

# PROCEEDINGS OF SPIE

[SPIEDigitalLibrary.org/conference-proceedings-of-spie](https://SPIEDigitalLibrary.org/conference-proceedings-of-spie)

## Automatic segmentation of the prostate on CT images using deep learning and multi-atlas fusion

Ling Ma, Rongrong Guo, Guoyi Zhang, Funmilayo Tade, David M. Schuster, et al.

Ling Ma, Rongrong Guo, Guoyi Zhang, Funmilayo Tade, David M. Schuster, Peter Nieh, Viraj Master, Baowei Fei, "Automatic segmentation of the prostate on CT images using deep learning and multi-atlas fusion", Proc. SPIE 10133, Medical Imaging 2017: Image Processing, 101332O (24 February 2017); doi: 10.1117/12.2255755

**SPIE.**

Event: SPIE Medical Imaging, 2017, Orlando, Florida, United States

# Automatic segmentation of the prostate on CT images using deep learning and multi-atlas fusion

Ling Ma<sup>a,b</sup>, Rongrong Guo<sup>a</sup>, Guoyi Zhang<sup>a</sup>, Funmilayo Tade<sup>a</sup>,  
David M. Schuster<sup>a</sup>, Peter Nieh<sup>c</sup>, Viraj Master<sup>c</sup>, Baowei Fei<sup>a,d,e,\*</sup>

<sup>a</sup>Department of Radiology and Imaging Sciences, Emory University, Atlanta, GA

<sup>b</sup>School of Computer Science, Beijing Institute of Technology

<sup>c</sup>Department of Urology, Emory University, Atlanta, GA

<sup>d</sup>Winship Cancer Institute of Emory University, Atlanta, GA

<sup>e</sup>The Wallace H. Coulter Department of Biomedical Engineering  
Georgia Institute of Technology and Emory University, Atlanta, GA.

\* E-mail: bfei@emory.edu ; Web: <http://feilab.org>

## ABSTRACT

Automatic segmentation of the prostate on CT images has many applications in prostate cancer diagnosis and therapy. However, prostate CT image segmentation is challenging because of the low contrast of soft tissue on CT images. In this paper, we propose an automatic segmentation method by combining a deep learning method and multi-atlas refinement. First, instead of segmenting the whole image, we extract the region of interesting (ROI) to delete irrelevant regions. Then, we use the convolutional neural networks (CNN) to learn the deep features for distinguishing the prostate pixels from the non-prostate pixels in order to obtain the preliminary segmentation results. CNN can automatically learn the deep features adapting to the data, which are different from some handcrafted features. Finally, we select some similar atlases to refine the initial segmentation results. The proposed method has been evaluated on a dataset of 92 prostate CT images. Experimental results show that our method achieved a Dice similarity coefficient of 86.80% as compared to the manual segmentation. The deep learning based method can provide a useful tool for automatic segmentation of the prostate on CT images and thus can have a variety of clinical applications.

**Keywords:** Prostate, image segmentation, computed tomography (CT), convolutional neural networks (CNN), multi-atlas segmentation, deep learning, prostate cancer

## 1. INTRODUCTION

Prostate cancer is one of the major causes of cancer mortality in American men. It is estimated that there were 180,890 new cases of prostate cancer in the United States in 2016 [1]. Accurate segmentation of the prostate on CT images has many applications in the diagnosis and therapy of this disease. However, manual segmentation is time-consuming and subjective to the experience of the physician. Some semi-automatic segmentation methods [2-4] need the user interaction. Automatic segmentation methods for the prostate on CT images have been proposed in recent years [5-10]. Some methods [5-8] need the information from the same patient for segmentation. The method in [9] used the shape of the prostate from one patient to guide the segmentation. As reported in [10], the distribution of the intensity histograms inside and outside the organ contours is important for prostate segmentation. The afore-mentioned methods use the handcrafted features for segmentation. However, it can be difficult to extract effective features. Deep learning based methods play an increasing role in image classification, natural language processing, and medical image analysis. For example, Guo et al. [11] proposed prostate MR segmentation method using deep features and sparse patch matching. They learned the latent feature representation from prostate MR images using the stacked sparse auto-encoder.

In this paper, we propose a deep learning based method to segment the prostate on CT images. We adopt the convolutional neural networks (CNN) to learn the deep feature and classify the pixels into the prostate and non-prostate for preliminary segmentation of the gland. We then use a multi-atlas method to segmentation the prostate.

## 2. METHOD

Our automatic segmentation method for the prostate on CT images contains three parts: extraction of regions of interest (ROI), CNN based prostate segmentation, and multi-atlas label fusion. Figure 1 shows the overview of the proposed method.

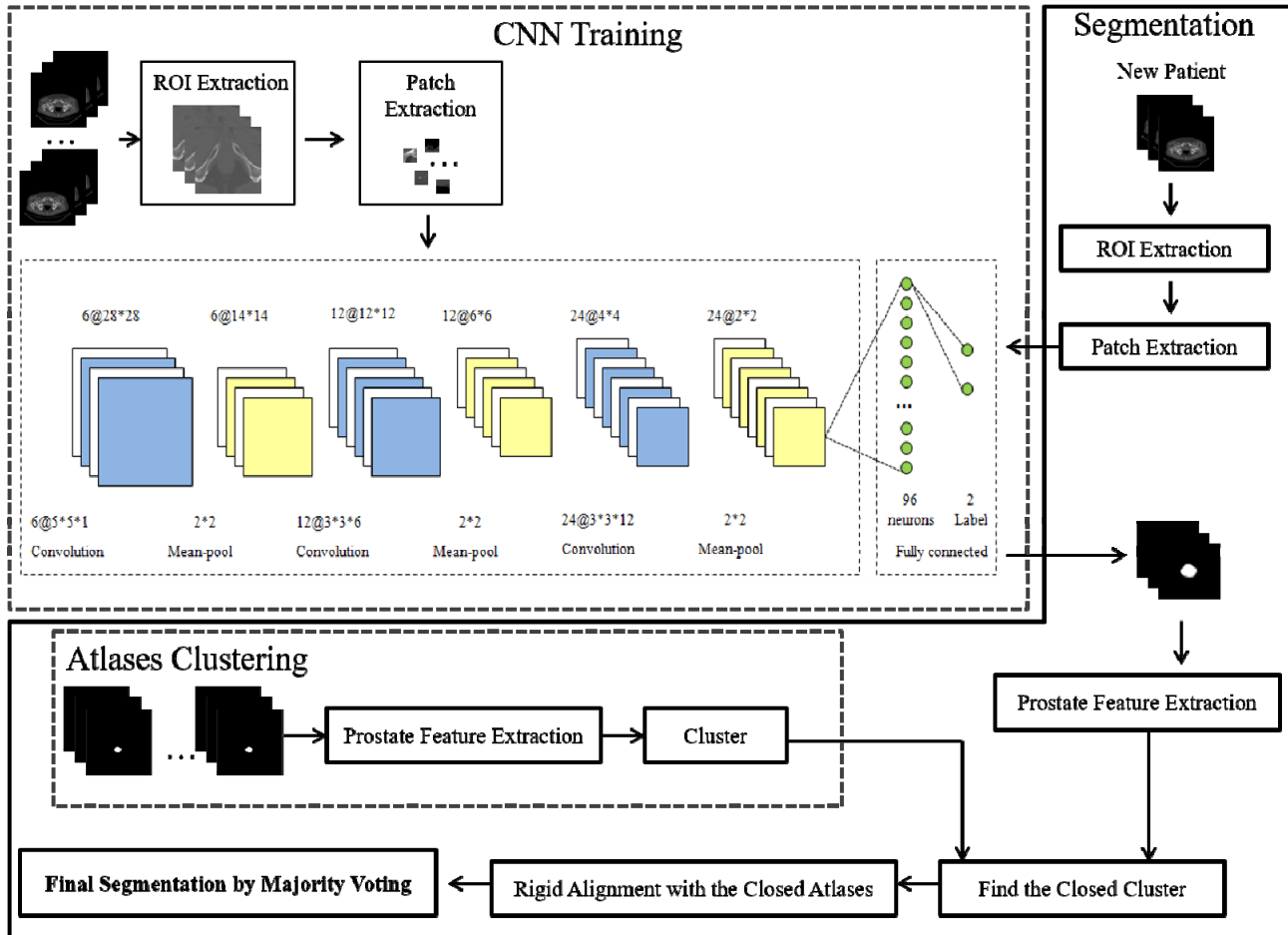


Figure 1. An overview of the proposed deep learning based segmentation method for 3D prostate CT images.

### 2.1 ROI extraction

Since the prostate is a small region of interest (ROI) on the pelvic CT image, the numbers of prostate and non-prostate pixels are not balanced. In order to extract the ROI, we use three processing steps. 1) We delete the bed of the scanner on the CT images. Because the bed on each slice locates in an approximately same position on the CT images, we overlap all the 2D slices to extract the bed. 2) After removing the bed from the images, we use a threshold to detect the pelvic bone on each slice. 3) We extract the ROI that includes the prostate region. The size of the extracted ROI is  $161 \times 161 \times 27$ .

### 2.2 Patch extraction

For each pixel, we use a patch of  $32 \times 32$  pixels centered at the pixel for the representation. Since a too big or too small size of the patch leads to adding noise or losing information, we choose  $32 \times 32$  as the batch size.

### 2.3 CNN based prostate segmentation

As described in the previous section, it is necessary to learn adaptive features. Thus, we introduce the convolutional neural networks (CNN) to learn the feature representation of each block and classify them into the prostate or non-prostate pixels for the segmentation. CNN is a type of feed-forward artificial neural network [12]. It can achieve a good performance by analyzing the stationarity of statistics and locality of pixel dependencies.

We construct a 2D CNN for the pixel classification. The 2D CNN consists of three consecutive convolutional layers and mean-pooling layers (Figure 1). The input is a  $32 \times 32$  patch of ROI extracted from the training images. The first, second, and third convolutional layer consists of 6 kernels with the size of  $5 \times 5 \times 1$ , 12 kernels with the size of  $3 \times 3 \times 6$ , and 24 kernels with the size of  $3 \times 3 \times 12$ , respectively. As shown in Figure 1, the outputs of the three layers are 6 blocks of  $28 \times 28$  (denoted as  $6 @ 28 \times 28$ ), 12 blocks of  $12 \times 12$ , and 24 blocks of  $4 \times 4$  for the three layers, respectively. The mean-pooling layers are followed by the convolutional layers. It can output the mean values in non-overlapping windows with the size of  $2 \times 2$  and stride of 2, thus reduces the size of patches by half. The last layer is a fully connected layer with 96 neurons and the output has two neurons, meaning the category: prostate and non-prostate.

### 2.4 Multi-atlas label fusion

We combine multiple atlases with the preliminary segmentation result from the CNN to generate the final segmentation result. Although the prostates from different patients have different appearances or sizes, their anatomical structure is similar and can be used to guide the segmentation. To select similar atlases, we cluster all the atlases into  $k$  groups according to the features of the atlases, including the size feature and shape feature [13]. We adaptively determine the value  $k$  according to the minimum distance within the group. The distance within the group  $DG(k)$  is computed by

$$DG(k) = \sum_{i=1}^k \sum_{j=1}^{N_i} (F_{ij} - \text{centroid}_i) \quad (1)$$

where  $F_{ij}$  is the features of the  $j$ -th atlas which is clustered into the  $i$ -th group,  $N_i$  is the number of atlases in the  $i$ -th group, and  $\text{centroid}_i$  is the centroid of the  $i$ -th group. The optimal number of groups,  $k^*$ , can be chosen according to the minimum distance within the group

$$k^* = \arg \min DG(k). \quad (2)$$

Based on the clustered atlases, we can find the similar group by clustering the preliminary segmentation result. We rigidly align the atlases in the similar group with the preliminary segmented prostate and conduct the majority voting for each pixel to achieve the final segmentation result.

### 2.5 Evaluation criterion

We use the Dice similarity coefficient (DSC) to evaluate the segmentation performance [15-18]. DSC is the relative volume overlap between the binary masks from our method and the manual segmentation by radiologists. The DSC can be computed as:

$$DSC = \frac{2 \times TP}{(FP + TP) + (TP + FN)}, \quad (3)$$

where TP, TN, FP, FN is the number of true positives, true negatives, false positives and false negatives, respectively. If a positive example for a prostate pixel can be recognized correctly, we call it "true positive". If a prostate pixel is classified as a non-prostate pixel incorrectly, we call it "false negative". The meanings of "true negative" and "false positive" are defined similarly.

## 3. EXPERIMENTS

### 3.1 Databases

The segmentation method was tested in CT image volumes from 92 patients. The slice thickness is 4.25 mm and the size of the images is  $512 \times 512 \times 27$  voxels. The in-plane resolution is  $0.977 \times 0.977 \text{ mm}^2$ . The prostate was manually

segmented by an experienced radiologist to produce the gold standard for segmentation evaluation. The set of the volumes from 92 patients is nearly evenly divided into five disjoint subsets. We conducted five-fold cross validation experiments for the segmentation. Specifically, we take each of the five subsets as the testing dataset, randomly select one subset as the validation set for parameter setting, and the three remaining subsets are the training set.

### 3.2 Result for the ROI extraction

Figure 2 shows the ROI extraction process. The images in the first column are the 2D original images. The images in the second and third columns show the detected bed region and bones. The images in the last column show the ROI in the 2D slices where the red rectangles are the prostate ROIs. We extracted the region inside the red box as our ROI and the blue curves are the prostate. It shows our ROIs are large enough to cover the whole prostate region.

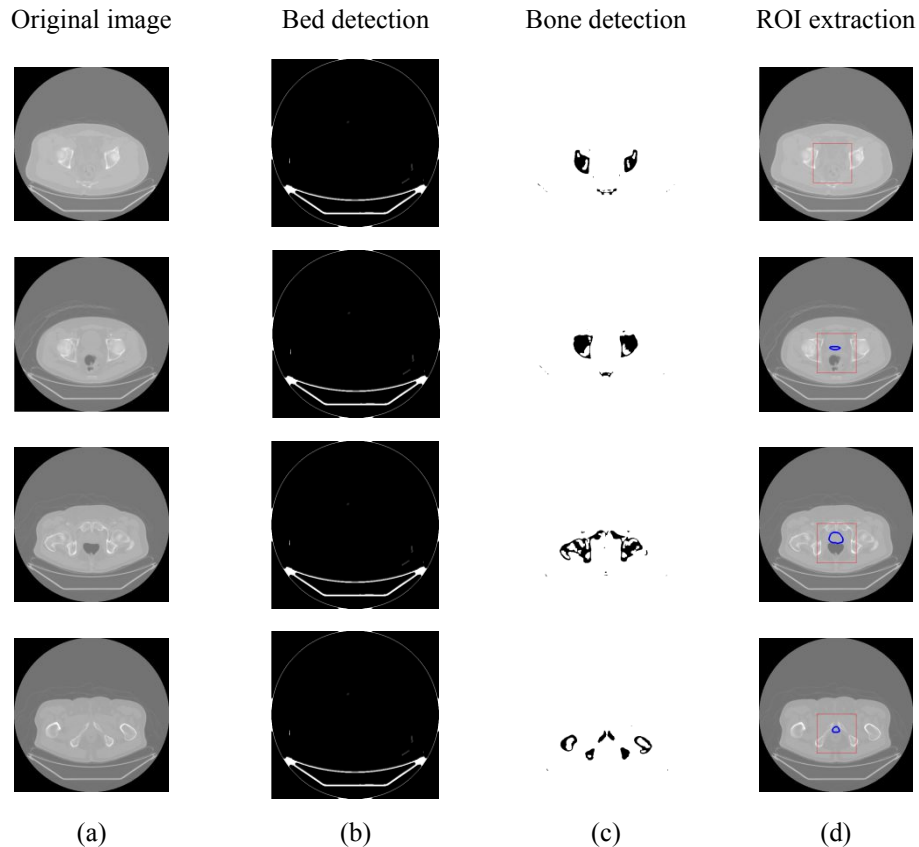


Figure 2. The extraction processing for the prostate region of interest (ROI). (a) The original CT images, (b) The detected bed, (c) The detected bones, and (d) The extracted ROIs, where the red rectangle boxes are the extracted ROIs and the blue contours are the prostate region segmented by the radiologist.

### 3.3 Parameters setting

There are two groups of parameter sets which are determined by experiments. One is the hyperparameters for CNN, and the other is the number  $k$  for the atlases clustering.

#### 3.3.1 The parameters for CNN training

The CNN framework is implemented using the method in [14]. We test the classification performance of CNN on the validation data set and determine three parameters for the CNN training, which include the learning rate (alpha), the batch size, and the number of epochs. We test the alpha from 0.1 to 1 with a step size of 0.1, the batch size from 10 to

100 with a step size of 10, and the epochs with 1 and values from 10 to 60 with step size of 10. When we test the performance influenced by one parameter on the validation set, we keep the other parameters stable. For example, if we test the influence of the alpha on the segmentation performance, we fix the epochs as 1 and the batch size as 50. In the same way, we fix the alpha as 1 and epochs as 1 to test the influence of the batch size, and we fix alpha as 1 and the batch size as 50 to test the influence of epochs. The effects of different parameters on the segmentation are shown in Figure 3-5. Finally, we select the best parameter with the highest DSC for the three parameters (alpha, batch size, epochs). They are (0.8, 10, 10), (0.6, 70, 40), and (0.1, 100, 50), for the five-fold validation, respectively. We then used them in all the following experiments.

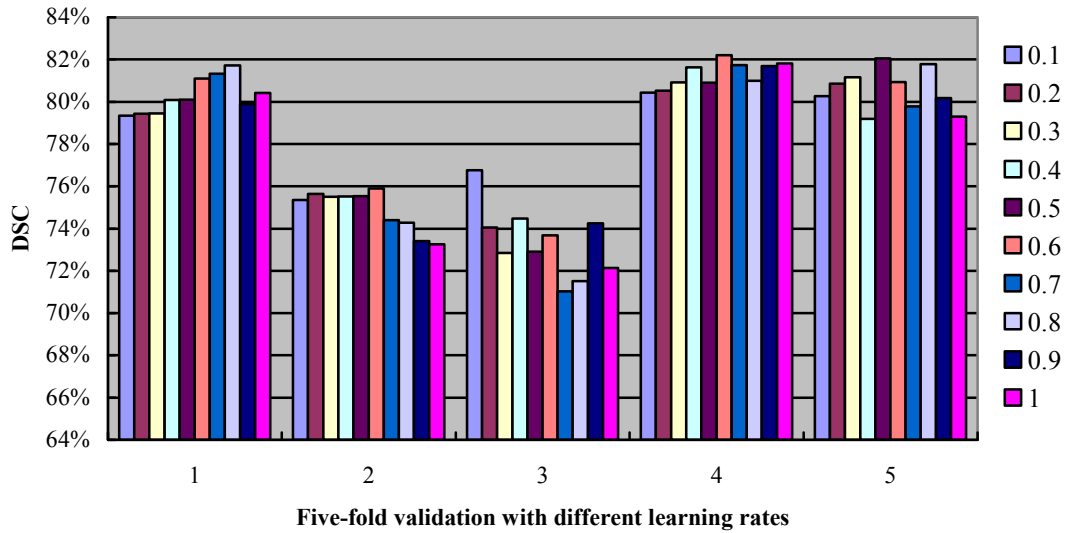


Figure 3. Five-fold segmentation validation with different alphas (from 0.1 to 1) with a fixed batch size of 50 and a fixed epoch of 1.

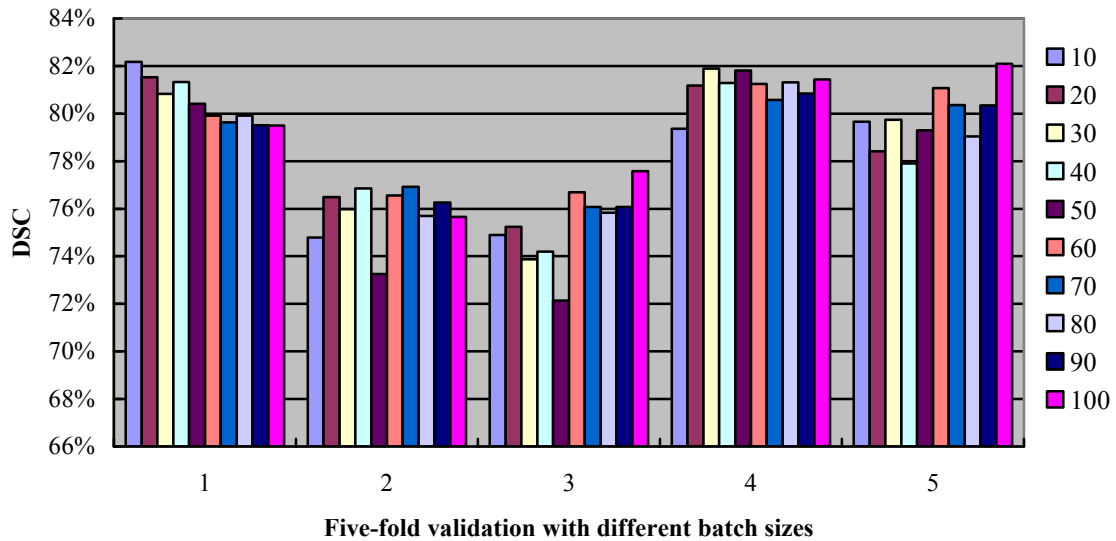


Figure 4. Five-fold segmentation validation with different batch sizes (from 10 to 100) with a fixed alpha of 1 and a fixed epoch of 1.

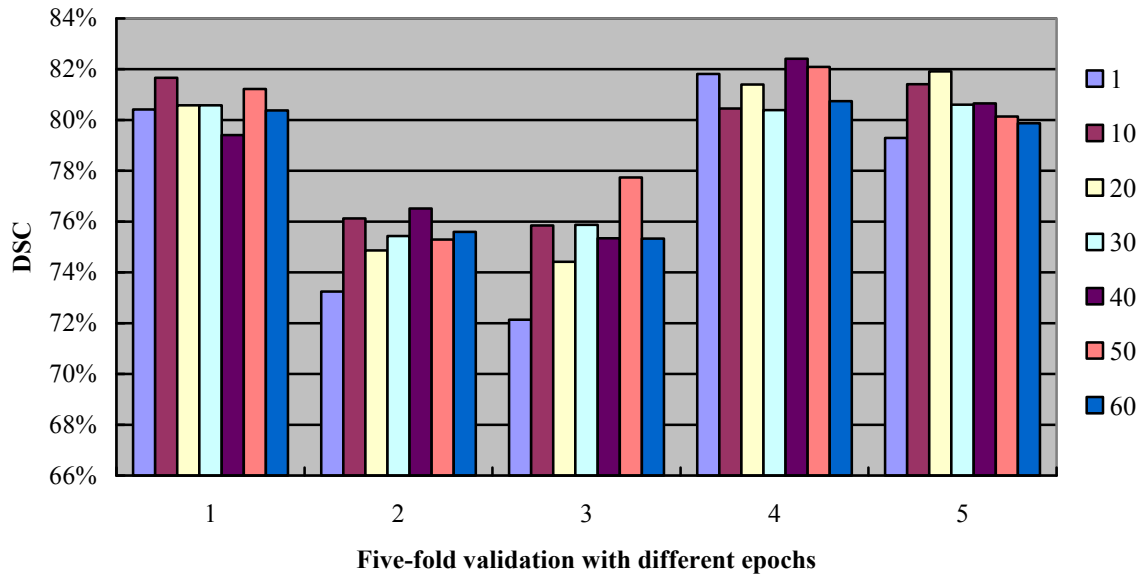


Figure 5. Five-fold segmentation validation with different epochs (from 1 to 60) with a fixed alpha of 1 and a fixed batch size of 50.

### 3.3.2 The parameters for the atlas clustering

To select the similar atlases for guiding the segmentation, we cluster the atlases into  $k$  groups, we determine the parameter  $k$  according to Eq. (2). We record the distance within groups under different  $k$  in the five-fold validation, as shown in Figure 6. We choose the value  $k$  in the five-fold validation as 15, 21, 21, 21, and 19 because they provide the smallest distances.

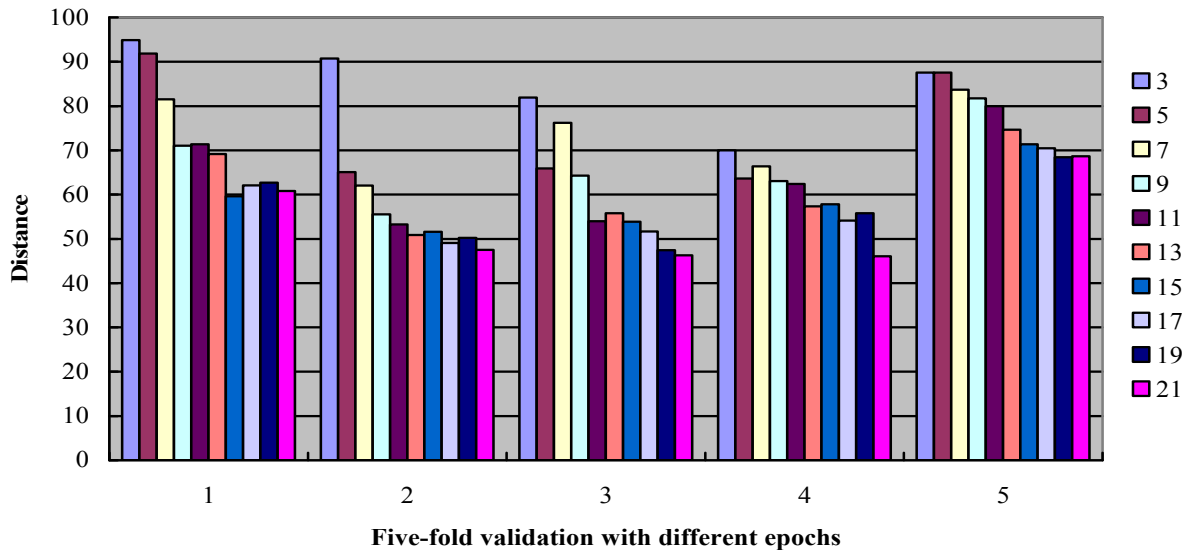


Figure 6. The distance in the five-fold segmentation validation with different  $k$  for the atlas clustering.

### 3.4 Qualitative results

Figure 7 shows the segmentation results of four patients. The contours by the automatic segmentation are close to those manually segmented by the radiologist, indicating the automatic method performs well.

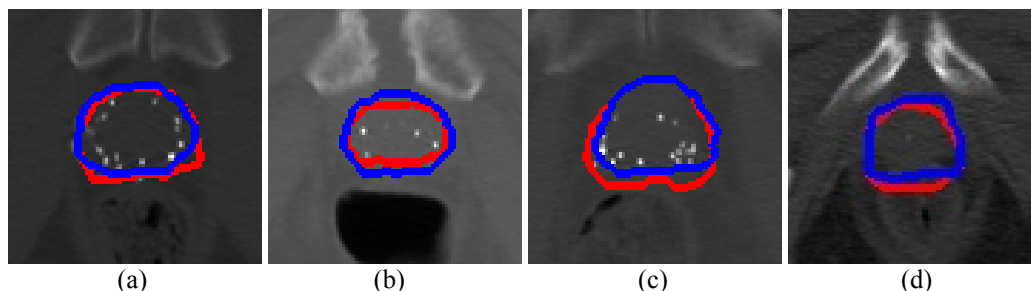


Figure 7. Qualitative evaluation of the prostate segmentation on four CT slices of different patients. The red curves are the prostate manually segmented by the radiologist, while the blue curves are the segmentation results by our automatic method.

### 3.5 Quantitative results

Table 1 shows the quantitative evaluation results from the five-fold validation with the 92 CT image volumes. We divided each volume into three sub-regions: the apex, middle, base regions, which contain 30%, 40%, and 30% slices of the prostate in the transverse direction, respectively. We measure the DSC for the whole volume and the three sub-regions. The DSC is 86.80% for the whole volume. The minimum DSCs are 79.91% and 75.26% for the apex and base regions, respectively. Note that it is usually difficult to segment the apex and base of the prostate.

Table 1. Segmentation performance for the whole gland and the three sub-regions (apex, mid-gland, and base) in 92 patients.

Five-fold	DSC (%)			
	whole	apex	middle	base
1	82.68±6.15	79.91±13.18	92.11±4.75	75.26±13.04
2	86.80±6.40	83.59±13.17	93.49±5.76	84.16±7.96
3	85.14±8.76	83.53±12.20	89.95±11.87	81.48±10.12
4	84.27±5.70	86.54±7.23	87.74±8.84	76.09±7.84
5	82.06±5.38	80.78±6.81	89.53±6.12	78.73±9.99

### 3.6 Effectiveness of the proposed method

In this paper, we adopt a two-stage method for the segmentation. Specifically, we use the CNN to learn the deep feature to obtain the preliminary segmentation result, and then use the multi-atlases to refine it for the final segmented prostate. In order to highlight the advantages of our method, we compare our method with the method involving only the deep learning algorithm and the multi-atlas algorithm, respectively. We use the M\_DL and M\_ATLASES to represent them, respectively. We tested the two methods on the same database and the results are shown in Table 2. By comparing the results in Tables 1 and 2, our method can take the advantage of both deep learning and multi-atlases and achieve the best segmentation performance, not only for the whole volume, but also for the apex, middle, base regions.

### 3.7 Comparison with the other methods

To prove the effectiveness of the deep learning method, we compare our method with two other methods [3, 8]. We changed the method in [3] to become automatic segmentation by removing the interaction part. The method [3] extracted the handcrafted features that include the CT histogram feature, multi-scale rotation invariant LBP feature, GLCM feature, HOG feature, and Haar-like feature. The method in [8] used a different database of 40 CT images. The mean DSCs are



0.759, 0.82, and 0.842 for Method [3], Method [8], and our deep learning method, respectively. The deep feature learned by CNN is better than the handcrafted features as used in [3]. The deep learning based segmentation method is more effective than other two segmentation methods.

Table 2. Segmentation performance of the two methods (M\_DL and M\_ATALAES) for the whole gland and the three sub-regions (apex, mid-gland, and base) in 92 patients.

Five-fold	DSC(M_DL)				DSC(M_ATALAES)			
	Whole	Apex	Middle	Base	Whole	Apex	Middle	Base
1	78.6±7.4	71.9±14.7	92.8±5.4	71.7±13.0	72.9±8.2	70.1±10.0	84.5±8.0	67.8±12.7
2	81.1±6.5	76.1±15.4	94.3±4.9	76.1±7.5	72.4±9.5	69.5±11.3	85.1±7.9	66.3±13.0
3	74.6±12.7	67.9±20.8	84.2±16.9	70.7±15.5	69.2±8.1	67.4±9.8	78.1±10.5	64.4±12.1
4	77.7±8.9	72.8±12.7	87.3±13.8	76.4±6.4	67.3±7.3	64.1±7.3	77.0±10.4	62.5±10.9
5	79.9±5.7	76.1±6.7	92.7±7.9	76.3±6.4	69.2±7.0	62.4±12.2	82.7±4.9	66.5±9.1

#### 4. CONCLUSIONS

In this paper, an automatic segmentation method has been proposed to segment the prostate on CT images. We used convolutional neural networks to learn the characteristic of the prostate and non-prostate pixels to obtain the preliminary segmented prostate. By selecting the similar atlases, we refine the segmentation and obtain the final contour of the prostate. The two-stage segmentation method is more effective as compared to the methods only involving the deep learning algorithm or multi-atlas label fusion. The proposed method also performs better than the method that only uses handcrafted feature. The automatic segmentation method can be applied to various applications in prostate cancer diagnosis and therapy.

#### 5. ACKNOWLEDGEMENTS

This research is supported in part by NIH grants (CA176684, R01CA156775, and CA204254).

#### REFERENCES

- [1] Siegel, R.L., Miller, K.D., and Jemal, A., "Cancer statistics, 2016," *CA: a cancer journal for clinicians*, 66(1), 7-30 (2016).
- [2] Park, S. H., et al., "Interactive prostate segmentation using atlas-guided semi-supervised learning and adaptive feature selection," *Medical physics*, 41(11), 111715 (2014).
- [3] Ma, L., et al., "Combining population and patient-specific characteristics for prostate segmentation on 3D CT images," in *SPIE Medical Imaging, International Society for Optics and Photonics*, 978427-978427 (2016).
- [4] Shi, Y., et al., "Prostate segmentation in CT images via spatial-constrained transductive lasso," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2227-2234 (2013).
- [5] Wu, Y., et al., "Prostate segmentation based on variant scale patch and local independent projection," *IEEE transactions on medical imaging*, 33(6), 1290-1303 (2014).
- [6] Li, W., et al. "Learning image context for segmentation of prostate in CT-guided radiotherapy," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 570-578 (2011).
- [7] Liao, S. and Shen, D., "A feature-based learning framework for accurate prostate localization in CT images," *IEEE transactions on image processing*, 21(8), 3546-3559 (2012).
- [8] Davis, B., et al., "Automatic segmentation of intra-treatment CT images for adaptive radiation therapy of the prostate," *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, 442-450 (2005).
- [9] Martínez, F., et al., "Segmentation of pelvic structures for planning CT using a geometrical shape model tuned by a multi-scale edge detector," *Physics in medicine and biology*, 59(6), 1471 (2014).

- [10] Chen, S., Lovelock, D.M., and Radke, R.J., "Segmenting the prostate and rectum in CT imagery using anatomical constraints," *Medical image analysis*, 15(1), 1-11 (2011).
- [11] Guo, Y., Gao, Y., and Shen, D., "Deformable MR prostate segmentation via deep feature learning and sparse patch matching," *IEEE transactions on medical imaging*, 35(4), 1077-1089 (2016).
- [12] Krizhevsky, A., Sutskever, I., and Hinton, G.E. "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 1097-1105 (2012).
- [13] Armon, S., *et al*, "Handwriting Recognition and Fast Retrieval for Hebrew Historical Manuscripts," Hebrew University of Jerusalem, (2011).
- [14] Palm, R.B., "Prediction as a candidate for learning deep hierarchical models of data," Technical University of Denmark, (2012).
- [15] Fei, B., *et al*, "MR/PET quantification tools: Registration, segmentation, classification, and MR-based attenuation correction," *Medical physics*, 39(10), 6443-6454 (2012).
- [16] Akbari, H, Fei, B. "3D ultrasound image segmentation using wavelet support vector machines," *Medical physics*, 39(6), 2972-2984 (2012).
- [17] Yang, X., Wu, S., Sechopoulos, I., and Fei, B. "Cupping artifact correction and automated classification for high-resolution dedicated breast CT images," *Medical physics*, 39(10), 6397-6406 (2012).
- [18] Wang, H., and Fei, B. "An MR image-guided, voxel-based partial volume correction method for PET images," *Medical physics*, 39(1), 179-194 (2012).