

Automatic Segmentation of Prostate Magnetic Resonance Imaging Using Generative Adversarial Networks

Wei Wang

Tongji Hospital of Tongji University School of Medicine

Mingang Wang

Huashan Hospital of Fudan University

Xiaofen Wu

Tongji Hospital of Tongji University School of Medicine

Xie Ding

School of Wonders information company

Xuexiang Cao

School of Wonders information company

Lei Wang

Tongji Hospital of Tongji University School of Medicine

Jingyi Zhang

School of Wonders information company

Peijun Wang (✉ pjwangtongji@sina.com)

Tongji Hospital of Tongji University School of Medicine <https://orcid.org/0000-0002-4012-9234>

Research article

Keywords: Automatic segmentation, Generative adversarial networks, Magnetic resonance imaging, Prostate

Posted Date: January 7th, 2020

DOI: <https://doi.org/10.21203/rs.2.12243/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Clinical Imaging on February 1st, 2021. See the published version at <https://doi.org/10.1016/j.clinimag.2020.10.014>.

RESEARCH ARTICLE

Automatic Segmentation of Prostate Magnetic Resonance Imaging Using Generative Adversarial Networks

Wei Wang¹, Mingang Wang², Xiaofen Wu³, Xie Ding⁴, Xuexiang Cao⁴, Lei Wang³, Jingyi Zhang⁴ and Peijun Wang^{1*}

Abstract

Background: Automatic and detailed segmentation of the prostate using magnetic resonance imaging (MRI) plays an essential role in prostate imaging diagnosis. However, the complexity of the prostate gland hampers accurate segmentation from other tissues. Thus, we propose the automatic prostate segmentation method SegDGAN, which is based on a classic generative adversarial network (GAN) model.

Methods: The proposed method comprises a fully convolutional generation network of densely connected blocks and a critic network with multi-scale feature extraction. In these computations, the objective function is optimized using mean absolute error and the Dice coefficient, leading to improved accuracy of segmentation results and correspondence with the ground truth. The common and similar medical image segmentation networks U-Net, fully convolution network, and SegAN were selected for qualitative and quantitative comparisons with SegDGAN using a 220-patient dataset and the publicly available dataset PROMISE12. The commonly used segmentation evaluation metrics Dice similarity coefficient (DSC), volumetric overlap error (VOE), average surface distance (ASD), and Hausdorff distance (HD) were also used to compare the accuracy of segmentation between these methods.

Results: SegDGAN achieved the highest DSC value of 91.66%, the lowest VOE value of 23.47% and the lowest ASD values of 0.46 mm with the clinical dataset. In addition, the highest DCS value of 88.69%, the lowest VOE value of 23.47%, the lowest ASD value of 0.83 mm, and the lowest HD value of 11.40 mm was achieved with the PROMISE12 dataset.

Conclusions: Our experimental results show that the SegDGAN model outperforms other segmentation methods

Keywords: Automatic segmentation, Generative adversarial networks, Magnetic resonance imaging, Prostate

Background

Prostate cancer (PCa) ranks second after lung cancer as the most common malignant tumor in men. According to global cancer statistics, 1.3 million new cases of prostate cancer and 359,000 associated deaths were recorded in 2018 [1]. Because early stage prostate cancer can be effectively diagnosed and controlled, accurate detection of emerging prostate cancers is particularly important. Currently, MRI is the most common imaging method for diagnosing PCa [2, 3]. Due to its relatively high soft-tissue resolution, MRI can clearly show internal structures and surrounding tissues of

the prostate. Hence, MRI is often used to diagnose PCa and effectively detects the locations of cancers [4].

Segmentation of prostate structures in MRI can facilitate clinical decisions. Thus, the accuracy of automatic prostate segmentation is critical for the detection and diagnosis of PCa. Traditionally, prostate MRI data were segmented manually in a time-consuming process that requires professional experience of the observer. To address this, machine learning and deep learning technologies have recently been developed considerably for image segmentation [5, 6, 7]. Moreover, in recent years, deep convolutional neural networks (CNN) have been frequently applied to medical image segmentation [8, 9, 10, 11]. For example, the fully convolution network (FCN) proposed by Shelhamer et al. [12] is an

*Correspondence: pjwangtongji@sina.com

¹Department of Radiology, Tongji Hospital of Tongji University School of Medicine, No.389 Xincun Road, 200065 Shanghai, China

Full list of author information is available at the end of the article

end-to-end network that classifies images at the pixel level, thereby solving the problem of semantic segmentation. As an alternative, Ronneberger et al. [13] applied a U-Net network to medical image segmentation. U-Net is performed with a contracting path to capture contextual information from images and an expanding path that is used to accurately localize the segmented target. To the same end, Milletari et al. [14] developed V-Net, which is a three-dimensional (3D) end-to-end medical segmentation method. This network introduces the Dice coefficient as a novel objective function and uses a residual learning method to accelerate convergence. Various other automatic algorithms have been proposed for the segmentation of prostate MRI data [15, 16, 17, 18]. Among these, Zhu et al. [19] developed a deeply supervised two-dimensional (2D) U-Net model, and Milletari et al. [14] introduced V-Net as a 3D prostate image segmentation model. Further, Zhu et al. [20] proposed a recursive neural networks (RNNs) UR-Net model, they treat prostate image slices as data sequences and use intra-slice features to improve the performance of prostate segmentation. Cheng et al. [21] presented an active appearance model (AAM) based on atlas and a deep learning model, and demonstrated higher precision for prostate segmentation on MRI. Vincent et al. [22] similarly developed a MRI segmentation system based on a fully automatic model, which was built on the AAM established by the MCCAII2012 team’s manual segmentation example using the minimum description length group. In a later study by Zhu et al. [23], a fully automated algorithm for peripheral zone and transition zone segmentation of prostate MRI achieved satisfactory performance. To date, these methods have been tested on clinical datasets and the public prostate dataset PROMISE12, and have achieved good performance. However, due to the complexity of MRI, these methods remain unable to actively fit natural structures of the prostate gland in MRI.

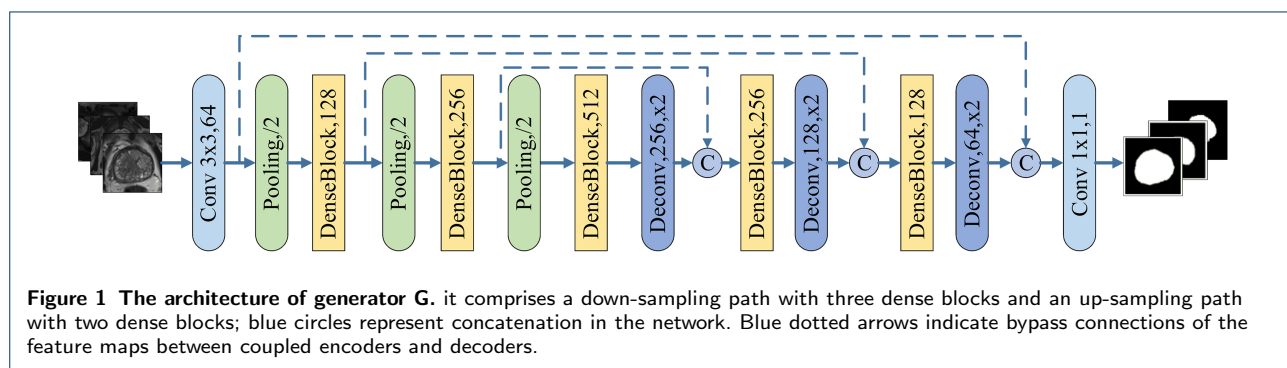
To address this shortcoming, we propose the new end-to-end architecture SegDGAN, which was inspired

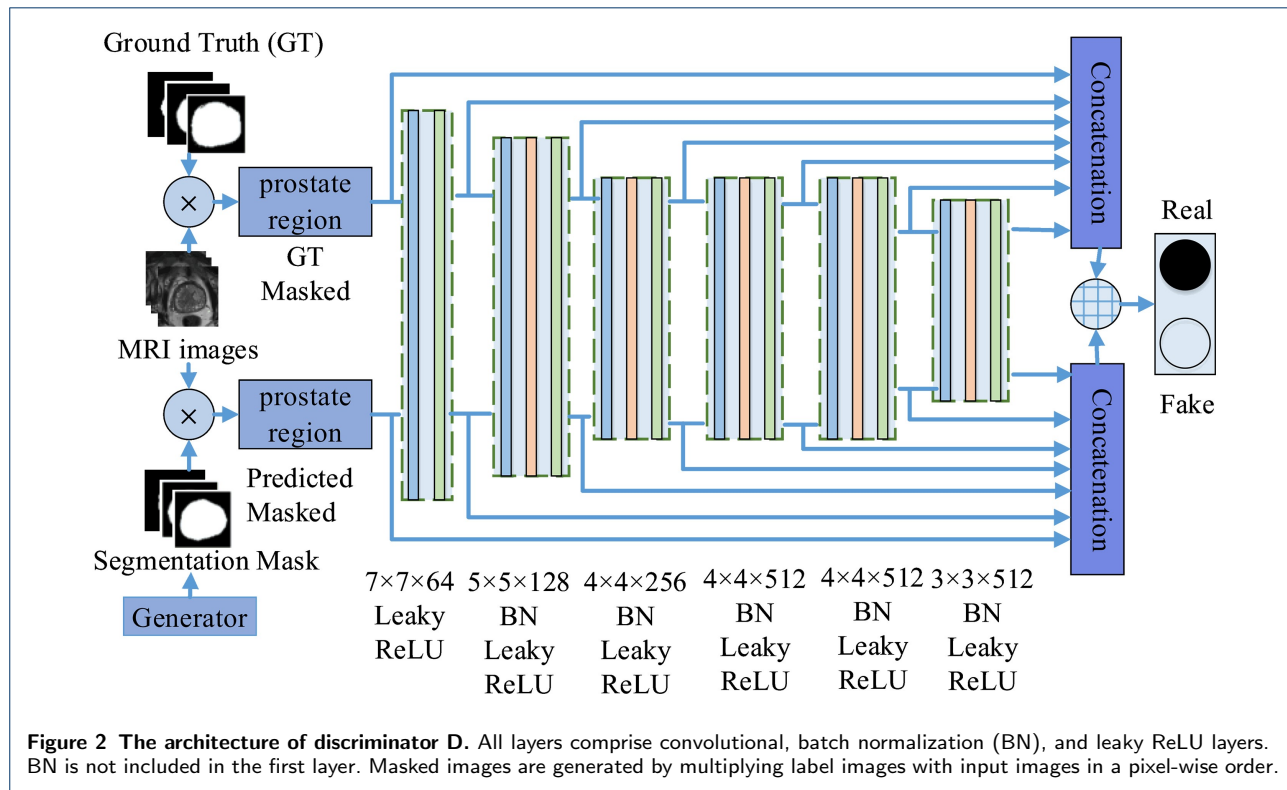
by the SegAN model [24] and was used to segment prostate glands. In this study, we enhanced the performance of prostate gland segmentation by optimizing the original GAN [25] network structure and the objective function. The resulting model overcomes the problems associated with natural prostate structures, and provides an architecture that performs better than previously described algorithms. The contributions of this work are listed as follows:

- In the generator network (segmentor network), we introduce a fully convolutional network that is similar to the U-Net codec structure, and then performed end-to-end training.
- We applied the dense connection block structure to achieve direct connections between single layers and all subsequent layers, thus contributing additional supervisory information. This design reduces redundant feature learning to alleviate the gradient vanishing problem and to prevent model over-fitting.
- We developed a multi-scale feature connection in the discriminator network and added Dice control to the objective function, leading to improved segmented binary mask images.

Methods

Our proposed prostate segmentation method SegDGAN comprises a generator \mathbf{G} network and a discriminator \mathbf{D} network, and these are trained concomitantly. The discriminator is designed to distinguish between ground truth images and those produced by the generator. In contrast, the generator generates images that are similar to ground truth images, and is hence designed to constrain the discriminator. Following adversarial training, both of these networks ensure that the output images from the generator are constantly approaching the ground truth of the prostate gland. The network structure is shown in Fig. 1 and Fig. 2.





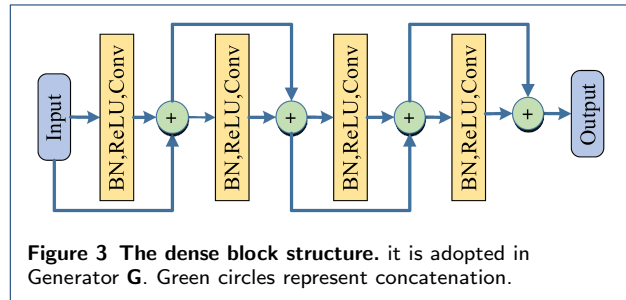
SegDGAN Architecture

Generator

The generator **G** is a segmentation network in which end-to-end training is performed. **G** uses an encoder-decoder structure of a previously described U-Net [13] based on a reported fully convolutional network [12], as shown in Fig. 1.

This network includes down-sampling and up-sampling processes. The down-sampling process includes a convolutional layer of 3×3 convolution kernels, three maximum pooling layers, and three densely connected blocks. The up-sampling process includes three deconvolution layers and three densely connected blocks followed by a 1×1 convolution kernel. The U-Net is applied to the generator **G** to add skip connections between the encoder and the decoder. Feature maps from the down-sampling path are connected to those from the up-sampling path in the symmetry pattern. This design allows extraction of image features at different scales during down-sampling, and provides a view of the same size as the input image in the up-sampling process.

The internal structure of the dense block is similar to DenseNet [26], as illustrated in Fig. 3, and contains four layers in a dense block. Each layer comprises batch normalization (BN), a rectified linear unit (ReLU), and a 3×3 convolution kernel. Direct connections



from a single layer are made with all subsequent layers to achieve additional supervision of the dense block structure.

Discriminator

Discriminator **D** is a multi-dimensional feature extraction network with six layers. Each layer comprises a convolution layer, a BN, and a leaky ReLU activation layer. But in the first layer, BN is not considered and only the other two layers are included. Convolution kernel sizes are 7×7 , 5×5 , 4×4 , and 3×3 . Fig. 2 shows the structural details of the discriminator **D** with the components of each convolutional layer.

Optimized Objective Function

The objective function of conventional GANs [25] is defined as follows:

$$\min_G \max_D L(G, D) = \mathbb{E}_{x \sim P_{data}} [\log D(x)] + \mathbb{E}_{z \sim P_z} [\log(1 - D(G(z)))] \quad (1)$$

where x represents the real images from distribution P_{data} and z is random data that is introduced to determine the distribution of the generator. z usually satisfies a random noise distribution P_z . $G(z)$ represents the differential function for the generator. Finally, $D(x)$ represents the probability that the input image x originates from the training data sample rather than the image generated by the generator.

In this study, we applied the GAN network to prostate segmentation. The generator \mathbf{G} is a mapping term that relates x original MRI images to y segmented binary mask images. The discriminator \mathbf{D} is used to classify the image and produces a binary classification $\{0, 1\}^k$ for each data point $\{x, y\}$. In this classification, 1 indicates that y is a label image from the training sample and 0 indicates that y is a \mathbf{G} generated image. Finally, k represents the number of decisions. Accordingly, the objective function of the GAN for this segmentation problem is defined as follows:

$$\min_G \max_D L(G, D) = \mathbb{E}_{x, y \sim P_{data}(x, y)} [\log D(x, y)] + \mathbb{E}_{x \sim P_{data}(x)} [\log(1 - D(x, G(z)))] \quad (2)$$

As suggested previously [24], we added a multi-scale feature connection and a Dice coefficient control to the objective function to achieve better training results as follows:

$$\min_G \max_D L(G, D) = \frac{\lambda}{N} \sum_{n=1}^N \ell_{dice}(G(x_n), y_n) + \frac{1}{N} \sum_{n=1}^N \ell_{mac}(f_D(x_n \circ G(x_n)), f_D(x_n \circ y_n)) \quad (3)$$

where N is the number of the training images and x_n and y_n refer to the MRI images and ground truth label maps, respectively. The symbol “ \circ ” in the formula represents the point multiplication of the matrix. The function of the discriminator f_D was used to extract hierarchical features from input data x , and ℓ_{mac} is the mean absolute error (MAE) between the predict-

ed prostate region and the real prostate region, and is defined as follows:

$$\ell_{mac}(x, x') = \frac{1}{M} \sum_{i=1}^M \|f_D^i(x) - f_D^i(x')\|_1 \quad (4)$$

where M is the number of layers of the discriminator \mathbf{D} , and $f_D^i(x)$ is the feature map of the i -th layer of \mathbf{D} . In this equation, ℓ_{dice} is the Dice coefficient that characterizes the degree of similarity between the generated binary mask image and the ground truth. λ is an adjustment coefficient that optimizes the weight effect between ℓ_{mac} and ℓ_{dice} . Accordingly, ℓ_{dice} is defined by the following equation:

$$\ell_{dice}(x, x') = 1 - \frac{2 \sum_{i=1}^S x_i x'_i + \varepsilon}{\sum_{i=1}^S (x_i + x'_i) + \varepsilon} \quad (5)$$

where S represents the total number of pixels in images and x_i is the predicted feature mapping probability. The term x'_i is the pixel value of the labeled segmentation sample and has a value of $\{0, 1\}$. Finally, ε is a stability constant that is used to prevent the denominator from being 0.

Segmentation Evaluation Metrics

We employed four widely used segmentation evaluation metrics to evaluate and compare the segmentation results of various neural networks. The metrics Dice similarity coefficient (DSC) [27] and volumetric overlap error (VOE) [28] are both indicators of volume similarities. The other two metrics, average surface distance (ASD) [29] and Hausdorff distance (HD) [30], are relevant to distance measurements. DSC is designed to calculate similarities between pairs of contour regions and is defined as follows:

$$DSC(M, N) = \frac{2|M \cap N|}{|M| + |N|} \quad (6)$$

where M and N correspond with segmentation results and ground truth masks, respectively. The VOE is used to calculate ratios between intersections and unions of the two images, and is calculated as follows:

$$VOE(M, N) = (1 - \frac{|M \cap N|}{|M \cup N|}) \times 100\% \quad (7)$$

ASD is used to compute average surface distances between binary objects in two images. ASD is defined as follows:

$$ASD(S(N), S(M)) = \frac{\sum_{n \in S(N)} \min_{m \in S(M)} d(n, M)}{|S(N)|} \quad (8)$$

$S(M)$ and $S(N)$ denote surface voxels of the segmentation results and the ground truth masks. $d(N, M)$ is the Euclidean distance between the two images. The HD is used to measure the distance between the two images, and $HD(M, N)$ is calculated using the following equation:

$$HD(M, N) = \max\left\{ \max_{m \in S(M)} d(m, S(N)), \max_{n \in S(N)} d(n, S(M)) \right\} \quad (9)$$

where $d(m, S(N))$ represents the Euclidean distance between m and $S(N)$, and $d(n, S(M))$ represents the Euclidean distance between n and $S(M)$.

Experiments

Dataset

The experimental data were collected during clinical prostate examinations in the hospital from July 2012 to June 2018. Data were collected for a total of 220 patients, including 121 healthy subjects and 99 prostate cancer patients, and were processed anonymously. All of prostate MRI images were acquired on a 3.0-T MR imaging system (SIEMENS Verio 3.0 Tesla (T)), and the running software version is syngo MR B17. Transverse T2-weighted (T2-w) images were acquired using Turbo spin-echo sequence with the following parameters: repetition time, 2900-4030 ms; echo time, 96-106 ms; slice thickness, 3-4.0 mm; intersection gap, 3.6-4.8 mm; and matrix, 640×640. Prostate boundary masks were cross-labeled by two expert clinicians with more than five years clinical experience, and labeled images used as the ground truth in the form of binary masks.

To verify the generalizability of the model, we used the public dataset PROMISE12, which was provided by 2012 MICCAI challenges [31]. Although the images in the dataset were collected using different devices with different resolutions and scan protocols, the data has greater universality, and it was not used in the training process. Thus, we considered 20 randomly selected MRI images from this dataset as suitable for verifying the generalizability of our model. These images were all as large as those in the clinical test dataset.

Data preparing

To facilitate model training, we resized the images to a fixed resolution of 0.3 mm×0.3 mm×2 mm. Subsequently, image sizes were fixed to 512×512 and 32 slices were selected to exclude those that do not contain prostate. The final dataset contained 7040 selected images. We then split the dataset randomly into a training dataset, a validation dataset, and a test dataset. Twenty patients with a total of 640 images were selected for testing, and the remaining data were used for verification and training. Hyperparameters were selected using the validation data set and were finally trained with the training data.

Because insufficient training samples would lead to over-fitting of the model, we performed data augmentation to increase the training sample artificially. The data augmentation transformations in our model included translation, rotation, distortion, and scaling.

Training

All model training was performed on a GTX 1080TI GPU with 64GB RAM and CUDA edition 9.0 software. The model was implemented on the open source deep learning framework PyTorch [32]. To expedite model training and save memory, input image sizes were reshaped to 128 × 128. Because the dense block structure used in our model shares weights among sub-networks, it can suppress over-fitting [26]. Thus dropout was not considered in our model. In addition, we used the Adam optimizer to achieve better training performance. The initial learning rate was set at 0.003, with decay every 40 rounds at a decay rate of 0.5 until the learning rate was fixed at 0.0000001. In consideration of machine performance and training speed, we set the batch size at 8 for the experimental training dataset.

Results

To evaluate the performance of our SegDGAN model, we compare it to the common segmentation network U-Net [13] and the FCN [12]. We also compared SegDGAN with the reference model SegAN, which was the precursor of our model [24]. All neural network models were trained and tested on the same dataset and were compared according to the same metrics.

Qualitative Comparison

Segmentation was performed using the SegDGAN model and the other neural networks U-Net, FCN, and SegAN based on experiment data, as shown in Fig. 4. Segmentation results are shown for various prostate locations, including the center, the apex, and the base.

In Fig. 5, we present segmentation results that were generated using the public dataset PROMISE12. As

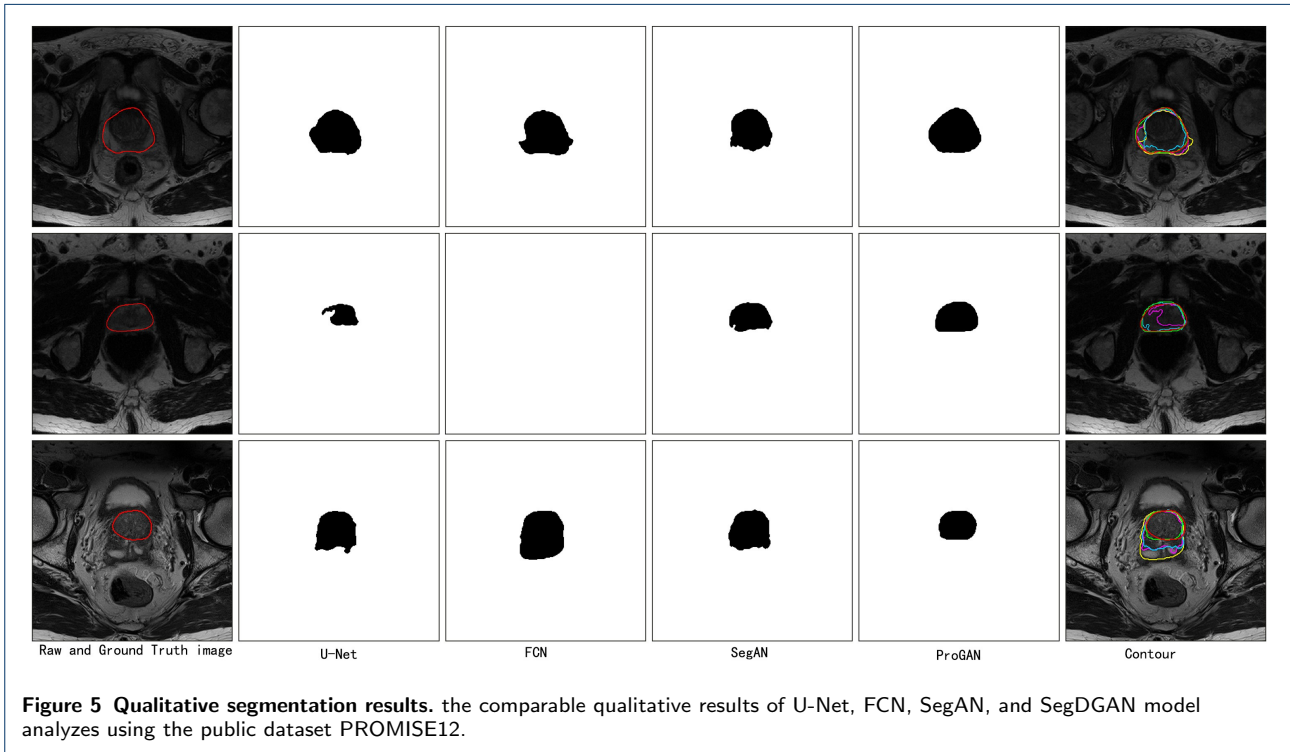
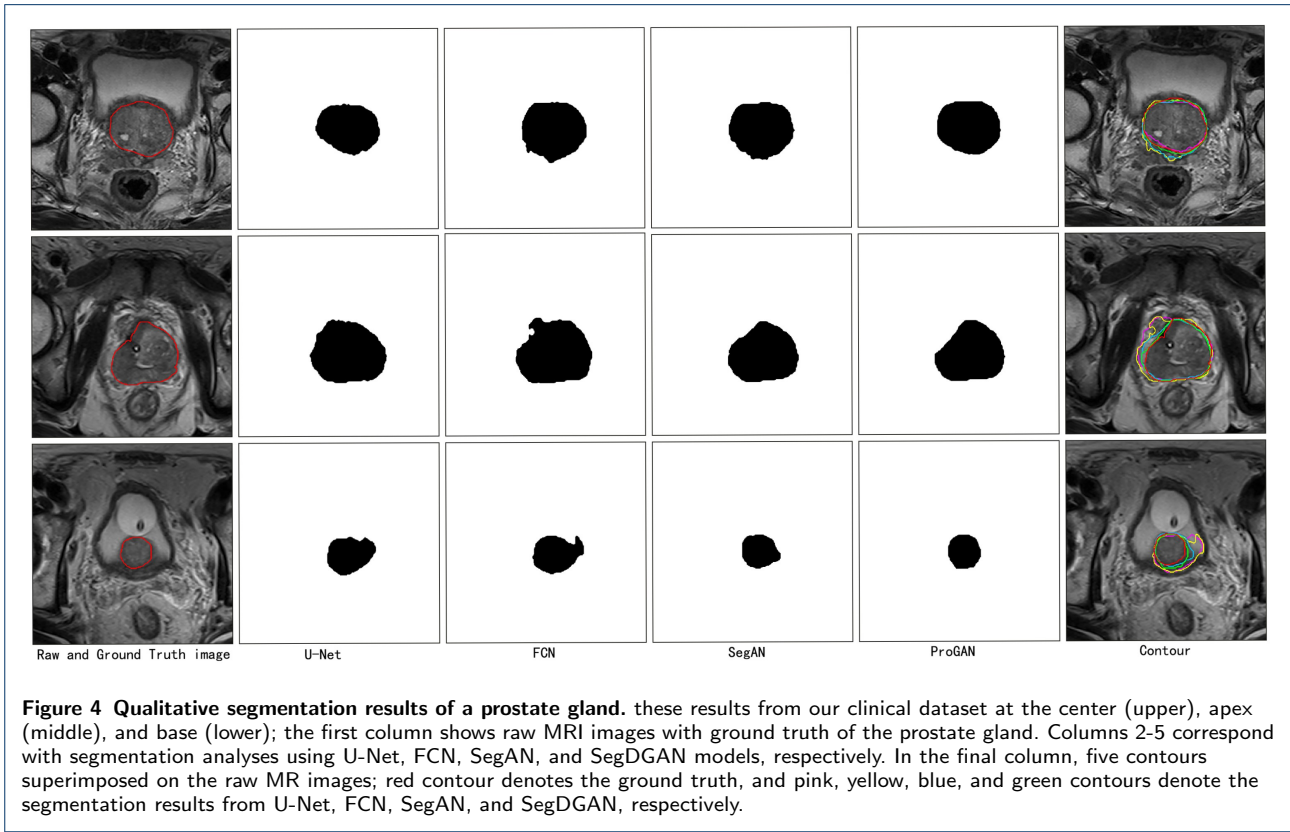


Table 1 Evaluation Metrics For Model Analyses Using Our Test Dataset And The Publicly Available Dataset PROMISE12

Database	Methods	ASD (mm)	VOE (%)	HD (mm)	DSC		
					Mean	Median	Maximum
Experiment	U-Net	0.62	89.43	11.05	0.8950	0.9053	0.9327
	FCN	1.28	24.81	18.03	0.8859	0.9026	0.9296
Data	SegAN	0.58	19.23	11.04	0.9014	0.9053	0.9332
	SegDGAN	0.46	15.98	13.89	0.9166	0.9259	0.9428
PROMISE12	U-Net	0.95	28.88	12.73	0.8596	0.8732	0.9382
	FCN	5.58	38.29	82.37	0.8413	0.8571	0.9530
	SegAN	0.997	29.28	15.68	0.8464	0.8603	0.9305
	SegDGAN	0.83	23.47	11.40	0.8869	0.8897	0.9467

indicated above, the images in this dataset were collected under differing clinical scanning conditions, and were not used during the training process. Hence, these data are more representative of the actual clinical application.

Segmentation results of SegDGAN model analyzes are shown in the fifth columns of Fig. 4 and Fig. 5. These figures indicate that our SegDGAN model outperforms the other methods at all locations. In particular, the second row and third column of Fig. 5 show no relevant results because FCN segmentation failed entirely and no prostate gland were detected.

Quantitative Comparison

We evaluated our model SegDGAN and the models U-Net, FCN and SegAN according to the metrics listed in section . All metrics for the four models were calculated using the experimental dataset. To compare and assess the generalizability of these methods, we also calculated evaluation metrics using the PROMISE12 dataset, as shown in Table 1. These data show that the highest DSC values, lowest VOE values, and lowest ASD values were obtained on our SegDGAN model with the experiment dataset. In contrast, the highest DCS values, lowest VOE values, lowest ASD values, and lowest HD values were achieved on the public dataset PROMISE12.

Discussion

In this study, we developed an automatic segmentation algorithm for prostate gland, based on generative adversarial networks (GANs). To test our algorithm, we compared qualities and quantities of segmentation results with those from other models. Currently, FCN [12] and U-Net [13] networks are the most widely used for medical image segmentation. Because our method was generated on the basis of published models [24], we made comparisons with the established segmentation algorithms U-Net, FCN, and SegAN. Generalizability of neural networks remains a challenging problem, especially for medical imaging analyses [33]. To demonstrate the clinical usability of the present four

models, we tested the models with our clinical dataset and with the public dataset PROMISE12. Because the PROMISE12 dataset contains images that were generated with differing scanning parameters and clinical conditions, it was omitted from model training, but its heterogeneity was favorable for model testing.

The experimental results presented in Fig. 4 and Fig. 5 show that the proposed SegDGAN method effectively performs automatic segmentation of prostate glands. Even with difficult prostate MRI images that had blurred borders and heterogeneous distributions of pixel intensity inside and outside the prostate, SegDGAN returned good segmentation results that were highly consistent with the ground truth. Hence, the present segmentation results benefited from the improved model design and the optimized objective function. In particular, we introduced a dense connection block structure into the generator network to realize direct connections between single layers and all subsequent layers, thus contributing additional monitoring information, reducing redundant feature learning, mitigating gradient disappearance issues, and preventing model over-fitting [26]. Multi-scale linking is widely used in discriminators to identify hierarchical features and calculate loss functions. Using these multi-layer features, loss functions can capture spatial relationships between pixels. We also added Dice control to our objective functions to improve training performance for SegDGAN. The quality results presented herein demonstrate that SegDGAN is more effective and robust for prostate gland segmentation than the other segmentation methods, and has better segmentation performance. Moreover, our segmentation results from SegDGAN were smoother and more continuous, and had a higher degree of coincidence with the ground truth.

Based on analyzes of evaluation metrics shown in Table 1, application of SegDGAN to the experiment dataset achieved the highest DSC value of 91.66%, the lowest VOE value of 15.98%, and the lowest ASD value of 0.46 mm. DSC values indicate degrees of similarity between two contour regions. Hence, the

high DSC value for segmentation using SegDGAN demonstrates greater similarity with the ground truth, and greater segmentation accuracy than that achieved by the other methods. Similarly, the low VOE value for SegDGAN indicates lower volumetric overlap error between segmentation results and reference masks. Finally, the ASD value of SegDGAN was the lowest among models, suggesting relatively low error of prostate segmentation results. Taken together, these metrics demonstrate that the SegDGAN method significantly improves the performance of prostate segmentation compared with previously reported methods. Primarily, this improvement reflects the use of a dense block structure, which increases the depth of the layer. Moreover, we used a GAN model with a generator and a discriminator. This generator model facilitates evaluations of distribution characteristics of marked images. Specifically, extra supervised layers were added to the network through the discriminator, leading to strong constraints on the network. This design improved the accuracy and completeness of feature learning from images. Finally, addition of MAE and Dice to the loss function improved the proximity of the segmentation result to the masked image. The present metrics analysis accordingly indicated that the segmentation results from our method were closer to the ground truth and had relatively low segmentation errors, suggesting greater stability and higher repeatability. Quantitative results analyses were also in good agreement with previous visual observations. In analyzes of images from the public dataset PROMISE12, the SegDGAN method produced the highest DCS value of 88.69%, the lowest VOE value of 23.47%, the lowest ASD value of 0.83 mm, and the lowest HD value of 11.40 mm. These data confirm that SegDGAN has improved robustness and repeatability compared with other methods, during application to both experimental and public datasets.

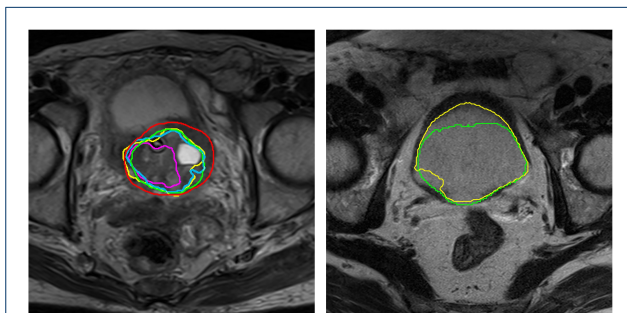


Figure 6 Segmentation results of the worst case.

In Fig. 6, we present worst results of our segmentation algorithm. The image on the right of this figure shows a case with lesions. The brightness of the

lesion area is relatively large, leading to a strong influence on the surrounding tissue. These conditions introduced error to our segmentation results. Our algorithm is also limited to two-dimensional information, and this could cause mis-segmentation, especially in tissues with structures and textures that are similar to those of the prostate gland, as shown in the left image of Fig. 6. In future studies, we will consider three-dimensional information and will optimize the algorithms for lesions. Finally, our method suffers disadvantages of longer training times and higher numbers of convergence rounds, again warranting further studies to optimize and improve future computational models.

Conclusion

In this paper, we present a novel segmentation network for GAN-based images. In this SegDGAN model, we used a fully convolutional network comprising densely connected blocks to construct a generator network. We also used a multi-scale and multi-level convolutional network to construct a discriminator network. Subsequently, we optimized corresponding objective functions to improve segmentation performance. When applied to MRI prostate image segmentation, SegDGAN segmentation results were close to the ground truth. Moreover, in comparisons with the established medical image segmentation networks U-Net, FCN, and SegAN, SegDGAN gave higher prostate segmentation accuracy when applied to our own clinical data and to that in the public dataset PROMISE12. In particular, analyzes with the PROMISE12 dataset, which was not used during model training, indicted that SegDGAN is more representative and generalizable than the other models. These qualitative and quantitative comparisons show that the SegDGAN model is a more effective and robust algorithm for prostate segmentation, and warrant consideration for automatic segmentation of prostate glands. This study is highly relevant to intelligent MRI analyses and diagnoses of prostate disease.

Abbreviations

MRI: Magnetic resonance imaging; GAN: Generative adversarial network; DSC: Dice similarity coefficient; VOE: Volumetric overlap error; ASD: Average surface distance; HD: Hausdorff distance; PCa: Prostate cancer; CNN: Convolutional neural network; FCN: Fully convolution network; 3D: Three-dimensional; 2D: Two-dimensional; RNNs: Recursive neural networks; AAM: Active appearance model; BN: Batch normalization; ReLU: Rectified linear unit; MAE: the mean absolute error; T2-w: T2-weighted

Acknowledgements

We acknowledge Tongji Hospital of Tongji University for providing the experimental dataset.

Author's contributions

PW and WW designed details of the research and carried out the majority of experiments. GW, XW and LW participated in the design and coordination of experimental work. JZ, LW, XW, XD and XC developed new algorithms. WW, GW, XD and XC contributed to data acquisition, interpretation and statistical analysis. WW and XC drafted the manuscript. JZ and PW made critical revision for important intellectual content. PW was the project administrator and handled funding and supervision. All authors have reviewed and approved the submitted manuscript for publication.

Funding

The work was supported by Science and Technology Commission of Shanghai Municipal (No. 17411952300).

Ethics approval and consent to participate

The dataset used in this study was approved by the Ethics Committee at Tongji Hospital of Tongji University. The requirement for informed patient consent was waived because it was a retrospective study without patient interactions.

Competing interests

The authors declare that they have no competing interests.

Availability of data and material

The datasets analysed during the current study are not publicly available due to relevant data protection laws.

Consent for publication

Not applicable.

Author details

¹Department of Radiology, Tongji Hospital of Tongji University School of Medicine, No.389 Xincun Road, 200065 Shanghai, China. ²Huashan Hospital of Fudan University, No.12 urumqi middle Road, 24105 Shanghai, China. ³Department of Information Section, Tongji Hospital of Tongji University School of Medicine, No. 389 Xincun Road, 200065 Shanghai, China. ⁴Department of Medical Big Data, School of Wonders information company, No. 1518 Lianhang Road, 201112 Shanghai, China.

References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. 2018 Sep;68(6):394–424.
- Miller KD, Siegel RL, Lin CC, Mariotto AB, Kramer JL, Rowland JH, et al. Cancer treatment and survivorship statistics, 2016. *CA: A Cancer Journal for Clinicians*. 2016 Jun;66(4):271–289.
- Litjens G, Debat O, Barentsz J, Karssemeijer N, Huisman H. Computer-Aided Detection of Prostate Cancer in MRI. *IEEE Transactions on Medical Imaging*. 2014 May;33(5):1083–1092.
- Ahmed HU, Bosaily AES, Brown LC, Gabe R, Kaplan R, Parmar MK, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *The Lancet*. 2017 Feb;389(10071):815–822.
- Noh H, Hong S, Han B. Learning Deconvolution Network for Semantic Segmentation. *CoRR*. 2015;abs/1505.04366. Available from: <http://arxiv.org/abs/1505.04366>.
- Mehta S, Mercan E, Bartlett J, Weaver DL, Elmore JG, Shapiro LG. Learning to Segment Breast Biopsy Whole Slide Images. *CoRR*. 2017;abs/1709.02554. Available from: <http://arxiv.org/abs/1709.02554>.
- Zreik M, Lessmann N, van Hamersvelt RW, Wolterink JM, Voskuil M, Viergever MA, et al. Deep learning analysis of the myocardium in coronary CT angiography for identification of patients with functionally significant coronary artery stenosis. *Medical Image Analysis*. 2018 Feb;44:72–85.
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*. 2017 Dec;42:60–88.
- Drozdal M, Chartrand G, Vorontsov E, Shakeri M, Jorio LD, Tang A, et al. Learning normalized inputs for iterative estimation in medical image segmentation. *Medical Image Analysis*. 2018 Feb;44:1–13.
- Xing F, Xie Y, Yang L. An Automatic Learning-Based Framework for Robust Nucleus Segmentation. *IEEE Transactions on Medical Imaging*. 2016 Feb;35(2):550–566.
- Zhu W, Huang Y, Tang H, Qian Z, Du N, Fan W, et al. AnatomyNet: Deep 3D Squeeze-and-excitation U-Nets for fast and fully automated whole-volume anatomical segmentation. *Medical Physics*. 2019 Feb;46(2):576–589.
- Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017 Apr;39(4):640–651.
- Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Lecture Notes in Computer Science*. vol. 9351; 2015. p. 234–241.
- Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE; 2016. .
- Tian Z, Liu L, Zhang Z, Fei B. PSNet: prostate segmentation on MRI based on a convolutional neural network. *Journal of Medical Imaging*. 2018 Jan;5(02):1.
- Ji D, Yu J, Kurihara T, Xu L, Zhan S. Automatic Prostate Segmentation on MR Images with Deeply Supervised Network. In: *2018 5th International Conference on Control, Decision and Information Technologies (CoDIT)*. IEEE; 2018. .
- To MNN, Vu DQ, Turkbey B, Choyke PL, Kwak JT. Deep dense multi-path neural network for prostate segmentation in magnetic resonance imaging. *International Journal of Computer Assisted Radiology and Surgery*. 2018 Aug;13(11):1687–1696.
- Liao S, Gao Y, Oto A, Shen D. Representation Learning: A Unified Deep Learning Framework for Automatic Prostate MR Segmentation. *Med Image Comput Comput Assist Interv*. 2013;16(Pt 2):254–261.
- Zhu Q, Du B, Turkbey B, Choyke PL, Yan P. Deeply-supervised CNN for prostate segmentation. In: *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE; 2017. .
- Zhu Q, Du B, Turkbey B, Choyke P, Yan P. Exploiting Interslice Correlation for MRI Prostate Image Segmentation, from Recursive Neural Networks Aspect. *Complexity*. 2018;2018:1–10.
- Cheng R, Roth HR, Lu L, Wang S, Turkbey B, Gandler W, et al. Active appearance model and deep learning for more accurate prostate segmentation on MRI. In: *Styner MA, Angelini ED, editors. Medical Imaging 2016: Image Processing*. vol. 9784. SPIE; 2016. .
- Vincent G, Guillard G, Bowes M. Fully automatic segmentation of the prostate using active appearance models. *MICCAI Grand Challenge: Prostate MR Image Segmentation*. 2012;2012:2.
- Zhu Y, Wei R, Gao G, Ding L, Zhang X, Wang X, et al. Fully automatic segmentation on prostate MR images based on cascaded fully convolution network. *Journal of Magnetic Resonance Imaging*. 2018 Oct;49(4):1149–1156.
- Xue Y, Xu T, Zhang H, Long LR, Huang X. SegAN: Adversarial Network with Multi-scale L1 Loss for Medical Image Segmentation. *Neuroinformatics*. 2018 May;16(3-4):383–392.
- Goodfellow IJ. NIPS 2016 Tutorial: Generative Adversarial Networks. *CoRR*. 2017;abs/1701.00160. Available from: <http://arxiv.org/abs/1701.00160>.
- Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE; 2017. p. 2261–2269.
- Thada V, Jaglan V. Comparison of jaccard, dice, cosine similarity coefficient to find best fitness value for web retrieved documents using genetic algorithm. *International Journal of Innovations in Engineering and Technology*. 2013;2(4):202–205.
- Dou Q, Yu L, Chen H, Jin Y, Yang X, Qin J, et al. 3D deeply supervised network for automated segmentation of volumetric medical images. *Medical Image Analysis*. 2017 Oct;41:40–54.
- Maier O, Schröder C, Forkert ND, Martinetz T, Handels H. Classifiers for Ischemic Stroke Lesion Segmentation: A Comparison Study. *PLOS ONE*. 2015 Dec;10(12):e0145118.
- Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1993;15(9):850–863.
- Litjens G, Toth R, van de Ven W, Hoeks C, Kerkstra S, van Ginneken

- B, et al. Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge. *Medical Image Analysis*. 2014 Feb;18(2):359–373.
32. Ketkar N. Introduction to PyTorch. In: *Deep Learning with Python*. .
 33. Ghafoorian M, Mehrtash A, Kapur T, Karssemeijer N, Marchiori E, Pesteie M, et al. Transfer Learning for Domain Adaptation in MRI: Application in Brain Lesion Segmentation. In: *Springer Proceedings in Physics*. .

Figures

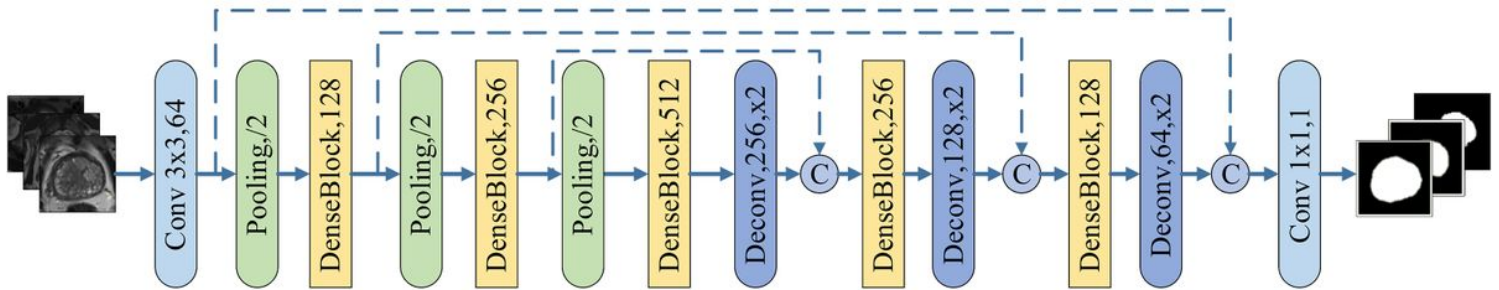


Figure 1

The architecture of generator G. it comprises a down-sampling path with three dense blocks and an up-sampling path with two dense blocks; blue circles represent concatenation in the network. Blue dotted arrows indicate bypass connections of the feature maps between coupled encoders and decoders

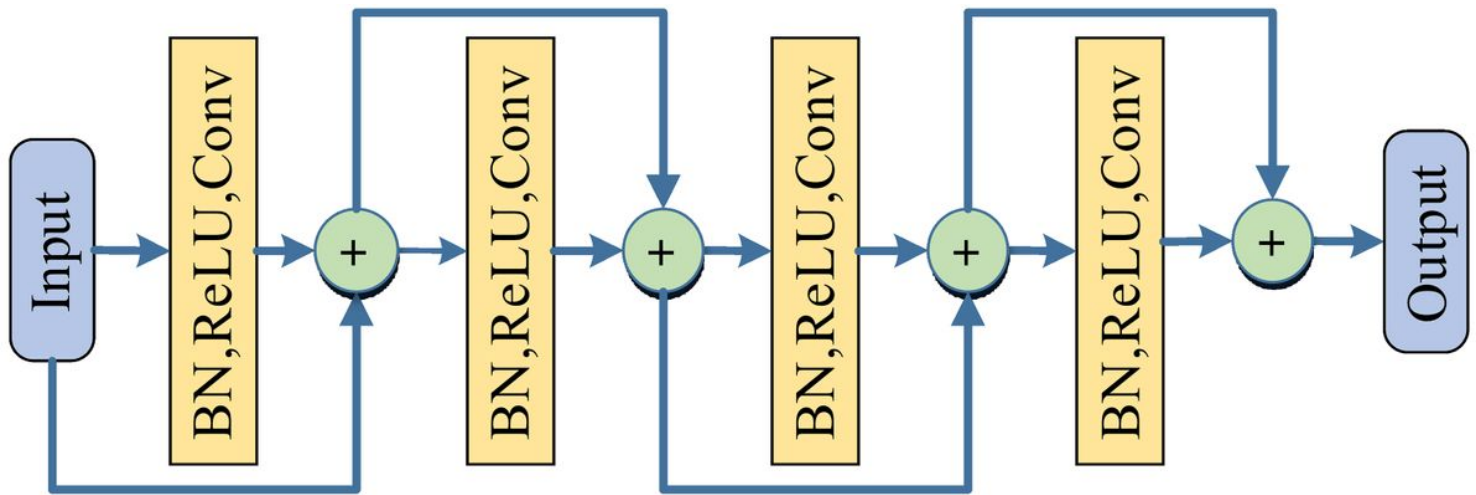


Figure 2

The architecture of discriminator D. All layers comprise convolutional, batch normalization (BN), and leaky ReLU layers. BN is not included in the rst layer. Masked images are generated by multiplying label images with input images in a pixel-wise orde

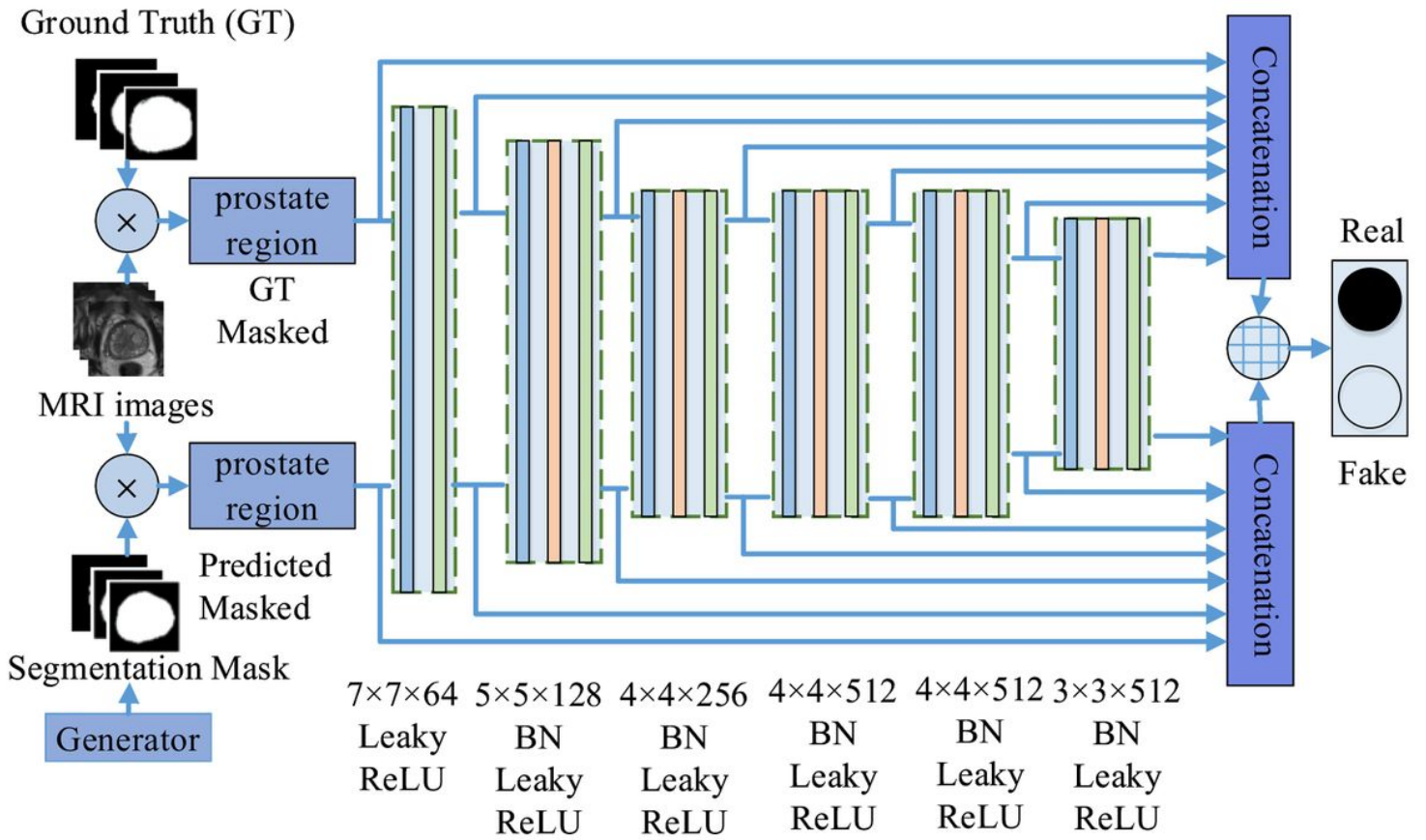


Figure 3

The dense block structure. it is adopted in Generator G. Green circles represent concatenation.

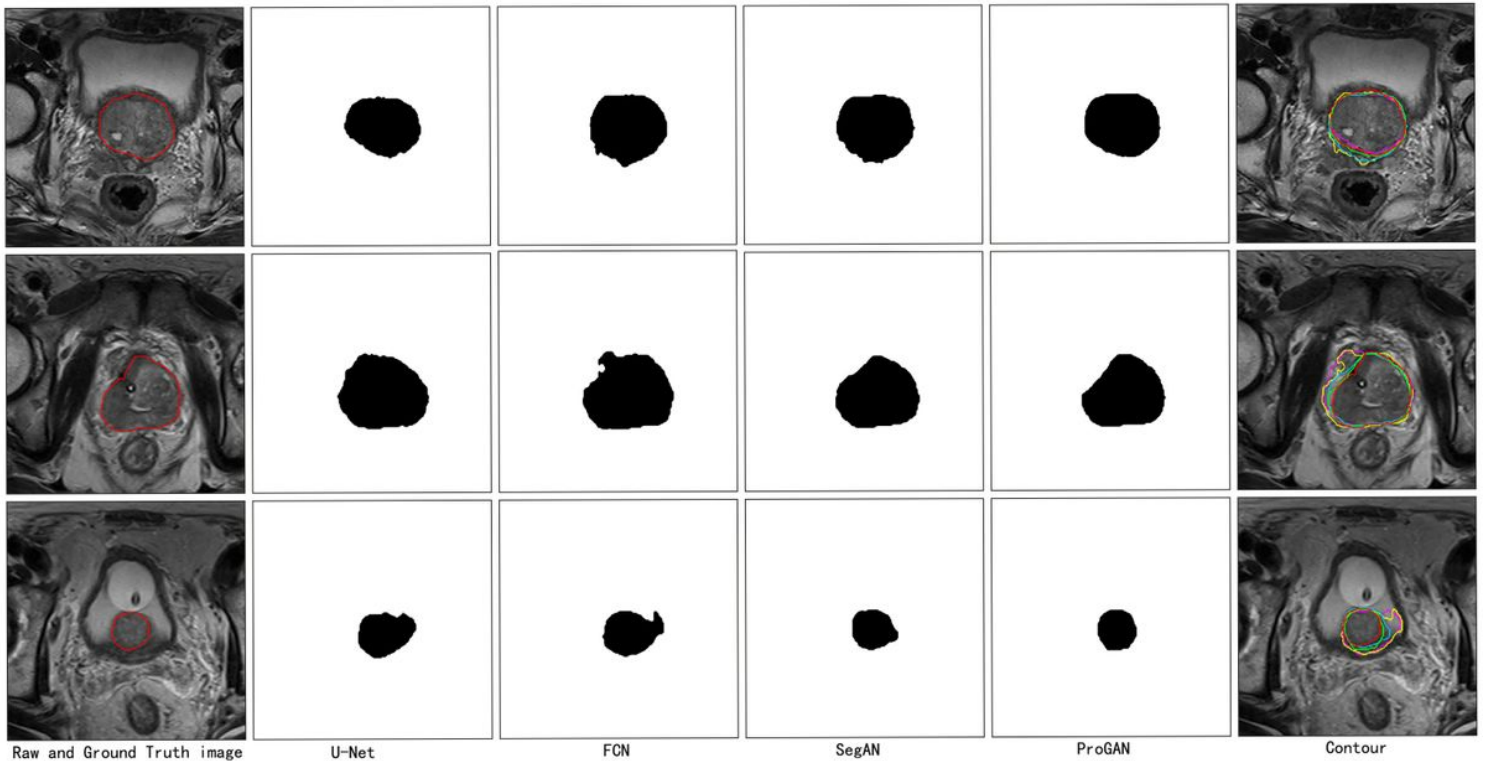


Figure 4

Qualitative segmentation results of a prostate gland. these results from our clinical dataset at the center (upper), apex (middle), and base (lower); the rst column shows raw MRI images with ground truth of the prostate gland. Columns 2-5 correspond with segmentation analyses using U-Net, FCN, SegAN, and SegDGAN models, respectively. In the nal column, ve contours superimposed on the raw MR images; red contour denotes the ground truth, and pink, yellow, blue, and green contours denote the segmentation results from U-Net, FCN, SegAN, and SegDGAN, respectively.

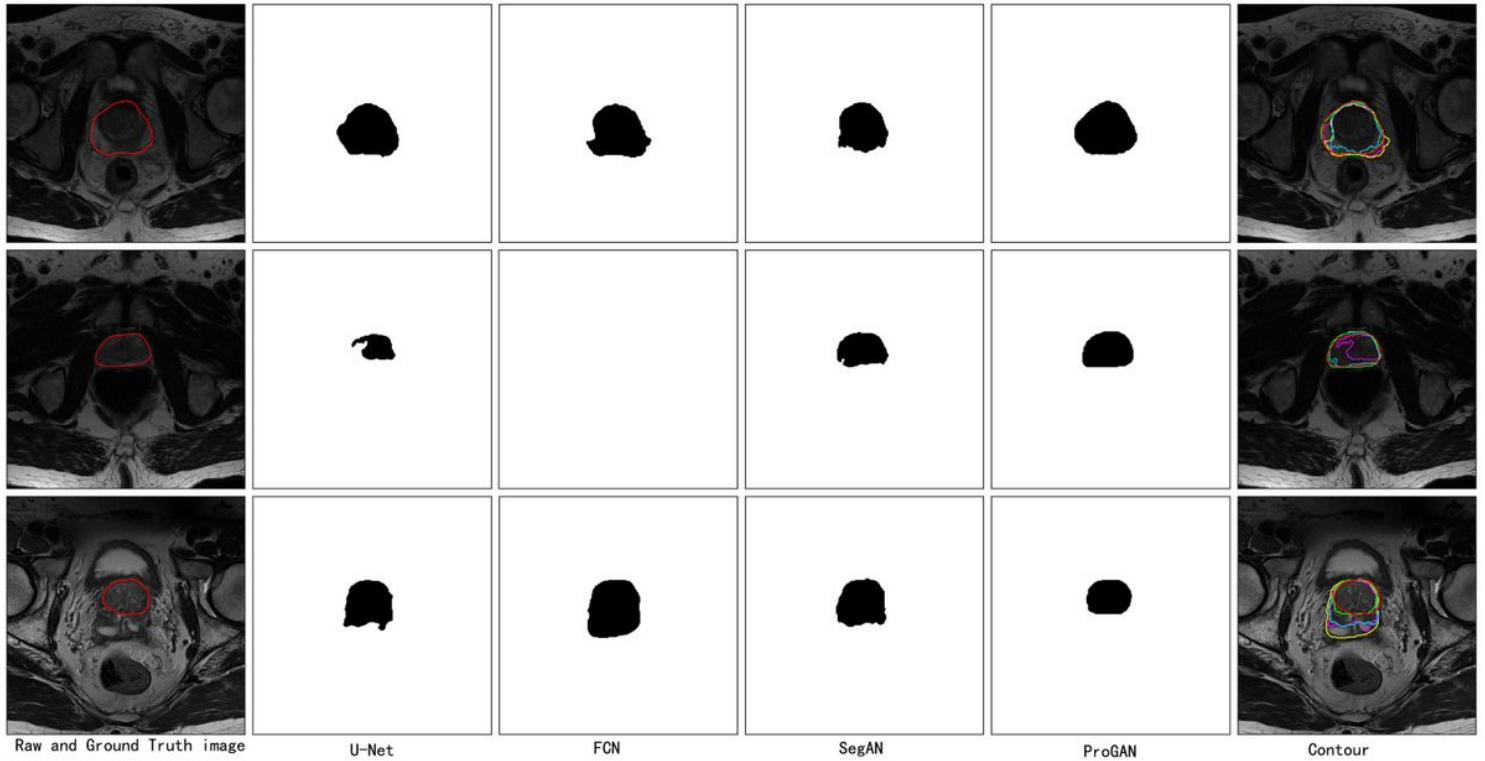


Figure 5

Qualitative segmentation results. the comparable qualitative results of U-Net, FCN, SegAN, and SegDGAN model analyzes using the public dataset PROMISE12.

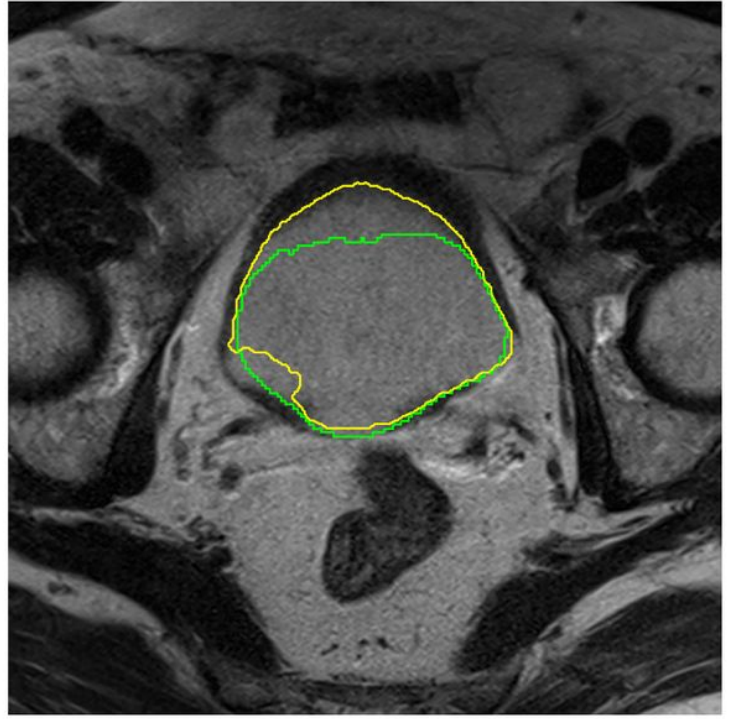
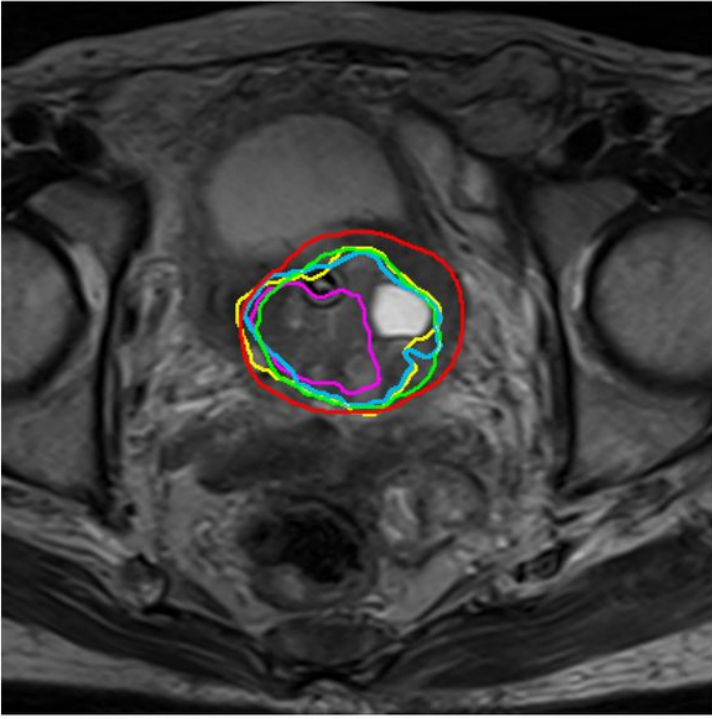


Figure 6

Segmentation results of the worst case