SHAPE INITIALIZATION WITHOUT GROUND TRUTH FOR FACE ALIGNMENT

Rizhen Qin *Ting Zhang*

National Laboratory of Pattern Recognition (NLPR)
Institute of Automation, Chinese Academy of Sciences, Beijing, China
rizhen.qin@nlpr.ia.ac.cn ting.zhang@nlpr.ia.ac.cn

ABSTRACT

The shape initialization is a crucial step for face alignment. In the literature, many approaches use the ground truth points to compute the bounding box. However, it is not always possible to detect an accurate bounding box in real applications due to various adverse factors. In this work, an effective initialization approach for face alignment is proposed. Firstly a modified Deformable Part Models (DPM) is used to estimate the face pose and the bounding box to obtain an initial shape. Then by detecting the two pupils, the roll rotation of the face is measured to correct the initial shape. To further increase the robustness and accuracy of face alignment, multiple initial shapes for each face are generated, then each one is refined by a cascade regression-based approach and we can get multiple shape estimations. Finally a better final shape is obtained by fusing the multiple estimations via the structured SVM learning. Experiments on challenging datasets and comparison with the state-of-the-art methods validate our proposed method in unconstrained environment.

Index Terms— shape initialization, face alignment, DPM, multiple initial shapes, structured SVM

1. INTRODUCTION

Face alignment is an essential step for many face based applications, such as face tracking and face animation. Thus a fully automatic, highly effective face alignment method is always desirable. Up to now, many approaches [1, 2, 3, 4, 5, 6, 7, 8] have been proposed to address this issue. However, accurate and robust face alignment is still a challenging task in unconstrained environments, due to large variations on facial appearance and occlusions.

For most face alignment methods, an initial shape is determined at first, and then progressively refined. In the literature, many methods assume that the used face detector can provide a reliable initialization, then what is needed is to merely refine this initialization. However this assumption is usually difficult to meet in practice due to large variations on facial appearance, illumination and partial occlusions which dramatically reduce the performance of the face detector. And the performance of many face alignment approaches degrades substantially in unconstrained environment.

In this paper, we propose several techniques to reliably construct an accurate initial shape for face alignment. In the literature, the mean shape is usually used for the initial shape construction from the bounding box. However, since faces usually have different yaw poses and roll rotations, using just a simple mean shape as the initial shape is not sufficient to acquire an accurate result. To obtain a more accurate shape initialization for face alignment, here we firstly use a modified Deformable Part Models (DPM) trained on a set of face images to estimate the bounding box and the yaw pose of the face. Then we set a specific initial shape for the face based on the estimated pose. After this, the positions of the two pupils are detected and used for estimating the roll rotation, by which the initial shape is rotated to make it more accurate.

After the above face shape initialization, the final shape is estimated under the cascade shape regression framework. At this stage, the method in [8] is adopted to extract local binary features representing the texture information around the estimated landmarks, and simple linear regressors are used for the shape bias estimation in each iteration.

Note that in some cases, the bounding box of a face and the positions of the two pupils cannot always be accurately detected, in particular, in the challenging unconstrained environment. To tackle such problem, we generate multiple hypotheses by shifting and re-scaling the initial shape, then estimate facial landmarks for each hypothesis, instead of using only one initial shape. Considering the existence of possible complementary information among these multiple estimations, the structured SVM is used to learn a better final shape from them.

2. RELATED WORK

Although many different methods have been proposed for face alignment, they appear sharing similar essential principles. Most previous works can be divided into two categories: *model based approaches* and *regression based approaches*.

Model based approaches mainly focus on training a shape model and applying it to fit a new face. The representative works include Active Shape Models (ASM) [9] and Active Appearance Models (AAM) [10], which both constrain the shape based on PCA. In recent years, many improvements over AAM and ASM have been proposed [2, 11, 12, 13] and these approaches are better for generalization and robustness. In [3], instead of modeling the holistic appearance as ASM, Constrained Local Models (CLM) learns a set of local detectors to capture the appearance in patches around facial landmarks and constrains them using a shape model. In [14], Saragih et al generalized the CLM by using a more sophisticated local model and the mean-shift was used for the matching.

Regression based approaches reflect the nature of the face alignment problem. Many approaches try to learn a regression function to directly map the image appearance to the shape. These methods often use boosted regression and random fern regressors. In recent years, some regression based methods have achieved significant progress. [7] used linear regression on SIFT features to predict the shape increment. [15] tested several local descriptors and found that HOG features performed the best. [4] used boosted ferns and pixel-

This work was supported by National Natural Science Foundation of China under the grant No(61273280, 61227804, 61305048).

different features to regress the shape increment. [8] generalized the ESR algorithm [4] by using the local regression and obtained more accurate result. Instead of using cascade simple regressors, [16] constructed a deep neural network to directly learn the regression function between the original image and the positions of landmarks.

3. ACCURATE SHAPE INITIALIZATION

3.1. Shape Initialization Based on DPM

Generally, most face alignment methods assume the face detector can provide a reliable bounding box, then set the initial shapes from it. However in practice, it is not always possible to find an accurate bounding box in unconstrained environment. As a result, the performance of many face alignment methods will degrade dramatically due to the sensitivity of their results to the shape initialization. Thus it is always desirable to provide a good initial shape. The DPM was first introduced in [17] to address the object detection problem and achieved enormous success. To solve the initialization problem, we use a modified DPM model in this work to detect the bounding box and the pose of a face to construct an accurate initial shape.

During the bounding box detection by the DPM, we use both the position and the scale information of every part filter to predict the bounding box. In the real world, many faces have a yaw pose. As shown in Figure 2(b), when the face is not a frontal one, significant differences may exist between the ground truth and the initial shape which is usually the mean value of the whole face dataset. To deal with this problem, we propose to use the DPM to detect the face pose, and provide a more accurate initial shape via the estimated pose.

(a) (b) (c) (d) (e) (f)

Fig. 1. (a) to (f) are the mean shapes derived from six groups of face images with different pose. Each of them is used to initialize one image for the face alignment based on the pose detected by the DPM.

In this paper, we use a DPM model to estimate the yaw pose and the bounding box of a face at the same time. For a face dataset containing the pose information of each face, we divide the training images into six groups according to their poses. Based on these training face images, we use the DPM algorithm to learn a mixture model which contains six components, and each one of them corresponds to a certain pose. When applying this model to process a face, it can determine which component is most fit to this face, which represents the pose of the face. Different from many other methods that use all samples in the face dataset to calculate the mean shape which is used to construct the initial shape, we use faces in different groups to calculate six different mean shapes respectively in order to provide a more accurate initial shape. Figure 1 shows such six mean shapes. According to the pose estimated by the DPM, we set an initial shape with the corresponding pose for a face. As shown in Figure 2(c), after applying the DPM, we can get a more reliable initial shape.

3.2. Initial Shape Refinement Based on Pupils

In face alignment, if the face has a roll rotation, it will be difficult to set an accurate initial shape. However, the rotation of a face could be estimated based on some image information, such as the positions of two pupils. Thus the key to face roll rotation estimation is the pupil detection. Fortunately, compared with other non-salient landmarks, a few salient landmarks, such as eye centres and mouth corners, can be reliably characterized by their image appearances. In this work, we propose to detect the positions of two pupils of a face firstly, and then calculate the face rotation angle according to the coordinates of the detected pupils. As described above, we get the yaw pose of the face using the DPM model, and then set an initial shape with the corresponding pose. Here we rotate this initial shape obtained in the proceeding section based on the rotation angle to further refine the initial shape. As shown in Figure 2, when using the DPM and the positions of two pupils, the initial shape is set step by step, and more and more accurate.

(a) (b) (c) (d)

Fig. 2. Results of the initialization. (a) the initial shape based on the Viola-Jones face detector. (b) the shape refined using the bounding box provided by the DPM. (c) a better shape based on the detected face pose. (d) the best shape by further refining the initial shape based on the roll rotation angle.

4. CASCADE REGRESSION

Face alignment can be naturally cast as a regression problem. It takes an image I as the input and outputs a shape S parameterized by the coordinates of a set of facial landmarks. As shown in Eq. 1, in the training process we want to learn a regression function f to minimize the mean square error.

$$f = \arg\min_{f} \sum_{i=1}^{N} \|f(S_i^0, I_i) - S_i\|_2 \tag{1}$$

where f returns a new shape based on the initial shape S_i^0 for each image I_i.

To simplify the problem, f is always divided into a series of simpler regression functions $\{f_1, f_2, \cdots, f_T\}$, which satisfy the following relationship :

$$f = f_T \circ f_{T-1} \circ \cdots \circ f_1, \tag{2}$$

where $f_l \circ f_{l-1}$ mean the input of f_l is the output of f_{l-1}.

In the shape regression, it is crucial to select an appropriate feature set to represent the appearance. In [8], Ren et al learned a set of highly discriminative local binary features for each facial landmark based on a locality principle and achieved impressive performance. In this paper, we use the approach proposed in [8] to get the shape estimation from the initial shape.

In [8], Ren et al progressively refine the S by predicting the shape increment ΔS in each cascade step.

$$\Delta S_i^t = W^t \Phi^t(S_i^{t-1}, I_i) \tag{3}$$

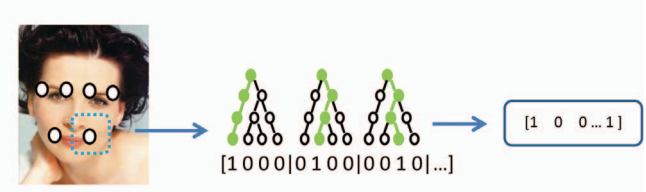

Fig. 3. The random forest is used as the local mapping function to extract local binary features. Every binary feature indicates whether the input image contains some kind of local patterns.

where I_i is the input image, S_i^{t-1} is the predicted shape in the previous stage, Φ^t is a feature mapping function, and W^t is a linear regression matrix.

In the stage of learning local binary features, [8] divided the feature mapping function Φ^t to a set of local feature mapping functions i.e., $\Phi^t = [\phi_1^t, \phi_2^t, \cdots, \phi_L^t]$ for each landmark, and a standard regression random forest was used to learn each local mapping function ϕ_l^t. Figure 3 illustrates the process of extracting local binary features. In [8], having learnt local binary features, they concatenated binary features into a global feature mapping function Φ^t, then learnt the global linear projection W^t by minimizing the following objective function:

$$\min_{W^t} \sum_{i=1}^{N} \|\Delta\hat{S}_i^t - W^t\Phi^t(I_i, S_i^{t-1})\|_2^2 + \lambda\|W^t\|_2^2 \qquad (4)$$

5. COMBINING MULTIPLE SHAPES

The effectiveness of cascade regression based face alignment method is largely based on a reliable initialization. However the DPM based initialization and refinement method described in section 3 cannot always find an accurate bounding box due to various factors such as large variations on facial appearance, illumination and partial occlusions. To further enhance the reliability of initialization, a multi-hypothesis scheme is used in this work. We generate multiple hypotheses $\{S_{i1}^0, S_{i2}^0, \cdots, S_{iM}^0\}$ by shifting and re-scaling the initial shape in the initialization stage and get multiple estimations of landmarks respectively through the cascade regression. Inspired by [15], a learning based fusion strategy from multiple estimations is adopted in our work for the finally estimation.

Shape estimations $\{\hat{S}_{i1}, \hat{S}_{i2}, \cdots, \hat{S}_{iM}\}$ can be estimated based on multiple initial shapes $\{S_{i1}^0, S_{i2}^0, \cdots, S_{iM}^0\}$ for an image I_i. However, we find even the best one out of these multiple estimations cannot meet the requirement in many applications. In fact, some complementary information does exist among different estimations we acquired. For example, some estimations may have better performance for landmarks in the cheek while other estimations in the eyes and the nose. By appropriately combining these multiple estimations, a better final estimation could be achieved. Considering dozens of landmarks are present in each shape estimation, in theory, each facial landmark of the final estimation can be chosen from all possible corresponding ones in the multiple estimations, then the best combination should be taken as the final one. To reduce the computational load and to enforce a local shape constraint, in this work, landmarks located in a local region such as eyes or the mouth are always selected from the same shape estimation.

After acquiring many potential outputs by combining different parts of multiple estimations, we should determine which one is the best. As described in [15], we define a mapping W and expect the best estimation to have the largest output among all estimations:

$$\hat{S}_i = \underset{l=1,\cdots,M}{arg\,max}\{\langle W^T \cdot \Phi(\hat{S}_{i1}, I_i)\rangle, \cdots, \langle W^T \cdot \Phi(\hat{S}_{iM}, I_i)\rangle\} \quad (5)$$

where $\hat{S}_i$ is the final output. regarding to the features Φ in Eq. 5, we use the same local binary features described in section 4.

Obviously this is a standard structured SVM problem, so we use the structured SVM algorithm [18, 19] to learn a mapping to automatically combine multiple estimations. As for the specific problem in this paper, we define the loss function between the estimated shape $\hat{S}_i$ and the ground-truth shape S_i as : $\Delta(S_i, \hat{S}_i) = \|S_i - \hat{S}_i\|_2^2$. As the dependency of the loss on W is very complex, we use a MATLAB wrapper of structured SVM [20] to solve this problem.

6. EXPERIMENT

In this section, we report our experiments on three widely used benchmark datasets. They present different challenges, such as different number of images, different variations in head pose, occlusions and illumination.

LFPW was first introduced in [21] and collected from the web. Its face images have large variations in pose, illumination and expressions, so it's a good benchmark to test the face alignment performance in unconstrained conditions. However due to some invalid URLs, we only use the 811 of the 1100 training images and 224 of the 300 testing images provided by [22]. In our experiments, we conduct evaluations on 68 and 49 points settings.

Helen was created in [23] and contains 2330 high resolution web images. Among these images, there are 2000 images for training and 330 images for testing. Compared with LFPW, Helen is more challenging because they contain more images with large variations. Here, we perform evaluations on 68 and 49 points settings.

300-W [24] is an extremely challenging dataset due to the large variation in pose, expression, background and image quality. It contains several existing datasets, such as LFPW, AFW, Helen and a challenging 135-image IBUG set. In our experiments, the training images consist of AFW dataset and the training sets of LFPW and Helen datasets. In addition, we perform testing on two parts: one is IBUG which contains 135 images, the other is the union of IBUG dataset, the test images set of LFPW and Helen dataset.

Evaluation In our experiments, when testing the performance of our method, we follow the standard [4, 8] to evaluate the shape estimation error for each sample using the standard landmarks mean error normalised by the inter-pupil distance.

6.1. Comparison with State-of-the-art Methods

We compare our method with several state-of-the-art ones. During the training stage, we use AFW dataset and training sets of LFPW and Helen to construct the training set. In the testing stage, we apply our method in IBUG dataset and testing sets of LFPW and Helen respectively. In our experiments, we use the code of the LBF [8] method released by its author to extract features which represent the appearance information around landmarks. During the training of our model, we set parameters as follows: T = 7, N = 700, D = 5, where T is the number of iterations, N is the number of trees in each stage, and D is the depth of each tree. In the training stage, we augment the training data by shifting and re-scaling the acquired initial shape to counter large variations. In our work, we create 15 different initial shapes and finally obtain 15 shape estimations. Then we

LFPW Dataset			Helen Dataset			300-W Dataset(All 68 points)		
Method	68 -pts	49 -pts	Method	68 -pts	49 -pts	Method	IBUG Subset	Fullset
Zhu et al [25]	8.29	7.78	Zhu et al [25]	8.16	7.43	Zhu et al [25]	18.33	10.20
DRMF [26]	6.57	-	DRMF [26]	6.70	-	DRMF [26]	19.79	9.22
						ESR [4]	17.00	7.58
RCPR [6]	6.56	5.48	RCPR [6]	5.93	4.64	RCPR [6]	17.26	8.35
SDM [7]	5.67	4.47	SDM [7]	5.50	4.25	SDM [7]	15.40	7.50
GN-DPM [27]	5.92	4.43	GN-DPM [27]	5.69	4.06			
						LBF [8]	11.98	6.32
						LBF fast [8]	15.50	7.37
CFSS [28]	4.87	3.78	CFSS [28]	4.63	3.47	CFSS [28]	9.98	5.76
CFSS Pratical [28]	4.90	3.80	CFSS Pratical [28]	4.72	3.50	CFSS Pratical [28]	10.92	5.09
LBF+box [8]	5.18	4.22	LBF+box [8]	5.59	4.51	LBF+box [8]	15.65	7.42
Ours	**4.96**	**4.05**	**Ours**	**5.24**	**4.14**	**Ours**	**12.11**	**6.49**

Table 1. Comparison with state-of-the-art methods in averaged errors. For most of the methods, the results are taken directly from the literature or obtained based on the authors' released codes. These comparative approaches use the ground truth points to compute the bounding box except for "LBF+box", where the initial shape is constructed based on the bounding box detected by the Viola-Jones face detection methods. For our method, we use the DPM algorithm and the positions of two pupils to refine the initial shape.

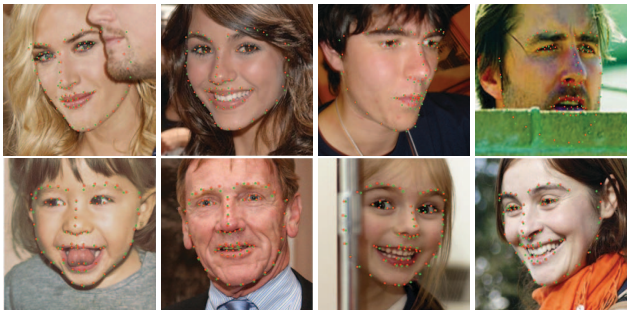

Fig. 4. Example results from the testing dataset. Red points are the ground-truth and green points are the output of our method.

use them to train the mapping of the structured SVM. In the testing stage, we follow the operation of the training to generate 15 shape estimations and combine them to get the best result.

We summarize the comparative results in Table 1. For initialization, those comparative approaches use the ground truth points to compute the bounding box. Compared with them, we use the DPM to detect the face position and use the head pose as well as the position of two pupils to refine the initial shape. Note that for some of the images(22 out of the total 689), our method failed to find faces, and for such images the ground truth bounding boxes are also used as other methods for the face initialization. As shown in Table 1, our method can achieve a comparable result compared with other methods on these datasets even if our face initialization does not use ground truth. It means that our DPM based initialization and multi-hypothesis fusion method can enhance the accuracy and robustness of face alignment method. In Figure 4, we show some example images and the face alignment results by our method.

Note in Table 1, "LBF+box" refers to the LBF face alignment method in [8] with the shape initialization by the bounding box detected by the Viola-Jones method. As the alignment step of our method also use LBF, we compare our method with "LBF+box" on all the three datasets. As shown in Table 1, our method consistently performs better on all the three datasets. In particular, our error reduction is more significant, about 3.54 for the IBUG dataset, which indicates that our method can cope with more complicated situation.

6.2. Discussion

In section 6.1, we demonstrate that using the head pose and positions of two pupils to refine the initial shape and integrating the multiple shape estimations can lead to a better result compared with merely using the bounding box. However, the face detection is still a challenging problem in unconstrained environment. In some cases, the DPM cannot find the appropriate position of the face and our method cannot achieve a good result either. Figure 5 shows some examples where our method cannot get good face alignment results.

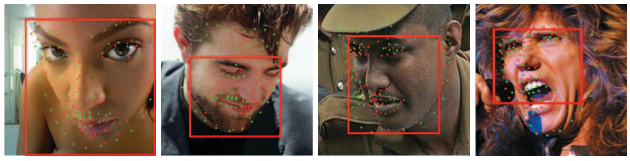

Fig. 5. Some failed examples by our method: Red points are the ground-truth and green points are the output of our method.

7. CONCLUSION

In this paper, we present an effective method for shape initialization to improve the accuracy and robustness of face alignment. In our method, the DPM is used to estimate the pose and the bounding box of the face, and the positions of two pupils are measured to calculate the rotation angle of the face. To further increase the accuracy and robustness, multiple initial shapes are generated by shifting and re-scaling the initial shape. In addition, the structured SVM is adopted to learn a combining strategy to fuse multiple estimations. Experiments on three benchmark datasets show that our proposed method can achieve good alignment results. Besides our proposed initialization method can be used for other face alignment methods.

8. REFERENCES

[1] Stephen Milborrow and Fred Nicolls, "Locating facial features with an extended active shape model," in *Computer Vision–ECCV 2008*, pp. 504–513. Springer, 2008.

[2] Iain Matthews and Simon Baker, "Active appearance models revisited," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.

[3] David Cristinacce and Timothy F Cootes, "Feature detection and tracking with constrained local models.," in *BMVC*. Citeseer, 2006, vol. 2, p. 6.

[4] Jian Sun, Fang Wen, Yichen Wei, and Xudong Cao, "Face alignment by explicit shape regression," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 2887–2894.

[5] Matthias Dantone, Juergen Gall, Gabriele Fanelli, and Luc Van Gool, "Real-time facial feature detection using conditional regression forests," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2578–2585.

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

[7] 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.

[8] Shaoqing Ren, Xudong Cao, Yichen Wei, and Jian Sun, "Face alignment at 3000 fps via regressing local binary features," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 1685–1692.

[9] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[10] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor, "Active appearance models," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 6, pp. 681–685, 2001.

[11] Ralph Gross, Iain Matthews, and Simon Baker, "Active appearance models with occlusion," *Image and Vision Computing*, vol. 24, no. 6, pp. 593–604, 2006.

[12] Patrick Sauer, Timothy F Cootes, and Christopher J Taylor, "Accurate regression procedures for active appearance models.," in *BMVC*, 2011, pp. 1–11.

[13] Philip A Tresadern, Patrick Sauer, and Timothy F Cootes, "Additive update predictors in active appearance models.," in *BMVC*. Citeseer, 2010, vol. 2, p. 4.

[14] Jason M Saragih, Simon Lucey, and Jeffrey F Cohn, "Deformable model fitting by regularized landmark mean-shift," *International Journal of Computer Vision*, vol. 91, no. 2, pp. 200–215, 2011.

[15] Junjie Yan, Zhen Lei, Dong Yi, and Stan Z Li, "Learn to combine multiple hypotheses for accurate face alignment," in *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*. IEEE, 2013, pp. 392–396.

[16] Yi Sun, Xiaogang Wang, and Xiaoou Tang, "Deep convolutional network cascade for facial point detection," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 3476–3483.

[17] Pedro Felzenszwalb, David McAllester, and Deva Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[18] Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun, "Support vector machine learning for interdependent and structured output spaces," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 104.

[19] Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, and Yasemin Altun, "Large margin methods for structured and interdependent output variables," in *Journal of Machine Learning Research*, 2005, pp. 1453–1484.

[20] A. Vedaldi, "A MATLAB wrapper of $\text{SVM}^{\text{struct}}$," `http://www.vlfeat.org/~vedaldi/code/svm-struct-matlab.html`, 2011.

[21] Peter N Belhumeur, David W Jacobs, David Kriegman, and Neeraj Kumar, "Localizing parts of faces using a consensus of exemplars," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 545–552.

[22] Christos Sagonas, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic, "A semi-automatic methodology for facial landmark annotation," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. IEEE, 2013, pp. 896–903.

[23] Vuong Le, Jonathan Brandt, Zhe Lin, Lubomir Bourdev, and Thomas S Huang, "Interactive facial feature localization," in *Computer Vision–ECCV 2012*, pp. 679–692. Springer, 2012.

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

[25] 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.

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

[27] Georgios Tzimiropoulos and Maja Pantic, "Gauss-newton deformable part models for face alignment in-the-wild," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*. IEEE, 2014, pp. 1851–1858.

[28] 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.