Original Paper

PSHop: A Lightweight Feed-Forward Method for 3D Prostate Gland Segmentation

Yijing Yang[1], Vasileios Magoulianitis[1*], Jiaxin Yang[1], Jintang Xue[1], Masatomo Kaneko[2], Giovanni Cacciamani[2], Andre Abreu[2], Vinay Duddalwar[3,2], C.-C. Jay Kuo[1], Inderbir S. Gill[2] and Chrysostomos Nikias[1]

[1]*Electrical and Computer Engineering Department, University of Southern California (USC), Los Angeles, CA, USA*
[2]*Department of Urology, Keck School of Medicine, University of Southern California (USC), Los Angeles, CA, USA*
[3]*Department of Radiology, Keck School of Medicine, University of Southern California (USC), Los Angeles, CA, USA*

ABSTRACT

Automatic prostate segmentation is an important step in computer-aided diagnosis of prostate cancer and treatment planning. Existing methods of prostate segmentation are based on deep learning models which have a large size and lack of transparency which is essential for physicians. In this paper, a new data-driven 3D prostate segmentation method on MRI is proposed, named PSHop. Different from deep learning based methods, the core methodology of PSHop is a feed-forward encoder-decoder system based on successive subspace learning (SSL). It consists of two modules: 1) encoder: fine to coarse unsupervised representation learning with cascaded VoxelHop units, 2) decoder: coarse to fine segmentation prediction with voxel-wise classification and local refinement. Experiments are conducted on the publicly available

*Corresponding author: magoulia@usc.edu

ISBI-2013 dataset, as well as on a larger private one. Experimental analysis shows that our proposed PSHop is effective, robust and lightweight in the tasks of prostate gland and zonal segmentation, achieving a Dice Similarity Coefficient (DSC) of 0.873 for the gland segmentation task. PSHop achieves a competitive performance comparatively to other deep learning methods, while keeping the model size and inference complexity an order of magnitude smaller.

1   Introduction

Prostate cancer (PCa) is reported as the second most frequent cancer among men in 2020, with an estimated of almost 1.4 million new cases and 375,000 deaths worldwide [52]. In 112 out of 185 countries of the world, it is even the most frequently diagnosed cancer in men. International guidelines recommend systematic 12-core transrectal ultrasound-guided biopsy (TRUSGB) in biopsy-naïve men with elevated prostate-specific antigen (PSA) serum levels of >3 ng/ml [18]. Compared to TRUSGB with limitations in the diagnosis, multiparametric magnetic resonance imaging (mpMRI) has been reported to reduce the detection of insignificant prostate cancer [59, 50, 32]. It has become the imaging method that is best able to detect clinically significant prostate cancer and guide biopsies [58], due to the superior resolution and contrast of imaging, without harming the human body [61].

Based on prostate MRI, prostate segmentation is an important step in the PCa diagnosis and treatment planning with various aims, such as localizing prostate boundaries for radiotherapy [43], automating the calculation of the prostate volume –key task for the Prostate-Specific Antigen Density (PSA-D) calculation– as well as track disease progression [56], or localizing the region of interest at the beginning of the computer-aided diagnosis (CADx) of PCa [60, 55, 41]. Moreover, zonal segmentation may enhance the PCa detection models [64], since the prostate zones have different visual features [17] and potentially zonal segmentation can be used to separate the feature extraction process.

However, manual MR segmentation takes extended time and labor. Radiologists need to mark slice by slice through visual inspection with a high demand of skills and expertise for accurate segmentation. It also comes with high intra- and inter-observer variation. Thus, automated MR prostate segmentation is needed to help improve the accuracy and efficiency in this routinely applied task from radiologists. Due to the various size and shape of

the prostate gland across different patients, low contrast between the gland and adjacent structures, imaging artifacts, as well as heterogeneity in signal intensity around endorectal coils (ERCs) [16, 22, 23], it is still a challenging task.

Existing prostate segmentation algorithms can be categorized into two classes: traditional and deep learning based methods. The traditional methods mainly include contour based [49, 14], atlas based [25, 35, 15], deformable models [11, 26], and machine learning based models such as c-means clustering [48, 47] and classification [2, 34].

With the development of convolutional neural networks (CNNs), deep learning based segmentation methods [53, 23, 1, 39, 51] are pioneered by the popular U-Net [46] and V-Net [42] architectures. They have achieved superior segmentation performance, comparable to that of expert radiologists [20].

Recently, a novel representation learning framework, named Green Learning (GL) was proposed by [30]. It provides a linear feature extraction model for image analysis, with key benefits a lightweight model size, as well as transparent and explainable pipeline. GL is based on the Successive Subspace Learning (SSL) methodology for feature extraction that was proposed in a sequence of papers [28, 27, 29, 31]. These features have multi-scale properties and are extracted in an unsupervised feed-forward manner using core signal processing operations [8]. GL has been applied for medical image analysis in the tasks of Amyotrophic lateral sclerosis (ALS) classification and cardiac MRI segmentation [38, 37].

In this work, we propose a novel prostate segmentation method named PSHop, which is built upon the methodology of SSL. The main contributions can be summarized in three folds:

1. To the best of our knowledge, PSHop is the first work which applies the Green Learning paradigm and SSL methodology on a 3D medical image segmentation problem, thus exploring a new direction in this field.

2. We propose a feed-forward encoder-decoder network for 3D medical image segmentation without the use of back-propagation that makes the entire pipeline transparent.

3. Our proposed PSHop method is lightweight which significantly reduces the model size compared to other deep learning based methods while keeping comparable performance.

The rest of the paper is organized as follows. Related work is reviewed in Section 2. The proposed PSHop method is presented in Section 3. The experimental setup and results are then discussed in Section 4 and Section 5. Conclusions are finally drawn in Section 6.

2   Related Work

2.1   Traditional Methods

Traditional methods mainly focus on atlas-based, deformable models, and graph cuts optimization. Atlas-based methods [25, 35, 15] improve the segmentation accuracy by measuring the similarity between target image and multiple atlases. [25] build two deformed atlas images for comparisons, stressing how important the atlas selected images are. For similarity measure they use the normalized mutual information (NMI) and a majority voting algorithm to combine multiple image segmentations. In a work of [35], a selective and iterative method for performance level estimation is proposed instead of majority voting.

For methods using graph cuts (GC) [5, 4], optimization is applied to find an optimal solution that separates different regions. Images are represented as graphs and segmentation is viewed as a graph cutting optimization process under certain constraints. For example, [7] combine graph cuts with active appearance models to improve the segmentation accuracy. Another work [45] solves the dual problem using convex optimization and specifically employs the flow-maximization algorithms in graphs. [54] propose supervoxel-based graph cuts and a 3D active contour model for segmentation refinement.

2.2   Learning-based Methods

With the increasing number of accessible data with ground truth annotations, learning based medical image segmentation methods have made a remarkable progress with the support of machine learning and deep learning. At first, fully convolutional network (FCN) was proposed by modifying the existing classification CNN for the segmentation task. Several FCN-based methods for medical image segmentation are then proposed [46, 42, 44, 53]. For example, U-Net [46] and V-Net [42] are two representative pioneer work for the 2D and 3D medical image segmentation. In U-Net, combines a contracting path of multiple convolutional layers and an expansive path of up-convolutional layers as an encoder-decoder network struction. The skip architecture in U-Net uses a simply concatenation operation that builds a bridge between encoder and decoder that take advantage of both coarse and fine features. Different from U-Net which processes 2D images, V-Net was generalized to 3D medical image segmentation based on a volumetric processing. It also introduced a novel loss layer based on the Dice coefficient. [44] proposed a 3D-GCN framework based on FCN where a Global Convolutional Network (GCN) was designed to address both the localization and classification for segmentation. A boundary refinement block was also proposed which models the boundary alignment as a residual structure.

With the promising performance of U-Net and V-Net, more algorithms are proposed having them as the backbone model in recent years. For example, DenseNet-like U-Net [1] took advantage of the strengths of both DenseNet and U-Net for segmentation of the prostate gland and its zones. [10] introduced a network for volumetric segmentation that learns from sparsely annotated volumetric images by extending the previous U-Net architecture by replacing all 2D operations with their 3D counterparts. [51] proposed a shape attentive U-Net (SAUNet) where both the texture information and shape information are used to learn the segmentation. [24] propose a bicubic intepolation for extracting the low frequencies of input MRI as a preprocessing step to V-Net for the segmentation task. On the other hand, unsupervised methods [36, 57, 65] perform domain adaptation using contrastive or predictive learning and can be also applied on medical imaging for segmenting several organs. Despite their limitations in performance comparing to the supervised ones, this is a prolific research area, especially in medical imaging, where annotated data are scarce and expensive to obtain.

Attention mechanisms have been proved quite efficient in learning better feature representations. [39] proposed a feature pyramid attention sub-module before the decoder in FCN, considering the semantic information from U-Net may not be sufficient to represent the heterogeneous anatomic structures for a clear boundary. [13] interleave the U-Net skip connections with a multi-scale self-attention mechanism for recalibrating the feature maps across multiple layers. [62] developed a two-stage approach, where they employ a Squeeze and Excitation (SE) CNN for detecting the prostate's existence in stage-1 and a Residual-Attention U-Net in stage-2 for segmenting the slices that include the prostate gland. [33] propose a dual attention mechanism using 3D convolutions to learn in an end-to-end manner both the gland and lesion segmentation tasks.

An interesting work of [23] proposes 3D APA-Net, a 3D adversarial pyramid anisotropic convolutional deep neural network for prostate segmentation, which has an encoder-decoder architecture, equipped with adversarial training for spatially consistent and continuous segmentation results. Recently, a novel attention mechanism among slices is proposed from [19] to learn cross-slices features at multiple scales using transformer blocks. A more generic approach of [9] proposes a semi-supervised learning method for segmenting medical images, using attention among slices to capture the common spatial layout of patients organs and a contrastive learning scheme to incorporate unlabeled data in training.

In some works, certain modules are proposed to help make the boundary more clear and accurate. For example, [21] proposed hybrid discriminative network named HD-Net, in which the decoder consists of two branches: a 3D segmentation branch and a 2D boundary branch to boost the shared encoder to learn features with more semantic discrimination. [66] proposed a

BOWDA-Net where a boundary-weighted segmentation loss was introduced to the transfer learning.

2.3   Successive Subspace Learning Methodology

Recently, being inspired by deep learning but being different, the successive subspace learning (SSL) methodology was proposed by Kuo *et al.* in a sequence of papers [28, 27, 31]. Instead of using back-propagation, feature representations in SSL-based methods are learnt in an unsupervised feedforward manner using multi-stage principal component analysis (PCA) for multi-scale subspace learning. The overall framework is named Green Learning, as it is meant to offer image analysis and understanding solutions with much less parameters and complexity.

Three variants of transforms were originally proposed, including Saak (subspace approximation with augmented kernels) transform [29], the Saab (subspace approximation via adjusted bias) transform [31], and the channel-wise (c/w) Saab transform [8]. Among them, the c/w Saab transform requires the smallest model size and has the best transform efficiency because it takes advantage of the weak correlation between channels so that filters are learnt from each channel separately. The details of c/w Saab transform is introduced in Section 3.1.

SSL methodology has been applied to different problems in medical imaging. A recent work which follows a similar feed-forward approach was published by [40]. Two other representative works of SSL-based medical image analysis can be found in [38] and [37], which solve ALS disease classification and cardiac MRI segmentation, respectively. Our proposed PSHop can find the closest shadows of [37] and is the first work that generalizes the SSL methodology on a 3D medical image segmentation.

3   Methods

A complete solution to prostate segmentation entails two tasks: (1) the whole gland segmentation and (2) the zonal segmentation. Each of those tasks has its importance within the prostate medical diagnosis pipeline. Our proposed PSHop method is used for both tasks separately and its architecture overview is shown in Figure 1, which has an encoder-decoder structure, inspired by U-Net structure. In the encoder part, unsupervised feature representations of different scales are extracted from fine to coarse. In the decoder part, segmentation mask is predicted in a coarse to fine fashion based on the representations from the encoder. The details of the method are described in Section 3.1 and Section 3.2.
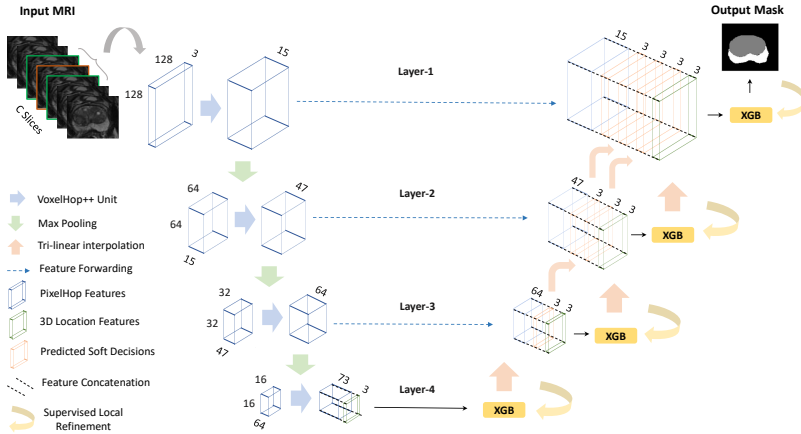
Figure 1: Illustration of the U-Net-like 3D architecture of PSHop for representation learning and feature extraction. Four scales (layers) are used in PSHop. Deeper layers correspond to coarser features, while shallower ones are meant to refine the segmentation result and output the segmentation mask at the input's scale. The process is shown for segmenting one slice from the entire sequence. It is repeated for every single slice in the MRI input.

3.1   Encoder: Fine to Coarse Representation Learning

We treat the segmentation task as a voxel-wise classification problem and extract the representation of a neighborhood for each voxel in different scales. Different from deep learning based methods where the convolutional filters are learnt through end-to-end optimization of the loss function, we use cascaded VoxelHop units [38] which is a statistical approach to extract feature vectors. The process is fully unsupervised, with no ground truth labels required during training. It is also a feed-forward learning process instead of using back-propagation in deep learning based encoders.

3.1.1   Feature Extraction Based on VoxelHop Units

Figure 2 illustrates the process of each VoxelHop unit. It consists of two consecutive steps: 1) neighborhood construction in the 3D space, and 2) representation learning through the c/w Saab transform [8].

Suppose the input tensor of the i-th VoxelHop unit is of dimension $H_i \times W_i \times C_i \times K_i$, where H_i, W_i and C_i represent the resolution in 3D space, and K_i represents the dimension of the feature vector for each voxel extracted from the $(i-1)$-th VoxelHop unit. Specifically, for the first VoxelHop unit where $i = 1$, $H_1 \times W_1 \times C_1$ corresponds to the input MRI data resolution, and $K_1 = 1$. We first gather the neighborhood in the 3D space centered at each
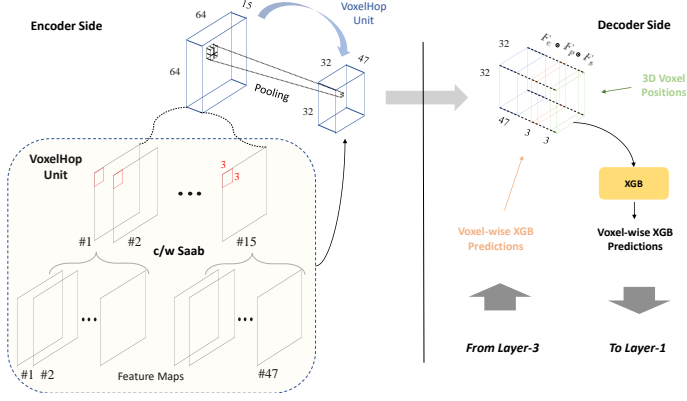
Figure 2: Feature representation in one VoxelHop layer within PSHop. Layer 2 is borrowed as example to show the connection between encoder-decoder and feature concatenation from the encoder F_e, the voxel-wise probabilities from the coarser scale after interpolation F_p and the 3D voxel positions F_s. The concatenated feature is fed in XGB classifier to predict the voxel-wise features for layer-1.

voxel. The neighborhood size is defined as $S_{Hi} \times S_{Wi} \times S_{Ci}$ in spatial. Each voxel in the neighborhood has a feature vector of dimension K_i, which results in a tensor of size $S_{Hi} \times S_{Wi} \times S_{Ci} \times K_i$.

The neighborhood tensor is then flattened in the spatial domain. Channel-wise Saab transform is performed in each of the C_i channels separately to learn spectral signals through subspace approximation at the current scale. Suppose the input vector is $\mathbf{x} \epsilon \mathbf{R}^N$, where $N = S_{Hi} \times S_{Wi} \times S_{Ci}$. Features can be extracted by projecting the input vector on to several anchor vectors, which can be expressed as an affine transform expressed as:

$$y_m = \mathbf{a}_m^T \cdot \mathbf{x} + b_m, m = 0, 1, \cdots, M-1, \tag{1}$$

where $\mathbf{a}_m$ is the m-th anchor vector of dimension N, and M is the total number of anchor vectors. Here, the channel-wise Saab transform is a data-driven approach to learn the anchor vectors from all the neighborhood tensors collected from the input data. First, it decomposes the input subspaces into the direct sum of two subspaces, i.e. DC and AC, expressed in Equation 2, where the terms are borrowed from the "direct circuit" and "alternating circuit" in the circuit theory.

$$S = S_{DC} \oplus S_{AC}. \tag{2}$$

S_{DC} and S_{AC} are spanned by DC and AC anchor vectors, defined as:

- DC anchor vector $\mathbf{a}_0 = \frac{1}{\sqrt{N}} (1, 1, \cdots, 1)^T$

- AC anchor vectors $\mathbf{a}_m$, $m = 1, \cdots, M-1$.

The two subspaces are orthogonal to each other, where the input signal $\mathbf{x}$ is projected on to $\mathbf{a}_0$ to get the DC component $\mathbf{x}_{DC}$. Then the AC component is extracted by subtracting DC component from the input signal, i.e. $\mathbf{x}_{AC} = \mathbf{x} - \mathbf{x}_{DC}$.

After that, AC anchor vectors are learnt by conducting principal component analysis (PCA) on the AC component. The first K principal components are kept as the AC anchor vectors. Thus, one can extract features by projecting $\mathbf{x}$ on to the above learnt anchor vectors based on Equation 1. The bias term is selected to ensure all features are positive by following [31]. An illustration of the feature extraction concept behind the VoxelHop unit using the derived subspaces from PCA is provided in Figure 3.
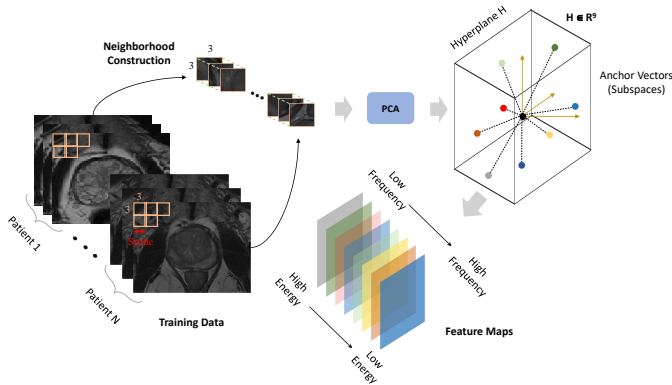

Figure 3: An illustration of the local neighborhood construction for unsupervised filter learning using the Saab transform based on PCA. Each anchor vector corresponds to a subspace on the N-dimensional plane (denoted in different colors). The input image is projected on to these subspaces to obtain its spectral decomposition at a certain scale that corresponds to the output feature map. In c/w Saab this decomposition is applied on every single feature map in a recursive manner, until the maximum number of layers is reached.

3.1.2 Representation Learning Through Cascaded VoxelHop Units

Each of the above presented VoxelHop unit extracts the representation in a certain resolution. For segmentation tasks, both local and global descriptions are important. Features that represent a small neighborhood serve for better localization, while features from a larger neighborhood provide a more accurate context semantic understanding.

In the PSHop encoder, we use cascaded VoxelHop units with max-pooling in between. Thus, the receptive field of the representations grows fast as more VoxelHop units are cascaded until a preset number of cascaded layers,

L, is reached. The detailed architecture of the proposed PSHop encoder is summarized in Table 1. We set $L = 4$, where VoxelHop units from Hop-1 to Hop-4 form a representation learning from fine to coarse. Here, we use "hop" to represent a certain neighborhood range. The four cascaded hops can be constructed as a tree-decomposed structure, where the i-the VoxelHop unit yields the i-th child nodes (see Figure 2). The number of Saab filters for each hop unit is decided by the input data using a pre-set energy threshold, where the energy of a child node refers to the multiplication of the energy of the parent node and the normalized energy from PCA among child nodes in the same level.

Table 1: Encoder Architecture of the Proposed PSHop

	Filter Size	Stride
VoxelHop 1	$(3 \times 3) \times 3$	$(1 \times 1) \times 1$
Max-pooling 1	$(2 \times 2) \times 2$	$(2 \times 2) \times 2$
VoxelHop 2	$(3 \times 3) \times 3$	$(1 \times 1) \times 1$
Max-pooling 2	$(2 \times 2) \times 2$	$(2 \times 2) \times 2$
VoxelHop 3	$(3 \times 3) \times 3$	$(1 \times 1) \times 1$
Max-pooling 3	$(2 \times 2) \times 2$	$(2 \times 2) \times 2$
VoxelHop 4	$(3 \times 3) \times 3$	$(1 \times 1) \times 1$

3.2   Decoder: Coarse to Fine Segmentation Prediction

The segmentation is conducted from coarse to fine based on the features from PSHop encoder. We first start from the deepest hop, Hop-L, and perform a one-level PSHop decoder unit. Then the process gradually move to a shallower hop unit until it outputs the segmentation prediction for the full input resolution. Each PSHop decoder unit consists of the following three steps: 1) feature aggregation; 2) segmentation and local refinement at the current scale; and 3) prediction upsampling.

3.2.1   Feature Aggregation

The feature for the segmentation at each scale comes from different sources. Besides the *encoder features* F_e^i at the i-th hop, a *position encoding* feature vector F_s^i is included where the 3D voxel coordinate is recorded, expressed as

$$F_s^i = [x, y, z]^T . \tag{3}$$

Also, the *predicted probability* vectors F_p^i upsampled from all the coarser hops is also included. The final feature vector for segmentation prediction is an aggregation of the above mentioned three features detailed as

$$F_{seg}^i = F_s^i \odot F_p^i \odot F_e^i, \tag{4}$$

where F_e^i is copied from encoder to the decoder at the i-th hop using a skip connection, and $\odot$ represents the voxel-wise concatenation operation. A detailed illustration can be found in Figure 2.

The three parts serve as different roles. First, the relative physical positions of different structures is similar among different patients in MRI images, for example, the prostate is always around the center region in the 2D plains. The position encoding F_s helps merge this prior knowledge into the prediction process. Second, F_p^i is propagated from all the coarser grids, which contains the probability of classes predicted using different receptive fields that are larger than that of the ith hop. Thus, F_p^i provides a memory of coarse to fine context semantic information. With these conditions, the role of F_e^i is to provide a representation of the local neighborhood of each voxel so that the segmentation prediction can result in a finer detail.

3.2.2 Segmentation Prediction and Local Refinement

We treat the segmentation at each scale as a voxel-wise classification problem. In the training process, the ground truth masks are first encoded as one-hot vectors representing the corresponding class. The ground truth of each voxel grid at the ith hop is then downsampled from ground truth in the original resolution using bilinear interpolation. The selection of the training samples is based on the confidence level of the interpolated ground truth mask. Only the voxels with high confidence is included as the training samples. Then, an eXtreme Gradient Boosting (XGBoost [6]) classifier is trained using the aggregated feature F_{seg}^i. The output is the predicted soft decision vectors at the current scale.

Since the segmentation requires smoothness in a local neighborhood while the classification solution is made for each voxel separately, we adopt a local refinement step after each XGBoost classifier based on the soft decisions. Inspired by the soft-label smoothing (SLS) technique proposed in [63], we gather a cubic of soft decisions in the 3D domain of size $3 \times 3 \times 3$. The aggregated neighborhood soft decisions are concatenated as the new feature to train another XGBoost classifier. This is repeated iteratively. In practice, we perform two iterations for the soft decision update. In this way, the resulted segmentation prediction is smoothed and the precision is improved. This is used at all scales of the prediction, up to layer-1 which gives the final segmentation output. To further refine the output and correct any segmentation artifacts,

we employ a median filter of size 7×7 as a post-processing step. In the experimental section we demonstrate the effectiveness of the post-processing filter.

3.2.3 Prediction Upsampling

To propagate the predicted soft decisions of a coarser grid to a finer grid, we perform a bilinear upsampling towards the target resolution. This process is cumulative, which means that the predicted probability vector F_p^i propagated to the i-th hop is from all the coarser hops, detailed as

$$F_p^i = \left\{ F_p^{i+1} \odot \hat{y}^{i+1} \right\} \uparrow, \quad \forall i \leq L - 1, \tag{5}$$

where $\hat{y}^{i+1}$ is the predicted soft decision at the $(i+1)$-th hop, and $\{\cdot\} \uparrow$ means the bilinear upsampling.

4 Experimental Setup

4.1 Database and Pre-processing

To demonstrate the effectiveness of the proposed PSHop method, we conduct experiments based on one public MR image database NCI-ISBI 2013 Challenge (Automated Segmentation of Prostate Structures [3]) and one private in-house USC-Keck dataset. The ISBI-2013 dataset consists of 60 training cases of axial T2-weighted MR 3D series, where half were obtained at 1.5T (Philips Achieva at Boston Medical Center) and the other half at 3T (Siemens TIM at Radboud University Nijmegen Medical Center). Since the ground truth segmentation includes background, peripheral zone (PZ), and transitional zone (TZ), we merge PZ and TZ as one class – prostate area, for the model training and evaluation since we consider the prostate segmentation task in this paper. The pixel spacing within each slice ranges from 0.39 mm to 0.75 mm, while the through-plane resolution ranges from 3.0 mm to 4.0 mm among different patients.

The USC-Keck dataset consists of a cohort of 260 patients collected in the Keck Medicine School in the University of Southern California. Besides T2-w 3D series, for each patient T2-Cube series is also available. It is known that T2-Cube has smaller pixel spacing, especially along the z-axis. That means, higher resolution and thinner slices. Specifically, for T2-w pixel spacing ranges from 0.5 mm to 0.7 mm and the through-plane resolution (z-axis) from 3.0 mm mm to 4.0 mm. For T2-Cube pixel spacing is set at 0.83 mm and the through-plane resolution (z-axis) at 1.4 mm. The scanner used to acquire those images is the GE 3T with 8ch Cardiac coil. In our experimental analysis with USC-Keck data, the T2-Cube series is used since it can presumably provide more

accurate segmentation results because of the higher perspicuity of the images (stemming from the smaller voxel spacing).

For both datasets, we first regularize the resolution of different images to the same physical resolution of $0.625 \times 0.625 \times 1.5 \, mm^3$. We use Lanczos interpolation where the factor is calculated based on the original pixel spacing and through-plane resolution of each image. Here, the through-plane resolution is increased so that the segmentation in the 3D space can be more accurate. After that, to reduce the artifacts while acquiring the images, contrast enhancement using CLAHE [67] is applied. Finally, before feeding PSHop input, for the whole gland segmentation task the input sequence is resized to 128×128. For the zonal segmentation task, a 256×256 centered crop around the segmented gland is resized to 128×128 to standardize the PSHop input.

4.2 Evaluation Metrics

To quantitatively evaluate the performance, we calculate the Dice Similarity Coefficient (DSC) [12] expressed in Equation (6) in a binary scenario, where X and Y represent the ground truth and the predicted segmentation mask, respectively. DSC is widely used in evaluating segmentation tasks for medical images. It measures the ratio of the intersection of two binary sets to the averaged cardinality.

$$DSC\left(X,Y\right) = \frac{2\left|X \cap Y\right|}{\left|X\right| + \left|Y\right|}.$$ (6)

Additionally to DSC, we include in our analysis the average Hausdorff Distance (HD) which is used to measure the distance between two finite sets of points (i.e. graound truth and prediction) and has been used in medical image analysis for segmentation performance comparison. The units we report the results is distance in voxels.

$$aHD = (\frac{GtoS}{G} + \frac{StoG}{S})/2$$ (7)

G and S are the voxel counts in the ground truth and segmentation, respectively. Also, GtoS is the HD distance from the ground truth to segmentation, while StoG is the HD distance from segmentation to ground truth.

4.3 Experimental Details

For both datasets, we apply the same experimental settings. For ISBI-2013 and USC-Keck we apply 5-fold cross validation on the 60 and 260 training images, respectively and calculate the mean and standard deviation of the evaluation scores. That is, for ISBI-2013 48 sequences are used for training PSHop from each fold and the rest 12 for validation. For USC-Keck dataset 206 sequences are used for training and the rest 54 for validation.

For benchmarking PSHop with other DL-based methods we conduct the same experiment setting for both datasets, and train V-Net and U-Net based architectures. Besides segmentation performance comparison using the DSC metric, we also compare the model size and complexity of the models, since the main target and motivation of this work is to offer a lightweight solution comparing to other methods.

5   Experimental Results

5.1   Segmentation Results

The DSC scores of ISBI-2013 and USC Keck datasets obtained by our proposed PSHop method for the prostate gland segmentation task are summarized in Table 2. The averaged validation performance over the 5-folds is reported and the standard deviation as well (shown in parenthesis). To begin with, PSHop has a very competitive performance among the two baseline supervised DL-based works. Also, comparing it with a state-of-the-art unsupervised method, such as CUTS [36] it outperforms by large margins in both datasets (see Tables 3 and 4) in both metrics. We benchmark with CUTS using two different versions for applying labels. From the results, one can infer that supervision definitely helps to achieve a higher performance, as it is hard to capture all the variations of prostate appearance without any supervision.

Table 2: Comparison of the whole gland segmentation performance with PSHop and two popular baseline deep learning models using the DSC metric.

	ISBI-2013	USC-Keck
V-Net	0.762 ($\pm$0.139)	0.906 ($\pm$ 0.009)
2D U-Net	0.684 ($\pm$0.031)	0.809 ($\pm$ 0.036)
PSHop (Ours)	0.826 ($\pm$0.018)	0.873 ($\pm$ 0.017)

Table 3:   Comparison of the whole gland segmentation performance with PSHop and a state-of-the-art unsupervised deep learning model using the DSC and HD metrics on the ISBI-2013 dataset.

	DSC	HD
CUTS + Spectral K-Means	0.585 ($\pm$ 0.010)	124.602 ($\pm$ 1.403)
CUTS + Diffusion	0.589 ($\pm$ 0.010)	121.638 ($\pm$ 1.486)
PSHop (Ours)	0.826 ($\pm$0.018)	14.206 ($\pm$ 4.967)

Table 4: Comparison of the whole gland segmentation performance with PSHop and a state-of-the-art unsupervised deep learning model using the DSC and HD metrics on the USC-Keck dataset.

	DSC	HD
CUTS + Spectral K-Means	0.653 ($\pm$ 0.009)	135.404 ($\pm$ 1.188)
CUTS + Diffusion	0.661 ($\pm$ 1.248)	131.696 ($\pm$ 1.288)
PSHop (Ours)	0.873 ($\pm$0.017)	14.159 ($\pm$ 7.991)

From the two supervised works, PSHop outperfoms V-Net on ISBI-2013 dataset and U-Net on the USC-Keck one. ISBI-2013 has considerably fewer patient data than USC-Keck. Therefore, V-Net performs better when it is given with sufficient training patients, while U-Net has a higher performance with fewer training samples. That is also evident from the high standard deviation on ISBI data from V-Net. Another important observation is that PSHop has a more stable performance (low standard deviation in both experiments), regardless the number of training samples. This underlines one of the GL framework advantages that is more stable even for fewer training samples, while large DL models fail to achieve a high performance when data are scarce. That also confirms GL's main assumption that statistical-based feed-forward models can still perform well even with a small number of training samples, which is usually the case for medical imaging datasets.

Steering our comparisons to the zonal segmentation performance, Table 5 shows the benchmarking on USC-Keck dataset, since it is fairly larger than the ISBI-2013 and thereby stronger conclusions can be drawn. For TZ, PSHop surpasses U-Net by large margins and maintains a small performance gap with V-Net. On the other hand, one can observe that the performance gap is larger on the PZ when compared with V-Net and this is an area for further improvement for GL. Yet, PSHop achieves a much higher DSC score comparing to the U-Net, which is the baseline architecture for many works throughout the literature. For the smaller ISBI-2013 dataset PSHop outperforms the other methods on the PZ (see Table 6), mainly due to the small number of training samples that PSHop has an advantage. On the TZ, PSHop surpasses the U-Net performance by large margins.

In the above comparisons, V-Net generally performs better when trained with a large number of samples and U-Net performs well only on the ISBI-2013, which has many fewer samples. This is sensible because V-Net has many more trainable parameters comparing to U-Net and thus can fit better the data diversity. Table 7 demonstrates the green advantages and benefits of GL when it comes to complexity and model size comparison. PSHop has an order of magnitude less parameters than the DL-based models. Also, in terms

Table 5: Comparison of the zonal segmentation performance with PSHop and two popular baseline deep learning models using the DSC metric on USC-Keck data.

	TZ	PZ
V-Net	0.878 (± 0.019)	0.747 (± 0.014)
U-Net	0.741 (± 0.041)	0.525 (± 0.032)
PSHop (Ours)	0.845 (± 0.025)	0.656 (± 0.012)

Table 6: Comparison of the zonal segmentation performance with PSHop and two popular baseline deep learning models using the DSC metric on ISBI-2013 data.

	TZ	PZ
V-Net	0.629 (± 0.064)	0.431 (± 0.030)
2D U-Net	0.563 (± 0.020)	0.316 (± 0.031)
PSHop (Ours)	0.667 (± 0.026)	0.366 (± 0.039)

Table 7: Comparison of model size and complexity in inference between PSHOP and the two deep learning baseline models.

	# of parameters	FLOPS
V-Net	45,603,934	379B (×5269)
2D U-Net	17,970,626	13.6B (×190)
PSHop (Ours)	235,206	72M (×1)

of complexity, it has ×190 less FLOPS than U-Net and ×5269 than V-Net. These comparisons stress the tremendous advantages of GL-based solutions for application deployment.

Overall, PSHop maintains a very competitive standing performance-wise with other DL baseline models, outperforming in general U-Net in both tasks and having a close performance with V-Net. Nevertheless, V-Net's performance comes at the expense of a higher complexity and model size.

In Figures 4 and 5, we provide qualitative comparisons for the whole gland and zonal segmentation, respectively. One observation is the segmentation refinement from the post-processing module for both tasks. Moreover, our module performs better on segmenting the TZ than PZ which is usually a more challenging task for various models, due to its irregular shape variations along slices. Looking at the qualitative results, one can see that unclear boundaries between TZ and PZ, or uneven and cluttered texture of the prostate challenges our method, particularly in the zonal segmentation task. PZ is relatively much smaller than TZ and hence it provides less information during training. This
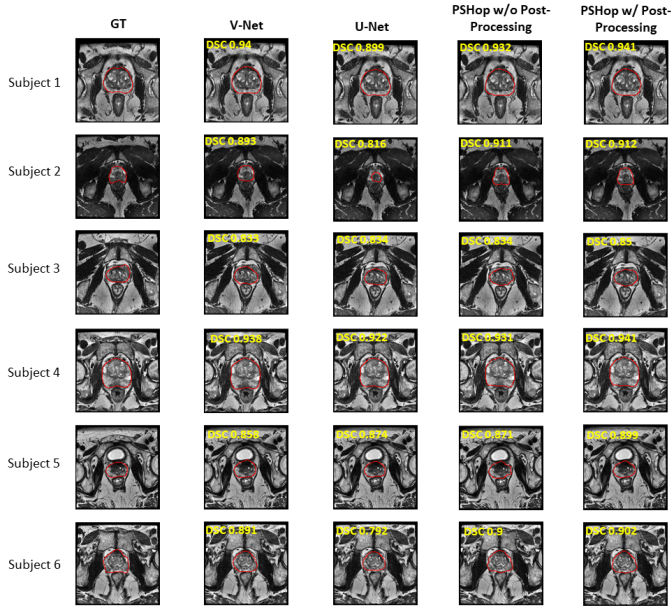
Figure 4: Qualitative comparison among DL methods and PSHop on the whole gland segmentation task. Results are also shown before and after the post-processing operation of the median filter.

is a limitation of PSHop that can be addressed in a future extension of this work by introducing an attention mechanism to enhance learning on the PZ, as well as include stronger topological features to leverage on the priors of prostate structure.

6   Conclusion

This work proposes the PSHop method for accurate prostate gland and zonal segmentation. Unlike other majority state-of-the-art works based on DNNs, such as U-Net or V-Net, PSHop follows the GL paradigm and adopts a feed-forward model for feature extraction. The model architecture is inspired by U-Net, where there are several multi-scale representations of the input MRI, nevertheless PSHop uses no back-propagation, but the SSL methodology to extract feature representations. PSHop has a competitive performance standing among other DL-based methods, outperforming U-Net on the larger dataset and has a small performance margin with V-Net. All in all, PSHop method
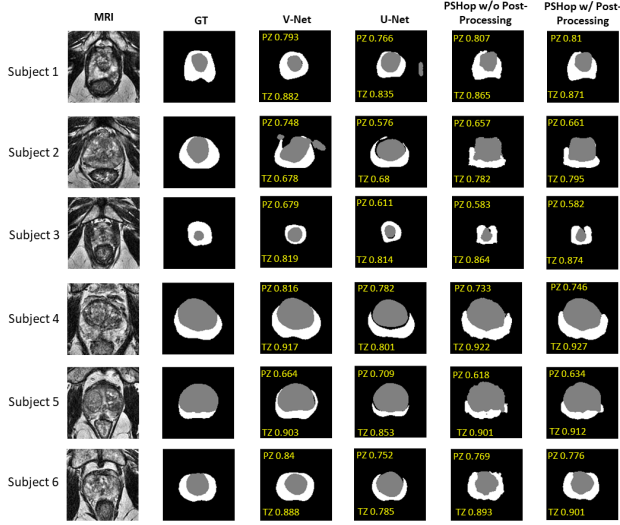
Figure 5: Qualitative comparison among DL methods and PSHop on the zonal segmentation task.

comes with a very lightweight model size and orders of magnitude less computational complexity, thus providing a green alternative for prostate segmentation task. Additionally, the linear feature extraction model provides more transparency in the pipeline and hence makes PSHop decisions more trustworthy to physicians.

References

[1] N. Aldoj, F. Biavati, F. Michallek, S. Stober, and M. Dewey, "Automatic prostate and prostate zones segmentation of magnetic resonance images using DenseNet-like U-net", *Scientific reports*, 10(1), 2020, 1–17.

[2] P. D. Allen, J. Graham, D. C. Williamson, and C. E. Hutchinson, "Differential segmentation of the prostate in MR images using combined 3D shape modelling and voxel classification", in *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.* IEEE, 2006, 410–3.

[3] N. Bloch, A. Madabhushi, H. Huisman, J. Freymann, J. Kirby, M. Grauer, A. Enquobahrie, C. Jaffe, L. Clarke, and K. Farahani, "NCI-

ISBI 2013 challenge: automated segmentation of prostate structures", *The Cancer Imaging Archive*, 370, 2015.

[4] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts", *IEEE Transactions on pattern analysis and machine intelligence*, 23(11), 2001, 1222–39.

[5] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in ND images", in *Proceedings eighth IEEE international conference on computer vision. ICCV 2001*, Vol. 1, IEEE, 2001, 105–12.

[6] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, *et al.*, "Xgboost: extreme gradient boosting", *R package version 0.4-2*, 1(4), 2015, 1–4.

[7] X. Chen, J. K. Udupa, U. Bagci, Y. Zhuge, and J. Yao, "Medical image segmentation by combining graph cuts and oriented active appearance models", *IEEE transactions on image processing*, 21(4), 2012, 2035–46.

[8] Y. Chen, M. Rouhsedaghat, S. You, R. Rao, and C.-C. J. Kuo, "Pixelhop++: A small successive-subspace-learning-based (ssl-based) model for image classification", in *2020 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2020, 3294–8.

[9] Z. Chen, W. Zhuo, T. Wang, J. Cheng, W. Xue, and D. Ni, "Semi-Supervised Representation Learning for Segmentation on Medical Volumes and Sequences", *IEEE Transactions on Medical Imaging*, 2023.

[10] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation", in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, Springer, 2016, 424–32.

[11] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, "The use of active shape models for locating structures in medical images", in *Biennial international conference on information processing in medical imaging*, Springer, 1993, 33–47.

[12] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species", *Ecology*, 26, 1945, 297–302.

[13] M. Ding, Z. Lin, C. H. Lee, C. H. Tan, and W. Huang, "A multi-scale channel attention network for prostate segmentation", *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2023.

[14] M. Ding, C. Chen, Y. Wang, I. Gyacskov, and A. Fenster, "Prostate segmentation in 3D US images using the cardinal-spline-based discrete dynamic contour", in *Medical Imaging 2003: Visualization, Image-Guided Procedures, and Display*, Vol. 5029, SPIE, 2003, 69–76.

[15] J. A. Dowling, J. Fripp, S. Chandra, J. P. W. Pluim, J. Lambert, J. Parker, J. Denham, P. B. Greer, and O. Salvado, "Fast automatic multi-atlas segmentation of the prostate from 3D MR images", in *International Workshop on Prostate Cancer Imaging*, Springer, 2011, 10–21.

[16] S. Ghose, A. Oliver, R. Martí, X. Lladó, J. C. Vilanova, J. Freixenet, J. Mitra, D. Sidibé, and F. Meriaudeau, "A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images", *Computer methods and programs in biomedicine*, 108(1), 2012, 262–87.

[17] S. B. Ginsburg, A. Algohary, S. Pahwa, V. Gulani, L. Ponsky, H. J. Aronen, P. J. Boström, M. Böhm, A.-M. Haynes, P. Brenner, *et al.*, "Radiomic features for prostate cancer detection on MRI differ between the transition and peripheral zones: preliminary findings from a multi-institutional study", *Journal of Magnetic Resonance Imaging*, 46(1), 2017, 184–93.

[18] A. Heidenreich, P. J. Bastian, J. Bellmunt, M. Bolla, S. Joniau, T. van der Kwast, M. Mason, V. Matveev, T. Wiegel, F. Zattoni, *et al.*, "EAU guidelines on prostate cancer. Part 1: screening, diagnosis, and local treatment with curative intentupdate 2013", *European urology*, 65(1), 2014, 124–37.

[19] A. L. Y. Hung, H. Zheng, Q. Miao, S. S. Raman, D. Terzopoulos, and K. Sung, "CAT-Net: A cross-slice attention transformer model for prostate zonal segmentation in MRI", *IEEE transactions on medical imaging*, 42(1), 2022, 291–303.

[20] L. J. Isaksson, M. Pepa, P. Summers, M. Zaffaroni, M. G. Vincini, G. Corrao, G. C. Mazzola, M. Rotondi, G. Lo Presti, S. Raimondi, *et al.*, "Comparison of automated segmentation techniques for magnetic resonance images of the prostate", *BMC medical imaging*, 23(1), 2023, 1–16.

[21] H. Jia, Y. Song, H. Huang, W. Cai, and Y. Xia, "HD-Net: hybrid discriminative network for prostate segmentation in MR images", in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, 110–8.

[22] H. Jia, Y. Xia, Y. Song, W. Cai, M. Fulham, and D. D. Feng, "Atlas registration and ensemble deep convolutional neural network-based prostate segmentation using magnetic resonance imaging", *Neurocomputing*, 275, 2018, 1358–69.

[23] H. Jia, Y. Xia, Y. Song, D. Zhang, H. Huang, Y. Zhang, and W. Cai, "3D APA-Net: 3D adversarial pyramid anisotropic convolutional network for prostate segmentation in MR images", *IEEE transactions on medical imaging*, 39(2), 2019, 447–57.

[24] Y. Jin, G. Yang, Y. Fang, R. Li, X. Xu, Y. Liu, and X. Lai, "3D PBV-Net: an automated prostate MRI data segmentation method", *Computers in biology and medicine*, 128, 2021, 104160.

[25] S. Klein, U. A. Van Der Heide, I. M. Lips, M. Van Vulpen, M. Staring, and J. P. Pluim, "Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information", *Medical physics*, 35(4), 2008, 1407–17.

[26] C. Knoll, M. Alcañiz, V. Grau, C. Monserrat, and M. C. Juan, "Outlining of the prostate using snakes with shape restrictions based on the wavelet transform (Doctoral Thesis: Dissertation)", *Pattern Recognition*, 32(10), 1999, 1767–81.

[27] C.-C. J. Kuo, "The cnn as a guided multilayer recos transform [lecture notes]", *IEEE signal processing magazine*, 34(3), 2017, 81–9.

[28] C.-C. J. Kuo, "Understanding convolutional neural networks with a mathematical model", *Journal of Visual Communication and Image Representation*, 41, 2016, 406–13.

[29] C.-C. J. Kuo and Y. Chen, "On data-driven saak transform", *Journal of Visual Communication and Image Representation*, 50, 2018, 237–46.

[30] C.-C. J. Kuo and A. M. Madni, "Green learning: Introduction, examples and outlook", *Journal of Visual Communication and Image Representation*, 2022, 103685.

[31] C.-C. J. Kuo, M. Zhang, S. Li, J. Duan, and Y. Chen, "Interpretable convolutional neural networks via feedforward design", *Journal of Visual Communication and Image Representation*, 2019.

[32] M. van der Leest, E. Cornel, B. Israël, R. Hendriks, A. R. Padhani, M. Hoogenboom, P. Zamecnik, D. Bakker, A. Y. Setiasti, J. Veltman, *et al.*, "Head-to-head comparison of transrectal ultrasound-guided prostate biopsy versus multiparametric prostate resonance imaging with subsequent magnetic resonance-guided biopsy in biopsy-naive men with elevated prostate-specific antigen: a large prospective multicenter clinical study", *European urology*, 75(4), 2019, 570–8.

[33] Y. Li, M. Huang, Y. Zhang, S. Feng, J. Chen, and Z. Bai, "A dual attention-guided 3D convolution network for automatic segmentation of prostate and tumor", *Biomedical Signal Processing and Control*, 85, 2023, 104755.

[34] G. Litjens, O. Debats, W. van de Ven, N. Karssemeijer, and H. Huisman, "A pattern recognition approach to zonal segmentation of the prostate on MRI", in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part II 15*, Springer, 2012, 413–20.

[35] G. Litjens, N. Karssemeijer, and H. Huisman, "A multi-atlas approach for prostate segmentation in MR images", *MICCAI Grand Challenge: Prostate MR Image Segmentation*, 2012, 2012.

[36] C. Liu, M. Amodio, L. L. Shen, F. Gao, A. Avesta, S. Aneja, J. C. Wang, L. V. Del Priore, and S. Krishnaswamy, " CUTS: A Deep Learning and Topological Framework for Multigranular Unsupervised Medical Image Segmentation ", in *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, Vol. LNCS 15008, Springer Nature Switzerland, October 2024.

[37] X. Liu, F. Xing, H. K. Gaggin, W. Wang, C.-C. J. Kuo, G. El Fakhri, and J. Woo, "Segmentation of cardiac structures via successive subspace learning with saab transform from cine mri", in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, 2021, 3535–8.

[38] X. Liu, F. Xing, C. Yang, C.-C. J. Kuo, S. Babu, G. El Fakhri, T. Jenkins, and J. Woo, "Voxelhop: Successive subspace learning for als disease classification using structural mri", *IEEE journal of biomedical and health informatics*, 26(3), 2021, 1128–39.

[39] Y. Liu, G. Yang, S. Afshari Mirak, M. Hosseiny, A. Azadikhah, X. Zhong, R. E. Reiter, Y. Lee, S. S. Raman, and K. Sung, "Automatic Prostate Zonal Segmentation Using Fully Convolutional Network With Feature Pyramid Attention", *IEEE Access*, 7, 2019, 163626–32, DOI: 10.1109/ACCESS.2019.2952534.

[40] V. Magoulianitis, J. Yang, Y. Yang, J. Xue, M. Kaneko, G. Cacciamani, A. Abreu, V. Duddalwar, C.-C. J. Kuo, I. S. Gill, *et al.*, "PCa-RadHop: A transparent and lightweight feed-forward method for clinically significant prostate cancer segmentation", *Computerized Medical Imaging and Graphics*, 2024, 102408.

[41] P. Mehta, M. Antonelli, H. U. Ahmed, M. Emberton, S. Punwani, and S. Ourselin, "Computer-aided diagnosis of prostate cancer using multiparametric MRI and clinical features: A patient-level classification framework", *Medical Image Analysis*, 73, 2021, 102153.

[42] F. Milletari, N. Navab, and S. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation", in *2016 Fourth International Conference on 3D Vision (3DV)*, 2016, 565–71, DOI: 10.1109/3DV.2016.79.

[43] D. Pasquier, T. Lacornerie, M. Vermandel, J. Rousseau, E. Lartigau, and N. Betrouni, "Automatic segmentation of pelvic structures from magnetic resonance images for prostate cancer radiotherapy", *International Journal of Radiation Oncology* Biology* Physics*, 68(2), 2007, 592–600.

[44] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters–improve semantic segmentation by global convolutional network", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, 4353–61.

[45] W. Qiu, J. Yuan, E. Ukwatta, Y. Sun, M. Rajchl, and A. Fenster, "Dual optimization based prostate zonal segmentation in 3D MR images", *Medical image analysis*, 18(4), 2014, 660–73.

[46] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", 2015, arXiv: 1505.04597 [cs.CV].

[47] L. Rundo, C. Militello, G. Russo, D. DUrso, L. M. Valastro, A. Garufi, G. Mauri, S. Vitabile, and M. C. Gilardi, "Fully automatic multispectral MR image segmentation of prostate gland based on the fuzzy c-means clustering algorithm", *Multidisciplinary Approaches to Neural Computing*, 2018, 23–37.

[48] L. Rundo, C. Militello, G. Russo, A. Garufi, S. Vitabile, M. C. Gilardi, and G. Mauri, "Automated prostate gland segmentation based on an unsupervised fuzzy C-means clustering technique using multispectral T1w and T2w MR imaging", *Information*, 8(2), 2017, 49.

[49] A. Salimi, M. A. Pourmina, and M.-S. Moin, "Fully automatic prostate segmentation in MR images using a new hybrid active contour-based approach", *Signal, Image and Video Processing*, 12, 2018, 1629–37.

[50] A. Stabile, F. Giganti, M. Emberton, and C. M. Moore, "MRI in prostate cancer diagnosis: do we need to add standard sampling? A review of the last 5 years", *Prostate Cancer and Prostatic Diseases*, 21(4), 2018, 473–87.

[51] J. Sun, F. Darbehani, M. Zaidi, and B. Wang, "SAUNet: Shape Attentive U-Net for Interpretable Medical Image Segmentation", 2020, arXiv: 2001.07645 [eess.IV].

[52] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries", *CA: a cancer journal for clinicians*, 71(3), 2021, 209–49.

[53] Z. Tian, L. Liu, Z. Zhang, and B. Fei, "PSNet: prostate segmentation on MRI based on a convolutional neural network", *Journal of Medical Imaging*, 5(2), 2018, 021208.

[54] Z. Tian, L. Liu, Z. Zhang, J. Xue, and B. Fei, "A supervoxel-based segmentation method for prostate MR images", *Medical physics*, 44(2), 2017, 558–69.

[55] P. Tiwari, J. Kurhanewicz, and A. Madabhushi, "Multi-kernel graph embedding for detection, Gleason grading of prostate cancer via MRI/MRS", *Medical image analysis*, 17(2), 2013, 219–35.

[56] R. Toth, B. N. Bloch, E. M. Genega, N. M. Rofsky, R. E. Lenkinski, M. A. Rosen, A. Kalyanpur, S. Pungavkar, and A. Madabhushi, "Accurate prostate volume estimation using multifeature active shape models on T2-weighted MRI", *Academic radiology*, 18(6), 2011, 745–54.

[57] M. Trombini, D. Solarna, G. Moser, and S. Dellepiane, "A goal-driven unsupervised image segmentation method combining graph-based processing and Markov random fields", *Pattern Recognition*, 134, 2023, 109082.

[58] B. Turkbey, A. M. Brown, S. Sankineni, B. J. Wood, P. A. Pinto, and P. L. Choyke, "Multiparametric prostate magnetic resonance imaging in the evaluation of prostate cancer", *CA: a cancer journal for clinicians*, 66(4), 2016, 326–36.

[59] W. Venderink, A. van Luijtelaar, J. G. Bomers, M. van der Leest, C. Hulsbergen-van de Kaa, J. O. Barentsz, J. M. Sedelaar, and J. J. Fütterer, "Results of targeted biopsy in men with magnetic resonance imaging lesions classified equivocal, likely or highly likely to be clinically significant prostate cancer", *European urology*, 73(3), 2018, 353–60.

[60] P. Vos, J. Barentsz, N. Karssemeijer, and H. Huisman, "Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis", *Physics in Medicine & Biology*, 57(6), 2012, 1527.

[61] P. I. Wang, S. T. Chong, A. Z. Kielar, A. M. Kelly, U. D. Knoepp, M. B. Mazza, and M. M. Goodsitt, "Imaging of pregnant and lactating patients: part 1, evidence-based review and recommendations", *American Journal of Roentgenology*, 198(4), 2012, 778–84.

[62] Z. Wang, R. Wu, Y. Xu, Y. Liu, R. Chai, and H. Ma, "A two-stage CNN method for MRI image segmentation of prostate with lesion", *Biomedical Signal Processing and Control*, 82, 2023, 104610.

[63] Y. Yang, V. Magoulianitis, and C.-C. J. Kuo, "E-pixelhop: An enhanced pixelhop method for object classification", in *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, IEEE, 2021, 1475–82.

[64] Y. Yuan, E. Ahn, D. Feng, M. Khadra, and J. Kim, "Z-SSMNet: A Zonal-aware Self-Supervised Mesh Network for Prostate Cancer Detection and Diagnosis in bpMRI", *arXiv preprint arXiv:2212.05808*, 2022.

[65] D. Zhang, C. Li, H. Li, W. Huang, L. Huang, and J. Zhang, "Rethinking alignment and uniformity in unsupervised image semantic segmentation", in *Proceedings of the AAAI conference on artificial intelligence*, Vol. 37, No. 9, 2023, 11192–200.

[66] Q. Zhu, B. Du, and P. Yan, "Boundary-weighted domain adaptive neural network for prostate MR image segmentation", *IEEE transactions on medical imaging*, 39(3), 2019, 753–63.

[67] K. J. Zuiderveld, "Contrast Limited Adaptive Histogram Equalization", in *Graphics Gems*, 1994.