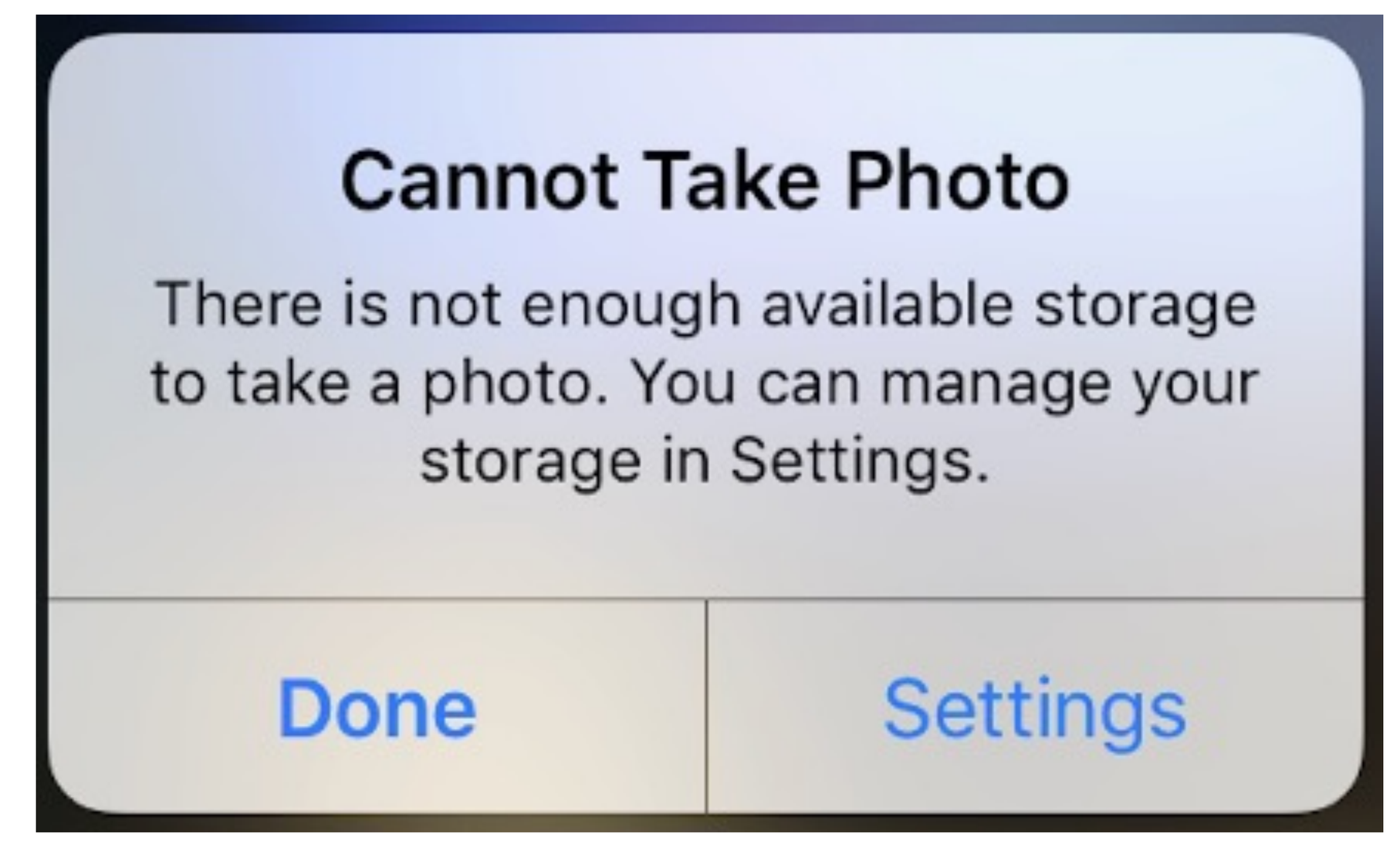
Duplicate Image Detection using Machine Learning

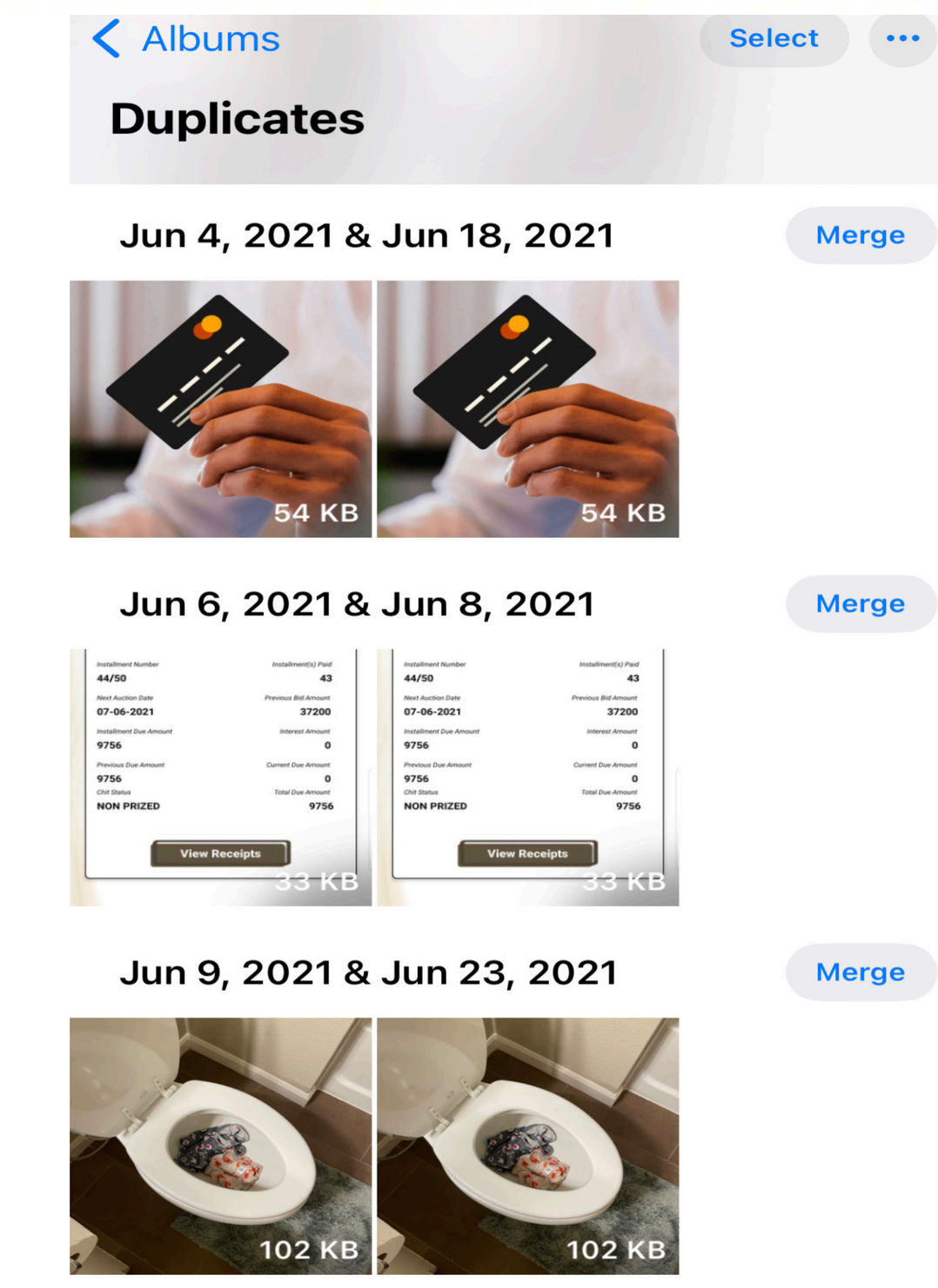
Presenter : **Karthi Balasundaram**
Advisors : Dr. Ahmed El Ouadrhiri and Dr. Phu H. Phung
Sponsor : **Synchrony Bank**

Objective

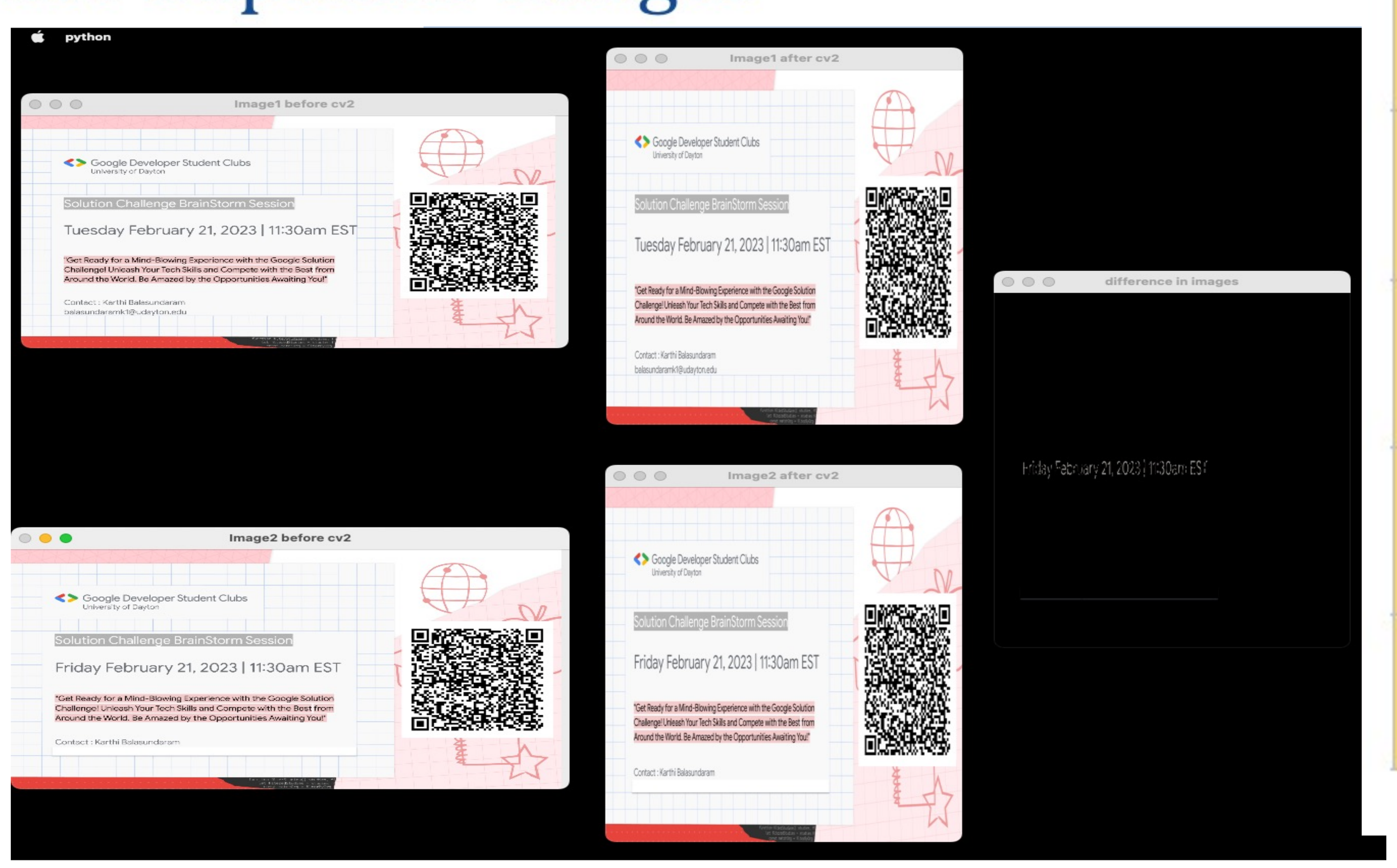
Compare two images and identify duplicates.



MOTIVATION



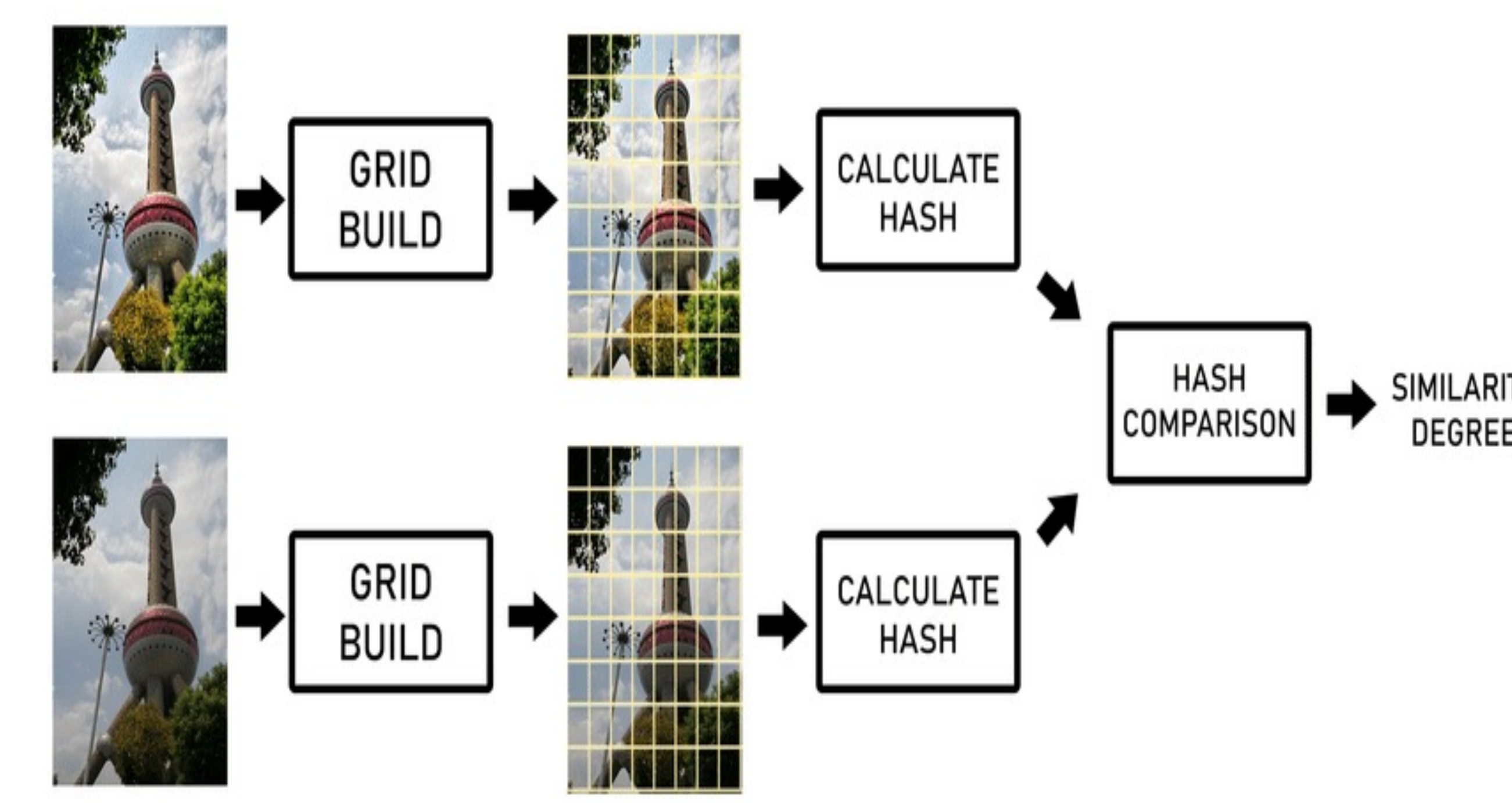
Computer Vision Technique to detect not-duplicate Images



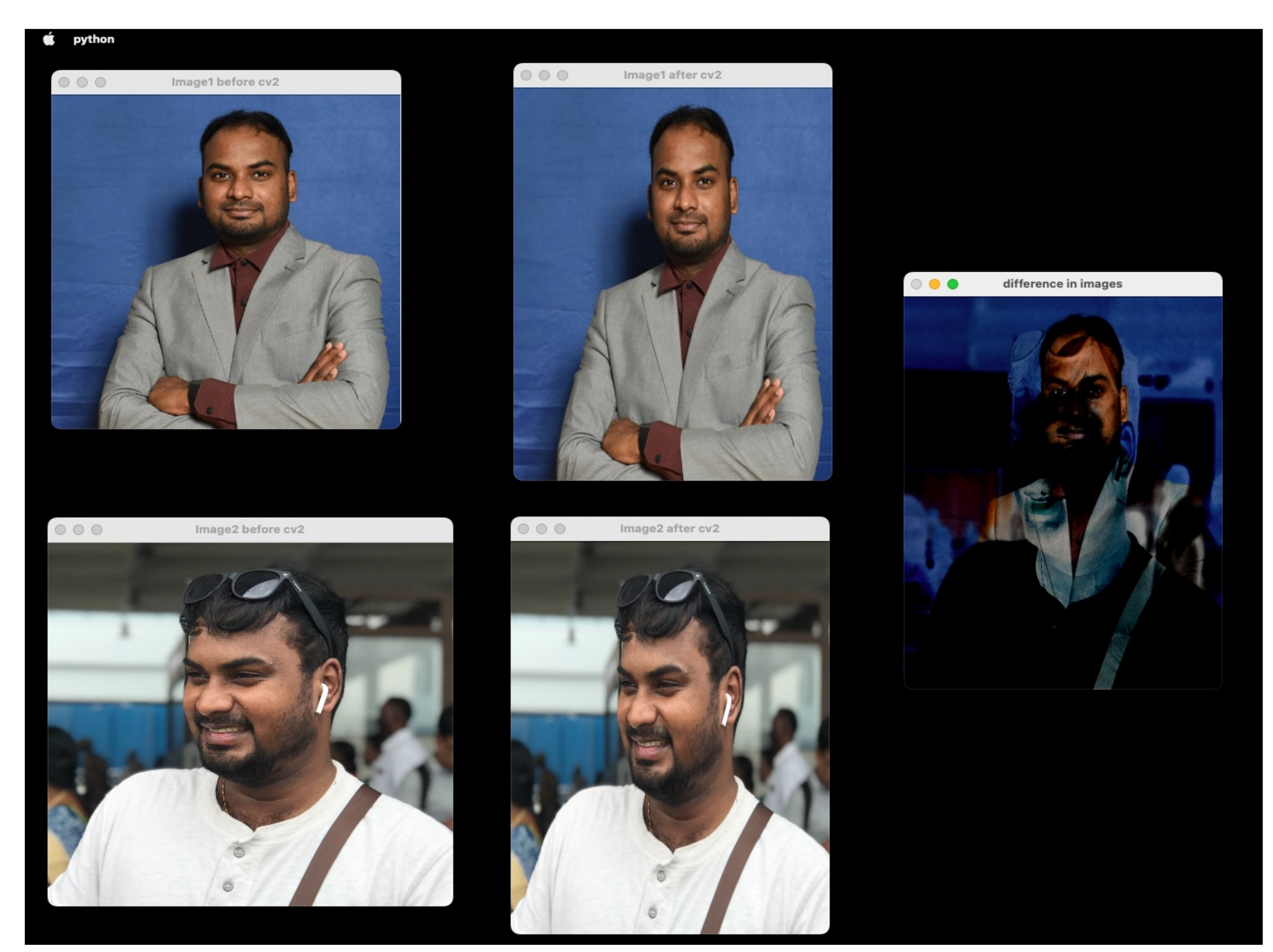
PROJECT OVERVIEW

PROBLEM OR ISSUE	The goal is to compare two images and identify duplicates. The System should be able to flag images with slight color variations, sizes, and formats to avoid duplicity.
PURPOSE OF PROJECT	Reducing the storage costs by identifying and removing duplicate images, enhancing data quality by flagging and removing duplicates, and by automating the process of identifying and managing duplicate images, the system can save users time and effort.
BUSINESS CASE	Financial companies; Synchrony that have large collections of images as part of their operations, could benefit from using a duplicate image detection system to manage and organize their images more effectively.
GOALS / METRICS	The goal is to compare two images and identify duplicates. The System will be able to flag images with slight color variations, sizes, and formats to avoid duplicity. This could include features such as the ability to compare images based on visual similarity, image hashing, or image recognition techniques.
EXPECTED DELIVERABLES	A web application that takes two images as inputs from the user, compares it and detects if it is duplicate or not.

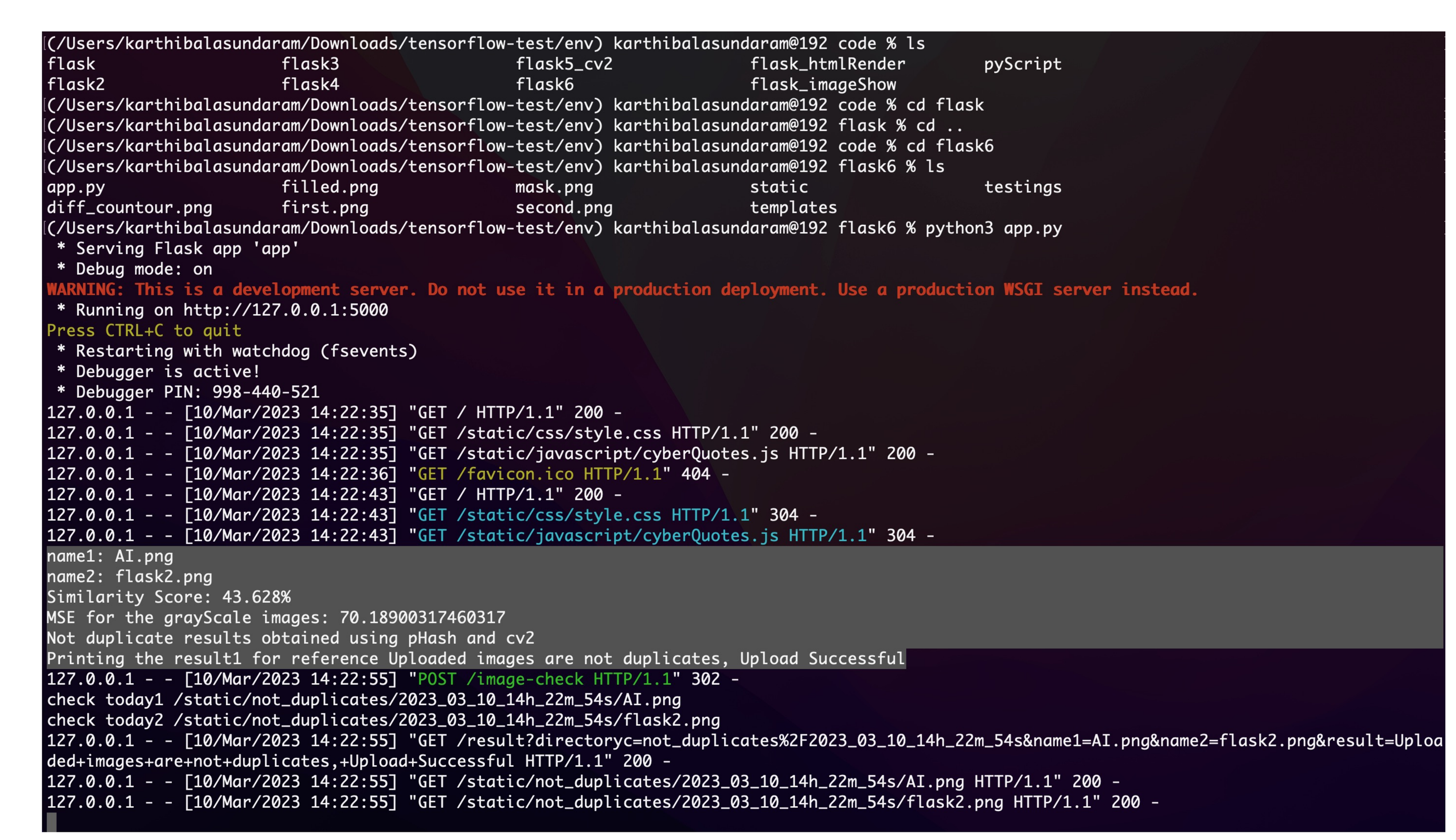
Perceptual Hash Logic



Highlighting the Difference!



Similarity Scores & System Log



```
(/Users/karthibalasundaram/Downloads/tensorflow-test/env) karthibalasundaram@192 code % ls
flask                  flask3                 flask5_cv2             flask_htmlRender       pyScript
flask2                 flask4                 flask6                 flask_imageShow
(/Users/karthibalasundaram/Downloads/tensorflow-test/env) karthibalasundaram@192 code % cd flask
(/Users/karthibalasundaram/Downloads/tensorflow-test/env) karthibalasundaram@192 flask % cd ..
(/Users/karthibalasundaram/Downloads/tensorflow-test/env) karthibalasundaram@192 code % cd flask6
(/Users/karthibalasundaram/Downloads/tensorflow-test/env) karthibalasundaram@192 flask6 % ls
app.py                 filled.png             mask.png               static                 testings
diff_countour.png      first.png              second.png             templates
(/Users/karthibalasundaram/Downloads/tensorflow-test/env) karthibalasundaram@192 flask6 % python3 app.py
 * Serving Flask app 'app'
 * Debug mode: on
WARNING: This is a development server. Do not use it in a production deployment. Use a production WSGI server instead.
 * Running on http://127.0.0.1:5000
Press CTRL+C to quit
 * Restarting with watchdog (fsevents)
 * Debugger is active!
 * Debugger PIN: 998-440-521
127.0.0.1 - - [10/Mar/2023 14:22:35] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [10/Mar/2023 14:22:35] "GET /static/css/style.css HTTP/1.1" 200 -
127.0.0.1 - - [10/Mar/2023 14:22:35] "GET /static/javascript/cyberQuotes.js HTTP/1.1" 200 -
127.0.0.1 - - [10/Mar/2023 14:22:36] "GET /favicon.ico HTTP/1.1" 404 -
127.0.0.1 - - [10/Mar/2023 14:22:43] "GET / HTTP/1.1" 200 -
127.0.0.1 - - [10/Mar/2023 14:22:43] "GET /static/css/style.css HTTP/1.1" 304 -
127.0.0.1 - - [10/Mar/2023 14:22:43] "GET /static/javascript/cyberQuotes.js HTTP/1.1" 304 -
name1: AI.png
name2: flask2.png
Similarity Score: 43.628%
MSE for the grayScale images: 70.18900317460317
Not duplicate results obtained using pHash and cv2
Printing the result1 for reference Uploaded images are not duplicates, Upload Successful
127.0.0.1 - - [10/Mar/2023 14:22:55] "POST /image-check HTTP/1.1" 302 -
check today1 /static/not_duplicates/2023_03_10_14h_22m_54s/AI.png
check today2 /static/not_duplicates/2023_03_10_14h_22m_54s/flask2.png
127.0.0.1 - - [10/Mar/2023 14:22:55] "GET /result?directoryc=not_duplicates%2F2023_03_10_14h_22m_54s&name1=AI.png&name2=flask2.png&result=Uploaded+images+are+not+duplicates,+Upload+Successful HTTP/1.1" 200 -
127.0.0.1 - - [10/Mar/2023 14:22:55] "GET /static/not_duplicates/2023_03_10_14h_22m_54s/AI.png HTTP/1.1" 200 -
127.0.0.1 - - [10/Mar/2023 14:22:55] "GET /static/not_duplicates/2023_03_10_14h_22m_54s/flask2.png HTTP/1.1" 200 -
```