

低空目标识别：
针对低空经济的无人机视觉传感设备
在快速和相对远的距离测试和实现视觉识别算法
——基于 TPH-YOLOv5

组号： 第 3 组

姓名： 徐民睿

学号： 1221001050

班级： XG22B1JK

目录

一、 引言与现状	1
二、 我们的工作	2
三、 数据集介绍	3
四、 算法模块	5
1. YOLOv5	5
2. TPH-YOLOv5	6
五、 实验结果	7
1. 不同批次对比实验	7
2. 不同轮次对比实验	8
3. 消融实验	8
六、 总结与展望	10
参考文献与资料	12
附录	13

一、 引言与现状

近年来，随着硬件技术成熟和成本降低，低空无人机行业快速发展，市场规模从 2015 年的 30 亿元增长至 2020 年的 273 亿元，增长约 9 倍。无人机视觉技术在能源、基建、农业、商业和公共安全等领域发挥重要作用，推动行业智能化升级。

在能源领域，无人机视觉可高效检测油管道泄漏、风力发电机叶片损坏及光伏板污损，替代高危人工作业，降低成本并提升效率。在基础建设方面，无人机视觉助力高精度地理测绘和气象探测，自动分析地形、建筑物及气象数据，加速城市信息化进程。农业植保中，无人机凭借成本低、作业范围广等优势，结合视觉技术实现农情监测、障碍规避和精准喷洒，适用于各类农作物。商业应用上，无人机视觉服务于影视拍摄、体育赛事和个人娱乐，提供灵活、低成本的航拍方案，并实现智能稳像和目标跟踪。在公共安全领域，无人机视觉助力抢险救灾（如火源定位、灾情建模）、交通管理（车流监测）及反恐安防（可疑目标追踪），突破传统监测限制，提升应急响应能力。

低空无人机飞行高度通常在 50~1000 米，其图像具有：目标尺度小，分辨率受限；低空背景复杂，干扰多；成像易受影响；数据质量差异大等特征。

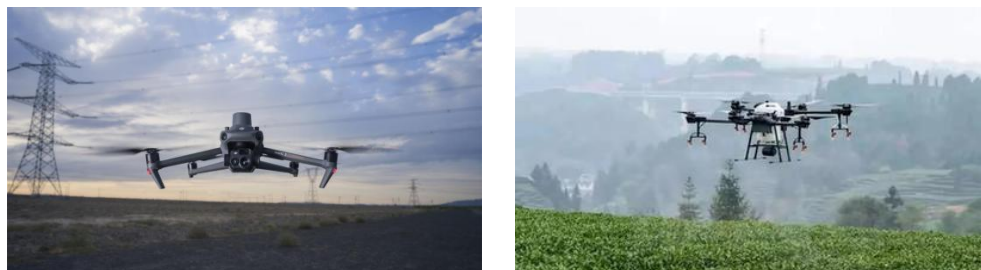


图 1 无人机图片

目前主流的目标检测算法主要是基于深度学习模型，其可以分成两大类：

(1) **two-stage 检测算法**，其将检测问题划分为两个阶段，首先产生候选区域（region proposals），然后对候选区域分类（一般还需要对位置精修），这类算法的典型代表是基于 region proposal 的 R-CNN 系算法，如 R-CNN，Fast R-CNN，Faster R-CNN 等；(2) **one-stage 检测算法**，其不需要 region proposal 阶段，直接产生物体的类别概率和位置坐标值，比较典型的算法如 YOLO 和 SSD。

目标检测模型的主要性能指标是检测准确度和速度，对于准确度，目标检测要考虑物体的定位准确性，而不单单是分类准确度。一般情况下，two-stage

算法在准确度上有优势，而 one-stage 算法在速度上有优势。不过，随着研究的发展，两类算法都在两个方面做改进。Google 在 2017 年开源了 TensorFlow Object Detection API，并对主流的 Faster R-CNN，R-FCN 及 SSD 三个算法在 MS COCO 数据集上的性能做了细致对比。Facebook 的 FAIR 也开源了基于 Caffe2 的目标检测平台 Detectron，其实现了最新的 Mask R-CNN，RetinaNet 等检测算法，并且给出了这些算法的 Baseline Results。可以说，准确度（accuracy）和速度（speed）是一对矛盾体，如何更好地平衡它们一直是目标检测算法研究的一个重要方向。

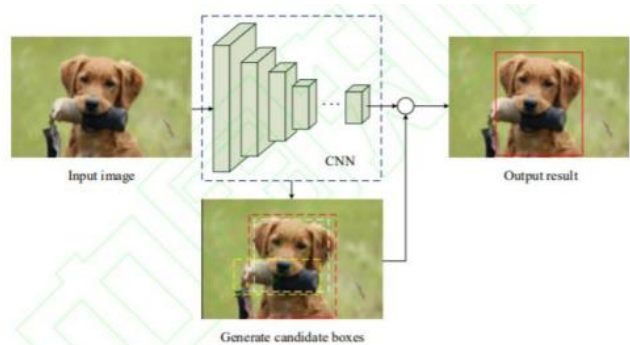


图 2 基于深度学习的两阶段目标检测算法
Fig. 2 Two-Stage object detection algorithm based on deep learning

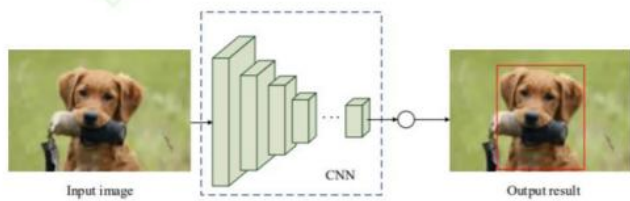


图 3 基于深度学习的单阶段目标检测算法
Fig. 3 One-Stage object detection algorithm based on deep learning

图 2 One-stage 与 Two-stage 算法示意图

特别的，YOLO 系列算法自推出以来，经过不断更新迭代，已发展至 YOLOv11 版本。YOLOv11 在继承历代版本核心优势的基础上，创新性地引入了改良的锚点生成机制和多尺度预测技术，对网络结构进行了深度优化，显著提升了目标检测的性能和效率，受到众多研究学者的广泛关注。

二、 我们的工作

在“低空目标识别：针对低空经济的无人机视觉传感设备，在快速和相对远的距离测试和实现视觉识别算法”这一研究课题中，本组成员首先对相关领域的研究现状及可用数据集展开了系统性调研。经综合评估，选定 VisDrone2019 数据集作为实验基础，并确定以 YOLO 系列算法为研究核心。

为实现不同算法间的横向对比与协同学习，团队成员分工负责 YOLO 系列中的特定算法，涵盖 YOLOv5、tph-YOLOv5、YOLOv7、YOLOv8。各成员深入探究所负责算法的理论原理，自主搭建实验环境并开展性能测试。实验阶段，成员们详细记录关键指标及现象，并且进行消融实验。在实验完成后，团队组织集中交流与讨论，共享实验成果与见解。

三、数据集介绍

VisDrone2019 数据集由天津大学机器学习与数据挖掘实验室的 AISKYEYE 团队收集，旨在为无人机场景下的视觉分析研究提供大规模数据支持。该数据集涵盖了城市和乡村环境的多种场景，包括不同的天气和光照条件，是目前无人机场景视觉分析领域最具代表性的数据集之一。官方提供的数据中训练集是 6471、验证集是 548、测试集 1610 张。数据集共提供了以下 12 个类，分别是：忽略区域、pedestrian、people、bicycle、car、van、truck、tricycle、awning-tricycle、bus、motor、others。VisDrone 数据集的 txt 文件的标签内容和 yolo 的不同，所以需要脚本文件来进行转换。

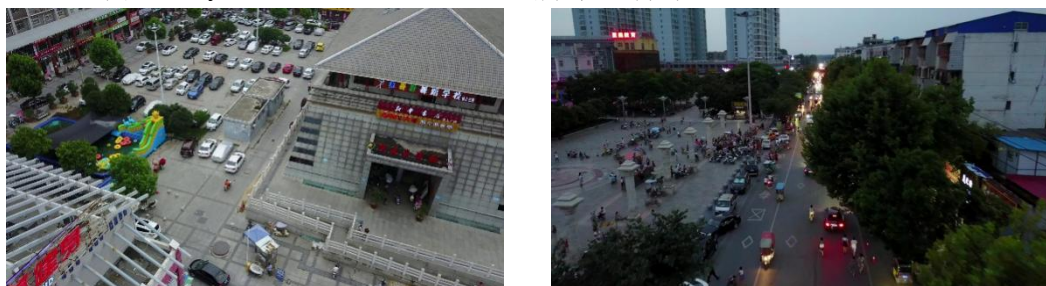


图 3 VisDrone2019 数据集图像

对 TPH-YOLOv5 在 VisDrone2019 数据集上的训练结果进行可视化，如图 4。

图中柱状图展示了不同类别目标的实例数量分布，横轴为目标类别，涵盖行人、人群、自行车、汽车、面包车、卡车、三轮车、带篷三轮车、公交车、摩托车等，纵轴为实例数量。从图中可以看出，数据集存在明显的类别不均衡现象，汽车类别的实例最多，超过 140,000 个，而带篷三轮车等类别的实例数量相对较少。这种不均衡的数据分布可能导致模型在训练过程中对少数类别目标的学习不足，影响其对这些类别目标的检测性能。

右上角小图呈现目标边界框示意图，绿色矩形框代表模型检测到的目标边界框，可直观观察其大小与形状，助力理解模型检测状况。左下角密度图揭示

目标在图像空间中的分布，横纵轴分别对应图像宽度与高度的归一化坐标，颜色深浅反映目标密度，深色区域指示目标聚集位置，与无人机拍摄时目标出现位置相关。**右下角散点图**描绘目标宽度和高度的分布关系，横轴为宽度，纵轴为高度，每个点对应一个目标的宽高组合。多数目标宽高集中在较小范围，契合无人机低空拍摄时目标尺度小的特征。

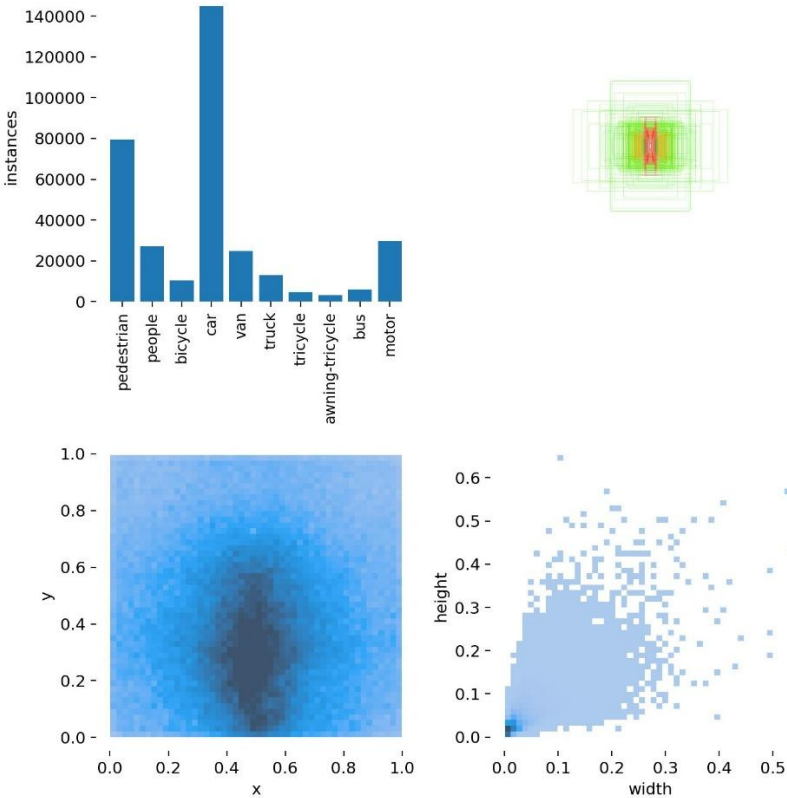


图 4 数据集训练结果可视化

在官方 Github 仓库中，我们主要关注 yolov5l-tph-plus、yolov5l-xs-tph、yolov5l-xs-tr-cbam-spp-bifpn 三个结构配置文件。其主要参数对比如图 5。

		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

图 5 官方文件主要参数对比

以上三个文件没有完全做到控制变量。所以在消融实验中，我们基于完整的 TPH-YOLOv5，即文件 yolov5l-xs-tr-cbam-spp-bifpn，对网络配置进行改写。

四、 算法模块

1. YOLOv5

YOLOv5 是一种基于深度学习的目标检测算法，属于 YOLO（You Only Look Once）系列的第五代版本。其背景主要源于目标检测在计算机视觉领域的重要作用和广泛应用需求。

在 YOLOv5 之前，YOLO 系列已经经历了多次迭代。YOLOv1 作为该系列的开山之作，提出了**端到端的目标检测框架**，首次实现了在一个神经网络中同时进行目标分类和定位，大大提高了检测速度。随后的 YOLOv2 和 YOLOv3 在精度和速度上不断改进，YOLOv3 特别强调了对小目标的检测能力。YOLOv4 则在模型结构和训练方法上进行了更深度的优化，引入了更多的新型网络结构和数据增强技术。

YOLOv5 是在 YOLOv4 的基础上进一步简化和优化而来。它主要针对实际应用中的需求，对模型结构进行了**轻量化设计**，使其更适合在资源受限的设备上运行，同时在速度和精度上取得了更好的平衡。这使得 YOLOv5 在众多目标检测任务中得到了广泛应用，特别是在实时性要求较高的场景，如安防监控、自动驾驶、无人机目标跟踪等领域。此外，YOLOv5 的开发者还提供了易于使用的开源代码和预训练模型，进一步降低了该算法的使用门槛，推动了目标检测技术在更广泛领域的发展。

YOLOv5 的模型架构由三个主要部分构成：**Backbone、Neck 和 Head**。Backbone 采用 CSPDarknet53，该网络基于 Cross Stage Partial（CSP）架构，通过将特征图在不同阶段进行分割和融合，有效增强梯度流并减少计算时间，从而提高特征提取的效率和质量。Neck 部分融合了 FPN（Feature Pyramid Network）和 PANet（Path Aggregation Network）结构。FPN 用于构建多尺度特征金字塔，使模型能够同时检测不同大小的目标；PANet 则进一步优化了特征的传递路径，增强特征的语义信息，提高小目标的检测能力。Head 部分负责最终的检测输出，它基于预定义的锚框（Anchor Box）生成候选边界框，并预测每个边界框的类别概率和位置坐标。为了提高检测的准确性，YOLOv5 采用了

一种基于 K-means 聚类 and 遗传算法的方法来自动学习锚框尺寸，这些锚框尺寸从数据集中边界框的分布中学习而来，能够更好地匹配实际目标的形状和大小，对初始化预测过程和模型性能尤为关键。

2. TPH-YOLOv5

TPH-YOLOv5 是针对实时目标检测而设计的一个深度学习框架。它的核心在于改进了原始 YOLOv5 模型的架构，特别是在处理小目标检测方面有显著提升，同时也优化了模型的训练流程，以提高整体性能。这个项目的目的是为研究人员和开发人员提供一个高效、易用且适应各种应用场景的目标检测解决方案。

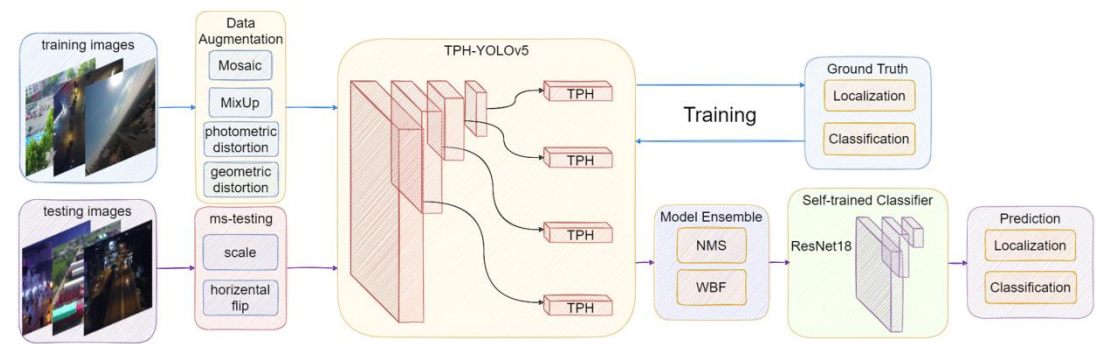


图 6 TPH-YOLOv5 算法流程图

对比 YOLOv5，TPH-YOLOv5 主要有以下改进：

- **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。
- **注意力机制：** 卷积块注意模块（CBAM）。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。

- 增加了模型集成和自训练分类器、更复杂的训练方式。

五、 实验结果

1. 不同批次对比实验

配置：预训练权重文件：yolov5l、Yaml 文件：yolov5l-xs-tph.yaml、
Epoch：10。

表 1 不同 batch size 实验结果对比

	precision	recall	mAP_0.5	mAP_0.5:0.95
Batch1	0.52554	0.43222	0.41252	0.2587
Batch2	0.46538	0.46844	0.44582	0.28236
Batch3	0.5152	0.4474	0.45958	0.28932

在机器学习和深度学习领域，批次大小（Batch Size）是一个关键的超参数，它决定了在进行一次梯度更新前模型处理的训练样本数量。批次大小对模型训练的稳定性和最终性能有着显著影响。较小的批次大小通常会引入较高的估计噪声，这可能导致训练过程中的权重更新不够稳定，损失函数下降轨迹呈现波动性。然而，这种噪声有时能帮助模型跳出局部最优解，从而提升泛化能力。此外，小批次训练的优势在于内存占用低，适合资源受限的设备。相对地，较大的批次大小能够降低估计噪声，使训练过程更加平稳，损失函数下降更为平滑。大批次训练通常需要采用预热策略以避免初始学习率过高带来的问题，并且可能使模型收敛到较为平坦的最小值，进而提升测试时的泛化性能，但其效果也受到优化器和学习率调整策略的显著影响。

本次实验聚焦于低空目标识别任务，采用 YOLO 系列算法处理无人机视觉传感数据，旨在探索不同批次训练对模型性能的影响。实验结果显示，不同批次在精确率（Precision）、召回率（Recall）、mAP_0.5 和 mAP_0.5:0.95 等关键指标上存在差异。具体来看，Batch1 的精确率表现最佳，表明该批次模型在预测无人机目标时较为保守，误报情况较少；而 Batch2 的召回率最高，意味着它在检测实际无人机目标方面更为敏感，但相应地，精确率有所下降。随着批次的

推进，从 Batch1 到 Batch3，mAP_0.5 和 mAP_0.5:0.95 呈现上升趋势，这表明模型在训练过程中逐渐学习到更有效的特征表示，整体检测性能得到提升。

综合分析这些实验数据，可以推断不同批次大小对模型性能的影响在本次实验中得到了一定程度的体现。尽管各批次之间的指标差异并非极端显著，但整体趋势反映出模型训练的渐进优化过程。实验结果也揭示了在实际应用中，选择合适的批次大小对平衡模型训练效率与性能的重要性。

2. 不同轮次对比实验

配置：预训练权重文件：yolov5l、Yaml 文件：yolov5l-xs-tph.yaml、Batch_size：2。

表 2 不同 epoch 大小实验结果对比				
	precision	recall	mAP_0.5	mAP_0.5:0.95
Epoch10	0.46538	0.46844	0.44582	0.28236
Epoch20	0.55638	0.52038	0.52888	0.33735
Epoch80	0.65793	0.54355	0.58968	0.38009

在机器学习和深度学习中，Epoch（轮次）是一个关键概念，它代表了模型在整个训练数据集上进行一次完整迭代的过程。增加 Epoch 数量可以让模型有更多机会学习数据中的模式，从而提高其性能。但是，Epoch 数量过多可能导致模型过拟合，即模型在训练数据上表现很好，但在测试数据上表现不佳。相反，Epoch 数量过少可能导致模型欠拟合，无法充分学习数据中的复杂模式。

本次实验展示了不同 Epoch 数量对模型性能的影响。实验结果显示，随着 Epoch 的增加，模型的精确率（Precision）、召回率（Recall）、mAP_0.5 和 mAP_0.5:0.95 等指标均呈现上升趋势。这表明随着训练轮次的增加，模型对低空目标识别的能力逐步增强，这可能是因为模型在更多的训练迭代中逐渐学习到了更丰富的特征表示。然而，需要注意的是，虽然性能指标在增加，但边际提升可能会逐渐减小。此外，过长的训练时间可能带来计算资源的浪费，并且需要关注是否存在过拟合的风险。

3. 消融实验

配置：预训练权重文件：yolov5l、epoch：80、batch_size：3。Yaml 文件

改自 yolov5l-xs-tr-cbam-spp-bifpn.yaml。分别去掉 cbma 注意力模块、去掉小检测头模块。实验结果如表 3:

表 3 消融实验结果对比

	precision	recall	mAP_0.5	mAP_0.5:0.95
Yolov5+detection	0.622	0.558	0.582	0.369
Yolov5+cbma	0.621	0.567	0.585	0.375
Tph-yolov5	0.649	0.58	0.61	0.393

从消融实验的结果来看，Yolov5+detection 和 Yolov5+cbma 的召回率（recall）和精确率（precision）相当，两者在 mAP_0.5 和 mAP_0.5:0.95 指标上也较为接近，表明这两个模块对模型性能的影响相似。Tph-yolov5 模型在所有指标上均优于前两者，意味着同时保留 cbma 注意力模块和小检测头模块时模型性能最佳，两者协同作用能够更好地提升模型的检测精度和召回能力。

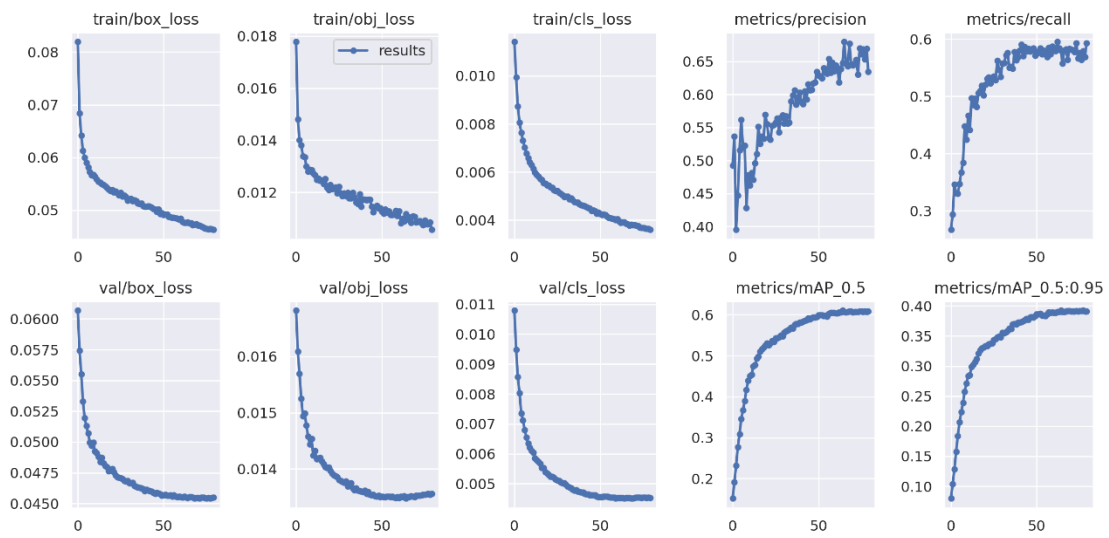


图 7 TPH-YOLOv5 训练 80 轮结果图

从 TPH-YOLOv5 模型训练 80 轮的结果图来看，模型展现了良好的收敛趋势和性能提升。训练和验证过程中的损失指标（包括边界框损失、目标损失和分类损失）均呈下降趋势，表明模型在不断优化，误差逐渐减小。与此同时，各项性能指标（精确率、召回率、mAP_0.5 及 mAP_0.5:0.95）稳步上升，说明模型对目标的检测能力不断增强，能够更精准地识别和定位目标。具体而言，训练损失曲线的平稳下降表明模型在训练集上逐渐拟合得更好，验证损失曲线的同步下降则说明模型在未见数据上也有较好的泛化能力，没有出现明显的过拟合现象。性能指标方面，精确率的提升表明模型的预测结果更加可靠，召回率的提高意味着模型能够识别出更多的真实目标。此外，mAP 指标的上升进一步验证了模型在不同检测难度下的综合性能提升。总体而言，Tph-

YOLOv5 模型在低空目标识别任务中表现出色，随着训练的进行，其检测性能和泛化能力都得到了有效提升。

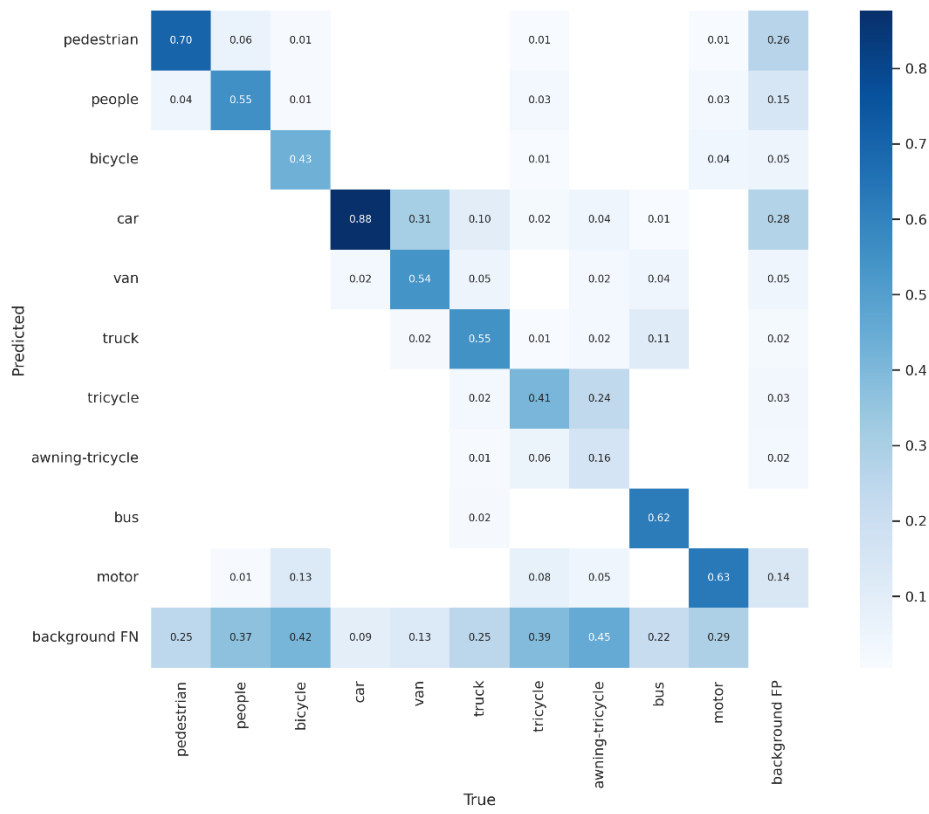


图 8 TPH-YOLOv5 训练 80 轮次混淆矩阵

从混淆矩阵的总体表现来看，模型在样本数量较多的类别（如 car、pedestrian、people 和 motor）上表现较好，对角线上的值较高，表明这些类别的正确分类率较高。对于样本数量较少的类别（如 bus、awning-tricycle 和 truck），模型的正确分类率相对较低，且存在一定的混淆现象。

六、 总结与展望

本文围绕低空经济下的无人机视觉传感设备，针对低空目标识别这一课题进行了深入研究与实践探索，主要取得以下成果：

- **开发环境学习：**在项目实践中，团队深入学习并熟练掌握了 PyCharm 开发环境的使用技巧，极大地提升了开发效率与代码质量。PyCharm 丰富的功能，如智能代码补全、自动化代码重构、实时语法检查与错误定位以及无缝集成的调试工具，显著加速了代码编写、调试和优化的流程，为复杂算法的实现和模型训练提供了有力支持，成为项目成功的重要助力。
- **理论层面：**系统梳理了低空无人机视觉技术的应用现状，对比分析了

主流目标检测算法（two-stage 与 one-stage）的优缺点，阐述了 YOLO 系列算法的发展脉络及其核心优势，为重点算法的选择与研究奠定了坚实理论基础。

- **实践层面：**基于 VisDrone2019 数据集，对 YOLOv5、TPH-YOLOv5 等算法进行了全面实验与性能评估。通过调整批次大小（Batch Size）、增加训练轮次（Epoch）等超参数优化策略，深入探究了不同条件对模型检测精度（Precision、Recall、mAP 等指标）的影响规律。同时，借助消融实验验证了 CBMA 注意力模块与小检测头模块对提升 TPH-YOLOv5 模型性能的关键作用，实验结果充分展示了该模型在低空目标识别任务中的优越性能，证明了其在精确率、召回率及 mAP 等关键指标上的显著提升，为低空目标检测领域提供了高效可靠的算法解决方案。

总的来说，通过本项目的学习与实践，团队掌握了系统性的新课题研究流程，包括从明确研究目标、深入调研相关技术背景与应用现状，到基于实际需求搭建实验环境、选择合适的算法与数据集，并通过大量的实验验证与参数调优逐步优化模型性能，最终实现创新性成果的输出。这一过程为后续开展复杂科研项目积累了宝贵经验，使团队对科研探索的全链条有了清晰认知。

尽管 TPH-YOLOv5 在本研究中表现出色，但仍有提升空间。一方面，可探索引入更先进的注意力机制或轻量化网络结构，进一步平衡模型精度与速度；另一方面，针对小目标检测难题，可研究多尺度特征融合与增强技术，以提高对微小低空目标（如微型无人机）的检测能力。

本次研究中，受限于算力资源，模型训练耗时较长，且 Batch Size 最大仅能设置为 3，这在一定程度上影响了训练效率。而在实际应用中，出于对计算资源高效利用和模型推理速度的综合考量，Batch Size 通常会设置为 2 的次方（如 1、2、4、8 等），这是由于计算机硬件架构对 2 的幂次运算具有天然的适配性，能够实现更高效的内存分配和计算资源管理。

参考文献与资料

[1]孙一铭,赵柯嘉,王硕,等.面向无人机的低空视觉数据集研究综述[J].数据采集与处理,2025,40(02):274-302.DOI:10.16337/j.1004-9037.2025.02.002.

[2]Zhao, Q., Liu, B., Lyu, S., Wang, C., & Zhang, H. (2023). TPH-YOLOv5++: Boosting Object Detection on Drone-Captured Scenarios with Cross-Layer Asymmetric Transformer. Remote Sensing, 15(6), 1687.

<https://doi.org/10.3390/rs15061687>

[3] X. Zhu, S. Lyu, X. Wang and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios," 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 2021, pp. 2778-2788, doi: 10.1109/ICCVW54120.2021.00312.

[4]https://blog.csdn.net/weixin_44298961/article/details/122764145?ops_request_misc=%257B%2522request%255Fid%2522%253A%2522801769e0653a2f3be2c196985a5a48fb%2522%252C%2522scm%2522%253A%25220140713.130102334..%2522%257D&request_id=801769e0653a2f3be2c196985a5a48fb&biz_id=0&utm_medium=distribute.pc_search_result.none-task-blog-2~all~sobaiduend~default-1-122764145-null-null.142^v102^pc_search_result_base2&utm_term=tph-yolov5%E4%BB%A3%E7%A0%81%E5%A4%8D%E7%8E%B0&spm=1018.2226.3001.4187

附录



图 9 数据集标签可视化



图 10 TPH-YOLOv5 预测结果可视化

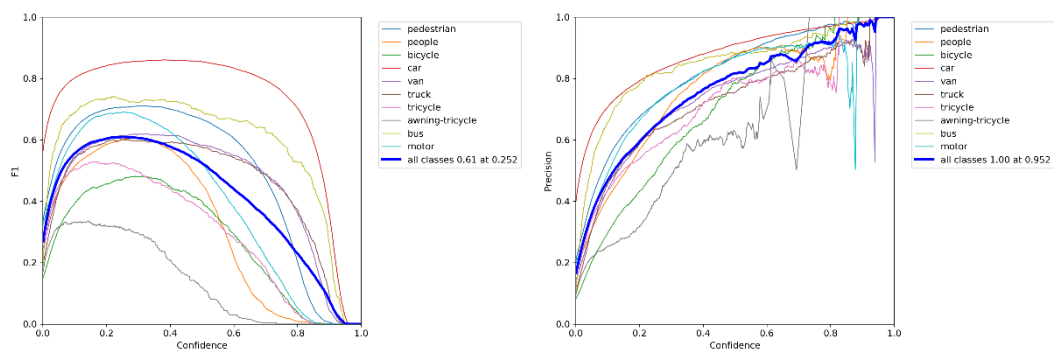


图 11 TPH-YOLOv5 训练结果图：左图为 F1_curve，右图为 P_curve

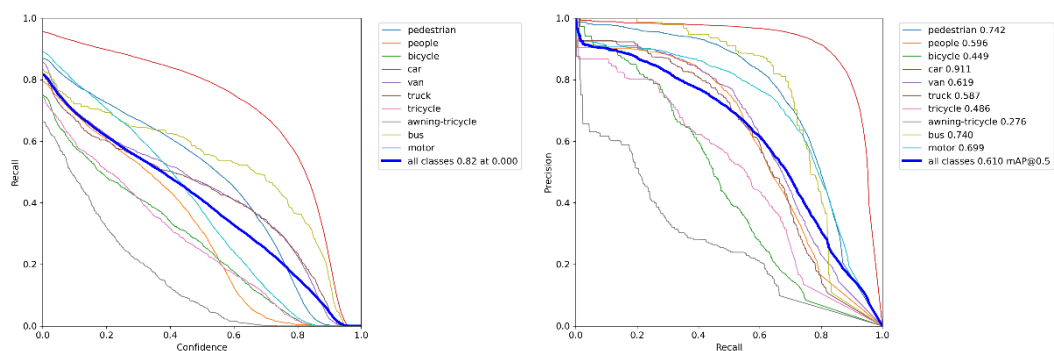


图 12 TPH-YOLOv5 训练结果图：左图为 R_curve，右图为 PR_curve