CS 335 Project Report

Balasubramanian, Karthikeya, Shubh, Teja

November 24, 2022

Contents

1 Problem Statement

Face Verification: To train a deep neural network model based on CNN architecture to verify if two different images(human faces) are of the same identity/person using state of the art ArcFace loss function.

2 Abstract

One of the main challenges in feature learning using Deep Convolutional Neural Networks (DCNNs) for large scale face recognition is the design of appropriate loss functions that enhance discriminative power (across classes/identities).

From ...s paper we can infer that the features learned by softmax loss have intrinsic angular distribution and thus the motivation for our loss function and also simultaneously countering the drawbacks of softmax-loss-based methods and triplet-loss-based methods as noted in ...a paper. The choice of ArcFace loss function over SphereFace(...s paper) loss function is that the former performs emperically better(...a paper).

3 Theory

3.1 Dataset

Preprocessed LFW dataset of images(*112 ∗ 112 ∗ 3*).

3.2 Model

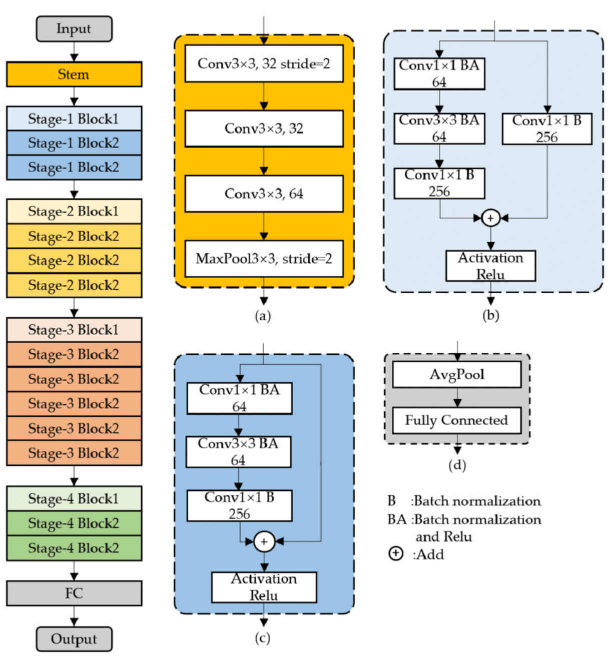

Figure 1: Resnet50's Architecture

Resnet50 and Resnet34 are used as backbones to produce embedded representations of the images.

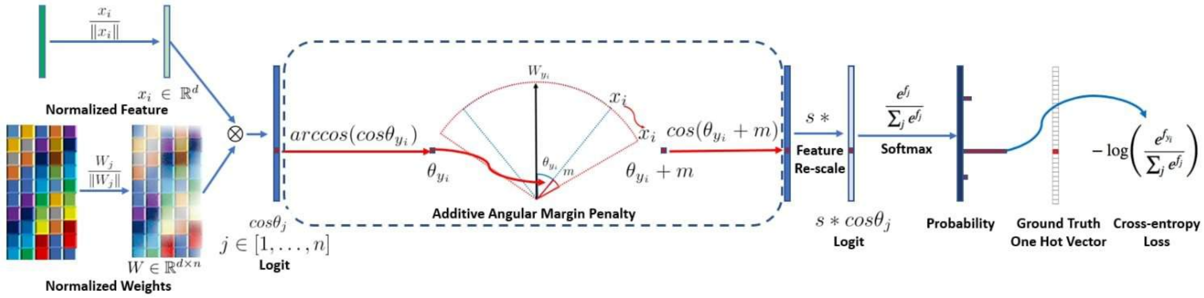

Figure 2: Final layer, prediction and loss evaluation

The output(normalized) of Resnet is passed through a Fully connected layer(weight's normalized) whose output is the cosine of the angle between a column of the weight vector(central class vector of an identity) and resnet's ouput referred henceforth as *the cosine distance*.

A vector containing the cosine distances for each column of the weight vector and resnet's output is passed through a Softmax layer to obtain the multiclass probability prediction.

3.3 Loss function

3.3.1 Naive loss function

$$L_2 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s\cos\theta_{y_i}}}{e^{s\cos\theta_{y_i}} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}}.$$

Figure 3: Cross entropy loss

Here N is the batch size, n is the number of classes(identities), s is a scaling factor for the dot product and y_i is the ground truth for the ith image.

The problem with cross entropy loss is it does not enforce extra intra-class compactness and inter-class discrepancy.

3.3.2 Arcface loss function

$$L_3 = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}}. \tag{3}$$

Figure 4: Arcface loss

Here m(hyper parameter) is the *angular margin*. Note that m is added only to θ_{y_i}. This loss function enforces extra intra-class compactness and inter-class discrepancy are discussed below.

3.4 Angular Margin

Consider the simple case of binary classification(only two classes)

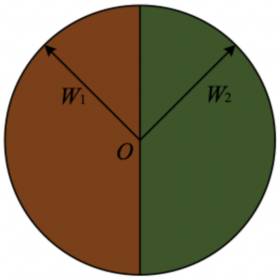

Figure 5: Binary decision boundary

The loss function penalises such that the decision boundary for the classes are separated like in support vector machines because of the misclassification error if any embedded vector lies outside the decision boundary of the two classes.

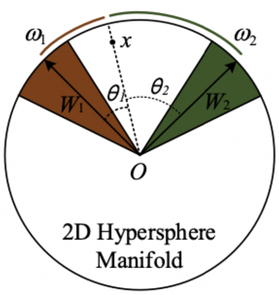

Figure 6: Separation of decision boundaries

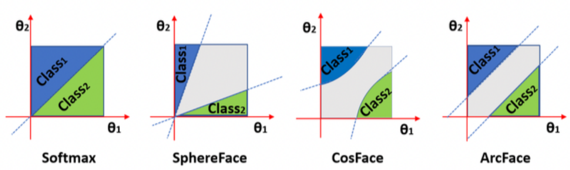

Figure 7: Separation of decision boundaries denoted in theta space

The above example captures the essence of how Arcface loss function achieves extra intra-class compactness and inter-class discrepancy.

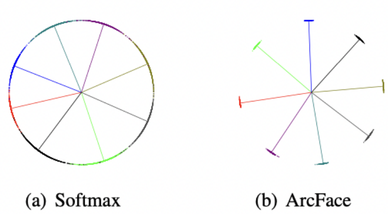

Figure 8: Multiclass examples under the softmax and ArcFace loss