Biomedical Signal Processing and Control 85 (2023) 104755

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

A dual attention-guided 3D convolution network for automatic segmentation of prostate and tumor

Yuchun Li [a], Mengxing Huang [a,*], Yu Zhang [b,*], Siling Feng [a], Jing Chen [c], Zhiming Bai [c]

[a] State Key Laboratory of Marine Resource Utilization in South China Sea, School of Information and Communication Engineering, Hainan University, Haikou 570288, China
[b] College of Computer Science and Technology, Hainan University, Haikou 570288, China
[c] Haikou Municipal People's Hospital and Central South University Xiangya Medical College Affiliated Hospital, Haikou 570288, China

ARTICLE INFO

Keywords:
Prostate segmentation
Tumor segmentation
Visual attention
Scale attention
DWI

ABSTRACT

Background: In middle-aged and older men, prostate cancer (PCa) is a common tumor disease with a mortality rate second only to lung cancer. The automatic and accurate segmentation of the prostate and tumor in magnetic resonance imaging (MRI) images can help doctors diagnose malignancies more efficiently. T2 weighted imaging (T2W) is now used in the majority of studies on prostate MRI image segmentation; however, diffusion-weighted imaging (DWI) is more valuable in the diagnosis of PCa. The morphological differences between the prostate and tumor regions are minimal, the tumor size is uncertain, the border between the tumor and surrounding tissue is hazy, and the categories separating normal regions from tumors are uneven. Consequently, it is challenging to segment prostate and tumor on DWI images.

Methods: For the segmentation of prostate and tumor regions on DWI images, this study offers a dual attention-guided 3D convolutional neural network (3D DAG-Net). A visual attention method is built into the encoder step to obtain the features of various receptive fields and deliver more detailed contextual information. A multiscale attention technique is proposed at the decoder stage to fuse multiscale features to acquire finer global and local details. To resolve the class discrepancies between the prostate, tumor, and background regions in segmentation tasks, we propose a hybrid loss function for handling class imbalance.

Results: We tested the algorithm on DWI images of PCa obtained from a nearby hospital, demonstrating the uniqueness and effectiveness of the method. Dice similarity coefficient (DSC) values for prostate and tumor DWI segmentation were 92.28% and 88.73%, respectively.

Conclusion: We present a unique dual-attention mechanism 3D segmentation network architecture for quantitative assessment of prostate and tumor volumes on DWI. The automatic segmentation results produced by our technology were highly correlated and consistent with expert manual segmentation findings.

1. Introduction

Prostatic cancer (PCa) is a frequent tumor condition in middle-aged and elderly men and the second leading cause of cancer death in males. According to statistics, the annual incidence rate of PCa is 1.41 million, accounting for 14.1% of all male cancer cases[1]. The diagnosis of PCa mostly depends on the experience of doctors, which is limited by their professional level and subjective judgment. Therefore, there is an urgent need for automatic diagnosis technology for PCa to help doctors diagnose tumors. Prostate-specific antigen (PSA) testing, needle biopsy, ultrasound, and magnetic resonance imaging (MRI) are some of the current clinical detection approaches for PCa (MRI). PSA can lead to unnecessary puncture biopsy[2], and puncture biopsies have a high false-negative rate owing to the uneven placement of cancer [3]. Multiparameter MRI (mpMRI) is widely used in the diagnosis of PCa because of its high sensitivity and rich image information. Therefore, MRI is crucial for the diagnosis, detection, and treatment of PCa [4]. Diffusion-weighted imaging (DWI) is an imaging technology of mpMRI, that also includes T1 weighted imaging (T1W), T2 weighted imaging (T2W), and dynamic contrast-enhanced imaging (DCE). DWI, a water molecule diffusion technique, allows examination of cellular and tissue structure of the prostate [5]. The water molecules in the human body, both free and bound, are the key contributing elements in DWI because they are always in random motion. PCa has a different spread coefficient

* Corresponding authors.
E-mail addresses: huangmx09@163.com (M. Huang), yuzhang_nwpu@163.com (Y. Zhang).

https://doi.org/10.1016/j.bspc.2023.104755
Received 2 April 2022; Received in revised form 7 February 2023; Accepted 18 February 2023
Available online 30 April 2023

value than normal prostate tissue, which makes it easier to diagnose PCa. DWI is the most common sequence determination method for PCa in the peripheral zone (approximately 70%) and has advantages in detecting PCa in the transition zone [6] (see Table 1).

Physicians currently quantitatively assess PCa by manually defining tissue regions on 3D MRI slices. However, this is a time-consuming and laborious task that adds to the workload of radiologist. Furthermore, personal experience and subjective consciousness had a significant impact on PCa accuracy, which is not conducive to PCa evaluation. Therefore, an automated, accurate, and efficient method for quantifying PCa is urgently required. Automatic segmentation technology is increasingly commonly employed in medical images as machine learning and deep learning technologies advance. Researchers have been interested in organ and lesion segmentation techniques based on computed tomography (CT) and MRI, such as brain tumor segmentation [7–10], lung cancer segmentation [11], breast tumor segmentation [12,13], nerve optic segmentation [14], COVID-19 segmentation [15] and prostate segmentation [16]. Because of its distinct internal tissue structure, T2W is used in the majority of prostate MRI image segmentation studies. The challenge of segmenting the prostate using T2W MRI was proposed by Montagne et al. [17]. The variability of extreme parts of the prostate gland is high, and the automatic segmentation of the prostate is greatly affected by changes in prostate morphology (volume, area, and intensity ratio). In the automatic segmentation task of prostate T2W, a two-step residual network [18] and a boundary coding network [19] were proposed to solve the boundary ambiguity problem. The multi-site network [20] and scenario learning in continuous frequency space (ELCFS) method [21] are proposed to solve the prostate segmentation problem of different source datasets. On T2W, most studies segment the prostate region and leave the tumor area out. Furthermore, studies on prostate and tumor segmentation approaches using DWI are limited. DWI, on the other hand, is more useful in the diagnosis of PCa [22]. The degree and the area on DWI with diverse characteristics, varying sizes, and hazy borders affect PCa. Prostate anatomy has a low signal-to-noise ratio and considerable variability, which makes DWI segmentation difficult. In DWI images. Fig. 1 depicts the fundamental morphology of the prostate and tumor in the DWI images. The original and annotated sequence images of the two examples are shown in (a) and (b), respectively. The prostate is represented by a yellow transparent area in the annotation image, whereas the tumor is represented by a blue transparent area. Fig. 1(a) shows the tumors in the transition zone (TZ) and peripheral zone (PZ) regions of case0 patient slices. The challenge of automatically segmenting the prostate and tumors increases in the presence of multiple tumors. The shape and size of the tumor area fluctuate with different slice photos, as shown in Fig. 1(b), and the complicated and changeable tumor area makes it more difficult to segment the task automatically. In addition, Fig. 1(a) and (b) contain areas that could be mistaken for tumors, and determining how to extract tumor areas effectively and reliably is critical.

For the identification and segmentation of early stage PCa, automatic detection and segmentation both extract handcrafted features and apply classifiers [23–26]. Chan et al. [23] suggested a method for extracting texture data and combining various classifiers to identify PCa using a co-occurrence matrix and a discrete cosine transform. They used the Fisher linear discriminant classifier to obtain the average receiver operator characteristic with the best detection of PCa, and the result reached 0.839. This time-honored feature classification method yields poor results. Based on the K-nearest neighbor and hidden Markov models, Llobet et al. [24] categorized PCa and non-cancer tissues with an accuracy of only 61.6%. Liu et al. [25] proposed an unsupervised PCa-detection approach. The MRF and class parameters are estimated alternately to enhance the clustering accuracy of multispectral PCa MRI segmentation based on the Fuzzy Markov random (MRF) field, and the MRF and class parameters are estimated alternately to improve the clustering accuracy. The Dice measure of the PCa segmentation results was 0.62. However, because this method is unsupervised, it will yield large errors, which will make the accurate diagnose of PCa difficult. McClure et al. [26] developed a three-dimensional framework for the automatic segmentation of PCa. The framework addresses the location and intensity of voxels while including the level-set deformation model and non-negative matrix decomposition (NMF) technology. Their method yielded an average value of 0.868. Although the conventional method of hand-drawn features has achieved some success, it still has problems. Convolutional neural networks (CNNs) [27] were used to compensate for the shortcomings of handwritten features. Wildeboer et al. [28] discussed the progress in computer-aided diagnosis of prostate cancer in the past few decades. The development of deep learning technology may alleviate the variability between and within observers. To distinguish carcinogenic tissue from non-cancerous tissue and establish whether it has clinical significance, Le et al. [29] developed an automatic diagnosis of PCa based on a multimodal CNN. The sensitivity and specificity of their method for distinguishing between PCa and non-cancerous tissues were 89.85% and 95.83%, respectively. Compared with traditional models, these models are more accurate. Yang et al. [30] proposed a multimodal multi-label CNN to locate prostate cancer in mpMRI images and trained it in a weak supervision manner by providing a group of prostate images with image-level labels. The recall rate of the automatic detection of PCa reached 80%. Cao et al. [31] developed a multi-task convolutional neural network for PCa detection and segmentation of malignant regions. They achieved 75.1% a PCa detection sensitivity. The conditional random field was used for post-processing, and its accuracy was 20% higher than that of the baseline. Abbasi et al. [32] used the transfer learning method to test the detection effect of PCa using decision trees, support vector machines, Bayesian, and CNN models (GoogleNet). The sensitivity of Google Net was 98.83%, which was the best classification result. Lai et al. [33] utilized a deep convolutional neural network (DCNN) model to diagnose PCa on dual-parameter MRI images. The Dice similarity coefficient in the PCa area was 52.73%. To segment the prostate region, Abdelmaksoud et al. [34] employed non-negative matrix factorization to combine three types of data. To detect PCa, the transfer learning of previously pretrained CNN models (AlexNet and VGGNet) was utilized, and deeper CNNs enhanced prostate segmentation and PCa detection accuracy (89.2% and 91.2%, respectively). Duran et al. [35] proposed a new end-to-end multiclass network for joint prostate segmentation and the detection and classification of cancer lesions. After encoding the information in the potential space, the network is divided into two branches for segmentation and detection. Malibari et al. [36] developed a new prostate cancer classification model based on deep learning (DTL-PSCC). Based on the feature extraction an efficient network, the fuzzy k-nearest neighbor (FKNN) model was used for classification after the krill-swarm algorithm (KHA) was used to optimize the membership value. The maximum accuracy of prostate cancer classification was 85.09%. Their approach was tested on MRI T2W images and performed satisfactorily. DWI images, on the other hand, provide significant hurdles in segmenting prostate regions and diagnosing PCa owing to their poor signal-to-noise ratio and substantial variability in prostate morphology. Although researchers' technology aids in the diagnosis of PCa, it is unable to precisely segment the PCa area and can only identify whether or not there is cancer. In addition, the volume of PCa could not be determined.

To address the issues of morphological differences, variable sizes, unclear boundaries, and unbalanced categories between the prostate and tumor regions. For the segmentation of prostate and tumor regions in DWI images, this research offers a dual attention-guided 3D convolutional neural network (3D DAG-Net). The visual attention method was established in the encoder and a multi-scale attention technique was

Table 1
Distribution of the number of DWI modalities in PCa patients.

Years	2013	2014	2015	2016	Total
Quantity	27	16	24	31	98

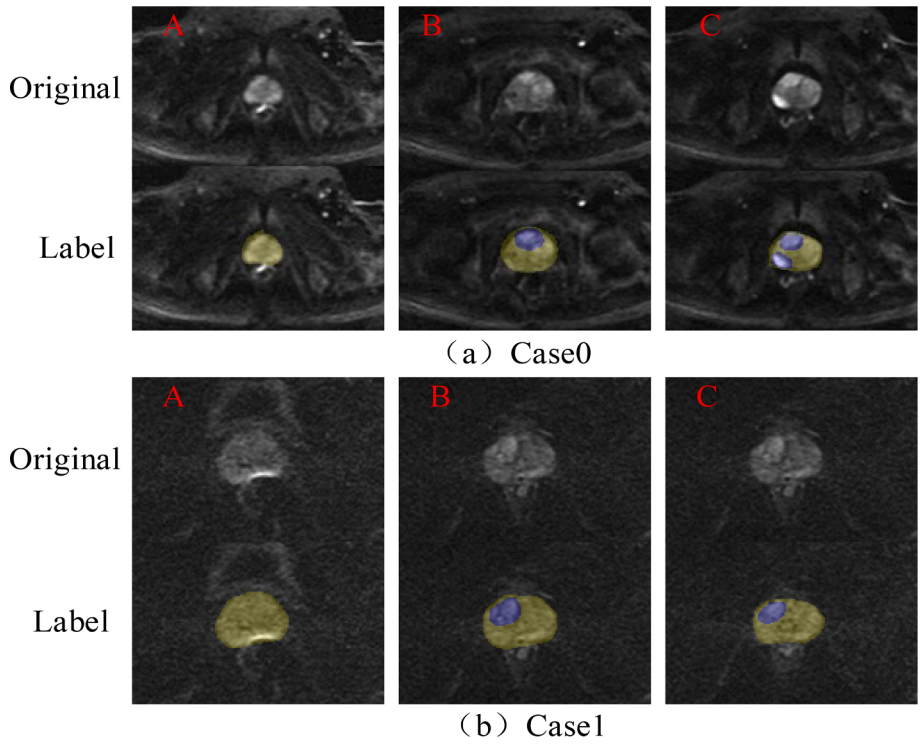

Fig. 1. Basic morphology of prostate and tumor in DWI 2D images. (a) and (b) show the original and annotated images of the sequence images of the two cases, respectively. The yellow transparent area in the annotation image is the prostate, and the blue transparent area is the tumor.

proposed in the decoder. We also built a hybrid loss function to deal with class imbalance. Dice similarity coefficient (DSC) values for prostate and tumor DWI segmentation were 92.28% and 88.73%, respectively. The contributions of this study are as follows:

(1) We offer a unique dual-attention mechanism 3D segmentation network architecture for quantitative assessment of prostate and tumor volumes on DWI.
(2) A visual attention method is built into the encoder step to obtain the features of various receptive fields and deliver more detailed contextual information.
(3) A multiscale attention technique is proposed in the decoder stage to fuse multiscale features and acquire finer global and local detailed information.
(4) To address the problem of class differences in the prostate, tumor, and background regions in segmentation tasks, we constructed a hybrid loss function for handling class imbalance.
(5) We tested the algorithm on PCa DWI images collected from local hospitals to ensure that it was both innovative and effective.

2. Materials and methodology

2.1. Datasets

From 2013 to 2016, we collected data from 98 patients with PCa at the Haikou People's Hospital of Central South University and the affiliated hospital of Xiangya Medical College. Doctors used multiparametric MRI (mp-MRI) in all of the patients and found malignancy. All tests were performed on a 3 T scanner using a 32-channel phased-array coil (Achieva 3 T; Philips Healthcare, Eindhoven, Netherlands). During this time, prostate biopsies were performed and PCa was discovered. Pathologists certified by the hospital board of directors made the pathologigal diagnoses using the Gleason grading system. Our data for 98 patients with an initial diagnosis of PCa DWI corresponding to an image voxel size of 256 × 256 × 22. The patient scanned a 400 mm × 400 mm field of vision (FOV) with a thickness of 4 mm. It should be noted that the dataset in question passed the relevant hospital's ethical assessment and was obtained with the informed consent of the patient.

2.2. Manual annotation

MR images of the prostate were analyzed by two radiologists with 10 to 20 years of experience in prostate MRI studies. The two experts had a face-to-face meetings and participated in training and internships. In two separate sessions, the two experts annotated all data. Table 2 summarizes the quantitative outcomes of inter-observer and intra-observer agreement evaluation for this dataset, where $A_i(i = 1,2)$ represents expert A's segmentation result for the *i-th* session, and $B_i(i = 1,2)$ represents expert B's segmentation result for the *i-th* session. The union and intersection of both sessions for each expert were used to calculate the inter-observer differences. $A_{1\&2}$ and $B_{1\&2}$ indicate the union of experts A's and B's segmentation results, respectively. The findings of the intra- and inter-observer comparisons indicate a very high correlation coefficient (CC), showing a very high linear correlation between different observers and the same observer. Because the mean overlap rate (overlap greater than 90%) and mean absolute area difference (ADD less than 5%) demonstrated good inter-observer and intra-observer agreement, we considered the union of the two experts' segmentation results as the ground truth.

Table 2
Results of intra-observer and inter-observer quantitative assessments.

	CC (mean)	ADD (%) [mean ± std]	Overlap (%) [mean ± std]
ExpertA_1-Expert A_2	0.997	3.15 ± 1.91	94.23 ± 4.27
ExpertB_1-Expert B_2	0.998	3.85 ± 2.46	95.44 ± 5.62
Expert$A_{1\&2}$-Expert$B_{1\&2}$	0.996	4.26 ± 3.15	92.37 ± 4.76

2.3. Data preprocessing

Our study included 98 patients who had been diagnosed with PCa and had a DWI image voxel size of 256 × 256 × 22. The patient scanned a 400 mm × 400 mm field of vision (FOV) with a thickness of 4 mm. The tumor region in the full-scale image is modest owing to the vast scanning field of view. Consequently, in each image, we counted the extent of the prostate and cut the image to include the prostate tissue structure and surrounding areas. As the model entered another branch of the created network, the resampled picture size was 192 × 192 × 22, and size selection was achieved by tuning trials. We also performed sample augmentation on all training data, flipping each prostate DWI image from left to right and top to bottom, and rotating the images by 90, 180, and 270 degrees to preserve the visual structure. The sample size can train the deep learning model and avoid overfitting after the data augmentation.

2.4. Method overview

Fig. 2 shows the structure of the dual-attention 3D segmentation network for prostate and cancerous tissues of our proposed method, which consists of four synergistic parts: encoder, visual attention, decoder, and multi-scale attention. We investigated the task of segmenting prostate and tumor regions using DWI scans. The model input was the PCa DWI volume. Let us consider the volume V = {V_1, V_2,.., V_N} containing PCa DWI $V_i \in \mathbb{R}^{d_m \times d_n \times d_p}$, i ∈ {1, 2, .., N} d_m, d_n and d_p are the dimensions of PCa DWI. Spatial feature representations were extracted for each of the PCa DWI of volume V. As shown in Fig. 2, we input the 3D DWI PCa image with an input block size of 64 × 64 × 64. This patch size was chosen to enable the learning of multiscale features while limiting memory requirements. For each 3D input patch, the network outputs a probability map of the same size in an end-to-end manner. In the segmentation stage, we selected an end-to-end trainable strong-structure 3D U-net [37] as the underlying backbone network. The 3D U-net contains an encoder and decoder, and has skip connections at each resolution level. Multiscale features incorporate rich semantic and detailed information. Therefore, prostate and tumor areas of different sizes can be treated effectively. The input to the encoder stage was first processed by two 3 × 3 × 3 convolutional layers, followed by batch normalization (BN) and parametric rectified linear unit (PReLU) [38] activation. The convolution process is expressed in formula (1). Formally, the value of a unit at pixel (x,y,z) in the *j*th feature map in the *i*th layer, denoted as$Pixel_{ij}^{xyz}$, is given by:

$$Pixel_{ij}^{xyz} = f\left(b_{ij} + \sum_{m}\sum_{p=0}^{p_i-1}\sum_{q=0}^{q_i-1}\sum_{r=0}^{r_i-1} w_{ijm}^{pqr} Pixel_{(i-1)m}^{(x+p)(y+q)(z+r)}\right) \quad (1)$$

where *f* is the PReLU, b_{ij} is the bias for this feature map, *m* refers to *m* indexes on the feature map set connected to the $(i-1)$th layer of the current feature map, w_{ijm}^{pqr} is the value at pixel (p, q, r) of the kernel connected to the *m*th feature map, and p_i, q_i and r_i are the height, width, and length of the kernel, respectively.

The PReLU function is demonstrated in formula (2) as follow:

$$f(y_k) = \begin{cases} y_k, & if \quad y_k > 0 \\ \alpha_k y_k, & if \quad y_k \leqslant 0 \end{cases} \quad (2)$$

where k represents different channels; that is, each channel has a PReLU function with different parameters. The maximum pooling layer was obtained after the convolution of different scales. We set downsampling of *Scale* 1, *Scale* 2, *Scale* 3 and *Scale* 4 by stride (1,2,2).

We designed a visual attention mechanism (VAM) in the encoder stage to extract features of different scales from *Scale* 1 to *Scale* 5 and applied them to the VAM module to obtain feature information with different receptive fields. Each feature map computed through the backbone network was then upsampled to the size of the input patch using trilinear interpolation. In the decoder stage, each stage inputs the features of the encoder and the VAM output from the same scale, as well as the features after the deconvolution of the previous layer and the characteristics of the VAM output of the previous stage. After the decoder stage, a scale attention mechanism (SAM) is designed, the features of scales 1–4 are fused in a combined form, and the SAM module is input to obtain finer multi-scale detail information. Then, perform 1 × 1 × 1 convolution on each obtained feature map. The purpose of the 1 × 1 × 1 convolution was to reduce the number of channels at the corresponding level. The output of the final network model was a 3D mask of the prostate and cancerous regions of the same size as the input. In the following subsections, we explain in detail the details of our model and its dual attention mechanism learning, which enables learning to segment the complex and variable structures of the prostate and tumor on DWI images with complex backgrounds.

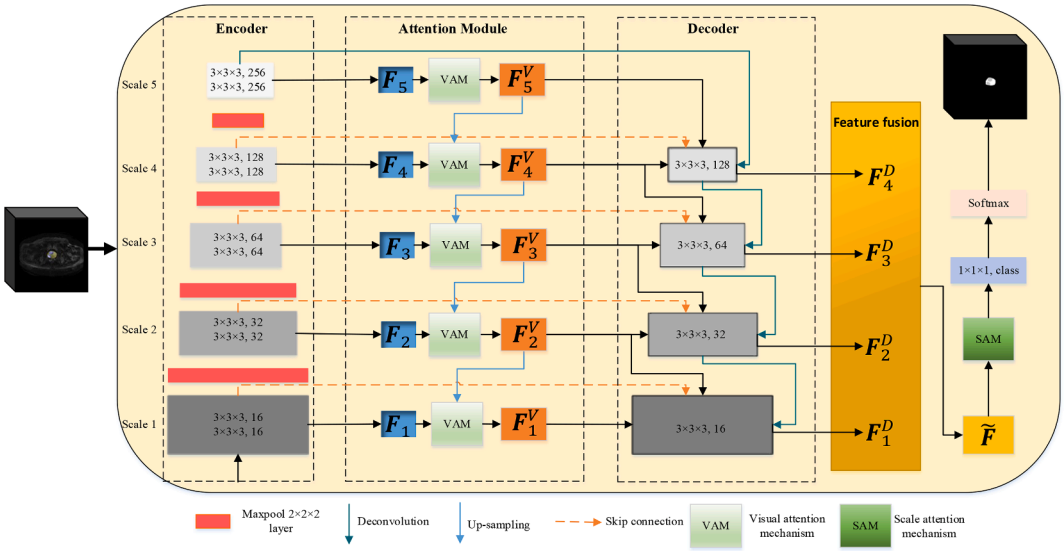

Fig. 2. Framework diagram of the proposed dual-attention mechanism 3D segmentation network. The network structure uses a powerful end-to-end and universal 3D U-net as the basic backbone network. A visual attention mechanism is designed at each resolution level in the encoder, and a scale attention mechanism is designed in the after decoder stage.

2.5. Visual attention mechanism

The visual selective attention of the human eye consists of two parts: blurry achromatic peripheral vision and high-acuity chromatic central vision [39]. This enbles humans to rapidly shift their foveal gaze to salient regions. The entire process includes parallel processing of various input features by the retina and the gradual fusion of different features with the participation of the visual attention mechanism. In general, the human visual perception process is a combination of two visual attention mechanisms and feature integration theory [40]. Fig. 3 shows the VAM structure of the 3D segmentation network for the prostate and tumor regions using the proposed method. The model inputs the feature maps F_i and F_{i+1}^V (i = 1, 2, 3, 4) of scale i to fuse the features. When i = 5, feature F_5 is the direct input, and the output is an attention feature F_i^V of the same size as F_i. We describe the perception process of the VAM layer under different scale features as follows:

$$F_i^V = \begin{cases} f_i(F_i), & i = 5 \\ f_i\left(\dfrac{\lambda_1 \cdot F_i + \lambda_2 \cdot Upsample\left(F_{i+1}^V\right)}{\lambda_1 + \lambda_2}\right), & i \neq 5 \end{cases} \tag{3}$$

where $\lambda_1 \geqslant 0$ and $\lambda_2 \geqslant 0$ represent the weights obtained by applying a rectified linear unit, *Upsample* represents the up-sampling operation.

In the designed VAM module, the convolutional layers of different scales represent the center of vision, and the dilated convolutional layers of different rates increase the visual receptive field. Inspired by [41], the VAM module aims to obtain global and local information through a combination of convolution and dilated convolution. The VAM contains two layers of group convolution blocks: the first layer is a standard convolution group of different sizes, and the second layer is a dilated convolution group of different rates. In our VAM, the input features were first processed using a group convolution block. The kernel sizes were $1 \times 1 \times 1$, $3 \times 3 \times 3$, $5 \times 5 \times 5$, and $7 \times 7 \times 7$, and each convolution operation was followed by BM and PReLU. This design enriches the learned representation and reduces the computational complexity [42]. The generated feature map was connected and passed through a second set of convolution blocks. The second group of convolution blocks is the expansion convolution blocks used to expand the receptive field, and the expansion rates are 1, 3, 5, and 7. The generated feature map is fused, and the feature map obtained by the sigmoid function activation method is multiplied and added to the input and intermediate features to obtain the final attention feature F_i^V.

2.6. Scale attention mechanism

Fig. 4 shows the SAM structure in the 3D network for the segmentation of prostate and cancerous tissues using the proposed method. To better utilize the refined multiscale features of feature consistency produced by feature learning, we propose a SAM feature-learning module. The module consisted of two parts. The first part inputs the features F_1^D, F_2^D, F_3^D and F_4^D generated by the decoder at different stages. Subsequently, the adjacent scale features are gradually aggregated instead of multi-branch prediction of the refined features [43 44], which has the advantage of obtaining a more accurate segmentation feature map. Specifically, the features between two adjacent scales are spliced from top to bottom, and the module transforms the high-level semantic information of the small-scale features into large-scale features. The second part of the input is the fused feature map **F** of all the decoder-generated features. First, we resampled the feature maps of different scales acquired by the encoder to the original input image size using trilinear interpolation. To reduce the computational cost, these feature maps were compressed into four channels using a $1 \times 1 \times 1$ convolution, and the compression results of different scales were spliced into a hybrid feature map **F**. Global average pooling ($Pool_{avg}$) and global max pooling ($Pool_{max}$) are then used to obtain global information. The results of the two poolings were added and input to the sigmoid function to obtain the scale coefficient attention vector. This vector is then multiplied by the input feature **F**, and the resulting feature performs a $3 \times 3 \times 3$ convolution and a $1 \times 1 \times 1$ convolution, and then enters the sigmoid function. The resulting feature map is multiplied with the features after $3 \times 3 \times 3$ convolution and added to the scale coefficient attention vector and the original feature **F** to obtain the output features of the second part. Finally, the output feature maps of the two parts were added to obtain the final attention feature map$\widetilde{F}$.

2.7. Loss function

To solve the class imbalance problem of the prostate, malignant, and background regions in the segmentation task, we built a hybrid loss function for the model, which comprises a weighted sum of the two functions in the proposed dual-attention-guided 3D segmentation learning network. Dice loss is the first loss function explicitly aimed at optimizing the segmentation performance assessment measure. Dice Loss was first proposed in the article VNet [45], and is widely used in medical image segmentation. Dice loss performs well in a scenario

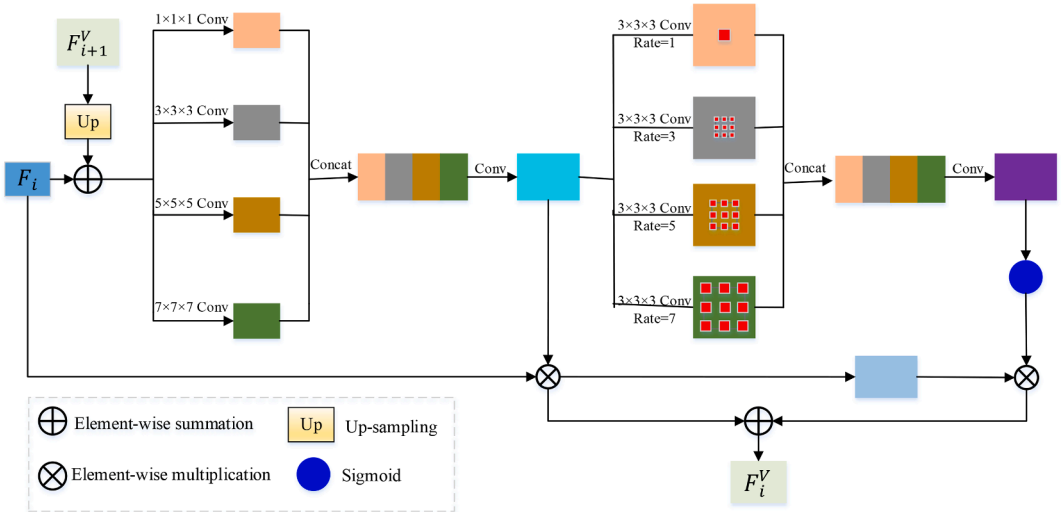

Fig. 3. The structure diagram of the proposed VAM. The model inputs the feature maps F_i and F_{i+1}^V (i = 1, 2, 3, 4) of Scale i to fuse the feature. When i = 5, the feature F_5 is directly input, and the output is an attention feature $F_i{}^V$ of the same size as F_i. VAM contains two layers of group convolution blocks, the first layer is a standard convolution group of different sizes, and the second layer is a dilated convolution group of different rates.

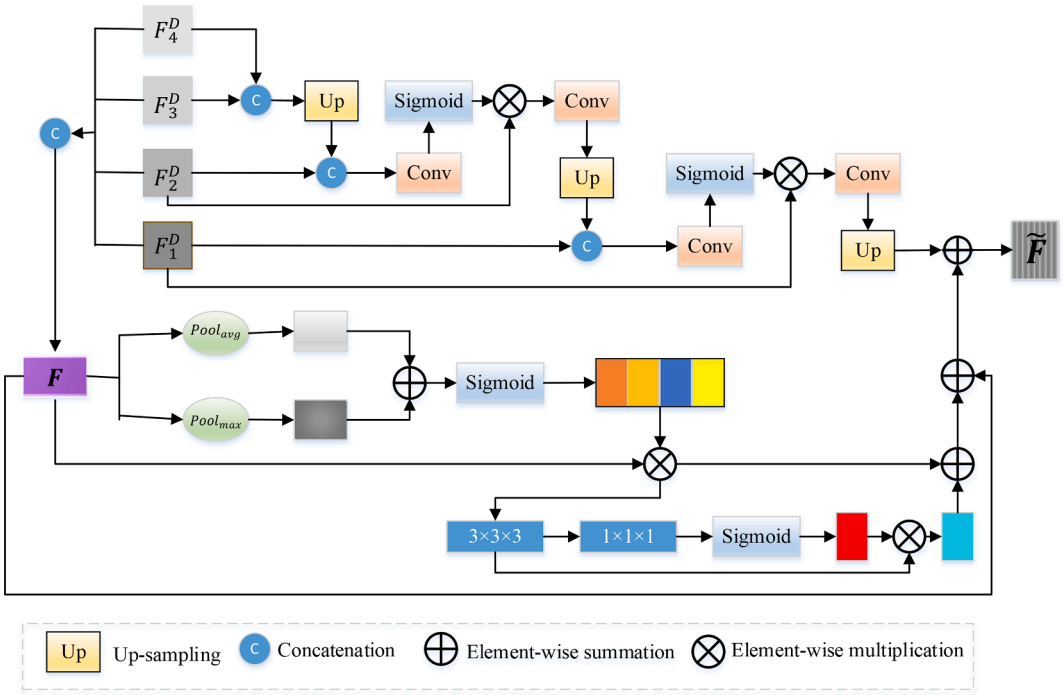

Fig. 4. The structure diagram of the proposed SAM. The module consists of two parts. The first part inputs the features F_1^D, F_2^D, F_3^D and F_4^D generated by the decoder in different stages. The second part of the input is the fused feature map **F** of all decoder-generated features. Finally, the output feature maps of the two parts are added to obtain the final attention feature map $\widetilde{F}$.

where the positive and negative samples are seriously unbalanced. In the training process, more emphasis is placed on the mining of foreground areas. The dice loss is defined as follows:

$$L_{DCE(X)} = \frac{2\sum_{x_i \in X} p(x_i) g(x_i) + \varepsilon}{\sum_{x_i \in X} p(x_i) + \sum_{x_i \in X} g(x_i) + \varepsilon} \tag{4}$$

where $p(x_i)$ represents the predicted probability of voxel x_i, and $g(x_i)$ indicates the corresponding ground truth on the same voxel. χ signifies the training image and ε denotes a small term that prevents the loss function from being divided by zero. The focal loss, which is improved by log loss to overcome the problem of sample imbalance, is the second loss function. Focal Loss is specifically defined for single-level target detection [46]. It can also effectively perform image classification tasks. Focal Loss reduces the weight of samples that are easy to classify but increases the weight of samples that are difficult to classify.

$$L_{Focal(X)} = -\frac{1}{n}\sum_{i-1}^{n}\left[\begin{array}{c}\varphi g(x_i)(1-p(x_i))^{\gamma}\log p(x_i) + \\ (1-\varphi)(1-g(x_i))p(x_i)^{\gamma}\log(1-p(x_i))\end{array}\right] \tag{5}$$

where φ represents the balance factor of the focal loss, which is set to 0.2. γ is the focus parameter of the smooth adjustment weight rate, which was set to 1.

Dice loss was used to optimize the evaluation metric of segmentation performance, and focal loss was used to guide the model segmentation of small target areas. Both can address the class-imbalance problem. Therefore, the loss function is expressed as follows:

$$L_{seg} = n_1 \cdot L_{DCE(X)} + n_2 \cdot L_{Focal(X)} \tag{6}$$

where, n_1 and n_2 represent the Dice loss, focal loss, and weight factor, which are set to 0.9 and 0.1, respectively.

3. Experiments and results

3.1. Implementation details

The experiment was run on a system with an Intel Xeon CPU, NVIDIA Tesla V100 PCIe GPU with 11 GB of memory, and 16 GB of RAM. The suggested model was trained using a 5-fold cross-validation technique on Python 3.7 and PyTorch platforms. The 3D Unet [37] is the backbone of the network structure. With an initial learning rate of 0.0001 and a weight decay of 0.00002, the Adam optimizer was used to stochastically optimize the proposed model and other models. In addition, the batch size was set to 8 and the number of training epochs was set to 300.

3.2. Evaluation metrics

To objectively analyze prostate and tumor segmentation results and model performance, we employed five performance evaluation metrics: correlation coefficient (CC), overlap rate (overlap), Hausdorff distance (HD), Dice similarity coefficient (DSC), and precision (ACC). CC is a statistical metric designed by the statistician Karl Pearson [47]. Correlation is an uncertain relationship, and CC is the quantity of linear correlation between research variables. The following is a list of definitions.

$$CC = \frac{\sum_{i=1}^{n}(A_i - \overline{A})(B_i - \overline{B})}{\sqrt{\sum_{i=1}^{n}(A_i - \overline{A})^2}\sqrt{\sum_{i=1}^{n}(B_i - \overline{B})^2}} \tag{7}$$

$$Overlap = \frac{\sum_{i=1}^{n} A_i \cap B_i}{\sum_{i=1}^{n} A_i \cup B_i} \tag{8}$$

$$HD = \frac{1}{n}\sum_{i=1}^{n}\max(h(A_i, B_i), h(B_i, A_i)) \tag{9}$$

$$h(A_i, B_i) = \max_{a_i \in A_i}\{\min_{b_i \in B_i}||a_i - b_i||\} \tag{10}$$

$$h(B_i, A_i) = \max_{b_i \in B_i}\{\min_{a_i \in A_i}|||b_i - a_i||\} \tag{11}$$

where A_i and B_i represent the ground truth and model output of the prostate or tumor segmentation region of the *i-th* scan DWI slice, respectively.

$$DSC = \frac{2TF}{FF + FN + 2TF} \tag{12}$$

$$ACC = \frac{TF + TN}{FF + FN + TF + TN} \tag{13}$$

where TP, TN, FP, and FN represent the true positive, true negative, false positive and false negative, respectively.

3.3. Results

Fig. 5 shows the consistency of our method's prostate and PCa volume estimation predictions and manual segmentation. Our proposed method's prostate prediction is significantly connected with expert manual segmentation of DWI image, according to correlation curve analysis. The volume correlation coefficient of CC = 0.9734 for the test set. The volume correlation coefficient of CC = 0.9458 shows that the automatic segmentation of PCa is consistent with expert manual segmentation. Furthermore, the p-values for prostate and PCa volumes were less than 0.05, indicating that our automatic and manual segmentation methods for prostate and PCa volumes were statistically significant.

Fig. 6 shows the outcomes of our method's segmentation, as well as the ground-truth prostate and tumor in a DWI image. The original image is in the first column, the ground truth is in the second column, our result mask is in the third column (the white area represents the prostate and the gray area indicates the tumor), the fourth column shows the comparison between our results and the ground truth, and the fifth column is the enlarged view of the prostate area in the fourth column in the yellow box. The red line indicates the ground truth and the green line represents the results of our method. The prostate and tumor regions of the six patients produced bright signals when compared to other tissues; however, the contours were fuzzy, and the forms were considerably different, as shown in Fig. 6. Owing to the development of PCa cells, the size, shape, and intensity distribution of the prostate also change. The tumor margin is difficult to discern because its shape is irregular, its placement is not fixed, the area is discontinuous, and its intensity varies significantly. Nevertheless, the automatic segmentation results of our proposed method are highly compatible with the ground truth of DWI images, which are notoriously difficult to segment. Our segmentation findings (green lines) have a high degree of overlap with the experts' manual segmentation results (red lines) in Fig. 6, column 5, with consistent border contours.

3.4. Ablation study

3.4.1. Ablation study of parameter settings

In the network structure we designed, there are four very important hyperparameters in the loss function; ϕ, γ, η_1 and η_2. To investigate the effect of different parameter settings on the network performance, we trained the network with different parameter values using the control variable method. As shown in Table 3, we first controlled (γ, η_1, η_2) to be (0.5, 0.5, 0.5) and explored the effect of increasing or decreasing the weight of φ on the network performance, so set φ to 0.1, 0.2, and 0.3. According to the evaluation indicators DSC and HD, the optimal parameter ratio is determined, and then the control (φ, η_1, η_2) is (0.2, 0.5, 05) to explore the influence of increasing or decreasing the weight of γ on the network performance; thus γ is set to 0.5, 1, and 2. The optimal parameter ratio is determined according to the evaluation metrics DSC and HD, then control (γ,φ) to the optimal parameter ratio and keep it unchanged, adjust the parameter ratio of (η_1, η_2) to (0.1, 0.9) and (0.9, 0.1) to increase or reduce the weight of DCE loss or focal loss, and then judge the network performance. Table 3 shows that the proposed method achieved the lowest HD and highest DSC for prostate and tumor segmentation when ϕ was set to 0.2, γ was set to 1, and (η_1, η_2) was set to (0.9, 0.1), respectively. The results of prostate segmentation showed an average CC of 0.9734, overlap of 90.82%, HD of 0.2854 mm, DSC of 92.28%, and ACC of 99.16%. The results of tumor segmentation showed an average CC of 0.9458, overlap of 86.64%, HD of 0.5746 mm, DSC of 88.73%, and ACC of 96.77%.

3.4.2. Ablation study of network structure

To demonstrate the accuracy of the network structure we designed for prostate and tumor region segmentation, we added VAM and SAM based on the input baseline (3D U-net) and aggregated different modules on the baseline. The detailed results are presented in Table 4. The addition of the proposed VAM (B + VAM) and SAM (B + SAM) to the baseline significantly improved the five evaluation metrics. Compared with the Baseline, the DSC of the prostate segmentation results in the VAM module increased by 5.62% (89.38% vs. 83.76%), and the DSC of the tumor segmentation results increased by 9.91% (86.25% vs. 76.34%). Therefore, the detailed information provided by the combination of standard convolution and dilated convolution of the VAM module plays a key role in prostate and tumor segmentation, particularly for small target regions (tumors). In addition, on the basis of the baseline, we added VAM (B + VAM_w/o_Dc) without dilated convolution, SAM with only the gradual feature fusion part (B + SAM_w/o_P2), and the module with only the feature splicing part (B + SAM_w/o_P1) for comparison with the segmentation performance of VAM and SAM designed by us.

From Table 4, the module we designed achieved a better segmentation performance for the prostate and tumor regions. The DSC values for prostate and tumor DWI segmentation were 92.28% and 88.73%, respectively. In our approach, two modules (VAM and SAM) work together to complete the segmentation task. The proposed visual attention method can obtain the features of multiple receptive fields and provide more detailed contextual information. The features of different

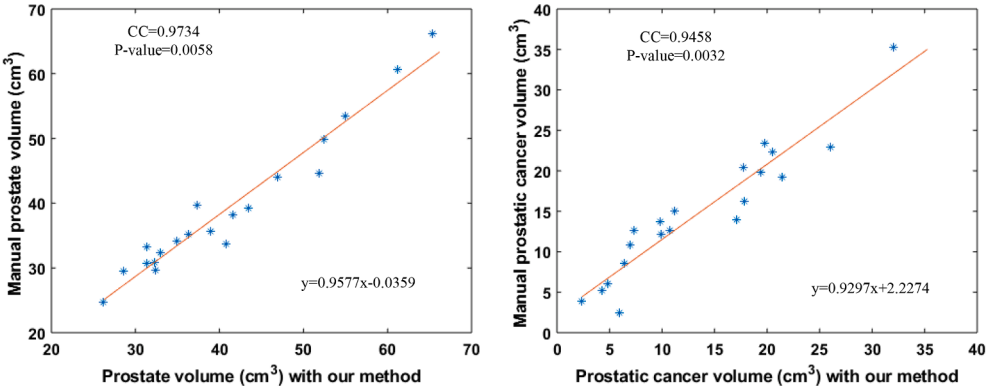

Fig. 5. Agreement of prostate and tumor volume assessments between our method predicts and manual segmentation. (a) Average prostate volume of DWI images on testing set (b) Average tumor volume of DWI images on testing set.

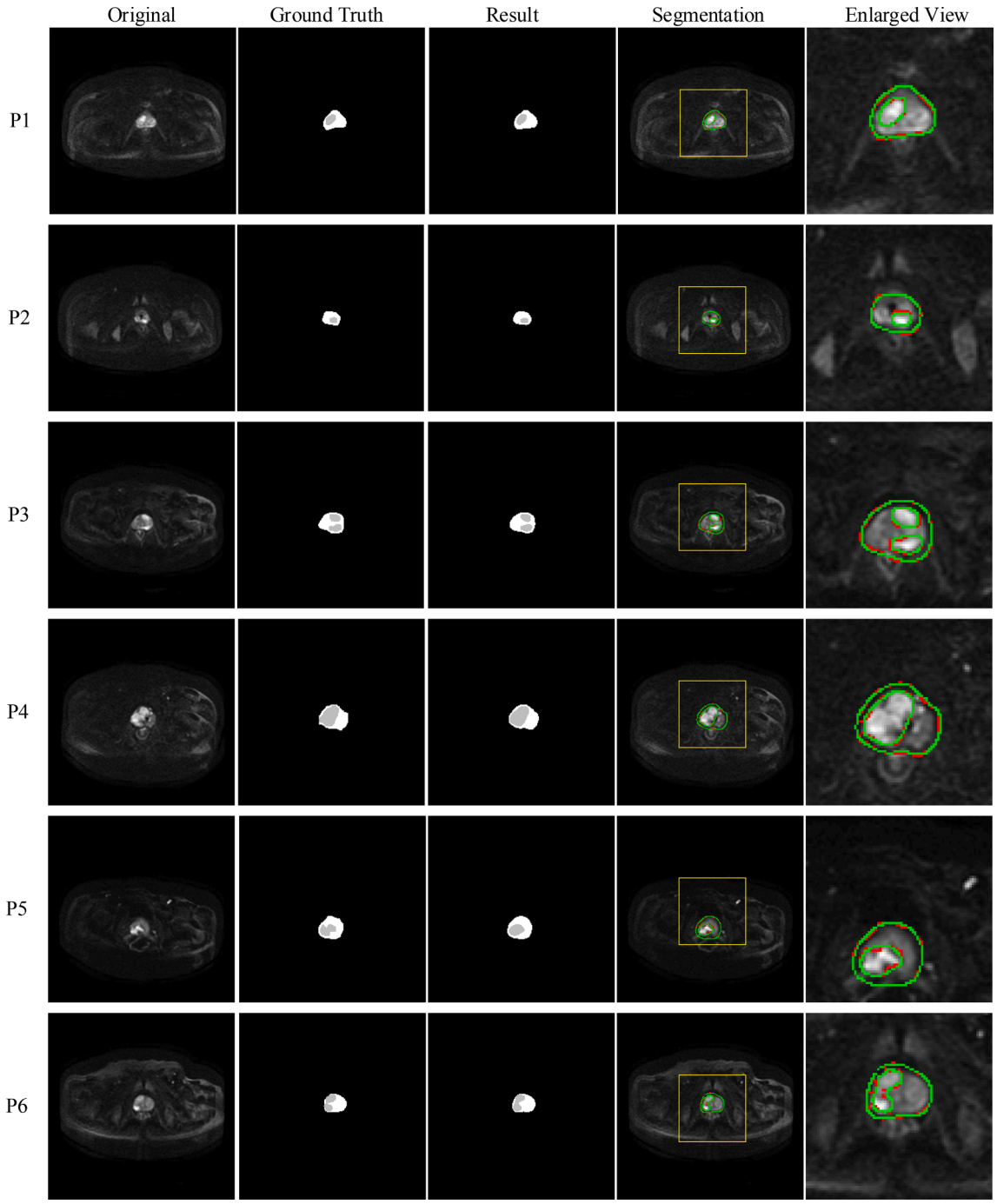

Fig. 6. Segmentation results of our method and the ground truth prostate and tumor in MR image. P1 ~ P6 represent 6 different patients respectively. The original image is in the first column, the ground truth is in the second column, our result mask is in the third column (the white area represents the prostate and the gray area indicates tumor), the fourth column is the comparison between our results and ground truth, and the fifth column is the enlarged view of the prostate area in the fourth column in the yellow box. The red line indicates the ground truth, and the green line represents the result of our method.

receptive fields and rich context information render the classification results more precise. Our proposed SAM feature learning module integrates multiscale features to obtain more detailed global and local details, which can make better use of the feature consistency generated by feature learning to refine multiscale features. From Table 4, the segmentation performance of our method is further improved compared to the addition of only one module. Compared with the baseline, our method improved the DSC value of prostate DWI segmentation by 8.52% (92.28% vs 83.76%), and the DSC value of tumor DWI segmentation by 12.39% (88.73% vs 76.34%). Therefore, both the VAM and SAM proposed in our method play an important role in target region segmentation, and neither method is indispensable. The fusion of the two attention modules makes segmentation of the prostate and tumor more accurate. The improved rate of tumor segmentation performance over prostate segmentation suggests that our method is beneficial for the segmentation of regions with irregular shapes, blurred boundaries, nonfixed locations, and small objects. The correlation coefficients between the results of our proposed method and the ground truth are both greater than 0.9 (0.9734 and 0.9458), indicating that the automatic segmentation results of the prostate and tumor and the expert manual segmentation results are highly correlated and consistent. The decrease in the average HD value also supported the superior segmentation

Table 3
Influence of different parameter settings on network performance (HD and DSC as evaluation metrics).

Parameters				Prostate		Tumor	
φ	γ	η_1	η_2	HD[mm]	DSC[%]	HD[mm]	DSC[%]
0.1	0.5	0.5	0.5	0.6025	87.38	0.8312	83.55
0.2	0.5	0.5	0.5	0.5715	88.66	0.7827	84.43
0.3	0.5	0.5	0.5	0.6218	86.53	0.9521	82.16
0.2	1	0.5	0.5	0.3218	90.85	0.6247	87.61
0.2	2	0.5	0.5	0.3671	90.14	0.6915	86.82
0.2	1	0.1	0.9	0.4627	89.65	0.7218	85.27
0.2	**1**	**0.9**	**0.1**	**0.2854**	**92.28**	**0.5746**	**88.73**

Table 4
Ablation analysis of different components in our network (baseline is 3D U-net).

Method	Prostate Segmentation				
	CC	Overlap[%]	HD[mm]	DSC[%]	ACC[%]
Baseline	0.9285	81.23	0.8761	83.76	92.65
B + VAM	0.9673	87.24	0.4438	89.38	96.19
B + SAM	0.9585	86.04	0.5165	87.92	95.78
B + VAM_w/o_Dc	0.9572	85.86	0.5826	87.14	95.84
B + SAM_w/o_P2	0.9519	85.58	0.6163	86.64	95.63
B + SAM_w/o_P1	0.9438	85.16	0.7084	86.05	95.18
Ours	**0.9734**	**90.85**	**0.2854**	**92.28**	**99.16**

	Tumor Segmentation				
	CC	Overlap[%]	HD[mm]	DSC[%]	ACC[%]
Baseline	0.8652	74.86	1.2736	76.34	85.67
B + VAM	0.9377	84.96	0.6353	86.25	94.61
B + SAM	0.9283	83.11	0.6732	84.56	93.94
B + VAM_w/o_Dc	0.9136	79.45	0.9544	81.35	90.28
B + SAM_w/o_P2	0.9018	78.83	1.0685	80.89	89.44
B + SAM_w/o_P1	0.9035	79.22	0.9761	82.55	90.74
Ours	**0.9458**	**86.64**	**0.5746**	**88.73**	**96.77**

performance of the proposed method.

3.5. Comparative experiment analysis

To verify the effectiveness of our proposed method, we divided the comparative methods into three categories.

(1) **Classical 3D Image Segmentation Networks**: 3D U-Net [37] and V-Net [45]. The 3D U-Net [37] has a symmetric encoder-decoder structure, and the feature map output by the final decoder incorporates more low-level features. V-Net [45] is an encoder-decoder 3D U-Net morphing network for prostate segmentation.
(2) **Attention Mechanism Segmentation Network**: Channel Attention Module [48] and Dual Attention Module [49]. The SE block proposed by Fu et al. [48] focuses on obtaining the channel relationship of feature information to improve the representational ability of the network. The dual attention network [49] is based on a channel attention mechanism that increases spatial relational attention to obtain feature information.
(3) **Prostate Segmentation Networks**: BiAA-Net [50] and ConvLSTMs [16]. BiAA-Net [50] developed a dual-attention adversarial network for PCa T2W image segmentation. ConvLSTMs [16] are an automatic and interactive prostate segmentation method based on a convolutional long short-term memory (convLSTM) module and a gated graph neural network (GGNN).

3.5.1. Quantitative analysis

On DWI images of the prostate and tumor regions, Table 5 demonstrates the segmentation performance scores of our method and the contrasting methods. For consistency, all approaches employ the same training and test data as well as the same evaluation metrics to produce segmentation performance scores. To acquire the best segmentation results using this method, our network runs parameter ablation studies, whereas other comparison methods also run parameter ablation studies. The results of prostate segmentation showed an average CC of 0.9734, overlap of 90.82%, HD of 0.2854 mm, DSC of 92.28%, and ACC of 99.16%. Among the comparative methods of classical 3D image segmentation networks, attention mechanism segmentation networks, and state-of-the-art prostate segmentation networks, our method yielded the best metrics. Compared with 3D U-Net, the most classic image segmentation network, our method improves the CC, Overlap, HD, DSC and ACC of prostate segmentation by approximately 0.045, 9.62 %, 0.59 mm, 8.52 %, and 6.51 %, respectively. The results of the tumor segmentation showed that the average CC, overlap, HD, DSC, and ACC were 0.9458, 86.64 %, 0.5746 mm, 88.73%, and 96.77 %, respectively. CC, Overlap, HD, DSC and ACC in tumor segmentation were improved by approximately 0.08, 11.78%, 0.70 mm, 12.39%, and 11.1%, respectively. Compared with the state-of-the-art prostate segmentation method ConvLSTMs, our method achieves approximately 0.04, 4.08%, 0.15 mm, 4.06%, and 2.78% improvement in prostate segmentation in CC, Overlap, HD, DSC and ACC, respectively. CC, Overlap, HD, DSC and ACC in tumor segmentation were improved by approximately 0.03, 5.81%, 0.25 mm, 6.18%, and 7.5%, respectively. This shows that our method outperforms the other methods in terms of the segmentation performance. The improved performance is mainly attributed to our proposed dual-attention segmentation structure, which provides more comprehensive visual information and multiscale feature information, thereby enhancing the performance of network learning. Compared with other attention methods, the VAM module we designed includes standard convolution and dilated convolution of different receptive fields. The combination of the two convolutions yields abundant global and local information. The features of different receptive fields and rich context information render the classification results more precise. Our proposed SAM feature learning module integrates multi-scale features to obtain more detailed global and local details, which can make better use of the feature consistency generated by feature learning to refine multi-scale features. In addition, to address the problem of class differences in the prostate, tumor, and background regions in segmentation tasks, we constructed a hybrid loss function for

Table 5
Performance comparison of our method with classical 3D image segmentation network (U-Net and U-Net++), attention network (SE-Net and DeepLab V3+) and prostate or tumor segmentation network (BiAA-Net and ConvLSTMs) on five evaluation metrics (CC, Overlap, HD, DSC and ACC).

Method	Prostate Segmentation				
	CC	Overlap[%]	HD[mm]	DSC[%]	ACC[%l]
3D U-Net [29]	0.9285	81.23	0.8761	83.76	92.65
V-Net [36]	0.9346	84.73	0.4058	86.17	95.31
SE-Net [37]	0.9285	82.04	0.7884	83.78	93.28
Dual attention [38]	0.9217	82.89	0.7028	84.61	94.87
BiAA-Net [39]	0.9248	85.85	0.4664	87.71	95.48
ConvLSTMs [11]	0.9367	86.77	0.4374	88.22	96.38
Ours	**0.9734**	**90.85**	**0.2854**	**92.28**	**99.16**

	Tumor Segmentation				
	CC	Overlap[%]	HD[mm]	DSC[%]	ACC[%l]
3D U-Net [29]	0.8652	74.86	1.2736	76.34	85.67
V-Net [36]	0.9166	83.05	0.6844	84.68	93.14
SE-Net [37]	0.8733	76.92	0.9525	78.63	87.86
Dual attention [38]	0.9038	79.47	0.8623	81.04	88.74
BiAA-Net [39]	0.9315	83.79	0.5734	85.12	93.81
ConvLSTMs [11]	0.9144	80.83	0.8316	82.55	89.27
Ours	**0.9458**	**86.64**	**0.5746**	**88.73**	**96.77**

handling class imbalance. This is reflected in Table 4, and an analysis of the ablation experiment was performed. Fig. 7 presents the distribution of DSC of prostate regions from 3D U-Net, V-Net, SE-Net, Dual attention, BiAA-Net, ConvLSTMs and our method in the test set. Fig. 7 shows that the proposed technique has a higher median and greater stability in the DSC value distribution. It outperformed other approaches in the treatment of tumors segmentation.

3.5.2. Qualitative analysis

Fig. 8 shows the qualitative comparison results of different methods and ground truth for DWI image segmentation of the prostate and tumor regions. Each row depicts a two-dimensional slice image of several examples, with red and green indicating the prostate and PCa areas, respectively. Each column represents the outcome of several segmentation method. The proposed method clearly produces segmentation results that are more compatible with the ground truth than the other six methods. The findings of segmenting the visible prostate and tumor confirmed the efficacy of our procedure. Fig. 8 shows how the prostate segmentation results of the seven patients ($P_A \sim P_G$) can be identified using different methods. However, in terms of boundary segmentation, the other six approaches diverge from the ground truth. While our proposed method can accurately identify the prostate region, the boundary and ground truth also maintain a high degree of coincidence, which is attributed to the effective capture of feature details and contextual information by our designed visual attention block and scale attention block. However, for tumor segmentation of small target regions, U-Net, V-Net, SE-Net and Dual attention methods have cases in which the cancerous region (P_B, P_D and P_F) is not recognized or the cancerous region is incorrectly identified (P_A, P_C and P_D). This is a very fatal error in clinical decision-making, which seriously affects doctors' judgment of PCa. Our proposed method can effectively identify PCa regions, whether as a single cancerous region or multiple cancerous regions. Although BiAA-Net and ConvLSTMs can effectively identify tumor locations, there is still a significant gap between essential data, such as the size and perimeter of the cancerous zone. and the ground truth. On the other hand, our proposed method can precisely segment the boundaries of cancerous zone while effectively recognizing the prostate region, while maintaining high consistency with the ground truth. This demonstrates that our dual-attention structure is capable of extracting visual and multi-scale features in prostate and tumor DWI image segmentation, as well as extracting comprehensive information and deeper learning representations.

4. Discussion

In this study, we propose a dual-attention 3D network to segregate the prostate and tumor in DWI images and offer tumor volume and localization information for subsequent prostatic cancer diagnosis and treatment. Experiments on a hospital-collected prostatic cancer DWI dataset revealed that our technique had high DSC values. The segmentation findings of our method are very similar to the labels manually annotated by professionals. With a strong correlation with manually segmented volumes, our algorithm predicted the quantitative volumes of the prostate and tumor.

The majority of available research focuses on prostate segmentation on T2W images [17–21], with only a few studies on tumor segmentation. Because DWI images are more useful for diagnosing PCa [22], this study investigates the simultaneous segmentation and volume quantification of the prostate and tumor using DWI images. This finding has many clinical implications for PCa detection, treatment, and prognosis. On the other hand, DWI images provide significant hurdles in segmenting prostate areas and tumors due to their poor signal-to-noise ratio and substantial variability in prostate morphology. The size, shape, and intensity distribution of the prostate also changes as PCa cells proliferate. The tumor margin is difficult to discern because its shape is irregular, its placement is not fixed, the area is discontinuous, and its intensity varies significantly. All these factors make automated DWI prostate and tumor segmentation difficult and challenging. We added a dual attention mechanism to the network structure of a 3D U-net with a powerful end-to-end encoder-decoder structure to overcome the aforementioned concerns. A visual attention method is built in the encoder step to obtain features of various receptive fields and deliver more detailed contextual information, and a multi-scale attention technique is proposed at the decoder stage to fuse multiscale features to acquire finer global and local details. Existing attention-based picture segmentation algorithms have yielded promising results [4849]. Although most attention mechanisms are channel- and spatial-based, minimal research has been conducted on visual and scale attention. The network for prostate and tumor DWI image segmentation was created with image identification and segmentation in mind, similar to how the human eye makes visual decisions. Two visual attention mechanisms and feature integration theories have been combined to make up the human visual perception process [40]. The center of vision is represented by convolutional layers of various scales in the VAM module that we built, and the visual receptive field is increased by dilated convolutional layers of various speeds. We propose a SAM feature learning module to better exploit the improved multi-scale features of feature consistency created by feature learning. Two attention methods were designed to offer 3D Unet with more detailed semantic information, which is advantageous for automatic and accurate prostate and tumor segmentation. In addition, we created a hybrid loss function for dealing with class imbalance in segmentation tasks to address class discrepancies across the prostate,

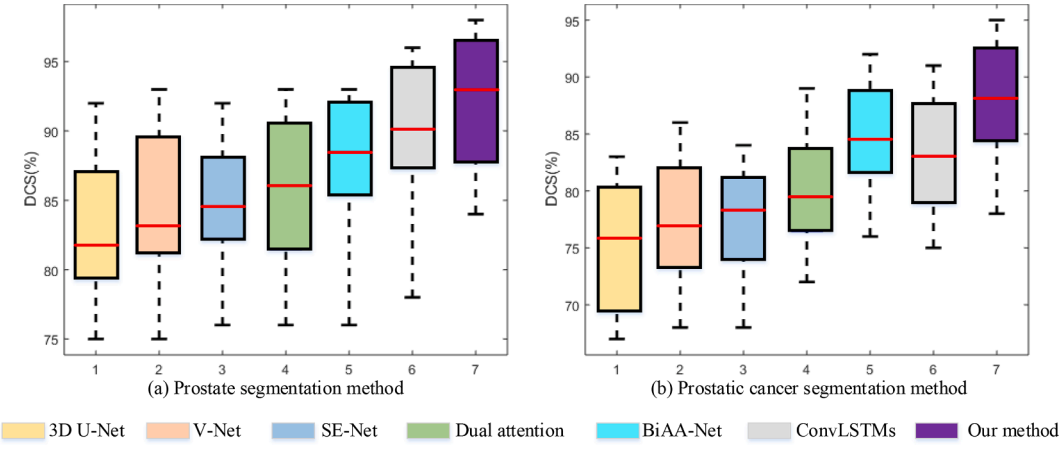

Fig. 7. Distribution of DSC of prostate and tumor regions from 3D U-Net, V-Net, SE-Net, Dual attention, BiAA-Net, ConvLSTMs and our method in the test set. The boxplot consists of five parts: the minimum value, the lower quartile, the median, the upper quartile, and the maximum value of DSC from different methods.

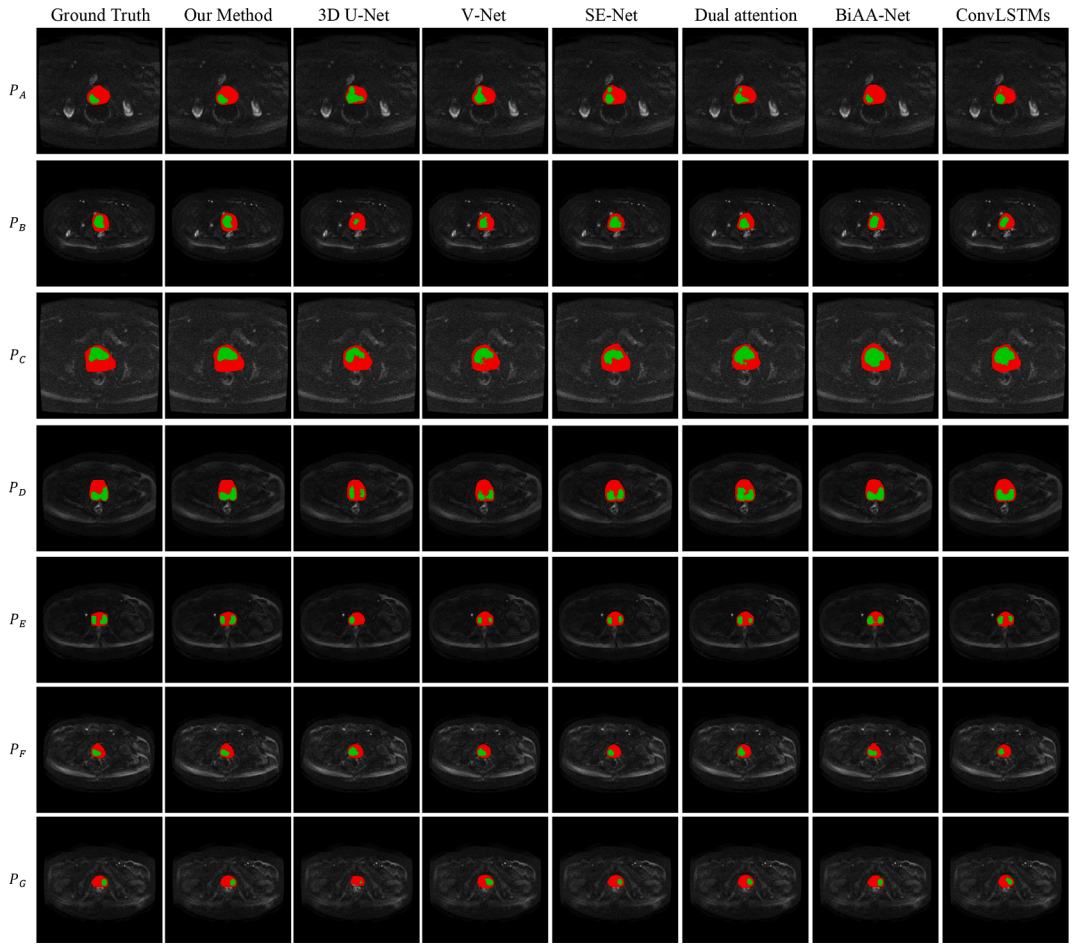

Fig. 8. Different methods for segmentation results of DWI images of prostate and tumor area from 7 patients. Each row is a two-dimensional slice image of different examples, in which red and green represent the prostate and the tumor area, respectively. Each column is the segmentation results of different methods. From left to right: ground truth, our methods, 3D U-Net, V-Net, SE-Net, Dual attention, BiAA-Net and ConvLSTMs.

tumor, and background regions. Ablation experiments on the network structure show that our proposed dual attention mechanism achieves good performance in DWI prostate and tumor segmentation, reaching high DSC values. Compared to existing classical 3D segmentation networks [37 45], our model provides more accurate segmentation results. Compared to the channel attention mechanism [48] and spatial attention mechanism [49], the prostate segmentation DSC value increased by 8.5% (92.28% vs. 83.78) and 7.67% (92.28% vs. 84.61), respectively, and our tumor segmentation DSC value increased by 10.1% (88.73% vs. 78.63%), and 7.69% (88.73% vs. 81.04%) respectively. Compared with the state-of-the-art prostate T2W segmentation method [16] and PCa segmentation method [50], our method achieved higher scores on all evaluation metrics. As can also be seen in the visualization results, our segmentation results were highly correlated and consistent with the expert manual annotation results.

Our strategy has certain limitations. Compared to the scenario of a small number of samples, our method increases the sample size during the experiment, and the segmentation results improve. This is a typical flaw in this and other deep-learning-based methods. After considering how to increase the sample size and enhance segmentation accuracy, we will continue to gather prostate MRI image data. Furthermore, all our data were obtained from Central South University's Haikou People's Hospital and Xiangya Medical College's Affiliated Hospital. In the future, we will examine gathering PCa MRI images from various locations, using various acquisition equipment and following various methods. Create an algorithm that can simultaneously detect and segment multisite data, as well as a solution to the image changes generated by data heterogeneity that affect prostate and tumor segmentation. Finally, because of the dual attention mechanism design in our 3D network, the network parameters and computational complexity have grown, reducing the model's segmentation efficiency of the model. We will improve the efficiency of the model in the future so that it can function more efficiently and rapidly while segmenting precisely and automatically.

5. Conclusions

To increase the accuracy of automatic segmentation and tackle the problems of morphological differences, changeable sizes, blurred boundaries, and class imbalances between the prostate and tumor regions on DWI images. This paper presents a dual-attention guided 3D DAG-Net for prostate and tumor segmentation. The VAM module is designed after the encoder stage. VAM has two levels of convolution blocks for each groups. The first layer is a convolution group of various sizes, whereas the second layer is a dilated convolution group with various speeds. Different features were merged into the visual attention module based on the joint cooperation of conventional and dilated convolutions to provide visual features rich in detailed information. The SAM module is designed after the decoder stage. The module was divided into two sections. The first half of the input aggregates the neighboring scale features generated by the decoders in stages. The second part of the input is the fusion feature map of all the features generated by the decoder, and then uses global average and global max pooling to extract global information. When the two-part structure was

combined, more global and local feature information was obtained. To address the class imbalance of the prostate, tumor, and background regions, we created a loss function. Our proposed network's prostate and tumor segmentation results on a large number of diversified male pelvic DWI datasets show that the segmentation results of our proposed algorithm are highly consistent with expert manual segmentation results and have better segmentation performance than existing methods.

Funding

This work was supported in part by the National Natural Science Foundation of China (Grant #: 82260362), in part by the National Key R&D Program of China (Grant #: 2021ZD0111000), in part by the Key R&D Project of Hainan province (Grant #: ZDYF2021SHFZ243), in part by the Major Science and Technology Project of Haikou (Grant #: 2020-009), in part by the Hainan Provincial Natural Science Foundation of China(Grant #: 621MS019).

Availability of data and materials

The data that support the findings of this study are available on request from the corresponding author, Mengxing Huang. The data are not publicly available due to ethical restrictions and privacy protection.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

The author sincerely thanks all the participants and staff of the school of information and communication engineering of Hainan University; We would also like to thank Haikou Municipal People's Hospital and Xiangya Medical College Affiliated Hospital of Central South University for providing data support for our research.

References

[1] B.S. Chhikara, K. Parang, Global Cancer Statistics 2022: the trends projection analysis, Chem. Biol. Lett. 10 (1) (2023) 451.
[2] A.B. Reed, D.J. Parekh, Biomarkers for Prostate cancer detection, Expert Rev Anticancer Ther 10 (1) (2010) 103–114.
[3] B. March, G. Koufogiannis, M. Louie-Johnsun, Management and outcomes of Gleason six Prostate cancer detected on needle biopsy: A single-surgeon experience over 6years - ScienceDirect, Prostate Int. 5 (4) (2017) 139–142.
[4] P.C. Moldovan, V.D.B. Thomas, R. Sylvester, et al., What Is the Negative Predictive Value of Multiparametric Magnetic Resonance Imaging in Excluding Prostate cancer at Biopsy? A Systematic Review and Meta-analysis from the European Association of Urology Prostate cancer Guidelines Panel, Eur. Urol. 72 (2) (2017) 250–266.
[5] C.K. Kim, B. Park, Diffusion-weighted MRI at 3T for the evaluation of prostate, Am. J. Roentgenol. 194 (6) (2010) 1461–1469.
[6] Z. Wang, W. Zhao, J. Shen, et al., PI-RADS version 2.1 scoring system is superior in detecting transition zone PCa: a diagnostic study. Abdominal, Radiology 45 (7) (2020).
[7] K.S.A. Kumar, A.Y. Prasad, J. Metan, A hybrid deep CNN-Cov-19-Res-Net Transfer learning architype for an enhanced Brain tumor Detection and Classification scheme in medical image processing, Biomed. Signal Process. Control 76 (2022), 103631.
[8] R. Ranjbarzadeh, A. Caputo, E.B. Tirkolaee, et al., Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools, Comput. Biol. Med. 106405 (2022).
[9] R. Ranjbarzadeh, A. Bagherian Kasgari, S. Jafarzadeh Ghoushchi, et al., Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images, Sci. Rep. 11 (1) (2021) 1–17.
[10] N. Tataei Sarshar, R. Ranjbarzadeh, S. Jafarzadeh Ghoushchi, et al., Glioma Brain Tumor Segmentation in Four MRI Modalities Using a Convolutional Neural Network and Based on a Transfer Learning Method[C]//Brazilian Technology Symposium, Springer, Cham, 2023, pp. 386–402.
[11] K. Huang, L. Yang, Y. Wang, et al., Identification of non-small-cell lung cancer subtypes by unsupervised clustering of CT image features with distinct prognoses and gene pathway activities, Biomed. Signal Process. Control 76 (2022), 103643.
[12] H. Pezeshki, Breast tumor segmentation in digital mammograms using spiculated regions, Biomed. Signal Process. Control 76 (2022), 103652.
[13] R. Ranjbarzadeh, N. Tataei Sarshar, S. Jafarzadeh Ghoushchi, et al., MRFE-CNN: multi-route feature extraction model for breast tumor segmentation in Mammograms using a convolutional neural network, Ann. Oper. Res. (2022) 1–22.
[14] R. Ranjbarzadeh, S. Dorosti, S. Jafarzadeh Ghoushchi, et al., Nerve optic segmentation in CT images using a deep learning model and a texture descriptor, Complex & Intelligent Systems (2022) 1–15.
[15] A. Oulefki, S. Agaian, T. Trongtirakul, et al., Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images, Pattern Recogn. 114 (2021), 107747.
[16] Z. Tian, X. Li, Z. Chen, et al., Interactive prostate MR image segmentation based on ConvLSTMs and GGNN, Neurocomputing 438 (2021) 84–93.
[17] S. Montagne, D. Hamzaoui, A. Allera, et al., Challenge of prostate MRI segmentation on T2-weighted images: inter-observer variability and impact of prostate morphology, Insights into imaging 12 (1) (2021) 1–12.
[18] K. Eppenhof, M. Maspero, M. Savenije, et al., Fast contour propagation for MR-guided prostate radiotherapy using convolutional neural networks, Med. Phys. 47 (3) (2020).
[19] S. Wang, M. Liu, J. Lian, et al., Boundary Coding Representation for Organ Segmentation in PCa Radiotherapy, IEEE Trans. Med. Imaging 40 (1) (2020) 310–320.
[20] Q. Liu, Q. Dou, L. Yu, et al., MS-Net: Multi-Site Network for Improving Prostate Segmentation with Heterogeneous MRI Data, IEEE Trans. Med. Imaging 39 (9) (2020) 2713–2724.
[21] Q. Liu, C. Chen, J. Qin, et al., FedDG: Federated Domain Generalization on Medical Image Segmentation via Episodic Learning in Continuous Frequency Space, CVPR (2021).
[22] X. Yin, F. Niu, L. Lin, et al., Diagnostic Value of Magnetic Resonance DWI in Prostate cancer, Modern Medical Imageology (2017).
[23] I. Chan, W. Wells, R.V. Mulkern, et al., Detection of PCa by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier, Med. Phys. 30 (2003).
[24] R. Llobet, J.C. Perez-Cortes, A.H. Toselli, A. Juan, Computeraided detection of Prostate cancer, Int J Med Imformatics (2007).
[25] X. Liu, D.L. Langer, M.A. Haider, et al., Prostate cancer Segmentation With Simultaneous Estimation of Markov Random Field Parameters and Class, IEEE Trans. Med. Imaging 28 (6) (2009) 906.
[26] P. McClure, F. Khalifa, A. Soliman, M.A. El-Ghar, A. El-Baz, A Novel NMF Guided Level-set for DWI Prostate Segmentation, J. Comput. Sci. Syst. Biol. 7 (6) (2014) 209–216.
[27] Y. LeCun, Y. Bengio, et al., Deep learning, Nat. Publ. Group 521 (7553) (2015) 436–444.
[28] R.R. Wildeboer, R.J.G. van Sloun, H. Wijkstra, et al., Artificial intelligence in multiparametric prostate cancer imaging with focus on deep-learning methods, Comput. Methods Programs Biomed. 189 (2020), 105316.
[29] M.H. Le, J. Chen, L. Wang, Z. Wang, W. Liu, K.T. Cheng, X. Yang, Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks, Phys Med Biol 62 (16) (2017) 6497–6514.
[30] X. Yang, Z. Wang, C. Liu, et al., Joint detection and diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks, in: International conference on medical image computing and computer-assisted intervention, Springer, Cham, 2017, pp. 426–434.
[31] R. Cao, X. Zhong, S. Shakeri, et al., Prostate cancer detection and segmentation in multi-parametric mri via cnn and conditional random field, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019, pp. 1900–1904.
[32] A.A. Abbasi, L. Hussain, I.A. Awan, et al., Detecting prostate cancer using deep learning convolution neural network with transfer learning approach, Cogn. Neurodyn. 14 (4) (2020) 523–533.
[33] C.-C. Lai, H.-K. Wang, F.-N. Wang, Y.-C. Peng, T.-P. Lin, H.-H. Peng, S.-H. Shen, Autosegmentation of Prostate Zones and Cancer Regions from Biparametric Magnetic Resonance Images by Using Deep-Learning-Based Neural Networks, Sensors 21 (8) (2021) 2709.
[34] I.R. Abdelmaksoud, A. Shalaby, A. Mahmoud, M. Elmogy, A. Aboelfetouh, M. Abou El-Ghar, M. El-Melegy, N.S. Alghamdi, A. El-Baz, Precise Identification of prostate cancer from DWI Using Transfer Learning, Sensors 21 (11) (2021) 3664.
[35] A. Duran, G. Dussert, O. Rouvière, et al., ProstAttention-Net: A deep attention model for prostate cancer segmentation by aggressiveness in MRI scans, Med. Image Anal. 77 (2022), 102347.
[36] A.A. Malibari, R. Alshahrani, F.N. Al-Wesabi, et al., Artificial intelligence based prostate cancer classification model using biomedical images, Comput. Mater. Contin 72 (2022) 3799–3813.
[37] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, et al., 3D U-Net: learning dense volumetric segmentation from sparse annotation[C]//International conference on medical image computing and computer-assisted intervention, Springer, Cham, 2016, pp. 424–432.
[38] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1026-1034.
[39] S. Yohanandan, A. Song, A.G. Dyer, et al., Saliency Preservation in Low-Resolution Grayscale Images, European Conference on Computer Vision (2018) 237–254.
[40] S.K. Ungerleider, LG, Mechanisms of visual attention in the human cortex, Annu. Rev Neurosci. 23 (1) (2000) 315–341.
[41] S. Liu, D. Huang, Receptive field block net for accurate and fast object detection, Munich, Germany, ECCV, 2018, pp. 385–400.

[42] Liu S, Huang D. Receptive field block net for accurate and fast object detection [C]//Proceedings of the European conference on computer vision (ECCV). 2018: 385-400.
[43] Q. Zhao, T. Sheng, Y. Wang, et al., M2det: A single-shot object detector based on multi-level feature pyramid network, in: Proceedings of the AAAI conference on artificial intelligence. 2019, 33(01): 9259-9266.
[44] H. Xu, J. Zhang, Aanet: Adaptive aggregation network for efficient stereo matching, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 1959-1968.
[45] F. Milletari, N. Navab, S.A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 fourth international conference on 3D vision (3DV), Ieee (2016) 565–571.
[46] T.Y. Lin, P. Goyal, R. Girshick, et al. Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.
[47] A.G. Asuero, A. Sayago, A.G. González, The correlation coefficient: An overview, Crit. Rev. Anal. Chem. 36 (1) (2006) 41–59.
[48] H. Jie, S. Li, S. Gang, Squeeze-and-Excitation Networks, in: 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.
[49] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual attention network for scene segmentation, IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2019 (2019) 3146–3154.
[50] G. Zhang, et al., A Bi-Attention Adversarial Network for Prostate cancer Segmentation, IEEE Access 7 (2019) 131448–131458, https://doi.org/10.1109/ACCESS.2019.2939389.