Computer Methods and Programs in Biomedicine 200 (2021) 105878

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

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

Medical image segmentation and reconstruction of prostate tumor based on 3D AlexNet

Jun Chen[a], Zhechao Wan[f], Jiacheng Zhang[b], Wenhua Li[c], Yanbing Chen[d], Yuebing Li[e,*], Yue Duan[a,*]

[a] Department of Urology, The Second Affiliated Hospital of Zhejiang Chinese Medical University, No.318 Chaowang Road, Gongshu District, Hangzhou 310005 China
[b] The 2nd Clinical Medical College, Zhejiang Chinese Medical University, 548 Bin Wen Road, Hangzhou 310053, China
[c] Department of Radiology, Xinhua Hospital affiliated to Shanghai Jiao Tong University School of Medicine, 1665 Kong Jiang Road, Shanghai 200092, China
[d] Computer Application Technology, School of Applied Sciences, Macao Polytechnic Institute, Macao SAR 999078, China
[e] Department of Anaesthesiology, The Second Affiliated Hospital of Zhejiang Chinese Medical University, No.318 Chaowang Road, Gongshu District, Hangzhou 310005 China
[f] Department of Urology, Zhuji Central Hospital, No.98 Zhugong Road, Jiyang Street, Zhuji City, 311800, Zhejiang Province, China

ARTICLE INFO

Article history:
Received 24 October 2020
Accepted 22 November 2020

Keywords:
Convolutional Neural Network
Prostate Cancer
Three-dimensional reconstruction
3D AlexNet

ABSTRACT

Background: Prostate cancer is a disease with a high incidence of tumors in men. Due to the long incubation time and insidious condition, early diagnosis is difficult; especially imaging diagnosis is more difficult. In actual clinical practice, the method of manual segmentation by medical experts is mainly used, which is time-consuming and labor-intensive and relies heavily on the experience and ability of medical experts. The rapid, accurate and repeatable segmentation of the prostate area is still a challenging problem. It is important to explore the automated segmentation of prostate images based on the 3D AlexNet network.

Method: Taking the medical image of prostate cancer as the entry point, the three-dimensional data is introduced into the deep learning convolutional neural network. This paper proposes a 3D AlexNet method for the automatic segmentation of prostate cancer magnetic resonance images, and the general network ResNet 50, Inception -V4 compares network performance.

Results: Based on the training samples of magnetic resonance images of 500 prostate cancer patients, a set of 3D AlexNet with simple structure and excellent performance was established through adaptive improvement on the basis of classic AlexNet. The accuracy rate was as high as 0.921, the specificity was 0.896, and the sensitivity It is 0.902 and the area under the receiver operating characteristic curve (AUC) is 0.964. The Mean Absolute Distance (MAD) between the segmentation result and the medical expert's gold standard is 0.356 mm, and the Hausdorff distance (HD) is 1.024 mm, the Dice similarity coefficient is 0.9768.

Conclusion: The improved 3D AlexNet can automatically complete the structured segmentation of prostate magnetic resonance images. Compared with traditional segmentation methods and depth segmentation methods, the performance of the 3D AlexNet network is superior in terms of training time and parameter amount, or network performance evaluation. Compared with the algorithm, it proves the effectiveness of this method.

1. Introduction

Prostate cancer is currently one of the most common cancers in men. In 2018, there were 164,690 new prostate cancer cases and 29,430 deaths in the United States [1]. From 1998 to 2008, the incidence of prostate cancer in China also showed a significant upward trend [2,3]. Some studies [4] predict that by 2030, there

* Corresponding authors.
E-mail addresses: lyb1853@zcmu.edu.cn (Y. Li), 20164919@zcmu.edu.cn (Y. Duan).

https://doi.org/10.1016/j.cmpb.2020.105878

will be 1,700,000 prostate cancer patients worldwide. Therefore, the early detection of prostate cancer is essential to improve the survival rate of prostate cancer patients and enhance the quality of life of prostate cancer patients.

The current clinical diagnosis methods for prostate cancer mainly include Prostate Specific Antigen (PSA) serum test, Digital Rectal Examination (DRE), Transrectal Ultrasound (TRUS), and guided by transrectal ultrasound. Needle biopsy and magnetic resonance imaging (MRI). Among them, transrectal ultrasound-guided needle biopsy is currently the most reliable way to diagnose prostate cancer. However, this traditional diagnosis method not only makes the patient very painful, but also may cause the sampling area to be not the diseased area [5,6], which may delay the treatment of the condition or cause over diagnosis and treatment.

MRI is a non-invasive and accurate way to detect prostate cancer [7]. However, accurate interpretation of MRI images requires radiologists to have a lot of professional knowledge and rich experience, and it is very time-consuming. Automated diagnosis of magnetic resonance images can effectively solve the above problems.

Usually, medical doctors diagnose patients through multimodal three-dimensional (3D) magnetic resonance imaging (MRI). MRI technology is a medical focus imaging technology that is realized through operations such as generating and acquiring magnetic resonance signals and performing spatial encoding and Fourier sampling. 3D MRI technology reconstructs a sequence of two-dimensional images arranged in sequence into 3D images, which is the most important imaging method in prostate cancer detection tasks. Compared with two-dimensional medical images such as plain X-ray films and angiography scans, 3D MRI can provide doctors with image information at any coordinate point in space, which is convenient for staff to perform quantitative and qualitative mathematical and medical analysis. In the 3D MRI process, MRI images of different modalities can be scanned due to the difference in auxiliary conditions such as contrast agents.

Convolutional Neural Network (CNN) has been used for the diagnosis of medical images of various diseases since the 1990s. In recent years, deep CNN has shown excellent results in the segmentation and classification of medical images. Studies [8,9] have shown that this method can be used to assist in the diagnosis of prostate cancer.

At present, the exploration of 3D image reconstruction and surgical navigation of prostate tumors in China is still in its infancy. There are studies on the 3D reconstruction of prostate based on multi-parameter magnetic resonance to assist in improving the accuracy of prostate biopsy [10]. With the overall development of deep learning research, especially the research of convolutional neural networks, more and more attention has been paid. Some scholars have used convolutional neural networks to segment prostate CT or magnetic resonance images as a whole [11].

In recent years, with the popularity of deep learning research, convolutional neural networks have received widespread attention. LeCun et al. [12] first proposed the use of multi-layer convolutional neural network structure to realize handwritten character discrimination in 1998, which is the cornerstone of convolutional neural network for image processing.

After decades of development, more researchers have applied convolutional neural networks to medical imaging data processing, and completed the segmentation and recognition of imaging organs through "pre-training + structural fine-tuning" [13]. Korez et al. [14] applied a three-dimensional convolutional neural network to segment the vertebral body of the magnetic resonance image. Brosch et al. [15] applied a deep three-dimensional convolutional network to perform multi-scale feature fusion, and then segmented multiple sclerosis lesions. Ranneberger et al. [16] proposed U-net, which performs good cell wall structure segmentation on the pathological tablet of He La cell suspension. In the Bayesian framework, Martínez et al. [17] used a preset ellipsoid shape to constrain, and then segmented the prostate.

With the improvement of image scanning accuracy, the image of the internal structure of the human body becomes clearer and clearer. Through the re-excavation of image data, the segmentation and reconstruction of organs and fine structures are realized. With the rise of material technology and the popularization of 3D printing technology, some scholars have printed 3D reconstruction models to help clinicians deepen their cognition, thereby improving operation accuracy and directivity. This non-real-time assistive technology is called Cognitive assisted navigation. Ebbing et al. [18] showed the 3D printed model to radiologists and achieved a 19% improvement in anatomical awareness.

2. Methodology

2.1. Improved AlexNet network

2.1.1. AlexNet network structure

AlexNet was proposed by Krizhevsky et al. [19] and won the championship in the 2012 Image Net competition, which more than doubled the recognition accuracy of Image Net. The model consists of an 8-layer structure and has its own outstanding advantages in image classification [20]. The specific structure of AlexNet is shown in Fig. 1.

The input format of the data source is 227 × 227 × 3 pixs, whereby 227 pixs represents the width and height of the input image, 3 represents the data source is a three-channel RGB mode, so it supports color pictures in commonly used formats, so there is no need for additional formats for the original data source collected cropped.

The computational processes of the first two layers are convolution (Conv), ReLu, maximum pooling (max-pooling) and normalization (normal). The output result of the second layer is subjected to 256 feature maps for convolution operation, in which kernel size is 5, stride is 1, and the remaining parameters are the same as the first layer. The third and fourth layers only performed convolution and ReLu operations. The fifth layer is similar to the first layer, except that it has not been normalized. Convert the result of the fifth layer into a long vector and input it into a traditional neural network using a three-layer fully connected structure. The kernels of the first two fully connected layers are 4,096 respectively. The last layer generates 1,000 nodes, and the label value can be obtained by using the Softmax regression function [21].

2.1.2. AlexNet network based on ResNet

In order to build a neural network model with high calculation accuracy and fast convergence speed, this paper proposes an improved AlexNet convolutional neural network model, which uses batch normalization and global maximum pooling algorithms. The biggest difference between the ordinary direct-connected convolutional neural network and the ResNet convolutional neural network is that ResNet has many bypass branches to directly connect the input to the subsequent layers, so that the subsequent layers can directly learn the residuals. In the traditional convolutional layer or fully connected layer, there will be more or less information loss, loss and other problems when information is transmitted. Therefore, the image information will be lost or depleted, which reduces the recognition efficiency of the algorithm [22].

However, ResNet solves this problem to some extent. By directly routing the input information to the output layer, thereby ensuring the integrity of different images, the entire network only needs to learn the two links of input and output differences, simplifying the learning goal And difficulty to improve the learning rate.

The activation function used by the traditional AlexNet is a modified linear unit activation function (rectified linear units,

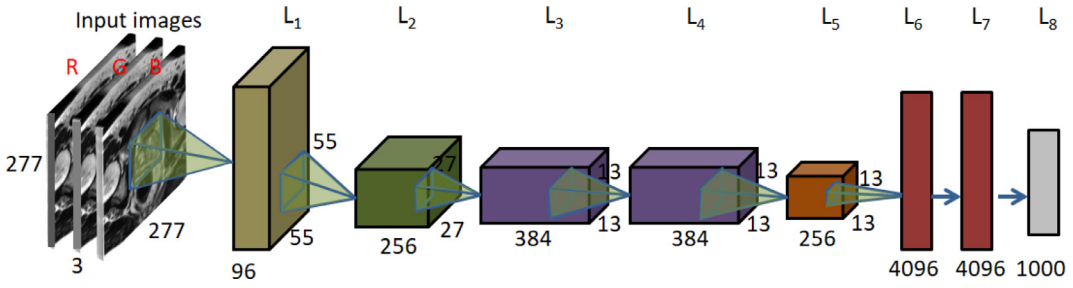

Fig. 1. AlexNet network structure diagram.

ReLu), which can prevent gradient dispersion during back propagation and accelerate the training of convolutional neural networks. He et al. proposed the PReLu activation function as an improved ReLu activation function, and changed the negative axis slope of the ReLu function from 0 to a variable parameter a, and achieved good results in model training. Eq. (1) and Eq. (2) are ReLu respectively. Activation function and PReLu activation function, where the variable slope is.

$$\mathrm{ReLu}(x) = \begin{cases} x, & \text{if } x > 0 \\ 0, & \text{if } x \le 0 \end{cases} \tag{1}$$

$$\mathrm{PReLu}(x) = \begin{cases} x, & \text{if } x > 0 \\ ax, & \text{if } x \le 0 \end{cases} \tag{2}$$

The AlexNet convolutional neural network based on residual connection proposed in this paper will use the PReLu activation function to observe the influence of different convolution kernel sizes on the accuracy of model recognition. The model's global pooling algorithm uses global maximum pooling and global average pooling to compare the impact of different pooling types on model recognition performance.

2.2. Data set and preprocessing

Multi-parameter magnetic resonance examination before prostate puncture, using Siemens 3.0 T magnetic resonance scanner, using abdominal phased array coil to receive signals, scanning includes peripheral, sagittal and coronal fast spin echo (turbo spin echo, TSE) T2WI and axial T1WI.

Although the accuracy of current image recognition algorithms based on deep learning is quite high, they are all realized after training based on a large number of samples. When the training sample is small, the learning effect of the algorithm will be difficult to guarantee. A small sample learning method based on data augmentation is a common method to solve such problems.

The easiest way to prevent overfitting is to increase the data unlimitedly, but in practice this is impossible. Reliable data requires a lot of manpower and material resources to obtain, especially in the field of medical image recognition, which takes longer to accumulate data. Therefore, the image needs to be artificially expanded. In this paper, two methods are mainly adopted for data amplification, namely image flipping and image noise adding.

2.2.1. Image flip

As shown in Fig. 2, one can find the horizontal mirror point (x_1, y_1) of the original point (x_0, y_0), according to the coordinate relationship.

x_1, y_1 are obtained as Eq. (5), and so can be expressed as Eq. (6) in matrix form.

$$\begin{cases} x_1 = x_0 + w \\ y_1 = y_0 \end{cases}$$

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & w \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix}$$

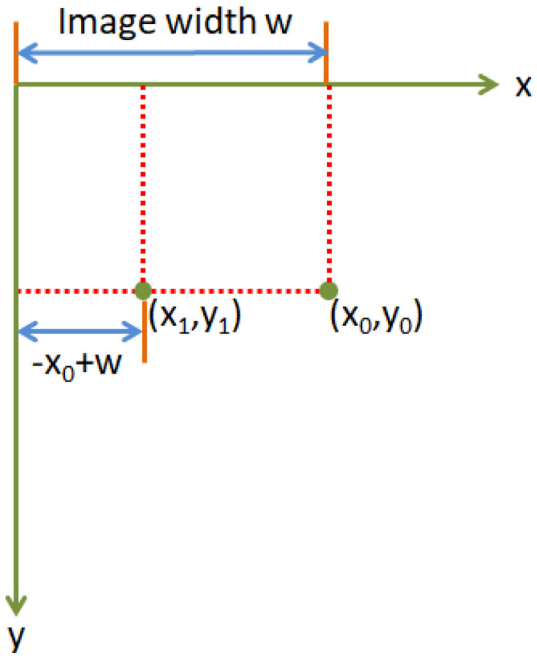

Fig. 2. Schematic diagram of image flip.

2.2.2. Add noise to the image

There are many algorithms for image noise enhancement. In this study, the additive zero-mean Gaussian noise method is used to add noise to the image. The way to increase noise is to add a noise value to the gray value of each point on the image, and the way to generate the noise value is to use the Box-Muller algorithm to generate Gaussian noise. The Box-Muller algorithm is based on two independent sets of uniformly distributed random numbers U and V in the interval (0,1) to generate two sets of independent standard normal distribution random variables X and Y, as shown in Eq. (7) and Eq. (8).

$$\mathrm{X} = \sqrt{-2\ln \mathrm{U}}\cos(2\pi \mathrm{V}) \tag{7}$$

$$\mathrm{Y} = \sqrt{-2\ln \mathrm{U}}\sin(2\pi \mathrm{V}) \tag{8}$$

2.3. Evaluation criteria

2.3.1. Network performance evaluation standards

An automated diagnosis system based on magnetic resonance images should meet the following two conditions: (1) low false positive rate (FR), the occurrence of false positives may cause overtreatment; (2) high true positive rate (TR), the high TPR can help patients receive timely treatment to increase the survival rate. Accuracy, sensitivity, specificity, and f1-score can reflect the model's judgment of true positive and false positive. The abscissa

of Receiver Operating Characteristic (ROC) curve plane is FR, the ordinate is TR, and the area under the curve (AUC) is the area at the bottom right of the ROC, which can intuitively reflect the overall performance of the model, The larger the value, the better the classification effect.

In this paper, accuracy, sensitivity (Sen), specificity (Spe), f1-score and AUC are used to jointly evaluate the performance of the 3D AlexNet network used in this paper, and the network efficiency is evaluated by comparing the training time and parameters of the network. The sensitivity and specificity calculation Equations are as Eq. (3) and Eq. (4).

$$\mathrm{Sen} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FR}} \tag{3}$$

$$\mathrm{Spe} = \frac{\mathrm{FR}}{\mathrm{TP} + \mathrm{FR}} \tag{4}$$

2.3.2. Evaluation method based on contour distance

For the same image, the distance between the corresponding pixel points between the result of manual segmentation and the result of algorithm segmentation is calculated to measure the difference between the two and compare them. The specific method of calculating the evaluation parameters based on the contour distance is as follows: First, record the coordinates of the contour line manually segmented by the doctor and the contour line segmented by the algorithm on the image according to the contour order on the image, and calculate the corresponding two coordinates of the two contour lines The distance between points. According to this method, the distance of all points between two contour lines can be calculated. Finally, by calculating the average of the distances of all contour points, the Mean Absolute Distance (MAD) is obtained, as shown in Eq. (5).

$$\mathrm{MAD} = \frac{1}{\mathrm{N}} \sum_{\mathrm{j}=1}^{\mathrm{N}} \left|\left|\mathrm{d}_j\right|\right| \tag{5}$$

In the formula, N represents the number of sampled contour points of the manually segmented image, and $||dj||$ represents the distance between the j-th point of the manually segmented contour and the corresponding point obtained by algorithm segmentation.

In addition, the Hausdorff distance (HD) is also an evaluation method based on contour distance. HD is a measure of the similarity between two sets of points. It can be used to evaluate the similarity between two sets of data points. When the HD is larger, it means that the similarity of the two sets of data is poor. On the contrary, it means that the similarity of the two sets of data is better.

$$\mathrm{HD}(\mathrm{A}, \mathrm{B}) = \max(\mathrm{h}(\mathrm{A}, \mathrm{B}), \mathrm{h}(\mathrm{B}, \mathrm{A})) \tag{6}$$

There are two sets of data points, namely point set $A=(a_1,a_2,...,ap)$ and point set $B=(b_1,b_2,...,b_q)$, then the HD between the two point sets is shown in Eq. (6) .

2.3.3. Evaluation method based on contour area

The evaluation method based on the contour area is to compare the area of the prostate contour, that is, the area of the prostate contour obtained by the doctor's manual segmentation and the area within the area of the prostate contour obtained by the algorithm. The evaluation method based on the contour area mainly compares the area in the common area of two different contours, and is not very sensitive to the change of the image contour shape.

Including specificity, sensitivity and Dice similarity coefficient (DSC), DSC is shown in Eq. (7).

$$\mathrm{DSC} = \frac{2\mathrm{TP}}{(\mathrm{FP} + \mathrm{TP}) + (\mathrm{TP} + \mathrm{FN})} \tag{7}$$

Table 1
Evaluation results based on contour distance.

Data set type	MAD (mm)	HD (mm)
Original data set	0.523	2.764
Enhanced data set	0.324	1.006

Table 2
Evaluation results based on contour area.

Data set type	DSC	SP	SN
Original data set	0.9257	0.9435	0.9146
Enhanced data set	0.9703	0.9867	0.9534

3. Results

3.1. Data enhancement experiment results and analysis

This paper proposes to optimize and filter the original data set first, and then to augment the filtered data set to improve the accuracy of training and the performance of the network. The images are from 25 patients in the hospital, and each patient selects 10 to 30 available samples for a total of 500 images. After 500 original images are labeled, 400 images are used for training, and the remaining 100 images are tested and evaluated. Among the 2000 enhanced images, 1700 images are used for training, and the remaining 300 images are used for algorithm testing and evaluation. In this paper, the manual segmentation results of doctors are used as the gold standard, and the segmentation results of different algorithms are compared to obtain different evaluation parameters.

The original data set and the data set after data augmentation are used for training respectively. During the training process, the accuracy rate and loss value of the network model on the validation set are shown in Fig. 3.

It can be seen from Fig. 3(a) that the loss value of the data set after data enhancement in the initial training period is close to that of the original data set on the validation set. As the number of iterations increases, the loss value of the data set after data enhancement is used. It is significantly smaller than the loss value of the original data set. It can be seen from Fig. 3(b) that as the number of iterations increases, the accuracy of the data set after data enhancement on the validation set is significantly higher than the accuracy of the original data set on the validation set. Therefore, data enhancement significantly improves the accuracy of network model training.

In order to further verify that data enhancement has improved the generalization ability and test accuracy of the network model, the model trained on the original data set and the model trained on the data enhanced data set are used to process the corresponding test set, and use MAD, HD, DSC, SP, SN. The segmentation results on the test set are quantitatively evaluated, and the results are shown in Table 1 and Table 2.

It can be seen from Table 1 and Table 2 that the MAD and HD of the original data set based on the contour distance evaluation result are 0.523mm and 2.764mm, respectively, and the corresponding MAD and HD of the enhanced data set are 0.324mm and 1.006mm, respectively. The profile distance is significantly reduced. Based on the evaluation results of the contour area of the original data, the DSC, SP, and SN are 0.9257, 0.9435, and 0.9146, respectively. The DSC, SP, and SN corresponding to the enhanced data set are 0.9703, 0.9867, and 0.9534, respectively. That is, the models trained after data enhancement is in two types. The indicators on the evaluation standard are better than the model trained on the original data set, which verifies the effectiveness of data increase in improving the generalization ability of the N-network model and the accuracy of the test.

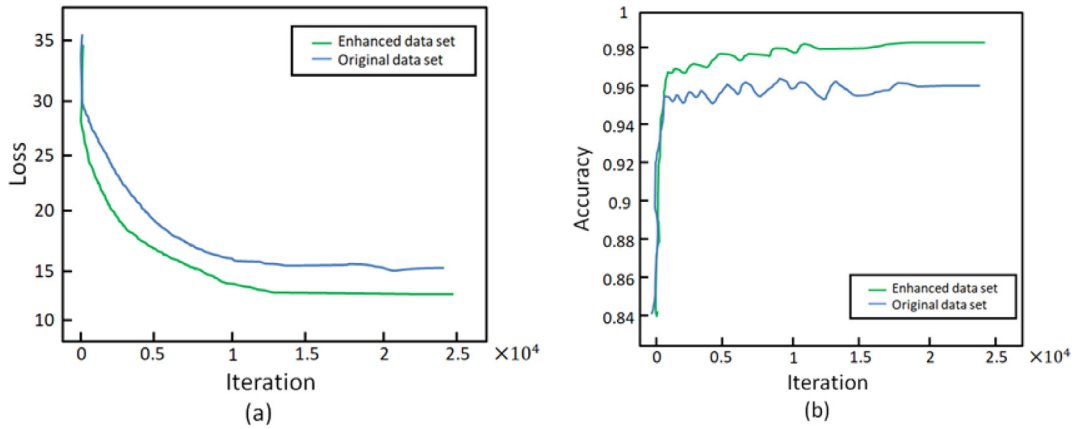

Fig. 3. Comparison of data augmentation training results based on (a) Loss comparison curve; (b) Accuracy contrast curve.

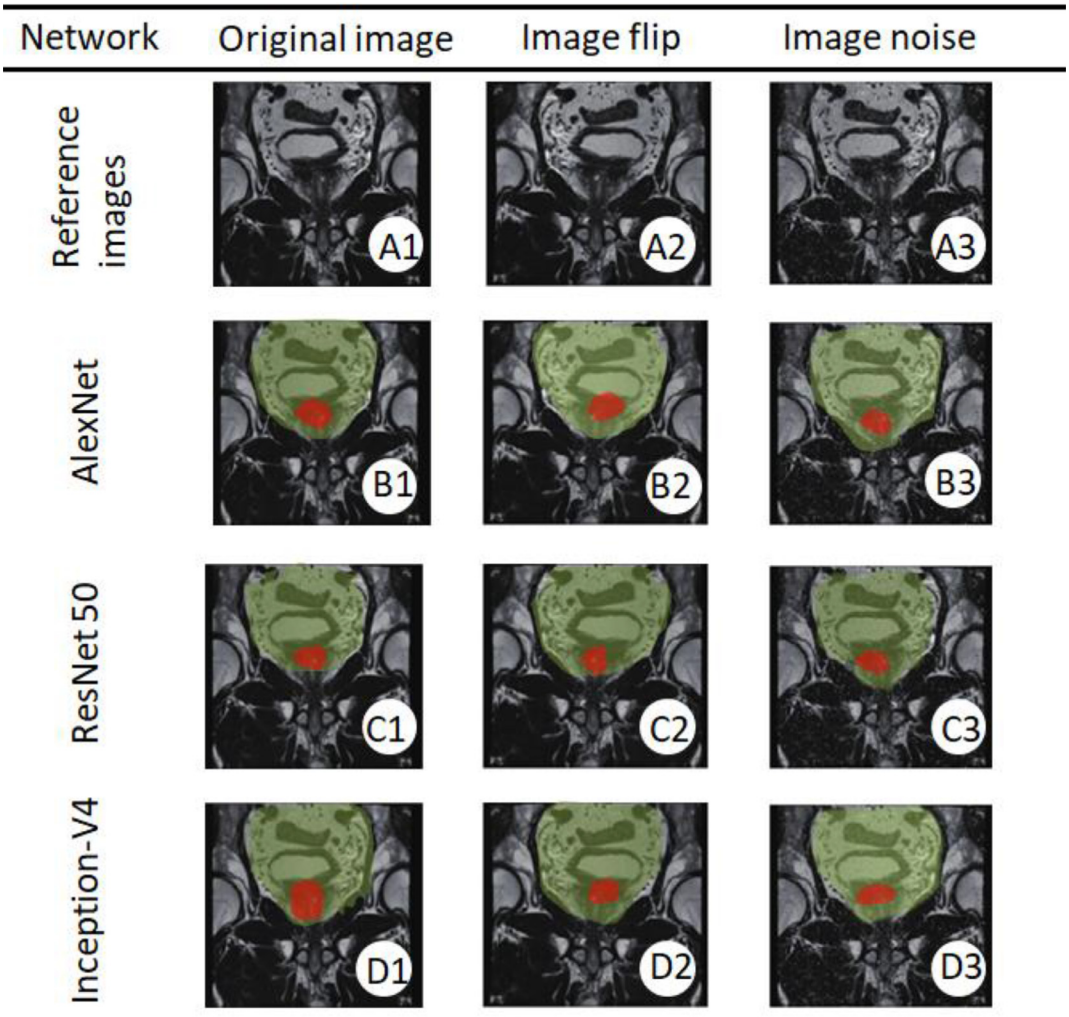

Fig. 4. The results of three kinds of network segmentation of prostate are based on (A) prostate image; (B) AlexNet; (C) ResNet 50; (D) Inception-V4. Note that the green part represents the prostate, and the red part represents the tumor.

3.2. High-precision 3D reconstruction

The first step is to segment the original image, flip the image, and add noise. We use AlexNet, ResNet 50, and Inception-V4 network to segment, compare the segmentation effect of the network, the green part represents the prostate, the red part represents cancer, the segmentation results are shown in Fig. 4.

It can be seen from Fig. 4 that the segmentation effect of AlexNet is better, the edges of the segmented image are smoother, and the location of the segmented cancer is more accurate.

We import the data obtained from the previous step of fully automatic 3D AlexNet segmentation [15], whereby the AlexNet framework is used for prostrate and tumor segmentation of the images pertaining to multiple cross-sectional MRI scans. Next, we

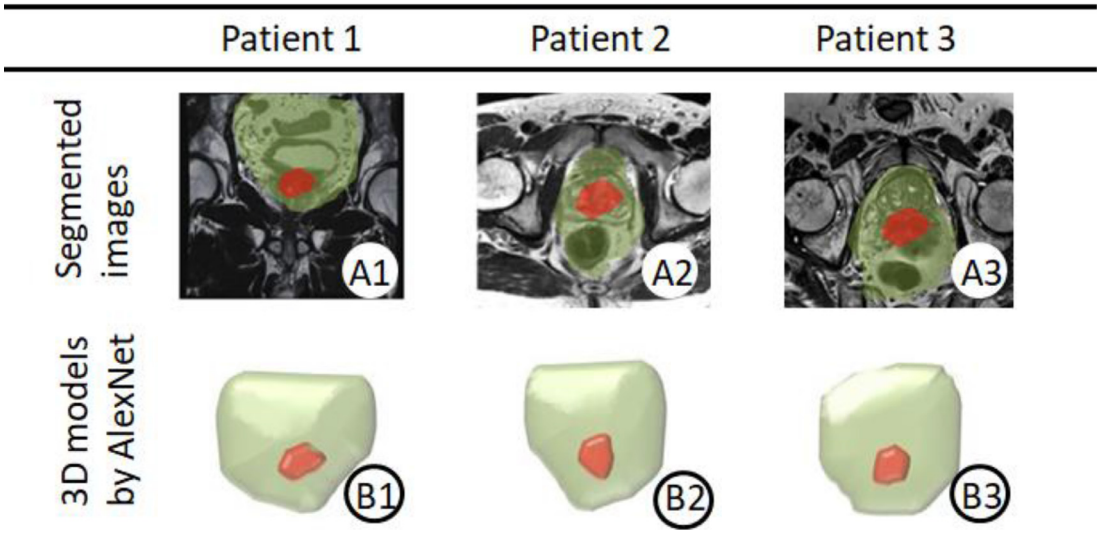

Fig. 5. Using AlexNet to segment the (A) prostrate scan images followed by (B) 3D reconstruction for three patients.

Table 3
Comparison of training time and parameter amount of different methods.

Method	Training time (s)	Parameter (MB)
3D AlexNet	4792	35.7
Inception-V4	51132	295.0
ResNet 50	24301	167.0

write the 3D reconstruction algorithm with the basic algorithm of boundary contour as the core for 3D reconstruction pertaining to three sample patients (Fig. 5). For the original magnetic resonance data the layer thickness is 5 mm and therefore, the model needs to be smoothed after reconstruction. In order to improve the visual perception, Gaussian smoothing is performed on the reconstructed graphics with a smoothing coefficient of 0.3, and 10 iterations. Because of the collapse of the organ volume after multiple smoothing, a 3.5% overall enlargement is performed to compensate. The structure data of each part is output in an independent STL format file.

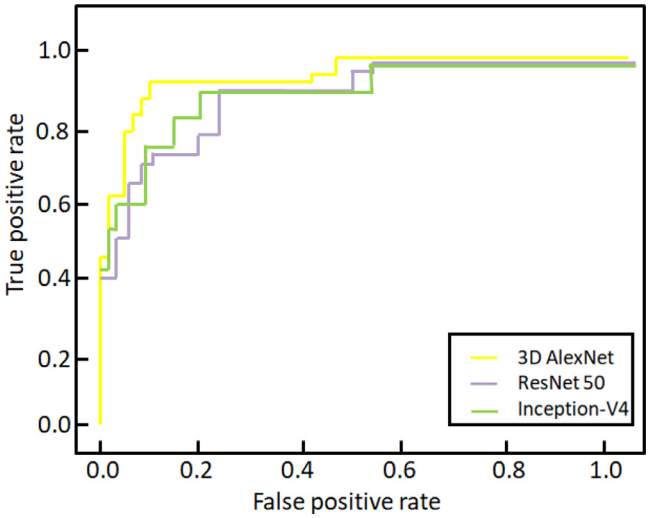

Fig. 6. ROC curves of the three models.

3.3. Comparison of different network models

3.3.1. Training time and parameter amount

The training time and parameters of the three networks are shown in Table 3. Compared with the two general models of ResNet50 and Inception-V4, 3D AlexNet's training time and parameters are significantly reduced. The training time of ResNet50 is about 6 times that of 3D AlexNet, and the amount of parameters is about 5 times that of 3D AlexNet, indicating that 3D AlexNet greatly reduces the number of layers and channels of the network. The above results show that the adjustment of the network architecture, as well as the application bottleneck structure and global average pooling, can shorten the training time and reduce the amount of parameters.

3.3.2. Network performance evaluation

Fig. 6 shows the ROC curve obtained on the data set of 3D AlexNet and general models ResNet 50 and Inception-V4. The final classification performance indicators obtained by the three models are shown in Table 4.

The AUC value obtained by the optimized network3D AlexNet proposed in this paper is 0.964. Compared with the AUC value of ResNet 50 of 0.901 and the AUC value of Inception-V4 of 0.923, the classification performance is better. In this paper, after reducing the number of layers and the number of channels, the classification performance is further improved. Therefore, we guess that on a data set with a low pixel density, a network with too many parameters may have too many feature inputs, but many of them the features are redundant and interfere with the classification of the classifier. The optimized network 3D AlexNet proposed in this paper is better than ResNet 50 and Inception-V4 in accuracy, sensitivity, specificity, and fl-score. It shows that 3D AlexNet has the best classification effect in the networks compared in this paper; the adjustment of the overall network structure and the application of bottleneck structure, batch standardization, and global average pooling can not only greatly reduce the training time and the amount of parameters, but also improve the classification performance.

In order to ensure that the data is statistically significant, SPSS 21.00 software is selected for analysis, and the P value is calculated. The P value of 3D AlexNet is 0.001 and $P < 0.05$ indicates that the difference is statistically significant.

Table 4
Performance comparison of different networks.

Method	Accuracy	Sensitivity	Specificity	fl-score	AUC	P value
3D AlexNet	0.921	0.896	0.902	0.897	0.964	0.001
Inception-V4	0.876	0.821	0.864	0.811	0.923	0.005
ResNet 50	0.857	0.798	0.842	0.873	0.901	0.005

Table 5
Evaluation results based on different network contour distance.

Method	MAD (mm)	HD (mm)
3D AlexNet	0.356	1.024
Inception-V4	0.486	1.968
ResNet 50	0.476	2.008

Table 6
Evaluation results based on different network contour area.

Method	DSC	SP	SN
3D AlexNet	0.9768	0.9889	0.9576
Inception-V4	0.9467	0.9854	0.9431
ResNet 50	0.9338	0.9810	0.9431

3.3.3. Network segmentation effect

From Tables 5 and 6, it can be seen that the contour distance-based MAD and HD corresponding to the 3D AlexNet algorithm are 0.356mm and 1.024mm, respectively, which are higher than Inception-V4 and ResNet 50. The DSC, SP and SN based on the contour area are 0.9768, 0.9889 and 0.9576, respectively, which are also better than Inception-V4 and ResNet 50. From the above analysis, it can be concluded that the 3D AlexNet algorithm combines the advantages of Inception-V4 and ResNet 50, which can quickly initialize the network model and maintain good network generalization performance.

4. Discussion

In the task of prostate cancer segmentation, how to deal with multi-modal 3D images is also one of the difficulties to be solved urgently. Currently, there are two main ways to process 3D images. One method is to use a 3D convolutional network for image processing. Due to the spatial characteristics of the 3D convolution kernel, the extracted features will be more sufficient and contain information in all dimensions of the data. Kamnitsas et al. [23] used a 3D convolutional neural network to extract the characteristics of brain tumor samples, and combined with post-processing methods such as fully connected conditional random fields to complete the segmentation task of multi-modal brain tumor MRI images. Chen et al. [24] extended the idea of residual learning to 3D convolution, designed the network structure of Voxrest, and also completed the task of 3D image segmentation. However, 3D images usually have a small amount of data, and in the actual training process, the graphics processor (GPU) occupied by the 3D image is too large, which is not conducive to the deepening of the network, so the segmentation accuracy of the 3D convolution method is difficult to improve.

The other method is to use a two-dimensional (2D) convolutional network for image processing. Pereira et al. [25] cropped the 3D image into several two-dimensional image sequences, and used the 2D convolution kernel to extract the features of each two-dimensional image, and obtained a better segmentation effect. Havaei et al. [26] applied the cascade idea to segmentation, and performed feature extraction through 2D convolution to achieve fully automatic brain tumor segmentation. Since each 3D image can be cropped into several two-dimensional images, the data volume of the 2D convolutional network is more sufficient than that of the 3D convolutional network [27–30]. For data-driven algorithms such as deep learning, the training effect of 2D convolutional networks is usually better. However, the 2D convolution kernel cannot extract spatial features. Simply splicing the 2D segmentation results into 3D segmentation results may cause jaggedness and faults, which affect the accuracy of segmentation. Therefore, improving the accuracy of 3D segmentation results is the focus of future work [31].

This paper integrates technologies and efforts in image segmentation, recognition, 3D reconstruction, neural network design, etc. However, the deficiencies faced in reality are reflected in each link, such as the segmentation link, the current neural network and artificial image segmentation [32,33]. The accuracy matching degree is less than 90%. How to optimize the algorithm or add multi-modal information to improve the accuracy and reduce the mismatch rate is the top priority of all work. Another example is the postoperative pathological section.

Due to the difference between the cutting angle and the actual MRI scan angle, and considering the influence of the tissue deformation after embedding and fixing and dehydration, how to accurately match the postoperative pathological section information with the image information is also related to segmentation. In addition, the leap from cognitive navigation to augmented reality navigation involves technical problems such as multi-screen fusion, visual real-time feedback, and automatic registration, which all needs to be further, explored and developed [34–36].

5. Conclusion

Based on the medical image diagnosis of prostate cancer as our scope of analysis, the three-dimensional data is introduced into the deep learning convolutional neural network. The construction of a specific three-dimensional convolutional neural network proves its feasibility and value for future development. When the actual data set size is difficult to support 3D convolutional neural network training, through the application of various over-fitting technologies, the required standards are reached as much as possible to achieve the effect of completing the project. In-depth research on the classic network model is not only to understand its structure and to a certain extent transform it in a direction suitable for new needs.

This paper proposes a 3D AlexNet method for automatic classification of prostate cancer magnetic resonance images, which reduces the number of network layers and channels to reduce model parameters and shortens training time. At the same time, batch normalization and global average pooling are used to optimize the network to improve classification effect. The results show that the classification network 3D AlexNet proposed in this paper can reach 0.964 in the classification of prostate benign and malignant lesions based on magnetic resonance images, which can effectively diagnose prostate cancer. At the same time, the network has low parameters and training speed. Fast and other characteristics can be used as the basis for larger data set training or multi-modal training.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to thank the radiologists at The Second Affiliated Hospital of Zhejiang Chinese Medical University for their assistance in imaging of the prostrate based on their collected patients.

References

[1] R L Siegel, K D Miller, A J Dvm, Cancer statistics, 2018 [J], Ca: A Cancer Journal for Clinicians 68 (1) (2018) 11.
[2] J Han S, W Zhang S, Q Chen W, et al., Analysis of the status and trends of prostate cancer incidence in China[J], Chinese Clinical Oncology 18 (4) (2013) 330–334.
[3] Y Yang J, Z Yang M, W Wei, Epidemiological study on the occurrence and development of prostate cancer [J], J Clin Urology 9 (2017) 74–78.
[4] J Manddams, M Utley, H Moller, Projections of cancer prevalence in the United Kingdom, 2010-2040 [J], Brit J Cancer 107 (7) (2012) 1195–1202.
[5] Y Peng, Y Jiang, C Yang, Quantitative analysis of multiparametric prostate MR images: differentiation between prostate cancer and normal tissue and correlation with Gleason score- - a computer -aided diagnosis development study [J], Radiology 267 (3) (2013) 787–796.
[6] M D Rooij, E H J Hamoen, J J Futterer, Accuracy of multiparametric MRI for prostate cancer detection: A meta- analysis[J], AJR Am J Roentgenol 202 (2) (2014) 343–351.
[7] D Fehr, H Veeraraghavan, A Wibmer, Automatic classification of prostate cancer Gleason scores from multiparametric magnetic resonance images [J], Proceeding National Academy Science USA 112 (46) (2015) 6265–6273.
[8] X Yang, Y Liu C, W Wang Z, Co-trained convolutional neural networks for automated detection of prostate cancer in multi- parametric MRI [J], Medicine Image Analysis 42 (2017) 212–227.
[9] W Wang Z, Y Liu C, P Cheng D, Automated detection of clinically significant prostate cancer in mp- MRI images based on an end-to- end deep neural network [J], IEEE Translational Medicine Imaging 37 (5) (2018) 1127–1139.
[10] C. Azim, Effect of imaging parameters on the accuracy of apparent diffusion coefficient and optimization strategies [J], Diagnostic & Interventional Radiology 22 (1) (2015) 3215–3324.
[11] W Wang, J Shen, L Shao, Video Salient Object Detection via Fully Convolutional Networks [J], IEEE Transactions on Image Processing 27 (1) (2017) 38–49.
[12] Y LeCun, L Bottou, Y Bengio, et al., Gradient-based learning applied to document recognition[J], Proceedings of the IEEE (1998) 2278–2324.
[13] D Mahapatra, J M Buhmann, Prostate MRI segmentation using learned semantic knowledge and graph cuts [J], IEEE Transactions on Biomedical Engineering 61 (3) (2014) 756–764.
[14] R Korez, B Likar, F Pernuš, et al., Model-based segmentation of vertebral bodies from MR images with 3D CNNs [C], International Conference on Medical Image Computing and Computer Assisted Intervention. Cham: Springer (2016) 433–441.
[15] T Brosch, LY Tang, Y Yoo, et al., Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation [J], IEEE Transactions on Medical Imaging 35 (5) (2016) 1229–1239.
[16] 0 Ronneberger, P Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation [C], International Conference on Medical image computing and computer-assisted inter-vention. Cham: Springer (2015) 234–241.
[17] F. Martínez, E. Romero, G Dréan, et al., Segmentation of pelvic structures for planning CT using a geometrical shape model tuned by a multi-scale edge detector [J], Physics in Medicine Biology 59 (6) (2014) 1471–1484.
[18] J Ebbing, F Jäderling, JW Collins, et al., Comparison of 3D printed prostate models with standard radiological information to aid understanding of the precise location of prostate cancer: A construct validation study[J], The Public Library of Science 13 (6) (2018) e199477.
[19] Y Keizhevsk, I Stuskever, E Hinton G, ImageNet classification with deep convolutional neural networks[C], in: Proceedings of the Advances in Neural Information Processing Systems. South Lake Tahoe, US, 2012, pp. 1097–1105.
[20] Payan A, Montana G. Predicting Alzheimer's disease: a neuro imaging study with 3D convolutional neural networks [J]. 2013, 45(4): 96-102.
[21] G Hinton, L Deng, D Yu, et al., Deep neural networks for acoustic modeling in speech recognition [J], IEEE Signal Processing Magazine 29 (6) (2012) 82–97.
[22] PF Neher, B Stieltjes, M Reisert, et al., MITK global tractography [C], in: Proceedings of SPIE: The International Society for Optical Engineering, 2012, p. 83144D, doi:10.1117/12.911215.
[23] K. Kamnitsas, C. Ledig, V.F.J Newcombe, et al., Efficient multi-scale 3D CNN with fully connected CRF for accurate prostate cancer segmentation [J], Medical Image Analysis 36 (4) (2017) 61–78.
[24] Chen H, Dou Q, Yu L, et al. Vox ResNet: deep voxel wise residual networks for volumetric prostate cancer segmentation [EB/OL]. (2016-04-21) [2018-06-30]. arXiv preprint arXiv:1608.05895.
[25] S Pereira, A Pinto, V Alves, et al., Prostate cancer segmentation using convolutional neural networks in MRI images [J], IEEE Transactions on Medical Imaging 35 (5) (2016) 1240–1251.
[26] M Havaei, A Davy, D Warde- Farley, et al., Prostate cancer segmentation with deep neural networks [J], Medical Image Analysis 35 (2017) 18–31.
[27] F. Volonté, F. Pugin, P. Bucher, et al., Augmented reality and image overlay navigation with OsiriX in laparoscopic and robotic surgery: not only a matter of fashion [J], J Hepatobiliary Pancreas Science 18 (4) (2011) 506–509.
[28] D Teber, S Guven, T Simpfendorfer, et al., Augmented reality: anew tool to improve surgical accuracy during laparoscopic partial nephrectomy? Preliminary in vitro and in vivo results [J], European Urology 56 (2) (2009) 332–338.
[29] F Porpiglia, C Fiori, E Checcucci, et al., Augmented reality robotassisted radical prostatectomy: Preliminary experience [J], European Urology 115 (5) (2018) 184.
[30] F Porpiglia, E Checcucci, D Amparore, et al., Augmented Hreality robot assisted radical prostatectomy using hyper-accuracy three-dimensional reconstruction (HA 3DTM) technology: a radiological and pathological study [J], BJU international 123 (5) (2018) 834–845.
[31] M Lin, Q Chen, S c Yan, Network in network [J], Computer Science (2013) 196–204.
[32] S Ioffe, C Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift [C], in: Proceedings of the 32nd International Conference on Machine Learning, Lille, 2015, pp. 448–456. 6-11 July.
[33] K.K.L. Wong, G. Fortino, D. Abbott, Deep learning-based cardiovascular image diagnosis: A promising challenge, Future Generation Computer Systems 110 (2020) 802–811.
[34] B Lo S, A Lous, S Lin J, et al., Artificial convolution neural network techniques and applications for lung nodule detection [J], IEEE Translation Medicine Imaging 14 (4) (1995) 711–718.
[35] F Chan T, Y Sandberg B, A Vese L, Active Contours without Edges for Vector–Valued Images [J], Journal of Visual Communication & Image Representation 11 (2) (2000) 130–141.
[36] A. Zaim, Automatic Segmentation of the Prostate from Ultrasound Data Using Feature Based Self Organizing Map [C], in: Scandinavian Conference on Image Analysis, Joensuu, Finland, 2005, pp. 1259–1265.