An Improved Dice Loss for Pneumothorax Segmentation by Mining the Information of Negative Areas

LU WANG[1], CHAOLI WANG [ID]2, ZHANQUAN SUN[2], AND SHENG CHEN [ID]2, (Member, IEEE)

[1]College of Science, University of Shanghai for Science and Technology, Shanghai 200093, China
[2]School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

Corresponding author: Chaoli Wang (clwang@usst.edu.cn)

ABSTRACT The lesion regions of a medical image account for only a small part of the image, and a critical imbalance exists in the distribution of the positive and negative samples, which affects the segmentation performance of the lesion regions. Dice loss is beneficial for the image segmentation involving an extreme imbalance of the positive and negative samples but it ignores the background regions, which also contain a large amount of information. In this work, we propose an improved dice loss that can mine the information in background areas and modify network architecture to improve performance. The improved dice loss called weighted soft dice loss (WSDice loss). Our loss function gives a small weight to the background area of the label, so the background area will be added to the calculation when calculating dice loss. It can also soft the hard label in the lesion area to increase the robustness of the model to noise label. What's more, we propose to cascade Focal loss and WSDice loss. Focal Loss is a Distribution-based loss function, WSDice Loss is a Region-based loss function, the optimization directions of them are different. The cascaded loss function can make full use of the advantages of both and greatly improve model performance. In addition, we add a simple but effective channel attention module to the decode module of U-net. We experimented on the ChestX-ray8 datasets. Compared with Dice loss, WSDice loss improves the dice coefficient by 1.59%, cascaded loss function can improve dice coefficient by 7.81%. The improved in model architecture can increase the dice coefficient by 1.36%.

INDEX TERMS Image segmentation, sample distribution, dice loss.

I. INTRODUCTION

The occurrence of pneumothorax is a medical condition that refers to the gas entrapment caused by a gas entering the pleural cavity. The occurrence of pneumothorax can affect the respiratory movement, cause breathing difficulties, affect the blood flow, reduce the blood pressure, lead to pleurisy and cardiovascular disorders, and may even be life threatening. Therefore, the occurrence of a pneumothorax needs to be discovered and treated as soon as possible [1]. X-ray examination is an important method for diagnosing a pneumothorax [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino [ID].

Currently, the detection of pneumothorax mainly relies on artificial observation. However, this method depends on the subjective experience of doctors, and the misdiagnosis rate is high. The traditional techniques of pneumothorax X-ray segmentation require the manual extraction of features. For example, Geva O *et al.* determine the position of the pneumothorax region by analyzing the local and global textural features [3]. Chan YH *et al.* use the local binary mode and support vector machine for pneumothorax detection [4]. However, these methods are time-consuming and laborious, and the features of the manual design contain incomplete information, which leads to a low accuracy of the pneumothorax segmentation.

With the successful application of deep learning methods represented by convolutional neural networks in the field of

computer vision, the methods based on neural network have demonstrated considerable success in the field of medical image segmentation [5], [6]. Wang *et al.* first used deep learning to perform pneumothorax segmentation [7], and achieved satisfactory results. In addition, Wang *et al.* published a large-scale lung disease dataset, namely, ChestX-ray8, contributing greatly to the field of automatic detection of lung diseases. Several researchers used deep-learning methods to detect pneumothorax regions [8], [9], and these approaches considerably improved the pneumothorax segmentation performance.

U-net [10] uses an encoder-decoder symmetric structure based on fully convolutional network (FCN) [11]. Milletari *et al.* proposed a 3D deformation structure of the U-net network structure, known as the V-net [12]. The V-net structure uses the dice coefficient loss function instead of the traditional cross entropy loss function. The Unet++ [13] improves the U-net in terms of the skip connection, introduces the idea of deep supervision and can achieve a high precision and speed through the implementation of model pruning.

In the cross entropy loss function, all the pixels in the image are treated equally. However, it leads to the network being dominated by the classes with more pixels, and it is difficult for the network to learn the features of small objects. Consequently, the network segmentation performance for the small objects is extremely poor. Researchers have performed considerable work in an attempt to overcome these problem. Tsung-Yi Lin *et al.* proposed an improvement in the cross entropy loss to address the problem of unbalanced data distribution, and this improved function was known as the focal loss [14]. Milletari *et al.* proposed the use of the dice loss [12] to solve the limitation. However, dice loss directly ignoring the background regions can lead to a considerable loss of information. Although the negative sample regions do not contain the lesion regions, they contain the characteristic information of the healthy regions. The network identifies the part of the information and compares it with the information of the lesion regions to better distinguish between the healthy and lesion regions.

Considering these disadvantages of dice loss, Sudre C H *et al.* proposed the generalized dice loss and explored the effects of different sample imbalance rates on the model performance [15]. However, an effective method to solve these problems was not proposed. Shen *et al.* generated the weight based on the number of label categories, number of samples and number of pixels in the same category, which was multiplied by the dice loss [16]. This method could effectively solve the problem of the imbalance in multiclass segmentation. However, in the semantic segmentation tasks with only one kind of lesions, this method could not address the imbalance in the distribution of the lesion and background regions. Liu Y C *et al.* simultaneously segmented the lesion regions and the organ regions in which the lesion regions were located [17]. They used the area ratio of lesion regions to the organ regions as the weight to strength the loss of lesion

regions. But this method needs the label of lesion regions and organ regions at the same time, the cost of label is expensive and it is not convenient to use. Therefore, we attempted to improve the dice loss by introducing a weighted soft dice loss (WSDice loss). In the proposed approach, according to the weight of the label generation, the negative sample regions are included in the calculation of the dice coefficient, which can considerably improve the performance of the network segmentation. What's more, we propose to cascade Focal loss and WSDice loss. The cascaded loss function can make full use of the advantages of both and greatly improve model performance.

The major contributions of this paper are summarized as follows:

1) We improve dice loss to mine more information of negative areas in medical images.
2) We propose cascading Focal loss and WSDice loss which can greatly improve model performance.
3) We improve the U-net structure by using the core idea of SE-ResNeXt [18].

The improved loss function can ensure that dice loss is used to address the unbalanced sample distribution problem, and it can deeply mine the information in the positive and negative samples. In this paper, we use the data of the ChestX-ray8 database to perform experiments, and the experiment results prove that the proposed method can effectively improve the segmentation accuracy.

II. WEIGHTED SOFT DICE LOSS
A. FOCAL LOSS
The lesion regions of the medical image account for only a small part of the image, and a critical imbalance exists in the distribution of the positive and negatives samples, which affects the segmentation performance of the lesion regions. In 2017, Tsung Yi Lin proposed the use of the focal loss [14] to address the problem of the imbalanced distribution of the positive and negative samples in the field of object detection. In medical image segmentation, because the labels are based on the doctor's experience, erroneous annotations may be present. The focal loss performs the classification of each pixel, owing to which, the erroneous annotations considerably influence the network. Although the focal loss has limitations, this loss function can effectively address the problem of imbalanced samples; therefore, this function is likely also effective in the field of medical image segmentation. In general, because the cross entropy loss treats all the samples equally, the loss value of a part may be large when the number of negative samples is large. This value may dominate the loss of the positive samples and lead the network to deeply mine the negative sample regions; however, this situation is in contrast to the goal of the lesion region segmentation. To solve this problem, the focal loss divides all the samples into easy and hard samples according to the confidence degree of the model. The number of pixels in the lesion regions are small; therefore, it is difficult to extract

features, which reduces the confidence of the network, and thus these samples are hard samples. The number of pixels in the negative sample regions are large and their features are obvious, and thus, these samples are easy samples. The focal loss uses the confidence of each sample to generate a dynamic weight to increase the loss value of the hard samples and reduce the loss value of the easy samples to ensure that the network can optimize the process considering the lesion regions. The focal loss can be expressed as

$$L_{fl} = -\alpha (1-p)^{\gamma} \log (p) \tag{1}$$

where p is the model's estimated probability, α, γ are two hyperparameters, α is used to adjust the distribution of the easy samples, and $(1-p)^{\gamma}$ is the dynamic scaling factor, which is used to adjust the distribution of the hard samples.

B. DICE LOSS
The dice loss, proposed in the context of the V-net, is derived from the Sørensen–Dice coefficient, which is a statistical indicator proposed by Thorvald Sørensen and Lee Raymond Dice in 1945. The dice loss has been often used for medical image segmentation tasks. The dice coefficient (DC) is a set similarity measure function, which is derived from the binary classification task. The value of this coefficient ranges from zero to one, which means that $DC \in [0, 1]$, and the dice loss is expressed as $1 - DC$.

The value of this coefficient ranges from zero to one, which means that, and the dice loss is expressed as one minus the dice coefficient.

For the binary classification of the image segmentation, the dice loss can be calculated as follows:

$$L_{dice} = \frac{1}{N} (1 - DC) \tag{2}$$

where N is the number of samples, DC is the dice coefficient, it can be expressed as

$$
\begin{aligned}
DC &= \frac{2 \left| \hat{Y} \bigcap Y \right|}{\left| \hat{Y} \right| + |Y|} \\
&= \frac{2 \sum_{i=1}^{n} \sum_{j=1}^{m} \hat{y}_{ij} \cdot y_{ij} + \varepsilon}{\sum_{i=1}^{n} \sum_{j=1}^{m} \hat{y}_{ij}^2 + \sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij}^2 + \varepsilon}
\end{aligned} \tag{3}
$$

where $\hat{Y}$ represents the predicted value of the network, Y represents the true label of the sample, $\hat{Y}$ and Y are both matrices, $\hat{y}_{ij}$ and y_{ij} represent the elements in $\hat{Y}$ and Y, respectively, n and m denote the number of rows and columns of the matrix, respectively, and ε is a minimum value that prevents the occurrence of a zero denominator. Since the overlap region between the predicted value $\hat{Y}$ and the true value Y is repeatedly calculated in the denominator, the coefficient is multiplied by two in the numerator when calculating the dice coefficient, which can lead to the formation of a loss function that can be minimized when calculating the dice loss.

C. WEIGHTED SOFT DICE LOSS
The lesion region labels of the samples have a value of one and the healthy regions have a value of zero, as shown in the first line in Fig. 1. The healthy regions have a value of zero when calculating the dice coefficient, and thus, the healthy regions are ignored. The advantage is that the loss of the model is calculated considering only the segmented objective regions, which can ensure that the model focuses more on extracting the lesion regions of the samples and the information of the lesion regions, which is beneficial for the segmentation involving an extreme imbalance of the positive and negative samples. However, it is unreasonable to ignore the background regions, which also contain a large amount of information, such as the features of the nonpneumothorax regions of the lung. If the model can recognize the features of the part and compare them with the features of the pneumothorax region, the effect of the pneumothorax segmentation can be considerably enhanced. In other words, the use of the dice loss causes the loss function of the negative samples being unable to be calculated and a part of the information is lost, which adversely affects the back propagation and destabilizes the training process.

The improved loss function can include the negative sample regions of the image into the loss calculation and retains the advantage of the dice loss in the problem of the imbalanced distribution of the positive and negative samples. The proposed loss function is known as the weighted soft dice loss, and it can be defined as

$$
\begin{aligned}
L_{WSDice} &= 1 - \frac{2 \left| W \left(2\hat{Y} - 1\right) \bigcap W (2Y - 1) \right|}{\left| W \left(2\hat{Y} - 1\right) \right| + |W (2Y - 1)|} \\
&= 1 - \frac{2 \sum_{i=1}^{n} \sum_{j=1}^{m} \hat{G} \cdot G + \varepsilon}{\sum_{i=1}^{n} \sum_{j=1}^{m} \hat{G}^2 + \sum_{i=1}^{n} \sum_{j=1}^{m} G^2 + \varepsilon}
\end{aligned} \tag{4}
$$

Here $\hat{G}$ is $w_{ij} \left(2\hat{y}_{ij} - 1\right)$, G is $w_{ij} \left(2y_{ij} - 1\right)$. Where, n and m denote the width and height of the image, respectively. w_{ij} is the weight generated according to the segmentation label, $w_{ij} = y_{ij} (v_2 - v_1) + v_1$, where v_1 and v_2 are the hyperparameters, and $0 \leq v_1 \leq v_2 \leq 1$, $v_2 = 1 - v_1$. As shown in Fig. 1, to include the negative sample regions in the calculation, we use items $2\hat{y}_{ij} - 1$ and $2y_{ij} - 1$ to change the [0, 1] range of the predicted values and labels to $[-1, 1]$. Similar to Dice Loss, when WSDice is used in multi-class segmentation tasks, for each class of segmentation object, we regard it as foreground and the rest area of the image as the background. After all category objects are segmented, we integrate them into the one mask.

The weight w_{ij} generated using the labels has two effects:

First, the negative samples are considered in the loss function to ensure that the network can focus on extracting the features of the negative sample regions. At the same time, due to the small weight, even if the range of the negative samples is large, they do not account for a large proportion of the total loss, and thus the advantage of the dice loss in terms of the imbalance of the positive and negative samples is retained.
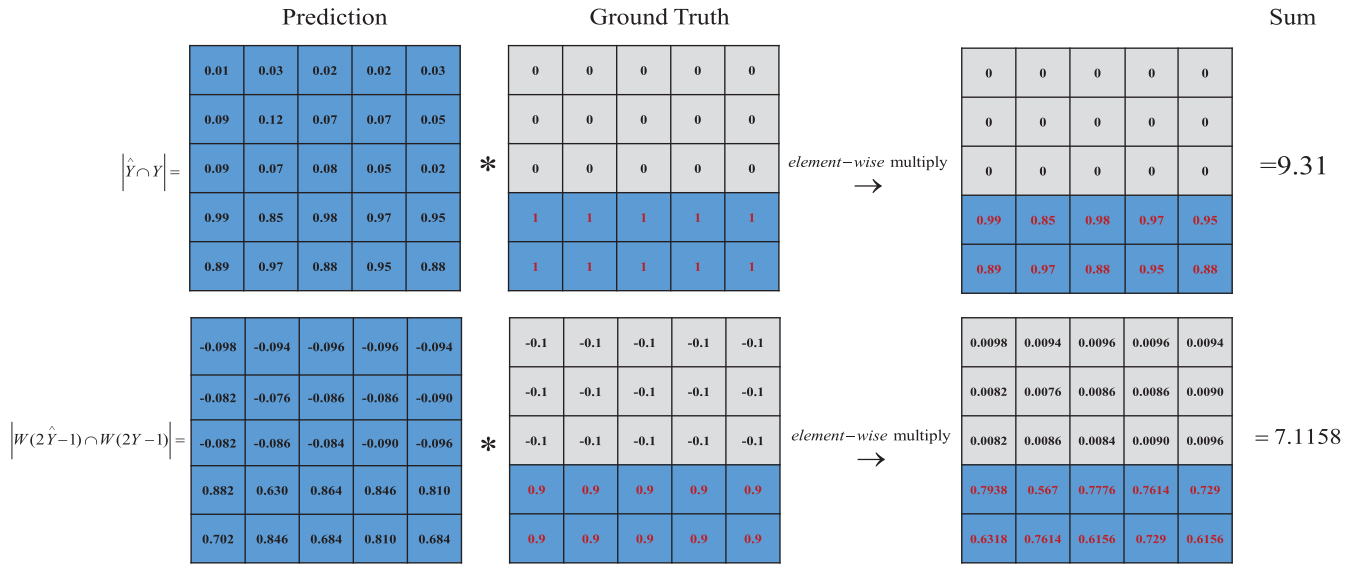
FIGURE 1. The first and second lines show simple examples of the dice loss and weighted soft dice loss, respectively. It is assumed that the hyperparameters are $v_1 = 0.1$ and $v_2 = 0.9$.

Second, the concept of label smoothing [19] is introduced. The corresponding value of the label is zero or one, which corresponds to the hard labels. This aspect makes the network over believe its judgment, and the network learns considering the large gap between the predicted value and the interpolation of the label. The small amount of data in the training set is not sufficient to represent all the sample features, which leads to overfitting. In our proposed loss function, the weight generated from the labels of the sample softens the one hot type hard label, which reduces the confidence of the positive samples in the label, increases the confidence of the negative samples, and suppresses the difference between the output of the positive and negative samples. This approach is equivalent to adding a certain amount of noise in the real distribution of the data, which can have a certain regularization effect and prevent the network from overfitting.

III. METHOD

A. DATASET

The dataset used in this paper was derived from the large scale lung disease open dataset named ChestX-ray8 [7] developed by Wang *et al.* This dataset contains more than 110, 000 lung X-ray images of more than 30, 000 patients, corresponding to 14 common lung diseases. We used the pneumothorax segmentation dataset. The training set of this part of the dataset includes 12, 089 lung X-ray images, including 2, 669 pneumothorax images and 3, 576 pneumothorax lesion areas (some images contain multiple lesion regions). There exist 9, 420 normal chest X-ray images, with the image size of 1024 × 1024. The regions of most lesions in the dataset have no more than 128 pixels, and the lesion regions in the largest X-ray image have less than 512 pixels, which means that represents a critical imbalance in the distribution of the

positive and negative samples in the dataset. In other words, the positive samples (sick regions) are considerably smaller than the negative samples (normal regions).

B. PREPROCESSING

Considering the influence of the network computation, video memory size and batch size on the model performance, the input image size of the proposed model is 576 × 576. For the data augmentation, we use the contrast limited adaptive histogram equalization (CLAHE) technique [20], horizontal flips, center crop, random brightness, random contrast adjustment, which are commonly used in the medical image segmentation. The upper threshold value for contrast limiting is selected randomly from (2.0, 12.0), and the size of grid for histogram equalization is selected randomly from (3, 15) in CLAHE. The ratio of center crop is selected randomly from (0, 0.25). The factor range for changing brightness is (0.1, 0.8) in random brightness, and the factor range for changing contrast is (0.2, 0.8) in random contrast adjustment. The probability of applying CLAHE is 0.6, and the probability of using other data augmentation methods is 0.5.

C. MODEL ARCHITECTURE

Compared with natural images, medical data is more difficult to obtain due to privacy issues, the amount of medical data is small. In addition, the complexity and individual differences in medical images will increase the difficulty of medical image segmentation. The U-net model [10] can solve these problems well, and its use has generated satisfactory results in medical image segmentation since it was proposed in 2015. This paper also uses the U-net structure albeit with certain improvements.
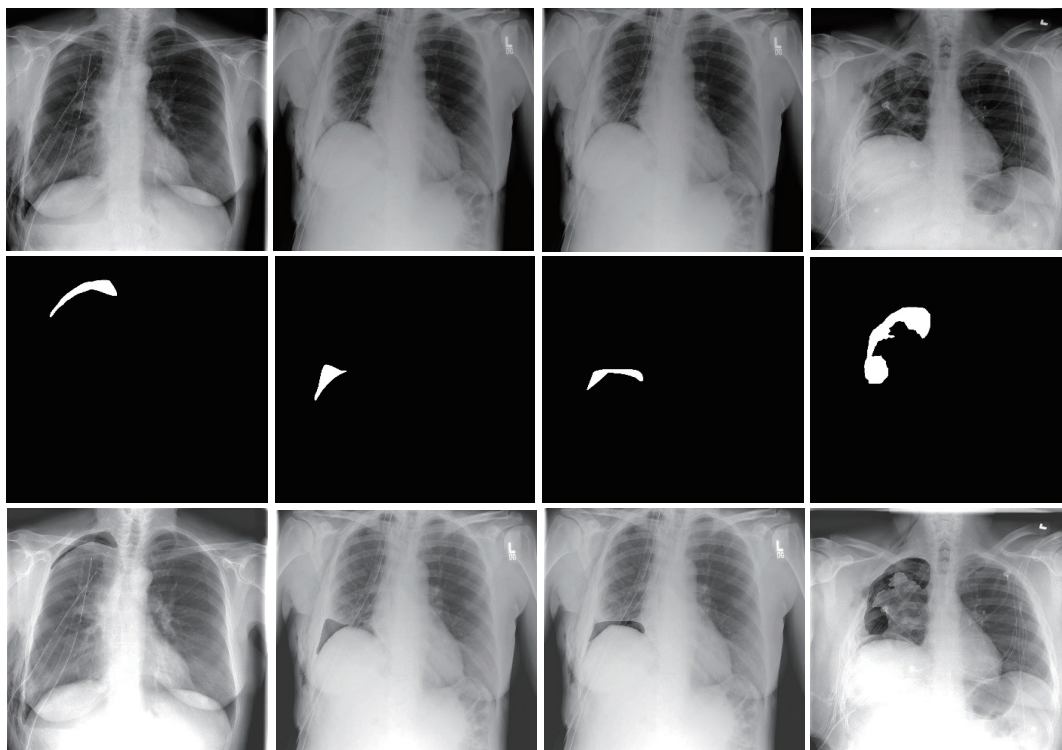
FIGURE 2. Pneumothorax image samples from the ChestX-ray8 dataset. The images in the first row are the chest X-ray images, the corresponding pneumothorax mask images are in the second line and the last row is the chest X-ray images that mark the pneumothorax areas.
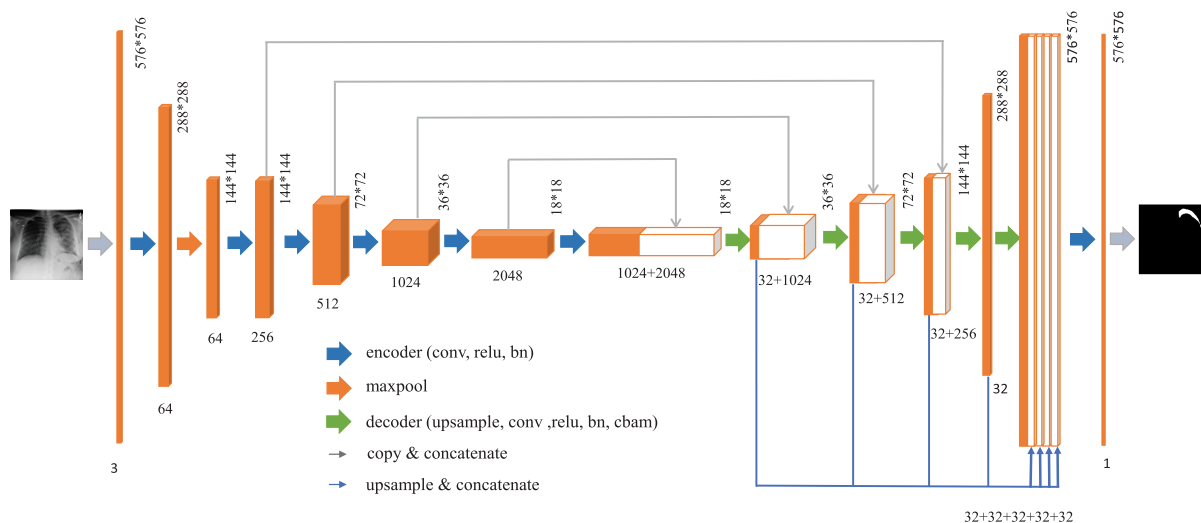

FIGURE 3. Network architecture. Each color box corresponds to a multichannel feature map. The number at the top of the box indicates the size of the feature map. The number at the bottom indicates the number of channels. The arrows indicate the different operations.

We make some improvements based on the U-net, as shown in Fig. 3. A certain amount of information is lost in the downsampling operation of the encoding module, and the upsampling operation can recover the dimension of the feature map. However, this operation cannot completely restore the lost information. The original U-net fuses the feature map in the downsampling process with the feature map in the upsampling process, which allows the complete utilization of the rich information in the encoding module, thereby compensating for the information that cannot be restored in the decoding module and improving the accuracy of the image segmentation. However, in the U-net decoding module,

the different depth layers also contain different information, and the U-net does not make better use of this information.

To more adequately mine the information of the feature map, we add a simple but effective channel attention module [21] to the decoding module. We use 1×1 convolution filter to compress the feature maps' channels in the decoding module, extract key features, and then concatenate the key features to the features before the output layer. Finally, the fused features are normalized and activated with Relu. The activated features become the input of the output layer.

Feature maps in convolutional neural networks (CNN) generally have many channels, the more feature channels, the stronger the expression capabilities of CNN, generally. But the importance of the features in different channels is different, only a small number of channels play a key role in the feature extraction, and most of channels only play an auxiliary role. However, if the number of feature channels used is too small, it will lead to weak feature extraction capabilities of the network. All feature channels are often considered to be equally important in CNN, which will lead to the significant loss of information in the key channels.

In our method, 1×1 convolution filter is used to compress the channels in the decoding module. After the loss convergence, we can regard the compressed feature channels as the key channels. We concatenate them before the output layer, which is equivalent to weighting these important feature channels, highlighting the importance of these feature channels. In addition, different depth features in the decode module has different receptive fields. The information of feature maps at different scales can be fully used to improve the accuracy of multi-scale object segmentation after cascading them. Finally, we use 1×1 convolution filter to compress the feature map channels to 32 in the decode module, and the ammount of parameters just have a slight increase.

We have tried the same method in the encode module, rather than increase the performance, this method aggravated it. The features in the shallow layers are not abstract enough, they can only extract simple local features such as textures and colors. So this method is only applicable to the decode module of U-net. The features of the decode module are deeply encoded by the encode module, the deeply encoded feature fusion can promote each other and improve performance.

The backbone network used in this paper is the SE-ResNeXt-50 [18]. As shown in Fig. 4, the SE-ResNext module is a combination of the SE block and ResNeXt. The SE block is not a complete network structure, which can be embedded in other network models. The core idea of this block is to establish the interdependence between channels. Through the loss function of the network, we can learn the feature weight. If the effective feature weight is large, the ineffective feature weight is small, and the model can achieve better results. The ResNeXt [22] proposed by Xie *et al.*, which is based on the ResNet [23] and reference of Inception [24], changes from a single path to multiple parallel paths, increases the width of the residual block and reduces
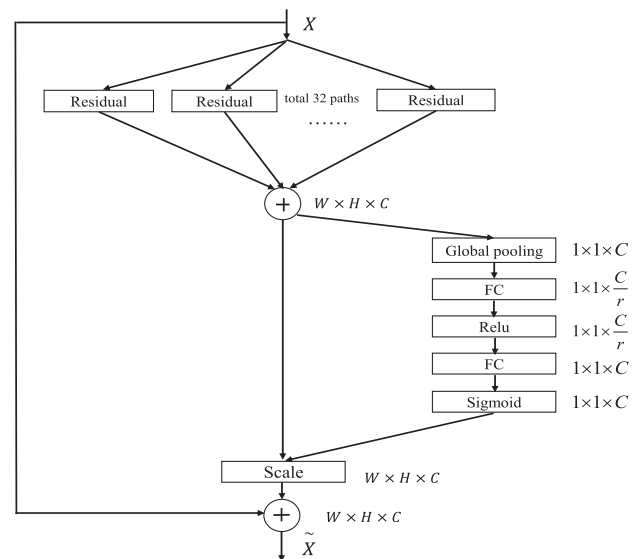

FIGURE 4. Se-ResNext module. W, H and C represent the width, height, and number of channels of the feature map, respectively.

the number of super parameters. When the ResNet parameters are used, the model depth of the ResNeXt is smaller, but the accuracy is higher. The SE-ResNeXt combines the advantages of the SE block and ResNeXt, as both of these blocks can weigh the channel of the model and increase the width of the network. Consequently, the performance of the model is high, and the amount of calculation does not increase considerably.

D. LOSS FUNCTION

As mentioned before, the dice loss ignores the negative sample regions and loses a large amount of information. Consequently, the medical image has the characteristics of fewer and smaller lesion regions and the positive and negative samples exhibit an unbalanced distribution in a single image. The proposed loss function uses an improved version of the dice loss called the weighted soft dice loss. The improved loss function can make the model focus on the lesion regions and overcome the problem of the unbalanced distribution of the positive and negative samples in a single sample of the medical image, and it can also avoid the loss of the feature information of the negative sample regions. These characteristics can considerably improve the accuracy of the medical image segmentation.

E. OPTIMIZER

We use the RAdam [25], which was recently proposed by Liyuan Liu. The new optimizer, RAdam, has many the advantages. It can not only ensure a fast rate of convergence, but it can also avoid falling into the local optimal solutions. The algorithm can dynamically turn the adaptive learning rate on or off, and the convergence result is not sensitive to the initial value of the learning rate. Only a few hyperparameters are

required to be adjusted. Therefore, this approach is convenient to use, and it exerts a desirable influence.

IV. EXPERIMENTS

A. PERFORMANCE EVALUATION

We evaluate the segmentation accuracy of the pneumothorax X-ray images considering the degree of overlap between the predicted regions and the regions of the ground truth. In our experiment, the dice coefficient was used to quantitatively evaluate the segmentation. A larger dice coefficient value corresponds to a higher segmentation accuracy. The dice coefficient can be calculated as follows:

$$DC = \frac{2\left|\hat{Y} \cap Y\right|}{\left|\hat{Y}\right| + |Y|} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (5)$$

Here, $\hat{Y}$ is the predicted result of the model, and Y is the label of the image. The true positive TP represents the common area of the manual segmentation and network prediction. The false positive FP represents the part of the model prediction area remaining after the removal of the common area of the manual segmentation. The false negative FN indicates that the artificial segmentation area removes the overlapped part with the segmentation area of the network prediction.

B. EXPERIMENTAL RESULTS

The experimental equipment used in this paper was a Tesla P100 with a 16G memory. Considering the data volume, image resolution, amount of model calculation, and computing power of the graphics card, the backbone network used in the experiment was the SE-ResNext-50. The size of the input image was 576×576, and the batch size during training was set as 16. We set epoch as 80, the initial learning rate was set as 0.001 and the learning rate schedule is gradual warmup, the warmup time is 2 epochs, the cosine learning rate schedule was used after warmup. We use the default parameters in the RAdam optimizer. The backbone was pretrained in ImageNet and we freeze the first three blocks. The threshold used in inference is 0.39, it's means that pixels with a probability greater than 0.39 are determined as the lesion area after sigmoid, and the predicted lesion area with an area less than 3500 is deleted. We divide the training set, validation set and test set according to the ratio of $6 : 2 : 2$.

To demonstrate the effectiveness of the proposed method, in this paper, the original U-net and the improved U-net were trained on the dataset by using the cross entropy loss, focal loss, dice loss, and weighted soft dice loss, while keeping the other conditions unchanged. The focal loss can effectively overcome the problem of segmentation of small object samples in the dataset, and its segmentation effect is better. The dice loss only focuses on the positive sample regions and is less affected by the imbalanced distribution of the positive and negative samples, which can alleviate the problem of the sample imbalance. The focal loss and dice loss thus have different regions of interest and advantages. Combining the advantages of both the losses, the segmentation effect can
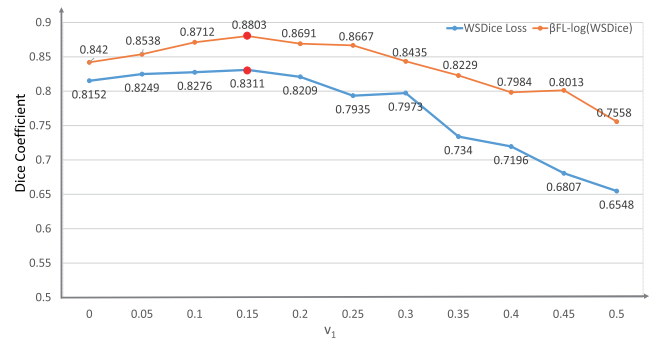

FIGURE 5. Comparison of segmentation performance with different settings of hyperparameter v_1 in the ChestX-ray8 dataset. $\beta FL - log(WSDice)$ is the combined loss function, and the β is 7.

be considerably improved. Therefore, this paper also uses different loss functions involving a weighted combination. In other words, the results of the focal loss and dice loss as a weighted combination, and those of the focal loss and the weighted soft dice loss as a combination, are compared. The experiments show that the method of using the combined loss function can considerably improve the performance of the network for pneumothorax segmentation.

In this paper, the improved dice loss function introduces the hyperparameters v_1 and v_2, which represents the weights of the negative and positive samples, respectively, $0 \le v_1 \le v_2 \le 1, v_2 = 1 - v_1$. In other words v_1 lies between 0 and 0.5. These two hyperparameters are adjusted according to the proportion of the positive and negative samples in different datasets. Fig. 5 shows a comparison of the dice coefficients of the models trained with different hyperparameters on the pneumothorax segmentation dataset used in this paper. The blue ployline indicates WSDice and the orange ployline is a cascade loss function of Focal loss and WSDice loss. In the cascade loss function, the coefficient of Focal loss is 8 and the WSDice loss is scaled by *log*. WSDice loss belong to [0, 1], it will approximately equal to the 8 times Focal loss after use *log* to scale. It can be seen that the network segmentation effect is the best when $v_1 = 0.15$ and $v_2 = 0.85$.

When using the loss function of the weighted combination, the results pertaining to different weight proportions of the two loss functions are also considerably different. When multiple loss functions are weighted and combined, it is necessary to maintain the proportion of the different loss functions to the total loss as constant. The hyperparameters can be adjusted according to this method by starting from the value that can maintain this proportion as nearly constant and later finetuning. The optimal weight value is near the starting value. As shown in Fig 7, a large number of experiments were performed using the combination of the focal loss and dice loss and the combination of the focal loss and weighted soft dice loss.

To optimize the experimental results, the improved network was used to determine the hyperparameter values of the focal loss. From Fig. 6, it can be seen that the segmentation is the best when $\alpha = 0.3, \gamma = 3.5$. In the dataset of the
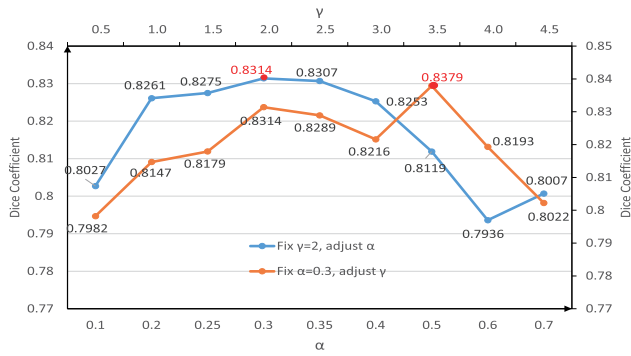
FIGURE 6. Hyperparameter adjustment of the focal loss with the improved network in the ChestX-ray8 dataset. First we fix $\gamma = 2$ and adjust α, when $\alpha = 0.3$, we get the highest dice coefficient. Then we fix $\alpha = 0.3$ and adjust γ, when $\gamma = 3.5$, we get the highest dice coefficient. Therefore, we regard $\alpha = 0.3$, $\gamma = 3.5$ as the optimal hyperparameter in Focal loss.
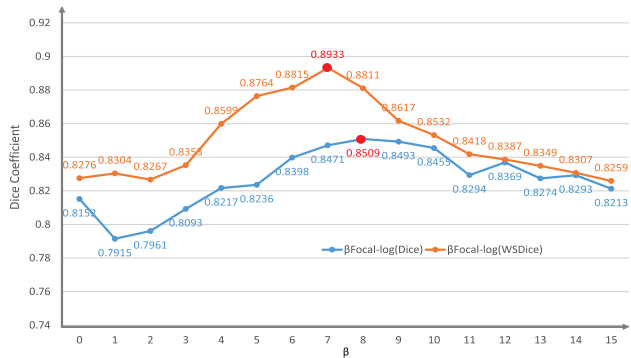

FIGURE 7. Adjustment of cascading coefficient of cascading loss function in the ChestX-ray8 dataset. $\beta FL - log(WSDice)$ has the best perfoemance when $\beta = 7$ and $\beta FL - log(Dice)$ get the highest dice coefficient when $\beta = 8$.

pneumothorax segmentation used in this paper, we fix the values of the hyperparameter of the focal loss as 10 to the weight of maintaining the balance of the two loss functions. Starting from this weight, finetuning is performed in the up and down directions. It can be seen from Fig. 7 that the optimal weights of the combination of the focal loss and dice loss and the combination of the focal loss and weighted dice loss are 8 and 7, respectively.

To demonstrate the effectiveness of our method, we used a variety of loss function training models for comparing the approaches in the U-net and improved U-net cases, the experimental results are presented in Table 1, and we can observe the segmentation results of different loss functions from Fig 8. Our method can increase the dice coefficient by approximately 10.6%. Due to the critical imbalance of the positive and negative sample distribution in the data, the performance of the cross entropy loss is the worst. Compared with that of the model of the cross entropy loss training, the performance of the focal loss is considerably improved. Because dice loss is less affected by the seriously unbalanced distribution of the positive and negative samples in the image, and its performance is considerably better than that of the cross entropy loss. However, a large number of small lesion regions is

TABLE 1. Comparison of performance of different loss functions on vanilla U-net and improved U-net in the ChestX-ray8 dataset, where $\alpha = 0.3$, $\gamma = 3.5$ in Focal Loss, $v_1 = 0.15$, $v_2 = 0.85$ in WSDice Loss, $\beta = 8$, $\alpha = 0.3$, $\gamma = 3.5$ in $\beta FL - log(Dice)$, and $\beta = 7$, $\alpha = 0.3$, $\gamma = 3.5$ $v_1 = 0.15$, $v_2 = 0.85$ in $\beta FL - log(WSDice)$.

Loss Function	Dice Coefficient	
	U-net	Improved U-net
Cross Entropy Loss	0.7831	0.7968
Focal Loss [14]	0.8237	0.8379
Dice Loss [12]	0.8076	0.8152
WSDice Loss	0.8113	0.8311
$\beta FL - log(Dice)$ [26]	0.8386	0.8509
$\beta FL - log(WSDice)$	**0.8792**	**0.8933**

TABLE 2. Comparison of pneumothorax segmentation performance of different papers on ChestX-ray8 dataset.

Methods	AUC	Dice Coefficient
Yao et al. [27]	0.841	0.767
Wang et al. [7]	0.806	0.7891
CheXNet [28]	0.8887	-
TieNet [29]	0.96	-
Li et al. [30]	-	0.8734
Mahajan et al. [31]	0.941	0.851
Tang et al. [32]	0.94	-
Ours	**0.9612**	**0.8933**

present in the dataset. When the segmented objective regions are too small, the prediction error has a significant impact on the dice loss, which easily destabilizes the training and to some extent affects the performance of the dice loss approach.

In addition, the dice loss can avoid the distribution imbalance of the negative and positive samples, but it also leads to considerable information loss, which affect the performance of the dice loss approach. The results in Table 1 indicate that the performance of the improved dice loss is considerably higher than that of the dice loss approach, which proves the effectiveness of the method used in this paper.

Although our method can solve the problem of the dice loss, the segmentation object considered in this paper is extremely small, and thus, the effect of the focal loss is better than that of the proposed method when segmenting extremely small objects. Moreover, the use of the dice loss can solve the problem of the unbalanced distribution of the negative and positive samples, and the focal loss can better segment small objects. Considering the characteristics of these two different loss functions, we formulated a weighted combination of the two loss functions [26]. As seen from Table 1, the performance of the combination of the focal loss and dice loss is better than that of the individual focal loss and dice loss functions. Besides, the combination of the focal loss and weighted soft dice loss is better than that of the focal loss and dice loss. This result further demonstrates the effectiveness of the proposed weighted soft dice loss. Moreover, Table 1 indicates that the improved network has a satisfactory segmentation effect, which demonstrates the excellent performance of the improved network.

We investigated some representative papers published from 2017 to 2020. All of these papers were involved in the research of pneumothorax region segmentation and
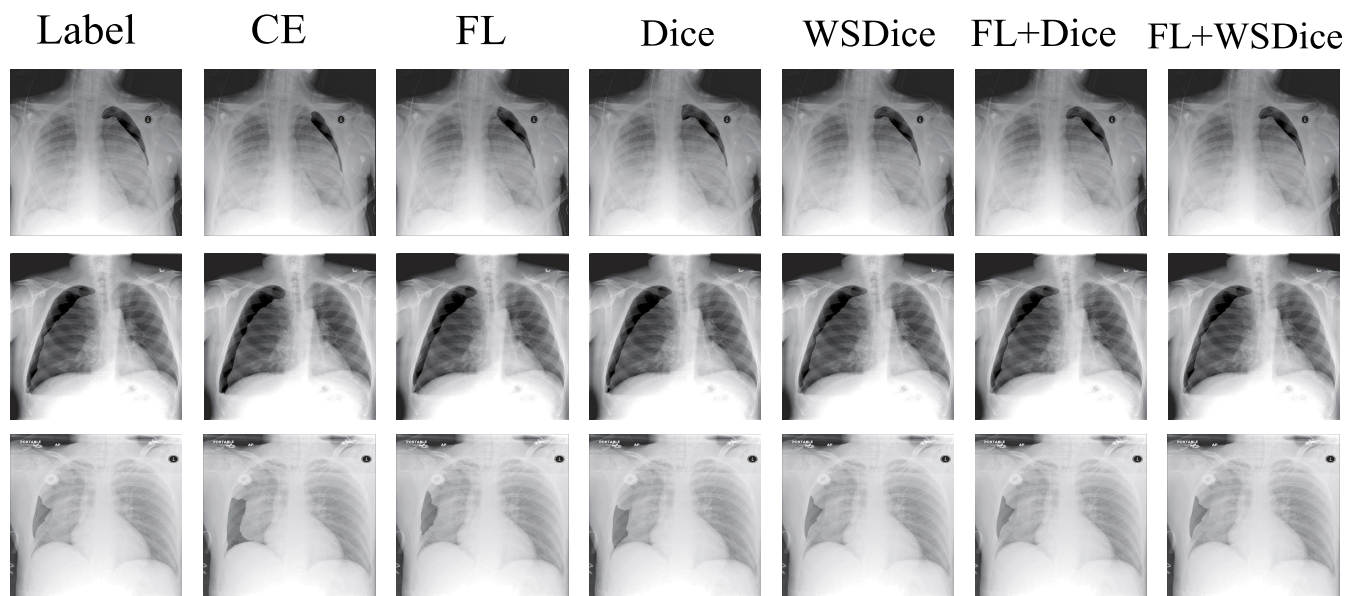
FIGURE 8. The segmentation results of different loss functions. The mask is overlaid on the image, so the area with deep color is the lesion area.

TABLE 3. The results of different loss function with the improved U-net in the JSRT dataset, where $\alpha = 0.26$, $\gamma = 2.3$ in Focal Loss, $v_1 = 0.1$, $v_2 = 0.9$ in WSDice Loss, $\beta = 10$, $\alpha = 0.26$, $\gamma = 2.3$ in $\beta FL - log(Dice)$, and $\beta = 8$, $\alpha = 0.26$, $\gamma = 2.3$, $v_1 = 0.1$, $v_2 = 0.9$ in $\beta FL - log(WSDice)$.

Loss Function	Heart	Left Clavicle	Left Lung	Right Clavicle	Right Lung	All
Cross Entropy	0.9370	0.8782	0.9588	0.8713	0.9731	0.9237
Focal Loss	0.9433	0.9376	0.9685	0.9364	0.9701	0.9512
Dice Loss	0.9746	0.9258	0.9731	0.9277	0.9733	0.9549
WSDice Loss	0.9750	0.9439	0.9729	0.9344	0.9706	0.9606
$\beta FL - log(Dice)$	0.9825	0.9721	0.9782	0.9739	0.9821	0.9767
$\beta FL - log(WSDice)$	**0.9874**	**0.9769**	**0.9834**	**0.9825**	**0.9878**	**0.9836**

experiments are performed on ChestX-ray8 dataset. The comparison result of the performance reported in those paper with the performance of our method is shown in Table 2. The score of our method in the table is obtained by using the combination of WSDice Loss and Focal Loss and the improved U-net structure.

In order to prove the generalization of the method proposed in this paper, we use the JSRT dataset, which contains 247 chest X-ray images and 5 kinds of segmentation object masks, to conduct experiments. We divide the training set, validation set and test set according to the ratio of 6 : 2 : 2. In our experiment, the epoch is 50, the initial learning rate is 0.005, and the learning rate schedule is gradual warmup, the warmup time is 1 epoch, the cosine learning rate schedule was used after warmup. Similar to the previous experiments, we use the default parameters in the RAdam optimizer, the backbone was pre-trained in ImageNet and we freeze the first three blocks. The experimental results are recorded in the Table 3.

It can be seen that the area of the clavicle is small, the unbalanced data distribution is more serious than other targets, so Focal loss improves the dice coefficient more than the cross-entropy loss function. The segmentation target

in the JSRT dataset is larger than the pneumothorax in the ChestX-ray8 dataset, the data distribution of JSRT is more balanced compared to the ChestX-ray8, so the Dice loss obtains a better performance than Focal loss. The area of the segmentation target in JSRT dataset is large, the edge is clear, so it is easy to segment. Dice loss get a good score and the average dice of the five categories can reach 0.9549, but the WSDice loss, proposed in this paper, still improves the average dice to 0.9606. This means that WSDice loss mines some background information and our method is effective. In addition, the cascaded loss function can still improve the segmentation score, and the performance of Focal loss and WSDice loss cascade is better than the performance of Focal loss and Dice loss cascade, which can further prove that the cascade loss function can make the model optimize the model from different direction to improve the performance. It also further illustrates the effectiveness of WSDice loss.

V. CONCLUSION

In this paper, we improve the dice loss in the form of the weighted soft dice loss to solve the problem of the imbalance between the positive and negative samples in the medical image segmentation and the disadvantage that the dice loss

only focuses on positive samples and ignores negative samples. The loss function uses the labels to generate weights, and the negative samples of the image are included in the calculation of the dice loss according to this weight. Owing to this process, the model can fully exploit the features of the negative sample regions and retain the characteristics of the dice loss of not being affected by the distribution imbalance of the negative and positive samples.

Besides, we use the pneumothorax segmentation dataset ChestX-ray8 to perform the experiments, build a U-net model with the SE-ResNeXt-50 as the backbone network, improve the U-net, make better use of the features of different scales of the U-net decoder module, and use a variety of different loss functions for the training and comparison. The experiments show that the proposed method can achieve better performance.

The focal loss and weighted soft dice loss are combined as the loss function of the model. The method proposed in this paper is simple and effective. Model training is performed on the dataset, and the hyperparameters of the focal loss are set as the best value of the experimental results. Various hyperparameters settings are tested for the loss function. The experimental results show that the network segmentation effect is the best when $v_1 = 0.15$ and $v_2 = 0.85$.

However, the proposed approach has some disadvantages. First, the use of the weighted soft dice loss introduces hyperparameters v_1 and v_2, which need to be carefully adjusted during use. Compared to the dice loss, the weighted soft dice loss is more inconvenient to implement. For this shortcoming, we can use Neural Architecture Search (NAS) [33] to search for the optimal parameters automatically. Second, our method does not solve the problem of the dice loss error being sensitive to the calculation of the loss function when it is used to segment small objects. This problem may cause fluctuations in the training process and affect the performance of the model. Finally, the improved dice loss has some defects when segmenting small objects, and the effect is not as satisfactory as that of the focal loss. For the last two weaks, we can use cascade loss function to solve them.

REFERENCES

[1] G. P. Currie, R. Alluri, G. L. Christie, and J. S. Legge, "Pneumothorax: An update," *Postgraduate Med. J.*, vol. 83, no. 981, pp. 461–465, Jul. 2007.

[2] Y. Wang, L. L. Sun, and Q. Jin, "Enhanced diagnosis of pneumothorax with an improved real-time augmentation for imbalanced chest X-Rays data based on DCNN," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Apr. 19, 2019, doi: 10.1109/TCBB.2019.2911947.

[3] O. Geva, G. Zimmerman-Moreno, S. Lieberman, E. Konen, and H. Greenspan, "Pneumothorax detection in chest radiographs using local and global texture signatures," *Proc. SPIE*, vol. 9414, Dec. 2015, Art. no. 94141P.

[4] Y.-H. Chan, Y.-Z. Zeng, H.-C. Wu, M.-C. Wu, and H.-M. Sun, "Effective pneumothorax detection for chest X-ray images using local binary pattern and support vector machine," *J. Healthcare Eng.*, vol. 2018, pp. 1–11, Apr. 2018.

[5] K. Liang, L. Zhang, H. Yang, Y. Yang, Z. Chen, and Y. Xing, "Metal artifact reduction for practical dental computed tomography by improving interpolation–based reconstruction with deep learning," *Med. Phys.*, vol. 46, no. 12, pp. 823–834, Dec. 2019.

[6] Y. Han and J. C. Ye, "One network to solve all ROIs: Deep learning CT for any ROI using differentiated backprojection," *Med. Phys.*, vol. 46, no. 12, pp. 855–872, Dec. 2019.

[7] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2097–2106.

[8] T. Joon Jun, D. Kim, and D. Kim, "Automated diagnosis of pneumothorax using an ensemble of convolutional neural networks with multi-sized chest radiography images," 2018, *arXiv:1804.06821*. [Online]. Available: http://arxiv.org/abs/1804.06821

[9] A. G. Taylor, C. Mielke, and J. Mongan, "Automated detection of moderate and large pneumothorax on frontal chest X-rays using deep convolutional neural networks: A retrospective study," *PLOS Med.*, vol. 15, no. 11, Nov. 2018, Art. no. e1002697.

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

[11] 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.

[12] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

[13] Z. Zhou, M. M. R. 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*. Granada, Spain: Springer, 2018, pp. 3–11.

[14] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[15] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Québec City, QC, Canada: Springer, 2017, pp. 240–248.

[16] C. Shen, H. R. Roth, H. Oda, M. Oda, Y. Hayashi, K. Misawa, and K. Mori, "On the influence of dice loss function in multi-class organ segmentation of abdominal CT using 3D fully convolutional networks," 2018, *arXiv:1801.05912*. [Online]. Available: http://arxiv.org/abs/1801.05912

[17] Y.-C. Liu, D. S. Tan, J.-C. Chen, W.-H. Cheng, and K.-L. Hua, "Segmenting hepatic lesions using residual attention U-Net with an adaptive weighted dice loss," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 3322–3326.

[18] 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.

[19] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[20] S. M. Pizer, R. E. Johnston, J. P. Ericksen, B. C. Yankaskas, and K. E. Müller, "Contrast-limited adaptive histogram equalization: Speed and effectiveness," in *Proc. 1st Conf. Visualizat. Biomed. Comput.*, 1990, pp. 337–345.

[21] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 1–4.

[22] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1492–1500.

[23] 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.

[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[25] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, "On the variance of the adaptive learning rate and beyond," 2019, *arXiv:1908.03265*. [Online]. Available: http://arxiv.org/abs/1908.03265

[26] W. Zhu, Y. Huang, L. Zeng, X. Chen, Y. Liu, Z. Qian, N. Du, W. Fan, and X. Xie, "AnatomyNet: Deep learning for fast and fully automated whole–volume segmentation of head and neck anatomy," *Med. Phys.*, vol. 46, no. 2, pp. 576–589, Feb. 2019.

[27] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," 2017, *arXiv:1710.10501*. [Online]. Available: https://arxiv.org/abs/1710.10501

[28] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," 2017, *arXiv:1711.05225*. [Online]. Available: https://arxiv.org/abs/1711.05225

[29] X. Wang, Y. Peng, L. Lu, Z. Lu, and R. M. Summers, "TieNet: Text-image embedding network for common thorax disease classification and reporting in chest X-Rays," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9049–9058.

[30] Z. Li, C. Wang, M. Han, Y. Xue, W. Wei, L.-J. Li, and L. Fei-Fei, "Thoracic disease identification and localization with limited supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8290–8299.

[31] V. Mahajan, V. K. Venugopal, M. Murugavel, and H. Mahajan, "The algorithmic audit: Working with vendors to validate radiology-AI algorithms—How we do it," *Academic Radiol.*, vol. 27, no. 1, pp. 132–135, Jan. 2020.

[32] Y.-X. Tang, Y.-B. Tang, Y. Peng, K. Yan, M. Bagheri, B. A. Redd, C. J. Brandon, Z. Lu, M. Han, J. Xiao, and R. M. Summers, "Automated abnormality classification of chest radiographs using deep convolutional neural networks," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–8, May 2020.

[33] B. Zoph and Q. V. Le, "Neural architecture search with reinforcement learning," in *Proc. 5th Int. Conf. Learn. Represent. (ICLR)*, Toulon, France: OpenReview.net, Apr. 2017. [Online]. Available: https://openreview.net/forum?id=r1Ue8Hcxg

CHAOLI WANG received the B.S. and M.Sc. degrees from the Mathematics Department, Lanzhou University, Lanzhou, China, in 1986 and 1992, respectively, and the Ph.D. degree in control theory and engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 1999. From 1999 to 2000, he was a Postdoctoral Research Fellow with the Robotics Laboratory, Chinese Academy of Sciences, Shenyang, China. From 2001 to 2002, he was a Research Associate with the Department of Automation and Computer-Aided Engineering, The Chinese University of Hong Kong, Hong Kong. Since 2003, he has been with the Department of Electrical Engineering, University of Shanghai for Science and Technology, Shanghai, China, where he is currently a Professor with the School of Optical-Electrical and Computer Engineering. His current research interests include nonlinear control, robust control, robot dynamic and control, visual servoing feedback control, and pattern identification.

ZHANQUAN SUN received the B.S. degree in mechanical design and theory from Shenyang Jianzhu University, Shenyang, China, in 2000, the M.Sc. degree in mechanical design and theory from the Dalian University of Technology, Dalian, China, in 2003, and the Ph.D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2007. From 2007 to 2018, he was a Researcher with the Shandong Computer Science Center, Jinan, China. Since 2018, he has been with the Department of Control Engineering, University of Shanghai for Science and Technology, Shanghai, China, where he is currently an Associate Professor with the School of Optical-Electrical and Computer Engineering. His current research interests include artificial intelligent, big data, and machine learning.

SHENG CHEN (Member, IEEE) received the B.S. and M.S. degrees in computer science and technology from Xidian University, Shanxi, China, in 1999 and 2002, respectively, and the Ph.D. degree in pattern recognition and intelligent system from Shanghai Jiao Tong University, Shanghai, China, in 2006. From 2008 to 2011, he was a Postdoctoral Research Fellow with The University of Chicago, USA. He has been an Associate Professor with the School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, since 2011. His current research interests include pattern recognition and intelligent systems, medical image processing and analysis, and computer-aided diagnosis research.

LU WANG received the B.S. degree from Pingxiang University, Jiangxi, China, in 2018. She is currently pursuing the M.Sc. degree in operations research and cybernetics with the University of Shanghai for Science and Technology, Shanghai, China. Her current research interests include computer vision and medicine image segmentation.

• • •