A weakly supervised registration-based framework for prostate segmentation via the combination of statistical shape model and CNN

Chunxia Qin*a,b,c, Xiaojun Chenb,∗, Jocelyne Troccazc

*aSchool of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China; bSchool of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai, China; cUniv. Grenoble Alpes, CNRS, Grenoble INP, TIMC-IMAG, Grenoble, France

Abstract

Precise determination of target is an essential procedure in prostate interventions, such as the prostate biopsy, lesion detection and targeted therapy. However, the prostate delineation may be tough in some cases due to tissue ambiguity or lack of partial anatomical boundary. To address this problem, we proposed a weakly supervised registration-based framework for the precise prostate segmentation, by combining convolutional neural network (CNN) with statistical shape model (SSM). To obtain the prostate region, an inception-based neural network (SSM-Net) was firstly exploited to predict the model transform, shape control parameters and a fine-tuning vector, for the generation of prostate boundary. According to the inferred boundary, a normalized distance map was calculated. Then, a residual U-net (ResU-Net) was employed to predict a probability label map from the input images. Finally, the average of the distance map and the probability map was regarded as the prostate segmentation. After that, two public dataset PROMISE12 and NCI-ISBI 2013 were utilized for the model computation and for the network training and testing. The validation results demonstrate that the segmentation framework using a SSM with 9500 nodes achieved the best performance, with a dice of 0.904 and an average surface distance of 1.88 mm. In addition, we verified the impact of model elasticity augmentation and fine-tuning item on the network segmentation capability. As a result, both factors have improved the delineation accuracy, with dice increased by 10% and 7% respectively. In conclusion, via the combination of two weakly supervised neural networks, our segmentation method might be an effective and robust approach for prostate segmentation.

Keywords: Weakly supervised, Registration-based segmentation, Statistical shape mode, probability map, boundary distance map

1. Introduction

With Magnetic Resonance (MR) imaging become an increasingly important non-invasive imaging modality [Püttner et al. 2015], prostate MR segmentation has been paid close attention in recent years, as it is crucial for the clinical diagnosis, therapeutic procedure and treatment planning of various prostate disorders (e.g., prostate cancer, prostatitis or prostatic hypertrophy) [Litjens et al. 2014a]. For example, prostate delineation is widely applied for the precise localization of prostate boundary in radiotherapy. Besides, in the image guided computer assisted surgery, the segmentation of prostate on preoperative MRI is an essential reference for the inter-operative low-quality image, like ultrasound image [Khalilaghi et al. 2015].

However, until now prostate on MR images is still mostly segmented manually by radiologist. The hand-crafted delineation of prostate boundary is a time-consuming and labor-intensive operation with an extremely low reproducibility because of its high dependence on medical experience. Moreover, those problems are further aggravated when the borderline is indistinct. As pointed out by Yu et al. (2017), automatic prostate segmentation is also a challenging task due to the issue of intensity inhomogeneity, variation of anatomical appearance and lack of boundary discriminability (as shown in Fig. 1).

In order to address this challenging task, different automatic or semi-automatic segmentation approaches have been reported in recent years. Martin et al. (2008), Martin et al. (2010) proposed a semi-automatic prostate segmentation method, in which a rigid intensity-based registration algorithm and a non-rigid hybrid registration framework were employed successively to align the atlas to patient image. In their work, 18 MRI series were involved to construct the atlas. Two accuracy metrics respectively based on volume and surface distance, were used to investigate the segmentation performance. Results showed that the segmentation accuracy of the apex region and the central region is higher than the base part. In another publication, to add additional knowledge into the segmentation procedure, Korsager et al. (2015) combined the spatial information of a prostate atlas with the intensity information in a graph cut segmentation framework to achieve automatic prostate delineation. Their validation experiment was investigated on 76 axial MR images. As a result, a mean Dice similarity coefficient (DSC) of 0.88 and a mean surface distance of 1.45 mm were reported. Besides, Tian et al. (2015) utilized a superpixel based graph cut framework to acquire the prostate surface on MRI. A superpixel is a group of pixels which have similar characteristics such as intensity or location. Due to the capacity to carry wider information, serving as a more convenient and compact representation of original image, the superpixel image has been widely used in image segmentation algorithms. In Tian study, a graph cuts
algorithm and an active contour model were integrated together for cross-promotion. According to their experiment results, the verification on 43 MRI examples obtained a mean dice of 0.893.

Recently, the performance of deep learning approaches has become comparable to state-of-the-art methods in many fields, especially in computer vision (Goodfellow et al. 2014; Ren et al. 2015) and medical image processing (Pereira et al. 2010; Tajbakhsh et al. 2016). In those researches, neural networks often work as information extractors to eliminate the tedious procedure of traditional feature choice and collection. For precise segmentation of MRI prostate, Guo et al. (2015) used more concise and effective hierarchical features from MRI prostate image by utilizing a stacked sparse auto-encoder. Based on the extracted features, a sparse patch matching method was employed to deduce the corresponding prostate likelihood map, which was further combined with a sparse shape model for the final segmentation. Besides, Mun et al. (2017) integrated encoding, bridge, decoding and classification modules to develop a baseline convolutional neural network to extract volumetric information. Meantime, Jia et al. (2017) researched a coarse-to-fine algorithm for MRI prostate segmentation through a deep learning method. In their algorithm, a registration-based segmentation was firstly used to obtain a rough prostate region. Then a pixel-wise recognizer based on neural network was further adopted to classify the prostate boundary from the image background. Finally, a refinement algorithm was applied to smooth the contour. Similar to Jia’s work, He et al. (2017) exploited another coarse-to-fine prostate segmentation system via different algorithms. They firstly proposed an adaptive feature learning probability boosting tree for prostate pre-segmentation. Next, a CNN method was developed to obtain the prostate profile model by the judgement of the inner, external and boundary points. For the last step, an active shape model was employed for the final surface refinement. In their work, the neural network was introduced for the extraction of latent image features and the prediction of prostate boundary. As a consequence, their method became more accurate and robust for prostate segmentation. In addition, Wang et al. (2019) introduced a 3D deep-supervised full convolutional network with group dilated convolution, aiming to preserve extra image information for prostate delineation. Their method achieved a dice of 0.86. Generally, compared with the traditional segmentation algorithms, the approaches based on deep learning method can delineate the target more accurate in less time. However, the aforementioned ways only involved the information of the specific input image while without any prior knowledge constraints which is potentially helpful to improve the prostate segmentation accuracy.

SSM is a geometric model containing a mean shape and multiple compressed primarily shape variations of a collection of similar shapes. Due to the ability to represent prior geometric information, SSM has been widely applied in different medical modality for the segmentation and registration of various anatomical structures, including brain, bone, liver, heart, prostate and so on (Shen et al. 2001; Seim et al. 2008; Zhang et al. 2010; Alb et al. 2016). For prostate segmentation, Karimi et al. (2018) exhibited a segmentation algorithm by combining SSM with CNN. In their study, a CNN predicted the model posture and shape control vector at the same time, to generated the prostate boundary from the employed SSM. However, this method was highly dependent on the shape representative ability of SSM. It might be powerless when the target is beyond the SSM deformation range. In this paper, based on a registration-based approach, we proposed a weakly supervised segmentation framework to tackle the prostate extraction problem, via a boundary predictor (SSM-Net) and a label classifier (ResU-Net). In this framework, SSM-Net, a GoogLeNet-based network was firstly involved to predict the prostate contour to obtain a boundary distance map, serving as the target contour constraints. Then, ResU-Net, a 3D residual U-net was employed to predict a probability label map from the input images, to judge the class possibility of each pixel. In the inference step, the average of the distance map and the probability map was regarded as the final prostate segmentation. In our validation experiment, six different SSMs with various nodes were built to investigate the accuracy and efficiency of the whole framework.

The structure of the remainder of the paper is organized as follows: Section 2 illustrates the methodology of the proposed algorithm, and Section 3 informs the verification experiment and the corresponding results. Discussion and conclusions are presented in the Section 4.

2. Methodology

In the proposed registration-based prostate segmentation framework, a boundary predictor SSM-Net is firstly exploited to yield the prostate contour. According to the generated boundary, a normalized distance map is calculated to work as the prostate contour constraints. Meanwhile, a label classifier ResU-Net is employed to predict a probability label map from the input images, instead of directly using its binarized result as the target region. Finally, the average of the distance map and the probability map is regarded as the final prostate segmentation.

Fig. 2 is an overview of the proposed prostate segmentation framework. As shown by Fig. 2A and B, SSM-Net and ResU-Net are trained separately with different loss functions. SSM-Net is used to deduce, from the input, three variables: the SSM global transform, shape control parameters and the point-wise fine-tuning vector. In its
training phase, a patient-specific deformation field is calculated based on the predicted variables, and a surface for SSM is further generated via a coordinate sampler. The parameters of SSM-Net can be optimized by maximizing the similarity between the generated SSM surface and the standard SSM surface, as shown in Fig. 2A. While in the inference phase, according to the generated contour obtained by driving SSM with the predicted deformation field, a distance map is directly calculated to serve as the target boundary constraints. As regards ResU-Net, it works in the same way during training and inference, deducing the probability label map directly from the input image volume.

The remainder of this section is structured as follows: Section 2.1 explains the SSM computation and its application in SSM-Net. Section 2.2 illustrates the structure of the two networks involved in this framework. Refereed image mapping and loss functions are informed in Section 2.3 and Section 2.4 respectively. And the prostate inference is shown in Section 2.5.

2.1. Statistical shape model

2.1.1. Building the SSM

Due to the capacity to carry prior geometric information of numerous examples in different medical modality, SSM has been widely applied in object recognition, image process, surgery implant design (Lüthi et al., 2012). Generally, it involves two parts to describe the statistical spatial information of a collection of objects: a geometric model for the representation of the mean shape, and a series of variation vectors to depict the principal components of divergences between the objects and the mean shape. PCA-based shape model as the most prevalent SSM type, can model the variability of various type of objects such as images, displacement fields, surface meshes and volumetric meshes. In our work, we built PCA SSMs based on triangulated surface meshes, to represent the prostate surface on patient image.

According to the theory of PCA-based model, an arbitrary shape can be represented by superimposing a deformation field to the mean shape. In our work, a deformation field is described as the sum of global transform, weighted variations and the mean shape. That means, according to the notations defined in Table 1, shape can be written as follows:

\[ u' = V(\psi_{3N \times M}, \kappa_M, \theta_M, t_3, \bar{u}_{3N}) = \bar{u}_{3N} + \psi_{3N \times M} \cdot \text{diag}(\kappa_M) \cdot \theta_M \cdot R + T \]  

(1)

2.1.2. Enlarging the flexibility of SSM

The deformation ability of SSM depends on its node number \( N \) and the variation matrix \( \psi_{3N \times M} \). As a SSM can only contain a limited number of nodes, it might be insufficiently expressive to represent all possible shapes. There are two approaches to enlarge the SSM flexibility: first, we can augment the example data by adding white noise to make the variation matrix \( \psi_{3N \times M} \) more representative; second, it is more convenient to directly supple-

---

Table 1: Mathematic notations of statistical shape model

| A. Variables: scalars, vectors, matrices |
|------------------------------------------|
| \( N \) | Number of SSM nodes |
| \( M \) | Number of SSM variations |
| \( \bar{u}_{3N \times 1}(\bar{u}_{3N}) \) | SSM mean shape |
| \( \psi_{3N \times M} \) | SSM variation / PCA basis |
| \( \kappa_{M \times 1}(\epsilon_M) \) | Variance of SSM variations |
| \( \theta_{M \times 1}(\theta_M) \) | SSM shape control parameters |
| \( t_3 : T, R \) | Transform parameters: translation and rotation |
| \( \text{Id}_{3 \times 3}(I) \) | Identity matrix |

| B. Operators and functions |
|---------------------------|
| \( \text{diag}(\vec{v}) \) | Diagonal matrix of vector \( \vec{v} \) |
| \( V(\psi_{3N \times M}, \kappa_M, \theta_M, \bar{u}_{3N}) \) | Model surface deformation |
mented a fine-free tuning item $\xi_{3N}$ to the deformed shape. The feasibility and efficiency of the two augmentation approaches are illustrated in the section of Experiments and results.

$$u' = V(\psi_{3N \times M}, \kappa_M, \theta_M, t_3, \bar{u}_{3N}, \xi_{3N})$$

$$= \bar{u}_{3N} + \psi_{3N \times M} \cdot \text{diag}(\kappa_M) \cdot \theta_M \cdot R + T + \xi_{3N} \quad (2)$$

### 2.1.3. Model space to image space

Surface model is commonly defined in physical spatial coordinate system to maintain the visualization invariance on different platforms. The transformation between model space and image space in this work is shown as follows:

$$P_i = [(P_m - P_0)/s + 0.5]$$

where $s$ is the image resolution, and $P_0$ is the position of image origin in the physical spatial coordinate system. $P_m$ indicates the coordinate of model node $m$ in spatial space and $P_i$ is the corresponding coordinate of $P_m$ in image space.

### 2.2. CNN architecture of the segmentation framework

This subsection informs the architecture of the networks involved in our segmentation framework, including the SSM-Net and the ResU-Net. Fig. 3A depicted the structure of SSM-Net. This distance map prediction network was not limited by a specific size, but the input image in our experiments was size of $(122, 120, 70, 1)$. We firstly adopted a GoogLeNet for the extraction of the prostate global transform. GoogLeNet (Szegedy et al., 2015) is a convolutional neural network originally designed for the ILSVRC (ImageNet Large Scale Visual Recognition Competition). Its adopted inception blocks are constituted by several convolutional filters in various sizes, facilitating its capability to explore the image details at different extents. On the basis of the original GoogLeNet, a dense layer with size of 128 is connected to its flatten layer, for the prediction of the global transform $(D_x, D_y, D_z, R_x, R_y, R_z)$, as shown by the left column of Fig. 3A. Then, two CNN with residuals are utilized to predict the shape control parameters or the fine-tuning item. The right column of Fig. 3A shows the network structure for the prediction of shape control parameters with size of $M \times 1$ ($M$ is the number of shape variations). The involved convolution layers with kernel size 3 3 3 used 1 pixel stride, and the employed max pooling layers used 2 pooling size. Similarly, the prediction of fine-tuning vector shared the same residual CNN structure, except the average pooling layer is replaced by a max pooling layer with pooling size of 2, stride of 2. In addition, the output including the transform, the shape control parameters and the fine-tuning vector, are input to the last spatial transformation layer of SSM-Net, to yield the prostate contour. For whole network, the shape of each layer is annotated as shown in Fig. 3A.

U-net (Ronneberger et al., 2015) is a widely used network with high accuracy for object segmentation. In our work, a residual U-Net is employed to infer the probability label map. As illustrated in Fig. 3B, each residual block consists two convolution layers with kernel size of 3 3 3, pixel stride of 1. The max pool layers use a pooling size equal to the stride and the size of the up-sample layers is set to 2. The sigmoid function is utilized as the activation function of the last layer to limit output values to [0,1].

### 2.3. Grid mapping

As shown in the overview of the segmentation framework (Fig. 2), the patient-specific prostate shape can be obtained by superimposing the predicted deformation field to the standard SSM surface in the inference procedure. While in the training phase, in order to calculate the value of loss function, a predicted binary surface image for SSM (annotated as generated model surface in Fig. 2) is generated by interpolating the input boundary based on the deformation field.

The output binary surface of SSM $g^t \in \mathbb{R}^{L' \times W' \times H'}$ is defined on a regular grid $G^t = \{G^t_i\} = \{(x_i^t, y_i^t, z_i^t)\}$, $i \in [L'W'\times H']$, where $L', W', H'$ represent the length, width and height of the output. Similarly, let $G_n = \{(x_i^n, y_i^n, z_i^n)\}$, $x_i^n \in [L\times L']$, $y_i^n \in [0, W']$, $z_i^n \in [0, H']$ be the input grid, where $L', W', H'$ are the length, width and height of the input binary mask respectively. The relationship between the output grid $G^t$ and the input grid $G^n$ can be written as follows. For $i \in [1...L'W'H']$,

$$G_i^n = D_i(G_i^t) = \begin{pmatrix} x_i^t + d_{i,x} \\ y_i^t + d_{i,y} \\ z_i^t + d_{i,z} \end{pmatrix} + \begin{pmatrix} x_i^n \\ y_i^n \\ z_i^n \end{pmatrix} \quad (4)$$

$D_i(G^t)$ is the deformation field predicted by SSM-Net, with size of $L' \times W' \times H' \times 3$. Based on Eq. 4 for
The partial derivatives with respect to gray value of the point $G_i^t$ and coordinate position $(x_i^t, y_i^t, z_i^t)$ for the backpropagation of loss can be written as follows ($\partial g_i^t / \partial y_i^t$, $\partial g_i^t / \partial z_i^t$ are similar with $\partial g_i^t / \partial x_i^t$):

$$
\frac{\partial g_i^t}{\partial g_{(n,m,p)}^t} = \sum_{n} \sum_{m} \sum_{p} g_{(n,m,p)}^t \cdot \max(0, 1 - |x_i^t - n|) \cdot \max(0, 1 - |y_i^t - m|) \cdot \max(0, 1 - |z_i^t - p|)
$$

$$
\frac{\partial g_i^t}{\partial x_i^t} = \sum_{n} \sum_{m} \sum_{p} g_{(n,m,p)}^t \cdot \max(0, 1 - |x_i^t - n|) \cdot \max(0, 1 - |z_i^t - p|)
$$

$$
\frac{\partial g_i^t}{\partial y_i^t} = \sum_{n} \sum_{m} \sum_{p} g_{(n,m,p)}^t \cdot \max(0, 1 - |y_i^t - m|)
$$

$$
\frac{\partial g_i^t}{\partial z_i^t} = \sum_{n} \sum_{m} \sum_{p} g_{(n,m,p)}^t \cdot \max(0, 1 - |z_i^t - p|)
$$

2.4.2. Loss function of SSM-Net

The following loss function is adopted by SSM-Net, where $\hat{\phi}$ is the network parameters and $P_{\text{ResU}}$ is the output probability map. The values of range from 0 to 1.

$$
L(\theta, S_{\text{mask}}) = 1 - 2 \times \frac{\|P_{\text{ResU}}(\hat{\theta}) \times S_{\text{mask}}\|}{\|P_{\text{ResU}}\| + \|S_{\text{mask}}\|}
$$

Thus, the parameterized SSM-Net can be optimized during the training procedure:

$$
\hat{\theta} = \arg \min \{ L(\theta, S_{\text{mask}}) \}.
$$
2.5. Inference of prostate region

In the prostate inference step, the binarized average of the deduced probability map and distance map is regarded as the final prostate segmentation. In term of probability label map, a bigger value means higher probability for a pixel to belong to the prostate region. The distance map is obtained according to the predicted boundary image by SSM-Net, in which only the pixel on the boundary has a value of 1, while others equal to 0. It is calculated according to the following equation:

\[ D_{ssm}(P_i) = 1 - \| P_i - \hat{P}_i \| / 10 \]  

(13)

Where, \( P_i \) represents pixel \( i \), \( D_{ssm}(P_i) \) is the value of \( P_i \) on the distance map. \( \hat{P}_i \) is the closest boundary point of \( P_i \) on the input boundary image, and \( \| P_i - \hat{P}_i \| \) is the Euclidean distance between \( P_i \) and \( \hat{P}_i \). 10 is the calculation range, which should be changed according to the image resolution because it decides the prostate voxel range in images. As the image volume were resampled to the same resolution, the calculation range is a constant in our work.

3. Experiment and results

3.1. Data acquisition and experiment set up

We validated the proposed SSM-Net on two public datasets, including MICCAI PROMISE12 challenge dataset [Litjens et al., 2014] and NCI-ISBI 2013 challenge dataset [Bloch et al., 2015]. The first database contains 50 prostate transversal T2 MRI for training and 30 prostate images for testing. And the second database respectively involves 60, 10 and 10 cases for training, leaderboard and testing. The two datasets share 11 common volumes, and the ground trues of PROMISE12 testing set are unavailable. Thus, totally 119 T2 MRI image cases are collected for our experiment. As the gathered data have different voxel spacing and image size, a cubic interpolation was employed firstly to uniform the image resolution to 1.00 mm \times 1.00 mm \times 1.00 mm. Then the processed images are further cropped to a homogeneous size of 122 120 70 for SSM-Net. While the input of ResU-Net was the simply resized image with shape of 176 128 48.

From the 119 volumes, 40 randomly selected cases contributed to the SSM establishment, and the rest data were used for the network optimization. Specifically, 63 out of the 79 image volumes (3/4) were applied for the network training and the remaining (1/4) for the testing. The modeling procedure and the network segmentation performance are illustrated in Section 3.2 and Section 3.3 respectively. And Section 3.4 analyzes and discusses the influence of SSM flexibility on the framework segmentation accuracy.

3.2. SSM establishment and analysis

The node number and the principle component number are two dominant factors for the flexibility of SSM. The former is decided according to the node number of the counted meshes, and the latter is determined by the compactness of SSM. The compactness is measured via the accumulation of SSM variations which are arranged according to their eigenvalues. In this work, we selected the first M principle components to keep 95% of the total eigenvalues. To investigate the influence of the two factors on the segmentation accuracy of the segmentation method, six SSMs with various nodes and variations were built according to the following establishment procedure:

1) Construct and refine 3D surface triangle mesh based on the provided segmentation;
2) Subdivide and decimate the meshes to specific number of points (N);
3) Transform the meshes to same posture and make them in correspondence;
4) Establish PCA-based SSMs based on the corresponded objects;
5) Augment the SSM flexibility by employed a Gaussian process.

Where, 3D slicer (https://www.slicer.org) was utilized for the refinement of 3D surface. In addition, the Gaussian process model building, model fitting, and the PCA-based model establishment are completed via Statismo library [Lüthi et al., 2012]. The details of SSMs are exhibited in Table 2. The second column shows the models after flexibility augmentation.

Table 2: The details of statistical shape models with various number of nodes

| Number of node | Original SSMs | Augmented SSMs |
|----------------|---------------|----------------|
|                | Variation(3N x M) | Variation(3N x M) |
| 1625           | 4875 49       | 4875 50        |
| 3250           | 9750 49       | 9750 50        |
| 6500           | 19500 49      | 19500 50       |
| 9750           | 29250 49      | 29250 50       |
| 13000          | 39000 48      | 39000 50       |
| 16250          | 48750 47      | 48750 50       |

![Figure 5: The flexibility of the statistical shape models with different node numbers.](image)

Figure 5 illustrates the flexibilities of the primary SSMs. The green models in the middle column depicts the mean models, and the left three columns and the right three columns respectively show the deformed models drove by 3\(\lambda\) and 3\(\sqrt{\lambda}\) times of the first three principal components of variations. \(\lambda\) is the corresponding deviation of each component.

The middle column with green color exhibits the mean models, and the left three columns and the right three columns respectively show the deformed models driven by 3\(\sqrt{\lambda}\) times of the first three principal components of variations. The green models in the middle column depict the mean shapes with increasing node numbers from top to bottom. From the middle to the rightmost, the three columns respectively exhibit the deformed shapes generated from the mean shape of SSM by weighting the first three principle variations with three times of their corresponding deviations. Accordingly, the three-column shapes on the left describe the generated models deformed by negative
triple deviations. Generally, each row horizontally reveals the influence of the first three principal variations on the deformation of each SSM, and each column vertically compares the different performances of SSMs. In conclusion, the last three SSMs with node numbers of 9750, 13000, 15250 had similar interior and exterior characteristics, while the first three SSMs behaved quite differently.

3.3. Accuracy evaluation and analysis

Two metrics were introduced to evaluate the performance of the proposed segmentation framework, including the dice similarity coefficient (DSC) and the average over the shortest distance between the boundary points of the volumes(ABD). The DSC is formulated as follows:

\[
\text{Dice} = 2 \times \frac{\| S_{\text{mask}} \|}{\| S_{\text{mask}} \| + \| S_{\text{pred}} \|} \quad (14)
\]

where \( S_{\text{pred}} \) and \( S_{\text{mask}} \) respectively present the predicted segmentation and the input ground truth. In this work, we firstly compared the performance of the proposed segmentation framework applying different values of regularization coefficients \( \lambda_1, \lambda_2, \lambda_3 \) in Eq. 4. The experiments showed that the best result in terms of Dice coefficient is achieved when \( \lambda_1, \lambda_2, \lambda_3 \) are equal to 0.01 and 0.01. In the following work, \( \lambda_1, \lambda_2, \lambda_3 \) are set to 0, 0.1, and 0.01. According to the record data, we concluded that the SSM with 9750 nodes maybe be capable of accurately delineation when the target is beyond the SSM deformation range.

![Figure 6: The DSC and ABD results of SSM-Net, ResU-Net and the whole framework.](Image)

The segmentation procedure and the delineation performance of the proposed framework are shown in Fig. 7. Subgraph A exhibits the whole operation process. The white line in (a) indicates our segmentation target, which is manually delineated in advance for visualization and testing. The purple model in (b) represents the SSM, whose center initially positioned at the origin (0,0,0). As shown in (b) and (c), according to the input image, SSM-Net respectively predicts the global transform, weight parameters and an offset vector for the calculation of deformation field. The 3d white surface in (d) represents the generated prostate boundary by applying the deformation field to the SSM. Subgraph B illustrates the delineation results of four images series. Four slices range from number 16 to number 46 with an interval of 10, are selected to display the recognition performances on different prostate zones. Each row stands one example, and accordingly each column shows the delineation results of the same layers of different examples. From the exhibition, we deduced that the segmentation on the prostate central zone has a higher accuracy, while the base of the prostate is more complex to delineate. For the case of severe hyperplastic prostate gland which is times larger than the mean shape of SSM, the segmentation framework has relative poor performance with the maximum ABD was 2.7mm and the dice coefficient was 0.83. According to the analysis of the output boundary distance map and the label probability map, the segmentation results basically depended on the output of ResU-Net while SSM-Net became less powerful for the boundary constraints prediction. Thus, compare to the segmentation approach only employed SSM method, our framework can achieve more accurate delineation when the target is beyond the SSM deformation range.

3.4. The influence of network flexibility on segmentation accuracy

According to the analysis in Section 2.2 network node number, flexibility augmentation of SSM and fine-tuning item (offset) are the three dominant factors to affect the network deformation ability. To investigate their impacts on the network delineation accuracy, we verified the segmentation performance of SSM-Net under different combinations of the three elastic determinants. Fig. 8 informs the DSC and ABD values of SSM-Net with various nodes and different utilization situation of augmentation and offset item. As shown by those statistical trend lines, regardless of the model augmentation or fine-tuning item, SSM-Net with 9750 nodes or more are outperformed other situations. In term of the model flexibility augmentation, its application has evidently improved the network accuracy. As shown by (A) and (B), compared with the results of the network without augmentation and offset, the best dice and distance of the network adopted model elastic augmentation improved to 0.81 and 2.36 mm respectively. Similarly, when the network employed both augmentation and offset has higher dice and superior distance than the network only utilized offset item, as informed by (C) and (D). In addition, we figure out that the employment of offset item also contributed to the improvement of the SSM-Net. Compared with the result (A) of the network without
Figure 7: The segmentation procedure and performance of SSM-Net. Subgraph A exhibits the whole operation process. White line in (a) indicates our segmentation target for visualization. The purple model in (b) represents the SSM, whose center initially positioned at the origin (0,0,0). And the white model in (d) shows the surface predicted by SSM-Net. Subgraph B illustrates the delineation results of four images. Each row stands one example, and each column shows their segmentation results of the same layers of different examples.
Figure 8: The influence of the deformation ability of SSM-Net on its segmentation performance. The network elastic ability is mainly determined by three factors: node numbers, model flexibility augmentation and the offset item.

4. Discussions and conclusion

Prostate segmentation facilitates the diagnosis and treatment of prostate diseases. For example, the determination and location of prostate region are essential information for radiotherapy or high-intensity focused ultrasound operation. However, its clinical application is still limited, because of the segmentation challenges like inhomogeneous intensity, various anatomical appearance and indiscernible boundary. Therefore, in this work we proposed a weakly supervised prostate segmentation framework, on the basis of convolutional neural network and SSM, which has been widely used in prostate segmentation (Heimann and Meinzer, 2009). In this framework, we firstly established a series of SSMs according to two public prostate MRI datasets. Then a GoogLeNet-based network was designed to predict the translation vector of model center, the weight parameters of shape variations and the point-wise fine-tuning vector, to obtain the prostate boundary from SSM. Next, a residual U-Net was utilized for the prostate probability class map. The final segmentation was decided by the distance map and the probability map together. To investigate the feasibility and capacity of this framework, we validated the segmentation accuracy by applying various models with different number of nodes. In addition, experiments to survey the influence of model augmentation and fine-tuning item on the segmentation performance were conducted respectively. The results demonstrate that, the network employed SSM with 9750 nodes has the optimal dice of 0.904 and DSC of 1.88 mm. And both model elastic augmentation and offset application have positive effects. The performance on the collected clinical data demonstrates that our prostate segmentation framework is feasible, and it might be a useful clinical tool for the diagnosis, treatment design and therapeutic procedure of variable prostate disease.

Table 3 summaries the related studies from literature and records their qualitative measure results on the promise12 testing datasets. In the past three decades, three major categories of automatic prostate segmentation have been introduced, including atlas-based algorithms, deformable model-based approaches and machine learning-based methods. For the firstly category, atlas-based strategy has been widely utilized in medical image segmentation and registration (Ou et al., 2012; Gao et al., 2012; Chilali et al., 2016). Its major principle is to align an atlas which contained spatial prior knowledge to the target images by registration approaches. Then apply the alignment information to deform the atlas label for the final target segmentation. Secondly, in term of model-based method, a deformable model is firstly constructed for the representation of standard prostate contour. Then the information extracted from the target image was further applied to drive the model to generate the specific shape. Several groups in the list (Maan and van der Heijden, 2012; Vincent et al., 2012; Kirschner et al., 2012; Karimi et al., 2018) employed deformable model (AAM, ASM, SSM) for the prediction of prostate. For the third group, learning-based approaches, especially convolutional neural networks are widely introduced in medical image processing because of its extraction capability of more latent features (Milletari et al., 2016; Yu...
Several teams have combined machine learning algorithms with atlas or deformable models in their researches. Similarly, our segmentation combined neural network with SSM for high precise MRI prostate segmentation. In this way, prior knowledge is collected to serve as a reference or boundary constraints, meantime more comprehensive information from the target image can be obtained by using neural network to recognize of target region more precise. Besides, as SSM is built based on lots of medical image data, the generated model deformed form it is reasonable, which means the surface smooth procedure could be eliminated. Therefore, our method required less time (approximately 3 s) while performed a satisfactory segmentation accuracy with higher robustness.

In medical image processing, the collection of the training dataset limits the application of various learning based approaches. Fortunately, SSM utilizes the geometric features rather than the intensity information. Therefore, all images in different modalities such as computed tomography, ultrasound and MRI can contribute to the construction of SSM. It is worth mentioning that, if the dataset is sufficient, active shape model (ASM) and statistical deformation models (SDM) could become superior training supervisors than SSM, as the former can carry intensity information from the target image can be obtained by using neural network to recognize of target region more precise. Therefore, our method required less time (approximately 3 s) while performed a satisfactory segmentation accuracy with higher robustness.

In medical image processing, the collection of the training dataset limits the application of various learning based approaches. Fortunately, SSM utilizes the geometric features rather than the intensity information. Therefore, all images in different modalities such as computed tomography, ultrasound and MRI can contribute to the construction of SSM. It is worth mentioning that, if the dataset is sufficient, active shape model (ASM) and statistical deformation models (SDM) could become superior training supervisors than SSM, as the former can carry intensity information from the target image can be obtained by using neural network to recognize of target region more precise. Therefore, our method required less time (approximately 3 s) while performed a satisfactory segmentation accuracy with higher robustness.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

This work was supported by grants from the PHC CAI YUANPEI Program (413606SA), Agence Nationale de la Recherche (Investissements d’Avenir Program grants ANR-11-LABX-0004 and ANR-19-P3IA-0003), Rgion Auvergne Rhne-Alpes (ProNaVIA project), National Key R&D Program of China (2017YFB1302900), National Natural Science Foundation of China (81971709; 81828003; M-0019; 8191101624), the Foundation of Ministry of Education of China and Science and Technology Development Center (2018C01038), the Foundation of Science and Technology Commission of Shanghai Municipality (19510712200), and Shanghai Jiao Tong University Foundation on Medical and Technological Joint Science Research (ZH20182DA15;YG2019ZDA06;ZH2018QNA23).

### References

Alb, X., Pereaez, M., Hoogendoorn, C., Swift, A.J., Wild, J.M., Frangi, A.F., Lekadir, K., 2016. An algorithm for the segmentation of highly abnormal hearts using a generic statistical shape model. IEEE Transactions on Medical Imaging 35, 845–859.

Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2019. Voxelmorph: A learning framework for deformable medical image registration. IEEE Transactions on Medical Imaging 38, 1788–1800.

Bloch, N., Madabhushi, A., Huisman, H., Freymann, J., Kirby, J., Grauer, M., Enqoubahirce, A., Jaffe, C., Clarke, L., Farahani, K., 2015. Ni-cibi 2013 challenge: automated segmentation of prostate structures. The Cancer Imaging Archive 370.

Chilali, O., Puech, P., Lakroum, S., Df, M., Mordon, S., Betrouni, N., 2016. Gland and zonal segmentation of prostate on t2w mr images. Journal of digital imaging 29, 730–736.

Fütterer, J.J., Briganti, A., De Visschere, P., Emberton, M., Giannarini, G., Kirkham, A., Taneja, S.S., Theony, H., Villers, G., Villers, A., 2015. Can clinically significant prostate cancer be detected with multiparametric magnetic resonance imaging? a systematic review of the literature. European urology 68, 1045–1053.

Gao, Q., Rueckert, D., Edwards, P., 2012. An automatic multi-atlas based prostate segmentation using local appearance-specific atlases and patch-based voxel weighting. MICCAI Grand Challenge: Prostate MR Image Segmentation.

Ghose, S., Mitra, J., Oliver, A., Marti, R., Lladó, X., Freixenet, J., Vilanova, J.C., Sidibé, D., Meriaudeau, F., 2012. A random forest based classification approach to prostate segmentation in mri. MICCAI Grand Challenge: Prostate MR Image Segmentation 2012, 125–128.
Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets, in: Advances in neural information processing systems, pp. 2672–2680.

Guo, Y., Gao, Y., Shen, D., 2015. Deformable nr prostate segmentation via deep feature learning and sparse patch matching. IEEE transactions on medical imaging 35, 1077–1089.

He, B., Xiao, D., Hu, Q., Jia, F., 2017. Automatic magnetic resonance image prostate segmentation based on adaptive feature learning probability boosting tree initialization and cnn-asn refinement. IEEE Access 6, 2005–2015.

Heimann, T., Meinzer, H.P., 2009. Statistical shape models for 3d medical image segmentation: a review. Medical image analysis 13, 543–563.

Jaderberg, M., Simonyan, K., Zisserman, A., et al., 2015. Spatial transformer networks, in: Advances in neural information processing systems, pp. 2017–2025.

Jia, H., Xia, Y., Cai, W., Fulham, M., Feng, D.D., 2017. Prostate segmentation in nr images using ensemble deep convolutional neural networks, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), pp. 762–765.

Jia, H., Xia, Y., Song, Y., Cai, W., Fulham, M., Feng, D.D.; 2018. Atlas registration and ensemble deep convolutional neural network-based prostate segmentation using magnetic resonance imaging. Neurocomputing 275, 1358–1369.

Karimi, D., Samei, G., Kesch, C., Nir, G., Salcudean, S.E., 2018. Prostate segmentation in mri using a convolutional neural network architecture and training strategy based on statistical shape models. International journal of computer assisted radiology and surgery 13, 1211–1219.

Khallaghi, S., Sánchez, C.A., Rasoulian, A., Nouranian, S., Romagnoli, C., Abdi, H., Chang, S.D., Black, P.C., Goldenberg, L., Morris, W.J., et al., 2015. Statistical biomechanical surface registration: application to mr-trus fusion for prostate interventions. IEEE transactions on medical imaging 34, 2535–2549.

Kirschner, M., Jung, F., Wesarg, S., 2012. Automatic prostate segmentation in nr images with a probabilistic active shape model. MICCAI Grand Challenge: Prostate MR Image Segmentation 2012.

Korsager, A.S., Fortunati, V., van der Lijn, F., Carl, J., Niessen, W., Østergaard, L.R., van Walsum, T., 2015. The use of atlas registration and graph cuts for prostate segmentation in magnetic resonance images. Medical physics 42, 1614–1624.

Litjens, G., Debats, O., Barentsz, J., Karssemeijer, N., Huisman, H., 2014a. Computer-aided detection of prostate cancer in mri. IEEE transactions on medical imaging 33, 1083–1092.

Litjens, G., Toth, R., van de Ven, W., Hoks, C., Kerkstra, S., van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al., 2014b. Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. Medical image analysis 18, 359–373.

Lüthi, M., Blanc, R., Albrecht, T., Gass, T., Goksel, O., Büchler, P., Kistler, M., Bousleiman, H., Reyes, M., Cattin, P., et al., 2012. Statismo-a framework for pca based statistical models. The Insight Journal 2012, 1–18.

Maan, B., van der Heijden, F., 2012. Prostate nr image segmentation using 3d active appearance models. MICCAI Grand Challenge: Prostate MR Image Segmentation 2012.

Mahapatra, D., Buhmann, J.M., 2015. Visual saliency based active learning for prostate mri segmentation, in: International Workshop on Machine Learning in Medical Imaging, pp. 9–16.

Martin, S., Daanen, V., Troccaz, J., 2008. Atlas-based prostate segmentation using an hybrid registration. International Journal of Computer Assisted Radiology and Surgery 3, 485–492.

Martin, S., Troccaz, J., Daanen, V., 2010. Automated segmentation of the prostate in 3d nr images using a probabilistic atlas and a spatially constrained deformable model. Medical physics 37, 1579–1590.

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 fourth international conference on 3D vision (3DV), pp. 565–571.

Mun, J., Jang, W.D., Sung, D.J., Kim, C.S., 2017. Comparison of objective functions in cnn-based prostate magnetic resonance image segmentation, in: 2017 IEEE International Conference on Image Processing (ICIP), pp. 3859–3863.

Ou, Y., Doshi, J., Erus, G., Davatzikos, C., 2012. Multi-atlas segmentation of the prostate: A zooming process with robust registration and atlas selection. MICCAI Grand Challenge: Prostate MR Image Segmentation 2012.

Pereira, S., Pinto, A., Alves, V., Silva, C.A., 2016. Brain tumor segmentation using convolutional neural networks in mri images. IEEE transactions on medical imaging 35, 1240–1251.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks, in: Advances in neural information processing systems, pp. 91–99.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241.

Seim, H., Kainmueller, D., Heller, M., Lamecker, H., Zachow, S., Hege, H.C., 2008. Automatic segmentation of the pelvic bones from ct data based on a statistical shape model. VCBM 8, 93–100.

Shen, D., Herskovits, E.H., Davatzikos, C., 2001. An adaptive-focus statistical shape model for segmentation and shape modeling of 3-d brain structures. IEEE transactions on medical imaging 20, 257–270.
Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9.

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J., 2016. Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging 35, 1299–1312.

Tian, Z., Liu, L., Zhang, Z., Fei, B., 2015. Superpixel-based segmentation for 3d prostate mr images. IEEE transactions on medical imaging 35, 791–801.

Tian, Z., Liu, L., Zhang, Z., Fei, B., 2018. Panet: prostate segmentation on mri based on a convolutional neural network. Journal of Medical Imaging 5, 021208.

Toth, R., Madabhushi, A., 2012. Deformable landmark-free active appearance models: application to segmentation of multi-institutional prostate mri data. MICCAI Grand Challenge: Prostate MR Image Segmentation 2012.

Vincent, G., Guillard, G., Bowes, M., 2012. Fully automatic segmentation of the prostate using active appearance models. MICCAI Grand Challenge: Prostate MR Image Segmentation 2012, 2.

Wang, B., Lei, Y., Tian, S., Wang, T., Liu, Y., Patel, P., Jani, A.B., Mao, H., Curran, W.J., Liu, T., et al., 2019. Deeply supervised 3d fully convolutional networks with group dilated convolution for automatic mri prostate segmentation. Medical physics 46, 1707–1718.

Yu, L., Yang, X., Chen, H., Qiu, J., Heng, P.A., 2017. Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images, in: Thirty-first AAAI conference on artificial intelligence.

Zhang, X., Tian*, J., Deng, K., Wu, Y., Li, X., 2010. Automatic liver segmentation using a statistical shape model with optimal surface detection. IEEE Transactions on Biomedical Engineering 57, 2622–2626.