



# DAS-Net: A lung nodule segmentation method based on adaptive dual-branch attention and shadow mapping

Shichao Luo<sup>1</sup> · Jina Zhang<sup>1</sup> · Ning Xiao<sup>1</sup> · Yan Qiang<sup>1</sup> · Keqin Li<sup>2</sup> · Juanjuan Zhao<sup>1</sup> · Liang Meng<sup>1</sup> · Ping Song<sup>1</sup>

Accepted: 24 November 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Quantitative analysis of pulmonary nodules is necessary for the early diagnosis and treatment of lung cancer, improving the possibility of patient survival. Although deep convolutional neural networks (DCNNs) have been widely used in the quantification of pulmonary nodules, these models generally suffer from poor detailed information learning and a large model size. In this paper, for solving the problem of insufficient learning on the surface detailed information of pulmonary nodules and excessive model parameters and computation, we propose a 3D Dual Attention Shadow Network (DAS-Net) to solve the problem of accurate pulmonary nodules segmentation. The model consists of a shadow mapping encoder, a shadow mapping decoder and an adaptive double-branch attention module. An encoder-decoder based on shadow mapping is designed, which greatly shrinks parameters and computation while ensuring the richness of feature mapping. We also design an adaptive double-branching attention module that focuses on learning the surface detailed information of nodules by using an attention mechanism. We evaluate our model on the LIDC-IDRI which is the largest publicly available dataset. Experimental results not only demonstrate the effectiveness of DAS-Net, but higher dice score and Hausdorff distance of segmentation results of our method compared with other deep learning methods.

**Keywords** Attention mechanism · Model compression · Pulmonary nodules · 3D Segmentation

## 1 Introduction

In 2020, the number of deaths caused by lung cancer reached 1.8 million, ranking first in the number of cancer deaths [1]. Lung cancer develops quickly and has a high mortality rate. There are no obvious symptoms in the early stage of lung cancer, but most patients are in the middle and late stages when it is found. Once lung cancer has spread, the five-year survival rate is only about 13% [2, 3]. Most of the early indications of lung cancer appear in the form of lung nodules. Pulmonary nodules are classified as solid, partially solid and ground glass, and the probability of canceration is about 7%, 63% and 18%, respectively [4]. Computed Tomography (CT) is increasingly used in the detection of pulmonary nodules, automatic screening and diagnosis of lung cancer because of

its detailed parameters, clear anatomical relationship and low cost. After communicating with physicians, we find that the shape and surface information of pulmonary nodules are valuable for the diagnosis of tumors (The information of tumor boundary in CT data is referred to as surface information in 3D space). [5] At present, the segmentation task of lung nodules are mostly done manually by radiologists with professional knowledge and experience, which is a tedious, time-consuming and experiential task. In contrast, the use of computer-aided diagnosis to complete the segmentation task of pulmonary nodules can significantly improve work efficiency of the radiologist and robustness of the nodule segmentation results. At the same time, accurate segmentation results are extremely helpful to assist physicians in diagnosis and treatment.

In the past few years, the segmentation methods for pulmonary nodules are mainly based on traditional image processing algorithms and deep learning. In the past ten years, the traditional image algorithm has played a major role in segmentation task of pulmonary nodules. The idea based on the traditional algorithms are to capture the position, brightness, contrast and pixel distribution in the image to complete the segmentation task. Savi et al. [6] carried out a segmentation

✉ Yan Qiang  
qiangyan@tyut.edu.cn

<sup>1</sup> College of Information and Computer, Taiyuan University of Technology, Taiyuan, China

<sup>2</sup> Department of Computer Science, State University of New York, 12561 New Paltz, NY, USA

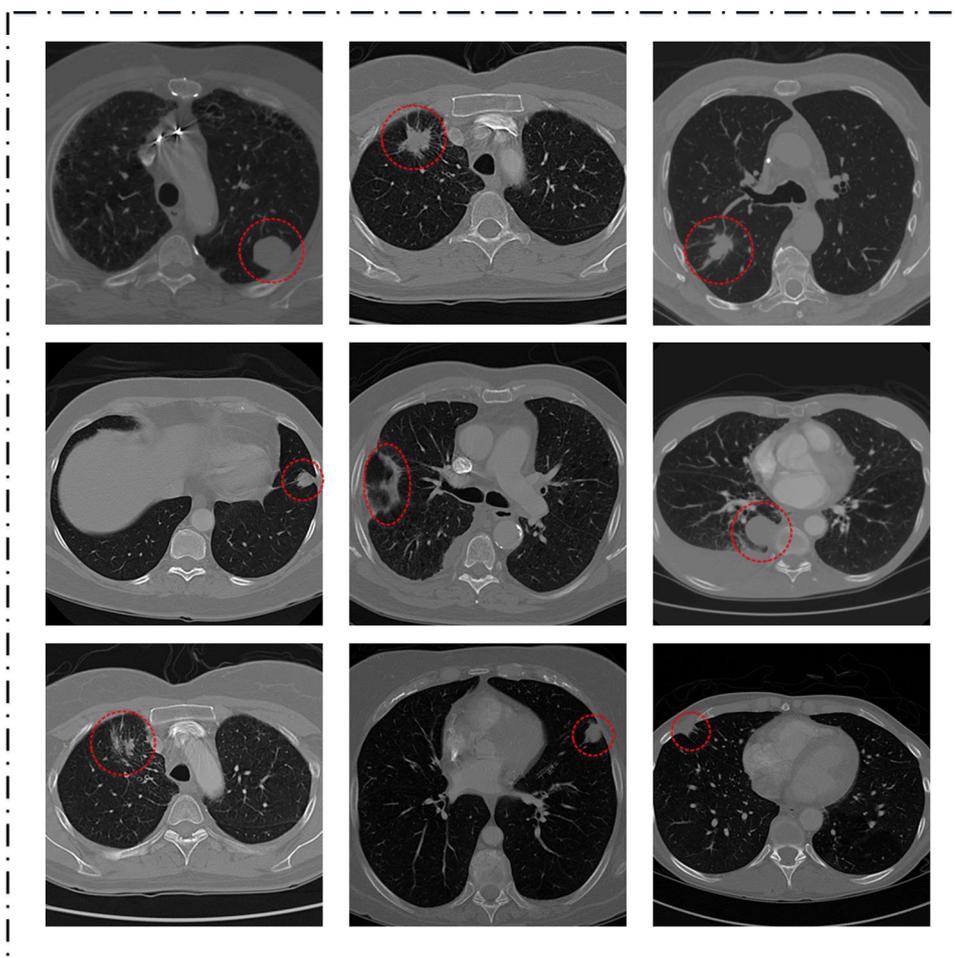
algorithm based on the fast-moving method, which divided the image into regions with similar features, and then merges the different regions through the k-means algorithm. This method was effective for round and irregular nodules. Suji et al. [7] used the optical flow algorithms Farneback, Horn-Schunck and Lucas-Kanade, which were commonly used in moving object segmentation tasks, to segment nodules from CT images. Traditional algorithms have achieved outstanding results in the nodules segmentation task, nevertheless, the segmentation accuracy of traditional algorithms is low due to the limitations of data quality, like nodules size small and the susceptibility to adhesion (see Fig. 1).

The deep learning-based approaches have superior performance in the pulmonary nodules segmentation task. Pawar et al. [8] extracted multi-scale intensive features from the images using encoder paths, and the decoder paths obtained lung segmentation images based on the multi-scale features. Wu et al. [9] constructed a coarse-to-fine segmentation of pulmonary nodules using a combination of image enhancement and dual-branch neural networks. Firstly, the images were pre-processed so that the purpose of roughly localizing the lesion area, eliminating background noise and focusing boundary

features could be achieved. Secondly, a double-branch network was designed to effectively explore the information of nodules in 2D sections and the relationship between adjacent sections, to achieve more accurate segmentation. Pezzano et al. [10] made all the background and minor vital elements in CT into two masks, and used DCNN to learn the two masks and encode them as context information. After that, the nodules were segmented by subtracting the context information from the CT images. Liu et al. [11] explored a fully automatic algorithm for accurately segmenting lungs from chest CT images. A DCNN model was first constructed to implement the classification of the chunked lung images to extract the initial lung regions. Then the pre-processed chest CT as a hyperpixel segmented and the lung contours were locally refined using the adjacency point statistics method. However, since the object of 2D deep learning method is a single image, the inter-layer context information of the focal organs is lost for medical images containing three-dimensional information. As the consequence, the improvement of segmentation accuracy is limited.

As deep learning techniques and computer hardware continue to evolve, it is possible to process CT images in 3D.

**Fig. 1** Pulmonary nodules have tissue adhesions, small capillary adhesions and ground glass shadows. The tumors have different shapes and large size spans, which bring tremendous challenges to the segmentation task



Hence, 3D deep learning method has been widely adopted in the field of segmentation of pulmonary nodules. Sahu et al. [12] showed a multi-stage algorithm for lung 3D segmentation in CT scans. The first stage used 3D CNN to obtain the coarse segmentation of the left and right lungs. In the second stage, 3D structural correction CNN was used to correct the shape of the segmented mask. Finally, a flood-fill operation was used to up-sampling and optimize the segmented mask after shape correction. Nishio et al. [13] described a model which based on the generative adversarial network to segment pulmonary nodules and generated a 3D result. To improve the segmentation accuracy, the nodule size was integrated into the model as the guidance information. Sun et al. [14] explained a semi-supervised 3D segmentation network for pulmonary nodules. This method retrieved multi-scale features from three different views and obtained spatial and semantic information at different scales. Then, for some samples that were difficult to identify, a mixed loss function with adjustment factor was proposed to make the network devote more focus to the learning process of these samples. It is easy to find that most of the research focuses on increasing the information abundance to strengthen the model's learning of nodules, but most methods ignore the small blood vessels and organ adhesions at the edges of the nodules, and do not make targeted improvements to the network. In addition, few studies have spotlighted on the explosion of parameters and computation difficulties caused by 3D methods in the medical segmentation field.

In this paper, we propose the Dual Attention Shadow Network (DAS-Net) to accurately segment 3D pulmonary medical images. Our DAS-Net adopts a UNet-like architecture, which is made up of an encoder-decoder based on shadow mapping and an adaptive double-branching attention module. Here, the encoder codes the spatial and surface feature information of the nodules contained in the volume data as high-level features. The adaptive dual-branch attention module calculates the long-distance dependence which contained in voxels in high-level features from two dimensions of spatial and channel. The results of the attention module, which can be seen as a kind of guidance information, and high-level features are input into the decoder simultaneously to improve the perceptual learning ability of the model for nodular boundary information (see Fig. 2). We evaluate the proposed DAS-Net model on the LIDC-IDRI dataset, and achieved the most advanced performance on this pulmonary nodules segmentation task. The main contributions of this paper are summarized as follows:

- We suggest a dual-attention module with adaptive capability and call it an adaptive dual-attention module, in which the adaptive weight matrix can enhance the model's ability to perceive the detailed information of 3D nodule surface and increase the segmentation accuracy.
- We construct the 3D shadow mapping layer and use it to build the basic structure of the whole network. This method can guarantee the feature richness extracted by the network while effectively shrinking the number of model parameters and computation. To the best of our knowledge this is the first time it is used in the field of 3D medical image segmentation.
- We present an accurate 3D lung nodule segmentation model with small number of parameters, called DAS-Net, which has superior performance to the most advanced methods in lung nodule segmentation tasks.

## 2 Related work

### 2.1 Medical image segmentation

During the past few years, Deep Convolution Neural Network (DCNN) has made remarkable achievements in the area of medical image processing [9, 15–17]. Ronneberger et al. [18] created U-Net based on the idea of an encoder-decoder, at the same time, skip-connection was added between the encoder and the decoder to maximize the retention of important information in the feature maps of different sizes during the feature extraction process. This method achieved excellent segmentation performance in a variety of 2D medical image segmentation tasks. Kushnure et al. [19] enhanced the multi-scale features in the CNN to improve the network receptive field by capturing global and local features at a finer-grained level in the image. They also realigned the channel responses between multi-scale features to enhance the segmentation capability of the network. Fang et al. [20] developed a kind of novel selective feature aggregation network that could learn the tumor boundary region and the central region separately to predict the target region and the boundary to improving the segmentation capability. These 2D methods cannot effectively segment the spatial information contained in medical images, resulting in limited segmentation accuracy. These 2D methods have a common problem. Because of the 2D method cannot effectively use the spatial information contained in the medical image for segmentation, the segmentation accuracy is limited.

With the rapid development of 3D convolution, 3D DCNN could take volume data as input, and the extracted spatial information could effectively enrich the information abundance. Compared with 2D DCNN, 3D DCNN has obvious advantages in medical image segmentation. Perslev et al. [21] designed a comprehensive 3D segmentation method to slice data from multiple angles. They used multiple encoders to extract lesion information from these images which cutted in different angles and integrated it into a single decoder to restore the final 3D segmentation results. Rickmann et al. [22]

extended the Squeeze-and-Excitation (SE) [23] module to three-dimensional space and carried out the project & excite (PE) module based on the idea of SE. In this method, more spatial information was retained by compressing the feature image and the information was added into the network decoding path in the excitation step to better complete the three-dimensional segmentation task of medical images. Usman et al. [24] relied on the residual U-Net [25] to complete the detection and segmentation of nodules. They offered adaptive RoI region algorithm to dynamically explore the nodules in the volume. After detected 3d nodules, the deep residual U-Net was used to complete the segmentation task from the three views of the nodules.

## 2.2 Attention learning

The aim of attention mechanism is that the model can ignore irrelevant information and focus on important information just like human. Flexible application of attention mechanism can improve the results of semantic segmentation task. Vaswani et al. [26] delivered to integrate the self-attention mechanism into the sequence model, and built (key, query, value) triad vectors to capture long-distance dependence through Scaled Dot-Product Attention module. This method achieved great success. Wang et al. [27] found an efficient non-local module on the basis of constructing triples. This module could improve the modeling effect of long-distance dependence. Huang et al. [28] created a novel cross-attention module, which collected the context information of all pixels in each pixel's cross-path and constructed the long-distance dependence of the whole image through further cyclic operation. Fu et al. [29] used two forms of attention on spatial and channel to learn long-distance dependence in slices, and then fused these dependency features, achieving final results. Sinha et al. [30] captured richer contextual relevance by using a guided self-attention mechanism. This method integrated the local features with their corresponding global dependencies, and highlighted the interdependent channel feature maps in an adaptive way, thus improving the segmentation accuracy. Park et al. [31] developed a simple and effective bottleneck attention module, which could generate attention diagrams for the features at each down-sampling (bottleneck) of the model along two different paths of channel and space to guide the segmentation task.

## 2.3 Balance the computation and model performance

The balance between computation and model performance is a problem that must be faced by every model. The principle is to decrease the number of parameters and computation of the model as much as possible without reducing the model performance. Yu et al. [32] employed dilation convolution for extending the perceptual field while acquiring multi-scale

contextual information. Dilated convolution could expand the receptive field of the model with the same image resolution without introducing additional parameters. Zeng et al. [33] proposed a 3D Tilted Convolution (3D-TC) method to solve the problem that the number of 3D segmented network parameters was too large. The parameters in multiple convolution cores were learned in the same layer, which could greatly reduce the amount of 3D model computation without reducing the model performance. Han et al. [34] developed a new Ghost module that could generate more features using fewer parameters on the basis of the original features.

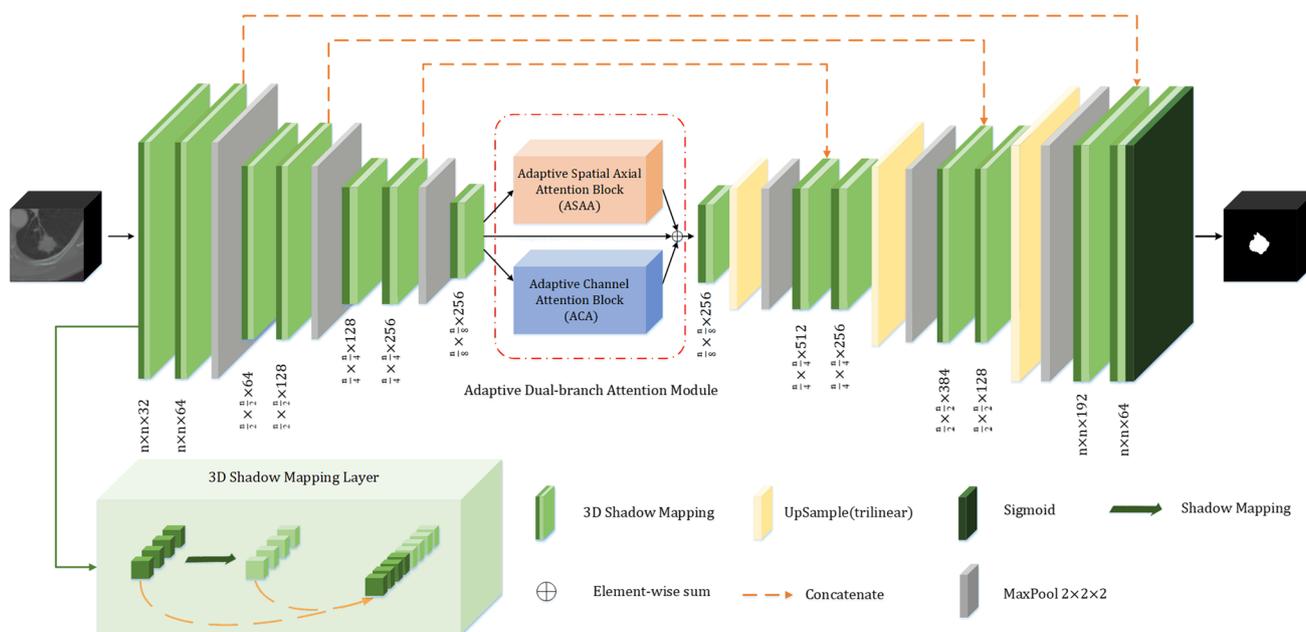
## 3 Method

### 3.1 Overview

Our proposed DAS-Net learns the long-range dependence in volume data through adaptive attention module, which decreases the network parameters while increasing the accuracy of the segmentation task. First of all, we develop an adaptive dual-attention module. By calculating the spatial and channel long-range dependencies between voxels, we can adaptively assign higher weights to important positions and channels, which solves the difficulty of learning the positional correlation of the previous attention model. Then, a shadow mapping encoder-decoder is built using a shadow mapping layer substituted for the convolutional layer in DCNN. This approach can make effective reduction in parameters and computation while guaranteeing feature diversity. Analogous to [35], we merge the outputs of the encoder and decoder layers which have the same feature scale by skip-connection to lower the missing information caused by the max-pooling operation. Figure 2 illustrates the architectures of our model, it composed of a shadow mapping encoder, a shadow mapping decoder and an adaptive dual-attention module.

### 3.2 Encoder-decoder based on shadow mapping

In the traditional CNN model, different high-level features can be extracted after the convolution calculation of images with different parameters. However, there are many convolution kernels with high similarity in the traditional convolution calculation process, and many similar feature mappings can be obtained. So, there is data and computational redundancy in CNN, which seems to be a common problem in CNN. Compared with 2D convolution, the number of parameters in 3D convolution method will increase by  $D$  times ( $D$  stands for convolutional kernel depth), and the computation will increase by  $D^2$  times. The 3D method enlarges the redundancy of data and computation. It also increases the difficulty of training the model. To solve this problem, inspired by [34],



**Fig. 2** DAS-Net structure diagram. It consists of an encoder, an adaptive dual attention module and a decoder.  $n$  is the height and width of the cropped volume during training

we extend this idea to 3D space and built 3D shadow mapping. Specifically, we suppose that a 3D convolution generates  $m$  feature mappings as  $Y \in R^{h \times w \times d \times m}$ :

$$Y = X * f \tag{1}$$

where  $*$  indicates convolution operation;  $f \in R^{c \times k \times k \times k \times m}$  is the core convolution kernel of this layer, and  $k \times k \times k$  means the length, width and height of the convolution kernel  $f$ . To enlarge the feature richness further, we carry out linear transformation operations on the basis of the core feature  $Y$  and obtain the final Shadow feature:

$$y_{ij} = \phi_{i,j}(Y_i), \forall i = 1, \dots, m \quad \forall j = 1, \dots, s \tag{2}$$

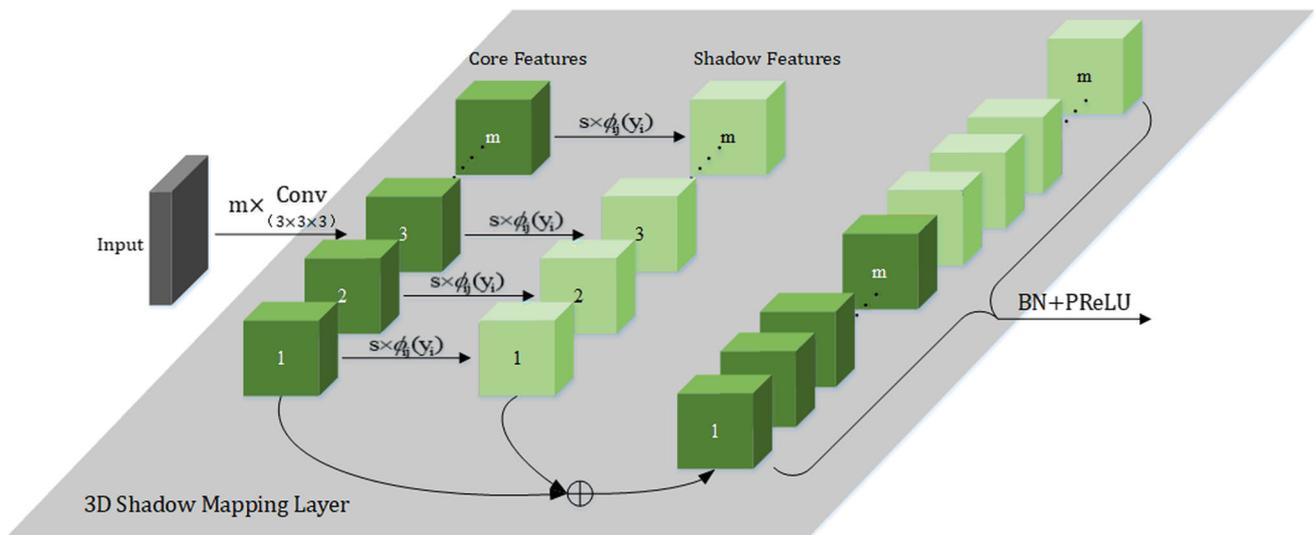
where  $\phi_{i,j}()$  represents a linear operation to generate the  $j^{th}$  shadow feature, and  $y_{ij}$  is the  $j^{th}$  shadow feature obtained from  $Y_i$ . Through the operation, we can get  $n = m \cdot s$  ( $s$  represents the number of linear operations) shadow features on the basis of  $m$  core features. We concatenate the core features and shadow features, and then perform batch normalization and PReLU activation. The linear operation  $\phi$  can make the task that originally required  $n k^3$  size convolution kernels to complete, now it only  $m$  can be completed. Compared with the traditional 3D convolution method, the number of parameters and the computation are reduced by  $k^3 \times (n - m)$  and  $2 \times k^3 \times (n - m)$  respectively (the input feature length, width, height and number of channels are the same as the traditional method). Yu and Koltun [32] and Zeng and Zheng [33] reduce the model size by using dilated convolution, the presence of empty holes in the convolution will lose a lot of surface detail information of the nodules. In contrast, each convolution in

the shadow mapping method extracts the fully surface detailed information in the voxels and ensures the feature integrity to the maximum extent.

Based on the shadow mapping, we build the shadow mapping layer(SML), as shown in Fig. 3. The encoding part codes volume data into high-level features. The decoding part restores extracted nodules surface detail information to a voxel-by-voxel representation. A shadow mapping unit is composed of  $3 \times 3 \times 3$  convolution operation, linear mapping, batch normalization and PReLU. The coding path has six shadow mapping units, and each of two mapping units constitutes one layer. The 3D Max Pooling operation is used for the down-sampling of the feature map for each layer. The decoding path also includes six shadow mapping units. The two units are combined into one layer. The feature maps are upsampled from the bottom up and connected to the same scale in the encoding path.  $1 \times 1 \times 1$  convolution and sigmoid activation function applied by us in the last layer of the decoding path to merge the features and derive the segmented results.

### 3.3 Adaptive dual-branch attention module

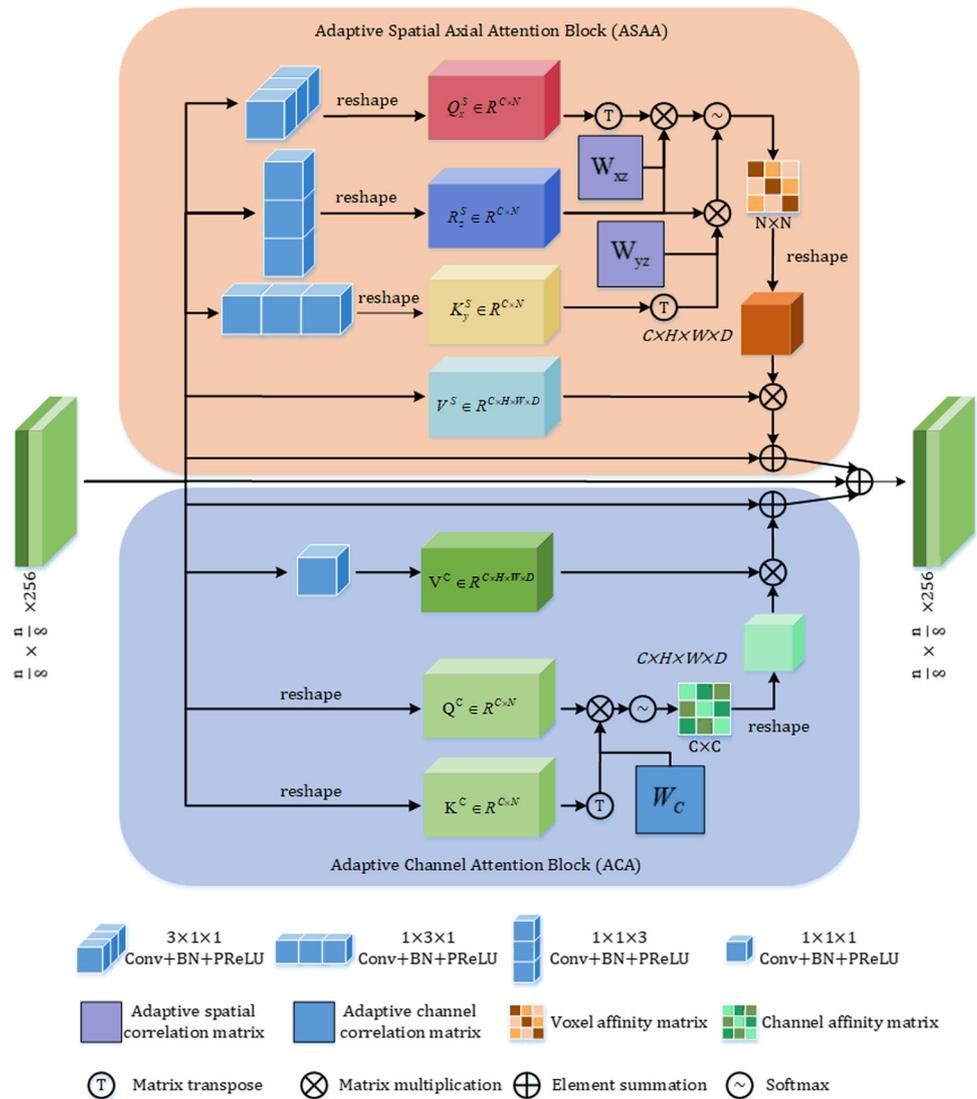
When DCNN learns each object (organ or tumor) in the three-dimensional medical image, the network has poor learning effect on the detail information of the object. Consequently, this situation leads to a network that is less sensitive to details of objects in adjacent slices, especially tissue adhesions and small blood vessels. Motivated by the success of the self-attentive mechanism [26] in mining long-range dependencies, we design an adaptive dual-attention module inspired by [29].



**Fig. 3** Shadow mapping layer ( $M$  represents the number of channels after convolution). The core features are obtained by normal convolution operation, then shadow features are obtained by linear mapping on the

basis of the core features. After concatenating these two features to get the output of the shadow mapping layer

**Fig. 4** Structure of adaptive dual-branch attention module



This module adaptively computes the positional correlation between voxels based on high-level features extracted by the encoder, enhancing the learning ability of object details by the network.

Figure 4 is a schematic diagram of our proposed attention module. It has two branches, that we named Adaptive Spatial Axial Attention Block (ASAA) and Adaptive Channel Attention Block (ACA) individually. These two blocks take the output result  $F$  of the encoder as input, then calculate the position correlation information and channel correlation information at the voxel level through the two attention modules of space and channel. In this way, spatial voxel feature  $F_{Spatial}$  and channel voxel feature  $F_{Channel}$  can be obtained respectively. At last, these two features and the high-level feature  $F$  extracted by the encoder are summed as the input of the decoder to better perceive and semantic segmentation of the relations between voxels.

**Adaptive spatial axial attention block** In order to construct global dependent information in features and better learn detailed features in medical images, we first need to construct spatial attention matrix, which can describe the correlation information among voxels. Since the relationship between the number of compatible parameters and the integrity of spatial information is needed, we choose to encode the spatial correlation information of high-level features after coding path processing. Inspired by [36], we use  $3 \times 1 \times 1$ ,  $1 \times 3 \times 1$ , and  $1 \times 1 \times 3$  convolution in 3D space to learn the correlation information in volume data from different directions. Batch normalization and PReLU activation are carried out after the convolution operation. PReLU makes tailored optimization for overfitting during network optimization than ReLU, which ensures the robustness of the deep model without increasing the computational effort. After calculation, we get three feature matrices  $Q_x^S \in R^{C \times H \times W \times D}$ ,  $K_y^S \in R^{C \times H \times W \times D}$  and  $R_z^S \in R^{C \times H \times W \times D}$ , where  $C$  represents the number of input channels.  $H$ ,  $W$ , and  $D$  represent the height, width, and depth of the input data, separately. These matrices take  $H$ ,  $W$ , and  $D$  as axes to learn the correlation information among voxels from three directions, and the constructed spatial attention matrix is used to capture the edge details of nodules. In the next step, we reshape these three feature maps to  $R^{C \times N}$ , where  $N = H \times W \times D$ . First,  $Q_x^{S^T}$  and  $R_z^S$  perform matrix multiplication to fuse the voxel correlation matrices calculated with height direction as axis and depth direction as axis. Next, the same operation is performed for  $K_y^{S^T}$  and  $R_z^S$  to fuse the voxel correlation matrices calculated with width direction as the axis and depth direction as the axis. However, we find that certain locational correlations are difficult to learn, and there are biases in encoding voxel affinity matrices. In the case that the voxel affinity matrix is not accurate enough,  $Q_x^S$ ,  $K_y^S$  and  $R_z^S$  can lead to insufficiently accurate long-range dependencies in the

computation of voxel correlations. As result, we add adaptive spatial correlation matrices  $W_{xz} \in R^{N \times N}$  and  $W_{yz} \in R^{N \times N}$  when performing matrix operations on  $Q_x^{S^T}$ ,  $R_z^S$  and  $K_y^{S^T}$ ,  $R_z^S$  respectively. It can adaptively optimize and control the bias in the calculation of the affinity between the current voxel and other locations in the back propagation, and reduce the influence in the process of voxel affinity coding. Where  $Q_x^{S^T}$  and  $K_y^{S^T}$  are  $Q_x^S$  and  $K_y^S$  transpose. In the end, the voxel affinity matrix  $VAM_{(x,y,z)}$  is calculated by Softmax function, as shown below:

$$VAM_{(x,y,z)} = \text{Softmax} \left[ \left( W_{xz} \times Q_x^{S^T} \times R_z^S \right) \cdot \left( W_{yz} \times K_y^{S^T} \times R_z^S \right) \right] \tag{3}$$

where  $VAM_{(x,y,z)}$  represents the affinity between the voxels at  $x^{th}$ ,  $y^{th}$  and  $z^{th}$  position and other positions. Next, we perform  $1 \times 1 \times 1$  convolution on the input high-level feature  $F$  to obtain a reduced-dimensional feature map  $V^S \in R^{C \times H \times W \times D}$ .

The voxel affinity matrix reshaped by us to obtain  $VAM_{(x,y,z)} \in R^{C \times H \times W \times D}$ , and then perform matrix multiplication between  $V^S$  and  $VAM_{(x,y,z)}$  to obtain the voxel-level attention enhancement feature  $F_S \in R^{C \times H \times W \times D}$ . Finally, the raw high-level features and voxel correlation features are element-wise summed to obtain the output of the ASAA block. In this part, voxel correlation in all directions of 3D data is taken into consideration, and spatial location information is effectively used for segmentation calculation to improve segmentation accuracy.

**Adaptive channel attention block** Each channel contains a kind of unique category characteristic in the high-level features obtained by the encoder part. The category features of different channels are interdependent. In order to calculate the interdependence information of different channels, an adaptive channel attention block is proposed by us to further synthesize the feature representation. We reshape the high-level feature  $F$  extracted from the encoder as  $R^{C \times N}$ . The reshaped features are expressed as  $Q^C \in R^{C \times N}$  and  $K^C \in R^{C \times N}$ , respectively. Next, the channel affinity matrix is calculated by the following formula:

$$CAM = \text{Softmax} \left( Q^C \times K^{C^T} \right) \tag{4}$$

where  $CAM \in R^{C \times C}$  represents the calculated channel affinity matrix. If the similarity of category features is high, the category features between channels can promote each other, otherwise, they can inhibit each other. Similar to the learning process of voxel affinity matrix, we find that there are biases in the learning process of channel similarity, which can lead to the performance degradation of the model. For reducing this kind of biases, similar to the ASAA block, an adaptive

channel correlation matrix  $W_C \in R^{C \times C}$  is added in the process of calculating channel affinity matrix. This matrix can adjust the inter-channel dependence during the back propagation of the model, effectively reduce the bias. At the same time, this matrix also enhances the model's ability to recognize nodule details and background. The calculation method is as follows:

$$CAM = \text{Softmax}(W_C \times Q^C \times K^{C^T}) \quad (5)$$

We use  $1 \times 1 \times 1$  convolution on the high-level feature  $F$  which is extracted by encoder to get  $V^C \in R^{C \times H \times W \times D}$ . At the same time, the channel affinity matrix  $CAM$  is reshaped into  $R^{C \times H \times W \times D}$ . Next,  $CAM$  is multiplied with  $V^C$  to obtain the attention feature of voxel-level channel  $F_C \in R^{C \times H \times W \times D}$ . Similar to the method of adaptive spatial axial attention block, we sum the channel attention feature with the original advanced feature to get the output of ACA block. The details of the model architecture information are shown in Table 5, which we put at the end of the paper due to the length of the table.

Compared with the original dual-attention mechanism [29] the proposed adaptive dual-attention module has a stronger learning ability for long-range dependencies. On the one hand, the convolution of the ASAA module is able to dig more intently into the dependencies on a single direction in the space and aggregate the information on the affinity of the constituent spatial voxels. On the other hand, the adaptive matrices in the ASAA and ACA blocks inhibit the impairment of dependencies by abnormal voxels and abnormal channels.

### 3.4 Loss function

In volume data, pulmonary nodules account for a small proportion compared to the background. To minimize the divergence between predicted and ground truth, we choose Dice Coefficient Loss  $L_{Dice}$  as the loss function, which can reduce the deviation generated in the learning of the central region of pulmonary nodules in the training process. In addition, we also add Boundary Loss [37]  $L_{Boundary}$  to promote the learning of nodules and blood vessels and other fine tissues. The overall loss function of our method is determined as:

$$\mathcal{L}_{Total} = L_{Dice} + L_{Boundary} \quad (6)$$

where Dice Coefficient loss  $L_{Dice}$  and Boundary loss  $L_{Boundary}$  are defined as:

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N p_i g_i + \epsilon}{\sum_{i=1}^N p_i^2 + \sum_{i=1}^K g_i^2 + \epsilon} \quad (7)$$

$$L_{Boundary} = \int_{\Omega} \phi_G(v) s_{\theta}(v) dv \quad (8)$$

In (7),  $N$  represents the number of voxels,  $p_i \in [0, 1]$  indicates the predicted probability of the  $i^{th}$  voxel, and  $g_i \in [0, 1]$  means

the ground truth of the  $i^{th}$  voxel. The parameter  $\epsilon$  is the *Laplacian* smoothing factor, it is used to avoid instabilities of model and to speed up the convergence during the training session ( $\epsilon = 1.0$ ). In (8),  $v$  represents the voxel in volume data,  $\Omega$  is the spatial domain of volume data, and  $\phi_G(v)$  calculates the distance of spatial distribution between the ground truth of the boundary voxel and the prediction, which is defined as:

$$\phi_G(v) = \|v - \text{Pred}(v)\| \quad (9)$$

where  $s_{\theta}(v)$  represents the probability that the voxel  $v$  is predicted to be foreground or background, and  $\|\cdot\|$  represents the  $L_2$  distance.

## 4 Experiments

### 4.1 Datasets

In this study, the LIDC-IDRI open dataset [38] is used for all experiments. The dataset is initiated and collected by the National Cancer Institute and include 1, 018 patients. The CT data of every patient are annotated by four experienced chest radiologists. We screen the dataset to remove the slice data which slice thickness greater than 2.5mm and missing labeling information. In the end, we obtain complete patient data for 862 patients with a total of 1160 nodules. We randomly select 162 patients data as the test set and the rest 700 cases as the training set. We perform a five-fold cross-validation on the training set, testing the model once per fold and taking the average of the five experiments as the final result.

### 4.2 Setting details

Our DAS-Net is realized by PyTorch framework and executed all experiments on a workstation fitted with an NVIDIA Titan XP GPU. The data are treated by some pre-processing method, such as linear mapping the HU values of all CT voxels to  $[0, 1]$ . Then, the nodules are cut as volume centers to size  $16 \times 128 \times 128$  for training and testing. The clipped size of  $16 \times 128 \times 128$  cover the entire range of nodule sizes. To reduce overfitting, we use a simple data enhancement technique involving random scaling and flipping in three directions. During the training process, the batch size and initial learning rate ( $lr$ ) are set at 8 and  $1e - 4$ , correspondingly. Stochastic gradient descent (SGD) is utilized as our optimizer by setting the weight decay to 0.8 and setting epoch to 500.

### 4.3 Assessment indicators

To better demonstrate the effectiveness of the proposed method in this paper, we measure the performance on the LIDC-

IDRI dataset [39] using the Dice coefficient, Sensitivity and Hausdorff distance.

**Dice coefficient** This is a measure of the similarity between the segmented prediction  $V_{seg}$  and the relative ground truth  $V_{GT}$ , defined as follows:

$$Dice = 2 * \frac{V_{seg} \cap V_{GT}}{V_{seg} + V_{GT}} \quad (10)$$

**Sensitivity** This index is used to measure the performance of voxel classification and the correctness of segmentation region. It is defined as follows:

$$Sen = \frac{V_{seg} \cap V_{GT}}{V_{seg}} \quad (11)$$

**Hausdorff distance** It evaluates the quality of segmented surface details by computing the maximized gap between the prediction and the ground truth, which is defined as follows:

$$H(V_{seg}, V_{GT}) = \max_{x \in V_{GT}} \{ \min_{y \in V_{seg}} \{ \|x, y\| \} \} \quad (12)$$

where  $\|*,*\|$  represents the *Euclidean* distance. When the Dice coefficient is large, the sensitivity is large or the Hausdorff distance is small, the segmentation results are more accurate.

## 4.4 Result

### 4.4.1 Comparative experiment

To better illustrate the capability of our method in extracting detailed features, our method is respectively compared with most advanced two kind of 2D segmentation methods and seven 3D methods. These methods are 2D U-Net [18], DA-Net [40], V-Net [39], 3D-CRF [41], 3D-Res2UNet [42], 3D TC U-Net [33], A-Roi algorithm [24], CoLe-CNN [10] and DB-ResNet [43]. DA-Net [40] replaces the convolution with an atrous convolution and uses a dense connection to extract rich features, completing the widening of the filter's field of view without producing losses. V-Net [39] is a classic model in 3D medical image segmentation. This method replaced 2D convolution with 3D convolution operation and introduced residual information to make up for the feature loss caused by the convolution operation. At the same time, convolution operations replaced the original pooling operations which used to up-sampling and down-sampling for reduce the loss when the image resolution changes. 3D-CRF [41] added 3D Conditional Random Field Post-processing algorithm to 3D U-Net for capturing global context information and assisting in better predicted accuracy. 3D-Res2UNet [42] is the combination of Res2Net and 3D-UNet, allowing the network to be more sensitive to small nodule targets. 3D TC U-Net [33] proposed a method to learn the parameters of multiple

convolution cores in the same layer, which effectively reduced the model parameters and calculation amount, and also improved the model segmentation performance. DB-ResNet [43] proposed a new pooling operation to select the intensity features of block center voxels and merge them with the multi-view and multi-scale features of the nodules to jointly solve the segmentation task. This method also added a weighted sampling strategy to select voxels at boundary locations to raise the accuracy of the model. The other three models have been described earlier and not repeated here.

The experimental results of the 2D work demonstrates that the 3D approach is superior in the lung nodule segmentation task, with the 3D approach outperforming them correspondingly by 18.75% and 11.02% in dice score and by 20.31% and 3.61% in sensitivity. Our analysis suggests that the 2D method cannot capture richer inter-layer contextual information to enhance the segmentation effect, so the 3D method is more advantageous. Compared to several other 3D methods, our proposed method is 3.88%, 3.76%, 1.89%, 4.85%, 4.5%, 5.95% and 2.71% higher than the other seven methods in the Dice score, respectively. It means that our proposed DAS-Net can learn the core region of the nodules in volume data better than the other methods. For HD evaluation index, our method is 7.8, 2.92, 1.48, 11.15, 0.85, 4.67 and 0.76 lower than other methods, respectively, which proves that our approach outperforms the other seven approaches for learning surface detail capabilities. The results of indicators Dice Score and HD demonstrated that our method offers better performance than other methods on aspects of surface detail learning. Results confirms that the method proposed by us is effective in resolving the complex issues of nodule segments, such as segmentation of small fine vessels and segmentation of tissue adhesions at the nodule margins. By comparing the standard deviation of Dice Score in Table 1, we consider our approach to be more stable than other approaches.

Figure 5 shows the segmentation results of 2D U-Net compared with our method on three cases of data. From the results, it can be found that the 2D method is more sensitive to the core region of the nodule which could lead failure of capture richer edge information. There are also cases of segmentation errors.

The segmentation results of the 3D method for the three case test data in the LIDC dataset are given in Fig. 6. The results show that all eight methods are very effective in learning the core structure of the nodules, whereas the other seven methods had weaknesses in learning the surface details of the nodules. A comparison of the results in Fig. 6 indicates that the other seven methods are incapable of accurate segmentation of tissue adhesions and small vascular adhesions. Our method extracts more precise features of nodule details, especially to distinguish the adhesion of blood vessels or tissues. In general, our method shows higher segmentation accuracy and better segmentation reliability in terms of Dice Score, Sensitivity and HD, which indicates that the segmentation

**Table 1** Performance comparison of different segmentation methods (mean  $\pm$  standard deviation)

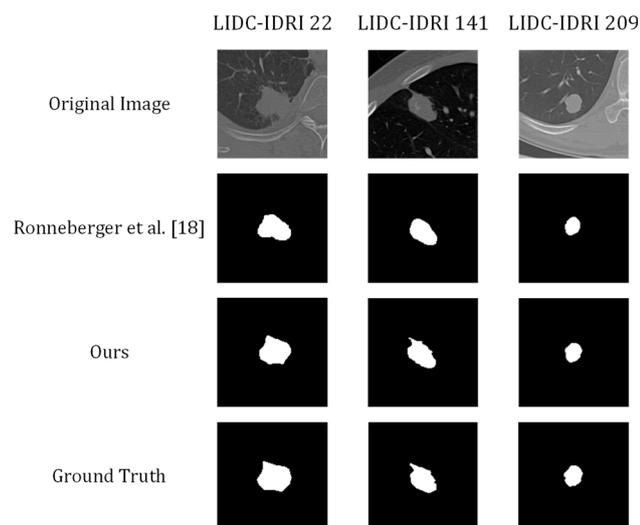
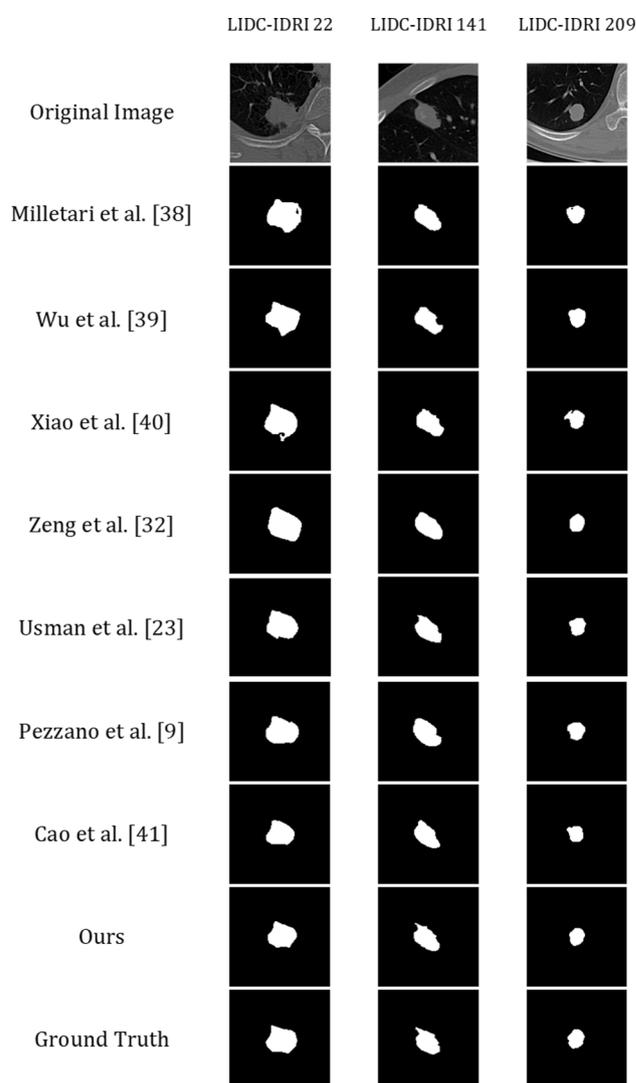
Method	Result		
	Dice score(%)	Sensitivity(%)	Hausdorff distance(mm)
Ronneberger et al. [18]	73.3 $\pm$ 8.36	70.5 $\pm$ 6.51	15.14 $\pm$ 7.24
Maqsood et al. [40]	81.00 $\pm$ 0.00	87.2 $\pm$ 0.00	-
Milletari et al. [39]	88.17 $\pm$ 7.93	91.52 $\pm$ 9.30	11.73 $\pm$ 6.62
Wu et al. [41]	88.29 $\pm$ 4.37	91.20 $\pm$ 6.58	5.48 $\pm$ 2.56
Xiao et al. [42]	90.16 $\pm$ 8.84	89.85 $\pm$ 10.24	5.41 $\pm$ 3.06
Zeng et al. [33]	87.20 $\pm$ 10.73	85.39 $\pm$ 10.68	15.08 $\pm$ 8.97
Usman et al. [24]	87.55 $\pm$ 10.58	<b>91.63<math>\pm</math>8.47</b>	4.78 $\pm$ 2.21
Pezzano et al. [10]	86.10 $\pm$ 8.59	90.00 $\pm$ 9.46	8.60 $\pm$ 7.10
Cao et al. [43]	89.34 $\pm$ 10.19	90.27 $\pm$ 11.79	4.69 $\pm$ 4.11
Ours	<b>92.05<math>\pm</math>3.08</b>	90.81 $\pm$ 6.35	<b>3.93<math>\pm</math>1.87</b>

effect of the proposed method is obviously superior to other methods.

#### 4.4.2 Ablation study

In this paper, we replace the traditional convolutional layer in CNN by using shadow mapping layer to diminish the model size, and introduce an adaptive dual-attention module to strengthen the model's perception of nodules detail information. The efficiency of our approach is evaluated on a local dataset.

Table 2 lists the performance of DAS-Net with or without the shadow mapping layer and the adaptive dual-attention module. The dual-branch attention module we proposed extracts feature information from two perspectives of channel and spatial to construct an attention matrix to enhance the model's learning of edge detail information. We also verify the contributions of the two attention modules separately and

**Fig. 5** Comparison of our method with 2D U-Net results**Fig. 6** The comparison results for 3d segmentation methods (we selected three cases from the test set for comparison, and the order of the methods is V-Net, 3D-CRF, 3D-Res2UNet, 3D TC U-Net, A-Roi algorithm, CoLe-CNN, DB-ResNet, Ours and Ground Truth)

**Table 2** Comparison of model performance with/without shadow mapping and adaptive dual attention module on the LIDC dataset (among the table, SML, ASAA and ACA represent shadow mapping layer, adaptive spatial axial attention block and adaptive channel attention block, separately)

Model name	SML	ASAA	ACA	Dice score (%)	Sensitivity (%)	Hausdorff distance (mm)
baseline	–	–	–	87.41±2.81	89.70±4.83	8.63±3.39
baseline+SML	✓	–	–	86.52±3.47	89.44±1.68	9.25±1.42
baseline + DA	–	✓	–	90.63±3.60	91.31±7.80	3.34±1.12
baseline + DA	–	–	✓	89.65±2.50	90.47±4.14	5.32±1.31
baseline + DA	–	✓	✓	91.21±2.33	92.12±5.84	2.54±1.11
DAS-Net	✓	✓	✓	<b>92.05±3.08</b>	<b>90.81±6.35</b>	<b>3.93±1.87</b>

present the results in Table 2. We use 3D U-Net as baseline and control variable method to keep all settings unchanged except for the contrast part.

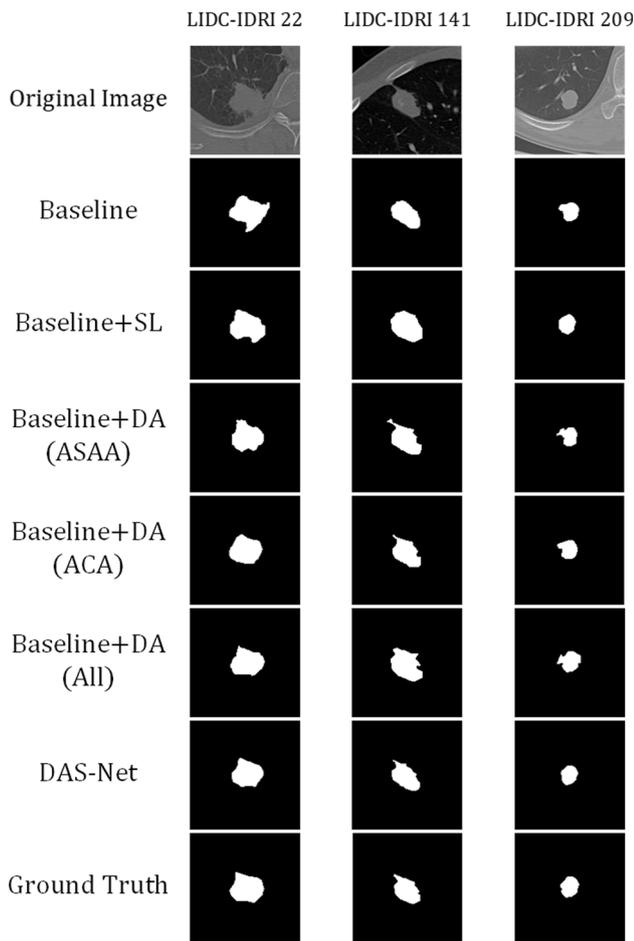
The results show that the designed adaptive dual attention module helps to improve the effectiveness of the model in segmentation of lung nodules. The role of shadow mapping layer will be analyzed in the next section. Compared with the baseline model, the model after adding the adaptive dual

attention module has improved 3.8% in Dice score and reduced 4.7 in Hausdorff distance. Using the two attention modules alone can improve the baseline performance to a certain extent. The model integrating the two attention modules outperforms the other three models in all evaluation metrics. These results prove the effectiveness of the designed dual attention module.

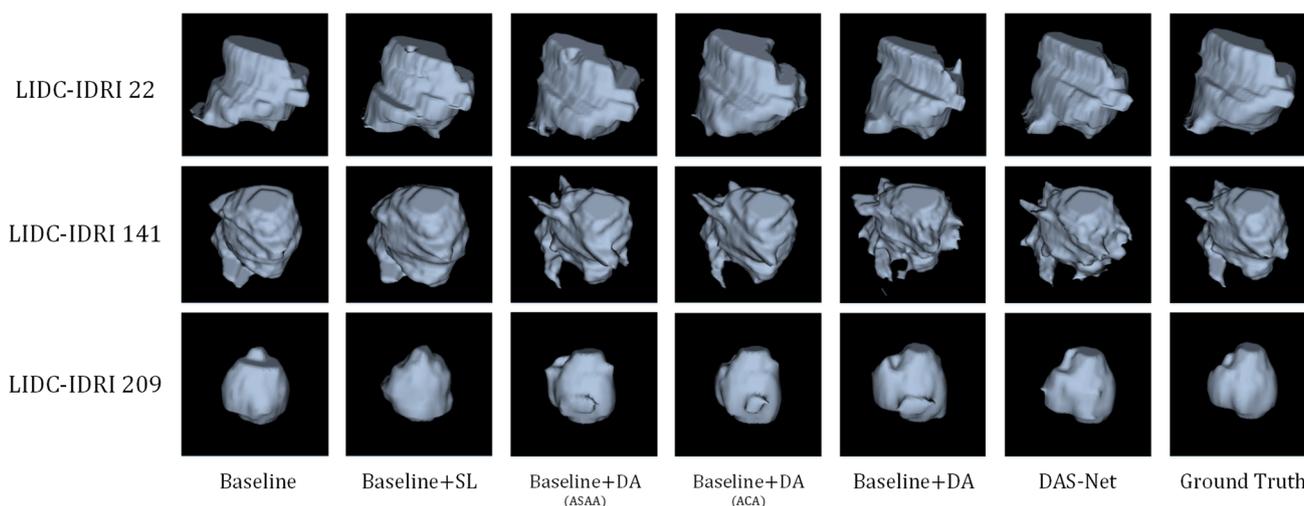
Adding the segmentation results and 3D visualization of each module as shown in Figs. 7 and 8. The baseline model is better for learning the core area of the nodules, but it loses a lot of detailed features of the nodules, and there are situations where the tissue adhesion cannot be distinguished. The model with the adaptive dual attention module can better learn the detailed characteristics of the nodules, especially for the case of tissue adhesion. However, we find that the model with the adaptive dual attention module learns too many details of the vascular tissue, resulting in a big difference from ground truth. This may be the reason for the lower Dice score. The results obtain by DAS-Net are closer to ground truth. The quantitative evaluation of Table 2 and the qualitative evaluation of Figs. 6 and 7 all prove the effectiveness of the DAS-Net proposed in this paper in the segmentation of three-dimensional lung nodules.

Ablation research of shadow mapping feature ratio Our DAS-Net changes the convolution operation to shadow mapping. The ratio of the core feature to the number of shadow features is controlled by the parameter  $r$ . We conduct 4 experiments on the LIDC dataset, and set  $r$  to 1, 2, 3, and 4, separately. The proportions of core features and shadow features were 1:0, 1:1, 1:2, and 1:3. We add shadow mapping with different proportions on the basis of the baseline. The comparisons are made from the four aspects of model parameters, calculation time per epoch, Dice score and Hausdorff distance. The results are shown in Table 3.

From the results in Table 3, it can be found that as the number of core features decreases, the amount of model parameters and Dice score decrease, while HD increases, and the instability of the model also increases. Through the analysis results, we believe that the reason for the fluctuation of the calculation time is that the linear mapping process of shadow



**Fig. 7** Visualized results of ablation study (one of the three samples was randomly selected from the test set for comparison)



**Fig. 8** 3D visualization of the results of the ablation study (the 3D visualization angles of different models of the same nodule are the same)

mapping is performed on the basis of convolution, which causes the calculation time to be slightly higher than the baseline. As the number of core features decreases, the diversity of features decreases. The network learns too much redundant information, which leads to a decrease in Dice score. We balance the feature diversity and model parameters. We think that the feature diversity is more abundant when the feature ratio is 1:1, which is more suitable for the advancement of segmentation tasks.

**Ablation research of attention module** In the process of constructing the adaptive dual attention module, we try three different combinations: connecting the adaptive spatial axial attention block and the adaptive channel attention block in parallel; connecting the two blocks serially in different sequences. We implement three combinations of these two blocks on the basis of the baseline, and the comparison results are given in Table 4.

The number in ASAA block and ACA block in Table 4 indicate the connection sequence. The results reveal that all three connection ways can enhance the performance of the model, but parallel connection has the strongest effect on the model performance. Therefore, we choose to construct our DAS-Net in parallel connection. As far as we are concerned, because the previous attention block will omit part of the

information that is useless for the current block in the process of extracting information, the effect of the serial connection method is worse than the parallel connection. However, the missing information is not useless for the latter block. The effect of the model will be reduced. In the end, we combine the two attention blocks in parallel connection to effectively help the model strengthen the learning of boundary information, thereby improving segmentation performance.

## 5 Conclusion

In this paper, we propose DAS-Net with 3D adaptive dual attention module and 3D shadow mapping to improve the ability of DCNN to learn the edges and details of nodules, thereby improving the accuracy of segmenting lung volume medical images. We evaluate our model on the LIDC-IDRI dataset. Our results demonstrate that our DAS-Net outperforms current most advanced methods in the task of lung nodule segmentation. The ablation study partially confirmed the validity of the designed module. In the future, we intend to detect nodule targets in lung space by video target detection algorithms and use them as input to our current method. On this basis, we will study how to design an end-to-end 3D lung

**Table 3** Core features and shadow feature ratio settings (In the table below, Feature Ratio means *core features* : *shadow features*)

Index	Feature Ratio	Parameters	Computation Time Per Epoch(second)	Dice score(%)	Hausdorff distance (mm)
1	1 : 0	$16.2 \times 10^6$	63	87.41±2.81	8.63±3.39
2	1 : 1	$8.13 \times 10^6$	69	86.52±3.47	9.25±1.42
3	1 : 2	$5.49 \times 10^6$	65	83.12±7.26	10.89±5.20
4	1 : 3	$4.10 \times 10^6$	50	72.56 ±11.38	14.76±6.26

**Table 4** Comparison results of the three combinations of adaptive dual attention modules

Index	ASAA	ACA	Dice score(%)	Sensitivity(%)	Hausdorff distance(mm)
1	1	1	91.21±2.33	92.12±5.84	2.54±1.11
2	1	2	90.47±3.40	92.01±4.72	3.73±2.02
3	2	1	89.76±1.85	90.50±4.50	4.18±1.84

**Table 5** Model architecture information

Layer Name	Output Size	Structure
Ghost_layer_1	[-1, 32, 4, 128, 128]	Conv3d(3 × -3 × -3, stride=1, padding=1), BatchNorm3d, PReLU
Ghost_layer_2	[-1, 64, 4, 128, 128]	Conv3d(3 × -3 × -3, stride=1, padding=1, groups=16), BatchNorm3d, PReLU Conv3d(3 × -3 × -3, stride=1, padding=1, groups=32), BatchNorm3d, PReLU
MaxPool3d_layer_3	[-1, 65, 4, 64, 64]	AdaptiveMaxPool3d
Ghost_layer_4	[-1, 64, 4, 64, 64]	Conv3d(3 × -3 × -3, stride=1, padding=1), BatchNorm3d, PReLU Conv3d(3 × -3 × -3, stride=1, padding=1, groups=32), BatchNorm3d, PReLU
Ghost_layer_5	[-1, 128, 4, 64, 64]	Conv3d(3 × -3 × -3, stride=1, padding=1), BatchNorm3d, PReLU Conv3d(3 × -3 × -3, stride=1, padding=1, groups=64), BatchNorm3d, PReLU
MaxPool3d_layer_6	[-1, 192, 4, 32, 32]	AdaptiveMaxPool3d
Ghost_layer_7	[-1, 128, 4, 32, 32]	Conv3d(3 × -3 × -3, stride=1, padding=1), BatchNorm3d, PReLU Conv3d(3 × -3 × -3, stride=1, padding=1, groups=64), BatchNorm3d, PReLU
Ghost_layer_8	[-1, 256, 4, 32, 32]	Conv3d(3 × -3 × -3, stride=1, padding=1), BatchNorm3d, PReLU Conv3d(3 × -3 × -3, stride=1, padding=1, groups=128), BatchNorm3d, PReLU
MaxPool3d_layer_9	[-1, 384, 4, 16, 16]	AdaptiveMaxPool3d
Conv1 × -1_layer_10	[-1, 256, 4, 16, 16]	Conv3d(1 × -1 × -1, stride=1)
ASAA_Block_11	[-1, 256, 4, 16, 16]	Conv3d(1 × -3 × -1, stride=1, padding=(0, 1, 0)) Conv3d(3 × -1 × -1, stride=1, padding=(1, 0, 0)) Conv3d(1 × -1 × -3, stride=1, padding=(0, 0, 1)) Softmax(dim=-1)
ACA_Block_12	[-1, 256, 4, 16, 16]	Conv3d(1 × -1 × -1, stride=1) Softmax(dim=-1)
Fusing_layer_13	[-1, 256, 4, 16, 16]	Conv3d(1 × -1 × -1, stride=1)
UpSample_layer_14	[-1, 256, 8, 32, 32]	Upsample(scale_factor=2.0, mode=trilinear)
MaxPool3d_layer_15	[-1, 256, 4, 32, 32]	AdaptiveMaxPool3d
Ghost_layer_16	[-1, 256, 4, 32, 32]	Conv3d(3 × -3 × -3, stride=1, padding=1), BatchNorm3d, PReLU Conv3d(3 × -3 × -3, stride=1, padding=1, groups=128), BatchNorm3d, PReLU
Ghost_layer_17	[-1, 256, 4, 32, 32]	Conv3d(3 × -3 × -3, stride=1, padding=1), BatchNorm3d, PReLU Conv3d(3 × -3 × -3, stride=1, padding=1, groups=128), BatchNorm3d, PReLU
UpSample_layer_18	[-1, 256, 8, 64, 64]	Upsample(scale_factor=2.0, mode=trilinear)
MaxPool3d_layer_19	[-1, 256, 4, 64, 64]	AdaptiveMaxPool3d
Ghost_layer_20	[-1, 128, 4, 64, 64]	Conv3d(3 × -3 × -3, stride=1, padding=1), BatchNorm3d, PReLU Conv3d(3 × -3 × -3, stride=1, padding=1, groups=64), BatchNorm3d, PReLU
Ghost_layer_21	[-1, 128, 4, 64, 64]	Conv3d(3 × -3 × -3, stride=1, padding=1), BatchNorm3d, PReLU Conv3d(3 × -3 × -3, stride=1, padding=1, groups=64), BatchNorm3d, PReLU
UpSample_layer_22	[-1, 128, 8, 128, 128]	Upsample(scale_factor=2.0, mode=trilinear)
MaxPool3d_layer_23	[-1, 128, 4, 128, 128]	AdaptiveMaxPool3d
Ghost_layer_24	[-1, 64, 4, 128, 128]	Conv3d(3 × -3 × -3, stride=1, padding=1), BatchNorm3d, PReLU Conv3d(3 × -3 × -3, stride=1, padding=1, groups=32), BatchNorm3d, PReLU
Ghost_layer_25	[-1, 2, 4, 128, 128]	Conv3d(1 × -1 × -1, stride=1, padding=1), BatchNorm3d, PReLU Conv3d(1 × -1 × -1, stride=1, padding=1), BatchNorm3d, PReLU
Activation_layer_26	[-1, 2, 4, 128, 128]	Sigmoid

nodule detection segmentation model to better complete the lung nodule segmentation task.

**Acknowledgements** We are very appreciative to every editor and reviewer for their comments and suggestions on this paper. These comments have greatly promoted the improvement of this paper. This work is supported by the National Natural Science Foundation of China grant numbers 61872261, the National Natural Science Foundation of China grant numbers 61972274 and the Taiyuan Science and Technology Project Plan numbers XE2020-5-04.

## Declarations

**Conflicts of interest** The authors declare that they have no conflict of interest.

## References

1. W. H. Organization, et al (2020) Latest global cancer data: cancer burden rises to 19.3 million new cases and 10.0 million cancer deaths in 2020, Press release
2. NLSTR Team (2011) Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine* 365(5):395–409
3. Liu Q (2019) Understanding the global cancer statistics 2018: implications for cancer control. *Science China Life Sciences* 64(6):1017–1020
4. Zieliński KW, Kulig A, Zieliński J (1984) Morphology of the microvascular bed in primary human carcinomas of lung: Part ii. morphometric investigations of microvascular bed of lung tumors. *Pathology-Research and Practice* 178(4):369–377
5. Nasim F, Sabath B, Eapen G (2019) Lung cancer. *Medical Clinics of North America* 103(3):463–473
6. Savic M, Ma Y, Ramponi G, Du W, Peng Y (2021) Lung nodule segmentation with a region-based fast marching method. *Sensors* 21(5): 1908
7. Suji RJ, Bhadouria SS, Dhar J, Godfrey WW (2020) Optical flow methods for lung nodule segmentation on ldr-ldr images. *Journal of Digital Imaging* 33(5):1306–1324
8. Pawar SP, Talbar SN (2021) Lungseg-net: Lung field segmentation using generative adversarial network. *Biomedical Signal Processing and Control* 64:102296
9. Wu Z, Zhou Q, Wang F (2021) Coarse-to-fine lung nodule segmentation in ct images with image enhancement and dual-branch network. *IEEE Access* 9:7255–7262
10. Pezzano G, Ripoll VR, Radeva P (2021) Cole-cnn: Context-learning convolutional neural network with adaptive loss function for lung nodule segmentation. *Computer Methods and Programs in Biomedicine* 198:105792
11. Liu C, Pang M (2020) Extracting lungs from ct images via deep convolutional neural network based segmentation and two-pass contour refinement. *Journal of Digital Imaging* 33(6):1465–1478
12. Sahu P, Zhao Y, Bhatia P, Bogoni L, Jerebko A, Qin H (2020) Structure correction for robust volume segmentation in presence of tumors. *IEEE Journal of Biomedical and Health Informatics* 25(4):1151–1162
13. Nishio M, Muramatsu C, Noguchi S, Nakai H, Fujimoto K, Sakamoto R, Fujita H (2020) Attribute-guided image generation of three-dimensional computed tomography images of lung nodules using a generative adversarial network. *Computers in Biology and Medicine* 126:104032
14. Sun Y, Tang J, Lei W, He D (2020) 3d segmentation of pulmonary nodules based on multi-view and semi-supervised. *IEEE Access* 8: 26457–26467
15. Hesamian MH, Jia W, He X, Wang Q, Kennedy PJ (2021) Synthetic ct images for semi-sequential detection and segmentation of lung nodules. *Applied Intelligence* 51(3):1616–1628
16. Dong X, Xu S, Liu Y, Wang A, Saripan MI, Li L, Zhang X, Lu L (2020) Multi-view secondary input collaborative deep learning for lung nodule 3d segmentation. *Cancer Imaging* 20(1):53–66
17. Zebin T, Rezvy S (2021) Covid-19 detection and disease progression visualization: Deep learning on chest x-rays for classification and coarse localization. *Applied Intelligence* 51(2):1010–1021
18. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. *International conference on medical image computing and computer-assisted intervention - MICCAI 2015*:234–241
19. Kushnure DT, Talbar SN (2021) Ms-unet: A multi-scale unet with feature recalibration approach for automatic liver and tumor segmentation in ct images. *Computerized Medical Imaging and Graphics* 89:101885
20. Fang Y, Chen C, Yuan Y, Tong K-Y (2019) Selective feature aggregation network with area-boundary constraints for polyp segmentation. *Medical image computing and computer assisted intervention - MICCAI 2019*:302–310
21. Perslev M, Dam EB, Pai A, Igel C (2019) One network to segment them all: A general, lightweight system for accurate 3d medical image segmentation. In: *Medical image computing and computer assisted intervention - MICCAI 2019*. pp 30–38
22. Rickmann A-M, Roy AG, Sarasua I, Navab N, Wachinger C (2019) 'project & excite' modules for segmentation of volumetric medical scans. *Medical image computing and computer assisted intervention - MICCAI 2019*:39–47
23. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. pp 7132–7141
24. Usman M, Lee B-D, Byon S-S, Kim S-H, Lee B-I, Shin Y-G (2020) Volumetric lung nodule segmentation using adaptive roi with multi-view residual learning. *Scientific Reports* 10(1):1–15
25. Zhang Z, Liu Q, Wang Y (2018) Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters* 15(5):749–753
26. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. In: *Proceedings of the 31st international conference on neural information processing systems*. pp 6000–6010
27. Wang X, Girshick R, Gupta A, He K (2018) Non-local neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. pp 7794–7803
28. Huang Z, Wang X, Huang L, Huang C, Wei Y, Liu W (2019) Ccnet: Criss-cross attention for semantic segmentation. In: *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*. pp 603–612
29. Fu J, Liu J, Tian H, Li Y, Bao Y, Fang Z, Lu H (2019) Dual attention network for scene segmentation. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*. pp 3141–3149
30. Sinha A, Dolz J (2020) Multi-scale self-guided attention for medical image segmentation. *IEEE Journal of Biomedical and Health Informatics* 25(1):121–130
31. Park J, Woo S, Lee J-Y, Kweon IS (2018) Bam: Bottleneck attention module. [arXiv:1807.06514](https://arxiv.org/abs/1807.06514)
32. Yu F, Koltun V (2015) Multi-scale context aggregation by dilated convolutions. [arXiv:1511.07122](https://arxiv.org/abs/1511.07122)
33. Zeng G, Zheng G (2019) 3d tiled convolution for effective segmentation of volumetric medical images. *Medical image computing and computer assisted intervention - MICCAI 2019*:146–154

34. Han K, Wang Y, Tian Q, Guo J, Xu C, Xu C (2020) Ghostnet: More features from cheap operations. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). pp 1577–1586
35. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O (2016) 3d u-net: learning dense volumetric segmentation from sparse annotation. *Medical image computing and computer-assisted intervention - MICCAI 2016*:424–432
36. Ho J, Kalchbrenner N, Weissenborn D, Salimans T, Axial attention in multidimensional transformers. [arXiv:1912.12180](https://arxiv.org/abs/1912.12180)
37. Kervadec H, Bouchtiba J, Desrosiers C, Granger E, Dolz J, Ayed IB (2021) Boundary loss for highly unbalanced segmentation. *Medical Image Analysis* 67:101851
38. Armato SG III, McLennan G, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, Zhao B, Aberle DR, Henschke CI, Hoffman EA et al (2011) The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical physics* 38(2):915–931
39. Milletari F, Navab N, Ahmadi S-A (2016) V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). IEEE, pp 565–571
40. Maqsood M, Yasmin S, Mehmood I, Bukhari M, Kim M (2021) An efficient da-net architecture for lung nodule segmentation. *Mathematics* 9 (13)
41. Wu W, Gao L, Duan H, Huang G, Ye X, Nie S (2020) Segmentation of pulmonary nodules in ct images based on 3d-unet combined with three-dimensional conditional random field optimization. *Medical Physics* 47(9):4054–4063
42. Xiao Z, Liu B, Geng L, Zhang F, Liu Y (2020) Segmentation of lung nodules using improved 3d-unet neural network. *Symmetry* 12(11):1787
43. Cao H, Liu H, Song E, Hung C-C, Ma G, Xu X, Jin R, Lu J (2020) Dual-branch residual network for lung nodule segmentation. *Applied Soft Computing* 86:105934

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.