Analysis of Medical Images

Registration, Segmentation and Classification

Landgren, Matilda

2016

Document Version:
Publisher's PDF, also known as Version of record

Analysis of Medical Images

Registration, Segmentation and Classification

Matilda Landgren

Lund University

Faculty of Engineering
Centre for Mathematical Sciences
Mathematics

Mathematics
Centre for Mathematical Sciences
Lund University
Box 118
SE-221 00 Lund
Sweden

http://www.maths.lth.se/

Doctoral Theses in Mathematical Sciences 2016:8
ISSN 1404-0034

ISBN 978-91-7623-988-9 (print)
ISBN 978-91-7623-989-6 (pdf)
LUTFMA-1059-2016

Printed in Sweden by MediaTryck, Lund 2016

Abstract

A large number of medical examinations involve images in some way. Images can be used for diagnostics, follow-up studies and treatment planning. In this thesis mathematical methods have been developed and adapted in order to analyze medical images. Several applications for different imaging modalities have been studied and the usefulness of such methods is demonstrated.

A complete system for detection and diagnosis of kidney lesions in scintigraphy images has been developed. We segment the kidneys with the use of an active shape model. The uptake of a biological molecule is then compared to the uptake in a healthy kidney and potential lesions are detected. A number of properties of the potential lesions are gathered and the lesions are classified as healthy or unhealthy with a linear classifier. We are able to correctly classify 86 % of the lesions.

Ultrasound images have also been studied. In the first case for the purpose of segmenting the left heart ventricle, which can be used for computing the ejection fraction. This was done using a region based snake with anchor points at each side of the cardiac valve. The second application in ultrasound images is also of the heart but with patients that, due to heart failure, have had a mechanical pump implanted. The septum wall between the ventricles is segmented using a shortest path approach and a measure of how much the septum bulges towards either of the ventricles is obtained. By studying this measure a more objective indication is given on whether the speed of the pump is correct for a patient than by only visually study the images.

In computed tomography (CT) whole-body images, several organs have been segmented using a multi-atlas approach. The fused labels are refined with a random forest classifier and a final graph cut segmentation. This method was evaluated in the VISCERAL Grand Anatomy Challenge and achieved the highest Dice

score for 13 out of 20 organs. A development of this approach was done in order to achieve qualitatively better segmentations of the organs. Instead of fusing organ labels, a map of corresponding landmarks is obtained and the segmentation is given by the robust average of these with similar refinement steps as in the origin work. The segmentation results using this method is on par with or better than state-of-the-art. Segmentation of organs is important in e.g. radiotherapy planning.

In another project with CT images, vertebrae have been detected and identified. This is useful in for instance surgical planning. The detection is done using convolutional neural networks. A shape model of the spine is fitted to the detections in order to correctly identify them. The task is difficult because, in general, only a limited part of the spine is visible. We are able to correctly identify 63 % of the vertebrae.

Populärvetenskaplig sammanfattning

Inom sjukvården tas en hel del bilder av olika slag för att undersöka patienter. Det kan till exempel vara för diagnostisering, uppföljning av en åkomma eller planering av en behandling. Dessa bilder analyseras med eller utan hjälpmedel av läkare och biomedicinska analytiker. Den här avhandlingen syftar till att utveckla matematiska metoder som kan ge stöd vid sådana analyser. Metoderna är inte till för att helt ersätta den medicinska personalen utan framför allt för att ge de ett hjälpmedel som kan belysa områden som är av större intresse att titta närmare på, ge ett objektivt råd samt minska det tidskrävande jobbet att t.ex. manuellt markera organ i en bild.

Ett av projekten i avhandlingen består av ett komplett system för datorstödd diagnostisering av ärr i njurarna hos barn som haft många urinvägsinfektioner. Systemet klarar av att göra en korrekt bedömning av möjliga ärr som antingen friska eller sjuka i ca 86 % av fallen. Avbildningsmetoden som används här är njurscintigrafi och innebär att man ger en injektion med ett radioaktivt ämne bundet till en biologisk molekyl som tas upp i de delarna av njurarna som är friska. Det radioaktiva ämnet avger strålning som registreras av en gammakamera och man får en bild av njurarna där de friska delarna av njuren lyser starkare än områden med ärr. Njurarna hittas och segmenteras, dvs. utlinjeras, automatiskt i bilden med hjälp av en metod som tar hänsyn till att njuren har en typisk form. Sedan jämförs upptaget i njurarna mot hur upptaget i en frisk njure brukar se ut för att undersöka om det finns avvikelser som kan vara potentiella ärr. Ett antal egenskaper för njurarna och områden med lägre upptag plockas ut och används sedan för att avgöra om det är ett skadligt ärr man har lokaliserat genom att träna en klassificeringsmetod med information från kända fall av sjuka och friska

områden.

Ultraljudsbilder är också vanligt förekommande som undersökningsmetod. Bilderna uppstår genom att ljud av hög frekvens skickas in i kroppen och signalerna sedan studsar olika mycket beroende på vad för typ av vävnad de går igenom. När man tittar på ultraljudsbilder av hjärtat så är det ofta av intresse att mäta hur mycket blod som hjärtat pumpar ut i kroppen vid varje hjärtslag. Detta kan göras genom att mäta storleken av vänster kammare. Då behöver man först segmentera kammaren vilket vi gör med en metod som kallas aktiva konturer.

När man vill ställa in hastigheten hos en hjärtpump som används av patienter som väntar på hjärttransplantation används också ultraljudsbilder. Pumpen behövs dels för att patienten ska överleva tills det finns ett hjärta som passar men också för att få patienten i bättre form inför operationen. Det är väggen mellan kamrarna, septum, som studeras och buktar den åt endera kammaren så betyder det antingen att hastigheten är för hög eller för låg. Genom att segmentera septum och automatiskt visa ett mått på hur mycket det buktar så fås en mer objektiv bedömning än vad endast en visuell undersökning ger.

Att segmentera organ i CT-bilder (datortomografi) kan vara användbart vid till exempel stråldosplanering då man vill veta hur de är placerade i kroppen för att undvika för mycket strålning. CT ger tredimensionella röntgenbilder där även organ är synliga. Vid en tävling i anslutning till en medicinsk bildanalyskonferens gav våran metod bäst segmentering för 13 av 20 utvalda organ och för mer än hälften av organen så stämde vår segmentering till över 80 % överens med manuella segmenteringar. Metoden som används bygger på att flera manuella segmenteringar matchas mot bilden där organet söks och den sammanvägda informationen ger en segmentering. Denna metod utvecklades senare till att ta mer hänsyn till formen på organen så att denna blir mer trolig. Detta gav lika bra eller bättre segmenteringsresultat.

En speciell typ av artificiella neurala nätverk, dvs. en metod som försöker härma hur nervcellerna i en mänsklig hjärna är kopplade och överför information, har anpassats för att lokalisera och identifiera vilka ryggkotor som syns i en CT-bild. Detta kan användas som information till ovan nämnda metod för organsegmentering då matchningen av manuella segmenteringar inte funkar lika bra för kotor på grund av likheten mellan intilliggande kotor.

Projekten i avhandlingen visar prov på vilka möjligheter det finns att med hjälp av matematiska metoder analysera medicinska bilder av olika slag. En hjälp som sjukvården har stor nytta av.

Preface

This thesis is based on the following six papers. My contributions to each paper is also listed below.

I: M. Landgren, K. Sjöstrand, M. Ohlsson, D. Ståhl, N. C. Overgaard, K. Åström, R. Sixt and L. Edenbrandt, "An Automated System for the Detection and Diagnosis of Kidney Lesions in Children from Scintigraphy Images", published at *Scandinavian Conference on Image Analysis*, Ystad, 2011, Springer.

I was involved in coming up with ideas of methods to use. I implemented the algorithms and conducted all the experiments (except the ANN). KS and I wrote the article together.

II: M. Landgren, N. C. Overgaard and A. Heyden, "Segmentation of the Left Heart Ventricle in Ultrasound Images Using a Region Based Snake", published at *SPIE Medical Imaging*, Orlando, 2013.

I contributed to the discussions about how to solve the problem together with the other authors. I did most of the implementation of the algorithms, conducted all experiments and wrote the paper with some input from NCO and AH.

III: M. Landgren, N. C. Overgaard and A. Heyden, "A Measure of Septum Shape using Shortest Path Segmentation in Echocardiographic Images of LVAD Patients", published at *International Conference on Pattern Recognition*, Stockholm, 2014, IEEE.

The idea to the solution of this problem came from me and AH. I did the implementation of the algorithms and conducted all experiments. I wrote the paper with some input from AH and NCO and a lot of help from Annika Ingvarsson, Skåne University Hospital, with the clinical background.

IV: F. Fejne, M. Landgren, J. Alvén, J. Ulén, J. Fredriksson, V. Larsson, O. Enqvist and F. Kahl, "Multi-Atlas Segmentation Using Robust Feature-Based Registration", will appear in the book *Cloud-Based Benchmarking of Medical Image Analysis*, Springer.

Me and FF assert equal contribution to this paper and joint first authorship. My main focus was the writing of the article, in particular the introduction and method parts. I developed the idea for obtaining corresponding landmarks and wrote the code for this.

V: J. Alvén, F. Kahl, M. Landgren, V. Larsson and J. Ulén, "Shape-Aware Multi-Atlas Segmentation", accepted to *International Conference on Pattern Recognition*, Cancun, 2016, IEEE.

This is a development of Paper IV and the code for corresponding landmarks is used here as well. I was involved in writing the introduction of the article and the part about the corresponding landmarks. I contributed to the ideas that shape ought to be incorporated in some way.

VI: M. Landgren, O. Enqvist, L. Svärm and F. Kahl, "Vertebra Detection and Identification using CNN-Based Detectors and Active Shape", *manuscript*.

I came up with the idea to identify sacral vertebrae and lock these positions in the active shape adaption. Me and FK came up with the idea to also detect rib joints in order to gain information for the active shape adaption. I wrote the code that deals with these two ideas and made some modifications to other parts of the code. I did the manual annotations of the rib joints for the entire dataset. I conducted all the experiments. The article is written by me with help by OE and LS on the technical details about the algorithms.

Acknowledgments

Five long years have quickly come to an end and there are a lot of people that deserve to be acknowledged. First I would like to thank my supervisors; Anders Heyden, Niels Chr. Overgaard and Fredrik Kahl, for discussions, support, comments and guidance through this work. I would also like to thank my co-authors on all papers. Furthermore, I am very grateful to the medical staff at Sahlgrenska University Hospital and Skåne University Hospital who have provided me with images and manual segmentations — your work have made my work possible. The discussions with you about the projects have also been invaluable.

I would like to thank the vision group and all colleagues at the Centre for Mathematics for a nice working environment. Special thanks to Hanna Källén, Kerstin Johnsson and Johan Fredriksson. Hanna, thanks for friendship, support and all discussions about anything and everything. Kerstin, thanks for friendship and support. Johan, who I shared office with for 3.5 years, thanks for the company and for the discussions about courses, research and how to optimize your life.

My friends and family, thank you for giving me valuable breaks from research and for your support. It has been reassuring to know that you always stand by my side. I'm forever grateful for that.

- - -

I came across an old school assignment written when I was around 10 years old. I had answered the question "What do you want to work with when you grow up?" with "Mathematics". Later, my dream was to become a medical doctor. I think I have found the perfect combination of those two dreams.

Contents

Introduction

This thesis deals with the analysis of medical images. This is a very broad field and includes subjects like image enhancement, segmentation, registration, quantification and classification. There have been numerous methods developed and adapted to various medical applications since the first clinical useful form of digital image analysis was developed in the early 1970s [16, 2], when the ejection fraction (the volumetric fraction of blood pumped out of the left ventricle in one heartbeat) was computed from cardiac nuclear medicine images in a semi-automated manner [39]. Still more work can be done and new developments of the imaging modalities, computer capacity etc. give new challenges.

The main contribution to this thesis is contained in the six papers. To prepare for these, this introduction aims to introduce and explain the methods that have been used. The methods have been divided into the subjects registration, segmentation and machine learning. These subjects are not independent of each other; registration can be a part of the segmentation and machine learning methods can be used for segmentation as well as classification. An overview of the papers is also given and the introduction ends with some concluding remarks. To begin with some background about medical imaging modalities is supplied in order to give some insight in their characteristics.

1 Medical Imaging Modalities

There exist a number of different imaging modalities and the most common ones will be presented here. Not only are there a lot of modalities, they can be used in various ways, with different parameters and aids which leads to a vast number of possible examinations to conduct. A more detailed description of the modalities is given in for example [21].

Ultrasound

The sound frequencies used in ultrasound are above 1 MHz. A transducer positioned against the body surface emits the ultrasound and receives the reflected waves. When the waves reach an interface between tissues with different densities they can be scattered, absorbed, reflected or transmitted depending on the circumstances. The characteristics of the waves that return to the transducer form the images. One special characteristic of ultrasound images is the speckle noise. It comes from the reflections back and forth between different tissues and makes the images quite noisy. There are three different presentation modes of ultrasound; A-mode, B-mode and M-mode. A-mode shows the echoes in the direction of the transducer, with amplitude on the y-axis and distance (assuming constant speed of sound) on the x-axis. In B-mode the sound beam is scanned across the patient and a two dimensional image is obtained. Echoes with a high amplitude are represented by high intensity values. M-mode shows each echo-producing interface as a function of time, which makes it possible to study movement of the structure. B-mode is the best known of the presentation modes, see an example image in Figure 1(a). Ultrasound is used e.g. in cardiology, gynecology, obstetrics, urology and neurology.

X-rays

X-rays are electromagnetic radiation that attenuates differently in different materials. There is a high attenuation in bones which means that a smaller amount of x-rays reach the detector behind the bones and they therefore show up clearly on the images. Different soft tissues in the body attenuate an equal amount of x-rays and it is hard to separate e.g. organs from each other. The lungs are visible because of the low attenuation in air compared to the attenuation in tissue. Some examples of examinations are skeletal x-ray, mammography and chest x-ray.

Computed Tomography

Computed tomography (CT) is an imaging technique where the x-ray tube rotates around the body and the rays are detected by a stationary circular array of detectors. The images are reconstructed using measurements of the transmitted x-rays through the body and mathematical models, see Figure 1(b) for an example of a CT image. In this way the body can be viewed in slices from different directions. The contrast resolution is higher in CT

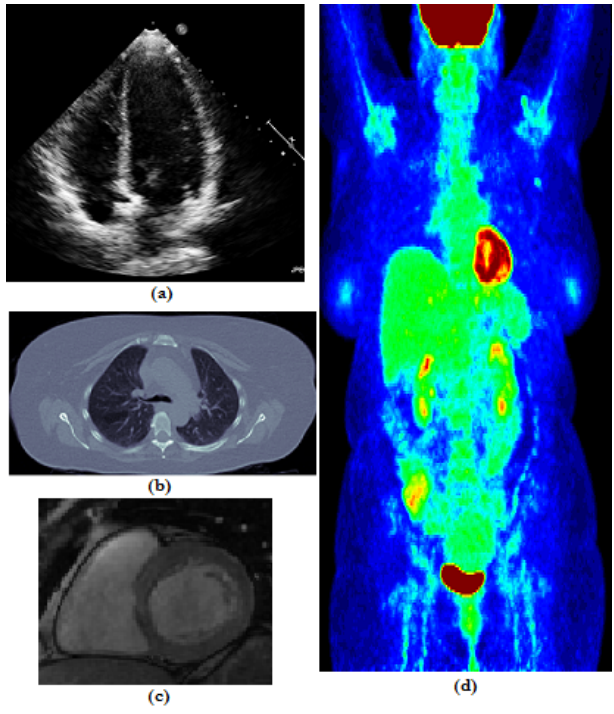

Figure 1: Examples of images from different modalities. (a) Ultrasound image, (b) CT image, (c) MR image [35] and (d) PET image.

images than in ordinary x-ray images which means that different organs can be separated in CT images. CT can be used for detecting tumors, head infarctions and abdominal diseases. Blood vessels can be examined using computed tomography angiography (CTA) where a contrast agent is injected in the blood.

Magnetic Resonance Imaging

In magnetic resonance imaging (MRI) the body is put into a magnetic field. By introducing an electromagnetic field to get a resonance with the protons (mainly in water molecules in the body) a radio frequency signal is emitted that can be detected by a receiver coil. By mathematical models, these signals are transformed into cross-sectional images of the body. Particularly soft tissue is interesting in these images because of the high contrast

resolution. In Figure 1(c) an MR image of the heart can be seen.

Nuclear Imaging

In nuclear imaging, photons from a radioactive substance are detected and then transformed into images. The radioactive substance is attached to a biological molecule that accumulates in the organ/tissue of interest after an injection into the blood. The detector can be a planar gamma camera in which case the imaging technique is called scintigraphy. This gives two dimensional images and some applications are lung, bone and renal scintigraphy. Single photon emission computed tomography (SPECT) also uses a gamma camera but, as with ordinary CT, images from a number of angles are captured and reconstructed into a three dimensional dataset. SPECT can be used for myocardial perfusion imaging and tumor detection. In positron emission tomography (PET) the detected radiation is annihilation radiation. This radiation occurs when positrons, from the injected radioactive substance, and electrons interact and the emitted photons are emitted in opposite directions. If a circular detector simultaneously detects photons in opposite directions it is assumed that the annihilation occurred on a straight line between the detectors. PET can for example be used in oncology and to diagnose Alzheimer's disease. An example of a PET image can be seen in Figure 1(d).

2 Registration

Registration is the task of aligning an image, often called the *target* image, to the same coordinate system as another image, called the *source* image. It can be used for comparing two images of the same patient acquired at different occasions in order to analyze the development of, for example, tumors. Another application is to overlay images from different modalities to gain additional information. PET images give us information if there are tumors present but they give very little anatomical information of their exact location. By acquiring a CT image and register it to the PET image, a better understanding is obtained. This can be a useful tool in radiotherapy planning. Registration between images from different patients can also be done for the purpose of comparing a disease or the appearance of some anatomical structure within the population.

In order to register two images some similarity between the images needs to be found and a technique for performing the actual transformation of one of the

images needs to be chosen. These steps will be explained below.

2.1 Similarity Functions

Registration methods can be divided into either intensity-based or feature-based methods. Intensity-based methods give in general accurate registrations but they may be sensitive to initialization and slow since they must be solved using time-consuming local search. Feature-based methods on the other hand are faster since they lead to relatively simple minimization problems. However, they have a risk of failing if there are too many false (outlier) correspondences between the images.

2.1.1 Intensity-based Methods

In intensity-based methods the similarity is measured using some correlation metric that focus on pixel intensities or colors. One such measure is the normalized mutual information (NMI) measure [40] where the information content is measured. If I_s and I_t are the source and target image, respectively, it can be defined as

$$Y(I_s, I_t) = \frac{H(I_s) + H(I_t)}{H(I_s, I_t)}, \tag{1}$$

where $H(\cdot)$ is the Shannon entropy of the signals,

$$H(I_s) = -\sum_{i=1}^{m} p_i \log p_i, \tag{2}$$

$$H(I_t) = -\sum_{j=1}^{n} q_j \log q_j, \tag{3}$$

$$H(I_s, I_t) = -\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} \log p_{ij}, \tag{4}$$

where p and q are the marginal probabilities of values occurring in the respective images and p_{ij} is the joint probability of values in the region of overlap of the two images. These probabilities are estimated from the images [40].

2.1.2 Feature-based Mehods

The similarity functions for feature-based methods depend on certain features in the images, such as edges, curves, points and surfaces. These features need to be detected in the images and then matched between the images that should be registered.

One popular method for feature detection is the Scale Invariant Feature Transform (SIFT) [32] that selects feature points at the maxima and minima of a difference of Gaussians in a scale-space pyramid. The feature points are described by a feature vector that holds information about image gradients and orientations. Feature vectors from different images are then matched by identifying the nearest neighbor using the Best-bin-first search method which is based on the k-d tree algorithm [4]. Outliers are removed by computing the affine transformation between the model and the detected object in the image.

SURF, Speeded-Up Robust Features [3], is another method for feature detection and description. This method is based on an approximation of the Hessian matrix applied to the integral image,

$$I_{\Sigma}(\mathbf{x}, \mathbf{y}) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j), \tag{5}$$

to detect feature points. The description of a feature point is based on Haar wavelet responses in a region around the point. SURF is both faster and more accurate than SIFT [3].

In the work by Svärm et al. [41] an approach similar to SIFT and SURF have been used for feature detection and matching. They present an algorithm for fast outlier removal that is based on the truncated L_2 loss of the residuals,

$$l(T) = \sum_{j=1}^{n} \min(r_j^2(T), \epsilon^2), \tag{6}$$

where $r_j(T) = |T(x_j) - y_j|$, T is the transformation matrix, x_j and y_j are the feature points in the source and target image, respectively, and ϵ is a threshold on the residuals. This algorithm is followed by a RANSAC [17] step to remove remaining outliers and estimate an accurate transformation.

2.2 Image Transformation

When we know a relationship between the images that should be registered we want to transform the target image to the source image. The transformation methods can be divided into *rigid* and *nonrigid* transformations. The rigid ones are global and transform the entire image in the same way. They include translations, rotations and reflections and the transformation can be written as

$$\mathbf{p}' = \mathbf{R}\mathbf{p} + \mathbf{t}, \tag{7}$$

where $\mathbf{R}$ is an orthogonal matrix and represent the rotation and reflection, $\mathbf{t}$ is the translation vector and $\mathbf{p}$ and $\mathbf{p}'$ are the original and transformed coordinates, respectively. Nonrigid transformations are elastic and can capture local geometrical differences. Examples of rigid and nonrigid transformations can be seen in Figure 2.

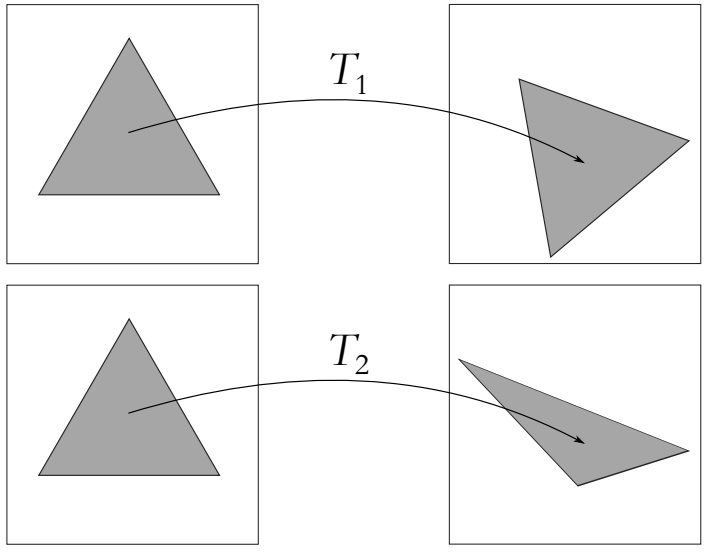

Figure 2: T_1 is an example of a rigid transformation where the triangle has been rotated and translated. T_2 has changed the shape of the triangle and is an example of a nonrigid transformation.

Thin-plate splines (TPS) are a nonrigid transformation method that was introduced by Duchon in 1977 [15] and for medical images by Bookstein in 1989

[5]. The mapping function, f, minimizes the following smooth energy function (in the 2D case, it can be extended to the 3D case),

$$E_{\text{TPS}}(f) = \sum_{i=1}^{N}(\mathbf{p}_{2i} - f(\mathbf{p}_{1i}))^2 + \lambda \iint \left(\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right) dxdy, \quad (8)$$

where $\{\mathbf{p}_{1i}\}_{i=1}^{N}$ and $\{\mathbf{p}_{2i}\}_{i=1}^{N}$ are the two point sets, with N points $\mathbf{p}_i = (x_i, y_i)$, that have been matched and the second part is the physical bending energy of a thin metal plate. According to [43] there exist a unique f that minimizes this energy and it is given by

$$f(x, y) = \beta_0 + \beta_x x + \beta_y y + \sum_{i=1}^{N} \alpha_i \eta(\|(x, y) - \mathbf{p}_i\|), \quad (9)$$

where β are parameters and

$$\eta(r) = \begin{cases} r^2 \log(r^2) & \text{if } r > 0, \\ 0 & \text{if } r = 0. \end{cases} \quad (10)$$

The parameters α must satisfy the following constraint:

$$\sum_{i=1}^{N} \alpha_i = \sum_{i=1}^{N} \alpha_i x_i = \sum_{i=1}^{N} \alpha_i y_i = 0.$$

The TPS deformation is smooth and captures both global and local deviations. The registrations have good accuracy and TPS is a popular method for registration of medical images.

3 Segmentation Methods

A segmentation of an image is a partitioning of the image into two or several regions of interest. For example in a CT image of the abdomen you can segment it into two parts; the liver and the rest of the abdomen, respectively, or you can segment several organs. Segmentations are often represented as a curve, that separates the region of interest from the rest of the image, or by assigning each pixel/voxel a label. This is the main subject of this thesis and several methods that can be used to obtain a segmentation will be presented here.

3.1 Thresholding

Thresholding is a quite simple segmentation method where an image, $I(x, y)$, is divided into two regions, one below the threshold and one above. The thresholded image, $T(x, y)$, can be defined as

$$T(x,y) = \begin{cases} 1, & I(x,y) \geq t \\ 0, & I(x,y) < t \end{cases} \tag{11}$$

where t is the chosen threshold. Using several thresholds the image can of course be segmented into more than two regions. The threshold can be chosen in several ways. One is to look at the histogram of the image. If the pixels corresponding to the object and background respectively, are grouped into two different groups the threshold is chosen as a value that separates the groups. Another way is to examine image intensities in the neighborhood of each pixel and calculate some statistics.

3.2 Active Contour Models: Snakes

Active contour models, or more commonly named snakes, were first introduced by Kass, Witkin and Terzopoulos [27]. A snake is an energy minimizing spline that can be used for finding salient image contours. The models are called snakes because of the way the contour moves while searching for image edges and they are active since the energy is always decreasing during the optimization, which gives the models a dynamic behavior. The snake needs to be initialized close to the contour of interest, but then finds its way to this contour by minimizing the energy.

3.2.1 The Classical Snake Model

A snake is a parametrized curve $\mathbf{u} : [0, 1] \to \mathbb{R}^2$ such that the point of the curve $\mathbf{u}(s) = (x(s), y(s))$ belongs to the image domain for all $0 \leq s \leq 1$. The energy functional to be minimized consists of two parts, one expression for the internal energy and one expression for the external energy,

$$E[\mathbf{u}] = \underbrace{\frac{1}{2}\int_0^1 \alpha|\mathbf{u}'(s)|^2 + \beta|\mathbf{u}''(s)|^2 ds}_{E_{\text{int}}[\mathbf{u}]} - \underbrace{\int_0^1 V(\mathbf{u}(s))ds}_{E_{\text{ext}}[\mathbf{u}]}, \tag{12}$$

where $\mathbf{u}(s)$ is required to be a twice differentiable curve, $V = V(x, y)$ is a potential function determined by the image and the positive parameters α and β are set by the user.

The internal energy controls the flexibility of the snake and it depends only on the contour and not on the image data. The first term in (12), with the first derivative, makes the snake act like a string and the second term, with the second derivative, makes it act like a thin beam.

The function V in the external energy can be expressed in different ways depending on what you are looking for in the image. The simplest function is the image intensity, then the snake will be attracted to lines. Edges can be found in the image by using the gradient of the image intensities as a function. In this thesis, we are interested in snakes that are region based. This means that the external energy is driven by statistical properties of the image data inside and outside the contour, see later subsection for further explanation.

The segmentation is given by the curve $\boldsymbol{\varphi} \in \mathcal{C}^2([0, 1], \mathbb{R}^2)$, which solves the variational problem

$$E[\boldsymbol{\varphi}] = \min_{\mathbf{u}} E[\mathbf{u}], \tag{13}$$

where $\boldsymbol{\varphi}$ and $\mathbf{u}$ are subject to boundary conditions. For instance $\boldsymbol{\varphi}$ and $\mathbf{u}$ could be open curves with fixed end points or closed, simple curves.

3.2.2 Finding the Optimal Solution

From now on we consider the case $\beta = 0$ when the snake acts like a string. The snake energy may be expressed in the general form

$$E[\mathbf{u}] = \int_0^1 L(\mathbf{u}(s), \mathbf{u}'(s))ds, \tag{14}$$

which includes (12) if the Lagrange function L is taken to be $L(\mathbf{u}, \mathbf{u}') = \frac{\alpha}{2}|\mathbf{u}'|^2 - V(\mathbf{u})$. In this section we consider closed contours. That is, we minimize the functional (14) over the following set of admissible curves,

$$\mathcal{A} = \{\mathbf{u} \in \mathcal{C}^1([0, 1], \mathbb{R}^2) : \mathbf{u}(1) = \mathbf{u}(0), \mathbf{u}'(1) = \mathbf{u}'(0)\}.$$

In order to minimize $E[\mathbf{u}]$ we first look at the differential of E defined as

$$dE[\mathbf{u}; \mathbf{v}] = \frac{d}{d\epsilon} E[\mathbf{u} + \epsilon\mathbf{v}]\Big|_{\epsilon=0}, \tag{15}$$

where $\mathbf{u}+\epsilon\mathbf{v}$ is a variation of $\mathbf{u}$ and $\mathbf{u}$, $\mathbf{v} \in \mathcal{A}$. When computing the differential we find that it is

$$\begin{aligned} dE[\mathbf{u};\mathbf{v}] &= \frac{d}{d\epsilon}E[\mathbf{u}+\epsilon\mathbf{v}]\Big|_{\epsilon=0} \\ &= \frac{d}{d\epsilon}\int_0^1 L(\mathbf{u}+\epsilon\mathbf{v},\mathbf{u}'+\epsilon\mathbf{v}')\,ds\Big|_{\epsilon=0} \\ &= \int_0^1 \nabla_{\mathbf{u}}L(\mathbf{u},\mathbf{u}')\cdot\mathbf{v}+\nabla_{\mathbf{u}'}L(\mathbf{u},\mathbf{u}')\cdot\mathbf{v}'ds. \end{aligned} \tag{16}$$

Now assume $\mathbf{u}$ is a twice differentiable curve. Then by using integration by parts we can write the differential as:

$$\begin{aligned} dE[\mathbf{u};\mathbf{v}] &= \int_0^1 \{\nabla_{\mathbf{u}}L(\mathbf{u},\mathbf{u}')-\frac{d}{ds}\nabla_{\mathbf{u}'}L(\mathbf{u},\mathbf{u}')\}\cdot\mathbf{v}\,ds \\ &= \langle\nabla_{\mathbf{u}}L(\mathbf{u},\mathbf{u}')-\frac{d}{ds}\nabla_{\mathbf{u}'}L(\mathbf{u},\mathbf{u}'),\mathbf{v}\rangle_{L^2} \\ &= \langle\nabla E[\mathbf{u}],\mathbf{v}\rangle_{L^2} \end{aligned} \tag{17}$$

where, in the last row we have defined the L^2-gradient of E at $\mathbf{u}$ as

$$\nabla E[\mathbf{u}] = \nabla_{\mathbf{u}}L(\mathbf{u},\mathbf{u}')-\frac{d}{ds}\nabla_{\mathbf{u}'}L(\mathbf{u},\mathbf{u}'), \tag{18}$$

using the definition of an inner product: $\langle\mathbf{u},\mathbf{v}\rangle_{L^2}=\int_0^1\mathbf{u}(s)\cdot\mathbf{v}(s)\,ds$.

If $\boldsymbol{\varphi}$ is a minimizer of $E[\mathbf{u}]$ over $\mathcal{A}$ then a necessary condition on $\boldsymbol{\varphi}$ is Euler's equation,

$$\nabla_{\boldsymbol{\varphi}}L-\frac{d}{ds}\nabla_{\boldsymbol{\varphi}}L=0, \tag{19}$$

where $\nabla_{\boldsymbol{\varphi}}L=\nabla_{\mathbf{u}}L(\boldsymbol{\varphi}(s),\boldsymbol{\varphi}'(s))$ and $\nabla_{\boldsymbol{\varphi}'}L=\nabla_{\mathbf{u}'}L(\boldsymbol{\varphi}(s),\boldsymbol{\varphi}'(s))$. Comparing with (18) we see that Euler's equation can be written concisely as

$$\nabla E[\boldsymbol{\varphi}]=0. \tag{20}$$

Applying our definition of the gradient of E on the Lagrangian for our snake model (12), $L(\mathbf{u},\mathbf{u}')=\frac{\alpha}{2}|\mathbf{u}'|^2-V(\mathbf{u})$, we get

$$\nabla E[\mathbf{u}]=-\nabla V(\mathbf{u})-\alpha\mathbf{u}''. \tag{21}$$

The minimization is done using the method of gradient descent, which evolves a curve $\mathbf{u} = \mathbf{u}(s,t)$ with respect to a fictitious time parameter $t \geq 0$ according to the PDE

$$\frac{\partial}{\partial t}\mathbf{u} = -\nabla E[\mathbf{u}]. \tag{22}$$

Thus we get the gradient descent scheme for the minimization of the snake functional E:

$$\begin{cases} \frac{\partial}{\partial t}\mathbf{u}(s,t) = \nabla V(\mathbf{u}(s,t)) + \alpha \mathbf{u}''(s,t), \\ \mathbf{u}(s,0) = \mathbf{u}_0(s), \end{cases} \tag{23}$$

where $\mathbf{u}_0(s)$ is the initial contour specified by the user.

3.2.3 Region Based Snakes

As mentioned before, in this thesis we will consider snakes that are region based. We consider a closed curve $\mathbf{u} \in \mathcal{A}$ which is simple and we write our snake model as

$$E[\mathbf{u}] = \int_0^1 \frac{\alpha}{2}|\mathbf{u}'(s)|^2 \, ds - \iint_{\Omega(\mathbf{u})} W(x,y) \, dxdy, \tag{24}$$

where $\Omega(\mathbf{u})$ denotes the domain enclosed by $\mathbf{u}$ with $\partial\Omega$ oriented counter clockwise and $W = W(x,y)$ is a function that depends on the image data.

Now we want to derive an expression for the gradient of E in (24), which can be used when minimizing the functional using the gradient descent method as described above. This has been done in [42] by using the divergence theorem of vector calculus. Another derivation, based on the interpretation for parametrized curves of the differential computed directly within the level-set framework, is given in [26]. In our approach we instead consider Green's theorem,

$$\int_{\partial\Omega} P \, dx + Q \, dy = \iint_{\Omega} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dxdy, \tag{25}$$

and assume that we can find a vector field (P,Q) such that $\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} = W$, e.g. by taking $Q = 0$ and $P = -\int_0^y W(x,\hat{y}) \, d\hat{y}$. Our Lagrangian in the energy functional (14) can now be written as

$$L = \frac{\alpha}{2}(x'^2 + y'^2) - P(x,y)x' - Q(x,y)y'. \tag{26}$$

Here we used the fact that we can write our curve integral as

$$\int_{\partial\Omega} P\,dx + Q\,dy = \int_0^1 P(x,y)x' + Q(x,y)y'ds. \tag{27}$$

Now we write (18) component wise , $\nabla E = ((\nabla E)_x, (\nabla E)_y)$, for easier computation when we have our region based term:

$$\begin{aligned}(\nabla E)_x &= \frac{\partial L}{\partial x} - \frac{\partial}{\partial s}\left(\frac{\partial L}{\partial x'}\right) \\ &= -\frac{\partial P}{\partial x}x' - \frac{\partial Q}{\partial x}y' - \alpha x'' + \frac{\partial P}{\partial x}x' + \frac{\partial P}{\partial y}y' \\ &= -\alpha x'' - \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)y' = -\alpha x'' - Wy', \end{aligned} \tag{28}$$

and similarly,

$$\begin{aligned}(\nabla E)_y &= \frac{\partial L}{\partial y} - \frac{\partial}{\partial s}\left(\frac{\partial L}{\partial y'}\right) \\ &= -\frac{\partial P}{\partial y}x' - \frac{\partial Q}{\partial y}y' - \alpha y'' + \frac{\partial Q}{\partial x}x' + \frac{\partial Q}{\partial y}y' \\ &= -\alpha y'' + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)x' = -\alpha y'' + Wx'. \end{aligned} \tag{29}$$

Summarizing this gives the gradient

$$\nabla E[\mathbf{u}(s)] = -\alpha\mathbf{u}''(s) + W(\mathbf{u}(s))\hat{\mathbf{u}}'(s), \tag{30}$$

where $\hat{\mathbf{u}}'(s) = (-y'(s), x'(s))$ is the inward unit normal vector to the curve $\mathbf{u}(s)$ since $\partial\Omega$ is oriented counter clockwise, see Figure 3. The term $W(\mathbf{u}(s))\hat{\mathbf{u}}'(s)$ can be interpreted as a force in the direction of the normal to the curve $\mathbf{u}(s)$.

3.3 Active Shape Models

Active shape models (ASMs) are statistical models of shape and it was first described by Cootes et al. in 1995 [11]. It can be applied to segmentation but also for other configurations where one want to ensure a plausible shape. The idea is that the contour should only be allowed to deform into shapes that are characteristic of the object of interest. To model the characteristic shapes of an object,

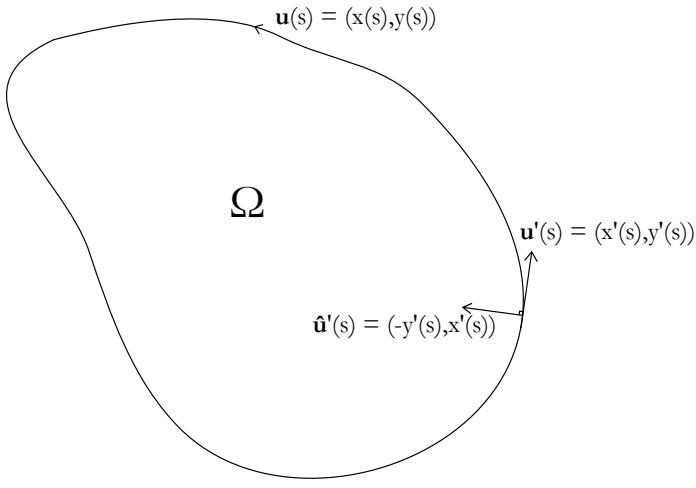

Figure 3: Interpretation of the terms in (30).

a training set is needed. The contours are represented by n landmarks and the ith contour is written $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in}, y_{i1}, y_{i2}, \ldots, y_{in})^T$. These points should be annotated in the same way for all shapes in the training set and then be aligned with for example Procrustes analysis [14]. By looking at the statistics of each point cloud, a Point Distribution Model can be derived. Each point cloud is modelled as a Gaussian distribution. The mean shape of N aligned shapes is given by

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_i. \tag{31}$$

The major axes are found by Principal Component Analysis (PCA) of the data. The eigenvalues are calculated together with the corresponding eigenvectors of the covariance matrix,

$$\Sigma = \frac{1}{N-1} \sum_{i=1}^{N} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T. \tag{32}$$

If the eigenvalues and eigenvectors are ordered as $\lambda_j \geq \lambda_{j+1}$, then the first eigenvector represents the first principal axis, i.e. the direction where the points move the most. The eigenvalues show how much of the variance that is described by each principal axis. In most cases it is not necessary to use all principal axes in the model. It is enough to use the k first principal axes that describe a sufficiently large proportion of the total variance. Each shape in the training set can be written as

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{V}\mathbf{s}, \tag{33}$$

where $\mathbf{V}$ is a matrix with the k first principal axes as columns and $\mathbf{s}$ are a k-dimensional vector containing the weights. The weights can be computed as

$$\mathbf{s} = \mathbf{V}^T(\mathbf{x} - \bar{\mathbf{x}}) \tag{34}$$

and by varying the elements of $\mathbf{s}$ we can vary the shape $\mathbf{x}$ using (33). It can be shown that each eigenvalue of Σ is the variance of the corresponding weight, s_j, and by restricting each weight to the interval $[-3\sqrt{\lambda_j},\ 3\sqrt{\lambda_j}\,]$ it is ensured that the shape is similar to those in the training set, since most of a normal population lies within three standard deviations from the mean.

3.3.1 Applying ASMs for Segmentation

When using ASMs the object of interest is found in an iterative process. First, each point of the contour is moved to a position in its vicinity that better represents the object border. For example, if the border is an edge one can search for the strongest edge along the normal in each point. The second step is to ensure that the current shape is allowed. This is done by computing the weights, $\mathbf{s}$, using (34) and restrict each s_j to be in the interval $[-3\sqrt{\lambda_j},\ 3\sqrt{\lambda_j}\,]$ as mentioned above. These two steps are repeated until there is a sufficiently low difference between two contour representations. In Figure 4 an example of how the contour adapts to the allowed kidney shape can be seen. The red contour represents the step where the object border is found and then, by the method described above, we obtain the cyan contour which is a more plausible shape of a kidney.

3.4 Graph Methods

Images can be represented by graphs and there exist several methods that can solve a segmentation problem using this representation. In this section two such methods will be presented after a general presentation of graphs.

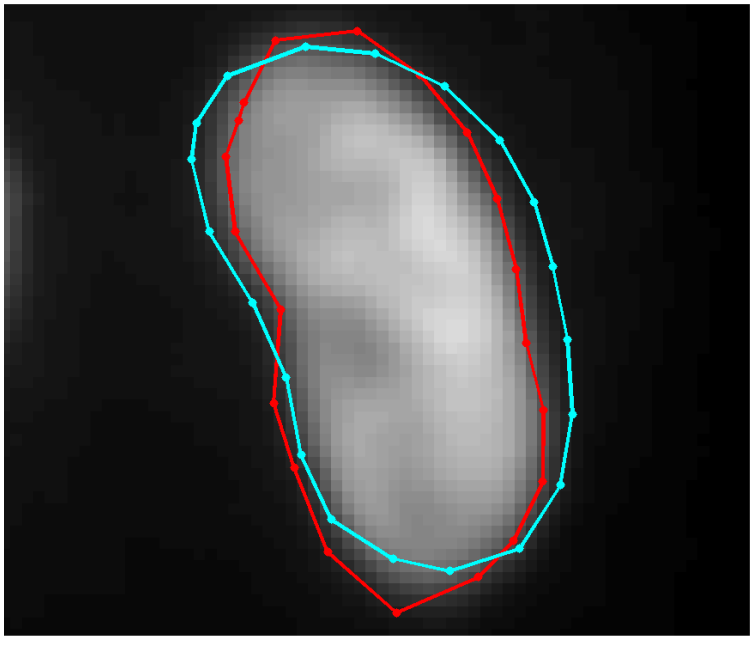

Figure 4: An example of how active shape models work in an early stage of the iterations. The red contour is the best fit to the object boundary and the cyan contour is the adaptation of the red contour to a more plausible shape of a kidney.

A graph is defined as $G = (V, E)$, where each vertex, $v \in V$, corresponds to a pixel in the image. Neighboring pixels will be adjacent in the graph with an edge, $e \in E$, connecting them. A neighborhood of a pixel can be defined in different ways; e.g. 4-connected and 8-connected in the 2D case and 6-connected in the 3D case, cf. Figure 5. Each edge can have a weight, w, which may be derived from e.g. the intensity difference between the neighboring pixels. The graph can be directed or undirected, i.e. the edges can point in both directions or be separated with different weights in different directions (or not exist in one direction).

3.4.1 Shortest Path Problem

The shortest path problem arises in many applications, e.g. finding the shortest route from one city to another, but segmentation can also be formulated as a

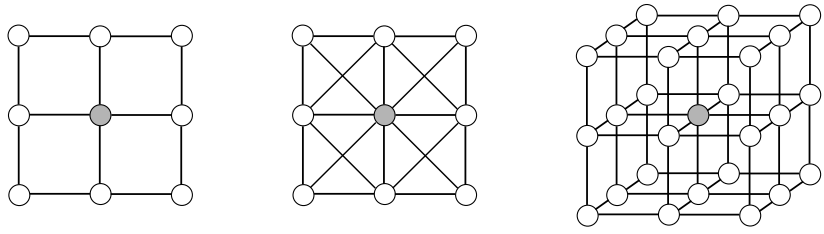

Figure 5: Graph representations of a 3×3 matrix with a 4-connected (*left*) and a 8-connected (*middle*) neighborhood, respectively, as well as a 6-connected neighborhood for a $3 \times 3 \times 3$ matrix (*right*).

shortest path problem. In this case the image is represented by a directed graph. Each edge has a weight, w, which in the case of finding the shortest route between cities corresponds to distances between intersections. The objective in the shortest path problem is to minimize the weight of a path, summarized as

$$\min_p w(p) = \sum_{i=1}^{n} w(v_{i-1}, v_i), \tag{35}$$

for a path $p = (v_0, v_1, \ldots, v_n)$, which is an ordered sequence of vertices where each vertex is adjacent to the previous one. Many of the algorithms that solve the shortest path problem rely on the property that each shortest path contains several shortest paths within it. This can be stated as in the following lemma:

Lemma. *Given a weighted, directed graph $G = (V, E)$ with the weight function $w : E \to \mathbb{R}$, let $p = (v_0, v_1, \ldots, v_n)$ be a shortest path from vertex v_0 to vertex v_n and, for any i and j such that $0 \leq i \leq j \leq n$, let $p_{ij} = (v_i, v_{i+1}, \ldots, v_j)$ be the subpath of p from vertex v_i to vertex v_j. Then p_{ij} is a shortest path from v_i to v_j.*

Proof. The path $p = (v_0, v_1, \ldots, v_n)$ can be decomposed into the subpaths $p_{0i} = (v_0, v_1, \ldots, v_i)$, $p_{ij} = (v_i, v_{i+1}, \ldots, v_j)$ and $p_{jn} = (v_j, v_{j+1}, \ldots, v_n)$ which give $w(p) = w(p_{0i}) + w(p_{ij}) + w(p_{jn})$. Now assume that there is a path p'_{ij} from v_i to v_j with $w(p'_{ij}) < w(p_{ij})$ then the path $p_{0i} \to p'_{ij} \to p_{jn}$ is a path from v_0 to v_n whose total weight, $w(p_{0i}) + w(p'_{ij}) + w(p_{jn})$, is less than $w(p)$, which contradicts that p is the shortest path from v_0 to v_n. □

Algorithms
There are some variants of the shortest path problem.

- Single-source shortest path problems, where one wants to find the shortest path from a source vertex $s \in V$ to all other vertices in V.
- Single-destination shortest path problems, here one wants to find the shortest path from each vertex in V to a destination vertex t.
- Single-pair shortest path problems, find the shortest path from vertex u to vertex v.
- All-pairs shortest path problems, find the shortest path from vertex u to vertex v for every pair of vertices u and v.

There is a number of algorithms that solve these different variants. Some of them are the Bellman-Ford algorithm, Dijkstra's algorithm, the Floyd-Warshall algorithm and Johnson's algorithm which are described in e.g. [12]. Next we will describe Dijkstra's algorithm in detail.

Dijkstra's algorithm [13] can be used on a weighted, directed graph with non-negative weights. It solves the single-source variant of the shortest path problem. To describe the algorithm we introduce two attributes of each vertex v; $v.d$ which is an upper bound on the weight from the source s to v (a shortest path estimate) and $v.\pi$ which is assigned either NIL (no vertex) or another vertex and it is the predecessor of the current vertex v. In Algorithm 1 the outline of the algorithm can be seen. The first for loop initializes the attributes d and π for each vertex and then the sets S (visited vertices) and Q (unvisited vertices) are initialized. In the while loop the algorithm picks out the closest vertex u and updates the adjacent vertices' ($G.Adj[u]$) attributes if these edges make their shortest path estimate smaller. This will give the shortest path from s to each vertex. An example of execution of Dijkstra's algorithm can be seen in Figure 6.

Algorithm 1 Dijkstra's algorithm

for each vertex $v \in G.V$ **do**
  $v.d = \infty$
  $v.\pi = \text{NIL}$
end for
$s.d = 0$
$S = \emptyset$
$Q = G.V$
while $Q \neq \emptyset$ **do**
  $u =$ the vertex v in Q with the lowest shortest path estimate
  $S = S \cup \{u\}$
  for each vertex $v \in G.Adj[u]$ **do**
    if $v.d > u.d + w(u, v)$ **then**
      $v.d = u.d + w(u, v)$
      $v.\pi = u$
    end if
  end for
  $Q = V \setminus S$
end while

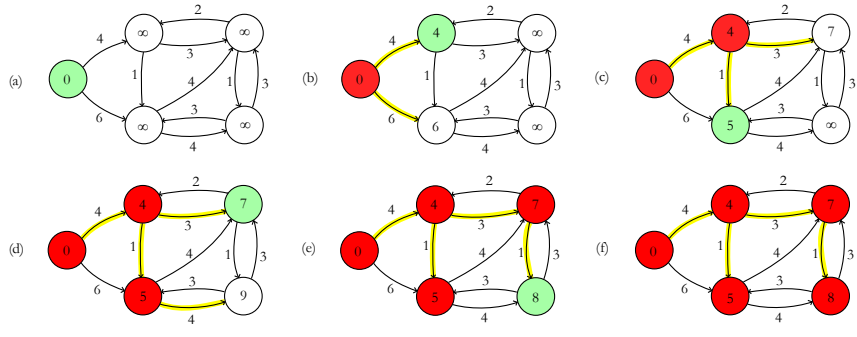

Figure 6: Execution of Djikstra's algorithm where the beginning of each iteration of the while loop in Algorithm 1 is represented by (a) to (f). The source vertex is the left most vertex. The green shaded vertices correspond to u in the algorithm and the red vertices are in the set S. The number within each vertex is the shortest path estimate and the edges marked with yellow are the predecessor values.

3.4.2 Graph Cuts

Using graph cuts we minimize an energy related to the segmentation problem, [7, 28]. The energy can be written as

$$E(L) = \sum_{u \in \mathcal{P}} D(L_u) + \sum_{u,v \in \mathcal{N}} V_{u,v}(L_u, L_v), \tag{36}$$

where $L = \{L_u | u \in \mathcal{P}\}$ is a labeling of image $\mathcal{P}$, D is a data term, V is a regularization term and $\mathcal{N}$ is the set of all pairs of neighboring pixels.

In graph cuts we also consider a directed graph. We define two terminal nodes called the *source*, s, and the *sink*, t, which in the segmentation problem refer to examples of the background and the foreground respectively. The source vertex has only outgoing edges and the sink has only incoming edges. Such a graph, that also have a capacity function, $c : E \to \mathbb{R}^+$, which map how much each edge can carry, is called a *network*. We want to find the cut of the graph that partitions the graph into one part containing the source and one part containing the sink with minimal cost, where the cost is defined as the sum of the weights of the boundary edges (cf. the data term in (36)). An example of the structure of a graph for imaging problems can be seen in Figure 7. The Ford-Fulkerson theorem [18] says that the minimal cut problem can be solved by computing the maximal flow from the source to the sink:

Theorem. *For any network the maximal flow value from s to t is equal to the minimal cut capacity of all cuts separating s and t.*

The *flow* in a network is given by a mapping $f : E \to \mathbb{R}^+$ such that

- $f(u, v) \leq c(u, v)\ \forall e_{u,v} \in E$ (capacity constraint)
- $\sum_{v \in \mathcal{N}(u)} f(u, v) = \sum_{v \in \mathcal{N}(u)} f(v, u)\ \forall u \in V \setminus \{s, t\}$ (flow conservation).

The minimal cut is given by a set of edges that are saturated by the maximal flow. This problem can be solved with different algorithms, see a comparison in [6].

3.5 Atlas Segmentation

An image with a manual segmentation of e.g. an organ can be called an *atlas*. By performing a registration, see Section 2, between the atlas and an image where we

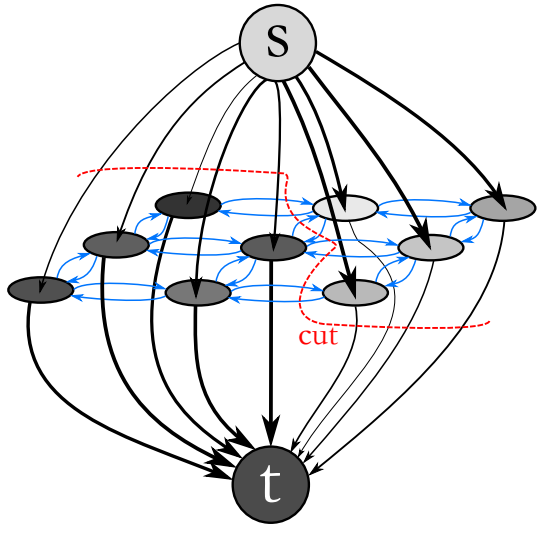

Figure 7: Basic graph structure used in graph cuts. The minimal cut of the graph is also shown.

want to segment the particular organ (the source image), the manual segmentation can be transformed to the source image and a segmentation can be obtained. This was the approach in the early works about atlas segmentation [34]. But the variation of the appearance of an organ within the population is not captured by using a single atlas. Early works that used several atlases created a probabilistic model of the atlases but in the approach normally referred to as *multi-atlas* segmentation the atlases are not summarized into a model [25].

Typically all atlas images are available in multi-atlas segmentation and pairwise registrations between each atlas image and the source image are computed. The labels from the manual segmentations are then propagated to the coordinate system of the source image according to this registration. There are different approaches to the fusion of the labels from several atlases. The simplest one is majority voting, i.e. for each pixel/voxel the label that most atlases assign to it is chosen. Other methods are weighted voting, where the weights may reflect the similarity between the atlas and source image, and probabilistic models.

4 Machine Learning

Machine learning methods are a group of methods where the computer learn to see patterns and draw conclusions from provided data by forming a model on its own. One can talk about three ways of learning; supervised, unsupervised and reinforcement learning, respectively. In supervised learning both input signals and the desired output are known whereas in unsupervised learning some cost function, that depends on the input signals and the output from the networks, is minimized. Reinforcement learning is based on interactions between the algorithm and the environment and the algorithm has to learn which actions give positive feedback. There exist several machine learning methods and three of them; discriminant analysis, artificial neural networks and random forests, will be described in the following sections.

4.1 Discriminant Analysis

Discriminant analysis is a method that can be used for classification. We model each class density as Gaussian,

$$f_j(\boldsymbol{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}_j|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)}, \tag{37}$$

where $\boldsymbol{x}$ is the feature vector of length p, $\boldsymbol{\mu}_j$ is the expected value for class j and $\boldsymbol{\Sigma}_j$ is the covariance matrix for the feature vector for class j. As will be shown below we get Linear Discriminant Analysis (LDA), a classifier that separates the classes with hyperplanes, if we assume that the classes have the same covariance matrix, $\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}$. The probability of class $C = j$ given the feature vector $\boldsymbol{x}$ can be expressed, using Bayes theorem, as

$$P(C = j \mid X = \boldsymbol{x}) = \frac{f_j(\boldsymbol{x})\pi_j}{\sum_{k=1}^{K} f_k(\boldsymbol{x})\pi_k}, \tag{38}$$

where π_j is the prior probability of class j. We can compare two classes by looking at the following log-ratio

$$\begin{aligned}\log \frac{P(C=j \mid X=\boldsymbol{x})}{P(C=k \mid X=\boldsymbol{x})} &= \log \frac{f_j(\boldsymbol{x})}{f_k(\boldsymbol{x})} + \log \frac{\pi_j}{\pi_k} \\ &= \log \frac{\pi_j}{\pi_k} - \frac{1}{2}(\boldsymbol{\mu}_j + \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_j - \boldsymbol{\mu}_k) \\ &\quad + \boldsymbol{x}^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_j - \boldsymbol{\mu}_k), \end{aligned} \tag{39}$$

which leads to a linear function in $\boldsymbol{x}$ since the quadratic terms cancels due to the same covariance matrix for all classes. This means that if the value of the log-ratio is larger than zero then class j is more probable, if it is smaller than zero then class k is more probable and if the classes have equal probability the value is zero. If we introduce the so-called linear discriminant function,

$$\delta_j(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j - \frac{1}{2} \boldsymbol{\mu}_j^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_j + \log \pi_j, \tag{40}$$

then (39) equals $\delta_j(\boldsymbol{x}) - \delta_k(\boldsymbol{x})$ and we see that an observation with feature vector $\boldsymbol{x}$ is assigned to the class for which the value of the discriminant function is the largest. If we have more than two classes then

$$C(\boldsymbol{x}) = \arg\max_j \delta_j(\boldsymbol{x}) \tag{41}$$

can be used as the decision function.

If we now assume that the classes have different covariance matrices then we get Quadratic Discriminant Analysis (QDA) and the discriminant functions become quadratic,

$$\delta_j(\boldsymbol{x}) = -\frac{1}{2} \log|\boldsymbol{\Sigma}_j| - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_j) + \log \pi_j. \tag{42}$$

Here (41) is also used in order to take a decision. In both cases the parameters of the Gaussian distribution and the prior probabilities are in general not known and need to be estimated from the training data.

4.2 Artificial Neural Networks

Artificial neural networks (ANNs) are a method that is designed to mimic the neurons of the human brain and their information processing. A biological neuron consists of a cell body, dendrites that receive signals from other neurons and

an axon that emits chemical substances from synapses which transmit the signal to following neurons. A model of a neuron can be seen in Figure 8, where we have input signals, x_j, that is weighted by synaptic weights, w_{ji}, and summed together before an activation function, $\phi(\cdot)$, which limits the amplitude of the output of the neuron. Mathematically the output of the model, y_j, can be written as

$$y_j = \phi(v_j) = \phi(b_j + \sum_{i=1}^{n} w_{ji} x_i), \tag{43}$$

where b_j is an externally applied bias.

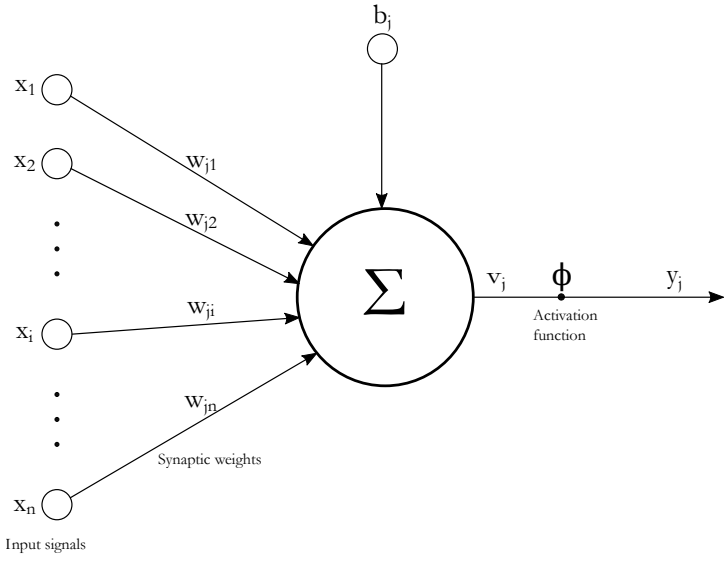

Figure 8: A model of neuron.

McCulloch and Pitts [33] were the first ones to propose a computational model for neural networks in 1943. In the McColluch-Pitts model of a neuron the activation function is a threshold function and hence the output of such a neuron is

$$y_j = \begin{cases} 1, \text{ if } v_j \geq 0 \\ 0, \text{ otherwise.} \end{cases} \tag{44}$$

The first algorithm for supervised learning using a neural network was the Rosenblatt's perceptron in 1958 [36], which is based on McColluch-Pitts model of a

neuron. It can classify patterns that are linearly separable by adjusting the synaptic weights and the bias. In 1974 Paul Werbos first introduced that the back-propagation algorithm could be used in ANNs [44]. But it was not until the mid-1980s that the use of ANNs became more popular when the back-propagation algorithm developed and made the training of multi-layer perceptrons (deeper networks) computationally efficient, see e.g. [37]. The back-propagation algorithm was used by LeCun et al. in 1989, [30], to train a special type of ANN called convolutional neural network (CNN) and in 1998 LeCun et al. showed that such a CNN outperforms all other machine learning methods in a standard handwritten digit recognition task [31]. CNNs have been increasingly popular during recent years due to increased computer power that enables training of deep networks.

4.2.1 Architectures and Learning Process

There are different kinds of network architectures. We have single-layer feedforward networks where the input layer projects directly onto the output layer. In multi-layer feedforward networks we have one or more hidden layers and the neurons detect features in the input data. If a network has at least one feedback loop it is called a recurrent network. Such networks may also be single-layered or contain hidden layers. If every node in each layer of a network is connected to every node in the next layer we say that the network is *fully connected*, otherwise it is *partially connected*. In Figure 9 examples of the different network architectures are given and from here on we will focus on feedforward networks.

Each neuron in a network has a differentiable non-linear activation function such as the sigmoid function,

$$\phi(v) = \frac{1}{1 + e^{-\alpha v}}, \tag{45}$$

where α is a positive slope parameter, or the hyperbolic tangent function,

$$\phi(v) = \beta \tanh(\gamma v), \tag{46}$$

where β and γ are positive constants.

We distinguish between two types of signals in the network; function signals that propagate forward through the network and error signals that propagate backward from the output neurons through the network and adjust the synaptic weights. The objective when training an ANN is to find the synaptic weights,

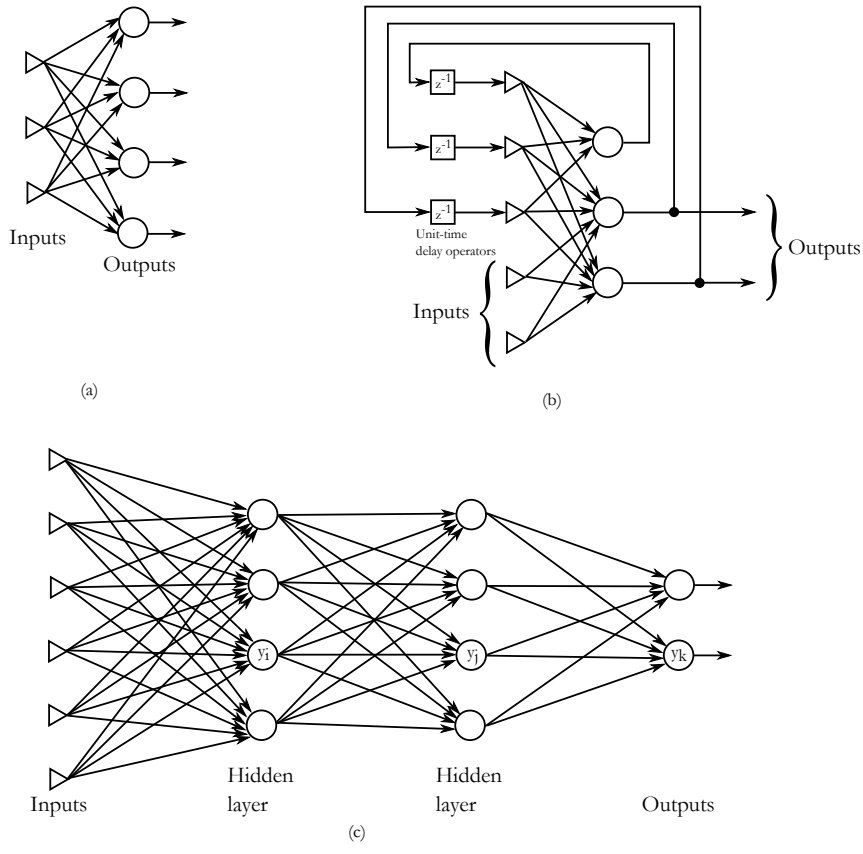

Figure 9: Examples of different network architectures; (a) a single-layer feedforward network, (b) a recurrent network and (c) a fully connected multi-layer feedforward network (y is the output of the neuron).

here collected in a vector $\boldsymbol{w}$, that minimize the total error energy from all output neurons in the network,

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} E(\boldsymbol{w}). \tag{47}$$

This error energy can be written as

$$E = \frac{1}{2}\sum_{j\in\mathcal{A}} e_j^2 = \frac{1}{2}\sum_{j\in\mathcal{A}}(d_j - y_j)^2, \tag{48}$$

where d_j is the desired response of output j, y_j is the function signal at the output of neuron j (which, of course, depends on the synaptic weights $\boldsymbol{w}$) and $\mathcal{A}$ is the set of output neurons. This energy is minimized using the back-propagation algorithm which is derived in for example [20] where also more detailed information about ANNs can be found. The corrections of the weights, obtained from the derivation, are computed as

$$\Delta w_{ji} = \eta\delta_j y_i, \tag{49}$$

where η is the learning-rate parameter, y_i is the output from neuron i in the prior layer and δ_j is the local gradient defined as

$$\delta_j = \begin{cases} e_j\phi_j'(v_j), & \text{if neuron } j \text{ is an ouput neuron,} \\ \phi_j'(v_j)\sum_{k\in\mathcal{L}}\delta_k w_{kj}, & \text{if neuron } j \text{ is hidden,} \end{cases} \tag{50}$$

where $\mathcal{L}$ is the set of neurons in the next layer (cf. Figure 9(c) for some assistance with the notation). The learning rate parameter can either be set to a low value, which will yield a smooth change of the weights but at the expense of slow learning, or it can be set to a higher value, which will speed up the learning rate but may lead to an unstable network.

4.2.2 Convolutional Neural Networks

In pattern recognition tasks within image analysis, the size of ordinary ANNs quickly becomes very large. Another problem is that they are very sensitive to local distortions or translation of the input. Then a CNN provides a different structure that reduces the size and gain some invariance to distortions and translations. It is inspired by the work of Hubel and Wiesel in 1962 [24] about the visual cortex and the way different cells detect different levels of features.

The main part of a CNN is, of course, the convolutional layer which consists of several feature maps that extract different features. Each feature map has neurons with weights, i.e. the kernel in a convolution, that are the same within each feature map but different between feature maps. The neurons obtain input from

small regions of neighboring neurons, *receptive fields*, in the previous layer. This means that the neurons detect the same kind of feature at different locations in the image, i.e., it reduces the sensitivity to translations of the input image.

The convolutional layer is followed by a layer with a non-linear activation function. The common choice in CNNs is the Rectifier Linear Unit (ReLU),

$$\phi(x) = \max(0, x), \tag{51}$$

since it makes the training faster and more efficient compared to other activation functions [29].

The invariance to shifts and distortions is achieved by a subsampling layer which reduces the spatial resolution. Historically a subsampling layer with an averaging operation was used but nowadays max pooling is used and it has been shown that it outperforms the former one in capturing invariances [38]. In a max pooling layer each neuron has a receptive field of e.g. 2×2 pixels in the previous feature map and the maximum value in this region is computed. A new feature map with a smaller resolution is constructed from theses maximum values.

A CNN usually consists of several repetitions of the layers described above, as shown in an architecture example in Figure 10. The final layer of a CNN is a fully connected layer and works like the ones in an ordinary ANN. In the case of classification the output of the CNN is usually converted to probabilities, p_k, using the softmax function,

$$p_k = \frac{e^{y_k}}{\sum_{i=1}^{n} e^{y_i}}, \tag{52}$$

where y_k is the output of output neuron k and n is the number of output neurons. The weights in the convolutional layers are learnt through back-propagation and since the neurons in a layer share the same weights the number of free parameters is reduced.

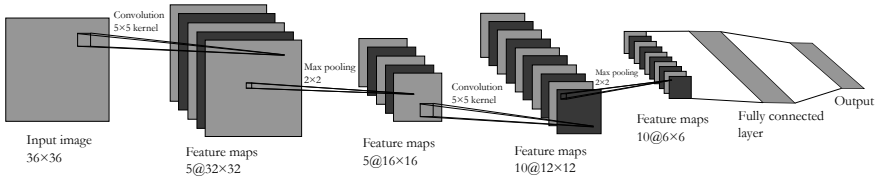

Figure 10: An example of a CNN architecture. In this figure the layers with the activation function are included in the convolutional layers.

4.3 Random Forests

Random forests are a method for regression and classification based on decision trees. The method that is referred to as random forests today was introduced by Breiman in 2001 [9] but decision trees have been used since the early 1980s when Breiman et al. (1984) [10] described its basics. The use of multiple decision trees and random feature selection at each node was introduced and developed by Amit & German [1] and Ho [22, 23] during the 1990s and influenced the work about random forests by Breiman.

4.3.1 Decision Trees

As mentioned above, random forests consist of decision trees. A tree is a graph structure where each node only has one incoming edge and, in the case of binary trees (which we focus on here), two outgoing edges, see Figure 11. There can not exist any loops in a tree. The tree starts with a root node, the following nodes are called internal nodes and the final nodes are called leaf nodes.

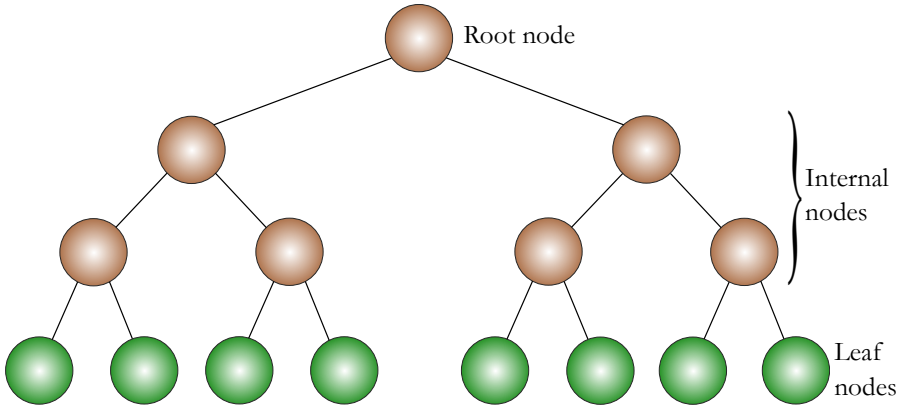

Figure 11: A tree structure.

A decision tree has questions at the root node and each internal node that will lead to a decision in the leaf nodes of, for example, what we see in an image. In Figure 12 a simple example of a decision tree can be seen where we want to decide which ball game someone is playing. In the example it is easy to come up with a number of questions that can solve the problem but for a larger problem this

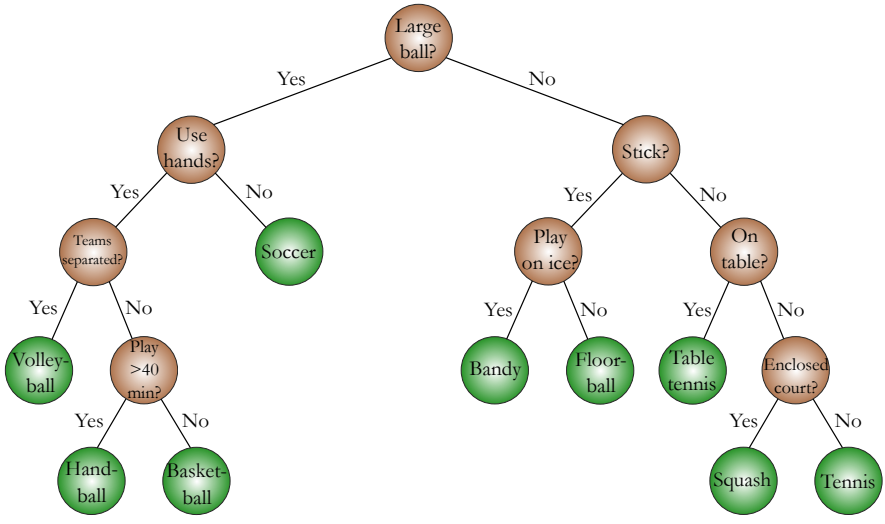

Figure 12: An example of a decision tree where we want to find out which ball game someone is playing. Note that the leaf nodes can be on different levels in the tree.

becomes more difficult and therefore the questions (splitting functions) and the tree structure are learnt automatically from training data.

Normally you do not send in an entire image into the tree, instead you have a feature vector, $\mathbf{f} = (x_1, x_2, ..., x_d)$, for each pixel where the x_i:s could represent e.g. the pixel value and the coordinates of the pixel. When growing the tree the task is to find the splitting criteria, i.e. how to decide which part of the training data that should go down to either child of the node. There exist different criteria but the usual one for classification trees is based on the Gini index,

$$G(n) = \sum_{k \neq k'} p_{nk} p_{nk'}, \tag{53}$$

where p_{nk} is the proportion of class k in node n, i.e.,

$$p_{nk} = \frac{N_k(n)}{N(n)}, \tag{54}$$

where $N_k(n)$ and $N(n)$ are the number of examples of class k in node n and the total number of samples in node n, respectively. The measure of impurity that

should be minimized is then the weighted sum of the child nodes' Gini index:

$$I = G(n_{cl})N(n_{cl}) + G(n_{cr})N(n_{cr}), \tag{55}$$

where n_{cl} and n_{cr} are the children of node n.

Another issue when growing the tree is how large the tree should grow. A too small tree might not capture the information we want and a too large tree might be overfitted. If we only have samples from one class in a node, then clearly we are done and this should be a leaf node. Otherwise there are different stopping criteria that could be used, such as a minimum number of samples in a node, a maximum depth of the tree or examine if the attributes in the particular node is similar to each other. One could also grow a large tree and prune it after training. This means that parts of the tree that do not improve the performance of the final classification to some large extent are removed.

4.3.2 Random Forests

In random forests a large collection of decision trees are used in order to reduce the variance. These trees are trained on different datasets drawn with replacement from the training set (bootstrap samples) and the predictions from all trees are averaged. This method is called bagging and it has been shown that it increases the accuracy [8]. Random forests also use out-of-bag (OOB) samples which means that for each observation, its random forest predictor is constructed by averaging only those trees where this sample was not included in the bootstrap sample.

The correlation between the trees in random forest is reduced by randomly select a subset of features in each node when growing the tree. This reduces the variance further. For classification the number of features chosen at each node is $\sqrt{d}$, where d is the number of available features. Algorithm 2 shows how a random forest is created. In Hastie et al. [19] decision trees and random forests are described and analyzed further.

5 Overview of the Papers

The papers, which are the main contribution to this thesis, are presented here with a short overview of the problems we solved and which methods that have been used. They show that mathematical methods can be applied with great potential on several different problems and modalities.

Algorithm 2 Random forest

1: **for** $b = 1$ to B **do**
2: Draw a bootstrap sample from the training data
3: Grow a tree, T_b, by:
4: **while** $\exists$ leaf nodes with $N > N_{min}$ **do**
5: i) Randomly select m features among the d available
6: ii) Pick the best split point
7: iii) Split the node into two child nodes
8: **end while**
9: **end for**
10: Output the group of trees $\{T_b\}$

Paper I — An Automated System for the Detection and Diagnosis of Kidney Lesions in Children from Scintigraphy Images

This paper presents a complete automated system for detection and diagnosis of kidney lesions in scintigraphy images. These lesions occur among children that have had repeated urinary tract infections. The healthy parts of the kidneys are shown as regions with high intensity values and lesions are often seen as wedge-shaped regions along the border of the kidney. We want to do a segmentation of the entire kidney and therefore active shape models are used that takes the characteristic shape of a kidney into account. The uptake of the injected substance is compared to the typical uptake in a healthy kidney. This is done by creating a statistical map of normal uptake using several healthy examples which uptake is transformed to a common coordinate system using thin-plate splines. Potential lesions are obtained by localizing regions with a lowered uptake. A number of features about the potential lesions and the kidney are gathered and used in order to classify them as healthy or unhealthy. Some examples of these features are the separate function of the kidney and the potential lesion position, size, shape and amount of deviation from the normal uptake. Three different classifiers are trained; linear and quadratic discriminant analysis (LDA and QDA) and artificial neural networks (ANN). We show that LDA has the best performance with a misclassification rate just over 14 % when the sensitivity is fixed to 96.5 % in order to lessen the risk of classifying an actual lesion as normal.

Paper II — Segmentation of the Left Heart Ventricle in Ultrasound Images Using a Region Based Snake

In this paper we segment the left heart ventricle in ultrasound images. This can be used e.g. for measuring the ejection fraction which is a measure of the left ventricular function. The segmentation method used here is a region based snake which takes the image data both inside and outside the contour into consideration when it evolves. There is a big risk that the contour gets stuck on the cardiac valve, that opens and closes during the heart cycle, and therefore we introduce an anchor point on each side of the cardiac valve. The snake is attached to these anchor points and across the cardiac valve a straight line is drawn as the segmentation boundary. The anchor points need to be initialized in the first frame but then a tracking algorithm follow them through the whole cycle to enable a segmentation in each frame. The tracking algorithm consists of three parts that include image resemblance, the distance to the anchor point in the previous frame and a prior model of the movement. By a visual assessment we can conclude that the final segmentation result is satisfying.

Paper III — A Measure of Septum Shape using Shortest Path Segmentation in Echocardiographic Images of LVAD Patients

This paper presents a system that can be used as decision support for physicians when tuning the speed of a mechanical heart pump in patients. The pump is implanted in patients with a failing heart in wait for a heart transplantation. Ultrasound images of the heart, echocardiographic images, are one of the aids the physician has when tuning the speed of the pump. By visually inspecting if the wall between the ventricles, the septum, bulges towards either of the ventricles, one can determine how the speed should be adjusted. We segment the septum in a semi-automated fashion and compute a measure of how it bulges that ought to be more objective than the visual inspection. The segmentation is done using a shortest path algorithm where each pixel in the image is seen as a vertex in a graph. The weight on the edges is based on the horizontal derivative of the image intensities since we want to capture the borders of the septum. As an initialization the user has to annotate the basal and apical part of the septum, respectively. These anchor points are then tracked through the cardiac cycle and a segmentation is obtained in each frame. By drawing a straight line between the anchor points we divide the septum into two parts and by computing the area of each

part, the septum measure is obtained as the area difference divided by the total area. If the septum bulges towards the right ventricle we get a positive value, if it bulges towards the left ventricle we get a negative value and if the speed is ok the value should be close to zero. The algorithm is tested on 38 examinations and it is shown that our septum measure can be a quite good indicator on how the speed of the pump should be adjusted.

Paper IV — Multi-Atlas Segmentation Using Robust Feature-Based Registration

The task in this paper is to segment 20 different organs in whole-body CT images and we solve it using a multi-atlas approach. This requires pairwise registrations between each atlas image and the target image. We determine organ specific features that are used when obtaining these registrations. The organ labels are transferred from the atlas images and fused into a probability for each voxel. This estimate is refined using a random forest classifier in order to also include the local appearance around the organ, which is done by including the image intensities among the features. Then the final segmentation is determined using graph cuts which takes neighboring voxels into account and further improves the segmentation. The approach was evaluated in the VISCERAL Grand Anatomy Challenge where we had the best performance in 13 out of 20 organs, measured with the Dice score. We have also evaluated the organ specific features for some of the organs and show that in most cases the segmentation is improved using these features compared to ordinary features.

Paper V — Shape-Aware Multi-Atlas Segmentation

This paper is a development of Paper IV. A multi-atlas approach is used here as well but in order to achieve more plausible segmentations in terms of shape we incorporate a regularization for this. Instead of fusing the organ labels for each voxel we transfer the landmarks of an organ after the registration. By computing the robust average of the transformed landmarks we obtain the segmentation in the target image. Local appearance is considered using a random forest classifier and the final segmentation is obtained with graph cuts in a similar way as in Paper IV. The method is evaluated both on the dataset from the VISCERAL Grand Anatomy Challenge and on 30 MR images of the brain. Our method is on par with or better than state-of-the-art methods and we also improve the quality with

more plausible segmentations.

Paper VI — Vertebra Detection and Identification Using CNN-Based Detectors and Active Shape

In this paper we detect and identify vertebrae in CT images. For the detection we use convolutional neural networks which also identify if any of the detections are thoracic or sacral vertebrae, respectively. To get more information about where we are along the spine (since the images can be focused on just a part of the spine) we also detect the rib joints using CNNs. The vertebrae and rib joint detections are adapted to a shape model of the spine and through this we get the identification of each vertebra and potential false detections are discarded. On a quite difficult dataset, containing patients with different pathologies concerning the spine, we are able to detect and identify 63 % of the vertebrae.

6 Concluding Remarks

Most of us will at some point in our lives be examined for a medical reason using images. They are an important tool for the physicians when diagnosing and planning treatments. They are also important for researchers when studying different medical conditions. The assistance of mathematical methods during these examinations is of great benefit for the user; in some cases reducing the workload by automated segmentation of organs, in other cases extracting more information from the images than a visual examination, etc.

This thesis gives several examples with different applications where mathematical methods are of great assistance. The system that detects and diagnoses kidney lesions in children, eases the examination for the physicians and let them concentrate on the particularly hard regions to diagnose. The measure of septum shape gives a more objective indication on how the pump speed should be adjusted than a visual inspection. Automated (or semi-automated) segmentation of organs or other structures reduces the workload compared to if the physician would do it manually. Different measurements, of e.g. the size of the structure, are also quickly obtained. For researchers this means that they can faster obtain segmentations from a larger population which give them a more substantial material when studying their particular medical question.

This is important work for all of us that will never be finished. New imaging possibilities and increased computer capacity will give new challenges ahead.

References

[1] Y. Amit and D. Geman. Shape quantization and recognition with randomized trees. *Neural Computation*, 9(7):1545–1588, 1997.

[2] I. N. Bankman, editor. *Handbook of Medical Image Processing and Analysis*. Academic Press, second edition, 2009.

[3] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.

[4] J. S. Beis and D. G. Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1000–1006, 1997.

[5] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.

[6] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1124–1137, 2004.

[7] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

[8] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[9] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[10] L. Breiman, JH. Friedman, R. Olshen, and CJ. Stone. *Classification and regression trees*. The Wadsworth statistics/probability series. Chapman & Hall, 1984.

[11] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models — their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[12] T. H. Cormen, C. H. Leiserson, R. L. Rivest, and C Stein. *Introduction to Algorithms*. MIT Press, 2009.

[13] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959.

[14] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, 1998.

[15] J. Duchon. Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In W. Schempp and K. Zeller, editors, *Constructive Theory of Functions of Several Variables: Proceedings of a Conference Held at Oberwolfach April 25 – May 1, 1976*, pages 85–100. Springer Berlin Heidelberg, 1977.

[16] J. S. Duncan and N. Ayache. Medical image analysis: Progress over two decades and the challenges ahead. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):85–106, 2000.

[17] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[18] L. R. Ford Jr and D. R. Fulkerson. *Flows in networks*. Princeton university press, 1962.

[19] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2009.

[20] S. Haykin. *Neural networks and learning machines*, volume 3. Pearson Upper Saddle River, NJ, USA:, 2009.

[21] W. R. Hendee and E. R. Ritenour. *Medical Imaging Physics*. John Wiley & Sons, Inc., 2003.

[22] T. K. Ho. Random decision forests. In *Proceedings of the Third International Conference on Document Analysis and Recognition*, volume 1, pages 278–282, 1995.

[23] T. K. Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.

[24] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1):106–154, 1962.

[25] J. E. Iglesias and M. R. Sabuncu. Multi-atlas segmentation of biomedical images: A survey. *Medical Image Analysis*, 24(1):205–219, 2015.

[26] A. Karlsson. Area-based active contours with applications in medical microscopy. Licentiate thesis, Lund University, 2005.

[27] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[28] V. Kolmogorov and R. Zabin. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2004.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[30] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989.

[31] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[32] D. G. Lowe. Object recognition from local scale-invariant features. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157, 1999.

[33] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[34] D. L. Pham, C. Xu, and J. L. Prince. Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2(1):315–337, 2000.

[35] P. Radau, Y. Lu, K. Connelly, G. Paul, A. Dick, and G. Wright. Evaluation framework for algorithms segmenting short axis cardiac MRI. *The MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*.

[36] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(5):386–408, 1958.

[37] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland, editors, *Parallel distributed processing: Explorations in the microstructure of cognition*, volume 1, chapter 8. Cambridge, MA: MIT Press, 1986.

[38] D. Scherer, A. Müller, and S. Behnke. Evaluation of pooling operations in convolutional architectures for object recognition. In K. Diamantaras, W. Duch, and L. S. Iliadis, editors, *Artificial Neural Networks – ICANN 2010: 20th International Conference, Thessaloniki, Greece, September 15-18, 2010, Proceedings, Part III*, pages 92–101. Springer Berlin Heidelberg, 2010.

[39] H. W. Strauss, B. L. Zaret, P. J. Hurley, T. K. Natarajan, and B. Pitt. A scintiphotographic method for measuring left ventricular ejection fraction in man without cardiac catheterization. *The American Journal of Cardiology*, 28(5):575–580, 1971.

[40] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, 1999.

[41] L. Svärm, O. Enqvist, F. Kahl, and M. Oskarsson. Improving robustness for inter-subject medical image registration using a feature-based approach. In *International Symposium on Biomedical Imaging*, 2015.

[42] A. Tsai. *Curve Evolution and Estimation — Theoretic Techniques for Image Processing*. PhD thesis, Massachusetts Institute of Technology, 2000.

[43] G. Wahba. *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, 1990.

[44] P. Werbos. *Beyond regression: New tools for prediction and analysis in the behavioral sciences*. PhD thesis, Harvard University, Cambridge, MA, 1974.