Multi-Channel Feature Pyramid Networks for Prostate Segmentation, Based on Transrectal Ultrasound Imaging

Lei Geng [1,2], Simu Li [2,3], Zhitao Xiao [1,2,*] and Fang Zhang [1,2]

[1] School of Life Sciences, Tiangong University, Tianjin 300387, China; genglei@tiangong.edu.cn (L.G.); hhzhangfang@126.com (F.Z.)

[2] Tianjin Key Laboratory of Optoelectronic Detection Technology and System, Tianjin 300387, China; lisimu123@163.com

[3] School of Electronics and Information Engineering, Tiangong University, Tianjin 300387, China

[*] Correspondence: xiaozhitao@tiangong.edu.cn

check for updates

Abstract: Accurate segmentation for transrectal ultrasound imaging (TRUS) is often a challenging medical image processing task. The problem of weak boundary between adjacent prostate tissue and non-prostate tissue, and high similarity between artifact area and prostate area has always been the difficulty of TRUS image segmentation. In this paper, we construct a multi-channel feature pyramid network (MFPN) based on deep convolutional neural network-based prostate segmentation method to process multi-scale feature maps. Each level enhances the edge characteristics of the prostate by controlling the scale of the channel. The optimized regression mechanism of the target area was used to accurately locate the prostate. Experimental results showed that the proposed method achieved the key indicator Dice similarity coefficient and average absolute distance of 0.9651 mm and 0.504 mm, which outperformed state-of-the-art approaches.

1. Introduction

Prostate cancer is the most common cancer among men in the United States in 2019. According to the latest survey by the American Cancer Society [1], there are approximately 174,650 new prostate cancer patients, and 31,620 people have died of prostate cancer In the past few decades, various methods have been used for prostate cancer detection. Transrectal ultrasound (TRUS) can also be referred to as prostate ultrasound or intrarectal ultrasound. It is used to observe the prostate and surrounding tissues. Ultrasonic transducers (also called probes) send sound waves through the rectal wall to the prostate and surrounding tissues. A computer analyzes the waveform (called an echo) reflected from an organ and converts it into an image that a doctor sees on a video screen [2]. In the treatment of prostate cancer, the use of brachytherapy to implant small radioactive particles in the prostate is made in the surgical environment. Before implanting radioactive particles, the volume and shape of the prostate area need to be calculated accurately. Therefore, it is very important to segment the prostate accurately in the TRUS image. At present, due to the real-time nature of TRUS images, low cost, and easy operation [3], it has become one of the important tools for doctors to diagnose prostate cancer clinically. The guided biopsy is the standard technique for obtaining whole body prostate tissue specimens. On the one hand, during the clinical diagnosis process, the probe was used to observe the prostate on all cross-sections and missing surfaces. It was found that the peripheral gland had low echo changes or other unexplained echo changes inside the glands, and the glands were not paired.

A series of changes in external features such as bulging or interruption of the capsule and dilation or disappearance of the seminal vesicles will all be considered that the detected prostate target may be cancerous. On the other hand, in the imaging process of TRUS images, the image itself has problems such as low signal-to-noise ratio and high similarity between the artifact area and the target area. This problem has caused a lot of prostate image segmentation algorithms to a certain extent. This is a big challenge. Therefore, the high-precision TRUS image real-time automatic segmentation method has become one of the important requirements for the clinical diagnosis of prostate cancer.

The current prostate segmentation is mainly based on contour and shape segmentation, region-based segmentation, and segmentation based on supervised and unsupervised classification methods. F. Arámbula Cosío proposed the active shape model (ASM) to generate a binary image from the initial pose, then Multilinear Principal Component Analysis (MPCA) was used to adjust the initial pose to a grayscale ultrasound image and finally adjust the ASM to a grayscale ultrasound image to generate the final prostate contour [4]. However, this method has a slow initialization speed due to the long initialization time. Although the accuracy is high, the real-time segmentation of the TRUS image still fails to meet the requirements.

Contour and shape-based segmentation method were proposed by Sahirzeeshan Ali and Anant Madabhushi [5]. This method is a collaborative boundary that includes shape priors and a region-based active contour model. In this work, the contour lines that segment the boundaries of objects are represented using the level set method and evolved by minimizing variational energy. The contribution of the new method of this work is to learn the priors of the shape of the target object of interest and integrate it into a hybrid active contours model to segment all overlapping and non-overlapping objects in the image simultaneously. However, this method can only be more accurate if the boundary of the sub-segmentation target is very clear. As shown in Figure 1, the TRUS image has weak borders between adjacent prostate tissue and non-prostate tissue, and features are very similar. As a result, the application of this method to TRUS images is very limited. The region-based segmentation method of Yiqiang Zhan et al., used a deformable shape model to segment the prostate boundary using shape statistics and image texture information. The image texture is represented by Gabor features [6] and is used to distinguish the boundary between the prostate and non-prostate regions. The use of kernel support vector machine (KSVM) is for measuring the probability of belonging to a prostate voxel to distinguish the prostate from surrounding tissues. However, the segmentation accuracy of this method relies more on the statistical information of prostate voxel boundaries. Not only is the manual workload of the statistical information large, but when the statistical data is biased, it will directly lead to accurate segmentation results.
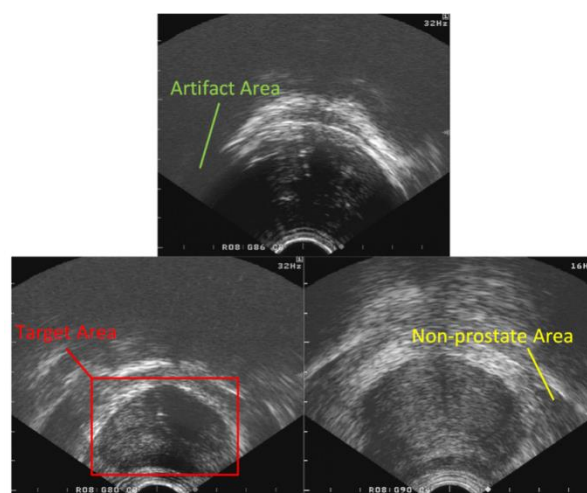
Figure 1. Target area, artifact area, and non-prostate area in transrectal ultrasound (TRUS) image.

In this paper, we propose a novel MFPN. The structure controls the scale and fuses multi-level feature maps to obtain feature fusion modules with rich semantic features. To avoid the artifact region from interfering with the prostate image segmentation, we base our algorithm on precise positioning of the target region before segmentation to solve this problem.

To summarize,

(1)　We propose that a MFPN network can strengthen a feature set by controlling the channel scale on a multi-scale feature map.
(2)　We solve the problem of inaccurate segmentation caused by weak edge features.
(3)　The Anchor size ratio of each anchor point in the Anchor mechanism was modified and a new bounding box regression algorithm was proposed, which not only improves the convergence speed of the target region. At the same time, the process effectively avoids the problem that the target region covers the inside of the prostate's contour and causes the segmentation accuracy to decrease.

2. Methods

The main component of our proposed network structure consists of a multi-scale feature extraction module and a target region screening module. While obtaining the feature map of the multi-scale feature set, our proposed MFPN also improves the network performance. A method to accurately locate the segmented gland target in the region of interest (ROI) is proposed; the method is based on 90 ovf on the characteristics of the TRUS image.

2.1. Multi-Scale Feature Extraction

The basic network used in this research is a deep convolutional neural network that is based on the ResNet101 [7] framework. Although the lower-level feature semantic information is relatively small, the target position is accurate. While the high-level feature semantic information is rich, the target position is rough. To find rich features and accurate target position concurrently, a basic network combined with a multi-scale feature detection structure MFPN is used to process prostate feature extraction in TRUS image segmentation. The network structure proposed in this paper is shown in Figure 2.
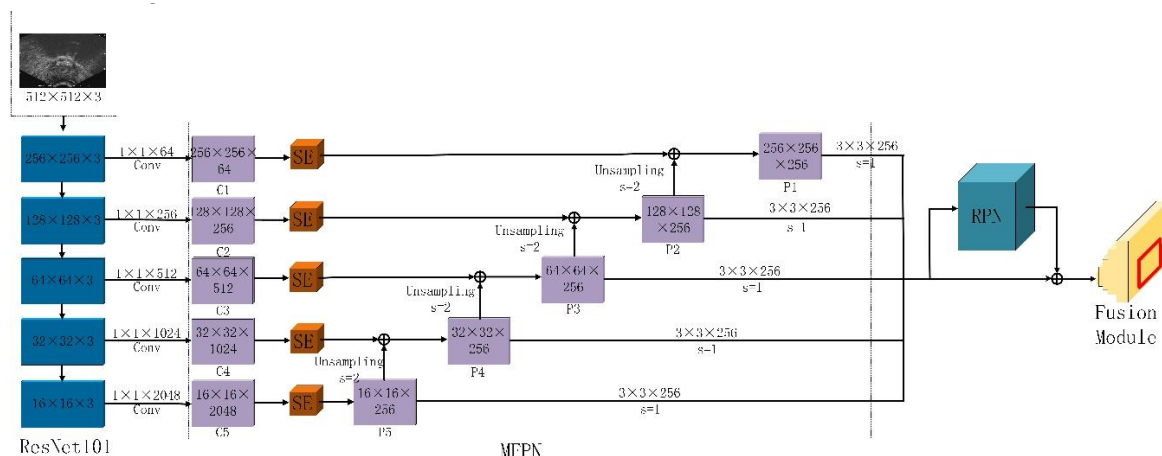
Figure 2. Network structure: The prostate image is the input of the ResNet101 structure. Then, the feature maps (of different sizes) of each layer are inputted into the multi-channel feature pyramid network (MFPN) for feature fusion. The output feature maps are finally inputted into the Region Proposal Networks (RPN) network. The resulting feature fusion module has rich semantic information in the region of interest (ROI).

At each stage, we use the features of the last residual structure to activate the output and represent these (residual module output) as C1, C2, C3, C4, and C5 after obtaining the corresponding output but before fusing C2, C3, C4, and C5. The purpose of the $1 \times 1 \times Cx$ convolution operation is to obtain the same number of channels that are needed to fuse with the upper layer features. That is, the C2 module is $128 \times 128 \times 256$ and the C3 module is $64 \times 64 \times 512$. After C3 undergoes a $1 \times 1 \times Cx$ convolution operation, it is then fused with the C2 module by upsampling with a stride of 2 to obtain the feature map P2. The same applies to the feature maps of P3, P4, and P5. Then, the squeeze-and-excitation (SE) module is added after the $1 \times 1 \times Cx$ convolution. The structure of the module is shown in Figure 3.
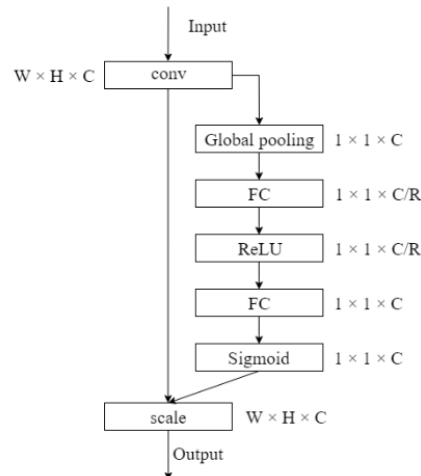

Figure 3. SE module structure.(FC: fully connected layer, ReLU: rectified linear unit).

This module is used to establish a nonlinear relationship between channels and adaptively recalibrate the characteristic response of the channel mode. By controlling the size of the scale, important features are enhanced, while unimportant features are weakened. This makes the directivity of the extracted features stronger. This not only improves the ability to learn features but can also use global information to selectively emphasize feature information and reduce useful functions by suppressing function recalibration. The 3 (a) operation fuses the spatial and channel information in the local area of the original convolution to extract feature information. In 3 (b), the feature is recalibrated using a three-step operation. (i) Compression: To obtain the global information of each channel, the feature is compressed using a global average pool after convolution. (ii) The threshold mechanism is used to set two fully connected layers. The first layer is used to reduce the size of the input to 1 / R of the original size. Then use ReLU to activate. The second layer restores the input to its original size. Normalize weights to 1 after activation. (iii) Weights: Normalized weights are weighted by scaling operations on each channel. We set W, H, C, and R to 1,1,256,16. The enhanced feature map obtained through the above operations is subjected to a $3 \times 3 \times 256$ convolution operation to obtain a multi-scale feature fusion module. Therefore, the module has rich feature information.

2.2. Target Area Screening Module

In Mask R-CNN [8], the Anchors mechanism is applied in RPN, and the role of RPN is to generate candidate regions from images. The original Anchor setup is shown in Figure 4 below:

(**a**)　　　　　　　　　　　　　　　　　　　　　　　(**b**)
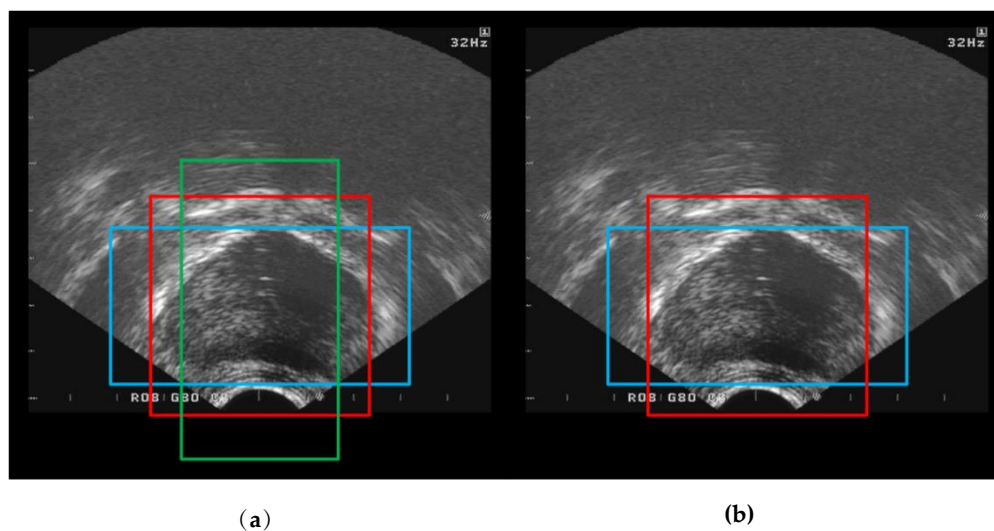
Figure 4. (**a**) The original Anchor mechanism (**b**) The modified Anchor mechanism.

In this paper, we first counted more than 2400 images, and fit the labeled image to the smallest circumscribed rectangle of the segmentation target. Table 1 shows the distribution of the aspect ratio of the smallest circumscribed rectangle. According to the characteristics of segmenting the TRUS image, the proportion of Anchor was changed to {2:1, 1:1}, and the Anchor size was {128, 256, 512}. In the case of the original Anchor mechanism, for a 3×3 sliding window, nine Anchors are generated at the center of the original image, and only six of them are active after modification. Therefore, the number of Anchors to get {p1, p2, p3, p4, p5} is {$256 \times 2 \times 3$, $128 \times 2 \times 3$, $64 \times 2 \times 3$, $32 \times 2 \times 3$, $16 \times 2 \times 3$}, which is {393216, 98304, 24576, 6144, 1536}. Among the obtained Anchors, we use the size of the Intersection-over-Union (IoU) to select samples, those with IoU > 0.9 are positive samples, those with IoU < 0.3 are subsamples, and the remaining Anchors are not used for final training.

Table 1. Statistics of the ratio of target minimum circumscribed rectangle length to width.

Aspect Ratio	>1:1	1:1	<1:1
Quantity	2249	145	6

The scores of the positive samples calculated through the forward propagation of the RPN network are sorted from high to low, and the first 1000 Anchors with the highest score are taken out. The offsets calculated by the forward propagation of the 1000 Anchors via the RPN network are accumulated to On Anchor box, get more accurate Anchor box coordinates. Before returning the box coordinates, the obtained 1000 Anchors are subjected to a non-maximum suppression operation, and duplicate Anchors are eliminated to obtain a non-repeating ROI.

2.3. Target Area Screening Module

During the experiment, the accuracy of the target region can affect the final segmentation result to a certain extent. Even when the target region has a higher accuracy rate, the target will still cause the prostate segmentation result to be poor, as shown in Figure 5.
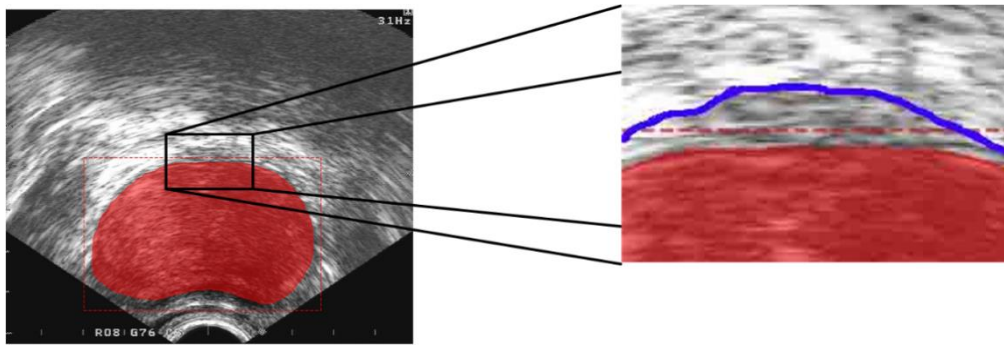
Figure 5. The border of the target area is inside the contour of the prostate. The blue line represents the outer contour of the prostate and the red dotted line represents the outer boundary of the target area.

The blue line in the figure is the external contour of the prostate. The external contour can be clearly seen in the figure. We can see the border of the target area inside the prostate's contour will cause a large deviation in the entire segmentation result. The size of the Ground Truth is the smallest circumscribed rectangle of the prostate segmentation target. Although the accuracy of the border frame of the obtained target area will have higher accuracy, if the size of the target area is smaller than the true value, the image segmentation will decrease in accuracy. In order to avoid this situation, we propose a new anchor regression calculation method based on the anchor mechanism adopted in Mask R-CNN. The positions of the positive sample and the Ground Truth are shown in Figure 6.
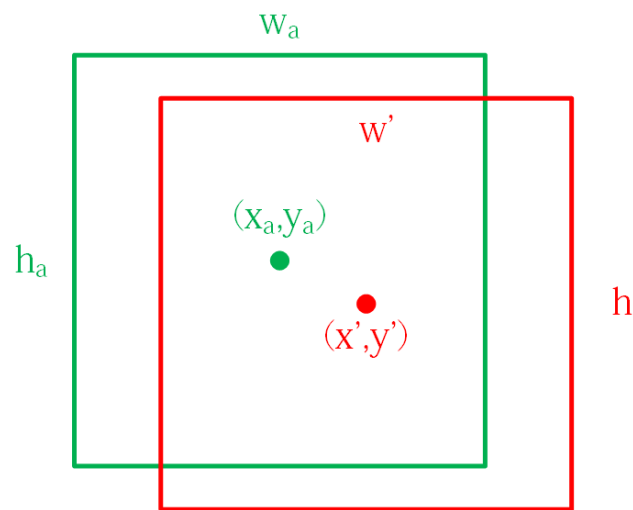
Figure 6. The red box represents the Ground Truth, the green box represents the positive sample. The (x,y) are the center coordinates of the box. The term w stands for width and h stands for height. Positive samples marked with angle *a* are the Ground Truth marked with angle '.

In the first step, a function is used to calculate the offset between the positive sample and the Ground Truth. This function, SmoothL1Loss, characterized as

$$Smooth_{L1}(x) = \begin{cases} 0.5x^2 & \text{if} |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}, \tag{1}$$

The corresponding loss function is shown in the formula

$$Loss = \sum_{i \in \{x,y,w,h\}} Smooth_{L1}(t_i), \tag{2}$$

We let the deviations between the predicted value and the real value be denoted by $|x' - x_a|$ and $|y' - y_a|$. Then, the calculation method of our proposed method is characterized by the set of equations

$$t_x = \frac{\Delta x}{w'}, \tag{3}$$

$$t_y = \frac{\Delta y}{h'}, \tag{4}$$

$$t_w = \log \frac{2w_a}{2w' + |\Delta x|}, \tag{5}$$

$$t_h = \log \frac{2h_a}{2h' + |\Delta y|}, \tag{6}$$

The term t_w can be used to calculate the logarithm of the ratio of the sum of w_a and w' to Δx. This can ensure the bounding box of the true value can increase with width and ensure that the bounding box of the target region is in the regression stage. Even if there are accidental errors, the bounding box will not encompass the entire contour of the prostate, resulting in reduced segmentation accuracy.

3. Results

3.1. Materials

The experimental platform for this research study was Tensorflow1.8 framework, NVIDIA GTX TiTan P40 graphics card, 64-bit Ubuntu 14.04 operating system. The data set was collected together with the cooperative agency, a total of 2400 images. Each image size is 512 × 512 and the pixel size is 0.34 mm. A total 2400 images with annotations were provide. Among them, 2200 images were used as the training set, and the remaining 200 images were used for testing and evaluation. The annotation of the data set passed the review of professional doctors.

3.2. Quality Assessment Methods

The picture shows the intuitive segmentation results of the segmentation method in this paper. By adjusting the size of the scale in the SE module, the outer contour of the final result will be extremely close to the true value: Thus, the problem of inaccurate segmentation caused by low-contrast prostate image edge features can be resolved. By using the precise position of the target region for gland location, the interference of the artifact region on the segmented prostate image is removed. This proves the effectiveness of the method.

3.2.1. Evaluation Method of the Target Area

The central idea of this paper is to use prior information to accurately locate a prostate in an image and then segment the prostate within the precise target area. Therefore, because of the high similarity between artifact areas of the TRUS image and the target region, the characteristics of the boundary features of the target contour are not obvious and the accuracy of the target area has a crucial effect on the proposed method. Next, we calculate the IoU value of the target area and the ground truth obtained from the output results. We calculate the average score of all the pictures in the test set. In this way, the average accuracy of the target area of a network can be calculated.

By combing the IoU results of the test frame and the real frame (p), the formula for calculating the average accuracy of the target area is

$$\beta = \frac{\sum\limits_{N} P_i}{N}, \tag{7}$$

where N represents the number of pictures in the test set and β represents the average accuracy of the target area.

3.2.2. Segmentation Evaluation Method

As in 3.2.1, our segmentation standard is based on manual segmentation by professional doctors. Therefore, in the doctor's standard image segmentation, there will inevitably be a small error term, which we do not discuss here. The area overlap accuracy and contour accuracy are used to evaluate the TRUS graphic segmentation results. These general evaluation methods include the Hausdorff distance [9] (HD), Mean Absolute Distance [10] (MAD), Dice Similarity Coefficient [11] (DSC), Sensitivity [12] (SN), specificity [13] (SP), and Relative Standard Deviation (RSD). Various evaluation indicators are

$$SP = \frac{TN}{TN + FP}, \tag{8}$$

$$SN = \frac{TP}{TP + FN}, \tag{9}$$

where TN represents the public area outside the true segmentation of standard and manual segmentation, and FN represents the algorithm segmentation region. These regions do not include non-public regions other than the standard segmentation region. TP is characterized by a common area representing standard segmentation and algorithmic segmentation, while FP is characterized by an algorithmic segmentation region other than standard segmentation.

$$DSC = \frac{2TP}{(TP + FP) + (TP + FN)}, \tag{10}$$

$$HD(A, B) = \max(h(A, B), h(B, A)), \tag{11}$$

$$h(A, B) = \max(a \in A)\min(b \in B)\|a - b\|, \tag{12}$$

$$h(B, A) = \max(b \in B)\min(a \in A)\|a - b\|, \tag{13}$$

$$A = \left\{a_1, a_2, L, a_p,\right\}, \tag{14}$$

$$B = \left\{b_1, b_2, L, b_p,\right\}, \tag{15}$$

$$MAD = \frac{1}{N}\sum_{i=1}^{N} \|d_j\|, \tag{16}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - 1)2}, \tag{17}$$

where A is the combined coordinates of the manually segmented contour points, B is the combined point coordinates of the segmented contour of the algorithm, N is the number of sampled contour points for manual segmentation, dj is the distance between the manual segmentation of the first j contour point and the corresponding point obtained by algorithmic segmentation. We use Equation (8) and Equation (9) to evaluate classification accuracy. Equations (10), (11), and (16) are used to compare the algorithm segmentation results with the standard. Equation (17) is used to compare the fluctuation of the segmentation results generated by different methods in the test set. σ represents the result of standard deviation, x represents the segmentation result of each picture and the IOU value of ground truth (GT).

3.3. Performance Comparisons

Different candidate frame selection mechanisms and bias calculations have a greater impact on the accuracy of the final target frame generation. Table 2 shows the results of the average accuracy of the target region on the same test set for the target region regression method in Mask R-CNN: This is the method used in this paper. Figure 7 shows the iterative process of the two methods in the same

environment. The results show that the regression method proposed by this method performs better on TRUS images.

Table 2. Target area border average accuracy.

Method	Accuracy (%)
Mask R-CNN	89.57
Our	91.28
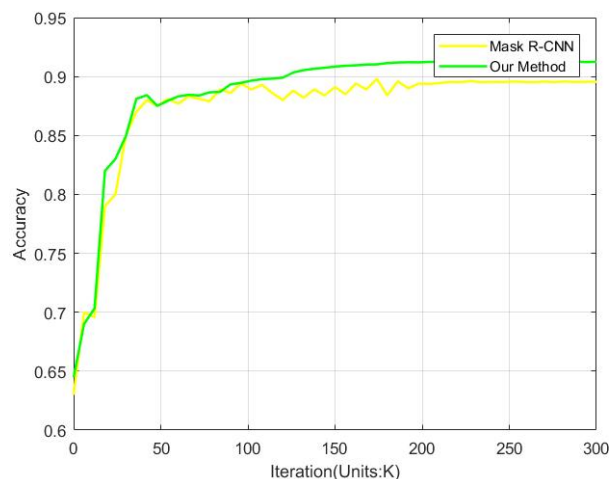
Figure 7. Mask R-CNN and our method's target region border average accuracy.

Table 3 is a detailed comparison of different methods. We can see that some algorithms, based on convolutional neural networks, perform worse than traditional algorithms. The reason is that these methods have not been modified to improve TRUS image segmentation. However, the low segmentation speed of traditional methods is also the reason why those methods cannot be sued for medical applications. Mask R-CNN and the method in this paper use the idea of precise positioning of the target position and segmentation of the TRUS image, within a narrow range, to make the segmentation performance more prominent. The average segmentation speed of Mask R-CNN is 0.330 s, which is 0.012 s faster than our method. The reason for this result is that the addition of the SE module increases the number of parameters and increases the amount of calculations, which adds time. Nevertheless, the time is acceptable in clinical medical applications.

Table 3. Quantitative evaluation of different methods.

| Method	DSC	SP	SN	MAD (mm)	HD (mm)	RSD (10^{-2})	Time(s)
ASM	0.9680	0.9884	0.9694	0.352	1.140	4.259	1.976
FCN [14]	0.8828	0.9275	0.8981	0.423	1.229	10.574	0.462
U-Net [15]	0.8953	0.9395	0.9069	0.405	1.220	9.636	0.483
SegNet [16]	0.9072	0.9501	0.9112	0.401	1.217	8.344	0.497
Deeplab v3 [17]	0.9439	0.9762	0.9386	0.372	1.185	6.015	0.348
Mask R-CNN [8]	0.9775	0.9974	0.9712	0.340	1.112	3.580	0.330
Our	0.9810	0.9983	0.9764	0.332	1.106	2.412	0.342

4. Discussion

We can conclude from the analysis of the experimental results that the reasons for the success of the method proposed in this article can be summarized by the following two points. First, according to the different convolution based feature detection (of different channels) steps, the feature map with strong edges is extracted by controlling the scale size. Finally, the fusion module with strong semantic

features is generated via a fusion technique. Second, according to the statistics of the segmentation target size in the TRUS image, a more reasonable anchoring mechanism and the regression method of the target area are designed to obtain a more accurate target area, which effectively avoids the artifacts and adjacent tissues to produce negative segmentation effects. The segmentation result is shown in Figure 8.

(**a**) (**b**) (**c**)
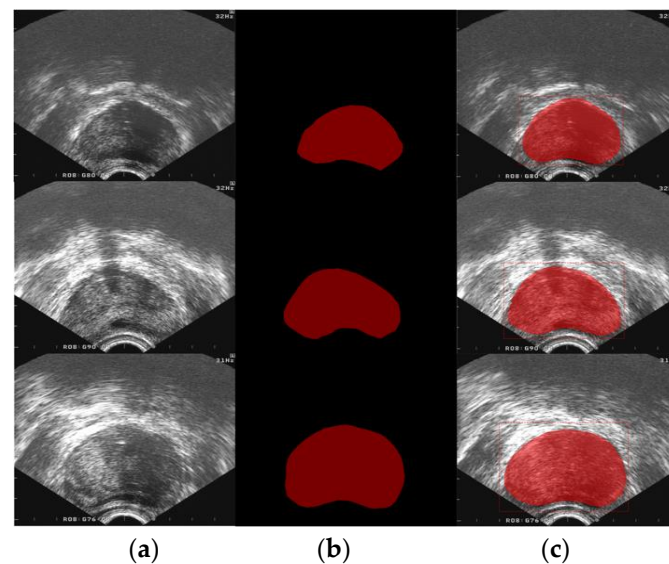
Figure 8. (**a**) Original prostate TRUS images. (**b**) Ground truth from the manual segmentation by a clinically experienced doctor. (**c**) Segmentation results of the proposed method.

We can see (in Table 3) that the effect improvement between the segmentation results of FCN, U-Net, SegNet, and Deeplabv3 in the TURS image has a small margin. Compared with the aforementioned method, Mask R-CNN has greatly improved the key indicators DSC and MAD by 0.336 and 0.039 mm. In addition, from Table 2, Table 3, we can see that the method used in this paper has not only improved 0.135 and 0.022 mm in DSC and MAD indicators compared to Mask R-CNN. We can produce a more intuitive observation effect for different segmentation according to Figure 9. In all the segmentation methods in Figure 9, except Mask R-CNN and the method proposed in this paper, the remaining methods are not instance segmentation. This has no detection function for the segmentation target so that the non-prostate tissue area and the artifact area will have a certain effect on the segmentation result. First, the ASM method uses the prior shape to add to the weighted model and searches for the edge of the target by initializing the shape. Although this method will have a good effect on the segmentation results. But because of its long time-consuming nature, it cannot meet the clinical application. FCN lacks the consideration of global context information, so the segmentation result is the worst. U-Net and SegNet, due to multiple down-sampling and up-sampling, cause detailed information to be lost to varying degrees, resulting in unsatisfactory segmentation results. On the contrary, the Deeplab v3 model using feature fusion means it effectively avoids the loss of detailed information, so the segmentation results have better performance than the above four methods. Finally, due to the limitations of semantic segmentation on prostate image segmentation, this method does not show better performance in segmenting TRUS images. Compared with Mask R-CNN, which also has the function of instance segmentation. In the method proposed in this article, the average accuracy of the target area border has better performance. The reason for this result is that our method uses a new Anchor mechanism and bounding box regression algorithm, thereby improving the accuracy of the border regression in the target area.
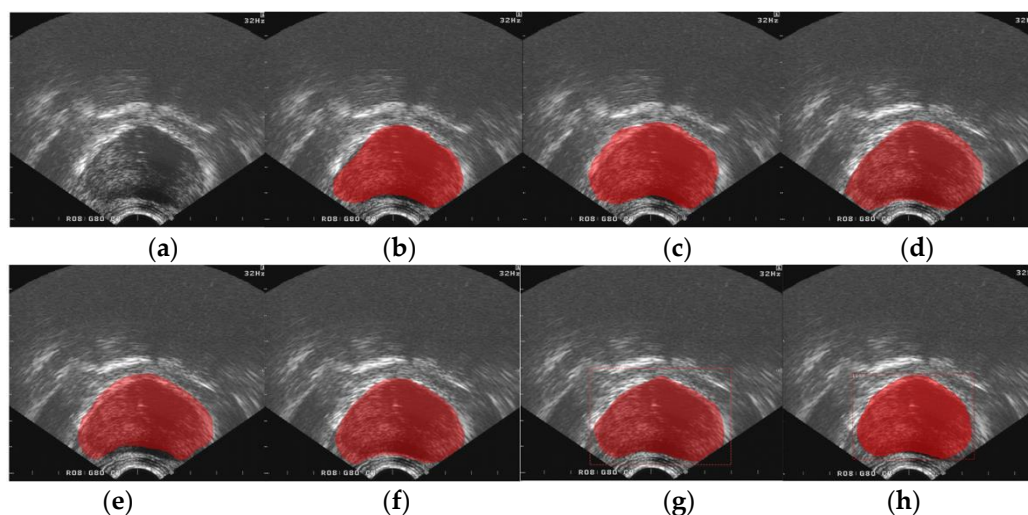
Figure 9. Segmentation results of different methods: (**a**) original prostate TRUS images; (**b**) active shape model (ASM); (**c**) fully convolutional networks (FCN) [14]; (**d**) U-Net [15]; (**e**) SegNet [16]; (**f**) Deeplab v3 [17]; (**g**) Mask R-CNN [7]; (**h**) our methods.

Thus, the network accurately locates and segments the target object, which is very suitable for the segmentation problem of TRUS image. Therefore, the method used in this paper can also be used to solve other types of ultrasound image segmentation problems. Although the network has achieved good results, there are still potential limitations, such as the problem of smaller data sets and insufficient training. Future work can begin with the method of augmenting the data set.

5. Conclusions

The difficulties of segmentation are effectively mitigated by the research method described in this paper. The proposed MFPN structure aims at the different performance of edge features in different channels. By controlling scale and fusing multi-level feature maps we can obtain feature fusion modules with rich semantic features, thus, solving the problem of weak edge features of TRUS images. A method of accurately positioning glandular tissue in a TRUS image and generating a segmentation target region is used, effectively avoiding the problem of interference of the artifact region on the prostate segmentation. Therefore, the algorithm proposed in this research can be effectively applied to clinical medicine.

Author Contributions: Conceptualization, L.G.; validation, L.G. and S.L.; investigation, Z.X.; resources, F.Z.; writing—original draft preparation, S.L.; writing—review and editing, L.G.; supervision, F.Z.; project administration, F.Z. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. American Cancer Society. Available online: http://www.cancer.org/ (accessed on 15 December 2019).
2. Canada Cancer Society. Available online: http://www.cancer.ca/ (accessed on 4 August 2019).
3. Tian, Z.; Liu, L.; Fei, B. Deep convolutional neural network for prostate MR segmentation. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*; SPIE Medical Imaging: Orlando, FL, USA, 2017; p. 101351L.
4. Arámbula Cosío, F.; Hèctor Acosta, G.; Conde, E. Improved edge detection for object segmentation in ultrasound images using Active Shape Models. In Proceedings of the International Symposium on Medical Information Processing & Analysis, Cartagena de Indias, Columbia, 1 January 2015; pp. 9287–9293.

5.   Ali, S.; Madabhushi, A. An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *IEEE Trans. Med. Imaging* **2012**, *31*, 1448–1460. [CrossRef] [PubMed]

6.   Zhan, Y.; Shen, D. Automated Segmentation of 3D US Prostate Images Using Statistical Texture-Based Matching. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention-MICCAI, Montréal, QC, Canada, 15–18 November 2003.

7.   He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.

8.   He, K.; Gkioxari, G.; Dollar, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *8*, 1–6.

9.   Belogay, E.; Cabrelli, C.; Molter, U.; Shonkwiler, R. Calculating the Hausdorff distance between curves. *Inf. Process. Lett.* **1997**, *64*, 17–22. [CrossRef]

10. Yan, P.; Xu, S.; Kruecker, J.; Turkbey, B. Optimal search guided by partial active shape model for prostate segmentation in TRUS images. In Proceedings of the Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling, Lake Buena Vista, FL, USA, 8–10 February 2009; pp. 7261–7272.

11. Dice, L.R. Measures of the Amount of Ecologic Association between Species. *Ecology* **1945**, *26*, 297–302. [CrossRef]

12. Badiei, S.; Salcudean, S.E.; Varah, J.; Morris, W.J. Prostate Segmentation in 2D Ultrasound Images Using Image Warping and Ellipse Fitting. In Proceedings of the International Conference on Computing and Computer-Assisted Intervention, Copenhagen, Denmark, 1 February 2006; pp. 17–24.

13. Diaz, K.; Castaneda, B. Semi-automated segmentation of the prostate gland boundary in ultrasound images using a machine learning approach. In Proceedings of the Medical Imaging 2008: Image Processing. International Society for Optics and Photonics, San Diego, CA, USA, 17–19 February 2008; pp. 9144–9147.

14. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *39*, 640–651.

15. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing & Computer-assisted Intervention, Munich, Germany, 18 May 2015; pp. 234–241.

16. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Scene Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]

17. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking Atrous Convolution for Semantic Image Segmentation. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1–14.