Pattern Recognition 48 (2015) 907–917

Contents lists available at ScienceDirect

Pattern Recognition

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

Deep sparse feature selection for computer aided endoscopy diagnosis

Yang Cong [a,b,*], Shuai Wang [a,e], Ji Liu [b], Jun Cao [c], Yunsheng Yang [d], Jiebo Luo [b]

[a] State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, China
[b] Department of Computer Science, University of Rochester, USA
[c] Department of Computer Science, Arizona State University, USA
[d] Chinese PLA General Hospital, China
[e] University of Chinese Academy of Sciences, China

ARTICLE INFO

Article history:
Received 10 April 2014
Received in revised form
2 July 2014
Accepted 10 September 2014
Available online 20 September 2014

Keywords:
Deep sparse
Group sparsity
Feature selection
Computer aided diagnosis
Endoscopy
Image representation

ABSTRACT

In this paper, we develop a computer aided diagnosis algorithm to detect and classify the abnormalities in vision-based endoscopic examination. We focus on analyzing the traditional gastroscope data and help the medical experts improve the accuracy of medical diagnosis with our analysis tool. To achieve this, we first segment the image into superpixels, then extract various color and texture features from them and combine the features into one feature vector to represent the images. This approach is more flexible and accurate than the traditional patch-based image representation. Then we design a novel feature selection model with group sparsity, Deep Sparse SVM (DSSVM) that not only can assign a suitable weight to the feature dimensions like the other traditional feature selection models, but also directly exclude useless features from the feature pool. Thus, our DSSVM model can maintain the accuracy while reducing the computation complexity. Moreover, the image quality is also pre-assessed. For the experiments, we build a new gastroscope dataset with a total of about 3800 images from 1284 volunteers, and conducted various experiments and comparisons with other algorithms to justify the effectiveness and efficiency of our algorithm.

1. Introduction

Nowadays, gastropathy is a common disease. There are about 24,000 stomach cancer occurrences, and about 4 million people are affected by stomach ulcer every year in America alone [1]. In this paper, we focus on various endoscopy lesions in esophagus and stomach, and intend to design a computer aided diagnosis algorithm to detect various esophagopathy and gastropathy abnormalities. Rather than making the final decisions and replacing human experts, our algorithm is intended to be used as a warning/support system to assist the medical experts to reduce their manual work and improve the accuracy of medical diagnosis. As the passive WCE [2–5] with lower resolution cannot control the position and altitude to make a clearly observation especially in stomach, it is only suitable in small intestine and colon but not in stomach. Therefore, we adopt the traditional gastroscope data with higher resolution and allows operations with more flexible.

Image representation is crucial for endoscopy image analysis, and various color and texture features have been designed, as shown in Fig. 1(a). Since no single feature can represent the image properly, multiple features are heuristically combined into one high-dimensional feature vector to complement each other, e.g., four kinds of features are combined in Fig. 1(b). One other problem is difficult to select the most discriminating features, and some useless features are also inevitably combined, which result in lower accuracy and increase complexity. To overcome these disadvantages, some feature selection models [6,7] are designed. For example [8] uses a neural network with feature selection to detect *Helicobacter pylori* infection, [7] designs a two-stage algorithm by first finding useful feature dimensions with sequential forward floating selection (SFFS) and then uses the SVM to train a classifier accordingly, and [6] makes a further step by using the recursive feature elimination based on SVM (SVM-REF) for feature selection. Generally, the idea of the above methods is to assign a greater weight to the more important feature dimensions, e.g., in Fig. 1(c), the deeper the color is, the greater the weight and the more important the corresponding feature dimension will be. However the problem is that the selected dimensions are always distributed in all feature dimensions as in Fig. 1(c), therefore we still need to extract all features in a time consuming way. Moreover, some noisy feature dimensions or units may be assigned with wrong weights as well. To overcome this, we can select the most relevant features and assign proper weights to the important feature dimensions simultaneously. In this way we not only improve the accuracy, but also reduce the computation complexity, because

* Corresponding author.
E-mail addresses: congyang81@gmail.com (Y. Cong), shuaiwang@sia.cn (S. Wang), jliu@cs.rochester.edu (J. Liu), jun.cao.1@asu.edu (J. Cao), sunny301ddc@126.com (Y. Yang), jiebo.luo@cs.rochester.edu (J. Luo).

http://dx.doi.org/10.1016/j.patcog.2014.09.010
0031-3203/

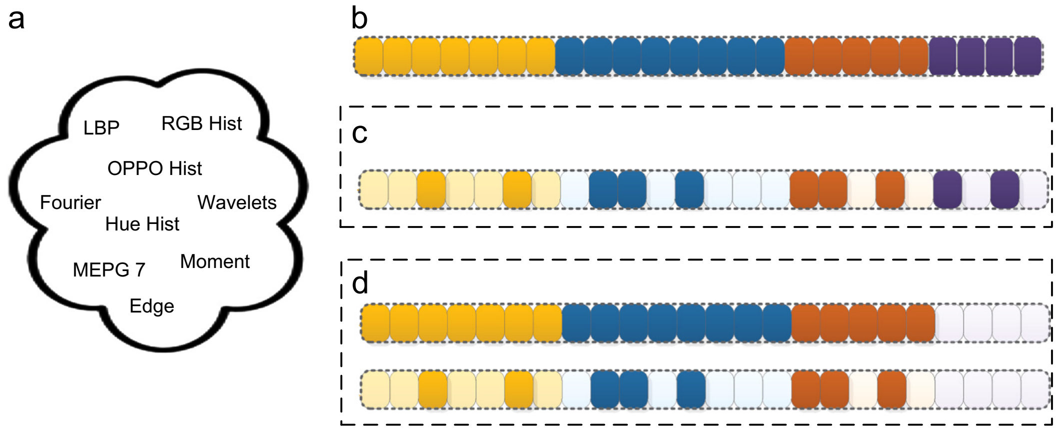

Fig. 1. The demonstration: (a) The pool of original individual feature units, i.e., single feature. (b) The heuristic feature combination, e.g., four kinds of feature units denoted by different colors are combined to complement each other and each feature unit contains several feature dimensions demonstrated by cubic. (c) Assigning a weight to each dimension by most of the traditional feature selection models, i.e., the deeper the color of the cubic is, the more important the corresponding feature dimension will be. (d) Selecting useful feature units and feature dimensions inside concurrently selected feature units by our Deep Sparse SVM (DSSVM) model. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

the useless feature units will not be extracted anymore. As shown in Fig. 1(d), the first three kinds of features are selected and the weights are assigned accordingly at the same time. To achieve this, we take into account the group sparsity of feature units and design a new model, i.e., Deep Sparse SVM (DSSVM). We call our model "deep sparse" because we aim to design a new model, which can not only select the feature units from the original feature set, but also assign proper weights to individual feature dimensions concurrently. In other words, "deep sparse" indicates that our DSSVM can select both feature units and feature dimensions, which are at two differently levels, at the same time.

Another concern in algorithm development is the feature extraction. Most current algorithms extract features from sub-image patches [4,2,9]. Since smaller patch sizes are not able to provide enough information and bigger patch sizes contain too many disturbed pixels, it is hard to decide on the trade-off of suitable sizes. Therefore, we adopt the superpixel method to segment the medical images in a more flexible and adaptive way, and we will show that the superpixel-based feature extraction leads to better results than patch-based one in the experiment. Additionally, since the image quality is often impacted by lighting conditions and internal structure of the human body, we need to assess and discard some regions with poor image quality. In summary, our contributions mainly lie in three aspects:

(i) We propose a new feature selection model via group sparsity, Deep Sparse Support Vector Machine (DSSVM). Our model can not only select the most relevant kind of features from feature set, but can also assign a suitable weight to each feature dimension concurrently.
(ii) A general framework for computer aided endoscopy diagnosis is designed, which adopts the superpixel segmentation method to achieve a more flexible and accurate feature extraction, and it also takes into account the image quality by pre-excluding regions with poor image quality.
(iii) In comparison with the current endoscopy datasets with no more than 3000 images, we collect and build a new dataset including more than 10,000 images from 1284 volunteers, and annotate about 3800 images of them with pixel-level and frame-level groundtruth.

The remainder of the paper is organized as follows. The related works are shown in Section 2. We overview our algorithm in Section 3, then describe the image representation in Section 4 and propose our model DSSVM in Section 5. Then the experiment results and conclusion are presented in Section 6 and Section 7, respectively.

2. Related works

For computer aided digestive endoscopy diagnosis [10,11] give a general review. Depending on the instruments for gastropathy examination, there are mainly two types: (a) the active flexible gastroscopes, including the traditional endoscopy [7], recent narrow-band imaging (NBI) endoscopy [12], zoom-endoscopy [13,14] and confocal laser endomicroscopy (CLE) [15], which are a thin, flexible fibre-optic instrument passed through the mouth to examine the inside of the gullet, stomach and duodenum; (b) the more recent passive and non-invasive technology, Wireless Capsule Endoscopy (WCE) [2–5], widely used for small intestine and gullet examination, which captures and sends out the internal images for diagnosis at rate of 2 fps. Depending on the areas in gastrointestinal tract (GI), the methods can be broken down for the esophagus [16], the stomach [17,7], the small intestine [2–5] and the colon [9,18]. Depending on the specific lesions, the diagnosis methods can be classified to handle bleeding [2], cancer [19,17], Celiac disease, *Helicobacter pylori* [7], polyps [20,14] and ulcers [4], motility assessment [21], tumors [6,7], Barrett's esophagus, Crohn's disease [9,18], and just classify the region into normal and abnormal [22]. Some other applications include detecting informative frames [3], WCE color video segmentation [23], summarization [24] and clustering [25]. In this paper, we intend to detect various esophagopathy and gastropathy abnormalities using traditional gastroscope.

Theoretically, the computer aided endoscopy diagnosis methods have two key technologies: (1) *image representation*, in which some color and texture feature units are extracted for endoscopy image analysis, such as Wavelet feature [14], Gabor feature [12], Fourier feature [13], LBP texture [26], various color histograms [27], edge feature [28], invariant feature [29] and feature combination [4,2]. (2) *Diagnosis/classification model*, in which the state-of-the-art methods [10,11] adopt various models, such as statistical learning models, e.g., Gaussian Mixture Models (GMM); supervised learning models, e.g., SVM [4], Artificial Neural Network (A-NN); the K-Nearest Neighborhood (K-NN) based methods by considering the problem as a retrieval issue. Recently, the feature combination [4,2,9,30,31] by adopting useful features to complement each other prevailed. Furthermore, the feature selection models [6,7] assigning greater weights to more