A Multi-view Fully Convolutional Neural Network for MR Brain Segmentation

Mauricio Orbes Arteaga[a,b,c],Mathias perslev[a], Akshay Pai[a,b,c]

[a]DIKU, University of Copenhagen, Denmark;

[b]Biomediq A/S, Copenhagen, Denmark [c]Cerebriu A/S, Copenhagen, Denmark

ABSTRACT

For fully convolutional neural networks (FCNs), the transition from 2D to 3D is computationally expensive and bears the risk of overfitting. We present a training and evaluation framework for multi-view fully convolutional networks(FCNs) applied to MR brain segmentation. To Achieve good generalization behavior, we combine the parameter efficiency of a 2D FCN with a systematic train-and test-time augmentation framework, which allows the 2D model to leverage a more information rich representation of the 3D image volume. To this end, we fit a single FCN to 2D image slices sampled along multiple viewing axes spanning the 3D Volume simultaneously. At the test time, the model predicts along each of the viewing directions used at training time. The resulting predictions are combined into a final segmentation using majority voting.

1. MULTI-VIEW MODEL

We propose a general semantic segmentation approach for medical image volumes that relies on a single 2D fully convolutional network (FCN) model fit image slices sampled along multiple viewing axes through the image volume. At test time, the model predicts along each view and a majority voting scheme merges the multiple volumes into one final segmentation. The proposed approach is illustrated in Figure 1.

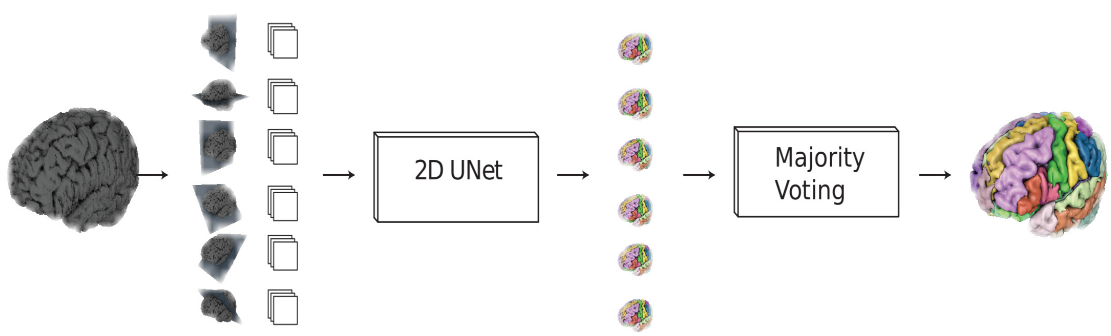
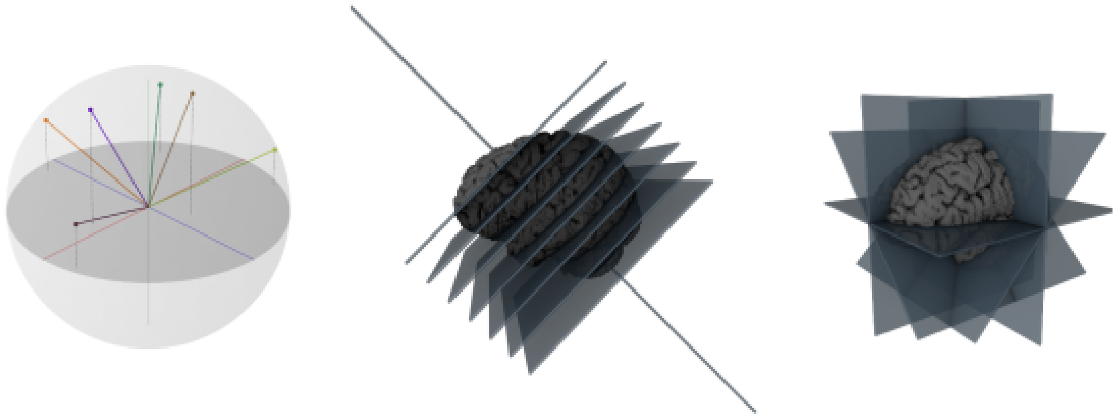


Figure 1: Model overview. In the inference phase, the input volume(left) is sampled on 2D isotropic grids along multiple view axes. The model predicts a full volume along each axis and maps the prediction into the original image space. A majority voting scheme combines the 6 proposed segmentation volumes into a single final segmentation

.

We use a FCN base on the **U-net** architecture as in [1]. The model $f(x, \theta)$ takes as input 2D image slices of C channels $x \in \mathbb{R}^{w \times h \times C}$ and outputs a probabilistic segmentation map $p \in \mathbb{R}^{w \times h \times K}$ by defining a 1×1 output convolution layer with K filters for predicting K classes. Prior to training we define a set $V = v_1, v_2, v_i$ of i randomly sampled unit vectors in $\mathbb{R}^3$. The set defines the axes through the image volume along which we sample 2D image plane inputs to the model, see figure 2. The model predicts along each view producing a set of i segmentation volumes P



(a) (b) (c)

Figure 2: (a) Visualization of a set V of sampled view axis unit vectors. (b) Illustration of images sampled along one view. (c) Illustration of multiple images sampled along multiple unique views.

2. AUGMENTATION

In addition to the extensive (affine) data augmentation, we also include non-linear augmented images in the training set. For each image in the dataset, we created 3 new volumes using random b-spline deformations. These were obtained by using a deformation field defined on a grid of $6 \times 6 \times 6$ number of control points and b-spline interpolation. The size of deformation was controlled by a Gaussian noise of with mean 0 and standard deviation Sd which is used to perturb the parameters of the deformation field. Specifically, we use $Sd = 8$.

.

References

[1] H. Li, G. Jiang, R. Wang, J. Zhang, Z. Wang, W.-S. Zheng, and B. Menze, "Fully convolutional network ensembles for white matter hyperintensities segmentation in mr images," *arXiv preprint arXiv:1802.05203* , 2018.