Segmentation of Pulmonary Nodules Based on MRBU-Net-WD Model

Zhuo Li, Xiaoxia Zhang, Bo Zhang

Abstract—It is important to diagnose lung nodules as early as possible in order to improve the cure rate of lung cancer patients. The CT technology is currently the most prevalent method of diagnosis and treatment in clinical medicine. However, the amount of data generated by CT diagnosis and treatment is increasing exponentially. In order to segment lung nodules from medical images, it is crucial to employ computer and artificial intelligence technology. The deep learning model proposed in this paper is MRBU-Net-WD model, a more effective and enhanced version of the U-Net. MRBU-Net-WD is distinguished by combining residual 3D convolution modules with multiscale densely connected modules. Moreover, In addition, in order to avoid the phenomenon of gradient disappearance when the network depth of the model is increasing. The Bi-FPN is therefore introduced to enhance the feature maps of the network at each depth, as well as to provide an effective fusion of features across depths. The weighted Dice loss function has a significant improvement for the pixel imbalance between lung nodules and background images in lung CT images. To test the proposed model, the LUNA-16 dataset has been extensively trained and evaluated. The performance of the comparison methods generally outperforms that of the original U-Net model and several other newly proposed models.

Index Terms—Lung nodule segmentation, Convolutional neural network, Residual 3D Convolution, Deep learning, Computer-aided diagnosis

I. INTRODUCTION

According to statistics, lung cancer has always been the most prevalent form of cancer worldwide. It typically manifests in the middle and later stages, resulting in a high mortality rate. Therefore, the earlier lung cancer is detected, the better the treatment for the patient. The pathology of early stage lung cancer is characterized by small size of nodules in the lung and have blurred edges, making them easy to misdiagnose and miss. Early and accurate detection and analysis of pulmonary nodules can result in timely and effective treatment, thereby significantly increasing patients' survival rates [1]. Although the advent of Computed Tomography (CT) technology can aid in the early detection of diseased nodules, the rising demand for

imaging diagnosis necessitates that physicians carefully consider and screen each patient individually, which is inefficient and can cause doctor fatigue. This increases the likelihood of an incorrect diagnosis. Computer-aided diagnosis (CAD) has been the main diagnostic modality for physicians in recent years, achieving rapid and accurate acquisition of patient's lesion information from a large number of CT images and reducing the possibility of missed and misdiagnosis. With the aid of computer image processing technology, physicians are able to precisely position and segment pulmonary nodules, analyze tissue components and identify pathological features within the lesion region of the nodule [2].

Traditional segmentation algorithms include threshold segmentation [3], edge detection segmentation [4], and region growth [5]. Kostis et al. [6] used a morphological approach in the problem of removing nodules attached to blood vessels and to separate the pulmonary nodules by identifying the attachment area. Sargent et al. [7] introduced morphological manipulations that combine shape hypotheses in place of morphological templates with a fixed size. According to Dehmeshki et al. [8], a region growth method that integrates fuzzy connectivity, distance, and intensity as growth mechanisms was proposed and applied to the segmentation of lung nodules. Based on Mumford-Shah partitioning techniques and level setting methods, Chan et al. [9] proposed an active contour model. It provides better detection and retention of boundary location information even when the initial image does not appear smooth, automatically detects internal contours beginning with a single initial curve and does not require the initial curve's position in an image to be fixed. The SBGF-new SPF algorithm presented by Nithila et al. [10] is capable of segmenting vascular adhesions and pleural adhesions from CT images, and it has a high accuracy and low error rate for segmenting lung nodules from CT images. A method was proposed by Ye et al. [11] that was able to successfully separate lesions with similar intensities but different shapes and correctly identify lesions with similar intensities but different shapes. According to Wu et al. [12], a hierarchical statistical learning method may be used to determine whether candidate nodules detected on CT scans are associated with pulmonary vessels, fissures, and lung walls. Using image features, Lu et al. [13] proposed a flexible hierarchical feature learning framework for the detection of all types of nodules and polyps. These methods are only applicable to simple scenarios. The performance and generalization capabilities are subpar and cannot be effectively transferred to new task scenarios. Later on, deep learning methods were improved to perfectly solve the limitations of machine learning in some problems.

Many medical image segmentation fields have used deep learning extensively and effectively in recent years. In

order to effectively extract, classify, and recognize images, deep neural networks employ a multilayer nonlinear mapping structure. Deep neural networks are typically modeled using CNNs. The convolutional layer, pooling layer and fully connected layer form a simple CNN model. Each layer performs a specific function. Kumar et al. [14] were the first to use convolutional neural networks (CNN) for the problem of extraction of lung nodules and also achieved classification and recognition to some extent. As a result of the CNN network model being used to classify an image, the final fully-connected layer compresses the 2D matrix information in the original image, resulting in the loss of image information and influencing the segmentation function. Shelhamer et al. [15] proposed a novel network structure (FCN) for the semantic level image segmentation problem, which differs most from the CNN in that it uses a convolutional layer instead of the original fully-connected layer. This change allows the network to make more accurate, dense predictions for images of any size. The high-level feature information obtained after convolution is fused with the upsa mpled output and passed to the convolution layer in order to produce a more precise output. With the advent of deep neural networks, techniques for the automatic segmentation of lung nodules have been accelerated. The system has achieved excellent results in retrieval, recognition, and semantic segmentation.

Due to the continuous advancement of computer hardware and software, researchers have been able to sift through large volumes of data in order to develop intelligent models and methods for solving a variety of practical problems. Natural image processing has made significant progress in areas such as improving diagnostic efficiency and reducing the likelihood of misdiagnosis and missed diagnoses. This technique has become a standard in the direction of medical image segmentation in particular. As of now, the most popular network model used in medical image segmentation are convolutional neural networks and full convolutional neural networks. In a full convolutional neural networks, the U-Net model is the most widely used network model.

Although there are some improved models for lung nodule image segmentation in recent years, the segmentation accuracy needs further research and improvement. In our study, we proposed a hybrid network model for segmentation of pulmonary nodules. The model proposed in this paper differs from the previous models in the following ways: First, the primary idea of this paper is to embed the 3D convolutional module into the U-Net model. Second, To further improve the segmentation accuracy of the model, the multi-scale modules are densely connected using the multi-scale dense connection module, so it allows the model to obtain more feature information during upsampling and downsampling. Third, as the depth of the network deepens during the training process, the gradient of U-Net model disappears. Hence Bi-FPN is introduced to enrich the feature maps of the network at each depth, and it also provides effective fusion of features across different depths. Finally, the defect of uneven target region and background pixels in lung CT images is solved by using the weighted Dice loss function. The hybrid model with these changes is what we call the MRBU-Net-WD model. To validate the effectiveness of the improved model in lung nodule segmentation, experiments will be conducted on the LUNA-16 dataset, and the experimental results show that the MRBU-Net-WD model performs better than the original U-Net model and some recently proposed new models.
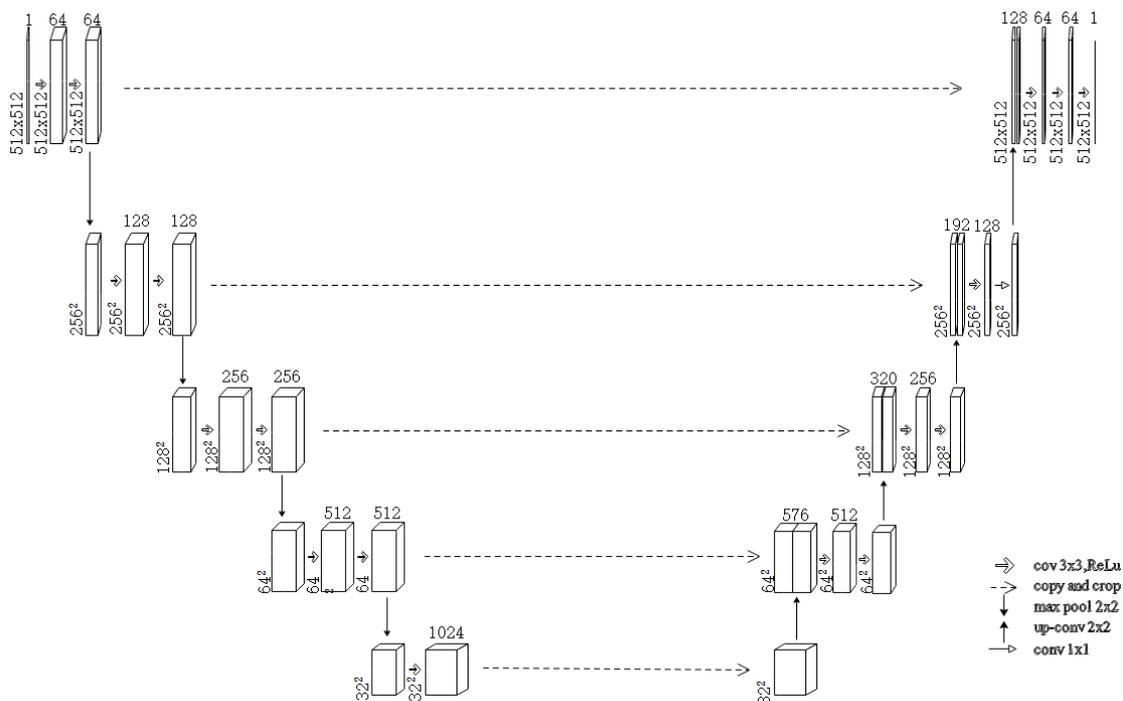
Fig. 1. Schematic diagram of U-NET network structure.

II. MATERIALS AND METHOD

As discussed in this section, there have been several strategies for dividing up lung nodules that have been proposed in recent years. These strategies also include the methods used in this improvement.

A. U-Net

The U-Net model was first proposed in 2015. Based on the end-to-end FCN [17], Ronneberger et al. [16] enhanced and proposed the U-Net model, which improved the expressiveness of features and the segmentation performance by using skip connections between encoding and decoding paths. A substantial amount of attention has also been paid to this method in the field of medical image segmentation. It is being improved by an increasing number of academics. The U-Net model has superior segmentation performance in the field of medical image segmentation compared to the FCN model.

The encoder-decoder structure is utilized by the U-Net model, and the overall structure is symmetrical. Upsampling and downsampling comprise the majority of the network's structure. The FCN network's convolutional and pooling layers are preserved in the U-Net model. In the connection layer portion, however, the full connection is converted to a skip connection, and feature fusion is achieved through a stacking operation. Consequently, the U-Net model is capable of recognizing object categories using both shallow and deep features. Therefore, the U-Net model can achieve more precise segmentation. Fig 1 is a schematic representation of U-network Net's architecture, which can be divided into three sections: left encoding, right decoding, and skip connection. The encoding path adheres to a typical convolutional network architecture with two repeated 3×3 convolutions, a ReLU activation function, a 2×2 maximum pool operation, and a span of 2 downsampling are connected. After the downsampling operation, the number of feature channels is increased. On the path of decoding, feature maps are upsampled. When the number of feature channels needs to be reduced, a 3×3 convolution operation is used to process the feature maps,

there will be a ReLU activation function in the last convolutional layer of the coding path, the final 1×1 convolution operation maps the feature vector to the output layer of the network. By means of the stacking operation, the skip connection part integrates the encoding and decoding paths on the left and right respectively, combining the low-level information derived from the encoding path with the high-level information derived from the decoding path. Due to this, the fused feature map has both shallow and deep semantic feature information, which allows the segmentation effect to be more precise.

B. Bi-FPN

Bi-FPNs (Feature Pyramid Networks) are complex bidirectional FPNs used by Efficient-Det. When building a top-down, bottom-up bidirectional channel, unify feature resolution scales by upsampling and downsampling information from different scales of the backbone network when merging information from different scales. By adding horizontal features between features connected at the same scale, to prevent excessive feature loss during training. During feature fusion, Bi-FPN also adds respective weights to each input feature separately, so that the network is able to determine the importance of each feature. A fast normalized fusion is used to provide dynamically learned behavior and accuracy. Bi-FPN enhances the feature maps at different depths in the network using bidirectional cross-scale connectivity and also performs effective feature fusion for features at different depths. The Bi-FPN network architecture is shown in Fig 2.

C. Residual 3D Convolution

With the continuous advancement of medical imaging, the data generated in clinical medical diagnosis no longer consists solely of 2D medical image data but rather a greater proportion of 3D medical image data. In order to raise the precision of segmentation, it is important to acquire the full feature data included in the 3D images. An effective method for improving the manifestation of the deep network model is to increase the network depth in
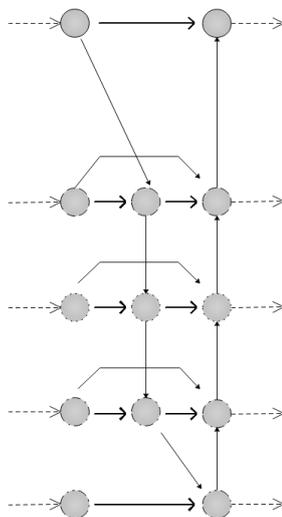
Fig. 2. Bi-FPN network structure diagram.
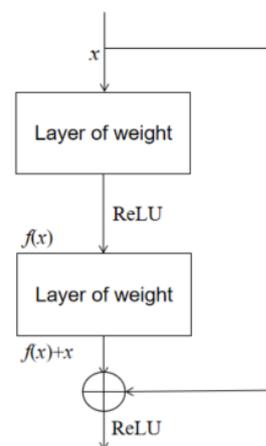
Fig. 3. Residual network diagram.

order to obtain the 3D spatial structure information in the image and to resolve the gradient disappearance problem during the training process. In theory, As the number of network levels grows, the network capabilities will improve, but in practice, this is not the case. The model's accuracy will decrease after a saturation point is reached as the number of network model layers is increased. This is because of the extent of Gradient disappearance is a problem that arises during network training. A series of residual blocks comprise the residual network. The direct mapping section and the residual section make up the two parts of the residual block, as shown by the following formula:

$$x_{l+1} = F(x_l + f(x_l + \omega_l)) \tag{1}$$

Among them, x_l and x_{l+1} denote the input and output data of the l layer, $f(\cdot)$ denotes the residual mapping function; w_l represents the convolution algorithm, and $F(\cdot)$ represents the ReLU function.

Fig 3 is a diagrammatic representation of the residual network. The result after direct mapping and the result after residual learning are added to the residual portion, which consists of convolution operations (represented in the figure as two convolution operations). Connect a ReLU function to the operation to obtain the final output result.

D. Multiscale Densely Connected Module

Reference [18] proposes treating each output of a convolution operation as a densely connected form of all inputs to enhance the efficacy and depth of feature extraction. In light of this, equations 2 through 5 show the multi-scale feature fusion encoding calculation process that is proposed in this paper. Equations 6 through 9 show the decoding calculation process.

$$X_d^{i-1} = D(X^{i-1}) \tag{2}$$

$$X_b^{i-1} = H(X_n^{(i-n)}, \Lambda\Lambda, X_n^{(i-1)}) \tag{3}$$

$$X_d^{(i-1)} = H(X_b^{i-1}, X_d^{i-1}) \tag{4}$$

$$X^i = F(X_d^{i-1}) \tag{5}$$

Among them, X^{i-1} denotes the input of the present level; X^i denotes the output of the current level; X_d^{i-1} denotes the output of X^{i-1} after downsampling, $D(\cdot)$ denotes downsampling, X_n^{i-n} denotes the adjustment of X_n^{i-n} to the same size as the X^{i-1} layer, X_b^{i-1} indicates the characteristic mask of the upper layer i-1, $X_d^{(i-1)}$ indicates the fusion of X_b^{i-1} and X_d^{i-1} two feature maps, H($\cdot$) denotes the concatenation operation and 1×1×1 convolution, $F(\cdot)$ denotes the activation function.

$$Y_q^{i-1} = O(Y^{i-1}) \tag{6}$$

$$X^{(i-1)} = H(X_n^{(i-n)}, \Lambda\Lambda X^i, , \Lambda\Lambda, X_n^{(i+n)}) \tag{7}$$

$$Y_{nc}^{i-1} = H(X^{i-1}, Y_q^{i-1}) \tag{8}$$

$$Y^i = F(Y_{nc}^{i-1}) \tag{9}$$

Among them, Y^{i-1} denotes the input of the current layer, Y^i denotes the output of the current stratum, Y_q^{i-1} denotes the output of Y^{i-1} after downsampling, $U(\cdot)$ denotes the upsampling, and X^{i-1} denotes the low-level containing two sets of feature information: i-1 layer to i-n layers information, the high-level feature information from layers i+1 to i+n; Y_{nc}^{i-1} denotes the fusion of two parts of X^{i-1} and Y_q^{i-1} feature maps, $H(\cdot)$ denotes the concatenation operation and 1×1×1 convolution, and $F(\cdot)$ represents the activation function.

E. Weighted Dice loss function

In medical image segmentation, Dice loss function is a well-known loss function. In CT images, the area of the lung nodule region to be segmented is much smaller than the area of the background, so there is a problem of pixel imbalance between the lung nodules and the background in lung nodule segmentation. However, the original Dice loss function is not an effective solution for this kind of problem. Therefore, an optimized weighted Dice loss function is presented in this paper, as shown in Equation 10.

$$L = -\frac{2}{i} \sum_{i=0}^{i-1} \frac{\sum_x u^k(x) \cdot v_w^k(x)}{\sum_x u^k(x) + \sum_x v_w^k(x)} \tag{10}$$

In Equation 10, u and v represent the prediction map outputted using the softmax classifier and the real picture encoded using one-hot. k represents the number of categories, $k \in \{0,1,...,K\text{-}1\}$. $u^k(x)$ represents the odds that pixel x is in class k in the output probability graph. $v_w^k(x)$ represents the weighted true picture, and the weighting is shown in Equation 11.

$$v_w^k(x) = w(x) \cdot v^k(x) \tag{11}$$

In Equation 11, $v^k(x)$ is the odds that pixel x is in class the k class in the real graph, $w(x)$ is the weight, and the weights are calculated as shown in Equation 12. where $w_{class}(x)$ is the class weight, according to the class to which pixel x belongs, it is given different weights, the class with more pixels has a relatively small weight, while the class with fewer pixels has a relatively large weight; $w_{area}(x)$ is the connected domain area weight, which is calculated as shown in Equation 13, where $volume(x)$ is the area of the linked field and c is a constant, from Equation The larger the area of a connected domain, the smaller the connected domain area weight $w_{area}(x)$, and vice versa; $w_{border}(x)$ is the border pixel weight, which gives a larger weight to the border pixels to raise the segmentation result of the border, which is gained by minus the corrupted image from the expanded image. is the balance factor, which measures the degree of influence of the boundary pixel weight on the overall weight.
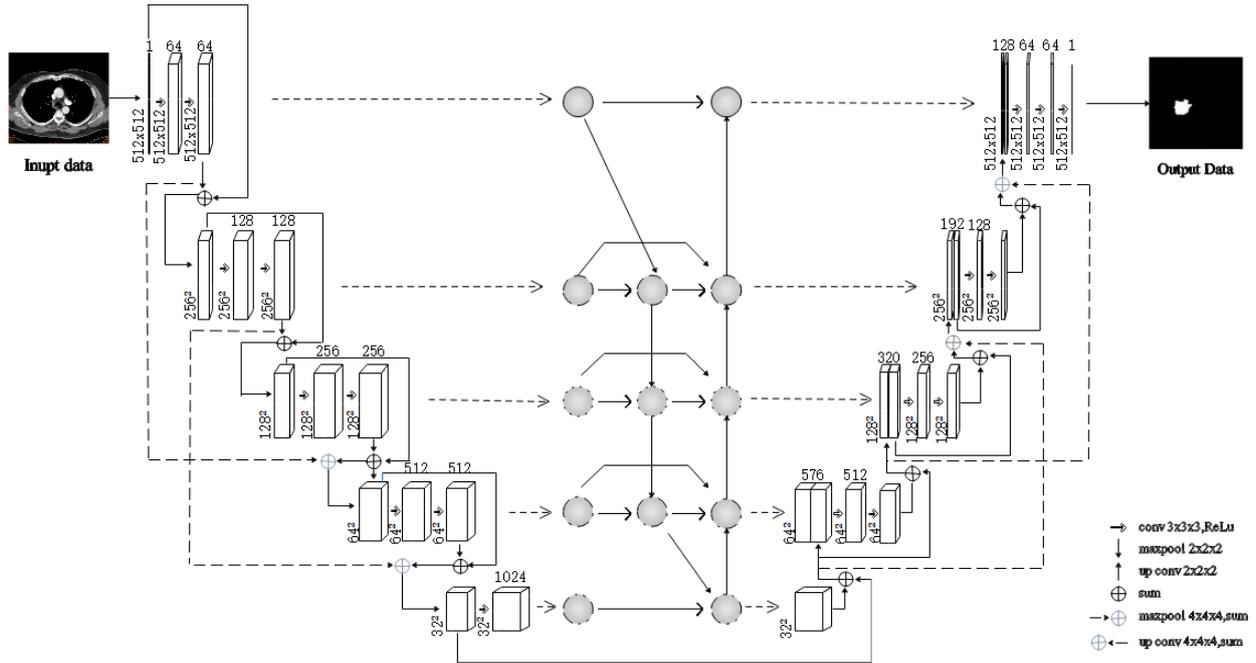
Fig. 4. Schematic diagram of MRBU-Net-WD model structure.

$$w(x) = w_{class(x)} \cdot w_{area(x)} + \gamma \cdot w_{border(x)} \quad (12)$$

$$w_{area(x)} = \exp\left(-\frac{volume(x)}{c^2}\right) \quad (13)$$

III. THE PROPOSED ALGORITHM

This paper proposes this model as a deep learning technique for segmenting lung nodules. This model is primarily an improvement to the U-Net model and retains the symmetrical properties of the U-Net network model. Fig. 4 depicts the MRBU-Net-WD architecture proposed.

The number of tiers and each network dimension of the MRBU-Net-WD model are shown in Table I. The encoding path is on the left. The decoding path is on the right. We introduce residual 3D convolution (R3D-CONV) and multi-scale dense connections, combine the U-Net model and residual network, and use Bi-FPN between the encoder and decoder.

As input, the MRBU-Net-WD model takes 512×512 images as input. The model's input data is a 3D-CT image, and the image's low-level features are extracted via the encoding route and then entered via the decoding route. In the decoding path, the image's high-level semantic features are extracted via an upsampling operation. The entire network's encoding and decoding paths consist of residual 3D convolution modules and multi-scale dense connection modules. The encoding and decoding paths are linked by a bidirectional feature fusion, and the segmentation results come from a 1×1×1 convolutional level.

In the case of input image, the MRBU-Net-WD encoder network consists of two 3×3×3 volumes, each followed by an activation feature and a 2×2×2 max-pooling action with stride 2. Features are reduced in size. The downsampling stage is composed of two 3D convolutional layers and activation functions, followed by a pooling layer for downsampling with a 2×2×2 max-pooling operation and stride 2. The feature sizes corresponding to the five encoding depths are 512×512×64, 256×256×128, 128×128×256, 64×64×512, 32×32×1024, of which 64, 128, 256, 512, and 1024 represent the number of channels. The

TABLE I
MRBU-NET-WD MODEL NETWORK PARAMETERS.

	Layer name	Filter dimension	Number of parameters
Contraction path	Conv×10, ReLu	3×3×3	$2.826×10^7$
	MaxPool×4	2×2×2	
Bi-FPN	Conv×5	3×3×3	$1.269×10^5$
	BatchNormalization×12		3072
	ReLU×12, MaxPool×3	2×2×2	
	DepthwiseConv×7		4032
Expansion path	Conv×9, ReLu	3×3×3	$1.023×10^7$
	ConvTrans×4, ReLu	2×2×2	$2.786×10^6$
Total parameters			$4.141×10^7$

encoder and decoder are symmetrical, and the process is performed in reverse. Equation 14-15 depicts the convolution operations that occurs in each stage of the network:

$$C[m,n] = (I \times k)[m,n] = \sum_i \sum_j k[i,j], I[m-i, n-j] \quad (14)$$

$$Z^{[l]} = W^{[l]}.A^{[l-1]} + b^{[l]} \quad (15)$$

$$A^{[l]} = f^{[l]}(Z^{[l]}) \quad (16)$$

denotes kernel convolution, while Equations 15 and 16 correspond to forward propagation. In Equation 14, I and k represent, respectively, the input picture and nucleus. In Equations 15 and 16, $A^{[l]}$, $w^{[l]}$, $b^{[l]}$, and $f^{[l]}$ represent the activation, weight, offset, and optimization feature of the lth layer, respectively.

In the residual stage, every convolution the input image is convolved by two identical 3×3×3 convolution operations; the ReLU optimization feature is connected after optimization feature; and the input feature map is obtained after each convolution. Using a 2×2×2 max-pooling action with a step size of 2, the input feature maps are added together to make the output feature map.

During the stage of multi-scale feature fusion, the encoding and decoding paths contain identical convolution operations, and each convolution module contains two identical 3×3×3 convolution kernels. The convolution operation is followed by the lot regularization operation and the ReLU activation function. In the encoding path, the output result from the first stage is downsampled by a factor of 4, and the output result from the second stage is downsampled by a factor of 2, before being combined as the input for the subsequent stage. The decoding path reverses the downsampling operation, while the fusion method remains the same.

The encoder features are input to Bi-FPN, and the decoder receives the output characteristics vector. Each stage of the decoder deconvolutes the feature map, which consists of a deconvolution operation with a stride of 2, and the corresponding feature vector from the feature network. The concatenation operation is followed by two 3×3×3 convolutions, while each deconvolution operation is followed by two repeated 3D convolutions and an activation function. In the last layer of deconvolution, the feature map goes through two 3×3×3 convolution and activation functions and a 1×1×1 convolutional layer with as many output channels as there are label categories.

MRBU-NET-WD receives CT images as input and generates 3D spatial features. And due to the residual module, multi-scale dense connection module, and Bi-FPN, it prevents the network from degenerating during the training process and prevents over-or undersegmentation. enhance the model's segmentation accuracy.

IV. EXPERIMENT

A. Data

Lung Nodule Analysis 2016, or LUNA16, is a dataset for lung nodule detection that was released as part of the 2016 Lung Nodule Analysis competition. The LUNA16 dataset is derived from the LIDC-IDRI dataset [19], which contains 1018 cases and 888 CT scans in total. All CT images are provided by a number of reputable medical institutions and well-established imaging companies. Lesion labeling on each CT image was done by four seasoned people chest radiologists. The process will have two stages. In the first stage, four doctors will independently annotate all the images in detail, including nodules of different sizes. In the second phase, each physician will review the results of the rest of the physicians for ensure accuracy and comprehensiveness.

The LUNA16 dataset contains approximately 1186 nodule annotations, which also include various attribute files of nodules. Each data in the LUNA16 dataset is a 3D image, including a range of multi-directional thoracic sections, and due to different scanning instruments, scanning equipment parameters, and the patient's own
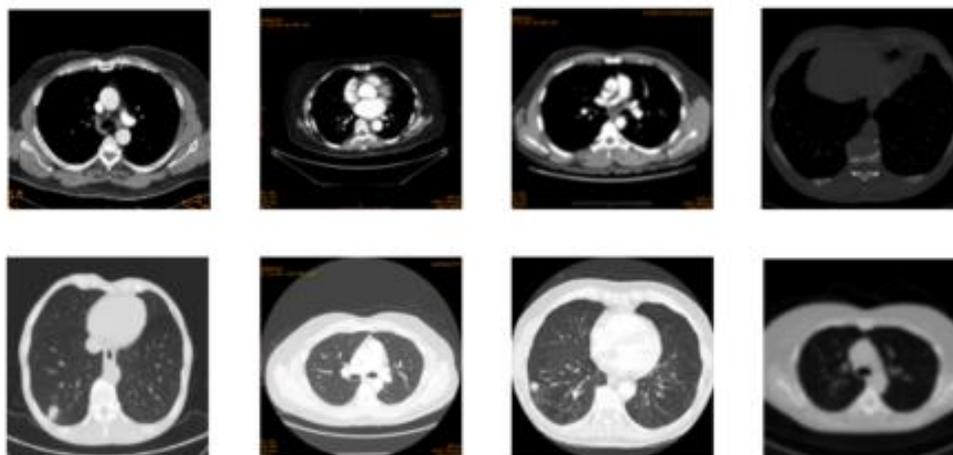
Fig. 5. LUNA16 data set example.

physiological conditions, the quantity of cuts in every picture will vary to some extent. Specifically, each 3D image in the dataset is composed of a different number of 2D images. Fig 5 shows the LUNA16 dataset.

B. Evaluation metrics

In order to test the segmentation effect of the MRBU-NET-WD model introduced in this paper, the Dice coefficient (DSC), Jaccard, and Sensitivity evaluation indicators are presented in this chapter (Sen). In the experiment, the segmented lung nodules can be classified as true/false positive or true/false negative pulmonary nodules. True Positives (TP) indicate that the model predicts a lung nodule that is in fact a lung nodule, whereas True Negatives (TN) indicate that the model predicts a non-pulmonary nodule that is in fact a lung nodule. Non-pulmonary nodules; false positive (FP), which means the model predicted a lung nodule when it was actually a non-pulmonary nodule; and false negative (FN), which means the model predicted a lung nodule when it was actually a non-pulmonary nodule.

The accuracy rate is used to evaluate the model's overall accuracy, allowing for a more intuitive analysis of the model's advantages and disadvantages. Accuracy is defined by equation 17:

$$Dice = \frac{2TP}{2TP + FP + FN} \tag{17}$$

The Jaccard similarity coefficient is often used to distinguish similarities and differences in data sets. Higher coefficients represent higher similarity. Jaccard is defined by equation 18:

$$Jaccard = \frac{TP}{TP + FN} \tag{18}$$

Sen is the proportion of correctly segmented tumor points in an image relative to the number of actual tumor points. The better the segmentation result, the higher the sensitivity. Sen is defined by equation 19:

$$Sen = \frac{TP}{TP + FN} \tag{19}$$

C. Experimental environment settings

In the experiments, a MRBU-Net-WD network is used to segment lung nodules in the LUNA16 dataset to prove the validity of this network for lung nodule segmentation. Moreover, to prevent the model from becoming overfit, an early stop training strategy is employed; if the model's performance does not improve after an additional ten training epochs, model training is terminated. In this experiment, the Tensorflow deep learning framework was adopted, Python 3.6 was used as the programming language, the processor was Intel Core i5-12600f, and the graphics card was NVIDIA Ge Force GTX 3060. The data sets used in this chapter are separated into two sections: training and testing. Repeat the training and test ten times in a cycle and average the segmentation index. In purpose of testing the

model, the batch amount is fixed to 16, the initial study speed is fixed to 0.001, and the optimization algorithm was Adam.

D. Experimental results

In the experiment, 704 images were selected at random as the training set, 96 images were used as the test set. The image size of both the training and test sets is 256×256. The transverse and longitudinal axes of Fig 6 indicate the number of iterations and the DSC in turn. During the training and validation phases, Fig. 6 demonstrates that when the MRBU-Net-WD network's epoch on the LUNA16 dataset reaches 40, the model converges on both the training and validation sets. When the epoch reaches 60, the training set and test set curves become stable, and training accuracy no longer increases.

Fig 7's abscissa and ordinate delegates the amount of iterations and the Loss value, respectively. The Loss curve of the training process of the MRBU-Net-WD network on the LUNA16 dataset is depicted in Fig 7. The MRBU-Net-WD network present in this study converges fundamentally when the training epoch of the training set and the validation set is approximately 40. When the epoch reaches 70, both the training set and the validation set have attained convergence, and the loss value is stable at 0.03. It
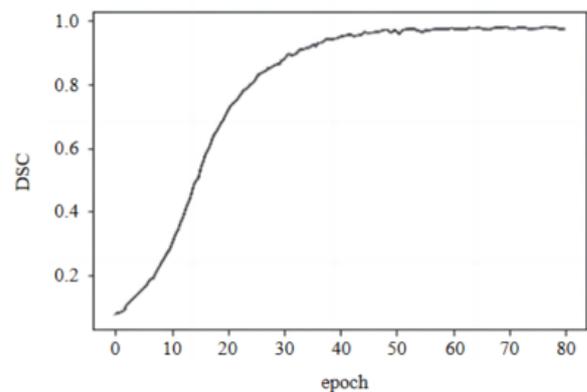
Fig. 6. MRBU-Net-WD network accuracy curve.
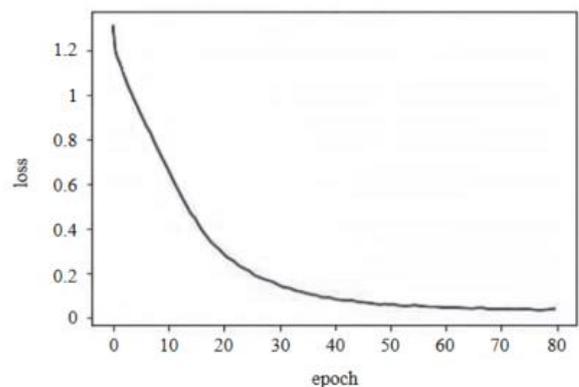
Fig. 7. MRBU-Net-WD Network training Loss curve.

demonstrates that the MRBU-Net-WD network can accommodate the distribution of samples in the training set and validation set, which are essentially the same.

TABLE II
COMPARISON OF SEGMENTATION EFFECTS OF EACH MODEL

	DSC	Sen	Jac
U-Net	78.82	78.04	65.94
U-NET3+	84.83	91.08	71.37
U-Det	88.54	95.03	74.28
RUNet	94.18	96.75	79.15
MRBU-Net-WD	**96.63**	**98.94**	**83.74**

As demonstrated in Table II, the MRBU-Net-WD network proposed in this paper denotes a great enhancement over the former U-Net model. The MRBU-Net-WD network has improved its DSC by 11.8%, 8.09% and 2.45% compared to some recently proposed models. The segmentation effect is much better than that of other models, which shows that the improved MRBU-Net-WD does a better job of generalization and doesn't overfit.

In the case of images containing negative samples, the acquisition of negative samples is conducted by intercepting them randomly from images without nodules. The acquisition is expanded to the same size as the positive samples. Introduce 5% of negative test samples. False-positive lung nodules with a radius of less than 2 pixels were masked out of the original 512×512 image to reduce the false-positive rate. In addition, it is assumed that false lung nodules will be identified in some non-pulmonary nodules in the positive sample, and the false positive rate is computed by taking the complete count of false lung nodules by the complete count of identified lung nodules. The sum of the precision rate and the rate of false positives is 1. The number of correct lung nodules out of the total number of lung nodules is the recall rate.

TABLE III
COMPARISON OF SEGMENTATION EFFECTS WITH 5% NEGATIVE SAMPLES

	RECALL	F1SCORE	JAC
U-Net	93.67	85.74	69.55
U-NET3+	92.96	91.64	75.36
U-Det	**97.83**	92.46	76.52
RUNet	94.35	88.24	83.61
MRBU-Net-WD	96.54	**93.91**	**85.67**

After adding 5% of negative samples, Table III shows that the improved model MRBU-Net-WD is only slightly less accurate than the U-Det model in terms of recall, but it

TABLE IV
COMPARISON OF SEGMENTATION EFFECTS WITH 10% NEGATIVE SAMPLES

	RECALL	F1SCORE	JAC
U-Net	93.71	84.06	66.53
U-NET3+	93.86	91.05	71.56
U-Det	**97.41**	90.14	74.04
RUNet	94.37	84.62	79.75
MRBU-Net-WD	96.54	**92.35**	**82.49**

is more accurate.

Based on Table IV, it is evident that the MRBU-Net-WD model outperforms the other models, based on the results of introducing 10% negative samples, similar to 5% negative samples. In terms of recall, the MRBU-Net-WD model is only slightly worse than the U-Det model, but it demonstrates certain advantages in all other performance indicators.

In order to verify the improvement of weighted Dice function, the performance of weighted cross entropy loss function, Focal loss function and weighted Dice loss function were compared and analyzed. Based on the data in Table V it is not difficult to analyze that among the three weighted loss functions, this change in the weighted Dice loss function has the best impact on the segmentation results, and its Dice coefficients on the segmentation of pulmonary nodules are 96.63% respectively. Thus, compared with the weighted cross entropy loss function and Focal loss function, the weighted Dice loss function has the best segmentation performance. It can not only solve the pixel imbalance problem of pulmonary nodules, but also increase the segmentation accuracy of pulmonary nodules. At the meantime, the weighted Dice loss function also has higher stability. In conclusion, compared with the weighted cross entropy loss function and Focal loss function, the weighted Dice loss function has higher accuracy and stability.

TABLE V
SEGMENTATION RESULTS FOR SEVERAL WEIGHTED LOSS FUNCTIONS (DSC:%)

	Cross Entropy Loss	Weighted Cross Entropy Loss	Focal Loss	Weighted Dice Loss
Mean	88.2	90.5	93.5	**96.6**
Std	7.0	3.7	5.3	4.3
Med	88.9	91.1	95.5	97.6
Min	69.0	85.5	85.5	91.1
Max	89.6	95.6	98.5	99.4

In this study, six groups of experiments were conceived to verify the suggested improved model against each other. From Table VI, it can be seen that the optimization effects of multi-scale fusion, Bi-FPN, Residual 3D Convolution (R3D-conv) and Weighted Dice loss function enhance the

Table VI
COMPARATIVE EXPERIMENTAL RESULTS

	Multi-scale feature fusion	Bi-FPN	R3D-conv	Weighted DIce loss function	DSC
U-Net	×	×	×	×	78.82
Model1	√	×	×	×	83.34
Model2	×	√	×	×	82.78
Model3	×		√		84.12
Model4	×	×	×	√	86.87
Model5	√	√	×	×	87.32
Model6	√	√	√	×	92.49
MRBU-Net-WD	√	√	√	√	**96.67**
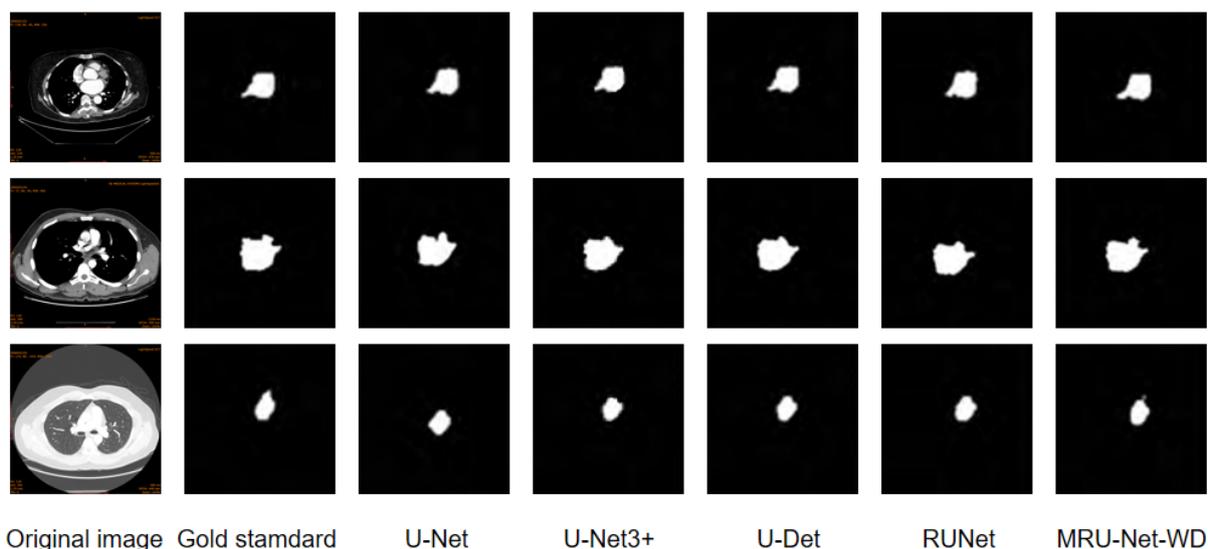
Fig. 8. Comparison of lung nodule segmentation results.

partitioning effect of the model very significantly. In terms of segmentation accuracy, the results of the multi-scale fusion, Bi-FPN, Residual 3D Convolution and Weighted Dice loss function can be superimposed and the final splitting precision is higher than that of the improved model alone.

Similarly, as can be seen from Figure 8, the algorithm in this study significantly outperforms other methods in segmenting fine nodules, nodules with unclear edges, and nodules with similar gray values in the image.

Using four sets of experiments, tables, graphs, and results of visual segmentation, this paper shows that the proposed MRBU-Net-WD network divides lung nodules more effectively than other models.

V. CONCLUSIONS

In this study, the architecture of U-Net model is improved and a new network structure. The MRBU-Net-WD model is presented by analyzing the features of medical pictures at the present stage and the current research status. Combining the residual 3D convolution module further enhances the U-Net network's segmentation performance. The multi-scale dense connection, which further improves the segmentation performance of the U-Net model, solves the problem of the gradient disappearing as the network depth increases during U-Net model training. On this basis, Bi-FPN is introduced to enrich the feature maps of each network depth and provide an effective fusion of features across depths. The defect of uneven target region and background pixels in lung CT images is solved by using the weighted Dice loss function. In the conclusion of this study, it is tested on the LUNA16 dataset. The experimental results demonstrate that the MRBU-Net-WD model segmentation effects are superior to those of some recently proposed models. In spite of this, the experiments presented in this paper do not overprocess the image data. It will be possible for researchers to process the data in more detail before training for the sake of improve the model's ability to segment lung nodules in the future.

Reference

[1] F. Bray, J. Ferlay, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol.70, pp. 313-313, 2018.

[2] K. Bhargavi, S. Jyothi, "A Survey on Threshold based Segmentation Technique in Image Processing," International Journal of Innovative Research and Development, vol.3, pp. 234-239, 2014.

[3] W.K. Zheng, K. Liu, "Research on Edge Detection Algorithm in Digital Image Processing," Materials Science, Machinery and Energy Engineering, pp. 1203-1208, 2013.

[4] N. Mesanovic, M. Grgic, H. Huseinagic, M. Males, E. Skejic, M. Smajlovic, "Automatic CT Image Segmentation of The Lungs with Region Growing Algorithm," Systems, Signals and Image Processing-IWSSIP, pp. 395-400, 2011.

[5] M.A. Taori, K.A. Chaudhari, S.S. Patankar, "Segmentation of Macula in Retinal Images Using Automated Seeding Region Growing Technique," Inventive Computation Technologies, pp. 1-5, 2017.

[6] J. Kostis, William, "Three-Dimensional Segmentation and Growth-Rate Estimation of Small Pulmonary Nodules in Helical CT Images," Medical Imaging, pp. 1259-1274, 2003.

[7] D. Sargent, P.Y. Sun, "Semi-automatic 3D Lung Nodule Segmentation in CT Using Dynamic Programming," Image Processing, vol. 10133, pp. 783-791, 2017.

[8] J. Dehmeshki, H. Amin, M. Valdivieso, "Segmentation of Pulmonary Nodules in Thoracic CT Scans: A Region Growing Approach," Medical Imaging, pp. 467-480, 2008.

[9] F.T. Chan, A.L. Vese, "Active Contours without Edges," Image Processing, pp.266-277, 2001.

[10] E. Ezhil, Nithila, "Segmentation of Lung Nodule in CT Data Using Active Contour Model and Fuzzy C-mean Clustering," Alexandria Engineering Journal, vol.3, pp. 2583-2588, 2016.

[11] X. Ye, G. Beddoe, G. Slabaugh, "Automatic Graph Cut Segmentation of Lesions in CT Using Mean Shift Superpixels," International Journal of Biomedical Imaging, vol. 547, pp. 870-886, 2012.

[12] D. Wu, L. Le, J. Bi, "Stratified Learning of Local Anatomical Context for Lung Nodules in CT Images," Computer Vision and Pattern Recognition, pp. 2791-2798 2010.

[13] L. Le, P. Devarakota, S. Vikal. "Computer Aided Diagnosis Using Multilevel Image Features on Large-Scale Evaluation," Medical Computer Vision, pp. 161-174, 2013.

[14] D. Kumar, A. Wong, A.D. Clausi, "Lung Nodule Classification Using Deep Features in CT Images," Computer and Robot Vision, pp, 133-138, 2015.

[15] E. Shelhamer, J. Long, T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," Pattern Analysis and Machine Intelligence, pp. 3431-3440, 2016.

[16] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Medical Image Computing and Computer-assisted Intervention, pp. 234-241, 2015.

[17] J. Long, E. Shelhamer, T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," Pattern Analysis and Machine Intelligence, pp. 640-651, 2015.

[18] G. Huang, Z. Liu, L. Van Der Maaten, K. Q. Weinberger, "Densely Connected Convolutional Networks," Computer Vision and Pattern Recognition, pp. 4700-4708, 2017.

[19] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, L. Pringle M, Tarbox, F. Prior. "Cancer Imaging Archives (TCIA): Maintaining and Operating a Public Information Repository Journal of Digital Imaging," Vol. 26, pp. 1045-1057, 2013.