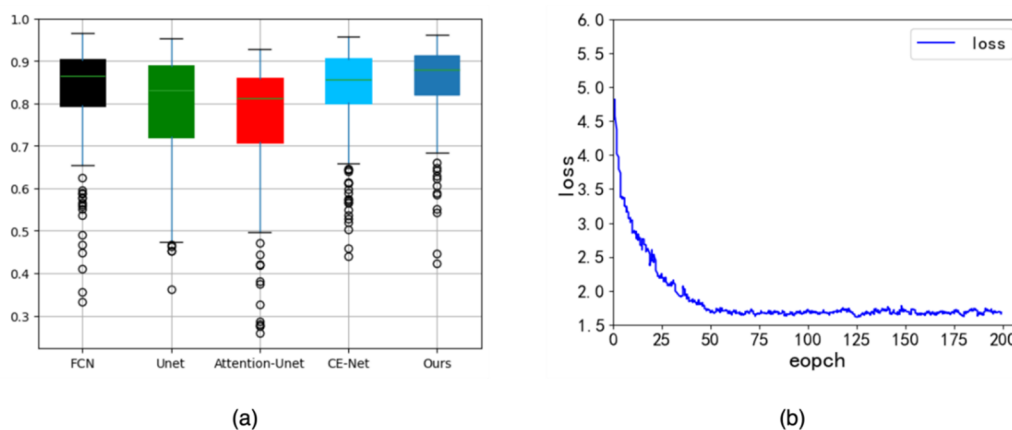


Figure 6. Comparison of different methods by the box plot. (a) The box plot includes different approaches of the FCN, UNet++, Attention-UNet, CE-Net and the proposed Dmbg-Net. (b) The training loss curve of the Dmbg-Net.

Our proposed approach demonstrates superior performance in these areas and showcases a heightened capability to identify intricate details. The eighth row of images validates the accuracy detection of infected area boundaries and effectively suppresses extraneous background noise. In addition, the proposed method has a more complete structure and sharper boundaries. In the second image, novel approaches, including GFNet and BSNet, fail to consistently and comprehensively identify infected lesions in the lower section of the left lung. Conversely, our innovative method successfully detects these regions with precision and completeness. Furthermore, compared with the currently existing methods, including UNet++ and Attention-UNet, our method results in clear

boundaries, which are crucial for both academic exploration and practical applications in the early identification of lung lesions.

In CT images, the segmentation results of the U-Net model are generally the least satisfactory, with relatively rough boundaries. On the other hand, FCN-8s demonstrate varying levels of image oversaturation. According to the experimental findings, the segmentation outcomes achieved by Dmbg-Net surpass those of other algorithms, including UNet++. This novel approach exhibits superior segmentation performance and enhances the overall image quality. To ensure an objective assessment, conducting a quantitative analysis of the segmentation results is crucial. This approach helps mitigate the potential biases introduced by subjective factors, ensuring a more objective assessment. The quantitative comparisons are reported in Table 1.

Table 1. Comparison of different methods on the segmentation of the COVID-19 dataset.

Methods	DSC	Sen.	Prec.	MAE	E_{ϕ}	S_{α}	FLOPs	Params
U-Net	0.796	0.763	0.871	0.021	0.891	0.789	5.660 G	1.814 M
Inf-Net	0.831	0.827	0.853	0.016	0.951	0.865	13.922 G	33.122 M
GF-Net	0.849	0.832	0.875	0.015	0.958	0.874	50.849 G	18.131 M
UNet++	0.774	0.752	0.846	0.021	0.885	0.831	65.938 G	9.163 M
Attention-UNet	0.803	0.832	0.812	0.021	0.922	0.864	31.730 G	8.727 M
FCN	0.835	0.824	0.868	0.020	0.952	0.854	48.209 G	18.644 M
CE-Net	0.834	0.814	0.869	0.017	0.915	0.860	16.828 G	29.003 M
BCS-Net	0.852	0.865	0.847	0.015	0.947	0.875	28.068 G	44.823 M
BS-Net	0.851	0.849	0.868	0.014	0.959	0.871	86.596 G	43.986 M
The Proposed	0.856	0.842	0.879	0.014	0.964	0.880	170.127 G	32.422 M

The Dmbg-Net model outperforms the other models across all five evaluation metrics when tested on the dataset, clearly demonstrating its superiority. Our proposed model's Dice coefficient reaches 85.6%, with a precision of 87.9%, and the segmentation effect is relatively more consistent. For example, compared to Inf-Net, GFNet has a percentage gain of 2.17% in DSC scores, while the proposed method has a percentage gain of 3.01%. In terms of E_{ϕ} , peak performance is attained. More precisely, the proposed method exhibits an 8.25% enhancement in performance compared to UNet++ and a 3.90% improvement compared to Attention-UNet. The proposed Dmbg-Net framework demonstrates superior performance in the segmentation method for COVID-19 infection. Moreover, compared to the runner-up method BCSNet, there is a noticeable increase of 3.78% in precision percentage. In terms of quantitative evaluations, our detection capabilities are generally superior, ensuring a high level of accuracy in identifying lung infections.

3.3. Fine-grained segmentation in a multiclass dataset

Our objective of this research is to develop a framework for segmenting lung infections associated with COVID-19. These lung infection areas have visible imaging manifestations caused by ground glass opacity (GGO), consolidation and pulmonary fibrosis. Segmenting infection regions in multiple classes will undoubtedly offer additional support for enhancing the assisted diagnosis. This study utilizes the dataset provided by Zhang et al. [39], which comprises 150 CT scans. It includes 750 slices featuring a lung field and templates for GGO and consolidation segmentation. In this study, we use 500 slices for the purpose of training, while 250 slices are allocated for testing. The proposed model demonstrates satisfactory segmentation results for GGOs and consolidation, which are shown in Table 2. These outcomes stand in contrast to the findings of Enet in Paszke et al. [40] and the two existing methods mentioned in Zhao et al. [32].

Table 2. Quantitative analysis of various networks for the segmentation of ground glass opacities and consolidations in multiclass datasets.

Methods	IOU (%)	
	GGO	Consolidation
U-Net	48.38	55.15
UNet++	51.40	58.23
Enet	51.54	53.99
SCOAT-Net	52.32	66.29
The Proposed	57.45	62.44

Table 2 demonstrates that when compared to alternative approaches, the proposed approach achieves significant performance in segmenting fine-grained infection areas, with a higher IOU for GGOs. The proposed method achieves an IOU of 57.45% for GGO segmentation, which is 9.81% higher than that of the second-ranked SCOAT-Net at 52.32%. Segmenting lung infection by multiclass segmentation poses a significant challenge due to the subtle variations in imaging manifestations observed between GGOs and consolidation. Conversely, the initial design of Dmbg-Net does not cater explicitly to consolidation segmentation but rather aims to identify abnormal regions within the lungs. Nevertheless, the proposed model consistently achieves high accuracy in segmenting both single-object and multi-class infection areas. Given the robustness of the segmentation method in handling variations in infection datasets, it is evident that the proposed model exhibits robustness and strong generalization capabilities.

3.4. Ablation study

Numerous ablation experiments are carried out to validate the performance and effectiveness of the dilated residual block and the EGP block. These experiments are conducted with the aim of examining the key components of our proposed model. The results are given in Table 3 and Figure 7. First, we examine the impact of the dilated residual block. The primary purpose of incorporating

dilated convolution layers is to enhance crucial spatial positional features and establish correlations between the various pixels. This approach guarantees improved precision and comprehensive segmentation capabilities. The absence of the dilated residual block results in a decline across all six metrics, with a particularly significant drop observed in the DSC score in Table 3.

Table 3. Quantitative evaluation results of Dmbg-Net.

Variations	DSC	Sen.	Prec.	MAE	E_{ϕ}	S_{α}
Proposed	0.856	0.842	0.879	0.014	0.964	0.880
W/o dilated residual block	0.837	0.825	0.866	0.019	0.877	0.863
W/o edge-attention guidance preservation block	0.848	0.825	0.885	0.015	0.903	0.873
W/o edge aggregation pile block	0.852	0.832	0.884	0.024	0.824	0.816

Upon the exclusion of the dilated residual block, some small infections in the top left of the image are not detected, and there are unsuppressed interference noises in the left lung of the first image in Column d. In the right region of the second image, there are visible instances of incorrect results. In the bottom left corner of the third image, there is a large false segmentation area compared to dilated convolution. The effectiveness of the dilated residual block is evident in these examples. Second, to assess the effectiveness of the edge-attention guidance preservation block, we attach them to the core network without establishing any connections between the boundary and the network. The final output of the network is obtained by utilizing characteristics from the second decoding phase, primarily focusing on the subsequent aspects. It maximizes the utilization of the attributes of the initial encoder layer, including E1 and E2, to complement boundary details for the higher-level decoder features. This dramatically outperforms the base network, proving that edge information is vital to segmentation.

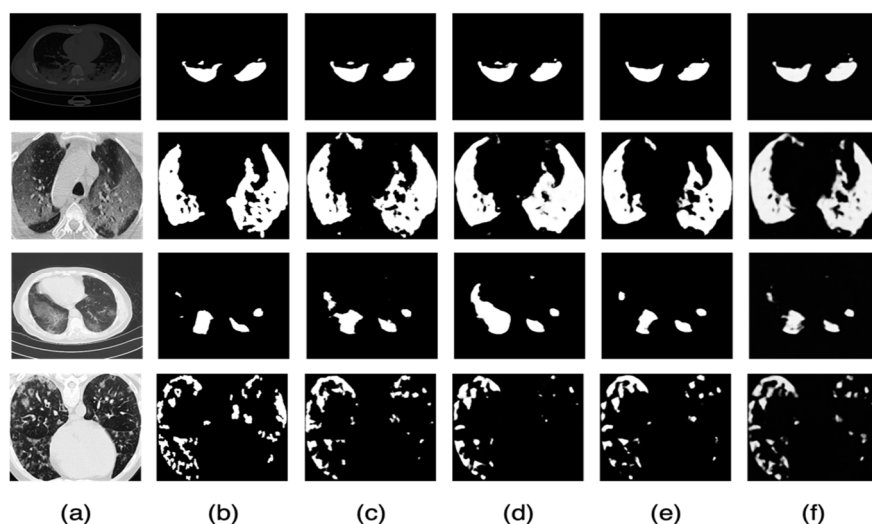


Figure 7. Comparing the visual aspects of various distinct modules of Dmbg-Net. (a) Images, (b) ground truth, (c) Dmbg-Net, (d) w/o dilated residual block, (e) w/o edge-attention guidance preservation block and (f) w/o edge aggregation pile block.

As shown in Figure 7, the first image in Column e has an obvious boundary over the segmentation, while the right area of the second image has blurred boundaries and obvious false detections. In the upper right section of the third picture, there are unsuppressed interference noises. When the foundational network does not include the EGP block, the DSC, MAE and E_{ϕ} values are 84.8%, 1.5% and 90.3%, respectively. When we attach the EGP block, the values are 85.6%, 1.4% and 96.4%, respectively. The fourth image w/o the EGP block in Figure 7 shows the lower portion of the left lung; the infection GT is mostly discrete and heavily dependent on the learning of the model for boundary information, and the infection spreads within the upper half of the lung. While the Dice coefficient of the w/o edge aggregation pile block is close to that of the proposed model for overall segmentation, the right lung region of the fourth image in Column f is missed in our global segmentation model. The evident improvements in performance for each of the three metrics demonstrate the effectiveness of the proposed EGP block. Through the integration of multiscale supervision, the segment of scattered and inconspicuous lesions is significantly improved by effectively leveraging contextual information.

4. Discussion

The distribution of infection is scattered due to the different GGO infection areas of the lungs and is distributed differently across the dataset. Background noise and low contrast make it challenging to detect edge information with traditional edge detection techniques. Dmbg-Net adds an edge-attention guidance preservation block to target regions for feature extraction and expands the receptive field by a dilated residual block while extracting multiscale contextual information. To evaluate the proposed models comprehensively, we implement the segmentation of two datasets obtained from publicly available medical datasets. Dmbg-Net performs better than the other models in identifying the precise borders and three assessment indicators.

In this study, we employ the dilated residual block architecture as a substitute for the conventional convolutional operations found in U-Net. By utilizing Dmbg-Net, we effectively augment the resolution of feature maps within the deeper layers, thereby enhancing the receptive fields of the input features. Additionally, the number of model parameters is slightly increased, effectively adjusting the requirements of the restricted target distribution in the affected area. The effect of this change is not statistically significant. In contrast to Bose et al. [16] and Fan et al. [27], we introduce border detection techniques that retain considerable border information in the E1 and E2 characteristics provided by the backbone network. In an encode-decode structure, edge detection enhances the sensitivity and precision of lung segmentation by suppressing background noise while generating finer semantic segmentation maps. Dmbg-Net offers great potential for quick identification of COVID-19 and quick separation of pulmonary infections, especially for distinguishing between the contours and intricate boundaries of small-scale target lesions. The distribution of classes in the second column of Figure 8 indicates an imbalance in the data categories. Within these categories, there is a limited representation of GGO and consolidation lesions, while most images predominantly depict unaffected regions of the lungs. The potential features in areas of interest can be seen in the slices (represented in Figure 8), and the manifestations of the shape, size, type and location of the infected areas during COVID-19 infection are highly diverse.

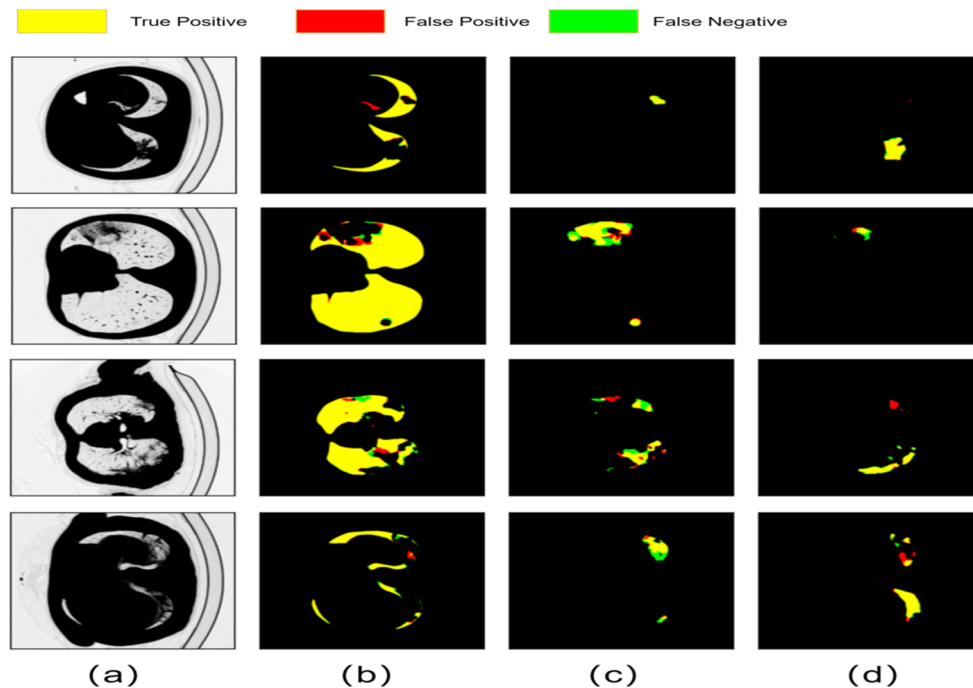


Figure 8. Visual comparison of lung field, GGO and consolidation segmentation with different examples in multiclass datasets. (a) Images, (b) lung field, (c) GGO and (d) consolidation. Yellow is true positives, red is false positives and green is false negatives.

Despite the promising results achieved by our Dmbg-Net in segmenting infected regions, there are certain limitations inherent in the current model. First, the focus of Dmbg-Net lies solely on the intact boundary of lung infection in COVID-19 patients. However, in clinical practice, it is often necessary to triage COVID-19 patients before proceeding with further treatment. In future work, we propose the integration of an automatic diagnosis system for lung lesions, COVID-19 detection, segmentation and quantification of lung infection into a unified framework. Second, in our multiclass infection segmentation framework, we adapt Dmbg-Net to guide and supervise the multiclass markers associated with different types of lung infections. Due to the scarcity of high-quality labelled data, this approach may result in suboptimal learning performance. We will study a viable option in which the semisupervised segmentation framework leverages unlabeled data to reduce the dependence on large amounts of labeled data.

5. Conclusions

Computer-aided COVID-19 infection segmentation is an effective approach for the early detection and diagnosis of lung lesions. A novel model, Dmbg-Net, is proposed and developed in this work, which uses an encoder-decoder framework with dilated convolution layers. These layers work together to refine boundaries and impose semantic constraints. Dmbg-Net is implemented on the multiclass segmentation of infections. The dilated residual block is designed to select the most crucial encoder features from the perspective of significant spatial information and contextual interdependence. The EGP block is designed to provide edge guidance, which can mitigate unclear boundaries effectively. Ablation studies have verified its effectiveness and accuracy. The proposed

method is used to analyze COVID-19 datasets, and the experimental results show that the proposed method has a Dice similarity coefficient of 85.6% and a sensitivity of 84.2%. Therefore, the proposed method has a potential application in the detection, labeling and segmentation of other lesion areas.

It is noteworthy that our Dmbg-Net model possesses a substantial number of parameters. However, in the context of training with deep neural networks for segmenting boundaries, it demonstrates a strong ability to effectively learn multiclass annotations of infected lesions. In future work, we will focus on the segmentation of multiclass infections with limited sample sizes, as well as more precise and automated COVID-19 infection recognition. This can be achieved through the advancement of network topologies and algorithms, coupled with enhancements in the quality of segmented edge contours.

Use of AI tools declaration

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

Acknowledgments

Supported by the Training Funding Program for The Youth Scholars of Shanghai Universities (Grant No. ZZGCD20019). We thank the anonymous reviewers and editors for their helpful and constructive suggestions on this paper.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. World Health Organization, WHO coronavirus (COVID-19) dashboard with vaccination data, 2023. Available from: <https://covid19.who.int/>.
2. A. Alhudhaif, K. Polat, O. Karaman, Determination of COVID-19 pneumonia based on generalized convolutional neural network model from chest X-ray images, *Expert Syst. Appl.*, **180** (2021). <https://doi.org/10.1016/j.eswa.2021.115141>
3. H. S. Shi, X. Y. Han, N. C. Jiang, Y. K. Cao, O. Alwalid, J. Gu, et al., Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study, *Lancet Infect. Dis.*, **20** (2020), 425–434. [https://doi.org/10.1016/S1473-3099\(20\)30086-4](https://doi.org/10.1016/S1473-3099(20)30086-4)
4. Z. Ye, Y. Zhang, Y. Wang, Z. X. Huang, B. Song, Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review, *Eur. Radio.*, **30** (2020), 4381–4389. <https://doi.org/10.1007/s00330-020-06801-0>
5. G. D. Rubin, C. J. Ryerson, L. B. Haramati, N. Sverzellati, J. P. Kanne, S. Raoof, et al., The role of chest imaging in patient management during the COVID-19 pandemic: a multinational consensus statement from the Fleischner society, *Radiology*, **296** (2020), 172–180. <https://doi.org/10.1148/radiol.2020201365>
6. M. Chung, A. Bernheim, X. Mei, N. Zhang, M. Huang, X. Zeng, et al., CT imaging features of 2019 novel coronavirus (2019-nCoV), *Radiology*, **295** (2020), 202–207. <https://doi.org/10.1148/radiol.2020200230>

7. H. Munusamy, K. J. Muthukumar, S. Gnanaprakasam, T. R. Shanmugakani, A. Sekar, FractalCovNet architecture for COVID-19 chest X-ray image classification and CT-scan image segmentation, *Biocybern. Biomed. Eng.*, **41** (2021), 1025–1038. <https://doi.org/10.1016/j.bbe.2021.06.011>
8. J. P. Kanne, Chest CT findings in 2019 novel coronavirus (2019-nCoV) infections from Wuhan, China: key points for the radiologist, *Radiology*, **295** (2020), 16–17. <https://doi.org/10.1148/radiol.2020200241>
9. O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, Springer, (2015), 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
10. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2015), 3431–3440. <https://doi.org/10.1109/CVPR.2015.7298965>
11. G. Rani, A. Misra, V. S. Dhaka, D. Buddhi, R. K. Sharma, E. Zumpano, et al., A multi-modal bone suppression, lung segmentation, and classification approach for accurate COVID-19 detection using chest radiographs, *Intell. Syst. Appl.*, **16** (2022), 200148. <https://doi.org/10.1016/j.iswa.2022.200148>
12. D. P. Fan, T. Zhou, G. P. 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>
13. R. Cong, H. Yang, Q. Jiang, W. Gao, H. Li, C. Wang, et al., BCS-Net: boundary, context, and semantic for automatic COVID-19 lung infection segmentation from CT images, *IEEE Trans. Instrum. Meas.*, **71** (2022), 1–11. <https://doi.org/10.1109/TIM.2022.3196430>
14. R. Cong, Y. Zhang, N. Yang, H. Li, X. Zhang, R. Li, et al., Boundary guided semantic learning for real-time COVID-19 lung infection segmentation system, *IEEE Trans. Consum. Electron.*, **68** (2022), 376–386. <https://doi.org/10.1109/TCE.2022.3205376>
15. X. Wang, Y. Yuan, D. Guo, X. Huang, Y. Cui, M. Xia, et al., SSA-Net: spatial self-attention network for COVID-19 pneumonia infection segmentation with semi-supervised few-shot learning, *Med. Image Anal.*, **79** (2022), 102459. <https://doi.org/10.1016/j.media.2022.102459>
16. S. Bose, R. S. Chowdhury, R. Das, U. Maulik, Dense dilated deep multiscale supervised u-network for biomedical image segmentation, *Comput. Biol. Med.*, **143** (2022). <https://doi.org/10.1016/j.compbimed.2022.105274>
17. Q. Mao, S. G. Zhao, D. B. Tong, S. C. Su, Z. W. Li, X. Cheng, Hessian-MRLoG: Hessian information and multi-scale reverse LoG filter for pulmonary nodule detection, *Comput. Biol. Med.*, **131** (2021). <https://doi.org/10.1016/j.compbimed.2021.104272>
18. Q. Mao, S. G. Zhao, L. J. Ren, Z. W. Li, D. B. Tong, X. Yuan, et al., Intelligent immune clonal optimization algorithm for pulmonary nodule classification, *Math. Biosci. Eng.*, **18** (2021), 4146–4161. <https://doi.org/10.3934/mbe.2021208>
19. G. Rani, A. Misra, V. S. Dhaka, E. Zumpano, E. Vocaturo, Spatial feature and resolution maximization GAN for bone suppression in chest radiographs, *Comput. Methods Programs Biomed.*, **224** (2022), 107024. <https://doi.org/10.1016/j.cmpb.2022.107024>
20. X. Yan, Y. Liu, M. Jia, Multiscale cascading deep belief network for fault identification of rotating machinery under various working conditions, *Knowledge-Based Syst.*, **193** (2020), 105484. <https://doi.org/10.1016/j.knosys.2020.105484>

21. X. Yan, D. She, Y. Xu, M. Jia, Deep regularized variational autoencoder for intelligent fault diagnosis of rotor-bearing system within entire life-cycle process, *Knowledge-based Syst.*, **226** (2021), 107142. <https://doi.org/10.1016/j.knosys.2021.107142>
22. X. Yan, D. She, Y. Xu, Deep order-wavelet convolutional variational autoencoder for fault identification of rolling bearing under fluctuating speed conditions, *Expert Syst. Appl.*, **216** (2023), 119479. <https://doi.org/10.1016/j.eswa.2022.119479>
23. F. Bougourzi, C. Distanto, F. Dornaika, A. Taleb-Ahmed, PDAtt-Unet: pyramid dual-decoder attention unet for COVID-19 infection segmentation from CT-scans, *Med. Image Anal.*, **86** (2023), 102797. <https://doi.org/10.1016/j.media.2023.102797>
24. L. Zhou, Z. Li, J. Zhou, H. Li, Y. Chen, Y. Huang, et al., A rapid, accurate and machine-agnostic segmentation and quantification method for CT-based COVID-19 diagnosis, *IEEE Trans. Med. Imaging*, **39** (2020), 2638–2652. <https://doi.org/10.1109/TMI.2020.3001810>
25. X. F. Wang, L. Jiang, L. Li, M. Xu, X. Deng, L. S. Dai, et al., Joint learning of 3D lesion segmentation and classification for explainable COVID-19 diagnosis, *IEEE Trans. Med. Imaging*, **40** (2021), 2463–2476. <https://doi.org/10.1109/TMI.2021.3079709>
26. X. Zhong, H. B. Zhang, G. L. Li, D. H. Ji, Do you need sharpened details? Asking MMDC-Net: multi-layer multi-scale dilated convolution network for retinal vessel segmentation, *Comput. Biol. Med.*, **150** (2022). <https://doi.org/10.1016/j.compbiomed.2022.106198>
27. C. Fan, Z. Zeng, L. Xiao, X. Qu, GFNet: automatic segmentation of COVID-19 lung infection regions using CT images based on boundary features, *Pattern Recognit.*, **132** (2022), 108963. <https://doi.org/10.1016/j.patcog.2022.108963>
28. S. Chakraborty, K. Mali, SUFEMO: a superpixel based fuzzy image segmentation method for COVID-19 radiological image elucidation, *Appl. Soft Comput.*, **129** (2022). <https://doi.org/10.1016/j.asoc.2022.109625>
29. N. Paluru, A. Dayal, H. B. Jenssen, T. Sakinis, L. R. Cenkeramaddi, J. Prakash, et al., Anam-Net: anamorphic depth embedding-based lightweight CNN for segmentation of anomalies in COVID-19 Chest CT Images, *IEEE Trans. Neural Networks Learn. Syst.*, **32** (2021), 932–946. <https://doi.org/10.1109/TNNLS.2021.3054746>
30. G. Wang, X. Liu, C. Li, Z. Xu, J. Ruan, H. Zhu, et al., A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images, *IEEE Trans. Med. Imaging*, **39** (2020), 2653–2663. <https://doi.org/10.1109/TMI.2020.3000314>
31. Y. H. Wu, S. H. Gao, J. Mei, J. Xu, D. P. Fan, R. G. Zhang, et al., Jcs: Anexplainable COVID-19 diagnosis system by joint classification and segmentation, *IEEE Trans. Image Process.*, **30** (2021), 3113–3126. <https://doi.org/10.1109/TIP.2021.3058783>
32. S. X. Zhao, Z. D. Li, Y. Chen, W. Zhao, X. Z. Xie, J. Liu, et al., SCOAT-Net: a novel network for segmenting COVID-19 lung opacification from CT images, *Pattern Recognit.*, **119** (2021). <https://doi.org/10.1016/j.patcog.2021.108109>
33. Z. J. Zhang, H. Z. Fu, H. Dai, J. B. Shen, Y. W. Pang, L. Shao, ET-Net: a generic edge-attention guidance network for medical image segmentation, in *Medical Image Computing and Computer Assisted Intervention—MICCAI 2019*, (2019), 442–450. https://doi.org/10.1007/978-3-030-32239-7_49
34. Z. Wu, L. Su, Q. Huang, Stacked cross refinement network for edge-aware salient object detection, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2019), 7264–7273.

35. Radiologists, COVID-19 CT segmentation dataset, 2020. Available from: <https://medical-segmentation.com/covid19>.
36. J. Ma, C. Ge, Y. Wang, X. An, J. Gao, Z. Yu, et al., COVID-19 CT lung and infection segmentation dataset, 2020. <https://doi.org/10.5281/zenodo.3757476>
37. Z. Zhou, M. 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*, Springer, (2018), 3–11. https://doi.org/10.1007/978-3-030-00889-5_1
38. 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, *arXiv preprint*, (2018), arXiv:1804.03999. <https://doi.org/10.48550/arXiv.1804.03999>
39. K. Zhang, X. Liu, J. Shen, Z. Li, Y. Sang, X. Wu, et al., Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography, *Cell*, **181** (2020), 1423–1433.
40. A. Paszke, A. Chaurasia, S. Kim, E. Culurciello, Enet: A deep neural network architecture for real-time semantic segmentation, *arXiv preprint*, (2016), arXiv:1606.02147. <https://doi.org/10.48550/arXiv.1606.02147>



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>)