Evaluating the Quality of Face Alignment without Ground Truth

ARTICLE *in* COMPUTER GRAPHICS FORUM · OCTOBER 2015

Impact Factor: 1.6

8 AUTHORS, INCLUDING:

Weiming Dong
Chinese Academy of Sciences
47 PUBLICATIONS **175** CITATIONS
SEE PROFILE

Yan Kong
Chinese Academy of Sciences
7 PUBLICATIONS **2** CITATIONS
SEE PROFILE

Xing Mei
Chinese Academy of Sciences
30 PUBLICATIONS **159** CITATIONS
SEE PROFILE

Bao-Gang Hu
InstiChinese Academy of Sciences
147 PUBLICATIONS **1,499** CITATIONS
SEE PROFILE

Pacific Graphics 2015
N. J. Mitra, J. Stam, and K. Xu
(Guest Editors)

Volume 34 (2015), Number 7

Evaluating the Quality of Face Alignment without Ground Truth

Kekai Sheng[1] Weiming Dong[†‡1] Yan Kong[1] Xing Mei[1] Jilin Li[2] Chengjie Wang[2] Feiyue Huang[2] Bao-Gang Hu[1]

[1]LIAMA-NLPR, Institute of Automation, Chinese Academy of Sciences [2]Tencent

Abstract

The study of face alignment has been an area of intense research in computer vision, with its achievements widely used in computer graphics applications. The performance of various face alignment methods is often image-dependent or somewhat random because of their own strategy. This study aims to develop a method that can select an input image with good face alignment results from many results produced by a single method or multiple ones. The task is challenging because different face alignment results need to be evaluated without any ground truth. This study addresses this problem by designing a feasible feature extraction scheme to measure the quality of face alignment results. The feature is then used in various machine learning algorithms to rank different face alignment results. Our experiments show that our method is promising for ranking face alignment results and is able to pick good face alignment results, which can enhance the overall performance of a face alignment method with a random strategy. We demonstrate the usefulness of our ranking-enhanced face alignment algorithm in two practical applications: face cartoon stylization and digital face makeup.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—

1. Introduction

Face alignment, or locating semantic facial landmarks such as the mouth, nose, eyes, eyebrows, and chin, is a long-standing problem in computer vision and computer graphics. It is an essential task in facial recognition, 2D/3D face animations, portrait photo editing, and many other face analysis applications [HMYL15, LTO*15]. The large variations in facial appearance caused by shape, pose, illumination, expression, occlusions and out-of-plane rotation make this problem challenging. A rich body of literature on face alignment now exists. The Active Shape Model (ASM) [CTCG95, CC07] is an early approach that can deform to fit the data consistently with a training set. The Active Appearance Model (AAM) [CET01, SCT11] approach reconstructs the entire face using an appearance model and estimates the shape by minimizing the texture residual. Recently, regression-based methods [SWT13, AZCP14, ZZD*14, KS14, CHZ14] have achieved the state-of-the-art performance for accurate and robust face alignment. These methods estimate landmark locations explicitly by regression using image features. Different feature functions have been used, such as SIFT [XDLT13], HOG [YLYL13], and binary feature [CWWS14, DWP10, DGFVG12].

Research on face alignment has been progressing quickly and has achieved good results statistically on public benchmarks, such as LFPW [BJKK11], HELEN [LBL*12], AFW [ZR12], AFLW [KWRB11] and iBUG [STZP13b, STZP13a]. Each individual method has its own advantages and disadvantages, but none of them can outperform the other methods for all the images. Moreover, for some regression-based methods, the regressor may generate results of different qualities with different initial shapes for a test image. For example, in the framework of [CWWS14], the initial shape is randomly sampled from the training shapes and no quality information exists for each result. Thus, they run the regressor five times and take the median result as the final estimation to improve accuracy. As shown in Figure 1, the regressor can produce good results for some initialization but not for others. In this case, the poor medium results may affect the quality of the final result. For a specific input image, selecting the good face alignment results from many results created by different methods or by different initializations is often

† Email: {wmdong,xmei,hubg}@nlpr.ia.ac.cn

‡ Kekai Sheng and Weiming Dong contributed equally to this work.

Figure 1: *Face alignment examples. The regression may generate face shapes of different qualities with different initializations.*

useful. This task is challenging, as the quality of different face alignment results need to be compared without knowing the ground truth.

Using a data-driven approach, this study attempts to address the problem of comparing face alignment results without ground truth. We study the factors that enable a good face alignment result and select a feasible feature descriptor to measure the quality of a face shape. We then use a learning-to-rank method to rank the face alignment results of the same input image. Specifically, we first train a ranking function via the minimization of a listwise loss function and then use the associated ranking function to rank all the face alignment results. *To the best of our knowledge, this study is the first to provide a method that is able to rank the quality of face alignment results as a permutation (a total order, instead of a partially ordered set) without any ground truth.* As shown in our experiments, our method can reliably evaluate a set of face shapes generated from an input image and get a good face alignment result for it.

2. Related Works

Only a few studies are related to the evaluation of face alignment results. Huang et al. [HLW04] use the AdaBoost approach to construct an evaluation function for face alignment. A nonlinear classification function is learnt from a training set of positive and negative training examples to distinguish between qualified and unqualified alignment results. Liu [Liu07] proposes a Boosted Appearance Model (BAM) based on boosting to learn the discriminative properties between correct and incorrect alignments. The classification score provides a way to describe the quality of the image alignment. Ranking-based Appearance Models (RAM) [WLD08, ZZCM08] are also investigated by boosting the score function in a pairwise ordinal classification. This model ensures that the score function returns a higher value if the current alignment is better than the others when comparing them with the ground truth in the shape parameter space. The above mentioned methods all use the pairwise learning-to-rank scheme. Gao et al. [GES12] exploit the pointwise regression trees-based ranking model to build appearance models for face alignment. This method of face alignment is equivalent to maximizing the score for the regression trees with the constraint of the prior knowledge on shapes.

Our problem is relevant to the research on no-reference image quality assessment, which estimates the quality of an image without ground truth [WBSS04, LWC*13]. Such research primarily detects and measures image artifacts, such as those from compression or deblurring, which are not fit for our problem. Mai and Liu [ML14] train a binary classifier to compare the quality between every two detection results and then aggregate these pairwise comparison results to rank all the detection results.

3. Method

Without loss of generality, the multiple face alignment results generated by an explicit regression based method for an input image are evaluated. For this method, the regressor can provide reasonable results with different initial shapes for a test image, but the quality is variant, so it provides a convenient approach for data collection. In our experiments, the face alignment framework in [CWWS14] is employed to generate the training and testing data. The proposed framework can be directly extended to evaluate face alignment results generated by different methods. In another words, although this paper does not provide a new face alignment method per se, it provides a way to better leverage the vast amount of face alignment methods provided by the community.

3.1. Face Alignment by Regression

The basic framework of explicit regression-based face alignment methods is treating the landmark localization as a re-

gression task. Let $\mathbf{p}_k \in \mathbb{R}^2$ be the x,y-coordinates of the kth facial landmark in an image I. Thereafter, the vector $s = (\mathbf{p}_1^T, \mathbf{p}_2^T, \ldots, \mathbf{p}_N^T)^T \in \mathbb{R}^{2m}$ denotes the coordinates of all the m facial landmarks in I. We refer to vector s as the face shape. Thus, for a given input image I with an initial shape estimation s^0, s is progressively refined by cascaded regressors $r_t(\cdot,\cdot)$ at stage t:

$$s^t = s^{t-1} + r_t(I, s^{t-1}),$$

where a different initialization s^0 may lead to alignment results with various qualities.

In this paper, the shape regression method in [CWWS14] is used for a test image to generate a list of face shapes, and the task is set to select good instances from the list by ranking.

3.2. Feature Design

Given a labeled training set of facial images, we first learn the mean shape $\bar{s}$ via Principle Component Analysis (PCA) [WLD08]. With the predefined mean shape $\bar{s}$, the normalized shape of an input shape s is obtained by a piecewise affine warping, which aligns the input shape to the mean shape.

Moreover, with this warping function $\mathbf{W}$, a facial image I can be warped to the mean shape domain, where a shape-normalized facial image $\mathbf{W} \circ I$ (see Figure 2) can be computed. The size of the warped image depends on the mean shape $\bar{s}$. As shown in the first row of Figure 2, some feature points are deviated from the ones of the ground truth, especially in some points on the upper lip. These artifacts cause unexpected deformations in the resulting image during the warping process, that is, if mistakes are found with the feature points, then artifacts will be present in the warped image. Through this process, the difficult problem of analyzing face alignment without ground truth is converted into a relatively easier problem of analyzing the warped image. This conversion is one of the core ideas of our approach. We compute the feature vector φ for each shape-normalized facial image. We choose Histogram of Oriented Gradients (HOG) [DT05] as the feature, for its desirable performance in describing of boundary information and computational efficiency. In our experiments, we divide each image into 8×8 spatial regions (cells) and accumulate 8-bins local 1-D HOG, that is, the size of the window is 8×8, and the number of bins in the histogram is 8, which results in a feature vector of 512 dimensionality. The orientation bins are spaced over 0° to 180°. Each pixel in the neighborhood is voted to its histogram based on gradient direction and magnitude. Although some flexibility in the choice of the parameters of HOG, we find that $8 \times 8 \times 8$ might achieve the best results.

3.3. Face Alignment Ranking

Given a facial image I and its n face shapes, we rank these shapes based on their quality without any ground truth. Different from the studies that use the pairwise-based ranking method to compare the input objects [HLW04, WLD08, ML14], this study adopts the listwise-based learning-to-rank methodology [CQL*07, XLW*08, LLML09] to perform the evaluation without ground truth. The reasons are threefold. First, previous experiments in machine learning have demonstrated that the listwise approach usually performs better than the pairwise and pointwise approaches [CQL*07, QZT*08], especially in information retrieval systems. In this case, our task is to model a ranking function that assigns scores to the shapes, which is similar to the score function in information retrieval. Second, most pairwise models can only generate a partially ordered set, but we want to build a model for a totally ordered set. The listwise ranking models meet this requirement by nature. Moreover, the third advantage of the listwise-ranking model is seen at the computation complexity: the time for a certain pairwise model to produce a totally ordered list is $O(n^2)$, which is significantly higher than the $O(n)$ complexity of listwise-based models. Therefore, instead of pairwise ranking schemes, listwise ranking models are used in our framework to evaluate the face shapes.

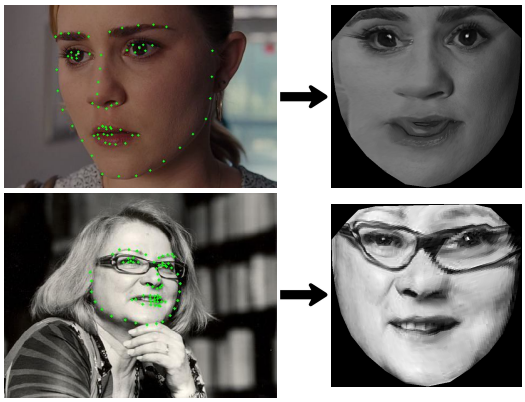


Figure 2: *Warping a face image to the mean shape domain. Such an approach is a good way to visualize errors in face alignment.*

We first describe how to obtain the labeled training data. For each image in the dataset, we use the face alignment method in [CWWS14] to generate n face shapes with different initializations (we set $n = 10$ in this case). We then use the manually annotated facial landmarks provided with each image to measure the quality of each face shape. This study employs the Root-Mean-Square Error (RMSE), which is a frequently used measure in face alignment, as the objective quality metric for face shapes. We normalize the RMSE scores to $[0, 1]$ by dividing each score with the maximum score. We obtain the ground truth ranking for each image by ranking its n face shapes in ascending order according to their RMSE scores.

We now elaborate our listwise-based ranking methods. The

framework is similar to the listwise learning-to-rank approach described in [XLW*08]. Suppose there are m facial images in the dataset and let S be the input space whose elements are sets of face shapes to be ranked, Y be the output space whose elements are permutations of face shapes, and P_{SY} be an unknown but fixed joint probability distribution of S and Y. Each set of face shapes $\{\mathbf{s}^{(i)}\}_{i=1}^{m} \in S$ contains n shape maps $\{s_j^{(i)}\}_{j=1}^{n}$ generated for the same image. Let $\mathbf{h} : S \to Y$ be a ranking function, $\mathbf{y} \in Y$, and $y^{(i)}$ be the index of the face shape ranked at position i. We learn a ranking function that can minimize the empirical loss $L(\mathbf{h})$ as follows:

$$L(\mathbf{h}) = \frac{1}{m}\sum_{i=1}^{m} \phi(\mathbf{h}(\mathbf{s}^{(i)}), \mathbf{y}^{(i)}), \tag{1}$$

where $\phi(\mathbf{h}(\mathbf{s}), \mathbf{y})$ is the loss function. The ranking function $\mathbf{h}$ assigns a score to each face shape (by employing a scoring function g), sorts the face shapes in descending order of the scores, and finally creates the ranked list, that is, $\mathbf{h}(\mathbf{s}^{(i)})$ is decomposable with respect to the face shapes and is defined as

$$\mathbf{h}(\mathbf{s}^{(i)}) = sort(g(s_1^{(i)}), \ldots, g(s_n^{(i)})),$$

where $sort(\cdot)$ denotes the sorting function, and $g(\cdot)$ is the scoring function. For simplicity, our framework defines the scoring function based on the linear network ω as

$$g_\omega(s_j^{(i)}) = \langle \omega, \varphi_j^{(i)} \rangle, \tag{2}$$

where $\varphi_1^{(j)}$ is the corresponding HOG feature vector of $s_j^{(i)}$. $\langle \cdot, \cdot \rangle$, which denotes an inner product. Furthermore, Equation (1) can be rewritten as

$$L(\mathbf{g}) = \frac{1}{m}\sum_{i=1}^{m} \phi(\mathbf{g}(\mathbf{s}^{(i)}), \mathbf{y}^{(i)}),$$

where $\mathbf{g}(\mathbf{s}^{(i)}) = (g(s_1^{(i)}), \ldots, g(s_n^{(i)}))$.

This study experimented with three learning methods for optimizing three different loss functions, including ListMLE for likelihood loss [XLW*08], RankCosine for cosine loss [QZT*08] and ListNet for cross entropy loss [CQL*07].

For the **ListMLE**, the likelihood loss function is defined as in [XLW*08]:

$$\phi(\mathbf{g}(\mathbf{s}), \mathbf{y}) = -\log P(\mathbf{y}|\mathbf{s}; \mathbf{g}), \tag{3}$$

where

$$P(\mathbf{y}|\mathbf{s}; \mathbf{g}) = \prod_{i=1}^{n} \frac{\exp(g(s_{y^{(i)}}))}{\sum_{k=i}^{n} \exp(g(s_{y^{(k)}}))}.$$

When a certain permutation is given based on specific method, a certain probability for the permutation can be calculated, and the total space of the possibility is the Plackett-Luce distribution. [Mur12]. With this defined distribution, we can use the Maximum Likelihood Estimation (MLE) to obtain the parameter of the model, and that is the key thought behind ListMLE.

We apply Stochastic Gradient Descent (SDG) as the algorithm for optimization to train the ListMLE (and for the two other ranking models), and use a linear neural network as a ranking model, which is parameterized by ω as mentioned in Equation (2). During the training process by SDG, the iteration is over when

$$\frac{\|\omega_{k+1} - \omega_k\|_2}{\|\omega_k\|_2} < 0.001,$$

or when the number of iterations is larger than 100.

For the second listwise ranking model, **RankCosine** [QZT*08], we use cosine loss as the loss function, which is based on the cosine similarity between the score vector of the ground truth and that of the predicted result [XLW*08]:

$$\phi(\mathbf{g}(\mathbf{s}), \mathbf{y}) = \frac{1}{2}(1 - \frac{\psi_\mathbf{y}(\mathbf{s})^T \mathbf{g}(\mathbf{s})}{\|\psi_\mathbf{y}(\mathbf{s})\| \|\mathbf{g}(\mathbf{s})\|}).$$

$\psi_\mathbf{y}$ is a mapping procedure used to produce the score vector of the ground truth. As mentioned before, RMSE is applied as the measure to evaluate the difference between the face alignment and the ground truth.

Finally, we let **ListNet** ranking model [CQL*07] and use cross entropy as the loss function, which is defined as:

$$\phi(\mathbf{g}(\mathbf{s}), \mathbf{y}) = D(P(\pi|\mathbf{s}, \psi_\mathbf{y}) \parallel P(\pi|\mathbf{s}, \mathbf{g})),$$
$$D(P||Q) = D_{KL}(P||Q) = \sum_i P(i) ln \frac{P(i)}{Q(i)},$$

where $\pi|\mathbf{s}, \psi_\mathbf{y}$ denotes the permutation of the n shapes sorted by comparing the ground truth with RMSE, and $\pi|\mathbf{s}, \mathbf{b}$ denotes the permutation of the shapes sorted through the ranking model. We use Plackett-Luce distribution in our model, aiming to reduce the difference between the two distributions. One distribution is acquired from the model, and the other is obtained from the reality.

To be efficient, we optimize the loss function for the ListNet method based on the top k probability [CQL*07], with the neural network as the model and gradient descent as the optimization algorithm. We set $k = 5$ in all our experiments.

3.4. Ranking Prediction

After the listwise model has been trained, it can be used to rank face alignment results of new images. Specifically, given an image I, a set $\{s_j\}_{j=1}^{n}$ of n face shapes generated from I, and $\{\varphi_j\}_{j=1}^{n}$ as the corresponding vector set, we compute a score for each face shape s_j as

$$g(s_j) = \omega \cdot \varphi_j.$$

We then obtain the rank of every s_j by sorting their scores.

4. Evaluations

We evaluate our method using the public face alignment benchmarks LFPW [BJKK11], HELEN [LBL*12],

Figure 3: *For each result of the same people, we produce 10 data via PCA to make full use of the result.*

AFW [ZR12], and IBUG [STZP13b, STZP13a]. To gain sufficient data for training, we combine the training sets of all the benchmarks to train our learning-to-rank models and to test the models on the combination of all their test sets. Therefore, a total of 1260 images are used for training and 991 images for testing. We did not use a validation set to adjust the parameters. To avoid the overfitting problem, we need to produce more data to train the model. As mentioned before, we run the explicit shape regression (ESR) method [CWWS14] 10 times for each image in the training set with different initializations to obtain 10 face alignment results. Thereafter, we apply the training set generation method in [WLD08] for each result to produce 9 extra data with PCA decomposition (see Figure 3), and a total of 100 data are obtained for each image. Finally, we have 1260×100 face alignment results in our training set, which is a considerable amount of data to train the ranking model.

In our experiments, we specifically first use ESR to generate 10 face shapes from each image, and warp the image to the mean shape to domain according to each face shape to obtain the warped images from which the HOG features are extracted. Second, we execute the ranking model with those features to calculate a score for each face shape. Finally, we sort the face shapes using the scores.

4.1. Baseline Methods

In addition to the three listwise models, the following two baseline ranking methods are compared against our methods.

4.1.1. Ranking by Pairwise Model

Despite those experience from previous literatures, we still try to use the pairwise model to do the ranking of shapes, in order to make the results from our experiments more persuasive.

In this method, the pairwise-ranking model is used to produce a certain score for the face shape. Given two face shapes s_i and s_j generated for the same image, we model the pairwise preference (quality comparison) as the probability that s_i has higher quality than s_j:

$$P_{s_i,s_j} = P(s_i \succ s_j),$$

where $s_i \succ s_j$ means that s_i has higher quality than s_j as evaluated by RMSE criteria. We model the task as a binary classification problem as [ML14], and use LibSVM [CCC14] to train the pairwise preference model P_{s_i,s_j}. Thus, given an image I, and a set $\{s_i\}_{i=1}^n$ of n face shapes generated from I, we compute a relative score for each face shape s_i as

$$r(s_i) = \sum_{j=1,\ldots,n,j\neq i} P_{s_i,s_j}.$$

The overall ranking for each face shape s_i can then be obtained by sorting their scores.

4.1.2. Predicting by Regression

Another baseline method is the random forest. In our experiments, a random forest [Bre01] is used to regress the face shape quality value with the extracted features and to be the baseline method for comparison. We note B as the number of trees in the random forest and M as the number of features extracted from the image. For each tree, two-thirds of the training face shapes are sampled as our bootstrapped training data set, and we use $\sqrt{M}$ features to fit the training set. Finally, given a face shape s_i we combine the prediction of B trees as its predicted quality:

$$f_{quality}(s_i) = \frac{1}{B}\sum_{b=1}^{B} f^b(s_i),$$

where $f^b(\cdot)$ represents the regression tree. In our experiments, B is set to 500. The overall ranking for each face shape s_i can then be obtained by sorting their qualities.

4.2. Ranking Accuracy

We examine the quality of our ranking results by assessing their correlation with the ground truth ranking. To the best of our knowledge, our study is the first to approach the problem of ranking face alignment results by the listwise method. We test the ranking accuracy on our three listwise models, including ListMLE, RankCosine and ListNet, and the two baseline methods.

4.2.1. Correlation with Ground-truth Ranking

To evaluate the extent to which ranking results agree with the ground-truth ranking, we compute their rank correlation with three rank correlation metrics.

Kendall's τ coefficient. Kendall's rank correlation coefficient [Ken38], commonly referred to as Kendall's τ coefficient, is a statistic used to measure the association between two measured quantities. For our problem, given a set of face shapes $\mathbf{s} = \{s_i, i = 1,\ldots,n\}$, a ranking function r, and the ground truth ranking function r' (sorting the face shapes by RMSE), the Kendall τ coefficient is computed as

$$\tau(r,r') = \frac{\sum_{ij}\delta[r(i,j) = r'(i,j)] - \sum_{ij}\delta[r(i,j) \neq r'(i,j)]}{0.5\cdot n\cdot(n-1)},$$

where δ denotes the indicator function. $r(i,j)$ outputs 1 if

Table 1: *Rank Correlation and Ranking Quality.*

	Kendall's τ	Spearman's ρ	$n\text{DCG}_1$	$n\text{DCG}_3$	$n\text{DCG}_5$	nDCG full list
ListMLE	0.2859	0.3607	0.8329	0.8736	0.8924	0.9474
ListNet	0.2131	0.2664	0.8108	0.8490	0.8717	0.9365
RankCosine	0.2189	0.2787	0.8087	0.8511	0.8737	0.9378
Random Forest	0.2147	0.2677	0.8096	0.8588	0.8702	0.9314
Pairwise SVM	0.2016	0.2328	0.7852	0.8129	0.8406	0.9021

the ranking function r results in i having a higher rank than j, and the function outputs 0 otherwise. This metric penalizes a pair of elements if their relative orders given by the two ranking functions disagree.

Spearman's ρ coefficient. In statistics, Spearman's rank correlation coefficient, or Spearman's ρ coefficient, is a non-parametric measure of statistical dependence between two variables. It assesses the extent to which the relationship between two variables can be described using a monotonic function. For our problem, Spearman's ρ is computed as

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot (n^2 - 1)}$$

where d_i is the difference between the rank of the ith face shape in the predicted results and in the ground truth.

The left part of Table 1 shows the average rank correlation on the test data. The results show that the alignment ranking from our listwise ranking methods have high correlation with the ground-truth ranking and that ListMLE achieves the best performance on both metrics.

4.2.2. Normalized Discounted Cumulative Gain

Discounted cumulative gain (DCG) is a measure of ranking quality that uses a graded relevance scale of elements in a ranking result set to measure the usefulness or gain of an element based on its position in the results list. The gain is accumulated from the top of the result list to the bottom with the gain of each result discounted at the lower ranks. For our problem, given a permutation of n face shapes generated by a ranking model, the DCG accumulated at a particular rank position p is defined as

$$\text{DCG}_p = rel_1 + \sum_{i=2}^{p} \frac{rel_i}{\log_2(i)},$$

where $rel_i = 1.0 - \text{RMSE}(s_i)$ is the graded relevance of the result at position i. $\text{RMSE}(s_i)$ is the RMSE score of ith face shape in the permutation, calculated by comparing with the ground truth face shape. We then use the normalized discounted cumulative gain (nDCG) to measure the ranking quality of our methods:

$$n\text{DCG}_p = \frac{\text{DCG}_p}{i\text{DCG}_p},$$

where iDCG is the ideal DCG, which is obtained by sorting the objects of a result permutation by relevance and producing the maximum possible DCG until position p. We note that in a perfect ranking algorithm, the DCG_p is the same as the $i\text{DCG}_p$ producing an nDCG of 1.0.

The right part of Table 1 shows the ranking quality on the test data. The results show that the face alignment quality ranking from our method has high ranking quality, whereas ListMLE achieves the best performance in all metrics.

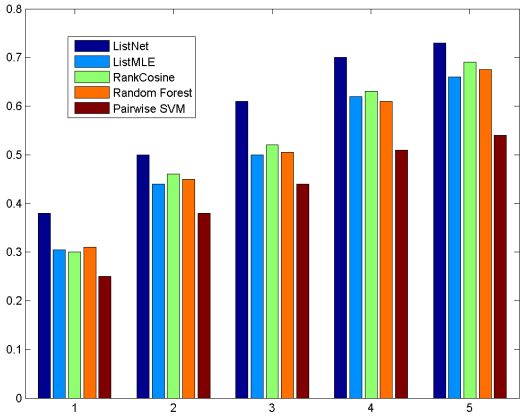
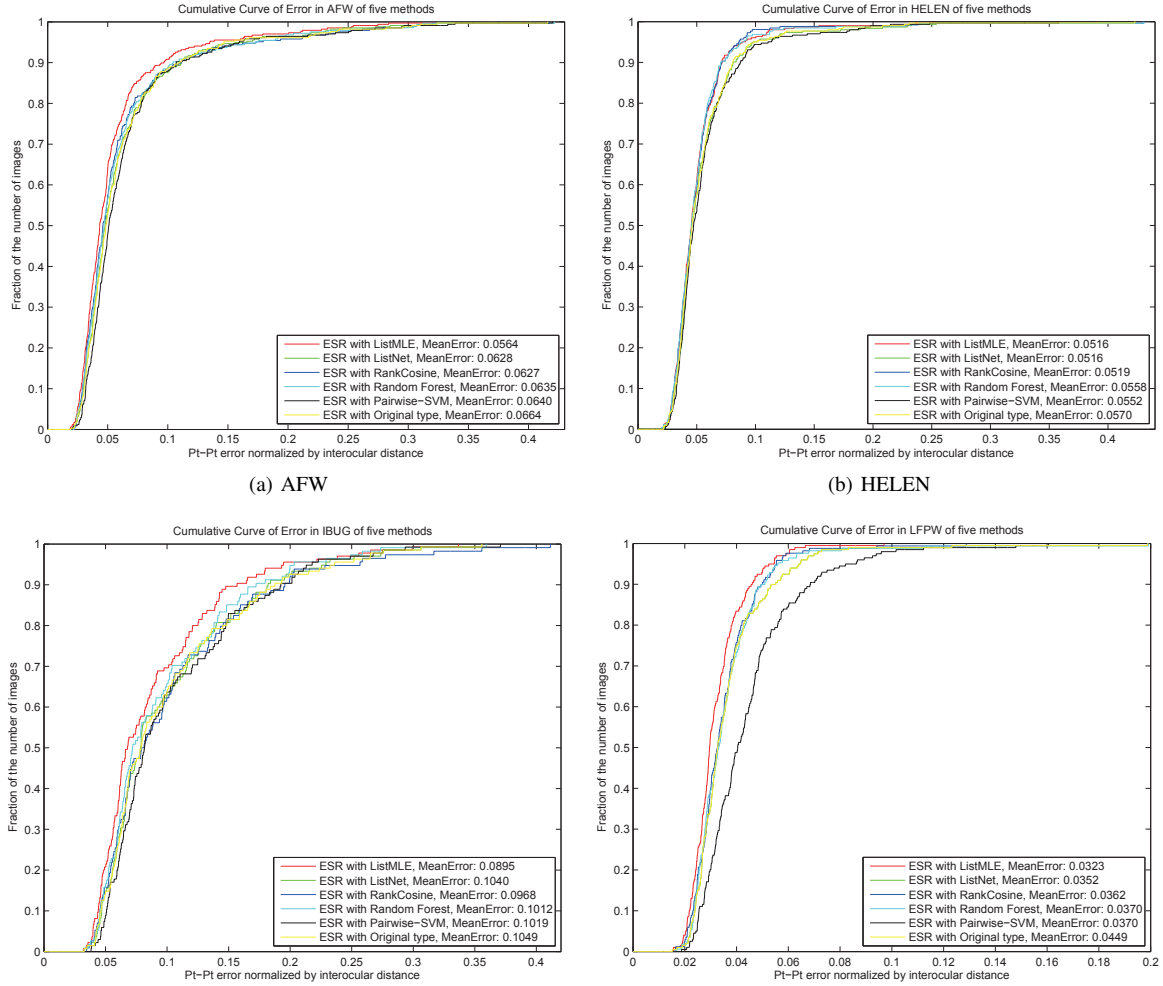


Figure 4: *R-Precision accuracy. The ranking predicted by our model is better than the two baseline methods (random forest and pairwise ranking).*

4.2.3. Quality Evaluation Performance

In this experiment, the performance of our face alignment quality evaluation results is examined for the task of retrieving the best face shape for a given image. For our evaluation, the R-precision [LO09] accuracy is measured. For a given image I, R-precision is the precision at R, where R is the number of actual best face shapes for I. In other words, if there are r actual face shapes among the top-R ranking results, then the R-precision is r/R. Given the ranking results for all testing images, the R-precision accuracy is computed as the average of the R-precision values of all the test data. Figure 4 shows 1-precision, 2-precision, 3-precision, 4-precision, and 5-precision accuracies from our methods on four benchmarks, where the ListMLE method achieves the highest face alignment quality ranking accuracy.



(a) AFW (b) HELEN

(c) IBUG (d) LFPW

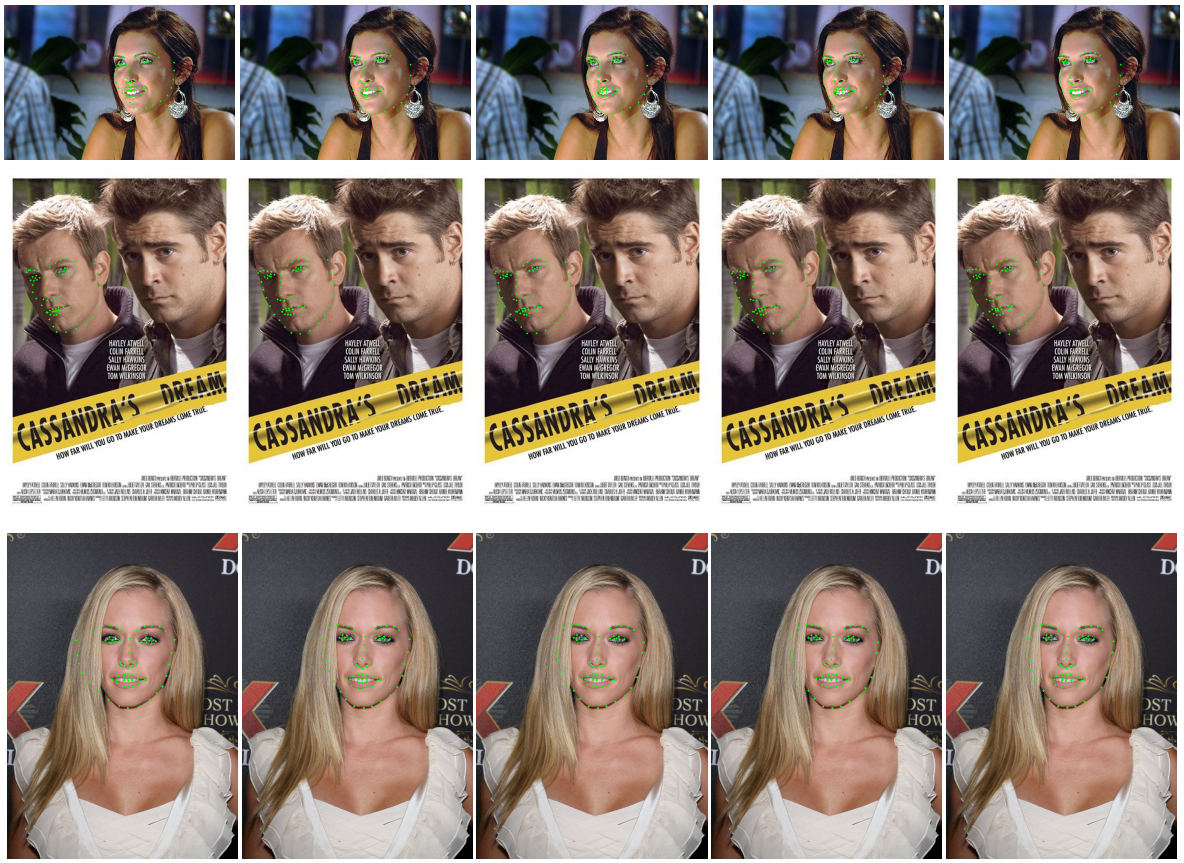
Figure 5: *The cumulative error rates of face alignment results on four benchmarks.*

5. Applications and Discussions

We implemented our method on a PC with Intel Core(TM) i7 950 CPU, 3.06 GHz, 16GB RAM, and nVidia GeForce GTX 770 GPU with 2048 MB video memory.

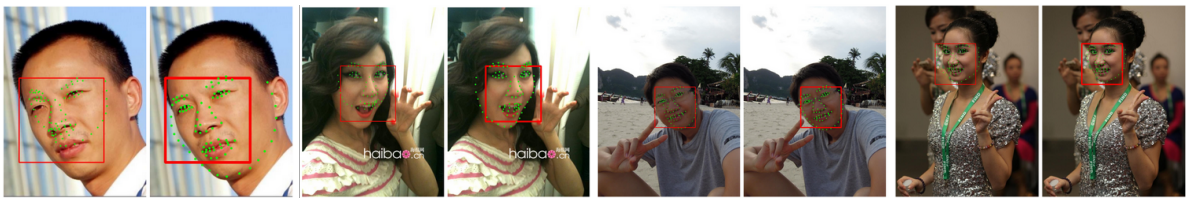
Face alignment enhancement. We integrate our ranking process into the face alignment framework of [CWWS14] to improve the face shape quality. In the original framework, the program run the regressor 5 times for an input image and takes the median result as the final estimation. We improve this framework by running the regressor 10 times in a parallel manner and using the average of the best 5 results (selected by our ranking method) as the final face shape. The total process can cost slightly more than the original framework but can achieve good enhancement while still achieving real-time performance. In our experiments, the average ranking time by using listwise models is 7 msec, while the average ranking time by using pairwise model is 68 msec. The cumulative error curves are shown in Figure 5. The cumulative errors become smaller after integrating our ranking mechanism into the original framework. Figure 6 has some examples showing that our method can generate better results than the original face alignment method in [CWWS14] after adding result ranking step.

In addition to the four standard benchmarks, our ranking model was also tested on another image set with 600 images collected from the Internet. By using $RMSE \leqslant 5$ as the standard, we obtained the result from the original ESR showing that the face shapes of 93.0% of the images are good enough to be accepted, whereas after integrating our ranking model, the percentage of the good face shapes is increased to 97.5%, which is an obvious improvement. More results are shown



(a) Original ESR [CWWS14] (b) Integrating ListNet (c) Integrating ListMLE (d) Integrating RankCosine (e) Integrating Random Forest

Figure 6: *Comparison of the face alignment results without and with ranking.*



(a) Avoid a bad case (b) Avoid a bad case (c) Result improvement (d) Result improvement

Figure 7: *Some typical examples improved by our ranking model, indicating that our model can enhance the original ESR. The red rectangular box is set for face detection.*

in Table 2. Figure 7 shows, some typical results that are obviously improved through our ranking model. Not only can our model help avoid some bad cases, but it also can improve some unsatisfactory results.

Although this paper does not provide a new face alignment method per se, it provides a way to better leverage multiple existing results. As shown in our experiments, our method can reliably select the good face alignment results for each

Table 2: *The Distribution of RMSE of the original ESR and the ESR enhanced with the ranking models.*

Method	<= 3	<= 5	<= 7	<=9
Original ESR	74.8%	93.0%	97.5%	98.3%
ESR with Ranking	80.5%	97.5%	99.8%	99.8%

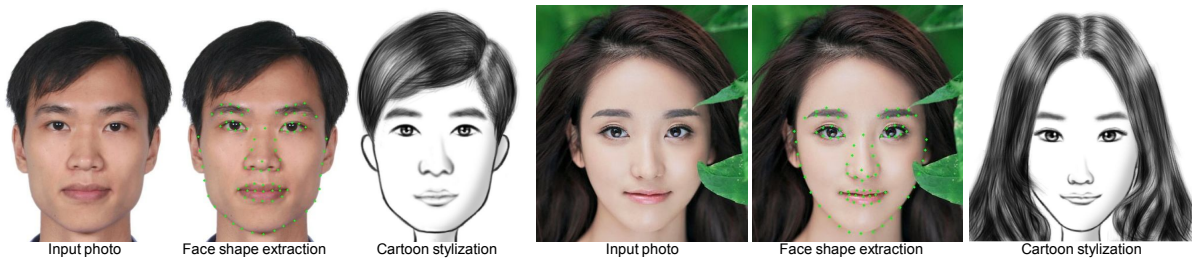


Figure 8: *Face cartoon stylization. Our ranking-enhanced face alignment method helps to accurately extract the facial components from the input portrait photo.*

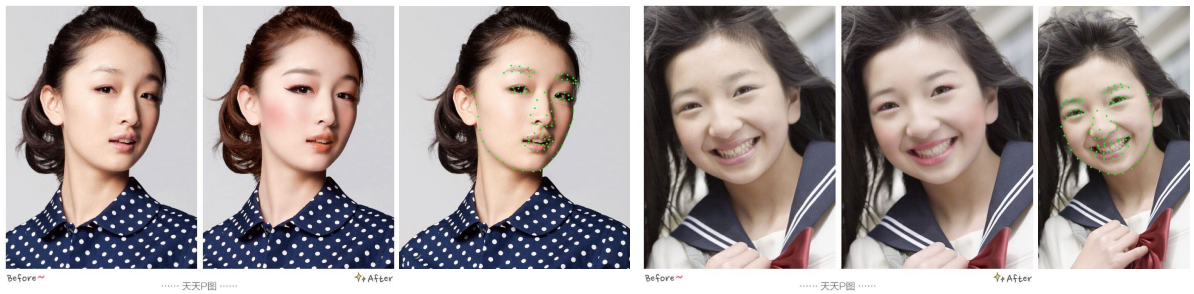


Figure 9: *Digital face makeup. Our face alignment method can accurately extract face shapes from images of different poses.*

input image and combine them together to obtain an optimal result.

Face cartoon stylization. Stylized cartoon faces are widely used as virtual personal images in social media such as instant chat, photo albums, or Twitter. An automatic system for cartooning based on an input photography is very useful for many practical multimedia applications. We enhance the face cartoon stylization system described in [ZDD*14] by integrating our face alignment algorithm into its face parsing step. Figure 8 shows two results of face cartoon stylization. Given an input portrait, we can automatically generate a cartoon face according to the facial features.

Digital face makeup. Face makeup is a technique used to change one's appearance with special cosmetics such as foundation, powder, cream, and other products. In most cases, especially for females, makeup is used to enhance one's appearance. Face alignment is a key technology in a digital face makeup system, which can create makeup on a face image [GS09, SPB*14]. We develop a mobile application for digital face makeup for Android and iOS platforms. Given an input portrait, with the aid of face alignment, our system first extracts the facial components including eyes, eyebrows, cheeks, lips and hair, by using the method in [ZDD*14]. Second, the user can interactively add virtual cosmetics (e.g., lip gloss, eye shadow, and cheek color), to decorate the face in the image. We show two digital face makeup results generated by our product in Figure 9. Our digital face makeup system can be downloaded from the website: http://tu.qq.com/ (the software interface is in Chinese).

Limitations. A limitation of our method is evident when used for face alignment: the approach is only able to rank existing face alignment results by using ranking models based on the assumption that at least one of them is good. The approach will fail when the quality of all the candidates are not good. As shown in Figure 10, the ranking-enhanced results are still not very satisfactory because all the candidates are not good. Another limitation of our approach is the top-k ranked results combined to generate the final face alignment result, by simply calculating the average value (point-by-point) of the points of the top-k results. This approach is not always the best scheme for results combination. We may need a better model to make a better face shape when given a ranking list that, will use more information of the face shapes and will create the final feature points more plausible.

6. Conclusions and Future Work

This study developed a data-driven approach for comparing face alignment results without the ground truth. We employed a HOG feature to capture face shape quality. The feature is used in our learning-to-rank framework to produce the face shape ranking for each input image. Experiments on several face alignment benchmarks show that our method can produce ranking results that correlate well with the ground

(a) Original ESR [CWWS14] (b) Integrating ListNet (c) Integrating ListMLE (d) Integrating RanCosine (e) Integrating Random Forest

Figure 10: *Comparison of the face alignment results without and with ranking.*

truth ranking. Our method can be used to select good face shapes for each input face image, which can improve the overall face alignment performance, especially for explicit regression based methods.

For future work, we will extend our method to rank the face alignment results generated by different methods. Removing poor candidates from face alignment enhancement application and combining the top-k ranked candidates to generate an optimal face alignment result are also challenging.

Acknowledgements

This work was supported by National Science and Technology Major Program for Core Electronic Devices, High-end Generic Chips and Basic Software Project of China under Grant No. 2012ZX01039-004-45, by National Natural Science Foundation of China under nos. 61172104, 61271430, 61201402, and 61372184), and by CASIA-Tencent BestImage joint research project.

References

[AZCP14] ASTHANA A., ZAFEIRIOU S., CHENG S., PANTIC M.: Incremental face alignment in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2014), pp. 1859–1866. 1

[BJKK11] BELHUMEUR P., JACOBS D., KRIEGMAN D., KUMAR N.: Localizing parts of faces using a consensus of exemplars. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2011), pp. 545–552. 1, 4

[Bre01] BREIMAN L.: Random forests. *Machine learning 45*, 1 (2001), 5–32. 5

[CC07] CRISTINACCE D., COOTES T.: Boosted regression active shape models. In *Proceedings of the British Machine Vision Conference* (2007), vol. 2, BMVA Press, pp. 880–889. 1

[CCC14] CHIH-CHUNG CHANG C.-J. L.: Libsvm. *ACM Transactions on Intelligent Systems and Technolotgy 2*, 27 (April 2014), 27:1 – 27:27. 5

[CET01] COOTES T., EDWARDS G., TAYLOR C.: Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence 23*, 6 (Jun 2001), 681–685. 1

[CHZ14] CAO C., HOU Q., ZHOU K.: Displaced dynamic expression regression for real-time facial tracking and animation. *ACM Transactions on Graphics 33*, 4 (July 2014), 43:1–43:10. 1

[CQL*07] CAO Z., QIN T., LIU T.-Y., TSAI M.-F., LI H.: Learning to rank: From pairwise approach to listwise approach. In *Proceedings of the 24th International Conference on Machine Learning* (New York, NY, USA, 2007), ICML '07, ACM, pp. 129–136. 3, 4

[CTCG95] COOTES T., TAYLOR C., COOPER D., GRAHAM J.: Active shape models-their training and application. *Computer Vision and Image Understanding 61*, 1 (1995), 38 – 59. 1

[CWWS14] CAO X., WEI Y., WEN F., SUN J.: Face alignment by explicit shape regression. *International Journal of Computer Vision 107*, 2 (2014), 177–190. 1, 2, 3, 5, 7, 8, 10

[DGFVG12] DANTONE M., GALL J., FANELLI G., VAN GOOL L.: Real-time facial feature detection using conditional regression forests. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2012), pp. 2578–2585. 1

[DT05] DALAL N., TRIGGS B.: Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2005), vol. 1, pp. 886–893. 3

[DWP10] DOLLAR P., WELINDER P., PERONA P.: Cascaded pose regression. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2010), pp. 1078–1085. 1

[GES12] GAO H., EKENEL H., STIEFELHAGEN R.: Face alignment using a ranking model based on regression trees. In *Proceedings of the British Machine Vision Conference* (2012), BMVA Press, pp. 118.1–118.11. 2

[GS09] GUO D., SIM T.: Digital face makeup by example. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2009), pp. 73–79. 9

[HLW04] HUANG X., LI S., WANG Y.: Statistical learning of evaluation function for asm/aam image alignment. In *Biometric Authentication* (2004), vol. 3087 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 45–56. 2, 3

[HMYL15] HSIEH P.-L., MA C., YU J., LI H.: Unconstrained realtime facial performance capture. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2015), pp. 1675–1683. 1

[Ken38] KENDALL M. G.: A new measure of rank correlation. *Biometrika 30*, 1/2 (June 1938), 81–93. 5

[KS14] KAZEMI V., SULLIVAN J.: One millisecond face alignment with an ensemble of regression trees. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2014), pp. 1867–1874. 1

[KWRB11] KOSTINGER M., WOHLHART P., ROTH P., BISCHOF H.: Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)* (Nov 2011), pp. 2144–2151. 1

[LBL*12] LE V., BRANDT J., LIN Z., BOURDEV L., HUANG T.: Interactive facial feature localization. In *Computer Vision - ECCV 2012* (2012), vol. 7574 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 679–692. 1, 4

[Liu07] LIU X.: Generic face alignment using boosted appearance model. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2007), pp. 1–8. 2

[LLML09] LAN Y., LIU T.-Y., MA Z., LI H.: Generalization analysis of listwise learning-to-rank algorithms. In *Proceedings of the 26th Annual International Conference on Machine Learning* (New York, NY, USA, 2009), ICML '09, ACM, pp. 577–584. 3

[LO09] LIU L., ÖZSU M. T.: *Encyclopedia of Database Systems*. Springer, 2009. 6

[LTO*15] LI H., TRUTOIU L., OLSZEWSKI K., WEI L., TRUTNA T., HSIEH P.-L., NICHOLLS A., MA C.: Facial performance sensing head-mounted display. *ACM Transactions on Graphics 34*, 4 (July 2015). 1

[LWC*13] LIU Y., WANG J., CHO S., FINKELSTEIN A., RUSINKIEWICZ S.: A no-reference metric for evaluating the quality of motion deblurring. *ACM Transactions on Graphics 32*, 6 (Nov. 2013), 175:1–175:12. 2

[ML14] MAI L., LIU F.: Comparing salient object detection results without ground truth. In *Computer Vision - ECCV 2014* (2014), vol. 8691 of *Lecture Notes in Computer Science*, Springer International Publishing, pp. 76–91. 2, 3, 5

[Mur12] MURPHY K. P.: *Machine Learning, A Probabilistic Perspective*. MIT, 2012. 4

[QZT*08] QIN T., ZHANG X.-D., TSAI M.-F., WANG D.-S., LIU T.-Y., LI H.: Query-level loss functions for information retrieval. *Information Processing and Management 44*, 2 (Mar. 2008), 838–855. 3, 4

[SCT11] SAUER P., COOTES T., TAYLOR C.: Accurate regression procedures for active appearance models. In *Proceedings of the British Machine Vision Conference* (2011), Hoey J., McKenna S., Trucco E., (Eds.), BMVA Press, pp. 30.1–30.11. 1

[SPB*14] SHIH Y., PARIS S., BARNES C., FREEMAN W. T., DURAND F.: Style transfer for headshot portraits. *ACM Transactions on Graphics 33*, 4 (July 2014), 148:1–148:14. 9

[STZP13a] SAGONAS C., TZIMIROPOULOS G., ZAFEIRIOU S., PANTIC M.: 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *IEEE International Conference on Computer Vision Workshops (ICCVW)* (Dec 2013), pp. 397–403. 1, 4

[STZP13b] SAGONAS C., TZIMIROPOULOS G., ZAFEIRIOU S., PANTIC M.: A semi-automatic methodology for facial landmark annotation. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (June 2013), pp. 896–903. 1, 4

[SWT13] SUN Y., WANG X., TANG X.: Deep convolutional network cascade for facial point detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2013), pp. 3476–3483. 1

[WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing 13*, 4 (Apr. 2004), 600–612. 2

[WLD08] WU H., LIU X., DORETTO G.: Face alignment via boosted ranking model. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2008), pp. 1–8. 2, 3, 5

[XDLT13] XIONG X., DE LA TORRE F.: Supervised descent method and its applications to face alignment. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2013), pp. 532–539. 1

[XLW*08] XIA F., LIU T.-Y., WANG J., ZHANG W., LI H.: Listwise approach to learning to rank: Theory and algorithm. In *Proceedings of the 25th International Conference on Machine Learning* (New York, NY, USA, 2008), ICML '08, ACM, pp. 1192–1199. 3, 4

[YLYL13] YAN J., LEI Z., YI D., LI S.: Learn to combine multiple hypotheses for accurate face alignment. In *IEEE International Conference on Computer Vision Workshops (ICCVW)* (Dec 2013), pp. 392–396. 1

[ZDD*14] ZHANG Y., DONG W., DEUSSEN O., HUANG F., LI K., HU B.-G.: Data-driven face cartoon stylization. In *SIGGRAPH Asia 2014 Technical Briefs* (New York, NY, USA, 2014), SA '14, ACM, pp. 14:1–14:4. 9

[ZR12] ZHU X., RAMANAN D.: Face detection, pose estimation, and landmark localization in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2012), pp. 2879–2886. 1, 4

[ZZCM08] ZHANG J., ZHOU S. K., COMANICIU D., MCMILLAN L.: Discriminative learning for deformable shape segmentation: A comparative study. In *Computer Vision - ECCV 2008* (2008), Forsyth D., Torr P., Zisserman A., (Eds.), vol. 5302 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 711–724. 2

[ZZD*14] ZHANG Z., ZHANG W., DING H., LIU J., TANG X.: Hierarchical facial landmark localization via cascaded random binary patterns. *Pattern Recognition* (2014). 1