

Human Detection

● Model Details

This model has approximately 260K parameters. The human detection model trained through YOLOv5 has only a single class, Human. It can capture images from a webcam in real-time and perform inference, supporting more than 5 different poses. On the VA8801, the inference speed is 50FPS, with an accuracy higher than 90%.



Detecting the position and number of people

● Model Specifications

Model Type: Convolutional Neural Network

Model Architecture: YOLOv5n with customized backbone for VA8801

Input: 224*224*1

Output: [class, x0, y0, x1, y1]

Class: Human

Number of people: 3 (suggestion) or more

Angle of view: Suggested to be at eye level

Distance: 1 ~ 4m (suggestion)

Posture:

stand, sit, walk, bend, crouch, raise hand, turn around, ...

● Application

Human detection products: surveillance equipment, people counting.

Application example: To enhance overall performance, motion detection can be used in preprocessing to first capture moving objects. Then, crop and resize these regions to the input of the model

● Limitation

- (1) Difficult to detect beyond a distance of four meters
- (2) Not supported in lighting conditions below 30 lux.
- (3) Special actions such as lying down, bending over, or raising hands may cause a decrease in recall

● Training Data

- (1) COCO-human about 31000 images
- (2) Surveillance dataset about 2000 images

● Testing Result

(1) mAP = 97.18% when the distance 1m ~ 2.5m

(2) mAP = 95.30% when the distance 2.5m ~ 4m

*confidence level = 0.6

● Reference

<https://cocodataset.org/#home>