|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | | |
| **低空目标识别** | | | | |
|  | | | | |
|  | | | | |
|  | 姓 名： | ： | 刘学梅 |  |
| 学 号： | ： | 1221001024 |
| 指导教师姓名： | ： | 刘铁 |
| 班 级： | ： | 计算机科学与技术 |
|  | ： |  |
| 二零二五年六月 | | | | |

**一、 研究背景**

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。

低空无人机视觉的广泛应用，不仅优化了各行业作业模式，更推动了社会管理的智能化和高效化发展。

低空无人机飞行高度通常在 50~1000 米，有以下特点：

1.目标尺度小，分辨率受限：在 500 米距离下，典型无人机相机中仅占 15~30 像素。在 1000 米外，目标可能仅 5~10 像素，接近传统 CNN 的检测极限。由于姿态变化，俯视/侧视视角下目标呈现不同形状，如十字形、星形、矩形等。

2.低空背景复杂，干扰多：动态背景干扰，如飞鸟、建筑物反射等。外界光照条件多变，对比度降低、阴影干扰等。

3.无人机高速特性：成像抖动、运动模糊等。

4.传感器限制：由传感器类型不同造成数据质量差别较大。



**二、 研究现状**

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Caffe2 的目标检测平台 Detectron，其实现了最新的 Mask R-CNNRetinaNet 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。

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[1]。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。

**三、数据集介绍**

为了在真实复杂环境下评估视觉算法在无人机平台上的性能，本文选用了 VisDrone 数据集作为主要实验基准。VisDrone（Visual Object Detection in Drone Imagery）是一个由中国合肥工业大学 AISKYEYE 团队构建的大规模无人机视觉数据集，旨在推动低空视角下目标检测与跟踪算法的研究与应用。

该数据集涵盖了多个城市（如天津、杭州等）的多样化户外场景，包括城市道路、住宅区、公园、广场等，具有极高的现实代表性。VisDrone 数据集不仅提供了静态图像和视频序列，还包含了详细的标注信息，为多种计算机视觉任务提供了统一的评估平台。

VisDrone 数据集主要包含以下几类任务与标注内容：

目标检测（Object Detection）：数据集中包含 10,209 张带有目标标注的静态图像，每张图像均附带边界框（bounding boxes）、类别标签和可见度信息。

单目标跟踪（Single Object Tracking）：数据集提供部分序列中的初始标注框，目标是持续跟踪该目标在视频中的位置。

多目标跟踪（Multi-Object Tracking）：视频序列中每个目标具有唯一的 ID 标签，支持基于检测的多目标跟踪任务研究。

实例分割与人群密度估计（Instance Segmentation & Density Estimation）：扩展任务包括目标分割与人群计数，用于提升算法在复杂人群场景中的适应性。

VisDrone 数据集中目标类别共计 10 类，包括：行人（pedestrian）、骑行者（people）、轿车（car）、面包车（van）、卡车（truck）、公交车（bus）、三轮车（tricycle）、摩托车（motor）、自行车（bicycle）和有遮挡三轮车（awning-tricycle）。每个标注框还包含 occlusion（遮挡程度）与 truncation（截断）信息，有助于算法更精确地建模遮挡问题。

VisDrone 数据集的显著特点包括：

多尺度与多角度视角：由于无人机拍摄的高度和角度不断变化，目标在图像中呈现出大尺度变化和复杂的几何形状。

遮挡严重与密集分布：城市环境下目标密集、人群拥挤，目标之间互相遮挡频繁，极大增加了检测和跟踪的难度。

背景复杂与光照多变：不同时间段与天气条件下采集的视频帧存在明显的光照变化、背景噪声较大，对模型泛化能力提出挑战。

动态视角导致的图像抖动：由于无人机在飞行中拍摄，画面存在频繁移动、抖动等现象，要求算法具有较强的鲁棒性。

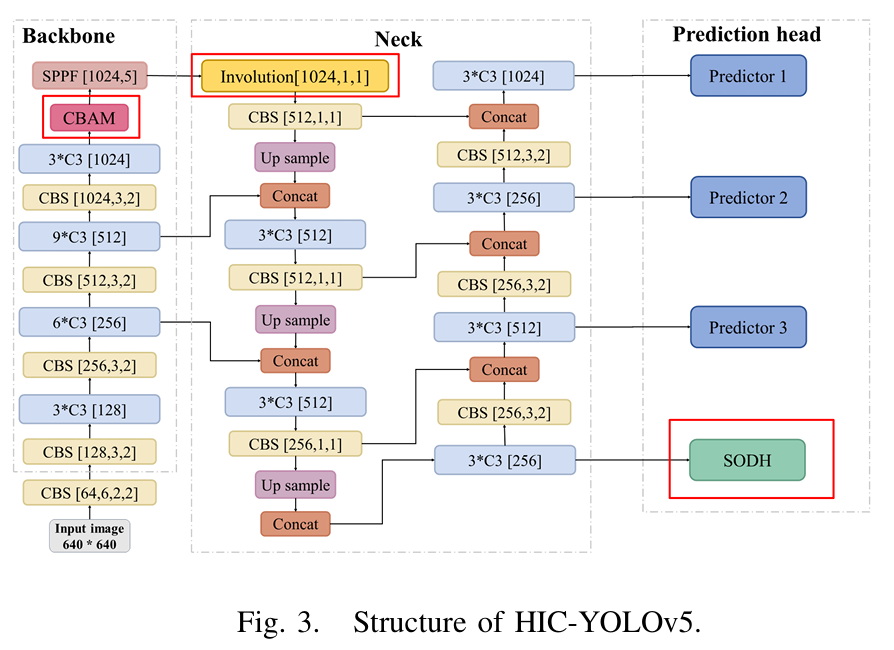


目前 VisDrone 已被广泛应用于目标检测、跟踪、分割等多个领域的重要竞赛和研究中（如 ICCV、CVPR Workshop: VisDrone Challenge），并成为评估无人机视觉系统性能的权威数据集之一。

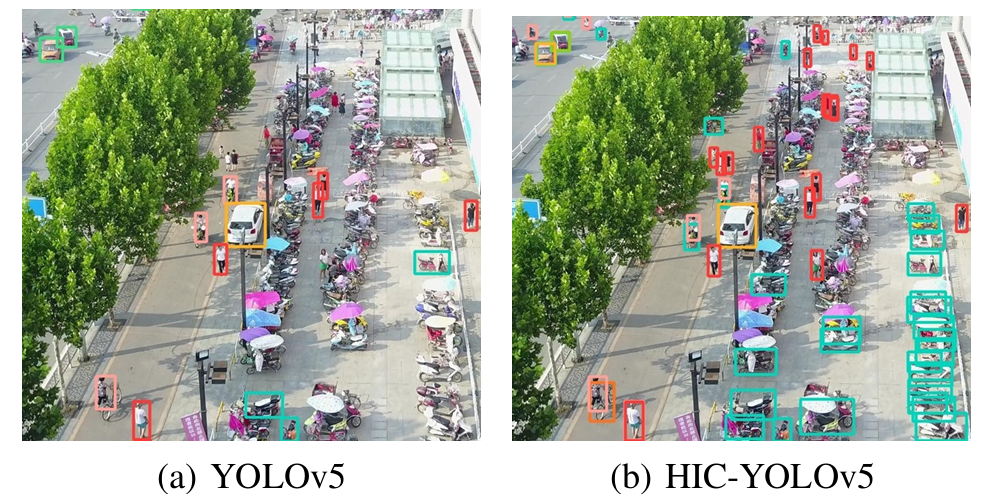
**四、算法部分**

4.1 HIC-YOLOv5算法

小目标检测在无人机图像中具有挑战性，例如物体远、小、模糊、遮挡多等，YOLOv5 尽管轻量，但在 VisDrone 这类小目标任务上效果欠佳，所以在不显著增加计算量的前提下，HIC-YOLOv5提高了YOLOv5对小目标的检测能力。



HIC-YOLOv5 是在 YOLOv5 基础上的改进模型，针对小目标检测性能不足的问题，引入了三大模块以提升检测精度：首先，增加小目标检测头（SODH），以增强对低层次特征中小目标信息的提取能力；其次，引入 Involution 通道特征增强模块，有效提升网络对局部细粒度特征的建模能力；最后，融合 CBAM 注意力机制，增强模型对关键目标区域的关注程度，从而在保持较低计算开销的同时，显著提升对小目标的检测性能。



同样使用采用 VisDrone-2019-DET 数据集对HIC-YOLOv5模型进行评估，得到结果表明改进效果显著的结论：mAP@0.5 从原始 YOLOv5 的 27.57 提升至 36.95，提升了 9.38%；mAP@[.5:.95] 从 14.43 提升至 20.85，提升了 6.42%。在引入中心裁剪策略和轻量级模块的辅助下，HIC-YOLOv5 在保持检测速度优势的同时，有效提升了小目标检测的精度，实现了精度与效率的良好平衡。

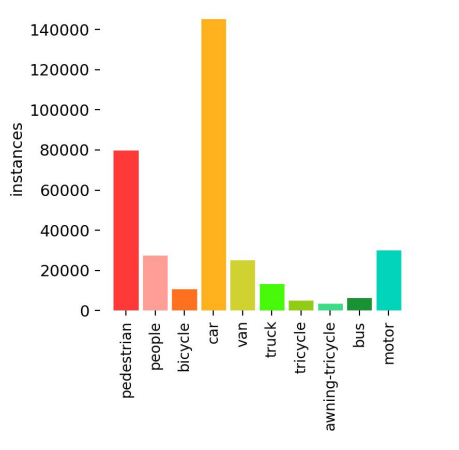
4.2 YOLOv5算法

新版本的yolov5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。

图形用户界面, 图表, 折线图

AI 生成的内容可能不正确。

许多乐高玩具

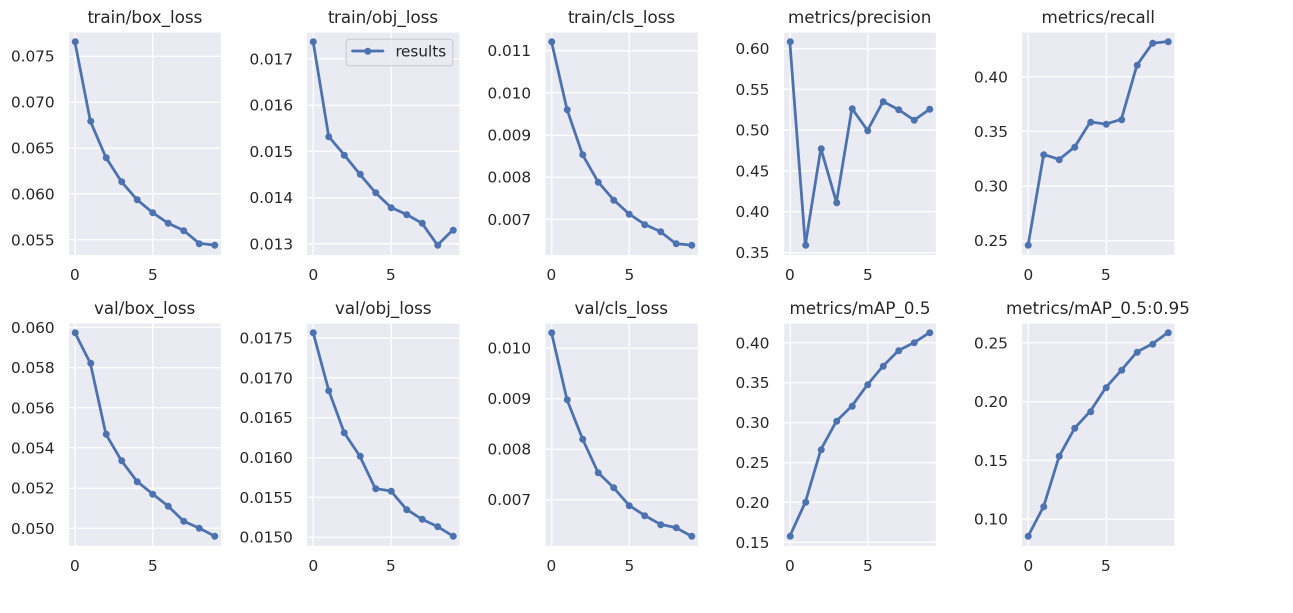
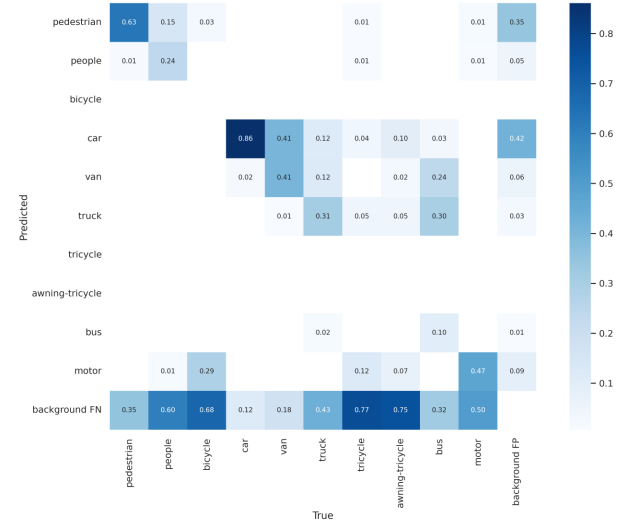
AI 生成的内容可能不正确。

4.3 TPH-YOLOv5

TPH-YOLOv5 是基于 YOLOv5 框架改进的一种专为复杂场景下小目标检测设计的目标检测模型。该模型在结构上进行了多方面的增强，旨在提升对微小物体及多尺度目标的检测能力。首先，TPH-YOLOv5 在原有三种预测头的基础上，新增了一个专门用于微小物体检测的预测头，与其他三个检测头协同工作，形成四头结构，有效缓解了由目标尺度剧烈变化带来的检测困难。其次，TPH-YOLOv5 在部分主干网络中用 Transformer 编码器块替换了原始的卷积块和 CSP 瓶颈结构。与传统瓶颈结构相比，Transformer 编码器具备更强的全局建模能力和上下文信息捕获能力。每个编码器模块由两个子层组成：第一子层为多头自注意力机制层，用于捕获特征之间的长距离依赖关系；第二子层为多层感知机（MLP），用于特征变换与增强；两个子层之间引入残差连接，从而提升信息传递效率和梯度流动性。该设计不仅增强了网络对局部与全局信息的综合建模能力，还提高了模型对复杂场景中细粒度特征的感知能力。

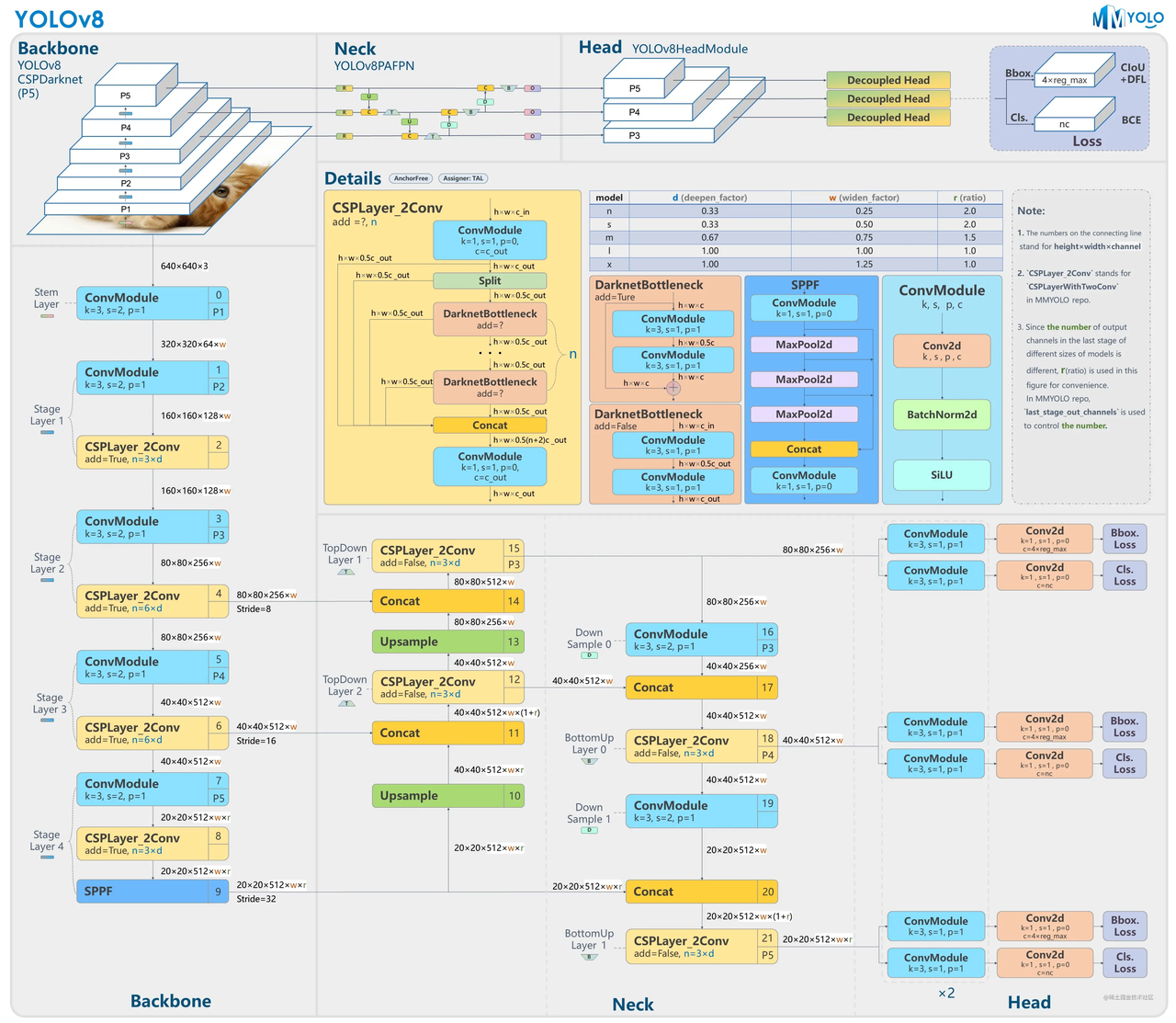
此外，为进一步优化特征表达，TPH-YOLOv5 引入了卷积块注意力模块（CBAM）。CBAM 是一种轻量级的注意力机制模块，能够以端到端方式无缝集成至现有的 CNN 架构中。它分别沿通道和空间两个维度顺序推理注意力权重，并与原始特征图相乘，实现对关键区域和关键信息的自适应特征增强。该机制在保持较低计算开销的同时，显著提高了模型对目标区域的聚焦能力。

在训练策略方面，TPH-YOLOv5 还采用了模型集成和自训练分类器等技术手段，并结合更复杂的训练方式以提升模型的泛化能力和鲁棒性。综合以上改进，TPH-YOLOv5 在小目标检测任务中表现出更高的准确性和稳定性，尤其适用于如 VisDrone 等具有复杂背景与尺度变化显著的实际场景。



模型在训练集和验证集上的表现较为一致，损失都在逐渐下降，而精确率、召回率和平均精度都在逐渐上升，这表明模型在持续学习和优化。

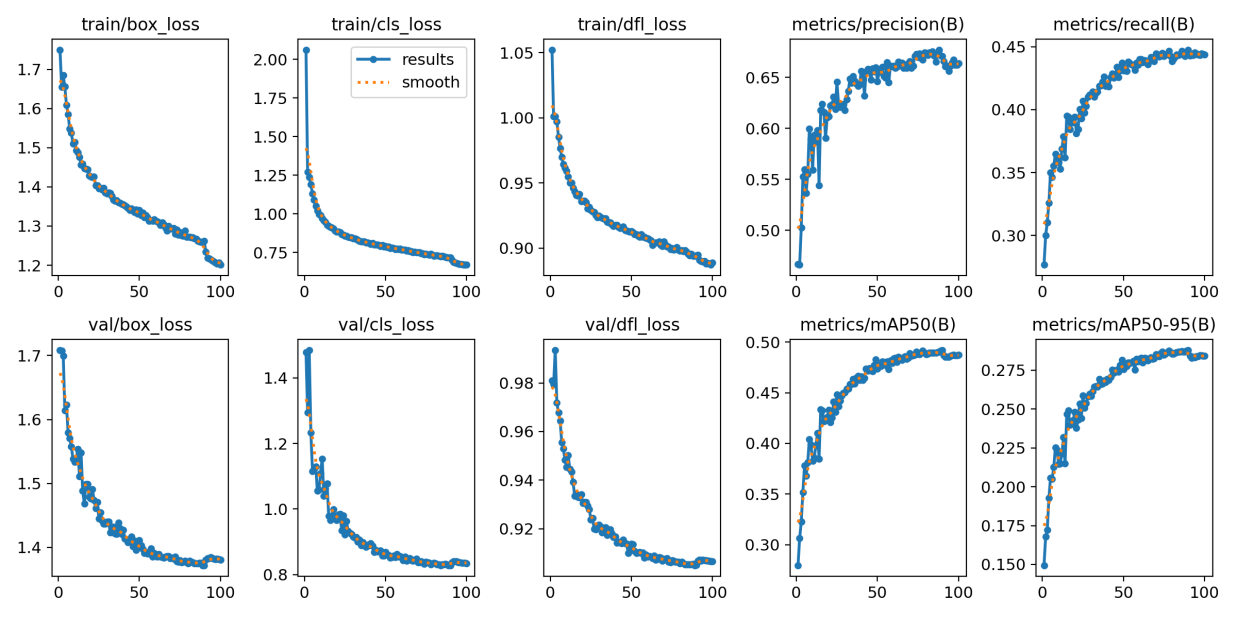
4.3 YOLOv8

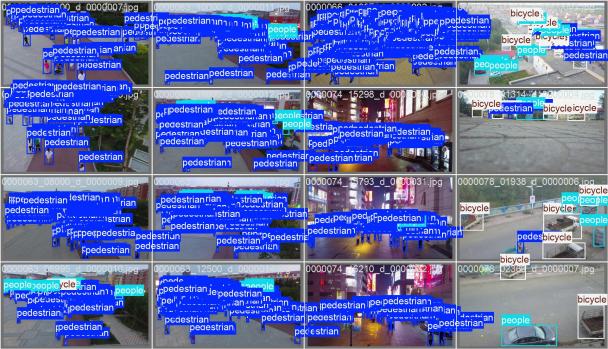
YOLOv8 是 YOLO 系列中最新一代的目标检测框架，继承了前几代模型的高效性与实用性，同时在结构设计与检测方式上进行了多项关键改进。整体架构依然遵循经典的三阶段设计，分别由 Backbone（主干网络）、Neck（特征融合模块） 和 Head（预测头） 组成。其中，主干网络用于提取图像的基础特征，Neck 结构负责多尺度特征的融合，而预测头则用于输出目标的位置和类别信息。

在主干网络部分，YOLOv8 用全新的 C2f（Cross Stage Partial Fusion）模块 替代了以往 YOLOv5 和 YOLOv7 中广泛使用的 CSP（Cross Stage Partial）模块。C2f 模块在结构上更加简洁，同时具备更强的特征表达能力，尤其在处理小目标和细粒度特征方面表现出更好的效果。这一设计简化了网络复杂度的同时，提升了特征建模效率。

在检测头部分，YOLOv8 摒弃了传统的 Anchor-Based（基于锚框）机制，采用了更先进的 Anchor-Free（无锚框）检测头。该检测头直接回归目标中心点及其宽高尺寸，无需预设锚框数量和尺寸，从而在处理密集分布的小目标时更加灵活且鲁棒。同时，Anchor-Free 机制简化了正负样本的匹配过程，进一步提高了模型的训练效率和小目标检测的准确率。

综上，YOLOv8 在继承 YOLO 系列高效率特性的基础上，通过结构优化与检测策略革新，显著提升了模型在小目标、细粒度任务中的表现，展现出良好的通用性与应用前景。





**五、总结与展望**

本文围绕低空无人机场景下的小目标检测问题，选取了 VisDrone-2019-DET 数据集作为研究平台，系统分析并对比了四种基于 YOLO 框架的目标检测算法：YOLOv5、HIC-YOLOv5、TPH-YOLOv5 和 YOLOv8。通过梳理各模型的结构特点及其在小目标检测中的应对策略，结合实际实验结果，验证了它们在复杂背景、多尺度目标、遮挡密集等无人机应用典型场景下的表现差异。

实验结果表明，YOLOv5 作为基准模型，具备良好的检测速度和基础精度；HIC-YOLOv5 在保持轻量化的前提下，通过引入小目标检测头等机制，在小目标检测精度上有明显提升；TPH-YOLOv5 则进一步引入多预测头和 Transformer 编码器，增强模型对复杂场景和细粒度目标的建模能力；YOLOv8 采用了更简洁高效的结构（如 C2f 模块）及 Anchor-Free 检测机制，在通用性和精度方面表现出良好平衡。

总体来看，各模型均为官方改进版本，侧重点不同，适用于不同实际需求场景。本文的工作为低空视觉任务中算法选型提供了参考，同时也为后续的无人机视觉系统部署提供了实验基础。

未来的研究可以在以下几个方向进行拓展：进一步结合实际任务需求，探索检测精度、模型体积与推理速度之间的更优平衡策略；考虑模型在动态环境、自适应光照或低分辨率输入下的鲁棒性表现；以及尝试将小目标检测与跟踪、分割等任务进行多任务联合建模，以提升系统整体智能感知能力。