Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI 2018)
IEEE Conference Record: # 42666; IEEE Xplore ISBN:978-1-5386-3570-4

Face recognition using Convolutional Neural Networks

V.K.N Kamlesh Pai
Department of Computer Engineering
St. John College Of Engineering and Management
Village Vevoor, Manor Road, Palghar (East) 401404
kamleshkarthik@gmail.com

Manoj Balraj
Department of Computer Engineering
St. John College Of Engineering and Management
Village Vevoor, Manor Road, Palghar (East) 401404
manojb912@gmail.com

Sachinkumar Mogaveera
Department of Computer Engineering
St. John College Of Engineering and Management
Village Vevoor, Manor Road, Palghar (East) 401404
sachinkumarmogaveera@gmail.com

Deepak Aeloor
Department of Computer Engineering
St. John College Of Engineering and Management
Village Vevoor, Manor Road, Palghar (East) 401404
deepakaeloor@gmail.com

Abstract— **Face recognition is a common biometric authentication technique used to analyse the face images and extract useful recognition information from them, which are always called as a feature vector that is used to distinguish the biological features. Face Recognition process begins with extracting the coordinates of features such as width of mouth, width of eyes, pupil, and compare it with a stored face template. The aim of the proposed system is to design an autonomous security system that performs face recognition based surveillance combined with a hardware mechanism to lockup the secured region. Haarcascade algorithm is used to detect and extract the face from an image thereby storing samples in order to train the system. The system consists of a buzzer alarm and two cameras diagonally placed. The camera locates, tracks people entering the secured room, recognize the individual and message is passed to the control room which is stored in the log file. Any unauthorized access is logged along with a buzzer alarm to notify the control room followed by locking the exit points of the system. This system focuses on system security using face recognition which can be installed at banking suits.**

Keywords— ***Haarcascade algorithm, PCA, LDA, Convolutional Neural Networks, Deep Learning, HMM, LFW.***

I. INTRODUCTION

Biometrics authentication (or realistic authentication) is a form of authentication which is used for identification and access control. Biometric authentication is mainly based on physiological and behavioural characteristics. The traits such as uniqueness, permanence, measurability, performance, acceptability and circumvention is checked in an individual for biometric verification [1].

There are various types of biometric authentication like fingerprint identification, Iris scan, retina scan, face recognition, voice analysis,etc. Fingerprint identification is most commonly used form of authentication in biometrics. But the disadvantage is that a person's fingerprint's pattern or form may change over time and fingerprint scanner does not take this into consideration.

In the current scenario, there are lot of face recognition techniques and algorithm found and developed around the world. Face recognition therefore, has received a great deal of attention in various applications in the field of image processing, computer vision, etc due to several advantages it has over other biometric method. For example, in public security system, it can identify the identity of the suspect; in the bank and customs control system, it can identify and prove the identity; it also helps users safeguard its own confidential information and experience more secure financial transaction.
The proposed system is trained to recognise a set of authorized person. Haarcascade system is used to create dataset of authorized person dynamically by identifying and extracting the facial features of face helping the system to recognise the face. All others who enter the guarded area are considered strangers. Neural network is used to train the system in order to identify the stranger by comparing the dataset of all authorized person**.** When an unauthorized person is detected in secured region, the buzzer alarm alerts the control room and the system triggers the hardware which closes all the exit points at the same time. Stranger is thus locked inside the room and immediate action is taken on the suspect.

II. RELATED WORKS

The algorithms commonly used for face recognition are active contour model [2] and deformable template model [3]. This model is based on the geometrical characteristic, which is first applied to the face recognition problem. Its basic idea is the difference of everyone's face because of difference in components of every face, like the eyes, noses, mouths and jaws are different. Thus the system uses the set of

architectures and shapes of these components to be the features for the face recognition problem.
There are five useful methods for face recognition developed in the past study.

The sub-space analysis method is often used in face recognition, which contains two methods such as **Principal Component Analysis(PCA) and Linear Discriminant Analysis(LDA)**. PCA is a technique used for identification of a smaller number of uncorrelated variables known as principal components from a larger set of data. The technique is widely used to emphasize variation and capture strong patterns in a data set. Principal component analysis is considered as a useful statistical method and used in predictive model and exploratory data analysis. The most classic method is PCA-based Eigenface which was put forward by Turk [4] in 1991. This method take the face images as random variables, which turns the N× N vector of a face image to a N2 × 1 vector, and after minuses the mean data vector, uses the K-L transformation to get a set of orthogonal basis, then after keeps part of the principal components, the reduced dimension vector space of face images is obtained. LDA[5] is aimed at the separability of the samples. It tries to find a projection direction, which can make the distance of within-class, is small and the distance of between-class is large based on the training samples' projection to that direction. Compared to the PCA method, only if the training sample is large, LDA can get a better result.

Researchers employed the **Hidden markov model (HMM**) to solve the problem that the different appearance of facial features and the connection of each other. Based on this model, the features observed are treated as a sequence of unobserved states. Different people use different HMM parameters, and for the same person, system uses the model with same parameters to represent the observed sequence of gestures and facial expressions. Samaria [6] first proposed the face model, which used a rectangular window sampling face images from top to bottom.

Another commonly practised method for face recognition is **Neural network (NN).** Neural network uses its ability of learning and classifying to extract and recognize face features. Lin, etc. [7] use the positive and negative samples for reinforcing learning to get an ideal probability result. And then increase the learning speed by applying a modular network.

Proposed system was inspired by Ya Wang's **Deep learning method** [8] for Face Recognition in Real-world. This system automatically generates dataset from real world surveillance video. This helps in generating dataset in various light illuminations, with different facial expressions, etc.

Another inspiration is Ze Lu's system [9] that performs extremely well.It improves face recognition performance of **Convolutional neural network (CNN's)** by using non-CNN features.The non-CNN features showcase the characteristics from a different perspective of the targeted face images.

In terms of results, Face recognition [10] based on deep neural network works the best. The system uses CNN which is a neural network capable of handling image data. It comprises of three layers, one convolution layer, one pooling layer and one fully connected layer.CNN can learn the variations of data without prior knowledge. This method also helps in identifying a person using additional features. The system uses Labeled Faces in the Wild (LFW) dataset for its implementation. A dataset of face photographs designed for studying the problem of unconstrained face recognition, known as LFW (Labelled Faces in the Wild),contains more than 13,000 images of faces collected from the web.

III. PROPOSED SYSTEM

1. Overall block diagram

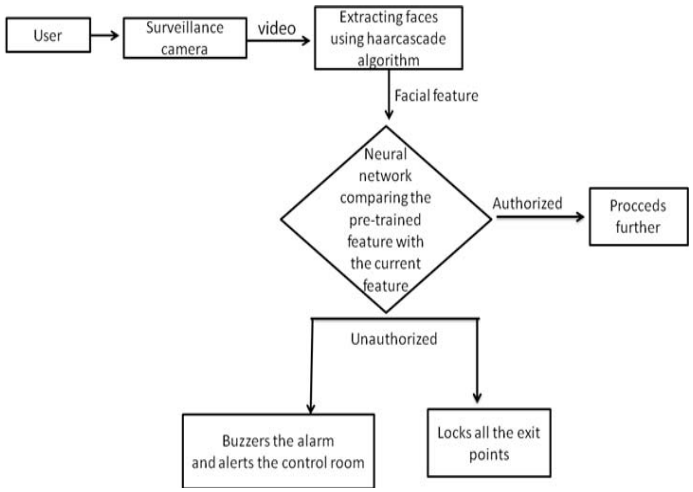


Fig 1: Block Diagram

The proposed system focuses on taking action against the intruder who enters the secured area of the banking system. The system uses face recognition technology to recognise the intruder thereby, taking appropriate action against intruder.
Only authorized persons are allowed to enter the room. When the user enters the room, input to the system is given by surveillance camera which extracts faces from the captured video using haarcascade algorithm. Haarcascade is an object detection algorithm used to locate faces, objects and facial expressions in an image and mainly used for face detection. These feature vectors are given to the neural network as a input which then compares with the pre-trained features. Based on the comparison it will give output as authorized person who is allowed to enter the secured area or unauthorized person classified as intruder against whom appropriate action is triggered by the system.

If he/she is an intruder then buzzer alarm alerts the control room and system initiates action by activating door-locking

system which inturn locks all the exit points of the room so that intruder cannot escape from the confined region.

2. Hardware implementation

When an intruder enters hardware is triggered to function by software(i.e by face recognition). In this system we have used alarm to alert the control room.

The system consists of following hardware components:

a. Servo motor: This is used to control locking system of the door. It helps in opening and closing the door

b. Solenoid latch: The system uses a 12V solenoid latch. This is used to prevent the intruder from escaping. The latch is closed when surveillance camera detects an unauthorized person.

c. Arduino: The system consists of open source arduino hardware used for building the project. Arduino is basically a microcontroller that can be used to control system functions.

d. Relay: This system uses a 5V relay.Relay helps in opening and closing of necessary contacts in the system.

e. Touch sensor: This is an ordinary sensor that helps in opening the door for any person who enters the banking suit.

3. Software implementation

Software system is used to identify the unauthorized person and trigger the hardware to take appropriate action. Now,when system has a clear view that there are limited number of images in some of the classes, the data can be bifurcated further into training, validation and testing datasets inclusive of few basic operations.. For a human to correctly recognize a new face, 50 images are said to be more than enough, whereas for a ideal machine learning, training set is considered a small sample.

Software implementation is carried out in following ways:

3.1. Image pre-processing

3.1.1. Labelled faces in the wild dataset

The selection of a appropriate dataset plays a very important role in the proposed system. The dataset should consist of valid labelled images in each class in order that the neural network can learn every label. Having no restriction to the number of images per class, better results can be obtained through a wide range of images in the LFW dataset. The main purpose of the LFW dataset is to verify whether two images are of the same individual or not as well as facial verification.

Fig 2: Sample of LFW dataset [11]

3.1.2. Elimination of classes which contains less images

Using all of the classes (individuals) would result in a useless model as many of the individuals have only a single image. It is visionary for even the most powerful convolution neural network to learn to identify an individual from a single image. Limiting the data to only 10 individuals with the most images in the dataset, is considered by the proposed system in order to give the model a chance to learn all the classes resulting in 10 classes with at least 50 images per class. . It is indeed considered a feasible idea to avoid using all the images from the LFW dataset. It is quite baffling for a neural network trained on such a dataset where 4096 individuals have only a single images of themselves.

3.1.3. Deep funneling for image transformation

Deep funneling is a technique of image alignment that seeks to reduce intra-class variability in order to allow the model to learn inter class differences which is adapted by the proposed system. Different transformations applied on the images are resulted from different versions of LFW dataset.

3.2. Training using Inception module

Considering the latest and updated Inception V3 model which comprises of the parameters learned through training on the ImageNet dataset, Google's pre-trained Inception Convolution Neural Network is selected to perform image recognition as building and training the CNN is not needed.

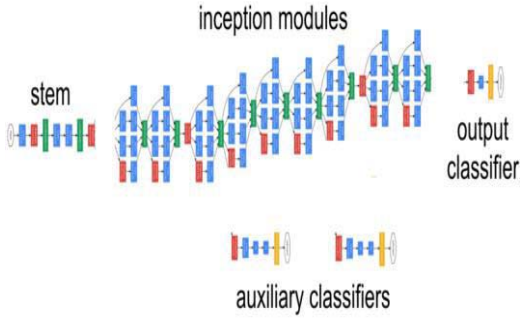


Fig 3: Inception Module [11]

3.2.1. Create sets

In the proposed system 70% of each class in training set, 5% in evaluation set and 25% in a testing set is considered because in order that the network learns the classes a training set is needed, also, to implement early stopping in training a validation set is needed and lastly to evaluate the performance of each model a testing set is required.

3.2.2. Processing the images

Images has to be provided to the ImageNet CNN in the form of arrays [batch_size, image_height, image_width, colour_channels]. The number of images in the training set or testing batch is represented by the batch_size, for Inception V3, the image of 299*299 is required, whereas the number of colour channels for Red-Green-Blue should be 3. For a colour channel, (x,y) denotes the pixel value between 0-255 which represents the intensity of the colour. ImageNet requires normalized pixel values between 0 and 1, which can be obtained by dividing the array by 255. Although there are built-in functions for processing the image to the required size and format,spicy and numpy method can also be used to fulfil the need. The LFW data has images of 255*255*3 which will be further transformed to 299*299*3 along with normalized pixel values between 0 and 1.

3.2.3. Define layer of Inception CNN to train.

In order to classify the objects, or to carry out image related process like facial verification, training of the parameters – connection weights and biases of at least one layer of the network is said to be mandatory. As the lower layers of the convolution neural network are well skilled at identifying lower level features like shapes, colours or textures as well as top layer can differentiate the higher level features like number of appendages or eyes on a human face. It is infeasible to train the entire network on a personal laptop but if the size of the dataset is limited and a Google Cloud GPU or a consumer-grade GPU is used then the last layer of the network can be trained in a reasonable amount of time. Although achievement of record results on the task is doubtful, but the principles involved in adapting an existing model to a new dataset can be visible.

Training of one layer of the network and a defined output layer is required so that the CNN can learn new classes. The parameters used by Inception V3 which is learned from ImageNet for all the layers excluding the one before the predictions hoping that all the weights and biases that are useful in differentiating objects can be applied to the face recognition task. In order to determine the trainable layer, having a view towards the structure of the network is advisable. About 12 million layers are comprised in Inception but all of them are not involved in this process for training.

Early stopping is used for training which is used to reduce over-fitting on the training set. It requires periodic testing of the network on a validation set to check the score of the cost function (average cross entropy). The training is halted if there is no decrease in the loss for specific number of epochs. For retaining the optimal model, whenever there is some improvement in the loss, the model has to be saved. Out of all the models achieved, the one which has the best loss on the validation set is then restored at the end of the training set. New instances are generalized as the model continues to learn the set in a much better way with each epoch, if early stopping method is avoided. Although this method has a variety of implementations, a single validation set and stop training is preferred if there is no improvement in the loss for 20 epochs. A new Tensor Board file is created and then the model parameters are saved to restore the evaluation every time a different model is implemented.

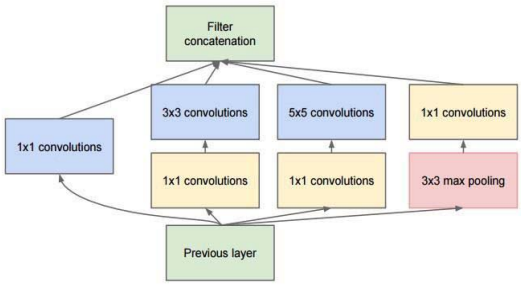


Fig 4: The Building Block of the Inception CNN [11]

3.2.4. Performance measure:

3.2.4.1. Loss function

To measure the performance of a classification model whose output has a probability value between 0 and 1, Cross-entropy loss or log loss is used. It increases if the predicted value diverges form the actual labels. If the actual observation label is 1 and a probability of 0.012 is predicted the it is not assumed to be good and may result in high loss value. A perfect model has log loss log 0.

For binary classification, the Cross-entropy can be calculated as follows [number of classes (M) equals to 2] :

$$-(y\log(p)+(1-y)\log(1-p))$$

For multi-classification [number of classes(M) > 2], a separate loss for class label per observation is carried out followed by summing the results:

$$-\sum c=1 Myo,c\log(po,c)$$

3.2.4.2. Optimization Function:

It is considered as one of the best optimizer till date. With a learning rate of 0.01, the proposed system uses Adam Optimization function as it is able to measure top-1 accuracy. In order to save the model during training and restore it for later, an initialize is used followed by a saver.

3.2.5. Result

After training the neural network, checking for usefulness is the next step. It can be done by carrying out evaluation against the test set, which is completely unseen by the model. The training will b done by going through the test set one batch at a time. 62% of individuals are correctly identified at the first attempt in this proposed system. Not quite impressive for a machine learning system (although it outperforms most humans), it is successful in displaying the fact that the network has learned during training. Even if not satisfies by the performance so far, various steps are available to improve the neural network which avoids building their own CNN.

IV. EXPERIMENT

4.1. Increasing the number of images by augmentation

The system increases the accuracy by expanding the dataset. The two main approaches used here are increasing the amount of training data and augmentation. The system gathers more labelled images to improve the performance of machine learning system. For image, we apply various shifts, that do not change the identifiable features in the image. Likewise, it is applied to the face as well. Other applied transformations include altering the background in the image, changing the contrast and lightning, adding noise, etc. The altered image is then appended to the original image with the correct label.

4.1.1. Shifting and flipping the images

Number of image of a particular class can be increased in two ways i.e. by shifting and flipping the existing images. Image can be shifted in four directions (left, right, up and down).Flipping is done by mapping the image present in left to right to right to left and vice versa.

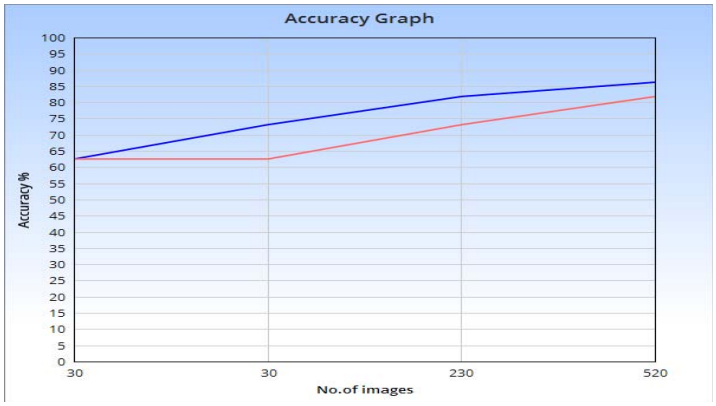
4.1.2. Training on augmented data

The new dataset created by augmentation is used again for training the system in order to increase the accuracy of face recognition.

The table below shows comparison in accuracy percentage obtained before and after augmentation. In this system it is found that in order to increase accuracy shifting, flipping and illumination can be performed on the images of stored dataset. This in turn helps in increasing the classes in dataset, thereby increasing the accuracy. Thus in the graph obtained after augmentation straight line moving continuously upwards is seen in comparison to the linegraph before augmentation.

	No. of images before augmentation	No. of images after Augmentation	Accuracy before augmentation	Accuracy after augmentation
Before Augmentation	30	-	62.5%	-
Shifting	30	230	62.5%	73.1%
Flipping	230	520	73.1%	81.8%
Illumination	520	750	81.8%	86.2%

Table 1: Comparison based on augmentation



Graph 1:Accuracy graph

V. CONCLUSION AND FUTURE WORK

In conclusion, it is found that this system identifies a person with improved accuracy, when compared with previous system from [8] and [10] and this can be used for safeguarding our banking system in an automated manner. In this system we have used buzzer alarm and door locking system for improvising security which results in taking immediate action against intruder. To further improve this system one should train system with more number of samples which will help in improved authentication. In future, with improved dataset one can use it for safeguarding army weapon rooms. The use of LFW datasets can be further improved with use of data augmentation.

REFERENCES

[1] Deepak Aeloor and Amrita A. Manjrekar, ”Securing biometric data and visual cryptography and steganography”

[2] B. Olstad and A H. Torp, “Encoding of a prior information in active contour models [J]”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 18, no. 9, (1996), pp. 863-872.

[3] A K. Jain and Z. Yu, “Deformable template models: A review [J]”, Signal Processing, vol. 71, no. 2, (1998), pp. 109-129.

[4] M A Turk and A P Pentland, “Eigenfaces for recognition [J]”, Journal of Cognitive Neuroscience, vol. 3, no. 1, (1991), pp. 71-86.

[5] J. Lu, K N. Plataniotis and A N. Venetsanopoulos, “Face recognition using LDA-based algorithms [J]”, IEEE Trans. On Neural Networks, vol. 14, no. 1, (2003), pp. 195-200.

[6] F. Samaria, “Face recognition using hidden Markov model [D]”, Cambridge, University of Cambridge, (1994).

[7] S H. Lin, S Y. Kung and L J. Lin, “Face recognition/detection by probabilistic decision based neural network [J]”, IEEE Trans. on Neural Networks, vol. 8, no. 1, (1997), pp. 114-132.

[8] Ya Wang, Ming Zhu, ”Face recognition in real-world surveillance videos with deep learning method”, International conference on image,vision and computing.

[9] Ze Lu, Alex Kot, ”Enhance deep learning performance in face recognition”, International conference on image,vision and computing.

[10] Li Xinhua, Yu Qian, ”Face recognition based on deep neural network”, International journal of signal processing,Vol.8,pp 29-38.

[11] http://medium.com/@williamkoehrsen/facial-recognition-using-googles-convolution-neural-network-aa752b4240e.