# Northumbria Research Link

Citation: Elharrouss, Omar, Almaadeed, Noor, Al-Maadeed, Somaya and Khelifi, Fouad (2022) Pose-invariant face recognition with multitask cascade networks. Neural Computing and Applications, 34 (8). pp. 6039-6052. ISSN 0941-0643

Published by: Springer

URL: https://doi.org/10.1007/s00521-021-06690-4 <https://doi.org/10.1007/s00521-021-06690-4>

This version was downloaded from Northumbria Research Link: https://nrl.northumbria.ac.uk/id/eprint/47589/

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: http://nrl.northumbria.ac.uk/policies.html

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)

Northumbria University NEWCASTLE

UniversityLibrary

# Pose-invariant face recognition with multitask cascade networks

Omar Elharrouss*[a], Noor Almaadeed[a], Somaya Al-Maadeed[a], Fouad Khelifi[b]

*omar.elharouss@qu.edu.qa, n.alali@qu.edu.qa, s_alali@qu.edu.qa, fouad.khelifi@northumbria.ac.uk*

[a] *Department of Computer Science and Engineering, Qatar University, Doha, Qatar, Doha, Qatar*
[b] *Department of Computer Science and Digital Technologies, Northumbria University, Newcastle Upon Tyne, UK, Newcastle, UK*

**Abstract**

In this work, a face recognition method is proposed for face under pose variations using a multi-task convolutional neural network (CNN). Furthermore, a pose estimation method followed by a face identification module are combined in a cascaded structure and used separately. In the presence of various facial poses as well as low illuminations, datasets that include separated face poses can enhance the robustness of face recognition. The proposed method relies on a pose estimation module using a convolutional neural network model and trained on three categories of face image capture such as the Left side, Frontal, and right side. Second, three CNN models are used for face identification according to the estimated pose. The Left-CNN model, Front-CNN model, and Right-CNN model are used to identify the face for the left, frontal, and right pose of the face, respectively. Because face images may contain some useless information (e.g. background content), we propose a skin-based face segmentation method using structure-decomposition and the Color Invariant Descriptor. Experimental evaluation has been conducted using the proposed cascade-based face recognition system that consists of the aforementioned steps (i.e., pose estimation, face segmentation, and face identification) is assessed on four different datasets and its superiority has been shown over related state-of-the-art techniques. Results reveal the contribution of the separate representation, skin segmentation, and pose estimation in the recognition robustness.

*Keywords:* Face recognition, Pose estimation, Pose-invariant, skin segmentation, Convolutional Neural Networks

## 1. Introduction

Person identification across multiple cameras is a crucial task for several applications including monitoring systems [3]. In fact, given the uniqueness of face features and the

*Email addresses:* omar.elharouss@qu.edu.qa (Omar Elharrouss*), n.alali@qu.edu.qa (Noor Almaadeed), s_alali@qu.edu.qa (Somaya Al-Maadeed), fouad.khelifi@northumbria.ac.uk (Fouad Khelifi)

abundance of information a face image can carry, face-based identification emerges as an effective task for video surveillance and biometric applications [4]. However, in some scenarios, a part of the face can not be captured by the camera or can be slightly occluded, thus limiting the performance of any recognition system. Moreover, the environmental conditions as well as the number of people in the surveilled areas represent a challenge for face recognition techniques. In addition, the speed of the recognition process can be important in many cases and real-time processing can be affected by these challenges.

For a long time, face recognition was the subject of many studies [8]. Different approaches have been proposed using statistical and sequential techniques, but the performance is limited when a large dataset is used [5–7, 9]. Some approaches used Deep learning models that improve the performance of recognition as well as reducing the cost of searching for suitable features. But, these methods required large-scale datasets and high computational cost. Also, pose variations represent a real challenge also for the two types of methods (sequential and deep learning-based methods). Because recognizing faces under different poses can affect the performance of the recognition. Some alternatives have been proposed, especially deep-learning-based models, to overcame these problems by proposing some methods for pose-invariant face recognition [10].

In this paper, face recognition under different pose conditions is conducted. A robust and efficient system is presented in which the pose is estimated first, then segmentation of the face skin is followed prior to the recognition step which is performed via a proposed multitask cascade CNN model.The multitask model that includes three parts consists of a Left-CNN model, a Front-CNN model, and a Right-CNN model trained on separated data into Left, Right, and Front sides of the face. Hence, the recognition process consists of classifying the pose of the face before identifying the person using one of the aforementioned models. This approach allows estimating the pose then recognize it using a specific CNN model from the pre-trained models for each pose. The main contributions of this work are listed in the following:

- A robust Multitask cascaded CNN-based system for face recognition under pose variation.
- Unlike the traditional approach that considers all face poses in training, this work proposes a pose-based approach that generates three separate CNN models corresponding to Left, Right, and Front poses. The trained models are fed by a pose estimation method.
- A skin segmentation method is proposed to discard irrelevant information in a face image prior to training.

Here, the face recognition problem is viewed as a trade-off between accuracy and processing time. The system includes three main stages: first the classification of the pose, which represents the categorization of the face poses. The second stage consists of the segmentation of the facial skin in order to discard irrelevant information in the face image. Then in the final stage, face recognition is performed via the corresponding model of the selected pose.

Table 1: Sequential-based methods

| Method | Used features | Problem handled |
|---|---|---|
| Padmanabhan et al. 2019 [8] | Hybrid optimized Kernel (ELM).<br>Hybrid particle swarm optimization-genetic algorithm (PSO-GA) | Pose variations<br>Low light conditions |
| Yi et al. 2013 [10] | PCA, 3D representation, Gabor feature | Pose variations |
| Crosswhite et al. 2018 [11] | SVM | Face categorization and verification |
| Abbad et al. 2018 [12] | Multidimensional ensemble empirical mode decomposition (MEEMD) | Dimensionality reduction |
| Dadi et al. 2017 [13] | GMM, HOG, SVM | Face Recognition |
| Xu et al. 2018 [14] | Sparse Representation | Multi-view recognition |
| Bhowmik et al. 2019 [15] | Log-ICA, ICA | Low resolution |

The analysis integration makes the system coherent in terms of computational complexity and simple to extract information.

The rest of the paper is organized as follows. In Section 2, an overview of related works is presented including sequential-based and CNN-based methods. A detailed description of the proposed face recognition approach is presented in section 3. Performance evaluation and comparison are provided in section 4. Section 5 draws a conclusion and discusses future works.

## 2. Related works

The last decades have shown an increased interest in computer vision tasks including face recognition due to the technical development in video surveillance and monitoring. Also, face recognition is able to recognize uncooperative subjects in a nonintrusive manner compared with other biometrics like fingerprint, iris, and retina recognition. Thus, it is applicable in different sectors, including border control, airports, train stations, and indoor like in companies or offices. Many works have been reported with high performance. Some of these methods have been incorporated with surveillance cameras for person identification exploiting large-scale datasets.

### *2.1. Sequential-based approaches*

Before deep learning techniques, sequential methods that consist of a list of processes using features, several face recognition methods have been proposed. These traditional methods provide face recognition in normal conditions but some of them also deal with pose variations of the faces. Accurate recognition is the main goal in the normal case or in the complex ones such as in the presence of illumination changes or bad lighting conditions [8], low-resolution [9] or variation of face poses [10]. To deal with bad conditions of the face during the detection phase, the authors in [8] proposed an approach based on hybrid optimized Kernel (extreme learning machine (ELM)) and Hybrid particle swarm optimization-genetic algorithm (PSO-GA). In order to classify the face for recognition or verification purposes, the authors in [11] propose a method for face classification and categorization based on CNN features and SVM classification. Reducing the directionality of the face images can reduce the processing time, the authors in [12] proposed a face recognition technique which starts with a preprocessing function using multidimensional ensemble empirical mode decomposition (MEEMD). For pose-invariants-based methods, they use different features in order to extract the suitable feature for the variation of the face. The authors in [10] use 3D reconstruction to estimate the pose of the face, followed by the combination of PCA and Gabor features and Cosine metrics to recognize (remove redundancy also for similarity perspective) the face with pose invariants. For this method, the 3D structure of the face enables an efficient preprocessing task for pose estimation. Also, 3D representation with facial symmetry can handle the problem of self-occlusion. For human tracking, the authors in [13] propose a face recognition technique using GMM, HOG features and SVM classifications. The recognition in the normal cases is good but not robust for pose variations which are essential for tracking. In the same context, the authors in [14] propose a face recognition algorithm that can be used for multimodal sensors. The authors exploit the experiments on Smart glasses (Inertial Measurement Unit (IMU) sensor). The proposed method named Multi-view Sparse Representation Classification (MVSRC) is less expensive in terms of processing time and computational complexity for inertial sensors. By applying the Logarithmic transformation on basic ICA(Logarithmic transformation on basic ICA) named Log-ICA, the authors in [15] enhance the Gaussianity of the detected faces in order to recognize those faces in bad conditions of lighting and different poses. In the Table 1, we summarize the methods as well as the features used, and the problems handled for face recognition.

### *2.2. CNN-based approaches*

Analyzing the face is the main task for several biometric and non-biometric applications including the recognition of facial expression [16], face identification in images under low-resolution [17], and face identification and verification under pose variations [18], represents the most subjects focusing on the analysis of the face. In this paper, we focused on the recognition of the face under pose variations.

Pose variations make the recognition of the face very difficult for different types of methods.Using convolutional neural network (CNN) based techniques, the performance of the existing methods is improved while attempting to handle the pose-invariants problem because of the power of deep learning techniques to learn from the used datasets. In early attempts, the face was recognized by introducing different face poses in a training database [19]. Nowadays, CNNs which are automatic learning techniques make possible the training and learning of different poses like DeepID [20] by using large-scale datasets like the one used for FaceNet [21]. Several types of neural networks are used for the same purpose, for example, Residual Neural Networks are used in [8] for face recognition under pose variations. In the same context, in [23] the authors proposed face recognition methods using a CNN model and trained on a well-known dataset LFW. Also, in [24] the authors use a CNN model for face classification and alignment of the near-face to a frontal view.

The use of a large-scale dataset that contains different poses was considered as one of the solutions for pose-invariant face recognition. For that, the authors in [18] use the CASIA-

WebFace dataset which composed of 500K images for training the proposed Pose-Aware Models (PAMs) method. To deal with the pose invariants problem, some authors reshape 2D face images to 3D representation for estimating the pose then recognize the face using a CNN model [25]. The same problem stimulated researchers to collect a very large-scale dataset [26], that contains more than 2.6 million images. The proposed dataset is used for training a CNN model inspired by [27]. Also, the authors in [28] use Transfer-Learning (TL) for CNN-based face recognition. Zhang et al. [29] proposed the CNN model using compact and discriminative features as well as two proposed loss functions. the evaluate the effectiveness of the proposed methods the authors use three CNN models including ResNet-50, CNN-M, and LeNet. In [30], the authors proposed a face recognition method using face images captured from two cameras, such as RGB camera and thermal cameras. The proposed techniques make the learning of invariant features from many types of data which can make the system more efficient. Combining the facial components and 3D facial characteristics the authors developed a CNN-based model for face recognition. The proposed method allows a large-scale dataset of poses for training.. In [32], the authors proposed a pose estimation method before straining the recognition process. In the same context, researchers in the last few years focused on the processing of pose invariance for face recognition [44–55]. In [44], the authors proposed a method for face recognition and clustering performed on the face image with different illumination and pose. Also, the authors in [45] proposed a face recognition method using transfer learning. In works reported in [44–46], the processing uses just one dataset, where other datasets can contain categories of images that should be tested like in [46–55].

The data structure can be considered to improve the learning, which not the case for almost all proposed methods. Deep face recognition methods use different loss functions to improve the classification resulsts [58–73]. Sofmax loss is one of the effective models for CNN-based face recognition[58]. Other methods, combine different features Softmax to enhance the recognition, including Sofmax+ contrastive Softmax Loss+Contrastive [66], L-Softmax Loss [62], Softmax+Center Loss [46] , Center Loss [61]. Where other approaches s use some loss function methods like Triplet Loss [43], Range loss [67], CosFace(LMCL) [66], FairLoss [73]. Our method provides a new data structure on which the proposed method can be performed with any loss function while maintaining high accuracy.

## 3. Proposed approach

Without individual cooperation, face recognition aims to identify a person using face images. It represents a crucial challenge in computer vision. Unlike the traditional method that recognizes faces represented with the frontal side, using the other side of the face including the left side and the right side makes the recognition very difficult. In order to overcome this challenge, pose-invariant face recognition is proposed in this work. Prior to the recognition stage, the face pose is estimated first. The recognition is conducted through a proposed multitask cascaded deep learning model. As illustrated in Figure 1, the proposed approach estimates the pose of the face using the LFR dataset [56] that contains the Left, Front, and Right side of each face as illustrated in Figure 2. Then, in order to recognize the

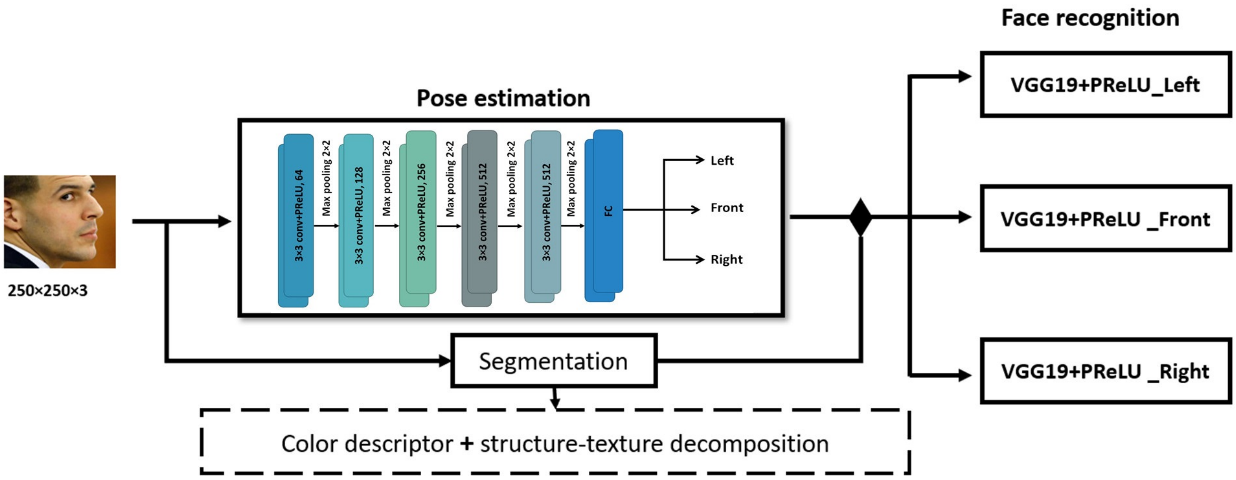


Figure 1: Multitask pose-invariant face recognition.

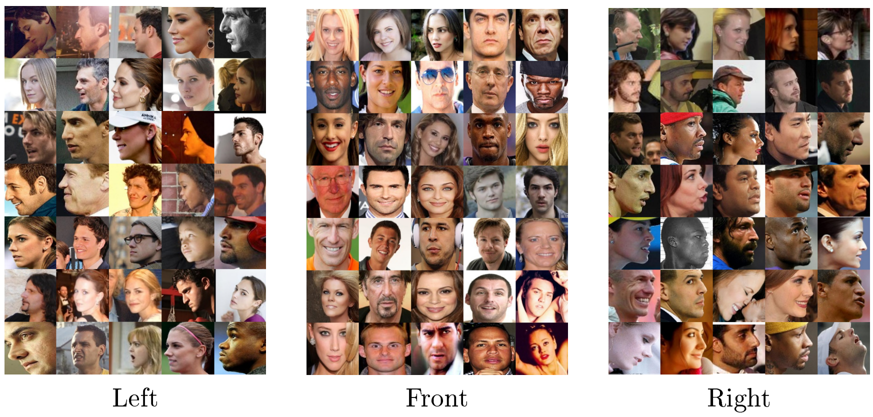


Figure 2: Data section from LFW, CFP, and CASIA-WebFace datasets.

face, a segmentation process of the face is performed in order to extract just the face skin from the image. The recognition process consists of building three CNN models; namely: Left-CNN, Front-CNN, and Right-CNN. One of the aforementioned models is finally used to recognize the face according to its classified pose. The following subsection describes the proposed approach in detail.

### *3.1. Pose estimation*

It is hard for a system to recognize a face under diverse pose conditions. Many datasets have been proposed mentioning the degree of pose variation captured from different angles using multiple cameras. Using these datasets many approaches have been proposed to recognize the faces and attempted to handle the pose variation problem. From these works, we can find a method named pose-directed CNN that was proposed in [14]. For the same purpose, and in order to handle pose variation problems, we propose a pose estimation CNN model trained on the LFR [53] dataset as illustrated in Figure 3.Once pose estimation is

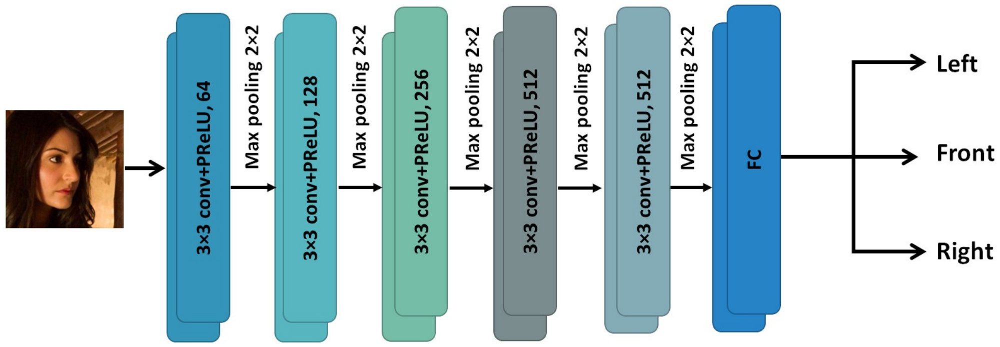


Figure 3: Proposed CNN model for face pose estimation.

performed, the face gets recognized using the corresponding one of the three trained CNN models where each model represents a face pose.

### *3.2. Face skin segmentation*

The face image may contain undesirable information in some regions of the image. These regions represent irrelevant information that cannot be used in a face recognition task. They can also affect the recognition accuracy if they are complex with different color intensities. For face recognition, only the region that covers the skin of the face is used. The skin is a homogenous part of the face that can be extracted to reduce the quantity of useless information in the face. In order to segment the region of the face skin and the part that will be analyzed, a face skin segmentation method is presented. This is based on the structure extraction of the homogenous part of the face image and color invariant descriptors. The proposed method is detailed in the following section.

#### *3.2.1. Structure decomposition*

First, the image smoothed prior to skin area segmentation. To this end, the structure-texture decomposition method proposed in [37] is adopted using the interval gradient, for adaptive gradient smoothing.This is a technique that splits the image into Structure and texture components. The structure of an image represents the bounded variation component while the texture represents the oscillating parts of the image. This strategy is applied to face images, and in order to segment the skin, the structure component which represents the homogeneous parts is exploited. The texture can contain much additional information that we do not need in skin segmentation, so it is ignored [38]. The extraction of the structure component uses gradient rescaling with interval gradients followed by a color handling operation.

In order to extract or remove the texture part from an input signal I, suppression of the texture region gradients must be performed. Moreover, For each pixel $p$ in all local windows $\Omega_p$, the signal should be either increasing or decreasing. To rescale the gradients of the input signal with the corresponding interval gradients, the following equation is used:

Original image Structure component Color descriptor Segmented image

Figure 4: Skin segmentation results.

$$(\nabla' I)_{\mathrm{p}} = \begin{cases} (\nabla I)_{\mathrm{p}}.\mathrm{w}_{\mathrm{p}} & if \;\; sign((\nabla I)_{\mathrm{p}}) = sign((\nabla_{\Omega} I)_{\mathrm{p}}) \\ 0 & otherwise \end{cases} \tag{1}$$

Where$(\nabla' I)_{\mathrm{p}}$ represents the rescaled gradient, and $w_{\mathrm{p}}$ is the rescaling weight:

$$w_{\mathrm{p}} = min\left(1, \frac{|(\nabla_{\Omega} I)_{\mathrm{p}}| + \varepsilon_{\mathrm{s}}}{|(\nabla I)_{\mathrm{p}}| + \varepsilon_{\mathrm{s}}}\right) \tag{2}$$

Where $\varepsilon_{\mathrm{s}}$ is instability prevention constant. The algorithm becomes sensitive to noise where The values of $\varepsilon_{\mathrm{s}}$ is too small, introducing unwanted artifacts to filtering results. By increasing $s$, the sensitivity to noise can be reduced, but textures may not be completely filtered if $\varepsilon_{\mathrm{s}}$ is too big. We set $\varepsilon_{\mathrm{s}} = 10^{-4}$ in our implementation.

For filtering color images, the gradient sums of color channels in the gradient rescaling step (Equations (1) and (2)) is used, that is:

$$w_{\mathrm{p}} = min\left(1, \frac{\sum\limits_{\mathrm{c}\in\{\mathrm{r,g,b}\}} |(\nabla_{\Omega} I^{\mathrm{c}})_{\mathrm{p}}| + \varepsilon_{\mathrm{s}}}{\sum\limits_{\mathrm{c}\in\{\mathrm{r,g,b}\}} |(\nabla I^{\mathrm{c}})_{\mathrm{p}}| + \varepsilon_{\mathrm{s}}}\right) \tag{3}$$

Figure 4 shows filtering examples to demonstrate the results of structure extraction using the same parameters $\varepsilon \in [0.01^2, 0.03^2]$.

*3.2.2. Color invariant descriptors*

The discrimination of objects in an image is performed using colors. For that, the color spatial configuration and spectral energy distribution should be used to describe colors [39]. The Kubelka–Munk theory as defined in Eq. (4) is established based on a physical model in terms of the spectral and spatial dimension of the color [12]. Parameters with properties independent of illumination intensity and viewpoint are defined as the color invariants.

$$E(\lambda, \overrightarrow{x}) = e(\lambda, \overrightarrow{x})(1 - \rho_{\mathrm{f}}(\overrightarrow{x}))^2 R_{\infty}(\lambda, \overrightarrow{x}) + e(\lambda, \overrightarrow{x})\rho_{\mathrm{f}}(\overrightarrow{x}) \tag{4}$$

where $\overrightarrow{x}$ denotes the position in the image plane, $\lambda$ denotes the wavelength, $e(\lambda, \overrightarrow{x})$ denotes the illumination spectrum, $\rho_{\mathrm{f}}(\overrightarrow{x})$ denotes the Fresnel reflectance at $\overrightarrow{x}$, and $R_{\infty}(\lambda, \overrightarrow{x})$

denotes the material reflectivity. The color invariants are defined in an ideal physical model, i.e. the color invariants are defined in the given dimensional spectrum at an infinitesimal small spatial neighborhood. We use here the formula that defined Equal energy, but uneven illumination as follows:

$$E(\lambda, \overrightarrow{x}) = i(x)\{\rho_{\mathrm{f}}(x) + (1-\rho_{\mathrm{f}}(x))^2 R_{\infty}(\lambda, x)\} \tag{5}$$

In order to get color invariants from RGB images, the spectral and spatial parameter of the Gaussian color model of $\lambda_0$ incident light at scale $\sigma_{\mathrm{x}}, \sigma_{\mathrm{y}}, \sigma_{\lambda}$ were derived. Colorimeter analysis is used to find out when $\lambda_0$= 520 mm and $\sigma_{\lambda}$=55 mm, the spectral structure of Gaussian color model is in perfect agreement with the human visual system of color perception [11]. Eq. (6) shows the linear transformation from RGB image value to spectral parameters of Gaussian color model:

$$\begin{bmatrix} \widehat{E} \\ \widehat{E_{\lambda}} \\ \widehat{E_{\lambda\lambda}} \end{bmatrix} = \begin{pmatrix} 0.06 & 0.63 & 0.27 \\ 0.3 & 0.04 & -0.35 \\ 0.34 & -0.6 & 0.17 \end{pmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{6}$$

The binary image elaboration is performed using the threshold of Eq. (4) which compute the maks for each pixel. When the value of $\widehat{E_{\lambda}}$ is significant the pixel is labeled by 1, and the threshold gets the value 0 where the value of D converge to 0.

$$D(x, y) = 1 - e^{-\widehat{E_{\lambda}}(x,y)} \tag{7}$$

Using the proposed technique for face segmentation, Figure 5 represents results performed on selected face images from CASIA-WebFace dataset showing the original image and the corresponding images after face segmentation. For each identity we illustrate three state of the face including left side frontal and right side. As seen, the results show high accuracy in segmenting face areas.

### *3.3. Face recognition*

The selection of an optimal CNN architecture is a challenging problem is normally dependent on the application. The proposed deep learning approach involves preprocessing and feeding it to a convolution neural network. The pre-processing stage consists of the extraction of skin face, followed by resizing the size of data prior to training. A multitask Convolution Neural Network (CNN) based on VGG19 [57] architecture, which is a supervised learning system with a multistage deep learning network, has been implemented. Multitask CNN could learn multiple stages of invariant features from the input images. For the pose, invariant faces are classified into three categories including Left, Front, and Right face images. Convolution and pooling Layer are the main layers in a CNN model. Any complex CNN can be constructed with a couple of combination convolution-pooling. Hence, by using images as inputs and back-propagating the errors, the learning process takes place.

The VGG19+PReLU architecture illustrated in figure 6 comprises 19 convolutional layers and 5 MaxPooling layers, and two fully connected layers. We introduce the PReLU layer after each convolution layer.

Figure 5: Face segmentation results.

Parametric Rectified Linear Unit (PReLU) has been employed as the activation function, which is a generalized parametric formulation of ReLU. This activation function adaptively learns the parameters of rectifiers and improves the accuracy at a negligible extra computational cost [40]. Only positive values are fed to the ReLU activation function, while all negative values are set to zero. PReLU assumes that a penalty should be applied for negative values, and it should be parametric. The PReLU function can be defined as:

$$f(y_{\mathrm{i}}) = \begin{cases} y_i & if\ y_{\mathrm{i}} > 0 \\ a_{\mathrm{i}} y_{\mathrm{i}} & if\ y_{\mathrm{i}} \leq 0 \end{cases} \tag{7}$$

where the $\mathrm{a_i}$ controls the slope of the negative part. When $\mathrm{a}_i = 0$, it operates as a ReLU; when $\mathrm{a}_i$ is a learnable parameter, it is referred to as a Parametric ReLU (PReLU). If $\mathrm{a}_i$ is a small fixed value, PReLU becomes LReLU ($\mathrm{a}_i = 0.01$). As shown in [41], PReLU can be

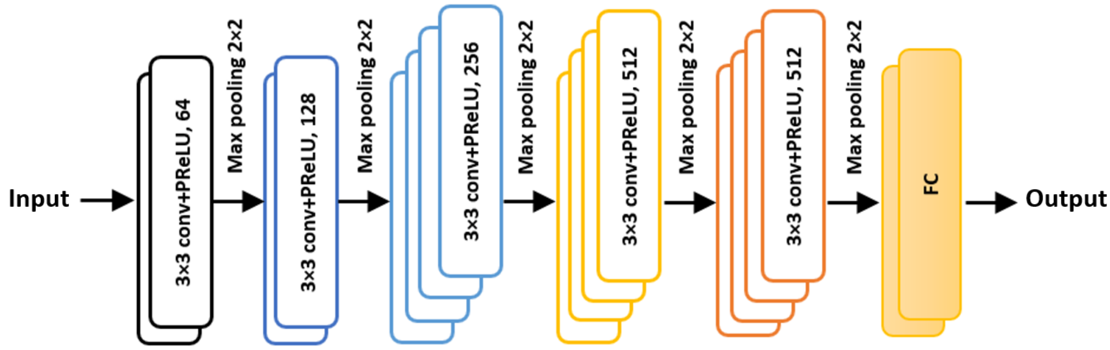


Figure 6: VGG19 + PReLU model adopted for face recognition.

Table 2: Training hyper parameters (General classifier)

| Optimizer | LR | Epsilon | Beta_1 | Beta_2 | Decay | Epochs | Batch size |
|---|---|---|---|---|---|---|---|
| Adam | 0.001 | 1e-08 | 0.91 | 0.999 | 0 | 50 | 32 |

trained using the backpropagation concept.

The input of the system is an image of a segmented face with a resolution of 60´60 pixels. For the training and testing, we use the preprocessed data from CFP, YTF, LFW, and CASIA-WebFace datasets. The model is trained using *CrossEntrpy* loss function with a batch size of 32 examples, and a learning rate of 0.001 as described in Table 2.

## 4. Experimental results

Experiments have been conducted to evaluate the proposed system in comparison with related state-of-the-art techniques. A number of well-known datasets have been used including LFW [33], YTF [34], CFP [35], CASIA-WEBFACE [36]. For each dataset, we compared the propsoed with some existing methods. For example, the experiemnts on LFW and YTF dataset are compared with Softmax Loss [58], Softmax Loss+Contrastive [19], FaceNet [43], Littwin et al. [44], MultiBatch [45], Softmax+Center Loss [46], Augmentation [59], LMLE [60], Center Loss [61], L-Softmax Loss [62], Center invariant loss [63], SphereFace [64], SphereFace(A-Softmax) [65], Wang et al. [47], MLT-CNN [31], CosFace(LMCL) [66], Range loss [67], Feature transfer [68], NRA+CD [69], CLMLE [70], RegularFace [71], ArcFace [72], and FairLoss [73]. By the follwoing, the propsoeed method evaluation and comparisions will be presented with details.

### *4.1. Experimental datasets*

The datasets used for training the proposed approach are summarized in Table 3 in which the characteristics of each dataset are listed.

Table 3: Fcaracteriscitcs of LFW and YTF, and WebFace dataset.

| Dataset | Size | identities | Angles | Resolution |
|---|---|---|---|---|
| LFW | 13233 | 5749 | diverse | 250x250 |
| YTF | 3425 videos | 1595 | diverse | 250x250 |
| CFPW (2015) | 7000 | 500 | Front, Profile | 250x250 |
| CASIA-WebFace | 494 414 | 10 575 | diverse | 250x250 |
| LFR [56] | 500 K | 11 000 | Left, Front, Right | 250x250 |

Table 4: Performance of pose estimation method.

| Method | Accuracy (%) |
|---|---|
| Patacchiola et al. (AlexNet 2017) [51] | 75 (FLW) |
| Sreekanth et al. (2018) [52] | 97.6 (FLW) |
| Zavan et al. (2019) [53] | 93.34 (FLW) |
| **Ours** | **99.36** (FLW, WebFace, CFP) |

*4.2. Pose estimation evaluation*

Pose estimation is an important step in face recognition analysis that can improve the performance of the recognition system. Several methods have been proposed as well as many datasets are collected such as the pointing'04 dataset [34] and Schneiderman dataset [35]. The two datasets contain a set of images that represents different face poses. While pointing'04 datasets [34] consist of 182 images of 15 identities. each image represents the face captured from different angle in the range of [-90°, +90°]. Where Schneiderman dataset contains 6660 images of 90 subjects. Each subject has 74 images that represent all the orientations of the head from the right (+90°) to the left (-90°).

In order to structure these datasets in accordance with the proposed approach, the datasets are split into three subsets : right, front and left. Regarding the yaw angle in the image the right pose is represented by the angles (-90°, -75°, -60°, -45°), the frontal is rerpesented by the images of angles (-30°, -15°, 0°, 15°, 30°), and the left profile is represented by the images of thee agnles: (45°, 60°, 75° , 90°). LFR dataset [56] used the same procedure fro collecting and grouping the images fron CASIA-WebFace dataset.

Once the proposed pose estimation model is trained on the LFR dataset [56], it has been tested and results obtained are presented in Table 4. As shown, the proposed method is efficient for pose estimation, while the accuracy reaches 99% for the Schneiderman dataset and 97% for pointing'04 dataset. The performance of the proposed pose estimation model is demonstrated also using Figure 7 that illustrates some examples and their respective probabilities. Irrespective of the orientation, the model overcomes the illumination and the image resolution problem.

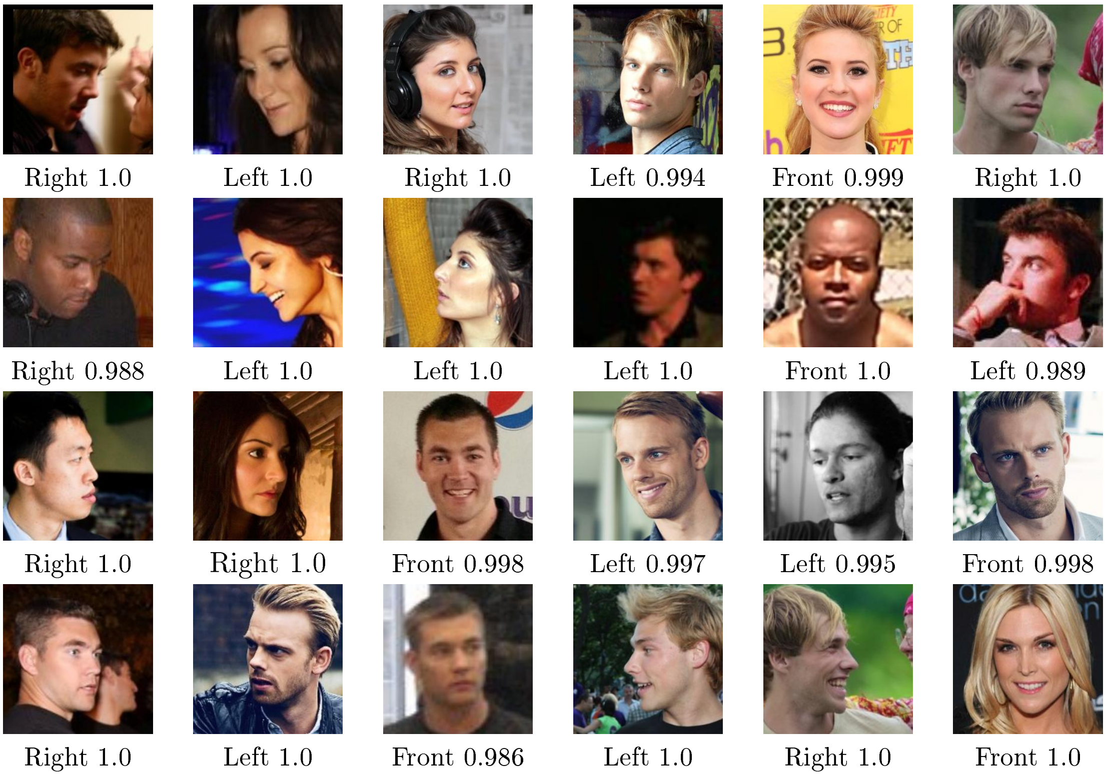


Figure 7: Obtained result of some tested face images for pose estimation with the estimated pose and the probabilities of each one

### *4.3. Face recognition evaluation*

The training set used in this paper is composed of CASIA-WebFace face images. For evaluating the proposed face recognition system, we use datasets including Poiting04 and Schneiderman, LFW, YTF, CFP, and IJB-A. The same architecture and parameters are used for all datasets. All models are trained from scratch for 500 epochs with a batch size of 128. The other parameter settings and training processes are the same as declared in section 3. Figure 8 shows the performance accuracy of the proposed pose estimation method for both Poiting04 and Schneiderman datasets. From the figure, we can observe that the frontal pose estimation is the highest accuracy due to the clarity of the frontal face images where the left and right side of the face can contain more expressions and information as well as the appearance differences from a face to another.

Similar to the comparative analysis made on the three datasets including LFW, YTF, CFP, and IJB-A, we have included experimental results of related state-of-the-art techniques on the same datasets for the sake of illustration. Tables 5, 6, and 7 depict the performance of the proposed system against its competitors.

#### *4.3.1. Evaluation on LFW and YTF dataset*

The evaluation of the proposed method is conducted on the widely used YTF and LFW datasets. While LFW contains about 5 000 images and YTF contains about 6000 face

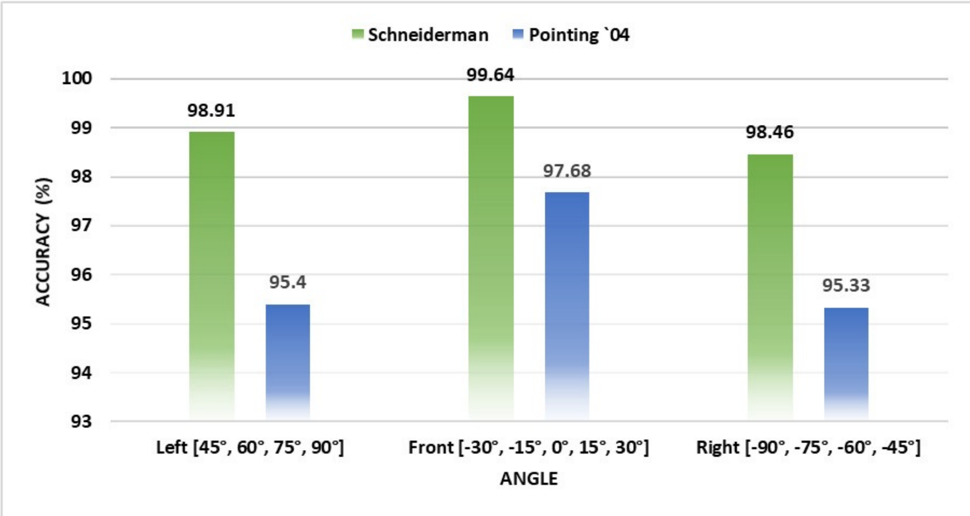


Figure 8: Pointing'04 and Schneiderman datasets performance resulsts

images. The two versions of our approach are compared with 23 state-of-the-art methods of the two datasets using the entire WebFace data in the training phase. Table 5 represents some state-of-the-art methods results. The list the state-of-the-art methods includes Softmax Loss [58], Softmax Loss+Contrastive [19], FaceNet [43], Littwin et al. [44], MultiBatch [45], Softmax+Center Loss [46], LMLE [60], Center Loss [61], SphereFace [64], SphereFace(A-Softmax) [65], Wang et al. [47], MLT-CNN [31], CosFace(LMCL) [66], Range loss [67], Feature transfer [68], NRA+CD [69], CLMLE [70], RegularFace [71], ArcFace [72], and FairLoss [73]. From the table, it can be seen that the techniques that use of the entire WebFace dataset seems to produce a higher error rate and hence underperform when compared with the techniques that use a portion of the dataset. For example, the ArcFace [72] method used 5.8M from the WebFace dataset and reached 99.83% as performance on LFW and 98.02% on the YTF dataset. the performance of ARcFace is more improved than the proposed Multitask-CNN by 0.4% on LFW and 2.3% on YTF. Note that the proposed Multitask-CNN model improves when used with the proposed segmentation method by 0.2% on LFW and 1.4% on YTF, which make the obtained results using the proposed method comparable with those obtained with ArcFAce although the data size used is different. Also, it can be seen that the proposed system outperforms CosFace(LMCL) [66], NRA+CD [69] by 0.2% and 0.3% respectively on the YTF dataset except for the ArcFace method which is better than the proposed method result on this dataset.

*4.3.2. Evaluation on CFP dataset*

Table 6 represents the accuracy of the proposed multitask system compared with related state-of-the-art techniques applied on CFP dataset. The representation of CFP dataset which is divided into Frontal and Profile sets minimizes the error rates coming from the variation of the poses in the dataset. In general, the recognition rate on the frontal side is more accurate than the profile side due to the information contained in it which is not

Table 5: Face verification accuracy (%) on LFW and YTF datasets trained on WebFace dataset.

| Method | Data | LFW | YTF |
|---|---|---|---|
| Softmax Loss [58] | WebFace | 97.88 | 93.1 |
| Softmax Loss+Contrastive [19] (2014) | WebFace | 98.78 | 93.5 |
| FaceNet [43] (2015) | WebFace | **99.63** | - |
| Littwin et al. [44] (2016) | WebFace | 98.14 | - |
| MultiBatch [45] (2016) | WebFace | 98.20 | - |
| Softmax+Center Loss [46] (2016) | WebFace | **99.28** | - |
| Augmentation [59] (2016) | WebFace | 98.06 | - |
| LMLE [60] (2016) | WebFace | 99.51 | 95.8 |
| Center Loss [61] (2016) | WebFace | 98.91 | 93.4 |
| L-Softmax Loss [62] (2016) | WebFace | 99.01 | 93.0 |
| Center invariant loss [63] (2017) | WebFace | 99.12 | 93.88 |
| SphereFace [64] (2017) | WebFace | 99.26 | 94.1 |
| SphereFace(A-Softmax) [65] (2017) | WebFace | 99.42 | 95.0 |
| Wang et al. [47] (2017) | WebFace | 96.2 | - |
| MLT-CNN [31] (2018) | WebFace | 98.27 | - |
| CosFace(LMCL) [66] (2018) | WebFace | 99.33 | 96.1 |
| Range loss [67] (2018) | 1.5M | 99.52 | 93.7 |
| Feature transfer [68] (2018) | 4.8M | 99.37 | - |
| NRA+CD [69] (2019) | 0.55M | 99.53 | 96.04 |
| CLMLE [70] (2019) | WebFace | 99.62 | 96.5 |
| RegularFace [71] (2019) | WebFace | 99.33 | 94.4 |
| ArcFace [72] (2019) | 5.8M | **99.83** | **98.02** |
| FairLoss [73] (2019) | WebFace | 99.57 | 96.2 |
| Our Multitask-CNN | WebFace | **99.46** | 95.72 |
| Our Multitask-CNN+ segmentation | WebFace | **99.61** | **96.38** |

the case in a profile face that represents just a part of the face. This can be observed in the table. For example, we can see that the state-of-the-art methods such as [48], DR-GAN [49], Pen et al. [50], and MLT-CNN [31] trained on frontal face set are better by an average of 3% to 7% than the same method trained on profile set. Also, the performance of the methods trained on the merged dataset like CosFace [66], ArcFace [72], and NRA+CD [69] is lower than the methods trained on each set by a difference of 2% to 3%. It is also worth mentioning that the proposed system trained on the new representation of Left, Right, and Front face sets delivers better performance in terms of the front set than the techniques that used merged frontal and profile datasets CosFace [66], ArcFace [72], and NRA+CD [69], also compared to some method trained on frontal and profile set like DR-GAN [49], and MLT-CNN [31] with a difference of 2 to 3% for the first and 1% for the others except Pen et al. [50] when we used Multitask-CNN+segmentation. Regarding the profile set, the two

Table 6: Evaluation comparison on CFP dataset.

| Method | Frontal (F) | Profile (P) | |
|---|---|---|---|
| Sankaranarayanan et al (2016) [48] | 96.93 | 89.17 | |
| DR-GAN [49] (2017) | 97.84 | 93.41 | |
| Pen et al.[50] (2017) | **98.67** | 93.76 | |
| MLT-CNN [31] (2018) | 97.79 | 94.39 | |
| CosFace(LMCL) [66] (2018) | 95.52 (FP) | | |
| ArcFace(LMCL) [72] (2019) | 95.44 (FP) | | |
| NRA+CD [69] (2019) | 95.50 (FP) | | |
| | **Front** | **Left** | **Right** |
| Our Multitask-CNN | 97.73 | 96.04 | 94.53 |
| our Multitask-CNN+ segmentation | 98.3 | **96.73** | **95.67** |

versions of the proposed approach is more accurate than the other method by 3% to 4%. The proposed approach reached 96.4% and 69.73% on the Left set using Multitask-CNN and Multitask-CNN+segmentation, respectively. While 95.67% is the achieved rate using Multitask-CNN+segmentation on the Right set.The two obtained results are better than the state-of-the-art methods on the same profile set using the new representation of data and feeding one of the three different CNN models according to the estimated pose.

#### *4.3.3. Evaluation on IJB-A dataset*

The proposed system is also assessed on IJB-A dataset and compared with relevant stat-of-the-art techniques. The face recognition model was retained by adding some identities from JIB-A datasets. There is a small similarity between CASIA-WebFace and IJB-A dataset. Therefore, the folder from IJB-A that is similar to CASIA-WebFace is removed. This contains about 26 subjects. The proposed model is then retained. The evaluation on IJB-B has been made using the proposed system and compared with some state-of-the-art methods. Table 7 represents the obtained results for verification and identification operations. Results with FAR=0.01 and FAR=0.001 are used for verification, while identification results are considered using Rank-1 and Rank-5 scores, accordingly.

By using the new presentation of IJB-A dataset into Left, Right, and Front sets, we trained using VGG19+PrLEU with and without segmentation operations. The obtained results are compared with a number of the recent state-of-the-art models including Masi et al [17], MLT-CNN [31], Wang et al. [47], DG-GAN [49], and Xie et al [55]. The same evaluation used for these methods for evaluation and identification is used for testing the accuracy of the proposed method. The obtained results for verification operations of the proposed method using the two approaches Multitask-CNN and Multitask-CNN+segmentation are the second-best results by a difference of 1.8 %(Multitask-CNN) and 0.9% (Multitask-CNN+segmentation) compared with the results obtained by Xie et al. [55] that achieved the best results for verification. The performance rates obtained with the proposed method and [55] outperform other competing techniques a difference of 18.4% for FAR=0.01 and

41.3% for FAR=0.001.

As for face identification, it can be seen from the presented results in Table 7 that all reported techniques managed to reach good performance unlike the results reported for face verification. The proposed Multitask-CNN system with and without segmentation clearly have the upper hand over other techniques for Rank-1 and Rank-5 identification. Overall, the proposed system delivers the highest performance with a simpler architecture when compared with state-of-the-art techniques. This is because of the separation of the face pose dataset into Left, Right and Front set, and the design of efficient CNN models of the three different poses unlike the traditional approach that merges all face poses together in a single system.

## 5. Conclusions

In this paper, a face recognition multitasks cascade networks is proposed. In uncontrolled environments, the variation of face pose can reduce the performance of such a system also make the recognition very difficult. In order to overcome this challenge, and using a CNN model a pre-estimation of face pose is performed before starting the recognition. For pose estimation, three poses are classified by the proposed model including Left, Front, and Right. We used for training the LFR dataset which is collected from CASIA-WebFace. Once the pose is estimated, the face is then recognized using one of three CNN models that represents the corresponding model for each category, namely: Left-CNN, Front-CNN, and Right-CNN. Indeed, the CNN-based face recognition model is selected according to the output of the pose estimation CNN model. However, because face images may contain much meaningless or noisy information in the background, for example, an efficient face skin segmentation method is proposed prior to face recognition. The proposed multitask CNN learning technique recognizes the face with the pose as the only side information regardless of the other features of the face such as expression, illumination, or image resolution. Intensive experiments have been conducted with the proposed system on different public datasets, namely: LFW, CFP, and IJB-A. The proposed multitask approach is compared with recent works to demonstrate the performance of each method as well as the efficiency.

## 6. Funding

This publication was made by NPRP grant # NPRP8-140-2-065 from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.

## 7. Conflict of interest

I certify that there is no actual or potential conflict of interest in relation to this article.

## 8. Competing interests

The authors declare that they have no competing interests.

Table 7: Evaluation comparison on IJB-A dataset

| Method | Verification | | Identification | |
|---|---|---|---|---|
| **Metrics** | **@FAR=0.01** | **@FAR=0.001** | **@Rank-1** | **@Rank-5** |
| Masi et al. [17] | 73.3±1.8 | 55.2±3.2 | 77.1±1.6 | 88.7±0.9 |
| DG-GAN [49] | 77.4±2.7 | 53.9±4.3 | 85.5±1.5 | 94.7±1.1 |
| Wang et al. [47] | 72.9±3.5 | 51.0±6.1 | 82.2±2.3 | 93.1±1.4 |
| MLT-CNN [31] | 77.5±2.5 | 53.9 ±4.2 | 85.8±1.4 | 93.8±0.9 |
| Xie et al. [55] | **96.2±0.01** | 92.0±0.005 | - | - |
| Our Multitask-CNN | 94.32±1.1 | 91.2±1.8 | 96.1±1.4 | 97.74±1.2 |
| Our Multitask-CNN+ segmentation | 95.1±1.6 | **92.5±1.1** | **96.7±0.9** | **97.3±1.2** |

## 9. Availability of data and materials

## 10. Financial interests

I certify that there is no financial interest in relation to this article.

## 11. List of abbreviations

[1] CNN: Convolutional Neural Networks

## 12. Acknowledgments

This publication was made by NPRP grant # NPRP8-140-2-065 from the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.

## References

[1] Becerra Arévalo, N. P., Calles Carrasco, M. F., Toasa Espinoza, J. L., & Velasteguí Córdova, M. (2020). Neutrosophic AHP for the prioritization of requirements for a computerized facial recognition system. Neutrosophic Sets and Systems, 34(1), 21.

[2] Shanmuganathan, M., & Nalini, T. (2021). AnInvestigation into Facial Attributes Prominence in Face Recognition: Using Hybrid Method AHP and TOPSIS Approach. International Journal of Grid and Distributed Computing, 14(1), 769-799.

[3] Elharrouss, O., Almaadeed, N. and Al-Maadeed, S., 2021. A Review of Video Surveillance Systems. Journal of Visual Communication and Image Representation, p.103116.

[4] Akbari, Y., Almaadeed, N., Al-maadeed, S., & Elharrouss, O. (2021). Applications, databases and open computer vision research from drone videos and images: a survey. Artificial Intelligence Review, 1-52.

[5] Yu, X., & Xu, F. (2020). Random inverse packet information and its acquisition. Applied Mathematics and Nonlinear Sciences, 5(2), 357-366.

[6] Xie, T., Liu, R.,& Wei, Z. (2020). Improvement of the fast clustering algorithm improved by k-means in the big data. Applied Mathematics and Nonlinear Sciences, 5(1), 1-10.

[7] Li, T.,& Yang, W. (2020). Solution to chance constrained programming problem in swap trailer transport organisation based on improved simulated annealing algorithm. Applied Mathematics and Nonlinear Sciences, 5(1), 47-54.
[8] Wang, Z., et al., Low-resolution face recognition: a review. The Visual Computer, 2014. 30(4): p. 359-386.
[9] Yi, D., Z. Lei, and S.Z. Li. Towards pose robust face recognition. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2013.
[10] Crosswhite, N., et al., Template adaptation for face verification and identification. Image and Vision Computing, 2018. 79: p. 35-48.
[11] Abbad, A., et al., Application of MEEMD in post-processing of dimensionality reduction methods for face recognition. IET Biometrics, 2018. 8(1): p. 59-68.
[12] Dadi, H.S., G.K.M. Pillutla, and M.L. Makkena, Face recognition and human tracking using GMM, HOG and SVM in surveillance videos. Annals of Data Science, 2017: p. 1-23.
[13] Xu, W., et al., Sensor-assisted multi-view face recognition system on smart glass. IEEE Transactions on Mobile Computing, 2018. 17(1): p. 197-210.
[14] Bhowmik, M.K., et al., Enhancement of robustness of face recognition system through reduced gaussianity in Log-ICA. Expert Systems with Applications, 2019. 116: p. 96-107.
[15] Li, Y., et al., Occlusion aware facial expression recognition using cnn with attention mechanism. IEEE Transactions on Image Processing, 2019. 28(5): p. 2439-2450.
[16] Ge, S., et al., Low-resolution face recognition in the wild via selective knowledge distillation. IEEE Transactions on Image Processing, 2019. 28(4): p. 2051-2062.
[17] Masi, I., et al. Pose-aware face recognition in the wild. in Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2016.
[18] Pentland, A., B. Moghaddam, and T. Starner, View-based and modular eigenspaces for face recognition. 1994.
[19] Sun, Y., X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
[20] Schroff, F., D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
[21] Taigman, Y., et al. Deepface: Closing the gap to human-level performance in face verification. in Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2014.
[22] Agrawal, Amrit Kumar, and Yogendra Narain Singh. "Unconstrained face recognition using deep convolution neural network." International Journal of Information and Computer Security 12.2-3 (2020): 332-348.
[23] Bashbaghi, Saman, et al. "Deep learning architectures for face recognition in video surveillance." Deep Learning in Object Detection and Recognition. Springer, Singapore, 2019. 133-154.
[24] Jourabloo, A., et al. Pose-invariant face alignment with a single cnn. in Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017.
[25] Parkhi, O.M., A. Vedaldi, and A. Zisserman. Deep face recognition. in bmvc. 2015.
[26] Simonyan, K. and A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[27] Banerjee, S. and S. Das, Mutual variation of information on transfer-CNN for face recognition with degraded probe samples. Neurocomputing, 2018. 310: p. 299-315.
[28] Zhang, M.M., K. Shang, and H. Wu, Deep Compact Discriminative representation for unconstrained face recognition. Signal Processing: Image Communication, 2019.
[29] He, R., et al., Wasserstein cnn: Learning invariant features for nir-vis face recognition. IEEE transactions on pattern analysis and machine intelligence, 2018.
[30] Hsu, G.-S., et al., Fast Landmark Localization With 3D Component Reconstruction and CNN for Cross-Pose Recognition. IEEE Transactions on Circuits and Systems for Video Technology, 2018. 28(11): p. 3194-3207.
[31] Yin, X. and X. Liu, Multi-task convolutional neural network for pose-invariant face recognition. IEEE

Transactions on Image Processing, 2018. 27(2): p. 964-975.
[32] Huang, G.B. and E. Learned-Miller, Labeled faces in the wild: Updates and new reporting procedures. Dept. Comput. Sci., Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep, 2014: p. 14-003.
[33] Sengupta, S., et al. Frontal to profile face verification in the wild. in 2016 IEEE Winter Conference on Applications of Computer Vision (WACV). 2016. IEEE.
[34] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR) ,2011.
[35] I. Masi, S. Rawls, G. Medioni, and P. Natarajan, "Pose-aware face recognition in the wild," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4838–4846.
[36] Lee, H., et al. Structure-Texture Decomposition of Images with Interval Gradient. in Computer graphics forum. 2017. Wiley Online Library.
[37] Elharrouss, O., Moujahid, D. and Tairi, H., 2015. Motion detection based on the combining of the background subtraction and the structure–texture decomposition. Optik-International Journal for Light and Electron Optics, 126(24), pp.5992-5997.
[38] Jan-Mark Geusebroek, R. van den Boomgaard, A.W.M. Smeulders, A. Dev, Color and scale: the spatial structure of color images, in: Proceeding of the 6th European Conference on Computer Vision (ECCV), vol. 1, 2000, pp. 331–341.
[39] Zhou, H., Chen, Y. and Feng, R., 2013. A novel background subtraction method based on color invariants. Computer Vision and Image Understanding, 117(11), pp.1589-1597.
[40] WANG, Liangliang, GE, Lianzheng, LI, Ruifeng, et al. Three-stream CNNs for action recognition. Pattern Recognition Letters, 2017, vol. 92, p. 33-40.
[41] Gourier, Nicolas, Daniela Hall, and James L. Crowley. "Estimating face orientation from robust detection of salient facial structures." In FG Net workshop on visual observation of deictic gestures, vol. 6, p. 7. FGnet (IST–2000–26434) Cambridge, UK, 2004.
[42] http://robotics.csie.ncku.edu.tw/Databases/FaceDetect_PoseEstimate.htm#Our_Database_
[43] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815–82.
[44] E. Littwin and L. Wolf, "The multiverse loss for robust transfer learning," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 3957–3966.
[45] O. Tadmor, Y. Wexler, T. Rosenwein, S. Shalev-Shwartz, and A. Shashua, "Learning a metric embedding for face recognition using the multibatch method," in Proc. NIPS, 2016, pp. 10–11.
[46] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in Proc. European Conference on Computer Vision (ECCV), 2016, pp. 499–515.
[47] D. Wang, C. Otto, and A. K. Jain, "Face search at scale," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1122–1136, Jan. 2017.
[48] Sankaranarayanan, Swami, Azadeh Alavi, Carlos D. Castillo, and Rama Chellappa. "Triplet probabilistic embedding for face verification and clustering." In 2016 IEEE 8th international conference on biometrics theory, applications and systems (BTAS), pp. 1-8. IEEE, 2016.
[49] L. Tran, X. Yin, and X. Liu, "Disentangled representation learning GAN for pose-invariant face recognition," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1–10.
[50] X. Peng, X. Yu, K. Sohn, D. N. Metaxas, and M. Chandraker, "Reconstruction-based disentanglement for pose-invariant face recognition," in Proc. IEEE International Conference on Computer Vision (ICCV), 2017, pp. 1–10.
[51] M. Patacchiola and A. Cangelosi, "Head pose estimation in the wild using Convolutional Neural Networks and adaptive gradient methods", Pattern Recognition, vol. 71, pp. 132-143, 2017.
[52] P. Sreekanth, U. Kulkarni, S. Shetty and M. S M, "Head Pose Estimation using Transfer Learning," 2018 International Conference on Recent Trends in Advance Computing (ICRTAC), Chennai, India, 2018, pp. 73-79.
[53] Zavan, F.H.D.B., Bellon, O.R., Silva, L. and Medioni, G.G., 2019. Benchmarking parts-based face

processing in-the-wild for gender recognition and head pose estimation. Pattern Recognition Letters, 123, pp.104-110.

[54] Ranjan, R., Patel, V.M. and Chellappa, R., 2017. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(1), pp.121-135.

[55] Xie, W. and Zisserman, A., 2018. Multicolumn networks for face recognition. arXiv preprint arXiv:1807.09192.

[56] Elharrouss, O., Almaadeed, N. and Al-Maadeed, S., 2020, February. LFR face dataset: Left-Front-Right dataset for pose-invariant face recognition in the wild. In 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT) (pp. 124-130). IEEE.

[57] Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[58] I. Kemelmacher-Shlizerman, S. M. Seitz, D. Miller, and E. Brossard, The MegaFace benchmark: 1 million faces for recognition at scale, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[59] I. Masi, A. T. an Tran, T. Hassner, J. T. Leksut, and G. Medioni, "Do we really need to collect millions of faces for effective face recognition?" in European Conference on Computer Vision (ECCV), 2016.

[60] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning deep representation for imbalanced classification," in CVPR, 2016.

[61] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In European Conference on Computer Vision., pages 499–515. Springer, 2016.

[62] Weiyang Liu, Yandong Wen, Zhiding Yu, and Meng Yang. Large-margin softmax loss for convolutional neural networks. pages 507–516, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016

[63] Y. Wu, H. Liu, J. Li, and Y. Fu, "Deep face recognition with center invariant loss," in Thematic Workshop of ACM-MM, 2017.

[64] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In IEEE conf Comput Vis Pattern Recog., volume 1, 2017.

[65] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) , 2017.

[66] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In CVPR, 2018.

[67] X. Zhang, Z. Fang, Y. Wen, Z. Li, and Y. Qiao, "Range loss for deep face recognition with long-tailed training data," in IEEE International Conference on Computer Vision (ICCV), 2017.

[68] X. Yin, X. Yu, K. Sohn, X. Liu, and M. Chandraker, "Feature transfer learning for deep face recognition with long-tail data," arXiv preprint, arXiv:1803.09014, 2018.

[69] ZHONG, Yaoyao, DENG, Weihong, WANG, Mei, et al. Unequal-training for deep face recognition with long-tailed noisy data. In : Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019. p. 7812-7821.

[70] HUANG, Chen, LI, Yining, CHEN, Change Loy, et al. Deep imbalanced learning for face recognition and attribute prediction. IEEE transactions on pattern analysis and machine intelligence, 2019.

[71] ZHAO, Kai, XU, Jingyi, et CHENG, Ming-Ming. Regularface: Deep face recognition via exclusive regularization. In : Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2019. p. 1136-1144.

[72] DENG, Jiankang, GUO, Jia, XUE, Niannan, et al. Arcface: Additive angular margin loss for deep face recognition. In : Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2019. p. 4690-4699.

[73] LIU, Bingyu, DENG, Weihong, ZHONG, Yaoyao, et al. Fair Loss: Margin-Aware Reinforcement Learning for Deep Face Recognition. In : Proceedings of the IEEE International Conference on Computer

Vision (ICCV). 2019. p. 10052-10061.