LITERATURE REVIEW

Introduction

Attendance is an important of classroom evaluation. In every class, attendance is usually checked by teacher but sometimes teacher may miss someone. The traditional attendance system consists of every student signing on a sheet of paper against their names where the sheet is passed around to each student. The faculty should have proper method to verify and maintain the students record especially in classes attended by large number of students. The concept of face recognition-based attendance system is to give a computer system the ability of finding and recognizing human faces fast and precisely in images or videos.

Academic Literature

Siamese Network:

Ref**:** http://www.cs.utoronto.ca/~gkoch/files/msc-thesis.pdf

Paper 1: Siamese Neural Networks for One-Shot Image Recognition by Gregory Koch

A siamese neural network consists of twin networks which accept distinct inputs but are joined by an energy function at the top. This function computes some metric between the highest-level feature representation on each side. The parameters between the twin networks are tied. The output of this network aids in simplified image classification.

This paper demonstrates how a Siamese network is:

a) capable of learning generic image features useful for making predictions about unknown class distributions even when very few examples from these new distributions are available;

b) easily trained using standard optimization techniques on pairs sampled from the source data;

c) able to provide a competitive approach that does not rely upon domain-specific knowledge by instead exploiting deep learning techniques.

This paper compares the performance of the Siamese network to an existing state-of-the-art classifier developed for an Omniglog data set.

This strategy has two key properties:

• It ensures the consistency of its predictions. Weight tying guarantees that two extremely similar images could not possibly be mapped by their respective networks to very different locations in feature space because each network computes the same function.

• The network is symmetric: if we present two distinct images to the twin networks, the top conjoining layer will compute the same metric as if we were to we present the same two images but to the opposite twins.

FaceNet:

Ref**:** https://arxiv.org/pdf/1503.03832.pdf

Paper 2: FaceNet: A Unified Embedding for Face Recognition and Clustering.

Background:

The Google Face Detection Algorithm, published in CVPR 2015, uses the same face to have high cohesion under different angles and different poses, and different faces have low coupling. It is proposed to use the cnn + triplet mining method on the LFW dataset. The accuracy is 99.63%, and the accuracy is 95.12% on the youtube face dataset, which is nearly 30% higher than the previous accuracy.

Method:

The CNN maps the face to the feature vector of the European space, calculates the distance of the face features of different pictures, and the distance of the faces of the same individual is always smaller than the a priori knowledge training network of the faces of different individuals.
Only the face features need to be calculated during the test, and then the distance threshold can be calculated to determine whether the two face photos belong to the same individual.

Implementation:

1. The feature vector is extracted from the image using the CNN network.
   There is a dimension selection problem of feature vectors. The faster the dimension is smaller, but it is difficult to distinguish different pictures when it is too small. The larger the dimension, the easier it is to distinguish different pictures, but the too large training model is not easy to converge, and the calculation is slow and the space is large. The authors have shown that the 128-dimensional feature can better balance this problem.
   The selection of CNN models, high-precision models often have many parameters and large calculations. It is best to use a model with a small size and a slightly lower precision on the mobile device; a high-precision, high-computation model can be used on the server.
2. Use the ternary loss function.
   Some people used the previous work to use the binary loss function. The goal of the binary loss function is to map the face features of the same individual to the same point in space, and the ternary loss function target is mapped to the same area, so that the class the distance is less than the distance between classes.
3. You can see that the above formula requires three input portraits. How to choose this one-tuple training? In order to ensure the speed of training convergence, we choose the same portrait that is the farthest distance, and train with the different portraits from the nearest. The author then makes the above selection in each mini-batch.
   Convolutional network. Choosing a model is a common problem, and the author did different tests for ZF/GoogLeNet.
   Final verification. By calculating the distance of the feature vectors of different pictures, the results are obtained using the threshold.

Triplet loss:

The triplet loss is the loss function used to train network like a siamese network. I triplet loss would help the Siamese network learn the parameters to understand the level of similarity or dissimilarity between two images. In triplet loss, we use three different images 1. Anchor images, a positive image which is of the same person as the anchor image and a negative image. The function is given below: Where **p** is positive image, **a** is the anchor image and **n** is the negative image and **α** is the margin. We need to minimize this loss function in order to converge the network and make it learn the similarity function.

$$Loss = \sum_{i=1}^{N} \left[\| f_i^a - f_i^p \|_2^2 - \| f_i^a - f_i^n \|_2^2 + \alpha \right]_+$$

Summary:

Extracting features directly calculates the distance, which is simpler than the previous PCA + SVM. The training loss function directly targets the actual error, and the end-to-end method can improve the accuracy. As long as there is a bounding box for the face of a person.

We made some studies and got the basics of the DeepFace and FaceNet:
Ref**:** https://www.cs.toronto.edu/~ranzato/publications/taigman_cvpr14.pdf

DeepFace is a deep learning facial recognition system created by a research group at Facebook. It identifies human faces in digital images. It employs a nine-layer neural net with over 120 million connection weights, and was trained on four million images uploaded by Facebook users.

FaceNet is a Deep Learning architecture consisting of convolutional layers based on GoogLeNet inspired inception models. FaceNet returns a 128 dimensional vector embedding for each face. This architecture is trained over the large dataset of faces from different people.

Industry Solutions:

Some of the organizations are working and have developed the attendance system using the face recognition. One of them is, Matrix Security Solutions: They built a Powerful Biometric User Identification based on Innovative, Deep Learning Technology. It accepts the user's mobile as a credential and evolves as per regular user interaction in different conditions. The technology checks liveness of a face with 99.53% accuracy. It is a facial recognition attendance system that ensures contactless authentication and identifies user's face in < 1 Sec.

Tool:

TensorFlow: We will be using tensor flow libraries to implement the network as tensor flow is an open source easy to use framework and has a great community support. Tensor flow has been around from 2017. This tool has thus evolved over time and has really helped developers to build their deep learning solutions effectively in short span of time.

References:

1. Siamese Neural Networks for One-Shot Image Recognition by Gregory Koch - http://www.cs.utoronto.ca/~gkoch/files/msc-thesis.pdf
2. FaceNet: A Unified Embedding for Face Recognition and Clustering - https://arxiv.org/pdf/1503.03832.pdf
3. DeepFace: Closing the Gap to Human-Level Performance in Face Verification - https://www.cs.toronto.edu/~ranzato/publications/taigman_cvpr14.pdf
4. https://www.tensorflow.org
5. Siamese Neural Networks for One-shot Image Recognition - https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf
6. https://www.matrixaccesscontrol.com/facerecognition.html & https://www.innovatrics.com/face-recognition-solutions/
7. https://www.youtube.com/watch?v=wr4rx0Spihs
8. Triplet loss function - https://towardsdatascience.com/lossless-triplet-loss-7e932f990b24
9. Siamese network and triplet loss - https://towardsdatascience.com/siamese-network-triplet-loss-b4ca82c1aec8
10. One shot learning - https://hackernoon.com/one-shot-learning-with-siamese-networks-in-pytorch-8ddaab10340e
11. http://llcao.net/cu-deeplearning17/pp/class10_FaceNet.pdf