Journal of Computer Science

Original Research Paper

Face Verification for Person Re-Identification from Surveillance Camera and Drone-based Videos

[1*]Vimala Mathew, [1]K. Ramesh, [2]Tom Toby and [3]Anu Mary Chacko

[1]Department of Computer Science, Hindustan Institute of Technology and Science, Chennai, India
[2]Information Technology Group, National Institute of Electronics and Information Technology, Calicut, India
[3]Department of Computer Science, National Institute of Technology, Calicut, India

Article history
Received: 16-05-2021
Revised: 28-06-2021
Accepted: 03-07-2021

Corresponding Author:
Vimala Mathew
Department of Computer Science, Hindustan Institute of Technology and Science, Chennai, India
Email: vimalamathew@gmail.com

Abstract: Person re-identification in surveillance camera videos is attracting widespread interest due to its increasing number of applications. It is being applied in the field of security, healthcare, product manufacturing, product sales and more. Though there are a variety of methods to do person re-identification, face verification-based methods are very much effective. In this study, a deep learning framework to perform face verification in videos is proposed. Face verification deep learning model development includes different stages like face recognition, cropping, alignment, augmentation, image enhancement and face image selection for model training. The authors have put forward innovative methods to be adopted in various stages of this sequence to improve the performance of the models. The focus of this study is on these image preprocessing stages of the process, rather than the deep learning part, which makes the approach unique. The overall model is improvised by increasing the efficiency of each of these stages by adopting methods like face recognition and cropping based on face landmarks, effective training image selection using face landmark symmetry, various image augmentation techniques including perspective transformation and image enhancement methods like contrast stretching and histogram equalization. An average two percent increase is obtained in the accuracy of the face verification models by applying these methods.

Keywords: Face Verification, Person Re-Identification, Face Landmarks, Drone Video Analytics, Histogram Equalization

Introduction

Human beings can identify a person in a photo or video, by observing it from a viewable distance and angle. This identification is based on the overall appearance of the person, facial features, pose, voice and complexion, or even based on an accessory the person carries. When this person re-identification task is automated, one or a combination of the above methods can be applied. Person re-identification through face recognition and verification (Koide *et al*., 2017) is one of the easiest ways to achieve the best results for this task.

Person identification methods based on face recognition can be divided into two categories, feature based and metric-based methods. In the first approach, effective descriptors are used for representing each person. In the second approach, an effective distance calculation method is used to minimize the distance between the images of the same person and increase the distance between the images of the different persons. (Wang *et al*., 2018a)

This study focuses on one of the methods under the first category, person re-identification, based on face verification from videos using deep learning methods. Convolutional Neural Network (CNN) based deep learning models are feature-based classification methods, (Jogin *et al*., 2018) even though the features are not listed out explicitly.

The biggest challenges in face recognition tasks from videos are the limitations due to the poor quality of images extracted from videos, ineffective object localization methods (Koc, 2021) and inefficient image preprocessing methods (Liao *et al*., 2012).

The person re-identification using face verification involves three stages (Mathew *et al*., 2019):

1. Data collection: Collect a set of videos or images of a person or persons to be identified

2. Face Image Extraction: Process the image frames in the video and do face detection using face detection algorithms and extract the faces of the person or persons to be identified
3. Model Development: Develop a deep learning model for face verification using the above face images after doing required image preprocessing on the images which include image alignment, augmentation, selection, image enhancement, etc.

The face images are collected from running video sequences. The quality and pose of the face images extracted have a big impact on the performance of the face verification model (An *et al*., 2019). Though many studies exist in fine-tuning the deep model parameters, (Deng *et al*., 2019) (Wang *et al*., 2018b) (Parkhi *et al*., 2015) we have focused on improving the pre-processing stages to improve the performance of the model. The studies in this direction are less compared to the model fine-tuning approach (Liao *et al*., 2012) (Koc, 2021) (An *et al*., 2019). An effective face verification algorithm in which importance is given to the face image preparation stages has been proposed in this study. This algorithm systematically collects suitable face images, preprocesses them properly and develops models for face verification. Innovative methods in the face image preprocessing stages are the main contributions of the paper. With the help of this algorithm, it is possible to develop face verification models and identify persons from videos collected from surveillance camera videos more accurately.

The remaining part of this study is organized as follows. The Literature Review describes the related work categorizing them into different sections based on the different stages of the person re-identification framework. The Materials and Methods section describes the details of the datasets used and the proposed person re-identification framework, including the algorithm. It also includes the model development stage and the results of the face verification or re-identification stage. Results related to the different stages of the algorithm like face detection, face alignment, selection, augmentation, symmetric face selection and perspective transformation is included next. A comparative study with other existing methods is carried out in the discussion section. Finally, the Conclusion and Future Scope are included.

Literature Review

Many studies have been carried out on automated person re-identification and activity detection to develop intelligent systems that reduce human effort in, home and public environments (Thakur and Han, 2021). The different methods used for person re-identif/ication tasks include Sequence Information (Hu *et al*., 2018). Saliency Learning, (Zhao *et al*., 2017), Pose-aware method (Cho and Yoon, 2018), Multi-shot ranking (Karanam *et al*., 2019), Correlation-based method (Hsu *et al*., 2018), PCA and Eigon Faces (Meedeniya and Ratnaweera, 2007), etc. A multimodal method of face and body matching is used in (Koo *et al*., 2018).

In this study, the focus is on the face detection-based person re-identification method. The other challenges in implementing person re-identification using face detection from videos are occlusion (Zhuo *et al*., 2018), poor quality of the video, (Hitesh *et al*., 2017), partial faces (Liao *et al*., 2012), etc. The nature of the input dataset is always a factor that affects the performance of any model (Mathew *et al*., 2018).

In (Deng *et al*., 2019), the authors have tried to improve the performance of the deep convolutional neural networks for face recognition model using angular loss function. (Wang *et al*., 2018b) have used a large margin cosine loss to improve the performance of the model.

A face verification model development process involves a sequence of steps and there exists a possibility of improving the model by incorporating improvements in any of these stages. A sequence of steps is applied in the proposed algorithm to make the input face images sharp and perfect before model training. The existing works in these stages are listed below.

Face Detection

Face detection is the identification of the location of one or more faces that exist in an image frame. In most cases, it returns the bounding rectangles in which the face image exists (Ranjan *et al*., 2019). Face detection algorithms have received much attention in the last decade and a variety of algorithms are generally accepted for the purpose. Viola and Jones (2001), Multitasking Convolutional Neural Networks (Zhang *et al*., 2016) and FaceNet (Schroff *et al*., 2015) are the most popular ones which provide considerable accuracy. A landmark-based algorithm is proposed in this study to improve face detection accuracy.

Face alignment and Augmentation

When face images are captured from videos and taken as input for model development, the factors that bring down the model accuracy are the lack of alignment of face images and the presence of partial faces (Liao *et al*., 2012). So, it is required to align all the face images captured to the same orientation and filter out partial faces. Some of the methods used for this purpose are Face alignment using regressing local binary features (Ren *et al*., 2016), Adaptive Pose Alignment for Pose-Invariant Face Recognition (Liao *et al*., 2012) and Viewpoint-Consistent 3D Face Alignment (Tulyakov *et al*., 2017).

Another important factor that affects the orientation of images is the position of the camera. They are fixed mostly on the walls of a room or traffic posts that are above the view level. Studies about image correction to nullify the effect of position and angle of the camera have

been conducted earlier. Mean shift clustering and Laplace linear regression were suggested for automatic radial distortion (Tang *et al*., 2019). When the camera is positioned above view level, perspective transformation is a useful option to cancel the distortion on object images introduced by the position of the camera. (Ansari and Shim, 2019).

Face alignment is an important image preparation stage that results in better performance of models. Joint face alignment is one such method where optimization was performed iteratively and sequentially (Zhang *et al*., 2020). Existing face alignment methods are explored and a new method with better performance is proposed in the study. The most aligned faces are selected based on this method for the model developed.

Enhancement of Face Images

Fuzzy-based illumination normalization of face images is applied in (Nasution *et al*., 2014) as a face enhancement step. Collaborative Random Faces Guided Encoders are used in (Shao *et al*., 2017).

Two face enhancement methods, Histogram Equalization (H. E.) and Contrast Stretching (C.S.) are proposed as effective methods for face enhancement in this study.

Compute Face Embedding

When the input face images used for model development are extracted from videos in different contexts and completely different attire and accessories, the direct deep learning method for face verification fails considerably. A method to overcome this scenario is to extract the face features first using the face embedding method and then apply deep learning for face verification. (Schroff *et al*., 2015).

Materials and Methods

Datasets

Five different datasets were used in the study which is listed below.

a. YouTube Faces Dataset

Originally collected from YouTube by its contributors, the purpose of this dataset is to do face recognition from videos. 3425 videos of 1595 people are included in the dataset. The videos are split into frames and stored in different folders. The smallest video includes 48 frames and the longest video includes 6070 frames with an average frame count of 181.3 for each video (Wolf *et al*., 2011). As the dataset is too large, a subset of this dataset was used for testing the algorithm. Removed the subjects having only very few images, especially to obtain a balanced training dataset.

b. Children Data Set

Using an 8MP phone camera, videos of five children were captured to do face detection and verification. The set of videos include 15 videos of five teenage children with 3 videos of each child. Some videos which include more than one of these children were also collected for verification and marking the person in the video.

Smartphone cameras are the next growing source of video and images. This way, the dataset is intended to represent real-world situations and hence make the model suitable for a wide range of applications from multi-type sources.

c. Choke Point Dataset

It is a dataset sponsored by NICTA (National Information and Communications Technology Australia) designed for carrying out person re-identification tasks (Wong *et al*., 2011). From this collection, the dataset used in the study includes cropped face images from videos of 30 people with an average of 50 images per person.

d. Film Star Dataset

It is a dataset of two popular film actors collected from the Kaggle website and it was enhanced by adding more images from the internet. It is an image dataset that includes around 1000 images of each actor. Since they are very popular film actors who have acted in several roles in different films, the images include a variety of appearances of the same actor. This dataset is being used for the development of a binary classification model for verifying the face of these two actors. The dataset includes two folders in which the two sets of images are stored.

e. Drone Face Dataset

Though many datasets are available for face identification (Yang *et al*., 2016), the dataset most suitable for performance analysis of the deep learning models based on the position of the surveillance camera is drone face (Hsu and Chen, 2017).

In the drone face dataset, it is possible to study the face recognition performances from drone videos. The part of the dataset used in the experiment includes:

- 1364 face images of 11 persons including 7 males and four female candidates. The face images are between 23 × 31 and 384 × 384 resolution. All are frontal face images.
- For each person, four sets of images are included. These four sets of images include images captured from cameras positioned at different heights 1.5, 3, 4 and 5 m from the person's head level. For each camera height, there are 31 images captured at 2 to 17 meters away from the person at 0.5 m difference.

- All images were taken in daylight.

f. Newsreader Dataset

It is a video dataset of ten television newsreaders collected from popular news channels. Different clippings of 10 newsreaders were collected. They were clipped into frames and stored in 10 different folders each containing minimum of 1000 images.

In this study, the main objective is to study the performance of the proposed preprocessing methods. In addition to the benchmark datasets, authors have created some additional datasets as listed above. While preparing the additional datasets, it was ensured that enough input videos are included to avoid class imbalance problems, after certain stages of preprocessing like landmark-based face detection and face symmetry-based image selection.

Proposed Person Re-Identification Framework

The proposed person re-identification framework comprises of different stages. (1) Face detection and extraction. (2) Face image selection, alignment and augmentation. (3) Face image quality enhancement. 4) Face embedding computation, if required based on the input dataset. (5) Model development using extracted images and (6) Face verification using the developed model. Figure 1 shows an overview of the person identification framework.

The different stages in the proposed person identification framework are explained below in detail.

Face Detection and Extraction

A video is a collection of image frames. To extract face images from the video, image frames are captured at regular intervals and each of these image frames is analyzed to detect face images using a face detection algorithm. The performance of different face detection algorithms was compared by developing models using face images extracted with each of these algorithms. Model, based on the Viola and Jones (2001) and model based on Multitasking Convolutional Neural Network (MTCNN) (Zhang *et al*., 2016) algorithm have produced comparatively better performance. The MTCNN based model was found to be the best out of these two algorithms (Mathew *et al*., 2019).

Algorithm 1: Person re-identification model development algorithm

1. Collect video dataset of the persons to be identified with frontal face videos and initialize the following
   Set limit L(x, y) as the minimum resolution for face images.
   Select two face detection algorithms alg1 and alg2 for the different stages of the process.
2. Face detection and extraction:
   a. From n videos vd_1, vd_2, vd_3,…vd_n of a person in training dataset, select each video vd_i and capture image frames fr_1, fr_2, fr_3 …fr_m at fixed time intervals.
   b. For each frame fr_k, apply alg1 to detect face image if any and perform steps i & ii
      i. If w and h are the width and height of the detected face image, extract the face image in an increased width and height by adding a constant c to w and h.
      ii. w1=w+c
      iii. h1=h+c
      iv. Save it as an image file to a folder PA1.
   c. Repeat step (a) and (b) for all the videos vd_1, vd_2, vd_3, .. , vd_n of the person P_i and store them in the folder PA_i.

   Repeat steps (a) (b) and (c) for all persons and store the images to separate folders PA_1, PA_2, PA_3,.., PA_k.
3. Face image selection:
   For each folder PA_i, in PA_1, PA_2, PA_3,..,PA_k, select each folder and repeat steps (a) to (d).
   a. Filter out the image, if the resolution of an image is less than the limit L(x,y).
   b. For each face-image, apply the second level of face detection using alg2, to filter out non face images if any.
   c. Apply face landmark-based face selection algorithm to ensure removal of all non-face images which were not detected during step 2 and 3b.
   d. Store the selected face images in folders PB_1, PB_2, PB_3,.., PB_k, respectively
4. Face image alignment:
   For each folder PB_i, in PB_1, PB_2, PB_3,.., PB_k, select each face image in PB_i and repeat the steps a to c.
   a. Rotate the face image to cancel any angular shift.
   b. Apply perspective transformation, if the images are taken from above in an angle.
   c. Store it in folders PC_1, PC_2, PC_3,.., PC_k.
5. Symmetric Image Selection:
   For each folder PC_i, in PC_1, PC_2, PC_3,..,PC_k select each face image in PC_i and repeat the steps a to c.
   a. Select the most symmetrical face images. Limit the number of images for all persons to uniform count, to have a balanced training dataset.
   b. Store it in folders PD_1, PD_2, PD_3,.., PD_k.
6. Face image enhancement:
   For each folder PD_i, in PD_1, PD_2, PD_3,..,PDk select each face image in PDi and repeat the steps (a or b) and c below.

a. Apply Contrast Stretching.
b. Apply histogram equalization
c. Store it in folders PE_1, PE_2, PE_3,.., PE_k.

7. Develop a face classifier model with face images from PE_1, PE_2, PE_3,.., PE_k folders.
8. Evaluate the algorithm using the datasets.

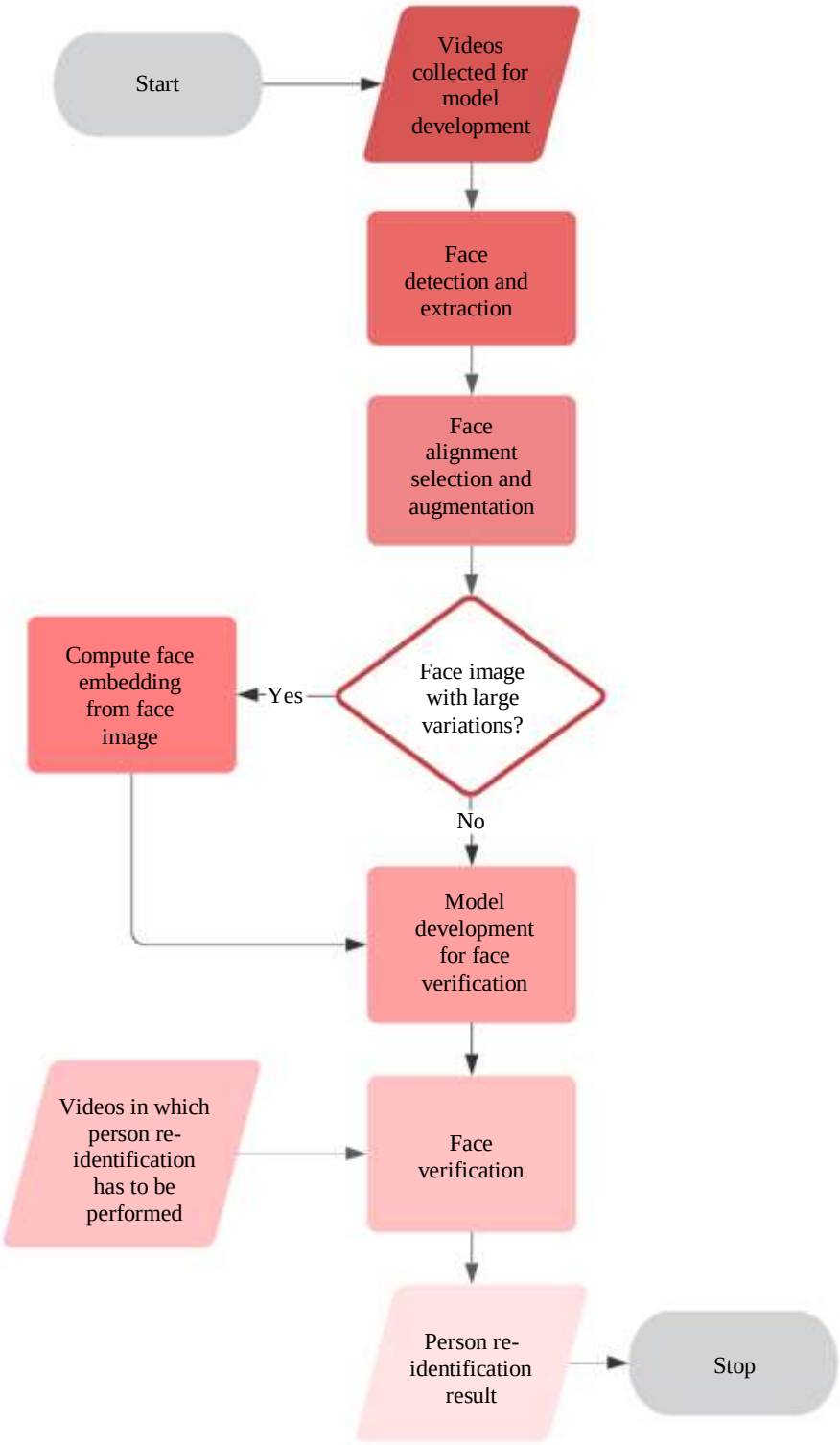


Fig. 1: Flowchart of the proposed person re-identification framework from videos

In this study, the Viola-Jones algorithm and MTCNN algorithms were explored for their suitability and performance for face detection from videos. A landmark-based algorithm that gives better performance in the context of person re-identification from videos is also proposed.

Face Image Selection

In the proposed person re-identification framework, face extraction and filtering are done in different stages. One of the algorithms, Viola-Jones is used in stage1 as shown in Fig. 2 for extracting face images and they are stored in different folders each corresponding to a different person in the dataset.

While analyzing raw video for face detection, it is found that many non-face images were detected as face images due to the deficiency of the face detection algorithm. After completing stage 1 in Fig. 2, many non face images were thus captured as face images into the folder. One effective mechanism to filter out such non face images is to apply a sequence of face detection algorithms to filter out the non-face images completely, as in Fig. 2. Irrespective of the algorithm being applied, it is most important to apply more than one algorithm in a sequence to filter out all non-face images from the dataset.

The quality and completeness of the face images used for model development are always a factor that affects the performance of the model used for face verification. The landmark-based face detection algorithm used in stage 3 of Fig. 2 is based on landmark point detection. In this algorithm, the face images are extracted only if the face images contain all the face landmarks in the face. A face with landmark points marked is shown in Fig. 3a.

If all face landmark points are not achievable in the landmark detection stage, drop the face from the folder or regenerate the face by flipping side by side based on the one-half available, if face images are less.

Different methods to improve the quality of input face images undergoing classification are proposed here. Remove the face images having poor resolution. The resolution of input face images is a factor that affects the performance of the face verification model. If the resolution is good it can be easily detected. When the face images with low resolution are removed the number of samples will be reduced. When the input video collection is large this can be ignored as there will be enough face images.

If the face is small, scale it with interpolation. If the input video collection is large, this can be ignored. There will be enough images even if not regenerated with scaling.

Another issue to be handled in the image selection stage is the number of samples. If there exist enough input videos with the person to be identified in it, it is easy to select the most suitable face images to make the model. But in some cases, there will be only very few videos of the person to be identified. In that case, utilize all the available face images without filtering out many faces even if they are not perfect.

All datasets will have male and female entities. In the model architecture, one more level can be added with a male-female detection first and then identify the person. This will reduce the classification problem size to half and better accuracy can be achieved.

Face Alignment and Symmetric Face Selection

A challenging problem related to Person identification from surveillance videos is the misaligned, distorted faces and faces twisted in angle in the detected face images. The faces extracted from videos are not always aligned. It can be side faces or rotated faces also. When these face images are used for model development, the accuracy of the models will be affected (An *et al*., 2019). Due to this, a set of face alignment methods are used to avoid these problems. Even though good quality face detection systems can detect slightly misaligned faces the performance of the face classifier is found to reduce when misaligned face inputs are also considered during model development. Face orientation correction and Frontal face detection need to be applied to avoid this issue.

a. Rotation

If the face is misaligned rotate it. Extract face using a face detection algorithm and find out all face landmarks. To align the face, use the face landmarks.

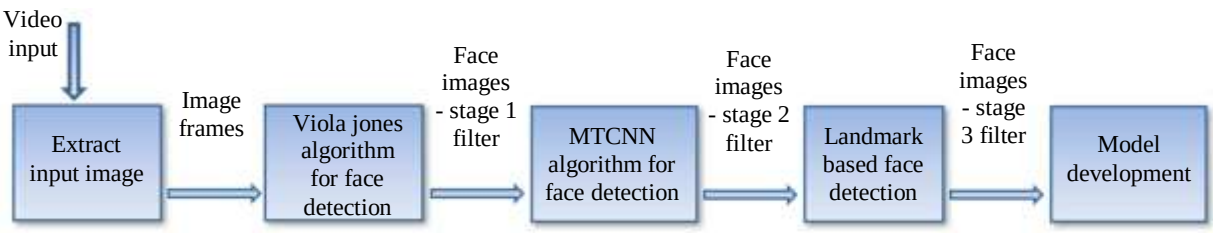


Fig. 2: Block diagram for the face detection stages of the proposed person re-identification framework

Face-landmarks are a set of points marked on the faces. Obtain the angle of rotation of the face from two face landmark points on a straight path. The two points from the corner of two eyes are taken here. Now calculate the angle of rotation of the face image as below:

$$\tan\varnothing = \left(\left(y2 - y1\right) / \left(x2 - x1\right)\right) \quad (1)$$

$$\varnothing = \tan^{-1}\left(y2 - y1\right) / \left(\left(x2 - x1\right)\right) \quad (2)$$

Where:
(x_1, y_1) = The left corner of the left eye
(x_2, y_2) = The right corner of the right eye

Rotate the face through this angle. A face that has been rotated in this way is included in Fig. 3. Face image before rotation and after rotation is depicted in Fig. 3a and 3b respectively.

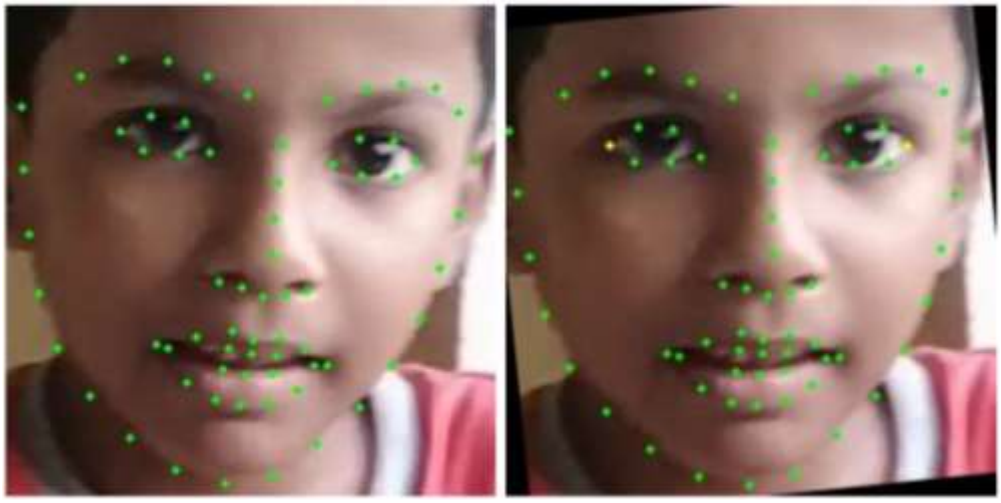
b. Symmetric Image Selection

Faces lacking symmetry in landmark points may be partial images. Therefore, remove it before model development. A mathematical method to check the symmetry of faces has been adopted to select the most suitable images to be included in the model. The mean square error from the axis can be calculated based on the position of the landmark points for finding the more symmetric images. Face images of different symmetry error values are described in Fig. 4a to 4e.

Face images before and after applying landmark points based symmetrical image selection is shown in Fig. 5.

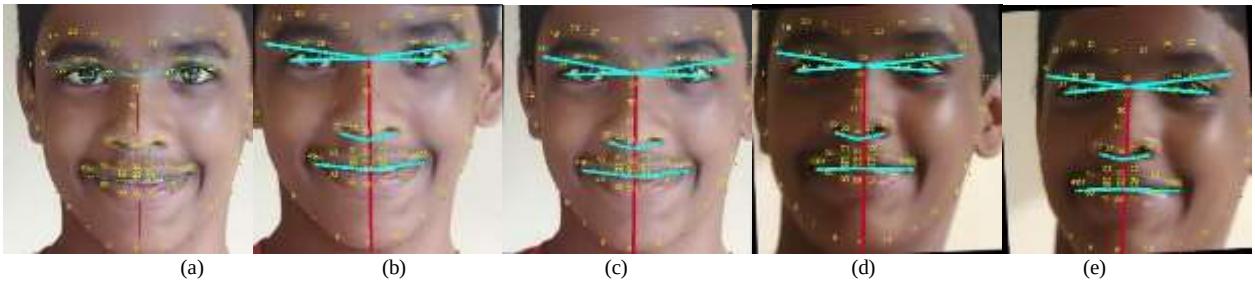
c. Image Distortion Correction using Perspective Transformation

In a practical scenario, the surveillance camera is fixed at an elevated position irrespective of whether they are indoor or outdoor. It is fixed on the sidewalls of the room, a traffic post, or even in a drone.



(a) (b)

Fig. 3: Face images before and after rotation; (a) Before rotation; (b) After rotation



(a) (b) (c) (d) (e)

Fig. 4: Symmetry Error (SE) values in points. From (a) to (f) the symmetry of the face image reduces; (a) SE = 0; (b) SE = 4; (c) SE = 24; (d) SE = 56; (e) SE = 84

Algorithm 2: Algorithm to obtain the most symmetrical face images

a. Input face images.
b. Initialize the count n, the number of face images to be selected.
c. For each face in the input face image collection, perform steps i and ii.
   i. Obtain the landmark points. Out of the 64 landmark points detected, select 4 horizontally opposite pairs of points.
   ii. For each such pair, compute their distance to the vertical axis aligned to the middle of the face passing through the face landmark points in the exact middle of eyes, nose and lips and compute the sum of the deviations, including the sign of the obtained result, using the following formula. $d = \sum_{i=1}^{4} X_i - X$, Where x corresponds to the vertical axis passing through the middle of the face and x_i that of landmark points.
d. In the available face image set, obtain n such faces with the minimum value for the d value.

In such cases, since the camera lens and the object are not always in parallel planes, image distortion occurs in the captured images. The important distortions that can occur are radial and tangential distortion. Camera calibration methods are used to obtain rectified images from distorted images (Beauchemin and Bajcsy, 2001).

The amount and type of distortion depends on the position and orientation of the camera against the object. In the case of face images, there can be a large variation between normal and distorted images as in Fig. 6. While applying face recognition models, this distortion introduced by the camera position will reduce the effectiveness of the face verification models.

The way the values of the projected coordinates change when images are captured is indicated in the Fig. 7.

The projections are obtained using the Perspective projection equations:

$$x = d * (X / Z) \quad (3)$$

$$y = d * (Y / Z) \quad (4)$$

Where:
(X, Y, Z) are co-ordinates of the scene point
(x, y, d) are coordinates of the image point

To obtain the corrected image, after capturing the images without camera calibration, perspective transformation can be applied (Ansari and Shim, 2019).

The perspective transformation matrix is calculated using 4 sets of points in source and destination images similar to the process of camera calibration using chessboard images (Beauchemin and Bajcsy, 2001). Then perspective transformation is applied to the distorted face images with this perspective transformation matrix to rectify this distortion.

After extracting face images from the video. perspective transformation is applied to reduce the distortion that occurred to the face image. When the perspective transformation was applied to the first level distorted images in the drone face dataset an accuracy improvement of 2% was achieved in the face verification model. Figure 8 describes some sample face images on which perspective transformation was applied.

Face Image Augmentation

To introduce variety in the input face images, suitable augmentation methods were applied during the model construction stage.

The following methods were tested using the programming framework before making the Convolutional Neural Network (CNN) model for face verification purposes:

- Horizontal shifting inside the image
- Vertical shifting inside the image
- Horizontal flip
- Vertical flip
- Varying brightness
- Rotating the images from different angles

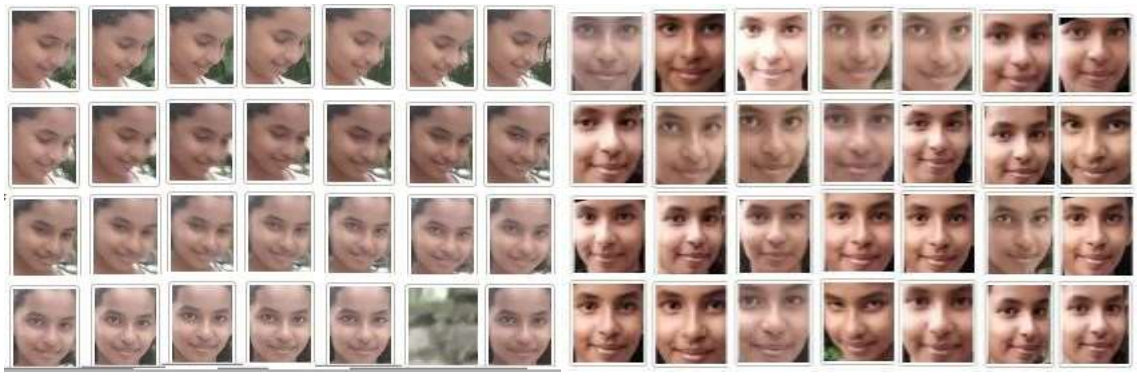


Fig. 5: Face images before and after applying symmetry-based selection (a) Face images before cleaning; (b) Face images after cleaning having frontal faces and same orientation

Fig. 6: Distortion of a face when viewed from different aerial positions

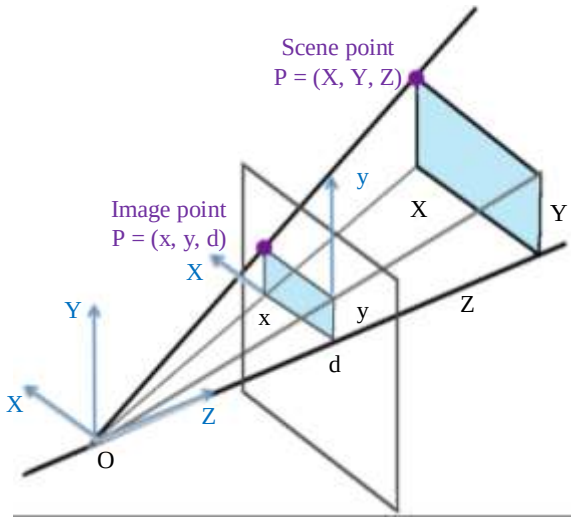


Fig. 7: Perspective transformation

Fig. 8: Original image and image after perspective transformation for 3 subjects in the children dataset

Face Image Quality Enhancement

Changes in lighting in the scenes may change the intensity of images captured in videos (Starovoitov *et al*., 2003). The videos and images captured from various sources will be of different quality and the quality of images may, in turn, affect the performance of the model being developed. Two methods are proposed to make the quality of images uniform and improve model performance.

a. Histogram Equalization (H.E.)

Histogram equalization is an important method for image enhancement. (Honnutagi and Maranur, 2018) (Akila, 2017) After applying histogram equalization, an image with uniform distribution of pixel values is obtained (Jeon and Kim, 2016). Image processing like Histogram equalization is found to increase the performance of the neural network model. (Bertalmío, 2019).

For an image, histogram equalization is done using the following procedure.

i) Calculate the histogram of the image pixels intensity values and obtain the L different values of $f[x, y]$ using the following formula:

$$h[i] = \sum_{x=1}^{N} \sum_{y=1}^{M} \begin{cases} 1, & if \ f[x,y] = i \\ 0, & otherwise \end{cases} \quad (5)$$

Where:

L = 256, the number of intensity levels of the image $M \times N$ is the size of the image.

ii) Calculate the Cumulative Distribution Function for the intensity values (CDF)

$$CDF[j] = \sum_{i=1}^{j} h[i] \quad (6)$$

Where:
$h[i]$ = Obtained using Eq. 5

Scale the input image with the cumulative distribution function to generate the output image.

$$g[x,y] = \frac{CDF[f[x,y]] - CDF\min}{(N \times M) - CDF\min} \times (L-1) \quad (7)$$

Where:
$CDF_{\min}$ = The minimum nonzero value of the cumulative distribution function
M × N = The size of the image L = 256, the number of intensity levels of the image

b. Contrast Stretching(C.S.)

Contrast stretching is applied to improve the contrast in an image by increasing the range of intensity values of the pixels. (Ruikar *et al*., 2018). It is applied to many image analysis algorithms as a preprocessing step. (Cao and Li, 2020). Increasing image contrast this way can improve the quality of the image and it is an important image enhancement method in various applications (Erwin and Ningsih, 2020).

Contrast stretching is obtained by the following formula:

$$g(x,y) = \frac{f(x,y) - f_{\min}}{f_{\max} - f_{\min}} * L - 1 \quad (8)$$

Where:
$f_{\min}$, $f_{\max}$ = The minimum and maximum pixel intensity of the pixels in the input image
L = 256 = The number of grey levels of the image

This method increases the image contrast and thereby improves the model performance. Better performance is obtained from models developed after applying any of these two image enhancements.

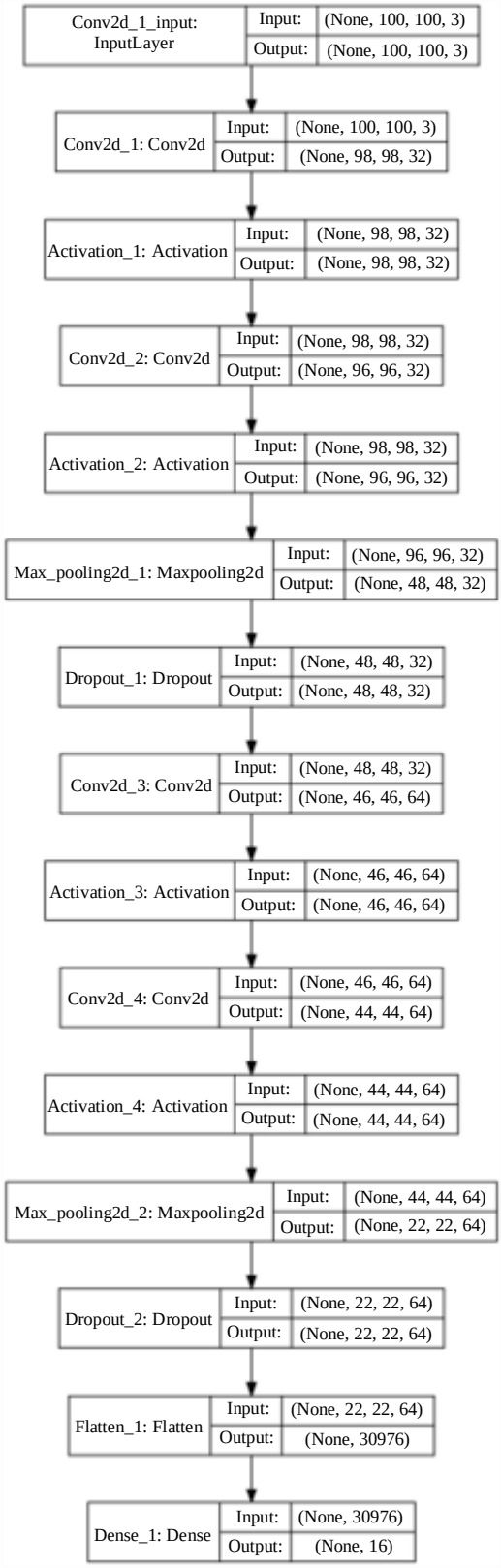


Fig. 9: The general structure of the CNN model adopted for face verification models

Model Development for Face Verification

The face images were prepared by adopting the various steps in the proposed algorithm for each of the datasets. The images were prepared into five different folders, based on the image processing steps applied. They are existing MTCNN algorithm-based face image cropping, Face landmark-based face image cropping, Rotation and Face landmark-based Symmetry, Histogram Equalization and Contrast Stretching. Five models each were prepared for each dataset to evaluate the performance of the stages in the algorithm by considering the above five categories of images in its training set. Use a convolutional neural network (Wang *et al*., 2018a) to develop a deep learning model for face image classification. Results of the performance comparison of these models are included and discussed in the Results and Discussion section.

The general structure of the Convolutional Neural Network used in all the models is shown in Fig 9. The effect of applying perspective transformation for face alignment was tested with the Drone Face dataset which is suitable for studying the performance of face alignment when the camera is fixed above view level. The models developed with images before and after applying perspective transformation were compared for performance.

Face Verification

During face verification, all the pre-processing stages applied during image preparation for model development have to be carried out first. It includes the sequence of face detection and cropping stages, face rotation, histogram equalization, contrast stretching, etc. After these stages, prediction is carried out. The confusion matrix and accuracy were computed by providing a test set of face images. The results are included in Table 1.

Compute Face Embedding

While experimenting with person re-identification from videos, it is found that CNN based image classifier does not perform well when the input face images are extracted from videos captured at a very different context other than the one being predicted or if the person being verified is wearing different face accessories or having the different facial appearance, etc. The face embedding calculation method described in the Face Net paper (Schroff *et al*., 2015), gives a better performance in this context. In this method, face embedding computation is performed first and after which a deep learning model is prepared from these computed results.

While doing face verification using face embedding, the process can be enhanced by incorporating a clustering layer to classify the face images of a scene first. Computed face embedding values can be used for this clustering. The number of clusters is identified using the elbow method (Marutho *et al*., 2018). This clustering procedure will help to fix the number of persons in a scene first and then the task of person verification is simplified.

Results

Figure 10a and 10b contains the accuracy and loss curves based performance comparison for the model developed using existing MTCNN based face image cropping against the model developed using Face landmark points-based face images for the YouTube Faces dataset. The proposed face verification pipeline recommends more than one face image detection and cropping algorithm during its different stages. But, while comparing the performance of proposed Landmark-based cropping against the MTCNN algorithm in the pipeline, the accuracy curve is found to be more stable and resulted in better accuracy as shown in Fig. 10.

Similarly, the performance comparison of the accuracy of the models for the five datasets and five different image enhancement methods is summarized in Table 1.

Model Performance for different Datasets and Image Enhancements

The performance improvement on the application of perspective transformation was tested using the drone face dataset and an accuracy of 99.5% was obtained after filtering out very low-resolution images from the dataset. Earlier work has reported 89.9% accuracy in the droneface dataset (Bustos *et al*., 2018).

For a collection of 5000 images of 100X100 resolution, the face-landmark-based symmetry calculation, histogram equalization and contrast stretching took around 15000ms, 22000ms and 900ms respectively in an i7 processor with 32GB, including the time required for storing the images created. A comparison of image preprocessing time for various methods used in this study with the multimodal method in (Koo *et al*., 2018) is included in the Table 2.

Table 1: The accuracy values of different models developed for different datasets with and without specific augmentation or enhancement methods

Dataset	MTCNN Based	Landmark based	Rotation and symmetry	H. E	C. S.
YTB	0.950	0.960	0.970	0.990	0.999
Children	0.964	0.978	0.985	0.986	1.000
ChokePoint	0.938	0.947	0.907	0.988	0.987
Film stars	0.868	0.860	0.888	0.915	0.898
Newsreader	1.000	1.000	1.000	1.000	1.000

Table 2: Time required for proposed face image pre-processing methods per image against the existing methods, in milli seconds (ms)

Method	Processing time(ms)
Multimodal method (Koo *et al*., 2018)	98
Face landmark-based symmetry calculation	3
Contrast stretching	1
Histogram equalization	5

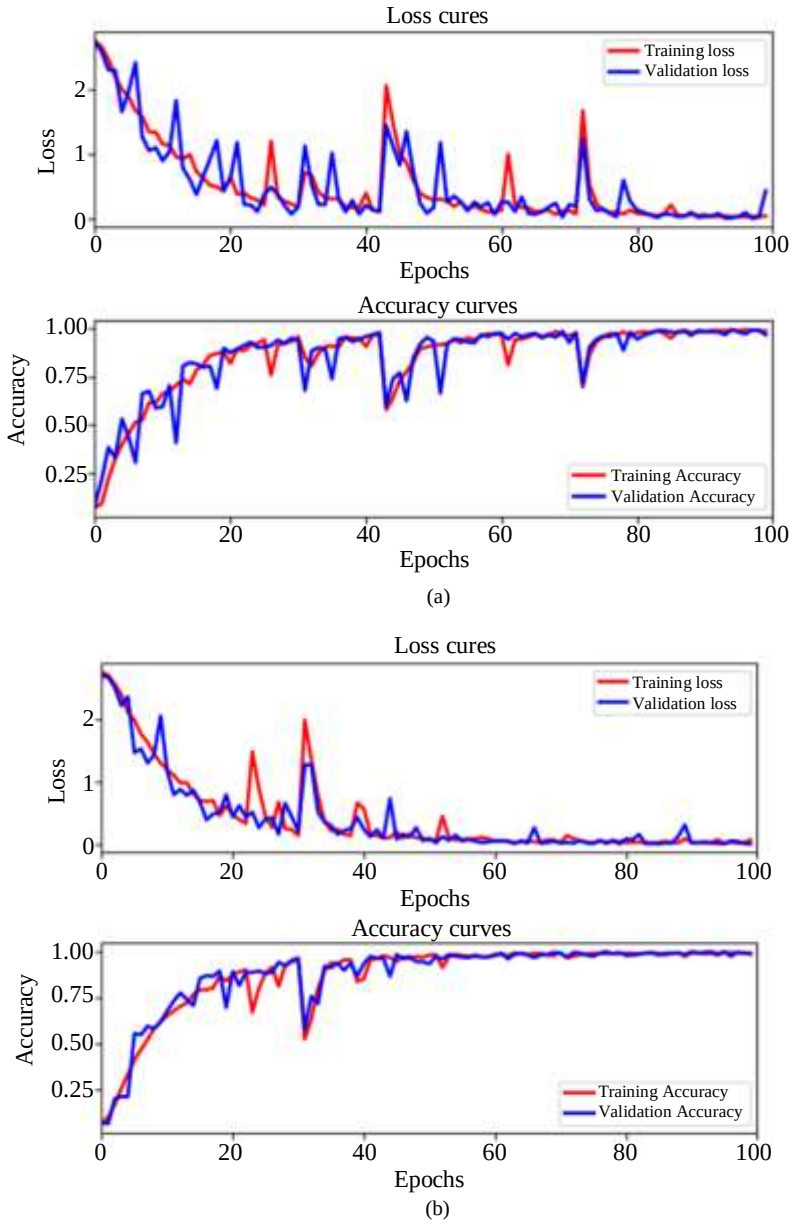


Fig. 10: Performance comparison of models developed using faces extracted with MTCNN face detection method and proposed landmark-based face image cropping method

Discussion

As per these results, a minimum one percent accuracy improvement is obtained when proposed landmark based face image detection and cropping is adopted compared to the Normal Multi-Tasking CNN-based algorithm.

It is also clear from Table 1 that the face alignment methods like Rotation and proposed Face Landmark Symmetry-based image selection also resulted in improvement in the accuracy by around two percent compared to the normal images.

The face image enhancement methods histogram equalization and contrast stretching also improved the accuracy by around three percent compared to the normal images as listed in Table 1.

The graphs in (Fig. 11a, 11b, 11c and 12a, 12b, 12c, 12d) show the comparison of accuracy curves of different datasets for five different image enhancement methods. To avoid cluttering of the images the comparison is done by splitting the graphs into two sets. Comparison of normal, landmark-based and symmetrical images is included in Fig. 11. Comparison of models developed using Normal images with Histogram equalization and Contrast stretching of images is included in Fig. 12.

Consistent performance improvement for the proposed methods in different datasets is clear in the graphs plotted. The accuracy curves for different datasets are not similar as the number of images and features of the images in the different datasets vary widely and the number of epochs for which the model training was carried out to get maximum performance also is different.

In addition to these results, the effect of applying perspective transformation on the drone face dataset was approximately two percent improvement compared to the normal dataset.

To further verify how the proposed algorithm performs in comparison with the existing algorithms a comparison study is performed. Out of the six datasets used, three are benchmark datasets for face verification tasks and the performance comparison of our algorithm with methods used in other studies where they have used the same datasets are shown in Table 3 and 4.

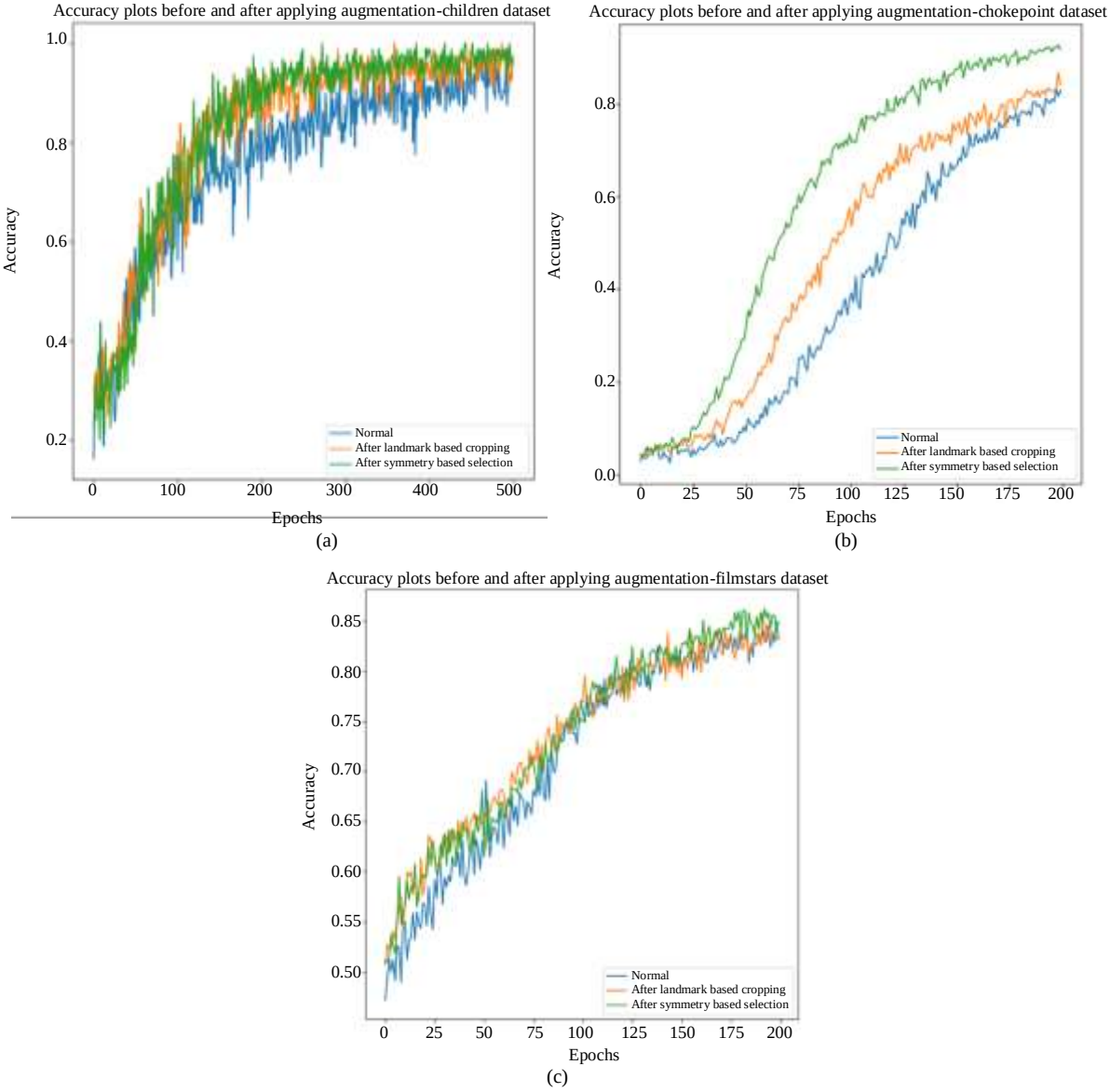


Fig. 11: Comparison of normal, landmark-based and symmetrical images

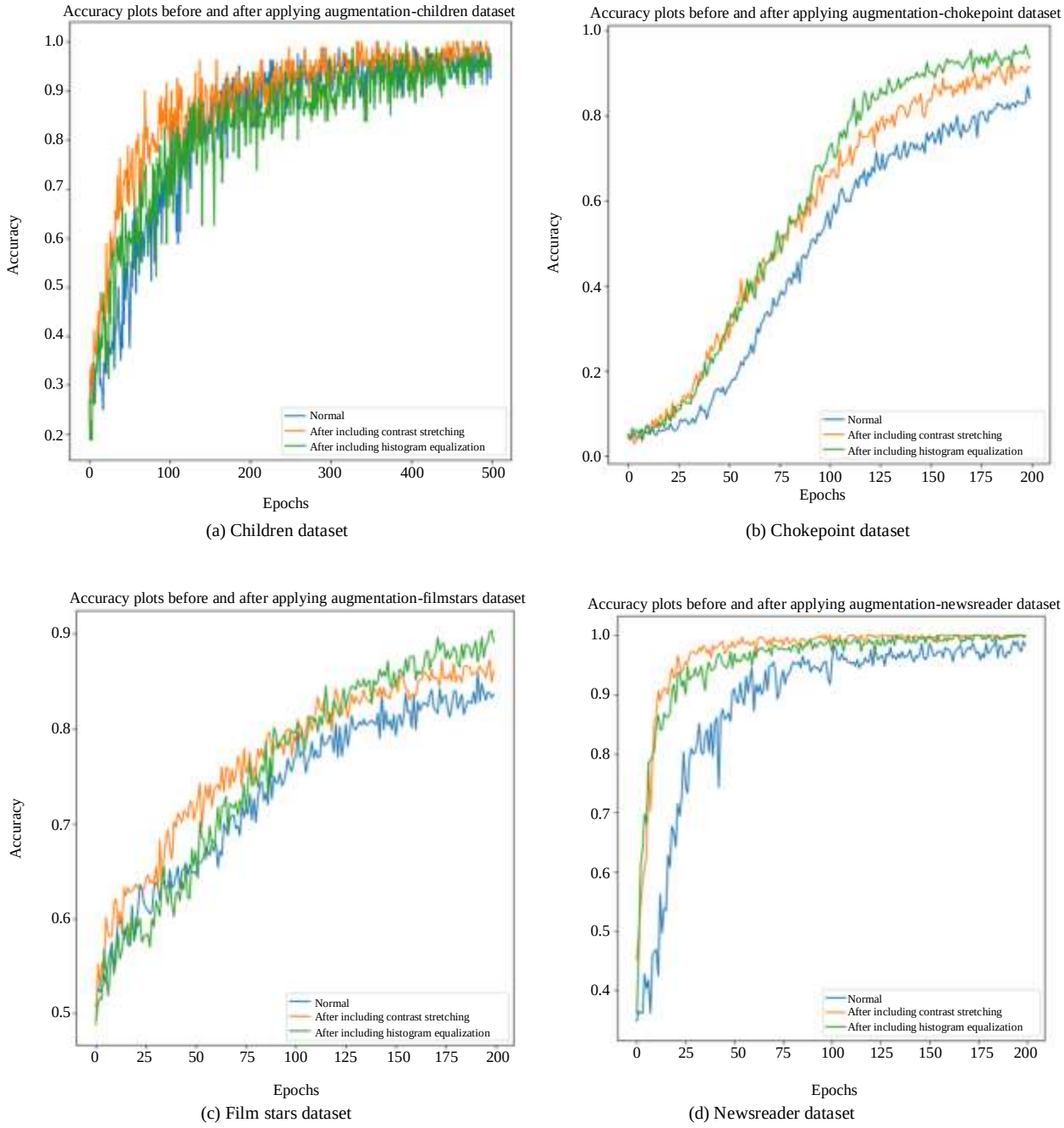


Fig. 12: Comparison of models developed using normal images against histogram equalized & contrast stretched images

Table 3. Accuracy Comparison of ChokePoint dataset with results obtained in other studies

Method	Accuracy percentage
QAF Method (Wong *et al*., 2011)	
S+V Model (Mokhayeri and Granger, 2019)	89.20
CCM-CNN (Parchami *et al*., 2017)	98.80
Face landmark-based symmetry	90.70
Contrast stretching	98.70
Histogram equalisation	98.80

Table 4: Accuracy Comparison of a subset of Youtube Faces dataset with results obtained in other studies

Method	Accuracy perc
Parkhi *et al*. (2015)	97.40
Wang *et al*. (2018a)	97.60
Deng *et al*. (2019)	98.02
Wang *et al*. (2018a; Hu *et al*., 2018)	98.12
Face landmark-based symmetry	97.00
Contrast stretching	99.00
Histogram equalization	99.90

One limitation of the algorithm is that while applying face landmark-based face image extraction and landmark symmetry-based selection of face images, a large number of images are required as input. Otherwise, face images that do not match the required criteria will be filtered out by the algorithm and dataset size will get reduced that may, in turn, reduce the performance of the algorithm.

Conclusion and Future Scope

The algorithm for person re-identification using face verification has been developed and tested using five different datasets. The performance of the face verification model is enhanced when the proposed sequence of face detection, face selection, alignment, augmentation and enhancement methods was used. The landmark-based symmetry calculation algorithm is highly efficient in the face selection process to select the most suitable face images for model development. The accuracy improvement obtained by applying landmark points symmetry-based face image selection was two percentage. It is observed that applying histogram equalization and contrast stretching of the face images, also improved the performance of the face verification model by around three percent. The important method used in the face-alignment stage is the application of perspective transformation. The improvement obtained by applying perspective transformation was also two percentage. The different possibilities to improve the face orientation, alignment and quality were tried rigorously and results were presented.

The proposed algorithm with its novel steps like face symmetry-based face image selection, preprocessing using perspective transformation, contrast stretching and histogram equalization helps the scientific community in implementing accurate face verification models. These concepts can also be extended to other areas like emotion recognition and activity detection where face image-based model construction is required. It is also planned to explore more preprocessing methods and the selection of better models and their parameters in future work.

Acknowledgement

The authors thank the National Institute of Electronics and Information Technology, Calicut, India for providing resources for carrying out the research.

Author's Contributions

Vimala Mathew: Conception, design, data collection and drafting the article.

K Ramesh: Conception and critical revision of the article.

Anu Mary Chacko: Conception and design.

Tom Toby: Conception and critical revision of the article.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Akila, S. B. (2017). Color Image Enhancement by Brightness Preservation using Histogram Equalisation Technique. International Journal of Engineering Research and, Technology 5(9), 1-3. https://doi.org/10.1016/j.ijleo.2015.08.173

An, Z., Deng, W., Hu, J., Zhong, Y., & Zhao, Y. (2019). APA: Adaptive pose alignment for pose-invariant face recognition. IEEE Access, 7, 14653-14670. https://ieeexplore.ieee.org/abstract/document/8620989

Ansari, I., & Shim, J. (2019). Vehicle Manufacturer Recognition using Deep Learning and Perspective Transformation. Journal of Multimedia Information System, 6(4), 235-238. https://doi.org/10.33851/JMIS.2019.6.4.235

Beauchemin, S. S., & Bajcsy, R. (2001). Modelling and removing radial and tangential distortions in spherical lenses. In Multi-Image Analysis (pp. 1-21). Springer, Berlin, Heidelberg. https://link.springer.com/chapter/10.1007/3-540-45134-X_1

Bertalmío, M. (2019). Histogram equalisation and vision models. In Vision Models for High Dynamic Range and Wide Colour Gamut Imaging (pp. 157-184). https://www.elsevier.com/books/vision-models-for-high-dynamic-range-and-wide-colour-gamut-imaging/bertalmio/978-0-12-813894-6

Bustos, J. F., de la Torre Gómora, M. Á., & Alvarez, S. C. (2018, October). Software engineering process for developing a person re-identification framework. In 2018 7th International Conference On Software Process Improvement (CIMPS) (pp. 69-77). IEEE. https://ieeexplore.ieee.org/abstract/document/8625627/

Cao, L., & Li, H. (2020). Enhancement of blurry retinal image based on non-uniform contrast stretching and intensity transfer. Medical and Biological Engineering and Computing, 58(3), 483-496. https://doi.org/10.1007/s11517-019-02106-7

Cho, Y. J., & Yoon, K. J. (2018). PaMM: Pose-aware multi-shot matching for improving person re-identification. IEEE Transactions on Image Processing, 27(8), 3739-3752. https://doi.org/10.1109/TIP.2018.2815840

Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4690-4699).

Erwin, E., & Ningsih, D. R. (2020). Improving Retinal Image Quality Using the Contrast Stretching, Histogram Equalization and CLAHE Methods with Median Filters. International Journal of Image, Graphics and Signal Processing, 12(2), 30-41. https://doi.org/10.5815/ijigsp.2020.02.04

Hitesh, G., Singh, V. J., Mahajan, S., Das, A., & Deepti, M. (2017). Identification of poor visibility conditions in Urban Settings. 8th International Conference on Computing Communication and Networking Technologies. Delhi. https://doi.org/10.1109/ICCCNT.2017.8204116

Honnutagi, P., & Maranur, J. (2018). Image Enhancement using Histogram Equalisation and DWT based image fusion. Journal of Emerging Technologies and Innovative Research, 5(9), 992-998. https://www.jetir.org/papers/JETIR1809797 Hsu,

Hsu, H. J., & Chen, K. T. (2017, June). DroneFace: an open dataset for drone research. In Proceedings of the 8th ACM on multimedia systems conference (pp. 187-192). https://dl.acm.org/doi/abs/10.1145/3083187.3083214

Hsu, H. W., Wu, T. Y., Wong, W. H., & Lee, C. Y. (2018, April). Correlation-Based Face Detection for Recognizing Faces in Videos. In 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 3101-3105). IEEE. https://ieeexplore.ieee.org/abstract/document/8461485

Hu, W., Huang, Y., Zhang, F., Li, R., Li, W., & Yuan, G. (2018). SeqFace: make full use of sequence information for face recognition. arXiv preprint arXiv:1803.06524. https://arxiv.org/abs/1803.06524

Jeon, H., & Kim, T. (2016). Grey-level context-driven histogram equalisation. IET Image Processing, 10(5), 349-358. https://digital-library.theiet.org/content/journals/10.1049/iet-ipr.2015.0491

Jogin, M., Madhulika, M. S., Divya, G. D., Meghana, R. K., & Apoorva, S. (2018, May). Feature extraction using convolution neural networks (CNN) and deep learning. In 2018 3rd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT) (pp. 2319-2323). IEEE. https://ieeexplore.ieee.org/abstract/document/9012507/

Karanam, S., Gou, M., Wu, Z., Rates-Borras, A., Camps, O., & Radke, R. J. (2019). A Systematic Evaluation and Benchmark for Person ReIdentification: Features, Metrics and Datasets. IEEE Transactions on Pattern Analysis and machine Intelligence, 41(3), 523-536. https://doi.org/10.1109/TPAMI.2018.2807450

Koc, M. (2021). A novel partition selection method for modular face recognition approaches on occlusion problem. Machine Vision and Applications, 32(1), 1-11. https://link.springer.com/article/10.1007/s00138-020-01156-4

Koide, K., Menegatti, E., Carraro, M., Munaro, M., & Miura, J. (2017, September). People tracking and re-identification by face recognition for rgb-d camera networks. In 2017 European Conference on Mobile Robots (ECMR) (pp. 1-7). IEEE. https://doi.org/10.1109/ECMR.2017.8098689

Koo, J. H., Cho, S. W., Baek, N. R., Kim, M. C., & Park, K. R. (2018). CNN-based multimodal human recognition in surveillance environments. Sensors, 18(9), 3040. https://www.mdpi.com/338310

Liao, S., Jain, A. K., & Li, S. Z. (2012). Partial face recognition: Alignment-free approach. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(5), 1193-1205. https://doi.org/10.1109/TPAMI.2012.191

Marutho, D., Handaka, S. H., & Wijaya, E. (2018, September). The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In 2018 International Seminar on Application for Technology of Information and Communication (pp. 533-538). IEEE. https://doi.org/10.1109/ISEMANTIC.2018.8549751

Mathew, V., Chacko, A. M., & Udhayakumar, A. (2018, December). Prediction of suitable human resource for replacement in skilled job positions using Supervised Machine Learning. In 2018 8th International Symposium on Embedded Computing and System Design (ISED) (pp. 37-41). IEEE. https://doi.org/10.1109/ISED.2018.8704120

Mathew, V., Toby, T., Chacko, A., & Udhayakumar, A. (2019, December). Person re-identification through face detection from videos using Deep Learning. In 2019 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS) (pp. 1-5). IEEE. https://ieeexplore.ieee.org/abstract/document/9117938/

Meedeniya, D. A., & Ratnaweera, D. A. A. C. (2007, August). Enhanced face recognition through variation of principle component analysis (PCA). In 2007 International Conference on Industrial and Information Systems (pp. 347-352). IEEE. https://ieeexplore.ieee.org/abstract/document/4579200/

Mokhayeri, F., & Granger, E. (2020). A paired sparse representation model for robust face recognition from a single sample. Pattern Recognition, 100, 107129. https://www.sciencedirect.com/science/article/pii/S0031320319304303

Nasution, A. L., Bayu, D. B. S., & Miura, J. (2014, August). Person identification by face recognition on portable device for teaching-aid system: Preliminary report. In 2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA) (pp. 171-176). IEEE. https://ieeexplore.ieee.org/abstract/document/7005935/

Parchami, M., Bashbaghi, S., & Granger, E. (2017, August). Cnns with cross-correlation matching for face recognition in video surveillance using a single training sample per person. In 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-6). IEEE. https://ieeexplore.ieee.org/abstract/document/8078554/

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. https://ora.ox.ac.uk/objects/uuid:a5f2e93f-2768-45bb-8508-74747f85cad1

Ranjan, R., Bansal, A., Zheng, J., Xu, H., Gleason, J., Lu, B., ... & Chellappa, R. (2019). A fast and accurate system for face detection, identification and verification. IEEE Transactions on Biometrics, Behavior and Identity Science, 1(2), 82-96. https://ieeexplore.ieee.org/abstract/document/8680708/

Ren, S., Cao, X., Wei, Y., & Sun, J. (2016). Face alignment via regressing local binary features. IEEE Transactions on Image Processing, 25(3), 1233-1245. https://doi.org/10.1109/TIP.2016.2518867

Ruikar, D. D., Santosh, K. C., & Hegadi, R. S. (2018, December). Contrast stretching-based unwanted artifacts removal from CT images. In International Conference on Recent Trends in Image Processing and Pattern Recognition (pp. 3-14). Springer, Singapore. https://doi.org/10.1007/978-981-13-9184-2_1

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 815-823). https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Schroff_FaceNet_A_Unified_2015_CVPR_paper.html

Shao, M., Zhang, Y., & Fu, Y. (2017). Collaborative random faces-guided encoders for pose-invariant face representation learning. IEEE Transactions on Neural Networks and Learning Systems, 29(4), 1019-1032. https://ieeexplore.ieee.org/abstract/document/7839179/

Starovoitov, V. V., Samal, D. I., Briliuk, D. V., & Kopendakov, A. (2003). Image enhancement for face recognition. In International Conference on Iconics.

Tang, Z., Lin, Y. S., Lee, K. H., Hwang, J. N., & Chuang, J. H. (2019). ESTHER: Joint camera self-calibration and automatic radial distortion correction from tracking of walking humans. IEEE Access, 7, 10754-10766. https://ieeexplore.ieee.org/abstract/document/8605504/

Thakur, N., & Han, C. Y. (2021). Framework for an intelligent affect aware smart home environment for elderly people. arXiv preprint arXiv:2106.15599. https://arxiv.org/abs/2106.15599

Tulyakov, S., Jeni, L. A., Cohn, J. F., & Sebe, N. (2017). consistent 3D face alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(9), 2250-2264. https://doi.org/10.1109/TPAMI.2017.2750687

Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001 (Vol. 1, pp. I-I). Ieee. https://doi.org/10.1109/CVPR.2001.990517

Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., ... & Liu, W. (2018a). Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5265-5274). http://openaccess.thecvf.com/content_cvpr_2018/html/Wang_CosFace_Large_Margin_CVPR_2018_paper.html

Wang, K., Wang, H., Liu, M., Xing, X., & Han, T. (2018b). Survey on person re-identification based on deep learning. CAAI Transactions on Intelligence Technology, 3(4), 219-227. https://doi.org/10.1109/ACCESS.2019.2957336

Wolf, L., Hassner, T., & Maoz, I. (2011, June). Face recognition in unconstrained videos with matched background similarity. In CVPR 2011 (pp. 529-534). IEEE. https://ieeexplore.ieee.org/abstract/document/5995566

Wong, Y., Chen, S., Mau, S., Sanderson, C., & Lovell, B. C. (2011, June). Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In CVPR 2011 WORKSHOPS (pp. 74-81). IEEE. https://doi.org/10.1109/CVPRW.2011.5981881

Yang, S., Luo, P., Loy, C. C., & Tang, X. (2016). Wider face: A face detection benchmark. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5525-5533). https://openaccess.thecvf.com/content_cvpr_2016/html/Yang_WIDER_FACE_A_CVPR_2016_paper.html

Zhang, G., Qin, H., Ke, Y., Chen, J., & Gong, Y. (2020). Phased Groupwise Face Alignment. IEEE Access, 8, 62415-62422. https://doi.org/10.1109/ACCESS.2020.2983722

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10), 1499-1503. https://doi.org/10.1109/LSP.2016.2603342

Zhao, R., Oyang, W., & Wang, X. (2016). Person re-identification by saliency learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(2), 356-370. https://ieeexplore.ieee.org/abstract/document/7437489/

Zhuo, J., Chen, Z., Lai, J., & Wang, G. (2018, July). Occluded person re-identification. In 2018 IEEE International Conference on Multimedia and Expo (ICME) (pp. 1-6). IEEE. https://doi.org/10.1109/ICME.2018.8486568