

ADAPTIVE CASCADED REGRESSION

Epameinondas Antonakos^{*,†}, Patrick Snape^{*,†}, George Trigeorgis[†], Stefanos Zafeiriou^{†,*}

[†]Department of Computing, Imperial College London, U.K.

^{*}Center for Machine Vision and Signal Analysis, University of Oulu, Finland

ABSTRACT

The two predominant families of deformable models for the task of face alignment are: (i) discriminative cascaded regression models, and (ii) generative models optimised with Gauss-Newton. Although these approaches have been found to work well in practise, they each suffer from convergence issues. Cascaded regression has no theoretical guarantee of convergence to a local minimum and thus may fail to recover the fine details of the object. Gauss-Newton optimisation is not robust to initialisations that are far from the optimal solution. In this paper, we propose the first, to the best of our knowledge, attempt to combine the best of these two worlds under a unified model and report state-of-the-art performance on the most recent facial benchmark challenge.

Index Terms— face alignment, cascaded regression, Gauss-Newton optimisation, descent directions

1. INTRODUCTION

Statistical deformable models have emerged as an important research field over the last few decades. The majority of ongoing research has focused on the task of facial alignment, due to the plethora of existing annotated databases [1, 2, 3, 4]. The most commonly-used face alignment methods can be separated in two families: (i) *discriminative* models that employ *cascaded regression*, and (ii) *generative* models that are iteratively optimised using a Newton-type algorithm.

Cascaded Regression models: The methods of this category aim to learn a regression function that regresses from the object’s appearance (e.g. commonly hand-crafted features) to the target output variables (the landmark coordinates). Recent works have shown that a single regression step (as previously proposed [5]) is not sufficient for accurate generic alignment and that superior performance can be achieved by employing a cascade of regression functions [6, 7, 8, 9, 10, 11, 12, 13]. The most representative example is Supervised Descent Method (SDM) [6, 7]. Cascaded regression-based techniques

^{*}Alphabetical order due to equal contribution.

The work of E. Antonakos and P. Snape was funded by the European Community Horizon 2020 [H2020/2014-2020] under grant agreement no. 688520 (TeSLA). G. Trigeorgis was funded by a DTA from Imperial College London. The work of S. Zafeiriou was funded by the FiDiPro program of Tekes (project number: 1849/31/2015).

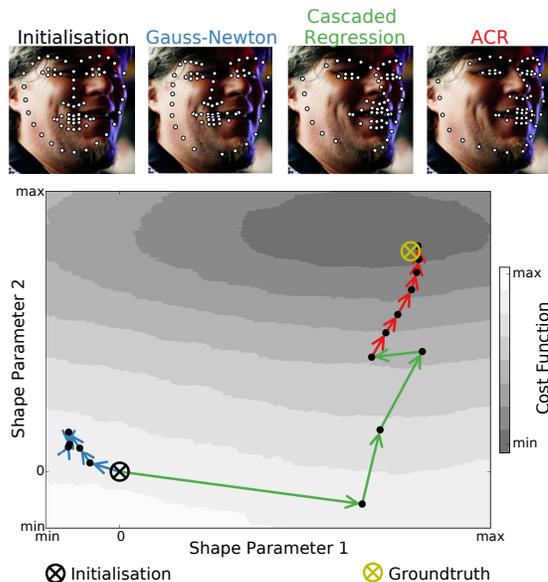


Fig. 1: Example descent directions. Cascaded-regression (green) does not reach the optimum. Gauss-Newton (blue) diverges due to the initialisation. However, applying Gauss-Newton after the final regression step (red) converges much closer to the optimum. This observation motivates the idea behind the proposed unified model.

generally return accurate results even with very challenging initialisations. Hence, they seek to learn averaged descent directions with good generalisation properties, as in the case of SDM [6]. However, since the descent directions are not specific to the test image, they are not always able to recover the fine details of the object. They also have no theoretical guarantee of local convergence on test images¹.

Generative models: The most common generative alignment method is the Active Appearance Model (AAM) [5]. AAMs are parametric linear models of shape and appearance of an object and their objective function involves the minimisation of the appearance reconstruction error with respect to the shape parameters. [14, 15] showed how to linearise the AAM objective function and optimise it using the Gauss-Newton algorithm. Numerous extensions have been published since

¹Theoretical guarantee for convergence exists only for the train set [6].

then, either related to the optimisation procedure [16, 17, 18] or the model structure [19, 20, 21]. Generative models have been shown to be more accurate than regression-based methods when initialised close to an optimum [16, 18, 19, 21]. However, the linearisation of the cost function required for Gauss-Newton optimisation causes generative models to be highly sensitive to their initialisations. In general, if Gauss-Newton is not initialised within close proximity of an acceptable local minima, the resulting alignment will be poor.

In this paper, we present a unified model that combines the generative and discriminative formulation via a weighted combination of their descent directions. Our motivation comes from the example of Figure 1, in which we plot the cost function that we aim to optimise based on a parametric shape model and a projected-out appearance subspace [14]. Note that the cost function is plotted w.r.t the first two shape parameters and is common for the discriminative and the generative models. Even though the initialisation is far from the ground-truth optimum, the cascaded regression manages to quickly converge towards the correct direction with large steps, but is not able to actually reach the optimum. In contrast, the generative model completely diverges due to the distance from the global optimum. However, by applying Gauss-Newton after cascaded regression, we reach a local optimum that is very close to the global optimum. Our proposed model employs a fully parametric cascade of regression-based descent directions, which are further adapted by the gradient descent directions provided by the Hessian of Gauss-Newton. This adaptation allows the model to be robust to very challenging initialisations and to converge to the local minimum which can recover accurate landmark localisation for the fine details of an object. Inspired by our method’s nature, we name it Adaptive Cascaded Regression (ACR). Finally, we report state-of-the-art performance on the task of face alignment, using the most recent benchmark challenge 300-W [1, 2, 3].

2. PROPOSED METHOD

Shape model We denote a landmark point for an image \mathbf{I} , as $\mathbf{x}_i = [x_i, y_i]^T$ and thus the *shape instance* of an object with n landmarks is given by $\mathbf{s} = [\mathbf{x}_1^T, \dots, \mathbf{x}_n^T]^T \in \mathbb{R}^{2n \times 1}$. Given a set of such shape samples $\{\mathbf{s}^i\}$, a parametric shape model can be modelled by aligning the shapes and then performing Principal Component Analysis (PCA). The shape subspace is further augmented with four eigenvectors that control the global similarity transform [14]. The resulting *shape model* $\{\mathbf{U}_s, \bar{\mathbf{s}}\}$ consists of the orthonormal basis $\mathbf{U}_s \in \mathbb{R}^{2n \times n_s}$ with n_s eigenvectors and the mean shape vector $\bar{\mathbf{s}} \in \mathbb{R}^{2n}$. A new shape instance is generated via $\mathbf{s}(\mathbf{p}) = \bar{\mathbf{s}} + \mathbf{U}_s \mathbf{p}$ where $\mathbf{p} = [p_1, \dots, p_{n_s}]^T \in \mathbb{R}^{n_s \times 1}$ are the *shape parameters*.

Appearance model Let us define the function $\phi(\mathbf{s})$ that returns a vector which concatenates the features (e.g. SIFT [22]) extracted from the patches centred at the landmarks of the provided shape instance \mathbf{s} . Given that each vector has

length m , then $\phi(\mathbf{s}) \in \mathbb{R}^{mn \times 1}$. Given a set of appearance samples $\{\phi^i\}$ and then applying PCA, we obtain a parametric *appearance model* $\{\mathbf{U}_a, \bar{\mathbf{a}}\}$ that consists of the orthonormal basis $\mathbf{U}_a \in \mathbb{R}^{mn \times n_a}$ with n_a eigenvectors and the mean appearance vector $\bar{\mathbf{a}} \in \mathbb{R}^{mn}$. An appearance instance can be generated as $\mathbf{a}(\mathbf{c}) = \bar{\mathbf{a}} + \mathbf{U}_a \mathbf{c}$ where $\mathbf{c} = [c_1, \dots, c_{n_a}]^T \in \mathbb{R}^{n_a \times 1}$ are the *appearance parameters*. We also define $\mathbf{P} = \mathbf{E} - \mathbf{U}_a \mathbf{U}_a^T$ as the orthogonal complement of the appearance subspace \mathbf{U}_a , where $\mathbf{E} \in \mathbb{R}^{mn \times mn}$ denotes an identity matrix.

Adaptive Cascaded Regression ACR combines the regressed descent directions with Gauss-Newton descent directions into a single unified cost function. Assume that we have a set of N training images $\{\mathbf{I}_1, \dots, \mathbf{I}_N\}$ and their corresponding annotated shapes $\{\mathbf{s}_1, \dots, \mathbf{s}_N\}$. By projecting each ground-truth shape onto the shape basis \mathbf{U}_s , we compute the ground-truth shape parameters $\{\mathbf{p}_1^*, \dots, \mathbf{p}_N^*\}$. During the training process of each cascade level, we generate a set of P perturbed shape parameters $\mathbf{p}_{i,j}^k$, $j = 1, \dots, P$, $i = 1, \dots, N$, which are sampled from a distribution that models the statistics of the detector employed for initialisation. By defining $\Delta \mathbf{p}_{i,j}^k = \mathbf{p}_i^* - \mathbf{p}_{i,j}^k$, $j = 1, \dots, P$, $i = 1, \dots, N$ to be a set of shape parameter increments, the training procedure of ACR aims to learn a cascade of K optimal linear regressors given the gradient descent directions of each training image at each level $k = 1, \dots, K$ by minimizing

$$\sum_{i=1}^N \sum_{j=1}^P \left\| \Delta \mathbf{p}_{i,j}^k - (\lambda^k \mathbf{W}^k - (1 - \lambda^k) \mathbf{H}_{i,j}^{-1} \mathbf{J}_{i,j}^T) \hat{\phi}_{i,j,k} \right\|_2^2 \quad (1)$$

with respect to \mathbf{W}^k . $\hat{\phi}_{i,j,k}$ denotes the projected-out residual defined as

$$\hat{\phi}_{i,j,k} \equiv \hat{\phi}_i(\mathbf{s}(\mathbf{p}_{i,j}^k)) = \mathbf{P} (\phi_i(\mathbf{s}(\mathbf{p}_{i,j}^k)) - \bar{\mathbf{a}}) \quad (2)$$

and $\mathbf{H}_{i,j}$ and $\mathbf{J}_{i,j}$ denote the Hessian and Jacobian matrices, respectively, of the Gauss-Newton algorithm per image $i = 1, \dots, N$ and per perturbation $j = 1, \dots, P$. λ_k is a hyper-parameter that controls the weighting between the regression-based and gradient descent directions at each cascade level k . The training procedure involves the following steps:

Step 1: Shape Parameters Increments Given the set of vectors $\mathbf{p}_{i,j}^k$, we formulate the set of shape parameters increments vectors $\Delta \mathbf{p}_{i,j}^k = \mathbf{p}_i^* - \mathbf{p}_{i,j}^k$, $\forall i = 1, \dots, N$, $\forall j = 1, \dots, P$ and concatenate them in an $n_s \times NP$ matrix $\Delta \mathbf{P}_k = [\Delta \mathbf{p}_{1,1}^k \dots \Delta \mathbf{p}_{N,P}^k]$.

Step 2: Projected-Out Residuals The next step is to compute the part-based appearance vectors from the perturbed shape locations $\hat{\phi}_i(\mathbf{s}(\mathbf{p}_{i,j}^k))$ and then the projected-out residuals of Eq. 2 $\forall i = 1, \dots, N$, $\forall j = 1, \dots, P$. These vectors are then concatenated in a single $mn \times NP$ matrix as $\hat{\Phi}_k = [\hat{\phi}_1(\mathbf{s}(\mathbf{p}_{1,1}^k)) \dots \hat{\phi}_N(\mathbf{s}(\mathbf{p}_{N,P}^k))]$.

Step 3: Gradient Descent Directions Compute the Gauss-

Newton solutions for all the images and their perturbed shapes and concatenate them in an $n_s \times NP$ matrix as

$$\mathbf{G}_k = (1 - \lambda^k) \begin{bmatrix} [\mathbf{H}_{1,1}^{-1} \mathbf{J}_{1,1}^T \hat{\phi}_1(\mathbf{s}(\mathbf{p}_{1,1}^k))]^T \\ \vdots \\ [\mathbf{H}_{i,j}^{-1} \mathbf{J}_{i,j}^T \hat{\phi}_i(\mathbf{s}(\mathbf{p}_{i,j}^k))]^T \\ \vdots \\ [\mathbf{H}_{N,P}^{-1} \mathbf{J}_{N,P}^T \hat{\phi}_N(\mathbf{s}(\mathbf{p}_{N,P}^k))]^T \end{bmatrix}^T \quad (3)$$

This involves the calculation of the Jacobian and Hessian per image as

$$\begin{aligned} \mathbf{J}_{i,j} &= \mathbf{J}_{i,j}^a \\ \mathbf{H}_{i,j} &= \mathbf{J}_{i,j}^T \mathbf{P} \mathbf{J}_{i,j} \end{aligned} \quad (4)$$

where

$$\mathbf{J}_{i,j}^a = \mathbf{J}_{\bar{\mathbf{a}}} + \sum_{l=1}^m c_{i,j}^l \mathbf{J}_l \quad (5)$$

is the model Jacobian that consists of the mean appearance Jacobian $\mathbf{J}_{\bar{\mathbf{a}}} = \frac{\partial \bar{\mathbf{a}}}{\partial \mathbf{p}}$ and the Jacobian of each appearance eigenvector denoted as \mathbf{J}_l , $l = 1, \dots, m$. Note that both $\mathbf{J}_{\bar{\mathbf{a}}}$ and \mathbf{J}_l can be pre-computed. Thus, Eq. 5 requires the calculation of the appearance parameters $\mathbf{c}_{i,j}$, given their current estimate from the previous cascade $\mathbf{c}_{i,j}^c$ and the shape increments $\Delta \mathbf{p}_{i,j}^k$, as

$$\mathbf{c}_{i,j} = \mathbf{c}_{i,j}^c + \mathbf{U}_a^T (\phi_i(\mathbf{s}(\mathbf{p}_{i,j}^k)) - \bar{\mathbf{a}} - \mathbf{U}_a \mathbf{c}_{i,j}^c - \mathbf{J}_{i,j}^a \Delta \mathbf{p}_{i,j}^k) \quad (6)$$

Step 4: Regression Descent Directions By using the previously above defined matrices, the cost function of ACR in Eq. 1 takes the form $\arg \min_{\mathbf{W}^k} \left\| \Delta \mathbf{P}_k - \lambda^k \mathbf{W}^k \hat{\Phi}_k + \mathbf{G}_k \right\|_2^2$. The closed-form solution of the above least-squares problem is

$$\mathbf{W}^k = \frac{1}{\lambda^k} (\Delta \mathbf{P}_k + \mathbf{G}_k) (\hat{\Phi}_k^T \hat{\Phi}_k)^{-1} \hat{\Phi}_k^T \quad (7)$$

Note that the regression matrix of this step is estimated only in case $\lambda_k \geq 0$. If $\lambda_k = 0$, then we directly set $\mathbf{W}_k = \mathbf{0}_{n_s \times mn}$.

Step 5: Shape Parameters Update The final step is to generate the new estimates of the shape parameters per training image as

$$\mathbf{p}_{i,j}^{k+1} = \mathbf{p}_{i,j}^k + (\lambda_k \mathbf{W}^k - (1 - \lambda_k) \mathbf{H}_{i,j}^{-1} \mathbf{J}_{i,j}^T) \phi_i(\mathbf{s}(\mathbf{p}_{i,j}^k)) \quad (8)$$

$\forall i = 1, \dots, N$ and $\forall j = 1, \dots, P$. After obtaining $\mathbf{p}_{i,j}^{k+1}$, steps 1-5 are repeated for the next cascade level.

Fitting In the fitting phase, given an unseen testing image \mathbf{I} and its initial shape parameters $\mathbf{p}^0 = [p_1^0, p_2^0, p_3^0, p_4^0, \mathbf{0}]^T$, we compute the parameters update at each cascade level k as

$$\mathbf{p}^k = \mathbf{p}^{k-1} + (\lambda_k \mathbf{W}^k - (1 - \lambda_k) \mathbf{H}^{-1} \mathbf{J}^T) \phi(\mathbf{s}(\mathbf{p}^{k-1})) \quad (9)$$

where the Jacobian and Hessian are computed as described in Step 3 of the training procedure (Eq. 4). The computational complexity per iteration is $\mathcal{O}(n_s mn(n_a + n_s + 1))$.

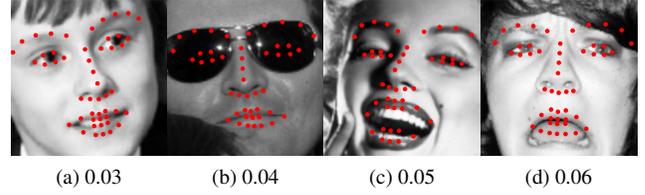


Fig. 2: Representative examples of normalised errors.

3. EXPERIMENTAL RESULTS

Datasets We use the 68-point annotations provided by [2, 1, 3] for LFPW [23], HELEN [24], AFW [25] and IBUG [2, 3] databases. For all experiments, we used the bounding boxes provided by the 300-W competition [2].

Error Metric To maintain consistency with the 300-W competition, we report cumulative error distribution (CED) graphs using the point-to-point error normalised by the interocular distance defined by the outer eye corners. We also report the failure rate, as well as the area under the curve (AUC). We fix a maximum error that we believe represents a failed fitting, and thus the higher the AUC, the more fittings are concentrated within this acceptable fitting area. In all experiments, CED curves and AUC errors are reported up to 0.06. Examples of different errors are given in Fig. 2, which shows that 0.06 represents an alignment failure. Finally, note that the error is computed on 49 points to be consistent with public implementations of state-of-the-art techniques.

Implementation Details 20 and 300 components were kept for the shape and appearance models, respectively. After performing cross-validation, we found that a cascade of 14 levels with $\lambda = [1, 0.75, 0.5, 0.25, 0, 0, \dots]$ was sufficient. Intuitively, this means that the regression-based descent directions need to dominate the optimisation for the first few iterations, as they move towards the optimum with steps of large magnitude. After that, the gradient descent steps are sufficient in order to converge to an accurate local minimum. The first two iterations are performed on the image at half scale, the rest at full scale (120×120). The patch sizes were $[32 \times 32, 24 \times 24, 24 \times 24, 16 \times 16, 24 \times 24, 24 \times 24, \dots]$. Dense SIFT [26, 22] features were used for all methods. When performing a regression, a ridge parameter of 100 was used. The training data is augmented by perturbing the provided bounding boxes of [2, 3] with uniform noise of 0.005 for scaling and 0.07 for translation. The same options were used for training the AAM and SDM using the Menpo Project [27].

Self Evaluation Fig. 3 compares ACR to both the generative AAM and the discriminative SDM. We trained all 3 methods on LFPW training set, HELEN training set and IBUG. The testing database was chosen as AFW as recent works (e.g. [12]) have shown that AFW is still a challenging dataset. Fig. 3 (left) shows the CED curve for the SDM, AAM and ACR. The curves clearly show the improved performance of

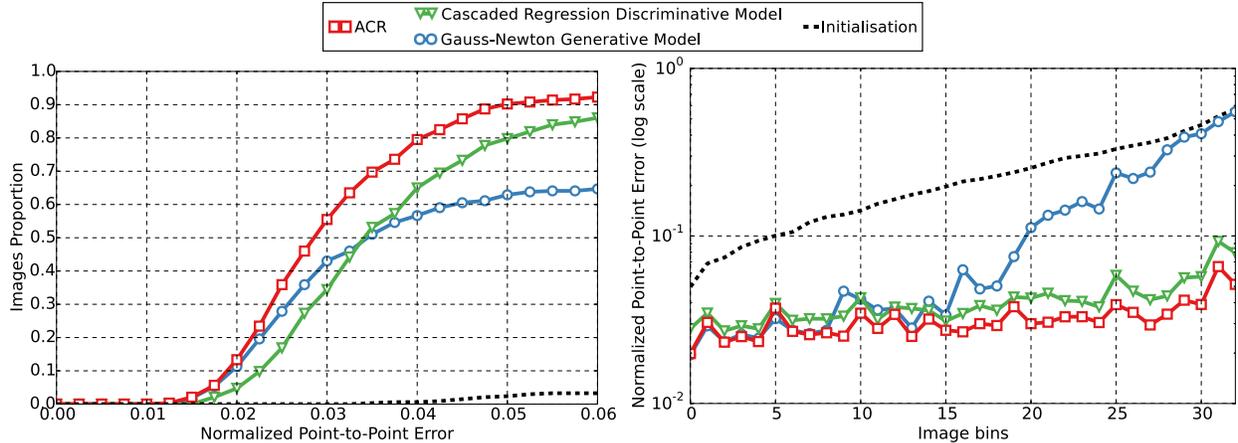


Fig. 3: ACR, AAM (Gauss-Newton) and SDM (Discriminative), trained identically, tested on the images of AFW. (Left) CED on 49 points using the bounding boxes of [2, 3]. (Right) Sorted initial errors of 10 random initialisations of each image. As the error increases, the AAM is unable to converge, whereas ACR is robust to initialisations and accurate.

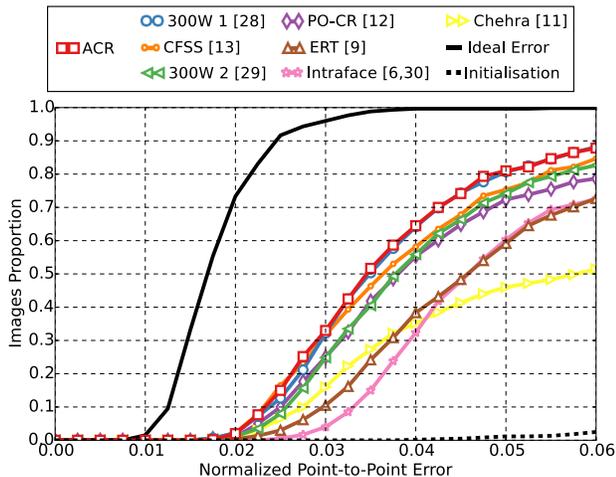


Fig. 4: Normalized error on 49 points for the testing dataset of the 300-W challenge [2, 3].

ACR over both SDM and AAM. Fig. 3 (right) demonstrates the sensitivity of generative methods to initialisations by generating 10 random initialisations per image of AFW and then sorting the initialisation errors (low-to-high). The initialisation errors were binned and the final error of the SDM, AAM and ACR is plotted with respect to increasing initial errors. As the initialisation error increases, the AAM is incapable of converging towards an acceptable local minima. In contrast, both the SDM and ACR perform well, though ACR outperforms SDM across all errors.

Comparison with state-of-the-art The testing dataset of the 300-W competition [2, 3] includes 600 “in-the-wild” images and is described as being drawn from the same distribution as the IBUG dataset. ACR was trained using LFPW training set, HELEN training set, AFW and IBUG. Fig. 4 compares the performance of ACR against the state-of-the-art methods

Method	AUC	Failure rate (%)
ACR	0.43	11.0
300W 1 [28]	0.42	9.3
CFSS [13]	0.40	13.5
300W 2 [29]	0.38	14.2
PO-CR [12]	0.37	17.7
ERT [9]	0.28	23.7
Intraface [6, 30]	0.27	23.8
Chehra [11]	0.24	46.8
Initialisation	0.01	96.8

Table 1: The area under the curve (AUC) and failure rate (% of images with error > 0.06) for the CED curve in Fig. 4.

of Zhou et al. (300W 1) [28], Yan et al. (300W 2) [29], CFSS [13], PO-CR [12], ERT [9], Intraface [6, 30] and Chehra [11]. In all cases, the testing was initialised using the bounding boxes from [2, 3]. It can be seen that ACR is the most accurate technique and slightly outperforms [28], which is a much more complex deep learning method provided by industry. Table 1 reinforces the results of Fig. 4 by showing that ACR is highly accurate and slightly less robust than the method of [28].

4. CONCLUSION

We have shown that by combining the descent directions of cascaded regression with the descent directions from Gauss-Newton optimisation, we can achieve both robustness to challenging initialisations and accuracy w.r.t fine details. We report state-of-the-art performance on the task of facial alignment, using the most recent benchmark challenge and have experimentally verified that ACR outperforms both AAM and SDM for a range of initialisations.

5. REFERENCES

- [1] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, “A semi-automatic methodology for facial landmark annotation,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPR’W)*, 2013.
- [2] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, “300 faces in-the-wild challenge: The first facial landmark localization challenge,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition Workshops (CVPR’W)*, 2013.
- [3] C. Sagonas, E. Antonakos, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, “300 faces in-the-wild challenge: Database and results,” *Image and Vision Computing*, 2016.
- [4] M. Koestinger, P. Wohlhart, P. M. Roth, and H. Bischof, “Annotated facial landmarks in the wild: A large-scale, real-world, database for facial landmark localization,” in *Proc. of IEEE Int’l Conf. on Computer Vision Workshops (ICCV’W)*, 2011.
- [5] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.
- [6] X. Xiong and F. De la Torre, “Supervised descent method and its applications to face alignment,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 532–539.
- [7] X. Xiong and F. De la Torre, “Global supervised descent method,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 2664–2673.
- [8] X. Cao, Y. Wei, F. Wen, and J. Sun, “Face alignment by explicit shape regression,” *Int’l Journal of Computer Vision*, vol. 107, no. 2, pp. 177–190, 2014.
- [9] V. Kazemi and J. Sullivan, “One millisecond face alignment with an ensemble of regression trees,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1867–1874.
- [10] S. Ren, X. Cao, Y. Wei, and J. Sun, “Face alignment at 3000 fps via regressing local binary features,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1685–1692.
- [11] A. Athana, S. Zafeiriou, S. Cheng, and M. Pantic, “Incremental face alignment in the wild,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1859–1866.
- [12] G. Tzimiropoulos, “Project-out cascaded regression with an application to face alignment,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3659–3667.
- [13] S. Zhu, C. Li, C. Loy, and X. Tang, “Face alignment by coarse-to-fine shape searching,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 4998–5006.
- [14] I. Matthews and S. Baker, “Active appearance models revisited,” *Int’l Journal of Computer Vision*, vol. 60, no. 2, pp. 135–164, 2004.
- [15] S. Baker and I. Matthews, “Lucas-kanade 20 years on: A unifying framework,” *Int’l Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, 2004.
- [16] J. Alabort-i-Medina and S. Zafeiriou, “Bayesian active appearance models,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014.
- [17] J. Alabort-i-Medina and S. Zafeiriou, “Unifying holistic and parts-based deformable model fitting,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [18] G. Tzimiropoulos and M. Pantic, “Gauss-newton deformable part models for face alignment in-the-wild,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1851–1858.
- [19] E. Antonakos, J. Alabort-i-Medina, G. Tzimiropoulos, and S. Zafeiriou, “Hog active appearance models,” in *Proc. of IEEE Int’l Conf. on Image Processing (ICIP)*, 2014, pp. 224–228.
- [20] G. Tzimiropoulos, J. Alabort-i-Medina, S. Zafeiriou, and M. Pantic, “Active orientation models for face alignment in-the-wild,” *IEEE Trans. on Information Forensics and Security*, vol. 9, no. 12, pp. 2024–2034, 2014.
- [21] E. Antonakos, J. Alabort-i-Medina, G. Tzimiropoulos, and S. Zafeiriou, “Feature-based lucas-kanade and active appearance models,” *IEEE Trans. on Image Processing*, vol. 24, no. 9, pp. 2617–2632, 2015.
- [22] D. Lowe, “Object recognition from local scale-invariant features,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 1999, pp. 1150–1157.
- [23] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, “Localizing parts of faces using a consensus of exemplars,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2930–2940, 2013.
- [24] V. Le, J. Brandt, Z. Lin, L. Bourdev, and T. S. Huang, “Interactive facial feature localization,” *Lecture Notes of Computer Science*, vol. 7574, pp. 679–692, 2012.
- [25] X. Zhu and D. Ramanan, “Face detection, pose estimation, and landmark localization in the wild,” in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2879–2886.
- [26] A. Vedaldi and B. Fulkerson, “Vlfeat: An open and portable library of computer vision algorithms,” in *Proc. of the Int’l Conf. on Multimedia*. ACM, 2010, pp. 1469–1472.
- [27] J. Alabort-i-Medina, E. Antonakos, J. Booth, P. Snape, and S. Zafeiriou, “Menpo: A comprehensive platform for parametric image alignment and visual deformable models,” in *Proc. of the ACM Int’l Conf. on Multimedia*, Orlando, Florida, USA, November 2014, pp. 679–682, ACM.
- [28] E. Zhou, H. Fan, Z. Cao, Y. Jiang, and Q. Yin, “Extensive facial landmark localization with coarse-to-fine convolutional network cascade,” in *Proc. of IEEE Int’l Conf. on Computer Vision Workshops (ICCV’W)*, 2013, pp. 386–391.
- [29] J. Yan, Z. Lei, D. Yi, and S. Z. Li, “Learn to combine multiple hypotheses for accurate face alignment,” in *Proc. of IEEE Int’l Conf. on Computer Vision Workshops (ICCV’W)*, 2013, pp. 392–396.
- [30] F. De la Torre, W.-S. Chu, X. Xiong, F. Vicente, X. Ding, and J. F. Cohn, “Intraface,” in *Automatic Face and Gesture Recognition*, 2015.