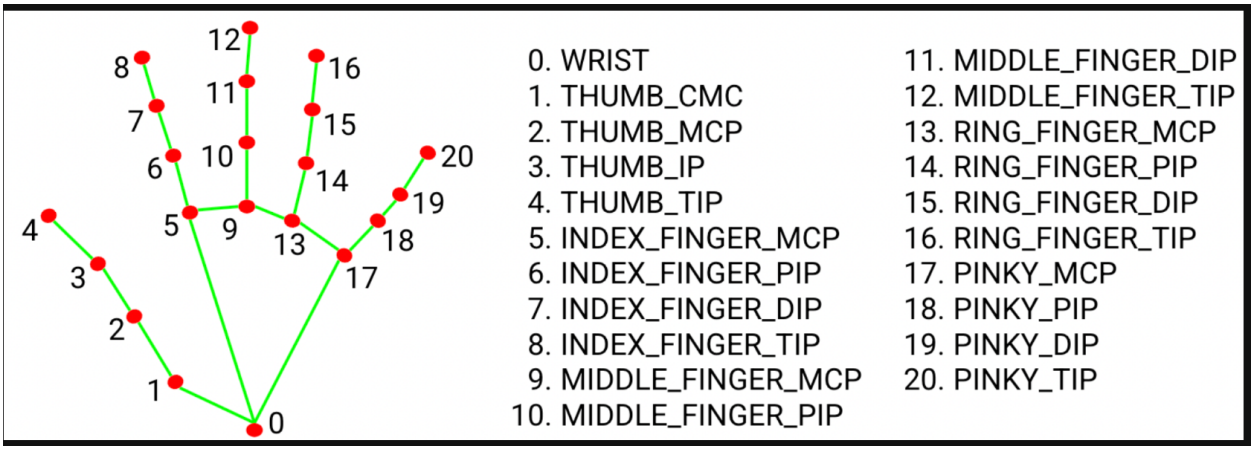
Readme

Mediapipe Models

hand landmark model

The hand landmark model bundle detects the keypoint localization of 21 hand-knuckle coordinates within the detected hand regions. The model was trained on approximately 30K real-world images, as well as several rendered synthetic hand models imposed over various backgrounds.

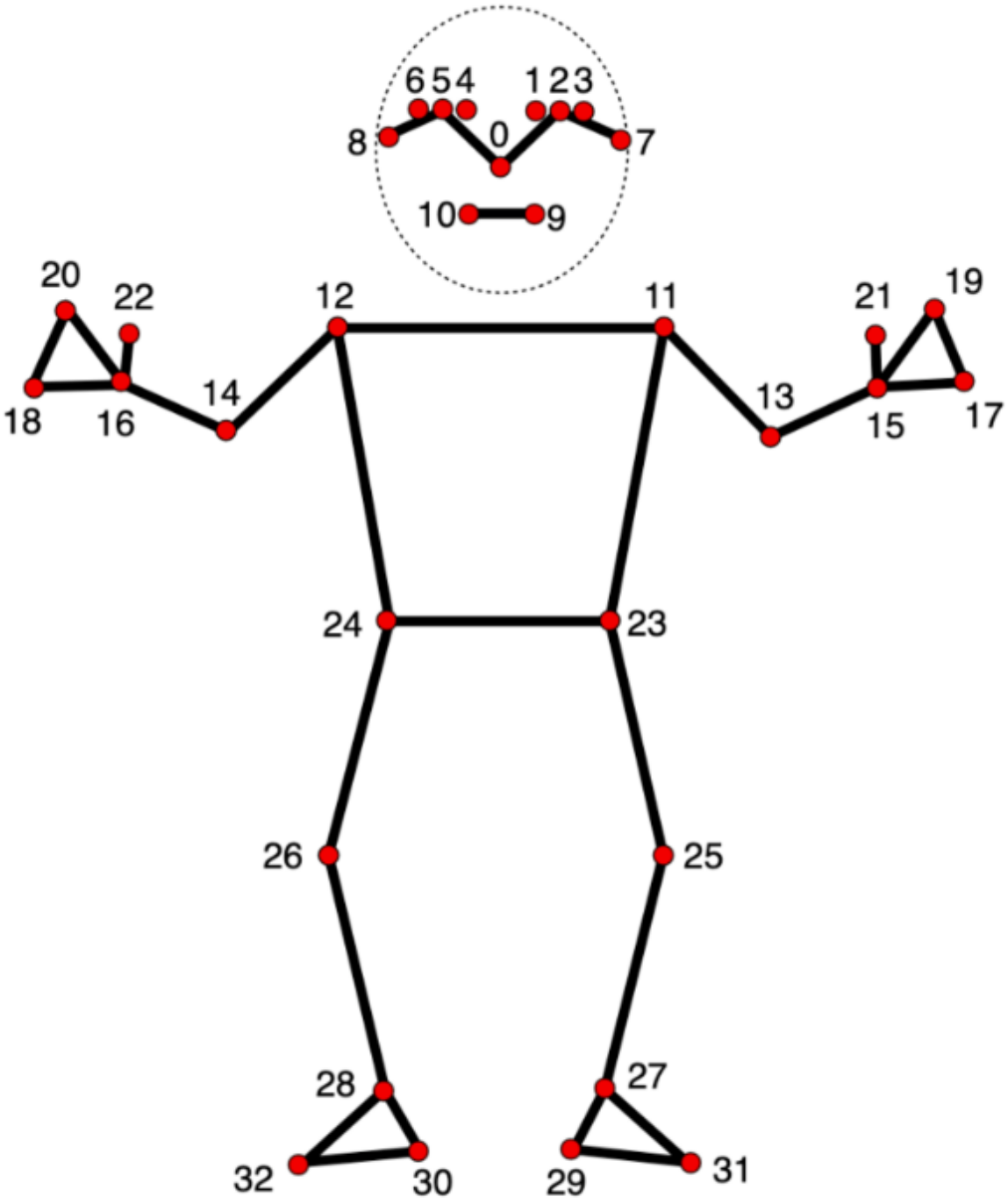


The hand landmarker model bundle contains a palm detection model and a hand landmarks detection model. The Palm detection model locates hands within the input image, and the hand landmarks detection model identifies specific hand landmarks on the cropped hand image defined by the palm detection model.

Since running the palm detection model is time consuming, when in video or live stream running mode, Hand Landmarker uses the bounding box defined by the hand landmarks model in one frame to localize the region of hands for subsequent frames. Hand Landmarker only re-triggers the palm detection model if the hand landmarks model no longer identifies the presence of hands or fails to track the hands within the frame. This reduces the number of times Hand Landmarker tiggers the palm detection model.

Pose landmarker model

The pose landmarker model tracks 33 body landmark locations, representing the approximate location of the following body parts:

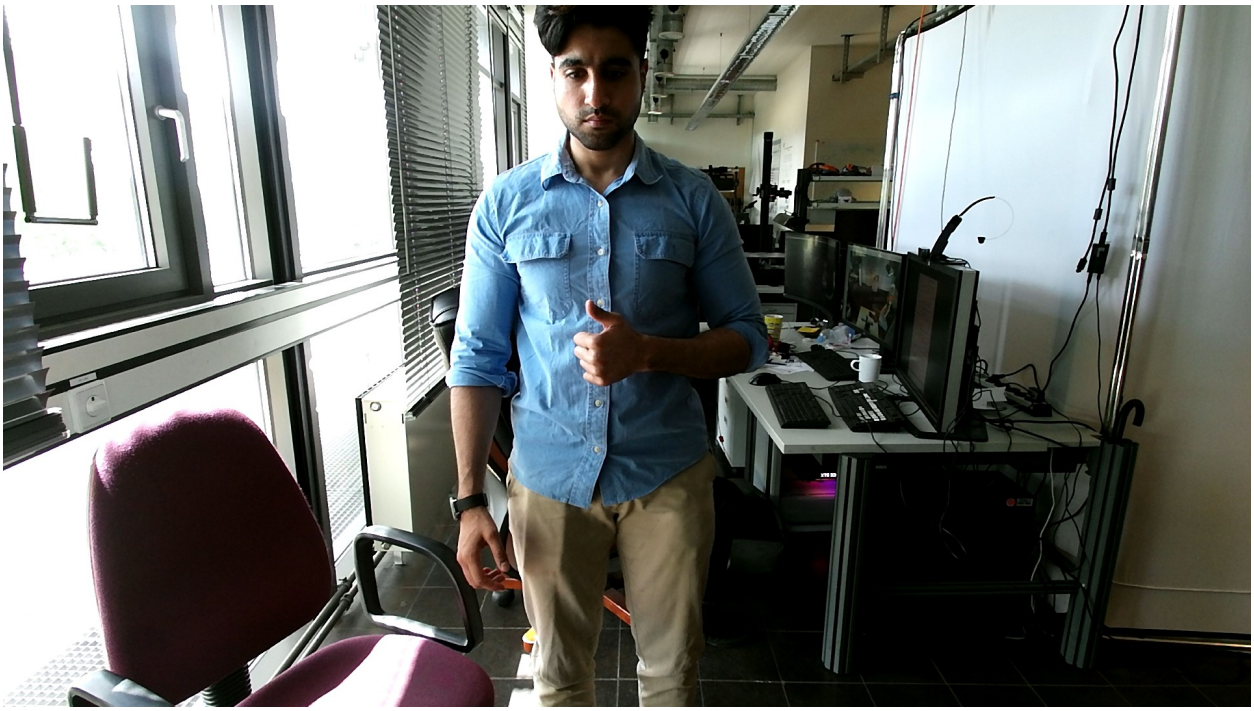


0 - nose
1 - left eye (inner)
2 - left eye
3 - left eye (outer)
4 - right eye (inner)

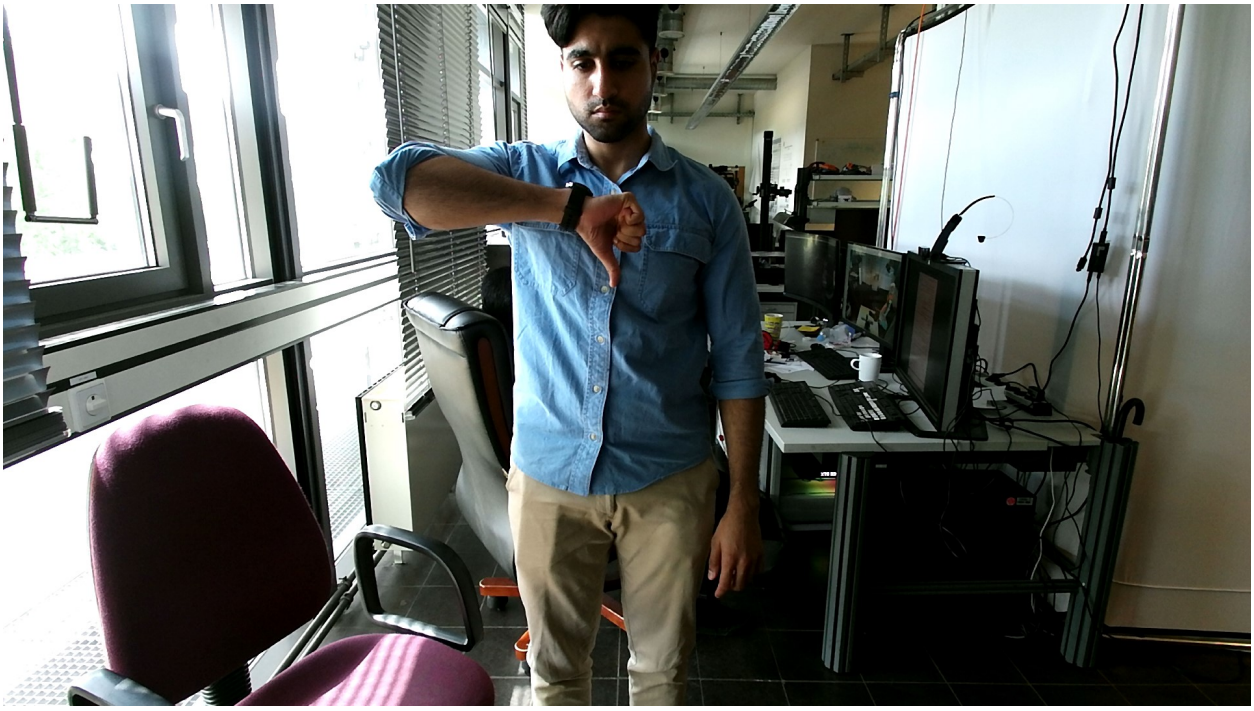
5 - right eye
6 - right eye (outer)
7 - left ear
8 - right ear
9 - mouth (left)
10 - mouth (right)
11 - left shoulder
12 - right shoulder
13 - left elbow
14 - right elbow
15 - left wrist
16 - right wrist
17 - left pinky
18 - right pinky
19 - left index
20 - right index
21 - left thumb
22 - right thumb
23 - left hip
24 - right hip
25 - left knee
26 - right knee
27 - left ankle
28 - right ankle
29 - left heel
30 - right heel
31 - left foot index
32 - right foot index

Data Collection for the training

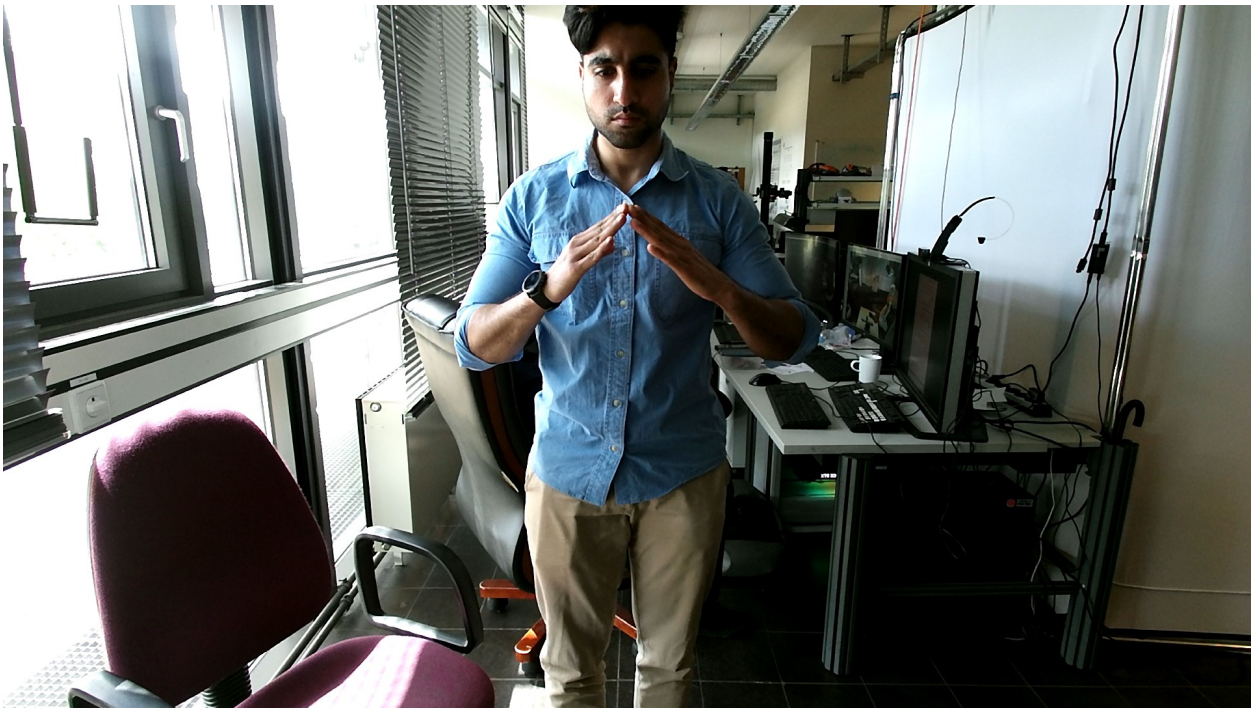
Pose 0: Lock Person



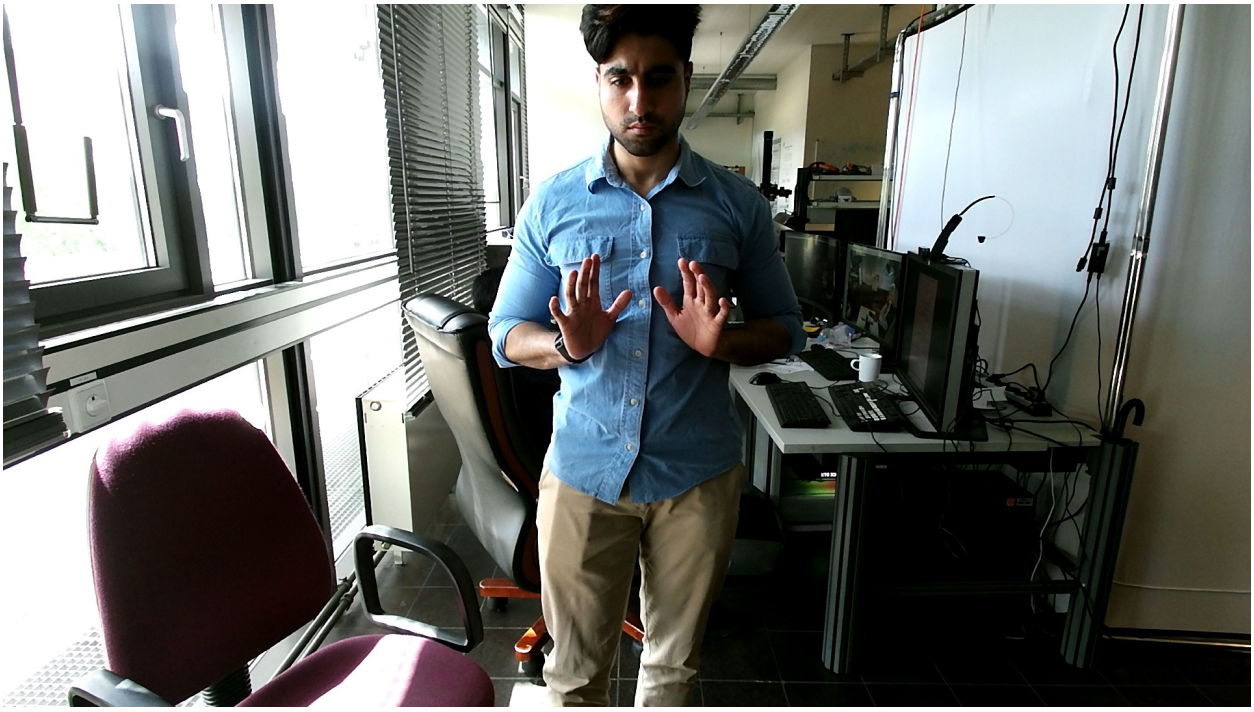
Pose 1: Unlock Person



Pose 2: Go to Base

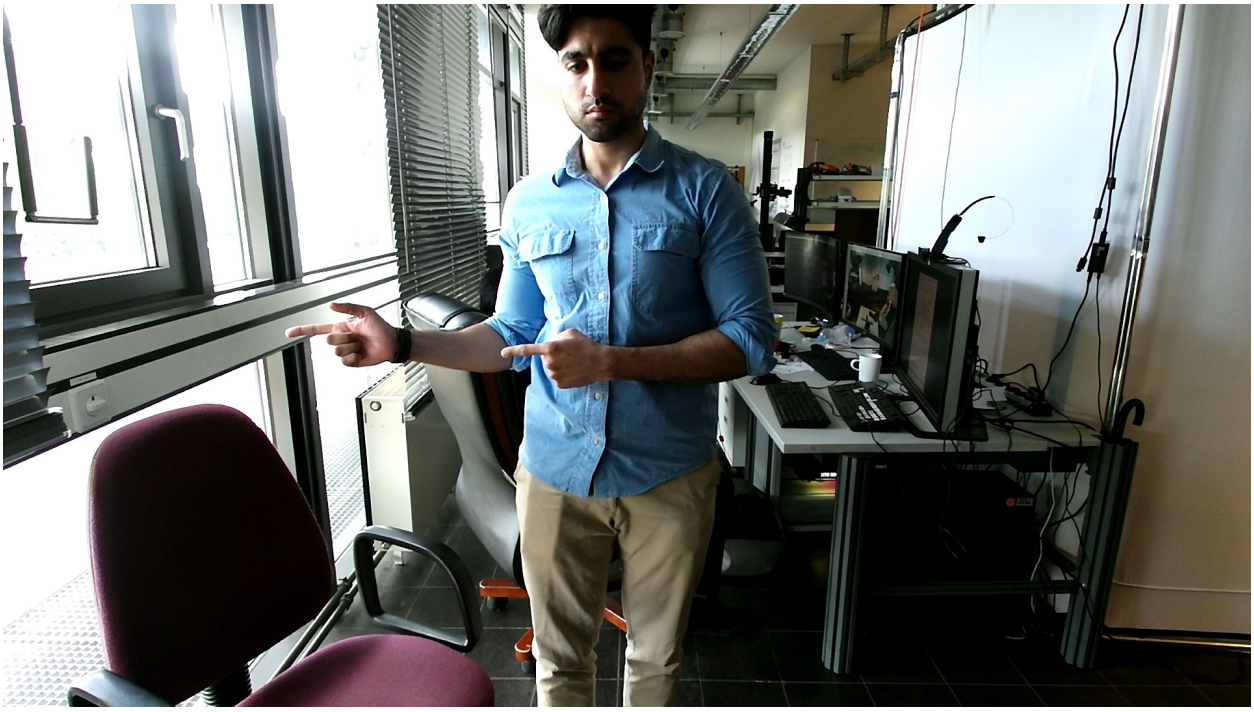


Pose 4: Stop

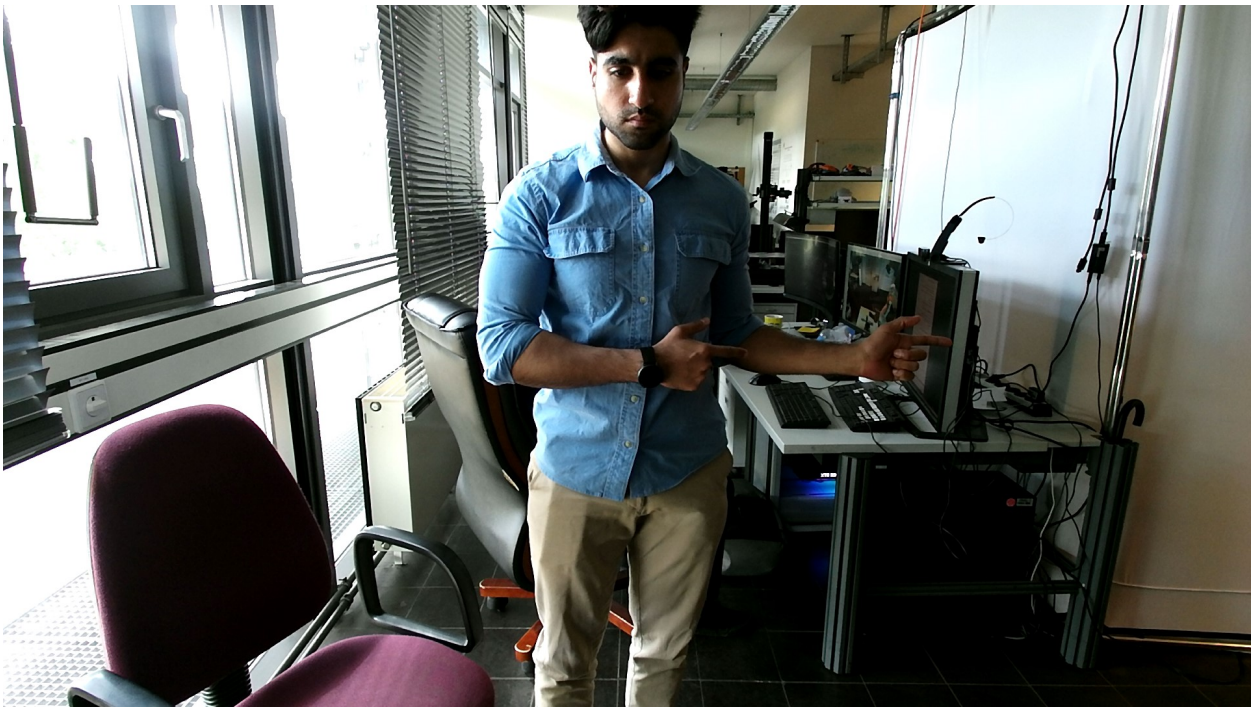


Pose 3: Follow

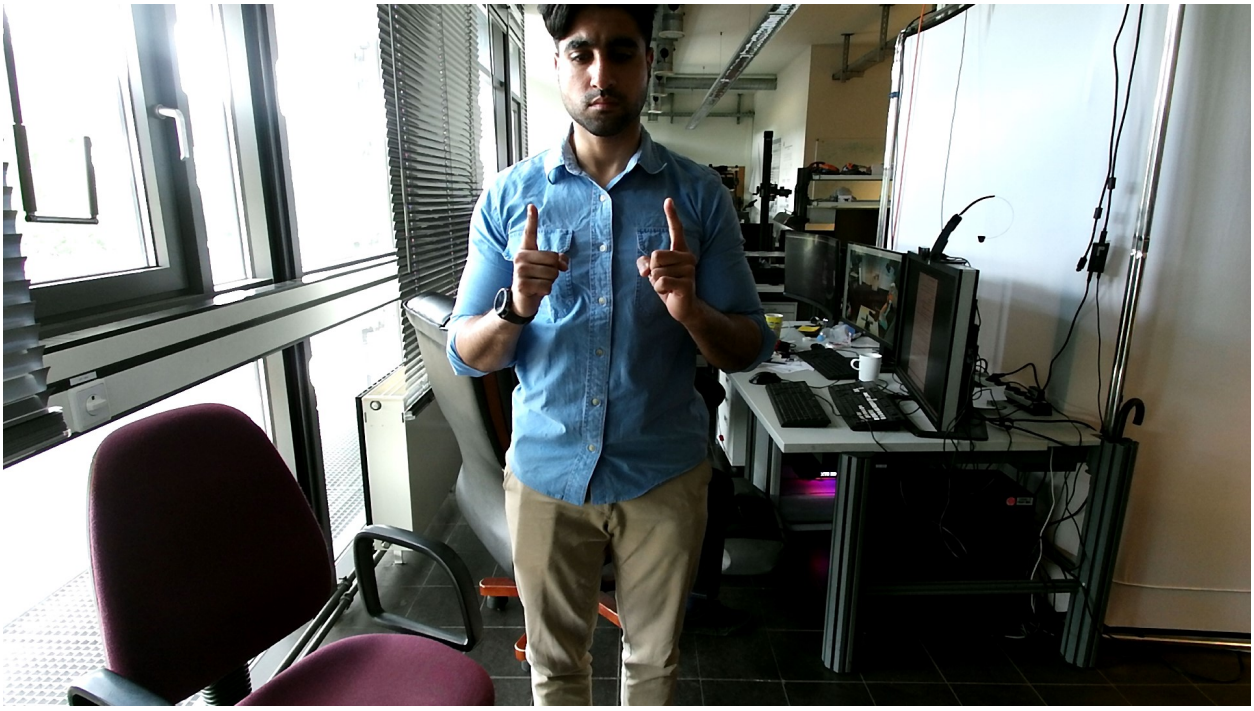
Pose 5: Turn Left



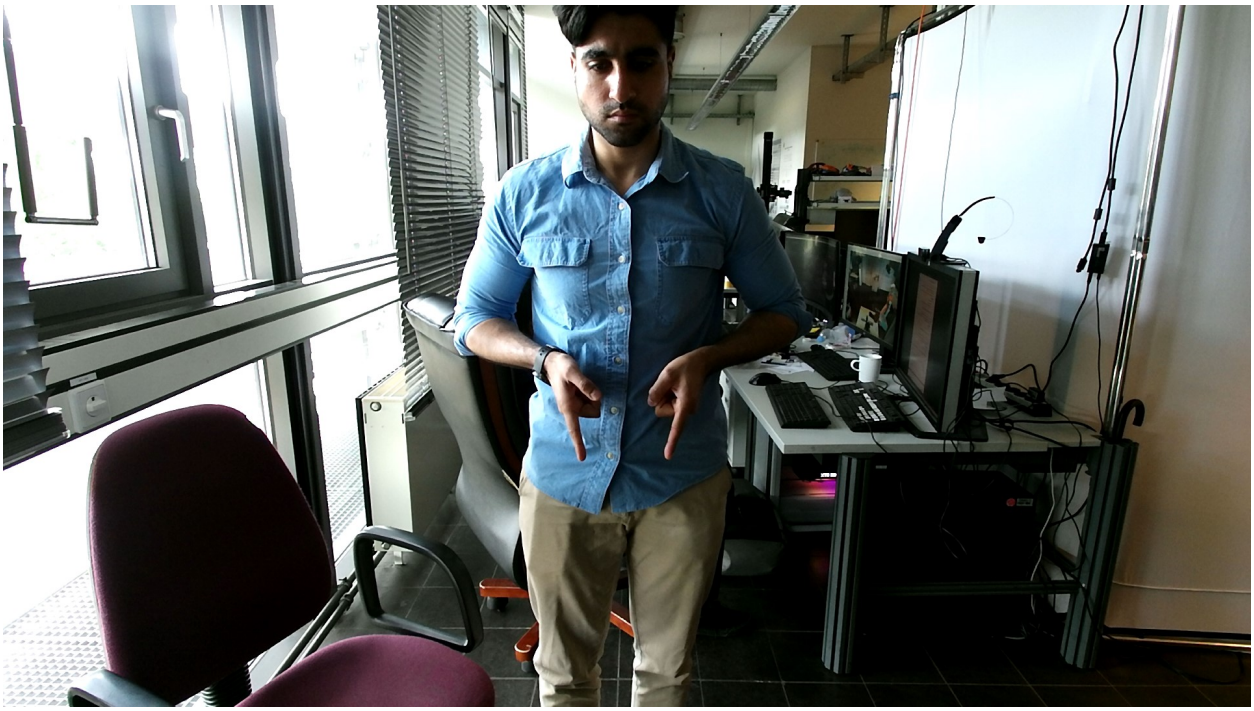
Pose 6: Turn Right



Pose 7: Move Forward



Pose 8: Move Backward



In this repository, I have created 5 models which are based on the Mediapipe holistic model for landmark KeyPoint detection and extraction. Using this we get:

- 21 KeyPoint for each hand with (x, y, z) coordinated making a total of 126 features
- 33 KeyPoint for the pose of the body with (x, y, z, visibility) coordinated making in total 132 features

I have used Preprocessing techniques such as:

- Normalization
- Relative Coordinate transformation

I have used data augmentation techniques to increase the data robustness as I collected the data myself like:

- Rescaling frame
- Rotating frame
- Shifting frame
- Adding Gaussian random noise etc.

I have added 5 Jupiter notebook files for code that are:

- DHGR_6_NPP_SM_30F
- DHGR_6_PP_LM_30F
- DHGR_9_NPP_LM_30F_hol
- DHGR_9_NPP_LM_60F
- DHGR_9_PP_LM_30F

Here, DHGR: Dynamic Hand Gesture Recognition, 6 or 9 represent the no. of gestures that it can recognize, PP or NPP means Preprocessing applied or not, LM or SM represent large LSTM model or small LSTM model, 30F or 60F represent the no. of frames used to predict a gesture.