

doi:10.13682/j.issn.2095-6533.2026.05.012

任务动态对齐的无人机目标检测网络

王殿伟,冯梦圆,房杰,李苑青,艾达

(西安邮电大学 通信与信息工程学院,陕西 西安 710121)

摘要: 大多数无人机目标检测模型通过两个平行分支实现分类和定位,任务间缺乏交互导致预测存在空间错位,并且大尺寸模型不满足无人机场景快速检测的需求。针对上述问题,提出基于任务动态对齐的轻量级检测模型。首先,构建了任务动态对齐检测头,通过边框和类别对齐统一定位和分类的最佳锚点;接下来,设计了高效特征提取模块,减少模型推理时间;最后,构造了基于尺度的特征筛选金字塔网络,策略性地融合多尺度特征,并创建了数据集验证所提方法的泛化性。在 VisDrone2019-test 数据集的结果表明,与 YOLOv8s 相比,所提网络模型尺寸减少了 43.1%,计算量减少了 14.2%,平均精度 mAP50 提高了 8.5%,在无人机航拍图像目标检测中取得了性能和资源消耗间的平衡。

关键词: 目标检测;无人机航拍图像;YOLOv8;任务动态对齐检测头;高效特征提取模块

中图分类号: TP391.4

文献标志码: A

文章编号: 2095-6533(2026)05-0126-11

Object detection network with task dynamic-aligned for UAVs

WANG Dianwei, FENG Mengyuan, FANG Jie, LI Yuanqing, AI Da

(School of Communications and Information Engineering, Xi'an University of Posts and
Telecommunications, Xi'an 710121, China)

Abstract: Most unmanned aerial vehicle (UAV) target detection models achieve classification and localization through two parallel branches. The lack of interaction between tasks leads to spatial misalignment of predictions, and the large-size model does not meet the demand for fast detection of UAV scenes. To address the above problems, a lightweight detection model based on task dynamic-aligned is proposed. Firstly, a task dynamic-aligned detection head is constructed to unify the optimal anchor points for localization and classification through edge and category alignment; next, an efficient feature extraction module is designed to reduce model inference time; finally, a scale-based feature screening pyramid network is constructed to strategically fuse multi-scale features, and a dataset is created to validate the generalization of the proposed method. The results on the VisDrone2019-test dataset show that the proposed model size is reduced by 43.1%, the computational resources are reduced by 14.2%, and the average accuracy mAP50 is improved by 8.5% compared to YOLOv8s, which achieves a balance between performance and resource consumption in object detection in UAV aerial images.

Keywords: object detection; unmanned aerial vehicle aerial images; YOLOv8; task dynamic-aligned detection head; effective feature extraction module

收稿日期: 2025-03-10

基金项目: 国家自然科学基金项目(62201454,62306235);西安市科学技术局社会发展创新示范项目(23SFSF0004);陕西省教育厅科学研究计划项目(23JK0670)

引文格式: 王殿伟,冯梦圆,房杰,等. 任务动态对齐的无人机目标检测网络[J]. 西安邮电大学学报,2026,31(5):126-136.

WANG Dianwei, FENG Mengyuan, FANG Jie, et al. Object detection network with task dynamic-aligned for UAVs[J]. Journal of Xi'an University of Post and Telecommunication, 2026,31(5):126-136.

近年来,无人机在电力线路巡检、交通监控和农作物分析等领域的应用越来越广泛^[1-5]。目标检测技术作为无人机执行任务的关键环节,具有十分重要的研究意义。

传统检测方法依赖手动设计,效率低下。近年来,深度学习提供了丰富的模型,使目标检测成为热点。目前,基于深度学习的目标检测算法分为两大类:双阶段检测算法和单阶段检测算法。基于区域的卷积神经网络(Region-based Convolutional Neural Network, R-CNN)^[6]是经典双阶段算法,提取候选帧,删除重复框并用非极大值抑制微调。基于双向特征金字塔的数据融合目标检测网络^[7-8]通过引入注意力残差提高多尺度目标检测鲁棒性。双阶段算法在精度上有优势,但检测速度不够理想。为解决这个问题,Redmon 等^[9]提出 YOLO (You Only Look Once)单阶段算法,该算法使用统一网络框架学习类别概率和边界框坐标。Ge 等^[10]在 YOLO 中引入无锚点概念来提升检测精度。YOLOv8 采用多尺度特征融合,增强了对不同尺度目标检测效果。YOLOv9^[11]引入可编程梯度信息确保目标重要特征保留。大多数检测模型在检测和定位这两个任务间缺乏交互,从而影响了整体性能。此外,随着检测算法不断改进,精度得到提高,但模型尺寸随之增加,造成了很大的负担。

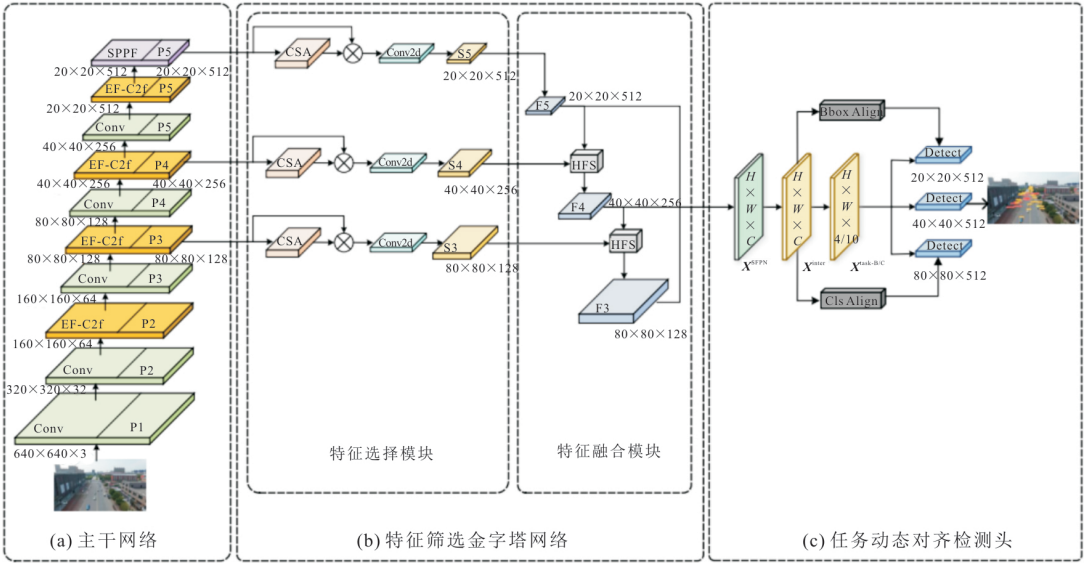
无人机航拍图像目标检测在计算机视觉领域得到越来越多关注,涌现大量研究工作。王殿伟等^[12]通过改进 Double-Head R-CNN 算法提高无人机航拍图像目标检测精度。江祥奎等^[13]通过添加轻量化卷积注意力模块来提取深层特征。TPH-YOLOv5^[14]使用 Transformer 预测头、卷积块注意力机制和数据增强策略来解决航拍图像俯视角度的尺度变化问题。这种方法有效改善多尺度目标

的检测效果,但使模型的结构复杂化。Du 等^[15]从轻量化角度设计稀疏卷积优化头部,减少计算量,但降低了模型检测精度。综上所述,现有目标检测算法在检测和定位两个任务间缺乏交互,很难获得检测精度和运算开销之间的平衡。

针对检测中分类和定位缺乏交互的问题,从推理时间和检测精度两方面综合考虑,提出基于任务动态对齐的轻量级检测模型(Task Dynamic-Aligned YOLO,TD-YOLO)。首先,构建任务动态对齐检测头(Task Dynamic-aligned Detection Head, TDDH)。通过边框对齐(Bbox Align)和类别对齐(Cls Align),平衡分类和定位任务的独立性。其次,设计高效特征处理块(Efficient Feature-extractor Block, EFB),用 EFB 重构 C2f (Cross Stage Partial Fusion with 2 Convolutions,C2f)得到快速特征提取模块(Efficient Faster-C2F, EF-C2f)。EFB以简单有效的方式利用特征冗余减少计算。EF-C2f 采用了梯度分流连接,丰富提取特征信息,减少模型尺寸。最后,构造基于尺度的特征筛选金字塔网络(Scale-based Feature-screening Pyramid Network,SFPN),将高级特征作为权重,筛选嵌入在低级特征的语义信息,增强模型的表达能力并降低模型推理时间。

1 任务动态对齐的目标检测网络

YOLOv8 的检测头采用两个解耦的独立分支进行分类和预测,这导致任务之间缺乏交互。针对上述问题,提出 TD-YOLO,主要由 3 部分组成:主干网络、特征筛选金字塔网络和任务动态对齐检测头,改进后的网络结构如图 1 所示,网络详细模块如图 2 所示。



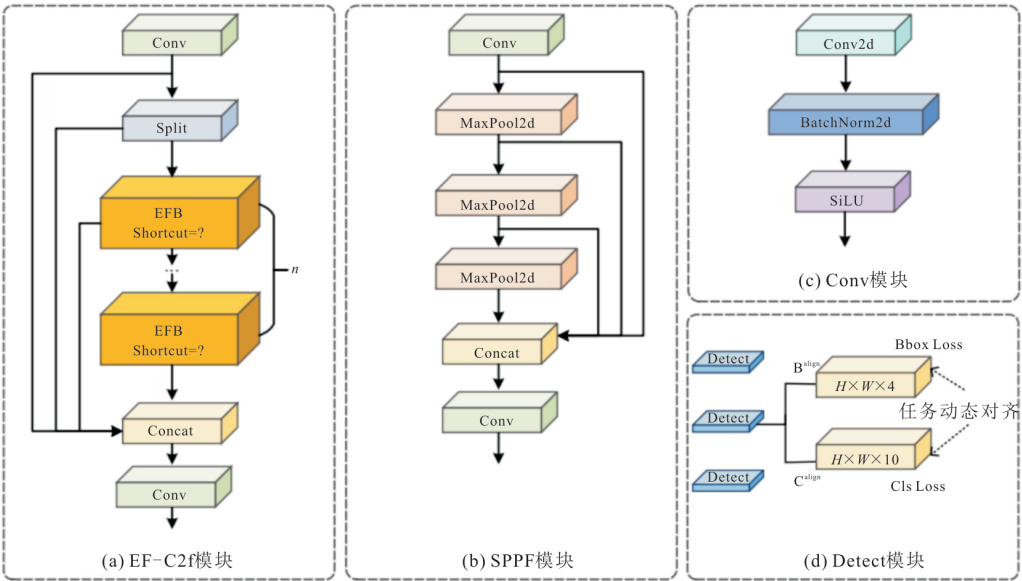


图 2 TD-YOLO 网络详细模块

主干网络由 Conv、EF-C2f 和 SPPF 模块组成。Conv 模块对输入特征进行 5 次下采样生成不同尺度的特征,记为 $P1 \sim P5$ 。EF-C2f 和 SPPF 模块进行快速特征提取,其结构如图 2(a)和图 2(b)所示。特征筛选金字塔网络首先通过特征选择模块对 $P3$ 、 $P4$ 和 $P5$ 层特征进行初步过滤得到 $S3$ 、 $S4$ 和 $S5$ 层特征,然后通过特征融合模块进行策略性融合生成 $F3$ 、 $F4$ 和 $F5$ 层特征。任务动态对齐检测头通过边框对齐和类别对齐,逐步拉近定位和

分类这两个任务的最优锚点,实现小、中、大目标多尺度预测。

1.1 任务动态对齐检测头

YOLOv8 使用独立解耦的检测头实现分类和定位,传统解耦检测头与任务动态对齐检测头结构如图 3 所示,但这两种任务的学习机制不同,在预测时存在空间错位。针对上述问题,构建了 TDDH,逐步拉近甚至统一两个任务的最优锚点,其结构如图 3(c)所示。

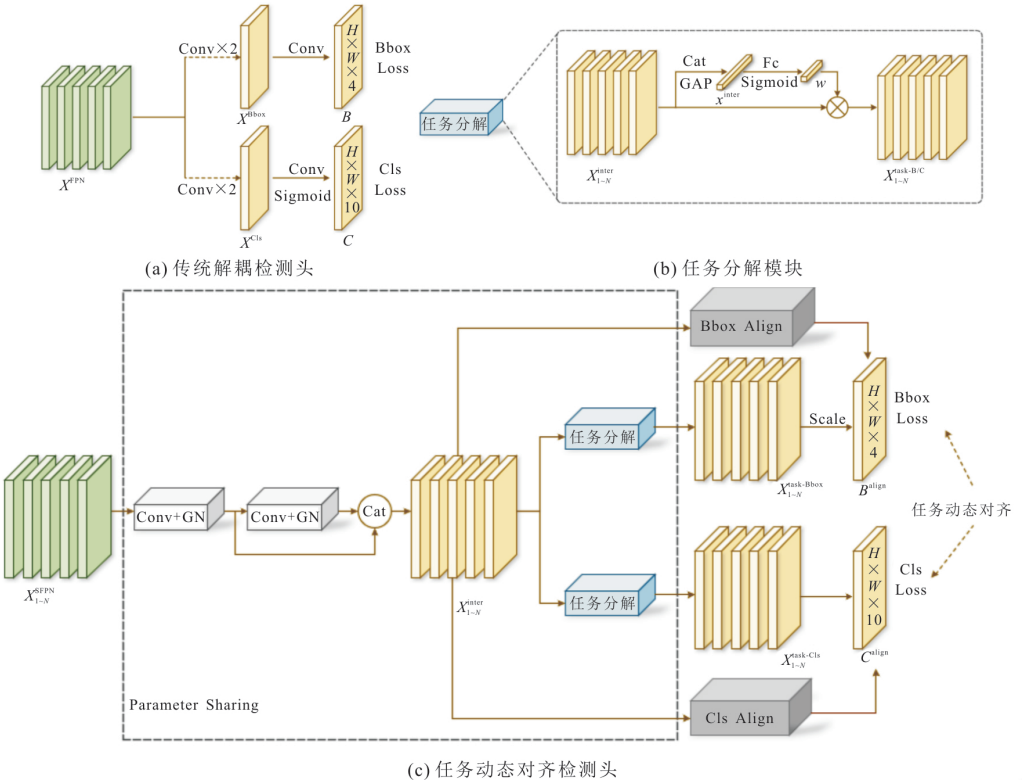


图 3 传统解耦检测头与任务动态对齐检测头结构

为增强交互性,使用两组卷积、组归一化和残差连接,从多尺度特征 $\mathbf{X}^{\text{SPFN}}$ 中学习交互特征 $\mathbf{X}^{\text{inter}}$ 。由于分支的单一性,特定任务无法进行有效分解。因此通过任务分解模块计算特定任务的特征,其结构如图 3(b)所示。针对分类或定位任务,分别计算任务特征

$$\mathbf{X}_k^{\text{task-Bbox}} = \mathbf{w}_k^{\text{Bbox}} \cdot \mathbf{X}_k^{\text{inter}}, k \in \{1, 2, \dots, N\} \quad (1)$$

$$\mathbf{X}_k^{\text{task-Cls}} = \mathbf{w}_k^{\text{Cls}} \cdot \mathbf{X}_k^{\text{inter}}, k \in \{1, 2, \dots, N\} \quad (2)$$

式中: $\mathbf{w}_k$ 是任务分解模块 $\mathbf{w} \in \mathbf{R}^N$ 第 k 个元素; $\mathbf{w}$ 由跨层交互特征 $\mathbf{X}^{\text{inter}}$ 计算得到^[16]

$$\mathbf{w} = \sigma(\text{ReLU}(\text{Fc}(\mathbf{x}^{\text{inter}}))) \quad (3)$$

式中: $\mathbf{x}^{\text{inter}}$ 是通过对 $\mathbf{X}_k^{\text{inter}}$ 的级联特征平均池化得到的; σ 为 Sigmoid 函数。通过坐标到真值的距离^[17]将 $\mathbf{X}_k^{\text{task-Bbox}}$ 转换为目标边界框 $\mathbf{B} \in \mathbf{R}^{H \times W \times 4}$, 并使用 Sigmoid 函数将 $\mathbf{X}_k^{\text{task-Cls}}$ 转换为分类分数 $\mathbf{C} \in \mathbf{R}^{H \times W \times 10}$ 。最后,从 $\mathbf{B}/\mathbf{C}$ 中预测定位和分类的结果。在预测时,由 $\mathbf{X}^{\text{inter}}$ 联合考虑两个任务来对齐这两个预测的空间分布: $\mathbf{B}$ 和 $\mathbf{C}$, 在两个任务上分别执行对齐方法,如图 3(c) 中 Bbox Align 和 Cls Align 所示。Bbox Align 使用可变型卷积^[18] (Deformable Convolution v2, DCNv2) 动态地生成 offset 和 mask, 自动调整卷积形状来改变尺度或感受野,从而调整定位预测为

$$\mathbf{B}^{\text{align}} = \mathbf{B} + \text{DCN}(\mathbf{X}^{\text{inter}}) \quad (4)$$

式中: $\text{DCN}(\cdot)$ 是可变型卷积,通过空间偏移量使最对齐的锚点识别其周围的最佳边界预测,最后用 scale 层进行特征缩放。Cls Align 进一步从 $\mathbf{X}^{\text{inter}}$ 中进行动态特征选择,对齐分类预测为

$$\mathbf{C}^{\text{align}} = \mathbf{C} \cdot \sigma(\text{ReLU}(\text{Conv}(\mathbf{X}^{\text{inter}}))) \quad (5)$$

$\mathbf{B}_{\text{box}}$ Align 和 Cls Align 与任务对齐学习^[16] (Task Alignment Learning, TAL) 协同工作,提高两个任务的对齐度。这里 TAL 利用预测结果在每个锚点计算任务对齐度量,为 TDDH 产生学习信号。TDDH 在反向传播过程中根据该学习信号动态调整分类和定位预测。针对任务对齐后参数量增加,采用权重共享^[17],用同一批参数共享权重来实现轻量化,如图 3(c) 虚线框所示,提高检测头的计算效率,降低模型复杂性。

1.2 高效特征提取模块

不同通道的特征图具有很高相似性,在提取特征过程中存在冗余计算。部分卷积 (Partial Convolution, PConv)^[19] 对连续通道数 C_P 的特征卷积,

其余通道恒等映射,有效减少了冗余计算。基于 PConv 设计的高效特征处理块,其结构如图 4 所示。

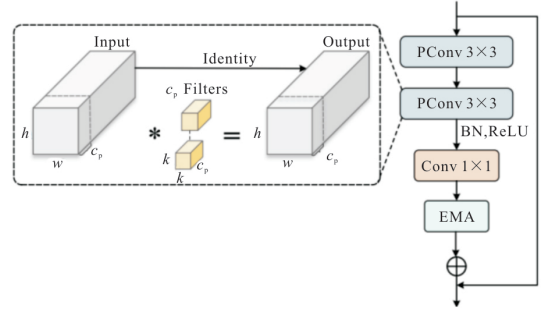


图 4 高效特征处理块结构

高效多尺度注意力^[20-21] (Efficient Multi-Scale Attention, EMA) 使空间语义信息均匀分布,保留每个通道信息并降低计算开销。在第二个 PConv 后使用批归一化和 ReLU 激活,保证特征多样性并实现更低的延迟。EFB 模块在特征处理过程中具有较少的计算和内存访问。模型用 EFB 重构 YOLOv8 主干网络的 C2f, EF-C2f 结构如图 2(a) 所示 (n 表示 EFB 模块数量)。EF-C2f 模块的结构简单,能高效快速地提取特征。

1.3 特征筛选金字塔网络

受多层次特征融合的血细胞精确检测方法^[22] 的启发,模型构造 SFPN 实现多级融合,其结构由两部分组成:1) 特征选择模块;2) 特征融合模块,如图 1(b) 所示。SFPN 中通道选择注意力模块 (Channel Selection Attention, CSA) 和高级特征筛选模块 (High-Level Feature Screening, HFS) 的结构如图 5 所示。

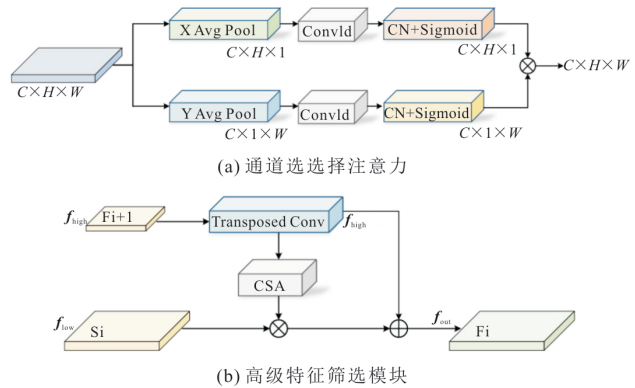


图 5 特征筛选金字塔网络模块结构

通道选择注意力模块对输入特征进行初步选择。特征图首先分别经过水平方向和垂直方向的平均池化,得到嵌入定位信息的序列。然后使用 1D 卷积处理此序列,再利用组归一化 (Group Normali-

zation, GN) 和 Sigmoid 得到位置注意力在水平和垂直方向上的表示,最后将其相乘得到注意力权重。将权重与对应尺度下的特征图相乘,生成筛选后的特征图 S3、S4 和 S5。

特征融合模块将高层特征作为权重对低层特征进行筛选,然后将高层特征与筛选后的低层特征融合。在特征筛选融合过程中,分别计算扩展特征和融合特征为

$$\widehat{f_{\text{high}}} = \text{TConv}(f_{\text{high}}) \tag{6}$$

$$f_{\text{out}} = f_{\text{low}} \cdot \text{CSA}(\widehat{f_{\text{high}}}) + \widehat{f_{\text{high}}} \tag{7}$$

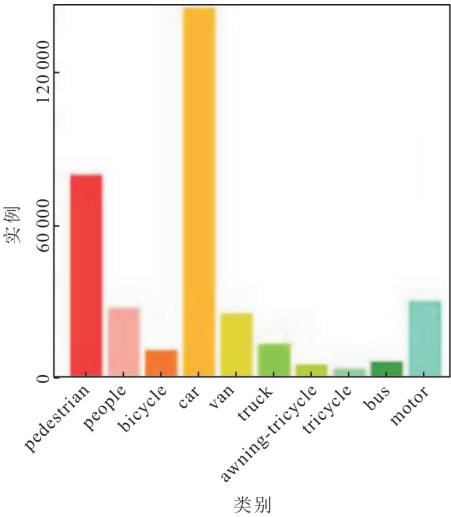
式中,转置卷积(Transposed Convolution, TConv)对高层特征 $f_{\text{high}} \in \mathbf{R}^{C \times H \times W}$ 扩展得到 $\widehat{f_{\text{high}}} \in \mathbf{R}^{C \times 2H \times 2W}$ 。高层特征通过 CSA 通道选择注意力模块过滤低层特征 $f_{\text{low}} \in \mathbf{R}^{C \times H_1 \times W_1}$,将过滤后的低层特征与高层特征进行融合得到 $f_{\text{out}} \in \mathbf{R}^{C \times H_1 \times W_1}$ 。最终将得到的多尺度特征图 F3、F4 和 F5 送到头部进行检测。

2 实验与结果分析

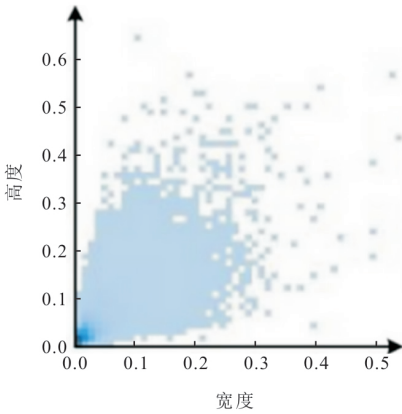
2.1 实验设置

2.1.1 数据集

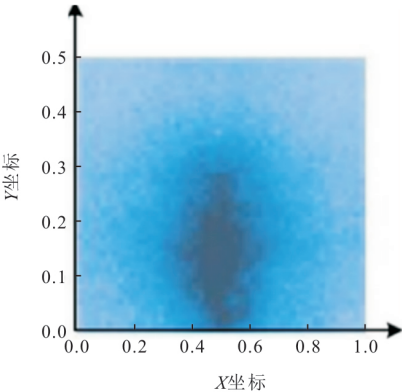
使用天津大学 Aiskyeye 团队采集的 VisDrone2019-Det 数据集^[23]来验证所提方法的有效性。该数据集包含 10 209 张不同位置、高度和光照条件的无人机航拍图像,其中 6 471 张用于训练,548 张用于验证,1 610 张用于测试。图 6 展示了 VisDrone2019-Det 数据集的标签信息。除了采用公开数据集进行评估,还用自建数据集 AerialDet 进一步验证模型泛化性。使用大疆无人机在不同时间、高度和光照条件下拍摄得到航拍图像并对所有图像进行标注,将其保存为 COCO 数据集格式。该数据集包含 8 个类别的 5 522 张图像(3 866 张用于训练,552 张用于验证,1 104 张用于测试)。图 7 展示了 AerialDet 数据集的标签信息。



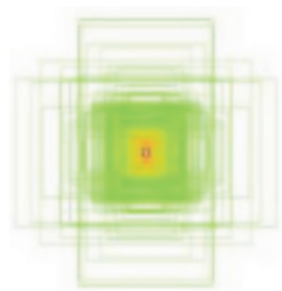
(a) 不同类别实例信息



(b) 目标相对于图像的高宽分布



(c) 目标位置空间分布



(d) 真值框大小分布

图 6 VisDrone2019-Det 数据集的标签信息

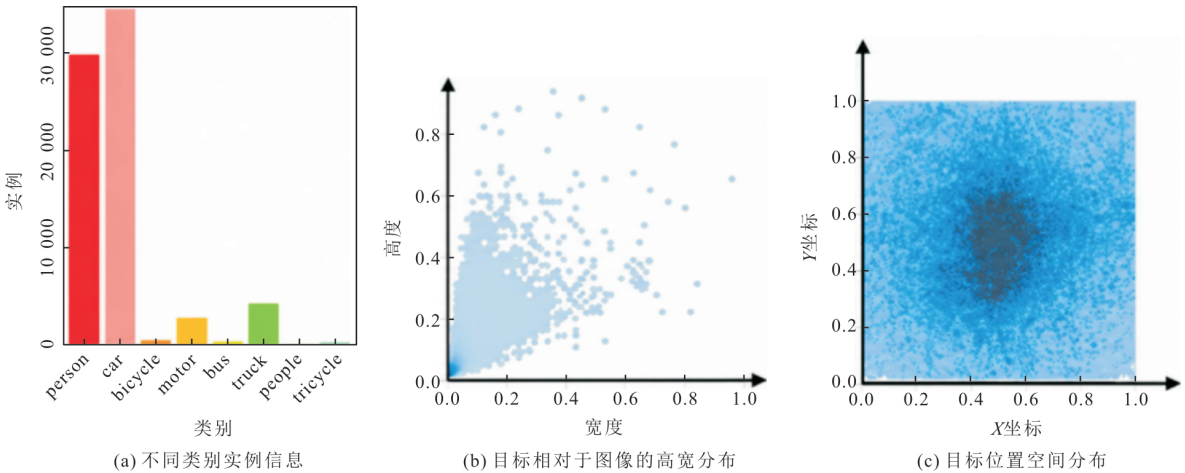


图 7 AerialDet 数据集的标签信息

2.1.2 实验平台、参数设置与评价指标

所提算法的实验平台为 CPU (Intel(R) Core (M) i7-12700K 3. 61GHz), GPU (Nvidia Titan XP), 实验环境为 CUDA11. 1、Python3. 8 和 Py-Torch1. 12. 0 框架。所提算法使用公开数据集 Vis-Drone2019 进行训练, 训练中将输入图像大小统一为 640×640 像素, Batch size 设置为 8, 学习率为 0. 001, 数据增强和水平翻转的应用率为 50 %, 使用 Adam 优化器, 共训练 200 个 epoch。

使用准确率(Precision, P)、召回率(Recall, R)、所有类别的平均精度的均值(Mean Average Precision, mAP)来评估模型的性能。此外, 还加入帧率(Frames Per Second, FPS)、模型大小(Model Size)和参数量(Parameters, Para)作为参考。

2.2 消融实验

为验证所提算法的有效性, 用 VisDrone2019 数据集进行消融实验, 实验结果如表 1 所示, √ 表示使用这种改进的策略。在骨干网络中用 EF-C2f

模块, 准确率提高 0. 6 %, 说明 EF-C2f 模块中 EMA 提高了对关键信息的关注度。EF-C2f 模块的简单结构也使模型尺寸减小 5. 5 MB。SFPN 利用无人机航拍图像中目标的尺度差距, 用高层语义信息选择性地过滤低层特征。用 SFPN 替换原模型颈部, 模型尺寸减小 7. 6 MB, FPS 增加 116, 模型推理速度大幅提升, 验证了 SFPN 特征融合的高效性。为弥补模型轻量化导致检测精度的下降, 使用 TDDH 检测头来提升精度。TDDH 模块显著地改善了检测结果, 准确率提高 2. 9 %, 召回率提高 2. 5 %。这验证了将分类和定位两个任务对齐可以提高检测性能。当结合 EF-C2f、SFPN 和 TDDH 时, 检测精度 P 提升 1. 4 %, 模型尺寸减少 7. 3 MB。TD-YOLO 模型在大部分指标上得到有效提升, 在精度和速度上达到较好的平衡。虽然加上 EF-C2f 和 SFPN 这两个轻量化的模块后对整体性能有一些影响, 但减轻了计算负担, 保证了模型的速度。

表 1 消融实验

模型	EF-C2f	SFPN	TDDH	P/%	R/%	mAP50/%	mAP50:95/%	FPS	Model Size/MB	Para /M
YOLOv8s				44. 4	34. 1	32. 6	18. 5	82	22. 5	11. 1
	√			45. 0	34. 8	32. 3	18. 4	120	17. 0	8. 3
		√		44. 3	34. 2	31. 7	18. 3	198	14. 9	7. 3
			√	47. 3	36. 6	34. 8	19. 5	147	17. 5	8. 6
	√	√		44. 1	34. 4	31. 6	18. 2	202	13. 1	6. 2
	√		√	46. 3	35. 6	34. 0	18. 9	149	17. 1	8. 0
		√	√	45. 8	35. 2	32. 6	18. 6	152	15. 2	7. 8
	√	√	√	48. 6	37. 0	35. 4	19. 8	142	12. 8	5. 7

2.3 对比实验与分析

为验证算法的性能, 在 VisDrone2019 数据集上选取 YOLOv3s^[24]、YOLOv5s、YOLOv6s^[25]、YOLOx-s^[10]、YOLOv7s^[26] 等进行对比, 并且还选取双阶段经典算法 Cascade R-CNN^[27], 单阶段经典

算法 TOOD^[16], Anchor-free 检测算法 Conner-Net^[28]、CenterNet^[29] 和 FSAF^[30], 自适应样本检测算法 ATSS^[17], 还有现阶段较为先进的 RT-DETR-R18^[31] 作为对比。对比实验的参数及评价指标与 2. 1 节所述一致, 对比实验结果如表 2 所示。

表 2 不同算法在 VisDrone2019-test 数据集上的实验结果

模型	P/%	R/%	mAP50/%	mAP50:95/%	FPS	Para /M
Cascade R-CNN	—	27.5	30.2	17.5	9	69.3
ConnerNet	—	24.5	22.1	14.3	11	36.8
CenterNet	—	27.0	29.2	18.4	22	32.2
FSAF	—	28.2	32.0	17.6	31	36.1
ATSS	—	30.6	31.6	18.5	30	31.9
TOOD	—	31.3	32.6	19.9	22	31.9
YOLOv3s	44.8	35.1	33.2	19.3	59	15.3
YOLOv5s	44.3	33.9	31.9	18.1	73	15.5
YOLOv6s	45.2	31.9	30.8	17.8	67	16.2
YOLOx-s	—	26.4	27.3	14.4	82	8.9
YOLOv7s	46.9	35.4	33.5	18.7	37	15.3
RT-DETR-R18	52.3	35.9	33.2	19.6	77	20.0
YOLOv8m	49.0	37.0	35.1	21.3	36	33.9
YOLOv8s	44.4	34.1	32.6	18.5	82	11.1
YOLOv8n	38.4	28.6	26.2	14.3	139	3.1
TD-YOLO(Ours)	48.6	37.0	35.4	19.8	142	5.7

根据表 2 的实验结果可知,提出的模型相比于其他优秀模型具有良好的检测性能和较快的推理速度。Cascade R-CNN 算法参数量为 69.3 M,FPS 仅为 9,检测速度较慢,不能满足无人机实时检测的需求。ConnerNet 提出的无锚框检测计算量降为 36.8 M,但对于密集场景小目标,检测精度 mAP50 仅为 22.1 %。FSAF 使用多尺度特征融合来考虑不同尺度下的物体,但召回率和帧率分别为 28.2 %和 31,检测效果不够理想。ATSS 提出的自适应训练样本选择机制,召回率和检测精度分别提高至 30.6 %和 31.6 %。TOOD 通过任务对齐机制,mAP50:95 达到 19.9 %,但增加了计算复杂度。YOLO 系列采用三尺度检测结构,但无法满足高比例小目标的检测需求。YOLOx 参数量减少至 8.9 M,但与所提算法相比,模型仍不够轻量。检测精度 mAP50 为 27.3 %,低于所提算法。RT-DETR 在检测精度上具有优势,但召回率并不理想。究其原因,该模型的规模远高于其他模型,而无人机航拍图像的目标尺寸较小,复杂的网络学习到大量无用背景特征,检测效果不理想。提出的 TD-

YOLO FPS 为 142,领先于其他模型,满足无人机场景下快速检测的需求。此外,由于任务对齐机制,召回率和平均精度分别为 37.0 %和 65.4 %,均位于其他模型的前列。而准确率为 48.6 %,仅比 RT-DETR 的准确率低 3.7 %,参数量仅比 YOLOv8n 高 1.7 M。总体而言,所提算法在检测精度和检测速度上取得较好的平衡,能够满足无人机场景下目标检测需求。

为进一步评估模型的泛化性能,在 AerialDet 数据集上进行实验。对比实验的参数及评价指标与 2.1 节所述一致,对比实验结果如表 3 所示。与基准模型 YOLOv8 相比,改进后的模型在 AerialDet 数据集上的准确率、平均精度 mAP50 和 mAP50:95 分别提高 4.3 %、1.6 %和 0.8 %,均位于其他 YOLO 系列模型的前列,并且参数量从 11.1 M 大幅减少至 5.7 M。由于 AerialDet 数据集中人(person)和汽车(car)这类小目标占比增加,背景更复杂,改进模型的召回率为 80.0 %,仅次于 YOLOv7s。改进后的模型在实现模型轻量化的同时,仍然具有较高的检测精度,表明所提算法泛化性较好。

表 3 不同算法在 AerialDet 数据集上的实验结果

模型	P/%	R/%	mAP50/%	mAP50:95/%	FPS	Para /M
YOLOv3s	79.0	73.9	78.7	47.0	51	15.3
YOLOv5s	79.6	69.0	78.6	47.8	70	15.5
YOLOv6s	79.2	67.4	75.2	44.4	67	16.2
YOLOv7s	80.9	80.0	79.6	49.1	35	15.3
YOLOv8s	80.3	77.5	82.0	52.9	75	11.1
TD-YOLO(Ours)	84.6	80.4	83.6	52.2	138	5.7

2.4 可视化结果

为直观地展示所提算法的检测效果,在 VisDrone2019 数据集上对 YOLOv7s、基准模型 YOLOv8s 和所提算法进行推理实验。选择交通路口、城市道路、和低光照人群密集场所 3 种具有代表性的场景作为实验数据。检测结果如图 8 所示。对于场景 1 交通路口,由于目标尺度变化过大, YOLOv7s 和 YOLOv8s 未能取得令人满意的检测效果。而所提算法由于策略性地融合多尺度特征,能有效识别各种尺度的行人,并通过增加分类和定位的一致性程度,使预测的边界框更加准确,从而

提高检测性能。根据场景 2 城市道路的实验结果可知 YOLOv8s 对小物体的注意力较差,对远处的物体不敏感。而所提的 TDDH 增强了分类和定位之间的交互,拉近这两个任务的最优锚点,从而对视野远端目标的检测精度大幅提升。根据场景 3 低光照人群密集场所的实验结果可以看出, YOLOv8s 可以在低光照光下检测出一些目标,但由于网络对无人机航拍目标的敏感度较低,无法检测出遮挡和密集物体。相比之下,所提的网络在低光照条件下可以准确检测和定位行人,这归因于 TDDH,使模型的注意力更加集中在物体的中心点上,增强对目标的聚焦。

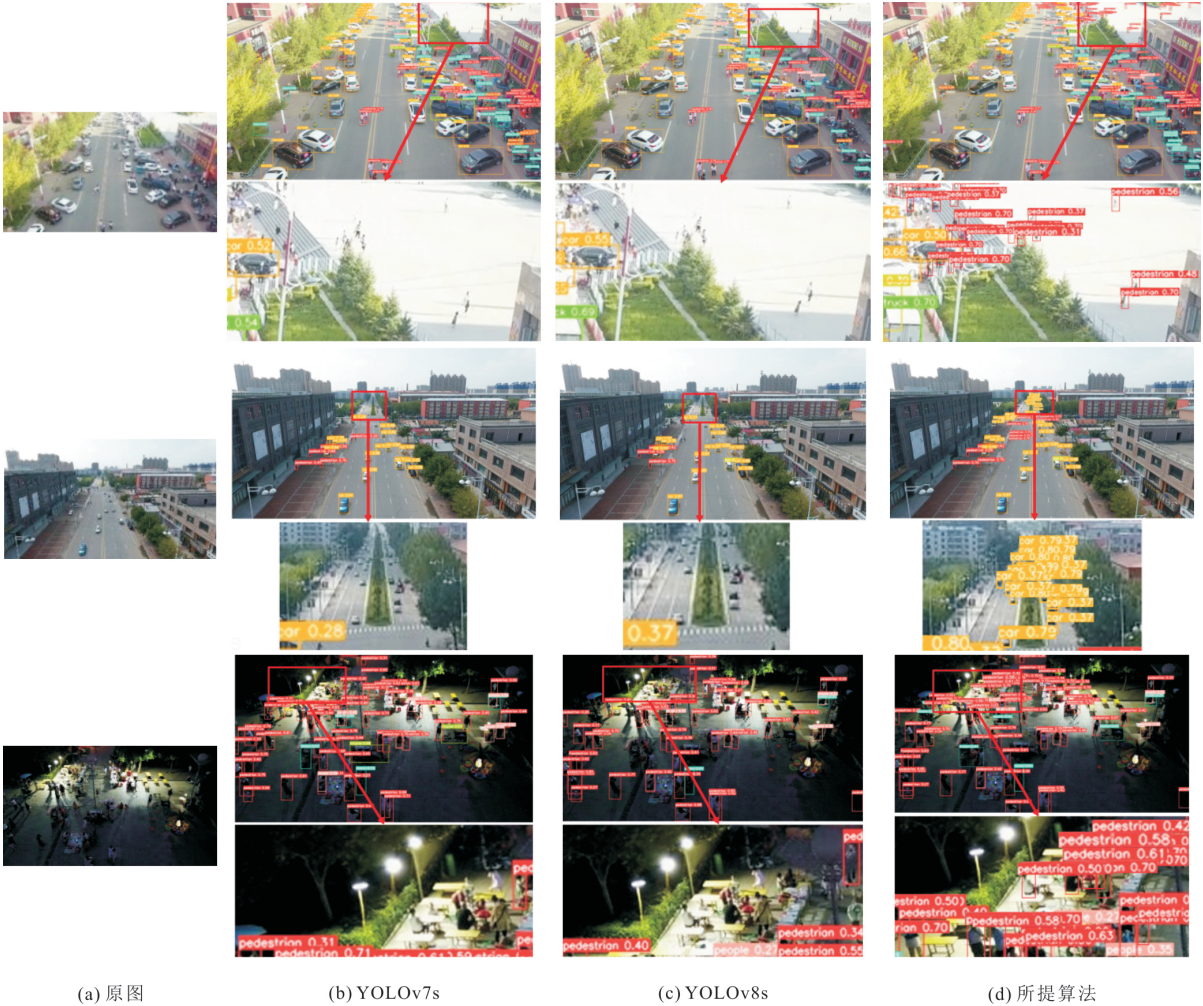
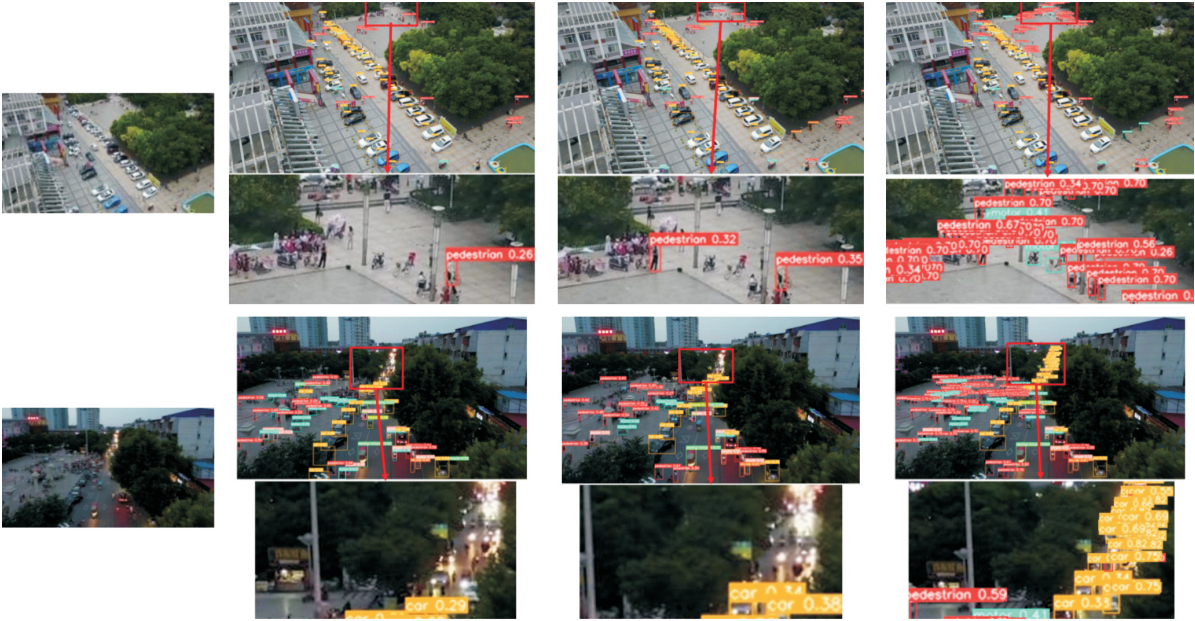


图 8 不同算法在 VisDrone2019 数据集上的检测结果

为进一步验证所提算法的泛化性,在 Aerial-Det 数据集上进行推理实验。选取具有代表性的城市交通场景,包括正常光照和低光照场景,具有一定的挑战性。图 9 展示对比实验的结果。从场景 1 实验结果可以看出,所提算法通过任务对齐,对右上方的密集小目标人群能够准确检测定位。从场景 2 实验结果可以看出,弱光条件的干扰并未对检测结果产生负面影响,证明了算法的抗干

扰能力。相比于其他模型,所提算法对视野远端目标的检测精度大幅提升。总体而言,所提算法通过将检测和分类两个任务进行对齐,有效地提高了检测性能。然而,改进后的模型虽然降低漏检率,但小类别目标被正确预测的百分比仍然较低。比如交通中的自行车类别较少,通常以密集遮挡的形式存在,在背景复杂的环境中难以被正确检测。



(a) 原图 (b) YOLOv7s (c) YOLOv8s (d) 所提算法

图 9 不同算法在 AerialDet 数据集上的检测结果

3 结 语

对于无人机航拍场景的目标检测网络,存在分类和检测两个任务之间缺乏交互的问题。考虑资源消耗并优化检测性能,在 YOLOv8 基础上提出 TD-YOLO。针对无人机快速检测的需求,EF-C2f 重构主干网络,减少了计算量和推理时间。SFPN 利用高级特征筛选低级特征,提升了对尺度变化大的目标的检测精度。TDDH 通过边框对齐和类别对齐,从联合优化的角度调整分类检测任务对齐度,使网络对航拍图像中密集小目标的检测更精准。在两个数据集上进行实验,相比基准模型,改进后的模型在保证平均检测精度小幅度提升的情况下,模型尺寸由 22.5 MB 降至 13.9 MB,参数量由 11.1 M 降至 5.7 M,为在无人机上进行精确快速地目标检测提供了可行性。

然而,计算资源的消耗仍有优化空间。同时,对于自行车等类别较少的物体,改进模型的检测精度仍然不高。因此,下一步的主要研究重点是在考虑资源消耗的同时,进一步优化模型的检测精度。

参 考 文 献

[1] 李岩,韩哲,张慕远. 电力巡检中多无人机并行式路径控制系统[J]. 电子设计工程, 2025, 33(3): 78-82.
Li Yan, Han Zhe, Zhang Muyuan. Parallel path con-

trol system for multiple unmanned aerial vehicles in electric power inspection[J]. Electronic Design Engineering, 2025, 33(3): 78-82.

[2] Li Z R, Zhang Y W, Wu H, et al. Design and application of a UAV autonomous inspection system for high-voltage power transmission lines[J]. Remote Sensing, 2023, 15(3): 865.
[3] Byun S, Shin I K, Moon J, et al. Road traffic monitoring from UAV images using deep learning networks [J]. Remote Sensing, 2021, 13(20): 4027.
[4] Bouguettaya A, Zarzour H, Kechida A, et al. A survey on deep learning-based identification of plant and crop diseases from UAV-based aerial images[J]. Cluster Computing, 2023, 26(2): 1297-1317.
[5] 朱超,孙万麟,韩晨,等. 棉花病虫害特征检测方法分析研究[J]. 电子设计工程, 2023, 31(13): 11-16.
Zhu Chao, Sun Wanlin, Han Chen, et al. Analysis and study on characteristics and detection methods of cotton diseases and pests[J]. Electronic Design Engineering, 2023, 31(13): 11-16.
[6] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580-587.
[7] 刘欣,张倩飞,刘成宇,等. 深度确定性策略梯度下运动目标识别及无人机跟随[J]. 西安工程大学学报, 2024, 38(4): 9-17.
Liu Xin, Zhang Qianfei, Liu Chengyu, et al. Moving

- target recognition and unmanned aerial vehicle followingbased on deep deterministic policy gradient[J]. Journal of Xi'an Polytechnic University, 2024, 38(4): 9-17.
- [8] 张浩钧, 郭建峰, 张锦忠, 等. 基于双向特征金字塔的数据融合目标检测网络[J]. 西安邮电大学学报, 2023, 28(1): 92-103.
- Zhang Haojun, Guo Jianfeng, Zhang Jinzhong, et al. Data fusion object detection network based on bidirectional feature pyramid[J]. Journal of Xi'an University of Posts and Telecommunications, 2023, 28(1): 92-103.
- [9] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 779-788.
- [10] Ge Z, Liu S T, Wang F, et al. YOLOX: Exceeding YOLO series in 2021[EB/OL]. [2024-10-20]. <https://doi.org/10.48550/arXiv.2107.08430>.
- [11] Wang C Y, Yeh I H, Mark Liao H Y. YOLOv9: learning what you want to Learn using programmable gradient information [C]//Computer Vision-ECCV 2024. Cham: Springer Nature Switzerland, 2025: 1-21.
- [12] 王殿伟, 胡里晨, 房杰, 等. 基于改进 Double-Head RCNN 的无人机航拍图像小目标检测算法[J]. 北京航空航天大学学报, 2024, 50(7): 2141-2149.
- Wang Dianwei, Hu Lichen, Fang Jie, et al. Small target detection algorithm based on improved Double-Head RCNN for UAV aerial images[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(7): 2141-2149.
- [13] 江祥奎, 杜遥遥, 胡浩昌. 一种改进 YOLOv5s 小目标无人机实时检测算法[J]. 西安邮电大学学报, 2023, 28(3): 88-96.
- Jiang Xiangkui, Du Yaoyao, Hu Haochang. An improved YOLOv5s real-time detection algorithm for small target UAVs[J]. Journal of Xi'an University of Posts and Telecommunications, 2023, 28(3): 88-96.
- [14] Zhu X K, Lyu S C, Wang X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]//2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal: IEEE, 2021: 2778-2788.
- [15] Du B W, Huang Y C, Chen J X, et al. Adaptive sparse convolutional networks with global context enhancement for faster object detection on drone images [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 13435-13444.
- [16] Feng C J, Zhong Y J, Gao Y, et al. TOOD: Task-aligned one-stage object detection [C]//2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021: 3490-3499.
- [17] Zhang S F, Chi C, Yao Y Q, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 9756-9765.
- [18] Zhu X Z, Hu H, Lin S, et al. Deformable convNets V2: More deformable, better results[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 9300-9308.
- [19] Chen J R, Kao S H, He H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 12021-12031.
- [20] Ouyang D L, He S, Zhang G Z, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing. Rhodes Island: IEEE, 2023: 1-5.
- [21] 刘宝宝, 李广哲, 王峰萍, 等. 基于 LSK-YOLO 算法的合成孔径雷达图像舰船目标检测[J]. 西安工程大学学报, 2025, 39(6): 48-56.
- Liu Baobao, Li Guangzhe, Wang Fengping, et al. Ship target detection for synthetic aperture radar images based on LSK-YOLO algorithm[J]. Journal of Xi'an Polytechnic University, 2025, 39(6): 48-56.
- [22] Chen Y F, Zhang C Y, Chen B, et al. Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases[J]. Computers in Biology and Medicine, 2024, 170: 107917.
- [23] Zhu P F, Du D W, Wen L Y, et al. VisDrone-VID2019: The vision meets drone object detection in video challenge results[C]//2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul: IEEE, 2019: 227-235.
- [24] Redmon J, Farhadi A. YOLOv3: An incremental improvement[EB/OL]. [2024-10-20]. <https://doi.org/10.48550/arXiv.1804.02767>.
- [25] Li C Y, Li L L, Jiang H L, et al. YOLOv6: A single-stage object detection framework for industrial applica-

tions[EB/OL]. [2024-10-20]. <https://doi.org/10.48550/arXiv.2209.02976>.

[26] Wang C Y, Bochkovskiy A, Liao H M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver; IEEE, 2023; 7464-7475.

[27] Cai Z W, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City; IEEE, 2018; 6154-6162.

[28] Law H, Deng J. CornerNet: detecting objects as paired keypoints[C]//Computer Vision-2018 European Conference on Computer Vision. Cham; Springer International Publishing, 2018; 765-781.

[29] Duan K W, Bai S, Xie L X, et al. CenterNet: Keypoint triplets for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul; IEEE, 2019; 6568-6577.

[30] Zhu C C, He Y H, Savvides M. Feature selective anchor-free module for single-shot object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach; IEEE, 2019; 840-849.

[31] Zhao Y A, Lv W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle; IEEE, 2024; 16965-16974.



冯梦圆(1999—),女,陕西西安人,西安邮电大学硕士研究生,主要研究方向为基于深度学习的目标检测与识别。
E-mail: fengmengyuankyxd@163.com



房杰(1993—),男,陕西彬州人,博士,西安邮电大学副教授,主要研究方向为计算机视觉、多模态认知计算等。
E-mail: fangjie@xupt.edu.cn



李苑青(1989—),女,河南洛阳人,博士,西安邮电大学讲师,主要研究方向为多模态信息处理、情感计算及强化学习等。E-mail: yuanqingli@xupt.edu.cn



艾达(1973—),男,陕西西安人,博士,西安邮电大学副教授,主要研究方向为数字视频图像处理等。E-mail: aida@xupt.edu.cn

[作者简介]

[责任编辑:蔡秀梅]



王殿伟(1978—),男,吉林松原人,博士,西安邮电大学副教授,主要研究方向为无人机智能态势感知技术。
E-mail: wangdianwei@xupt.edu.cn