
《物联网与嵌入式技术》

大作业课题报告

基于 YOLO 算法的手势识别

学 生 姓 名： 董天零

学 科、 专 业： 智慧交通与运载工程

学 号： 12306071

完 成 日 期： 2024.05.03

大连理工大学

Dalian University of Technology

目 录

1	绪论.....	1
1.1	YOLO 算法简介.....	1
1.2	YOLO 算法原理.....	1
1.3	个人所做工作.....	2
2	基于 YOLO 算法的手势识别.....	3
2.1	方案选择.....	3
2.2	训练流程.....	3
3	结论.....	4

1 绪论

1.1 YOLO 算法简介

在目标检测算法的发展过程中,人们一开始采用 Proposal+图像分类的思路,Proposal 用于预测目标位置,分类用于识别目标的类别。这类算法被称为 Two-stage 算法,例如 R-CNN, Faster R-CNN 算法。2015 年,Joseph Redmon, Santosh Divvala 等人在《You Only Look Once: Unified, Real-Time Object Detection》论文中提出 YOLO 模型,开启了 YOLO 算法的研究热潮。经过后来人们的不断改进,YOLO 算法已经发展成为一个庞大的家族,后来人们把第一个 YOLO 模型称为 YOLOv1 模型。

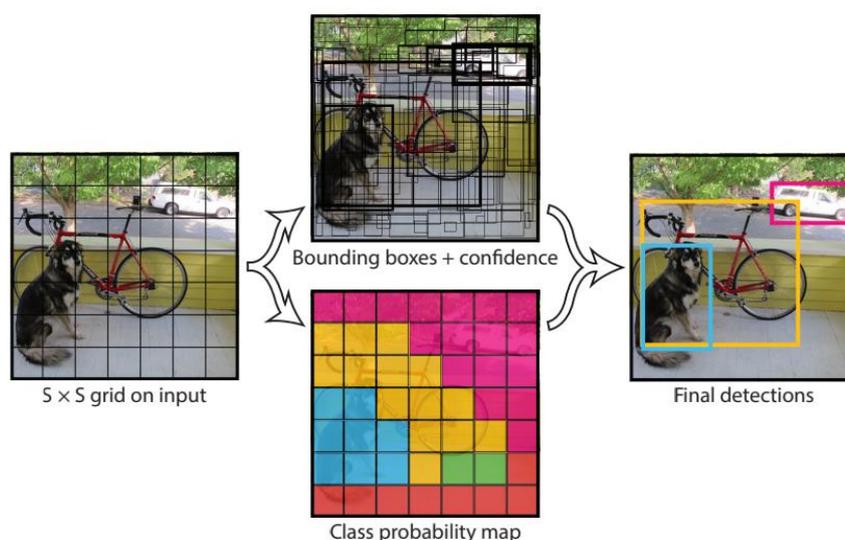
YOLO 模型使用单个卷积网络同时预测多个边界框,以及对应框的类别概率。与传统的目标检测方法相比,这种统一的模型有以下优点:

(1) YOLO 的预测速度快,由于模型将检测框转化为回归问题,因此只用一个网络同时输出目标的位置与分类信息;

(2) YOLO 模型在进行预测时对图像进行全局推理。与基于滑动窗口和区域建议的技术不同,YOLO 在训练和测试期间看到整个图像,因此它隐式地学习关于类别及其形状的上下文信息。

1.2 YOLO 算法原理

YOLO 算法把输入图像分成 $S \times S$ 个方格,每个方格输出一个 $B \times 5 + C$ 维的张量。这里, B 是每个方格预测方框(Bounding box)的数目, C 表示需要检测的对象类别数目。如果检测目标的中点落在某个方格内,那么目标物体就由该方格进行检测与输出。



YOLO 算法采用了相对坐标的方法，并使用 Logistic 激活函数，使每个坐标的取值落在区间 0 到 1，如图 2 所示。目标检测网络在每个边界框处预测 5 个值，分别是 t_x ， t_y ， t_w ， t_h 和 t_o 。如果单元格相对图像左上角偏移记为 c_x ， c_y ，并且先验边界框宽度、高度分别记为 p_w ， p_h ，对应的预测结果为：

$$b_x = \sigma(t_x) + c_x$$

$$b_y = \sigma(t_y) + c_y$$

$$b_w = p_w e^{t_w}$$

$$b_h = p_h e^{t_h}$$

$$Pr(object) * IOU(b, object) = \sigma(t_o)$$

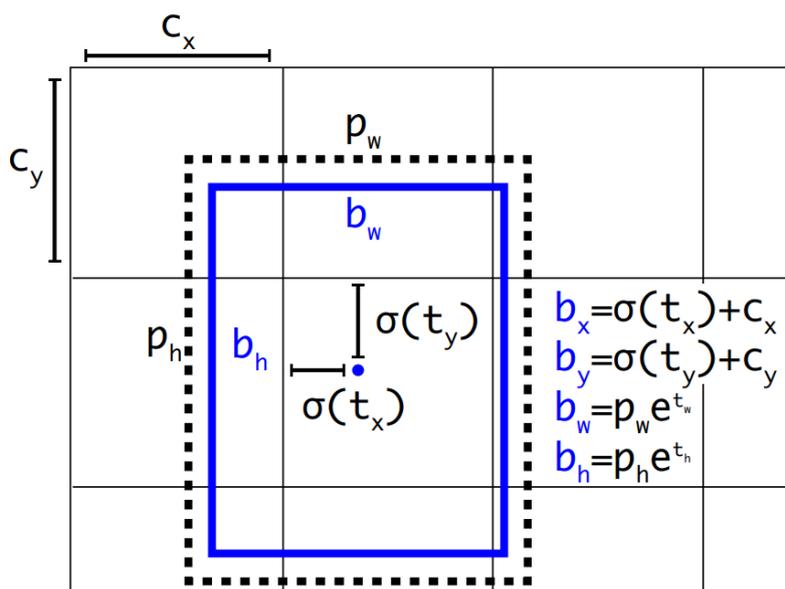


图 2 目标检测网络

1.3 个人所做工作

我参与的工作主要有进行 YOLO 算法基础理论的学习，协助组长杜志远进行手势识别代码的调试，以及手势识别所需数据集的制作和训练。

2 基于 YOLO 算法的手势识别

2.1 方案选择

本项目的手势识别部分是使用 YOLOv8 算法来实现的，YOLOv8 是最新版本的 YOLO 系列模型，用于进行实时对象检测。由于其速度快和效率高并且参数量低，被广泛用于计算机视觉任务中，常用于小图像的目标检测中。具有如下的优点：

(1) 更高的识别准确率：YOLOv8 在保持高速处理能力的同时，通过优化模型架构和训练过程，进一步提高了识别准确性。

(2) 适用性强：YOLOv8 旨在提高在不同环境和条件下的泛化能力，使其能够在多种场景下都有很好的表现。

(3) 易于部署：YOLOv8 进一步优化了模型的实现，使其更加易于在各种设备上部署，包括边缘设备和云端服务器。

本次数据集选取的是 HaGRID 数据集，HaGRID 数据集种类非常丰富，包含 One, Two, ok 等 18 种常见的通用手势，HaGRID 数据集数量特别大，有 716GB 的大小，包含 552,992 个 FullHD(1920×1080)RGB 图像。标注了手势框和手势类别标签，可以用于图像分类或图像检测等任务。

本次选取了其中两种手势（分别为 like、stop）共 11000 张图片作为训练数据集。将训练集、测试集、验证集按照 7: 2: 1 的比例进行划分后进行训练。

2.2 训练流程

训练流程包括 5 个部分：制作数据集进行标签标注、图像预处理、手势图像的训练、模型评价、识别模型生成（见图 3）。

(1) 进行标签标注工作，使用 labelImg 对收集到的手势数据集进行标注。

(2) 对手势识别图像进行一定的预处理，以排除因环境因素造成的干扰。

(3) 下载预训练权重模型，设定训练停止阈值，分别是训练满 50 次以及训练精度达到 90%，执行训练操作，根据设定的停止条件终止训练得到最优模型，即 best.pt 文件。

(4) 对每一轮训练后的模型进行精确度计算，即模型的评价操作。

(5) 判断阈值是否达到，达到则输出训练好的识别模型，未到达则执行误差反向传播算法，使学习参数朝着精确度更高的方向更新变化，再进行训练。

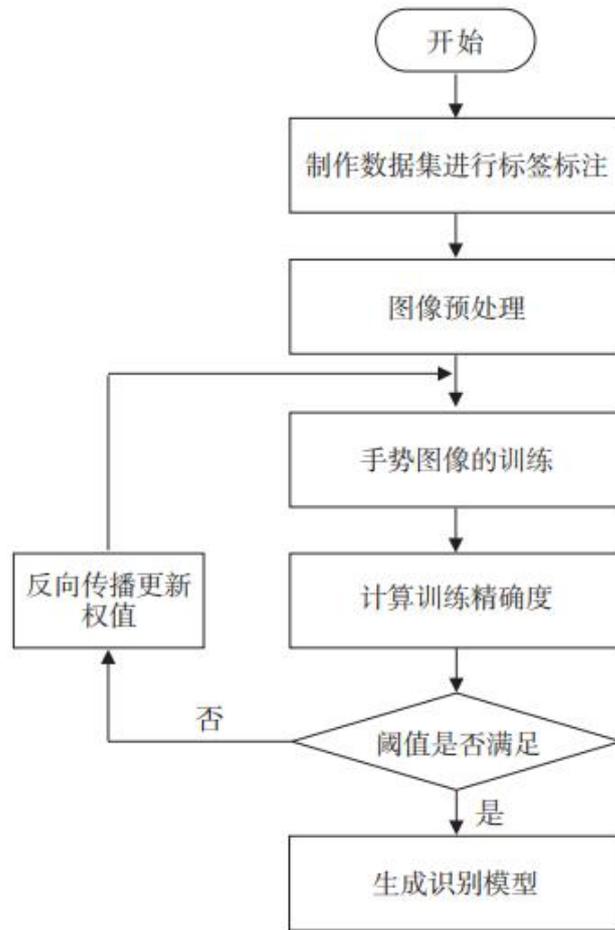


图3 训练程序流程图

3 结论

将 YOLO 算法应用到手势识别中，取得了非常不错的效果，在模糊图像、背景近肤色和光线较暗等多种情况下，识别的准确率都较高。YOLOv8 能够很好地实现对于手势的实时识别，并且对于非静态手势图片也有着较高的精确度，能够很好地实现手势的快速识别，在对手势图片及摄像头手势实时识别方面有着良好的应用价值。在今后的研究中也会进一步加大数据样本量，优化识别系统对复杂环境的识别处理，增强模型的泛化能力和鲁棒性，提高模型的识别准确率。