Get input type

1: Image
2: video
3: Live camera stream

1   2   3

Get input path

Get camera ID

Check if input is image or video

Image   Video

read Image

extract image frames

resie image to required size

process image frame to remove noise

parse the processed image through pretrained YOLO model

draw bounding boxes for the detected humans in the image

transform the coordinates of the center of the bounding boxes from image plane to robot's inertial reference frame

store the output coordinates in a text file