FACE RECOGNITION ATTENDANCE SYSTEM FRAMEWORK

Cos30082: Applied Machine Learning

MAY 26, 2025

NAME: HONG ANH NGUYEN
STUDENT ID: 104359382

1. Introduction

In this project, we introduce an end-to-end face recognition system which is a crucial technique for wide range of applications, for instance, security, smartphones face id… We focus on building a face recognition system with liveness detection for anti-spoofing and emotion detection. The system using CNN (convolutional neural networks) for this face recognition task. This report discusses methodologies and results of the models, including training schemes, evaluation metrics and performance comparisons.

2. Methodology

The dataset has approximately 380k image of faces with a 64x64 size, considering the small size of image, ResNet-18 model was implemented as based model for embedding for 2 methods (self-supervised learning or metric learning using cosine similarity and Euclidean) and Classification based (supervised learning). The Resnet-18 model was replicated same as original paper, but we plugged a classification head with 512 dim to output the face embedding vectors. In this project, metric learning (cosine similarity and Euclidean) using Triplet loss and classification-based learning using categorical entropy loss were implemented.

Method 1: Classification-based learning

Using original embedding model of ResNet-18, a FC (fully connected) layer with 1000 dim was implemented following by batch norm layer and ReLU activation to present the vector of face embedding, and a one hot layer mapping to 4000 nodes layer (4000 people id)

Training Scheme

A 64x64 dataset of human faces were used to train the CNN, it was divided into batch size of 16 for efficiency training, it was resized, decode image to uint8 tensor and applied ResNet preprocessing layer before training.

The model was trained with Adam optimizer, learning rate 1e-4 with sparse categorical cross-entropy loss function on 10 epochs.

Method 2: Metric learning

Siamese Model was used to learn the similarity between pairs of face images, we used 2 common metrics which are: Euclidean distance – calculating the sum of squared differences between the coordinates of 2 vectors, Cosine distance – 1 minus cosine of the angle between 2 vectors. Since the cosine range from [-1,1], 1 - cosine will range from [0,2].

Training Scheme

The same dataset was used but we constructed the set of anchors and positive were chosen by sliding a sliding window with size of 2 and stride of 2, ensuring that all images are learned, with each of this window, a random negative image is random chosen from other class for diversity. The idea that model should maximizing the distance between the anchor and negative while minimize the distance between the anchor and positive. The used margin value was 0.5 for Euclidean and 1 for Cosine.

The loss function then return max of (anchor-positive distance minus anchor-negative distance + margin) and 0. The models were compiled same with Adam optimizer and learning rate 1e-4 but only with 5 epochs due to time complexity and computational cost.

Anti-Spoofing

A robust Eye blink detection based on Eye Landmarks taking idea from [1], the eye detector will calculate horizontal and vertical Euclidean distance of eyes, form a formular to calculate EAR (eyes aspect ratio) and if it toggles the threshold mean they blink [2]. This approach is computational reasonable compared to other light/contrast or texture analysis approaches. The eye facial landmark is constructed as below:

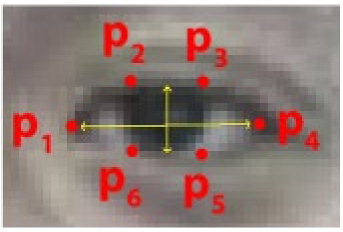

Figure 1. Mark points

Then the EAR is calculated by:

$$\mathrm{EAR} = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|},$$

When the p2-p6 + p3-p5 small showing that they closed eye and otherwise. To anti spoofing, whenever a new face come in, we require user to show 2 states of closing and opening eyes, then it will be considered as real face which have liveness.

Emotion detection

https://www.kaggle.com/code/abhisheksingh016/emotion-detection-jupyter-notebook

Pre-trained model was downloaded from this source due to the time constraint; the model was trained with **FER2013 (Facial Expression Recognition 2013) dataset.** Model able to detect 7 types of emotion, 'Angry', 'Disgusted', 'Fear', 'Happy', 'Sad', 'Surprise', 'Neutral'. A simple pipeline of load model, making prediction real-time was implemented but the result was not good since the confidence of emotion always return the highest is Sad regardless of face's emotion.

Conclusion and Discussion

Metrics:

For the Siamese models, Triplet loss with corresponding metric was used as loss and AUC was used to assess the face recognition performance while the classification model was used sparse categorical cross-entropy as loss and Top 1-Accuracy used as performance metric.

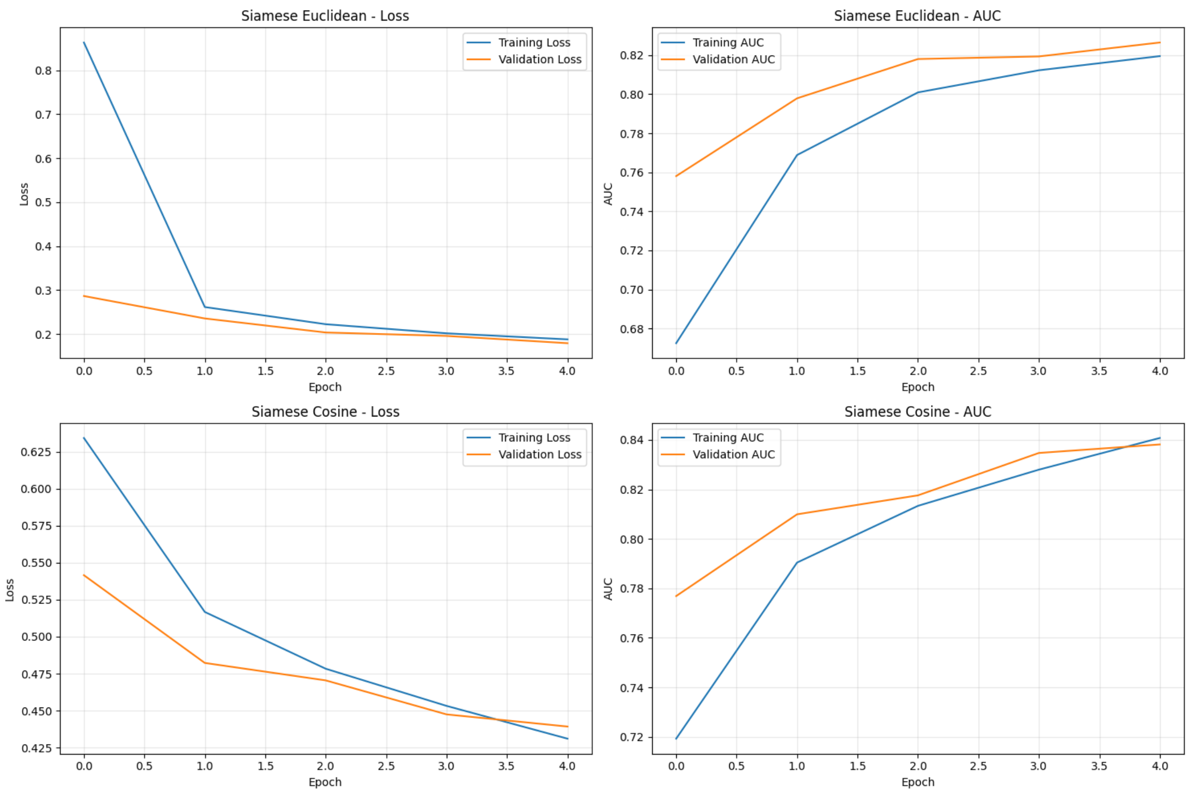

Figure 2. Siamese Model Training History

Both Euclidean and Cosine showed good learning capabilities, but the Euclidean model is slightly better. Euclidean model showed potential of improving since the AUC and triplet loss of training are not surpassed validation yet while the Cosine model had some signal of overfitting. But due to the time constrains and computational limitations, these models were only trained with 5 epochs.

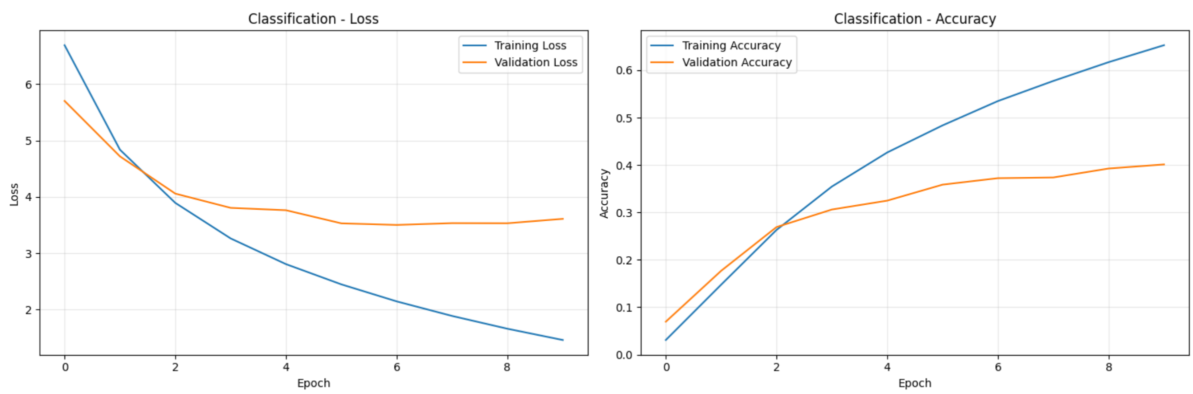

Figure 3. Classification model Training History

The classification model observed a clear overfitting issue at epoch 6-10 when the training accuracy keeps significantly increased while the validation accuracy slightly increased, because of that only 10 epochs was trained with this model.

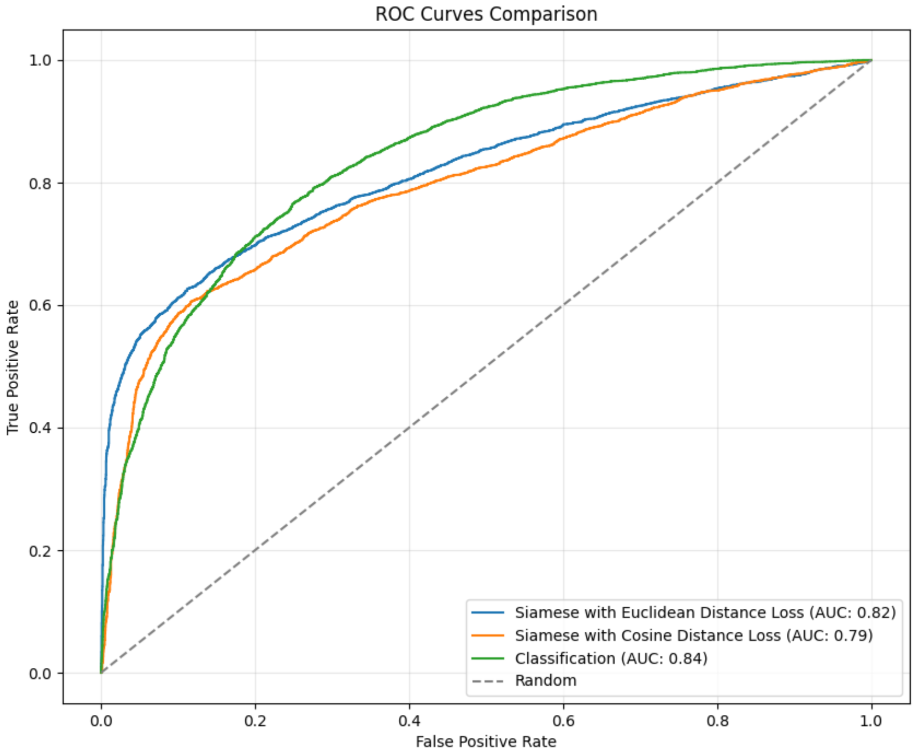

Figure 4. Models Performance Comparison

While the AUC point out that the Euclidean is better than Cosine which is reasonable considering the training history, but the classification perform the best AUC. But when inference on GUI to test with user face, the classification model perform bad result when it can't tell face difference, most of the faces are same while the Euclidean can achieve certain performance and tell the difference between face as well as recognize the user face.

Reference

1. Al-gawwam, S., & Benaissa, M. (2018). Robust Eye Blink Detection Based on Eye Landmarks and Savitzky–Golay Filtering. *Information*, *9*(4), 93. https://doi.org/10.3390/info9040093

2. Soukupová, T. (2016). *Real-Time Eye Blink Detection using Facial Landmarks*. https://vision.fe.uni-lj.si/cvww2016/proceedings/papers/05.pdf