

基于改进 YOLOv11 的无人机航拍小目标车辆检测算法

丁美云¹ 杜杰¹ 高扬² 丁四景^{3*}
DING Meiyun DU Jie GAO Yang DING Sijing

摘要

针对航拍小交通目标分布密集、检测精度低、漏检和错检的问题,文章提出一种改进的 YOLOv11 的航拍小交通目标检测算法。在骨干网络中引入风车状卷积 PConv,使其对齐小目标特征图像的高斯空间分布,抑制杂波信号;在颈部网络中引入 EVC 模块,能够同时捕获全局长依赖和保留角落区域信息,提高密级预测任务检测性能;在头部网络中引入分离和增强注意力模块 SEAM,建立多尺度并行子网络、避免维度缩减、融合跨空间信息,增强网络对关键特征的敏感度。在 VisDrone2019 数据集中选择汽车、面包车、卡车、公共汽车和摩托车 5 类车辆目标,相比于 YOLOv11m,改进的 YOLOv11m 算法的准确率增加 3.32%,平均检测精度 mAP@0.5/% 和 mAP@0.5-0.95/% 分别增加了 3.73%、2.46%。

关键词

航拍图像;小目标检测;YOLOv11;风车状卷积

doi: 10.3969/j.issn.1672-9528.2026.06.004

0 引言

近年来,随着无人机技术的发展,高空无人机航拍也得到越来越多的关注与研究^[1]。在智能交通领域,利用图像识别技术进行车辆交通目标检测,从而为自动驾驶提供支持。然而,无人机航拍高度较高、车辆等目标尺度较小、高空拍摄有建筑和树木等遮挡,导致航拍小目标车辆的检测能力较低^[2-3]。

随着机器视觉和深度学习算法技术的发展,基于深度学习的车辆目标检测也逐渐应用于智慧交通领域。主流深度学习算法主要包括两阶目标检测算法和一阶目标检测算法。两阶检测算法有 R-CNN、Fast R-CNN、Mask R-CNN^[4-6] 等。两阶检测算法需要两次特征提取,检测速度低,不能满足快速动态目标检测的需求。单阶段算法有 YOLO (you only look once)^[7] 和 SSD^[8] (single shot MultiBox detector)、RetinaNet^[9]。

小目标图像检测,图像像素低和可以利用的信息少,因而识别难度大。Yi 等^[10]改进 YOLOv7 模型,通过融合注意力、增强 FPN 拓扑的方式,增强关注小目标特征信息,然而过于复杂的 FPN 网络,可导致训练数据过拟合。奉志强等^[11]基于 YOLOv5,提出空间-通道注意力特征融合模块,该模块提高对中小目标聚集区域的关注度,并根据特征图尺度

大小自动调整注意力权重,但多批次调用跨步卷积层进行下采样,会丢失目标细粒度特征。Dong 等^[12]提出一种轻量化 YOLOv5 模型,通过引入注意力机制 CBAM,并改进损失函数 CIoU,进而提高检测效率和检测精度。何湘杰等^[13]提出一种改进的轻量级的 YOLOv4 算法,提出并使用 CSATT 注意力机制,提高算法对空间目标的识别能力。何植仟等^[14]基于 YOLOv8,提出适用于无人机航拍图像的目标检测模型,引入可变形卷积网络,扩大特征提取时的感受野,提高模型对小目标的检测效果。尽管众多学者对小目标检测投入了大量研究,但由于航拍图像的场景复杂、目标数量多、目标尺度变化大等原因,其检测精度仍然较低。

1 改进 YOLOv11 算法

YOLOv11 是由 Ultralytics 公司于 2024 年 9 月 30 日发布的目标检测模型^[15],YOLOv11 模型由输入(Input)、骨干网络(Backbone)、颈部网络(Neck)、检测头(Head) 4 个部分构成。YOLOv11 在骨干网络中引入 C3k2 模块代替 C2f 优化浅层特征提取、结合 C2PSA 注意力机制提升特征捕获能力,在检测头中加入深度可分离卷积(DWConv),在降低模型参数和开销的同时提升推理效率。

本文在 YOLOv11 骨干网络中添加 PConv 卷积,在颈部网络中增加 EVC,在头部网络中增加 SEAM 注意力机制对航拍小交通目标进行检测。如图 1 所示。

(1) 在骨干网络引入风车卷积 PConv。PConv 能够对齐小目标特征图像的高斯空间分布,相比标准卷积,能够更加有效地提取小目标的底层特征,增强了目标与背景对比,抑

1. 青岛新诚志卓轨道交通装备股份有限公司 山东青岛 266000
2. 青岛科林铁路设备有限公司 山东青岛 266000
3. 四川大学青岛研究院 山东青岛 266000

制了杂波信号。

(2) 在颈部网络中引入 EVC 模块, 通过轻量级 MLP 和可学习视觉中心机制 LVC 能够同时捕获全局长依赖和保留角落区域信息, 提高密集预测任务检测性能。

(3) 在头部网络部分引入分离和增强注意力模块 SEAM, 能够建立多尺度并行子网络、避免维度缩减、融合跨空间信息, 增强网络对关键特征的敏感度。

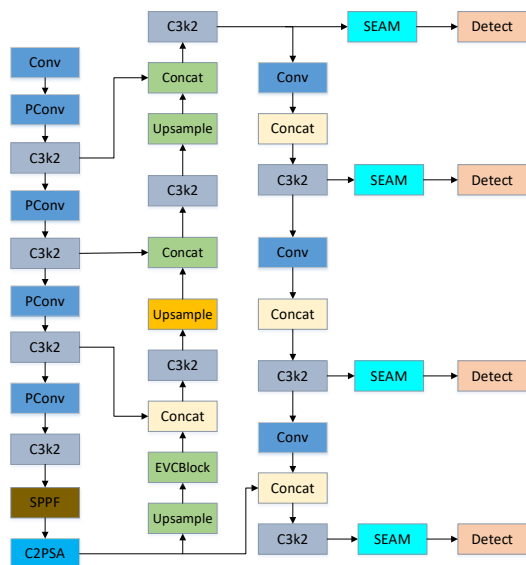


图 1 改进 YOLOv11 网络结构图

1.1 风车卷积 PConv

传统神经网络的检测方法往往使用标准卷积, 忽略了小目标像素的空间高斯分布特性。风车卷积 (pinwheel-shaped convolution, PConv) [16] 模块的结构与传统卷积层有所不同。该模块通过非对称填充机制, 为图像的不同区域分别生成水平与垂直方向的卷积核。这些卷积核的感受野向外扩散, 其中 h_1 、 w_1 、 i_1 分别表示输入张量的高度、宽度和通道数。PConv 每个卷积层之后应用批归一化和 SiLU 激活函数来提升训练稳定性和速度。PConv 的首层通过以下方式实现并行卷积操作如图 2 所示。

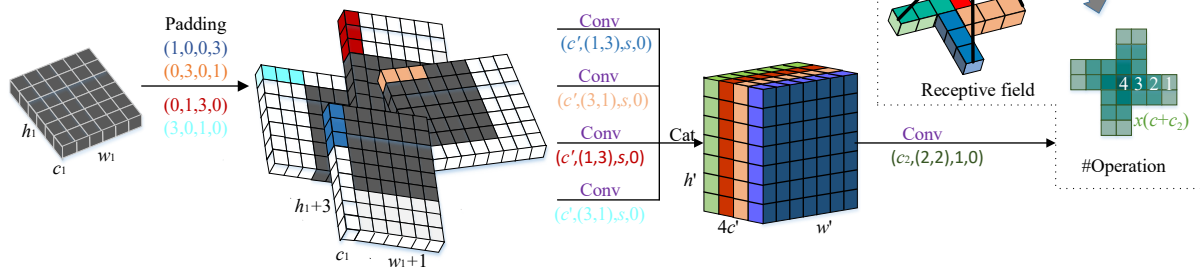


图 2 PConv 结构图

$$\begin{aligned} X_1^{(h',w',c')} &= \text{SiLU}(\text{BN}(X_{P(1,0,0,3)}^{(h_1,w_1,c_1)} \otimes W_1^{(1,3,c')})) \\ X_2^{(h',w',c')} &= \text{SiLU}(\text{BN}(X_{P(0,3,0,1)}^{(h_1,w_1,c_1)} \otimes W_2^{(3,1,c')})) \\ X_3^{(h',w',c')} &= \text{SiLU}(\text{BN}(X_{P(0,1,3,0)}^{(h_1,w_1,c_1)} \otimes W_3^{(3,1,c')})) \\ X_4^{(h',w',c')} &= \text{SiLU}(\text{BN}(X_{P(3,0,1,0)}^{(h_1,w_1,c_1)} \otimes W_4^{(3,1,c')})) \end{aligned} \quad (1)$$

式中: $\otimes$ 表示卷积运算符, $W_1^{(1,3,c')}$ 表示 1×3 、输出通道数为 c' 的卷积核, $P(1,0,0,3)$ 分别表示左、右、上、下四个方向的像素填充量。

经过第一层交错卷积后, 输出特征图的高度 (h')、宽度 (w') 和通道数 (c') 与输入特征图的关系用公式表示为:

$$h' = \frac{h_1}{s} + 1, w' = \frac{w_1}{s} + 1, c' = \frac{c_2}{4} \quad (2)$$

式中: c_2 表示 PConv 模块最终输出特征图的通道数, s 表示卷积步长。

第一层交错卷积的结果将进行拼接 (Cat(.)) 操作, 其输出计算结果为:

$$X^{(h',w',4c')} = \text{Cat}(X_1^{(h',w',c')}, \dots, X_4^{(h',w',c')}) \quad (3)$$

通过一个无填充的卷积核 $W^{(2,2,c_2)}$ 对拼接后的张量进行归一化处理。输出特征图的高度和宽度被调整至预设值 h_2 和 w_2 , 这使得 PConv 层可与标准卷积层互换使用, 并作为通道注意力机制来分析不同卷积方向的贡献。最终输出 $Y^{(h_2,w_2,c_2)}$ 的计算公式为:

$$h_2 = h' - 1 = \frac{h_1}{s}, w_2 = w' - 1 = \frac{w_1}{s} \quad (4)$$

$$Y^{(h_2,w_2,c_2)} = \text{SiLU}(\text{BN}(X^{(h',w',4c')} \otimes W^{(2,2,c_2)})) \quad (5)$$

感受野的有效性由内向外逐渐减弱, 呈现类似高斯分布的特征。此外, 目标越小, 其特征越集中, 凸显了中心特征的重要性。PConv 采用了分组卷积技术, 在显著扩大感受野的同时最大限度减少了参数量。

1.2 EVC 模块

航拍交通图片很多有密集目标被遮挡的场景, 利用注意力或视觉变换器学习层内特征的方法, 忽略了密集预测任务

的角落区域,进而导致漏检,为此在颈部网络中添加 EVC-Block^[17]。EVCBlock 模块主要由轻量级 MLP 和可学习的视觉中心机制(LVC)。采用轻量级 MLP 架构捕捉远程依赖关系,同时通过并行的可学习视觉中心机制聚合输入图像的局部角部区域。集成后的特征兼具两大模块优势,用于下游识别。

首先,提取输入图像包含五个层级特征的金字塔结构 X ,其中每层特征图 X_i ($i=0,1,2,3,4$) 的空间尺寸分别为输入图像的 $1/2$ 、 $1/4$ 、 $1/8$ 、 $1/16$ 和 $1/32$ 。然后通过轻量级 MLP 架构捕捉 X_4 层的全局长程依赖关系,EVC 使用 MLP 层替代标准 Transformer 编码器中的多头自注意力模块,与 Transformer 相比,MLP 不仅结构简洁,还具有更轻量的体积和更高的计算效率。此外,配合轻量级 MLP, EVC 还采用可学习视觉中心机制来聚合输入图像的局部角部区域。 X_4 经过 Stem 模块的输出通道数为 256 的 7×7 卷积层、批归一化层和激活函数层进行特征平滑处理,使得 EVCBlock 模块能够更加准确地处理这些特征。

EVCBlock 模块的计算公式为:

$$X = \text{Cat}(\text{MLP}(X_{in}); \text{LVC}(X_{in})) \quad (6)$$

式中: X 表示 EVC 模块的输出, Cat 表示沿通道维度的特征图拼接操作, $\text{MLP}(X_{in})$ 和 $\text{LVC}(X_{in})$ 分别表示轻量级 MLP 与可学习视觉中心机制输出的特征, X_{in} 表示 Stem 模块的输出,其计算方式如图 3 和式 (7) 所示。

$$X_{in} = \sigma(\text{BN}(\text{Conv}7 \times 7(X_4))) \quad (7)$$

式中: $\text{Conv}7 \times 7$ 表示一个 7×7 卷积,步长为 1,通道数为 256; $\text{BN}()$ 表示批归一化层; $\sigma()$ 表示 ReLU 激活函数。

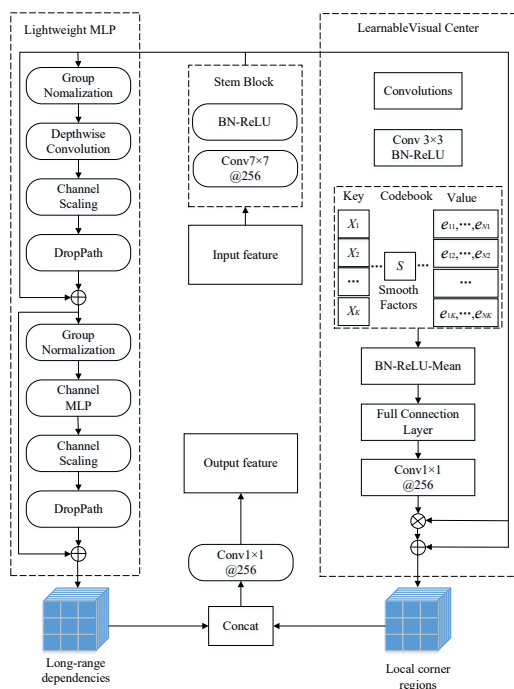


图 3 EVCBlock 模块结构图

1.3 SEMA 分离和增强注意力模块

SEAM (separated and enhancement attention module) 分离和增强注意力模块^[18]的设计是为了解决复杂场景下的遮挡可能引起特征缺失,导致的精确度下降等问题。SEAM 算法通过增强未遮挡区域的特征响应并增强遮挡区域的信息损失,提升模型对遮挡小目标的检测能力。SEAM 算法首先逐个通道深度可分离卷积并结合残差连接,不同深度的卷积层通过点卷积融合,进而减少参数冗余并缓解梯度消失问题。然后,利用两层全连接网络融合通道间的信息,从而整合遮挡区域与周围区域的关联。最后,通过指数归一化操作将全连接层输出映射到 $[1,e]$ 区间,减少算法的位置误差,如图 4 所示。

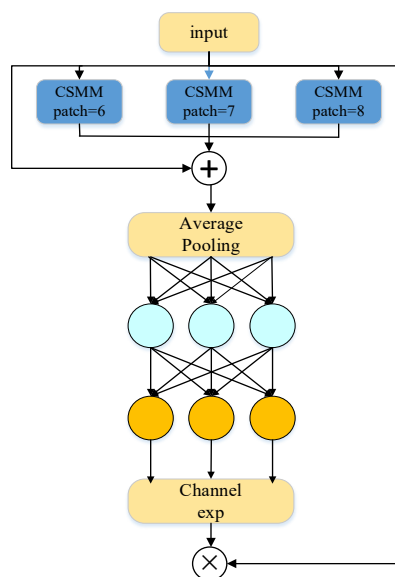


图 4 SEMA 结构图

2 实验结果与分析

2.1 实验参数和数据集

本文实验环境的操作系统为 Windows10, 实验环境配置如表 1 所示。

表 1 实验环境配置

名称	环境参数
操作系统	Windows10
GPU	NVIDIA GeForce RTX 4080, 16 G
Python 版本	Python3.9.23
学习框架	PyTorch 2.4.1
CUDA 版本	12.1

在实验中均设置相同的参数设置,优化器为 SGD, 实验训练迭代次数为 150epoch, 初始学习率为 0.01, 图片大小设置是 640 px×640 px, 模型最优超参数 batch-size 设置为 4。本文采用的数据集是天津大学的研究团队公开的无人机视角

的 VisDrone 数据集，包含 6 471 张训练集图像、1 610 张测试集图像和 548 张验证集图像。本文所识别的目标包括汽车、面包车、卡车、公共汽车和摩托车。

2.2 消融实验

为了验证本文改进 YOLOv11 算法是否合理，选用准确率（precision, P ），召回率（recall, R ），平均精度均值 mAP（mean average precision, mAP）、浮点运算量（giga floating-point operations per second, GFlops）和参数量（number of parameters, Params）等指标进行评估。

$$P = \frac{TP}{TP+FP}$$

(8)

$$R = \frac{TP}{TP+FN}$$

(9)

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

(10)

$$mAP = \frac{1}{n} \sum_{i=0}^n AP_{(i)}$$

(11)

式中：TP 为预测目标的正样本数量，FP 为预测目标的负样本数量，FN 为未能正确预测的个数， n 为航拍交通目标个数。

本文分别增加 Pconv、EVC、SEAM 算法进行实验，检测引入后的算法检测航拍小交通目标指标，结果如表 2 所示。

表 2 消融实验图

单位：%				
模型	P	R	mAP@0.5	mAP@0.5-0.95
YOLOv11m	68.66	55.07	58.94	40.1
YOLOv11m + PConv	68.57	56.59	61.09	41.8
YOLOv11m + EVC	68.44	57.77	61.98	42.15
YOLOv11m + SEAM	69.31	57.95	62.25	42.21
YOLOv11m + PConv + EVC	68.19	55.77	60.42	40.53
YOLOv11m + PConv + SEAM	69.57	56.63	61.15	41.72
YOLOv11m + EVC + SEAM	70.37	57.49	62.12	42.24
YOLOv11m + PConv + EVC + SEAM	71.98	57.13	62.67	42.56

从表 2 中可以看出，在骨干网络中引入 PConv 风车状卷积后，准确率降低 0.09%，但平均精度均值 mAP@0.5/% 为 61.09%，与原始的 YOLOv11m 相比增加了 2.15%，mAP@0.5-0.95/% 为 41.8%，较 YOLOv11m 增加了 1.7%。在颈部网络中引入 EVC 后，准确率虽然降低了 0.22%，

平均精度均值 mAP@0.5/%，mAP@0.5-0.95/% 分别增加了 3.04%、2.05% 为 61.98% 和 42.15%，较原始 YOLOv11m 对航拍小目标图像检测平均精度有一个提升。在头部网络中增加分离和增强注意力模块 SEAM 后，准确率增加了 0.65%，mAP@0.5/%，mAP@0.5-0.95/% 分别增加了 3.31% 和 2.14%。在 YOLOv11m 中引入 PConv、EVC 和 SEAM 后，准确率为 71.98%，较原始 YOLOv11m 增加 3.32%，mAP@0.5/% 和 mAP@0.5-0.95/% 分别增加了 3.73%、2.46%，说明测算对航拍小目标图像的检测性能有提升。

2.3 对比实验

为进一步检测改进 YOLOv11m 算法的有效性，在实验环境相同的情况下，将本文算法与主流算法 YOLOv5m、YOLOv8m、YOLOv8n、YOLOv9s、YOLOv10n 对比，得到如表 3 所示。

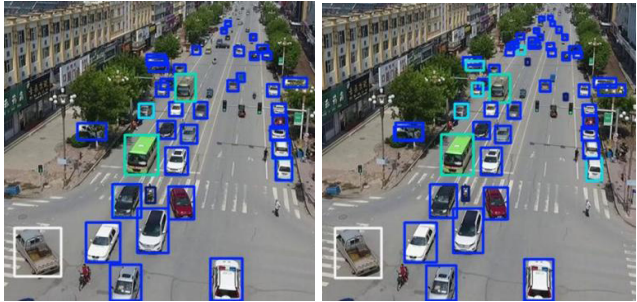
表 3 不同算法对比表

单位：%				
模型	P	R	mAP@0.5	mAP@0.5-0.95
YOLOv5m	66.91	47.30	52.22	33.89
YOLOv8m	68.89	51.83	56.58	38.02
YOLOv9m	67.09	52.72	57.13	38.4
YOLOv10m	69.23	52.63	57.22	38.6
YOLOv11m	68.66	55.07	58.94	40.1
改进 YOLOv11m	71.98	57.13	62.67	42.56

实验结果表明，随着 YOLO 算法的升级，准确率和平均精度也相应地提高。改进 YOLOv11m 算法，在精确度，召回率和平均精度 mAP@0.5、mAP@0.5-0.95/% 上表现优越。

2.4 可视化结果与分析

图 5（a）为原 YOLOv11m，图 5（b）为改进 YOLOv11m 算法。可以看出，原始的 YOLOv11m 在小目标车辆的检测上漏检更多，改进的算法在车辆数量密集，图像目标极小的情况下，检测效果更高。误检、漏检、检测精度值。



(a) YOLOv11m (b) 改进 YOLOv11m

图 5 可视化结果对比图

3 结论

本文提出了一种改进的 YOLOv11m 航拍交通小目标检测算法。在骨干网络中添加 PConv 卷积,使 YOLOv11 能够对齐小目标特征图的高斯空间分布,抑制杂波信号,增强航拍小交通目标与背景的对比;在颈部网络中增加 EVC,捕获全局长依赖、保留在角落处的信息,提高小目标在密集场景下的检测性能;在头部网络中增加 SEAM 注意力机制增强网络对关键特征的检测敏感度,从而提高对航拍小交通目标的检测精度。实验结果表明,改进 YOLOv11m 算法的准确率增加 3.32%,平均检测精度 $mAP@0.5\%$ 和 $mAP@0.5-0.95\%$ 分别增加了 3.73%、2.46%。

参考文献:

- [1] 宋晓茹,刘康,高嵩,等.复杂战场环境下改进 YOLOv5 军事目标识别算法研究[J].兵工学报,2024,45(3):934-947.
- [2] 罗旭东,吴一全,陈金林.无人机航拍影像目标检测与语义分割的深度学习研究方法研究进展[J].航空学报,2024,45(6):241-270.
- [3] Khan N A, Jhanjhi N Z, Brohi S N, et al. Smart traffic monitoring system using unmanned aerial vehicles (UAVs)[J]. Computer Communications, 2020, 157: 434-443.
- [4] Yang Fan, Fan Heng, Chu Peng, et al. Clustered object detection in aerial images[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 8310-8319.
- [5] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6):1137-1149.
- [6] He Kaiming, Gkioxari G, Dollar P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2980-2988.
- [7] 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 (CVPR). Piscataway: IEEE, 2016: 779-788.
- [8] Liu Wei, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//Computer Vision – ECCV 2016. Berlin: Springer, 2016: 21-37.
- [9] Wang Qinglang, Mao Jingchao, Zhai Xu, et al. Improvements of YOLOv3 for road damage detection[C]//2021 3rd International Conference on Computer Information and Big Data Applications (CIBDA). Bristol: IOP Publishing Ltd, 1903: 012008.
- [10] Yi Weiguang, Wang Bo. Research on underwater small target detection algorithm based on improved YOLOv7[J]. IEEE Access, 2023, 11: 66818-66827.
- [11] 奉志强,谢志军,包正伟,等.基于改进 YOLOv5 的无人机实时密集小目标检测算法[J].航空学报,2023,44(7):251-265.
- [12] Dong Xudong, Yan Shuai, Duan Chaoqun. A lightweight vehicles detection network model based on YOLOv5[J]. Engineering Applications of Artificial Intelligence, 2022, 113: 104914.
- [13] 何湘杰,宋晓宁. YOLOv4-Tiny 的改进轻量级目标检测算法[J]. 计算机科学与探索, 2024, 18(1): 138-150.
- [14] 何植仟,曹立杰. UAVAI-YOLO: 无人机航拍图像的小目标检测模型[J]. 智能科学与技术学报, 2024, 6(2):262-271.
- [15] 彭晏飞,陈炎康,赵涛,等.基于重参数化注意力的无人机航拍目标检测方法[J]. 光电与控制, 2024, 31(9):81-86.
- [16] Yang Jiangnan, Liu Shuangli, Wu Jingjun, et al. Pin-wheel-shaped convolution and scale-based dynamic loss for infrared small target detection [PP/OL]. (2024-12-22)[2026-05-23]. <https://doi.org/10.48550/arXiv.2412.16986>.
- [17] Quan Yu, Zhang Dong, Zhang Liyan, et al. Centralized feature pyramid for object detection[J]. IEEE Transactions on Image Processing, 2023, 32: 4341-4354.
- [18] Yu Ziping, Huang Hongbo, Chen Weijun, et al. YOLO-FaceV2: a scale and occlusion aware face detector[J]. Pattern Recognition, 2024, 155: 110714.

【作者简介】

丁美云(1984—),女,山东菏泽人,本科,工程师,研究方向:轨道交通建设管理, email: dingmeiyun123456@163.com。

杜杰(1986—),男,山东临沂人,本科,工程师,研究方向:轨道交通建设管理, email: famouswood1986@126.com。

高扬(1981—),男,山东日照人,本科,工程师,研究方向:轨道交通建设管理, email: gaoyang2356@163.com。

丁四景(1990—),通信作者(email: dingsijing0505@163.com),女,山东菏泽人,硕士研究生,工程师,研究方向:深度学习与目标检测、数据挖掘、智能交通系统。

(收稿日期: 2025-11-11 修回日期: 2026-06-01)