胡文慧, 李智. 基于改进 YOLOv5 的无人机航拍小目标检测模型[J]. 智能计算机与应用,2026,16(5):106-112. DOI:10.20169/j.issn.2095-2163.25111702

基于改进 YOLOv5 的无人机航拍小目标检测模型

胡文慧, 李 智
(四川大学 电子信息学院, 成都 610065)

摘 要: 在数字化与智能化高速发展的当下,无人机凭借高机动性、灵活性以及成本效益优势,在多个领域中展现出广阔的应用前景。针对无人机视角下小目标特征不明显、分辨率低等特点导致的检测性能差误检、漏检问题,提出一种改进的 YOLOv5 检测模型以提升无人机视角下小目标的检测精度。首先,增加小目标层检测头 P2,使模型能够学习到更全面、更具代表性的特征;其次,采用 DySample 替换模型中的上采样,并同步嵌入 GAM 注意力机制,在保持模型高效率的同时提升模型性能;最后,设计结合 NWD 的 InnerSIoU 改进损失函数,使模型更适用于高精度对称边界框场景。在公开的 RGBT-Tiny 数据集上的实验结果表明,改进后的 YOLOv5s 模型在无人机航拍小目标检测任务中,相较于原始模型,精确率(Precision)、召回率(Recall)、mAP50、mAP50-95 分别提升了 2.9、7.2、7.9、5.0 个百分点,有效改善了识别效果,具有较高的实用价值和应用潜力。

关键词: 无人机航拍; 小目标检测; YOLOv5; 注意力机制; InnerSIoU
中图分类号: TP391 **文献标志码**: A **文章编号**: 2095-2163(2026)05-0106-07

A drone aerial small target detection model based on improved YOLOv5

HU Wenhui, LI Zhi
(College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China)

Abstract: The current rapid development of digitalization and intelligence, drones have shown broad application prospects in many fields with their high mobility, flexibility and cost-effectiveness. To address the issues of low detection performance, false detections, and missed detections caused by the indistinct features and low resolution of small targets from a UAV perspective, this paper proposes an improved YOLOv5 detection model to enhance the detection accuracy of small targets under UAV view. Firstly, a small target detection head P2 is added to enable the model to learn more comprehensive and representative features. Secondly, DySample is used to replace the upsampling in the model, and the GAM attention mechanism is simultaneously embedded, which improves the model performance while maintaining high efficiency. Finally, an improved InnerSIoU loss function combined with NWD is designed, making the model more suitable for scenarios requiring symmetric high-precision bounding boxes. Experimental results on the public RGBT-Tiny dataset show that, in the task of small target detection in UAV aerial photography, the improved YOLOv5s model achieves increases of 2.9, 7.2, 7.9, and 5.0 percentage points in Precision, Recall, mAP50, and mAP50-95 respectively compared with the original model. It effectively improves the recognition effect and has high practical value and application potential.

Key words: UAV aerial photography; small target detection; YOLOv5; attention mechanism; InnerSIoU

0 引 言

在数字化与智能化高速发展的当下,无人机技术凭借高机动性、部署灵活性及成本效益优势异军突起,在交通监控[1]、区域巡视[2]、航空摄影[3]等领域展现出巨大应用潜力。在这些应用任务中,从无人机捕获的图像中检测出目标信息是核心环节。然而,无人机多采用顶部俯视拍摄,这也导致其图像中存在小目标密集、特征模糊、遮挡重叠等问题,导致现有检测模型普遍面临精度不足、误检漏检率高等挑战,因此研发适配无人机场景的小目标检测模型具有重要意义。

近年来,深度学习驱动的目标检测算法快速发展,主要分为双阶段与单阶段两类。双阶段算法先生

作者简介: 胡文慧(2001—),女,硕士研究生。主要研究方向:多模态航拍小目标联合检测算法研究。
通信作者: 李 智(1974—),男,博士,教授。主要研究方向:物联网与边缘计算,压缩感知与频谱感知,群体智能。Email:286086522@qq.com。

收稿日期: 2025-11-17

成候选区域框，再对目标进行精细分类与位置调整。这种算法虽检测精度较高，但需经历两阶段处理，存在计算量大、推理速度慢、帧率低的问题，难以满足无人机设备算力有限、实时性强的实际部署需求。

单阶段算法无需生成候选区域，直接从特征图输出物体类别概率与位置坐标，处理速度快，更契合无人机目标检测的实时性与轻量化要求。目前已涌现出 RetinaNet [4]、SSD(Single Shot Multibox Detector)[5]系列、YOLO(You Only Look Once)系列等代表性模型。其中，YOLO 系列应用广泛且已经历过多版本的迭代，YOLOv3 [6]通过多尺度预测与深度残差网络提升小目标检测能力；YOLOv5 基于 PyTorch 框架简化部署，并通过自动学习锚点优化检测效率。同时，众多学者针对 YOLO 模型展开持续优化。李斌等[7]在特征提取网络中引入特征金字塔实现多尺度预测，结合 GIoU 与焦点损失改进损失函数，减少漏检误检；刘辉等[8]提出融合注意力机制与上下文信息的算法，通过多尺度空洞可分离卷积模块 MDSCM 提取多尺度特征，并以 Soft-NMS 替代 NMS 降低漏检率；祝星馗等[9]将 TransforNer 引入 YOLOv5，利用自注意力整合上下文信息，同时以 BiFPN 替换 PANet 增强特征提取能力，提升无人机航拍数据集检测精度。Liang 等[10]构建创新的轻量级跨层三分支并行融合网络 CTPFNet，通过强化低层次原始特征与深层次语义特征间的跨尺度迁移融合，成功实现了无人机图像中小目标实时检测精度的有效提升。贾星宇等[11]提出了一种基于 YOLOv7-tiny 的多模型融合目标检测网络 YOLO-DA，该网络通过增加小目标和极小目标检测层、引入 ASFF-L 检测头及带扩展变形建模范围调制机制的 DCNS 可变形卷积，成功实现了无人机图像中小目标检测性能的有效提升。李博豪[12]通过加入小目标检测头，并创新性地将空间注意力机制引入 C2f 模块中，精准捕捉小目标的微弱特征，显著提升了不同区域小目标的区分能力。代涛等[13]则通过添加 BiFormer 双层路由机制，构建高效的特征传递通道，专门强化小目标特征的提取与保留，以此增加小目标在整体检测流程中的重要性。尽管上述优化推动了目标检测技术发展，但针对无人机视角小目标检测的模型仍存在精度不足、复杂场景下漏检误检率高、模型参数量大等问题亟待解决。针对上述问题，本文从以下几个方面对 YOLOv5s 模型进行改进：

1)在检测头方面，增加小目标专属检测层 P2，专门针对航拍场景中的极小目标，采用四检测头架构以覆盖更细粒度的特征尺度，通过提升小目标特征在检测阶段的响应权重，解决传统三检测头对小目标的漏检问题。

2)采用 DySample[14]上采样替换模型中的上采样操作，在减少特征信息损失的同时提升特征映射精度；同步嵌入 GAM 注意力机制[15]，通过通道维度与空间维度的双重权重分配，强化小目标与背景的特征区分度，抑制云层、地面纹理等复杂背景的干扰。

3)将 NWD[16]结合 InnerSIoU[17-18]构成一种新的损失函数替换模型原有的 CIoU 损失函数，其中，SIoU 融入了角度参数与规模敏感性，Inner 思想强化边界框在对称场景下的定位精度，配合 NWD 对小目标的距离度量优势，有效缓解小目标因坐标偏移导致的误检问题。

1　YOLOv5 模型

YOLOv5 模型按照模型宽度和深度的不同分为 YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 五个版本，模型的复杂度依次递增。综合考虑检测效果与无人机有限的算力，YOLOv5s 作为 YOLOv5 系列中典型的单阶段轻量化目标检测模型，其架构精巧，能在检测精度与推理速度间实现良好权衡，因此本文选择 YOLOv5s 模型作为基线模型。

2　YOLOv5s 模型改进

2.1　改进的 YOLOv5s 模型

改进后的 YOLOv5s 网络结构如图 1 所示。首先，在原有检测头基础上新增 P2 小目标专属检测层，构建四检测头架构形成更精细的尺度覆盖。其次，针对颈部网络特征融合过程中尺度转换易丢失小目标信息的问题，采用 DySample 上采样替换传统上采样，并嵌入 GAM 注意力机制，在特征融合过程中减少信息损失并强化小目标与背景的区分度；最后，NWD 提升小目标距离度量精度，InnerSIoU 通过角度与对称场景适配优化定位效果，使用二者结合形成的新损失函数 NWD_InnerSIoU 替换原有的 CIoU 以改进损失函数，缓解漏检、误检问题。

2.2　P2 小目标检测头

YOLOv5s 模型在特征提取过程中，随着骨干网络层级推进，通过卷积操作实现特征图通道数增加与空间尺寸缩小。这种变化虽能整合图像信息，但也让深层特征图对应原始图像的范围变大，小目标特征被弱化，从而影响检测效果。

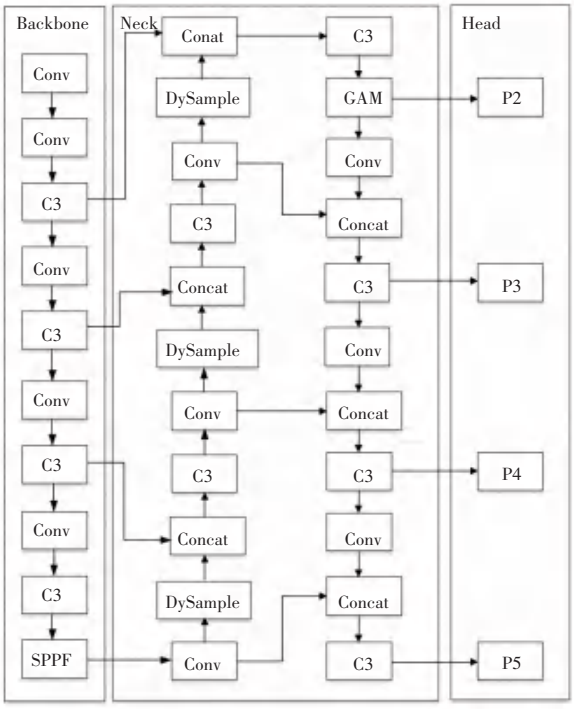


图 1 改进 YOLOv5s 网络结构图

Fig. 1 Improved YOLOv5s network structure diagram

针对此问题本文新增小目标检测头 P2,该检测头的特征源于骨干网络中经较少下采样的特征提取模块,所得特征图空间尺寸较大,可保留小目标更多细节。同时,该检测头融入了骨干网络中浅层的 C3 模块输出特征,通过 Concat 操作实现特征互补,让小目标特征更突出,使模型在处理航拍场景时表现更优。

2.3 改进的颈部网络

2.3.1 DySample 上采样

在 YOLOv5s 模型的特征融合过程中,上采样操作是实现不同尺度特征匹配的关键环节,其性能直接影响特征传递的完整性与模型的检测效率。传统上采样方法基于固定内核进行插值,容易导致小目标特征在尺度转换中丢失细节。

为此,本文采用 DySample 动态上采样器替换原有上采样模块。该采样器基于点重采样机制,分为采样点生成、特征重映射两阶段,其流程如图 2 所示:首先由采样点生成器创建采样集,其核心在于通过静态或动态范围因子方法生成偏移量。静态方法结合线性层、像素洗牌技术与固定范围因子生成偏移量,动态方法则额外引入 Sigmoid 函数生成动态范围因子调节偏移量,最终将偏移量与原始网格位置相加,得到采样点;随后利用 grid_sample 函数对输入特征进行重采样,输出上采样特征。这种设计一方面通过动态生成采样点保留小目标的关键特征,避免了传统插值导致的特征模糊,提升特征融合后的小目标辨识度;另一方面,摒弃耗时的动态卷积与冗余子网络,通过轻量级架构设计维持高效推理速度。从而在航拍小目标检测场景中既能适配无人机的算力限制,又能通过精准的特征传递增强模型对远距离小目标的检测稳定性。

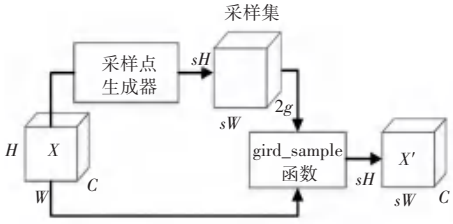


图 2 DySample 动态上采样流程图

Fig. 2 DySample dynamic upsampling flowchart

2.3.2 GAM 注意力机制

GAM 是一种全局注意力机制,采用顺序的通道-空间注意力机制,可减少信息损失并放大全局维度互动特征,其结构如图 3 所示。在通道注意力子模块中,使用 3D 排列跨 3 个维度保留信息,并通过一个两层的多层感知机 MLP 增强跨维度的通道-空间依赖性,能够更全面地提取小目标的特征,提升对小目标信息的捕捉能力。在空间注意力子模块中,采用了两个卷积层进行空间信息融合,同时去除了可能导致信息减少的最大池化操作,更好地保留了空间信息,避免了下采样过程中对小目标信息的丢失。

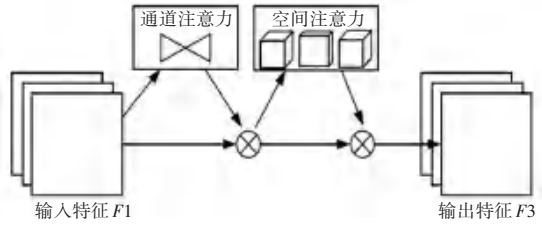


图 3 GAM 注意力机制结构图

Fig. 3 GAM Attention Mechanism structure diagram

2.4 NWD_InnerSIoU 损失函数

在目标检测任务中,损失函数作为量化预测框与真实框差异的核心指标,通过精确度量二者间的位置偏差、尺寸误差和类别预测准确性,为模型优化提供方向。YOLOv5s 中采用 CIoU 损失函数来衡量预测边界框与真实边界框之间的差异,其计算公式如下:

$$L_{\mathrm{CIoU}} = 1 - \mathrm{IoU} + \frac{\rho^2(b, b^{\mathrm{gt}})}{c^2} + \alpha\nu \qquad (1)$$

式中：$\rho^2(b,b^{gt})$ 表示预测框中心 b 与真实框中心 b^{gt} 之间的欧氏距离的平方，c 表示同时包含预测框与真实框的最小外接矩形的对角线长度，α 表示权重系数，v 表示用于衡量预测框与真实框的长宽比一致性的参数。

当预测框与真实目标框完全不重叠的情况下，CIoU 无法有效反映二者之间的紧密关系，这对于小目标检测来说是很不利的。SIoU 是在传统 IoU 的基础上通过引入尺度感知和角度感知机制进行创新改进，通过多维度联合优化能够显著提升小目标检测效果，同时加速模型收敛，其表达式为：

$$L_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \tag{2}$$

式中：Δ 为尺度损失项，用于衡量预测框与真实框的尺寸差异；Ω 为度损失项，角衡量预测框与真实框的尺寸差异。

由于 IoU 损失计算在实际应用中无法根据需要检测的对象进行自我调整，泛化能力弱，引入 Inner-IoU 替代 SIoU 式中的 IoU，通过辅助边框而不是两框的交并比计算损失值，并使用 ratio 系数控制辅助边框的尺度，其表达式如下：

$$IoU^{Inner} = \frac{inter}{union} \tag{3}$$

$$inter = [\min(b_r^{gt} - b_r) - \max(b_l^{gt} - b_l)] \times [\min(b_b^{gt} - b_b) - \max(b_t^{gt} - b_t)] \tag{4}$$

$$union = (w^{gt} \times h^{gt}) \times (ratio)^2 + (w \times h) \times (ratio)^2 - inter \tag{5}$$

$$L_{InnerSIoU} = 1 - IoU^{Inner} + \frac{\Delta + \Omega}{2} \tag{6}$$

式中：inter 为两个框的交集面积，union 为两个框的并集面积，b_l、b_r、b_t、b_b 为预测框的左、右、上、下边界，b_l^{gt}、b_r^{gt}、b_t^{gt}、b_b^{gt} 为真实框的左、右、上、下边界，w、h、w^{gt}、h^{gt} 为预测框及真实框的宽和高。

NWD 为一种针对小目标改进的损失函数，其采用新的度量方法来计算框与框之间的相似度，使两个框即便完全不重叠，或者重叠很少，仍能度量出相似度出来。表达式如下：

$$NWD(b,b^{gt}) = \frac{\sqrt{\rho^2 + \left(\frac{w - w^{gt}}{2}\right)^2 + \left(\frac{h - h^{gt}}{2}\right)^2}}{\sqrt{\left(\frac{w - h}{4}\right) + \left(\frac{w^{gt} - h^{gt}}{4}\right)}} \tag{7}$$

$$L_{NWD} = 1 - NWD \tag{8}$$

对 NWD 和 InnerSIoU 进行加权融合，得到新损失函数表达式为：

$$L_{NWD_InnerSIoU} = (1 - \theta)E[L_{NWD}] + \theta \cdot E[L_{InnerSIoU}] \tag{9}$$

其中，$E[\cdot]$ 代表对样本取均值；θ 为平衡 InnerSIoU 与 NWD 的约束，本文取值为 0.5。

3　实验结果与分析

3.1　数据集

本文实验采用 RGBT-Tiny[19] 中的可见光数据集，数据集作者用无人机从 60~100 m 的高度收集 RGBT 视频序列。数据集在 4 个城市的 8 种场景下（海洋、湖泊、桥梁、城市道路、乡村道路、操场、机场和天空）拍摄并标注 7 类目标（船、汽车、自行车、行人、公交车、无人机和飞机）。数据集将小尺度分为 3 个级别：极其微小 $\{1^2,8^2\}$、微小 $\{8^2,16^2\}$、小型 $\{16^2,32^2\}$，其中超过 97% 的目标属于小型或更小型。数据集中共包含 46 701 张图片，本文将数据集按 8∶2 的比例划分为训练集与验证测试集，训练集含 37 360 张图片，验证测试集含 9 341 张图片。

3.2　环境配置与评价指标

实验平台为 ubuntu18.04 系统，CPU 为 E5-1650，显卡为 NVIDIA GeForce GTX 1080 Ti，11 GB 显存。环境配置：python 版本为 3.8.20，torch 版本为 2.4.1，cuda 版本为 11.8。模型训练参数见表 1。

表 1　模型训练相关参数

Table 1　Model training related parameters

参数项	参数值
lrf	0.010
lr0	0.010
momentum	0.937
workers	8
imgsz(pixel×pixel)	640×640
epochs	60
batch_size	8

输入图片尺寸为 640×640，训练迭代次数为 60 轮，批处理数 batch_size 设置为 8。其余实验参数均为系统的默认设置。

本实验选取精确率（Precision）、召回率（Recall）、mAP_{50}、$mAP_{50\text{-}95}$ 四种评价指标来衡量模型的性能：

1）精确率：表示在模型预测为正确的结果中，实际正确的结果所占的比例。

2）召回率：表示在实际为正确的结果中，被模

型成功预测为正确的结果所占的比例。

3) mAP_{50}:表示当模型预测的 IoU 阈值为 0.5 的条件下,目标检测模型的平均精确率。

4) mAP_{50-95}:表示当模型预测 IoU 阈值从 0.5 起以 0.05 为间隔逐步提升至 0.95 的条件下,模型平均精确率的均值。

3.3 对比实验

3.3.1 损失函数对比实验

为了体现出改进的 NWD_InnerSIoU 损失函数具有优越性,将其与一些其他损失函数进行一组对比实验,实验结果见表 2。

表 2 不同损失函数对比实验

Table 2 Comparison experiment of different loss functions %

损失函数	精确率	召回率	mAP_{50}	mAP_{50-95}
CIoU	91.2	77.6	83.0	49.4
InnerDIoU	91.3	77.5	83.1	49.4
FocalerIoU	87.6	73.5	78.9	46.0
NWD_InnerSIoU	**91.2**	**78.1**	**83.8**	**49.6**

3.3.2 消融实验

为了验证每个改进策略的有效性,对所有改进策略以原始的 YOLOv5s 模型为基础进行消融实验,实验结果见表 3。模型 A 表示未经改进过的原始 YOLOv5s 模型;模型 B 表示添加了小目标层检测头的模型;模型 C 表示用 DySample 模块替换原有上采样函数的模型;模型 D 表示添加了 GAM 注意力机制的模型;模型 E 表示使用 NWD_InnerSIoU 损失函数的模型。随着改进策略的逐步融合,模型的精确率、召回率、mAP_{50}、mAP_{50-95} 指标逐步优化。模型 H 集成全部策略后,精确率达 94.1%、mAP_{50} 达 90.9%。

3.3.3 与现有其他模型对比

为验证改进模型在目标检测任务中的性能优势,针对数据集特性,将改进后的 YOLOv5s 模型与其他目标检测模型 YOLOv3 - tinyu、YOLOv5s、YOLOv7 - Tiny[20]、YOLOv8s[21] 以及 YOLOv11s[22] 进行对比,实验数据见表 4 ,以此定量分析改进模型的有效性与合理性。

表 3 消融实验

Table 3 Ablation Experiments

模型	P2	DySample	GAM	NWD_InnerSIoU	精确率/%	召回率/%	mAP_{50}/%	mAP_{50-95}/%
A					91.2	77.6	83.0	49.4
B	√				93.6	83.8	89.4	54.0
C		√			91.9	77.7	83.3	49.6
D			√		91.5	77.8	83.3	49.6
E				√	91.0	78.1	83.8	49.6
F	√	√			93.3	84.7	90.1	54.0
G	√	√	√		94.0	85.2	90.4	54.3
H	√	√	√	√	**94.1**	**84.8**	**90.9**	**54.4**

表 4 对比实验

Table 4 Comparative experiment %

模型	精确率	召回率	mAP_{50}	mAP_{50-95}
YOLOv3-tinyu	78.1	42.9	52.0	30.2
YOLOv5s	91.2	77.6	83.0	49.4
YOLOv7-Tiny	80.1	68.8	59.7	22.6
YOLOv8s	89.0	75.3	81.6	50.7
YOLOv11s	88.6	75.0	81.2	50.2
本文	**94.1**	**84.8**	**90.9**	**54.4**

图 4 为本文算法的训练过程可视化结果。从曲线趋势看,随着训练迭代次数的增加,训练集与验证集的各类损失均逐步下降并趋于稳定,表明模型对数据的拟合与泛化能力持续提升;同时,精确率、召回率及 mAP 系列指标稳步上升,说明模型在迭代过程中不断优化检测性能,最终达到较好的训练效果,验证了改进策略在模型训练阶段的有效性。

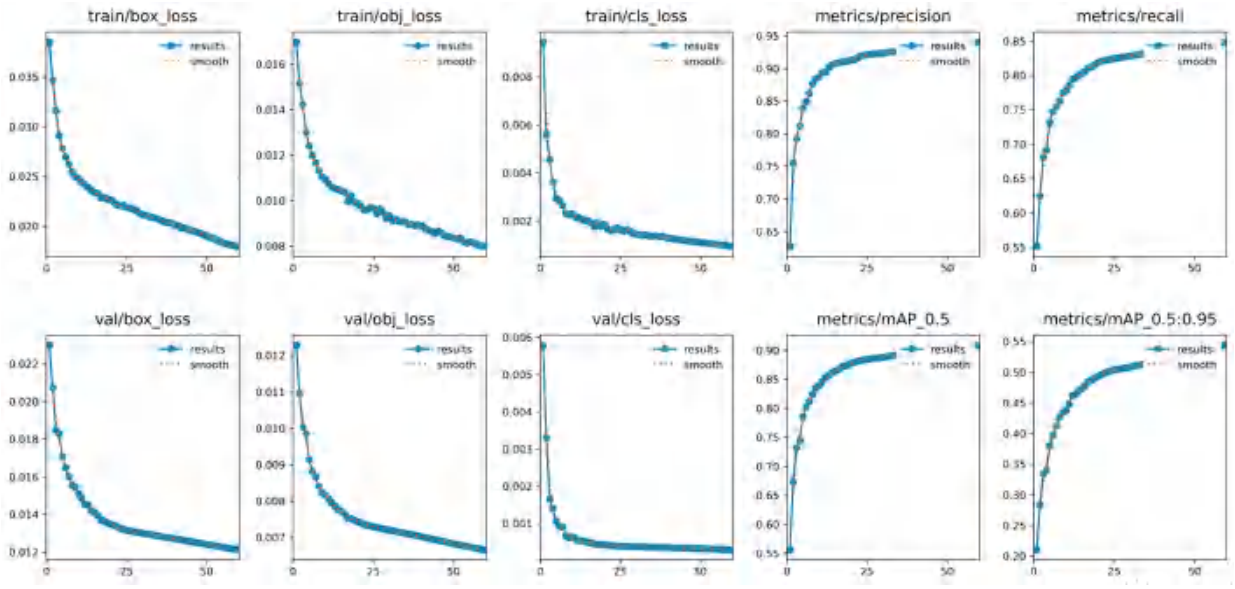
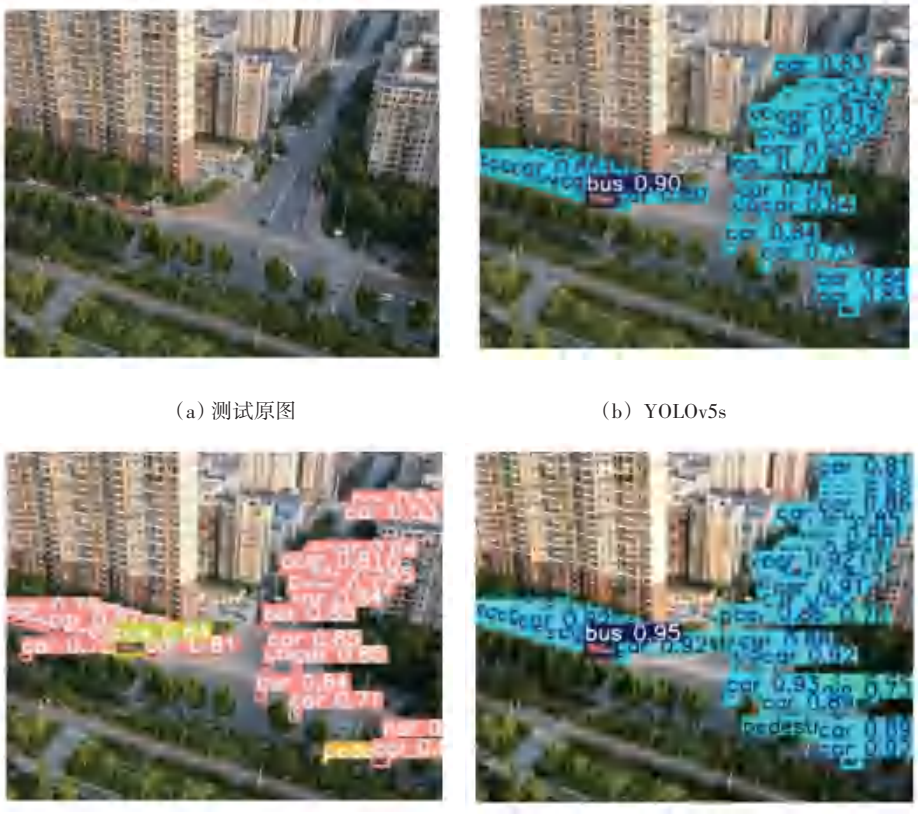


图 4　训练过程可视化结果图

Fig. 4　Visualization of the training process

3.3.4　检测结果的可视化对比

在测试集中对模型进行可视化对比，结果如图 5 所示，YOLOv5s 与 YOLOv8s 都存在行人小目标漏检的情况，而本文模型识别出了更多的行人与车辆，且置信度总体上显著高于原始模型。



（a）测试原图　　（b）YOLOv5s

（c）YOLOv8s　　（d）本文模型

图 5　检测结果可视化对比图

Fig. 5　Comparison of detection result visualization

4 结束语

实验数据显示,本文改进方案对模型性能产生了显著的提升效果。在公开的 RGBT-Tiny 数据集测试中,经过优化的 YOLOv5s 模型应用于无人机航拍小目标检测任务时,各项核心指标较原始模型均有明显进步:精确率提升 2.9 个百分点,召回率提升 7.2 个百分点,mAP_{50} 提升 7.9 个百分点,mAP_{50-95} 提升 5.0 个百分点。这一结果表明模型的识别效果得到有效改善,且与其他目标检测模型相比,具备更为突出的性能优势,拥有较高的实用价值和广阔的应用前景。然而该模型仍具有一定的局限性,未来的研究应进一步拓展模型的泛化能力和适用性,在保证精度的情况下进一步降低模型参数量,使其能够适配无人机检测场景下的实时运行需求。

参考文献

[1] KANISTRAS K, MARTINS G, RUTHERFORD M J, et al. A survey of unmanned aerial vehicles (UAVs) for traffic monitoring [J]. Handbook of UAVs, 2015, 34(2): 2643-2666.

[2] KORKI M, SHANKAR N D, SHAH R N, et al. Automatic fault detection of power lines using unmanned aerial vehicle (UAV) [C]//Proceedings of the 1st International Conference on Unmanned Vehicle Systems-Oman (UVS). Piscataway, NJ: IEEE, 2019: 1-6.

[3] 郭辰阳, 敖万忠, 吕宜宏. 低空经济与通用航空、无人机、UAM 的关系分析[J]. 财经界, 2023(28): 30-32.

[4] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 42(2): 318-327.

[5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shotmultibox detector [C]// Proceedings of the European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[6] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. arXiv preprint arXiv,1804.02767,2018.

[7] 李斌, 张彩霞, 杨阳, 等. 复杂场景下深度表示的无人机目标检测算法[J]. 计算机工程与应用, 2020, 56(16): 118-123.

[8] 刘辉, 张琳玉, 王复港, 等. 基于注意力机制和上下文信息的目标检测算法[J]. 计算机应用, 2023,43(5):1557-1564.

[9] 祝星旭, 蒋球伟. 基于 CNN 与 Transformer 的无人机图像目标检测研究[J]. 武汉理工大学学报(信息与管理工程版), 2022, 44(2): 323-331.

[10] LIANG B, SU J, FENG K, et al. Cross-layer triple-branch parallel fusion network for small object detection in UAV images [J]. IEEE Access, 2023, 11: 39738-39749.

[11] 贾星宇, 李大鹏. 基于 YOLO11 的无人机航拍图像小目标检测算法[J]. 无线电工程, 2025, 55(8): 1560-1570.

[12] 李博豪, 李智. 基于改进 YOLOv8n 的无人机航拍小目标检测算法[J]. 智能计算机与应用,2025,15(5):173-179. DOI:10.20169/j.issn.2095-2163.24120902.

[13] 代涛, 黎青松, 邓鹏毅. 基于改进 YOLOv5 的小目标检测算法研究[J]. 智能计算机与应用,2025,15(9):132-138. DOI:10.20169/j.issn.2095-2163.250921.

[14] LIU W, LU H, FU H, et al. Learning toupsample by learning to sample [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2023: 6004-6014.

[15] LIU Y, SHAO Z, HOFFMANN N. Global attention mechanism: Retain information to enhance channel-spatial interactions [J]. arXiv preprint arXiv,2112.05561,2021.

[16] WANG J, XU C, YANG W, et al. A normalized Gaussian Wasserstein distance for tiny object detection[J]. arXiv preprint arXiv,2110.13389,2021.

[17] ZHANG H, XU C, ZHANG S. Inner-IoU: More effective intersection over union loss with auxiliary bounding box[J]. arXiv preprint arXiv,2311.02877,2024.

[18] GEVORGYAN Z. SIoU loss: More powerful learning for bounding box regression[J]. arXiv preprint arXiv,2205.12740, 2022.

[19] YING X, WANG M, LI W, et al. Visible-thermal tiny object detection: A benchmark dataset and baselines [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(7):6088-6096. DOI: 10.1109/TPAMI.2025.3544621.

[20] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 7464-7475.

[21] VARGHESE R, SAMBATH M. Visible-thermal tiny object detection: A benchmark dataset and baselines[C]// Proceedings of 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). Piscataway, NJ: IEEE, 2024: 1-6.

[22] KHANAM R, HUSSAIN M. YOLOv11: An overview of the key architectural enhancements[EB/OL]. arXiv preprint arXiv,2410.17725,2024. https://arxiv.org/abs/2410.17725.