International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 05 | May 2020 www.irjet.net p-ISSN: 2395-0072

Face Detection and Recognition through Viola Jones Algorithm and CNN

Aiswarya Devadas[1], Shine P James [2]

[1]PG Student, Dept. of Electronics and Communication Engineering, College of Engineering Poonjar, Kerala, India
[2]Assistant Professor, Dept. of Electronics and Communication Engineering, College of Engineering Poonjar, Kerala, India

---

Abstract - *In this work home security through face detection and recognition method is explained. The system consist of two components: one is face detection and the other is face recognition. Face detection is done by Viola Jones algorithm and recognition is done through CNN. For face detection and recognition we use real time dataset created using a webcam and LFW datasets also are used. Here recognition is done through splitting the total images into 80% training and 20% testing images. Some augmentation techniques are used to enlarge the dataset. Live image testing is also performed. The face recognition success rate based on this network is 98.86%.*

Key Words: **face recognition, face detection, CNN, Viola Jones**

1. INTRODUCTION

The need of security system for home is considered as one of the important aspects of our modern life. The aim of this paper is to help users for improvement of the door security by using face detection and recognition. Face is considered as a complex multidimensional structure and needs good computing techniques for detection and recognition. Nowadays face recognition system is becoming an increasing trend across the world in providing extremely safe and reliable security technologies. Moreover this system is providing vast benefits when compared to other biometric methods like palm print, fingerprint etc. The system will capture the image of the person who intended to enter the house. From the captured image face is detected using Viola Jones algorithm. Face detection is the process of detecting the region of the face in an image. The detected face is then recognized using CNN. Face recognition is the process of identification of a person. Before recognition CNN are trained using 80% of the total images stored in the dataset. After training rest of the 20% of images are used to test the network. Some augmentation techniques are provided to enlarge the dataset. The accuracy of the system is also measured using LFW datasets.

2. METHODOLOGY

2.1 Face Detection

Face is detected from the captured images using viola jones algorithm. It is a haar feature based algorithm. Its main property is its fast detection. Here a detector object is created and then this detected object is applied to each captured images. Here Frontal face cart classification model is used for detecting the face from any captured images. The detection happens inside a bbox. After finding the four coordinates of bbox, the detected face is enclosed in that bbox.

2.2 Face Recognition

After detecting the face from the images, the next step is to identify the detected face. Here CNN is used for face recognition. Before training the total images in each datasets, they are split into two. 80% of the total images are taken for training the CNN and rest of the 20% images are used for testing.

2.3 Accuracy Measurement

Accuracy is measured using confusion matrix. A confusion matrix is a summary of prediction results on a classification problem. Here the no. of correct and incorrect predictions is summarized with count values. The confusion matrix shows the ways in which our classification model is confused when it makes predictions or it is a technique for summarizing the performance of a classification algorithm. Here the diagonal elements indicate the no. of faces correctly classified in each class. Sum of all elements indicates the total no. of images used for testing.

$$\text{Accuracy} = \frac{\text{Sum of diagonal elements}}{\text{Sum of all elements}}$$

3. PROPOSED METHOD

In this section we give a detailed description of the architecture and the different layers of CNN.

A. Database creation
B. Augmentation
C. Architecture
D. Different layers

3.1 Database Creation

Created dataset using webcam and LFW datasets are also used in the face recognition system. The datasets are a collection of real world imaging conditions that varies in appearance, noise, posture, light and so on. For creating dataset, 100 images of ten people each are captured using

webcam. Face is detected from each image with the help of viola jones algorithm, which uses bounding box (bbox) for enclosing the detected face. Then the detected face is cropped and resized to 256x256. After deleting the irrelevant detection the faces are stored in the dataset.

Labeled Faces in the Wild (LFW) is a database of face photographs designed for studying the problem of unconstrained face recognition. This database was created and maintained by researchers at the University of Massachusetts, Amherst. 13,233 images of 5,749 people were detected and centered by the Viola Jones face detector and collected from the web. 1,680 of the people pictured have two or more distinct photos in the dataset.

3.2 Augmentation

Image augmentation is a very powerful technique used to artificially create variations in existing images to expand an existing image dataset. This creates new and different images from the existing image dataset that represents a comprehensive set of possible images. This is done by applying different transformation techniques like rotating the existing image by a few degree, translating the existing set of images in x and y directions etc. Deep learning convolutional neural networks need a large number of images for the model to be trained effectively. This helps to increase the performance of the model by generalizing better and thereby reducing overfitting. Image augmentation creates a rich, diverse set of images from a small set of images for image classification or object detection and recognition.

3.3 Architecture

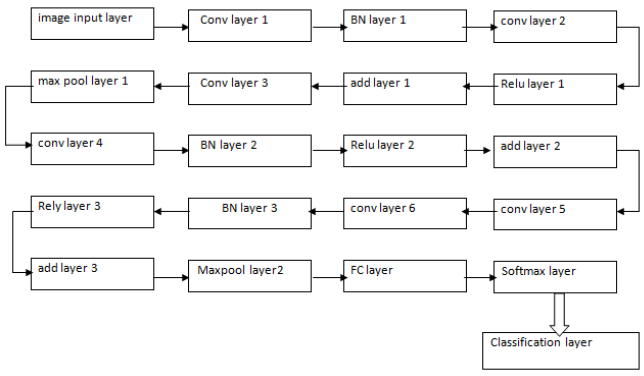

Fig -1: Architecture of CNN

3.4 Different Layers

(i) Convolution layer

Convolution layer is the first layer used in CNN for extracting features from the input image. Convolution is a mathematical operation that takes two inputs and a filter or kernel. Here first the images are resized to an optimal size and fed as an input to the convolution layer. The filter actually slides over the input images, thus it is multiplying the values in the filter with the original pixel values of the images (i.e, element wise multiplication). The multiplication are summed up and generating a single value for that particular receptive field. Finally it mapped to an array known as feature map. Convolution layer can be considered as the eyes of the CNN. The neuron in this layer looks for specific features.

(ii) Pooling layer

Pooling layers will reduce the number of parameters when the images are too large. Pooling operations are also called down sampling, which reduces the dimensionality of each feature map, but it will retain the important information. Pooling operation can be of different kinds

- Max pooling
- Average pooling
- Sum pooling

Max pooling takes the largest element from the rectified feature map. Average pooling returns the average of all the values from the portion of the image covered by the kernal. Sum of all elements in the feature map are called as sum pooling. Pooling layer will decrease the computational power required to process the data through dimensionality reduction. Furthermore it is used for extracting dominant features which are rotational and positional invariant, thus maintaining the process of effectively training the model.

(iii) ReLu layer

ReLu stands for Rectified Linear Unit for a nonlinear operation. The output is f(x)= max(0,x). ReLu's purpose is to introduce nonlinearity into our ConvNet. Since real world data would want our ConvNet to learn would be non-negative linear values. There are other non-linear functions such as tanh or sigmoid which can also be used instead of ReLu. Most of the data scientists use ReLu, since performance wise ReLu is better than other two.

(iv) Softmax layer

The way this layer works is that it looks at the output of the previous layer and the number of classes. Basically a softmax layer looks at what high level features most strongly correlate to a particular class and has particular weights, so that when you compute the products between the weights and the previous layer, you get the correct probabilities for the different classes.

(v) Fully connected layer

Fully connected layers are essential component of CNN which determines the number of hidden layers used in the network. This layer has been proven very successfully in recognizing and classifying images for computer vision. The

CNN process begins with convolution and pooling, breaking down the image into features and analyzing them independently. The result of this process feeds into a fully connected neural network structure that drives the final classification decision. The objective of FC is to take the result of the convolution/pooling process and use them to classify the image into a label.

4. RESULTS

4.1 Layer Graph

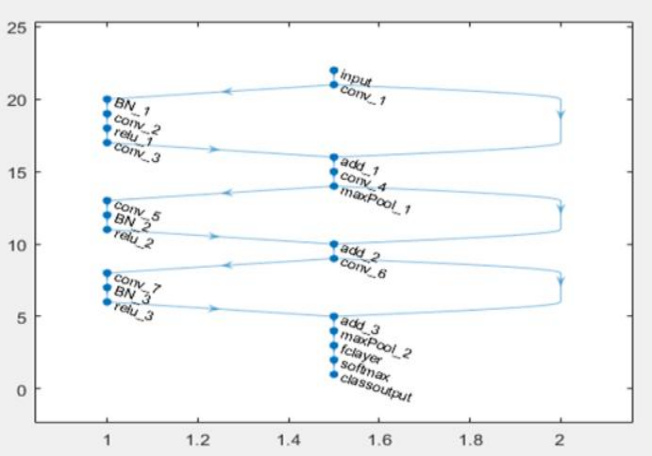

Chart -1: Layer graph

4.2 Training process

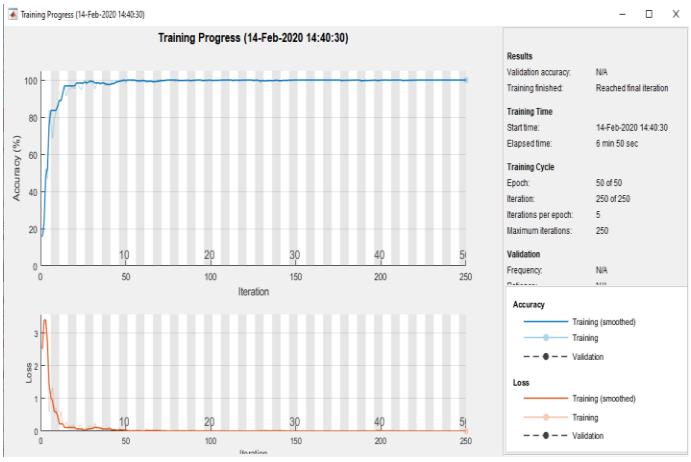

Chart -2: Training progress using created dataset

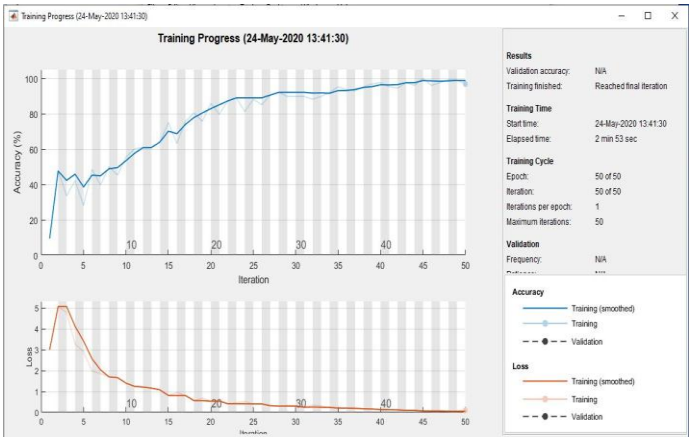

Chart -3: Training progress using LFW dataset

Table -1: Datasets and accuracy

	datasets	Identity	Accuracy
Face Recognition	Created dataset using webcam	1000	98.3%
	LFW	5749	80.6%

4.3 Confusion Matrix

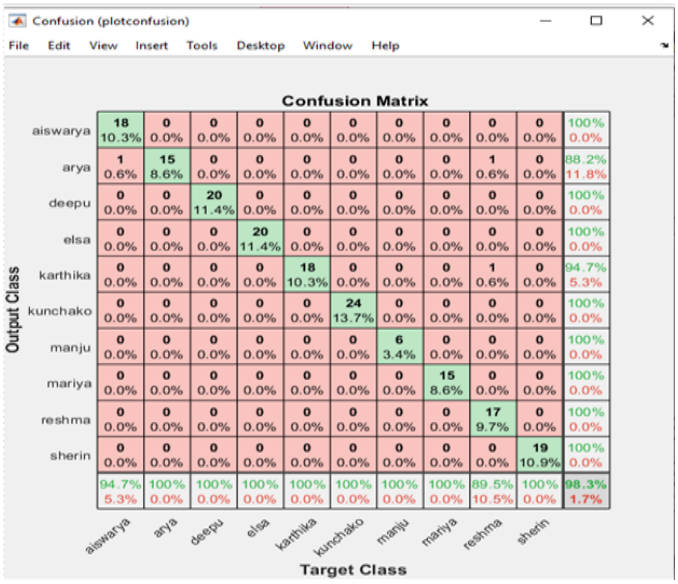

Fig -2: Confusion matrix using created dataset

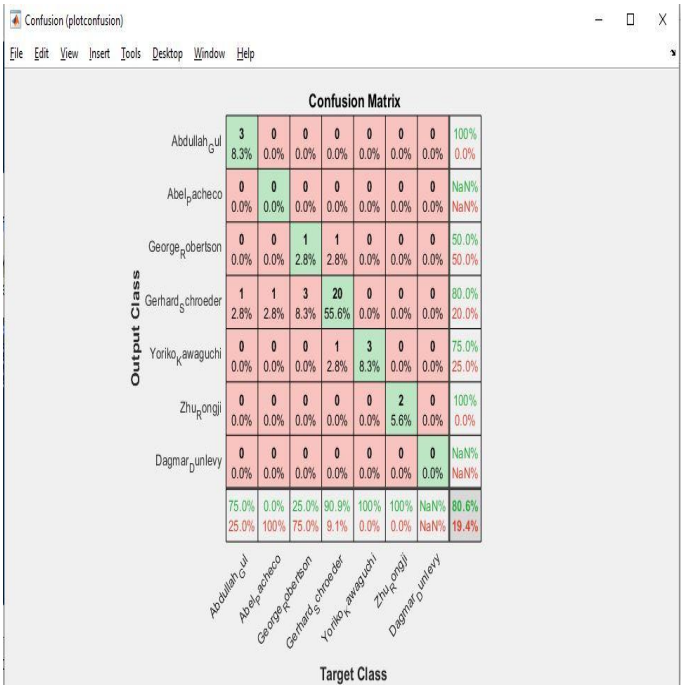

Fig -3: Confusion matrix using LFW dataset

5. CONCLUSIONS

A face recognition method based on convolution neural network (CNN) is presented in this paper. Here the network has 21 layers in which 6 layers are convolutional layer and one is softmax layer. Some augmentation techniques are used before training and testing process. The network achieves an accuracy of 98.6% comparing with other face recognition techniques. Also working of the system with live images is tested. Moreover the network has excellent convergence and strong robustness. This paper presents a face detection and recognition algorithm based on viola jones and CNN. CNN use softmax layer to classify different faces in appropriate classes based on probability. The algorithm has excellent performance among the existing algorithms. The test result on the live image dataset also proves the effectiveness of the proposed method. Traditional feature extraction based methods becomes time consuming for selecting appropriate features for the learning of the model. CNN automatically learns those features efficiently and that is why CNN become very convenient for real-world scenarios. This paper uses a webcam to capture the image of the person to create the dataset and LFW datasets also are used. Face is detected from it using viola jones algorithm. Then detected face is used for training CNN for face recognition. CNN uses softmax layer to classify each faces to appropriate classes. The network achieves an accuracy of 98.85 by using created dataset and 95% by using LFW datasets. The test results on the live image datasets also prove the effectiveness of the proposed method.

REFERENCES

[1] Yuxiang Zhou and Hongjun Ni, "Face and Gender Recognition System Based on Convolutional Neural networks", Proceedings of IEEE International Conference on Mechatronics and Automation,2019.

[2] Kewen Yan1, Shaohui Huang2," Face Recognition Based on Convolution Neural Network", Proceedings of the 36th Chinese Control Conference,2017.

[3] Minjun Wang , Zhihui Wang,"Deep Convolutional Neural Network Applies to Face Recognition in Small and Medium Databases", The 4th International Conference on Systems and Informatics,2017.

[4] Beatrice Dorothy, Dr. S. Britto Ramesh Kumar," IoT based Home Security through Digital Image Processing Algorithms", World Congress on Computing and Communication Technologies, 2017.

[5] Zi wang1, Chengcheng li1, Huiru sha," Eye recognition with mixed convolutional and residual network (micore-net)", special section on information security solutions for telemedicine applications",2018.

[6] Nurulhuda ismail, Mas idayu md. Sabri, "Review of existing algorithms for face detection and recognition", recent advances in computational intelligence, man-machine systems and cybernetics,2018.

[7] Aditi Upadhyay1, Sudhir Kumar Sharma, "Face Recognition and Verification: A Literature Review", International Journal of Emerging Technology and Advanced Engineering,2017.

[8] Ravinder Pal Singh1, DaljitKaur, "Review of Face Detection and Recognition Techniques", International Journal of Engineering Research and General Science Volume 3,2015.

[9] Marko Arsenovic, Srdjan Sladojevic, "FaceTime – Deep Learning Based Face Recognition Attendance System", IEEE 15th International Symposium on Intelligent Systems and Informatics,2017.

[10] Ankit S. Vyas, Harshadkumar B. Prajapati," Survey on Face Expression Recognition using CNN", International Conference on Advanced Computing & Communication Systems,2019.

[11] Vinita Bhandiwad, Bhanu Tekwani, "Face Recognition and Detection using Neural Networks", International Conference on Trends in Electronics and Informatics,2017.

[12] Ya Wang, Tianlong Bao, Chunhui Ding," Face Recognition in Real-world Surveillance Videos with Deep Learning

Method", 2nd International Conference on Image, Vision and Computing,2017.

[13] Mehul K Dabhi1, Bhavna K Pancholi, "Face Detection System Based on Viola - Jones Algorithm", International Journal of Science and Research,2016.

[14] Ole Helvig Jensen, Kongens Lyngby," Implementing the Viola-Jones Face Detection Algorithm", Technical University of Denmark Informatics and Mathematical Modelling,2008.