3D Image Segmentation of MRI Prostate Based on a Pytorch Implementation of V-Net

Yizhou Liu
Shanghai Southwest Weiuy Middle School, Shanghai, China
hzhe_wjdi@163.com

Abstract. Prostatitis is one of the common diseases in adult males. At present, the MRI prostate volumetric medical image analysis mainly relies on the radiologist's naked eye localization recognition diagnosis, which is easy to be misdiagnosed due to the fatigue of the doctor. Although the traditional technology uses the Support Vector Machine (SVM) to identify the image, it exists. Manual design features are incomplete and detection accuracy is not high. Recently, with the popularity and breakthrough of convolutional neural networks in the field of computer vision and medical image analysis, the author uses the MRI Prostate MR Image Segmentation 2012 combined with the Vnet method to exploit the depth of open source. Learning framework Pytorch, to achieve 3D image segmentation. The experimental results show that the Dice coefficient of the model on the training set reaches 0.9238 that is, the similarity between the image after cutting and the original image of the original label is 0.9238, and the error value is only 6.42, which is fast and automatic and accurate for intelligent target area painting, thus facilitating the doctor. Diagnosing medical imaging work is conducive to alleviating the imbalance of the proportion of doctors and patients in China.

1. Introduction
At present, benign prostatic hyperplasia is a common disease in middle-aged and elderly people. There are numerous patients, and patients usually go to the hospital for treatment. In the case of uneven doctor-patient ratio in China, computer technology can effectively reduce misdiagnosis caused by doctors due to fatigue.

Meanwhile, in clinical medicine, the prostate-based diagnostic technical solution favors the traditional machine learning solution, that is, the professional hand-designed characteristic index, and the doctor judges according to the domain knowledge with the help of computer imaging. Since the 1980s, MR has become a non-invasive method for assessing lesions in the prostate and its surrounding tissues. Initially, prostate MR relied solely on T1WI, T2WI sequences for morphological evaluation, and its role was mainly for patients with prostate cancer who had been confirmed by biopsy. To a certain extent, MR has the ability to distinguish between benign prostatic lesions, prostate cancer without significant clinical significance, and highly malignant prostate cancer. With advances in MR technology (including MR hardware and software), multiparametric MR (mpMRI) has been gradually applied clinically, combining anatomical (T2WI), functional and pathological assessments, including diffusion-weighted imaging (DWI) and Derived apparent diffusion coefficient (ADC), dynamic enhancement (DCE) MRI, and other techniques such as MR hydrogen proton spectroscopy (MRS). Advances in these
technologies, as well as the accumulation of mpMRI diagnostic experience, have made a qualitative leap in clinical diagnosis of prostate cancer.

Although the above methods have made certain progress, there are still incompleteness of the artificial design features, time-consuming and labor-intensive, and lack of sensitivity to new situations, and the real automation of the detection process is not realized.

In recent years, with the deep success of deep neural networks in the field of image and speech, especially in some aspects, even more than human performance, I naturally think of whether you can try to apply deep learning to medical image analysis. The field is to achieve complete automation of medical diagnostic processes and breakthroughs in high precision and reliability.

Deep learning is the basic tool for the modeling and training of artificial neural networks. Since the early 1980s, artificial neural network technology has been rapidly spread and spread on a global scale. Artificial neural networks are based on the simulation of the structure and thinking of the human brain. The established mathematical model can handle many problems like image recognition, speech recognition, and natural speech processing. In 2006, Hinton et al. proposed the concept of deep learning based on artificial neural network. Deep learning is a deep neural network with many hidden layers. It has better ability of feature learning, and can more abstract and express data. The layer data is initialized to optimize the data model training process, thereby improving the accuracy of the model prediction classification.

2. Method

2.1. Vnet Model

Nassir Navab et al. published a paper on IEEE 3D Vison, V-Net, which is a 3D version of U-Net. In fact, U-Net authors themselves published 3D U-Net.

Contributions: First, 3D image segmentation end2ent model (based on 3D convolution) for MRI prostate volume medical image segmentation. Second, the new objective function is based on Dice coefficient. Third, data expansion methods: random non-linear transformations and histogram matching. Fourth, join the residual learning to improve convergence.

MRI prostate 3D image segmentation is difficult: first, the large appearance variation between different scans (the transformation and change of grayscale distribution). Second, artifacts and distortion caused by main magnetic field inhomogeneities.

Method (V-Net): as shown in Figure 1.
The entire network is divided into a compressed path and an uncompressed path, that is, the feature maps are reduced and enlarged, and each stage reduces the feature by half, that is, \(128-128-64-32-16-8\), and the channel is \(1-16-32-64-128-256\). Each stage joins residual learning to speed up convergence.

The circle plus cross in Figure 1 represents the convolution kernel with \(5*5*5\) and stride of 1 convolution. It can be seen that padding \(2*2*2\) can keep the feature size unchanged. The end of each stage uses a convolution kernel of \(2*2*2\), stride is a convolution of 2, and the feature size is half done (the advantage is that there is no need to save the pooling switches, so the smaller memory footprint). The entire network is a PReLU nonlinear unit proposed by Keiming et al. Add a \(1*1*1\) convolution at the end of the network, process it into the same size as the input, and then connect a softmax.

The idea of network learning U-Net, sending the underlying features of the reduced end to the corresponding position on the magnifying end helps to reconstruct high quality images and accelerates model convergence.

Summary: 3D + residual learning + U-Net thinking

2.2. Dice loss layer

The network prediction consisting of two volumes having the same resolution as the original input data is processed by a soft-max layer that outputs the probability that each voxel belongs to the foreground and the background. In medical quantities such as the work we are working on, it is not uncommon for an anatomical structure of interest to occupy only a very small area of the scan. This often causes the learning process to fall into the local minimum of the loss function, resulting in a network whose predictions are strongly biased towards the background. As a result, the foreground area is often lost or only partially detected. The previous methods used a loss function based on sample re-weighting, where the foreground region is more important during learning than the background region. In this work, we
propose a new objective function based on the Dice coefficient, which is a quantity between 0 and 1, and our goal is to maximize. The Dice coefficient $D$ between two binary volumes can be written as:

$$D = \frac{2 \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2}$$

The sum is run on $N$ three-dimensional pixels, the predicted binary partition volume $p_i \in P$ and the basic real binary volume $g_i \in G$, the form of this dice can be distinguished, and a gradient is generated.

Calculated relative to the predicted $j$th voxel. Using this formula, we don't need to assign weights to different categories of samples to establish the correct balance between foreground voxels and background voxels, and the results we obtain are re-weighted by the sample, and the experimental observations are better than the optimized polynomials. The result of the same network calculation of logic is much better.

3. Experiment Process

3.1. DataSet
We trained our Vnet model on 50 MRI raw prostate raw files and the corresponding real tags after cutting from the "PROMISE2012" challenge dataset, which contains medical data acquired in different hospitals using different devices and different acquisition protocols. The data in this data set represents clinical variability and the challenges encountered in the clinical setting. In training, we add a large amount of this data set to each small batch fed to the network by a random transformation performed in each training iteration. The small batch used in our implementation contained two original images of the prostate, mainly due to the high memory requirements of the model during training. We use a momentum of 0.99 and an initial learning rate of 0.001, which is an order of magnitude reduction per 25K iteration.

The hyperparameters trained on Pytorch are shown below:

| Hyperparameters | Initial Value | Method |
|-----------------|---------------|--------|
| Momentum        | 0.99          | N/A    |
| Learning rate   | 0.001         | Learning rate is reduced by an order of magnitude every 20k iterations |

3.2. Data augumentation
The original data set is 50 MRIs. Randomly distorted the original data through a dense 2x2x2 grid.

During each training iteration, we use a densely deformed field obtained by $2 \times 2 \times 2$ grid control points and B-spline interpolation as input to input a randomly deformed version of the training image $b$ to the network. This enhancement has been performed "on the fly" before each optimization iteration to mitigate excessive storage requirements. In addition, we change the intensity distribution of the data by using histogram matching to adjust the intensity distribution of the training volume used in each iteration to other randomly selected scans belonging to the data set.

3.3. Training process
Our CNN performs end-to-end training of prostate scan data sets in MRI. An example of typical content of such a volume is shown in Figure 2.
All volumes processed by the network have a fixed size of 128 x 128 x 64 voxels and a spatial resolution of 1 x 1 x 1.5 mm. Since one or more experts are required to manually mark the correct cut image label and there is a cost associated with its acquisition, it is not easy to obtain an annotated medical amount. In this work, we found it necessary to add raw training data sets to achieve robustness and higher accuracy on the test data set.

During each training iteration, we input a randomly deformed version of the training image to the network by using a densely deformed field obtained by 2 x 2 x 2 grid control points and B-spline interpolation as input. This enhancement has been performed "on the fly" before each optimization iteration to mitigate excessive storage requirements. In addition, we change the intensity distribution of the data by using histogram matching to adapt the intensity distribution of the training volume used in each iteration to other randomly selected scans belonging to the data set.

During the run, we set the batch_size equal to 2, which is fixed for 1 round. First, in Vnet's Pytorch code, we set 1000, iteration to observe the dice coefficient and the change in the error value, but only selected 26 of the 50 MRI original images. In order to further verify the relationship between the similarity correlation coefficient and iteration, we set the iteration equal to 2000 under the condition of a batch sample size, the number of training set samples and the number of rounds unchanged, but the number of test set samples is 30 MRI original prostate image.

At the same time, in order to verify, the relationship between the number of data sets in the training set and the dice coefficient, we further experiment, each batch is still 2 original volume prostate images, 1 round, the test set size is equal to 30, respectively set iteration Equal to 2000,4000.

3.4. Testing process
The previously invisible MRI volume can be segmented by processing it in a network feedforward manner. After soft-max, the output of the last convolutional layer consists of a probability map of the background and foreground. Voxels with a higher probability (> 0.5) belonging to the foreground rather than the background are considered part of the anatomy.

4. Data Analysis
Finally, we infer the 3D prostate image segmentation results as shown below:

Figure 3. The lower left corner is a perspective view of the segmented prostate.
Our implementation is implemented in Python, a custom version on the Pytorch framework. A total of 72 hours of training were conducted, and 5 experiments were performed, and different experiments were performed for 26 samples and 50 samples, respectively. Experiments show that the dice coefficient has a positive correlation with the number of iterations. Compared with the number of samples in the training set, the average dice coefficient reaches 0.887 on the original volume prostate image of 30 test sets, and the score of "Promise 12" reaches 0.9046.

The data set is first normalized using the N4 bias field correction function of the ANT framework and then resampled to a common resolution of 1 x 1 x 1.5 mm. We applied random deformation to the scan for training by changing the position of the control points, where the random quantities were obtained from the Gaussian distribution with zero mean and 15 voxel standard deviation. The qualitative results are as follows:

![Figure 4. Qualitative results of PROMISE 2012 data set](image)

5. Conclusion and Discussion

This paper is based on a volumetric convolutional neural network implementation on the Pytorch deep learning framework, which performs MRI prostate volume segmentation in a fast and accurate manner. We introduce a new objective function that we optimize during training based on the dice overlap coefficient between the predicted segmentation and the real segmentation label. When the number of background and foreground pixels is very unbalanced, our Dice loss layer does not require sample re-weighting and is indicated for binary segmentation tasks. Although we have borrowed our architecture from the U-net network, we divide it into stages of learning residuals and, based on experience, they can improve results and convergence time. The final experimental results show that our dice coefficient reaches 0.9238 on 26 3d original MRI prostate images, indicating that the model can accurately segment 3d images. Future work will split volumes of multiple regions containing other modalities (such as ultrasound and higher resolution) by dividing the network into multiple GPUs.

References

[1] Cha, K.H., Hadjiiski, L., Samala, R.K., Chan, H.P., Caoili, E.M., Cohan, R.H.: Urinary bladder segmentation in ct urography using deep-learning convolutional neural network and level sets. Medical Physics 43 (4), 1882-1896 (2016).

[2] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. arXiv preprint arXiv: 1512.03385 (2015)

[3] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. arXiv preprint arXiv: 1408.5093 (2014)

[4] Kamnitsas, K., Ledig, C., Newcombe, V.F., Simpson, J.P., Kane, A.D., Menon, D.K., Rueckert, D., Glocker, B.: Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. arXiv preprint arXiv: 1603.05959 (2016)