Prostate T2 MRI segmentation with Convolutional Neural Networks

García-Ferrando, Germán A.
gergarfe@upv.es

Juan-Albarracín, Javier
jajuaal1@ibime.upv.es

Fuster-García, Elies
elfusgar@mat.upv.es

García-Gómez, Juan Miguel
juanmig@ibime.upv.es

December 2017

Abstract

We propose a two steps model with Convolutional Neural Networks for the segmentation of the prostate in T2 weighted MRI. The first step receives the T2 images and generates a coarse heat map of the prostate location. The second step uses the heat map and the T2 image to perform the final segmentation. In order to train the model we used the dataset provided by the PROMISE12 challenge.

1 Method

1.1 Preprocess

Firstly, we resize all the volumes to have spacing $0.625 \times 0.625 \times 1.5$ (1; 2), then each volume is normalized to have average zero and standard deviation one (3).

We loop over all the volumes extracting patches with size $128 \times 128 \times 16$ and $64 \times 64 \times 16$, with an step size of $32 \times 32 \times 8$ and $16 \times 16 \times 4$ respectively. Then, all the patches where at least the 1% of the pixels are prostate, are classified as *positive*, while all the others are classified as *negative*. We use this classification to balance the representation of both classes in each batch (4; 5).

Finally, we perform data augmentation on the fly, having each batch a probability of 50% of being rotated (90°, 180°, 270°) or flipped among the vertical axis in the axial plane.

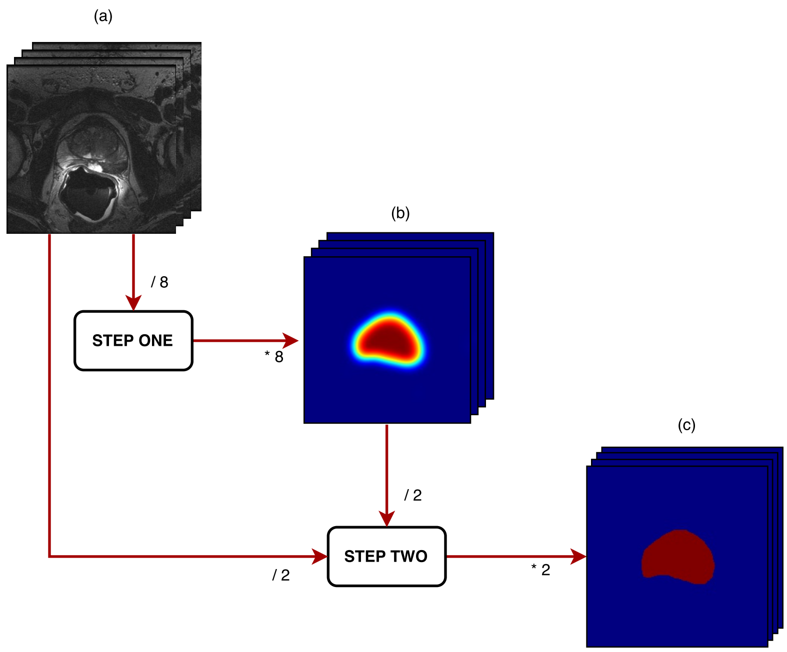

Figure 1: General description of the proposed model. Note that in the step one only the T2 image (a) is used, whilst in the step two both, the T2 image (a) and the heat map (b) are used in order to generate the final segmentation (c).

1.2 Model architecture

The proposed model consists in two steps. *Firstly*, a Convolutional Neural Network (CNN) receives a patch of $128 \times 128 \times 16$ from the T2 images, reduces its resolution by a factor of 8, and then performs a prediction. The generated map is then resized by bilinear upsampling to its original resolution, and a Gaussian filter is applied in order to smooth the probabilities. *Secondly*, another CNN receives a patch of $64 \times 64 \times 16$ with the T2 images and the previous heat map; then reduces its resolution by a factor of 2 and performs a prediction. Finally the prediction is upsampled in order to match the original resolution. The Figure 1 shows a graphical definition of the model.

The first step consists in a CNN with the next architecture: one first convolution with a kernel size of $11 \times 11 \times 11$, stride of $8 \times 8 \times 8$ and 32 channels is applied to the input patches. It is followed by three Residual Blocks (6) with kernel size of $3 \times 3 \times 3$, stride of $1 \times 1 \times 1$ and 32 channels. Then a last convolution is performed in order to get the final 2 maps. Finally the maps are upsampled with bilinear interpolation to the original resolution.

The second step consists in a CNN with the next architecture: one first convolution with a kernel size of $11 \times 11 \times 11$, stride of $2 \times 2 \times 2$ and 32 channels is applied to the input. It is followed by four Residual Blocks (6) with kernel size of $3 \times 3 \times 3$, dilation rate of $2 \times 2 \times 1$ and 32 channels. Each two residual block we perform a long residual connection. A last convolution is performed in order to get the final 2 maps. Finally the maps are upsampled with bilinear interpolation to the original resolution and the largest connected region is selected.

We follow the standard scheme of Batch Normalization + Convolution + ReLU in both architectures.

2 Evaluation

2.1 Training protocol

The same training protocol have been followed in both networks. The networks were trained with Adam Optimizer with a mini-batch size of 16 and equal representation of *negative* and *positive* samples. The initial learning rate was set to 1×10^{-3}, and the models were trained for 15×10^{3} iterations. L2 regularization was set to 5×10^{-4}. Both models were trained using cross entropy as cost function. In test phase, we used overlapped sliding windows to merge all the probability maps generated for an specific volume. The stride used was $64 \times 64 \times 8$ for the first step and $16 \times 16 \times 4$ for the second one.

2.2 Results

To check the model behavior we split the dataset into 5 cases for validation and 45 cases for training. The 5 cases of validation were Case(03/04/18/23/42). We obtained a Dice average of 95.14 in the training set, and 86.25 in the validation set.

Acknowledgment

We are grateful to the NVIDIA corporation for supporting our research. We are using a donated Titan Xp GPU to train our models.

References

[1] 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 *AAAI*, pp. 66–72, 2017.

[2] F. Milletari, N. Navab, and S. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *CoRR*, vol. abs/1606.04797, 2016.

[3] Y. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural networks: Tricks of the trade*, pp. 9–50, Springer, 1998.

[4] K. Kamnitsas, C. Ledig, V. F. J. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3d CNN with fully connected CRF for accurate brain lesion segmentation," *CoRR*, vol. abs/1603.05959, 2016.

[5] K. Kamnitsas, W. Bai, E. Ferrante, S. G. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. C. H. Lee, B. Kainz, D. Rueckert, and B. Glocker, "Ensembles of multiple models and architectures for robust brain tumour segmentation," *CoRR*, vol. abs/1711.01468, 2017.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015.