Mining Anatomical, Physiological and Pathological Information from Medical Images

Xiang Sean Zhou, Yiqiang Zhan, Vikas C. Raykar, Gerardo Hermosillo, Luca Bogoni
Siemens Medical Solutions USA, Inc., 51 Valley Stream Parkway, Malvern, PA 19355, USA
xiang.zhou@siemens.com

ABSTRACT

The field of medical imaging has shown substantial growth over the last decade. Even more dramatic increase was observed in the use of machine learning and data mining techniques within this field. In this paper, we discuss three aspects related to information mining in the domain of medical imaging: the target user groups ("for whom"), the information to mine ("what"), and technologies to enable mining ("how"). Specifically, we focus on three types of information: anatomical, physiological and pathological, and present use cases for each one of them. Furthermore, we introduce representative methods and algorithms that are effective for solving these problems. We conclude the paper by discussing some major trends in the related domains for the coming decade.

1. INTRODUCTION

The application of machine learning and data mining techniques in the domain of medical imaging has gained momentum in the last decade. This is illustrated in Figure 1, where we show the statistics gathered through a simple data mining experiment on Google Scholar, by searching for terms "medical image", "machine learning", and "data mining", and their combinations. Comparing the period from 2005 to 2011 to the same period a decade earlier, the number of articles related to "medical image" has doubled from 14600 to 32500. However, some topics increased much more than others: "medical image" articles which are further related to "machine learning" or "data mining" increased more than 10-fold, from 447 to 6150 and 417 to 4420, respectively.

A medical image is worth a thousand words; in some cases, a thousand *dollars* as well. Indeed, it can be *priceless* if it reveals an early sign of a disease which might be cured or at least properly managed thus increasing the quality of life if detected and treated early. However, several factors affect the ability to mine information out of medical images:

1. Medical images are most useful when acquired using a suitable protocol, and with the best parameters. This is often not easy to achieve. Data mining algorithms can provide guidance and automate some aspect of the image acquisition process based on, for example, learning from a set of "good" data.

2. There is an ever-increasing amount of data produced by a modern medical imaging scan [14]. It is not unusual for a study to produce more than 1000 images with each image containing 512 by 512 pixels. Additionally to this "data explosion", there is also mounting pressure to accomplish the same task in less and less time, in order to reduce the per-case cost.

3. There are many intrinsic subtleties and variations of the manifestation of anatomical structures and lesions. For example, a lesion in a breast (lung / colon) can be extremely inconspicuous in a mammogram (X-Ray / Computer Tomography (CT)).

4. The lack of automatic and quantitative tools is another substantial limitation. Insightful information (e.g., angulations of bones, alignment of vertebrae, diameters of nerve paths [4]) can be more easily and adequately reported and followed-up with the help of such tools.

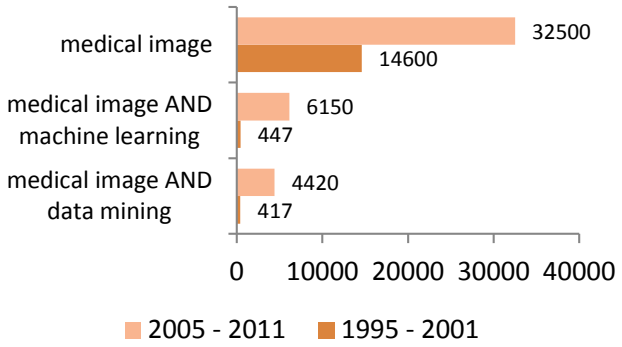

Figure 1: Increase of research activities in related fields as evidenced by the number of articles returned in Google Scholar, comparing the period 2005-2011 with a decade earlier 1995-2001. The exact query was formulated using a custom range and excluding patents and citations

In this paper, we provide an overview of target user groups (Section 2), discuss what information to mine from medical images (Section 3) and highlight some of the most effective methods and algorithms for mining anatomical, physiological, and pathological information from medical images (Section 4). We discuss four major trends that will affect the domain of medical image mining in the concluding section.

2. THE TARGET USER GROUPS

The improved access to medical image data using data mining techniques will benefit various consumers, including clin-

icians (technicians, radiologists, cardiologists, etc.), patients, and researchers.

2.1 Clinicians

As the primary interpreter and consumer of medical images, radiologists, cardiologists, and referring physicians will benefit from any additional meta-data or information that will help them *see more*, *see better*, or *see faster*. For example, for lung CT scans, it is a challenge for a radiologist to scroll through 500 or more images to look for lung nodules as small as 3-5mm. An automated algorithm can propose *candidates* of lung nodules to the radiologist, reducing the chance of misses. For cardiologists, an automated algorithm can detect and track the heart muscle in an ultrasound cardiac scan, and provide more accurate and consistent estimate for cardiac function [21].

Technicians, operating a scanner to acquire medical images, can also benefit from information mining, since it can help in achieving faster acquisitions as well as higher consistency and quality. The acquisition of medical images is often a multi-step procedure which begins with a scout scan intended to provide a rough overview of the patient. An algorithm that can detect anatomical information in the scout image can automatically steer the scanner to the correct angulations and focusing for subsequent scans, thus speeding up the scanning process and improving consistency [11].

2.2 Patients

In the era of *digital* imaging, it is increasingly common for patients to get access to their medical images in addition to the report from clinicians. Moreover, with the explosion of medical and health information on the Internet, and the unprecedented accessibility through search engines and specialized web sites, patients (as well as the general population) are becoming more proactive towards healthcare, and demanding more information and explanations. Image databases of skin lesions, for example, and associated matching tools can help the search of similar lesions with special appearances, and help educating the patient with regard to the underlying clinical condition. In the coming years, we believe that image mining tools can enable broader and deeper involvement of patients in more medical imaging fields.

2.3 Researchers

Undoubtedly, population models and statistics extracted from a large collection of medical images will be valuable for clinical, medical, biological, social, and public policy researchers. Large-scale population study and quantitative analysis can also provide public policy makers with adequate data for their policy decisions. For example, whether public health programs should reimburse for cancer screening using a particular imaging modality is a very complex decision. Information mining from large cohorts of clinical and image data could potentially provide the necessary evidence to support the most cost-effective option [3].

3. WHAT INFORMATION TO MINE ?

In this section, we broadly discuss three classes of information—*anatomical*, *physiological*, and *pathological*—that can be mined from medical images, and elaborate on their usefulness. This is not intended to be an exhaustive list. Anatomical information pertains to the structure of the human body, while physiological information is related to the function of the body. The former is directly visualized in most medical images, whereas the latter is less direct and less accessible since its manifestation and quantification are more dynamic in nature and require specific techniques. With the introduction of more functional imaging modalities (functional MRI, PET etc.), today we can also extract an abundance of physiological information from medical images. Pathological information refers to the abnormalities related to or caused by diseases that manifest in the medical images.

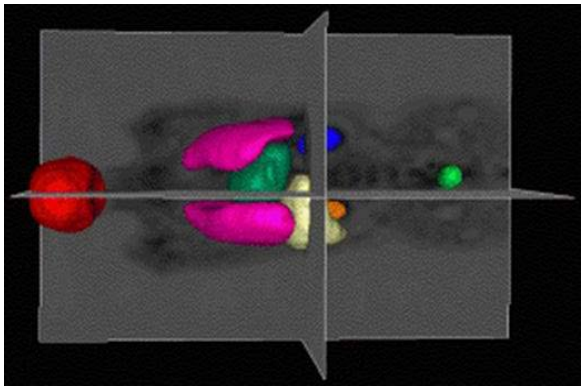

Figure 2: Automatic organ segmentation based on CT data which is overlaid on PET data for quantitative analysis.

3.1 Anatomical Information

Two basic questions can be answered related to an anatomical structure: 'Where is it in the image?' and 'What is it?'. Although an expert human user can answer these seemingly fundamental questions with a quick glance of the image, automatic answers are very useful as stepping-stones to other intelligent, organ-specific post-processing and analysis. Furthermore, while anatomical localization is part of the training of each physician, it may be a tedious and time consuming process.

3.1.1 Organ Location and Delineation

The first level of information that can be extracted from a medical image is the location and extent of organ structures. This may not be useful for a radiologist initially because they are trained to see such structures and further detect anomalies. However, automatic and precise 3D delineation can enable organ-specific quantitative analysis, and improve the reporting process. Figure 2 illustrates organs delineated using a CT image overlaid on a PET image to provide organ-specific FDG uptake statistics. This is useful in cancer detection and further follow-up [17].

3.1.2 Anatomical Labeling

In addition to general organ delineation, automatic labeling of anatomical structures such as vertebrae, ribs, or the blood vessels can speed up the reading and reporting workflow. To label such kind of repetitive (e.g., vertebrae) or complex structures (e.g., vessel tree), the relationship or configuration of multiple structures has to be considered in order to correctly label each component (see Figure 3). This is a rather challenging task since the variations in the population are quite high. For example, in the case of vertebrae: there are many asymptomatic variations [15], and correct la-

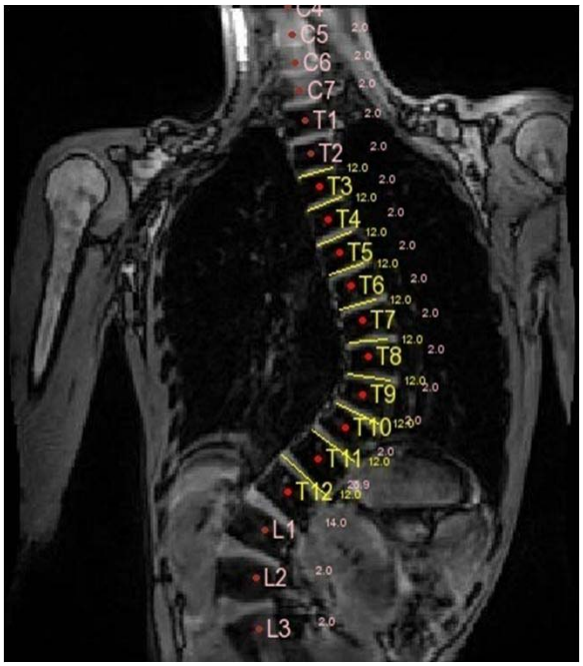

Figure 3: Automatic spine labeling and curved reformat (of a scoliotic patient) in a MR image.

beling requires a careful *walk-through* which can be tedious and error-prone in the cases of scoliosis or kyphosis. An automatic labeling algorithm would be a great help [18].

3.2 Physiological Information

In addition to static anatomical information, medical imaging can capture dynamic, functional information. The list of capabilities and possibilities is enormous and growing. In this section we explore only a few examples.

3.2.1 Motion

The first effective check of cardiac function is usually obtained through a cardiac ultrasound examination, in which cardiac contractions and relaxations are captured in a 2D or 3D "movie". An automatic motion-tracking algorithm can quantitatively measure the speed and extent of muscle contraction and relaxation within different coronary territories (see Figure 4). Since blocked coronary arteries result in reduced muscle activity, such motion analysis can aid the detection of cardiovascular disease [21].

3.2.2 Glucose or oxygen consumption

The functioning of the brain and the body requires oxygen and nutrients (glucose), and there are medical imaging modalities that can capture exactly these activities. Positron emission tomography measures the concentration of glucose at various locations in the body by "tracing" radioactive glucose molecules. Since cancer cells tend to consume a higher amount of glucose, they appear bright in a PET image. Functional MRI, on the other hand, can "read our brains" by exploiting the fact that local concentrations of oxygen associated with changes in blood flow become measurable using MRI.

For mining PET images, the challenge is the consistency and comparability of different acquisitions, even after standardization using the so-called "standardized uptake values" or SUV. The challenge of mining functional MRI (fMRI) images is that of obtaining a large, statistically significant number of well-controlled imaging studies.

Figure 4: Automatic analysis of cardiac function using a motion tracking algorithm and an ischemia classification algorithm. The red section in the right image highlights an ischemic segment of the myocardium.

3.2.3 Biochemical data

MR-spectroscopy provides tissue characterization by imaging the metabolite status. By using color-coding, we can visualize both morphology and biochemical status in the body. Such metabolite imaging can help to characterize lesions when a biopsy is difficult or impossible. It can also help the evaluation of therapy response before anatomical changes occur.

3.3 Abnormal and pathological information

During interpretation of medical images, clinicians typically have to *search* for abnormal structures and then *characterize* them in order to form a clinical conclusion.

The *search* for abnormalities is often very challenging due to large amounts of data and the fact that some of the findings are subtle and thus easy to overlook. Mining algorithms can learn from a large image database of a particular type of abnormality, and prompt the clinician with candidate locations in a new image. Examples of such abnormalities include lesions or micro-calcifications in a mammogram; lung nodules in chest X-rays or chest CT; and colon polyps in CT colonography.

The quantitative characterization of an abnormality can be challenging and time-consuming as well, either due to the need to compare against a normal population model—e.g., quantitative assessment of amyloid plaque deposition and neuronal tangle formation in the brain for the analysis of Alzheimer's disease using PET; or due to the 3D or 4D (i.e., 3D in time) nature of the problem—e.g., estimating the ejection fraction of the heart in echocardiogram [21].

4. METHODS AND ALGORITHMS

In this section we provide examples of novel medical image mining algorithms, covering different aspects of the domain, from image analysis to pattern recognition and machine learning.

4.1 Discovering anatomical signatures

Given a 3D medical image, an experienced radiologist is able to quickly localize a small anatomical structure, while an untrained college student might be completely clueless. The

"extraordinary" capability of radiologists comes from their knowledge of anatomical signatures, which is extracted from a large number of medical images during their medical training. More specifically, anatomical signature is the anatomical commonality across population in terms of appearance, shape and geometric configuration. Machine learning is an ideal tool to discover them from a large number of medical images.

4.1.1 *Appearance signatures*

Appearance signature originates from low-level image information. It consists of the combination of a set of image features, which optimally identifies specific anatomical primitives, e.g., anatomical landmark and organ boundary.

Anatomical landmarks Landmark is the most fundamental anatomical primitive. Two principles are employed to "discover" appearance signatures of landmarks from a large number of datasets. (1) A feature pool is constructed that contains an extremely large number of features (10,000+). (2) A machine leaning algorithm is used to select the most distinctive features and combine them. To be more precise, feature construction starts from a set of mother functions, each of which consists of one or more 2D/3D rectangle functions of the form

$$H(x) = \sum_{i=1}^{N} p_i R(x - a_i), \tag{1}$$

where $p_i \in \{-1, +1\}$ is the polarity of the rectangle function defined as $R(x) = 1$ if $\|x\|_\infty \leq 1$ and zero otherwise. a_i is the translation and x is the image location where the feature is being computed. Figure 5 shows examples of some representative mother functions. By scaling and convolving these mother functions with the original image I, a set of spatial-frequency spaces F are constructed as: $F_l = H_l(sx) * I(x)$, where s and l denote the scaling factor and the inter of the mother function respectively. For any pixel/voxel x_0 under study, its feature vector $\Im(x_0)$ is built by sampling these spatial-frequency spaces F in its neighborhood.

$$\Im(x_0) = \cup_l \{F_l(x, s_j) \mid x \in N(x_0), s_{\min} < s < s_{\max}\}, \tag{2}$$

where $N(x_0)$ defines the neighborhood of x_0, and $s_{\min}$ and $s_{\max}$ define the minimum and maximum values of the scale used. By tuning the sampling scheme, one can easily construct a feature space with 10,000+ features for any pixel/voxel.

Given such a large feature pool, Adaboost is used to select the features and combine them for landmark identification. In principle, the problem is formulated as binary classification, where pixels/voxels close to manually annotated landmarks are considered as positive training samples and remaining pixels/voxels are regarded as negatives. The output of the learnt classifier is the likelihood of a specific landmark at position x. This is in fact the learned appearance signature of an anatomical landmark [19].

Organ boundary In the same spirit, the signature of organ boundary can also be learnt. However, the challenge here is that some organs, e.g., liver, exhibit highly heterogeneous appearance characteristics along boundaries (see Figure 6). Hence, it is almost impossible to capture the boundary signature with a single classifier.

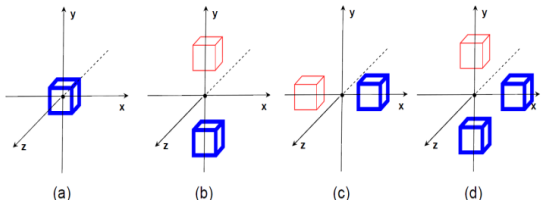

Figure 5: Some representative examples of mother functions. Blue/thick and red/thin boxes denote the non-zero ranges of 3D rectangle functions with positive and negative polarities, respectively. (Figure reprinted from [16])

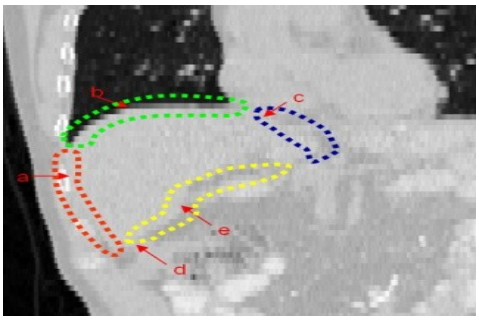

Figure 6: Heterogeneous appearance characteristics along the liver boundary.

We proposed a "divide-and-conquer" strategy [17] to learn spatially adaptive boundary signatures. The basic idea is to divide the organ surface/contour into several partitions, whose boundary appearance characteristics are relatively more homogenous. Now, it is easy to learn the local boundary "signatures" of these partitions. To achieve this goal, we designed an iterative approach, in which Adaboost based learning and Affinity Propagation clustering are alternatively performed. This approach starts by training an Adaboost classifier with all samples around the organ boundary. The selected features are then used by Affinity Propagation to cluster vertices into different groups. Spatial proximity of vertices is also incorporated to ensure continuous partitions. As shown in Figure 7, an organ surface/contour can be eventually divided into several sub-divisions. Each sub-division is attached to a learnt Adaboost classifier, which captures the boundary signature in the local region.

4.1.2 *Geometric/shape signatures*

Different from photographs taken in daily life, where objects often have arbitrary context, e.g., a person can stand by a car, a house or another person, anatomical structures in medical image have strong geometric/shape correlations. For example, heart is located in-between two lungs. Hence, geometric/shape signatures play an important role in the interpretation of medical images, especially when appearance cues become weak or misleading due to severe diseases or

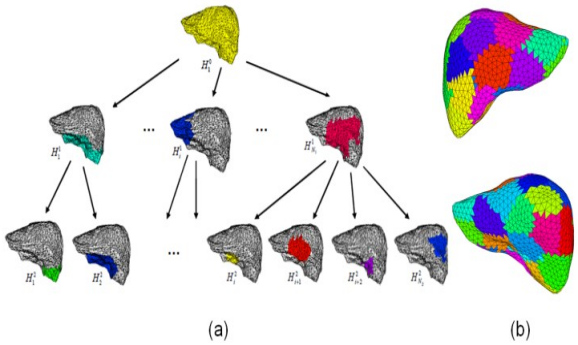

Figure 7: (a) Hierarchy of liver surface partition. (b) Details of partitioned liver surface, where different colors denote different sub-division. (Figure reprinted from [17])

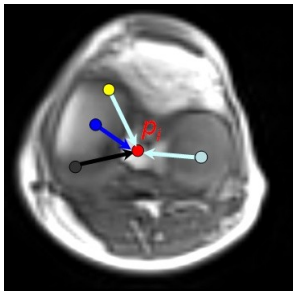

Figure 8: Schematic explanation of voting scheme. Detection p_i receives votes from other landmarks.

imaging artifacts. In addition, geometric/shape signatures themselves are critical diagnostic evidence for various clinical studies.

Spatial configurations Spatial correlation across different anatomical primitives is one strong geometric signature, which can be employed to correct detections derived by misleading appearance cues. Accordingly, the robustness of this signature with respect to erroneous/missing detections becomes very critical. We proposed a group-wise configuration model to characterize the geometric signature [11]. Instead of modeling the spatial correlations across a large group of landmarks, we learn a large number of spatial correlations across small groups (2, 3 or 4). At run-time, each detected location p_i receives a set of votes from other landmarks in different groups (see Figure 8). All votes received by p_i will be combined in a non-linear way to evaluate the eligibility of p_i. Erroneous detections are then iteratively removed and predicted by learned group-wised spatial configurations. Since the learned geometric signature is embedded in a consensus of a large number of small landmark groups, it is robust to individual landmarking errors that are often caused by diseases/imaging artifacts.

Shape Shape is a more sophisticated description of spatial relations across different anatomical primitives.

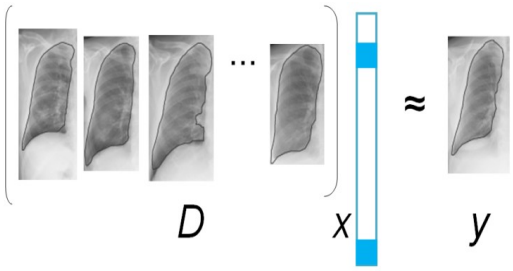

Figure 9: Schematic explanation of sparse shape composition. (Figure reprinted from [20])

The limitation of existing shape signature/prior modeling mainly comes from the parametric assumptions of shape statistics. To alleviate this limitation, we designed a parameter-free approach to characterize shape signature [20]. Our method is designed according to two observations about sparsity in medical images: (1) A shape instance can be approximated by a sparse combination of other shapes in the same category. (2) A shape instance derived by low-level appearance information might include gross errors. However, these gross errors are also sparse.

Therefore, for any shape instance derived by appearance cues, we approximate it by a sparse composition of shape instances of the same organ (see Figure 9). In this manner, the shape signature/prior is applied on-the-fly. Mathematically, our objective function is formulated as:

$$\arg_{x,e,\beta} \min \|T(y,\beta) - Dx - e\|_2^2, , \|x\|_0 \leq k_1, \|e\|_0 \leq k_2, \quad (3)$$

where y is the input shape derived by appearance cues, which might include gross errors. D is the shape repository containing all training shape instances. $T(y,\beta)$ is the transformation that brings y to the training shape space. x and e denote the composition coefficients and gross errors, respectively. Both of them are imposed with sparsity constraints, which is in accordance with the two sparsity observations identified above.

The objective function can be optimized efficiently using sparse learning theory. Compared to classic shape prior/signature modeling, our method is robust to gross errors and flexible to arbitrary shape statistics. In addition, given new training shapes, the shape repository can be easily updated without extensive re-training, which satisfies real-world medical imaging applications.

4.2 Feature engineering

One effective way to mine the information contained in images is by explicitly computing numbers that quantify certain properties of the structures of interest. This is often what is done for automatic detection of abnormalities since the abnormality is described by experts as possessing specific characteristics, or features. For example, a breast MR radiologist may describe a mass as a structure having definable margins with a separable distinct edge from the surrounding glandular tissue and having no normal tissue within it. He or she may proceed to describe ductal and segmental

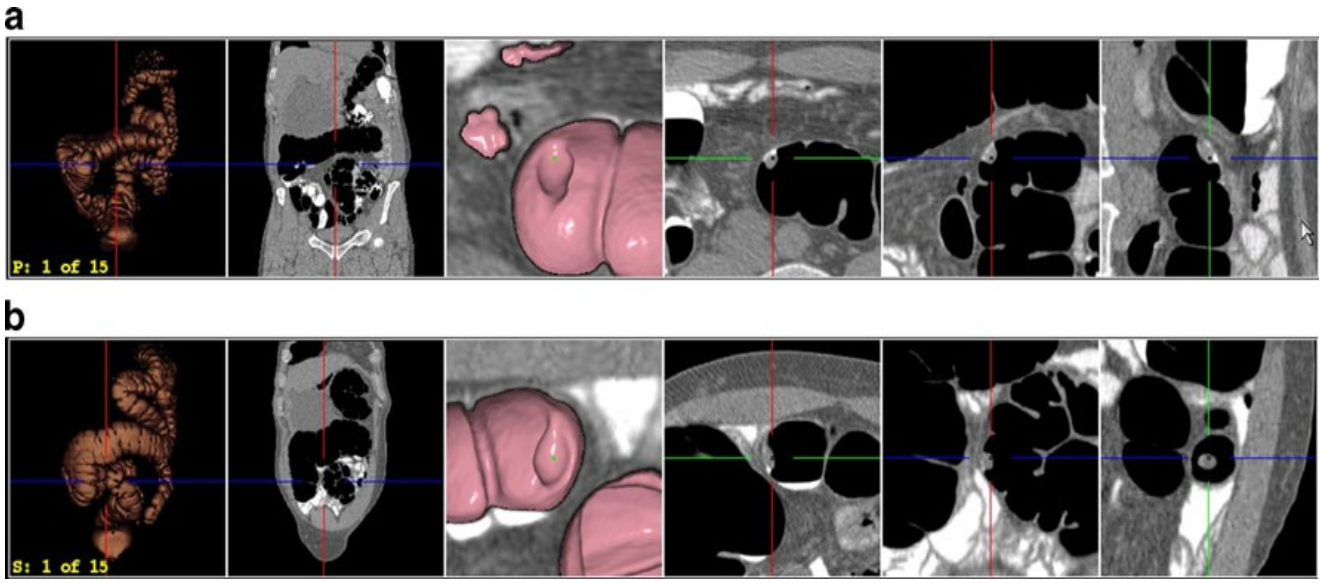

Figure 10: Illustration of various components for polyp detection in CT images. The colon wall needs to be segmented out from the image, liquid tagging must be removed and candidate structures need to be described with shape and texture features to distinguish them from false positives. The candidates should be visible in both prone (a) and supine (b) views.

enhancements as linear or sheet-like enhancements not definitely in a duct and that cannot be otherwise characterized, probably representing the same process at two different resolutions, with or without discernible margins and with or without branches.

Similarly, an abdominal radiologist may describe a colonic polyp as a protuberance in the colonic wall at least 6 mm in diameter and at least 3 mm in height that can be flat, sessile or pedunculated in shape, its pattern of attenuation being characteristic of soft, non-fatty tissue, usually homogenous. In order to design features that are useful in detecting and characterizing such structures, it is often necessary to solve various intermediary problems. For example the simple fact that a colonic polyp is a structure on the colon wall implies that an automatic detection algorithm should first identify and separate (segment) the colon wall from the entire image. Without attempting to be exhaustive, we discuss specific examples of such features in the context of the detection and characterization of colonic polyps and breast MR lesions.

4.2.1 Detection and characterization of polyps

The detection of polyps in CT images requires the segmentation of the colon wall from the image. In modern preparations, this requires the virtual removal of tagging liquid, a process known as electronic cleansing (Figure 10). Once the colon is segmented out, a candidate-generation stage computes a set of geometric and texture features on each location of the wall. A cascading approach is used to progressively reduce the number of candidate locations before computing more computationally expensive features. Morphological features include global shape features, analysis of curvature patterns and specific properties of the gradient vector fields around each candidate. A classifier is then trained to distinguish true polyps from false positive structures based on a set of annotated polyps. The segmental location of the finding can be automatically detected by counting the folds from the rectum and by distance from it. Also the size of each candidate is estimated by linear regression from a subset of the features. Figure 11 shows typical examples of a sessile and a flat polyp. Figure 12 illustrates some examples of false-positive structures commonly misinterpreted as polyps: folds, tagged stool residuals and the ileocecal valve.

4.2.2 Characterization of lesions in Breast MRI

Contrast enhanced MR sequences are a powerful diagnostic tool for the detection of lesions in breast. Typically, the diagnosis begins by identifying suspicious regions of enhancement in post contrast acquisitions with respect to a pre-contrast one. Automating this process is therefore required for a computer-aided system.

Segmentation of vascular structures Because lesions usually contain a high number of newly created vessels, perfusion of a contrast agent makes the lesions appear brighter than the background. Automatically segmenting and measuring the lesions saves the radiologist time and makes these measurements more consistent across readers.

One set of features extracted from breast MR images aims at distinguishing locations on vascular structures in order to help the segmentation and characterization processes. These features are based on the local differential structure of the image. Figure 13 illustrates the result of separating the enhancing structures on vessels from ones outside vessels (potential tumors).

Shape characterization and pharmacokinetic analysis In breast MR, lesions are described both in terms of their morphology and their pattern of enhancement. Lesions are classified into focus of enhancement, regular or irregular mass or non-mass-like enhancement.

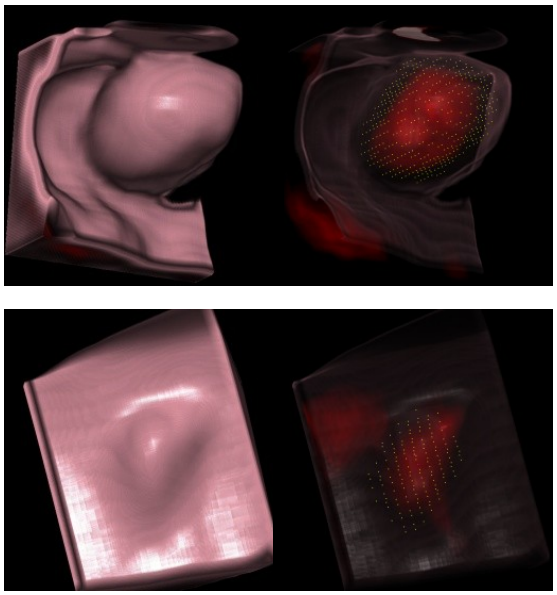

Figure 11: A typical sessile polyp (top row) and a typical flat polyp (bottom row). In addition to their shape, polyps are characterized by their internal texture, which is usually homogeneous non-fatty soft tissue (muscle-like), illustrated here by the red colored area under the transparent rendering of the colonic tissue.

Masses are described by their shape which can be round, oval, lobulated or irregular, their margin, which can be smooth, irregular or spiculated and their internal enhancement, which can be homogeneous or heterogeneous. Non-mass-like enhancement are categorized as focal, linear, ductal, segmental, regional, or diffuse, while their pattern of enhancement can be described as homogeneous, heterogeneous, stippled, clumped, patchy, dendritic, reticular, symmetric or asymmetric.

Each of these image characteristics can be translated into one or more imaging features, extracted from the local structure of the image and from texture analysis. Machine learning methods can then be applied to map these features into two or more classes of lesions.

One additional dimension of the study of breast MR imaging is the analysis of the way different areas of the anatomy absorb and release contrast agents. This is called pharmacokinetic analysis and its goal is to provide a framework where the kinetics of the contrast agent within the tissue of interest can be quantitatively described and compared across data sets from one or more patients and/or MR systems. Malignant tissues differ from benign tissues in how contrast agents flow in and leak out. Pharmacokinetic analysis aims to quantify the wash-in and wash-out of the contrast agent towards differentiating malignant and benign lesions [8] (Figure 14).

Both the morphological and temporal analysis of lesions are affected by artifacts due to patient motion during the acquisition of the temporal sequence. Motion artifacts may introduce misinterpretations of the lesion intensity over time, as well as changes in the lesion morphology [5]. Motion correction can reduce false categorization of lesion kinetics and provide better lesion delineation (Figure 15).

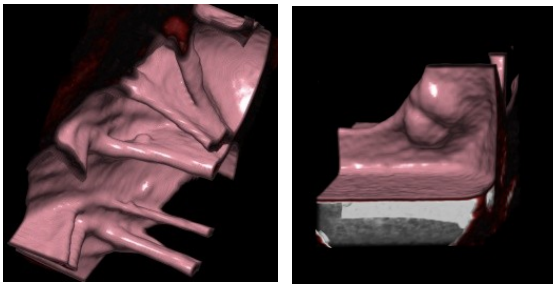

Figure 12: Two types of frequent false positives in the detection of polyps: folds and tagged stool (left) and the ileocecal valve (right). The small tagged stool on the left and the ileocecal valve roughly illustrate the range of sizes that are targeted for detection.

4.3 Non-standard learning paradigms

From a machine learning perspective, early stage detection of cancer from medical images can be abstracted as a simple supervised binary classification problem. One could potentially use standard off-the-shelf classifiers like logistic regression, support vector machine, neural network, etc. However most of the standard off-the-shelf supervised learning algorithms are generally developed for an ideal world. They often make strong assumptions which make them less suitable to be applied directly to real world messy data. For example training points are often noisily labeled, training samples are not independent and identically distributed, it is difficult or impossible to acquire the objective ground truth, the desired performance metric may be quite different etc. For these reasons most of the basic assumptions in developing classification algorithms have to be questioned. Suitable modifications must be introducced to model these deviations from the ideal scenario which can give a significant improvement in performance over off-the-shelf standard classification algorithms to account for more realistic conditions. In this section we will describe some of the assumptions that can possibly break down and how we can re-engineer some standard learning algorithms to suit our needs. Specifically we will discuss three non-standard learning paradigms which can deal with noisy, subjective, and partial label information.

4.3.1 Multiple instance learning

In a conventional supervised learning scenario it is always assumed that the label is provided for every instance. However for many practical applications labels are available at a much higher granularity and are not available for every instance. For example in medical imaging applications the radiologist who provides us the ground truth just marks the location of the lesion. The lesions are often irregular in shape and are of different sizes. The computer algorithm designed to detect these lesions produces a lot of training examples which are spatially close to each other (see Figure 16 for an illustration). All these examples point to the same ground truth. A single instance classifier considers all these examples as positive. In practice it often happens that

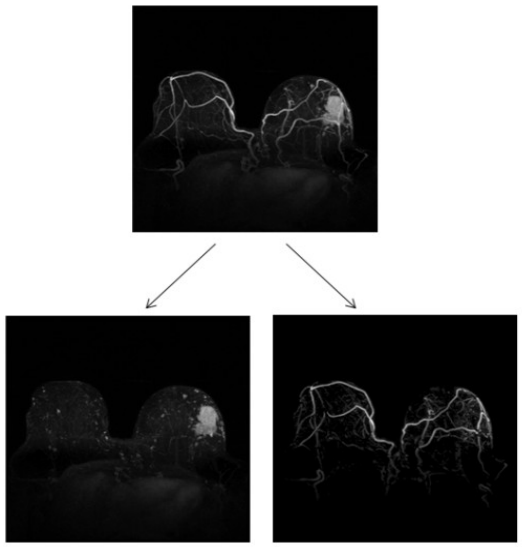

Figure 13: Local differential features are computed that allow distinguishing locations of enhancement on vascular structures from other structures such as tumors.

there will be a lot of negatives which mistakenly are labeled as positives.

In the multiple instance learning (MIL) framework the training set consists of bags. A bag contains many instances. All the instances in a bag share the same bag-level label. A bag is labeled positive if it contains at-least one positive instance. A negative bag means that all instances in the bag are negative. The goal is to learn a classification function that can predict the labels of unseen instances and/or bags. Figure 16 illustrates that MIL can yield very different classifiers compared to conventional single instance learning. The single instance classifier on the left is trying to reject as many negative candidates as possible and detect as many positives as possible. The MIL classifier on the right tries to detect at-least one candidate in a positive bag and reject as many negative candidates as possible.

There is another important reason why MIL is a natural framework for medical imagining applications. The candidate generation algorithm produces a lot of spatially close candidates. Even if one of these is highlighted to the radiologist and other adjacent or overlapping candidates are missed, the underlying lesion would still have been detected. Hence while evaluating the performance of such systems we use the bag level sensitivity, that is, a classifier is successful in detecting a lesion if at least one of the candidates pointing to it is predicted as a lesion. MIL lends itself to model our desired accuracy measure during training itself.

We incorporate the definition of a positive bag to modify the link function used in logistic regression [12; 7]. Standard logistic regression uses a sigmoid link function to model the probability of the positive class. For MIL since we have the notion of a positive bag the probability that a bag contains at-least one positive instance is one minus the probability that all of them are negative. The proposed algorithm selects features and designs the classifier jointly. One interesting empirical outcome was that the multiple instance model was able to select many fewer features; almost half the number of features selected by the single instance approach.

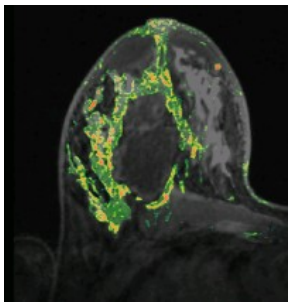

Figure 14: Pharmacokinetic analysis aims to quantify the wash-in and wash-out of the contrast agent towards differentiating malignant and benign lesions.

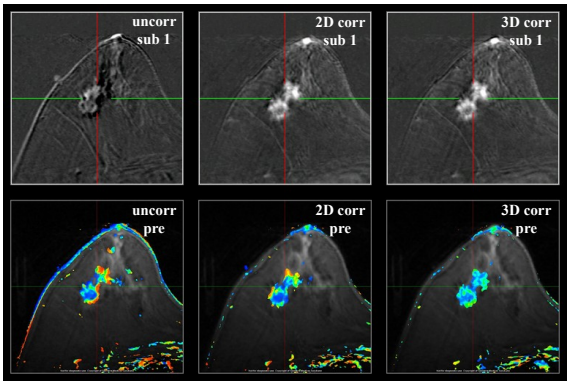

Figure 15: There are 2D and 3D motions resulting in prominent ridge-and-valley artifact, best seen on the uncorrected images. Uncorrected and 2D-corrected images each show different areas of artifactual washout, in part related to 3D motion. 3D-registered images correctly show plateau and progressive enhancement without washout.

4.3.2 Multiple expert learning

In many medical imaging applications it is actually quite difficult to obtain the ground truth. The actual gold standard (whether it is cancer or not) can be obtained from biopsies, but since it is an expensive and an invasive process, often data mining systems are built from labels assigned by multiple radiologists who identify the locations of malignant lesions. Each radiologist visually examines the medical images and provides a subjective (possibly noisy) version of the gold standard. The radiologists come from a diverse pool including luminaries, experts, residents, and novices and very often there is lot of disagreement among the annotations. For many tasks the labels provided by the annotators are inherently subjective and there will be substantial variation among different annotators.

With the advent of crowdsourcing services like Amazons Mechanical Turk it is quite inexpensive to acquire labels from a

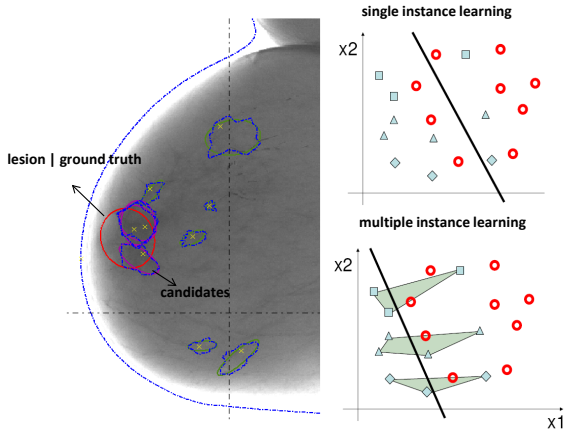

Figure 16: *Illustration of multiple instance learning* (left) A mammogram of the right breast illustrating the concept of multiple candidates pointing to the same ground truth. The red ellipse is the lesion as marked by the radiologist (ground truth). The blue contours are the candidates generated by our algorithm. (right) Illustration of single-instance learning (top) and multiple instance learning (bottom) for a toy problem. The red circles are negative candidates. The blue shapes are positives. There are three positive bags (square, triangle, and diamond).

large number of annotators (possibly thousands) in a short time. In situations like these, the performance of different annotators can vary widely (some may even be malicious), and without the actual gold standard, it may not be possible to evaluate the annotators. In [13] we proposed a probabilistic approach for supervised learning which addresses the following three issues simultaneously: (1) how to adapt conventional supervised learning algorithms when we have multiple annotators providing subjective labels but no objective gold standard?; (2) how to evaluate systems when we do not have an absolute gold-standard?; (3) how to estimate how reliable/trustworthy each annotator is ?. The last problem is particularly relevant when there are a large number of annotators.

The commonly used majority voting scheme uses the labels on which the majority agree as an estimate of the actual gold standard. We proposed a Bayesian approach that jointly learns the classifier, the annotator accuracy, and the unknown true label. The final estimation is performed by an Expectation Maximization algorithm that iteratively establishes a particular gold standard, measures the performance of the experts given that gold standard, and refines the gold standard based on the performance measures. Experimental results indicate that the proposed method is superior to the commonly used majority voting baseline. A novel feature is that the proposed algorithm learns the classifier and the ground truth jointly in a way the classifier is allowed to influence the ground truth. The method was successfully applied to a model for prediction of malignancy for breast tumors in MR with subjective assessments from multiple radiologists in the absence of biopsy results.

4.3.3 Multiple task learning

We are often faced with a shortage of training data for learning classifiers for a task. However we may have additional data for closely related, albeit non-identical tasks. For example our data set includes images from CT scanners with two different reconstruction kernels. While training the classifier we could ignore this information and pool all the data together. However, there are some systematic differences that make the feature distributions slightly different. We could also train a separate classifier for each kernel, but a large part of our data set is from one particular kernel and we have a smaller data set for the other. Alternatively we could use the framework of multi-task learning [12] that tries to estimate models for several classification tasks in a joint manner. Multi-task learning can compensate for small sample size by using additional samples from related tasks, and exchanging statistical information between tasks.

5. TRENDS

We discussed the increased research activities in medical image mining in the last decade and gave examples of use cases and the methods employed. We conclude by identifying some trends that will shape the research directions of this domain.

1. **Personalization** Advances in science and medicine are clearly pointing in the direction of personalization. Medical imaging plays important roles in this transformation. Already there are many functional and molecular imaging modalities and computing techniques that are opening doors to new ways of personalized diagnosis, treatment and care [6].

2. **Specialization** Highly specialized, high performance, and value-added image mining software systems will continue to grow, together with the growth of new imaging modalities and hardware, and new diagnosis and surgical techniques in cardiology, oncology and neurology. Domain knowledge is the key to success in this direction.

3. **Generalization** On the other hand, advances in algorithms in the domain of general machine learning, data mining, and statistics will facilitate the development of more powerful and more general tools that can deal with different imaging modalities in a unified way. Disruptive innovation here could sweep away old generations of specialized solutions in multiple domains, or at least significantly transform their landscape. The data mining community should aspire to facilitate such revolutionary changes in the medical imaging domain.

4. **Cost reduction** The cost pressure on the healthcare system will continue to rise. As a result, productivity gain will be a near- to mid-term priority, and *do more with what you have* will be the driver for image mining algorithm research, which should focus on extracting more clinical value out of medical images.

6. ACKNOWLEDGEMENTS

We appreciate the comments and suggestions from the editors which have helped to improve the quality and the presentation of the paper.

7. ADDITIONAL AUTHORS

Zhigang Peng, Maneesh Dewan, Matthias Wolf, Yoshihisa Shinagawa, Sangmin Park, Marcos Salganicoff (Siemens Medical Solutions USA, Inc., 51 Valley Stream Parkway, Malvern, PA 19355, USA) and Lakshminarasimhan Raghupathi, M.S. Dinesh, Pandu Devarakota (Siemens Technology and Services, 84 Keonics Electronic City, Hosur Road, Bangalore 560 100, India)

8. REFERENCES

[1] P. Angelini, J. A. Velasco, and S. Flamm. Coronary Anomalies: Incidence, Pathophysiology, and Clinical Relevance. *Circulation*, 105:2449–2454, 2002.

[2] J. G. Fletcher, F. Booya, R. M. Summers, D. Roy, L. Guendel, B. Schmidt, C H McCollough, and J L Fidler. Comparative Performance of Two Polyp Detection Systems on CT Colonography. *American Journal of Roentgenology*, 189(2):277–282, 2007.

[3] T. W. Freer and M. J. Ulissey. Screening Mammography with Computer-aided Detection: Prospective Study of 12,860 Patients in a Community Breast Center. *Radiology*, 220:781–786, 2001.

[4] R. Guggenberger, P. Eppenberger, D. Markovic, D. Nanz, A. Chhabra, K. P. Pruessmann, and G Andreisek. MR neurography of the median nerve at 3.0T: Optimization of diffusion tensor imaging and fiber tractography. *European Journal of Radiology*, 81(7):e775–782, 2012.

[5] G. Hermosillo, C. Chefdhotel, K. H. Herrmann, G. Bousquet, L. Bogoni, K. Chaudhuri, D. R. Fischer, C. Geppert, R. Janka, and A. Krishnan. *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*, chapter Image Registration in Medical Imaging: Applications, Methods, and Clinical Evaluation.

[6] M. F. Kircher, H. Hricak, and S. M. Larson. Molecular imaging for personalized cancer care. *Molecular Oncology*, 6(2):182–195, 2012.

[7] B. Krishnapuram, J. Stoeckel, V. C. Raykar, R. B. Rao, P. Bamberger, E. Ratner, N. Merlet, I. Stainvas, M. Abramov, and A. Manevitch. Multiple instance learning improves CAD detection of masses in digital mammography. In *Proceedings of the 9th international workshop on Digital Mammography*, pages 350–357. 2008.

[8] H. B. W. Larsson, T. Y. Lee, N. A. Mayr, G. J. M. Parker, R. E. Port, J. Taylor, and R. M. Weisshoff. Estimating kinetic parameters from dynamic contrast-enhanced t1-weighted mri of diffusable tracer: a common global language for standardized quantities and symbols. *Journal of Magnetic Resonance Imaging*, 19:223–232, 1999.

[9] L. Lu, J. Bi, S. Yu, Z. Peng, A. Krishnan, and X. S. Zhou. Hierarchical learning for tubular structure parsing in medical imaging: A study on coronary arteries using 3D CT Angiography. In *IEEE 12th International Conference on Computer Vision*. 2009.

[10] J. Patterson II, A. Minagar, N. Natarajan, and A. Takalkar. Alzheimer's disease detection with objective statistical evaluation of FDG-PET brain scans: essential methodology for early identification. *Future Neurology*, 5(2):259–276, 2010.

[11] Z. Peng, Y. Zhan, X. S. Zhou, and A. Krishnan. Robust anatomy detection from CT topograms. In *Proceedings of SPIE*. 2009.

[12] V. C. Raykar, B. Krishnapuram, J. Bi, M. Dundar, and R. B. Rao. Bayesian multiple instance learning: Automatic feature selection and inductive transfer. In *Proceedings of the 25th International Conference on Machine Learning*, pages 808–815. 2008.

[13] V. C. Raykar, S. Yu, L. H. Zhao, G. H. Valadez, C. Florin, L. Bogoni, and L. Moy. Learning from crowds. *Journal of Machine Learning Research*, 11:1297–1322, April 2010.

[14] G. D. Rubin. Data explosion: the challenge of multidetector-row CT. *European Journal of Radiology*, 36(2):74–80, 2000.

[15] G. K. Thawait, A. Chhabra, and J. A. Carrino. Spine segmentation and enumeration and normal variants. *Radiol Clin North Am*, 105:587–98, 2012.

[16] Y. Zhan, M. Dewan, M. Harder, A. Krishnan, and X. S. Zhou. Robust Automatic Knee MR Slice Positioning Through Redundant and Hierarchical Anatomy Detection. *IEEE Transactions on Medical Imaging*, 30(12):2087–2100, 2011.

[17] Y. Zhan, M. Dewan, and X. S. Zhou. Cross Modality Deformable Segmentation Using Hierarchical Clustering and Learning. In *Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2009.

[18] Y. Zhan, M. Dewan, and X. S. Zhou. Robust MR spine detection using hierarchical learning and local articulated model. In *International Conference on Medical Image Computing and Computer Assisted Intervention (to appear)*. 2012.

[19] Y. Zhan, X. S. Zhou, Z. Peng, and A. Krishnan. Active scheduling of organ detection and segmentation in whole-body medical images. In *Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2008.

[20] S. Zhang, Y. Zhan, M. Dewan, K. Huang, D. N. Metaxas, and X. S. Zhou. Towards robust and effective shape modeling: Sparse shape composition. *Medical Image Analysis*, 16(1):265–277, 2012.

[21] X. S. Zhou, A. Gupta, and D. Comaniciu. An information fusion framework for robust shape tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):115–129, 2005.