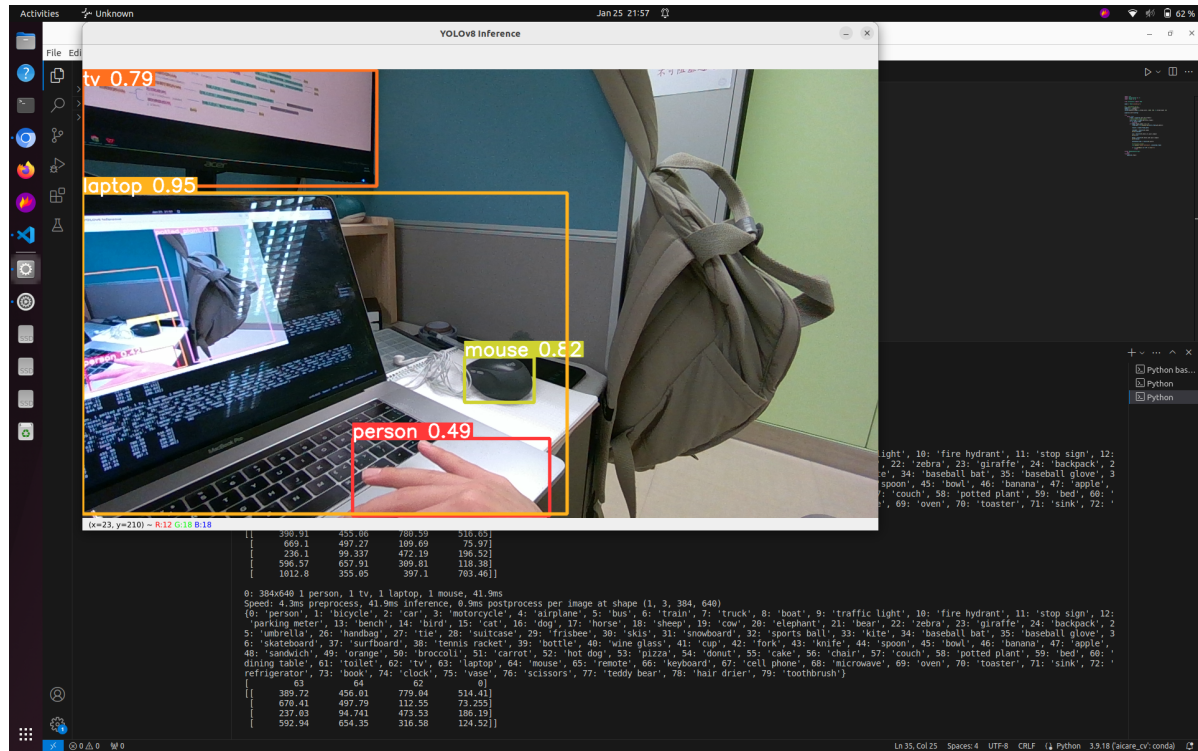


Structure2Description (01.26)

1. 现状

1.1 样例



1.2 'hand'识别正在开发中(Rong Xiang)。

2. 开发ToDo:

2.1 脚本修改, 实现以下功能: 快照, 保存image-structure pairs到文件。

3. 数据集构建的操作步骤:

3.1 挑选这80个类别中的物品---暂定47种。

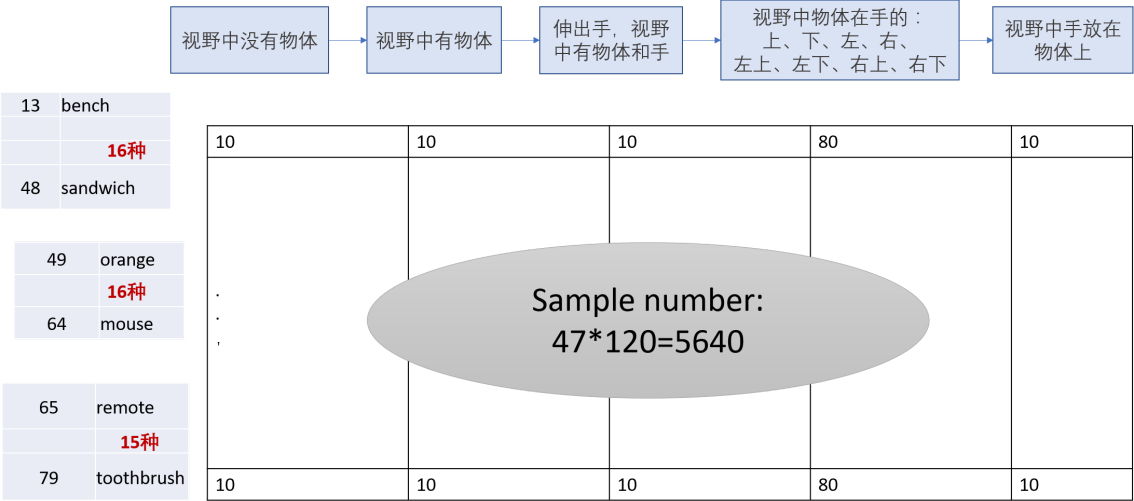
{0:'person', 1:'bicycle', 2:'car', 3:'motorcycle', 4:'airplane', 5:'bus', 6:'train', 7:'truck', 8:'boat', 9:'traffic light', 10:'fire hydrant', 11:'stop sign', 12:'parking meter', 13:'bench', 14:'bird', 15:'cat', 16:'dog', 17:'horse', 18:'sheep', 19:'cow', 20:'elephant', 21:'bear', 22:'zebra', 23:'giraffe', 24:'backpack', 25:'umbrella', 26:'handbag', 27:'tie', 28:'suitcase', 29:'frisbee', 30:'skis', 31:'snowboard', 32:'sports ball', 33:'kite', 34:'baseball bat', 35:'baseball glove', 36:'skateboard', 37:'surfboard', 38:'tennis racket', 39:'bottle', 40:'wine glass', 41:'cup', 42:'fork', 43:'knife', 44:'spoon', 45:'bowl', 46:'banana', 47:'apple', 48:'sandwich', 49:'orange', 50:'broccoli', 51:'carrot', 52:'hot dog', 53:'pizza', 54:'donut', 55:'cake', 56:'chair', 57:'couch', 58:'potted plant', 59:'bed', 60:'dining table', 61:'toilet', 62:'tv', 63:'laptop', 64:'mouse', 65:'remote', 66:'keyboard', 67:'cell phone', 68:'microwave', 69:'oven', 70:'toaster', 71:'sink', 72:'refrigerator', 73:'book', 74:'clock', 75:'vase', 76:'scissors', 77:'teddy bear', 78:'hair drier', 79:'toothbrush'}

3.2 存在的情况按照顺序:

- “视野中没有物体”
- “视野中找到了物体”
- “伸出手, 视野中有物体和手”
- “视野中物体在手: 上、下、左、右、左上、左下、右上、右下”

- “视野中手放在物体上”

3.3 一共需要抓取的image-structure数目： $47 * (10+10+10+80+10) = 5640$



4.工作计划 0128~0202:

- Yi: 1. 调整脚本便于样本收集; 2.收集13-48中可用的类别的5种情况的样本 (一部分)
- Jason: 1. 搭建LlaMA-2模型环境, 设计description prompt, 并打通fine-tuning; 2.收集49-64序号的5种情况样本 (一部分)