Computer Vision Base Attendance System

Denesh Kumar Manoharan,
The Pennsylvania State University
dzm5964@psu.edu

***Abstract*—Artificial Intelligence can be trained over several billion images to predict and classify necessary information. Some of the major applications of this methods includes face detection and face recognition which can be used for marking attendance in universities or organization and can be used for surveillance of facility or workplace against intruders from gaining access. In this project I have created a vision bases attendance system which detects and recognize the people presence. A deep learning architecture, multi-task cascaded neural network (MTCNN) is used to perform the person detection and ArcFace machine learning model is used for face recognition to mark attendance.**

Keywords—Attendance system, Surveillance, Face recognition, Face detection, deep learning, Machine learning

I. Introduction

A key factor of improving the quality of education is having students attend classes regularly. Traditionally students are stimulated to attend classes using attendance points which at the end of a semester constitute a part of a student's final grade. However, traditionally this presents additional effort from the teacher, who must make sure to correctly mark attending students, which at the same time wastes a considerable amount of time from the teaching process. Furthermore, it can get much more complicated if one must deal with large groups of students. Every organization requires a robust and stable system to record the attendance of their student. All organization has their own method, taking attendance manually by rollcall and some have adopted biometrics system such as fingerprint, RFID card reader, Iris system to mark the attendance. The conventional method of calling the names of students manually is time consuming event. The RFID card system, each student assigns a card with their corresponding identity but there is chance of card loss or unauthorized person misuse the card for fake attendance. While in other biometrics such as fingerprint, iris, or voice recognition, they all have their own flaws and not 100% accurate. This paper introduces a new automatic attendance management marking system, without any interference with the regular teaching process.

Use of face recognition for the purpose of attendance marking is the smart way of attendance management system. Face recognition is more accurate and faster technique among other techniques and reduces chance of proxy attendance. Face recognition provide passive identification that is a person who is to be identified does not need to take any action for their attendance. The system can be used also during exam sessions or other teaching activities where attendance is obligatory. This system eliminates classical student identification such as calling student names, or checking respective identification cards, which can not only interfere with the teaching process, but also can be stressful for students during exam sessions. This system only requires a good quality camera to capture the images of students, The images capture by the camera is normalized and sent to system for recognizing and mark attendance. The model faces challenges as image obtained from real time over head camera are mostly occluded, rotated, or zoomed out faces, which requires image correction data pre-processing. The face recognition must contend with uncontrolled lighting conditions, large pose variation, face expression, makeup, change in face hair, ageing and partial occlusion. The literature survey part gives a brief review of all the previous works on face recognition-based attendance systems. The methods, features, and limitations in practical implementation have been mentioned.

II. RELATED WORK

Currently most of the management system implements attendance management system by Moodle [1] automates the process by using RFID or barcode scanner. Classrooms are equipped with a barcode / RFID scanner which scans and enrolls students entering classroom. Also, other management system uses LMS system such that require students to login into a webpage with a special one-time temporary key to mark their attendance. All these interfere with teaching process.

Other management systems used biometrics [2] (fingerprint recognition, iris recognition) to identify students however installing this system in every classroom in university is a bigger financial burden. It would also require all the students to submit their

.

biometric data to university which raises privacy concern. These systems are also subjected to physical damage from user. Therefore, they need additional maintenance costs.

Using Face Recognition system for attendance is not new, In [3], author has proposed a model of student attendance management using face recognition with an additional feature of percentage analyzer. The proposed system used Artificial Neural Network (ANN) with Principal Component Analysis (PCA) for face recognition. In this model, maximum varied face points were identified and feature vectors at those points were created. Also, the covariance column matrix was created using PCA for creating facial features. They used the extracted features for computation of eigenfaces, which served as an input to ANN for training. The ANN once trained could identify faces with a threshold. If the value of the result was less than the threshold, it discarded the result.

In [4], the authors proposed a prototype of a model for classroom attendance using a deep learning model. In the proposed model, they used a fixed camera in the classroom for image acquisition. Open-source machine learning library, Torch was used for building the detection system. They performed the face detection using Convolutional Neural Network (CNN) cascade of twelve CNNs. They trained deep CNN using deep learning based OpenFace library with a pretrained FaceNet face recognition system's network. They performed classification using a Linear Support Vector Machine (SVM). They tested it on images having less than 20 faces and had an accuracy of 92.02%.

Developing this attendance system is divided in two tasks, the face detection and face recognition. This task is implemented using deep learning and machine learning algorithm. Face recognition is implemented using Naimish Net algorithm [5], Haar Cascaded Neural Network [6] and the Multi-task Cascaded Neural Network etc. Naimish Net is involved predicting coordinates of the facial key points (FKP) such as nose tip, center of eye, etc. for each given face. This uses LeNet adapted deep CNN model. This network has good accuracy of classification, but this network is complex and becomes slower when implemented in real time. This algorithm leads to lower accuracy of our model. Haar cascade neural network is designed for face detection. This algorithm has less accuracy since it only identifies the matching shape and size of eye, nose, lips using cascade window. It tries to calculate feature for every window and classify positive and negative.

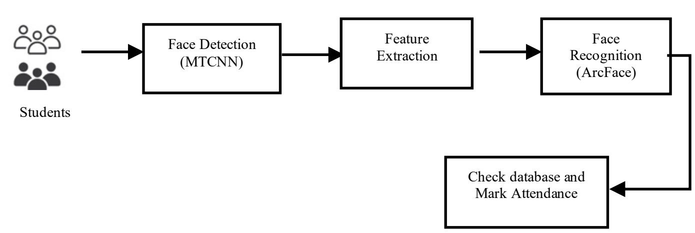
The CovNets or deep CovNets have shown good results for face verification. Many algorithms have been implemented, like the use of neural networks [7,8], geometrical features, Eigen faces [9], template matching [10], and graph matching [11]. CovNets has shown many promising results for FR. Automatic feature extraction method using ratios of distances, by Kanade [12] used geometrical features and reported a recognition rate between 45-75% with a database of 20 people.

The FaceNet algorithm [13] is used for face recognition. The advantage of this algorithm is the different lightings on input image does not affect the accuracy and can also recognize faces with markings. This algorithm also detects the person even if face is partially cropped. The disadvantage of this algorithm is it does not recognize the person if wearing eyeglasses, this algorithm is also biases to certain race and skin colors.

The MobileNet-v2 is used for Face recognition, this involves the Depth wise Separable Convolution structure into three sequential layers and integrated two new features: inverted residual connections, and linear bottlenecks. This algorithm is less computationally complex and achieves good classification rate only when face to recognize is not occluded and with good lighting conditions. The images with low light intensity are misclassified.

III. System Implementation

To implement the system a Face detection system which identifies and extracts the face of students entering the classroom. This extracted face image of the student is the region of interest which further will be input to face recognition system and will be compared to our database and mark the presence of students. The video of students entering the classroom will be converted as digitalized image using OpenCV library in python and given as input to face detection algorithm and the extracted region of interest will be used in face recognition system. To implement this multi-task convolution neural network (MTCNN) and Arc Face Machine learning model is used.



A. Multi-task Convolution neural network (MTCNN):

MTCNN is a multitask neural network model for face detection. It uses small model to generate target region candidate box with certain possibility and uses more complex model for fine classification and higher precision region box regression and makes this step recursive to form a three-layer network, p-net, R-Net, o-net, to achieve fast and efficient face detection. In the input, image pyramid is used to transform the scale of the initial image, and p-net is used to generate a large number of candidate target area frames. After that, R-Net is used for the first selection and border regression of these target area frames, and most of the negative examples are excluded. Then, the more complex and higher precision network o-net is used to classify and regress the remaining target area frames.

MTCNN performs the image resizing operation to scale the original image to different scales to generate image pyramid. Then the images of different scales are sent to the three sub networks for training to detect different sizes of human faces and realize multi -scale target detection.

P-net is the first subnet in MTCNN. It is known as proposal network. The structure of the algorithm is a full connected convolution. The image pyramid constructed in the previous step of the algorithm is used to extract the preliminary features and calibrate the frame through a FCN, and roughly obtain the face candidate frame and frame regression vector. Then the candidate frames are regressed by the frame, and finally the candidate frames with high coincidence are merged by NMS algorithm. Most of the windows are filtered through the bounding box regression adjustment window and NMS.

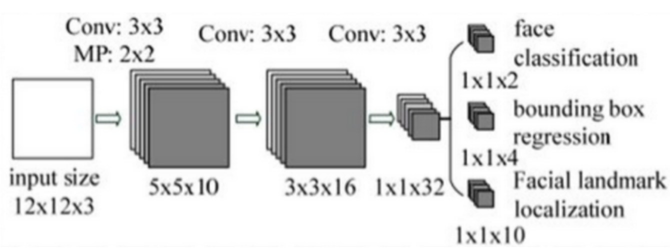


Figure 1: P-net MTCNN

The R-Net is Refine-network, in Figure 2. Its structure is more complex than the p-net network structure of the upper layer. The constraint conditions are mainly added, and the face prediction frame is screened again by the added constraints.

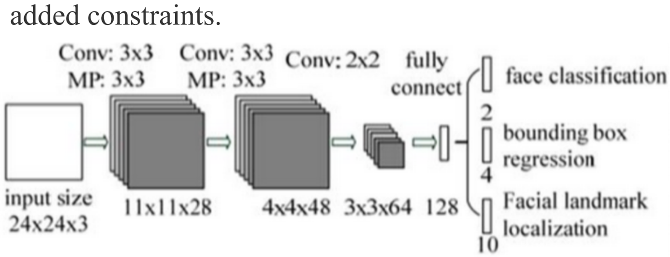


Figure 2: R-net MTCNN

The R-Net network makes further decision on the output window of the upper layer and uses border regression and NMS algorithm to discard the face candidate frames with low score, to select several groups of locally optimal face candidate frames. It can be seen from Figure 2 that the R-Net network has one more full connection layer than the p-net network in the end. The function of the full connection layer is to output a 128 dimensions vector. Because of the full connection classification, R -Net will further filter the prediction box.

The third layer of network is o-net, is also known as output network shown in figure 3. The function of o-net is to select the best candidate frame and output the five feature key points detected. The o-net is deeper than of its previous layers.

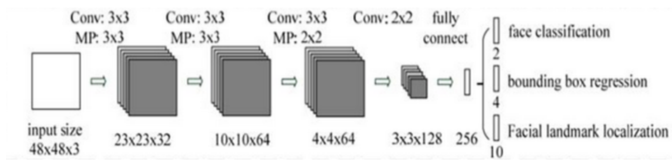


Figure 3: O-net MTCNN

To train the MTCNN model VGGFace-2 data set is used, this dataset contains 3.31 million images of 9131 subjects with an average of 362 images per subject. Data set annotation is divided into four categories, if IOU < 0.3 is negative sample, if IOU > 0.65 is positive sample, IOU is 0.4 - 0.65 is part of the face, IOU is 0.3 - 0.4 is an unclear area.

Two different learning rates, 0.001 and 0.0001, were used to train the model. The learning rate is 0.005, and the learning rate is reduced once every 100 backward propagations. From the results analyzed when the larger learning rate of 0.001 is used, the model iteration is faster and the training time is shorter, when using 0.0001 as the basic learning rate, the iteration time of the model is delayed, but the final effect of the model is ideal. Large batch can let the model see more samples in the same iteration, which can make learning more stable and achieve better results. But with the increase of batches, the mean value of the overall sample noise remains unchanged, but the variance decreases, and the sample noise helps the optimizer to avoid the local optimum and improve the overall generalization ability; in this training, the batch size is set to 32. Training the model with VGGFace-2 dataset accuracy of 98.77% is achieved.

Figure 4: Accuracy and Loss MTCNN

B. Additive Angular Margin Loss:

Additive Angular Margin Loss is also known as ArcFace. It is a loss function used in face recognition. The SoftMax generally used in these tasks. However, the SoftMax loss function does not perfectly optimize the feature embedding to enforce higher similarity for intraclass samples and diversity for inter-class samples, which results in low performance for deep face recognition. The ArcFace loss transforms the logits $W_j^T x_i = \left|\left|W_j\right|\right| \, ||x_i|| \cos\theta_j$, where θ_j is the angle between the weight W_j and the feature x_i. The individual weight W_j||=1 is fixed by l2 normalization. The individual weight $||W_j||$=1 is fixed by l_2 normalization. The embedding feature $||x_i||$ is fixed by l_2 normalization and re-scaled to s. The normalization step on features and weights makes the predictions only depend on the angle between the feature and the weight. The learned embedding features are thus distributed on a hypersphere with a radius of s. Finally, an additive angular margin penalty m is added between x_i and W_{y_i} to simultaneously enhance the intra-class compactness and inter-class discrepancy.

$$L_3 = -\frac{1}{N}\sum_{i=1}^{N} log \frac{e^{s(\cos(\theta+m))}}{e^{s(\cos(\theta+m))+\sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}}}$$

The ArcFace views each Face as a class and it represents each class as a smooth vector which makes it efficient and stable during training, however in certain cases due to camera pose, The largest intra-class distance between samples is illumination, blur and low resolution can significantly change intra-class distance. larger than the largest image-to-class distance, and the smallest inter-class distance between samples is smaller than the smallest image-to-class distance. Therefore, we can go back to image-to-image comparison method to further improve the ArcFace model. So, to further improve the ArcFace model the intra loss and inter loss with the ArcFace function. This inter and intra losses helps in enhance the discriminative power of the deeply learned features. Since these losses help in improving the discriminative power, they have a good classification rate and produce good accuracy, So In order to evaluate the improvement in the classification accuracy we performed this model with different dataset with different learning rate.

Table 1: Different ArcFace with different learning rate

Methods	1e-06	1e-05	1e-04	1e-03	1e-02	1e-01
ArcFace	19.45	21.16	83.33	91.76	95.11	97.52
ArcFace + Intra	18.67	19.77	85.94	92.46	95.66	97.86
ArcFace + Inter	22.29	23.91	84.52	94.70	97.60	98.15
ArcFace + Intra + Inter	18.93	20.52	86.54	94.31	97.96	99.32

From table1 we can clearly see that the different variations of the Arc Face algorithm are compared with different learning rates. These algorithms were tested on the VGG Face 2 dataset, and the highest accuracy was obtained for the algorithm ArcFace + Inter Loss + Intra loss and when learning rate is set to e^-01 , the accuracy obtained is 99.32% and ROC of 0.99.

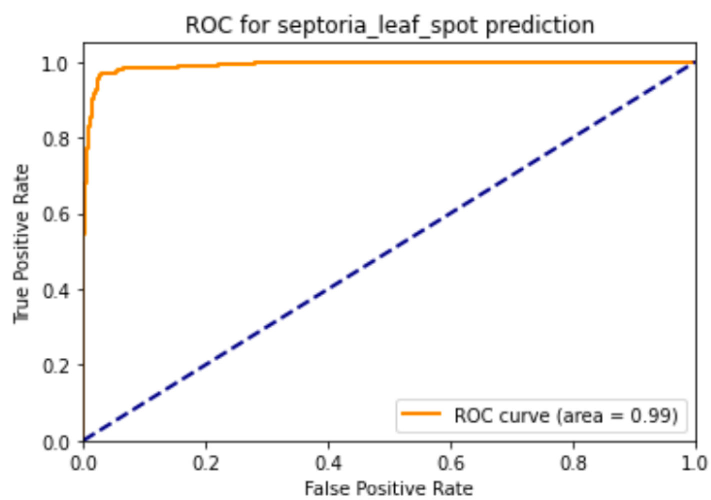


Figure 5: Accuracy and Loss ArcFace

IV. Compare Techniques and Approaches

Different approaches and their combinations have been used in face recognition-based attendance systems to boost performance. Various methods of face detection such as Haar Cascaded Neural Network, Naimish Net algorithm, LeNet adapted deep CNN model is used. The comparison of all the techniques and the drawbacks is mentioned in table 2.

Table 2: Comparison of Face detection algorithm

S.No	Reference	Methods	Accuracy	Drawback
1	[9]	Eigenface algorithm, PCA and ANN	80%	High computational cost due to combining PCA and ANN
2	[14]	LBP, HOG and SVM	92%	Light sensitivity issues because of LBP
3	[13]	CNN, FaceNet and SVM	95.40%	High computational cost due to combining SVM and CNN
4	[4]	Viola-Jones and CNN	80%	Initial performance on lookalike faces was poor because of the self-learning feature
5	[6]	Haar cascade and LBPH	98.20%	Light sensitivity issues because of LBPH, and Accuracy of 98.20% at 24 lx light and distance of 40 cm
6	[15]	Viola-Jones, PCA, LDA and LBP	90%	Light sensitivity issues and overall model not cost-effective
7	[16]	KNN and LBP	95%	Low accuracy in low lighting as pixels change the value with light

As we can see from the table the algorithms have accuracy in the range of 80% to 98.20%. The eigenface algorithm and viola jones algorithm have the lowest accuracy of 80 precent. The CNN and Haar cascade have achieved an accuracy of 95.40% and 98.20 percent respectively but Haar cascade system has light sensitive issue, the algorithm fails to detect the face if the image is in low light intensity. Similarly in CNN with SVM model the computation cost was high as it was combination two model. The accuracy of the MTCNN system implemented has more accuracy compared to any of previously used system for face detection.

Similarly various face recognition algorithm has been implemented to improve the accuracy of face recognition, the comparison of algorithm is presented in table 3.

Table 3: comparison of face recognition algorithm

S.No	Methods	Accuracy	Drawbacks
1	Eigenface and PCA	90%	Sensitive to variation in lighting, pose, and expression
2	Fisherface	95%	Sensitive to light and pose change
3	LBP	85%	Sensitive to change in illumination and face expression
4	CNN	98%	Requires large amount of data and computation power
5	Siamese network	96%	Computationally complex and take more time to compute

The local binary pattern algorithm is lowest accuracy, it works by comparing the intensity values of a central pixel with its surrounding pixels in a circular or rectangular neighborhood, generating a binary pattern, and converting it into a decimal number. This process is repeated for all pixels in the image, this produces a new image where each pixel has been replaced by its corresponding LBP value. So this algorithm is highly sensitive to light intensity and so achieved low accuracy rate. The CNN has achieved accuracy of 98% but this requires large amount of data to be trained and needs more computational power. The ArcFace algorithm we have used has achieved higher accuracy and is not computationally expensive like CNN models.

V. Conclusion

In conclusion, the development of a vision-based attendance system has the potential to revolutionize attendance management systems in schools, colleges, and workplaces. The system offers numerous advantages over traditional methods, including increased accuracy, efficiency, and convenience. With the use of advanced technologies such as facial recognition, machine learning, and computer vision, the system can accurately and quickly identify individuals and record their attendance in real-time. Furthermore, the system can be easily integrated with existing systems, reducing the need for additional hardware and software. Overall, the vision-based attendance system is an innovative solution that can help institutions manage attendance more effectively, leading to increased productivity and better student and employee outcomes.

VI. Reference

[1] Behera, Sourish, and Rajesh Kumar Kushwaha. "RFID based people management system using UHF tags." *Proceedings of ASCNT, CDAC, Noida, India* (2009): 231-238.

[2] Ujan, Imran Anwar, and Imdad Ali Ismaili. "Biometric attendance system." In *The 2011 IEEE/ICME International Conference on Complex Medical Engineering*, pp. 499-501. IEEE, 2011.

[3] Sunaryono, Dwi, Joko Siswantoro, and Radityo Anggoro. "An android based course attendance system using face recognition." *Journal of King Saud University-*

Computer and Information Sciences 33, no. 3 (2021): 304-312.

[4] Hartanto, Rudy, and Marcus Nurtiantoro Adji. "Face recognition for attendance system detection." In *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, pp. 376-381. IEEE, 2018.

[5] Agarwal, Naimish, Artus Krohn-Grimberghe, and Ranjana Vyas. "Facial key points detection using deep convolutional neural network-NaimishNet." *arXiv preprint arXiv:1710.00977*(2017).

[6] Padilla, Rafael, C. F. F. Costa Filho, and M. G. F. Costa. "Evaluation of haar cascade classifiers designed for face detection." *World Academy of Science, Engineering and Technology* 64 (2012): 362-365.

[7] Li, Haoxiang, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua. "A convolutional neural network cascade for face detection." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5325-5334. 2015.

[8] Guo, Guanjun, Hanzi Wang, Yan Yan, Jin Zheng, and Bo Li. "A fast face detection method via convolutional neural network." *Neurocomputing* 395 (2020): 128-137.

[9] Wagh, Priyanka, Roshani Thakare, Jagruti Chaudhari, and Shweta Patil. "Attendance system based on face recognition using eigen face and PCA algorithms." In *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, pp. 303-308. IEEE, 2015.

[10] Jin, Zhong, Zhen Lou, Jingyu Yang, and Quansen Sun. "Face detection using template matching and skin-color information." *Neurocomputing* 70, no. 4-6 (2007): 794-800.

[11] Pan, Zhiqun, Yongxiong Wang, and Sunjie Zhang. "Joint face detection and Facial Landmark Localization using graph match and pseudo label." *Signal Processing: Image Communication*102 (2022): 116587.

[12] Tian, Y-I., Takeo Kanade, and Jeffrey F. Cohn. "Recognizing action units for facial expression analysis." *IEEE Transactions on pattern analysis and machine intelligence* 23, no. 2 (2001): 97-115.

[13] Nyein, Thida, and Aung Nway Oo. "University classroom attendance system using facenet and support vector machine." In *2019 International conference on advanced information technologies (ICAIT)*, pp. 171-176. IEEE, 2019.

[14] Mehta, Preeti, and Pankaj Tomar. "An efficient attendance management sytem based on face recognition using Matlab and Raspberry Pi 2." *International Journal of Engineering Technology Science and Research IJETSR* 3, no. 5 (2016): 71-78.

[15] Mothwa, Louis, Jules-Raymond Tapamo, and Temitope Mapati. "Conceptual model of the smart attendance monitoring system using computer vision." In *2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, pp. 229-234. IEEE, 2018.

[16] Salim, Omar Abdul Rhman, Rashidah Funke Olanrewaju, and Wasiu Adebayo Balogun. "Class attendance management system using face recognition." In *2018 7th International conference on computer and communication engineering (ICCCE)*, pp. 93-98. IEEE, 2018.