

Boundary Coding Representation for Organ Segmentation in Prostate Cancer Radiotherapy

Shuai Wang, Mingxia Liu, *Senior Member, IEEE*, Jun Lian, and Dinggang Shen, *Fellow, IEEE*

Abstract—Accurate segmentation of the prostate and organs at risk (OARs, e.g., bladder and rectum) in male pelvic CT images is a critical step for prostate cancer radiotherapy. Unfortunately, the unclear organ boundary and large shape variation make the segmentation task very challenging. Previous studies usually used representations defined directly on *unclear* boundaries as context information to guide segmentation. Those boundary representations may not be so discriminative, resulting in limited performance improvement. To this end, we propose a novel boundary coding network (BCnet) to learn a discriminative representation for organ boundary and use it as the context information to guide the segmentation. Specifically, we design a two-stage learning strategy in the proposed BCnet: 1) Boundary coding representation learning. Two sub-networks under the supervision of the dilation and erosion masks transformed from the manually delineated organ mask are first separately trained to learn the spatial-semantic context near the organ boundary. Then we encode the organ boundary based on the predictions of these two sub-networks and design a multi-atlas based refinement strategy by transferring the knowledge from training data to inference. 2) Organ segmentation. The boundary coding representation as context information, in addition to the image patches, are used to train the final segmentation network. Experimental results on a large and diverse male pelvic CT dataset show that our method achieves superior performance compared with several state-of-the-art methods.

Index Terms—Image Segmentation, Fully Convolutional Network, Boundary Representation, Prostate, CT Image

I. INTRODUCTION

PROSTATE diseases, such as prostatitis, enlarged prostate, and prostate cancer, are common diseases in adult males [1], [2]. With age, men are more susceptible to prostate diseases, among which prostate cancer is the most harmful one to men’s health. Nowadays, external beam radiation therapy (EBRT) is one of the most commonly applied treatments for prostate cancer, in which high-energy radiation beams are delivered to destroy cancer cells while sparing surrounding normal tissues [3]. To guarantee the radiotherapy effect, accurate delineation of the prostate and organs at risk (OARs, e.g., bladder and rectum) in CT images is critical in EBRT.

While manual delineation is laborious and often suffers large inter-operator variability, many studies on automated

This work was supported in part by the NIH under Grant CA206100 (Corresponding authors: J. Lian and D. Shen).

S. Wang, M. Liu, and D. Shen are with the Department of Radiology and Biomedical Research Imaging Center, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA (E-mail: shuaiwang.tai@gmail.com; mxliu@med.unc.edu; Dinggang.Shen@gmail.com). Dr. Shen is also with Department of Artificial Intelligence, Korea University, Seoul 02841, Republic of Korea. S. Wang and M. Liu contribute equally to this work.

J. Lian is with the Department of Radiation Oncology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA (E-mail: jun_lian@med.unc.edu).

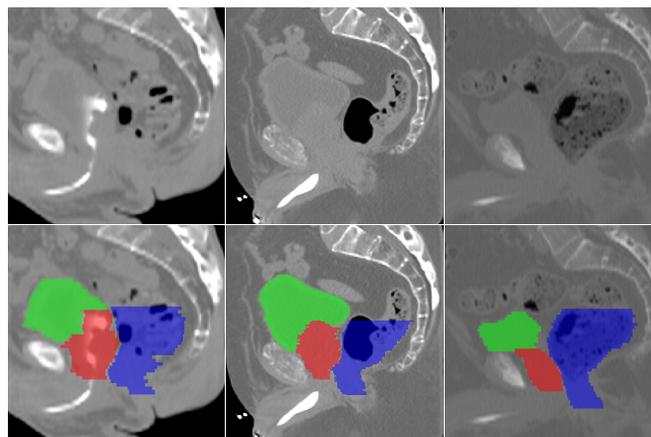


Fig. 1: Exemplars of male pelvic CT images. The first row shows the sagittal slices of three subjects. The second row denotes the ground truth segmentation results for three types of organs, with the prostate, bladder, and rectum represented by red, green, and blue, respectively.

segmentation of male pelvic organ have been proposed to assist clinicians in various clinical applications [4]–[10]. Nevertheless, accurate segmentation of the male pelvic organ in CT images is still an open and challenging task, due to the unclear organ boundary caused by the low contrast of CT images (as shown in Fig. 1). Especially when two organs are adjacent to each other, it is difficult to distinguish the boundary of the organ under consideration from its neighboring organs. To address this problem, some existing approaches use diverse boundary representations as context information to guide the organ segmentation, such as auto-context [11], [12], landmark [13], and curve [14]. However, these boundary representations are generally defined on *unclear* boundaries, which may not be discriminative for organ segmentation and make the learned model easily misled by regions with similar characteristics to boundaries.

In this paper, we propose a novel boundary coding network (BCnet) to address the challenging problem of segmenting the prostate and OARs in male pelvic CT images. To deal with unclear organ boundary and large shape variation, we propose a representation method that encodes the organ boundary based on the learned dilation and erosion characteristics, which provides discriminative spatial-semantic context information to help distinguish the organ from the background. Specifically, the proposed boundary coding representation can not only indicate the location of the boundary but also encode

the morphological relationship of the boundary. Moreover, to improve the robustness of the proposed boundary coding representation, a multi-atlas based refinement method is further designed by transferring the knowledge learned in the training phase to the inference phase. Besides original image patches, the learned boundary coding representation (treated as context information) is also used as input for the final segmentation. The paper is extended from our conference paper [2] and has been improved from both the method and the experiments. Specifically, a more simplified coding network is proposed, and a multi-atlas based refinement is further to improve the robustness. Moreover, instead of only performing the segmentation of the prostate under the 2D architecture, we also verify the effectiveness on the segmentation of the bladder and rectum using both 2D and 3D architectures.

The main contributions of this work can be summarized as follows. *First*, a boundary coding representation method is proposed to address the unclear organ boundary of male pelvic organs in CT images. Different from conventional methods that are directly defined on the unclear boundary, the proposed representation contains more discriminative spatial-semantic context information. *Second*, to perform accurate segmentation of the prostate and OARs in CT images, we design a two-stage learning strategy, where the learned boundary coding representation is used as the context information to guide the organ segmentation. *Finally*, comprehensive experiments have been performed on 313 pelvic CT planning images collected from male patients with prostate cancers, with results demonstrating the superiority of our method over several state-of-the-art methods.

II. RELATED WORK

Automated segmentation of prostate and OARs has been studied for many years. Existing methods can be roughly divided into three categories: multi-atlas based methods, deformable model based methods, and learning based methods, where outperforms the former two types of approaches. A detailed review can be found in [8], [15]. Here, we overview learning based methods that are most relevant to our work and especially the current state-of-the-art deep learning methods.

A. Conventional Learning based Segmentation

Compared with multi-atlas based methods and deformable model based methods, conventional learning based methods are relatively more efficient and robust [8], [15]. Theoretically, conventional learning based methods have two key components to build accurate models: (a) extraction of hand-crafted features to represent target organs, and (b) classification/regression model for segmentation.

Currently, much effort has been devoted to the feature extraction component. For example, Feng et al. [16] combined gradient and probability distribution function features to guide the prostate segmentation in CT images. Yang et al. [17] introduced the discriminant analysis to make the scale and variance of the learned SIFT features adaptive to the locations. Duan et al. [18] used volume based features to detect the bladder wall. Garapati et al. [19] first extracted

a 91-dimensional representation (26 morphological features and 65 texture features) to segment bladder lesions and then performed classification. Xie et al. [20] propagated rectum contour from planning CT to cone-beam CT using a feature based registration method. Chen et al. [21] used a feature selection method to select the most useful dimensions from the whole feature set to improve the segmentation robustness. For the segmentation component, various classification/regression models have been developed. Glocker et al. [22] proposed a forest model that jointly performs the pixel classification and shape regression. In [13], Lay et al. integrated local and global image context information to detect landmarks of the target organ and then deformed the organ shape using discriminative boundary detectors. Gao et al. [15] jointly learned the displacement regressor and organ classifier via multi-task random forests to perform the segmentation of the prostate, bladder, rectum, left and right femoral heads. Lu et al. [23] combined Jensen-Shannon divergence with a learning based boundary detector for the segmentation of the prostate, bladder, and rectum. Besides population knowledge learned from the whole data, several studies also introduced patient-specific information into the segmentation to improve the performance. Wu et al. [24] constructed a feature dictionary based on the daily CT scans that have been manually segmented and then automatically updated the dictionary using the segmentation results of new treatment scans. Ma et al. [25] trained a population model using the whole set and a patient-specific model using the data from a specific patient. These two models were combined to output the probability of a voxel belonging to the target organ. Since these methods generally use low-dimensional hand-crafted features extracted from low-level cues, their performance may be limited, especially when the organ boundaries in male pelvic CT images are unclear.

B. Deep Learning based Segmentation

Recently, deep learning has drawn much attention in many fields, such as computer vision, natural language processing, and computer-aided disease diagnosis. And many deep learning models have been developed for automated segmentation of medical images. Compared with conventional learning methods using hand-crafted features and shallow models, deep learning methods can automatically learn hierarchical feature representations from the input image itself and make the prediction in an end-to-end manner. Anas et al. [26] employed residual neural nets (ResNets) and dilated convolution at deeper layers to delineate the 2D ultrasound images for low dose-rate prostate brachytherapy of early-stage prostate cancer treatment. Brosch et al. [27] formulated the boundary detection problem as a regression task and used a convolutional neural network (CNN) to localize the organ boundary. Liang et al. [28] introduced the pyramid module and attention mechanism into ResNet-101 [29] to perform the segmentation of 50 pelvic organs in MRI scans. Wang et al. [8] proposed a boundary-sensitive multi-task network to jointly segment the prostate, bladder, and rectum in CT images. Anas et al. [30] presented a real-time prostate segmentation technique during the biopsy procedure based on recurrent

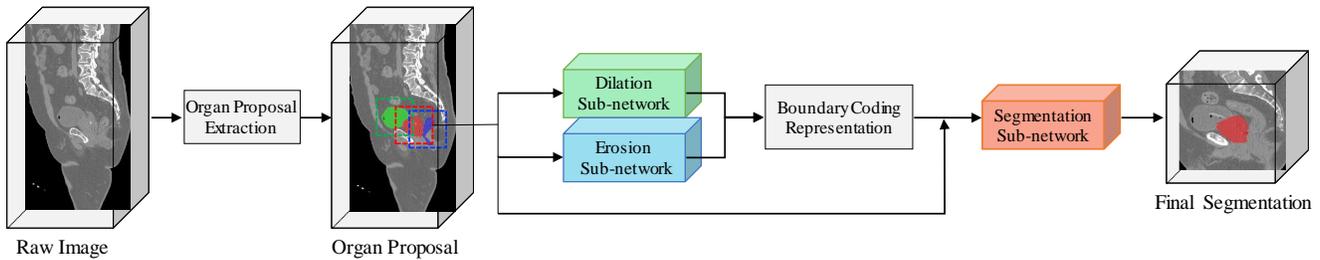


Fig. 2: Pipeline of the proposed boundary coding representation for organ segmentation in CT images, with three major components: (1) organ proposal extraction, (2) boundary coding representation learning via a dilation sub-network and an erosion sub-network, and (3) a segmentation sub-network. Note that the organ proposal is extracted for each organ under consideration and only the prostate is taken as an example in the final segmentation.

neural networks (RNNs). Léger et al. [31] trained a 3D U-net for segmenting the prostate, bladder, and rectum on cone beam CT by augmenting the training data with CT scans.

Even though the existing deep learning approaches have shown remarkable performance compared with conventional learning methods, they usually focus on using the image intensity features, without considering the fact that even the organ boundary is generally unclear but it contains rich contextual information (e.g., the relationship between nearby voxels). Therefore, it should be useful to explicitly capture this contextual information conveyed in the organ boundaries and take advantage of this knowledge in the segmentation model for boosting the learning performance.

III. METHOD

To achieve accurate and automatic segmentation of the prostate and OARs in CT images, we propose a boundary coding network (BCnet), with the whole pipeline illustrated in Fig. 2. Our proposed BCnet is developed for joint feature representation learning and organ segmentation, consisting of three components, i.e., (1) organ proposal extraction, (2) boundary coding representation learning via a dilation sub-network and an erosion sub-network, and (3) a segmentation sub-network. For a given raw CT image, we first perform organ proposal extraction to remove the uninformative background. Specifically, a multi-class coarse segmentation network is trained to detect the rough location of each target organ, and then sufficiently large organ proposals are extracted from the original image based on the rough centroid of each organ. Then, the dilation and erosion sub-networks are used to learn the spatial-semantic context information near the organ boundary, supervised by the morphological dilation and erosion masks transformed from the manually delineated organ mask, respectively. Using the generated predicted dilation and erosion masks, in the boundary coding representation learning module, we encode the boundary and design a multi-atlas based refinement method to improve the coding robustness by transferring the training knowledge to the inference stage. Finally, we train the segmentation sub-network using the boundary coding representation (as context information) and the proposal patches as input data. The details of each component are elaborated as follows.

A. Extraction of Organ Proposals

For the purpose of dose calculation of radiation therapy, the original pelvic CT images are scanned with the larger field-of-view including target organs (i.e., prostate, bladder, and rectum) and skin. Therefore, it is not reasonable to directly perform organ segmentation on the original images, because the uninformative/noisy background would degrade the segmentation performance and also increase the computational complexity. To address this problem, following [14], we propose to first train a coarse detection network to roughly locate each target organ in the original CT image. In our implementation, since there are three target organs to be segmented, we train a four-class (prostate, bladder, rectum, and background) segmentation network under the supervision of the manually delineated organ mask (with ground truth segmentation of three organs). Based on the predicted results, the rough centroid of each target organ can be determined. And then, a sufficiently large proposal is extracted from the original image by taking the organ centroid as the cube center. Here, the size of each organ proposal is set as $160 \times 160 \times 160$, which ensures complete coverage of each target organ.

B. Learning of Boundary Coding Representation

In this section, we give details about how to learn the discriminative representation of organ boundaries using our proposed method. To achieve this goal, the dilation and erosion sub-networks are first trained under the supervision of the morphological dilation and erosion masks transformed from the manually delineated organ mask (denoting as ground truth in the following), respectively, as shown in Fig. 3. Then, the predicted dilation and erosion masks are used to encode discriminative spatial-semantic information for the boundary. Also, to improve the robustness of the proposed boundary coding representation, a multi-atlas based refinement is designed to refine the representation by transferring the knowledge learned from the training data to the inference stage.

1) *Supervision Mask Generation*: For accurate organ segmentation, the boundaries between different organs actually provide very important clues. Especially for pelvic CT images with low contrast in terms of image intensity, if we can represent and highlight unclear organ boundaries, it will be very helpful to delineate organs. While the representation

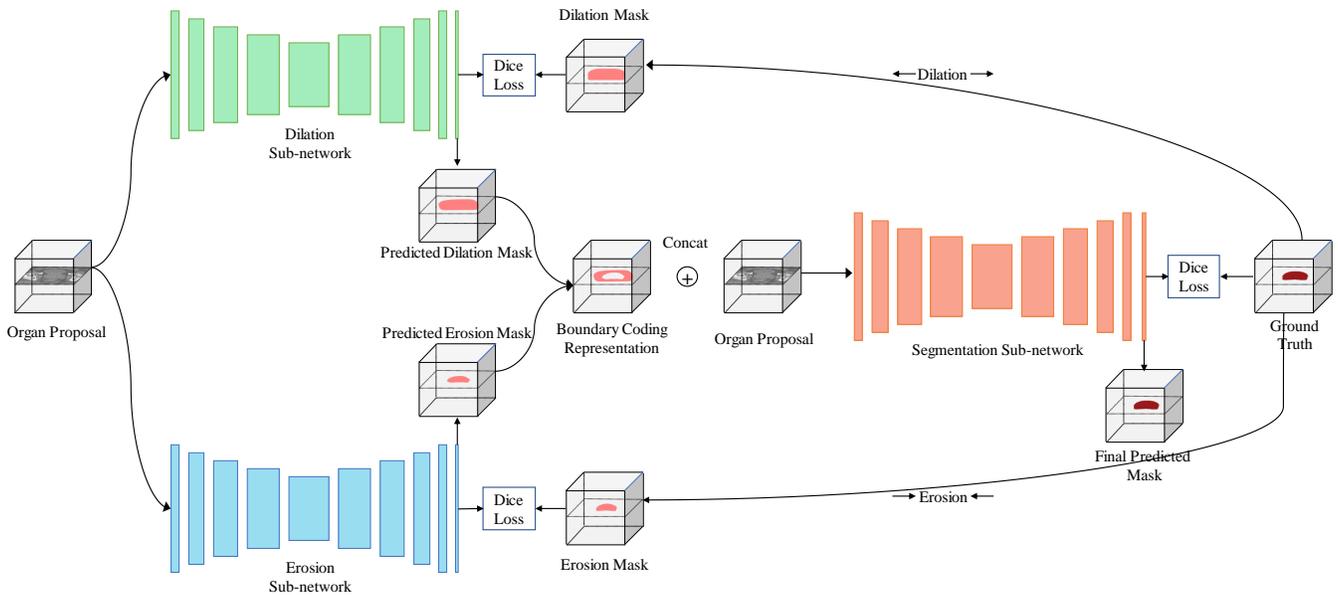


Fig. 3: The architecture of the proposed network for learning boundary coding representation and organ segmentation (with prostate as an example), which contains two stages and three sub-networks. The dilation and erosion sub-networks are trained under the supervision of the morphological dilation and erosion masks transformed from the ground truth, respectively. The predicted masks of the dilation and erosion sub-networks are used to encode the spatial-semantic information of organ boundaries, yielding the boundary coding representation. Then the boundary coding representation, as well as the organ proposals, are input into the segmentation sub-network to achieve the final segmentation. In the implementation, any FCN-like architecture can be used as the backbones for the dilation, erosion, and segmentation sub-networks.

defined directly on the boundary may lack discrimination capability due to the fact that the boundary itself is unclear, we will utilize the *context information* around the boundary instead of focusing only on the boundary. In this study, two sub-networks, i.e., the dilation sub-network and the erosion sub-network, are trained to learn the structured dilation and erosion masks, respectively. We perform the dilation and erosion individually for each of multiple organs individually. These masks indicate the spatial-semantic context information of the boundary, namely, the true organ boundary between the predicted dilation and erosion masks and the morphological relationship between them (a structured distance related to the morphological kernel size K when learning).

To obtain the supervision masks for training these two sub-networks, we perform the 2D binary morphological operation along the axial slice (an imaginary plane that divides the body into superior and inferior parts) of the ground truth mask. The reason for not performing a 3D operation on the whole volume or 2D operation along other plane slices (sagittal and coronal) is that the shape of the organ on the axial slice is more regular and smooth. Let G_{n_i} denote ground truth of the i -th axial slice of the training image I_n . We can generate its corresponding dilation mask D_{n_i} and erosion mask E_{n_i} by performing $G_{n_i} \oplus S[K_D]$ and $G_{n_i} \ominus S[K_E]$, respectively, where \oplus and \ominus denote the binary morphological dilation and erosion operations, respectively. Here, $S[K_D]$ and $S[K_E]$ are two binary structuring elements, each with a morphological kernel (size: $K_D = 27$, $K_E = 11$) for dilation and erosion

operations, respectively. Note that we keep the smallest organ region if the morphological erosion completely erodes the organ region.

In this work, the resulting dilation mask and erosion mask are used as the supervision information for training the proposed dilation sub-network and the erosion sub-network, respectively. Considering the efficacy of U-net in segmentation tasks, we use both 2D U-net [32] and 3D U-net [33] as the backbones for the dilation and erosion sub-networks.

2) *Boundary Coding Representation Generation*: Under the supervision of dilation and erosion masks, the dilation and erosion sub-networks can be trained separately. Based on the predicted dilation and erosion masks (i.e., the outputs of dilation and erosion sub-networks), we can construct the boundary coding representation. Let \hat{D}_n and \hat{E}_n denote the predicted dilation and erosion masks of the n -th image. Our boundary coding representation is formulated as:

$$Z_n = \hat{D}_n \vee \hat{E}_n, \quad (1)$$

where \vee is the XOR operation.

Using Eq. 1, one can fuse the spatial-semantic context information of the boundary conveyed in both the dilation and erosion masks into the final boundary coding representation. In this way, contextual knowledge around the boundaries of the CT image can be explicitly captured. In Fig. 4, we visually illustrate an example of the proposed boundary coding representation. From this figure, one can observe that the boundary can be easily localized under the guidance of our

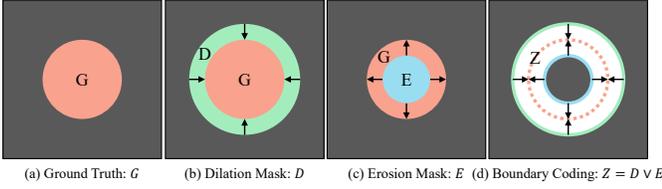


Fig. 4: An example for our proposed boundary coding representation. Here, the dilation mask D and the erosion mask E are generated based on the ground truth G . The proposed boundary coding representation $Z = D \vee E$ is used to capture the spatial-semantic information of organ boundaries.

proposed boundary coding representation Z , given the perfect dilation mask D and the erosion mask E .

3) *Boundary Coding Representation Refinement*: To improve the robustness of the proposed boundary coding representation, we propose to further refine Z_n , since not all \hat{D}_n and \hat{E}_n are perfect predictions. In our implementation, we will deal with the following two cases of \hat{D}_n and \hat{E}_n to refine the representation: (1) beyond case, in which some voxels outside an organ are incorrectly classified as belonging to the organ; and (2) missing case, in which no voxels in an organ slice are classified as the organ. In the training phase, we can simply reset them based on D_n and E_n .

In the inference phase, we adopt minimal user interaction to deal with the beyond case, and introduce a multi-atlas based method to deal with the missing case. Specifically, for the *beyond case*, the user is required to indicate the apex and base axial slices of the organ under consideration. In real clinical applications, the oncologist also needs to confirm the apex and base slices of the organ, and hence, our adopted minimal user interaction will not increase the burden of oncologists. For the *missing case*, we design a multi-atlas based method to refine/impute the missing prediction in the predicted dilation/erosion mask for a new unseen image based on the predictions of training images. Fig. 5 shows the proposed strategy for dealing with the missing case. As shown in Fig. 5, we first register the predicted masks (e.g., \hat{D}_n^T) of the training images to the predicted mask (e.g., \hat{D}^q) of the target image via just scaling and translation transformations. Then, an axial neighborhood window (denoted by a blue dotted rectangle) with H -slice centered at the slice with the missing prediction (denoted by a yellow line in \hat{D}^q) is extracted from the target predicted mask, and the registered images and the Dice Similarity Coefficient (DSC) value is computed between the extracted window and the window from each registered image. Finally, L -top similar registered images with the largest DSC values are selected, and the majority-voting slice of the corresponding slices from the selected registered images is transferred as the prediction of the target slice with the missing case. In our implementation, the values of H and L are set to 5 and 3, respectively. In this way, we can refine the predicted dilation and erosion maps with missing cases.

C. Organ Segmentation

To employ the learned boundary coding representation as the context information, we treat the concatenation of the organ proposal and boundary coding representation as the input of the proposed segmentation sub-network. Also, as shown in Fig. 3, the manual delineation mask (i.e., ground truth) is used as the supervision information of the segmentation sub-network. Specifically, in the implementation, we use both 2D U-net [32] and 3D U-net [33] as the backbones for the segmentation sub-network.

To make the learned boundary representation more robust, we adopt a two-stage learning strategy to train our BCnet. Specifically, in the first stage, we will train the dilation and erosion sub-networks to generate the dilation and erosion masks, respectively. And then based on the predicted masks, we build the boundary coding representation. In the second stage, the segmentation sub-network is trained by taking the proposal patches and the corresponding boundary representation as input data for segmenting three organs separately, i.e., prostate, bladder, and rectum.

Specifically, to train each sub-network, we use the Dice loss [34] instead of cross-entropy loss for optimization as

$$L = -\frac{2|P_n \cap Q_n|}{|P_n| + |Q_n|}, \quad (2)$$

where P_n indicates the ground truth (dilation/erosion/manual annotation mask) and Q_n is its corresponding prediction.

D. Implementation

At the *training* stage, the dilation and erosion sub-networks are first trained separately, in which the proposal patches (with size: $128 \times 128 \times 1$ for 2D networks and $96 \times 96 \times 64$ for 3D networks) are used as the input and the dilation and erosion masks transformed from the ground truth are used as the output. After that, the multi-atlas based refinement is applied on the predicted dilation and erosion masks and then the boundary coding representation is formed based on Eq. (1). Finally, the proposal patches with their corresponding boundary coding representation masks are used as the input to train the segmentation sub-network under the supervision of the ground truth.

At the *inference* stage, the predicted dilation and erosion masks are first obtained by inputting the proposal patches into the trained dilation and erosion sub-networks. Then, the boundary coding representation is generated based on the predicted dilation and erosion masks after applying the multi-atlas based refinement. Finally, the proposal patches with the boundary coding representation are input into the trained segmentation sub-network to get the segmentation results.

IV. EXPERIMENTS

A. Dataset

The dataset used in our experiments was collected from the North Carolina Cancer Hospital, consisting of 313 CT planning images scanned from 313 patients with prostate cancer. Compared with some pelvic datasets that each patient

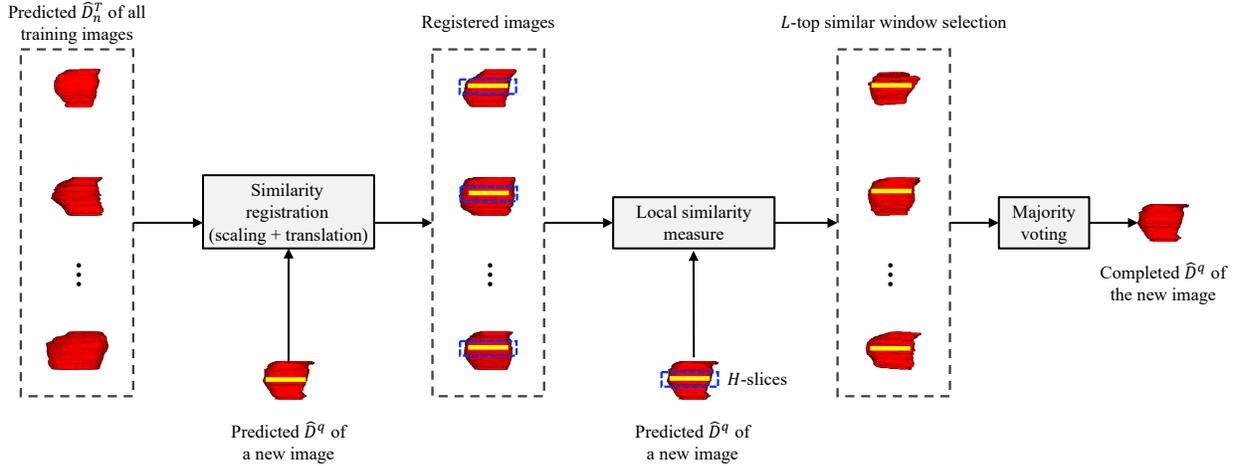


Fig. 5: The pipeline of our method to impute the missing prediction in the predicted dilation/erosion mask for a new unseen image. Here, we take the predicted dilation masks of training images (i.e., \hat{D}_n^T) and that of a new image (i.e., \hat{D}^q) as an example. The yellow line indicates the location of one slice with the missing prediction, while the blue dotted rectangles denote the neighborhood windows.

TABLE I: Segmentation results for three male pelvic organs achieved by eight different methods, in terms of DSC (%), ASD (mm), PPV (%), and SEN (%). The best results are marked in red, and the second-best results are marked in blue. Asterisk indicates whether our best result (2D/3D) is better than all other compared methods with statistical significance ($p < 0.05$).

Metrics		DenseNet [35]	DeepLabv3+ [36]	V-net [34]	2D U-net [32]	3D U-net [33]	nnU-net [37]	2D Ours	3D Ours
Prostate	DSC	81.91±3.28	87.36±1.30	87.20±1.52	87.97±1.40	86.29±1.84	88.06±1.83	89.68±0.78*	89.48±0.87
	ASD	2.42±0.70	1.53±0.19	1.65±0.23	1.50±0.22	1.79±0.31	1.48±0.29	1.22±0.09*	1.29±0.11
	PPV	87.53±1.37	89.12±1.22	90.76±1.16	89.19±1.09	90.38±1.20	89.71±1.50	88.87±1.13	89.87±1.10
	SEN	80.41±4.33	86.76±2.29	85.21±2.50	87.82±2.30	84.19±2.90	88.80±2.72	90.67±1.44*	89.72±1.62
Bladder	DSC	91.50±1.63	93.80±1.06	93.24±1.06	94.06±1.10	92.15±2.69	93.74±1.22	94.08±1.03	93.81±1.02
	ASD	1.41±0.20	1.07±0.18	1.23±0.17	1.08±0.24	3.63±4.78	1.10±0.21	1.04±0.19	1.09±0.17
	PPV	95.34±1.22	94.60±1.33	93.68±1.07	94.49±1.33	94.62±1.12	94.12±1.61	94.61±1.30	95.09±1.38
	SEN	88.52±2.29	93.38±1.29	93.18±1.58	93.96±1.29	91.45±2.94	94.13±1.21	93.88±1.22	92.90±1.17
Rectum	DSC	82.75±2.50	87.31±1.32	86.35±1.62	88.85±1.60	88.61±1.20	88.87±1.21	92.22±0.51*	92.08±0.49
	ASD	2.71±0.59	1.72±0.29	1.97±0.33	1.62±0.35	1.66±0.31	1.44±0.20	0.91±0.06*	0.94±0.05
	PPV	93.86±0.69	92.90±0.86	91.03±0.84	93.49±0.80	92.17±0.89	92.52±0.94	93.06±0.79	92.94±0.81
	SEN	75.76±3.60	82.99±2.17	83.16±2.66	85.41±2.45	85.85±2.00	86.13±2.02	91.57±0.82*	91.41±0.79

has one planning and sequence treatment images [38], [39], our adopted dataset is more diverse and challenging since the difference among different unique images/patients is generally greater. Two experienced radiologists helped to delineate the prostate, bladder, and rectum in each CT image. The final ground truth is confirmed after careful checking for the inconsistent regions. While these images were scanned using different configurations, they have different in-plane resolutions and inter-slice thicknesses. Specifically, the in-plane resolutions and inter-slice thicknesses of the original images range from 0.932 mm to 1.365 mm and 1 mm to 3 mm, respectively. To process the data on the same scale, we resampled the intensity images and also their corresponding ground truth to the same resolution of $1 \times 1 \times 1 \text{ mm}^3$. After resampling, the average image size is $583 \times 583 \times 354$. The prostate, bladder, and rectum are with sizes of $35 \sim 84 \times 30 \sim 86 \times 30 \sim 79$, $48 \sim 143 \times 48 \sim 131 \times 27 \sim 139$, and $24 \sim 91 \times 35 \sim 134 \times 45 \sim 156$, respectively. Moreover, the min-max normalization was performed on each image to normalize the intensity to the range

of $[0, 1]$. Specifically, the intensity value of each voxel first subtracts the minimum value of the whole image and then is divided by the maximum value of the whole image after subtracting the minimum value.

B. Evaluation Metrics

Four metrics widely used in medical image segmentation are computed between the automated segmentation results and ground truth for performance evaluation, i.e., DSC, Average Symmetric Surface Distance (ASD), Positive Predictive Value (PPV), and Sensitivity (SEN) [40]. For DSC, PPV, and SEN, larger values indicate better performance, while for ASD vice versa. Three-fold cross-validation is performed in the experiments. Both the mean values and the 95% confidence intervals of all evaluation metrics are reported for each method.

C. Comparison Methods

Based on the input, CNN-like networks can be divided into two groups with 2D and 3D architectures, respectively.

Compared with 2D networks, 3D networks introduce more spatial information into the learning process. However, for different tasks, the performance of 3D networks may not be better than that of 2D networks, which sometimes depends on the characteristics of the to-be-analyzed data. Therefore, in the implementation, we compare our BCnet with both 2D and 3D methods. For 2D case, DenseNet [35], 2D U-net [32], and DeepLabv3+ [36] are compared with the proposed 2D version of BCnet (denoted as “2D Ours”). For 3D case, 3D U-net [33] and V-net [34] are compared with the 3D version of BCnet (denoted as “3D Ours”). Moreover, nnU-net [37] which adaptively deploys itself to any given dataset is also compared. Specifically, 2D U-net, 3D full resolution U-net, and 3D cascaded U-net are first trained. Then the best U-net configuration is identified to use for test prediction.

Both our proposed methods and all comparison methods (except nnU-net [37]) are built and trained based on an open-source package Keras [41], using TensorFlow [42] as the backend. For nnU-net [37], we use the implementation proposed in [37] and no configuration needs to be manually set. Since our proposed boundary coding representation can be implemented based on any architectures of fully convolutional networks (FCNs), we take 2D U-net [32] and 3D U-net [33] as the backbones to implement the proposed three sub-networks (i.e., for dilation, erosion, and segmentation) in BCnet, respectively. For a fair comparison, the adopted 2D U-net and 3D U-net backbones use the same configurations as the reported architectures in [32] and [33]. The main difference is that the Dice loss [34] is used as the objective function for all methods. We use “Adam” optimizer for stochastic optimization, and set the initial learning rate as 10^{-3} and decrease it by a factor 0.1 after 5 epochs until the performance on the validation dataset has no further improvement. For each method, 100 epochs are run for each fold and the model with the best performance on the validation data is used for performance comparison.

At the training phase, overlapping patches with sizes of $128 \times 128 \times 1$ and $96 \times 96 \times 64$ randomly cropped from the organ proposals are used as the inputs of 2D and 3D methods, respectively. To augment the training data, left-right flip on the axial plane and three random rotations within 15 degrees are performed. At the inference phase, a window with the same size as the training phase is moved across the full image by a stride of $32 \times 32 \times 1$ for 2D methods and $16 \times 16 \times 16$ for 3D methods, and passed through the trained models. The average results are used as the final prediction. For a fair comparison, the metrics of all methods are reported after performing the minimal user interaction we use.

D. Comparison with State-of-the-Art Approaches

Table I reports the mean results with 95% confidence intervals of eight different methods with respect to four evaluation metrics. Based on different backbone architectures, we can conduct two different comparisons between our methods and the comparison methods, with details given below.

1) *Ours vs. Methods with Different Backbones:* First of all, we compare our methods with DenseNet [35], DeepLabv3+ [36], V-net [34], and nnU-net [37], while our

implementations adopt U-net-like backbones that are different from them. From Table I, we can see that our method achieves better performance than these methods in terms of most evaluation metrics, especially in DSC and ASD. For prostate segmentation, DenseNet [35] obtains poor results whose DSC value is only 81.91% and ASD value is 2.42 mm, which demonstrates the challenge of the prostate segmentation. DeepLabv3+ [36] and V-net [34] achieve similar performance and are with great improvement compared with DenseNet [35]. nnU-net [37] achieves slightly better results than DeepLabv3+ [36] and V-net [34]. The best result is achieved by the 2D implementation of our method, which improves the DSC value by 1.62% from the best result of the compared methods. For bladder segmentation, the performance of all methods is better than prostate segmentation, which is mainly due to the wall structure of the bladder. Nonetheless, our implementations still can make an improvement. For rectum segmentation, the improvement is particularly significant. The DSC value is improved by at least 3.21% compared with the best result of the compared methods (achieved by nnU-net [37]), while the ASD value is reduced by 0.50 mm.

2) *Ours vs. Methods with Same Backbones:* In this work, we implement our method using 2D U-net [32] and 3D U-net [33] as the backbones, respectively. Hence, we directly compare two versions of our methods with the original 2D U-net [32] and 3D U-net [33]. As shown in Table I, 2D U-net [32] and 3D U-net [33] achieve stable performance. But after introducing our proposed boundary coding representation, the DSC values of 2D U-net [32] and 3D U-net [33] are improved from 87.97% to 89.68% and 86.29% to 89.48% for prostate segmentation, respectively. For bladder segmentation, the improvement is not very significant, which may be because the regular wall structure of the bladder has provided similar context information as our boundary coding representation. For rectum segmentation, the model capability benefits hugely from our proposed boundary coding strategy, i.e., at least 3.00% improvement in terms of DSC and 0.7 mm reduction in terms of ASD. We also found that the 2D implementation produces better results than the 3D implementation, for both our method and the original U-net, which is caused by two possible reasons. 1) The sample size for training the 2D network is larger than that of the 3D network, because more 2D slices are available for sampling compared with 3D volumes. 2) Due to memory limitations, the in-plane patch size used by 3D networks (96×96) is a bit smaller than 2D networks (128×128), which limits the receptive field size and the learning of global information.

Besides the quantitative results, Fig. 6 and Fig. 7 show several 3D segmentation results obtained by methods with different and the same backbones as ours, respectively. It is obvious that our proposed methods can achieve more consistent results with ground truth, compared with six competing methods. These visualization results further verify the effectiveness of our method.

E. Comparison with Auto-Context Approach

Compared with conventional U-net architecture, our proposed method first encodes the boundary information and then

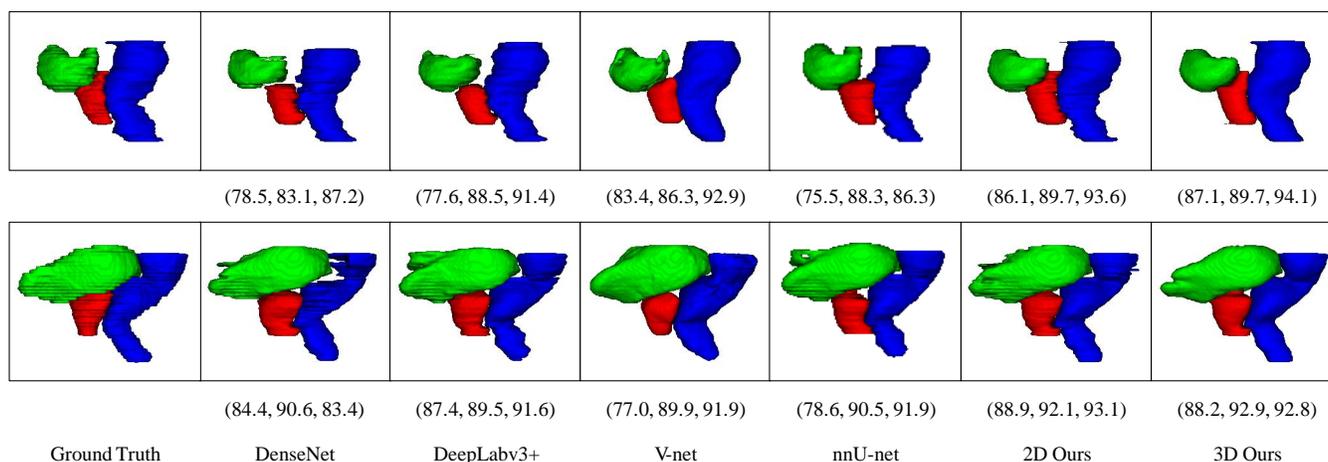


Fig. 6: 3D visualization of the segmentation results of the prostate, bladder, and rectum obtained by methods with different backbones as ours. Each row corresponds to a representative case. The triple array at the bottom of each image indicates the DSC values of the prostate, bladder, and rectum, respectively. Red, green, and blue denote the segmentation results of the prostate, bladder, and rectum, respectively.

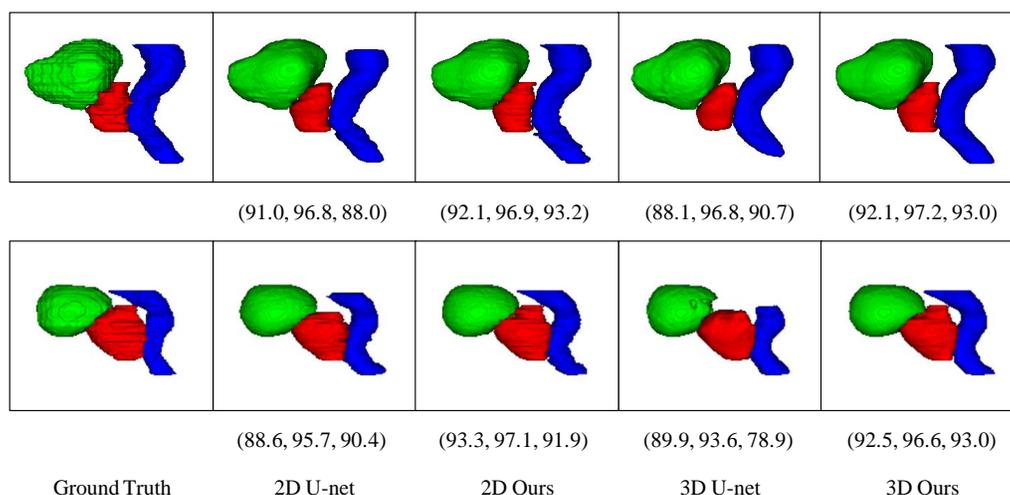


Fig. 7: 3D visualization of the segmentation results of the prostate, bladder, and rectum obtained by methods with the same backbones as ours. Each row corresponds to a representative case. The triple array at the bottom of each image indicates the DSC values of the prostate, bladder, and rectum, respectively. Red, green, and blue denote the segmentation results of the prostate, bladder, and rectum, respectively.

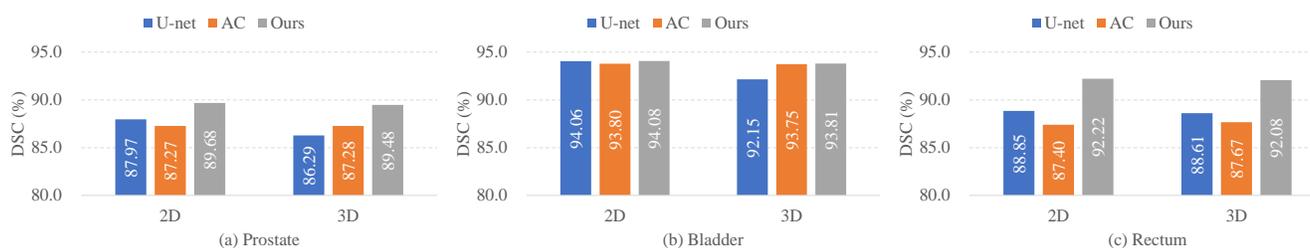


Fig. 8: DSC values achieved by the original U-net, auto-context approach (AC), and ours in terms of both 2D and 3D implementations. The number in the middle of each bar indicates the DSC value achieved by the corresponding method.

combines the encoded context with proposal patches as the input for segmentation. To better demonstrate the effectiveness of our boundary coding representation, we further compare our method with an auto-context (AC) approach. In the AC

method, a segmenter is first learned based on local image patches, and then the predicted masks generated by the learned segmenter are then used as the context information. That is, such predicted masks and the proposal patches are finally fused

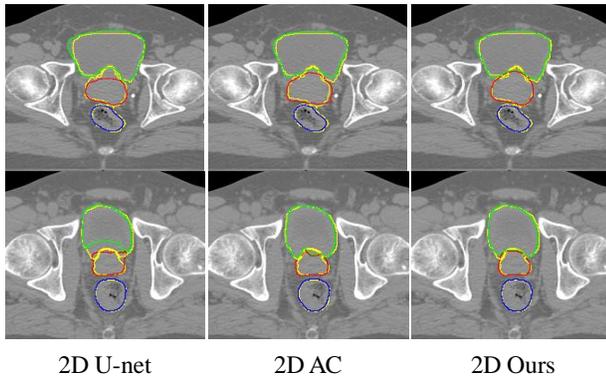


Fig. 9: Comparisons between the ground truth (yellow) and automated segmentation of the original U-net, auto-context approach (AC), and ours (prostate: red, bladder: green, rectum: blue) in terms of the 2D implementation.

to train the final segmentation model [43]. That is, our method and AC make use of different types of context information for segmentation. For the fair comparison, AC also adopts U-net architecture as the backbone and both 2D and 3D versions have been implemented.

The DSC values obtained by the original U-net, AC, and our method with 2D and 3D architectures are shown in Fig. 8. From this figure, one can have the following observations. First, our proposed method based on boundary coding representation generally achieves relatively better results for the segmentation of three organs, compared with U-net and AC. Also, the AC method does not work well (especially for the rectum segmentation) and the original U-net is inferior to our method in most cases. These results imply that the boundary coding representation could effectively boost the segmentation performance, by providing rich spatial-semantic information for the boundaries.

In Figs. 9-10, we further illustrate several representative comparisons between the ground truth and automated segmentation of the original U-net, AC, and our method, respectively. As can be observed, our proposed methods can facilitate the segmentation and delineate a more accurate organ boundary. For AC, the extracted context information may not provide useful guidance for segmentation because it is very challenging to directly encode the unclear boundary.

V. DISCUSSION

A. Comparison with Previous Studies

We now compare our method with two recently published studies [8], [15] on the same dataset and two studies [39], [44] on different datasets, with results shown in Table II.

As can be seen from Table II, our proposed method consistently outperforms these two methods on the same dataset in terms of both DSC and ASD metrics for all three organs. For instance, for rectum segmentation, our method achieves an improvement of 3.22% in terms of DSC, compared with the second-best result achieved by [8]. These results further validate the effectiveness of the proposed boundary coding

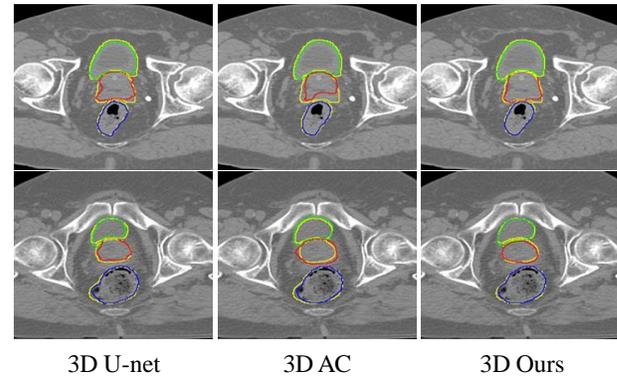


Fig. 10: Comparisons between the ground truth (yellow) and automatic segmentation of the original U-net, auto-context approach (AC), and ours (prostate: red, bladder: green, rectum: blue) in terms of the 3D implementation.

TABLE II: Comparison with previous studies in terms of DSC (%) and ASD (*mm*). Asterisk indicates whether our result is better than all other compared methods on the same dataset with statistical significance ($p < 0.05$).

Metrics		Different datasets		Same dataset		
		Qin [44]	Elmahdy [39]	Gao [15]	Wang [8]	2D Ours
Prostate	DSC	76.66	85.00	86.60	89.00	89.68
	ASD	N/A	1.13	1.77	1.32	1.22*
Bladder	DSC	94.49	86.00	92.10	94.00	94.08
	ASD	N/A	2.29	1.37	1.15	1.04
Rectum	DSC	79.80	81.00	88.40	89.00	92.22*
	ASD	N/A	2.21	1.38	1.53	0.91*

representation learning strategy for boosting the segmentation performance.

Qin [44] and Elmahdy [39] are two deep learning based methods but using different datasets smaller than our used. The results show that our method achieves the best performance, except that the ASD value of our method in the prostate segmentation is a little worse than Elmahdy [39]. The poor results of Qin [44] in the segmentation of the prostate and rectum and Elmahdy [39] in the segmentation of the rectum demonstrate the challenging of the CT pelvic organ segmentation. So, our excellent performance in all three organs verifies the robustness of our method.

B. Analysis of Boundary Coding Representation

To encode the organ boundary, we design a dilation sub-network and an erosion sub-network to learn the dilation and erosion masks. The comparison between our method and other state-of-the-art methods has verified the effectiveness of the boundary coding representation. Here, we further investigate such representation conveyed in the predicted dilation and erosion masks. Specifically, we compute the predictive value (PPV) and the Sensitivity (SEN) value by comparing the predicted dilation and erosion masks with ground truth, respectively. Note that a large PPV value indicates that the

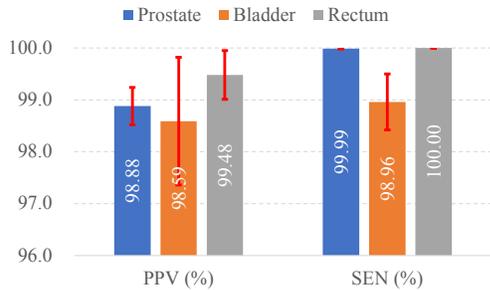


Fig. 11: PPV and SEN values by comparing the predicted dilation and erosion masks (generated by our dilation and erosion sub-networks, respectively) with ground truth.

TABLE III: Comparison between our implementations without (w/o) and with (w) boundary coding representation refinement, in terms of DSC (%) and ASD (mm) values.

Method	Prostate		Bladder		Rectum		
	2D	3D	2D	3D	2D	3D	
w/o	DSC	88.17	88.43	94.01	93.28	91.44	90.65
	ASD	1.41	1.38	1.16	1.37	1.14	1.52
w	DSC	89.68	89.48	94.08	93.81	92.22	92.08
	ASD	1.22	1.29	1.04	1.09	0.91	0.94

organ can be well covered by the prediction, while a large SEN value indicates that the prediction can be well covered by the organ. The experimental results for the segmentation of three organs are shown in Fig. 11. From Fig. 11, we can see that the yielded PPV and SEN values are very high (i.e., $> 98\%$), suggesting that the proposed dilation and erosion sub-networks can effectively capture the dilation and erosion characteristics, respectively.

C. Influence of Boundary Coding Representation Refinement

In the proposed method, a refinement strategy is performed on the learned bounding coding representation. To verify its effectiveness, we further compare our implementations without and with using the refinement strategy. The corresponding DSC results achieved by 2D and 3D architectures are shown in Table III. From Table III, in both 2D and 3D settings, our method with the refinement strategy consistently outperforms its variant without refinement. For instance, for segmenting rectum, our method (3D version) with and without refinement yields the DSC values of 92.08 and 90.65, respectively. These results suggest that the refinement strategy helps boost the performance of our learned boundary coding representation in segmenting three organs in both 2D and 3D settings.

D. Analysis of Clinical Significance

In the previous subsections, we have conducted comprehensive comparisons to verify the effectiveness of our proposed method. Here, based on these compared results, we discuss the clinical significance of our method. In EBRT, the radiation dose gradient at the boundary of the prostate is intentionally optimized very sharp, about 10% of prescription dose per mm . The EBRT of the prostate is a difficult case since the part of the

bladder and rectum often abuts the prostate which is covered by a high therapeutic dose. The overall treatment accuracy of Intensity-Modulated Radiotherapy (a modern type of EBRT) recommended by the Task Group 142 of American Association of Physicists in Medicine (AAPM) is 1 mm , which accumulates the uncertainty of all factors [45]. Compared with the best ASD results of the competing methods, our BCnet improves the overall segmentation of the prostate, bladder, and rectum respectively by 0.26 mm , 0.03 mm , and 0.53 mm in Table I, and by 0.10 mm , 0.11 mm , and 0.47 mm in Table II. Even though the overall improvement is less than 1 mm in terms of ASD, the results are still with clinical significance. As shown in Figs. 9-10, our method achieves similar performance to compared methods in most areas, but our method significantly improves the delineation results of boundary regions that other methods cannot handle well. Boundary regions are the most difficult areas to spare OARs while providing sufficient coverage on the tumor. Therefore, using our method, the treatment effect for areas with better delineation will be improved.

E. Limitations and Future Work

Even though our proposed method has achieved excellent performance in CT male pelvic organ segmentation, there are still several technical issues to be considered in the future. First, in our current implementation, the boundary coding representation is formed and the segmentation network is trained for each organ separately since the dilation masks will be overlapped between different organs if performing multi-organ operations at the same time. And we found that our method achieves poor performance in some adjacent regions between different organs. This is mainly because if the shape distribution of adjacent organs is extremely irregular, the robustness of the proposed boundary coding representation may be weakened. To address this problem, we plan to design an adaptive coding method that can represent the overlapped regions using more discriminative representation and make the segmentation of all organs in a unified framework. Second, in our work, only 2D binary morphological operation along the axial slice is used for learning boundary coding representation. Actually, we can also apply the same operation along with the sagittal and coronal directions and fuse the results from three directions to further improve the performance. Finally, we would like to improve the framework to make it implement in an end-to-end manner while in the current framework, the refinement module must be achieved in an offline manner.

VI. CONCLUSION

In this work, we propose a novel fully convolutional network (i.e., BCnet) for improving male pelvic organ segmentation in CT images by learning a discriminative boundary coding representation. Our proposed method is directly based on the morphological dilation and erosion masks transformed from the manually delineated mask, which is easy to implement. Extensive experiments in terms of both 2D and 3D implementations on a diverse and large pelvic CT dataset demonstrate the superiority of our approach that improves the state-of-the-art segmentation results by large margins.

REFERENCES

- [1] D. Parikesit, C. A. Mochtar, R. Umbas, and A. R. A. H. Hamid, "The impact of obesity towards prostate diseases," *Prostate International*, vol. 4, no. 1, pp. 1–6, 2016.
- [2] S. Wang, K. He, D. Nie, S. Zhou, Y. Gao, and D. Shen, "Hierarchical representation for CT prostate segmentation," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 1501–1504.
- [3] A. V. D'Amico, R. Whittington, S. B. Malkowicz, D. Schultz, K. Blank, G. A. Broderick *et al.*, "Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer," *Jama*, vol. 280, no. 11, pp. 969–974, 1998.
- [4] S. S. Chandra, J. A. Dowling, K.-K. Shen, P. Raniga, J. P. Pluim, P. B. Greer *et al.*, "Patient specific prostate segmentation in 3-D magnetic resonance images," *IEEE Transactions on Medical Imaging*, vol. 31, no. 10, pp. 1955–1964, 2012.
- [5] G. Litjens, R. Toth, W. van de Ven, C. Hoeks, S. Kerkstra, B. van Ginneken *et al.*, "Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge," *Medical Image Analysis*, vol. 18, no. 2, pp. 359–373, 2014.
- [6] L. Yu, X. Yang, H. Chen, J. Qin, and P. A. Heng, "Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d MR images," in *Thirty-first AAAI Conference on Artificial Intelligence*, 2017.
- [7] B. Wang, Y. Lei, T. Wang, X. Dong, S. Tian, X. Jiang *et al.*, "Automated prostate segmentation of volumetric CT images using 3D deeply supervised dilated FCN," in *Medical Imaging 2019: Image Processing*, vol. 10949. International Society for Optics and Photonics, 2019, p. 109492S.
- [8] S. Wang, K. He, D. Nie, S. Zhou, Y. Gao, and D. Shen, "CT male pelvic organ segmentation using fully convolutional networks with boundary sensitive representation," *Medical Image Analysis*, vol. 54, pp. 168–178, 2019.
- [9] X. Dong, Y. Lei, S. Tian, T. Wang, P. Patel, W. J. Curran *et al.*, "Synthetic mri-aided multi-organ segmentation on male pelvic ct using cycle consistent deep attention network," *Radiotherapy and Oncology*, vol. 141, pp. 192–199, 2019.
- [10] Y. Lei, T. Wang, S. Tian, X. Dong, A. B. Jani, D. Schuster *et al.*, "Male pelvic multi-organ segmentation aided by cbct-based synthetic mri," *Physics in Medicine & Biology*, vol. 65, no. 3, p. 035013, 2020.
- [11] D. Andreasen, J. M. Edmund, V. Zografos, B. H. Menze, and K. Van Leemput, "Computed tomography synthesis from magnetic resonance images in the pelvis using multiple random forests and auto-context features," in *Medical Imaging 2016: Image Processing*, vol. 9784. International Society for Optics and Photonics, 2016, p. 978417.
- [12] C. Qian, L. Wang, Y. Gao, A. Yousuf, X. Yang, A. Oto *et al.*, "In vivo MRI based prostate cancer localization with random forests and auto-context model," *Computerized Medical Imaging and Graphics*, vol. 52, pp. 44–57, 2016.
- [13] N. Lay, N. Birkbeck, J. Zhang, and S. K. Zhou, "Rapid multi-organ segmentation using context integration and discriminative models," in *International Conference on Information Processing in Medical Imaging*. Springer, 2013, pp. 450–462.
- [14] K. He, X. Cao, Y. Shi, D. Nie, Y. Gao, and D. Shen, "Pelvic organ segmentation using distinctive curve guided fully convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 38, no. 2, pp. 585–595, 2018.
- [15] Y. Gao, Y. Shao, J. Lian, A. Z. Wang, R. C. Chen, and D. Shen, "Accurate segmentation of CT male pelvic organs via regression-based deformable models and multi-task random forests," *IEEE Transactions on Medical Imaging*, vol. 35, no. 6, pp. 1532–1543, 2016.
- [16] Q. Feng, M. Foskey, W. Chen, and D. Shen, "Segmenting CT prostate images using population and patient-specific statistics for radiotherapy," *Medical Physics*, vol. 37, no. 8, pp. 4121–4132, 2010.
- [17] M. Yang, X. Li, B. Turkbey, P. L. Choyke, and P. Yan, "Prostate segmentation in MR images using discriminant boundary features," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 2, pp. 479–488, 2012.
- [18] C. Duan, K. Yuan, F. Liu, P. Xiao, G. Lv, and Z. Liang, "Volume-based features for detection of bladder wall abnormal regions via MR cystography," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 9, pp. 2506–2512, 2011.
- [19] S. S. Garapati, L. Hadjiiski, K. H. Cha, H.-P. Chan, E. M. Caoili, R. H. Cohan *et al.*, "Urinary bladder cancer staging in CT urography using machine learning," *Medical Physics*, vol. 44, no. 11, pp. 5814–5823, 2017.
- [20] Y. Xie, M. Chao, P. Lee, and L. Xing, "Feature-based rectal contour propagation from planning CT to cone beam CT," *Medical Physics*, vol. 35, no. 10, pp. 4450–4459, 2008.
- [21] C. Chen and G. Zheng, "Fully automatic segmentation of ap pelvis x-rays via random forest regression and hierarchical sparse shape composition," in *International Conference on Computer Analysis of Images and Patterns*. Springer, 2013, pp. 335–343.
- [22] B. Glocker, O. Pauly, E. Konukoglu, and A. Criminisi, "Joint classification-regression forests for spatially structured multi-object segmentation," in *European conference on computer vision*. Springer, 2012, pp. 870–881.
- [23] C. Lu, Y. Zheng, N. Birkbeck, J. Zhang, T. Kohlberger, C. Tietjen *et al.*, "Precise segmentation of multiple organs in CT volumes using learning-based approach and information theory," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2012, pp. 462–469.
- [24] Y. Wu, G. Liu, M. Huang, J. Guo, J. Jiang, W. Yang *et al.*, "Prostate segmentation based on variant scale patch and local independent projection," *IEEE transactions on medical imaging*, vol. 33, no. 6, pp. 1290–1303, 2014.
- [25] L. Ma, R. Guo, Z. Tian, R. Venkataraman, S. Sarkar, X. Liu *et al.*, "Combining population and patient-specific characteristics for prostate segmentation on 3d ct images," in *Medical Imaging 2016: Image Processing*, vol. 9784. International Society for Optics and Photonics, 2016, p. 978427.
- [26] E. M. A. Anas, S. Nouranian, S. S. Mahdavi, I. Spadinger, W. J. Morris, S. E. Salcudean *et al.*, "Clinical target-volume delineation in prostate brachytherapy using residual neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2017, pp. 365–373.
- [27] T. Brosch, J. Peters, A. Groth, T. Stehle, and J. Weese, "Deep learning-based boundary detection for model-based segmentation with application to MR prostate segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 515–522.
- [28] T.-T. Liang, S. Tsutsui, L. Gao, J.-J. Lu, and M. Sun, "Combining pyramid pooling and attention mechanism for pelvic MR image semantic segmentation," *arXiv preprint arXiv:1806.00264*, 2018.
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [30] E. M. A. Anas, P. Mousavi, and P. Abolmaesumi, "A deep learning approach for real time prostate segmentation in freehand ultrasound guided biopsy," *Medical Image Analysis*, vol. 48, pp. 107–116, 2018.
- [31] J. Léger, E. Brion, P. Desbordes, C. De Vleeschouwer, J. A. Lee, and B. Macq, "Cross-domain data augmentation for deep-learning-based male pelvic organ segmentation in cone beam ct," *Applied Sciences*, vol. 10, no. 3, p. 1154, 2020.
- [32] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.
- [33] Ö. eCicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.
- [34] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 565–571.
- [35] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
- [36] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 801–818.
- [37] F. Isensee, P. F. Jäger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "Automated design of deep learning methods for biomedical image segmentation," *arXiv preprint arXiv:1904.08128*, 2020.
- [38] Y. Shi, S. Liao, Y. Gao, D. Zhang, Y. Gao, and D. Shen, "Prostate segmentation in ct images via spatial-constrained transductive lasso," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2227–2234.

- [39] M. S. Elmahdy, J. M. Wolterink, H. Sokooti, I. Ivsgum, and M. Staring, "Adversarial optimization for joint registration and segmentation in prostate ct radiotherapy," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 366–374.
- [40] A. A. Taha and A. Hanbury, "Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool," *BMC Medical Imaging*, vol. 15, no. 1, p. 29, 2015.
- [41] F. Chollet *et al.*, "Keras," <https://keras.io>, 2015.
- [42] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.
- [43] Z. Tu, "Auto-context and its application to high-level vision tasks," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–8.
- [44] Y. Qin, K. Kamnitsas, S. Ancha, J. Nanavati, G. Cottrell, A. Criminisi *et al.*, "Autofocus layer for semantic segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 603–611.
- [45] E. E. Klein, J. Hanley, J. Bayouth, F.-F. Yin, W. Simon, S. Dresser *et al.*, "Task group 142 report: Quality assurance of medical accelerators," *Medical physics*, vol. 36, no. 9Part1, pp. 4197–4212, 2009.