P-DNN: A Deep Combination of Mutilevel Features for Prostate Segmentation from MR Image

Ke Yan, Xiuying Wang, Jinman Kim, Changyang Li, Dagan Feng

School of Information Technologies, University of Sydney

1. Introduction

In this paper, we will propose an automated method which combines multilevel features in deeper manner for prostate segmentation from MR image. As the paper is preparing, we will roughly describe our algorithm and the detailed full paper will be uploaded after publication.

2. Method

Our proposed propagation deep neural network (P-DNN) adopts prostate MR transversal image as input, and directly outputs pixel-wise segmentation map to estimate the likelihood of being foreground/background pixel by pixel. To segment a 3D prostate image volume, the corresponding transverses can be fed into P-DNN iteratively and the segmentation result is the stack of the outputs from P-DNN.

2.1. Architecture of P-DNN

The proposed P-DNN is consisted of three blocks: convolution and pooling layers (CP-layers), propagation layer (P-layer), and F-measure loss layer (L-layer). The convolution layer convolves input image to produce output feature maps, with a set of learnable filters. Each filter is small spatially along width and height, thus captures and convolves local information slidingly at all locations of the input image. More deeply and intuitively, the convolution is conducted layers by layers so that the obtained final feature maps are more intrinsic and comprehensive. The pooling layers are usually added during the forward pass in the convolution layers, which decrease the resolution of the feature maps to make them less sensitive to input shift and distortions [1]. There are many off-the-shelf convolutional DNN models which can be employed in this work, including normal models (e.g. AlexNet [2] and Overfeat [3]), deeper models (e.g. VGG [4]) and extremely deeper models (e.g. ResNet [5]). Considering both the accuracy and efficiency of our method, we typically employ one of the most popular deeper models, i.e. VGG16 [4], as the base structure to form our P-DNN in this work. In order to ease the way of error backpropagating from our later proposed P-layer to CP-layers in the training phase, we tend to train an end-to-end convolutional network which directly produces the rough pixel-wise recognition maps from CP-layers. Fully convolutional networks (FCN) [6] makes such end-to-end training possible through converting the

fully connected layers of off-the-shelf DNN model into convolutional ones with 1×1 filter. By doing so, the structure of FCN-VGG16 [6] is shared to form the CP-layers.

After the CP-layers, the obtained recognized regions will propagate toward their surroundings through P-layer to have a finer segmentation map. This finer segmentation map will be then evaluated by manual labels in our proposed L-layer, and the corresponding errors will be backpropagated through the whole P-DNN to enable the filters update.

2.2. Propagation Layer

In the CP-layers, the input image and the sequent feature maps are forward passed into totally 5 pooling layers to explore high-level semantic information of image. However, such pooling process causes the obtained feature maps to be of 32× subsampled resolution, in which the image details especially the prostate boundary information may be lost. In order to solve this problem, we (1) combine the predictions from the penultimate layer of CP-layers and the POOLING4 layer as suggested in [6]; and (2) propose the propagation layer to finely delineate the prostate boundaries as follows.

While the feature maps from CP-layers are the highly convolved information of original images, we utilize appearance cues in P-layer since they are less sensitive to spatial variance but more intuitively describe the observations of original images. Specifically, the image is segmented into N superpixels via SLICO algorithm [7]. For each superpixel p, we calculate the three types of low-level cues, i.e. intensity, texture and gradient. Intensity cue of p, noted as $IC(p)$, is the intensity histogram with 32 bins, thus we have $IC(p) \in \mathbb{R}^{1\times 32}$. Texture and gradient cues of p, noted as $TC(p)$ and $GC(p)$, are described by the rotation-invariant Gabor-LBP (RGLBP) feature [8] and multi-coordinate HOG (MCHOG) feature [8] respectively. As RGLBP and MCHOG features are originally designed for biomedical image patches, we extract them from the bounding box of p to approximate $TC(p)$ and $GC(p)$. According to [8], we have $TC(p) \in \mathbb{R}^{1\times 108}$ and $GC(p) \in \mathbb{R}^{1\times 36}$. The three cues are then normalized to have a common sum value 1, and concatenated as $C(p) \in \mathbb{R}^{1\times 176}$ to represent the appearance of p.

Afterwards, an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ can be constructed over the input image IM. $\mathcal{V}$ is the node set consisted of all superpixels of IM. $\mathcal{E}$ is the edge set encoding the distance of two superpixels, which is measured by the histogram intersection using the appearance cues, and thus the associated adjacency matrix $W \in \mathbb{R}^{N\times N}$ of $\mathcal{G}$ is defined as

$$W_{ij} = \delta(\langle p_i, p_j\rangle \in \mathcal{N}) \times \left(1 - \frac{\sum_z \min(C(p_i)^z - C(p_j)^z)}{3}\right) \quad (1)$$

where $\mathcal{N}$ is the set of all neighboring superpixels pairs in IM and $C(p_i)^z$ is the value of the z-th bin in $C(p_i)$. $\delta(\cdot)$ is 1 if the condition inside the parentheses is true and 0 otherwise. Since the graph $\mathcal{G}$ and the corresponding associated adjacency matrix W are

specified, the obtained recognized regions (those superpixels with high probabilities being prostate) from CP-layers can be propagated into other regions to re-estimate the likelihood of being prostate within superpixel scale. By doing so, we have a pixel-wised propagation map G which encodes the image appearance cues on the basis of the output from the previous highly convolutional layers.

2.3. F-measure Loss Layer

The output of P-layer, denoted as $S_{out}^{P-layer}$, is the corresponding segmentation map of IM in our method. To update the learnable filters, $S_{out}^{P-layer}$ will be forward passed into a loss layer, with a particular loss function, for the calculation of the segmentation errors according to manual labels. Theoretically, most classical loss functions (such as multi-nomial logistic loss and L2 loss) can be employed here. However, as the foregrounds only cover the relatively small regions across the entire prostate MR images, either multinomial logistic loss or L2 loss may lead the model to get trapped into local minima since they treat the false-positive and false-negative pixels equally. Inspired by the F-measure theory [9] in statistical analysis, we propose an F-measure based loss function to get rid of the local minima.

References

1. Li, H., R. Zhao, and X. Wang, *Highly efficient forward and backward propagation of convolutional neural networks for pixelwise classification.* arXiv preprint arXiv:1412.4526, 2014.
2. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks*. in *Advances in neural information processing systems*. 2012.
3. Sermanet, P., et al., *Overfeat: Integrated recognition, localization and detection using convolutional networks.* arXiv preprint arXiv:1312.6229, 2013.
4. Simonyan, K. and A. Zisserman, *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556, 2014.
5. He, K., et al., *Deep residual learning for image recognition.* arXiv preprint arXiv:1512.03385, 2015.
6. Long, J., E. Shelhamer, and T. Darrell. *Fully convolutional networks for semantic segmentation*. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015.
7. Achanta, R., et al., *SLIC superpixels compared to state-of-the-art superpixel methods.* IEEE transactions on pattern analysis and machine intelligence, 2012. **34**(11): p. 2274-2282.
8. Song, Y., et al., *Feature-based image patch approximation for lung tissue classification.* Medical Imaging, IEEE Transactions on, 2013. **32**(4): p. 797-808.

9. Borji, A., et al., *Salient object detection: A benchmark.* IEEE Transactions on Image Processing, 2015. **24**(12): p. 5706-5722.