A VISION-BASED INTELLIGENT SYSTEM FOR EYE CLINIC AUTOMATION

FINAL PRESENTATION

Team Members

Zhong Xiaohui
Zheng Xiaolan
Lim Chang Siang
Li Zhenghao, Kelvin

Agenda

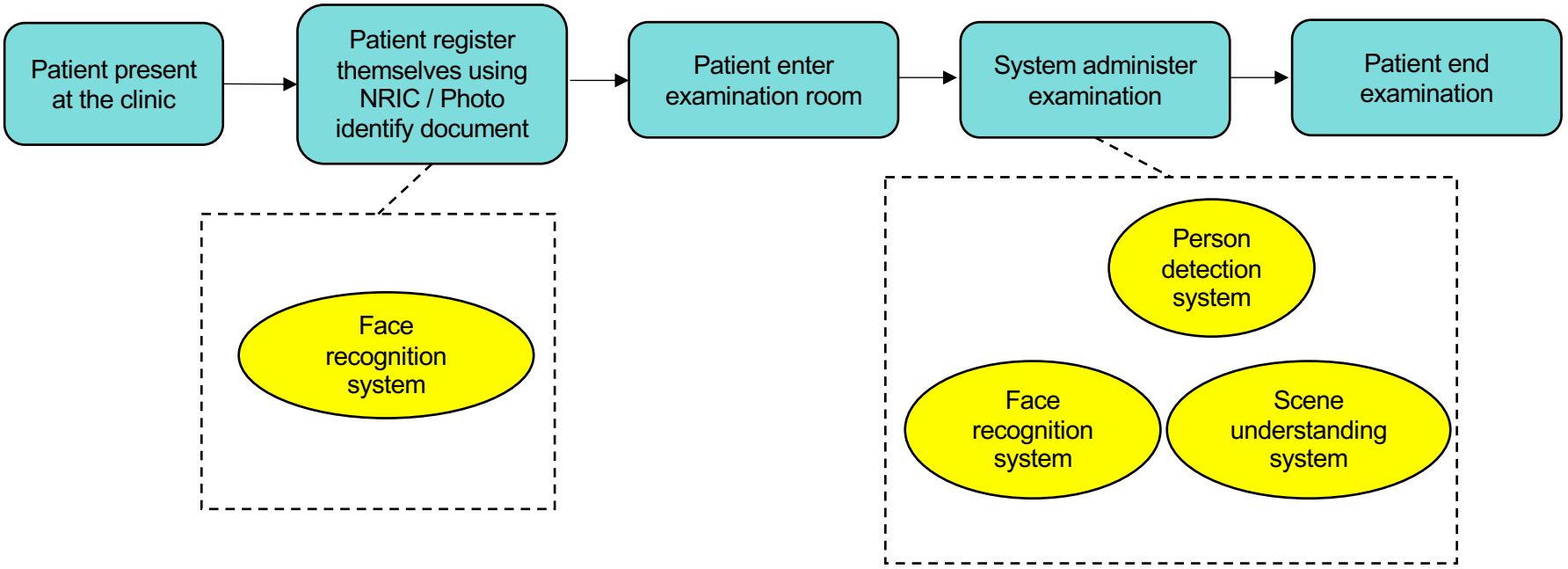
- Introduction
- Data collection process
- Facial Recognition System
- Face Detection System
- Scene Understanding System
  - Glasses Detection
  - Action Detection
- Conclusion

Introduction

- We developed a vision-based system to facilitate the development of an automation system in the eye clinic
- The objective of our solution is to,
  - To perform facial recognition and re-identify person
  - To detect number of persons in the scene
  - To identify the objects present on person's face and action perform by the person

Introduction

Overview of the System



Data Collection

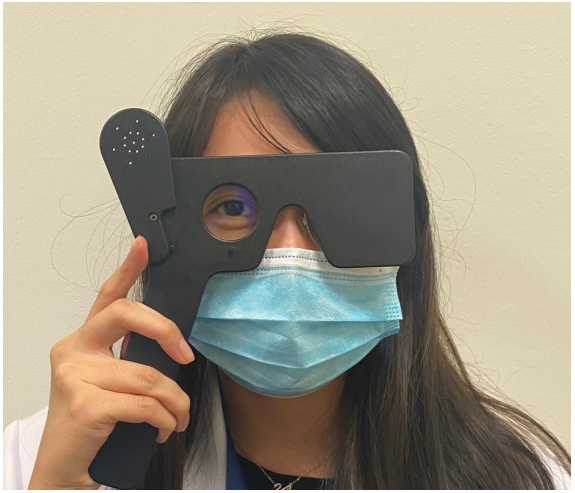
- Face Recognition System and Face Detection System
  - Labeled Face in the Wild
    - 1140 Images
    - 5 Person
- Scene Understanding System
  - Self created dataset

Data Collection

The eye occluder is a necessary piece of the eye examination equipment, to ensure that one eye is correctly assessed at a time.

The eye occluder consist of a
- handle,
- an aperture,
- a movable perforated accessory which acts as a pinhole.

Principles for dataset creation
1. As close to real-world use as possible
2. Encompass all 4 positions of eye occluder



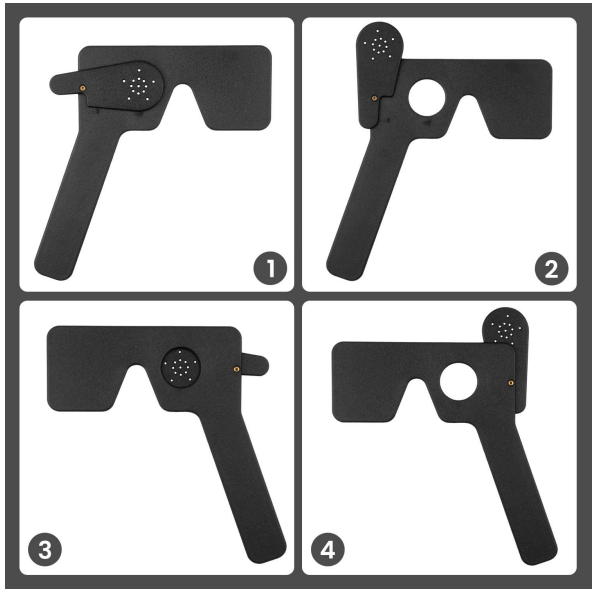
Data Collection

Real-world representation

- Presence of facial mask
- Presence of spectacles
- Long hair (e.g., females) vs short hair (e.g., males)Holding posture
- Young vs old

All positions of occluder needed in the dataset

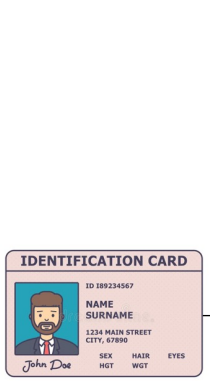
- Right eye occluder without pinhole
- Left eye occluder without pinhole
- Right eye occluder with pinhole
- Left eye occluder with pinhole



Data Collection

- All images were captured using an Apple iPhone 11 camera (Cupertino, California, United States).
- The images were taken in portrait mode, at a resolution of 3024 x 4032 pixels, in RGB mode.
- Files were saved as .jpeg format in the database.
- A total of 40 images were collected from 10 participants in 4 poses

Person Recognition System



Frame Decoder

Face Detector

Pretrained-CNN[2]

HoG Face Detection[1] (pretrained)

Feature Extractor (Face)

Save

Database

Optical Character Recognition

Save

Retrieve

Distance Measure

Siamese Network

Compare

Frame Decoder

Face Detector

Feature Extractor (Face)

HoG Face Detection[1] (pretrained)

Pretrained-CNN[2]

Existing Face data

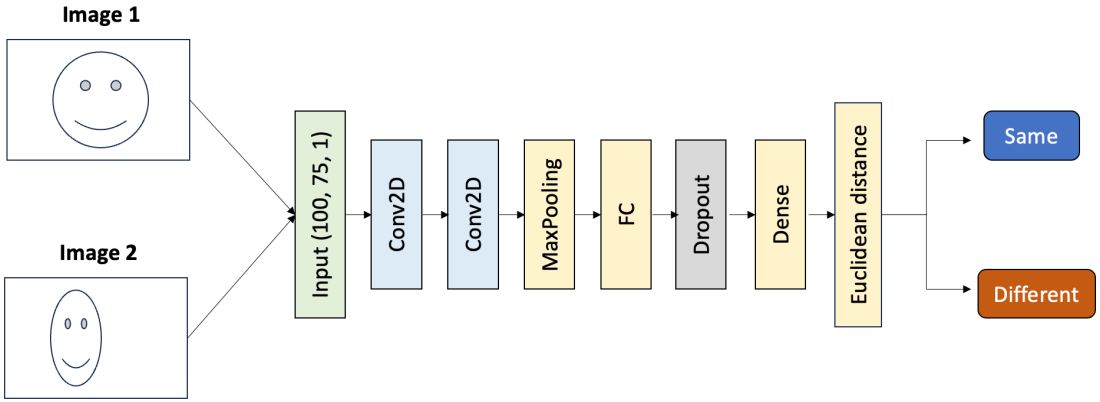
Patient Identity

Person Recognition System

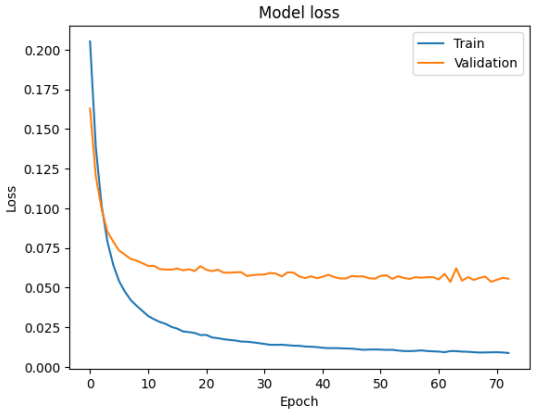
- Dataset: Labeled-Face in the Wild (LFW)
  - 1140 images from 5 individuals were used
- Approaches
  - Histogram of Gradient (HoG)
  - ResNet50 trained using ImageNet
  - VGG19 trained using ImageNet
  - VGGFace
  - CNN model trained using Siamese Network
- Distance Measure: Euclidean Distance
- Re-identification using gallery method

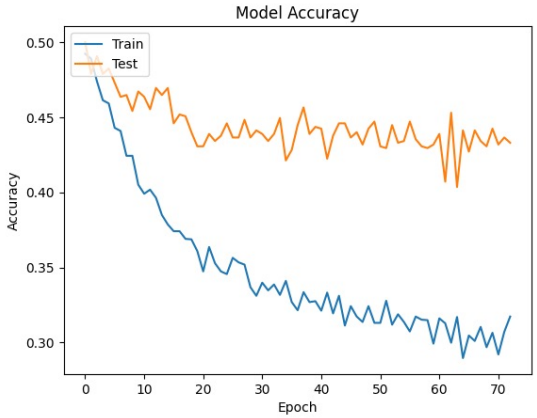
Person Recognition System

Siamese Network Architecture



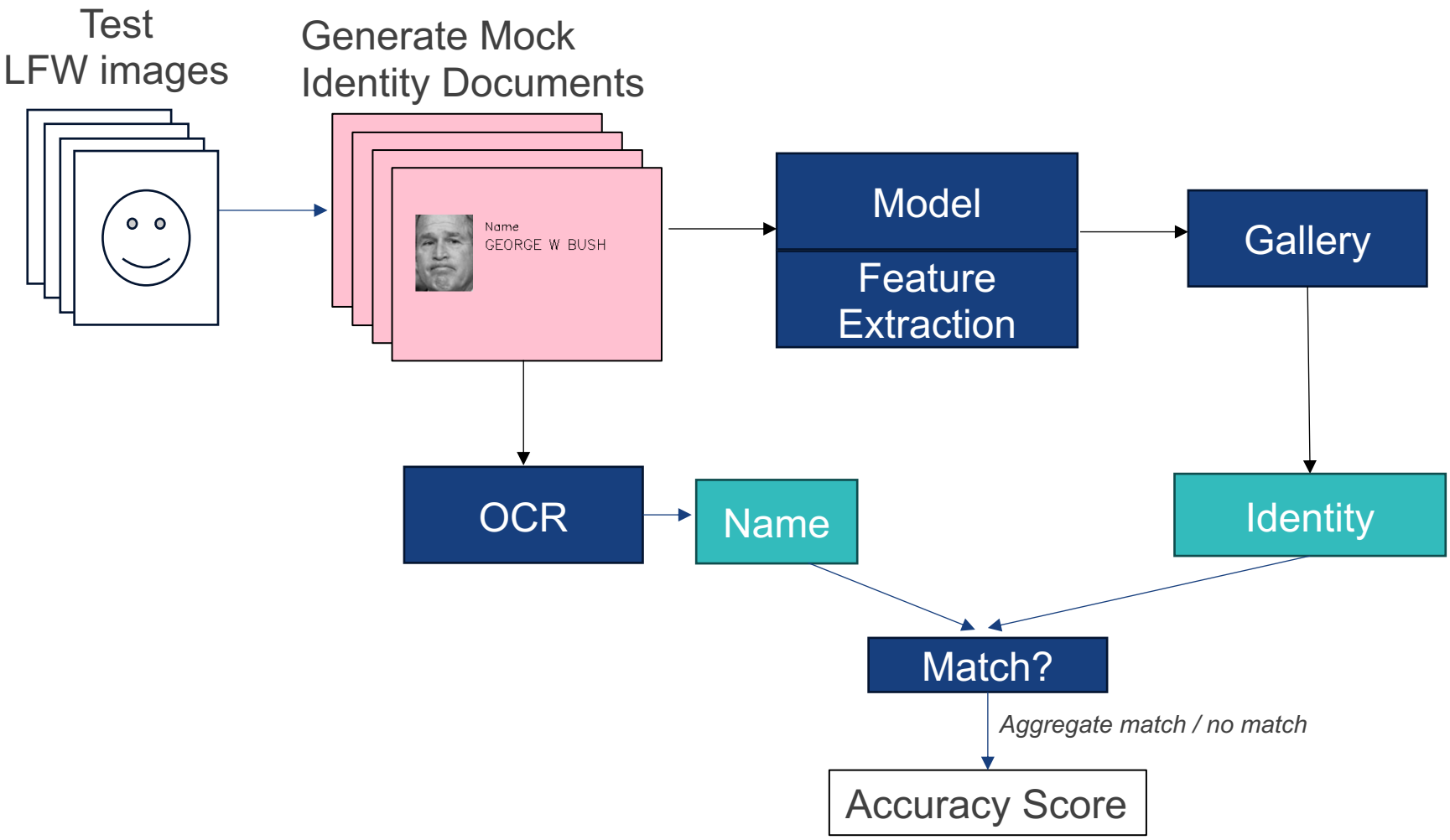
Training Loss and Accuracy





Person Recognition System

- Experiment Approach



Person Recognition System

- We compared various feature extraction methods and found that VGGFace approach produce the best accuracy score

Approach	Accuracy Score
HoG	0.4647
ResNet50	0.5017
VGG19	0.3160
VGGFace	0.6690
Custom CNN	0.1101

Person Detection System

In this project, face detection is required to count the number of people in the scene and ensure that only one person is present before proceeding with the subsequent process.

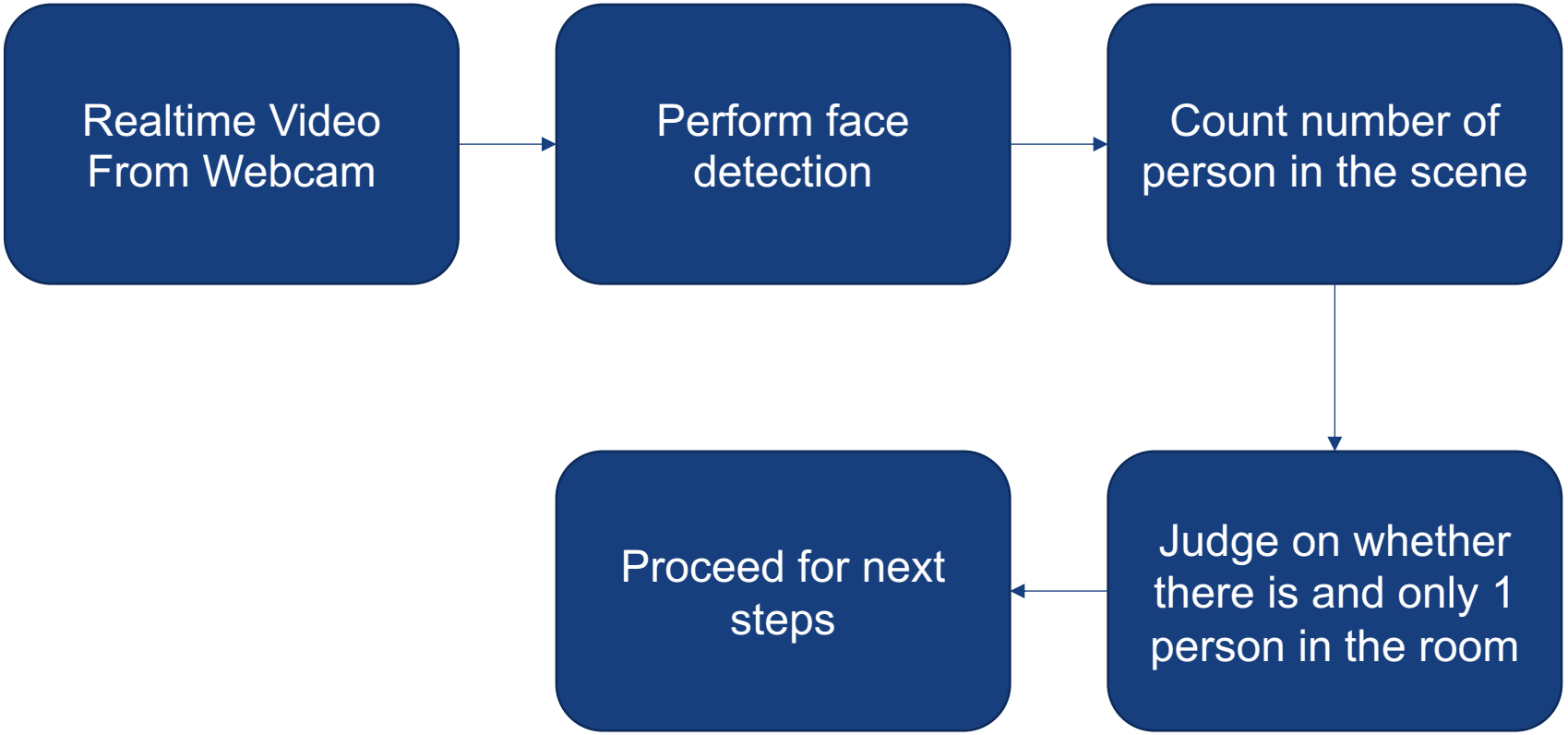
Methods for face detection method:

- CNN
- Eigenfaces
- Fisherfaces
- PCA and SVM
- Haar Cascade

We selected Haar Cascade for our project because this method is easy to deploy, fast and accurate in response and high efficiency in computing.

Person Detection System

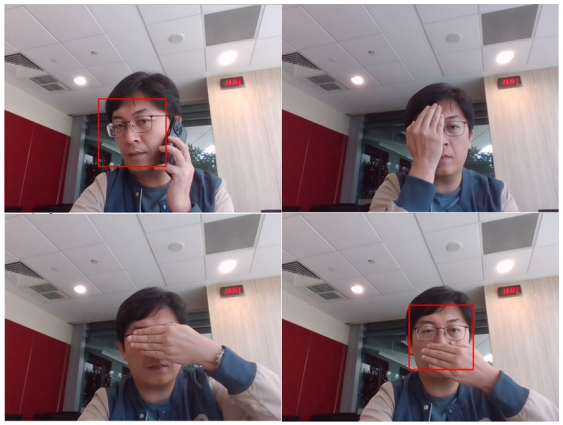
Process flow for Face Detection



Person Detection System

Weakness and Limitations

Different angles of the face and movement



Occlusion of face

Different lighting conditions

Test result :

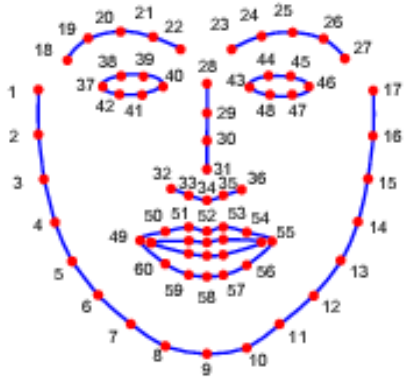
1. Able to perform face detection up to a certain angel, and when face moves very fast, the box not able to display at actual location. However, in actual scenario we will request patient not to perform fast movement and, face to the camera to avoid this problem.
2. Able to perform face detection up to cover a larger part of face, If the coverage of the face is less than half, the detection is accurate.
3. Works well under both weak light environment and bright light environment, however in actual scenario we will recommend enough brightness.

Scene Understanding System

Glasses Detection

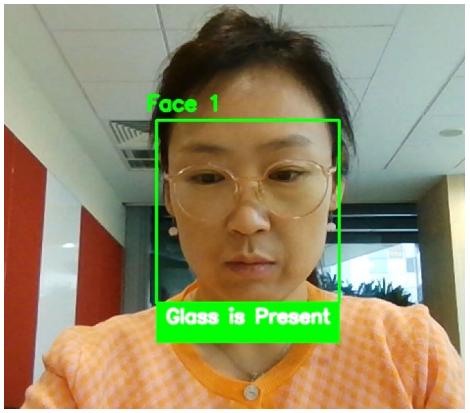
We used hand-crafted techniques

1) Face landmark detection
2) Canny edge detection
3) Determine glasses present or not



68 points facial landmark

Canny Edge Detection
after Gaussian Blur



Determine glasses
present or not

Scene Understanding System

Segment Anything Mask Generation For Occluder

- To determine whether the patient operate eye occluder following the instruction or not, we need detect its orientation.
- We adopted segment anything model to generate occluder masks and created a SVM classifier to determine the orientation from mask.
- From the original dataset, we generated 60 occlude masks and 1092 non occlude masks by SAM. Which provide training data for SVM classifier.

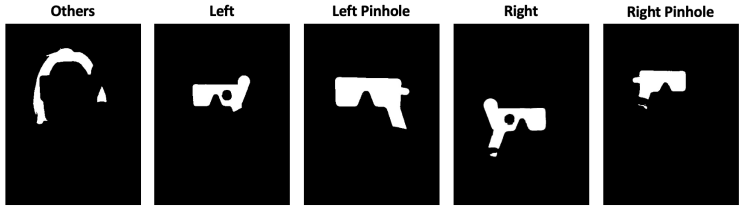
Masks for occluder

Masks for non-occluder

Masks generated from single image

Scene Understanding System

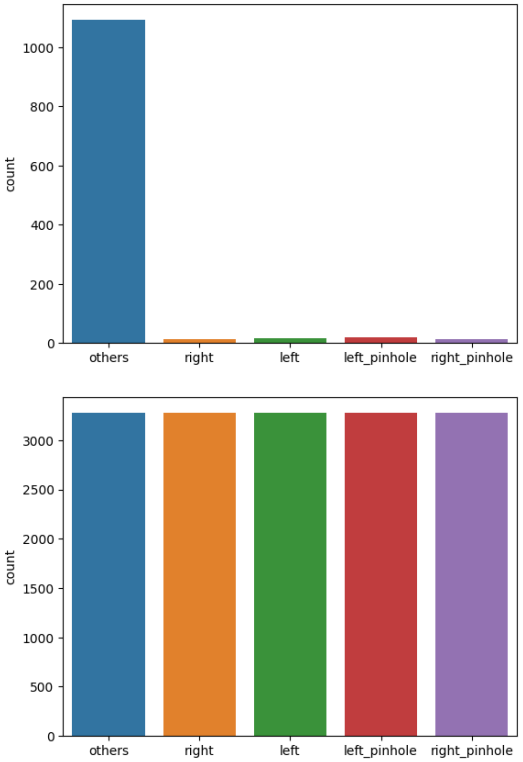
Masks generated using SAM model



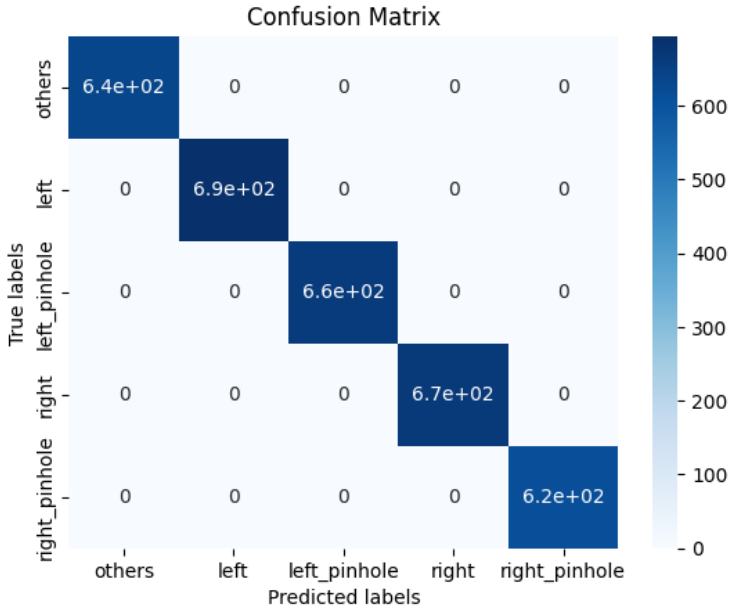
Train SVM classifier

Model for occluder action detection

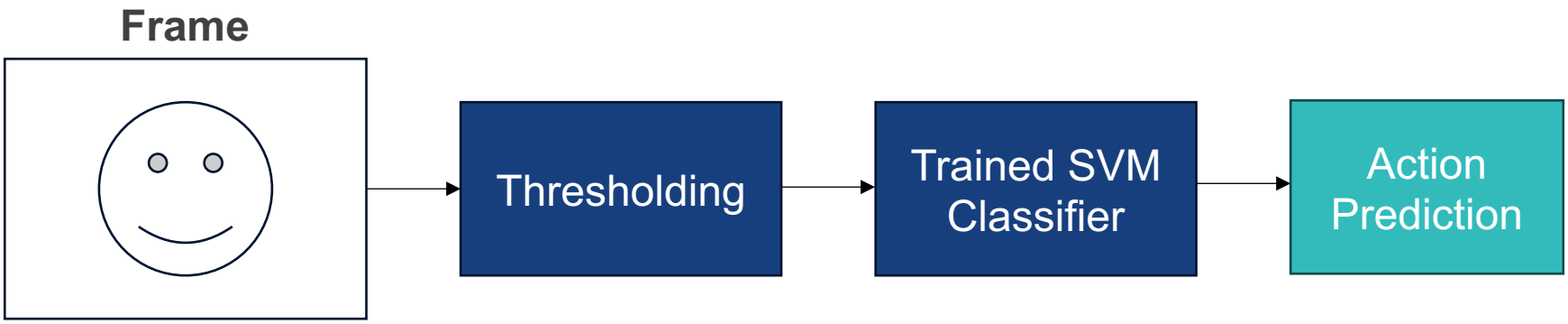
Resample under-represented class



Model evaluation



Scene Understanding System



Predicted “right”

Predicted “left”

Conclusion

- We have delivered with the following features
  - Facial Recognition System
    - Using OCR and VGGFace feature extractor
  - Face Detection System
    - Using Haar Cascade Classifier and report the count
  - Scene Understanding System
    - Glasses Detection: Edge Detection
    - Action Detection: Image Segmentation and SVM classifier

Thank you