Facial Recognition based Attendance System Using CNN and Raspberry Pi

Muhammad Owais
Department of Electrical
Engineering
DHA Suffa University
Karachi, Pakistan
ovaisbabai@hotmail.com

Aireen Amir Jalal
Department of Electrical
Engineering
DHA Suffa University
Karachi, Pakistan
aireenamirjalal98@gmail.com

Muhammad Moiz Hassan
Department of Electrical
Engineering
DHA Suffa University
Karachi, Pakistan
moizhassan@ieee.org

Ammara Shaikh
Department of Electrical
Engineering
DHA Suffa University
Karachi, Pakistan
shaikhammara832@gmail.com

Abstract—This work aims to present face recognition solution using deep learning based facial recognition algorithm. Convolutional Neural Network (CNN) along with triplet loss function has been used to tweak the neural network weights in a way to make the vectors closer via distance metric. The 128-d embedding of each image constitute the feature vectors while K-Nearest Neighbors (KNN) model classifier has been used along with maximum vote count to classify face images. A varying dataset which has been used in this work is the DSU Dataset. The DSU Dataset has been generated locally at DHA Suffa University, Karachi, Pakistan. The implemented algorithm is developed on Python and ensures an overall efficiency of around ninety five percent which is then implemented on Raspberry pi hardware along with an addition of digital attendance management through email.

Keywords—CNN, Convolutional Neural Network, Triplet Loss Function, 128-d embedding, K-Nearest Neighbors, KNN Model Classifier.

I. INTRODUCTION

Each Face has its own unique traits and features that differentiate a person from others. The ability to correctly identify a person based on their characteristics is inherent to humans. Nowadays, multiple applications are being initiated that makes use of these traits in order to speed up the normal day to day procedures. One of the most commonly followed procedures is attendance management; practiced by almost every institute or organization. The process manually practiced on learning management system (lms) in most educational institutes, can be time taking and impose chances of error. With the growing need and desire to automate and speed up processes for the facilitation, saving time and efforts and reducing manual errors, development of these processes is undertaken by incorporating latest technologies.

This paper introduces a system that uses facial recognition technology to record attendance automatically by acquiring images through a high-resolution digital camera. The data acquired is fed to the computer for classification. However, for a machine to be able to identify a person based on their characteristics, it needs to be trained by the use of different algorithms suitable for the purpose. The algorithm defined in the paper, is therefore able to recognize faces by means of comparing the test images; acquired on runtime, with the face images stored in training database and decisions are made using suitable classifiers. Once the test face matches a stored image, attendance is marked.

II. RELATED WORK

Substantial amount of research has been established in the domain of facial recognition and its associated challenges. Beginning with the work of Turk and Pentland in 1991 [1], where emphasis has been laid on the extraction of unique features, termed as eigenfaces through a dimensionality reduction algorithm known as principal component analysis (PCA). Similar approach has been demonstrated in 2019 [2], on locally generated dataset with slight pose and expression variations. However, for the development of attendance system, the limitation of pose, expression and illumination variation needs to be overcome. This has led to the introduction of more diversified algorithms like, Deep face [3] in 2014, which includes several connected layers with multiple parameters, responsible for 3D face alignment and its representation on a commonly known, Labeled faces in the wild dataset (LFW) dataset. Following year in 2015, FaceNet [4] introduced the concept of training the dataset through triplets, generating 128-d embedding and thus improving the classifiers' efficiency. 99.63% efficiency for LFW dataset has been recorded. This concept has then been adopted along with the incorporation of support vector machine; SVM classifier into radio-frequency identification; RFID based attendance system and proposed in related work by Arsenovic [5] in 2017. Later on, other researchers [6] in 2017, worked on the implementation of the idea with a smaller dataset. The systems achieved an efficiency of 95.02%. Work has also been done in the incorporation of the neural network along with other Algorithms to improve face detection process [7]. It has given a boost to the object detection algorithms previously introduced by Viola and Jones [8] in 2004 from 78.4% to 90%.

This work primarily focuses on finding out attendance management [9] solution by incorporating facial recognition technology in real-time, using deep learning

algorithm. Dataset for a group of five people has been gathered and trained using triplet loss function. Triplet loss function allows generation of 128-d Embedding i.e. 128 real valued feature vector that quantifies a face image. The final classification has been made using KNN model and maximum vote count. The KNN model aims on identifying the minimum Euclidean distance, also referred to as L_2 Norm, between the test image and the training image, whereas maximum vote count allows person with maximum number of votes to be identified as the recognized face. The algorithm is finally implemented on Raspberry Pi [10] hardware, as it is cheap, has a built-in DSP module dedicated for image processing and allows project execution in real-time. The final automated attendance data is stored as a Comma-separated value (CSV) file and shared with participants via email using a simple mail transfer protocol (SMTP).

III. FACIAL DATASETS

The proposed work has been carried out using "DSU Dataset" which has been locally generated by students of Electrical Engineering department of DHA Suffa University, Karachi, Pakistan.

The DSU dataset comprises of the data of 05 persons (Fig.1) with 100 images per person totaling up to 500 images. All the images captured are of greater pose and expression variations. In case of testing with individual test images, a ratio of 70:30 has been maintained i.e. 70 images from the dataset of each person have been kept for training and the rest of the 30 images have been saved for testing. For group test images, complete dataset i.e. 100 images of each person are used for training and 39 group images are used for testing. A sample of our DSU Dataset is shown in Fig. 2 for individual images and in Fig. 3 for group test images.

Fig. 1 Selected Person from DHA Suffa University

Fig. 2 Sample Images from the DSU Dataset

Fig. 3 Sample Images from the DSU Dataset for Group Test

IV. EXPERIMENTAL SETUP AND RESULTS

The CNN based algorithm has been applied using Python on DSU dataset during experimentation. The algorithm is first trained and then subjected to test dataset. Python libraries including 'Dlib library' and 'face recognition library' are used on the system. These are used to construct the face embeddings used for the actual recognition process. After the installation of these two libraries, the network is trained using triplet loss function which is already done on the system i.e. the system is pre trained. The idea of Triplet loss function is to tweak the weights of the neural network so that the 128-d Embedding measurements for the two same persons (Anchor and Positive) come closer and farther away from the measurements of the other person (Anchor and Negative) as shown in Fig.4 [4].


Fig.4 Triplet Loss [4]

Once the training is done, 128-d embedding for each face in the dataset is created and these embedding are used to recognize the faces of the characters for both the datasets. When a test image is inferred for testing, the minimum distance or the L_2 norm between the test embedding and the training embedding is calculated. True is returned if the Euclidean distance is less than a certain tolerance indicating a matched face whereas a False is returned for a distance greater than a certain tolerance indicating a rejected face. The experimental result for the DSU dataset is presented separately in the next section.

Once the training has been done and the 128-d embeddings for each face present in the dataset are generated, the next step is face localization, i.e. creation of bounding boxes around the faces using HOG as the face detection algorithm. After that, a KNN model classifier has been used where minimum Euclidean distance between the test and the training encodings is found. By default, faces are considered a match if the Euclidean distance between the face vectors is 0.6 or less. This is controlled by passing a tolerance parameter to the system. A representation for 128-d embedding of a single image can be seen in Fig.5.

A. The DSU Dataset

The experimental results for both individual persons as well as for group photos using DSU dataset have been generated.

1- Individual Test Images Results

For individual test images, 350 images are used for training therefore the returned list will have 350 Boolean values indicating either True or False. A test image has been shown in Fig. 6 as an example. The True values for this particular test image can be seen in Fig.7 and the indexes on which they occur, the character name with the maximum number of votes or True values can be seen in Fig.8. It can be realized that the maximum number of votes were for Moiz, which are 62, Fig.8. Hence the test image is recognized as Moiz, Fig. 9.
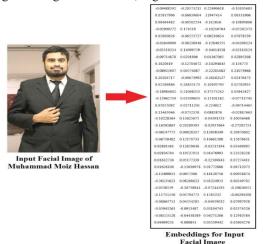
Fig.5 Representation of 128-d embeddings of an Image

Fig.6 Test image of Moiz

Fig.7 Array containing True/False values for the training image

Fig. 8 Array containing indexes of generated True values for the training image and the maximum number of votes for that character

Fig. 9 Recognized Image of Moiz

Next, the results based on different tolerance values have been analyzed. As evident from Fig.10, smaller the tolerance value, the greater the number of false negatives (incorrect rejections) occurs. Higher the tolerance value, the larger the number of false positives (incorrect matches) occurs. It can also be observed that at tolerance value 0.5 and 0.6, an efficiency of around 95% is achieved. At tolerance value 0, none of the images can be recognized, mentioning those images to be unknown thus increasing the count of false negatives, Fig.12. At tolerance value 1, the test images were recognized incorrectly thus increasing the count of false positives (refer Fig.13). The tolerance values 0 and 1 are the two extremes which should never be considered. It has been chosen heuristically here so as to reduce the count of false positives and false negatives.

Threshold	Success	Failure	Efficiency in %
0.7	109	39	72.6667
0.6	142	8	94.6667
0.5	145	5	96.6667
0.4	133	17	88.6667
0.3	75	75	50.000
0.2	13	137	8.6667
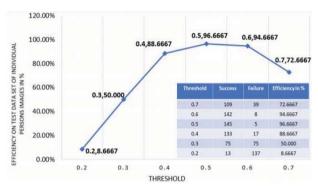
Fig. 10 Effect of increasing the number of Tolerance values on Individual Test images

Fig. 11 Sample Recognition Results at tolerance =0.5

Fig. 12 Recognition Results at tolerance =0 Increased False Negatives

Fig. 13 Recognition Results at tolerance =1 Increased False Positives

2- *Group Test Images Results*

As already mentioned above that the encoding files have been generated for all the faces and based on these embeddings the recognition process occurs. However, the same variation of tolerance values will occur for the group test images but the efficiency will somehow differ from the individual test images. Analyzing Fig.14, it can be concluded that a higher efficiency is achieved at tolerance 0.4. At tolerance values 0 and 1, the count of false positives and false negatives tend to increase respectively, thus affecting the efficiency of the system. Some Recognition sample results for varying tolerance values can be seen in Fig.15, 16, 17 and 18.

Threshold	Success	Failure	Efficiency in %
0.7	85	104	44.9736
0.6	99	90	52.381
0.5	132	57	69.8413
0.4	178	11	94.1799
0.3	123	66	65.0794
0.2	101	88	53.4392
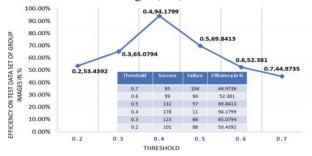
Fig. 14 Effect of increasing the tolerance value on group test images

Fig. 15 Sample Recognition Results at tolerance =0

Fig. 16 Sample Recognition Results at tolerance =0.4

Fig. 17 Sample Recognition Results at tolerance =1

Fig. 18 Sample Recognition Results at tolerance =0.4

V. HARDWARE SETUP

The Hardware setup, as shown in Fig.19, has been implemented using Raspberry Pi 4 having 4 GB RAM with 32 GB SD Card and has been used in real- time. It is a Broadcom BCM2831 Quad core Processor, low-cost device having a clock speed of 1.2 GHz and dedicated DSP hardware called NEON which enables images to be processed with greater speed in real time i.e. the FPS of our pipeline is increased dramatically. Also, the algorithm being implemented on Raspberry Pi allows one-time training of the dataset. Once the dataset is trained, a model file is generated and is saved on the disk, which contains the weights and biases of each image (128-d embedding). Thus, whenever a new image is inferred for test; the process of training the dataset is avoided and thus minimizes computational requirements.
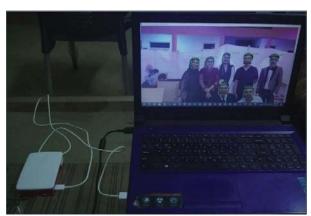

Fig.19 Facial Recognition Hardware Setup

VI. SYSTEM ARCHITECTURE
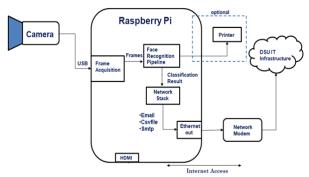
Fig. 20 System Architecture

Fig. 20 illustrates the system Architecture of our Project. The main hardware of our Project is Raspberry pi. The Frame acquisition block acquires the pictures from the camera and passes it on as a frame to face recognition pipeline. The Face recognition pipeline involves the algorithms (CNN) which will be run on Raspberry Pi. The Recognition results have been passed to a network stack which contains email or CSV file to ensure that a student is marked present or absent for a record. The results are then transferred to Ethernet cable which is connected to the local university IT infrastructure.

VII. DIGITAL ATTENDANCE MANAGEMENT THROUGH EMAIL

To keep record of the student attendance on daily basis, CSV file, containing complete student data, has been generated as shown in Fig. 21.

Fig. 21 CSV file

A python program to send emails to students through SMTP has been implemented. Successful implementation result can be seen below:

Fig. 22 Recognition results of Group test image#1 at tolerance =0.4, implemented on Raspberry Pi
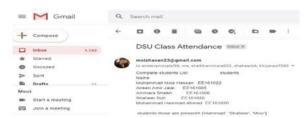

Fig. 23 Received email at tolerance =0.4 for Group test image#1

VIII. CONCLUSION

A successful implementation of deep learning algorithm for real-time attendance system using facial recognition along with KNN model classification and maximum vote count has been established. The DSU dataset with great expression, pose and illumination variation have been used. Results with efficiency of up to 95% have been received against the tolerance value 0.4. Effect of varying the tolerance value on the number of false positives and false negatives has also been evaluated. It has been found that increasing the tolerance value beyond optimum results in false positives and vice versa. The Final implementation has been successfully done on Raspberry Pi along with the addition of digital attendance management through email.

ACKNOWLEDGMENT

REFERENCES

[1] M. Turk, A. Pentland, "Face Recognition using Eigenfaces", Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91, IEEE Computer Society Conference on. IEEE, 1991.

[2] Muhammad Owais, Ammara Shaikh, Aireen Amir Jalal, Muhammad Moiz Hassan, "Human Face Recognition using PCAEigenfaces"6th IEEE international conference on Engineering Technologies and Applied Sciences (ICETAS), 2019.

[3] Yaniv Taigman, Ming Yang, Marc" Aurelio Ranzato, Lior Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification", IEEE conference on Computer Vision and Pattern Recognition,2014.

[4] Florian Schroff, Dmitry Kalenichenko, James Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering" IEEE Computer Society Conference on Computer Vision and Pattern Recognition2015.

[5] Marko Arsenovic, Srdjan Sladojevic, Andras Anderla, Darko Stefanovic, "FaceTime – Deep Learning Based Face Recognition Attendance System" IEEE 15th International Symposium on Intelligent Systems and Informatics • September 14-16, 2017 • Subotica, Serbia.

[6] Nusrat Mubin Ara, Nishikanto Sarkar Simul, Md. Saiful Islam, "Convolutional Neural Network Approach for Vision Based Student Recognition System", 20th International Conference and Information Technology (ICCIT),2017.

[7] Mrs. Monali Nitin Chaudhari, Ms. Gayatri Ramra khiani, "Face Detection using Viola Jones Algorithm and Neural Networks", 978-1-5386-5257-2/18/$31.00 ©2018IEEE.

[8] P Viol, M Jones, "Robust Real-Time Face Detection",Int'lJ. Computer Vision, Vol. 57, no. 2, pp137-154, May2004.

[9] Hari krishnan J, Arya Sudarsan, "Vision-Face Recognition Attendance Monitoring System for Surveillance using Deep Learning Technology and Computer Vision". International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN),2019.

[10] A.S.Hasban, N.A.Hasif, Z.I.Khan, M.F.Husin, N.E.A.Rashid, K.K M. Sharif, N.A. Zakaria, "Face Recognition for Student Attendance using Raspberry Pi. IEEE Asia-Pacific Conference on Applied Elecromagnetics (APACE) 25-27 November 2019 at Malacca, Malaysia.