



## 3D PBV-Net: An automated prostate MRI data segmentation method

Yao Jin<sup>a,1</sup>, Guang Yang<sup>b,c,1,\*</sup>, Ying Fang<sup>a</sup>, Ruipeng Li<sup>d</sup>, Xiaomei Xu<sup>a</sup>, Yongkai Liu<sup>e</sup>, Xiaobo Lai<sup>a,\*\*</sup>

<sup>a</sup> College of Medical Technology, Zhejiang Chinese Medical University, Hangzhou, 310053, China

<sup>b</sup> Cardiovascular Research Centre, Royal Brompton Hospital, London, SW3 6NP, UK

<sup>c</sup> National Heart and Lung Institute, Imperial College London, London, SW7 2AZ, UK

<sup>d</sup> Hangzhou Third People's Hospital, Hangzhou, 310009, China

<sup>e</sup> Department of Radiological Sciences, David Geffen School of Medicine, The University of California at Los Angeles, Los Angeles, CA, USA



### ARTICLE INFO

**Keywords:**

Prostate cancer

MRI

Automated segmentation

Telehealth care

Enabling technology

### ABSTRACT

Prostate cancer is one of the most common deadly diseases in men worldwide, which is seriously affecting people's life and health. Reliable and automated segmentation of the prostate gland in MRI data is exceptionally critical for diagnosis and treatment planning of prostate cancer. Although many automated segmentation methods have emerged, including deep learning based approaches, segmentation performance is still poor due to the large variability of image appearance, anisotropic spatial resolution, and imaging interference. This study proposes an automated prostate MRI data segmentation approach using bicubic interpolation with improved 3D V-Net (dubbed 3D PBV-Net). Considering the low-frequency components in the prostate gland, the bicubic interpolation is applied to preprocess the MRI data. On this basis, a 3D PBV-Net is developed to perform prostate MRI data segmentation. To illustrate the effectiveness of our approach, we evaluate the proposed 3D PBV-Net on two clinical prostate MRI data datasets, i.e., PROMISE 12 and TPHOH, with the manual delineations available as the ground truth. Our approach generates promising segmentation results, which have achieved 97.65% and 98.29% of average accuracy, 0.9613 and 0.9765 of Dice metric, 3.120 mm and 0.9382 mm of Hausdorff distance, and average boundary distance of 1.708, 0.7950 on PROMISE 12 and TPHOH datasets, respectively. Our method has effectively improved the accuracy of automated segmentation of the prostate MRI data and is promising to meet the accuracy requirements for telehealth applications.

### 1. Introduction

Prostate cancer is one of the diseases with a wide range of tissue patterns that result in high morbidity and mortality rate [1]. The most effective non-invasive diagnostic method of prostate cancer is magnetic resonance (MR) imaging, which provides excellent soft-tissue contrast. It is also one of the most accurate prostate imaging methods [2], giving the doctor intuitive and comprehensive observations of the disease. The automated segmentation of the prostate gland in the medical imaging dataset is often considered a challenging task for radiotherapy treatment [3]. In automated prostate segmentation, traditional approaches are mainly based on anatomical atlas registration (Guo Y et al., 2016). However, the obtained images may be affected by the field strength, scanner type, and coil profile, which can increase the difficulty of

automated segmentation and also computationally intensive [4]. Moreover, due to technical limitations, medical institutions usually rely on manual segmentation of prostate MRI data. The whole process is time-consuming and less accurate [5]. The manual segmentation method can no longer meet the clinical needs with an increased number of patients. Therefore, a technique that can automatically and efficiently segment prostate MRI data is highly in demand.

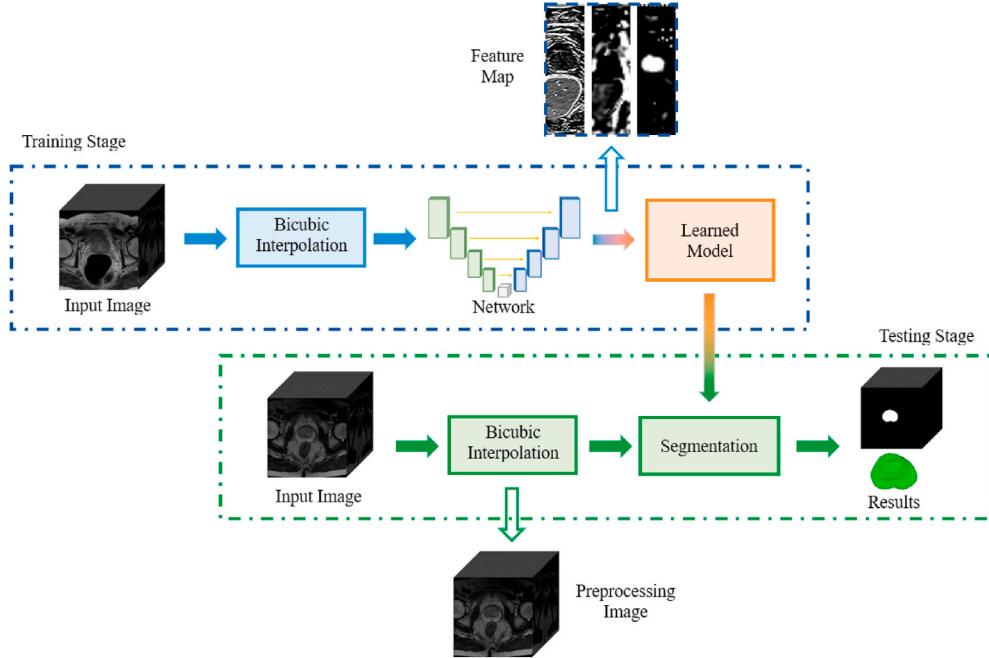
More than 50% of the WHO allied countries have a specific national telehealth policy in their country (Global diffusion of eHealth). The overall development of telehealth and the application of remote radiological image diagnosis have made the requirements more stringent for image processing. For instance, wearable sensing and telehealth technology have been used in coronavirus pandemic diagnosis and elderly health checks [6,7]. Telehealth can skillfully manage and distribute

\* Corresponding author. Cardiovascular Research Centre, Royal Brompton Hospital, London, SW3 6NP, UK.

\*\* Corresponding author.

E-mail addresses: [g.yang@imperial.ac.uk](mailto:g.yang@imperial.ac.uk) (G. Yang), [dmia\\_lab@zcmu.edu.cn](mailto:dmia_lab@zcmu.edu.cn) (X. Lai).

<sup>1</sup> Co-first authors contributed equally.



**Fig. 1.** Schematic illustration of our proposed 3D PBV-Net for the prostate MRI data segmentation.

emergency medical services in remote areas. Besides, it can lift the geographical restrictions by sharing the local medical records and diagnosis images of the patients. In the context of MRI data diagnosis in telehealth care, medical image segmentation is still a tremendous amount of work [8]. A method such as three-level coarse-to-fine segmentation can split and extract parts or regions of interest for clinicians, thereby strengthening the efficiency and precision of the disease observations [9]. Deep neural networks-based approaches have been vastly applied to medical image segmentation, replacing traditional image segmentation technologies with the rise of the available computing power. Nevertheless, most of these deep learning-based methods have a common shortcoming that they excessively rely on manual segmentation and are less capable of handle full 3D medical images directly. Existing methods normally convert 3D data into 2D slices via either dimensionality reduction or image slicing before training the deep learning models. This is not conducive to efficient segmentation of the 3D images. Therefore, it is desirable to present a more effective method for 3D images in order to obtain more in-depth information to meet the needs of clinical diagnosis better.

The major contributions of our work are four folds that can be summarized as follows.

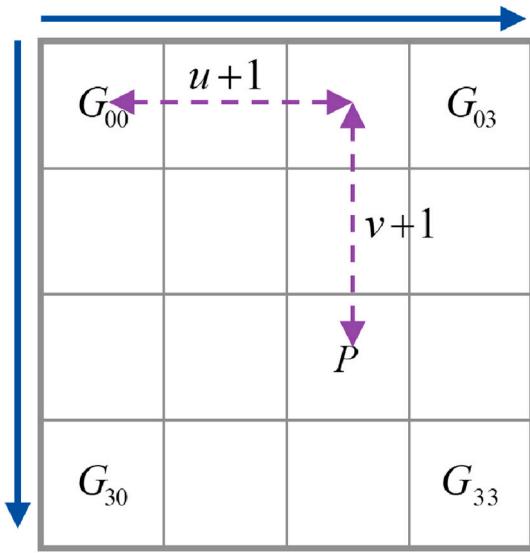
- We design an automated prostate MRI data segmentation method based on bicubic interpolation coupled with an improved 3D V-NET (3D PBV-Net). The expansion of the pixel information of the prostate MRI data uses three-dimensional convolution to quickly, accurately, and efficiently segment the prostate regions. It also relies less on manual segmentation, which is a time-consuming labeling task.
- We test our novel 3D PBV-Net on both open accessible PROMISE 12 dataset and our in-house private dataset (collected from the Third Peoples Hospital of Hangzhou), i.e., TPHOH dataset, respectively. We have achieved high segmentation accuracy of 0.9613 and 0.9765 measured by the Dice metric (DM).
- Based on the segmentation results, 3D reconstruction and rendering of the delineated results are performed. The visualization can provide clinicians with an intuitive diagnostic foundation in telehealth to improve the quality of patient management, treatment planning, and prognosis.

The remainder of this paper is organized as follows. Section 2 expounds on the previous studies related to the prostate MRI data segmentation. In section 3, our proposed method is detailed, including data preprocessing, model description, and parameter settings. In Section 4, our experimental results are represented. We then discuss and analyze our results in Section 5 and draw a conclusion in Section 6.

## 2. Related works

In this section, we briefly review the related studies on prostate MRI data segmentation. Traditional methods used image priors or shape priors to segment the target gland in prostate MRI data. For example, Kharote et al. used a k-nearest neighbor classifier to locate the tumor area by voxel, the results showed that the latent features were more prominent in the segmentation (Kharote P et al., 2019). Yuan et al. adopted a model that consisted of two paths, including one with a dense coding path, which could learn advanced features, and the other one with a dense decoder path, which could achieve pixel-level segmentation [10]. Kharote et al. incorporated image patches to construct probability maps. On this basis, a sparse autoencoder was used to extract potential features from prostate MRI data. Besides, they combined with a sparse patch matching method to achieve final segmentation (Kharote P et al., 2019). For prostate cancer, Reda et al. proposed a computer-aided diagnosis framework for the early detection of prostate cancer spread. The proposed framework could segment prostate cancer without preprocessing [11]. He et al. studied a new Active Shape Modeling (ASM)-based method, which divided voxels into three types, i.e., near the boundary, outside the boundary, and the inner core boundary. Finally, the three-layer ASM of the convolutional neural network (CNN) boundary model was used for refinement to achieve better segmentation results [12].

With the powerful learning and generalization capabilities, deep neural networks have played an increasingly important role in many medical image analysis applications and have also been developed for prostate MRI data analysis. For example, Guo et al. proposed a deformable segmentation method combining deep feature learning with sparse patch matching (Guo Y et al., 2016). Zhu et al. designed a boundary-weighted neural network that used a boundary-weighted loss and an advanced boundary-weighted transfer tilt method to make the



**Fig. 2.** Bicubic interpolation algorithm.

network more sensitive during segmentation [13]. H. Öcal et al. reported a model that combined a three-dimensional volumetric convolutional neural network with a two-dimensional volumetric convolutional neural network, which had certain robustness and accuracy [14]. Khan et al. applied DeeplabV3 and semantic segmentation methods to distinguish the prostate, and the results were better than the Fully Convolutional Network (FCN) and the CNN model of prostate segmentation [15].

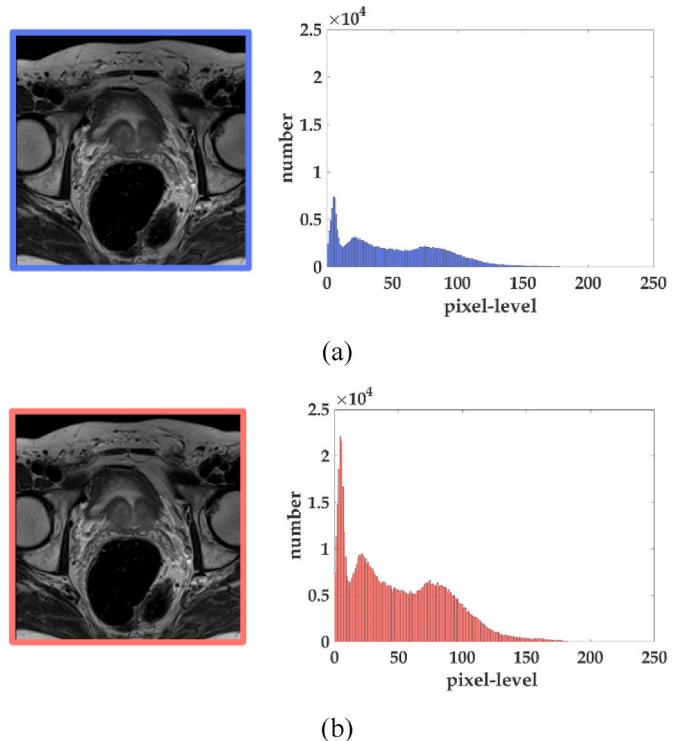
To sum up, the above-mentioned methods have some merits in delineating 3D prostate MRI data, but some limitations may still exist. In particular, full 3D information has not been extracted and effectively used. Therefore, we propose a novel 3D PBV-Net method for automated prostate MRI data segmentation. We use a bicubic interpolation algorithm to enrich the details of MRI data in a preprocessing procedure. We hypothesize that this simple, yet effective preprocessing can boost the performance of the following deep neural network based segmentation. Besides, we improve the 3D V-Net by obtaining more effective and deep features. We believe that our solution can further enhance the efficiency and quality of remote diagnosis of the prostate MRI data analysis in telehealth.

### 3. Materials and methods

Here we detailed our methodology and experimental study for the automated prostate MRI data segmentation using the proposed 3D PBV-Net as follows 1) datasets used for training and testing 3D PBV-Net, 2) preprocess the prostate MRI data via a bicubic interpolation and image rotation to enrich the generalization capability, 3) a detailed description of our proposed 3D PBV-Net, and extended experiment, 4) model training, testing and evaluation metrics. A diagram that summarizes the proposed method is given in Fig. 1.

#### 3.1. Datasets

The proposed method is evaluated on two clinical prostate MRI datasets, which are used to train, evaluate, and objectively compare the performance of standardized prostate MRI data segmentation algorithms. First, we use the prostate MRI data segmentation challenge (PROMISE 12) dataset. A set of 100 representative prostate MRI data is shared on the challenge website (<http://promise12.grandchallenge.org/>). PROMISE 12 includes 50 training data, 30 test data, and 20 challenge data. In addition, the second in-house private dataset is provided by the Third Peoples Hospital of Hangzhou (TPHOH dataset). In total, 106



**Fig. 3.** Bicubic interpolation results.

clinical prostate MRI data are provided with manual segmentation labels to verify the model.

#### 3.2. Data preprocessing and data augmentation

The main intention of the image preprocessing is to eliminate irrelevant information, restore rich and useful information, and enhance the relevant information, which can improve feature extraction and image segmentation reliability. In this study, the preprocessing steps of prostate MRI data are mainly composed of two parts. First, the spatial resolution of the prostate MRI data is enhanced using the bicubic interpolation algorithm. Second, a rotation operation has been performed to augment the dataset.

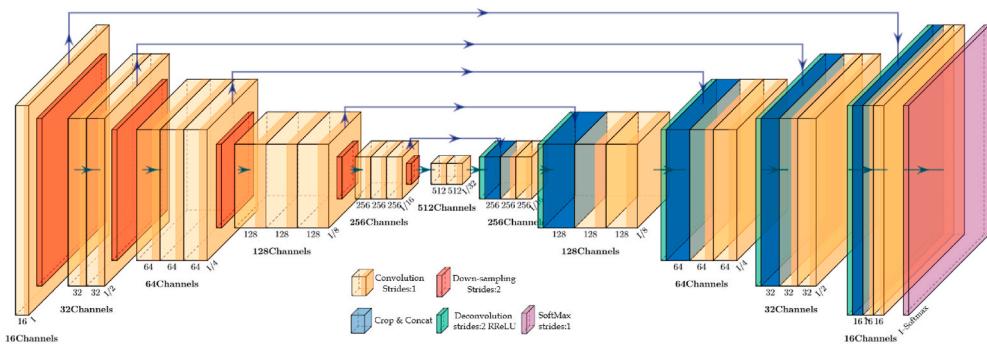
##### 3.2.1. Bicubic interpolation algorithm

The purpose of using a bicubic interpolation algorithm is to widen the low-frequency components [16]. A bicubic interpolation algorithm can make the image with smoother edges and minimal loss of image quality. As the pixels are two-dimensional, the rows and columns are calculated separately. As shown in Fig. 2, the coordinate of the target point  $P$  is  $(x+u, y+v)$  and  $G_{00}$  is  $(x, y)$ . The horizontal and vertical distances between them are  $u_{00} = u + 1$  and  $v_{00} = v + 1$ , respectively.

Furthermore, according to the basis function, the weights of their contributions can be obtained as  $S(u+1)$  and  $S(v+1)$ . Through product operation, the estimation formula for target pixel is as follows

$$f(P) = \sum_{i=0}^3 \sum_{j=0}^3 f(G_{ij}) S(u_{ij}) S(v_{ij}), \quad (1)$$

where  $G$  is the position of the pixel,  $u$  is the longitudinal distance between  $P$  and  $G$ ,  $v$  is the horizontal distance between  $P$  and  $G$ ,  $f(P)$  is the pixel value of the point and  $S$  is the interpolation basic function. The most commonly used interpolation basis function  $S$ , which can be denoted as



**Fig. 4.** The overall architecture of the proposed 3D PBV-Net.

$$S(w) = \begin{cases} (a+2)|w|^3 - (a+3)|w|^2 + 1, & |w| \leq 1 \\ a|w|^3 - 5a|w|^2 + 8|w| - 4a, & 1 < |w| < 2 \\ 0, & |w| \geq 2 \end{cases} \quad (2)$$

where,  $a = 2$ , the pixels need to be estimated using the image type, and corresponding pixels are at the sub-pixel level. For different value ranges, corresponding values can be obtained through the interpolation basis function. It is similar to perform a function mapping from the independent variable.

According to [formula \(1\)](#) and [formula \(2\)](#), bicubic interpolation is performed on PROMISE 12 and TPHOH datasets. For demonstration purposes, we take an example image from the PROMISE 12 datasets and apply the bicubic interpolation algorithm. [Fig. 3](#) shows the 2 Gy histograms of the images before and after performing bicubic interpolation. [Fig. 3](#) (a) shows the original image and the corresponding histogram. The size of the original image is  $512 \times 512$ . [Fig. 3](#) (b) is a processed image using the bicubic interpolation. The size of the processed image is  $1024 \times 1024 \times 3$ . The abscissa represents pixel-level, and the ordinate is the number of grayscales.

### 3.2.2. Dataset augmentation

To maximize the model's generalization ability and prevent overfitting, we augment the datasets. The image rotation operation is used because CNN does not have good rotation invariance. After mapping on PROMISE 12 and TPHOH, they are extended to more than 2000 images for training by the image rotation operation, respectively. Besides, we use the center point of the  $1024 \times 1024 \times 16$  prostate MR image as the rotation center and take different random angles as the rotation angle to rotate the image. The range of the rotation angle is  $5\text{--}60^\circ$ . The extra part of the image is cropped directly, and the missing part is supplemented with the pixel value of 0.

### 3.3. Network architecture design and configurations

#### 3.3.1. Overall framework

A typical deep convolutional network usually consists of an input layer, several convolutional layers, and several downsampling layers. Besides, a classifier is trained to output the corresponding category probability. The convolutional layer in the structure is used to learn image features, and the local connection and weight sharing are used to reduce the number of parameters and reduce the computation complexity. With the deepening of the network, layer-by-layer convolution can extract more abstract image features. As one of the key steps of pattern recognition, the quality of feature extraction directly affects the accuracy of image recognition.

Our proposed 3D PBV-Net is better to suit the 3D structure of the prostate MRI data. The overall framework of the 3D PBV-Net is displayed in [Fig. 4](#). Yellow represents the block that uses convolutions with the stride of 1. The red block represents downsampling with the stride of 2, and the gray block represents the concatenation. Blue indicates the block that uses deconvolutions with the stride of 2 and the RReLU

**Table 1**  
Parameter details of the proposed 3D PBV-Net.

Layer	Input Size	Components
Left-S 1	1024	$3 \times 3 \times 16 \times 16$ , stride (1, 2, 2, 2, 1)
Left-S 2	512	$3 \times 3 \times 32 \times 32$ , stride (1, 2, 2, 2, 1)
Left-S 3	256	$3 \times 3 \times 64 \times 64$ , stride (1, 2, 2, 2, 1)
Left-S 4	128	$3 \times 3 \times 128 \times 128$ , stride (1, 2, 2, 2, 1)
Left-S 5	64	$3 \times 3 \times 256 \times 256$ , stride (1, 2, 2, 2, 1)
Left-S 6	32	$3 \times 3 \times 512 \times 512$ , stride (1, 2, 2, 2, 1)
Right-S 5	64	$3 \times 3 \times 256 \times 256$ , stride (1, 2, 2, 2, 1)
Right-S 4	128	$3 \times 3 \times 128 \times 128$ , stride (1, 2, 2, 2, 1)
Right-S 3	256	$3 \times 3 \times 64 \times 64$ , stride (1, 2, 2, 2, 1)
Right-S 2	512	$3 \times 3 \times 32 \times 32$ , stride (1, 2, 2, 2, 1)
Right-S 1	1024	$3 \times 3 \times 16 \times 16$ , stride (1, 2, 2, 2, 1)
Output	1024	$1 \times 1 \times 32$ , stride (1, 1, 1, 1, 1)

activate function. Purple represents the SoftMax layer. Moreover, the parameter details of the 3D PBV-Net are provided in [Table 1](#).

In the input layer, we input the image with the size of  $1024 \times 1024 \times 3 \times 16$ , where the number of channels is 3. Next, we use  $5 \times 5 \times 5$  voxel core to compress the data along with the model structure. In this model, the first step is to convolve the input image. The strides of the convolution kernel are [1, 1, 1, 1, 1], the convolution kernel is  $R$  and the padding method is 'SAME', which add pixel 0 to the missing parts to keep the feature map stable. On this basis, standardization and dropout processing with 0.8 are performed. The subsequent convolution operation is the same. Then through the down-sampling, the stride of the convolution kernel is set to [1, 2, 2, 2, 1], and the features of images are extracted in depth. After continuous convolution and downsampling, an image of the sixth layer is obtained, the size of which is  $16 \times 16 \times 1024 \times 1024$ . After deconvolution operation, the stride of the deconvolution kernel is [1, 2, 2, 2, 1], and the node is processed using RReLU activation function. In doing so, the processing model structure retains the compatibility of the original network and lays the foundation for rapid iterative optimization. In the up-sampling stage, the concatenate calculation is performed after each deconvolution operation. With this operation, the last layer uses SoftMax to predict 0 and 1 for each pixel. Finally, the size of the image is recovered back to  $1024 \times 1024 \times 16$ .

#### 3.3.2. Configurations

We consider the following configurations parameters for our 3D PBV-Net:

- 1)  $R$ - the size of convolution kernel,
- 2)  $NL$ - the number of model's layer, and
- 3)  $NC$ - the number of each layer's convolution.

In computer vision, the convolution kernel is one of the most important concepts. It has the advantages of weight sharing and translation invariance, and its size will affect the number of model parameters and the information of the extracted image. Therefore, we designed

**Table 2**

Experiments with different convolution kernels and model structure.

Experiment Name	Model Structure the Number of Layer & Each Layer's Convolution		
	11 & (1,2,3,3,3,2,2,2,2,2)	11 & (1,2,3,3,3,2,3,3,3,2,1)	9 & (1,2,3,3,2,2,2,2)
Convolution Kernel Size	S1 S4 S7	S2 S5 S8	S3 S6 S9

the model with  $3 \times 3 \times 3$ ,  $5 \times 5 \times 5$  and  $7 \times 7 \times 7$  convolution kernel, respectively. Besides, minor adjustments to the overall structure may also affect the final result, including the number of model layer and the number of convolutions per layer. The depth of the model structure and the number of convolutions affect the depth of the extracted features. The deeper information can reflect the more accurate information of the target.

### 3.3.3. Dice loss layer

In terms of image segmentation, the target region usually only occupies a small area, which easily leads to poor prediction accuracy. In this study, we borrow the idea of the Dice coefficient in the loss layer [17]. This method achieves preprocessing by giving the background area a much smaller weight than the foreground area. The value of the objective function lies between 0 and 1, and our goal is to maximize it. The formula for the Dice coefficient between two binary convolutions is given as follows

$$D = 2 \sum_i^N p_i g_i / \left( \sum_i^N p_i^2 + \sum_i^N g_i^2 \right), \quad (3)$$

By differentiating formula (3), formula (4) can be obtained as

$$\frac{\partial D}{\partial p_j} = 2 \left[ \frac{g_j \left( \sum_i^N p_i^2 + \sum_i^N g_i^2 \right) - 2p_j \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \right], \quad (4)$$

where  $p$  and  $g$  are the predicted and standard image pixel values, respectively. Besides, this formula calculates the voxel, and it is no longer necessary to assign weights to achieve a balance between background and foreground pixels.

In the next stage of function optimization, we use stochastic gradient descent to perform the optimization. The weight  $W$  is updated by using a negative gradient  $\nabla L(W)$  and the previous weight is to continuously update the linear combination. The potential energy  $\mu \in [0, 1]$  represents the weight value of the previous iteration, and the learning rate  $\eta > 0$  represents the negative gradient weight. The definition of the weight matrix can be obtained as

$$\Delta_{i+1} = \mu \Delta_i + \eta \nabla L(W_i^l), \quad (5)$$

$$W_{i+1}^l = W_i^l + \Delta_{i+1}, \quad (6)$$

where  $l$  is the number of layers and  $i$  is the number of iterations. We can use formulas (5) and (6) to iterate  $\Delta_{i+1}$  and  $W_{i+1}^l$ , respectively. Moreover, this paper uses the Nesterov Accelerated Gradient (NAG) to update the weights [18]. The method is similar to the stochastic gradient descent method, and the specific formula is as follows

$$\Delta_{i+1} = \mu \Delta_i + \eta \nabla L(W_i^l + \mu \Delta_i), \quad (7)$$

For NAG, we perform a partial update of  $W_i^l$  and calculate the similarity of  $W_i^l + \mu \Delta_i$ . This is similar to update  $W_{i+1}^l$  but this process has better adaptability to unknown corrections.

### 3.3.4. Activation function

The conventional linear correction unit (ReLU) has an extensive

application and is first used in Boltzmann [19]. It is defined as follows

$$y = \max(0, Wx + b), \quad (8)$$

where  $W$  and  $b$  are the coefficients of the first-order linear equation. When the linear correction unit is used, only information greater than zero can be retained, and the information in other areas is also compressed to a certain degree. As a result, we can clearly find a series of defects such as too large amplitude, unstable network, and extremely high volatility.

To avoid the problem of excessive amplitude, we use a random nonlinear correction unit (RReLU) here, which is shown as

$$y = \max(0, W_1 * x + b_1) + \min(0, W_2 * x + b_2) / a, \quad (9)$$

where  $a$  represents the result of a uniform distribution,  $W_1$  represents the filtering that contains  $n_1$  filters of size  $c^*f^*f$ ,  $c$  represents the number of channels of the input layer image and  $*$  represents the convolution calculation.

### 3.4. Model training

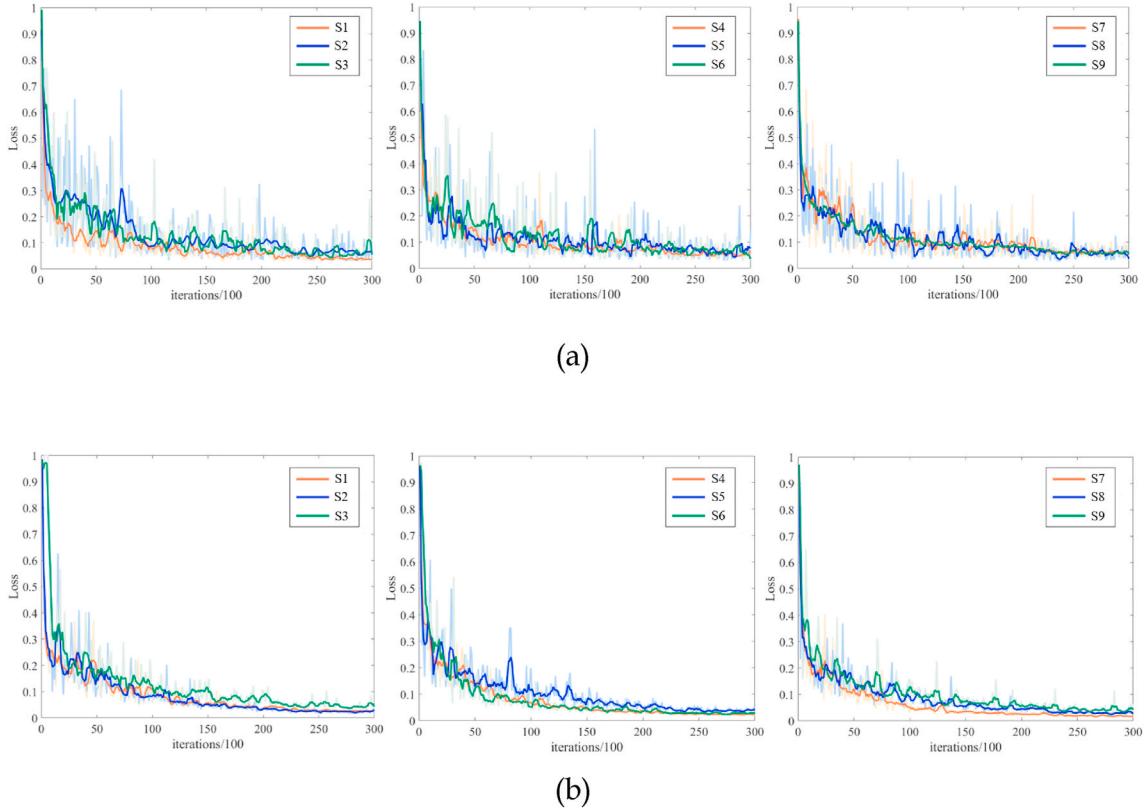
We design ablation studies for the design of different convolution kernels and model structures, and experimental settings are shown in Table 2. According to Table 2, we develop a total of 9 experiments from experiment S1 to experiment S9 for PROMISE 12 and TPHOH datasets, respectively.

We use the cross-validation method to make our results more convincing and robust [20]. In this study, we randomly selected 80% of the data from two datasets for training and cross-validation and the remaining 20% for testing. For PROMISE 12 datasets, after data preprocessing, there are 1692 images, among which 1311 are used for training and 381 for testing. For TPHOH datasets, there are 530 images after data preprocessing, among which 424 are used for training and 106 for testing. For PROMISE 12 and TPHOH datasets, we have carried out 30,000 iterations of training to obtain the expected results. The learning rate drops from 0.00005 at the beginning to 0.00001 after 20,000 steps to increase the training refinement and convergence speed. We also set the dropout rate to 0.8 to reduce the interdependence between nodes, thereby realizing the network's regularization and reducing its structural risk [21]. The batch size is set to 4. In this process, the probability greater than 0.5 of the pixel value is regarded as the target region, filled with white color. Otherwise, it is set to black color.

### 3.5. Evaluation metrics

To evaluate the results of the proposed model objectively, we use the Dice metric (DM), the Hausdorff distance (HD), and the average boundary distance (ABD) for quantifications. The metrics presented in this study are widely used in the evaluation of image segmentation. Both the boundary and volume metrics are chosen to estimate the accuracy and efficiency of segmentation results from multiple aspects. In addition, for the 3D PBV-Net, a volume measurement is more critical than a plane. It can not only reflect the size of the segmented target region but can also provide a reliable guarantee of the global diagnosis for the clinicians.

The DM measures the relative change between the automatically segmented regions and the manually delineated areas. The larger the



**Fig. 5.** Training loss curves on PROMISE 12 (a) and TPHOH (b) datasets.

calculated DM value is, the more accurate the segmentation is. DM is calculated as

$$DM = \frac{2 \times \text{Area}(A \cap B)}{\text{Area}(A) + \text{Area}(B)}, \quad (10)$$

where DM is the Dice metric,  $\text{Area}(A)$  is the region of resulting partition,  $\text{Area}(B)$  is the mark region and  $\text{Area}(A \cap B)$  is the region where the result intersects the mark.

In addition, HD is a measure to describe the similarity between two sets of points, which the unit is mm. As for the sets of  $Y$  and  $X$ ,  $HD(X, Y)$  is a directed distance that can be calculated by

$$HD_{\text{asym}}(X_S, Y_S) = \max_{x \in X_S} \left( \min_{y \in Y_S} d(x, y) \right), \quad (11)$$

where  $d(x, y)$  is the Euclidean distance operator, and  $Y_S$  and  $X_S$  are the sets of algorithm segmented points and the reference points, respectively. Then the HD is the maximum of the directed distances that can be defined as

$$HD(X_S, Y_S) = \max(HD_{\text{asym}}(X_S, Y_S), HD_{\text{asym}}(Y_S, X_S)), \quad (12)$$

After sorting the calculated distances, only 95% of them are taken to exclude the influence of the outliers of the results. Finally, the ABD is defined as the average shortest distance between two-point sets, which is a particular classic shape-based metric. ABD is given by

$$ABD(X_S, Y_S) = \frac{1}{N_{X_S} + N_{Y_S}} \left( \sum_{x \in X_S} \min_{y \in Y_S} d(x, y) + \sum_{y \in Y_S} \min_{x \in X_S} d(y, x) \right), \quad (13)$$

where  $N_{X_S}$  is the number of reference points and  $N_{Y_S}$  is the number of algorithm segmented points.

#### 4. Experimental results

##### 4.1. Implementation details

To illustrate the training process, we show the iterations and their corresponding training loss in Fig. 5. Fig. 5 (a) shows the training loss curves of the PROMISE 12 dataset, and Fig. 5 (b) shows the training loss

**Table 3**

Quantitative evaluation results of the PROMISE 12 dataset.

Experiment Name	the Model Structure	Convolution Kernel Size	Dice metric (DSC)	95% HD	ABD	Comprehensive Score	Time(h)
S1	11 & (1,2,3,3,3,2,2,2,2,2)	3*3*3	<b>0.9613 ± 0.020</b>	<b>3.120 ± 2.91</b>	1.708 ± 1.59	<b>0.3564</b>	8.3
S4		5*5*5	0.9394 ± 0.028	7.020 ± 13.5	1.867 ± 1.63	0.05331	8.2
S7		7*7*7	0.9378 ± 0.027	6.470 ± 3.66	1.834 ± 1.04	0.03212	8.4
S2	11 & (1,2,3,3,3,2,3,3,3,2,1)	3*3*3	0.9457 ± 0.032	7.904 ± 12.4	1.575 ± 2.47	0.1405	8.2
S5		5*5*5	0.9407 ± 0.031	5.637 ± 8.04	1.450 ± 2.07	0.07123	8.3
S8		7*7*7	0.9429 ± 0.031	4.970 ± 3.88	<b>1.603 ± 1.25</b>	0.1017	8.3
S3	9 & (1,2,3,3,2,2,2,2)	3*3*3	0.9482 ± 0.027	4.507 ± 6.08	1.678 ± 2.26	0.1751	7.2
S6		5*5*5	0.9406 ± 0.025	5.562 ± 6.79	1.703 ± 1.89	0.06990	7.4
S9		7*7*7	0.9356 ± 0.031	5.879 ± 3.42	2.056 ± 1.19	10e-04	7.3

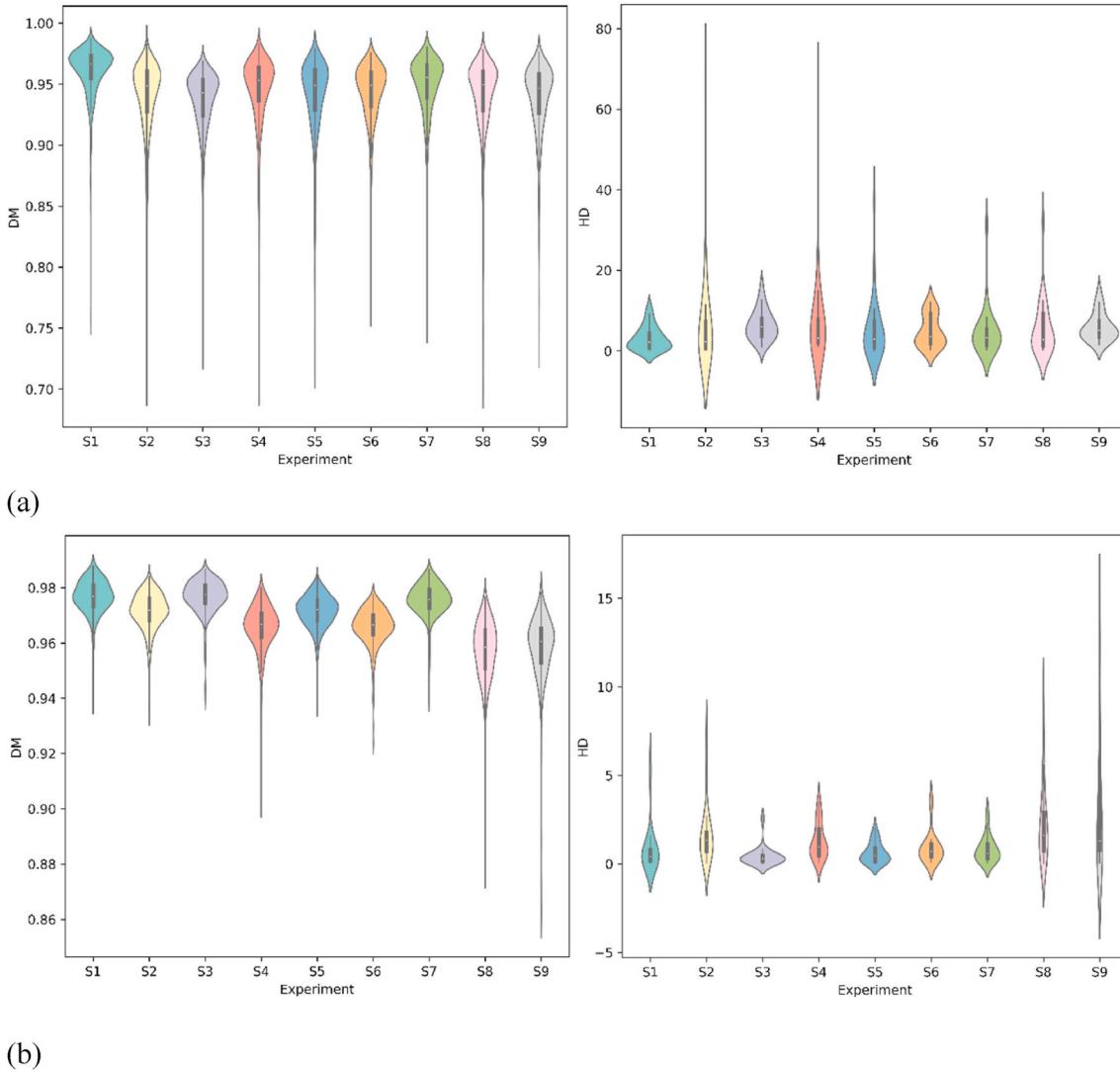
Bold numbers indicate the best results.

**Table 4**

Quantitative evaluation results of the TPHOH dataset.

Experiment Name	the Model Structure	Convolution Kernel Size	Dice metric (DSC)	95% HD	ABD	Comprehensive Score	Time(h)
S1	11 & (1,2,3,3,3,2,2,2,2,2)	3*3*3	0.9765 ± 0.007	0.9382 ± 1.45	<b>0.7950 ± 1.23</b>	<b>0.3522</b>	7.4
S4		5*5*5	0.9713 ± 0.008	1.636 ± 1.66	0.9402 ± 0.95	0.2016	7.4
S7		7*7*7	<b>0.9769 ± 0.006</b>	<b>0.4062 ± 0.52</b>	0.9744 ± 1.25	0.2368	7.6
S2	11 & (1,2,3,3,3,2,3,3,3,2,1)	3*3*3	0.9658 ± 0.007	1.276 ± 0.99	1.457 ± 1.13	10e-04	7.3
S5		5*5*5	0.9714 ± 0.006	0.6131 ± 0.57	1.238 ± 1.15	0.05774	7.4
S8		7*7*7	0.9660 ± 0.011	0.9123 ± 0.90	1.090 ± 1.08	0.1118	7.4
S3	9 & (1,2,3,3,2,2,2,2,2)	3*3*3	0.9756 ± 0.006	0.8093 ± 0.72	0.8351 ± 0.75	0.3153	6.8
S6		5*5*5	0.9573 ± 0.006	2.338 ± 2.24	1.545 ± 1.67	10e-04	6.9
S9		7*7*7	0.9585 ± 0.012	3.093 ± 3.87	1.609 ± 1.88	10e-04	7.0

Bold numbers indicate the best results.

**Fig. 6.** Violin plot of the DM and HD indicators for PROMISE 12 (a) and TPHOH (b).

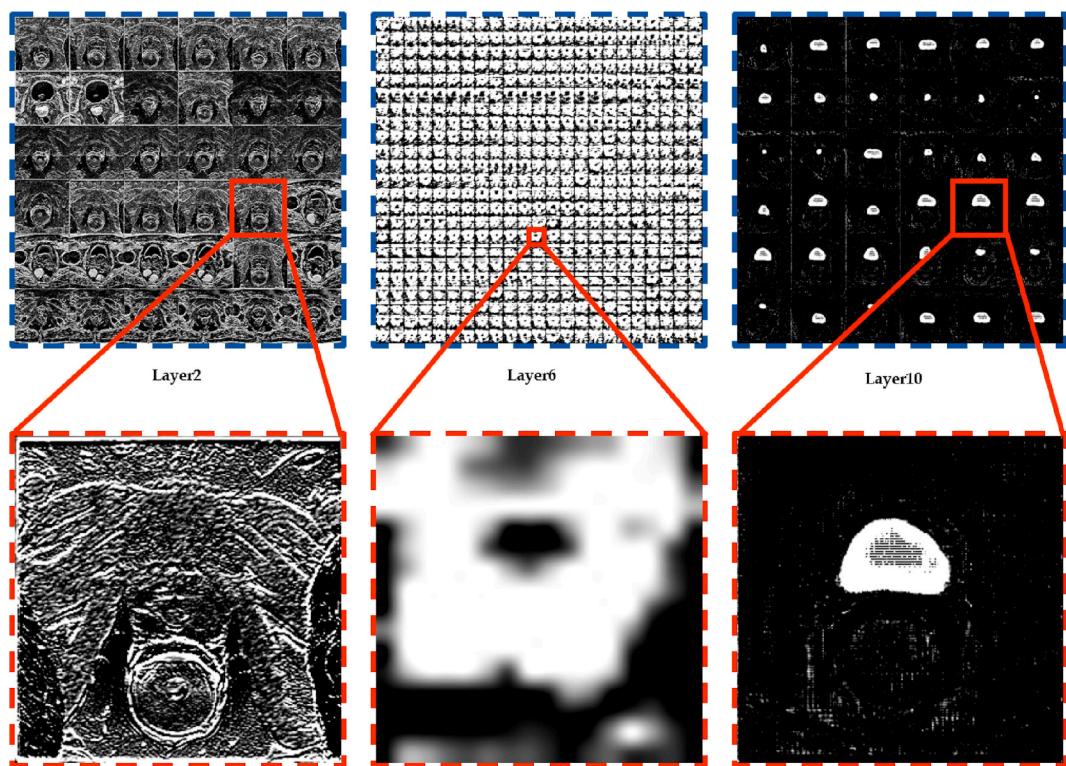
curves of the TPHOH dataset. For the loss value, we normalize the result value between 0 and 1.

For this study, we used the python-based TensorFlow framework to implement the preprocessing and our 3D PBV-Net model, including 1.18.4 NumPy, 4.1.2 OpenCV, and the GPU version of 1.4.0 TensorFlow. The workstation is installed with the Window10 system, with one AMD Ryzen 7 2700  $\times$  3.70 GHz\*16 CPUs, one NVIDIA GeForce RTX 2080 16 GB GPUs, and the 32 GB of running memory. The training time costs of the 3D PBV-Net on PROMISE 12 and TPHOH are around 8 h for each dataset. Besides, each prostate MR image takes only a few seconds to

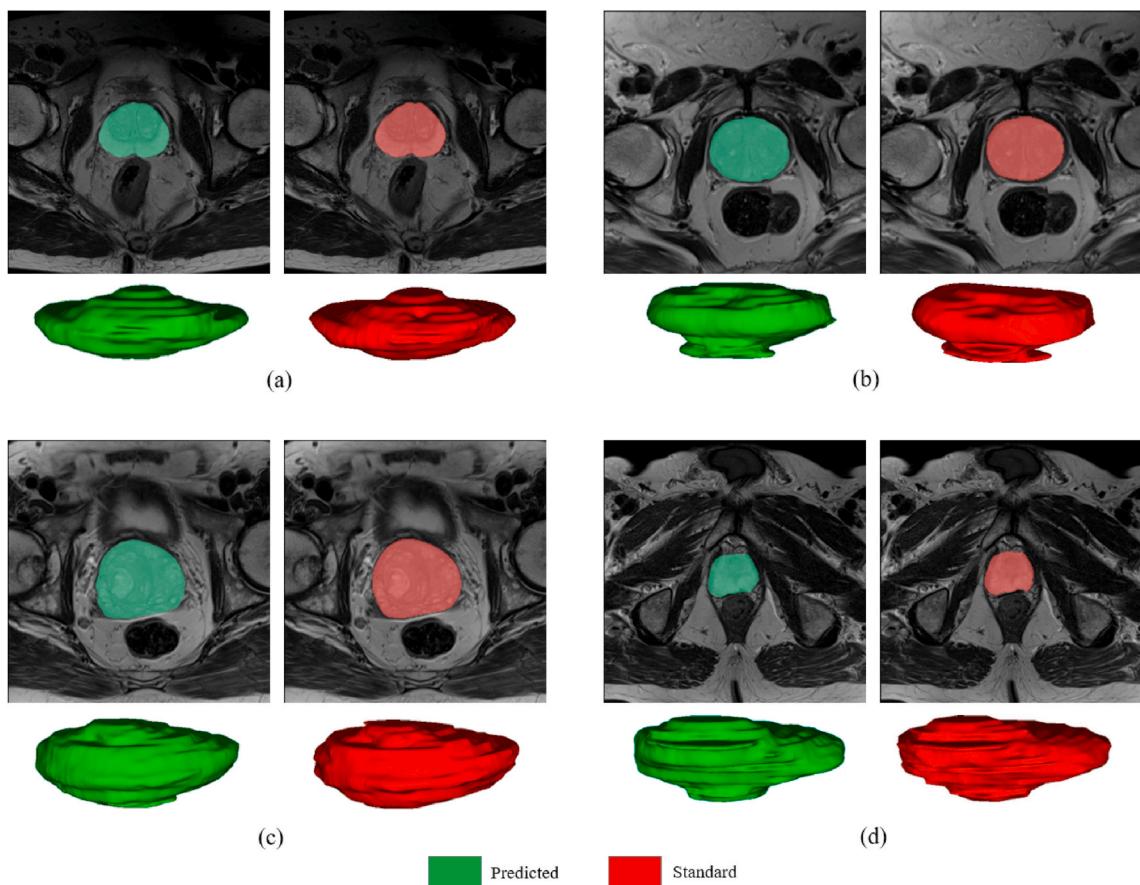
predict.

#### 4.2. Segmentation performance

We evaluate our proposed 3D PBV-Net on PROMISE 12 and TPHOH datasets using DM, HD, and ABD. For the two datasets, we present the results of 9 experiments in Table 3 and Table 4. The value means mean  $\pm$  std in the table. The comprehensive score in Tables 3 and 4 is calculated by the entropy weight method of DM, HD, and ABD. The weight value of each index is obtained by calculating the information entropy,



**Fig. 7.** Visualization of the feature maps.



**Fig. 8.** Prostate segmentation results.

**Table 5**

Prostate segmentation results of classic methods on PROMISE 12.

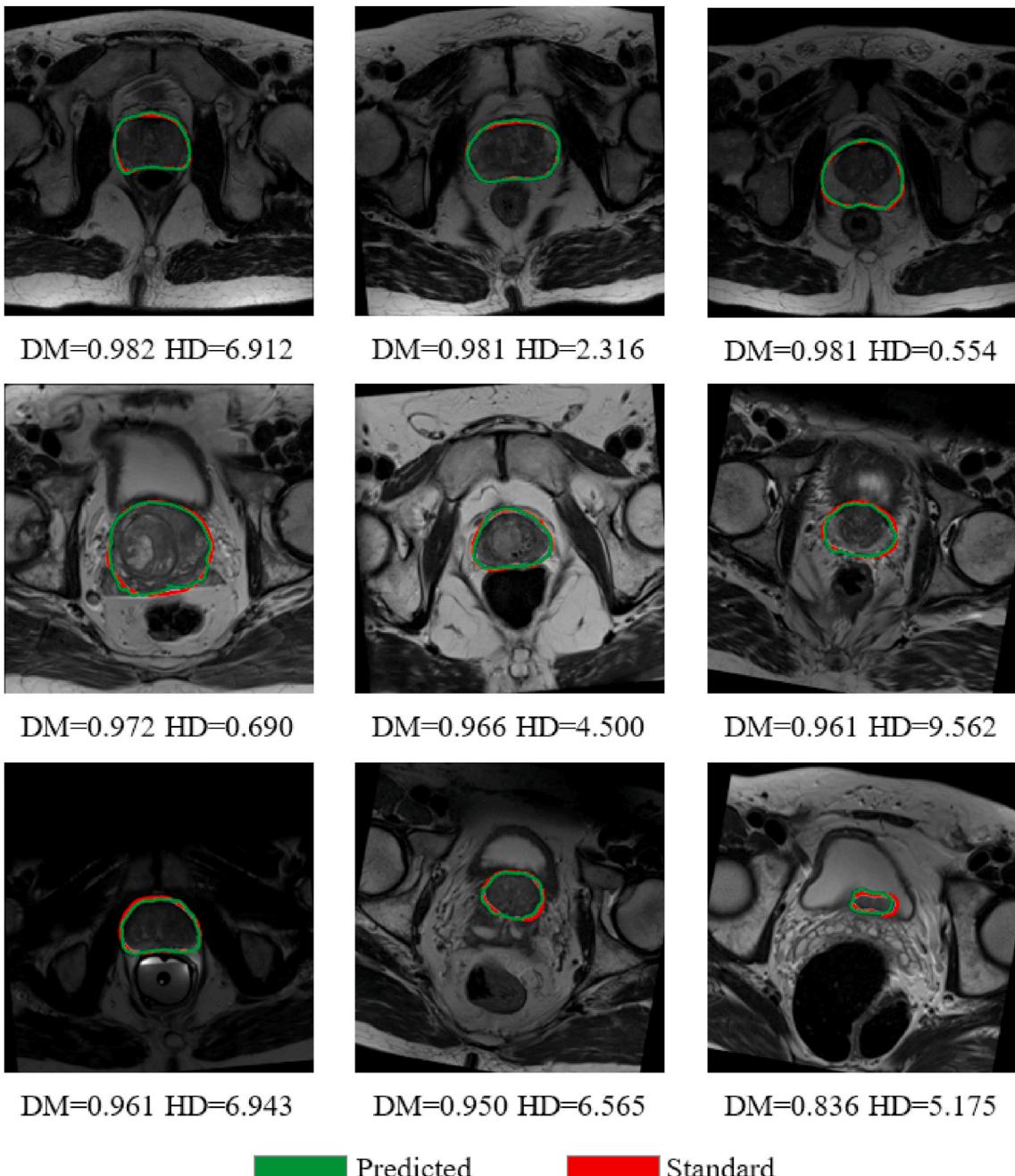
Provenance	Algorithm	Dice metric (DSC)	95% HD	ABD
Ours	3D PBV-NET	0.9613	3.120	1.708
Milletari et al.	V-NET + Dice loss	0.8720	4.510	/
Vincent et al.	BOWDA-Net	0.8790	3.785	/
Jia et al.	DCNN	0.8810	4.995	1.736
Zhu et al.	BOWDA-Net	0.9254	6.420	1.580
Jia H et al.	3D APA-Net69	0.9060	4.128	1.454
Ghavami et al.	HighRes3dNet	0.8900	/	2.330

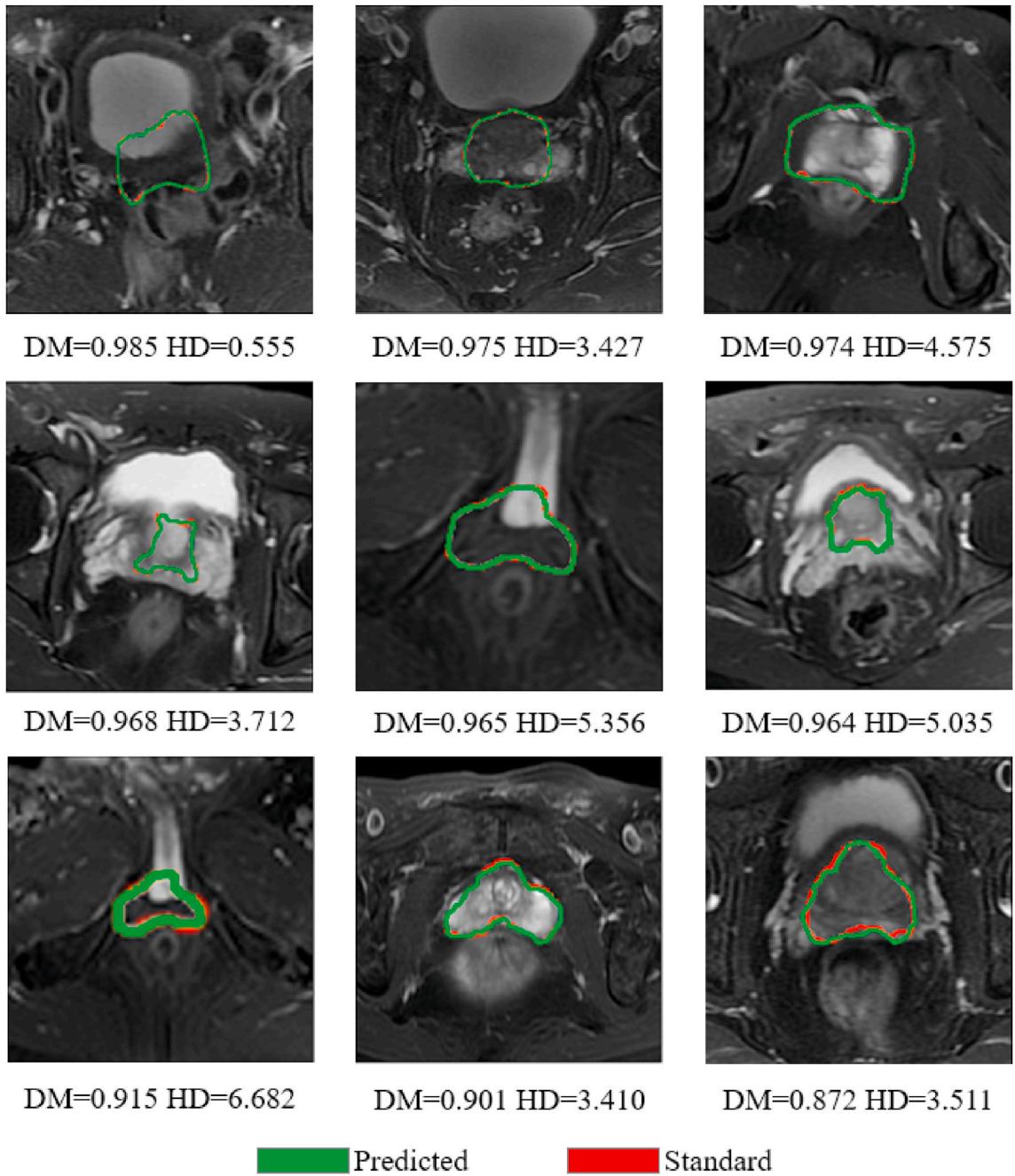
and the closer the index is, the smaller the value is [22]. In the column of the comprehensive score, the score of 10e-04 indicates that the value closes to 0. We can also observe that experiment S1 achieves better results for both datasets. Reducing the depth and number of image

information extraction is not conducive to the effective segmentation of the target region. Moreover, a too-large convolution kernel does not significantly improve the overall result but increases the calculation time.

On this basis, we make a statistical analysis of the results for each experiment. The violin plot is a statistical chart used to display a set of data dispersion information. It is mainly used to reflect the original data distribution characteristics, and it can also compare multiple sets of data distribution characteristics and probability density. The result is given in Fig. 6, where (a) is the violin plot of the DM and HD (mm) indicators for PROMISE 12, and (b) is the violin plot of the DM and HD indicators for TPHOH.

To sum up, for PROMISE 12, the experimental results show that the recognition accuracy rate of prostate MRI data reaches 97.65%. The accuracy rate is the proportion of the target region that is predicted

**Fig. 9.** Visualization results of the prostate segmentation using the proposed 3D PBV-Net on the PROMISE 12.



**Fig. 10.** Visualization results of the prostate segmentation using the proposed 3D PBV-Net on the TPHOH dataset.

correctly divided by the total regions. In addition, DM reaches 0.9613, HD reaches 3.120 mm and ABD reaches 1.708. For TPHOH, the accuracy rate reaches 98.29%, DM reaches 0.9765, HD reaches 0.9382 mm and ABD reaches 0.7950. From the calculated results, it can be seen that our model shows promising results in multiple levels, such as boundary, accuracy, and three-dimensional space. This further verifies the correctness of our preprocessing and the effectiveness of model tuning.

Fig. 7 shows the hidden layer output of our 3D PBV-Net. It suggests that the first layer network and the second layer network have learned simple features, such as edges, connections, and corners. The later layers learn advanced features to produce highly abstract output images. The second column is the image display of each layer, where the target area becomes more and more evident.

On this basis, we also perform the segmentation results and 3D reconstruction and rendering to provide a more intuitive visualization

using the 3D Slicer software. As shown in Fig. 8, there are four groups ((a)~(d)) in total, among which the green stereogram is the predicted results and the red stereogram is standard manual segmentation. It can be seen that the results obtained by our model are very similar to the given label in general and details, which can provide a more reliable remote medical diagnosis.

## 5. Discussion

We conduct an in-depth analysis of the results obtained. To get a better view of the results on PROMISE 12, we have compared the results of prostate MRI data segmentation with other studies. The corresponding comparison is shown in Table 5. As can be seen from the table, Milletari et al. used the V-NET model to train on the PROMISE 12 dataset with a DM reached 0.8720 and an HD reached 4.510 [17]. Vincent et al.

presented the BOWDA-Net, in which DM reached 0.8790 and HD reached 3.785 [23]. In addition, Jia et al. proposed DCNN to get good results, which DM reached 0.8810, HD reached 4.995 and ABD reached 1.736 [24]. Zhu et al. used the BOWDA-Net, in which DM reached 0.9254, HD reached 6.420 and ABD reached 1.580 [13]. Jia H et al. tried the BOWDA-Net to segment prostate MRI data, which DM reached 0.9060, HD reached 4.128 and ABD reached 1.454 [25]. Ghavami et al. presented the HighRes3dNet, which DM reached 0.8900 and ABD reached 2.330 [26].

By comparing the results of different algorithms in Table 5, we can see that our model has certain advantages with a DM and 95% HD indicators. It shows that the target region of our prediction has a good overlap with the given standard manual segmentation. But our method has relatively poor ABD results. This is due to the problem of too many or too few segmentation boundaries caused by the complexity of the prostate tissue.

Further visualization has been shown for the PROMISE 12 datasets. In Fig. 9, the boundary of the target region is bolded. There is a total of 9 results shown, each of which is the result of zoomed-in and cropped target region in order to observe our results more clearly. The red line in the figure is the standard manual segmented area provided by the dataset, and the green line is the area predicted by our optimal model. It can be seen that our model has a certain degree of robustness and accuracy for large, medium, and small regions and regions of various shapes.

After discussing the results on PROMISE 12, we also visualized the results of TPHOH. The region enclosed by the red line is the standard area of the prostate MRI data segmentation, while the region enclosed by the green line is the area predicted by our optimized model (Fig. 10.). Our model has achieved satisfactory results on this clinical dataset.

Our research might still have some limitations. Due to limited testing datasets, our developed model may provide various results when dealing with data collected from different clinical centers and different scanners. However, based on the current experiments on two datasets, we can prove the efficacy of the proposed method and we believe with some fine-tuning and transfer learning strategies we can mitigate the performance variations. Moreover, our current evaluations are based on the most widely used segmentation scores; however, there is currently no validation using clinically relevant metrics. For future work, we will focus on multimodal prostate MR image segmentation and continue to expand the current work to a large multi-center and multi-scanner study. In addition, the design and implementation of a clinical level automatic prostate segmentation system with clinical validations will be carried out.

## 6. Conclusion

This study has been developed based on the most popular neural networks and deep learning approaches to achieve automated segmentation for prostate MRI data. Most previously proposed deep learning models first convert 3D MRI data into 2D image slices, and then use them to train and optimize the network. This undoubtedly wastes a lot of time on training the model and does not conform to the stereoscopic characteristics of the 3D slice itself. For our proposed 3D PBV-Net, the input layer is a 3D matrix, and the number of feature maps of the original input is 1, which can speed up the training progress and improve the training efficiency. On this basis, we also incorporate some methods of image preprocessing, namely the bicubic interpolation algorithm and image rotation-based augmentation. In addition, we test the model on TPHOH and PROMISE 12, all of which show the reliability of our model. The experimental results show that the prostate MRI data segmentation framework based on our 3D PBV-Net has significant advantages such as higher accuracy and less time for segmentation compared to other algorithms. As the accurate segmentation results have been achieved, we can envisage more intuitive and accurate prostate MRI data segmentation results to be provided for telehealth. It may also alleviate the

problem of the high referral rate and the high cost for patients in remote areas in the future.

## Authors contribution

Y. Jin, G. Yang, and X. Lai conceived and designed the study. Y. Jin, G. Yang, Y. Liu, Z. Niu, and X. Lai contributed to the literature search. Y. Jin, G. Yang, R. Li, and X. Xu contributed to data analysis and data curation. Y. Jin, G. Yang, R. Li, X. Xu, and X. Lai contributed to data visualization. Y. Jin and X. Xu contributed to software implementation. Y. Jin, G. Yang, Y. Fang, R. Li, X. Xu, and X. Lai contributed to the tables and figures. Y. Jin, G. Yang, Y. Fang, R. Li, X. Xu, and X. Lai contributed to the writing of the report. Y. Jin, G. Yang, Y. Liu, Z. Niu, and X. Lai contribute to review and editing. All the authors have read and approved the publication of this work.

## Declaration of competing interest

The authors declare no conflict of interest.

## Acknowledgment

We would like to thank Ruipeng Li from the Hangzhou Third People's Hospital to collect the new prostate MRI data dataset for our study. Furthermore, we appreciate the organizers of PROMISE 12 to share the prostate MRI data for us to test our models and methods. This work is funded in part by the National Natural Science Foundation of China (Grant No. 62072413), and also supported in part by the AI for Health Imaging Award 'CHAIMELEON: Accelerating the Lab to Market Transition of AI Tools for Cancer Management' [H2020-SC1-FA-DTS-2019-1 952172].

## References

- [1] Y. Shao, J. Wang, B. Wodlinger, S.E. Salcudean, Improving prostate cancer (PCa) classification performance by using three-player minimax game to reduce data source heterogeneity, *IEEE Trans. Med. Imag.* 1 (1) (2020) 99, <https://doi.org/10.1109/TMI.2020.2988198>.
- [2] W. Yang, Y. Shi, S.H. Park, M. Yang, Y. Gao, D. Shen, An effective MR-guided CT network training for segmenting prostate in CT images, *IEEE Journal of Biomedical and Health Informatics* 24 (8) (2019) 2278–2291, <https://doi.org/10.1109/JBHI.2019.2960153>.
- [3] T. Hassanzadeh, L.G.C. Hamey, K. Ho-Shon, Convolutional neural networks for prostate magnetic resonance image segmentation, *IEEE Access* 7 (2019) 36748–36760, <https://doi.org/10.1109/ACCESS.2019.2903284>.
- [4] D. Bo, J. Liao, B. Turkbey, P. Yan, Multi-task learning for registering images with large deformation, *IEEE J Biomed Health Inform* (2020), <https://doi.org/10.1109/JBHI.2020.3016699>.
- [5] Q. Liu, Q. Dou, L. Yu, P. Heng, MS-net: multi-site network for improving prostate segmentation with heterogeneous MRI data, *IEEE Trans. Med. Imag.* 39 (9) (2016) 2713–2724, <https://doi.org/10.1109/TMI.2020.2974574>.
- [6] X. Ding, D. Clifton, N. Ji, N.H. Lovell, P. Bonato, W. Chen, et al., Wearable sensing and telehealth technology with potential applications in the coronavirus pandemic, *IEEE Rev Biomed Eng* (2020), <https://doi.org/10.1109/RBME.2020.2992838>.
- [7] S. Macis, D. Loi, A. Ulgheri, D. Pani, G. Solinas, S. Manna, et al., Design and usability assessment of a multi-device SOA-based telecare framework for the elderly, *IEEE J Biomed Health Inform* 24 (1) (2020) 268–279, <https://doi.org/10.1109/JBHI.2019.2894552>.
- [8] Z. Gao, Y. Li, Y. Sun, J. Yang, S. Li, Motion tracking of the carotid artery wall from ultrasound image sequences: nonlinear state-space approach, *IEEE Trans. Med. Imag.* 37 (1) (2018) 273–283, <https://doi.org/10.1109/TMI.2017.2746879>.
- [9] Q. Yu, Y. Shi, J. Sun, Y. Gao, J. Zhu, Y. Dai, Crossbar-Net: a novel convolutional neural network for kidney tumor segmentation in ct images, *IEEE Trans. Image Process.* (2019), <https://doi.org/10.1109/TIP.2019.2905537>.
- [10] Y. Yuan, W. Qin, X. Guo, Prostate segmentation with encoder-decoder densely connected convolutional network (Ed-Densenet), in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019, pp. 434–437.
- [11] I. Reda, A. Shalaby, M. Elmogy, M. Ghazal, A. El-Baz, A new fast framework for early detection of prostate cancer without prostate segmentation, in: 2018 IEEE International Conference on Imaging Systems and Techniques (IST), 2018, pp. 1–5.
- [12] B. He, D. Xiao, Q. Hu, F. Jia, Automated magnetic resonance image prostate segmentation based on adaptive feature learning probability boosting tree initialization and CNN-ASM refinement, *IEEE Access* 6 (2018), <https://doi.org/10.1109/ACCESS.2017.2781278>, 2005.

- [13] Q. Zhu, B. Du, P. Yan, Boundary-weighted domain adaptive neural network for prostate MRI data segmentation, *IEEE Trans. Med. Imag.* 39 (3) (2020) 753–763, <https://doi.org/10.1109/TMI.2019.2935018>.
- [14] H. Ócal, N. Barışçı, Prostate segmentation via fusing the nested-V-net3d and V-net2d, in: 2019 1st International Informatics and Software Engineering Conference (UBMYK), 2019, pp. 1–4.
- [15] Z. Khan, N. Yahya, K. Alsaith, F. Meriaudeau, Zonal segmentation of prostate T2W MRI using atrous convolutional neural network, in: 2019 IEEE Student Conference on Research and Development (SCORED), 2019, pp. 95–99.
- [16] J. Koljonen, V.A. Bochko, S.J. Lauronen, J.T. Alander, Fast fixed-point bicubic interpolation algorithm on FPGA, in: 2019 IEEE Nordic Circuits and Systems Conference (NORCAS): NORCHIP and International Symposium of System-On-Chip (SoC), 2019, pp. 1–7.
- [17] Fausto Milletari, V-net: fully convolutional neural networks for volumetric medical image segmentation, in: Fourth International Conference on 3D Vision, 2016, pp. 565–571, <https://doi.org/10.1109/3DV.2016.79>.
- [18] P. Gu, S. Tian, Y. Chen, Iterative learning control based on Nesterov accelerated gradient method, *IEEE Access* 7 (2019) 115836–115842, <https://doi.org/10.1109/ACCESS.2019.2936044>.
- [19] Z. Hu, Y. Li, Z. Yang, Improving convolutional neural network using pseudo derivative ReLU, in: 2018 5th International Conference on Systems and Informatics (ICSAI), 2018, pp. 283–287.
- [20] K. Pal, B.V. Patel, Data classification with k-fold cross validation and holdout accuracy estimation methods with 5 different machine learning techniques, in: 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 83–87.
- [21] E.C. Medina, C.B. Chunga, J. Armas-Aguirre, E.E. Grandón, Predictive model to reduce the dropout rate of university students in Perú: bayesian Networks vs, in: Decision Trees. 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), 2020, pp. 1–7.
- [22] C. Qiao, W. Zhang, Research on classification method of CRH maintenance parts based on entropy weight-clustering analysis, in: 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), 2019, pp. 517–521.
- [23] G. Vincent, G. Guillard, M. Bowes, Fully Automated Segmentation of the Prostate Using Active Appearance Models. MICCAI Grand Challenge PROMISE 2012, 2012.
- [24] H. Jia, Y. Xia, W. Cai, M. Fulham, D.D. Feng, Prostate segmentation in MRI data using ensemble deep convolutional neural networks, in: IEEE 14th International Symposium on Biomedical Imaging, 2017, pp. 762–765.
- [25] H. Jia, Y. Xia, Y. Song, 3D APA-net: 3D adversarial pyramid anisotropic convolutional network for prostate segmentation in MRI data, *IEEE Trans. Med. Imag.* 39 (2) (2020) 447–457, <https://doi.org/10.1109/TMI.2019.2928056>.
- [26] Nooshin Ghavami, Automated segmentation of prostate MRI using convolutional neural networks: investigating the impact of network architecture on the accuracy of volume measurement and MRI-ultrasound registration, *Med. Image Anal.* 58 (2019).