Medical Image Registration and Applications

Lab 3: Atlas Based Segmentation
(Integration to EM Algorithm)

28 November, 2019

Abstract:

This report summarises the results of our work in the construction of the **Atlas and Tissue Model** which were combined with **Expectation Maximization** and utilized for Segmentation.

We constructed three different types of Atlas techniques (**Individual, Most Similar and Average Shape**). Peak and Histogram Matching Techniques were applied in an effort to improve the constructed Tissue Model.

Authors:

Zohaib Salahuddin

Isaac Llorente Saguer

MSc. Medical Imaging and Applications (MaIA)
Universitat de Girona

Contents

1 Introduction

The segmentation of anatomical structures in brain is a challenging task due to the presence of intensity variations as well as artifacts. It can also be attributed to the fact that these structures may share closely associated intensity distributions. In such circumstances, the use of *A Priori Information* is of paramount importance for segmentation purposes. Two types of such information are utilized commonly; **tissue properties** and **anatomical properties**. In this lab, we explored the use of such anatomical priors in terms of an atlas. Tissue Intensity Properties were also utilized by the construction of intensity based histogram tissue models.

1.1 Objectives

- Familiarization with the steps involved in the construction of Atlas.
- Choosing the best method of registration for the construction of Atlas.
- Experimenting with different methods of Atlas construction.
- Exploring the different ways of constructing the tissue model highlighting the intensity profile of each tissue.

2 General Steps in Atlas Construction

2.1 Pre-Processing

Aniostropic Diffusion was applied to all the brain volumes in the training set before the construction of the atlas and registration. All the testing images were also exposed to similar preprocessing techniques.

We tested different parameters for the anisotropic diffusion, following the work that we did for a previous lab, and although about half of the resulting dice scores saw a slight improvement, the overall average dice score was slightly worse, hence we stopped this line of work.

Bias removal was also considered to help with registration; it is already applied with the elastix parameters that we use.

2.2 Discussion of Parameters for Registration

The training data consisted of 3D MRI T1 brain volumes. For this reason, we used the parameter files already tuned for generating atlases from Data originated from the Rotterdam Scan Study. [1] These parameters files were used as a starting points, however, further tuning of parameters didn't give better registration performance. Table 1 below discusses the choice of the parameters for registration purposes.

Registration Parameters	Choice and Discussion
Multi Resolution Registration	As seen in the first registration lab, multi-resolution frameworks help in faster convergence. For this reason, we employ this along with gaussian smoothing and downsampling. The number of resolutions were set as 5. (Registration "MultiResolutionRegistration"), (FixedImagePyramid "FixedRecursiveImagePyramid") and (MovingImagePyramid "MovingSmoothingImagePyramid") are the set parameters.
Optimizer	The optimizer used for our registration was a variation of Stochastic Gradient Descent known as Adaptive Stochastic Gradient Descent. It is same as SDG but it has a better initialization approach.
Metric	There are various available choices for metrics such Normalized Cross Correlation, Square differences etc but the preferred metric employed for our registration was a variation of mutual information i.e. Advanced Mattes Mutual Information.
Transform	The transformation type was set as Affine and B-spline according to the transformation desired.
Interpolator	B-Spline Interpolator was used while transforming images with intensity values. However, for transformation of labels Nearest Neighbor interpolator was used.
Result Image Pixel Type	The result image pixel type while performing transformix on the atlas probabilities using the parameter files generated from the registration of mean image to the testing image, we set the result image pixel type as float instead of short.
Image Sampler	This is the parameter of the optimizer. Instead of using all the values for optimization, we can choose subset of the values at each iteration which ensures faster convergence. Hence, in our case we used Random Sampling for the optimizer.

Table 1: The table discusses different parameters used for Affine and B-spline registration as well as the parameters of the transformation used for label transformation.

2.3 Registration of the Training Images

The first step involved in atlas construction is registration of all the training images in the common space. The registration is done using **elastix**. The corresponding transformation parameters obtained as a result of registration are then utilized for label transformation using **transformix**.

2.3.1 Choice of Registration Type

Rigid (Affine) and Non - Rigid (BSpline) Registrations as well as their combinations were used for performing the step of transforming all the images to the common space.

Registration Type	NCC between Fixed and Registered Images
Affine	0.8217
BSpline	0.9461
Affine + BSpline	0.9452

Table 2: The table shows the Mean Normalized Cross Corelation Values between the fixed image i.e 1000.ni and the registered images.

Table 2 shows that Affine Registered is much inferior for registering the images to common space as compared to using B-Spline or Affine followed by B-Spline Registration. This is due to the limitations of affine transformations, since they only influence scaling, translation, rotation and shear, as a general picture, whereas B-Spline can bend the transformation field locally, so it can adapt much better. **We chose to apply Affine registration followed by B-Spline registration.** Even though B-Spline Registration seems to give a very slight superior performance, we chose a combination of Affine followed by B-Spline Registration because it is advisable to perform a rigid registration before applying non-rigid registration [as per the lecture we had on this topic]. This was also primarily influenced by the fact that *elastix* crashes for some of the cases when B-Spline registration is applied alone.

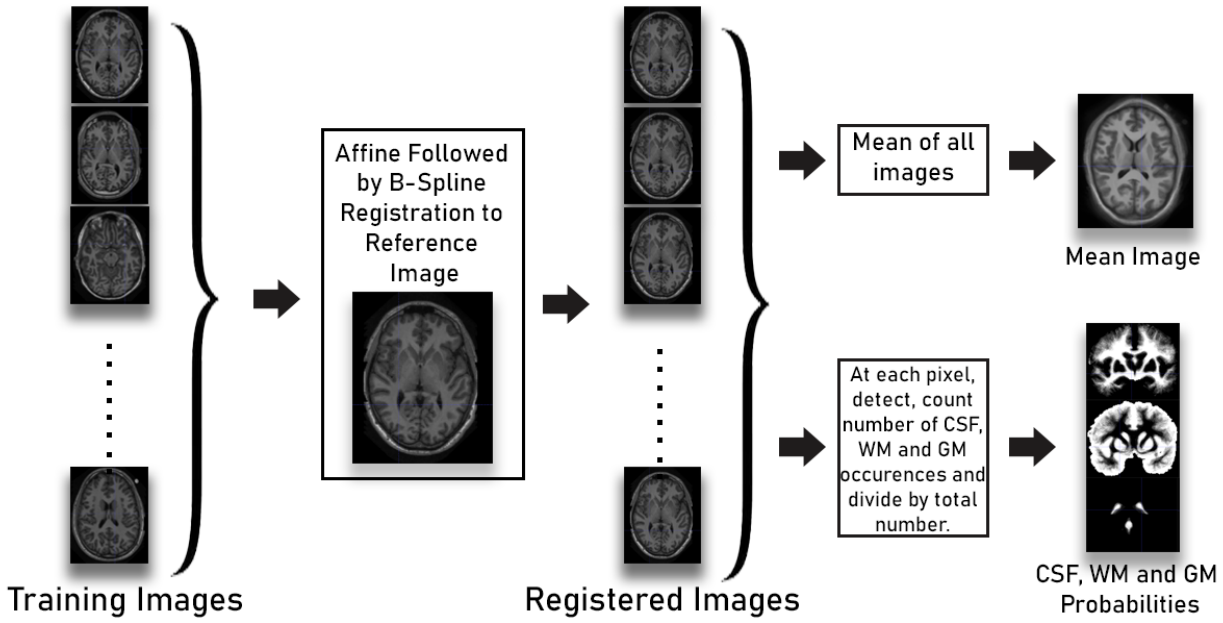

Figure 1: The figure shows the whole pipeline of the Atlas Construction Process.

2.3.2 Registration Result after Masking and Un-Masking

After the choice for registration has been made, we experimented with performing registration using image masks. Since the mask information is available, we perform registration to the common space with using the fixed and moving images masks. After the comparison of segmentation results, it was evident that registration with image masks yields superior performance than without image masks.

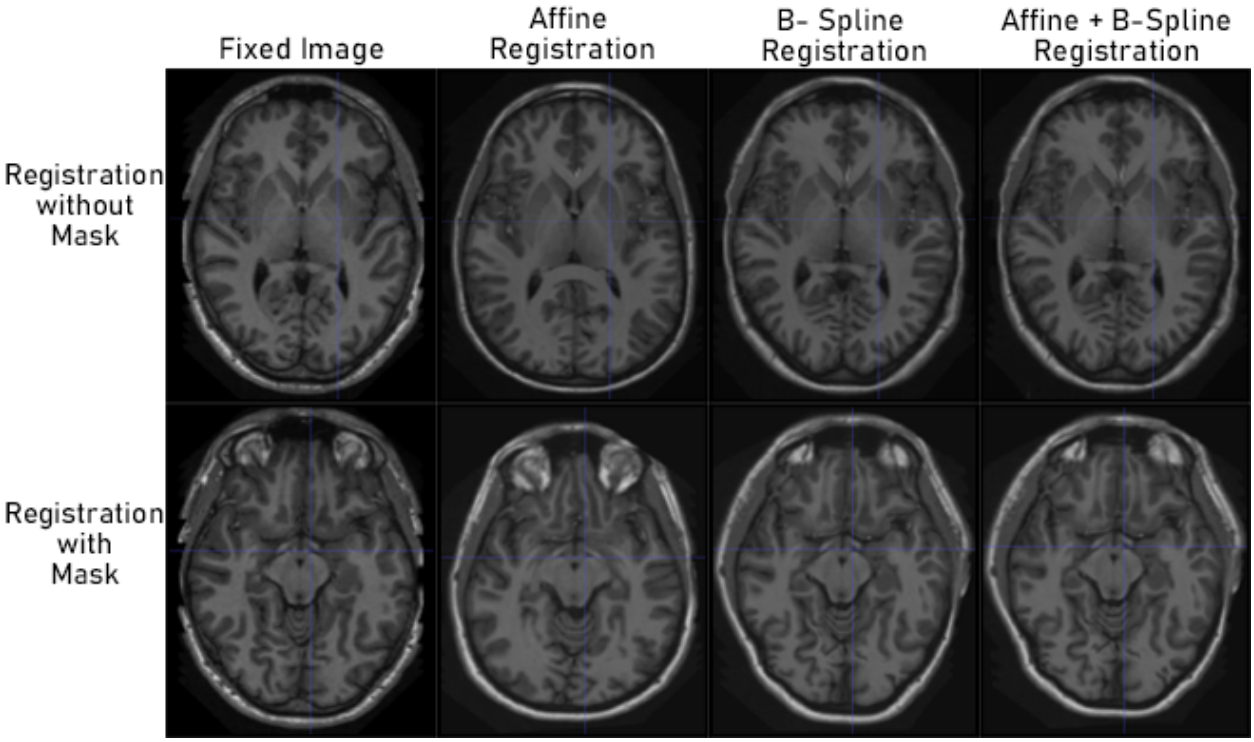

Figure 2: Registration results for different types of registration using mask and unmasked images

Figure 2 clearly emphasizes the use of Affine followed by B-Spline Registration and the usage of mask images.

2.4 Mean Image and Atlas Construction

2.4.1 Construction of Mean Image

The mean image was constructed by normalization of all the registered images. This mean image is useful as it was utilized later for registration of the atlas probabilistic models to the testing images to be used for label propagation.

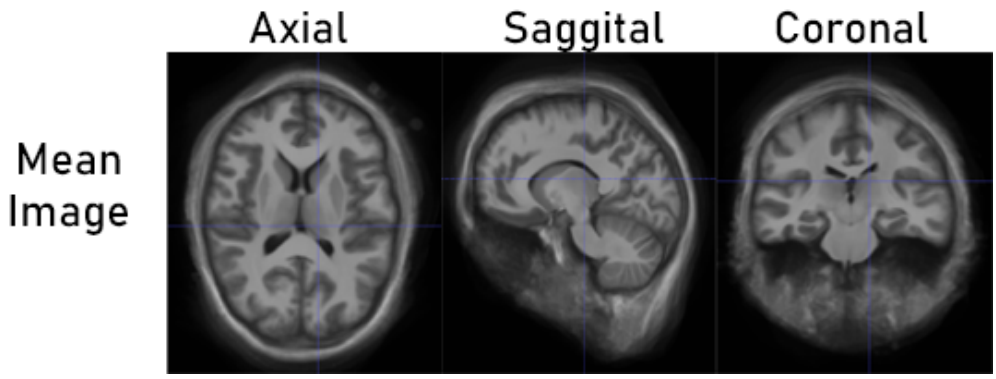

Figure 3: Mean Image Construction after Registration of training images to common space

2.4.2 Construction of Atlas Probabilistic Model

After the registration of the images, we calculate at each pixel, the probability of it belonging to either CSF, White Matter or Grey Matter. We make use of the segmentation of the training images and detect at each pixel position, the number of occurrences of each tissue type. These number of occurrences are divided by the total number of training images. The result of this process are three probabilistic distributions for each of the tissue types.

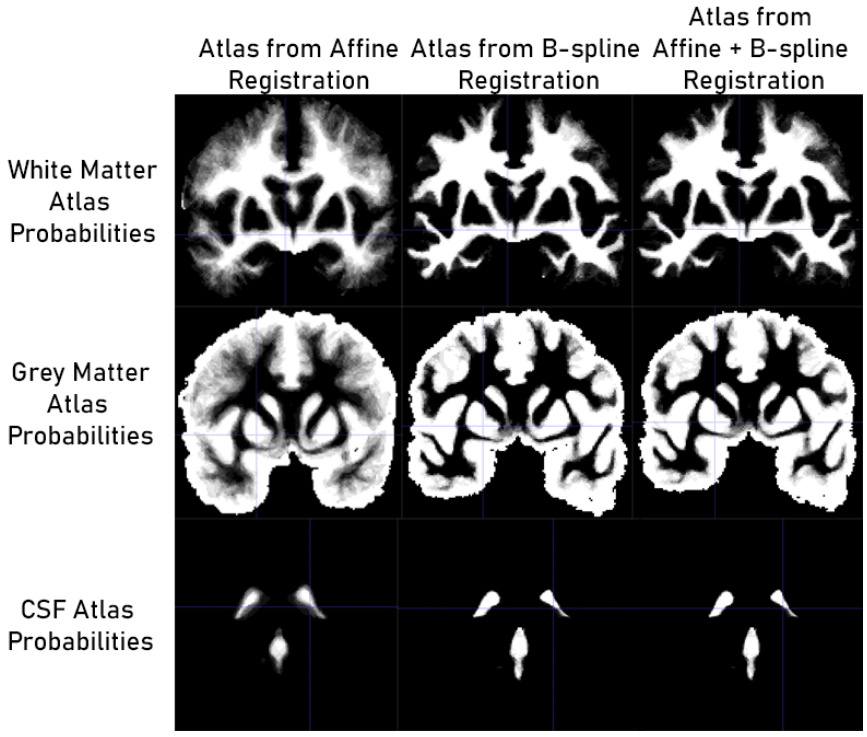

Figure 4: The figure shows the constructed atlas probabilities for 3 registration types. The boundaries of atlas with Affine registration is more blurred than the other two types.

3 Types of Constructed Atlas

3.1 Individual Atlas

In this type of atlas, we construct only one atlas. The choice of reference is usually arbitrary after some qualitative analysis. We chose *1000.ni* after observing that it had no obvious lesions or artifacts. All the other images were registered to it and atlas was constructed. Figure 4 above shows such an individual atlas.

3.2 Most Similar Atlas

We have 15 images in the training set. 15 atlas were constructed. We then calculated the normalized cross correlation of each fixed image with all the other images.

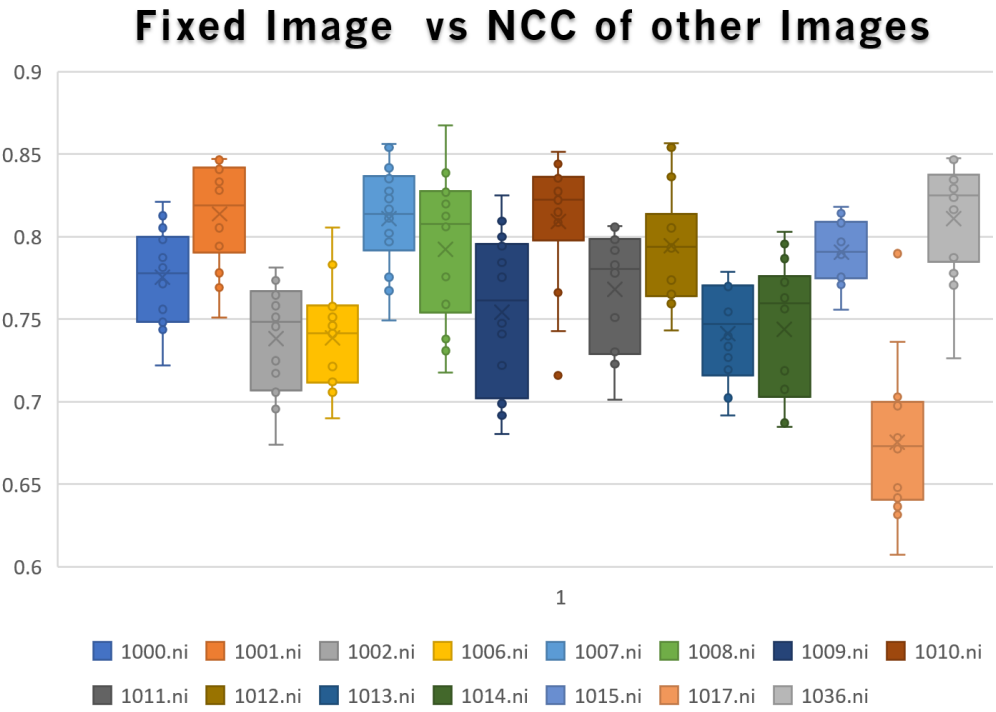

Figure 5: The figure shows the normalized cross co-relation of each fixed image with all the other registered images.

Figure 5 clearly depicts that the atlas with *1001.ni*, *1007.ni* and *10036.ni* could be choices for most similar atlases. We choose 1007.ni as the most similar atlas.

3.3 Average Shape Atlas

The following steps were involved in the construction of average shape atlas:

1. Each training image is used as a reference image in order to construct 15 different atlases.

2. All the labels corresponding to these 15 atlases are transformed to the common space of the most similar atlas.
3. After this transformation to common space, average atlas is constructed by taking the average of all these atlases.

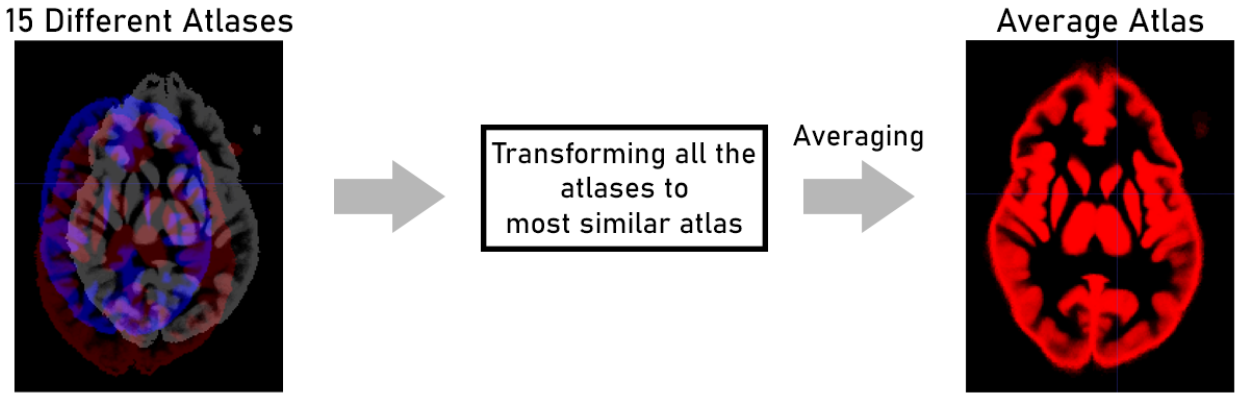

Figure 6: The figure shows the pipeline for average atlas. 15 different atlases are constructed but only 3 of them are shown in the figure. They are transformed to common space and then averaged.

3.4 MNI space

We have explored different MNI spaces. Some of them were discarded because of various reasons (Colin27 has no CSF defined, 6th generation MNI152 has no associated definite labels, etc.). We took the one provided by the course via moodle, which did not have the definite labels or mask, so we built them using the probability maps. For an initial evaluation, we visualised the material with ITK-SNAP. As we can observe in the following Figure 7, it displays a pretty diffused version of a brain. This diffusion could play a big role for determining the actual tissue: there are no folds in the outer region of the labels image.

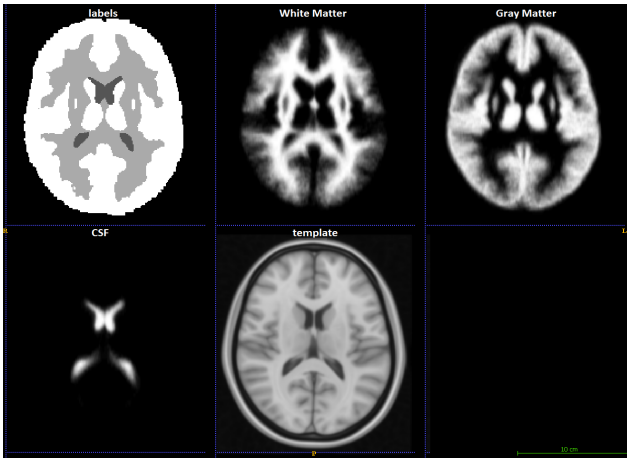

Figure 7: Labels and tissue probabilities for the given MNI atlas

3.5 Single image as Atlas

For the sake of experimentation, and also to follow the intuition behind this strategy, we used the sole information contained on a single brain to perform label propagation. The average dice scores (average of the 3 tissue dice scores for all test brains) varied between 0.85 and 0.87, as we will see in the results section. To further expand this line of research, we could have different types of brain with a good resolution and segmentation, and use the most similar one to the test to register and propagate the labels.

4 Tissue Model

Another type of A Priori Information that can be used for segmentation is the tissue intensity information. We used the training images in order to calculate the probability of each intensity belonging to the one of the three tissue classes. This probability distribution of intensity in each of the tissues can be used for segmentation purposes. Three types of tissue models were explored: Without pre-processing, with peak match and histogram matching.

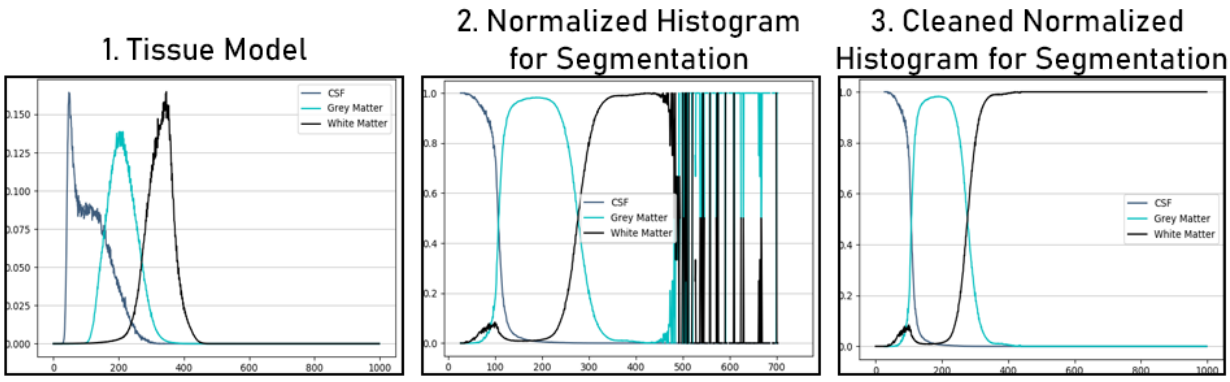

Figure 8: The first image shows the tissue model probability distribution.The second image shows normalized histogram for segmentation. The sum of probabilities at any intensity is equal to one. The third image is obtained after removing noise from the second one.

For the peak match, we realised that the different histograms of the images in the training set could be identified and normalised by the two main peaks. These are automatically detected (as we can see in Figure 9) and used to normalise the intensities.

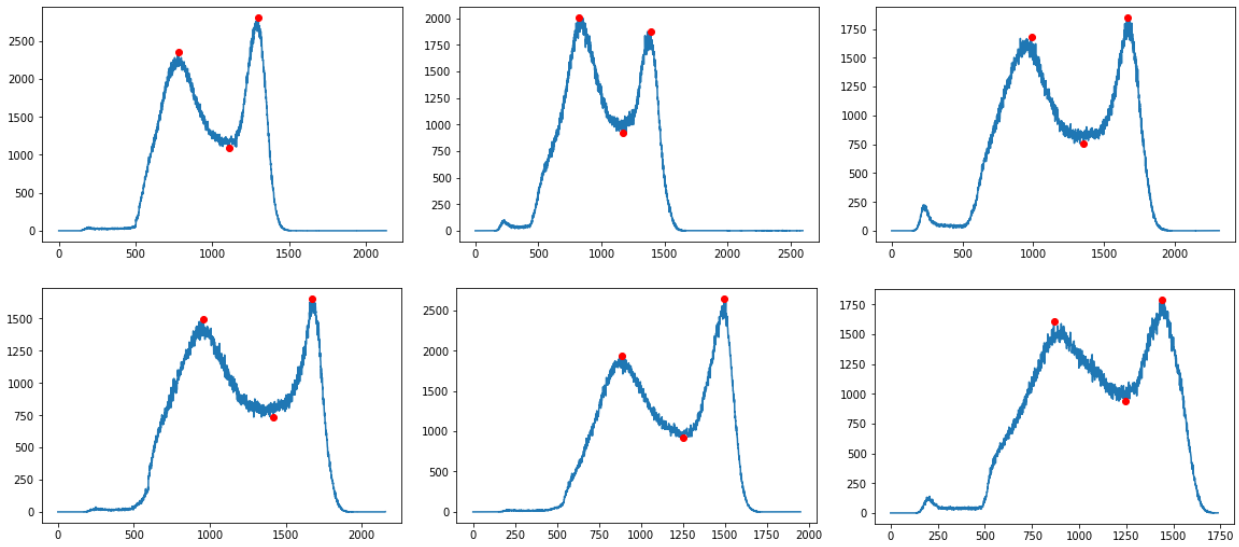

Figure 9: Automatic detection of the characteristic peaks of the histograms of T1 images

In the following Figure 11 we can see how we matched all the peaks of all the images in the training set, by just analysing their histograms. We can clearly see the advantage of peaks versus using the min-max histogram stretching.

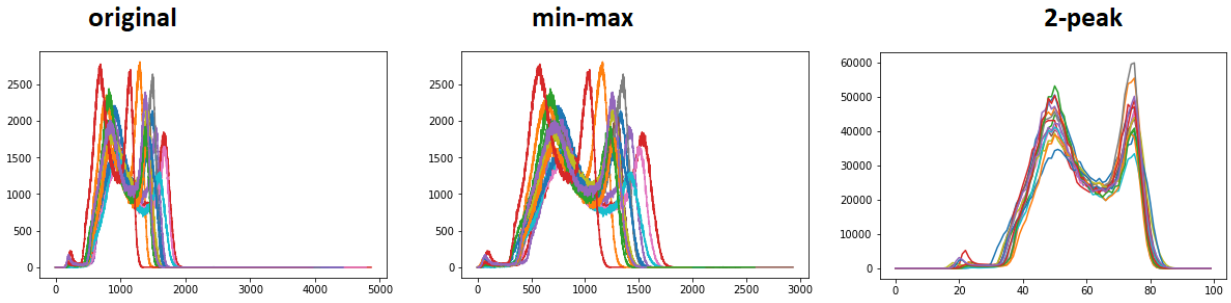

Figure 10: Comparison between raw histograms, min-max stretched and 2peak-adapted of all the images in the training set

5 Registration Results and Discussion

In this project we had the opportunity to further explore the field of registration using the powerful tool that is elastix.

As seen in the results of Section 2.3.2, We obtained the best results for registration using Affine and then B-Spline transformation. We favor using the masked brain and not the whole head because of various reasons:

- Different MRI acquisitions might cut the head in different planes
- Our final objective is the segmentation of the brain, so we don't need to adapt any differences that occur outside of it

We explored multiple types of tissue models:

- Raw (no pre-processing of the intensities)
- Max intensity stretching
- min-Max intensity stretching
- Peak stretching

The aim of these models was to help normalise the intensities of different brain acquisitions, as well as having a probability baseline.

Both the visual evaluation of the registration and the more quantitative metrics such as the normalised cross correlation or in an indirect way, the dice scores, show that we achieved a 'good' registration. Otherwise, label propagation could not achieve such dice scores (close to 0.9).

Further work could be focused on comparing other registration frameworks, pre-processing the brains to help the registration, or building more atlases to help with generalisation (different age ranges, brains with and without atrophy...)

6 Annex

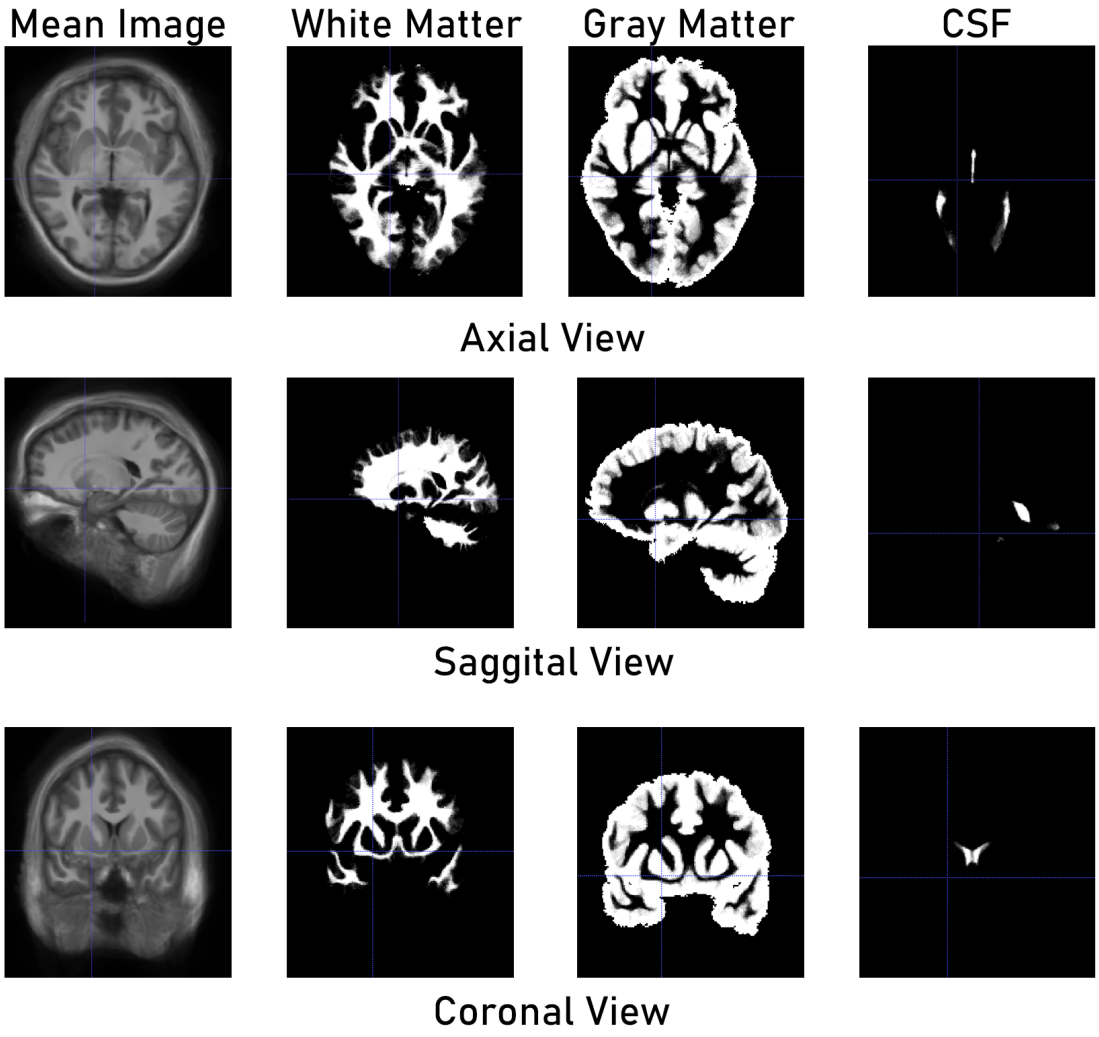

Figure 11: Atlas Probability Maps being displayed in Axial, Coronal and Saggital Views

7 Segmentation Results

Having worked with different kinds of atlases and tissue models, we then went on to compute the segmentation and evaluation of the brains in the test set.

These are the type of strategies we tried:

- Atlas label propagation
- Most Similar atlas label propagation
- Multi-atlas average label propagation
- Single reference label propagation
- Tissue model prediction
  - Raw (no pre-processing of the intensities)
  - Max intensity stretching
  - min-Max intensity stretching
  - Peak stretching
- EM algorithm (developed at the previous lab)
  - K-Means initialisation
  - Tissue Model initialisation
  - Atlas initialisation
  - Atlas weighed integration
- Combination of different of the mentioned methods

7.1 Individual Atlas, Most Similar Atlas and Multi Atlas Average Shape Segmentation

The segmentation results for Individual Atlas *1000.ni* and Most Similar Atlas *1007.ni* were very similar. 15 atlas were generated previously, registered to a common space of Most Similar Atlas. Finally it was averaged to give the multi-atlas average shape segmentation. This average shape atlas had similar performance for CSF and gray matter. However, in case of white matter it showed better performance. Average Shape atlas would perform better generalisation.

7.2 Late Fusion Based on Bayes Rule

The probability maps generated from Label Propagation, Expectation Maximisation and Tissue Model were combined together using Bayes Rule. We can combine spatial information together with intensity information in the following two ways using Bayes Rule:

- Expectation Maximization (Intensity) and Atlas (Spatial)
- Tissue Model (Intensity) and Atlas (Spatial)

After the probability multiplication, we perform label assignment taking the Argmax. These frameworks gave the best result as they incorporated spatial as well as intensity based information. The spatial discrepancies in EM and TM based models were overcome with atlas based models. We observe that the errors in EM is especially on the borders. This is because these intensity values are on a borderline of the intensity distribution of Gray matter and CSF. The confidence level of EM is not high at this region. Atlas on the other hand has high confidence for Gray matter at this region. In this way, this error of CSF is averted from EM because of incorporation of high confidence Spatial information.

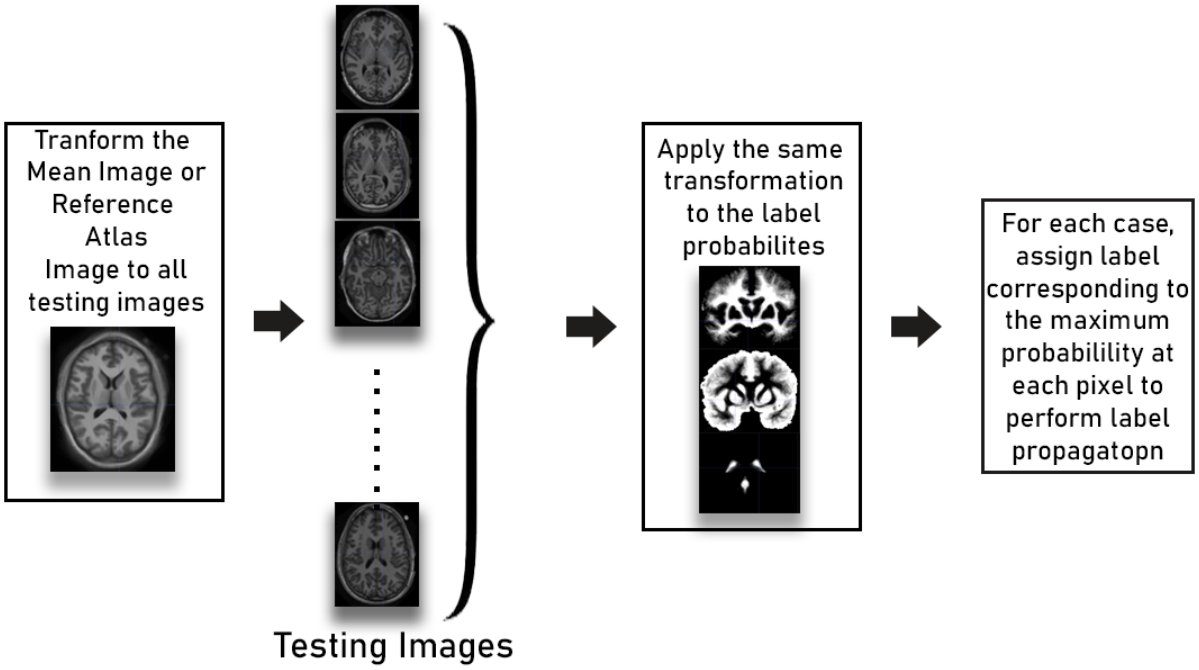

Figure 12: Label Propagation Pipeline

7.3 Expectation Maximization

7.3.1 Initialization

As we observed in the previous labs, the Expectation Maximization Algorithm converges to different solutions depending on initialization. For this reason, we tried three different kinds of initializations:

- K-means Initialization
- Atlas Initialization
- Tissue Model Initialization

K-means wrongly identifies CSF at the boundaries which is labelled as Gray Matter in the ground truth. EM builds on this and the dice plunges to low value. Atlas and TM initializations give a better starting point to EM close to the ground truth. Even still, EM drifts away but the error is less as compared to K-means.

7.3.2 Atlas incorporation inside EM

The incorporation of spatial information at each step of EM would help to keep the performance of EM in check. After the update of mean and covariance parameters of the Gaussian mixture model, we scale the output of the mixture model with the atlas probabilities with the corresponding class. For this reason, the initialization of EM has to conform with atlas labels for proper scaling at each iteration.

In contrast to the previous EM algorithm approach, we scale the Gaussian mixture model output with atlas probabilities instead of the class distribution. It is achieved in the following way as demonstrated by the code.

```
prob_dist[0] = prob_dist[0] * CSF[t1_idx_i,t1_idx_j,t1_idx_k]
prob_dist[1] = prob_dist[1] * GM[t1_idx_i,t1_idx_j,t1_idx_k]
prob_dist[2] = prob_dist[2] * WM[t1_idx_i,t1_idx_j,t1_idx_k]

total = weight_dist[0] + weight_dist[1] + weight_dist[2]

weight_dist[0] = np.divide(prob_dist[0],total)
weight_dist[1] = np.divide(prob_dist[1],total)
weight_dist[2] = np.divide(prob_dist[2],total)
```

Listing 1: Incorporating Spatial Information Inside EM

7.4 Single reference label propagation

The following boxplot shows the DICE scores obtained by simply propagating the labels of one of the training images to each of the test set images. TO achieve these numbers, we register the reference training image to the test image, and apply the transformation to the reference labels. These results can be obtained by executing the LabelProp.py file. We can see that with this simple method alone we are able to reach scores close to 0.9.

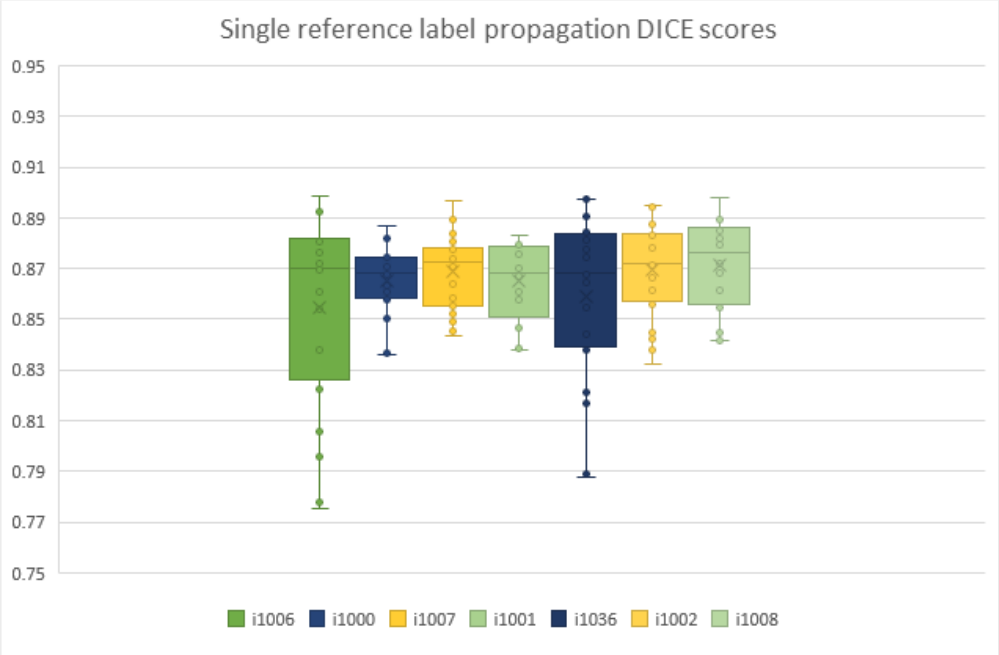

Figure 13: Dice scores of different training images as reference, applying label propagation

7.5 Tissue Model prediction

Here we can see different results for different tissue model strategies applied directly to segment the brain. We can observe how the peak method looks more robust that a simple histogram stretching. We can also see that applying anisotropic diffusion didn't turn up to be a significant boost for this model. The results can be obtained by executing the file getTM.py.

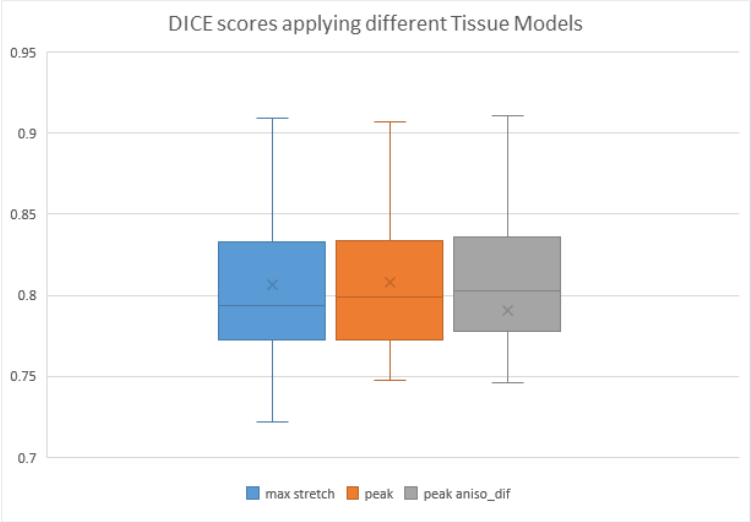

Figure 14: Dice scores of different training images as reference, applying label propagation

7.6 Method comparison

In this subsection we compare the best version of each method against each other. All sub-figures in the following Figure 15 have the same method order.

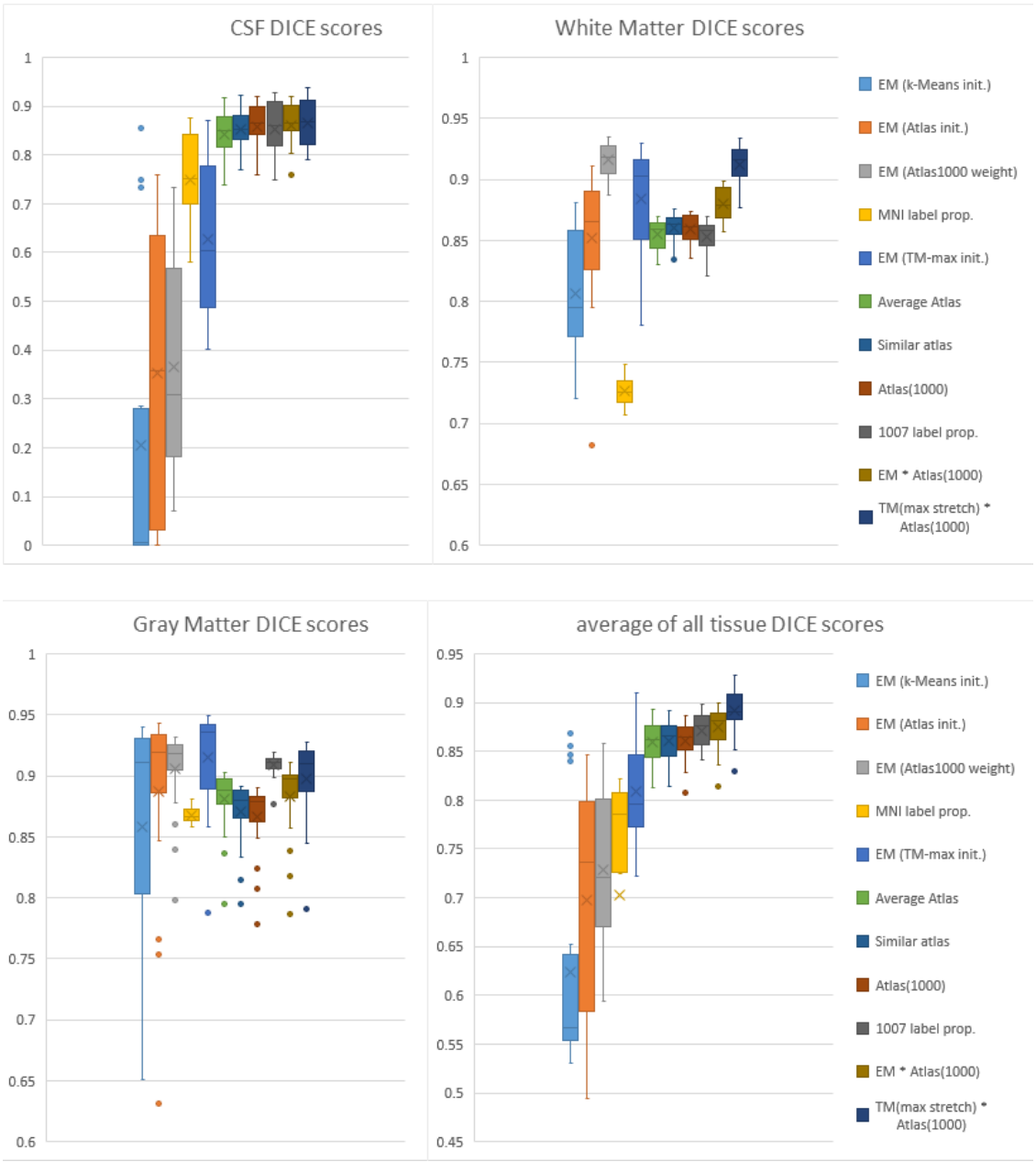

Figure 15: Dice scores for the three tissues separately, and together. Different strategies evaluated.

We can see in the previous figure how powerful the combination of different methods can be. This could be a future field of research, if Deep Learning was not present in the picture.

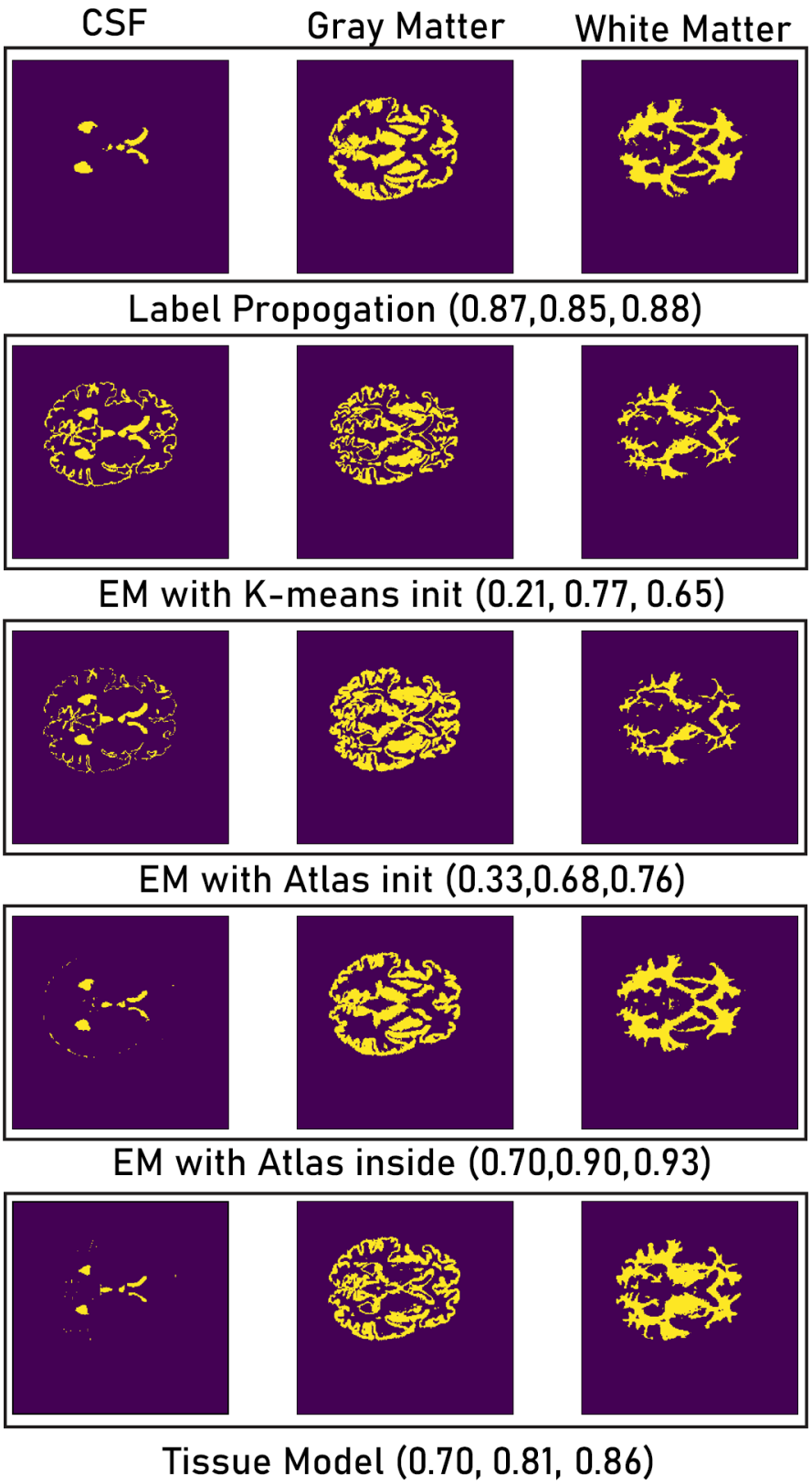

Figure 16: Comparison of Segmentation Techniques without Late Fusion for testing image 1128.ni

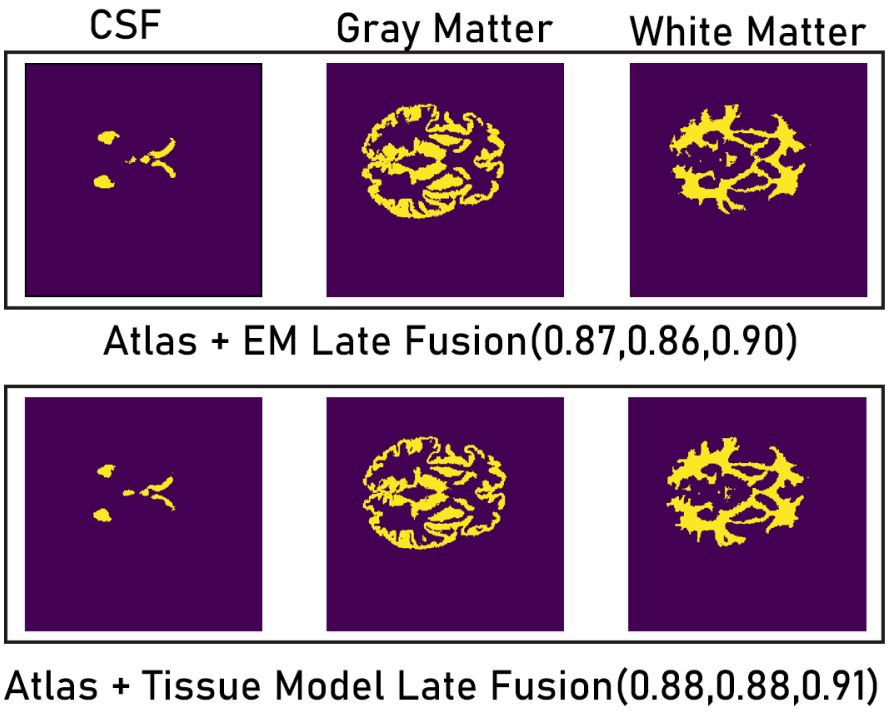

Figure 17: Comparison of Segmentation Techniques with Late Fusion for testing image 1128.ni

A representative testing case of 1128.ni was chosen for qualitative analysis Figure 16 shows comparison of methods without bayesian late Fusion. The first row shows label propagation which does not take into account any intensity information. The result is reasonable. The second row shows result of EM with K means init which shows visible errors in CSF at the boundaries. The third row shows EM with Atlas init which shows that this spatial initialization somewhat reduces the boundary errors. The fourth row shows that incorporating Spatial information inside EM further reduces the error. The last row shows that tissue Model provides acceptable segmentation but it is noisy.

Figure 17 shows the result of bayesian late fusion which can improves the performance of segmentation drastically because of high confidence values from the involved probabilities. Atlas corrects the CSF misrepresentation at the boundaries in case of EM fusion and corrects noisy segmentation in case of tissue Model fusion.

8 Segmentation Conclusions

This lab gave us a good opportunity to experiment with different strategies for "traditional" segmentation.

We think that the results obtained are hugely affected by the data set that we have. This would explain in part why the MNI spaces didn't work as good as using ground truth information from the same homogeneous set. Also, the CSF in the ground truth of this set

is very different from the one we had in the previous lab where we coded the EM algorithm, which could explain the different CSF results from one lab to the other.

As for the peak adaptive method for intensity homogenisation, it worked on all but the last two test images, the reason being because they are brains with a significant level of atrophy, which were not represented in the training set. We can see the peaks detected in the following Figure 18

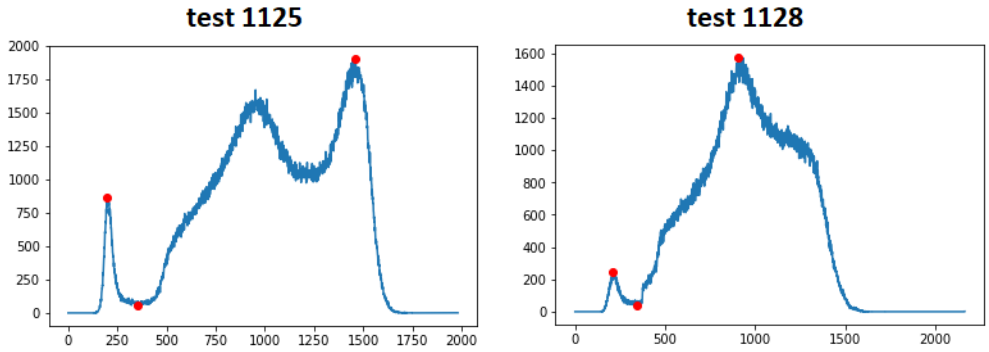

Figure 18: Histogram of brains with high atrophy

Even if the current implementation of the peak adaptive method is not robust to severe atrophy, the following Figure 19 shows how interesting it can be by simply agreeing to it when the probability of the tissue is above a certain threshold (0.8 in this example). The average white matter and gray matter dice scores for the images without severe atrophy are 0.92 and 0.94 respectively. However, the CSF lacks some good segmentation, so this would be a line of future work to refine.

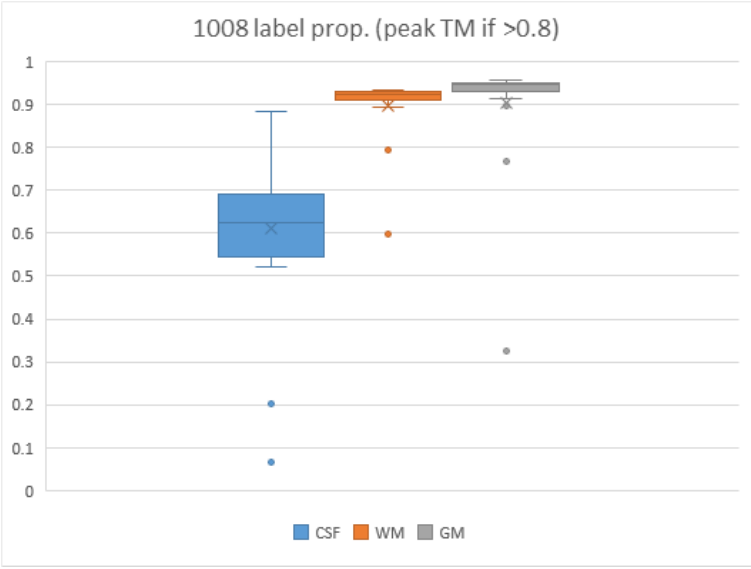

Figure 19: Experimental combination of peak adaptation Tissue Model and the label propagation of the training image 1008

We can also see how the ground truth of this set was most probably annotated on the coronal slices, exemplified in the following Figure 20. This could have a huge effect on the evaluation of our algorithms, since we don't evaluate the segmentation on the real truth, but on the given annotations. If that was the case, we could apply the different segmenting algorithms on these neat coronal slices, and we would probably reach a better consensus with the given ground truth.

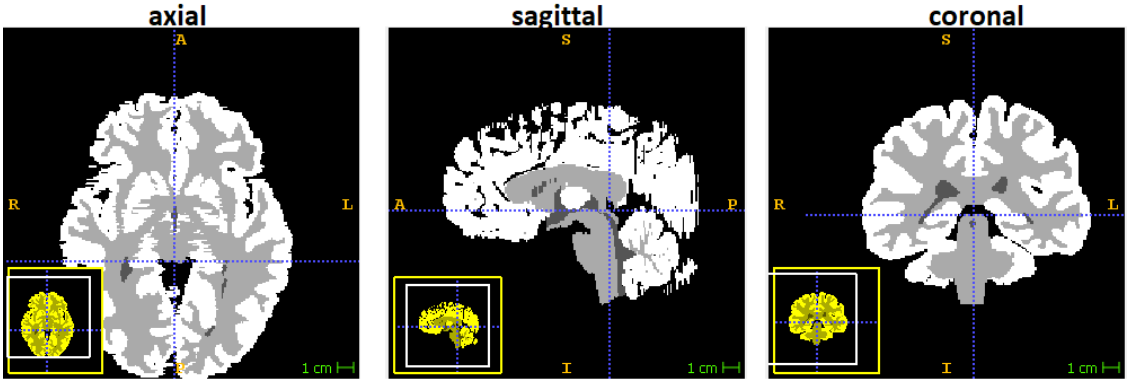

Figure 20: Different anatomical views show different tissue homogeneity

References

[1] F. van der Lijn, M. de Bruijne, Y.Y. Hoogendam, S. Klein, K. Hameeteman, M. M. B. Breteler and W.J. Niessen, Cerebellum Segmenta- tion in MRI Using Atlas Registration and Local Multi-Scale Image Descriptors, Proceedings of IEEE International Symposium on Biomedical Imaging: Macro to Nano, 2009, in pre