



Polar transform network for prostate ultrasound segmentation with uncertainty estimation

Xuanang Xu^a, Thomas Sanford^b, Baris Turkbey^c, Sheng Xu^d, Bradford J. Wood^d,
Pingkun Yan^{a,*}

^a Department of Biomedical Engineering and the Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

^b Department of Urology, The State University of New York Upstate Medical University, Syracuse, NY 13210, USA

^c Molecular Imaging Program, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

^d Center for Interventional Oncology, Radiology & Imaging Sciences at National Institutes of Health, Bethesda, MD 20892, USA

ARTICLE INFO

Article history:

Received 24 August 2021

Revised 8 December 2021

Accepted 7 March 2022

Available online 17 March 2022

Keywords:

Prostate segmentation

Ultrasound image

Polar transform

Fully convolutional network

Uncertainty estimation

ABSTRACT

Automatic and accurate prostate ultrasound segmentation is a long-standing and challenging problem due to the severe noise and ambiguous/missing prostate boundaries. In this work, we propose a novel polar transform network (PTN) to handle this problem from a fundamentally new perspective, where the prostate is represented and segmented in the polar coordinate space rather than the original image grid space. This new representation gives a prostate volume, especially the most challenging apex and base sub-areas, much denser samples than the background and thus facilitate the learning of discriminative features for accurate prostate segmentation. Moreover, in the polar representation, the prostate surface can be efficiently parameterized using a 2D surface radius map with respect to a centroid coordinate, which allows the proposed PTN to obtain superior accuracy compared with its counterparts using convolutional neural networks while having significantly fewer (18%~41%) trainable parameters. We also equip our PTN with a novel strategy of centroid perturbed test-time augmentation (CPTTA), which is designed to further improve the segmentation accuracy and quantitatively assess the model uncertainty at the same time. The uncertainty estimation function provides valuable feedback to clinicians when manual modifications or approvals are required for the segmentation, substantially improving the clinical significance of our work. We conduct a three-fold cross validation on a clinical dataset consisting of 315 transrectal ultrasound (TRUS) images to comprehensively evaluate the performance of the proposed method. The experimental results show that our proposed PTN with CPTTA outperforms the state-of-the-art methods with statistical significance on most of the metrics while exhibiting a much smaller model size. Source code of the proposed PTN is released at <https://github.com/DIAL-RPI/PTN>.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

Prostate segmentation in transrectal ultrasound (TRUS) images is a fundamental prerequisite for many prostate cancer-related clinical procedures, such as image-guided biopsy, brachytherapy needle placement, and interventional surgery planning. Accurate delineation of the prostate boundary in the TRUS images benefits the final efficacy of cancer diagnosis and treatment. However, manual tracing of the prostate volume not only consumes tremendous of time and labor, but also relies on professional expertise and experience. The resulting contour may also suffer from significant inter- and intra-observer variabilities, which seriously affect

the reproducibility of treatment planning. Therefore, computer-aided prostate segmentation in TRUS images has been highly desired by the clinicians. On the other hand, the TRUS images often exhibit low signal-to-noise ratio and inhomogeneous intensities, which severely hinder the discrimination of the prostate from the background. Shadow artifacts caused by the ultrasonic absorption and reflection even occlude some parts of an image, leading to the missing boundaries. The large variability of the prostate shape and size across different patients further complicates the segmentation.

Currently, the state-of-the-art methods for prostate ultrasound segmentation are all built upon deep learning (DL), more specifically, the convolutional neural networks (CNNs). Benefiting from the powerful representation ability of the hierarchical features and the data-driven end-to-end learning manner, the DL-based methods have substantially extended the frontier of the prostate ultra-

* Corresponding author.

E-mail address: yanp2@rpi.edu (P. Yan).

sound segmentation. The Dice similarity coefficient (DSC) of the segmented whole prostate is generally over 90% (Yang et al., 2017; Anas et al., 2018; Wang et al., 2018; 2019b; Lei et al., 2019b; 2019a; Orlando et al., 2020a; 2020b; Geng et al., 2020). However, three significant challenges still exist in the current DL-based methods for prostate ultrasound segmentation.

1. The previous works mainly focused on the segmentation of the main gland or the whole prostate but pay less attention to the apex and base sub-areas. According to our observation, the majority of the segmentation error comes from the prostate apex and base, where the boundary suffers from even worse contrast than that in the main gland region. Improving the performance in the prostate apex and base sub-area could be an effective way to further push the limits of the DL-based prostate ultrasound segmentation.
2. Despite the advances in segmentation accuracy, DL segmentation models are growing larger with more complex architectures, which brings higher memory overhead, computation complexity, and risk of over-fitting. This is unfavorable for the modern DL paradigms (e.g., federated learning (McMahan et al., 2017) and swarm learning (Warnat-Herresthal et al., 2021)), where the DL models could be deployed on some low-end computation devices with limited training data.
3. Existing DL models tend to produce overconfident segmentation but lack a quantitative assessment on the reliability of these predictions. Since segmentation error inevitably appears in the area suffering from severe noise and low boundary contrast (e.g., the prostate apex and base sub-areas), the overconfident segmentation could mislead the subsequent clinical procedures. Therefore, a quantitative estimation of the model uncertainty is highly desired in the prostate ultrasound segmentation, which can guide clinicians when they are amending the auto-generated contours.

In this work, we propose a novel *polar transform network* (PTN) to address the above three issues in 3D prostate ultrasound segmentation. Instead of solving the prostate segmentation in the Cartesian coordinates as a pixel-wise classification problem, our core idea is to handle this task in the polar coordinates and convert it into a regression problem. To achieve this goal, we first use a localization sub-network to estimate the centroid of the prostate in the TRUS image. Then, the estimated centroid and the input image are fed to a polar transform layer, which is designed to conduct polar transformations on the image and feature map through a differentiable way. Last, we exploit a surface regression sub-network to regress the prostate surface radius, which is subsequently used to reconstruct the prostate volume.

As shown in Fig. 1e, through the polar transformation, both the apex and base areas as well as the central region around the prostate centroid can get denser samples than the other regions. The densely sampled voxels largely belong to the prostate and can thus effectively rebalance the foreground and background samples for network training. This helps prostate segmentation since it can naturally increase the weight of the prostate samples, i.e., the foreground tissue. Moreover, the denser samples around the apex and base regions further emphasize the weight of these two areas in the prostate segmentation. As a result, the model trained on the transformed polar image and feature maps learns to pay more attention to lower the segmentation error in these areas (addressing Issue #1). At the same time, the 3D shape of the prostate can be parameterized as a 2D surface radius map combined with a centroid coordinate in polar space, which is much more efficient than the conventional binary mask representation in image space. Benefiting from this property, we convert the prostate segmentation to a centroid localization task jointed with a surface radius regression task, which can be efficiently solved by a regression network (PTN)

with much fewer parameters than its FCN counterparts (addressing Issue #2). Furthermore, to indicate the segmentation errors from DL models, we propose a novel strategy of *centroid perturbed test-time augmentation* (CPTTA) to estimate the uncertainty map of the PTN segmentation. This mechanism can not only refine the PTN prediction but also provide crucial guidance to physicians when they modify or approve the auto-generated contours (addressing Issue #3).

It is worth noting that, although polar transform has been applied to medical image analysis before (Zahoor and Fraz, 2017; Fu et al., 2018), our method is fundamentally different from the previous works in three aspects. First, the previous works merely utilize the polar transformation as a pre-processing in their pipelines, while our PTN embeds the polar transformation into the network architecture. This innovation allows gradient back-propagation through polar transformation to achieve end-to-end training, which significantly improves the training and testing efficiency. Second, the previous methods segment the target organ as a 3D binary mask while our PTN parameterizes the segmentation result as a 2D surface radius map along with a centroid coordinate. This simplified representation significantly improves the parametric efficiency of our PTN and substantially reduces its model size. Third, the previous methods simply conduct the polar transformation using the image center as the polar origin, while our PTN can automatically detect the prostate centroid as the polar origin. This makes our method more robust in the clinical scenarios, where the target organ centroid could vary a lot across different cases.

The rest of this paper is organized as follows. Section 2 gives a brief review of the previous literature related to this work. Section 3 presents the details of the proposed method. In Section 4, we conduct extensive experiments on a clinical dataset to evaluate the performance of the proposed method and justify our designs. Some specific issues are discussed in Section 5. Finally, we conclude this work in Section 6.

2. Related works

Since our work involves prostate segmentation in ultrasound images and uncertainty estimation of DL models, we will first introduce the related works in these two research areas before we present our method.

2.1. Prostate segmentation in ultrasound images

At the early stage of prostate ultrasound segmentation, researchers focused on designing various of hand-crafted features to use with statistical shape models (Shen et al., 2003; Yan et al., 2010; 2011) or conventional machine learning (Ghose et al., 2013). The performance of these methods highly depends on the designed features, which tend to fail in the segmentation of the non-contrast prostate boundaries in the shadow regions.

In recent years, due to the expeditious development of deep learning (DL) in computer vision, convolutional neural network (CNN) has become the dominant methodology for image segmentation. The powerful representation ability makes the fully convolutional networks (FCNs) (Long et al., 2015) overwhelmingly outperform the conventional hand-crafted feature based methods. The current benchmark method for medical image segmentation is U-Net (Ronneberger et al., 2015). With proper fine-tuning and hyperparameter seeking, this U-shape-like network can achieve state-of-the-art performance in a large diversity of medical image segmentation tasks (Isensee et al., 2021). For example, to reduce the prostate appearance variation in transverse slices, Orlando et al. (2020a,b) resampled 2D slices radially around the superior-inferior axis of the 3D TRUS image, and then utilized a standard U-Net to predict 2D contours on the slices to reconstruct the 3D prostate

volume. Yang et al. (2017) serialized static 2D ultrasound images into dynamic sequence and exploited recurrent neural networks (RNNs) to infer the prostate shape sequentially, aiming to bridge the missing boundaries through long short-term memory (LSTM). Wang et al. (2018, 2019b) leveraged attention mechanisms to selectively extract multi-level features from 2D TRUS slices, facilitating the prostate segmentation by suppressing irrelevant background noise while enhancing prostate structural details. Their following work (Wang et al., 2019b) extended this method to 3D manner and performed more comprehensive evaluations on a larger dataset. Lei et al. (2019b,a) integrated multi-view ensemble learning and deep supervision strategy into 3D V-Net (Milletari et al., 2016) to refine the prostate segmentation with limited training data.

Overall, the advanced DL techniques have substantially extended the frontier of the prostate ultrasound segmentation to a very high level, where the DSC of the whole prostate can generally surpass 90% (Yang et al., 2017; Anas et al., 2018; Wang et al., 2018; 2019b; Lei et al., 2019b; 2019a; Orlando et al., 2020a; 2020b; Geng et al., 2020). However, the existing algorithms' performance is still unsatisfactory in the prostate apex and base sub-areas, which are only occasionally reported in the literature. Furthermore, along with the improvement in segmentation accuracy, the DL model size and complexity have increased significantly, resulting in high demand of computational hardware and elevated over-fitting risks.

2.2. Uncertainty estimation of DL models

DL models often tend to produce overconfident predictions (Lakshminarayanan et al., 2016). Quantitatively assessing the uncertainty or reliability of the DL models is indispensable for the practical applications, especially for the healthcare systems, where the overconfident incorrect predictions could be harmful even fatal to the patient. The basic idea to quantify the DL model uncertainty is to introduce stochasticity at the inference stage. In consequence, with multiple inferences, the DL model could produce multiple different predictions on the same input. The inconsistency of these predictions can be utilized to gauge the uncertainty.

Currently, there are two representative methods widely used for uncertainty estimation in medical image analysis, i.e., the Monte Carlo dropout (MC-dropout) method (Gal and Ghahramani, 2016) and the test-time augmentation (TTA) method (Wang et al., 2019a). The former imposes stochasticity to the DL model and the latter does it to the input data. For example, Balagopal et al. (2021) embedded the MC-dropout layers in their DL model to estimate the uncertainty of the predicted segmentation of prostate bed, which is a hardly visible clinical target volume with non-contrast boundaries. Yu et al. (2019) combined the MC-dropout mechanism with mean-teacher (Tarvainen and Valpola, 2017) learning framework, aiming to utilize the estimated uncertainty map to guide the semi-supervised learning of left atrium segmentation. Xia et al. (2020a,b) proposed a similar idea to use the MC-dropout uncertainty as the weights for the pseudo labels in semi-supervised medical image segmentation. Wang et al. (2019a) utilized the TTA strategy to quantify the CNNs' uncertainty in medical image segmentation. They also provided a theoretical formulation of the TTA uncertainty estimation. Moshkov et al. (2020) applied the same idea of TTA to assess uncertainty of the cell segmentation in microscopy images.

Due to the severe imaging noise and ambiguous boundaries in the TRUS images, the DL-based methods inevitably exhibit some errors in the predicted contours for prostate ultrasound segmentation, especially in the apex and base sub-areas. Estimating the uncertainty map of the predicted segmentations would be of great clinical interest, since it can highlight where the generated contour is less reliable and needs further modifications or approval. Moti-

vated by this clinical need, we incorporate a novel CPTTA strategy into our PTN to realize an uncertainty-aware prostate ultrasound segmentation. Compared with the MC-dropout method and regular image-based TTA method, our CPTTA can not only estimate the uncertainty map but also act as a refinement on the PTN segmentation as demonstrated by our ablation study in Sec 4.

3. Method

Fig. 2 gives an overview of the proposed PTN for 3D prostate ultrasound segmentation. It consists of three components: 1) a *localization sub-network* estimating the prostate centroid position in the original TRUS image coordinates, 2) a novel *polar transform layer* conducting polar transformation on the input image and intermediate feature maps, and 3) a *surface regression sub-network* regressing the prostate surface radius with respect to its centroid in the polar coordinates. The final prostate segmentation is reconstructed from the predicted centroid and the surface radii. As the polar transform layer is implemented in a differentiable manner, the entire PTN can be trained and tested end-to-end while directly taking the original TRUS image as input without any pre-processing. In the remaining part of this section, we will present the details about the three major components in Section 3.1, Section 3.2, and Section 3.3, respectively. Afterwards, we introduce CPTTA for uncertainty estimation in our PTN in Section 3.4. Lastly, some implementation details are provided in Section 3.5.

3.1. Prostate centroid localization

As the first step of our PTN, we employ a localization sub-network to estimate the prostate centroid position in the input TRUS image. The architecture of this localization sub-network derives from the ResNet-18 (He et al., 2016) as illustrated in Fig. 2. It takes a 3D TRUS image as input and outputs three scalars (c_x, c_y, c_z) to regress the prostate centroid coordinate in the original image space. We use a *Tanh* activation function as the output layer to normalize the regressed coordinates to a range of $[-1, 1]$, which corresponds to the extents of the 3D TRUS volume in each direction. The localization sub-network is trained by minimizing the mean squared error (MSE) loss L_{loc} between the estimated prostate centroid (c_x, c_y, c_z) and the ground-truth center of mass ($\hat{c}_x, \hat{c}_y, \hat{c}_z$):

$$L_{loc} = \frac{1}{3}[(c_x - \hat{c}_x)^2 + (c_y - \hat{c}_y)^2 + (c_z - \hat{c}_z)^2]. \quad (1)$$

3.2. Differentiable 3D polar transformation

Given the estimated prostate centroid $\mathbf{c} = (c_x, c_y, c_z)$ as the polar center (i.e., the origin of polar coordinate system), the correlation between the Cartesian coordinates (x, y, z) and the polar coordinates (u, v, r) can be expressed as:

$$\begin{cases} x = c_x + r \cdot \cos(v) \cdot \sin(u) \\ y = c_y + r \cdot \cos(v) \cdot \cos(u) \\ z = c_z + r \cdot \sin(v) \end{cases} \quad \text{and} \quad \begin{cases} u \in [-\pi, \pi) \\ v \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ r \in [0, 1] \end{cases}, \quad (2)$$

where u , v , and r are the longitude angle, latitude angle, and radius of a polar coordinate, respectively. Let $\mathbf{I} \in \mathbb{R}^{D \times H \times W}$ denote a 3D image or feature map¹, our goal is to design a polar transform layer:

$$\mathbf{I}_p = \mathcal{F}_{polar}(\mathbf{I}, \mathbf{c}) \quad (3)$$

¹ For brevity in the following contents, we use "image" to refer to both image and feature map.

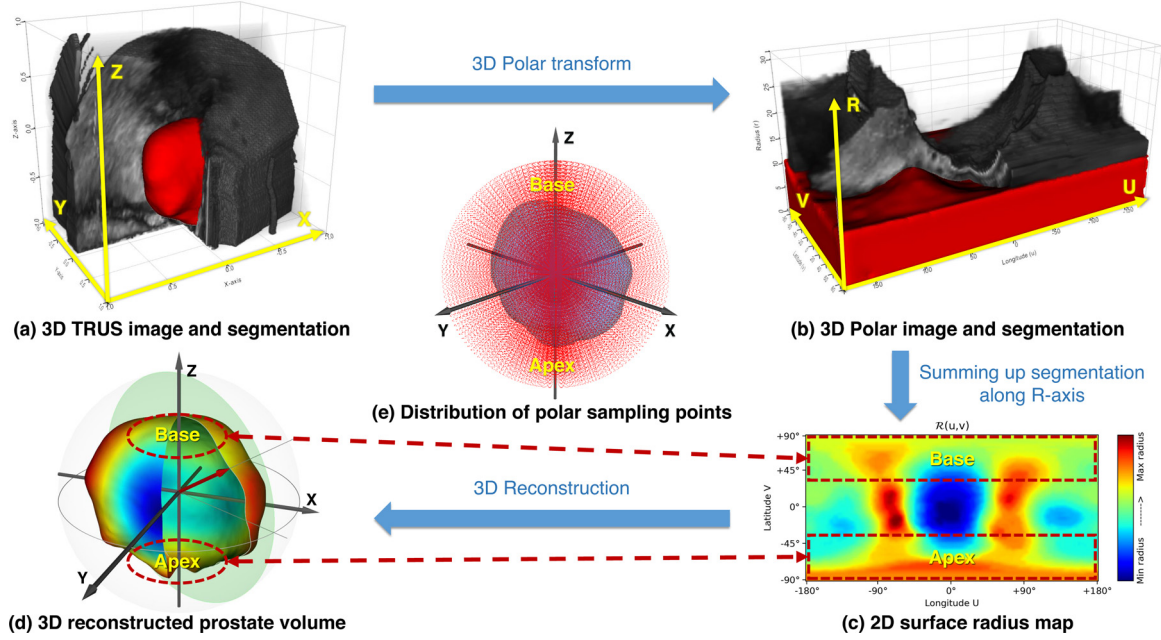


Fig. 1. 3D polar transform of TRUS image and prostate volume.

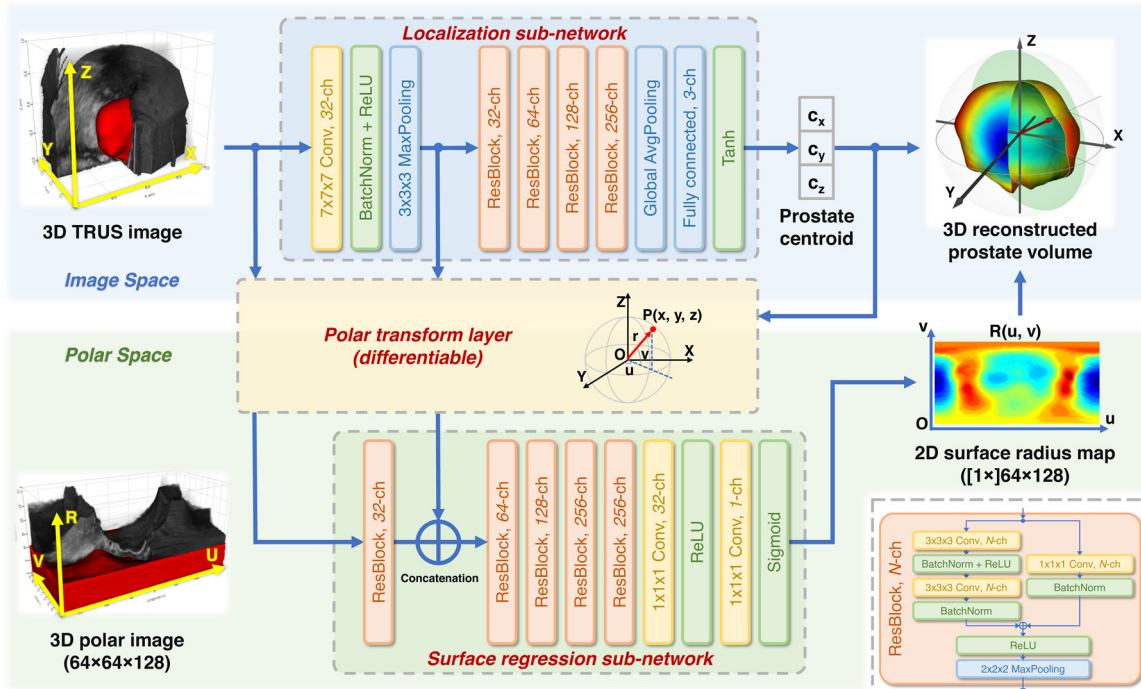


Fig. 2. Diagram of the proposed polar transform network (PTN), which is composed of a localization sub-network and a surface regression sub-network connected by a polar transform layer. It is worth noting that the polar transform layer is implemented in a differentiable way, which makes the entire PTN can be trained and tested in an end-to-end manner. The bottom-right legend shows the inner structure of the residual blocks used in the networks.

which takes I and c as inputs and outputs the transformed polar image I_p defined in the polar space. Inspired by the spatial transformer network (Jaderberg et al., 2015), we formulate this polar transform layer as a grid sampling procedure. Specifically, we first generate a sampling grid² $G_p = \{(u, v, r)\}$ in the polar space to define the size and scope of the output polar image $I_p \in \mathbb{R}^{R \times V \times U}$. This

can be done by sampling $R \times V \times U$ evenly spaced points in the polar space defined in Eq. (2). Then, through polar coordinate transform, each sampling point $(u, v, r) \in G_p$ is mapped to a spatial location (x, y, z) in the input image I , where a sampling kernel is applied to extract the intensity value to fill the corresponding voxel in the output polar image $I_p(u, v, r)$. This procedure can be formulated as:

$$I_p(u, v, r) = \frac{\sum_d^D \sum_h^H \sum_w^W I(w, h, d) \cdot k(x, w) \cdot k(y, h) \cdot k(z, d)}{\sum_d^D \sum_h^H \sum_w^W k(x, w) \cdot k(y, h) \cdot k(z, d)}, \quad (4)$$

² By sampling grid we refer to a point grid, which consists of a number of sampling points uniformly distributed in space. Each sampling point has a 3D coordinate and corresponds to a voxel in the associated image.

where W , H , and D are the width, height, and depth (in pixel) of an input image \mathbf{I} , respectively. (x, y, z) is the Cartesian coordinates derived from the polar coordinates (u, v, r) and the polar center \mathbf{c} through Eq. (2). The denominator $\sum_d^D \sum_h^H \sum_w^W k(x, w) \cdot k(y, h) \cdot k(z, d)$ is used to normalize the sampling output. $k(a, b)$ is a sampling kernel function for linear interpolation, which is defined as:

$$k(a, b) = \max(0, 1 - |a - b|). \quad (5)$$

As the denominator $\sum_d^D \sum_h^H \sum_w^W k(x, w) \cdot k(y, h) \cdot k(z, d)$ is designed to be a unit value, Eq. (4) can be further simplified as:

$$\mathbf{I}_p(u, v, r) = \sum_d^D \sum_h^H \sum_w^W \mathbf{I}(w, h, d) \cdot k(x, w) \cdot k(y, h) \cdot k(z, d), \quad (6)$$

Note that, before the kernel sampling, the normalized Cartesian coordinates (x, y, z) will be converted to pixel indices to fit the interpolation operation.

On the other hand, the partial derivatives of the output polar image \mathbf{I}_p with respect to the input image \mathbf{I} and centroid coordinate c_x can be calculated as:

$$\frac{\partial \mathbf{I}_p(u, v, r)}{\partial \mathbf{I}(w, h, d)} = k(x, w) \cdot k(y, h) \cdot k(z, d), \quad (7)$$

$$\frac{\partial \mathbf{I}_p(u, v, r)}{\partial c_x} = \sum_{d, h, w}^{D, H, W} \mathbf{I}(w, h, d) \cdot \frac{\partial k(x, w)}{\partial c_x} \cdot k(y, h) \cdot k(z, d), \quad (8)$$

respectively, where

$$\frac{\partial k(x, w)}{\partial c_x} = \frac{\partial k(x, w)}{\partial x} = \begin{cases} 1, & |x - w| < 1 \text{ and } x < w \\ -1, & |x - w| < 1 \text{ and } x \geq w, \\ 0, & |x - w| \geq 1 \end{cases} \quad (9)$$

The partial derivatives $\frac{\partial \mathbf{I}_p(u, v, r)}{\partial c_y}$ and $\frac{\partial \mathbf{I}_p(u, v, r)}{\partial c_z}$ have similar forms to $\frac{\partial \mathbf{I}_p(u, v, r)}{\partial c_x}$ since the variables x , y , and z are symmetrical in Eq. (6). These partial derivatives allow the loss back-propagate through the polar transform layer, resulting in a (sub-)differentiable procedure that can be integrated into any end-to-end trained deep network.

In this study, we empirically set $R \times V \times U = 64 \times 64 \times 128$ to define the polar sampling grid G_p with sufficient resolution to represent the prostate volume. During training, the proposed polar transform layer is also used to transform the ground-truth prostate mask $\hat{\mathbf{Y}}$, resulting in the polar mask $\hat{\mathbf{Y}}_p$. Fig. 1a and 1b give an illustration of the polar transformation from the image space to the polar space, using both TRUS image volume (in gray color) and the associated prostate mask (in red color). Through the polar transformation, the region inside the prostate can get denser sampling points than the outside regions (see Fig. 1e), which naturally rebalances the foreground and background by increasing the prostate samples for training. Moreover, the two poles of the prostate volume, which generally correspond to the base and apex areas, can have much denser samples than the region around its equator (i.e., the main gland). This effectively emphasizes the weight of the prostate base and apex in the input image and intermediate feature maps, enforcing the network to pay more attention to these areas to lower the training loss.

Since we choose the prostate centroid as the polar origin, the prostate voxels will distribute in the bottom part of the transformed polar image (see Fig. 1b), while the top part will be filled by background voxels. Due to this special label distribution, we can compress the 3D polar mask $\hat{\mathbf{Y}}_p(u, v, r) \in \{0, 1\}$ into a 2D surface radius map (SRM), as shown in Fig. 1c, by summing along the radial direction (r -axis):

$$\hat{\mathcal{R}}(u, v) = \frac{1}{R} \sum_{r_i} \hat{\mathbf{Y}}_p(u, v, r_i), \quad (10)$$

where R is the size (in pixel) of the polar mask $\hat{\mathbf{Y}}_p(u, v, r)$ along r -axis. Each element $\hat{\mathcal{R}}(u, v) \in [0, 1]$ of this SRM represents the normalized distance from the polar center \mathbf{c} to a prostate surface point

along the radial direction specified by (u, v) . Inversely, given the polar center \mathbf{c} and the 2D SRM $\hat{\mathcal{R}}(u, v)$, we can reconstruct the prostate volume in the original TRUS image space (see from Fig. 1c to Fig. 1d). The representation of the prostate centroid jointed with the SRM is much more efficient than the conventional voxel-wise binary mask, hence it can largely reduce the model parameters required to represent the discriminative features for prostate segmentation.

3.3. Prostate surface radius regression

Following the polar transform layer, we exploit a surface regression sub-network to regress the SRM based on the transformed polar image and feature maps. The architecture of the surface regression sub-network is shown in Fig. 2, which consists of five successive residual blocks (He et al., 2016) followed by two $1 \times 1 \times 1$ convolutional layers. The first four residual blocks end with a down-sample layer that has an anisotropic kernel shape of $2 \times 3 \times 3$ with a stride of $2 \times 1 \times 1$. This structure can gradually extract the global semantic features from the input polar image along the radial direction (r -axis) while keep the other two spatial dimensions (u -/ v -axis) unchanged. The last residual block ends with an average pooling layer with $4 \times 1 \times 1$ size kernel, which finally produces “2D” feature maps in a spatial size of $(1 \times) 64 \times 128$. We use a *Sigmoid* activation as the output layer to limit the regressed value in a range of $[0, 1]$, representing the estimated boundary location between the prostate center (0) and the polar space extent (1). For all the convolutional layers in the surface regression sub-network, a circular padding mode is specified when padding the feature maps along the longitude direction (u -axis), which mimics the circular loop nature of the polar transformation. The surface regression sub-network is trained by minimizing the MSE loss L_{sur} between the regressed SRM $\mathcal{R}(u, v)$ and the ground-truth SRM $\hat{\mathcal{R}}(u, v)$.

$$L_{sur} = \frac{1}{UV} \sum_u^U \sum_v^V [\mathcal{R}(u, v) - \hat{\mathcal{R}}(u, v)]^2 \quad (11)$$

Finally, the localization sub-network and the surface regression sub-network are jointly trained by minimizing the following multi-task loss function:

$$Loss = L_{loc} + L_{sur}. \quad (12)$$

With the prostate centroid (c_x, c_y, c_z) estimated by the localization sub-network and the SRM $\mathcal{R}(u, v)$ regressed by the surface regression sub-network, we can reconstruct the prostate volume in the original TRUS image space through Eq. (2) and convert it to a segmentation mask \mathbf{Y} with the same size as the input image \mathbf{I} .

3.4. Uncertainty estimation

To quantitatively assess the uncertainty of the segmentation results, based on the previous work by Wang et al. (2019a), we propose a new strategy of *centroid perturbed test-time augmentation* (CPTTA). The underlying assumption of CPTTA is that, by imposing random perturbations (e.g., data augmentation) on the input, the DL model tend to produce inconsistent output when the model is uncertain about its predictions. Unlike the regular TTA method performing data augmentation on the network input, our CPTTA conducts augmentation on the intermediate output of the network, i.e., the detected prostate volume centroid. Specifically, given a testing image \mathbf{I} , we first locate the prostate centroid \mathbf{c} using the localization sub-network. Based on the estimated centroid \mathbf{c} , we generate K perturbed centroids by injecting a zero-mean Gaussian noise $\mathbf{n}_i \sim \mathcal{N}^3(0, \sigma^2)$ to it:

$$\tilde{\mathbf{c}}_i = \mathbf{c} + \mathbf{n}_i, \quad i \in \{1, 2, \dots, K\}. \quad (13)$$

These noisy centroids $\{\tilde{\mathbf{c}}_i\}_{i=1}^K$ are then used to conduct the following polar transformation and surface radius regression in our PTN, resulting in a set of prostate segmentation $\{\tilde{\mathbf{Y}}_i\}_{i=1}^K$. Since the prostate shape is independent on the choice of polar center, the generated segmentation should be consistent across different $\tilde{\mathbf{c}}_i$. Otherwise, if $\tilde{\mathbf{Y}}_i$ exhibits inconsistency in some local regions, the result should be considered as less reliable with high uncertainty. Therefore, by calculating the standard deviation of $\tilde{\mathbf{Y}}_i$, we can get a pixel-wise uncertainty map \mathbf{U} to quantify the reliability of the segmentation:

$$\mathbf{U} = \sqrt{\mathbb{E}[\tilde{\mathbf{Y}}_i^2] - \mathbb{E}[\tilde{\mathbf{Y}}_i]^2}, \quad (14)$$

where a pixel assigned with a higher uncertainty value indicates a less reliable result.

In addition, we propose to use the expectation $\mathbb{E}[\tilde{\mathbf{Y}}_i]$ as the final segmentation, instead of the regular segmentation \mathbf{Y} derived from the unperturbed centroid \mathbf{c} . Averaging the results derived from noisy centroids can act as a compensation to the potential localization error in the estimated centroid \mathbf{c} and, therefore, improve the segmentation accuracy. We will demonstrate this efficacy in an ablation study in Section 4.

In our experiments, we empirically set the number of augmentations to $K=20$ and the perturbation standard deviation $\sigma=0.05$ to get a smooth uncertainty estimation. Since the TTA strategy merely impacts the inference stage of the PTN, it will not bring any extra computational cost to the training procedure. By parallel computation, the inference time of a segmentation as well as the corresponding uncertainty map can also be limited within one forward pass.

3.5. Implementation details

The proposed PTN is fully implemented in 3D using PyTorch framework. We center-crop and resample the input TRUS images to a uniform size of $256 \times 256 \times 256$ with a spacing of $0.5 \times 0.5 \times 0.5 \text{ mm}^3$. The image intensities are normalized from $[0.0, 255.0]$ to $[0.0, 1.0]$. Random translation ($[-5, 5] \text{ mm}$) and rotation ($[-0.05, 0.05] \text{ rad}$) are used to augment the training data. Model parameters are initialized using Xavier algorithm (Glorot and Bengio, 2010) and optimized by Adam optimizer (Kingma and Ba, 2014) with a base learning rate of 0.001 for 400 epochs and a batch size of 16. We evaluate the model on the validation set every epoch and save the model achieving the highest accuracy as the final model to be tested on the testing set. During training, we directly use the ground-truth prostate centroid as the input of the polar transform layer since it can accelerate the convergence of the subsequent surface regression sub-network. To alleviate the risk of over-fitting, we also add a uniformly distributed random translation ($[-5, 5] \text{ mm}$) to the input centroid coordinate. The source code of our PTN is released at <https://github.com/DIAL-RPI/PTN>.

4. Experiments

4.1. Dataset and metrics

We evaluate our method on a clinical dataset consisting of 315 3D TRUS images with the whole prostate segmentation mask annotated by two experienced radiologists. The TRUS image volumes are reconstructed from 2D TRUS frame sequence acquired by a Philips iU22 ultrasound scanner with C9-5 probes. To make full use of the data, we conduct a 3-fold cross validation to evaluate the performance. Specifically, the 315 data samples are randomly divided into 3 folds with 105 samples in each fold. In each iteration of the 3-fold cross validation, one fold is used for testing and the rest two folds are used for training and validation (with a fixed split of 180 and 30 samples for training and validation, respectively).

We use Dice similarity coefficient (DSC), average symmetric surface distance (ASD), and Hausdorff distance (HD) to quantitatively evaluate the segmentation performance. We also conduct paired t-test between the results of different models. A p -value lower than 0.05 is considered as an evidence of statistical significance.

4.2. Comparison with other methods

4.2.1. Benchmarking methods

To justify the performance of the proposed PTN, we compare our method with three state-of-the-art methods for medical image segmentation (Milletari et al., 2016; Ronneberger et al., 2015; Isensee et al., 2021) and two representative methods dedicated to prostate ultrasound segmentation (Orlando et al., 2020b; Wang et al., 2019b), which are briefly introduced below:

- **V-Net** (Milletari et al., 2016) and **U-Net** (Ronneberger et al., 2015): Two popular FCNs with U-shape-like architectures that are widely used for various of medical image segmentation tasks.
- **nnU-Net** (Isensee et al., 2021): The current state-of-the-art method in many benchmark medical image segmentation tasks, which is essentially a standard U-Net trained with specially tuned hyper-parameters. The value of the hyper-parameters are determined by a set of pre-defined guidelines regarding the dataset properties.
- **2.5D-UNet** (Orlando et al., 2020b): A 2.5D deep learning-based method for prostate ultrasound segmentation, in which the TRUS volume is radially resampled to a set of slices around the superior-inferior axis and a 2D U-Net is trained to segment the prostate in these radial slices to reconstruct the 3D prostate volume.
- **DAF-Net** (Wang et al., 2019b): A 3D deep learning method for prostate ultrasound segmentation, where a feature pyramid network combined with a special-designed attention module is trained through deep supervision to deal with the complex background condition in TRUS images.

Both U-Net and nnU-Net are implemented in 3D. All the competing methods are trained and evaluated using the same 3-fold cross validation and pre-/post-processing as our method.

4.2.2. Comparison results

In Table 1, we list the comparison results. The best results are marked in bold. The underlined results indicate statistically non-significant difference with our result ($p > 0.05$). Fig. 3 visualizes some 3D segmentation results rendered by the surface distance error.

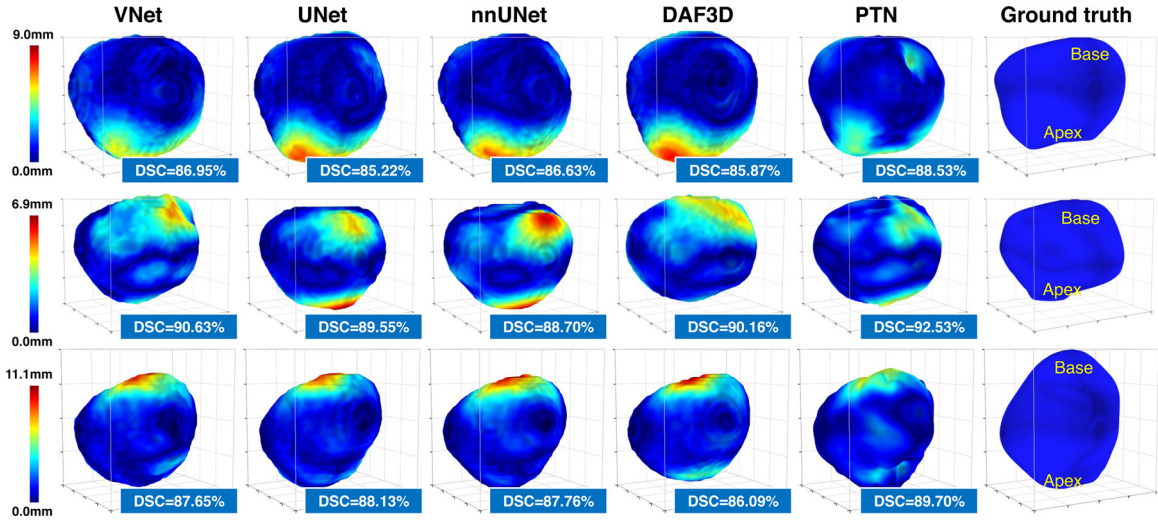
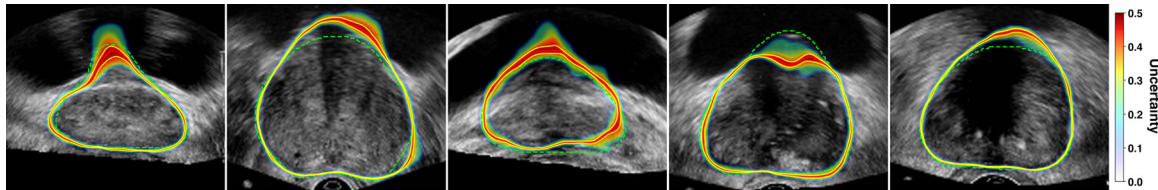
It can be seen that, for the whole prostate segmentation, our method outperforms the competing methods on all entries except the HD metric slightly higher than DAF-Net but without statistical significance ($p=0.3178$). To further investigate the segmentation performance in different sub-areas of the prostate, we divide the prostate volume into three sections, i.e., the apex, base, and main gland³, and calculate the DSC metric for them separately (see the right part of Table 1). For all the three sections, our PTN consistently achieves higher DSC than other competing methods. Most of the outperformance margins are statistically significant except the apex section against UNet ($p=0.0760$) and the base section against nnUNet ($p=0.2283$). This superior performance on the apex and base sub-areas demonstrates the effectiveness of our PTN, in which the learning of apex and base features can be emphasized by

³ We evenly split the ground-truth prostate volume into three parts with equal length along the superior-inferior (head-foot) direction. The top, middle, and bottom part of the prostate volume are roughly considered as the base, main gland, and apex, respectively.

Table 1

Comparison of different methods for prostate ultrasound segmentation. The values in the brackets denote the standard deviations. Best results are marked in bold. The underlined results indicate no statistically significant difference with our result ($p > 0.05$).

Methods	Whole prostate [mean (std)]			DSC of sub-areas [mean (std) %]			Parameter number
	DSC [%]	ASD [mm]	HD [mm]	Apex	Base	Main gland	
2.5D-UNet (Orlando et al., 2020b)	86.13 (5.49)	1.80 (1.07)	8.38 (4.08)	80.46 (9.93)	81.91 (8.00)	90.84 (4.63)	15M
VNet (Milletari et al., 2016)	88.15 (4.57)	1.51 (0.78)	7.86 (4.22)	82.01 (9.99)	84.43 (7.89)	92.97 (3.32)	68M
UNet (Ronneberger et al., 2015)	89.28 (4.55)	1.34 (0.68)	7.36 (3.65)	84.57 (7.78)	85.18 (7.36)	93.73 (3.88)	43M
nnUNet (Isensee et al., 2021)	89.36 (4.56)	<u>1.35</u> (0.81)	7.09 (3.97)	84.29 (8.11)	<u>85.68</u> (7.01)	93.68 (4.04)	43M
DAF-Net (Wang et al., 2019b)	89.07 (4.15)	1.39 (0.74)	6.92 (3.59)	83.81 (8.38)	85.26 (6.75)	93.49 (3.61)	29M
PTN	89.78 (3.41)	1.32 (0.61)	7.35 (3.13)	84.84 (7.56)	85.71 (5.76)	94.30 (2.58)	12M
PTN+CPTTA	89.97 (3.47)	1.30 (0.61)	7.07 (3.19)	85.09 (7.67)	86.00 (5.75)	94.37 (2.69)	12M

**Fig. 3.** 3D visualization of the segmentation results rendered by surface distance error.**Fig. 4.** Visualization of the segmentation results superimposed with the estimated uncertainty map by the proposed PTN. The red solid line and green dashed line represent the predicted and ground-truth prostate boundary, respectively. The yellow lines inside and outside the predicted boundary indicate the low- and upper-bound of the 95% confidence interval band, respectively.

denser sampling grid through the polar transformation. CPTTA further improves the segmentation accuracy of our method as shown in Table 1.

It is worth noting that the number of learnable parameters in our PTN is around 12M, which is much smaller than other 3D models (varying from 29M to 68M) and even the 2D model proposed by Orlando et al. (2020b) (15M). This clearly demonstrates the parametric efficiency of our method by decomposing the 3D voxel-wise segmentation into a centroid localization task jointed with a 2D surface radius regression task.

4.3. Uncertainty estimation of prostate segmentation

Uncertainty estimation is a featured function of our PTN that distinguishes it from other prostate ultrasound segmentation methods. To justify the clinical significance of the uncertainty estimation, in Fig. 4, we visualize some segmentation results superimposed with the estimated uncertainty map. It can be seen that the estimated uncertainty map effectively highlights the regions where the prostate is suffering from severe noise and low-contrast

even missing boundaries. This observation is in line with our common sense that the physicians often need to pay extra efforts in such areas to carefully delineate the prostate contour. Therefore, the uncertainty map can provide valuable guidance for the clinicians when they are modifying the contours generated by our PTN.

To analysis the correlation between the segmentation accuracy and uncertainty, we define a metric, called *95% confidence interval (CI) band width*, to quantify the overall uncertainty level for each case. Specifically, since the multiple TTA predictions approximately follow a Gaussian distribution, the lower- and upper-bound of the generated segmentation $\{\tilde{Y}_i\}_{i=1}^K$ can be calculated by:

$$\begin{cases} \mathbf{Y}_{lower} = \mathbb{E}[\tilde{Y}_i] - Z * \frac{U}{\sqrt{K}} \\ \mathbf{Y}_{upper} = \mathbb{E}[\tilde{Y}_i] + Z * \frac{U}{\sqrt{K}} \end{cases} \quad (15)$$

where the Z-score is set to 1.96 to correspond a 95% confidence interval (CI). By a naive 0.5 thresholding on \mathbf{Y}_{lower} and \mathbf{Y}_{upper} , we can get the lower- and upper-bound contour (illustrated as the yellow lines inside and outside the predicted contour, respectively, in Fig. 4). The region between the lower- and upper-bound contours is the 95% CI band, where the predicted boundaries most

Table 2

Prostate localization and segmentation results of our PTN when using different localization methods to estimate the prostate centroid. The values in the brackets denote the standard deviations. Best results are marked in bold.

Prostate localization methods	Centroid localization error [mm]				Segmentation performance		
	x-axis	y-axis	z-axis	dist.	DSC [%]	ASD [mm]	HD [mm]
Image center	2.46 (1.89)	4.22 (2.57)	5.82 (4.21)	8.43 (3.83)	87.83 (6.15)	1.62 (0.99)	8.97 (4.63)
Statistical centroid	1.96 (1.54)	2.65 (2.01)	5.19 (4.01)	6.81 (3.73)	88.57 (4.72)	1.50 (0.78)	8.24 (3.95)
Ground-truth centroid	-	-	-	-	90.30 (2.82)	1.24 (0.51)	6.96 (2.76)
Localization sub-network	1.14 (0.93)	1.25 (1.11)	1.66 (1.47)	2.74 (1.53)	89.78 (3.41)	1.32 (0.61)	7.35 (3.13)

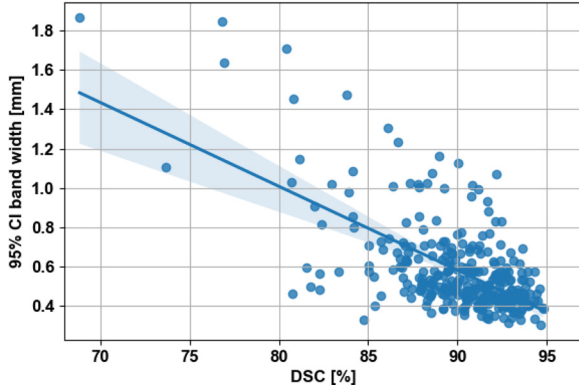


Fig. 5. Correlation between the segmentation Dice similarity coefficient (DSC) and the 95% confidence interval (CI) band width. Each point represents one case.

likely present. Obviously, the mean width of this 95% CI band indicate the case-wise uncertainty level. The wider the 95% CI band is, the higher uncertainty the predicted contour would exhibit. Here, we use the ASD between the lower- and upper-bound contours as the surrogate of the mean 95% CI band width. We plot the segmentation DSC vs. the 95% CI band width of all the testing cases in Fig. 5. It can be seen that the segmentation accuracy generally shows a negative correlation with the uncertainty level. This result suggests the clinical value of our estimated uncertainty map, with which the physicians can effectively focus on the cases or regions associated with wider 95% CI band when they are amending the auto-generated contours.

4.4. Effects of different prostate localization methods

In our PTN method, the prostate centroid is estimated by the localization sub-network, whose accuracy will directly impact the subsequent polar transformation as well as the final segmentation result. There are some other optional solutions to achieve the goal of prostate centroid localization. To justify our choice of the localization sub-network, in this ablation study, we replace the localization sub-network with other three prostate localization methods, including: 1) using the image volume center to approximate the prostate centroid, 2) using the statistical value (*i.e.*, the mean value) of the prostate centroid over the training samples, and 3) directly using the ground-truth prostate centroid during inference time. We successively evaluate the localization accuracy of these centroid localization methods and the segmentation accuracy of our PTN when cooperating with them. The centroid localization accuracy is assessed by the absolute distance between the predicted and ground-truth centroid coordinates. Table 2 lists the results of this experiment. Our localization sub-network significantly outperforms the other two localization methods in terms of localization accuracy. Furthermore, the localization error along the z-axis (*i.e.*, the apex-base direction) is higher than that along the x- and y-axes. This result is in line with the fact that the prostate bound-

Table 3

Segmentation accuracy of PTN combined with different uncertainty estimation mechanisms. The values in the brackets denote the standard deviations. Best results are marked in bold.

Model	DSC [%]	ASD [mm]	HD [mm]
PTN	89.78 (3.41)	1.32 (0.61)	7.35 (3.13)
PTN+TTA on image	89.83 (3.40)	1.31 (0.61)	7.28 (3.14)
PTN+MC dropout	89.33 (4.36)	1.40 (0.85)	7.54 (3.33)
PTN+CPTTA	89.97 (3.47)	1.30 (0.61)	7.07 (3.19)

ary near the apex and base regions is harder to recognize compared with the main gland. The segmentation performance follows a similar trend. Using image volume center and statistical centroid as the polar coordinate origins results in the whole prostate segmentation DSC dropping from 89.78% to 87.83% and 88.57%, respectively. This result demonstrates the efficacy of our localization sub-network and the importance of an accurate prostate centroid to the PTN. On the other hand, although using the ground-truth centroid can further improve the segmentation accuracy to a mean DSC of 90.30%, the localization sub-network has already helped our PTN to achieve a comparable result to that and outperform other competing methods in the task as shown in Table 1. It also suggests that the accuracy of our PTN can be further improved by incorporating more advanced localization strategies or manual interactions when feasible.

4.5. Comparison between CPTTA and other uncertainty estimation methods

As aforementioned in Section 2, there are other two representative methods commonly used for uncertainty estimation, *i.e.*, the Monte Carlo dropout (MC-dropout) (Gal and Ghahramani, 2016) and the test-time augmentation (TTA) (Wang et al., 2019a). To justify our design of CPTTA, we evaluate the segmentation accuracy of our PTN when using TTA, MC-dropout, and our proposed CPTTA, respectively. In this ablation study, both the TTA method and the MC-dropout method have the same inference time $K=20$ as our CPTTA to generate multiple segmentations on each testing image and use the mean of these segmentations as the final result. For TTA, we perturb the input image using the same random translation ($\sim \mathcal{N}^3(0, 0.05^2)$) as we applied on the prostate centroid in CPTTA. For the MC-dropout method, we insert dropout layers (with a dropout probability of 0.5) to the end of each residual block (before the MaxPooling layer) in the surface regression sub-network. According to the experimental results shown in Table 3, CPTTA ("PTN+CPTTA") outperforms both the TTA ("PTN+TTA on image") and the MC-dropout ("PTN+MC dropout") in terms of segmentation accuracy. We attribute this performance gain to the ensemble of multiple results derived from noisy centroid, which is essentially a compensation to the localization error introduced by the localization sub-network.

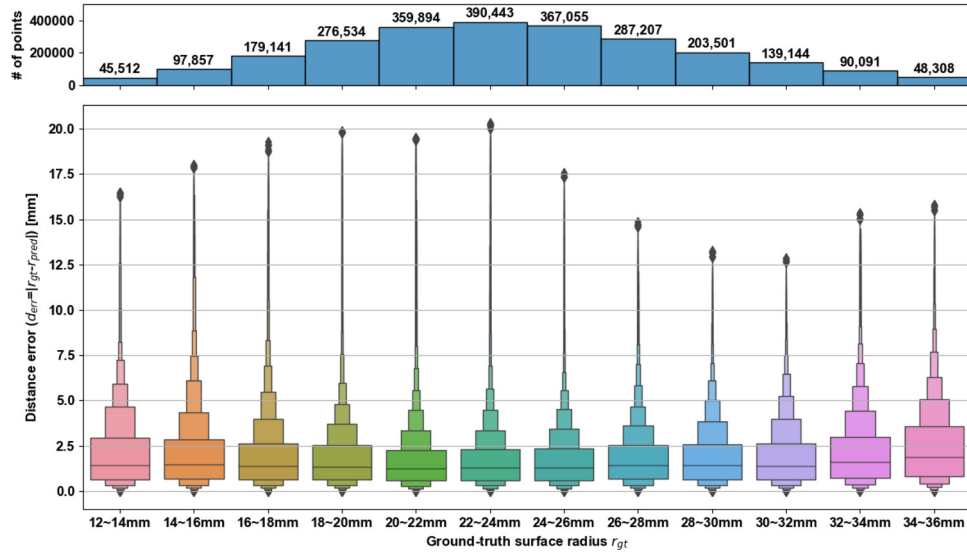


Fig. 6. Distance error of surface points with different surface radii. The surface points are grouped by their surface radii. The top histogram shows the number of surface points in each group. The bottom boxplot shows the distribution of the distance errors of each group.

5. Discussion

5.1. Generalization ability to other medical image segmentation tasks

Since our PTN method does not rely on any specific ultrasound image property, it is modality-agnostic and can deal with the prostate segmentation tasks in other image modalities, such as MRI (Zhu et al., 2019) and CT (Xu et al., 2020, 2021). The core idea of segmenting the prostate in the polar coordinates is also suitable for other bubble-shaped organs, such as the bladder (Xu et al., 2018), as long as the organ surface can be represented in and reconstructed from a 2D SRM with respect to a center point.

5.2. Estimation accuracy of SRM regarding different radii of surface points

In the polar coordinate system, since different points on the prostate surface may have different distances to the centroid, the sampling density around the surface points can vary depending on their locations. Points closer to the centroid have a denser sampling density than those located further away from the centroid. To determine how this inconsistent sampling density impacts the accuracy of the predicted SRM, we conduct the following quantitative analysis. As the sampling density of the surface points can be well surrogated by their distances to the centroid, i.e., the surface radius r_{gt} in the ground-truth SRM, we first split the surface points into groups based on their radii and plot the number of the surface points in each group, resulting in the histogram shown at the top of Fig. 6. We can see that the radii of the surface points generally follow a Gaussian distribution, which demonstrates the variation of the sampling density across different surface points. We then quantify the estimation accuracy of the surface points in SRM by the absolute error between the ground-truth surface radius r_{gt} and the predicted surface radius r_{pred} as $d_{err} = |r_{gt} - r_{pred}|$, resulting in the boxplot in the bottom of Fig. 6. Although the radius (sampling density) of the surface points varies in a wide range from 12mm to 36mm, the absolute error (estimation accuracy) generally keeps consistent. This result indicates that the sampling density of the surface points has less impact on the estimation of the SRM in our PTN method.

We also observed that, when the surface points get further away from the centroid (indicating sparser sampling densities

around the points), the reconstructed surfaces will have sparser mesh grids. However, by converting the surface mesh to the pixel-wise binary mask in the final step of our method, this effect will be alleviated since all the surface mesh grids are pixelated in isotropic resolutions.

5.3. Significance of CPTTA strategy for prostate ultrasound segmentation

The main purpose of CPTTA is to allow our PTN to quantitatively estimate the uncertainty of its predictions. This ability is of great clinical value since it can highlight the regions in the segmentation results that are less reliable and need manual modifications and approval. As a byproduct, the CPTTA can also bring some accuracy gain to the PTN by averaging multiple predicted segmentations yielded from different polar origin points. Although the improvement is somewhat marginal, it still outperforms other uncertainty estimation methods (i.e., the Monte Carlo dropout and test-time augmentation method) as we showed in Table 3.

5.4. Limitations of PTN

One potential limitation of our PTN method is that the polar transformation may not accurately represent non-convex shapes. Certain surface points may share identical radial angles with different radii to the centroid point, such as the points p_a , p_b , and p_c illustrated in Fig. 7a. This causes confusion in our polar representation, since each point on SRM should have a unique radius. However, such a limitation is not a concern for the prostate segmentation, because the prostate is generally in a convex shape. For the reader who is interested in extending our PTN to other segmentation tasks, the shape complexity of the target organ should be considered.

Another drawback of our PTN in its current form is the lack of smoothness constraint on the predicted SRMs. The top and bottom rows of the SRM should have similar values since they are sampled from a small region around the poles of the prostate volume. Some inconsistent values that appear in this region may lead to irregularities in the reconstructed 3D surface. Fig. 7b shows an example of such artifact. This problem can be solved by applying a smoothing filter (e.g., Gaussian filter) on the SRMs before reconstructing

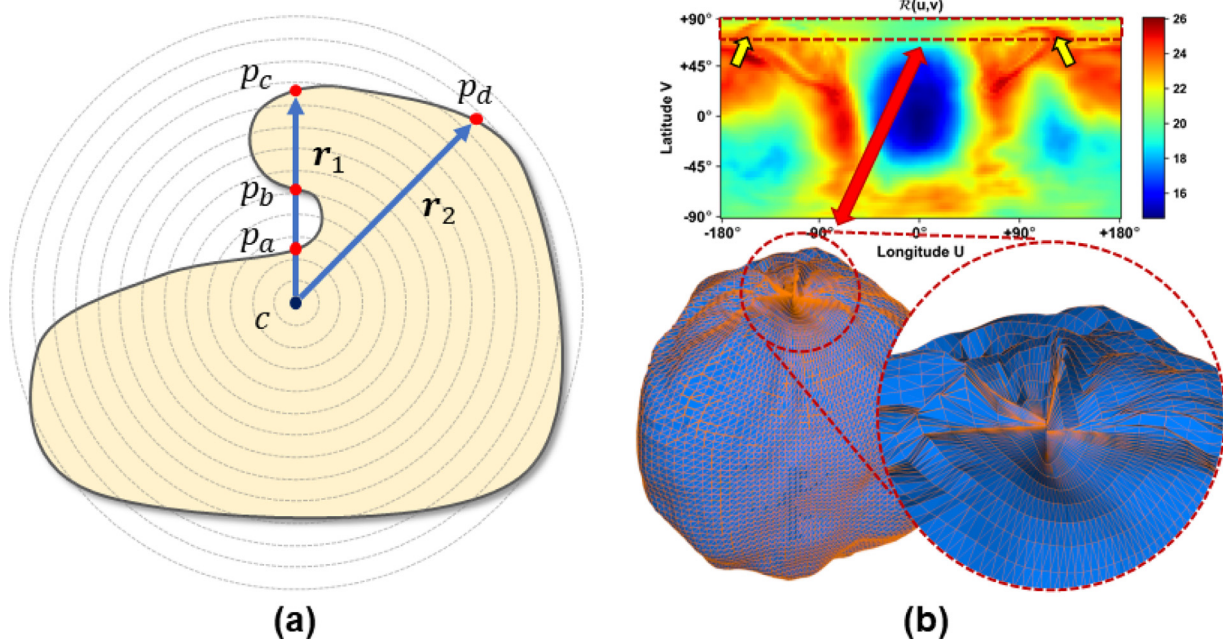


Fig. 7. (a) An example of non-convex surface. The surface points p_a , p_b , and p_c share identical radial angle but different radii to the polar center c . This causes confusion in our polar representation, where each point on SRM should have a unique radius. (b) An example of irregularity in the pole region caused by the value variations in the top rows of SRM at the locations pointed by the yellow arrows. r in this figure legend, the reader is referred to the web version of this article.)

the prostate volume. We can also impose a regularization term in the training loss of PTN to enforce smoothness.

6. Conclusion

This work addresses the problem of 3D prostate ultrasound segmentation from a fundamentally new perspective, in which the segmentation task is handled in the polar coordinate space with a transformed representation of the TRUS images and the prostate volume. Benefiting from this polar representation, the foreground prostate volume, especially the most challenging apex and base sub-areas, get emphasized through the feature learning of the segmentation network, leading to higher model accuracy. Following this idea, we propose a novel *polar transform network* (PTN), which converts the problem of prostate segmentation into a centroid localization task jointed with a surface radius regression task. The two tasks are handled by two sub-networks for localization and surface regression, respectively. The differentiable design of the polar transform layer enables the end-to-end training and testing of the entire PTN. Moreover, the *centroid perturbed test-time augmentation* (CPTTA) strategy shows effective in refining the segmentation result and conducting uncertainty estimation. Experimental results on a clinical dataset consisting of 315 TRUS images show that the proposed PTN can achieve very competitive performance with the state-of-the-art methods while using much fewer (approximately 18%~41% of) model parameters. The uncertainty map estimated by CPTTA effectively reflects the unreliable segmentation results and highlights where the auto-generated contours may need manual modification or approval. This practical function substantially improves the clinical value of our method.

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.

CRediT authorship contribution statement

Xuanang Xu: Methodology, Data curation, Software, Visualization, Writing – original draft. **Thomas Sanford:** Formal analysis, Data curation, Writing – review & editing. **Baris Turkbey:** Formal analysis, Data curation, Writing – review & editing. **Sheng Xu:** Data curation, Formal analysis, Investigation, Writing – review & editing. **Bradford J. Wood:** Investigation, Resources, Data curation, Writing – review & editing. **Pingkun Yan:** Conceptualization, Investigation, Resources, Writing – review & editing, Supervision, Project administration.

Acknowledgments

This work was partially supported by National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health (NIH) under awards R21EB028001 and R01EB027898, and through an NIH Bench-to-Bedside award made possible by the National Cancer Institute.

References

- Anas, E.M.A., Mousavi, P., Abolmaesumi, P., 2018. A deep learning approach for real time prostate segmentation in freehand ultrasound guided biopsy. *Med. Image Anal.* 48, 107–116.
- Balogopal, A., Nguyen, D., Morgan, H., Weng, Y., Dohopolski, M., Lin, M.-H., Barkousaraie, A.S., Gonzalez, Y., Garant, A., Desai, N., et al., 2021. A deep learning-based framework for segmenting invisible clinical target volumes with estimated uncertainties for post-operative prostate cancer radiotherapy. *Med. Image Anal.* 72, 102101.
- Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X., 2018. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Trans. Med. Imaging* 37 (7), 1597–1605.
- Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In: *International Conference on Machine Learning*. PMLR, pp. 1050–1059.
- Geng, L., Li, S., Xiao, Z., Zhang, F., 2020. Multi-channel feature pyramid networks for prostate segmentation, based on transrectal ultrasound imaging. *Applied Sciences* 10 (11), 3834.
- Ghose, S., Oliver, A., Mitra, J., Martí, R., Lladó, X., Freixenet, J., Sidibé, D., Vılanova, J.C., Comet, J., Meriaudeau, F., 2013. A supervised learning framework of statistical shape and probability priors for automatic prostate segmentation in ultrasound images. *Med. Image Anal.* 17 (6), 587–600.

- Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings*, pp. 249–256.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.
- Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2021. NnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* 18 (2), 203–211.
- Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K., 2015. Spatial transformer networks. *arXiv preprint arXiv:1506.02025*.
- Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Lakshminarayanan, B., Pritzel, A., Blundell, C., 2016. Simple and scalable predictive uncertainty estimation using deep ensembles. *arXiv preprint arXiv:1612.01474*.
- Lei, Y., Tian, S., He, X., Wang, T., Wang, B., Patel, P., Jani, A.B., Mao, H., Curran, W.J., Liu, T., et al., 2019. Ultrasound prostate segmentation based on multidirectional deeply supervised V-Net. *Med Phys* 46 (7), 3194–3206.
- Lei, Y., Wang, T., Wang, B., He, X., Tian, S., Jani, A.B., Mao, H., Curran, W.J., Patel, P., Liu, T., et al., 2019. Ultrasound prostate segmentation based on 3D V-Net with deep supervision. In: *Medical Imaging 2019: Ultrasonic Imaging and Tomography*, Vol. 10955. International Society for Optics and Photonics, p. 109550V.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440.
- McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A., 2017. Communication-efficient learning of deep networks from decentralized data. In: *Artificial Intelligence and Statistics. PMLR*, pp. 1273–1282.
- 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)*. IEEE, pp. 565–571.
- Moshkov, N., Mathe, B., Kertesz-Farkas, A., Hollandi, R., Horvath, P., 2020. Test-time augmentation for deep learning-based cell segmentation on microscopy images. *Sci Rep* 10 (1), 1–7.
- Orlando, N., Gillies, D.J., Gyacskov, I., Fenster, A., 2020. Deep learning-based automatic prostate segmentation in 3D transrectal ultrasound images from multiple acquisition geometries and systems. In: *Medical Imaging 2020: Image-Guided Procedures, Robotic Interventions, and Modeling*, Vol. 11315. International Society for Optics and Photonics, p. 113152I.
- Orlando, N., Gillies, D.J., Gyacskov, I., Romagnoli, C., D'Souza, D., Fenster, A., 2020. Automatic prostate segmentation using deep learning on clinically diverse 3D transrectal ultrasound images. *Med Phys* 47 (6), 2413–2426.
- 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*. Springer, pp. 234–241.
- Shen, D., Zhan, Y., Davatzikos, C., 2003. Segmentation of prostate boundaries from ultrasound images using statistical shape model. *IEEE Trans. Med. Imaging* 22 (4), 539–551.
- Tarvainen, A., Valpola, H., 2017. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. *arXiv preprint arXiv:1703.01780*.
- Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S., Vercauteren, T., 2019. Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. *Neurocomputing* 338, 34–45.
- Wang, Y., Deng, Z., Hu, X., Zhu, L., Yang, X., Xu, X., Heng, P.-A., Ni, D., 2018. Deep attentional features for prostate segmentation in ultrasound. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 523–530.
- Wang, Y., Dou, H., Hu, X., Zhu, L., Yang, X., Xu, M., Qin, J., Heng, P.-A., Wang, T., Ni, D., 2019. Deep attentive features for prostate segmentation in 3D transrectal ultrasound. *IEEE Trans. Med. Imaging* 38 (12), 2768–2778.
- Warnat-Herresthal, S., Schultze, H., Shastry, K.L., Manamohan, S., Mukherjee, S., Garg, V., Sarveswara, R., Händler, K., Pickkers, P., Aziz, N.A., et al., 2021. Swarm learning for decentralized and confidential clinical machine learning. *Nature* 594 (7862), 265–270.
- Xia, Y., Liu, F., Yang, D., Cai, J., Yu, L., Zhu, Z., Xu, D., Yuille, A., Roth, H., 2020. 3D semi-supervised learning with uncertainty-aware multi-view co-training. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3646–3655.
- Xia, Y., Yang, D., Yu, Z., Liu, F., Cai, J., Yu, L., Zhu, Z., Xu, D., Yuille, A., Roth, H., 2020. Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. *Med. Image Anal.* 65, 101766.
- Xu, X., Lian, C., Wang, S., Wang, A., Royce, T., Chen, R., Lian, J., Shen, D., 2020. Asymmetrical multi-task attention u-net for the segmentation of prostate bed in CT image. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 470–479.
- Xu, X., Lian, C., Wang, S., Zhu, T., Chen, R.C., Wang, A.Z., Royce, T.J., Yap, P.-T., Shen, D., Lian, J., 2021. Asymmetric multi-task attention network for prostate bed segmentation in computed tomography images. *Med. Image Anal.* 72, 102116.
- Xu, X., Zhou, F., Liu, B., 2018. Automatic bladder segmentation from CT images using deep CNN and 3D fully connected CRF-RNN. *Int J Comput Assist Radiol Surg* 13 (7), 967–975.
- Yan, P., Xu, S., Turkbey, B., Kruecker, J., 2010. Discrete deformable model guided by partial active shape model for TRUS image segmentation. *IEEE Trans. Biomed. Eng.* 57 (5), 1158–1166.
- Yan, P., Xu, S., Turkbey, B., Kruecker, J., 2011. Adaptively learning local shape statistics for prostate segmentation in ultrasound. *IEEE Trans. Biomed. Eng.* 58 (3), 633–641.
- Yang, X., Yu, L., Wu, L., Wang, Y., Ni, D., Qin, J., Heng, P.-A., 2017. Fine-grained recurrent neural networks for automatic prostate segmentation in ultrasound images. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31, pp. 1633–1639.
- Yu, L., Wang, S., Li, X., Fu, C.-W., Heng, P.-A., 2019. Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, pp. 605–613.
- Zahoor, M.N., Fraz, M.M., 2017. Fast optic disc segmentation in retina using polar transform. *IEEE Access* 5, 12293–12300.
- Zhu, Q., Du, B., Yan, P., 2019. Boundary-weighted domain adaptive neural network for prostate MR image segmentation. *IEEE Trans. Med. Imaging* 39 (3), 753–763.