Biometrics (COMP 388-002/488-002)
Loyola University Chicago, Fall 2025
Assignment 3: Face Recognition
Due date: November 14, 11:59 PM CT
Total: 10 points

1. Introduction

The purpose of this assignment is to train and evaluate the students' capabilities to use, set the operating point, and perform biometric identification with a third-party library of face recognition, over a selected partition of the Yale Faces Database (https://tinyurl.com/mpb42a7j).

1.1. Face dataset

The face dataset is available at https://tinyurl.com/4bwsx53m and is a modified subset of the Yale Faces Database (https://tinyurl.com/mpb42a7j). The data is organized into two folders, namely "dataset" and "queries". The "dataset" folder contains 90 images, each depicting a single face, whose names follow the format "`subject<ID>.<CAPTURE_CONDITION>.png`". As a consequence, two files with the same `<ID>` depict the same individual, hence comprising a genuine pair. Figure 2 provides a summary of the content of "dataset" with 14 images and two individuals.

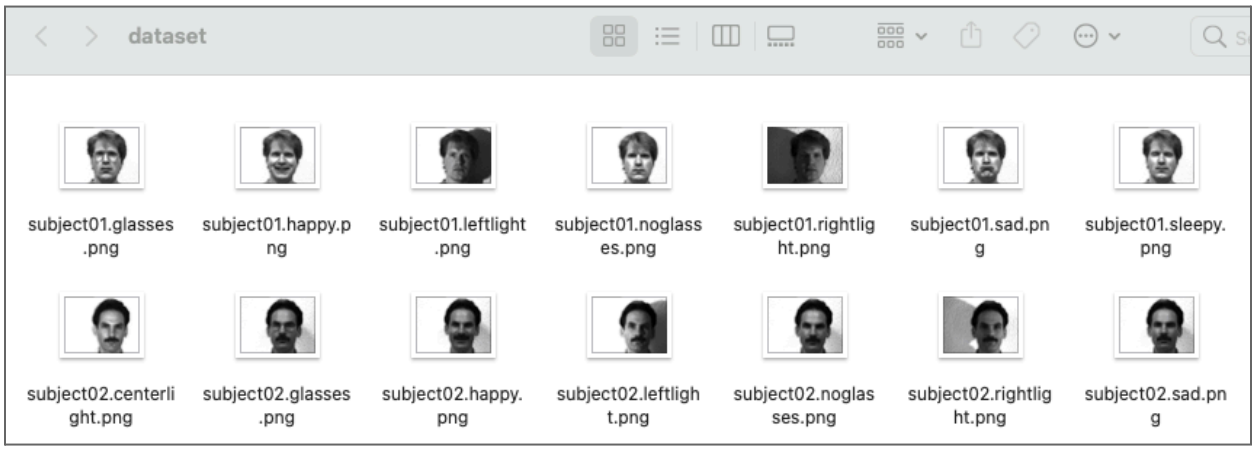

Figure 2. Example of the content of the "dataset" folder.

The "queries" folder, in turn, contains 15 image files whose names are random numbers with no particular meaning. These images do not have exact copies within the "dataset" folder, and the depicted individuals may or may not have images with their faces within the "dataset" folder. Figure 3 depicts the full content of "queries".

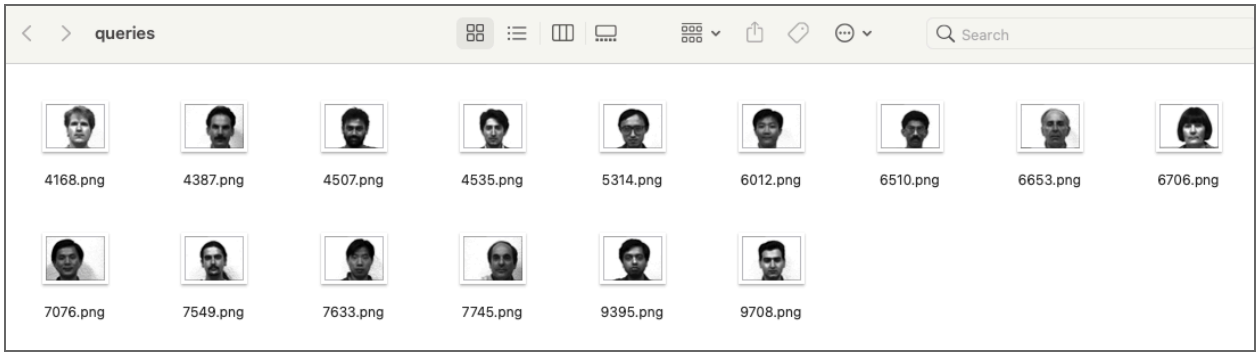

Figure 3. Example of the content of the folder “queries”.

1.2. Face recognition and metrics libraries

The face recognition library is available as a Google Colab notebook (https://tinyurl.com/26y9rcrh), which was explained in class. In addition to this implementation, you will also need either the first assignment’s notebook (metrics’ implementation, available at https://tinyurl.com/53knr8y5) or the *scikit-learn* (https://tinyurl.com/nsyu2k9b) library to compute the proper metrics and analyze the performance of the system.

1.3. Assignment directions

To complete this assignment, access the Google Colab notebook (https://tinyurl.com/26y9rcrh) and make your own copy. After downloading and unzipping the dataset from Sakai to your local computer, upload the contents to your Google Colab notebook copy. Lastly, follow the instructions and answer the questions presented in Sec. 2.

There is no formal template for providing your answers. You may use the editor you like. The following option should work fine:

- A single PDF file or Word document containing all your answers and generated figures.

Please submit your file through the respective open assignment in Sakai by November 14, 2025, 11:59 PM CT.

2. Questions

2.1. As explained in class, the provided third-party face recognition library can extract a 512-dimensional float feature vector from a given face image, as well as calculate the angular distance between two feature vectors, using ArcFace (https://arxiv.org/abs/1801.07698). The expected behavior of the software is to generate small distances for two face images depicting the same individual (a genuine pair) and large distances for two images depicting different individuals (an impostor pair).

Leveraging the content of **only** the “dataset” folder within the provided data, the third-party face recognition library, and metrics learned in class, please determine **what a good**

angular distance threshold is to separate genuine from impostor pairs. While providing your answer for the distance threshold, please explain in detail how you computed it. (1 point)

2.2. **What is the AUC** of the face recognition system you are using? In addition, please provide a graph with the system's **ROC curve**. (1 point)

2.3. By leveraging the face recognition system and the distance threshold previously computed, and by either capturing your face with your webcam or providing an image with your face, find within the "dataset" folder which individual is the most similar to you. **Please provide the subject ID and the angular distance between your face and theirs**. In your *opinion*, do you have anything in common with this subject (e.g., gender, ethnicity, age, pose, etc.)? If yes, what do you think it is? (2 points)

2.4. By leveraging the face recognition system and the distance threshold previously computed, **please provide the subject ID** (or "**UNKNOWN**", if the individual does not have a face within the "datasets" folder), as well as **the respective angular distances** that supported your decision, for each one of the 15 images provided within the "queries" folder. (6 points)