Medical Image Segmentation and Applications:

Brain tissue segmentation

Authors: Valerio Di Sano, Zafar Toshpulatov, Antoine Merlet

Advisors: Pr. Robert Marti, Pr. Xavier Llado

January 11, 2019

All the work presented and its results can be viewed at `https://drive.google.com/open?id=1mVCaHIlUJ8ixB05esgwYHPfwI7IHpjea`

I. INTRODUCTION

Brain tissue segmentation is one of the most sought after research areas in medical image processing. It provides detailed quantitative brain analysis for accurate disease diagnosis, detection, and classification of abnormalities. It plays an essential role in discriminating healthy tissues from lesion tissues. Therefore, accurate disease diagnosis and treatment planning depend on the performance of the segmentation method used [1]. In the brain MRI image, the gray matter (GM) is basically the cortex, the white matter (WM) lies mainly under it. The cortex is the place where all the higher mental processing takes place. Cerebrospinal fluid (CSF) is a clear, colorless body fluid found in the brain and spinal cord. In this project, the goal is to develop automatic tissue (WM, GM and CSF) segmentation method in brain MRI images. We have proposed a method based on 3D-Unet deep learning model for brain tissue segmentation. Our strategy was to do some pre-processing through registration, histogram matching and histogram equalization.

II. DATASET

In this project, the proposed method has been evaluated on IBSR18 dataset using Dice (DSC). The dataset consists of 18 MRI volumes and was divided into ten volumes for training (1, 3-9, 16, 18), five for validation (11-14, 17) and three for testing (2, 10, 15) which are described in Figure 1. The dataset includes skull-stripped T1-w images, heterogeneity in image intensities and also dataset contains cases with different voxel spacing which complicates the segmentation. Figure 1 shows that voxel size of box under red line is **0.9375 x 1.5 x 0.9375** mm^3, orange is **1 x 1.5 x 1** mm^3, and blue is **0.8371 x 1.5 x 0.8371** mm^3. We have used the training and validation set to train our network with groundtruth. The test set has been used to evaluate the method.

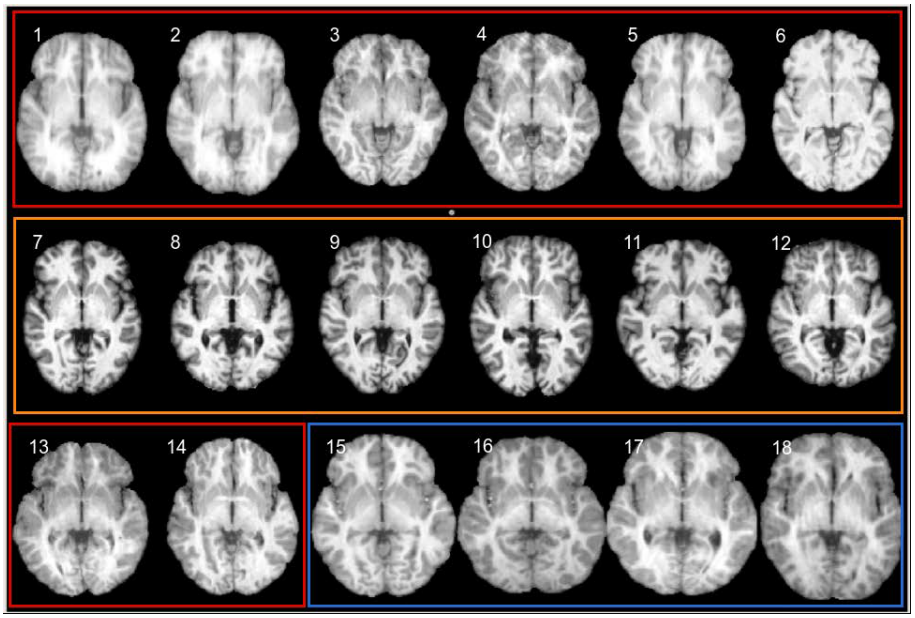


FIG. 1: Axial slice of the 18 cases from the IBSR18 dataset with different voxel spacing

III. METHODS

In order to try to solve this segmentation problem, we chose to use a Deep Learning approach. Given the problem and the type of data, one could use a U-net as Convolutional Neural Network (CNN). U-net is one State-of-the-art network architecture for data segmentation, especially in 2D, with many pretrained versions. This network consists of an encoder (VGG11, ResNet, others...) and of a decoder. In the case of 3D data such as ours, it is possible to use a 3D implementation of the U-net architecture for volumetric segmentation. We chose to base our network on one of two implementation proposed at `https://github.com/ellisdg/3DUnetCNN`. The first is a 3D U-net implementation of Çiçek et al. [2]. The second is an adaptation of the previous one proposed by Isensee et al. [3] for the BRATS Challenge [4, 5]. The topic of this challenge is brain tumor segmentation, which is a part of brain tissue segmentation. We therefore chose to focus our work on this second architecture. This network can be first trained with the dataset as given. However, it is known that in our case, the dataset images are in limited quantity, and very diverse regarding their voxel spacing, intensities and orientations. For that matter, in Section IV propose a Registration process to solve data orientation and voxel spacing problem, while in Section V are suggested

methods for Data intensity normalisation. In both cases, it is important to still be able to obtain a segmentation result on any test image base reference frame.

IV. REGISTRATION

As mentioned in **??** section, IBSR 18 dataset contains cases with different voxel spacing (0.9375 x 1.5 x 0.9375 mm^3, 1 x 1.5 x 1 mm^3, 0.8371 x 1.5 x 0.8371 mm^3) and it arises real problem for our network. It is important to match the voxel spacing and orientation. In this step, we discuss our approach to solving this problem. In order to normalize all the voxels in the same space, we have decided to register the dataset using elastix 4.9.0 toolbox. We have chosen a IBSR07.nii (**1 x 1.5 x 1** mm^3) volume as reference (fixed) due to its quality and intensity. All the train, validation and test volumes have been registered in Python by calling elastix on command line as follows:

```
elastix -f fixedImage.ext -m movingImage.ext -out outputDirectory -p parameterNonRigid.txt
```

The b-spline parameters files have been used from elastix wiki for the brain case. After getting the transformation parameters, we applied it respectively to moving and label (segmentation) volumes to gain resolution as follows:

```
transformix -in inputImage.ext -out outputDirectory -tp TransformParameters.txt
```

The transformation parameters have been modified to get .nii format volume and FinalB-SplineInterpolationOrder 3 to 0.

Moreover, after getting predicted validation volumes from net, we have needed to recover the original voxel space of the predicted validation volumes in order to test the accuracy with segmentation. To do this, we have performed the inverse registration. The registered validation volume has been aligned on their original volume to get the inverse transformation parameters. The inverse transformation has been applied to predicted validation volumes to convert the original voxel space.

For the test dataset, we have done same process. After getting our predicted test volumes, the registered test volumes, which were done at initial stage has been registered again on their original space, just to get inverse transformation parameter and to go back the predicted test volumes to original voxel space.

V. DATA INTENSITY NORMALIZATION AND HISTOGRAM MATCHING

In this section, we aim to reduce the disparity between dataset images intensities. We performe first an intensity normalisation to rescale all the dataset intensities between values 0 and 255. Having a uniform intensity scale is preferable for training the network. In addition to this step, a histogram matching process is performed. Since the images have been acquired in different conditions, the intensities distributions are different across images. This induces a noticeable difference in contrast between images. Histogram matching allows us to fit all images Cumulative Distribution Function (CDF) to a common CDF, matching the data according the the intensity distributions. However, the choice of the reference image CDF is important, as poor contrast dataset might lead to poor results in training the network. We chose one image in the dataset (in our case *IBSR_07* to be consistent witht he registration reference) as a reference and ensure acceptable contrast by using a Adaptive Histogram Equalization (AHE). The rest of the images are then modified by the of usage histogram matching, with the AHE enhanced reference image. These steps are performed using MATLAB R2018b and its toolboxes(*Pre_Processing.m*).

VI. NETWORK

The network is implemented in Python 3.6 with the Keras framework and TensorFlow as back-end. The experiments where performed under Windows 10 Pro, with NVidia 1080 8Go GPU.

The full structure of the Isensee network can be seen in the file *isensee2017.png*. This network size is then modified according to our data and needs.

First, our training data is a T1 MRI image of size (256,128,256) and its corresponding grountruth segmentation image. As it might be difficult to train a network on an image this large, a patch-based approach is prefered. Generators are used to extract this patches on the run. We chose to not perform data augmentation such as permutations and distortion. This choice is motivated by the fact that our dataset will be registered, inducing a high coherence regarding class-wise tissue spatial localisation.

In order to use the generators, the dataset is converted to a HDF5 file. The overall data shape is (15,2,256,128,256) under the form (s,m,,x,y,z) where s is the number of training images, validation included (15 in our case), m the number of channels(intensities and grountruth), and

x,y,z the dimensions of the 3D images. It is to be noticed that one could train this network in a multimodal manner by changing the size of the second dimension. The size of one patch is chosen to be (64,64,64). This value is convenient as it is big enough to keep spatial features as much as possible, but is is also small enough to fit in a deep network with limited hardware capabilities. All patches are selected to containing at least non-background pixel. Given our data, and the fact that patches data is not overlapping within a given batch, we have 320 patches for the training set and 80 patches for validation.

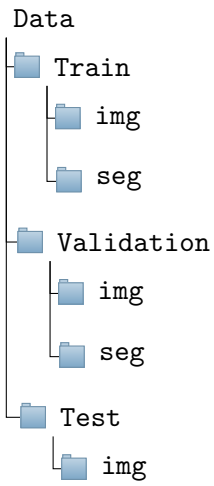
The model encoder part on the model is of depth 5, each depth being 2 instance normalized convolution follow by a LeakyRelu for activation and finally a dropout of 0.3. As the input size of the first layer is 8 (empirically determined) for our network, the output of the final layer of the decoder is of size (b,128,2,2,2). This last layer is connected to a decoder of same depth 5, based on deconvolutions and learning (concatenation) of the data held in the corresponding depth in the encoder part.

The output of the network is a patchwise segmentation of size (b,1,64,64,64) with b the batch_size, i.e to each voxel is assigned a class chosen from: 0. Backgroung (BG: label_0); 1. CerebroSpinal Fluid (CSF: label_1); 2. Grey Matter (GM: label_2); 3. White Matter(WM: label_3). One could use overlapping patches for prediction followed by a voting scheme. We took a non-overlaping approach, leading to an simpler 3D image segmentation reconstruction from the patches.

This network has been trained using the adam optimizer and the inverse Dice Similarity Coefficient (DSC) loss.

One could perform the training of this network by using the *Project_Isensee2017.ipynb* file, given that the dataset stay the same. Few modification in the code are needed in the case of dataset change. For testing on new data, images should be processed as described in Sections IV and **??**. It might also be prefered to project the prediction back to the original image space after training.

The dataset organisation is assumed to be as follow:



VII. RESULTS

In this part we present the results obtain for the different steps of the project.

A. Registration

Figure 3 represents the registration of axial slice 126 of train volume.

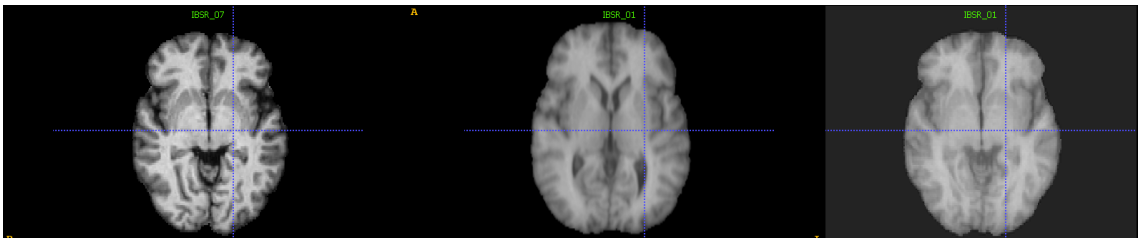


FIG. 2: Axial slice 126 of the train volumes (from left fixed, moving and registered) in the same voxel space

Figure 3 shows the axial slice 126 of the registered train (red), validation (orange) and test (blue) volumes in the same voxel space. In the last stage, validation and test volumes have been converted to original voxel space to get inverse transformation parameter and apply for the predicted validation and test volumes.

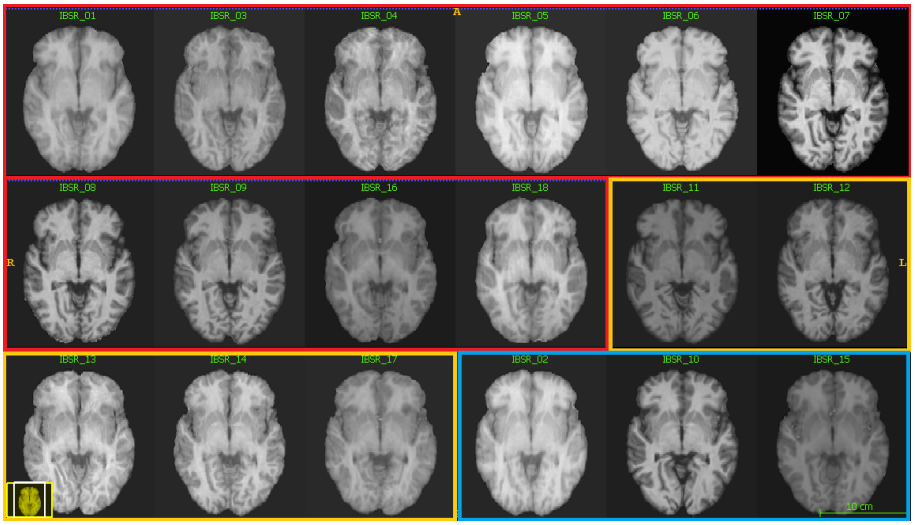


FIG. 3: Axial slice 126 of the registered train (red), validation (orange) and test (blue) volumes in the same voxel space

B. Histogram matching

The first thing to see is the improvement of the contrast of the reference image (*IBSR_07*) after the use of AHE with for parameters alpha=0.7 and beta = 0.7. This result can be seen in Figure 4.

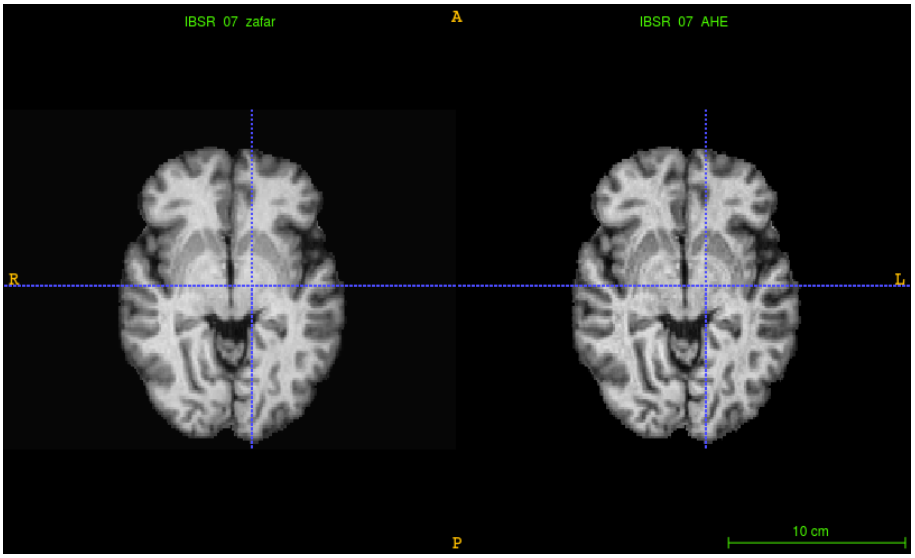


FIG. 4: Axial slice 126 of registered image (left) and registered image with AHE (right)

The first point here is the general improment of contrast and instensity values range over all the other images of the dataset when their histogram is matched onto the AHE image. In Figure 5 is show a very noticable improvement for image *IBSR_06*.

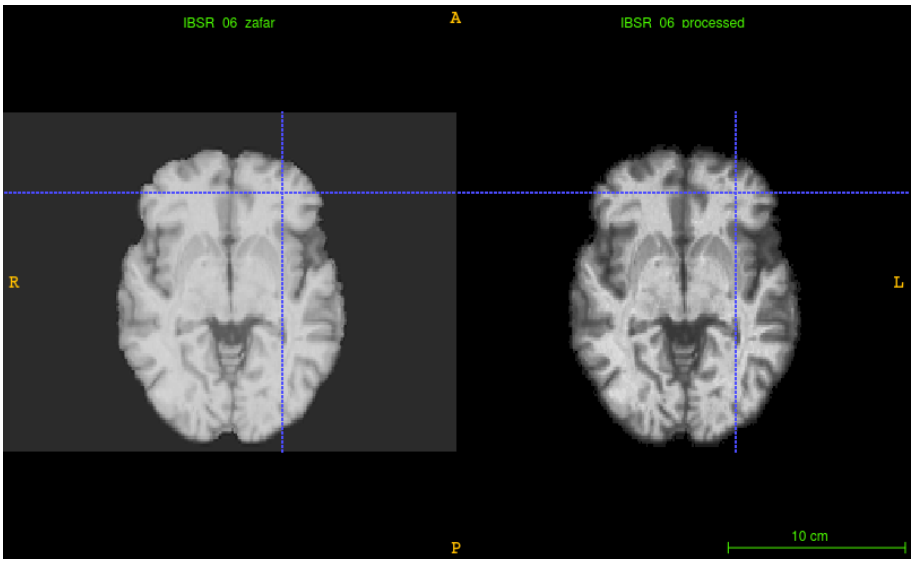


FIG. 5: Axial slice 126 of registered image (left) and registered image with histogram matching (right)

From this visual assessement, we can deduce that the histogram mlatching step design and implementation are both sucessful. Is to be determined its influence on the training phase. This is commented in the two following sections.

C. First model

This model has been trained with images which are only registered as described in Section IV. No other preprocessing (such as histogram matching) was performed. The network used is exactly the same as presented in Section VI. The results obtained during the training are presented in Figure 9

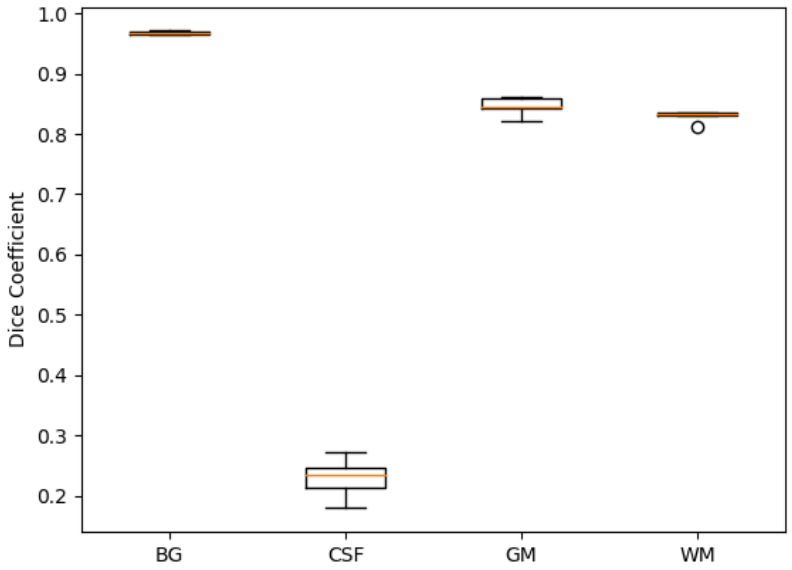


FIG. 6: Dice indexs of validation-set segmentation results

It is clear that the learning process on the CSF is not performing as expected. The DSC as a loss metric does not take into account data unbalance. In fact, the CSF class is underrepresented in all brain MRI images. The weigthed version of the DSC allows to give important weight to CSF although it is only a small part of the volume. We empirically verified that using simple DSC loss lead to poor segmentation of the CSF class. The weighted DSC loss has been used in all models tried after this one. To recall, the data used to train this network was not preprocessed.

D. Final model

In this part, we present the model that gave the best results for the Validation set predictions. Between the first model and this final model, many structure were tried, and most of them unsuccessful. Some aspects learned during these tries will be discussed in the next section.

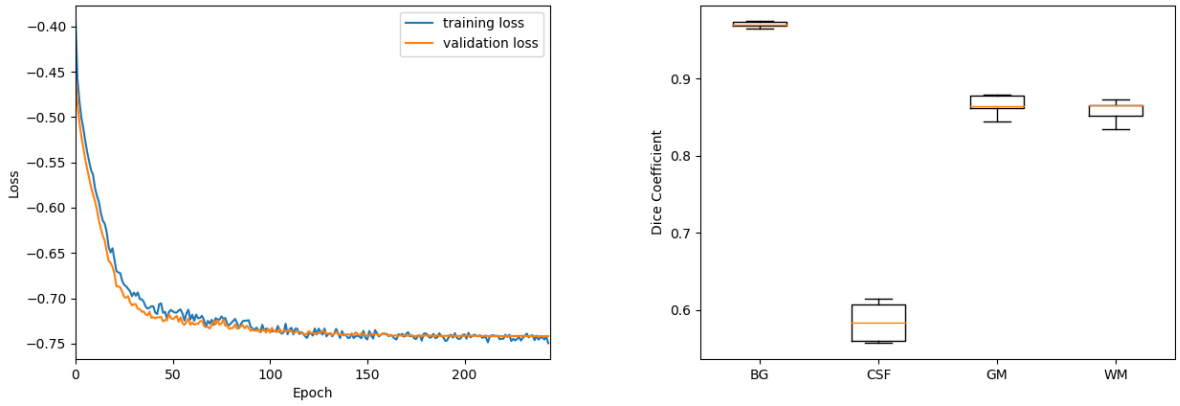


FIG. 7: Output metrics of the training phase

We can note the value of 0.58 DSC for the CSF class. This is low compared to standards results. However we can deduce that the methods used to preprocess the data, and specifically focused on the CSF class segmentation, has been an enhancement. Regarding the White Matter and Grey Matter classes, the median DSC is of 0.86 for both classes. These values are in accordance with the ones observed in the training phase.

Below are DSC valueson segmentations projected back to the original image space. It is expected for this results to be worse than in the training space as we have to the inverse registration process described in IV (on the segmentation image only). Therefore, the segmentation result, by the use of interpolation, is modified.

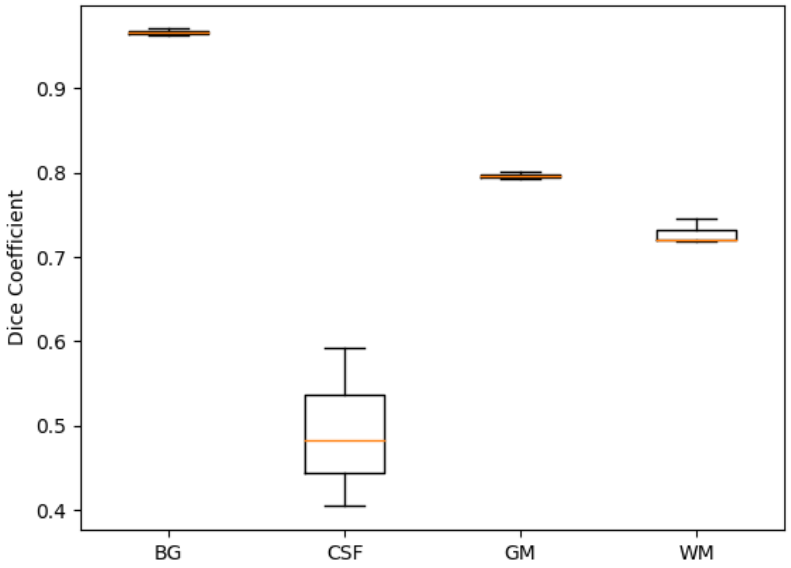


FIG. 8: Final DSC values for training set in its original space

One could note that all class-wise DSC have, as espected, dropped. For the CSF, the relative decrease of the DSC is 17%(10 points, from 0.8 to 0.48). There has also ben a noticeable decrease in the White Matter class. Below are shown examples of segmentation, compared to their groundtruth.

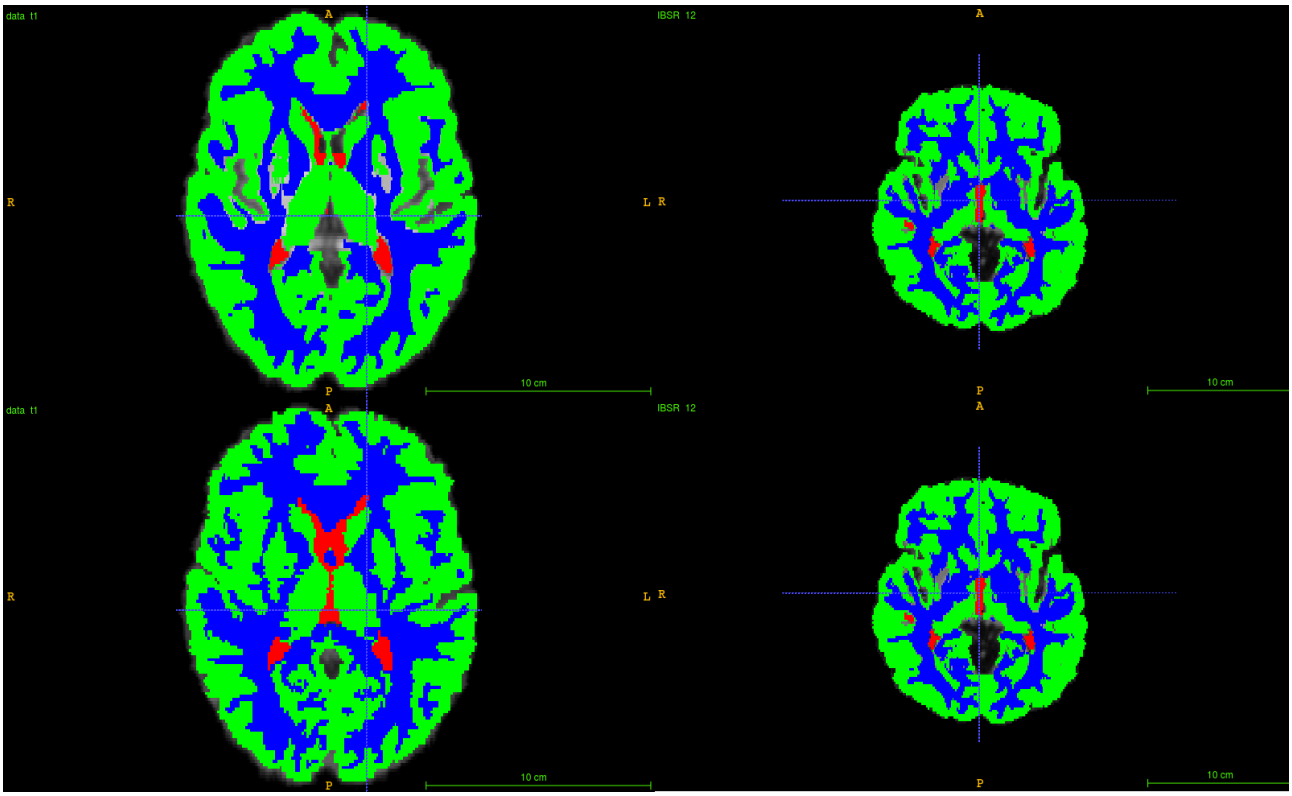


FIG. 9: Segmentation results of IBSR_12 slice 129 (Axial). Top:Segmentation prediction; Bottom: Groundtruth segmentation; Left: Registered space; Right: Original image space.

VIII. DISCUSSION

Overall, we expected our best network to give a much better CSF segmentation. In the aspects that we explored to try to improve this segmentation, some are to be mentionned:

- The use of weighted DSC loss. With this particular network structure, we had to hange the loss metrics. By using a simple DSC, the CSF class was not learned at all during the training phase.

- This bring the issue of data augmentation and data balancing. This project has been our first use of generators, and it was a challenging task. Although we are able to select only non-empty patches, we were unabe to modify the data extraction in order to balance the data (ina class-wise manner) through means of data augmentation.

- We tried data augmentaton on the all data-set (permutations, flip, axis swaps) and gave only worse resuts. In fact, they are not used to try to solve the balance poblem mentionned above, simply to artificially increase dataset size. Since our data is registered into a common space (voxel spacing size and data location/orientation), using permutations or distortions decreqses the spatial coherence of the dataset.

- As always, the more data the better. This case, if the dataset was way larger, it might be better to skip the registartion process totally, as many spaces could ealisy represented by many instances in a big training set. Also, we realised in the final parts of the project that this network structure might have been too ambitious compared to our amount of data. The Isensee implementation of the 3D u-net includes droupout and instance normalisation (improved batch normalisation for small batches) with 32 base filters. This concepts are generally having a positive inpact on segmentation results. However, we empiricaly saw that reducing the number of input filters in the first layers inproved our segmentation. We went from 33 millions parameters to 2 millions, which were easier to train with our data.

- The final thing we tried is to modify more (than the input layer) the network. This task was challenging to us, as we only previously used predefined networks. We first unlooped it, and removed dropout and instance normalisation. We also tried to look for faulty box linking, as the predicted results in the first network were containing unaceptable artifacts (see Figure 10. Due to either inexperience or missunderstanding of the problems' source, we were not able to correctly and meaningfuly modify the network architecture.

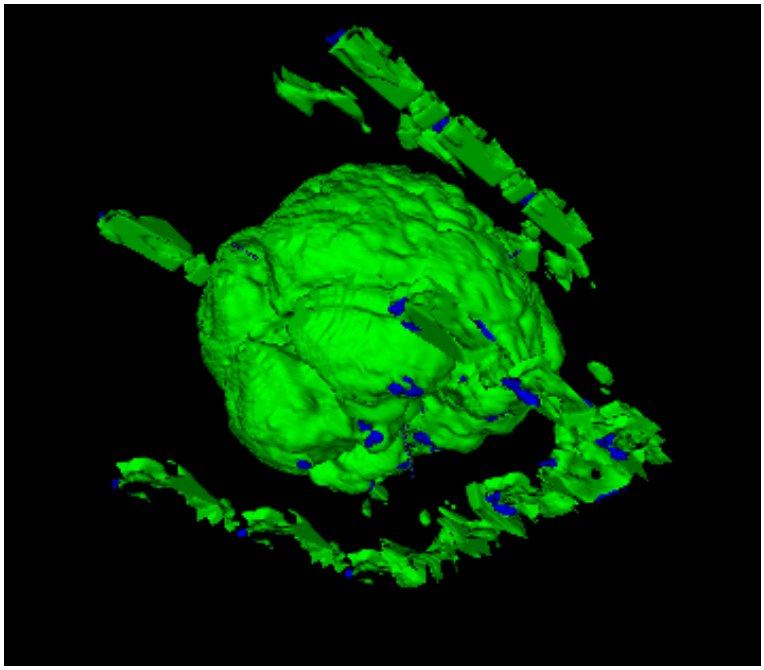


FIG. 10: Presence of artifact in network segmentation predictions. Most of them are on the edges of the images.

IX. CONCLUSION

In this report we presented our work on brain tissue segmentation for T1 MRI images using a 3D U-net CNN, adapted from previous work on brain tissue lesion segmentation. Using the Isensee 3D U-net architecture we managed to reach a DSC on validation set for CSF of 0.48. This value has been increased when using the weigthed inverse DSC as loss function for the training. We also pointed the important role of intensity normalisation and contrast enhancement methods in the improvement of the training phase results. The Overall Dice Similarity Coefficient over all classes is 0.94. However, this metric does contain the background information, and doesnot show the low prediction acuracy on CSF class. Finally, voxel spacing and data registration might be of lesser importance for way bigger datasets, but in our case they have given good improvement over the segmentation accuracy. As a final note, we belive that the network choice was mitigated. While its presentation and implementation were showing good results for brain tissue lesion segmentation, we never managed to obtain the results we would have liked to on the CSF class. Even if our placement

in the MISA challenge is to be expected, we feel we improved a lot or knowledge on the deep learning concepts and implementation. A pretrained approach with fine tuning or transfer learning might have given better results, but we chose at the beginning of this project to work first in a 2D patches approach (which we abandonned rapidly), to then move to 3D U-net implementations, which were new approaches for all of us. Finally, withthe chosen network base, we did our best to search and implement new idea to improve the segmentation results

X. REFERENCES

[1] *State-of-the-Art Methods for Brain Tissue Segmentation: A Review. Dora L, Agrawal S, Panda R, Abraham A. Send to IEEE Rev Biomed Eng. 2017;10:235-249. doi: 10.1109/RBME.2017.2715350. Epub 2017 Jun 14.*

[2] *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation.Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, Olaf Ronneberger. arXiv:1606.06650v1 [cs.CV] 21 Jun 2016*

[3] *Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the BRATS 2017 Challenge. Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, Klaus H. Maier-Hein. arXiv:1802.10508v1 [cs.CV] 28 Feb 2018*

[4] *Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber MA, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp R, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SM, Ryan M, Sarikaya D, Schwartz L, Shin HC, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)", IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015) DOI: 10.1109/TMI.2014.2377694*

[5] *Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features", Nature Scientific Data, 4:170117 (2017) DOI: 10.1038/sdata.2017.117*