3D Prostate Segmentation from MR Images using FCNN

Aman Agarwal, Aditya Mishra, and Priyanka Sharma

Institute of Technology, Nirma University, Ahmedabad
{15bce006, 15bce003, priyanka.sharma}@nirmauni.ac.in

Abstract. This paper presents a novel architecture using a fully convolutional neural network (FCNN) for prostate segmentation has been proposed. The FCNN model presented here is a modified version of V-net with smart factorization of convolutional layers for reducing the number of parameters, and Batch Normalization for better and faster convergence. Unlike previous approaches which focus on segmentation using 2D slices, whole volume in 3D has been considered which saves time during inference and additionally, the algorithm can take advantage of the 3D spatial information. Various on-the-fly augmentation methods have been applied to prevent overfitting of the network. The model achieved a Dice Similarity Coefficient (DSC) of 0.92 and 0.80 on training and validation set respectively.

Keywords: Prostate, Segmentation, FCNN, Dice Similarity Coefficient, Batch Normalization.

1 Data Preparation

The dataset was taken from PROMISE12 challenge consisting of 50 volumes from various sources. They were first brought to the same shape. Each slice was converted to size 128x128, and the number of slices was fixed to 64 using interpolation. This made the size of each volume as 128x128x64. The spacing was set to 1.0, 1.0, 1.5. The intensities of volumes were normalized in the range of 0 to 1, and then the dataset was split into training and validation set using an 80:20 ratio. To prevent overfitting during training, data augmentation methods were used including, random flipping, rotation, intensity smoothing, noise, and histogram matching with random images from the training set.

2 Model Architecture

The model architecture was inspired by V-net [1], which uses an encoder-decoder structure involving downsampling and upsampling layers to limit the number of parameters [3]. A Convolution-Batchnorm-PReLu (CBP) block is defined, which is used repetitively in this paper, consisting of two 3x3x3 convolutions followed by Batch Normalization [4] and PReLu activation [5]. Two 3x3x3 convolution

gives the same receptive field as one 5x5x5 convolution but with less number of parameters as suggested in [6]. All convolutions in these blocks were applied with appropriate padding.

The architecture was analogous to a multi-storeyed building with each floor consisting of many CBP blocks. Downsampling was performed between floors using strided convolutions [7] with stride size 2x2x2, thereby, reducing the volume by a factor of 2 each time.

After getting an intermediate representation of size 8x8x4, upsampling was performed using transposed convolutions [9]. During upsampling, corresponding encoder layers were concatenated with the decoder layers to promote feature forwarding [10]. Same CBP blocks were used in these floors as well. As suggested in [8], a skip connection was added before every downsampling and upsampling layer to learn a residual function which helps in training deep networks. The output of the last convolutional layer was passed through a 1x1x1 convolution followed by sigmoid activation function for detecting background and foreground voxels.

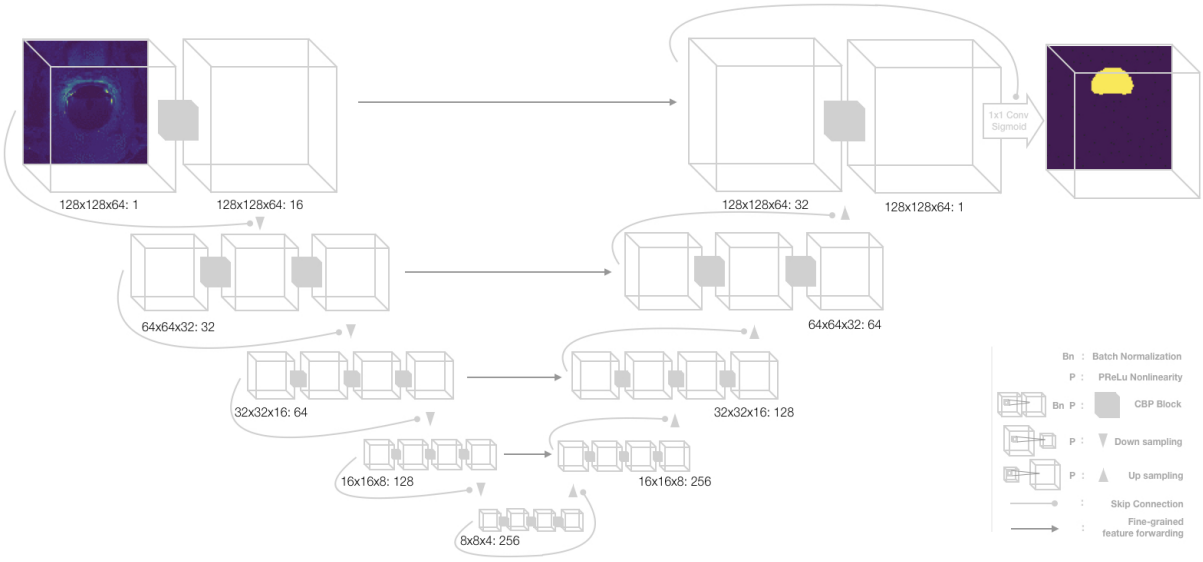

Fig. 1. A schematic representation of our FCNN model architecture.

3 Training

For training the network, Adam [11] optimizer was applied. The base learning rate was set to $1e-6$ in the beginning and then changed during the course of training. The batch size was fixed to 5 volumes. The weights of the network were

updated according to DSC, also known as Dice loss, given in equation 1.

$$DSC = \left(\frac{2 * \sum(X.Y)}{\sum X + \sum Y + \epsilon} \right) \tag{1}$$

where, X and Y are the predicted and ground truth volumes respectively, ϵ is the smoothing parameter set to 0.00001.

4 Results

After training for 5000 epochs on NVIDIA P100 GPU system, An average DSC of 0.92 and 0.80 was achieved on the training and validation set respectively. The results were evaluated on additional metrics [2] viz. Absolute Relative Volume Difference (ARVD) [12], Average Boundary Distance (ABD) [12], and 95% Hausdorff distance (95HD) [13]. Fig. 2 shows the mean and standard deviation of 95HD, Dice loss, ABD, and ARVD on training and validation sets. Fig. 3 shows the segmentation results on the test set.

Training		Validation	
Hausdorff	: 2.17 ± 0.82	Hausdorff	: 11.91 ± 18.63
Dice	: 0.92 ± 0.02	Dice	: 0.80 ± 0.10
Boundary dist	: 0.80 ± 0.27	Boundary dist	: 2.54 ± 2.01
Volume diff	: 5.34 ± 4.71	Volume diff	: 12.50 ± 54.34

Fig. 2. Mean and standard deviation on various metrics for training and validation set.

Fig. 3. Segmentation results on test dataset for (a) Case05, (b) Case15, and (c) Case25 (from left to right).

5 Acknowledgement

We would like to thank Institute of Technology, Nirma University for providing us with the resources required to carry out this work.

References

1. Milletari, Fausto, Navab, Nassir & Ahmadi, Seyed-Ahmad. (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation.
2. Litjens, G.J., Toth, R., Ven, W.J., Hoeks, C., Kerkstra, S., Ginneken, B.V., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., Strand, R., Malmberg, F., Ou, Y., Davatzikos, C., Kirschner, M., Jung, F., Yuan, J., Qiu, W., Gao, Q., Edwards, P.J., Maan, B., Heijden, F.V., Ghose, S., Mitra, J., Dowling, J., Barratt, D.C., Huisman, H.J., & Madabhushi, A. (2014). Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge. Medical image analysis, 18 2, 359-73.
3. Evan Shelhamer, Jonathan Long, and Trevor Darrell. (2017). Fully Convolutional Networks for Semantic Segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39, 4 (April 2017), 640-651. DOI: https://doi.org/10.1109/TPAMI.2016.2572683
4. Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ICML.
5. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015 IEEE International Conference on Computer Vision (ICCV), 1026-1034.
6. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2818-2826.
7. Springenberg, J.T., Dosovitskiy, A., Brox, T., & Riedmiller, M.A. (2014). Striving for Simplicity: The All Convolutional Net. CoRR, abs/1412.6806.
8. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.
9. Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning. CoRR, abs/1603.07285.
10. Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., & Pal, C.J. (2016). The Importance of Skip Connections in Biomedical Image Segmentation. LABELS/DLMIA@MICCAI.
11. Kingma, Diederik & Ba, Jimmy. (2014). Adam: A Method for Stochastic Optimization. International Conference on Learning Representations.
12. Heimann, T., van Ginneken, B., Styner, M., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., Bello, F., Binnig, G., Bischof, H., Bornik, A., Cashman, P., Chi, Y., Cordova, A., Dawant, B., Fidrich, M., Furst, J., Furukawa, D., Grenacher, L., Hornegger, J., Kainmuller, D., Kitney, R., Kobatake, H., Lamecker, H., Lange, T., Lee, J., Lennon, B., Li, R., Li, S., Meinzer, H.P., Nemeth, G., Raicu, D., Rau, A.M., van Rikxoort, E., Rousson, M., Rusko, L., Saddi, K., Schmidt, G., Seghers, D., Shimizu, A., Slagmolen, P., Sorantin, E., Soza, G., Susomboon, R., Waite, J., Wimmer, A., & Wolf, I., (2009). Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE Trans. Med. Imaging 28, 12511265.

13. Chandra, S.S., Dowling, J.A., Shen, K.K., Raniga, P., Pluim, J.P.W., Greer, P.B., Salvado, O., & Fripp, J., (2012). Patient specific prostate segmentation in 3-D magnetic resonance images. IEEE Trans. Med. Imaging 31, 19551964.