Project Title: Facial Anti-Spoofing using Deep Neural Network Approaches

Participants: Catherine Lyu (hl3553), Lara Karacasu (lk2859), Jasmine Chen (jjc2328), YeongWoo Kim (yk2920)

Project Description

Facial anti-spoofing is the problem of preventing false facial authentication using a photo, video, or mask. Given that many modern authentication methods require users to display their faces in order to gain access to their device, facial spoofing remains a prominent security threat. We aim to develop a neural network that accurately predicts whether or not an image of a face is spoofed.

Objectives

This is essentially a classification problem with binary labels "real" and "spoof". The goal is to reach a high success rate of detecting real vs spoof images using methods we have seen in relevant papers, and possibly improve on them. Such papers include "A Personalized Benchmark for Facial Anti-Spoofing" (Belli et al.), "CelebASpoof: A Large-Scale Face Anti-Spoofing Dataset with Rich Annotations" (Zhang et al.), and others.

Plan for Completion

- Data Collection: Gather a large dataset of celebrity images and spoof images on Kaggle. The linked dataset has approximately 500,000 images that are classified into different types of spoofing.
- Data Curation and Augmentation: Perform augmentation techniques to improve the robustness of the model, such as noise injection or removal, or geometric transformations.
- Framework: Utilize Pytorch for model development.
- Network Architecture: Use models including transformer and CNN.
- Training and Validation: Split the dataset, preprocess the images, and train the neural network using the curated and augmented dataset. Validate the model's performance using appropriate evaluation metrics such as accuracy, precision, recall, and F1-score.
- Testing and Evaluation: We will reserve a portion of the data for testing, around 20%. The aforementioned papers include several evaluation metrics – we will use some of the same ones for ease of comparison. Namely, Area Under Curve (AUC), Area Under Curve till False Negative Rate of 10% (AUC10), and Equal Error Rate (EER) are three evaluation metrics used in Belli et al. These metrics were used due to their indifference toward choice of threshold values, and we will likely use some of the same metrics to compare. Additionally, because this is a binary classification problem where the cost of false negatives is higher than the cost of false positives (eg. your device incorrectly marking someone else's face as real would be worse than your device incorrectly marking your face as spoofed), we will optimize for recall.

- Final Demonstration: Showcase the model's capabilities to identify spoof images.

Milestones

- ~~Data selection~~ and cleaning: 2/12
- Start researching ML techniques for preliminary model: 2/16
- Start developing preliminary model: 2/29
- Finish developing preliminary model: 3/1
- Finish midterm presentation: 3/4
- Finish final model: 4/19
- Finish final presentation: 4/21

Risks & Mitigations

Spoofing Method: Four main categories of spoofing are used within the dataset, namely print (printed photo), papercut (photo cut-out), replay (image on digital devices) and 3D (photo worn by a real person). Each of the four categories have unique attributes and challenges, making it challenging to train a highly accurate model to detect spoofing. As a mitigation, we will focus on print and replay as the initial scope of the project. If time permits, we will expand on all types of spoofing included in the dataset. These two categories are chosen as the starting point since they are the easiest and most common methods of spoofing.

Illumination condition: One possible feature of the dataset that may introduce challenges to the spoofing detection task is illumination condition. Specifically, the dataset has introduced extremely strong and weak lighting conditions, in addition to indoor and outdoor environments. These can make certain distinguishing features such as borders of fake images harder to recognize. The first possible mitigation for this would be to only introduce images that have normal lighting conditions to the model, then expand and finetune with different illumination once the model is able to achieve a good level of spoofing detection. The second mitigation would be to include additional steps within the model that will first process the images such that contrast and exposure can be appropriately adjusted.

References:

- "Domain Invariant Vision Transformer Learning for Face Anti-spoofing" (Liao, etc. 2023)
- CelebA-Spoof: Large-Scale Face Anti-Spoofing Dataset with Rich Annotations (Kaggle, 2020)
- Large Crowdcollected Face Anti-Spoofing Dataset (Kaggle, 2020)
- CelebASpoof: A Large-Scale Face Anti-Spoofing Dataset with Rich Annotations (Zhang et al.)
- A Personalized Benchmark for Facial Anti-Spoofing (Belli et al.)