Automatic Detection of Anatomical Regions in
Three-Dimensional Medical Images

Thesis booklet

by

Márton József Tóth

Department of Control Engineering and Information Technology
Budapest University of Technology and Economics

Supervisor

Balázs Csébfalvi, Ph.D.

Industrial Supervisor

László Ruskó, Ph.D.

2020
Budapest
1. Introduction

In modern clinical practice three-dimensional medical imaging has an important role. Different imaging modalities like Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) are routinely applied in diagnostics, radiotherapy planning, tumor monitoring and in many other clinical workflows. By the analysis and evaluation of the collected data, radiologists and other therapists can suggest special treatments that can support the recovery of patients.

In the recent years, the number of the acquired images has been increasing rapidly [1]. More and more images are captured and even more resources are used to evaluate these images. To support the work of the radiologists several automatic image processing methods have been developed. These automatic tools are used in various clinical workflows and are usually specialized to a specific region of interest (ROI). There are different applications for lung nodule detection, virtual colonoscopy, cardiac or cerebral vessel analysis or even a simple visualization tool can have different settings for different anatomical regions.

The success of these automatic tools depends on the correct identification of the ROI they can work on. The DICOM standard involves tags to specify the anatomy location of a medical image, but the corresponding field is rarely filled in by the scanner and the manual specification is time consuming and error prone. According to Güld [2] 15% of the images are affected by an incomplete or even wrong indication of the represented anatomy location.

It is easy to see that an automated pre-processing function cannot rely on the DICOM tag only as it may lead to an incomplete processing of the medical image. Such errors can lead to a time-consuming error-correction process that requires serious efforts from the therapist.

It would be advantageous to define a method that can reliably recognize the anatomical regions in medical images based only on the image content itself. It could prevent several errors in the automatic pre-processing steps and therefore would save significant time for the therapists. Beside this benefit, it would also provide several potential automation or optimization possibilities in numerous clinical applications.

The usage of such a method would be transparent for the therapists and therefore the need for it does not directly comes from the users, but its benefits would be definitely welcome in medical image processing.
In this thesis work, a method is presented to classify the axial slices of three-dimensional medical images according to the anatomical region they belong to. The proposed method has two variations. The first one uses traditional image processing, feature extraction and machine learning tools to provide an initial labeling. The second version replaces these steps and uses a convolutional neural network for this task. As the initial labeling, in both cases, is unreliable and contains major classification errors, a post-processing is applied to ensure a correct and continuous labeling.

2. The Data Set

In this thesis work, two data sets were used to evaluate the proposed methods, an MRI and a CT data set. The MRI data set was used to develop the first version of the method and later, when it became clear that the proposed workflow could handle the recognition task reliably, a more detailed CT data set was assembled. In this way the method could be challenged with more regions to recognize and, as the data were collected in clinical practice, a more detailed overview could be obtained about its usability in real-world cases.

The first data set contained 49 full body MRI cases. It was acquired for research purposes by GE Healthcare in the University Hospital Zurich. The acquisitions were performed with a Discovery MR750w MRI scanner using the same protocol. This MRI data set was manually labeled. The axial slices were annotated as LEG, PELVIS, ABDOMEN, CHEST or HEAD&NECK.

The second data set contains CT examinations. They were originally collected to evaluate various types of clinical applications in the area of oncology, cardiology, surgery, neurology. The cases involve whole-body as well as partial scans. The images were acquired with whole (50-70 cm) as well as small (20-35 cm) reconstruction diameter. The dataset shows significant variance in patient's sex, age, and level of obesity. It includes native and contrast-enhanced cases, and exams demonstrating various types of abnormality (pathology, implant, noise). Thus, it is a representative set of 480 CT images. In this data set the axial slices were divided into eleven different regions. The definition of the anatomy regions were driven by the needs of organ segmentation algorithms and considered the following regions: FOOT, SHIN, THIGH, PELVIS-LOWER, PELVIS-CENTER, PELVIS-UPPER, ABDOMEN-LOWER. ABDOMEN-UPPER, CHEST, HEAD and BRAIN.
3. Method Overview

The presented work offers a method to assign labels to the axial slices of three-dimensional medical images according to the body part they belong to. The method has three main parts. The first one is an image pre-processing and normalization step, the second part provides an initial classification while the third step filters and finalizes the results.

The workflow starts with a classical image processing step. In this phase the axial slices are normalized and the most interfering imaging artifacts are removed. As two different data sets are used in this work, this preprocessing step is modality dependent. Different normalization functions are used for the MR and CT images.

To calculate the initial classifications, two solutions are presented. The first one uses the tools of the traditional image processing methods while the second relies on modern machine learning and uses convolutional neural networks. The output of both approaches have the same format but have different artifacts and accuracies.

The third step is independent from the underlying workflow elements. It takes the initial classification and the provided classification confidence values and filters the outliers and guarantees a continuous and anatomically correct labeling by fitting region membership functions.

An overview of the proposed method is presented in Figure 1. As it was stated before, the method can be divided into three main parts. In step 1, the axial slices of the image are separately pre-processed. This is required to eliminate several disturbing image artifacts and to reduce image complexity. A normalization is also introduced, so the slices can be used in a uniform way in the further processing steps.

Steps 2-4 represent the image recognition part that uses 2-dimensional feature extraction and a machine-learning approach to classify the individual axial slices of the image by assigning the probability of each label to each slice.

In method 3(a), to calculate the feature vectors, a global shape feature extraction algorithm is applied, called the Zernike transform. It is followed by a feature vector size reduction method using Principal Component Analysis (PCA). This way a compact representation is generated for each slice. These feature vectors are further processed by a Random Forest classifier. This
A machine-learning approach assigns a probability for each label to an input slice. The parameters of the PCA transform and the classifier are estimated during the training process. Method 3(b) realizes the same functionality, but uses a convolutional neural network that integrates all the above mentioned features.

Step 4 applies the previously trained model to the test images. Finally, in the 5th step, a post-processing method is applied. A three dimensional coherence inspection method, which takes the valid sequence and the mean size of the anatomy regions into consideration to eliminate the false slice classifications and guarantees the correct order of labels and the valid size of the recognized anatomy regions.

1. Figure. The schematic overview of the method.
My first thesis point summarizes the above described algorithm overview.

**Thesis 1:**
I have defined a workflow to classify the axial slices of CT and MRI images. The workflow consist of three stages. The first step is an image preprocessing stage, that normalizes the axial slices independently for further processing. The second step of the workflow classifies the axial slices and provides information about the class confidence values as well. The third step of the workflow applies a post processing method on the initial classification based on the confidence values. This method incorporates anatomical knowledge to filter unrealistic results from the previous step of the workflow.

*Related publications: [J1], [J2], [C1], [C3], [C4], [C5]*

4. **Initial Classification Using the Classical Approach**

The first step of the method is the normalization of the axial slices of the MR images. Since the MRI intensities are not standardized like in case of CT images, the intensity range can vary on a large scale, even if they were acquired using the same (e.g. T1) protocol. To deal with this phenomenon the intensity range of the MRI scan is normalized to the range $[0, 255]$ using a histogram-based and a linear technique.

Differences caused by the different imaging settings have to be also handled carefully. The input images may have different resolutions, field of view (FOV), and even the positioning of the patients can vary. To handle these problems a multistep process is applied. The resolution of the axial slices is down-sampled to 128x128 pixels. The next step of the normalization is the correction of the patient position. Different positions can cause translation of the body on the axial slices. To correct this effect, the Hu image moment [3] is computed for each axial slice.

The next step is the so-called feature extraction. As the input image resolution is 128x128 pixels, it can be considered as a 16384 dimensional vector. It means that the primary classifier has to find an optimal subdivision of a 16384 dimensional space to represent the five classes of the anatomical regions. This task would require a complex structure with many trainable parameters. To reduce this complexity, the idea is to reduce the dimension of the input while maintaining its information content. This procedure is called feature vector extraction.
The computation of these feature vectors consists of two steps. First, a global image shape descriptor is applied to capture the structural content of the image and subsequently a dimension reduction method is applied to the generated feature descriptor to minimize its size and the complexity of the machine learning model. In the first step, the Zernike transform [4], [5] is used, while in the second step a Principal Component Analysis (PCA) [6] is applied.

In many image recognition works various machine learning tools are applied to classify images based on extracted features. In this work the Random Forest classifier [7] was used. This classifier creates multiple, independently trained decision trees. It was shown in [7] that this composition of decision trees has better generalization ability than the classical decision trees through the applied randomness during the training process.

To calculate the parameters of the PCA and to train the random forest classifier, the following training process was performed (illustrated in Figure 1.). First, the image database is randomly split into two parts. The smaller part (including 22 patients) is used to train the system while the larger part (including 27 images), is used for the evaluation of the trained system. The training process estimates the parameters of the PCA and the random forest and provides a model that is used to label the test data. The initial classification is achieved by assigning the most likely label to each slice. After the labeling is performed, its result is compared with the manually assigned reference labeling. The result of the comparison is displayed using a confusion matrix in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>true LEG</th>
<th>true PELVIS</th>
<th>true ABDOMEN</th>
<th>true CHEST</th>
<th>true HEAD&amp;NECK</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. LEG</td>
<td>1514</td>
<td>81</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>94.21%</td>
</tr>
<tr>
<td>pred. PELVIS</td>
<td>117</td>
<td>1663</td>
<td>86</td>
<td>0</td>
<td>0</td>
<td>89.12%</td>
</tr>
<tr>
<td>pred. ABDOMEN</td>
<td>0</td>
<td>212</td>
<td>1735</td>
<td>116</td>
<td>1</td>
<td>84.06%</td>
</tr>
<tr>
<td>pred. CHEST</td>
<td>3</td>
<td>50</td>
<td>88</td>
<td>1665</td>
<td>19</td>
<td>91.23%</td>
</tr>
<tr>
<td>pred. HEAD&amp;NECK</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>43</td>
<td>2098</td>
<td>97.90%</td>
</tr>
<tr>
<td>class recall</td>
<td>92.66%</td>
<td>82.90%</td>
<td>90.55%</td>
<td>90.93%</td>
<td>99.06%</td>
<td>91.27%</td>
</tr>
</tbody>
</table>

1. Table: Accuracy of the initial slice classification using the classical approach.
The columns of the table specify the known ground truth labeling while the rows represent the labeling assigned by the trained system. The summarized accuracy of the random forest classifier measured on the test set. According to the table the overall accuracy was 91.27%, which means that 91.27% of the slices involved in the test set were correctly labeled.

The main diagonal of the table contains the correctly labeled slices for each class. The elements directly above and below of the diagonal represent slices which were confused with a neighboring region, while the remaining elements represent large confusion in the initial labeling. The first type of the confusion errors is not considered as a major problem, because it usually means border shift between the neighboring regions or some alternating labels near the region borders. The second type of confusion is considered as greater mismatch which means the classifier completely missed the labeling (e.g. 50 PELVIS slices were labeled as CHEST by the initial classification). This effect could be caused by several factors. As the random forest classification is a machine-learning tool that performs statistical operations, it has the possibility that the generated model could not handle all feature vector occurrences correctly.

Examples for the above-mentioned errors are illustrated in Figure 2. The left side of both examples shows the initial labeling (based on maximum confidence) and the confidence values provided by the random forest are plot on the right side. As it can be seen, even the most accurate classification result (2\(a\)) includes some outliers (some slices in the middle of the ABDOMEN are labeled as PELVIS). This indicates that the independent processing of slices, even if the overall accuracy seems high enough, cannot be reliable, as it does not incorporate the neighborhood of the slices. When 2\(b\) is considered, the results are more confusing. Although the majority of slices are correctly classified, the result contains several misclassified slices (e.g. chest detected in the pelvis). Some labels are confused with labels of other regions, so such a result cannot be considered as useful information in clinical applications.
My second thesis point summarizes the image normalization and initial classification method definition for the MRI data set using the classical approach.

**Thesis 2:**
According to the requirements of the classification workflow, I have developed a method to normalize the input axial images and to provide the initial classification for MRI data. The normalization step performs translation correction, field of view and resolution normalization using classic image processing methods. To generate the initial classification, conventional machine learning tools are used. For image feature extraction the Zernike transform is applied, whose result is further processed with the PCA method to reduce its size. The feature vector is classified by using the Random Forest method, which also returns the classification confidence values.

*Related publications: [J1], [C1], [C3], [C4], [C5]*
5. **Initial Classification Using a Convolutional Neural Network**

Similarly to the classical approach the processing starts with an image normalization step that eliminates some imaging artifacts and reduces the image variety to provide a more easily processable input to the method.

The position of the patient with respect to image center can vary significantly among examinations, which introduces unwanted variation. Even if the patient is acquired in supine position, the table can be lifted in anterior or posterior direction. In order to compensate that, the weight center of non-air voxels is computed that is used to define the body axis for the whole CT exam.

After this point, each slice is separately processed. First, a squared region is extracted, such that its center is located on the body center axis (that is defined as the average of the weight center for all slices), its size (both width and height) is equal to 35 cm, and its resolution is equal to 256x256 pixels. Since the pixel size of CT images varies among examinations, interpolation is used. If the original input slice covers larger or smaller image region than 35 cm, the pre-processed image is cropped or padded (with air voxels), respectively. Then, the pixel intensities (from the original 16-bit signed integer value) are rescaled to the range \([0,255]\), such that the range \([-500,500]\) HU is linearly mapped to \([0,255]\). This step enhances those tissue types which are key important from anatomy point of view and prevents the classifier taking very high or low density pixels into account.

In the recent years, the image classification tasks are usually solved with artificial neural networks, as they offer a more comfortable way to achieve same or better results compared to the classical image processing approaches.

In this work AlexNet [8] was used for the classification of the 2D grayscale images. The structure of the network follows the originally introduced topology. The network contains eight layers. The first five are convolutional while the remaining are fully connected. Only the last softmax output is modified to fit the current problem size.

The CT image dataset was split to three nearly equal subsets, such that the training, the cross-validation, and the test sets included 66 000, 55 000, and 60 000 labeled images, respectively. The first set was used to train the classifier, the second was used to monitor the accuracy.
during the training, and the last one was used to evaluate the CNN model (as well as the post-processed result).

In order to simulate all possible patient positions, and to perform some data augmentation, the axial slices of the training dataset (that included images acquired in supine position only) were rotated 0, 90, 180, and 270 degrees.
### Table Accuracy of the initial slice classification using the neural network.

<table>
<thead>
<tr>
<th>Pred. \ True</th>
<th>FOOT</th>
<th>SHIN</th>
<th>THIGH</th>
<th>PE-LO</th>
<th>PE-CR</th>
<th>PE-UP</th>
<th>AB-LO</th>
<th>AB-UP</th>
<th>CHEST</th>
<th>HEAD</th>
<th>BRAIN</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOOT</td>
<td>1670</td>
<td>94</td>
<td>57</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>9</td>
<td>3</td>
<td>90.40%</td>
</tr>
<tr>
<td>SHIN</td>
<td>64</td>
<td>5662</td>
<td>78</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>7</td>
<td>97.40%</td>
</tr>
<tr>
<td>THIGH</td>
<td>17</td>
<td>89</td>
<td>5273</td>
<td>318</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>15</td>
<td>4</td>
<td>92.10%</td>
</tr>
<tr>
<td>PE-LO</td>
<td>0</td>
<td>0</td>
<td>133</td>
<td>3377</td>
<td>106</td>
<td>4</td>
<td>146</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>89.70%</td>
</tr>
<tr>
<td>PE-CR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>399</td>
<td>3782</td>
<td>257</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>85.10%</td>
</tr>
<tr>
<td>PE-UP</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>51</td>
<td>518</td>
<td>15140</td>
<td>0</td>
<td>130</td>
<td>34</td>
<td>12</td>
<td>0</td>
<td>95.30%</td>
</tr>
<tr>
<td>AB-LO</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>61</td>
<td>193</td>
<td>549</td>
<td>14335</td>
<td>801</td>
<td>40</td>
<td>11</td>
<td>0</td>
<td>89.50%</td>
</tr>
<tr>
<td>AB-UP</td>
<td>21</td>
<td>0</td>
<td>15</td>
<td>25</td>
<td>0</td>
<td>11</td>
<td>239</td>
<td>12455</td>
<td>186</td>
<td>45</td>
<td>0</td>
<td>95.80%</td>
</tr>
<tr>
<td>CHEST</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>28</td>
<td>11</td>
<td>265</td>
<td>436</td>
<td>15087</td>
<td>143</td>
<td>17</td>
<td>94.30%</td>
</tr>
<tr>
<td>HEAD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>18</td>
<td>39</td>
<td>19</td>
<td>67</td>
<td>4033</td>
<td>68</td>
<td>94.80%</td>
</tr>
<tr>
<td>BRAIN</td>
<td>30</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>121</td>
<td>5968</td>
<td>97.40%</td>
<td></td>
</tr>
<tr>
<td>Recall</td>
<td>92.50%</td>
<td>96.90%</td>
<td>94.20%</td>
<td>79.60%</td>
<td>81.70%</td>
<td>94.60%</td>
<td>95.40%</td>
<td>89.90%</td>
<td>97.80%</td>
<td>91.90%</td>
<td>98.40%</td>
<td>93.40%</td>
</tr>
</tbody>
</table>
Table 2 presents the confusion matrix of the applied CNN network evaluated on the test data set. The columns of the table specify the known ground truth labeling while the rows represent the labeling assigned by the trained system.

The overall accuracy of the initial classification was 93.4% and the precision varied between 85.1% and 97.4%. The values in the main diagonal represent the correctly labeled slices. The values directly below or above the main diagonal indicate some minor confusion, usually occurring at the border of the anatomical regions. For example 143 HEAD slices were labeled as CHEST. Other confusions mean serious misclassification errors. This effect indicates the independent processing of the slices, as the labeling process does not incorporate the neighborhood information at this point.

Figure 3 shows two examples from the test set. In the images the central coronal slice of the CT exams are shown and each row of the images represent an axial slice of the corresponding CT exam. In Figure 3(a) the initial classification accuracy is 84.7% and it is free from significant misclassifications while Figure 3(b) is one of the worst initial results (accuracy: 67.8%) and contains examples for all major errors. As the labeling of the slices are independent, the CNN can produce alternating labeling near the region borders (e.g. between PELVIS-CENTER and PELVIS-UPPER) or can completely miss the labeling as can be seen in the HEAD and CHEST regions.

My third thesis point summarizes the initial classification method that uses a convolutional neural network to classify the axial slices of CT images.
I have defined an alternative solution for the image normalization and initial classification. This approach performs translation correction, field of view and resolution normalization for CT images. The initial classification is performed using a convolutional neural network that integrates the feature extraction and the classification process as well. I have used a pretrained AlexNet structure whose last fully connected layer is modified according the needs of the slice classification workflow.

Related publications: [J1], [C2], [C5]

6. Establishing Coherent Labeling

Up to this point in the proposed method, the slices are handled independently in the initial classification phase. This approach leads to some errors which make the recognition process less reliable. These errors are corrected by a post-processing step described in this section. This step incorporates some a priori information about the human anatomy in the recognition process to establish a coherent final labeling.

The first and most obvious fact that is exploited is the order of the anatomical regions. The labels shall follow each other in the right order: LEG, ..., LEG, PELVIS, ..., PELVIS, ABDOMEN, ..., ABDOMEN, CHEST, ..., CHEST, HEAD&NECK, .... in case of the MRI data set. In the CT data set the correct order is FOOT, ... FOOT, SHIN, ..., SHIN, THIGH, ..., THIGH, PELVIS_LOWER, ..., PELVIS_LOWER, PELVIS_CENTER, ..., PELVIS_CENTER, PELVIS_UPPER, ..., PELVIS_UPPER, ABDOMEN_LOWER, ..., ABDOMEN_LOWER, ABDOMEN_UPPER, ..., ABDOMEN_UPPER, CHEST, ..., CHEST, HEAD, ..., HEAD, BRAIN, ... . These orders cannot be changed, as they are a major property of the human body.

The second consideration is that the size of the anatomy regions cannot be arbitrarily large or small. It can vary among patients but each region has its expected length and variance. The post-processing can incorporate these lengths to estimate the final labeling. The mean size of each region was calculated based on the manually labeled test databases that are large enough to have a good estimation for these statistical values.

The main goal of this step is to produce a continuous and reliable labeling for the whole image series. A labeling can be considered acceptable if it meets the following requirements:
• The labeling must be continuous. No alternating labeling can occur.

• The anatomical regions should appear in their correct order e.g. the CHEST region should be followed by the HEAD

• The size of the anatomical regions should be reasonable. The regions cannot be arbitrarily small or large.

To obtain the desired properties of the labeling, a region membership function-fitting algorithm is proposed. In this method the anatomical regions are represented with generalized normal distributions (GDF) [9] (1).

\[
GDF(x) = \frac{\beta}{2\alpha \Gamma(1/\beta)} \cdot e^{(-|x-\mu|/\alpha)^\beta}
\]

(1)

Each membership function can be customized to have region specific parameters. To estimate the position (\(\mu\)) and range of the membership functions (\(\alpha\)), the confidence values, provided by the random forest or by the CNN, can be utilized. The \(\beta\) shape parameter is set to a fixed value (\(\beta = 6\)) in this work.

In the first step of the method, a set of regions, which are possibly involved in the image, is selected. This provides a list of candidate regions. From this list the anatomically correct sequences are assembled and the most likely one is chosen for further processing.

In the next step, a membership function set is produced to represent the selected region sequence. Based on the confidence values of the initial labeling, the position and range parameters of the functions are optimized. In this way the functions are fit on the image series and a realistic final labeling can be obtained.

To estimate the visible regions in the image, the confidence weighted size of each region is calculated (2). If this value is greater than a predefined limit, the region is considered visible in the image. In this way the small and misclassified regions can be eliminated from the further processing.

\[
S_i = \sum_{k=1}^{N} conf(k, i) T_k,
\]

(2)
where \( N \) is the number of the slices, \( \text{conf}(k, i) \) is the confidence value provided by the initial classifier describing the likelihood of that slice \( k \) is located in region \( i \), and \( T_k \) is the physical thickness of slice \( k \) (\( T_k \) is always available in the DICOM header).

As soon as the visible regions are available, all of the anatomically correct sequences of these regions are assembled. For example, if ABDOMEN, CHEST and HEAD&NECK regions are visible in the MRI scan, the following valid region combinations are generated: (ABDOMEN); (CHEST); (HEAD&NECK); (ABDOMEN, CHEST); (CHEST, HEAD&NECK) and (ABDOMEN, CHEST, HEAD&NECK)

To select the most likely sequence the Summarized Accepted Confidences (SAC) is calculated:

\[
R_{ci} = \frac{\sum_{k=1}^{N} \text{conf}(k, i) \cdot k}{\sum_{k=1}^{N} \text{conf}(k, i)}
\]

\[
SAC = \sum_{i=1}^{M} \sum_{k=1}^{N} \text{conf}(k, i) \{R_{ci-1} < k < R_{ci+1}\},
\]

where \( M \) is the number of regions in the sequence, \( R_{ci-1} \) and \( R_{ci+1} \) are the confidence weighted center of the previous and next regions, respectively. This means the confidence value \( \text{conf}(k, i) \) is accepted only if the position of the \( k \)-th slice is between the centers of the neighboring regions. This eliminates the most interfering outliers from the SAC calculations. This way the SAC value indicates the summarized confidence values of the slices that are positioned and labeled correctly according to the processed sequence.

Once all the SAC values are calculated, the region sequence with the highest SAC can be selected as the most likely combination of the candidate regions.

After selecting the most likely region sequence, a set of membership functions are fitted on the confidence values. This process estimates the position and range parameters of the functions to provide a continuous and reliable labeling. Regarding these calculations, the following considerations are taken:

- The membership functions should maximize the correctly covered confidence values.
- The membership functions should minimize the incorrectly covered confidence values.
• The membership functions should minimize the overlap between each other.

To perform the parameter estimation an iterative method is used. The position of the membership functions are initialized with the confidence weighted center of each region and their range property is set to the average size of the represented region. Figure 4(a) shows the initial position and range estimations for a given confidence distribution.

To meet the requirements, a cost function is defined that consists of three components. The first one, defined in (5), penalizes if a membership function $GDF_j$ does not cover slices that belong to region $j$.

$$cf_1 = \sum_{j=1}^{M} \sum_{k=1}^{N} (1 - GDF_j(k)) \cdot (1 + conf(k,j))^2$$

(5)

The second component (6) penalizes if a membership function covers slices that do not belong to that specific region.

$$cf_2 = \sum_{j=1}^{M} \sum_{k=1}^{N} \sum_{i=1, i\neq j}^{M} GDF_j(k) \cdot (1 + conf(k,i))^2$$

(6)

The third component (7) of the cost function tries to minimize the overlap between the region membership functions.

$$cf_3 = \sum_{j=1}^{M} \sum_{k=1}^{N} \sum_{i=1, i\neq j}^{M} \min\left(GDF_j(k), GDF_i(k)\right)$$

(7)

The final cost function is defined in (8) as the sum of the three components.

$$cf = cf_1 + cf_2 + cf_3$$

(8)

To perform the optimization and to find the desired values for the $\alpha$ and $\mu$ parameters of the region membership functions a gradient descent iterative optimization method is applied.

Figure 4(b) shows the membership functions after the post-process. It can be observed that the functions are positioned according to the input confidence values and their range fit the confidences as well. The correct ordering of the regions is maintained during the iterations and their size is in the acceptable range.
My fourth thesis point summarizes the post processing method that filters the results of the initial classification methods and guarantees a reliable and reasonable final classification.

**Thesis 4:**

I have defined a post processing method for the slice classification workflow. This method incorporates anatomical knowledge and can operate on the results of both initial classification methods. It fits region membership functions on the classification confidence values. The fitting method guarantees the correct order of the anatomical regions and estimates the parameters of the membership functions to correctly cover the confidence values.

*Related publications: [J1], [J2], [C1] [C3]*

**7. Evaluation**

The proposed method was evaluated first on the whole body MRI dataset. The results are presented in Table 3. According to the table the overall accuracy increased to 94.48% and the most important issue of the initial labeling (confusion with non-neighboring region) was completely eliminated. The confusion matrix has nonzero values only in the main diagonal (94.48%) and in cells directly below and above the diagonal (5.52%). This means the majority of the slices are correctly labeled and there are confusions only between neighboring regions. There were no serious mistakes in contrast to Table 1 where 0.72% of the slices were significantly misclassified. As the correct sequence of the labels is always ensured in the result,
the confusion between the neighboring regions mean only small displacements of the borders between two regions.

<table>
<thead>
<tr>
<th></th>
<th>true LEG</th>
<th>true PELVIS</th>
<th>true ABDOMEN</th>
<th>true CHEST</th>
<th>true HEAD&amp;NECK</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>pred. LEG</td>
<td>1579</td>
<td>55</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>96.63%</td>
</tr>
<tr>
<td>pred. PELVIS</td>
<td>92</td>
<td>1865</td>
<td>49</td>
<td>0</td>
<td>0</td>
<td>92.97%</td>
</tr>
<tr>
<td>pred. ABDOMEN</td>
<td>0</td>
<td>101</td>
<td>1735</td>
<td>80</td>
<td>0</td>
<td>90.55%</td>
</tr>
<tr>
<td>pred. CHEST</td>
<td>0</td>
<td>0</td>
<td>46</td>
<td>1693</td>
<td>92</td>
<td>92.46%</td>
</tr>
<tr>
<td>pred. HEAD&amp;NECK</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>2108</td>
<td>99.53%</td>
</tr>
<tr>
<td>class recall</td>
<td>94.49%</td>
<td>92.28%</td>
<td>94.81%</td>
<td>94.95%</td>
<td>95.82%</td>
<td>94.48%</td>
</tr>
</tbody>
</table>

3. Table. Accuracy of slice classification after post-processing using whole body MRI cases.

Figure 5 illustrates the accuracy improvement for one selected case, the same image can be seen in Figure 2b. It shows both the initial classification and the effect of the post-processing method. As it can be seen the most serious misclassifications were corrected, the correct sequence of the labels was established and the all detected regions have a reasonable size and fit the real anatomical structure of the human body.

5. Figure. Effect of the post-processing in an MR image.

Since the acquisition of full body scans is not so frequent in the clinical practice (especially in case of MRI modality), the method was tested on partial images as well. The test images were created from the full body scans by using a sliding window sampling. Using this technique,
different samples were cut out from the whole-body scans at different locations with different sample sizes. The cut size varied from 350 mm to 1150 mm (with 100 mm steps), and all possible subseries were generated. This means that all together 36,681 different partial body scans were generated which contained 10,513,282 slices.

The evaluation was performed similarly to the previous cases. The slices were independently labeled with the initial method and then, the post-process was applied, but in these cases, the post-process had significantly less information to work with.

<table>
<thead>
<tr>
<th>true LEG</th>
<th>true PELVIS</th>
<th>true ABDOMEN</th>
<th>true CHEST</th>
<th>true HEAD&amp;NECK</th>
<th>class precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>943614</td>
<td>88995</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>91.38%</td>
</tr>
<tr>
<td>91020</td>
<td>2447127</td>
<td>118326</td>
<td>0</td>
<td>0</td>
<td>92.12%</td>
</tr>
<tr>
<td>0</td>
<td>238429</td>
<td>2921987</td>
<td>60043</td>
<td>0</td>
<td>90.73%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>217707</td>
<td>2195947</td>
<td>20135</td>
<td>90.23%</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>83181</td>
<td>1086771</td>
<td>92.89%</td>
</tr>
<tr>
<td>91.20%</td>
<td>88.20%</td>
<td>89.69%</td>
<td>93.88%</td>
<td>98.18%</td>
<td>91.27%</td>
</tr>
</tbody>
</table>

4. Table: Accuracy measured on partial MRI scans.

Table 4 summarizes the accuracy of the anatomy labeling measured on the partial body scans. The numbers show the overall accuracy was 91.27% that is somewhat lower than in case of whole body scans. The table shows that the majority of the slices are correctly labeled, there were confusions only between the neighboring regions (8.73%), and there were no serious misclassifications.
In the clinical practice the acquisition of pathological cases is part of the daily examination routines. The method shall be prepared and robust in the recognition of such cases. Figure 6 includes some examples from the data set containing pathological distortions.

6. Figure. Pathological MRI cases.

The initial classification of the CT dataset is further refined with post-processing method. The artifacts of the initial labeling can be eliminated and a continuous and reliable labeling can be obtained. The post-processed results are summarized in Table 5. As one can see the overall accuracy increased to 94.13%. Furthermore, only the main diagonal and values below and above that are not zero, so the most interfering errors, confusion of non-neighboring regions, were completely eliminated thanks to the post-process. As the post-process guarantees the continuous labeling, the values directly below and above the main diagonal of the confusion matrix mean only small displacements of the borders between the neighboring regions.

The effect of the post-process is illustrated in Figure 7(a). Compared to Figure 3(b) it can be seen, that the initial misclassification errors are eliminated and a continuous labeling is obtained without alternating labels near the borders of the regions. Figure 7(b) shows the ground truth labeling of the examined CT series. It can be observed that the order of the anatomical regions is correct, most of the regions have the correct size and only the PELVIS-LOWER region became smaller than the original one in the ground truth image.
<table>
<thead>
<tr>
<th>Pred. \ True</th>
<th>FOOT</th>
<th>SHIN</th>
<th>THIGH</th>
<th>PE-LO</th>
<th>PE-CR</th>
<th>PE-UP</th>
<th>AB-LO</th>
<th>AB-UP</th>
<th>CHEST</th>
<th>HEAD</th>
<th>BRAIN</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOOT</td>
<td>1769</td>
<td>138</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>92.76%</td>
</tr>
<tr>
<td>SHIN</td>
<td>36</td>
<td>5616</td>
<td>113</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>97.42%</td>
</tr>
<tr>
<td>THIGH</td>
<td>0</td>
<td>91</td>
<td>5419</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>91.79%</td>
</tr>
<tr>
<td>PE-LO</td>
<td>0</td>
<td>0</td>
<td>64</td>
<td>3264</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>93.79%</td>
</tr>
<tr>
<td>PE-CR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>490</td>
<td>3942</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>83.68%</td>
</tr>
<tr>
<td>PE-UP</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>533</td>
<td>15292</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>95.34%</td>
</tr>
<tr>
<td>AB-LO</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>428</td>
<td>14525</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>91.70%</td>
</tr>
<tr>
<td>AB-UP</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>95.32%</td>
</tr>
<tr>
<td>CHEST</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>531</td>
<td>15088</td>
<td>249</td>
<td>95.08%</td>
</tr>
<tr>
<td>HEAD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>4025</td>
<td>89</td>
<td>96.94%</td>
<td></td>
</tr>
<tr>
<td>BRAIN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>115</td>
<td>5978</td>
<td>98.11%</td>
</tr>
<tr>
<td>Recall</td>
<td>98.01%</td>
<td>96.08%</td>
<td>96.84%</td>
<td>78.69%</td>
<td>85.20%</td>
<td>95.58%</td>
<td>96.55%</td>
<td>89.76%</td>
<td>97.77%</td>
<td>91.71%</td>
<td>98.53%</td>
<td>94.13%</td>
</tr>
</tbody>
</table>

5. Table. Accuracy of slice classification after post-processing using CT cases
Further results and statistics can be found in corresponding chapter of the dissertation.

8. Conclusion

In this work I have presented a method with two possible initial classification techniques to solve the problem of the recognition of axial slices in 3D medical images.

The complete method starts with an image pre-processing stage that corrects some translation differences caused by the patient laying and a field of view normalization. This step reduces the variance of the images and makes the work of the following steps easier.

In the next phase, two methods are presented to perform the initial slice-by-slice classification. The first method uses traditional image-processing tools. The normalized axial slices are processed with the Zernike transform that creates a lower-dimensional representation of the images. This transform uses circular basis function to represent the images. The structure of the basis functions fits well to the human anatomy and it is rotationally invariant therefore the representation contains all the required information to solve the initial classification task.

To perform the actual classification a Random Forest classifier is used. This machine-learning tool uses multiple decision trees each one trained with a random subset of the training data. As it can be seen from the classification results, it can solve the task with good accuracy.
The second method I have presented solves the classification with a convolutional neural network. It omits all the intermediate steps of the previous method and gives the classification result directly from the input image.

By considering the structure and the working mechanism of the neural network, it is more precise saying that the neural network incorporates the intermediate steps. It also performs a feature extraction and creates a compact representation, but these steps are not separated. The network learns these transforms during the training phase. It selects the optimal transform, which fits the data best, automatically.

This version of the initial classification performs better in general even in a more complex environment with more regions to recognize.

If we check the results in numbers only, we can see that the achieved accuracy is promising in both methods. However, if some of the MRI or CT scans are examined as a whole, we can see some major classification errors. This can happen because the axial slices of a scan are processed independently. The continuity of the achieved labeling is not guaranteed and no anatomical knowledge is incorporated in the methods to check the validity of the results.

These insufficiencies are corrected in the presented post-processing method. It selects the anatomical regions that are most likely present in the image, based on the confidence values of the initial classification, and fits the regions on the whole scan. During the fitting it maintains the correct order of the anatomical regions, selects their appropriate positions and gives a good estimation about their sizes. This process takes the anatomical knowledge into consideration to ensure a continuous labeling without outliers, too small or too large regions.
9. Publications

<table>
<thead>
<tr>
<th>Number of publications</th>
<th>27</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of journal papers</td>
<td>3</td>
</tr>
<tr>
<td>Number of articles in journals indexed by WoS or Scopus</td>
<td>2</td>
</tr>
<tr>
<td>Number of independent citations:</td>
<td>8</td>
</tr>
</tbody>
</table>

**Journal Papers Related to the Theses**

[J1] **Tóth, Márton József**; Ruskó, László ; Csébfalvi, Balázs
Automatic Recognition of Anatomical Regions in Computed Tomography Images
PERIODICA POLYTECHNICA-ELECTRICAL ENGINEERING AND COMPUTER SCIENCE

[J2] **Tóth, Márton József**; Ruskó, László ; Csébfalvi, Balázs
Automatic recognition of anatomical regions in three-dimensional medical images
COMPUTERS IN BIOLOGY AND MEDICINE 76 pp. 120-133. , 14 p. (2016)

**Conference Papers Related to the Theses**

[C1] **Tóth, Márton József**; Ruskó, László ; Csébfalvi, Balázs
Változó méretű anatómiai régiók szegmentálása MRI felvételeken

[C2] **Tóth, Márton József** ; Csébfalvi, Balázs ; Ruskó, László
Convolutional Neural Networks for Anatomical Region Detection

[C3] **Tóth, Márton József** ; Ruskó, László ; Csébfalvi, Balázs
Automatic Recognition of Anatomical Regions
**C4**  Tóth, Márton József; Blaskovics, Tamás; Ruskó, László; Delso, Gaspar; Csébfalvi, Balázs

Automated Detection of Anatomical Regions in Magnetic Resonance Images
Jan, Bender; Arjan, Kuijper; Tatiana, von Landesberger; Holger, Theisel; Philipp, Urban (szerk.) Vision, Modeling & Visualization

**C5**  Tóth, Márton József; Ruskó, László; Csébfalvi, Balázs; Blaskovics, Tamás

Anatómiai régiók automatikus detektálása
In: KÉPAF 2013 pp. 1-10. , 10 p.

<table>
<thead>
<tr>
<th>Thesis</th>
<th>Related Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>[J1], [J2], [C1], [C3], [C4], [C5]</td>
</tr>
<tr>
<td>2</td>
<td>[J1], [C1], [C3], [C4], [C5]</td>
</tr>
<tr>
<td>3</td>
<td>[J1], [C2], [C5]</td>
</tr>
<tr>
<td>4</td>
<td>[J1], [J2], [C1] [C3]</td>
</tr>
</tbody>
</table>

**Other Publications, not Related to the Theses**

**O1**  Gergely, Ferenc Rácz; Ágota, Kacsó; Márton, Tóth; Balázs, Tóth

Fast and accurate initial parameter fitting in Positron Emission Tomography using Neural Networks
Kiss, Bálint; Szirmai-Kalos, László – WAIT 2019 pp. 65-71. , 7 p

**O2**  Ágota, Kacsó; László, Szécsi; Márton, Tóth; Balázs, Benyó; Tamás, Umenhoffer

Finite volume blood flow simulation for highly deformable boundaries
Bálint, Kiss; László, Szirmai-Kalos (szerk.) Proceedings of the Workshop on the Advances of Information Technology : WAIT 2018

**O3**  Gergely, Ferenc Rácz; Ágota, Kacsó; Márton, Tóth; Balázs, Tóth

PET Image Denoising using a Deep Neural Network
Bálint, Kiss; László, Szirmay-Kalos (szerk.) Proceedings of the Workshop on the Advances of Information Technology : WAIT 2018

[O4] Márton, Tóth ; Tamás, Umenhoffer ; László, Szécsi ; Ágota, Kacsó ; Balázs, Benyó
Aortic Root Simulation Using Smoothed Particle Hydrodynamics
Bálint, Kiss; László, Szirmay-Kalos (szerk.) Proceedings of the Workshop on the Advances of Information Technology : WAIT 2018

[O5] Rácz, Gergely ; Kacsó, Ágota ; Tóth, Márton ; Tóth, Balázs
Enhanced PET Reconstruction with Neural Networks
Szirmay-Kalos, László; Renner, Gábor (szerk.) IX. magyar számítógépes grafika és geometria konferencia, GRAFGEKO 2018

[O6] Tamas, Umenhoffer ; Marton, Toth ; Agota, Kacso ; Laszlo, Szecsi ; Akos, Szlavecz ; Peter, Somogyi ; Laszlo, Szilagyi ; Aniko, Kubovje ; Tamas, Szerafin ; Laszlo, Szirmay-Kalos et al.
Modeling and simulation framework of aortic valve for hemodynamic evaluation of aortic root replacement surgery outcomes

[O7] Tóth, Balázs ; Tóth, Márton József ; Kacsó, Ágota Enikő ; Rácz, Gergely Ferenc ; Szirmay-Kalos, László
Controlling TV Regularization with Deep Learning

[O8] Umenhoffer, Tamás ; Tóth, Márton ; Szécsi, László ; Kacsó, Ágota ; Benyó, Balázs
Aortic Root Simulation Framework for Valve Sparing Aortic Root Replacement Surgery
Bálint, Kiss; László, Szirmay-Kalos (szerk.) Proceedings of the Workshop on the Advances of Information Technology : WAIT 2018
[O9] Tóth, Márton József
Progresses in a Fluid Mechanics Simulation Engine on the GPU
Kiss, Bálint; Szirmay-Kalos, László (szerk.) Proceedings of the Workshop on the Advances of Information Technology : WAIT 2017

[O10] Tóth, Márton József ; Csébfalvi, Balázs
Distribution Interpolation of the Radon Transforms for Shape Transformation of Gray-Scale Images and Volumes

[O11] Tóth, Márton József ; Csébfalvi, Balázs
Agyszegmens detektálása CT felvételeken inverz anizotróp diffúzióval

[O12] Józsa, Péter ; Tóth, Márton József ; Csébfalvi, Balázs
Analytic Isosurface Rendering and Maximum Intensity Projection on the GPU
Vaclav, Skala (szerk.) WSCG 2014 Full Papers Proceedings

[O13] Márton, József Tóth ; Balázs, Csébfalvi
Recent Results on Shape-Based Interpolation
Szirmay-Kalos, László; Renner, Gábor (szerk.) VII. Magyar Számítógépes Grafika és Geometria Konferencia

[O14] Márton, József Tóth ; Balázs, Csébfalvi
Shape Transformation of Multidimensional Density Functions using Distribution Interpolation of the Radon Transforms

[O15] Tóth, Márton József ; Csébfalvi, Balázs
Mass-Spring Models for Anisotropic Diffusion
Török, Marianna; Tóth, Márton József; Szöllősi, Alexandra
Foundations and perspectives of mathability in relation to the CogInfoCom domain

Márton, Tóth; Dávid, Dvorszki; Balázs, Csébfalvi
Robust Volume Segmentation using an Abstract Distance Transform

Márton, Tóth; Dávid, Dvorszki; Balázs, Csébfalvi
GPU-Accelerated Segmentation of Medical Volume Data

Rácz, Gergely Ferenc; Kacsó, Ágota; Tóth, Márton; Szirmay-Kalos, László
Dynamic PET Reconstruction from Sinogram Data using Neural Networks

Tóth, Márton; Kacsó, Ágota; Magdics, Milán; Salvi, Péter
Advances in the visualization framework for vehicle intelligence simulations
10. **Bibliography**


