HL2027 3D image reconstruction and analysis in medicine

MiniProject2 - Segmentation and analysis of pelvic bone in CT images.

Julia Onerud, Kenza Bouzid
Group 3

May 31st, 2021

Abstract

Rapid advances in the field of medical imaging are revolutionizing medicine. Disease diagnosis and surgery planning is becoming more and more accurate thanks to computer-aided diagnosis (CAD) in which Medical Image Segmentation and Classification play a crucial role. In this report, we address the problem of segmentation and analysis of pelvic bone in CT images in order to facilitate hip surgery planning. More specifically, we apply atlas based segmentation to localize the right femur and right hip bone in 3 common images. Evaluation is performed by the means of Dice Coefficient and Hausdorff distance. We also design a deep learning solution in order to detect the particular axial slice of a CT image which is most likely to contain the pubic symphysis.

Code and demo notebooks are available at :
https://github.com/KenzaB27/segmentation-of-pelvic-bone.

1 Problem

In this project, we aim to solve two main problems of great importance in hip surgery planning. First we would like to segment the pelvic bone and the femur in CT scans. Then, we aim to detect the particular axial slice of a CT image which is most likely to contain the pubic symphysis. The pubic symphysis [11] is a secondary cartilaginous joint between the left and right superior rami of the pubis of the hip bones (e.g Fig. 1).

Figure 1: Pubic symphisis in an axial CT scan.

Given three 3d CT scans manually segmented by an expert radiologist denoted as D_c (for common data) and three 3d CT scans specific to our group denoted as D_g, we aim to create atlases of D_g that can further on be used for multi-atlas based segmentation of D_c using registration techniques.

We define 5 regions of interest (ROI) in the CT scans as shown in Fig. 2 :

- Right femur (in red).
- Left femur (in green).
- Right hip bone (in blue).
- Left hip bone (in yellow).
- Sacrum (in cyan).

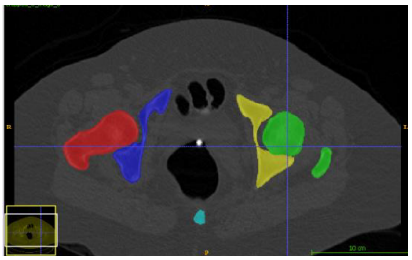

Figure 2: Axial Slice of a CT Scan with all ROIs.

As a first milestone, we perform segmentation of the right femur and right hip bone only. The left femur and left hip bone are segmented as bonus. D_g is used as training set for pubic symphysis detection. Testing is performed on D_c.

2 Methods

2.1 Multi-Atlas Based Segmentation

Segmentation in medical imaging is really challenging due to problems linked to low contrast images, fuzzy object-contours, similar intensities with adjacent objects of interest, etc. On way to encounter these problems is by the use of prior knowledge [1]. A widely used method consists of extracting this prior knowledge from a reference image so called atlas. During segmentation, atlases and corresponding label maps are warped to the target image space using image *registration*[6].

Atlases of D_g were created by manual/partially automatic segmentation using ITK-Snap [12]. A mix of threshold based segmentation and edge-finding segmentation was used to first find basic label maps that we manually correct in order to get atlases of high precision. The resulting 3 group atlases are used for multi-atlas segmentation of the common images. 3 simple segmentations of the common images were also made using only automatic threshold and edge based methods in ITK-Snap with no manual corrections.

Let, $i, j \in \{1, 2, 3\} \times \{1, 2, 3\}$. We define x_i a CT image from D_c and (x_j,m_j) an atlas from D_g. We first use a *Linear Affine Transformation*[3] to register every image x_j to x_i and obtain $\mathcal{F}_{aff}$ the affine transformation. We apply $\mathcal{F}_{aff}$ to x_j, m_j respectively and obtain lx_j, lm_j. Now we register lx_j to x_i using *Free-Form Deformations* [8] and obtain $\mathcal{F}_{ffd}$ that we apply to lm_j to obtain the final registration mask rm_j. The final segmentation label map is obtained by applying majority voting among labels predicted by each rm_j.

For reference a *Linear Affine Transformation* [3] is a linear registration method that generalizes rigid-body (rotation+translation) transformations by including scaling (3 DOF) and shears (3 DOF) which results in (12 DOF) in 3D. On the other hand, *Free-Form Deformations*[8] is a non linear parametric registration method that uses priors. It relies on moving the control points in a grid instead of moving the image, and then interpolate with B-splines.

The ground truth segmentation masks serve as reference to evaluate the quality of our segmentation quantified by the means of Dice Coefficient and Hausdorff distance.

Motivation: The elected methods correspond perfectly to our problematic. Affine Linear Transformation along with Free Form Deformations are both parametric registration techniques that were easy to integrate in our segmentation pipeline with a reasonable computational cost. Furthermore, they gave very satisfactory results as will be shown in the experiments section.

2.2 Image Classification

In order to detect the particular axial slice of a CT image which is most likely to contain the pubic symphysis, we designed a *deep learning* based solution. We exploit the power of *Convolutional Neural Network* (so called CNN) and *Transfer Learning* in extracting high level feature maps to discriminate with high accuracy between negative and positive classes.

Deep Learning is considered as a subset of machine learning methods based on artificial neural networks which are inspired by the function and structure of the brain. Among all kind of neural networks, we utilize CNNs. *Convolutional Neural Networks* were first introduced by Yann LeCun in 1998 [7]. The architecture is similar to the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. The objective of the Convolution Operation is to extract the high-level features such as edges, from the input image. A CNN is able to successfully capture the Spatial and Temporal dependencies in an image through the application of relevant filters which is very useful for our use case. A more powerful technique is Transfer Leaning. *Transfer Learning* [13] improves the performance of target learners on target domains by transferring the knowledge contained in different but related source domains. The idea is to fine-tune previously trained neural networks for a specific task instead of training from scratch.

Motivation: We chose a deep learning based method over a simple machine learning solution as neural networks can be trained in an end to end fashion and take care of feature engineering and extraction through convolutions. In fact a NN learns an explainable high level representation of the problem which can be exploited for further analysis of the most important features in detecting the pubic symphysis. Furthermore, our experiments showed that machine learning based methods performed poorly in comparison to a neural network.

3 Experiments & Findings

3.1 Multi-Atlas Based Segmentation of pelvic bone

3.1.1 Experimental Settings:

We use the SimpleITK software [2] in order to implement multi-atlas segmentation of the CT images based on registration techniques. This segmentation in two steps. First, with the use of ITK Snap [12] an automatic thresholding between intensities 114 and maximum within a square area around the femur and hip bone. Then manual correction of the segmentations by separation and addition of parts that were missed. These segmentations were then used to perform the next steps multi-atlas based segmentation of the regions of interest.

Affine Linear Registration was performed by optimizing Mutual Information (MI) between the reference image from D_c and a moving image from D_g. Whereas we optimize Sum Squared Distance (SSD) for Free Form Deformations. We use gradient descent optimizer to solve both registration problems with a gradient descent step of 1 and minimum convergence value of 10^{-6} for 200 iterations in the case of affine transformation and 20 iterations for FFD.
In order to make registration better and faster, a mask covering the non anatomy parts was used during registration. Since our main interest is making sure the segmentations from D_g fit correctly onto the new image and not the entire image from D_c.

3.1.2 Results & Findings:

We report the following results for all regions of interest jointly.

- **Common Image 40**

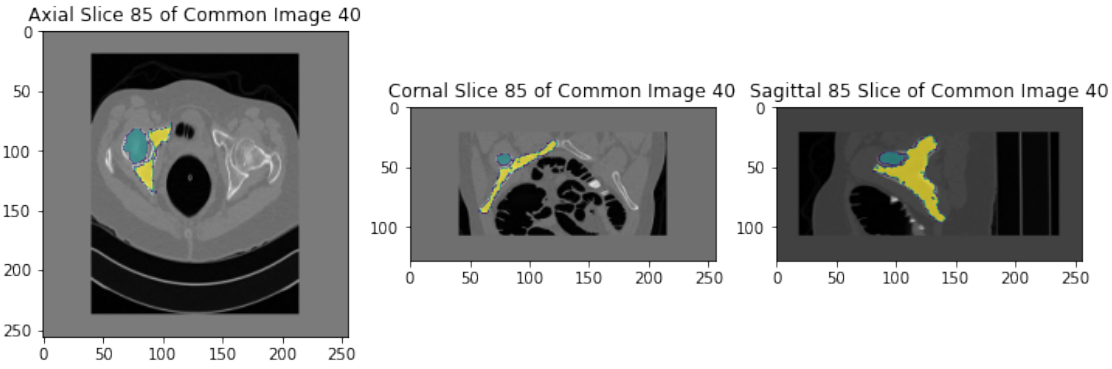

Figure 3: Atlas Segmentation of Right pelvic bone and femur Image 40.

Evaluation Metric	Right Hip Bone	Right Femur	Left Hip Bone	Left Femur
Dice Score	73.46	79.00	78.38	78.31
Hausdorff Distance	18.05	7.48	17.03	7.0

Table 1: Atlas Segmentation Evaluation of Image 40.

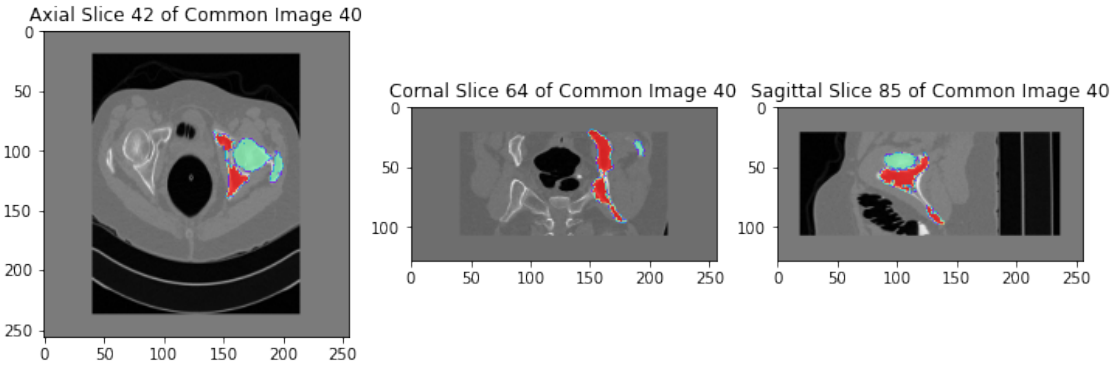

Figure 4: Atlas Segmentation of Left pelvic bone and femur Image 40.

- **Common Image 41**

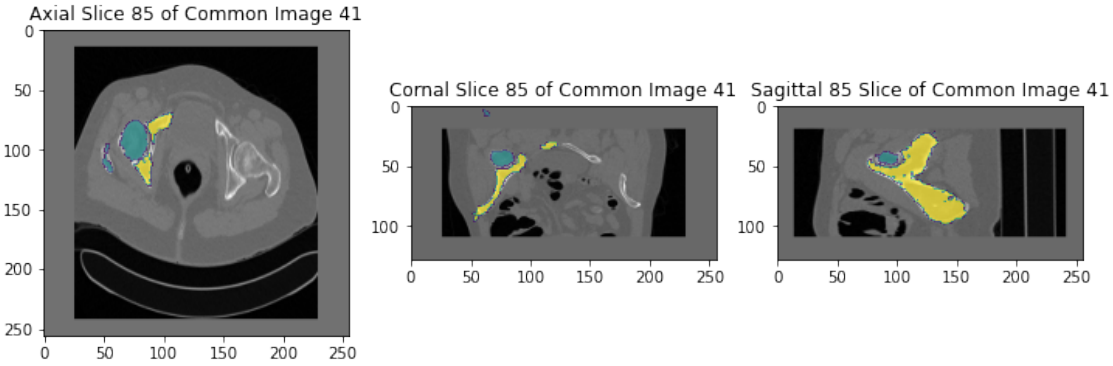

Figure 5: Atlas Segmentation of Right pelvic bone and femur Image 41.

Evaluation Metric	Right Hip Bone	Right Femur	Left Hip Bone	Left Femur
Dice Score	79.46	78.56	83.88	82.49
Hausdorff Distance	20.12	8.66	11.22	6.32

Table 2: Atlas Segmentation Evaluation of Image 41.

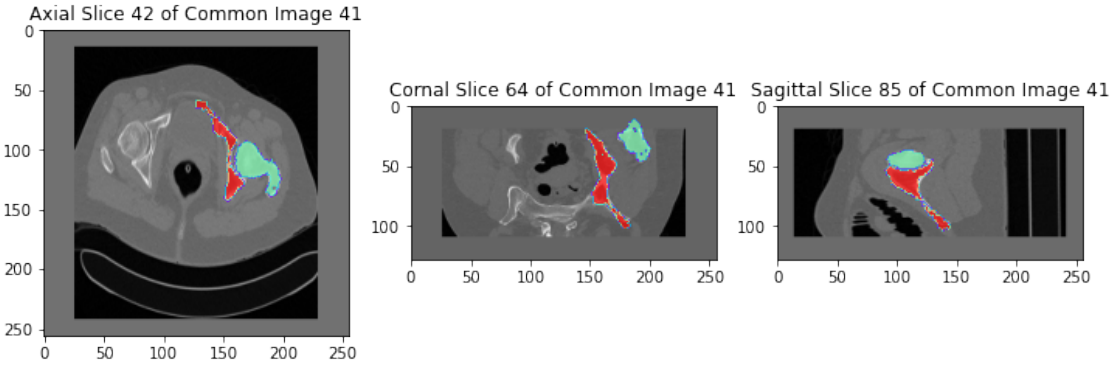

Figure 6: Atlas Segmentation of Left pelvic bone and femur Image 41.

- **Common Image 42**

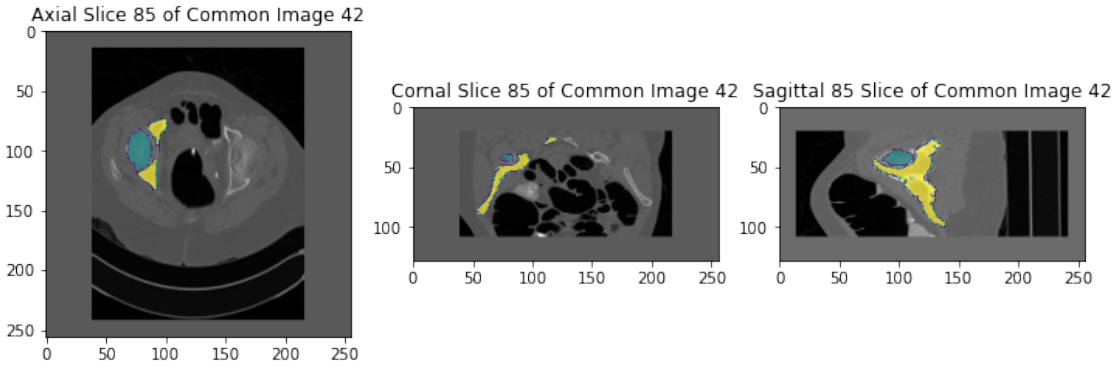

Figure 7: Atlas Segmentation of Right pelvic bone and femur Image 42.

Evaluation Metric	Right Hip Bone	Right Femur	Left Hip Bone	Left Femur
Dice Score	74.68	82.13	75.87	81.78
Hausdorff Distance	14.87	5.19	15.0	6.78

Table 3: Atlas Segmentation Evaluation of Image 42.

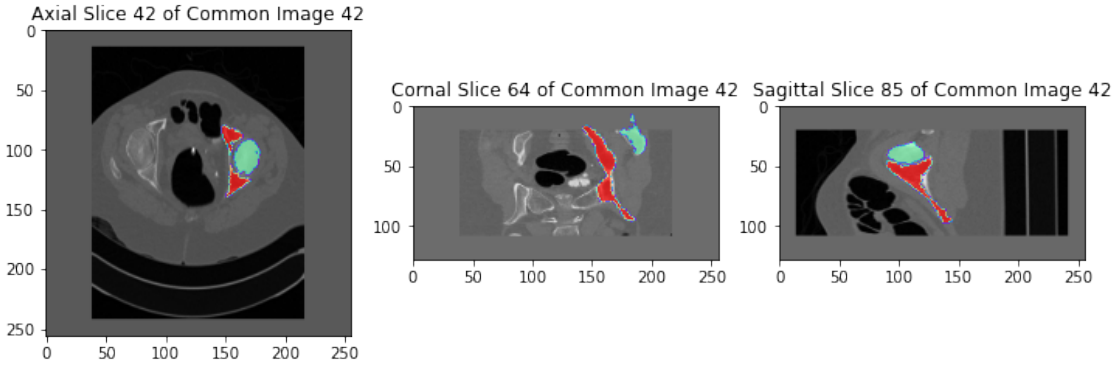

Figure 8: Atlas Segmentation of Left pelvic bone and femur Image 42.

Comparison to automatic segmentation: The automatic segmentations of the common-images resulted in segmentations that were unsatisfactory in many regards. The automatic segmentations could not separate the hip bone and femur as two segments and also contained parts of the sacrum if it was not manually blocked. The resulting segmentations were also missing large parts of the femur and hip bone, both inside the bones and on some edges. The Dice score and Hausdorff Distance for these segmentations were abyssmal, and so they were omitted from this report.

3.2 Detection of pubic symphysis

3.2.1 Experimental settings

We experimented with both neural networks and machine learning algorithms such as Support Vector Machines, Random Forests and Nearest Neighbors. Our benchmark study showed that neural networks perform way better in detecting the pubic symphysis in CT scans. We therefore detail only the experiments with neural network implemented in Tensorflow.

- **Data Annotation:** Data Labelling was performed manually by the use of ITK Snap [12]. We went through all slices of all images from D_g and check if the pubic symphysis is visible. In general, the pubic symphysis was present in the first third of the CT Slices. We therefore expect this region to have higher probability in test time.

Case Id	Number of Slices	First Appearance of PS	Last Appearance of PS
59	275	63	89
60	541	52	84
61	339	97	138

Table 4: Pubis Symphysis Detection in Group CT images.

- **Data Preprocessing:** As we decided to use a neural network based solution, the only processing we performed was standardizing the images to 0 mean and unit variance in order to avoid exploding gradients effect with ReLu activation function. We cropped images to 300x300 widows in order to avoid the use of misleading/irrelevant information to discriminate between positive and negative classes. We use 90% of group slices for training and 10% for validation. The common images are used for testing.

- **Network Design:** We conducted multiple experiments in order to define the best architecture for our task. First we experimented with different pretrained models: ResNet[4], Inception [10], VGG[9], and DenseNet[5]. VGG16 and DenseNet121 perfomed best in extracting relevant features for the classification task. Next, we experimented with number of hidden layers and number of nodes for the dense layers following our pretrained model.
The best architecture is composed of a pretrained model DenseNet121 on Imagenet dataset followed by a dense layer of 200 nodes with ReLu activation and 0.4 Dropout. Finally an output dense layer with a single node and sigmoid activation as we would like to get the probability of each CT slice containing the pubic symphysis. The loss function to minimize is binary cross entropy. Adam optimizer was used with 10^{-4} learning rate and 10^{-3} decay. We train the network for 30 epochs with batches of size 32.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_5 (InputLayer)         [(None, None, None, 3)]   0
_________________________________________________________________
densenet121 (Functional)     (None, 1024)              7037504
_________________________________________________________________
dense_6 (Dense)              (None, 200)               205000
_________________________________________________________________
dropout_4 (Dropout)          (None, 200)               0
_________________________________________________________________
dense_7 (Dense)              (None, 1)                 201
=================================================================
Total params: 7,242,705
Trainable params: 7,159,057
Non-trainable params: 83,648
_________________________________________________________________
```

Figure 9: Best Network Architecture.

3.2.2 Results & Findings

We hereby report the best test results with DenseNet and VGG. Note: the highest probabilities are around 75% for denseNet. Higher probabilities were obtained (See Appendix) but we chose to stick with this network as it was able to generalize well on all 3 common images in contrary to the other networks where common image 40 were misclassfied (same for reported VGG).

- **DenseNet121 Results:**

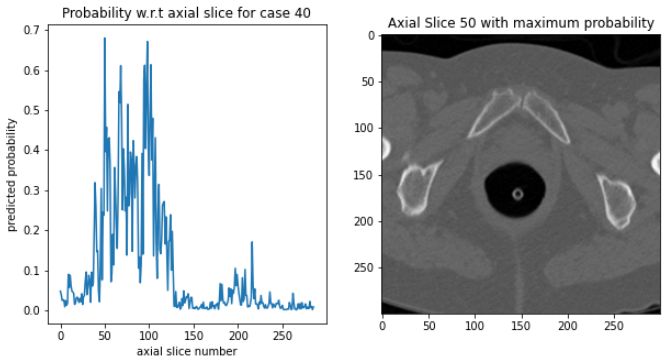

Figure 10: Probabilities of containing the pubic symphysis for Common Image 40.

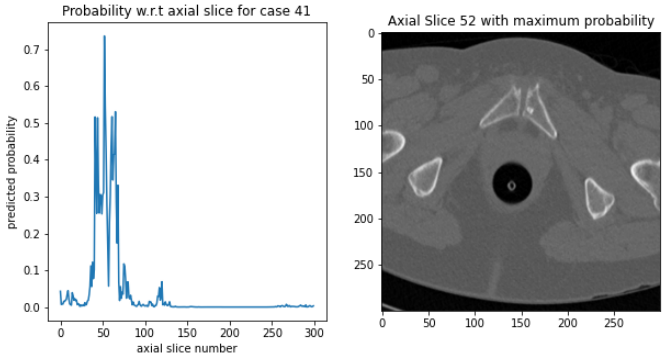

Figure 11: Probabilities of containing the pubic symphysis for Common Image 41.

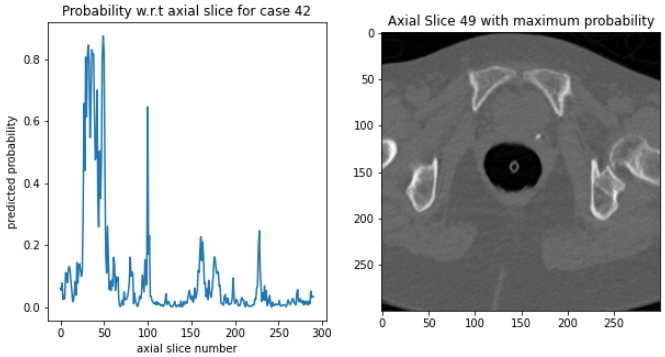

Figure 12: Probabilities of containing the pubic symphysis for Common Image 42.

- **VGG16 Results:** We obtained higher precision and certainty with VGG. However, the network was unable to generalize for common image 40. This is probably due to the difference in intensities in each case and the restricted training set at our disposal.

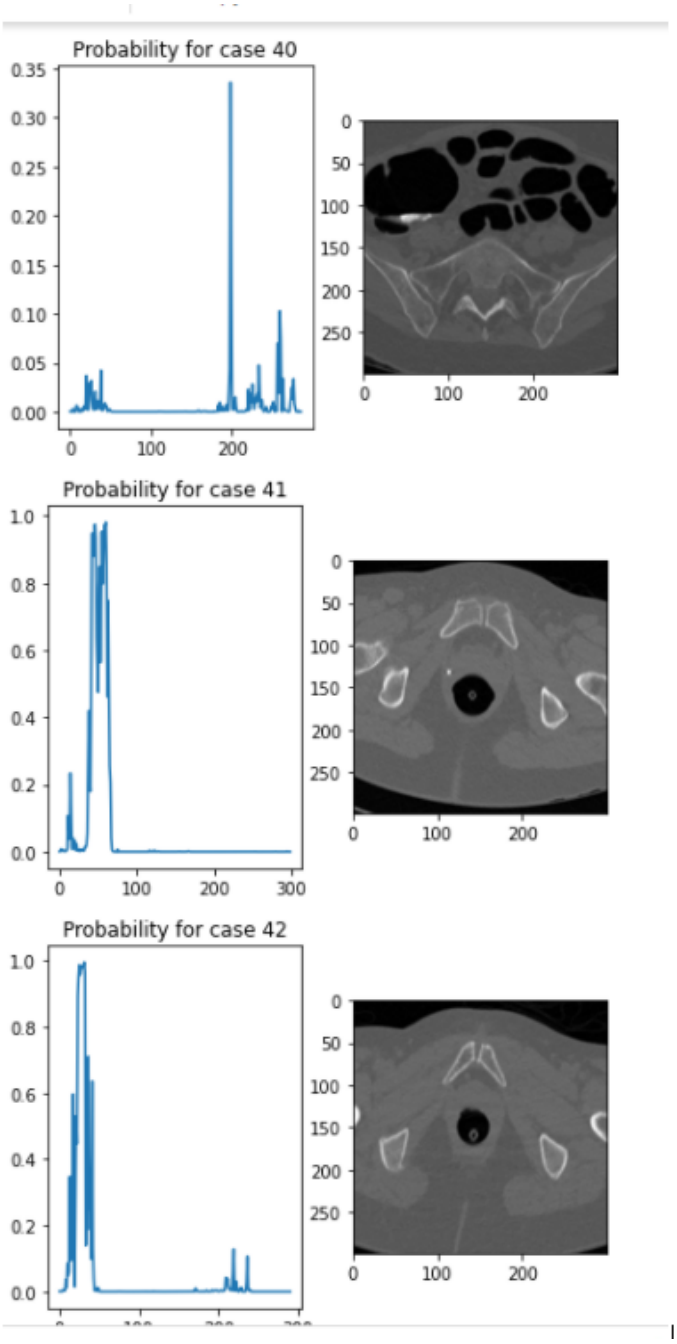

Figure 13: Probabilities of containing the pubic symphysis for Common Image 40, 41, 42.

4 Strategies to solve unexpected problems/difficulties

4.1 Atlas Segmentation

Before creating the atlas for each of the 3 D_c images using our segmentations. We found that the images and masks had a varying sizes, pixel spacing and origins. This causes a lot of issues with atlas creations and registrations. So to fix it, we first created a reference domain based on all of

the images different parameters. Then all of the images and their respective segmentation masks were resampled onto this reference domain. Some trimming of the images was also done to make the resulting resampled images better.

The resampling of the images also had another beneficial effect. In that it allowed for a reduction in the sizes of the images. The smaller size made the registration much faster, which made debugging and testing much quicker. Without the resampling it would have been much more difficult to finish the project. Since it took up to half an hour for each atlas segmentation creation even with the resampling, and up to one or two hours without it.

4.2 Image Classification

Regarding the detection of the pubic symphysis in CT slices, the main difficulties that we encountered were linked to generalization performance for common image 40. We solved this problem by fine designing our network and trying different pre-trained models, different learning schedules and optimizers (SGD, Adam, RmsProp, cyclical learning, ...) and by varying the number of hidden nodes until we found the perfect trade of between overfitting and under fitting. We also experimented with different degrees of regularization with Dropout and Batch Normalization to stabilize the training.

5 Discussion & Conclusion

Segmentation of the pelvic bone using fully automatic methods, while fast and easy to do, did not get good results. The resulting segmentations were not useful without a lot of further manual corrections. That could take a long time to do. Making the total time to get a good segmentation still quite long. Atlas segmentations on the other hand was difficult to get working correctly. But the results were good and did not require manual correction apart from the creation of the first segmentations. And when it was done so that it worked for one image, for just the right femur and hip bone. It was quick and easy to do the same procedure for other bones such as the left femur and hip bone. And it would probably have been similarly quick and easy to do segmenting of bones on many more new pictures.

Atlas segmentation is not worth it in terms of speed and ease when it is just a few images that need segmenting. But once the work has been done to enable atlas segmentation, it can be scaled up quite well to more images and it can work for segmenting of other bones of similar type. While other types of automatic or semiautomatic segmentation such as thresholding and edge detection would not scale up as well.

Transfer Learning is very powerful in extracting high level features however given the restricted amount of data at out disposal, it was pretty hard to get a working solution that generalizes well on all common images. It would have been nice to have more training data then just three 3d CT scan (a total of 878 2D slices). Nevertheless, we were able to get good results with design engineering and fine tuning the model. Machine Learning algorithms on the other hand require more focus on feature extraction and engineering whereas deep learning offer an end to end solution.

References

[1] M. Bach Cuadra, V. Duay, and J.-Ph. Thiran. *Atlas-based Segmentation*, pages 221–244. Springer US, Boston, MA, 2015.

[2] Richard Beare, Bradley Lowekamp, and Ziv Yaniv. Image segmentation, registration and characterization in r with simpleitk. *Journal of Statistical Software, Articles*, 86(8):1–35, 2018.

[3] Rami Hagege and Joseph Francos. Parametric estimation of affine transformations: An exact linear solution. *Journal of Mathematical Imaging and Vision*, 37:1–16, 05 2010.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[5] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks, 2018.

[6] Shweta Jain and Navdeep Kanwal. Overview on image registration. In *2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom)*, pages 376–381, 2014.

[7] Yann Lecun, Leon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. In *Proceedings of the IEEE*, pages 2278–2324, 1998.

[8] Thomas Sederberg and Scott Parry. Free-form deformation of solid geometric models. volume 20, pages 151–160, 08 1986.

[9] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

[10] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, 2014.

[11] Wikipedia contributors. Pubic symphysis — Wikipedia, the free encyclopedia, 2020. [Online; accessed 30-May-2021].

[12] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, 2006.

[13] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning, 2020.

Appendices

A DenseNet121 and 64 nodes for hidden layer.

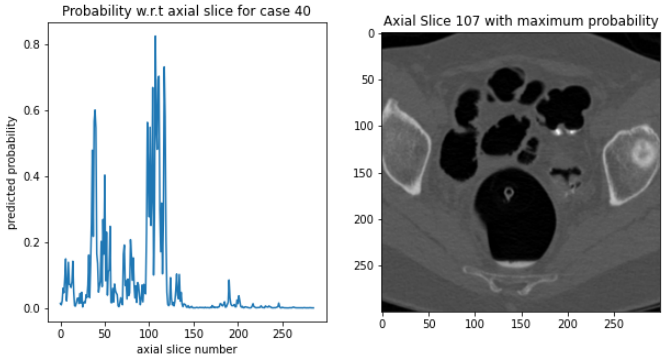

Figure 14: Probabilities of containing the pubic symphysis for Common Image 40.

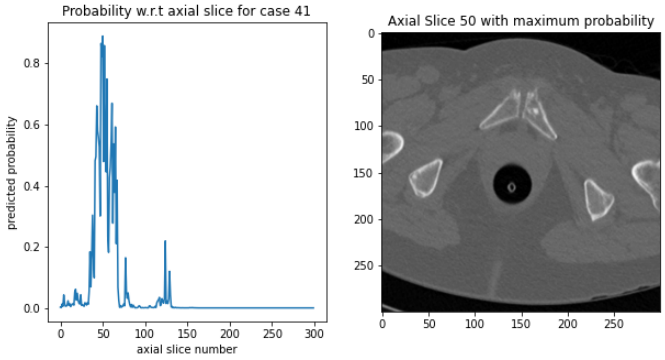

Figure 15: Probabilities of containing the pubic symphysis for Common Image 41.

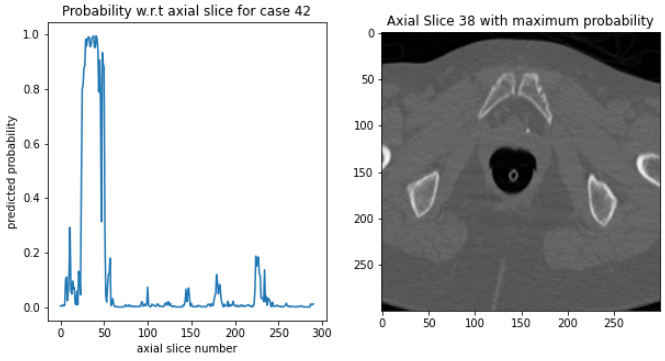

Figure 16: Probabilities of containing the pubic symphysis for Common Image 42.

B DenseNet121 and 128 nodes for hidden layer.

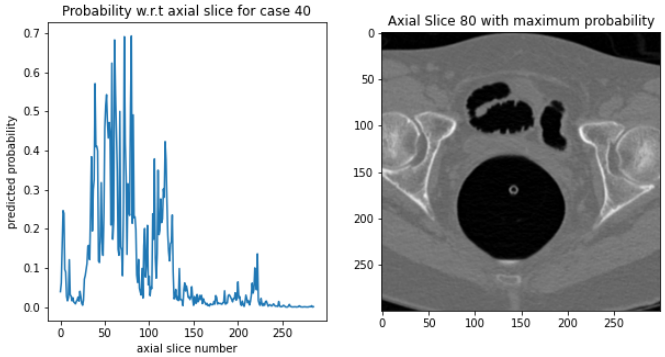

Figure 17: Probabilities of containing the pubic symphysis for Common Image 40.

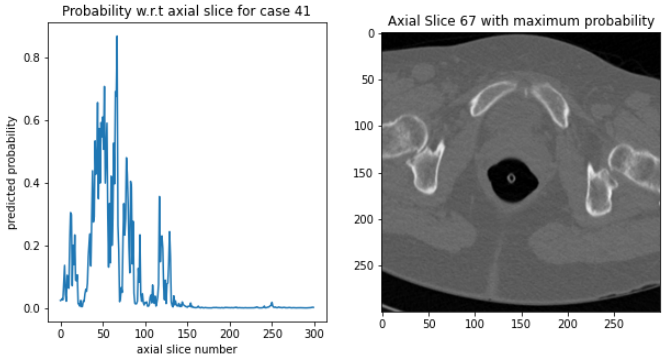

Figure 18: Probabilities of containing the pubic symphysis for Common Image 41.

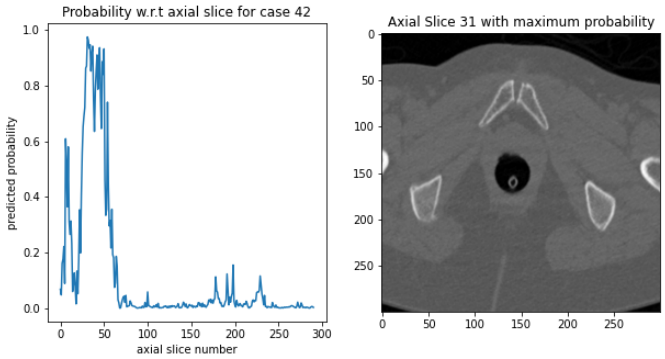

Figure 19: Probabilities of containing the pubic symphysis for Common Image 42.

C DenseNet121 and 512 nodes for hidden layer.

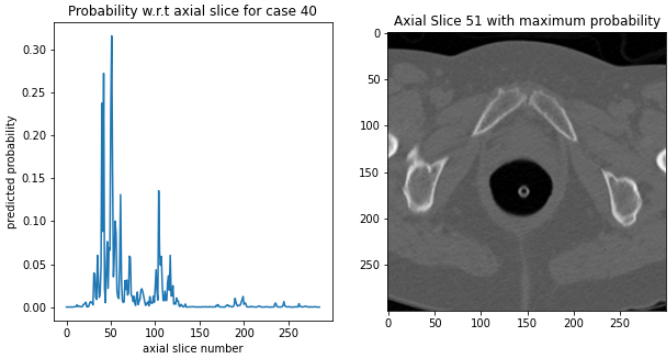

Figure 20: Probabilities of containing the pubic symphysis for Common Image 40.

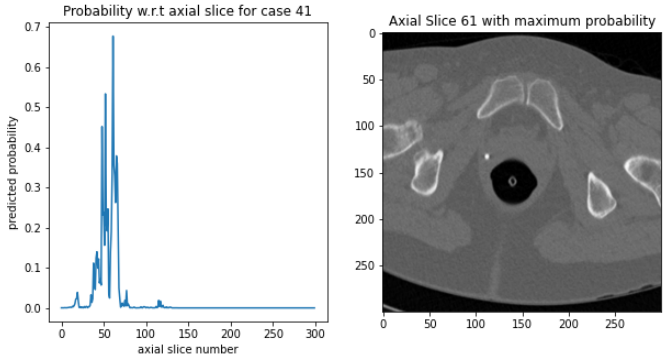

Figure 21: Probabilities of containing the pubic symphysis for Common Image 41.

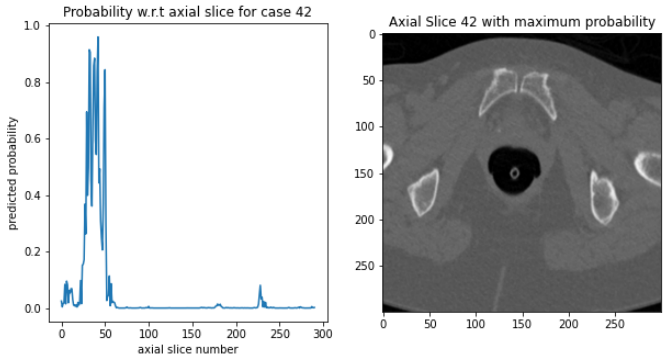

Figure 22: Probabilities of containing the pubic symphysis for Common Image 42.

D DenseNet121 and 1024 nodes for hidden layer and Batch Normalization.

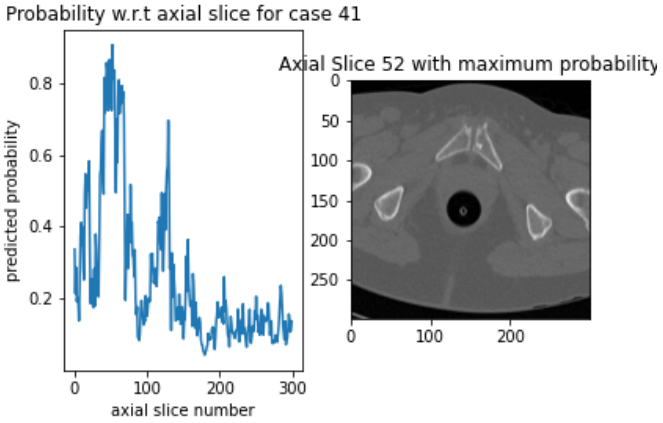

Figure 23: Probabilities of containing the pubic symphysis for Common Image 40.

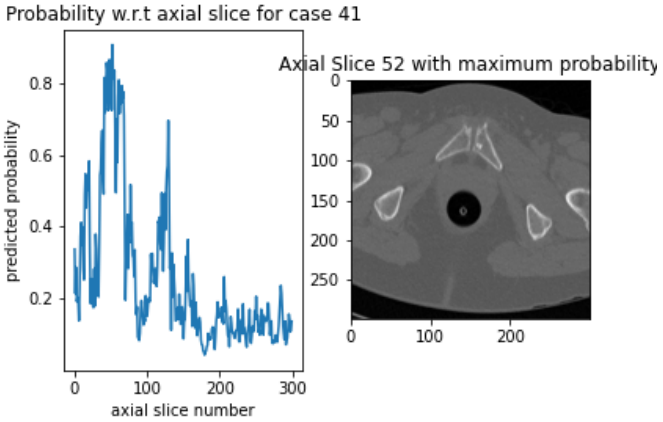

Figure 24: Probabilities of containing the pubic symphysis for Common Image 41.

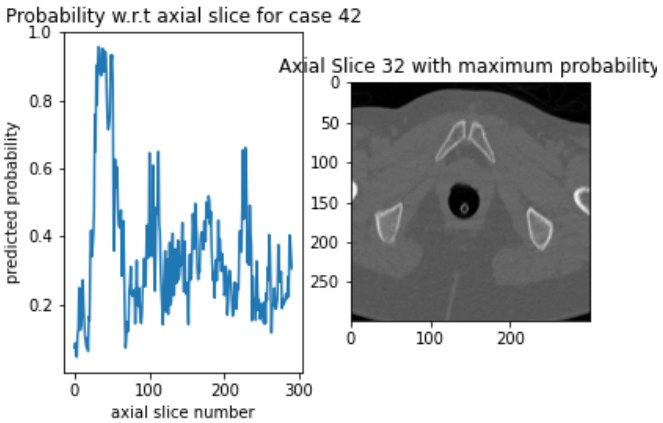

Figure 25: Probabilities of containing the pubic symphysis for Common Image 42.