International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
***Article Received:** 25 November 2023 **Revised:** 12 December 2023 **Accepted:** 30 January 2024*

Survey on Face Detection & Recognition Techniques

Janhvi Suman
Dept. of Computer Science & Engineering
Indira Gandhi Delhi Technical University for Women
Delhi, India
janhvimisc@gmail.com

Prof. Ela Kumar
Dept. of Computer Science & Engineering
Indira Gandhi Delhi Technical University For Women
Delhi, India
ela_kumar@igdtuw.ac.in

Abstract—Face detection is an extensively investigated subject in the realm of computer vision and holds considerable significance in diverse applications, encompassing human-computer interfaces, video surveillance, security access control systems, video surveillance, and image database management. Numerous face detection methods have already been devised, like Viola-Jones, RCNN, SSD among others. This paper discusses some additions that have been done on the existing models and systems in a bid to produce better outputs, using the standard datasets like WIDERFACE, FDDB etc. The face detection and recognition techniques discussed here employ the following approaches: (i) Mask R-CNN (ii) MTCNN, (iii) Local Binary Pattern Histogram (LBPH), (iv) PCA with Eigenfaces (iv) Weighted Kernel PCA, and (v) VGG architectures such as Siamese-VGG.

Keywords- Face detection, Face recognition, Loss functions, CNNs, YOLOs, PCA

I. INTRODUCTION

Face Detection has been a major theme after the creation of image detection algorithms and models. After the development of image detection algorithms and models, researchers have worked on improving their efficiency. Early research into face detection primarily involved manual feature engineering and the utilization of conventional machine learning algorithms to build effective detectors for face identification. Nevertheless, these approaches have drawbacks, including the intricacy of feature design and sub-optimal detection accuracy [1].

A typical face recognition method consists of these major steps, i.e. face recognition, feature extraction, face recognition and detection processes [2].

As a state-of-the-art field of face recognition, the modeling of loss functions that are margin-based can enhance the differentiation among various classes.while mining-based strategies can highlight misclassified samples and lead to promising results.

The extensive training data, architectural network, and loss functions are the major contributors to the advancement of Convolutional Neural Networks(CNN). In face recognition with CNNs, the network architecture and loss functions are critical factors for success. Recent efforts are directed at development of effective loss functions to enhance discriminative power, particularly in training deep face CNNs.

Even after the creation and calculation of various different algorithms and models with increased efficiency, everyday a new improvement is made to amplify the efficiency obtained in the previous results.

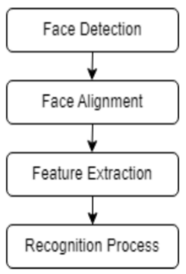

Fig. 1: Face Detection Flowchart

The aim of this paper is to study the accuracy of the face recognition and detection algorithms and models mentioned so as to try to come up with the most efficient technique.

II. BASIC PRINCIPLE

The first and foremost requirement before selecting any algorithm or model for the face detection or recognition is to understand the need of it in the practical world, understanding where its requirement lies and the way it will be used once designed. A standard face recognition system conventionally encompasses four principal phases: face detection, facial alignment, feature extraction, and the recognition process.

Feature selection process should derive important features while reducing redundancy and increasing discriminatory power.

Face tracking predicts upcoming video frames from the preceding frames of the video sequence [3]. It's achieved through methods like skin-color [3] [4] or motion-based [5] approaches, which utilize heuristic knowledge to delineate the search space for efficient tracking. Although many methods for face tracking use only a portion of the face information and rely on certain constraints, such as good lighting or background features of the video, to improve tracking, they still struggle to perform well in complex and dynamic scenes.

There are a lot of techniques through which best results can be obtained but realizing the extent of precision is important.

Afterwards, obtaining the right dataset in the right size which fulfills the demands of the model and the algorithm is necessary. Usually, getting the perfect quantity and quality of the dataset is not possible hence it needs to be altered to requirement.

A. PCA (Principle Component Analysis)

The use of machines to recognize faces from still images is a growing field of research that involves multiple disciplines, including pattern recognition, neural networks, image processing and computer vision [6][7]. WKPCA enhances Kernel PCA with a weight system involving N weight vectors of M dimensions, forming N X N weight and weighted kernel matrices. Optimal weights can be selected using optimization methods to boost recognition performance, similar to PCA but with sorted eigenvalues. While PCA [8] and Kernel PCA [9] are used for feature extraction, this algorithm introduces non-linearity via a Weight Matrix in the Covariance/Kernel Matrix, termed the Kernel Trick.

B. CNN (Convolutional Neural Network)

Object detection models like R-CNN, Fast R-CNN, and Faster R-CNN, while being efficient face detection techniques, often suffer from accuracy issues due to noisy features. Mask R-CNN, an enhanced model, combines face detection and segmentation. It incorporates a mask-branch to predict segmentation of masks for each Region of Interest (RoI), introducing the G-Mask method to improve accuracy. Utilizing ResNet-101 and the Region Proposal Network (RPN), it produces RoIs while preserving spatial data through RoIAlign. Multiple feature maps with varying scales and aspect ratios are used. The Mask Branch employs Full Convolutional Net (FCN) for pixel-wise segmentation. The GIoU metric is introduced for bounding box regression, and two distinct loss functions are employed in the G-Mask method: GIoU for bounding boxes and average binary cross-entropy for segmentation.

Since 2012, CNNs have resurged in face detection, Li et al. [10] introduced a cascaded Convolutional Neural Networks (CNN) method however it was time-consuming, Yang et al. [11] developed Faceness-Net with five CNNs for feature detection, improving accuracy but it was slowing computation, and Zhan et al. [12] combined Adaboost with CNN for face detection. Most methods overlook the inherent link between face detection and landmark localization.

C. LBPH- Based

Today, computer vision is a complex area of programming that involves analyzing image inputs and videos to perform jobs such as recognition, recognition and detection automatically. Here's a new technique for recognition of human faces in real-time in high as well as low-level images, it is based on the Local Binary Pattern Histogram (LBPH) method and is used in a facial recognition system. It encodes the edges in a cost-effective way and maximizes the variation that is important for open edges and facial expression. These highly successful features are then named Local Binary Pattern Histogram(LBPH).

D. VGG- Based

Siamese-Net is an improved method of face tracking that curb the ineffectiveness caused due to fast movement, rotation and occlusion and illumination.

The proposed method, called Siamese-VGG, utilizes the first two convolutional layers of the VGG-16 model to extract features. The researchers experimented with a pre-trained VGG Face model, which was fine-tuned for face tracking. During training in the framework, two branches receive input crops of the same size, and smaller template feature maps are generated to minimize offset losses. Furthermore, the loss function incorporates L2 regularization to enhance the model's capacity for generalization. The results of the proposed improved technique surpasses the original algorithm in terms of robustness & generalization performance. In composite cases, the improved technique produced an average improvement of almost 11%.

Even convolutional neural networks method is not efficient for face tracking purposes in a dynamic setting which includes various disturbances, redundant figures, hence is not used in face tracking.

The Siamese-VGG Network aims at predicting the face frame of the next frame using the face bounding box horizontal and vertical coordinates. This method calculates a similarity score map by analyzing similarities in feature maps. Utilizing smaller and deeper networks along with convolution kernels significantly enhanced the accuracy of both image classification and recognition. In this research, the number of convolutional layers was reduced to one or two due to the replacement of the algorithm's CNN layer with VGG-16.

E. YOLOs

Two techniques, weighted-boxes-fusion (WBF) & non-maximum weighted (NMW), were employed to combine YOLO versions 1 to 4. These methods required assigning weights to each YOLO model, reflecting their effectiveness in the ensemble. This discussion covers four YOLO models implementing WBF and NMW techniques for the WIDER FACE benchmark, resulting in an enhancement of the mean Average Precision (mAP). WBF surpasses NMW, increasing mAP by 7.81%, 22.91%, and 12.96% across easy, medium, and hard instances versus the top YOLOv1 to v4 mAP. Future work involves ensembling YOLOv5 using test-time-augmentation (TTA) techniques. Two techniques, weighted-boxes-fusion (WBF) & non-maximum weighted (NMW) were employed to combine YOLO versions 1 to 4. These methods required assigning weights to each YOLO model, reflecting their effectiveness in the ensemble.

YOLOv1	YOLOv2	YOLOv3	YOLOv4
WIDERFACE	WIDERFACE	WIDERFACE	WIDERFACE
LeakyReLu	LeakyReLu	Logistic Func. [13]	Linear Func. [14]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 12 Issue: 2
***Article Received:** 25 November 2023 **Revised:** 12 December 2023 **Accepted:** 30 January 2024*

GoogleNet [15]	DarkNet	DarkNet53	DarkNet

TABLE 1: Features of each model used

III. METHODOLOGY

a. ***PCA (Principle Component Analysis)***

1) ***PCA using Eigenfaces**:* The eigenface method, commonly integrated with PCA, is a widely adopted technique for developing efficient facial recognition systems. It condenses facial features into a compact set of essential traits known as eigenfaces, which are derived from the primary components of the initial training image series [16].

Fig. 2: Eigenfaces [17] extracted from image

PCA computes the covariance matrix from various sections of a training face image dataset and employs it for facial image recognition, effectively reducing dimensionality.
The outcomes of a facial recognition experiment using PCA and eigenface are evaluated by calculating accuracy, recall, and precision metrics [18].

2) ***Weighted- Kernel PCA**:* The introduction of a weight matrix (W) matching the dimensions of the covariance/kernel matrixleads to the formation of a refined weighted kernel matrix (K'). This transformation improves the subspace for pattern recognition through careful weight selection. Subsequently, eigenvalues and their associated eigenvectors of this adjusted kernel matrix are computed and normalized, creating a nonlinear subspace that can be applied to extract features from training and test datasets.
In Kernel PCA, a weight system is integrated, which involves N weight vectors, each with a dimension of M, corresponding to N face images. A weight matrix in an N x N configuration, along with a weighted kernel matrix are formulated. The paper utilizes Genetic Algorithms (GAs) for global optimization, stressing the significance of optimizing weights in training and testing datasets to enhance pattern recognition, particularly focusing on classification accuracy.
A Simple Encoding Scheme is adopted, representing chromosomes as strings of scalar values, each encoded with 10 bits. During the evolutionary process, chromosomes are selected for a mating pool based on their fitness if the termination criterion is not met. The mating pool undergoes uniform crossover, and mutation, with a typically low probability (Pm), has a minor effect.
The optimal scalars for WKPCA are provided by the string with the highest fitness within the last generation's population.

b. ***CNN (Convolutional Neural Networks)***

1) ***Improved Mask R-CNN**:* This approach introduces G-Mask, which improves image segmentation accuracy. G-Mask comprises two branches, one to detect faces, and the other for segmenting the face along with its surroundings. It uses ResNet-101 for facial feature extraction and employs Region Proposal Network (RPN) for Region of Interest (RoI) generation on the feature map; spatial details are preserved through RoIAlign. At the completion of the process, a bounding box is identified & a Fully Convolutional Network (FCN) in the segmentation branch create resultant face mask, which is then applied to the image.

The RoIAlign layer is introduced to eliminate the coarse quantization of the feature map [19]. It achieves this by utilizing bilinear interpolation to maintain fractional coordinates, ensuring accurate alignment between the Region of Interest (RoI) and the extracted features.

2) ***MTCNN**:* It is a three-stage network for face detection and landmark localization, using a cascaded design for accuracy. It incorporates multi-task learning, fast proposal generation, and post-processing techniques like bounding box regression and non-maximum suppression (NMS) algorithm for efficiency. This framework excels in complex environments. It enhances accuracy with its multi-step approach, combining tasks, speeding up with a proposal network, and refining results through bounding box regression and NMS.

It eliminates non-face areas using the Refinement Network (R-Net), improves candidates, and the Output Network (O-Net) refines further, providing 5 facial landmark positions. The paper introduced an accurate MTCNN-based face detection and landmark localization algorithm, addressing large-angle face detection, highlighting MTCNN's effectiveness in various face-related applications.

c. LBPH- Based

This method integrates the LBPH and HoG algorithms to extract image patterns, set probability thresholds for face detection, and perform face recognition using the sliding window approach. The algorithm operates on grayscale images obtained by converting RGB images. It uses the LBP operator to compare pixel values, minimizing illumination impact and enhancing texture retrieval.

Here, Face recognition is achieved using the Local Binary Pattern (LBP) Algorithm, which reduces the complexity of facial features by applying the LBP operator to local binary patterns [20]. LBP computes a binary ratio of pixel intensities within the central pixel, taking eight adjacent pixels into account.

The face S matrix extracts features from an image by comparing them to the center pixel values, generating binary codes. Specific areas are analyzed using the LBP operator, excluding irrelevant border sections. The feature vector for an image of size CxD is created by calculating LBP codes for all pixels.

To assess image features, comparisons are made between feature vectors obtained from a sample (H) and a model (I) to quantify the differences. Along with the help of histograms, the Log-Likelihood Statistics and Ch-Square Statistics are compared. The value of X2 is taken, greater the value, greater the similarity.

d. VGG

This paper substitutes the original algorithm's CNN layer with the VGG-16 network. Ian's "Deep Learning" [21] proposed the addition of L2 regularization terms to the objective function. This adjustment brings the weights closer to the origin, diminishing their influence, and thereby reducing the model's complexity. This aligns with the Occam's razor principle [22] and results in improved model fitting. L2 regularization is introduced into the Logistic objective function. Experimental results demonstrate that this enhances the model's generalization capabilities and yields superior tracking performance.

e. YOLOs

1) ***YOLOv1**:* YOLOv1, from 2016, inspired by GoogleNet [23], features 24 convolutions, 4 pools, and 2 fully connected layers with Dropout. Leaky ReLU used, except for the final sigmoid layer. Takes 448x448x3 input, yields 7x7x30 grid output. Loss is mean squared error for bounding box coordinates, class probablities, and confidence scores.

2) ***YOLOv2**:* In 2016, YOLOv2 by Joseph Redmon [24], based on DarkNet19, incorporates 23 convolution layers, 5 max-pooling layers, leaky ReLU activation, and Batch normalization. It improves upon YOLOv1 with more bounding box predictions, anchor boxes, and fine-grained features. Input: 416x416x3 image, Output: 13x13x40 grid.

3) ***YOLOv3**:* YOLOv3 [25] employs a 53-layer Darknet framework with skip and residual connections, upsampling, and downsampling. Darknet53 handles extraction of features, while an additional 53 convolutional layers follow. Input is a 416x416x3 image, yielding grids of 13x13x255, 26x26x255, and 52x52x255 in the output.

4) ***YOLOv4**:* The YOLOv4 model includes improvements such as CIoU loss, Mish activation, DropBlock regularization, WRC, CMBN, Mosaic data augmentation,CSP, SAT, and [26]. These enhancements notably boost its accuracy, surpassing prior versions.

IV. RESULTS & CONCLUSIONS

A. PCA (Principle Component Analysis)

The assessment of WKPCA involves a composite database that combines the Cambridge ORL, Yale, and UMIST databases. This merged dataset contains 575 testing samples and 555 training images, covering 75 different classes.
The experimental result outcomes from a combined database demonstrate that the WKPCA approach beats PCA and KPCA in terms of recognition accuracy for face recognition.

B. CNN (Convolutional Neural Networks)

This comparative experiment involved three popular face benchmark datasets, one of them being FDDB. It is a widely recognized dataset and benchmark for evaluating face detection, containing a total of 5,171 human faces. Comparisons made using various methods on the FDDB dataset, including Faster R-CNN [27], Mask R-CNN [28], Pico [29], Viola-Jones [30], and Koestinger [31].
G-Mask on FDDB showed excellent face segmentation and outperformed other methods with only 1500 false positives & a true positive rate of 88.80%.

Algorithm for landmark localization & face detection based on MTCNN, achieved high accuracy in evaluations with benchmark datasets. A non-frontal face detector addressed the missing large-angle faces through training, enhancing multi-view detection. The study underscores MTCNN's effectiveness and its potential applications in various face-related technologies.
The frontal face detection method, when compared with methods like RCPR [32], TSPM [33], Luxand face SDK [34], ESR [35], CDM [36], SDM [37], and TCDCN [38] on ALFW and other advanced methods on FDDB outperformed all others, achieving over 90% accuracy.

C. LBPH- Based

The face recognition algorithm is divided into three separate and independent sections. All 2000 images of four subjects are organized in a shared folder, each image receives a sample number & a subject ID. The proposed LBPH-Based approach for face detection & image recognition after being put into practice on surveillance cameras, consistently yielded positive outcomes in various experimental assessments, including challenging scenarios involving occlusion, changes in facial pose, and fluctuations in lighting conditions.

D. VGGs

VGG-16, featuring three fully connected layers & five convolutional layers, was employed solely for feature extraction in the study. Owing to GPU memory constraints, the comparison of face tracking performance was limited to 2 to 4 convolution layers. The proposed face tracking algorithm achieved an average overlap of up to 80% across various complex backgrounds, significantly surpassing the performance of the original algorithm.

In future research on the suggested algorithm, enhanced GPU performance could potentially yield superior results by leveraging the full set of available convolutional layers in the algorithm.

E. YOLOs

We assessed YOLO versions 1 through 4 by comparing NMW and WBF ensembles using mAP on the WIDER FACE dataset. YOLOv3 showed the best performance. YOLOv4 achieved higher mAP when we used a larger input size. YOLOv1 had the smallest output and depth. YOLOv2 and YOLOv3 improved by using anchor boxes and larger outputs. YOLOv4 gained even better results with mish activation and a larger output grid. The ensemble of these versions resulted in a higher mAP, indicating their combined strengths.

NMW and WBF improved mAP. After trying different weights and thresholds, we settled on 0.6 IoU, 0.1 skip, and weights of 0.1, 0.5, 0.8, and 0.9 for YOLO v1 to v4. YOLOv4 had the highest weight due to its top mAP. WBF outperformed NMW, increasing mAP by 7.81% for easy cases, 22.91% for medium cases, and 12.96% for difficult cases., surpassing individual YOLOv1 to v4 by 6.25%, 20.83%, and 11.11%.

Given the efficacy of existing methods and algorithms in delivering precise outcomes with notable performance, the primary area of focus for forthcoming research should be enhancing the efficiency of implementation to achieve faster execution.

REFERENCES

[1] N. Liu, H. Wang, and W.-Y. Yau, “Face recognition with weighted kernel principal component analysis,” in 2006 9th International Conference on Control, Automation, Robotics and Vision, 2006, pp. 1–5.
[2] Y. Zhong, J. Chen, and B. Huang, “Toward end-to-end face recognition through alignment learning,” IEEE Signal Processing Letters, vol. 24, no. 8, pp. 1213–1217, 2017.
[3] S. Yuan, X. Yu, and A. Majid, “Robust face tracking using siamese-vgg with pre-training and fine-tuning,” in 2019 4th International Conference on Control and Robotics Engineering (ICCRE), 2019, pp. 170–174.
[4] S. Xia, J. Li, and L. Xia, “Robust face tracking using self-skin color segmentation,” in
[5] 2006 8th international Conference on Signal Processing, vol. 2, 2006.
[6] L.-K.Lee,S.-Y.An, and S.-Y.Oh, “Efficient face detection and tracking with extended camshift and haar-like features,” in 2011 IEEE International Conference on Mechatronics and Automation , 2011, pp. 507–513.
[7] M. Shafiq and Z. Gu, “Deep residual learning for image recognition: A survey,” Applied Sciences, vol. 12, no. 18, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/18/8972
[8] M. Grudin, “On internal representations in face recognition systems,” Pattern Recognition, vol. 33, pp. 1161–1177, 07 2000.
[9] M. Turk and A. Pentland, “Face recognition using eigenfaces,” in Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1991, pp. 586–591.
[10] B. Scholkopf, A. Smola, and K.-R. Muller, “Nonlinear component analysis as a kernel eigenvalue problem,” Neural Computation, vol. 10, pp. 1299–1319, 07 1998.
[11] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, “A convolutional neural network cascade for face detection,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 5325–5334.
[12] S. Yang, P. Luo, C. C. Loy, and X. Tang, “From facial parts responses to face detection: A deep learning approach,” 12 2015, pp. 3676–3684.
[13] S. Zhan, Q.-Q. Tao, and X.-H. Li, “Face detection using representation learning,” Neurocomputing, vol. 187, pp. 19–26, 2016.
[14] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” 042018.
[15] A. Bochkovskiy, C.-Y. Wang, and H.-y. Liao, “Yolov4: Optimal speed and accuracy of object detection,” 04 2020.
[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.
[17] P. Sandhu, I. Kaur, A. Verma, S. Jindal, I. Kaur, and S. Kumari, “Face recognition using eigen face coefficients and principal component analysis,” International Journal on Electrical and Electronics Engineering, vol. 3, 01 2009.
[18] A. Rosebrock, “Opencv eigenfaces for face recognition,” Dec 2022. [Online]. Available: https://pyimagesearch.com/2021/05/10/opencv-eigenfaces-for-face-recognition/
[19] M. A. Firdaus, R. Nadia, and B. Adhi Tama, “Detecting major disease in public hospital using ensemble techniques,” in 2014 International Symposium on Technology Management and Emerging Technologies, 2014, pp. 149–152.
[20] B. Scholkopf, A. Smola, and K.-R. Muller, “Nonlinear component analysis as a kernel eigenvalue problem,” Neural Computation, vol. 10, no. 5, pp. 1299–1319, 1998.
[21] T. Chen, W. Yin, X. Zhou, and D. Comaniciu, “Total variation models for variable lighting face recognition,” IEEE transactions on pattern analysis and machine intelligence, vol. 28, pp. 1519–24, 10 2006.
[22] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, “Deep learning earth observation classification using imagenet

pretrained networks," IEEE Geoscience and Remote Sensing Letters, vol. 13, no. 1, pp. 105–109, 2016.

[23] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning .USA, 2016
[24] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.
[25] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6517–6525. 5
[26] "Yolov3: An incremental improvement," 04 2018. A. Bochkovskiy, C.-Y. Wang, and H.-y. Liao, "Yolov4: Optimal speed and accuracy of object detection," 04 2020.
[27] H. Jiang and E. Learned-Miller, "Face detection with the faster r-cnn," in 2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017), 2017, pp. 650–657.
[28] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask r-cnn," in 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp.2980–2988.
[29] N. Markus, M. Frljak, I. Pandzic, J. Ahlberg, and R. Forchheimer, "A method for object detection based on pixel intensity comparisons," 052013.
[30] A. Ghorbel, N. Ben Amor, and M. Abid, "Gpgpu-based parallel computing of viola and jones eyes detection algorithm to drive an intelligent
[31] Wheelchair," J. Signal Process. Syst., vol. 94, no. 12, p. 1365–1379, dec 2022. [Online]. Available: https://doi.org/10.1007/s11265-022-01783-2
[32] K. Lin, H. Zhao, J. Lv, C. Li, X. Liu, R. Chen, and R. Zhao, "Face detection and segmentation based on improved mask r-cnn," Discrete Dynamics in Nature and Society, vol. 2020, p. 1–11, 2020.
[33] X. P. Burgos-Artizzu, P. Perona, and P. Dollar, "Robust face landmark estimation under occlusion," in 2013 IEEE International Conference on Computer Vision, 2013, pp. 1513–1520.
[34] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 2879–2886. [Online]. Available: https://www.luxand.com/
[35] X. Cao, Y. Wei, F. Wen, and J. Sun, "Face alignment by explicit shape regression," International Journal of Computer Vision, vol. 107, no. 2, p. 177–190, 2013.
[36] X. Yu, J. Huang, S. Zhang, W. Yan, and D. N. Metaxas, "Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model," in 2013 IEEE International Conference on Computer Vision, 2013, pp. 1944–1951.
[37] X. Xiong and F. De la Torre, "Supervised descent method and its applications to face alignment," in 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 532–539.
[38] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial landmark detection by deep multi-task learning," in Computer Vision – ECCV 2014, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 94–108.