TP segmentation - IMA204

Júlia Togashi de Miranda

January 2021

The goal of this practical session is to test the different segmentation techniques on several medical images. We have at our disposal: CT renal scan, brain MRI, heart MRI and dermoscopic images.

1 Threshold segmentation

1.1 Manually Found Threshold

1.1.1 Abdominal CT

For the first modality, an abdominal CT scan will be analyzed. The segmentation provided corresponded to the subject kidney, and if existent, a renal tumor. In 1 an example of the image provided can be seen (subject with renal tumor).

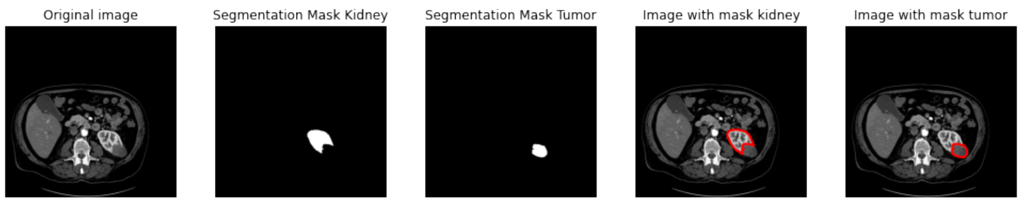

Figure 1: Image showing abdominal CT scan from a subject with renal tumor. Them respectively the mask of the kidney and the tumor, and their segmentation in the original image.

In 2 the histogram for this subject can be seen. As the image has a black background, it was expected that there would be a lot of values equal to zero, which difficulties the visualization. For that reason, the histogram excluding the zero values is also displayed, where it can be seen that the image has two distinct classes.

Below, in 3 the result for different manually tries of thresholds. First, a threshold that divides the two classes of the histogram shown above. That gives a good segmentation of all the organs from the patient.

As the kidney has a larger gray-level value than most organs, it was tried to increase the threshold. With a 140 threshold, the resulting segmentation shows the subject kidney and spine. As the spine has a even larger gray-level, if the threshold is even larger, we will only segment it, failing to segment the kidney. Using the Dice function with the kidney reference mask and threshold 150, we have a 0.42 score.

To conclude, for the CT renal scan we can manually find a thresholds that segments the kidney plus the subject spine, but not just the kidney (another method could be used to segment the just the kidney from this resulting image). As the tumor has a gray-level even similar to the other organs, it cannot be segmented using this method.

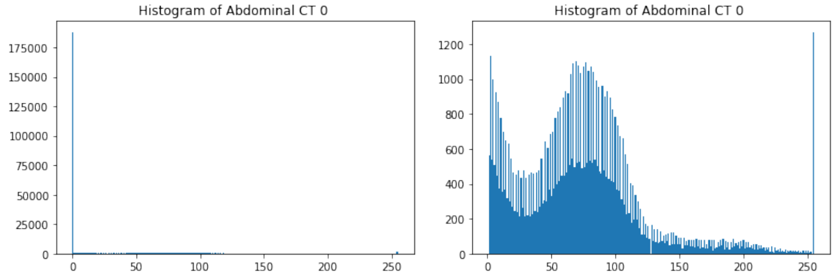

Figure 2: Image showing abdominal CT histogram for the patient above. On the left, we have the histogram including the 0 value, on the right, excluding the 0 value for better visualization.

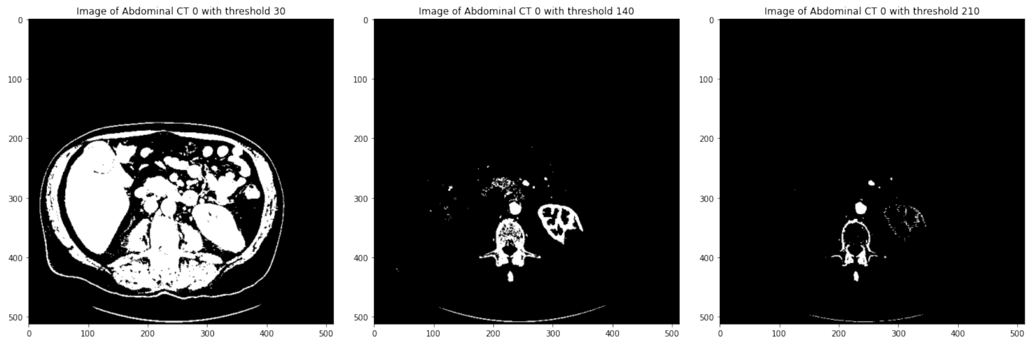

Figure 3: Image showing abdominal CT using respectively thresholds 30, 140 and 210

In the following, it's displayed the renal CT scan for another subject, this time without a renal tumor. It can be seen that the histogram for this patient has a similar distribution from the first one, an the remarks mad above are still valid. For this, a threshold equals to 150 gives a 0.46 score.

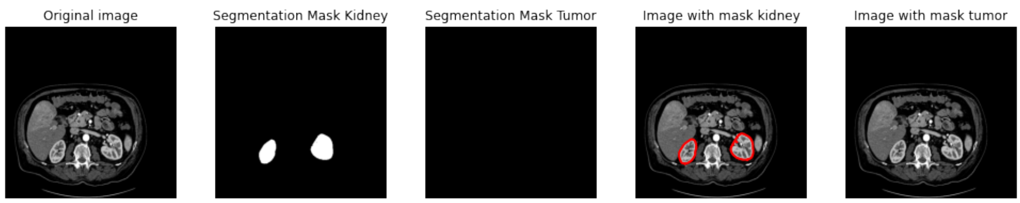

Figure 4: Image showing abdominal CT scan from another subject without renal tumor.

1.1.2 Brain MRI

For the following modality, it will be presented simultaneously two subjects Brain MRIs, where the goal is to segment the Corpus Callosum. Just looking at the images provided, we can see that the Corpus Callosum has a lighter gray tone that the gray matter around it, but a similar tone of the cerebellum and spinal cord.

As can be seen in 7, the histograms for the brain MRI seems to vary more than for the abdominal CT scan. For the first subject, the image is cropped to only show the patient scull, that explains why it has more values equal to 0. As consequently, the lighter area is also cropped, which results

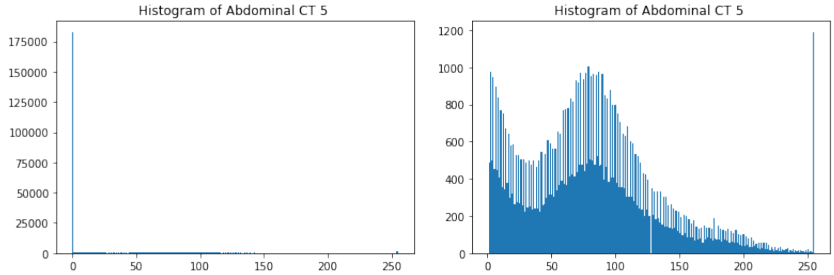

Figure 5: Image showing abdominal CT histogram for the patient above. On the left, we have the histogram including the 0 value, on the right, excluding the 0 value for better visualization.

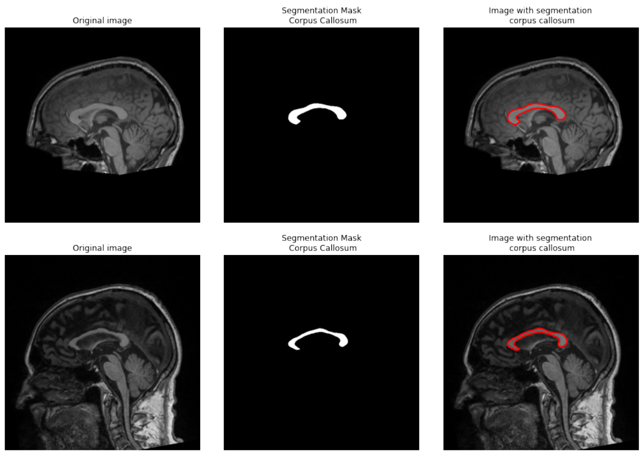

Figure 6: Image showing two examples of brain MRIs, with the segmented Corpus Callosum.

in the values concentrated in the 50-100 gray-level range (excluding the 0). The second patient has a wider range of values.

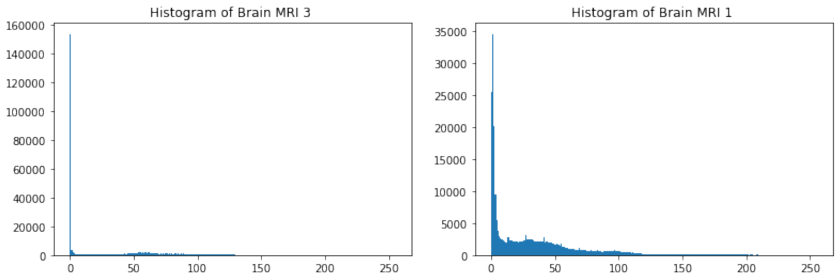

Figure 7: Image showing Brain MRI histogram respectively for the two subjects above.

Similar to the abdominal CT scan, there is no value of threshold that successfully just segments the Corpus Callosum, but there are good values that segments it from the gray matter around it. As the distribution for the histograms are different, it can be seen that the manual thresholds that give good results for both images is significantly different. For the segmentations above, the dice score is respectively equals to 0.38 and 0.11.

It can be remarked that the Corpus Callosum seems to have a similar C shape for all subjects, so using for example a mathematical morphology method with this element above the segmented results might give good results. But, given the limited forms available through skimage (diamond, disk, square and line), the results trying to use this tools didn't improve the result.

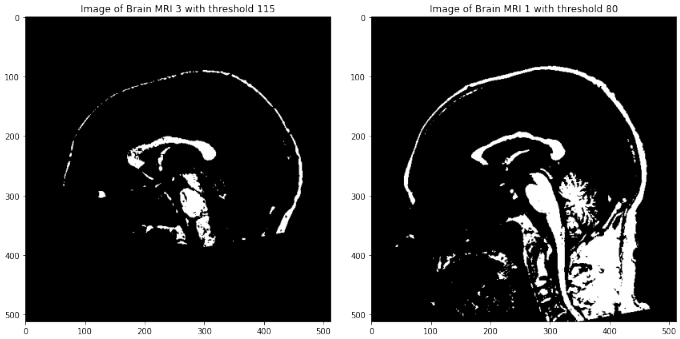

Figure 8: Image showing Brain MRI segmentation for patients above, each one with reasonable threshold value.

1.1.3 Skin lesions

For the following modality, it will be presented simultaneously three subjects skin photos, as the skin lesions vary a lot between subjects. The goal is to segment the lesion.

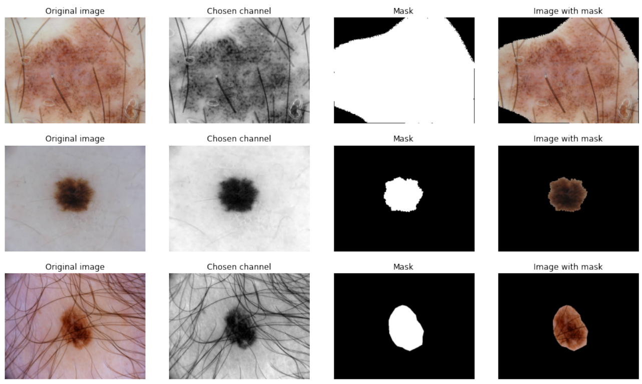

Figure 9: Image showing three examples skin lesions, with the segmented lesion.

As we can see in 10, the histograms vary even more. This time, different than for the brain MDI, the variability of the distribution is due to the large difference between lesions (shape and color) and the presence of hair.

For the subject 7 for example, it can easily be seen two classes in the histogram, that difference is less cleat for subject 10 and even lesser for 0 (which seems to have only one class).

For the segmentation, the results in case similar of subject 7 (very dark lesion, with well defined borders and no hair) are very good. For subject 10, the lesion also can be seen well, but the hair (dark) is segmented with the lesion). For subject 0, the segmentation of the lesion has some faults due to it being lighter. The threshold values change within a range between cases due to the difference in color of the lesion. For the segmentations above, the dice score is respectively equals to 0.76, 0.93 and 0.53. One remark is that in this modality, the segmentation obtained is the negative

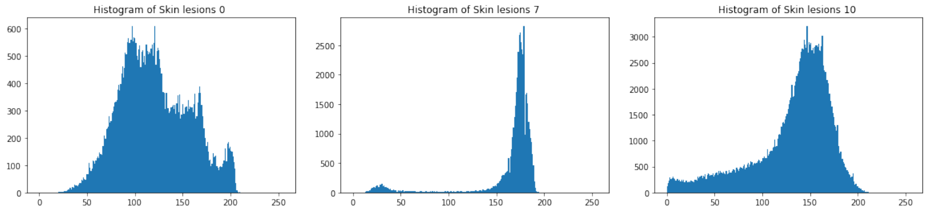

Figure 10: Image showing skin histogram respectively for the three subjects above.

of the expected mask, but we can easily obtain the image compatible with the mask by doing (im ¡ thresh) *255.

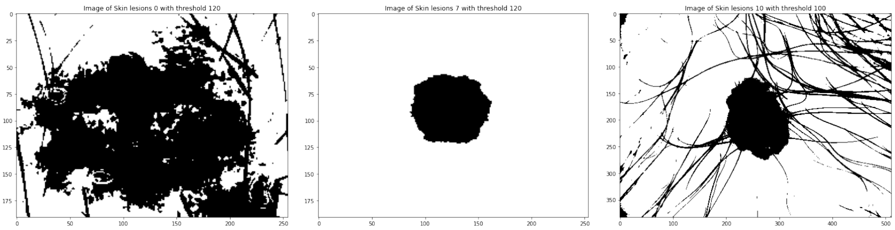

Figure 11: Image showing skin lesion segmentation (negative of the expected mask) for patients above, each one with reasonable threshold value.

For the hair case, as in the brain MDR case, mathematical morphology method might give good results. As for the light lesion, a regularization method as seen in IMA203 also will improve results.

Below, the results of the hair case after using a filter ”Chapeau haut-de-forme”, that is, doing the binary image above less the opening of that image. For the element structurant, we used a disk with radius equal to 9. The results improved drastically, getting a dice score of 0.92, almost as good as the case where there was no hair. This means that this is an effective technique to get better results, and can be applied for all the variations of threshold segmentation in this section.

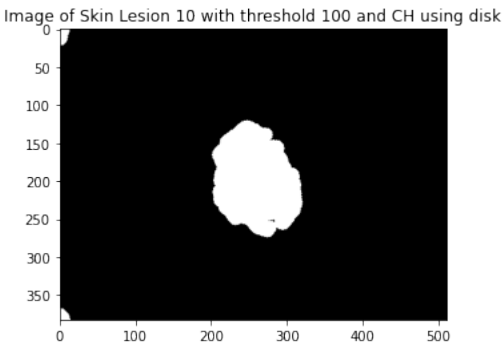

Figure 12: Image showing skin lesion segmentation after Chapeau haut-de-forme.

A last remark is that this manual segmentation won’t probably work as well for a subject with darker skin.

1.1.4 MRI heart

For the following modality, it will be presented simultaneously two subjects heart MRIs, where the goal is to segment the left ventricle.

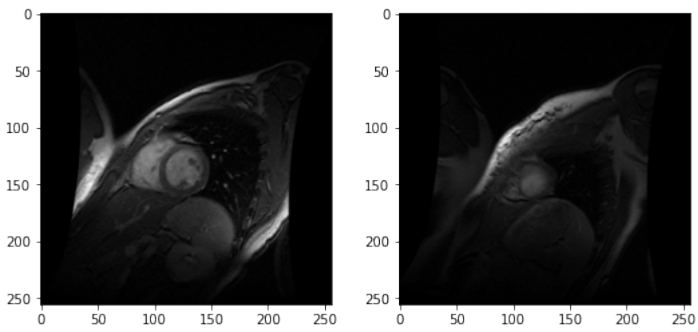

Figure 13: Image showing example of two different slices for the same time for heart MDI.

It can be seen that for this modality, the histograms between examples look similar, with a lot of values equal to zero, ans most of the concentrated in the interval 0-30.

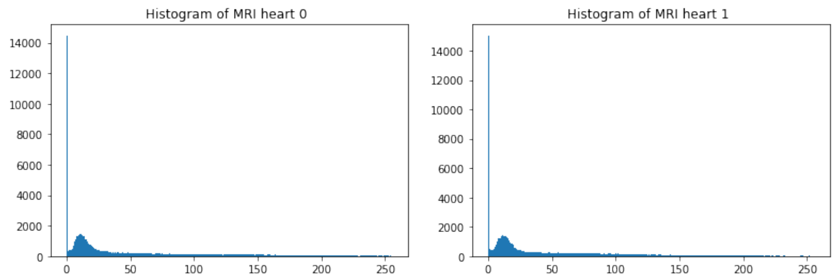

Figure 14: Image showing heart MRI histogram respectively for the examples above.

As can be seen bellow, the similar gray-level of the whole heart makes it difficult to correctly just segment the left ventricle, such that other close areas are also segmented. An mathematical morphology method might also help, but the left ventricle seems to change its form in the different slices, which can cause problems.

1.1.5 Comparison

The histograms fond for each different modality vary a lot. The manually found threshold that gives good results cannot be used across modalities and in some cases, not even for different subjects from a same modality.

1.2 Otsu's algorithm

As stated in the TP "Otsu's method looks for a threshold to split a gray-level image into two separate regions, based on their grey-level values". So it is expected that this method will have the same segmentation problems that stated in the previous subsection. Below are the threshold found for the same examples above and their respective dice score.

- Abdominal CT 0: Otsu threshold is 49 (dice score: 0.11)

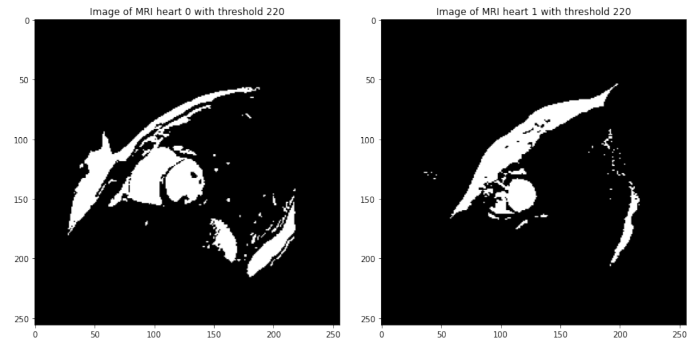

Figure 15: Image showing heart MRI segmentation for slices above, each one with reasonable threshold value.

- Brain MRI 0: Otsu threshold is 38 (dice score: 0.06)
- Skin lesion 0: Otsu threshold is 128 (dice score: 0.83)
- Skin lesion 7: Otsu threshold is 109 (dice score: 0.96)
- Skin lesion 10: Otsu threshold is 107 (dice score: 0.50)
- Heart MRI: Otsu threshold is 228

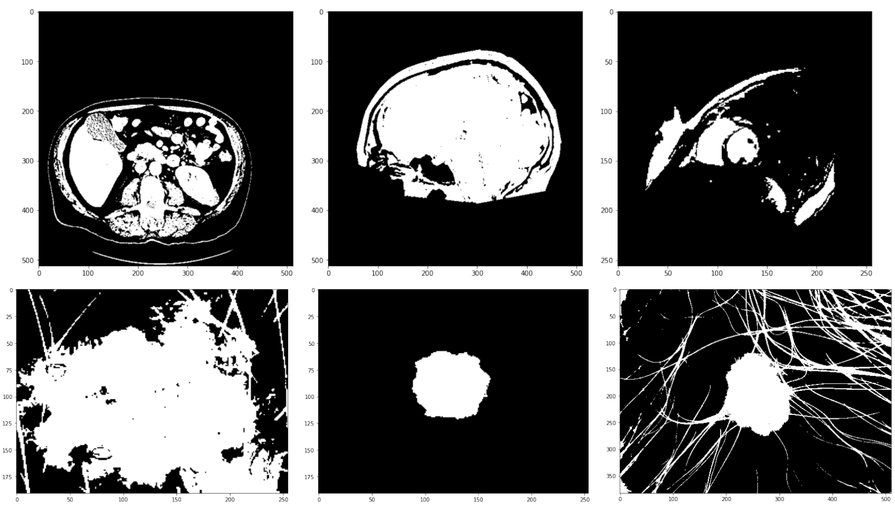

Figure 16: Images showing Otsu results for respectively the kidney CT; Brain MRI;Heart MRI and the three skin lesion examples.

We can see that the results vary a lot between modality. For the Kidney CT and the Brain MRI, the Otsu method has really poor results (worst that the ones found manualy). This occurs because the threshold that minimises the intra-class intensity variance is actually the one that divides the subjects organs from the background. For the skin lesion, the results as expected are good (the threshold is slight different giving better results). At last, for the Heart MRI, the result is similar from what was estimated manually, but it isn't so good.

1.3 K-means

For the kidney case, a small number of clusters results in the same problem from the manual selection of threshold, that is, the kidney and spine are in the same class. Increasing the number of clusters improve a little bit the results (0.48 score). The identification of the tumor is still very hard.

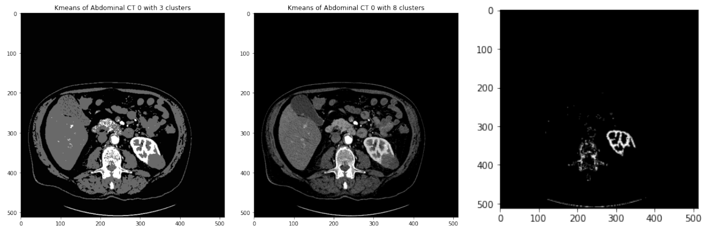

Figure 17: Image showing kidney segmentation using kmeans for different number of clusters. The mask is the obtained using 8 clusters.

The situation for the brain MRI is similar than for the kidney. Increasing the number of clusters improve a little bit the results (0.42), but the Corpus Callosum and spine cord still have too similar gray values, in a way that if we keep increasing the number of clusters, we will start to divide the Corpus Callosum into different classes, before separating it from the spine.

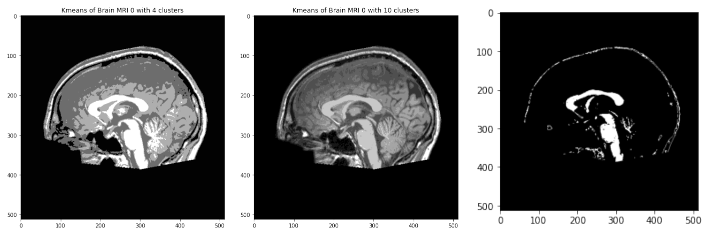

Figure 18: Image showing brain segmentation using kmeans for different number of clusters. The mask is the obtained using 10 clusters.

For the skin case, the Otsu method already gave really good results in most cases, so, it's expected that the k-means with 2 clusters will give similar good results. One case where the results were not so good, was the one with a lot of hair. Increasing the number of clusters allows to partially differentiate the target lesion from the hair, improving the results (0.65 score).

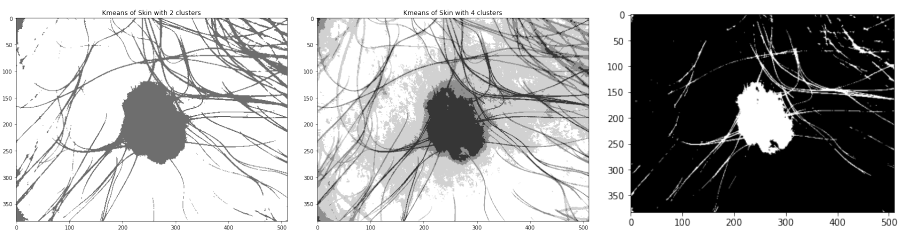

Figure 19: Image showing skin segmentation using kmeans for different number of clusters. The mask is the obtained using 4 clusters.

At last, for the heart MRI, increasing the number of clusters also improves the results, but as the kidney and brain cases, it still doesn't solve the problem of miss-detection.

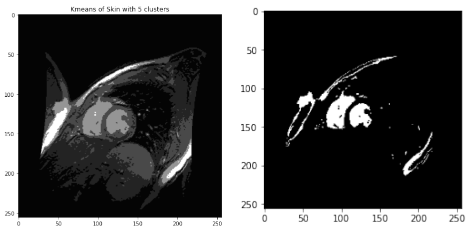

Figure 20: Image showing heart segmentation using kmeans for5 clusters and its mask.

Summarizing, using the k-means method generally improves the results for all methods, but it still doesn't solve the problem of miss-detection from different structures with similar gray-level value. The ideal number of clusters is not the same for all modalities.

For the kidney and heart case, it gives a better automatic result than Otsu, but it's similar for the one manualy predict. The hair problem of the skin was significantly improved.

To select the mask, we could observe that choosing a good number of clusters, the class that represent the mask tend to be the same (for example last less one for kidney, brain and heart and first for skin). Even though, to make sure the right class is selected, a coordinate can be given to select the class (mask will correspond to all points in the same class of the one given).

2 Region growing

2.1 Abdominal CT

It can be seen below that this method gave considerable better results for the kidney segmentation, starting with a point in the kidney. Using a tau equal to 60, we got a 0.81 score. Also, for the first time we were successfully able to detect the tumor, using a starting point inside the tumor and a tau equal to 20, we got a 0.79 score.

A remark is that using the previous obtained mask gave us worst results than the given mask. In the case where we use the mask as starting region, the expansion occurs correctly in the kidney, but also in the spine, making the numerical score worst. Using the region growing as pos-processing, that is, using the mask as image, also doesn't help, because the previously obtained mask only contained the kidney contour.

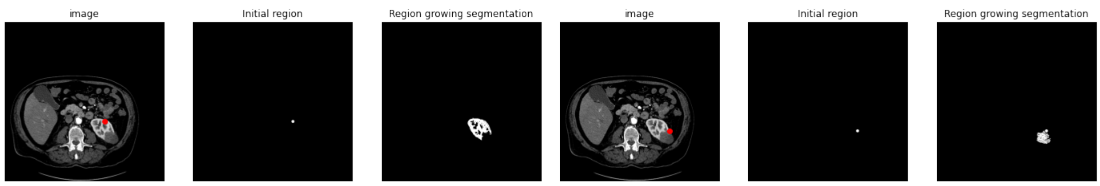

Figure 21: The first 3 images correspond to the initialization point in the kidney and the 3 last for an initialization in the tumor.

The results can be improved using a dilatation in the obtained segmentation, as it will fill the interior of the kidney and the tumor, to resemble more the intended mask. By doing that, we improve our dice score to 0.88 (using a disk with radius equal to 9 as element structurant) for the kidney, and 0.93 for the tumor (using a disk with radius equal to 1 as element structurant).

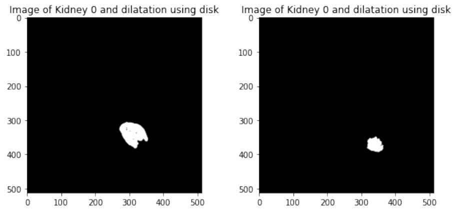

Figure 22: Dilatation result in the Region growing for kidney and tumor segmentation.

2.2 Brain MRI

Similar to the abdominal CT, this method gives much better results for the Corpus Callosum segmentation. Starting with a point inside it, and a tau equal to 15, we obtain a 0.88 score (we miss-detect a small light area close to it). Like the kidney case, using the mask obtained in k-means as initial doesn't give good results (similar reason), but, in contrast, using the mask as initial images help the results to be even better. Using also a tau equal to 15, the score is equal to 0.94. This happens because unlike for the kidney, the mask fully contains the Corpus Callosum.

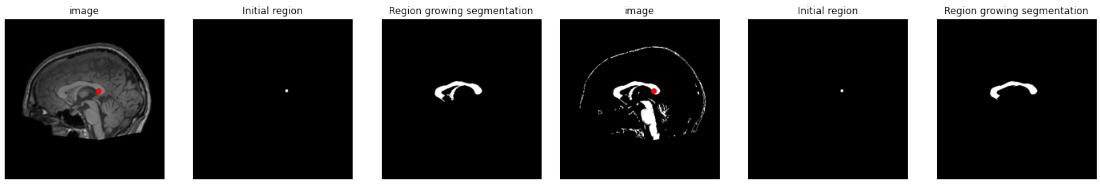

Figure 23: All the images correspond to the initialization point in the Corpus Callosum, the first 3 are results using the original image, the last 3 are using the previously obtained mask.

2.3 Skin lesion

For the skin lesion case with hair, this method also improves the results (although less drastically). If we use a smaller tau (for the example equals to 25), we manage to just detect the lesion, but it is smaller than the reference (the score is equal to 0.70). Increasing tau means we will segment a small portion of hair, but may numerically make the score better. Using the previously obtained mask gives similar results.

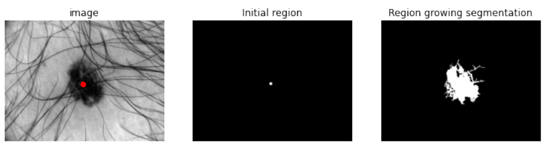

Figure 24: Images corresponding to initialization in skin lesion

2.4 Heart MRI

Similar to the previous cases, the region growing method also improves the results for the left ventricle. One remark, is that for this modality, a considerably larger tau was needed to give the result (tau=110). The results using the previously obtained masks also don't vary much, for the same reasons as described before.

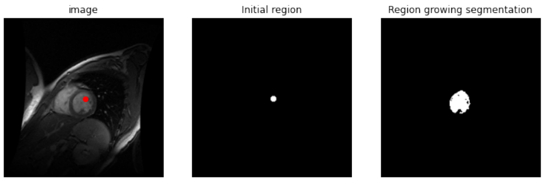

Figure 25: Images corresponding to initialization in left ventricle

3 Watershed segmentation

3.1 Abdominal CT

As can be seen below, the Watershed segmentation gives really good results both for the kidney segmentation and the tumor segmentation (when we have only one kidney). The score obtained is equal to 0.95 and 0.96 respectively. This results are specially due to the clean initial selection made over the binary segmentation image. This method can be considered more advantageous over the Region growing as we can get both segmentations in one run of the algorithm.

In the case when we have two kidneys, selecting both with just one initialization won't give good results, as it's difficult to have an initialization that includes both kidneys and excludes the spine. One possible solution is running the algorithm for the left kidney, then for the right. The final segmentation will be equal to the logical or between the two.

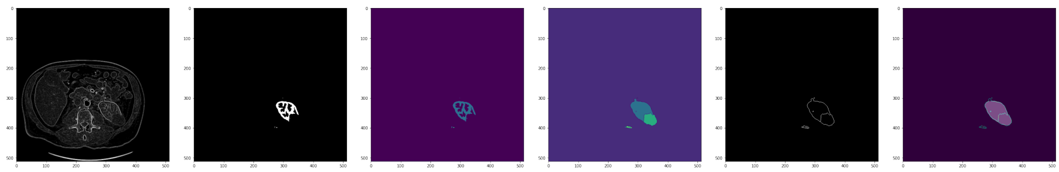

Figure 26: Images corresponding to Watershed segmentation process: first the difference between an dilatation and erosion to the image; the initialization; markers; labels; the resulting segmentation and at last the comparison between the result and the reference.

3.2 Brain MRI

As can be seen , this method gives an acceptable result for the Corpus Callosum (score equal to 0.73), specially if compared to one of the first methods, such as Otsu. But the results are inferior to the Region growing. Different from the kidney selection, the C form of the Corpus Callosum and the artifacts with close gray-level inside this C makes it difficult to obtain a initial selection that only contains the target area.

3.3 Skin lesion

The results for the skin lesion modality can be considered poor both visually and numerically (score equal to 0.55). Even only selecting the lesion as initialization, the hair still represents a big problem in the result of the method.

3.4 Heart MRI

At last, the results for the heart MRI can also be considered good, and similar to the Region growing method.

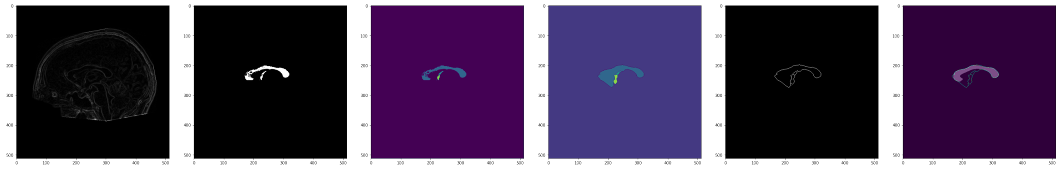

Figure 27: Images corresponding to Watershed segmentation process: first the difference between an dilatation and erosion to the image; the initialization; markers; labels; the resulting segmentation and at last the comparison between the result and the reference.

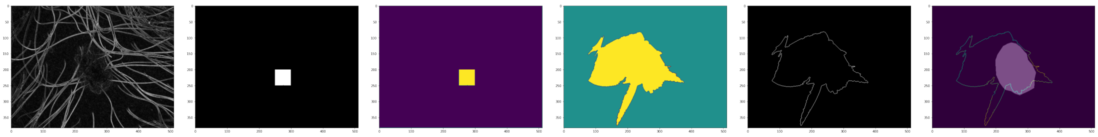

Figure 28: Images corresponding to Watershed segmentation process: first the difference between an dilatation and erosion to the image; the initialization; markers; labels; the resulting segmentation and at last the comparison between the result and the reference.

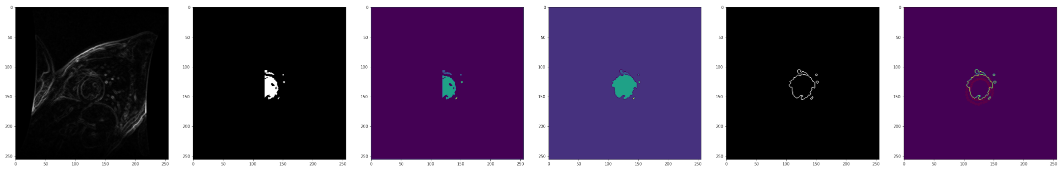

Figure 29: Images corresponding to Watershed segmentation process: first the difference between an dilatation and erosion to the image; the initialization; markers; labels; the resulting segmentation and at last the comparison between the result and the reference.

4 Proposed Algorithm

We would like to propose an algorithm to segment a sequence of images temporally correlated. In **??** it can be seen that the evolution in one time frame is very subtle. Comparing the first and last time frame examples, we can see that in fact the heart is contracting in this sequence.

To use the result of one time frame as initialization of the next, the Watershed segmentation is proposed, applying a closing to the result and using it as initialization. This way, we only need to worry to find a good initialization in the first time frame.

This technique could be used in different slices, given that they are close. Comparing different slices that are in 13, it can be seen that the shape of the left ventricle varies, while for different time frames it contracts/expands.

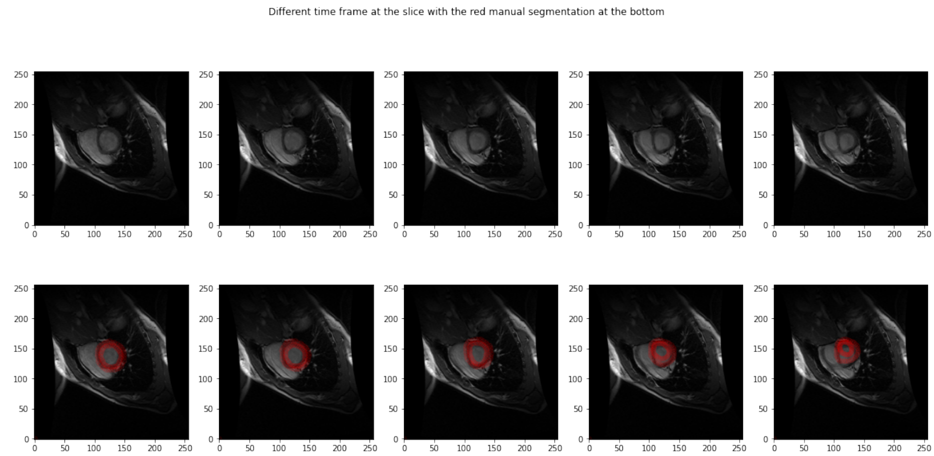

Figure 30: Image showing example of time evolution for a single slice (2) for the Heart MRI.

5 Conclusion

As it can be seen in the course of this TP, the methods give different results for the different modalities. Even for a same modality, there are special cases which also makes results vary a lot. There is no method or parameter that is universally good, and some times, the combination of methods might give the best results.

For the kidney CT for example, the Watershed segmentation gave the best results. For the brain MRI, the Region growing method. For the skin trauma, which is a simpler case, the Otsu and k-means gives good results, with a remark to the presence of hair, which can be solved using a mathematical morphology method or even Region growing in the post-processing. For the heart MRI, both Watershed segmentation and Region growing gives acceptable results, and we can also take advantage of the temporal correlation between images to get better results.