IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 July 2024, accepted 8 August 2024, date of publication 12 August 2024, date of current version 23 August 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3442380*

RESEARCH ARTICLE

# Enhancing Pulmonary Nodule Detection Rate Using 3D Convolutional Neural Networks With Optical Flow Frame Insertion Technique

**WEI ZHANG[1], ABDERRAHMANE SALMI[2], FENG JIANG[2], AND CHI-FU YANG[1], (Member, IEEE)**
[1]School of Mechanical and Electrical Engineering, Harbin Institute of Technology, Harbin 150001, China
[2]School of Computer Science, Harbin Institute of Technology, Harbin 150001, China

Corresponding author: Wei Zhang (19B908079@stu.hit.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0806802 and Grant 2018YFC0832105.

**ABSTRACT** The lung nodule detection technology plays a vital role in the diagnosis and treatment of early lung cancer. Deep learning is currently one of the main technologies applied in computer vision related fields. Therefore, this topic is to combine deep learning and lung nodule detection technology. Different from some research methods, we improved from two-dimensional space to three- dimensional space and applied it to lung nodule detection. After statistically analyzing the lung CT data, we find that the lung CT data has inconsistent scales on the vertical axis. Aiming at the problem of lung CT data, this article is based on the deep learning method. First, the lung interpolation network based on voxel flow is used to achieve the same scale, and then the lung nodule detection network based on three-dimensional convolution is used to complete the detection of lung nodules. The entire network combines U- Net-like and RPN-like network structures. Through data slice input, it avoids the limitation of the display memory of the computing platform. The network structure also introduces prior knowledge of the coordinate information of the input slices to improve the classification accuracy of lung nodules. The experimental results show that the lung slice data of consistent scale is achieved through interpolation, and then through the three-dimensional convolution of the lung nodule detection network, the state- of-the-art detection effect is achieved. Because of the introduction of the interpolation network, the time- consuming has increased. Of course, the overall speed stillmeets the actual use value.

**INDEX TERMS** Lung, pattern recognition and classification, computer-aided detection and diagnosis, X- ray imaging, computed tomography.

## I. INTRODUCTION

The incidence and mortality of lung cancer are very high, which greatly affects human health [1].Compared with other cancers, the biological characteristics of lung cancer are complex. Many patients have no obvious special symptoms in the initial stage of the disease. Most of the clinical manifestations that can attract the attention of patients are advanced lung cancer, and the best treatment opportunity has been lost.The American Cancer Society's research report shows that the current survival rate of lung cancer patients within 5 years is less than 17%. If early detection and treatment can be obtained, the survival rate can be increased to more than 55% [2]. Therefore, the early detection and diagnosis of lung cancer is of vital importance to patients and can effectively improve the survival rate of patients. In clinical medicine, the early lesions of lung cancer are mainly manifested by the appearance of lung nodules. If lung nodules can be detected earlier and more accurately, it can help patients diagnose lung cancer earlier. Therefore, the

The associate editor coordinating the review of this manuscript and approving it for publication was Cristian A. Linte.

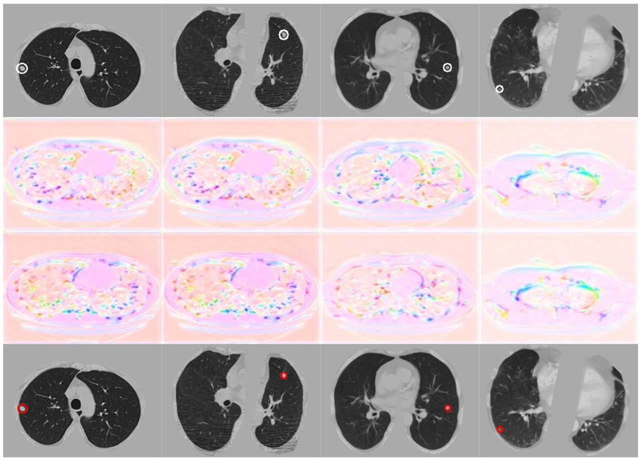


FIGURE 1. Lung nodule detection examples with proposed method. This paper innovatively starts with the problem that the inconsistency of the vertical axis scale of lung CT data leads to the decrease in the accuracy of lung nodule detection, and proposes an end-to-end framework that combines the interpolation network and the detection network. Figure 1 shows the results of four experiments. The first row is the label data (white circle), the second row and the third row are the forward voxel flow and the backward voxel flow during the interpolation process, and the fourth row is the test result.

correct detection of lung nodules has a great effect and significance for the diagnosis and treatment of early lung cancer. In the diagnosis of lung cancer, computed tomography (CT) images are a fairly common visualization tool [3]. CT images of the lungs will visualize all tissues based on the absorption of X-rays. The lung lesions are collectively called lung nodules. Nodules usually have the same suction as normal tissueIt is harvested horizontally, but has its unique shape: the bronchi and blood vessels are continuous pipe systems with thick roots and thin branches, while the nodules are usually spherical and in an isolated form. However, the diagnosis of nodules usually requires experienced doctors to spend a long time to thoroughly examine the patient. This is because some nodules are very small and difficult to find.In addition, the shape of the nodules is also different, and the risk of cancer will vary with the shape of the nodules. Doctors can estimate the benign and malignant nodules by their morphology, but the accuracy of diagnosis depends very much on the doctor's experience.In most cases, doctors need to look up images and identify lung nodules with the naked eye, which is more difficult, and different doctors may give different predictions [4].

Computer aided diagnosis (CAD) is a very important solution to this problemSuitable method, because the computer vision model can quickly scan every part of the CT image of the same quality for analysis, and will not be affected by fatigue and emotion. Through the computer's automatic processing of CT images, it is possible to frame the spatial position of the suspected lung nodules on the CT image and provide the doctors for reference.The auxiliary system provides detection results, reduces the work intensity of doctors who originally need to consult a large number of image sequences, improves the efficiency of image reading, and also avoids errors that may be caused by long-term human eye fatigue. The importance of CAD has also prompted more and more researchers to study lung nodule detection algorithms and propose advanced algorithms to improve the accuracy of lung nodule detection. The latest advances in deep learning enable computer vision models to help doctors diagnose various diseases, and in some cases, models have shown greater competitiveness than doctors. However, due to the three-dimensional complex structure of lung CT data and the diversity of lung nodule shapes and sizes, and the rich variety of similar structures, the correct detection of lung nodules is still a challenge.

## II. RELATED WORK

How to assist in the diagnosis of lung cancer through medical image processing has always been a concern of academic circles. In recent years, many research institutions and scholars have successively carried out research on the detection technology of lung nodules, and have done a lot of outstanding work in this direction. Traditional lung nodule detection methods mainly set features and descriptors manually [5]. In recent years, due to the very successful application of deep learning technology in the field of computer vision and image processing, as well as the labeling and disclosure of a large number of lung CT data, many researchers have also used deep learning to do research work on lung nodule detection.

Drawing on the method of natural image target detection, Xie et al. [6] believed that using the original three-dimensional lung CT as input would consume a lot of computational cost, and proposed to directly merge three adjacent lung CT slices into one two-dimensional image. Then use the improved Faster R-CNN network structure to perform target detection on the two-dimensional image. They use the VGG [7] network as the feature extraction network, and use the deconvolution structure in the RPN structure to generate candidate regions. They added a deconvolution layer to the RPN structure, which made the network have more fine-grained features and obtained better detection results than the original Faster R-CNN network structure. The disadvantage of this method is that for target detection on a two-dimensional image, only the spatial information on the lung slice is used. The lung CT and the lung nodules themselves have three-dimensional information, so the convolution on the two-dimensional image cannot capture some other characteristics of the lung nodules in the three- dimensional space, such as size and shape.

Some work is not based on Anchor-Based. Khosravan and Bagci proposed S4ND [8]. They used the dense convolutional network DenseNet [9] as the main structure to achieve feature reuse through dense connections between layers. S4ND merges 8 lung CT slices as input. After passing through a full convolutional network, it outputs a probability density map of lung nodules corresponding to the spatial region of the original image. Different from the candidate area network structure, because the input of the network is the CT image

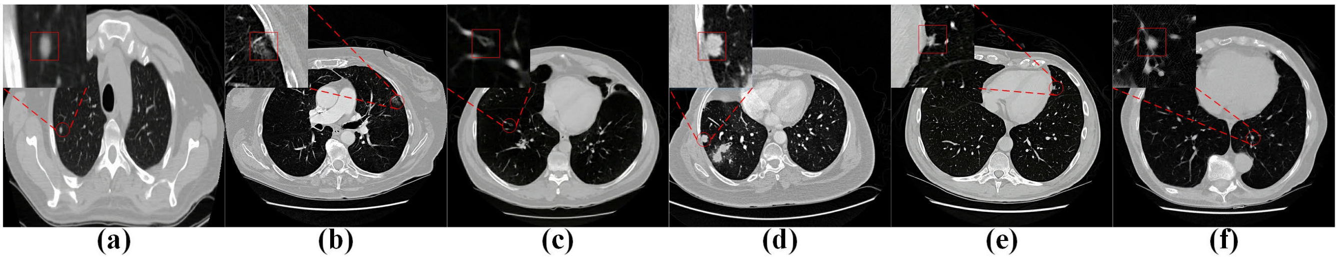


**FIGURE 2.** Lung nodule detection examples with proposed method. This paper innovatively starts with the problem that the inconsistency of the vertical axis scale of lung CT data leads to the decrease in the accuracy of lung nodule detection, and proposes an end-to-end framework that combines the interpolation network and the detection network. Figure 1 shows the results of four experiments. The first row is the label data (white circle), the second row and the third row are the forward voxel flow and the backward voxel flow during the interpolation process, and the fourth row is the test result.

of the entire lung, the global texture information of the CT image can be captured. Similarly, Dou et al. [10] also input lung CT images into a fully convolutional network, and the network outputs a probability density map of lung nodules, corresponding to the grid of the original image space. For regions with high probability, they are then input into the convolutional network to subdivide their categories and locations through block dicing. The network also adopts the method of cutting out difficult samples. Compared with other samples, the difficult samples will get higher loss scores through network prediction. Therefore, the difficult samples can be extracted online according to this judgment method in the forward network transmission process during each training process. Samples are trained for many times to improve the ability to detect difficult samples.

The morphology of the lung nodule itself in the three-dimensional space is quite different, and a single lung nodule is also different from various perspectives. If these feature information can be used, the detection of lung nodules can be improved, so there is a lot of work to directly extract features using three-dimensional convolutional networks on three- dimensional lung CT images. The entire network uses a three- dimensional convolution-like U-Net structure, which can capture the multi-scale information of the image, which is conducive to the detection of nodules of different sizes. The final output of the network uses an RPN-like structure to directly output the location information and probability of the lung nodules. Similarly, Tang et al. [11] used a residual network structure based on the U-Net-like [12]structure to avoid the disappearance of the gradient caused by the network being too deep. Zhu et al. proposed DeepLung [13], based on the U-Net- like structure, using a simple, efficient and highly modular two- way network DPN [14] network structure. The dual path network (DPN) has both the advantages of ResNet [14] to prevent the gradient from disappearing and DenseNet [15] of feature reuse, and it obtains a better detection effect than the previous residual network with fewer parameters. After passing the detection network, the gradient boosting machine (GBM) is also used to reduce false positives of lung nodules. However, the three-dimensional convolution method also has shortcomings. Since the input of the network is a three-dimensional structure lung CT image, the three- dimensional convolution network is used to extract the features, which has a large demand limit on the size of the computing platform's video memory, making it often impossible Realize high-resolution lung CT image input.

In order to extract more effective features of lung nodules, many works have also improved the backbone feature extraction network. Tang et al. proposed NoduleNet [16], [17], [18]. The three tasks of lung nodule detection, false positive reduction, and nodule semantic segmentation share the same feature extraction network. The task is trained in an end-to-end manner. Different from processing each task independently, the need to train multiple deep neural networks is a waste of resources and time. Multi-task training also enables the transfer of information flow between each other, allowing the network to learn a more intrinsic feature representation, that is, the segmentation mask of the nodule can provide powerful guidance for the neural network to learn to distinguish and detect features, thereby improving the nodule Detection performance. Liu et al. proposed 3DFPN-HS [19], borrowing from the natural image target detection feature pyramid networks (FPN) structure, changing two-dimensional convolution to three-dimensional convolution, and using dense connections for feature fusion of different layers. Pooling and deconvolution between layers of different sizes achieve size uniformity. Through the multi-scale structure, the detection ability of small nodules and lung nodules with relatively large scale distribution is improved. In addition, the article pointed out that some normal tissues whose appearance and appearance are similar to the nodules in lung CT images are often mistakenly detected as nodules by the network, leading to high false positives. If the real nodule has a round shape, and the false positive tissue has a straight line shape, The calculated features are input into the auxiliary network to learn to judge true and false lung nodules. Liu et al. [19] proposed a method for self-supervised learning of lung nodule orientation features. The lung CT orientation characteristics learned by this unsupervised method improve the robustness of the lung nodule network algorithm to the data collected by different CT scanning instruments. This study [20], [21], [22] used attention mechanisms to improve the resolution of stereoscopic endoscopic images and employed saliency-aware methods to focus on the most informative regions in endoscopic images. These methods more effectively increased the resolution of these regions,

**TABLE 1.** Comparative analysis of lung nodule detection methods.

| Study | Techniques | Datase | Evaluation Metrics | Weaknesses |
|---|---|---|---|---|
| Ding[6] | Faster R-CNN, 2D CT slices | LIDC-IDRI | Sensitivity, Specificity | Limited to 2D, misses 3D context |
| Setio[7] | Multi-view Convolutional Networks | LUNA16 | Sensitivity, Specificity | Limited by view aggregation complexity |
| Bagci[8] | S4ND, DenseNet, 2D CT slices | Private dataset | AUC, FROC | Requires large labeled datasets |
| Huang[9] | Mask R-CNN, 3D CT slices | LIDC-IDRI | Precision, Recall | High computational cost for 3D data |
| Dou[10] | Fully Convolutional Network | LUNA16 | Dice Score, FROC | High computational cost |
| Tang[11] | U-Net, 3D Convolution | LIDC-IDRI | Precision, Recall | High GPU resources |
| Brox [12] | 3D U-Net, Residual Connections | LIDC-IDRI | Dice Score, Sensitivity | Computationally intensive |
| Zhu[13] | DPN, 3D Convolution | LUNA16 | Sensitivity, Specificity | High memory usage for 3D CT input |
| Chen [14] | 3D CNN, Hybrid Loss Functions | Private dataset | AUC, Precision | Requires extensive tuning of loss functions |
| Tang[16] | Multi-task learning | Private dataset | AUC, FROC | Complexity in multi-task training |
| Liu [18] | 3DFPN-HS, FPN with 3D Convolution | LUNA16 | FROC | High false positives due to similar tissues |

thereby improving the overall image quality and ensuring that key edges and boundaries are maintained, which is essential for accurate medical diagnosis. This study [23] explored the automatic segmentation of CT images for body composition analysis. The methods used here can be directly used to segment lung CT images to isolate regions of interest (e.g., lung nodules) before further processing and analysis.

Compared with the traditional lung nodule detection method, the lung nodule detection method based on deep network has higher performance.However, these solutions still have the following disadvantages:(1) There is a lot of redundant information in the original CT data graph, which increases the search space of the model, and the model training cannot reach the optimal level. (2) Although the effect of the scheme based on the three-dimensional convolutional neural network is better than that of the two-dimensional scheme, it relies on a large number of GPUs, which is difficult to meet under normal circumstances, which greatly limits the application of the algorithm. (3) The data used in the existing schemes are inconsistent in scale on the vertical axis, which has a significant impact on the detection effect. Table1 provides a comparative analysis of important previous studies, summarizing their techniques and shortcomings. Despite the advances in lung nodule detection, existing methods still face several limitations: firstly, redundant information, raw CT data contains a large amount of redundant information, which increases the search space and hinders optimal model training. Secondly, resource requirements. 3D convolutional neural network-based methods, although more efficient, require large GPU resources, which limits their usefulness. 2. There is also the problem of scale inconsistency, where inconsistent vertical axis scales in CT data can seriously affect detection accuracy.

In order to overcome the shortcomings of the above methods, we propose a lung nodule detection network based on three- dimensional convolution of voxel flow interpolation. As shown in Figure 8, it is composed of three modules: a lung interpolation network module based on voxel flow (LIVF), a lung data segmentation processing module (LDCP), and a lung nodule detection module based on a three-dimensional convolutional neural network (LNDT). In the LIVF module, first obtain the voxel flow between adjacent slices through a multi-layer network, and then obtain the interpolated lung slices through the forward warping (wrap) operation according to the extracted voxel flow and the required interpolation coefficients. The vertical axis resolution of lung CT data is consistent. In the LDCP module, first extract the three-dimensional slices on the CT image of the entire lung, and then input each three- dimensional slice separately into the detection network. Finally, after the LNDT module, it integrates U-Net-like and region proposal network (RPN) to achieve better lung nodule detection results. The main contributions are summarized as follows:

(1) We innovatively started with the inconsistency of the vertical axis scale of lung CT data, and proposed a lung interpolation network based on voxel flow, which can achieve consistent resolution of the vertical axis of lung CT data.

(2) This paper proposes a lung nodule detection network based on three-dimensional convolution, which combines U-Net-like and RPN-like network structures, and combines lung CT data with consistent scale as input. The experimental results show that a better lung is obtained. Nodule detection effect.

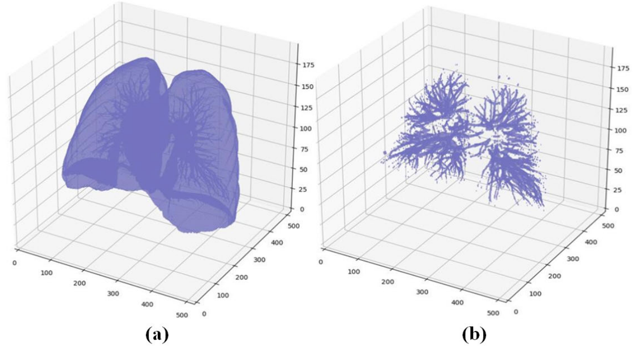
(3) The lung CT interpolation module proposed in this paper realizes the consistent resolution of the vertical axis of lung CT data. It can not only be applied to lung nodule detection, but also can be inspired to be applied to other medical imaging tasks, such as lungs CT semantic segmentation, pathological detection of brain and heart.

## III. LUNG INTERPOLATION NETWORK BASED ON VOXEL FLOW

The lung CT data set is a three-dimensional data, including a series of multiple horizontal axial slices of the chest cavity. Before input to the lung detection network, the original lung CT data needs to be preprocessed to reduce the influence of irrelevant noise on the network detection. Reduce the search space of the model to make the model train the best. This section first introduces the lung CT data set and the preprocessing method of the data, then analyzes the lung CT data, and finally proposes a lung CT interpolation network based on voxel flow, which solves the lung CT data The impact of inconsistent vertical axis resolution scale on the detection of lung nodules.

### A. LUNG CT DATA SET AND DATA PREPROCESSING

In order to improve the early detection of lung cancer, the American Cancer Institute initiated the collection of the LIDC-IDRI data set. This data set is a large chest medical image annotation data set, composed of chest medical image files and corresponding medical diagnosis results, including 1018 research examples. The lung CT data is displayed in the three dimensional space as shown in Figure 3(a), and the bronchus and lung nodules are shown in Figure 3(b). Each CT image consists of a series of multiple axial slices of the thoracic cavity, usually containing hundreds of slices. Each CT image includes 512 × 512 pixels. The pixel values are all integers.



(a)
(b)

**FIGURE 3.** CT three-dimensional space display of lung.

The lung CT data used for training comes from different hospitals and different scanning equipment. The scan interval, scan initial coordinate points and other parameters are different for each CT case. When the data is acquired, it is inevitable that the equipment will produce errors and interference. In addition, lung CT data has the following problems compared to natural image data: (1) The data occupies a large space; (2) the features are not obvious; (3) the scale of lung nodules is small. For deep learning models, the quality of the input training data directly affects the fit of the final model. In addition, the amount of lung CT data is large. If it is directly input to the detection network structure without preprocessing, it will cause errors in system parameters. In order to make the trained model reach the best, the detection algorithm is reliable and stable, and consider the consumption of hardware settings during model training.

In order to optimize the trained model, the detection algorithm is reliable and stable, and consider the consumption of hardware settings during model training, it is necessary to preprocess and standardize lung CT data so that lung CT data reflects lung nodules as much as possible The information needed for diagnosis reduces the search space for lung nodule detection. The preprocessing process of lung CT data is shown in Figure 4. Therefore, the lung CT data needs to be preprocessed and standardized, so that the lung CT data reflects the information needed for lung nodule diagnosis as much as possible, and reduces the search space for lung nodule detection. The preprocessing process of lung CT data is as follows: (1) Data reading. Process CT data by using functions corresponding to the SimpleITK library. The main information obtained from the read data includes: the original three-dimensional image data of lung CT in the form of an array, whether the image is flipped, the initial coordinate information in the world coordinate system corresponding to the image, and the scan interval of the image on the x, y, and z axes. The read slice image of the original lung CT data is displayed in the heat map space as shown in Figure 5 (a), and the corresponding gray space display is shown in Figure 5(b).

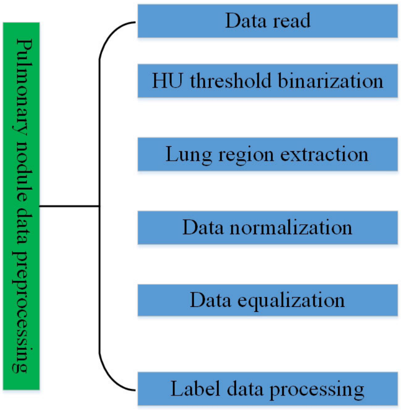
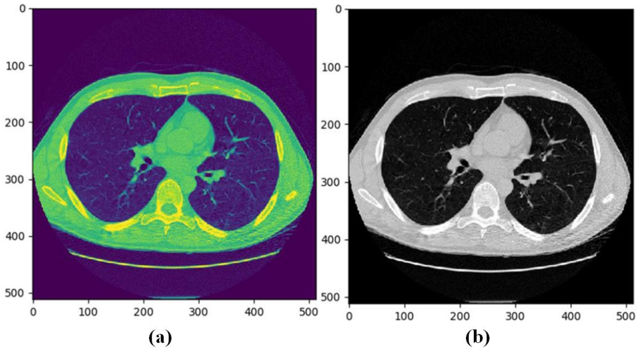


**FIGURE 4.** The preprocessing process of lung CT data.

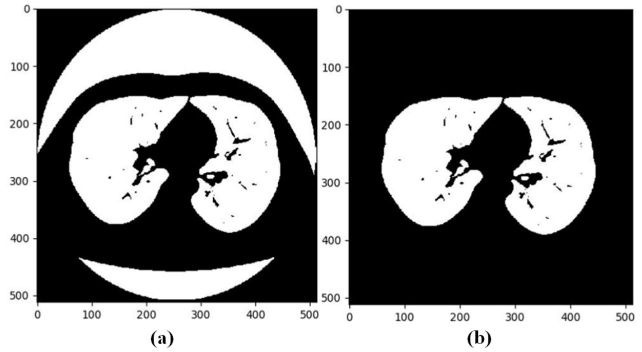


(a)
(b)

**FIGURE 5.** Lung CT data reading.

(2) Binarization of HU threshold. For lung CT image data, not only the lungs are included, but also some other tissues, some of which have a spherical shape similar to nodules in appearance. For data without a lung area mask, to avoid the influence of these tissues on the detection of lung nodules, it is necessary to extract the lung area mask. The image can be segmented by setting a reasonable HU threshold. The CT image of Figure 6(a) is filtered with a Gaussian filter with a standard deviation of 1, and then the image is binarized with a threshold of HU = −320, and the result of Figure 6(b) is obtained.

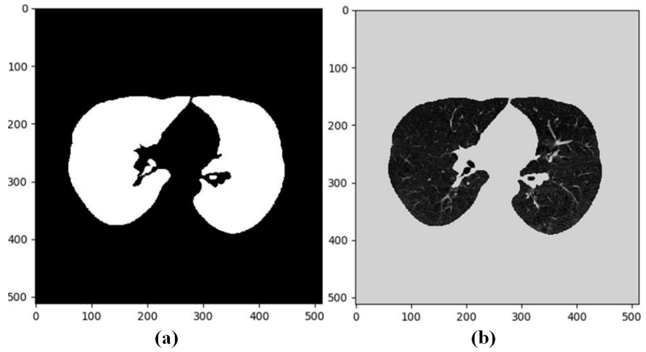
(3) Extraction of lung area. After threshold binarization, a lung and other air tissues will be obtained. Compared with these other tissues, the lung tissues generally occupy the center of the image. By calculating the minimum distance from all connected domain tissues on each lung slice to the



(a) (b)

**FIGURE 6.** CT data HU threshold binarization.

image center and their area, the tissues with an area greater than 6000 mm and the average minimum distance greater than 62 mm are removed, and the rest is the lung area, As shown in Figure 6(b). For the holes in the lung area in Figure 5(b), the connected domains are filled to fill the holes, and the final lung region mask Figure 6(a) is obtained. Multiply the lung area mask Figure 7(a) and the original image Figure 5(b) to obtain the result of the lung area extraction image Figure 7(b).



(a) (b)

**FIGURE 7.** CT data lung region extraction.

(4) Data normalization. For the obtained lung region image, the CT value data range is large, and data normalization is required. First, we limit the original data HU value to the interval [−1200, 600], and then linearly transform it to the interval [0, 255]. Fill the area outside the lung area mask with a filling value of 170. The original data is image, the normalized data is new_image, the limit assigns the maximum value max_ bound = 600, and the limit assigns the minimum value min_bound = −1200, the specific formula is as follows:

$$\text{image} = \begin{cases} \text{max\_bound} & \text{image} > \text{max\_bound} \\ \text{min\_bound} & \text{image} < \text{min\_bound} \\ \text{image} & \text{others} \end{cases} \tag{1}$$

$$\text{new\_image} = \frac{\text{image}}{\text{max\_bound} - \text{min\_bound}} \times 255 \tag{2}$$

(5) Data equalization. The size distribution of lung nodules in the data set is not balanced, and the number of small nodules is much greater than the number of large nodules. But in medical diagnosis, the correct detection of large nodules is very important. Therefore, the sampling frequency of large nodules has been increased in the training set. Specifically, for nodules between 3 mm and 10 mm, keep one copy; for nodules from 10 mm to 20 mm, the sampling frequency is 3 times; for nodules larger than 20 mm, the sampling frequency is 5 times.

(6) Label data processing.The data label information corresponds to the original image data. After the lung CT image data is processed, the label information also needs to be processed accordingly. For negative sample data, the label data is set to [0, 0, 0, 0, 0]; for the positive sample data, the processing process is the same as the original image data preprocessing method, including the change of corresponding coordinate system, the change of the resolution, etc. Finally, the label data input into the training model are: whether there are nodules, the correction amount of the nodule center coordinate, the correction amount of the diameter of the nodule.

### B. ALGORITHM DESIGN AND NETWORK STRUCTURE

In recent years, research work has shown that deep convolutional networks can capture the correlation between adjacent frames of video, and realize interpolation between adjacent frames of video [24], [25], [26], [27]. The network can obtain the corresponding interpolation frame $I_t$ by the inputting images $I_0$, $I_1$ and time coefficient $t \in (0, 1)$. A straightforward and concise method is to directly output the RGB pixels of the interpolation frame $I_t$ by training the convolutional network, but this requires the network to learn the internal connection and the appearance of the two frames of images. This method is difficult to generate high-quality interpolation. frame. The method based on optical flow is a commonly used method of video interpolation [28], [29], [30], optical flow is instantaneous velocity of the pixels corresponding to the moving object on the imaging plane. It is the change of pixels in the video sequence in the time domain, reflecting the mapping relationship between the previous frame and the current frame. Through amplification, this paper learns the voxel flow between adjacent slices of lung CT data through a convolutional network, and uses the obtained voxel flow on the vertical axis of lung CT to achieve lung CT slices through warping operations interpolation [31].

For the original lung CT, there are n slices, and the vertical axis scan interval is $k$ (mm/voxel), the actual physical length corresponding to the vertical axis of the lung CT is $(n-1) \times k$ mm. In order to achieve the same vertical axis resolution scale for lung CT data, interpolation is performed on the vertical axis of the lung CT, so that the distance between adjacent slices of the lung CT after interpolation is 1 mm. That is, $k - 1$ slice is interpolated between two adjacent slices, and the lung CT after interpolation has $(n - 1) \times k$ slices.

The lung interpolation network based on voxel flow proposed in this topic is shown in Figure 8. First, according to the vertical axis scan interval $k$, the integer value $k'$

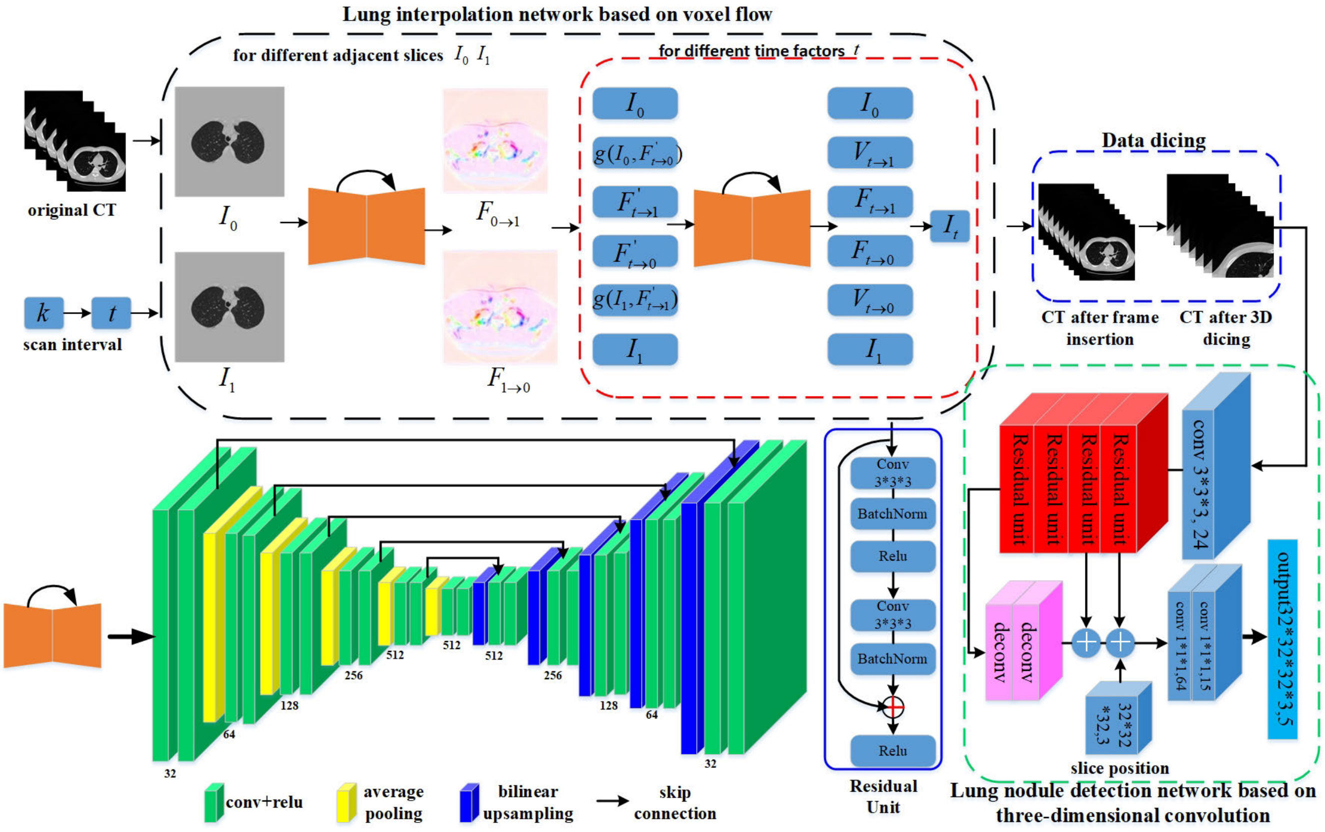


**FIGURE 8.** Structure diagram of lung nodule detection network based on three-dimensional convolutional neural network. Our proposed lung nodule detection network structure includes three parts: Lung CT interpolation network based on voxel flow, Lung CT data dicing operation and a deep detection network based on 3D convolutional neural network.

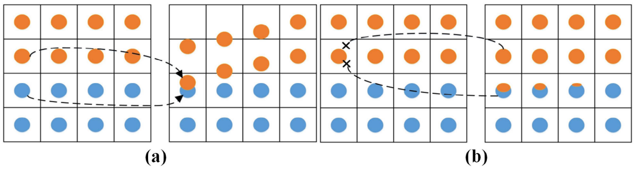


**FIGURE 9.** Warping operation.

is obtained:

$$k' = round(k) \tag{3}$$

Interpolate $k' - 1$ slices between adjacent slices to obtain the corresponding interpolation time coefficient:

$$\{t_1, t_2, t_3, \ldots\} = \frac{range(1, k')}{k'} \tag{4}$$

For each pair of different lung CT adjacent slices $I_0$ and $I_1$, they go through a convolutional network with a similar structure to U-Net [31], the network parameters are shown in Table 2. The network adopts a self-encoding structure, and realizes down-sampling through encoding and decoding to realize up-sampling. The entire structure realizes the spatial size of input and output data. The network includes feature shrinkage paths and feature amplification paths. Among them, the feature shrinking path is realized by the convolutional network and pooling operation. Firstly, the two-dimensional convolution feature extraction is performed on the input image, and then the spatial size of the feature is reduced through the pooling operation. The feature amplification path is realized by convolutional network and bilinear interpolation operation. First, two-dimensional convolution is used to further extract the previous features to high-level semantic features and global features, and then the upsampling process is completed through bilinear interpolation. The jump structure is also used in the network connection, the feature shrinkage path is used to splice the feature maps, and the spliced features are then convolved. In this way, the network will have richer global information and later features, and have more local and detailed features.The shallow network features of the section information are combined, and finally the network prediction results are output.

After passing through the U-Net network, we get the voxel flow $F_{0\to1}$ from slice $I_0$ to slice $I_1$ and the voxel flow $F_{1\to0}$from slice $I_1$ to slice $I_0$. Corresponding to different time coefficients $t$, after forward warping operation [32], the voxel flow $F'_{t\to1}$ from slice $I_t$ to slice $I_1$ and the voxel flow

**TABLE 2.** Network structure parameters.

| Network layer | Parameters | AF | Output |
|---|---|---|---|
| Input | | | 512×512 |
| 2D Conv | {3×3, 32}×2 | LeakyRelu | 512×512×32 |
| Ave pooling | 2×2 | | 256×256×32 |
| 2D Conv | {3×3, 64}×2 | LeakyRelu | 256×256×64 |
| Ave pooling | 2×2 | | 128×128×64 |
| 2D Conv | {3×3, 128}×2 | LeakyRelu | 128×128×64 |
| Ave pooling | 2×2 | | 64×64×128 |
| 2D Conv | {3×3, 256}×2 | LeakyRelu | 64×64×256 |
| Ave pooling | 2×2 | | 32×32×256 |
| 2D Conv | {3×3, 512}×2 | LeakyRelu | 32×32×512 |
| Ave pooling | 2×2 | | 16×16×512 |
| 2D Conv | {3×3, 512}×2 | LeakyRelu | 32×32×512 |
| Upsample | | | 64×64×512 |
| Connet+2D Conv | {3×3, 256}×2 | LeakyRelu | 64×64×256 |
| Upsample | | | 128×128×256 |
| Connet+2D Conv | {3×3, 128}×2 | LeakyRelu | 128×128×128 |
| Upsample | | | 256×256×128 |
| Connet+2D Conv | {3×3, 64}×2 | LeakyRelu | 256×256×64 |
| Upsample | | | 512×512×64 |
| Connet+2D Conv | {3×3, 2}×2 | LeakyRelu | 512×512×2 |
| Upsample | | | 512×512×2 |
| Connet+2D Conv | {3×3, 2}×2 | LeakyRelu | 512×512×2 |

$F'_{t\to 0}$ from slice $I_t$ to slice $I_0$are obtained by the network after forward warping operation [32]. The network does not directly fit the voxel flow $F'_{t\to 1}$ from slice $I_t$ to slice $I_1$ and the voxel flow $F'_{t\to 0}$ from slice $I_t$ to slice $I_0$ directly through the network, because the interpolation slice $I_t$ cannot be accessed during network training, the method generated directly in one step is inaccurate. We first predict the voxel flow $F_{0\to 1}$ and $F_{1\to 0}$ of the two slices, and then synthesize the middle voxel flow $F'_{t\to 0}$ and $F'_{t\to 0}$. The forward warping operation is (.,.) as shown in Figure 7(a), the specific operation process is as follows:

$$u = p - (1 + F_{0\to t}[q]) \tag{5}$$

$$b(u) = \max(0, 1 - |u_x|) \cdot \max\left(0, 1 - \left|u_y\right|\right) \tag{6}$$

$$I_t^{\Sigma}[p] = \sum_{\forall q \in I_0} b(u) \cdot I_0[q] \tag{7}$$

$$\vec{\sum}(I_0, F_{0\to t}) = I_t^{\Sigma} \tag{8}$$

$$vec\Phi(I_0, F_{0\to t}) = \frac{\vec{\sum}(I_0, F_{0\to t})}{\vec{\sum}(I_1, F_{0\to t})} \tag{9}$$

$p$ is the pixel coordinate in $I_0$, $q$ is the pixel coordinate in $I_t$, $u$ is the pixel mapping coordinate difference, and $bu()$ is the pixel warping weight. The network uses slice $I_0$, slice $I_1$, $F'_{t\to 1}$ and $F'_{t\to 0}$ to peform backward warping operation [12] $g(I_0, F'_{t\to 0})$, $g(I_1, F'_{t\to 1})$ ..$g(.,.)$ is shown in Figure 7(b), it is a backward value mapping function, realized by nonlinear interpolation, and is differentiable and differentiable. The network input the voxel flow $F'_{t\to 1}$, $F'_{t\to 0}$ which have holes, $g(I_0, F'_{t\to 0})$ and $g(I_0, F'_{t\to 1})$ into a U-Net similar structure [33], and obtain the voxel flow $F_{t\to 1}$, $F_{t\to 0}$, the visual variables $V_{t\leftarrow 0}$ and $V_{t\leftarrow 1}$ that use to complete the holes. Finally, the lung interpolation slice $I_t$ can be obtained according to the following expression:

$$I_t = \frac{1}{Z} \odot ((1-t)V_{t\leftarrow 0} \odot g(I_0, F_{t\to 0}) + tV_{t\leftarrow 1} \odot g(I_1, F_{t\to 1})) \tag{10}$$

$\odot$ is the dot product of the corresponding element. If the time coefficient $t$ is closer to the time $T = 0$, slice $I_0$ provides more contribution to generate the interpolation slice $I_t$. The principle is equivalent to $I_1$. The visual variables $V_{t\leftarrow 0}$ and $V_{t\leftarrow 1}$ take values in [0,1], which represent the weight of the interpolation slice $I_t$ generated by slices $I_0$ and $I_1$. The $Z$ variable is equal to $(1-t)V_{t\leftarrow 0} + tV_{t\leftarrow 1}$, which is a regularized variable.

### C. LOSS FUNCTION

Given the input lung slices $I_0$ and $I_1$, and the label slice $\{I_{ti}\}_{i=0}^{N}$ between the two slices, the interpolated slice output by the network is $\{\widehat{I_{ti}}\}_{i=0}^{N}$. Then the loss function of the lung interpolation network based on voxel flow is $l$, which is the linear weighted combination of voxel flow loss $l_w$, interpolation slice $L_1$ norm loss $l_r$, and perceptual loss of interpolation slice [34] Linear weighted combination of $l_p$:

$$l = \lambda_w l_w + \lambda_r l_r + \lambda_p l_p \tag{11}$$

$l_w$ is the voxel flow loss. The bidirectional voxel flow is trained in an unsupervised manner. Unlike supervised training, each pair of input adjacent slices has a corresponding bidirectional voxel flow label. In unsupervised training, the parameter update method of bidirectional voxel flow is to map the pixels of adjacent slices through backward warping operation, and supervise the mapped pixels with L1 norm constraint. Because in real lung CT data, the true mapping labels of adjacent slices cannot be obtained. The method of unsupervised training reduces the need for real mapping labels, and can still achieve high accuracy without manually labeling the corresponding data, making it more conducive to practical applications.Where $g(.,.)$ is the backward warping operation [35], including the voxel flow $F_{0\to 1}$ from $I_0$ to $I_1$ and the voxel flow $F_{1\to 0}$ from $I_1$ to $I_0$:

$$l_w = \|I_0 - g(I_1, F_{0\to 1})\|_1 + \|I_1 - g(I_0, F_{1\to 0})\|_1 \tag{12}$$

$l_r$ is the $L_1$ norm loss of the interpolation slice, which can directly reflect the difference between pixels. The $L_1$ norm is the sum of the absolute values of the data between the pointing quantities. For image generation tasks, the $L_1$ norm is used as the loss in the experiment to obtain a clearer image result, so the $L_1$ norm loss is used here in this paper:

$$l_r = \frac{1}{N} \sum_{i=1}^{N} \left\| I_{ti} \hat{-} I_{ti} \right\|_1 \tag{13}$$

$l_p$ is s the perceptual loss of the interpolation slice [36], [37], [38]. Perceptual loss is the feature extracted from the pre-training model as a part of the target function. If only the $L_1$ norm loss is used to optimize the interpolation slice, it is easy to cause the generated interpolation slice to have fuzzy problems. Adding the perceptual loss makes the generated interpolation slice and label slice still similar in high-level semantics. Perceptual loss enables the network to learn more high-frequency and detailed information. With the addition of perceptual loss, the texture and details of the generated image can be significantly improved. The specific calculation is as follows:

$$l_p = \frac{1}{N}\sum_{i=1}^{N}\left\|\phi\left(\hat{I}_t\right) - \phi\left(I_t\right)\right\|_2 \tag{14}$$

## IV. LUNG NODULE DETECTION NETWORK BASED ON THREE-DIMENSIONAL CONVOLUTION

After interpolation, the lung slice data with the same scale is realized, which is still data with a three-dimensional structure. In order to capture the three-dimensional features of lung slice data, this paper uses a three-dimensional convolutional depth detection network to detect nodules [39], [40], [41]. Since the input of the network is 3D lung data, the 3D convolutional network is used to extract the features.In order to avoid the limitation of the size of the graphics memory of the computing platform, the lung data is input in blocks.

### A. DATA DICING

The lung CT obtained by scanning is a three-dimensional data with $x, y, z$ axes. In order to enable the network to directly capture the three-dimensional information of lung CT, the entire lung nodule detection network uses three-dimensional convolution. The task of detecting lung nodules is for the entire lung CT image, but due to the limitation of GPU memory, the entire lung CT image cannot be directly input into the detection network. This paper first extracts three-dimensional slices from the CT image of the entire lung, and then inputs each three-dimensional slice separately into the detection network. The length of a single slice in the axial direction is $l$, and the size of each slice is $(l \times l \times l)$ voxel.

In the network training stage, set $l = 128$, and at least one nodule is included in a ratio of 70%, and no nodules are included in a ratio of 30%, and all three-dimensional blocks are extracted. The extraction method is shown in 'ure 4, In order to avoid over-fitting, the extraction process sets a scaling factor for data enhancement. First, randomly select the set zoom ratio range $s \in [0.75, 1.25]$, and get the length of a single axis of the extracted block after zooming is $l'$, which satisfies:

$$l' = l \times s \tag{15}$$

The two clipping vertices of the extracted block after scaling are $c_1 = (x_1, y_1, z_1)$, $c_2 = (x_2, y_2, z_2)$, satisfying:

$$\begin{aligned} x_2 &= x_1 + l' \\ y_2 &= y_1 + l' \\ z_2 &= z_1 + l' \end{aligned} \tag{16}$$

The center coordinates of the nodules are $(x, y, z)$ and the radius is $r$. In order to make the nodules in the cut slices not necessarily located in the center, they are randomly distributed, while ensuring that the nodules and the boundaries of the cut $A$ certain distance margin. Set the boundary margin $m = 12$, then $c_1$ is selected in the space range composed of vertices $b$ and $e$, where $b = (x_s, y_s, z_s)$, $e = (x_e, y_e, z_e)$, satisfying:

$$\begin{aligned} x_s &= x = r = m \\ y_s &= y - r - m \\ z_x &= z - r - m \\ x_e &= x + r + m - l' \\ y_e &= y + r + m - l' \\ z_e &= z + r + m - l' \end{aligned} \tag{17}$$

For the portion beyond the CT boundary of the entire lung, we use a fixed value of 170 to fill. After cropping the entire lung CT image, the zoomed block is extracted and then according to the zoom ratio of$1/s$, a slice block with side length $l = 128$ is obtained. In order to avoid over-fitting, the data were flipped on the $x$, $y$, and $z$ axes to enhance the data. In the network test reasoning stage, first the entire lung CT is equally divided into slice blocks for input, and then the output of the network is spliced to get the result. The length of the single-axis upward cutting block input by the network is $l$, which satisfies:

$$l = i + 2 \times v \tag{18}$$

$i$ is the segmentation distance and $v$ is the expanded field of view. The expanded field of view can prevent lung nodules from appearing on multiple adjacent segments at the same time. For each segmentation block, the output part of the network is stitched according to the segmentation distance $i$ on single axis.

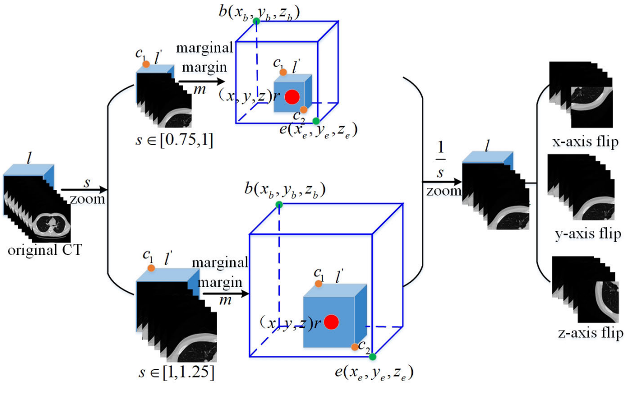


**FIGURE 10.** Data dicing.

### B. ALGORITHM DESIGN AND NETWORK STRUCTURE

The network structure of lung nodule detection based on three-dimensional convolution is shown in Figure 11, which mainly includes a U-Net similar structure of the backbone

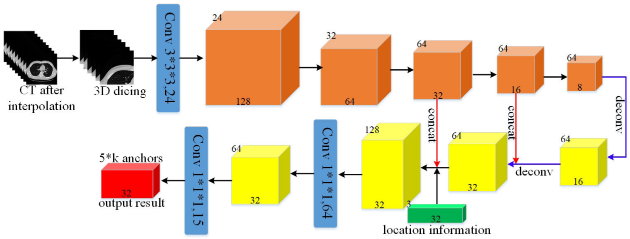


**FIGURE 11.** Lung nodule detection network based on three-dimensional convolutional neural network. The network includes feed-forward network and feed-back network, which are mainly composed of convolution operation, deconvolution operation and jump connection operation.

network and an RPN [42], [43] similar structure of the output network. The size change of the network structure data is shown in Figure 11. The U-Net backbone network structure can capture the multi-scale information of the image, which is conducive to the detection of nodules of different sizes. The network structure combines high-level semantic features and shallow detail features at the same time through jump connections, which improves the extraction of lung nodule features ability. The RPN network structure can directly output the location information of lung nodules. The backbone network includes a feed-forward network and a feed-back network.

In the feedforward network, it first passes through a convolution kernel including $(3 \times 3 \times 3)$ and a three-dimensional convolution with a channel of 24, and then passes through four three-dimensional residual convolution units. The entire feedforward network has four residual convolution units inserted in the middle of four downsampling, using four three-dimensional pooling layers, the size of the pooling layer is $(2 \times 2 \times 2)$, and the step size is 2. Each residual unit includes three residual blocks, and the residual block is extended to the three-dimensional convolution structure by ResNet in the two-dimensional convolution structure. Compared with the traditional convolutional network structure, the residual network structure is mainly to solve the problem of gradient disappearance. Generally speaking, the depth of the network structure plays an important role in improving the performance of the network model. When the number of network layers deepens, it means that different levels of data features can be extracted. At the same time, the features extracted by deeper networks are more abstract and have richer semantic information, so deep networks have better results in theory. It is that when the traditional convolutional network is continuously improving, the deeper network does not converge, mainly because the gradient disappears, so that the parameters of the shallow network cannot be updated.

In the ResNet structure, this problem is solved by increasing the identity mapping of the network layer. For the input $x$ of the neural network structure, the expected fitting target is $H(x)$. Then the ResNet structure is realized by learning the identity mapping function, and the network structure is designed as $H(x) = F(x) + x$. This is equivalent to directly transmitting the input $x$ to the output as the initial result. For the network structure, the goal to be learned becomes the residual function $H(x) = F(x) - x$. If the residual function $F(x)$ is zero, the network structure is an identical mapping $H(x) = x$. $F(x) + x$ in the ResNet structure is the addition operation of the calculated value, and the number of channels remains unchanged. When the network input and output are the same, the input is directly added to the output. When the network input and output are inconsistent, a new mapping relationship needs to be added, and the weight matrix is transformed to a new space, usually through unit convolution. In the feedback network, it first passes through two convolution kernels with size 2 and transposed convolution with step size 2 and two jump connections. The jump connections are shown in Figure 12. Then it passes through two convolution kernels $(1 \times 1 \times 1)$, and the number of channels is 64 and 1 three-dimensional convolution. The final output of the network is the feature quantity of $(32 \times 32 \times 32 \times 3 \times 5)$, and the last two dimensions correspond to the Anchor and the fitting quantity in RPN respectively. Anchor includes 3 cubes of different scales with side lengths of 5, 10 and 20 respectively. The fitting quantity is $(\hat{p}, \hat{d}_x, \hat{d}_y, \hat{d}_z, \hat{d}_r)$, among which the first one is the probability of predicting lung nodules, and the last four are the correction amounts of predicting the center position of the positioning nodule frame and the length of the positioning frame. For probability, the last output of the network is followed by a *sigmoid* activation function, *sigmoid* maps variables to [0, 1], the specific expression is as follows:

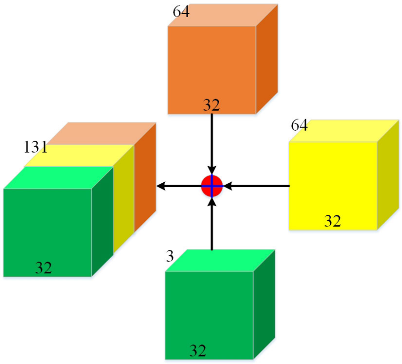


**FIGURE 12.** The concat operation in the feedback network (the lower layer in Figure 11) in the lung nodule detection network (Figure 11) combines the information in the feedforward network (the upper layer in Figure 10) with Combine the information in the feedback network and the slice information of lung nodules (green module).

$$\hat{p} = \frac{1}{1 + \exp(-x)} \tag{19}$$

### C. LOSS FUNCTION

The bounding box target label of the lung nodule is $(G_x, G_y, G_z, G_r)$, corresponding to the three-dimensional coordinates and side length of the nodule, Anchor is $(A_x, A_y, A_z, A_r)$, corresponding to the three-dimensional coordinates and side length of the Anchor, and the bounding box fitting label is $(d_x, d_y, d_z, d_r)$. The specific calculation formulas for the bounding box fitting label and the target label are as follows. For the nodule probability label $p$, it is obtained by calculating the area cross ratio (Intersectionover Union, IOU ) between the bounding box target label and the Anchor, IoU is the regional intersection ratio, where IoU greater than 0.5 is a positive sample, and less than 0.02 is a negative sample, that is, $p \in (0, 1)$ (1 represents a positive sample, and 0 represents a negative sample). The final output of the network structure is $(\hat{d}_x, \hat{d}_y, \hat{d}_z, \hat{d}_r)$.

$$d_x = \frac{G_x - A_x}{A_x} d_y = \frac{G_y - A_y}{A_y} d_z = \frac{G_z - A_z}{A_z} d_r = \frac{G_r - A_r}{A_r} \tag{20}$$

The expression of the loss function of the entire network is as follows, the $\lambda$ is the weight coefficient, and the value is 1.

$$l = l_{cls} + \lambda l_{reg} \tag{21}$$

Classification loss uses cross entropy loss:

$$l_{cls} = p \log(\hat{p}) + (1 - p) \log(1 - \hat{p}) \tag{22}$$

The bounding box fitting loss uses the smoothing loss $S$:

$$l_{reg} = \sum_{k \in \{x,y,z,r\}} S\left(d_k, \hat{d}_k\right) \tag{23}$$

The smoothing loss calculation process is as follows:

$$S(d, \hat{d}) = \begin{cases} |d - \hat{d}|, & if\ |d - \hat{d}| > 1 \\ (d - \hat{d})^2, & \text{else} \end{cases} \tag{24}$$

## V. EXPERIMENTAL CONFIGURATION AND EXPERIMENTAL RESULTS

### A. EXPERIMENTAL DATASET

The LIDC-IDRI (Lung Image Database Consortium Image Collection) dataset is widely used in the research community for developing and evaluating lung nodule detection algorithms. It contains a diverse collection of thoracic CT scans with annotated lung nodules. The dataset includes: 1018 thoracic CT scans from a variety of scanners and imaging protocols. Annotations from four experienced radiologists for each scan, providing detailed information on the location, size, and characteristics of lung nodules.

The CT scans in the LIDC-IDRI dataset are obtained from different scanners and protocols, resulting in variations in image resolution, slice thickness, and contrast. These inconsistencies pose a challenge for developing algorithms that need to generalize well across different imaging settings. The annotations provided by the four radiologists exhibit inter-observer variability, as different radiologists may have different opinions on the presence and characteristics of nodules. This variability can affect the training and evaluation of detection algorithms [44]. The nodules in the dataset vary significantly in size, shape, and density. They range from small, well-defined nodules to larger, ill-defined ones, making it challenging to develop algorithms that can accurately detect and classify all types of nodules. The CT scans also contain various non-nodule structures, such as blood vessels, airways, and other anatomical features, which can be mistaken for nodules. Differentiating between nodules and these structures is a key challenge for detection algorithms.

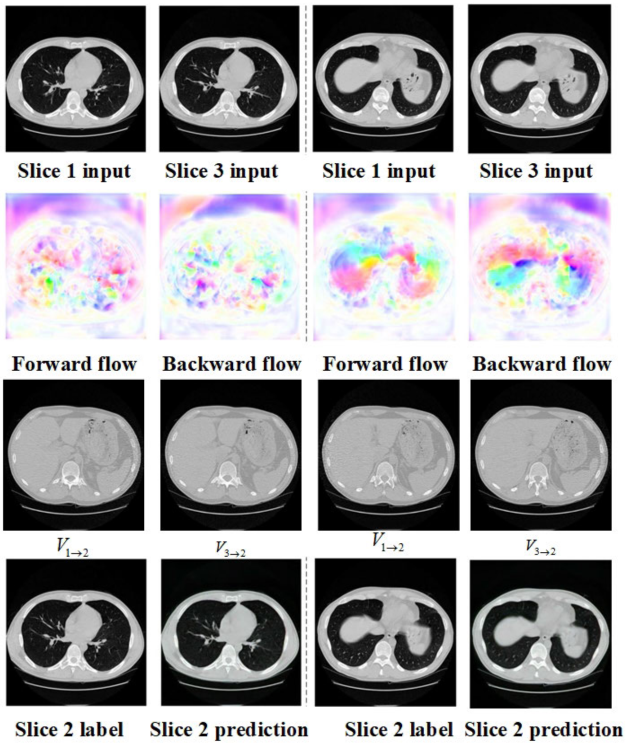


**FIGURE 13.** Lung CT interpolation results. We use CT slice 1 and slice 3 to fit slice 2 prediction. By comparing with label 2, there is almost no difference between the two and the detailed information is very consistent. (Label 2 is the Ground Truth which is Included in the data set.)

### B. LUNG INTERPOLATION NETWORK EXPERIMENT

The results of lung CT interpolation are shown in Figure 13. The adjacent slice 1 and slice 3 in the lung CT are input to the network to obtain the corresponding forward voxel flow, backward voxel flow and predicted slice 2. Compared with the second label of slice, it can be seen that the interpolation slice generated by the network is clearer and retains more image detail information. In this paper, several deep learning-based frame interpolation methods are selected for comparative experiments. Among them, DVF [20] is one of the representative work based on deep voxel stream interpolation frames, and SepConv [21] learns the convolution of corresponding pixels through convolutional networks. Kernel to achieve interpolation.

The experimental results are shown in Table 3. Due to the convolution kernel size constraint in SepConv method, but because there is no large pixel displacement mapping between the lung CT data slices, SepConv shows a better interpolation effect than DVF. The DVF method is to directly output the interpolated slice to the voxel stream of adjacent slices in one step. Compared with the method in DVF, the

**TABLE 3.** PSNR and SSIM test results.

| Predictive model | PSNR | SSIM |
|---|---|---|
| DVF [20] | 27.86 | 0.73 |
| SepConv [21] | 28.23 | 0.76 |
| EpicFlow [22] | 27.32 | 0.71 |
| LD Flow [23] | 27.53 | 0.71 |
| FlowNet [24] | 28.36 | 0.78 |
| B,Basics [25] | 28.62 | 0.80 |
| Our model | 29.19 | 0.82 |

method proposed in this paper is refined into two steps. First obtain the voxel stream of adjacent slices, and then Convert to the voxel stream of the interpolated slice to the adjacent slice, so that higher PSNR and SSIM values are obtained, and a better interpolation effect is achieved.

## C. LUNG NODULE DETECTION TEST

### 1) EXPERIMENTAL SETUP

There are far more examples of negative samples than positive samples. For some negative samples, the network can easily classify correctly, but for some negative samples with similar appearance and nodules, the network is not easy to classify correctly. To deal with this problem, online negative sample mining is used in training. First, the lung slices are input to the network, and the network outputs the corresponding bounding box and confidence. Then randomly collect N negative samples to form a candidate pool, and sort them in order of decreasing confidence. Then select the highest n as the difficult negative samples, and remove other negative samples without being included in the calculation of the loss function. By randomly selecting negative samples to form candidate regions, the correlation between negative samples can be reduced, and controlling the size of the candidate region $N$ and the number of selection n each time can help improve the model's detection capability.

The network training uses the SGD [45] optimizer to update the gradient, and the setting parameter of SGD is: the momentum item is set to 0.9. In the experiment, the training batch size is set to 8, and the number of iterations is 150 epochs. The initial learning rate $r = 0.01$, when the number of iterations is half of the total number of times, $r = 0.001$, and when the number of iterations is 120 times, $r = 0.0001$. For the deep learning framework, Pytorch is used for training and testing. In the experiment, the data in Luna16 [34] is divided into training set and test set in proportion, and the ratio of training set to test set is 8:2. Data enhancement is carried out in the following ways: (1) the image is randomly flipped in each dimension; (2) the image is scaled according to the coefficient of [0.75, 1.25].

### 2) LUNG NODULE DETECTION TEST

The evaluation index used in the lung nodule detection algorithm model in this paper is free-response receiver operating characteristic (FROC) [46]. The FROC curve is a correction to the receiver operating characteristic (ROC) curve. In the ROC curve, the row coordinate is the false positive rate, and the ordinate is the sensitivity rate; In the FROC curve, the row coordinate is the average number of false positives for each CT sample, and the ordinate is the sensitivity rate. The method for determining the true positive (TP) is that the center of the nodule output by the network detection is located within the corresponding radius from the center of the labeled nodule. If it is not within the corresponding radius, it is determined as a false negative (FN). The final result of FROC takes (the row coordinates are 0.125, 0.25, 0.5, 1, 2, 4, 8) the average value of the sensitivity rate corresponding to a total of 7 data points.

**TABLE 4.** FROC value test results.

| Predictive model | Test set FROC value (%) |
|---|---|
| DSB | 83.59 |
| DeepLung | 84.49 |
| NoduleNet | 83.14 |
| ZNET | 81.16 |
| Aidenc | 80.73 |
| DIAG_CONVNET | 84.52 |
| No interpolation method | 84.60 |
| Interpolation method | 85.16 |

The results of lung nodule detection are shown in Figure 15. The upper area in the figure corresponds to the nodule label, the position of the lung nodule is framed by a white circle, and the middle area in the figure corresponds to the detection of the lung nodule detection network of this subject, the position of the lung nodule is framed by a red circle. The lower area in the figure corresponds to the enlarged area of the detection result of the lung nodule detection network and the corresponding predicted probability. It can be seen that the proposed lung nodule detection network can correctly detect nodules, and the detection frame is close to the nodule boundary area, and the predicted slice position is close to the true label value.

In this paper, several lung nodule detection models based on deep learning are selected for comparative experiments. Among them, DSB [35] is one of the representative work of lung nodule detection using three-dimensional deep convolution. DeepLung [13] improves the structure of DPN [14] into the backbone network. NoduleNet [17] combines the backbone network of lung nodule detection and semantic segmentation tasks, and simplifies the feature extraction network. The non-interpolation model of this subject is the result of removing the lung interpolation network structure based on voxel flow. The FROC value results of these methods and the method proposed in this paper on the test set are shown in Table 4. The specific FROC curve is shown in Figure 14. The Figure 14(a) is the result of the DSB method, (b) Is the result of the DeepLung method, (c) is the result of the NoduleNet method, (d) is the result of the non-interpolation method proposed for this topic, and (e) is the result of the method proposed for this topic. It can be seen that the highest FROC value is obtained, especially after the introduction of the interpolation module based on voxel flow, the performance is greatly improved. Although the

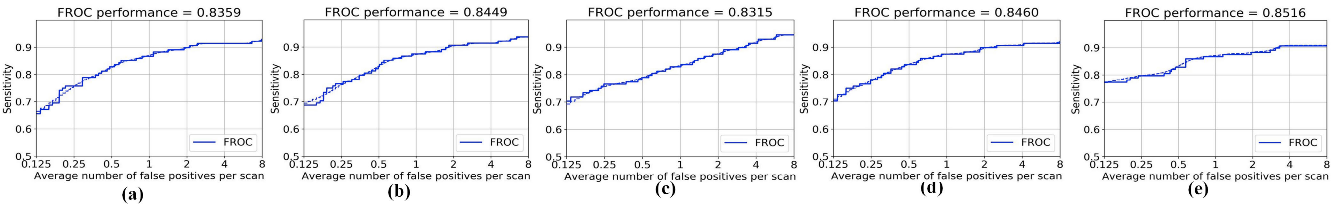


**FIGURE 14.** FROC curve result graph.

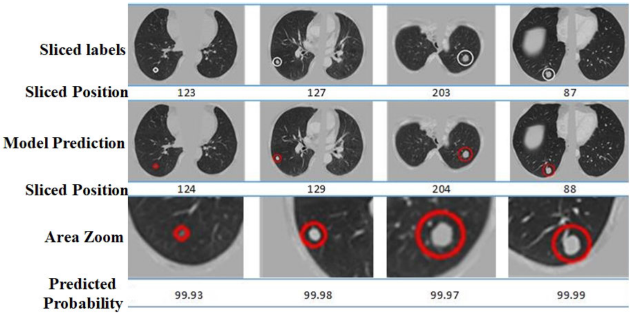


**FIGURE 15.** Lung nodule test results.

NoduleNet method achieves the combination of multi-tasks, it also reduces the performance of lung nodule detection. As shown in Figure 14, comparing several methods, the detection model proposed in this paper, in the FROC curve, especially under the average false positive number of low CT samples, greatly improves the sensitivity rate. The inference time of these methods and the method proposed in this paper on the test set is shown in Table 5. Because the three-dimensional deep convolutional network is used, and the detection process needs to input the lung CT slices, and finally the network The output results are spliced, so several methods are time-consuming. In the NoduleNet method, due to the simplified feature extraction network, the network reasoning time is the shortest. In this subject, the interpolation network based on voxel flow takes 0.15 seconds, so it takes longer than the non-interpolated network, but it still meets the needs of practical medical and clinical use.

**TABLE 5.** Test time results.

| Predictive model | Time (s) |
|---|---|
| DSB | 6.73 |
| DeepLung | 6.82 |
| Nodulenet | 4.81 |
| ZNET | 4.32 |
| Aidenc | 5.95 |
| DIAG_CONVNET | 6.23 |
| No interpolated method | 6.77 |
| Interpolated method | 6.92 |

## VI. DISCUSSION

Although the voxel flow-based interpolation network is expected to achieve consistent resolution on the vertical axis of lung CT data, there are still several potential limitations to consider. For lung CT interpolation, design a more reasonable and concise network to implement lung CT interpolation, balance the compression model from the perspectives of network depth, width, resolution, etc., reduce the lung CT interpolation time, and improve the practical application value. The voxel flow-based interpolation network may have difficulty processing highly irregular or noisy data. Irregularities such as severe anatomical abnormalities or artifacts can adversely affect the accuracy of voxel flow estimation, thereby affecting the quality of interpolated slices. For the lung nodule detection network, improve the feature extraction module, such as the currently proposed network structure search (NAS) method, and obtain a more reasonable and efficient lung nodule feature extraction network structure through search. With deep learning as the core, the lung interpolation network and the detection network are combined, and a fully end-to-end network is set to train the model, so that the lung interpolation network and the detection network can have information exchange during training, promote each other, and further improve the lung nodule detection effect.

Future work can focus on enhancing the robustness of the voxel flow-based interpolation network to handle noisy and highly irregular data. This may involve incorporating noise reduction techniques or developing more sophisticated interpolation models to accommodate varying degrees of irregularity. Improving the efficiency of interpolation algorithms is critical for practical applications. Exploring optimization techniques such as model pruning, quantization, or efficient architecture design can help reduce the computational burden and processing time. Combining the voxel flow-based interpolation network with other advanced image processing techniques, such as GANs (generative adversarial networks) for data augmentation or unsupervised learning methods for better feature extraction, can further improve the overall performance of the system. By addressing these potential limitations and exploring future research directions, the proposed method can be improved to be more robust, efficient, and suitable for practical applications in lung nodule detection and other medical imaging tasks.

Our proposed method can be clinically applied through, Integration with existing systems For practical application of the proposed method in clinical work, its integration with existing radiology systems and workflows needs to be considered. One is the gradual introduction of various AI-assisted diagnosis platforms in modern hospitals, which can automatically analyze and assist in diagnosis of medical images. The proposed method can be used as part of these platforms to provide more accurate pulmonary nodule detection capabilities. Integrate the algorithm into the existing AI platform through API or plug-in form to ensure compatibility and collaborative work with other AI tools. Another is to integrate the proposed method into imaging workstation software, allowing doctors to directly access and use denoising and nodule detection functions when viewing and analyzing CT images. This integration method can be seamlessly integrated into doctors' daily workflow and improve diagnostic efficiency.

## VII. CONCLUSION

This thesis combines deep learning and lung nodule detection technology. Aiming at the problem of inconsistent scales on the vertical axis of lung CT data, this article is based on the deep learning method, First, the lung interpolation network based on voxel flow is used to achieve the same scale, and then the lung nodule detection network is completed by the lung nodule detection network based on three-dimensional convolution, and a better lung nodule detection effect is obtained. This paper innovatively starts with the inconsistency of the vertical axis scale of lung CT data, and proposes a lung interpolation network based on voxel flow. This paper proposes a lung nodule detection network based on three-dimensional convolution, which combines U-Net-like and RPN-like network structures. Combining the lung CT data with consistent scale as input, the experimental results show that a better lung nodule detection effect is obtained. The lung CT interpolation module proposed in this paper achieves consistent vertical axis resolution of lung CT data. Not only can it be applied to lung nodule detection, this module can be inspired to be applied to other vision tasks in lung CT, such as lung CT semantic segmentation.

## REFERENCES

[1] C. F. Mountain, "Revisions in the international system for staging lung cancer," *Chest*, vol. 111, no. 6, pp. 1710–1717, Jun. 1997.

[2] L. C. M. Gomez, M. Kondratova, N. Sompairac, C. Lonjou, J.-M. Ravel, E. Barillot, A. Zinovyev, and I. Kuperstein, "Atlas of cancer signaling network: A resource of multi-scale biological maps to study disease mechanisms," in *Systems Medicine: Integrative Qualitative and Computational Approaches* (Reference Module in Biomedical Sciences), O. Wolkenhauer, Ed., Amsterdam, The Netherlands: Elsevier, 2020.

[3] M. Infante, S. Cavuto, F. R. Lutman, G. Brambilla, G. Chiesa, G. Ceresoli, E. Passera, E. Angeli, M. Chiarenza, G. Aranzulla, U. Cariboni, V. Errico, F. Inzirillo, E. Bottoni, E. Voulaz, M. Alloisio, A. Destro, M. Roncalli, A. Santoro, and G. Ravasi, "A randomized study of lung cancer screening with spiral computed tomography: Three-year results from the dante trial," *Amer. J. Respiratory Crit. care Med.*, vol. 180, no. 5, pp. 445–453, 2009.

[4] S. P. Singh, D. S. Gierada, P. Pinsky, C. Sanders, N. Fineberg, Y. Sun, D. Lynch, and H. Nath, "Reader variability in identifying pulmonary nodules on chest radiographs from the national lung screening trial," *J. Thoracic Imag.*, vol. 27, no. 4, pp. 249–254, 2012.

[5] E. L. Torres, E. Fiorina, F. Pennazio, C. Peroni, M. Saletta, N. Camarlinghi, M. E. Fantacci, and P. Cerello, "Large scale validation of the M5L lung CAD on heterogeneous CT datasets," *Med. Phys.*, vol. 42, no. 4, pp. 1477–1489, Apr. 2015.

[6] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in CT images using deep convolutional neural networks," *Pattern Recognit.*, vol. 85, pp. 109–119, Jan. 2019.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[8] N. Khosravan and U. Bagci, "S4ND: Single-shot single-scale lung nodule detection," in *Proc. 21st Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Granada, Spain: Springer, Sep. 2018, pp. 794–802.

[9] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[10] Q. Dou et al., "Automated pulmonary nodule detection via 3D ConvNets with online sample filtering and hybrid-loss residual learning," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Quebec City, QC, Canada: Springer, Sep. 2017, pp. 630–638.

[11] H. Tang, D. R. Kim, and X. Xie, "Automated pulmonary nodule detection using 3D deep convolutional neural networks," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 523–526.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Munich, Germany: Springer, Oct. 2015, pp. 234–241.

[13] W. Zhu, C. Liu, W. Fan, and X. Xie, "DeepLung: Deep 3D dual path nets for automated pulmonary nodule detection and classification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2018, pp. 673–681.

[14] Y. Chen, J. Li, H. Xiao, X. Jin, S. Yan, and J. Feng, "Dual path networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–9.

[15] P. Monkam, S. Qi, H. Ma, W. Gao, Y. Yao, and W. Qian, "Detection and classification of pulmonary nodules using convolutional neural networks: A survey," *IEEE Access*, vol. 7, pp. 78075–78091, 2019.

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

[17] H. Tang, C. Zhang, and X. Xie, "NoduleNet: Decoupled false positive reduction for pulmonary nodule detection and segmentation," in *Proc. 22nd Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Shenzhen, China: Springer, Oct. 2019, pp. 266–274.

[18] H. Tang, X. Liu, and X. Xie, "An end-to-end framework for integrated pulmonary nodule detection and false positive reduction," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 859–862.

[19] J. Liu et al., "3DFPN-HS$^2$: 3D feature pyramid network based high sensitivity and specificity pulmonary nodule detection," in *Proc. 22nd Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*. Shenzhen, China: Springer, Oct. 2019, pp. 513–521.

[20] M. Hayat, S. Aramvith, and T. Achakulvisut, "Combined channel and spatial attention-based stereo endoscopic image super-resolution," in *Proc. IEEE Region 10 Conf. (TENCON)*, Oct. 2023, pp. 920–925.

[21] M. Hayat and S. Aramvith, "Saliency-aware deep learning approach for enhanced endoscopic image super-resolution," *IEEE Access*, vol. 12, pp. 83452–83465, 2024.

[22] M. Hayat and S. Aramvith, "E-SEVSR—Edge guided stereo endoscopic video super-resolution," *IEEE Access*, vol. 12, pp. 30893–30906, 2024.

[23] N. Ahmad, R. Strand, B. Sparresäter, S. Tarai, E. Lundström, G. Bergström, H. Ahlström, and J. Kullberg, "Automatic segmentation of large-scale CT image datasets for detailed body composition analysis," *BMC Bioinf.*, vol. 24, no. 1, pp. 346–359, Sep. 2023.

[24] B. Zhang, S. Qi, P. Monkam, C. Li, F. Yang, Y.-D. Yao, and W. Qian, "Ensemble learners of multiple deep CNNs for pulmonary nodules classification using CT images," *IEEE Access*, vol. 7, pp. 110358–110371, 2019.

[25] J. Liu, L. Cao, O. Akin, and Y. Tian, "Robust and accurate pulmonary nodule detection with self-supervised feature learning on domain adaptation," *Frontiers Radiol.*, vol. 2, Dec. 2022, Art. no. 1041518.

[26] Z. Liu, R. A. Yeh, X. Tang, Y. Liu, and A. Agarwala, "Video frame synthesis using deep voxel flow," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4463–4471.

[27] S. Niklaus, L. Mai, and F. Liu, "Video frame interpolation via adaptive separable convolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 261–270.

[28] G. Li, W. Zhou, W. Chen, F. Sun, Y. Fu, F. Gong, and H. Zhang, "Study on the detection of pulmonary nodules in CT images based on deep learning," *IEEE Access*, vol. 8, pp. 67300–67309, 2020.

[29] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "EpicFlow: Edge-preserving interpolation of correspondences for optical flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1164–1172.

[30] T. Brox and J. Malik, "Large displacement optical flow: Descriptor matching in variational motion estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 500–513, Mar. 2011.

[31] A. Dosovitskiy, P. Fischer, E. Ilg, P. Häusser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox, "FlowNet: Learning optical flow with convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2758–2766.

[32] Y. Sun, J. Tang, W. Lei, and D. He, "3D segmentation of pulmonary nodules based on multi-view and semi-supervised," *IEEE Access*, vol. 8, pp. 26457–26467, 2020.

[33] J. J. Yu, A. W. Harley, and K. G. Derpanis, "Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness," in *Computer Vision—ECCV 2016 Workshops*. Amsterdam, The Netherlands: Springer, Oct. 2016, pp. 3–10.

[34] Y. L. Liu, Y. T. Liao, Y. Y. Lin, and Y. Y. Chuang, "Deep video frame interpolation using cyclic frame generation," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 8794–8802.

[35] W. Bao, W.-S. Lai, C. Ma, X. Zhang, Z. Gao, and M.-H. Yang, "Depth-aware video frame interpolation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3698–3707.

[36] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8934–8943.

[37] H. Jiang, D. Sun, V. Jampani, M.-H. Yang, E. Learned-Miller, and J. Kautz, "Super SloMo: High quality estimation of multiple intermediate frames for video interpolation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9000–9008.

[38] S. Niklaus and F. Liu, "Softmax splatting for video frame interpolation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5436–5445.

[39] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, Mar. 2011.

[40] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*. Amsterdam, The Netherlands: Springer, Oct. 2016, pp. 694–711.

[41] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.

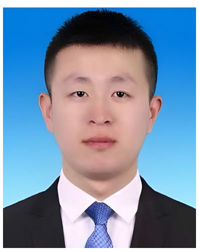
[42] W. Wang, J. Luo, X. Yang, and H. Lin, "Data analysis of the lung imaging database consortium and image database resource initiative," *Academic Radiol.*, vol. 22, no. 4, pp. 488–495, Apr. 2015.

[43] G. Montavon, G. Orr, and K.-R. Müller, Eds., *Neural Networks: Tricks of the Trade*, vol. 7700. Springer, 2012.

[44] L. Cheng et al., "A tensor processing framework for CPU-manycore heterogeneous systems," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 41, no. 6, pp. 1620–1635, 2021.

[45] A. A. A. Setio et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," *Med. Image Anal.*, vol. 42, pp. 1–13, Dec. 2017.

[46] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the malignancy of pulmonary nodules using the 3-D deep leaky noisy-OR network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3484–3495, Nov. 2019.



**WEI ZHANG** received the bachelor's and master's degrees from Harbin Engineering University, in 2016 and 2019, respectively. He is currently pursuing the Ph.D. degree with Harbin Institute of Technology. During the period, he has published many journal articles and conference papers. His research interests include computer vision and medical image processing. In 2019, he won the Best Creative Champion of the Robotx Challenge in Hawaii, USA, and won a number of national scholarships and national championships. He served as a Reviewer for ICME, ICASSP, and several journals.

**ABDERRAHMANE SALMI** received the bachelor's and master's degrees from the University Frères Mentouri Constantine 1, in 2015 and 2017, respectively. He is currently pursuing the Ph.D. degree in computer science with Harbin Institute of Technology. During this period, he has published research articles. His research interests include machine learning, deep learning, computer vision, image compression, and medical image processing.

**FENG JIANG** received the Bachelor of Engineering, Master of Engineering, and Ph.D. of Engineering degrees in computer science and technology from Harbin Institute of Technology, in 2001, 2003, and 2008, respectively. He is the Professor of the School of Computer Science, Harbin Institute of Technology. He is the Chief Expert of Heilongjiang VR Alliance. He has published more than 100 articles in related fields, of which the first author has published more than 20 articles in international journals, such as *Journal of Machine Learning Research* and IEEE Transactions; a single article has been cited up to 210 times; and three monographs and textbooks have been published in Chinese and English. In the past five years, the first person in charge has presided over more than ten scientific research projects, such as sub-projects of national key research and development projects, the National Natural Science Foundation of China, natural fund youth, and international cooperation; and participated in the national key research and development plan, the national natural science fund key, the national 973 plan, and 863. 20 international cooperation projects. He won the second prize of Ministerial Science and Technology Progress Award (ranked second), the first prize of Heilongjiang Provincial University Science and Technology Award, the National Championship of the 2017 Sohu Graphic and Text Matching Competition, and the Fourth ChaLearn Gesture Recognition Challenge (the first domestic competition), and more than ten college-level awards and other honors. He is the Chairperson of IEEE Harbin Signal Processing Branch.

**CHI-FU YANG** (Member, IEEE) was born in October 1982. He received the Ph.D. degree in engineering. He is currently a top Associate Professor with Harbin Institute of Technology, the Doctoral Supervisor, the Deputy Director of the Electro-Hydraulic Servo Institute, Harbin Institute of Technology, the Head of the Robotics and Rehabilitation Laboratory, Harbin Institute of Technology, and a senior visitor from Columbia University in the United States Scientist, "Young Top Talent" of Harbin Institute of Technology, Scientific Chinese Person of the Year. He is a member of American Mechanical Engineering Association, a Senior Member of Chinese Mechanical Engineering Association, an Expert Member of Chinese Geriatrics Association Medical Care Promotion Committee, the Director of the Intelligent Manufacturing and Robotics Alliance of the Yangtze River Basin President, the Vice Chairperson of China Medical and Nursing Integration Alliance, a special reviewer of international top magazines and editorial board of international magazines, and the Chairperson and a Chief Scientist of Harbin Institute of Technology Robotics Group Tianyu Rehabilitation Medical Robot Co., Ltd., Harbin Institute of Technology Robotics Group Harbin Institute of Technology Tianyu (Zhongshan) Robot Executive director of a limited company.