Neurocomputing 187 (2016) 19–26

Contents lists available at ScienceDirect

Neurocomputing

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

# Face detection using representation learning

Shu Zhan *, Qin-Qin Tao, Xiao-Hong Li

*Hefei University of Technology, Hefei, China*

ARTICLE INFO

*Article history:*
Received 1 May 2015
Received in revised form
29 July 2015
Accepted 30 July 2015
Available online 3 December 2015

*Keywords:*
Face detection
Convolutional neural network
Deep learning
Support vector machine
Adaboost

ABSTRACT

Face representation is a crucial step of face detection system. In this paper, we present a fast face detection algorithm based on representation learnt using convolutional neural network (CNN) so as to explicitly capture various latent facial features. Firstly, in order to improve the speed of detection in the system, we train an Adaboost background filter which can remove the background most quickly. Secondly, we use the CNN to extract more distinctive features for those face and non-face patterns that have not been filtered by Adaboost. CNN can automatically learn and synthesize a problem-specific feature extractor from a training set, without making any assumptions or using any hand-made design concerning the features to extract or the areas of the face pattern to analyze. Finally, support vector machines (SVM) are used to detect instead of using the classification function of CNN itself. Extensive experiments demonstrate the robustness and efficiency of our system by comparing it with several popular face detection algorithms on the widely used CMU+MIT frontal face dataset and FDDB dataset.


## 1. Introduction

Face detection is the foundation of computer vision and pattern recognition technology [1,2]. It plays an important role in the face recognition, facial point detection, facial expression analysis and other topics [3]. However, because of the illumination, head pose, partial occlusion, facial expressions and other reasons, the face detection problem remains a challenge.

The first step in face detection system is to represent the face images as feature vectors. After obtaining the representation, various learning algorithms can be applied to perform the classification task [4]. Therefore, the performance of face detection algorithm mainly depends on the selected features. As for features, many studies proposed numerous hand-crafted features. The encoding methods of these hand-crafted features are designed manually based on the prior knowledge of face images (e.g., LBP or SIFT). For example, after Viola and Jones [5] proposed the first real-time face detector, Haar-like features have been adopted as the standard feature representation for face detection. Ahonen et al. [6] proposed to use the LBP features to describe the microscopic structure of the face. In addition to using a single feature, many researches use heterogeneous feature types come together to describe the human face [7–9]: Pan et al. [7] used heterogeneous feature types, including Haar feature, LBP feature, SURF feature, to represent face patterns from various aspects, which greatly improves the performance.

These articles all used hand-crafted characteristics to represent the human face; although these features also achieved good results, considerable room for improvement still exists. On one hand, Chen et al.'s [10] experiments showed that most hand-crafted features only gave similar results under the high-dimensional learning framework. It claimed that traditional hand-crafted representations suffered from a visible performance bottleneck and most of them were making different tradeoffs between discriminative ability and robustness. On the other hand, manually acquiring the optimal feature from data is very difficult. To avoid the drawbacks of handcrafted encoding methods, a lot of deep learning algorithms start to look for a new type of feature. For example, CNN could be employed to obtain simple and effective facial features [11,12].

In this paper, we propose a novel and effective face detection system based on CNN learning facial features automatically. First, we train an Adaboost classifier which can roughly find the position of faces, filter most background regions quickly, and consequently improve the detection speed of the system. Then, a feature extractor called CNN is trained to learn and extract features automatically. Making use of the obtained features, we train a SVM classifier for the final classification. The powerful and complex SVM makes the classification better than the CNN itself, which can carefully remove those remaining complex non-face patterns that cannot be ejected by Adaboost.

The rest of this paper is organized as follows: Section 2 presents an overview of some popular techniques applied to face detection. Section 3 describes the proposed face detection system. This is followed by the experimental results and performance

* Corresponding author.

http://dx.doi.org/10.1016/j.neucom.2015.07.130
0925-2312/

analysis, presented in Section 4. Finally, Section 5 presents concluding remarks.

## 2. Related works

There are some significant previous studies about face detection. These studies can be grouped into four categories: knowledge-based methods, feature invariant methods, template matching methods, and appearance-based methods. Among various face detection approaches, appearance-based methods are able to learn distinctive face characteristics, so these methods have attracted much attention. In appearance-based face detection methods, the general practice is to collect a large set of face and non-face examples, and adopt certain machine learning techniques to learn a face model for classification. The key issues are what feature to extract and what learning algorithm to apply.

The boosting cascade framework by Viola and Jones [5] is a milestone in face detection. The amazing real-time speed and high detection accuracy of the face detector can be attributed to three factors: the integral image representation, the cascade framework, and the use of Adaboost to train cascade nodes. But it still has several limitations: First, the number of Haar features is too large, which is usually in hundreds of thousands level for a typical $20 \times 20$ sample. Selecting several effective weak classifiers takes a long time in so many features. Second, the feature representation capacity of Haar feature is very limited. It cannot well handle viewpoint, pose and illumination variations. Li et al. [13] proposed a boosting cascade based face detection framework using SURF features to outperform Viola and Jones' work. SURF feature is more distinctive and the number is smaller, so that the feature selection time is shortened and the performance is improved. Shih et al. [14] presented a novel face detection method by applying discriminating feature analysis (DFA) and SVM. DFA derived a discriminating feature vector by combining the input image, its 1-D Haar wavelet representation, and its amplitude projections. In addition to the above hand-crafted features, there are some learning-based features. For example, unlike many previous manually encoding methods, Cao et al. [15] used unsupervised learning techniques to learn an encoder from the training examples. And then they applied PCA to get a compact face descriptor. Although this scheme upgrades the performance, the careful tuning of each individual module is very labor-intensive. More important, it is unclear how to ensure the performance of the whole system by optimizing each module individually.

Usually, how to use these features to achieve best performance is a process to constantly correct errors and regulate parameters. In addition, these features are usually effective only when they are high-dimensional. And the algorithm is relatively complicated. So in order to extract effective features simply, CNN began to be widely studied. In fact, before the Viola and Jones's [5] detector was published, neural network had been a very popular approach and achieved state-of-the-art performance at that time [16]. Garcia et al. [11] applied CNN to face detection. CNN performed self-driven feature extraction and classification of the extracted features in a single integrated scheme. Chen et al. [17] added a pre-processing step and a single convolutional feature map based on Garcia's work, which can quickly filter more than 75% of backgrounds. The rest of the complex patterns were passed to CNN to deal with. Tivive et al. [18] applied Shunting inhibitory convolutional networks to face detection, which used shunting inhibitory neurons as feature detectors. It showed the proposed face detector based on a hierarchical neural network that can classify in-plane rotated faces in an image, regardless of their orientation. A few research works have been reported to apply CNN on face related problems. For instance, Zhang et al. [19] built a CNN that can simultaneously learn face/non-face decision, the face pose estimation problem, and the facial landmark localization problem. Sun et al. [12] used CNN to extract the global high-level characteristics and detect facial landmarks.

About the research on learning algorithm, AdaBoost has been proven to be an effective algorithm in the area of face detection since the milestone work of Viola and Jones. After that, variants of AdaBoost are proposed for improving the performance of face detector, such as OtBoost [20] . Recently, SVM is an effective classifier with high accuracy, which is commonly-adopted. It can determine the best discriminative support vectors for face detection and classification. However, the detection could not be executed in real time when only a single SVM-based detector was used. So, Pan et al. [7] adopted a coarse-to-fine classifier: in early stage of the system, it employed GH features to remove simple non-face patterns as soon as possible. In the middle stage, MB-LBP descriptors were applied to filter out as many as non-face patterns efficiently. More discriminative and slower SVM classifier used SURF descriptors performing the final detection in the last stage of cascade classifiers to separate face patterns from the remaining difficult non-face patterns that are similar to each other. Base on the above method, we also adopt a coarse-to-fine classifier. Firstly, Adaboost removes most backgrounds quickly. Then only a small part of candidate faces are passed to the CNN and SVM. This framework can ensure the speed of the whole system and the detection rate at the same time.

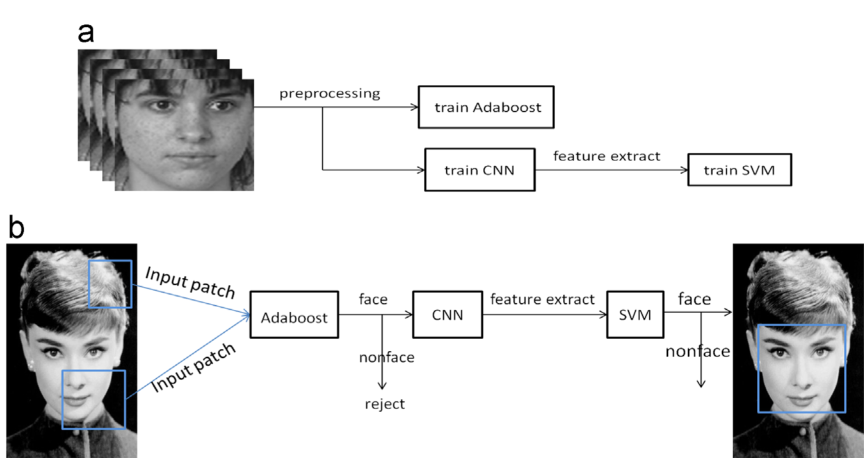


**Fig. 1.** (a) Training of face detection framework and (b) testing of face detection framework.