Brain Structure Segmentation using Adversarial Learning

Motivation: Semantic Segmentation in medical imaging is especially used for identification of disorders or diseases. Segmentation of MRI scans is quite challenging due to poor contrast, artifacts, and subject variations. In cases where limited annotations are available, pre-trained networks are used. However, adopting this strategy to segment medical images is challenging since distribution of MRI scans is quite different from natural images. This work aims to leverage unannotated MRI scans to improve the segmentation results. We aim to alleviate the need for large sets of labeled samples by exploiting available non-annotated data. Recently, approaches based on GANs, have shown great potential for improving semantic segmentation in a semi-supervised setting. We apply them on 3D medical images (multiple modalities) to improve the segmentation results.

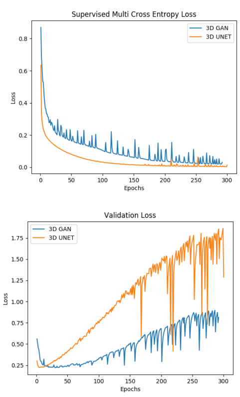
Dataset: We evaluate our results on brain MRI scans. Two datasets are used. The annotated dataset is provided by MRBrainS18 [**1**]. It consists of 7 sets of annotated brain MR images (T1, T1-IR, and T2-FLAIR) with manual segmentations by experts. These subjects include patients with diabetes, dementia and Alzheimers, and white matter lesions (age $\geq$ 50). The segmentation task is done on 8 classes : Cortical gray matter, Basal ganglia, White matter, White matter lesions, Cerebrospinal fluid, Ventricles, Cerebellum, Brain stem. The unannotated dataset is provided by WMH Segmentation Challenge [**2**]. This data consists of brain MR images (T1 and T2-FLAIR). So we used only two modalities (FLAIR and T1) to train our network. We used total of 8 subjects (4 annotated and 4 unannotated) for training, 1 annotated subject for validation and 2 annotated subjects for testing.

Pre-processing: This is done to remove noise and reduce variation across subjects. Each scan is of size 240 $\times$ 240 $\times$ 48. To pre-process, every scan is first bias field corrected using the N4ITK algorithm. After this, the scans are cropped such that 10 pixels from each side is removed to get rid of black background borders. Then the scans are Z-normalized. After pre-processing, the final dimension of each scan is 220$\times$220$\times$ 48.

Technical Details: Training is done on fixed size patches of size 32 $\times$ 32 $\times$ 32 as scans are too big to fit into memory. While training patches have fixed size, test images may have arbitrary size. To address this issue the segmentation network is fully-convolutional. To compare the results when unannotated images are also used to train the network, a baseline is required to evaluate segmentation quality. For baseline, 3D-UNet network was trained with multi-class entropy loss. To include unlabelled images during training, GAN is used. The task of the discriminator(segmentation) network is to determine whether the image patch is real (labelled or unlabelled) or fake (generated by the generator). So, the discriminator predicts (K+1) classes, where K is the number of class labels for the real samples and the additional class corresponds to fake samples from the generator. Given a 3D image patch $x_{H\times W\times D}$, the network predicts segmentation mask $y_{H\times W\times D\times(K+1)}$. The loss for discriminator L_{disc} is given by equation below[**3**]. Similar number of labelled, unlabelled and fake images are used for equal importance in training. For training generator network, we just minimize fake loss L_{fake}. We used two metrics to evaluate the quality of generated segmentation masks : Dice Score and Volumetric Symmetry.

$$L_{disc} = L_{labelled} + L_{unlabelled} + L_{fake}, L_{labelled} = -E_{x,y\sim p_{data}(x,y)} \sum_{i=1}^{H\times W\times D} log[p_{model}(y_i|x, y_i < K+1)],$$
$$L_{unlabelled} = -E_{x\sim p_{data}(x)} \sum_{i=1}^{H\times W\times D} log[1 - p_{model}(y_i = K+1|x)],$$
$$L_{fake} = -E_{z\sim Noise} \sum_{i=1}^{H\times W\times D} log[p_{model}(y_i = K+1|G_{\theta_G}(z))]$$

It is evident from the validation loss curves that in case of 3D-UNET (orange), due to limited training data (only 4 MRI scans), the network overfits within first few epochs as the loss increases very rapidly. But for 3D-GAN (blue), even though the training curve has some fluctuations but validation curves is relatively more stable.



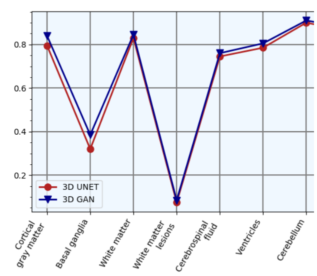


Figure 1: Dice Score

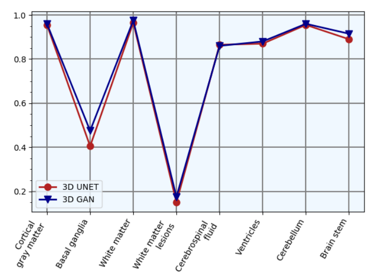


Figure 2: Volume Similarity

5. Conclusions: We used GAN in a semi-supervised setting to improve segmentation results. Results in figure 1 and 2 show that there predicted segmentation masks are slightly improved.

6. References:

[1] *Grand Challenge on MR Brain Segmentation at MICCAI 2018*, MRBrainS18.

[2] *WMH Segmentation Challenge. 2017*, WMH Segmentation Challenge.

[3] *Few-shot 3D Multi-modal Medical Image Segmentation using Generative Adversarial Learning*, arXiv:1810.12241.