Neurocomputing 194 (2016) 87–94

Contents lists available at ScienceDirect

Neurocomputing

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

Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset

Jun Shi [a], Shichong Zhou [b,c], Xiao Liu [a], Qi Zhang [a], Minhua Lu [d,*], Tianfu Wang [d]

[a] School of Communication and Information Engineering, Institute of Biomedical Engineering, Shanghai University, Shanghai, China
[b] Department of Ultrasound, Fudan University Shanghai Cancer Center, China
[c] Department of Oncology, Shanghai Medical College, Fudan University, China
[d] National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Department of Biomedical Engineering, School of Medicine, Shenzhen University, China

ARTICLE INFO

Article history:
Received 31 August 2015
Received in revised form
2 December 2015
Accepted 25 January 2016
Communicated by: Yue Gao
Available online 22 February 2016

Keywords:
Stacked deep polynomial network
Deep learning
Ultrasound image
Tumor classification
Texture feature
Small dataset

ABSTRACT

Ultrasound imaging has been widely used for tumor detection and diagnosis. In ultrasound based computer-aided diagnosis, feature representation is a crucial step. In recent years, deep learning (DL) has achieved great success in feature representation learning. However, it generally suffers from the small sample size problem. Since the medical datasets usually have small training samples, texture features are still very commonly used for small ultrasound image datasets. Compared with the commonly used DL algorithms, the newly proposed deep polynomial network (DPN) algorithm not only shows superior performance on large scale data, but also has the potential to learn effective feature representation from a relatively small dataset. In this work, a stacked DPN (S-DPN) algorithm is proposed to further improve the representation performance of the original DPN, and S-DPN is then applied to the task of texture feature learning for ultrasound based tumor classification with small dataset. The task tumor classification is performed on two image dataset, namely the breast B-mode ultrasound dataset and prostate ultrasound elastography dataset. In both cases, experimental results show that S-DPN achieves the best performance with classification accuracies of $92.40 \pm 1.1\%$ and $90.28 \pm 2.78\%$ on breast and prostate ultrasound datasets, respectively. This level of accuracy is significantly superior to all other compared algorithms in this work, including stacked auto-encoder and deep belief network. It suggests that S-DPN can be a strong candidate for the texture feature representation learning on small ultrasound datasets.

1. Introduction

Ultrasound imaging is very commonly used for tumor detection and diagnosis in clinical practice. Ultrasound-based computer-aided diagnosis (CAD) for tumor diseases provides an effective decision support and second opinion tool for radiologists [1].

Feature representation plays an essential role in ultrasound-based CAD [1]. Texture information is among the most commonly used features in ultrasound image [1,2]. Numerous texture feature extraction methods, such as wavelet transform [1,2], multiscale geometric analysis methods [3–5], fractal theory [1,6], and other statistical descriptors [1,2,7,8], have been used for ultrasound images.

As the most popular method for data representation learning, deep learning (DL) has gained its good reputation in computer vision and pattern recognition [9–12]. After DL was first introduced in 2006 by Hinton [14,15], many DL algorithms have been proposed, such as stacked auto-encoder (SAE), restricted Boltzmann machine (RBM), deep belief network (DBN), and convolutional neural network (CNN) [9–13]. DL has been successfully applied to the field of medical image analysis, such as medical image classification, detection and segmentation [15–23].

DL also has been effectively used for ultrasound image processing in recent years. Carneiro et al. proposed a DBN-based algorithm to segment left ventricle of the heart from ultrasound images [24], and they also proposed to combine multiple dynamic models and deep neural network for tracking the left ventricle endocardium in ultrasound image sequences [25]. Wu et al. employed the DBN algorithm for the classification of benign and malignant focal liver lesions based on contrast enhanced ultrasound (CEUS) video [26]. Fasel et al. also used DBN to automatically extract tongue contours on a large scale ultrasound image dataset [27]. Menchon-Lara and Sancho-Gomez applied AE to learn feature for the segmentation of common carotid artery in longitudinal B-mode ultrasound images [28].

* Correspondence to: Department of Biomedical Engineering, School of Medicine, Shenzhen University, Nanhai Avenue 3688, Guangdong 518060, China. Tel.: +86 755 86671915.
E-mail address: luminhua@szu.edu.cn (M. Lu).

http://dx.doi.org/10.1016/j.neucom.2016.01.074
0925-2312/

The great success of DL is based on the fact that a large amount of training samples are required to achieve excellent learning performance, and most commonly used DL algorithms, such as SAE, RBM, DBN and CNN, suffer from the issue of a small sample dataset. However, in the field of machine learning based medical image analysis, the medical datasets usually have small training samples, because labeling is generally expensive and time consuming, which results in limited ground truth data [29,30]. In order to successfully apply DL algorithms to medical images, the local patches from 2D or 3D images, sequence images or video are usually selected as samples so as to enlarge the training set from small dataset with small size subjects [23,24,26].

It is worth noting that the local patch based feature is not the preferred choice for ultrasound image, as it is found that local patch-level features do not always work as well as in other medical imaging modalities. For example, the breast tumor in B-mode ultrasound images reflects a lower level of echoes than surrounding tissues, and appears relatively darker (hypoechoic) [31]. In this case, local patch is only able to provide less information. Therefore, it is the texture features extracted from the whole image rather than the local patch based features, which are very commonly used in ultrasound image, especially for a small dataset. It is still a challenge task to extract and represent discriminative feature from small ultrasound datasets for tumor detection and classification.

The deep polynomial network (DPN) is a novel DL algorithm, which is devoted to learn polynomial predictors with a deep network architecture to provide a good approximate basis for the values attained by all polynomials of bounded degree [32]. DPN does not rely on complicated heuristics, and is easy to be implemented with simple parameters. The experimental results on some commonly used large-scale image datasets show that DPN have comparable or even better performance compared with DBN and SAE algorithms [32]. It is worth noting that the proper way of building the deep networks in DPN allows data representation learning on finite samples in a compact way. From this aspect, DPN has the potential to be applied for learning feature representation from small ultrasound image dataset.

In order to improve the performance of feature representation in DL, it has been observed that layer-wise stacking of feature extraction can yield better representations [10]. Motivated by the successful application of stack method in DL, such as DBN (a stack of RBMs) [12], SAE [33], stacked predictive sparse coding [34], and tensor deep stacking networks [35], we believe that DPN can also be stacked up to build a much deeper structure, because the output layer in the current DPN is the concatenated feature vector of all the features from different layers, which can be further learned to improve the representation ability.

In this work, we propose a stacked DPN (S-DPN) algorithm and then apply it to learn texture feature representation for ultrasound image based tumor classification. The main contributions of this work are twofold: (1) DPN is employed to improve the representation performance of the initially extracted texture feature for small ultrasound dataset and (2) the S-DPN algorithm is proposed to further improve the deep learning ability of the original DPN.

2. Stacked deep polynomial networks

2.1. Deep polynomial network algorithm

The original DPN algorithm is introduced here, whose network architecture is shown in Fig. 1 [32].

Let $X=\{x_1, x_2, \ldots, x_m\} \in R^{m\times d}$ be a set of m training samples, where x_i is a d-dimensional sample. When constructing the first layer network in DPN, the value set from degree-1 polynomial (linear) function over the training samples is given by

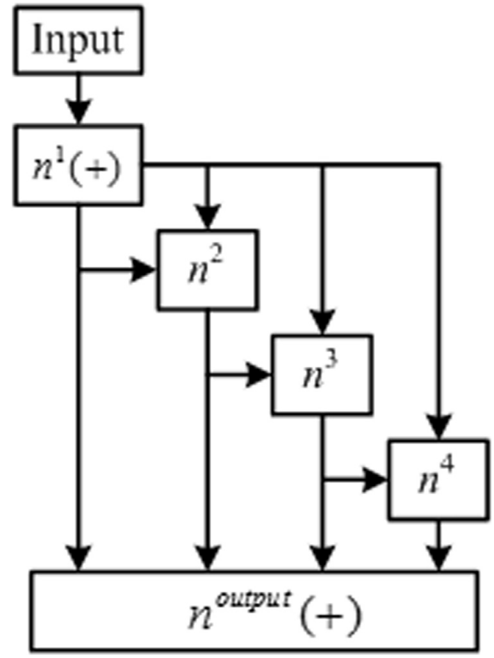

Fig. 1. Schematic diagram of the network's architecture in a degree-4 DPN with each n^i representing a layer of nodes. Here, $(+)$ means a layer of nodes, which calculate functions of the form $n(z)=\sum w_i z_i$, while other layers consist of nodes that compute functions of the form $n(z)=n\left((z_{i(1)}, z_{i(2)})\right)=wz_{i(1)}z_{i(2)}$.

$$\left\{(\langle w[1\ x_1]\rangle, \ldots, \langle w, [1\ x_m]\rangle) : w \in R^{d+1}\right\} \tag{1}$$

which is the $(d+1)$-dimensional linear subspace of R^m. The singular value decomposition (SVD) is then used to build a basis, which thus generate the linear independent set of vectors $\left\{\left(\langle w_j, [1\ x_1]\rangle, \ldots, \langle w_j, [1\ x_m]\rangle\right)\right\}_{j=1}^{d+1}$. Moreover, the consequent linear transformation (specified by a matrix W) will map $[\mathbf{1}\ X]$ into the constructed basis, where $\mathbf{1}$ is the all-ones vector. The columns in W define the d+1 linear functions forming the 1st-layer in DPN, in which the j-th node of first layer is the function by

$$n_j^1(x) = \langle W_j, [1\ X]\rangle \tag{2}$$

where $(n_j^1(x_1), \ldots, n_j^1(x_m))_{j=1}^{d+1}$ forms a basis for all values obtained by degree-1 polynomials over training samples. Define F^1 as a $m\times(d+1)$ matrix, then $F^1_{i,j} = n_j^1(x_i)$ means that the columns in F^1 are the vectors of this set. A one-layer network is then finished, whose outputs span all values by the linear functions on the training samples.

It has been proved that any degree t polynomial can be written as [32]

$$\sum_i g_i(x)h_i(x) + k(x) \tag{3}$$

where $g_i(x)$ is the degree-1 polynomial, $h_i(x)$ is the degree-$(t-1)$ polynomial, and $k(x)$ is a polynomial of degree at most $(t-1)$. The nodes in the 1st-layer network span all degree-1 polynomials, and therefore, they also span the polynomials g_i, h_i, and k. As a result, any degree-2 polynomial can be given by

$$\begin{aligned}&\sum_i\left(\sum_j \alpha_i^{(g_i)} n_j^1(x)\right)\left(\sum_j \alpha_s^{(h_i)} n_s^1(x)\right) + \left(\sum_j \alpha_j^{(k)} n_j^1(x)\right)\\ &\quad = \sum_{j,i} n_j^1(x) n_s^1(x)\left(\sum_i \alpha_j^{(g_i)}\alpha_s^{(h_i)}\right) + \sum_j n_j^1(x)\left(\alpha_j^{(k)}\right)\end{aligned} \tag{4}$$

where α is a scale factor. Eq. (4) indicates that the vector of values obtained by any degree-2 polynomial span the vector of values