Deep Learning Assignment

Class Project

Submitted in partial fulfillment of the requirements of
CS F425 Deep Learning

By

Mayank Jain
ID No. 2019A7PS0141P

Under the supervision of:

Dr. Kamlesh Tiwari

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, PILANI CAMPUS
March 2022

0.1 Methodology

0.1.1 Data collection

The first step was collecting the data. Every person in the class uploaded images of 4 different actors in 6 different movies, separated by 4-6 years each. In each movie, 25 images were to be taken, giving 150 images per person. These images were then stored in a database, and after some cleaning and ROI cropping, a subset of the data was provided to us. A total of 510 images were provided. These images were of 15 different actors.

0.1.2 Preprocessing

Following the methods used in face recognition papers like DeepTeeth, images were resized to 75x75 for uniformity and faster processing. The images were then subjected to greyscale conversion and image enhancement.

0.1.2.1 Image Enhancement

I tested 2 famous techniques -

1. Histogram Equalization - It's purpose is to equalize the contrast of a given grayscale image and normalize the brightness.
2. CLAHE - It stands for Contrast Limited Adaptive Histogram Equalization and is used to improve the contrast of images. It works by creating several histograms of the image and uses all of these histograms to redistribute the lightness of the image.

The performance by CLAHE came out to be better than Histogram Equalization and significantly better than with no enhancement across all feature extractors I tested (discussed in next section).

0.1.3 Feature Extractors used

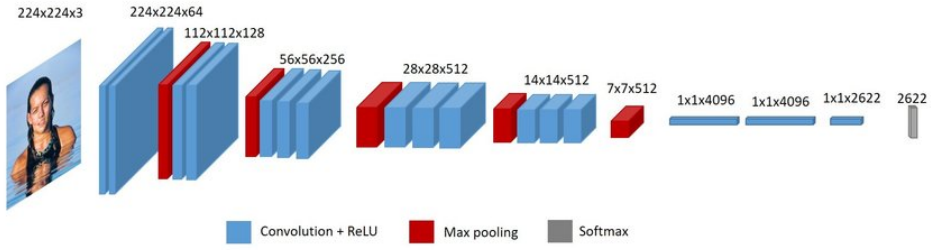
The next step was N x N matching using different feature extractors. The ones used are as follows -

1. SIFT - It stands for Scale Invariant Feature Transform and it is a feature detection algorithm. We extract the features from the test image and use them to match the test image. One major advantage is that its performance is independent of the orientation of the image.

2. ORB - ORB is a combination of the FAST keypoint detector and BRIEF descriptor with some additional features to improve performance. It is termed brief because it doesn't compute the orientation and descriptors for the features. ORB is known to be the fastest algorithm which we realized while performing the experiments. The speed reflects in the performance which showed a dip.

3. Arcface - ArcFace is a machine learning model that takes in 2 images as input and returns the distance between them to see how likely they are to be the same person. It can be used for face search and recognition. ArcFace uses a similarity learning mechanism that allows distance metric learning to be solved in the classification task by introducing Angular Margin Loss to replace Softmax Loss.

4. VGGFace - VGGFace is deep convolutional neural network trained on millions of images to recognize labelled faces in the wild (LFW). The original model takes an image in WildFace dataset on which VGGFace was trained and classifies/recognize person in image. It outputs 2622 embeddings for an image,we take these 2622 embeddings for each cropped image for later classification of image.



(A) VGGFace Network Architecture

0.2 Metrics

The matching scores were then used to plot the genuine/imposter histograms along with ROC curves. Along with the curves, we also obtained the CRR, ERR, Accuracy and the best threshold corresponding to each feature extractor/method.

1. Correct Recognition Rate (CRR): Ratio of number of subjects correctly classified to the total number of test subjects.

$$CRR = \frac{Number\ of\ matches\ correctly\ recognized}{Total\ number\ of\ matches} \tag{1}$$

2. Equal error rate (ERR) - It is the point on the ROC curve where proportion of false acceptances is same as the proportion of false rejection. The lower this value, the better the system. Higher CRR and low EER is better.

3. Accuracy: It is maximum value of $(100 - (FRR + FAR)/2)$ across all thresholds in the ROC Curve.

4. Discriminative Index(DI) - It represents the separation between genuine and imposter scores.

$$DI = \frac{|\mu_{Genuine} - \mu_{Imposter}|}{\sqrt{\sigma^2_{Genuine} + \sigma^2_{Imposter}}} \tag{2}$$

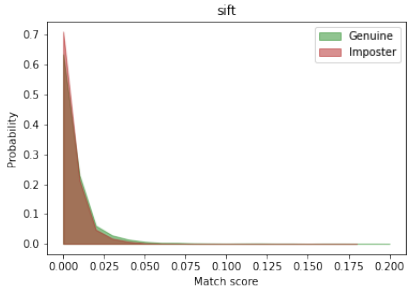
0.3 Results

The results have been summarized in the following table.

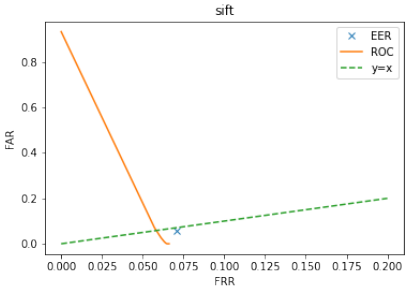
Method Used	EER(%)	Accuracy(%)	CRR(%)	DI
SIFT	0.01	99.967	68.80	0.15
AKAZE	0.01	99.624	90.00	0.09
ORB	0.02	99.576	82.17	0.04
ArcFace	0.81	99.933	78.33	0.55
VGGFace	0.75	99.967	86.64	1.41

0.4 Result plots

The plots obtained are shown below -

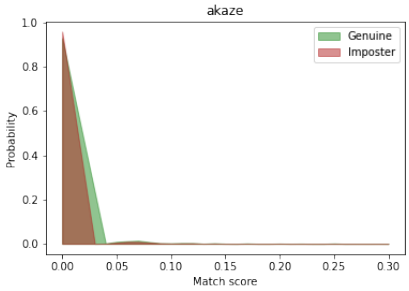


(A) Genuine-Imposter Matching Score Histogram

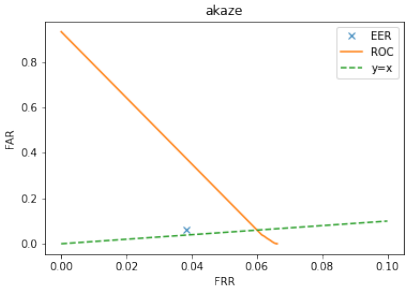


(B) ROC Curve

FIGURE 1: SIFT Results

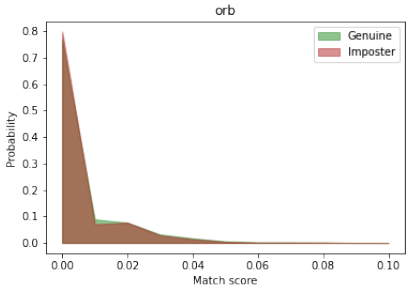


(A) Genuine-Imposter Matching Score Histogram

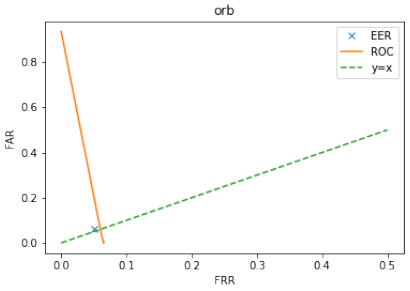


(B) ROC Curve

FIGURE 2: AKAZE Results

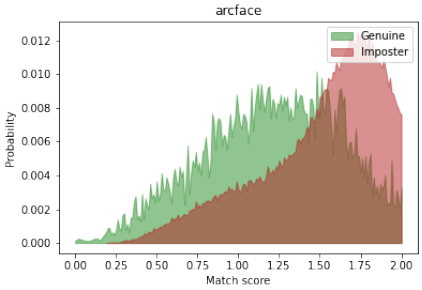


(A) Genuine-Imposter Matching Score Histogram

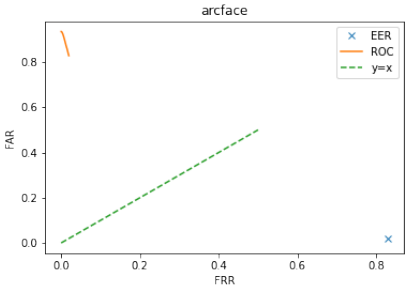


(B) ROC Curve

FIGURE 3: ORB Results



(A) Genuine-Imposter Matching Score Histogram

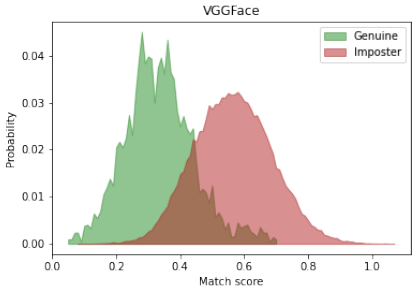


(B) ROC Curve

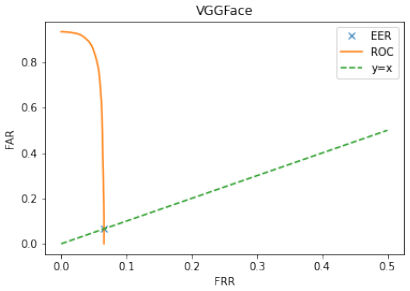
FIGURE 4: ArcFace Results

0.5 Ablation Study & Conclusions

Experiments were carried out using three different feature extractors: SIFT, AKAZE, ORB and the VGGFace network. Various combinations of preprocessing along with internal feature extractor settings were changed to obtain the best results.



(A) Genuine-Imposter Matching Score Histogram



(B) ROC Curve

FIGURE 5: VGGFace Results

Enhancing the images with Greyscaling and CLAHE gave the best results across all the feature extractors. Image Enhancement Techniques like Histogram Equalization and CLAHE improve the performance significantly as they improve the quality of the images and reduce the irrelevant noise.

VGGFace algorithm after image enhancement gave the best performance across Accuracy, EER and CRR metrics with a maximum accuracy of 99.967% and CRR of 86.64%. AKAZE, SIFT and ORB are much faster than VGGFace and ArcFace but perform marginally worser than them.

The histogram generated using AKAZE as a feature extractor has many gaps and a high percentage of imposter scores with a score of 1.0, which is incorrect. Although AKAZE has a better CRR, VGGFace is the best of the four methods as it has a higher CRR accuracy with a better distribution of the scores.

The Area under the ROC curve is also a measure of the performance of a system. VGGFace again leads here, followed by Arcface,SIFT,ORB and AKAZE. The threshold value(EER) for AKAZE is low compared to others, attributing to the ambiguities in the histogram due to its inability to adequately capture the features in the enhanced image.

Overall, VGGFace with image enhancement and resizing gives decent results and a proper age invariant face recognition system.