An End-to-end Approach with CNN and Posterior-CRF in White Matter Hyperintensities Segmentation

Shuai Chen[1], Marleen de Bruijne[1,2]

[1] Erasmus MC, Biomedical Imaging Group Rotterdam, Departments of Radiology and Nuclear Medicine, The Netherlands
[2] University of Copenhagen, Machine Learning Section, Department of Computer Science, Denmark

1 Data

We use 60 MRI scans(T1 and FLAIR) from *WMH 2017 Challenge*. The images were acquired from three hospitals and manually annotated with three labels: background, white matter hyperintensities and other pathology. The first two labels are used as ground truth. We use 54 scans for training and 6 scans for validation.

2 Preprocessing

2.1 Patches extraction

Due to the limitation of GPU memory, we use patches of the size 200*200*16 voxels as the input to our networks. The patches are extracted from the 3D region of interest (cropping the background outside the center of each slice in axial direction, from the size 240*240*16 into 200*200*16) of the 3D MR scans, which contains most of the whole brain. There is 75% overlap between neighbor patches in training.

2.2 Data augmentation

Several 3D data augmentation strategies were applied on the training patches, including 3D rotation, shifting, as well as flipping in 3 directions (axial, sagittal, coronal). For detailed parameters, rotation in 3 directions are [10, 5, 5] degrees, shifting range in 3 directions are [20, 20, 7] voxels.

3 Method

3.1 3D UNet

We use 3D UNet [2] as the baseline architecture in this paper. The details of the network can be found in Fig 1. All convolution layers in UNet use *ReLU*

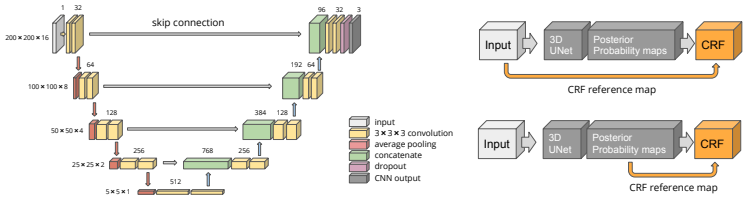

Fig. 1: **End-to-end training networks.** For each graph: 3D UNet baseline (left), Intensity-CRF (upper right) and Posterior-CRF Neural Network (lower right).

as activation function except for the last output layer, which use *softmax* to produce the final CNN probability maps. We use categorical cross-entropy as the loss function.

3.2 Posterior-CRF

In the fully-connected CRF model (**X**,**I**), the corresponding Gibbs energy w.r.t the label segmentation **x** is

$$E(\mathbf{X}=\mathbf{x}|\mathbf{I}) = \sum_i \varphi_u(x_i|\mathbf{I}) + \sum_{i<j} \varphi_p(x_i, x_j|\mathbf{I}) \tag{1}$$

where i and j range from 1 to N, which is the number of voxels in the random field **X** and 3D input patch **I**. For convenience, the conditioning on **I** will be omitted in the rest of the paper. The first term $\varphi_u(x_i)$ is the unary potential, which is set to be the CNN posterior probability maps. The second term $\varphi_p(x_i, x_j)$ is the pairwise potential:

$$\varphi_p(x_i, x_j) = \mu(x_i, x_j)[\omega^{(1)}\mathbf{exp}(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|f_i - f_j|^2}{2\theta_\beta^2}) + \omega^{(2)}\mathbf{exp}(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2})] \tag{2}$$

where $\mu(x_i, x_j)$ is the label compatibility function that captures the compatibility between different pairs of labels. ω is the linear combination weight of different predefined kernels.

The first kernel in Eq 2 is defined by the positions vectors p_i and p_j and feature vectors f_i and f_j, and the second kernel is the *smoothness kernel* which is only controlled by the voxel positions. θ_α, θ_β and θ_γ are the parameters that control the sensitivity to the corresponding feature space. In the previous methods [3], people usually use the intensity I of the input image as the feature (or reference map) f, which we call Intensity-CRF methods (Fig 1). However, the Intensity-CRF is very sensitive to the parameter θ_β because the intensity varies a lot between different medical images as well as the random noise. Therefore, we replace the intensity I by the posterior probability x as the new reference maps, which we call Posterior-CRF method [1] (Fig 1).

References

1. Chen, S., de Bruijne, M.: An end-to-end approach to semantic segmentation with 3d cnn and posterior-crf in medical images. arXiv preprint arXiv:1811.03549 (2018)
2. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
3. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.: Conditional random fields as recurrent neural networks. In: Proceedings of the IEEE international conference on computer vision. pp. 1529–1537 (2015)