Healthy Abdominal Organ Segmentation

Francesco Caserta
Biomedical Computer Vision -Passion in Action, 2023
Politecnico di Milano, Italy

1 INTRODUCTION

The aim of this project was to perform image segmentation of thoracic sections obtained through magnetic resonance imaging, accurately identifying and segmenting the liver, kidneys, spleen, and background. This report presents the various steps that led to the completion of this task, starting from the dataset analysis, through its division, to the construction of a neural network capable of segmenting the different regions. Finally, the report provides insights into the testing phase conducted on the designated test set.

2 PREPROCESSING AND DATASET PREPARATION

2.1 DATASET USED

The dataset employed for this phase is the same as the one used for the CHAOS 2019-Combined (CT-MR) Healthy Abdominal Organ Segmentation challenge. The data, available at this link, comprises both CT and magnetic resonance imaging (MRI) volumes from 20 different patients, along with their respective Ground Truth annotations. The MRI data sets was acquired with two different sequences (T1-DUAL and T2-SPIR).

Specifically for this project, only the T1-DUAL MRI volumes within the following folder were utilized:

T1DUAL/DICOM_anon/InPhase/. The corresponding Ground Truth labels from "T1DUAL/Ground/" were employed. Essential properties for preprocessing, namely Slice Spacing and Pixel Spacing in the horizontal and vertical dimensions, were extracted alongside the slices of each volume in .dcm format.

2.2 PREPROCESSING

The obtained volumes had varying and incompatible dimensions (e.g., (50,256,256), (30,288,288)). To prepare them for neural network training, they were initially resized to a uniform size of 1x1x1 through resampling. Subsequently, resizing was performed to bring all volumes to a consistent shape of 192x192x192, using a cubic interpolator for actual data and a zero interpolator for the masks (ground truth).

Following that, the data was normalized between 0 and 1 (only the data, not the masks) using the OpenCV library.

Subsequently, all volumes were concatenated into a single numpy array with a shape of (20*192,192,192,1). The last dimension was added for compatibility with the U-net architecture used later. This operation was repeated for the masks (labels), with an additional step involving mapping to restrict labels to the range from 0 to 4. In the original masks, labels ranged from 0 to 255. The mapping was applied based on guidelines provided in the CHAOS dataset.

Examining the label distribution across the dataset reveals a noticeable imbalance. The 'Background' class is significantly larger than the other classes (135 million), followed by 'Liver' (4.6 million). 'Right kidney' and 'Left kidney' each have approximately half a million instances, while 'Spleen' has 800 thousand.

The labels 0 to 4 correspond to organs within the thoracic region:

- 0: Background
- 1: Liver
- 2: Right kidney
- 3: Left kidney
- 4: Spleen

Given this pronounced imbalance between the background class and the others, the final performance of the model was assessed by calculating the Intersection over Union (IoU) for each class. Otherwise, the background class would have significantly inflated the accuracy, making it a less descriptive metric of the network's performance.

2.3 TRAIN, VALIDATION, TEST

Two methods were implemented to split the dataset and their respective masks into training, validation, and test sets. Initially, the array containing all slices (3840,192,192,1) was randomly divided, thus mixing

some slices from the same patient across the training, validation, and test sets.

However, this approach could introduce bias during the testing phase, as the network may have already encountered some slices from the same patient during training. In the final solution, the data was split based on patients, using all slices from the first 14 patients for the training set, 3 patients for the validation set, and 3 for the test set.

Dataset	1st dim	2nd dim	3rd dim
X_{train}	2688	192	192
y_{train}	2688	192	192
X_{val}	576	192	192
y_{val}	576	192	192
X_{test}	576	192	192
y_{test}	576	192	192

Table 1: Dataset shape

3 MODEL ARCHITECTURES

The architecture of the neural network used is a classic U-Net. In both the encoder and decoder, U-Net blocks were employed, consisting of a 2D convolutional layer (with a variable number of filters), followed by a Batch Normalization layer and an activation layer using the ReLU function. In the encoder, each block is followed by a MaxPooling2D layer, while in the decoder, the blocks are preceded by UpSampling2D layers. Additionally, there are skip connections between corresponding blocks of the encoder and decoder (refer to the complete architecture code on GitHub).

During training, a learning rate of 1e-3, a batch size of 16, and a maximum number of epochs set to 1000 were used. Early stopping was configured with a patience of 5 epochs and automatic reduction of the learning rate if it reached a plateau, with a reduction factor of 0.1, down to a minimum of 1e-5.

In the notebook on github you can see the training history for Cross Entropy, Accuracy, and Mean Intersection over Union. It is observed that the Mean Intersection over Union reaches almost 80% on the validation set.

4 RESULTS

Initially, the model was tested using overall accuracy and the Mean Intersection over Union applied to the entire test set. However, as mentioned earlier to gain a better perspective on the actual performance of the model, given the class imbalance where the background prevails, the Intersection over Union was calculated for each individual class. The results obtained are presented in the Table1. Upon observing the predicted mask images generated by the model, it is apparent that, for the majority of the data, the model can effectively distinguish between various organs and segment them with reasonably good precision. However, the segmentation borders are not always extremely accurate, and in the initial and final slices of the volume (where the background prevails), it may not consistently segment the various organs as effectively.

Table 2: Model Performance

Organs	**IoU**
All	0.744
Background	0.9762
Liver	0.8064
Right kidney	0.6205
Left kidney	0.7818
Spleen	0.3677

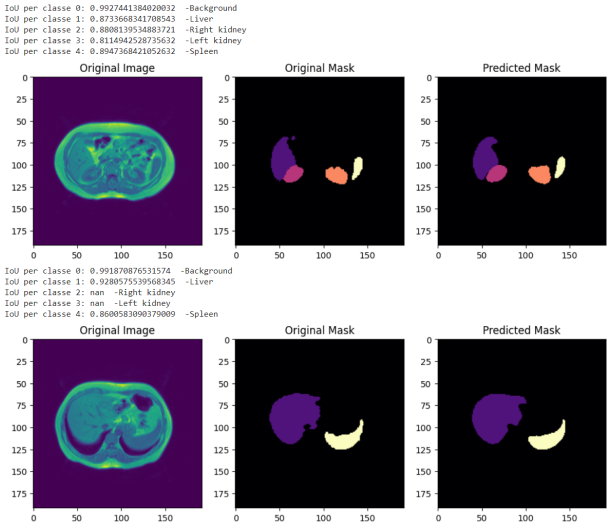

Figure 1: Examples of segmentation on Test Set