3D EAGAN: 3D edge-aware attention generative adversarial network for prostate segmentation in transrectal ultrasound images

Mengqing Liu[1,2]^, Xiao Shao[3]^, Liping Jiang[4], Kaizhi Wu[5]

[1]School of Computer and Information Engineering, Nantong Institute of Technology, Nantong, China; [2]School of Information Engineering, Nanchang Hangkong University, Nanchang, China; [3]School of Computer Science, Nanjing University of Information Science and Technology, Nanjing, China; [4]Department of Ultrasound Medicine, The First Affiliated Hospital of Nanchang University, Nanchang, China; [5]School of Information Engineering, Nanchang Hangkong University, Nanchang, China

Contributions: (I) Conception and design: M Liu, X Shao; (II) Administrative support: M Liu, K Wu; (III) Provision of study materials or patients: L Jiang, K Wu; (IV) Collection and assembly of data: M Liu, L Jiang; (V) Data analysis and interpretation: M Liu, X Shao; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Kaizhi Wu, PhD. School of Information Engineering, Nanchang Hangkong University, 696 Fenghenan Street, Honggutan District, Nanchang, China. Email: kaizhiwu@163.com.

Background: The segmentation of prostates from transrectal ultrasound (TRUS) images is a critical step in the diagnosis and treatment of prostate cancer. Nevertheless, the manual segmentation performed by physicians is a time-consuming and laborious task. To address this challenge, there is a pressing need to develop computerized algorithms capable of autonomously segmenting prostates from TRUS images, which sets a direction and form for future development. However, automatic prostate segmentation in TRUS images has always been a challenging problem since prostates in TRUS images have ambiguous boundaries and inhomogeneous intensity distribution. Although many prostate segmentation methods have been proposed, they still need to be improved due to the lack of sensibility to edge information. Consequently, the objective of this study is to devise a highly effective prostate segmentation method that overcomes these limitations and achieves accurate segmentation of prostates in TRUS images.

Methods: A three-dimensional (3D) edge-aware attention generative adversarial network (3D EAGAN)-based prostate segmentation method is proposed in this paper, which consists of an edge-aware segmentation network (EASNet) that performs the prostate segmentation and a discriminator network that distinguishes predicted prostates from real prostates. The proposed EASNet is composed of an encoder-decoder-based U-Net backbone network, a detail compensation module (DCM), four 3D spatial and channel attention modules (3D SCAM), an edge enhancement module (EEM), and a global feature extractor (GFE). The DCM is proposed to compensate for the loss of detailed information caused by the down-sampling process of the encoder. The features of the DCM are selectively enhanced by the 3D spatial and channel attention module. Furthermore, an EEM is proposed to guide shallow layers in the EASNet to focus on contour and edge information in prostates. Finally, features from shallow layers and hierarchical features from the decoder module are fused through the GFE to predict the segmentation prostates.

Results: The proposed method is evaluated on our TRUS image dataset and the open-source µRegPro dataset. Specifically, experimental results on two datasets show that the proposed method significantly improved the average segmentation Dice score from 85.33% to 90.06%, Jaccard score from 76.09% to 84.11%, Hausdorff distance (HD) score from 8.59 to 4.58 mm, Precision score from 86.48% to 90.58%, and

---

^ ORCID: Mengqing Liu, 0000-0002-0491-9108; Xiao Shao, 0000-0003-1194-240X.

Recall score from 84.79% to 89.24%.

Conclusions: A novel 3D EAGAN-based prostate segmentation method is proposed. The proposed method consists of an EASNet and a discriminator network. Experimental results demonstrate that the proposed method has achieved satisfactory results on 3D TRUS image segmentation for prostates.

Keywords: Prostate segmentation; generative adversarial network; edge-aware segmentation network (EASNet); detail compensation module (DCM); edge enhancement module (EEM)

Introduction

Background

Prostate cancer is one of the most commonly diagnosed cancers in men (1). Since early-stage prostate cancer can be effectively controlled, early detection and interventions are crucial to the diagnosis and treatment planning of prostate diseases (2). Conventionally, experienced physicians have manually segmented prostate imagery by visually inspecting transrectal ultrasound (TRUS) images, a process that is time-consuming and laborious, relying heavily on the doctor's expertise. Hence, the development of computer algorithms capable of automatically performing accurate prostate segmentation from TRUS images is of significant value for improving medical practices, alleviating the workload of physicians, and enhancing the quality of patient care. With this context, accurate segmentation-the precise identification and isolation of prostate tissue boundaries through advanced techniques is particularly critical. Precise segmentation plays a vital role in the diagnosis, treatment planning (3), biopsy needle placement (4), and cryotherapy (5) of prostate cancer as it can enhance clinical outcomes and reduce unnecessary treatments and interventions due to inaccuracies.

Traditionally, prostate segmentation methods have relied on hand-crafted features (6-18), such as shape statistics, to differentiate between healthy tissue and cancerous areas. Yet, these manually extracted features are low-level semantic representations and often fail to accurately characterize the complexities of actual prostate tissues, inherently limiting their effectiveness and potentially leading to missed detection or misclassification of critical areas. The advent of deep convolutional neural networks (DCNN) (19-22) has revolutionized the field of semantic segmentation in recent years. These methods, powered by DCNN, can automatically learn and recognize complex patterns at the pixel level, which enables them to assign categories to each pixel with a high degree of precision. Long *et al.* (23) proposed the fully convolutional network (FCN)-based method for image segmentation tasks, which is an end-to-end architecture to automatically classify images into different classes. Ronneberger *et al.* (24) proposed an encoder-decoder-based U-Net architecture for medical semantic segmentation, which utilizes the skip connection to integrate low-level features extracted by the encoder into the decoder. Inspired by these novel architectures, a large number of DCNN-based prostate segmentation methods (25-29) were proposed.

Although these methods achieved great improvements over traditional methods, further improvements are still lacking. Different from other semantic segmentation tasks (e.g., indoor scenes and street scenes), TRUS images have weak boundaries, low signal-to-noise ratio, and large differences in contrast and resolution. Specifically, TRUS images have ambiguous boundaries caused by poor contrast between the prostate and surrounding tissues. Hence, current methods that adopt semantic segmentation models (e.g., FCN and U-Net) to segment prostates would lack sensitivity to ambiguous boundaries and inhomogeneous intensity distribution of prostates. Therefore, it is quite challenging to accurately segment the boundary of prostates.

In this paper, a novel three-dimensional (3D) edge-aware attention generative adversarial network (3D EAGAN)-based prostate segmentation method is proposed. The proposed method consists of an edge-aware segmentation network (EASNet) and a discriminator network. The EASNet aims to produce prostate segmentation results and the discriminator is designed to distinguish the predicted prostates from the ground-truth prostates. The EASNet is composed of an encoder-decoder-based U-Net backbone network, a detail compensation module (DCM), four 3D

spatial and channel attention modules (3D SCAM), an edge enhancement module (EEM), and a global feature extractor (GFE). Since the down-sampling of the encoder in EASNet would cause information loss, the DCM is proposed to introduce rich detail contextual information to the encoder, which is pre-trained on a large-scale medical data set 3DSeg-8 (30) to learn rich details and texture information. Since the DCM contains some irrelevant features with prostates, the 3D SCAM is proposed to selectively utilize the features that can reflect more prostate details from the channel and spatial dimensions. To further assist the EASNet in generating more accurate prostate margins, an EEM is proposed to guide shallow layers in the EASNet to focus on contour and edge information in prostates. Finally, the enhanced low-level features from the encoder and hierarchical features from the decoder are fused to the GFE to obtain the final segmentation results. In summary, this paper has the following main contributions:

(I) A novel framework 3D EAGAN for improving prostate segmentation is proposed, which adopts a DCM to learn rich detail information of prostates and an EEM to guide the network focus on edge information of prostates.

(II) Since prostates in TRUS images have ambiguous boundaries, an EEM is introduced to further guide shallow layers in the encoder to focus on the prostate edges without adding extra computation cost during the inference process.

(III) A 3D spatial and channel attention module is proposed to adaptively enhance the features that can reflect prostate details by considering interdependencies among channel and spatial dimensions.

Related work

Traditional prostate segmentation methods

According to the way of feature extraction, the segmentation method can be divided into hand-crafted-based prostate segmentation methods and deep learning-based ones. Traditional hand-crafted-based prostate segmentation methods utilize carefully designed hand-crafted features to detect the shape and edge of prostates. The shape statistics belong to the mainstream of traditional segmentation methods. Ladak *et al.* (6) proposed a semi-automatic segmentation method based on 2D ultrasound images, which first utilized shape statistics to detect prostates. To detect the edge of prostates, Pathak *et al.* (7)

developed an edge detection algorithm to depict the prostate edges. Shen *et al.* (8) employed Gabor filter sets to characterize prostate boundaries and reconstructed Gabor features to guide deformable segmentation. Yan *et al.* (9) learned the shape statistical information of the local domain to segment prostates. Santiago *et al.* (10) employed an active shape model (ASM) to improve the robustness in the presence of outliers. Although these methods have achieved more promising segmentation performance than traditional manual segmentation methods, these methods are performed on 2D TRUS images, which would lack the correlation between different image slices and 3D image context.

To effectively enhance the correlation between different TRUS image slices and 3D image context, many 3D prostate segmentation methods were proposed. Ghanei *et al.* (11) proposed a 3D deformable surface model to segment ultrasound images. Wang *et al.* (12) proposed two semi-automatic segmentation methods by using 2D ultrasound images to achieve 3D prostate. Hu *et al.* (13) employed a semi-automatic segmentation by using an efficient deformable mesh. Gong *et al.* (14) used deformable models for the automatic segmentation of prostates. Qiu *et al.* (15) proposed a novel globally optimized method to segment 3D prostate images. Previously methods utilized shape information of prostates to enhance the segmentation performance, but the shape of prostates varies greatly, which would lose the specificity of individual cases and lead to a decrease in prediction accuracy. Different from these shape statistics-based methods, many other prostate segmentation methods treat the segmentation task as a classification task. Ghose *et al.* (16) applied the principal component analysis and random forest classification in machine learning to implement prostate segmentation. Zhan *et al.* (17) proposed a deformable model for automatic prostate segmentation by shape and texture statistics. To augment training samples, Yang *et al.* (18) proposed a 3D TRUS image segmentation method for the prostate based on a patch-based feature learning framework. Although these hand-crafted-based methods have achieved promising prediction accuracy, the hand-crafted features are shallow and not capable of obtaining high-level semantic information in images, resulting in the lack of prostate boundary information.

Deep learning-based prostate segmentation methods

Recently, deep learning technology (19-24) has achieved great success in various image processing tasks, including image classification, image enhancement, and semantic

segmentation. Benefiting from features automatically learned by convolutional neural networks (CNNs), many deep learning-based prostate segmentation methods have been proposed. Ghavami *et al.* (25) employed a U-Net-based method for automatic prostate segmentation, which replaces the convolutional layers in the original U-Net with residual network unit blocks to enhance the feature representation ability. To solve the problem of information loss in traditional shape models, Yang *et al.* (26) used the recurrent neural network to learn the shape prior information of prostates. Wang *et al.* (27) proposed a 3D deep neural network-based prostate segmentation method, which first utilized the 3D feature pyramid network (FPN) to extract multi-level features. Then, an attention mechanism was proposed to adapt and fuse different features and pay attention to the prostate region. Lei *et al.* (28) employed the V-Net-based backbone network to extract primary features, and the 3D supervision mechanism was integrated into the network training to speed up the network convergence. Pellicer-Valero *et al.* (29) proposed a Densenet-resnet-based 3D prostate segmentation method. In addition, to improve the robustness of the network, some techniques, e.g., deep supervision, checkpoint ensembling, and neural resolution enhancement, are also integrated into the network training process.

Currently, generative adversarial networks (GANs) have made impressive progress in many computer vision tasks (31,32). Generally, GAN is composed of two parts: a generator network G and a discriminator network D. G aims to generate more real samples and D is designed to distinguish the real samples from the generated samples by G. The training stage of G and D is to optimize the minimax game by using the objective function,

$$\min_G \max_D F(D,G) = E_{x \sim P_{data}}\left[\log\left(D(x)\right)\right] + E_{y \sim P_y}\left[\log\left(1 - D\left(G(y)\right)\right)\right] \quad [1]$$

where G denotes the generator network; D denotes the discriminator network; P_y represents the distribution of the noise; P_{data} represents the distribution of real data. To further enhance the prostate segmentation performance, methods (33,34) adopted the GAN for the prostate segmentation task. Dong *et al.* (33) utilized the adversarial training strategy for the prostate segmentation task. The generator is composed of a set of U-Nets and the discriminator is the FCN. Wang *et al.* (34) employed the GAN to automatically segment prostates, which consists of a Densenet-based generator and a multi-scale discriminator.

Methods

Due to ambiguous boundaries and inhomogeneous intensity distribution of prostates in TRUS images, segmenting prostates from TRUS images is still a challenging task. To effectively segment prostates from TRUS images, a 3D EAGAN-based prostate segmentation method is proposed in this paper. The architecture of the proposed 3D EAGAN is shown in *Figure 1*, it consists of two parts: an edge-aware segmentation network and a discriminator network. The edge-aware segmentation network aims to segment more accurate prostates to fool the discriminator, and the discriminator network is expected to distinguish predicted prostates from real prostates. Then, the edge-aware segmentation network and the discriminator network are described in detail.

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the Ethics Committee of the Western University. Participants provided informed consent before taking part in the study.

Edge-aware segmentation network

The network architecture of the proposed edge-aware segmentation network is shown in *Figure 1*, which is composed of an encoder-decoder-based U-Net, a DCM, four 3D SCAM, an EEM, and a GFE. For the input TRUS image I, the encoder decoder-based U-Net is first used to extract primary features from the input. To compensate for the loss of detail information during the U-Net forward propagation, the input TRUS image I is also fed to the DCM to introduce rich local detail information to the encoder module. This process is shown as follows:

$$E_i = G_U(I), \quad i \in \{1,2,3,4\} \quad [2]$$

$$SQ_i = G_{DCM}(I), \quad i \in \{1,2,3,4\} \quad [3]$$

where I represents the input TRUS image, $G_U()$ and $G_{DCM}()$ represent the U-Net and DCM respectively, E_i and SQ_i ($i \in \{1,2,3,4\}$) represent the multi-level features extracted by the U-Net and the DCM respectively.

To reduce the computational complexity, $1 \times 1 \times 1$ convolutional layers Sq-i ($i \in \{1,2,3,4\}$) are adopted to the output of the DCM to squeeze the channel dimensions. Then, the extracted detail information is refined by the
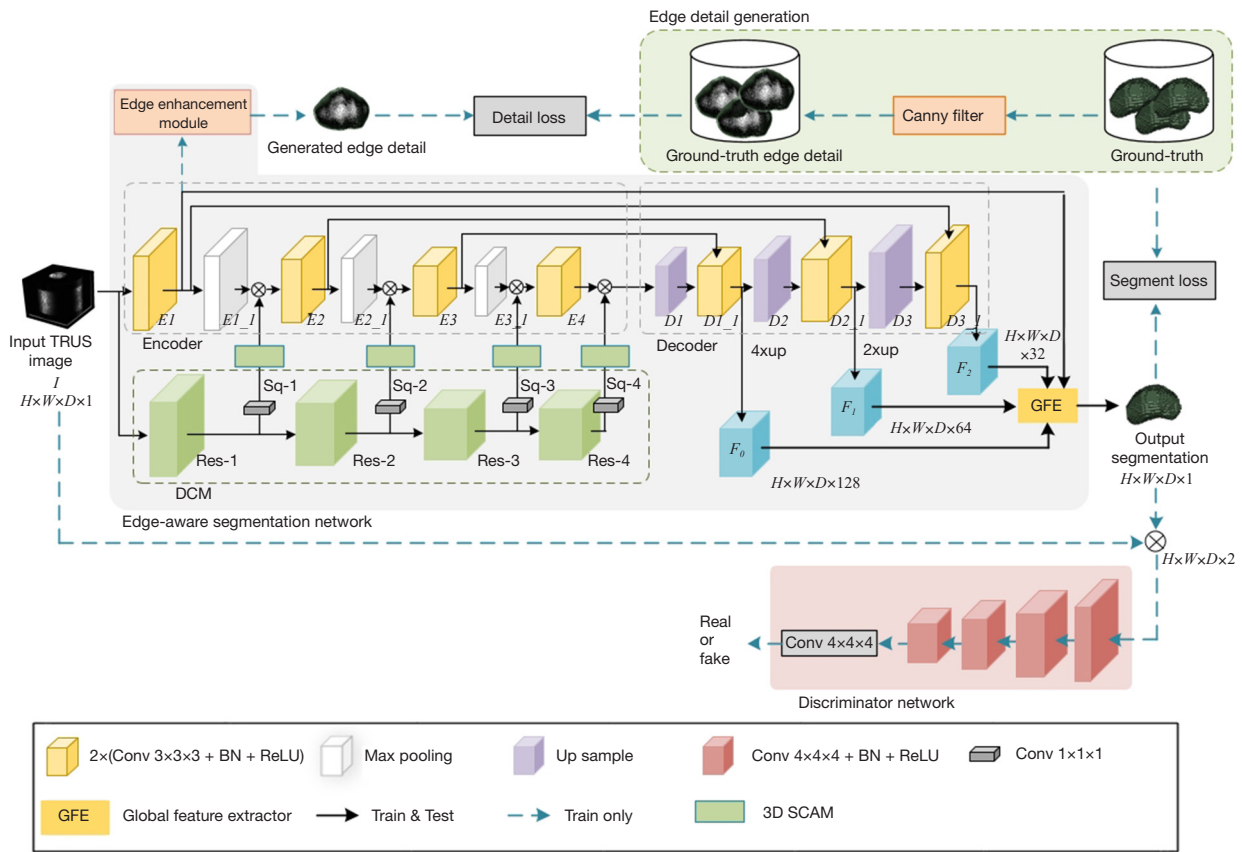
Figure 1 The framework of the proposed method. "H×W×D×1" denotes the dimension of the input image. "H" represents height, "W" represents width, "D" represents depth. TRUS, transrectal ultrasound; DCM, detail compensation module; GFE, global feature extractor; BN, batch normalization; ReLU, rectified linear unit; 3D SCAM, three-dimensional spatial and channel attention module.

3D SCAM, and feature maps from the encoder module and 3D SCAM at each resolution stage are fused through concatenation operation. This process is shown as follows:

$$SQ_i^{'} = G_{SCAM}\left(Conv\left(SQ_i\right)\right), \quad i \in \{1,2,3,4\} \tag{4}$$

$$Q_i = E_i \otimes SQ_i^{'}, \quad i \in \{1,2,3,4\} \tag{5}$$

where $Conv()$ represents the convolution layer with the kernel size of 1×1×1, $G_{SCAM}()$ represents the 3D SCAM, $\otimes$ represents the concatenation operation, $Q_i\left(i \in \{1,2,3,4\}\right)$ represents the multi-level features after fusion. In addition, an EEM is utilized to guide the shallow layers in the encoder to focus on contour and edge information. The working mechanism of the EEM is shown as follows:

$$E_1^{'} = G_{EEM}\left(E_1\right) \tag{6}$$

where $G_{EEM}()$ represents the EEM. Finally, the GFE is used

to integrate hierarchical features from the decoder and shallow layers from the encoder to generate more realistic prostates. Finally, a GFE is used to fuse hierarchical features from the decoder and shallow features from the encoder to generate more realistic prostate images. This process is shown as follows:

$$O = G_{GFE}\left(E_1, F_0, F_1, F_2\right) \tag{7}$$

where E_1 represents the shallow features of the encoder part, F_i ($i \in \{0,1,2\}$) represents the hierarchical features of the decoder, and $G_{GFE}()$ represents the GFE.

Encoder-decoder-based U-Net

Since the U-Net structure (24) has shown strong feature representation ability in the image segmentation task, the encoder-decoder-based U-Net structure is adopted

as the backbone network. The encoder module aims to extract low-level and high-level semantic features from TRUS images, while the decoder module is designed to progressively combine contextual information from different levels to generate the output segmentation.

More specifically, the encoder module is composed of four Conv-BN-ReLU (CBR) modules and three max-pooling layers. Each CBR module consists of two groups of 3×3×3 convolutional layers followed by a batch normalization (BN) layer and a ReLU activation. Max pooling layers are applied to gradually down-sample the resolutions of feature maps to half to reduce the computational complexity and improve the inference speed of the network. The decoder module consists of three CBR modules and three up-sampling layers. The up-sampling layers are used to gradually recover the resolutions of feature maps to match the resolutions of the input image.

DCM

Due to the requirement of fast inference and low resource consumption, the U-Net structure utilizes the down-sampling operation in the encoder to progressively decrease the resolutions of the input images. However, the down-sampling operation would cause the loss of detailed contextual information. Since prostates in TRUS images have ambiguous boundaries, the loss of detailed contextual information would inevitably cause the degradation of segmentation ability.

Hence, the transfer learning-based DCM is proposed, and it is built on a ResNet-34 (35) pre-trained on the large-scale medical dataset 3DSeg-8 (30). The network-based transfer learning technique aims to solve the problem of limited training data (36). Its underlying assumption is that the internal layers of a CNN are not specific to a particular task, e.g., the shallow layers from an image classification CNN are sensitive to the detail information (e.g., edge and texture features). Specifically, the network-based transfer learning technique usually pre-trains a network on the source task T_s, and features learned from the source task are transferred to the target task T_s to enhance the robustness of the network. The architecture of the DCM is shown in *Figure 1*. First, the ResNet-34 is trained on the 3DSeg-8 to learn the features of different organs. Then, the internal layers Res-i ($i\in\{1,2,3,4\}$) of the pre-trained ResNet-34 are transferred to the prostate segmentation task to learn more abundant detailed contextual information.

3D spatial and channel attention module

Recently, attention mechanisms have been widely used in many computer vision tasks, which can effectively boost the performance of DCNNs. In medical segmentation tasks, many methods have applied attention mechanisms to make the network focus on target regions. Since the features extracted from DCM contain rich detailed information, but also include some non-prostate features. To further enhance and refine more important features, attention mechanisms are introduced to adaptive filter out non-prostate features and focus on important features by exploring the relationship of features between the spatial and channel dimensions. The architecture of the proposed 3D SCAM is shown in *Figure 2*, and it is composed of a spatial attention module and a channel attention module.

Specifically, the squeezed feature maps N_i from Res-i ($i\in\{1,2,3,4\}$) in DCM are fed into the 3D SCAM. Given the input feature maps $N_i\in R^{H\times W\times D\times C}$ ($i\in\{1,2,3,4\}$), they are first fed into the spatial attention module and the channel attention module to generate the weight scores, respectively. The spatial attention module consists of a convolutional layer with a kernel size of 7×7×7 and a Sigmoid layer. The 7×7×7 convolutional layer is used to calculate the spatial weight scores W_s from input feature maps N_i. Then, the Sigmoid layer is adapted to constrain the weight scores W_s to be between [0, 1] to obtain W_s'. Finally, the calculated spatial weight scores W_s' are multiplied by the input feature maps N_i to obtain the adaptive feature maps N_i^s. The working mechanism of the spatial attention module is shown as follows,

$$\begin{cases} W_s = Conv_1(N_i), & i \in \{1,2,3,4\} \\ W_s' = \alpha(W_s) \\ N_i^s = W_s' \times N_i, & i \in \{1,2,3,4\} \end{cases} \quad [8]$$

where $Conv_1()$ denotes the convolutional layer with kernel size 7×7×7. $\alpha()$ denotes the Sigmoid layer. × denotes the element-wise multiply. For the channel attention module, it consists of a global max pooling (GMP) layer, a global average pooling (GAP) layer, a fully connected (FC) layer, and a Sigmoid layer. Given the input feature maps N_i, the GMP and GAP layers are first used to squeeze the features maps to the $N_i'\in R^{H\times W\times D\times 1}$ along the channel dimensions. Then, the feature maps N_i' are fed into the FC layer to calculate the channel weight scores W_c. The Sigmoid layer is adapted to constrain the W_c to be between [0, 1] to obtain W_c'. Finally, the input feature maps N_i are calculated with
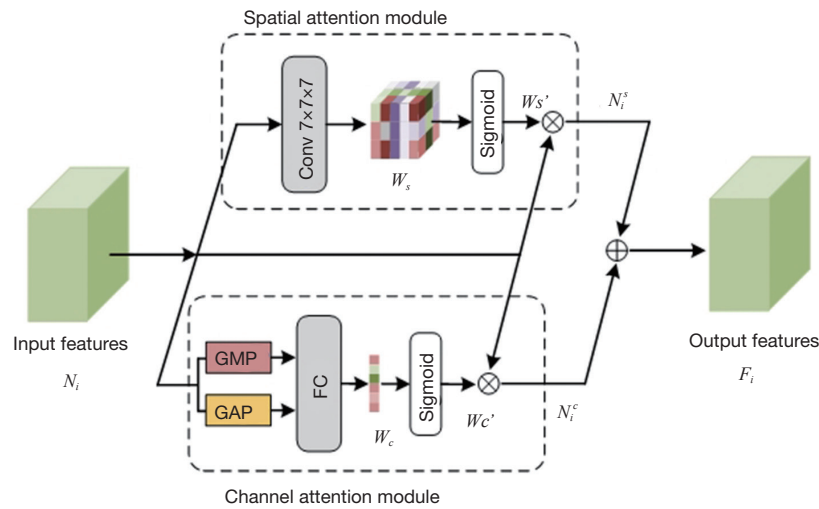
Figure 2 The network structure of the 3D SCAM. GMP, global max pooling layer; GAP, global average pooling layer; FC, fully connected layer; 3D SCAM, three-dimensional spatial and channel attention module.

the channel weight scores $W_c^{'}$ to obtain the adaptive feature maps N_i^c. The working mechanism of the channel attention module is shown as follows,

$$\begin{cases} W_c = GMP(N_i) \otimes GAP(N_i), & i \in \{1,2,3,4\} \\ W_c^{'} = \alpha(W_c) \\ N_i^s = W_c^{'} \times N_i, & i \in \{1,2,3,4\} \end{cases} \quad [9]$$

where $GMP()$ and $GAP()$ denote the GMP layer and GAP layer, respectively. $\otimes$ denotes the concatenation operation. $\alpha()$ denotes the Sigmoid layer. $\times$ denotes the element-wise multiply. Finally, feature maps N_i^s and N_i^c ($i \in \{1,2,3,4\}$) are fused through the element-wise addition operation. This process can be described as,

$$F_i = N_i^s + N_i^c, \quad i \in \{1,2,3,4\} \quad [10]$$

where + denotes the element-wise addition operation.

Edge generation guidance of low-level features

Since prostates in TRUS images have ambiguous structure boundaries, current methods fail to accurately predict the structure boundary of prostates. To enhance the sensitivity to edge details, an EEM is proposed to guide shallow layers in the encoder module to focus on the edge details of prostates.

To accurately obtain the ground-truth edge maps of prostates, two ways are tested to obtain the edge maps by using the Canny algorithm (37). First, edge maps are

obtained from TRUS images. Second, edge maps are obtained from the ground-truth semantic segmentation prostate images. To intuitively show the difference between the edge maps calculated from TRUS images and the ground-truth semantic segmentation prostate images, the visualization of the different edge maps is shown in *Figure 3*. It can be observed that edge maps directly achieved from TRUS images contain useless information. On the contrary, edge maps obtained from the ground-truth semantic segmentation prostate images can accurately reflect the edge of prostates. Hence, the ground-truth edge maps are generated from the ground-truth semantic segmentation prostate images by using the Canny algorithm.

With the generated ground-truth edge maps, the EEM is used to guide the low-level layers in the encoder module to focus on learning the prostate boundary. The EEM is composed of a 3×3×3 convolutional layer for feature extraction and a 1×1×1 convolutional layer to reduce the channel dimension. Finally, the learned edge features are fused with the hierarchical features in the decoder module for the final prediction.

GFE

To obtain more accurate segmentation performance and achieve a precise prostate edge, a GFE is proposed. The architecture of the proposed GFE is shown in *Figure 4*. Specifically, multi-layer features F_i ($i \in \{0,1,2\}$) are first fused through the concatenation operation. The 1×1×1
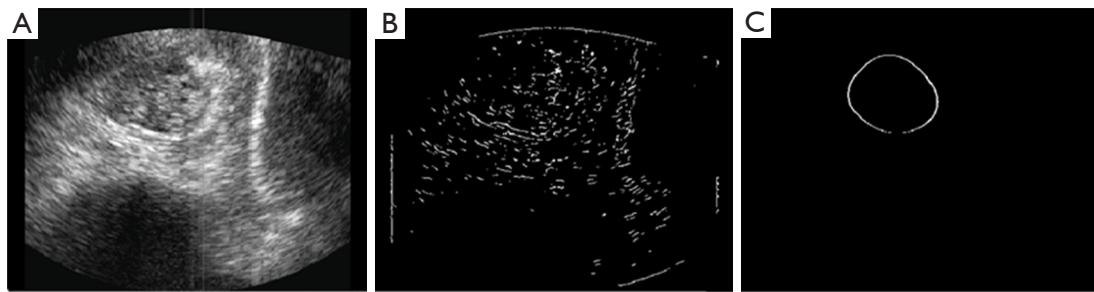
Figure 3 Visual comparisons of different ways to obtain edge maps. (A) TRUS image; (B) edge maps achieved from TRUS image; (C) edge maps achieved from the semantic segmentation of prostates. TRUS, transrectal ultrasound.
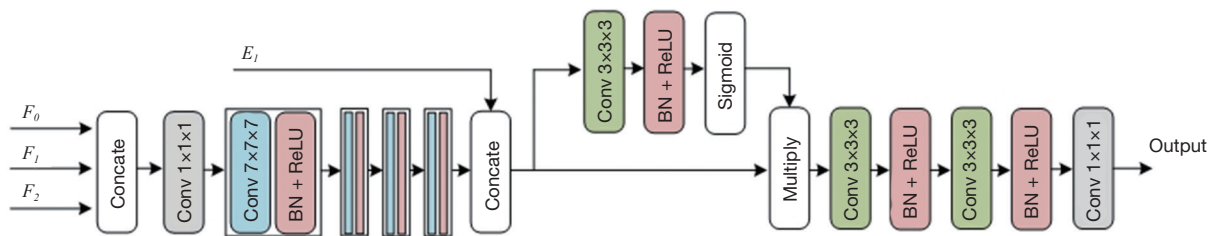

Figure 4 The network architecture of the global feature extractor. BN, batch normalization; ReLU, rectified linear unit.

convolutional layers are adopted to decrease the number of the channel dimensions, which aims to reduce the computational complexity. Then, four 7×7×7 convolutional blocks are designed to build a density connection between feature maps and per-pixel classifier, which enhances the capability to handle different shapes and sizes. Motivated by previous work (38), to enhance the sensitively to edge information, low-layer features E_1, which are enhanced by the EEM, are also introduced to the GFE. To further help the network select more important features, a spatial attention structure is adopted, which is similar to the spatial attention module in the 3D SCAM. In the spatial attention structure, the 3×3×3 convolutional block is utilized for the feature extraction and the Sigmoid operation is used to constrain the value of the weight scores to be between [0, 1]. The weight scores calculated by the spatial attention structure are multiplied by the fused features to obtain the selective features. Finally, two 3×3×3 convolutional blocks are used for the final feature extraction and the 1×1×1 convolutional layer aims to map the channel dimensions to match the channel of predicted prostates.

The discrimination network

The discriminator of the traditional GAN utilizes the whole image as input to conduct the discrimination, which only outputs one value to determine whether the generated image is real or false. Different from mapping the whole image to one value, PatchGAN (39) extracts features from the input image, and then maps the input image into the N×N matrix by the full convolution structure. Benefiting from the PatchGAN structure, it can effectively enhance the attention to each area of the image. Hence, to achieve a better discriminative effect, PatchGAN is adopted as the discriminator network of the proposed 3D EAGAN.

The loss function

In deep learning tasks, the loss function plays a vital role in the neural network model training process. An elaborately designed loss function can effectively speed up the convergence of the model training and improve the prediction accuracy of the model. In the training process, the edge-aware segmentation network G and the discriminator network D are optimized by the minimax game. The objective function for training the edge-aware segmentation network is defined as:

$$minL(D) = \left[D(x,y) - l_r \right]^2 + \left[D(x, G(x)) - l_f \right]^2 \quad [11]$$

where $G()$ denotes the edge-aware segmentation network

and $D()$ denotes the discriminator network. x and y denote input TRUS images and ground-truth labels, respectively. l_r and l_f represent the real label and fake matrix label with constant elements one and zero, respectively. The objective function for training the discriminator network is defined as:

$$maxL(G) = \left[D(x,G(x)) - l_t \right]^2 + \alpha * l_{dice} \left[y, G(x) \right] + \beta * l_{dice} \left[y_e, \hat{y}_e \right] \quad [12]$$

where l_t is the matrix with constant elements one. y_e and $\widehat{y_e}$ denote the predicted edge maps and ground-truth edge maps, respectively. l_{dice} represents the Dice loss, which is widely used in medical image segmentation tasks. α and β represent the hyper-parameters that control the impact of the loss function. According to extensive experiments, when α and β are set to 1 and 0.5, respectively, the proposed method achieves the best prediction performance.

Results

In this section, the experimental setups are first introduced, including experimental environments and implementation tools, datasets, and evaluation metrics. Then, experiments are performed to compare the proposed method with other medical segmentation methods. Finally, the ablation study is conducted to verify the effectiveness of components in our method.

Experimental setups

Experimental environments and implement tools
The proposed method is programmed with Python 3.7 and implemented by PyTorch 1.2.0. To train the network, the training and testing process are performed on NVIDIA GeForce RTX 3090 GPU.

Implementation details
For the training stage, due to the limited GPU memory, input TRUS images are down-sampled with the size of 88×112×112. For the proposed 3D EAGAN, both the edge-aware segmentation network and the discriminator network are trained using the Adam optimizer (40) with the parameters λ_1=0.9, λ_2=0.999, and the learning rate is initialized as 0.00001.

Dataset
To verify the effectiveness of the proposed method, the experiments have conducted on our dataset and open-source dataset μRegPro (41). For our dataset, the TRUS images are obtained through a mechanically assisted biopsy system used by collaborating radiologists (CRs) (42) of Western University. The TRUS image dataset consists of 56 patients. We acquired one 3D TRUS image from each patient. These 3D TRUS images are acquired with an end-firing 5–9 MHz TRUS transducer probe (Philips Medical Systems, Seattle, WA, USA). The 3D TRUS image contains 350×448×448 voxels with a voxel size of 0.19×0.18×0.18 mm³. The data is processed using spatial and intensity distribution normalization. For the μRegPro dataset, we only choose the TRUS images as the training and validation samples.

Compared methods
To verify the effectiveness of the proposed method, seven state-of-the-art medical segmentation methods are used to conduct the experiments, including 3D FCN (23), 3D U-Net (24), Skip-Densenet (43), Deeplabv3+ (44), deep attentive features for three-dimensional prostate segmentation (DAF 3D) (27), Vox2Vox (45), and Chen *et al.* (46).

Evaluation metrics
As following previous work (25-29), five evaluation metrics are used to measure the segmentation performance of our proposed method, including Dice similarity coefficient (Dice), Jaccard index (Jaccard), Hausdorff distance (HD, in voxel), Precision, and Recall.

The Dice is used to evaluate the similarity between predicted prostates and the ground truth ones,

$$Dice(P,G) = \frac{2|P| \cap |G|}{|P| + |G|} \quad [13]$$

where P and G denote the predicted prostates and the ground-truth prostates. $|\cdot|$ represents the number of voxels. The value of Dice is in the range of [0, 1], the higher value denotes better segmentation performance.

The Jaccard is used to evaluate the similarity between predicted prostates and the ground truth ones,

$$Jaccard(P,G) = \frac{|P| \cap |G|}{|P| \cup |G|} \quad [14]$$

where P and G denote the predicted prostates and the ground-truth prostates. $|\cdot|$ represents the number of voxels. The value of Dice is in the range of [0, 1], the higher value denotes better segmentation performance.

The HD is utilized to evaluate the distance between predicted prostates and the ground truth ones,

Table 1 Quantitative results for prostate segmentation of different methods on the proposed dataset

Method	Dice (%)	Jaccard (%)	HD (mm)	Precision (%)	Recall (%)
3D FCN	84.12±3.02	72.46±2.40	11.96±5.89	86.49±2.68	84.08±2.87
Chen *et al.*	85.32±2.62	75.04±2.27	10.82±3.62	87.39±2.56	85.08±2.74
3D U-Net	86.34±2.07	76.24±2.64	10.77±4.05	86.83±3.22	85.19±2.11
Skip-Densenet	88.90±1.87	80.56±2.19	8.95±2.63	90.25±1.94	88.13±1.90
Deeplabv3+	89.29±2.27	80.96±2.53	6.89±1.82	90.17±2.12	88.39±1.75
DAF 3D	89.92±1.75	81.75±2.67	6.48±1.61	90.78±1.27	90.12±1.58
Vox2Vox	90.32±1.57	82.76±1.99	7.23±2.20	91.04±1.17	91.34±1.46
3D EAGAN	92.80±0.75*	87.01±0.42*	4.64±0.69*	93.11±0.62*	92.42±1.00*

Data are presented as mean ± standard deviation. *, the best results. Dice, Dice similarity coefficient; HD, Hausdorff distance; 3D FCN, three-dimensional fully convolutional network; 3D, three-dimensional; DAF 3D, deep attentive features for three-dimensional prostate segmentation; 3D EAGAN, three-dimensional edge-aware attention generative adversarial network.

$$h(A,B) = \max_{a \subset A} \left\{ \min_{b \subset B} \|a - b\| \right\} \qquad [15]$$

$$h(B,A) = \max_{b \subset B} \left\{ \min_{a \subset A} \|b - a\| \right\} \qquad [16]$$

$$HD(A,B) = \max \left\{ h(A,B), h(B,A) \right\} \qquad [17]$$

where $\|\cdot\|$ represents the distance paradigm between predicted prostates and the ground truth ones. The lower value of HD represents better segmentation performance.

The Precision is used to evaluate the proportion of samples with a predicted value of one and a true value of one among all samples with a predicted value of one,

$$precision(P,G) = \frac{Area(P \cap G)}{Area(P)} \qquad [18]$$

where P and G denote the predicted prostates and the ground-truth prostates. The value of Precision is in the range of [0, 1], the higher value denotes better segmentation performance.

The Recall is used to evaluate the proportion of samples with a predicted value of one and a true value of one among all samples with a true value of one,

$$Recall(P,G) = \frac{Area(P \cap G)}{Area(G)} \qquad [19]$$

where P and G denote the predicted prostates and the ground-truth prostates. The value of Recall is in the range of [0, 1], the higher value denotes better segmentation performance.

Comparison to state-of-the-art methods

To verify the effectiveness of the proposed method, the segmentation performance of the proposed method is quantitatively evaluated through comparisons to seven state-of-the-art segmentation methods. The experimental results on the proposed dataset are shown in *Table 1*. It can be observed that the proposed 3D EAGAN outperforms other methods on the used metrics. Specifically, 3D EAGAN achieves the mean Dice of 92.80%, Jaccard of 87.01%, HD of 4.64 mm, Precision of 93.11%, and Recall of 92.42%, respectively. Compared with the traditional method of 3D FCN, our proposed method outperforms it by a large margin. The 3D FCN utilizes progressive down-sampling to reduce the resolution of the input image, which would lead to the loss of detailed information and degrades the segmentation performance. The 3D U-Net utilizes skip connection layers to effectively combine different features of the encoder and the decoder, which makes it achieve better segmentation performance than 3D FCN. The proposed DCM can effectively introduce abundant detailed information to the encoder, which compensates for the loss of detailed information caused by the down-sampling process. Hence, our proposed method improves the 3D U-Net by 6.46%, 10.77%, 6.13 mm, 6.28%, and 7.23% on Dice, Jaccard, HD, Precision, and Recall, respectively. Compared with the DAF 3D, which also utilizes the attention mechanism to make the network focus on prostates, our proposed 3D EAGAN improves it for 2.88% on Dice, 5.26% on Jaccard, 1.84 mm on HD, 2.33% on Precision, 2.30% on Recall, respectively. The improvements

benefit from the proposed 3D SCAM can not only focus on features in the spatial domain but also stress essential features in the channel domain. Compared with the GAN-based Vox2Vox, our proposed 3D EAGAN improves it by 2.48% on Dice, 4.25% on Jaccard, 2.59 mm on HD, 2.07% on Precision, and 1.08% on Recall, respectively. Benefiting from the proposed EEM, the proposed 3D EAGAN can pay attention to edge information, which leads to the improvement of segmentation performance.

To further verify the effectiveness of the proposed method, the 2D slice visualization results of prostate segmented by different methods are shown in *Figure 5*. The first row of images is the prostate TRUS images of different samples, the second row of images is the real prostate label image, and the rest of the images are the visualization of the segmentation effect of different methods. For 3D FCN and Chen *et al.* methods, these methods use continuous downsampling to reduce the image resolution, which causes the loss of detailed information. Hence, the results of these methods have a large difference between the predicted prostate and the ground-truth one. For the Skip-Densenet and Deeplabv3++ models, there is a substantial discrepancy between the segmentation results and the ground-truth ones, indicating a notable deficiency in their segmentation performance. The Vox2Vox method achieves better segmentation results than other methods, the reason is that it uses generative discriminator training to train the whole network. Different from these methods, the prostate images segmented by the proposed method are closer to the ground-truth ones.

In order to evaluate the statistical significance of the proposed method over other compared methods on all metrics, the paired *t*-tests are performed, and the P values are reported. The results are shown in *Table 2*. It can be observed from *Table 2*, it becomes evident that the null hypotheses for the first six comparing pairs on all metrics are not accepted at the 0.05 level, which means that the proposed method is better than these six comparing methods on all metrics. For the comparison between the proposed method and Vox2Vox, the P value on metric Recall is beyond the 0.05 level, which indicates the proposed method achieves similar performance with Vox2Vox on the Recall metric.

The experimental results on the µRegPro dataset are shown in *Table 3*. It can be observed that the proposed method achieves the best results on all metrics compared with other methods. Moreover, to evaluate the statistical significance of the proposed method over other compared

methods on all metrics, the paired *t*-tests are performed, and the P values are reported in *Table 4*. It can be observed that the null hypotheses for the first six comparing pairs on all metrics are not accepted at the 0.05 level, which means that the proposed method is better than almost all methods on all metrics. For the comparison between the proposed method and Vox2Vox, the P values on metric HD and Precision are beyond the 0.05 level, which indicates the proposed method achieves similar performance with Vox2Vox on HD and Precision metrics.

In summary, the reasons for satisfactory segmentation results of the proposed method are: (I) the proposed EEM can effectively enhance the perception of shallow features for prostate edge information, thereby improving the network's segmentation accuracy for prostates. (II) The proposed 3D SCAM enhances more important features in the spatial and channel dimensions through the attention mechanism, which makes the network pay more attention to the prostate region.

Ablation study

Ablation study on the proposed DCM

As discussed before, the DCM is adopted in the edge-aware segmentation network to introduce abundant detail information to the network. The ablation study experiments are conducted to verify the effectiveness of the DCM. The experimental results are shown in *Table 5*. The proposed method without using the DCM is denoted as "3D EAGAN w/o DCM". On the contrary, "3D EAGAN w/ DCM" represents the DCM used in the proposed method. It can be observed that with the use of the DCM, the segmentation performance is significantly improved.

To further verify the effectiveness of the DCM, feature maps are extracted from the encoder module in the edge-aware segmentation network. The visualization of the feature maps is shown in *Figure 6*. It can be observed that with the use of the DCM, abundant detail information can be introduced to the encoder module, which can enhance the robustness of the proposed method, and it consists with experimental results in *Table 1*.

Ablation study on the proposed 3D spatial and channel attention module

The 3D spatial and channel attention module is added to the edge-aware segmentation network to selectively leverage the useful prostate features. To verify the effectiveness of the 3D spatial and channel attention module, it is compared
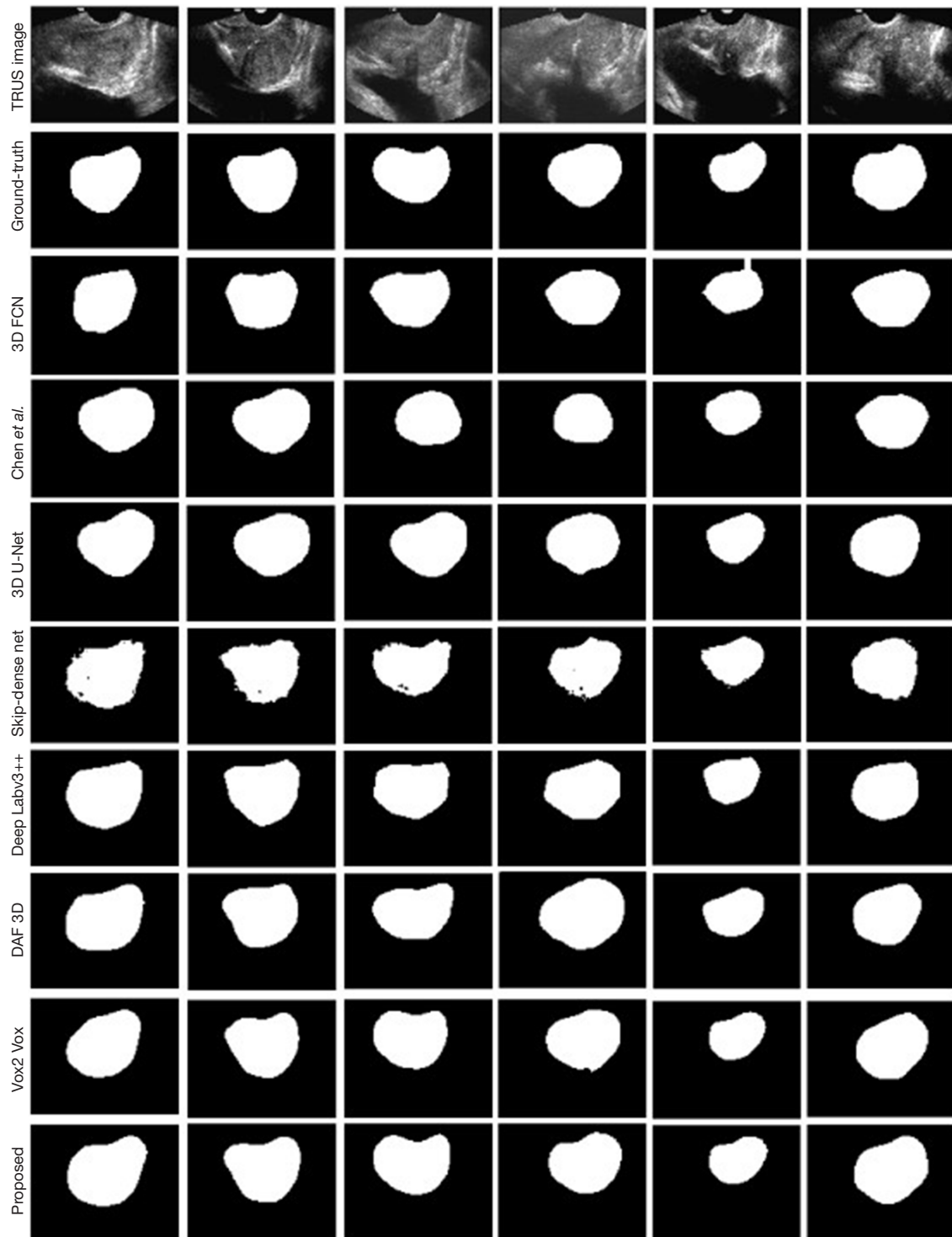
Figure 5 Visualization results of different methods on TRUS dataset. 3D FCN, three-dimensional fully convolutional network; DAF 3D, deep attentive features for three-dimensional prostate segmentation; TRUS, transrectal ultrasound.

Table 2 P values from a paired *t*-test between the proposed method and other methods on our dataset

Method	Dice	Jaccard	HD	Precision	Recall
3D FCN	10^{-5}	10^{-9}	10^{-3}	10^{-5}	10^{-5}
Chen *et al.*	10^{-5}	10^{-12}	10^{-4}	10^{-5}	10^{-7}
3D U-Net	10^{-5}	10^{-7}	10^{-3}	10^{-4}	10^{-7}
Skip-Densenet	0.02	10^{-5}	10^{-3}	10^{-3}	10^{-4}
Deeplabv3+	0.02	10^{-5}	0.01	10^{-3}	10^{-4}
DAF 3D	0.04	10^{-4}	0.02	10^{-3}	10^{-3}
Vox2Vox	0.03	10^{-6}	0.04	10^{-3}	0.08

Dice, Dice similarity coefficient; Jaccard, Jaccard index; HD, Hausdorff distance; 3D FCN, three-dimensional fully convolutional network; 3D, three-dimensional; DAF 3D, deep attentive features for three-dimensional prostate segmentation.

Table 3 Quantitative results for prostate segmentation of different methods on the µRegPro dataset

Method	Dice (%)	Jaccard (%)	HD (mm)	Precision (%)	Recall (%)
3D FCN	76.21±4.01	65.23±2.14	11.50±3.62	80.13±3.14	75.22±2.63
Chen *et al.*	80.21±2.04	69.42±2.51	10.07±3.82	77.24±3.06	82.82±2.81
3D U-Net	82.34±2.83	71.55±3.04	8.67±2.96	85.49±2.58	81.14±3.16
Skip-Densenet	83.02±3.16	74.22±2.82	7.34±3.11	86.25±2.66	82.50±3.22
Deeplabv3+	85.23±3.53	75.48±3.11	7.61±3.53	86.63±3.04	83.26±3.52
DAF 3D	86.12±3.24	79.51±2.83	6.28±3.62	84.22±2.43	85.22±2.31
Vox2Vox	87.22±2.80	80.01±2.93	5.73±2.32	87.81±3.01	84.53±2.14
3D EAGAN	87.31±1.61*	81.21±1.77*	5.12±2.03*	88.04±2.06*	86.06±1.36*

Data are presented as mean ± standard deviation. *, the best results. Dice, Dice similarity coefficient; Jaccard, Jaccard index; HD, Hausdorff distance; 3D FCN, three-dimensional fully convolutional network; 3D, three-dimensional; DAF 3D, deep attentive features for three-dimensional prostate segmentation; 3D EAGAN, three-dimensional edge-aware attention generative adversarial network.

Table 4 P values from a paired *t*-test between the proposed method and other methods on the µRegPro dataset

Method	Dice	Jaccard	HD	Precision	Recall
3D FCN	10^{-9}	10^{-8}	10^{-5}	10^{-5}	10^{-6}
Chen *et al.*	10^{-5}	10^{-9}	10^{-5}	10^{-5}	10^{-5}
3D U-Net	10^{-4}	10^{-7}	10^{-4}	10^{-4}	10^{-5}
Skip-Densenet	10^{-4}	10^{-5}	10^{-3}	10^{-5}	10^{-3}
Deeplabv3+	10^{-4}	10^{-3}	0.02	10^{-4}	10^{-3}
DAF 3D	10^{-3}	0.03	0.01	10^{-5}	0.03
Vox2Vox	0.03	0.02	0.06	0.09	10^{-4}

Dice, Dice similarity coefficient; Jaccard, Jaccard index; HD, Hausdorff distance; 3D FCN, three-dimensional fully convolutional network; 3D, three-dimensional; DAF 3D, deep attentive features for three-dimensional prostate segmentation.

Table 5 Evaluation of using the DCM

Method	Dice (%)	Jaccard (%)	HD (mm)	Precision (%)	Recall (%)
3D EAGAN w/o DCM	91.08±1.51	83.64±1.72	5.46±0.93	90.43±1.62	91.79±1.45
3D EAGAN w/ DCM	92.80±0.75*	87.01±0.42*	4.64±0.69*	93.11±0.62*	92.42±1.00*

Data are presented as mean ± standard deviation. *, the best results. DCM, detail compensation module; Dice, Dice similarity coefficient; Jaccard, Jaccard index; HD, Hausdorff distance; 3D EAGAN, three-dimensional edge-aware attention generative adversarial network.
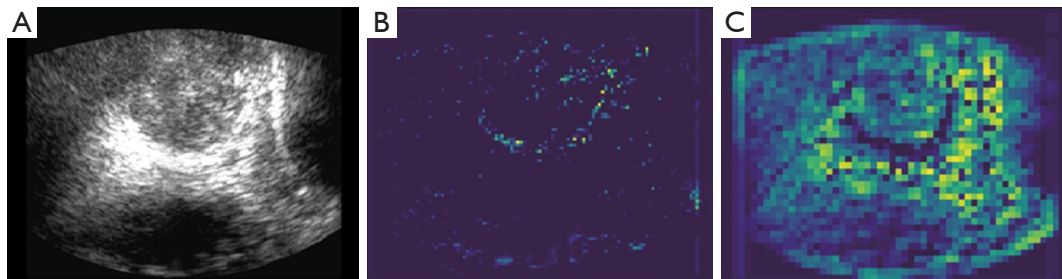

Figure 6 Visualization results of the feature maps extracted from the encoder module in the edge-aware segmentation network. (A) TRUS image; (B) EAGAN w/o DCM; (C) EAGAN w/ DCM. TRUS, transrectal ultrasound; EAGAN, edge-aware attention generative adversarial network; DCM, detail compensation module.

Table 6 Evaluation of using the 3D spatial and channel attention module

Method	Dice (%)	Jaccard (%)	HD (mm)	Precision (%)	Recall (%)
3D EAGAN w/o 3D SCAM	92.07±0.91	85.04±1.04	5.20±0.92	92.04±1.26	91.75±1.76
3D EAGAN w/ 3D SCAM	92.80±0.75*	87.01±0.42*	4.64±0.69*	93.11±0.62*	92.42±1.00*

Data are presented as mean ± standard deviation. *, the best results. 3D, three-dimensional; Dice, Dice similarity coefficient; Jaccard, Jaccard index; HD, Hausdorff distance; 3D EAGAN, three-dimensional edge-aware attention generative adversarial network; 3D SCAM, three-dimensional spatial and channel attention module.

Table 7 Evaluation of using the EEM

Method	Dice (%)	Jaccard (%)	HD (mm)	Precision (%)	Recall (%)
3D EAGAN w/o EEM	91.62±0.88	84.46±1.28	5.12±0.63	91.21±1.43	92.03±0.90
3D EAGAN w/ EEM	92.80±0.75*	87.01±0.42*	4.64±0.69*	93.11±0.62*	92.42±1.00*

Data are presented as mean ± standard deviation. *, the best results. EEM, edge enhancement module; Dice, Dice similarity coefficient; Jaccard, Jaccard index; HD, Hausdorff distance; 3D EAGAN, three-dimensional edge-aware attention generative adversarial network.

with the network without using the 3D spatial and channel attention module. The experimental results are shown in *Table 6*. "3D EAGAN w/o 3D SCAM" represents the 3D spatial and channel attention module removed from the 3D EAGAN. "3D EAGAN w/ 3D SCAM" is kept in the 3D EAGAN. It can be observed that the use of the 3D spatial and channel attention module can slightly improve the used metrics.

Ablation study on the proposed EEM
The EEM is utilized to guide the shallow layers of the edge-aware segmentation network to focus on the contour and edge information of prostates. To verify the effectiveness of the EEM, ablation study experiments are performed. The experimental results are shown in *Table 7*. It can be observed that the use of the EEM can improve the performance of the 3D EAGAN.
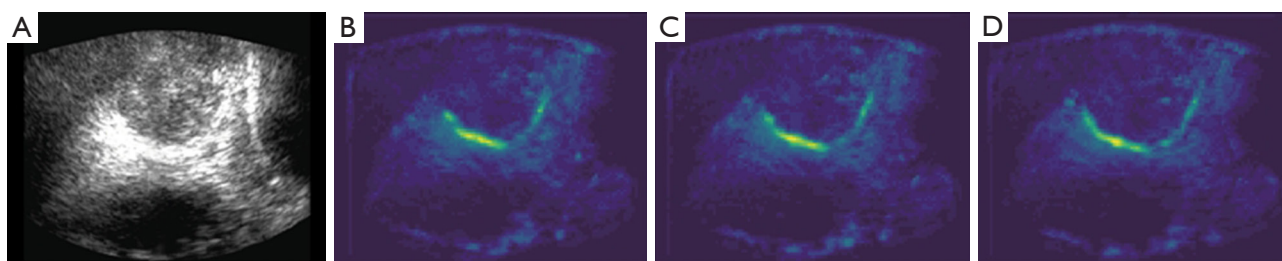
Figure 7 Visualization results of the feature maps extracted from the shallow layers of the edge-aware segmentation network. (A) is the input TRUS images; (B-D) are feature maps extracted from the shallow layers in the encoder module. TRUS, transrectal ultrasound.

Table 8 Evaluation of numbers of channel dimensions in the edge-aware segmentation network

Method	Dice (%)	Jaccard (%)	HD (mm)	Precision (%)	Recall (%)	Params (MB)
{8,16,32,64}	90.48±0.89	84.33±0.79	8.11±0.83	89.48±0.69	88.34±1.28	73.92*
{16,32,64,128}	92.80±0.75*	87.01±0.42	4.64±0.69*	93.11±0.62*	92.42±1.00*	75.10
{32,64,128,256}	92.76±0.82	87.26±0.48*	4.71±0.57	93.21±0.66	92.40±0.87	78.77

Data are presented as mean ± standard deviation. *, the best results. Dice, Dice similarity coefficient; Jaccard, Jaccard index; HD, Hausdorff distance.

To further verify the effectiveness of the EEM, the visualization of feature maps extracted from the shallow layers of the edge-aware segmentation network is shown in *Figure 7*. It can be observed that with the help of the EEM, the prostate edge is more distinctive than the surrounding features in feature maps. Hence, shallow layers of the edge-aware segmentation network can pay attention to the edge of prostates.

Number of channel dimensions in the edge-aware segmentation network

In the proposed 3D EAGAN, 3D convolutional layers are adopted to extract prostates in the 3D spatial domain. However, 3D convolutional layers inevitably increase computational complexity. To balance the segmentation performance and computational complexity, ablation study experiments on the number of channel dimensions in the edge-aware segmentation network are conducted. The experimental results are shown in *Table 8*. Hence, the number of channel dimensions in the edge-aware segmentation network is limited to {16,32,64,128} according to the experimental results.

In summary, to statistically evaluate the performance of each component in the proposed method, the paired *t*-tests are performed, and the P values are reported. The results are shown in *Tables 9* and *10*. For the ablation study on the proposed DCM, 3D SCAM, and EEM, it can be observed that the null hypotheses on most metrics are not accepted at the 0.05 level, which means that the DCM, 3D SCAM, and EEM can effectively improve segmentation accuracy. For the ablation study on the number of channel dimensions in the edge-aware segmentation network, it can be observed that the null hypotheses for the comparison between {8,16,32,64} and {16,32,64,128} on all metrics are not accepted at the 0.05 level, which means the channel dimensions of {16,32,64,128} achieves better performance. But for the comparison between {16,32,64,128} and {32,64,128,256}, the P value on almost all metrics is beyond the 0.05 level, which means that the channel dimensions of {16,32,64,128} and {32,64,128,256} achieve the similar performance.

Discussion

This paper proposed a 3D EAGAN-based prostate segmentation method, which achieved good segmentation performance on two datasets. Since prostates in TRUS images may have missing and ambiguous boundaries, accurately segmenting prostates from TRUS images is still a challenging task. Many traditional segmentation methods utilize prior information to extract the boundary of prostates, which failed in accurately segmenting prostates

Table 9 Ablation study (DCM, 3D SCAM, and EEM) of P values from a paired *t*-test between the proposed method and other methods on our dataset

Metric	3D EAGAN w/o DCM *vs.* 3D EAGAN w/ DCM	3D EAGAN w/o 3D SCAM *vs.* 3D EAGAN w/ 3D SCAM	3D EAGAN w/o EEM *vs.* 3D EAGAN w/ EEM
Dice	10^{-4}	0.08	10^{-3}
Jaccard	10^{-3}	10^{-4}	10^{-4}
HD	0.02	0.01	0.13
Precision	10^{-5}	10^{-3}	10^{-5}
Recall	0.03	0.04	0.12

DCM, detail compensation module; 3D SCAM, three-dimensional spatial and channel attention module; EEM, edge enhancement module; 3D EAGAN, three-dimensional edge-aware attention generative adversarial network; Dice, Dice similarity coefficient; Jaccard, Jaccard index; HD, Hausdorff distance.

Table 10 Ablation study (number of channel dimensions) of P values from a paired *t*-test between the proposed method and other methods on our dataset

Metric	{8,16,32,64} *vs.* {16,32,64,128}	{32,64,128,256} *vs.* {16,32,64,128}
Dice	10^{-3}	0.3
Jaccard	10^{-6}	0.5
HD	0.01	0.05
Precision	10^{-6}	0.04
Recall	10^{-6}	0.58

Dice, Dice similarity coefficient; Jaccard, Jaccard index; HD, Hausdorff distance.

due to the complex background of TRUS images.

Recently, DCNN have proved great performance and robustness in many computer vision tasks, which utilize backpropagation algorithms to learn features automatically. To achieve higher segmentation performance, many deep CNN-based prostate segmentation methods have been proposed. Although these methods can achieve better performance than traditional prostate segmentation methods, it still lacks sensitivity to the boundary of prostates, which may cause inaccuracies in the edges of the segmentation results, see *Figure 5*. Hence, how to effectively and accurately segment the edge of the prostate has become a key issue to be addressed in this field. To address this issue and accurately segment the edges of the prostate, an EEM is proposed to guide shallow layers to pay attention to edge information prostates. *Tables 1*,*3*,7, and *Figures 5*,7 all present that the proposed EEM can effectively improve the accuracy of prostate edge segmentation. Moreover,

the DCM is proposed to introduce abundant detailed information to the network.

Although the proposed method achieves satisfactory performance results, there is still some room for improvement. Since actual implementation relies heavily on hardware performance, a more lightweight network (47) is very important for actual application scenarios. The edge-aware segmentation network used for segmentation in this method has reached a network parameter of 75.1 MB. In the future, we will consider lightweighting the network to achieve better real-time effects.

Conclusions

In this paper, a 3D EAGAN-based prostate segmentation method is proposed, which consists of an edge-aware segmentation network and a discriminator network. In the edge-aware segmentation network, the DCM is proposed to introduce abundant detailed information to the network. In addition, an EEM is proposed to guide shallow layers to pay attention to edge information prostates. Experimental results demonstrate the proposed method has achieved satisfactory results in 3D TRUS image segmentation of prostates.

Acknowledgments

Footnote

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims. amegroups.com/article/view/10.21037/qims-23-1698/rc). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was approved by the Ethics Committee of the Western University. Participants provided informed consent before taking part in the study. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019;69:7-34.
2. Pinto F, Totaro A, Calarco A, Sacco E, Volpe A, Racioppi M, D'Addessi A, Gulino G, Bassi P. Imaging in prostate cancer diagnosis: present role and future perspectives. Urol Int 2011;86:373-82.
3. Wang Y, Cheng JZ, Ni D, Lin M, Qin J, Luo X, Xu M, Xie X, Heng PA. Towards Personalized Statistical Deformable Model and Hybrid Point Matching for Robust MR-TRUS Registration. IEEE Trans Med Imaging 2016;35:589-604.
4. Yan P, Xu S, Turkbey B, Kruecker J. Discrete deformable model guided by partial active shape model for TRUS image segmentation. IEEE Trans Biomed Eng 2010;57:1158-66.
5. Bahn DK, Lee F, Badalament R, Kumar A, Greski J, Chernick M. Targeted cryoablation of the prostate: 7-year outcomes in the primary treatment of prostate cancer. Urology 2002;60:3-11.
6. Ladak HM, Mao F, Wang Y, Downey DB, Steinman DA, Fenster A. Prostate boundary segmentation from 2D ultrasound images. Med Phys 2000;27:1777-88.
7. Pathak SD, Chalana V, Haynor DR, Kim Y. Edge-guided boundary delineation in prostate ultrasound images. IEEE Trans Med Imaging 2000;19:1211-9.
8. Shen D, Zhan Y, Davatzikos C. Segmentation of prostate boundaries from ultrasound images using statistical shape model. IEEE Trans Med Imaging 2003;22:539-51.
9. Yan P, Xu S, Turkbey B, Kruecker J. Adaptively learning local shape statistics for prostate segmentation in ultrasound. IEEE Trans Biomed Eng 2011;58:633-41.
10. Santiago C, Nascimento J, Marques J. 2D Segmentation Using a Robust Active Shape Model With the EM Algorithm. IEEE Trans Image Process 2015;24:2592-601.
11. Ghanei A, Soltanian-Zadeh H, Ratkewicz A, Yin FF. A three-dimensional deformable model for segmentation of human prostate from ultrasound images. Med Phys 2001;28:2147-53.
12. Wang Y, Cardinal HN, Downey DB, Fenster A. Semiautomatic three-dimensional segmentation of the prostate using two-dimensional ultrasound images. Med Phys 2003;30:887-97.
13. Hu N, Downey DB, Fenster A, Ladak HM. Prostate boundary segmentation from 3D ultrasound images. Med Phys 2003;30:1648-59.
14. Gong L, Pathak SD, Haynor DR, Cho PS, Kim Y. Parametric shape modeling using deformable superellipses for prostate segmentation. IEEE Trans Med Imaging 2004;23:340-9.
15. Qiu W, Yuan J, Ukwatta E, Sun Y, Rajchl M, Fenster A. Prostate segmentation: an efficient convex optimization approach with axial symmetry using 3-D TRUS and MR images. IEEE Trans Med Imaging 2014;33:947-60.
16. Ghose S, Oliver A, Mitra J, Martí R, Lladó X, Freixenet J, Sidibé D, Vilanova JC, Comet J, Meriaudeau F. A supervised learning framework of statistical shape and probability priors for automatic prostate segmentation in ultrasound images. Med Image Anal 2013;17:587-600.
17. Zhan Y, Shen D. Deformable segmentation of 3-D ultrasound prostate images using statistical texture matching method. IEEE Trans Med Imaging 2006;25:256-72.
18. Yang X, Rossi PJ, Jani AB, Mao H, Curran WJ, Liu T. 3D Transrectal Ultrasound (TRUS) Prostate Segmentation Based on Optimal Feature Learning Framework. Proc

SPIE Int Soc Opt Eng 2016;9784:97842F.

19. Xun S, Li D, Zhu H, Chen M, Wang J, Li J, Chen M, Wu B, Zhang H, Chai X, Jiang Z, Zhang Y, Huang P. Generative adversarial networks in medical image segmentation: A review. Comput Biol Med 2022;140:105063.

20. Shao X, Liu M, Li Z, Zhang P. CPDINet: Blind image quality assessment via a content perception and distortion inference network. IET Image Processing 2022;16:1973-87.

21. Zhang P, Shao X, Li Z. Cycleiqa: Blind image quality assessment via cycle-consistent adversarial networks. 2022 IEEE International Conference on Multimedia and Expo (ICME), Taipei, 2022;1-6.

22. Liu M, Wu K, Jiang L. ADC-Net: adaptive detail compensation network for prostate segmentation in 3D transrectal ultrasound images. Medical Imaging 2023: Ultrasonic Imaging and Tomography 2023;12470:211-20.

23. Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. IEEE Trans Pattern Anal Mach Intell 2017;39:640-51.

24. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A. editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science(), Springer, 2015; 234-241.

25. Ghavami N, Hu Y, Bonmati E, Rodell R, Gibson E, Moore C, Barratt D. Automatic slice segmentation of intraoperative transrectal ultrasound images using convolutional neural networks. Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling (SPIE) 2018;10576:1057603.

26. Yang X, Yu L, Wu L, Wang Y, Qin J, Heng P. Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images. Proceedings of the AAAI Conference on Artificial Intelligence 2017. doi: 10.1609/aaai.v31i1.10761.

27. Wang Y, Dou H, Hu X, Zhu L, Yang X, Xu M, Qin J, Heng PA, Wang T, Ni D. Deep Attentive Features for Prostate Segmentation in 3D Transrectal Ultrasound. IEEE Trans Med Imaging 2019;38:2768-78.

28. Lei Y, Tian S, He X, Wang T, Wang B, Patel P, Jani AB, Mao H, Curran WJ, Liu T, Yang X. Ultrasound prostate segmentation based on multidirectional deeply supervised V-Net. Med Phys 2019;46:3194-206.

29. Pellicer-Valero OJ, Gonzalez-Perez V, Ramn-Borja JLC, Garcia IM, Benito MB, Gomez PP, Rubio-Briones J, Ruperez MJ, Martin-Guerrero JD. Robust resolution-enhanced prostate segmentation in magnetic resonance and ultrasound images through convolutional neural networks. Applied Sciences 2021;11:844.

30. Chen S, Ma K, Zheng Y. Med3d: Transfer learning for 3d medical image analysis. arXiv:1904.00625. 2019.

31. Liu Q, Zhou H, Xu Q, Liu X, Wang Y. PSGAN: A generative adversarial network for remote sensing image pan-sharpening. IEEE Transactions on Geoscience and Remote Sensing 2021;59:10227-42.

32. Zhu C, Xu J, Feng D, Xie R, Song L. Edge-based video compression texture synthesis using generative adversarial network. IEEE Transactions on Circuits and Systems for Video Technology 2022;32:7061-76.

33. Dong X, Lei Y, Wang T, Thomas M, Tang L, Curran WJ, Liu T, Yang X. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. Med Phys 2019;46:2157-68.

34. Wang W, Wang G, Wu X, Ding X, Cao X, Wang L, Zhang J, Wang P. Automatic segmentation of prostate magnetic resonance imaging using generative adversarial networks. Clin Imaging 2021;70:1-9.

35. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016:770-8.

36. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. In: Kůrková V, Manolopoulos Y, Hammer B, Iliadis L, Maglogiannis I. editors. Artificial Neural Networks and Machine Learning – ICANN 2018. Lecture Notes in Computer Science(), Springer, 2018;11141:270-9.

37. Canny J. A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 1986;8:679-98.

38. Fan M, Lai S, Huang J, Wei X, Chai Z, Luo J, Wei X. Rethinking BiSeNet for real-time semantic segmentation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021;9716-25.

39. Isola P, Zhu J Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017;1125-34.

40. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv:1412.6980. 2014.

41. Available online: https://muregpro.github.io/data.html

42. Bax J, Cool D, Gardi L, Knight K, Smith D, Montreuil J, Sherebrin S, Romagnoli C, Fenster A. Mechanically assisted 3D ultrasound guided prostate biopsy system. Med Phys 2008;35:5397-410.

43. Bui TD, Shin J, Moon T. Skip-connected 3D DenseNet for volumetric infant brain MRI segmentation. Biomed Signal Process Control 2019;54:101613.
44. Chen L C, Zhu Y, Papandreou G, Schroff D, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European conference on computer vision (ECCV), 2018;801-18.
45. Cirillo M D, Abramian D, Eklund A. Vox2Vox: 3D-GAN for brain tumour segmentation. International MICCAI Brainlesion Workshop 2020;274-84.
46. Chen J, Wan Z, Zhang J, Li W, Chen Y, Li Y, Duan Y. Medical image segmentation and reconstruction of prostate tumor based on 3D AlexNet. Comput Methods Programs Biomed 2021;200:105878.
47. Tan M, Le QV. Efficientnet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning, ICML 2019, Long Beach, 2019;6105-14.