Signal Processing 105 (2014) 1–11

Contents lists available at ScienceDirect

Signal Processing

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

Exploring biologically inspired shallow model for visual classification

Tang Tang, Hong Qiao *

Institute of Automation, Chinese Academy of Sciences, State Key Laboratory of Management and Control for Complex Systems, 95 Zhongguancun East Road, Beijing 100190, China

ARTICLE INFO

Article history:
Received 29 January 2014
Accepted 14 April 2014
Available online 30 April 2014

Keywords:
Biologically inspired model
Visual classification
Max-pooling
Shallow model

ABSTRACT

Visual classification has long been a major challenge for computer vision. In recent years, biologically inspired visual models have raised great interests. However, most of the related studies mainly focus on learning features and representations from very large scale dataset relying on deep network architecture, which is doomed to fail with limited training samples due to its high complexity. In this paper, it is found that aside from the deep architecture, two other biologically inspired mechanisms, the pooling and nonlinear operations, also contribute to the improvement of classification performance. Based on this perspective, a new classifier of shallow architecture is proposed, in which the both mechanisms are implemented with max operation. Moreover, the architecture is derived in a probabilistic perspective to further explain the underlying rationale thereof. To train the classifier, a supervised learning algorithm is devised to minimize the hinge loss function under the new architecture. Based on the manifold assumption of continuously transforming features, an unsupervised learning algorithm is also presented to learn the features used by the classifier. Finally, the method is compared against other classifiers on several image classification benchmarks. The results demonstrate the strength of the proposed method when the training data source is limited.

1. Introduction

Classification of visual objects is a long term interest in computer vision field. One of the pioneering contributions in visual classification dates back to the framework developed by Papageorgiou et al. [1]. In this work, they first proposed to combine a Support Vector Machine (SVM) classifier with features of Haar wavelet basis to form a system for object detection. Following this idea, Viola and Jones developed the first face detection system running in realtime [2], leading to the milestone success for face detection, as well as for generic visual classification. Although the techniques introduced by them help to speed up the system greatly, the basic structure of a classifier being appended after Haar feature extraction is still remained, with the only difference being that the SVM is replaced by adaBoost.

The two-component architecture, in which a classifier and a feature extraction process are designed separately, has been continuously evolving and applied to other visual classification tasks since then. To improve the performance of the framework, the design, selection, and organization of features have been widely investigated. This effort leads to the proposal of many important visual features, including Gabor basis [3,4], SIFT descriptor [5,6], Histogram of Oriented Gradients [7,8], Local Binary Patterns [9,10], etc. When a large number of features are extracted from the visual objects with high variation, a clustering algorithm of

* Corresponding author. Tel.: +86 10 82544616.
E-mail address: hong.qiao@ia.ac.cn (H. Qiao).

http://dx.doi.org/10.1016/j.sigpro.2014.04.014
0165-1684/

local patches is usually used to learn a compact codebook of features [11]. In some recently developed visual classification systems, sophisticated coding algorithms also help to further extract discriminative information from the raw features as a post processing step [12,13].

Meanwhile, regarding the classifier, linear classifiers are still the first choice in the most cases for their simplicity and effectiveness. When a large number of visual features carrying enough discriminative information are provided, some visual classification problems can be regarded as being approximately linearly separable. In such a case, a linear SVM can be directly applied even with a small number of training samples.

However, due to the high complexity and variation of the visual objects, hardly any visual classification problems are completely linearly separable. Therefore, theoretically speaking, a non-linear classifier would be more appealing than a linear one for visual classification. K-Nearest Neighbor classifier (k-NN) is perhaps the most simple classifier falling into this category, and several studies have been done to adapt k-NN to visual classification [14,15]. Zhang et al. proposed SVM-KNN algorithm, in which a locally trained SVM is applied in case of the nearest neighbors do not agree in class labels. Boiman et al. proposed the so-called Naive–Bayes Nearest Neighbor classifier (NBNN), in which quantization step in feature extraction process is removed and the image-to-class distance is used instead of image-to-image distance [15]. Random Forest (RF) is another popular choice for non-linear classification [16]. Generally, it is a set of decision trees grown on randomly sampled training sets, in which feature space is split by node from roots to leaves repeatedly. To classify an image, each tree makes a decision independently and votes to decide the predicted label of the sample. Both the k-NN and RF are non-parametric methods for non-linear classification. At the other end, kernel mapping based classifier provides a parametric and elegant solution to the same problem. By applying a proper feature mapping function, original feature space can be mapped into a very high or even infinite dimensional space, on which the complex distributions of different classes can be classified linearly. Owing to the favorable dual formulation of SVM training where explicit feature mapping is omitted, kernel based SVM can be easily realized by simply replacing inner product by the corresponding kernel function. More importantly, the kernel trick provides an easy yet very effective way of embedding prior knowledge into the classifier design. Following this idea, many kernels specialized for visual classification problem have been proposed and significant improvement in visual classification has been made then [17–19]. Nevertheless, the kernel based methods seem not to be the thorough solution to the problem. In one hand, restricted by the form of combining raw features or kernel mapped features linearly, these classifiers do not possess natural invariance to typical variations of visual objects, such as changes in position, illumination and orientation. In the other hand, designing features or kernels of complex invariance has been proved a very challenging task, if not impossible, especially when the visual objects appear in the cluttered real world surroundings.

In recent years, biologically inspired visual models attract much attention from the researchers by promising new insights to improve current image classification systems [20]. With inherently embedded mechanisms inspired by the neural systems, these models are well adapted to spatial correlation and compositional structure of visual stimuli. Furthermore, the flexibility brought by these mechanisms allow the biologically inspired models to learn more and more complex invariance along the hierarchy composed of repeatedly stacked blocks. When being properly configured and trained on the large dataset, the biologically inspired model can output more meaningful representation than that based on hand-crafted features.

In spite of the big success achieved by the biologically inspired models, limitations still exist in the studies and applications of such methods. First, most previous works mainly focused on how to improve the performance with very deep architectures [20,23,27], while other important biologically inspired mechanisms employed are only considered as accessories, of which the detailed analysis and discussion are often missing. Second, related studies are generally driven by the benefits of biologically inspired models to the visual feature extraction and representation, while the discriminative model on top of the system is largely neglected. In the biologically inspired visual systems completely trained in a supervised way, the final decision is usually made by a Multi-Layer Perceptron (MLP) [21,22], which only works well with a large size training set. When the available training samples are limited, the deep learning models suffer from over-fitting problem severely due to their high complexity, and a generic classifier is more preferable to a discriminatively trained neural network in this case [24,23].

This paper proposes a scheme of improving visual classifier design with biologically inspired mechanisms other than deep architecture. Actually, the new classifier can be viewed as a special case of typical deep neural networks by restricting the structure to two adjacent pooling and matching layers. The new structure differs from commonly used classifiers with two features borrowed from biologically inspired models, i.e., non-linear transformation and lateral inhibition, which are both implemented with max operation. The classifier is then interpreted in a probabilistic perspective to further explain the rationale underlying the work. The new structure also raises the question of how to learn the model properly. To sufficiently take advantage of the structure, both the training algorithm of the classifier and unsupervised learning algorithm of visual codebooks are specially designed by considering new features of the structure and important concepts of commonly used classifiers simultaneously. The proposed classifier is compared against other commonly used ones on several visual classification benchmarks. Experimental results show that the introduced mechanisms produce meaningful improvement in performance.

The remainder of this paper is organized as following. In Section 2, several biologically inspired models and the mechanisms motivating this work are briefly introduced and discussed. In Section 3, the formulation of proposed classifier is given and analyzed in the probabilistic perspective of visual classification. In Section 4, the learning algorithms of the system are presented. In Section 5,