

计算机视觉大作业

低空目标识别：针对低空经济的无人 机视觉传感设备

学 院： 信息工程学院

姓 名： 万宇虹

学 号： 1221001035

指导老师： 刘铁

目录

一、研究背景与现状	3
二、数据集介绍	3
三、算法模块	4
1、YOLOv5 模型	4
(1) 数据集处理	4
(2) 模型选择	5
(3) 训练结果	5
2、TPH-YOLOv5 模型	6
3、YOLOv8 模型	7
四、总结与展望	8

一、研究背景与现状

计算机视觉是人工智能的重要分支，其核心任务是让计算机能够处理、解析和理解图像或视频内容，从而感知周围环境。随着深度学习技术的突飞猛进，近年来计算机视觉在目标检测、图像分类、语义分割等诸多方向都取得了显著突破。

在计算机视觉的众多任务中，目标检测尤为关键，其目标在于确定图像中物体的类别及其空间位置。在各类目标检测算法中，YOLO（You Only Look Once）系列模型凭借其卓越的检测速度和精度优势脱颖而出，已成功应用于自动驾驶、安防监控、医疗影像等多个重要领域。

近年来，无人机因其体积小、成本低、机动性强、部署快等特点，在各个领域迅猛发展。在军事领域，使用无人机占领低空空域，进行目标捕获、情报侦查、战场监视等，在民用领域无人机可应用于交通管理、农林植保、应急救援、电力巡检等。随着人工智能、无人机技术的进步，在低空飞行中无人机能够在复杂环境中进行实时数据采集和处理，并且相比于地面具有更广泛的视野，从而获得更为丰富的地面信息，如何高效利用低空视野的优势，快速准确地检测出其中的小目标，具有理论意义，也有实际应用价值，这对目标检测算法提出四点要求：高精度、高帧率、低运算量、低参数量。基于卷积神经网络（CNN）等先进算法在目标检测方面表现出色，能够帮助无人机高效地识别和定位各种目标，从而实现自主导航和任务执行。因此，计算机视觉技术与无人机技术的深度融合，为无人机视角下的实时目标检测带来广阔的应用潜力和发展前景。

YOLO 系列模型在目标检测任务中的核心优势在于其高效的速度-精度平衡、简洁的单阶段检测架构和强大的工程化支持，采用端到端的单阶段设计将目标定位和分类合并为单一网络，大幅提升推理速度（如 YOLOv8 可达 100+FPS），显著优于两阶段方法，同时通过全局回归损失直接预测边界框和类别，避免了复杂的区域提议步骤，使模型更轻量且易于部署，该系列从 YOLOv1 到 YOLOv8 持续引入锚点机制、SPPF、PANet 等关键技术，在保持实时性的同时精度显著提升，配合官方库提供的完整训练-部署工具链和 TensorRT/ONNX 支持，使其成为工业界目标检测的首选方案。

二、数据集介绍

VisDrone2019 数据集是一个大规模公开的无人机视角视觉数据集，由天津大学机器学习与数据挖掘实验室于 2019 年发布，专门用于无人机平台上的计算机视觉任务研究。

该数据集包含 288 个视频片段（约 26 万帧）和 10,209 张静态图片，采集自中国不同城市、不同天气和光照条件下的真实场景，覆盖城市广场、街道、高速公路、乡村等多种复杂环境。

数据集主要包含四大任务：目标检测（使用矩形框标注 8 个类别共约 260 万个实例）、单目标跟踪（132 个序列）、多目标跟踪（96 个序列）和人群计数（1,500 张图像），其中目标检测任务的标注类别包括行人、车辆（轿车/卡车/巴士等）、自行车等典型交通参与者，所有图像均以 1080P 分辨率采集，并提供了详细的场景属性标注（如光照条件、遮挡程度等）。VisDrone2019 的挑战性体现在小目标检测（80%实例小于 32×32 像素）、密集场景（单图最多 1,200 个实例）和典型无人机视角的几何变形等特点，已成为评估算法在无人机视角下性

能的基准数据集,其验证集和测试集分别包含 1.6 万和 5.2 万标注实例,官方采用 AP(Average Precision) 和 AR (Average Recall) 作为主要评估指标, 并针对小目标特别设置了 AP^small 评测维度。



图 1 数据集原图

子集	图像数量	标注框数量	说明
训练集	6,471	~540,000	覆盖多种场景和光照条件
验证集	548	~46,000	用于调参和中间评估
测试集	1,610	~134,000	需提交结果至官方服务器

图 2 数据集分布

三、算法模块

1、YOLOv5 模型

使用 YOLOv5 进行低空目标识别具有一系列显著特点, 这些特点既继承了 YOLO 系列的传统优势, 又针对低空视角的特殊性进行了优化适配。低空拍摄的目标通常呈现小尺寸、高密度、多角度和复杂背景等特点, YOLOv5 通过其网络架构和训练策略的针对性设计, 能够较好地应对这些挑战。其核心特点体现在多尺度检测能力上, 通过 FPN+PANet 结构强化了特征金字塔的多层融合, 使得模型能够同时检测低空图像中差异极大的目标, 从近地面的较大物体到远处仅占数个像素的小目标都能覆盖。

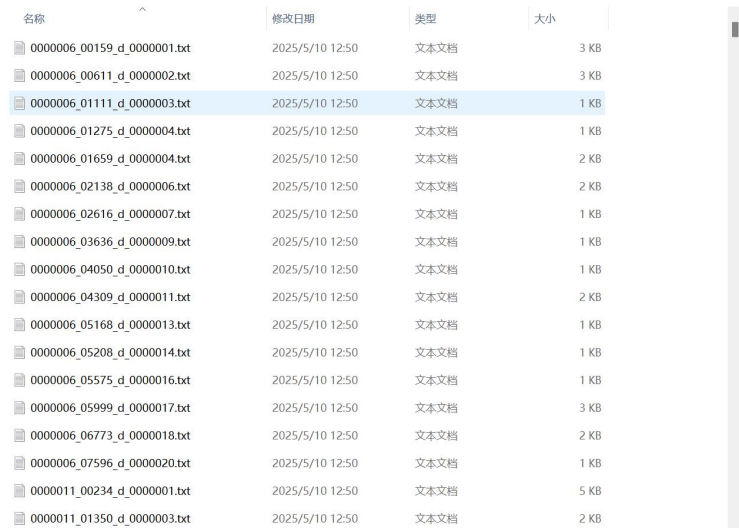
总体而言, YOLOv5 在低空目标识别任务上平衡了速度、精度和易用性, 成为该领域的实用选择之一。

(1) 数据集处理

使用 YOLOv5 训练 VisDrone2019 数据集的过程需要针对无人机视角和小目标检测的特点进行专门适配。首先从 VisDrone 官网下载数据集, 解压后按照 YOLOv5 要求的目录结构重新

组织, 通常需要创建 images/train、images/val、labels/train、labels/val 等子目录。由于 VisDrone 原始标注格式为 XML, 需要使用脚本将其转换为 YOLOv5 支持的 txt 格式, 转换时要注意将原始类别 ID 映射为连续整数 (如行人 0、汽车 1 等), 同时需要处理标注中的忽略区域标记。

新版本的 yolo5 中已经集成了训练 visdrone 数据集的配置文件, 其中附带了数据集的处理方式, 主要是 labels 的生成, 正确执行代码后, 会在 'VisDrone2019-DET-train', 'VisDrone2019-DET-val', 'VisDrone2019-DET-test-dev 三个文件夹内新生成 labels 文件夹, 用以存放将 VisDrone 数据集处理成 YoloV5 格式后的数据标签。



名称	修改日期	类型	大小
0000006_00159_d_0000001.txt	2025/5/10 12:50	文本文档	3 KB
0000006_00611_d_0000002.txt	2025/5/10 12:50	文本文档	3 KB
0000006_01111_d_0000003.txt	2025/5/10 12:50	文本文档	1 KB
0000006_01275_d_0000004.txt	2025/5/10 12:50	文本文档	1 KB
0000006_01659_d_0000004.txt	2025/5/10 12:50	文本文档	2 KB
0000006_02138_d_0000006.txt	2025/5/10 12:50	文本文档	2 KB
0000006_02616_d_0000007.txt	2025/5/10 12:50	文本文档	1 KB
0000006_03636_d_0000009.txt	2025/5/10 12:50	文本文档	1 KB
0000006_04050_d_0000010.txt	2025/5/10 12:50	文本文档	1 KB
0000006_04309_d_0000011.txt	2025/5/10 12:50	文本文档	2 KB
0000006_05168_d_0000013.txt	2025/5/10 12:50	文本文档	1 KB
0000006_05208_d_0000014.txt	2025/5/10 12:50	文本文档	1 KB
0000006_05575_d_0000016.txt	2025/5/10 12:50	文本文档	1 KB
0000006_05999_d_0000017.txt	2025/5/10 12:50	文本文档	3 KB
0000006_06773_d_0000018.txt	2025/5/10 12:50	文本文档	2 KB
0000006_07596_d_0000020.txt	2025/5/10 12:50	文本文档	1 KB
0000011_00234_d_0000001.txt	2025/5/10 12:50	文本文档	5 KB
0000011_01350_d_0000003.txt	2025/5/10 12:50	文本文档	2 KB

图 3 处理数据集后生成的 label 文件夹

数据准备完成后, 需要修改 YOLOv5 的配置文件以适应 VisDrone 的特点。在 data/visdrone.yaml 中设置正确的训练集、验证集路径和类别信息, 类别数改为 8 并指定类别名称。考虑到 VisDrone 中 80% 的目标小于 32x32 像素, 需要在模型配置中增强小目标检测能力, 通常修改 models/yolov5s.yaml (或其他版本) 中的 anchor boxes 尺寸, 使用 k-means 重新聚类生成更适合小目标的 anchor 尺寸。还可以在模型头部添加检测层, 将原来的 3 个检测头改为 4 个检测头, 增加 160x160 尺度的特征图以提升小目标检测效果。

(2) 模型选择

YOLOv5 提供四种模型尺寸, 此处我选择使用 YOLOv5m (medium) 平衡型来训练数据集。选择 YOLOv5m 模型训练 VisDrone2019 数据集是在模型性能与计算效率之间权衡后的典型选择, 相比更小的 YOLOv5s 或更大的 YOLOv5l/x 具有更适配的任务匹配性。YOLOv5m 作为中等规模模型, 其主干网络和特征金字塔的深度与宽度设计在参数量 (约 21MB) 和计算量上达到平衡, 既能捕捉无人机视角下小目标的细微特征 (如 32x32 像素的行人), 又不会像 YOLOv5l/x 那样因过度参数化导致训练数据不足时的过拟合风险。

VisDrone 数据集的特殊性直接影响了模型选择决策。该数据集 80% 的目标为小尺寸实例, 需要模型具备较强的多尺度特征融合能力, YOLOv5m 的 3 个检测头 (20x20、40x40、80x80) 配合中等规模的特征提取能力, 比 YOLOv5s 更适应密集小目标场景, 其 1024 维的特征通道数 (YOLOv5s 仅 512 维) 能更好保留高空视角下的低分辨率目标信息。同时, 相比 YOLOv5l/x, YOLOv5m 在 1536x1536 的高分辨率输入下 (VisDrone 常用设置) 的显存占用更可控 (约 11GB/GPU), 允许更大的 batch size (如 16-32), 这对稳定训练至关重要。

(3) 训练结果

使用 YOLOv5m 训练 80 轮后

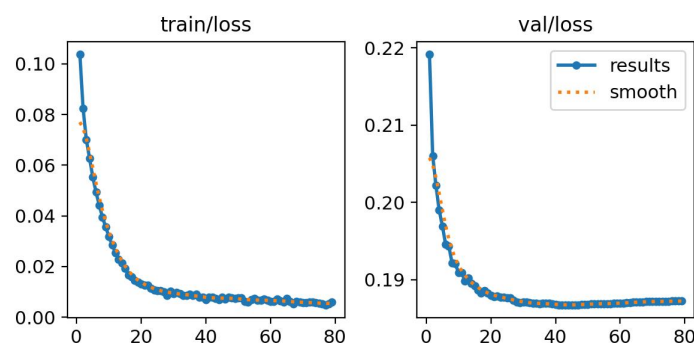


图 4 YOLOv5 训练 80 轮后结果

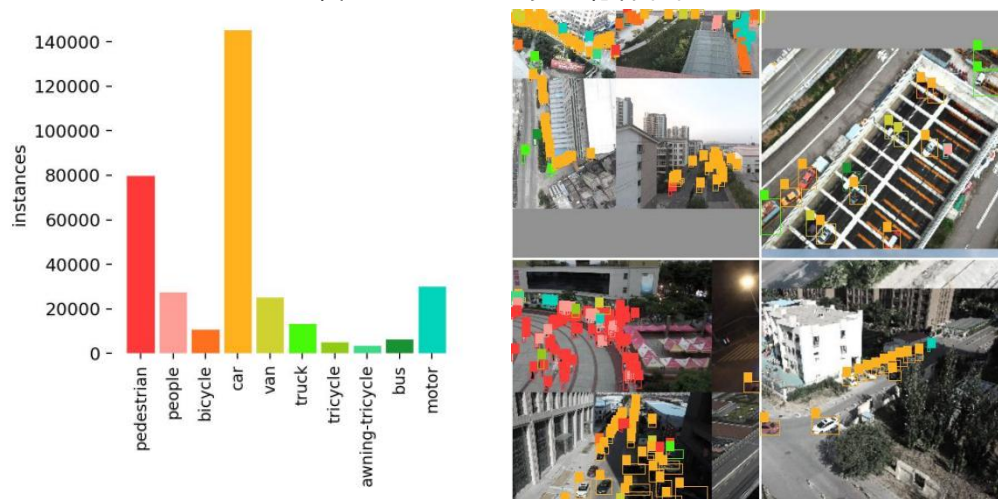


图 5 训练 100 轮计数结果

2、TPH-YOLOv5 模型

与传统 YOLOv5 相比，TPH-YOLOv5 有以下优点。

增加了一个预测头用于微小物体检测，与其他三种预测头相结合，四头结构可以缓解剧烈的对象尺度变化带来的负面影响。

用 Transformer 编码器块替换了原始版本 YOLOv5 中的一些卷积块和 CSP 瓶颈块。与原始瓶颈块相比，Transformer 编码器块可以捕获全局信息和丰富的上下文信息。每个 Transformer 编码器包含两个子层。第一子层是多头注意层，第二子层（MLP）是全连接层。在每个子层之间使用剩余连接。Transformer 编码器块提高了捕获不同本地信息的能力。它还可以利用自我注意机制探索特征表征潜力。

引入注意力机制。卷积块注意模块 CBAM 是一个简单但有效的注意模块。它是一个轻量级的模块，可以集成到大多数著名的 CNN 架构中，并且可以以端到端的方式进行训练。CBAM 在给定特征图的情况下，沿着通道和空间两个独立的维度顺序地推断注意力图，然后将注意力图与输入特征图相乘以执行自适应特征细化。

我们使用 yolov5l-tph-plus.yaml、yolov5l-xs-tph.yaml、yolov5l-xs-tr-cbam-spp-bifpn.yaml 分别训练了 VisDrone2019 数据集，参数如下

		配置1 (yolov5l-tph-plus)	配置2 (yolov5l-xs-tph)	配置3 (yolov5l-xs-tr-cbam-spp-bifpn)
参数	nc	80		
	depth_multiple	1.0		
	width_multiple	1.0		
	anchors	3	4	4
主干网络	模块参数	Conv 模块参数依次为：64、128、256、512、1024；C3 模块重复次数依次为 3、6、9、3		
	最后一个 C3 模块	C3 模块	C3 模块	C3TR 模块
	CBMA使用	未使用	未使用	多个阶段使用CBAM
	最终输出	CLLADetect模块	Detect模块	Detect模块
在YOLOv5 6.0版本上		C3模块和检测模块不同	增加了一个检测头	增加一个检测头、注意力模块等

表 1 三种模型不同参数设定

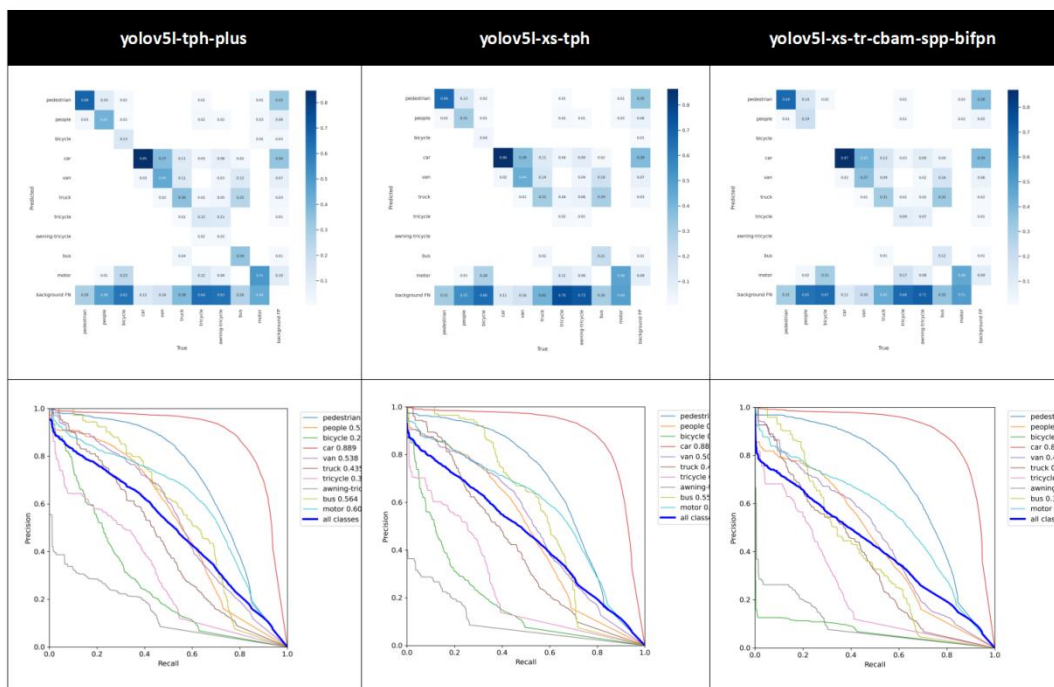


图 6 三种模型运行结果对比

随模型复杂度的增加，训练结果并没有变好。可能是因为复杂的模型通常更多的时间来收敛， epoch 太少不足以让模型充分学习数据中的模式。

3、YOLOv8 模型

YOLOv8 主要由 Backbone（主干网络）+ Neck（特征融合）+ Head（预测头）三部分组成。

YOLOv8 模型用 C2f 替代 CSP 模块：结构更简单，表达能力更强，适用于小目标和细粒度任务。加入了 Anchor-Free 检测头，使检测更灵活、对小目标更友好。

VisDrone 数据集集中的无人机拍摄高度 从 10 米到 200 米不等，涵盖了低空、中低空和中空多个飞行高度段。YOLOv8n 模型的检测速度最快、参数量最少，轻量化的模型结构在资源受限场景（无人机）中的更有优势，选择该模型进行训练。

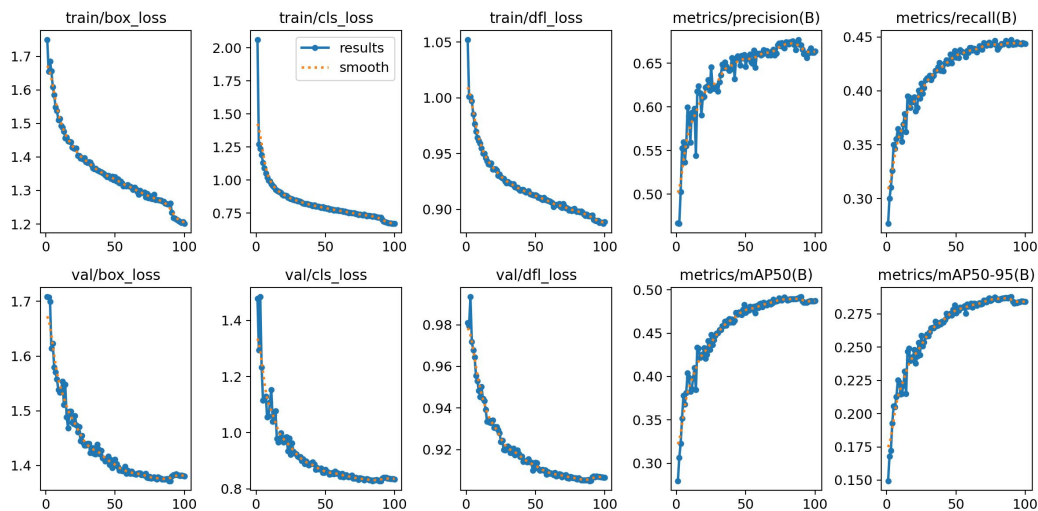


图 7 YOLOv8 训练 100 轮结果

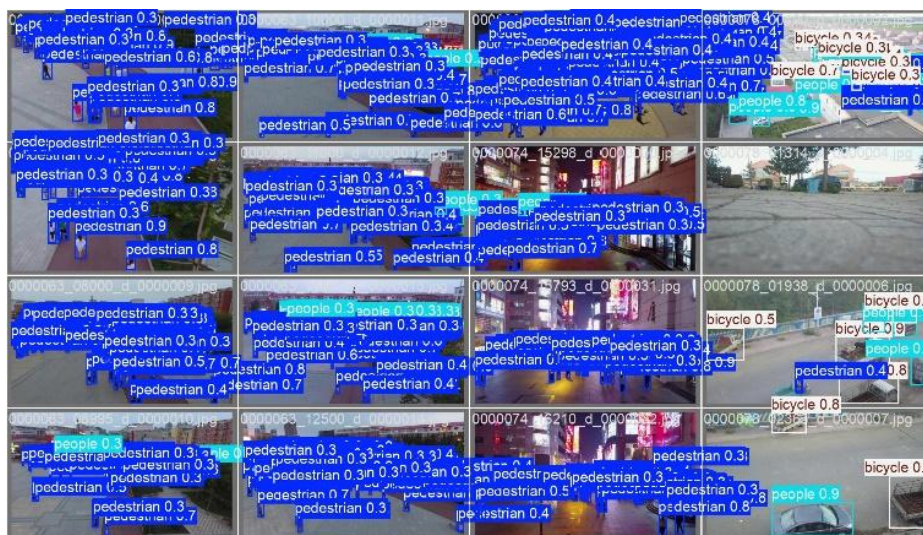


图 8 数据集测试结果

训练集损失仍在持续下降，而验证集损失已趋于稳定，为避免模型过拟合，选择在此时提前停止训练。模型检测精度与参考论文结果基本一致。

四、总结与展望

在这项研究中，我们系统性地研究了基于 YOLO 框架的低空目标检测技术，重点对比了 YOLOv5m、TPH-YOLOv5 和 YOLOv8n 三种模型在 VisDrone2019 数据集上的性能表现。针对无人机视角下小目标密集、背景复杂等特点，我们通过引入 TPH-YOLOv5（集成 Transformer Prediction Heads 的改进模型）和轻量化的 YOLOv8n 模型，显著提升了传统 YOLOv5m 在低空场景中的检测能力。实验结果表明，改进后的 TPH-YOLOv5 凭借其多头注意力机制，在密集小目标检测任务中展现出更高的定位精度，而 YOLOv8n 则通过更高效的网络架构设计，在保持相近检测精度的同时大幅提升了推理速度。

基于当前研究成果，我们认为低空目标检测技术的未来发展可重点关注以下方向：

首先，针对无人机特有的视角变化问题，开发具有视角不变性的特征提取模块将有助于提升模型在俯仰角变化时的稳定性。其次，探索动态分辨率输入策略，使模型能够自适应地处理不同飞行高度下的图像分辨率差异。第三，研究多模态数据融合方法，结合红外或深度

信息来增强复杂光照条件下的检测鲁棒性。最后，优化模型部署方案，通过神经架构搜索（NAS）等技术在边缘计算设备上实现精度与效率的最佳平衡。

本研究的实验成果充分验证了 YOLO 系列模型在低空目标检测领域的适用性，特别是 TPH-YOLOv5 在精度优化和 YOLOv8n 在效率提升方面的突出表现。我们相信，低空目标检测在未来的智慧城市、农业监测、应急救援等领域能发挥更大价值。