Large-scale Retrieval for Medical Image Analytics: A Comprehensive Review

Zhongyu Li[a], Xiaofan Zhang[a], Henning Müller[b], Shaoting Zhang[a,*]

[a] *Department of Computer Science, University of North Carolina at Charlotte, USA*
[b] *University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland*

Abstract

Over the past decades, medical image analytics was greatly facilitated by the explosion of digital imaging techniques, where huge amounts of medical images were produced with ever-increasing quality and diversity. However, conventional methods for analyzing medical images have achieved limited success, as they are not capable to tackle the huge amount of image data. In this paper, we review state-of-the-art approaches for large-scale medical image analysis, which are mainly based on recent advances in computer vision, machine learning and information retrieval. Specifically, we first present the general pipeline of large-scale retrieval, summarize the challenges/opportunities of medical image analytics on a large-scale. Then, we provide a comprehensive review of algorithms and techniques relevant to major processes in the pipeline, including feature representation, feature indexing, searching, etc. On the basis of existing work, we introduce the evaluation protocols and multiple applications of large-scale medical image retrieval, with a variety of exploratory and diagnostic scenarios. Finally, we discuss future directions of large-scale retrieval, which can further improve the performance of medical image analysis.

Keywords: Medical image analysis, information retrieval, large scale, computer aided diagnosis

*Corresponding author, rutgers.shaoting@gmail.com

1. Introduction

Medical image analytics plays a central role in clinical diagnosis, image-guided surgery and pattern discovery. Many protocols and modalities of digital imaging techniques have been adopted to generate medical images, including magnetic resonance imaging (MRI) (Slichter, 2013), computed tomography (CT) (Hsieh, 2009), photon emission tomography (PET) (Bailey et al., 2005), ultrasound (Szabo, 2004), fluorescence microscopy (Lichtman and Conchello, 2005), X-ray (Lewis, 2004) and others. Generally, these medical images reflect specific aspects (anatomy, function) of tissue types/organs that require an accurate interpretation and analysis from either domain experts or computer-aided decision support. In comparison with domain expert analysis that is labor intensive and time-consuming, computer-aided approaches are efficient and its accuracy has increased continuously with the rapid development of computer vision, machine learning and related fields (Doi, 2014; Katouzian et al., 2012; May, 2010). To support computer-aided medical image analytics, one important task is content-based image retrieval (CBIR) (Akgül et al., 2011; Lehmann et al., 2004; Müller et al., 2004), i.e., indexing and mining images that contain a similar visual content (e.g., shape, morphology, structure, etc). For a new medical image to be analyzed, a CBIR system can first retrieve visually similar images in an existing dataset. Then, its high-level descriptions and interpretations can be explored based on the retrieved images.

Over the past 25 years, CBIR has been one of the most vivid research topics in the field of computer vision. Many CBIR methods were developed for accurate and efficient image retrieval. Especially in recent years, with the ever-increasing number of digital images (e.g., ImageNet (Russakovsky et al., 2015), COCO (Lin et al., 2014), PASCAL VOC (Everingham et al., 2010), etc), CBIR has moved towards the era of big data. Massive amounts of images can provide rich information for comparison and analysis, and thus facilitate the generation of new algorithms and techniques that can tackle image retrieval in large databases. In general, large-scale image retrieval can be divided into two stages, i.e., feature extraction to represent images and feature indexing. Deep learning (LeCun et al., 2015) is one of the most popular methods for feature representation that is particularly suitable for large image databases, where massive amounts of data can boost the retrieval performance by training deep and complex neural networks with millions of parameters (Babenko and Lempitsky, 2015; Wan et al., 2014). For the feature indexing at a large-scale, the key problem is computational efficiency, i.e., similarity searching in millions of images with thousand dimensional features vectors. Methods such as vocabulary trees (Nister and Stewenius, 2006) and hashing (Wang et al., 2016) can efficiently tackle this problem, either through changing the indexing structure or compressing the original features.

Despite the current large-scale methods having achieved many successes in generic image retrieval problems, how to best tackle the retrieval in large-scale medical image databases is still a very challenging topic (Zhang and Metaxas, 2016). On the one hand, the meaning of large-scale in the medical image field is somewhat different from large-scale in the generic image domain. Generally, each patient can generate hundreds to thousands of image slices using different protocols, modalities (e.g., CT, MRI, X-ray) and multiple dimensions (e.g., volumetric 3D, time series). These volumes are usually stored in many single images (as slices) in the DICOM (Digitla Imaging and Communications in Medicine) format (Kahn et al., 2007). Besides this, the size of some medical images can be extremely large. For example, the whole-slide histopathological images can include more than $100,000 \times 100,000$ pixels and thus each is usually split into millions of small patches for processing. On the other hand, medical images are usually more difficult to analyze compared to generic images. The complex imaging parameters (contrast agents, machine settings), anatomic difference and interactions between different diseases result in a more complex analysis compared with natural images, where broad object categories are recognized and used for similarity calculations. The relevant changes of some medical images can be very subtle, which require more fine-grained and detailed analysis. Therefore, directly employing traditional CBIR methods may not suitable for the large-scale medical image retrieval problem. In recent years, many efforts have been made to achieve large-scale medical image analytics, aiming to improve the efficiency and accuracy of image retrieval.

1.1. Related Work

There have been multiple reviews focusing on content-based medical image retrieval. The first review in the field was (Tang et al., 1999) but the text only contained few systems with a limited scope. Muller et al. (Müller et al., 2004) presented a first complete review that concentrates on image retrieval in the medical domain, where the techniques used in medical image retrieval, including visual feature extraction, image comparison, system evaluation, etc. are summarized. Subsequently, Long et al. (Long et al.,

2009) introduced four medical CBIR systems, i.e., CervigramFinder (Xue et al., 2008), SPIRS (Hsu et al., 2007), IRMA (IRMA), SPIRS-IRMA (Antani et al., 2007). The authors also discussed future directions of medical image retrieval. Akgul et al. (Akgül et al., 2011) presented a comprehensive review about recent techniques of content-based image retrieval in radiology until 2011, including image features/descriptors, similarity measures and state-of-the-art systems. Additionally, they discussed challenges and future directions for the coming decade. Hwang et al. (Hwang et al., 2012) reviewed both text-based and content-based medical image retrieval systems, drawing a conclusion that the image retrieval service will be more effective if CBIR and semantic systems are combined. In 2013, Kumar et al. (Kumar et al., 2013) surveyed several applications and approaches to medical CBIR that focus on clinical imaging data that are multidimensional or acquired using multiple modalities such as combined PET-CT images.

Besides the abovementioned survey articles, the image retrieval task of the Conference and Labs of the Evaluation Forum, named ImageCLEF (ImageCLEF; Müller et al., 2010), has held several medical image retrieval tasks from 2004-2014. ImageCLEF provides a platform for research groups submitting results and competing on the performance of their medical image retrieval methods. After each ImageCLEF medical image retrieval task, an overview is provided to summarize the methods and results of each competition groups (de Herrera et al., 2013; Kalpathy-Cramer et al., 2015, 2011; Müller et al., 2012), which demonstrates the state-of-the-art in the medical image retrieval field. A benchmark for case-based retrieval including full volumetric images of more than 300 patients was run as part of the VISCERAL benchmark Jimenez-del-Toro et al. (2015).

1.2. Contributions and Organization of this Article

This survey provides a structured and extensive overview of large-scale retrieval for medical image analytics. Despite existing reviews having summarized varieties of medical retrieval systems and methods, none of them focused on the retrieval techniques for large-scale medical data, which is currently the main challenge in the field of medical analytics. This survey offers a focused overview of the retrieval approaches for the large-scale medical image data by expanding multidisciplinary components that involve a nexus of the idea from machine learning, computer vision, information retrieval, and bioinformatics. It explains the entire process from scratch and presents a comprehensive pipeline that discusses every processing step from feature

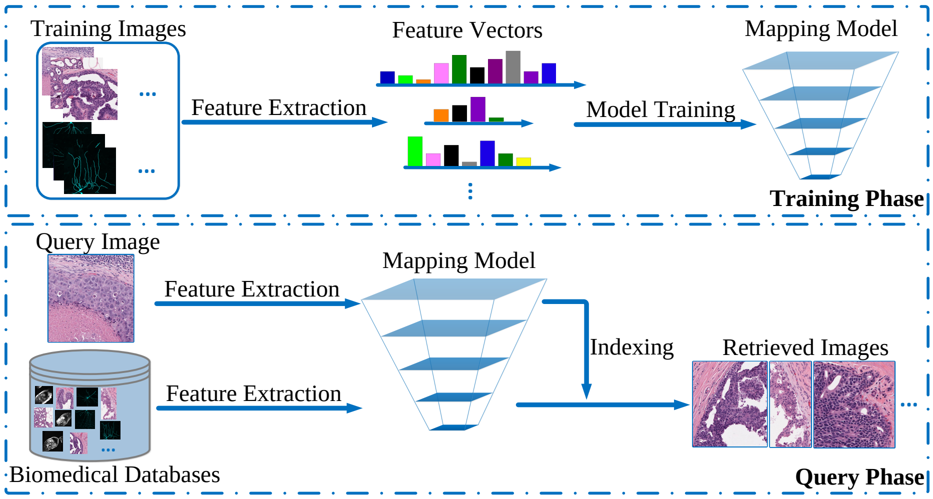

Figure 1: A general pipeline of large-scale medical image retrieval.

extraction to knowledge discovery and decision support. Fig. 1 illustrates a general pipeline of large-scale medical image retrieval. Given a set of medical images (e.g., MRI, CT, microscopy, etc.), feature extraction methods are employed to represent each image. Unlike traditional medical retrieval methods that directly compare the image similarity via original feature vectors, large-scale approaches often first train a retrieval model, e.g., organizing and transforming image features that can improve the performance of feature indexing. In the query phase, the query image is compared only to similar images based on the well-designed retrieval model rather than an exhaustive search of the whole database. The retrieval results can be provided to users for further analysis. According to Fig. 1, retrieval with large-scale medical image databases is different compared with classical CBIR systems. In recent years, many researchers in the medical domain have moved their attention to the analytical questions of large-scale image analysis (Zhang and Metaxas, 2016). Therefore, in this era of big data, it is necessary to present a comprehensive review of recent advances in large-scale medical image analytics.

In this paper, we organize the survey into five parts: challenges/opportunities, methodology review, evaluation protocols, applications, and future directions. In Section 2, challenges and opportunities related to big data in medical image analytics are provided. Section 3 and 4 discuss the methodology details relevant to the large-scale medical image retrieval, which mainly includes two parts, i.e., feature representation, feature indexing and search. Following Section 5 introcudes evaluation protocols in medical image retrieval. Based on the existing approaches, Section 6 reviews several applications of large-scale medical image retrieval. Finally, Section 7 explores potential directions for future work on large-scale medical retrieval.

2. Challenges and Opportunities

The challenges of large-scale medical image retrieval can be summarized as a good trade-off between efficiency and accuracy. Despite traditional methods having already achieved good performance in many very specific medical scenarios, keeping efficiency and accuracy in large-scale approaches still faces many problems. Additionally, in the era of big data, large-scale medical image analysis provides many opportunities for both academia and industry.

2.1. Challenges

One major concern in the big data era is system efficiency. Given the huge amount of medical image data (WPS, 2010), how to represent and search in an efficient way still has many challenges. Counting the data from The Cancer Imaging Archive (TCIA), a large-scale medical image repository, Fig. 2 illustrates the number of images with the six most common anatomical sites. According to Fig. 2, these data sets have hundreds of thousands to millions of medical images, which are hard to analyze in real-time. For medical image retrieval, each image is usually represented by a feature vector with often thousands of dimensions. An exhaustive search of millions of images with large feature vectors is very time-consuming (Zhang et al., 2015c). In clinical applications, for a single patient tens to hundreds/thousands of images are collected

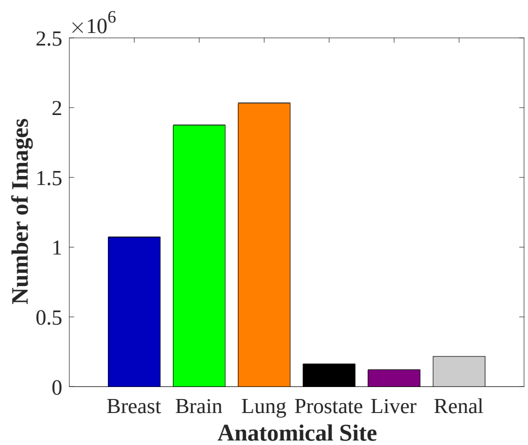

Figure 2: Number of medical images for the six most common anatomical area in the TCIA (The Cancer Imaging Archive) repository.

(large MRI studies can easily contains tens of thousand of single image slices for a single patient) and an efficient retrieval of these images is required for computer-assisted diagnosis. Accordingly, to achieve medical retrieval with massive amounts of images, two aspects need to be explored for improvement , i.e., 1) reducing the dimension of the feature vectors (or creating very sparse spaces), 2) improving the strategy of similarity search or data indexing. Both challenges are hard to tackle using conventional methods.

Another concern of medical retrieval is system accuracy. In the information retrieval field, precision is one of the most important criteria for performance evaluation, which is defined as the fraction of retrieved images that are relevant to the query image (Powers, 2011). For a query image, higher retrieval precision indicates more reliable analysis and exploration results, since most of the retrieved images share (hopefully) similar semantic content with the query image. Retrieval precision plays a critical role in medical analytics, where clinical diagnoses can depend on decision support that is based on the retrieved images. However, achieving high precision in medical retrieval is not an easy task, especially with the large amount of volumetric image data, where most parts of the images/volumes are not important for similarity calculations but small, local anomalies are. Fig. 3 illustrates a common problem in the classification of histopathological images, which are obtained from intraductal breast lesions in this case. The two images to the left (with the blue bounding box) belong to the same category, i.e., both are actionable (indicating the cells/tumors are pathogenic). However, they have quite different expressions. On the other side, for the bottom two images (with the red bounding box), despite the visual similarity, they belong to different categories (the right image is benign, indicating the cell's/tumor's lack of the ability to invade neighboring tissue and create metastasis). This problem can be summarized as large intra-class variation and small

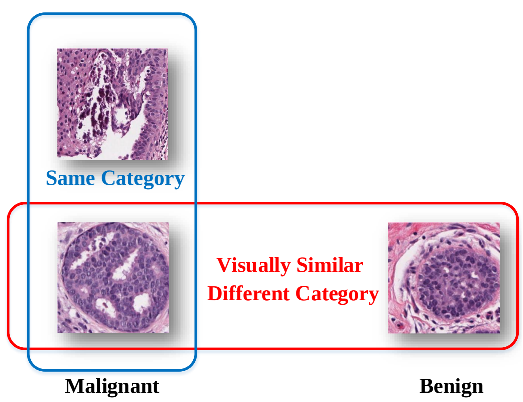

Figure 3: Three histopathology images of intraductal breast lesions. Classifying the breast histopathology images into benign or malignant is a challenge due to their large intra-class variation and small inter-class variation.

inter-class variation (Zhang et al., 2016b). Not only in histopathological image analysis, most medical image analytics tasks encounter similar problems. More critically, when dealing with massive medical data, this problem becomes more challenging since more noisy images are included and influence the retrieval performance.

In addition to measuring efficiency and accuracy, the detailed evaluation protocol is also a challenging question in large-scale medical image retrieval. Most of the traditional methods simply use class labels to evaluate the retrieval performance, which is not suitable for large-scale medical image databases, as they are most often not fully labeled and there can be different relevance expectations depending on the query images. Besides this, the data storage, access, organization, and computing techniques may also influence the retrieval performance of large-scale medical images. In this article, we review relevant methods and techniques that can tackle large-scale medical image retrieval.

2.2. Opportunities

Leaving aside the above challenges, large-scale image data brings unprecedented opportunities to the medical field. In 2014, Siemens released a report saying that the market for medical imaging systems will grow from 32.3 billion in 2014 to 49 billion in 2020 (Siemens). Without doubt, in the era of big data the development of large-scale medical analytics will accelerate this process. In a medical retrieval system, massive image data generally provides more samples for similarity search, which can improve the accuracy and reliability of the system (Fang et al., 2016). More importantly, it also facilitates the research of knowledge discovery and pattern exploration in biomedical informatics. We illustrate two major opportunities that benefit from large-scale medical retrieval, i.e., computer-aided diagnosis and visual pattern exploration:

1. computer-aided diagnostics (CAD): CBIR methods have been proposed as an effective technology for CAD systems, which have the capacity of relieving the workload of doctors and to offer more reliable and consist analysis of medical images (Akgül et al., 2011; Depeursinge et al., 2011). Despite most retrieval systems are not routinely used, CBIR based CAD are rather research prototypes for medical image analytics. Given an image database with diagnosis information, CBIR methods aim to retrieve and visualize images with morphological profiles most

relevant to and consistent with the query image. This can provide decision support, for example for pathologists (Kumar et al., 2013; Müller et al., 2004). When the CBIR-based CAD systems meet large-scale image databases, this benefit is enlarged by searching more relevant images with fine-grained content and morphologies. Retrieval results from large-scale databases can help pathologists to have accurate and deep understanding of query images, when they are unsure about specific patterns.

2. visual pattern exploration: medical images contain a wealth of structures and patterns that may convey information about underlying mechanisms in biology (Peng et al., 2010; Schindelin et al., 2012). Generally, individuals with similar structures, shapes, morphologies will also express similar functions and properties, such as neurons, tissue cells, etc. (Li et al., 2017a; Xing and Yang, 2016). By establishing large medical databases of visual data, CBIR systems can be used to identify and explore unknown individuals based on the retrieval results. Massive image data are the basic requirement for such a medical exploration. As individuals usually have complex shapes and varieties in the images, large-scale databases can provide more reliable results for pattern exploration, as it is more likely that similar patients exist of which images were taken with similar protocols.

Large-scale image databases bring new opportunities to innovate the traditional medical retrieval systems, and some of the large-scale medical systems have already achieved good performance in clinical practice. In Section 6, we review relevant applications of large-scale medical retrieval.

3. Feature Representation

To achieve medical analytics from large-scale image databases, the first step is visual feature extraction, i.e., using feature vectors to represent each digital image. Generally, feature vectors are representing the low-level image content and can be linked to high-level perceptions of the images. A good feature representation is the prerequisite to achieve good performance in medical image retrieval. In recent years, a variety of feature representations have been developed based on computer vision and machine learning. This section reviews recent advances in feature vectors in medical images. Specifically, the feature representation is classified into two categories, i.e., hand-crafted and learned features. This is mainly based on whether the features are obtained through domain expert knowledge (model-driven) or a purely data-driven procedures.

3.1. Hand-crafted Features

Generally, hand-crafted features are sequentially extracted from each image according to algorithms based generally on expert knowledge (Antipov et al., 2015), where each feature models a specific information such as color, texture or shape. Before the strong use of deep learning, hand-crafted methods dominated the feature extraction field for several decades. Most current medical retrieval systems still employ hand-crafted methods for feature representation. In this subsection, we review typical hand-crafted features that have been used in medical image retrieval.

The most widely used hand-crafted features for image retrieval are based on the Scale-Invariant Feature Transform (SIFT) (Lowe, 2004). SIFT detects scale-invariant key points by finding local extrema in the difference-of-Gaussian (DoG) space. It describes each key point by a 128-dimensional gradient orientation histogram. Subsequently, all SIFT descriptors are modeled/quantized using a bag-of-words (BoW) (Sivic and Zisserman, 2003). The feature vector of each image is computed by counting the frequency of the generated visual words in the image. SIFT is a local texture feature that has achieved success in medical image retrieval (e.g., it was the most popular feature in the ImageCLEF medical image retrieval task (Müller et al., 2012)). Besides SIFT descriptors, many local descriptors can use the BoWs to generate local features for medical images, such as SURF (Speeded Up Robust Features) (Bay et al., 2008), LBP (Local Binary Patterns) (Ojala et al., 1996) and others. In contrast to features extracted locally, holistic features are also widely adopted in medical image retrieval. These kinds of features can directly represent the global information of the entire image. For example, GIST (Oliva and Torralba, 2001) is a holistic feature which is based on a low dimensional representation of the scene that does not require any form of segmentation, and it includes a set of perceptual dimensions (naturalness, openness, roughness, expansion, ruggedness) that represent the dominant spatial structure of a scene (Douze et al., 2009). GIST has been applied in many medical image retrieval problems (Kalpathy-Cramer and Hersh, 2008; Liu et al., 2014a). Other holistic features such as HOG (Histogram of Gaussians) (Dalal and Triggs, 2005), color histograms (Siggelkow, 2002) are also frequently used in medical image retrieval (Müller and Deserno, 2010; Yu

Method	Category	Application
SIFT (Lowe, 2004)	Local, texture	Breast cancer (Zhang et al., 2015c), Basal-cell carcinoma (Wang et al., 2011a), etc
SURF (Bay et al., 2008)	Local, texture	Lung CTs (Haas et al., 2011), Body portion (Feulner et al., 2011), etc
LBP (Ojala et al., 1996)	Local, texture	2D-HeLa (Nanni et al., 2010), Brain MR (Murala et al., 2012), etc
GIST (Oliva and Torralba, 2001)	Holistic, shape	Mammogram (Liu et al., 2014a), Breast-tissue (Jiang et al., 2016a), etc.
HOG (Dalal and Triggs, 2005)	Holistic, texture	Cortical (Unay and Ekin, 2011), Lung (Song et al., 2012), etc
Color Histogram (Siggelkow, 2002)	Holistic, color	Organ (Caicedo et al., 2007), Dermatology (Bunte et al., 2011), etc.
Moments (Stricker and Orengo, 1995)	Holostic, shape	Multi-modalities (Rahman et al., 2007), Liver CT (Gletsos et al., 2003), etc.
Gabor filters (Manjunath and Ma, 1996)	Local, texture	Multi-modalities (Lim and Chevallet, 2005), Prostate Histopathology (Doyle et al., 2007), etc.
Tamura (Tamura et al., 1978)	Local, texture	Mammogram (Zhou et al., 2012), Multi-modalities (Güld et al., 2005), etc.
3D Riesz (Chenouard and Unser, 2011)	Local, texture	Epileptogenic Lesion (del Toro et al., 2013), 3D Multi-modalities (Jiménez-del Toro et al., 2015), etc

Table 1: Commonly used hand-crafted features and their applications in medical image retrieval.

et al., 2013). Table 1 lists some of the most commonly used hand-crafted features and their corresponding applications in medical image retrieval.

In addition to the common features mentioned above that can be used for the retrieval of both natural and medical images, there are many other hand-crafted features that are designed specifically for medical image data. In histopathology image analysis, the shape and texture information play an important role in the representation of cell/nuclei. Basavanhally et al. (Basavanhally et al., 2010) designed three graph-based features, i.e., Voronoi diagram, Delaunay triangulation, and minimum spanning tree, to describe the arrangement of the lymphocytes. Filipczuk (Filipczuk et al., 2013) employed 25 kinds of features to represent cytological images, including the size of the nuclei, the texture features based on gray-level pixels, and the distribution of nuclei in the image. In general, these specific features are more discriminative than the general hand-crafted features. They achieved good performance in the detection, retrieval and analysis of cells and nuclei (Xing and Yang, 2016). Besides the histopathological images, specific features are also widely used for the representation of 3D medical image data, such as 3D brain tumors, neuronal morphology. For example, Cai et al. (Cai et al., 2010) developed PCM-based volumetric texture features for 3D neurological image retrieval, and Wan et al. (Wan et al., 2015) employed quantitative measurements and geometrical moments as features to represent the 3D neuron morphological data. Both achieved good performance in the retrieval task. A more general system that creates many quantitative measurements of the brain including shape features is FreeSurfer (Fischl, 2012).

In order to achieve better retrieval performance, many researchers employ multiple hand-crafted features and combine them to represent each image. For example, Song et al. (Song et al., 2012) employed HOG and LBP features for retrieval and to recognize lung lesions. In general, combining multiple features (e.g., local and holistic features, common and specific features) obtains better performance compared with single feature systems (Lisin et al., 2005; Zhang et al., 2016a). Many groups in the ImageCLEF medical retrieval tasks have adopted this strategy (Simpson et al., 2012). However, when dealing with massive amounts of medical images, the combined features are often too large for scalable retrieval and may adversely affect the retrieval efficiency. Although a variety of features has been discussed above, for the medical retrieval problem, there are no universal features that are suitable for all kinds of medical images. This is the case, as medical images are generated by different imaging techniques and tissues/organs usually have specific colors, textures and shapes. Even for the same tissue/organ, features may visually differ under multiple dimensions and modalities (Kumar et al., 2013). Therefore, employing suitable hand-crafted features for a given kind of image data is an important and challenging step during medical retrieval. Feature selection can also be a step to create a subset of the features for a specific task.

Despite hand-crafted features having achieved many good results in medical image retrieval, they have shortcomings when tackling large-scale medical data:

1. hand-crafted features need expert knowledge but expert knowledge usually does not work well when the dataset is large as there may be outliers and cases not covered by standardized rules;
2. feature extraction using hand-crafted methods is time-consuming and computationally expensive, especially when dealing with massive amounts of images;
3. many hand-crafted methods are only designed for specific medical data and can not be extended to other domains.

Accordingly, more automatic, efficient and extensible feature representation methods are required for the large-scale medical retrieval.

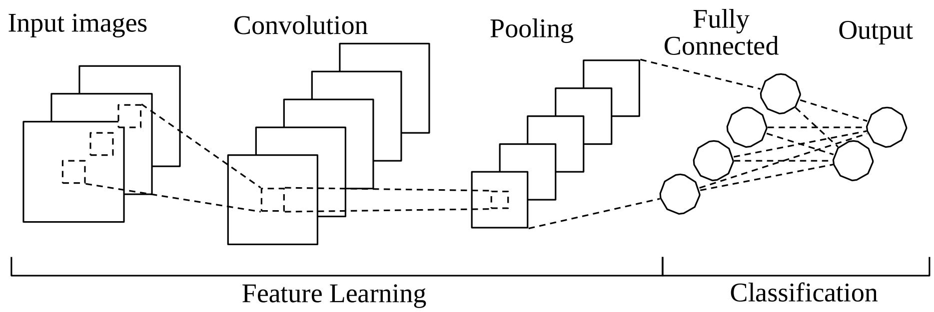

Figure 4: A general framework of convolutional neural networks.

3.2. Learned Features

In recent years, deep learning has become a hot topic and has achieved very good results in feature representation, image classification, retrieval, detection and other related fields. Compared with hand-crafted methods using domain expert knowledge, deep learning requires only a set of training data that allows to discover the feature representations in a self-taught manner (Bengio, 2009; LeCun et al., 2015). For the learned feature representation, a variety of deep neural networks are designed nonlinearly and hierarchically, i.e., mapping features from fine to abstract with multiple layers of neural networks (e.g., tens to hundreds) and a large number of parameters (e.g., thousands to millions) (Shen et al., 2016). In general, the prevalence of deep learning mainly benefits from the availability of large training data sets that make it possible to optimize the parameters. Accordingly, due to the availability of current large-scale medical image databases, deep learning can also be adopted to solve analytics tasks of medical images. Specifically, both supervised and unsupervised deep neural networks have been explored for creating feature representations of medical images.

Fig. 4 illustrates a general framework of a supervised deep neural network, i.e., a Convolutional Neural Network (CNN) (LeCun et al., 1998). The input images with fixed size are convolved with multiple learned kernels using shared weights. Then, the pooling layers down-sample the input representation nonlinearly and preserve the feature information in each subregion. Afterwards, the extracted features are weighted and combined in the fully-connected layer, and these features are sent to a pre-defined classifier for prediction. Finally, by comparing the output class with the image label, the CNN parameters (e.g., kernels, weights, bias) are updated in each iteration. Recent results, as on the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2015) have shown the excellent performance of very deep neural networks, where more convolution, pooling and fully connected layers are employed than before, and more complicated network structures are developed, e.g., AlexNet (Krizhevsky et al., 2012), GoogLeNet (Szegedy et al., 2015), VGG Net (Simonyan and Zisserman, 2014) and ResNets (He et al., 2015).

Supervised deep neural networks require a large amount of labeled images to train the parameters in each layer. However, in the medical field, the amount of labeled images is typically limited. Simply training deep neural networks from scratch using small-sized labeled data can easily result in overfitting (Srivastava et al., 2014). Thus, researchers have proposed several methods to accommodate medical image analysis with deep neural networks. For example, Bar et al. (Bar et al., 2015) learned features for chest pathology detection using a Decaf pre-trained CNN model (Donahue et al., 2014), and the parameters are trained from non-medical datasets such as ImageNet (Deng et al., 2009). In ImageCLEFmed 2016, NovaSearch adopted CNN models that are trained from scratch using only the provided medical data (Semedo and Magalhães). They employed several techniques (e.g., Dropout (Srivastava et al., 2014), data augmentation) to deal with the unbalanced and small data sets. According to (Shin et al., 2016), there are three major techniques that can successfully learn feature representation of medical images through CNNs:

1. pre-training the CNN model on natural images and fine-tuning on medical target images; this technique has been used for lung images (Hofmanninger and Langs, 2015; Li et al., 2014a; Schlegl et al., 2014), brain MRI (Li et al., 2014b), etc.;
2. training the CNN model from scratch using only medical images, and employing several measures to avoid overfitting; this technique has been used in cardiac CT (Wolterink et al., 2015), on lung nodules (Shen et al., 2015d), etc.;
3. using a pre-trained CNN model to extract features, employing these features as complementary information and combining them with hand-crafted features; these combined features have been used on chest X-rays (Bar et al., 2015), pulmonary peri-fissural nodules (Ciompi et al., 2015), etc.

Although supervised deep neural networks have demonstrated excellent performance in feature representation, they require a large amount of manually labeled training data. However, unlike the annotation of natural images that is easy to achieve, the labels of many medical images can only be annotated by physicians or domain experts, which is expensive. In many cases, the ground truth labels are simply unavailable, as the exact patterns of some abnormalities are still unidentified or very subjective in nature (e.g., neuron images, precise tumor regions). To overcome the limitations of supervised feature learning, multiple unsupervised deep neural networks have been proposed for feature representation (Bengio et al., 2012). Fig. 5 illustrates a typical unsupervised neural network, i.e. an Auto-Encoder (Bourlard and Kamp, 1988). Given the input images $\mathbf{X}_m$, it learns the feature representations $h^{(2)}$ by minimizing the reconstruction error between the input and the output, i.e., $\mathbf{Y}_m \approx \mathbf{X}_m$, which indicates the decoder results should approximate the input. Despite the single layer auto-encoder being too shallow to learn features, the representation power improves significantly when several auto-encoders are stacked to form deep stacked auto-encoders (SAEs). For example, Wu et al. (Wu et al., 2013, 2016) developed an unsupervised feature selection method using a convolutional stacked auto-encoder to identify intrinsic deep feature representations in image patches. The method is demonstrated on 7.0-tesla brain MR images, validating that unsupervised feature learning is effective for brain MR registration. Besides this, Shin (Shin et al., 2013) employed stacked auto-encoders for unsupervised feature learning and organ identification in magnetic resonance images, where visual and temporal hierarchical features are learned to categorize object classes from an unlabeled multimodal DCE-MRI data set (Collins and Padhani, 2004).

In addition to auto-encoders, restricted Boltzmann machines (RBM) (Smolensky, 1986) can also construct unsupervised deep neural networks, e.g. deep belief networks (Hinton and Salakhutdinov, 2006) and deep Boltzmann machines (Salakhutdinov, 2015). These deep neural networks are also the common choice to tackle medical feature representations and other medical analytics tasks. For example, Brosch and Tam (Brosch et al., 2013) performed

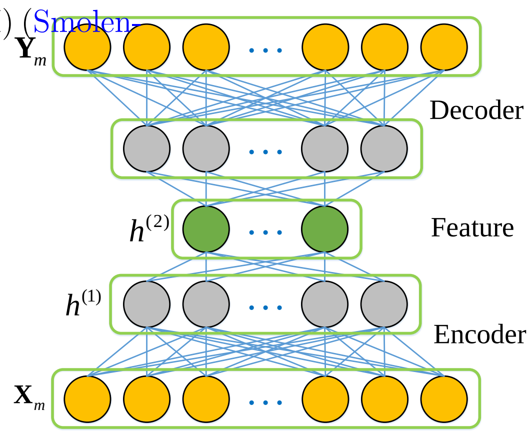

Figure 5: The hierarchical structure of an auto-encoder.

manifold learning by reducing the dimensionality of brain images using a deep belief network that can discover patterns of similarity in groups of images. Cao et al. (Cao et al., 2014) developed a multimodal approach for medical image retrieval that is based on deep Boltzmann machines. Experimental results demonstrate that the new deep Boltzmann machine-based multimodal learning model is a promising solution for next-generation medical image indexing and retrieval systems.

For large-scale medical image analytics, learned feature representations are a clear trend, since more and more images are available to train the deep neural networks. However, the usage of deep learning for medical image retrieval is not frequent. One reason is that previously most medical image retrieval tasks only had to tackle small-sized data sets (e.g., hundreds to thousands of images at the most), which does not allow the training of deep neural networks. The other reason is that for some specific medical images the hand-crafted features designed by domain experts can achieve very good performance when the data sets are not too large (e.g., the holistic features of histopathological images (Basavanhally et al., 2010)). Due to the multi-modality, complexity (e.g., diverse medical imaging techniques, complex structures and morphology of tissues/organs) and also quickly changing image acquisition devices, the specified hand-crafted features are still useful in many medical image retrieval scenarios. Additionally, the deep-learning based methods are capable to learn different types of features compared with hand-crafted methods. Thus the learned features also play a critical role in the feature representation of medical images, particularly when the data sets are large. In the ImageCLEF Challenges (García Seco de Herrera et al., 2016), many groups employed both learned features and hand-crafted features to represent medical images. Then, these features are fused for more accurate retrieval and classification results.

4. Feature Indexing and Search

After feature extraction, each image is represented by a feature vector. The medical image retrieval problem can now be treated as a nearest-neighbor search among these feature vectors, i.e., computing and ranking the distance between the query image(s) or volume(s) and all images in the databases. However, when handling large-scale databases, exhaustive search among long feature vectors is time-consuming. Sequentially computing the

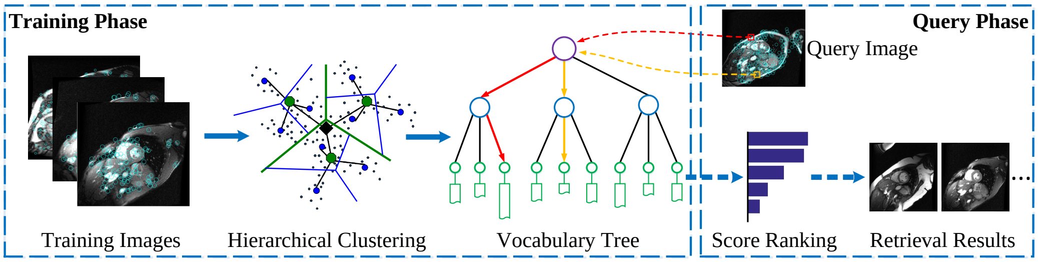

Figure 6: A framework for vocabulary tree based image retrieval.

distance of millions of high-dimensional feature vectors is unfeasible. In this section, we review recent advances that can efficiently and accurately tackle feature indexing in large-scale medical retrieval.

4.1. Vocabulary Tree

The vocabulary tree was first proposed by Nistér and Stewénius (Nister and Stewenius, 2006). It is widely used for scalable image retrieval (Wang et al., 2011b; Zhang et al., 2015b). It builds a tree-structure to accelerate similarity indexing. Compared with traditional methods based on exhaustive search of image features, vocabulary tree based methods employ a hierarchical tree and inverted files that can significantly improve the retrieval efficiency. Fig. 6 presents the framework of vocabulary tree based image retrieval. The framework can be divided into two phases, i.e., the training phase (offline) and the query phase (online). The training phase builds the indexing model (hierarchical tree-structure) from given image sets and the query phase returns images that are similar to the query image.

Training Phase: For a set of training data, vocabulary tree methods first detect key points in each image (denoted as the cyan circles in Fig. 6). The key points can be defined as corners with scale and rotation invariance, as well as interest points specified by domain experts. Subsequently, these key points are represented by local feature vectors (e.g., SIFT (Lowe, 2004)), and the descriptors from all training images are collected for hierarchical k-means clustering. Specifically, instead of defining k as the final number of clusters, k is defined as the number of children centers in each cluster. After L recursive clustering, a tree-structure of depth L and branch factor k is built, where each tree node (also referred to as the visual word) corresponds to a cluster center. Each leaf node includes several key points that are close to each other visually. Accordingly, all images in the database are added to inverted files attached to the leaf nodes with respect to their corresponding key points. Afterwards, the vocabulary tree-structure and the inverted file are used for the indexing of the images.

Query Phase: Given a query image q, its key points are extracted and set as the input in the vocabulary tree. By comparing with nodes in each hierarchy, each key point can reach a leaf node attached to an inverted file. As each inverted file records images relevant to the leaf node, the similarity scores can be computed between q and the images in corresponding inverted files. Normally, the term frequency-inverse document frequency (TF-IDF) (Salton and Buckley, 1988) is adopted as the similarity score to balance the importance of a visual word to an image in a collection. By ranking all the similarity scores in descending order, the top ranked images can be considered as the retrieval results. Unlike previous methods simply comparing the similarity of all the key points between two images, vocabulary tree methods construct the hierarchical tree-structure and index similar images using the inverted files. For each key point vector, only a total of kL dot products are needed, which is very efficient if k is not large. More importantly, the inverted file strategy can significantly improve the indexing process since it does not need to traverse the whole image database.

Vocabulary trees and its variants have been applied for large-scale medical image retrieval. They do not only improve the computational efficiency but are also often more accurate compared with traditional retrieval methods. For example, Jiang et al. (Jiang et al., 2015a,c) proposed an adaptive weighting strategy in the vocabulary tree based framework to tackle mammogram image retrieval. As the features with high frequencies in a mammogram are less informative than those with low frequencies, to avoid overcounting, they incorporate mammogram-specific node frequencies into the IDF scheme to down-weight the high-frequency features. The adaptive weighting technique is very effective to retrieve these specific images, i.e., mammographic masses. Wang et al. (Wang et al., 2015) designed a discriminative and generative vocabulary tree for the authentication and recognition of finger vein images. This method considers both the discriminative appearance of local image patches and their generative spatial layout. The training process remains the same as building a conventional vocabulary tree, while the prediction process uses a proposed point set matching method to support non-parametric patch layout matching. This joint discriminative and generative model can achieve good performance in finger vein images, since the employed vocabulary tree model can retain the efficiency for the whole system. More importantly, the point set matching strategy considers the geometrical layout of local image patches, which is more accurate compared with previous vocabulary tree based methods that only consider the description of local key points.

By changing the similarity indexing strategy, vocabulary tree based methods have achieved efficient retrieval in large-scale databases. As these kinds of methods directly employ local feature descriptors instead of the global features, it can be applied to most medical images, including both 2D and 3D images where local key points can be detected and described. However, vocabulary tree based methods also have several limitations. For example, simply using local features is not enough to represent and discriminate some specific medical types of images, e.g., for some lung images, the global shape should be considered during retrieval. In addition, the training phase in building the hierarchical vocabulary tree is usually time-consuming, especially when tackling very large image databases (search on a database with millions of images). In practical applications, to achieve good results, vocabulary tree based methods also rely heavily on parameter tuning, i.e., the number of each cluster center k, total levels of the hierarchical tree L. Thus, more efficient and accurate methods need to be developed for large-scale medical image retrieval.

4.2. Hashing

In recent years, hashing methods have been intensively investigated in the machine learning and computer vision fields for indexing big data (Wang et al., 2016). Instead of directly searching nearest neighbors from an original data set, hashing methods first compress the original data into short binary codes (e.g., tens to hundreds of bits) based on the defined hashing functions. Then, the nearest-neighbor search is more efficient by computing the similarity distances in binary Hamming space rather than in the high-dimensional feature space.

4.2.1. Hashing Frameworks

Fig. 7 presents the framework of hashing-based image retrieval. Assuming we have n medical images in the database, after feature representation these n images are represented by d dimensional feature vectors, i.e., $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\} \subset \mathbb{R}^{d\times n}$ (denoted as the blue points in Fig. 7). For the image $\mathbf{x}_i \subset \mathbb{R}^{d\times 1}$, its feature space can be split by a set of hashing functions $H = \{h_1, h_2, \ldots h_K\} \subset \mathbb{R}^{d\times K}$, and each hashing function encodes $\mathbf{x}_i$ into one

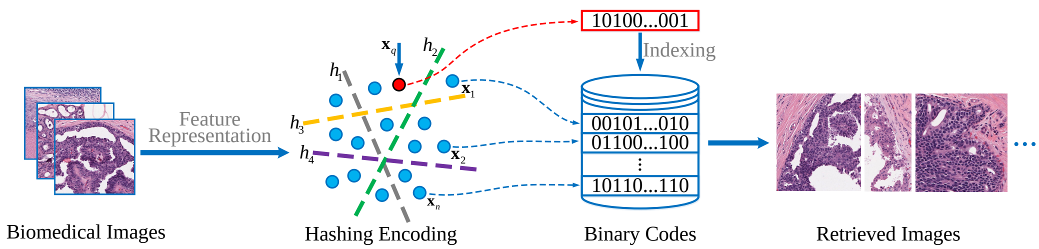

Figure 7: The framework of hashing-based image retrieval.

bit of binary code $h_k(\mathbf{x}_i)$. Therefore, the corresponding K bits of binary code of $\mathbf{x}_i$ can be denoted as:

$$\mathbf{y}_i = H(\mathbf{x}_i) = \{h_1(\mathbf{x}_i), h_2(\mathbf{x}_i), \ldots, h_K(\mathbf{x}_i)\} \tag{1}$$

In practice, for computational convenience, the above hashing functions are usually substituted by the projected matrix $\mathbf{w} \subset \mathbb{R}^{d\times K}$ and the intercept vector $\mathbf{b} \subset \mathbb{R}^{K\times 1}$:

$$\mathbf{y}_i = \mathrm{sgn}\left(f(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + \mathbf{b})\right) \tag{2}$$

where $f(\cdot)$ is a pre-specified function that can be linear or nonlinear. Then, all images in the database are represented by the mapped binary codes. The query image $\mathbf{x}_q$ (denoted as the red point in Fig. 7) can also be mapped into binary codes through Eq. 2. Subsequently, the similarity search between the query and each image in the database is transformed as the Hamming distance ranking of their corresponding binary codes, which is very fast. The key question of hashing methods is how to obtain good hashing functions that can not only split the feature space via binary encoding but also keep similarities and diversity among the original data.

4.2.2. Categories of Hashing Methods

The methods to compute hashing functions can be roughly divided into two categories, i.e., data-independent and data-dependent. Data-independent methods usually design generalized hashing functions that can compact any given data set into binary codes. Locality-Sensitive Hashing (LSH) and its variants are the most popular data-independent methods (Gionis et al., 1999; Kulis et al., 2009; Raginsky and Lazebnik, 2009). LSH-based methods compute hashing functions via maximizing the probability of collision for similar items, which can keep the originally nearby data points mapping into the same bit with high probability. However, this type of method often needs long binary codes and many hashing functions to ensure the desired retrieval precision, which dramatically increases the storage costs and the query time. More importantly, as these hashing functions are designed independently from the training data sets, it is hard to ensure the retrieval performance for any given data set.

Another category are the data-dependent methods (also called learning to hash methods) that learn the hashing functions from a given training data set. In general, compared with data-independent methods, data-dependent methods can achieve comparable or even better retrieval accuracy with shorter binary codes. Currently, many learning-based hashing methods have been applied for large-scale medical image retrieval, including but not limited to, Iterative Quantization (ITQ) (Gong et al., 2013), Kernel-Based Supervised Hashing (KSH) (Liu et al., 2012), Anchor Graph Hashing (AGH) (Liu et al., 2011), Asymmetric Inner-product Binary Coding (AIBC) (Shen et al., 2015a) and others. Accordingly, the taxonomy of data-dependent hashing methods can be defined in multiple viewpoints. For example, based on whether the training data sets have labels or not, hashing methods can be divided into supervised, unsupervised and semi-supervised methods. Supervised methods employ advanced machine learning techniques such as kernel learning, metric learning, and deep learning to compute the hashing functions from labeled training data. Many supervised hashing methods have achieved good performance since they can shorten the semantic gap between the compacted binary codes and the image labels (Fan, 2013; Gordo et al., 2014; Norouzi et al., 2012; Shen et al., 2015b). Without label information, unsupervised methods explore the properties of training data sets such as distributions and manifold structures to design effective hashing functions. Representative methods include spectral hashing (Weiss et al., 2009), graph hashing (Liu et al., 2014b), manifold hashing (Shen et al., 2013), etc. Additionally, semi-supervised methods design hashing functions using both labeled and unlabeled data. These kinds of methods can improve the binary encoding performance by leveraging semantic similarity with limited image labels while remaining robust to overfitting (Jain et al., 2009, 2008; Wang et al., 2012). Another taxonomy of data-dependent methods is based on the form of the hashing functions, i.e., linear and nonlinear. Linear hashing functions separate and map the original feature space with simple projections (as shown in Fig. 7, $\{h_1, h_2, \dots h_K\}$). They are computationally efficient and easy to

Method	Taxonomy	Application
PCA Hashing (Gong and Lazebnik, 2011; Yu et al., 2013)	Unsupervised Linear	Multi-modality images (Yu et al., 2013)
Kernelized Hashing (Liu et al., 2012)	Supervised Nonlinear	Brest histopathology (Zhang et al., 2015c) Cell-level histopathology (Zhang et al., 2015d)
Composite Hashing (Gong et al., 2013; Liu et al., 2011)	Unsupervised Nonlinear	Digital mammogram (Liu et al., 2016b)
Hashing Forest (Conjeti et al., 2016a)	Unsupervised Linear	Neuron morphology (Mesbah et al., 2015; Yu and Yuan, 2014)
MIPS Binary Coding (Shen et al., 2015a)	Unsupervised Nonlinear	Neuron morphology (Li et al., 2017a)
Deep Autoencoder (Sze-To et al., 2016; Vincent et al., 2010)	Unsupervised Nonlinear	X-ray images (Sze-To et al., 2016)

Table 2: Existing hashing based large-scale medical image retrieval methods with their taxonomies and applications

optimize (Gong et al., 2012; He et al., 2012; Trzcinski and Lepetit, 2012). However, linear hashing functions can not handle the situation when the difference among image data are subtle and linearly inseparable. Thus, nonlinear hashing was developed to override such limitations. Such methods learn hashing functions based on kernel matrixes or manifold structures and can embed the intrinsic structure in a high-dimensional space and nonlinearly map feature vectors into binary codes (Kulis and Grauman, 2012; Liu et al., 2012; Shen et al., 2015c).

4.2.3. Methodology Review

Table. 2 summarizes the existing hashing-based medical retrieval methods, as well as their corresponding taxonomies and applications. According to Table. 2, both supervised and unsupervised, linear and nonlinear hashing methods have been developed for medical retrieval. In this subsection, we briefly review the above hashing methods and also discuss their advantages and drawbacks.

PCA Hashing (Yu et al., 2013): it first linearly projects raw image features into uncorrelated dimensions via Principal Component Analysis (PCA), where each new feature dimension is orthogonal to each other. Then, it learns the hashing function (i.e. a rotation matrix) by minimizing the binarization error between the new feature matrix and the corresponding binarized feature matrix (Gong and Lazebnik, 2011). PCA Hashing demonstrates high computational efficiency and comparable retrieval precision compared with traditional feature-based nearest-neighbor search. However, since both PCA projection and hashing function optimization are linear, PCA hashing cannot achieve good performance when tackling medical images that are complex (e.g., image differences are subtle, the feature space is not linearly separable).

Kernelized Hashing (Zhang et al., 2015c): for most medical images, linear inseparability is a critical constraint that needs to be taken into account during large-scale retrieval. To tackle this challenge, Kernelized Hashing considers the hashing function with kernels, since kernel methods can map the feature vectors into a high-dimensional space and make the linearly inseparable images easy to differentiate. Accordingly, the learned binary codes from kernelized hashing are also able to differentiate complex medical images. In addition, Kernelized Hashing designs a supervised framework by collaborating kernel functions with medical labels (e.g., labeling the histopathological image with benign or malignant). The supervised information significantly boosts the retrieval performance since it can bridge the semantic gap between low-level features and high-level clinical analytics.

Composite Hashing (Liu et al., 2016b): this algorithm can generate more effective hash codes by integrating global features (e.g. GIST (Oliva and Torralba, 2001)) and local features (e.g. SIFT (Lowe, 2004)) with different distance metrics. In general, single types of features can not comprehensively represent a medical image. On the other side, simply combining multiple features may also fail to achieve accurate image retrieval, since each type of feature has its specific meaning and representation. Thus, Composite Hashing improves the Anchor Graph with multiple features and fuses them by distance metric and local manifold. Then, it learns the hashing function using iterative quantization.

Hashing Forests (Conjeti et al., 2016a): this approach learns binary codes by training independent hashing trees. For the internal node in each tree, locality preserving projections are employed to project data into a latent subspace, where separability between dissimilar points is enhanced. For each input image, each trained tree generates several bits of binary codes, and the combination of these binary codes in the forest is used to represent the input image. Additionally, it employs an inverse-lookup search scheme to improve the efficiency of similarity comparisons. Hashing Forests can generate any given length of binary codes, which is particularly suitable for low-dimensional image features.

MIPS Binary Coding (Li et al., 2017a): as demonstrated in (Liu et al., 2012; Shen et al., 2015a), the Hamming distance and the inner code product have a one-to-one correspondence. Thus, unlike the above methods based on the Hamming distance metric, MIPS (Maximum Inner Product Search)

Binary Coding introduces a new objective based on the inner code product, which is more likely to learn non-linear hashing functions. Specifically, by adopting an alternative iteration strategy, it learns two asymmetric hashing functions for the image database and the query image respectively. This strategy can make the inner product based objective easy to optimize. It also promotes the hashing functions to map binary codes into a high-dimensional non-linear space.

Deep Autoencoders (Sze-To et al., 2016): this algorithm employs deep architectures to hash medical images into binary codes without class labels. Specifically, it uses a specific unsupervised deep architecture, namely deep denoising autoencoders (DDA) (Vincent et al., 2010) to enhance feature learning and binary coding with four steps: image pre-processing, unsupervised layer-by-layer training, unsupervised fine-tuning with dropout, and decoder removal. Finally, a threshold (> 0.5) is applied on the real-valued feature vectors to obtain binary codes. Deep Autoencoders learn binary codes without using any supervised information, which is suitable for medical images where labels are hard to obtain.

When using hashing methods to tackle large-scale medical image retrieval problems, we should not only focus on the hashing methods itself but also need to consider their possible adaptations for different medical image data sets. When the annotation of all medical images in data sets are available, supervised hashing methods are more suitable and are generally more accurate than unsupervised and semi-supervised hashing. For example, Kernel-Based Supervised Hashing (KSH) (Liu et al., 2012), Supervised Discrete Hashing (SDH) (Shen et al., 2015b), Deep Supervised Hashing (DSH) (Liu et al., 2016a) can achieve excellent performance in many public data sets. However, in many cases when the medical image annatations are not easy to acquire, semi-supervised/unsupervised hashing is a more reasonable choice (e.g., Discrete Graph Hashing (DGH) (Liu et al., 2014b), MIPS (Shen et al., 2015a), Semi-Supervised Hashing (SSH) (Wang et al., 2012)). In addition, for some medical images that are not easy to differentiate, non-linear hashing methods can usually achieve much better retrieval performance, such as Inductive Manifold-Hashing (IMH) (Shen et al., 2013), AGH (Liu et al., 2011), despite training non-linear hashing functions being more time-consuming than training linear hashing functions.

4.3. Other Methods

Besides the vocabulary tree and hashing, there are many other methods that have been designed to tackle the feature indexing of large-scale medical image databases. These methods can be either accelerating the similarity search or improving the retrieval accuracy. We briefly introduce and discuss these methods.

4.3.1. Feature Compression

Indexing in large medical databases is usually very time-consuming, especially when the images are represented by high dimensional features. To accelerate the indexing process, one kind of methods is feature compression, which can compress long image features into a smaller size. Hashing belongs to the category of feature compression that is discussed above. In addition to hashing, many other compression methods have been employed for efficient medical image retrieval.

Principal components analysis (PCA) is one of the most popular method for feature compression. After feature extraction, medical images can be represented by single or multiple feature vectors that have high dimension. Many medical image retrieval methods have employed PCA to reduce the feature dimensionality. For example, Tian et al. (Tian et al., 2008) first presented a global and local texture feature combination for the description of medical images. Then, they adopted PCA to reduce the dimension of the combined features. In the analytics of histopathological images, Sertel et al. (Sertel et al., 2009) introduced a novel color-texture analysis approach that combines a model-based intermediate representation with low level texture features. Then, PCA and LDA (linear discriminant analysis (Fukunaga, 2013)) are employed for feature dimensionality reduction. PCA-based medical image retrieval can significantly reduce the feature dimensionality and usually demonstrates comparable performance with the methods using the original features.

In addition to PCA, multiple methods have been proposed for medical feature compression in recent years. In (Foncubierta-Rodríguez et al., 2013), Foncubierta-Rodriguez et al. presented a medical image retrieval method using a bag of meaningful visual words. As visual vocabularies are often redundant, over-complete and noisy, they presented a pruning technique based on probabilistic latent semantic analysis (PLSA) (Hofmann, 2001). The PLAS pruning can enormously reduce the feature dimension when describing a medical image data set with no significant effect on accuracy. More recently, Lan and Zhou (Lan and Zhou, 2016) proposed a simple yet discriminant feature, called histogram of compressed scattering coefficients (HCSCs) for medical image retrieval. They first performed a particular variation of deep convolutional networks, i.e., the scattering transform, to yield high dimensional features. Then a compression operation is carried out on the obtained coefficients for a dimensionality reduction.

4.3.2. Re-ranking

After the similarity indexing through feature compression and other large-scale methods, a set of top ranked medical images can be efficiently computed based on a distance measure. However, these retrieved images may not always correspond to what a human would want and the retrieval precision can vary strongly using different features. Therefore, re-ranking of the coarse results is expected to further improve the retrieval performance for more accurate retrieval. Particularly, re-ranking methods can reorder the initially retrieved images to move the most relevant images to the top or optimise diversiy in the top results.

In recent years, multiple methods have been proposed for re-ranking in different image retrieval applications. In the medical domain, based on the information employed for re-ranking, the re-ranking methods can be roughly divided into three categories, i.e., textual-visual based, multi-feature based and user-feedback based. In the following, we briefly review relevant articles about the three categories respectively:

1. Textual-visual based: these kinds of methods first retrieve relevant medical images through textual indexing, then the initial results are re-ranked by considering the visual similarity. Textual-visual based re-ranking was adopted by many groups in the ImageCLEF medical image retrieval tasks. For example, Radhouani et al. (Radhouani et al., 2009) introduced their work at ImageCLEF 2009. They first leveraged textual data to search relevant images in three domain dimensions, anatomy, pathology and modality. Then, they employ the visual data to re-rank the document lists based on the extracted features, including a color and intensity histogram, gray-level co-occurrence matrices and other features. Besides this, Depeursinge and Müller (Depeursinge and Müller, 2010) described several fusion techniques for combining textual and visual information that were used in ImageCLEF.
2. Multi-feature based: this kind of method first computes the retrieval results from multiple kinds of features, then the final results are ob-

tained by fusing and re-ranking the above retrieved images. Recently, Zhang et al. (Zhang et al., 2016a) presented a method for histopathology image analysis by re-ranking the results from multiple features. Specifically, after obtaining several top ranked relevant images from multiple kinds of features, they employed a graph-based query-specific fusion approach where multiple retrieval results are integrated and reordered based on a fused graph (Zhang et al., 2015b). In general, such re-ranking methods can significantly improve the retrieval performance since they consider the image similarity and discrimination from several viewpoints using multiple features, e.g., local and holistic features.

3. User-feedback based: after receiving the initial results, this kind of method re-ranks the retrieved images based on relevance feedback from users. The relevance feedback can specify which image is relevant/irrelevant. Agarwal and Mostafa (Agarwal and Mostafa, 2011) employed the user-feedback based re-ranking for the Alzheimer's disease detection. They first described a content-based image retrieval system, i.e., ViewFinder Medicine (vfM), to combine visual and textual features for initial indexing. Then the retrieval system employed the user-provided feedback to perform re-ranking, including inter-session and intra-session re-ranking. This re-ranking process improved the system precision from 0.8 to 0.89. The importance of negative feedback in this process is highlighted in (Muller et al., 2000).

In most cases, re-ranking methods are only required to consider the top ranked initial retrieval results, e.g., most truly relevant images are included in the top-K results, and K is much smaller than the number of images in the whole database. Therefore, re-ranking can be very efficient as it only needs to process a few images. More importantly, by considering and comparing the similarity using multiple information sources, the retrieval precision can be improved for further exploration and analysis.

4.3.3. High-performance Computing

In addition to the above large-scale methods which belong to the fields of image processing, computer vision and machine learning, High-performance Computing (HPC) also plays an important role in medical image analytics. HPC is the use of parallel processing techniques to execute programs efficiently, reliably and quickly. The HPC techniques include parallel computing, distributed computing, cloud computing, etc. that are useful for tackling large databases. Particularly in the medical field, some large databases are usually stored in different locations and they are essential to be processed based on parallel systems.

Recently, HPC techniques have been widely employed for the large-scale medical retrieval. Foran et al. (Foran et al., 2011) proposed a software system based on parallel and distributed computing, namely ImageMiner, to efficiently retrieve and analyze the expression patterns of tissue microarrays (TMAs). The ImageMiner system embedded a data analysis component for efficient retrieval, i.e., DataCutter (Kumar et al., 2006), which the data processing pipeline can be composed as a network of interacting components. Images received by ImageMiner were distributed and processed by the computation cluster using a master-slave parallelization scheme. Subsequently, Qi et al. (Qi et al., 2014) investigated large-scale histopathological image retrieval using the CometCloud (Kim et al., 2011), an automatic cloud framework that allows dynamic, on-demand federation of distributed infrastructures. They first formulated the histopathological image retrieval problem as a set of heterogeneous and independent tasks. Then these tasks can be parallelized and solved using the aggregated computational power of distributed resources. More recently, Markonis et al. (Markonis et al., 2015b) proposed solutions for the large-scale medical image analysis based on parallel computing and algorithm optimization. Specifically, a MapReduce framework is employed to speed up the medical image analysis in three tasks, i.e., lung texture segmentation using support vector machines, content-based medical image indexing and 3D directional wavelet analysis for solid texture classification.

High-performance computing can well be used to handle large-scale retrieval tasks, especially for clinical systems, where the parallelized processing can achieve similarity retrieval in real-time. More importantly, as presented in Fig. 1, high-performance computing can be adopted in both the feature extraction/indexing and retrieval, which can dramatically improve the retrieval efficiency in these time-consuming steps.

5. Evaluation

After receiving similar samples from medical image retrieval systems, evaluating the retrieval performance and the whole retrieval system are also critical tasks. Especially for large-scale medical image sets, simply using class labels is usually not adequate to evaluate the retrieval performance in fine-grained levels. In the past decades, challenges and tasks such as ImageCLEF, VISCERAL, etc. have made great efforts for the evaluation of medical image retrieval (Kalpathy-Cramer et al., 2015; Langs et al., 2012). This section reviews related work of evaluation protocols which are relevant to medical image retrieval, including evaluation measures, criteria, and public medical image data sets.

5.1. Evaluation Measures

We first introduce the evaluation measures for medical image retrieval that can provide a quantitative analysis, comparison, and validation of different retrieval methods. In general, the evaluation measures in large-scale medical image retrieval are similar with the measures in generic information retrieval, i.e., evaluating the precision, recall, efficiency and several other measures.

Precision: retrieval precision is the main indicator for performance evaluation, which can be denoted as the fraction of the images retrieved that are relevant to the query image:

$$precision = \frac{|\{relevant\ images\} \cap \{retrieved\ images\}|}{|\{retrieved\ images\}|} \tag{3}$$

In information retrieval, precision can evaluate the capability of a method for searching similar or relevant samples. It has also been widely used for the evaluation of medical image retrieval methods, especially for some medical analytical tasks where the image used as query can be better interpreted with similar/relevant images (Li et al., 2017b; Zhang et al., 2015c,d). This is similar to asking a colleague for help or searching similar images/patterns in books.

Besides precision, mean average precision (MAP) is most commonly used for the evaluation of retrieval methods and for the comparison of search in large-scale medical image sets. MAP is relatively stable and include aspects of precision and recall, as it averages over positions of all relevant items. It is defined as the mean of the average precision scores of all relavnt items of a query averaged over all queries. The MAP can be formulated as:

$$MAP = \frac{1}{|M|}\sum_{m=1}^{M}\frac{1}{|K|}\sum_{k=1}^{K} precision(Q_{m,k}) \tag{4}$$

where M is the number of query images (i.e., testing data), K indicates the top-K ranked relevant images for each query image, and $Q_{m,k}$ denotes the top-k retrieval precision of the mth query image. For large-scale retrieval methods, the MAP can evaluate their performance with massive testing data (e.g., hundreds to thousands of query images), and thus alleviate the bias during precision evaluation.

Recall: in image retrieval, recall is the fraction of relevant retrieved images with all relevant images in databases, i.e.:

$$recall = \frac{|\{relevant\ images\} \cap \{retrieved\ images\}|}{|\{all\ relevant\ images\}|} \tag{5}$$

Recall reflects the sensitivity of a retrieval system, i.e., whether it can completely find all relevant samples in top-K ranked results, keeping K as small as possible. Thus, for medical retrieval tasks that need to find all relevant samples for analysis (such as a systematic review), recall is a critical evaluation criterion. Normally, recall is associated with precision, i.e., precision-recall curve, for the evaluation and comparison of different retrieval methods with a global view on the performance (Davis and Goadrich, 2006; Müller et al., 2001).

Efficiency: as directly indexing massive images with high dimensional features are usually very time-consuming, one important evaluation indicator for large-scale retrieval is efficiency. Currently, in most large-scale retrieval cases, efficiency is denoted as the time for the feature indexing phase, i.e., given a query image (or its features), the time for returning a set of relevant images after searching in large-scale databases. For medical image retrieval with many testing images, their accumulated and average run time are the commonly used efficiency measures, where the average run time can be formulated as:

$$AvgTime = \frac{1}{M}\sum_{m=1}^{M} t_{m,K} \tag{6}$$

$t_{m,K}$ indicates the time cost of retrieving K relevant images for the mth query image. The average/accumulated run time has been widely adopted for the evaluation, comparison and validation of large-scale medical image retrieval (Jiang et al., 2016a, 2015c; Zhang et al., 2015c,d). Still, run times need to be put in relationship to hardware resources available and are thus not always easy to interpret. Sometimes the run time for the offline parts (data indexation) and the online parts (interactive search) are separately compared.

Additionally, there are several other commonly employed measures for medical image retrieval evaluation. For example, the precision after the first N_R images are retrieved (i.e., $P(N_R)$), recall at 0.5 precision, rank first relevant, etc.). These measures were discussed in previous articles (Müller et al., 2001; Muller et al., 2004).

5.2. Evaluation Criteria

In addition to evaluation measures, the criteria of deciding similarity/relevance are also important and challenging tasks in large-scale medical image retrieval. Here, we introduce two kinds of evaluation criteria: annotation-based and user-based, which are the commonly employed criteria in medical image retrieval.

Annotation-based Criteria: when class labels of medical images are available, their annotations are a commonly used evaluation criterion. As the class labels of all medical images in the database are provided, the similar or relevant images can be determined quickly by comparing their class labels. Thus, given testing images, the retrieval precision and recall can be measured by sequentially comparing the labels between each test image and the retrieved images. Currently, several large-scale medical image retrieval cases adopted annotation-based criteria for performance evaluation. For example, Zhang et al. (Zhang et al., 2015d) evaluated the large-scale histopathological image retrieval through the class label of two type lung cancers (i.e., adenocarcinoma and squamous carcinoma) for each image. The annotation-based evaluation criteria are only suitable for the cases that image classes are identified and the similarity of images are simply determined by class labels.

User-based Criteria: despite the annotation-based criteria being a simple way for retrieval evaluation, it may not suitable in many practical cases of large-scale medical image retrieval. One reason is that the annotation of medical images is usually hard to obtain. Some medical images are still classified and do not have unified classification rules. Moreover, annotating every medical image in large databases is extremely labor expensive,time-consuming, and sometimes impossible. Another reason is the similarity/relevance measure. In large-scale medical image retrieval, one query image may have thousands of images with the same label. For some analytical tasks, simply using class labels is not adequate to identify relevant images.

Compared with annotation-based criteria, users or domain experts can provide more fine-grained retrieval evaluations in the form of relevance judgements for specific tasks. Many medical image retrieval systems have employed users for the performance evaluation. In general, users can observe the retrieved images and assign them with different relevance levels during evaluation. For example, the medical ImageCLEF challenges used three levels of relevance, i.e., relevant, partly-relevant, and non-relevant (Kalpathy-Cramer et al., 2011; Müller et al., 2012, 2009). These relevance judgments were employed for the retrieval performance evaluation of the database with $300,000$ medical images. Besides ImageCLEF challenges, considering neurons is usually hard to classify and identify, Wan et al. (Wan et al., 2015) asked two users for the visual comparison of morphological neuron retrieval results. In medical image retrieval, user-based criteria rely on user's domain knowledge and may be partly subjective based on the user's background. Thus, the retrieval results are usually judged by two or more users for more reliable evaluation.

The evaluation of system design also plays an important role in medical image retrieval, especially for the retrieval systems where users are interactively involved. Markonis et al. (Markonis et al., 2015a) reported the user-orientied evaluation of a text- and content-based medical image retrieval system. In total, 16 radiologists participated in the user tests with a working image retrieval system in an iterative manner. Such analyses in clinical practice are really needed to advance the practical use of image retrieval in hospitals

5.3. Public Datasets

With the increasing availability of digital imaging techniques, a large number of medical images are generated and well organized in many repositories. Some of the repositories are publicly available for users and researchers. The medical image repositories usually include thousands to millions of images. Images are collected for different purposes, such as cancer grading/staging and treatment planning. We briefly introduce some of the public data sets that are widely used for medical image retrieval:

- ImageCLEF (ImageCLEF): ImageCLEF provides an evaluation forum for the cross-language annotation and retrieval of images. ImageCLEF has held 14 years of medical image retrieval challenges, with the number of images in the dataset having increased from $6,000$ to $300,000$ (Kalpathy-Cramer et al., 2015). The datasets in ImageCLEF include multiple modalities of medical images, e.g., radiology, microscopy and also general photography.

- DDSM (of South Florida): The digital database for screening mammography (DDSM) is a public mammogram database. It includes $2,604$ breast cases and every case consists of four views, with two cranio-caudal views and two mediolateral oblique views. The mammographic masses have different shapes, sizes, margins and breast densities as well as the patient race and age, which provide rich information for diagnosis.

- MedPix (of Medicine): MedPix is a fully web-enabled cross-platform database, integrating images and text information. This medical image database includes over $53,000$ indexed and curated images, from more than $13,000$ patients. The merit of this database is that it records detailed descriptions of patients and their corresponding diagnosis.

- TCGA (Institute, a): The Cancer Genome Atlas (TCGA) collects a huge amount of cancer images (currently around $10,000,000$ images and increasing quickly) from multiple projects funded by National Cancer Institute. It records many types of cancer images, including but not limited to, brain, esophageal, lung, thyroid and rectum. All TCGA data reside in the Genomic Data Commons (Institute, b).

- TCIA (TCIA): The Cancer Imaging Archive (TCIA) is organized into collections with a variety of cancer types and/or anatomical areas. Similar to TCGA, it collects cancer images from many projects and institutes. The cancer types include breast, prostate, liver, lung, brain, etc. and the image modalities include CT, MR, PET and others.

- VISCERAL (VISCERAL): VISCERAL is the abbreviation for Visual Concept Extraction Challenge in Radiology, which provides a benchmark for the retrieval in the medical domain. This dataset consists $2,311$ medical 3D volumes originating from various modalities (CT, MRT1, MRT2 with and without contrast agent) and each volume consists $200-2000$ images (slices). The VISCERAL project has organized

Public data sets	Number of images or or size	Image category
ImageCLEF (ImageCLEF)	300,000	Multi-modalities
DDSM (of South Florida)	10,480, 231GB	Mammogram
MedPix (of Medicine)	53,000	Multi-modalities
TCGA (Institute, a)	470TB	Cancer Images, Multi-modalities
TCIA (TCIA)	10,000,000, 3TB	Cancer Images, Multi-modalities
Retinopathy (EyePACS)	35,000, 82GB	Retinal Photographs
DREAM (Bionetworks)	640,000	Screening Mammograms
VISCERAL (VISCERAL)	2,300	3D CT, MRI volume
LIDC-IDRI (Armato III et al., 2011)	240,000, 124GB	Lung CT, DX, and CR
ADNI (of Southern California)	Unknown	Alzheimer's MR, PET, etc
NBIA (NBIA)	76,000	Cancer Images, Multi-modalities
CAMELYON 17 (, DIAG)	1,000, 2TB	Whole-slide Histopathological Images
PubMed Center (NCBI)	4,000,000	Multi-modalities
NLST (Institute, c)	76,000	Lung CT, Pathology Images

Table 3: Current publicly available medical image data sets.

several challenges, workshops and provided multiple benchmarks related to large-scale data in medical image analysis and retrieval (Langs et al., 2012; Müller et al., 2014; Zhang et al., 2015a).

In addition to the above data sets, Table. 3 presents a summary of publicly available data sets with many medical images, including number of images, size and categories if available. Due to the fast growth of medical images, we only provide a small subset of commonly used data sets in Table. 3.

6. Applications

After reviewing the above large-scale techniques, we introduce their applications for medical image analytics in this section. Large-scale retrieval methods have demonstrated impressive improvement on many medical image types, including CT, MRI, X-ray, microscopy and others. In the following, we illustrate their applications in clinical diagnosis, cancer grading, and neuron exploration.

6.1. Mammographic Retrieval and Segmentation

Breast cancer remains the second leading cause of cancer-related death among women (Society, 2013). Early diagnosis based on mammography is a widely adopted approach to improving the chances of recovery, which is recognized as a gold standard for breast cancer detection by the American Cancer Society (Society, 2013). However, the detection of masses in a mammogram is a challenging task, as masses have a large variation in shape, margin, and size. They are often indistinguishable from surrounding tissue (Cheng et al., 2006; Oliver et al., 2010). For an undetected mammogram, computer-aided diagnosis (CAD) with content-based image retrieval (CBIR) is an effective solution by returning a limited number of the most similar mammograms in the pre-built image database, where the retrieved mammograms were already annotated with the class labels of mass and normal. Nevertheless, with the ever increasing number of mammograms generated and added to the pre-built database, scalable CBIR techniques have become one of the important problems for mammogram based breast cancer diagnosis (Langs et al., 2012).

Jiang et al. (Jiang et al., 2015c) successfully solved the scalable mammogram retrieval problem based on a vocabulary tree with adaptive weighting. For a query with a mammographic region of interest (ROI), it can achieve efficient retrieval in a dataset with $11,553$ ROIs. Specifically, in the experiment, this method reported an 88.4% retrieval precision with 500 mass ROIs and 500 normal ROIs as queries. This demonstrates good accuracy compared with other methods including NMI (Tourassi et al., 2007), BoW (André et al., 2012), and VocTree (Nister and Stewenius, 2006). The method also achieved highest classification accuracy (90.8%) for whether the query ROIs are masses or normal. Additionally, this method is 3 to 10 times faster than other methods and the advantage is larger when the size of image database increases.

(Jiang et al., 2016b) propose to learn online shape and appearance priors via image retrieval, i.e., setting an input mass as the query, its visually similar training masses can be obtained by image retrieval. Then, the query mass can be segmented using the retrieval priors and graph cuts. Extensive experiments on a mammography database demonstrate that the method can improve the segmentation accuracy and outperform several widely used mass segmentation methods.

6.2. Cell-Level Histopathological Image Analysis

Histopathological image analysis is widely used for cancer grading. Compared to mammography, CT and others, histopathology slides provide more comprehensive information for diagnosis and the diseases are analyzed by detecting tissue and cells in lesions (Gurcan et al., 2009). On the other hand an invasive biopsy is necessary, which is often tried to be avoided. CBIR systems are commonly employed to analyze histopathological images (Caicedo et al., 2009, 2011; Doyle et al., 2007). In CBIR systems, the returned visually similar images can be used to identify and classify the query images (e.g.,

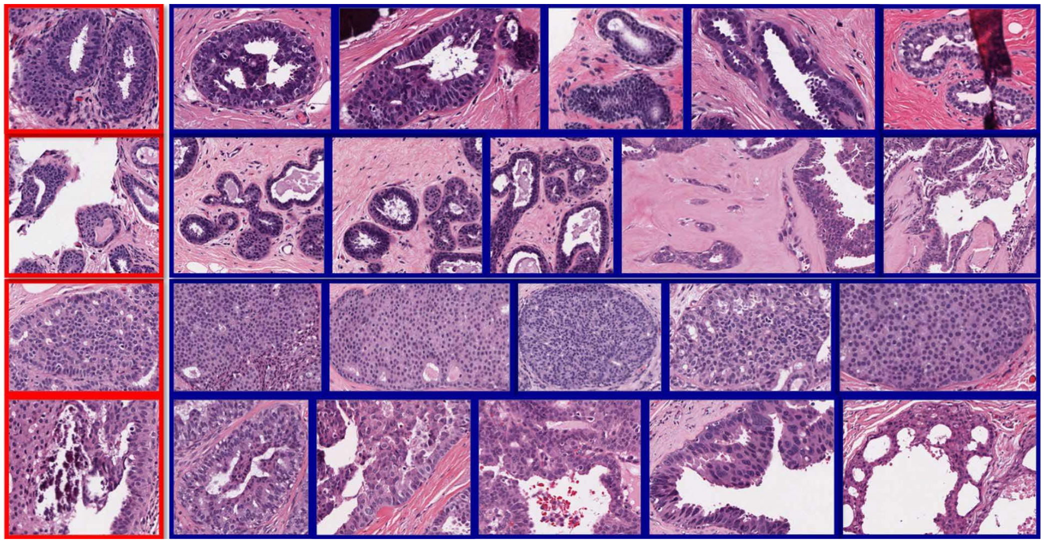

Figure 8: Examples of hashing-based histopathological image retrieval illustrated in (Zhang et al., 2015c) (query marked in red and retrieved images marked in blue). The first two rows are benign tissue; the last two rows are malignant tissue.

classifying them as benign or malignant), and further assist pathologists to describe the tissue samples.

Hashing methods were first employed by Zhang et al. (Zhang et al., 2015c, 2014) to tackle large histopathological image databases for CBIR. They designed a comprehensive CBIR framework to analyze histopathological images by leveraging high-dimensional texture features and kernelized hashing with supervised information. In the experiment, this hashing method demonstrated significant improvement in histopathological image classification and retrieval tasks. Compared to methods such as SVM (Caicedo et al., 2009), Adaboost (Doyle et al., 2012), KNN (Tabesh et al., 2007), and Graph Embedding (Basavanhally et al., 2010), its accuracy was 5 to 10 percent higher. The method achieved histopathological retrieval for 700-900 images within 0.01 seconds (3121 images in the database), which is 1000 times faster than the given baseline. Fig. 8 illustrates four queries (two benign images, two malignant images) and their corresponding top five retrieval results based on this hashing-based CBIR framework. Despite the difference between benign and malignant images being subtle, the proposed method is effective to retrieve images in the same category. The authors extended the CBIR system for more accurate diagnosis by examining the cells in histopathology images (Zhang et al., 2015d). As each histopathology image usually includes thousands of cells, examining every cell by traditional retrieval methods is almost impossible when the image databases are large. Thus, a hashing-based framework is proposed that enables cell-level analysis in real-time with high accuracy, i.e., indexing $96,000$ cells within 1.68 seconds (the whole database includes $484,136$ cells), and achieving 87.3% accuracy for the classification of histopathology lung images (i.e., two types of lung cancers, adenocarcinoma or squamous carcinoma).

In histopathological image analysis, it is a common practice to employ multiple features to improve performance. To embed multiple features in a hashing framework, Jiang et al. (Jiang et al., 2015b, 2016a) employed joint kernel-based supervised hashing (JKSH) for scalable histopathological image analysis, where multiple features are linearly combined by individual kernels (Liu et al., 2014c). Experiments on breast cancer histopathology images demonstrate the effectiveness in both retrieval and classification.

6.3. Exploration of a Neuron Databases

Analyzing single neuron properties, such as cell types, brain regions, functions and development stages is usually a fundamental task to understand the nervous system and brain mechanisms. In general, neuron morphology plays a major role in determining the neuron's functional and physiological properties. Recent approaches in neuroscience (e.g., BigNeuron (big, a)) have facilitated the research in neuron morphology. An increasing number of neurons are reconstructed and added to the public repositories (big, b; NeuroMorpho). Therefore, given an unknown neuron, it is reasonable to explore its properties through the morphological retrieval in neuron databases.

Conjeti et al. (Conjeti et al., 2016b; Mesbah et al., 2015) developed an advanced tool for morphological search and retrieval in large-scale neuroscientific image databases, namely Neuron-Miner. Neuron-Miner first employs quantitative measurements as neuron features, such as soma surface, the number of branches and the neuron's total length. Then, it adopts a novel hashing method, i.e., hashing forests, to compact the features into binary codes. In the experiment, Neuron-Miner demonstrates the effectiveness in morphological retrieval with a database including $31,266$ neurons. Given a query, this tool is able to return several visually similar neurons from the database. The ground truth (using normalized Euclidean distance) shows that returned neurons are relevant to the query.

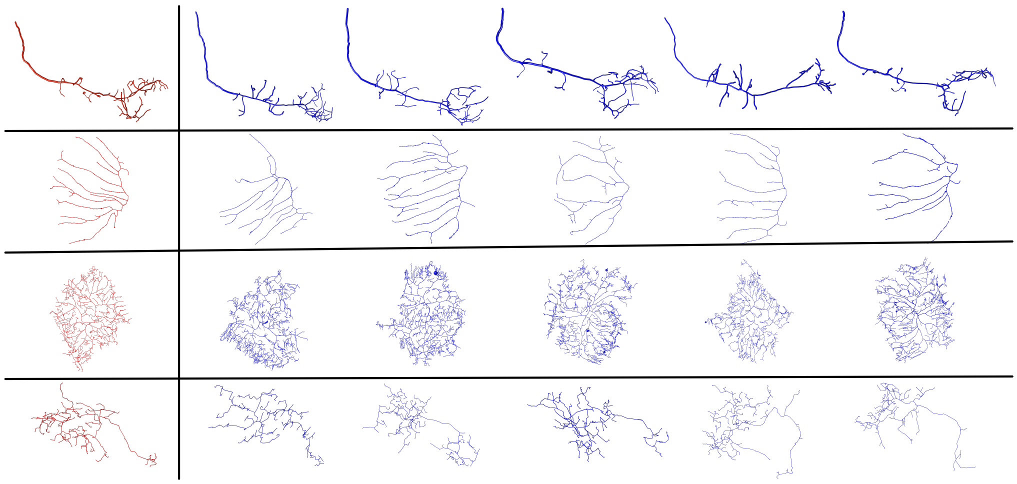

Figure 9: Results of morphological neuron retrieval shown in (Li et al., 2017a). For each neuron on the left (red), the top-5 retrieved neurons on the right (blue) are shown. This illustrates the morphological similarity between query neurons and retrieved neurons.

More recently, Li et al. (Li et al., 2017a, 2016) explored large-scale morphological neuron databases based on a novel search strategy, the maximum inner product search (MIPS). Based on MIPS, nonlinear hashing functions are learned by embedding the inner code product rather than the conventional Hamming distance. The nonlinear hashing functions are particularly suitable for the morphological neuron retrieval problem, since the neurons' tree-topological structure makes them hard to be discriminative in low-dimensional linear space. Fig. 9 demonstrates that the MIPS-based method is able to retrieve morphologically similar neurons in the large-scale database. To evaluate the retrieval precision, it employed projection neurons in the olfactory bulb as queries. The retrieval results validated that most returned neurons have the same properties as the queries (with a reported 90.48% average precision in the top-5 relevant neurons). Additionally, the authors demonstrated the application of morphological retrieval in neuron exploration. By collecting properties of the top-K relevant neurons (e.g., a neurons' brain regions, cell types, transmitters). Properties of the query neuron can be inferred in real-time based on this MIPS hashing framework.

7. Future Directions

After reviewing the above methods and applications of large-scale medical image analytics, we discuss possible future directions in this section. Despite varieties of advanced large-scale techniques being employed for retrieval, there are still many directions to explore and improve the retrieval performance.

Multi-features: in general, only employing a single kind of feature is not enough to represent and discriminate medical images. Especially when a database is large, the difference with some irrelevant images can be subtle. One intuitive solution is using multiple features to represent each image, e.g., local, holistic, and texture features. These features can be fused and embedded in a large-scale retrieval framework. According to existing work, multi-feature fusion can be conducted on three levels during retrieval, i.e., feature level (Atrey et al., 2010), training level (Liu et al., 2014c), and decision level (Zhang et al., 2012). Jiang et al. (Jiang et al., 2016a) fuse three types of features (SIFT (Lowe, 2004), HOG (Dalal and Triggs, 2005), and GIST (Oliva and Torralba, 2001)) in the training level when learning hashing functions; Zhang et al. (Zhang et al., 2016a) employ a graph-based query-specific fusion approach to integrating local and holistic features at the decision level. Despite the two methods having achieved good performance, these are far from enough for large-scale medical image retrieval. With the ever-increasing techniques in feature representation, employing more features to retrieve complex medical images is a clear trend (e.g. the ImageCLEF medical image retrieval tasks in recent years). However, as diverse features have different meanings and representations, deciding on the importance of each feature is a challenging task. User specified feature importance is usually not reliable, and automatically computing each feature's importance is time-consuming, especially when dealing with many features in a large database. Thus, successfully handling multi-feature fusion in a large-scale database further improves the accuracy and efficiency of medical image retrieval.

Online updating: as more medical images are being generated, the size of the corresponding databases are continuously increasing. For example, the aforementioned ImageCLEF database increased the number of images from 600 to $300,000$, and the NeuroMorpho database usually releases $1,000$ to $2,000$ reconstructed neuron cells in each update. The newly added images should be considered to train new models for retrieval, since employing more training data can accordingly improve the retrieval accuracy. However, if we re-train a large-scale model every time from scratch, using both the original and the newly added images, it is time-consuming and adversely affects the efficiency of medical retrieval. On the other side, when the medical image databases are extremely large (e.g., including millions of images), current storage techniques are not able to arrange and process all the images within one batch. More importantly, both the vocabulary tree and hashing based methods cannot efficiently train models for the huge amount of images at a given time, e.g., building a hierarchical tree or learning hashing functions with millions of feature vectors. To tackle these problems, one possible solution is to divide huge databases into several batches, and then develop an online updating strategy to train the retrieval model with one-by-one image batches in a streaming manner. The newly added images can also be treated as a batch to update the retrieval model. Currently, several online hashing methods have been developed for computer vision tasks (Cakir and Sclaroff, 2015; Huang et al., 2013; Leng et al., 2015). In medical image analytics, the merit of the online updating strategy is beneficial in the future with a continuously increasing number of images and extremely large databases.

Bringing humans in the loop: for the retrieval of large-scale medical image databases, lacking label information is the main limitation to achieve good retrieval results. As medical images usually have different modalities and appearances, their intra-class variations can be large, and their inter-class variations can be small. The image labels are useful to handle this problem, since it can embed supervised information in retrieval models and bridge the low-level features with high-level image semantics. However, labeling images is not an easy task. Especially for some medical images, assigning their labels requires domain experts with proper training. Crowdsoucing can be used when very precisely defined tasks allow for quick training times (Foncubierta Rodríguez and Müller, 2012). Deciding whether a histopathology image contains benign or malignant lesions is complex and time-consuming, for example. Moreover, large-scale databases make this task even harder. To tackle these problems, one feasible solution is to bring humans in the retrieval loop. They can interactively give feedback to improve the retrieval performance (Feng et al., 2013; Rui et al., 1998). After acquiring a set of similar images from unsupervised retrieval, users/domain experts can specify images relevant to the query and those that are not. Such feedback can be returned to the retrieval system to improve the final results (Bulo et al., 2011; Sahbi et al., 2007). The feedback can be treated as supervised information but it is more efficient than labeling all medical images. Theoretically, such an interactive strategy can achieve two goals: 1) it presents retrieval results to users/domain experts to help them analyze medical images; 2) it receives and uses the interactive feedback to improve the retrieval system.

8. Conclusions

In this review, we summarize recent advances of large-scale retrieval for medical image analytics. By introducing the pipeline of large-scale retrieval, we presented a comprehensive review of relevant techniques that can improve the efficiency and accuracy of medical image analysis, including feature representation, feature indexing and searching. We also reviewed clinical applications and discussed the future directions of large-scale medical analytics. With the ever-increasing amount of newly generated medical images, we believe that the algorithms and methods of large-scale medical image analytics will lead to new ideas for knowledge discovery and decision support.

Currently, only few systems have been exposed to detailed user testing (Markonis et al., 2015a) and such user tests are clearly needed for very large scale systems. Many currently CBIR systems only use small databases and not update mechanisms and this is required for real application including an integration of the systems into the standard clinical workflow, which is often neglected. Many technical approaches are now available for large-scale applications but more work is needed to actually integrate the tools for clinical impact, an this includes the use of deep learning and explaining these results to physicians.

Acknowledgement

This work is partially supported by the National Science Foundation under grant ABI-1661280, CNS-1629913 and IIP-1439695.

References

, a. Bigneuron project. http://www.alleninstitute.org/bigneuron/. Accessed June 28, 2016.

, b. Bigneuron released data. https://github.com/BigNeuron/Data/releases. Accessed June 28, 2016.

, 2010. Riding the wave: How europe can gain from the rising tide of scientific data. Submission to the European Comission, available online at http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf. URL: http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf.

Agarwal, M., Mostafa, J., 2011. Content-based image retrieval for alzheimer's disease detection, in: Content-Based Multimedia Indexing (CBMI), 2011 9th International Workshop on, IEEE. pp. 13–18.

Akgül, C.B., Rubin, D.L., Napel, S., Beaulieu, C.F., Greenspan, H., Acar, B., 2011. Content-based image retrieval in radiology: current status and future directions. Journal of Digital Imaging 24, 208–222.

André, B., Vercauteren, T., Buchner, A.M., Wallace, M.B., Ayache, N., 2012. Learning semantic and visual similarity for endomicroscopy video retrieval. IEEE Transactions on Medical Imaging 31, 1276–1288.

Antani, S.K., Deserno, T.M., Long, L.R., Güld, M.O., Neve, L., Thoma, G.R., 2007. Interfacing global and local cbir systems for medical image retrieval, in: Bildverarbeitung für die Medizin 2007. Springer, pp. 166–171.

Antipov, G., Berrani, S.A., Ruchaud, N., Dugelay, J.L., 2015. Learned vs. hand-crafted features for pedestrian gender recognition, in: Proceedings of the 23rd ACM International Conference on Multimedia, ACM. pp. 1263–1266.

Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., et al., 2011. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. Medical Physics 38, 915–931.

Atrey, P.K., Hossain, M.A., El Saddik, A., Kankanhalli, M.S., 2010. Multimodal fusion for multimedia analysis: a survey. Multimedia Systems 16, 345–379.

Babenko, A., Lempitsky, V., 2015. Aggregating local deep features for image retrieval, in: Proceedings of the IEEE international conference on computer vision, pp. 1269–1277.

Bailey, D.L., Townsend, D.W., Valk, P.E., Maisey, M.N., 2005. Positron emission tomography. Springer.

Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., Greenspan, H., 2015. Chest pathology detection using deep learning with non-medical training, in: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 294–297.

Basavanhally, A.N., Ganesan, S., Agner, S., Monaco, J.P., Feldman, M.D., Tomaszewski, J.E., Bhanot, G., Madabhushi, A., 2010. Computerized image-based detection and grading of lymphocytic infiltration in her2+ breast cancer histopathology. IEEE Transactions on Biomedical Engineering 57, 642–653.

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008. Speeded-up robust features (surf). Computer Vision and Image Understanding 110, 346–359.

Bengio, Y., 2009. Learning deep architectures for ai. Foundations and trends® in Machine Learning 2, 1–127.

Bengio, Y., Courville, A.C., Vincent, P., 2012. Unsupervised feature learning and deep learning: A review and new perspectives. CoRR, abs/1206.5538 1.

Bionetworks, S., . The digital mammography dream challenge. https://www.synapse.org/Synapse:syn4224222/wiki/. Accessed October 23, 2016.

Bourlard, H., Kamp, Y., 1988. Auto-association by multilayer perceptrons and singular value decomposition. Biological Cybernetics 59, 291–294.

Brosch, T., Tam, R., Initiative, A.D.N., et al., 2013. Manifold learning of brain mris by deep learning, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 633–640.

Bulo, S.R., Rabbi, M., Pelillo, M., 2011. Content-based image retrieval with relevance feedback using random walks. Pattern Recognition 44, 2109–2122.

Bunte, K., Biehl, M., Jonkman, M.F., Petkov, N., 2011. Learning effective color features for content based image retrieval in dermatology. Pattern Recognition 44, 1892–1902.

Cai, W., Liu, S., Wen, L., Eberl, S., Fulham, M.J., Feng, D., 2010. 3d neurological image retrieval with localized pathology-centric cmrglc patterns, in: 2010 IEEE International Conference on Image Processing, IEEE. pp. 3201–3204.

Caicedo, J.C., Cruz, A., Gonzalez, F.A., 2009. Histopathology image classification using bag of features and kernel functions, in: Conference on Artificial Intelligence in Medicine in Europe, Springer. pp. 126–135.

Caicedo, J.C., Gonzalez, F.A., Romero, E., 2007. Content-based medical image retrieval using low-level visual features and modality identification, in: Workshop of the Cross-Language Evaluation Forum for European Languages, Springer. pp. 615–622.

Caicedo, J.C., González, F.A., Romero, E., 2011. Content-based histopathology image retrieval using a kernel-based semantic annotation framework. Journal of Biomedical Informatics 44, 519–528.

Cakir, F., Sclaroff, S., 2015. Adaptive hashing for fast similarity search, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 1044–1052.

Cao, Y., Steffey, S., He, J., Xiao, D., Tao, C., Chen, P., Müller, H., 2014. Medical image retrieval: A multimodal approach. Cancer Informatics 13, 125.

Cheng, H., Shi, X., Min, R., Hu, L., Cai, X., Du, H., 2006. Approaches for automated detection and classification of masses in mammograms. Pattern Recognition 39, 646–668.

Chenouard, N., Unser, M., 2011. 3d steerable wavelets and monogenic analysis for bioimaging, in: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE. pp. 2132–2135.

Ciompi, F., de Hoop, B., van Riel, S.J., Chung, K., Scholten, E.T., Oudkerk, M., de Jong, P.A., Prokop, M., van Ginneken, B., 2015. Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2d views and a convolutional neural network out-of-the-box. Medical Image Analysis 26, 195–202.

Collins, D.J., Padhani, A.R., 2004. Dynamic magnetic resonance imaging of tumor perfusion. IEEE Engineering in Medicine and Biology Magazine 23, 65–83.

Conjeti, S., Katouzian, A., Kazi, A., Mesbah, S., Beymer, D., Syeda-Mahmood, T.F., Navab, N., 2016a. Metric hashing forests. Medical Image Analysis .

Conjeti, S., Mesbah, S., Negahdar, M., Rautenberg, P.L., Zhang, S., Navab, N., Katouzian, A., 2016b. Neuron-miner: An advanced tool for morphological search and retrieval in neuroscientific image databases. Neuroinformatics 14, 369–385.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), IEEE. pp. 886–893.

Davis, J., Goadrich, M., 2006. The relationship between precision-recall and roc curves, in: Proceedings of the 23rd international conference on Machine learning, ACM. pp. 233–240.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database, in: Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE. pp. 248–255.

Depeursinge, A., Fischer, B., Müller, H., Deserno, T.M., 2011. Suppl 1: Prototypes for content-based image retrieval in clinical practice. The open medical informatics journal 5, 58.

Depeursinge, A., Müller, H., 2010. Fusion techniques for combining textual and visual information retrieval, in: ImageCLEF. Springer, pp. 95–114.

(DIAG), D.I.A.G., . The camelyon challenge. https://camelyon17.grand-challenge.org/. Accessed March 23, 2017.

Doi, K., 2014. Current status and future potential of computer-aided diagnosis in medical imaging. The British journal of radiology 78.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T., 2014. Decaf: A deep convolutional activation feature for generic visual recognition., in: ICML, pp. 647–655.

Douze, M., Jégou, H., Sandhawalia, H., Amsaleg, L., Schmid, C., 2009. Evaluation of gist descriptors for web-scale image search, in: Proceedings of the ACM International Conference on Image and Video Retrieval, ACM. p. 19.

Doyle, S., Feldman, M., Tomaszewski, J., Madabhushi, A., 2012. A boosted bayesian multiresolution classifier for prostate cancer detection from digitized needle biopsies. IEEE Transactions on Biomedical Engineering 59, 1205–1218.

Doyle, S., Hwang, M., Naik, S., Feldman, M., Tomaszeweski, J., Madabhushi, A., 2007. Using manifold learning for content-based image retrieval of prostate histopathology, in: MICCAI 2007 Workshop on Content-based Image Retrieval for Biomedical Image Archives: Achievements, Problems, and Prospects, Citeseer. pp. 53–62.

Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2010. The pascal visual object classes (voc) challenge. International journal of computer vision 88, 303–338.

EyePACS, . Diabetic retinopathy detection. https://www.kaggle.com/c/diabetic-retinopathy-detection/. Accessed October 23, 2016.

Fan, L., 2013. Supervised binary hash code learning with jensen shannon divergence, in: Proceedings of the IEEE International Conference on Computer Vision, pp. 2616–2623.

Fang, R., Pouyanfar, S., Yang, Y., Chen, S.C., Iyengar, S., 2016. Computational health informatics in the big data age: a survey. ACM Computing Surveys (CSUR) 49, 12.

Feng, D., Siu, W.C., Zhang, H.J., 2013. Multimedia information retrieval and management: Technological fundamentals and applications. Springer Science & Business Media.

Feulner, J., Zhou, S.K., Angelopoulou, E., Seifert, S., Cavallaro, A., Hornegger, J., Comaniciu, D., 2011. Comparing axial ct slices in quantized n-dimensional surf descriptor space to estimate the visible body region. Computerized Medical Imaging and Graphics 35, 227–236.

Filipczuk, P., Fevens, T., Krzyżak, A., Monczak, R., 2013. Computer-aided breast cancer diagnosis based on the analysis of cytological images of fine needle biopsies. IEEE Transactions on Medical Imaging 32, 2169–2178.

Fischl, B., 2012. Freesurfer. Neuroimage 62, 774–781.

Foncubierta-Rodríguez, A., García Seco de Herrera, A., Müller, H., 2013. Medical image retrieval using bag of meaningful visual words: unsupervised visual vocabulary pruning with plsa, in: Proceedings of the 1st ACM International Workshop on Multimedia Indexing and Information Retrieval for Healthcare, ACM. pp. 75–82.

Foncubierta Rodríguez, A., Müller, H., 2012. Ground truth generation in medical imaging: a crowdsourcing-based iterative approach, in: Proceedings of the ACM multimedia 2012 workshop on Crowdsourcing for multimedia, ACM. pp. 9–14.

Foran, D.J., Yang, L., Chen, W., Hu, J., Goodell, L.A., Reiss, M., Wang, F., Kurc, T., Pan, T., Sharma, A., et al., 2011. Imageminer: a software system for comparative analysis of tissue microarrays using content-based image retrieval, high-performance computing, and grid technology. Journal of the American Medical Informatics Association 18, 403–415.

Fukunaga, K., 2013. Introduction to statistical pattern recognition. Academic press.

Gionis, A., Indyk, P., Motwani, R., et al., 1999. Similarity search in high dimensions via hashing, in: VLDB, pp. 518–529.

Gletsos, M., Mougiakakou, S.G., Matsopoulos, G.K., Nikita, K.S., Nikita, A.S., Kelekis, D., 2003. A computer-aided diagnostic system to characterize ct focal liver lesions: design and optimization of a neural network classifier. IEEE Transactions on Information Technology in Biomedicine 7, 153–162.

Gong, Y., Kumar, S., Verma, V., Lazebnik, S., 2012. Angular quantization-based binary codes for fast similarity search, in: Advances in Neural Information Processing Systems, pp. 1196–1204.

Gong, Y., Lazebnik, S., 2011. Iterative quantization: A procrustean approach to learning binary codes, in: Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE. pp. 817–824.

Gong, Y., Lazebnik, S., Gordo, A., Perronnin, F., 2013. Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 2916–2929.

Gordo, A., Perronnin, F., Gong, Y., Lazebnik, S., 2014. Asymmetric distances for binary embeddings. IEEE Transactions on Pattern Analysis and Machine Intelligence 36, 33–47.

Güld, M.O., Thies, C., Fischer, B., Lehmann, T.M., 2005. Content-based retrieval of medical images by combining global features, in: Workshop of the Cross-Language Evaluation Forum for European Languages, Springer. pp. 702–711.

Gurcan, M.N., Boucheron, L.E., Can, A., Madabhushi, A., Rajpoot, N.M., Yener, B., 2009. Histopathological image analysis: A review. IEEE Reviews in Biomedical Engineering 2, 147–171.

Haas, S., Donner, R., Burner, A., Holzer, M., Langs, G., 2011. Superpixel-based interest points for effective bags of visual words medical image retrieval, in: MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support, Springer. pp. 58–68.

He, J., Kumar, S., Chang, S.F., 2012. On the difficulty of nearest neighbor search, in: Proceedings of the 29th international conference on machine learning (ICML-12), pp. 1127–1134.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 .

García Seco de Herrera, A., Schaer, R., Bromuri, S., Müller, H., 2016. Overview of the ImageCLEF 2016 medical task, in: Working Notes of CLEF 2016 (Cross Language Evaluation Forum), pp. 1–13.

de Herrera, A.G.S., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S., Müller, H., 2013. Overview of the imageclef 2013 medical tasks., in: CLEF (Working Notes).

Hinton, G.E., Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. Science 313, 504–507.

Hofmann, T., 2001. Unsupervised learning by probabilistic latent semantic analysis. Machine Learning 42, 177–196.

Hofmanninger, J., Langs, G., 2015. Mapping visual features to semantic profiles for retrieval in medical imaging, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 457–465.

Hsieh, J., 2009. Computed tomography: principles, design, artifacts, and recent advances, SPIE Bellingham, WA.

Hsu, W., Long, L.R., Antani, S., et al., 2007. Spirs: a framework for content-based image retrieval from large biomedical databases. MedInfo 12, 188–192.

Huang, L.K., Yang, Q., Zheng, W.S., 2013. Online hashing., in: IJCAI, Citeseer.

Hwang, K.H., Lee, H., Choi, D., 2012. Medical image retrieval: past and present. Healthcare Informatics Research 18, 3–9.

ImageCLEF, . Image retrieval task of the conference and labs of the evaluation forum. http://www.imageclef.org/. Accessed October 23, 2016.

Institute, N.C., a. The cancer genome atlas. https://tcga-data.nci.nih.gov/. Accessed October 23, 2016.

Institute, N.C., b. Genomic data commons. https://gdc.cancer.gov/. Accessed October 23, 2016.

Institute, N.C., c. The national lung screening trial (nlst). https://biometry.nci.nih.gov/cdas/nlst/. Accessed March 23, 2017.

IRMA, . Image retrieval in medical applications project. http://ganymed.imib.rwth-aachen.de/irma/. Accessed October 23, 2016.

Jain, P., Kulis, B., Dhillon, I.S., Grauman, K., 2009. Online metric learning and fast similarity search, in: Advances in Neural Information Processing Systems, pp. 761–768.

Jain, P., Kulis, B., Grauman, K., 2008. Fast image search for learned metrics, in: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE. pp. 1–8.

Jiang, M., Zhang, S., Fang, R., Metaxas, D.N., 2015a. Leveraging coupled multi-index for scalable retrieval of mammographic masses, in: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 276–280.

Jiang, M., Zhang, S., Huang, J., Yang, L., Metaxas, D.N., 2015b. Joint kernel-based supervised hashing for scalable histopathological image analysis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 366–373.

Jiang, M., Zhang, S., Huang, J., Yang, L., Metaxas, D.N., 2016a. Scalable histopathological image analysis via supervised hashing with multiple features. Medical Image Analysis 34, 3–12.

Jiang, M., Zhang, S., Li, H., Metaxas, D.N., 2015c. Computer-aided diagnosis of mammographic masses using scalable image retrieval. IEEE Transactions on Biomedical Engineering 62, 783–792.

Jiang, M., Zhang, S., Zheng, Y., Metaxas, D.N., 2016b. Mammographic mass segmentation with online learned shape and appearance priors, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 35–43.

Jimenez-del-Toro, O., Hanbury, A., Langs, G., Foncubierta-Rodríguez, A., Müller, H., 2015. Overview of the VISCERAL Retrieval Benchmark 2015, in: Multimodal Retrieval in the Medical Domain: First International Workshop, MRMD 2015, Vienna, Austria, March 29, 2015, Revised Selected Papers, Springer. pp. 115–123.

Kahn, C.E., Carrino, J.A., Flynn, M.J., Peck, D.J., Horii, S.C., 2007. Dicom and radiology: past, present, and future. Journal of the American College of Radiology 4, 652–657.

Kalpathy-Cramer, J., de Herrera, A.G.S., Demner-Fushman, D., Antani, S., Bedrick, S., Müller, H., 2015. Evaluating performance of biomedical image retrieval systemsan overview of the medical image retrieval task at imageclef 2004–2013. Computerized Medical Imaging and Graphics 39, 55–61.

Kalpathy-Cramer, J., Hersh, W., 2008. Effectiveness of global features for automatic medical image classification and retrieval–the experiences of ohsu at imageclefmed. Pattern Recognition Letters 29, 2032–2038.

Kalpathy-Cramer, J., Müller, H., Bedrick, S., Eggel, I., de Herrera, A.G.S., Tsikrika, T., 2011. Overview of the clef 2011 medical image classification and retrieval tasks., in: CLEF (notebook papers/labs/workshop), pp. 97–112.

Katouzian, A., Angelini, E.D., Carlier, S.G., Suri, J.S., Navab, N., Laine, A.F., 2012. A state-of-the-art review on segmentation algorithms in intravascular ultrasound (ivus) images. IEEE Transactions on Information Technology in Biomedicine 16, 823–834.

Kim, H., El-Khamra, Y., Rodero, I., Jha, S., Parashar, M., 2011. Autonomic management of application workflows on hybrid computing infrastructure. Scientific Programming 19, 75–89.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks, in: Advances in Neural Information Processing Systems, pp. 1097–1105.

Kulis, B., Grauman, K., 2012. Kernelized locality-sensitive hashing. IEEE Transactions on Pattern Analysis and Machine Intelligence 34, 1092–1104.

Kulis, B., Jain, P., Grauman, K., 2009. Fast similarity search for learned metrics. IEEE Transactions on Pattern Analysis and Machine Intelligence 31, 2143–2157.

Kumar, A., Kim, J., Cai, W., Fulham, M., Feng, D., 2013. Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data. Journal of Digital Imaging 26, 1025–1039.

Kumar, V.S., Rutt, B., Kurc, T., Catalyurek, U., Saltz, J., Chow, S., Lamont, S., Martone, M., 2006. Large image correction and warping in a cluster environment, in: SC 2006 Conference, Proceedings of the ACM/IEEE, IEEE. pp. 38–38.

Lan, R., Zhou, Y., 2016. Medical image retrieval via histogram of compressed scattering coefficients. IEEE Journal of Biomedical and Health Informatics .

Langs, G., Hanbury, A., Menze, B., Müller, H., 2012. Visceral: Towards large data in medical imagingchallenges and directions, in: MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support, Springer. pp. 92–98.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86, 2278–2324.

Lehmann, T.M., Gold, M., Thies, C., Fischer, B., Spitzer, K., Keysers, D., Ney, H., Kohnen, M., Schubert, H., Wein, B.B., 2004. Content-based image retrieval in medical applications. Methods of Information in Medicine 43, 354–361.

Leng, C., Wu, J., Cheng, J., Bai, X., Lu, H., 2015. Online sketching hashing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2503–2511.

Lewis, R., 2004. Medical phase contrast x-ray imaging: current status and future prospects. Physics in medicine and biology 49, 3573–3584.

Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D.D., Chen, M., 2014a. Medical image classification with convolutional neural network, in: Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on, IEEE. pp. 844–848.

Li, R., Zhang, W., Suk, H.I., Wang, L., Li, J., Shen, D., Ji, S., 2014b. Deep learning based imaging data completion for improved brain disease diagnosis, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 305–312.

Li, Z., Fang, R., Shen, F., Katouzian, A., Zhang, S., 2017a. Indexing and mining large-scale neuron databases using maximum inner product search. Pattern Recognition 63, 680–688.

Li, Z., Metaxas, D.N., Lu, A., Zhang, S., 2017b. Interactive exploration for continuously expanding neuron databases. Methods 115, 100–109.

Li, Z., Shen, F., Fang, R., Conjeti, S., Katouzian, A., Zhang, S., 2016. Maximum inner product search for morphological retrieval of large-scale neuron data, in: Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on, IEEE. pp. 602–606.

Lichtman, J.W., Conchello, J.A., 2005. Fluorescence microscopy. Nature Methods 2, 910–919.

Lim, J.H., Chevallet, J.P., 2005. Vismed: a visual vocabulary approach for medical image indexing and retrieval, in: Asia Information Retrieval Symposium, Springer. pp. 84–96.

Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollar, P., 2014. Microsoft coco: Common objects in context. arXiv preprint arXiv:1405.0312 .

Lisin, D.A., Mattar, M.A., Blaschko, M.B., Learned-Miller, E.G., Benfield, M.C., 2005. Combining local and global image features for object class recognition, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops, IEEE. pp. 47–47.

Liu, H., Wang, R., Shan, S., Chen, X., 2016a. Deep supervised hashing for fast image retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2064–2072.

Liu, J., Zhang, S., Liu, W., Deng, C., Zheng, Y., Metaxas, D.N., 2016b. Scalable mammogram retrieval using composite anchor graph hashing with iterative quantization. IEEE Transactions on Circuits and Systems for Video Technology PP, 1–11.

Liu, J., Zhang, S., Liu, W., Zhang, X., Metaxas, D.N., 2014a. Scalable mammogram retrieval using anchor graph hashing, in: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 898–901.

Liu, W., Mu, C., Kumar, S., Chang, S.F., 2014b. Discrete graph hashing, in: Advances in Neural Information Processing Systems, pp. 3419–3427.

Liu, W., Wang, J., Ji, R., Jiang, Y.G., Chang, S.F., 2012. Supervised hashing with kernels, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE. pp. 2074–2081.

Liu, W., Wang, J., Kumar, S., Chang, S.F., 2011. Hashing with graphs, in: Proceedings of the 28th International Conference on Machine Learning (ICML-11), pp. 1–8.

Liu, X., He, J., Lang, B., 2014c. Multiple feature kernel hashing for large-scale visual search. Pattern Recognition 47, 748–757.

Long, L.R., Antani, S., Deserno, T.M., Thoma, G.R., 2009. Content-based image retrieval in medicine: retrospective assessment, state of the art, and future directions. International Journal of Healthcare Information Systems and Informatics 4, 1.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–110.

Manjunath, B.S., Ma, W.Y., 1996. Texture features for browsing and retrieval of image data. IEEE Transactions on Pattern Analysis and Machine Intelligence 18, 837–842.

Markonis, D., Holzer, M., Baroz, F., De Castaneda, R.L.R., Boyer, C., Langs, G., Müller, H., 2015a. User-oriented evaluation of a medical image retrieval system for radiologists. International journal of medical informatics 84, 774–783.

Markonis, D., Schaer, R., Eggel, I., Müller, H., Depeursinge, A., 2015b. Using mapreduce for large-scale medical image analysis. arXiv preprint arXiv:1510.06937 .

May, M., 2010. A better lens on disease. Scientific American 302, 74–77.

of Medicine, T.N.L., . Medpix. https://medpix.nlm.nih.gov/home. Accessed October 23, 2016.

Mesbah, S., Conjeti, S., Kumaraswamy, A., Rautenberg, P., Navab, N., Katouzian, A., 2015. Hashing forests for morphological search and retrieval in neuroscientific image databases, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 135–143.

Müller, H., Clough, P., Deselaers, T., Caputo, B., CLEF, I., 2010. Experimental evaluation in visual information retrieval. The Information Retrieval Series 32.

Müller, H., Deserno, T.M., 2010. Content-based medical image retrieval, in: Biomedical Image Processing. Springer, pp. 471–494.

Müller, H., de Herrera, A.G.S., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S., Eggel, I., 2012. Overview of the imageclef 2012 medical image retrieval and classification tasks., in: CLEF (online working notes/labs/workshop), pp. 1–16.

Müller, H., Kalpathy-Cramer, J., Eggel, I., Bedrick, S., Radhouani, S., Bakke, B., Kahn, C.E., Hersh, W., 2009. Overview of the clef 2009 medical image retrieval track, in: Workshop of the Cross-Language Evaluation Forum for European Languages, Springer. pp. 72–84.

Müller, H., Menze, B., Langs, G., Montillo, A., Kelm, M., Zhang, S., Cai, W.T., Metaxas, D., 2014. Overview of the 2014 workshop on medical computer visionalgorithms for big data (mcv 2014), in: International MICCAI Workshop on Medical Computer Vision, Springer. pp. 3–10.

Müller, H., Michoux, N., Bandon, D., Geissbuhler, A., 2004. A review of content-based image retrieval systems in medical applicationsclinical benefits and future directions. International Journal of Medical Informatics 73, 1–23.

Muller, H., Muller, W., Marchand-Maillet, S., Pun, T., Squire, D.M., 2000. Strategies for positive and negative relevance feedback in image retrieval, in: Pattern Recognition, 2000. Proceedings. 15th International Conference on, IEEE. pp. 1043–1046.

Müller, H., Müller, W., Squire, D.M., Marchand-Maillet, S., Pun, T., 2001. Performance evaluation in content-based image retrieval: overview and proposals. Pattern Recognition Letters 22, 593–601.

Muller, H., Rosset, A., Vallee, J.P., Geissbuhler, A., 2004. Comparing features sets for content-based image retrieval in a medical- case database, in: Proceedings of SPIE, pp. 99–109.

Murala, S., Maheshwari, R., Balasubramanian, R., 2012. Directional binary wavelet patterns for biomedical image indexing and retrieval. Journal of Medical Systems 36, 2865–2879.

Nanni, L., Lumini, A., Brahnam, S., 2010. Local binary patterns variants as texture descriptors for medical image analysis. Artificial Intelligence in Medicine 49, 117–125.

NBIA, . National biomedical imaging archive. https://imaging.nci.nih.gov/ncia/login.jsf. Accessed October 23, 2016.

NCBI, . Pubmed central. https://www.ncbi.nlm.nih.gov/pmc/. Accessed March 23, 2017.

NeuroMorpho, . Neuron morphological database. http://neuromorpho.org/. Accessed October 23, 2016.

Nister, D., Stewenius, H., 2006. Scalable recognition with a vocabulary tree, in: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), IEEE. pp. 2161–2168.

Norouzi, M., Fleet, D.J., Salakhutdinov, R.R., 2012. Hamming distance metric learning, in: Advances in Neural Information Processing Systems, pp. 1061–1069.

Ojala, T., Pietikäinen, M., Harwood, D., 1996. A comparative study of texture measures with classification based on featured distributions. Pattern Recognition 29, 51–59.

Oliva, A., Torralba, A., 2001. Modeling the shape of the scene: A holistic representation of the spatial envelope. International Journal of Computer Vision 42, 145–175.

Oliver, A., Freixenet, J., Marti, J., Pérez, E., Pont, J., Denton, E.R., Zwiggelaar, R., 2010. A review of automatic mass detection and segmentation in mammographic images. Medical Image Analysis 14, 87–110.

Peng, H., Ruan, Z., Long, F., Simpson, J.H., Myers, E.W., 2010. V3d enables real-time 3d visualization and quantitative analysis of large-scale biological image data sets. Nature Biotechnology 28, 348–353.

Powers, D.M., 2011. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. Journal of Machine Learning Technologies 2, 37–63.

Qi, X., Wang, D., Rodero, I., Diaz-Montes, J., Gensure, R.H., Xing, F., Zhong, H., Goodell, L., Parashar, M., Foran, D.J., et al., 2014. Content-based histopathology image retrieval using cometcloud. BMC Bioinformatics 15, 1.

Radhouani, S., Kalpathy-Cramer, J., Bedrick, S., Bakke, B., Hersh, W.R., 2009. Multimodal medical image retrieval: Improving precision at imageclef 2009., in: CLEF (Working Notes).

Raginsky, M., Lazebnik, S., 2009. Locality-sensitive binary codes from shift-invariant kernels, in: Advances in Neural Information Processing Systems, pp. 1509–1517.

Rahman, M.M., Bhattacharya, P., Desai, B.C., 2007. A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. IEEE Transactions on Information Technology in Biomedicine 11, 58–69.

Rui, Y., Huang, T.S., Ortega, M., Mehrotra, S., 1998. Relevance feedback: a power tool for interactive content-based image retrieval. IEEE Transactions on Circuits and Systems for Video Technology 8, 644–655.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV) 115, 211–252. doi:10.1007/s11263-015-0816-y.

Sahbi, H., Audibert, J.Y., Keriven, R., 2007. Graph-cut transducers for relevance feedback in content based image retrieval, in: 2007 IEEE 11th International Conference on Computer Vision, IEEE. pp. 1–8.

Salakhutdinov, R., 2015. Learning deep generative models. Annual Review of Statistics and Its Application 2, 361–385.

Salton, G., Buckley, C., 1988. Term-weighting approaches in automatic text retrieval. Information Processing & Management 24, 513–523.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al., 2012. Fiji: an open-source platform for biological-image analysis. Nature Methods 9, 676–682.

Schlegl, T., Ofner, J., Langs, G., 2014. Unsupervised pre-training across image domains improves lung tissue classification, in: International MICCAI Workshop on Medical Computer Vision, Springer. pp. 82–93.

Semedo, D., Magalhães, J., . Novasearch at imageclefmed 2016 subfigure classification task, in: CLEF2016 Working Notes. CEUR Workshop Proceedings, Évora, Portugal, CEUR-WS. org (September 5-8 2016).

Sertel, O., Kong, J., Catalyurek, U.V., Lozanski, G., Saltz, J.H., Gurcan, M.N., 2009. Histopathological image analysis using model-based intermediate representations and color texture: Follicular lymphoma grading. Journal of Signal Processing Systems 55, 169–183.

Shen, D., Wu, G., Suk, H.I., 2016. Deep learning in medical image analysis. Annual Review of Biomedical Engineering 19.

Shen, F., Liu, W., Zhang, S., Yang, Y., Shen, H.T., 2015a. Learning binary codes for maximum inner product search, in: 2015 IEEE International Conference on Computer Vision (ICCV), IEEE. pp. 4148–4156.

Shen, F., Shen, C., Liu, W., Tao Shen, H., 2015b. Supervised discrete hashing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 37–45.

Shen, F., Shen, C., Shi, Q., van den Hengel, A., Tang, Z., Shen, H.T., 2015c. Hashing on nonlinear manifolds. IEEE Transactions on Image Processing 24, 1839–1851.

Shen, F., Shen, C., Shi, Q., Van Den Hengel, A., Tang, Z., 2013. Inductive hashing on manifolds, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1562–1569.

Shen, W., Zhou, M., Yang, F., Yang, C., Tian, J., 2015d. Multi-scale convolutional neural networks for lung nodule classification, in: International Conference on Information Processing in Medical Imaging, Springer. pp. 588–599.

Shin, H.C., Orton, M.R., Collins, D.J., Doran, S.J., Leach, M.O., 2013. Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 1930–1943.

Shin, H.C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE Transactions on Medical Imaging 35, 1285–1298.

Siemens, . Great growth potential for medical imaging systems. http://www.siemens.com/innovation/en/home/pictures-of-the-future/health-and-well-being/medical-imaging-facts-and-forecasts.html. Accessed October 23, 2016.

Siggelkow, S., 2002. Feature histograms for content-based image retrieval. Ph.D. thesis. Universität Freiburg.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

Simpson, M.S., You, D., Rahman, M.M., Demner-Fushman, D., Antani, S., Thoma, G.R., 2012. Iti's participation in the imageclef 2012 medical retrieval and classification tasks., in: CLEF (Online Working Notes/Labs/Workshop).

Sivic, J., Zisserman, A., 2003. Video google: A text retrieval approach to object matching in videos, in: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, IEEE. pp. 1470–1477.

Slichter, C.P., 2013. Principles of magnetic resonance. volume 1. Springer Science & Business Media.

Smolensky, P., 1986. Information processing in dynamical systems: foundations of harmony theory, in: Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol. 1, MIT Press. pp. 194–281.

Society, A.C., 2013. Breast Cancer Facts & Figures 2013-2014. volume 1. American Cancer Society.

Song, L., Liu, X., Ma, L., Zhou, C., Zhao, X., Zhao, Y., 2012. Using hog-lbp features and mmp learning to recognize imaging signs of lung lesions, in: Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on, IEEE. pp. 1–4.

of South Florida, U., . Usf digital mammography homepage. http://marathon.csee.usf.edu/Mammography/. Accessed October 23, 2016.

of Southern California, U., . Alzheimer's disease neuroimaging initiative. http://adni.loni.usc.edu/. Accessed October 23, 2016.

Srivastava, N., Hinton, G.E., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research 15, 1929–1958.

Stricker, M.A., Orengo, M., 1995. Similarity of color images, in: IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology, International Society for Optics and Photonics. pp. 381–392.

Szabo, T.L., 2004. Diagnostic ultrasound imaging: inside out. Academic Press.

Sze-To, A., Tizhoosh, H.R., Wong, A.K., 2016. Binary codes for tagging x-ray images via deep de-noising autoencoders, in: Neural Networks (IJCNN), 2016 International Joint Conference on, IEEE. pp. 2864–2871.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9.

Tabesh, A., Teverovskiy, M., Pang, H.Y., Kumar, V.P., Verbel, D., Kotsianti, A., Saidi, O., 2007. Multifeature prostate cancer diagnosis and gleason grading of histological images. IEEE Transactions on Medical Imaging 26, 1366–1378.

Tamura, H., Mori, S., Yamawaki, T., 1978. Textural features corresponding to visual perception. IEEE Transactions on Systems, Man, and Cybernetics 8, 460–473.

Tang, L.H.Y., Hanka, R., Ip, H.H.S., 1999. A review of intelligent content–based indexing and browsing of medical images. HIJ 5, 40–49.

TCIA, . The cancer imaging archive. http://www.cancerimagingarchive.net/. Accessed October 23, 2016.

Tian, G., Fu, H., Feng, D.D., 2008. Automatic medical image categorization and annotation using lbp and mpeg-7 edge histograms, in: 2008 International Conference on Information Technology and Applications in Biomedicine, IEEE. pp. 51–53.

Jiménez-del Toro, O.A., Cirujeda, P., Cid, Y.D., Müller, H., 2015. Radlex terms and local texture features for multimodal medical case retrieval, in: Multimodal Retrieval in the Medical Domain. Springer, pp. 144–152.

del Toro, O.A.J., Foncubierta-Rodriguez, A., Gómez, M.I.V., Müller, H., Depeursinge, A., 2013. Epileptogenic lesion quantification in mri using contralateral 3d texture comparisons, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 353–360.

Tourassi, G.D., Harrawood, B., Singh, S., Lo, J.Y., Floyd, C.E., 2007. Evaluation of information-theoretic similarity measures for content-based retrieval and detection of masses in mammograms. Medical Physics 34, 140–150.

Trzcinski, T., Lepetit, V., 2012. Efficient discriminative projections for compact binary descriptors, in: European Conference on Computer Vision, Springer. pp. 228–242.

Unay, D., Ekin, A., 2011. Dementia diagnosis using similar and dissimilar retrieval items, in: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, IEEE. pp. 1889–1892.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A., 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of Machine Learning Research 11, 3371–3408.

VISCERAL, . Visual concept extraction challenge in radiology. http://www.visceral.eu/benchmarks/retrieval2-benchmark/. Accessed October 23, 2016.

Wan, J., Wang, D., Hoi, S.C.H., Wu, P., Zhu, J., Zhang, Y., Li, J., 2014. Deep learning for content-based image retrieval: A comprehensive study, in: Proceedings of the 22nd ACM international conference on Multimedia, ACM. pp. 157–166.

Wan, Y., Long, F., Qu, L., Xiao, H., Hawrylycz, M., Myers, E.W., Peng, H., 2015. Blastneuron for automated comparison, retrieval and clustering of 3d neuron morphologies. Neuroinformatics 13, 487–499.

Wang, J., Kumar, S., Chang, S.F., 2012. Semi-supervised hashing for large-scale search. IEEE Transactions on Pattern Analysis and Machine Intelligence 34, 2393–2406.

Wang, J., Li, Y., Zhang, Y., Wang, C., Xie, H., Chen, G., Gao, X., 2011a. Bag-of-features based medical image retrieval via multiple assignment and visual words weighting. IEEE Transactions on Medical Imaging 30, 1996–2011.

Wang, J., Liu, W., Kumar, S., Chang, S.F., 2016. Learning to hash for indexing big dataa survey. Proceedings of the IEEE 104, 34–57.

Wang, J., Xiao, J., Lin, W., Luo, C., 2015. Discriminative and generative vocabulary tree: With application to vein image authentication and recognition. Image and Vision Computing 34, 51–62.

Wang, X., Yang, M., Cour, T., Zhu, S., Yu, K., Han, T.X., 2011b. Contextual weighting for vocabulary tree based image retrieval, in: 2011 International Conference on Computer Vision, IEEE. pp. 209–216.

Weiss, Y., Torralba, A., Fergus, R., 2009. Spectral hashing, in: Advances in Neural Information Processing Systems, pp. 1753–1760.

Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I., 2015. Automatic coronary calcium scoring in cardiac ct angiography using convolutional neural networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 589–596.

Wu, G., Kim, M., Wang, Q., Gao, Y., Liao, S., Shen, D., 2013. Unsupervised deep feature learning for deformable registration of mr brain images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 649–656.

Wu, G., Kim, M., Wang, Q., Munsell, B.C., Shen, D., 2016. Scalable high-performance image registration framework by unsupervised deep feature representations learning. IEEE Transactions on Biomedical Engineering 63, 1505–1516.

Xing, F., Yang, L., 2016. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. IEEE Reviews in Biomedical Engineering 9, 234–263.

Xue, Z., Long, L.R., Antani, S., Jeronimo, J., Thoma, G.R., 2008. A web-accessible content-based cervicographic image retrieval system, in: Medical Imaging, International Society for Optics and Photonics. pp. 691907–691907.

Yu, G., Yuan, J., 2014. Scalable forest hashing for fast similarity search, in: Multimedia and Expo (ICME), 2014 IEEE International Conference on, IEEE. pp. 1–6.

Yu, X., Zhang, S., Liu, B., Zhong, L., Metaxas, D., 2013. Large scale medical image search via unsupervised pca hashing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 393–398.

Zhang, F., Song, Y., Cai, W., Depeursinge, A., Müller, H., 2015a. Usyd/hes-so in the visceral retrieval benchmark, in: Multimodal Retrieval in the Medical Domain. Springer, pp. 139–143.

Zhang, S., Metaxas, D., 2016. Large-scale medical image analytics: Recent methodologies, applications and future directions. Medical Image Analysis 33, 98–101.

Zhang, S., Yang, M., Cour, T., Yu, K., Metaxas, D.N., 2012. Query specific fusion for image retrieval, in: Computer Vision–ECCV 2012. Springer, pp. 660–673.

Zhang, S., Yang, M., Cour, T., Yu, K., Metaxas, D.N., 2015b. Query specific rank fusion for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 37, 803–815.

Zhang, X., Dou, H., Ju, T., Xu, J., Zhang, S., 2016a. Fusing heterogeneous features from stacked sparse autoencoder for histopathological image analysis. IEEE Journal of Biomedical and Health Informatics 20, 1377–1383.

Zhang, X., Liu, W., Dundar, M., Badve, S., Zhang, S., 2015c. Towards large-scale histopathological image analysis: Hashing-based image retrieval. IEEE Transactions on Medical Imaging 34, 496–506.

Zhang, X., Liu, W., Zhang, S., 2014. Mining histopathological images via hashing-based scalable image retrieval, in: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 1111–1114.

Zhang, X., Xing, F., Su, H., Yang, L., Zhang, S., 2015d. High-throughput histopathological image analysis via robust cell segmentation and hashing. Medical Image Analysis 26, 306–315.

Zhang, X., Zhou, F., Lin, Y., Zhang, S., 2016b. Embedding label structures for fine-grained feature representation, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Zhou, J., Feng, C., Liu, X., Tang, J., 2012. A texture features based medical image retrieval system for breast cancer, in: Computing and Convergence Technology (ICCCT), 2012 7th International Conference on, IEEE. pp. 1010–1015.