Neurocomputing 197 (2016) 221–231

Contents lists available at ScienceDirect

Neurocomputing

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

A deep feature based framework for breast masses classification

Zhicheng Jiao, Xinbo Gao*, Ying Wang, Jie Li

Lab of Video and Image Processing Systems, School of Electronic Engineering, Xidian University, Xi'an, China

ARTICLE INFO

Article history:
Received 26 November 2015
Received in revised form
29 January 2016
Accepted 9 February 2016
Available online 17 March 2016

Keywords:
Deep learning
Convolutional neural network
Breast mass classification
Computer-aided diagnosis
Feature visualization

ABSTRACT

Characteristic classification of mass plays a role of vital importance in diagnosis of breast cancer. The existing computer aided diagnosis (CAD) methods used to benefit a lot from low-level or middle-level features which are not that good at the simulation of real diagnostic processes, adding difficulties in improving the classification performance. In this paper, we design a deep feature based framework for breast mass classification task. It mainly contains a convolutional neural network (CNN) and a decision mechanism. Combining intensity information and deep features automatically extracted by the trained CNN from the original image, our proposed method could better simulate the diagnostic procedure operated by doctors and achieved state-of-art performance. In this framework, doctors' global and local impressions left by mass images were represented by deep features extracted from two different layers called high-level and middle-level features. Meanwhile, the original images were regarded as detailed descriptions of the breast mass. Then, classifiers based on features above were used in combination to predict classes of test images. And outcomes of classifiers based on different features were analyzed jointly to determine the types of test images. With the help of two kinds of feature visualization methods, deep features extracted from different layers illustrate effective in classification performance and diagnosis simulation. In addition, our method was applied to DDSM dataset and achieved high accuracy under two objective evaluation measures.

1. Introduction

1.1. Breast mass classification

Being the second cause of death, breast cancer is one of the most common cancers in women. According to a world health organization (WHO) report, breast cancer accounts for 22.9% of diagnosed cancers and 13.7% of cancer related death worldwide [1]. To improve the five-year and ten-year survival rate and to relieve great suffering of patients, the early diagnosis is of crucial importance. Being a process of utilizing low-energy X-rays to examine the human breast, mammography is the most widely used screening and diagnostic tool in both clinical and scientific fields. In order to analyze such an amount of mammograms generated daily in medical centers and hospitals, traditional solution for this challenge is that radiologists have to browse all these images day and night. The next several diagnosis processes also exhaust the physicians, causing the diagnosis to be highly susceptible to errors. This situation also troubles physicians in other fields, and computer-aided diagnosis (CAD) systems have been playing more and more important parts in assisting and improving physicians' work. In previous works, Doi [2] considered that CAD had become one of the major research subjects in medical imaging and diagnostic radiology. Ginneken et al. [3] pointed out that CAD systems were of great help in diagnosis of chest radiography. Jiang et al. [4] and Chan et al. [5] obtained the conclusion that CAD could be used to improve radiologists' performance in breast cancer diagnosis. Identifying benign and malignant masses is among the core contents in diagnosis using mammography. Meanwhile, the building of systems which can effectively assist to do mass classification is one of the hotspots in the mammography related CAD field. Therefore, designing better classification algorithms and frameworks has been attracting more and more attention.

However, as shown in Fig. 1, owing to the diversity in appearance, it is difficult to distinguish the malignant masses from benign ones. A number of researchers and their research teams have been devoted to designing learning and classifying framework to overcome this difficulty. Rangayyan et al. [6] proposed using morphological features to characterize the roughness of tumor boundaries and applied them in classification tasks; Mavroforakis et al. [7] used linear, neural and SVM classifiers to classify masses with the help of textural features and conducted fractal dimension analysis; Timp et al. [8] came up with a novel method exploring the temporal change features among mammography series by regional registration; Rojas-Domínguez et al. [9] performed the analysis of the gradient orientation, fuzziness,

* Corresponding author.
E-mail address: xbgao@mail.xidian.edu.cn (X. Gao).

http://dx.doi.org/10.1016/j.neucom.2016.02.060
0925-2312/

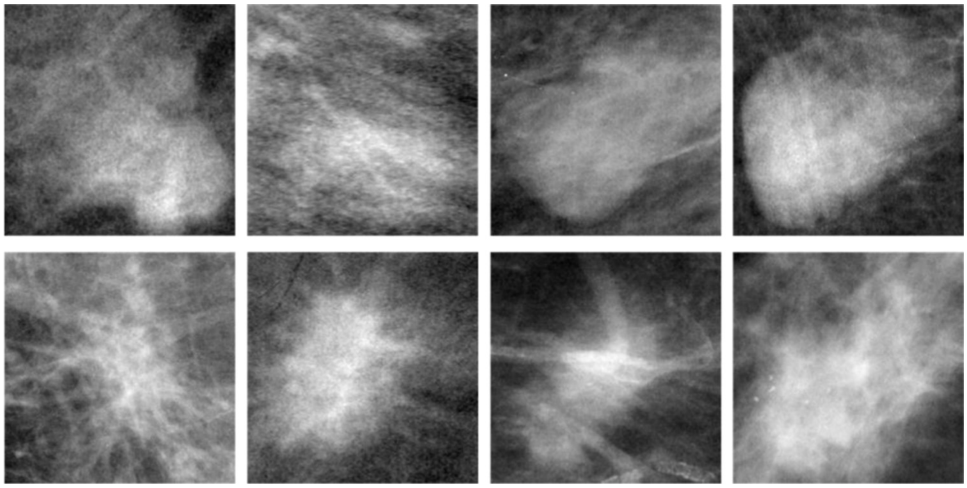

Fig. 1. Examples of benign and malignant breast mass images. Instances in the first row are benign masses while ones in the second row are malignant.

speculation, and mutual information of mass margins; Employing BI-RADS mammographic features with SVM-REF classifier, Yoon et al. [10] achieved a good performance in DDSM; Ramirez-Villegas et al. [11] chose SVM and neural-based classification methods combining Wavelet packet energy, Tsallis entropy and statistical parameterization feature analysis; benefitting more from data structure high accuracy was also reported by Wang et al. [12] in the way of formulating this task into one second order cone programming problem. Verma et al. [13,14] extracted various kinds of features such as density, morphology, abnormity assessment rank, and so on, for description of the masses. Then the soft neural network and soft clustered based direct learning method were employed to do the classification. Wang et al. [15] proposed a latent feature mining based method which characterized spatial and marginal information effectively and achieved good results. More recently, Beura et al. [16] proposed a scheme utilizing 2D-DWT and GLCM in succession to derive feature matrix form mammograms for further classification. Besides, Xie et al. [17] applied extreme learning machine method to improve the performance of mass classification tasks.

Methods which were mentioned above used to extract and utilize low-level features such as margin, texture and so on, or middle-level features such as shape and some variants of bag of words (BOW) [18–20], which has been proven to be effective by Avni et al. [21]. Then the features are introduced into different kinds of classifiers to categorize masses. Although the wide range of traditional handcraft features seem like building a good description of an image, there has been a significant gap existing between these features and cognitive behaviors of physicians. And they do not seem to cover the basic strategies [2] for development of CAD methods and techniques. Strategies aiming at achieving detection and quantitation of lesions in medical images should be based on the understanding of image readings by radiologists. In the real diagnosis process, doctors usually glance at the X-ray first to get preliminary understanding of it. Then several regions that might contain lesions would attract more attention, and the overall look of these regions and details of the entire image would leave impression on and result in different levels of knowledge in the physician's brain. To make the judgement of whether a mass is benign or malignant, doctors used to combine the varying levels of knowledge and awareness with previous experience in similar tasks. The procedure above is similar to that described by the attending doctor we consult from, and it agrees with two representational diagnosis methods: symptom comparisons and anti-diastole. Among these processes, the hierarchical impression of mass images and information processing of human brain are difficult to be specified with traditional features and related methods. Nevertheless, all these things which could only be unspeakable account for a lot in the real diagnosis. So, methods utilizing hierarchical representations and a similar decision mechanism to the real diagnosis may be better choices.

In addition, various types of traditional features can improve classification performance in most situations, but they may have some negative impacts owing to the incompatibility. For example, incompatible extraction methods and corresponding features are less explicable when they are combined directly in a unified framework. However, designing an effective feature fusion strategy is also exhausting in previous papers. On the contrary, hierarchical frameworks could put features extracted from different levels together to form a more explainable and unified structure, avoiding fusing features directly with different models.

1.2. Deep learning and deep learning on biomedical image

From the year of 2006 on, a new machine learning paradigm, named deep learning [22–24], has been playing a much more important role in the academic community. And it has become a huge tide of technology trend in the field of big data and artificial intelligence. Simulating the hierarchical structure of human brain and its data processing mechanism which transfers information from lower level to higher level, deep learning introduces more semantic information to the final representations. Thus deep structure makes significant breakthroughs on image understanding, speech recognition, natural language processing and many other areas [22]. The fever of deep learning has been sweeping the world and has been attracting more attention of top researchers. As results of all these efforts, a few outstanding deep structures are proposed and prove to be successful, such as convolutional neural network (CNN) [25], sparse autoencoder (SAE) [26], restricted boltzmann machine (RBM) [27,28], and so on.

As aforementioned, a convolutional neural network (CNN) is a representative structure of deep models. It is a type of feedforward artificial neural network where individual neurons are tiled in a way that they respond to overlapping regions called receptive fields [47] in the visual field. And it is inspired by biological processes and is a variation of multilayer perceptrons which are designed to reduce preprocessing.

Having been developed for more than three decades, CNN has become an outstanding method. The powerful structures were introduced by Fukushima [29]. CNN was later improved by LeCun et al. [30]. The famous leNet-5 [31] being in form of CNN obtained huge success in recognizing checking numbers. However, given