Available online at www.sciencedirect.com

ScienceDirect

Procedia Technology 24 (2016) 1349 – 1357

Procedia Technology

International Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015)

Texture-based estimation of age and gender from wild conditions

Aswathy Unnikrishnan*,Ajesh F, Dr. Jubilant J Kizhakkethottam

*MTech Scholar, MCET,Pathanamthitta
Department of CSE, MCET, Pathanamthitta

Abstract

The paper concerns the estimation of facial attributes—namely, age and gender—from images of faces acquired in challenging, in the wild conditions. This problem has received far less attention than the related problem of face recognition, and in particular, has not enjoyed the same dramatic improvement in capabilities demonstrated by contemporary face recognition systems. Here, this problem is addressed by making the following contributions. First, in answer to one of the key problems of age estimation research—absence of data—a unique data set of face images, labelled for age and gender is offered, acquired by smart-phones and other mobile devices, and uploaded without manual filtering to online image repositories. The images in this collection are more challenging than those offered by other face-photo benchmarks. Second, a dropout-support vector machine approach is described used by this system for face attribute estimation, in order to avoid overfitting. Inorder to make classification of age using kNN more easy, texture features are extracted. Finally, a robust face alignment technique is presented, which explicitly considers the uncertainties of facial feature detectors.

Peer-review under responsibility of the organizing committee of ICETEST – 2015

Keywords: Face recognition; Dropout; Support vector machine; KNN;overfitting

* Aswathy Unnikrishnan. Tel.: 9946723757,fax:0468231702
*E-mail address:*aswathy.unni19@gmail.com

1. Introduction

1.1 Age and gender estimation

Among several faces that come across our lives, some might be familiar and the rest unfamiliar. While addressing these faces, probably “Sir/Madam” is used based on their gender. Similarly age classification is also

Peer-review under responsibility of the organizing committee of ICETEST – 2015
doi:10.1016/j.protcy.2016.05.145

an important factor in formal addressing of a person. An elder one would be more formally addressed than the younger ones. Recently, there has been growing interest in human age estimation due to its potential applications, such as human-computer interaction and electronic customer relationship management (ECRM). The task of human age estimation is to estimate a person's exact age or age-group based on face images. If these tasks of age and gender estimation get computerized with similar accuracy and effortlessness as humans, it would be more easy and convenient, since role of computers grow in our lives. Many efforts have been executed previously in this area. But most of the works were done using images from Public figures dataset and group photos which purposefully posed for the click. Compared to this, considering images from wild conditions has been a challenging task. As it may possess poor lighting, sideways facing subjects, motion blur etc, it could reflect the challenges of real-world age and gender estimation tasks. Such images can be obtained from online image albums captured by smart-phones and automatically uploaded to Flickr before being manually filtered or curated.

The issue of over-fitting i.e., when predictor is too complex or it fits 'noise' in the training data or it makes more mistakes in the training data, could be avoided by training the classifier. In short, dropout training for Support Vector machine (SVM) washes out the problem of over-fitting.

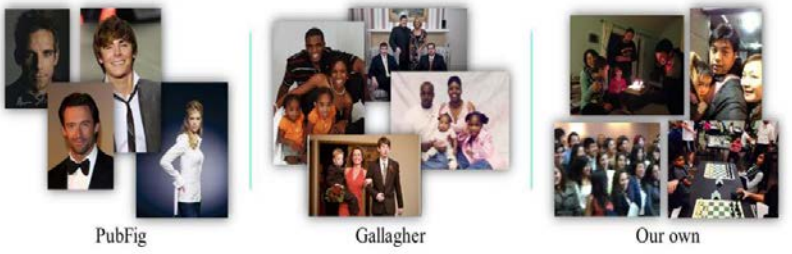


Fig.. 1. Example images from two existing relevant collections and Adience set. Left: PubFig benchmark images. Despite being considered "in the wild", these images are often clean in terms of viewing conditions, and enjoy participation from the subjects being photographed. Mid: The Gallagher collection provides images with an intentional bias towards groups of people, typically facing the camera and posing for their shots. Right: Images from our collection automatically uploaded to Flickr, without manual pre-filtering by their owners. Consequently, they include sideways facing subjects, motion blur, poor lighting and more, all of which present additional challenges to automated face analysis systems.

2. Literature review

Estimating the age and gender of a person appearing in a photo from that person's facial features, has been studied at length in the past though, far less than the related problem of age-gender recognition.

Face detection: Performing face detection and facial trait in realistic scenarios presents significant challenges[2] particularly in terms of achieving robustness over large changes in a person's viewpoint (head pose), various face scales, non-uniform illumination conditions, partial occlusion and overcoming potentially noisy images or false face detection. Age appearances and variations can be modeled using example images. In particular, [3] used local binary patterns (LBP) [4] which is a powerful feature for texture classification. Gabor features were used by [5], both Gabor and LBP features were recently used. Face alignment is another technique used in order to detect faces more clearly. Affine transformation, which contributes geometric transformations such as scaling, translation, rotation is used for this alignment purpose.

Face discrimination: Some of the most basic classifications regarding humans are gender and age. There is an important difference between age and gender in classification: for gender, there are two clearly distinguished classes (male and female), while for age estimation, the values to predict are continuous, thus the estimation method needs to be different. For this reason, the gender is estimated with a classification method and the age with a regression method. Age estimation, depending on the application domain, can be a regression problem or a multi-label classification task.

Earlier works concentrated on neural networks for classification. Deep neural networks contain multiple non-linear hidden layers and this makes them very expressive models that can learn very complicated relationships between their inputs and outputs. Support vector machines were used to classify in [7], where rank relationship of ages is learned. When a test image given to ranking SVM a weight vector is obtained and projects images with that weight vector to get ranking value. More recently, some have suggested different ways of partitioning the space of face images based on age like the ordinal subspaces approach of [8], which uses a flat partitioning scheme. Finally, others have proposed hierarchical partitioning models, including [9], which use an "AND-OR" graph partitioning of age progression. An exemplar aging sequence represented with a parse graph and face aging is modelled as Markov model.

All methods go through the difficulty of dealing with over-fitting. To tackle the problem, 'Dropout' technique was defined in [10] where the idea was to randomly drop units from neural networks during training so as to prevent units from co-adapting too much. Then the technique was tried for SVM in [11] in which IRLS(Iteratively reweighted least square) algorithm was used which minimize variational bounds. It was proved that dropout training can significantly boost the classification performance for simple linear SVM.

Gender classification: Gender is an important demographic attribute of people. Gender classification has received considerable attention over the years. For a rigorous survey of the methods developed for this problem over the years, refer to [12] or the more recent [13]. Face image intensities were directly classified using SVM in [14] and later again using AdaBoost in [15] where Adaptive boosting (Adaboost) select only those features which are relevant but is sensitive to noise and outliers. Also using AdaBoost, [16] used local binary patterns (LBP) [4] rather than intensities. Real-time performance when classifying gender in real-world images was the emphasis in [17]. Somewhat related to the work here, LBP was used along with SVM classifiers in [18].

Benchmarks: As mentioned earlier several works have been executed on estimation of age and gender. They all experimented with different sort of datasets, but mostly camera-posed images. Possibly the most well-used benchmark for age estimation has been FG-NET aging set [19]. It consists of about 1,000 images of 82 subjects, labeled for accurate age. It reflected a fewer challenges than those expected of modern face recognition systems.

Another popular benchmark used by many in the past is the MORPH set [20], collected by the Face Aging Group at the University of North Carolina at Wilmington. It contains over 55,000 images of 13,000 individuals. It too contains images under highly controlled viewing conditions. Over the years, performance on this set has also saturated, with systems demonstrating performances reaching near-perfect scores.

The UIUC-IFP-Y Internal Aging extensively used by the SMILE lab at Northeastern University, is not publicly available due to intellectual property limitations. It offers 8,000 images of 1,600 voluntary Asian

subjects (half male, half female) in outdoor settings. This set too, was produced under lab-controlled conditions and so unsurprisingly, performance measured by mean average age prediction error on this set has been reported to be near perfect.

Recently, following the shift towards face recognition "in the wild" (e.g., the LFW set [21]), benchmarks for age and gender estimation have also been assembled using unconstrained images. The first, originally designed for face recognition, is the Public Figures benchmark (PubFig). It includes images from news and media websites which are typically of high quality, with subjects collaborating with the camera, posing for the shot [22]. Its construction emphasized many images for each individual, and so it includes nearly 60,000 photos of only 200 celebrity faces.

In [23], Gallagher and Chen proposed a benchmark for the study of groups of people, posing for the camera (e.g., family photos). Photos in this collection therefore typically present multiple subjects, in forward facing (towards the camera) poses, each face in relatively low resolution. The age labels provided in this set make it a convenient choice for studying age estimation.

Finally, the VADANA set was recently proposed in [24]. With 2,298 images of 43 subjects, it is substantially smaller than its recent predecessors, but unlike them, provides multiple images of the same subjects in different ages, allowing for the study of age progression of the same face.

3. Existing system

In the work done in [1], the system for age and gender estimation follows a pipeline consisting of detection, alignment and identification. For detection and alignment purposes the standard Viola and Jones face detector [25] and affine transformation are applied respectively. The representation method holds with encoding the aligned faces using several popular global image representations specifically, local binary patterns (LBP) and Four Patch LBP codes (FPLBP). Classification is performed using the feature vector representations and each descriptor is examined either independently or by combining multiple descriptors by concatenating them into single long feature vectors. Training is performed using dropping out many of the values from the training instances thus, reducing the risk of over-fitting. Simpler classifiers have shown well for these problems in the past [26]. Classification of gender is performed using a single linear SVM classifier, for the multi-label age classification, a one-vs-one linear SVM arrangement is used.

4. Proposed System

4.1 Overview of the approach

In the proposed system too it follows a pipeline which consists of detection, alignment, and identification which include representation and classification. As a key design choice, the system is modelled after similar systems successfully applied for face recognition.

Detection and alignment- Given a photo, the process begins by applying the standard Viola and Jones face detector [25]. Detected faces are then aligned to a single reference coordinate frame using Affine transformation.

Face segmentation-Active contour model face segmentation is applied to the training images of dataset as well as on the test image. Using active contour based segmentation the 2-D grayscale image is segmented into

foreground (object) and background regions. The output image is a binary image where the foreground is white (logical true) and the background is black (logical false). It provides a binary image that specifies the initial state of the active contour. The boundaries of the object region(s) (white) in this image define the initial contour position used for contour evolution to segment the image.

Representation-In this step, aligned faces are encoded using several popular global image representations. This system applies local binary patterns (LBP) of [27] the related Four Patch LBP codes (FPLBP). These were selected due to their successful application to face recognition problems, as well as their efficient computation and representation requirements.

Addition of texture features- Gray level co-occurrence matrices (GLCM), which have been used very successfully for texture classifications in evaluations is been used here. It is used to extract second order statistics from an image. It is a matrix of frequencies at which two pixels, separated by a certain vector, occur in the image. Once the GLCM has been created, various features can be computed from it. Here following features can be calculated from this co-occurrence matrix.

- Contrast -Measures the local variations in the gray-level co-occurrence matrix.
- Correlation-Measures the joint probability occurrence of the specified pixel pairs.
- Energy- Provides the sum of squared elements in the GLCM. Also known as uniformity.
- Homogeneity-Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

Classification- Since gender has only two distinguished classes, it is better to choose standard linear SVM trained using the feature vector representations. Here, each descriptor is examined independently or multiple descriptors are combined by concatenating them into single long feature vectors. Training is performed in order to avoid over-fitting, using the dropout scheme. Age classification, since it is a multi-class problem, k-nearest neighbor classification method is used. When given an unknown tuple, a k-nearest neighbor (k-NN) classifier searches the pattern space for the k training tuples which are closest to the unknown tuple. These k training tuples are the k-nearest neighbours of the unknown tuple. Here, since k>1 and dropout mechanism is applied, the system is less prone to over-fitting. The results show that excellent results may be obtained without apparent over-fitting by training these classifiers by dropout SVM.

4.2Face Alignment with Uncertainty

Here, the robust facial detector proposed by Zhu and Ramanan [28] is employed. It detects 68 specific facial features, including the corners of the eyes and mouth, the nose and more. By selecting ideal coordinates for each of these points an affine transformation can presumably be obtained and the images aligned. In practice, however, errors in point localizations as well as the variability of face shapes can often result in unstable alignment results. To address this, it is noted that some detections are more reliable than others: The corners of the eyes, for examples, are easier to localize than, say, the cheekbones. In order to accurately align faces, these uncertainties should be accounted for. Doing so requires knowledge of the uncertainty associated with each of the 68 features, but this information can only be estimated once the faces are already aligned. In order to resolve this chicken-and-egg problem, we take an Iterative Re-weighted Least Squares (IRLS) approach [29].

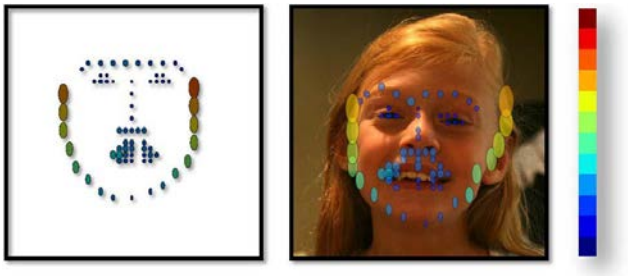


Fig. 2. **Visualization of the uncertainties associated with the 68 facial feature detections.** Left: Uncertainties on the reference coordinate frame. Mid: Uncertainties on a sample face image. Note that ellipses are aligned with the image axes, as we assume variance along the axes is uncorrelated. In both images color codes the amount of uncertainty; color-bar provided on the right.

Specifically we assume facial feature points $\{r_i\}_{i=1..68}=\{(x^r_i, y^r_i)\}_{i=1..68}$ in a frontal-facing face. Given corresponding feature points, $\{q_{i,j}\}_{i=1..68,j=1..N} = \{(x^q_{i,j}, y^q_{i,j})\}_{i=1..68,j=1..N}$ (for training N photos), detected on a query photo. First H^0_j, an initial time-0 affine transformation, relating facial feature points $q_{i,j}$ is detected in the j'th query photo with their corresponding reference points, r_i, by using standard least squares. Feature points in all queries are then projected using the standard

$$q'_{i,j} = H^0_j q_{i,j}$$

For a given facial feature i , for all images j , we consider the variance along the x axis and the variance along the y axis of these projected points as the uncertainty values associated with this feature. These are used to estimate a new aligning transformation, H^1_j at time-1, this time, by weighted least squares. This process is repeated until convergence.

4.3Dropout-SVM

The issue of over-fitting i.e., when predictor is too complex or it fits 'noise' in the training data or it makes more mistakes in the training data, could be avoided by training the classifier. In short, dropout training for Support Vector machine (SVM) [11] washes out the problem of over-fitting. This approach was inspired after the recent success of dropout learning for deep neural networks [10]. The idea was to randomly drop units from neural networks during training so as to prevent units from co-adapting too much. While training, dropout essentially omits neurons from the network with some probability p_{drop} for each feature, in each sample, to be dropped. By doing so, it was claimed that neurons must better adapt to the input data, relying less on other neurons in the network and the representations thus obtained are distributed and better generalized.

Then the technique was tried for SVM, in which IRLS (Iteratively reweighted least square) algorithm was used which minimize variational bounds. Since SVM can be considered as a single-layer neural network, a similar dropout procedure can be conceivably applied to train SVM classifiers. It was proved that dropout training can significantly boost the classification performance for simple linear SVM.

Here, each training features are applied with the value zero at random. This random selection is applied to each training instance separately; different features are randomly selected and set to zero for different training instances. Dropping out many of the values from the training instances requires that the obtained model be modified accordingly. Specifically, following training with a dropout of rate of p_{drop}, we divide all the

coefficients of the resulting linear SVM model by $(1 - p_{drop})$. This compensates for the dropped-out values, and provides a model suitable for test instances which include all their values.

4.4The Adience benchmark

To facilitate the study of age and gender recognition, the key design principle is to use the data which is as true as possible to challenging real-world conditions. As such, it should present all the variations in appearance, noise, pose, lighting and more, that can be expected of images taken without careful preparation or posing. The database collection process and testing protocols can be discussed as follows:-

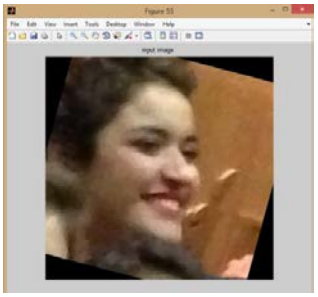
Database Preparation and Contents: Photos from challenging real-world conditions could be collected from Flickr albums produced by automatic upload from iPhone 5 or later smart-phones. The following steps are included in image collection. Photos downloaded from Flickr albums were processed by first running Viola and Jones face detector and then detecting facial feature points. Presumably due to the recent "selfie" trend, many faces in these albums appear at different roll angles. To avoid missing these faces, the face detection process was applied to each image, rotated 360° degrees in 5° increments.

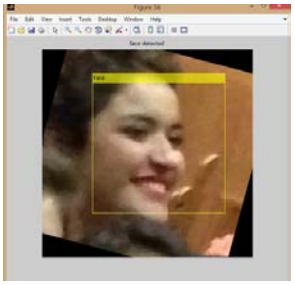
Benchmark Protocols: Some test protocols are defined to benchmark the performance of gender and age estimation techniques, using Adience collection. Two variations of data set can be used: the frontal and the complete sets. The frontal set includes only roughly frontal facing faces; that is, faces which were determined to be within ±5° yaw angle from a forward facing face. The complete set includes faces with up to ±45° degrees of yaw. Training and testing is performed using 5-fold cross validation with splits pre-selected to eliminate cases of images from the same Flickr album appearing in both training and testing sets in the same fold. Same splits for both the gender and the age classification tasks are used. Results therefore include both mean classification accuracy (age or gender), including, ± standard error over the five folds.

5. Conclusion

This paper addresses the problem of automatic age, gender, and estimation from real-life face images acquired in unconstrained conditions. A new and extensive data set and benchmark for the study of age and gender estimation and a classification pipeline is contributed in this paper. The pipeline consists of detection, alignment, segmentation, texture feature addition and identification. In addition a novel, a robust face alignment technique based on iterative estimation of the uncertainties of facial feature localizations is described. Here face segmentation is performed using active contour based segmentation which makes classification easier. Along with this, texture features, the GLCM features are added which contributes to age estimation. Age and gender classification is performed using dropout trained KNN and SVM respectively.

Appendix

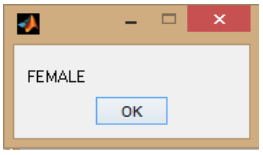




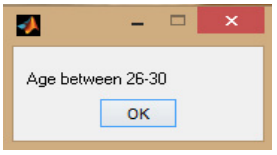
Screenshot 1: Input image

Screenshot 2: Face detected

Screenshot 3: Segmented image



Screenshot 4: Gender detected



Screenshot 5: Age detected

References

1. Eidinger, E., Enbar, R., & Hassner, T. (2014). Age and gender estimation of unfiltered faces. *Information Forensics and Security, IEEE Transactions on*,*9*(12), 2170-2179.
2. S. K. Zhou, R. Chellappa, and Chao, W. L., Liu, J. Z., & Ding, J. J. (2013). Facial age estimation based on label-sensitive learning and age-oriented regression. *Pattern Recognition*, *46*(3), 628-641.1W. Zhao. Unconstrained FaceRecognition. Springer-Verlag New York, Inc., 2005
3. Chao, W. L., Liu, J. Z., & Ding, J. J. (2013). Facial age estimation based on label-sensitive learning and age-oriented regression. *Pattern Recognition*, *46*(3), 628-641.
4. Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *28*(12), 2037-2041.
5. Gao, F., & Ai, H. (2009). Face age classification on consumer images with gabor feature and fuzzy lda method. In *Advances in biometrics* (pp. 132-141). Springer Berlin Heidelberg.
6. Choi, S. E., Lee, Y. J., Lee, S. J., Park, K. R., & Kim, J. (2011). Age estimation using a hierarchical classifier based on global and local facial features. *Pattern Recognition*, *44*(6), 1262-1281.
7. Cao, D., Lei, Z., Zhang, Z., Feng, J., & Li, S. Z. (2012). Human age estimation using ranking SVM. In *Biometric Recognition* (pp. 324-331). Springer Berlin Heidelberg.
8. Chang, K. Y., Chen, C. S., & Hung, Y. P. (2011, June). Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (pp. 585-592). IEEE.
9. Suo, J., Zhu, S. C., Shan, S., & Chen, X. (2010). A compositional and dynamic model for face aging. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *32*(3), 385-401.
10. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929-1958.
11. Chen, N., Zhu, J., Chen, J., & Zhang, B. (2014). Dropout training for support vector machines. *arXiv preprint arXiv:1404.4171*.
12. Mäkinen, E., & Raisamo, R. (2008). Evaluation of gender classification methods with automatically detected and aligned faces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *30*(3), 541-547.
13. Reid, D. A., Samangooei, S., Chen, C., Nixon, M. S., & Ross, A. (2013). Soft biometrics for surveillance: an overview. *Machine learning: theory and applications. Elsevier*, 327-352.
14. Moghaddam, B., & Yang, M. H. (2002). Learning gender with support faces.*Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *24*(5), 707-711.
15. Baluja, S., & Rowley, H. A. (2007). Boosting sex identification performance.*International Journal of computer vision*, *71*(1), 111-119.
16. Sun, N., Zheng, W., Sun, C., Zou, C., & Zhao, L. (2006). Gender classification based on boosting local binary pattern. In *Advances in Neural Networks-ISNN 2006* (pp. 194-201). Springer Berlin Heidelberg.

17. Chen, D. Y., & Lin, K. Y. (2010). Robust gender recognition for uncontrolled environment of real-life images. *Consumer Electronics, IEEE Transactions on*,*56*(3), 1586-1592.
18. Shan, C. (2010, January). Gender classification on real-life faces. In *Advanced Concepts for Intelligent Vision Systems* (pp. 323-331). Springer Berlin Heidelberg.
19. Lanitis, A., & Cootes, T. (2002). FG-Net aging database. *Cyprus College*.
20. Ricanek Jr, K., & Tesafaye, T. (2006, April). Morph: A longitudinal image database of normal adult age-progression. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on* (pp. 341-345). IEEE.
21. Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007). *Labeled faces in the wild: A database for studying face recognition in unconstrained environments* (Vol. 1, No. 2, p. 3). Technical Report 07-49, University of Massachusetts, Amherst.
22. Kumar, N., Berg, A. C., Belhumeur, P. N., & Nayar, S. K. (2009, September). Attribute and simile classifiers for face verification. In *Computer Vision, 2009 IEEE 12th International Conference on* (pp. 365-372). IEEE.
23. Gallagher, A., & Chen, T. (2009, June). Understanding images of groups of people. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 256-263). IEEE.
24. Somanath, G., Rohith, M. V., & Kambhamettu, C. (2011, November). Vadana: A dense dataset for facial image analysis. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on* (pp. 2175-2182). IEEE.
25. Viola, P., & Jones, M. J. (2004). Robust real-time face detection. *International journal of computer vision*, *57*(2), 137-154.
26. Bekios-Calfa, J., Buenaposada, J. M., & Baumela, L. (2011). Revisiting linear discriminant techniques in gender recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *33*(4), 858-864.
27. Ojala, T., Pietikäinen, M., & Mäenpää, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *24*(7), 971-987.
28. Zhu, X., & Ramanan, D. (2012, June). Face detection, pose estimation, and landmark localization in the wild. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (pp. 2879-2886). IEEE.
29. D. B. Rubin, "Iteratively reweighted least squares," in Encyclopedia of Statistical Sciences, vol. 4. Hoboken, NJ, USA: Wiley, 1983, pp. 272–275