IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 April 2024, accepted 7 May 2024, date of publication 13 May 2024, date of current version 30 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3400119

RESEARCH ARTICLE

LMA-Net: Lightweight Multiple Attention Network for Multi-Source Heterogeneous Pulmonary CXR Segmentation

TURGHUNJAN MAMUT[1], LUN MENG[2], ZIYI PEI[3], TENGFEI WENG[2], QI HAN[2], (Member, IEEE), KEPENG WU[2], XIN QIAN[2], HONGXIANG XU[2], ZICHENG QIU[2], YUAN TIAN[2], AND YANGJUN PEI[2]

[1]College of Information Engineering, Tarim University, Alar, Xinjiang 843300, China
[2]College of Intelligent Technology and Engineering, Chongqing University of Science and Technology, Chongqing 401331, China
[3]College of Materials Science and Engineering, Chongqing University of Arts and Sciences, Chongqing 402160, China

Corresponding author: Lun Meng (melomeng@qq.com)

This work was supported in part by the West Light Foundation of Chinese Academy of Science; in part by the Research Foundation of the Natural Foundation of Chongqing City under Grant cstc2021jcyj-msxmX0146 and Grant cstc2021jcyj-msxmX1212; in part by Scientific and Technological Research Program of Chongqing Municipal Education Commission under Grant KJQN202301517, Grant HZ2021015, Grant KJZD-K202100104, and Grant KJQN202301543; in part by Chongqing Science and Technology Military-Civilian Integration Innovation Project, in 2022; in part by Bingtuan Science and Technology Program in China under Grant 2021AB026; and in part by Shanxi Province Applied Basic Research Program, China, under Grant 202203021211116.

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT The automatic pulmonary segmentation for chest X-ray(CXR) plays an important role in assisting diagnosis. Many deep learning methods have the problems of high computational complexity and low segmentation accuracy, which hinder the application to clinical workstations. Therefore, this paper proposes a lightweight multiple attention network(LMA-Net), which improved U-Net by using the progressive dilated convolution(PDC) for lightweight. A reinforced channel attention(RCA) and a multiscale attention(MSA) are embedded in the decoder to further improve the network segmentation performance. We fuse four types of pulmonary disease CXR from the COVID-QE-Ex dataset to generate a multi-source heterogeneous dataset. Effectiveness of LMA-Net is shown by achieving Intersection over Union(*IoU*) of 96.28%, *Dice* of 96.95%, Average symmetric surface distance(*ASSD*) of 13.11mm and Hausdorff Distance 95th percentile(*HD*95) of 81.12mm, respectively. It can be seen that lightweight of LMA-Net is achieved according to parameter(Param) of 2.89M and floating-point operations(FLOPs) of 2.64G. This method can effectively improve segmentation performance and speed.

INDEX TERMS Convolutional neural network, attention, lightweight, multi-source heterogeneous dataset, medical image segmentation.

I. INTRODUCTION

Due to the pandemic of COVID-19, pulmonary diseases have received more attention. It is highly sensitive and efficient to use medical image for pulmonary diseases. Compared to CT, chest X-ray(CXR) is widely applied to diagnose various pulmonary diseases due to lower cost, lower radiation and faster speed [1].

As shown in FIGURE 1, not every CXR is standardized. Pulmonary segmentation becomes challenging due to several factors: (1) non-pathological changes: the shape and size of the pulmonary vary with age, gender and heart size [2]; (2) pathological changes:the opacity caused by severe pulmonary disease reaches a high-intensity value [3];

The associate editor coordinating the review of this manuscript and approving it for publication was Diego Oliva.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

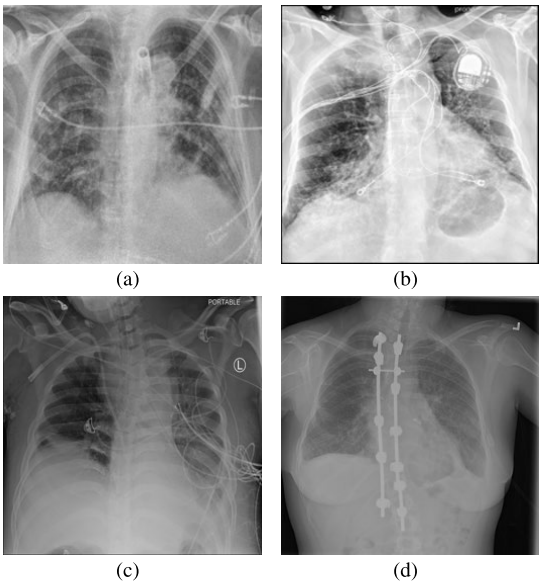


FIGURE 1. There are four non-standard CXR images.

(3) foreign body coverage: the pulmonary field is obscured by the patient's clothes or medical equipment (pacemaker, infusion line, medical catheter) [4]. For example, the medical device implanted in the body affects the lung imaging in FIGURE 1(a) and FIGURE 1(b). The boundary between the albino lung area and the normal lung area in FIGURE 1(b) and FIGURE 1(c) is blurred. In Figure 1(d), there is a significant difference in lung morphology between females and males. These unstable factors can cause delays and misdiagnosis. Artificial intelligence partnering with massive data will improve this dilemma [5].

Pulmonary segmentation is a crucial step in quantitative analysis of CXR in computer-aided medical diagnostic systems [6], [7]. This initial step significantly affects the performance of downstream analysis, such as anomaly detection or classification of lung diseases, such as cancer [7], [8].

With the development of deep learning methods, state-of-the-art convolutional neural network(CNN) can automatically learn ROI from a large dataset [9], which can solve the delay and misdiagnosis caused by non-standard images. However, the existing CNNs commonly used in medical image processing usually have high computing costs, especially on graphic processing unit(GPU). Deploying deep learning models to medical workstations with limited resources would lead to various restrictions for real-time analysis. Therefore, on the premise of the accurate prediction, it is necessary to further promote the development of models towards lightweight and reduce the complexity in time and space.

To address these challenges, this paper proposes a lightweight multiple attention network(LMA-Net) for pulmonary segmentation. Experimental results show that the proposed method can effectively segment pulmonary region from CXRs and achieve lightweight. The main contributions of this paper are as follows.

- We design LMA-Net that incorporates four types of pulmonary CXR to automatically segment pulmonary region. This model can solve the delay and misdiagnosis caused by non-standard images.
- Based on the encode-decode structure of U-Net [10], we extend dilated convolution [11] to progressive dilated convolution(PDC). In LMA-Net, PDC gradually expands with the network deepening. Small-dilated convolution effectively captures fine details in low-level features, and large-dilated convolution captures semantic information in high-level features. What's more, PDC develops the structure of depthwise separable convolution [12] to achieve ultra lightweight, which is crucial aspect in solving the problem of limited clinical computing resources.
- To make full use of channel information in feature map, we propose reinforced channel attention(RCA). Every layer of decoder are embeded with RCA to handle the concatenation of high-level and low-level features. The RCA compresses the concatenated feature maps to obtain the weights for each channel. Subsequently, residual connection [13] is introduced to combine the concatenated feature maps with the corresponding weights several times to self-adaptively reinforce the feature channel information and suppress irrelevant features. The RCAs in different layers enable LMA-Net to focus on the target features.
- Multi-scale attention(MSA) is introduced to subtly fuse features at different scales to cope with feature loss, which allows the fine details in low-level features and the semantic information in high-level features to be fully utilized. In addition, in order to overcome the problem of class imbalance within feature channels, a Channel Factor Enhancement(CFE) module is proposed based on dimension transformation, which can automatically calibrates the target region prediction to improve segmentation performance.

II. RELATED WORK

In this paper, CNN and attention mechanism are introduced for medical image segmentation.

A. U-NET VARIANTS FOR MEDICAL IMAGE SEGMENTATION

Deep learning methods such as CNN [14] have excellent performance for medical image segmentation tasks in recent years. Fully convolutional network(FCN) [15] is a successful network in image segmentation. Inspired by the encoder-decoder architecture of FCN, subsequently, the 2D U-Net [10] has been developed and widely implemented for medical image segmentation tasks. Since the invention of U-Net, many improved networks based on U-Net have great performance for medical image segmentation tasks.

In [16], R2U-Net referred to the idea of recurrent CNN and residual CNN to gain the precise segmentation results. The U-Net has also been extended into attention modules such as Attention U-Net [17], which introduced the Attention gates(AGs) and replaced hard-attention with soft-attention. U-Net++ [18] is a nested architecture of U-Net, which adopts dense connection to eliminate gradient problems, reuse feature and enhance feature propagation. Trans U-Net [19] is the combination of transform and U-Net, where transform can provide global self-attention based on the U-Net. These methods show quite high performance for medical image segmentation tasks. However, these state-of-the-art models have problems of high complexity and parameter quantity. There are various limitations in deploying above models to workstations with limited computing resources.

B. ATTENTION MECHANISM

Attention mechanism originates from human vision research and is often used in computer vision research [20]. The attention mechanism simulates the human behavior of paying attention to a few more important words in the process of reading [21], [22]. In the research of computer vision, in order to make full use of limited resources and focus on specific relevant feature information, the attention mechanism is realized through dynamic adaptive weighting of feature information [23].

SEnet [24] automatically strengthens channel information of features through learning, and uses the obtained importance to enhance features and suppress features that are not important to the current task. Woo et al. [25] creatively proposed CBAM, which combines the attention mechanism of channels and spaces, and automatically obtains the importance of each feature space to enhance features related to segmented targets and suppress unimportant features in the current task. With the development of deep learning, recently many new attention mechanisms not only focus on performance but also on lightweight. Wang et al. [26] not only proposed an effective channel attention module ECA, which realized lightweight, but also achieved significant performance gains. Hou et al. [27] proposed a coordinate attention(CA) that is new efficient attention mechanism, which can encode the horizontal and vertical location information into the channel attention, so that the network can focus on a wide range of location information without too much computation. Dai et al. [28] proposed a trainable second-order channel attention (SOCA) module, which adaptively rescale the channel-wise features by using second-order feature statistics for more discriminative representations. Yang et al. [29] proposed parameter-free attention module(SimAM), which inferred 3-D attention weights for the feature map in a layer without adding parameters to the original networks. D2-Net [30] address the problem of finding reliable pixel-level correspondences under difficult imaging conditions. LCNet [31] introduces a partial-channel transformation (PCT) strategy to minimize computing latency and hardware requirements of the basic unit.

In conclusion, existing attention mechanisms not only achieve high performance, but are also moving towards lightweight. Therefore, following the current development steps, we propose a lightweight network based on U-Net and multiple attentions. This method can effectively improve segmentation performance and speed.

III. METHODS

A. LMA-NET: LIGHTWEIGHT MULTIPLE ATTENTION NETWORK

In medical image segmentation, U-Net preserves high-level semantic features and low-level spatial details by using a skip connected symmetric encoder-decoder architecture, which is crucial for accurately dividing organ boundaries and fine structures [10]. The proposed methodology utilizes its basic architecture to construct a lightweight multiple attention network (LMA-Net). The LMA-Net has been specifically designed to facilitate the accurate localization of the elusive lung area, and subsequently, deftly execute CXR image segmentation with clinical precision. This network has skip connections which can fuse different scale features, improves the convolutional modules to achieve lightweight, and uses multiple attention modules to improve segmentation performance.

The structure of LMA-Net is shown in FIGURE 2, where LMA-Net innovatively proposes progressive dilated convolution(PDC) [11] and reinforced channel attention(RCA) [24] to replace the classical convolutional modules of five scales. The proposed PDC improves the strategy in [11] to achieve lightweight and avoid the problem about the local information loss due to the excessive expansion of the dilated convolutional kernel. The PDC captures more global information while preserving local information loss. The proposed RCA improves the dual-channel feature fusion, and adopts residual connection [13] to avoid gradient disappearance and explosion caused by network deepening. To avoid the loss of significant information during decoding, the LMA-Net aggregates the multi-scale feature maps by using channel connection. Then the target features can be extracted by using the multiscale attention (MSA).

The proposed LMA-Net uses five PDCs to replace the classical convolutional modules in the encoder, and the dilated rate d increases with the network deepening. The proposed RCA accepts the cascade features of the low-level features from the encoder and high-level features from the decoder, so as to obtain more relevant channel weight factors. In addition, four output feature maps from RCA are spliced, and the proposed MSA is used to process the cascade feature map to obtain target features map.

B. PROGRESSIVE DILATED CONVOLUTION

Due to the complexity and uncertainty of the CXR images, in general, CXR images will contain many different organs or tissues. It is not easy to distinguish these organs and tissues

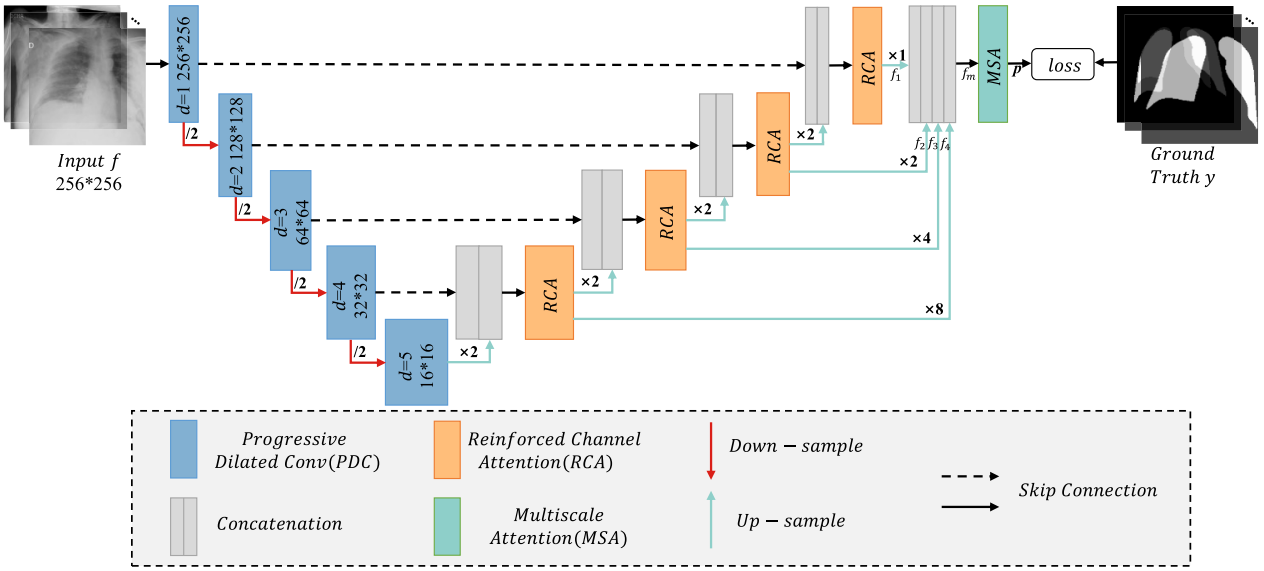


FIGURE 2. The proposed lightweight multiple attention network(LMA-Net). Blue rectangle and *d* correspond the progressive dilated convolution and expansion rate, respectively. We use four reinforced channel attention(RCA) to replace the classical convolutional modules. Green rectangle MSA represents multiscale attention.

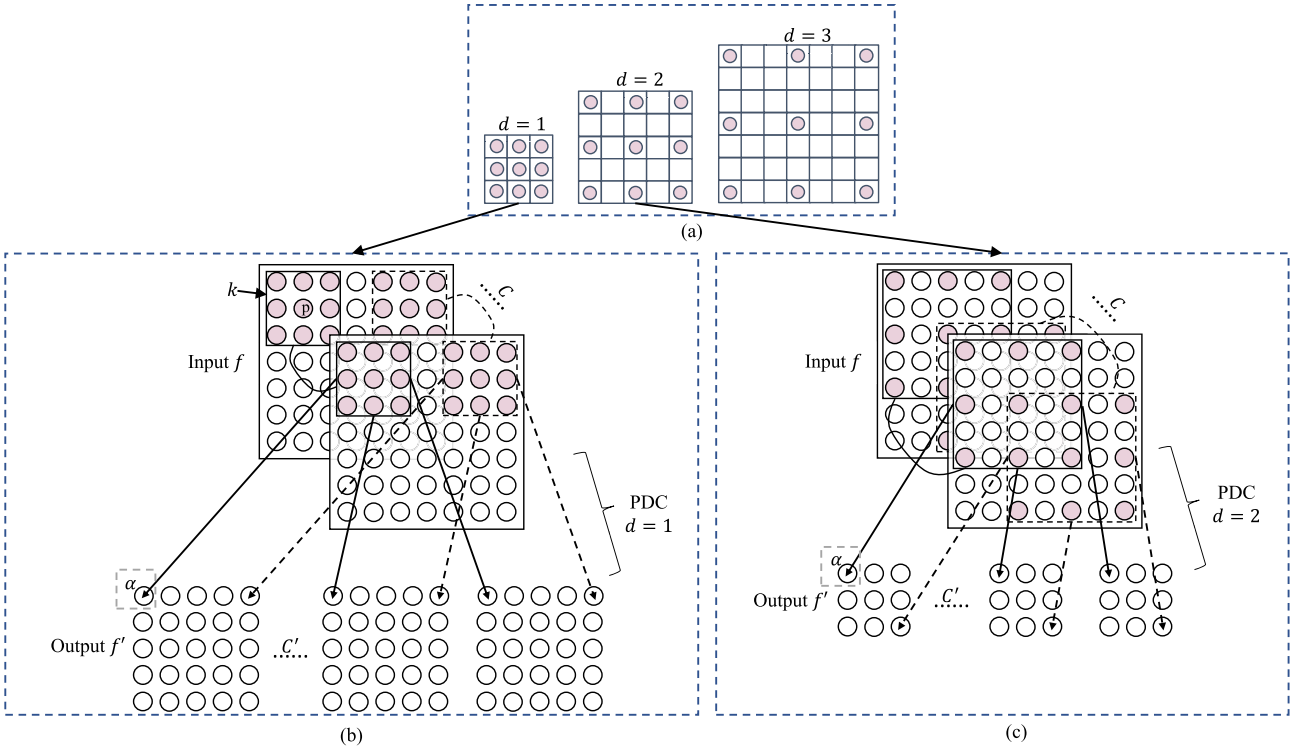


FIGURE 3. (a)The dilated convolutional kernel with the shape of 3 ∗ 3 under dilated rate *d* of 1, 2 and 3. (b)The classical convolutional operation. (c)The dilated convolutional operation with $d = 2$.

in clinical situations. The classical convolutional kernel would ignore the organ features with large receptive field while processing medical images in the forward propagation process.

Therefore, the LMA-Net uses dilated convolution [11] to improve the classical convolutional kernel to solve the problem of large receptive field feature loss, but retains the basic structure of depthwise separable convolution. We set the *groups* of depthwise separable convolution as the greatest common divisor of input channel number and output channel number to minimize the Param and FLOPs.

As shown in FIGURE 3(a), the dilated convolutional kernel can be seen as an expansion of the classical convolutional kernel, where the dilated rate is expressed by a positive integer d. As shown in FIGURE 3(b), the process of the classical convolution can be expressed as:

$$\alpha = k \circledast f = \sum_{l=1}^{C} \sum_{i=-r}^{r} \sum_{j=-r}^{r} k(i,j) f_l(p_x + i, p_y + j) \quad (1)$$

where k represents the convolutional kernel of size $(2r + 1)(2r + 1)$. The input $f \in \mathbb{R}^{C*H*W}$ in FIGURE 3 contains C feature channels, and H, W represent the height and width of input, respectively($C = 1$ in section III-B). $\circledast$ represents the convolutional operation. $k(i, j)$ represents an element in the convolutional kernel k, where $(i, j) \in \{(i, j)|i \in [-r, r], j \in [-r, r], r \in N\}$. (p_x, p_y) represents the coordinates of a pixel in the input f. $f_l(p_x + i, p_y + j)$ represents a pixel of the l-th channel in input f. Eventually, the result α represents a pixel point in the output f', where α is obtained by sliding the convolutional kernel k once in the input f.

As shown in FIGURE 3(a), the dilated rate d can separate and expand receptive field of the convolutional kernel k. The dilated convolution helps the network learn more contextual features, such as lung and lung lesion area, without increasing the time and space complexity.

In LMA-Net, the dilated rate d increases with the network deepening. The process of PDC is shown in FIGURE 3(c) when $d = 2$, which can be expressed as:

$$\alpha = k \circledast_{d=2} f = \sum_{l=1}^{C} \sum_{i=-2r}^{2r} \sum_{j=-2r}^{2r} k(i,j) f_l(p_x + i, p_y + j) \quad (2)$$

where $\circledast_{d=2}$ is convolutional operation with dilated rate of $d = 2$.

From FIGURE 2, FIGURE 3(c) and EQUATION 2, we can see that the excessive increase in dilated rate inevitably leads to the loss of local image information [11]. To address this issue, we propose an improved strategy where the dilated rate slowly increases with the network deepening, ensuring that more global image information is captured while retaining more local fine detail.

C. REINFORCED CHANNEL ATTENTION

Inspired by CBAM [25], we propose a reinforced channel attention(RCA) module which is shown in FIGURE 4. The proposed RCA is used to replace the classical convolutional module in the decoder, which can make full use of the channel information. The decoding process often involves upsampling and concatenation operations, which can lead to a loss of fine-grained details and spatial context. RCA modules strategically placed in the decoder can help mitigate this issue by recalibrating channel-wise feature responses based on their global dependencies, ensuring that important details are preserved and emphasized during the upsampling process.

First of all, let $f_r \in \mathbb{R}^{2C*H*W}$ represent the concatenated feature map. A $3*3$ convolution is used to smooth f_r obtained

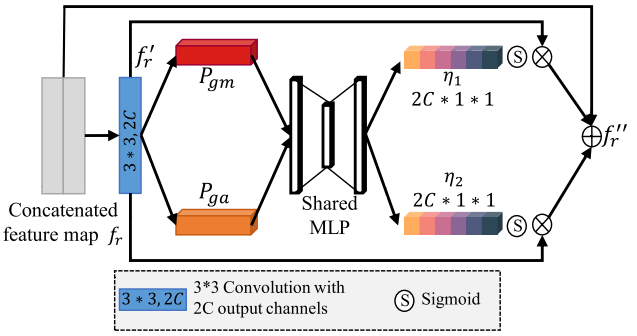


FIGURE 4. Structure of reinforced channel attention(RCA) module with residual connection. The η_1 and η_2 represent the channel factors.

of two different f_p' to get f_r', and the process is expressed as:

$$f_r' = conv(f_r) \quad (3)$$

Subsequently, RCA implements global average pooling(P_{ga}) and global maximum pooling(P_{gm}) to compress the dimension of f_r' to $2C * 1 * 1$ in parallel. After pooling, two different feature maps with size $2C * 1 * 1$ are sent to the shared multilayer perceptron(MLP) to get two channel factors η_1 and $\eta_2(\eta_1, \eta_2 \in [0, 1]^{2C*1*1})$. MLP consists of two fully connected layers, with the ReLU function after the first layer and the Sigmoid function after the second layer. The above process is represented as:

$$\{\eta_i\}_{i=1}^{2} = MLP(P_{gm}(f_r'), P_{ga}(f_r')) \quad (4)$$

The channel factors can guide feature maps to automatically highlight the relevant feature channels and restrain the irrelevant feature channels. Different from the channel attention in CBAM [25], we introduce residual connection [13] in RCA to avoid gradient explosion and gradient disappearance due to network deepening. Through residual connection, η_1 and η_2 are multiplied by f_r' from convolutional operation respectively. Eventually, the results of multiplication is added by f_r pixel-wisely to obtain $f_r'' \in \mathbb{R}^{2C*H*W}$, and the above process is expressed as:

$$f_r'' = \eta_1 * f_r' + \eta_2 * f_r' + f_r \quad (5)$$

D. MULTISCALE ATTENTION

Benefiting from multiscale architecture of U-Net, the nearest interpolation is utilized to up-sample four feature maps from RCA of different scales, as shown in FIGURE 2. After up-sampling, four feature maps $\{f_i\}_{i=1}^{4}$ are gotten with a same dimension of $C * H * W(C = 4$ in section III-D). Concatenating $\{f_i\}_{i=1}^{4}$ channel-wisely and getting $f_m \in \mathbb{R}^{4C*H*W}$, the f_m represents the input feature map of multiscale attention(MSA) module. The process of generating f_m is specifically expressed as:

$$f_m = Concat(f_1, f_2, f_3, f_4) \quad (6)$$

The first module of MSA is the ChannelGate, as shown in FIGURE 5(a). To highlight the channel correlation of

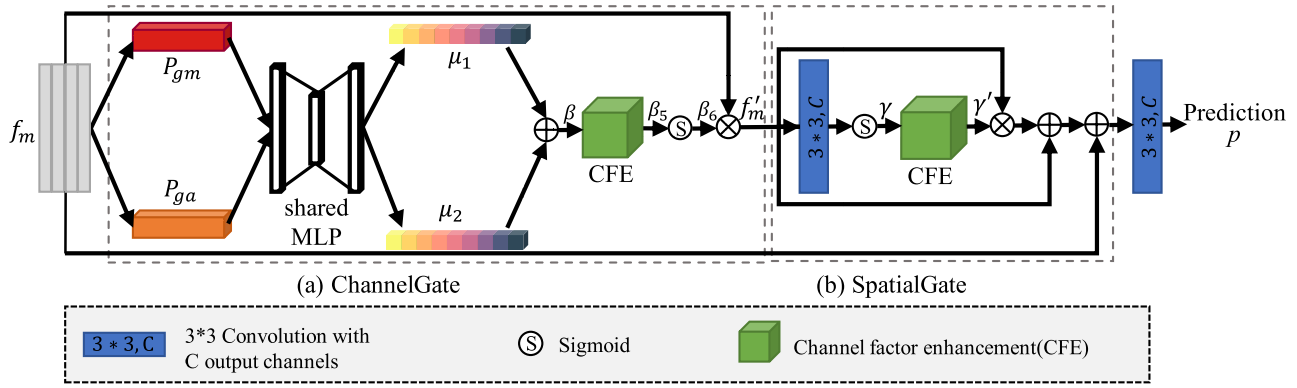


FIGURE 5. Structure of the proposed multiscale attention(MSA) module with residual connection. Red rectangle represents global maximum pooling(P_{gm}). Orange rectangle represents global average pooling(P_{ga}). The μ_1 and μ_2 represent the channel factors. Green rectangle represents channel factor enhancement(CFE) module.

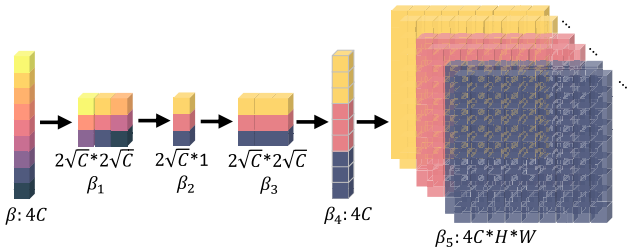


FIGURE 6. The overlap-adding procedure.

concatenated feature maps f_m, f_m is sent to global average pooling(P_{ga}), global maximum pooling(P_{gm}) and MLP to obtain channel factors $\{\mu_i\}_{i=1}^{2} \in \mathbb{R}^{4C}$. The channel factors can be used to distinguish the importance of channels. Subsequently, μ_1 and μ_2 are added to obtain $\beta \in \mathbb{R}^{4C}$, which serves as input to channel factor enhancement(CFE) module. The transformation of β is shown in FIGURE 6.

Specifically, CFE utilizes compression and expansion operations to enhance the information of different dimensions in the feature map f_m. Firstly, β is reshaped to obtain $\beta_1 \in \mathbb{R}^{2\sqrt{C}*2\sqrt{C}}$($C = 4$ in section III-D) without changing the amount and value of the data. Subsequently, all elements of $\beta_1(i,:)$ in β_1 are summed and averaged to obtain $\beta_2 \in \mathbb{R}^{2\sqrt{C}*1}$, $i = 1, 2, \ldots, 2\sqrt{C}$. After the above compression, each pixel in β_2 captures more contextual information. Next, β_2 is replicated $2\sqrt{C}$ times to obtain $\beta_3 \in \mathbb{R}^{2\sqrt{C}*2\sqrt{C}}$, and then β_3 is reshaped to obtain $\beta_4 \in \mathbb{R}^{4C}$ without changing the amount and value of the data. Therefore, the dimensions of β_4 are the same as the dimensions of β. To fit the dimensions of the MSA input f_m, each $\beta_4(j,0)$ are replicated $H * W$ times, so C single-channel feature maps of size $H * W$ is obtained, where $H = W$ and $j = 1, 2, \ldots, 4C$. Finally, the single-channel feature maps are concatenated channel-wisely to obtain $\beta_5 \in \mathbb{R}^{4C*H*W}$. The process of channel factor enhancement(CFE) is specifically described in Algorithm 1.

After dimension-based transformation of CFE, β_5 is fed into *sigmoid* to obtain the enhanced channel factor $\beta_6 \in [0,1]^{4C*H*W}$. Eventually, we use Hadamard product to multiply the input f_m and the β_6 pointly to obtain $f'_m \in \mathbb{R}^{4C*H*W}$.

The second module of MSA is the SpatialGate, which is used to achieve spatial attention, as shown in FIGURE 5(b). MSA consists of two $3 * 3$ convolutional modules and one CFE module. Firstly, the SpatialGate uses $3 * 3$ convolutional module to smooth f'_m, which is then activated by sigmoid to obtain $\gamma \in [0,1]^{C*H*W}$. Subsequently, the γ is processed by use of the CFE to obtain $\gamma' \in [0,1]^{4C*H*W}$. In addition, we use residual connections to recover feature information loss as the network deepening. Eventually, γ' is processed using $3 * 3$ convolutional module to obtain prediction $p \in \mathbb{R}^{C*H*W}$, which is used to calculate the loss function with the ground truth, as shown in FIGURE 2.

E. LOSS

BCEWithLogitsLoss is a loss function for binary classification. It combines the sigmoid function with the binary cross entropy(BCELoss). p is the feature map predicted by the LMA-Net. Before calculating the loss, we activate p using sigmoid to obtain $\bar{p} \in [0,1]^{C*H*W}$. In binary classification problems, the pixel y_i in the ground truth y is usually 0 or 1. The *BCEWithLogitsLoss* formula is expressed as:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot log(\sigma(p_i)) + (1 - y_i) \cdot log(1 - \sigma(p_i))\right] \tag{7}$$

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{8}$$

where $\mathcal{L}$ is the *BCEWithLogitsLoss*, y_i is the i-th pixel in the ground truth y and p_i is the i-th pixel in the prediction p, and N represents the number of samples and $\sigma()$ is the sigmoid function.

IV. EXPERIMENTS

A. DATASETS

Experiments are conducted on the public dataset COVID-QU-Ex. The researchers of Qatar University have compiled the COVID-QU-Ex dataset, which consists of 34,613 chest

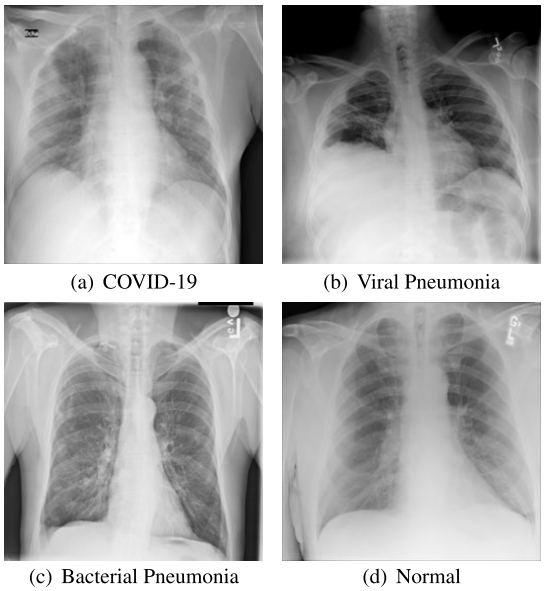
Algorithm 1 Channel Factor Enhancement(CFE)

Input: feature map $\beta \in \mathbb{R}^{4C}$
Output: feature map $\beta_4 \in \mathbb{R}^{4C*H*W}$

$count = 0$
for i = 0 to $2\sqrt{C} - 1$ **do**
  for j = 0 to $2\sqrt{C} - 1$ **do**
    $\beta_1[i, j] = \beta[count, 0]$
    $count = count + 1$
  end for
end for
$count = 0$
for i = 0 to $2\sqrt{C} - 1$ **do**
  for j = 0 to $2\sqrt{C} - 1$ **do**
    $count = count + \beta_1[i, j]$
  end for
  $\beta_2[i, 0] = count \ / \ j$
end for
for i = 0 to $2\sqrt{C} - 1$ **do**
  for j = 0 to $2\sqrt{C} - 1$ **do**
    $\beta_3[i, j] = \beta_2[i, 0]$
  end for
end for
$count = 0$
for i = 0 to $2\sqrt{C} - 1$ **do**
  for j = 0 to $2\sqrt{C} - 1$ **do**
    $\beta_4[count, 0] = \beta_3[i, j]$
    $count = count + 1$
  end for
end for
for i = 0 to 4C **do**
  for j = 0 to H **do**
    for k = 0 to W **do**
      $\beta_5[i, j, k] = \beta_4[count, 0]$
    end for
  end for
  if i > 0 **then**
    Channelconcatenation
  end if
end forreturn β_5

X-ray (CXR) images, including 11,956 lung images of COVID-19 infection, 5,897 lung images of Viral Pneumonia(VP), 6,059 lung images of Bacterial Pneumonia(BP) and 10,701 normal lung images. Four typical images of these four categories are shown in FIGURE 7. Ground truths are provided for the entire dataset. This is the largest ever created lung mask dataset. The download address of dataset COVID-QU-Ex is https://www.kaggle.com/datasets/anasmohammedtahir/covidqu.

The number of training set, validation set and testing set is shown in TABLE 1 for four types of CXR images. In our study, we carefully considered the balance between these subsets to ensure that the model could efficiently and quickly learn from the training set, avoid overfitting through validation, and provide a fair assessment of its usability via the testing set. Therefore, we randomly collect and split the dataset into 4000, 480 and 480 for training, validation and testing in three different datasets respectively.



(a) COVID-19 (b) Viral Pneumonia (c) Bacterial Pneumonia (d) Normal

FIGURE 7. (a) to (d) represent four types of input images, which are COVID-19, Viral Pneumonia(VP), Bacterial Pneumonia(BP) and Normal respectively.

TABLE 1. Distribution of dataset.

Set	COVID-19	VP	BP	Normal	Total
Training	1000	1000	1000	1000	4000
Validation	120	120	120	120	480
Testing	120	120	120	120	480

To avoid training termination due to inconsistent images, we preprocess the entire dataset before training, such as grayscale, tensor transformation, and random cropping. Preprocessing ensures pulmonary image size of 1*256*256.

B. IMPLEMENTATION DETAILS

The LMA-Net is implemented using Pytorch framework. For network training, the BCEWithLogitsLoss is used as the loss function, while adaptive moment estimation(Adam) optimization with initial learning rate of 1e-3, standard beta values of (0.9, 0.999) and eps of 1e-8 is applied to minimize this loss. The LMA-Net is trained for 300 epochs with a batch size of 8. All experiments are conducted on GeForce RTX 3090 with 24 GB of memory.

C. EVALUATION INDICATORS

Specifically, the LMA-Net is evaluated by use of intersection over union(IoU), dice coefficient($Dice$), average symmetric surface distance($ASSD$) and 95th percentile of the hausdorff distance($HD95$). The IoU and $Dice$ are respectively

defined as

$$IoU = \frac{|p \cap y|}{|p \cup y|} \tag{9}$$

$$Dice = \frac{2|p \cap y|}{|p| + |y|} \tag{10}$$

where p and y denote prediction and ground truth respectively,

Both *ASSD* and *HD*95 are used to calculate the surface distance and measure the accuracy of the segmentation boundary. The *ASSD* calculates the average distance between predicted boundary and ground truth boundary, while *HD*95 calculates the maximum distance between the two boundaries. *ASSD* and *HD*95 are defined as

$$ASSD = \frac{\sum_{a \in S_a} min_{b \in S_b} d(a,b) + \sum_{b \in S_b} min_{a \in S_a} d(b,a)}{len(S_a) + len(S_b)} \tag{11}$$

$$HD95 = 0.95max[\max_{a \in S_a} \min_{b \in S_b} d(a,b), \max_{b \in S_b} \min_{a \in S_a} d(a,b)] \tag{12}$$

where S_a and S_b indicate the predicted segmentation boundary and the manual segmentation boundary respectively. Both a and b indicate pixels on the boundary, and $d(\cdot,\cdot)$ is distance function. $len()$ represents the sum of pixels that make up the boundary S_a or S_b.

D. EXPERIMENTAL RESULTS

This paper takes COVID-QU-Ex as the research object and carries out four groups of experiments. In order to verify the effectiveness of each module in LMA-Net, ablation experiments are conducted. In addition, LMA-Net is compared with the current advanced segmentation network.

TABLE 2. Ablation experiments for segmentation results.

Network	*IoU*(%)	*Dice*(%)	*ASSD*(mm)	*HD*95(mm)
Baseline	95.85	96.26	16.63	83.37
PDC	95.80	96.77	16.90	83.59
RCA	95.31	96.80	13.29	81.07
MSA	94.08	96.77	**12.32**	**79.88**
PDC+RCA	95.29	95.24	14.17	81.74
PDC+MSA	95.22	96.77	14.80	81.76
RCA+MSA	94.75	96.17	13.44	81.14
LMA-Net	**96.28**	**96.95**	13.11	81.12

1) ABLATION EXPERIMENTS FOR SEGMENTATION RESULTS

The proposed LMA-Net takes the U-Net as the baseline, where LMA-Net includes progressive dilated convolution (PDC), reinforced channel attention(RCA) and multiscale attention(MSA). To verify the effectiveness of combining different network, we compared LMA-Net with six variants of different combinations of PDC, RCA and MSA. Specifically, PDC means progressive dilated convolution used only in the encoder of the baseline. RCA represents reinforced channel attention used only in the decoder of the baseline. MSA represents multiscale attention used only in decoder of the baseline.

TABLE 2 presents quantitative comparison of the LMA-Net and other variants lung segmentation, where *IoU*, *Dice*, *ASSD* and *HD*95 are adopted to evaluate the segmentation effect. It can be observed that LMA-Net has the highest score of 96.28% and 96.95% respectively in *IoU* and *Dice*. At the same time, in the comparison between *ASSD* and *HD*95, LMA-Net is also close to the best MSA, reaching 13.11mm and 81.12mm respectively. FIGURE 8 shows the visual comparison of different CNNs dealing with CXR segmentation task.

2) ABLATION EXPERIMENTS FOR COMPUTING COST

We randomly generate a tensor $f \in \mathbb{R}^{1*256*256}$ as the input of the segmentation models to test computing cost and inference time. Then, the parameters(Param), floating point operations(FLOPs) and inference time of networks can be tested. As show in TABLE 3, the LMA-Net outperforms all other variants in Param, FLOPs and inference time, and the corresponding values are 2.89M, 2.64G and 609.31ms, respectively. It follows that LMA-Net can effectively reduce the number of parameters and floating point operations in the segmentation process. This method can effectively improve segmentation speed.

TABLE 3. Ablation experiments for computing cost.

Network	Param(M)	FLOPs(G)	Time(ms)
U-Net	3.91	7.72	724.38
PDC	3.80	7.52	701.87
RCA	3.59	4.96	779.68
MSA	3.92	7.99	781.04
PDC+RCA	3.59	4.69	763.52
PDC+MSA	3.92	7.99	813.43
RCA+MSA	3.60	4.96	903.81
LMA-Net	**2.89**	**2.64**	**609.31**

3) COMPARISON WITH OTHER METHODS FOR SEGMENTATION RESULTS

LMA-Net is compared with seven state-of-the-art methods which are U-Net, FCN, DenseNet, U-Net++, Attention U-Net, R2U-Net and MFM-Net. These models are all retrained on COVID-QU-Ex.

As shown in TABLE 4, segmentation results of the segmentation models are listed in the contrastive

TABLE 4. Comparison with other methods for segmentation results.

Network	*IoU*(%)	*Dice*(%)	*ASSD*(mm)	*HD*95(mm)
U-Net [10]	93.85	94.26	17.63	85.37
FCN [15]	92.70	93.70	16.60	85.21
DenseNet [13]	92.73	93.21	16.57	84.22
U-Net++ [18]	93.88	93.91	15.36	83.54
Attention U-Net [17]	94.84	94.11	16.69	83.68
R2U-Net [16]	95.20	95.88	13.76	82.30
MFM-Net [32]	93.45	92.78	14.59	84.12
LMA-Net	**96.28**	**96.95**	**13.11**	**81.12**

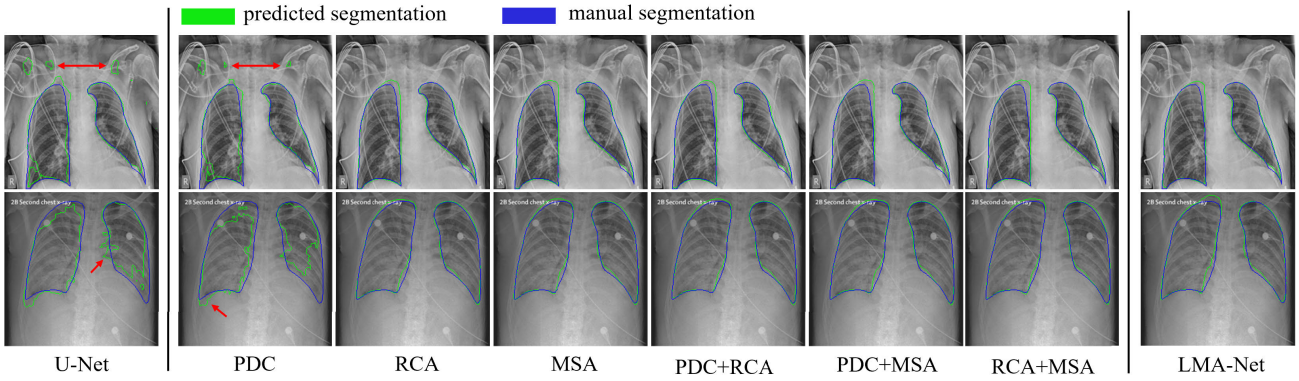


FIGURE 8. Visual comparison of ablation experiments for lung segmentation. The red arrows highlight some mis-segmentations. The blue border is the result of manual segmentation. The green border is the predicted segmentation result.

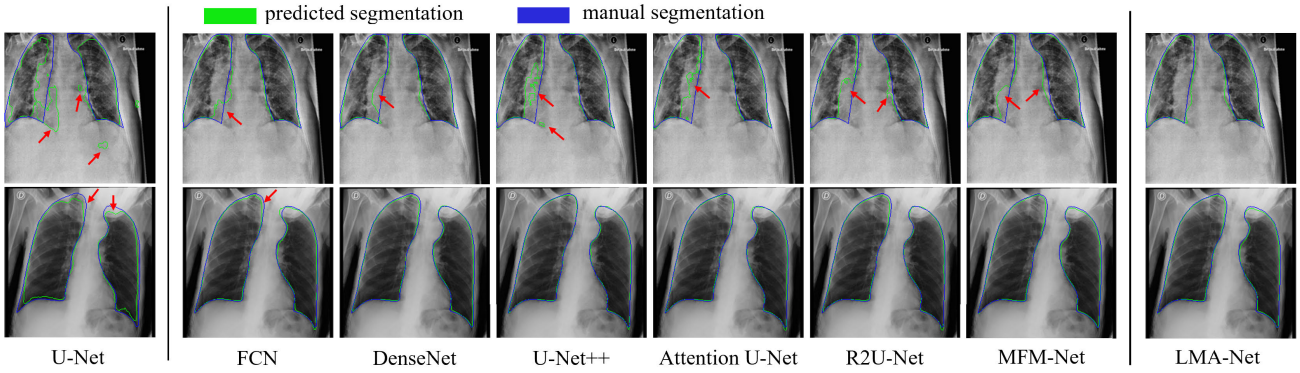


FIGURE 9. Visual comparison between LMA-Net and state-of-the-art networks for lung segmentation. The red arrows highlight some mis-segmentations. The blue border is the result of manual segmentation. The green border is the predicted segmentation result.

TABLE 5. Comparison with other methods for computing cost.

Network	Param(M)	FLOPs(G)	Time(ms)
U-Net [10]	3.91	7.72	724.38
FCN [15]	30.02	801.97	1119.81
DenseNet [13]	7.97	3.64	1372.41
U-Net++ [18]	36.63	138.52	726.72
Attention U-Net [17]	9.83	19.91	980.13
R2U-Net [16]	39.09	152.81	879.82
MFM-Net [32]	**0.31**	8.77	817.66
LMA-Net	2.89	**2.64**	**609.31**

TABLE 6. Generalization comparison with other methods for segmentation results.

Network	*IoU*(%)	*Dice*(%)	*ASSD*(mm)	*HD*95(mm)
U-Net	95.96	96.36	32.62	158.62
FCN	93.35	93.52	37.33	143.89
DenseNet	93.59	94.03	29.58	**125.22**
U-Net++	95.10	95.40	36.74	163.37
Attention U-Net	95.44	95.83	34.10	180.24
R2U-Net	94.29	94.90	34.06	170.64
MFM-Net	94.85	95.69	**27.73**	155.47
LMA-Net	**97.10**	**97.16**	34.70	138.22

experiment. It shows that the LMA-Net has good segmentation performance. The LMA-Net obtains a *IoU* of 96.28%, which is a great improvement compared with 93.85% of U-Net. Although the LMA-Net is similar to the Attention U-Net and R2U-Net in segmentation performance, our model is far less than other models for comparative experiments in terms of Param and FLOPs, as shown in TABLE 5. FIGURE 9 shows the visual comparison of different CNNs dealing with CXR segmentation task.

4) COMPARISON WITH OTHER METHODS FOR COMPUTING COST

We randomly generate a tensor $f \in \mathbb{R}^{1*256*256}$ as the input of the segmentation models to test computing cost. Then, the parameters(Param), floating point operations(FLOPs) and inference time of networks can be tested. As show in TABLE 5, the LMA-Net outperforms other state-of-the-art networks in Param, FLOPs and inference time, and the corresponding values are 2.89M, 2.64G and 609.31ms, respectively. Although the LMA-Net has more parameters than MFM-Net, its segmentation performance is much higher than MFM-Net. It follows that the LMA-Net achieves lightweight while ensuring superior segmentation performance.

5) COMPARISON FOR GENERALIZATION

LMA-Net is compared with seven state-of-the-art methods which are U-Net, FCN, DenseNet, U-Net++, Attention

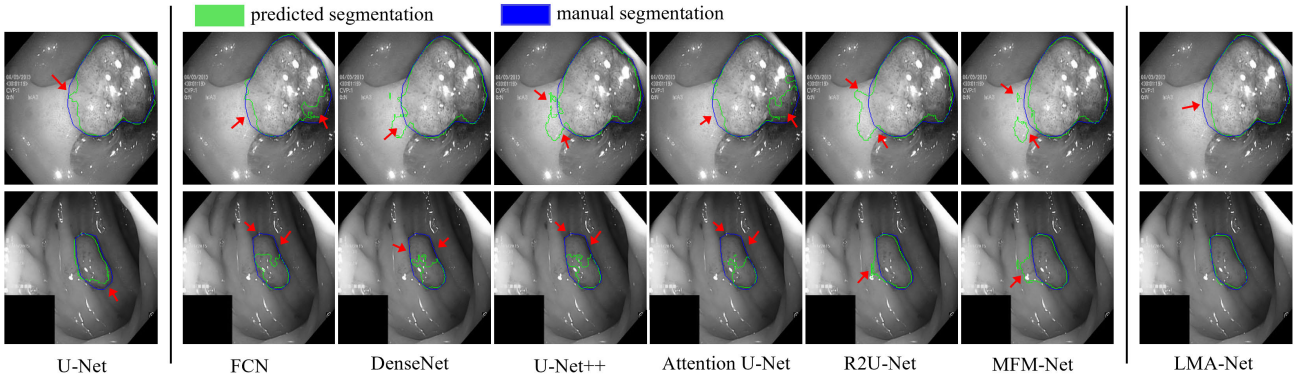


FIGURE 10. Visual comparison between LMA-Net and state-of-the-art networks for colonic polyps segmentation. The red arrows highlight some mis-segmentations. The blue border is the result of manual segmentation. The green border is the predicted segmentation result.

U-Net, R2U-Net and MFM-Net. These models are all retrained on the colon polyp dataset Kvasir-SEG.

As shown in TABLE 6, segmentation results of the segmentation models are listed in the contrastive experiment. It shows that the LMA-Net has good segmentation performance. The LMA-Net obtains a *IoU* of 97.10%, which is a great improvement compared with 95.96% of U-Net. Although the LMA-Net is similar to the U-Net++ and Attention U-Net in segmentation performance, our model is far less than other models for comparative experiments in terms of Param and FLOPs, as shown in TABLE 6. FIGURE 10 shows the visual comparison of different CNNs dealing with Kvasir-SEG segmentation task.

V. DISCUSSION AND CONCLUSION

In this paper, we explore the possibility of deep learning to assist in medical diagnosis. To address the two key issues of limited computing resources in clinical medical workstations and low accuracy of deep learning networks for medical image segmentation, a lightweight multiple attention network(LMA-Net) is proposed to achieve pulmonary segmentation of CXR images, which is used to assist diagnosis. This method improves the classical convolutional module to achieve lightweight, and innovatively proposes reinforced channel attention and multiscale attention to heighten segmentation accuracy. In addition, the COVID-QU-Ex dataset is improved and multiple lung lesion images are fused to improve the generalization of the segmentation network. Compared with other approaches, it verifies that the proposed method is superior through comprehensive experiments. To conclude, we will explore imagedriven methods for lesion recognition in future work.

VI. DECLARATIONS

The authors declare that there are no conflict of interests, we do not have any possible conflicts of interest.

REFERENCES

[1] B. Gecer, S. Aksoy, E. Mercan, L. G. Shapiro, D. L. Weaver, and J. G. Elmore, "Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks," *Pattern Recognit.*, vol. 84, pp. 345–356, Dec. 2018.

[2] W. Liu, J. Luo, Y. Yang, W. Wang, J. Deng, and L. Yu, "Automatic lung segmentation in chest X-ray images using improved U-Net," *Sci. Rep.*, vol. 12, no. 1, p. 8649, May 2022.

[3] E. Skoura, A. Zumla, and J. Bomanji, "Imaging in tuberculosis," *Int. J. Infectious Diseases*, vol. 32, pp. 87–93, Mar. 2015.

[4] S. Candemir and S. Antani, "A review on lung boundary detection in chest X-rays," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 4, pp. 563–576, Apr. 2019.

[5] S. Secinaro, D. Calandra, A. Secinaro, V. Muthurangu, and P. Biancone, "The role of artificial intelligence in healthcare: A structured literature review," *BMC Med. Informat. Decis. Making*, vol. 21, no. 1, pp. 1–23, Dec. 2021.

[6] N. Delfan, H. A. Moghaddam, M. Modaresi, K. Afshari, K. Nezamabadi, N. Pak, O. Ghaemi, and M. Forouzanfar, "CT-LungNet: A deep learning framework for precise lung tissue segmentation in 3D thoracic CT scans," 2022, *arXiv:2212.13971*.

[7] J. Paul Cohen, P. Morrison, L. Dao, K. Roth, T. Q Duong, and M. Ghassemi, "COVID-19 image data collection: Prospective predictions are the future," 2020, *arXiv:2006.11988*.

[8] S. G. Armato III and W. F. Sensakovic, "Automated lung segmentation for thoracic CT: Impact on computer-aided diagnosis," *Academic Radiol.*, vol. 11, no. 9, pp. 1011–1021, 2004.

[9] G. Litjens, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.

[11] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.

[12] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[14] Y. LeCun, "Generalization and network design strategies," *Connectionism Perspective*, vol. 19, nos. 143–155, p. 18, 1989.

[15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[16] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*.

[17] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[18] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and 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*. Cham, Switzerland: Springer, 2018, pp. 3–11.

[19] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "TransUNet: Transformers make strong encoders for medical image segmentation," 2021, *arXiv:2102.04306*.
[20] M. Guo, "Attention mechanisms in computer vision: A survey," *Comput. Vis. Media*, vol. 8, no. 3, pp. 331–368, 2022.
[21] G. Cheng, P. Lai, D. Gao, and J. Han, "Class attention network for image recognition," *Sci. China Inf. Sci.*, vol. 66, no. 3, Mar. 2023, Art. no. 132105.
[22] L. Giusti, C. Battiloro, L. Testa, P. Di Lorenzo, S. Sardellitti, and S. Barbarossa, "Cell attention networks," 2022, *arXiv:2209.08179*.
[23] C. H. Song, H. J. Han, and Y. Avrithis, "All the attention you need: Global-local, spatial-channel attention for image retrieval," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 439–448.
[24] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[25] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 3–19.
[26] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11531–11539.
[27] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13708–13717.
[28] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11057–11066.
[29] L. Yang, R. Zhang, L. Li, and X. Xie, "SimAM: A simple, parameter-free attention module for convolutional neural networks," in *Proc. 38th Int. Conf. Mach. Learn.*, vol. 139, 2021, pp. 11863–11874.
[30] M. Dusmanu, I. Rocco, T. Pajdla, M. Pollefeys, J. Sivic, A. Torii, and T. Sattler, "D2-Net: A trainable CNN for joint description and detection of local features," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8084–8093.
[31] M. Shi, S. Lin, Q. Yi, J. Weng, A. Luo, and Y. Zhou, "Lightweight context-aware network using partial-channel transformation for real-time semantic segmentation," *IEEE Trans. Intell. Transp. Syst.*, early access, Jan. 22, 2024, doi: 10.1109/TITS.2023.3348631.
[32] Z. Cao, W. Zhang, X. Wen, Z. Dong, Y.-s. Liu, X. Xiao, and B. Yang, "KTNet: Knowledge transfer for unpaired 3D shape completion," 2021, *arXiv:2111.11976*.

TURGHUNJAN MAMUT was born in Awat, Aksu, Xinjiang, China, in 1970. He received the B.S. degree in applied mathematics from Xinjiang Normal University, in 1993. From 1993 to 1997, he was a Teaching Assistant with the Basic Teaching Department, Tarim University, Alar, Xinjiang, China. In 1997, he studied with the Department of Statistics, East China Normal University, Shanghai, for one year. From 1998 to 2006, he was a Lecturer with the College of Arts and Sciences, Tarim University, where he has been an Associate Professor with the College of Information Engineering, since 2006. His research interests include AI, computer vision, Uyghur speech recognition, and natural language processing of Uyghur.

LUN MENG is currently pursuing the master's degree with Chongqing University of Science and Technology. His current research interests include computer vision and semantic segmentation.

ZIYI PEI is currently pursuing the degree in material forming and control engineering with Chongqing University of Arts and Sciences. His current research interests include artificial intelligence and deep learning.

TENGFEI WENG received the B.S. and M.S. degrees in chemistry and chemical engineering from Chongqing University, China, in 2007 and 2010, respectively. She is currently with Chongqing University of Science and Technology. Her current research interests include artificial intelligence and neural networks.

QI HAN (Member, IEEE) received the B.S. degree in computer science and technology from Shandong University, China, in 2005, and the M.S. and Ph.D. degrees from Chongqing University, China, in 2009 and 2012, respectively. He is currently an Associate Professor with Chongqing University of Science and Technology. His current research interests include artificial intelligence, system optimization, neural networks, and chaos control.

KEPENG WU is currently pursuing the master's degree with Chongqing University of Science and Technology, Chongqing, China. His research interests include image segmentation, image super-resolution, image refinement reconstruction, medical image noise reduction, and neural network parameter optimization. At present, the main research directions are MRI medical image segmentation and medical auxiliary diagnosis.

XIN QIAN received the bachelor's degree in computer science from Jinggangshan University, China, in 2021. He is currently pursuing the master's degree with Chongqing University of Science and Technology. His research interests include machine learning and medical image analysis.

HONGXIANG XU received the B.S. degree from Sanjiang University, Jiangsu, China, in 2020. He is currently pursuing the master's degree with Chongqing University of Science and Technology. His research interests include deep learning, brain–computer interface, and medical image analysis.

ZICHENG QIU was born in Wuhan, Hubei, China, in 1983. He received the B.S. degree in optoelectronic engineering from Huazhong University of Science and Technology, Wuhan, in 2005, and the Ph.D. degree in optical engineering (advanced lithography technologies) from Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai, China, in 2010. From 2010 to 2011, he was a Software Engineer with Synopsys. From 2011 to 2015, he was an Assistant Professor with Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Science, Chongqing, China. From 2015 to 2020, he was an Associate Professor with the College of Information Engineering, Tarim University, Alar, Xinjiang, China. He is currently an Associate Professor with the College of Intelligent Technology of Engineering. He is the author of more than ten articles and three inventions. His research interests include AI, intelligent speech technologies, and security of AI.

YUAN TIAN received the B.S. degree in computer science and technology from Chongqing Normal University, Chongqing, China, in 2009, the M.S. degree in computer application technology from Chongqing University, Chongqing, in 2012, and the Ph.D. degree in computational intelligence and information processing from Southwest University, Chongqing, in 2020. She was with Chongqing University of Science and Technology, Chongqing. Her current research interests include impulsive systems, discontinuous dynamical systems, and multi-agent systems.

YANGJUN PEI received the bachelor's degree in technical economics and the master's degree in computer software and theory from Chongqing University, in 2001 and 2005, respectively. He is currently a Lecturer with the School of Intelligent Technology and Engineering, Chongqing University of Science and Technology. His career has been dedicated to teaching and research with Chongqing University, since June 2007, focusing on intelligent technology and engineering. He has actively contributed to research projects, particularly in steel surface defect classification and power systems. He has also authored articles in computational intelligence and neural networks and holds patents for inventions, including intelligent drinking water machines and parallel watering devices.