Vision Based Attendance System

M.Denesh Kumar, S.Aravind Raamasamy
{dzm5964,afs6349}@psu.edu

1. Task

In this project, an automatic people detection and counting system using data collected from an over-head camera is proposed. The purpose of this research is to develop a fast and accurate intelligent people counting technique for attendance monitoring systems in offices and lecture rooms. The proposed method includes two stages working sequentially. First, the face detection task is executed to find any person presented in the current frame. A deep learning architecture, Multi-task Cascaded Neural Network (MTCNN) [5][6], is used to carry out the detecting phase. If there is any detected person, the face detection will be initialized, The ArcFace [1], Additive angular margin loss for deep face recognition is used to obtain highly discriminative features for face recognition. Based on this we can built an effective attendance marking system.

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.

2. Related Work

This project is divided into 2 tasks, first is face detection and the next step is face recognition. For face detection there were many algorithms that worked well like the Naimish Net [7], Haar Cascaded Neural Network and the Multi-task Cascaded Neural Network. Among these algorithms after extensive literature survey found out that the MTCNN algorithm is very fast in real time and the accuracy is also nearly perfect. Now coming to the Face Recognition algorithm, found out that Arc Face, FaceNet [2] and FaceNet2ExpNet [3] performs really well. Among these Arc Face is a state of the art algorithm and it outperforms the other face recognition algorithms. Also from literature survey we got a good idea of the dataset that needs to be used to train these algorithms.

3. Approach

In the initial phase of the project, we will be creating an GUI interface using Tkinter library in python. This interface allows the user to register themselves for attendance [8] . During the registration we will be collecting the sample image by requesting the user to move his face around the camera which will be later used for face recognition. In the second phase of our project, we will be implementing MTCNN model using the VGGFace2 dataset. Based on the literature survey we figured out MTCNN model works better with VGGFace2 dataset. If we have time in our project we will also try implementing RetinaFace model with VGGFace2 dataset.

In the third phase we will be implementing the ArcFace loss function for face recognition and compare it with our data to mark attendance. The architecture of ArcFace can be seen in fig 2 , it uses an angular marginal loss function which is much better than the softmax loss function.

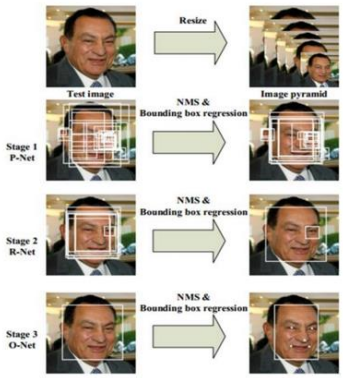

Figure 1. MTCNN

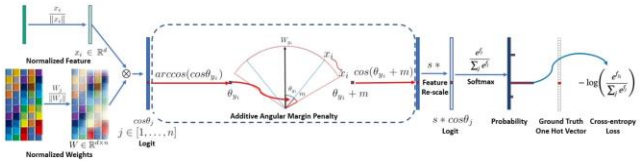

Figure 2. ArcFace

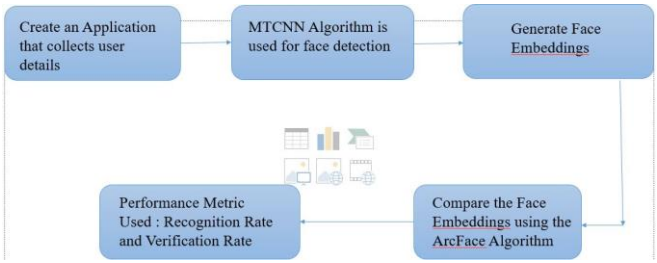

Figure 3 : Approach Flowchart

4. Dataset and Metric

For the face detection algorithm, we will be using VGGFace2 [4]. The dataset contains 3.31 million images of 9131 subjects, with an average of 362.6 images for each subject. Images are downloaded from Google Image Search and have large variations in pose, age, illumination, ethnicity. The dataset was collected with three goals in mind: (i) to have both a large number of
identities and also a large number of images for each identity; (ii) to cover a large range of pose, age and ethnicity; and (iii) to minimize the label noise. The dataset can be seen in figure 4.

Figure 4 3.31 million images divided into 9131 classes, each representing a different person identity

Figure 5. Faces with high degree of variability

We also use WIDER FACE data set. The dataset contains rich annotations, including occlusions, poses, event categories, and face bounding boxes. Faces in the proposed dataset are extremely challenging due to large variations in scale, pose and occlusion, as shown in Fig. 5. Furthermore, we show that WIDER FACE dataset is an effective training source for face detection. The link below is the source for our dataset.

- https://www.kaggle.com/mksaad/wider-face-a-face-detection-benchmark
- https://www.robots.ox.ac.uk/~vgg/data/vgg_face/

Metrics:
In the first phase of the project, the accuracy of face detection model using MTCNN is measured by Recognition rate. And in the second phase, face detection by using ArcFace, the SoftMax loss function does not explicitly optimize the feature embedding to enforce higher similarity for intra- class samples and diversity for inter-class samples, which results in a performance gap for deep face recognition under large intra-class appearance variations so we will be trying to improve the accuracy and analyze the efficiency using confusion matrix. In this project we use two performance metric that is the Recognition Rate and the Verification rate, these both are similar to precision and recall. So our aim is to reduce it as much as possible so that our working prototype is robust.

5. Approximate Timeline
Make a plan with approximate deadlines, e.g.

Task	**Deadline**
GUI Interface for Registration	03/15/2022
MTCNN Model Implementation	04/02/2022
Arc Face Model Implementation	04/15/2022
Prepare report and presentation	04/25/2022

References

1) Jiankang Deng, Jia Guo, Niannan Xue, Stefanos Zafeiriou, “ArcFace: Additive Angular Margin Loss for Deep Face Recognition” , CVPR 2019.
2) Florian Schroff, Dmitry Kalenichenko, James Philbin , “FaceNet: A Unified Embedding for Face Recognition and Clustering”, CVPR 2015.
3) Hui Ding, Shaohua Kevin Zhou, Rama Chellappa,“FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition”, arXiv.org 2015.
4) Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, Andrew Zisserman, “VGGFace2: A dataset for recognising faces across pose and age” , arXiv.org 2017.
5) Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, Yu Qiao ,“Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks” , arXiv.org 2016.
6) Jinhyung Park, Xinshuo Weng, Yunze Man, Kris Kitani ,“Multi-Modality Task Cascade for 3D Object Detection” , arXiv.org 2021.
7) Naimish Agarwal, Artus Krohn-Grimberghe, Ranjana Vyas,“Facial Key Points Detection using Deep Convolutional Neural Network – NaimishNet” , arXiv.org 2017.
8) Haldar, R., Chatterjee, R., Sanyal, D.K., Mallick, P.K.: Deep learning based smart attendance monitoring system. In: Proceedings of the Global AI Congress 2019. Advances in Intelligent Systems and Computing, vol. 1112. Springer, Singapore (2019).