Journal of Visual Communication and Image Representation 55 (2018) 243–252

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

journal homepage: www.elsevier.com/locate/jvci

Multiple disjoint dictionaries for representation of histopathology images[☆]

Shujin Zhu[a,b,*], Yuehua Li[a], Shivam Kalra[b], H.R. Tizhoosh[b]

[a] School of Electronic & Optical Eng., Nanjing University of Science & Technology, China
[b] Laboratory for Knowledge Inference in Medical Image Analysis (KIMIA Lab), University of Waterloo, Waterloo, Canada

ARTICLE INFO

Keywords:
Image retrieval
Image representation
Histopathology
Wholeslide imaging
Bag-of-words
Dictionary learning
LBP
SVM
Deep learning

ABSTRACT

With the availability of whole-slide imaging in pathology, high-resolution images offer a more convenient disease observation but also require content-based retrieval of large scans. The bag-of-visual-words methodology has shown a high ability to describe the image content for recognition and retrieval purposes. In this work, a variant of the bag-of-visual-words with *multiple* dictionaries for histopathology image classification is proposed and tested on the image dataset *Kimia Path24* with more than 27,000 patches of size 1000×1000 belonging to 24 different classes. Features are extracted from patches and clustered to form multiple codebooks. The histogram intersection approach and support vector machines are exploited to build multiple classifiers. At last, the majority voting determines the final classification for each patch. The experiments demonstrate the superiority of the proposed method for histopathology images that surpasses deep networks, LBP and other BoW results.

1. Introduction

Pathological analysis of both microscopic and gross tissue samples is a common clinical procedure to identify the presence and type of different diseases. The traditional pathology analysis of cancer tissues has been limited to a few variables such as stage, grade and some clinical markers [1]. With the availability of whole-slide imaging (WSI), the tissue samples can be scanned digitally at very high resolutions. The images produced by the sophisticated WSI are not subject to decay (i.e., stains fade over time) and can be investigated by multiple pathologists at the same time. As well, WSI scans can be integrated into existing information systems of hospitals, providing a better observation and helping to learn more about disease variables [1]. However, visual inspection of a highly detailed tissue scan can be very time-consuming and inefficient in the daily practice of pathology laboratories [2]. Thus, developing computer-aided diagnosis (CAD) to enhance the efficiency and to provide a reliable analysis of digital pathology images is becoming increasingly important for expediting the adoption of digital pathology. The study of CAD systems to quantify spatial histopathology structures can be traced back to 1990s. The recent studies have focused on analyzing high-resolution pathology scans due to the emergence of WSI technology [2–5].

Although digital pathology systems are proposed for variable purposes, most of the efforts have been dedicated to cell detection, segmentation, retrieval and classification [6]. The texture and color information is usually used as features for detection and segmentation tasks [7–10]. While for search and classification of pathology images, the features such as scale-invariant feature transform (SIFT) or local binary patterns (LBP) are often used under a content-based image retrieval (CBIR) framework [11–13,5].

For most image classification tasks, the images in the training dataset usually come with different types of shape or texture (or from various classes). When applying the bag-of-words (BoW) methodology, it is believed that some codewords can be more informative and discriminative than others when describing a specific type of image. Thus, if the training dataset is imbalanced, the most observed types of texture samples are likely to be well clustered into corresponding centroids and dominate the codebook formation. In such a way, the less frequent texture samples may not be well represented by the biased codebook during codeword assignment. To address this issue, in this work, a histopathology image retrieval framework is proposed that creates multiple disjoint dictionaries that uses histogram intersection kernel SVM (IKSVM) for classification. This framework helps to build a more expressive and discriminative dictionary to describe the samples from the same class as training ones.We validate our approach on *Kimia Path24* image dataset [14] which is comprised of 24 WSI scans of different tissue textures which yields up to 25,000 to 50,000 patches of size 1000×1000. A total of 1325 patches have been designated for testing. The main contribution of this study is the design of the “multiple BoW” framework with disjoint training that avoids building the

☆ This paper has been recommended for acceptance by Caifeng Shan.

* Corresponding author at: Laboratory for Knowledge Inference in Medical Image Analysis (KIMIA Lab), University of Waterloo, Waterloo, Canada.
E-mail address: tizhoosh@uwaterloo.ca (S. Zhu).
URL: http://kimia.uwaterloo.ca (S. Zhu).

https://doi.org/10.1016/j.jvcir.2018.06.001
Received 9 October 2017; Received in revised form 22 February 2018; Accepted 2 June 2018
Available online 15 June 2018
1047-3203/

biased codebook caused by the imbalanced types of training samples in the BoW methodology.

This paper is organized as follows: Section 2 presents a brief review of relevant literature. Section 3 explains the proposed framework in detail. Experimental results and discussions are provided in Section 4. Section 5 concludes the paper.

2. Related works

This section covers the literature on three major related fields: *i)* image classification in digital pathology, *ii)* the bag-of-words methodology, *iii)* and the LBP descriptor.

2.1. Classification and retrieval of pathology images

Image classification and retrieval is one of the research focuses in digital pathology. Classification and retrieval systems typically use statistical pattern recognition methods and feature extraction [15,16]. For instance, Ozdemir and Demir [6] proposed a digital pathology hybrid classification model based on structural and statistical pattern recognition techniques. Their model locates and characterizes biological structures, and defines the attributed graph which is used for searching and classification. Different from the conventional tissue classification approaches, which quantify tissue deformations by extracting global features, it only extracts features from the identified key regions most relevant for cancer diagnosis. But the hybrid model uses four simple features to quantify textural changes of the key regions, which may be difficult to recognize the complex structure. Unlike the previous structural methods which quantify a tissue considering the spatial distributions of its cell nuclei, Altunbay et al. [17] relies on the use of distributions of multiple tissue components to represent and classify the colon cancer in pathology images.

Numerous pathology image retrieval algorithms have been proposed based on image content, in contrast to text-based search, using features such as color, texture or shape, and more advanced descriptors. Zheng et al. [11] used color histograms, image texture, and Fourier and wavelet coefficients to build a CBIR system for histopathology. It concatenates all these features to a long vector encoding primitive features, which may be unreliable if one of four features is not accurate (the case of pseudo-color histopathology image). Under the framework of CBIR, a kernel-based supervised hashing model that encoded the SIFT-based descriptors (shift invariant feature transform) to binary bits, and a block LBP-based probabilistic latent semantic analysis model have been introduced [5,13]. Multiple clustered instance learning (MCIL) [18], and convolutional neural networks (CNNs) [14,19,20] have also been developed in recent years to develop intelligent CBIR systems for digital pathology.

2.2. Bag of words approach

The bag-of-words (BoW) approach, or *dictionary learning*, was initially designed for text document classification [21]. Expanding the idea on *visual* words, the error of extracted local descriptors can be minimized and a visual analogue of a word can be formed along with an over-complete dictionary (i.e., the codebook) which is then used to describe the input image with the occurrence frequency of “visual” words captured in a histogram [22,23]. The BoW model has proven to be a powerful recognition tool in computer vision. To construct the BoW model, the following steps should be considered [24–28]:

- Locating regions of interest which is either many selected locations (i.e., keypoint detection), or basically all locations (i.e., dense sampling);
- Extraction of local descriptors, or recording raw (unprocessed) pixels over those regions;
- Quantization of descriptors/pixels and codebook construction;
- Counting frequency of word occurrences to form histograms.

The bag-of-words approach has shown a powerful performance for image annotation, classification and retrieval [29–32]. Numerous studies have been reported regarding its applications in medical or biomedical image analysis [33–35] including histopathology [36]. Diamant et al. [37] proposed a task-driven dictionary medical image classification method by encoding the most relevant codewords per task using the mutual entropy information such that the input image histogram encoded by relevant codewords is more compact than the one encoded by the fully generated dictionary. As it uses the pairwise binary SVM classification, $m(m-1)/2$ binary classifiers should be constructed for m training classes, leading to a problem for applications with a large number of classes. A similar idea can be seen in approaches where several sets of sub-dictionaries are extracted from a fully generated dictionary to encode different image types [38]. Besides, the Pearson’s coefficient of variation weighting scheme was also exploited before similarity measurement and final retrieval.

The idea of using multiple dictionaries has been sporadically proposed in different variations. Pyun et al. [39] generated separate codebooks or Gauss mixtures for each texture using the generalized Lloyd algorithm with a minimum discrimination information distortion, and classified the input into the class whose corresponding codebook results in the smallest average distortion. As these codebooks are only trained by the certain types of texture image, they work well when classifying the texture images that are similar to the training samples, but may fail for other types of texture (different from the training) or for images with complex structure. Sujathaa et al. [40] compute several dictionaries from the overlapping subsets randomly selected from all training samples, and concatenate the histograms encoded by each dictionary to a final histogram before feeding it to a k-NN classifier. But this study is only applied on a small dataset (800 training samples and 200 testing samples) to show the superiority of multiple BoW over the traditional BoW, which may be an issue for large scale dataset as the subsets of a large scale dataset contain sufficient samples (textures) that may lead to similar codebooks.

We will provide a novel design of the concept ”multiple dictionaries” which includes not only disjoint training for individual classes but also multiples representations that requires multiple classifications and a majority vote for final decision.

2.3. LBP – local binary patterns

LBP is a simple but powerful texture feature descriptor with the ability to compete with state-of-the-art learning algorithms like deep networks [41,42]. LBP is the particular case of the texture spectrum model and was first introduced by Ojala et al. [43]. It tends to threshold the differences between the center pixel and its neighbors in spatial windows, and considers the result as a binary number that can be converted into an integer and counted to assemble a histogram. The LBP histogram can be seen as a unified approach to the traditionally divergent statistical and structural models of texture analysis [41]. Several LBP versions with invariance capabilities have been proposed for different applications [44–47].

Considering the success of LBP as a handcrafted descriptor, we will compare our proposed approach against LBP particularly because histopathology images do indeed constitute a class of texture patterns.And another low computational cost feature, namely local directional pattern (LDP) [48], is also implemented for comparison in Section 4.

3. Proposed method

In this section, the pathology image retrieval framework based on multiple bag-of-LBP-words is described. The overall structure of the proposed framework is depicted in Fig. 1. Although the color information has been considered to be one of the most expressive visual