Analysis of Methods for Facial Landmark Detection

Shiyi Liu[1,a]

[1]College of Food Science & Technology,Shanghai Ocean University,Shanghai, China

[a]18800295167@163.com

Keywords: Component; facial alignment, regression, deep learning

Abstract. In this paper, we propose an analysis of some methods for facial landmark detection (FLD). Recently, facial landmark detection has become a popular topic due to its importance in computer vision area. There has been some novel methods achieving breakthrough in different challenges such as accuracy improvement, detection with severe occlusions and large head poses. These FLD are majorly based on regression process. However, they have their own characteristics in the algorithm by adopting diverse approaches to optimize the regression such as deep learning, Gaussian process, local binary features and so nonexperimental results show all of them can achieve improvements compare to state-of-art methods. With facial point detection, information of human face can be utilized for facial analysis applications, such as face recognition, face synthesis and age estimate.

1. Introduction

Facial landmarks detection (FLD) refers to locate facial landmark points, such as eye comers, nose tip and chin, in a face image [1]. Recently, facial landmarks detection has become a popular topic due to its importance in achieving the goals of various face related applications, such as face recognition, facial expressions classification and age estimation [1]-[3].

As most of FLD methods can achieve a promising result, there are still practicing challenges for FLD applications. The two major challenges are severe occlusion and large head pose. Besides, computer expense and time consumption are evaluated. Last but not the least, high accuracy is the foundation of the efficiency of a method.

FLD methods can be categorized into three major classes, including discriminative fitting, regression based, and deep learning [4]. As for now, regression based method has become a robust, accurate and fast framework for FLD. Based on regression framework, these FLD methods have their own characteristics by combining regression with other methods to improve the performance against different challenges. Regression based framework can be divided into two stages [4]: initialization and iteration. In initialization, it handles the pose prediction .As for the iteration, the solution update those landmarks iteratively. In this paper, we are going to introduce sixes methods which outperform in solving face alignment problem. Ren has proposed a learning based approach utilizing discriminative shape regression. It regulates the learning with a ‘locality’ principle which extracts information from local binary features. Then it concatenates all binary features to learning global linear regression for updating facial landmarks iteratively [5]. Weng proposed a cascaded deep auto-encoder networks. It consists of a global exemplar-based deep auto-encoder network and a series of localized deep auto-encoder networks (LDAN) in a cascaded fashion. The former is used to initialize facial poses and the later updates the pose [4]. Lee proposed a cascaded Gaussian regression trees. He utilizes a kernel function to measure the similarity between two input images in each Gaussian regression tree. The input features are deigned through shape-indexed difference of Gaussian features [6]. Baddar proposed a deep learning regression based methods. It bases on deep convolutional neural networks (DCNN). Baddar utilizes two DCNNs to detection landmarks on facial components and facial contour respectively [7]. Wu proposed a method using a general framework aiming to improve the performance for solving problem with severe occlusions and large

head poses. During the literation, this method considers the interaction of landmark locations and landmarks visibility possibilities [8]. Excepts for regression framework methods, Sagonas proposed a method of statistical frontalization. It is an effective algorithm for robust statistical face frontalization which achieves landmark localization and face frontalization jointly. [9]

The rest of the paper is organized as follows. Section II introduces the regression based methods. Section III describes deep learning approach based on regression process. The statistical frontalization is shown in Section IV. Finally, conclusions are drawn in Section V.

2. Regression Based Methods

In this section, we are going to introduces three of these six methods. They utilize shape regression as their core algorithm. A regression local binary feature is a typical type of regression based method. It regulates the learning procedure through a series of local binary features. Robust cascaded regression framework can handle severe occlusion and large head pose in the meantime. Cascade Gaussian Process Regression Trees combines cascaded regression trees with Gaussian process to improve the accuracy of FLD.

A. Face Alignment via Regression Local Binary Features

This method is a kind of discriminative shape regression approach. It predicts facial shape in a cascaded manner of two learning processes: local binary features and global linear regression.

The shape regression approach predicts facial shape S in a cascaded manner. Beginning with an initial shape S^0, S is progressively refined by estimating a shape increment $\triangle S$ stage-by-stage. In a generic form, a shape increment $\triangle S^t$ at stage t is regressed as:

$$\triangle S^t = W^t \Phi^t \left(I, S^{t-1} \right) \tag{1}$$

where I is the input image, S^{t-1} is the shape from the previous stage, Φ^t is a feature mapping function, and W^t is a linear regression matrix. Note that Φ^t depends on both I and S^{t-1}.

The feature mapping function is essential in shape regression, by which binary features are induced in the corresponding region. The binary features encode the intrinsic structure in a local region, for predicating the landmark position. After concatenating all local binary features to form the feature mapping, it discriminatively learn linear regression matrix for global shape estimation.

To regularize the mapping function, this method proposes two approaches: 1) the function is decomposed into a set of feature mapping functions.2) each feature mapping function predicts the landmark in the corresponding local feature. A standard regression random forest is utilized to learn the feature mapping functions. By experimental results, the radius of local region in each stages gradually shrinks, because the variation of regressed face shapes decreases during the cascade.

Local regression has its advantages: 1) the feature pool in local learning is less noisy. 2) Features learned by local learning have a better estimation for independent landmarks. 3) Local learning is adaptive in different stages (the radius of local region in every stage). Next the global prediction is through a linear regression process based on the binary features. As a result, the accuracy is improved.

B. Unified Robust Cascade Regression Framework

This is a unified robust cascade regression framework targeting to handle both of two challenge: severe occlusion and large head poses. It regards the landmarks on self-occluded facial parts as occluded points, where the face itself is the occlude. Hence, it can consider images with large head poses as special cases of images with occlusion and treat them similarly.

It initializes the locations using the mean face shape and assumes all the points are visible. With this intuition, the algorithm updates the visibility probabilities and the landmark locations across iterations to achieve convergence. When updating the visibility probability, it introduces a supervised regression model based on the information that depends on the current estimated detections. In addition, the occlusion pattern is embedded as a constraint. When updating the landmark locations, we considers the information depend on the image, the current landmark locations and the landmark visibility. Through this interaction between these two elements, this

general framework can handle FLD problems with both severe occlusion and large head poses.

There are two detractions to improve. First, extending the detection algorithm for real-time tracking. Second, improving the algorithm to handle more challenges in real conditions such as illumination change, low resolution etc.

C. Face Alignment using Cascade Gaussian Process Regression Trees

In a shape regression procedure, input shape is initialized and iteratively updated through a cascaded regression trees (CRT). There are two key elements of shape regression that impact for the prediction performance are gradient boosting for learning the CRT and the shape-indexed features which the trees are based. Instead of using gradient boosting. However, overfitting occurs when the fitting rates are not coordinated with prediction. Instead of designing regularization methods, Lee abandons gradient boosting but proposed cascade Gaussian process regression trees (cGPRT), which can be incorporated as a learning method for a CRT prediction framework.

To descending computational complexity of a Gaussian process, Lee used a special kernel function to measure the similarity between two input images. The proposed cGPRT is formed by a cascade of Gaussian process regression trees, each of which considers a kernel function. In the other hand, input features to cGPRT are designed through shape-indexed difference of Gaussian (DoG) features computed on local retinal patterns. The shape-indexed DoG features are extracted in three steps: (1) smoothing face images with Gaussian filters at various scales to reduce noise sensitivity, (2) extracting pixel values from Gaussian-smoothed face images indexed by local retinal sampling patterns, shape estimates, and smoothing scales, and (3) computing the differences of extracted pixel values.

The experimental results show that this method performs a promising generalization with an acceptable computational complexity.

3. Deep Learning Regression Methods

In this section, we introduce two methods utilized deep learning approach based on regression process.

A. Learning Cascaded Deep Auto-Encoder Networks for Face Alignment

This is a new cascaded deep auto-encoder networks (CDAN) approach for FLD.

During the initialization, most of approaches employ the mean shape as the initial pose. This strategy performs poorly, especially to a near-profile pose. As for iteration process, real-time detection is a corner stone that most methods fail to extract information when facing facial variations. As a consequence, this method utilizes a deep learning strategy to model the complex and nonlinear relationship between facial features and face poses, as well as to extract pose-informative features for FLD.

The proposed framework consists of a global exemplar-based deep auto-encoder network (GEDAN) and a series of localized deep auto-encoder networks (LDAN) in a cascaded fashion. In this way, both global facial structural information and local landmark feature information are jointly led to local repressors, exploited for facial landmark detection. Consequently, the accuracy of landmark alignment is improved.

For GEDAN model, it aims to improve the pose estimation accuracy on holistic facial images. It takes a holistic facial images as an input and generates a roughly aligned face configuration. GEDAN incorporates several exemplars at the top layer to form a non-linear regression model, which enlarges the deep auto encoder network's capacity in pose estimation. The hallmark of the GEDAN model lies on the top regression layer which consists of two modules, namely a linear regression module and a nonlinear exemplar-based regression module.

For LDAN model, it aims to improve the alignment accuracy basing on the result of GEDAN. The first layer consists of individual LAEs, each of which aims to extract localized pose-informative features from the corresponding pose-indexed patches. It concatenates the outputs of these LAEs into a global feature vector and then lead global features into the local repressors to impose global facial structure constraints.

Experimental results have shown that this approach consistently outperforms state-of-the-art face alignment methods. Furthermore, it achieves real-time performance with Matlab on a common desktop without special optimization for speed up.

B. Novel Cascaded Deep Convolutional Neural Network (DCNN) Structure

With recent success of deep learning, deep learning regression-based FLD methods have been proposed. This method consists of two deep convolution neural networks: (1) DCNN-C is to detect landmarks constrained on facial contour (2) DCNN-I is to detect landmarks constrained on facial components. Two DCNNs are trained separately. By separately, it improves the detection on facial components. In addition, by learning the landmarks on facial contour with the components jointly, it improves the detection on facial contour. Besides, a novel DCNN structure is proposed for detection on facial components constraints, which branches networks at higher layers in order to capture the intricate local facial components features, and a novel learning strategy to learn the DCNN for detection on facial contour, which exploits the relationship between landmarks on facial contour and components.

1. DCNN-I for facial components

This is a linear regression problem, where a loss function is the total loss from all landmarks in a batch of N images. This regression function assumes that the loss from all the landmarks collectively contributes to learn the hyper parameters of the DCNN and relationships between all landmarks on facial components are considered jointly. The relationships can be characterized by inter-facial components relationship and intra-facial component relationship. In inter-facial components relationship, landmarks on each facial component are constrained by the location of the corresponding facial component. In intra-facial components relationship, facial components variations can affect its own local landmarks but do not directly affect other facial component landmarks. To that end, a novel DCNN structure is composed of shared layers that fork into branches at the high layers dedicated for landmarks on different facial components. The hyper parameters of separate higher layers are learned only by corresponding facial components, making the network more robust to partial occlusion, expression variations. In the meantime, the lower layers maintain detection of the overall shape in facial landmarks.

2. DCNN-C for facial contour

Due to facial background noise, head posed variations or appearance variations, facial contour landmarks are significantly more difficult to detect. A separate DCNN is utilized to detect facial contour in order to avoid affecting DCNN-I for facial components. And learning landmarks for facial contour jointly with facial components can improve the detection, because components can depict the head posed and a rough estimate of the location of facial contour. A network similar in structure to DCNN-I is devised to jointly learn all the facial landmarks and hyper parameters of shared lower layers are transferred to lower layers of DCNN-C. For higher layers, only hyper parameters of facial contour are transferred to DCNN-C.

Experimental results showed that this method improve the accuracy and robustness compared to state-of- the-art methods. However, cascaded DCNNs structure causes a higher computational burden.

4. Statistical Face Frontalization

This is a novel method for joint frontal view reconstruction and landmark localization using a small set of frontal images only.

This method regards a statistical element as a key motivation: for the facial images lying in a linear space, the rank of a frontal facial image, due to the approximately structure of human face, is much smaller than the rank of facial images in other poses. Thus, input image is warped into a reference frontal-pose frame and the nuclear norm was computed. The contributions of this method can be summarized as follows: technical contributions: 1) jointly achieves landmark localization and face frontalization; 2) an effective algorithm for the RSF is developed. Applications in computer vision: 1) this is the first landmark localization method using a model of frontal images only; 2) it

can improve in pose invariant face recognition and unconstrained face verification; 3) it can handle all human faces, cat faces, and face sketches.

5. Conclusion

In this paper, we introduced six FLD methods which have elaborate structure utilizing diverse algorithms. Three of them employ shape regression method and the other two utilizing deep learning method to optimize the regression process. They include local binary features, Gaussian process, cascade framework, DCNN and deep auto-encoder networks. Through different strategies, they improve the results of FLD, especially targeting to some "cornerstone challenges" The last one uses a statistical strategy to achieve face frontalization, which has contributions in both of technical aspect and applications of computer vision.

References

[1] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic, "Robust discriminative response map fitting with constrained local models," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2013, pp. 3444–3451.

[2] P. N. Belhumeur, D. W. Jacobs, D. Kriegman, and N. Kumar, "Localizing parts of faces using a consensus of exemplars," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2011, pp. 545–552.

[3] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., London, U.K., Jun. 2012, pp. 3642–3649.

[4] Weng R, Lu J, Tan Y P, et al. Learning Cascaded Deep Auto-Encoder Networks for Face Alignment[J]. IEEE Transactions on Multimedia, 2016, 18(10):2066-2078.

[5] Ren S, Cao X, Wei Y, et al. Face Alignment at 3000 FPS via Regressing Local Binary Features[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2014:1685-1692.

[6] Lee D, Park H, Chang D Y. Face alignment using cascade Gaussian process regression trees[C]// Computer Vision and Pattern Recognition. IEEE, 2015:4204-4212.

[7] Baddar W J, Son J, Kim D H, et al. A deep facial landmarks detection with facial contour and facial components constraint[C]// IEEE International Conference on Image Processing. IEEE, 2016:3209-3213.

[8] Wu Y, Ji Q. Robust Facial Landmark Detection Under Significant Head Poses and Occlusion[C]// IEEE International Conference on Computer Vision. IEEE, 2016:3658-3666.