COUPLED CASCADE REGRESSION FOR SIMULTANEOUS FACIAL LANDMARK DETECTION AND HEAD POSE ESTIMATION

Chao Gou[*], *Yue Wu*[†], *Fei-Yue Wang*[*], *Qiang Ji* [†]

[*]Institute of Automation, Chinese Academy of Sciences, China
[†] Rensselaer Polytechnic Institute, USA
[*] Qingdao Academy of Intelligent Industries, China

ABSTRACT

Current approaches for facial landmark detection and head pose estimation first perform landmark detection, followed by fitting 3D face model or regression model to estimate head pose. Different from the existing methods, in this paper, we propose a unified method, called Coupled Cascade Regression (CCR), for simultaneous facial landmark detection and head pose estimation. At each cascade level, two separate regressors are learned to update the landmark locations and 3D face model parameters based on the local appearance features, respectively. Since 2D facial landmark locations and head pose parameters are related, we further apply the projection model to refine the prediction results in each cascade iteration and make them consistent. As a result, CCR can leverage both the learning methods and the projection model to simultaneously perform facial landmark detection and pose estimation to enhance the performances of both tasks. Experimental results on 300-W and BU datasets indicate that our proposed CCR method outperforms many conventional methods both for landmark detection and head pose estimation.

Index Terms— Facial landmark detection, head pose estimation, coupled cascade regression

1. INTRODUCTION

Facial landmark detection aims to predict the facial key points (e.g. eye corners, mouth corners and nose tip) in a given image. It is becoming an increasingly important research topic due to its wide applications, such as face recognition, facial action unit recognition, and 3D face reconstruction[1, 2]. Head pose estimation in a given image aims to predict the orientation of the head with respect to the camera coordinate frame. It has been widely used in many scenarios, such as capturing the visual attention of subjects in human-machine intersection, estimating the gaze direction of the driver and analyzing social event interaction [3, 4].

Related work. Recently, cascade regression based methods [5, 6, 7] have achieved impressive performance for facial landmark detection, where they learn regression models at each cascade level to iteratively map the discriminative local features around landmarks to the ground truth landmark positions. Supervised Descent Method (SDM) [5] is one of representative methods. It learns a cascade of liner regressors, each of which further optimizes the facial landmark locations given an initial positions. Head pose estimation approaches in computer vision can be generally categorized into learning-based and model-based methods. The learning-based methods try to map the extracted appearance features (e.g. HOG, SIFT) to 3D head poses (e.g. pitch, yaw and roll). The model-based methods estimate the head pose by linking the 2D observation and 3D face model through the computer vision projection model. These methods typically first perform 2D landmark detection, followed by fitting the 3D face model to estimate the head pose. In [8, 9], head pose is estimated by fitting the related cylindrical models. Some other effective methods [10, 11] decode head pose from 3D face model parameters, which are estimated by minimizing the misalignment error between the ground truth locations and projected locations of 3D face model on the image plane.

By treating the 3D morphable model (3DMM) and corresponding 3D face model parameters as a representation of 2D facial shape [12], some research works [13, 14, 15] proposed to iteratively update the 3D face model parameters for facial landmark detection. Liu *et al.* [14] propose a regression based method to simultaneously detect 2D facial landmarks and reconstruct 3D facial shape. It alternatively optimize the facial landmark locations and 3D facial shape by applying corresponding learned regressors at the same cascade iteration. Typically, 3D face model parameters consist of projection matrix parameters and 3D deformable parameters. The head pose information can be extracted from the projection matrix parameters. Tulyakov and Sebe [16] estimate the 3D facial shape from a single image based on cascade regression framework using the shape invariant features. They extract

This work was performed while the first author visited Rensselaer Polytechnic Institute (RPI), supported by a scholarship from University of Chinese Academy of Sciences (UCAS). This work was also supported in part by National Science Foundation under the grant 1145152 and by the National Natural Science Foundation of China under Grant 61304200 and 61533019.

head pose from the defined face basis vector which denotes the rotation angle of face.

Most of works have been done for facial landmark detection and head pose estimation separately. However, since the landmark positions and head pose are projectively related to each other, they should be tackled jointly. In this paper, we propose a unified framework called Coupled Cascade Regression (CCR). It first leverages the benefits from the cascade regression framework to separately estimate the landmark positions and to estimate head pose. Since 2D facial landmark and 3D pose parameters are related, we then link them through the projection model to further improve their estimations. As a result, CCR can exploit the interactions between landmark points and head pose to perform simultaneous landmark detection and head pose estimation. The major contributions of this work come from three folds:

Simultaneity: By exploiting joint relationships among facial landmarks and head pose, the proposed CCR can simultaneously detect the landmark locations and estimate the head pose. In addition, CCR can improve the performances of both tasks.

Learning-with-model: The proposed method combines learning and model. The power of learning comes from the cascade regression step, while the projection model can leverage the relationship among facial landmark locations and head pose to ensure their consistency.

Effectiveness: The proposed CCR, which applies linear regression models with hand-crafted features, is simple but effective.

2. PROPOSED METHOD

The overall framework of the proposed CCR method that estimates the target facial shape and head pose is shown in Algorithm 1. By leveraging the cascade regression framework, we jointly perform two tasks of landmark detection and 3D face model parameter estimation by learning the cascade regression models separately. It begins with an initialization of mean 3D face model parameters $\mathbf{p}^0$, which can map face to mean landmark locations $\mathbf{x}^0$ by the projection model. For an facial image $\mathbf{I}$, we use $\Phi(\mathbf{x}^t, \mathbf{I}))$ to denote the local appearance features. At each cascade level t, we learn one regerssor R^t for updating the landmark locations $\mathbf{x}^{t*}$ and learn another regressor Q^t for updating the 3D face model parameters $\mathbf{p}^{t*}$ (see *step* 1 and *step* 2 in Algorithm 1). We further refine $\mathbf{p}^{t*}$ and $\mathbf{x}^{t*}$ through the projection model to ensure their consistency (see *step* 3 in Algorithm 1). As a result, they are updated to new values of $\mathbf{p}^t$ and $\mathbf{x}^t$ for next iteration. As a result, CCR can effectively learn the cascaded regressors by incorporating local appearance and the projection model to improve the performance of both landmark detection and head pose estimation. In the following, we describe our proposed Coupled Cascade Regression (CCR) method in details.

Algorithm 1 Coupled cascade regression for landmark detection and head pose estimation.

Input:

Give the facial image $\mathbf{I}$. 68 key point locations $\mathbf{x}^0$ are initialized with mean 3D face model parameters $\mathbf{p}^0$.

Do cascade regression:

for t=1,...,T **do**

step 1: Estimate the landmark location updates $\Delta\mathbf{x}^t$ given the current landmark locations $\mathbf{x}^{t-1}$,

$$\Delta\mathbf{x}^t = R(\Phi(\mathbf{x}^{t-1}, \mathbf{I})),$$

Update the landmark locations,

$$\mathbf{x}^{t*} = \mathbf{x}^{t-1} + \Delta\mathbf{x}^t,$$

step 2: Estimate the parameter updates $\Delta\mathbf{p}^t$ given the current landmark locations $\mathbf{x}^{t-1}$,

$$\Delta\mathbf{p}^t = Q(\Phi(\mathbf{x}^{t-1}, \mathbf{I})),$$

Update the parameters,

$$\mathbf{p}^{t*} = \mathbf{p}^{t-1} + \Delta\mathbf{p}^t,$$

step 3: Make $\mathbf{p}^t$ and $\mathbf{x}^t$ consistent to be corresponding to each other based on projection model, with initialization of $\mathbf{p}^{t*}$ and $\mathbf{x}^{t*}$,

$$\mathbf{x}^t, \mathbf{p}^t = arg\min_{\mathbf{x},\mathbf{p}} \varepsilon(\mathbf{x}, \mathbf{p}).$$

end for

Output:

Acquire the landmark locations $\mathbf{x}^T$ and head pose.

2.1. Update the Landmark Locations

Different regression functions (e.g. linear regression model, random forest) can be applied in the general cascade regression framework. One of most widely used models is liner regression model [5], which is effective and efficient. In this work, we use linear model in general cascade regression framework for landmark detection. The regressor R^t for predicting the updates of landmark location is :

$$\Delta\mathbf{x}^t = R^t(\Phi(\mathbf{x}^{t-1}, \mathbf{I})) = \mathbf{R}^t\Phi(\mathbf{x}^{t-1}, \mathbf{I}) + \mathbf{b}^t, \qquad (1)$$

where $\mathbf{R}^t$ and $\mathbf{b}^t$ are the parameters of cascade regressor R^t at iteration t.

For training at each cascade level t, given K training facial images with ground truth landmark locations $\mathbf{x}^*$, the ground truth updates of landmark location $\Delta\mathbf{x}_i^{t,*}$ in *i*-th image can be acquired by subtracting the current landmark locations $\mathbf{x}_i^{t-1}$ from the ground truth landmark locations $\mathbf{x}_i^*$. In addition, given the training images with estimated key point locations $\mathbf{x}^{t-1}$, the local appearance features $\Phi(\mathbf{x}_i^{t-1}, I_i)$ of i-th image can be calculated. Then we can learn the corresponding liner model parameters $\mathbf{R}^t$ and $\mathbf{b}^t$ for updates of landmark location by the standard least-square formation with closed form solution:

$$\mathbf{R}^{t^*}, \mathbf{b}^{t^*} = arg \min_{\mathbf{R}^t, \mathbf{b}^t} \sum_{i=1}^{K} \parallel \Delta \mathbf{x}_i^{t,*} - \mathbf{R}^t \Phi(\mathbf{x}_i^{t-1}, I_i) - \mathbf{b}_i^t \parallel^2 . \tag{2}$$

For landmark location at iteration t, given the facial image with current landmark positions $\mathbf{x}^{t-1}$ and the learned regressor with parameters $\mathbf{R}^t$ and $\mathbf{b}^t$, we can estimate the updates $\Delta\mathbf{x}^t$ of landmark locations by Eq. 1. Then we add the updates $\Delta\mathbf{x}^t$ to current landmark locations $\mathbf{x}^{t-1}$ to get a new locations by $\mathbf{x}^t = \mathbf{x}^{t-1} + \Delta\mathbf{x}^t$.

2.2. Update the 3D Face Model Parameters

Similar to directly updating the landmark locations, we apply the cascade regression framework to update the 3D face model parameters based on the local appearance features (e.g. SIFT).

For the training with benchmark dataset, the ground truth 3D face model parameters are not provided. We apply the method [12] to generate the related 3D face model parameters $\mathbf{p}$, which consists of deformable parameters $\mathbf{q}$ and face pose parameters $\mathbf{h}$. The face pose parameters $\mathbf{h}$ encode the head pose (pitch, yaw, roll). And the projected 2D facial landmark based on 3D face model parameters $\mathbf{p}$ is estimated by $\mathbf{x} = g(\mathbf{p})$. Refer to [12] for more details.

For the cascade regression at iteration t, we use the linear regression model to predict the updates of 3D face model parameters by:

$$\Delta\mathbf{p}^t = \mathbf{Q}^t \Phi(g(\mathbf{p}^{t-1}), \mathbf{I}) + \mathbf{c}^t, \tag{3}$$

where $g(\mathbf{p}^{t-1})$ denotes the projected 2D landmark locations $\mathbf{x}^{t-1}$ based on current estimated 3D face model parameters $\mathbf{p}^{t-1}$, $\Phi(g(\mathbf{p}^{t-1}), \mathbf{I})$ is the extracted image features at current location $g(\mathbf{p}^{t-1})$, and $\mathbf{Q}^t$ and $\mathbf{c}^t$ are the parameters of cascade regressor Q^t at iteration t.

For training, given K training images with ground truth 3D face model parameters $\mathbf{p}^*$, we can learn the model parameters $\mathbf{Q}^t$ and $\mathbf{c}^t$ by standard least-square formation with closed form solution as below:

$$\mathbf{Q}^{t^*}, \mathbf{c}^{t^*} = arg \min_{\mathbf{Q}^t, \mathbf{c}^t} \sum_{i=1}^{K} \parallel \Delta\mathbf{p}_i^{t,*} - \mathbf{Q}^t \Phi(g(\mathbf{p}_i^{t-1}), I_i) - \mathbf{c}_i^t \parallel^2 . \tag{4}$$

For *i*-th image, we can get the ground truth updates of 3D face model parameters $\Delta\mathbf{p}_i^{t,*}$ by subtracting the current model parameters $\mathbf{p}_i^{t-1}$ from the ground truth 3D face model parameters $\mathbf{p}^*$. Given the *i*-th training image with current estimated model parameters $\mathbf{p}_i^{t-1}$, we can estimate the current landmark locations $\mathbf{x}_i^{t-1}$ from projection function $g(\cdot)$ where $\mathbf{x} = g(\mathbf{p})$. Hence, the related local appearance features $\Phi(g(\mathbf{p}_i^{t-1}), I_i)$ can be calculated. It should be noted that we initialize the landmark location based on the mean 3D face model parameters $\bar{\mathbf{p}}$ from the training data by $g(\bar{\mathbf{p}})$.

For pose estimation at iteration t, given the facial image $\mathbf{I}$, projection model $g(\cdot)$ with current 3D face model parameters $\mathbf{p}^{t-1}$ and learned regressor with parameters $\mathbf{Q}^t$ and $\mathbf{c}^t$, we calculate the updates $\Delta\mathbf{p}^t$ of 3D face model parameters using the learned regressor by Eq. 3. The new 3D face model parameters can be estimated by $\mathbf{p}^{t*} = \mathbf{p}^{t-1} + \Delta\mathbf{p}^t$.

2.3. Coupled Cascade Regression

Since 2D facial landmark detection and 3D head pose are related, to exploit their dependencies, we further propose to add a third step to make them consistent with each other. We name the whole framework as coupled cascade regression (CCR) since the third step couples the first and second step through the projection model. Specifically, for the third step, we define the objective function as Euclidean distance of landmark locations from two tasks as below:

$$\mathbf{x}^*, \mathbf{p}^* = arg \min_{\mathbf{x},\mathbf{p}} \varepsilon(\mathbf{x}, \mathbf{p}) = arg \min_{\mathbf{x},\mathbf{p}} \frac{1}{2}(\mathbf{x} - g(\mathbf{p}))^2, \tag{5}$$

where both $\mathbf{x}$ and $\mathbf{p}$ are unknowns, and $g(\mathbf{p})$ is the projection function. For iteration $t+1$, we solve this optimization problem by gradient descent method with initialization of $\mathbf{p}^{t*}$ and $\mathbf{x}^{t*}$. We alternatively update 3D face model parameters $\mathbf{p}$ and the location $\mathbf{x}$ as below:

$$\begin{aligned} \mathbf{p}^{t+1} &= \mathbf{p}^{t*} - \eta \frac{\partial \varepsilon(\mathbf{x}^{t*}, \mathbf{p})}{\partial \mathbf{p}}|_{\mathbf{p}^{t*}} \\ \mathbf{x}^{t+1} &= \mathbf{x}^{t*} - \xi \frac{\partial \varepsilon(\mathbf{x}, \mathbf{p}^{t*})}{\partial \mathbf{x}}|_{\mathbf{x}^{t*}}, \end{aligned} \tag{6}$$

where η and ξ are the learning rate parameters. From the estimated $\mathbf{p}$, we can obtain face pose parameters $\mathbf{h}$ and deformable parameters $\mathbf{q}$.

3. EXPERIMENTS

3.1. Datasets

We conduct experimental comparisons with conventional learning based methods on landmark detections and pose estimations on two benchmark datasets, 300-W [17] and BU [9]. 300-W combine images from AFW, IBUG, LFPW, Helen and XM2VTS. In particular, we follow the same protocol as [18]: the whole set of AFW, training images of LFPW and Helen are used for training which consists of 3,148 faces in total. The IBUG (challenging), testing images of LFPW and Helen (common) are used for testing which consists of 689 faces in total (full). The Boston University (BU) head tracking database [9] consists two subsets with uniform-lighting and varying-lighting conditions. We follow the same protocol in [11] and [8], where they use the 45 videos (each video with

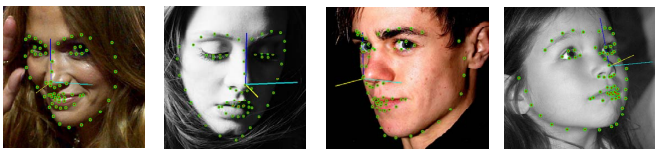

Fig. 1. Qualitative results of our proposed CCR method for 300-W dataset (best view in color).

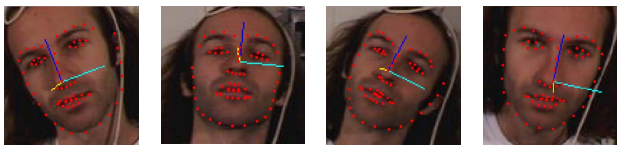

Fig. 2. Qualitative results of our proposed CCR method for BU dataset (best view in color).

Table 1. Landmark detection comparison of point-to-point error (%) normalized by intra-occular distance on 300-W dataset (68 facial points), with best result highlighted.

Method	Common	Challenging	Full
Zhu *et. al*[19]	8.22	18.33	10.20
DRMF [20]	6.65	19.79	9.22
ESR [6]	**5.28**	17.00	7.58
RCPR [21]	6.18	17.26	8.35
SDM [5]	5.57	15.40	7.50
CR-Landmark	5.75	15.56	7.64
CR-3DMP	6.47	18.08	8.78
CCR(proposed)	5.55	**14.01**	**7.24**

Table 2. Head pose estimation comparison of mean absolute error in degree on BU dataset, with best result highlighted.

Method	Pitch	Yaw	Roll	Average
Rigid model [10]	11.9	5.2	2.8	6.6
Cylindrical [9]	6.6	**3.3**	9.8	6.4
AAM+Cylindrical[8]	5.6	5.4	**3.1**	4.7
SDM+Deformable[11]	**4.3**	6.2	3.2	4.6
CCR(proposed)	4.8	5.1	3.3	**4.4**

200 frames) of 5 subjects with uniform-lighting conditions to evaluate the head pose estimation.

3.2. Experimental Results

For facial landmark detection, we conduct experimental comparisons with other methods. To demonstrate the effectiveness of our proposed method, we first discard the third step in Algorithm 1 without capturing any joint relationship using the cascade regression framework, where we call this method as CR-Landmark. In addition, we conduct another experiments where we do cascade regression to update 3D face model parameters only (only the second step in Algorithm 1). We call it as CR-3DMP. Both CR-Landmark and CR-3DMP are baselines for comparisons. We report the normalized inter-ocular error in Table 1. Some image examples are shown in Fig. 1.

Results for a large and challenging dataset named 300-W are shown in Table 1. From Table 1, CCR can achieve preferable results on Challenging and Full subset of 300-W. For the Common subset, CCR achieve a normalized error of 5.55% and the improvement is marginal over the baseline of CR-Landmark. While for the Challenging subset, CCR significantly outperforms the SDM by 9.02%. The reason comes from that the challenging subset contains images with larger head pose and CCR leverages 3D face model with pose information. This is also one of our major contributions.

Since there is little work on simultaneous head pose estimation and landmark detection. We conduct comparisons for head pose estimation separately in Table 2 on BU dataset. Since BU dataset consists of 45 videos, we do the detection and tracking on the BU dataset and test on each frame. Some detection image examples are shown in Fig. 2. It should be noted that, the ground truth of 3D face model parameters (head pose and deformable parameters) are generated by model base method. We compare our model for head pose with other model-based approaches. As shown in Table 2, CCR can also achieve competitive results with an average mean absolute error of 4.4 in degree. Different from conventional methods that sequentially perform landmark detection and head pose estimation, CCR achieves two tasks in one step using the appearance features. It is worth nothing that, though the improvement compared with the state-of-the-art is marginal, CCR can simultaneously predict landmark locations and estimate the 3D head pose. In addition, since we approximately generate the 3D face model parameters as ground truth, it can further be improved for head pose estimation with accurate 3D face model parameters.

4. CONCLUSION

In this paper, we propose a novel method called Coupled Cascade Regression (CCR) for simultaneous landmark detection and head pose estimation. CCR performs cascade regression for two different tasks to update the 3D face model parameters and landmark locations separately, followed by combining the landmark position and head pose through the projection model to further improve the estimations for both tasks. It allows to incorporate local textural and projection model at each cascaded iteration to simultaneously improve the performance both on landmark detection and head pose estimation. Experiments demonstrate that CCR achieve preferable results on benchmark datasets.

5. REFERENCES

[1] Yue Wu and Qiang Ji, "Constrained joint cascade regression framework for simultaneous facial action unit recognition and facial landmark detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. CVPR, 2016.

[2] Joseph Roth, Yiying Tong, and Xiaoming Liu, "Adaptive 3d face reconstruction from unconstrained photo collections," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. CVPR, 2016.

[3] Athi Narayanan, Ramachandra Mathava Kaimal, and Kamal Bijlani, "Estimation of driver head yaw angle using a generic geometric model," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, no. 99, pp. 1–15, 2016.

[4] Erik Murphy-Chutorian and Mohan Manubhai Trivedi, "Head pose estimation in computer vision: A survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 4, pp. 607–626, 2009.

[5] Xuehan Xiong and Fernando De la Torre, "Supervised descent method and its applications to face alignment," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 532–539.

[6] Xudong Cao, Yichen Wei, Fang Wen, and Jian Sun, "Face alignment by explicit shape regression," *International Journal of Computer Vision*, vol. 107, no. 2, pp. 177–190, 2014.

[7] Shizhan Zhu, Cheng Li, Chen Change Loy, and Xiaoou Tang, "Unconstrained face alignment via cascaded compositional learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. CVPR, 2016.

[8] Jaewon Sung, Takeo Kanade, and Daijin Kim, "Pose robust face tracking by combining active appearance models and cylinder head models," *International Journal of Computer Vision*, vol. 80, no. 2, pp. 260–274, 2008.

[9] Marco La Cascia, Stan Sclaroff, and Vassilis Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 4, pp. 322–336, 2000.

[10] Yan Tong, Yang Wang, Zhiwei Zhu, and Qiang Ji, "Robust facial feature tracking under varying face pose and facial expression," *Pattern Recognition*, vol. 40, no. 11, pp. 3195–3208, 2007.

[11] Francisco Vicente, Zehua Huang, Xuehan Xiong, Fernando De la Torre, Wende Zhang, and Dan Levi, "Driver gaze tracking and eyes off the road detection system," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 16, no. 4, pp. 2014–2027, 2015.

[12] Chao Gou, Yue Wu, Fei-Yue Wang, and Qiang Ji, "Shape augmented regression for 3d face alignment," in *European Conference on Computer Vision Workshops*. Springer, 2016, pp. 604–615.

[13] Amin Jourabloo and Xiaoming Liu, "Large-pose face alignment via cnn-based dense 3d model fitting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. CVPR, 2016.

[14] Feng Liu, Dan Zeng, Qijun Zhao, and Xiaoming Liu, "Joint face alignment and 3d face reconstruction," in *European Conference on Computer Vision*. Springer, 2016, pp. 545–560.

[15] Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, and Stan Z Li, "Face alignment across large poses: A 3d solution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. CVPR, 2016.

[16] Sergey Tulyakov and Nicu Sebe, "Regressing a 3d face shape from a single image," in *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2015, pp. 3748–3755.

[17] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2013, pp. 397–403.

[18] Shizhan Zhu, Cheng Li, Chen Change Loy, and Xiaoou Tang, "Face alignment by coarse-to-fine shape searching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4998–5006.

[19] Xiangxin Zhu and Deva Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2879–2886.

[20] Akshay Asthana, Stefanos Zafeiriou, Shiyang Cheng, and Maja Pantic, "Robust discriminative response map fitting with constrained local models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3444–3451.

[21] Xavier P Burgos-Artizzu, Pietro Perona, and Piotr Dollár, "Robust face landmark estimation under occlusion," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1513–1520.