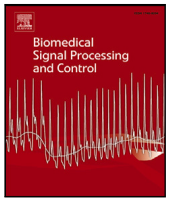
Biomedical Signal Processing and Control 93 (2024) 106183

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



A detail-oriented super-2D network for pulmonary artery segmentation

Lufei Lou, Yu Xin *, Jiangbo Qian, Yihong Dong

Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China

ARTICLE INFO

Keywords:
Deep learning
Medical image segmentation
Pulmonary vascular segmentation
Super-2D
Fine vessel

ABSTRACT

In medical image segmentation, automatic pulmonary vascular segmentation is very important for the diagnosis of pulmonary vascular lesions. However, the pulmonary vascular has problems such as a small cross-sectional area and morphological approximation to other pulmonary interstitial tissues, which results in low segmentation accuracy for fine vessels. Moreover, the 3D and 2D networks applied to medical image segmentation have high computational costs and low segmentation accuracies, respectively. To address these issues, we propose a detail-oriented super-2D network, named DS2Net, that focuses on the semantic reinforcement of fine vessels and computational cost optimization. Aiming at the high computational costs of 3D networks and the low segmentation accuracies of 2D networks, a super-2D segmentation pattern is adopted, which uses consecutive computed tomography (CT) slices as inputs to build 3D correlations in 2D network and provides volume spatial information to the network with a lower computational cost. Then, to achieve improved segmentation accuracy for fine vessels in the pulmonary vascular segmentation task, we propose an OffsetConv algorithm, which extends the feature scope of fine vessels via convolutional diffusion combined with morphological diffusion and solves the problem regarding the ease of fine vessel feature loss during multi scale transformation. Qualitative and quantitative experimental results obtained on the public Parse22 pulmonary artery segmentation dataset show that our proposed DS2Net is superior to other 2D networks that are widely used in medical image segmentation and achieves segmentation accuracy close to that of the 3D benchmark network under the premise of a much lower computational cost. Concretely, our DS2Net achieves average DSC scores of 78.16% for the overall CT slices and 61.53% for a single CT slice on Parse22, and it significantly surpasses other networks in terms of fine vessel segmentation. Code is available at https://github.com/LuffyMonsterB/DS2Net.

1. Introduction

In recent years, pulmonary vascular disease and pulmonary hypertension have become some of the diseases with the highest morbidity and mortality rates. Traditional lung examination methods are very invasive to patients and only reflect the health of the whole lung, making it difficult to describe localized pulmonary vessel lesions. This is detrimental to the early diagnosis and treatment of lung tissue diseases. With the consecutive development of CT imaging technology, thoracic CT has also become one of the main tools for diagnosing pulmonary vessel lesions. However, diagnosing pulmonary vessel diseases via manual reading alone is very time-consuming and inefficient, and it requires quite a high professional ability level, which is greatly influenced by human factors. Therefore, it is of great clinical significance to use automatic pulmonary vascular segmentation to obtain a three-dimensional reconstruction of the pulmonary vessel structure.

Automatic pulmonary vascular segmentation first requires pulmonary vascular label datasets, but pulmonary arteries need to be manually annotated by professional doctors on a slice-by-slice basis, which is a time-consuming and labour-intensive process that demands high professionalism. As shown in Fig. 1's first row, the pulmonary vascular has a small cross-sectional area in planar space and is similar to interstitial tissues, which makes automatic segmentation prone to the loss of fine vessel features and the misclassification of pulmonary vessels as other interstitial tissues. Secondly, in the second row of Fig. 1, a 3D thoracic CT slice group is shown together with 3D pulmonary artery label and 3D segmentation results obtained using 3D U-Net. It can be observed that the pulmonary artery has a complex tree-like branching structure in 3D space, with the lung region containing a large number of intricate fine vessels and the pericardial region having a larger branching arterial trunk. The traditional 3D segmentation network finds it difficult to consider the segmentation of large and small vessels simultaneously, resulting in branch omission and discontinuity in the segmentation results at the distal fine vessels, and a complete segmentation result cannot be obtained. Therefore, automatic pulmonary vascular segmentation is a very challenging task.

* Corresponding author.
E-mail address: xinyu@nbu.edu.cn (Y. Xin).

https://doi.org/10.1016/j.bspc.2024.106183
Received 9 May 2023; Received in revised form 18 February 2024; Accepted 28 February 2024
Available online 11 March 2024

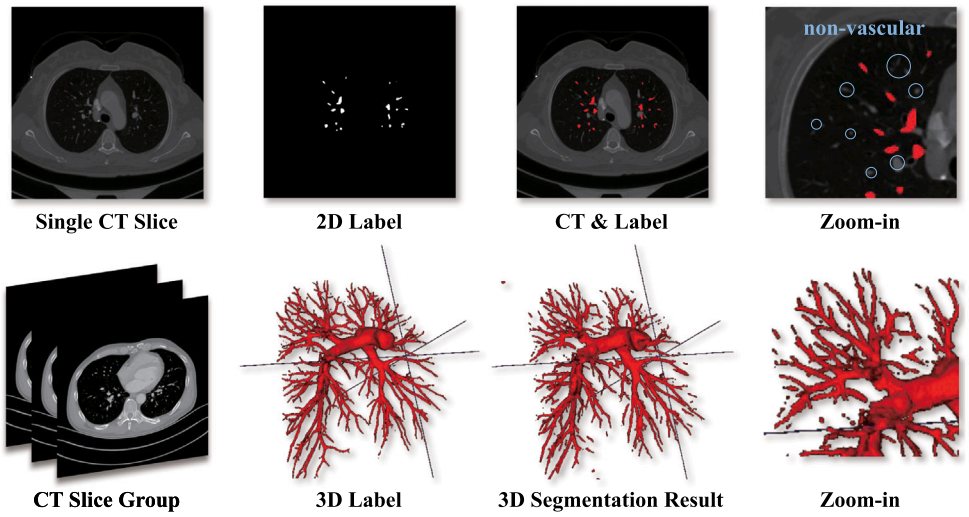


Fig. 1. Thoracic CT slice and pulmonary vascular label. The first row is a single CT slice and its 2D vascular label. The second row is the group of CT slices and their 3D vascular label, as well as the 3D segmentation results from 3D U-Net.

In the early years, machine learning algorithms based on image processing were mainly used for the task of pulmonary vessel segmentation, and they started with texture morphology extraction and CT intensity value discrimination for pulmonary vessels. Moccia et al. [1] summarized the applications of machine learning in pulmonary vessel segmentation. These works mainly included AdaBoost classifier training [2], the implementation of vessel segmentation via a logistic regression classifier [3], the application of a scale-space enhancement filter based on a Hessian matrix [4], and the pioneering use of a convolutional neural network for the classification of pulmonary veins and arteries by Nardelli et al. [5]. However, such methods have poor construction capabilities in planar space, cannot handle complex nonlinear relationships and cannot extract spatial volume information from CT images, resulting in low accuracy and efficiency for pulmonary vessel segmentation. Therefore, an increasing number of studies have started to use deep learning methods to solve the task of pulmonary vascular segmentation.

Recently, medical image segmentation methods based on deep learning, such as high-precision brain tumour segmentation based on the implementation of multimodal medical images and retinal vessel segmentation based on the implementation of morphology, have also made good achievements. They mainly use U-Net [6] and FCNs [7] as the base structures for improvement, and some studies have also used cascading frameworks that operate from coarse to fine [8].

Despite the progress attained by such methods, applying them to the task of pulmonary vascular segmentation still suffers from the following problems. (1) 2D networks are limited by the lack of information between slices and cannot build 3D feature correlations, resulting in pulmonary vascular segmentation results with poor spatial continuity; (2) 3D networks or two-stage networks require high computational costs; and (3) such methods require multi scale transformations during feature extraction, and the fine vessel features are easily lost during their downsampling process, resulting in poor segmentation accuracy for the fine vessels. In this regard, a detail-oriented super-2D segmentation network is proposed in this paper, and the contributions of this paper are listed as follows.

(1) In this paper, we design a segmentation network called DS2Net based on a super-2D input, which can solve the low segmentation accuracy of 2D networks and the high computational costs of 3D segmentation networks, achieving segmentation accuracy close to that of the 3D benchmark network with a lower computational cost.

(2) To solve the problem that the features of fine vessels are easily lost during the multi scale transformation of the segmentation network, an OffsetConv algorithm is designed for the semantic reinforcement of fine vessels, which significantly improves the segmentation accuracy achieved for fine vessels.

(3) To improve the differentiation between the pulmonary vascular and other pulmonary interstitial tissues, we use the PoolFormer module based on the transformer architecture to enhance the shape-position significance of pulmonary vessel features.

(4) To extract spatial volume information from the super-2D input, a squeeze-and-excitation block (hereinafter referred to as the SE block) is used to build 3D correlations in the 2D network.

2. Related work

2.1. Medical image segmentation

Research on 3D networks mainly utilizes U-Net for improvement purposes, and many variants have been extended on this basis. 3D U-Net [9] was proposed by Cicek et al. This model uses 3D convolution to achieve 3D medical image segmentation by inputting consecutive CT slices, and the emergence of 3D U-Net has been a great inspiration for other 3D medical image segmentation methods. Miletari et al. followed up by proposing a 3D variant of the U-Net architecture, V-Net [10], which uses 3D convolution to extract features from 3D inputs and controls the channel dimensionality via $1 \times 1 \times 1$ convolution. Although V-Net yields further improved 3D medical image segmentation performance, it also requires a larger number of parameters and greater computational costs, which is not amenable to general hardware. The nn U-Net [11] is a medical image segmentation framework based on U-Net that was proposed by Isensee et al. in 2018. The nn U-Net does not design a new network architecture but focuses on data pre-processing and training strategies. The emergence of nn U-Net is very beneficial for the rapid completion of medical image segmentation tasks, but its complex configuration and long training time are not conducive to further in-depth research.

Recently, research has also been dedicated to efficient medical image segmentation methods utilizing 2D networks. Zettler et al. [12] observed that the performance of 3D convolution-based U-Nets is not necessarily better than their 2D counterparts. For example, in their study, liver and kidney were significantly better processed using the faster and GPU memory saving 2D U-Nets. In addition, TransFuse [13] was proposed as a CNN combined with a 2D transformer network architecture; this approach runs a CNN-based encoder and a transformer-based encoder in parallel and then uses the proposed fusion module

(BiFusion) to fuse the multilevel features of both branches. However, the transformer branch of TransFuse uses a large-scale compression strategy, which enhances the overall target information, such as its category and contours, but leads to the loss of a large amount of detailed information. For a small target segmentation task in which the target and background are extremely unbalanced, the transformer branch will directly lose the target feature information.

Many studies have proposed using 2.5D segmentation methods to supplement 3D information in 2D networks or applying 2D methods in 3D networks to improve their segmentation accuracy under the premise of controlling computational costs. H-Dense U-Net [14] was applied to liver and liver tumor segmentation; this approach uses 2D segmentation results to provide a priori shape information for 3D CNNs, then uses 3D CNNs to extract spatial information, and finally fuses the 3D features and 2D features as the final segmentation result. APAU-Net [15] is a 3D medical image segmentation network for small targets, but unlike traditional 3D networks, APAU-Net introduces a projection strategy that projects 3D features onto three orthogonal 2D planes to capture the context from different views. In this way, APAU-Net filters out redundant feature information and reduces the loss of critical information for small lesions in 3D scans. APAU-Net achieves improved 3D medical image segmentation performance for small targets and reduces the amount of redundant information contained in 3D features using 2D projections, but the problem is that APAU-Net only performs self-attention learning for 2D planes, and it will inevitably fail to extract sufficient spatial volume information. If it is applied to the pulmonary vascular segmentation task, the impact of insufficient extracted spatial volume information is greater for pulmonary vessel structures with strong spatial continuity.

2.2. Pulmonary vascular segmentation

The main challenge faced by automatic segmentation algorithms for pulmonary vessels is that pulmonary vessels have very similar CT intensity (HU) values to those of other pulmonary interstitial tissues, which is especially difficult for fine vessel segmentation. Although traditional medical image segmentation methods based on HU values are simple and fast, it is very easy to identify other tissues with similar HU values as pulmonary vascular because of the difficulty of dealing with nonlinear data and the effect of image noise [16]. Other studies have optimized HU values by combining them with other methods. Kaftan et al. [17] proposed the concept of fuzzy segmentation and used fuzzy concatenation and thresholding [18] for pulmonary vessel segmentation, but the thresholding method is greatly affected by image noise, and its segmentation effect is poor when the differences between the HU values are small or there is more overlap. Orkisz et al. [19] proposed a variational region growth algorithm with the help of the eigenvalues of the Hessian matrix, which can effectively describe common geometry information, and different geometries present different eigenvalues. However, due to the large-scale differences in pulmonary vessels, a suitable multi scale range must be selected to effectively distinguish pulmonary vessels from non-vessel tissues. Therefore, this method can only enhance pulmonary vessels and cannot directly extract them. MSI-U-Net [20] is a newly released deep learning technique for pulmonary blood vessel 3D segmentation. It uses three decoder branches to extract small-scale, mesoscale, and large-scale blood vessels, correspondingly. These branches are used to increase the accuracy of small blood vessel segmentation by improving the characterization of small blood vessels. Among the three branches of the decoder, we designed a multi-scale information interaction strategy based on migration learning. This approach effectively enhances the correlation of multi-scale vessels in lung CT images. Moreover, to address the issue of small vessel information loss due to downsampling, the method introduces a cross-layer aggregation attention mechanism in the cross-layer connection.

2.3. Pulmonary vessel segmentation

Considering the relatively limited research on pulmonary vascular segmentation in the field of computer vision, we conducted a series of investigations and referred to studies in other categories of tree branching structure segmentation.Yang et al. [21] proposed a retinal blood vessel segmentation network based on the U-Net architecture. To address the challenge of segmenting coarse and fine blood vessels in retinal images, they constructed a multi-task segmentation network that uses a modified U-Net network for feature extraction of both coarse and fine blood vessels. To address the imbalance in the ratio of coarse and fine vessels, they designed a loss function to handle these two different scales. Furthermore, Nasr et al. [22] proposed using pixel block evaluation of images to extract blood vessel regions. This method effectively extracts the blood vessel regions by enhancing contrast through pre-processing and training the CNN with a large number of pixel blocks.

Another study [23] proposed a highly sensitive airway segmentation method for peripheral fine bronchioles to address the severe imbalance between anterior and posterior views in airway segmentation. The method enhances the 3D UNet by introducing a recalibration module [24,25] based on the channel compression incentive mechanism to refine volumetric spatial correlations and improve the semantic feature strength of the small bronchioles with the assistance of an attentional distillation module [26,27]. The method maintains high overall segmentation accuracy and excels in the Branch Detected and Tree-Length Detected metrics, exhibiting higher sensitivity to complex fine bronchial tubes, albeit with a higher false positive rate.

Additionally, Kai et al. [28] observed the effects of local discontinuities and scale differences in substructures when identifying fine bronchi and proposed a 3D multi-scale feature aggregation network (MFA-Net). MFA-Net utilizes multi-scale feature aggregation blocks and parallel null convolution to capture contextual information at various scales, thus enhancing the sensitivity of small bronchial segmentation. Meanwhile, Yun et al. [29] introduced an improved method for automatic airway segmentation of 3D CT images using a 2.5D approach to address the challenging recognition problem posed by small peripheral bronchial targets. The method combines 2D and 3D information in a 2.5D CNN architecture, considering three adjacent slices in the axial, sagittal, and coronal orthogonal directions, respectively. It employs 2D mapping to filter out redundant background information in the 3D features, thus enhancing the focus on small targets. However, this method only learns features in the 2D plane, which is insufficient for extracting volumetric spatial information. If this method is applied to the task of pulmonary vascular segmentation, the lack of volumetric spatial information extraction will significantly impact pulmonary vascular structures with strong spatial continuity.

In summary, a 2.5D segmentation method should serve as the foundation for medical image segmentation due to its practical feasibility. Moreover, addressing the challenging issue of fine vessel segmentation is pivotal for accomplishing the task of pulmonary vascular segmentation. To tackle this challenge, this paper introduces the DS2Net algorithm, which aims to overcome the high computational costs associated with 3D networks, the low segmentation accuracies of 2D networks, and the tendency for fine vessel features to be lost during multi scale transformations.

3. Method

3.1. DS2Net

Thoracic CT images are 3D data consisting of consecutive CT slices, in which the pulmonary vascular exhibits significant spatial volume correlations. Traditional 2D networks only extract features from the spatial planar dimension, so they cannot utilize the correlation information between adjacent slices when performing pulmonary vascular

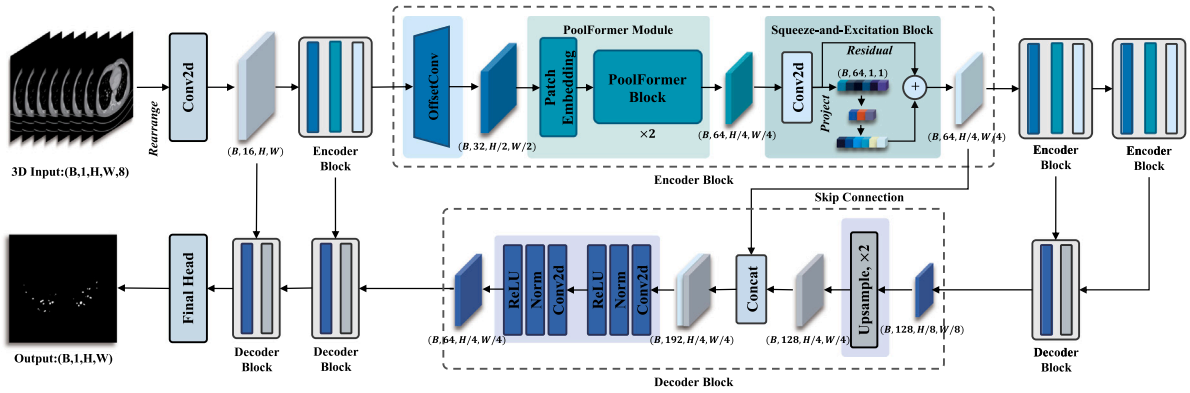


Fig. 2. Overview of the proposed DS2Net for pulmonary vessel segmentation. The rearrange operation is used to convert 3D input into 2.5D feature. DS2Net is designed with four Encoder Blocks and four Decoder Blocks. Each Encoder Block is composed of three components: OffsetConv, PoolFormer Module and SE Block. Additionally, each Decoder Block is comprised of an upsample and a skip connection. The feature size is indicated below the features in each layer.

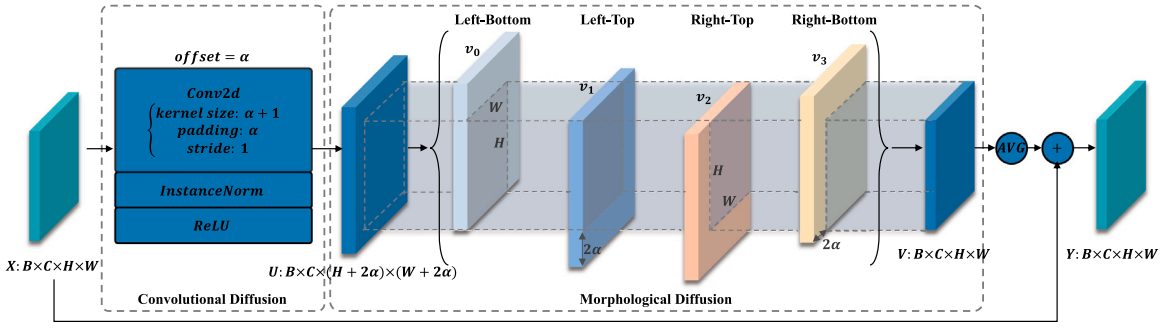


Fig. 3. Overview architecture of the proposed OffsetConv algorithm. When an input feature is provided, convolution is performed using a set α to achieve convolutional diffusion. Subsequently, four tensors are derived, each matching the size of the top-left, top-right, bottom-left, and bottom-right corners, respectively. These four tensors are then aggregated. Finally, the aggregated tensors undergo addition, averaging, and superimposition onto the original map to achieve morphological diffusion.

segmentation and cannot guarantee the spatial continuity of the segmentation results. In contrast, although the 3D processing mechanism of traditional 3D networks can extract more spatial volume information, these methods also require higher computational costs due to the increased dimensionality. In addition, a large number of fine vessels, which have the characteristics of small cross-sectional areas and a scattered distribution, are contained in the pulmonary vascular, and the previously developed medical image segmentation networks cannot accurately segment fine vessels. In this regard, DS2Net is designed to address the loss of fine vessel features and the lack of inter-slice information in the 2D segmentation model for the pulmonary vascular segmentation task. DS2Net uses a super-2D segmentation pattern to enhance the 3D correlations of features and applies semantic enhancement to the fine vessel features to solve the fine vessel feature loss problem and address the high computational costs caused by pure 3D segmentation patterns.

The super-2D segmentation adopted by DS2Net uses consecutive CT slices as inputs and a 2D processing method to achieve vascular segmentation for the most intermediate slice. In this pattern, the input consecutive CT slices are 3D tensors containing five dimensions (batch, channel, length, width, and thickness dimensions), which are not directly applicable to a 2D network. However, the thickness dimension can be directly mapped to the channel dimension due to the single channel feature of a CT slice. As shown in Fig. 2, eight consecutive CT slices as network inputs, and the segmentation target of the network is the label of the fourth CT slice. The original input dimensions of the network are $1 \times H \times W \times D$, where 1 represents the number of original input channels, H represents the height, W represents the width and D represents the slice thickness. Through the "rearrange" operation, the original input combines the channel dimension and the slice thickness dimension into one dimension to obtain the super-2D input required by the network, and its size is $D \times H \times W$. Then, the super-2D input is mapped to C through a 1×1 2D convolutional layer, which contains not only planar spatial information but also rich spatial volume information in the features. As shown in Fig. 4, The first row is the traditional 2D segmentation, where the input is a single CT slice with dimension $C \times H \times W$, which undergoes the 2D segmentation network to obtain a 2D segmentation result with dimension $1 \times H \times W$. The second row is the super-2D segmentation proposed in this paper, where the input is multiple CT slices but the channel dimension and thickness dimension are fused, and the final input dimension is expressed as $(C * D) \times H \times W$, which is passed through the DS2Net, and a 2D segmentation result with the dimension of $1 \times H \times W$ is also obtained. The last line is the 3D segmentation, where the input is a group of 3D CT slices with dimensions $C \times H \times W \times D$, which goes through the 3D segmentation network and gets a 3D segmentation result with dimensions $1 \times H \times W \times D$. Compared with the traditional 2D input, the super-2D input provides the missing inter slice information, which is useful for maintaining the spatial continuity of the pulmonary vascular during the segmentation process. Compared with the traditional 3D input, the super-2D input avoids 3D convolution processing, which greatly reduces the parametric and computational costs.

DS2Net uses U-Net as its basic framework. The encoder–decoder structure of U-Net can efficiently extract the multi scale semantic features of the pulmonary vascular. The encoding phase of DS2Net consists of four encoder blocks at different scales, each of which consists of three components.

(1) OffsetConv Block: This module uses the Offset-Convolution (OffsetConv) algorithm to semantically reinforce the morphological features of fine vessels, avoiding the loss of fine vessel features during multiscale transformations.

(2) PoolFormer Module: The PoolFormer block is used to learn the morphological and spatial characteristics of the pulmonary vascular,

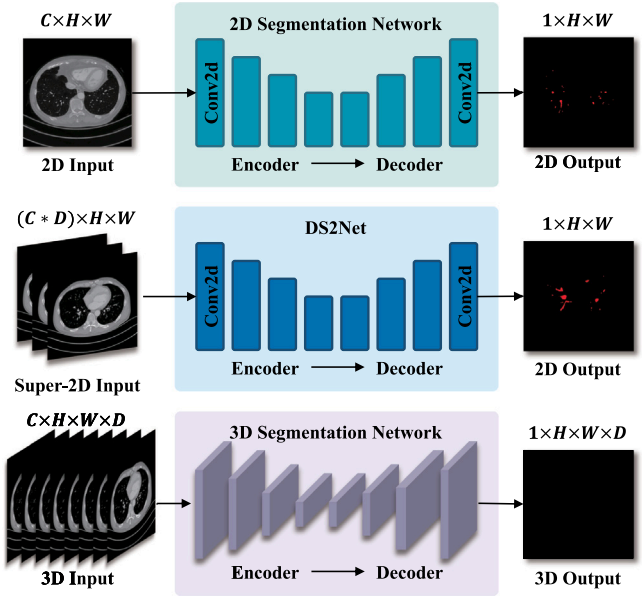


Fig. 4. Comparison of three different types of segmentation patterns.

reinforce the morphological-positional saliency of pulmonary vascular features, and to improve the differentiation of the pulmonary vascular from other tissues (e.g., connective and nerve tissues) in the interstitial lung.

(3) SE Block. This module uses a squeeze-and-excitation (SE) block to incorporate inter slice information into features using a channel SE mechanism to enhance the 3D correlations within the super-2D features.

Encoder blocks semantically reinforce fine vessel features, pulmonary vascular morphological-locational saliency, and 3D correlations using the three OffsetConv block, PoolFormer module, and SE block components mentioned above for thoracic CT images.

The decoder of DS2Net is symmetric to the encoder and is fused with the encoded features of the corresponding level via skip-layer connections. The decoder block upsamples the deeper features and splices them with the encoder block features, after which the spliced features are fed to a convolutional block consisting of two 3×3 convolutional layers and normalized by InstanceNorm. Finally, the 1×1 convolution layer and sigmoid activation function are used to predict the pulmonary vascular.

3.2. OffsetConv in the encoder

The analysis of the 2D pulmonary vascular label shown in Fig. 1 reveals that the pulmonary vascular mainly consists of fine vessels in the lungs and a pulmonary arterial trunk in the mediastinum. The pulmonary vascular trunk has a large cross-sectional area and a fixed position, so segmenting the pulmonary vascular trunk is less difficult. However, the cross-sectional areas of the fine vessels are small and scattered, and because segmentation networks need to extract multi scale features, the fine vessel features are very easy to lose during the downsampling process of the multi scale transformation, resulting in poor segmentation accuracy for the fine vessels. To preserve the features of the fine vessels during the multi scale transformation, we use convolutional diffusion combined with morphological diffusion to extend the feature domain of the fine vessels and avoid the loss of fine vessel features. Convolutional diffusion refers to upsampling the features using convolutional methods, where the sizes of the features are scaled up and the features are diluted. While morphological diffusion refers to hard information scattering via offset superimposed features, such a method does not dilute the features but rather leads to feature unsmoothing. Therefore, the combination of soft convolutional diffusion and hard morphological diffusion is used to propose an OffsetConv algorithm, which can semantically reinforce the morphological features of fine vessels and improve the segmentation accuracy achieved for fine vessels.

The core of the OffsetConv algorithm is a two-dimensional convolution with an additional parameter offset, which is used to control the kernel size and padding size. Fig. 3 shows an intuitive example. Given an input feature $X \in \mathbb{R}^{B\times C\times H\times W}$, we convolve X by convolution with $offset$ set to α to achieve convolutional diffusion, obtaining a feature $U \in \mathbb{R}^{B\times C\times(H+2\alpha)\times(W+2\alpha)}$. Then, four tensors with the same size as X in the top-left, top-right, bottom-left, and bottom-right corners are taken on U to form four tensors v_0, v_1, v_2, v_3, respectively. Finally, the four tensors are added, averaged and then superimposed on the original map X to achieve morphological diffusion, and the superimposed result is used as the output of the algorithm $Y \in \mathbb{R}^{B\times C\times H\times W}$, which can be expressed as:

$$Y = AVG(OffsetConv(X, \alpha)) + X \tag{1}$$

Where $OffsetConv(*, \alpha)$ represents the OffsetConv algorithm and represents the averaging operation. In Fig. 5, the computational procedures for convolutional diffusion and morphological diffusion is illustrated. A 2D convolution with special padding is applied to the feature X, resulting in a set of enlarged features denoted as U. This process, known as convolutional diffusion, induces the diffusion of features from the small blood vessels in X, thereby expanding the feature scope. However, the individual features of these vessels are also diluted. To address this issue, U is divided into four features—v_1, v_2, v_3, and v_4—each of the same size as X. After these features are obtained, their average is calculated, and the result is added back to the original feature X. This process, termed morphological diffusion, effectively mitigates the problem of feature dilution. Furthermore, the feature strength of the fine blood vessels is enhanced while preserving the expanded feature scope.

In DS2Net, the OffsetConv algorithm acts as a semantic fine vessel feature reinforcement mechanism. We use the ratio of the cross-sectional area to the diffusion range as the measure. For features with larger areas, the diffusion range is small in proportion to the area and does not cause a feature coverage problem. For fine vessel features, the diffusion range can reach several times its own value, which can effectively extend the feature scope.

DS2Net contains four encoder blocks for multiscale feature extraction, and the input aspect sizes of the four encoder blocks are $[512 \times 512, 256 \times 256, 128 \times 128, 64 \times 64]$. To guarantee the effectiveness of the offset, the larger the input aspect size is, the larger the offset that needs to be set by the OffsetConv block. Therefore, we set the offset values of the OffsetConv block to $[8, 4, 2, 1]$.

3.3. PoolFormer module in the encoder

In thoracic CT images, both the pulmonary artery and the connective tissues exhibit dot patterns, and their morphological features are relatively significant, but it is difficult to distinguish them from each other only from their morphologies. This leads to the misclassification of the pulmonary artery and connective tissue. Although their morphologies are similar, they possess position differences, so they can be discriminated by location factors. The traditional CNN emphasizes local feature extraction and morphological characteristic learning, and its global construction ability is weak, so it cannot effectively capture position characteristics. A transformer-based network [30] can learn position relationships with its excellent global construction ability and can make full use of the spatial location features of the pulmonary artery and the connective tissues for discrimination purposes. Therefore, we adopt a transformer-based network instead of a CNN-based network. However, when applying the traditional transformer-based

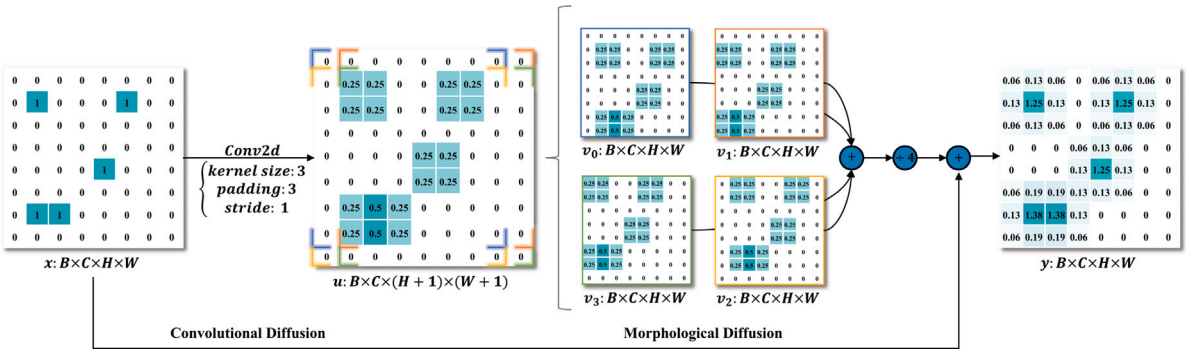


Fig. 5. Computational procedures for convolutional diffusion and morphological diffusion in the OffsetConv algorithm.

network to the pulmonary artery segmentation task, the following two problems need to be solved. (1) For a target segmentation problem with significant morphological features, the token mixer should emphasize the learning of morphological and spatial characteristics; (2) the computational cost brought by the self-attention operation in the transformer is too large. Therefore, we draw on the method proposed by Yu et al. [31] to control the computational complexity by using the pooling operator as a token mixer instead of self-attention. Compared with self-attention, the pooling operation greatly reduces the number of computations needed by the network and does not require training parameters. Moreover, pooling itself is a filtering mechanism that can filter out insignificant features and retain significant pulmonary artery features to a greater extent. Therefore, we adopt the PoolFormer module based on pooling to learn the morphological and spatial characteristics of the pulmonary artery, enhance the morphological-positional saliency of pulmonary artery features, and improve its differentiation from connective tissue.

The PoolFormer module contains a patch embedding (PE) layer and two PoolFormer blocks (PFBs). This component can be represented as:

$$y_{PFB}^{En} = \mathcal{F}_{PFB}^{\times 2}\left(PE\left(y_{OC}^{En}\right)\right) - y_{OC}^{En} \tag{2}$$

where the input $y_{OC}^{En} \in \mathbb{R}^{B\times C\times H\times W}$ of the PoolFormer module is the feature obtained after the OffsetConv block. $PE(*)$ denotes the PE layer, and $\mathcal{F}_{PFB}^{\times 2}(*)$ denotes the two PFBs. The PE layer is a 2D convolution with a kernel size of 2 × 2 and a stride of 2. The main function is to divide the input features into blocks and map more channels. In each PFB, we use AvgPooling as the token mixer in the structure, and the window size is the same as the divided block size of the PE layer. The output of the PoolFormer module is represented as $y_{PFB}^{En} \in \mathbb{R}^{B\times(2C)\times(H/2)\times(W/2)}$, and feature downsampling is implemented in the PE layer.

3.4. SE block in the encoder

Our DS2Net adopts a super-2D pattern with continuous CT slices as inputs, where each layer of the channel dimension represents each slice rather than different mappings of the same image. Therefore, the 3D correlations of pulmonary artery features can be enhanced by extracting the volume space information embedded in the channel dimension. To this end, we draw on the SE block proposed by Hu et al. [32] as the interslice information extraction module in the encoder block. The core idea of the SE block is to fuse the information between channels through its squeeze-excitation mechanism. The traditional 2D input mainly plays the role of building relationships between channels. For the super-2D input, we can rely on the squeeze-excitation mechanism to fuse and refine the information between slices and build the spatial volume relationships, which only brings a minimal computational cost.

In the SE block, we input the features $y_{PFB}^{E} \in \mathbb{R}^{B\times C\times H\times W}$ extracted by the PoolFormer module, which are first compressed into a channel descriptor on the spatial dimension by using global average pooling. To make full use of the compressed feature information, interchannel fusion is then modelled using two fully connected (FC) layers, where the first is a downscaled layer with a scale of 0.5, and the second an upscaled layer that recovers to the original number of input channels. Finally, after a layer of weights is generated by a sigmoid activation function, this layer is then superimposed on the input features via broadcast addition. The output of the coding block $z \in \mathbb{R}^{B\times C\times H\times W}$ can be expressed as follows:

$$Z = \mathcal{F}_{SE}\left(y_{PFB}^{En}\right) + y_{PFB}^{En} \tag{3}$$

where $\mathcal{F}_{SE}(*)$ represents the SE block, and the effectiveness of the SE block is also fully verified in an ablation experiment.

3.5. Loss function

In pulmonary artery mask images, the area of the pulmonary artery is usually small, with a small cross-section and a scattered distribution. During the training process, a large number of background regions may dominate the training direction of the network, leading the network to bias its predictions towards background regions and thus fall into local optimal solutions. Regarding the voxel imbalance between the foreground and background, the traditional solution is to use the Dice loss [10] as the loss function, which only calculates the proportions of positive samples and positive predictions during the training process and can guide the network to favour the prediction of the foreground region. However, the Dice loss still has a serious problem: when only a few masks are contained in the given image, once a partial prediction error is induced, it will lead to a large change in the loss value, resulting in a drastic change in the gradient. The extreme case assumes that there is only one fine vessel in the input CT image, and if the fine vessel is correctly predicted, the loss value will be infinitely close to 0 regardless of how the other regions are predicted, while the loss value will be infinitely close to 1 when an error occurs. If only the Dice loss is used for training guidance, it will lead to unstable loss value oscillation, and the network training process will not converge. Therefore, the focal loss [33] is introduced to solve the loss oscillation problem.

The focal loss (FL) function is defined as:

$$FL(p,g) = \begin{cases} -\sum_{i=1}^{N} \alpha(1-p)^{\gamma}\log(p), & \text{if } g = 1 \\ -\sum_{i=1}^{N} (1-\alpha)p^{\gamma}\log(1-p), & \text{if } g = 0 \end{cases} \tag{4}$$

In the formula, $\alpha \in [0,1]$ represents the balance factor, which is used to balance the uneven proportions of positive and negative samples. We set α to 0.25; that is, the proportion of positive samples is smaller than that of negative samples. $\gamma \in [0,5]$ represents the modulation factor, which can adjust the rate of simple sample weight reduction. When γ is 0, the cross-entropy loss function is utilized. We take the default value 2 for γ. The focal loss is an improvement of the binary cross-entropy function. It adds a modulation factor to the original function,

reduces the extent to which the easy samples contribute to the total loss, and makes the network pay more attention to difficult samples that are difficult to classify.

The combination of the Dice and focal losses can not only use the foreground pertinence of the Dice loss but also use the focal loss to smooth the training loss curve. As shown below, we use an adjustable hyperparameter $\omega \in [0, 1]$ to control the weights of the two loss functions, which can solve the voxel imbalance between the foreground and background in the pulmonary artery segmentation task and effectively alleviate the oscillation of unstable loss values.

$$L(p, g) = \omega \times DL(p, g) + (1 - \omega) \times FL(p, g) \tag{5}$$

4. Experiment

4.1. Datasets and validation metrics

Our DS2Net is evaluated on the Parse22 [34] pulmonary artery segmentation dataset. The dataset contains 3D CT groups with pulmonary vascular labels. The image sizes are between 512 × 512 × 228 and 512 × 512 × 376. Among them, 100 images are available as the training dataset. 70 of the CT slice groups are used for network training and the other 30 CT slice groups to verify the effectiveness of DS2Net and compare it with other models. For the validation metrics, we used the following evaluation metrics: (1) Dice Similarity Coefficient (DSC) [10], (2) Accurate (Acc), (3) False Positive Rate (FPR), (4) False Negative Rate (FNR), (5) Branch Detected (BD) [35], (6) Tree-length Detected (TD) [35]. In this context, the DSC metric measures similarity by calculating the ratio between the overlap of two regions and the total volume. The Pre metric is used to measure how many of the positive cases predicted by the model are truly positive. The FPR metric is used to indicate the rate of negative cases that the model incorrectly predicts as positive, and the FNR metric is used to indicate the rate of positive cases that the model incorrectly predicts as negative. In addition, the BD metric is one of the metrics used to assess the tree structure. It refers to the number of branches correctly predicted by the model and is typically used to compare the accuracy of the model's prediction of branching nodes in the tree structure. The TD metric is another metric used to assess the tree structure. It measures the total length of the tree structure correctly detected by the model and is typically used to compare the completeness of the model in representing the tree structure.

To reflect the segmentation performance of the network in a more comprehensive and detailed way, not only calculate the overall DSC scores based on the group of CT slices (3D DSC) are calculated, but also the DSC scores of a single slice (2D DSC). 3D DSC is usually used to evaluate the segmentation results of 3D medical images and is a commonly used metric in 3D medical segmentation networks. It represents the degree of similarity between the 3D segmentation results and the real segmentation results. On the other hand, 2D DSC is usually used to evaluate the segmentation results of 2D images. When evaluating pulmonary vascular segmentation performance, more attention is paid to the segmentation accuracy on the details of fine vessels. Higher 2D DSC values indicate that the model is more capable of segmenting fine vessels.

Furthermore, since the resource cost consumption of the network is also an important factor for evaluating its segmentation performance, we also calculated the number of parameters, computational cost, and inference time of the network as comparison indexes.

4.2. Implementation details

Our DS2Net is implemented using PyTorch and Monai. Due to the multi-scale transformations involved in the DS2Net network, random Gaussian fuzzy and random Gaussian noise methods are employed during data pre-processing to enhance the segmentation network's generalization performance and prevent excessive model complexity and overfitting. Additionally, a Dropout layer is implemented in the neural network to deactivate neurons at random during each training iteration. And due to the network's lightweight feature, a full-size input can be utilized with a batch size of 16 and an image size of 512 × 512 × 8, and there is not need to randomly crop the input to control the number of calculations. This makes the position of the target area in the input relatively fixed, which is more conducive to the network process of constructing the position characteristics of the target. Different from RGB images, the voxel values of CT image data are represented by CT intensity values ranging from [-1000, +1000], and different intensity value ranges represent different organs and tissues. To remove the irrelevant information about other organs and tissues from the data, the input values located in the range of [-968, + 512] are intercepted during the data pre-processing stage and then standardized to the range of [0,1]. During the training phase, stochastic gradient descent (SGD) is employed as the optimizer, with an initial learning rate of 0.001, a momentum of 0.9, and a weight decay of 1e-6 with cosine annealing. In the verification stage, the 3D and 2D DSC scores are calculated simultaneously, and the 3D DSC score is taken as the final evaluation index. The model is trained for a total of 200 epochs, and all models are trained on a NVIDIA 3090 GPU.

Table 1
Quantitative comparison between our proposed method and other 2D/2.5D methods.

Model	Type	2D DSC	3D DSC	Acc	FPR	FNR	BD	TD
DeepLab v3[36]	2D	49.85	70.45	67.23	0.221	34.63	44.92	49.82
U-Net [6]	2D	56.23	72.51	72.84	0.196	24.02	42.94	58.93
VNet [10]	2D	59.31	74.89	75.16	0.167	25.48	43.28	61.83
AttentionU-Net [37]	2D	58.61	75.84	76.36	0.183	25.94	46.74	59.45
LamdaU-Net [38]	2.5D	40.58	48.94	46.29	0.270	53.71	35.46	44.29
MNet [39]	2.5D	60.99	68.25	70.03	0.107	29.97	42.41	58.05
DU-Net [40]	2.5D	61.13	73.61	**79.60**	0.179	**15.40**	46.88	55.60
DS2Net(Ours)	2.5D	**61.53**	**78.16**	77.40	**0.077**	23.00	**47.15**	**65.13**

Table 2
Quantitative comparison between our proposed method and other 3D methods.

Model	Type	3D DSC	Acc	FPR	FNR	BD	TD
AHNet [41]	3D	77.63	74.42	0.052	25.58	39.71	55.48
KiUNet [42]	3D	77.89	75.47	0.056	24.53	48.24	63.83
ResNet [43]	3D	78.78	76.83	0.053	24.32	45.29	62.49
U-Net	3D	**80.06**	**77.97**	**0.050**	22.03	48.35	63.98
DS2Net(Ours)	2.5D	78.16	77.40	0.077	23.00	47.15	**65.13**

Table 3
The calculation and parameter quantities of different networks.

Model	Type	Params(M)	FLOPs(G)	Times(s)
AHNet	3D	38.13	281.79	7.92 ± 0.74
KiU-Net	3D	4.33	141.94	5.56 ± 0.55
U-Net	3D	4.81	75.89	3.36 ± 0.39
U-Net	2D	1.62	4.69	16.36 ± 0.75
VNet	2D	9.36	58.45	28.54 ± 0.64
Attention U-Net	2D	1.99	14.35	17.34 ± 0.84
LamdaU-Net	2.5D	3.61	287.64	36.73 ± 1.47
MNet	2.5D	2.20	183.79	49.50 ± 1.94
DU-Net	2.5D	9.51	128.9	39.05 ± 2.87
DS2Net(Ours)	2.5D	2.66	17.76	20.42 ± 1.73

4.3. Comparison with other methods

To quantitatively illustrate the superiority of DS2Net, several other methods that are widely used in the field of medical image segmentation are selected and compared with DS2Net, including U-Net (2D), VNet (2D), DeepLab v3 (2D), AttentionU-Net (2D) and ResNet (3D). The results are shown in Tables 1 and 2. Since the original papers of the other methods did not evaluate the results obtained on the Parse22 dataset, we replicate their segmentation networks and maintain the same data preprocessing method as that of DS2Net. It should be noted

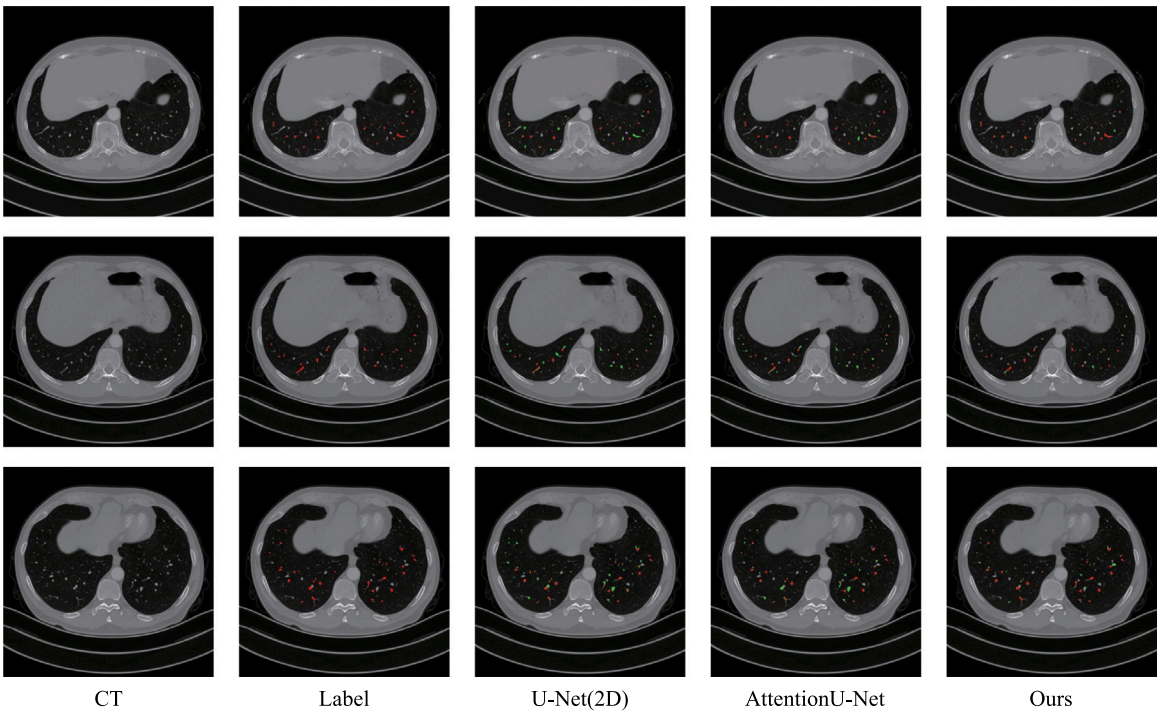


Fig. 6. 2D Segmentation results of different methods. Red for correct segmentation results, green for incorrect segmentation results (including omissions and mis-segmentations).

that the 3D DSC score is selected as the final evaluation index for the networks, and the corresponding 2D DSC scores may not optimal.

In the quantitative comparison, as shown in Table 1, our DS2Net achieves the best 2D DSC score of 61.53% and the best 3D DSC score of 78.16% compared with those of the 2D networks and 2.5D networks; these values are better than those of the second-best method (Attention U-Net) by 2.93% and 2.32%, respectively. In terms of segmentation accuracy and computational cost, compared with the 2D networks, although DS2Net increases the computational cost and parameter quantity, it also achieves improved segmentation accuracy, and the segmentation performance of the network is effectively enhanced. Experiments show that compared with other 2D networks and 2.5D networks, our proposed DS2Net network can yield significantly improved segmentation accuracy based on its superior morphological and spatial characteristic learning ability and its spatial volume information extraction ability with only small increases in the number of parameters and the computational cost.

In the comparison experiment with the 3D networks, due to the high memory requirements of the 3D networks, we use the commonly employed random clipping input method to control the number of calculations when training the 3D networks. The input size of all 3D networks is set to 256 × 256 × 96, and the batch size is set to 4. Although the 3D networks have larger input data and more computational power, they do not have a significant advantage over DS2Net in terms of segmentation accuracy. Taking ResNet (3D) as an example, the number of parameters is approximately 8.25 times that of DS2Net, and the number of calculations is approximately 97.93 times that of DS2Net. When the numbers of parameters and calculations are far greater than those of DS2Net, the 3D DSC score of this method is only 0.62% higher than that of DS2Net. These experimental results show that many redundant computations are utilized by the 3D networks when they are applied to the field of medical image segmentation, which not only leads to a waste of computing resources but also leads to the overfitting problem for 3D networks. Our DS2Net can extract spatial volume information in its 2D network via the super-2D segmentation pattern, thereby not only approaching the segmentation accuracy of the 3D benchmark network but also avoiding its high computational cost.

The computation time, parameters, and flops of DS2Net and other compared methods have been summarized in Table 3. In terms of parameters and computational flops, DS2Net shows significant advantages over 3D methods. However, in terms of computational time, the computational time of 3D segmentation methods is less than that of DS2Net because DS2Net obtains one vascular segmentation result per computation, whereas 3D segmentation methods are computed directly on the whole group of CT slices. It is worth noting that although the 3D segmentation method has some advantages in terms of computation time, its computational burden is also quite large, and ordinary hardware cannot even run the common 3D segmentation model. In contrast, DS2Net, although slightly slower, has the advantage of running on lightweight hardware, and we have improved the segmentation performance of DS2Net to reach the mid-range of 3D segmentation methods.

4.4. Visualization results

In order to show the details of the pulmonary vascular segmentation, we performed 3D visualization and 2D visualization of the lung vessel segmentation results in this section. Fig. 6 shows some 2D segmentation results obtained by DS2Net on the Parse22 dataset and compares them with those of Attention U-Net and U-Net, where the red labels represent correct segmentation and the green labels represent omission or misclassification. It can be seen that for fine vessels the segmentation of Attention U-Net and U-Net is not satisfactory, with more omissions and mis-segmentations. In contrast, the segmentation results obtained by DS2Net for the fine vessels are clearer and more accurate, which further verifies the effectiveness of DS2Net. As shown in Fig. 7, from left to right are the vascular labels, the 3D segmentation results of DS2Net, the 3D segmentation results of MNet and the 3D segmentation results of D U-Net. As can be seen from the 3D visualization results, especially in most of the fine vessel regions in the figure, the DS2Net proposed in this paper has obvious advantages over other 2.5D methods.

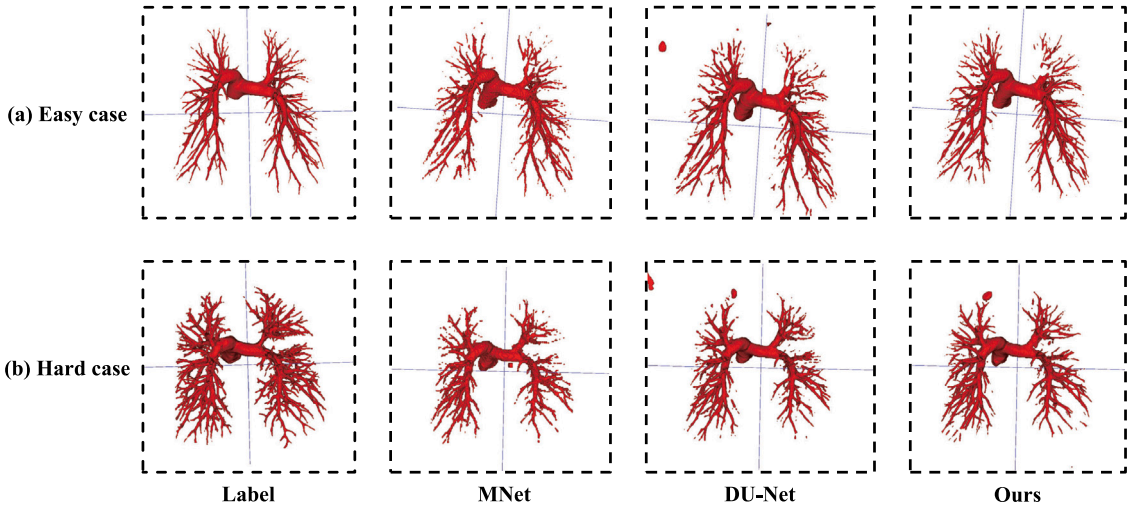


Fig. 7. Rendering of pulmonary airway segmentation results on (a) easy and (b) hard testing cases. In the easy case, the 3D DSC scores were 83.76% for DS2Net, 82.98% for DU-Net and 79.97% for MNet; in the hard case, the 3D DSC scores were 71.74% for DS2Net, 69.37% for DU-Net and 71.35% for MNet.

Table 4
Ablation studies concerning different encoder block configurations.

Setting	Params(M)	FLOPs(G)	2D DSC	3D DSC
Baseline + PoolFormer Module	2.34	15.07	56.39	74.17
Baseline + OffsetConv + PoolFormer Module	2.49	17.71	60.87	76.57
Baseline + OffsetConv + PoolFormer Module + SE Block	2.66	17.76	**61.53**	**78.16**

Table 5
DS2Net and UNet 2D T-Test Results from 5-fold randomized cross-validation experiments using 7:3 division of the 100 images into training and validation data. We use statistical standard notation for found significances: *; **; ***: $p < 0.05$; $p < 0.01$; $p < 0.001$.

Metric	DS2Net Mean ± standard deviation	UNet 2D	Difference	t	df	P	Cohen's d
2D DSC	**61.526 ± 0.881**	57.008 ± 1.124	4.518 ± −0.243	14.969	4	0.000***	6.694
3D DSC	**78.16 ± 0.443**	73.04 ± 1.163	5.12 ± −0.719	13.518	4	0.000***	6.046
Acc	**77.404 ± 1.108**	73.242 ± 1.115	4.162 ± −0.007	14.551	4	0.000***	6.507
FPR	**0.077 ± 0.007**	0.195 ± 0.054	−0.118 ± −0.048	−4.484	4	0.011**	2.005
FNR	**22.996 ± 1.008**	23.506 ± 0.76	−0.51 ± 0.248	−1.174	4	0.305	0.525
BD	**47.146 ± 2.442**	43.724 ± 2.64	3.422 ± −0.199	4.98	4	0.008***	2.227
TD	**65.128 ± 2.3**	59.162 ± 1.654	5.966 ± 0.646	10.745	4	0.000***	4.805

4.5. Ablation studies and analysis

We conducted a series of ablation experiments in which we individually removed specific modules or components from DS2Net and assessed the impact of these changes on model performance. This allowed us to determine which components were critical to the effectiveness of our proposed methodology, making our study more convincing. As shown in Table 4, first, we test the removal of the OffsetConv and SE blocks in DS2Net and only retain the PoolFormer module, that is, , and the network can only reach the level of common 2D networks. Then, we add the OffsetConv block in it. Compared with that of , the 2D DSC score is increased by 4.48%, and the 3D DSC score is increased by 2.4%. The significant improvement in the 2D DSC score of proves that the semantic reinforcement of fine vessel features via the OffsetConv algorithm can effectively improve the segmentation accuracy achieved for fine vessels on a single CT slice. Finally, compared with before, the complete DS2Net provides a 0.66% 2D DSC score improvement and a 1.59% 3D DSC score improvement, which also proves that after the SE block adds interslice feature information, the segmentation accuracy of the whole CT slice group is significantly improved.

In addition, we performed hyperparametric experiments on the number of CT slices to investigate the effect of the number of input CT slices on the segmentation performance of DS2Net. As shown in Fig. 8, we performed model training and evaluation with the number of slices ranging from 2 to 16, and counted the 2D DSC scores and 3D DSC scores with different numbers of CT slices. The results show

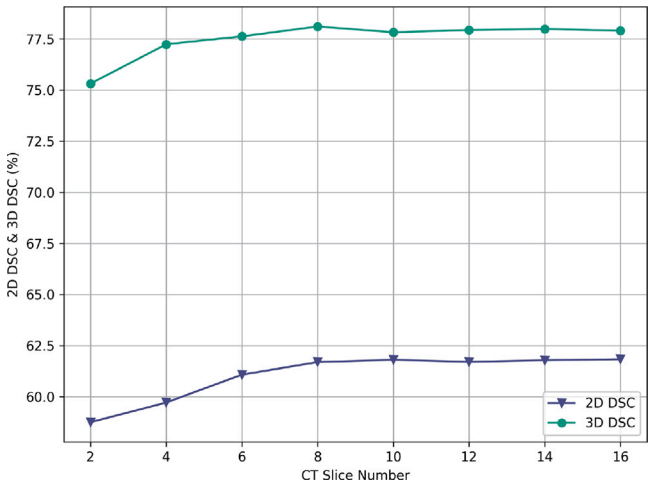


Fig. 8. Hyperparametric experiment on CT slice number.

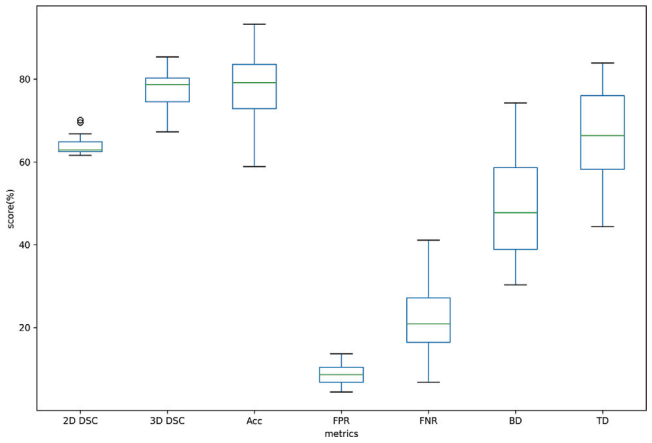


Fig. 9. Overview metrics boxplot for DS2Net, includes metrics 2D DSC, 3D DSC, Acc, FPR, FNR, BD and TD.

that both 3D DSC scores and 2D DSC scores tend to increase as the number of slices increases. Specifically, the 3D DSC score reaches the highest value when the number of slices is 8, while the 2D DSC score starts to stabilize at the number of slices of 8. The experiments show that the segmentation performance of the model is better when the number of hyperparameter slices is set to 8, while it does not waste computational resources, which is a cost-effective choice. Finally, We have supplemented DS2Net with a five-fold cross-validation experiment and verified the significant improvement of DS2Net by performing a t-test with the UNet 2D. We have summarized the results in Table 5. Through this series of five-fold cross-validation experiments, we have further validated the effectiveness of DS2Net. The t-test confirmed the significant improvement of DS2Net on the 2D DSC, 3D DSC, Acc, FPR, BD and TD metrics. However, no significant improvement was observed in the FNR metrics and this aspect may require further in-depth research and refinement.

5. Conclusion

In this paper, a powerful super-2D segmentation network for pulmonary vascular segmentation called the detail-oriented super-2D segmentation network (DS2Net) is proposed. To solve the high computational costs of 3D networks and the low segmentation accuracies of 2D networks, a super-2D segmentation pattern is adopted, and the input pattern is optimized with consecutive CT slices as inputs, thereby successfully extracting spatial volume information in a 2D network. Aiming at the difficulty of segmenting fine vessels in the pulmonary vascular segmentation task, an OffsetConv algorithm is designed for the semantic reinforcement of fine vessel features. As a part of the DS2Net encoder, the OffsetConv algorithm effectively improves the segmentation performance of the network for fine vessels on a single CT slice. Extensive experiments conducted on the Parse22 pulmonary artery dataset show that our DS2Net achieves excellent results, and its excellent fine vessel segmentation performance validates the effectiveness of our proposed OffsetConv algorithm. In our experiments, DS2Net outperforms all 2D segmentation networks used for comparison, with a 3D DSC score of 78.16%. It achieves segmentation accuracy close to that of the 3D benchmark network with a much lower computational cost. The effectiveness of the super-2D segmentation pattern is fully validated, which will motivate further investigation into the surprising performance of the super-2D segmentation pattern. In clinical medicine, the application of 3D reconstructive imaging to examine the pulmonary vasculature is crucial for detecting problems such as haemangiomas and vascular lesions. However, this process currently requires manual labelling, which is time-consuming. In order to reduce the workload, deep learning-based pulmonary vessel segmentation techniques are widely used. These techniques are designed to help labelling blood vascular. However, these techniques usually require limited high performance hardware equipment, which restricts their wide application. The significance of DS2Net for accelerating, simplifying and optimizing pulmonary vascular labelling aids is therefore clear. This approach has the potential to accelerate and simplify pulmonary vascular labelling, making it more accessible to healthcare practitioners and ultimately improving patient diagnosis and treatment. For future work, we will further explore more powerful dimensional fusion strategies and explore more efficient semantic reinforcement algorithms for fine vessels to further improve the segmentation accuracy achieved for fine vessels on a single CT slice.

CRediT authorship contribution statement

Lufei Lou: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Yu Xin:** Supervision, Resources, Funding acquisition. **Jiangbo Qian:** Writing – review & editing, Validation, Supervision, Funding acquisition. **Yihong Dong:** Writing – review & editing, Validation, Supervision, Resources.

Declaration of competing interest

The authors declared no potential conflicts of interest with respect to the research, authorship or publication of this article.

Data availability

I have shared the code link at the end of abstract.

Acknowledgements

We acknowledge the support of the Natural Science Foundation of Zhejiang Province, China (Grant No. LY22F020001,No. LZ20F020001), the 3315 Plan Foundation of Ningbo (Grant No. 2019B-18-G), China Natural Science Foundation under Grant 62271274.

References

[1] S. Moccia, E. De Momi, S. El Hadji, L.S. Mattos, Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics, Comput. Methods Programs Biomed. 158 (2018) 71–91.
[2] R.A. Ochs, J.G. Goldin, F. Abtin, H.J. Kim, K. Brown, P. Batra, D. Roback, M.F. McNitt-Gray, M.S. Brown, Automated classification of lung bronchovascular anatomy in CT using AdaBoost, Med. Image Anal. 11 (3) (2007) 315–324.
[3] R. Kiros, K. Popuri, D. Cobzas, M. Jagersand, Stacked multiscale feature learning for domain independent medical image segmentation, in: Machine Learning in Medical Imaging: 5th International Workshop, MLMI 2014, Held in Conjunction with MICCAI 2014, Boston, MA, USA, September 14, 2014. Proceedings 5, Springer, 2014, pp. 25–32.
[4] A.F. Frangi, W.J. Niessen, K.L. Vincken, M.A. Viergever, Multiscale vessel enhancement filtering, in: Medical Image Computing and Computer-Assisted Intervention—MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings 1, Springer, 1998, pp. 130–137.
[5] P. Nardelli, D. Jimenez-Carretero, D. Bermejo-Pelaez, G.R. Washko, F.N. Rahaghi, M.J. Ledesma-Carbayo, R.S.J. Estépar, Pulmonary artery–vein classification in CT images using deep learning, IEEE Trans. Med. Imaging 37 (11) (2018) 2428–2440.
[6] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, 2015, pp. 234–241.
[7] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
[8] Z. Zhu, Y. Xia, W. Shen, E. Fishman, A. Yuille, A 3D coarse-to-fine framework for volumetric medical image segmentation, in: 2018 International Conference on 3D Vision, 3DV, IEEE, 2018, pp. 682–690.
[9] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: learning dense volumetric segmentation from sparse annotation, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19, Springer, 2016, pp. 424–432.

[10] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision, 3DV, Ieee, 2016, pp. 565–571.
[11] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P.F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, et al., nnu-net: Self-adapting framework for u-net-based medical image segmentation, 2018, arXiv preprint arXiv:1809.10486.
[12] N. Zettler, A. Mastmeyer, Comparison of 2D vs. 3D U-net organ segmentation in abdominal 3D CT images, 2021, arXiv preprint arXiv:2107.04062.
[13] Y. Zhang, H. Liu, Q. Hu, Transfuse: Fusing transformers and cnns for medical image segmentation, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24, Springer, 2021, pp. 14–24.
[14] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, P.-A. Heng, H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes, IEEE Trans. Med. Imaging 37 (12) (2018) 2663–2674.
[15] Y. Jiang, Z. Zhang, S. Qin, Y. Guo, Z. Li, S. Cui, APAUNet: Axis projection attention unet for small target in 3D medical segmentation, in: Proceedings of the Asian Conference on Computer Vision, 2022, pp. 283–298.
[16] W. Tan, Y. Yuan, A. Chen, L. Mao, Y. Ke, X. Lv, An approach for pulmonary vascular extraction from chest CT images, J. Healthc. Eng. 2019 (2019).
[17] J.N. Kaftan, A.P. Kiraly, A. Bakai, M. Das, C.L. Novak, T. Aach, Fuzzy pulmonary vessel segmentation in contrast enhanced CT data, in: Medical Imaging 2008: Image Processing, 6914, SPIE, 2008, pp. 585–596.
[18] M.E. Yueksel, M. Borlu, Accurate segmentation of dermoscopic images by image thresholding based on type-2 fuzzy logic, IEEE Trans. Fuzzy Syst. 17 (4) (2009) 976–982.
[19] M. Orkisz, M.H. Hoyos, V.P. Romanello, C.P. Romanello, J. Prieto, C. Revol-Muller, Segmentation of the pulmonary vascular trees in 3D CT images using variational region-growing, Irbm 35 (1) (2014) 11–19.
[20] R. Wu, Y. Xin, J. Qian, Y. Dong, A multi-scale interactive U-Net for pulmonary vessel segmentation method based on transfer learning, Biomed. Signal Process. Control 80 (2023) 104407.
[21] L. Yang, H. Wang, Q. Zeng, Y. Liu, G. Bian, A hybrid deep segmentation network for fundus vessels via deep-learning framework, Neurocomputing 448 (2021) 168–178.
[22] E. Nasr-Esfahani, S. Samavi, N. Karimi, S.R. Soroushmehr, K. Ward, M.H. Jafari, B. Felfeliyan, B. Nallamothu, K. Najarian, Vessel extraction in X-ray angiograms using deep learning, in: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, IEEE, 2016, pp. 643–646.
[23] Y. Qin, H. Zheng, Y. Gu, X. Huang, J. Yang, L. Wang, Y.-M. Zhu, Learning bronchiole-sensitive airway segmentation CNNs by feature recalibration and attention distillation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2020, pp. 221–231.
[24] A.-M. Rickmann, A.G. Roy, I. Sarasua, N. Navab, C. Wachinger, 'Project & excite'modules for segmentation of volumetric medical scans, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22, Springer, 2019, pp. 39–47.
[25] W. Zhu, Y. Huang, L. Zeng, X. Chen, Y. Liu, Z. Qian, N. Du, W. Fan, X. Xie, AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy, Med. Phys. 46 (2) (2019) 576–589.
[26] Y. Hou, Z. Ma, C. Liu, C.C. Loy, Learning lightweight lane detection cnns by self attention distillation, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1013–1021.
[27] S. Zagoruyko, N. Komodakis, Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer, 2016, arXiv preprint arXiv:1612.03928.
[28] K. Zhou, N. Chen, X. Xu, Z. Wang, J. Guo, L. Liu, Z. Yi, Automatic airway tree segmentation based on multi-scale context information, Int. J. Comput. Assist. Radiol. Surg. 16 (2021) 219–230.
[29] J. Yun, J. Park, D. Yu, J. Yi, M. Lee, H.J. Park, J.-G. Lee, J.B. Seo, N. Kim, Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net, Med. Image Anal. 51 (2019) 13–20.
[30] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).
[31] W. Yu, M. Luo, P. Zhou, C. Si, Y. Zhou, X. Wang, J. Feng, S. Yan, Metaformer is actually what you need for vision, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 10819–10829.
[32] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.
[33] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.
[34] L. Wang, Q. Dou, P.T. Fletcher, S. Speidel, S. Li, Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI, vol. 13436, Springer Nature, 2022.
[35] P. Lo, B. Van Ginneken, J.M. Reinhardt, T. Yavarna, P.A. De Jong, B. Irving, C. Fetita, M. Ortner, R. Pinho, J. Sijbers, et al., Extraction of airways from CT (EXACT'09), IEEE Trans. Med. Imaging 31 (11) (2012) 2093–2107.
[36] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 801–818.
[37] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention u-net: Learning where to look for the pancreas, 2018, arXiv preprint arXiv:1804.03999.
[38] Y. Ou, Y. Yuan, X. Huang, K. Wong, J. Volpi, J.Z. Wang, S.T. Wong, Lambdaunet: 2.5 d stroke lesion segmentation of diffusion-weighted mr images, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24, Springer, 2021, pp. 731–741.
[39] Z. Dong, Y. He, X. Qi, Y. Chen, H. Shu, J.-L. Coatrieux, G. Yang, S. Li, MNet: rethinking 2D/3D networks for anisotropic medical image segmentation, 2022, arXiv preprint arXiv:2205.04846.
[40] Y. Zhou, W. Huang, P. Dong, Y. Xia, S. Wang, D-UNet: a dimension-fusion u shape network for chronic stroke lesion segmentation, IEEE/ACM Trans. Comput. Biol. Bioinform. 18 (3) (2019) 940–950.
[41] S. Liu, D. Xu, S.K. Zhou, O. Pauly, S. Grbic, T. Mertelmeier, J. Wicklein, A. Jerebko, W. Cai, D. Comaniciu, 3D anisotropic hybrid network: Transferring convolutional features from 2d images to 3d anisotropic volumes, in: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11, Springer, 2018, pp. 851–858.
[42] J.M.J. Valanarasu, V.A. Sindagi, I. Hacihaliloglu, V.M. Patel, Kiu-net: Overcomplete convolutional architectures for biomedical image and volumetric segmentation, IEEE Trans. Med. Imaging 41 (4) (2021) 965–976.
[43] A. Myronenko, 3D MRI brain tumor segmentation using autoencoder regularization, in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part II 4, Springer, 2019, pp. 311–320.

Lufei Lou is a graduate student of Faculty of Electrical Engineering and Computer Science at Ningbo University. His research interests include deep learning and medical image segmentation and pulmonary vessel segmentation.

Yu Xin received the Ph.D. degree in computer science and technology from Harbin Engineering University, China. He is currently an Associate Professor with the Faculty of Electrical Engineering and Computer Science, Ningbo University, China. His current research interests include multiple classifier and prediction systems, processing and modelling of uncertainty in predictive modelling, recommendation systems, diagnostic analysis, and decision support systems.

Jiangbo Qian received the Ph.D. degree in computer science from Southeast University, China, in 2006. He was a Visiting Scholar with the Department of Computer and Information Science, The University of Michigan–Dearborn, USA. He is currently a Professor with the Faculty of Electrical Engineering and Computer Science, Ningbo University, China. His research interests include database management, streaming data processing, deep learning, computer vision, and hardware/software codesign.

Yihong Dong received the Ph.D. degree in computer science from Zhejiang University, China, in 2007. He is currently a Professor with the Faculty of Electrical Engineering and Computer Science, Ningbo University, China. His research interests include big data, data mining, and artificial intelligence.