

# Face Recognition using CNN:A Systematic Review

Aneesa M P  
CSE Department

MGM College Of Engineering And Pharmaceutical  
Sciences Velachery, India

Sabina N  
CSE Department

MGM College Of Engineering And Pharmaceutical  
Sciences Velachery, India

Meera K  
Assistant Professor  
CSE Department  
MGM College Of Engineering And Pharmaceutical Sciences  
Velachery, India

**Abstract**— With the constant development of computer technology, human dependence on network technology have grown, which leads to the importance of security issues. User authentication is an important thing to avoid attacks and security vulnerabilities. There are different authentication methods such as fingerprint scanning, voice recognition, SMS one-time passcodes, and face recognition. Face recognition is one of the important applications of image processing in still images and video. It's a true challenge to build an automated system which equals human ability to recognize faces. The main objective of this paper is to analyze importance of CNN, different datasets used in face recognition system and discuss about the different models of CNN. The deep learned CNN can be used for face recognition to provide more security in authentication purpose.

**Keywords**—CNN,Face recognition,DeepLearning

## I. INTRODUCTION

Due to the demand of computer technology, instead of using pencil and paper or face to face everyday actions are increasingly being handled electronically. With the advent of computer technology that results in great demand for fast and accurate user identification and authentication. Understanding user authentication is pivotal because it's a key step in the process that keeps unauthorized users from gaining access to sensitive information. There are different biometric authentication methods are available such as fingerprint recognition, facial pattern, voice or typing writing. Fingerprint recognition provide less accuracy due to skin distortion. Voice authentication is difficult for proceeding due to background noise and also if the user has a cold, then by the design, the person would not be recognized as a match with enrollee.

Now facial recognition is popular and widely used for person identification. A human face characteristic is different for each person to person. Camera is the only device that needed for face recognition.so it provides inexpensive and reliable personal identification which is applicable in many fields. An efficient face recognition system provides fast and accurate user identification and authentication. It has important role in many applications such as government use, commercial use, security gates, attendance management, smart cards, access control and biometrics.

Facial recognition is a technique that used to authenticate humans based on facial features.[2] Facial Recognition has been done in the past with different algorithms such as Gabor wavelet-based solutions, Face descriptor-based methods and Eigen face-based methods [1]. Facial recognition has been done using CNN due to their high frequency and virtuous recognition rate.

This paper describes the important CNN and different models of CNN used in face recognition. This will help the researchers to utilise the best solution for further improvement in this field.

The idea of face recognition system is the ability to recognize human face from image /video.

Face recognition system has mainly two parts,

- (i) Face Detection
- (ii) Face Authentication

**Face detection:** It is the process of finding human face in an image or video.

**Face authentication:** Facial recognition is a way of identifying or confirming an individual's based on facial features.



Fig 1:Face Detection

## A. Structure of CNN

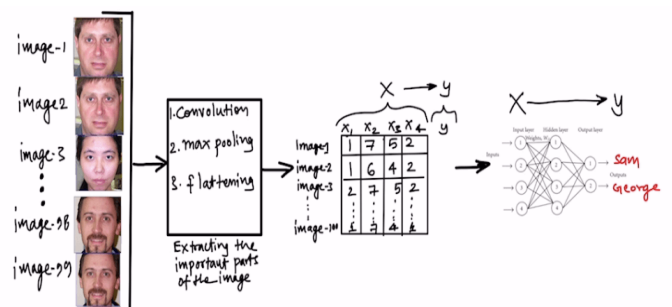


Fig 2: Structure of CNN

A Convolutional neural network (CNN) is a type of artificial neural network that has one or more convolution layers and are used mainly for image processing, classification, segmentation and also for other auto correlated data. Deep learning is a machine learning based artificial neural network that recognize objects in image by progressively extracting features from data through higher layers. As shown in figure in order to recognize face in an image we have to train the CNN with human faces. The benefit of using CNNs is their ability to develop an internal representation of a two-dimensional image. This allows the model to learn position and scale of faces in an image. After train the CNN it can able to recognize face in an image One can effectively use Convolutional Neural Network for Image data. CNN that extracts features in an image [2].

**B. How does CNN works?**

**Step 1:** An image is nothing but the 2-dimensional array. Before training an image, we need to process the dataset. By processing the dataset, we mean converting each image in to NumPy array. Each row represents an image. NumPy package is inbuilt function. Datasets is completely ready to be trained by the model.

**Step 2:** Neural networks are like layers. Each layer of neural network contains nodes which calculates some values based on characteristics or weights. Activation function are Relu for hidden layers and either sigmoid or SoftMax for output layers.

**Step 3:** Convolution layer is a fundamental mathematical operation that is highly useful for to detect features of an image. In this layer we pass kernel. i.e., n\*n matrix over the image pixel. Kernel has values in each of cell. It processed with original image help to produce some characteristics which help to identify images of the same object while predicting.

**Step 4:** Max Pooling operation involves sliding a 2-dimensional filter over each channel of features map and extract maximum features from image. Pooling layer used to reduce the dimension of feature map. It reduces the number of parameters to learn and amount of computation to perform. Pooling layer

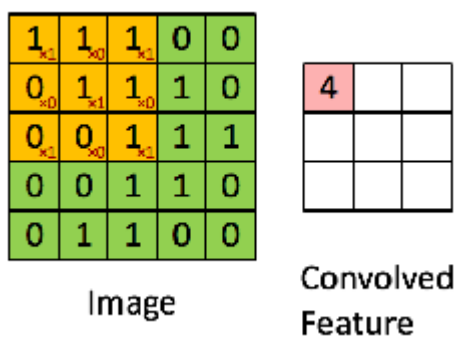


Fig 3: Convolutional operation

summarises the feature present in a region of the feature map generated by the convolution layer.

**Step 5: Flattening**

Flattening operation is performed when we got multidimensional output and we want to convert in to a single long continuous linear vector.

The flattened matrix is fed as input to the fully connected layer

**Step 6: Fully Connection Layer**

It is one of the fully feed forward neural network. It formed by last few layers. Once the image is convolved, pooled and flattened, the result is a vector. This vector act as the input layer for an ANN which then works normally to detect the image. It assigns random weights to each synapse; the input layer is weight adjusted and put in to an activation function. Every single neuron has a connection to every single neuron in next layer. The output is then compared with true values and the error generated is back-propagated, i.e., the weights are re-adjusted and all the processes repeated. This is done until the error is reduced or get correct output.

One of the greatest challenges of developing CNNs is adjusting the weights of the individual neurons to extract the right features from images. The process of adjusting these weights to get correct output is called training.

**C. Datasets used for Face recognition:**

Data is an important part in any Machine learning applications. Without data we can't able to train any model.

1. Flickr-Faces-HQ Dataset (FFHQ): It is a dataset contain of human faces in terms of age, ethnicity and image background. The images where crawled from Flickr and then automatically aligned and cropped. It consists of 70,000 high quality with 512\*512 PNG images.
2. Tufts-Face-Database: It is the most comprehensive, large-scale face dataset that contains visible, near-infrared, thermal, computerised sketch, LYTRO, recorded video, and 3D images. It consists of 10,000 images.
3. Real and Fake Face Detection: This dataset mainly used for fake face detection. It provides more application in the field of security. This dataset contains expr-generated high-quality photoshopped face images are composite of different faces, separated by eyes, nose, mouth, or whole face. The size of the dataset is 215MB.
4. Google Facial Expression Comparison Dataset: This dataset consists of face image triplets along with human annotations that specify, which two faces in each triplet form the most similar pair in terms of facial expression. The size of dataset is 200MB.
5. Face Images with marked landmark points: It is a Kaggle dataset to predict key point positions on face images. The size of the dataset is 497MP and contains 7049 facial images.
6. Labelled Faces in the wild home (LFW)Dataset: It is a database of face photographs designed for studying the problem of unconstrained face recognition is a public benchmark for face verification, also known as pair matching. The size of the dataset is 73MB and it consists of over 13,000 images of faces collected from the web
7. UTKFace Large Scale Dataset: It is a large-scale face dataset with long age span, which ranges

from 0 to 116 years old. The images cover large variations in pose, facial expression, illumination, occlusion, resolution and other such. The dataset consists of over 20k images with annotation of age, gender and ethnicity.

8. YouTube Faces Dataset with Facial Key points: This dataset is a processed version of the YouTube faces Dataset, that basically contained short videos of celebrities that are publicly available and were downloaded from YouTube. The size of dataset is 10GB, and it include approximately 1293 videos.
9. Yale Face Database: It contain 165 grayscale images in GIF format of 15 individuals' size of dataset is 6.4MB and contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions
10. Large-scale CelebFaces Attributes (Celeb A) Dataset :This data set contain quite 200k celebrity pictures, every with forty attribute annotations. The photographs in this dataset cover large pose variations and background muddle. The dataset can be used for the following computer vision tasks: face attribute recognition, face recognition, face detection, landmark localization, and face editing & synthesis.

## II USE OF CNN MODELS FOR FACE RECOGNITION IN DIFFERENT SCENARIOS:

AlexNet and GoogLeNet:

AlexNet and Google Net are pretrained CNN models. AlexNet and GoogLeNet can be for face recognition due to their excellent accuracy performances in computer vision. AlexNet is a deeper architecture with 8 layers, consist of 5 convolutional layers,3 max-pooling layers,2 normalization layers,2 fully connected layers, and one SoftMax layer. I is one of the 1st major CNN models that used GPU's for training. Due to the deeper architecture it is better able to extract features when compared to LeNet. For the three large linear layers, it was the first design to use max-pooling layers, ReLu activation functions, and dropout. The network was used to classify images in to 1000 different categories. GoogLeNet architecture is very different from the architectures such as AlexNet. It uses many different kinds of methods such as one by one convolution and global average pooling that enables it to create deeper architecture.

By using Celebrity face dataset for training where it stores at most 200,000 and 40 attributes. Different face expressions, views and background are the sample of 40 attributes indicated in this dataset. The AlexNet and GoogLeNet have better accuracy which is 100% compared to CNN which is 99.72%[3]. It shows that the machine has perfectly recognized all celebrity images in the dataset using AlexNet and GoogLeNet. This is due to the training of millions of data. CNN completes the execution or converges after 48 seconds while AlexNet achieved 100% accuracy in 9 minutes and 8 seconds. GoogLeNet requires 14 minutes and 47 seconds to converge or complete the execution.

### VGGNet:

The VGGNet is a deep convolutional neural network which extracts deeper features in the image and has higher feature extraction capabilities [2]. VGG stands for Visual Geometry Group; it is a standard deep convolutional neural network architecture with multiple layers. The VGG-16 network model is used to extract the features of face images. VGGNet model with SVM provide better accuracy when evaluated with LFW and CelebFaces Attributes Datasets (CelebA). VGG is a specific convolutional network designed for classification and localization and in which the characteristics of face image are extracted by convolution neural network VGGNet model, then the extracted feature dimensions are reduced by PCA, and finally face recognition is carried out by SVM classification method. VGGNet with SVM classifier provide an accuracy of 97.47%.

### MobileNet:

Researchers have designed more and more deep neural network models, such as VGGNet16/19, GoogLeNet, ResNet50 and so on. Compared with traditional classification algorithms, these have been excellent [4].Inorder to get more accuracy , people continue to deepen the network which leads to huge storage pressure and computational burden. Traditional CNN has large memory requirements and a large computational quantity, due to this it impossible to run on mobile devices and embedded devices. For the purpose of use in mobile phones google has proposed a lightweight deep neural network called MobileNetV1. It is a CNN model with a smaller model size, less trainable parameters and calculation amount, and is suitable for mobile devices. It takes full advantage of its computing resources and improves the accuracy of the model to the best extent.

### ResNet:

ResNet-50, with 50 layers is one of the variants of ResNet, a convolutional neural network. It has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer [5]. ResNet 50 has best time and memory performance compared to VGGNet. Identifying the identity of masked faces is a challenging problem for computer vision models since the features required to accurately predict the identity of an individual is reduced from the whole face to just the eye and sometimes the forehead. This study is built on existing pre-trained ResNet-50 architecture trained on human faces to solve the problem of identifying a person's identity when wearing a face mask, a ResNet-50 based architecture that performs well at recognizing masked faces and Networks with more number of layers can be trained easily without increasing the training error percentage and also it help in tackling the vanishing gradient problem using identity mapping .The outcome of this study could be seamlessly integrated into existing face recognition programs that are designed to detect faces for security verification purposes.

## III CONCLUSION

This work has presented a review about face recognition using CNN in different scenarios. CNN can take real-time pictures from surveillance camera to detect the presence of human face. CNN models such as ResNet, Mobile Net, VGG, Alex Net and Google Net can be used for face

recognition. This architectures can be used in both image and real time video. In the covid situation, ResNet-50 based architecture can be used for recognizing masked faces. In order to provide high accuracy, we can use transfer learning. We can increase accuracy by increasing the training set. As the number of layers increases computation time also increases. As the technology grows it become imperative to provide an efficient and accurate authentication system. This paper will help to provide an efficient authentication system using face recognition by CNN.

#### REFERENCES

- [1] K Jegabharathi 1, Dr.V. Latha Jothi2 “A survey on Face recognition system” ISSN: 2456-0197, ©Malla Reddy Engineering College (Autonomous), Volume 3, Issue3, September 2017.
- [2] Suleman Khan, M. Hammad Javed, Ehtasham Ahmed “Facial Recognition using Convolutional Neural Networks and Implementation on Smart Glasses” 2019 International Conference on Information Science and Communication Technology (ICISCT)
- [3] Rohit Halder, Rajdeep Chatterjee “A deep learning based smart attendance monitoring system” uploaded by Rajdeep Chatterjee on 11 April 2020
- [4] Nur Ateqah Binti Mat Kasim, Nur Hidayah Binti Abd Rahman, Zaidah Ibrahim, Nur Nabilah Abu Mangshor “Celebrity Face Recognition using Deep Learning “Indonesian Journal of Electrical Engineering and Computer Science Vol. 12, No. 2, November 2018.
- [5] Hongling Chen<sup>1\*</sup> and Chen Haoyu<sup>2</sup> “Face Recognition Algorithm Based on VGG Network Model and SVM “IOP Conf. Series: Journal of Physics: Conf. Series 1229 (2019) 012015 IOP Publishing doi:10.1088/1742-6596/1229/1/012015.
- [6] Yahui Nan , Jianguo Ju <sup>\*</sup>, Qingyi Hua a , Haoming Zhang a , Bo Wang” A-MobileNet: An approach of facial expression recognition”Received 26 August 2021; revised 18 September 2021; accepted 27 September 2021 Available online 19 October 2021.
- [7] Bishwas Mandal, Adaeze Okeukwu, Yihong Theis” Masked Face Recognition using ResNet-50” Preprint · April 2021.
- [8] G. Kaur, R. Sinha, P.K. Tiwari” Face mask recognition system using CNN model” Neuroscience Informatics 2 (2022) 100035.
- [9] Rohit Halder, Rajdeep Chatterjee “A deep learning based smart attendance monitoring system” uploaded by Rajdeep Chatterjee on 11 April 2020.
- [10] Isunuri B Venkateswarlu, Jagadeesh Kakarla, Shree Prakash” Face mask detection using MobileNet and Global Pooling Block” Authorized licensed use limited to: IEEE Xplore. Downloaded on May 30,2022 at 07:18:04 UTC from IEEE Xplore.
- [11] Showkat A. Dara , and S.Palanivelb “Neural Networks (CNNs) and Vgg on Real Time Face Recognition System” Turkish Journal of Computer and Mathematics Education Vol.12 No.9 (2021), 1809-1822.