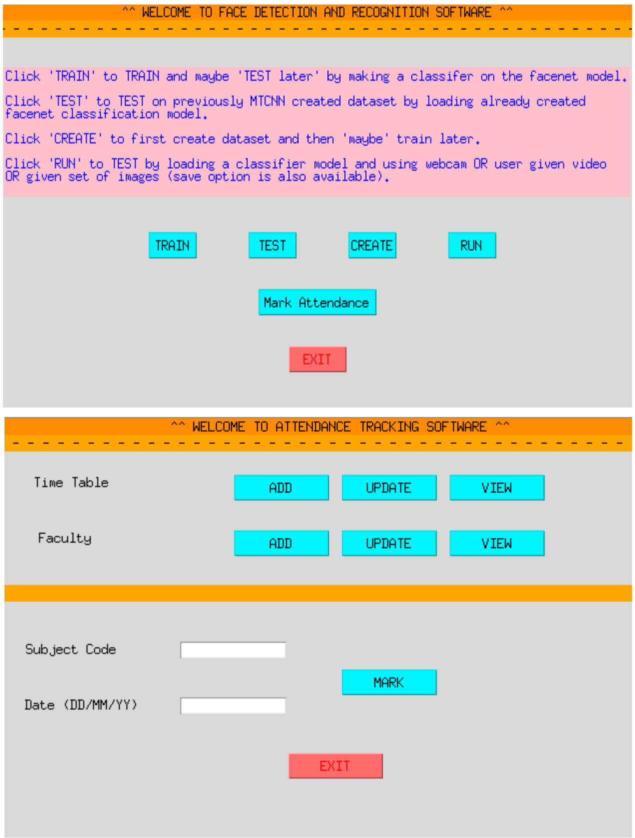
An End-to-End Real-Time Face Identification and Attendance System using Convolutional Neural Networks

Aashish Rai, Rashmi Karnani, Vishal Chudasama, Kishor Upla
Electronics Engineering Department, SVNIT, Surat, India
{aashishrai3799, karnanyr, vishalchudasama2188, kishorupla}@gmail.com

***Abstract*—Carrying out the attendance process in any academic organization is a very significant task. However, the manual attendance process is very tedious and time-consuming. Hence, an automatic face attendance system using CCTV camera may be helpful by reducing the manpower and it also makes the attendance process faultless. There are some automated systems available commercially, but most of them deploy near frontal faces and processes them one by one, which is again a prolonged task. Some deep learning-based face attendance approaches have been proposed in the literature and improving the efficiency of the face attendance is still under research. In this paper, we propose an end-to-end face identification and attendance approach using Convolutional Neural Networks (CNN), which processes the CCTV footage or a video of the class and mark the attendance of the entire class in a single shot. One of the main advantages of the proposed solution is its robustness against usual challenges like occlusion (partially visible/covered faces), orientation, alignment and luminescence of the classroom. The proposed method obtained a real-time accuracy of 96.02% which is better than that of the existing end-to-end face attendance systems.**

I. INTRODUCTION

Artificial Intelligence is a technology that is profoundly changing the world and also significantly improving the state-of-the-art in many applications including healthcare [1], security [2] and marketing [3]. People prefer to use intelligent systems instead of sluggish traditional methods. But still, attendance is being carried out manually, which indeed is quite slow and erroneous. This paper describes a smart approach to mark the attendance of the entire class in one shot. It is not required for every student to go in front of the camera and wait for their face to get recognized. This paper introduces a stand-alone software with its own interactive user interface, as shown in Fig. 1, to make the attendance procedure seamless and accurate. The software allows the user to create, train and/or test the dataset effortlessly. The software process the video/CCTV footage and detect the faces using state-of-the-art face detection method of Multi-Task Cascade Convolutional Neural Network (MTCNN) [4], recognize these detected faces by using the FaceNet module [5] and mark the attendance of all the recognized students. The attendance is saved in a spreadsheet corresponding to the entered subject and date. The proposed approach, imroves the performance of the entire end-to-end attendance system. The efficiency of proposed approach is examined in experimental results section.



Click 'TRAIN' to TRAIN and maybe 'TEST later' by making a classifer on the facenet model.

Click 'TEST' to TEST on previously MTCNN created dataset by loading already created facenet classification model.

Click 'CREATE' to first create dataset and then 'maybe' train later.

Click 'RUN' to TEST by loading a classifier model and using webcam OR user given video OR given set of images (save option is also available).

Fig. 1: User Interface.

II. LITERATURE REVIEW

In order to make end-to-end face attendance management system, in 2013 Chintalapati et al. [6] proposed a method to carryout the attendance system using the Viola Johns method [7] for face detection, followed by histogram equalization for feature extraction and SVM classifier for face recognition. Later in 2017, Rathod et al. [8] proposed an end-to-end face attendance system using same technique for face detection and classification for the face recognition. However, these methods were based on classical machine learning based algorithms. In 2017, Arsenovic et al. [9] proposed the state-of-the-art method called FaceTime using CNN cascade for face detection and CNN for generating face embeddings which were then used for face recognition. In this paper, the MTCNN framework [4] has been deployed for face detection and the prevalent FaceNet module [5] is used to extract high-quality features from the face and predict a 128 element vector to represent those features, which is called face embedding. Further discussion on proposed method is done under Section III and IV.

978-1-7281-2327-1/19/$31.00 ©2019 IEEE

III. End-to-end Proposed Approach

In this section, the three main tasks of proposed end-to-end approach are discussed: creating and training the dataset, working of recognizer and the attendance manipulator. The first and the foremost step of training is to have a proper dataset. To create dataset, the user has to take a 5-10 seconds video of all the students of a class and save them in a folder with their respective names. The user needs to specify the path to the dataset folder and run the software. The software loads the video successively and segment it into frames. Once the segmentation is over, the software starts processing the frames successively, as shown in Fig. 2.

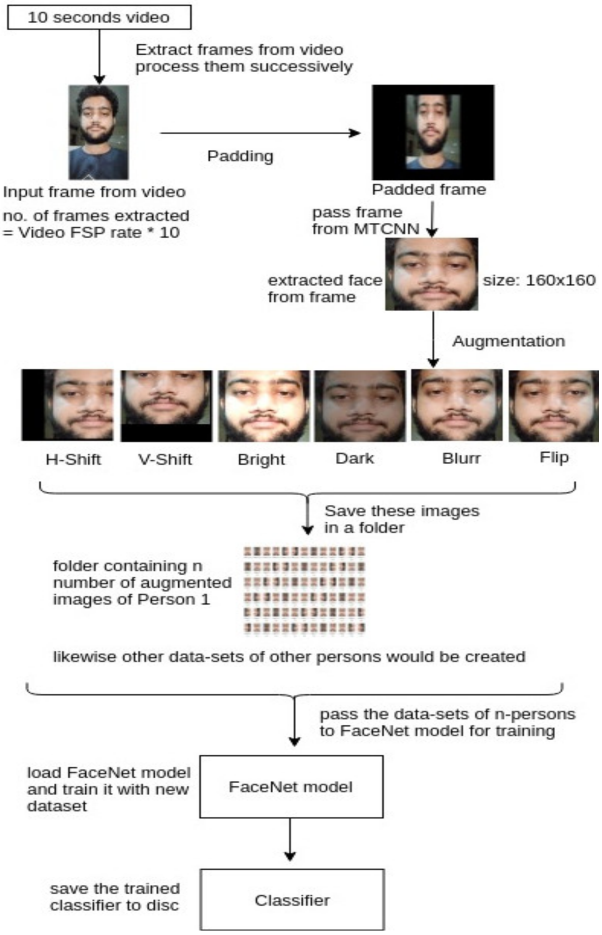

10 seconds video
Extract frames from video process them successively
Padding
Input frame from video
Padded frame
no. of frames extracted = Video FSP rate * 10
pass frame from MTCNN
extracted face from frame
size: 160x160
Augmentation
H-Shift
V-Shift
Bright
Dark
Blurr
Flip
Save these images in a folder
folder containing n number of augmented images of Person 1
likewise other data-sets of other persons would be created
pass the data-sets of n-persons to FaceNet model for training
load FaceNet model and train it with new dataset
FaceNet model
save the trained classifier to disc
Classifier

Fig. 2: Working of Trainer.

The processing includes padding the frame from all sides with black color and resizing it into proper dimensions. It is then passed through MTCNN [4], which gives the coordinates of the bounding box enclosing the face. With the help of the coordinates, the face is cropped from the frame and is passed for augmentation, which includes flipping, blurring, adding Gaussian noise, salt & pepper noise, horizontal shifting, vertical shifting, increasing and decreasing the brightness. The augmented faces are then resized to 160×160 and saved in a folder corresponding to the name of the video. Likewise, the other frames of the video are processed. Once the video ends, the software loads another video from dataset folder and iterates the same process. Once the dataset is created, another significant task is to train it. The software facilitates the user to assign the percentage for the dataset to be used for training, the rest will be used for testing the trained classifier. To train the FaceNet model [5], at first, it needs to be loaded, and re-trained. In this approach, the pre-trained model 'Inception ResNetV1' architecture [10] which is trained on 'VGGFace2' dataset is used. After loading the pre-trained model, it is re-trained on the generated dataset. The model is trained until its accuracy is found to be greater than 99%. After training, it is saved and later used for the recognition task.

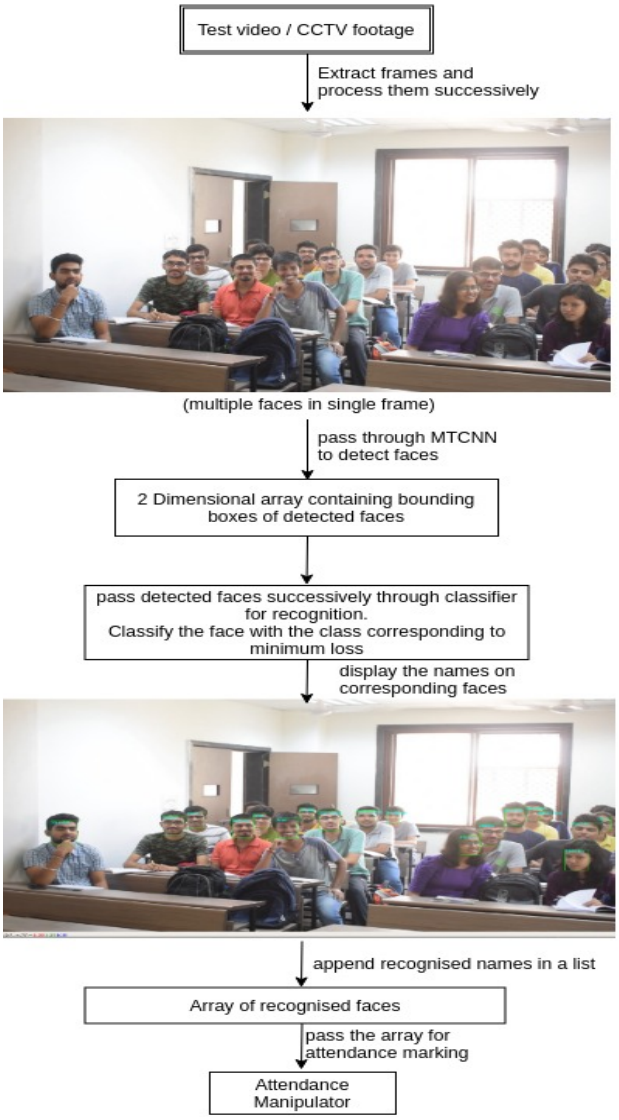

Test video / CCTV footage
Extract frames and process them successively
(multiple faces in single frame)
pass through MTCNN to detect faces
2 Dimensional array containing bounding boxes of detected faces
pass detected faces successively through classifier for recognition.
Classify the face with the class corresponding to minimum loss
display the names on corresponding faces
append recognised names in a list
Array of recognised faces
pass the array for attendance marking
Attendance Manipulator

Fig. 3: Working of Recognizer.

First, the video is taken from the source and is segmented into frames. The frame is then passed through MTCNN, which returns a 2-D array containing the coordinates of the bounding boxes of all the detected faces as shown in Fig. 3. The faces are then extracted from the frame and after being resized, they are processed for recognition using the trained classifier. The confidence scores corresponding to all the classes are recorded and the class with maximum score is assigned to the face. Similarly, all the faces are classified on the basis of their respective scores. The processed frame is then displayed with the names corresponding to every face. All the recognized names are stored in a list and likewise other frames are processed. The names recognized in the other frames are appended to the list. The list containing the names of the students is then passed for attendance. The attendance manipulator creates subjectwise attendance sheets and feed the names of all the registered students in them. Once this module gets the list containing the names of the recognized students, it extracts unique names from it. The manipulator then loads the spreadsheet corresponding to the entered subject in append mode and creates a new column for the entered date, and marks the attendance.

IV. *EXPERIMENTAL RESULTS*

In this section, we discuss the testing and validation of the proposed end-to-end approach. We created the dataset of 18,000 images of 18 different students, that is around 1000 images per person. The performance of the proposed approach is compared with the existing face attendance approaches. In our proposed approach, we choose a pre-trained model of MTCNN framework [4] for face detection and the FaceNet module [5] is used for face recognition. The generated dataset was trained on GeForce GTX 1070 GPU. For testing, we used 1900×900, 25 FPS, Video Codec H.264 (High Profile), 5 seconds video. The software was able to recognize 17/18 people accurately on a normally illuminated video. The real-time performance of the proposed system is mentioned in Table 1 in terms of benchmark evaluation metrics. The proposed end-to-end approach achieves 96.02% accuracy in real-time face attendance. In Table 2, the real-time accuracy of the proposed approach along with existing end-to-end approaches [6, 9] are mentioned. However, we have not compared our result with that of the papers [8, 11] as they have not mentioned the end-to-end accuracy of their proposed methods. Here, one can observe that the proposed approach outperforms the other existing end-to-end face attendance approaches.

Table 1: Measurement evaluation metrics of system.

Accuracy	96.02%
Recall/Sensitivity	99.61%
Precision	96.26%
F1 Score	97.90%

Table 2: Performance comparison with other methods.

Sr. no	Method	Accuracy
1.	Chintalapati et al. [6]	95.00%
2.	FaceTime [9]	95.02%
3.	Proposed Approach	96.02%

We tested the proposed approach by varying different parameters and adding different noises to the video. We altered the various parameters of the video that outweighed the practical scenarios. Such measurement shows the performance of the proposed approach in different real-time scenarios. The effect of these various factors on the proposed approach in terms of their accuracy is mentioned in Table 3.

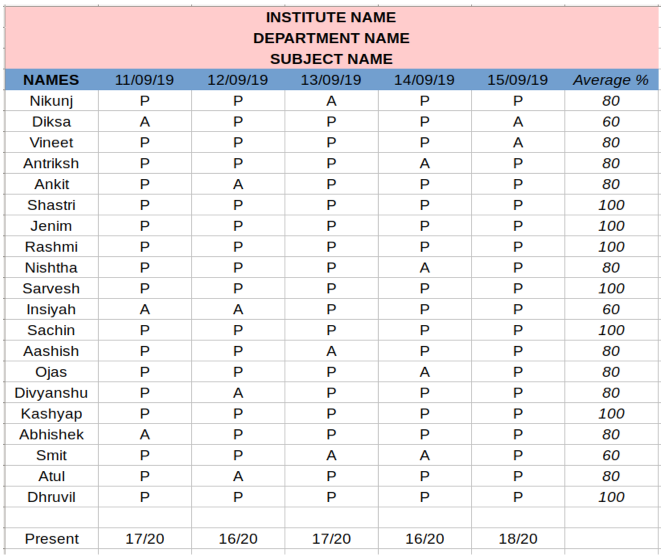
Table 3: Effect on accuracy on various factors.

Sr. no	Factor	Accuracy
1.	Normal Brightness	96.02%
2.	High Brightness	92.45%
3.	Low Brightness	95.17%
4.	Hue	89.54%
5.	Saturation	90.10%
6.	Gaussian Noise	93.77%
7.	Poisson Noise	93.45%
8.	Salt & Pepper Noise	91.67%

In order to validate the effectiveness of zero padding and data augmentation, we further test the proposed approach with and without zero padding and augmentation. The quantitative comparison on these conditions are mentioned in Table 4. From Table 4, one can observe that padding the video frame with black color helps to detect all face images from the CCTV video frames. Also, one can conclude that the data augmentation increase the diversity of data for training models and helps to improve the accuracy of the end-to-end system.

Table 4: Effect of Padding and Augmentation.

Sr. no	Attribute	Result
1.	Without padding	Face detected = 500/600
2.	With padding	Face detected = 600/600
3.	Without augmentation	Accuracy = 38.88%
4.	With augmentation	Accuracy = 96.02%



INSTITUTE NAME
DEPARTMENT NAME
SUBJECT NAME

NAMES	11/09/19	12/09/19	13/09/19	14/09/19	15/09/19	*Average %*
Nikunj	P	P	A	P	P	*80*
Diksa	A	P	P	P	A	*60*
Vineet	P	P	P	P	A	*80*
Antriksh	P	P	P	A	P	*80*
Ankit	P	A	P	P	P	*80*
Shastri	P	P	P	P	P	*100*
Jenim	P	P	P	P	P	*100*
Rashmi	P	P	P	P	P	*100*
Nishtha	P	P	P	A	P	*80*
Sarvesh	P	P	P	P	P	*100*
Insiyah	A	A	P	P	P	*60*
Sachin	P	P	P	P	P	*100*
Aashish	P	P	A	P	P	*80*
Ojas	P	P	P	A	P	*80*
Divyanshu	P	A	P	P	P	*80*
Kashyap	P	P	P	P	P	*100*
Abhishek	A	P	P	P	P	*80*
Smit	P	P	A	A	P	*60*
Atul	P	A	P	P	P	*80*
Dhruvil	P	P	P	P	P	*100*
Present	17/20	16/20	17/20	16/20	18/20	

Fig. 4: Sample attendance sheet.

Fig. 4 shows the sample attendance sheet of a particular subject for five different days along with other essential information. It shows the respective attendance of particular student on different dates along with its overall attendance information.

V. CONCLUSION

In this paper, we proposed an end-to-end approach using CNN to carry out the attendance process accurately and seamlessly. The proposed system achieved appreciable results when tested in a real classroom environment. The system was able to overcome the usual challenges of occlusion, alignment, orientation, and luminescence. Experimental results demonstrate that the proposed system is able to achieve appreciable results in practical environments, and hence it can be implemented at various places to carryout attendance process. The limitation of the proposed system is that it gets perplexed for distant faces and also for low resolution videos. The MTCNN architecture is impotent for faces with their eyes closed which makes this system restricted to faces with eyes open analogous to camera. The future work may include detecting and recognizing faces irrespective of resolution of the video coverage by preprocessing it using a super-resolution module.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

REFERENCES

[1]. Pandit, Hardik, and Dr M. Shah. "Application of Digital Image Processing and analysis in healthcare based on Medical Palmistry." In International Conference on Intelligent Systems and Data Processing (ICISD), pp. 56-59. 2011.

[2]. Soutar, Colin, Danny Roberge, Alex Stoianov, Rene Gilroy, and Bhagavatula Vijaya Kumar. "Biometric encryption using image processing." In Optical Security and Counterfeit Deterrence Techniques II, vol. 3314, pp. 178-188. International Society for Optics and Photonics, 1998.

[3]. Snyder, Mark, and Kenneth G. DeBono. "Appeals to image and claims about quality: Understanding the psychology of advertising." Journal of personality and Social Psychology 49, no. 3 (1985): 586.

[4]. Zhang, Kaipeng, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. "Joint face detection and alignment using multitask cascaded convolutional networks." IEEE Signal Processing Letters 23, no. 10 (2016): 1499-1503.

[5]. Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 815-823. 2015.

[6]. Chintalapati, Shireesha, and M. V. Raghunadh. "Automated attendance management system based on face recognition algorithms." In 2013 IEEE International Conference on Computational Intelligence and Computing Research, pp. 1-5. IEEE, 2013.

[7]. Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57, no. 2 (2004): 137-154.

[8]. Rathod, Hemantkumar, Yudhisthir Ware, Snehal Sane, Suresh Raulo, Vishal Pakhare, and Imdad A. Rizvi. "Automated attendance system using machine learning approach." In 2017 International Conference on Nascent Technologies in Engineering (ICNTE), pp. 1-5. IEEE, 2017.

[9]. Arsenovic, Marko, Srdjan Sladojevic, Andras Anderla, and Darko Stefanovic. "FaceTime—Deep learning based face recognition attendance system." In 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY), pp. 000053-000058. IEEE, 2017.

[10]. Szegedy, Christian, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. "Going deeper with convolutions." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1-9. 2015.

[11]. Fu, Rong, Dan Wang, Dongxing Li, and Zuying Luo. "University classroom attendance based on deep learning." In 2017 10th International Conference on Intelligent Computation Technology and Automation (ICICTA), pp. 128-131. IEEE, 2017.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE