第五章 臂载机器人的人体识别研究

5.1 YOLO算法版本的选定

因为YOLO在实时检测的目标识别任务中，相比其他模型是非常成功并且得到大家认可的，所以本文选择了YOLO系列作为目标识别的主框架。而YOLO系列版本中，每个版本都有自己的特色，并不是高的YOLO版本一定是最优的，首先考虑YOLOv1版本算法，其模型只是简单暴力的将输入图像分成7×7一共49个的网格，其中每个网格只会设定2个候选框去进行最终的预测，这代表只会有98个区域被涉及到，有可能导致正样本的缺失，召回率的下降。该版本为初代版本，与现在对比虽然效果差一些，但在当时确实风靡一时，带动了视觉领域的发展。其算法简单有效，体量小适合移植，对背景简单负样本少的图片检测还是有效的，但在灾难环境中人们，背景复杂且镜头遮挡严重，正样本召回率和准确率均不满足实验要求，因此该版本不适合做灾害复杂场景下的识别救援。

YOLOv2版本加入了聚类先验框，可以根据数据集提前做出合适的先验框尺寸，在三种不同大小的感受野下的特征图均有较为合适的预测框。在先验框的预测位置边界也做了一定的范围限制，使训练过程中模型收敛变得更稳定，减少了模型震荡，此模型将输入图像进行了5次下采样，也就是将图像分为了13×13一共169个网格，特征通道增加到1024，对正样本的识别有了一定提升，但在激活函数以及损失函数的选择上仍有一些瑕疵，故该版本先作为替补模型。

YOLOv3版本是YOLO原作者编写的最后一个版本，也有着不错的实际应用场景，流传也较为广泛，在提升了mAP的基础上仍然保持了其单阶段检测极其优秀的速度优势。从网络结构方面来考虑，YOLOv3提供2种方案，如果选择性能可以选择darknet-53作为提取特征的骨架结构。其准确率与经典网络RestNet相近，但速度提升较多；同时也可以选择更加轻量高速的tiny-darknet作为提取特征的骨架结构，比较灵活。相比上个版本可以输出多尺度通道检测网络的3个尺寸分别为1/8,、1/16、1/32，所以可以提取和利用更细腻的特征，可以解决不同尺度的目标识别，同时基于3个大小不同的感受野用kmeans聚类出3类共9种预测框，在本实验的复杂环境下的人体识别有着不错的效果体验，但由于其输入数据的通道优化较少，导致对设备的显存要求较高，尽管我选择了epoch为最小1的情况下，3g显存仍然无法训练此模型，故本版本仍只作为替补模型。

YOLOv5版本是Ultralytics公司在2020年公开发表的，其效果也得到了原作者的认可。YOLOv5是基于YOLOv3架构的基础上改进提升得来的，其性能与稳定性均为当下最优秀的目标识别模型之一。该版本增强了性能的同时也减少了一定的计算量，在灵活性能上远高于之前的几代版本，同时也支持多种格式数据的输入。由于是公开项目，于github中有专业的团队一直进行维护，且其知名度、流传度、资料丰富性较高，经过近2两年的沉淀效果得到了大家的认可，无论从代码工作量、代码整洁性、兼容性都是非常优秀的项目。YOLOv5模型分为三部分由骨干网络、颈部网络、头部网络组成，又从模型的大小来分为了好几个版本，包括YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x等不同层数深度的模型结构，YOLOv5s为最小架构版本，其层数深度为标准版本的1/3，每层参数量为标准版本的1/2，更适合小模型学习和快速训练，YOLOv5x是较大模型，其层数深度为标准版本的1.33倍，每层参数量为标准版本的1.25倍，更适合高精识别度、更高识别效果要求的任务目标。其对感受野的不同尺度也做了很好的特征拼接与融合，对正样本的框定也做了优化，所以总体来考虑YOLOv5是不错的应用版本。