Atlas-based Image Segmentation

Medical Image Registration and Applications

Joaquin Oscar Seia Kaouther Mouheb

I. Introduction

Multi-Atlas based segmentation aims to improve the performance of segmentation algorithms using prior knowledge about the task. This prior knowledge is usually derived from a set of training images with available ground truths. There exist multiple ways to embed this information into a segmentation framework. For instance, one can use the labels of the atlas as initialization to an expectation-maximization based method, use the tissue models and the probability maps to derive the labels or condition the posterior probabilities generated by the Expectation Maximization (EM) segmentation algorithm by multiplying them by the atlas tissue probability maps .

II. Objective

In this second part of the laboratory work, we used the multi-atlas that we generated in the first part to segment brain tissues. We compared the different possible ways to perform the segmentation, we inspected the benefits of using expectation maximization algorithm against using simple approaches that rely only on the atlas information. Finally, we analyzed not only the effect of introducing the atlas in different parts of EM algorithm but also the difference in performance induced by using the standard Montreal Neurological Institute (MNI) Atlas versus the dataset built with the training set coming from the same distribution as the provided test set.

III. Methodology

The implementation of the task of atlas-based segmentation involves two major steps:

A. Registration

In the first step, the atlas images need to be registered to the target images that we need to segment. The resulting optimal registration parameters, were used to propagate the atlas probability maps to the space of each one of the cases.

B. Segmentation

Having obtained the atlas in the space of the target image, it can be used to perform the segmentation in multiple ways. Hereafter we describe the ways of using an atlas to perform segmentation that we explored in this assignment.

1) Non-Expectation Maximization methods

An atlas alone can provide two types of information that can be used to obtain a segmentation in three different ways.

- *Intensity-based segmentation*: Atlases can be used to build the tissue models. This models assign to each possible intensity value a probability to belong to each of the tissue classes. Each voxel can then be assigned to the class that has the highest probability for its intensity (maximum a posteriori - MAP appoach).

$$Y(x,y,z) = \max_k TM_k(I(x,y,z)), \quad k = \{CSF, GM, WM\}, \tag{1}$$

  where TM denotes the tissue maps and Y the classification label assigned to the voxel at position (x, y, z) with intensity $I(x, y, z)$.
- *Position-based segmentation*: The propagated atlas probability maps assign to each voxel the probability of belonging to each of the tissues based on its position. Again, each voxel can be assigned to the class that has the highest probability for its location.

$$Y(x,y,z) = \max_k TPM_k(x,y,z), \quad k = \{CSF, GM, WM\}, \tag{2}$$

  where TPM_k denotes the probability map of tissue k obtained from the muli-atlas, Y the classification label assigned to the voxel at position (x, y, z).
- *Combined segmentation approach*: Multiplying the probability maps generated using the tissue model with the propagated atlas probability maps can provide a segmentation result that takes into account both intensity and position information.

$$Y(x,y,z) = \max_k \{TPM_k(x,y,z) * TM_k(I(x,y,z))\}, \quad k = \{CSF, GM, WM\}, \tag{3}$$

2) Atlases as initialization for Expectation Maximization algorithm

Another way to use the atlases is to include them as initialization to a segmentation method such as EM+GMM. This can be done by using one of the membership probability maps obtained with any of the methods described in the previous section as the weights used to estimate the initial parameters (mean, covariance and priors) for each cluster in the Gaussian Mixture. Then the optimization of EM algorithm takes place until a convergence criteria is fulfilled.

3) Atlases in Expectation Maximization algorithm

Atlases can further be used to improve the segmentation performance by embedding them into the EM process. This can be achieved in different ways. In this work we focus on the following two:

- *After EM*: After obtaining the final membership probability maps (posteriors) from the EM+GMM process, they are multiplied with the probability maps of the atlas. This helps to use the location information in the atlas to refine the intensity-based results obtained using the EM algorithm.
- *Inside EM*: Since EM is an iterative approach, the atlases can be used at each iteration by multiplying them with the

probability maps of EM. This will guide the segmentation throughout the EM process by introducing a location conditioning into each iteration to avoid the intensity-based results from diverging into a non-plausible result.

$$w_{ik} = \frac{p_k(v_i|\theta_k).\alpha_k}{\sum_{m=1}^{K} p_m(p_m(v_i|\theta_m).\alpha_m)} \tag{4}$$

$$w'_{ik} = w_{ik} * TPM_k(v_i), \tag{5}$$

$$Y(v_i) = \max_k w'_{ik}, \tag{6}$$

where TPM_k denotes the probability map of tissue k obtained from the muli-atlas, Y the classification label assigned to the voxel v_i.

IV. Project Management

This work was conducted in two main phases. First, we focused on obtaining the correct registrations for the atlases with respect to each of our test images. Then, we modified the segmentation method based on the feedback that we obtained from the previous lab. Next, we implemented the different segmentation methods, tested them on a subset of cases, and proceeded to run them on the complete provided test set. Finally, we reported the results, measuring both the segmentation performance by dice score and the computational cost by time and iterarions required. Because of the workload and the long time required for the registration step and the segmentation to converge, this work took longer than the time assigned for it, we would suggest that for next editions of these labs more than one week should be assigned to it.

V. Implementation

In this lab work, Python was used as a programming language. We used the EM+GMM segmentation framework that we have previously developed after including the changes suggested in the correction. Elastix was used to perform the registration process. The source code can be found in https://github.com/kakou34/misa_lab

A. Registration

The registration process was performed in the jupyter notebook registration_elastix.ipynb under the registration folder. We used the same functions elastix_wrapper and transformix_wrapper as in the first part of this laboratory to perform the registration. We use each of the test images as a fixed image and both MNI and our atlas templates as moving image.

Differently from previous laboratory, we utilized the Parameter Map file named Par0010affine[1], instead of Parameters.Par0009.elastic[2]. This was done in order to avoid points sampling errors introduced by the latter.

The resulting parameter maps returned by elastix were then used to apply the registration transformation on the atlas probability maps of each tissue to perform the 'label' propagation step. For this task, we used a BSplineInterpolator of degree 2 in the label propagation process and saved the results as floats. The results of interpolation are not in the range [0, 1] because a spline of degree 2 can incorrectly interpolate values out of the range. Therefore, the values higher than 1 in the resulting probability maps were mapped to 1 and the values lower than 0 were mapped to 0.

B. Segmentation

In the segmentation part we implemented each of the previously mentioned segmentation approaches using the atlas that we built in the first part of this laboratory work and also MNI atlas. This section provides the implementation details of each method.

1) EM+GMM Updates

We updated our EM class according to the feedback provided and based on the needs of this new task. The changes applied can be summarized as follows: (all the lines reported are in in the src.py file unless stated differently)

- Priors made ourselves sure that the priors were updated in each iteration in the maximization step (line 322). They were given by the sum of weights of all the voxels for each class divided by the total number of voxels.
- Instead of using only the labels inside each region (z_k vector according to the lecture slides) to calculate the parameters in the maximization step, all the weights/posteriors associated to the tissue were considered.
- A first M-step is applied at the start of each segmentation process to estimate the initial parameters. (lines 163, 165)
- Two more initialization options are added:
  - Using tissue models (lines [142, 148]) by initializing the posteriors (weights) of each voxel to the probability of the tissue given the voxel's intensity according to the corresponding TM.
  - Using tissue probability maps (lines [150-151]) by setting the initial posteriors (membership weights) to the tissue probability maps provided by the atlas.
- Atlas after EM: we add the option of using an atlas to refine the segmentation results obtained by EM after the algorithm meets a stopping criterion. This is done by multiplying the final posteriors obtained by EM with the tissue probability maps provided by the atlas (lines [244-246]).
- Atlas into EM: we provide the option of using the atlas within the EM process to guide the segmentation. This is achieved by multiplying the posteriors (membership weights) obtained at each iteration with the atlas probability maps (lines [264-267]).
- label matching: when using the atlas inside the EM process or in the "after" mode, we need to make sure that the categorical label assigned to each class in the atlas match with the ones assigned by EM (line 271). This problem is particular to the case of k-means initialization, where the categorical label given to each component can vary randomly. The matching is done in the match_labels function (lines [274-283]) by first obtaining the categorical segmentation results for both the probability maps of the atlas and the partial EM segmentation result, then inside each region in the categorical atlas, we count the number of voxels belonging to each class based on the partial EM result, and the region of the atlas is matched with the one that provides the maximum overlap from EM.

[1]https://elastix.lumc.nl/modelzoo/par0010/

[2]https://elastix.lumc.nl/modelzoo/par0009/

2) Segmentation Without EM

We implemented the three possible methods to apply segmentation using the information provided by the atlas only. All line numbers reported here are in utils.py unless stated differently.

- *Intensity-based, using tissue models* (lines [155-174]): The voxels inside the brain mask are selected. Then probabilities provided by the tissue models are used to assign each voxel to the tissue that has the highest probability for its intensity value.
- *Position-based, using tissue probability maps* (lines [177-189]): An argmax operation is applied on the probability maps provided by the atlas to define the most probable tissue for each pixel position. Then the brain region is extracted using the provided brain masks.
- *Combined approach segmentation* (lines [192-220]): a combination of the two previous approaches. For each voxel, the final probabilities for each tissue are given by the probabilities obtained from the tissue models for the voxel's intensity and the ones given by the probabilistic atlas for its position. The segmentation is obtained by taking the argmax of the final probabilities after masking the brain with the provided brain masks.

3) Segmentation With EM

The function brain_tissue_segmentation_em (line [105-152]) was used to perform the segmentation using both the atlas and EM+GMM algorithm using the previously described ExpectationMaximization class.

- Using the tissue models as initialization: this is achieved by setting the mean_init attribute to 'tissue_models' and the tissue_models attribute to the tissue models generated using the atlas. The latter will be used to initialize the EM algorithm as previously explained.
- Using the tissue probability maps as initialization: this is achieved by setting the mean_init attribute to 'label_prop' and the atlas_map attribute to the probability maps given my the atlas. The latter will be used to initialize the EM algorithm as previously explained. It is important to highlight that since the background class was not going to be used in EM algorithm runs due to the in-brain mask we applied to select the voxels, we took probabilities of csf, gm, wm for each voxel and normalized again in order to make a probability map out the image.
- Using the atlas after EM: this is achieved by setting the atlas_use attribute to 'after' and the atlas_map attribute to the probability maps of the atlas.
- Using the atlas into EM: this is achieved by setting the atlas_use attribute to 'into' and the atlas_map attribute to the probability maps of the atlas.

Note that these methods can be used in combination. For instance, we can use the tissue labels to initialize the EM components and use the atlas into the EM process in the same run. Finally, for all the EM experiments, the maximum number of iterations was set to 300 by default.

VI. Results

In this section we will give details about the quantitative and qualitative results obtained during the different experiments conducted in this laboratory work.

A. Segmentation without EM

For each test image, we performed the segmentation using only the information provided by the atlas without the EM optimization (tissue models, tissue probability maps, and both) as explained in the previous sections. A global visualization of the results can be appreciated in the boxplots in figure 1.[3]

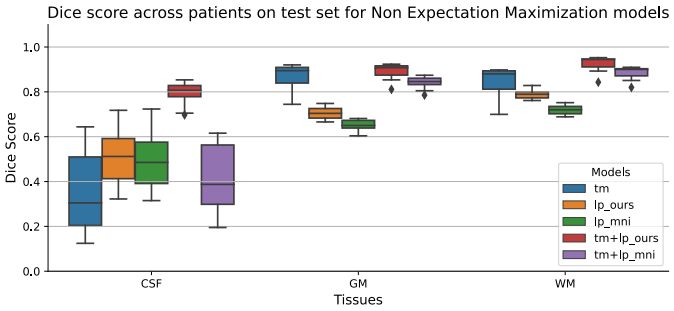


Fig. 1. Boxplots of Dice scores across all patient for each segmentation approach using **only atlas information**.

More details about the experiments are reported in table I

Tissue	Algorithm	mean_dice	std_dice
CSF	lp_mni	0.4952	0.118
	lp_ours	0.5092	0.1212
	tm	0.354	0.1718
	tm+lp_mni	0.4089	0.1484
	tm+lp_ours	**0.7967**	0.0424
GM	lp_mni	0.6514	0.0237
	lp_ours	0.7036	0.0244
	tm	0.8721	0.0531
	tm+lp_mni	0.8432	0.0232
	tm+lp_ours	**0.8956**	0.0301
WM	lp_mni	0.7208	0.0201
	lp_ours	0.7899	0.0199
	tm	0.8526	0.0557
	tm+lp_mni	0.8889	0.0236
	tm+lp_ours	**0.9295**	0.0282

TABLE I
MEANS AND STANDARD DEVIATIONS OF THE DICE SCORE OF EACH **ATLAS-ONLY SEGMENTATION** APPROACH ACROSS ALL PATIENTS.

The results show that across tissues, in all the methods CSF was the hardest to segment which resulted in a very low Dice score with high dispersion except in the case of using both tissue models and probability maps from our atlas.

When comparing the methods, we can see that using intensity information (tissue models) provide more discriminative information to perform the segmentation compared to the results obtained using just position information. However, combining both information resulted in an improvement of the dice score for all tissue types.

B. EM Initialization

To compare the different possible initialization methods of the EM procedure (K-means, tissue models, probability maps) three experiments were performed on the test set for each atlas. The boxplots in figure 2 summarize the results whereas table II provides more details. In this figure, we provide the results of the best model obtained without EM as a baseline (*tp+lp_ours* in blue).

[3]In all the figures TM refers to tissue models, LP refers to Label Propagation, KM refers to k-means

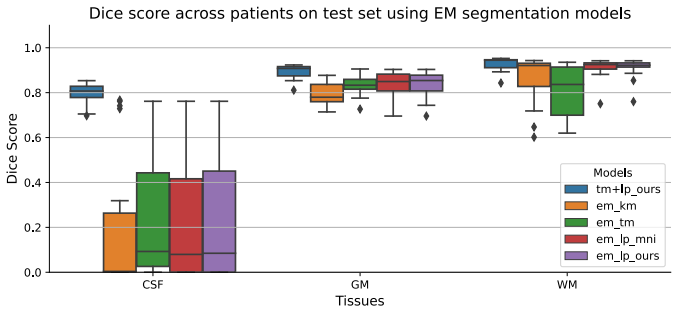


Fig. 2. Boxplots of Dice scores across all patient for each **EM initialization** approach.

Tissue	Algorithm	mean_dice	std_dice	mean_iter	std_iter
CSF	tm+lp_ours	0.7967	0.0424	–	–
	em_km	0.2015	0.2992	291	29
	em_lp_mni	0.2271	0.2759	275	54
	em_lp_ours	0.242	0.2927	**260**	65
	em_tm	**0.2542**	0.2797	280	41
GM	lp_mni	0.6514	0.0237	–	–
	em_km	0.7961	0.0521	291	29
	em_lp_mni	0.8364	0.0575	275	54
	em_lp_ours	**0.8364**	0.0567	**260**	65
	em_tm	0.8325	0.0396	280	41
WM	lp_mni	0.7208	0.0201	–	–
	em_km	0.8668	0.1009	291	29
	em_lp_mni	0.9121	0.0419	275	54
	em_lp_ours	**0.9121**	0.0415	**260**	65
	em_tm	0.8021	0.1156	280	41

TABLE II
MEANS AND STANDARD DEVIATIONS OF THE DICE SCORES AND ITERATIONS OF **EM SEGMENTATION USING DIFFERENT INITIALIZATION TECHNIQUES**.

We can see from the given figures that using EM with any initialization type does not provide better results than the methods using only the atlas. This further highlights how powerful can be the inclusion of atlas information in segmentation models. We can notice that the CSF in particular is very poorly segmented compared to the other tissues for which the mean performance is almost 0.8 Dice score.

We noticed that all initialization methods provide slightly different results. Label propagation initialization, done with our atlas, provides slightly higher and less dispersed Dice scores on the GM and WM classes. The worst segmentation method varies according to the tissue analyzed, being k-means initialization the worse in CSF and GM, whereas tissue maps initialization works the worst in white matter.

Knowing that the maximum number of iterations was set to 300 during these experiments, we can notice that the algorithm is stopping mostly because of reaching the maximum number of iterations rather than converging to an optimal log-likelihood value. From the results we conclude that it is hard to choose an initialization method as an optimal one and favoring one with respect to the other is a trade-off between the different tissues, however initialization with label propagation from our atlas seems to be the best choice.

C. Atlas After EM

In the next experiments, we investigate the effect of using the atlas probability maps to refine the segmentation results obtained using the EM algorithm after convergence using the different possible initializations. The global results are shown in figure 3 and details about the segmentation performance are given in table III.

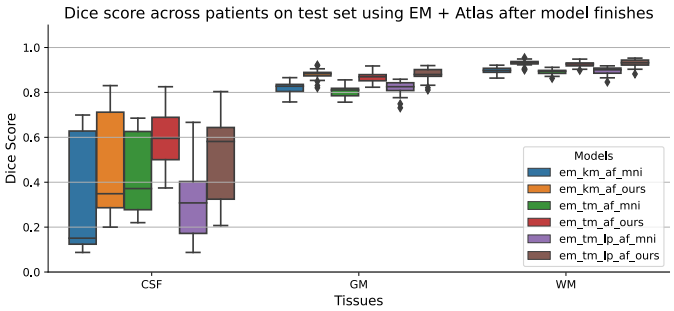


Fig. 3. Boxplots of Dice scores across all patient for segmentation using **atlas after EM** with different initialization methods. 'af' stands for after use of the atlas in EM.

Tissue	Algorithm	mean_dice	std_dice
CSF	em_km_af_mni	0.3322	0.2601
	em_km_af_ours	0.4735	0.2359
	em_tm_af_mni	0.4247	0.1739
	em_tm_af_ours	**0.5972**	0.1326
	em_tm_lp_af_mni	0.3239	0.1947
	em_tm_lp_af_ours	0.5046	0.1848
GM	em_km_af_mni	0.8193	0.0296
	em_km_af_ours	0.8789	0.0254
	em_tm_af_mni	0.8031	0.0251
	em_tm_af_ours	0.8659	0.0231
	em_tm_lp_af_mni	0.8182	0.0348
	em_tm_lp_af_ours	**0.8793**	0.0313
WM	em_km_af_mni	0.8954	0.0168
	em_km_af_ours	**0.9302**	0.0139
	em_tm_af_mni	0.8907	0.0129
	em_tm_af_ours	0.925	0.0123
	em_tm_lp_af_mni	0.8946	0.0193
	em_tm_lp_af_ours	0.9292	0.0181

TABLE III
MEANS AND STANDARD DEVIATIONS OF DICE SCORES OF SEGMENTATION USING **ATLAS AFTER EM** WITH DIFFERENT INITIALIZATION METHODS. 'AF' STANDS FOR AFTER USE OF THE ATLAS IN EM.

Following the same trend as the results from the previous experiments on EM initialization (figure 2 and table II), we noticed a relatively low performance for segmentation of CSF compared to the other tissues. After a qualitative assessment of the results we found out that this might be due to the matching errors between the EM segmentation results and the atlas labels in the CSF region surrounding the brain. The ground truth masks seem to be a little coarse on the outer boundary of GM, so including some CSF as GM which is correctly classified by our algorithms but miss-matching the ground truth. Since CSF is the class with the lowest prevalence, this may have a higher impact in CSF than in GM. Taking that into account, we can say that overall, using the atlas after the segmentation with EM improved the segmentation results in terms of Dice score. One more time, identifiying the best algorithm is a matter of the tissue taken into account. Acording to the boxplots, for CSF initialization with tissue maps and inclusion of our atlas led to highest dice score. For GM and WM initialization with tissue models plus label propagation from our atlas, and the addition of our atlas in EM, led to higher performance just slightly surpassed by k-means initialization in the second tissue.

One clear outcome of these experiments is the difference in the results of using MNI atlas vs using our atlas. It can be seen

both in the boxplot and in table III that the use of our atlas led to higher performance in all tissues.

D. Atlas Into EM

In this set of experiments, we evaluate the segmentation performance when embedding the atlas information into the EM process by multiplication at the end of each iteration. The results are summarized in figure 4 and more details are given in IV

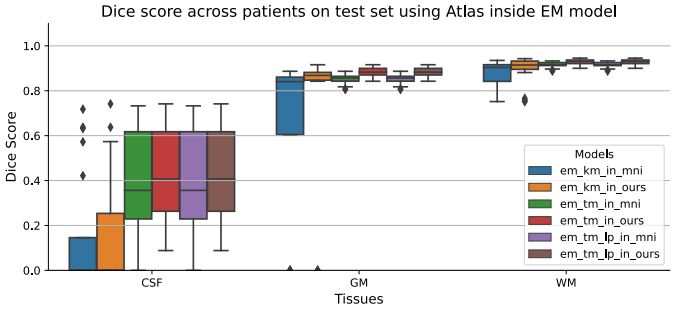


Fig. 4. Boxplots of Dice scores across all patient for segmentation using **atlas into EM** with different initialization methods.

Tissue	Algorithm	mean_dice	std_dice	mean_iter	std_iter
CSF	em_km_in_mni	0.1536	0.2674	110	88
	em_km_in_ours	0.1683	0.2508	101	75
	em_tm_in_mni	0.3797	0.2546	84	23
	em_tm_in_ours	**0.4216**	0.2186	**83**	14
	em_tm_lp_in_mni	0.3779	0.2572	105	71
	em_tm_lp_in_ours	**0.4216**	0.2186	84	13
GM	em_km_in_mni	0.6356	0.3771	110	88
	em_km_in_ours	0.7002	0.3597	101	75
	em_tm_in_mni	0.8509	0.0228	84	23
	em_tm_in_ours	**0.8818**	0.0208	**83**	14
	em_tm_lp_in_mni	0.8509	0.0228	105	71
	em_tm_lp_in_ours	**0.8818**	0.0208	84	13
WM	em_km_in_mni	0.8712	0.0697	110	88
	em_km_in_ours	0.8874	0.0685	101	75
	em_tm_in_mni	0.9147	0.0135	84	23
	em_tm_in_ours	**0.9267**	0.0136	**83**	14
	em_tm_lp_in_mni	0.9147	0.0135	105	71
	em_tm_lp_in_ours	**0.9267**	0.0136	84	13

TABLE IV
MEANS AND STANDARD DEVIATIONS OF THE DICE SCORE AND ITERATIONS OF **ATLAS INTO EM SEGMENTATION** WITH DIFFERENT INITIALIZATIONS.

We can notice that although the average dice score is not significantly higher than the previous experiments, adding the atlas into EM helped the algorithm to converge in less iterations. Consistently with the previous experiments, the CSF tissue was poorly segmented in this set up as well, especially when using the k-means segmentation, due to the label matching problem that we previously mentioned. To address this issue, we tried to match the categorical labels with mean or median intensity or by label frequency counting (as explained before), however, no significant improvement was obtained, mainly because k-means is not a good segmentation.

E. MNI Atlas

From the results of all previous experiments and the summary given in figure 5, we can see that in most cases, the atlas that we built provides better results in terms of Dice score compared to the MNI atlas for all tissue types. This support the idea that using atlases of distributions closer to the data evaluated in test time may have a positive impact on the segmentation performance. However, this results might be biased, a further inspection at the MNI atlas showed that it didn't contain the segmentation for the cerebellum region, which is on the contrary contained in the provided labels for the test set and also in our atlas. Because of this, the overall use of this atlas might have had a lower performance in the mentioned region and therefore the comparison is not completely fair.

Qualitative segmentation results are for all previous experiments are available in the appendix.

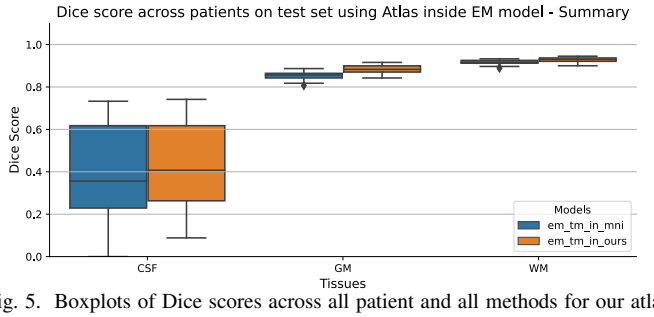


Fig. 5. Boxplots of Dice scores across all patient and all methods for our atlas and MNI atlas.

F. Comparison of the best results for each kind of segmentation

Figure 6 shows an overall summary of the best results obtained in each experiment. We can see a final comparison of the methods that had the best overall performance in each part of this laboratory. The figure shows that using the simple model of segmenting based on tissue maps and label propagation probabilities without EM, is the best alternative. The performance of this method is either equal or superior for all the tissues, and since the method is not iterative, the computational cost is minimal compared to EM algorithm.

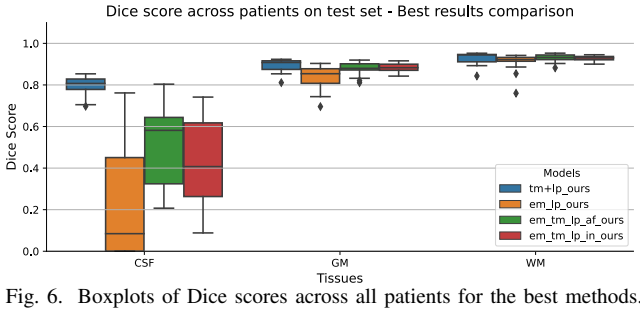
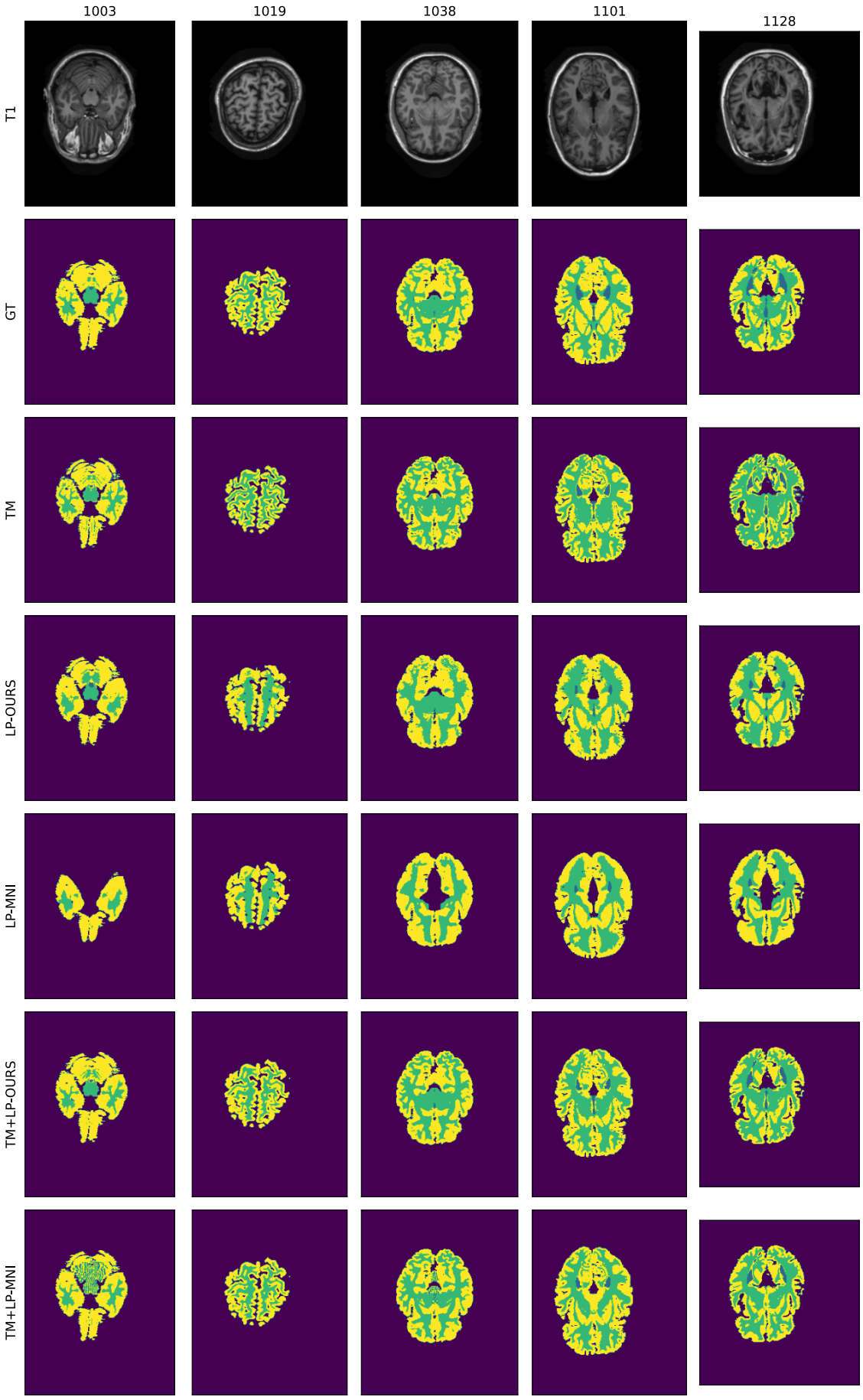


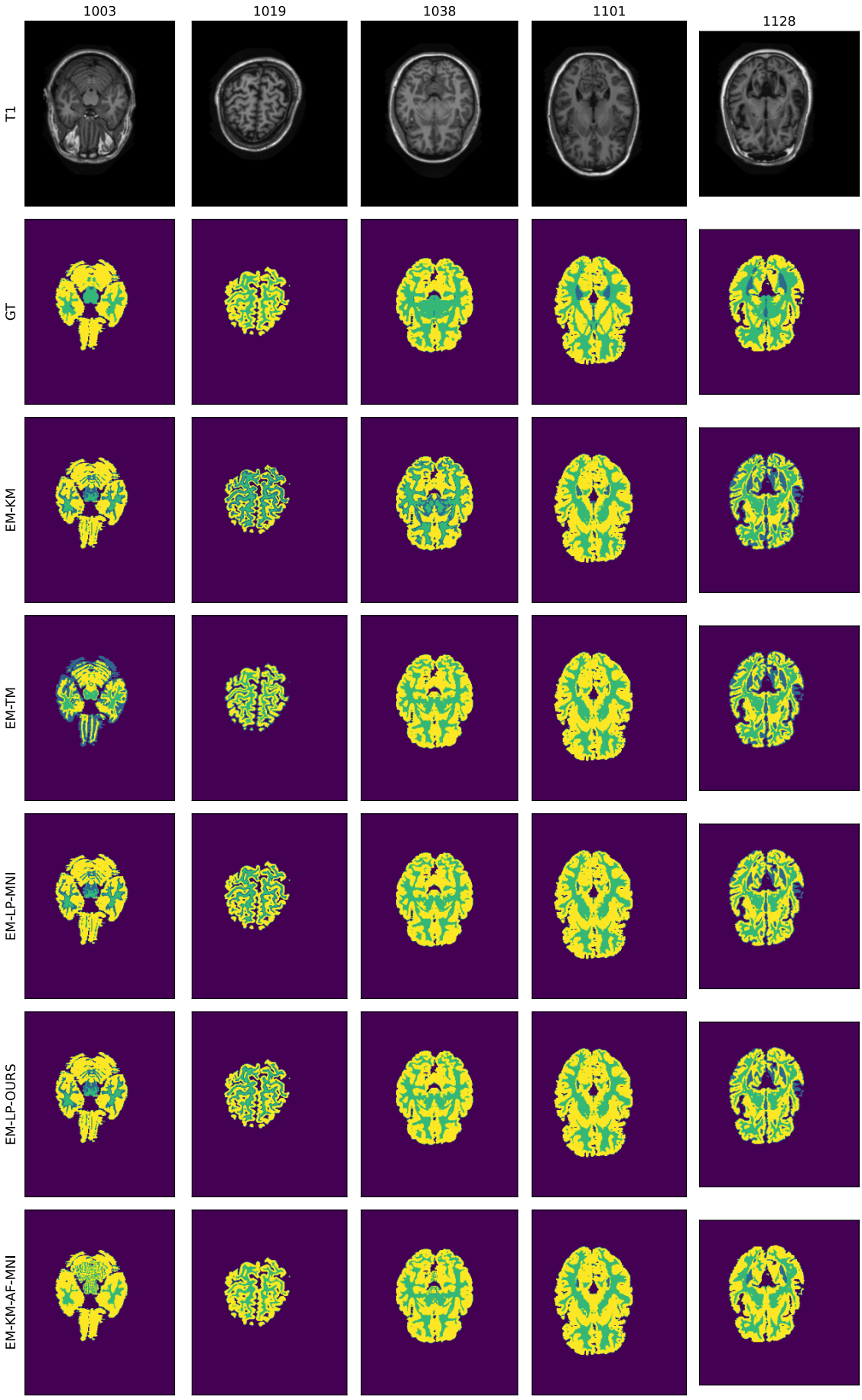
Fig. 6. Boxplots of Dice scores across all patients for the best methods.

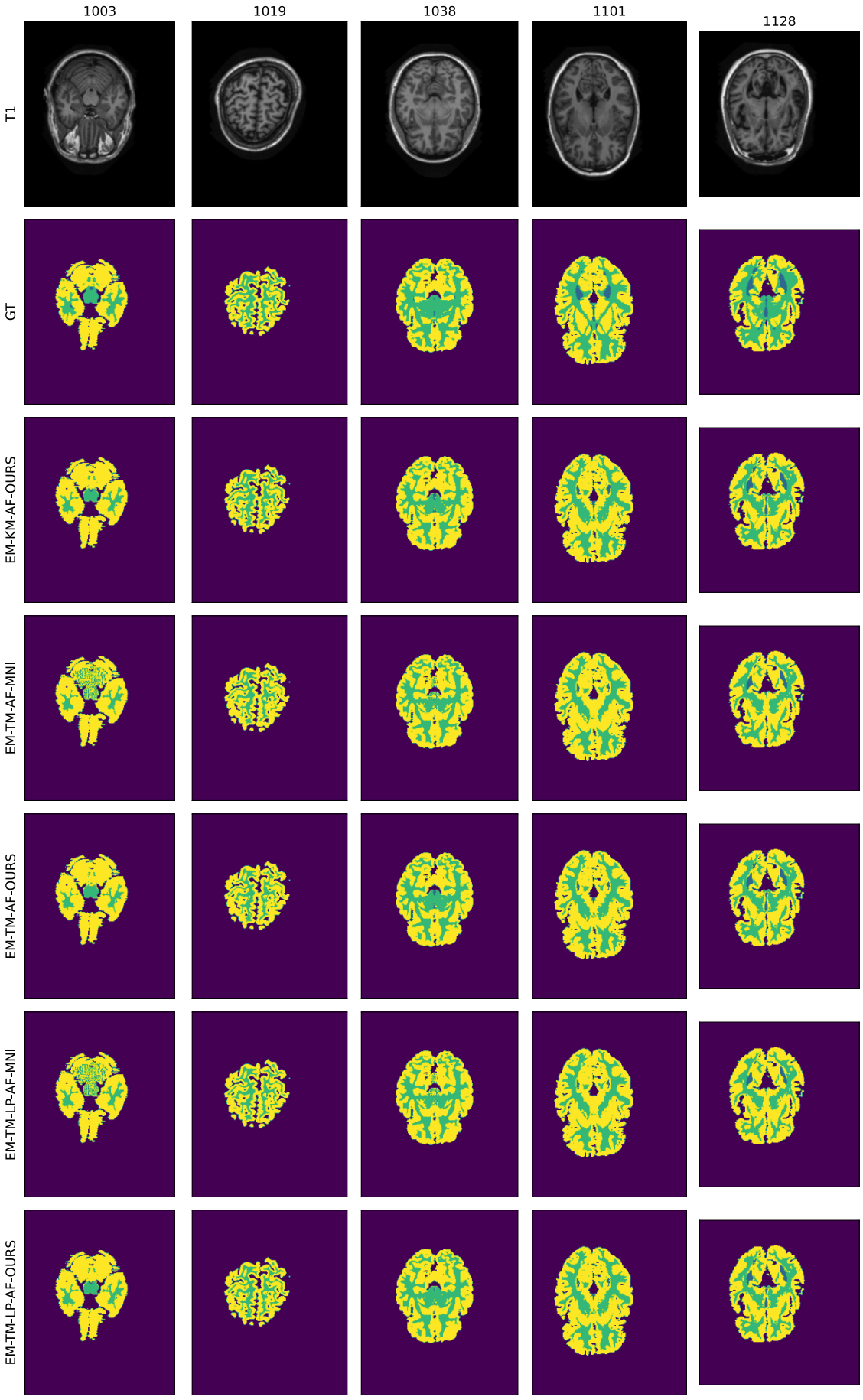
VII. CONCLUSION

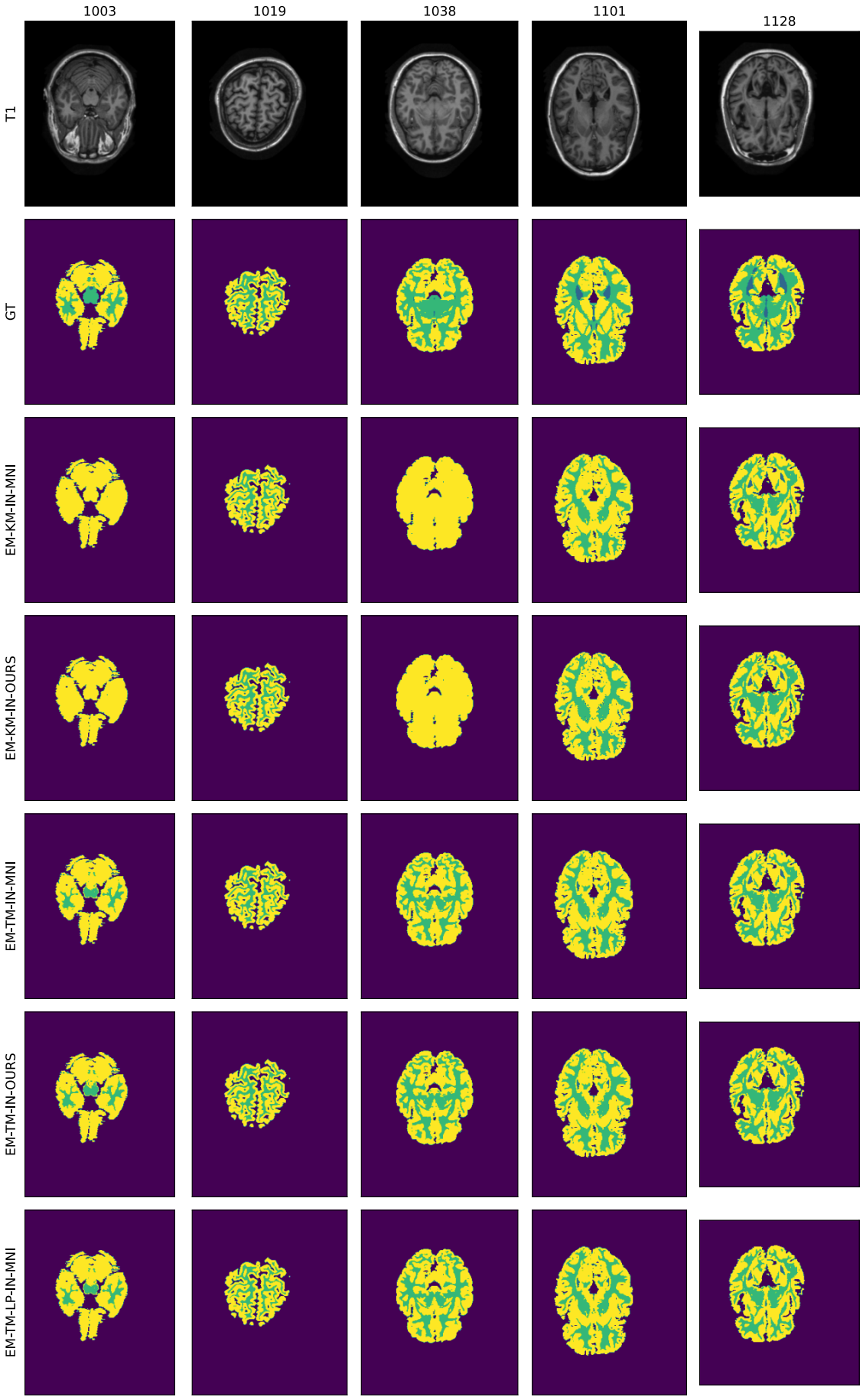
Across this laboratory, we could see the importance of the inclusion of atlas information in the segmentation of brain tissues. As previously stated, the simple model of segmenting based on tissue maps and label propagation probabilities without EM, resulted the best alternative both in segmentation performance and computational cost. In addition, the analysis of the different atlases available showed important differences, however further studies should be performed in order to determine if the higher performance of our multi-atlas was due to the closer resemblance of the test set or due to the inclusion of the cerebellum areas.

APPENDIX


1003
1019
1038
1101
1128
T1
GT
TM
LP-OURS
LP-MNI
TM+LP-OURS
TM+LP-MNI


1003
1019
1038
1101
1128
T1
GT
EM-KM
EM-TM
EM-LP-MNI
EM-LP-OURS
EM-KM-AF-MNI


1003
1019
1038
1101
1128
T1
GT
EM-KM-AF-OURS
EM-TM-AF-MNI
EM-TM-AF-OURS
EM-TM-LP-AF-MNI
EM-TM-LP-AF-OURS


1003
1019
1038
1101
1128
T1
GT
EM-KM-IN-MNI
EM-KM-IN-OURS
EM-TM-IN-MNI
EM-TM-IN-OURS
EM-TM-LP-IN-MNI

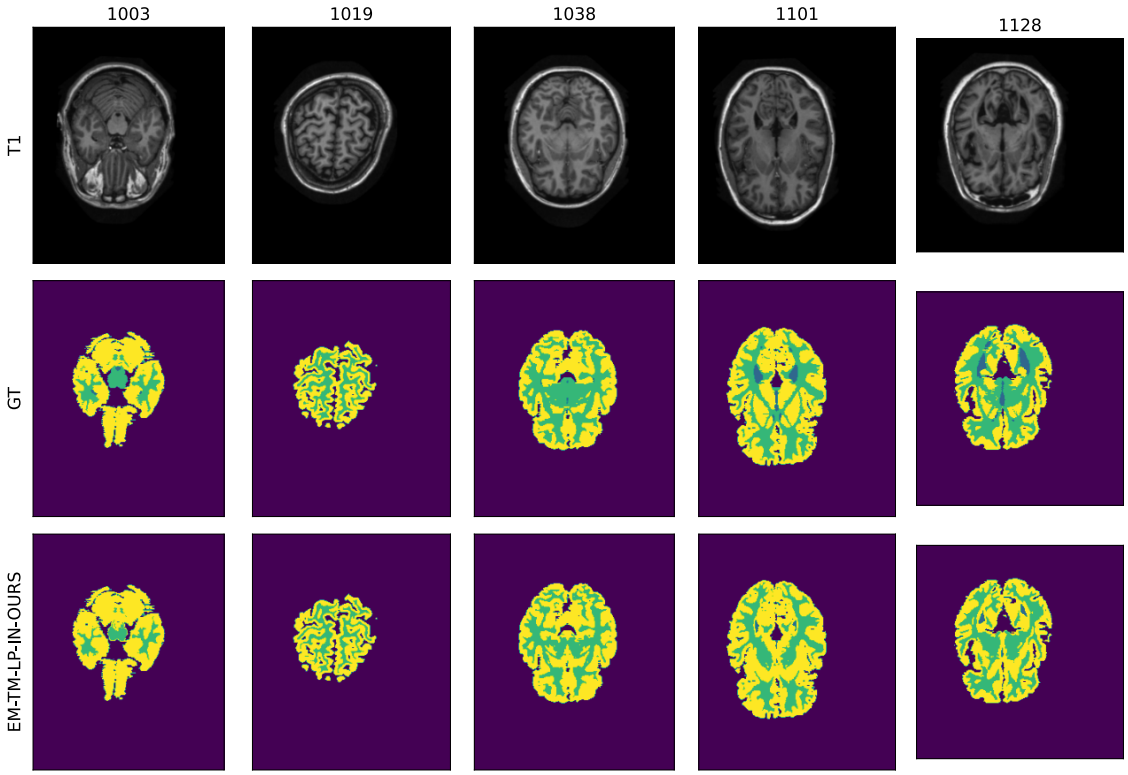


Fig. 7. Segmentation results using all methods for example cases (slice n°125)