Magnetic Resonance Imaging 99 (2023) 98–109

Contents lists available at ScienceDirect

Magnetic Resonance Imaging

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

Prostate lesion segmentation based on a 3D end-to-end convolution neural network with deep multi-scale attention

Enmin Song [a], Jiaosong Long [a], Guangzhi Ma [a,*], Hong Liu [a], Chih-Cheng Hung [b], Renchao Jin [a], Peijun Wang [c], Wei Wang [c]

[a] School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China
[b] College of Computing and Software Engineering, Kennesaw State University, Atlanta, USA
[c] Department of Radiology, Tongji Hospital, School of Medcine, Tongji University, Shanghai 200065, China

ARTICLE INFO

Keywords:
Mp-MRI
Prostate cancer segmentation
Convolution neural network
Attention

ABSTRACT

Prostate cancer is one of the deadest cancers among human beings. To better diagnose the prostate cancer, prostate lesion segmentation becomes a very important work, but its progress is very slow due to the prostate lesions small in size, irregular in shape, and blurred in contour. Therefore, automatic prostate lesion segmentation from mp-MRI is a great significant work and a challenging task. However, the most existing multi-step segmentation methods based on voxel-level classification are time-consuming, may introduce errors in different steps and lead to error accumulation. To decrease the computation time, harness richer 3D spatial features, and fuse the multi-level contextual information of mp-MRI, we present an automatic segmentation method in which all steps are optimized conjointly as one step to form our end-to-end convolutional neural network. The proposed end-to-end network DMSA-V-Net consists of two parts: (1) a 3D V-Net is used as the backbone network, it is the first attempt in employing 3D convolutional neural network for CS prostate lesion segmentation, (2) a deep multi-scale attention mechanism is introduced into the 3D V-Net which can highly focus on the ROI while suppressing the redundant background. As a merit, the attention can adaptively re-align the context information between the feature maps at different scales and the saliency maps in high-levels. We performed experiments based on five cross-fold validation with data including 97 patients. The results show that the Dice and sensitivity are **0.7014** and **0.8652** respectively, which demonstrates that our segmentation approach is more significant and accurate compared to other methods.

1. Introduction

Prostate Cancer (PCa) is one of the deadliest cancers. There are about 180,890 newly diagnosed cases and 26,120 deaths of PCa per year [1]. It is estimated that there will be more than 170,000 PCa patients worldwide by 2030, and it may cause more than 500,000 patients facing of mortality annually [2]. Early detection and reliable diagnosis of PCa with careful treatment can suppress the metastatic progressing, this may greatly enhance the survive rate of patients and relieve their huge suffering caused by PCa.

Normally, patients with PCa are divided into low, intermediate, or high risk group according to their prostate-specific Antigen (PSA) level, pathological assessment/Gleason Score (GS) [3,4], and clinical stage [5]. In general, 90% PCa cases are defined as benign as the Gleason Score (GS) [6] is smaller than and equal to 6. The benign lesion will deteriorate without noticed, thus adequate follow-up for patients is very important. Once the GS is higher than 7, it becomes very serious and even leads to death. To avoid over-treatment in the low-risk stage and to suppress the continuous deterioration, the demand for automatic prostate lesion analysis is highly eagerly required.

To precisely interpret prostate lesion, abundant image information is very useful for doctors. The multi-parameter magnetic resonance imaging (mp-MRI), which includes Diffusion-Weighted Imaging (DWI), KTrans imaging and Dynamic Contrast Enhanced (DCE) MRI, can capture both the structure and the pathological information of a prostate, so it is widely used in PCa segmentation and detection [7–13].

However, interpreting mp-MRI data manually is an onerous task, which requires considerable amount of expertise skills and is quite time consuming. In addition, manual analyzing is easy to be affected by some subjective factors such as wrong marking. Therefore, developing an

* Corresponding author.
E-mail address: maguangzhi@hust.edu.cn (G. Ma).

https://doi.org/10.1016/j.mri.2023.01.015
Received 13 October 2021; Received in revised form 6 July 2022; Accepted 14 January 2023
Available online 18 January 2023

automatic segmentation system for PCa diagnosis from mp-MRI to alleviate the burden of radiologist and to reduce the risk of over−/under-treatment becomes the current primary target.

In the past decade, computer-aided (CAD) PCa segmentation, detection and diagnosis are gradually developing [10–13]. Generally, the existed CAD system can be divided into two types: (1) computer-aided segmentation and detection system, which mainly concerns on identifying the present of prostate lesion and localizing the suspicious lesion in mp-MRI sequences, and (2) computer-aided diagnosis system, which aimed at diagnosing PCa benign or malign according to the lesions manually selected by radiologists or automatically segmented by the prior CAD systems. In this paper, we focus on the study of the first one.

Prostate lesion segmentation and detection typically includes two steps: (1) prostate segmentation, which focuses to find the boundary of prostate region [14], and (2) suspicious lesion segmentation and detection, which concern on finding the boundary for the suspicious PCa lesion. Then the PCa can be diagnosed according to the suspicious PCa lesion.

In [11,15–19], they manually segmented the prostate then studied to find suspicious lesions. Manually delineating prostate boundary of each slice layer was time-consuming and easy to introduce errors. Later, many automatic prostate segmentation methods [20–22] were proposed by using multi-atlas-based method for prostate lesion segmentation.

To segment and detect prostate cancer, Chan et al. [23] applied co-occurrence matrix and discrete cosine transform to extract texture features from the line-scan diffusion, T2, and T2-W images. Then, a SVM [24] classifier is employed to generate a malignancy likelihood map on the peripheral zone of the prostate for the segmented suspicious PCa lesion.

Tiwari et al. [25] proposed to combine the magnetic resonance spectroscopy and KTrans images to find the voxels with pathological feature, then the voxels of PCa are represented by a series of features including intensities and blobness of different sequences, texture strength and homogeneity. To segmentation PCa regions, Litjen et al. [7] applied blob detection based on Hessian matrix to classify each voxel of a multi-scale ADC sequence.

To improve the accuracy of PCa lesion localization, Artan et al. [26] proposed to use the cost-sensitive SVM to segment PCa lesions, and it resulted in an improved localization accuracy compared to the classical SVM. They further integrated conditional random field [27] with the cost-sensitive SVM and got more accurate prostate lesion segmentation than only the cost-sensitive SVM is used.

Most of the above lesion segmentation methods were based on voxel-level classification with two steps. However, the misclassified voxel would affect the subsequent steps, and reduce the overall segmentation accuracy. Moreover, voxel-level classification would naturally ignore the global information of the prostate, and lead to train a flawed network with a large number of false positives.

The traditional two step segmentation can be simplified as one step segmentation by using deep learning method. The benefit of one step segmentation is that it can prevent information loss of the first step, thus improving the segmentation accuracy and increasing the segmentation efficiency.

In recent years, deep learning technology [28] has been widely employed in computer vision and machine learning filed due to their success of high accuracy. Despite deep learning has superior performance in many medical image applications such as breast mammography, lung CT and brain MRI, development of prostate cancer CAD is still in very slow pace.

Based on deep learning, some studies on prostate lesion segmentation and detection have been presented. Kiraly et al. [29] proposed to use a multi-channel image-to-image convolutional encoder-decoder for PCa lesion detection. Wang et al. [30] presented a cascaded training strategy to train FCN with a weighted cross entropy loss function for tumor segmentation from mp-MRI including KTrans, DWI, ADC. Abbasi et al. [31] applied a transform learning approach to learn cancer images with GoogleNet to detect prostate cancer. Yoo et al. [32] developed an automatic CNN-based pipeline to detect clinically significant prostate cancer for a given DWI image of a patient. Q Lee et al. [33] proposed a new model that amalgamated the U-Net and the convolutional gated recurrent unit (convGRU) neural network architecture, aiming at interpreting DCE time-series on the temporal and spatial basis for segmenting PCa in MR images. Cao et al. [34] employed a focal loss to adjust the imbalance between the normal and cancerous areas, and developed a selective dense conditional random field method, to refine CNN prediction into lesion segmentation based on the intensity pattern of a specific imaging component of a mp-MRI.

The above methods show the proposed methods can get better precision. But still there is room for further improvement. While studying lesion segmentation for other diseases, we discover that the accuracy of using 3D deep learning architecture is higher than that of using 2D deep learning architecture. We think the higher accuracy is due to the 3D deep learning ability in capturing complex spatial structures and texture features from high-dimensional data.

However, fully 3D CNNs come with an increased number of parameters and significant memory and computation requirements, leading to the limitation to the network depth and thus weakening the learning ability of the 3D deep network.

To address the limitation of learning ability caused by the excessive parameters of 3D deep networks produced, the attention mechanism is used to solve the problem of learning ability. Human vision studies [35,36] found that the visual attention mechanism can quickly guide the network components to find the interesting objects in images.

Generally, attention mechanism is divided into soft-attention and hard-attention categories. For the former, the attention weight was assigned by the using of continuous function to the input. For the latter, hard attention is used to highlight specific locations by assigning sampling weights. Recently, attention mechanisms have been applied to medical image applications with lots of successes.

For instance, Schelemper et al. [37] proposed an attention-gated framework for real-time automatic plane scan detection in fetal ultrasound screening. Jin et al. [38] proposed a 3D hybrid residual attention-aware segmentation method to segment the tumors from the patient's liver. As the above attention mechanisms have achieved great success in medical image segmentation, Zhang et al. [39] proposed to use channel attention and position attention simultaneously in a single adversarial 2D network to complete prostate cancer region segmentation.

In order to capture richer semantic information and the spatial feature with attention of multi-scale to improve network discrimination, we proposed a deep multi-scale attention 3D-V-Net (DMSA-V-Net) based framework, which is an automatic end-to-end PCa lesion segmentation network for mp-MRI. This network integrates 3D architecture for complex spatial features with attention mechanism to optimize the deep network for learning performance. The main contributions of this article include the following:

① We start first to employ the fully deep encoder-decoder 3D-CNN to segment the region of PCa, it can capture the global information representations and spatial features to improve the network learning ability.
② We propose a deep multi-scale attention mechanism for PCa region segmentation , enabling the network focusing on the interest regions of feature maps and reducing the weakened learning ability of 3D network due to its numerous parameters and low computing efficiency.
③ We put forward DMSA-V-Net to extract multi-scale attention information features which are connected with corresponding low-level high-resolution feature maps to compensate semantic information to improve the segmentation accuracy.

The content of this paper is listed as follows: in Section 2, the

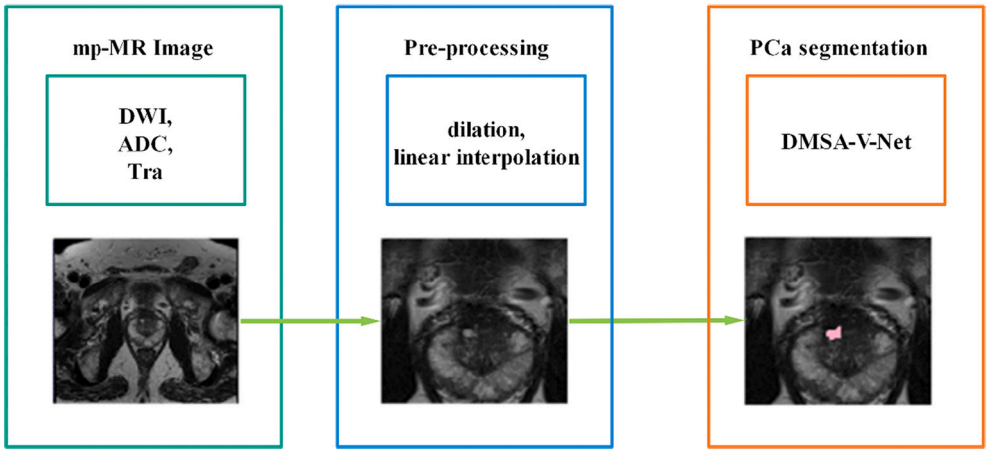

Fig. 1. Flow chart of the proposed algorithm including pre-processing and PCa lesion segmentation.

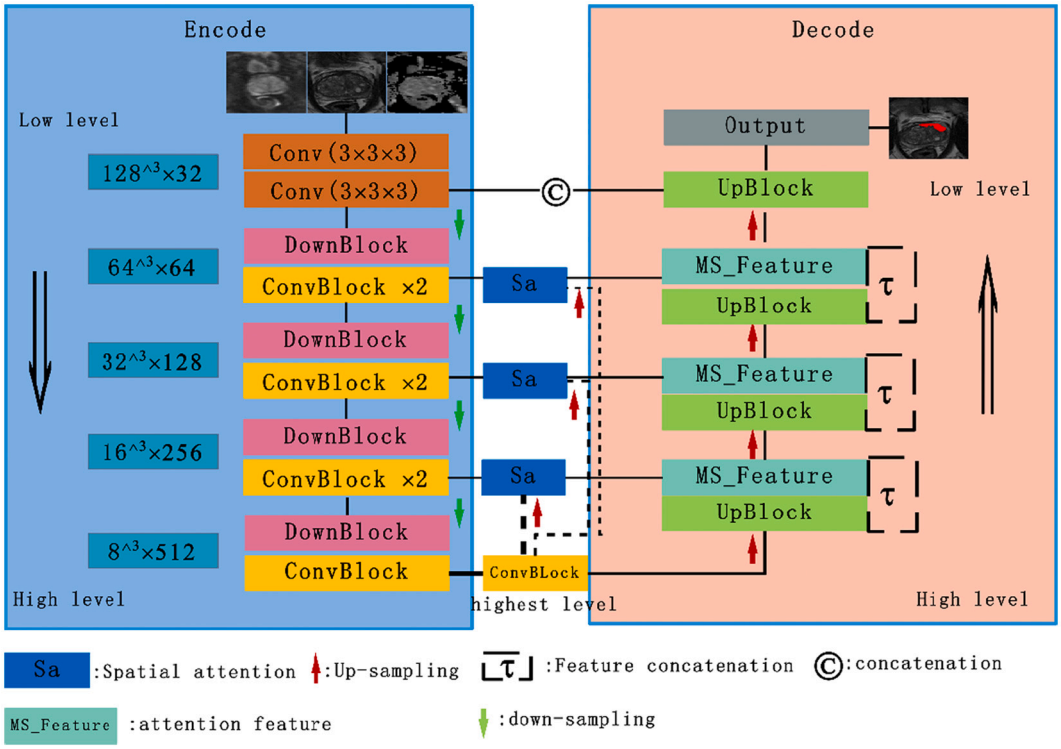

Fig. 2. The framework of our proposed PCa lesion segmentation. ADC, T2-weighted and DWI images stacked as different channels of the DMSA-V-Net input.

proposed methods are described in detail, including data pre-processing, multi-scale attention mechanism and 3D-V-Net; in Section 3, the experimental demonstration is presented and the analysis about the proposed methods is illustrated; in Section 4, discussion and conclusion is conducted.

2. Method

The mp-MRI data including T2-weighted, DWI, DCE, KTrans, ADC etc., have been widely applied in PCa detecting and segmenting. The results of studies [8,9] have shown that the combination of ADC and T2-weighted can significantly improve PCa lesion segmentation performance. Therefore, in this paper, three image sequences ADC, T2-weighted and DWI are used to complete PCa lesion segmentation to achieve better performance. The structure of this paper is divided into two parts: (1) image preprocessing to unify the mp-MRI data formats for all patients; (2) 3D-V-Net constructing with deep multi-scale attention (DMSA-V-Net) for PCa lesion segmentation. The flow chart of the proposed method is shown in Fig. 1.

2.1. Data preprocessing

The existing problems of mp-MRIs of patients are unclear image, noise image, inconsistent resolution, inconsistent spacing and partial prostatic cavity. To deal with the prostatic cavity, we apply affine registration to register different modalities to T2-weighted image data. For each patient, we use morphological preprocessing to fill the cavity with kernel of size of 5×5. Due to the large gaps between the slices on axial, leading the blurry imaging on sagittal and coronal, hence the T2-weighted sequence is selected as the reference, then other sequences are linearly interpolated according to the size of the T2-weighted to improve the image quality. For the sampling of PCa region, we cropped the

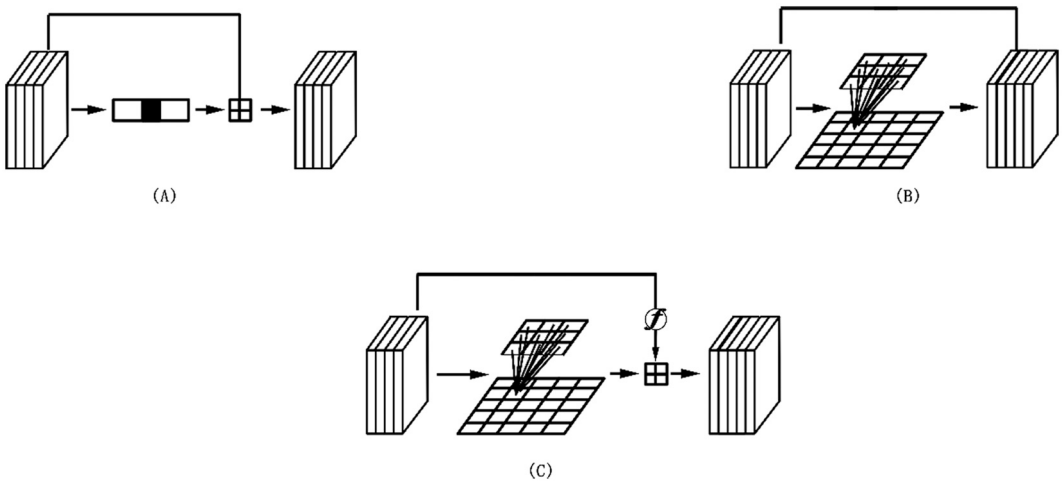

Fig. 3. Different feature connection for information compensation in CNN. (A) Residual unit connection; (B) Feature maps concatenation; (C) Gated block connection.

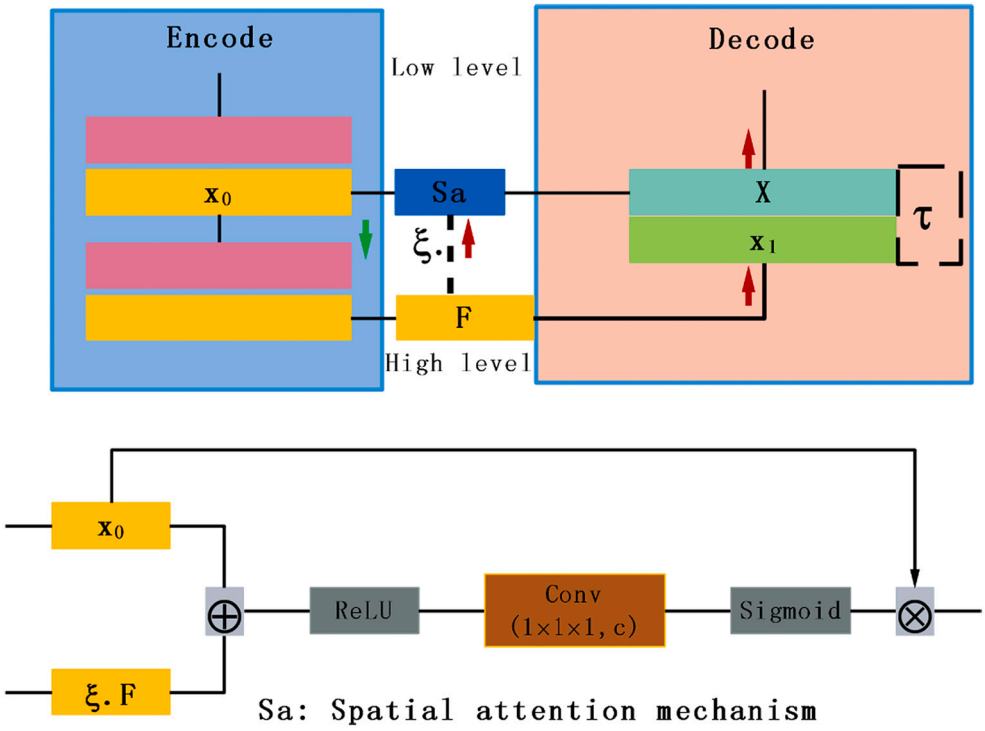

Fig. 4. The components of spatial attention mechanism.

region of interest (ROI) including the prostate with a fixed size of 128 × 128 × 32, and adopt histogram equalization to enhance local contrast. Finally, zero mean normalization are carried out on the ROI images.

2.2. Deep multi-scale attention 3D-V-net

2.2.1. Model overview

The multi-scale attention 3D-V-Net comprises three modules: feature connection, deep multi-scale spatial attention and DMSA-V-Net, as shown in Fig. 2. Concretely, we constructed the entire 3D network based on 3D-V-Net with the popular encoder-decoder manner and introduced spatial wise attention to enhance the learning of spatial target. The attention mechanism serves as a bridge to gradually fuse low-level high-resolution features from bottom layers with high-level but low-resolution features from top layers, which is expected to be useful for the decoder to generate high-resolution semantic results. With additional high-resolution features embedded, high-level features may have a potential to refine itself by aligning to the nearest low-level boundary.

2.2.2. Feature connection

Similar to skip connection, feature connection is regarded as the compensation of information in CNN. Evidently, the proper fusion of different-level features has shown effectiveness for semantic information compensation in the area of computer vision [40]. Further, it has been proved that the loss of information during the training will lead to a severe degradation problem [41] when constructing a deeper network.

Here, we classify the commonly used information compensation operations into 3 categories: feature map concatenation, residual unit connection and gated block connection [37], as shown in Fig. 3. In this paper, we adopt feature maps concatenation to concatenate the low-

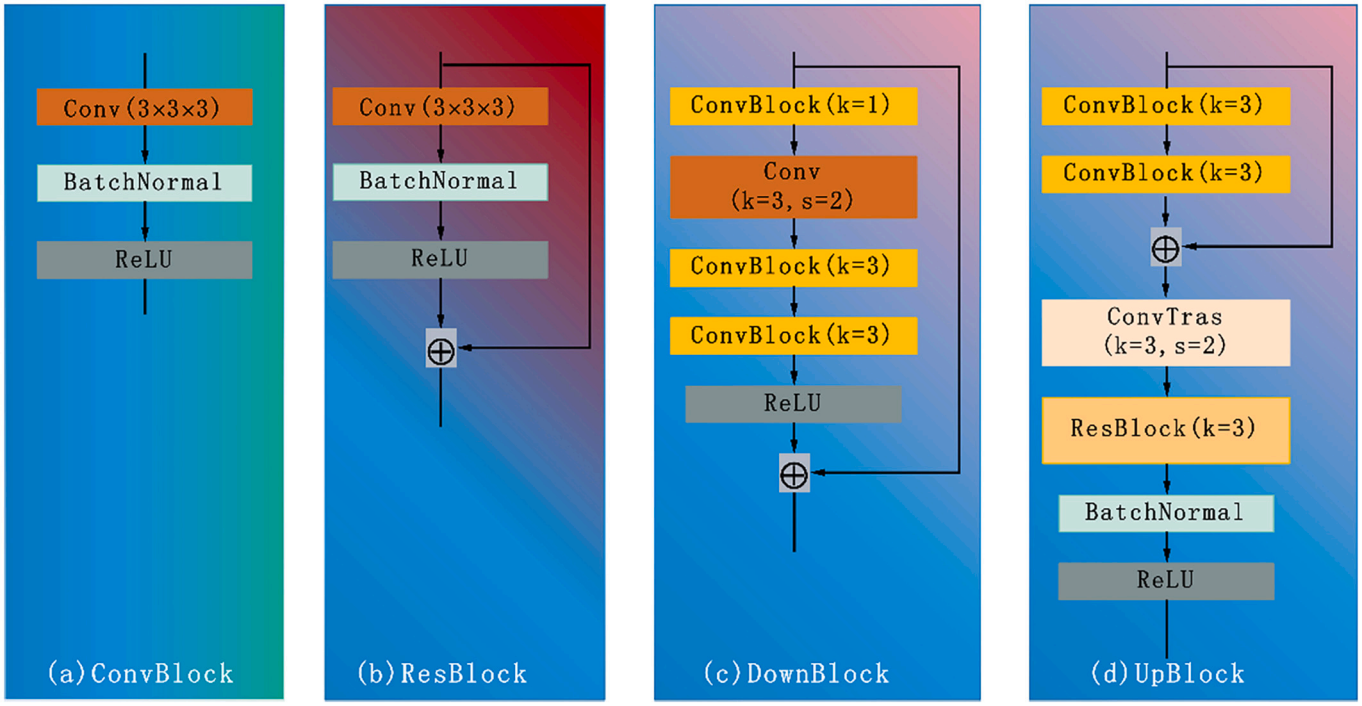

Fig. 5. Illustration of the used components in DMSA-V-Net.

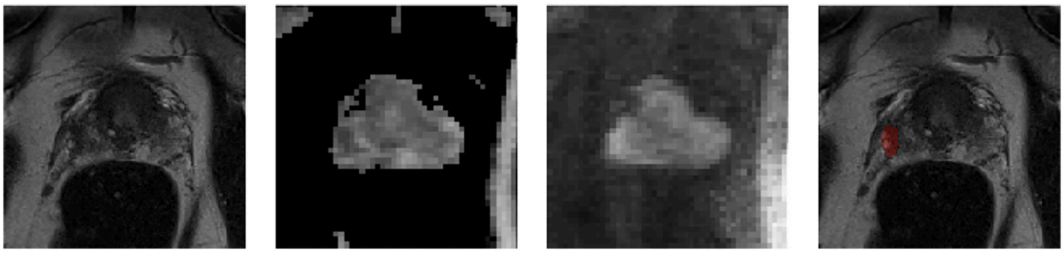

Fig. 6. Example of the patient no.46 MRI with different image modalities and the manual annotation by experts; the first three images from left to right are the sequences of T2-weighted, ADC and DWI respectively, the fourth is the ground truth in sequence of T2-weighted where the red color represents the lesion. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Performance of the three different architectures A, B and C is shown in Fig. 7 by using 5-fold cross-validation. The 3D-V-Net denotes the backbone network shown in Fig. 7 (A); the A-3D-V-Net denotes the improved architecture shown in Fig. 7 (C); the DMSA-V-Net denotes our proposed architecture shown in Fig. 7 (B).

	Sensitivity	Specificity	Dice
3D-V-Net	0.7425 ± 0.06	0.99	0.6603 ± 0.02
A-3D-V-Net	0.8343 ± 0.05	0.99	0.6795 ± 0.02
DMSA-V-Net	**0.8652 ± 0.03**	**0.99**	**0.7014 ± 0.02**

level and high-level feature maps of corresponding layers in each scale, utilizing short connection for information compensation because of its excellent effect and simple connection way under encoder-decoder manner.

2.2.3. Deep multi-scale spatial attention

The spatial attention can focus on specific target of the image to learn the features of "where". When observing a complex scene involving multiple objects, the human visual system can selectively locate and analyze the regions of interest in the object. The visual selective attention can solve the object ambiguities for different tasks. In the field of computer vision, many researchers have developed similar attention mechanisms and demonstrated that they c-ould significantly improve performance. Moreover, according to the conclusion [42], multi-scale feature learning can adaptively aggregate the features of different layers to enrich the representation; simultaneously inspired by [43], we presented our multi-scale spatial attention.

(1) Spatial attention mechanism. We define an alternative slicing of the input tensor $\mathbf{F} = [f^{1,1,1}, f^{1,1,2}, f^{1,1,3}, \ldots, f^{H,W,D}]$, where $f^{ijk} \in R^c$ is corresponding to the spatial location (i,j,k) with $i \in \{1,2,3,\ldots,H\}$, $j \in \{1,2,3,\ldots,W\}$, $k \in \{1,2,3,\ldots,D\}$. The spatial attention theory is achieved through a convolution $a = \boldsymbol{W_{oa}} \times \boldsymbol{F}$ to generate a projection tensor $a \in R^{H \times W \times D}$, each a_{ijk} of the projection indicates the feature after convolution for all C channels at a spatial location (i,j,k). This projection tensor is mapped by sigmoid layer $\sigma(.)$ to rescale activations into $[0,1]$, which is applied to highlight the important spatial location of $\mathbf{F}$,

$$S_a = \left[\sigma\left(a_{1,1,1}\right)f^{1,1,1}, \ldots, \sigma\left(a_{i,j,k}\right)f^{i,j,k}, \ldots, \sigma\left(a_{H,W,D}\right)f^{H,W,D}\right]. \quad (1)$$

Each value $\sigma(a_{i,j,k})$ corresponds to the relative information importance of a spatial location (i,j,k) of a given feature map. It assigns higher weight to the relevant spatial location and lower to the irrelevant.

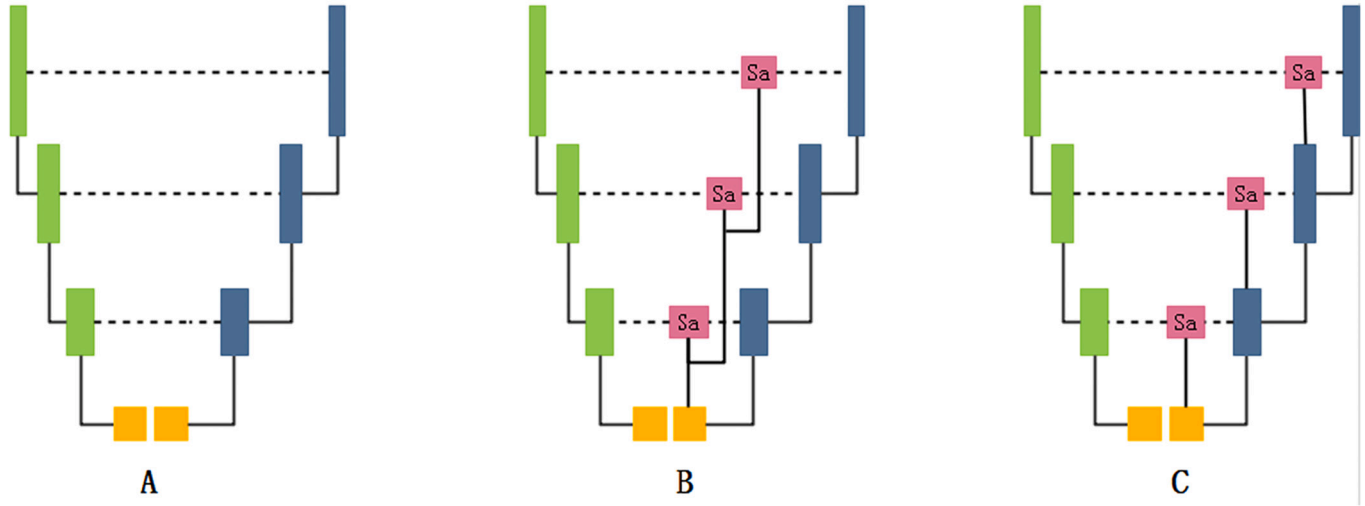

Fig. 7. Three different ways of experiments for comparison. A: the original 3D-V-Net architecture; B: the 3D-V-Net connecting with deep multi-scale spatial attention (DMSA-V-Net); C: the 3D-V-Net connecting with spatial attention mechanism in others way (A-3D-V-Net).

(2) Deep multi-scale attention. There is a difference between high- and low-level feature maps in the encoder-decoder architecture. Therefore, it is worth to fuse the low-level features with rich spatial details and high-level features with rich semantic information, with a view to modulating the information passed to the decoding stage. Hence, we rescale the feature maps of highest-level in encoding stage to the same size of each low-level for generating multi-scale attention feature maps and concatenating with each level of high-resolution features in decoding stage for information compensation. Finally, the multi-scale attention feature maps X can be obtained by multiplying each low-level feature x_0 with S_a, then concatenating with the corresponding high-resolution features of decoding stage x_1 for information compensating by feature concatenation $\tau[\ ]$:

$$X = \tau[x_0 \otimes S_a(\zeta.\mathbf{F}), x_1] \quad (2)$$

where x_0 denotes the feature maps of each low-level high-resolution in encoding stage, ζ. **F** denotes to multiply the different scaling factors to rescale the size of highest-level feature maps **F** the same as x_0, $\otimes$ denotes the operation of element-wise multiplication, x_1 denotes the feature maps of decode stage de-convoluted to the same size as x_0, $\tau[,]$ denotes the feature maps concatenation between the spatial attention features and x_1.

The components of the deep multi-scale spatial attention is shown in the Fig. 4.

2.2.4. DMSA-V-net

Aimed at learning the characteristic of MR medical images, which had small target and complex background, the 3D-V-Net had been successfully applied in prostate segmentation [44].Therefore, we utilized the 3D-V-Net as the backbone network and proposed our DMSA-V-Net to accomplish the prostate lesion segmentation task.

The overview of the architecture of DMSA-V-Net is depicted in Fig. 2. The encoder and the decoder are symmetrically presented on the two sides of the DMSA-V-Net. The encoder propagates the contextual information with skip connections which enable extracting much richer feature representation. The decoder receives features from multiple levels and regenerates features in a coarse-to-fine manner.

The main layers used to construct the DMSA-V-Net consist of convolution layer, batch-normal layer and activation layer. These layers can be stacked on top of each other to form a hierarchy feature maps. For the 'ConvBlock' and the 'ResBlock', both of them contain convolution layer, batch-normalization layer and ReLU activation layer, but the latter has a skip connection, as shown in Fig. 5 (a) and (b). For the 'DownBlock', we adopted a convolution layer with kernels of size 1 to form the cross-channel parametric pooling layer, then following by connecting a convolution layer with kernel of size $3 \times 3 \times 3$ and stride of $2 \times 2 \times 2$ to reduce the feature map size for contracted feature representation, and followed by two 'ConvBlock' and ReLU, as shown in Fig. 5 (c). However, we used convolution layer with stride of $2 \times 2 \times 1$ in the first two down-sampling steps because the input image has fewer number of slice in axial. For the 'UpBlock', we used two stacked 'ConvBlock' with shortcut, then the following layers are 'ConvTranspose' layer, 'ResBlock', batch-normal layer and ReLU, as shown in Fig. 5 (d).

We utilized the features maps of 'ConvBlock' generated in first three level of encode to generate the multi-scale attention feature maps 'MS_Feature', then connecting with the feature maps 'UpBlock' generated by feature concatenation for information compensation. Through the aggregation of high- and low-level features the network's learning ability can be enhanced.

3. Experiments and results

3.1. Dataset and experiment setup

Our dataset for experiments is from the part of PROSTATEX challenge, in order to segment the clinically significant lesion, we select all patients whose Gleason scores are greater than or equal to 6. In the experiment, we evaluate the performance of PCa lesion segmentation based on 5-fold cross-validation. All the data contains 97 patients and totally 107 PCa lesions. Three radiologists are invited to manually segment the lesions, and the intersected segment area is used as the final region. The mp-MRI of each patient includes T2-weighted, Apparent Diffusion Coefficient (ADC), Diffusion Weighted Images (DWI), KTrans, but only the three sequences e.g. T2-weighted, ADC, DWI, were used because the best experimental results could be achieved. The three sequences of T2-weighted, ADC, DWI are shown in Fig. 6.

Our network was implemented with Keras based on Tensorflow and the proposed network was trained and tested on Nvidia Geforce GTX 1080Ti GPU. According to [44], we built the network based on the original frames. Our segmentation network took the cropped ROI image in size of $128 \times 128 \times 32$ as its input. During the training, we augmented data by using rotation, scaling original volumes and random shift; and the Adam [45] method was used to optimize our network with batch size 1, $\beta_1 = 0.9$, $\beta_2 = 0.999$. We set the initial learning rate to $2 \times e^{-5}$ and

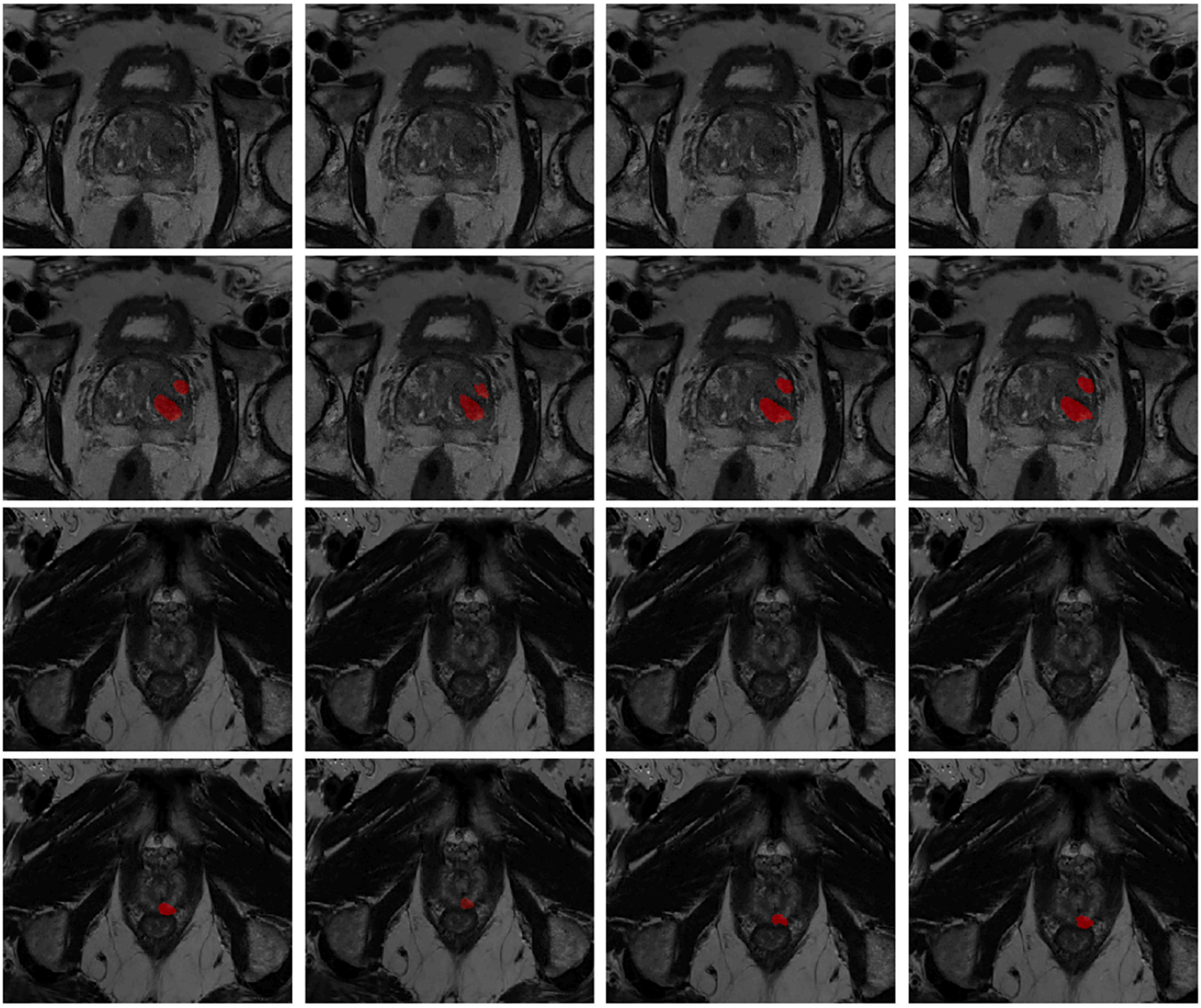

Fig. 8. The DMSA-V-Net is more useful for PCa lesion segmentation. The first row and the third row respect a slice of T2-weighted sequence from two different patient respectively; From left to right in the row second and the row fourth, each of which respects ground truth, and the results segmented by 3D-V-Net, A-3D-V-Net and DMSA-V-Net.

Table 2
Cross-validation results of MRI PCa lesion segmentation using different MR image sequences.

	Sensitivity	Specificity	Dice
T2-weighted, DWI, ADC, KTrans	0.8284 ± 0.04	0.99	0.6769 ± 0.02
T2-weighted	0.6258 ± 0.08	0.99	0.5495 ± 0.06
T2-weighted, DWI, ADC	**0.8652 ± 0.03**	0.99	**0.7014 ± 0.02**

decrease the learning rate by 20% after by 20 epochs. The soft Dice loss function was used to train the proposed network:

$$\text{Dice} = 1 - \frac{2\sum_i^N g_i p_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{3}$$

where $p_i \epsilon [0, \ldots, 1]$ denotes the prediction probability of lesion and $g_i \epsilon \{0, 1\}$ denotes the binary ground truth value for each pixel i.

Table 3
Comparison on Dice, Sensitivity and Specificity between our proposed DMSA-V-Net and the state-of-the-art approaches for PCa lesion segmentation on our dataset. The bold values correspond to the best metric scores.

	Sensitivity	Specificity	Dice
Kiraly [29]	0.5865 ± 0.07	0.99	0.6012 ± 0.11
Wang [30]	0.4823 ± 0.08	0.99	0.4458 ± 0.08
nn-U-Net [46]	0.7963 ± 0.04	0.99	0.6286 ± 0.03
Deeplabv3+ [47]	0.6978 ± 0.05	0.99	0.5940 ± 0.04
Proposed	**0.8652 ± 0.03**	**0.99**	**0.7014 ± 0.02**

3.2. Evaluation measures

To fully evaluate the PCa lesion segmentation performance, three wildly-used metrics are adopted in our experiments, i.e., sensitivity, dice and specificity. Sensitivity and specificity indicate the proportion of correctly classified positive and negative pixels in the image respectively. Dice indicates the similarity between the predicted and the ground truth pixels. They are given in Eqs. (3), (4) and (5) respectively.

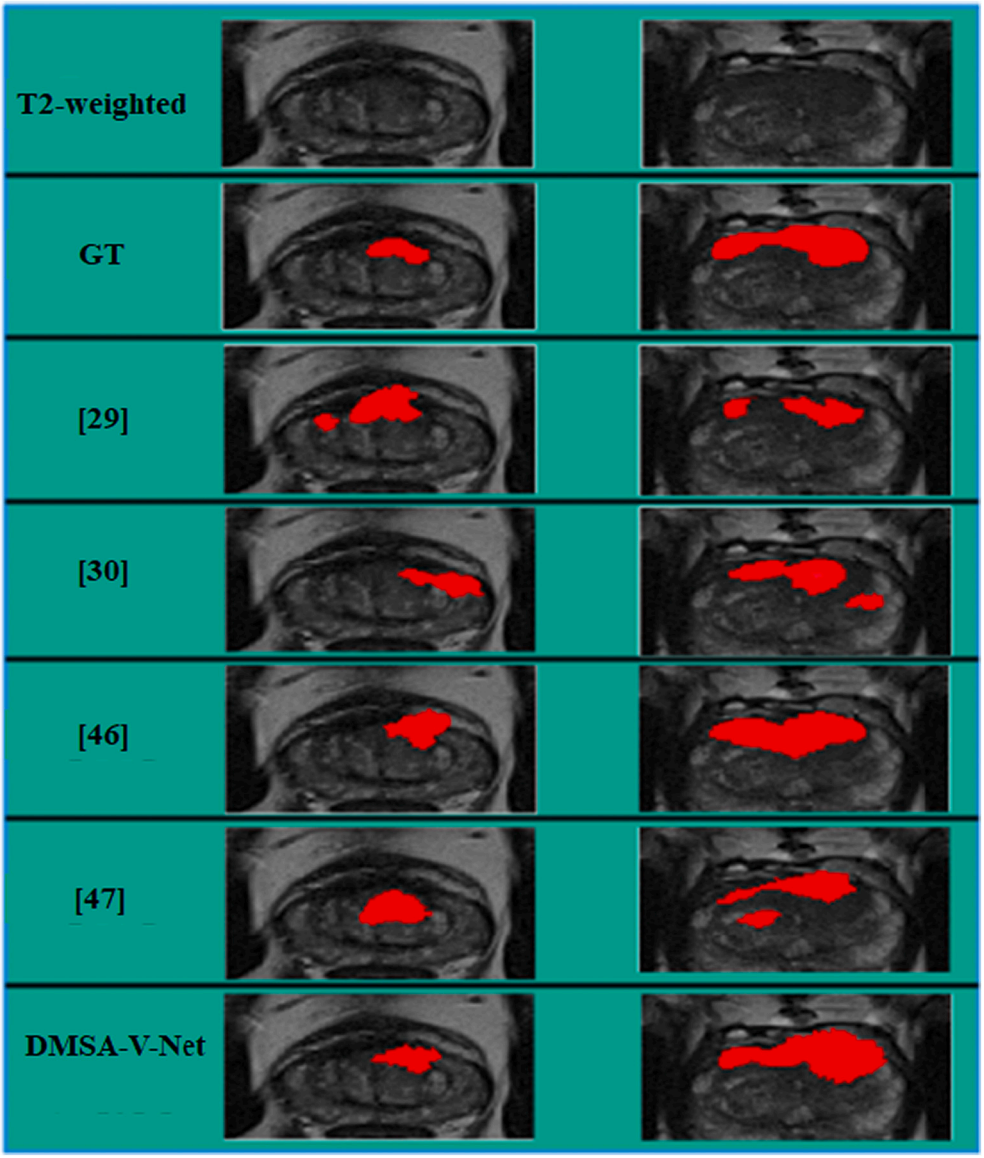

Fig. 9. Examples of the segmented results for different methods. From the first column to the second column, it shows the segmentation results of the 8th slice of the axial view of ProstateX-0105 and the 13th slice of the axial view of ProstateX-0164. From the first row to last row: the image data of T2-weighted, ground truth, results of methods [29,30,46,47] and our DMSA-V-Net.

Table 4
Comparison on recall and false-positive between our proposed DMSA-V-Net and [29,30] for the PCa detection for per patient. The bold values correspond to the best scores for recall and false-positive.

	Recall	False positive
Kiraly [29]	0.8747	0.28
Wang [30]	0.8186	0.21
Proposed	**0.8841**	**0.10**

$$\text{Dice}(P_1 G_1) = 2 \times \frac{|P_1 \cap G_1|}{|P_1| + |G_1|} \tag{4}$$

$$\text{Sensitivity}(P_1 G_1) = \frac{|P_1 \cap G_1|}{|G_1|} \tag{5}$$

$$\text{Specificity}(P_n G_n) = \frac{|P_n \cap G_n|}{|G_n|} \tag{6}$$

where P_1 stands for the segmented region for the PCa lesion, G_1 the PCa lesion region of ground truth, P_n the segmented region of health tissue, G_n the manually segmented region of health tissue and $|P_1 \cap G_1|$the overlap area between P_1 and G_1.

3.3. Quantitative results of ablation studies

To quantify the benefits of DMA-V-Net, we make step-wise comparisons to evaluate the effectiveness of component proposed in Section 2.

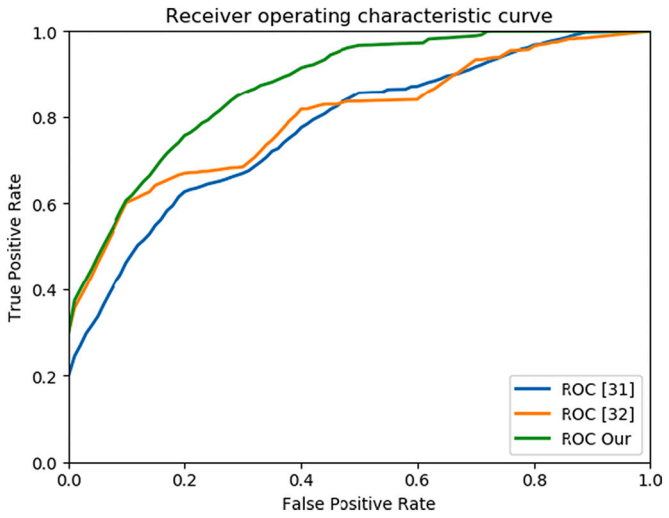

Fig. 10. Slice-level ROC curve of our method and the other two state-of-the-art PCa detection methods.

To investigate the performance of deep multi-scale spatial attention on PCa segmentation, we trained two networks: the original 3D-V-Net and DMSA-V-Net. Experimental results (see Table 1) show that the performance of DMSA-V-Net is better than that of 3D-V-Net, the averaged Dice on all lesions is **0.7014**, **0.6603** for DMSA-V-Net, 3D-V-Net respectively. In addition, we consider another spatial attention mechanism connection manner which is connected to the deeper layer feature maps without connecting to the deepest layer (see Fig. 7. C). The results show that DMSA-V-Net performs better than A-3D-U-Net (see Fig. 7. C), the averaged Dice of A-3D-U-Net on all lesions is 0.6795. The Fig. 8 shows the schematic of our proposed DMSA-V-Net, A-3D-V-Net and 3D-V-Net.

In order to probe the influence of different imaging sequences for PCa lesion segmentation, MR images of each sequence were tested in the DMSA-V-Net for PCa lesion segmentation. As shown in Table 2, based on different sequences (T2-weighted, ADC, DWI and KTrans) of MR images, cross-validation results on some of the four sequences are presented. Obviously, the segmentation performance using the T2-weighted, DWI and ADC sequences is better than the performance using T2-weighted, ADC, DWI and KTrans sequences or the single sequence T2-weighted. The results demonstrate the DMSA-V-Net's effectiveness in aggregation of complementary information from multi sequences of MR images.

3.4. Comparison with the state-of-the-art

To evaluate the performance of our proposed model, we compare our method with state-of-the-art segmentation methods [29,30,46,47]. For a fair comparison, we re-implemented their approaches on our dataset and the detailed comparison result is shown in Table 3. Note that Kiraly et al. [29] utilized similar yet simpler architecture which only consists of several convolutional block in encoder and decoder. Additionally, they used four different modalities including T2, ADC, High B-value and KTrans as input into the network, but we used the following four sequences including DWI, ADC, KTrans and T2-weighted, as these sequences were used as the input of an improved network by Wang [30]. They employed the FCN architecture similar to U-Net with residual blocks to segment PCa lesion. Both of them didn't performed very well in the PCa segmentation task on our dataset. The network of Deeplab v3+ [47] with spatial dilated pooling could capture more multi-scale features, so we compared with them, according to the result of experiment, we get a better performance.

The recently proposed network “nnU-Net” [46] was widely used in medial image segmentation, but this 3D network had never been used for prostate lesion segmentation, so we compared our DMSA-V-Net with the nnU-Net. The results show that our network performed better than the nnU-Net, even it can achieve better results than other networks because both DMSA-V-Net and nnU-Net can harness richer spatial information due to the advantage of 3D network in 3D medical image. The Fig. 9 shows some predicted examples of state-of-the-art methods.

Further, we also adopt the evaluation metrics of recall/false-positive as used in their papers [29,30]. The result is shown in Table 4. It can be seen that, the recalls of ours and [29,30] are 0.8841, 0.8747 and 0.8186 respectively, and ours has a less averaged false positive per patient.

Furthermore, we compared our method with two state-of-the-art PCa detection methods [31,32]. It was important to notice that we finished the slice level PCa classification instead of the patient level, because all the patients were defined as PCa as they have Gleason score greater than or equal to 6. The AUC and ROC curve were used as evaluation metrics. The proposed DMSA-V-Net, [31,32] achieved the slice-level AUC of 0.87, 0.77 and 0.79, respectively. Fig. 10 shows the ROC curve of their performance.

For the characteristics of medical image in high dimension, 3D network architecture achieved a higher accuracy contrast to 2D. Up till now, there is rarely 3D network being applied on PCa lesion segmentation, because the target of PCa lesion is too small and it is difficult to learn the feature information. We used the 3D architecture network with deep spatial attention mechanism to segment the region of PCa and achieved positive results.

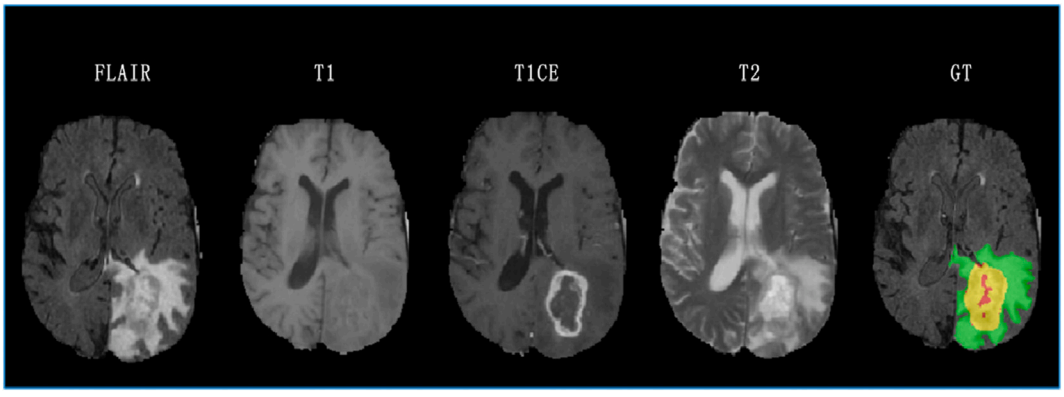

Fig. 11. Example of MR images from different image modalities and the manually marked labels by experts; the first four images from left to right are the sequence of FLAIR, T1, T1CE and T2, the fifth image is the ground truth label where the color is to denote different regions of the tumor: red represents necrosis and non-enhancing, yellow edema, and green enhancing. Others denote the healthy tissues. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Architecture of the proposed DMSA-V-Net for brain tumor segmentation. [] represents the connection with long range; [,] represents the operation of layer concatenation; 'Conv' denotes the operation of convolution; 'DownBlock', 'ConvBlock', 'Sa', 'MS_Feature' and 'UpBlock' have been clearly explained in detail in Section 2.

Encode	Output size	Decoder		Output size
Input	128^3 × 4	UpBlock_1	[ConvBlock5] deep = 1	16^3 × 256
Conv1 × 2	128^3 × 32	MS_Feature1	Sa (ConvBlock_5, ConvBlock_3)	16^3 × 256
DownBlock_1	64^3 × 64	τ_1	[UpBlock_1, MS_Feature1]	16^3 × 512
ConvBlock_1 × 2	64^3 × 64	UpBlock_2	[τ_1] deep = 2	32^3 × 128
Down_Block_2	32^3 × 128	MS_Feature2	Sa (ConvBlock_5, ConvBlock_2)	32^3 × 128
ConvBlock_2 × 2	32^3 × 128	τ_2	[UpBlock_2, MS_Feature2]	32^3 × 256
Down_Block_3	16^3 × 256	UpBlock_3	[τ_2] deep = 3	64^3 × 64
ConvBlock_3 × 2	16^3 × 256	MS_Feature3	Sa (ConvBlock_5, ConvBlock_1)	64^3 × 64
Down_Block_4	8^3 × 512	τ_3	[UpBlock_3, MS_Feature3]	64^3 × 128
ConvBlock_4	8^3 × 512	UpBlock_4	[τ_3] deep = 4	128^3 × 32
ConvBlock_5	8^3 × 512	Concate_1	[UpBlock_4, Conv1]	128^3 × 64
		Output	Conv (Concate_1)	128^3 × 2

Table 6
Performance of brain tumor segmentation with 5-fold cross-validation on the BraTS training dataset.

	Dice	Specificity	Sensitivity
U-Net [48]	0.8413 ± 0.01	0.9978	0.8965 ± 0.02
Kayalibay [49]	0.8794 ± 0.02	0.9979	0.8613 ± 0.03
DMSA-V-Net	**0.8677 ± 0.02**	**0.9979**	**0.8583 ± 0.03**
CHR-U-Net [50]	0.8926 ± 0.01	0.9994	0.8826 ± 0.01

3.5. Extension to brain tumor segmentation

Our proposed DMSA-V-Net is extendable for other tumor segmentation of mp-MR image and proves its strong generalization ability when we apply the Brain tumor Segmentation Challenge (BraTs) 2017 dataset to validate our network. The dataset has 285 brain MR images, including 210 subjects with high grade and 75 subjects with low grade. Each subject, which is associated with 4 modality channels (i.e. T1, T1C, T2, FLARI) in size of 155 × 240 × 240, and the manually marked labels of 4 classes, namely, necrosis, edema, non-enhancing, and enhancing, is shown in Fig. 11. To prove the generalization capability of our DMSA-V-Net, 5-fold cross-validation is carried out on the 285 subjects.

We concatenate all the modality-channel data, and pre-process it by normalizing the brain-tissue intensities of each sequence to zero-mean and unit variance. We do not perform other preprocessing strategy. This experiment aims at showing the extension and generalization ability of DMSA-V-Net which segments the whole tumor from the brain MRI data. Thus, we regard the 4 different tumor labels as the single label of the total tumor region. We extract a 128 × 128 × 128 resolution for each patient. Compared to DMSA-V-Net, we add more convolution filters for brain tumor segmentation for feature enrichment. Detailed network setting is listed in Table 5. The other hyper parameter settings are the same as those in PCa segmentation.

We compare our method with state-of-the-art methods [48–50]. In Table 6, it shows that our network performs well and reaches the state-of-the-art performance. The key factor is that our network uses spatial attention mechanism to harness multi-scale features. Representative segmentation results are depicted in Fig. 12, which explain that the proposed DMSA-V-Net can be competent in brain tumor segmentation, and has a high extension ability.

4. Discussion

In this study, an automated prostate cancer lesion segmentation approach is presented based on 3D-V-Net and deep multi-scale spatial attention mechanism. In the experiments, two different connection manners of multi-scale spatial attention mechanism are compared, we can see that the result of connection manner (Fig. 7-B) outperforms the connection manner (Fig. 7-C) from Table 1, the difference between them is that the former uses the deepest layer feature to form attention map and the latter uses the feature of decoder part to form attention map, it shows that the deepest layer feature is better to capture richer spatial information.

In addition, we compare our method with state-of-the-art segmentation methods, in Table 3, it shows that our DMSA-V-Net and "nnU-Net" can achieve better results than other networks because both DMSA-V-Net and nnU-Net can harness richer spatial information due to the advantage of 3D network in 3D medical image. While the proposed method has taken effect, the accuracy of Dice is still not very high. Although data quality is an important factor affecting the segmentation accuracy, we should still look for better network architecture to improve the performance.

In order to prove the proposed DMSA-V-Net is extendable for other tumor segmentation of mp-MR image. We apply the Brain tumor Segmentation Challenge (BraTs) 2017 dataset to validate our network. From the Table 6, we can see that the 3D U-Net with deep multi-scale spatial attention mechanism has a higher accuracy than the basic 3D U-Net on the evaluation index of Dice, the reason is that the module of deep multi-scale spatial attention mechanism has performed on the glioma which has multi scale size.

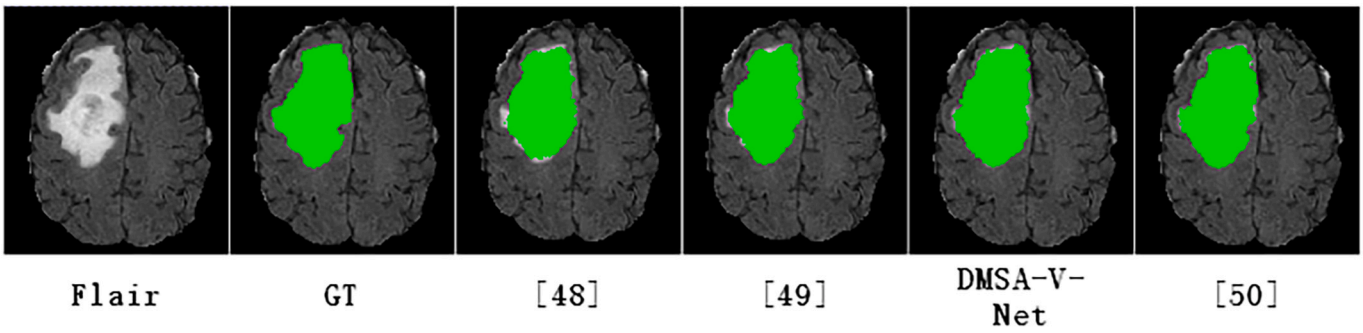

Fig. 12. Examples of the segmented results for different models. From the first column to the last column: the image data of Tra, ground truth, methods [48,49], our proposed DMSA-V-Net and [50].

In future work, there are several problems which can be further explored in order to improve the performance of PCa segmentation. Because the characteristics of PCa is small in volume and complex in image background, the PCa lesion segmentation is still a huge task. We will study generative adversarial network to augment our data and strengthen the analyzing ability of DMSA-V-Net for small targets and complex image background.

5. Conclusions

Existing PCa lesion segmentation methods which focus more on the local features and ignore the global features of the image are based on voxel-level classification. In this work, we propose an end-to-end network architecture (DMSA-V-Net) which focus on both the local and the global features for the automatic segmentation of PCa lesion.

We are the first who employ 3D network architecture in PCa lesion segmentation. Our proposed method can provide a benefit to learn spatial structure features and fuse multi-level contextual information of mp-MRI, therefore increasing the image comprehension ability.

Moreover, in order to address the limitation of 3D network depth caused by a large number of parameters, our proposed deep multi-scale attention patch up the defect of network depth and focus on learning the semantic features of PCa lesion. Final results show that the proposed DMSA-V-Net with deep multi-scale attention certainly has an effect on PCa lesion segmentation.

Funding

This research was financially supported by the National Key R&D program of China (Grant No. 2017YFC0112804), the National Natural Science Foundation of China (No. 81671768) and Science and Technology Commission of Shanghai Municipal (No. 17411952300).

CRediT authorship contribution statement

Enmin Song: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing – original draft. **Jiaosong Long:** Conceptualization, Methodology, Writing – original draft, Investigation. **Guangzhi Ma:** Funding acquisition, Investigation, Writing – review & editing, Software. **Hong Liu:** Resources, Supervision. **Chih-Cheng Hung:** Writing – review & editing. **Renchao Jin:** Software, Validation. **Peijun Wang:** Visualization, Supervision. **Wei Wang:** Visualization, Supervision.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

References

[1] Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics, 2011. CA Cancer J Clin 2011;61(4):212–36.
[2] Maddams J, Utley M, LlerH M. Projections of cancer prevalence in the United Kingdom, 2010-2040. Brit J Cancer 2012;107(7):1195–202.
[3] Chaddad A, Niazi T, Probst S, Bladou F, Anidjar M, et al. Predicting Gleason score of prostate cancer patients using radiomic analysis. Front Oncol 2018;8. https://doi.org/10.3389/fonc.2018.00630.
[4] Ahmad C, Michael K, Tamim N. Multimodal radiomic features for the predicting gleason score of prostate cancer. Cancers 2018;10(8):249.
[5] D'Amico AV, Moul J, Carroll PR, Sun L, Lubeck D, et al. Cancer-specific mortality after surgery or radiation for patients with clinically localized prostate cancer managed during the prostate-specific antigen era. J Clin Oncol 2003;21(11):2163.
[6] Gu Z, Thomas GJ, Shintaku IP, Dorey F, Raitano A, et al. Prostate stem cell antigen (PSCA) expression increases with high Gleason score, advanced stage and bone metastasis in prostate cancer. Oncogene 2000;19(10):1288–96.
[7] Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-aided detection of prostate cancer in MRI. IEEE T Med Imaging 2014;33:1083–92. https://doi.org/10.1109/TMI.2014.2303821.
[8] Peng Y, Jiang Y, Yang C, Brown JB, Antic T, et al. Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score–a computer-aided diagnosis development study. Radiology 2013;267(3):787–96.
[9] Duc F, Harini V, Andreas W, Tatsuo G, Kazuhiro M, et al. Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images. Proc Natl Acad Sci U S A 2015;112(46):6265–73.
[10] Turkbey B, Choyke PL. Multiparametric MRI and prostate cancer diagnosis and risk stratification. Curr Opin Urol 2012;22(4):310.
[11] Niaf E, Rouvière O, Mège-Lechevallier F, Bratan F, Lartizien C. Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. Phys Med Biol 2012;57(12):3833.
[12] Liu P, Wang S, Turkbey B, Grant K, Pinto P, et al. A prostate cancer computer-aided diagnosis system using multimodal magnetic resonance imaging and targeted biopsy labels. Proc Spie 2013;8670(4). 86701G.
[13] Turkbey B, Xu S, Kruecker J, Locklin J, Pang Y, et al. Documenting the location of prostate biopsies with image fusion. BJU Int 2011;107(1):53–7.
[14] Wang S, Burtt K, Turkbey B, Choyke P, Summers RM. Computer aided-diagnosis of prostate cancer on multiparametric MRI: a technical review of current research. Biomed Res Int 2016;2014(13). 789561.
[15] Lukasz M, Jansen JFA, Louisa B, Hebert Alberto V, Oguz A, et al. Anatomic segmentation improves prostate cancer detection with artificial neural networks analysis of 1H magnetic resonance spectroscopic imaging.J. Magn Reson Imaging 2015;40(6):1414–21.
[16] Ozer S, Haider MA, Langer DL, THVD Kwast, Evans AJ, et al. Prostate cancer localization with multispectral MRI based on relevance vector machines. 2009-01-01. p. 2009.
[17] Puech P, Betrouni N, Makni N, Dewalle AS, Villers A, et al. Computer-assisted diagnosis of prostate cancer using DCE-MRI data: design, implementation and preliminary results. Int J Comput Assist Radiol Surg 2009;4(1):1–10.
[18] APC Vos, Hambrock T, Barentsz JO, Huisman HJ. Combining T2-weighted with dynamic MR images for computerized classification of prostate lesions. 2008-01-01. p. 2008.
[19] Vos P, Hambrock T, Hulsbergen-Van-De-Kaa C, Futterer J, Barentsz J, et al. Computerized analysis of prostate lesions in the peripheral zone using dynamic contrast enhanced MRI. Med Phys 2008;35(3):888–99.
[20] Litjens GJS, Vos PC, Barentsz JO, Karssemeijer N, Huisman HJ. Automatic computer aided detection of abnormalities in multi-parametric prostate MRI. Proc SPIE - Int Soc Opt Eng 2011;7963(1):554–61.
[21] Stojanov D, Koceski S. Topological MRI prostate segmentation method. 2014-01-01. p. 2014.
[22] Klein S, Van-Der-Heide U, Lips I, Van-Vulpen M, Staring M, et al. Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Med Phys 2008;35(4):1407–17.
[23] Chan I, Haker S, Zhang J, Zou KH, Maier SE, et al. Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier. Med Phys 2003;30(9):2390–8.
[24] Burges C. A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 1998;2:121–67. https://doi.org/10.1023/A:1009715923555.
[25] Tiwari P, Viswanath S, Kurhanewicz J, Sridhar A, Madabhushi A. Multimodal wavelet embedding representation for data combination (MaWERiC): integrating magnetic resonance imaging and spectroscopy for prostate cancer detection. NMR Biomed 2012;25(4):607–19.
[26] Artan Y, Haider MA, Langer DL, Kwast THVD, Evans AJ, et al. Prostate Cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields. IEEE Trans Image Proc A Public IEEE Signal Proc Soc 2010;19(9):2444–55.
[27] Lafferty JD, Mccallum A, Pereira FCN. Conditional random fields: probabilistic models for segmenting and labeling sequence data. Proc ICML 2001;3(2):282–9.
[28] Schmidhuber J. Deep learning in neural networks: An overview. Neural Netw 2015;61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003.
[29] Kiraly AP, Nader CA, Tuysuzoglu A, Grimm R, Kiefer B, et al. Deep convolutional encoder-decoders for prostate cancer detection and classification. 2017-01-01. p. 2017.
[30] Y. W, B. Z, D. G, J. W. Fully convolutional neural networks for prostate cancer detection using multi-parametric magnetic resonance images: an initial investigation2018. In: 24th international conference on pattern recognition (ICPR), 20183814–3819; 2018-01-01. https://doi.org/10.1109/ICPR.2018.8545754.
[31] Abbasi AA, Hussain L, Awan IA, Abbasi I, Majid A, et al. Detecting prostate cancer using deep learning convolution neural network with transfer learning approach. Cogn Neurodyn. 2020;14(4):523–33. https://doi.org/10.1007/s11571-020-09587-5.
[32] Yoo S., Gujrathi I., Haider M.A., Khalvati F. Prostate cancer detection using deep convolutional neural networks. Scientific Reports 2019; 9:Articale Number 19518. doi: 10.1038/s41598-019-55972-4.
[33] Peter QL, Alessandro G, Steve P, Thomas T, Sharon EC. Model-free prostate cancer segmentation from dynamic contrast-enhanced MRI with recurrent convolutional networks: a feasibility study. Comput Med Imaging Graph 2019;75:14–23.
[34] Cao R, Zhong X, Shakeri S, Mohammadian Bajgiran A, Afshari Mirak S, et al. Prostate cancer detection and segmentation in multi-parametric MRI via CNN and conditional random field. 2019. https://doi.org/10.1109/ISBI.2019.8759584.
[35] L. C, H. Z, J. X, L. N, J. S et al T. C. SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning2017 IEEE conference on computer vision and pattern recognition (CVPR), 20176298–6306. 2017-01-01. https://doi.org/10.1109/CVPR.2017.667.

[36] Woo S, Park J, Lee J, Kweon IS. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. CBAM: convolutional block attention module. Cham: Springer International Publishing; 2018-01-01. p. 3–19. 2018.
[37] Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, et al. Attention gated networks: learning to leverage salient regions in medical Images. Med Image Anal 2019;53. https://doi.org/10.1016/j.media.2019.01.012.
[38] Jin Q, Meng Z, Sun C, Wei L, Su R. RA-UNet: a hybrid deep attention-aware network to extract liver and tumor in CT scans. 2018.
[39] Guokai Z, Weigang W, Dinghao Y, Jihao L, Jianwei L. A Bi-attention adversarial network for prostate cancer segmentation. IEEE Access 2019;PP(99):1.
[40] Zhang Z, Zhang X, Peng C, Xue X, Sun J. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. ExFuse: enhancing feature fusion for semantic segmentation. Cham: Springer International Publishing; 2018-01-01. p. 273–88. 2018.
[41] He K, Zhang X, Ren S, Jian S. Deep residual learning for image recognition. 2016-01-01. p. 2016.
[42] Ji Y, Zhang H, Jonathan Wu QM. Salient object detection via multi-scale attention CNN. Neurocomputing 2018;322:130–40. https://doi.org/10.1016/j.neucom.2018.09.061.
[43] Roy AG, Navab N, Wachinger C, Concurrent Spatial and Channel "Squeeze & Excitation" in Fully Convolutional Networks, Cham. Medical image computing and computer assisted intervention – MICCAI 2018. Springer International Publishing; 2018. p. 421–9.
[44] Milletari F, Navab N, Ahmadi SA. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. 2016-01-01. p. 2016.
[45] Kingma DP, Ba J. Adam: A method for stochastic optimization.Computer science. 2014.
[46] Isensee F, Jager PF, Kohl SAA, Petersen J, Maier-Hein K. Automated design of deep learning methods for biomedical image segmentation.arXiv: Computer vision and pattern recognition. 2019.
[47] Chen L, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. ECCV; 2018.
[48] Ronneberger O, Fischer P, Brox T. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. U-net: convolutional networks for biomedical image segmentation. Cham: Springer International Publishing; 2015-01-01. p. 234–41. 2015.
[49] Kayalibay B, Jensen G, Smagt PVD. CNN-based segmentation of medical imaging data. 2017 [Available].
[50] Long J, Ma G, Liu H, et al. Cascaded hybrid residual U-net for glioma segmentation [J]. Multimed Tools Appl 2020;79(33–34).