ARTICLE IN PRESS
JID: CAEE [m3Gsc;January 19, 2017;19:54]

Computers and Electrical Engineering 000 (2017) 1–12

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Richer feature for image classification with super and sub kernels based on deep convolutional neural network☆

Pengjie Tang[a,b,c], Hanli Wang[a,b,*]

[a] Department of Computer Science and Technology, Tongji University, Shanghai 201804, P. R. China
[b] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 200092, P. R. China
[c] College of Math and Physics, Jinggangshan University, Ji'an 343009, P. R. China

ARTICLE INFO

Article history:
Received 29 January 2016
Revised 13 January 2017
Accepted 13 January 2017
Available online xxx

Keywords:
Deep convolutional neural network
Super convolutional kernel
Sub convolutional kernel
Parallel crossing
Image classification

ABSTRACT

Deep convolutional neural network (DCNN) has obtained great successes for image classification. However, the principle of human visual system (HVS) is not fully investigated and incorporated into the current popular DCNN models. In this work, a novel DCNN model named parallel crossing DCNN (PC–DCNN) is designed to simulate HVS with the concepts of super convolutional kernel and sub convolutional kernel being introduced. Moreover, a multi-scale PC–DCNN (MS-PC-DCNN) framework is designed, with which a batch of PC–DCNN models are deployed and the scores from each PC–DCNN model are fused by weighted average for the final prediction. The experimental results on four public datasets verify the superiority of the proposed model as compared to a number of state-of-the-art models.

1. Introduction

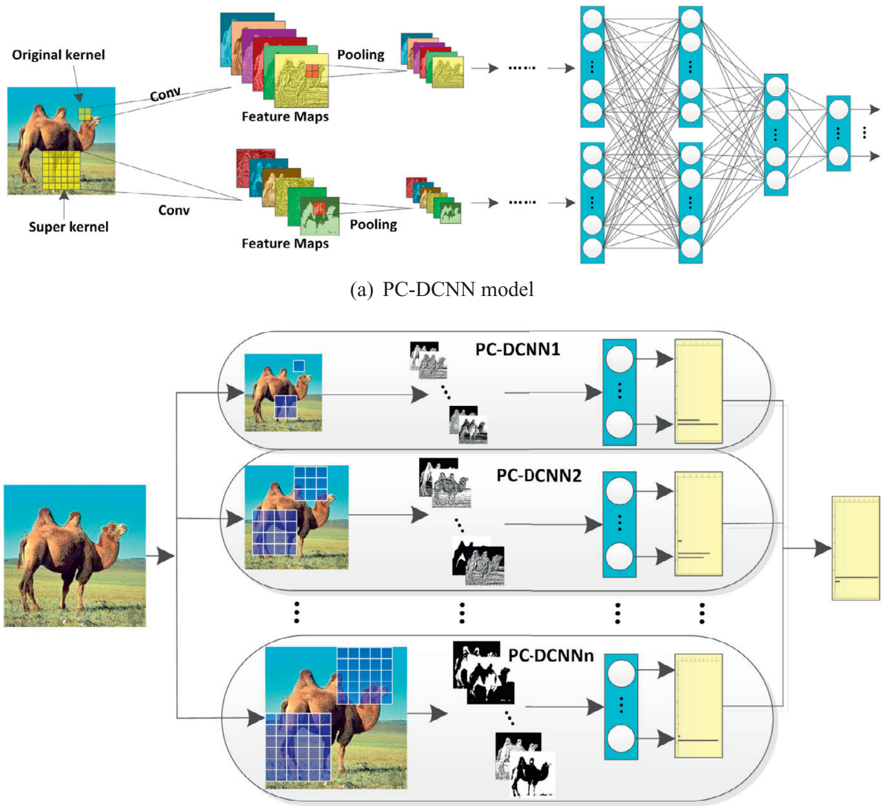
Image classification plays an important role in computer vision. Traditionally, the histogram of orientation gradients (HOG) [1], the scale invariant feature transform (SIFT) [2], and other biological features [3] are extracted from image first. Then, several techniques such as the principle component analysis (PCA) and the linear discriminant analysis (LDA) are usually employed for dimension reduction. Afterwards, the bag of features (BoF) or fisher vector (FV) approach is applied to encode the descriptors as feature vectors. At last, all the feature vectors are fed to a classifier such as the support vector machine (SVM) to predict the image class. Many effective works, such as the spatial pyramid matching (SPM) [4] and sparse coding along with pooling and spatial pyramid matching (Sc^+SPM) [5], have emerged and achieved wonderful performances for image classification. However, these handcrafted features generally possess less semantic and structural information, and the image classification performance can be further improved.

Nowadays, deep convolutional neural network (DCNN) has attracted a lot of research attentions because of its amazing performance [6–8]. It simulates the human visual system (HVS) and the brain multi-level architecture. A number of DCNN models(*e.g.*, Alex-Net [6], VGG16 [8], GoogLeNet [7]) have been designed and obtained astonishing results on a number of visual tasks (*e.g.*, image classification, object detection, human action recognition). In DCNN, the depth is one of the key factors to enhance the discriminative ability of features. Generally speaking, the deeper the model is, the better the

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. M. Senthil Kumar.

* Corresponding author.
E-mail address: hanliwang@tongji.edu.cn (H. Wang).

http://dx.doi.org/10.1016/j.compeleceng.2017.01.011
0045-7906/© 2017 Elsevier Ltd. All rights reserved.



(a) PC-DCNN model

(b) MS-PC-DCNN model

Fig. 1. Overview of the proposed PC–DCNN and MS-PC-DCNN models.

performance becomes. Many research efforts are devoted to optimize the DCNN model architecture such as [9–11]. The works mentioned above achieve great breakthroughs on the task of image classification. However, many of these works pay much more attentions to the model depth, and neglect the fact that with increasing the number of layers, the model complexity is dramatically increased and the performance is not always improved [12]. In addition, most of the current models ignore another fact that human eyes have different visual fields, and the types of information they collect are also different.

As we know, the visual information goes into the brain through two visual pathways, and then it comes into being more comprehensive information via the optic chiasma. The comprehensive information at last is more discriminative and abstract. Therefore, in this work, we simulate human eyes with two types of convolutional kernels with different sizes at the bottom layer; and then, the extracted information forms two streams and is forwarded via each other's pathway. When the features arrive at the top of the proposed model, these two streams will be fused, which is similar to the mechanism of optic chiasma. According to this process, we design a novel architecture called parallel crossing DCNN (PC–DCNN) as shown in Fig. 1(a).

Actually, the ability of human eyes is limited, and most often people have to turn to some optical equipments for more information about the objects they want to understand or recognize. For example, people use telescope for observing the macro features of objects, and they use microscope to get the information of microstructure. The more information about the objects can be used, the higher precision can be obtained. We simulate this process by multi-scale convolutional kernels and develop the multi-scale PC–DCNN (MS-PC-DCNN) model as shown in Fig. 1(b). In MS-PC-DCNN, we use super convolutional kernel to simulate the process of using telescopes, and employ sub convolutional kernel to simulate the process of using microscopes. By this way, a batch of trained DCNN models and a few groups of scores are obtained, then, we fuse all the models' scores by computing their weighted average. The experimental results demonstrate that the proposed models can improve the performance greatly for image classification. Meanwhile, the proposed framework is expandable, so if we employ more and smaller sub convolutional kernels, more PC–DCNN modules can be generated, and the performance can be further improved.

The main contributions of this work include the following three folds. First, inspired by the principle of human vision, we propose a novel model which has a reasonable architecture with low complexity called PC–DCNN for the task of image classification. Second, the concepts of super convolutional kernel and sub convolutional kernel are proposed in accordance

Please cite this article as: P. Tang, H. Wang, Richer feature for image classification with super and sub kernels based on deep convolutional neural network, Computers and Electrical Engineering (2017), http://dx.doi.org/10.1016/j.compeleceng.2017.01.011