WILEY | Hindawi

Research Article

An Edge-Assisted Computing and Mask Attention Based Network for Lung Region Segmentation

Yong Wang,[1] Like Zhong,[1] Weihong Huang,[2] and Xiaoyu He [1]

[1]*School of Automation, Central South University, Changsha 410083, China*
[2]*Mobile Health Ministry of Education-China Mobile Joint Laboratory, Changsha 410008, China*

Correspondence should be addressed to Xiaoyu He; hexiaoyu@csu.edu.cn

Recent years have witnessed the success of encoder-decoder structure-based approaches in lung region segmentation of chest X-ray (CXR) images. However, accurate lung region segmentation is still challenging due to the following three issues: (1) inaccurate lung region segmentation boundaries, (2) existence of lesion-related artifacts (e.g., opacity and pneumonia), and (3) lack of the ability to utilize multiscale information. To address these issues, we propose an edge-assisted computing and mask attention based network (called EAM-Net), which consists of an encoder-decoder network, an edge-assisted computing module, and multiple mask attention modules. Based on the encoder-decoder structure, an edge-assisted computing module is first proposed, which integrates the feature maps of the shallow encoding layers for edge prediction, and uses the edge evidence map as a strong cue to guide the lung region segmentation, thereby refining the lung region segmentation boundaries. We further design a mask attention module after each decoding layer, which employs a mask attention operation to make the model focus on lung regions while suppressing the lesion-related artifacts. Besides, a multiscale aggregation loss is proposed to optimize EAM-Net. Extensive experiments on the JSRT, Shenzhen, and Montgomery datasets demonstrate that EAM-Net outperforms existing state-of-the-art lung region segmentation methods.

1. Introduction

Lung is one of the most important organs in the human body. There are many types of lung diseases with high incidence [1]. Among them, lung cancer is the most common and deadliest tumor in China, with a 5-year survival rate of 16.4% [2]. Chest X-ray (CXR) is a widely used technique to identify lung diseases. However, manual interpretation of CXR images is usually time-consuming, laborious, and subjective. To this end, many efficient image processing methods have been proposed for the computer-aided diagnosis of lung diseases, in which lung region segmentation is a critical task. It provides basic information on lung shape and size measurements, which can be used for pathological analysis [3]. Therefore, automatic lung region segmentation of CXR images deserves an in-depth study.

Traditional image processing methods use edge detection, thresholding, and clustering to segment the target area of an image [4]. These methods are simple to implement but suffer from poor generalization performance. With the development of deep learning technology, the application of convolutional neural networks (CNNs) for image segmentation is increasing. Compared with traditional image processing methods, CNNs can automatically learn meaningful features from data, thus achieving improved segmentation performance.

Currently, the encoder-decoder structure is one of the most effective CNN models in the field of image segmentation because it preserves the detailed features of images [5]. For example, Long et al. [6] proposed a fully convolutional network, which can take an image of any size as input and produce a pixel-wise segmentation prediction map. Ronneberger et al. [7] proposed a U-shaped network (called UNet), which has been widely used in medical image analysis. UNet consists of three parts: a shrinking path, an expanding path, and skip connections. The input image is

downsampled four times on the shrinking path for feature extraction, and upsampled four times on the expanding path to restore to its original input size. Between the shrinking and expanding paths, feature maps at the same level are connected by skip connections to supplement the loss of image details during successive downsampling.

Although encoder-decoder structure-based approaches have achieved remarkable results in the field of image segmentation, precise lung region segmentation is still challenging due to the following three issues. First, the calculation of various pathological indicators relies on accurate lung region segmentation boundaries. However, the edge of the lung region often has some problems such as oscillation, deformation, and noise. As a result, the edge cannot be preserved well. Second, in more complex cases, patients may suffer from some diseases that affect the lung region segmentation, such as opacity, consolidation, tuberculosis, pulmonary nodules, and pneumonia, as shown in Figure 1. From Figure 1, the high-intensity abnormal pixels overlap with the real lung region, thus reducing the contrast of the lung region boundaries. During segmentation, lung regions that overlap with opacity may be incorrectly predicted as lung boundaries. Third, due to the loss of information caused by successive downsampling operations, the model cannot make full use of the effective information of features at different scales during training, including semantic features for lung region identification and detailed features for edge localization.

To address the abovementioned issues, this paper proposes an edge-assisted computing and mask attention based network (called EAM-Net) for lung region segmentation of CXR images. EAM-Net follows the classic encoder-decoder structure, where the encoder extracts the feature map of an input image, and the decoder restores the feature map to the original input size. In order to learn robust lung region segmentation boundaries, an edge-assisted computing module is designed, which fuses the feature maps of the first three encoding layers to generate an edge evidence map. Note that this edge evidence map is used as a strong cue to guide the lung region segmentation, thus refining the edge parts of the lung region segmentation result. We further design a mask attention module after each decoding layer, which employs a spatial attention mechanism to learn important regions on the segmentation features while ignoring irrelevant artifacts. EAM-Net is optimized by a multiscale aggregation loss, which consists of a multiscale lung region segmentation loss and an edge prediction loss. Overall, EAM-Net has the capability to deal with the abovementioned three issues in lung region segmentation of CXR images. The main contributions of this paper are summarized as follows:

(i) We propose the edge-assisted computing module, which integrates the feature maps of the shallow encoding layers for edge prediction and transfers meaningful edge information to guide the lung region segmentation.

(ii) We design a mask attention module after each decoding layer, which utilizes the feature map output from the decoding layer for segmentation prediction, and enhances the lung regions on the segmentation features through a mask attention operation.

(iii) A multiscale aggregation loss is proposed to jointly supervise the lung region segmentation task and edge prediction task in EAM-Net, which is beneficial to explore the correlation between these two tasks and can make full use of the multiscale information through a deep supervision method.

(iv) Extensive experiments are conducted on the JSRT, Shenzhen, and Montgomery datasets to verify the effectiveness of EAM-Net. The results show that EAM-Net can provide more accurate lung region segmentation results than all its competitors.

The rest of this paper is organized as follows: Section 2 introduces the related work. The proposed approach, i.e., EAM-Net, is elaborated in Section 3. Section 4 and Section 5 present the experimental setup and the experimental results, respectively. Finally, Section 6 concludes this paper.

2. Related Work

Since our method is based on the encoder-decoder structure and we also make use of attention mechanism to enhance features, their related work for lung region segmentation is briefly introduced in this section.

2.1. Encoder-Decoder Structure. In the field of medical image segmentation based on deep learning [8–10], the encoder-decoder structure is one of the most commonly used network structures [11–13]. UNet [7] and SegNet [14] are two representatives of the encoder-decoder structure-based methods. In the following, we introduce the related work on these two methods and their variants for lung region segmentation.

UNet is a segmentation network proposed for medical images. The core of UNet is the skip connection, which combines high-level semantic information with low-level image details. Novikov et al. [15] proposed an improved UNet called InvertedNet to segment the clavicle, lung, and heart. They added a Gaussian decay after each convolutional layer, and used the exponential linear unit [16] instead of rectified linear unit (ReLU) to speed up model training. Yahyatabar et al. [17] proposed a lightweight UNet model called Dense-Unet for lung region segmentation, which utilizes dense connections among layers for feature reuse. Rahman et al. [18] proposed a two-stage framework for lung region segmentation. In the first stage, a UNet model is trained to generate initial segmentation results for lung regions. In the second stage, the morphological processing is performed on the initial segmentation results to obtain the final refined segmentation results. Abas Hasan and Mohsin Abdulazeez [19] proposed an improved UNet with less parameters and network training time. Compared with the original UNet, it reduces the number of convolution kernels in each layer and the network training time by 75% and 70%,

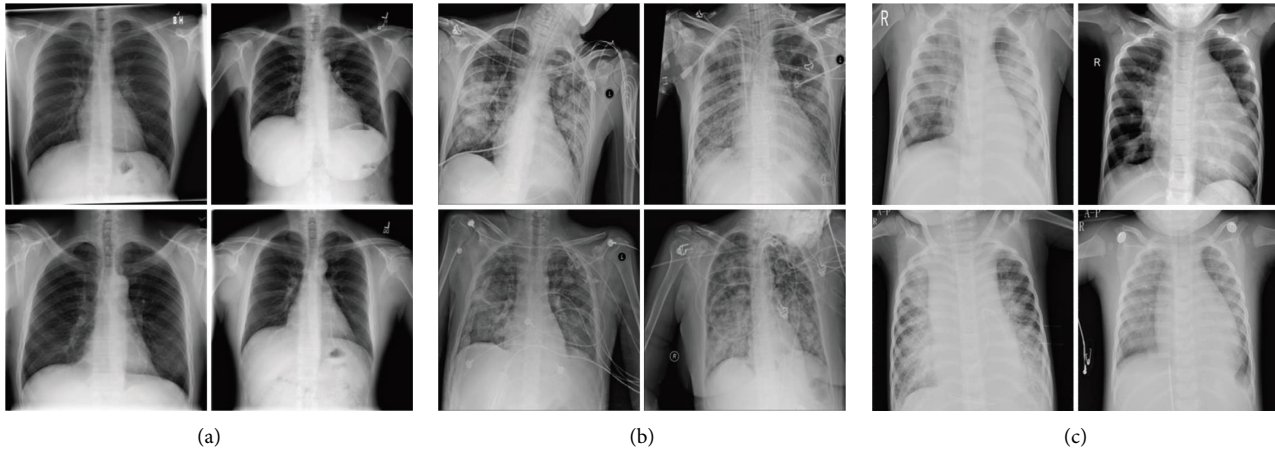|		
(a)	(b)	(c)

FIGURE 1: Some examples of CXR images of normal people (a) and patients with lung diseases such as opacity (b) and pneumonia (c).

respectively. Zhang et al. [20] proposed an edge attention guided network called ET-Net for lung region segmentation. ET-Net employs an edge-guided module to learn edge representations, and a weighted aggregation module to integrate the edge representations. Solovyev et al. [21] proposed a Bayesian-based deep learning framework for lung region segmentation using a standard ResNet50 as the encoder and a Bayesian feature pyramid network as the decoder. Kholiavchenko et al. [22] proposed a contour-aware segmentation framework for lung region segmentation, which combines the advantages of UNet, LinkNet, and Tiramisu architectures. Arsalan et al. [23] proposed an improved UNet called X-ray-Net for multiclass segmentation of the lung, heart, and clavicle bones. X-ray-Net uses a residual mesh in its network to preserve the spatial edge information. Milletari et al. [24] proposed a U-shaped approach called CFCM for lung region segmentation. The core idea of CFCM is to use long short-term memory networks to fuse features from different encoding layers.

Different from UNet, SegNet utilizes the maximum pooling index for upsampling, with the aim of better preserving the boundary feature information. In SegNet, the index of the maximum value in the pooling layer is recorded, which is directly used for the unpooling operation during upsampling. Kalinovsky and Kovalev [25] combined SegNet with histogram equalization and local contrast normalization (LCN) for lung region segmentation. The histogram equalization is employed on the input image in the preprocessing stage, and the LCN operations are implemented in each encoding layer. In [25], the segmentation accuracy is 96.2% on 354 CXR images. Saidy and Lee [26] employed a modified SegNet for lung region segmentation, which performs batch normalization and ReLU activation functions after each convolutional layer, and adopts pooling indices to restore the feature maps to the original input size. They achieved the Dice coefficient of 96% on a test set of 35 unseen images. Mittal et al. [27] proposed LF-SegNet for lung region segmentation, which upsamples feature maps by

simple replication and uses dropout layers for regularization and avoiding overfitting. On 199 images from the JSRT and Montgomery datasets, the average accuracy and Jaccard of LF-SegNet are 98.73% and 95.10%, respectively.

2.2. Attention Mechanism. The attention mechanism in deep learning is similar to human visual attention, which can focus on information that is more important to the current task. Many attention-based methods have been proposed to enhance the feature representation capabilities of CNNs [28, 29]. For example, Hu et al. [28] proposed SENet that models the dependencies among channels to improve the performance of the network. Fu et al. [29] proposed a dual attention network called DANet for scene segmentation. DANet jointly employs a position attention module and a channel attention module to learn spatial and channel dependencies.

There are also some attention-based approaches for lung region segmentation. Tang et al. [30] proposed a criss-cross attention [31] based network for lung region segmentation, in which the criss-cross attention is used to capture global contextual information and enhance pixel-level representations in both horizontal and vertical directions, thus boosting the lung segmentation performance. Kim and Lee [32] proposed a UNet with a self-attention module for lung region segmentation. They utilized the self-attention mechanism for feature optimization and achieved improved performance on several medical image segmentation datasets. Cao and Zhao [33] proposed a lung region segmentation method to address the issue of opacity regions in patient CXR. They designed a three-terminal attention mechanism, which integrates both the channel and spatial attention mechanisms to make the model focus on target regions. Li et al. [34] introduced the SE module [28] into a fully convolutional neural network for lung region segmentation. Given an input feature map, the SE module can automatically learn and generate the channel attention weight. Then, the input feature map can be adaptively

calibrated and optimized along the channel dimension by using this weight.

From the above introduction, we can find that existing methods usually make use of the original encoder-decoder architecture or its attention-based variants for segmentation. Accurate lung region segmentation is still challenging due to inaccurate lung region segmentation boundaries, existence of lesion-related artifacts, and lack of the ability to utilize multiscale information. Therefore, to further improve the performance of lung region segmentation, it is necessary to propose a unified framework that can address the above-mentioned three issues.

3. Proposed Approach

3.1. Overview. Based on the analysis in Section 2, we propose EAM-Net for lung region segmentation of CXR images, which combines the advantages of the encoder-decoder structure and attention mechanism to achieve accurate segmentation of lung regions. Figure 2 shows the architecture of EAM-Net, which consists of an encoder-decoder network, an edge-assisted computing module, and five mask attention modules. The edge-assisted computing module is placed on the side of the encoder, which utilizes the features of shallow encoding layers for edge prediction and uses the edge information to achieve better segmentation. The mask attention module is added after each decoding layer, which employs a mask attention operation to highlight important regions on segmentation features and ignore irrelevant artifacts on them.

3.2. Encoder-Decoder Network. The encoder-decoder network is constructed based on the classic UNet and consists of an encoder and a decoder. In order to extract more representative features, we adopt ResNet [35] as the backbone of the encoder. As shown in Figure 2, the encoder consists of five encoding layers: conv1, conv2_x, conv3_x, conv4_x, and conv5_x. The encoding layers at different stages are used to process feature maps with different resolutions. In the experiments of this paper, we used ResNet18, ResNet34, and ResNet50 as the backbone of the encoder, respectively, to study the effect of different network depths on the segmentation performance. Table 1 describes the structures of ResNet18, ResNet34, and ResNet50. Note that ResNet18 and ResNet34 use basic block as the residual module, while ResNet50 adopts bottleNeck block instead. Figure 3 presents the specific structures of these two blocks. From Figure 3, the basic block consists of two 3×3 convolutions and a residual connection, while the bottleNeck block consists of two 1×1 convolutions, a 3×3 convolution, and a residual connection. Assuming that x is the input of a residual module, the output of the residual module (denoted as y) is calculated as follows:

$$y = F(x, W) + x, \tag{1}$$

where $F(\cdot)$ represents the residual mapping of the stacked convolutional layers in the residual module, and W represents the parameter weights of these stacked layers. In EAM-

Net, the use of residual module makes the network stack deeper, which increases the representation ability of the network and avoids the vanishing gradient problem.

The decoder consists of five decoding layers that gradually restore the feature map output by the encoder to the original input size. Note that between two adjacent decoding layers, there is a mask attention module, which is proposed to further optimize the decoding features. Except for the first decoding layer (i.e., the decoding layer connected to conv_5), each decoding layer contains two inputs, i.e., the feature map output by the previous mask attention module and the feature map output by the encoding layer at the corresponding level. The former is first 2× upsampled by the bilinear interpolation operation, and then concatenated with the latter, followed by a 3×3 convolution and a ReLU activation function for feature fusion.

3.3. Edge-Assisted Computing Module. In existing encoder-decoder based segmentation methods, the edge and structure information of an image may be lost due to successive downsampling operations, thus leading to inaccurate segmentation edges. Generally, the detailed edge and structure information are included in the feature maps of the early stages of the network. Based on this, we design an edge-assisted computing module, which integrates the feature maps of the shallow encoding layers for edge prediction and transfers meaningful edge information to help better segmentation.

The structure of the edge-assisted computing module is shown in Figure 4. Considering the trade-off between model performance and computation complexity, we adopt the feature maps of the first three encoding layers to compute the edge evidence map. We denote these three feature maps as $E_1 \in \mathbb{R}^{H/2 \times W/2 \times C_1}$, $E_2 \in \mathbb{R}^{H/4 \times W/4 \times C_2}$, and $E_3 \in \mathbb{R}^{H/8 \times W/8 \times C_3}$, respectively, where C_1, C_2, and C_3 are their corresponding numbers of channels, and H and W represent the height and width of the original input image, respectively. From Figure 4, E_1, E_2, and E_3 are processed by three parallel branches to generate E'_1, E'_2, and E'_3. Each branch contains a 1×1 convolution and a 3×3 convolution. The 1×1 convolution is used to adjust the numbers of channels of E_1, E_2, and E_3 to 32, and the 3×3 convolution is used to perform feature extraction. Afterward, E'_1, E'_2, and E'_3 are upsampled to the resolution of the input image and concatenated together to generate a fused feature map, denoted as $F_{\text{fusion}} \in \mathbb{R}^{H \times W \times 96}$:

$$F_{\text{fusion}} = \mathscr{C}\left(U^{2\times}(E'_1), U^{4\times}(E'_2), U^{8\times}(E'_3)\right), \tag{2}$$

where $U^{2\times}(\cdot)$, $U^{4\times}(\cdot)$, and $U^{8\times}(\cdot)$ represent 2×, 4×, and 8× bilinear interpolation operations, respectively, and $\mathscr{C}$ denotes the concatenation operation. Because F_{fusion} is further used for edge prediction, this fused feature map is also called the edge evidence map. Subsequently, on the one hand, F_{fusion} is processed by a 1×1 convolution to reduce the number of channels to 1 for edge prediction, which generates the edge prediction map, denoted as F_{edge}. On the other hand, as shown in Figure 2, F_{fusion} is concatenated with the output feature map of the last mask attention module for
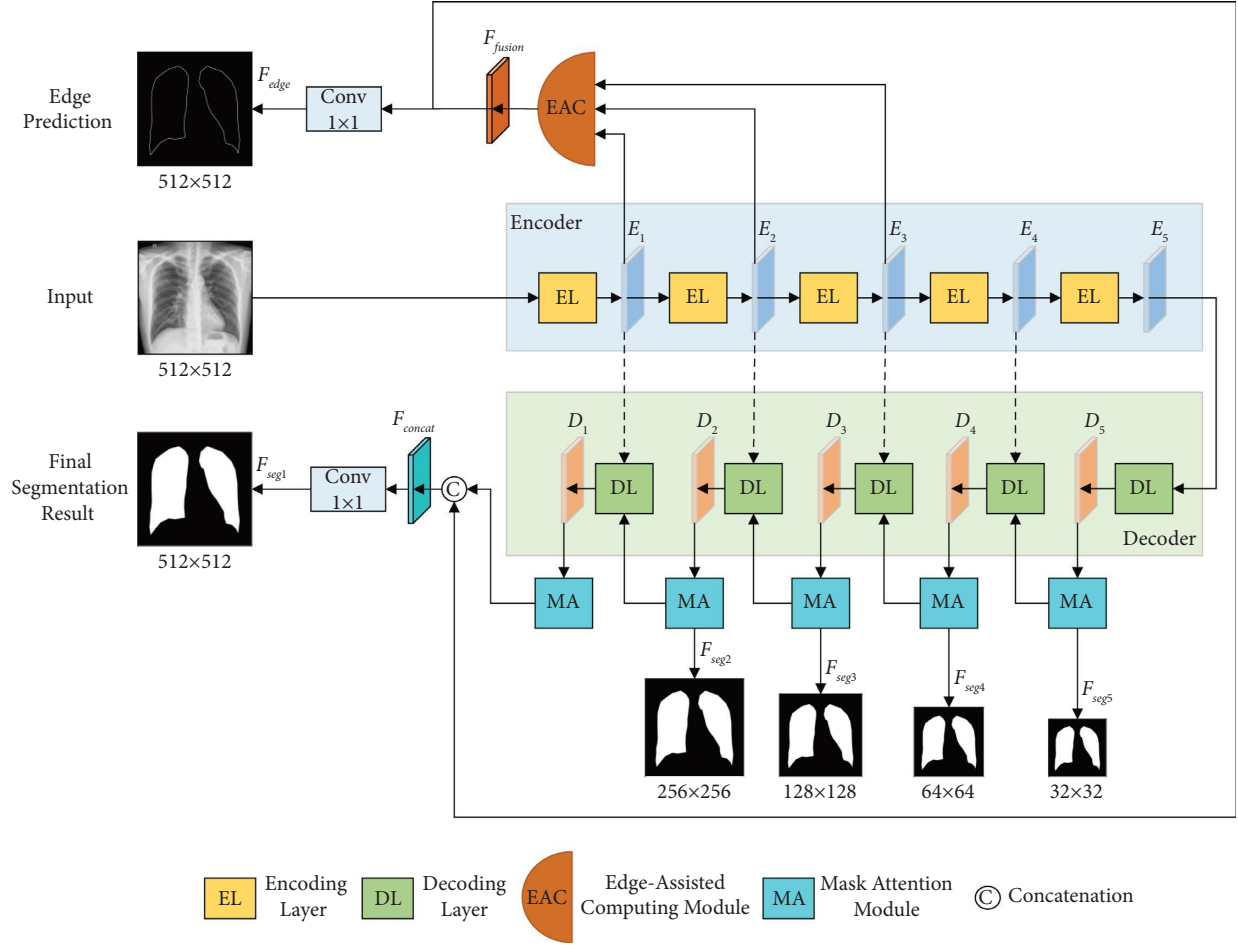
FIGURE 2: Overview of the proposed EAM-Net.

TABLE 1: Encoder structures of EAM-Net with three network depths.

Layer name	ResNet18	ResNet34	ResNet50
conv1	Conv, 7×7, stride 2	Conv, 7×7, stride 2	Conv, 7×7, stride 2
conv2_x		3×3 max pooling, stride 2	
	(Basic block)×2	(Basic block)×3	(BottleNeck block)×3
conv3_x	(Basic block)×2	(Basic block)×4	(BottleNeck block)×4
conv4_x	(Basic block)×2	(Basic block)×6	(BottleNeck block)×6
conv5_x	(Basic block)×2	(Basic block)×3	(BottleNeck block)×3

lung region segmentation. The purpose of this feature concatenation operation is to connect the edge prediction task with the lung region segmentation task, and to introduce meaningful edge information included in F_{fusion} to enhance the lung region segmentation performance.

3.4. Mask Attention Module.
Existing attention-based methods have achieved promising results in various segmentation tasks. However, they usually rely on global average pooling operation or global max pooling operation to learn spatial attention weights [36, 37]. These global pooling operations roughly merge the features of each channel, which may lead to the loss of some key information. To this

end, we design a mask attention module, which employs a supervised learning branch to learn a mask feature map representing lung regions, and adopts this mask feature map as an attention weight to enhance the lung regions on segmentation features.

The structure of the mask attention module is shown in Figure 5. The input to this module is the output feature map of the decoder at the ith stage, denoted as $D_i (i \in \{1, 2, 3, 4, 5\})$. For D_i, we first construct a supervised learning branch consisting of a 3×3 convolution and a 1×1 convolution to obtain a mask feature map A_i:

$$A_i = \mathscr{F}_i^{1\times1}\left(\mathscr{F}_i^{3\times3}\left(D_i; \theta_i^{3\times3}\right); \theta_i^{1\times1}\right), \tag{3}$$
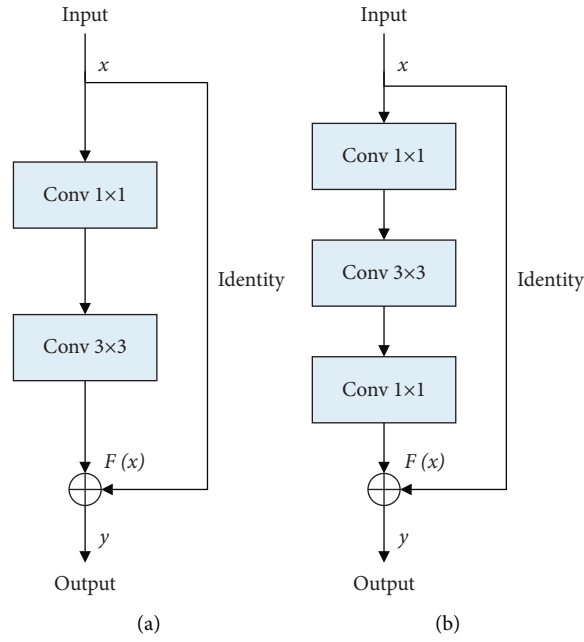
FIGURE 3: Two types of residual modules: (a) basic block and (b) bottleNeck block.
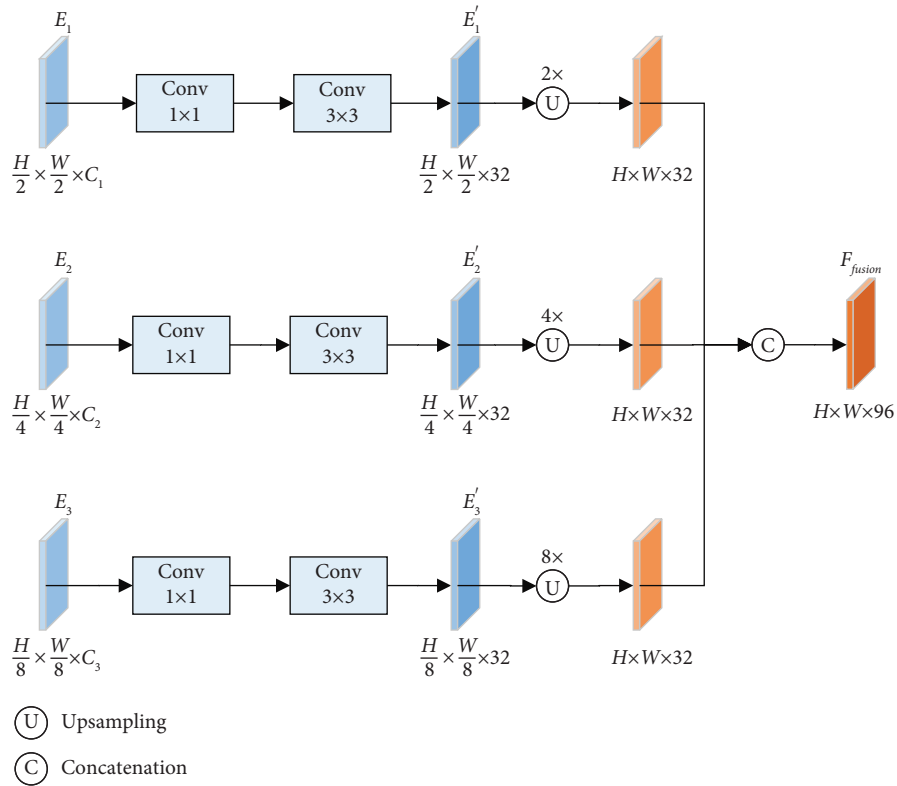

U   Upsampling

C   Concatenation

FIGURE 4: Structure of the edge-assisted computing module.

where $\mathscr{F}_i^{3\times3}$ and $\mathscr{F}_i^{1\times1}$ indicate the 3×3 and 1×1 convolutions, respectively, and $\theta_i^{3\times3}$ and $\theta_i^{1\times1}$ are their corresponding parameters. Herein, the 3×3 convolution is employed to extract image features and the 1×1 convolution is used to adjust the number of channels to 1. As

shown in Figure 5, on the one hand, A_i is used for lung region segmentation, which can receive the optimization of lung region information during training. On the other hand, A_i is processed by the Sigmoid activation function and used as the attention weight to multiply D_i. This multiplication
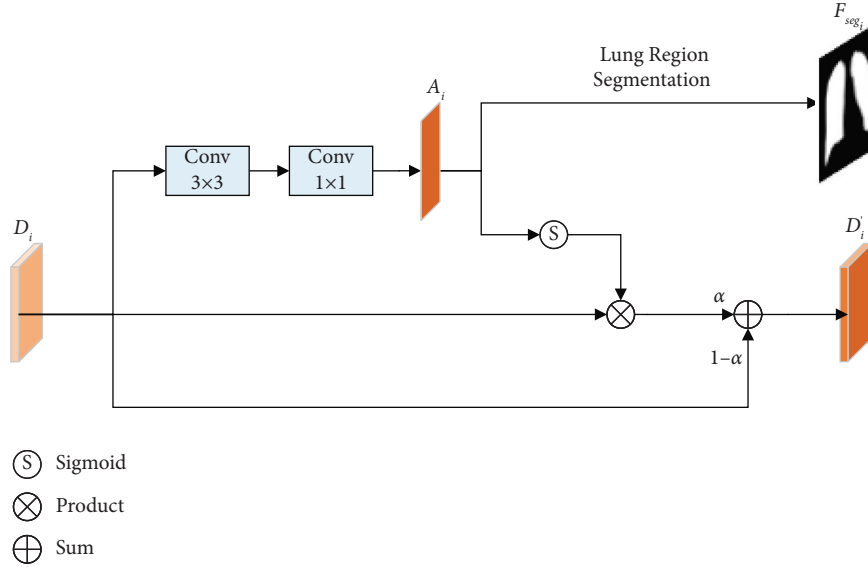
FIGURE 5: Structure of the mask attention module.

operation is called the mask attention operation. Since A_i is optimized to characterize lung regions, the mask attention operation can improve the lung region parts of D_i while suppressing irrelevant lesion-related distractions. To ensure the convergence of the model, a residual connection from D_i to the output is added, which generates the final refined feature map, denoted as D_i'. The abovementioned mask attention operation can be formulated as follows:

$$D_i' = \alpha \cdot (\sigma(A_i) \odot D_i) + (1 - \alpha) \cdot D_i, \quad (4)$$

where σ represents the Sigmoid activation function, $\odot$ denotes the element-wise product operation, and α is a learnable weighting factor to adjust the contribution of the attention part in the final refined feature map. α is initialized to 0 and its value can be dynamically optimized during network training.

3.5. Multiscale Aggregation Loss Function. Existing encoder-decoder based segmentation methods usually design a network to directly learn the mapping from an input image to the segmentation result, and only supervise the prediction at the last stage of the network [38]. Therefore, they cannot make full use of the effective information at early and immediate stages of the network. Unlike these methods, EAM-Net adds an edge prediction task as the auxiliary task to improve the segmentation boundaries, and adopts a deep supervision method to supervise the segmentation prediction at each stage of the network. In order to train EAM-Net, we propose a multiscale aggregation loss function, which is defined as follows:

$$\mathscr{L}_{\text{total}} = \mathscr{L}_{\text{edge}} + \mathscr{L}_{\text{seg}}^{\text{multi-scale}}, \quad (5)$$

where $\mathscr{L}_{\text{edge}}$ and $\mathscr{L}_{\text{seg}}^{\text{multiscale}}$ denote the edge prediction loss and multiscale segmentation loss, respectively. To deal with the class imbalance problem in medical image segmentation tasks, we jointly use the binary cross-entropy loss, Dice loss,

and Jaccard loss. Specifically, $\mathscr{L}_{\text{edge}}$ and $\mathscr{L}_{\text{seg}}^{\text{multiscale}}$ are calculated as follows:

$$\mathscr{L}_{\text{edge}} = L_{\text{BCE}}(F_{\text{edge}}, G_{\text{edge}}) + L_{\text{Dice}}(F_{\text{edge}}, G_{\text{edge}})$$
$$+ L_{JA}(F_{\text{edge}}, G_{\text{edge}})$$
$$\mathscr{L}_{\text{seg}}^{\text{multi-scale}} = \sum_{i \in I} L_{\text{BCE}}(F_{\text{seg}_i}, G_{\text{seg}_i}) + L_{\text{Dice}}(F_{\text{seg}_i}, G_{\text{seg}_i}) \quad (6)$$
$$+ L_{JA}(F_{\text{seg}_i}, G_{\text{seg}_i}), I = \{1, 2, 3, 4, 5\},$$

where L_{BCE}, L_{Dice}, and L_{JA} represent the binary cross-entropy, Dice, and Jaccard losses, respectively, F_{edge} and G_{edge} are the edge prediction result and edge ground-truth label, respectively, and F_{seg_i} and G_{seg_i} are the segmentation result and segmentation ground-truth label at the ith stage of the decoder, respectively. As shown in Figure 2, assuming that the resolution of the input image is 512×512, the dimensions of the feature maps output from the first to fifth stages of the decoder are 512×512, 256×256, 128×128, 64×64, and 32×32, respectively.

4. Experimental Setup

4.1. Datasets. The performance of EAM-Net was examined on three public datasets, i.e., the JSRT, Shenzhen, and Montgomery datasets. The JSRT dataset [39] is commonly used for lung nodule detection and lung region segmentation, containing 247 CXR images from 14 institutions in Japan and USA. These images are 12-bit images with a resolution of 2048×2048 and include the labels of nodule location (on 154 images) and diagnosis (malignant or benign). Note that the lung region segmentation labels for these images were obtained from [40]. The Shenzhen dataset [41] contains 662 CXR images, of which 326 images belong to normal and 336 images represent pulmonary tuberculosis. These images were collected at the Third People's Hospital of Shenzhen, Guangdong Province, China. They are 8-bit

images with the tuberculosis annotations and lung region segmentation labels. The Montgomery dataset [41] contains 138 CXRs images, including 80 normal and 58 pulmonary tuberculosis images. These images were collected by the Tuberculosis Control Program at the Montgomery County Department of Health and Human Services, Maryland, USA, and annotated with nodule location and lung region segmentation labels.

From the above description, all the three datasets have the interference of lung diseases, which makes lung region segmentation more difficult. Many researchers have tested their methods on these three datasets. Therefore, it is convenient to compare EAM-Net with other methods on them. However, due to the fact that there is no official training and test data separation for these three datasets, existing methods employ different data separation methods to evaluate their performance. In the following experiments, we adopted a ratio of 7 : 3 to randomly split the training and test sets, and conducted 10 independent experiments on the proposed method. This separation ratio is commonly used in the machine learning community and can reflect the general performance of the model on a dataset. In addition, we averaged the results of the 10 independent experiments as the model performance, which can greatly reduce the accidental error caused by the one-time data separation to the model performance. The detailed data distributions of the JSRT, Shenzhen, and Montgomery datasets are shown in Table 2.

4.2. Algorithms for Comparison. The compared methods of EAM-Net on the JSRT, Shenzhen, and Montgomery datasets can be simply divided into two categories: the baseline method and the state-of-the-art lung region segmentation methods. Res-UNet is chosen as the baseline method, which is a U-shaped network with ResNet as the encoder backbone. EAM-Net also adopts ResNet as the encoder backbone, and additionally proposes two new modules, i.e., the edge-assisted computing module and the mask attention module. The state-of-the-art lung region segmentation methods refer to those that have reported excellent performance on the three selected datasets over the past five years. The details of these methods have been introduced in Section 2. Note that for a fair comparison with these state-of-the-art methods, we reproduced the results of these methods using the same training and test data separation as the proposed EAM-Net. For [9, 20, 21, 24], we directly used their source codes to reproduce the results under our data separation. For those methods without source codes available (i.e., [8, 10, 15, 17, 18, 22, 23, 32]), we also reproduced their results under our data separation according to the technical details of the corresponding papers.

4.3. Implementation Details. During the training phase, all training images of EAM-Net were prepared in the form of edge images and multiscale segmentation images, which have been introduced in Section 3.5. We first resized the original input images to the size of 512×512. Note that we obtained the edge mask (i.e., the ground truth of the edge) by

TABLE 2: Data distributions of the JSRT, Shenzhen, and Montgomery datasets.

Dataset	Disease	Total	Training	Test
JSRT	Nodule	154	108	46
	Non-nodule	93	65	28
Shenzhen	Tuberculosis	336	235	101
	Nontuberculosis	326	228	98
Montgomery	Tuberculosis	58	41	17
	Nontuberculosis	80	56	24

applying a contour detection algorithm to the segmentation mask, where the thickness of the edge contour was set to 8. Then, we performed the online data augmentation, including random horizontal flip and random vertical flip, with the aim of preventing the model from overfitting. The Adaptive Moment Estimation (i.e., Adam) algorithm was used as the optimizer. The initial learning rate was set to 0.001, and we reduced the learning rate by 1/10 when the loss does not decrease for three consecutive epochs. The batch size of the training set was set to 8. We loaded the pretrained weights of residual blocks on ImageNet for the encoder of EAM-Net. The maximum number of training epochs was set to 50. In the test phase, for an input CXR image, EAM-Net could generate multiscale segmentation results, and the segmentation result with the size of 512×512 was used as the final prediction result to evaluate the performance of the model. The source codes of the proposed EAM-Net can be downloaded from the publication page (https://intleo.csu.edu.cn/publication.html) of our research group for reproducing the experimental results of this paper.

4.4. Evaluation Metrics. Three commonly used evaluation metrics were adopted to measure the performance of lung region segmentation: pixel Accuracy (PA), Dice, and Jaccard (JA). They are defined as follows:

$$PA = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|} = \frac{2TP}{2TP + FP + FN} \quad (7)$$

$$JA = \frac{|X \cap Y|}{|X \cup Y|} = \frac{TP}{TP + FP + FN},$$

where TP, TN, FP, and FN are the true positives, true negatives, false positives, and false negatives, respectively, and X and Y represent a segmentation map and a ground-truth map, respectively. PA is used to calculate the percentage of correctly classified pixels in an image, while Dice and Jaccard are used to measure the similarity between the segmentation map and the ground-truth map.

5. Results and Discussion

5.1. Comparion with the Baseline Method. First, we compared EAM-Net with the baseline method (i.e., Res-UNet) on the JSRT, Shenzhen, and Montgomery datasets. Table 3 summarizes the results of EAM-Net and Res-UNet using

different encoder backbones (i.e., ResNet18, ResNet34, and ResNet50). From Table 3, no matter which encoder backbone is used, EAM-Net achieves better segmentation performance than Res-UNet in terms of all the three evaluation metrics on the three datasets. Compared with Res-UNet18, EAM-Net18 improves JA by 1.0%, 0.9%, and 1.27% on the JSRT, Shenzhen, and Montgomery datasets, respectively. Compared with Res-UNet50, EAM-Net50 improves Dice by 0.42%, 0.6%, and 0.82%, respectively, and JA by 0.66%, 0.87%, and 0.92%, respectively, on these three datasets. Since EAM-Net is the network that adds the edge-assisted computing module and mask attention module to Res-UNet, the abovementioned results demonstrate the effectiveness of these two modules proposed in this paper.

5.2. Comparion with the State-of-the-Art Lung Region Segmentation Methods.

We also compared the performance of EAM-Net with that of the state-of-the-art lung region segmentation methods. Herein, ResNet50 was used as the backbone in EAM-Net. The results of these methods on the JSRT, Shenzhen, and Montgomery datasets are reported in Tables 4–6, respectively.

The results in Table 4 suggest that on the JSRT dataset, EAM-Net achieves the highest overall performance among all the compared methods. Compared with the Bayesian feature pyramid network [21], EAM-Net improves the performance by a large margin, with a 4.98% improvement in terms of Dice and a 9.01% improvement in terms of JA. Compared with the second-best model (i.e., SED [9]), EAM-Net improves Dice by 0.60% and JA by 1.01%.

As can be seen from Table 5, among all the compared methods, EAM-Net obtains the highest performance in terms of all the three evaluation metrics on the Shenzhen dataset. Specifically, in terms of Dice and JA, EAM-Net outperforms the second best method (i.e., Kim and Lee [32]) by 0.68% and 1.01%, respectively. Compared with X-ray-Net [23], EAM-Net improves Dice and JA by 1.21% and 1.94%, respectively.

From Table 6, on the Montgomery dataset, EAM-Net achieves the highest PA of 99.16%, highest Dice of 98.23%, and highest JA of 96.52% among all the compared methods. Compared with X-ray-Net, which performs the second best, EAM-Net improves PA, Dice, and JA by 0.59%, 0.83%, and 1.59%, respectively. Compared with ET-Net [20] that also uses the edge attention mechanism, EAM-Net obtains a 0.65% improvement in terms of PA and a 2.20% improvement in terms of JA.

The abovementioned results demonstrate the superiority of EAM-Net over the state-of-the-art lung region segmentation methods.

5.3. Visualization of Segmentation Results.

In order to show the segmentation performance more intuitively, we visualized some segmentation results of EAM-Net (using ResNet50 as the backbone) in Figure 6. For each CXR image in Figure 6, we used red and green curves to denote the segmentation prediction and ground truth, respectively, and the corresponding Dice and JA values were also provided.

Figure 6(a) depicts the segmentation results of five CXR images on the JSRT dataset, in which the first four images represent lungs with lung nodules and the last image represents a normal lung. It can be observed that even in the presence of lung nodules, EAM-Net can still obtain accurate lung region segmentation boundaries. Figures 6(b) and 6(c) show the segmentation results on the Shenzhen and Montgomery datasets, respectively. The images in the first four columns represent lungs with tuberculosis, and the images in the last column represent normal lungs. From Figures 6(b) and 6(c), the presence of tuberculosis causes a slight segmentation performance drop, but EAM-Net still achieves promising Dice and JA values. Overall, the abovementioned visualization results suggest that EAM-Net can achieve precise lung region segmentation despite the existence of lesion-related artifacts.

5.4. Ablation Study.

In EAM-Net, we proposed two kinds of modules, i.e., the edge-assisted computing module and the mask attention modules. Furthermore, we designed a multiscale aggregation loss to optimize EAM-Net, which consists of an edge prediction loss and a multiscale segmentation loss. To investigate their effectiveness, we conducted the ablation experiments on each of them. Table 7 gives the comparison results on the Montgomery dataset. Note that all methods took ResNet50 as the backbone of the encoder for a fair comparison.

5.4.1. Effectiveness of the Edge-Assisted Computing Module.

To validate the effectiveness of the edge-assisted computing module, we compared EAM-Net with and without the edge-assisted computing module. For EAM-Net without the edge-assisted computing module, we removed the edge-assisted computing module from the model and directly used the output feature of the last mask attention module to generate the final segmentation result. Meanwhile, the multiscale aggregation loss was degraded to a multiscale segmentation loss. From Table 7, the edge-assisted computing module improves Dice and JA by 0.44% and 0.47%, respectively. The performance gain of EAM-Net can be attributed to the fact that the edge-assisted computing module could transfer meaningful edge information from the edge prediction task to the segmentation task, thereby improving the lung region segmentation performance.

5.4.2. Effectiveness of the Mask Attention Modules.

We also compared EAM-Net with and without the mask attention modules. For EAM-Net without the mask attention modules, five mask attention modules were removed from the model, i.e., two adjacent decoding layers in the decoder were directly connected. In addition, we only used the decoding feature from the last stage for segmentation; thus, the loss of the model only included the edge prediction loss and the final segmentation loss. The results in Table 7 suggest that compared with EAM-Net, EAM-Net without the mask attention modules reduces PA, Dice, and JA by 0.32%, 0.45%, and 0.54%, respectively. The abovementioned results suggest

TABLE 3: Results of EAM-Net and Res-UNet using different encoder backbones on the JSRT, Shenzhen, and Montgomery datasets.

Methods	JSRT			Shenzhen			Montgomery		
	PA (%)	Dice (%)	JA (%)	PA (%)	Dice (%)	JA (%)	PA (%)	Dice (%)	JA (%)
Res-UNet18	98.40	97.37	94.88	97.69	95.37	91.17	98.56	97.01	95.03
EAM-Net18	98.71	97.88	95.88	98.00	95.92	92.07	99.09	98.11	96.30
Res-UNet34	98.43	97.42	94.98	97.77	95.42	91.31	98.58	97.34	95.39
EAM-Net34	98.83	97.99	96.00	98.04	96.00	92.18	99.15	98.22	96.51
Res-UNet50	98.59	97.69	95.49	97.85	95.53	91.40	98.65	97.41	95.60
EAM-Net50	**98.96**	**98.11**	**96.15**	**98.20**	**96.13**	**92.27**	**99.16**	**98.23**	**96.52**

Bold values represent the the highest performance for each performance metric.

TABLE 4: Results of EAM-Net and the state-of-the-art lung region segmentation methods on the JSRT dataset.

Methods	PA (%)	Dice (%)	JA (%)
Bayesian feature pyramid network [21]	96.24	93.13	87.14
Kim and Lee [32]	98.17	96.68	93.57
Kholiavchenko et al. [22]	98.38	97.06	94.29
Yahyatabar et al. [17]	98.58	97.42	94.97
Novikov et al. [15]	98.55	97.36	94.86
SED [9]	98.62	97.51	95.14
EAM-Net	**98.96**	**98.11**	**96.15**

Bold values represent the the highest performance for each performance metric.

TABLE 5: Results of EAM-Net and the state-of-the-art lung region segmentation methods on the Shenzhen dataset.

Methods	PA (%)	Dice (%)	JA (%)
Feature selection with BN [10]	81.22	67.85	51.09
Feature selection with MLP [10]	87.82	77.85	64.74
Feature selection with RF [10]	89.19	80.25	68.15
Feature selection and vote [10]	91.06	83.65	73.04
Bayesian feature pyramid network [21]	96.17	93.04	87.00
X-ray-Net [23]	97.21	94.92	90.33
Kim and Lee [32]	97.49	95.45	91.26
EAM-Net	**98.20**	**96.13**	**92.27**

Bold values represent the the highest performance for each performance metric.

TABLE 6: Results of EAM-Net and the state-of-the-art lung region segmentation methods on the Montgomery dataset.

Methods	PA (%)	Dice (%)	JA (%)
Feature selection with BN [10]	78.06	62.31	43.97
Feature selection with MLP [10]	79.16	64.17	46.04
Feature selection with RF [10]	80.81	66.32	49.27
Feature selection and vote [10]	83.44	69.89	53.72
Bayesian feature pyramid network [21]	96.19	93.07	87.04
Souza et al. [8]	97.01	94.12	88.27
Rahman et al. [18]	96.84	94.25	89.13
ET-Net [20]	98.51	97.29	94.32
CFCM18 [24]	98.19	96.67	93.61
CFCM34 [24]	98.30	96.91	93.99
CFCM50 [24]	98.35	97.01	94.17
CFCM101 [24]	98.46	97.18	94.55
Yahyatabar et al. [17]	98.52	97.30	94.74
X-ray-Net [23]	98.57	97.44	94.93
EAM-Net	**99.16**	**98.23**	**96.52**

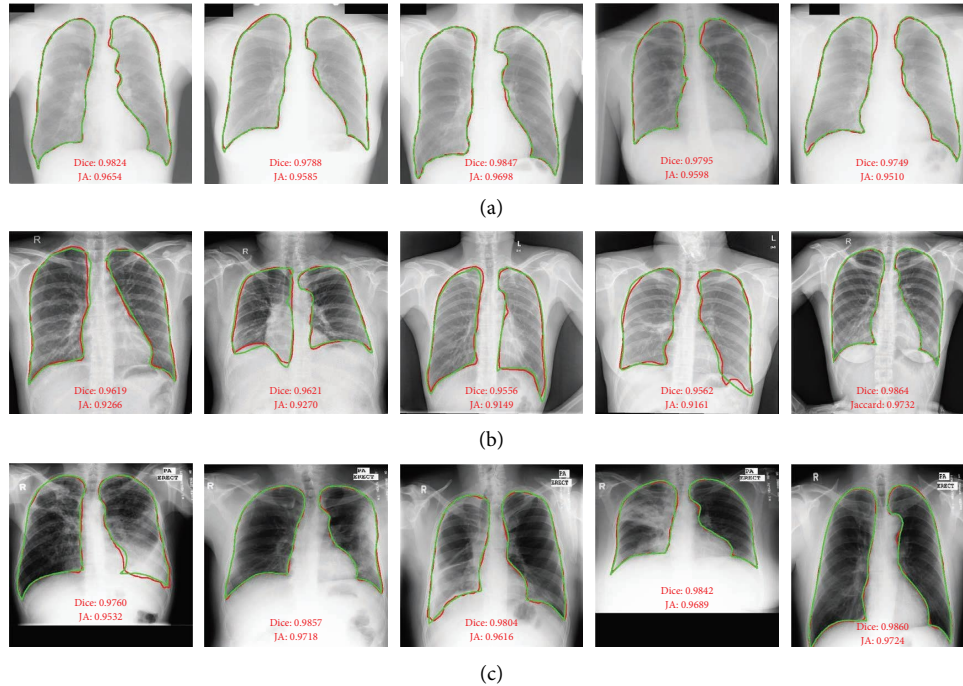Bold values represent the the highest performance for each performance metric.

FIGURE 6: Visualization of some segmentation results of EAM-Net on the JSRT, Shenzhen, and Montgomery datasets.

TABLE 7: Ablation study for the edge-assisted computing module, mask attention modules, and multiscale aggregation loss on the Montgomery dataset.

Methods	PA (%)	Dice (%)	JA (%)
EAM-Net without the edge-assisted computing module	98.81	97.79	96.05
EAM-Net without the mask attention modules	98.84	97.78	95.98
EAM-Net replacing the multiscale aggregation loss with the single-scale loss	98.89	97.85	96.07
EAM-Net	**99.16**	**98.23**	**96.52**

Bold values represent the the highest performance for each performance metric.

that the mask attention modules can highlight the lung regions on the segmentation features while suppressing irrelevant lesion-related artifacts, which is beneficial for the lung region segmentation task.

5.4.3. Effectiveness of the Multiscale Aggregation Loss. To investigate the effectiveness of the multiscale aggregation loss, we compared it with the single-scale loss, which includes an edge prediction loss and a final segmentation loss. Note that we still kept the mask attention operations in the decoder, but only performed the segmentation prediction at the last stage of the network instead of multiscale segmentation prediction. From Table 7, replacing the multiscale aggregation loss with the single-scale loss, drops Dice and JA by 0.38% and 0.45%, respectively. The abovementioned results demonstrate the effectiveness of the multiscale aggregation loss. In principle, the proposed multiscale segmentation loss belongs to deep supervision, which adds additional supervision (intermediate segmentation predictions) to the model, assisting the model in learning discriminative features to obtain performance gains.

6. Conclusion

Lung region segmentation is a fundamental and critical medical image analysis task. Existing lung region segmentation methods still suffer from inaccurate lung region segmentation boundaries, lesion-related artifacts, and limited ability to exploit multiscale information. In this paper, we proposed EAM-Net for lung region segmentation, which included an encoder-decoder structure and two new modules to address the abovementioned issues. First, an edge-assisted computing module was proposed, which integrated the shallow feature maps of the encoder for edge prediction, and transferred the edge prediction information to the segmentation task to optimize the edge parts of the segmentation results. We further designed a mask attention module after each decoding layer, which used the mask attention operation to enhance the lung regions while suppressing lesion-related artifacts. In addition, a multiscale aggregation loss was proposed to exploit multiscale segmentation information and jointly optimize the edge prediction and lung region segmentation tasks.

The results on the JSRT, Shenzhen, and Montgomery datasets confirmed that EAM-Net achieved better lung region segmentation performance than the baseline and state-of-the-art lung region segmentation methods. Our future work includes the following two aspects. On the one hand, we intend to extend EAM-Net to other medical image segmentation tasks. On the other hand, we plan to explore an end-to-end deep learning model for joint lung region segmentation and lung disease classification.

Abbreviations

CXR: Chest X-ray
CNNs: Convolutional neural networks
ReLU: Rectified linear unit
PA: Pixal accuracy.

Data Availability

The data that support the findings of this study are available on request from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] H. Kenmotsu, T. Naito, M. Kimura, A. Ono, T. Shukuya, and Y. Nakamura, "The risk of cytotoxic chemotherapy-related exacerbation of interstitial lung disease with lung cancer," *Journal of Thoracic Oncology*, vol. 6, no. 7, pp. 1242–1246, 2011.

[2] H. Zeng, W. Chen, R. Zheng, S. Zhang, J. S. Ji, and X. Zou, "Changing cancer survival in China during 2003–15: a pooled analysis of 17 population-based cancer registries," *Lancet Global Health*, vol. 6, no. 5, pp. e555–e567, 2018.

[3] S. Candemir and S. Antani, "A review on lung boundary detection in chest X-rays," *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, no. 4, pp. 563–576, 2019.

[4] J. D. Sweetlin, H. K. Nehemiah, and A. Kannan, "Feature selection using ant colony optimization with tandem-run recruitment to diagnose bronchitis from CT scan images," *Computer Methods and Programs in Biomedicine*, vol. 145, pp. 115–125, 2017.

[5] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: a Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2022.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, Boston, MA, USA, September, 2015.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, Singapore, September, 2015.

[8] J. C. Souza, J. O. Bandeira Diniz, J. L. Ferreira, G. L. França da Silva, A. Corrêa Silva, and A. C. de Paiva, "An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks," *Computer Methods and Programs in Biomedicine*, vol. 177, pp. 285–296, 2019.

[9] W. Yang, Y. Liu, L. Lin et al., "Lung field segmentation in chest radiographs from boundary maps by a structured edge detector," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 3, pp. 842–851, 2018.

[10] K. Santosh and S. Antani, "Automated chest X-ray screening: can lung region symmetry help detect pulmonary abnormalities?" *IEEE Transactions on Medical Imaging*, vol. 37, no. 5, pp. 1168–1177, 2018.

[11] H. Wang, H. Zhang, J. Hu, Y. Song, S. Bai, and Z. Yi, "DeepEC: an error correction framework for dose prediction and organ segmentation using deep neural networks," *International Journal of Intelligent Systems*, vol. 35, no. 12, pp. 1987–2008, 2020.

[12] B. Li, S. Liu, F. Wu, G. Li, M. Zhong, and X. Guan, "RT-Unet: an advanced network based on residual network and transformer for medical image segmentation," *International Journal of Intelligent Systems*, vol. 37, no. 11, pp. 8565–8582, 2022.

[13] B. Li, F. Wu, S. Liu et al., "CA-Unet++: an improved structure for medical CT scanning based on the Unet++ Architecture," *International Journal of Intelligent Systems*, vol. 37, no. 11, pp. 8814–8832, 2022.

[14] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[15] A. A. Novikov, D. Lenis, D. Major, K. Hlad Buhler, M. Wimmer, and K. Bühler, "Fully convolutional architectures for multiclass segmentation in chest radiographs," *IEEE Transactions on Medical Imaging*, vol. 37, no. 8, pp. 1865–1876, 2018.

[16] D. A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," 2015, https://arxiv.org/abs/1511.07289.

[17] M. Yahyatabar, P. Jouvet, and F. Cheriet, "Dense-Unet: a light model for lung fields segmentation in Chest X-Ray images," in *Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 1242–1245, IEEE, Montreal, QC, Canada, July, 2020.

[18] M. F. Rahman, T. L. B. Tseng, M. Pokojovy, W. Qian, B. Totada, and H. Xu, "An automatic approach to lung region segmentation in chest x-ray images using adapted U-Net architecture," in *Medical Imaging 2021: Physics of Medical Imaging*, vol. 11595, pp. 894–901, International Society for Optics and Photonics, 2021.

[19] D. Abas Hasan and A. Mohsin Abdulazeez, "A modified convolutional neural networks model for medical image segmentation," *Test Engineering and Management*, vol. 83, pp. 16798–16808, 2020 07.

[20] Z. Zhang, H. Fu, H. Dai et al., "A generic edge-attention guidance network for medical image segmentation," in

Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 442–450, Springer, Lima, Peru, October, 2019.

[21] R. Solovyev, I. Melekhov, T. Lesonen, E. Vaattovaara, O. Tervonen, and A. Tiulpin, "Bayesian feature pyramid networks for automatic multi-label segmentation of chest X-rays and assessment of cardio-thoracic ratio," in *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems*, pp. 117–130, Springer, Auckland, New Zealand, February, 2020.

[22] M. Kholiavchenko, I. Sirazitdinov, K. Kubrak et al., "Contour-aware multi-label chest X-ray organ segmentation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 3, pp. 425–436, 2020.

[23] M. Arsalan, M. Owais, T. Mahmood, J. Choi, and K. R. Park, "Artificial intelligence-based diagnosis of cardiac and related diseases," *Journal of Clinical Medicine*, vol. 9, no. 3, p. 871, 2020.

[24] F. Milletari, N. Rieke, M. Baust, M. Esposito, and N. Navab, "CFCM: segmentation via coarse to fine context memory," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 667–674, Springer, Granada, Spain, September, 2018.

[25] A. Kalinovsky and V. Kovalev, "Lung image segmentation using deep learning methods and convolutional neural networks," in *Proceedings of the Pattern Recognition and Information Processing*, pp. 21–24, Las Vegas, NV, USA, June, 2016.

[26] L. Saidy and C. C. Lee, "Chest X-ray image segmentation using encoder-decoder convolutional network," in *Proceedings of the 2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*, pp. 1-2, IEEE, Yilan, Taiwan, May, 2018.

[27] A. Mittal, R. Hooda, and S. Sofat, "LF-SegNet: a fully convolutional encoder–decoder network for segmenting lung fields from chest radiographs," *Wireless Personal Communications*, vol. 101, no. 1, pp. 511–529, 2018.

[28] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, Anchorage, Alaska, June, 2018.

[29] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, and Z. Fang, "Dual attention network for scene segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3146–3154, Long Beach, CA, USA, June, 2019.

[30] Y. B. Tang, Y. X. Tang, J. Xiao, and R. M. Summers, "XLSor: a robust and accurate lung segmentor on chest x-rays using criss-cross attention and customized radiorealistic abnormalities generation," in *Proceedings of the International Conference on Medical Imaging with Deep Learning PMLR*, pp. 457–467, Zurich, Switzerland, July, 2019.

[31] Z. Huang, X. Wang, L. Huang, C. Huang, Y. Wei, and W. Liu, "CCNet: criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 603–612, Seoul, Korea (South), November, 2019.

[32] M. Kim and B. D. Lee, "Automatic lung segmentation on chest X-rays using self-attention deep neural network," *Sensors*, vol. 21, no. 2, p. 369, 2021.

[33] F. Cao and H. Zhao, "Automatic lung segmentation algorithm on chest X-ray images based on fusion variational autoencoder and three-terminal attention mechanism," *Symmetry*, vol. 13, no. 5, p. 814, 2021.

[34] B. Li, G. Kang, K. Cheng, and N. Zhang, "Attention-guided convolutional neural network for detecting pneumonia on chest x-rays," in *Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4851–4854, IEEE, Berlin, Germany, July, 2019.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June, 2016.

[36] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19, Munich, Germany, September, 2018.

[37] J. Park, S. Woo, J. Y. Lee, and I. S. Kweon, "BAM: bottleneck attention module," in *Proceedings of the British Machine Vision Conference (BMVC) British Machine Vision Association (BMVA)*, Newcastle, UK, September, 2018.

[38] J. Wang, P. Lv, H. Wang, and C. Shi, "SAR-U-Net: squeeze-and-excitation block and atrous spatial pyramid pooling based residual U-Net for automatic liver segmentation in Computed Tomography," *Computer Methods and Programs in Biomedicine*, vol. 208, Article ID 106268, 2021.

[39] J. Shiraishi, S. Katsuragawa, J. Ikezoe et al., "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, 2000.

[40] B. Van Ginneken, M. B. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," *Medical Image Analysis*, vol. 10, no. 1, pp. 19–40, 2006.

[41] S. Jaeger, S. Candemir, S. Antani, Y. X. J. Wáng, P. X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative Imaging in Medicine and Surgery*, vol. 4, no. 6, pp. 475–477, 2014.