

Weakly Supervised Learning from Medical Images and Radiology Reports

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Medizinische Informatik

eingereicht von

Markus Krenn

Matrikelnummer 0725992

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung: a.o.Univ.-Prof. Dipl.-Ing. Dr.techn. Robert Sablatnig

Mitwirkung: Ass.Prof. Dipl.-Ing. Dr. Georg Langs (CIR Lab, Medical University of Vienna)

Wien, 20.04.2015

(Unterschrift Verfasserin)

(Unterschrift Betreuung)

Weakly Supervised Learning from Medical Images and Radiology Reports

MASTER'S THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Medical Computer Science

by

Markus Krenn

Registration Number 0725992

to the Faculty of Informatics
at the Vienna University of Technology

Advisor: a.o.Univ.-Prof. Dipl.-Ing. Dr.techn. Robert Sablatnig

Assistance: Ass.Prof. Dipl.-Ing. Dr. Georg Langs (CIR Lab, Medical University of Vienna)

Vienna, 20.04.2015

(Signature of Author)

(Signature of Advisor)

Erklärung zur Verfassung der Arbeit

Markus Krenn
Wattgasse 35/5, 1160 Wien

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.

(Ort, Datum)

(Unterschrift Verfasserin)

Acknowledgements

At this point I would like to thank everybody who supported me during my master thesis. First of all, a big thank you goes to Georg Langs, my advisor from the Medical University of Vienna. For all his patience, expertise, support and most important the opportunity to work in the field of medical image processing at the CIR lab. He was the one who aroused my interest for scientific work in this domain, for which I am most grateful.

I also want to thank Robert Sablatnig from the Computer Vision Lab at Vienna University of Technology for his supervision. His seminars and feedback, also during his seminar "Scientific Presentation and Communication", helped a lot to improve my scientific writing skills and thus to successfully finish this thesis.

Many thanks to all my colleagues from the CIR lab for the great cooperation and support during development and writing this thesis. Thanks Albert, Joachim, Karl-Heinz, Markus and Matthias for your technical and most important mental support during hard times. Special thanks to one of my closest friends, former study and now working colleague Johannes, who went through all semester of my studies with me. It has been a great time.

A huge thank you to my wonderful partner Kristina. For your patience, understanding and unconditional support during the last couple of months.

The biggest thank you goes to my parents, Maria and Walter. For your unconditional support trough my whole studies and especially in the last months of my master thesis, the opportunity to study and to find my own way in life. Thank you!

Abstract

Computer Aided Diagnosis (CAD) systems are an important tool to guide radiologists during detection and diagnosis of clinical findings in medical images to improve their quality and productivity. Typical components of such systems involve the localisation of the structure of interest (Segmentation), digital encoding of visual information (Feature extraction) and the identification of healthy and pathological observations (Classification). The task of segmentation and classification in this context is often addressed by supervised machine learning approaches where annotated training data is required, which is usually time consuming and expensive to acquire.

The aim of the work in this thesis is to address the problems of segmentation and classification in medical images by learning methods that do not require manually annotated data, but instead learn from data that is generated during clinical routine in hospitals.

The first approach takes a set of medical images as input and aims at the unsupervised identification and segmentation of anatomical structures. The unsupervised segmentation approach consists of four main processing steps: the registration of all images to a central reference space (atlas), learning image region feature prototypes, learning a segmentation in the atlas space using Markov Random Fields (MRF) and combining the atlas segmentation with local image features to segment novel target images, again using MRFs.

The second approach aims at the classification of healthy and pathological image regions within an organ by learning from a set of medical images, where each image is assigned with a set of weak textual labels that describe clinical findings and pathologies occurring in an image. The approach is based on clustering image region features, learning the distribution of weak labels in the partitioned feature space, computing a probability table to predict single labels for clusters and using this knowledge to classify image regions in unseen images.

Evaluation shows that the unsupervised segmentation approach is able to locate three organs (lung, heart, liver) across all images, but as expected, shows clear limitations in comparison to supervised learning approaches. The approach for weakly supervised classification is evaluated on 300 chest CT scans and yields sensitivity/specificity values of 0.9/0.98 for healthy, 0.77/0.96 for ground glass, 0.91/0.98 for reticular, 0.37/0.99 for honeycombing and 0.9/0.96 for emphysema image regions.

Kurzfassung

Computerunterstützte Diagnosesysteme (CAD) spielen eine wichtige Rolle während der Befundung von medizinischen Bilddaten. Ihr Ziel ist es, Ärzte während der Befundung zu unterstützen, indem die für eine Diagnose relevante Informationen rasch und zuverlässig detektiert und visualisiert werden. Typische Aufgaben eines CAD Systems umfassen die Lokalisierung von anatomischen Regionen (Segmentierung), das Kodieren von visueller Information (Merkmalsextraktion) und die Identifikation von pathologischem Gewebe (Klassifizierung). Segmentierungs- und Klassifizierungsprobleme werden oft von Algorithmen gelöst, die dem Konzept des überwachten maschinellen Lernens zugrunde liegen. Dieses Konzept setzt jedoch annotierte Trainingsdaten voraus, welche meist nur mit hohem Zeitaufwand und hohen Kosten generiert werden können. Ziel dieser Arbeit ist daher die Entwicklung von Methoden zur Segmentierung und Klassifizierung in medizinischen Bilddaten, die aus Daten lernen, die im klinischen Alltag erzeugt werden.

Der erste Teil dieser Arbeit stellt eine Methode zur Segmentierung von anatomischen Strukturen in medizinischen Bilddaten vor, die lediglich aus einer Menge von Bilddaten lernt (unüberwachtes Lernen). Diese besteht aus vier Verarbeitungsschritten: die Registrierung aller Bilder zu einem zentralen Atlas, das Erlernen von Prototypen von Bildmerkmalen, das Erlernen einer Segmentierung in dem Atlas, wofür Markov Random Fields (MRF) benützt werden, und das Generieren von Segmentierungen in neuen Bildern durch die Kombination der Atlassegmentierung und lokalen Bildmerkmalen, wiederum mit Hilfe von MRFs.

Im zweiten Teil wird eine Methode zur Klassifikation von pathologischen Regionen in einem Organ beschrieben. Diese Methode lernt aus einer Menge Bildern, wobei zu jedem Bild textuelle Labels zur Verfügung stehen, welche die im Bild auftretenden Pathologien beschreiben und besteht aus vier Hauptteilen: das Partitionieren von Bildmerkmalen, das Erlernen der Verteilung von textuellen Labels in diesen Partitionen, das Generieren einer Wahrscheinlichkeitstabelle um einzelne textuelle Labels zu Partitionen zuzuordnen, um diese Information schließlich für die Klassifikation von Pathologien in neuen Bildern verwenden zu können.

Die Evaluierung der Methode zum unüberwachten Segmentieren zeigt, dass der Ansatz drei Organe identifizieren kann (Lunge, Herz, Leber), aber wie erwartet limitiert ist im Vergleich zu überwachten Lernverfahren. Die zweite Methode wurde auf einem Datensatz von 300 Computer Tomographie (CT) Bildern mit Annotierungen von fünf pathologischen und gesunden Strukturen in der Lunge evaluiert. Hierbei wurden Sensitivitäts- und Spezifitätswerte von 0.9/0.98 für gesunde, 0.77/0.96 für Milchglas, 0.91/0.98 für Reticular, 0.37/0.99 für Honeycombing und 0.9/0.96 für Emphysem Regionen erreicht.

Contents

1	Introduction	1
1.1	Problem Statement	2
1.2	Methodical Approach	2
1.3	Contribution of the Thesis	3
1.4	Thesis Outline	3
2	State of the Art	5
2.1	Texture Descriptors in Medical Images	5
2.1.1	Local Binary Patterns	6
2.1.2	Bag of Visual Words	7
2.1.3	Texture Bags - A Multiscale BVW Approach to describe 3D Medical Images	8
2.1.4	Haralick Features of Grey Level Co-occurrence Matrices	9
2.1.5	Discussion and Relation to Present Work	11
2.2	Data Analysis - Unsupervised Learning	13
2.2.1	k-Means Clustering	13
2.2.2	Gaussian Mixture Model Clustering	15
2.2.3	Dimensionality Reduction using Principal Component Analysis	19
2.2.4	Discussion and Relation to Present Work	19
2.3	Medical Image Registration	20
2.3.1	Image Transformations	21
2.3.2	Similarity Functions	23
2.3.3	Registration of Volumes to a Reference Space	24
2.3.4	Discussion and Relation to Present Work	26
2.4	Markov Random Fields in Medical Image Segmentation	26
2.4.1	Theoretical overview	27
2.4.2	Discussion and Relation to Present Work	28
2.5	Summary	29
3	Methodical Approach	31
3.1	Unsupervised Medical Image Segmentation on Supervoxel Level	31
3.1.1	Objects, Notation and Problem Definition	32
3.1.2	Image Normalization Pipeline	33

3.1.3	Learning Supervoxel Texture Prototypes	35
3.1.4	Finding a Latent Atlas Labeling	37
3.1.5	Finding Individual Labelings using the Latent Atlas as Prior	39
3.2	Weakly Supervised Classification of Pathologies	43
3.2.1	Problem Definition	44
3.2.2	Feature Extraction and Clustering	44
3.2.3	Mapping Terms to Clusters	45
3.2.4	Classifying Novel Image Regions	46
3.3	Summary	47
4	Experiments and Results	49
4.1	Unsupervised Medical Image Segmentation on Supervoxel Level	49
4.1.1	Data in Use	50
4.1.2	Evaluation Metric	51
4.1.3	Experimental Setup	52
4.1.4	Accuracy of Latent Atlas Space Labeling	54
4.1.5	Accuracy of Labeling Individual Volumes	58
4.2	Weakly Supervised Classification of Pathologies	63
4.2.1	Data in Use	63
4.2.2	Experimental Setup	64
4.2.3	Evaluation Metric	65
4.2.4	Evaluation of Classification Performance	65
4.3	Summary	68
5	Discussion, Conclusion and Future Work	71
5.1	Discussion	71
5.1.1	Unsupervised Medical Image Segmentation on Supervoxel Level	71
5.1.2	Weakly Supervised Classification of Pathologies	72
5.2	Conclusion	73
5.2.1	Unsupervised Medical Image Segmentation on Supervoxel Level	74
5.2.2	Weakly Supervised Classification of Pathologies	75
5.3	Future Work	75
5.3.1	Unsupervised Medical Image Segmentation on Supervoxel Level	75
5.3.2	Weakly Supervised Classification of Pathologies	76
A	Appendix	77
	Bibliography	79

Acronyms

BVW	Bag of Visual Words
BVW-LBP	Bag of Visual Words of Local Binary Patterns
CBIR	Content Based Image Retrieval
CAD	Computer Aided Diagnosis
CT	Computed Tomography
DICOM	Digital Imaging and Communications in Medicine
EM	Expectation Maximization
FFD	Free Form Deformation
GLCM	Grey Level Co-occurrence Matrices
GMM	Gaussian Mixture Model
ILD	Interstitial Lung Diseases
kNN	k nearest neighbors
LBP	Local Binary Pattern
MAP	Maximum posterior
MI	Mutual Information
MIP	Maximum Intensity Projection
MRF	Markov Random Field
MRI	Magnetic Resonance Image
NCC	Normalized Cross Correlation
NIfTI	Neuroimaging Informatics Technology Initiative
NMI	Normalized Mutual Information
PACS	Picture Archiving and Communication System
PCA	Principal Component Analysis

Introduction

Daily routines in hospitals produce large amounts of medical imaging data during diagnosis and therapy. The university Hospital of Geneva for instance produced 25 000 images per day in 2012 [44]. Finding relevant information in medical images for diagnosis (*image reading*) is often tedious, time consuming, needs expertise, is expensive [16] and additionally suffers from intra- and inter-reader variability [3], [6].

In this context, Content Based Image Image Retrieval (CBIR) and Computer Aided Diagnosis (CAD) Systems are tools to support radiologists during detection and diagnosis of clinical findings in medical images to improve quality and productivity [18], [49]. Such systems are designed to find similar cases in large medical databases (CBIR), to analyse large amounts of data in reasonable time, locate and visualize relevant information (such as pathologies) for diagnosis and provide objective and repeatable results (CAD) [19], [67], [44].

Recently, CAD systems for various clinical tasks such as early breast cancer detection [5], [39], lung nodule detection [41], polyp detection in colonography Computed Tomography (CT) [87] or tissue classification in chest CT scans [84], [49] have been developed. According to Sluimer et al. a typical CAD system is set up of four components. Preprocessing, segmentation, feature extraction and classification [66]. Classification in this context is often performed in supervised manner [39], [41], [87], [84], which requires from clinical experts annotated training data, where the acquisition of training data has the same drawbacks as the task of image reading for finding a diagnosis [16]. In particular for learning tasks that need large numbers of training examples, to represent the variability in patient data, expert annotation is not feasible.

In this thesis methods are proposed that can support learning from imaging- and textual data that is generated in clinical routine. The approaches explore what can be learned when relying on this existing data, instead of additional detailed manual annotations. To overcome the problem of acquisition of annotated training data, we emphasize the idea of CAD systems that learn from data that is available from clinical routine. Specifically, these are (1) large sets of medical images and (2) corresponding radiology reports that describe occurrences of pathological observations and clinical findings in these images. Once findings and observations in these reports are extracted and mapped to ontologies such as RadLex [46], the task of learning from

this data can be interpreted as weakly supervised learning task. Learning the correspondence between weak text labels and images has been studied in several fields outside the medical domain [24], [79] [69]. It has also been shown that learning these correspondences enables CBIR systems to improve accuracy [50].

1.1 Problem Statement

The aim of this thesis is to implement and evaluate learning algorithms for two typical components (segmentation of anatomical structures and classification of clinical findings in these structures [66]) of a CAD system. The proposed approaches do not require manually annotated training data since their acquisition is normally time consuming and expensive [16]. Instead we only use data that is available from clinical routine. The problems of segmentation and classification in medical images are addressed separately in this thesis, which leads to a two-folded problem statement:

1. The development of an unsupervised learning algorithm that is able to identify visually coherent regions and thus to compute group wise segmentations of anatomical structures in a set of medical images.
2. The development of a weakly supervised learning algorithm that is able to classify healthy and pathological regions within a previously segmented anatomical structure.

Whereas the first approach requires only a set of medical images as input, the latter learns to predict tissue classes from information that is available in radiological reports. For this purpose we assume that the content of a report contains descriptions of clinical findings and pathological observations that occur in an image. Furthermore, we assume that this relevant information can be extracted and mapped to ontologies such as RadLex [46] so that the extracted terms form weak textual labels for each image.

It has to be investigated which anatomical structures can be identified and segmented to which accuracy by the approach proposed for unsupervised medical image segmentation as well as if the method proposed for weakly supervised classification is able to identify pathological and healthy image regions. For this purpose two publicly available datasets are used that carry voxel-wise labeling of (1) anatomical structures [32] and (2) multiple (healthy and pathological) tissue classes within the lungs [37].

1.2 Methodical Approach

Both methods analyse and classify small patches of neighboring voxels with similar texture properties, so called supervoxels [38], separately. The main idea is that unsupervised clustering of supervoxel features sampled from image regions covering the whole body results in partitions of the feature space that represent prototypes of anatomical structures. The assignment of supervoxels to these prototypes is then used to generate an initial segmentation of all images in the training data. After registration of all images to a central reference space or *atlas*, we learn a segmentation of the atlas using Markov Random Fields (MRF) to combine initial segmentations

and model the assumption that spatially neighboring supervoxels are likely to belong to the same segmentation class. The final segmentation of an image is then obtained by the combination of the previously learned atlas segmentation together with the initial segmentation of an image and modelling relations between spatially neighboring supervoxels, again using MRFs.

The main idea of the approach proposed for weakly supervised classification approach is again based on unsupervised clustering of supervoxel features. Here we sample supervoxel features of a specific organ and expect that the clustering results in partitions that represent prototypes of healthy and pathological tissue classes. Weak labels that describe occurring pathologies of a whole image are then assigned to each supervoxel of an image. By assigning supervoxels and their weak labels to clusters, we establish a mapping of clusters to single labels and use this knowledge to classify supervoxels of novel images.

1.3 Contribution of the Thesis

The main contribution of this thesis is the development and detailed evaluation of approaches for the unsupervised segmentation of anatomical structures and the classification of pathological and healthy tissue in medical images. Both methods learn from data sets that are available from clinical routine instead of relying on manually annotated training data and are based on the assumption that unsupervised clustering of features results in partitions representing prototypes of anatomical structures within the first, and prototypes of healthy and pathological tissue classes within the second approach.

The main contributions of this thesis are thus outlined as follows:

- Unsupervised segmentation of anatomical structures in medical images based on unsupervised supervoxel feature clustering, atlas labeling and regularization based on MRFs,
- Classification of healthy and pathological tissues within an organ using weak text labels available from radiological reports,
- Comparative evaluation of two unsupervised clustering methods to find prototypes of anatomical structures and different tissue types,
- Comparative evaluation of two supervoxel texture descriptors and their ability to describe pathological and healthy tissue types in medical images.

1.4 Thesis Outline

The thesis is divided into five chapters and outlined as follows:

1. **Introduction:** Provides a general introduction, formulates the problem statement and summarizes the main contributions of this thesis.
2. **State of the Art:** Introduces, describes and discusses techniques that are used within this thesis.

3. **Methodical Approach:** This chapter describes the methods proposed for unsupervised image segmentation and weakly supervised classification in detail. To facilitate reading, both methods are described in separated sections of this chapter.
4. **Experiments and Results:** Describes experiments performed to evaluate both methods proposed and shows their results. This chapter is again divided into two separated sections to facilitate reading.
5. **Discussion, Conclusion and Future Work:** The final chapter provides a discussion of evaluation, draws a conclusion and closes with thoughts on future work and possible improvements of both methods.

State of the Art

The present chapter describes methods and techniques that are used and applied as components of the processing pipelines we propose for unsupervised medical image segmentation and weakly supervised classification of pathologies. Section 2.1 describes texture descriptors used to encode visual information of images. In Section 2.2 two methods for unsupervised clustering and a method for dimensionality reductions are presented. Medical image registration methods and a framework for registration of images covering different parts of the human to a central reference space are described in Section 2.3. Markov Random Fields (MRF) in the context of medical image segmentation are presented in Section 2.4 of this chapter. Finally, Section 2.5 gives a summary of the methods described within this chapter of this work.

2.1 Texture Descriptors in Medical Images

In order to analyse and classify objects such as different pathologies or organs in medical images, methods that numerically encode structural and statistical properties of an object are required. The process of extracting such information from objects in digital images is referred to as *feature extraction*. The extracted information about an object is then called *feature* or *descriptor* of an object. In the scope of this thesis we will refer to this information also as *texture descriptor* since we aim to classify different texture types in medical images.

Features are categorized depending on the area they are describing. *Voxel features* describe the local neighborhood of each voxel in an image, where *Region features* describe a patch of voxels and their structure. Furthermore we differentiate as well between *statistical* and *structural* feature extractors as suggested from Xie and Majid in [82].

Statistical features are derived from the statistical distribution of intensity values of an image region. First order statistics, such as intensity histograms, consider only one pixel per observation whereas higher order statistics encode relationships of more than one pixel [82].

Structural features describe texture using *texture primitives* and their spatial appearance in an image region. This is achieved in a two step approach. In the first step texture primitives are identified where in the second step the original image content is replaced following a replacement rule with the derived primitives. [82].

The following sections describe two statistical feature extractors and an approaches to derive structural features that describe image regions that are used within the scope of this thesis.

2.1.1 Local Binary Patterns

The Local Binary Pattern (LBP) operator introduced by Ojala et al. [57] is a higher order statistics feature. It describes the local structure of a pixels c neighborhood by using the intensity value I_c as a threshold, multiplying the thresholded neighborhood values with a weighting scheme and summing up the results as illustrated in Figure 2.1.

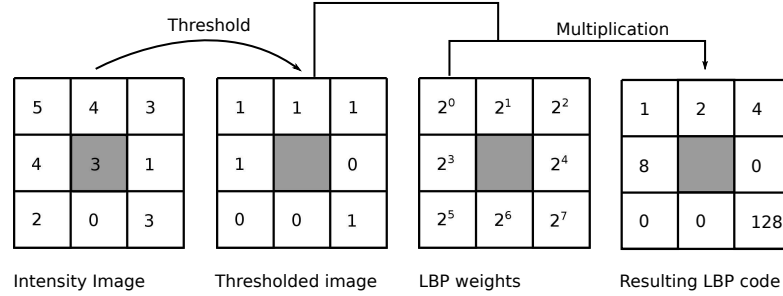


Figure 2.1: Computation of a LBP code, adapted from [52].

Suppose we have a center pixel c with P neighboring pixel and the intensity values i_c and i_p where $p = 0 \dots P - 1$. The LBP_c value is formally defined as follows [58]

$$LBP_c = \sum_{p=0}^{P-1} s(i_p - i_c) 2^p \quad (2.1)$$

where $s(x)$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2.2)$$

By assigning the binomial factor 2^p to each sign $s(i_p - i_c)$ a unique LBP_c value describing the local image structure is computed. The name Local Binary Pattern is derived from the principle approach of the operator, which generates an eight bit binary code for an eight pixel neighborhood. The LBP operator is per definition invariant to gray scale and intensity range and can be extended by a contrast measure LBP_C and a intensity measure LBP_I . LBP_C is computed by the difference of the average intensities of neighboring pixels brighter than the center pixel and pixels darker as the center pixel [52], [82]. The local intensity extension LBP_I proposed by Burner et al. [10] simply depicts the average intensity off all neighboring pixel.

2.1.2 Bag of Visual Words

The concept of Bag of Visual Words (BVW) is an approach to build image region descriptors based on voxel features. The name is derived from its application in textual information retrieval where a document is described by a normalized histogram of word counts [85]. The representation of an image region following BVW is analogous. A visual vocabulary is constructed by vector quantization or clustering of local voxel features, sampled from a set of training images. This vocabulary contains prototypes, also referred to as *visual words* or *textons* [28]. An image region is then described by extracting local features and replacing them with their nearest visual word. Classification techniques according to the vector quantization or clustering method such as k Nearest Neighbor (kNN) search are required. The resulting feature vector is a normalized histogram of occurring visual words in an image region. BVW approaches are, as the name indicates, invariant to the spatial distribution of words since only the appearance of words is depicted in the histogram.

Computing a BVW representation of an image region requires three steps [85], which are illustrated in Figure 2.2.

1. **Build Vocabulary** Features are extracted from all training images. To receive a discrete vocabulary this feature space is vector quantized or clustered, where each cluster depicts one visual word and the set of visual words is called visual vocabulary.
2. **Assign Terms** Extract features from a novel image and use feature classification techniques, such as kNN to assign the nearest visual word to each feature vector.
3. **Generate Term Vector** Generates the representation of an image region, which is a normalized occurrence histogram of visual words.

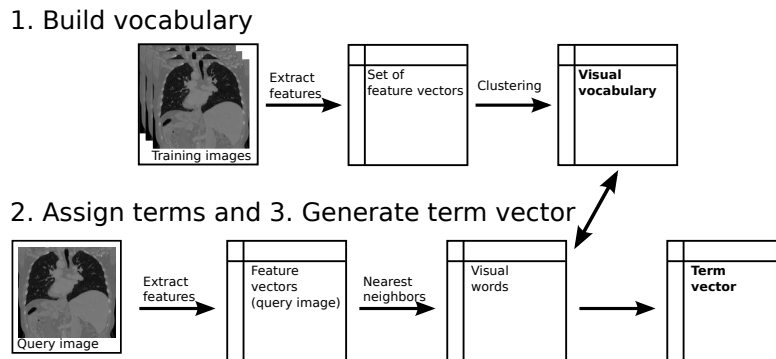


Figure 2.2: Overview of the BVW approach. First a visual vocabulary is learned by clustering features sampled from a set of training images. In the second phase features from a query image are extracted and assigned to the most similar visual word. Finally the term vector is computed in step three by building a normalized histogram of the image regions visual words.

2.1.3 Texture Bags - A Multiscale BVW Approach to describe 3D Medical Images

In the context of this thesis we follow an approach proposed by Burner et al. [10] to describe different texture types in medical images. They implement a CBIR system for pathologies and anomalies in 3D medical image data, in which they describe texture by building BVW features of so called *supervoxels* using LBP features. Supervoxels are patches of neighboring voxels, in their work computed by an oversegmentation algorithm proposed from [26] which aims to merge voxels into homogeneous regions such that the boundaries of objects in an image are preserved [38].

Burner et al. use a 3D adaption of the LBP operator, a local contrast measure and a local intensity measure as base feature extractor. As Figure 2.3 illustrates, the extension of the LBP operator from 2D to 3D increases the dimensionality of the resulting descriptor. While the 2D operator is based on a 3x3 grid that results in an 8 bit vector, using the 3D adaption based on a 3x3x3 grid results in a descriptor of 26 bits [10]. The final voxel descriptor \mathbf{D} as denoted in Equation 2.3 has 28 dimensions. 26 LBP bits \mathbf{D}_{LBP3d} , one contrast dimension \mathbf{D}_C , and one intensity dimension \mathbf{D}_I . Contrast and intensity descriptor are scaled to the range $[0, 1]$ and weighted according to the image modality by factors c_c and c_i [10].

$$\mathbf{D} = [\mathbf{D}_{LBP3d}, c_c \mathbf{D}_C, c_i \mathbf{D}_I] \quad (2.3)$$

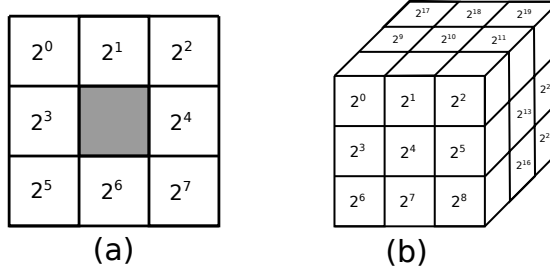


Figure 2.3: Weights of the two dimensional LBP descriptor, introduced in [57] *a*. Three dimensional adaption as used in the scope of this thesis *b*, suggested and adapted from [10].

Burner et al. compute supervoxel descriptors following the BVW approach as described in Section 2.1.2. In a first step they randomly sample a set of features from a training set of volumes and cluster this feature space using k -means clustering (see also Section 2.2) to build a visual vocabulary. This process is illustrated in Figure 2.4. The resulting k cluster centers specify the visual words \mathbf{W}_k . This process is performed on multiple resolutions of the training volumes to which the authors refer to as *scales*, so that the final set of visual words is denoted by \mathbf{W}_k^s , where s depicts the scale factor and k the index of the visual word. [10].

Figure 2.5 illustrates how texture of a novel volume i is described. All voxels are represented with the closest visual word of the voxels feature on each scale, i.e. with the index of the closest cluster center. After computing a supervoxel oversegmentation \mathbf{R}_{ij} following an approach in-

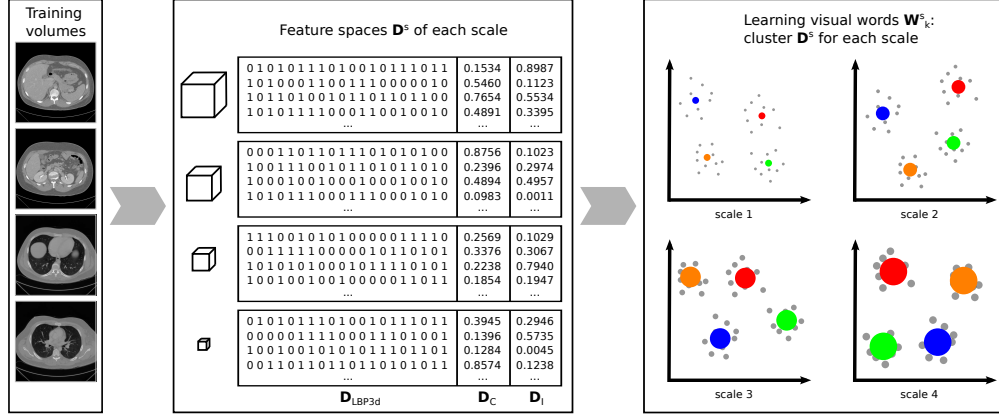


Figure 2.4: Learning of visual words as proposed in [10]. Voxel features are sampled from the training volumes on different volume scales, building the feature spaces D_s . A vocabulary for each scale is built by clustering each feature space using k -means. The cluster centres represent the resulting visual words W_k^s . Figure adapted from [10].

roduced in [81], a supervoxels texture j is described by concatenating normalized visual word occurrence histograms of each scale.

The resulting feature vector of a supervoxel to which we will refer as f^{LBP} encodes tissue prototype patterns occurring in a supervoxel that are learned in unsupervised manner from the training data. The dimensionality of f^{LBP} depends on the number of scales s chosen and the size k of the visual vocabulary on each scale. In context of this thesis, implementations in C are used for the LBP operator, k -means clustering and nearest neighbor assignment of features to visual words. The implementation for sampling visual word histograms of supervoxels is implemented in Matlab.

2.1.4 Haralick Features of Grey Level Co-occurrence Matrices

The second texture descriptor that is used in the scope of this thesis has been proposed from Haralick et. al. in [33]. They suggest to use statistical properties of Grey Level Co-occurrence Matrices (GLCM) in order to create second order statistic image features. GLCMs depict the number of all pair wise combinations of grey levels that occur in the neighborhood of a center pixel [73], [11]. The $G \times G$ GLCM $M_{d,G}$ of an n dimensional image I is defined by a displacement vector $d = \{d_1 \dots d_n\}$ and grey levels G , where a linear grey level quantization function maps each grey value of the image to the range of grey levels G and d describes the offset of pixel pairs in each dimension n . An entry $M_{d,G}(i, j)$ then contains the number of occurring pairs of grey levels (i, j) with offset d . Consider for instance an two dimensional image I of

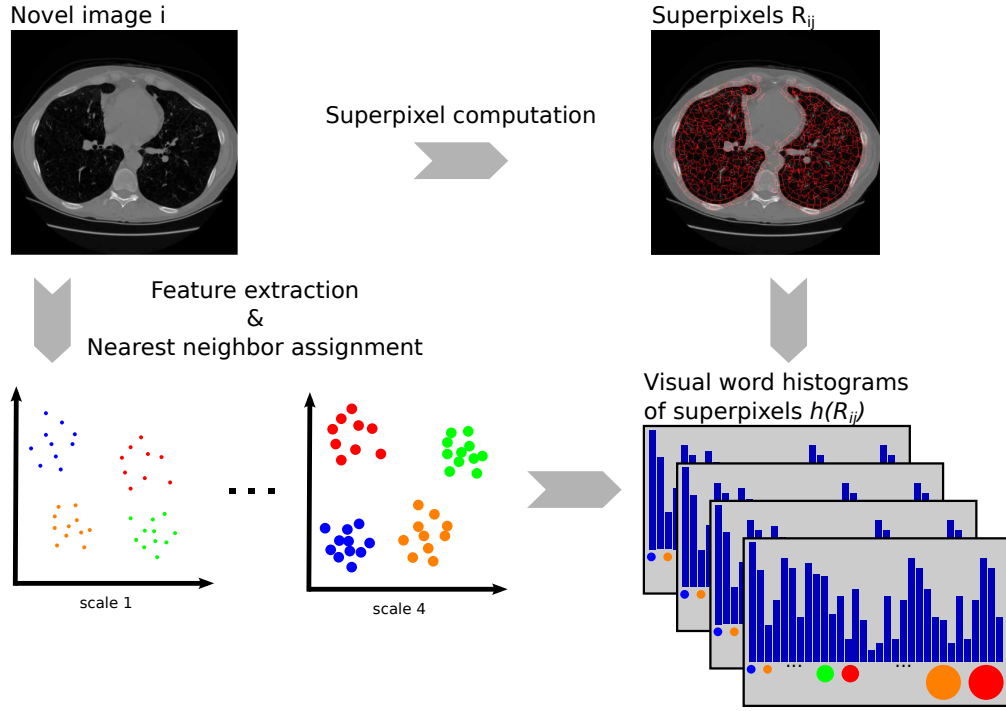


Figure 2.5: Building supervoxel descriptors. All voxels in a novel volume are represented with their closest visual word. A supervoxel is then described by the concatenated occurrence visual word histograms of all scales [10].

size 5×5 as follows

$$I = \begin{bmatrix} 2 & 2 & 0 & 0 & 0 \\ 2 & 1 & 1 & 1 & 0 \\ 0 & 1 & 3 & 1 & 2 \\ 0 & 1 & 1 & 1 & 2 \\ 0 & 0 & 2 & 2 & 2 \end{bmatrix} \quad (2.4)$$

with $d = (1, 1)$ and $G = 1 \dots 4$, $M_{d,G}$ results in

$$\mathbf{M}_{d,G} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 2 & 2 & 3 & 1 \\ 0 & 5 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad (2.5)$$

Haralick et. al. furthermore propose a set of features that is calculated from GLCMs in [33]. After introducing statistical parameters we give formal definitions of the features used in the scope of this thesis in Table 2.1.

Let p_x and p_y define the marginal distributions of $\mathbf{M}_{d,G}$ as

$$p_x(i) = \sum_j \mathbf{M}(i, j) \quad p_y(j) = \sum_i M(i, j) \quad (2.6)$$

while μ_x, μ_y depict the means and σ_x, σ_y the standard deviation of p_x and p_y respectively. Furthermore let H_X and H_Y be the entropies of p_x and p_y , and

$$H_{XY} = - \sum_i \sum_j p(i, j) \log(p(i, j)) \quad (2.7)$$

$$H_{XY1} = - \sum_i \sum_j p(i, j) \log(p_x(i)p_y(j)) \quad (2.8)$$

$$H_{XY2} = - \sum_i \sum_j p_x(i)p_y(j) \log(p_x(i)p_y(j)) \quad (2.9)$$

Finally we define p_{x+y} and p_{x-y} as

$$p_{x+y}(k) = \sum_i \sum_j \mathbf{M}(i, j) \quad , k = i + j \quad (2.10)$$

$$p_{x-y}(k) = \sum_i \sum_j \mathbf{M}(i, j) \quad , k = |i - j| \quad (2.11)$$

In order to compute Haralick features that describe texture of supervoxels, we calculate GLCMs for each supervoxel of a volume by extracting cubic patches with a fixed side length l centred around the supervoxels center. We furthermore define 13 independent displacement vectors to sample pixel-pairs not only in 2D but as well in 3D as suggested in [74]. The direction of a displacement vector is described using two angles θ and ϕ , where θ denotes the angle between X axis and Y plane and ϕ between X axis and the Z plane. Table 2.2 shows the displacement vectors and their directions used in this thesis, similar to [74].

While using Haralick features in order to describe texture the following parameters have to be chosen according to the task requirements. G , the number of grey levels, which affects the dimensionality of resulting GLCMs. l , patch side length of voxel patches extracted from a supervoxels center and D the pixel offset used to sample pixel pairs while building the GLCMs. GLCMs and Haralick features are then calculated for each value of D separately so that the final super voxel feature vector f_H as denoted in Equation 2.12 has $13 \cdot |D|$ dimensions.

$$f_H = \{f_{H1,1}, f_{H2,1}, \dots, f_{H13,1}, f_{H1,2}, f_{H2,2}, \dots, f_{H13,2}, f_{H1,D}, f_{H2,D}, \dots, f_{H13,D}\} \quad (2.12)$$

2.1.5 Discussion and Relation to Present Work

We have introduced two feature extractors and approaches that allow texture description of supervoxels. Both feature extractors have been used in literature to describe texture in medical

Name	Formula
Energy	$f_{H1} = \sum_i \sum_j \mathbf{M}(i, j)^2$
Contrast	$f_{H2} = \sum_i \sum_j (i - j)^2 \mathbf{M}(i, j)$
Correlation	$f_{H3} = \frac{\sum_i \sum_j (i - \mu_x)(j - \mu_y) \mathbf{M}_{d,G}(i, j)}{\sigma_x \sigma_y}$
Sum of squares	$f_{H4} = \sum_i (i - \mu_x)^2 p_x(i)$
Inverse difference moment	$f_{H5} = \sum_i \sum_j \frac{p(i, j)}{1 + (i - j)^2}$
Sum average	$f_{H6} = \sum_{n=2}^{2N} n p_{x+y}(n)$
Sum Variance	$f_{H7} = \sum_{n=2}^{2N} (n - f_{H6})^2 p_{x+y}(n)$
Sum entropy	$f_{H8} = \sum_{n=2}^{2N} (p_{x+y} \log(p_{x+y}(n)))$
Entropy	$f_{H9} = - \sum_i \sum_j \mathbf{M}(i, j) \log(\mathbf{M}(i, j))$
Difference Variance	$f_{H10} = \sum_j (j - \sum_i i p_{x-y}(i))^2 p_{x-y}(j)$
Difference Entropy	$f_{H11} = \sum_{n=0}^N -1 p_{x-y}(n) \log(p_{x-y}(i))$
Information of correlation 1	$f_{H12} = \frac{H_{XY} - H_{XY1}}{\max(H_X, H_Y)}$
Information of correlation 2	$f_{H13} = \frac{1}{1 - \exp(-2(H_{XY2} - H_{XY}))}$

Table 2.1: Haralick features proposed in [33].

image data [75], [10], [25], [70], [45], [83]. Both methods require task specific parameter settings. Using BOV-LBP features one has to define the number of visual words and the number of scales, which influence the resulting dimensionality of the feature vector as well as weighting factors for intensity and contrast measures. Using Haralick features one has to define the patch size of the area to be considered around a super voxels center, number of grey levels and the offset between pixel pairs.

BVW-LBP features are sensitive to texture orientation even though the BVW approach itself is reported to be independent to texture orientation [28], since the bagged LBP features are sensitive to texture orientation [57]. Haralick features [33] describe statistical properties of GLCMs which are sampled in multiple directions so that these features are not sensitive to texture orientation.

Both supervoxel feature extraction methods discussed are used within this work. BVW-LBP features are used to describe and learn prototypes of supervoxels within the unsupervised segmentation approach, whereas we use and evaluate both methods (BVW-LBP and Haralick features) within the weakly supervised classification part of this work.

angles (θ, ϕ)	displacement vector \mathbf{d}
(0°, 45°)	(D,0,D)
(0°, 90°)	(D,0,0)
(0°, 135°)	(D,0,-D)
(45°, 45°)	(D,D,D)
(45°, 90°)	(D,D,0)
(45°, 135°)	(D,D,-D)
(90°, 45°)	(0,D,D)
(90°, 90°)	(0,D,0)
(90°, 135°)	(0,D,-D)
(135°, 45°)	(-D,D,D)
(135°, 90°)	(-D,D,0)
(135°, 135°)	(-D,D,-D)
(0°, 0°)	(0,0,D)

Table 2.2: Displacement vectors used to sample pixel pairs in a cubic neighborhood, as suggested in [74]. D depicts the offset between pixel pairs.

2.2 Data Analysis - Unsupervised Learning

Clustering methods make it possible to discover homogeneous classes in a quantity of objects based on the object’s similarities [31]. Given a set of objects, clustering methods aim to find a reduced representation of the data describing the underlying classes, ensuring that objects of the same class show similar properties and objects from different classes have dissimilar properties.

In the scope of this work so called *Partitional clustering* methods are applied on from training data sampled sets of supervoxel features to detect clusters that represent feature prototypes of anatomical structures and subclasses of healthy and pathological image regions.

Partitional clustering methods find partitions in a set of objects such that objects in one partition are more similar to each other than to objects in other clusters and one object is assigned to exactly one cluster. Clusters are for instance modelled by mean vectors, which are not necessarily members of the dataset, where each object belongs to the cluster with the smallest distance [31].

The following sections describe and discuss two approaches of partitional clustering that are used in this work. Furthermore, a method for dimensionality reduction is introduced that tackles the problem of the *curse of dimensionality*, which states that clustering in high dimensional feature spaces is difficult, because all pairs of points tend to have the same distance to each other [34].

2.2.1 k-Means Clustering

Suppose we have a dataset of N points $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ from a random D -dimensional variable \mathbf{X}_N and we want to partition the dataset into a given number of K clusters. K-Means clustering represents one cluster as mean vector of the points that belong to the cluster. The idea is to find

an assignment of points to clusters that minimizes the sum of all distances between points and cluster centers [9].

This is formalized by introducing the D -dimensional vectors \mathbf{c}_k , $k = 1 \dots K$ describing the cluster centers and the binary assignment matrix $\mathbf{A}_{n,k} \in \{0, 1\}$ depicting the assignments of points to clusters, i.e. $\mathbf{A}_{n,k} = 1$ if point n is assigned to cluster k and $\mathbf{A}_{n,k} = 0$ otherwise. Equation 2.13 defines the so called *distortion function* J , depicting the sum of square distances from all points to its cluster centers \mathbf{c}_k , which is going to be minimized in order to find an optimal cluster assignment.

$$J = \sum_{n=1}^N \sum_{k=1}^K \mathbf{A}_{n,k} \|\mathbf{x}_n - \mathbf{c}_k\|^2 \quad (2.13)$$

Cluster centers \mathbf{c}_k are calculated as described in Equation 2.14 [9].

$$\mathbf{c}_k = \frac{\sum_n \mathbf{A}_{n,k} \mathbf{x}_n}{\sum_n \mathbf{A}_{n,k}} \quad (2.14)$$

K-means clustering algorithms proceed as follows to find an assignment of $\mathbf{A}_{n,k}$ that minimizes the distortion function J [31].

1. Initialize \mathbf{c}_k by randomly selecting k points from \mathbf{X}_n
2. Repeat until there is convergence
 - a) Assign each point n to its nearest cluster center, i.e. update $\mathbf{A}_{n,k}$
 - b) Recalculate \mathbf{c}_k as described in Equation 2.14.

Since the steps *a* and *b* reduce the distortion function J in each iteration, convergence is assured [9]. Note that k-means algorithms are sensitive to the the initialization of the cluster centers and can thus converge to a local minimum. This effect is in practise avoided by running k-means several times with different initializations and keeping the assignment with the smallest distortion function [9]. To illustrate the characteristics of the following approaches we generate a two-dimensional synthetic dataset $\mathbf{X}^S = \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ originated from three Gaussian distributions denoted as $\mathbf{X}_i \sim \mathcal{N}(\mu_i, \Sigma_i)$ as follows:

$$\mathbf{X}_1 \sim \mathcal{N}\left(\begin{pmatrix} 2 \\ 11 \end{pmatrix}, \begin{pmatrix} 2.5 & 0.2 \\ 0.2 & 2.5 \end{pmatrix}\right) \quad (2.15)$$

$$\mathbf{X}_2 \sim \mathcal{N}\left(\begin{pmatrix} 6 \\ 4 \end{pmatrix}, \begin{pmatrix} 0.4 & 0 \\ 0 & 0.4 \end{pmatrix}\right) \quad (2.16)$$

$$\mathbf{X}_3 \sim \mathcal{N}\left(\begin{pmatrix} 11 \\ 7.5 \end{pmatrix}, \begin{pmatrix} 1.5 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}\right) \quad (2.17)$$

$$(2.18)$$

Figure 2.6 illustrates the process of k-means clustering when partitioning \mathbf{X}^S into $k = 3$ clusters. Plot *a* shows the dataset in blue and the initial cluster centers. Plots *b* - *d* illustrate the partitioning and the updated cluster centers after each iteration. Please note that the algorithm terminates after four iterations, plots *e* - *f* illustrate the cluster assignment convergence.

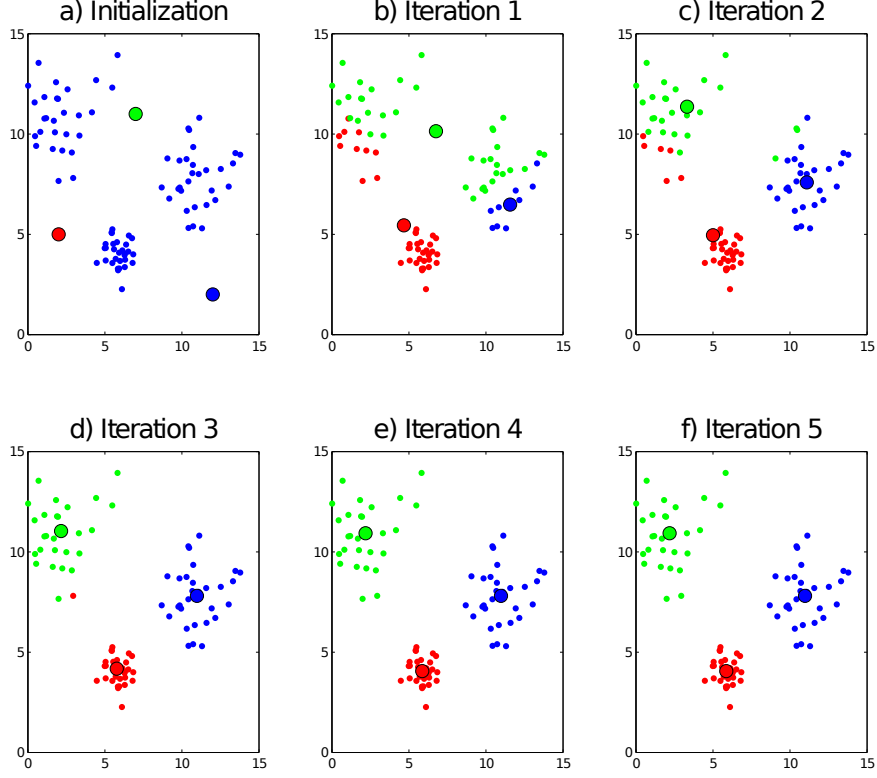


Figure 2.6: Illustration of the k-means algorithm while clustering the synthetic data \mathbf{X}^S into three clusters. The figure shows the current cluster centers (bold dots) and the assignment of data points in red, green and blue.

2.2.2 Gaussian Mixture Model Clustering

An alternative to k-means clustering where one represents one cluster as mean vector of the assigned points, is to fit a Gaussian Mixture Model (GMM) to the given data. Suppose again we have a dataset of N points $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ from a random D -dimensional variable \mathbf{X}_N and we want to partition the dataset into a given number of K clusters, we formalize a GMM as linear superposition of k Gaussian distributions as [9]:

$$p(\mathbf{x}) = \sum_{k=1}^K \pi_K \mathcal{N}(\mathbf{x} \mid \mu_k, \Sigma_k) \quad (2.19)$$

Each Gaussian density $\mathcal{N}(\mathbf{x} \mid \mu_k, \Sigma_k)$, also called a *component* of the mixture, describes one cluster and has its own mean μ_k and covariance Σ_k . The parameters π_k are referred to as *mixing coefficients* satisfying $\sum_{k=1}^K \pi_k = 1$ and $0 \leq \pi_k \leq 1$ and are interpreted as the *prior* probability that an observation is drawn from component k [9].

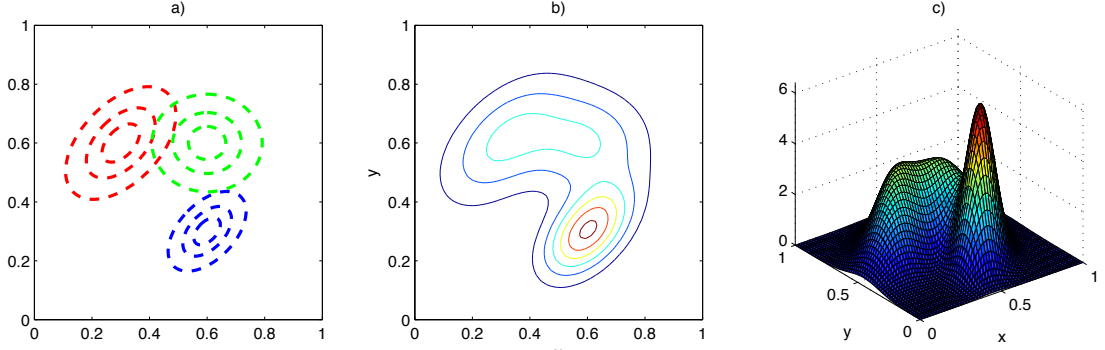


Figure 2.7: Gaussian Mixture Model of three components, adapted from [9].

Figure 2.7 illustrates a mixture of three Gaussian components in two dimensional space. (a) shows the contours of the constant probability, (b) contour plot of the marginal probability density of the mixture model and (c) 3D surface plot of the 2D density function. Figure adapted from [9].

The model is fitted to the dataset \mathbf{X}_N by maximizing the log likelihood function as stated in Equation 2.20.

$$\ln p(\mathbf{X} \mid \pi, \mu, \Sigma) = \sum_{n=1}^N \ln \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}_n \mid \mu_k, \Sigma_k) \quad (2.20)$$

One method to find the maximum likelihood estimate of parameters describing an underlying distribution from a given dataset where the parameters are missing, is to use Expectation Maximization (EM) algorithms [8]. We introduce a K -dimensional binary random variable \mathbf{z} , where $z_k \in \{0, 1\}$ and $\sum_k z_k = 1$ which is interpreted as latent indicator variable that depicts which mixture component a data point comes from. The marginal distribution over \mathbf{z} is expressed by the mixing coefficients π_k as follows:

$$p(z_k = 1) = \pi_k \quad (2.21)$$

Since \mathbf{z} uses a 1-of- K representation in which one particular element z_k is equal to 1 and all other elements are 0, we can formulate the distribution as

$$p(\mathbf{z}) = \prod_{k=1}^K \pi_k^{z_k} \quad (2.22)$$

and the conditional distribution of \mathbf{x} given a specific value for \mathbf{z} as

$$p(\mathbf{x} \mid z_k = 1) = \mathcal{N}(\mathbf{x} \mid \mu_k, \Sigma_k) \quad (2.23)$$

or similar to 2.22 in form of

$$p(\mathbf{x} | \mathbf{z}) = \prod_{k=1}^K \mathcal{N}(\mathbf{x} | \mu_k, \Sigma_k)^{z_k} \quad (2.24)$$

The joint distribution is given by $p(\mathbf{z})p(\mathbf{x} | \mathbf{z})$ and the marginal distribution of \mathbf{x} is then obtained by summing the joint distribution over all possible states of \mathbf{z} . Using equations 2.22 and 2.24 we can reformulate the Gaussian mixture model from 2.19 showing that for every data point \mathbf{x}_n exists a corresponding latent variable \mathbf{z}_k .

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{x} | \mathbf{z}) = \sum_{\mathbf{z}} p(\mathbf{z})p(\mathbf{x} | \mathbf{z}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x} | \mu_k, \Sigma_k) \quad (2.25)$$

We are now able to calculate the conditional probability of \mathbf{z} given \mathbf{x} using Bayes' theorem [9]:

$$\gamma(z_k) \equiv p(z_k = 1 | \mathbf{x}) = \frac{p(z_k = 1)p(\mathbf{x} | z_k = 1)}{\sum_{j=1}^K p(z_j = 1)p(\mathbf{x} | z_j = 1)} = \frac{\pi_k \mathcal{N}(\mathbf{x} | \mu_k, \Sigma_k)}{\sum_{j=1}^K \pi_j \mathcal{N}(\mathbf{x} | \mu_j, \Sigma_j)} \quad (2.26)$$

While we see π_k as the prior probability of $z_k = 1$ meaning that a data point \mathbf{x} comes from Gaussian component k , we interpret $\gamma(z_k)$ as posterior probability once \mathbf{x} has been observed. $\gamma(z_k)$ is also referred to as *responsibility* that component k takes for explaining the observation \mathbf{x} [56], [9].

In order to fit the GMM to a given dataset we have to maximize the log likelihood function of the mixture model with respect to the observed data. This is done by setting the derivative of the log likelihood function from 2.20 to zero with respect to each parameter separately. Using Equation 2.26, we can calculate the parameters π , μ and Σ as follows:

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{n,k}) \mathbf{x}_n \quad (2.27)$$

with

$$N_k = \sum_{n=1}^N \gamma(z_k), \quad (2.28)$$

$$\Sigma_k = \frac{1}{N_k} \sum_{n=1}^N (\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T \quad (2.29)$$

and

$$\pi_k = \frac{N_k}{N} \quad (2.30)$$

EM algorithms as described in [15] and [8] use these parameters to iteratively evaluate the posterior probabilities $\gamma(z_{n,k})$ of the model in a first, so called *Expectation* or *E* step. Where they update the model parameters (μ_k , Σ_k and π_k) according to the posterior probabilities in the *Maximization* or *M* step until the method converges. The EM algorithm for GMM clustering is thus outlined as follows in four steps [9]:

1. Initialize means μ_k , covariances Σ_k and mixing coefficients π_k , which can for instance be achieved by running k-means clustering and compute the initial value of the log likelihood function [31].
2. **E-step** Calculate responsibilities $\gamma(z_{n,k})$ using current values μ_k , Σ_k , π_k and Equation 2.26.
3. **M-step** Re-estimate parameters using the calculated responsibilities $\gamma(z_{n,k})$ of step 2.
4. Evaluate Log-Likelihood using Equation 2.20. If either the parameters or the log likelihood converged, terminate the algorithm. Return to step 2 otherwise.

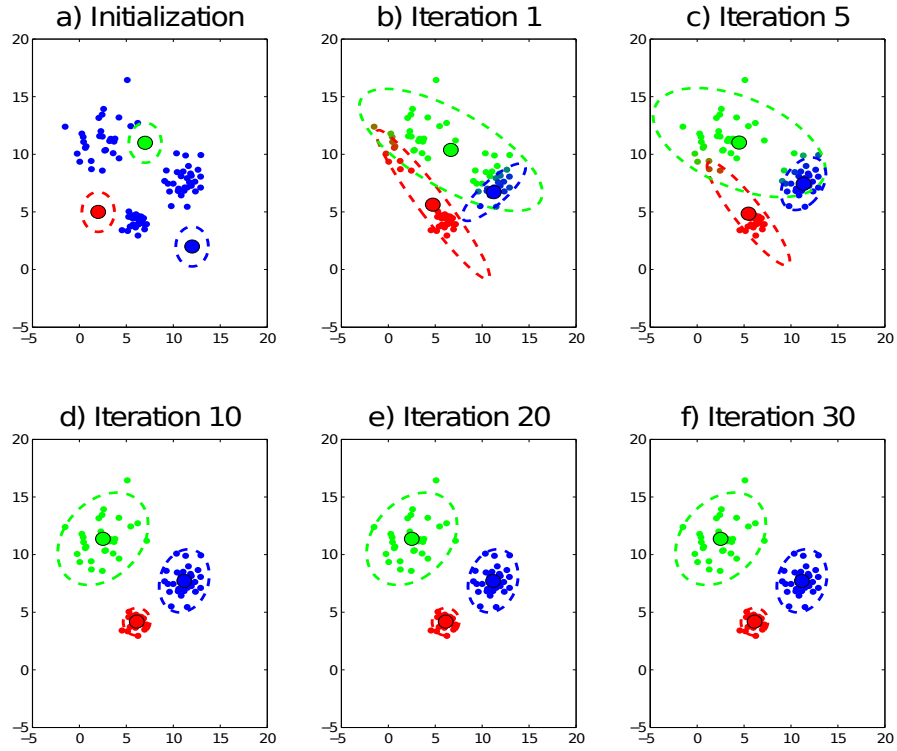


Figure 2.8: Illustration of GMM clustering on the synthetic dataset \mathbf{X}^S . Plot *a* shows the generated data and the randomly initialized Gaussian mixtures. Plots *b-f* depict the current cluster assignment in blue, green and red as well as the clusters current distribution parameters.

Figure 2.8 illustrates the EM clustering algorithm. Plot *a* shows the data the contours of the initial mixture model in red, green and blue. Plots *b-f* show results after 1, 5, 10, 20 and 30 complete iterations of the EM algorithm. The color of each point depicts the probability of having been generated from the red, green and blue mixture component.

2.2.3 Dimensionality Reduction using Principal Component Analysis

Clustering data in a high dimensional feature space is a particular challenge because of the so called *curse of dimensionality*, which states that all pairs of points in a high-dimensional space tend to have the same distance to each other so that distance based clustering approaches may not be able to find meaningful clusters in the original feature space [56]. Additionally some dimensions might be irrelevant for clustering because of correlation or redundancy to other dimensions [56].

One approach to overcome this problem is to find an embedding of the original data in a lower dimensional subspace and to seek for clusters in this subspace. Principal Component Analysis (PCA) in this context is a technique to reduce the dimensionality of data by projecting the data into a lower dimensional linear subspace that is orthogonal to the original space so that the variance of the projected data is maximized [34].

Assume we have a $N \times D$ dimensional data matrix \mathbf{X} holding a set of N observations drawn from a random D -dimensional variable \mathbf{x}_i and we want to project the data into a space having $M < D$ dimensions. Let μ be the D -dimensional mean vector and Σ denote the $D \times D$ dimensional covariance matrix of \mathbf{X} . In order to derive the directions and shares of maximal variances, an Eigenvalue decomposition of Σ using the formula

$$\Sigma = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1} \quad (2.31)$$

is performed [9]. By solving this equation we receive the matrix \mathbf{U} holding the D Eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_D$ and the diagonal matrix $\mathbf{\Lambda}$, which holds the corresponding Eigenvalues $\lambda_1, \dots, \lambda_D$. The Eigenvectors \mathbf{u}_i are orthogonal to each other and point in direction of maximum variances. The corresponding Eigenvalues λ_i indicate the share of variance in percent of each Eigenvector, i.e. $\sum_{i=1}^D \lambda_i = 1$. In other words, the Eigenvalues depict the proportional amount of variance that is expressed by an Eigenvector [9].

The dimensionality of the original data is then reduced by projection \mathbf{X} into the M dimensional embedding space Φ , which is build by using the first M Eigenvectors according to Equation 2.32.

$$\Phi = \mathbf{U}^T (\mathbf{X} - \mu) \quad (2.32)$$

Similar, one can define to reduce the dimensionality of the data so that a specific amount of variance of the data is preserved by selecting the Eigenvectors until the sum of their Eigenvalues exceeds the target amount of variance. Figure 2.9 illustrates the PCA on the data from the synthetic dataset \mathbf{X}_3 . Next to the original data we see the Eigenvectors on the right side which build the coordinate system for the embedding space Φ .

2.2.4 Discussion and Relation to Present Work

We have introduced two approaches for unsupervised partitional clustering of data. Both methods discussed (k -means and GMM) require to define the number of clusters a priori. K -means models its clusters as mean vector of each classes assigned observations, where a GMM models one class as Gaussian distribution, having means and variances for each cluster. GMM provides a so called *soft* cluster assignment which means that a posterior probability for each cluster given

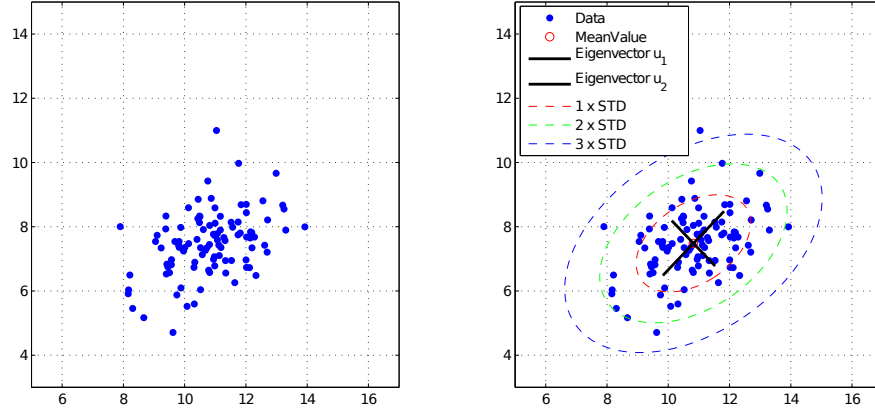


Figure 2.9: *Left*: Data obtained from X_3 . *Right*: Illustration of PCA on X_3 . Eigenvectors in black and their standard deviation curves.

an observation is calculated, where k -means only provides a *hard* cluster assignment (i.e. one observation belongs to one and only one class). We have furthermore shown a method to tackle the *curse of dimensionality*, which states that all points in high dimensional spaces tend to have similar distances. PCA is a method to overcome this problem by projecting the observations in a lower dimensional feature space while preserving maximal variance of the original data.

Unsupervised clustering methods are used in this work to find prototypes of supervoxel image features, which are expected to represent anatomical structures in the image segmentation approach and subclasses of healthy and pathological supervoxel features in the weakly supervised classification approach of presented in this work. PCA is applied within the weakly supervised classification approach to reduce the dimensionality of BVW-LBP features.

2.3 Medical Image Registration

The goal of image registration is to transform a set medical images into one common coordinate system so that a spatial relationship in all images is ensured. The relationship in the context of this thesis covers anatomical correspondences (i.e. lining up the same anatomical structures in all images). In other contexts the spatial relation covers functional correspondences (i.e. functional equivalent regions of medical images are aligned) or functional-spatial correspondences (i.e. lining up functional information on structural images) [12].

Approaches that use a common coordinate system to overcome anatomical variability in image collections are also referred to as atlas based methods where the term *atlas* depicts the shared coordinate system [23]. Such approaches are among others used to propagate annotations from an atlas to the underlying image population, also referred to as *atlas based segmentation* [64], [76] or to differentiate healthy and pathological subjects of an image collection [40].

The following sections describe the transformations applied to the images to achieve anatomical correspondences, as well as image similarity measurements which give an opportunity to estimate the performance of an image alignment and finally an approach that registers medical

images covering different body parts to one shared reference space [23].

2.3.1 Image Transformations

An image registration system consists of three components. A static target or reference image, a source image that should be aligned to the target image and finally the transformation model that describes the alignment between source and target image. Transformation models are categorized depending on whether the transformation deforms the source image *locally* or *globally* and whether the transformation is *rigid* or *non-rigid* [12], [54], [63]. We discuss the transformation models used in the scope of this thesis in the following sections.

Rigid and Affine Transformation

Rigid and affine transformations, both also called *global* deformation models since the transformation is applied to the entire source image, cover rotation and translation (rigid transformation) as well as scaling and shearing (affine transformation) of the source image to achieve alignment [12]. Rigid and affine transformations are described using a transformation matrix T to transform a coordinate vector from the source image coordinate system \mathbf{x} to the target image coordinate system \mathbf{x}^T as follows:

$$\mathbf{x}^T = T\mathbf{x}, \quad (2.33)$$

where T is a composition of affine and rigid transformations. For the two dimensional case T is defined as

$$T = \begin{pmatrix} t_{11} & t_{12} & 0 \\ t_{21} & t_{22} & 0 \\ t_{31} & t_{32} & 1 \end{pmatrix} \quad (2.34)$$

The parameters t_{11} - t_{32} are derived by matrix multiplications of the transformation matrices applied, which are illustrated in Table 2.3. Please note that matrix multiplications are associative $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$ but not commutative, i.e. $\mathbf{AB} \neq \mathbf{BA}$, meaning that the order of the applied transformation influences the resulting transformation matrix.

Since affine and rigid transformation are global transformation methods it is not possible to model local deformations. To improve image alignment rigid transformations can be used as initial registration step to reduce the global alignment error while in a second step a non-rigid registration is used to model local deformations [4], [23], [63].

Non Rigid Transformation

Rueckert et al. [63] propose a Free-Form Deformation (FFD) model that is based on B-splines [47], [48]. The idea of their approach is to deform an image or object by moving a mesh grid of control points while maximizing an image similarity function. This approach results in a smooth C^2 continuous transformation [63].

B-spline based FFD has, in contrast to thin-plate splines or elastic-body splines [47], [48], the advantage that B-splines are locally controlled which makes the computation even for a

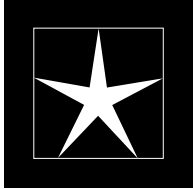
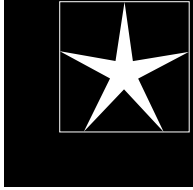
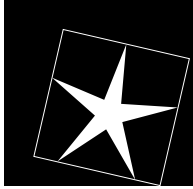
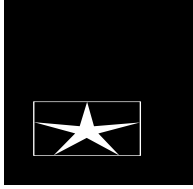
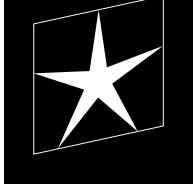
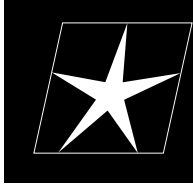
Transformation name	Matrix	Illustration
Identity	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	
Translation	$\begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix}$	
Rotation	$\begin{pmatrix} \cos \alpha & \sin \alpha & 0 \\ -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}$	
Scaling	$\begin{pmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{pmatrix}$	
Shearing (vertical)	$\begin{pmatrix} 1 & 0 & 0 \\ s_v & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	
Shearing (horizontal)	$\begin{pmatrix} 1 & s_h & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	

Table 2.3: Affine and rigid transformation matrizes

large number of control points efficient, since moving one control point affects only neighboring control points [63]. Figure 2.10 illustrates the non-rigid deformation of a CT image covering thorax and the abdomen using a B-spline FFD model. Subplot (a) shows the source image with its underlying mesh grid of control points, where (b) illustrates the target image, while (c) shows the deformed image and its control points.

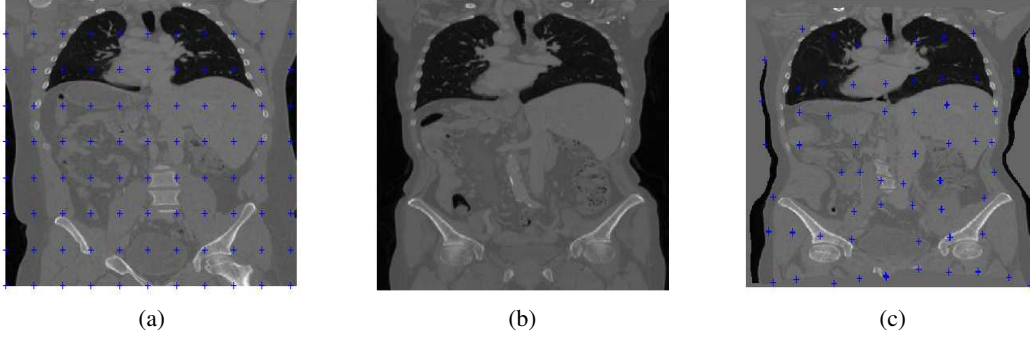


Figure 2.10: Source image with underlying mesh grid of control points (a), target image (b) and registered source image with deformed control points (c).

2.3.2 Similarity Functions

Similarity functions are measurements to quantify the similarity between two images. In the scope of image registration similarity functions serve as tool to evaluate the quality of image alignment and are used as a component of cost functions during an optimization process [63]. The following sections describe two *voxel based* similarity measurements, meaning that no additional information like landmarks or personal judging than intensity information is necessary to quantify image similarity [53].

Starting with Normalized Cross Correlation (NCC), which is a metric to measure *inner-modality* similarity between images, meaning that both images require to be recorded with the same recording technique (e.g. MR-MR or CT-CT). Followed by Normalized Mutual Information (NMI) which is used to quantify *multi-modal* similarity (e.g. MR-CT).

Normalized Cross Correlation

NCC is the sum of products of pairwise intensity values subtracted from their mean intensity, normalized by the product of standard deviations of both images. Considering a source image I_S and a target image I_T , where the intensity of a voxel v is given by $I(v)$, NCC is then defined by Equation 2.35 [88].

$$NCC = \frac{\sum_{v=1}^N (I_T(v) - \mu_T)(I_S(v) - \mu_S)}{\sqrt{\sum_{v=1}^N (I_T(v) - \mu_T)^2} \sqrt{\sum_{v=1}^N (I_S(v) - \mu_S)^2}} \quad (2.35)$$

NCC lies in the range of $[0, 1]$ where 0 depicts minimal and 1 depicts maximal similarity between the compared images.

Normalized Mutual Information

Mutual information (MI) proposed by Maes et al. [53] is used to measure similarity between images originated from different modalities [63]. It is defined based on the interpretation of the images I_S and I_T as random variables, having marginal entropy $H(I_S)$, $H(I_T)$ and joint entropy $H(I_S, I_T)$, as follows [53]:

$$MI(I_S, I_T) = H(I_S) + H(I_T) - H(I_S, I_T) \quad (2.36)$$

where $H(I_S)$ is defined over the marginal probability distribution $p_{I_S}(i_s)$ as

$$H(I_S) = - \sum_i p_{I_S}(i_s) \log p_{I_S}(i_s) \quad (2.37)$$

and $H(I_S, I_T)$ is defined over the joint probability distribution $p_{I_S I_T}(i_s, i_t)$ as

$$H(I_S, I_T) = - \sum_{i_s, i_t} p_{I_S I_T}(i_s, i_t) \log p_{I_S I_T}(i_s, i_t) \quad (2.38)$$

The normalized version of MI (NMI) was proposed by Studholme et al. and is given in Equation 2.39 [72].

$$NMI = \frac{H(I_S) + H(I_T)}{H(I_S, I_T)} \quad (2.39)$$

2.3.3 Registration of Volumes to a Reference Space

The previous sections described approaches to register images to each other where both images cover the same anatomical regions of the human body. In the scope of this thesis we follow an approach proposed by Dorfer et al. [23] to register medical images covering different parts of the human body (i.e. thorax, abdomen, thorax and abdomen) to a whole body CT image, to which we will refer to as *atlas* \mathbf{A} from now on. The aim of this approach is to find a transformation $T_{i,A}(\mathbf{x})$ which maps each position \mathbf{x} of an image \mathbf{I}_i to a position \mathbf{x}' in \mathbf{A} . Figure 2.11 illustrates the goal of the proposed approach.

Dorfer et al. propose a three step process to register an image \mathbf{I}_i to a whole body CT image serving as atlas \mathbf{A} . In a first step the center position c_i with respect to \mathbf{A} is estimated following an miniature similarity based approach proposed in [20]. Secondly the atlas region $\mathbf{R}_{i,A}$ which is expected to cover the anatomical region of \mathbf{I}_i , is estimated using the images estimated center position c_i and affine image transformations. The final transformation is computed by performing a non-rigid registration of affine transformed image \mathbf{I}_i to $\mathbf{R}_{i,A}$ [23].

First, the center position \mathbf{c}_i of a novel volume \mathbf{I}_i has to be estimated. This is achieved by following a image miniature similarity approach suggested in [20]. This method requires a set of training volumes \mathbf{I}_j^c with annotated center positions \mathbf{c}_j^c in the atlas. Miniatures of $32 \times 32 \times 32$

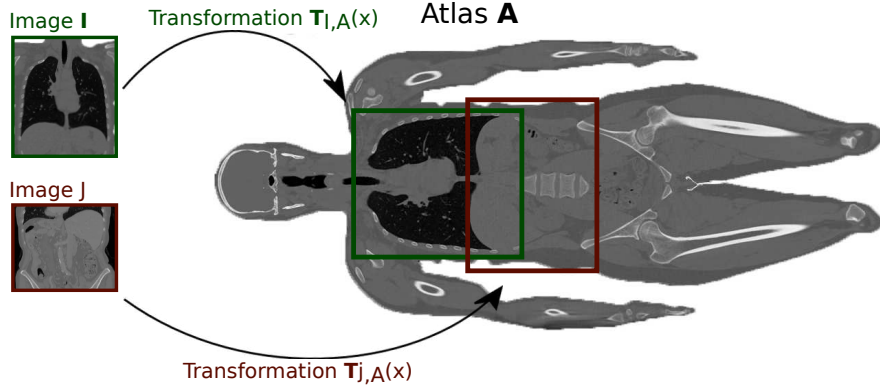


Figure 2.11: Illustration of the aim of the registration framework. Finding non rigid transformations from volumes that cover different regions of the human body to a common reference space or atlas. Figure adapted from [23].

voxels are computed for all \mathbf{I}_j^c and \mathbf{I}_i . A k-nearest neighbor (k-nn) search is performed to select the center positions $\mathbf{c}_1^c, \dots, \mathbf{c}_k^c$ of the k most similar miniatures. Miniature similarity is measured using NCC. To be robust against outliers only the 50 percent closest to the median of the selected positions are used to calculate the final position estimate. This set is denoted as Region of Trimmed Estimates (RTE) [23]. The position estimate \mathbf{c}_i is then calculated as denoted in Equation 2.40. Figure 2.12 shows the principle of the robust center estimation.

$$\mathbf{c}_i = \frac{\sum_{\mathbf{c}_j^c \in RTE} \mathbf{c}_j^c}{k/2} \quad (2.40)$$

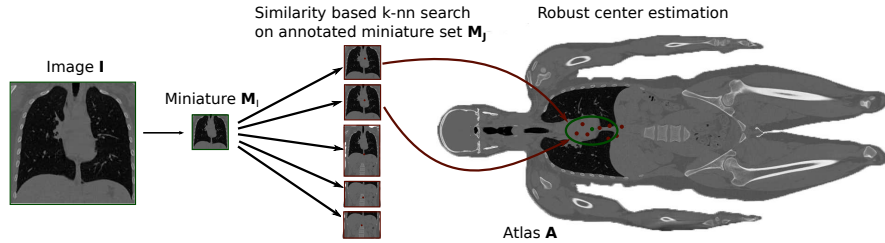


Figure 2.12: Robust center estimation algorithm. The miniature of a novel volume is compared to a set of miniatures where the center coordinates in the atlas are annotated. The center position is then computed by considering the annotations of the top k-nearest neighbors. Figure adapted from [23].

The region $\mathbf{R}_{i,A}$ of the atlas that covers the anatomical region of image \mathbf{I}_i is estimated using an iterative strategy. $\mathbf{R}_{i,A}$ is defined by two of points that form a bounding box $\mathbf{x}_{cor1,i}$ and $\mathbf{x}_{cor2,i}$, where $\mathbf{x}_{cor1,i}$ denotes the offset between the origin of \mathbf{A} and $\mathbf{R}_{i,A}$. After initialization of $\mathbf{x}_{cor1,i}$ and $\mathbf{x}_{cor2,i}$ with the dimension of \mathbf{I}_i the algorithm alternates between

1. Perform an affine registration to obtain the transformation $T_{i,A}^a$ of \mathbf{I}_i to $\mathbf{R}_{i,A}$

2. Updating $\mathbf{R}_{i,A}$ by applying the inverse affine transformation $T_{i,A}^{a,-1}$ to the corner coordinates, i.e. $\mathbf{x}_{cor1,i} = T_{i,A}^{a,-1}(\mathbf{x}_{cor1,i})$.

These two steps are iterated until the region estimation converges or a number of predefined maximum iterations is exceeded. [23].

A non-rigid registration from image \mathbf{I}_i to the final atlas region $\mathbf{R}_{i,A}$ using B-spline based FFD is performed to obtain the transformation $T_{i,A}^{nr}$ in the last step. The final result is a nonrigid transformation $T'_{i,A}$ of image \mathbf{I}_i to the atlas region $\mathbf{R}_{i,A}$ by computing

$$T'_{i,A} = T_{i,A}^{nr} \circ T_{i,A}^{a,-1}. \quad (2.41)$$

$T'_{i,A}$ then allows a mapping of a coordinate \mathbf{y} in I_i to the coordinate system of $\mathbf{R}_{i,A}$. This transformation is finally used to map coordinates from I_i to the coordinate system \mathbf{A} as

$$T_{i,A}(\mathbf{y}) = T'_{i,A}(\mathbf{y}) + \mathbf{x}_{cor1,i}. \quad (2.42)$$

Which results in a transformation $T_{i,A}$ that maps \mathbf{I}_i to \mathbf{A} so that

$$\mathbf{I}_i(\mathbf{y}) \approx \mathbf{A}(T_{i,A}(\mathbf{y})) \quad (2.43)$$

Please note that the transformation $T_{i,A}$ is not bijective, i.e. there is a corresponding position in the atlas for each coordinate of the source volume, but there might not be a corresponding coordinate in the source image for all coordinates of the atlas. We are thus able to transform coordinates only from the source volume to the atlas and hence as well to transform label volumes from the atlas to the source volume.

2.3.4 Discussion and Relation to Present Work

This section describes different components of a medical image registration system. Starting with image transformation that are used to deform a medical images, we introduce similarity measures such as NCC and NMI that are used to evaluate the quality of image alignment after a registration step. Finally we describe an approach that makes use of all described components that yields in finding a non-rigid transformation between medical images, that cover different parts of the human body (i.e. chest scans, abdominal scans), to a common reference space also called *atlas*. This approach allows the propagation of coordinates or landmarks from all registered source volumes to an atlas, as well as the propagation of a voxel wise labeling in the atlas to all volumes of the population. We will use the introduced concept to propagate segmentation estimations across the population through the atlas. These segmentation methods are also referred to as *atlas based* segmentation methods [61].

2.4 Markov Random Fields in Medical Image Segmentation

Within the scope of this thesis we propose an unsupervised atlas based segmentation method on super voxel basis that uses Markov Random Fields (MRFs) to encode spatial relationships between neighboring supervoxels. MRFs in this context provide a statistical framework that we

use to assign one label from a discrete set of segmentation labels $\mathcal{L} = \{l_1, \dots, l_M\}$ to each supervoxel of a volume.

2.4.1 Theoretical overview

Similar to [21] and [80] we view a MRF as a weighted undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, having N vertices $\mathcal{V} = \{v_1, \dots, v_N\}$ where each vertex has M states or labels and each state carries a weight that indicates the likelihood that a vertex is assigned with a specific label. Those weights are also referred to as *qualities* [21], [80]. All states of two adjacent vertices are fully connected by M^2 edges, again carrying label assignment qualities. An example with $N = 6$ vertices and $M = 3$ states for each vertex, resulting in $M^2 = 9$ edges for adjacent vertices and one possible label assignment is illustrated in Figure 2.13.

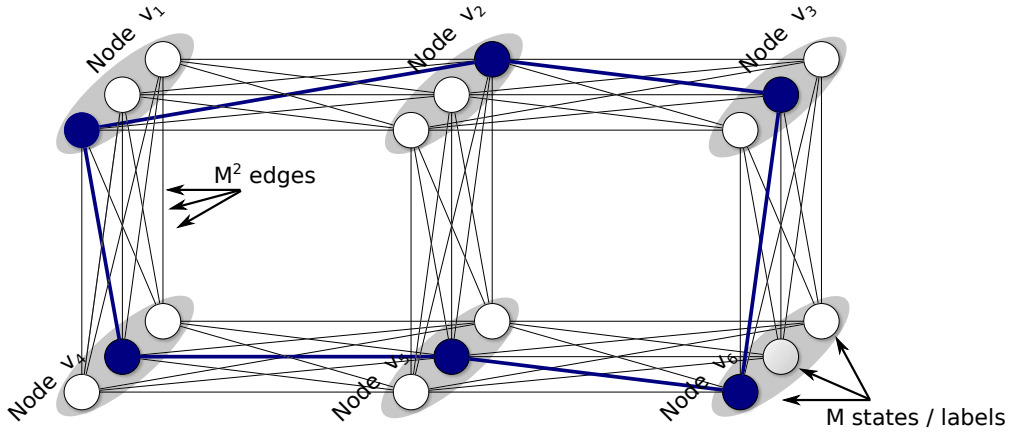


Figure 2.13: A MRF graph as used in this thesis. Each of the $N = 6$ vertices has $M = 3$ possible label assignments. Adjacent vertices are fully connected by $M^2 = 9$ edges reflecting energies of a specific label assignment between neighboring vertices. Blue vertices and edges indicate a possible label assignment. Figure adapted from [80] and [21].

Two vertices $v_i, v_j \in \mathcal{V}$ are said to be adjacent if $(v_i, v_j) \in \mathcal{E}$, we can thus denote the neighborhood of a vertex v_i as $\mathcal{N}(v_i) = \{v_j : (v_i, v_j) \in \mathcal{E}, \text{ where } i \neq j\}$ [36]. In other words the neighborhood of a vertex contains all vertices that are connected by at least one edge, i.e. $\mathcal{N}(v_2) = \{v_1, v_3, v_5\}$ of the illustrated graph in Figure 2.13. If we view interpret such a graphs vertices and their states as random variables $\mathbf{X} = \{X_1, X_2, \dots, X_N\}$, with one specific state configuration $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ from the set of discrete labels \mathcal{L} , then \mathbf{X} is said to be a random field having Markov property if

$$P(x_i) > 0, \quad \forall x_i \quad (2.44)$$

$$P(x_i | \mathbf{x}_{\mathcal{V} \setminus \{i\}}) = P(x_i | \mathbf{x}_{\mathcal{N}(i)}), \quad (2.45)$$

where $P(x_i)$ denotes the probability that X_i takes label x_i and $P(\mathbf{x})$ denotes the joint probability that a set of random variables \mathbf{X} is assigned with state configuration \mathbf{x} [36]. That means

a set of random variables is called a MRF if and only if that the probability of all states to be assigned to a random variable must be greater than zero and that a random variable conditionally depends only on its neighboring random variables [36].

Solving such a MRF, i.e. selecting one state for each vertex so that the sum of selected state and their connected edge qualities becomes maximal, means finding a Maximum A Posterior (MAP) distribution of $P(\mathbf{x})$, which is equivalent to finding a configuration of a Gibbs distribution that is maximal [21]. This problem is in general NP-hard, therefore different methods, such as second order cone programming [43], convergent tree-reweighted message passing [42] or belief propagation [86], that find an approximated solution are described in literature [21]. In this thesis the OpenGM toolbox ¹, that provides a C++ implementation of loopy belief propagation, proposed in [2], is used.

Suppose we have a $N \times M$ matrix U depicting the state qualities for each vertex. Furthermore we consider a graph with A connected vertices resulting in an $A \times M^2$ matrix B holding the edge qualities between neighboring vertices. The sum of qualities E of a state configuration \mathbf{x} is then defined as

$$E(\mathbf{x}) = \sum_{n=1}^N U(n, \mathbf{x}(n)) + \sum_{a=1}^A \mathbf{B}(a, \beta(\mathbf{B}, \mathbf{x}, a)) \quad (2.46)$$

where $\beta(\mathbf{B}, \mathbf{x}, a)$ identifies the column in B that represents the quality of the connected states of edge a [21].

Loopy belief propagation is capable of finding a state configuration \mathbf{x}^* so that the sum of qualities becomes minimal

$$\mathbf{x}^* = \arg \min_x E(\mathbf{x}). \quad (2.47)$$

Hence, U and B must be transformed so that they express label assignment costs rather than qualities in order to find an optimal labeling of the graph. The computational complexity of the loopy belief propagation algorithm depends linear on the number of edges A and quadratic on the number of possible states M of the graph [9].

Similar to [60] we will refer to state costs U for each vertex as *unary terms* since they affect only single vertices and to the edge costs B as *binary terms* since they hold costs of two (neighboring) vertices in the following sections of this thesis.

2.4.2 Discussion and Relation to Present Work

Several approaches that use MRFs to segment anatomical structures in medical image data have been recently proposed. Most of them in the context of brain matter segmentation in MR images [35], [76], [30], [62], [60]. But as well to segment cardiac structures [51] and bones in MR images [60] or organs of the abdominal part of the human body in CT data [60].

Unary terms are amongst others used to incorporate atlases by medical image registration [76], [30] and to model the intensity and feature distributions of tissue classes [62], [35], [51]. Binary terms are used to model the assumption that neighboring voxels should belong to the

¹<http://hci.iwr.uni-heidelberg.de/opengm2/>

same tissue class leading to smooth segmentation contours noise reduction [35], [76], [30], [62], [60], as well as to model relationships among neighboring time frames of temporal recorded medical image data [51].

MRFs are used in this work to find an optimal segmentation labeling of supervoxels. Unary terms are used to propagate segmentation estimates across spatially corresponding supervoxels of all volumes in the training data, where binary terms are used to model the assumption that spatially neighboring supervoxels within an image are likely to belong to the same segmentation class.

2.5 Summary

In this section techniques for texture description, unsupervised data analysis, medical image registration and MRFs in context of medical image segmentation that are used within the scope of this thesis have been described. Please note that the present chapter does not aim at providing an extensive survey of the addressed research fields but to give an overview of the components used in this work.

Two approaches to describe the texture of supervoxels have been described in Section 2.1. BVW-LBP [10] features build histograms of visual words trained in unsupervised manner from a set of LBP features sampled from multiple scales. Haralick features [33] encode different statistical properties of GLCMs which are sampled in different directions and offsets so that the resulting feature vector is independent to texture orientation. Both methods are used within the scope of this work to describe the texture of supervoxels.

Within Section 2.2, two unsupervised clustering approaches (k-means [31] and GMM clustering [9]) have been described to detect homogeneous classes in a set of objects. These methods are used in this work to find groups of similar supervoxel features that represent tissue prototypes occurring in medical images. Furthermore a method to decrease the dimensionality of a feature space has been introduced (PCA) to tackle the *curse of dimensionality* [56].

In Section 2.3 components required for a medical image registration framework have been shown. Transformation methods (affine, rigid and non-rigid) to register images that cover the same anatomical region of the human body and similarity functions (NCC and NMI) have been introduced, followed by a framework to register medical images covering different parts of the human body to a central reference space [23]. This framework is used to establish voxel-wise spatial correspondences between all images of the training data to a central reference space and furthermore to propagate supervoxels and segmentation labels across the image population.

Finally, Section 2.4 has given a theoretical overview of MRFs and shown how they are used in the context of medical image segmentation to combine atlas based segmentations with local image information and to model constraints on spatially neighboring voxels or supervoxels.

Methodical Approach

This chapter describes the main contributions of this thesis in detail. In Section 3.1 we propose a method for the unsupervised segmentation of anatomical structures on a supervoxel level. Section 3.2 describes an approach that learns to predict healthy and pathological tissues from data that is available from clinical routine. Finally, the key components of both methods are summarized in Section 3.3.

3.1 Unsupervised Medical Image Segmentation on Supervoxel Level

In this section we propose an atlas based medical image segmentation method, that identifies anatomical structures in a set of images in an unsupervised manner, i.e. without having prior segmentation knowledge such as manual expert annotations of anatomical structures. The method takes a set of medical images as input and learns prototypes of image regions by texture feature clustering. These prototypes are used to compute an initial segmentation estimate for all images. After registration of all images to an atlas, a labeling in the atlas is learned based on a majority vote of all training images. The final segmentation of an image is then obtained by combining the computed labeling of the atlas as a prior together with the individual segmentation of an image. Figure 3.1 provides an overview of the workflow.

In the following sections we describe the whole image processing pipeline in detail. We start by giving a formal problem definition in Section 3.1.1. After describing the image normalization pipeline, in which all volumes are registered to a central template space or *atlas* to obtain supervoxels in all images of the population in Section 3.1.2, we describe feature extraction and identification of prototypes of supervoxels in the data set in Section 3.1.3. Section 3.1.4 covers the usage of MRFs to find a latent atlas labeling, where Section 3.1.5 describes the labeling of anatomical structures in novel images, where all previous described components are used.

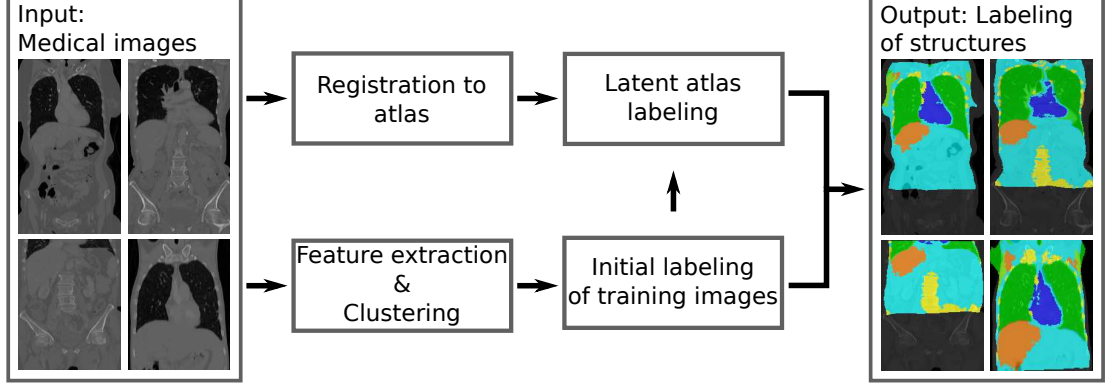


Figure 3.1: Workflow overview of the method proposed for unsupervised atlas based anatomical structure segmentation in medical images.

3.1.1 Objects, Notation and Problem Definition

Notation We use the following notation in the remaining part of this section: We have a set of N volumes $\mathbf{I}_i \in \mathbb{R}^{X_i \times Y_i \times Z_i}$, where $i = 1, \dots, N$ and an atlas volume $\mathbf{A} \in \mathbb{R}^{X_A \times Y_A \times Z_A}$. The constants $(X, Y, Z) \in \mathbb{N}^+$ denote the dimensions of an image in each direction. A voxel coordinate is specified by a vector $\mathbf{x} = (x, y, z) \in \mathbb{N}^+$, where $x \leq X, y \leq Y, z \leq Z$.

After registration of all images to the atlas, we assume correspondence of voxels across all images. We view \mathbf{A} as a graph with K nodes, where each node is a supervoxel, and supervoxels are linked by a connectivity structure that expresses the spatial neighborhood of a supervoxel. Let k be the index of each node in \mathbf{A} . From the registration, each node k of the atlas has spatially corresponding supervoxels in a sub-set of J^k images with indices $\mathbf{i}_j^k \subseteq \{1, \dots, N\}$, where $j = 1 \dots J^k$.

Correspondingly we view each image \mathbf{I}_i as a graph with K_i^I nodes that correspond to from the atlas propagated supervoxels, where each node k_i^I spatially corresponds to a node k in the atlas. For simplicity we will use only the index k from now on, and will refer to the node in image i corresponding to the atlas node k as $\langle k, i \rangle$.

Problem statement We want to solve a labeling problem that assigns each node in each volume a label corresponding to an anatomical region. That is we want to find the labeling

$$\mathcal{L} : \langle k, i \rangle \mapsto l \quad (3.1)$$

where $l \in \{1, \dots, L\}$ is an anatomical region label. Therefore we have a labeling for each node $\langle k, i \rangle$ in each image. We also learn a labeling in the template space that serves as a latent prior shared across all individual volumes. We will call the entirety of the labels assigned to the nodes in the template space \mathcal{L}^* the *latent atlas*:

$$\mathcal{L}^* : k \mapsto l \in \{1, \dots, L\} \quad (3.2)$$

3.1.2 Image Normalization Pipeline

The normalization pipeline includes data acquisition and image reorientation and aims at finding a non-rigid mapping of all images to the central reference space so that we can propagate supervoxels to all images through the atlas. Figure 3.2 gives an overview of all components of the image normalization pipeline. A detailed description of data acquisition and volume reorientation is given in Section 4.1.3 of this thesis.

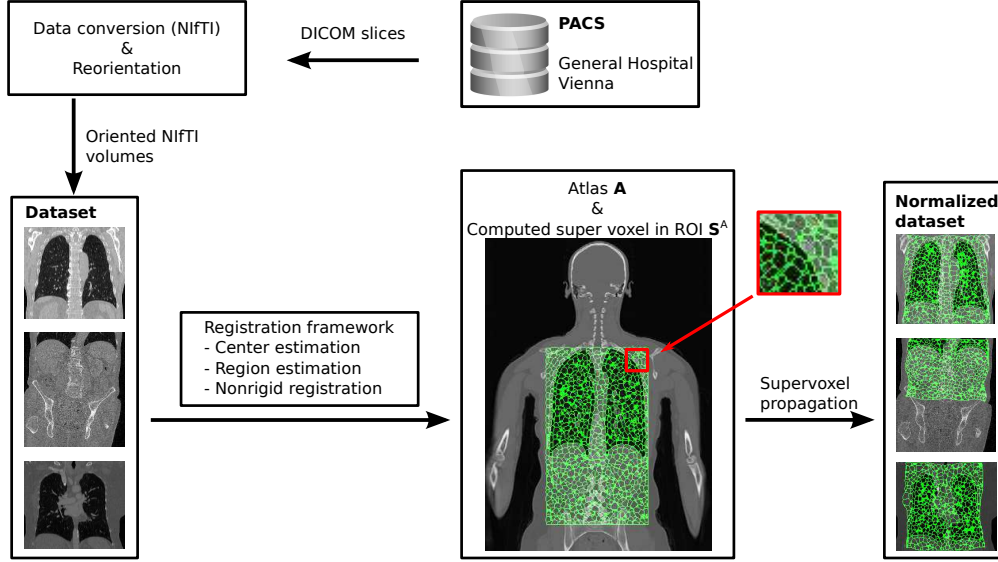


Figure 3.2: Overview of the image normalization pipeline used in this work. All images are extracted from a PACS system. After conversion from DICOM to NIfTI file format and reorientation to *ALS* convention, all volumes are registered to an atlas. The atlas is quantized into supervoxels, which are propagated to all volumes. This results in sets of supervoxels that correspond across parts of the population, and allow for group-wise learning of organ maps.

Registration to a shared reference space In order to achieve spatial correspondence between anatomical regions across the image population we register all volumes of the dataset to one central reference space, to which we refer as *atlas A*. Since the dataset contains volumes that cover different parts of the human body, such as the chest, the abdomen or both of them, a method that locates the area of an image in an atlas is required. We use an approach proposed by Dorfer et. al. in [23], to overcome this problem.

This method yields a transformation $T_{i,A}$ that maps all coordinates \mathbf{x} from a source volume \mathbf{I}_i to the coordinate system of \mathbf{A} . Finding this transformation is based on a three step process, which is described in detail in Section 2.3.3 of this thesis and summarized as follows:

1. Estimate the center coordinates \mathbf{c}_i of the novel volume in \mathbf{A} .
2. Find the region $\mathbf{R}_{i,A}$ in \mathbf{A} that is covered by \mathbf{I}_i and compute an affine transformation $T_{i,A}^a$.

3. Compute a non-rigid transformation $T_{i,A}^{nr}$ from the affine transformed image \mathbf{I}_i to $\mathbf{R}_{i,A}$ and generate $T_{i,A}(\mathbf{y})$ that transforms a coordinate \mathbf{y} of \mathbf{I}_i as

$$T_{i,A}(\mathbf{y}) = (T_{i,A}^{nr}(\mathbf{y}) \circ T_{i,A}^{a,-1}(\mathbf{y})) + \mathbf{x}_{cor1,i}, \quad (3.3)$$

where $\mathbf{x}_{cor1,i}$ is the offset between the origin of \mathbf{I}_i and $\mathbf{R}_{i,A}$

The resulting transformation $T_{i,A}$ maps each coordinate of \mathbf{I}_i to the coordinate system of \mathbf{A} so that

$$\mathbf{I}_i(\mathbf{y}) \approx \mathbf{A}(T_{i,A}(\mathbf{y})) \quad (3.4)$$

For this purpose, center positions in \mathbf{A} in a set of 200 volumes, covering all occurring body regions in the dataset, are annotated. Figure 3.3 shows the atlas volume and the spatial distribution of annotated center positions. Using this set of annotations and the method described allows the registration of all volumes in the dataset to \mathbf{A} .

The NiftyReg-toolbox ¹ is used to perform the registrations which implements an approach from Ourselin et. al. [59] for affine registration and provides an implementation of B-spline based registration from Rueckert et. al. [63].

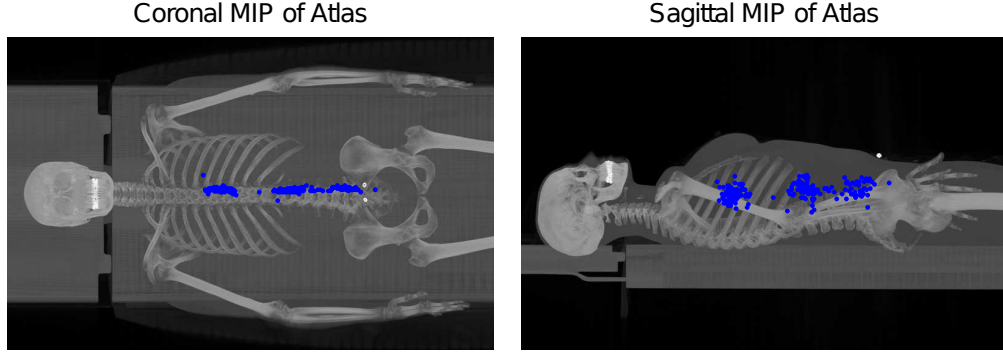


Figure 3.3: Annotated center position annotations in blue, shown in Maximum Intensity Projections (MIP) of the cropped atlas volume in coronal (left) and sagittal view (right).

Deriving supervoxels in the reference space In order to reduce the number of objects to analyse in an image we quantize \mathbf{A} into so called *supervoxels*. A supervoxel algorithm computes an oversegmentation of a volume that merges neighboring voxels into homogeneous groups of voxels (supervoxels) so that voxels in one group have similar texture properties while the edges of supervoxel preserve the boundaries of objects in a volume [38]. We use an algorithm proposed by Holzer et al. in [38] to compute an oversegmentation \mathbf{S}^A of \mathbf{A} into a set of K disjunct supervoxels, so that

¹<http://www.nitrc.org/projects/niftyreg/>

$$\mathbf{S}_u^A \subset \mathbf{A} \quad (3.5)$$

$$\mathbf{S}_u^A \cap \mathbf{S}_v^A = \emptyset, \quad \forall u \neq v : \mathbf{A}, \quad (3.6)$$

where u, v depict indices of supervoxels. The method of Holzer et al. uses the phase of the monogenic signal, which contains local structural information of a volume to detect edges and then applies k -means clustering based on the edge cues and the spatial location of voxels to derive supervoxels. The number of clusters represents the number of desired supervoxels, which enables us to control the average size of resulting supervoxels. Using the monogenic signal rather than voxel intensities leads to being independent to brightness and contrast as well as being robust against noise in the volume [38].

Deriving supervoxels in the image population In order to derive supervoxels with spatial correspondences to \mathbf{A} in a volume \mathbf{I}_i we use the computed transformations $T_{i,A}$ to propagate the oversegmentation volume of the atlas \mathbf{S}^A . The supervoxel volume \mathbf{S}_i of an image \mathbf{I}_i is thus computed by

$$\mathbf{S}_i(\mathbf{y}) = \mathbf{S}^A(T_{i,A}(\mathbf{y})), \quad (3.7)$$

where \mathbf{S}_i contains a set of K_i^I disjunct supervoxels. $K_i^I \leq K$ since the volumes are only partially overlapping with the region of interest in the atlas. Analysing supervoxels rather than voxels of medical images reduces the computational complexity substantially, since $K_i^I \ll X_i \times Y_i \times Z_i$.

The propagation of supervoxels from \mathbf{A} to \mathbf{I}_i assures that for each supervoxel k_i^I exists exactly one spatially corresponding supervoxel k in \mathbf{A} . And furthermore that for each node k in the atlas exists a sub set of images that have a spatially to k corresponding supervoxel.

We will thus refer to supervoxels in all images and the atlas with index k and refer to a node in \mathbf{I}_i that spatially corresponds to k in the atlas as $\langle k, i \rangle$. Furthermore, we denote the indexes of J^k images that have a spatially to k corresponding supervoxel as \mathbf{i}_j^k , where $j = 1 \dots J^k$.

3.1.3 Learning Supervoxel Texture Prototypes

After the the registration step and the propagation of supervoxels to all images, we aim to find classes of supervoxels with similar texture properties that reflect anatomical structures in the image data. To achieve this goal we extract BVW-LBP features as described in detail in Section 2.1.3 of this thesis, since this approach allows to control the impact of local intensity and contrast on the resulting feature vector.

Feature extraction We compute LBP3d/CI descriptors, denoted as

$$\mathbf{d}(\mathbf{x}) = [\mathbf{d}(\mathbf{x})_{LBP3d}, c_c \mathbf{d}(\mathbf{x})_C, c_i \mathbf{d}(\mathbf{x})_I] \quad (3.8)$$

for a voxel \mathbf{x} in a volume. We denote the entirety of feature space as \mathcal{D} , and the respective entirety of intensity and contrast features as \mathcal{D}_I and \mathcal{D}_C . Since \mathbf{d}_C and \mathbf{d}_I are independent variables having different ranges, both measures are standardized to the range of $[0, 1]$ by

$$\mathbf{d}_C(\mathbf{x}) = \frac{\mathbf{d}_C(\mathbf{x}) - \min(\mathcal{D}_C)}{\max(\mathcal{D}_C) - \min(\mathcal{D}_C)} \quad \mathbf{d}_I(\mathbf{x}) = \frac{\mathbf{d}_I(\mathbf{x}) - \min(\mathcal{D}_I)}{\max(\mathcal{D}_I) - \min(\mathcal{D}_I)}. \quad (3.9)$$

A randomly sampled subset of 150000 features is clustered using k -means to obtain the visual vocabularies \mathbf{W}_K in the size of $J = 300$ visual words. Using LBP3d/CI leads to feature vectors and visual words having $d^{lbp3D} = 28$ dimensions. All voxels are replaced with the index of to their LBP3d/CI feature vectors nearest cluster. The distance between a feature vector $\mathbf{d}(\mathbf{x})$ and a visual word \mathbf{W}_j is measured using the Euclidean distance $d_{\mathcal{E}}$, given in [78] as

$$d_{\mathcal{E}} = \sqrt{\sum_1^{d^{lbp3D}} (\mathbf{d}(x) - \mathbf{W}_j)^2}. \quad (3.10)$$

The resulting feature vector $\mathbf{f}_{\langle k, i \rangle}^{LBP}$ of a supervoxel is then depicted as the histogram of visual words \mathbf{W}_J that occur in the supervoxel. Let $|\langle k, i \rangle|$ depict the number of voxels that belong to $\langle k, i \rangle$, we normalize the visual word histogram $\mathbf{f}_{\langle k, i \rangle}^{LBP}$ by

$$\mathbf{f}_{\langle k, i \rangle}^{LBP} = \frac{\mathbf{f}_{\langle k, i \rangle}^{LBP}}{|\langle k, i \rangle|} \quad (3.11)$$

in order to be invariant to the supervoxels size.

Inspired by the work of Burner et al. we compute visual words on $s = (1, 2, 3, 4)$ different scales to be able to perceive the granularity of the texture [10]. This is achieved by down sampling the volumes by the factor of the respective scale, extracting LBP3d/CI descriptors, clustering features to build visual words and building normalized occurrence histograms on each scale. The final feature vector is then built by concatenation of occurrence histograms of each scale. We will refer to the entire feature space that holds all supervoxel features as \mathcal{F}^{LBP} .

Unsupervised clustering In the next step we aim to identify groups of supervoxels with similar texture properties in an unsupervised manner. Those groups are expected to reflect tissue prototypes of anatomical structures in the underlying image population. To achieve that we partition a randomly sampled subset of $M = 100000$ supervoxel features $\mathcal{F}'^{LBP} \subseteq \mathcal{F}^{LBP}$ into L anatomical structure classes \mathcal{C}_l , where $l \in \{1, \dots, L\}$ using GMM clustering.

Here, one cluster \mathcal{C}_l of is described as Gaussian distribution $\mathcal{N}(\mu_l, \Sigma_l)$ having mean μ_l and covariance matrix Σ_l as well as π_l , which depicts the prior probability that an observation is drawn from cluster l .

Finding the parameters μ_l , Σ_l and π_l of all classes, so that the sampled feature space \mathcal{F}'^{LBP} is explained as good as possible requires maximizing the following log likelihood function

$$\ln p(\mathcal{F}'^{LBP} | \pi, \mu, \Sigma) = \sum_{m=1}^M \ln \sum_{l=1}^J \pi_l \mathcal{N}(\mathcal{F}'_m^{LBP} | \mu_l, \Sigma_l). \quad (3.12)$$

An approach to optimize this function is given in Section 2.2.2 of this work. After fitting the model to the sampled feature space and estimating the underlying Gaussian mixtures \mathcal{C}_L , the posterior probability that a supervoxel $\langle k, i \rangle$ with feature vector $\mathbf{f}_{\langle k, i \rangle}^{LBP}$ is drawn from class \mathcal{C}_l is given by

$$p(\langle k, i \rangle \mapsto l) = \frac{\pi_l \mathcal{N}(\mathbf{f}_{\langle k, i \rangle}^{LBP} \mid \mu_l, \Sigma_l)}{\sum_{i=1}^L \pi_i \mathcal{N}(\mathbf{f}_{\langle k, i \rangle}^{LBP} \mid \mu_i, \Sigma_i)}. \quad (3.13)$$

Hence we calculate $K_i^I \times L$ dimensional assignment probability matrices \mathbf{M}^i for each volume, which stores the probability that a supervoxel $\langle k, i \rangle$ belongs to the anatomical structure l as follows

$$\mathbf{M}_{\langle k, i \rangle, l}^i = p(\langle k, i \rangle \mapsto l). \quad (3.14)$$

Figure 3.4 gives an overview of the processing steps required to learn supervoxel feature prototypes and to obtain anatomical region label assignment probability matrices. Assigning each supervoxel of a volume with the index of the cluster with highest posterior assignment probability can also be interpreted as an initial segmentation estimate of a volume. While we use the collection of all cluster assignment probabilities to find a labeling in the atlas space in Section 3.1.4, initial segmentation estimates build one component while finding a labeling of individual volumes in Section 3.1.5.

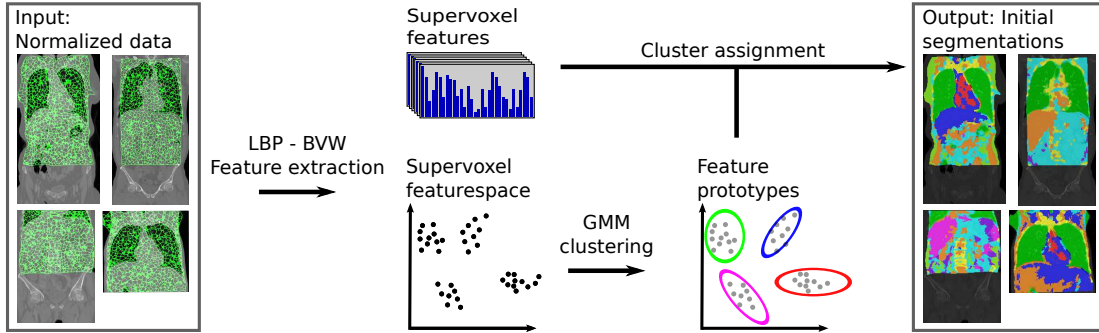


Figure 3.4: Overview of the processing steps to learn supervoxel feature prototypes that are expected to reflect anatomical structures. In the first step supervoxel texture features are extracted from the training data. GMM clustering is applied to the sampled feature space in the second step. Assigning each supervoxel with the cluster index of highest probability leads to an initial segmentation of each volume. These cluster assignment probabilities are also used during labeling of the atlas.

3.1.4 Finding a Latent Atlas Labeling

In the next step we compute a segmentation labeling for each node in the template space, which serves as prior segmentation estimation when labeling an individual image in the final step of

the image segmentation process.

We construct a MRF with K nodes, representing supervoxels in the atlas, connected by E undirected, weighted edges. Each node has L possible states which reflect the label assignment of a node. An edge between two nodes models spatial neighborhood of supervoxels. The label assignment probability matrices \mathbf{M} are used to derive qualities for unary terms $\mathbf{U} \in \mathbb{R}^{K \times L}$, while we binary terms $\mathbf{B} \in \mathbb{R}^{K \times K \times E}$ are used to encode qualities of label configurations of two spatially neighboring nodes. The following paragraphs describe how the components of the MRF are derived.

Topology The topology of the MRF models the neighborhood of the supervoxels in the atlas and is thus based on the spatial distance of the center of gravity of a supervoxel. Given a node k . We build the neighborhood \mathcal{N}_k by forming edges between u and its T nearest neighbors, where distance between two nodes is derived by calculating the Euclidean distance of their centers of gravity. Iterating over all nodes and removing duplicates results in a set of \mathcal{E}_e undirected edges, where $e \in \{1, \dots, E \leq K \cdot T\}$ and an edge is defined by the pair of nodes it connects. I.e. $\mathcal{E}_e = \{u, v\}$, where $u, v \in K$.

Unary Terms Unary terms are used to encode the prior probability that a node k is assigned with label l . Since each node k of the atlas has a spatially corresponding node $\langle k, i \rangle$ in a subset of J^k images with indexes \mathbf{i}_j^k , we build unary terms by sampling the probabilities of these corresponding nodes. An entry in the $K \times L$ dimensional unary term \mathbf{U} matrix is thus computed as

$$\mathbf{U}(k, l) = \frac{\sum_{j=1}^{J^k} \mathbf{M}^{\mathbf{i}_j^k}(\langle k, \mathbf{i}_j^k \rangle, l)}{J^k} \quad (3.15)$$

so that

$$\sum_{l=1}^L \mathbf{U}(k, l) = 1, \forall k. \quad (3.16)$$

In other words, the label assignment qualities of a node k in the atlas are derived by a majority vote of all nodes in the image population that spatially correspond to k .

Binary Terms Binary terms \mathbf{B} are used to fully connect all possible label configurations of two neighboring nodes. This allows to model relationships between spatially neighboring supervoxels. We use the binary terms to model the assumption that two spatially neighboring supervoxels are likely to belong to the same anatomical structure, i.e. we encourage similar label assignments of neighboring nodes. This is achieved by creating the $L \times L \times E$ matrix \mathbf{B} , where an $L \times L$ entry that holds qualities all possible label configurations of two nodes that are connected by an edge e and is computed as

$$\mathbf{B}_e = \alpha \cdot \mathbf{I}_L, \quad (3.17)$$

where \mathbf{I}_L depicts the identity matrix of size L and α is a scalar weighting factor that is used to control the impact of binary terms.

Solving the MRF An implementation of loopy believe propagation [86] proposed by Andres et al. in [2] is used in this work for finding an approximately optimal solution of the MRF. This implementation allows to find a configuration of a MRF with a minimum sum of qualities, which requires that unary and binary terms reflect costs rather than probabilities of label assignments. \mathbf{U} and \mathbf{B} are thus converted to costs by

$$\mathbf{U} = \lfloor 1000 * (1 - \mathbf{U}) \rfloor \quad \mathbf{B} = \lfloor 1000 * (1 - \mathbf{B}) \rfloor. \quad (3.18)$$

The cost function $\mathcal{Q}(\mathcal{A})$ of a configuration \mathcal{A} that assigns one label to each node of the MRF is then given by

$$\mathcal{Q}(\mathcal{A}) = \sum_{k=1}^K \mathbf{U}(k, \mathcal{A}(k)) + \sum_{e=1}^E \mathbf{B}(\mathcal{A}(e_1), \mathcal{A}(e_2), e) \quad (3.19)$$

where e_1, e_2 depict the nodes that are connected from e .

The latent atlas labeling

$$\mathcal{L}^* : k \mapsto l \in \{1, \dots, L\} \quad (3.20)$$

that holds the assignment of one of L anatomical region labels for all nodes k in the atlas \mathbf{A} is then obtained by finding a configuration with minimal costs, denoted as

$$\mathcal{L}^* = \arg \min_{\mathcal{A}} \mathcal{Q}(\mathcal{A}). \quad (3.21)$$

3.1.5 Finding Individual Labelings using the Latent Atlas as Prior

In the final step of the image segmentation process we use the information computed in the previous steps to obtain a labeling of all nodes in an individual image \mathbf{I}_i .

Again, we construct a MRF having K_i^I nodes representing the supervoxel that occur in \mathbf{I}_i . Similar to the MRF for labeling the atlas, edges are formed to encode spatially neighboring supervoxels. Unary terms are used to combine the a priori estimated atlas labeling \mathcal{L}^* of a node and the label assignment probability $\mathbf{M}_{\langle k, i \rangle}^i$. Binary terms are used to model the assumption that neighboring supervoxels are likely to belong to the same anatomical structure. In comparison to the former MRF here we add knowledge about the average supervoxel intensity into the model. The following paragraphs describe in detail how each component of the MRF is computed.

Topology The topology of the MRF for labeling an individual volume is derived similar to forming the topology when constructing the MRF for labeling the atlas. We construct edges to the T spatially nearest neighbors of each node $\langle k, i \rangle$, where spatial distance is measured between the centers of gravity of two nodes. After removing duplicates we derive a set of undirected edges \mathcal{E}_e , where $e \in \{1, \dots, E \leq K_i^I \cdot T\}$ and an edge is defined by the pair of nodes it connects. I.e. $\mathcal{E}_e = \{u, v\}$, where $u, v \in \langle k, i \rangle$.

Unary Terms The $K_i^I \times L$ unary terms \mathbf{U} for labeling an individual volume are based on two components. With the first component we incorporate the a priori computed latent atlas labeling \mathcal{L}^* to the model as

$$\mathbf{U}_{\langle k, i \rangle, l}^{\mathcal{L}^*} = \begin{cases} 1, & \mathcal{L}_k^* == l \\ 0, & \text{otherwise.} \end{cases} \quad (3.22)$$

The second component is derived from the anatomical structure class assignment probabilities, denoted as

$$\mathbf{U}_{\langle k, i \rangle, l}^{\mathcal{C}} = \mathbf{M}^i(\langle k, i \rangle, l). \quad (3.23)$$

Finally both components are combined using a the mixing coefficient β as

$$\mathbf{U}_{\langle k, i \rangle, l} = \beta \cdot \mathbf{U}_{\langle k, i \rangle, l}^{\mathcal{L}^*} + (1 - \beta) \cdot \mathbf{U}_{\langle k, i \rangle, l}^{\mathcal{C}}. \quad (3.24)$$

Binary Terms The $L \times L \times E$ binary terms \mathbf{B} are used to encourage similar label assignments between two neighboring nodes u, v connected by edge e . In comparison to labeling two nodes in the atlas, we take intensity information of the connected supervoxels into account. I.e. we encourage similar label assignments of two neighboring nodes with similar average intensity, but penalize similar label assignments on edges that connect nodes with high average intensity differences.

Let $f_{\langle k, i \rangle}^I$ depict the average intensity of a supervoxel and $\Delta f_e^I = |f_u^I - f_v^I|$ the absolute difference of the average intensities of two supervoxels u, v that are connected by edge e . Let $\Delta \mathcal{F}^I$ denote the entirety of average intensity differences sampled over all edges $e \in \mathcal{E}$. The intensity difference indicator d_e of an edge is then computed by

$$d_e = \frac{\Delta f_e^I - \min(\Delta \mathcal{F}^I)}{\max(\Delta \mathcal{F}^I) - \min(\Delta \mathcal{F}^I)}, \quad (3.25)$$

so that d_e describes the intensity difference with respect to all other differences of connected nodes and is in the range of $[0, 1]$. To obtain a weighting of d_e so that similar label assignments are only encouraged if the intensity difference is low compared to all sampled intensity differences the final edge weighting factor w_e is calculated as given in Equation 3.26. The weighting function is illustrated in Figure 3.5.

$$w_e = e^{-\frac{d_e}{0.2}} \quad (3.26)$$

Binary terms of an edge e are then computed as

$$\mathbf{B}_e = \gamma \cdot w_e \cdot \mathbf{I}_L, \quad (3.27)$$

where \mathbf{I}_L again depicts the identity matrix of size L and γ denotes a scalar weighting factor to control the impact of binary terms of the MRF.

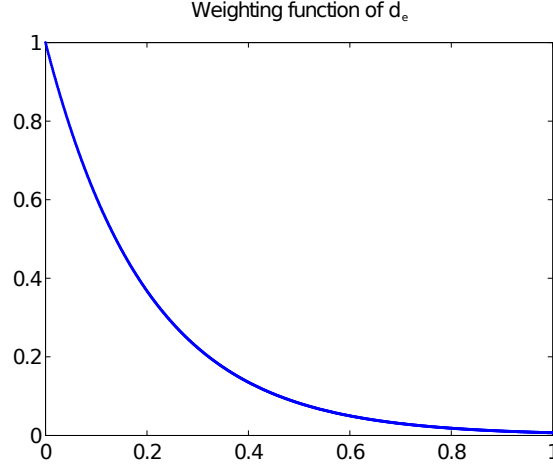


Figure 3.5: Weighting function that is applied to d_e , so that similar label assignments between nodes are only encouraged if the difference of average intensities is low with respect to all sampled intensity differences.

Solving the MRF Solving the MRF of individual volumes is performed similar to solving the MRF of the reference space, described in Section 3.1.4. Binary and Unary terms are transformed from carrying probabilities to costs in a first step. The labeling $\mathcal{L} : \langle k, i \rangle \mapsto l \in \{1, \dots, L\}$ which assigns one of L anatomical structure labels to each node $\langle k, i \rangle$ of the MRF and thus to each supervoxel of a volume is then derived by finding a configuration \mathcal{A} of the MRF with minimal costs $\mathcal{Q}(\mathcal{A})$, given as

$$\mathcal{Q}(\mathcal{A}) = \sum_{k=1}^{K_i^I} \mathbf{U}(k, \mathcal{A}(k)) + \sum_{e=1}^E \mathbf{B}(\mathcal{A}(e_1), \mathcal{A}(e_2), e), \quad (3.28)$$

where e_1, e_2 depict the nodes that are connected from e . Again, we apply loopy believe propagation [86] to find an optimal solution for \mathcal{L} .

$$\mathcal{L} = \arg \min_{\mathcal{A}} \mathcal{Q}(\mathcal{A}). \quad (3.29)$$

Figure 3.6 illustrates the resulting workflow of the approach proposed in this section for the unsupervised segmentation of anatomical structures in medical images.

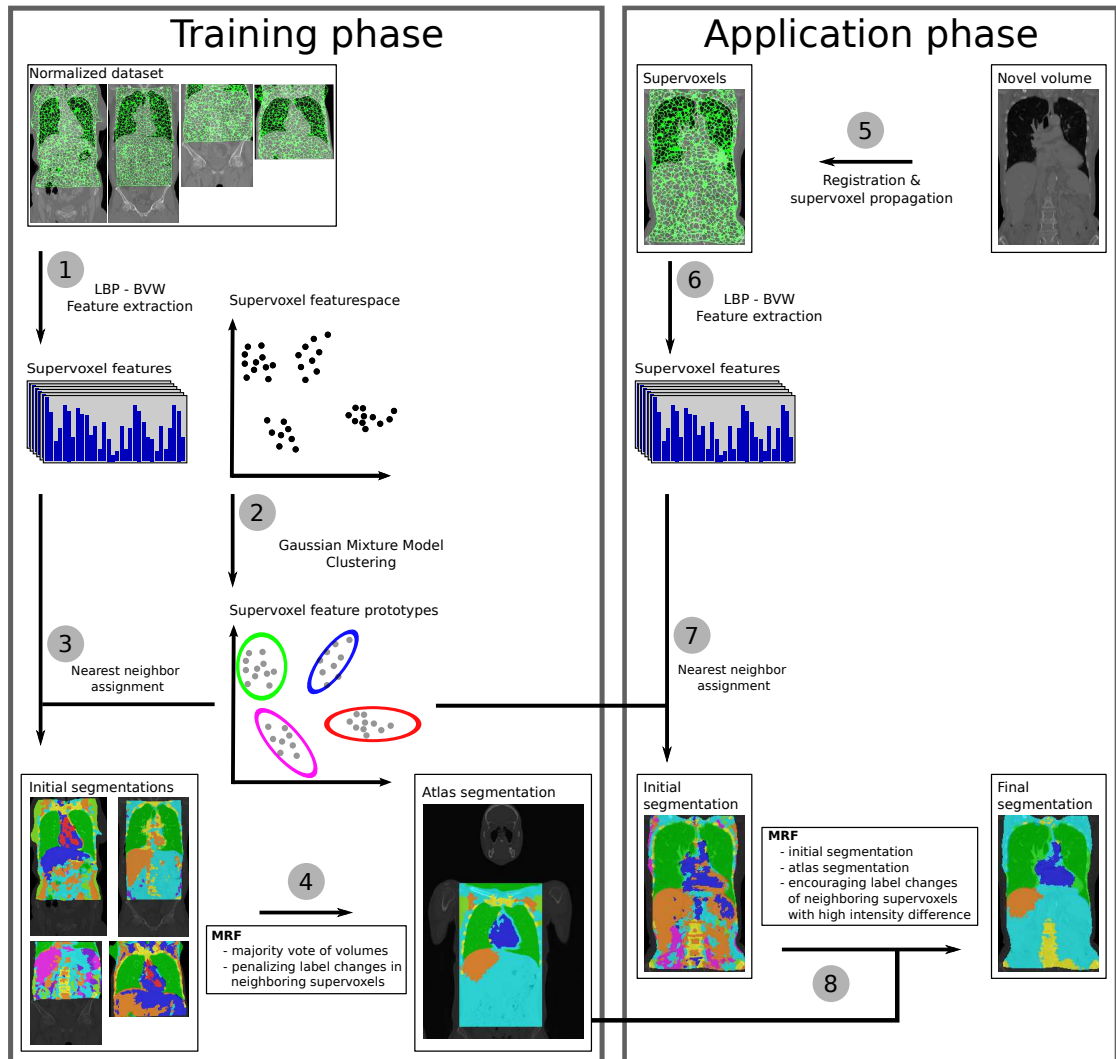


Figure 3.6: Overview of the image segmentation framework. In the training phase supervoxel features are extracted after all images are registered to the atlas (1). The resulting feature space is clustered using GMM clustering to find prototypes of occurring supervoxel features (2). In (3) all supervoxels are assigned with the index of the cluster with maximal assignment probability to compute initial segmentations of all volumes of the training data. In the final step of the training phase a MRF is created to find a labeling of the atlas (4). Segmenting a novel volume requires the registration to the atlas (5), feature extraction (6), cluster assignment to compute the initial segmentation (7) and finally a MRF that incorporates the atlas segmentation, the initial segmentation and models spatial neighborhood assumptions of supervoxels (8).

3.2 Weakly Supervised Classification of Pathologies

In this section we propose a weakly supervised learning method that is capable of classifying tissues in medical images based on the data that is generated during clinical routine:

1. A set of medical images, where each image typically shows different pathologies
2. A corresponding set of pathology terms or *clinical findings* that describe the underlying diseases and pathological observations. They are extracted from the radiological reports corresponding to the images.

From this data we know which pathologies occur in an image, but we do not know which regions are affected by each of the pathologies. The aim of the algorithm proposed is to establish this link based on the imaging data and corresponding pathology terms. Figure 3.7 illustrates the aim of the approach proposed (a), as well as the problem setting and given data in (b).

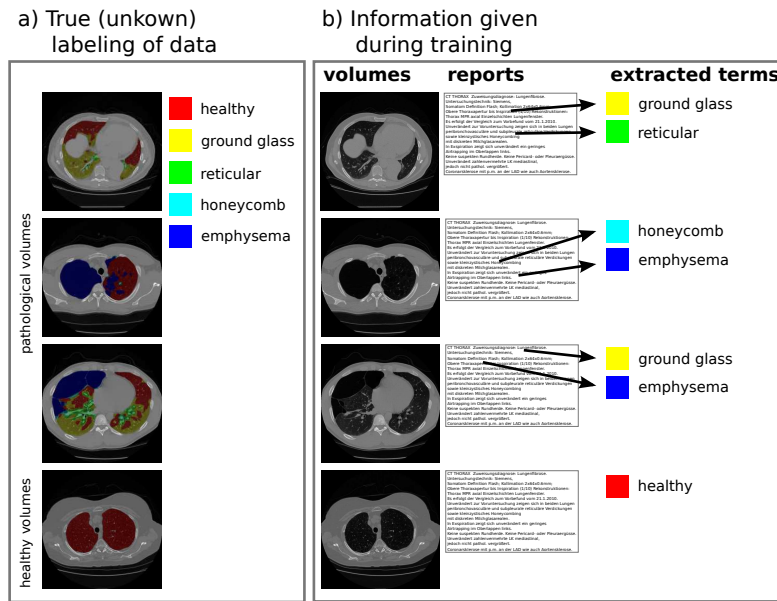


Figure 3.7: Illustration of the problem setting. We aim in finding the true (unknown) labeling of tissues in a medical volume. The dataset contains volumes with multiple tissue types as well as healthy controls (a). Figure (b) illustrates the data given during training. A set of medical images with corresponding reports, where pathological terms are extracted that describe the occurrence of pathologies in the images.

The problem of finding correspondences between image regions and textual image labels has been addressed by Duygulu et al. in [24]. Here we briefly review their approach since our method is based on their methodology. They have a set of 2-dimensional images and to each image a corresponding set of keywords that describes objects that occur in an image. Their approach aims at establishing correspondences between keywords and objects so that both entities

can be linked to each other and furthermore that objects in novel images can be identified and annotated with the established correspondences. They segment each image into regions using Normalized Cuts [65] and describe each region following the principle of BVW. After that all sampled region features are clustered to obtain a discrete set of image region feature prototypes. Finally, they use co-occurring prototypes and aligned keywords in images of the training data to establish a probability table that predicts a keyword given an image region prototype [24]. We follow their approach and adapt it to fit our problem setting.

To map pathologies occurring in an images report to specific regions of an image (classification), the images are quantized into supervoxels, which build the basis for further analysis. Our approach has two phases.

In the **training phase** each supervoxel of an image is labeled with all terms that occur in the report of an image. We extract texture features of all supervoxels in the training data and partition this sampled feature space using clustering techniques as introduced in Section 2.2 of this work. By mapping supervoxels and their aligned (pathological term) labels to the computed clusters based on their feature vectors, we derive a distribution of pathological labels in each partition of the feature space. We use this information to create a probability table that holds conditional probabilities of pathological labels given an observation of a specific cluster.

In the **application phase**, supervoxels in a novel (unseen) volume are then classified by mapping their extracted texture feature vectors to the computed clusters and assigning the labels with highest probabilities.

3.2.1 Problem Definition

We are given a set of N images (volumes) $\mathcal{I} = \{\mathbf{I}_1, \dots, \mathbf{I}_N\}$, where each image \mathbf{I}_i carries a set of tissue class labels (e.g. healthy tissue, ground glass, emphysema,...) \mathcal{T}_i , indicating which classes of tissue occur in the image. The set of T unique tissue labels is denoted as \mathcal{T} , so that $\mathcal{T}_i \subseteq \mathcal{T}$ holds. Furthermore, each image is quantized into S_i supervoxels, where supervoxel j in image i is denoted as $s_{i,j}$. The true tissue class label $l_{i,j}$ of a supervoxel is unknown during training. Instead each supervoxel is assigned with a set of *weak* labels $\mathcal{T}_{i,j}$, derived from the labels that are assigned to its volume $\mathcal{T}_{i,j} = \mathcal{T}_i$.

We aim at estimating the true labeling $l'_{u,j}$ of each supervoxel j for a novel (unseen) target volume \mathbf{I}_u with index u , i.e. we want to find a mapping that assigns one of T tissue classes:

$$l'_{u,j} : s_{u,j} \mapsto \mathcal{T} = \{1, \dots, T\} \quad (3.30)$$

3.2.2 Feature Extraction and Clustering

Since we assume that a volume contains multiple rather than only one type of tissue classes (healthy, ground glass,...) in different expansions, we quantize each image in supervoxels and classify supervoxels individually. For this purpose the MonoSLIC algorithm [38] is applied to each image, that computes an oversegmentation of \mathbf{I}_i into a set of S_i disjunct supervoxels (see Section 3.1.2 for a detailed explanation). A supervoxel j in image i is identified by $s_{i,j}$.

To describe the visual content of supervoxels we extract texture features as introduced in Section 2.1 for each supervoxel. Both descriptors introduced (BVW-LBP, Haralick) are evaluated in this work. To facilitate reading, we describe the remaining part of our method independent from the type of extracted feature and denote the D dimensional feature vector of a supervoxel j in image i as $\mathbf{f}_{i,j}$ and the entirety of all features as

$$\mathcal{F} = \{\mathbf{f}_{1,1} \dots \mathbf{f}_{1,S_1}, \dots, \mathbf{f}_{N,1} \dots, \mathbf{f}_{N,S_N}\}. \quad (3.31)$$

In the next step we partition a randomly selected subset $\mathcal{F}' \subset \mathcal{F}$ of this feature space using clustering algorithms as described in Section 2.2. Two clustering approaches are used and evaluated in this work (k-means, GMM). Since partitioning clustering algorithms aim at finding groups of similar objects, we expect the computed clusters to represent prototypes of different tissue classes (healthy and pathological) that occur in the training data. Independent from the clustering approach in use, we denote the set of K computed clusters as $\mathcal{C} = \{C_1, \dots, C_K\}$, with index $k = 1 \dots K$.

Furthermore we define the function that assigns an observed supervoxel feature vector $\mathbf{f}_{i,j}$ to a specific cluster as $m(\mathbf{f}_{i,j}) = \{1..K\}$. Depending on the clustering approach, m assigns a feature vector to a cluster based on minimal Euclidean distance to the mean vectors clusters (k-means) or on the maximal posterior probability of a cluster given a feature vector (GMM). A detailed description of both methods is given in Section 2.2 of this work.

3.2.3 Mapping Terms to Clusters

For each image, we have an aligned set of terms \mathcal{T}_i that describes occurring tissue types in the whole image. This means that there is at least one existing supervoxel in the image for each of the assigned labels. We thus generate labels $\mathcal{T}_{i,j}$ for each supervoxel in the training set as

$$\mathcal{T}_{i,j} = \mathcal{T}_i \quad \forall j \in S_i. \quad (3.32)$$

We refer to these labels as *weak* labels, since each supervoxel is assumed to belong to one and only one tissue class, but can be assigned (depending on the volume labels) with multiple labels. Correspondingly, from the set of all supervoxels labeled with term t , only a subset carries the true label t .

We use the weakly labeled supervoxels $\mathcal{T}_{i,j}$ and the cluster assignment of all supervoxel feature vectors $m(\mathbf{f}_{i,j})$ to build a histogram H that reflects occurrences of labels in each cluster. An entry in the $K \times T$ dimensional occurrence histogram H for a cluster k and term t is thus computed as

$$H_{k,t} = \#\{\mathbf{f}_{i,j} \mid m(\mathbf{f}_{i,j}) = k\}, \quad \forall (i,j) \mid t \in \mathcal{T}_{i,j}. \quad (3.33)$$

Since we do not expect the labels to be uniformly distributed across the training data, we normalize each entry of H by its term frequency, to reduce an overrating of the most dominant labels in the data and define the normalized occurrence histogram H' as

$$H'_{k,t} = \frac{H_{k,t}}{\sum_{k=1}^K (H_{k,t})} \quad (3.34)$$

From H' we compute a $K \times T$ dimensional probability table L that holds the conditional probability of a term t when observing a specific cluster C_k as

$$L_{k,t} = p(t | C_k) = \frac{H'_{k,t}}{\sum_{t=1}^T (H'_{k,t})}, \quad (3.35)$$

where k identifies a cluster and t a specific term.

Figure 3.8 illustrates the processing pipeline of the approach proposed to learn the probability table L that predicts tissue classes for observations of clusters.

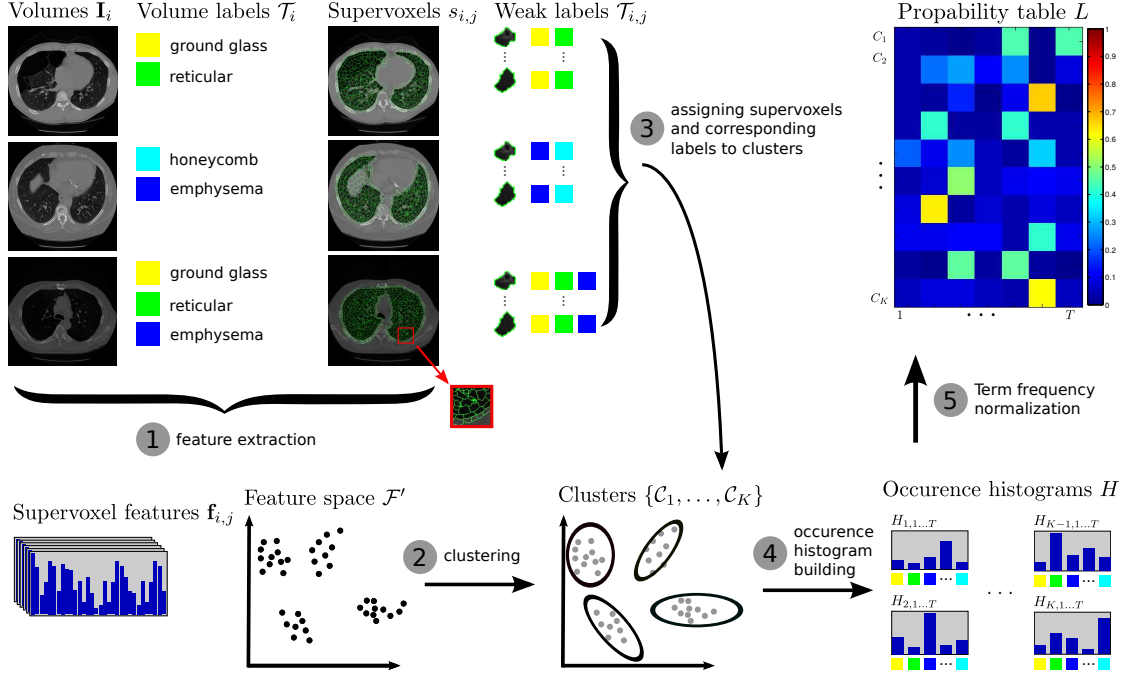


Figure 3.8: Overview of learning a probability table for predicting tissue classes. Supervoxel features are extracted (1), followed by clustering a randomly sampled subset of this feature space (2). Supervoxels and assigned weak labels are mapped to the computed clusters (3). Based on this information, occurrence histograms are calculated (4). After normalizing occurrence histograms by term frequencies of tissue class labels (5), a probability table that predicts tissue classes given a specific cluster is computed.

3.2.4 Classifying Novel Image Regions

The probability table L computed in the previous step is finally used to classify supervoxels $s_{u,j}$ of an unseen target volume \mathbf{I}_u , which is achieved as illustrated in Figure 3.9 with the following computation steps:

1. Compute supervoxels $s_{u,j}$ of \mathbf{I}_u

2. Extract supervoxel texture features $\mathbf{f}_{u,j}$
3. Assign supervoxels to cluster \mathcal{C} using mapping function $m(\mathbf{f}_{u,j})$ to establish clusters with highest probabilities given a feature vector.
4. Generate the true tissue class label estimate $l'_{u,j}$ for each supervoxel j of image with index u by assigning the label with highest conditional probability given the supervoxels cluster assignment $m(\mathbf{f}_{u,j})$.

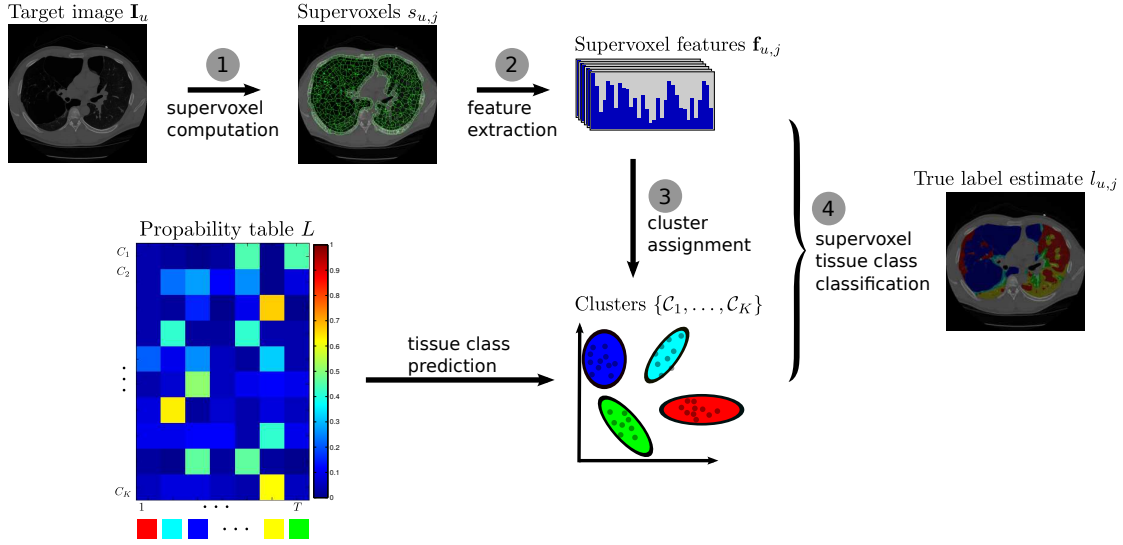


Figure 3.9: Overview of the classification of supervoxels in a novel volume. This process involves the computation of supervoxels (1), feature extraction (2), assignment of supervoxel to clusters (3) and finally labeling each supervoxel with the term of highest prediction probability given its cluster assignment (4).

3.3 Summary

Within this chapter approaches for two components (segmentation and classification) of a CAD system have been described. Both methods do not rely on manually annotated training data. Instead they are designed to learn from data that is available from clinical routine.

Unsupervised Medical Image Segmentation on Supervoxel Level The approach proposed to segment anatomical structures in medical image data on supervoxel level consists of two phases:

1. The training phase, which contains the learning of supervoxel feature prototypes to compute an initial segmentation of all images and furthermore the labeling of the atlas space.

Here each supervoxel in the reference space is labeled using a MRF that incorporates the majority vote of initially estimated segmentations of the image population as well as penalizes label changes of spatially neighboring supervoxels.

2. A novel volume is then segmented in the application phase. After registration of the volume to the atlas and propagation of supervoxels to the volume, texture features are extracted and used to find an initial segmentation. Finally a MRF that incorporates the initial segmentation, the learned atlas segmentation and spatial constraints is created to find a segmentation labeling of the novel volume.

Weakly Supervised Classification of Pathologies To classify healthy and pathological tissues in medical images, a weakly supervised learning method has been proposed. The approach takes a set of medical images as input, where each image is aligned with a set of weak labels, that describe occurring pathologies of an image. From this labels, we know which pathologies occur in an image, but we do not know which regions of an image are affected by which pathology and which regions contain healthy tissue.

The approach learns prototypes of occurring tissue types by clustering super voxel features of the training images. By assigning volume labels to all supervoxels of a volume and mapping these weak labels to the computed clusters we learn the distribution of tissue labels in each cluster. We assume that each cluster represents on specific tissue class and use the label distribution in the clusters to compute a probability table that predicts a tissue class of each cluster.

Supervoxels of a novel volume are then classified by identifying the cluster of each supervoxel and assigning the tissue label with highest probability.

Experiments and Results

In the previous chapter we have introduced two approaches that learn from data that is typically available during clinical routine. (1) A latent atlas based unsupervised segmentation approach that aims in finding anatomical structures in medical images and (2) a weakly supervised learning approach that classifies different tissue types within an organ based on textual labels that are obtained from radiological reports corresponding to an image.

The present chapter provides an evaluation of both approaches on publicly available data sets that carry voxel wise ground truth annotations for evaluation purposes. This is the VISCERAL data set [32], that provides medical images with manual annotations of anatomical structures and the LTRC data set [37] which contains chest CT scans of patients affected by Interstitial Lung Diseases (ILD) [55] with voxel wise annotations for healthy and pathological tissues within the lungs.

Section 4.1 provides the evaluation of the method proposed for unsupervised anatomical structure segmentation, where Section 4.2 contains the evaluation of the approach proposed for weakly supervised image region classification. Section 4.3 summarizes the experiments performed and their results.

4.1 Unsupervised Medical Image Segmentation on Supervoxel Level

This section provides an evaluation of the approach proposed for unsupervised anatomical structure segmentation in medical images. We start with describing data set that the framework (1) requires for learning and (2) is used for evaluation in Section 4.1.1. Section 4.1.2 describes and discusses the evaluation metric used to measure segmentation performance, where Section 4.1.3 describes the data acquisition, parameter settings and features in use during learning. In 4.1.4 we provide a detailed evaluation of all components of the framework that influence the resulting latent atlas labeling, where Section 4.1.5 covers the evaluation of all components that influence the resulting labeling in novel images that are not part of the training.

4.1.1 Data in Use

Three datasets are required to perform experiments with the segmentation framework proposed and evaluate their results:

1. **Training data:** A set of medical images from which the algorithm learns occurring supervoxel prototypes and a segmentation in the atlas space.
2. **Reference space:** One medical image that serves as reference space or atlas. The atlas is used to provide a common coordinate system across the image population.
3. **Test data:** A set of, from the training data disjunct, medical images that have voxel wise expert annotations of anatomical structures. These volumes are used as ground truth data.

The following paragraphs describe each of the datasets in detail. Finally, Table 4.1 gives a summary and provides representative illustrations of all datasets in use.

Training data A set of 450 CT volumes originated at the radiology department of the General Hospital of Vienna (AKH) is used as training data for the unsupervised image segmentation part of this work. There are three types of body regions that are covered by a volume: chest scans, abdominal scans and chest + abdominal scans. The in-slice resolution lies between $0.5 \text{ mm} \times 0.5 \text{ mm}$ and $1.2 \text{ mm} \times 1.2 \text{ mm}$, where the resolution between slices varies from 0.7 mm to 2 mm . Within the image normalization pipeline all volumes are oriented to the same direction and transformed to isotropic volumes so that a pixel resolution of $0.7 \text{ mm} \times 0.7 \text{ mm} \times 0.7 \text{ mm}$ in x-, y- and z- direction is assured. The dataset contains both, contrast enhanced CT (CTce) scans and CT scans without contrast enhancement.

Reference space As reference space, a whole body CT volume with pixel resolutions of $1.3 \text{ mm} \times 1.3 \text{ mm} \times 1.3 \text{ mm}$ is used. 9 anatomical structures, including the brain, parts of the spine, left and right lung, trachea, heart, liver as well as left and right kidneys have been annotated by a medical expert. These annotations are used to evaluate the segmentation performance in the atlas. For the center estimation within the registration process, volume center positions in the atlas volume in a set of 200, from the training dataset distinct CT volumes have been annotated.

Test data For the evaluation of anatomical region segmentation we use a subset of the *VISCERALanatomy*¹ benchmark training data [32]. VISCERAL provides medical image data in four modalities (CT, CTce, MRT1 and MRT1 contrast enhanced fat saturated) with voxel wise manual expert annotations of 20 anatomical structures including the trachea, lung, pancreas, gallbladder, urinary bladder, sternum, kidneys, aorta, thyroid gland, liver, adrenal glands and the first lumbar vertebra. The dataset in use to evaluate the segmentation performance of the method proposed contains 7 CT and 7 CTce volumes, all of them covering chest and abdomen.

¹<http://www.visceral.eu/>

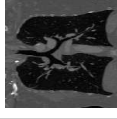
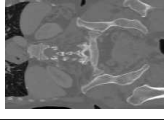
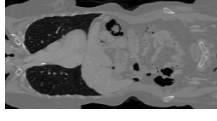
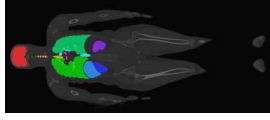
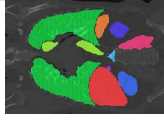

Region & modality	# Volumes	# Annotations	Origin	Illustration
Training data				
Chest CT	192	-	AKH	
Abdominal CT	150	-	AKH	
Chest & abdomen CT	108	-	AKH	
Reference space				
Whole body CT	1	9 structures		
Test data				
Chest & abdomen CTce	7	20 structures per volume	VISCERAL	
Chest & abdomen CT	7	20 structures per volume	VISCERAL	

Table 4.1: Overview of data used for the unsupervised image segmentation part of this work.

4.1.2 Evaluation Metric

The performance of the segmentation framework is measured using the Dice coefficient [17], calculated between computed segmentations and manual annotations of anatomical structures in the atlas space and in volumes of the test dataset. The Dice coefficient is a metric that expresses the spatial overlap of two binary label images.

Given a binary labeled ground truth segmentation \mathbf{S}^{GT} and a computed binary segmentation \mathbf{I}^C , the Dice coefficient is defined as

$$DICE = \frac{2 * |\mathbf{S}^{GT} \cap \mathbf{S}^C|}{|\mathbf{S}^{GT}| + |\mathbf{S}^C|} \quad (4.1)$$

The resulting coefficient lies in the range of $(0, 1)$ where 0 indicates no overlap and 1 depicts full overlap of the two involved binary label images [89]. The Dice coefficient incorporates both, the proportion of correctly positive segmented (*Sensitivity*) and correctly negative segmented (*Specificity*) voxels. Sensitivity and Specificity are calculated by

$$Sensitivity = \frac{TP}{TP + FN} \quad Specificity = \frac{TN}{TN + FP}, \quad (4.2)$$

where TP indicates true positive voxel segmentations, TN true negatives, FP false positives and FN false negatives respectively.

Figure 4.1 illustrates the relation of these three segmentation similarity measures, on a synthetic dataset. The Dice coefficient (in red) is sensitive to false negative segmented pixels (synthetic image indices 1 - 4) as well as to false positive segmented pixels (synthetic images indices 1 - 4), where sensitivity and specificity are only influenced by one of those. The Dice coefficient is thus used in the remaining part of this work to measure segmentation performance of computed segmentations.

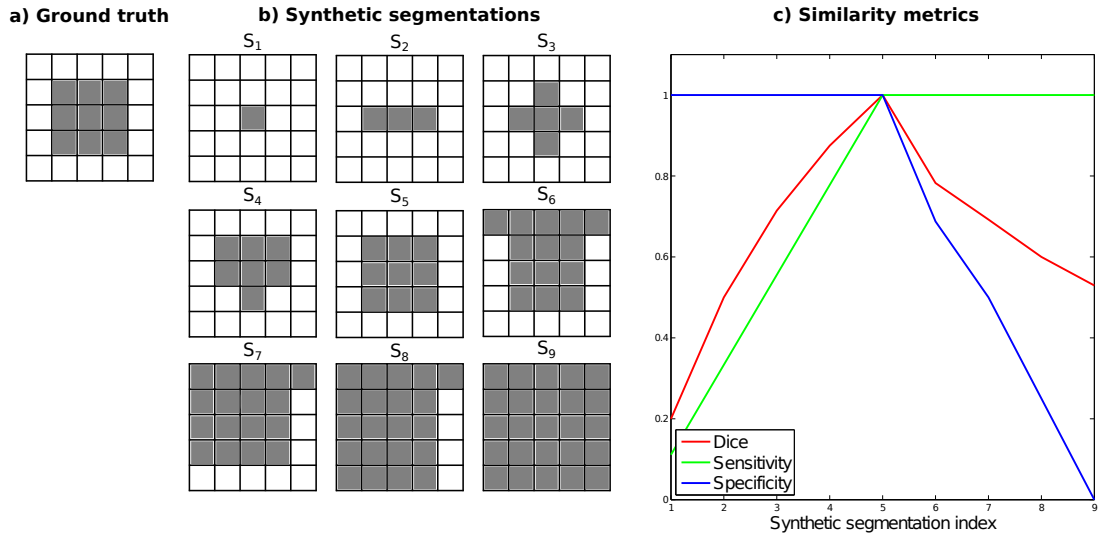


Figure 4.1: Relation of Dice coefficient, Sensitivity and Specificity. Ground truth segmentation is shown in *a*. Plot *b* illustrates synthetic segmentations $S_1 \dots S_9$. Plot *c* shows sensitivity, specificity and Dice coefficient values, computed between ground truth image and synthetic segmentations.

4.1.3 Experimental Setup

All training images have been recorded during clinical routine at the Radiology Department of the General Hospital of Vienna, where the images are stored in a Picture Archiving and

Communication System (PACS). The DICOM [7] format is used to export images from the PACS. Here each slice of a volume is stored separately. Since we aim at processing 3-d volumes rather than a stack 2-d slides of a patient we convert those files to the Neuroimaging Informatics Technology Initiative (NIfTI)² file format, which stores only one file per volume. Since it is not ensured that each volume exported from the PACS is stored in the same orientation we reorient each volume using the FSL³ toolbox [68] to ALS orientation. This means that every volume is oriented and stored in a matrix so that the x-coordinate increases from Anterior to posterior, y-coordinate from Left to right and the z-coordinate from Superior to inferior. Figure 4.2 visualizes the orientation convention used in this work, utilizing the matVTK⁴ toolbox.

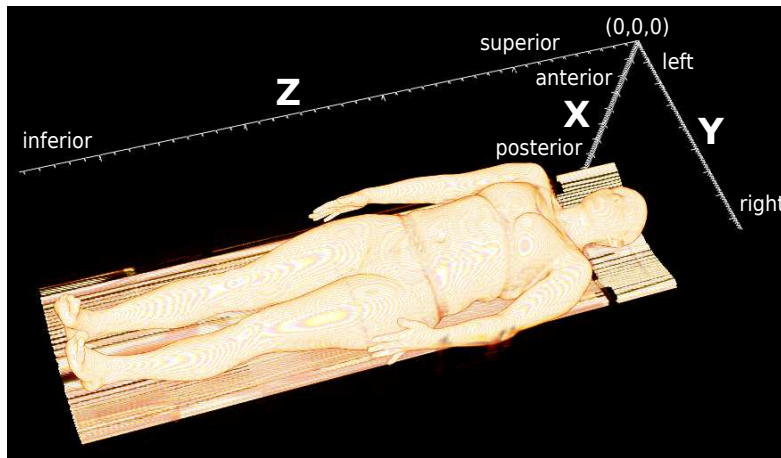


Figure 4.2: ALS volume orientation as used in this work. X-axis increases from Anterior to posterior, y-axis from Left to right and Z-axis from superior to inferior point of view.

The segmentation framework requires the registration of all images to the atlas. For this purpose the registration framework described in Section 2.3.3, developed by Dorfer in [22] has been used. Inspired by his work, the atlas and all volumes are down sampled to an isotropic pixel resolution of 2 mm in x-, y-, and z-direction before performing the registration of a volume to the atlas, which reduces the amount of voxels in the involved volumes, causing less computation time and memory costs [63]. Again inspired by Dorfers work in [22] a B-spline grid with 10 mm spacing and NMI as similarity measure are used for non rigid registration.

The atlas is quantized into supervoxels using the MonoSLIC algorithm proposed by Holzer et al. in [38]. Results in this work show that the average size of a supervoxel influences the recall rate of anatomical structure boundaries in medical images. According to their results, we choose an average superpixel size of 1 cm³.

After the transformation of the oversegmentation in the atlas to each volume of the training set, supervoxel texture descriptors are extracted. Inspired by Burner et. al. [10] we use the BVW-LBP operator. They suggest a weighting factor for the average intensity and contrast measure

²<http://nifti.nimh.nih.gov/>

³<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>

⁴<http://www.cir.meduniwien.ac.at/team/birngruber/matvtk/>

of $w_c = w_i = 10$ to retrieve pathological supervoxels of lungs CT scans. Since we want to distinguish anatomical regions rather than pathologies in a certain anatomy we assume that the impact of average intensity and contrast in a supervoxel is of more relevance and set $w_c = w_i = 20$. Again inspired by their work we compute $k = 300$ visual words on $s = (1, 2, 3, 4)$ different scales, by random sampling of 150000 features and performing k -means clustering on each scale.

4.1.4 Accuracy of Latent Atlas Space Labeling

The present section describes experiments that have been applied to obtain segmentations of anatomical structures in the atlas. The segmentation labeling for each node k in the atlas that represents a supervoxel is obtained by solving a MRF that is constructed as described in Section 3.1.4. Here, we review the components and parameters of the segmentation framework that influence the cost function of the MRF and thus the resulting labeling of the atlas:

- L , the number of clusters used to partition the feature space. Each cluster is expected to represent an anatomical structure.
- Unary terms \mathbf{U} , which hold the probabilities that node k belongs to one of the L computed clusters. Unary terms for a specific node k are calculated based on a majority vote of cluster assignment probabilities of all supervoxels in the training volumes that spatially correspond to k .
- Binary terms \mathbf{B} with impact weighting parameter α , that penalize label changes in neighboring supervoxels.
- T , the number of neighbors to which an edge of each node k is created.

Results show that the approach is able to detect three anatomical structures in the reference space that correspond to organs in the human body. The best performing parameter setting results in segmentations of the lungs with a Dice score of 0.94, the heart with 0.8 and the liver with 0.45. In the following sections we provide a detailed stepwise evaluation of all components of the framework for labeling the reference space.

Impact of unary terms on segmentation accuracy The aim of the first experiment is to reveal how many clusters are formed during the feature space partition that correspond to organs of the human body. Hence we set the impact of binary terms to $\alpha = 0$, which causes the resulting cost function of the MRF being only dependent on unary terms \mathbf{U} . This means that the labeling of one node k in the atlas depends only on the cluster assignment probabilities of all spatially to k corresponding supervoxels in the training data.

Figure 4.3 illustrates four slices of the atlas volume and two corresponding segmentations computed with $\alpha = 0$ and $L = 12$ and $L = 18$ cluster centers. The colors in the plots illustrate the assignment of a supervoxel to one of the L clusters. Results show that the clustering of both configurations causes segmentations of three organs (lungs, liver heart) as well as anatomical

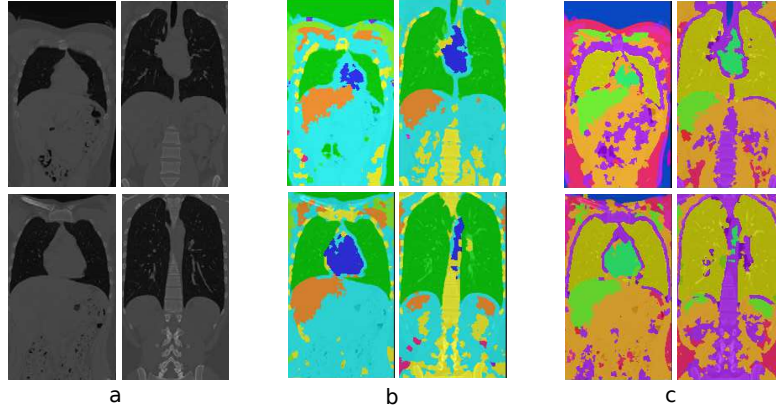


Figure 4.3: Four coronal slices of the cropped atlas volume (a). Supervoxel labeling of atlas, obtained with $\alpha = 0$ while clustering the feature space into $L = 12$ (b) and $L = 18$ (c) clusters.

structures with similar texture properties (regions that contain bones such as the spine, and the ribs on the borders of the lungs).

In Figure 4.4, unary terms \mathbf{U} are illustrated in form of heat maps for a subset of clusters in coronal slices of the atlas, with $L = 18$ computed clusters. The first column shows the color coded labeling of the atlas in two slices. The remaining plots show slices of the atlas with an overlaying color coding of \mathbf{U} for a specific cluster or labeling class l . The frame color of visualizations in the first row corresponds to the clusters color coding in the atlas segmentation. In other words, the plots show probabilities of segmentation classes to occur in specific regions of the atlas.

Figure 4.5 shows a bar plot of the performances of computed segmentations that correspond to the lungs, liver and heart. The y-axis depicts the Dice coefficient from the segmentations where the x-axis depicts the number of clusters $L = (4, 6, 8, 10, 12, 14, 16, 18, 20, 25, 30)$ used to partition the feature space. The Dice coefficients of liver segmentations are denoted in blue, the lungs in green and the heart in red.

Best results for the three detected organs (lungs, heart, liver) have been obtained using $L = 18$ clusters, leading to Dice coefficients of 0.92, 0.72, and 0.54 respectively. This setting is thus used for all further experiments.

Impact of binary terms on segmentation accuracy Within this experiment the incorporation of binary terms to the cost function of the MRF is evaluated. Binary terms are used to model the assumption that two neighboring nodes in the atlas are likely to belong to the same anatomical structure. Hence we encourage similar label assignments in neighboring nodes of the MRF using weighting factor α , i.e. the higher α the more we encourage similar label assignments.

Figure 4.6 illustrates four slices of the atlas volume in *a*, together with computed segmentations for $\alpha = (0, 0.05, 0.4)$. Here, all segmentations have been computed using $L = 18$ clusters, since the three addressed organs have shown most promising results with this configuration.

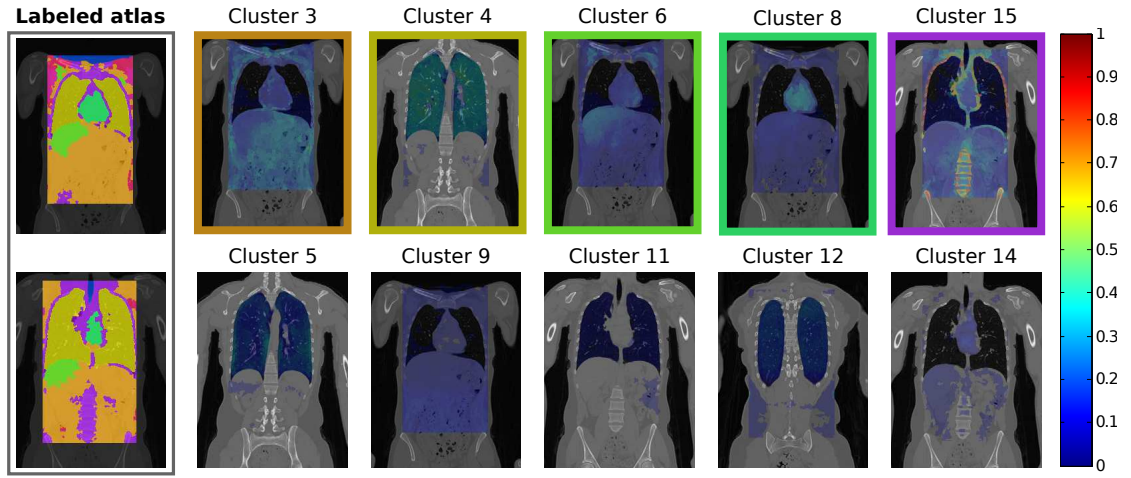


Figure 4.4: Computed atlas segmentation (left) and probability heat maps of unary terms for different clusters in the remaining plots. The frame color of the first rows images depicts the color coding of the clusters resulting segmentation in the atlas.

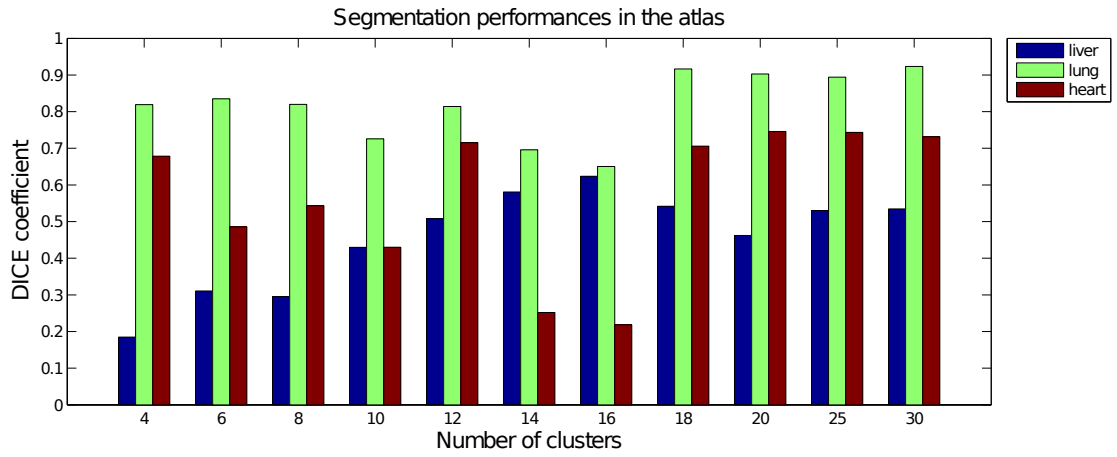


Figure 4.5: Dice coefficients of segmentation labels in the atlas that correspond to the liver, heart and the lungs for different numbers of clusters $L = (4, 6, 8, 10, 12, 14, 16, 18, 20, 25, 30)$.

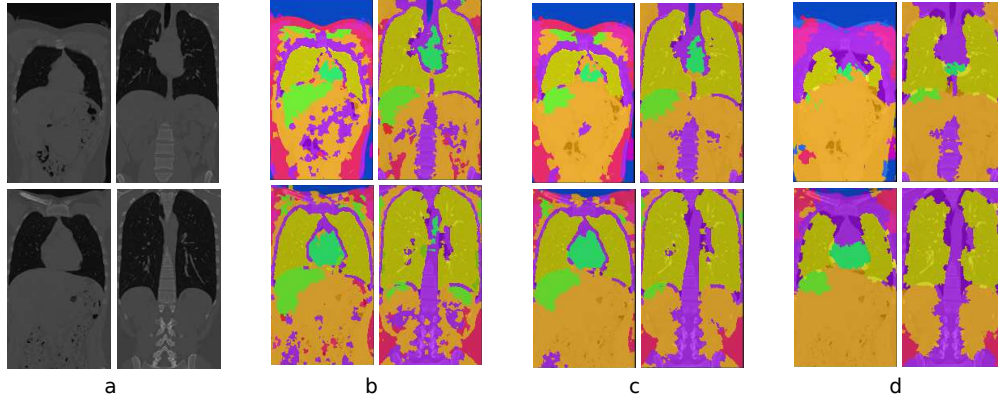


Figure 4.6: Four coronal slices of the cropped atlas volume (a). Supervoxel labeling of atlas, obtained with increasing number of $\alpha = 0$ (b), $\alpha = 0.05$ (c), and $\alpha = 0.4$ (d), while partitioning the feature space into $L = 18$ clusters.

Figure 4.7 shows a bar plot holding the Dice coefficients of computed lung (green), liver (blue) and heart (red) segmentations on the y-axis, while different weighting factors α are denoted on the x-axis. Experiments are performed while connecting each node to its $T = 6$ spatially nearest neighbors.

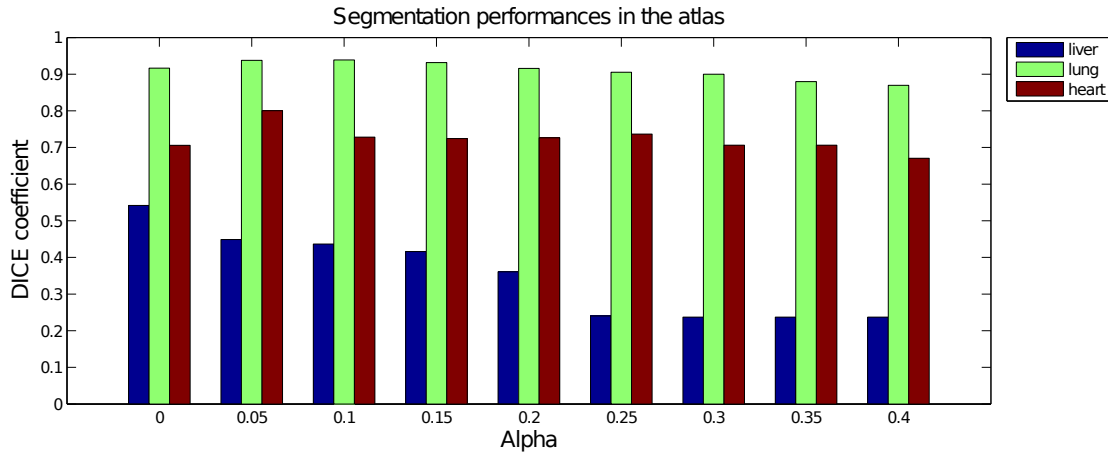


Figure 4.7: Dice coefficients of segmentation labels in the atlas that correspond to the liver, heart and the lungs for different weighting factors α of binary terms while segmenting the atlas using a MRF.

Table 4.2 lists segmentation performances of the addressed organs while varying the neighborhood size $T = (6, 9, 12)$ of a node and setting $\alpha = 0.05$ during MRF construction.

Results show that segmentation performances do not improve with increasing size of neigh-

L	α	T	liver	lung	heart	average
18	0.05	6	0.45	0.94	0.80	0.59
18	0.05	9	0.48	0.95	0.75	0.57
18	0.05	12	0.37	0.95	0.78	0.58

Table 4.2: Lung, liver and heart segmentation Dice coefficients obtained with $\alpha = 0.05$, $L = 18$ and different sizes of neighborhood T .

boring supervoxels. Hence we set $T = 6$ for all further experiments.

Results furthermore show that incorporating binary terms into the cost function of the MRF increases the segmentation performance of lungs and the heart, whereas the segmentation performance of the liver decreases with an impact factor of $\alpha = 0.05$. The dice coefficient of the lung segmentation increases from 0.92 to 0.94, the heart from 0.72 to 0.80, where the liver segmentation performance is reduced from 0.54 to 0.45. All other tested values ($\alpha > 0.05$) do not improve segmentation performances compared to results when setting $\alpha = 0.05$.

4.1.5 Accuracy of Labeling Individual Volumes

This section describes experiments that have been performed to evaluate the accuracy of the proposed framework when segmenting anatomical structures in novel medical images. As described in Section 3.1.5, a novel image is segmented by solving a MRF that assigns a segmentation label to each supervoxel of an image.

Similar to the previous section, we review the components of the MRF that influence its cost function and thus the resulting segmentation.

- $\mathbf{U}^{\mathcal{L}^*}$, the first component of unary terms, which holds the labeling vote of the atlas in each supervoxel.
- $\mathbf{U}^{\mathcal{C}}$, the second component of unary terms, which hold the assignment probabilities of a supervoxels feature vector to belong to each of the L clusters \mathcal{C}_l .
- β , the mixing coefficient of unary terms. The lower β , the higher is the impact of the atlas and the lower the impact of the cluster center assignments on the resulting segmentation.
- γ , the weighting factor of binary terms that model the assumption that spatially neighboring supervoxels are likely to belong to the same class. γ has similar functionality as α for labeling the atlas space. As described in Section 3.1.5 the encouragement of similar label assignments is dependent on the intensity difference. I.e. two connected supervoxels with low intensity differences are encouraged to have a similar label assignment, whereas neighboring supervoxels with high intensity differences are not encouraged to have similar label assignments.

Similar to the previous section we provide a detailed evaluation of all components within the following paragraphs. We investigate the impact of unary terms on the resulting segmentations in a first step, whereas the impact of binary terms is evaluated in a second step.

Impact of unary terms on segmentation accuracy Within this experiment, the impact of unary terms to the resulting segmentations is investigated. The MRF parameters to obtain the latent atlas labeling, which serves as prior segmentation estimate, are chosen according to the most promising results of the previous section, i.e. we set $L = 18$ clusters and $\alpha = 0.05$. To show the impact of unary terms only we set $\gamma = 0$, causing the cost function of the MRF being only dependent on unary terms.

As described in the previous section, the framework is capable to detect the lungs, liver and the heart in medical images. Since the VISCERAL dataset does not provide manual annotations for the heart, we focus on the lungs and the liver in the remaining parts of the evaluation section.

The impact of unary term components $U^{\mathcal{L}^*}$ and $U^{\mathcal{C}}$ is controlled by the mixing coefficient β . Segmentations obtained with $\beta = 1$ are referred to as *initial segmentations*, since the cost function of the MRF in this setting is derived only by cluster assignment probabilities.

Figure 4.10 shows ground truth annotations (a), initial segmentations (b) and segmentations computed by setting $\beta = 0.6$ (c) of one CT (first row) and CTce volume (second row) as well as the a priori computed atlas segmentation (d). The visualizations in b show that the initial segmentation labeling partially corresponds to the addressed organs. The lung is mainly labeled in blue in the CT volume and in yellow and green in the CTce volume. The label that is dominant in the liver in both modalities occurs next to the liver also in the heart and in other abdominal regions. Plot c illustrates the effect of incorporating the atlas labeling to the unary terms. Here, the addressed organs have to the atlas corresponding, consistent labels in the volumes of both modalities.

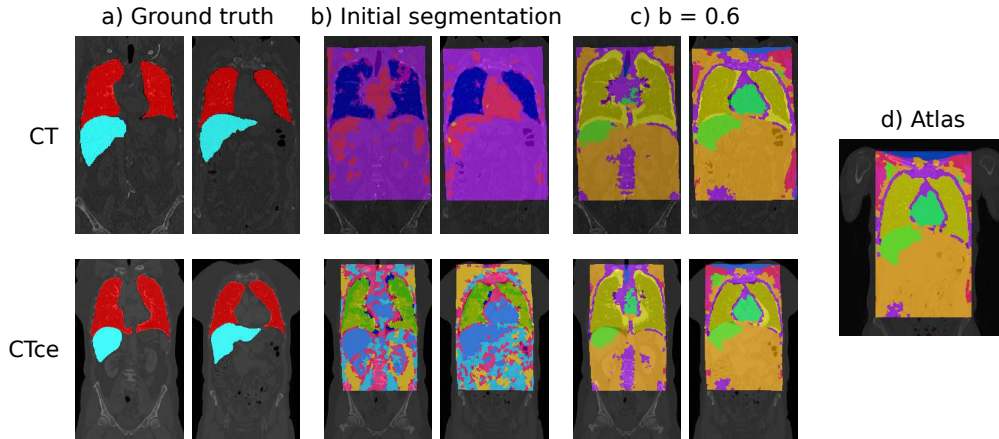


Figure 4.8: Ground truth annotations of the two addressed organs, initial segmentations and segmentations computed with $\beta = 0.6$.

Figure 4.9 illustrates average Dice coefficients of the two addressed organs in both modalities

that are part of the test set (CT & CTce). The x-axis depicts the increasing impact of the atlas segmentation. Starting with $\beta = 1$, where the computed segmentation is based on the cluster assignment probabilities only, β is decreased until the segmentation is influenced only by the atlas segmentation, $\beta = 0$.

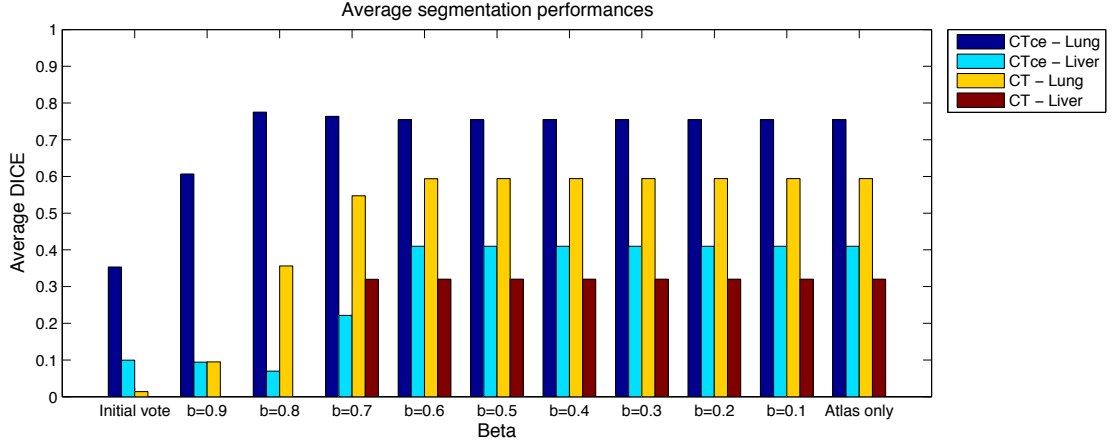


Figure 4.9: Dice coefficients of region labels that correspond to the liver and the lungs in both modalities that occur in the test set (CT, CTce). X-axis depicts decreasing β values, which increases the impact of the atlas.

Results show that the proposed framework is able to find consistent labels of the addressed organs in all tested volumes. Combination of the atlas segmentation with the initial segmentations leads to average Dice coefficients of 0.75 and 0.51 of lungs and 0.41 and 0.32 of the liver in CTce and CT volumes respectively using the mixing coefficient $\beta = 0.6$.

Impact of binary terms on segmentation accuracy Within the final experiment, the impact of binary terms when labeling individual volumes is evaluated. In comparison to binary terms of MRF for labeling the atlas, an additional weighting is applied by considering the average intensity difference of neighboring supervoxels. Here, similar label assignments are only encouraged if the intensity difference of the involved supervoxels is low with respect to all occurring intensity differences of neighboring supervoxels in a volume.

According to the best performing parameter settings in the previous sections we set $L = 18$, $\alpha = 0.05$, $\beta = 0.6$ and vary the binary impact factor γ in the following experiments. The number of neighbors to which a node is connected is set to $T = 6$ since results while labeling the atlas have shown that a higher number of considered neighbors does not increase segmentation performance.

Figure 4.10 illustrates ground truth annotations (a) and computed segmentations in a CT and a CTce volume while increasing the impact of binary terms (b,c,d,e). The corresponding average Dice coefficients are shown in Figure 4.11. Here the y-axis depicts average Dice coefficients, the x-axis depicts the increasing impact of binary terms. Grouped bars indicate the modality and organ that is evaluated.

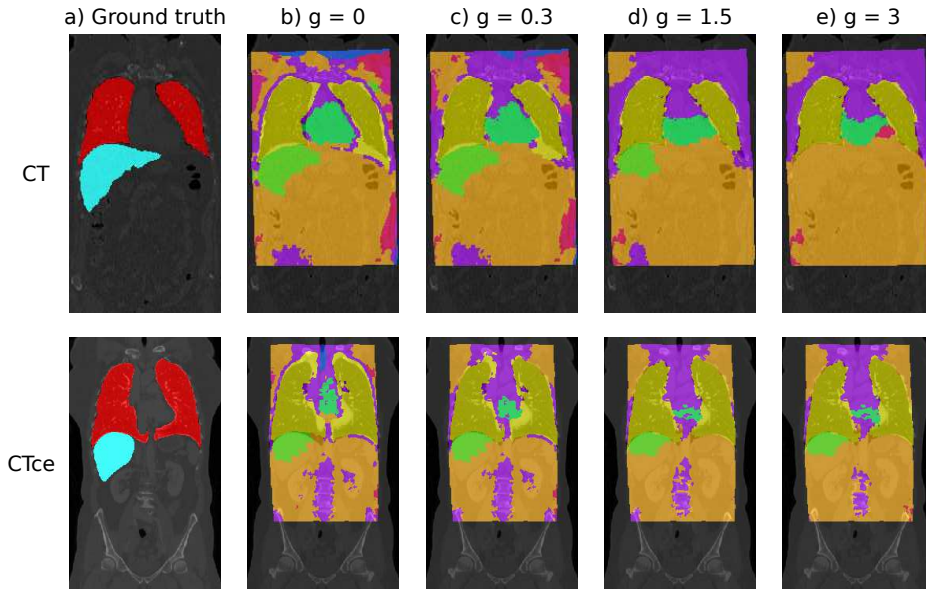


Figure 4.10: Ground truth annotations of lungs and liver in one CT and one CTce volume (a), Computed segmentations of the corresponding volumes while increasing γ , the impact of binary terms (b-e).

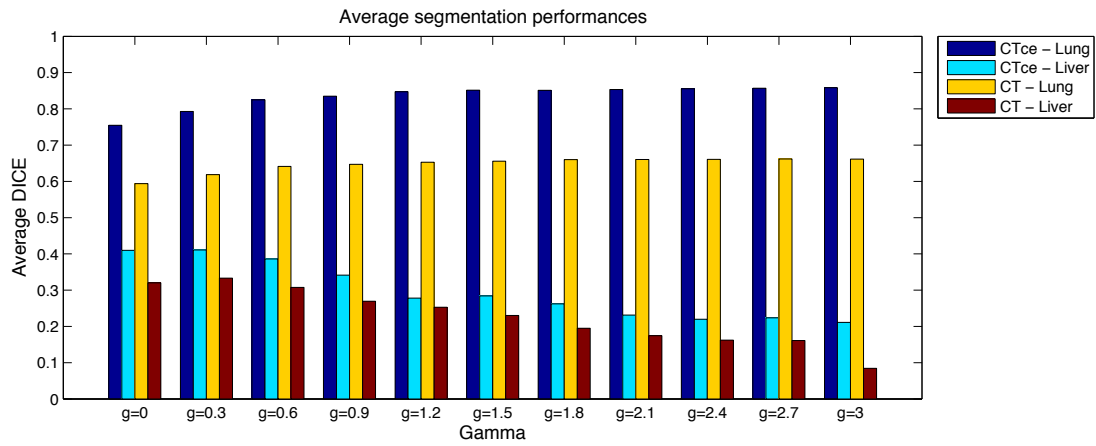


Figure 4.11: Dice coefficients of region labels that correspond to the liver and the lungs while increasing the impact of the atlas segmentation.

By incorporation of binary terms, segmentation performances increase within the lungs from 0.75 and 0.59 to 0.86 and 0.66 (CT and CTce) but decreases within the liver from 0.41 and 0.32 to 0.21 and 0.08 respectively when setting $\gamma = 3$.

4.2 Weakly Supervised Classification of Pathologies

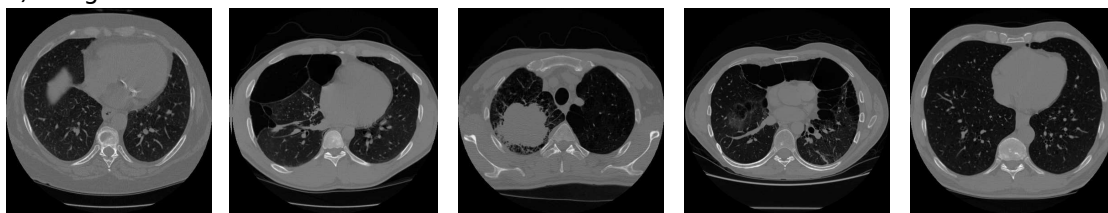
This section describes experiments that have been performed to evaluate the method proposed for classifying medical image regions by learning from weak image labels. We apply our method with four combinations of texture descriptors and clustering methods, show their strengths and weaknesses and discuss main characteristics and limitations of the method proposed.

Section 4.2.1 describes the LTRC [37] data set and how it is prepared to fit our problem statement. Section 4.2.2 describes parameter settings, features and clustering methods in use, while Section 4.2.3 introduces the metric used to measure classification performance. Finally, Section 4.2.4 shows results of the experiments performed.

4.2.1 Data in Use

All experiments are performed using the LTRC [37] dataset, which consists 300 lung CT scans, all affected by ILD. Each volume carries a binary lung mask and a voxel-wise labeling of five tissue classes within the lungs: healthy, emphysema, ground glass, honeycombing and reticular. Figure 4.12 shows axial slices of five volumes and corresponding annotations of the dataset.

a) Images



b) Voxel wise annotations

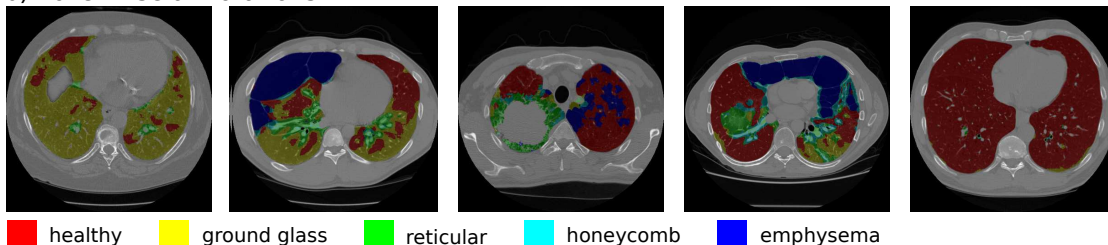


Figure 4.12: Illustration of the LTRC [37] dataset used in this work. (a) Axial slices of five chest CT scans affected by ILD. (b) Corresponding voxel-wise annotation of healthy and pathological tissues.

The volumes are recorded with pixel resolutions between $0.6 - 0.7 \text{ mm} \times 0.6 - 0.7 \text{ mm} \times 0.6 - 0.7 \text{ mm}$ and resampled so that an isotropic resolution of 0.7 mm in each direction is assured. We compute supervoxels with an average size of 1 cm^3 using the MonoSLIC algorithm proposed in [38].

To obtain ground truth labeling on supervoxel level, a supervoxel is assigned with the most dominant label of its voxels. To avoid partial volume effects, supervoxels with less than 75% of

one label are excluded from learning and testing. To simulate a data set that is originated during clinical routines, each supervoxel is assigned with one or two pathological labels that occur in its volume during training, so that each possible combination of pathological terms occurs equally often. The class of healthy supervoxels is a special case. First, we include 25 volumes with no pathological findings (only healthy labels) during training, which is feasible since healthy volumes can be obtained from scans without pathological observations in the report. Second, we take all healthy supervoxels of a volume with a pathological label into account during training, but do not assign the label *healthy* to those supervoxels since this information is not given in practise. Instead those supervoxels are labeled with to the volume assigned pathological terms. We also add 15 volumes assigned with ground glass, reticular and honeycombing and 8 assigned with emphysema only based on the assumption that cases with the finding of a single pathology can be obtained from radiology reports. Table 4.3 lists resulting co-occurrences of volume labels and Table 4.4 the corresponding distribution of supervoxel labels in the training data set.

	Healthy	Ground glass	Reticular	Honeycomb	Emphysema
Healthy	25	0	0	0	0
Ground glass	0	15	38	38	38
Reticular	0	38	15	38	32
Honey combing	0	38	38	15	38
Emphysema	0	38	32	38	8

Table 4.3: Co-occurring volume labels in the training data.

		True labels					Σ	TP
		Healthy	Ground glass	Reticular	Honey-combing	Emphy-sema		
Weak labels	Healthy	174610	0	0	0	0	174610	100
	Ground glass	514600	12665	1668	1281	20134	550348	2.09
	Reticular	358340	4055	5104	1464	4492	373455	1.37
	Honeycombing	468200	2189	2141	4601	25731	502862	0.91
	Emphysema	359730	1751	790	1615	58841	422727	13.92

Table 4.4: Tissue label distribution of supervoxels in the training data. A line depicts the distribution of true labels for all supervoxels of a tissue class. Total amount of supervoxels labeled with a specific tissue class (Σ) and True Positives (TP) in % are given in the last two columns.

4.2.2 Experimental Setup

We extract both texture descriptors described in Section 2.1 for all supervoxels. (1) Texture bags of Local Binary Patterns as proposed in [10] on (1,2,3,4) four scales with 300 visual words on

each scale resulting in a 1200-dimensional feature vector. To overcome the curse of dimensionality we apply PCA and keep 95% of the feature spaces variance, which results in a mapping to a 46-dimensional feature space. Texture bags of Local Binary Patterns are abbreviated with BVW-LBP in the remaining part of this section. And (2), Haralick features [33] of GLCM on $21 \times 21 \times 21$ voxel patches around the center of a supervoxel, binned to 32 grey levels with 1 and 3 pixels offset, resulting in a 26-dimensional feature vector.

We furthermore apply both methods for unsupervised partitional clustering described in Section 2.2 (GMM,k-means) to the sampled feature spaces. Depending on the applied texture descriptor and clustering method in use, classification performances are identified by the combination of their abbreviation (BVW-LBP - k-means, BVW-LBP - GMM, Haralick - k-means, Haralick - GMM).

Experiments are performed by 10-fold cross validation, meaning that we split the data set ten times in 30 test and 270 training volumes so that each volume is exactly once part of the test and nine times part of the training set.

4.2.3 Evaluation Metric

Classification performance is measured by sensitivity and specificity values of classified supervoxels of a tissue class as given in Equation 4.3.

$$Sensitivity = \frac{TP}{TP + FN} \quad Specificity = \frac{TN}{TN + FP} \quad (4.3)$$

Here, TP depict the amount of true positive, TN true negative, FP false positive and FN false negative classified supervoxels of a specific tissue class.

4.2.4 Evaluation of Classification Performance

The evaluation of the method proposed shows the four out of five tissue classes are classified with reasonable accuracy (healthy, ground glass, reticular, emphysema). Best results in all tissue classes are obtained by the combination of Haralick features with k-means clustering. In the following paragraphs we provide a detailed description of performed experiments and obtained results.

Overall sensitivity and specificity Figure 4.13 compares sensitivity and specificity values of tested texture descriptors and clustering approaches, averaged over all tissue classes. Experiments are performed for ten different numbers clusters (15, 20, 25, 50, 75, 100, 150, 200, 300, 400), depicted on the x-axis. Classification baseline is illustrated in dashed grey, obtained by random guessing of tissue labels for each tested supervoxel.

Classification performances on tissue class level Figure 4.14 shows corresponding sensitivity and specificity values separately for all tested tissue classes (healthy, ground glass, reticular, honeycombing, emphysema). Best classification performances are reached using Haralick features and k-means clustering to 400 clusters. Table 4.5 shows the confusion matrix, sensitivity

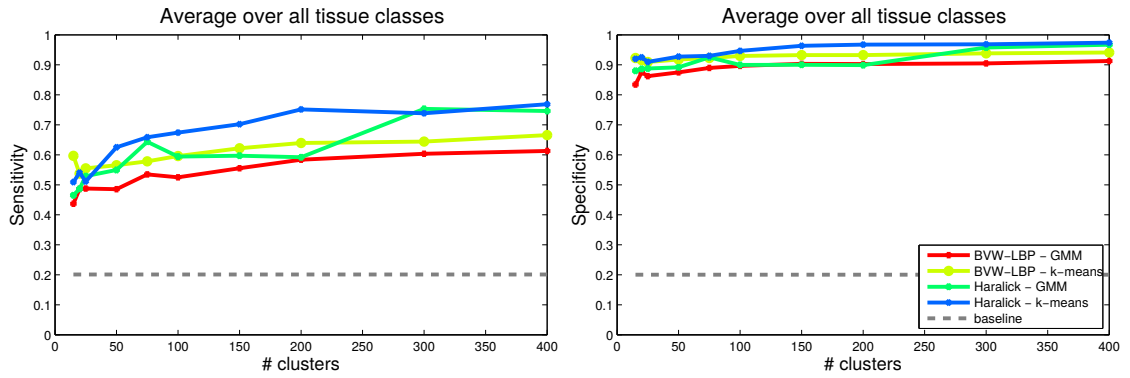


Figure 4.13: Sensitivity (left) and Specificity (right) values averaged over all tissue classes of tested combinations of texture descriptors and cluster approaches for increasing numbers of clusters (15 - 400). Performance baseline (dashed grey) is depicts random guessing.

and specificity values of classified supervoxels for this setting. As the matrix shows, supervoxels affected by honeycombing are more often classified as reticular than as honeycombing. Also ground glass supervoxels are often incorrectly labelled with reticular. Corresponding confusion matrices, sensitivity and specificity values of all other tested combinations of texture features and clustering methods are given in Tables A.1, A.2, A.3 in the appendix of this work.

		Predicted classes				
		Healthy	Ground glass	Reticular	Honeycombing	Emphysema
True classes	Healthy	1020128	40096	19840	11926	4223
	Ground glass	410	27647	6989	687	10
	Reticular	24	264	5071	204	0
	Honeycomb	389	1071	4259	3395	135
	Emphysema	2984	8007	277	5222	138180
Sensitivity		0.90	0.77	0.91	0.37	0.9
Specificity		0.98	0.96	0.98	0.99	0.96

Table 4.5: Confusion matrix, Sensitivity and Specificity values of supervoxel classification using Haralick features and k-means clustering with $k = 400$ clusters.

Qualitative results Figure 4.15 illustrates the classification of supervoxels on axial slices of six tested volumes. The image slice is shown in the first row, where the second row illustrates the ground truth (GT) annotations. The third row illustrates the corresponding supervoxel classification, computed using Haralick features and k-means clustering on 400 clusters. We observe that most regions affected by emphysema (blue), reticular (green) and ground glass (yellow), as well as healthy supervoxels (red) are correctly labelled. Classification errors are observed in honeycombing regions (turquoise label) of volumes in columns 3,5 and 6, where supervoxels

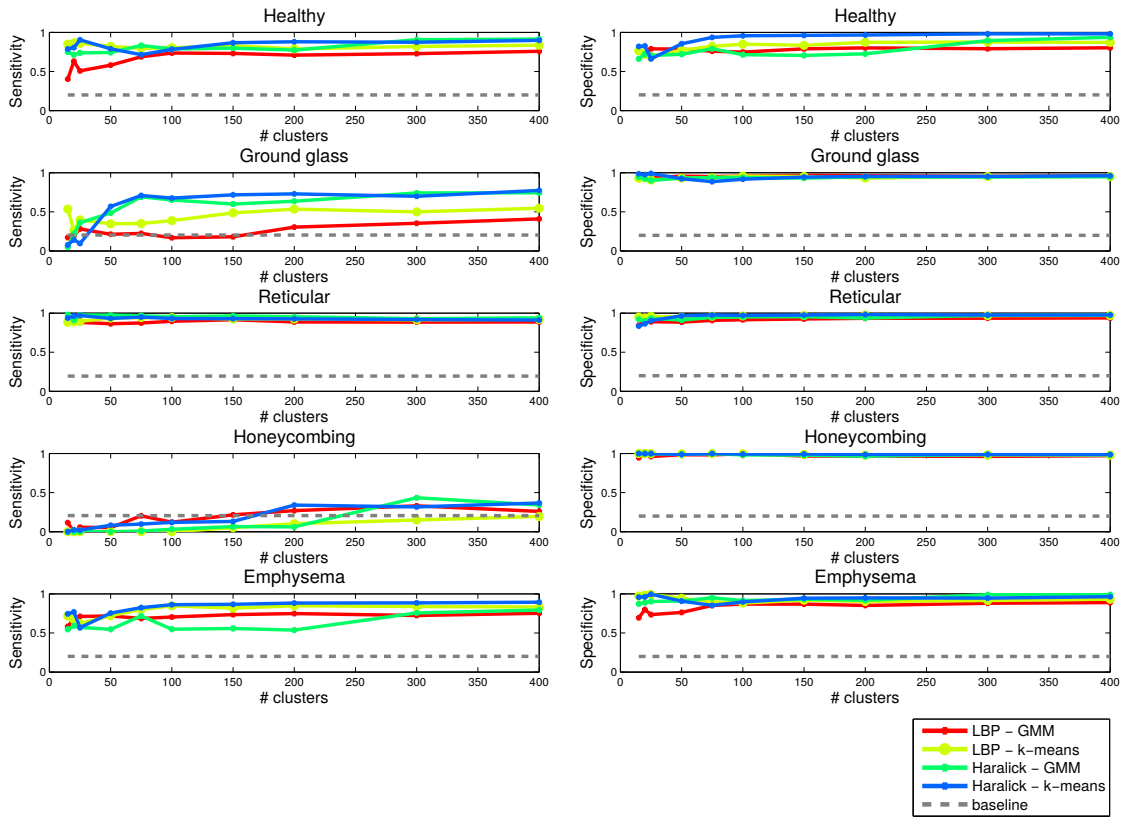


Figure 4.14: Sensitivity (left column) and Specificity (right column) values for each tissue class (rows) obtained while increasing the number of computed clusters (x-axis) from 15 to 400. Classification performances are given for each tested combination of texture descriptors and cluster approaches (color coded). Performance baseline (dashed grey) is obtained by random guessing.

are incorrectly identified as reticular patterns. Furthermore, ground class supervoxels incorrectly classified as reticular as indicated in the confusion matrix are observed in the volume of column 2.

Characteristics of clustering image region features Since the main idea of the approach proposed is based on the assumption that the partitioning of features results in clusters that represent prototypes of tissue classes, we expect that the true tissue classes of supervoxels close to a cluster center match with the clusters predicted tissue class.

Figure 4.16 thus shows the cluster centers nearest supervoxels for two clusters of each predicted tissue class. The predicted tissue class and its color coding is shown in the left column, where the middle column holds patches of nearest supervoxels obtained using Haralick features and the right column supervoxels obtained by BVW-LBP features. Clustering is performed using k-means. Distances are measured between the center (mean vector) of a cluster and the super-

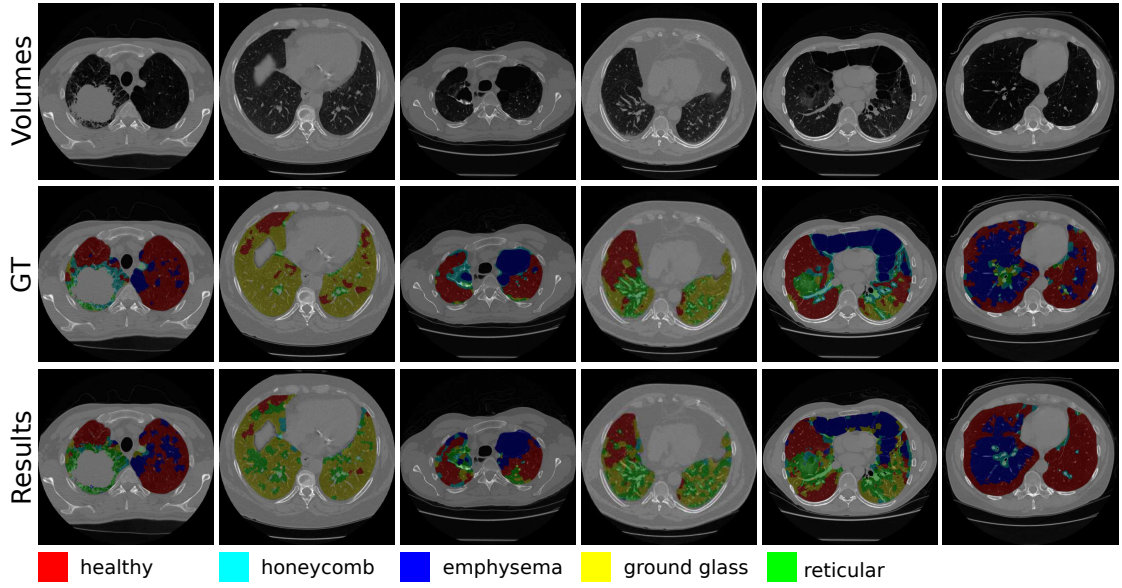


Figure 4.15: Axial slices of five volumes (top), ground truth (GT) annotations (not available during training, middle) and classification results (bottom), using Haralick features and k-means clustering with $k = 400$ clusters.

voxels feature vector. The frame color of an image patch indicates the ground truth labeling of the supervoxel. Please note that only a representative subset of all computed clusters (400) is shown to highlight characteristics of the texture descriptors and their classification performance.

Computation time The runtime of both texture descriptors has been tested on a Intel Xeon 2.67GHz CPU providing 24 cores. For Haralick features, a C implementation has been available, where the BVW-LBP feature extractor has been partially implemented in C and Matlab. On an average volume size of $512 \times 512 \times 490$ voxels, where the lungs are over segmented in average into 7105 supervoxels, the average computation time of BVW-LBP (15 minutes) is significantly higher than using Haralick features (45 seconds).

4.3 Summary

In this section experiments performed to evaluate both methods proposed in this thesis and their results have been described.

The approach proposed for unsupervised segmentation of anatomical structures in medical images has been evaluated in Section 4.1. Within the experiments in Section 4.1.4 the methods ability to identify structures in the reference space has been investigated in two steps. First the impact of unary terms to the resulting segmentation has been shown. Second the impact of binary terms which are used to model the assumption that spatially neighboring supervoxels are

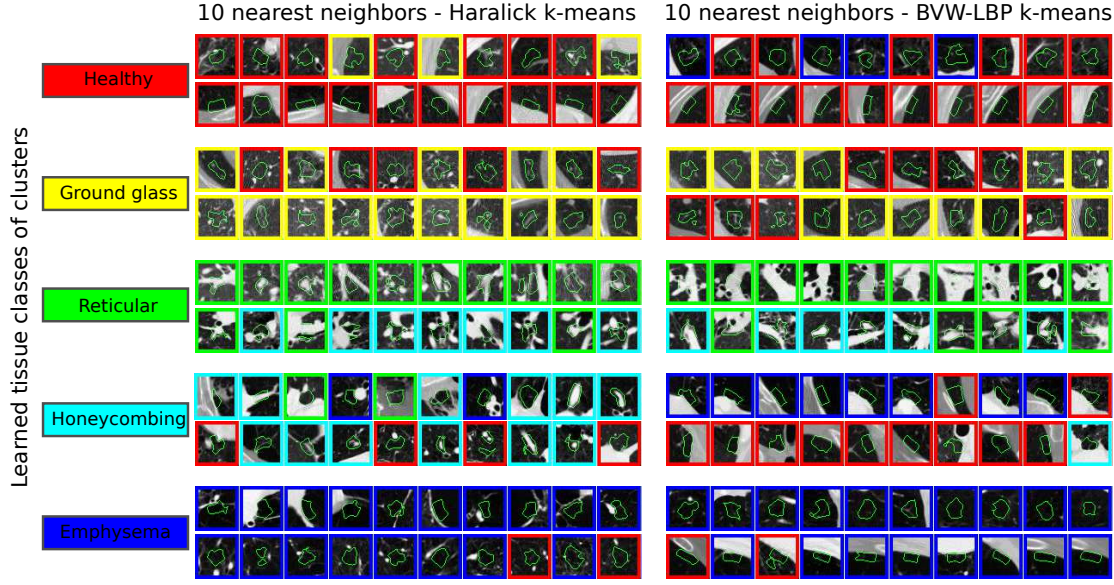


Figure 4.16: Ten nearest neighbors (supervoxels) of a subset of computed clusters (k-means, $k = 400$). Tissue class and the respective color coding with highest probability for a cluster is given in the right column. Two sets of nearest neighbors are shown for Haralick (middle) and BVW-LBP (right) features for each tissue class. The frame color coding of an image patch indicates the true labeling of supervoxels.

likely to belong to the same segmentation class has been investigated. For the evaluation purpose nine manual annotations of anatomical structures in the atlas volume have been available.

Experiments in Section 4.1.5 have been performed to evaluate the segmentation performance of the method proposed when labeling individual volumes. First we have investigated the impact of the a priori computed atlas segmentation and the initially computed segmentation of a volume on the resulting segmentation. Followed by the evaluation of the influence of binary terms, which again model the assumption that spatially neighboring supervoxels are likely to belong to the same segmentation class. For this purpose 14 volumes of the VISCERAL [32] data set have been used which carry manual annotations of 20 anatomical structures.

The method proposed for weakly supervised classification of healthy and pathological tissues in medical images has been evaluated in Section 4.2. The LTRC data set that provides 300 volumes with voxel-wise ground truth annotations of five tissue classes within the lungs has been used for this purpose. Evaluation has been performed in a 10 fold cross validation scenario. Four combinations of supervoxel texture features (BVW-LBP, Haralick features) and clustering methods (k-means, GMM) have been tested with ten different numbers of computed clusters while partitioning the sampled feature space to learn feature prototypes.

A detailed discussion of the experiments performed and their results is given in Section 5.1 of this thesis.

Discussion, Conclusion and Future Work

The final chapter discusses results of the performed experiments in Section 5.1, draws a conclusion of the work in this thesis including a summary of the methods proposed, the data in use for learning and evaluation as well as the main results in Section 5.2 and closes with thoughts on future work and possible improvements for both methods proposed in Section 5.3.

5.1 Discussion

The present section provides a discussion of the results shown in the previous sections of this work. Section 5.1.1 addresses results of the evaluation of the unsupervised segmentation of anatomical structures, where Section 5.1.2 discusses results of the weakly supervised classification of healthy and pathological tissue classes.

5.1.1 Unsupervised Medical Image Segmentation on Supervoxel Level

Results have shown that the approach proposed is able to identify the lungs, heart and liver (yellow, turquoise and light-green labels in Figure 4.3 *c*) and additionally several regions with similar texture properties that do not correspond to a specific organ or anatomical structure (see for instance the pink labeled region in Figure 4.3 *c* that contains tissue with bone structures).

Best overall Dice coefficients of the three detected organs in the reference space have been obtained using $L = 18$ clusters and a binary impact factor of $\alpha = 0.05$, leading to Dice coefficients of 0.94 (lungs), 0.8 (heart) and 0.45 (liver).

The illustration in Figure 4.4 indicates that multiple clusters correspond to one anatomical structure and that segmentation classes being dominant in one organ can also occur in other parts of the human body. This is explained by the diversity of the training data. First, there are two different modalities present in the training data (CT & CTce volumes) resulting in different in-

tensity values of the same anatomical structures. Second, the training data consists pathological and healthy volumes resulting in different visual appearing tissue types within an organ.

The incorporation of the latent atlas segmentation while labeling novel volumes ensures a consistent labeling of organs across all volumes. This effect can be observed in Figure 4.8. Here, the initial segmentation shows different labelings of the same organ in two volumes, where segmentations obtained by a combination of the latent atlas and the initial segmentation results in the same segmentation labeling of organs in both volumes.

Taking binary terms into account increases segmentation performances of the reference space and individual volumes in structures with high contrast to its surrounding tissue (lungs and the heart) but decreases segmentation accuracy of structures with low contrast to its surrounding tissue (liver). This effect is illustrated in Figures 4.6 and 4.10, corresponding Dice values are shown in Figures 4.7 and 4.11.

Best segmentation results while segmenting the lungs in individual volumes are obtained by setting $\beta = 0.6$ and $\gamma = 3$, resulting in average Dice values of 0.86 (CTce) and 0.66 (CT). Whereas liver segmentations reach average Dice values of 0.21 and 0.08 respectively within the same setting. The best results for liver segmentations are obtained with a binary term impact factor of $\gamma = 0.3$ resulting in average Dice values of 0.41 (CTce) and 0.34 (CT), whereas the lungs are segmented with average Dice values of 0.79 and 0.61 respectively.

Within the *VISCERALanatomy 2* benchmark [13] several approaches addressing the problem of segmenting multiple anatomical structures in medical imaging data have been proposed. Segmentation approaches of participating algorithms significantly outperform our approach when comparing obtained average Dice scores with reported results in [29], [14], [71], [77]. Here, all approaches segment the lungs with a minimum average Dice of 0.95 in both modalities and the liver with values 0.9 in CTce and > 0.82 in CT volumes.

The main difference between their approaches and the method proposed is that they require voxel-wise annotated training data. Goksel et al. [29] and Del Toro et al. [14] use annotations for atlas based label fusion, Spanier et al. [71] generate prior location and appearance knowledge from annotated training images and Wang et al. [77] propose a combination of atlas based segmentation and shape modelling, learned on annotated training data.

Comparing our method to the participants of *VISCERALanatomy 2* highlights the limitations of our approach. Due to the nature of unsupervised learning, the amount of detected structures and their segmentation quality highly depends on the chosen image features and their capability to form clusters that correspond to anatomical structures in the feature space.

5.1.2 Weakly Supervised Classification of Pathologies

Results show that the method proposed is capable of classifying healthy and pathological tissues with reasonable success in four (healthy, emphysema, ground glass, reticular) out of five tissue classes. Where the best performing method (Haralick features & k-means clustering) yields sensitivity values > 0.96 for all tissue classes, performance differences become visible on sensitivity values. Here, the best method setting reaches a sensitivity of 0.91 for reticular, 0.9 for healthy, 0.9 for emphysema and 0.77 for ground glass supervoxels, whereas the sensitivity when classifying honeycombing supervoxels is significantly lower (0.37).

Results in Figure 4.14 show that the ranking of the tested methods is consistent across all tissue classes. I.e. Haralick in combination with k-means performs better than the remaining methods in all tissue classes. Results furthermore indicate that k-means clustering outperforms GMM clustering independent from the feature descriptor as well as that Haralick features outperform BVW-LBP features in our context and are additionally cheaper to compute in terms of computation time.

The distribution of true positive labeled supervoxels of a tissue class during training as given in Table 4.4 and their corresponding classification performances in Figure 4.14. Results have shown that classes with the highest TP share rates and absolute amount of TP labeled supervoxels in the training set have sensitivity values significantly higher than the baseline (random guessing) (healthy, ground glass, reticular, emphysema), whereas the performance of honeycombing (TP share 0.91%) classification is comparable low for the best method (sensitivity 0.37) and even lower than baseline for all other tested methods.

Experiments have furthermore shown that classification performance in all classes increases with an increasing amount of computed clusters (see Figures 4.13 and 4.14). This indicates that some subtypes of tissue classes cover only a small region and can be located close to more dominant subtypes of other tissue classes in the feature space. When increasing the number of clusters, each partition is expected to cover smaller areas of the feature space and can thus distinguish such classes and overcome this problem. One could argue that best classification performance can thus be obtained by choosing a very high number of clusters to obtain small partitions in the feature space. This would result in over fitting the model. In the most extreme case partitions would cover only single observations, so that learning the label distribution in those partitions is not feasible since we assume to have multiple, weak labels for each observation. A partition with a single observation would then always predict the label of the observation with lowest term frequency.

Results in Figure 4.16 show that the majority of cluster centers nearest true supervoxel labels correspond to the predicted class of a cluster. Several observations from the confusion matrix in Table 4.5 can be observed as well. (1) Clusters predicting reticular classes contain several honeycombing supervoxels, which means that honeycombing supervoxels are likely to be incorrectly classified as reticular. (2) Reticular supervoxel occur almost only in reticular predicting clusters, which means that they are rarely predicted incorrectly. (3) If a supervoxel is incorrectly predicted as emphysema, it is most likely a healthy supervoxel. (4) Incorrectly classified healthy supervoxels are most likely ground glass supervoxels.

Comparing supervoxels of healthy and emphysema classes also highlight the fact that BVW-LBP features are sensitive to texture orientation, where Haralick features are rotation invariant. Here, BVW-LBP feature clusters contain supervoxels close to the border region of the lungs all oriented in similar directions, where Haralick feature clusters contain also supervoxels on the lung borders, but with independent orientation.

5.2 Conclusion

In this thesis two approaches that address two components (Segmentation and Classification) of typical CAD systems have been proposed. Inspired by the fact that CAD systems aid radiolo-

gists during clinical tasks to improve accuracy and productivity [18], [49] but suffer from the limitation that training data often requires manual annotations [39], [41], [87], [84], which is usually time consuming and expensive to acquire [16], the methods proposed are designed so that only data that is created during clinical routine is required.

5.2.1 Unsupervised Medical Image Segmentation on Supervoxel Level

The first method addresses the unsupervised segmentation of anatomical structures in medical images. It takes a set of medical images as input and computes an across all images consistent labeling of anatomical structures on supervoxel level. The method learns prototypes of occurring supervoxels in the training data by unsupervised clustering of supervoxel texture descriptors. Since these prototypes are expected to represent anatomical structures, the assignment of supervoxel to clusters is used to generate initial segmentations in all images. The registration of all images to an atlas allows us to learn a labeling in the atlas space, the *latent* atlas.

The final segmentation of an image is obtained by combining the initial computed segmentation based on supervoxel cluster assignments with the latent atlas, which is shared across all images. Both labeling procedures (in the atlas and in individual images) use MRFs to include relations between spatially neighboring supervoxels in an image.

The approach has been trained on a set of 450 CT and CTce volumes recorded at the radiology department of the AKH. For evaluation purposes manual annotations of 9 anatomical structures in the atlas have been available. To evaluate segmentation accuracy in novel images 14 volumes of the VISCERAL [32] have been used, that carry manual annotations of 20 structures.

Results have shown that the approach is able to identify three anatomical structures that correspond to organs in the human body. The lungs, the heart and the liver have been segmented with Dice coefficients of 0.94, 0.8 and 0.54 respectively in the atlas space. Since the VISCERAL data set does not provide ground truth annotations for the heart, the evaluation of segmentation accuracy of individual volumes has focused on the lungs and the liver. The incorporation of the assumption that spatially neighboring supervoxels are likely to belong to the same anatomical region increased segmentation performances in structures with high contrast to neighboring tissues (lungs and heart) but resulted in decreased performance values of liver segmentations where the contrast to neighboring tissue is comparable low.

We have shown that clustering supervoxel features results in partitions that represent feature prototypes of anatomical structures, which can be used to learn segmentations consistent across all volumes. We have furthermore shown that using MRFs to find a segmentation labeling on supervoxel level that combine atlas segmentations with local image information and incorporate constraints between spatially neighboring supervoxels has the potential to outperform segmentations obtained by only one of both components combined.

It is furthermore important to note that the unsupervised learning approach aims to discover anatomical structures with coherent appearance. It cannot compete with supervised approaches [29], [14], [71], [77] that rely on training with voxel-level expert annotations of anatomy. However, it hints at the potential of successful training with minimal supervision and future work will explore how additional clinical information can be incorporated.

5.2.2 Weakly Supervised Classification of Pathologies

The second method addresses the weakly supervised classification of healthy and pathological image regions of single organs. The method proposed is based on the idea that clustering supervoxel features results in partitions that represent prototypes of tissue classes. Weak labels, such as pathology terms that are extracted from a radiological report, that describe occurring pathologies in an image are assigned to all supervoxels of an image. By assigning supervoxels and their weak labels to clusters, we learn a probability table that predicts a single label given an observed cluster. This knowledge is then used to classify supervoxels of a novel image.

The method has been trained and evaluated in a 10 fold cross validation scenario on 300 chest CT scans of the LTRC [37] data set. Here, each volume carries voxel-wise ground truth annotations of five different lung tissue classes (healthy, emphysema, ground glass, reticular, honeycombing) and a binary segmentation mask of the lungs. During training the data set has been prepared to fit the addressed scenario, with at most two pathological findings per volume. Four combinations of texture descriptors (Haralick and BVW-LBP) and clustering techniques (k-means and GMM) has been evaluated.

Four out of five tissue classes have been classified with reasonable accuracy (healthy, ground glass, reticular, emphysema). Results have shown that the classification performance of the method proposed is related to the true positive rate of labeled supervoxles in the training data, since the worst classification performance is obtained for the tissue class with lowest rate of true positive labeled supervoxels (honeycombing).

It has been shown that Haralick features in combination with k-means clustering yield best overall classification results. Haralick features furthermore outperform BVW-LBP features independent from the clustering method within our setting and are additionally cheaper to compute in terms of time resources. It has also been shown that k-means clustering outperforms GMM clustering independent from the texture descriptor in use in the context of this work.

We have shown that the unsupervised clustering of supervoxel features leads to partitions of the feature space that are representative for subclasses of healthy and pathological tissue types. We have also shown that weak labels can be used to learn the underlying tissue classes of clusters and that this knowledge can then be used to classify regions of novel images. We have furthermore shown that supervoxels located in centers of clusters are likely to have the true labeling of the clusters predicted tissue class. This makes the clustering approach a suitable preprocessing step of CBIR systems, to reduce the search space a CBIR system has to evaluate when finding similar cases to a given query.

5.3 Future Work

Several aspects and components of both methods proposed within this thesis have possibilities for further improvement, which are addressed in the following two sections of this chapter.

5.3.1 Unsupervised Medical Image Segmentation on Supervoxel Level

In the present work only one supervoxel texture descriptor (BVW-LBP) has been used within the unsupervised image segmentation part. The framework is designed so that the feature extraction

technique is an interchangeable component. Future work would thus include the evaluation of different texture descriptors or the combination of different texture descriptors on the resulting segmentation performance.

Results have shown that the unsupervised nature of the method proposed clearly limits its segmentation performance compared to recently supervised image segmentation methods. However results have also shown that the usage of MRFs to combine the latent atlas with local image information and the incorporation of local spatial constraints on supervoxel levels is able to perform better than segmentations observed from only one of both components. Since large data sets carrying multiple annotations of anatomical structures recently became publicly available [32], future work includes the redesign of the system so that the latent atlas labeling is received from multiple annotated atlases, also the distribution of features can be learned from the annotated training set in supervised manner, which is expected to significantly improve segmentation results.

5.3.2 Weakly Supervised Classification of Pathologies

The method proposed shows promising results for tissue classes with high ratios of true positive labeled supervoxels in the training data. Future work include the adaption of the method so that the probability table that predicts tissue labels for clusters is estimated within an iterative EM strategy as suggested in [24], to improve classification performance especially for tissue classes with low TP labeling rates.

The addressed problem of unsupervised learning from text and images can also be interpreted as Multiple Instance Learning (MIL) [27], [1]. Future work would include the implementation and evaluation of MIL approaches in context of our problem setting.

The method proposed has been evaluated on a data set of 300 volumes carrying five different tissue classes, that has been prepared to fit our problem setting. Future work would include experiments on larger data sets carrying more tissue classes or data sets and weak labels obtained from clinical routine. The question of ground truth data and how to evaluate such a system if manual voxel-wise annotation is too expensive to acquire in this context addresses another important point of future work in this field. Furthermore the performance of the method proposed using training data where no single as well as more than two pathological findings are given has to be evaluated. Another possibility to test the method would include the usage of during clinical routine acquired training data and the evaluation on independent test cases such as the LTRC data set [37].

Appendix

Tables A.1, A.2, A.3 provide confusion matrices, sensitivity and specificity values of supervoxel classification performed on the LTRC [37] data set for the following combinations of texture features and clustering approaches: Haralick - GMM, BVW-LBP k-means, BVW-LBP - GMM.

		Predicted classes				
		Healthy	Ground glass	Reticular	Honeycombing	Emphysema
True classes	Healthy	1039047	57701	21182	8279	11650
	Ground glass	2856	26288	61219	136	0
	Reticular	60	180	5259	101	0
	Honeycomb	9276	264	5269	3150	55
	Emphysema	10449	10333	375	11614	121120
Sensitivity		0.91	0.74	0.94	0.34	0.79
Specificity		0.94	0.95	0.98	0.98	0.99

Table A.1: Confusion matrix, sensitivity and specificity values of supervoxel classification using Haralick features and GMM clustering with $k = 400$ clusters.

		Predicted classes				
		Healthy	Ground glass	Reticular	Honeycombing	Emphysema
True classes	Healthy	651601	58517	23596	23906	80595
	Ground glass	3838	19299	11678	499	0
	Reticular	56	310	4888	61	0
	Honeycomb	1880	886	4542	1811	206
	Emphysema	20189	879	125	4484	126865
Sensitivity		0.84	0.55	0.92	0.19	0.83
Specificity		0.87	0.93	0.97	0.98	0.93

Table A.2: Confusion matrix, sensitivity and specificity values of supervoxel classification using BVW-LBP features and k-means clustering with $k = 400$ clusters.

		Predicted classes				
		Healthy	Ground glass	Reticular	Honeycombing	Emphysema
True classes	Healthy	864155	47230	62460	33820	132807
	Ground glass	5822	14762	14940	447	65
	Reticular	121	367	4731	114	4
	Honeycomb	1879	436	4382	2429	318
	Emphysema	32771	428	963	4343	166214
Sensitivity		0.76	0.41	0.89	0.19	0.75
Specificity		0.8	0.96	0.94	0.97	0.88

Table A.3: Confusion matrix, sensitivity and specificity values of supervoxel classification using BVW-LBP features and GMM clustering with $k = 400$ clusters.

Bibliography

- [1] J. Amores. Multiple instance classification: Review, taxonomy and comparative study. *Artificial Intelligence*, 201:81–105, 2013.
- [2] B. Andres, T. Beier, and J. H. Kappes. Opengm: A C++ library for discrete graphical models. *CoRR*, abs/1206.0111, 2012.
- [3] S. G. Armato, M. F. McNitt-Gray, A. P. Reeves, C. R. Meyer, G. McLennan, D. R. Aberle, E. A. Kazerooni, H. MacMahon, H. J. R. MacMahon, and D. Yankelevitz. The lung image database consortium (LIDC): an evaluation of radiologist variability in the identification of lung nodules on CT scans. *Academic radiology*, 14(11):1409–1421, 2007.
- [4] B. B. Avants, Nicholas J. T., G. Song, P. A. Cook, A. Klein, and J. C. Gee. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, 54(3):2033 – 2044, 2011.
- [5] T. Ayer, M. US Ayvaci, Z. X. Liu, O. Alagoz, and E. S. Burnside. Computer-aided diagnostic models in breast cancer screening. *Imaging in medicine*, 2(3):313–323, 2010.
- [6] Z. A. Aziz, A. U. Wells, D. M. Hansell, G. A. Bain, S. J. Copley, S. R. Desai, S. M. Ellis, F. V. Gleeson, S. Grubnic, A. G. Nicholson, S. P. G. Padley, K. S. Pointon, Reynolds J. H., R. J. H. Robertson, and M. B. Rubens. HRCT diagnosis of diffuse parenchymal lung disease: inter-observer variation. *Thorax*, 59(6):506–511, 2004.
- [7] W. D. Bidgood, S. C. Horii, F. W. Prior, and D. E. Van Syckle. Understanding and using DICOM, the data interchange standard for biomedical imaging. *Journal of the American Medical Informatics Association*, 4(3):199–212, 1997.
- [8] J. A. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. *International Computer Science Institute*, 4(510):126, 1998.
- [9] C. M. Bishop and N. M. Nasrabadi. *Pattern recognition and machine learning*, volume 1. Springer New York, 2006.
- [10] A. Burner, R. Donner, M. Mayerhoefer, M. Holzer, F. Kainberger, and G. Langs. Texture bags: anomaly retrieval in medical images based on local 3d-texture similarity. In *Medical Content-Based Retrieval for Clinical Decision Support*, pages 116–127. Springer, 2012.

- [11] D. A. Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of remote sensing*, 28(1):45–62, 2002.
- [12] W. R. Crum, T. Hartkens, and D. L. G. Hill. Non-rigid image registration: theory and practice. *British Journal of Radiology*, 77:140–153, 2004.
- [13] O. A. J. del Toro, O. Goksel, B. Menze, H. Müller, G. Langs, M.-A. Weber, I. Eggel, K. Gruenberg, M. Holzer, A. Jakab, G. Kotsios-Kontokotsios, M. Krenn, T. Salas Fernandez, R. Schaer, T. Abdel Aziz, M. Winterstein, and A. Hanbury. Visceral-visual concept extraction challenge in radiology: ISBI 2014 challenge organization. In *Proceedings of the VISCERAL Challenge at ISBI, CEUR Workshop Proceedings*, pages 6–15, 2014.
- [14] O. A. J. del Toro and H. Müller. Hierarchical multi-structure segmentation guided by anatomical correlations. In *Proceedings of the VISCERAL Challenge at ISBI, CEUR Workshop Proceedings*, pages 32–36, 2014.
- [15] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.
- [16] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P.-A. Poletti, and H. Müller. Building a reference multimedia database for interstitial lung diseases. *Computerized medical imaging and graphics*, 36(3):227–238, 2012.
- [17] L. R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [18] K. Doi. Current status and future potential of computer-aided diagnosis in medical imaging. *The British Journal of Radiology*, 78:3–19, 2005.
- [19] K. Doi. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4):198–211, 2007.
- [20] R. Donner, S. Haas, A. Burner, M. Holzer, H. Bischof, and G. Langs. Evaluation of fast 2d and 3d medical image retrieval approaches based on image miniatures. In *Medical Content-Based Retrieval for Clinical Decision Support*, pages 128–138. Springer, 2012.
- [21] R. Donner, G. Langs, B. Mičušik, and H. Bischof. Generalized sparse MRF appearance models. *Image and Vision Computing*, 28(6):1031–1038, 2010.
- [22] M. Dorfer. A framework for medical-imaging-fragment based whole body atlas construction. Master’s thesis, Technical University of Vienna, 2013.
- [23] M. Dorfer, R. Donner, and G. Langs. Constructing an un-biased whole body atlas from clinical imaging data by fragment bundling. In Kensaku Mori, Ichiro Sakuma, Yoshinobu Sato, Christian Barillot, and Nassir Navab, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2013*, volume 8149 of *Lecture Notes in Computer Science*, pages 219–226. Springer Berlin Heidelberg, 2013.

- [24] P. Duygulu, K. Barnard, J. de Freitas, and D. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In Anders Heyden, Gunnar Sparr, Mads Nielsen, and Peter Johansen, editors, *Computer Vision, ECCV*, volume 2353 of *Lecture Notes in Computer Science*, pages 349–354. Springer Berlin / Heidelberg, 2006.
- [25] J. C. Felipe, A. J. M. Traina, and C. Traina Jr. Retrieval by content of medical images using texture for tissue identification. In *16th IEEE Symposium Proceedings, Computer-Based Medical Systems*, pages 175–180. IEEE, 2003.
- [26] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.
- [27] G. Fung, M. Dundar, B. Krishnapuram, and R. B. Rao. Multiple instance learning for computer aided diagnosis. *Advances in neural information processing systems*, 19:425, 2007.
- [28] M. J. Gangeh, L. Sørensen, S. B. Shaker, M. S. Kamel, M. De Bruijne, and M. Loog. A texton-based approach for the classification of lung parenchyma in ct images. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pages 595–602. Springer, 2010.
- [29] O. Goksel, T. Gass, and G. Szekely. Segmentation and landmark localization based on multiple atlases. In *Proceedings of the VISCERAL Challenge at ISBI. CEUR Workshop Proceedings, Beijing, China*, pages 37–43, 2014.
- [30] A. F. Goldszal, C. Davatzikos, D. L. Pham, M. X. H. Yan, R. N. Bryan, and S. M. Resnick. An image-processing system for qualitative and quantitative volumetric analysis of brain images. *Journal of computer assisted tomography*, 22(5):827–837, 1998.
- [31] G. Govaert. *Clustering and the Mixture Model*, pages 257–287. ISTE, 2010.
- [32] A. Hanbury, H. Müller, G. Langs, M. A. Weber, B. H. Menze, and T. S. Fernandez. Bringing the algorithms to the data: cloud-based benchmarking for medical image analysis. In *Information Access Evaluation. Multilinguality, Multimodality, and Visual Analytics*, pages 24–29. Springer, 2012.
- [33] R. M. Haralick, K. Shanmugam, and I. H. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6):610–621, Nov 1973.
- [34] W. Härdle and L. Simar. *Applied multivariate statistical analysis*, volume 22007. Springer, 2007.
- [35] K. Held, E. R. Kops, B. J. Krause, W. M. Wells III, R. Kikinis, and H.-W. Müller-Gärtner. Markov random field segmentation of brain MR images. *arXiv preprint arXiv:0903.3114*, 2009.

- [36] M. Hödlmoser. *Towards Exploiting Redundancy for 3D Scene Understanding from Videos*. PhD thesis, Technical University of Vienna, 2013.
- [37] D. R. Holmes III, B. J. Bartholmai, R. A. Karwoski, V. Zavaletta, and R. A. Robb. The lung tissue research consortium: an extensive open database containing histological, clinical, and radiological data to study chronic lung disease. In *The Insight Journal, MICCAI Open Science Workshop*, pages 1–5, 2006.
- [38] M. Holzer and R. Donner. Over-segmentation of 3d medical image volumes based on monogenic cues. In *Proceedings of the 19th CVWW*, pages 35–42, 2014.
- [39] Z. Huo, M. L. Giger, C. J. Vyborny, D. E. Wolverton, R. A. Schmidt, and K. Doi. Automated computerized classification of malignant and benign masses on digitized mammograms. *Academic Radiology*, 5(3):155–168, 1998.
- [40] S. Joshi, B. Davis, M. Jomier, and G. Gerig. Unbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23:S151–S160, 2004.
- [41] S. Kakeda, J. Moriya, H. Sato, T. Aoki, H. Watanabe, H. Nakata, N. Oda, S. Katsuragawa, K. Yamamoto, and K. Doi. Improved detection of lung nodules on chest radiographs using a commercial computer-aided diagnosis system. *American Journal of Roentgenology*, 182(2):505–510, 2004.
- [42] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1568–1583, 2006.
- [43] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Solving markov random fields using second order cone programming relaxations. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 1045–1052. IEEE, 2006.
- [44] R. S. Kumar and M. Senthilmurugan. Content-based image retrieval system in medical applications. *International Journal of Engineering*, 2(3), 2013.
- [45] Michael Lam, Tim Disney, Mailan Pham, Daniela Raicu, Jacob Furst, and Ruchaneewan Susomboon. Content-based image retrieval for pulmonary computed tomography nodule images. In *Medical Imaging*, pages 65160N–65160N. International Society for Optics and Photonics, 2007.
- [46] C. P. Langlotz. Radlex: A new method for indexing online educational materials 1. *Radio-graphics*, 26(6):1595–1597, 2006.
- [47] S. Lee, G. Wolberg, K.-Y. Chwa, and S. Y. Shin. Image metamorphosis with scattered feature constraints. *IEEE Transactions on Visualization and Computer Graphics*, 2(4):337–354, 1996.
- [48] S. Lee, G. Wolberg, and S. Y. Shin. Scattered data interpolation with multilevel b-splines. *IEEE Transactions on Visualization and Computer Graphics*, 3(3):228–244, 1997.

- [49] F. Li, H. Arimura, K. Suzuki, J. Shiraishi, Q. Li, H. Abe, R. Engelmann, S. Sone, H. MacMahon, and K. Doi. Computer-aided detection of peripheral lung cancers missed at CT: Roc analyses without and with localization 1. *Radiology*, 237(2):684–690, 2005.
- [50] Y. Liu, D. Zhang, G. L., and W.-Y. Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1):262 – 282, 2007.
- [51] M. Lorenzo-Valdés, G. I. Sanchez-Ortiz, A. G. Elkington, R. H. Mohiaddin, and D. Rueckert. Segmentation of 4d cardiac mr images using a probabilistic atlas and the em algorithm. *Medical Image Analysis*, 8(3):255 – 265, 2004.
- [52] T. Mäenpää and M. Pietikäinen. Texture analysis with local binary patterns. *Handbook of Pattern Recognition and Computer Vision*, 3:197–216, 2005.
- [53] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, 1997.
- [54] J. B. Maintz and M. A. Viergever. A survey of medical image registration. *Medical image analysis*, 2(1):1–36, 1998.
- [55] C. Mueller-Mang, C. Plank, H. Ringl, A. Dirisamer, and C. J. Herold. Interstitial lung diseases. In *Multislice CT*, pages 333–355. Springer, 2009.
- [56] M. Nadif and G. Govaert. *Cluster Analysis*, pages 215–255. ISTE, 2010.
- [57] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1):51 – 59, 1996.
- [58] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
- [59] S. Ourselin, A. Roche, G. Subsol, X. Pennec, and N. Ayache. Reconstructing a 3d structure from serial histological sections. *Image and vision computing*, 19(1):25–31, 2001.
- [60] S. H. Park, S. Lee, I. D. Yun, and S. Uk Lee. Hierarchical MRF of globally consistent localized classifiers for 3d medical image segmentation. *Pattern Recognition*, 46(9):2408 – 2419, 2013.
- [61] D. L. Pham, C. Xu, and J. L. Prince. Current methods in medical image segmentation 1. *Annual review of biomedical engineering*, 2(1):315–337, 2000.
- [62] J. C. Rajapakse, J. N. Giedd, and J. L. Rapoport. Statistical approach to segmentation of single-channel cerebral MR images. *IEEE Transactions on Medical Imaging*, 16(2):176–186, 1997.

- [63] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Non-rigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.
- [64] M. R. Sabuncu, B. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland. A generative model for image segmentation based on label fusion. *IEEE Transactions on Medical Imaging*, 29(10):1714–1729, 2010.
- [65] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:888–905, 1997.
- [66] I. Sluimer, A. Schilham, M. Prokop, and B. van Ginneken. Computer analysis of computed tomography scans of the lung: a survey. *IEEE Transactions on Medical Imaging*, 25(4):385–405, 2006.
- [67] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, Dec 2000.
- [68] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, M. Brady, and P. M. Matthews. Advances in functional and structural mr image analysis and implementation as FSL. *Neuroimage*, 23:S208–S219, 2004.
- [69] R. Socher and L. Fei-Fei. Connecting modalities: Semi-supervised segmentation and annotation of images using unaligned text corpora. In *CVPR*, pages 966–973. IEEE, 2010.
- [70] L. Sørensen, S. B. Shaker, and M. Bruijne. Texture classification in lung ct using local binary patterns. In Dimitris Metaxas, Leon Axel, Gabor Fichtinger, and Gábor Székely, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*, volume 5241 of *Lecture Notes in Computer Science*, pages 934–941. Springer Berlin Heidelberg, 2008.
- [71] A. B. Spanier and L. Joskowicz. Rule-based ventral cavity multi-organ automatic segmentation in CT scans. In *Medical Computer Vision: Algorithms for Big Data*, pages 163–170. Springer, 2014.
- [72] 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.
- [73] M. Tuceryan and A. K. Jain. Texture analysis. *The handbook of pattern recognition and computer vision*, 2:207–248, 1998.
- [74] A. Valentinitisch. *Computational Assessment of Bone Microarchitecture in the Diagnosis of Osteoporosis*. PhD thesis, Medical University of Vienna, 2014.

- [75] A. Valentinitich, J. Patsch, D. Mueller, F. Kainberger, and G. Langs. Texture analysis in quantitative osteoporosis assessment: Characterizing microarchitecture. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1361–1364, April 2010.
- [76] K. Van Leemput, F. Maes, D. Vandermeulen, and P. Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging*, 18(10):897–908, 1999.
- [77] C. Wang and O. Smedby. Automatic multi-organ segmentation using fast model based level set method and hierarchical shape priors. In *Proceedings of the VISCERAL Challenge at ISBI, CEUR Workshop Proceedings*, pages 25–31, 2014.
- [78] L. Wang, Y. Zhang, and J. Feng. On the euclidean distance of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1334–1339, 2005.
- [79] Y. Wang and G. Mori. A discriminative latent model of image region and object tag correspondence. In *Advances in Neural Information Processing Systems*, pages 2397–2405, 2010.
- [80] T. Werner. A linear programming approach to max-sum problem: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7):1165–1179, 2007.
- [81] H. Wildenauer, B. Mičušík, and M. Vincze. Efficient texture representation using multi-scale regions. In *Computer Vision—ACCV 2007*, pages 65–74. Springer, 2007.
- [82] X. Xie and M. Mirmehdi. A galaxy of texture features. *Handbook of texture analysis*, pages 375–406, 2008.
- [83] Y. Xu, M. Sonka, G. McLennan, J. Guo, and E. A. Hoffman. MDCT-based 3-d texture classification of emphysema and early smoking related lung pathologies. *IEEE Transactions on Medical Imaging*, 25(4):464–475, 2006.
- [84] Y. Xu, E. J. R. van Beek, Y. Hwanjo, J. Guo, G. McLennan, and E. A. Hoffman. Computer-aided classification of interstitial lung diseases via MDCT: 3d adaptive multiple feature method (3D AMFM). *Academic Radiology*, 13(8):969 – 978, 2006.
- [85] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo. Evaluating bag-of-visual-words representations in scene classification. In *Proceedings of the international workshop on Workshop on multimedia information retrieval*, pages 197–206. ACM, 2007.
- [86] J. S. Yedidia, W. T. Freeman, and Y. Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51(7):2282–2312, 2005.
- [87] H. Yoshida, J. Napi, P. MacEneaney, D. T. Rubin, and A. H. Dachman. Computer-aided diagnosis scheme for detection of polyps at ct colonography 1. *Radiographics*, 22(4):963–979, 2002.

- [88] B. Zitova and J. Flusser. Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000, 2003.
- [89] K. H. Zou, W. M. Wells, R. Kikinis, and S. K. Warfield. Three validation metrics for automated probabilistic image segmentation of brain tumours. *Statistics in medicine*, 23(8):1259–1282, 2004.