

DOI:10.19651/j.cnki.emt.2518410

# 融合动态多尺度轻量化无人机小目标检测算法<sup>\*</sup>

孙冰男<sup>1,2</sup> 于思恺<sup>1,2</sup> 张宇<sup>1,2</sup> 刘俊<sup>1,2</sup> 王军<sup>1,2</sup>

(1. 沈阳化工大学计算机科学与技术学院 沈阳 110142; 2. 辽宁省化工过程工业智能化技术重点实验室 沈阳 110142)

**摘要:** 针对无人机航拍图像中因小目标尺度微小、背景复杂干扰以及多尺度特征融合效率低下等问题,提出轻量化检测模型 YOLO-DAS。通过构建动态多尺度感知卷积模块 DMSConv,增强目标特征捕获能力;设计上下文感知特征重组上采样 ADEPT,优化特征图重建过程以提升上下文信息整合精度;采用双向全局-局部空间注意力 SCOPE 重构颈部网络,通过双向特征交互突破单路径融合局限;增设浅层小目标检测层以强化低层级特征的定位信息提取。实验基于 VisDrone2019 数据集验证,模型在 mAP0.5 和 mAP0.5:0.95 指标上分别达到 39.8% 和 23.7%,较基准 YOLOv8n 分别提升了 8.4% 和 5.1%,精准率与召回率同步提升 8.1% 和 7%,参数量减少 0.49 M,为无人机航拍图像中小目标检测提供有效解决方案。

**关键词:** 无人机; 小目标; YOLO; 轻量级; 多尺度卷积; 注意力机制

**中图分类号:** TN911.73    **文献标识码:** A    **国家标准学科分类代码:** 520.6040

## Lightweight UAV small target detection algorithm based on dynamic multi-scale fusion

Sun Bingnan<sup>1,2</sup> Yu Sikai<sup>1,2</sup> Zhang Yu<sup>1,2</sup> Liu Jun<sup>1,2</sup> Wang Jun<sup>1,2</sup>

(1. School of Computer Science and Technology, Shenyang University of Chemical Technology, Shenyang 110142, China;

2. Key Laboratory of Industrial Intelligent Technology of Chemical Process of Liaoning Province, Shenyang 110142, China)

**Abstract:** A lightweight detection model YOLO-DAS was proposed to solve the problems of small target size, complex background interference and low efficiency of multi-scale feature fusion in UAV aerial images. A dynamic multi-scale sensing convolution module DMSConv is constructed to enhance the feature capture capability. The context-aware feature recombination upsampling ADEPT was designed to optimize the feature map reconstruction process to improve the integration accuracy of context information. The neck network is reconstructed using the bidirectional global-local spatial attention SCOPE, and the single path fusion limitation is broken through the bidirectional feature interaction. A shallow small target detection layer is added to strengthen the localization information extraction of low-level features. Based on the VisDrone2019 dataset, the model achieved 39.8% and 23.7% in mAP0.5 and MAP0.5:0.95 indexes, respectively, which increased by 8.4% and 5.1% compared with the benchmark YOLOv8n. The accuracy and recall rate increased by 8.1% and 7% simultaneously, and the number of parameters decreased by 0.49 M. It provides an effective solution for small and medium-sized target detection in UAV aerial images.

**Keywords:** unmanned aerial vehicle; small goals; YOLO; lightweight; multi-scale convolution; attention mechanism

## 0 引言

目标检测作为计算机视觉领域的核心任务之一,在农业检测<sup>[1]</sup>、智慧医疗<sup>[2]</sup>和遥感影像分析<sup>[3]</sup>等领域发挥着关键作用。基于深度学习的单阶段检测器 YOLO 系列<sup>[4-8]</sup>凭借其高效性成为实时检测主流方案。然而在无人机航拍等复杂场景中,现有模型面临小目标检测精度与速度难以平

衡的严峻挑战<sup>[9-10]</sup>。无人机视角具有动态变化特性,目标尺度差异显著:近景目标可达数百像素,而远景目标常不足 50 pixel,且受复杂背景干扰严重<sup>[11]</sup>。目标空间分布不均衡:如交通场景中的车辆聚集<sup>[12]</sup>,进一步加剧了检测难度,要求检测模型具备更强的多尺度感知和背景抑制能力,而现有算法在此类场景下的性能仍有显著不足。

近年来,研究者们针对小目标检测提出了一系列改进

收稿日期:2025-03-22

\* 基金项目:辽宁省自然科学基金(2024-BSLH-214)、辽宁省教育厅基本科研项目(LJKMZ20220781,LJKMZ20220783)资助

方案。Yang 等<sup>[13]</sup>通过扩展 E-ELAN 结构改进 YOLOv7 的特征融合能力,但浅层特征提取能力不足导致小目标漏检;Zhang 等<sup>[14]</sup>设计的三层 PAFPN 结构虽增强多尺度融合但缺乏动态尺度适应性。在注意力机制<sup>[15]</sup>应用领域,Ning 等<sup>[16]</sup>在 YOLOv8 中引入区域注意力模块,虽降低了计算量,却削弱了微小目标的定位精度;Luo 等<sup>[17]</sup>采用 GAM 注意力机制提升特征融合,但固定尺度的注意力分配难以适应目标动态分布。特征提取改进方面,Xu 等<sup>[18]</sup>提出的 YOLOv8-MPEB 模型虽提升检测能力,其固定尺度金字塔难以应对动态变化;Chang 等<sup>[19]</sup>用 SPD-Convolution 避免池化特征损失,但对背景噪声抑制不足。此外,Qi 等<sup>[20]</sup>设计的 CA-Trans 模块通过稀疏局部注意力优化多尺度融合,但计算复杂度较高;Zhuo 等<sup>[21]</sup>的 DAMP-YOLO 模型利用可变形卷积提升定位精度,却牺牲了实时性。这些方法普遍存在多尺度融合适应性不足、上采样细节丢失以及注意力机制失效等问题,在密集小目标场景下的泛化能力仍有显著缺陷<sup>[22-23]</sup>。

针对上述挑战,本文以 YOLOv8n 为基础框架,提出一种融合动态多尺度感知的轻量化模型(YOLO dynamic multiscale target detection model, YOLO-DAS)。主要创新包括:

1)设计动态多尺度感知卷积模块(dynamic multi-scale sensing convolution module, DMSConv),通过自适应卷积核调整与跨尺度特征交互增强细节捕捉能力。

2)构建特征重组上采样(adaptive detail-enhanced pixel transform, ADEPT),利用局部-全局上下文信息优化特征分辨率缓解细节丢失问题。

3)采用双向全局-局部空间注意力(selective content optimization with pyramid enhancement, SCOPE),重构颈部网络以抑制背景噪声的同时强化目标区域的上下文关联。

4)增设浅层小目标检测层,使网络在不同尺度上获得更丰富的语义信息,通过对小目标进行特定尺度的处理,增强网络对小目标的感知能力。

## 1 YOLOv8 网络结构

YOLOv8 是由 Wang 等<sup>[7]</sup>于 2023 年 1 月提出的一种实时目标检测算法。YOLOv8 系列根据模型复杂度分为 n、s、m、l、x 五个版本,本研究针对无人机场景选用最轻量的 YOLOv8n 进行优化。其网络结构由输入端(Input)、骨干网络(Backbone)、颈部网络(Neck)以及头部网络(Head)组成,如图 1 所示。

## 2 YOLO-DAS 算法设计

本文提出的轻量化小目标检测模型 YOLO-DAS,旨在解决无人机航拍场景下小目标检测中存在的尺度敏感性高、复杂背景干扰强以及多尺度特征融合不足等核心问题。

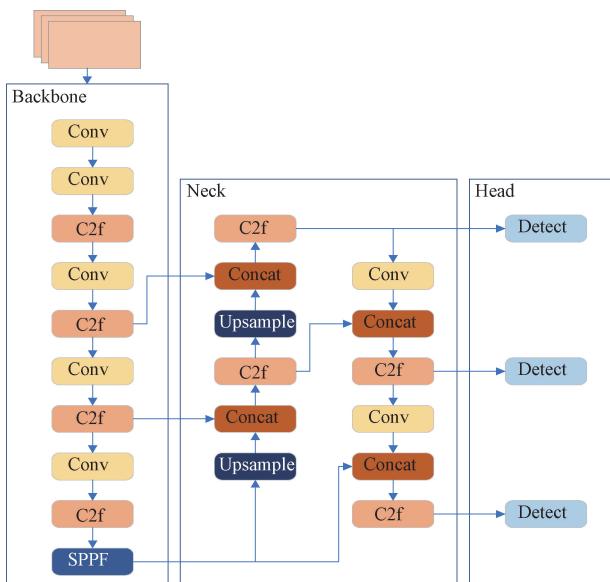


图 1 YOLOv8 结构

Fig. 1 The structure of YOLOv8

YOLO-DAS 模型结构如图 2 所示,通过引入以下改进模块,有效提升了复杂场景下的检测性能。

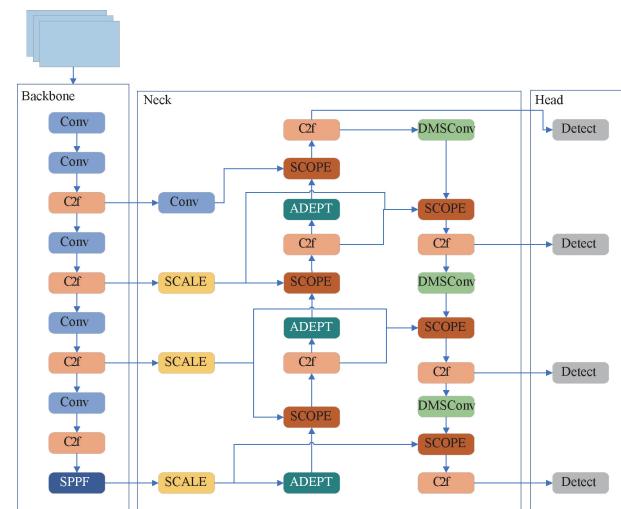


图 2 YOLO-DAS 模型结构

Fig. 2 Model structure of YOLO-DAS

### 2.1 DMSConv 卷积模块

针对 YOLOv8 骨干网络在无人机航拍场景中因重复卷积导致小目标细粒度信息丢失的问题,本文提出轻量化动态多尺度感知卷积模块 DMSConv(dynamic multi-scale sensing convolution module),如图 3 所示。

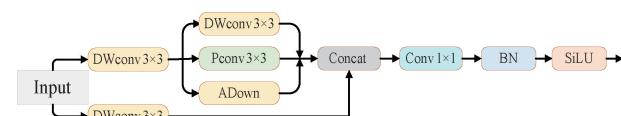


图 3 DMSConv 结构

Fig. 3 Structure of DMSConv

首先,输入特征经过一个 $3 \times 3$ 卷积核的深度可分离卷积<sup>[24]</sup>(DWConv)进行初步的特征提取,以捕捉输入特征的空间维度信息。接着,这些初步提取的特征会分别送入3个并行处理模块,分别是采用 $3 \times 3$ 卷积核的深度可分离卷积(DWConv)、 $3 \times 3$ 卷积核的普通卷积<sup>[25]</sup>(PConv)以及Adown卷积<sup>[26]</sup>,以进一步深入提取特征。PConv卷积能够在保持常规卷积有效提取空间特征能力的同时,显著减少冗余计算,从而提升效率。而ADown模块则通过精简参数量,有效降低模型的复杂度。同时,DMSConv也融入了残差结构,其中残差分支的输入特征会先经过一个 $5 \times 5$ 卷积核的深度可分离卷积(DWConv)进行进一步的特征提取。之后将这些特征融合在一起,经过一个 $1 \times 1$ 卷积核的标准卷积来调整输出通道数,最后对输出进行标准化处理(batch normalization)操作后通过激活函数,综合考虑精度与速度,选择SiLu作为激活函数。

DMSConv通过 $3 \times 3$ 和 $5 \times 5$ 的卷积核协同工作形成多尺度感知机制:前者能够精准捕捉图像的细节特征,而后者则更适用于捕获图像中广泛分布的上下文信息,有效增强模型的特征表达能力。同时,残差结构的设计有效解决了深度神经网络中常见的梯度消失问题,并允许信息更容易地在网络中流动。

## 2.2 上采样方式改进

ADEPT(adaptive detail-enhanced pixel transform)模块是一种基于内容感知的特征重组上采样架构,通过双分支协同工作机制实现动态特征重建。如图4所示,其工作流程包含4个核心阶段:

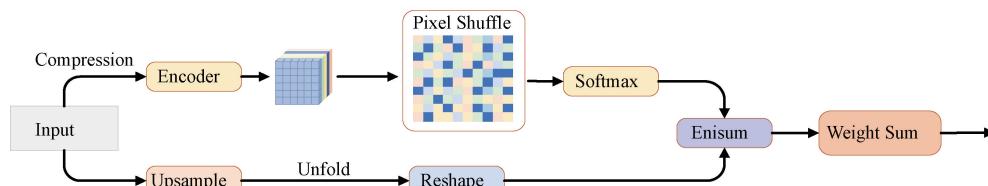


图4 ADEPT 结构

Fig. 4 Structure of ADEPT

## 2.3 使用 SCOPE 网络重构颈部网络架构

SCOPE(selective content optimization with pyramid enhancement)网络采用双向多尺度特征融合架构,如图5所示。其创新性设计包含以下3部分:

### 1) 特征金字塔流程

以P2~P5多尺度特征为输入,通过自顶向下与自底向上路径实现交叉融合。自顶向下路径中,P5经SCALE模块增强后上采样与P4在Fusion模块融合,逐级传递至P3;自底向上路径则将P3融合特征下采样与P4二次融合,进而再次下采样与P5结合,各阶段均通过C2f模块优化特征。最终,经过双向交叉融合强化的多尺度特征(P3、P4、P5)被输入Detect模块完成目标检测。

### 2) SCALE模块结构

SCALE(spatial context-aware local-global embedding)模

块将输入特征分为两部分。全局分支(上半部分)通过 $1 \times 1$ 卷积与Transpose、Softmax操作提取上下文信息,使用IN-ReLU(实例归一化+ReLU)激活,再经Contact模块捕获全局上下文信息。局部分支(下半部分)使用DWConv进行高效特征提取,并通过Sigmoid激活函数控制特征重要性。双路输出融合后生成兼具全局语义与局部精度的增强特征。

### 3) 自适应特征融合(Fusion)模块

该模块采用可学习的动态加权机制处理多源特征输入。对于N个输入特征(Feature 1, Feature 2, ..., Feature N),使用Weight模块为每个输入特征分配权重,经ReLU和归一化处理后执行加权融合,促进跨层级语义信息高效流动。

SCOPE网络采用双向路径实现不同尺度的特征充分

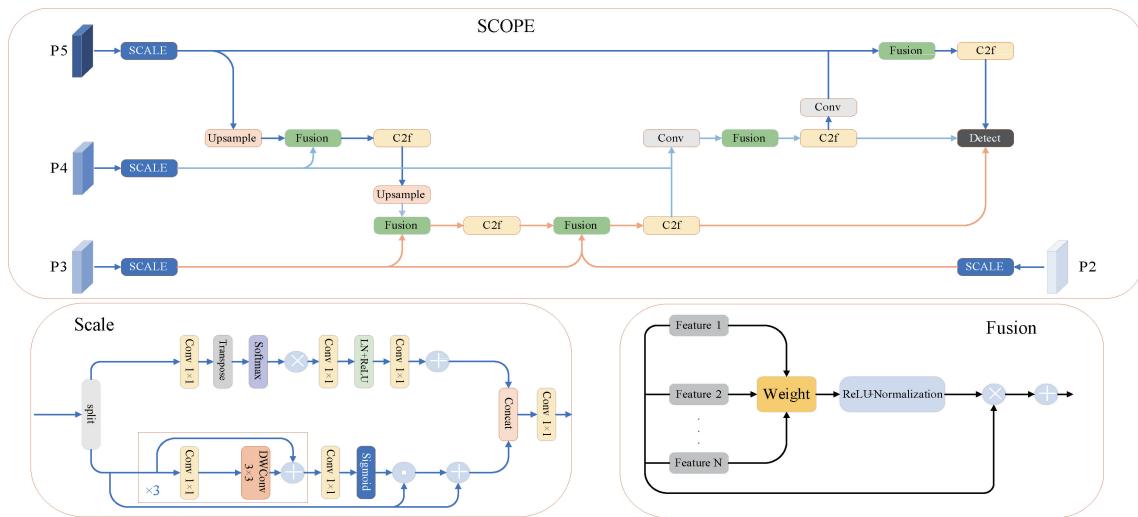


图 5 SCOPE 网络结构图

Fig. 5 SCOPE network structure diagram

融合, SCALE 模块通过并行全局上下文与局部细节, 增强特征表示能力。Fusion 模块通过学习的权重自适应地融合不同来源的特征, 使语义信息可以在整个金字塔网络中自由流动, 实现最优特征组合。这种设计使 SCOPE 网络能够有效处理不同尺度的目标检测, 特别适合处理复杂场景中的多尺度目标识别任务。

#### 2.4 小目标检测层

在 YOLOv8 中, 特征融合网络通过将不同尺度的特征图进行结合来提升目标检测的准确性和鲁棒性。原始的 YOLOv8 算法在对  $640 \times 640$  分辨率的输入图像进行 8 倍、16 倍和 32 倍下采样时, 只能生成  $80 \times 80$ 、 $40 \times 40$  和  $20 \times 20$  大小的预测特征图, 用于检测  $8 \times 8$ 、 $16 \times 16$ 、 $32 \times 32$  以上的目标。然而, 在无人机航拍图像中, 目标的尺寸普遍小于原尺寸的 5%, 并且由于 YOLOv8 模型采用了较多的下采样操作, 随着下采样过程的不断进行, 目标的尺度逐渐减小, 导致小目标的特征信息在此过程中逐渐丢失。

因此, 本文在保持其他特征图尺度不变的情况下, 增设了一个  $160 \times 160$  的小目标检测层。该层的高分辨率预测特征图能够更充分地保留小目标的位置信息及丰富的特征信息, 可以从更浅层的网络获得小目标的类别和位置信息。

### 3 实验结果与分析

#### 3.1 数据集

本文基于 VisDrone2019 数据集<sup>[27]</sup>验证算法有效性, 该数据集由天津大学 AISKY EYE 团队构建, 包含行人、自行车等 10 类目标, 共 10 209 幅静态图像, 划分为训练集(6 471 张)、验证集(578 张)、和测试集(3 190 张)。与其他数据集相比, VisDrone2019 数据集具有小目标数量更多、尺度差异显著等特点, 精准匹配无人机航拍场景的检测需求。

#### 3.2 实验环境及参数配置

本研究在训练阶段选用了当前主流的先进模型 YOLOv8n 作为基准模型, 并基于 Ubuntu 20.04 操作系统进行, 实验参数的批量大小为 8, 训练轮数为 200, 学习率为 0.01, 权重衰减系数为 0.000 5, 输入图片尺寸为  $640 \times 640$ , 优化器采用 SGD 优化器, 深度学习框架使用 Python 3.8、PyTorch 2.0.0 和 Cuda 11.8 的环境, 在 NVIDIA GeForce RTX 3090 显卡上进行训练。

#### 3.3 评价指标

为了更好的评估模型的性能并与基线模型进行对比, 本文选取精准率 P(Precision)、召回率 R(Recall)、平均精度均值(mAP)和模型参数量(Params)作为模型性能的评价指标。

1) 精准率 P(Precision): 表示正确预测为正样本的数量占所有预测为正样本数量的比例。公式如下:

$$P = \frac{TP}{TP + FP} \quad (1)$$

2) 召回率 R(recall): 表示正确预测为正样本的数量占所有实际为正样本数量的比例。公式如下:

$$R = \frac{TP}{TP + FN} \quad (2)$$

3) 平均精度均值(mAP): mAP<sub>0.5</sub> 和 mAP<sub>0.5~0.95</sub> 衡量的分别是当 IoU 阈值为 0.5 时和 0.5~0.95 范围内时模型的平均精度。公式如下:

$$mAP = \frac{1}{N} \sum_{i=1}^N AP(i) \quad (3)$$

$$AP = \int_0^1 P(r) dr \quad (4)$$

4) 模型参数量(Params): 通常用于评估模型的复杂度。参数量越大, 模型可能越复杂, 拟合能力越强, 但也可能导致过拟合。相反, 参数量较少可能会削弱模型的表达

能力,使其难以有效捕捉数据中的复杂关联性。

### 3.4 消融实验

为验证 YOLO-DAS 模型的无人机目标检测性能,在

固定实验环境与参数配置下,基于 YOLOv8n 基准模型于 VisDrone2019 数据集开展模块消融实验,结果如表 1 所示。

表 1 基于 VisDrone2019 消融实验结果

Table 1 Based on the ablation results of VisDrone2019

模型	P/%	R/%	mAP <sub>50</sub> /%	mAP <sub>50,95</sub> /%	Params/M
YOLOv8n	41.6	31.7	31.4	18.6	3.01
YOLOv8n+DMSConv	43.1	33.6	32.7	19.0	2.99
YOLOv8n+DMSConv+SCOPE	43.3	33.3	33.4	19.5	2.13
YOLOv8n+DMSConv+SCOPE+ADEPT	44.3	33.6	34.0	20.0	2.26
YOLOv8n+DMSConv+SCOPE+ADEPT+layer	49.7	38.7	39.8	23.7	2.52

从表 1 的消融实验结果中可见,使用 DMSConv 卷积模块之后,因其分层特征融合机制通过差异化卷积核增强了小目标的局部纹理与全局语义表征能力,使精准率 P、召回率 R、平均精度均值(mAP<sub>0.5</sub>、mAP<sub>0.5:0.95</sub>)分别提高了 1.5%、1.9%、1.3%、0.4%;添加 SCOPE 模块,建立了双向特征交互路径,全局分支通过空间注意力聚焦目标显著性区域,局部分支利用空洞卷积强化边界特征响应,二者的协同作用在降低误检率的同时减少 0.88 M 参数量,精准率 P 提升了 1.7%,mAP<sub>0.5</sub> 值提升了 2%,在提升检测精度前提下实现模型轻量化,证明 SCOPE 模块的改进有效;添加 ADEPT 算法,通过通道重校准策略动态分配特征权重,结合可变形卷积核自适应调整感受野分布,优化边界特征重建精度,使参数量下降的同时精准率 P、召回率 R、mAP<sub>0.5</sub>、mAP<sub>0.5:0.95</sub> 分别提高了 1%、0.3%、

0.6%、0.5%;最后引入一个小目标检测层,直接利用高分辨率特征图的细粒度空间信息,在仅增加 0.26 M 参数量的情况下使精准率 P 提高了 5.4%,mAP<sub>0.5</sub> 值提高了 5.8%,有效地获取了小目标的位置以及类别信息。

相较于 YOLOv8n,改进算法参数量下降了 0.49 M 的同时精准率 P、召回率 R、mAP<sub>0.5</sub>、mAP<sub>0.5:0.95</sub> 分别提升了 8.1%、7%、8.4%、5.1%,证明本文所提出的方法不仅有效提升了检测精度,还实现了模型的轻量化。

从表 2 中与 YOLOv8n 对比可见,改进后的模型在所有检测类别上的精度均有大幅提高,尤其是 Motor 类、Pedestrian 类和 People 类分别提升了 14.1%、13.6% 和 12.5%;此外,尽管 Bicycle 类和 Tricycle 类因特征信息有限而检测难度大,但它们的检测精度仍分别提升了 6.3% 和 6.4%。

表 2 VisDrone2019 各类别 mAP50 详情

Table 2 Comparison of VisDrone2019's accuracy by category

模型	Pedestrian	People	Bicycle	Car	Van	Truck	Tricycle	Awning-tricycle	Bus	Motor
YOLOv8n	33.1	26.0	7.23	74.2	37.5	26.6	19.5	11.6	45.0	33.3
YOLO-DAS	46.7	38.5	13.5	81.9	45.9	30.3	25.9	13.7	52.6	47.4

### 3.5 对比实验

为了进一步验证 YOLO-DAS 模型检测无人机航拍图像的有效性,将其与 YOLOv5、Drone-YOLO、YOLOv8n、LW-YOLOv8<sup>[28]</sup>、YOLOv10n 和文献[29]等主流目标检测算法进行对比实验,实验结果如表 3 所示。

从表 3 可以看出,本文提出的 YOLO-DAS 模型与 YOLOv6、Drone-YOLO 和 LW-YOLOv8 算法相比,在拥有更低参数量的同时,mAP<sub>0.5</sub> 分别提升了 12.1%、8.8%、3.5%。虽然 YOLOv5 和 YOLOv10n 算法参数量较少,但检测性能各项指标不及所提算法。而相比于基准算法 YOLOv8n,本文所提模型 YOLO-DAS 在保持参数量降低的同时,mAP<sub>0.5</sub> 提升了 8.4%。

图 6 通过 PR 曲线与 mAP 指标对比主流算法与 YOLO-DAS 模型的检测性能。分析曲线覆盖面积表明,

该模型在各类别检测精度均显著提升,验证了改进模型的有效性及复杂场景下的综合优势。对比实验显示,改进算法在保持参数量优化的同时,显著提高平均检测精度,充分体现其在无人机目标检测任务中的优越性。

### 3.6 结果可视化

为评估所提算法在真实场景中的性能,本研究从 VisDrone2019 测试集中筛选出低光照、高密度及高空视角等典型场景的样本图像进行验证。这些样本中的待检测目标均呈现小尺度特征,图 7 展示了相应检测结果的可视化效果。

图 7(a)展示了低光照条件下的检测效果,该样本存在照度不足导致的视觉特征弱化问题,尽管目标细节呈现模糊化特征,本文提出的 YOLO-DAS 仍能有效识别图像中全部待检目标。图 7(b)对应高密度遮挡场景的检测结果,

表 3 不同算法在 VisDrone2019 数据集上的目标检测结果对比

Table 3 Comparison of target detection results of different algorithms on VisDrone2019 dataset

模型	P/ %	R/ %	mAP <sub>50</sub> / %	mAP <sub>50,95</sub> / %	Params/ M
YOLOv3-tiny	36.8	23.3	22.3	12.4	12.1
Faster-RCNN	—	—	35.2	21.3	42.36
YOLOv5	39.3	31.0	29.7	17.5	2.50
Drone-YOLO	—	—	31.0	17.5	3.05
YOLOv6	35.6	28.9	27.7	16.5	4.23
YOLOv8n	41.6	31.7	31.4	18.6	3.01
文献[29]	46.9	38.0	36.4	20.1	—
YOLOv9s	<b>50.9</b>	37.3	38.7	23.4	7.17
SSD	—	—	21.3	13.1	26.13
LW-YOLOv8	47.6	35.5	36.3	—	2.83
YOLOv10n	44.0	33.0	33.3	19.2	<b>2.26</b>
YOLO-DAS	49.7	<b>38.7</b>	<b>39.8</b>	<b>23.7</b>	2.52

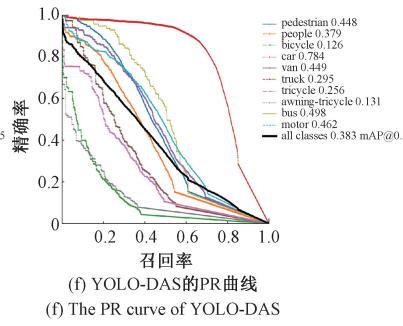
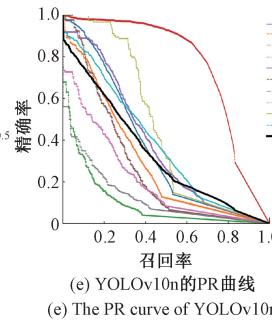
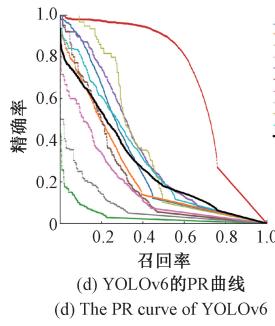
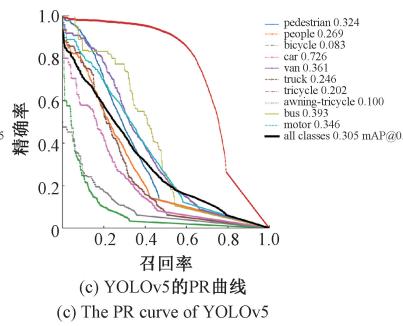
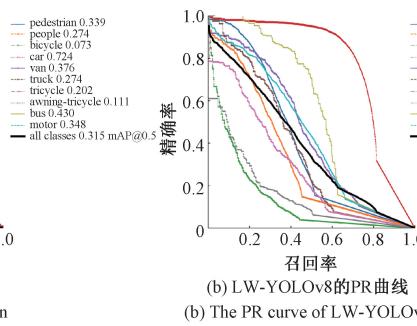
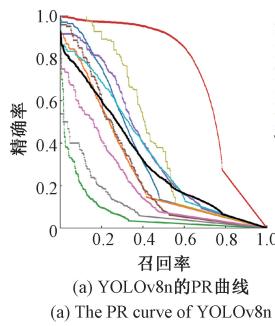


图 6 主流先进模型和 YOLO-DAS 在 VisDrone2019 数据集上的 PR 曲线

Fig. 6 The PR curves of mainstream advanced models and YOLO-DAS on the VisDrone2019 dataset

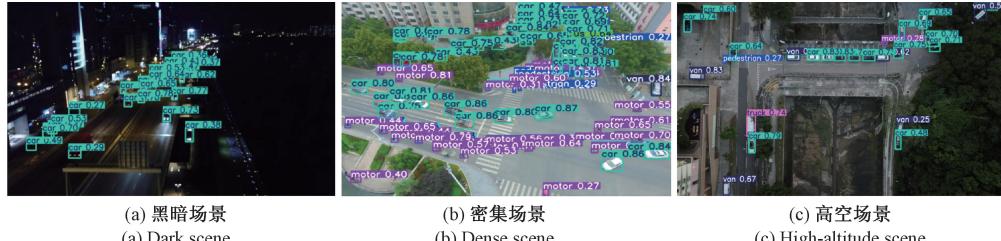


图 7 本文提出模型 YOLO-DAS 在不同场景下的检测效果图

Fig. 7 This paper presents the detection effect of YOLO-DAS model in different scenarios

面对高密度车流导致的严重目标遮挡现象,本文模型成功实现了汽车与行人的准确检测。图 7(c)呈现高空航拍视角的检测效果,在远距离拍摄造成的目标微小化挑战下(目标像素占比均小于  $32 \times 32$ ),YOLO-DAS 仍能准确捕捉货车、摩托车及行人等多尺度目标。

YOLOv8n 和本文提出的 YOLO-DAS 模型的检测结果对比如图 8 所示,图 8(a)为基准算法 YOLOv8n 检测效果图,图 8(b)为本文模型的检测效果图。为直观呈现算法改进前后的性能差异,实验采用矩形框对检测差异区域进行标注。

从第 1 组图可以看出,在目标尺度小且相互遮挡场景中,基线模型对密集分布的自行车和行人目标存在明显漏检,而本文提出的 YOLO-DAS 模型能够有效识别多数目标。尽管在极端背景干扰和重度遮挡情况下仍出现个别行人漏检案例,但相较于基线模型,本文模型在漏检率和误检率指标上均取得显著优化。第 2 组目标重叠场景的检测对比表明,基线模型因特征混淆导致部分目标未被识别,而本文模型通过增强特征判别能力实现了完整目标框

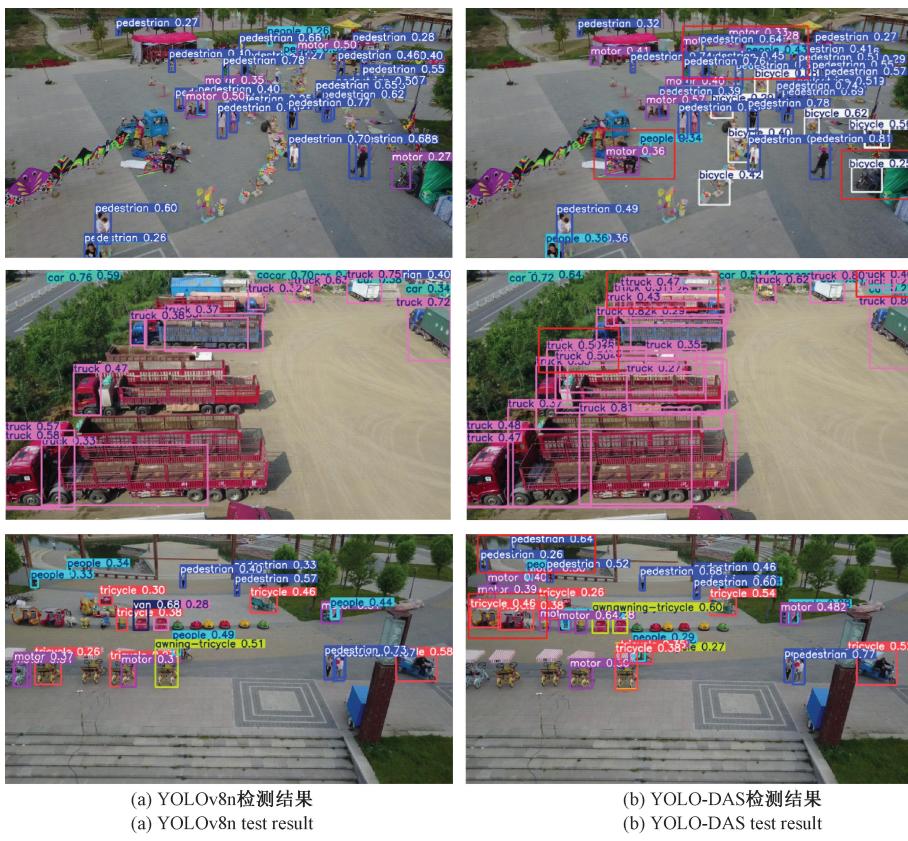


图 8 改进前后检测效果对比图

Fig. 8 Comparison of detection results before and after improvement

的连续检测。第 3 组植被遮挡与小目标复合场景的实验进一步验证,本文模型对树丛间隙等复杂遮挡下的行人目标检测精度显著优于基线模型。

可视化对比证实,本文提出的模型在保持实时检测速度的同时,有效缓解了小目标特征丢失和遮挡干扰问题,为无人机航拍场景提供了更可靠的检测方案。

## 4 结 论

针对无人机航拍小目标检测中特征弱、上下文缺失与多尺度融合不足等问题,本研究提出轻量化模型 YOLO-DDE。通过动态多尺度卷积模块(DMSConv)增强细粒度特征提取,结合上采样机制(ADEPT)优化目标定位,采用双向全局-局部注意力(SCOPE)与浅层检测策略提升多尺度特征融合效率。该模型在保持轻量化的同时显著提升了小目标的检测精度,为无人机实时检测提供了实用解决方案。

## 参考文献

- [1] LI T, ZHANG L, LIN J. Precision agriculture with YOLO-Leaf: Advanced methods for detecting apple leaf diseases[J]. Frontiers in Plant Science, 2024, 15: 1452502.
- [2] MEHTA P, VAGHELA R, PANSURIYA N, et al. Benchmarking YOLO variants for enhanced blood cell detection[J]. International Journal of Imaging Systems and Technology, 2025, 35(1): e70037.
- [3] SILPALATHA G, JAYADEVA T. Accelerating fast and accurate instantaneous segmentation with YOLO-v8 for remote sensing image analysis [J]. Remote Sensing Applications: Society and Environment, 2025, 37:101502.
- [4] VIJAYAKUMAR A, VAIRAVASUNDARAM S. YOLO-based object detection models: A review and its applications [J]. Multimedia Tools and Applications, 2024, 83(35): 83535-83574.
- [5] ANGUCHAMY K K, PALANISAMY V. Real-time object detection using improvised YOLOv4 and feature mapping technique for autonomous driving[J]. Expert Systems with Applications, 2025, 280:127452.
- [6] XI AI G, MEI F M, LI H Q, et al. An improved YOLOv5 model for aeolian saltating particle recognition in high-speed videos[J]. Measurement, 2025, 253(Part A): 117321.
- [7] WANG G, CHEN Y, AN P, et al. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J].

- Sensors, 2023, 23(16): 7190.
- [8] ZHANG Q, WANG X, SHI H, et al. BRA-YOLOv10: UAV small target detection based on YOLOv10[J]. Drones, 2025, 9(3): 159.
- [9] GAO P, LI Z. YOLO-S3DT: A small target detection model for UAV images based on YOLOv8 [J]. Computers, Materials & Continua, 2025, 82 (3): 4555-4572.
- [10] WANG S, LIU Y, WANG X, et al. An improved YOLO algorithm for UAV detection in formation flight[J]. Signal, Image and Video Processing, 2025, 19(3): 195.
- [11] JI J, ZHAO Y, LI A, et al. Dense small object detection algorithm for unmanned aerial vehicle remote sensing images in complex backgrounds[J]. Digital Signal Processing, 2025, 158:104938.
- [12] WU Y, ZHANG T, NIU J, et al. YOLO-based lightweight traffic sign detection algorithm and mobile deployment [J]. Optoelectronics Letters, 2025, 21(4): 249-256.
- [13] YANG M, TONG X, CHEN H. Detection of small lesions on grape leaves based on improved YOLOv7 [J]. Electronics, 2024, 13(2): 464.
- [14] ZHANG Z. Drone-YOLO: An efficient neural network method for target detection in drone images[J]. Drones, 2023, 7(8): 526.
- [15] SOYDANER D. Attention mechanism in neural networks: where it comes and where it goes [J]. Neural Computing and Applications, 2022, 34 (16): 13371-13385.
- [16] NING T, WU W, ZHANG J. Small object detection based on YOLOv8 in UAV perspective[J]. Pattern Analysis and Applications, 2024, 27(3): 103.
- [17] LUO H, LIU W, XU P, et al. Recognition algorithm for laboratory protective equipment based on improved YOLOv7[J]. Heliyon, 2024, 10(16): e36264.
- [18] XU W, CUI C, JI Y, et al. YOLOv8-MPEB small target detection algorithm based on UAV images[J]. Heliyon, 2024, 10(8): e29501.
- [19] CHANG Y, LI D, GAO Y, et al. An improved YOLO model for UAV fuzzy small target image detection[J]. Applied Sciences, 2023, 13(9): 5409.
- [20] QI Z, BINGHAO L, SHUCHANG L, et al. TPH-YOLOv5++: Boosting object detection on drone-captured scenarios with cross-layer asymmetric transformer [J]. Remote Sensing, 2023, 15(6): 1687.
- [21] ZHUO S CH, ZHANG X M, CHEN Z Y, et al. DAMP-YOLO: A lightweight network based on deformable features and aggregation for meter reading recognition[J]. Applied Sciences, 2023, 13(20): 11493.
- [22] LUO X D, WU Y Q, WANG F Y. Target detection method of UAV aerial imagery based on improved YOLOv5[J]. Remote Sensing, 2022, 14(19): 5063.
- [23] MA P, HE X, CHEN Y, et al. ISOD: Improved small object detection based on extended scale feature pyramid network[J]. The Visual Computer, 2024, 41(1): 1-15.
- [24] WANG Z J, HE X W, LI Y, et al. EmbedFormer: Embedded depth-wise convolution layer for token mixing[J]. Sensors, 2022, 22(24): 9854.
- [25] PARK S, YEO Y J, SHIN Y G. PConv: Simple yet effective convolutional layer for generative adversarial network[J]. Neural Computing and Applications, 2022, 34(9): 7113-7124.
- [26] ZHANG G, CAO H, JIN Y, et al. YOLOv8n-DDA-SAM: Accurate cutting-point estimation for robotic cherry-tomato harvesting [J]. Agriculture, 2024, 14(7): 1011.
- [27] HE ZH Q, CAO L J. SOD-YOLO: Small object detection network for UAV aerial images[J]. IEEJ Transactions on Electrical and Electronic Engineering, 2024, 20(3): 431-439.
- [28] HUANGFU ZH M, LI SH Q. Lightweight you only look once v8: An upgraded you only look once v8 algorithm for small object identification in unmanned aerial vehicle images [J]. Applied Sciences, 2023, 13(22): 12369-12382.
- [29] SHANG J CH, WANG J S, LIU SH B, et al. Small target detection algorithm for UAV aerial photography based on improved YOLOv5s[J]. Electronics, 2023, 12(11): 2434.

## 作者简介

孙冰男,硕士,主要研究方向为图形图像处理、目标检测。

E-mail:Z2023488@stu.sjtu.edu.cn

于思恺,硕士,主要研究方向为人工智能。

E-mail:2205238232@qq.com

张宇,博士,讲师,主要研究方向为深度学习与计算机视觉。

E-mail:zytriumph@163.com

刘俊,学士,副教授,主要研究方向为工业互联网。

E-mail:computer\_lj@163.com

王军(通信作者),博士,教授,硕士研究生导师,主要研究方向为人工智能技术及智能制造、工业大数据技术。

E-mail:wj\_software@hotmail.com