

# CIS 5810 Extra Project Mid-term Report – Celebrity Facial Recognition

Angela Wang (shuw119@seas.upenn.edu)

Zhengtao Hu (huzhengt@seas.upenn.edu)

Spring 2023

## Project Summary

Facial recognition technologies are widely used across various industries. For this extra project, our team aims to develop an algorithm that can accurately detect a given celebrity's face from sets of videos or images. The team plans to leverage some widely adopted facial recognition models (such as VGGFace, FaceNet, etc.) and utilize deep learning techniques to extract features and identify the target celebrity's face in the given dataset.

## Goals and Objectives

The main goal of the project is to build a facial recognition algorithm that can detect a given celebrity's face from sets of videos or images with a high level of accuracy.

To achieve this, our objectives are:

- Assemble an impactful dataset for training, validation, and testing purposes.
- Accurately label the data to facilitate more effective training.
- Develop a functional facial recognition algorithm capable of detecting faces in videos or images.
- Evaluate the algorithm's performance in terms of accuracy and speed.

## Proposed Approach

The proposed approach consists of the following steps:

1. Collect a large dataset of images containing a certain celebrity's face from various sources, such as movies, interviews, and public appearances.
2. Conduct a review of the widely adopted pre-trained deep learning models, such as VGGFace and FaceNet, and select a model that best fits the purpose of the project.
3. Process and annotate a portion of the dataset for training purposes, ensuring the model makes the right decision in subsequent supervised learning processes. Reserve the rest of the dataset for testing and validation.
4. Build the algorithm by implementing the selected pretrained deep learning model. Freeze the early layers of model, which usually detect basic features like edges, corners, and textures. Subsequently add new layer for classification tasks. Train model once after.
5. Fine-tuning it on the training dataset, and optimizing the results using additional deep learning techniques, such as face alignment
6. Evaluate the algorithm's performance in terms of accuracy and speed.

## Progress

Over the past two weeks, our team has progressed as scheduled in implementing our celebrity facial recognition project using PyTorch, a powerful deep learning framework. We have successfully completed the following tasks:

1. **Data Collection and Preprocessing:** We chose Brad Pitt as our target celebrity for facial recognition for the ease of collecting images. We collected a dataset of 1,000 images containing Brad Pitt's face and 4,500 containing other people's faces. We allocated 3,700 images as the training set and 1,800 as the validation set. The images were mostly single-person shots, so we separated them into "brad" photos and "non-brad" photos. We preprocessed the training set images using various data augmentation techniques, such as random cropping, horizontal flipping, rotation, affine transformations, and color jittering to increase the robustness of our model. We used the *"torchvision"* library to apply these transformations.
2. **Model Selection:** We selected the VGG16 model as our base model due to its proven performance and popularity in facial recognition tasks. The VGG16 model is built with a simple architecture with a series of convolutional layers, which enables the model to learn a hierarchical representation of features and makes it suitable for facial recognition problems. We utilized the *"torchvision.models"* library to load a pre-trained VGG16 model. To adapt the model for our binary classification problem, we modified the last layer of the classifier to have two output nodes corresponding to our two classes (*"brad\_pitt"* and *"ppl"*). We froze the early layers of the model, which detect basic features like edges, corners, and textures, to preserve the pre-trained weights.
3. **Training and Validation:** We set up the training and validation data loaders using the *"torch.utils.data.DataLoader"* class to load and process the training and validation datasets. So far, we trained our model using the *"CrossEntropyLoss"* loss function and the *"Adam"* optimizer with a learning rate of  $10e-5$  and weight decay of 0.001. We trained the model for 35 epochs and monitored the training and validation losses and accuracies.
4. **Model Evaluation:** After training, we evaluated our model's performance with a batch of validation images. We visualized the images along with their predicted and true labels, highlighting correct predictions in green and incorrect predictions in red (see **Figure 1**). Based on our work so far, our model is showing that training loss stabilized at 0.0001 and validation loss stabilized at 0.0028. To encounter the overfitting issues, we have fine-tuned the hyperparameters and applied the L2 regularization. The model is showing an accuracy of 97% in recognizing Brad Pitt's face using the validation dataset.



Figure 1. Visualization of Results

## Challenges and Improvement Strategy

The main challenge our team has faced so far is that we are still experiencing overfitting issues, as the training loss and validation loss do not seem to have converged to the same minimal value. Specifically, the validation loss stops decreasing and fluctuates around a certain value, even with an increased number of training epochs (please refer to **Figure 2**).

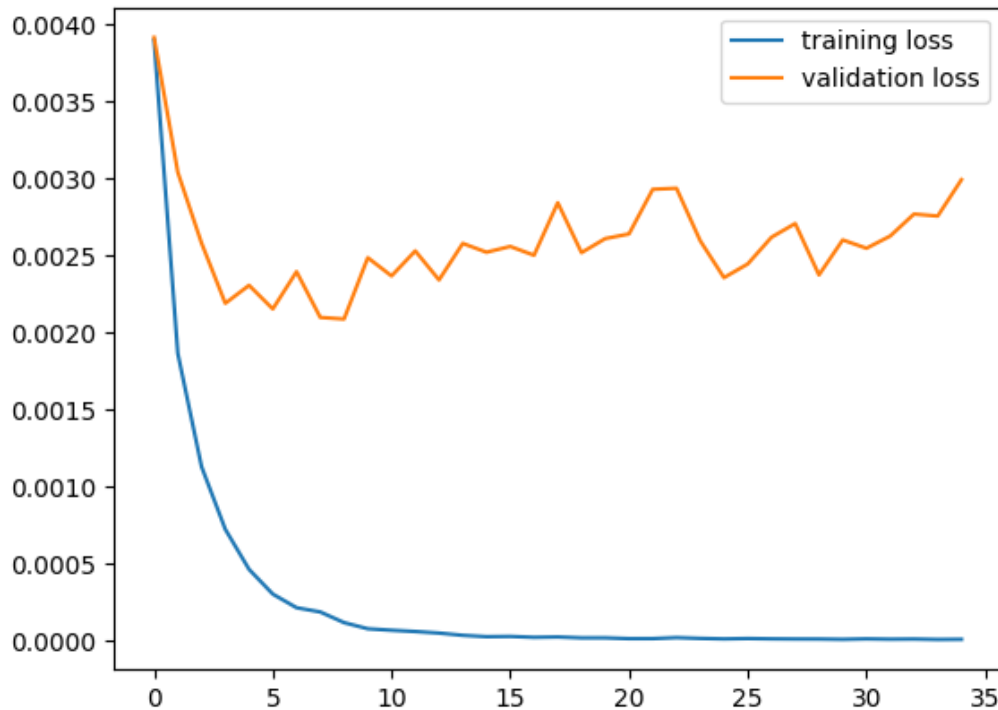


Figure 2. Training Loss and Validation Loss

Our team has identified several areas for improvement that can potentially further enhance the accuracy and performance of our celebrity facial recognition algorithm.

1. **Expanding the dataset:** We will continue to expand our dataset by adding more images to both “brad” and “non-brad” groups with photos taken in various lighting conditions, poses, and expressions. Ideally, we would like to expand the size of the dataset to 5,000 for “brad” and 10,000 for “non-brad”. A more comprehensive and diverse dataset will help the model learn more nuanced features, potentially increasing the model’s accuracy and robustness. A possible method of obtaining a large number of images could be capturing pictures from Brad Pitt’s interview videos.
2. **Implementing a more sophisticated augmentation strategy:** We will look into developing a more sophisticated data augmentation strategy to improve the model's ability to generalize to new data. This may include additional transformations, such as random rotations, zooming, and changes in brightness and contrast.
3. **Investigating alternative models:** We will continue to explore other pre-trained models to determine if they offer better performance or are more suited to our specific task.

4. **Training our model on multi-person photos:** we will investigate the approach of manually annotating Brad Pitt's presence in the multi-person images to improve the model's performance in more challenging scenarios.
5. **Hyperparameter tuning:** We will continue to try different hyperparameters, such as learning rate and weight decay, to optimize performance in terms of speed and accuracy.
6. **Incorporating face alignment:** We will investigate incorporating face alignment techniques during the preprocessing stage to ensure that the faces in the dataset are consistently aligned. This could help improve the algorithm's performance by making it easier for the model to extract meaningful features and recognize the celebrity.

By addressing these potential improvements, we hope to further enhance the performance of our celebrity facial recognition algorithm and make it more robust and versatile for real-world applications.

## Related Work

1. Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). **Deep face recognition. Proceedings of the British Machine Vision Conference (BMVC).**

This paper introduces the VGGFace model and demonstrates its effectiveness in facial recognition.

2. Schroff, F., Kalenichenko, D., & Philbin, J. (2015). **FaceNet: A unified embedding for face recognition and clustering. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).**

This paper introduces the FaceNet model and its application in face recognition and clustering.

3. Xi Yin, Xiang Yu, Kihyuk Sohn, Xiaoming Liu and Manmohan Chandraker Michigan State University NEC Laboratories America §University of California, San Diego: **Feature Transfer Learning for Face Recognition with Under-Represented Data.**

This paper proposed center-based feature transfer framework to augment the feature space of under-represented subjects from the regular subjects that have sufficiently diverse samples.

## TimeLine

- Week 1-2
  - Reviewing and selecting pre-trained models, and setting up development environment
  - Data collection and annotation
- Week 3
  - Model training
- Week 4-5
  - Model fine-tuning, algorithm optimization, and performance evaluation

## Duties

Zhengtao Hu	Angela Wang
- Dataset collection and annotation	- Dataset collection and annotation
- Model training	- Pretrained model research and selection
- Model validation and test	- Model fine-tune