

RESEARCH

Open Access



# HR-MPF: high-resolution representation network with multi-scale progressive fusion for pulmonary nodule segmentation and classification

Ling Zhu<sup>1</sup>, Hongqing Zhu<sup>1\*</sup> , Suyi Yang<sup>2</sup>, Pengyu Wang<sup>1</sup> and Yang Yu<sup>1</sup>

\*Correspondence:

hqzhu@ecust.edu.cn

<sup>1</sup> School of Information Science & Engineering, East China University of Science and Technology, Mei Long Road, 200237 Shanghai, China

Full list of author information is available at the end of the article

## Abstract

Accurate segmentation and classification of pulmonary nodules are of great significance to early detection and diagnosis of lung diseases, which can reduce the risk of developing lung cancer and improve patient survival rate. In this paper, we propose an effective network for pulmonary nodule segmentation and classification at one time based on adversarial training scheme. The segmentation network consists of a High-Resolution network with Multi-scale Progressive Fusion (HR-MPF) and a proposed Progressive Decoding Module (PDM) recovering final pixel-wise prediction results. Specifically, the proposed HR-MPF firstly incorporates boosted module to High-Resolution Network (HRNet) in a progressive feature fusion manner. In this case, feature communication is augmented among all levels in this high-resolution network. Then, downstream classification module would identify benign and malignant pulmonary nodules based on feature map from PDM. In the adversarial training scheme, a discriminator is set to optimize HR-MPF and PDM through back propagation. Meanwhile, a reasonably designed multi-task loss function optimizes performance of segmentation and classification overall. To improve the accuracy of boundary prediction crucial to nodule segmentation, a boundary consistency constraint is designed and incorporated in the segmentation loss function. Experiments on publicly available LUNA16 dataset show that the framework outperforms relevant advanced methods in quantitative evaluation and visual perception.

**Keywords:** Pulmonary nodule, Segmentation and classification, High-resolution network, Multi-scale progressive fusion, Generative adversarial network

## 1 Introduction

Timely diagnosis and treatment of lung cancer can reduce mortality rate of patients, and early manifestation of lung cancer is mainly pulmonary nodules [1]. Popularity of Computed Tomography (CT) enables radiologists to diagnose pulmonary nodules with more convenience. However, it is a time-consuming and labor-intensive task for radiologists to make accurate judgements from a number of CT scans. Moreover, the volumes of pulmonary nodules are relatively small, and the shapes are easily confused with surrounding blood

vessels in most cases. Nowadays, Computer-Aided Diagnosis (CAD) has been widely used to assist radiologists for diagnosis, and pulmonary nodule segmentation and classification of benign and malignant are crucial steps in CAD systems [2]. The shape, size, growth position and other characteristics can be observed from precise segmentation results of various types of pulmonary nodules in CT images, which also provide a reference for the classification of benign and malignant pulmonary nodules. Therefore, the accurate segmentation and classification crucial in early detection of pulmonary nodules have raised great research value.

In this paper, we design a network based on adversarial training scheme for pulmonary nodule segmentation and classification at one time. High-Resolution network with Multi-scale Progressive Fusion (HR-MPF) is proposed based on High-Resolution Network (HRNet) [3]. In comparison with HRNet, HR-MPF is a high-resolution network reduced to only three stages with several modifications. In HR-MPF, modified boosted module is inserted in the network in multi-scale progressive feature fusion manner to deliver spatial and context information from different resolutions. Specifically, the modified boosted module engages residual blocks that can resolve vanishing gradient problem and promote effective feature learning. Corresponding to the feature extraction network HR-MPF, a Progressive Decoding Module (PDM) is proposed to recover the pixel-wise segmentation prediction from the output of HR-MPF. Then, fused feature map is fed to classification module to distinguish types of pulmonary nodules. In the adversarial training scheme, a discriminator is integrated to optimize HR-MPF and PDM, inspired by the generative adversarial network. Hence, the optimized segmentation module enhances segmentation and classification performance in general. In addition, a simple but practical loss function with boundary consistency constraint is proposed in the segmentation loss. This constraint measures the inconsistency between the boundaries of segmentation prediction and ground truth such that accuracy of boundary prediction is improved. Experiments on LUNA16 dataset show superior results both for segmentation and classification in quantitative evaluation and visual perception.

The main contributions of this paper are as follows:

- An HR-MPF enhancing feature communication of all scales is proposed in which the boosted module is introduced to HRNet based on multi-scale progressive fusion strategy;
- A PDM is proposed to recover final pixel segmentation prediction in progressive fusion manner and refine the output from HR-MPF;
- A discriminator is set to provide additional supervision to the training process of HR-MPF and PDM;
- A loss function with boundary consistency constraint is proposed that improves the accuracy of boundary segmentation. At the same time, a reasonably designed multi-task loss function jointly optimizes the whole framework.

## 2 Related works

### 2.1 Pulmonary nodule segmentation

With the rapid development of deep learning in recent years, Convolutional Neural Network (CNN) has been widely applied to the field of medical image processing [4] and computer vision [5]. By quick features extraction network trained in supervised manner, nodule segmentation and classification are improved by deep learning-based methods. Previously, encoder-and-decoder architectures were widely utilized in these semantic segmentation tasks. Chen et al. [6] established a framework, in which the encoder extracted features of nodules, Atrous Spatial Pyramid Pooling (ASPP) captured multi-scale information through atrous convolution at different atrous rates, and decoder recovered spatial resolution. In U-Det [7], Bidirectional Feature Pyramid Network (BiFPN) was added between the encoder and decoder for multi-scale feature fusion. Mish activation function and class weight of mask were also used to improve segmentation accuracy. Some other works also focus on the context information of pulmonary nodules indicated by the relationship between pixels. For instance, CoLe-CNN [8] accessed the context information of nodules by generating two masks of all background and secondary elements, and introduced an asymmetric loss function that could automatically compensate for errors in annotations of nodules. CDP-ResNet [9] combined the multi-view and multi-scale features of nodules and extracted rich local features and context information. Yan et al. [10] introduced a Mask R-CNN-based segmentation method to deal with class imbalance problem. At present, more and more studies pay attention to 3D segmentation of pulmonary nodules due to the significance of spatial structure characteristics for medical diagnosis. Wang et al. [11] adopted a 3D segmentation network for pulmonary nodules to obtain the three-dimensional global features of pulmonary nodules. Sun et al. [12] and Dong et al. [13] took the slices of different views in CT scans as input to realize multi-view collaborative learning.

### 2.2 Generative adversarial network-based medical image segmentation

Since medical images involve privacy of patients and labeling is labor-consuming, most of the medical image datasets are limited in quantity and annotation information. Among pulmonary nodule datasets, the number of benign nodules also exceeds malignant nodules. In order to generate more diverse CT images for training, Generative Adversarial Network (GAN) [14] is widely used as a data augmentation method. For example, Conditional Generative Adversarial Network (CGAN) was adopted by Qin et al. [15] to synthesize CT images, and a 3DCNN network with residual units was established for pulmonary nodule segmentation, so that the training speed and segmentation accuracy were greatly improved. Onishi et al. [16] synthesized pseudo-pulmonary nodules on the basis of the three-dimensional regions of pulmonary nodules extracted from CTs. Considering of the diversity of pulmonary nodules, Shi et al. [17] introduced a style-based GAN to synthesize the pulmonary nodules with different styles, and the experiments proved that using augmented samples can obtain more accurate and robust segmentation results.

GAN can not only be used to synthesize images and augment datasets, but also effectively improve the quality of segmentation results. Some researchers [18–20] combine a segmentation network, which is regarded as a generator, with a discriminator to process

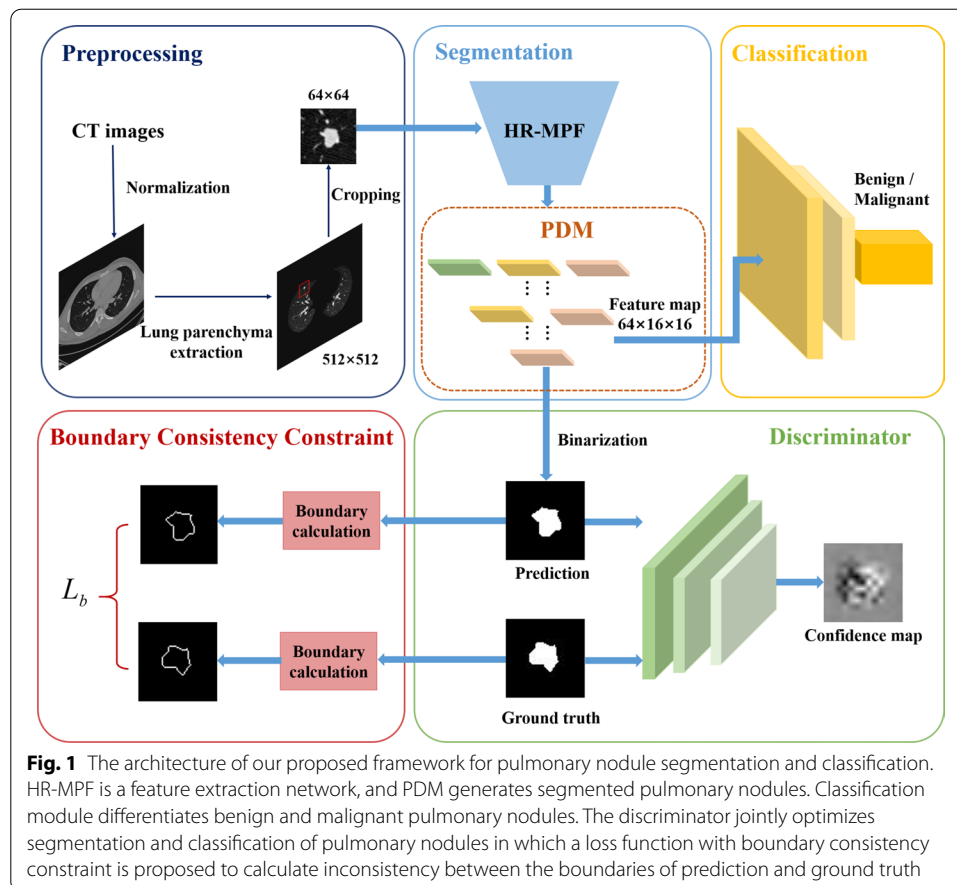
the segmentation task. In the works of Nie et al. [18] and Decourt et al. [19], the generator outputted segmentation predictions and the discriminator outputted confidence maps. The regions with high confidence in the segmentation results were used to further guide the training process of segmentation network. The discriminator in Spine-GAN [20] outputted 0 or 1 representing whether the input was ground truth or prediction result. In some other GAN-based segmentation methods, there is not only a segmentation network, but also a generator. For instance, in Parasitic GAN [21], the segmentation network generated pixel-wise segmentation predictions, the generator synthesized supplementary label maps based on the input random noise, so that the discriminator could learn the more accurate boundaries of ground truth.

### 2.3 Pulmonary nodule classification

Accurate segmentation of pulmonary nodules is of great importance to CAD systems, and the classification of benign and malignant nodules is essential to timely treatment of lung diseases. In order to cope with the variant characteristics of pulmonary nodules, some modifications are made on CNN-based framework and applied to nodules classification. Wang et al. [22] proposed a CNN based on multi-path and multi-scale network robustness of variance in pulmonary nodule volumes and shapes. In STM-Net [2], the scale transfer module and multi-feature fusion operation could enlarge small targets and adapt to images of different resolutions. Multi-planar analyses of pulmonary nodules, that take the features from different places into consideration, are also adopted in some classification methods. For example, the methods of pulmonary nodule classification proposed by Zhang et al. [23] and Onishi et al. [24] extracted features from CT images in three planes (coronal plane, sagittal plane and axial plane). In addition, researchers consider that the malignancy of pulmonary nodules is related to not only its morphological characteristics, but also patients' personal condition. Tong et al. [25] put forward an automated pulmonary nodule diagnosis system and multiple kernel learning algorithm, combined patients' age, medical history and other personal information to classification based on the shape characteristics of nodules.

## 3 Methodology

The architecture of the proposed framework is shown in Fig. 1, which is designed for pulmonary nodule segmentation and classification all at once. In this paper, the proposed HR-MPF with multi-scale progressive fusion scheme is used for segmentation and an adversarial strategy is applied to optimize segmentation module through back propagation. Specifically, the input CT images are preprocessed to assist efficient segmentation of HR-MPF. Meanwhile, through confrontation training between segmentation network and discriminator, the segmentation results can gradually approach to ground truth. To improve boundary pixel prediction accuracy which is a main challenge of pulmonary nodule segmentation, a loss function of boundary consistency constraint is proposed. Finally, the feature map from PDM would be delivered to classification module to discriminate types of the pulmonary nodule (benign/malignant).

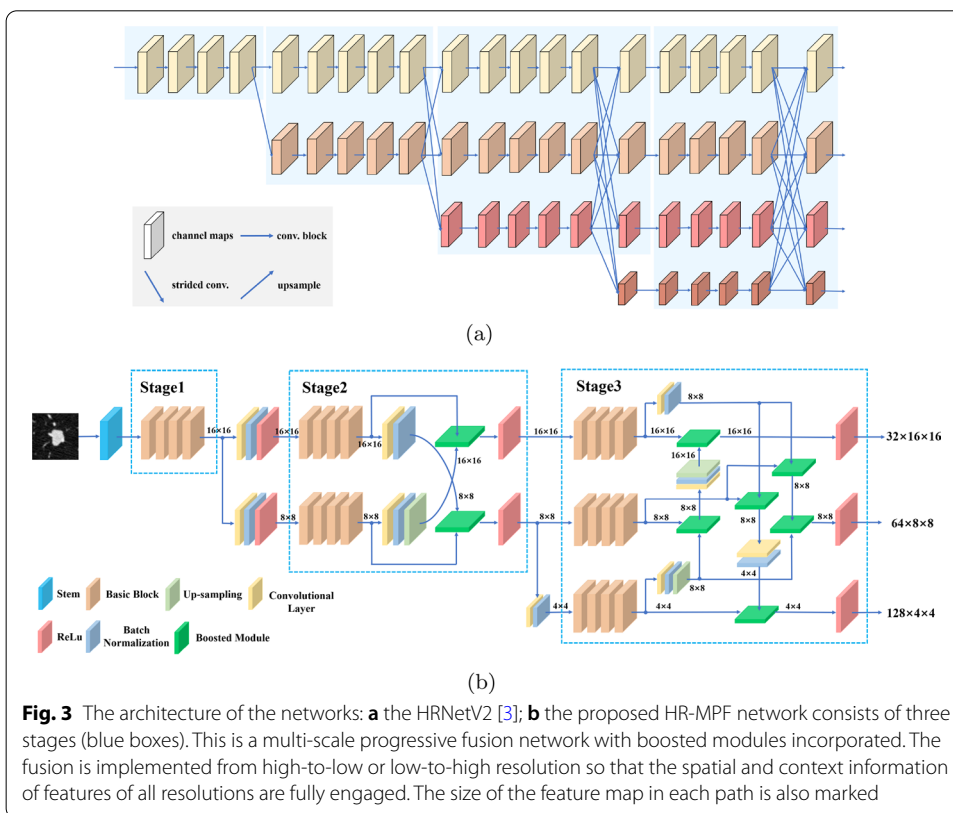
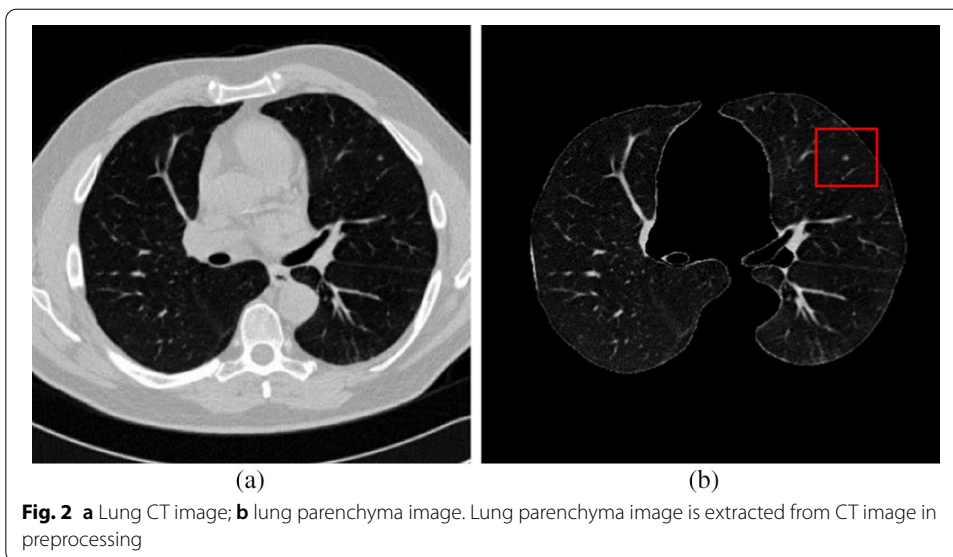


### 3.1 Preprocessing

As a preliminary step, all images are re-sampled to make the pixel interval uniform, because of the different pixel intervals in different CT scans. The preprocessing process includes three steps: (i) CT scan with an effective value of Hounsfield unit between  $[-1000, +400]$  is transformed into the range of 0–255 through linear mapping; (ii) lung CT images contain not only lung tissues, but also blood vessels, lung trachea, and other external tissues. However, pulmonary nodules are located inside or at the edge of the lung in most cases, so it is necessary to remove the tissue other than lung parenchyma to reduce interference with pulmonary nodule segmentation. Therefore, we apply a mask to the normalized CT scan for obtaining lung parenchyma shown in Fig. 2; (iii) considering that pulmonary nodule accounts for a small proportion in a whole CT image, a  $64 \times 64$  region of interest (ROI) in the lung parenchyma is cropped, as the part in the red box in Fig. 2, and the cropped images will be input to the proposed feature extraction network.

### 3.2 Pulmonary nodule segmentation module

Pulmonary nodule segmentation is performed through the proposed HR-MPF and PDM. As shown in Fig. 3b, the segmentation module of our method adopts classical HRNet (HRNetV2) [3] as baseline, and some boosted modules are merged with



multi-scale progressive fusion scheme. By this network architecture, spatial and context information of different resolutions could be comprehensively fused.

As shown in Fig. 3a, stage 2, stage 3 and stage 4 of HRNetV2 [3] contain a multi-resolution group convolution and a multi-resolution convolution. In multi-resolution

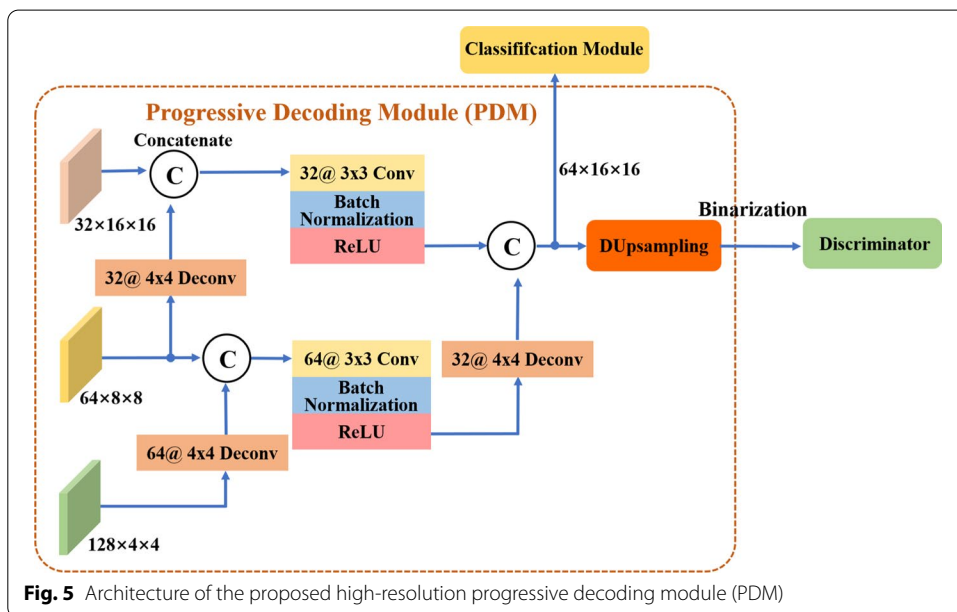
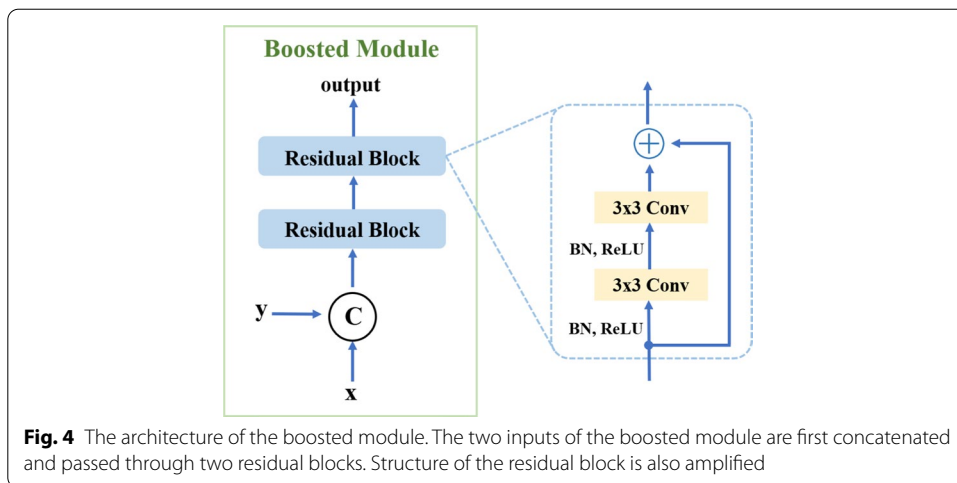


convolution, the resolution of each input subset is adjusted by strided convolution or bilinear up-sampling, and each output subset fuses the features by adding input subsets of same resolution and channels. This feature fusion manner could merge features from different input subsets, but this pure summation scheme may cause loss of information from different resolutions. Therefore, HR-MPF that incorporates boosted module to HRNet in a progressive fusion manner is designed to eliminate loss of different resolution information to the greatest extent. Then, a corresponding PDM is also designed to generate the final pixel-wise predictions of the same size of as input images recovering from coarse feature maps outputted by HR-MPF.

1) *HR-MPF* Considering that low-level features with higher resolution contain much spatial and detailed information, while the high-level features are rich in semantic information. Therefore, this study introduces a modified boosted module to original basic units of HRNet and concatenates them through progressive feature fusion manner. In this case, features of different scales can well deliver their information to each other.

The proposed HR-MPF extracts features of different resolutions gradually in three stages and maintains a high-resolution representation throughout the network. Different from HRNetV2 with four stages, the size of the input pulmonary nodule image is  $64 \times 64$  such that three stages are enough for feature extraction. As shown in Fig. 3b, HR-MPF starts at a high-resolution stage with only one branch of feature map size  $16 \times 16$ , and a branch whose resolution is  $1/2$  of the lowest resolution in previous stage is added in each stage subsequently. By the progressive fusion strategy applied, semantic information from low-resolution branches and spatial information from high-resolution branches are delivered to each branch with modified boosting method incorporated. Progressive feature fusion is implemented from the highest resolution to the lowest resolution or the lowest to highest. In stage 3 for example, the strided convolution or up-sampling is used for changing the resolution of input to ensure both inputs of the boosted module have the same resolution. Fusion is implemented from high-to-low or low-to-high resolutions so that the output feature map would maximally preserve structural details and semantic information from all three branches. In low-to-high fusion pathway, the  $4 \times 4$  input first undergoes a convolution and up-sampling and is fused with the  $8 \times 8$  input by a boosted module. Then, the fusion result undergoes a convolution and up-sampling, and is finally fused with the  $16 \times 16$  input. Similarly, in the high-to-low fusion pathway, the  $16 \times 16$  input first undergoes a strided convolution for down-sampling so that its resolution becomes  $8 \times 8$ . Then, it is fused with the  $8 \times 8$  input with a boosted module, and the output of the boosted module is  $8 \times 8$ . The output also undergoes a strided convolution for down-sampling and is finally fused with the  $4 \times 4$  input with a boosted module. Ultimately, three feature maps engaged feature information of all scales would be conveyed to the proposed PDM for decoding. By this progressive feature fusion strategy, cooperative representations can be promoted [26] and the spatial and context information between multi-scale features are effectively extracted and integrated.

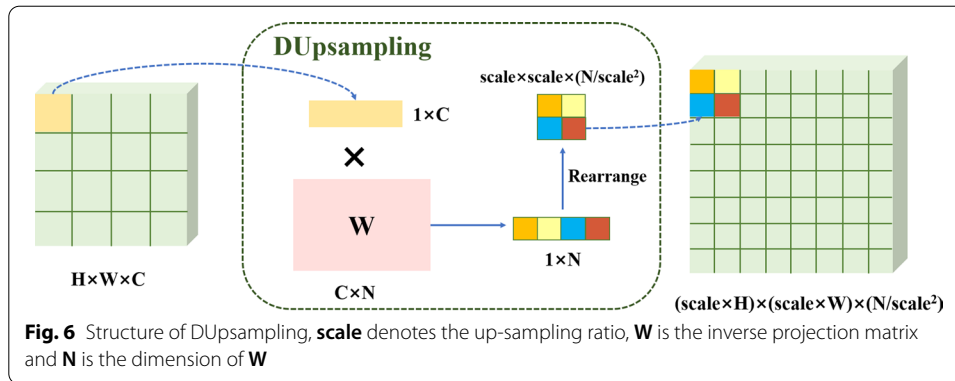
The boosting method in our progressive fusion network refines results by fusing the current and another level of features. Boosted module was initially used in image denoising, but it has also been applied to built a boosted decoder for task such as dehazing [27]. The concatenation style of boosted module we use (see Fig. 4) belongs to “U-Net module” classified as an alternative of the SOS boosted module used in [27]. Considering



the size of the input image, our boosted module is a stack of two residual blocks and the computation can be simultaneously compressed. Each residual block consists of two 3×3 convolutional layers, batch normalization and ReLU layers. The connection method of residual blocks adopts the method proposed in [28] that the batch normalization and ReLU layers are placed before the convolution layers. The residual learning mechanism is suggested to well solve problem of gradient vanishing in back propagation when layers of the network become deeper.

2) *Progressive decoding module* As shown in Fig. 5, a new decoder module PDM for HR-MPF is designed to progressively recover integrated feature maps to original sizes. Different from the representation head in HRNetV1 [29] and HRNetV2 [3] that recover final pixel-wise prediction results by direct up-sampling, PDM integrates feature maps by feature fusion among all levels progressively. Generated feature maps with resolutions of 16×16, 8×8 and 4×4 are taken as input, of which 8×8 and 4×4 feature maps go through



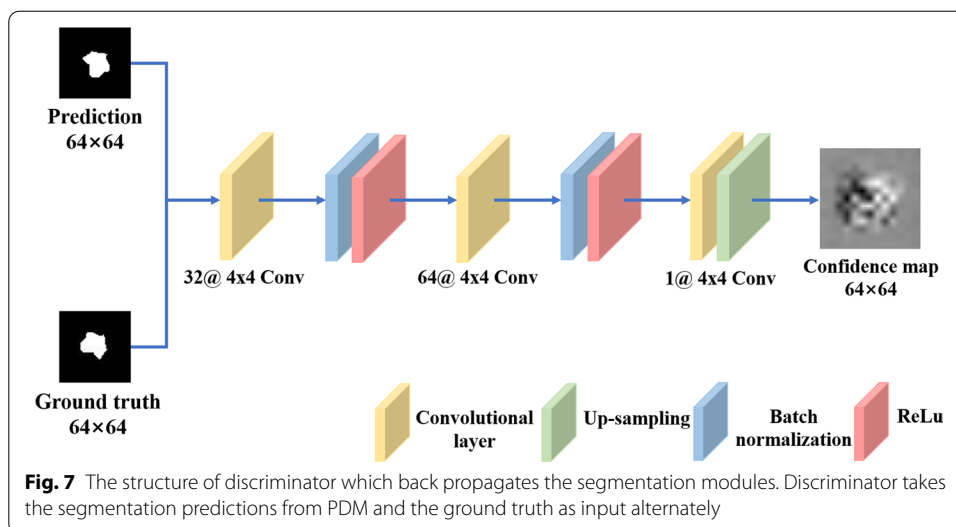


a deconvolution layer and are concatenated with feature maps of higher resolution ( $16 \times 16$  and  $8 \times 8$ ), respectively. Next, the two concatenated feature maps are separately convolved, and then concatenated together after  $8 \times 8$  feature map undergoes a deconvolution layer. Finally, a  $16 \times 16$  feature map with 64 channels is obtained and would be delivered to the classification module to discriminator types of pulmonary nodule. The optimized segmentation prediction would be then passed through a DUpsampling for binarization and fed into the discriminator.

3) *DUpsampling* The last layer of the many semantic segmentation networks with encoder–decoder architectures [30, 31] typically exploit bilinear up-sampling to recover final segmentation prediction. However, this data-independent and over-simple bilinear up-sampling may result in sub-optimal results, whose capability in recovering detailed edge and texture features is limited. Our segmentation module adopts DUpsampling [32], which is a data-dependent up-sampling that considers the correlation among prediction of each pixel. It recovers pixel-wise prediction from final high-resolution representation in the PDM. As shown in Fig. 6, DUpsampling takes effect the same as applying a  $1 \times 1$  convolutional layer along spatial dimensions, and the convolutional kernels are stored in a learnable reconstruction matrix  $\mathbf{W}$ . DUpsampling makes use of the redundancy in label space of pulmonary nodule segmentation that can be compressed considerably almost with no loss. Although two matrixes computed in pre-training are suggested [32], matrix  $\mathbf{P}$  is used to compress segmentation labels with linear projecting and  $\mathbf{W}$  is the corresponding inverse projection matrix. Considering that two approaches could be selected to compute the segmentation loss in our adversarial framework, one is calculating loss between coarse outputs of the encoder and compressed labels, the other is between the decompressed coarse outputs and segmentation labels. The PDM takes the second strategy that matrix  $\mathbf{W}$  is utilized for recovering the pixel-wise segmentation prediction from the final  $64 \times 16 \times 16$  feature map.

### 3.3 Pulmonary nodule classification module

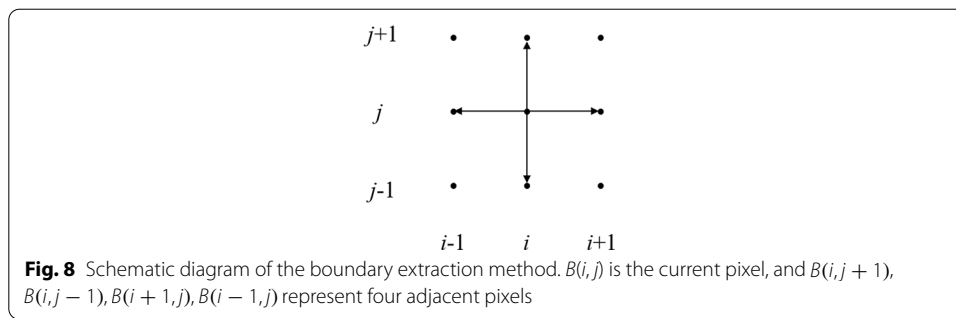
Since pulmonary nodules usually distribute in very small span, accurate segmentation is appreciated for classification training. Follow-up classification module would classify the types of pulmonary nodule based on the largely improved segmentation results from previous segmentation module. The input size of classification module is  $64 \times 16 \times 16$  and would be transformed into a 128-dimensional feature vector after a global average



pooling layer and a  $1 \times 1$  convolutional layer. Finally, the possibilities of benign and malignant are output through a fully connected layer.

### 3.4 Discriminator

In current segmentation networks, cross-entropy loss and dice loss are commonly used to minimize the differences between prediction result and ground truth. However, a discriminative network can more efficiently guide the learning of segmentation towards desirable results by back propagating mismatches between prediction result and ground truth [33]. Inspired by GAN, the proposed framework takes the segmentation network as generator and adopts a discriminator at the end for adversarial training. In the training process of our discriminator, ground truth or a prediction result of PDM would be input, and the output is a confidence map with the same size as input. As a supervisory signal, confidence map indicates the quality of the segmentation and helps segmentation module to know the regions it can trust during the training [11]. Specifically, the discriminator network is composed of three convolutional layers of 64, 128 and 1 channels, respectively, and convolution kernels of them are all  $4 \times 4$ . The structure of discriminator is shown in Fig. 7. The convolution stride of the first two layers is 2, while the stride of the last layer is 1. Batch normalization and ReLU activation function are used after the first two convolutional layers, and the output is scaled to the size of the input through bilinear up-sampling. Finally, the discriminator outputs a  $1 \times 64 \times 64$  confidence map, and each pixel represents the probability of ground truth or prediction results from segmentation module [20]. This process is carried out on a pixel-by-pixel basis, which can improve the accuracy of segmentation.



### 3.5 Loss function of networks

In the adversarial training of our method, a segmentation loss with our proposed boundary consistency constraint loss is used for segmentation optimization, and a cross-entropy loss is applied for classification optimization.

1) *Segmentation loss* Our segmentation task is to classify all pixels of the input images into nodules and background, which is a pixel-level binary classification problem. The loss function  $\mathcal{L}_{\text{seg}}$  is used to measure the gap between prediction results and ground truth. It adopts cross-entropy loss, which is a practical and effective loss function and often applied in classification problems. In fact, it is a convex function convenient for optimization during training as follows:

$$\mathcal{L}_{\text{seg}} = -\frac{1}{N} \sum_{i,j} \sum_{c \in C} y(i, j) \log(\hat{y}(i, j)), \quad (1)$$

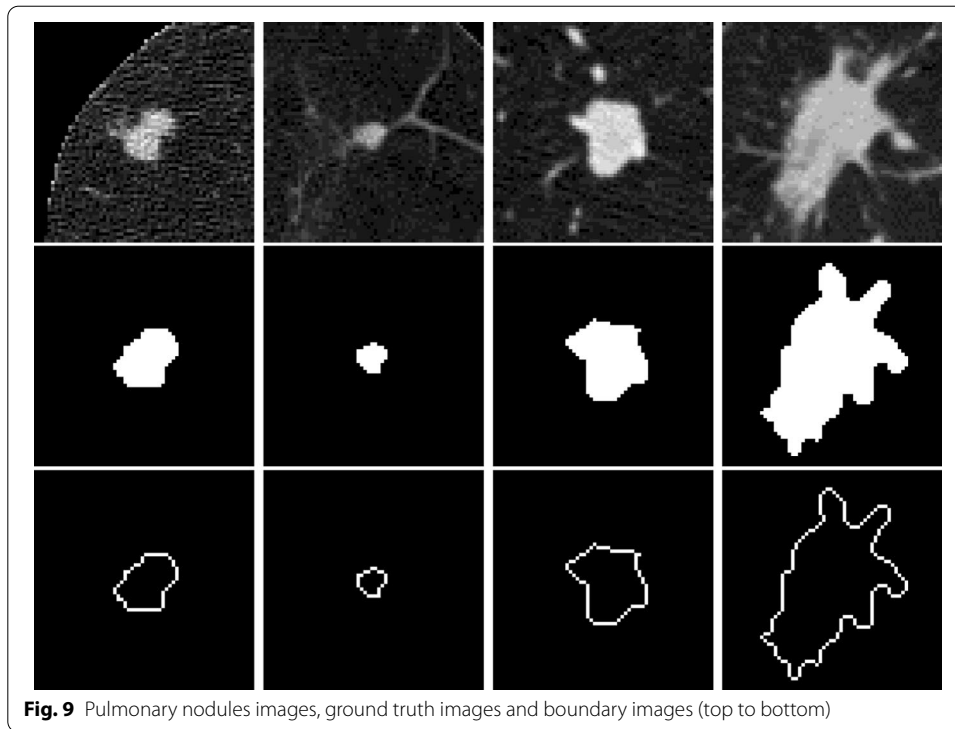
where  $y$  is ground truth and  $\hat{y}$  is the prediction results of segmentation module. The pixel value at coordinate  $(i, j)$  in  $y$  is expressed as  $y(i, j)$ .  $N$  is the total number of pixels in  $y$ , and  $C$  denotes the number of categories.

The adversarial loss function is used to train the segmentation module and fool the discriminator  $D$  by maximizing the probability of segmentation prediction, thus the segmentation module output can be closer to the ground truth. Our adversarial loss is defined as:

$$\mathcal{L}_{\text{adv}} = -\mathbb{E}[D(\hat{y})], \quad (2)$$

where  $D(\hat{y})$  means the output of the discriminator for input segmentation prediction results. As indicated by [34] that it's difficult to handle the prediction error at the boundaries of segmented objects well and the number of error pixels increases with the distance from the boundary getting closer. This means that the prediction of boundary pixels is relatively unreliable in overall segmentation. Besides, shapes and sizes of pulmonary nodules vary and most of them are irregular, which also increase difficulty of boundary prediction. As a result, the refinement of boundary prediction affects significantly in accurate segmentation results.

Therefore, this paper applies a loss function with boundary consistency constraint to calculate the inconsistency between the boundaries of prediction result and ground truth to improve boundary pixel segmentation accuracy. Since the segmentation results and ground truth of pulmonary nodules are binary images, we only need to make judgement on each pixel. If the current pixel  $B(i, j)$  is 1, and one or more of its four adjacent



pixels is 0, the pixel is a boundary pixel. As shown in Fig. 8, the schematic diagram of the boundary extraction method,  $B(i, j)$  is the current pixel, and  $B(i, j + 1)$ ,  $B(i, j - 1)$ ,  $B(i + 1, j)$ ,  $B(i - 1, j)$  are the four adjacent pixels. Figure 9 shows the several boundary images obtained by the proposed method.

The boundary of prediction results  $B_{\text{pre}}$  and ground truth  $B_{\text{gt}}$  are obtained, respectively, the loss function with boundary consistency constraint  $\mathcal{L}_b$  is calculated by:

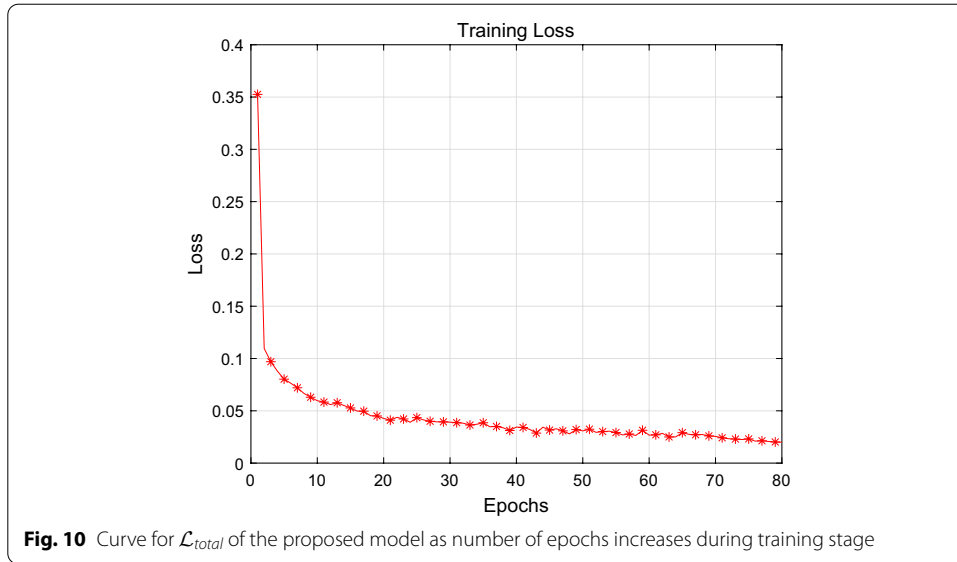
$$\mathcal{L}_b = \frac{\sum_{i,j} |B_{\text{pre}}(i, j) - B_{\text{gt}}(i, j)|}{\sum_{i,j} B_{\text{pre}}(i, j) + \sum_{i,j} B_{\text{gt}}(i, j)}. \quad (3)$$

In order to improve the performance of pulmonary nodule segmentation, two consistency constraints are introduced in this study. The loss functions  $\mathcal{L}_{\text{seg}}$  and  $\mathcal{L}_b$  measure the consistency of pixels and boundaries between the prediction results and ground truth, respectively. Therefore, the loss function needs to be minimized in the training process of the segmentation module, and it is given by:

$$\mathcal{L}_S = \mathcal{L}_{\text{seg}} + \lambda_1 \mathcal{L}_{\text{adv}} + \lambda_2 \mathcal{L}_b, \quad (4)$$

where  $\mathcal{L}_{\text{seg}}$  is segmentation loss,  $\mathcal{L}_{\text{adv}}$  is adversarial loss, and  $\mathcal{L}_b$  is the loss function of boundary consistency constraint. Both  $\lambda_1$  and  $\lambda_2$  are two weights to balance  $\mathcal{L}_{\text{adv}}$  and  $\mathcal{L}_b$ .

2) *Classification loss* Since the classification task judges the type of nodule as benign or malignant, we use cross-entropy usually applied for binary classification of discrete target



**Fig. 10** Curve for  $\mathcal{L}_{total}$  of the proposed model as number of epochs increases during training stage

variables to determine how close the actual output is to the expected. Therefore, the loss of classification training could be defined as:

$$\mathcal{L}_{cls} = - \sum_{k=1}^n p_k \log \hat{p}_k, \tag{5}$$

where  $p_k \in \{0, 1\}$  is ground truth,  $p_k=0$  represents that nodule is benign, while  $p_k=1$  represents nodule is malignant.  $\hat{p}_k \in [0, 1]$  is prediction of classification module.  $n$  is the batch size, and  $k$  is the index of samples in a batch.

Since this network architecture achieves segmentation and classification in one model, a multi-task loss function is used for joint optimization as follows:

$$\mathcal{L}_{total} = \mathcal{L}_S + \lambda_3 \mathcal{L}_{cls} = \mathcal{L}_{seg} + \lambda_1 \mathcal{L}_{adv} + \lambda_2 \mathcal{L}_b + \lambda_3 \mathcal{L}_{cls}. \tag{6}$$

The values of  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  will be discussed in Section 4.3. Figure 10 displays the loss function mentioned in (6), which indicates the convergence performance of the proposed method.

3) *Loss function of discriminator* Although the training of GAN is the process of confrontation between generator and discriminator, the segmentation module outputting segmentation results could be regarded as a generator. The objective of the discriminator is to accurately determine the source of input and make the output of generator approach real data distribution during the confrontation training. Therefore, the design of a reasonable loss function of discriminator is also crucial to the training process of GAN. WGAN [35] improves the traditional GAN loss function and solves the problem of unstable training and collapse mode. In our method, the loss function adopts that in WGAN, which is defined as follows:

$$\mathcal{L}_D = -\mathbb{E}[D(y)] + \mathbb{E}[D(\hat{y})], \tag{7}$$

where  $D(y)$  means the output of the discriminator for input ground truth.  $\mathcal{L}_D$  would maximize  $\mathbb{E}[D(y)]$  when the input is ground truth and minimize  $\mathbb{E}[D(\hat{y})]$  when the input is segmentation result. The smaller the  $\mathcal{L}_D$  is, the smaller the Wasserstein distance between the real distribution and the generated distribution is. After each update of the parameters of discriminator, the absolute values of them are truncated to no more than a fixed constant  $c$ , which is set to 0.01 in our experiments. The segmentation task is to classify all the pixels of the input images into nodules and backgrounds, which is a pixel-level binary classification problem.

## 4 Results and discussion

### 4.1 Datasets and evaluation metrics

In this paper, we employ publicly available LUNA16<sup>1</sup> dataset for evaluating all methods. LUNA16 excludes CT images with slice thickness larger than 3mm and pulmonary nodules with diameters less than 3mm from the LIDC-IDRI dataset [36]. LUNA16 contains 888 CT scans where 1186 pulmonary nodules are annotated by at least three radiologists. The degree of malignancy of each pulmonary nodule is evaluated with a score of 1 ~ 5, the higher the score, the higher the degree of malignancy. Nodules with a mean of 1 or 2 are classified as benign, with a mean of 4 or 5 are classified as malignant, and with a mean of 3 are ignored [23]. Ultimately, 835 nodules are obtained, including 539 benign nodules and 296 malignant nodules. At the same time, the dataset is augmented with the operation of flip, rotation and transpose and then enlarged by 6 times. There are a total of 5845 nodules, of which 4968 are used for training (3207 benign nodules, 1761 malignant nodules) and 877 are used for testing (566 benign nodules, 311 malignant nodules).

Some standard metrics namely Accuracy (Acc), Dice Similarity Coefficient (DSC), Mean Intersection over Union (MIoU), Precision (Prec), Sensitivity (SE) and Specificity (SP) are used to validate the performance of our method. The less popular evaluation criteria MIoU is defined as:

$$\text{MIoU} = \left( \frac{\text{TP}}{\text{FN} + \text{TP} + \text{FP}} + \frac{\text{TN}}{\text{FP} + \text{TN} + \text{FN}} \right) / 2, \quad (8)$$

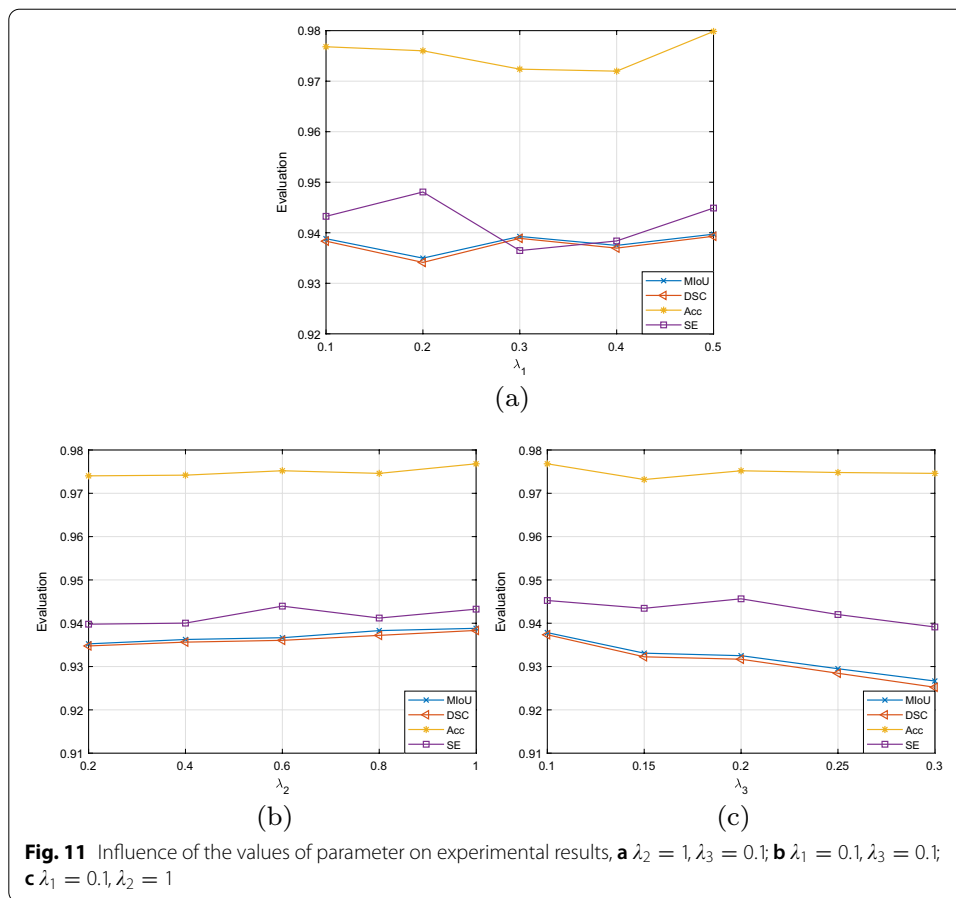
where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. For the pulmonary nodule segmentation task, the nodular area is positive, the non-nodular area is negative. For the pulmonary nodule classification task, we define that benign nodules are negative, and malignant nodules are positive.

### 4.2 Implementation details

The proposed method is trained on the Pytorch platform with Python 3.6. The training and validation of overall network are performed on a computer with Intel(R) Core (TM) i7-10700 CPU (2.9GHz) with 16G RAM memory, and NVIDIA GeForce GTX 2070 SUPER with 8G memory.

<sup>1</sup> <https://luna16.grand-challenge.org/>.

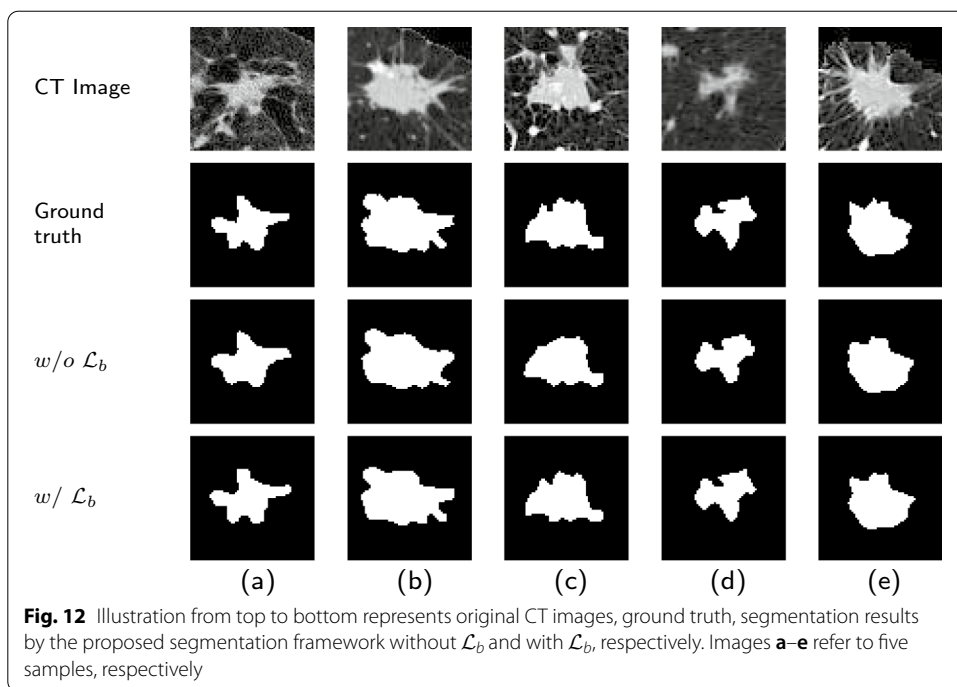




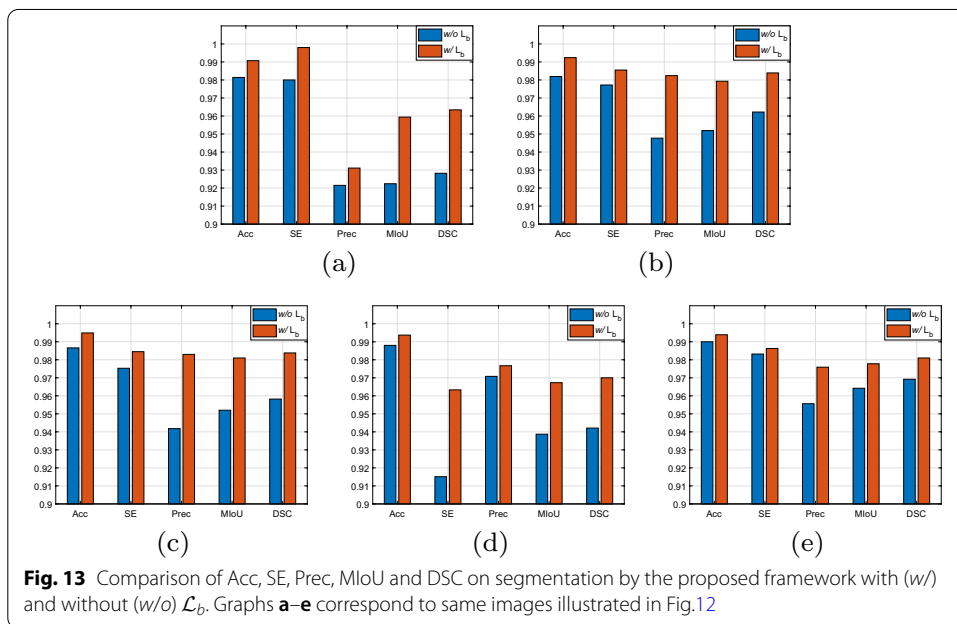
During the experiment, the sizes of all input images are  $64 \times 64$  and the number of epoch and batch size are set to 80 and 16, respectively. AdaBelief (learning rate= $2.5 \times 10^{-4}$ , eps= $10^{-6}$ , Betas=(0.5, 0.999)) is used as optimization algorithm of segmentation network, which is of fast convergence and high accuracy, and performs high stability when training a GAN [37]. In addition, the optimizer of discriminator is RMSProp (learning rate=0.002, eps= $10^{-8}$ ).

### 4.3 Parameter setting

As a one-off pulmonary nodule segmentation and classification method, total loss function  $\mathcal{L}_{total}$  defined in (6) is balanced by three trade-off parameters. Specifically,  $\lambda_1, \lambda_2$  and  $\lambda_3$  control the weights of adversarial loss  $\mathcal{L}_{adv}$ , boundary consistency constraint  $\mathcal{L}_b$  and loss function of classification  $\mathcal{L}_{cls}$ , respectively. To evaluate how individual loss contributes to total loss, we fix any two parameters and measure MIoU, DSC for segmentation task and Acc, SE for classification task when varying the third parameter. As shown in Fig. 11, weight of adversarial loss is less variable compared to the other two as expected. For instance, the value of SE drops from 0.9481 to 0.9365 as  $\lambda_1$  varies from 0.2 to 0.3 most noticeably. By contrast, larger  $\lambda_3$  for  $\mathcal{L}_{cls}$  seems to cause a decrease in both segmentation and classification performance, although quantitative results fluctuate insignificantly. In addition, it could be observed that more boundary loss applied to total adversarial



**Fig. 12** Illustration from top to bottom represents original CT images, ground truth, segmentation results by the proposed segmentation framework without  $\mathcal{L}_b$  and with  $\mathcal{L}_b$ , respectively. Images a–e refer to five samples, respectively



**Fig. 13** Comparison of Acc, SE, Prec, MioU and DSC on segmentation by the proposed framework with ( $w/$ ) and without ( $w/o$ )  $\mathcal{L}_b$ . Graphs a–e correspond to same images illustrated in Fig.12

training results in a slightly better segmentation results in general. Classification result presents a similar trend that Acc and SE peak at 0.9768 and 0.9440 while  $\lambda_2$  setting at 0.6 and 1, respectively. In general, quantitative results fluctuate no more than 0.01 from a wide range of selection for each parameter. By evaluating experimental results on both segmentation and classification,  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$  are set to 0.1, 1, and 0.1, respectively.

#### 4.4 Evaluation on loss function $\mathcal{L}_b$

To verify effectiveness of the proposed boundary loss function  $\mathcal{L}_b$ , experimental results on five instances are conducted whether using  $\mathcal{L}_b$  or not, as shown in Fig. 12. Noticeably, visual results shown in this figure present a much similar outline to ground truth (such as  $w/\mathcal{L}_b$  for Fig. 12a, c). Qualitative results of segmentation task using Acc, SE, Prec, MIoU, DSC are reported in Fig. 13. In all cases, boundary loss contributes to an increase in quantitative results to some extent. Generally, segmentation seems to be affected more significantly since high attention to boundary areas contributes straightforward to segmentation accuracy.

#### 4.5 Evaluation on pulmonary nodule segmentation

In order to evaluate the segmentation performance of our proposed network for pulmonary nodules, we compare it with nine different segmentation algorithms based on deep neural networks, ENet<sup>2</sup> [38], SegNet<sup>3</sup> [30], PSPNet<sup>4</sup> [39], UNet++<sup>5</sup> [40], DeepLabV3+<sup>6</sup> [31], DFANet<sup>7</sup> [41], Fast SCNN<sup>8</sup> [42], FANet<sup>9</sup> [43], SPNet<sup>10</sup> [44]. The implementation of these algorithms is based on publicly published codes. As shown in Fig. 14, DeepLabV3+ and UNet++ achieve relatively better performance among all approaches. The reason might be that the atrous convolutions can capture multi-scale information at different atrous rates, and the encoder–decoder architecture can extract features and recover spatial resolution better in DeepLabV3+. The encoder and decoder are connected through nested and dense skip pathways, which can eliminate semantic gap between feature maps as well by UNet++. However, segmentation results of the proposed model are more precise than comparative methods because of the progressive fusion strategy and loss function with boundary consistency constraint. It could be seen from nodules with complex boundaries especially that the proposed network locates the exact region with very similar boundaries to ground truth (e.g. the 5th and 8th images of Fig. 14).

All quantitative experiments for segmentation on LUNA16 dataset obtained by the proposed HR-MPF and recent deep learning-based models are shown in Table 1. Different from Acc and SE in parameter experiment, Acc and SE in Table 1 measure segmentation results in this comparative experiment. It is obvious that Acc for all methods are of minor differences and the values stay at relatively high level. This is mainly because non-nodular area makes up the majority of CT images in general, which is easily divided into non-nodular area. MIoU and DSC both measure region intersection with ground truth, and DeeplabV3+, UNet++ and SPNet all show relatively competitive results similar as in visual comparison. In detail, DeeplabV3+ achieves 0.9222 and 0.9204, and UNet++ achieves even better results at 0.9263 and 0.9227 for MIoU and DSC, respectively. The

---

<sup>2</sup> <https://github.com/davidtvs/PyTorch-ENet>.

<sup>3</sup> <https://github.com/ZijunDeng/pytorch-semantic-segmentation>.

<sup>4</sup> <https://github.com/kazuto1011/pspnet-pytorch>.

<sup>5</sup> <https://github.com/4uiiur21/pytorch-nested-unet>.

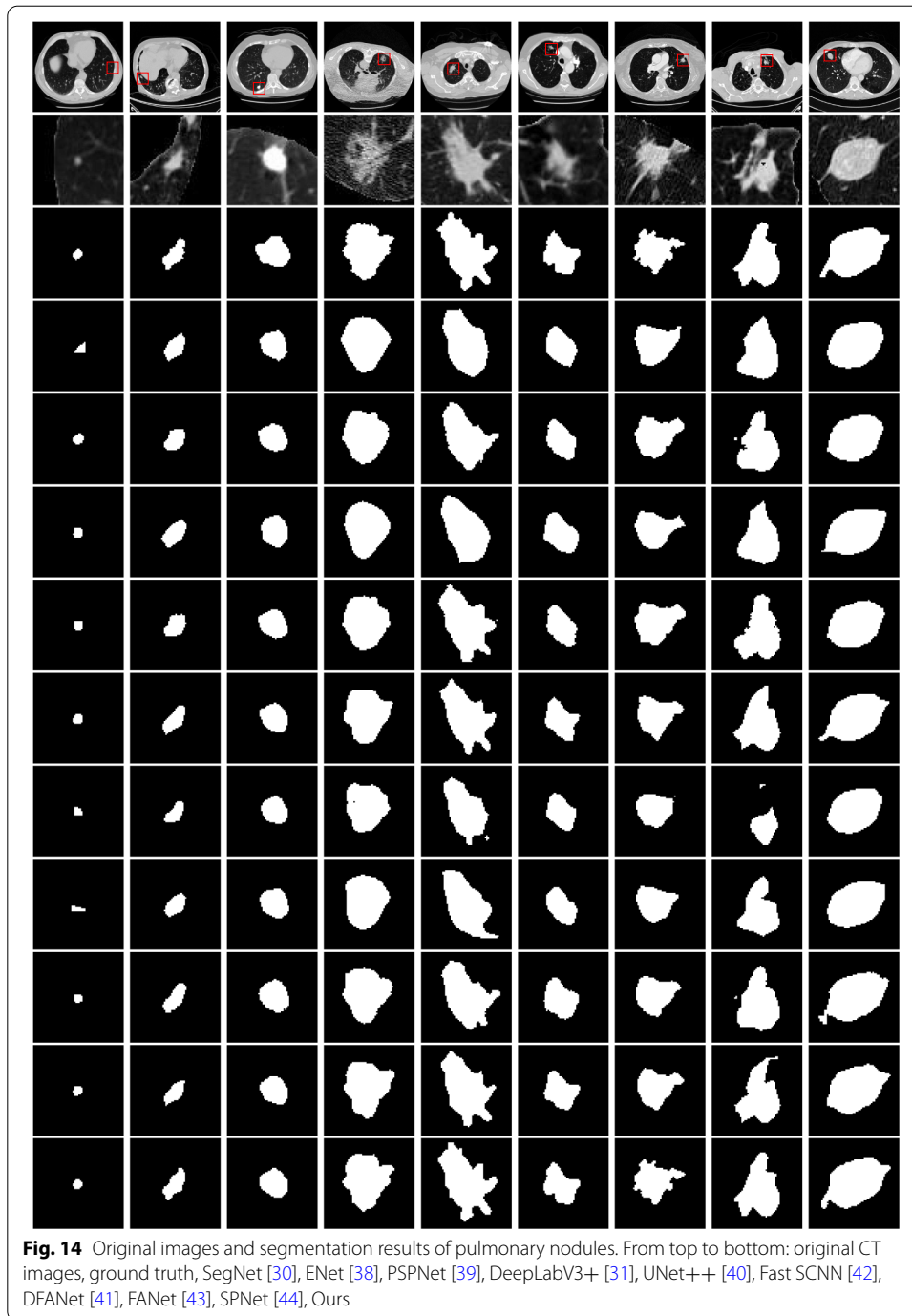
<sup>6</sup> <https://github.com/hualin95/Deeplab-v3plus>.

<sup>7</sup> <https://github.com/huaifeng1993/DFANet>.

<sup>8</sup> <https://github.com/DeepVoltaire/Fast-SCNN>.

<sup>9</sup> <https://github.com/feinanshan/FANet>.

<sup>10</sup> <https://github.com/Andrew-Qibin/SPNet>.



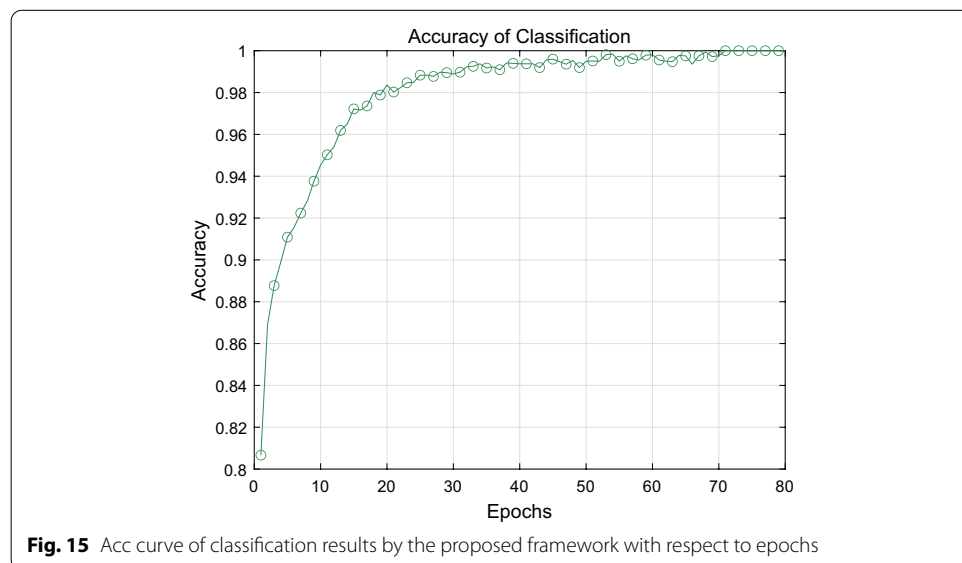
proposed HR-MPF with progressive feature fusion and adversarial training scheme achieves slightly better results than comparative methods in all aspects. It could be seen that the proposed method could qualitatively locate the area of nodules as MIoU and DSC both stay above 0.937. In addition, the proposed method seems to be particularly sensitive rather than missing suspicious areas. In general, the proposed architecture is

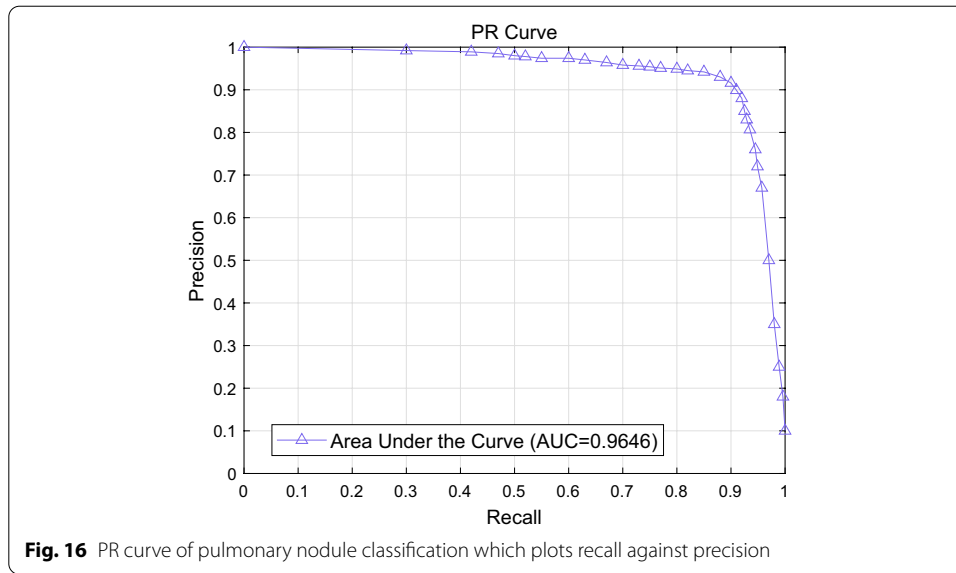
**Table 1** Quantitative results of pulmonary nodule segmentation in which highlighted values represent the best result of each metric

Methods	Acc	SE	Prec	MIoU	DSC
ENet	0.9895	0.8787	0.9032	0.8960	0.8908
SegNet	0.9873	0.8914	0.8852	0.8906	0.8859
PSPNet	0.9888	0.8742	0.8936	0.8900	0.8838
DeeplabV3+	0.9923	0.9168	0.9241	0.9222	0.9204
UNet++	0.9918	0.9150	0.9285	0.9263	0.9227
Fast SCNN	0.9859	0.8130	0.8978	0.8605	0.8477
DFANet	0.9874	0.8208	0.9115	0.8716	0.8612
FANet	0.9901	0.9012	0.9127	0.9049	0.8934
SPNet	0.9931	0.9165	0.9117	0.9106	0.9096
Ours	<b>0.9937</b>	<b>0.9377</b>	<b>0.9427</b>	<b>0.9378</b>	<b>0.9373</b>

**Table 2** Comparison of parameters (M), training time (m) and FLOPs (G) between our method and other segmentation networks, in which highlighted values represent the best result of each criteria

Methods	Parameters (M)	Training time (m)	FLOPs (G)
ENet	<b>0.36</b>	51	0.04
SegNet	53.54	62	4.5
PSPNet	85.86	80	5.05
DeeplabV3+	59.33	71	1.38
UNet++	9.16	42	2.16
Fast SCNN	1.2	<b>38</b>	<b>0.02</b>
DFANet	2.16	69	0.03
FANet	13.65	40	0.09
SPNet	3.8	55	0.17
Ours	3.46	70	0.16





**Table 3** Quantitative results of pulmonary nodule classification in which highlighted values represent the best result of each metric

Methods	Acc	SE	SP	AUC
Zhang et al. [23]	0.8353	0.8046	0.8599	–
Zheng et al. [47]	0.8867	0.8222	0.9317	–
Tong et al. [25]	0.9129	0.9101	0.9140	–
Zuo et al. [46]	0.9733	<b>0.9726</b>	–	<b>0.9954</b>
Agnes et al. [48]	0.8726	0.8100	0.9190	0.9440
Ours	<b>0.9768</b>	0.9453	<b>0.9789</b>	0.9646

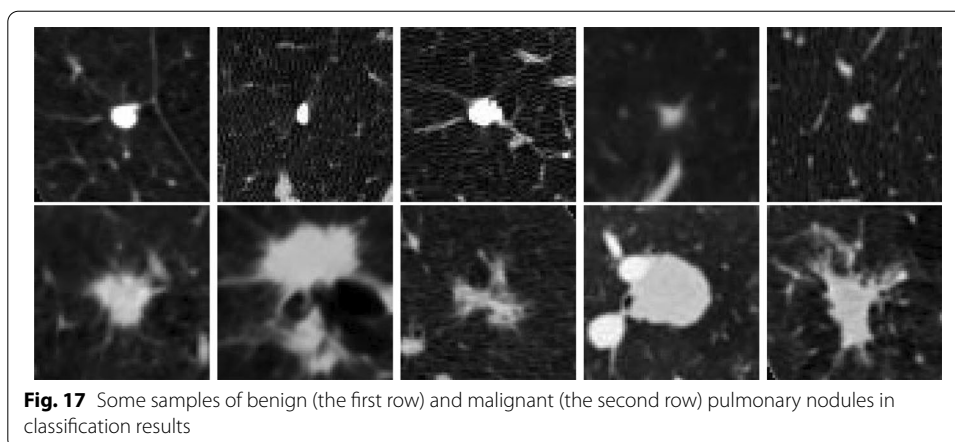
verified effective on improving performance for pulmonary nodule segmentation, while DeeplabV3+, UNet++ and SPNet also achieve qualified results.

In this sub-section, computational cost of this method is evaluated in terms of parameters, training time and FLOPs [45] as shown in Table 2. Input size for all comparative methods is set to the same size ( $64 \times 64$ ) and all experiments are implemented under the same conditions. It could be seen that Fast SCNN uses the least training time (38 min) and has the least model complexity as indicated by FLOPs (0.02G). However, ENet has the least number of parameters (0.36M) though number of parameters for Fast SCNN also stay at a low level (1.2M). Model complexity of the proposed method stays at a moderate level (3.46M parameters, 0.16G FLOPs), but relatively long training time is spent due to the adversarial training scheme we introduce for better results.

#### 4.6 Evaluation on pulmonary nodule classification

Figure 15 demonstrates the influence of iteration on classification accuracy. It could be observed that the Acc curve raises relatively quickly. In detail, Acc extremely approaches 1 as the epoch increases to 30. Therefore, it could be concluded that the training of classifier is relatively convergent and stable. Figure 16 illustrates the precision–recall curve which indicates the influence of thresholds on classifier





**Table 4** Comparison of our method with other versions of HRNet, the highlighted values represent the best results

Networks	Acc	SE	Prec	MIoU	DSC
HRNetV1 [29]	0.9914	0.9212	0.9326	0.9208	0.9162
HRNetV2 [3]	0.9921	0.9231	0.9329	0.9216	0.9197
Ours	<b>0.9923</b>	<b>0.9240</b>	<b>0.9347</b>	<b>0.9286</b>	<b>0.9252</b>

performance. It could be observed that the precision–recall curve stays close to the top-right corner where the area under curve (AUC) arrives at 0.9646.

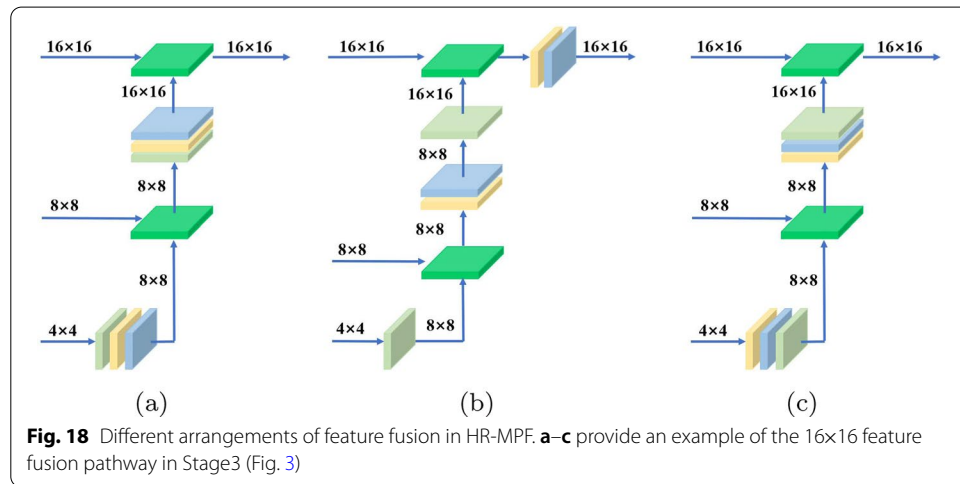
In order to evaluate the classification performance of proposed method, we compare it with five classification networks in terms of Acc, SE, SP and AUC as listed in Table 3. Evaluation results are taken from cited papers and the result for each metric differs noticeably. However, it could be seen that the values of SP are greater than that of SE. This might be because that most malignant nodules have more prominent shape characteristics than benign pulmonary nodules, such as the more irregular shape and larger volume as shown in Fig. 17. In general, Zuo et al. [46] achieves surpassing results where Acc, SE are both over 0.97. The network established by Zuo et al. only conducts classification and uses a multi-resolution convolutional neural network to extract features. In addition, the proposed method also achieves qualified results among comparative methods. Specifically, highest Acc and SP values are achieved (0.9768 and 0.9789) because the proposed framework that jointly conducts segmentation and classification could well capture malignant nodules if it is in prominent shape characteristics. Overall, the proposed method for segmentation and classification provides relatively competitive results among methods for classification only.

#### 4.7 Ablation study

The proposed HR-MPF is modified based on HRNetV2 [3]. To verify the effectiveness of several improvement proposed by HR-MPF, comparative experiments with HRNetV1 [29] and HRNetV2 [3] are implemented. HRNetV1 and HRNetV2 with four stages originally have been reduced to three stages for equivalent comparison

**Table 5** Quantitative results on segmentation and classification by different arrangements of feature fusion in HR-MPF

Categories	Segmentation		Classification		Parameters (M)
	MIoU	DSC	Acc	SE	
Fig. 18a	0.9366	0.9354	0.9759	0.9448	3.46
Fig. 18b	0.9315	0.9309	0.9697	0.9355	3.46
Fig. 18c	<b>0.9378</b>	<b>0.9373</b>	<b>0.9768</b>	<b>0.9453</b>	3.46



**Table 6** Ablation study of each module in the whole framework, the highlighted values represent the best results

Networks	Segmentation					Classification		
	Acc	SE	Prec	MIoU	DSC	Acc	SE	SP
S	0.9923	0.9240	0.9347	0.9286	0.9252	–	–	–
S + C	0.9925	0.9237	0.9353	0.9283	0.9249	0.9650	0.9323	0.9735
S + C + D	<b>0.9937</b>	<b>0.9377</b>	<b>0.9427</b>	<b>0.9378</b>	<b>0.9373</b>	<b>0.9768</b>	<b>0.9453</b>	<b>0.9789</b>

with our network. As shown in Table 4, HRNetV2 achieves slightly higher results than HRNetV1 for all metrics. In comparison, the proposed HR-MPF brings an even noticeable increase especially on MIoU (0.9286) and DSC (0.9252). This experiment generally verifies advantage of the proposed progressive fusion architecture with boosted module.

We also make a comparison between different architectures of the proposed HR-MPF with multi-scale progressive fusion. Because the progressive fusion basically fuses feature maps of different resolutions with a boosted module input by same resolution features, up-sampling and convolution need to precede the boosted module. As shown in Fig. 18, the three available combinations are: (a) an up-sampling, a convolution+Batch Normalization, and a boosted module followed by; (b) an up-sampling, a boosted module, and a convolution+Batch Normalization; (c) ours, a convolution+Batch

Normalization, an up-sampling, and a boosted module followed by. As shown in Table 5, change of architecture has mild effect on both segmentation and classification performances, while model parameters are the same. Quantitative results of (a) and (c) are similar, which indicates that the order of convolution and up-sampling effects the results insignificantly. However, the scores of (b) are relatively lower than (a) and (c). In the architecture of (b), an individual up-sampling set before the boosted may not guarantee an explicit and constrained fusing process of two inputs for boosted modules [27].

To verify the effectiveness of each module in the whole framework, several ablation experiments are conducted and the results are recorded in Table 6, where  $S$  represents only the segmentation network (including HR-MPF and PDM),  $C$  denotes the classification module and  $D$  stands for the discriminator. Ablation model “S+C” represents the proposed framework ablating the discriminator and adversarial loss in Eq. (6). Experimental results show that performing segmentation and classification together does not affect segmentation results very significantly. However, the discriminator we introduced improves both segmentation and classification results significantly. For instance, segmentation Prec rises from 0.9353 to 0.9427 and Acc for classification increases from 0.9650 to 0.9768. This might be benefited from the adversarial training mechanism that guides the training process of HR-MPF and PDM such that segmentation results more approaching to ground truths are generated. Hence, the accurate segmentation brings an improvement to subsequent classification results as well. Therefore, the architecture with two sub-networks and adversarial training strategy is verified effective.

## 5 Conclusion

In this paper, an effective multi-task framework is designed for pulmonary nodule segmentation and classification, which can contribute to clinical diagnosis of pulmonary nodules. Specifically, a widely applicable feature extraction network HR-MPF is proposed. This architecture attributes progressive fusion strategy to HRNet with modified boosted modules incorporated. Corresponding PDM decoding predictions from HR-MPF is also designed which recovers the final pixel-wise segmentation predictions in progressive fusion manner. Then, a feature map from PDM is fed into the classification module to determine the benign and malignant of pulmonary nodules. Joint training of pulmonary nodule segmentation and classification is realized with discriminator established and reasonably designed multi-task loss function. Specifically, a boundary consistency constraint is designed in the segmentation loss which further enhances boundary segmentation crucial in pulmonary nodule segmentation tasks. In comparison with latest segmentation and classification methods individually, the proposed method shows superior results in segmentation and competitive classification behavior in general.

### Abbreviations

GAN: Generative adversarial network; HR-MPF: High-resolution network with multi-scale progressive fusion; PDM: Progressive decoding module; HRNet: High-resolution network; CT: Computed tomography; CAD: Computer-aided diagnosis; Acc: Accuracy; DSC: Dice similarity coefficient; MIoU: Mean intersection over union; Prec: Precision; SE: Sensitivity; SP: Specificity; TP: True positive; TN: True negative; FP: False positive; FN: False negative; AUC: Area under curve.

### Acknowledgements

No additional acknowledgements.

**Authors' contributions**

LZ and PYW came up with the idea of the work and implemented the proposed method. LZ and YY performed the experiments and drafted the manuscript. HQZ, SYY took part in writing the manuscript. All authors read and approved the final manuscript.

**Authors information**

**Ling Zhu** received the B.S. degree of electronic information science from Jiangsu University of Science and Technology, Jiangsu, China, in 2019. She is currently pursuing the M.S. degree of information and communication engineering in East China University of Science and Technology, Shanghai, China. Her research interests include medical image processing, deep learning and object detection and segmentation.

**Hongqing Zhu** received the Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 2000. From 2003 to 2005, she was a Post-Doctoral Fellow with the Department of Biology and Medical Engineering, Southeast University, Nanjing, China. She is currently a professor with the East China University of Science and Technology, Shanghai. Her current research interests include medical image processing, deep learning, computer vision, and pattern recognition. She is a member of IEEE and IEICE.

**Suyi Yang** is currently pursuing the B.Sc. degree with the Department of Mathematics, Natural, Mathematical & Engineering Sciences, King's College London. Her interests include mathematical modeling and mathematical problems in image processing, especially partial differential equations.

**Pengyu Wang** is currently pursuing the Ph.D. degree with the Department of Electronics and Communication Engineering, East China University of Science and Technology, Shanghai, China. In 2015 and 2018, he received the B.S. degree in automation and M.S. degree in control engineering from Hebei University in Baoding. His research interests include image processing, deep learning, computer vision, and pattern recognition.

**Yang Yu** is currently working towards his Ph.D. in East China University of Science and Technology, Shanghai, China, and has received B.S. degree in electronic Information Science from Jiangsu University of Science and Technology in 2018. His research domains include medical image processing, deep learning, and image segmentation.

**Funding**

This work was supported by the National Nature Science Foundation of China under Grant 61872143.

**Availability of data and materials**

The dataset and materials can be downloaded from the provided hyperlinks.

**Declarations****Consent for publication**

Approved.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>School of Information Science & Engineering, East China University of Science and Technology, Mei Long Road, 200237 Shanghai, China. <sup>2</sup>Department of Mathematics, Natural, Mathematical & Engineering Sciences, King's College London, WC2R 2LS London, United Kingdom.

Received: 3 May 2021 Accepted: 25 October 2021

Published online: 13 November 2021

**References**

1. S. Blandin Knight, P.A. Crosbie, H. Balata, J. Chudziak, T. Hussell, C. Dive, Progress and prospects of early detection in lung cancer. *Open Biol.* **7**(9), 170070 (2017)
2. J. Zheng, D. Yang, Y. Zhu, W. Gu, B. Zheng, C. Bai, L. Zhao, H. Shi, J. Hu, S. Lu et al., Pulmonary nodule risk classification in adenocarcinoma from CT images using deep CNN with scale transfer module. *IET Image Process.* **14**(8), 1481–1489 (2020)
3. K. Sun, Y. Zhao, B. Jiang, T. Cheng, B. Xiao, D. Liu, Y. Mu, X. Wang, W. Liu, J. Wang, High-resolution representations for labeling pixels and regions. *arXiv preprint arXiv:1904.04514* (2019)
4. S. Baghersalimi, B. Bozorgtabar, P. Schmid-Saugeon, H.K. Ekenel, J.-P. Thiran, DermoNet: Densely linked convolutional neural network for efficient skin lesion segmentation. *EURASIP J. Image Video Process.* **2019**(1), 1–10 (2019)
5. R. Miao, Y. Gao, L. Ge, Z. Jiang, J. Zhang, Online defect recognition of narrow overlap weld based on two-stage recognition model combining continuous wavelet transform and convolutional neural network. *Comput. Ind.* **112**, 103115 (2019)
6. S. Chen, Y. Wang, Pulmonary nodule segmentation in computed tomography with an encoder-decoder architecture. In: *International Conference on Information Technology in Medicine and Education (ITME)*, pp. 157–162 (2019)
7. N.V. Keetha, C.S.R. Annavarapu, et al., U-Det: A modified U-Net architecture with bidirectional feature network for lung nodule segmentation. *arXiv preprint arXiv:2003.09293* (2020)
8. G. Pezzano, V.R. Ripoll, P. Radeva, Cole-CNN: Context-learning convolutional neural network with adaptive loss function for lung nodule segmentation. *Comput. Meth. Prog. Bio.* **198**, 105792 (2021)
9. H. Liu, H. Cao, E. Song, G. Ma, X. Xu, R. Jin, Y. Jin, C.-C. Hung, A cascaded dual-pathway residual network for lung nodule segmentation in CT images. *Phys. Medica* **63**, 112–121 (2019)

10. H. Yan, H. Lu, M. Ye, K. Yan, Y. Xu, Q. Jin, Improved Mask R-CNN for lung nodule segmentation. In: International Conference on Information Technology in Medicine and Education (ITME), pp. 137–141 (2019)
11. W. Wang, R. Feng, J. Chen, Y. Lu, T. Chen, H. Yu, D.Z. Chen, J. Wu, Nodule-plus R-CNN and deep self-paced active learning for 3D instance segmentation of pulmonary nodules. *IEEE Access* **7**, 128796–128805 (2019)
12. Y. Sun, J. Tang, W. Lei, D. He, 3D segmentation of pulmonary nodules based on multi-view and semi-supervised. *IEEE Access* **8**, 26457–26467 (2020)
13. X. Dong, S. Xu, Y. Liu, A. Wang, M.I. Saripan, L. Li, X. Zhang, L. Lu, Multi-view secondary input collaborative deep learning for lung nodule 3D segmentation. *Cancer Imaging* **20**(1), 1–13 (2020)
14. I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks. arXiv preprint [arXiv:1406.2661](https://arxiv.org/abs/1406.2661) (2014)
15. Y. Qin, H. Zheng, X. Huang, J. Yang, Y.-M. Zhu, Pulmonary nodule segmentation with CT sample synthesis using adversarial networks. *Med. Phys.* **46**(3), 1218–1229 (2019)
16. Y. Onishi, A. Teramoto, M. Tsujimoto, T. Tsukamoto, K. Saito, H. Toyama, K. Imaizumi, H. Fujita, Investigation of pulmonary nodule classification using multi-scale residual network enhanced with 3DGAN-synthesized volumes. *Radiol. Phys. Technol.* **13**(2), 160–169 (2020)
17. H. Shi, J. Lu, Q. Zhou, A novel data augmentation method using style-based GAN for robust pulmonary nodule segmentation. In: 2020 Chinese Control and Decision Conference (CCDC), pp. 2486–2491 (2020)
18. D. Nie, Y. Gao, L. Wang, D. Shen, ASDNet: Attention based semi-supervised deep networks for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 370–378 (2018)
19. C. Decourt, L. Duong, Semi-supervised generative adversarial networks for the segmentation of the left ventricle in pediatric MRI. *Comput. Biol. Med.* **123**, 103884 (2020)
20. Z. Han, B. Wei, A. Mercado, S. Leung, S. Li, Spine-GAN: Semantic segmentation of multiple spinal structures. *Med. Image Anal.* **50**, 23–35 (2018)
21. Y. Sun, C. Zhou, Y. Fu, X. Xue, Parasitic GAN for semi-supervised brain tumor segmentation. In: 2019 IEEE International Conference on Image Processing (ICIP), pp. 1535–1539 (2019)
22. Y. Wang, H. Zhang, K.J. Chae, Y. Choi, G.Y. Jin, S.-B. Ko, Novel convolutional neural network architecture for improved pulmonary nodule classification on computed tomography. *Multidimens. Syst. Signal Process.* **31**(3), 1163–1183 (2020)
23. Y. Zhang, J. Zhang, L. Zhao, X. Wei, Q. Zhang, Classification of benign and malignant pulmonary nodules based on deep learning. In: International Conference on Information Science and Control Engineering (ICISCE), pp. 156–160 (2018)
24. Y. Onishi, A. Teramoto, M. Tsujimoto, T. Tsukamoto, K. Saito, H. Toyama, K. Imaizumi, H. Fujita, Multiplanar analysis for pulmonary nodule classification in CT images using deep convolutional neural network and generative adversarial networks. *Int. J. Comput. Assist. Radiol. Surg.* **15**(1), 173–178 (2020)
25. C. Tong, B. Liang, Q. Su, M. Yu, J. Hu, A.K. Bashir, Z. Zheng, Pulmonary nodule classification based on heterogeneous features learning. *IEEE J. Sel. Areas Commun.* **39**(2), 574–581 (2020)
26. Jiang, K., Wang, Z., Yi, P., Chen, C., Huang, B., Luo, Y., Ma, J., Jiang, J.: Multi-scale progressive fusion network for single image deraining. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 8346–8355 (2020)
27. H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, M.-H. Yang, Multi-scale boosted dehazing network with dense feature fusion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2157–2167 (2020)
28. K. He, X. Zhang, S. Ren, J. Sun, Identity mappings in deep residual networks. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 630–645 (2016)
29. K. Sun, B. Xiao, D. Liu, J. Wang, Deep high-resolution representation learning for human pose estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5693–5703 (2019)
30. V. Badrinarayanan, A. Kendall, R. Cipolla, SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017)
31. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818 (2018)
32. Z. Tian, T. He, C. Shen, Y. Yan, Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3126–3135 (2019)
33. Z. Zhao, Q. Sun, H. Yang, H. Qiao, Z. Wang, D.O. Wu, Compression artifacts reduction by improved generative adversarial networks. *EURASIP J. Image Video Process.* **2019**(1), 1–7 (2019)
34. Y. Yuan, J. Xie, X. Chen, J. Wang, Segfix: Model-agnostic boundary refinement for segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 489–506 (2020)
35. M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks. In: International Conference on Machine Learning (ICML), pp. 214–223 (2017)
36. S.G. Armato III, G. McLennan, L. Bidaut, M.F. McNitt-Gray, C.R. Meyer, A.P. Reeves, B. Zhao, D.R. Aberle, C.I. Henschke, E.A. Hoffman et al., The lung image database consortium (LIDC) and image database resource initiative (IDRI): A completed reference database of lung nodules on CT scans. *Med. Phys.* **38**(2), 915–931 (2011)
37. J. Zhuang, T. Tang, S. Tatikonda, N. Dvornik, Y. Ding, X. Papademetris, J.S. Duncan, Adabelief optimizer: Adapting stepsizes by the belief in observed gradients. arXiv preprint [arXiv:2010.07468](https://arxiv.org/abs/2010.07468) (2020)
38. A. Paszke, A. Chaurasia, S. Kim, E. Culurciello, ENet: A deep neural network architecture for real-time semantic segmentation. arXiv preprint [arXiv:1606.02147](https://arxiv.org/abs/1606.02147) (2016)
39. H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2881–2890 (2017)

40. Z. Zhou, M.M.R. Siddiquee, N. Tajbakhsh, 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*, pp. 3–11 (2018)
41. H. Li, P. Xiong, H. Fan, J. Sun, DFANet: Deep feature aggregation for real-time semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9522–9531 (2019)
42. R.P. Poudel, S. Liwicki, R. Cipolla, Fast-SCNN: Fast semantic segmentation network. arXiv preprint [arXiv:1902.04502](https://arxiv.org/abs/1902.04502) (2019)
43. P. Hu, F. Perazzi, F.C. Heilbron, O. Wang, Z. Lin, K. Saenko, S. Sclaroff, Real-time semantic segmentation with fast attention. *IEEE Robot. Autom. Lett.* **6**(1), 263–270 (2020)
44. Q. Hou, L. Zhang, M.-M. Cheng, J. Feng, Strip pooling: Rethinking spatial pooling for scene parsing. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4003–4012 (2020)
45. O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241 (2015)
46. W. Zuo, F. Zhou, Z. Li, L. Wang, Multi-resolution CNN and knowledge transfer for candidate classification in lung nodule detection. *IEEE Access* **7**, 32510–32521 (2019)
47. G. Zheng, G. Han, N.Q. Soomro, An inception module CNN classifiers fusion method on pulmonary nodule diagnosis by signs. *Tsinghua Sci. Technol.* **25**(3), 368–383 (2019)
48. S. Akila Agnes, J. Anitha, Automatic 2D lung nodule patch classification using deep neural networks. In: *International Conference on Inventive Systems and Control (ICISC)*, pp. 500–504 (2020)

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---