

# Multi-path Aggregation U-Net for Lung Segmentation in Chest Radiographs

Ling Ma

Hunan University of Information Technology

Xiaomao Hou

Hunan University of Information Technology

Zhi Gong (✉ [gongzhi19822@163.com](mailto:gongzhi19822@163.com))

Hunan University of Information Technology

---

## Research

**Keywords:** Lung segmentation, Chest X-ray, Atrous convolution, Deep layer aggregation, Fully convolutional networks

**Posted Date:** April 6th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-365278/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Multi-path Aggregation U-Net for Lung Segmentation in Chest Radiographs

Ling Ma, Xiaomao Hou, Zhi Gong\*

School of Electronic Institute of Information, Hunan University of Information Technology, Changsha, 410451, Hunan, China

Corresponding author: Zhi Gong; gongzhi19822@163.com

**Abstract** Lung segmentation from chest X-ray images is a fundamental and crucial step for computer-aid diagnosis (CAD) system. Although many techniques for this problem have been proposed, it still remains as a challenge. Recently, Fully Convolutional Networks (FCNs) especially U-Net has been hugely successful for many image segmentation tasks. In this paper, we propose a revised variant of U-Net, specifically, we design two main components. The first component is multi-path dilated convolutions with different dilation rate to extract multi-scale features. It was used to replace the basic convolutions used in the original U-Net. The second component is skip connections with dense deep layer aggregation to further aggregate features across different scales. We perform extensive experiments on three publicly available datasets (in total 951 images). Our proposed method outperforms many other segmentation methods and achieves state-of-the-art segmentation performance (Dice's coefficient of 96.5%, 97.9% and 96.7% on the three datasets, respectively).

**Keywords:** Lung segmentation, Chest X-ray, Atrous convolution, Deep layer aggregation, Fully convolutional networks

## 1 Introduction

As a cost-effective and most common used medical imaging method, chest radiographs provide extensive and valuable information for diagnosis of many lung diseases, for example, lung cancer, pneumonia and pneumothorax. Accurate lung segmentation is a critical and fundamental step for Computer-Aided Diagnosis (CAD) of most lung diseases in automatic analysis of chest radiographs. The accuracy of lung segmentation methods greatly affects the accuracy of downstream tasks. Although many lung segmentation approaches have been presented, this task still has some unsolved problems and challenging scenarios. For example, the shape of lung area is highly related to gender, age, and health condition of patients. In addition, segmentation of lung in chest radiographs is much more difficult when there exist some external objects such as sternum wire, heart pacemaker and surgical clips [1].

Over the past decades, a large variety of traditional lung segmentation methods have been proposed. These lung segmentation approaches can be roughly grouped into four types: rule based, deformable model based, pixel classification based and hybrid methods. These methods have achieved certain results, however, they have some disadvantages and limitations such as heavily rely on hand-craft features and often lead to over-segmentation and under-segmentation [19].

Recently, deep learning techniques are undergoing a renaissance, due to the large amount of data for training and high performance computation power [16]. Convolutional Neural Networks (CNNs), as one important component of deep learning, have dominated the computer vision field. For image segmentation task, Fully Convolutional Network (FCN) [18] is one of the first neural networks which adopts fully convolutional network to tackle image segmentation problem. Based on FCN, U-Net [22] is a segmentation network which was specifically designed for medical images. Since U-Net was proposed, it

has been extensively adopted for segmenting medical image and achieved superior results [9]. There have been numerous segmentation works which are based on U-Net and some innovations or improvements are made upon. Li et al. [17] proposed a hybrid U-Net based on the dense connection which was introduced in DenseNet [11] for tumor segmentation from CT images. Zhou et al. [29] proposed a nested U-Net architecture to better extract semantic features. Ibtehaz et al. [12] designed MultiResUNet which is an enhanced version of U-Net and showed excellent performance. On the other side, many approaches for lung segmentation have been proposed, for example, Inf-Net [5] adopts a segmentation network based on a parallel partial decoder and dual attention mechanism for CT images.

Inspire by the above-mentioned research works, we propose a novel lung segmentation approach in chest X-ray images based on CNNs. Specifically, we design an encoder-decoder style segmentation network on the basis of U-Net. In order to extract multi-scale image features, we design a multi-path convolution block named Multi-Path Dilated Residual Block (MPDRB) to replace the original convolution block in U-Net. In addition, with the purpose of enlarge the receptive field of convolution operations in bottleneck layers, we introduce Dilated Residual Block (DRB) which utilizes dilate/atrous convolution [27] to accomplish that. Although some previous works have introduced atrous convolution for medical image segmentation, for instance, Hesamian et al. [10] propose a deep residual CNN with atrous convolution for lung nodule segmentation from CT images, the effect of atrous convolution for lung segmentation from X-ray images are still not well exploited and studied. Moreover, motivated by Yu et al. [28], we extend the original skip connections in U-Net with Dense Deep Layer Aggregation (DDLDA). We evaluate our proposed models on three publicly available chest X-ray image datasets and show that our proposed approach achieves state-of-the-art performance compared

with other segmentation methods.

The remaining of the paper is organized as following: Section 2 describes the details of our proposed method. Section 3 shows the results of the experiments and the conclusion is given in Section 4.

## 2 Methodology

### 2.1 Overview

Our proposed lung segmentation method is an end-to-end deep learning framework. The only pre-processing step is just histogram equalization and no post-processing is needed. During training phase, the chest X-ray images after histogram equalization and the associated lung segmentation mask are resized and feed into our deep CNN model. In the testing phase, once the model was trained, the lung segmentation map can be directly obtained through prediction of trained model and no other processing steps are needed.

### 2.2 Pre-processing

The pre-processing step in our approach including histogram equalization and image resizing. Since the pixel intensity is varies greatly for some chest X-ray images, histogram equalization operation is applied for every input chest radiographs to enhance the image contrast.

Histogram equalization operation is one of the most widely used and classical method for image enhancement. The main concept of histogram equalization is to convert the histogram distribution of an image into an approximately uniform distribution, thereby enhancing the contrast of the image. Specifically, the cumulative distribution function of the input image is firstly computed based on the probability distribution of pixels. Afterwards, the equalized image can be produced by a transformation:

$$S_k = \frac{L-1}{N} \sum_{j=1}^k n_j \quad k = 0, 1, 2, \dots, L-1 \quad (1)$$

where  $k$  is the input gray level and  $S_k$  is the output gray level. The total number of gray levels is represented by  $L$  and the total amount of pixels is indicated by  $N$ .

### 2.3 Network Architecture

The overall architecture of our network is illustrated in Figure 1, which is mainly based on encoder-decoder fully convolutional neural networks (e.g. U-Net), with two

principle significant improvements: multi-path dilated convolution and dense deep layer aggregation.

In order to illustrate the improvements of our proposed network more clearly and efficiently, Figure 2 shows the architecture of the conventional U-Net. Compared with the conventional U-Net which uses two

consecutive  $3 \times 3$  convolutions to extract semantic features, we propose multi-path dilated convolutions with

diverse dilation rate to enlarge the receptive fields of convolution operations thus can better learn features at multiple scales. As shown in Figure 1, three different building blocks of multi-path dilated convolutions were presented: Multi-Path Dilated Residual Block (MPDRB), Multi-Path Block (MPB) and Dilated Residual Block (DRB).

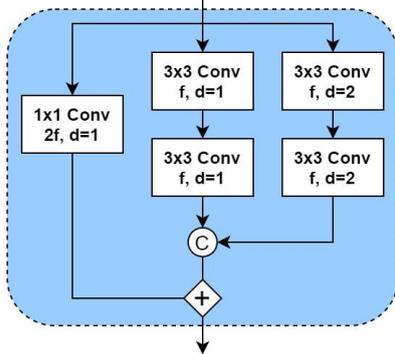
In the original U-Net, the skip connection was utilized to fuse image features between encoder and decoder at corresponding scales. However, these connections are shallow and linear [28]. There have many previous works intend to tackle this issue [14]. Based on these works, in our approach we present a novel feature aggregation paradigm: dense deep layer aggregation to further strengthen the skip connection.

### 2.4 Multi-path Dilated Convolution

In the original U-Net, for every convolution block, two successive  $3 \times 3$  convolutions are used both for the encoder and decoder. However, it has some disadvantages such as it may not capable of extract spatial multi-scale features. In addition, in the case of the resolution of input image is large (such as 256 or 512), the receptive field of the convolution operation in bottleneck layers may not cover the entire input image, thus may not very efficient to segment objects which is pretty large.

Considering these above reasons, we specifically design three components based on dilated convolutions. In particular, the first one is MPDRB in which we design two branches with different dilation rates and one residual connection. The first branch employs two ordinary convolutions to small scale semantic features. The success of design philosophy of VGG network [25] indicate that the effect of successive convolution operations with small kernel (e.g.  $3 \times 3$ ) is equivalent to the effect of convolution operations with large kernel (e.g.  $5 \times 5$ ), but with less computation and parameters. For example, the receptive field of two successive  $3 \times 3$  convolutions is the same with the receptive field of one  $5 \times 5$  convolution, but with 28% less computation.

## Multi-Path Dilated Residual Block (MPDRB)



## Dilated Residual Block (DRB)

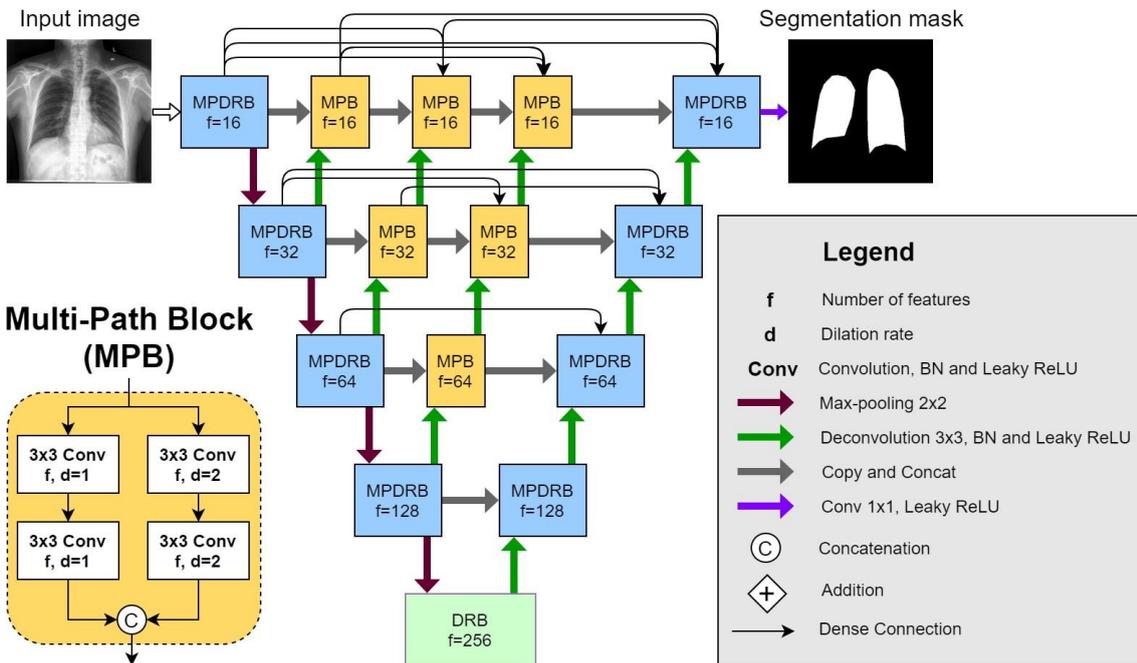
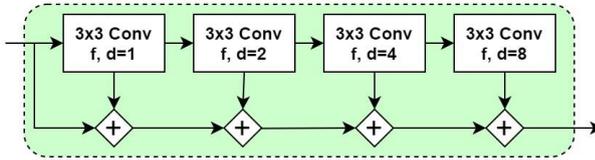


Figure 1: The proposed segmentation network architecture.

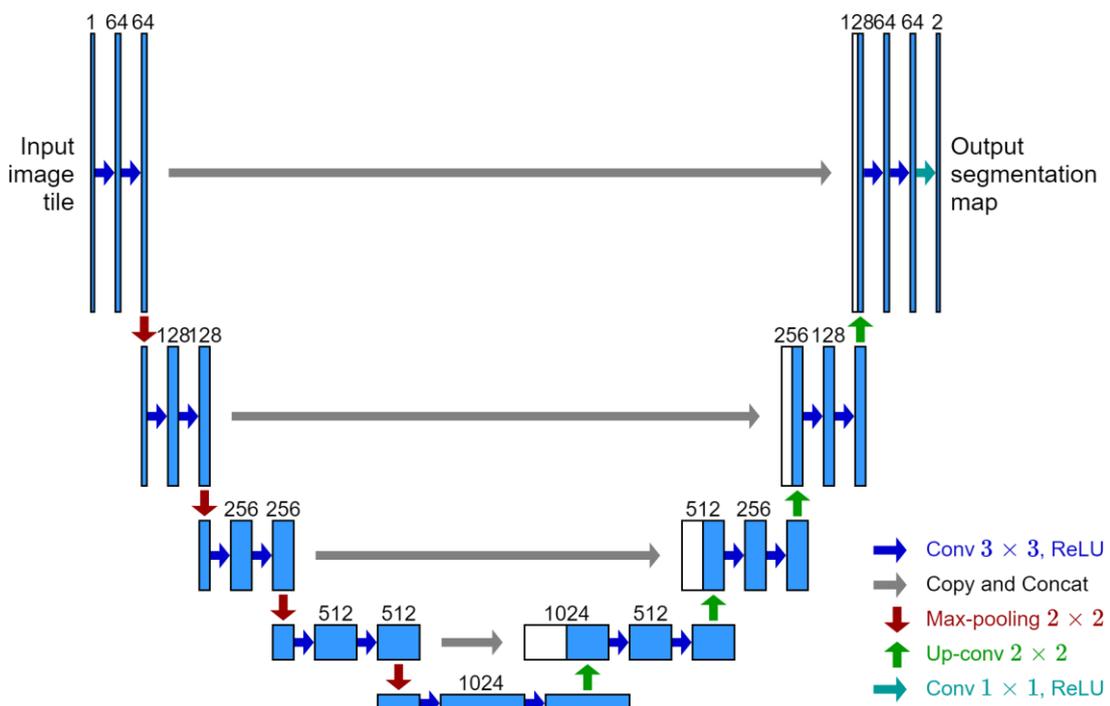


Figure 2: The conventional U-Net architecture. Blue boxes represent feature maps. Different operations are denoted by arrows with different color.

Meanwhile, in order to extract semantic features of large objects, we also design dilation branch, which is convolutions with dilation rate 2. Figure 3 illustrates the receptive fields of these two branches used in MPDRB, where Figure 3 (a) shows that two successive ordinary  $3 \times 3$  convolutions which has a receptive field of 5, and Figure 3 (b) demonstrates that two successive dilated convolutions with dilation rate 2 has a receptive field of 9. For the residual connection in MPDRB, we adopt  $1 \times 1$  convolution motivated by the work of He et al.<sup>[8]</sup>.

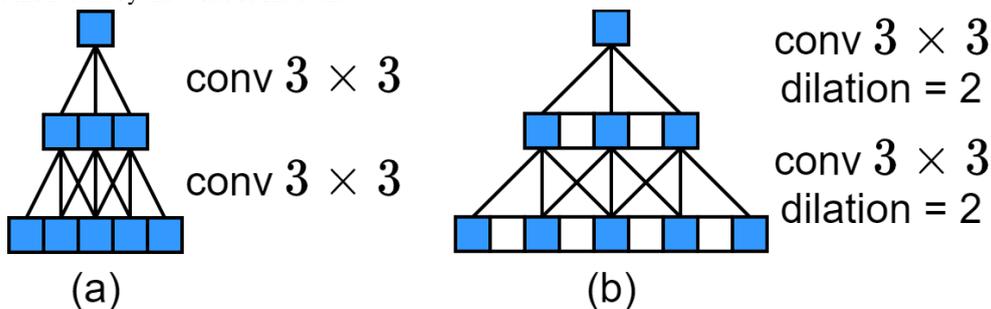


Figure 3: The receptive fields of two branches used in MPDRB block.

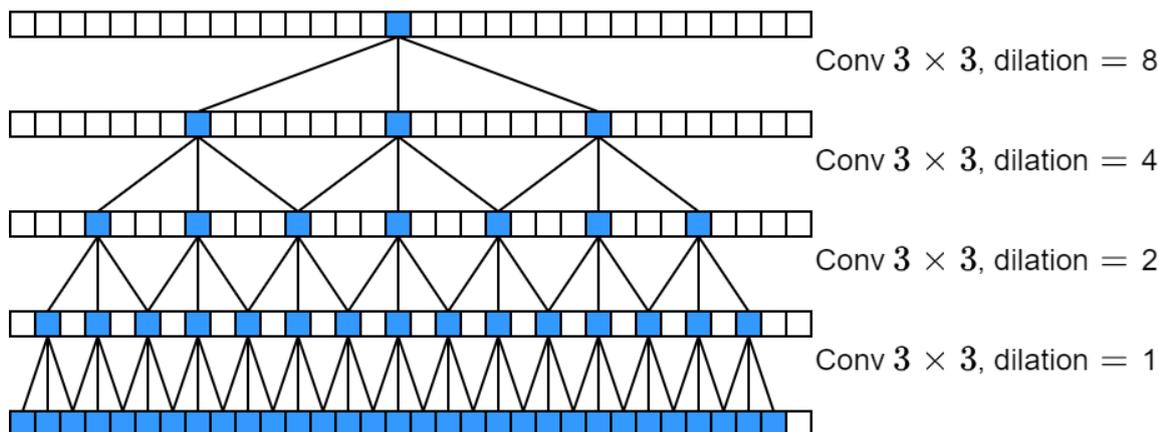


Figure 4: The receptive field of one convolution operation in the DRB.

## 2.5 Dense Deep Layer Aggregation

The main purpose of skip connection in the conventional U-Net is merging features at corresponding scales between encoder and decoder. Additionally, it also provides shortcut path for back propagates gradients. Nevertheless, the skip connection is just pure copy and paste, therefore it is still linear and shallow<sup>[28]</sup>. In [14], the skip connection is extended by DLA. However, it eliminates the shortcut path for gradients propagation. In conclusion, we design DDLA which combines DLA and dense connection. Specifically, direct connections are introduced between any layer to all following layers in DLA.

## 3 Experimental Results

### 3.1 Datasets

We employ three publicly available lung segmentation datasets for this study: Montgomery County (MC)<sup>[13]</sup>, Japanese Society of Radiological Technology (JSRT)<sup>[2, 24]</sup> and Shenzhen Dataset<sup>[26]</sup>.

### 3.2 MC Dataset

The MC dataset consists of 138 frontal chest radiographs and each has either  $4892 \times 4020$  or  $4020 \times 4892$  resolution. This set collected over many years from

The second one is Multi-Path Block (MPB) which is MPDRB without the residual connection. The last one is DRB which we stack four dilated convolutions with dilation rate 1,2,4,8 respectively. The receptive field of one convolution operation in the DRB is 31, which is illustrated in Figure 4. Even in the case of the resolution of input image is 512, this receptive field can approximately cover the entire input image (the resolution of the features in bottleneck layer is 32).

the department of Health and Human Services, Montgomery County, Maryland, USA. The corresponding lung segmentation golden standard are annotated under the supervision of experienced radiologists.

### 3.3 JSRT Dataset

The JSRT dataset contains 247 chest X-rays and was gathered from 14 medical centers and assembled by the Japanese Society of Radiological Technology (JSRT). All images are with  $2048 \times 2048$  resolution. The annotated lung segmentation mask was performed and introduced by Ginneken et al.<sup>[6]</sup>.

### 3.4 Shenzhen Dataset

The Shenzhen dataset is collected from Shenzhen No.3 People's Hospital, Shenzhen, China. In total 662 images are in this dataset. The dimensions of these images are varied but around  $3000 \times 3000$ . The corresponding lung segmentation mask are provided by Stirenko et al.<sup>[26]</sup>. However, only a portion of images (566) have their associated lung segmentation mask, therefore only 566 images from this dataset are used in our experiments.

Table 1 summarizes the details of each dataset used in this study and some visual examples of these three datasets are illustrated in Figure 5. The example images in Figure 5 were obtained through the weighted summation of the green channel of the chest image and the associated lung

segmentation mask image.

## 4. Evaluation Metrics

For the sake of performance evaluation of our proposed approach and make a comparison with other lung

segmentation methods, we used three evaluation metrics in this study: the Jaccard similarity coefficient (JSC, also referred as overlap measure), and Dice’s score (DCS).

**Tab.1 Details of each dataset used in this study.**

Dataset	Number of Images	Resolution
MC	138	4892 × 4020 or 4020 × 4892
JSRT	247	2048 × 2048
Shenzhen	566	around 3000 × 3000



**Figure 5: Chest radiographs and corresponding manual lung segmentation (shown in green) from 3 datasets used in our study (From left to right: MC, JSRT and Shenzhen dataset).**

Given the ground truth mask  $GT$  and the predicted segmentation mask  $S$ , the  $JSC$  and  $DCS$  are formally defined as:

$$JSC = \frac{|GT \cap S|}{|GT \cup S|} = \frac{|TP|}{|TP| + |FP| + |FN|} \quad (2)$$

$$DCS = \frac{2|GT \cap S|}{|GT| + |S|} = \frac{2|TP|}{2|TP| + |FP| + |FN|} \quad (3)$$

where  $TP$ ,  $FP$ ,  $TN$  and  $FN$  stands for true positives, false positives, true negatives and false negatives, respectively.

### 4.1 Implementation Details

The data splitting protocol for all the three datasets is the same, specifically, the portion of data used for training, validation and testing is 50%, 10% and 40%, respectively. In order to achieve reliable and steady performance results and eliminate the dataset random splitting bias, we randomly generate 10 different splitting training/validation/testing sets. Following previous studies, the resolution of all input images is resized to  $256 \times 256$ . Since the pixel intensity is varies greatly for some images, the histogram equalization operation is applied for every input chest radiographs as the pre-processing step.

Since we made two main improvements compared with previous works, we implement three segmentation models, namely, MPDC U-Net which represents the U-Net extended with multi-path dilated convolution, DDLA U-Net which stands for the U-Net extended with dense deep layer aggregation, and MPDC DDLA U-Net which combines both. Moreover, we also implement the DLA U-Net model which is the U-Net extends DLA as described in [14].

All of the models were implemented by PyTorch framework [20] and optimized by Adam optimizer with learning rate  $1e-3$  which is suggested by the original literature [15]. The weights of all models were initialized by MSRA initialization strategy [7] and trained from scratch. The batch size for both training and validation is 4. In order to prevent overfitting and accelerate convergence during model training, we use numerous strategies including Dropout with drop rate 0.2 and early stopping with patience 15 epochs. We use the model which achieves the lowest validation loss to evaluate the performance on test set.

### 4.2 Study on Loss Functions

The design and selection of different loss functions is critical for the performance of deep learning segmentation models. The loss function is used to measure the error between prediction and ground truth which can be back propagated to network layers and update the weights. For segmentation problems, Cross Entropy (CE) loss and Dice loss are the most two widely used options for loss function. In this study, we investigate these two above-mentioned loss functions in our model.

CE loss perhaps is the most widely used loss function for image segmentation, it is a pixel-wise based loss function. This loss examines each individual pixel separately and compares the class prediction vector with the ground truth class vector. In binary segmentation situation, the formal definition of CE is:

$$CE = -\frac{1}{m} \sum_{i=1}^m (y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)) \quad (4)$$

where  $y_i$  and  $\hat{y}_i$  are the ground truth segmentation and predicted segmentation, respectively. The pixel of image is

indexed by  $i$ .

Dice loss is another popular loss function which is based on overlap measure between the ground truth and prediction for image segmentation tasks. The Dice coefficient is often used to evaluate the segmentation performance and ranges from 0 to 1. The Dice loss is defined as:

$$DICE = 1 - \frac{2|y \cap \hat{y}|}{|y| + |\hat{y}|} \quad (5)$$

In order to find the most suitable loss function for the lung segmentation task, we firstly design experiments on Shenzhen dataset using our proposed model. Specifically,

we use three loss functions: CE loss, Dice loss and their addition, then we trained three models with the corresponding loss. The accuracy curves on validation set are shown in Figure 6, in which we can clear observe that the combination of CE and Dice loss achieved the highest accuracy. Therefore, for the rest of experiments, we use the addition of these two losses for all our proposed models:

$$Loss = CE + DICE \quad (6)$$

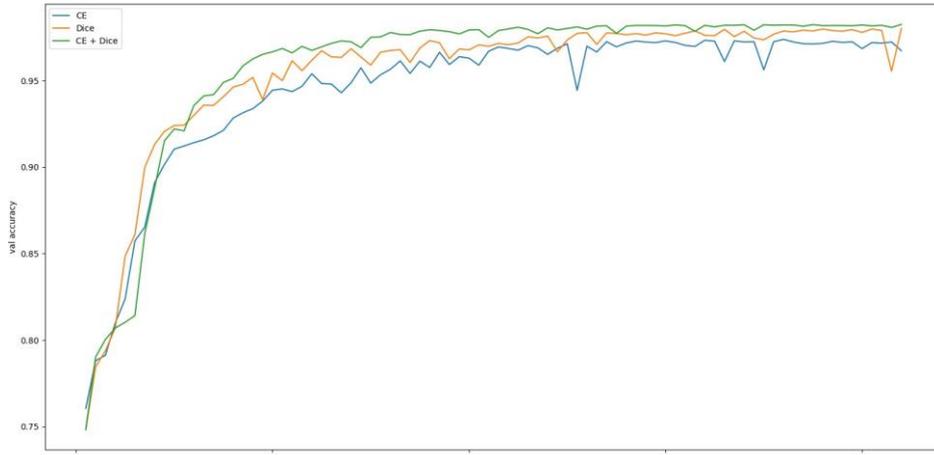


Figure 6: Accuracy curves on validation set for different loss functions.

## 5. Results and Discussion

The performance comparison of our proposed approach with other methods on the MC dataset, JSRT dataset and Shenzhen dataset is illustrated in Table 2. We make extensive comparison with many renowned segmentation networks such as FCN-8<sup>[18]</sup>, U-Net<sup>[22]</sup> and Deeplabv3<sup>[4]</sup>. The results of these three models are from

Tab.2 Performance comparison of our proposed method with other methods on the MC, JSRT and Shenzhen dataset.

Results showing mean with standard deviation.

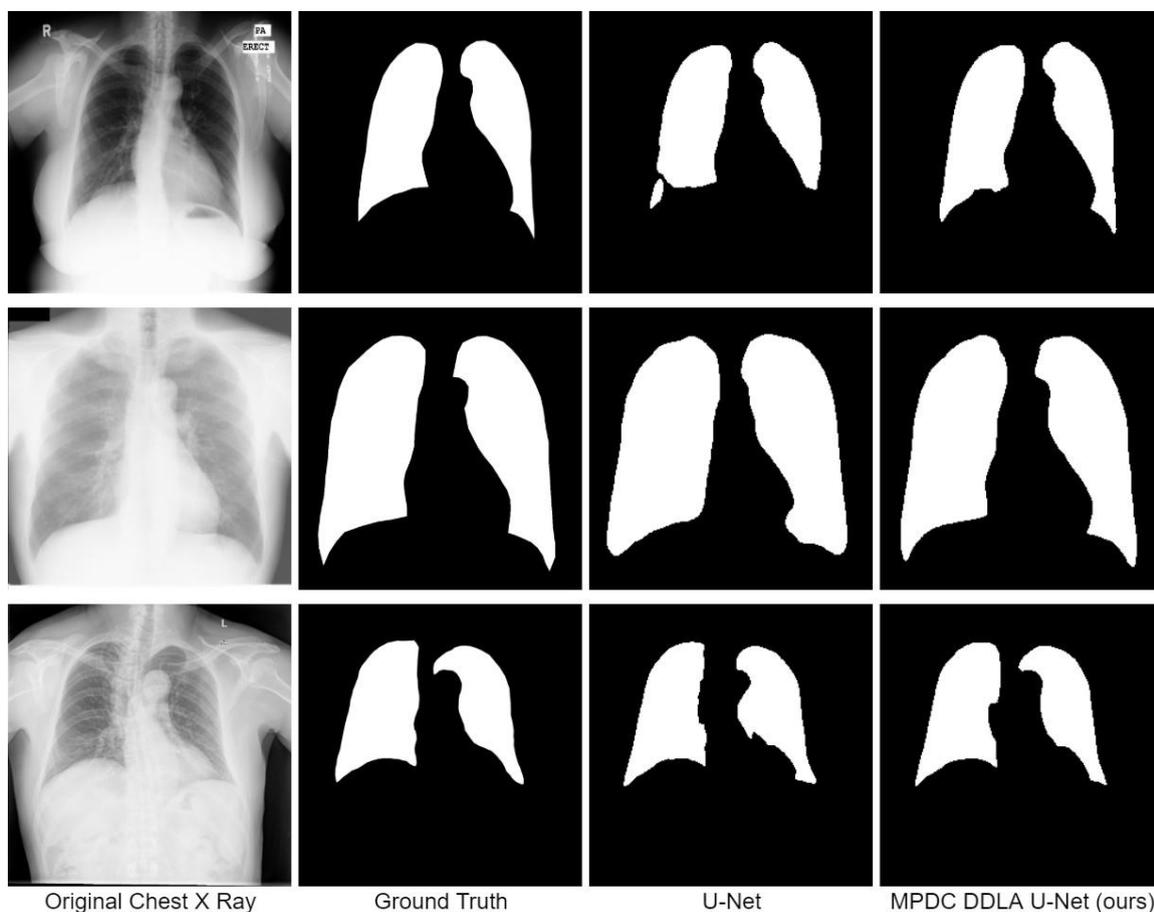
Method	MC		JSRT		Shenzhen	
	JSC(%)↑	DCS(%)↑	JSC(%)↑	DCS(%)↑	JSC(%)↑	DCS(%)↑
FCN-8 <sup>18</sup>	91.1±4.1	95.1±2.4	94.2±2.3	97.0±1.2	91.0±4.4	95.2±2.3
Candemir <i>et al.</i> <sup>3</sup>	94.1±3.4	96.0±1.8	95.4±1.5	96.7±0.8	-	-
Rashid <i>et al.</i> <sup>21</sup>	-	95.4±2.1	-	95.1±1.7	-	-
Novikov <i>et al.</i> <sup>19</sup>	94.1±3.3	96.1±1.8	94.8±1.8	97.3±1.0	92.2±2.9	96.0±1.3
MultiResUNet <sup>12</sup>	94.2±3.1	96.3±1.7	94.7±1.8	97.4±1.0	92.4±2.8	96.0±1.4
AG-net <sup>23</sup>	94.3±3.2	96.3±1.8	95.0±1.6	97.5±0.9	92.5±2.8	96.1±1.4
Deeplabv3 <sup>4</sup>	93.5±3.5	95.4±1.9	94.7±1.8	97.4±0.9	92.2±3.1	95.8±1.7
U-Net <sup>22</sup>	93.1±3.6	95.2±2.0	94.5±1.9	97.4±1.0	92.2±3.1	95.8±1.6
U-Net (DLA)	93.8±3.4	95.6±1.9	94.6±1.8	97.5±1.0	92.4±2.8	96.1±1.3

our implementation. We also make a comparison with some recently proposed approach for lung segmentation like Candemir *et al.*<sup>[3]</sup>, Rashid *et al.*<sup>[21]</sup> and Novikov *et al.*<sup>[19]</sup>. The performance results are directly come from the original literature. We also compare with some medical image segmentation such as MultiResUNet<sup>[12]</sup> and AG-net<sup>[23]</sup>. The performance results of these two methods are also from our implementation. A visual comparison of lung segmentation results is demonstrated in Figure 5.

MPDC U-Net (ours)	94.3±3.1	96.2±1.7	95.2±1.5	97.7±0.8	92.5±2.6	96.4±1.2
DDLA U-Net (ours)	93.9±3.3	95.8±1.9	94.8±1.7	97.5±1.0	92.4±2.7	96.2±1.3
MPDC DDLA U-Net (ours)	<b>94.8±2.9</b>	<b>96.5±1.5</b>	<b>95.6±1.3</b>	<b>97.9±0.7</b>	<b>92.9±2.4</b>	<b>96.7±1.1</b>

From these three tables, we can observe that our best performing model (MPDC DDLA U-Net) reached 94.8%, 95.6%, 92.9% in terms of JSC and 96.5%, 97.9%, 96.7% in terms of DCS on the three datasets, respectively. Compared with the performance of our baseline model (U-Net), our proposed best model increases the JSC with 1.7%, 1.1%,

0.5%, and DCS with 1.3%, 0.5%, 0.9% on the MC, JSRT and Shenzhen dataset. Notably, our best model achieved the highest score in terms of all two metrics on all the three datasets, surpassing numerous other methods including renowned segmentation models and recently proposed lung segmentation approaches.



**Figure 7: A visual comparison of lung segmentation results between U-Net and our proposed MPDC DDLA U-Net model.**

Further, even the performances of our models with just one improved component (MPDC U-Net (ours) or DDLA U-Net (ours)) are higher than the majority of other methods. Specifically, our proposed model with multi-path dilated convolution (MPDC U-Net) performs significant better than the original U-Net, which proves that the multi-path dilated convolution can improve the lung segmentation performance against the original U-Net. In addition, our DDLA U-Net model performs better than U-Net extends with DLA consistently, which shows that the benefits of adding dense connections into DLA.

To conclude, these results indicate that by incorporating multi-path dilated convolution and dense deep layer aggregation into the original U-Net will significantly improve performance for lung segmentation in

chest radiographs.

## 6 Conclusion

In this paper, we propose a multi-path aggregation U-Net for segmenting lung in chest X-ray images. Specifically, we propose two improvements upon U-Net, the first one is multi-path dilation convolution with different dilation rate, this enables the model has the ability to extract multi-scale semantic features and the receptive field of the model can cover the whole spatial dimension of input image. The second one is dense deep layer aggregation, which further strength the original skip connections in the U-Net. The extensive experiments on three publicly available datasets indicate that our proposed method achieves the best performance compared with other

standard segmentation models and chest X-ray lung segmentation methods. One limitation of our approach is that it only performs one task, i.e. lung segmentation. In the future, we plan to integrate with more downstream tasks such as pneumonia classification to provide more comprehensive pipeline for CAD of chest radiographs.

## Ethics approval and consent to participate

Our manuscript does not involve research manuscripts of human participants, human data or human tissues, so our manuscript does not require the statement of ethical approval and ethical consent.

## Consent for publication

Our manuscript does not contain any individual person's data in any form, so we do not need the consent of others.

## Data Availability

The data included in this paper are available without any restriction.

## Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

## Funding Statement

This work was supported by the Research Foundation for Outstanding Young of Education Bureau of Hunan Province (No.18B571).

## Authors' contributions

The main idea of this paper is proposed by Zhi Gong. The algorithm design and experimental environment construction are jointly completed by Ling Ma and Xiaomao Hou. The experimental verification was completed by all the three authors. The writing of the article is jointly completed by Ling Ma and Xiaomao Hou. And the writing guidance, English polish is completed by Zhi Gong.

## Acknowledgement

We wish to express their appreciation to the reviewers for their helpful suggestions which greatly improved the presentation of this paper.

## References

- [1] M. R. Arbabshirani, A. H. Dallah, C. Agarwal, A. Patel, and G. Moore. Accurate segmentation of lung fields on chest radiographs using deep convolutional networks. In *Medical Imaging 2017: Image Processing*, volume 10133, page 1013305, (2017).
- [2] J. A. G. W. Blajer and M. Krawczyk. The Inverse Simulation Study of Aircraft Flight Path Reconstruction. *Transport*, XVII(3),103–107, (2002).
- [3] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE transactions on medical imaging*, 33(2),577–590, (2013).
- [4] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, (2017).
- [5] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao. Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images. *IEEE Transactions on Medical Imaging*, (2020).
- [6] B. V. 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*, 10(1),19–40, (2006).
- [7] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, (2015).
- [8] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In *European conference on computer vision*, pages 630–645, (2016).
- [9] M. H. Hesamian, W. Jia, X. He, and P. Kennedy. Deep learning techniques for medical image segmentation: Achievements and challenges. *Journal of digital imaging*, 32(4),582–596, (2019).
- [10] M. H. Hesamian, W. Jia, X. He, and P. J. Kennedy. Atrous Convolution for Binary Semantic Segmentation of Lung Nodule. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1015–1019, (2019).
- [11] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, (2017).
- [12] N. Ibtehaz and M. S. Rahman. MultiResUNet: Rethinking the U-Net architecture for multimodal

- biomedical image segmentation. *Neural Networks*, 121,74–87, (2020).
- [13] 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*, 4(6),475, (2014).
- [14] Q. Kang, Q. Lao, and T. Fevens. Nuclei Segmentation in Histopathological Images Using Two-Stage Learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 703–711, (2019).
- [15] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. (2015).
- [16] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553),436–444, (2015).
- [17] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng. H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE transactions on medical imaging*, 37(12),2663–2674, (2018).
- [18] 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*, pages 3431–3440, (2015).
- [19] A. A. Novikov, D. Lenis, D. Major, J. Hlad, M. Wimmer, and K. Bühler. Fully convolutional architectures for multiclass segmentation in chest radiographs. *IEEE transactions on medical imaging*, 37(8),1865–1876, (2018).
- [20] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshain, and L. Antiga. PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, pages 8024–8035, (2019).
- [21] R. Rashid, M. U. Akram, and T. Hassan. Fully convolutional neural network for lungs segmentation from chest X-rays. In *International Conference Image Analysis and Recognition*, pages 71–80, (2018).
- [22] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, (2015).
- [23] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert. Attention gated networks: Learning to leverage salient regions in medical images. *Medical image analysis*, 53,197–207, (2019).
- [24] J. Shiraiishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. ichi Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi. 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*, 174(1),71–74, (2000).
- [25] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, (2014).
- [26] S. Stirenko, Y. Kochura, O. Alienin, O. Rokovy, Y. Gordienko, P. Gang, and W. Zeng. Chest X-ray analysis of tuberculosis by deep learning with segmentation and augmentation. In *2018 IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO)*, pages 422–428, (2018).
- [27] F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, (2015).
- [28] F. Yu, D. Wang, E. Shelhamer, and T. Darrell. Deep layer aggregation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2403–2412, (2018).
- [29] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6),1856–1867, (2019).

## Author biography



**Ling Ma.** Master of Computer Science, Associate Professor. Graduated from Central South University in 2006. Worked in Hunan University of Information Technology. Her research interests include big data and computer image.



**Xiaomao Hou.** Master of Computer Engineering, Professor. Graduated from Hunan University in 2014. Worked in Hunan University of Information Technology. His research interests include computer technology and image processing.

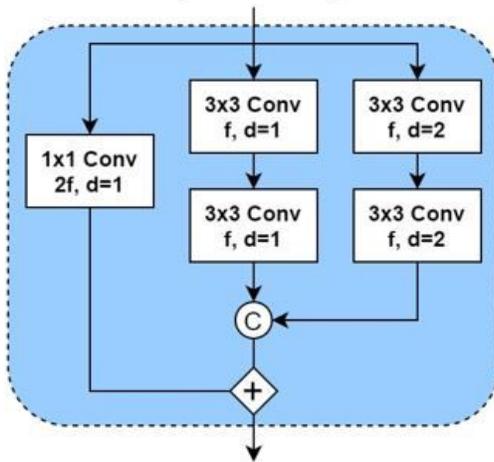


**Zhi Gong.** Master of Software Engineering, Senior Experimenter. Graduated from Hunan University in 2017. Worked in Hunan University of Information Technology. Her research interests include machine learning and data mining.

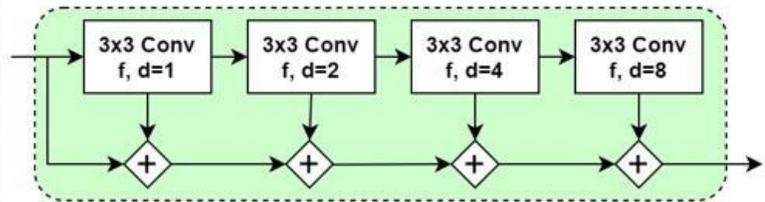


# Figures

## Multi-Path Dilated Residual Block (MPDRB)



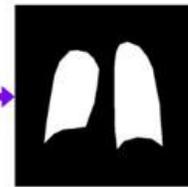
## Dilated Residual Block (DRB)



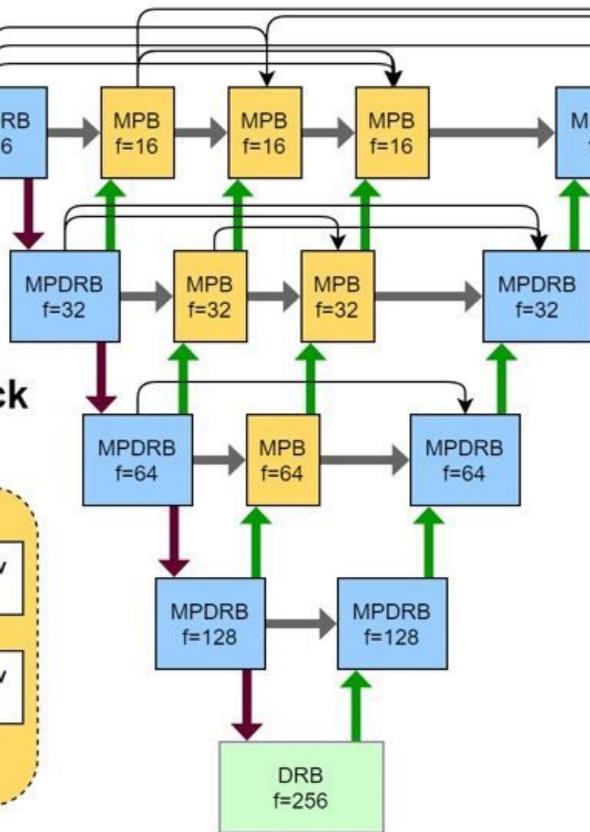
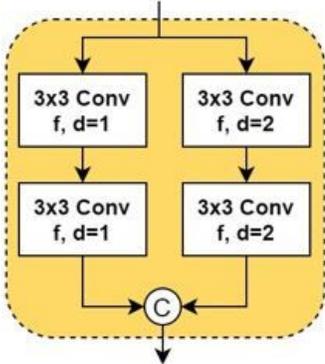
Input image



Segmentation mask



## Multi-Path Block (MPB)



### Legend

- f** Number of features
- d** Dilation rate
- Conv** Convolution, BN and Leaky ReLU
- Max-pooling 2x2
- Deconvolution 3x3, BN and Leaky ReLU
- Copy and Concat
- Conv 1x1, Leaky ReLU
- Concatenation
- Addition
- Dense Connection

Figure 1

The proposed segmentation network architecture.

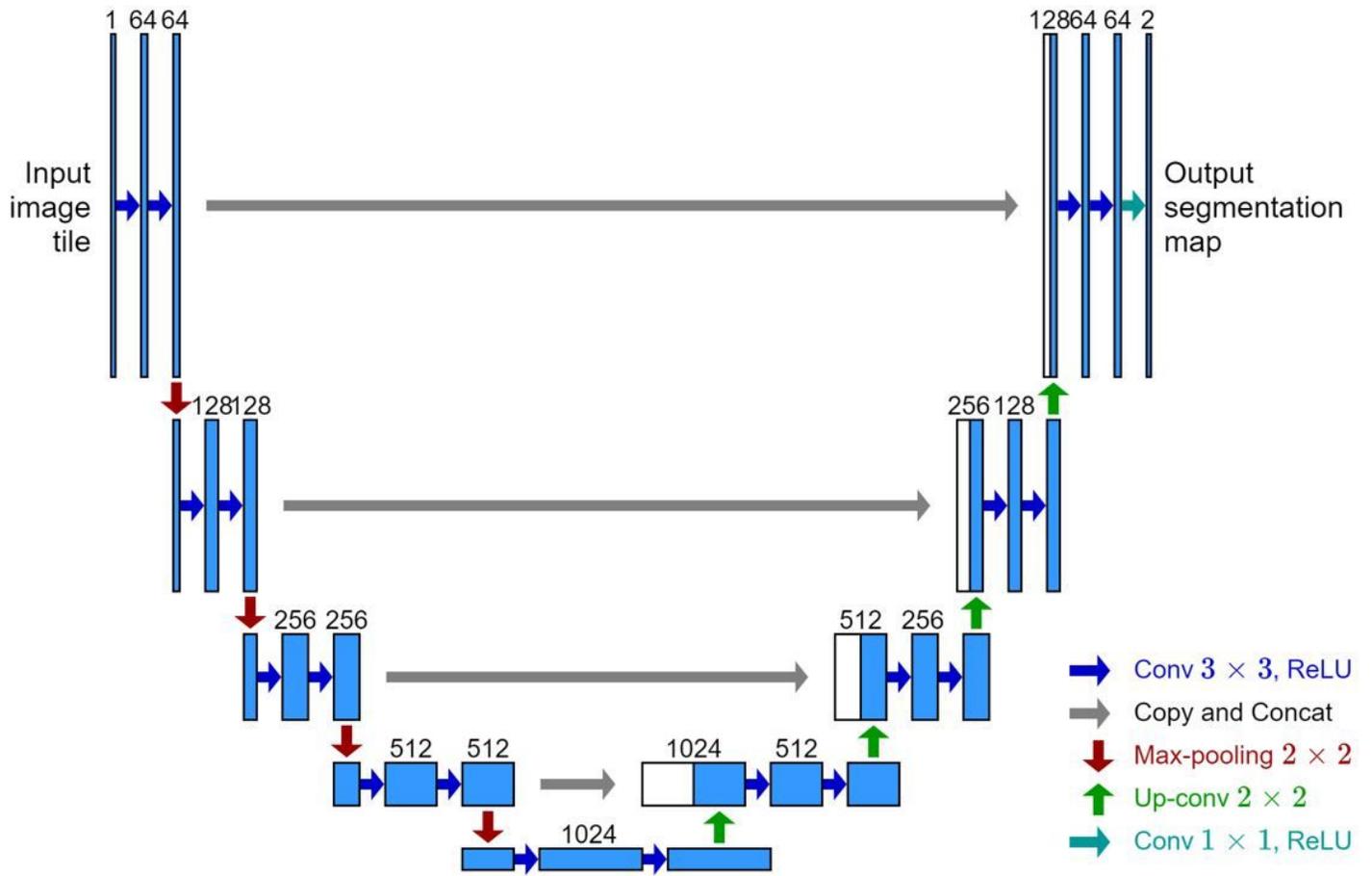


Figure 2

The conventional U-Net architecture. Blue boxes represent feature maps. Different operations are denoted by arrows with different color.

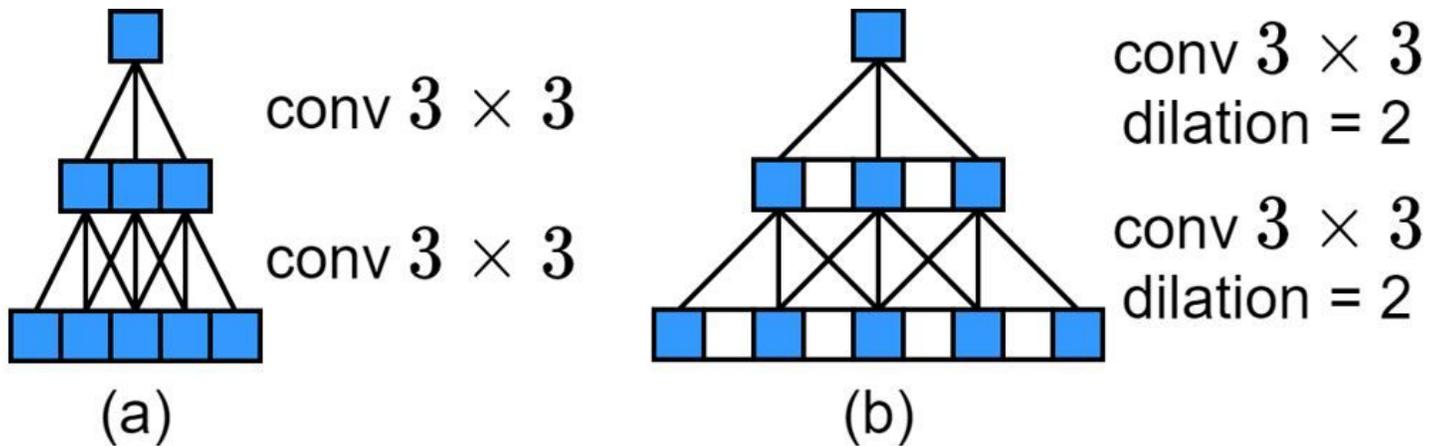
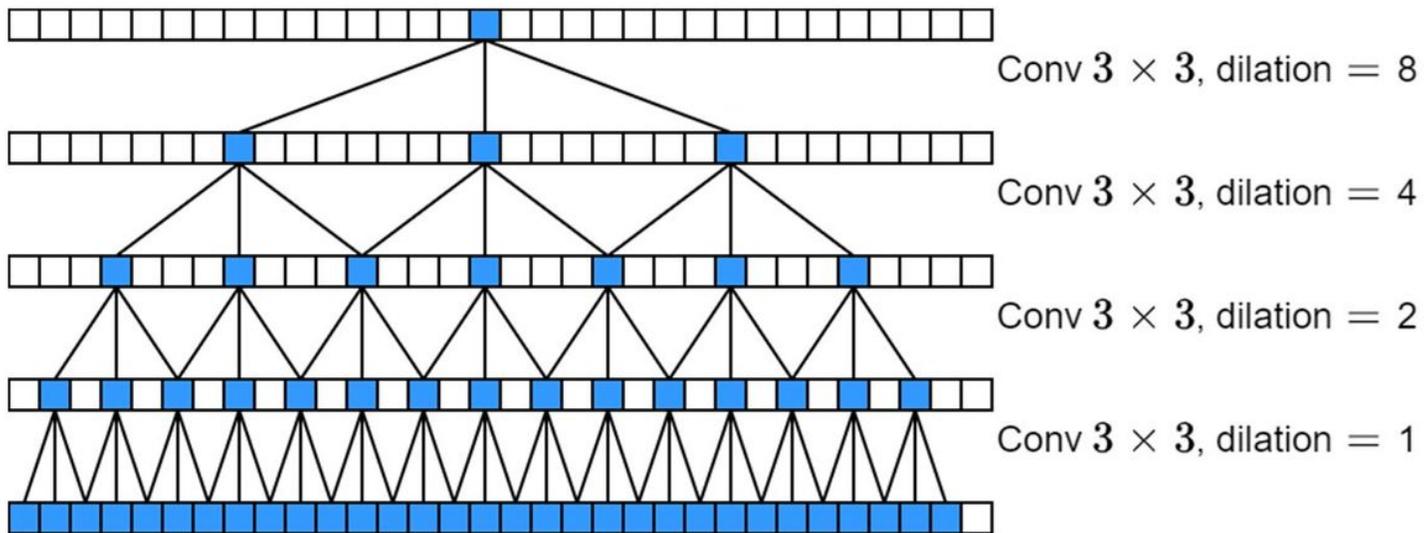


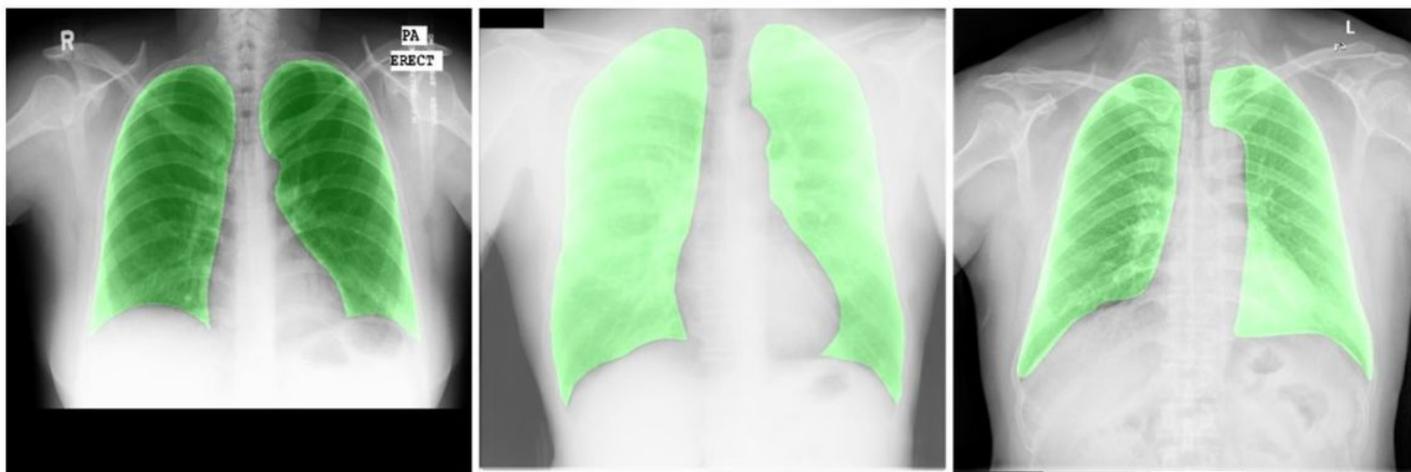
Figure 3

The receptive fields of two branches used in MPDRB block.



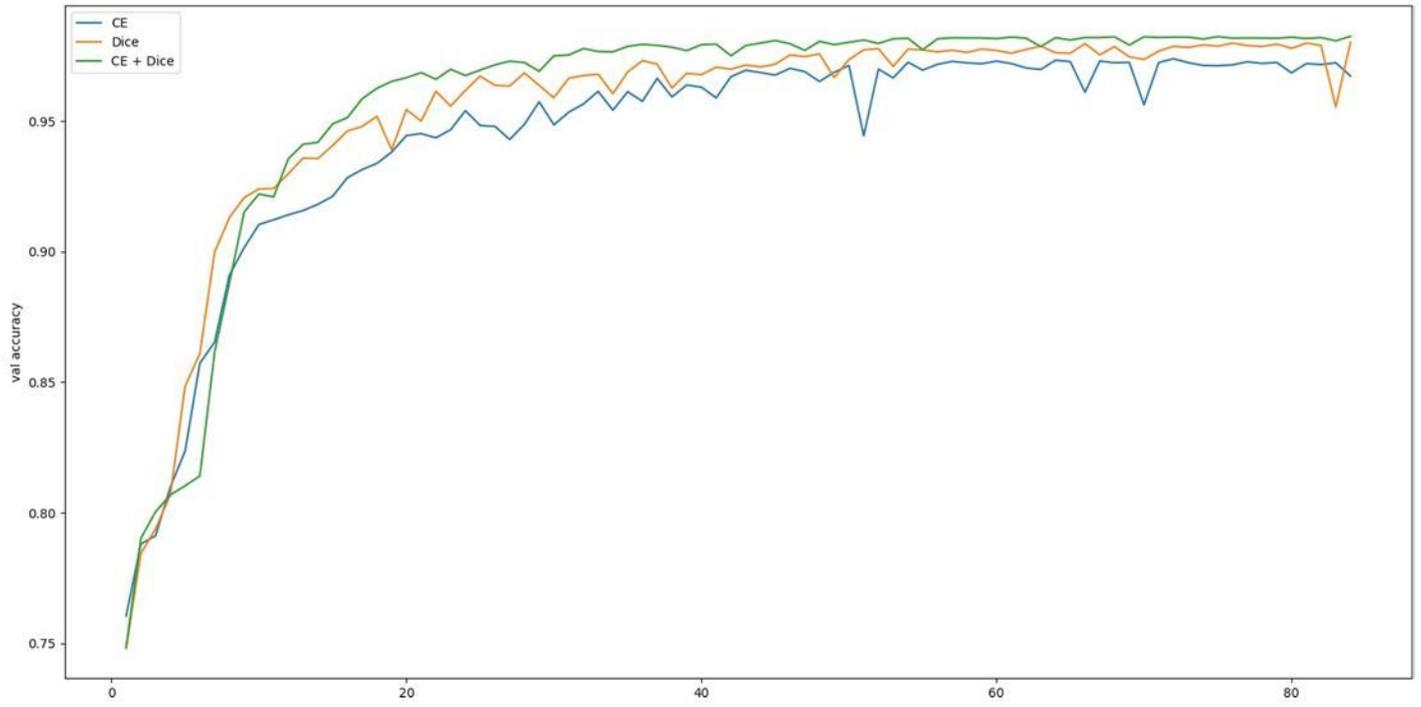
**Figure 4**

The receptive field of one convolution operation in the DRB.



**Figure 5**

Chest radiographs and corresponding manual lung segmentation (shown in green) from 3 datasets used in our study (From left to right: MC, JSRT and Shenzhen dataset).



**Figure 6**

Accuracy curves on validation set for different loss functions.



**Figure 7**

A visual comparison of lung segmentation results between U-Net and our proposed MPDC DDLA U-Net model.