Applied Computing Journal https://doi.org/10.52098/acj.202247
ISSN: 2788-9688 Volume 2, Issue 2, pp 193-212, April 2022
Received: 19/10/2022 Revised: 04/02/2022 Accepted: 16/04/2022

Face Recognition based on Convoluted Neural Networks: Technical Review

Basil Ismail Mirghani[1,*] and Sara Ali K.M AL-Mazruii[2]
[1,2] Sohar University, FCIT, 311 Sohar, Oman
* Corresponding author: Basil Ismail Mirghani[1], 191148@students.su.edu.om

Abstract

Human beings recognize and classify objects with biological senses and brain that processes the input into meaningful information. Other than that humans have come to recognize each other in multiple ways one of which is visual recognition of faces. As a biological trait human faces are certainly a biometric such they are universal, distinctive, mostly permanent and collectable. With that a computerized face recognition system can constructed relying on visual information present on each face uniquely. Generally a face recognition system consists of two main phases, face detection phase where presence of a human face is verified on visual input and face recognition phase where detected face is processed for identification. One of the most sought after methods in field image processing for face recognition is CNN (Convoluted Neural Networks). CNNs have proved its effectiveness and accuracy in many CNN based face detection and face recognition systems. As such in this paper the architecture of CNN is presented. Then different techniques for face detection and face recognition based on CNNs are reviewed. In reviewed papers CNNs have repeatedly demonstrated effectiveness and accuracy on multiple benchmarks for face recognition application.

Keywords: biometrics authentication; face recognition; face detection; Convoluted Neural Networks

1. Introduction

For human beings, recognizing and classifying objects (animated or not) is done by capturing the object through multiple available biological senses and then the information is passed to the brain that recognizes (or learn of) the object and classifies it instantly based on traits captured from that object. Furthermore, objects' traits could also be measured using measurement tools which provide distinctive data that can be translated into information to be used to describe or uniquely identify that object (Alblushi A., 2021; Hassin & Abbood, 2021). With that certain biological traits could be measured and used to uniquely identify an individual among human beings. Such biological traits are known as biometrics. According to (Jain et al., 2004) in order for a biological trait to be eligible as biometric it must be universal (common among humans), distinctive (measured uniquely between different humans to sufficient extend), permanent (largely unchangeable over time) and collectable (measurable quantitatively). One of the biological traits that are eligible as biometric is human face. Human faces satisfy all the requirements of biometric; they are certainly universal, highly distinctive in large scale, largely permanent over long periods of time and collectable. As such it's possible to construct a biometric system based on human face biometric.

A computerized biometric system based on human faces is essentially a face recognition system that relies on visual information present in each face uniquely. Image enhancement is the process of altering a digital image to be more appropriate for identifying and classifying the correct objects (Al-Hatmi & Yousif, 2017; Hasson et al., 2011)). According to (Li et al., 2020) face recognition is a visual pattern recognition problem where visual inputs presented as matrixes in computer needs to be distinguished in terms of whether data contains a face then identify who the face belongs to. (Oloyede et al., 2020) explains that a face recognition system structure is similar in essence to structure of biometric system it involves face detection, face image preprocessing, facial feature extraction and feature classification which is a common step in biometric systems as stated by (Oloyede & Hancke, 2016). (Oloyede et al., 2020) further explains the stages involved in face recognition system:

- Face detection is verification of presence of human face in visual input data.
- Face image preprocessing is preparing the image so that it contains important facial visual data only. Approaches include normalization (face images are transformed to same scale), face alignment (defined by (Jin & Tan, 2017) as locating fiducial points on face image) and enhancement of image (stated by (Karamizadeh et al., 2016) as processing the face image into an enhanced version which has the potential to enhance face recognition system performance).

- Facial feature extraction is extraction of most relevant facial visual data that identify face uniquely - while minimizing noise and unrelated information - into sufficient description vector.
- Feature classification is recognition stage of facial images where facial images are compared for verification or identification of facial images from database. As mentioned by (Oloyede & Hancke, 2016) this is a common stage in biometric systems and it involves verification and identification. Verification is achieved through a one-to-one search between an input and a target as for identification is one-to-many search between input and entire database of targets (Coventry et al., 2003) (Ganorkar & Ghatol, 2007) (P Tripathi, 2011) (Muhtahir et al., 2013) (Ahmad et al., 2012).

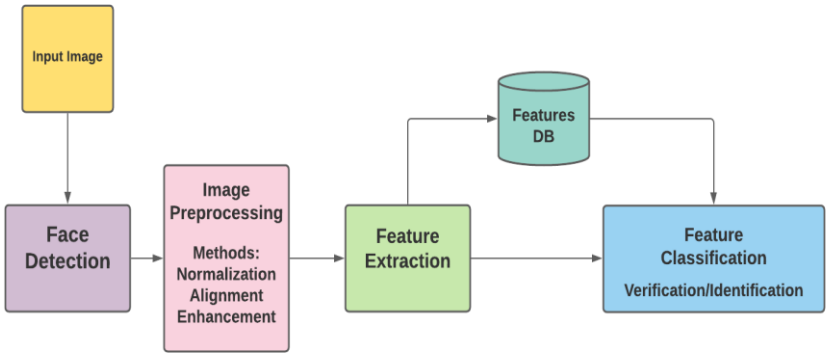

Figure 1: General Steps Involved in Facial Recognition System

Facial recognition systems are deployed in wide range of applications. Some of the applications include control of attendance access (S. Manjula & S. Santhosh Baboo, 2012), security (Lander et al., 2018), finance, education, smartphones, retail, transportation and network information security (Hu et al., 2010).

2. Problem Statement:

As mentioned face recognition systems are deployed in various applications, making it a critically needed computer vision technology that attracted interest for further development and enhancement. There are multiple techniques used in main subsystems (face detection and feature classification) involved in overall structure of face recognition system. All of subsystems collectively have techniques that use deep learning (DL) convolutional neural networks (CNN) method to fulfill their purposes. As such the purpose of this study is to introduce the CNN method and present some of the CNN based techniques for face recognition subsystems.

3. Convolutional Neural Networks:

Neural networks are powerful mathematical models that aim to mimic the human brain in solving complex problems in multidimensional space and convert them to a lower dimension (Yousif J., 2015; Yousif & Kazem, 2021; Alattar et al., 2019). Convolutional neural networks are type of artificial neural networks (Lecun et al., 1998) that are specifically applied in applications that involve processing of visual information. Some of CNN applications include face recognition (Taigman et al., 2014), detection of objects (Ren et al., 2017), image segmentation and classification (Farabet et al., 2013). Visual data in images is typically contained in form of an array or multiple of which. CNNs translate visual data into meaningful visual information using sequential layers of convolution filters to detect edges, detect portion of objects and finally detect the whole object shape (LeCun et al., 2015). Convolution filters are classified in terms of their function in CNN to convolution layer filters, pooling layer filters and fully connected layers filters (Bezdan & Bačanin Džakula, 2019).

3.1. Layers of CNN:

As mentioned mainly CNN consists of three layers which are convolution layer, pooling layer and fully connected layer (Bezdan & Bačanin Džakula, 2019). Ultimately processing visual data through CNN layers is done by extracting feature maps from input 2D image using kernels (filters) (Salomon et al., 2017).

3.2. Convolution Layer:

Convolution layer as its name implies relies on convolution operation between image pixels and set of learning kernels. Kernels typically have small size of $n \times n$ and depth d equal to input image channels, if image is grayscale $d = 1$ and $d = 3$ if image is RGB color and so on. As input visual data is passed to convolution layer, frame pixels at defined positions are convoluted with kernel filter yielding a convoluted frame; and this process is repeated for each kernel (Bezdan & Bačanin Džakula, 2019). Convoluted frames are then processed by activation function to generate feature maps. Some of activation functions include sigmoid logistic function, hyperbolic tangent Gaussian function and Rectified Linear unit (ReLU). Similar to activation functions in neural networks (NN) a bias value can be introduced to shift activation function input for generation of feature maps, therefore for feature map A, $A = f(Conv.frame + bias)$ (Salomon et al., 2017).

According to (Bezdan & Bačanin Džakula, 2019) size of generated feature maps depend on three convolution-related parameters which are stride, depth and padding. Stride is position shift parameter that defines next position of frame pixels to be convoluted with kernel i.e., for pixel at position n the next pixel to be convoluted is at position

$n + s$ where s is stride value. Depth refers to number of unique kernel filters applied to input frame. Padding is adding zeros to boards of input image such that required pixels are convoluted and information is preserved. With that output feature map size can be calculated as $(n + 2p - f)/s + 1$ where n is filters number, p is padding layers number, f is kernel size and s is stride.

3.3. Pooling Layer:

Features maps are processed in pooling layer for reduction of maps' dimensions by down-sampling them (Bezdan & Bačanin Džakula, 2019) and reducing variance among feature maps pixels (Salomon et al., 2017). In down-sampling process feature maps are divided into smaller regions of equal dimensions $d \times d$ then in each region either the average or maximum of pixels values is taken as representative of the region (Salomon et al., 2017). Pooling process also depends on stride and size of pooling region. Overlapping between to-be-pooled regions can be controlled using stride value and to prevent occurrence of any overlapping between regions stride value can be set as d where d is feature map dimension (Salomon et al., 2017).

3.4. Fully Connected Layer:

Fully connected layer is last layer of CNN. Here processed features maps are converted into vectors that are fed to artificial neural network neurons as input (Bezdan & Bačanin Džakula, 2019) for classification. Deep learning methods can discover many complex relations between training data and outputs due to non-linearity of its intermediate hidden layers. However in case of limited training data DL network may formulate relationships that might be valid in context of training data only and not on real testing data. This is known as overfitting (Srivastava et al., 2014). One of techniques that can be applied to prevent overfitting on CNN is dropout method proposed by (Srivastava et al., 2014). In dropout method neural network nodes are dropped randomly from network temporally along with its incoming and outgoing connections.

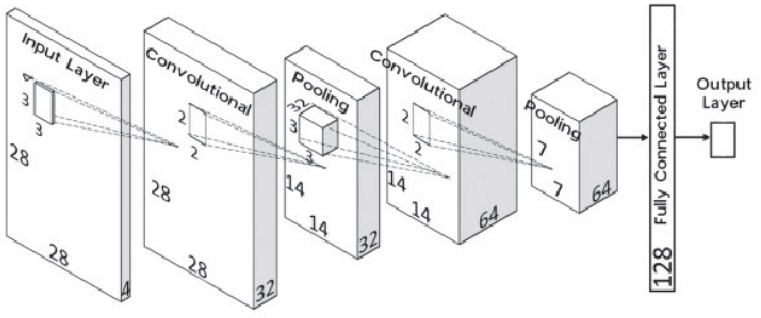

Figure 2: Example of CNN architecture (Ignjatić et al., 2018)

4. Face Recognition Subsystems Methods:

4.1. Face Detection Methods

(Triantafyllidou & Tefas, 2016) proposed a light model for face detection based on CNN using 113,864 parameters only. Despite lesser complexity of model its result showed that it can be deployed for real world applications using standard processing power. The model consists of two CNNs that were combined in a single architecture. The first CNN was trained to detect major facial features such as mouth, eyes, nose and so on. The second CNN was trained for full detection of face. Face detection CNN contains seven convolution layers and used images of dimensions $32 \times 32 \times 3$ for training. Face parts CNN contains three convolution layers and used $16 \times 16 \times 3$ images for training. First CNN was evaluated in terms of successful full face detection whereas the second CNN was evaluated in terms of detection of relevant face parts successfully. CNNs were combined by parallel processing of first three layers of first CNN and second CNN fully then results were stacked as inputs for convolution four to seven on first CNN. The performance of CNN was tested on FDDB (Face Detection Data Set and Benchmark) dataset. The detector achieved a recall rate of 88.9% outperforming most of recent face detection methods.

(Farfade et al., 2015) presented a method based on CNN named Deep Dense Face Detector (DDFD) for multiple faces detection in various poses. DDFD model has lesser complexity as it doesn't require bounding-box regression (for reduction of localization errors (Girshick et al., 2014)), semantic segmentation, or support vector machines classifiers. DDFD was constructed based on principle of maximizing CNN capacity for classification and feature extraction for detecting faces from various orientations while simplifying its architecture to reduce computational complexity. DDFD consists of five convolutional layers followed by three fully connected ones. Fully connected layers are converted into convolutional layers by reshaping parameters on layers (Felzenszwalb et al., 2010) which allowed CNN to process images of any dimensions effectively and generate heat map. From heat maps regions of highest probability of containing face are detected and then processed with non maximal suppression to localize faces accurately. DDFD was tested on three libraries PASCAL, AFW and FDDB datasets and non-maximal suppression model was implemented on maximum and average. Firstly implementing DDFD based on average non-maximal suppression (NMS-avg) had higher average precision than maximum non-maximal suppression (NMS-max). From which DDFD NMS-avg was tested and compared with different detectors using mentioned datasets.

DDFD had average precision of 91.79% on PASCAL dataset scoring highest among face detectors and average precision of 96.26% on AFW dataset coming in third. As for FDDB dataset DDFD had recall rate of 84%.

(H. Li et al., 2015) builds a cascade CNN face detector that rejects false detections during early stages where input is processed in low resolution and verifies truthiness of detection at high resolution stages. Face detector was designed to feature fast detection of faces, accelerated cascade CNN, localization with high quality and multi-resolution architecture for detection verification. Cascade is composed of six CNNs three of which are for face classification and others are calibrating bounding boxes for faces. CNNs are based on Alexnet architecture and use ReLU activation function. As input image is passed to CNN cascade detector 12-net CNN scans image on different scales and reject more than 90% of detected windows. 12-calibration-net CNN processes remaining windows as 12×12 images adjusting their location and size to approach potential face. NMS is then applied for elimination of highly overlapped detection windows, and then remaining windows are resized into 24×24 images. Generated images becomes input for 24-net CNN and subsequently to 24-calibration-net CNN and processes that occurred on first two CNNs are repeated outputting 48×48 windows images. 48-net CNN receives new windows and evaluates detection and NMS eliminates overlapped windows. Lastly 48-calibration-net CNN calibrates bounding boxes for detected output faces. Cascade CNN detector was tested on AFW and FDDB datasets. On AFW the detector achieved average precision of 96.72% and had recall rate of 85.1% on FDDB dataset.

(Yang et al., 2018) created face detector utilizing capabilities of supervised CNN by capturing facial features based on common attributes of face rather than standard bounding box. Authors show that this approach has more robustness in detecting faces under server oscillations or pose variations. Face detector was based on three principles. The first principle is uniqueness of human face parts structure where CNN can be trained to detect and classify different face parts without explicit supervision. The second principle is evaluation of detect parts based on their spatial arrangements on faces through a score to find likelihood of detection actually being a face or not. The third principle is refining output of bounding boxes detection of potential faces by CNN that recognizes true faces and estimates face locations more precisely. Based on those principles face detector named Faceness-Net was constructed and it consists of two stages; the first stage is detection of facial parts to generate face proposals that are ranked according to faceness score and second stage is enhancement of face proposals for detecting faces. On first stage attribute-aware CNNs are used to generate facial parts maps from inputs images. Those maps show locations of hair, eyes, nose, mouth and bread face components then maps are combined on face label map. Generated face

proposals are ranked based on their faceness scores which are determined from face parts maps own faceness scores that are determined from spatial configuration of detected face part. NMS is applied on face parts to reduce number of detected windows then average faceness score of parts is taken as faceness score of face proposal. Another NMS is applied to reduce face proposals based on faceness score to eliminate false positive detections. On second stage CNNs for optimizing face classification and bounding boxes regression are used to enhance generated face proposals. Authors have implemented three more variations of Faceness-Net which are Faceness-Net-SR, Faceness-Net-TP and Faceness-Net-SR-TP. SR means that variant uses single attribute-aware CNN not five and TP means varies uses template technique was used to generate candidate windows not external generic object. Faceness-Net and its variants were tested on AFW, PASCAL and FDDB datasets. On AFM Faceness-Net-SR-TP, Faceness-Net-SR, Faceness-Net-TP and Faceness-Net had average precision of 98.05%, 97.38%, 97.25% and 97.2% respectively. On PASCAL dataset average precisions were 92.11% for Faceness-Net, 91.79% for Faceness-Net-SR-TP, 91.65% for Faceness-Net-SR and 91.23% for Faceness-Net-TP. As for FDDB recall rates were 92.84% for Faceness-Net-SR-TP, 91.72% for Faceness-Net-TP, 91.31% for Faceness-Net-SR and 90.98% for Faceness-Net.

(Qin et al., 2016) made modifications on cascade CNNs approach to obtain better performance from network by jointly training CNNs. Authors showed that back propagation algorithm can be used in training cascaded CNN and joint training approach can be implemented on more complex cascade CNNs architectures. On joint training architecture named FaceCraft image of size 48×48 is input for three branch networks x12, x24 and x48 and image is resized according to branch name. Activation function ReLU is used on non-linear layers and dropout is implemented before regression or classification layer. Output of network is one joint loss of three branches and its optimized using back propagation. Joint network also use control threshold layers to determine how loss is contributed from proposals coming from up branches to down branches. FaceCraft was tested on AFW and FDDB datasets. FaceCarft scored an average precision of 98% on AFW dataset and had recall rate of 88.2% on FDDB dataset.

(Garg et al., 2018) proposed a face detection system based on YOLO-Face CNN detector. YOLO (You Only Look Once) is deep learning CNN approach (Redmon et al., 2016) that demonstrated its elevated face detection performance in standard datasets such as PASCAL VOC and COCO (Garg et al., 2018). Authors list features of YOLO as being comparatively faster in face detection in real time, maintains accurate detection performance regardless of input image size and capable of extracting features from arbitrary image sizes. Architecture of

proposed model takes color images of size 448×448 as input and it consists of seven convolution layers for features extraction each is followed by pooling layer that performs max pooling using 2×2 down-sampling kernels. Following that are three fully connected layers and output layer where NMS (Non-Maximal Suppression) is used for classifying detection according to extracted features and bounding box position. FDDB was used for training and testing model where 70% of selected samples were used for training and remaining for testing.

Table 1: Review of CNN based face detectors

Author	Year	Methodology Highlight	Results summary
(Triantafyllidou & Tefas, 2016)	2016	Light CNN model using 113,864 parameters only. Consists of two CNNs for detecting major face parts (mouth, nose, etc.) and overall face detection	Achieved a recall rate of 88.9% on FDDB
(Farfade et al., 2015)	2015	DDFD for multiple faces detection in various poses. Model has lesser complexity as it doesn't require bounding-box regression, semantic segmentation, or support vector machines classifiers.	DDFD had average precision of 91.79% on PASCAL, 96.26% on AFW and recall rate of 84% on FDDB.
(H. Li et al., 2015)	2015	Cascade CNN face detector that rejects false detections during early stages and verifies detection at later stages.	Average precision was 96.72% on AFW and recall rate of 85.1% on FDDB.
(Yang et al., 2018)	2017	Faceness-Net. Supervised CNN that captures facial features based on common attributes of face.	On AFM, Faceness-Net-SR-TP had average precision of 98.05%. On PASCAL dataset average precisions was 92.11% for Faceness-Net. As for FDDB recall rates were 92.84% for Faceness-Net-SR-TP.
(Qin et al., 2016)	2016	FaceCraft. Modifications on cascade CNNs approach to obtain better performance from network by jointly training CNNs.	FaceCarft scored an average precision of 98% on AFW dataset and had recall rate of 88.2% on FDDB dataset.
(Garg et al., 2018)	2018	Face detection system based on YOLO-Face CNN detector	92.2% accuracy on FDDB.

On training phase gradient decent optimizer algorithm was used, model was run for 25 epochs and different learning rates values were tested. It was found that accuracy remained constant after 20 epochs and optimal learning rate was 0.0001. Running model on testing set resulted in achieving 92.2% accuracy which is higher than accuracies achieved by other face detection algorithms tested by authors which are 89.6% on R-CNN and 83.8% on Haar Cascade.

4.2. Face Recognition Methods:

(Liu et al., 2021) proposed a lightweight CNNs architecture for face recognition. The reasoning behind this architecture is despite current CNN based face recognition systems being highly accurate they are complex and require extensive computation resources which make them unsuitable for computationally limited devices (e.g. mobile devices). Also there have been previous attempts to build lightweight CNN face recognition system however despite systems showed efficiency their results were not accurate enough. As such the authors build compressed face recognition CNN model while maintaining accuracy for computationally limited devices. Improvements were made in design of network structure, training methodology and loss function. In terms of network design structure, three structures based on channel attention mechanism are proposed which are depthwise squeeze and excitation model, depthwise separable squeeze and excitation model and linear squeeze and excitation model. Squeeze and excitation approach reduces computational costs for processing feature maps and improves CNNs based architecture performance (J. Hu et al., 2018). Those structures were applied on light CNN with small set of parameters and tested on datasets. In terms of training methodology authors implement teacher-student training method that is based on additive angular margin loss function (loss function for distinguishing faces (Deng et al., 2019)) and knowledge distillation for transferring knowledge between CNNs. Deep CNN that is superior in feature extraction and fitting capabilities called teacher is used to guide and train a light CNN called student. Using knowledge distillation superior performance and capabilities of teacher can be transferred to student. With that lightweight CNN model can be improved while maintaining model compression. Different models were constructed with mentioned SE (Squeeze and Excitation) structures and teacher-student training method. Models were trained and tested on several datasets and achieved highest accuracy of 99.67% using a combined model of depthwise SE and linear SE structure on LFW dataset with 5.36 MB storage space and 1.35 million parameters.

(Nimbarte & Bhoyar, 2018) presents age invariant face recognition model based on CNNs. The main goal is for network to recognize matching face for input from gallery of face images despite the changes occurring in face features due to age difference. AIFR (Age Invariant Face Recognition)-CNN architecture has seven layers and it accepts images of size 32×32 to reduce computational costs. Architecture of AIFR-CNN consists of three stages: image preprocessing, feature extraction and classification. On image preprocessing stage Viola Jones face detection algorithm is applied to crop image into face-focused image, then image is transformed to grayscale and resized to 32×32. As for feature extraction stage, here image is passed to AIFR-CNN seven layers. Layers are arranged as

two convolution layers followed by pooling layers for each then a convolution layer followed by two fully connected layers. Kernels used on architecture are size 5×5 and pooling filters are size 2×2. On last stage of classification output of last fully connected layer is passed to SVM classifier for face identification. AIFR-CNN was trained and tested on FGNET and MORPH (album-II) datasets. On FGNET 980 images were used, 852 of which for training and remaining 128 for testing. Testing on FGNET resulted on network having a recognition percentage of 76.6%. As for MORPH (album-II) dataset total of 1005 images were used, 750 for training and 255 images for testing. Testing on MORPH (album-II) resulted on network having recognition rate of 92.5%.

(Tang et al., 2020) proposes face recognition system architecture based on local binary pattern (LBP) and parallel ensemble learning of CNNs. The reasoning behind this architecture is to address issues that degrades face recognition systems success rate such as face expression, pose orientation, illumination and occlusion. Those issues raise mainly due to single CNN low generalization abilities. On architecture face features are extracted firstly using LBP on input image. Following that ten CNNs based on five different structures extract features further for training and improvement of parameters (weights and biases) values. Those CNNs also obtain classification for input after fully connected layer using Softmax function. To obtain final face recognition result parallel ensemble learning is used to get the result with majority voting. Method was tested on ORL and Yale-B face datasets and achieved recognition rates of 100% and 97.51% respectively. Experiments on model showed its tolerance to mentioned face recognition issues in addition to elevation of face recognition accuracy and generalization performances. More to that a detection hybrid model consisting of proposed face recognition model and pedestrian detection model was tested for improvement of detection rate. It achieved 11.2% increase in detection rate performance.

In a study conducted by (Khalajzadeh et al., 2013) a hybrid face recognition system consisting of CNN and LRC (Logistic Regression Classifier) was presented. CNN component was trained for detection and recognition of face images. Features extracted by CNN are then passed to LRC component for classification of output. CNN structure consists of two convolution layers each followed by pooling layer then a fully connected layer. Images of size 64×64 are passed to convolution layer where $7 \times 7 \times 6$ kernel is applied resulting in six 58×58 feature maps. Following pooling layer applies $2 \times 2 \times 6$ sub-sampling kernel resulting in six 29×29 feature maps. The second convolution layer applies $8 \times 8 \times 16$ kernel generating sixteen 22×22 feature maps that are passed to pooling layer where $2 \times 2 \times 16$ sub-sampling kernel is applied, down-sampling sixteen feature maps to 11×11. On fully connected layer feature maps are downsized to fifteen 1×1 using 11×11 kernels. For CNN training, five hundred

epochs were applied due to complexity and time for computation constrains. Learning rate for CNN was set dynamically decreasing as a function of number of epochs. To address issues of illumination and varying pose orientations of face images for recognition input images were normalized using pixels mean and standard deviation. Other techniques applied to CNN are back propagation algorithm and dynamic update of weights during feature presentation (to keep number of parameters within data range) rather than after passing training set (Batch update). For evaluation of network performance Yale dataset was used for training and testing of CNN structure and several classifiers were applied for final recognition. Out of tested classifiers the model had highest accuracy and least time when using SimpleLogistic classifier on Yale dataset with 86.06% accuracy and 1.22 seconds recognition time.

(Ramaiah et al., 2015) presented a facial recognition system based on CNN that contributes to tackling face recognition systems performance degrading issue of illumination variations in input face images. Authors take advantage of feature extraction capabilities of CNN for processing correct recognition of face images and further enhance CNN performance by considering symmetrical face information present in horizontal reflection of facial image. Architecture of CNN consists of five layers, two convolution layers followed by pooling layer for each and finally a fully connected layer. Input face images are rescaled to size of 28×32 and passed as input CNN. On first convolution layer kernel of size $5 \times 5 \times 6$ is applied on input face image generating six 24×28 feature maps that are down-sampled on pooling layer using $2 \times 2 \times 6$ down-sampling kernel to size 12×14. Then on second convolution layer kernel of size $5 \times 5 \times 12$ is convoluted with feature maps generating new twelve 8×10 sized feature maps. New feature maps are down-sampled to size 4×5 on last pooling layer using $2 \times 2 \times 12$ down-sampling feature maps. Generated feature maps are converted and combined into 240×1 column vector using row major order. Column vector is input to fully connected layer where classifications of facial image to one of thirty output classes occur. CNN classifier was trained using back-propagation algorithm with batch mode. Experiments on CNN were conducted using extended Yale Face Database B. From dataset thirty subjects (classes) were selected and for each subject sixty two face images with different illuminations were taken. Face images were then organized into five different sets according to lighting degree. Training CNN was implemented using back-propagation method, batch size 2 and 500 epochs. Five-fold cross validation was used for training and testing. Running five sets on constructed CNN face recognition system resulted in average accuracy of 89.05%. To boost CNN performance, images on sets were enhanced by adding horizontal reflection to face images which provide classifier with

additional information relating to shadows on face image. This enhancement resulted in increasing CNN average accuracy of classification inputs on five sets to 94.01%.

(Nakada et al., 2017) constructed an active face recognition system named AcFR. AcFR is a viewpoint-dependent system means that the system outputs certain behavior depending on recognition result similar to human behavior when attempting to recognize face of another individual. AcFR implements its proposed tasks through two components. The first component is a face recognition model consisting of VGG-Face CNN coupled with nearest-neighbor identity recognition criterion. First component evaluates recognition (identifies subject) and provides information required for second component which is a control model to take decisions. Decisions made by control model determine output behavior of AcFR which belong to set $\{greet(x),\ \ ChangeView(x),\ \ ignore(x)\}$ where x is individual extracted information. For face recognition component on AcFR it follows conventional architecture of face recognition system steps. On first step preprocessing (detection and alignment) authors follow (Mathias et al., 2014) face detection algorithm. On feature extraction step VGG-Face CNN was implemented which has sixteen layers and was trained with two million images. On classification stage authors experiment with different classifiers such as SVM, Linear and Regression and Nearest Neighbor classifier. The first two had low accuracy below 20% whereas Nearest Neighbor classifier achieved 90% accuracy. Nearest Neighbor classifier uses extracted features from feature maps, stored feature maps and Euclidean distance to compute classification. Euclidean distance is also used in control model to output behavior. As mentioned control model makes decisions according to information provided from first model. Control model is given two initial threshold distances (t_1 and t_2) that are compared with euclidean distance (d) to output certain behavior. If distance is lesser than or equal to first threshold value output is greet, if it's higher than or equal to second threshold distance output is ignore and if it falls in between view is changed. When changing view features are extracted for same subject however input image is taken from different orientation. Experiments on AcFR was conducted on PIE dataset and for each individual nine different pictures from nine different view angles in range of -90 and 90 degrees were used. On face recognition component views closer to frontal views (0 degrees) had highest accuracy (can reach to 100%) and least Euclidean distance which showed the robustness of VGG-Face CNN and AcFR being view dependent. The hypothesis of AcFR being view dependent was tested further by changing feature vectors in gallery from frontal view to -45 degrees. Similarly highest recognition accuracy and least Euclidean distance were achieved for views nearing -45 degrees and 45 degrees as well due to symmetrical nature of human face. To test AcFR behavior when subject is a stranger, authors removed ten subjects'

features from gallery and reevaluated system response. AcFR computed Euclidean distance in range of 286 to 350 similar to views at extremes (-90 and 90 degrees) which showed AcFR ability to distinguish strangers from recognized individuals. As for control system component AcFR behavior was dependent on input image characteristics (illumination, expression and mainly view) and computed Euclidean distance. To minimize impact of change in image characteristics illumination was set as constant on different subject's images. Results showed that when setting higher first threshold (t_1) AcFR would greet more often and when setting lower second threshold (t_2) AcFR would ignore more often. As such first and second thresholds were set to 250 and 325 respectively.

(Schroff et al., 2015) present a face recognition system that overcomes scaling and efficiently requirements in such systems. System named FaceNet is based on principle of calculating Euclidean space from face images. From distances in Euclidean space a face similarity measure can be computed. Euclidean spaces are features vectors generated from FaceNet as such; FaceNet can be combined with other techniques to implement face recognition, verification or clustering system as well. FaceNet uses a trained deep CNN that directly optimizes how features are extracted rather from classical bottleneck approach used in other CNN based features extractors. CNN was trained using multiple three-similar-sets of approximately aligned matching and non-matching face patches. Those sets were mined using an online triplets mining tool. This training approach resulted in achieving high performance with much greater efficiency using 128 bytes face images. FaceNet was tested on LFW and YouTube Faces DB. On LFW FaceNet achieved 99.63% accuracy and 95.12% accuracy on YouTube Faces DB. FaceNet also highly reduces error rate by 30% in comparison to results achieved by (Sun et al., 2015) on same datasets.

(Sanchez-Moreno et al., 2021) presents a face recognition system mainly composed of FaceNet (FaceNet implements features extraction using deep CNNs (William et al., 2019)) and known classifiers such as SVM, K Nearest Neighbor and Random Forest. The reasoning behind building the system is address the need for having low cost and efficient face recognition system that can operate in unconstrained environment. As face recognition systems involve two main stages face detection and face recognition, authors implement real-time high speed face detector YOLO-Face (one of most popular CNN face detectors in recent years (Garg et al., 2018)) based on YOLOv3. On face recognition stage FaceNet along with supervised classifiers are used as mentioned previously. Experiments on model were carried for face detector and face recognition stages. On face detection stage YOLO-Face based detector was able to reach 89.6% accuracy on Honda/USCD dataset which is mainly composed of images taken in unconstrained environment. It's worth noting that experiments carried on YOLO-Face for face

detection showed that the model can detect small faces and had better performance when detecting partly blocked or differently pose oriented faces. Tests on face recognition stage were conducted using LFW dataset on FaceNet and different combinations of FaceNet and other classifiers. Highest accuracy was 99.7% achieved using combination of FaceNet + SVM. Accuracies of 99.5%, 85.1% and 99.6% were achieved using FaceNet + K Nearest Neighbor, FaceNet + Random Forest and FaceNet respectively. Overall face recognition system (composed of two stages) had 99.1% recognition rate and runtime of 49 milliseconds.

(Khan et al., 2019) presents a face recognition system framework for smart glasses using CNN. The method adds flexibility and portability attributes to such system and good capturing capabilities on frontal view. The overall face recognition system is presented in two stages face detection stage and face recognition stage. Face detection uses Haar classifier which is composed of series of weak classifiers that form one strong classifier. Here face detector was able to achieve 98% accuracy using 3099 features samples. On face recognition stage authors used AlexNet CNN that includes five convolution layers, three fully connected layer and ReLU as activation function. Transfer learning ability of AlexNet was used for facial recognition on smart glasses. The system was able to reach 98.5% accuracy after training it with 2500 various images per class.

Table 2: Review of CNN based face recognition systems

Author	Year	Methodology Highlight	Results summary
(Liu et al., 2021)	2021	Compressed face recognition CNN model that maintains accuracy for computationally limited devices.	Different models were constructed with SE structures and teacher-student training method. Highest accuracy was 99.67% using a combined model of depth wise SE and linear SE structure on LFW dataset with 5.36 MB storage space and 1.35 million parameters.
(Nimbarte & Bhoyar, 2018)	2018	Age invariant face recognition model based on CNNs	On FGNET recognition percentage was 76.6%. On MORPH (album-II) recognition rate was 92.5%.
(Tang et al., 2020)	2020	Face recognition system architecture based on LBP and parallel ensemble learning of CNNs	Method was tested on ORL and Yale-B face datasets and achieved recognition rates of 100% and 97.51% respectively.
(Khalajzadeh et al., 2013)	2013	Hybrid faces recognition system consisting of CNN and LRC.	Model had 86.06% accuracy on Yale dataset and 1.22 seconds recognition time.
(Ramaiah et al., 2015)	2015	Facial recognition system based on CNN and model is enhanced by considering symmetrical face	On Yale Face Database B base model had average accuracy of 89.05% and increased to 94.01% after enhancing

		information present in horizontal reflection of facial image.	model.
(Nakada et al., 2017)	2017	AcFR. A viewpoint-dependent system that uses VGG-Face CNN on recognition stage and control model to output behavior.	AcFR with VGG-Face CNN achieves high recognition accuracies (can reach 100%) for views nearest to frontal views and less Euclidean distance. Control model behavior is dependent on computed Euclidean distance for input against stored feature maps.
(Schroff et al., 2015)	2015	FaceNet. Model is based on principle of calculating Euclidean space from face images. FaceNet uses a trained deep CNN that directly optimizes how features are extracted.	On LFW FaceNet achieved 99.63% accuracy and 95.12% accuracy on YouTube Faces DB
(Sanchez-Moreno et al., 2021)	2021	Face recognition system mainly composed of FaceNet and classifiers such as SVM, K Nearest Neighbor and Random Forest. YOLO-Face based on YOLOv3 was used on detection stage.	YOLO-Face reached 89.6% accuracy on Honda/USCD dataset. Highest accuracy was 99.7% achieved using FaceNet + SVM. Overall face recognition system had 99.1% recognition rate and 49 ms runtime.
(Khan et al., 2019)	2019	Face detection uses Haar classifier and face recognition stage uses AlexNet	The system was able to reach 98.5% accuracy after training it with 2500 various images per class.

5. Discussion

In this paper a total of fifteen papers were reviewed on implementation of CNN on face recognition applications. Six of reviewed papers focused on various implementations of CNN on face detection whereas the rest focused on face classification/recognition aspect. The overall trends on papers were a focus on improving accuracy on various databases or compression of CNN required resources to run on computationally limited devices. Figures 3 and 4 show highest accuracies achieved for face detection and face recognition systems.

Figure 3 showed that significant improvements had been made on CNN detectors over the years. Highest recall rates were achieved on later years which show the ongoing improvement process on CNN face detectors. The same trend can be observed on CNN face recognition subsystems. Despite the differences on testing datasets face recognition CNNs have had higher accuracies with passing of years, reaching near to or 100% accuracies on conducted tests.

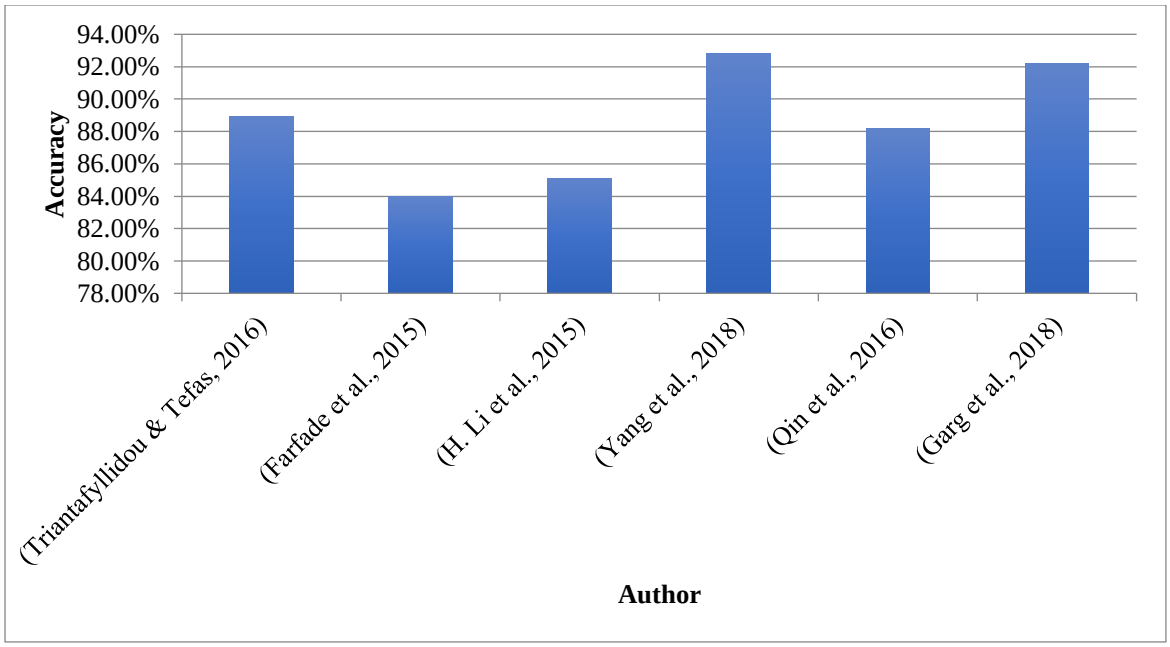

Figure 3: Comparison between CNN face detector recall rates on FDDB dataset

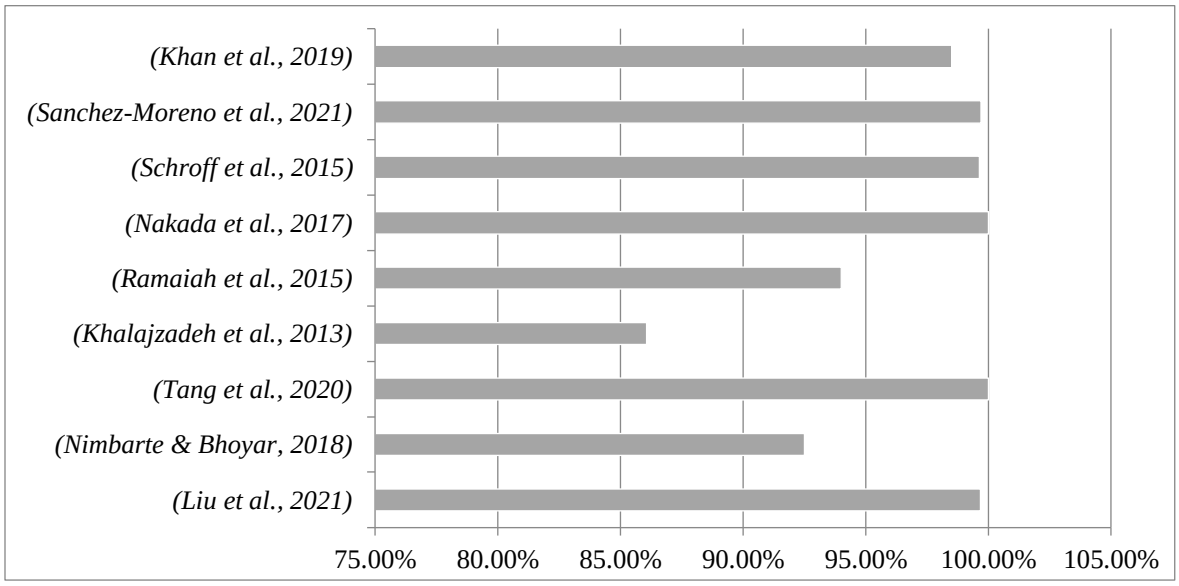

Figure 4: Comparison between CNN face recognition highest accuracies on various dataset

6. Conclusion

In conclusion, face recognition systems are of great importance as they are deployed on various applications including attendance control, security, finance, education, smartphones, retail, transportation and network information security. Overall face recognition system consists of two main stages face detection and face recognition subsystems. Both of those systems have methods that utilize CNNs on carrying their respective

purposes. CNN are a special type of ANN for processing on visual data. CNNs use convolutional layers and pooling layers for extraction of required features for fully connected layer which is a neural network classifier. On face detection CNN feature extractions focuses on extracting features that are unique for a human being then classifier decides on result being a face or not. On face recognition CNN features extraction focuses on extracting features that are unique to a person then classifier decides identity of result. Deployment on CNNs for face detection and face recognition showed continuous improvements over the years, reaching higher than 90% and near 100% on some cases respectively.

Acknowledgment

The research leading to these results has received No Research Project Grant Funding.

References

[1]. Ahmad, S. M. S., Ali, B. M., & Adnan, W. A. W. (2012). Technical Issues and Challenges of Biometric Applications as Access Control Tools of Information Security. International Journal of Innovative Computing, Information and Control, 8(11), 7983–7999.

[2]. Alattar, N. E. B. R. A. S., Yousif, J. A. B. A. R., Jaffer, M. O. O. S. A., & Aljunid, S. A. (2019). Neural and Mathematical Predicting Models for Particulate Matter Impact on Human Health in Oman. WSEAS Trans Env & Dev, 15, 578-585.

[3]. Alblushi, A. (2021). Face Recognition Based on Artificial Neural Network: A review. Artificial Intelligence & Robotics Development Journal, 116-131.

[4]. Al-Hatmi, M. O., & Yousif, J. H. (2017). A review of Image Enhancement Systems and a case study of Salt &pepper noise removing. International Journal of Computation and Applied Sciences (IJOCAAS), 2(3), 171-176.

[5]. Bezdan, T., & Bačanin Džakula, N. (2019). Convolutional Neural Network Layers and Architectures. Proceedings of the International Scientific Conference - Sinteza 2019. Published. https://doi.org/10.15308/sinteza-2019-445-451

[6]. Coventry, L., de Angeli, A., & Johnson, G. (2003). Honest it's me! Self service verification. Paper Presented at Workshop on Human-Computer Interaction and Security Systems, Fort Lauderdale, Florida, United States, 1–4. https://www.andrewpatrick.ca/CHI2003/HCISEC/HCISEC-papers.html

[7]. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive Angular Margin Loss for Deep Face Recognition. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Published. https://doi.org/10.1109/cvpr.2019.00482

[8]. Farabet, C., Couprie, C., Najman, L., & LeCun, Y. (2013). Learning Hierarchical Features for Scene Labeling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1915–1929. https://doi.org/10.1109/tpami.2012.231

[9]. Farfade, S. S., Saberian, M. J., & Li, L. J. (2015). Multi-view Face Detection Using Deep Convolutional Neural Networks. Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. Published. https://doi.org/10.1145/2671188.2749408

[10]. Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2010). Object Detection with Discriminatively Trained Part-Based Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9), 1627–1645. https://doi.org/10.1109/tpami.2009.167

[11]. Ganorkar, S. R., & Ghatol, A. A. (2007). Iris Recognition: An Emerging Biometric Technology. Proceedings of the 6th WSEAS International Conference on Signal Processing, Robotics and Automation, Corfu Island, Greece. Published.

[12]. Garg, D., Goel, P., Pandya, S., Ganatra, A., & Kotecha, K. (2018). A Deep Learning Approach for Face Detection using YOLO. 2018 IEEE Punecon. Published. https://doi.org/10.1109/punecon.2018.8745376

[13]. Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Published. https://doi.org/10.1109/cvpr.2014.81

[14]. Hasoon, F. N., Yousif, J. H., Hasson, N. N., & Ramli, A. R. (2011). Image enhancement using nonlinear filtering based neural network. Journal of Computing, 3(5), 171-176.

[15]. Hassin, A., & Abbood, D. (2021). Machine Learning System for Human–Ear Recognition Using Scale Invariant Feature Transform. Artificial Intelligence & Robotics Development Journal, 1-12.

[16]. Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Published. https://doi.org/10.1109/cvpr.2018.00745
[17]. Hu, Y., An, H., Guo, Y., Zhang, C., Zhang, T., & Ye, L. (2010). The Development Status and Prospects on the Face Recognition. 2010 4th International Conference on Bioinformatics and Biomedical Engineering. Published. https://doi.org/10.1109/icbbe.2010.5517197
[18]. Ignjatić, J., Nikolić, B., Rikalović, A., & ĆUlibrk, D. (2018). Deep Learning for Historical Cadastral Maps Digitization: Overview, Challenges and Potential. WSCG 2018 - Poster Papers Proceedings. Published. https://doi.org/10.24132/csrn.2018.2803.6
[19]. Jain, A., Ross, A., & Prabhakar, S. (2004). An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, 14(1), 4–20. https://doi.org/10.1109/tcsvt.2003.818349
[20]. Jin, X., & Tan, X. (2017). Face alignment in-the-wild: A Survey. Computer Vision and Image Understanding, 162, 1–22. https://doi.org/10.1016/j.cviu.2017.08.008
[21]. Karamizadeh, S., Abdullah, S. M., Zamani, M., Shayan, J., & Nooralishahi, P. (2016). Face Recognition via Taxonomy of Illumination Normalization. Intelligent Systems Reference Library, 139–160. https://doi.org/10.1007/978-3-319-44270-9_7
[22]. Khalajzadeh, H., Mansouri, M., & Teshnehlab, M. (2013). Face Recognition Using Convolutional Neural Network and Simple Logistic Classifier. Advances in Intelligent Systems and Computing, 197–207. https://doi.org/10.1007/978-3-319-00930-8_18
[23]. Khan, S., Javed, M. H., Ahmed, E., Shah, S. A. A., & Ali, S. U. (2019). Facial Recognition using Convolutional Neural Networks and Implementation on Smart Glasses. 2019 International Conference on Information Science and Communication Technology (ICISCT). Published. https://doi.org/10.1109/cisct.2019.8777442
[24]. Lander, K., Bruce, V., & Bindemann, M. (2018). Use-inspired basic research on individual differences in face identification: implications for criminal investigation and security. Cognitive Research: Principles and Implications, 3(1). https://doi.org/10.1186/s41235-018-0115-6
[25]. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539
[26]. Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. https://doi.org/10.1109/5.726791
[27]. Li, H., Lin, Z., Shen, X., Brandt, J., & Hua, G. (2015). A convolutional neural network cascade for face detection. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Published. https://doi.org/10.1109/cvpr.2015.7299170
[28]. Li, L., Mu, X., Li, S., & Peng, H. (2020). A Review of Face Recognition Technology. IEEE Access, 8, 139110–139120. https://doi.org/10.1109/access.2020.3011028
[29]. Liu, W., Zhou, L., & Chen, J. (2021). Face Recognition Based on Lightweight Convolutional Neural Networks. Information, 12(5), 191. https://doi.org/10.3390/info12050191
[30]. Mathias, M., Benenson, R., Pedersoli, M., & van Gool, L. (2014). Face Detection without Bells and Whistles. Computer Vision – ECCV 2014, 720–735. https://doi.org/10.1007/978-3-319-10593-2_47
[31]. Muhtahir, O. O., Adeyinka, A. O., & Kayode, A. S. (2013). Fingerprint Biometric Authentication for Enhancing Staff Attendance System. International Journal of Applied Information Systems, 5(3).
[32]. Nakada, M., Wang, H., & Terzopoulos, D. (2017). AcFR: Active Face Recognition Using Convolutional Neural Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Published. https://doi.org/10.1109/cvprw.2017.11
[33]. Nimbarte, M., & Bhoyar, K. (2018). Age Invariant Face Recognition using Convolutional Neural Network. International Journal of Electrical and Computer Engineering (IJECE), 8(4), 2126. https://doi.org/10.11591/ijece.v8i4.pp2126-2138
[34]. Oloyede, M. O., & Hancke, G. P. (2016). Unimodal and Multimodal Biometric Sensing Systems: A Review. IEEE Access, 4, 7532–7555. https://doi.org/10.1109/access.2016.2614720
[35]. Oloyede, M. O., Hancke, G. P., & Myburgh, H. C. (2020). A review on face recognition systems: recent approaches and challenges. Multimedia Tools and Applications, 79(37–38), 27891–27922. https://doi.org/10.1007/s11042-020-09261-2
[36]. P Tripathi, K. (2011). A Comparative Study of Biometric Technologies with Reference to Human Interface. International Journal of Computer Applications, 14(5), 10–15. https://doi.org/10.5120/1842-2493
[37]. Qin, H., Yan, J., Li, X., & Hu, X. (2016). Joint Training of Cascaded CNN for Face Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Published. https://doi.org/10.1109/cvpr.2016.376
[38]. Ramaiah, N. P., Ijjina, E. P., & Mohan, C. K. (2015). Illumination invariant face recognition using convolutional neural networks. 2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES). Published. https://doi.org/10.1109/spices.2015.7091490
[39]. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Published. https://doi.org/10.1109/cvpr.2016.91
[40]. Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149. https://doi.org/10.1109/tpami.2016.2577031

[41]. S. Manjula, V., & S. Santhosh Baboo, L. D. (2012). Face Detection Identification and Tracking by PRDIT Algorithm using Image Database for Crime Investigation. International Journal of Computer Applications, 38(10), 40–46. https://doi.org/10.5120/4741-6649
[42]. Salomon, M., Couturier, R., Guyeux, C., Couchot, J. F., & Bahi, J. (2017). Steganalysis via a convolutional neural network using large convolution filters for embedding process with same stego key: A deep learning approach for telemedicine. European Research in Telemedicine / La Recherche Européenne En Télémédecine, 6(2), 79–92. https://doi.org/10.1016/j.eurtel.2017.06.001
[43]. Sanchez-Moreno, A. S., Olivares-Mercado, J., Hernandez-Suarez, A., Toscano-Medina, K., Sanchez-Perez, G., & Benitez-Garcia, G. (2021). Efficient Face Recognition System for Operating in Unconstrained Environments. Journal of Imaging, 7(9), 161. https://doi.org/10.3390/jimaging7090161
[44]. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Published. https://doi.org/10.1109/cvpr.2015.7298682
[45]. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. The Journal of Machine Learning Research, 15(1), 1929–1958.
[46]. Sun, Y., Wang, X., & Tang, X. (2015). Deeply learned face representations are sparse, selective, and robust. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Published. https://doi.org/10.1109/cvpr.2015.7298907
[47]. Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. 2014 IEEE Conference on Computer Vision and Pattern Recognition. Published. https://doi.org/10.1109/cvpr.2014.220
[48]. Tang, J., Su, Q., Su, B., Fong, S., Cao, W., & Gong, X. (2020). Parallel ensemble learning of convolutional neural networks and local binary patterns for face recognition. Computer Methods and Programs in Biomedicine, 197, 105622. https://doi.org/10.1016/j.cmpb.2020.105622
[49]. Triantafyllidou, D., & Tefas, A. (2016). Face detection based on deep convolutional neural networks exploiting incremental facial part learning. 2016 23rd International Conference on Pattern Recognition (ICPR). Published. https://doi.org/10.1109/icpr.2016.7900186
[50]. William, I., Ignatius Moses Setiadi, D. R., Rachmawanto, E. H., Santoso, H. A., & Sari, C. A. (2019). Face Recognition using FaceNet (Survey, Performance Test, and Comparison). 2019 Fourth International Conference on Informatics and Computing (ICIC). Published. https://doi.org/10.1109/icic47613.2019.8985786
[51]. Yang, S., Luo, P., Loy, C. C., & Tang, X. (2018). Faceness-Net: Face Detection through Deep Facial Part Responses. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(8), 1845–1859. https://doi.org/10.1109/tpami.2017.2738644
[52]. Yousif, J. H. (2015). Classification of mental disorders figures based on soft computing methods. International Journal of Computer Applications, 117(2), 5-11.
[53]. Yousif, J. H., & Kazem, H. A. (2021). Prediction and evaluation of photovoltaic-thermal energy systems production using artificial neural network and experimental dataset. Case Studies in Thermal Engineering, 27, 101297.