Augmented V-Net for White Matter Hyperintensities segmentation

Adrià Casamitjana, Marc Combalia, Irina Sánchez and Verónica Vilaplana

Signal Theory and Communications Department, Universitat Politècnica de Catalunya. BarcelonaTech, Spain
{adria.casamitjana,veronica.vilaplana}@upc.edu

1 Method

In this work, we propose an extension of the V-Net architecture [1] called Augmented V-Net. The main changes with respect to the V-Net model are the following:

- *Augmented path:* we use an upsampled version of the input to exploit high resolution features. This is done by upsampling by repetition the input (factor of 2) and stacking several convolutional layers after the upsampling. The resultant features are concatenated in the last layers of the network.
- *Modified residual connections:* the residual connections are reformulated such that the propagation of the input signal through the network is minimally modified.
- *Mask:* a mask is used before the final prediction in order to constrain the network to train on relevant voxels.
- *Input concatenation*: the raw input image is used as feature map in the last stages of the network.

The key part of the network is the augmented path, which has been shown to boost the performance of the standard V-Net for detecting white matter hyperintensities. It provides high-resolution features by keeping small filter sizes and adding redundancy in the input, helping to detect finer regions such as boundaries. Later in the network, we use the input image as raw features, since voxel's intensities already contain valuable information. Finally, the mask is used to train/predict only on voxels of brain tissue. No further post-processing is used.

2 Training

Flair and T1 modalities are used in the system. Both are normalized to zero mean and unit variance. From the normalized Flair image, a mask is created to mask-out background voxels.

[*] This work has been partially supported by the projects BIGGRAPH-TEC2013-43935-R and MALEGRA TEC2016-75976-R financed by the Spanish Ministerio de Economía y Competitividad and the European Regional Development Fund (ERDF). Adrià Casamitjana is supported by the Spanish "Ministerio de Educación, Cultura y Deporte" FPU Research Fellowship.

[1] Milletari et al. https://arxiv.org/abs/1606.04797, 2016

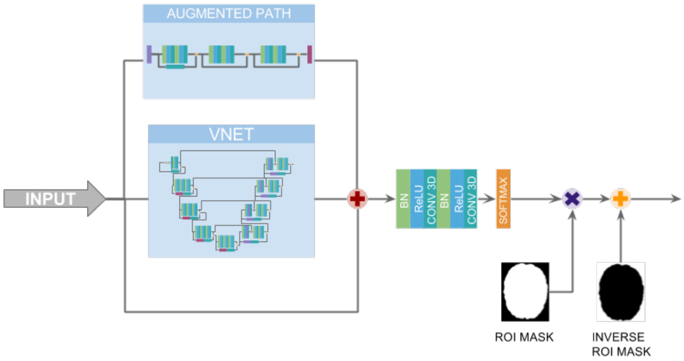

Fig. 1: Masked VNet

When training such a big and deep network, we face two main problems: GPU memory constraints and the scarcity of data. Patch-wise training arises as a possible solution for the first issue. The memory required to train Augmented V-Net does not allow using dense-training, which is also discouraged when data is scarce. Then, patch-wise training is the only solution. Larger patches are preferred because they can encode localization features (brain structures) across the network, while smaller patches allow increasing the batch size in the optimization process. For this task in particular, we use a mixed strategy to obtain benefits from both techniques. We first train the network using patches of size 32x32x32 and sample patches such that 60% of patches are centered in a WMH region, while the other 40% are centered in a non-lesion brain tissue. Then we fine-tune the network using increasingly growing patches (64x64x32 and 96x96x32) in order to benefit from structural information in the brain. This last step is optimized using the Dice as loss function, sampling only those patches centered in WMH. Note that due to the reduced number of slices (z-dimension) in the FLAIR modality, we increase the patch size only in the axial plane. We use data augmentation to increase the size of the training set, by making sagittal reflections of each subject. In the optimization process, we use Adam optimizer with initial learning rate of $LR = 0.0005$.

3 Inference

In inference time, we can use the whole subject, performing dense inference and using the mask to indicate foreground. The method is fully automatic, taking from 5 to 7 seconds to infer each subject.