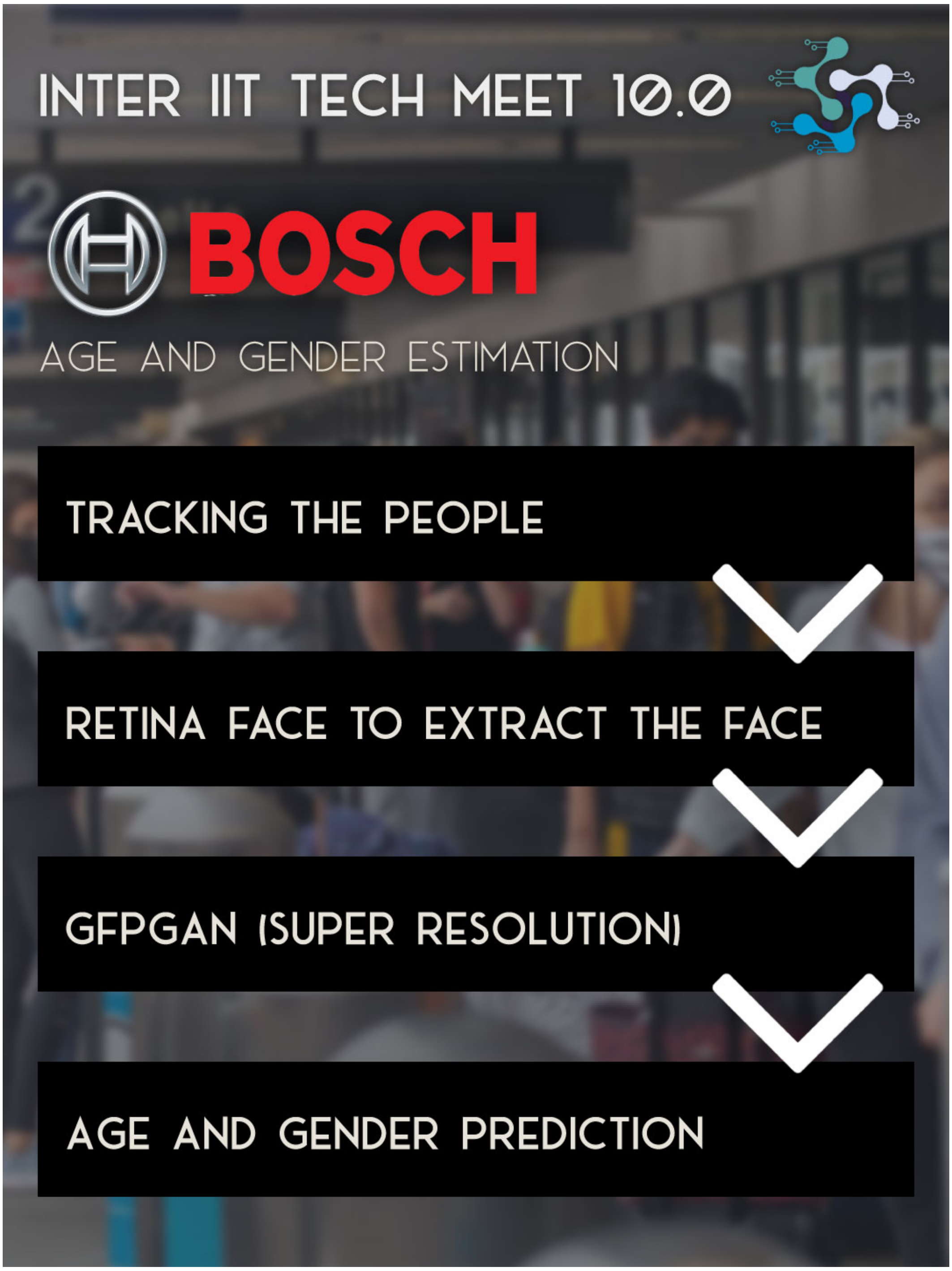
Inter-IIT Bosch’s Age and Gender Estimation



INTER IIT TECH MEET 10.0

BOSCH

AGE AND GENDER ESTIMATION

TRACKING THE PEOPLE

RETINA FACE TO EXTRACT THE FACE

GFPGAN (SUPER RESOLUTION)

AGE AND GENDER PREDICTION

Introduction

The scenes obtained from a surveillance video are usually with low resolution. Most of the scenes captured by a static camera are with minimal change of background. Objects in outdoor surveillance are often detected in far-fields. Most existing digital video surveillance systems rely on human observers for detecting specific activities in a real-time video scene. However, there are limitations in the human capability to monitor simultaneous events in surveillance displays.
Hence, human motion analysis in automated video surveillance has become one of the most active and attractive research topics in the area of computer vision and pattern recognition.

Problem Statement

Build a solution to estimate the gender and age of people from a surveillance video feed (like mall, retail store, hospital etc.). Consider low resolution cameras as well as cameras put at a height for surveillance.

Approach:

1. Person Tracking and Re-identificaion using DeepSort based on YOLO-v5
2. Face Detection using RetinaFace
3. Facial Super-Resolution using GFPGAN based on ESRGAN
4. Age & Gender Estimation models trained based on FaceNet as a feature extractor

Project Tree:

```
— . / MP_BO_T2_CODE
├── 0.2.0.3
├── Age_Gender_Estimation
├── GFPGAN
├── Results
├── Retina_Face
├── Yolov5_DeepSort
├── pycache
├── age_gender.py
├── face_detection.py
├── face_sr.py
├── gfpgan.egg-info
├── inference
├── main.py
├── make_csv.py
├── person_extraction.py
├── requirements.txt
├── tree.txt
├── utkface-new
└── yolov5m.pt

    9 directories, 10 files
```

How to run

Note: This project needs **cuda** to run the code

1. Change the working directory to the **MP_BO_T2_CODE**
2. Make sure that you fulfill all the requirements: Python 3.8 or later with all requirements.txt dependencies installed, including torch>=1.7. To install, run:

   ```
   pip install -r requirements.txt
   ```

3. To generate the csv files as an output for a given video

```
python3 main.py --video 'path/../video_name.mp4' --visual 1
```

Note: 'path/../video_name.mp4' is relative path to the video file with **video_name** being the name of video file

- —visual is the optional argument to save the resulting inferences with the annotation video in the **results/** folder with the name **video_name_infer.mp4. (0** flag for not saving, **1** for saving the video**)**

4. To generate the csv files as an output for a given image

```
python3 main_image.py --image 'path/../image_name.mp4' --visual 1
```

Note: 'path/../image_name.mp4' is relative path to the video file with **image_name** being the name of image file

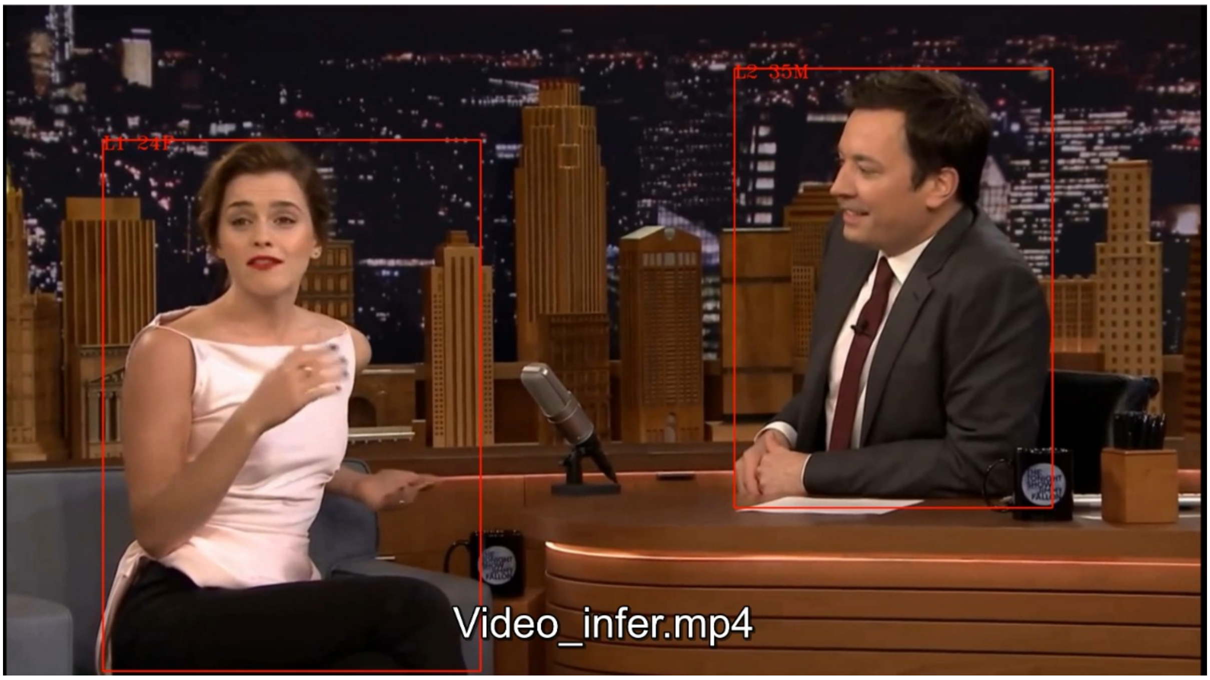
- —visual is the optional argument to save the resulting inferences with the annotation video in the **results/** folder with the name **video_name_infer.png (0** flag for not saving, **1** for saving the video**)**

Running above command in terminal will generate the csv file of the inferences in the **results/** folder with the name **video_name.csv**

Format of the output csv:

`frame num,person id, bb_xmin, bb_ymin, bb_height, bb_width, age_min, age_max, age_actual, gender`

Sample Inferences



Sample results can be found in the given folder

Bosch_age_gender_sample_results - Google Drive

https://drive.google.com/drive/folders/1sT-X4NJQyJZMsg2ytsyZZz3s7zCWabOC?usp=sharing

Detailed Approach

- We are using **DeepSort** algorithm for person tracking and Re-identification built on the backbone of the YOLO_v5. The images of people tracked along with bounding boxes and ids of unique people for each frame are saved.
- From the images of person, face is extracted using RetinaFace model, which subsequently goes through super-resolution to be fed into the age and Gender prediction model.
- RetinaFace is cutting edge technology for for face detection which uses facial landmarks for face detection
- For super resolution we have used GFPGan which is built upon ESRGan and and is used for real world face restoration.
- Face restoration helps recover details of facial features lost due to interpolation and while cropping the face out of the low quality security feed.

- The enhanced image is then passed through FaceNet which is has the architecture of an Inception ResNet V1 model trained with softmax loss on VGGFace2 dataset consisting on ~3.3M images.
- FaceNet extracted features from the image of face and makes embeddings which which will be used for prediction. We have trained two separate Neural Networks, one of which performs regression for age estimation and other one performs binary classification for gender prediction.
- At each point the ids of unique persons are maintained and at the end average of all predictions is taken. Output is a csv with bounding boxes, person ids for each frame along with age and gender estimates. These are also used to annotate the video for visual output on the input video.