New York University

Project Report

Prepared by

Rahul Sankar (rrs6684)
Harshita Kukreja (hk3203)

1. Problem Statement

To create a multi-modal biometric system using facial images and handwritten signatures and verify the user of the system.

2. Background Work

Advancements in technology have led to a rise in the threat to the users' privacy. It is imperative that the identity of the user not be compromised in any way. Conventional systems like passwords, patterns, or personal identification numbers usually fail due to a lack of uniqueness and universality. To make the system more secure and to eliminate the threat to the privacy of the users, it is essential to employ multiple biometric modalities to verify the identity of the user.

Wang et al. [1] proposed a layer-by-layer deep learning method that solves the face detection problem and presented a Siamese Convolutional Neural Network that trained on the different parts of the face. The feature representation was then concatenated and the system reached an accuracy of 91% on the ORL database and 81% on the LFW database. Abdulrazzaq [2] presented a solution to the fusion of incompatible features of different modalities and issues with dimensionality using a Siamese CNN with two identical sub-networks that utilized a shared set of weights. The multimodal system fuses the features learnt from palm veins and the face making the system robust against imposters. Chakladar [3] combined handwritten signatures and EEG signals as the two biometric traits using a multimodal Siamese Neural Network (mSNN) for improved user verification. The mSNN employed a distance metric based on the similarity and the dissimilarity of the input features and achieved a classification accuracy of 98.75%. Wu et al. [4] proposed a convolutional siamese neural network for recognizing only the faces of users. The architecture used face detection to get the position of the face in an image. Srinivasan et al. [5] proposed a system involving handwritten signatures for user identification. The approach compares genuine and forged signatures using two learning methods, person-independent and person-dependent. The former learns differences between all individuals'

signatures while the latter learns a person's signature using only that person's multiple signature samples.

3. Solution Approach

3.1 Dataset

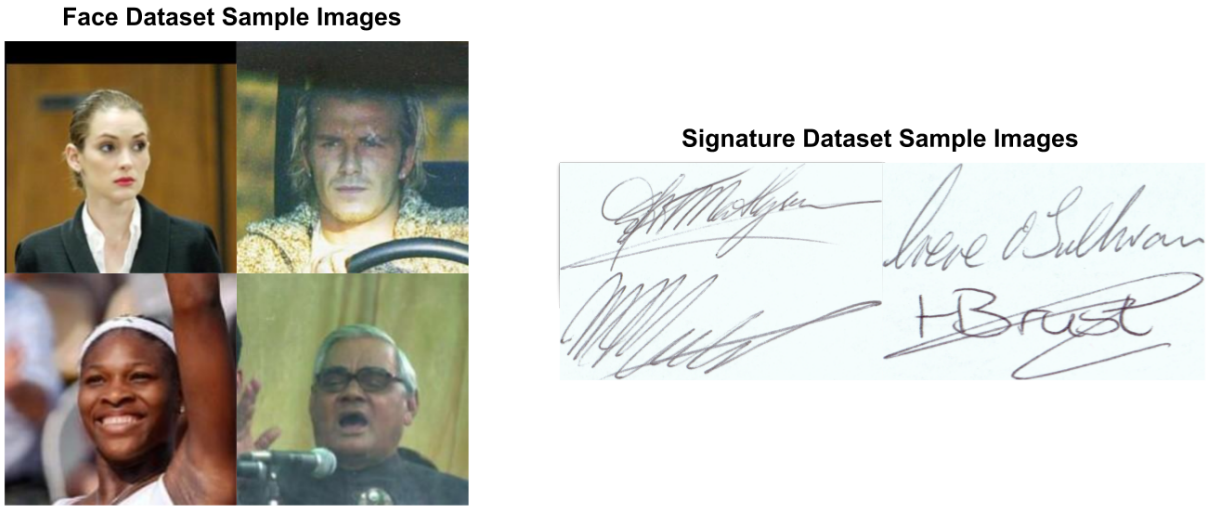

Figure 1 - Sample images from both the datasets

For creating our dataset, we had to combine two heterogeneous datasets, namely the 'Labeled Faces From the Wild' and 'ICDAR 2011 Signature Verification' datasets to form a single dataset. The LFW dataset was non-uniform, i.e. not every person had the same number of faces. We sorted the dataset based on the number of faces that each person had, and only selected people with the most faces to maximize the size of our final dataset. Each signature from the ICDAR dataset was uniquely mapped to each person from the LFW dataset. The final dataset consisted of 64 samples with 18 faces and 12 signatures each.

3.2 Model Architecture

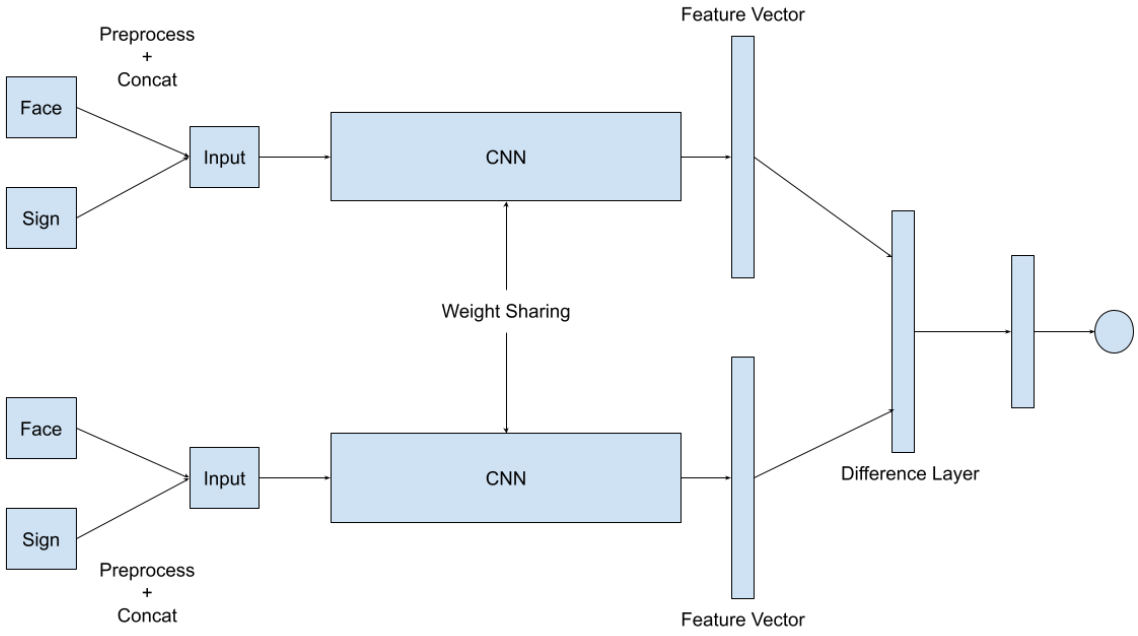

Figure 2 - Architecture of proposed model

We used a Siamese Convolutional Neural Network to estimate how similar two inputs are. A siamese CNN consists of two identical CNNs that share weights with each other. This enables them to learn how to properly perform feature extraction. Input to the model was an RGB image of the face, appended with a grayscale image of the signature. The input was resized to 128x128x4 and augmented since the size of the dataset is small. Augmentations included random horizontal flip, random rotation, random zoom, and random contrast.

The model takes two samples, performs features extraction, and compares the similarity of the two obtained feature vectors. Similarity of feature vectors was compared using the L1 norm. Other distance metrics such as the L2 norm and cosine similarity were also explored, but the L1 norm gave the best performance due to high dimensionality of the feature vector.
Training was done by selecting an anchor image, and sampling a batch from the dataset such that half of the batch was the same class as the anchor, and half was different.

4. Evaluation

All of our experiments were performed on a GPU + High RAM runtime on Google Colab Pro with an Intel(R) Xeon(R) CPU @ 2.30GHz, an NVIDIA P100, 25GB RAM and 160GB Disk memory. We used Python 3.8 with Tensorflow v2.7.0 to write the code for our approach.

The model was trained for 20,000 iterations. The Adam optimizer was used with a learning rate of 0.0001, along with the binary cross-entropy loss function. The model was evaluated every 1000 iterations on 250 20-way one-shot tasks.

An N-way one-shot task refers to the process where an anchor image is selected, and a support set is formed with 1 image belonging to the same class as the anchor image, and N-1 different images. If our model is working as expected, it should give the highest similarity for the image in the support set that is similar to the anchor image.

4.1 Results

All of our experiments were performed on a GPU + High RAM runtime on Google Colab Pro with an Intel(R) Xeon(R) CPU @ 2.30GHz, an NVIDIA P100, 25GB RAM and 160GB Disk memory. We used Python 3.8 with Tensorflow v2.7.0 to write the code for our approach.

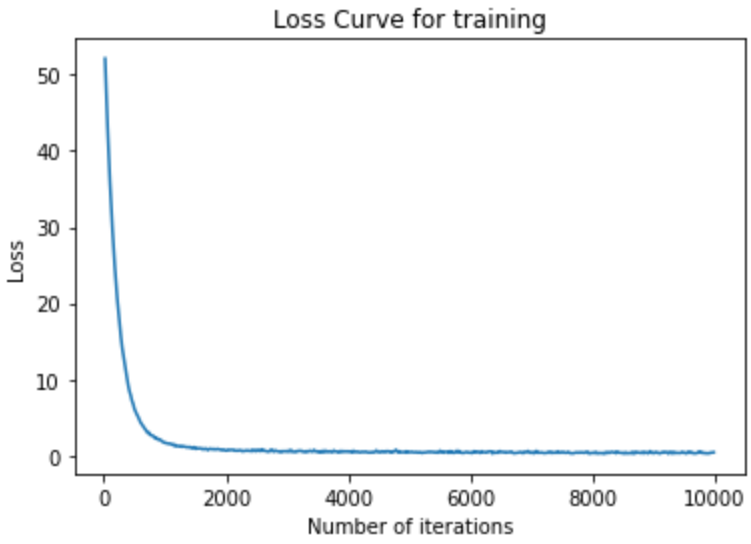

Figure 3 - Training Loss Curve

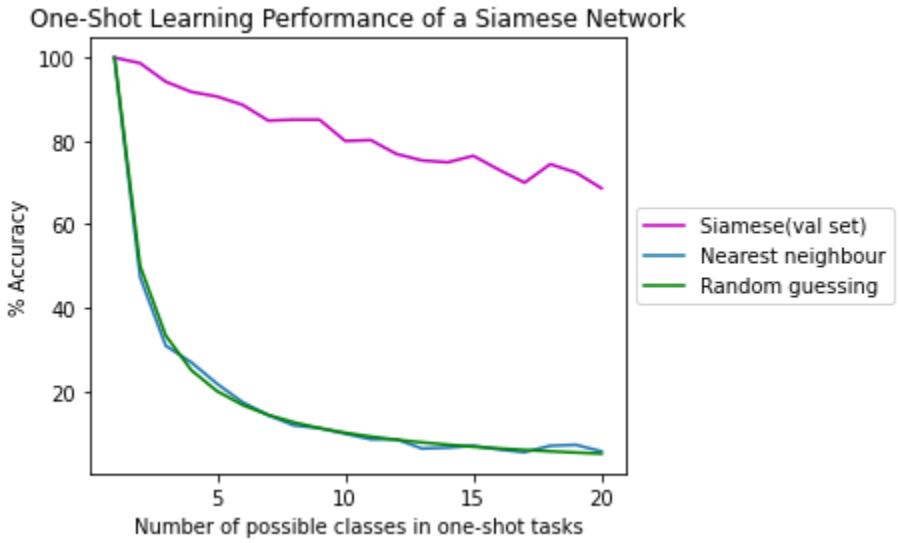

Figure 4 - Model performance compared to naive methods

As shown in Figure 3, the training loss curve decreases as the number of iterations increases. The model seems to converge rather quickly at around 1500 iterations and remains stable.

Figure 4 shows the performance of our model compared with random guessing and naive nearest neighbour clustering as we increase the value of N for 450

N-way one shot tasks. The random guessing method randomly picks an image from the support set, and the nearest neighbour clustering directly compares the anchor image with each image in the support set by taking their difference, and then picking the image with the least difference. It can be observed that our model consistently outperforms both the methods, indicating that the model is working as expected.

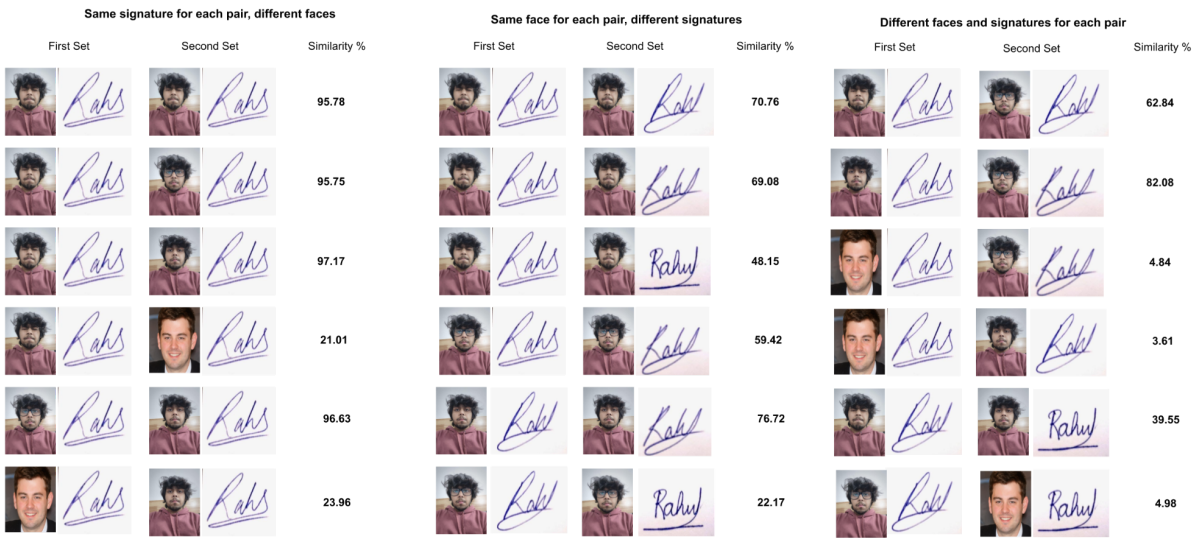

Figure 4 - Model performance on unseen face-signature pairs

We further evaluated our model on unseen face-signature pairs to confirm if the model hsa not overfitted our dataset. Three different scenarios were considered, i.e. same signature for each pair but different faces, same face for each pair but different signatures, and each pair with both the faces and signature as different.

The results of the experiments are shown in Figure 4. The model consistently performs as we expect. For all three scenarios, the model outputs a high similarity score if both pairs of images are from the same source, and a low similarity score if either the face, the sign, or both of them are from different sources. The threshold to decide if both pairs of the same system is left to the user of the system, but as per our experiments, the model seems to output a similarity of

more than 50% if both pairs belong to the same person, and lower than 50% otherwise.

5. Conclusion

From our experiments, we conclude that our solution outperforms random guessing and naive nearest neighbour clustering, and the similarity score generated by the model falls within expectations. Although more rigorous testing needs to be done to evaluate the performance of our model, since this belongs to the security domain, where we require highly reliable and predictable systems, the initial results look very promising.

Future work can include training our model on a larger dataset, and leveraging our solution as part of a biometric verification pipeline that sends an alert if the similarity score goes below a set threshold.

6. References

[1] Wang, W., Yang, J., Xiao, J., Li, S., & Zhou, D. (2015). Face Recognition Based on Deep Learning. Human Centered Computing, 812-820.

[2] Abdulrazzaq, H.I., & Hassan, N.F. (2019). Modified Siamese Convolutional Neural Network for Fusion Multimodal Biometrics at Feature Level. 2019 2nd Scientific Conference of Computer Sciences (SCCS), 12-17.

[3] Chakladar, D., Kumar, P., Roy, P., Dogra, D., Scheme, E., & Chang, V. (2021). A multimodal-Siamese Neural Network (mSNN) for person verification using signatures and EEG. Information Fusion, 71, 17-27.

[4] H. Wu, Z. Xu, J. Zhang, W. Yan and X. Ma, "Face recognition based on convolution siamese networks," 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017, pp. 1-5.

[5] Srinivasan, H., Srihari, S., & Beal, M. (2006). Machine Learning for Signature Verification. Computer Vision, Graphics And Image Processing, 761-775.