



# SURVEY ON FACE RECOGNITION USING CNN

**Pavan S<sup>1</sup>, Thanuja N<sup>2</sup>**

MTECH, CSE, 1st SEM, BIT<sup>1</sup>

Asst. Professor, CSE, BIT<sup>2</sup>

**Abstract**– Face recognition is the process of identifying a human person by their face using current, highly developed technology. Face recognition has gained interest in the field of biometric in this post COVID era where face biometric is considered more than finger biometric. Face recognition has several uses, including attendance tracking, identity verification, law enforcement, and surveillance. Facial recognition has various steps involved, which is to recognize the face area from the RAW images, alignment of the face recognized, feature extraction and finally classification of the face. So in this survey paper, we are going to discuss and compare the various techniques which are involved in facial recognition using Convolutional Neural Network(CNN) and also about the accuracy which we get when using those methods.

## I. INTRODUCTION

Identification or recognition of a person based on their visual traits is called face recognition. For humans to recognize a person it will take a few seconds, because the human brain will have innate ability to recognize small differences in facial features. Thus it will provide remarkable accuracy. But for the computer to recognize they will be facing a lot of complications because they are not like humans. They will only recognize based on some calculations which are contrived on the facial features. The mouth, nose, and eyes are a few examples of face characteristics. The face recognition field is mainly the intersection of the three main domains which are Image Processing, Artificial Intelligence and Cyber Security[1].

Facial recognition will occur in three steps which are preprocessing, training the dataset and testing the dataset. The preprocessing of the images will have various steps which are involved. The first one is to find the face points from the entire image. Next, to align the face in a standard format for all the images. Noise reduction is also considered one of the steps in preprocessing, which will reduce the distortion. Gray scaling of the image will help the system to recognize the image in a much faster and efficient manner. Like that, there will be many steps which will be involved in preprocessing of the images from the given dataset. After the preprocessing of the dataset, we are going to split the dataset into two parts: the training data and the testing data. The training data is a labeled dataset which is used for training our model. This consists of input data and the corresponding output or target values. The testing data is used for evaluation of our model. This consists of input data but no corresponding output or target values. After all the steps are completed, the accuracy of the model is estimated.

In this survey paper, we are going to discuss the various techniques for facial recognition through using Convolutional Neural Network(CNN). Different approaches for face recognition are using Prewitt operator [2] and improved neural network which can reduce the recognition time and can boost the recognition rate, CNN-based KNN classifier and SIFT-based KNN classifier [1], CNN-PCA based feature extraction which involves calculating the mahalanobis distance [4], comparing convolutional neural based models AlexNet, VGG16, VGG19 and MobileNet [5] and so on.

## II. BACKGROUND KNOWLEDGE

### Face Recognition

Face recognition is the process of identification or verification of individuals based on the facial features they possess. Pratul Kumar et al [7] states about assigning the correct label to the face under different studies. Face will have features, such as eyes, nose, mouth etc. These features are considered while extracting information out of the face for the recognition process. But considering the surveillance footage, the face images will have low resolution which may hinder the process of face recognition [3].

### Face Recognition using CNN



A deep learning neural network known as a convolutional neural network (CNN) excels at analysing visual data, such as pictures and videos. CNN has developed interest in computer vision such as object detection, image recognition and many more.

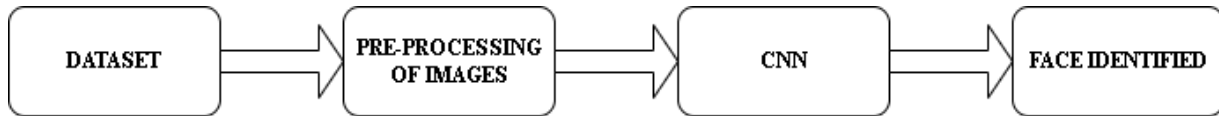


Fig 1. Steps involved performing Face recognition using CNN

(a) Dataset

Fig1 shows the steps involved in performing Face Recognition using CNN. The face recognition system is comprehensively dependent on the dataset which is used. Shyam Singh Rajput et al [3] have used CMU-PIE face database and Extended Yale database which has 98.84% and 91.01% respectively.

(b) Pre-Processing of dataset

After considering a particular dataset for the model. The preprocessing of the dataset needs to be done. The images in the dataset will be of different format, size, resolution, angles and so on.

In [1] the researcher has used Adaptive Median Filter (AMF) which removes impulse, smoothing of other noises and reduction of distortion. Edy Winaro et al [4] uses the most common techniques, such as cropping, resizing, RGB-gray, and histogram equalisation, that are also used by many others.

(c) CNN

As mentioned above CNN is a deep neural network which is majorly used for face and object detection. Fig 2 represents the architecture diagram of Convolutional Neural Network. The key characteristic of CNN is to automatically learn and extract relevant information from the dataset. These are done by using convolution layer, pooling layer and fully connected layer.

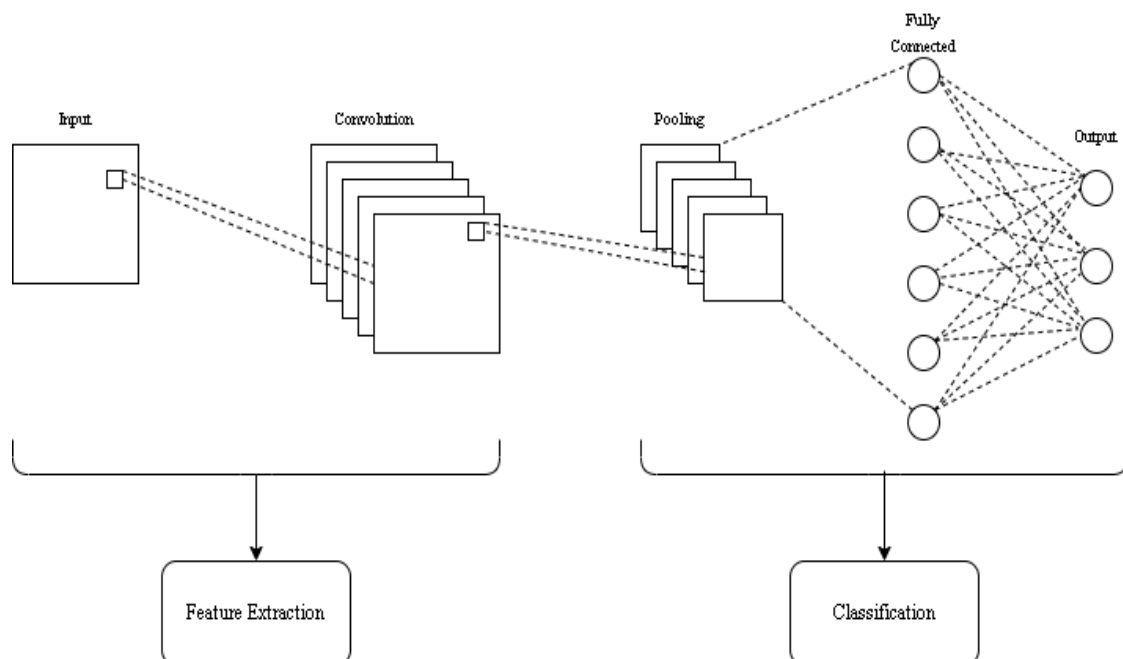


Fig 2. Convolutional Neural Network Architecture diagram

Convolutional layers are also called as the feature detectors for the input data. These filters scan the images in a sliding window manner, performing element wise multiplications and summations to extract local patterns and features. The output of the convolution layer is called feature maps.



Pooling layer reduces the spatial dimension of the feature maps. They down sample or aggregate the information with the local region of the feature map. The most common pooling is max-pooling, which selects the maximum value within each region.

Activation function the activation function of a node defines the output of that node given an input or set of inputs. Rectified Linear Unit (ReLU) is a commonly used activation function in CNNs, which sets negative values to zero and keeps positive values unchanged.

The ultimate predictions or classifications based on the features discovered from the preceding levels are made by fully connected layers. The outputs from the previous layers are flattened into a vector and connected to the fully connected layers, which perform weighted summations and apply activation functions to produce the final outputs.

(d) Face Identified

The face will be recognised at the conclusion of the procedure, and the model's accuracy will be assessed. So the evaluation metric in Jamilah Almari et al [1], will be based on performance and AI based metrics. For performance total processing time will be considered. For AI based, accuracy will be considered.

### III. SURVEY

In [2] Jiang Xiao et al, face recognition is based on the Prewitt operator and improved CNN, which is the proposed algorithm by Jiang Xiao et al. The dataset will be preprocessed by the histogram equalization and Prewitt operator. To prevent overfitting, they have used L2 regularization and Dropout. The activation function used here is Tanh and ReLu. The proposed algorithm will produce an accuracy of 98.1% with a performance time of 0.2 sec.

In [1] Jamilah Almari, face recognition is based on two feature extraction methods which are Scale Invariant and Convolutional Neural Network. For classification of the feature they have used KNN classifier. ORL is the database used by [1] which has 400 samples of 40 subjects with a preprocessing method of Adaptive Median Filter(AMF). The models proposed are KNN-CNN and KNN-SIFT which will have the accuracy of 97% and 95% respectively.

Researcher	Dataset	Samples	Subjects	Pre-Processing methods	Model	Result
Jamilah Alamri et al	ORL	400	40	Adaptive Median Filter(AMF)	KNN-CNN	97%
					KNN-SIFT	95%
Shayam Singh Rajput et al	CMU-PIE face database	41,368	68	Low Resolution-Face Recognition(LR-FR)	CNN	98.84%
	Extended Yale face database	16,128	38			91.01%
Edy Winaro et al	Own Database	N/A	N/A	Cropping method, re-sizing, RGB-gray, histogram equalization	PCA-CNN	98%
Tanvir Ahmed et al	N/A	1300	13	N/A	CNN-AlexNet	57%
					CNN-VGG16	71%
					CNN-VGG19	56%



					CNN- MobileNet	84%
Pratul Kumar et al	Webcam	7767	10	N/A	CNN	96%
Haofei Wang et al	Extracted from Google	540	36	Electroencephalography(E EG) Electrooculogram(EOG)	CNN	95%
Yang Zhiqi et al	Casia Webface database	1,20,000	2000	Microface	CNN- VGGNET	96.26%
Jiang Xia et al	N/A	N/A	N/A	Prewitt operator	Improved CNN	98.1%

Table 1. Tabulation of summary

In [3] Shayam Singh Rajput et al, where the main focus is based on low resolution face detection where they have used two sets of database that are CMU-PIE face database and Extended Yale database with sample of 41,368 of 68 subjects and sample of 16,128 of 38 subjects. This model resnet18 and ReLu is chosen as the activation function. Relu function is a linear function, where the images are more precisely categorized. The proposed algorithm will produce an accuracy of 98.84% when using CMU-PIE database and 91.01% when using Extended Yale Database.

In [4] Edy Winaro et al, the focus here is based on an attendance system which uses face recognition. This model uses CNN-PCA, when combined together will produce a more accurate feature extraction method. The preprocessing methods used here are more general such as cropping method, resizing, RGB gray, etc. This model reconstructs 2D face images to 3D by using CNN. The mahalanobis distance is used as a classification method. The accuracy produced by this model is 98%.

In [5] Tanvir Ahmed et al, have used CNN based models such as AlexNet, VGG16, VGG19 and MobileNet. The dataset has not been disclosed here. The dataset is made up of face images of 10 celebrities, 130 images for each celebrity. 100 images to train, 20 images for validation and 10 images to test. As a deep neural network we require more data hence they perform data augmentation to avoid overfitting. AlexNet will have eight layers- five convolutional and three fully connected layers. ReLu is used as an activation function here. Whereas in VGG, there are 12 convolutional layers some of them are followed by maximum pooling and 4 fully connected layers. The accuracy of the CNN based models AlexNet, VGG16, VGG19 and MobileNet will have accuracy of 57%, 71%, 56%, 84%

In [6] Yang Zhiqi et al, they have used VGGNET-CNN as their model for face recognition. Deep learning algorithm VGGNET is used for image classification and proposes a facial recognition method called MicroFace. They have used the CASIA webface database as the testing and training samples. Dataset will have 1,20,000 images of 2,000 subjects. VGGNET has 16-19 convolutional layers which has an impact on the accuracy (more the layers higher the accuracy). The proposed model will have 96.26% accuracy.

In [8] Haofei Wang et al, proposed a system which is based on electroencephalography(EEG) and single trial target detection using rapid visual presentation (RSVP) paradigm. The face dataset is extracted from google. Here a person with no physical condition and no mental condition will be used. The person will get NeuroScan, which is a cap which has 60 Ag/AgCl electrodes which is used to record brain impulse. Electrooculogram (EOG) is the data being recorded from the place above and below the left eye. The target images will be placed somewhere in the queue along with the non target images. The person has to click the image within the given timespan of 500 ms before the image vanishes. The proposed system will have an accuracy of 95%.



In [7] Pratul Kumar et al, they proposed a model where it uses CNN to classify and identify the images. The dataset used here is from the system webcam which has 10 subjects and 7767 images of different lighting, illumination, background etc. The accuracy obtained in this model is 96%.

#### IV. CONCLUSION

After going through the research paper, we can come to a conclusion that we require images in the image data that we can properly train our model. In [3] Shayam Singh Rajput et al has the maximal accuracy of 98.84% for the CMU-PIE face database which has been used. Since the aim of [3] is to find the high resolution image from the image which has low resolution. The dataset used here is the most if we see the Table 1 for reference. Next comes the model which was proposed by Edy Winaro et al [4], which has an accuracy of 98%. This model is based on PCA-CNN. This survey paper's conclusion is that in order to improve accuracy and performance, we need to add additional data to the dataset and use one or more algorithms with CNN.

#### REFERENCES

- [1] Jamilah Alami et al, "Face Recognition based on Convolution Neural Network and Scale Invariant Feature Transform" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol.12, No. 2, 2021.
- [2] Jiang Xiao et al, "Face recognition algorithm based on Prewitt and convolutional neural network" 2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE).
- [3] Shyam Singh Rajput et al, "CNN Classifier based Low-Resolution Face Recognition Algorithm" 2020 International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET).
- [4] Edy Winaro et al, "Attendance System Based on Face Recognition System CNN-PCA Method and Real-time Camera" 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI).
- [5] Tanvir Ahmed et al, "A Comparative Study on Convolutional Neural Network Based Face Recognition" 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT).
- [6] Yang Zhiqi, "Face recognition based on improved VGGNET convolutional neural network", 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC).
- [7] Pratul Kumar et al, "Convolutional neural network based face recognition approach", TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON).
- [8] Haofei Wang et al, "Convolutional Neural Network for Target Face Detection using Single-trial EEG signal", 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).