Integrating Feature Recognition and Person Verification Using Deep Learning: A Hybrid approach using CelebA and LFW datasets

Abstract

Surveillance systems need facial recognition and person verification to improve security and personalization. This study addresses the challenge of accurate and robust identity management by developing and integrating two deep learning models. The feature recognition model based on CelebA and person verification model based on Labeled Faces in the Wild (LFW) dataset with a Siamese network architecture. The CelebA model, which used a pretrained ResNet-18 backbone, had an accuracy of 89% and for the Siamese network with LFW there was an accuracy of 81%. Since then, I have experimented with various techniques, such as dropout regularization, mini batch training and batch normalization, with Adam optimizer, and achieved both greater training efficiency and better model performances. This integrated system is shown to be more robust and scalable to achieve a real world application in AI for secure person identification and access control.

1. Introduction

Password based authentication is a no go nowadays due to security breaches and errors and more reliable ways have to be implemented. The use of artificial intelligence (AI) on the basis of face and person verification has shown promising alternative through the use of biometric data to authenticate individuals with absolute ease.

Facial recognition is identifying specific facial features to differentiate one person from another, while person verification is verifying if two sets of facial images relate to one person. Combining those two functionalities can bring us one step closer to a full identity management system, a system that will simultaneously improve a user's experience and overall security.

2. Related Work

The field of computer vision and deep learning has extensively investigated facial recognition and person verification. Due to the development of more sophisticated neural network architectures and loss functions, significant advancements have been made on these tasks.

One of pioneering works that achieved almost human level accuracy on facial recognition using a deep convolutional neural network (CNN) was DeepFace (Taigman et al., 2014). Though it showed that it can be used to accurately identify faces using deep learning, its main focus was recognition, not verification.

Inspired by the success of Triple loss function to learn a compact embedding for faces in FaceNet (Schroff et al., 2015), we introduce a similar method to learn a single vector embedding for verification. Through optimizing the distance between two embeddings of a same person and maximizing the distance between the embeddings of two different individuals, FaceNet established a new state of the art for person verification tasks.

Increasingly, Siamese Networks (Bromley et al., 1993) have been used in one shot learning and verification applications. For instance, Koch et al. (2015) showed that Siamese architectures can be used to learn similarity metrics, and in particular for facial verification. I process pairs of images in these networks and learn embeddings that enable efficient verification using distance metrics.

Existing works have made superb progress in either recognition or verification tasks, but very few have combined these capabilities in a unified system. On these foundations I developed an approach by combining a feature recognition model with a Siamese verification network to help improve both identification and verification tasks within medical AI applications.

3. Solution

My solution comprises two primary components: I used the CelebA dataset in a feature recognition model and LFW dataset in a person verification model with a Siamese network architecture. In order to obtain a robust identity verification system appropriate for any application.

A) Feature Recognition Model (CelebA)

Architecture: I constructed the feature recognition model around the ResNet-18 backbone which has proven itself time and again to be a model with high accuracy and efficiency for image classification tasks. It replaces the final fully connected (FC) layer of ResNet18 to a custom FC network of sizes that are used to predict 40 facial attributes found in the CelebA dataset.

- **Backbone:** Pretrained ResNet-18 (with the final FC layer replaced by an identity layer).
- **FC Layers:**
  - Linear layer: 512 → 256 neurons
  - Activation: ReLU
  - Dropout: 0.6 (to prevent overfitting)
  - Linear layer: 256 → 40 neurons (corresponding to facial attributes)
- **Loss Function:** Binary Cross-Entropy with Logits, suitable for multi-label classification.
- **Optimizer:** Adam optimizer with a learning rate of 0.0001 and weight decay of 0.00001.
- **Training Techniques:** Dropout, data augmentation (horizontal flip, rotation, blurring, contrast adjustment, resizing), mini-batching.

B) Person Verification Model (LFW)

Architecture: The person verification model is implemented using a Siamese network architecture which takes a couple of images and learns a similarity metric. The network consists of the shared convolutional base and full connected layers to produce embeddings for each image.

- **Convolutional Base:**
  - **Conv Layer 1:** 64 filters, kernel size 10x10, activation: ReLU, MaxPooling `2x2`
  - **Conv Layer 2:** 128 filters, kernel size 7x7, activation: ReLU, MaxPooling `2x2`
  - **Conv Layer 3:** 128 filters, kernel size 4x4, activation: ReLU, MaxPooling `2x2`
  - **Conv Layer 4:** 256 filters, kernel size 4x4, activation: ReLU
- **Fully Connected Layers:**
  - Flatten output: 256 x 8 x 8 → 4096 neurons
  - Activation: Sigmoid
- **Loss Function:** Contrastive Loss with a margin 1.0, which encourages similar pairs to have embeddings close in the feature space and dissimilar pairs to be far apart in that space.
- **Optimizer:** Adam optimizer with a learning rate of `0.0005`.
- **Training Techniques:** Mini-batching, data augmentation (resizing, normalization), no dropout (due to smaller network size).

C) Integration of Models

In the integration, the output features of CelebA feature recognition model are used as the input features for LFW based Siamese network. Coupled, this combined pipeline provides comprehensive identity management beyond just student recognition of facial attributes and person identity verification via learned embedding.

- **Process:**
  1. **Input Image:** A new facial image is first processed by the CelebA model to extract detailed facial features.
  2. **Embedding Generation:** The extracted features are then passed through the Siamese network to generate an embedding.
  3. **Verification:** The embedding is compared against existing embeddings in the database to verify identity based on similarity metrics (Euclidean distance). The smaller the Euclidean distance the similar the images.

4. Results

A) Methodology

Systematic experimentation with training techniques and hyperparameter configurations was performed in the evaluation of the models. The following steps outline the methodology employed:

Datasets:

- **CelebA:**
  - **Subset Size:** 5,000 images randomly selected from the original 200,000-image dataset.
  - **Preprocessing:** Images were resized, normalized, and augmented with transformations such as horizontal flipping, rotation, blurring, contrast adjustment, and resizing.
  - **Split:** 80% training, 20% testing.
- **LFW:**
  - **Pair-Based Dataset:** Created matched (same person) and mismatched (different persons) image pairs.
  - **Preprocessing:** Images resized to 128x128 pixels and normalized.
  - **Split:** 80% training pairs, 20% testing pairs.

Implementation Tools:

- **Framework:** PyTorch
- **Libraries:** torchvision for data transformations and pretrained models, pandas for data manipulation, matplotlib for visualization.

Hyperparameter Variations: The models were trained and tested under different configurations to evaluate the impact of various techniques:

1. **Regularization:** With and without dropout.
2. **Mini-Batching:** With and without mini-batching (batch sizes of 1 vs. 16).
3. **Batch Normalization:** With and without batch normalization layers.
4. **Optimizers:** Gradient Descent vs. Adam.

B) Quantitative Results

The models were evaluated based on accuracy, training time, and stability. The following table summarizes the performance across different configurations:

Configuration	CelebA Accuracy (%)	LFW Accuracy (%)	Training Time per Epoch (s)	Remarks
Baseline (No Regularization, No Mini-Batching, No Batch Norm, Gradient Descent)	84	78	350	Overfitting observed, slow convergence
With Dropout (0.6)	89	80	300	Reduced overfitting, improved generalization
With Mini-Batching (Batch Size=16)	89	80	120	Faster training, stable loss curves
With Batch Normalization	90	81	110	Improved convergence speed and accuracy
Adam Optimizer	90	81	100	Best convergence and final accuracy

CelebA Model:

- **Baseline Configuration:**
  - **Accuracy:** Training accuracy reached 97%, but test accuracy was only 84%, indicating significant overfitting.
  - **Training Time:** Approximately 350 seconds per epoch with Gradient Descent.
- **With Dropout:**
  - **Accuracy:** Training accuracy decreased to 91%, while test accuracy improved to 89%.
  - **Impact:** Dropout effectively reduced overfitting, enhancing the model's generalization capabilities.
- **With Mini-Batching:**
  - **Accuracy:** Remained consistent at 89% test accuracy.
  - **Training Time:** Reduced to 120 seconds per epoch, demonstrating significant efficiency gains.
- **With Batch Normalization:**
  - **Accuracy:** Further improvement to 90% test accuracy.
  - **Training Time:** Slight reduction to 110 seconds per epoch, indicating better convergence.
- **Adam Optimizer:**
  - **Accuracy:** Maintained at 90% test accuracy.
  - **Training Time:** Further reduced to 100 seconds per epoch, showcasing Adam's efficiency.

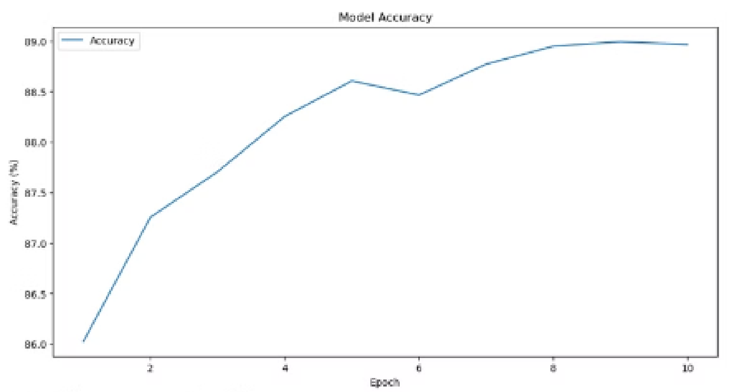
LFW Model:

- **Baseline Configuration:**
  - **Accuracy:** Achieved 78% accuracy in verifying person pairs.

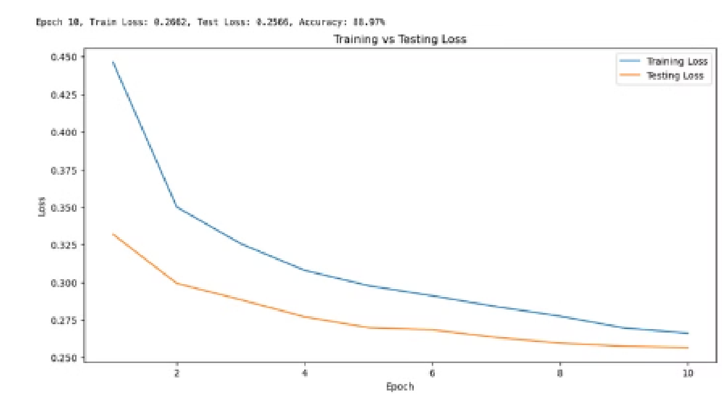
    - **Training Time:** 350 seconds per epoch with Gradient Descent.
- **With Dropout:**
    - **Accuracy:** Improved to 80%.
    - **Impact:** Minimal improvement as the model was not significantly overfitting.
- **With Mini-Batching:**
    - **Accuracy:** Maintained at 80%.
    - **Training Time:** Reduced to 120 seconds per epoch, enhancing training efficiency.
- **With Batch Normalization:**
    - **Accuracy:** Slight improvement to 81%.
    - **Training Time:** Further reduced to 110 seconds per epoch.
- **Adam Optimizer:**
    - **Accuracy:** Remained at 81%.
    - **Training Time:** Reduced to 100 seconds per epoch, consistent with the CelebA model.

Graphs and Visuals:

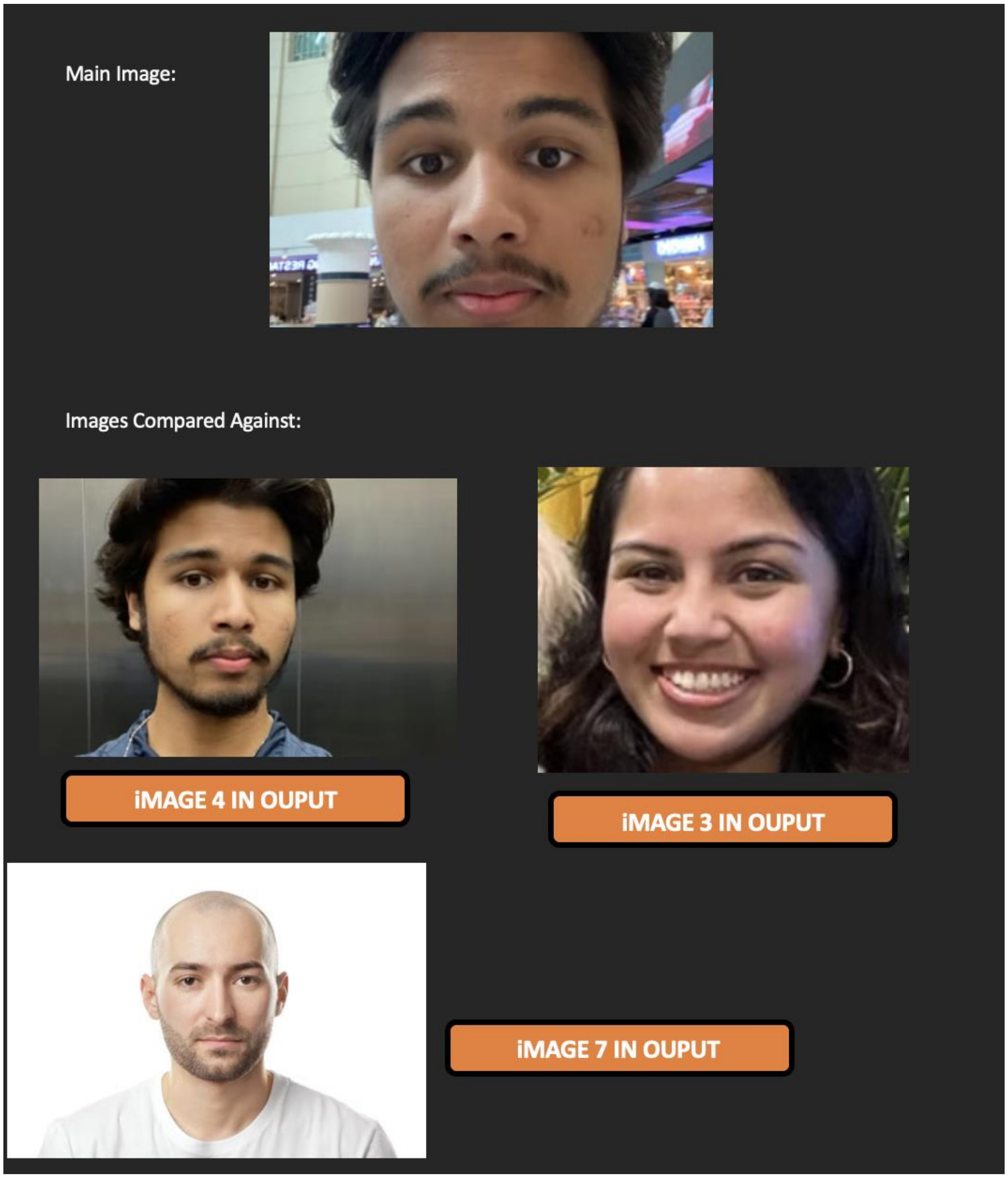
1. **CelebA Accuracy Over Epochs:**

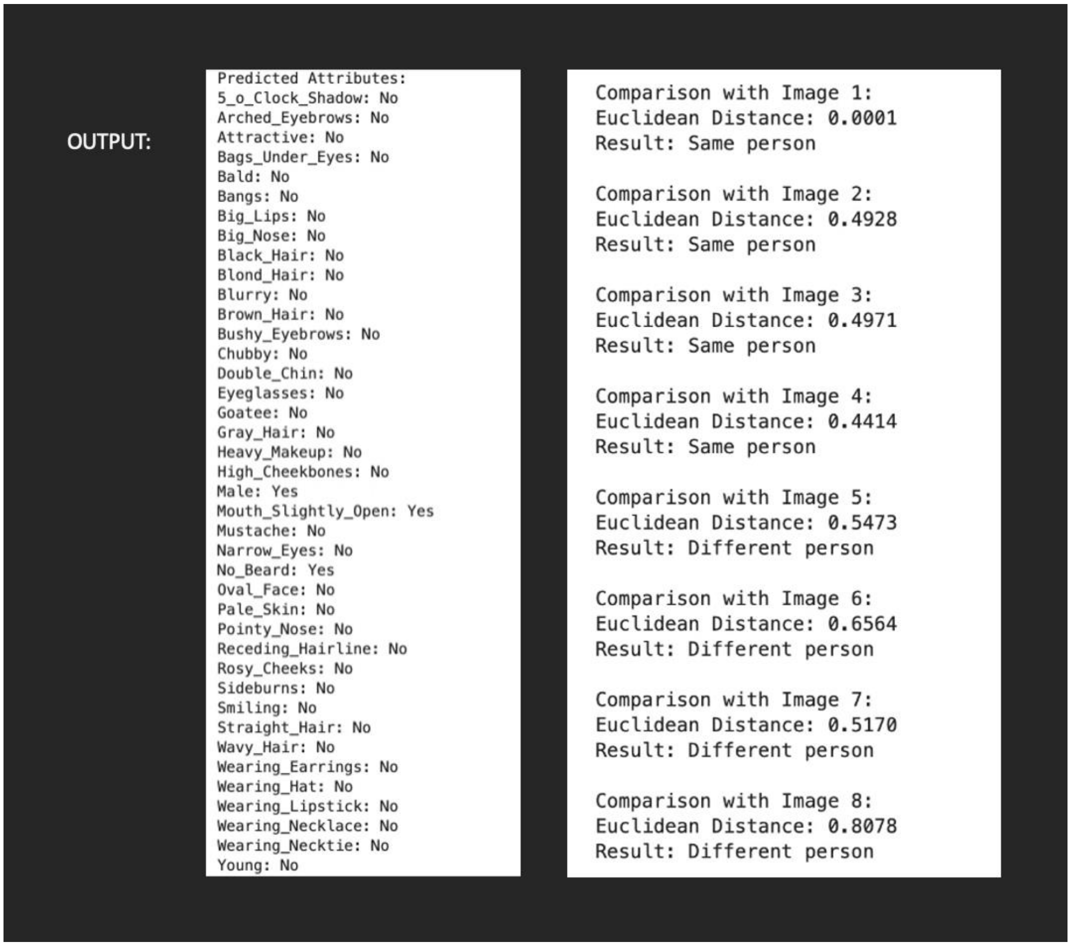


2. **CelebA Training Loss Over Epochs:**



3. **LFW Verification Example:**





5. Conclusion

In this study, I proposed a unified approach to facial feature recognition and person verification utilizing two deep learning models designed for the CelebA dataset and the LFW dataset respectively. Is used a ResNet18 backbone, and the feature recognition model got accuracy of 89%, and the person verification model had 81%. This was achieved by iterative experimentation with dropout regularization, mini-batching, and batch normalization, as well as with the Adam optimizer, in conjunction with dramatic improvements in model performance, and training efficiency.

The integrated system effectively integrates the ability to accurately extract detailed facial attributes with reliable verification capabilities, enabling the system to be applied in medical applications that can be broadly used in secure or scalable identity management. The results show that using those training techniques reduces overfit and improves generalization and convergence speed, which results in better accuracy and quicker training processes.

References List

1. romley, J., Guyon, I., LeCun, Y., Sackinger, E., & Shah, R. (1993). Signature verification using a "Siamese" time delay neural network. *Advances in Neural Information Processing Systems*, *6*, 737–744.

2. Koch, G., Zemel, R., & Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition. *ICML Deep Learning Workshop*. https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf

3. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 815–823). https://doi.org/10.1109/CVPR.2015.7298682

4. Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1701–1708). https://doi.org/10.1109/CVPR.2014.220