

PROSTATE SEGMENTATION WITH ENCODER-DECODER DENSELY CONNECTED CONVOLUTIONAL NETWORK (ED-DENSENET)

Yixuan Yuan^a, Wenjian Qin^b, Xiaoqing Guo^a, Mark Buyyounouski^c, Steve Hancock^c, Bin Han^c, Lei Xing^c

^a Department of Electronic Engineering, City University of Hong Kong, Hong Kong

^b Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

^c Department of Radiation Oncology, Stanford University, Stanford 94305, US

ABSTRACT

Prostate cancer is a leading cause of mortality among men. Prostate segmentation of Magnetic Resonance (MR) images plays a critical role in treatment planning and image guided interventions. However, manual delineation of prostate is very time-consuming and subjects to large inter-observer variations. To deal with this problem, we proposed a novel Encoder-Decoder Densely Connected Convolutional Network (ED-DenseNet) to segment prostate region automatically. Our model consists of two interconnected pathways, a dense encoder pathway, which learns discriminative high-level image features and a dense decoder pathway, which predicts the final segmentation in the pixel level. Instead of using the convolutional network as the basic unit in the encoder-decoder framework, we utilize Densely Connected Convolutional Network (DenseNet) to preserve the maximum information flow among layers by a densely-connected mechanism. In addition, a novel loss function that jointly considers the encoder-decoder reconstruction error and the prediction error is proposed to optimize the feature learning and segmentation result. Our automatic segmentation result shows high agreement (DSC 87.14%) to the clinical segmentation results by experienced radiation oncologists. In addition, comparison with state-of-the-art methods shows that our ED-DenseNet model is superior in segmentation performance.

Index Terms— DenseNet, Encoder-Decoder network, Prostate segmentation, reconstruction error and prediction error.

1. INTRODUCTION

According to statistics of American Joint Committee on Cancer, 164,690 new cases of prostate cancer will be examined, and 29,430 cases of cancer death will occur in the United States in 2018 [1]. Magnetic resonance imaging (MRI) is the main imaging technology to evaluate the prostate region with superior contrast and spatial resolution. Accurate delineation of the prostate gland is crucial for diagnosis, treatment planning and prognosis of the prostate cancer. However, the current segmentation of the prostate gland is mainly annotated

by physicians on a frame-by-frame basis, which is tedious, time-consuming, subject to inter- and intra-reader variations.

Recently, deep learning based computational methods [2, 3, 4, 5, 6] have been developed to automatically segment prostate region in MRI images. These methods are usually divided into two categories: patch based method and pixel based methods. The patch based method extracts image patches, trains a patch-wise convolutional neural network (CNN), and predicts the center pixel of the testing patches based on the trained models. Jia et al. [2] constructed coarse prostate segmentation with a probabilistic atlas algorithm, and then trained a set of deep CNNs to obtain fine segmentation results utilizing the extracted patches from the previous coarse boundary regions. Chen et al. [3] utilized Holistically-Nested Networks for prostate segmentation. However, the performance of patch-wise CNN models is affected by the patch size. A large patch size may decrease the localization accuracy while a small patch size could not represent complete texture information of the prostate. In addition, when the number of patches is large, there is a high redundant computation that needs to be performed for neighboring patches. The pixel based method utilized the end-to-end pixel neural network to conduct the segmentation task. Tian et al. [4] utilized the fully convolutional neural network (FCN) for prostate segmentation. The authors in [5] proposed a deeply supervised CNN utilizing the convolutional information to accurately segment the prostate from MRI images. Clark et al. [6] modified the Unet [7] by replacing the regular convolution blocks with inception reduction blocks to conduct the experiment.

The recently proposed Convolutional Encoder-Decoder Network [8] achieved superior segmentation performance compared with FCN network. The decoder layer utilizes the encoder features to calculate final segmentation maps with the same resolution of the input images. It eliminates the scale issues in the original FCN-based methods and could identify fine details of an object [8]. Moreover, the Densely Connected Convolutional network (DenseNet) [9] attracts attentions with higher classification performance compared with tradition CNN model. It connects every other preceding

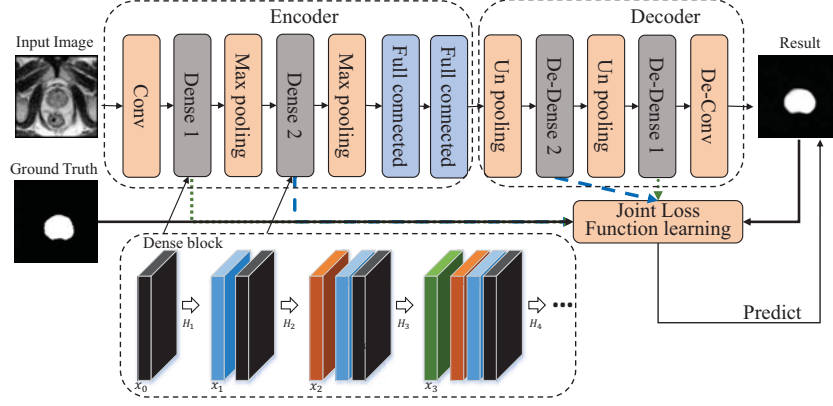


Fig. 1. Workflow of the proposed ED-DenseNet. It includes encoder-decoder procedure and a novel joint loss function. The left bottom represent the detail structure of denseblock.

layer to learn features. Thereby, it enables maximum information flow among layers in the model, strengthens feature propagation, mitigates the vanishing-gradient problem and encourages feature reuse.

Inspired by these two deep learning models, we propose a novel Encoder-Decoder Densely Connected Convolutional Network (ED-DenseNet) to segment prostate region in this paper. Instead of using convolutional layer in the Encoder-Decoder network, we utilize the DenseNet as the basic computational unit. Our network is composed of three parts: a dense encoder component that extracts features from the input MRI images, a dense decoder that utilizes the obtained features to infer a prostate mask and a joint loss function that evaluates segmentation performance of the network based on the reconstruction error and the prediction error. The encoder and decoder layers are trained together to enable the discriminative feature learning and good segmentation performance.

2. METHOD

In this paper, we proposed a novel ED-DenseNet model to segment prostate with MRI images. The corresponding workflow is illustrated in Fig. 1. Our model has an dense-encoder procedure and a corresponding dense-decoder procedure, followed by a final pixelwise classification layer. The encoder network consists of two denseblocks, which connect a layer to every other layer in a feedforward fashion. The decoder network corresponds to the encoder one and hence also has two denseblocks. The final output of the ED-DenseNet is a probability map with the same size to input image, indicating probability of each pixel that belongs to the prostate region. In the training stage, this output is fed into a novel joint loss function to iteratively evaluate prostate segmentation result and update parameters of the network. Finally, with the optimized segmentation model, we can obtain the final prostate segmentation result in the testing stage.

2.1. ED-DenseNet Model

In this work, we implement the DenseNet as the basic unit in our ED-DenseNet. DenseNet [9] is a recently proposed neural network model, which is inspired by basic ideas of the residual network [10]. It provides dense connections among all layers, in order to improve the information gradient flow within the network to enable better parameter efficiency.

2.1.1. Dense-encoder Pathway

In this work, we use a chain of DenseNet layers as the stacked encoders. Our dense-encoder pathway consists of one convolutional layer, two denseblocks, two max pooling layers and two fully connected layers. Specific, the input MRI T2-weighted (T2w) images are first fed into the convolution layer with a kernel size $k = 3$, stride $s = 1$ and zero padding $p = 2$. Then the extracted feature maps are passed through two densely-connected blocks (Fig. 1), named dense blocks. Each denseblock consists of 8 densely connections transformation layers (Fig. 1 shows 4 layers for simplicity). The max pooling with a 2×2 window is followed with each dense block to filter noisy activations by abstracting activations with a single representative value in a given receptive field. Following the last max pooling strategy, we set two fully connected layer to output the high-level coarse feature maps extracted from the images.

2.1.2. Dense-decoder Pathway

The dense-decoder pathway tends to map the calculated features in the encoder procedure to full input resolution segmentation maps, thus the encoders and decoders layers are symmetric in the proposed ED-DenseNet network. To be specific, our dense-decoder pathway includes two de-dense block, two unpooling layer and one deconvolutional layer.

The main transformation in the dense-decoder includes unpooling and deconvolutional. In the max pooling of encoder procedure, spatial information that is critical for precise prostate segmentation within a receptive field is lost. To

deal with this problem, we employ unpooling transformation (the reverse operation of pooling) in dense-decoder pathway to reconstruct the original activations. The deconvolutional transformations densify the sparse activations obtained by unpooling and output an enlarged and dense activation map. The learned filters in deconvolution can be utilized to reconstruct the shape of an input object. The filters in lower layers capture global shape of the prostate while the filters in higher layers obtain and recover the fine details.

2.2. Joint Loss Function

The proposed network is an end-to-end network from original prostate images to the segmented images. The cross entropy [2] is usually utilized to define the loss function and train the deep learning model. This measure works well if different input categories are fairly balanced. However, in our prostate segmentation problem, the background occupies the large region of the scan while the anatomy of interest (prostate) exists in only a very small region. Therefore, the learning process of this loss function usually gets trapped in local minima, and the prediction of the network tends to be biased towards background. As a result, the prostate is often missing or only partially detected.

To deal with this problem, we propose a novel joint loss function to train the ED-DenseNet model. Our loss function integrates two objectives. First, we use Dice Similarity Coefficient (DSC) to minimize the misclassification error for each pixel in the given training samples. It is represented as the ratio between the intersection and union of the predict map and ground truth. Second, we emphasize the reconstruction ability of encoder-decoder network, which enables the errors between the denseblocks in the encoder and the corresponding denseblocks in the decoder procedure to be small.

Given $p_i \in P$ presents the predicted binary segmentation image and $g_i \in G$ is the ground truth binary pixel, the joint loss function is represented as

$$L = \frac{2 \sum_i^N p_i g_i}{\sum_i^N p_i^2 + 2 \sum_i^N g_i^2} + (f(d_1) - f(e_1))^2 + (f(d_2) - f(e_2))^2, \quad (1)$$

where N is the number of pixels in the image. $f(d_j)$ and $f(e_j)$ define the learned features and the reconstructed features for the denseblock j . In our work, $j = 2$ since our model includes two denseblocks. We can easily see that the objective function defined by Eq. (1) not only minimizes the segmentation loss between the predict labels and the ground truth, but also imposes an encoder-decoder constraint information to achieve good reconstruction ability and preserve discriminative information for features. Therefore, our novel joint loss function enhances the discriminative power of the deeply learned features. In practice, we solve this optimization problem by using Adaptive Moment Estimation (Adam) method, which has been widely used in complicated optimization problems of neural networks training.

3. RESULTS

3.1. Dataset and Experiment Setup

We applied two datasets to evaluate the robustness and effectiveness of the proposed ED-DenseNet model. The first one came from PROSTATEx-2 Challenge, which contains 112 prostate T2w MRI images. The prostate MRI was performed in the Prostate MR Reference Center of the Radboud University Medical Center and the detailed description of this data set was previously recorded in [11]. The second dataset was collected from the Stanford Hospital. This dataset consists of 132 prostate cancer cases, obtained under ethical Institutional Review Board (IRB) approval. The boundary of the collected T2w images was labeled by two experienced radiologists manually. We combined these two datasets together and randomly divided them into three subsets: training, validation and testing set according to the percentage of 80:10:10 to conduct experiment. In order to enlarge training samples for the proposed ED-DenseNet model, data augmentation was implemented, including randomly rotation, scaling and cropping [12].

We implemented our model with Tensorflow and trained and tested on a desktop with 8 kernel processors (Intel Xeon(R)@3.50 GHz) and an NVIDIA GeForce 1080 with 128 GB of RAM. Each mini-batch includes 80 samples in training steps. The learning rate, momentums and weight decay of the ED-DenseNet model were set as 0.002, 0.9 and 0.005, respectively. The performance of prostate segmentation was evaluated by four common criteria: Average Symmetric Surface Distance (ASD), Hausdorff Distance (HD), Dice Similarity Coefficient (DSC) and Volume Overlap Error (VOE). The ASD and HD calculate the distance errors between the predict results and the ground truth.

3.2. Experiment Result

Qualitative segmentation results of our approach are illustrated in Fig. 2. The obtained prostate boundary of our ED-DenseNet is showed in green color while the ground truth boundary is highlighted in red. This result indicates that the segmentation obtained by the proposed method well overlaps with the ground truth. By simultaneously learning image descriptors and segmenting the prostate region in the encoder-decoder model, our method can not only successfully infer ambiguous and deficient boundaries in low contrast prostate images, but also conquer the large inter-variance of prostate shape and size. Importantly, our ED-DenseNet model is robust in distinguishing the inhomogeneous prostate tissues, and recognizing them as a whole part.

We then analyzed the effects of proposed items for the prostate segmentation. The first comparison experiment was Encoder-Decoder convolutional Network, which utilized the traditional convolutional layer to substitute our introduced DenseNet module for prostate segmentation. The second one applied the ED-DenseNet method with cross entropy to segment the prostate region instead of utilizing the joint loss

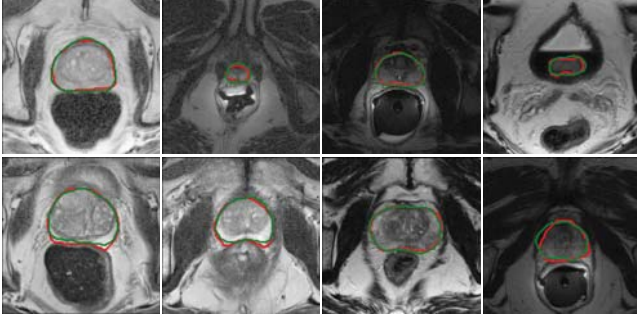


Fig. 2. The qualitative results of the proposed ED-DenseNet model. The green curves represent our segmentation results, while the red curves show the ground truths obtained from manual segmentation.

function. The corresponding criteria of segmentation results are recorded in Table 1. The average VOE of our method is 16.17%, demonstrating that the segmentation result of the proposed method achieves a good balance between over-segmentation and under-segmentation. Our method achieves an inspiring ASD 2.23 and shows significant improvements compared with the other two comparison experiments. This result validates the introduced DenseNet and joint loss function have critical roles in learning discriminative features for improving segmentation performance.

We further compared our ED-DenseNet model with the state-of-the-art segmentation models: FCN, U-Net and the method in [13]. We utilized our collected prostate datasets to implement these methods and the criteria obtained by existing methods and ours are shown in Table 1, respectively. The average of DSC values of our model, FCN and U-Net are 87.14%, 78.84% and 85.46%, respectively. The relative low standard deviation of the computed four metrics demonstrates the robustness and repeatability of the ED-DenseNet model. So it can be proved that the proposed method obtains significant improvement on the prostate segmentation compare with the other methods. This improvement can be attributed to the deeper network with the DenseNet and the novel proposed joint loss function.

Table 1. Comparison with different methods.

	ASD (mm)	HD (mm)	DSC (%)	VOE (%)
Baseline 1	3.19 ± 2.28	8.67 ± 5.89	81.85 ± 11.49	29.13 ± 16.49
Baseline 2	4.02 ± 1.85	9.16 ± 3.67	75.42 ± 12.11	37.99 ± 15.24
Our Method	2.23 ± 1.06	6.12 ± 2.16	87.14 ± 6.63	16.17 ± 7.55
FCN	3.76 ± 1.51	8.91 ± 3.21	78.84 ± 7.44	34.31 ± 10.14
U-Net	2.76 ± 1.09	7.71 ± 2.19	85.46 ± 5.86	24.94 ± 8.81
[13]	2.49 ± 1.32	7.21 ± 2.35	86.09 ± 6.43	21.47 ± 8.29

4. CONCLUSION

We proposed a novel ED-DenseNet model to automatically segment prostate region with MRI T2w images. Instead of utilizing the convolutional layer as the basic unit in the deep learning model, we are the first to introduce the DenseNet

to the Encoder-Decoder framework. We further proposed a novel loss constraint to evaluate the segmentation result and optimize the feature learning. It includes the Dice overlap coefficient evaluating the errors between the predicted prostate segmentation results and the ground truth annotation, the encoder-decoder error that emphasizes the reconstruction ability. This loss function could address well for the segmentation problem when the amount of background and prostate pixels is strongly unbalanced. Our results suggest that the ED-DenseNet model can substantially achieve high segmentation accuracy and produce the highest criteria when compared to the state-of-the-art approaches. Therefore, our model could further provide valuable information for treatment planning and precision radiation therapy.

5. REFERENCES

- [1] American Cancer Society, “Key statistics for prostate cancer,” <https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html/>, Accessed Feb 11, 2018.
- [2] Haozhe Jia, Yong Xia, Weidong Cai, Michael Fulham, and David Dagan Feng, “Prostate segmentation in mr images using ensemble deep convolutional neural networks,” in *IEEE ISBI*, 2017, pp. 762–765.
- [3] Ruida Cheng, Holger R Roth, Nathan Lay, Le Lu, Baris Turkbey, William Gandler, Evan S McCreedy, Peter Choyke, Ronald M Summers, and Matthew J McAuliffe, “Automatic mr prostate segmentation by deep learning with holistically-nested networks,” in *IPMI*, 2017, vol. 10133, p. 101332H.
- [4] Zhiqiang Tian, Lizhi Liu, Zhenfeng Zhang, and Baowei Fei, “Psnnet: prostate segmentation on mri based on a convolutional neural network,” *J. Med. Imaging*, vol. 5, no. 2, pp. 021208, 2018.
- [5] Qikui Zhu, Bo Du, Baris Turkbey, Peter L Choyke, and Pingkun Yan, “Deeply-supervised cnn for prostate segmentation,” in *IEEE IJCNN*, 2017, pp. 178–184.
- [6] Tyler Clark, Alexander Wong, Masoom A Haider, and Farzad Khalvati, “Fully deep convolutional neural networks for segmentation of the prostate gland in diffusion-weighted mr images,” in *ICIAR*. Springer, 2017, pp. 97–104.
- [7] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *MICCAI*. Springer, 2015, pp. 234–241.
- [8] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han, “Learning deconvolution network for semantic segmentation,” in *IEEE CVPR*, 2015, pp. 1520–1528.
- [9] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten, “Densely connected convolutional networks,” in *IEEE CVPR*, 2017, pp. 4700–4708.
- [10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” in *IEEE CVPR*, 2016, pp. 770–778.
- [11] Geert Litjens, Oscar Debats, Jelle Barentsz, Nico Karssemeijer, and Henkjan Huisman, “Computer-aided detection of prostate cancer in mri,” *IEEE Trans. Med. Imag.*, vol. 33, no. 5, pp. 1083–1092, 2014.
- [12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolutional neural networks,” in *NIPS*, 2012, pp. 1097–1105.
- [13] Tom Brosch, Jochen Peters, Alexandra Groth, Thomas Stehle, and Jürgen Weese, “Deep learning-based boundary detection for model-based segmentation with application to mr prostate segmentation,” in *MICCAI*. Springer, 2018, pp. 515–522.