Neurocomputing 297 (2018) 22–32

Contents lists available at ScienceDirect

Neurocomputing

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

Facial landmark detection by semi-supervised deep learning

Xin Tang, Fang Guo, Jianbing Shen*, Tianyuan Du

Beijing Laboratory of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing 100081, PR China

ARTICLE INFO

Article history:
Received 27 September 2017
Revised 8 January 2018
Accepted 22 January 2018
Available online 10 February 2018

Communicated by Dr Jiwen Lu

Keywords:
Facial landmark
Semi-supervised
Convolutional neural network

ABSTRACT

In this paper, we propose a semi-supervised facial landmark detection algorithm (SEMI) based on convolutional neural network (CNN), which can detect facial components and landmarks simultaneously. Unlike previous coarse-to-fine algorithms, our model does not need extra input such as initial landmark prediction. It also solves the occlusion problem of large area by detecting the visible facial components while existing face detectors failed to detect faces. Semi-supervised learning algorithm is also an effective data augmentation method. In our experiment, each image has two types of ground truth, one is bounding-box related (classification and coordinates) and the other is landmark coordinates inside the bounding-box. The supervised data have both two types of ground truth while the semi-supervised data only have the bounding-box. Our model was trained by the merge of two parts of data. Extensive evaluations on Helen, LFPW and 300-W show that our algorithm is able to complete the landmark task and performs better than many state-of-the-art facial landmark detecting algorithms.

1. Introduction

Facial landmark detection, or face alignment is a significant procedure in many face analysis tasks, such as face recognition [1,4,5], face reconstruction [6,7] and facial attributes analysis [2,3,8,9,58], such as age, gender, smile and wearing-glasses. Since deep learning algorithm especially CNN made a great progress for various computer vision applications [15,28,35,44,45,52,53,55,59], the accuracy of landmark detection task also has the significant improvement. However, the task of detecting landmarks with severe occlusion or extreme expressions are still a difficult challenge.

Traditional landmark detection algorithms can be categorized into two types, one typical type is the template fitting approaches [10–12] and the other is regression-based methods [13,14,16,17]. Though these methods perform well in laboratory environment, they get in trouble when detect facial landmarks in real world because of the extreme poses, lightings and occlusions. In recent years, facial landmark detection algorithms based on deep learning [8,18,19,21–23] had achieved remarkable results. Sun et al. [19] proposed a cascaded convolutional network model with 23 CNN models in total, which requires huge compute complexity while training and testing. Zhou et al. [21] proposed a coarse-to-fine framework with multi-level network models and each model has a different task such as detecting face, generating initial landmark estimation, rotating landmarks to fit the face and so on. Kumar et al. [23] also proposed a coarse-to-fine framework where the input is not only raw pixels but also a set of given landmarks. They train their model four times with input images at different scales, and these algorithms either have high computing complexity.

We believe that human detect facial landmarks, they neither use template fitting approaches, which means put an initial estimation on the test face and fine-tune it, nor regression-based methods, which means detect landmarks on the face simultaneously. Our algorithm first detects facial components (*i.e.* eyebrow, eye, nose and mouth) firstly, then estimate landmarks based on the detection result (position and classification) as shown in Fig. 1. We believe that our method is closer to the way how human brains work. We can also treat this method as a multi-task learning algorithm. Zhang et al. [8] proposed a multi-task learning model to detect facial landmarks and estimate head pose, gender and other facial attributes simultaneously. They are also the first to combine facial attributes estimation with landmark task. Zhou et al. [21] also separated landmarks into different classes and fine-tune them in each network model. But their network can not recognize facial components, which means they only use specific model to predict corresponding landmarks without detecting others.

In this paper, we propose a semi-supervised facial landmark detection algorithm, which is implemented by our specifical data format. In our experiment, each image has two types of ground truth, one is bounding-box related (each box's classification and coordinates) and the other is landmark coordinates inside the

* Corresponding author.
E-mail address: shenjianbing@bit.edu.cn (J. Shen).

https://doi.org/10.1016/j.neucom.2018.01.080
0925-2312/© 2018 Published by Elsevier B.V.

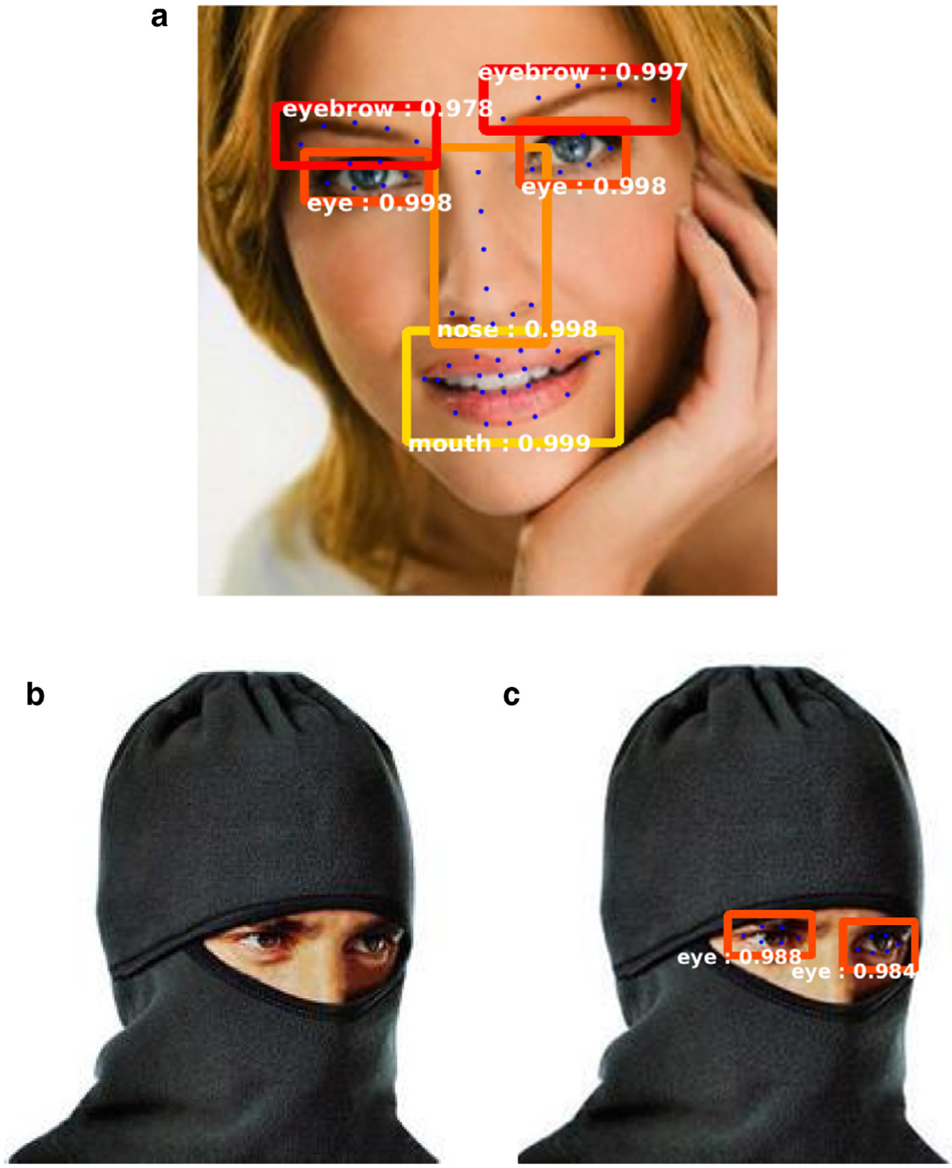

Fig. 1. Illustration of the proposed method. (a) Result by our semi-supervised facial landmark detection method. The text and number pairs denote the probabilities of bounding boxes belong to the corresponding categories. The predicted landmark coordinates are plotted by blue points. (b) An example with severe occlusion face. It is difficult for existing face detectors to find the correct face. (c) The detection result by our algorithm. Our algorithm can detect the visible parts, *i.e.* two eyes with corresponding landmarks plotted in blue. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

bounding-box. The supervised data have two types of ground truth while the semi-supervised data only have the bounding-box. Training data consist of these two parts of data, and our experiment result shows that the semi-supervised part not only improves the accuracy of facial component bounding-box but also enhances the landmarks detection result greatly. Overall, the main contributions of this paper are summarized below:

- We propose a semi-supervised facial landmark detection algorithm, which can augment training data effectively and also brings better detection result.
- Our method can solve large area occlusion problem by detecting the visible facial components while existing face detectors failed to detect the face in the complicated scene. As shown in Fig. 1, our landmark result contains eyebrows and eyes region with accurate prediction.
- Our proposed method is the first attempt to incorporate object detection algorithm with facial landmark detection task under the framework of semi-supervised deep learning.

2. Related work

2.1. Facial landmark detection

Traditional landmark detection algorithms can be roughly divided into two categories, which was named as template fitting approaches and regression-based methods. Template fitting algorithms aim to learn a shape model during training and fit input images when testing. The pioneering works of template fitting algorithms are ASM [10] and AAM [11]. In ASM, the shape of face is represented by the linear combination of basic shapes learning via PCA and appearance of face is modeled by different pre-trained templates. As for AAM, the shape representation is similar with ASM while the appearance is modeled by PCA in regular coordinate system that eliminates shape changes. Zhu et al. [12] show that face detection, facial landmark prediction and pose estimation can be addressed simultaneously.

Usually, regression-based methods estimate landmark locations explicitly by regression using image features. For example,