Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2025, 14(3), 519-526
Published Online May 2025 in Hans. https://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2025.143051

Hans汉斯

基于改进YOLOv5的行人目标检测算法研究

肖黄梅，张　宇，杨会芳*，程　静

重庆对外经贸学院大数据与智能工程学院，重庆

收稿日期：2025年3月15日；录用日期：2025年4月30日；发布日期：2025年5月12日

摘　要

本文针对行人目标检测中的挑战，提出了一种基于改进YOLOv5的行人检测算法。该算法融合了Ghost模块和SE注意力机制，旨在提高特征提取能力的同时，保持模型的轻量性。在面对密集场景和遮挡问题时，改进的YOLOv5能有效提取重要特征并提升检测精度。通过对比实验和模拟分析，验证了该算法在提升检测性能的同时，仍能保持较低的计算复杂度和较高的实时性。实验结果表明，所提算法在行人检测任务中具有较好的表现，尤其是在低照度和复杂背景条件下。

关键词

行人目标检测，YOLOv5，Ghost，SE

Research on Pedestrian Target Detection Algorithm Based on Improved YOLOv5

Huangmei Xiao, Yu Zhang, Huifang Yang*, Jing Cheng

School of Big Data and Intelligent Engineering, Chongqing College of International Business and Economics, Chongqing

Received: Mar. 15th, 2025; accepted: Apr. 30th, 2025; published: May 12th, 2025

Abstract

Aiming at the challenges in pedestrian target detection, this paper proposes a pedestrian detection algorithm based on improved YOLOv5. The algorithm combines the Ghost module and SE attention mechanism to improve the feature extraction capability while maintaining the lightweight of the model. When faced with dense scenes and occlusion problems, the improved YOLOv5 can effectively extract important features and improve detection accuracy. Through comparative experiments and simulation analysis, it is verified that the algorithm can maintain low computational complexity and high real-time

*通讯作者。

文章引用: 肖黄梅, 张宇, 杨会芳, 程静. 基于改进 YOLOv5 的行人目标检测算法研究[J]. 人工智能与机器人研究, 2025, 14(3): 519-526. DOI: 10.12677/airr.2025.143051

performance while improving detection performance. Experimental results show that the proposed algorithm has good performance in pedestrian detection tasks, especially under low illumination and complex background conditions.

Keywords

Pedestrian Target Detection, YOLOv5, Ghost, SE

1. 引言

行人检测是智能监控、交通管理和自动驾驶等领域的重要研究方向。然而，在密集场景下，由于行人间距较小、频繁遮挡以及复杂背景的干扰，检测任务变得极具挑战性。行人间的遮挡减少了可见区域特征，导致特征提取困难，同时背景干扰加剧，使得网络模型难以准确分类和定位，进而引发误检与漏检[1]。因此，优化行人检测算法，提升检测精度的同时保证实时性，是未来研究的核心方向。

Wang 等[2]通过更换新的骨干网络并结合预训练模型，增强了目标检测的特征提取能力，从而提供了更丰富的特征信息。Bodla 等[3]采用优化目标预测帧分数策略的方法，增加了检测帧的数量，从而提升了对遮挡目标的检测性能。然而，提升模型的泛化能力仍然是一个重要的研究课题。Xue 等[4]提出了一种创新的实时行人检测算法——多模态注意力融合 YOLO，该算法基于 DarkNet53 框架优化夜间行人检测性能。

为了解决上述问题，本文提出了一种改进型 YOLOv5 行人目标检测算法[5]，该算法集成了 Ghost 模块和 SE 注意力机制，在提升特征提取能力的同时，保持模型的轻量化。并在 INRIA 数据集上进行了系统性实验分析。与 FasterR-CNN、原始 YOLOv5 及其他 YOLOv5 改进方案相比，本方法在遮挡、小目标检测等复杂场景下表现更优，且计算成本较低，验证了所提算法的有效性和优越性。

2. 方法介绍

YOLOv5 系列网络根据残差结构的宽度和深度分为 YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 五类，结构相似但复杂度不同。表 1 展示了各模型在 MS COCO 数据集上的表现，分析表明 YOLOv5 在准确率与复杂度间实现了良好平衡。

Table 1. Performance of each model on the MS COCO dataset
表 1. 各模型在 MS COCO 数据集上的表现

Method	Image size	*mAP*0.5 (%)	*mAP*0.5:0.95 (%)	FLOPs (G)
YOLOv5n	640 × 640	45.7	28.0	4.5
YOLOv5s	640 × 640	56.8	37.4	16.5
YOLOv5m	640 × 640	64.1	45.4	49.0
YOLOv5l	640 × 640	67.3	49.0	109.1
YOLOv5x	640 × 640	68.9	50.7	205.7

YOLOv5s 由 Backbone、Neck 和 Prediction 三个主要组件组成[6]，各模块的高效协作提升了模型性

能。该网络引入了多项优化，如 Mosaic 数据增强和自适应锚框计算。Backbone 采用 Focus 和 CSP_X 结构，Neck 结合 CSP2 结构，以增强特征融合能力。此外，空间金字塔池化(SPP)用于融合不同感受野特征，边界框损失函数采用 CIOU，NMS 非极大值抑制优化了重叠目标检测。YOLOv5s 的整体结构如图 1 所示。

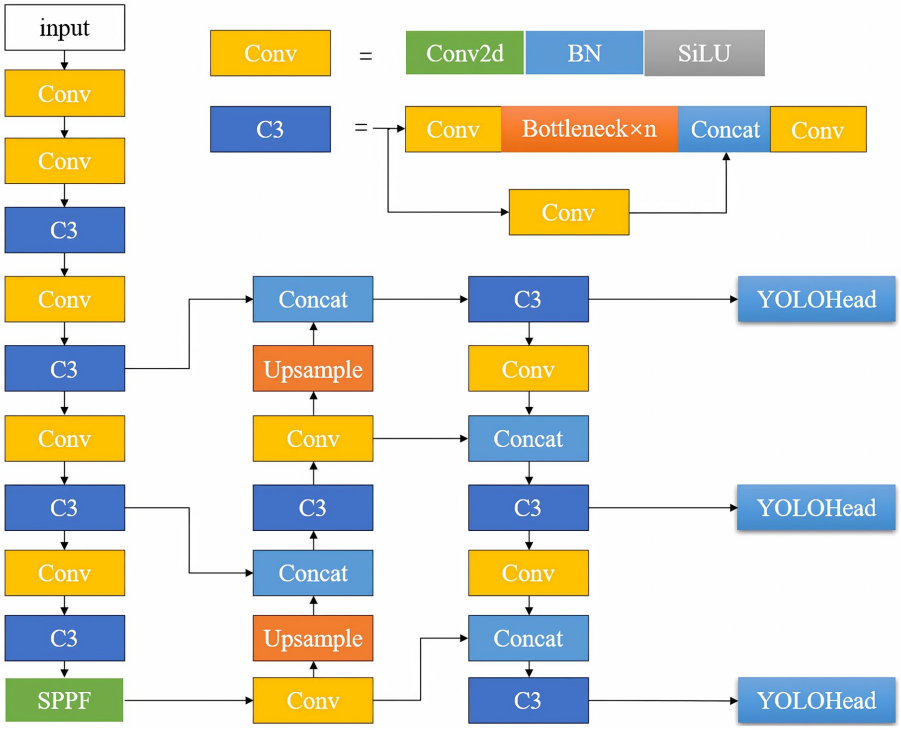

Figure 1. YOLOv5 network structure
图 1. YOLOv5 网络结构

2.1. Ghost 模块

本文引入了 Ghost 模块以优化模型的轻量化设计[7]。Ghost 卷积通过将输入通道划分为两部分：一部分采用标准卷积运算，另一部分使用较小的核(如 5 × 5)进行简化计算，以减少计算复杂度。该方法能有效降低模型的计算量和存储需求，同时保持良好的特征提取能力。此外，Ghost 卷积利用组卷积机制，将输入通道划分为多个独立组，使每个组分别执行卷积计算，从而减少参数数量和计算开销，同时提升计算效率。Ghost 卷积的示意图如图 2 和图 3 所示。

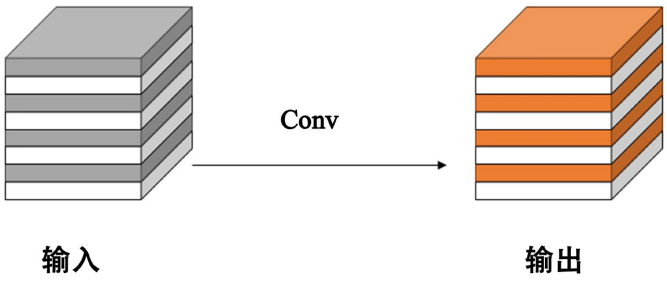

Figure 2. Ordinary convolution
图 2. 普通卷积

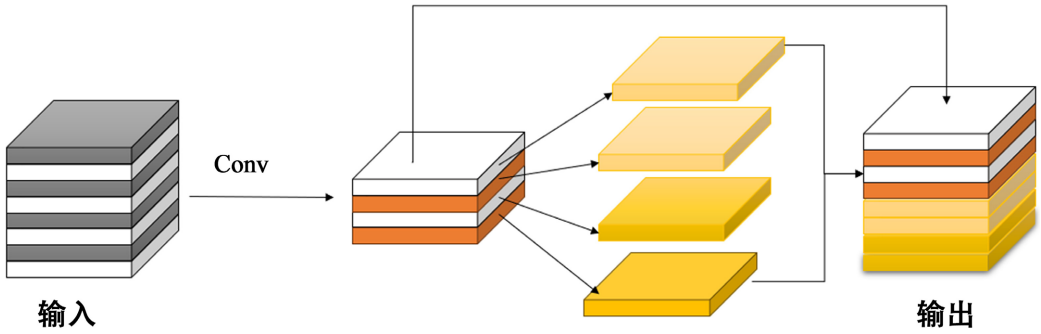

Figure 3. Ghost convolution
图 3. Ghost 卷积

Ghost 卷积的计算过程包括多个步骤。首先，对输入特征图 $X \in R^{c\times h\times w}$ 进行标准卷积操作，提取初步特征并生成较小的输出张量。随后，利用 Ghost 卷积生成额外的特征图，以扩展输出通道。最后，将 Ghost 卷积生成的特征与初始卷积输出沿通道维度拼接，得到最终的输出特征图 $Y \in R^{c\times h\times w}$。在此过程中，标准卷积的计算量可表示为 $c\times h'\times w'\times c'\times k\times k$。相比之下，在生成相同数量特征的情况下，Ghost 卷积减少了冗余计算，有效降低了计算成本，并加速了推理过程。两者的计算速率关系可由公式(1)推导得出。

$$r_s = \frac{c\times h\times w\times c'\times k\times k}{\frac{c}{s}\times h\times w\times c'\times k\times k + (s-1)\times\frac{c}{c'}\times h\times w\times d\times d} \approx \frac{s\times c'}{s+c-1} \approx s \tag{1}$$

Ghost 模块通过减少冗余计算，大幅降低计算成本并压缩模型大小。其核心思想是利用较小的标准卷积核 $d\times d$ 提取初步特征，并通过简单变换操作 s 生成额外特征，满足 $s \ll c'$。根据等式(1)，Ghost 模块的计算量相比标准卷积减少了 $1/s$ 倍，从而有效提升计算效率。

2.2. 引入注意力机制模块

为了增强全局信息利用率，YOLOv5s 在主干网络中引入了 SE (Squeeze-and-Excitation)通道注意模块[8]，其特征变换过程如图 4 所示。首先，SE 模块通过变换算子 F_{tr} 处理输入特征图 $X \in \mathbb{R}^{H'\times W'\times C'}$，生成输出 $U \in \mathbb{R}^{H\times W\times C}$。其中，$F_{tr}$ 作为卷积算子，包含滤波核集合 $M = [m_1, m_2, \cdots, m_c]$，用于提取关键特征，得到映射后的结果 $U = [u_1, u_2, \cdots, u_c]$。

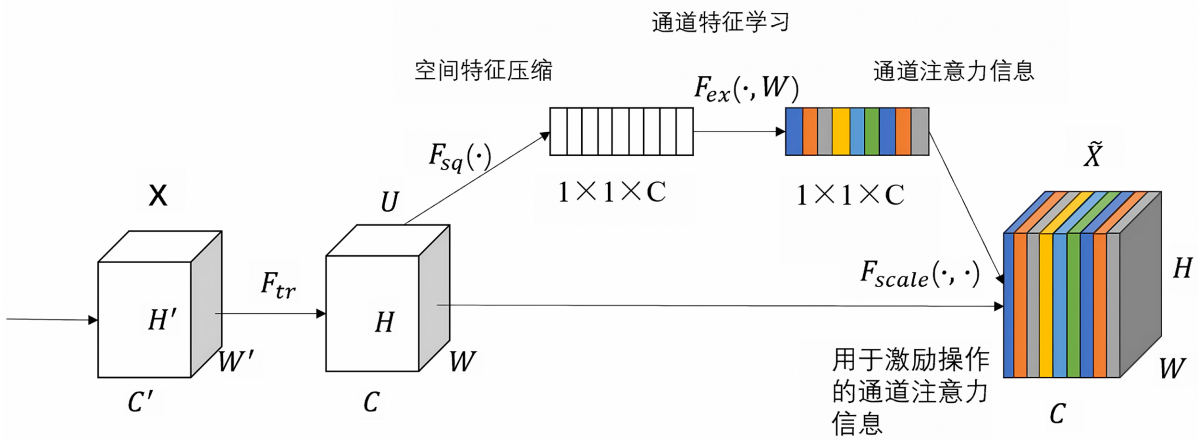

Figure 4. SE channel attention mechanism
图 4. SE 通道注意力机制

$$u_c = m_c * X = \sum_{s=1}^{c'} m_c^s * x^s. \tag{2}$$

其中，$*$代表卷积运算，滤波核 m_c 由参数矩阵 $\left[m_c^1, m_c^2, \cdots, m_c^{c'}\right]$ 组成，而输入特征 X 包含通道集合 $\left[x^1, x^2, \cdots, x^{c'}\right]$。其中，每个通道 x^s 对应于滤波核 m_c^s，该滤波核是一个二维空间核。

随后，该特征图经过全局平均池化(F_{sq})，将其空间维度 $H \times W \times C$ 压缩为 $1 \times 1 \times C$。接着，利用全连接层(FC)学习通道注意力，使特征保持相同维度但增强通道权重。最终，原始特征图与生成的 $1 \times 1 \times C$ 注意力权重相乘，恢复 $H \times W \times C$ 形态，同时强化通道信息。

2.3. 损失函数

YOLOv5 网络[9]的总损失函数由三个部分组成：定位损失 L_{box}、置信度损失 L_{obj} 和分类损失 L_{cls}。损失函数如公式(3)~(6)所示：

$$L_{\text{total}} = L_{\text{box}} + L_{\text{obj}} + L_{\text{cls}} \tag{3}$$

$$L_{\text{box}} = \lambda_{\text{IoU}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} l_{ij}^{\text{obj}} L_{\text{CIoU}} \tag{4}$$

$$L_{\text{obj}} = \lambda_{\text{cls}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} l_{ij}^{\text{obj}} \lambda_c \left(C_i - \hat{C}_i\right)^2 + \lambda_{\text{cls}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} l_{ij}^{\text{noobj}} \lambda_c \left(C_i - \hat{C}_i\right)^2 \tag{5}$$

$$L_{\text{cls}} = -\sum_{i=0}^{S^2} \sum_{j=0}^{B} l_{ij}^{\text{obj}} \sum_{c \in \text{classes}} \lambda_c \tag{6}$$

其中，λ 表示总体损失平衡系数，l 表示每一类样本的损失权重，C_i 表示置信度得分。

3. 实验验证

本研究在 CUDA 11.2 版本的 PyTorch 深度学习框架下进行模型训练，硬件环境包括 AMD Ryzen 7 5800H 处理器和 NVIDIA GeForce RTX 3060 GPU。本文采用 INRIA 数据集，该数据集包含标注的站立或行走行人图像，由 NavneetDalal 在研究图像和视频中直立行人检测时收集。INRIA 数据集的训练集包括 614 张正样本图像(共包含 1237 个行人)和 1218 张负样本图像；测试集包含 288 张正样本图像(589 个行人)和 453 张负样本图像。如图 5 所示，展示了 INRIA 数据集的部分样本。

Figure 5. Some samples of the INRIA dataset
图 5. INRIA 数据集的部分样本

为了从不同角度比较和分析不同检测模型的优缺点，本文使用精度 P、召回率 R、平均精度 mAP 作

为所提算法的主要评价指标，见下式(7)~(10)：

$$P=\frac{TP}{TP+FP} \tag{7}$$

$$R=\frac{TP}{TP+FN} \tag{8}$$

$$AP=\int_0^1 P(R)\mathrm{d}R \tag{9}$$

$$mAP=\frac{1}{n}\sum_{i=1}^{n} AP \tag{10}$$

其中，TP 表示预测正确的正样本数，FP 表示预测错误的正样本数，FN 表示预测错误的负样本数；n 表示物体检测类别数。引入交集比，即预测帧与真实帧的交集与并集之比，本实验中设置为 0.5。

为了验证改进算法的有效性，本研究进行了消融实验，分别是包含 YOLOv5s、YOLOv5-Ghost 与 YOLOv5-Ghost-SE。如表 2 所示，展示了改进的 YOLOv5s 算法的消融实验结果。

Table 2. Comparison results of ablation experiments
表 2. 消融实验对比结果

YOLOv5	Ghost	SE	*P*/%	*R*/%	*mAP*/%
√			89.16	88.58	89.31
√	√		90.65	90.12	92.18
√	√	√	93.28	91.49	92.24

表 2 展示了消融实验的对比结果，以评估 Ghost 模块和 SE (Squeeze-and-Excitation)注意力机制对 YOLOv5 行人检测性能的影响。基础 YOLOv5 模型的精确率(*P*)为 89.16%，召回率(*R*)为 88.58%，*mAP* (均值平均精度)为 89.31%。当引入 Ghost 模块后，*P*、*R* 和 *mAP* 分别提升至 90.65%、90.12%和 92.18%，表明 Ghost 模块有效降低计算成本的同时，增强了特征表达能力，提高了检测性能。在此基础上进一步加入 SE 注意力机制后，精确率、召回率和 *mAP* 进一步提升至 93.28%、91.49%和 92.24%。这表明 SE 机制能够有效增强网络对重要特征的关注，提高检测的准确性和鲁棒性。综合来看，Ghost 模块和 SE 注意力机制的结合，使得 YOLOv5 的行人检测能力得到显著增强，优化了检测精度的同时，保持了模型的轻量化。

之后进一步对该改进的模型进行对比，实验选择了 YOLOv5 和 Faster RCNN [10]进行对比实验。表 3 给出了这三个模型在测试集上的具体实验结果。

Table 3. Comparative experimental data
表 3. 对比实验数据

模型	*P*/%	*R*/%	*mAP*0.5/%
Faster RCNN	87.89	86.94	87.39
YOLOv5	89.16	88.58	89.31
Ours	93.28	91.49	92.24

表 3 展示了不同目标检测模型在行人检测任务上的性能对比，包括 FasterR-CNN、YOLOv5 以及本文提出的改进模型(Ours)。从实验数据来看，FasterR-CNN 的精确率(*P*)为 87.89%，召回率(*R*)为 86.94%，*mAP*0.5 为 87.39%，说明其检测能力较为稳定，但由于其两阶段检测框架，可能在实时性方面存在一定

的局限性。相比之下，YOLOv5 作为单阶段检测模型，精确率、召回率和 *mAP* 分别提升至 89.16%、88.58%和 89.31%，在保证检测精度的同时，具有更高的计算效率。本文提出的改进模型(Ours)在YOLOv5 的基础上进一步优化，*P*、*R* 和 *mAP* 分别达到了 93.28%、91.49%和 92.24%，显著优于前两者。这表明，通过引入 Ghost 模块和 SE 注意力机制，模型的特征提取能力得到了增强，从而提高了检测精度和召回率，同时保持了较高的计算效率，使其在行人检测任务中表现更优。图 6 展示了不同的模型之间的性能对比。

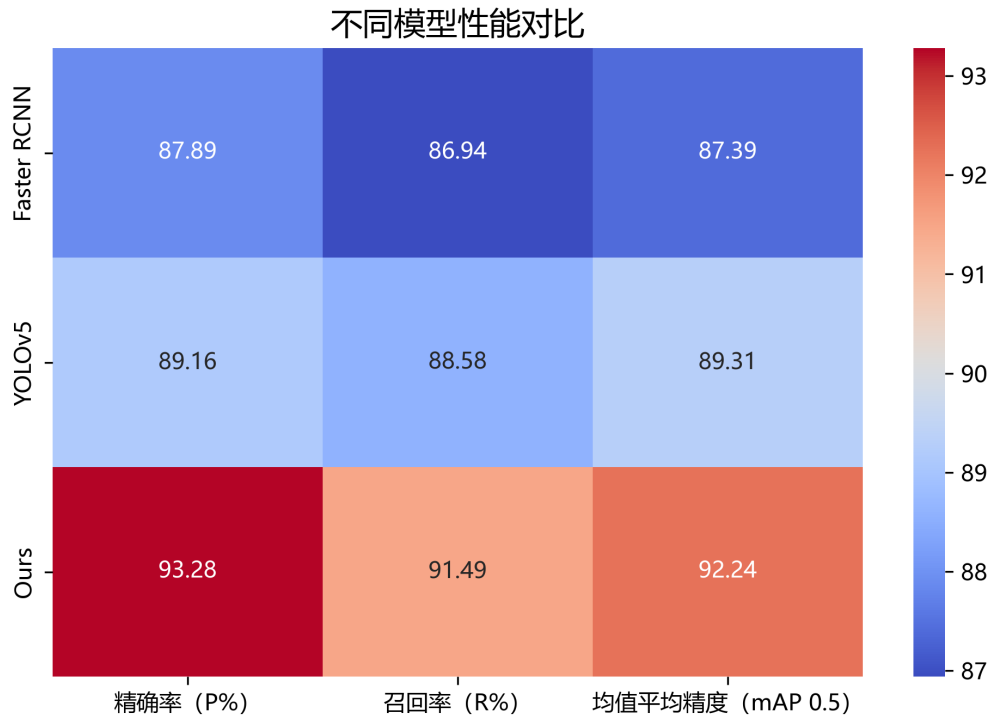

Figure 6. Heat map display
图 6. 热力图展示

图 6 展示了不同行人检测模型在精确率、召回率和 *mAP* 三项指标上的性能对比。横轴表示评价指标，纵轴代表 FasterR-CNN、YOLOv5 和改进模型。颜色从蓝色到红色变化，数值越高，颜色越偏向红色，数值较低则表现为蓝色。从实验结果来看，FasterR-CNN 在所有指标上的数值最低，整体呈现较深的蓝色。YOLOv5 在各项指标上相较 FasterR-CNN 略有提升，精确率、召回率和 *mAP* 分别达到 89.16%、88.58%和 89.31%，颜色也较浅，表明检测性能有所增强。相比之下，改进模型 Ours 在所有指标上均表现最佳，精确率达到 93.28%，召回率为 91.49%，*mAP* 为 92.24%，呈现最红的色块，直观地反映了其在实验范围内表现更优。

4. 结语

本研究提出了一种基于改进 YOLOv5 的行人目标检测系统，旨在提高行人检测的准确性和实时性。本实验对 FasterRCNN、YOLOv5 及改进模型在行人检测任务中的性能进行了系统对比。实验结果表明，相较于 FasterRCNN，YOLOv5 由于其单阶段检测架构，在检测精度与效率上均具备优势。进一步引入 Ghost 模块进行特征解耦重构，并结合 SE 注意力机制进行动态特征校准，使改进模型在保持实时推理能力的同时，实现 93.28%的精确率、91.49%的召回率、92.24%的 *mAP*0.5 的提升。消融实验验证了轻量化

特征提取和通道注意力机制的有效性。

基金项目

1) 2024-2025 年重庆对外经贸学院《智能算法应用》课程教学改革(项目编号：KG2024067)；

2) 2024-2025 年重庆对外经贸学院科研项目：人工智能图像识别技术助力农产品品质分级与分拣(项目编号：KYZK2024028)；

3) 2024-2025 年重庆对外经贸学院科研项目：股票市场波动的统计特征分析与预测模型研究项(项目编号：KYZK2024042)。

参考文献

[1] Ren, Z., Lam, E.Y. and Zhao, J. (2020) Real-time Target Detection in Visual Sensing Environments Using Deep Transfer Learning and Improved Anchor Box Generation. *IEEE Access*, **8**, 193512-193522. https://doi.org/10.1109/access.2020.3032955

[2] Wang, T., Anwer, R.M., Cholakkal, H., Khan, F.S., Pang, Y. and Shao, L. (2019) Learning Rich Features at High-Speed for Single-Shot Object Detection. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 1971-1980. https://doi.org/10.1109/iccv.2019.00206

[3] Bodla, N., Singh, B., Chellappa, R. and Davis, L.S. (2017) Soft-NMS—Improving Object Detection with One Line of Code. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 5562-5570. https://doi.org/10.1109/iccv.2017.593

[4] Xue, Y., Ju, Z., Li, Y. and Zhang, W. (2021) MAF-YOLO: Multi-Modal Attention Fusion Based YOLO for Pedestrian Detection. *Infrared Physics & Technology*, **118**, Article ID: 103906. https://doi.org/10.1016/j.infrared.2021.103906

[5] 王维锋, 邵琳骞, 黄建鑫, 等. 基于改进 YOLOv5 的双模态融合行人检测方法[J/OL]. 吉林大学学报(工学版): 1-11. https://doi.org/10.13229/j.cnki.jdxbgxb.20241288, 2025-03-12.

[6] 薛仁政, 吴乾龙, 叶宝丰. 基于可变形卷积的 YOLOv5 行人检测算法[J]. 齐齐哈尔大学学报(自然科学版), 2025, 41(2): 14-21, 27.

[7] 舒密, 王占刚. 基于 Ghost 卷积与自适应注意力的点云分类[J]. 现代电子技术, 2025, 48(6): 106-112.

[8] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. https://doi.org/10.1109/cvpr.2018.00745

[9] 许皓翔, 殳国华. 基于轻量化的 YOLOv5 的 PCB 缺陷检测算法[J]. 电气自动化, 2024, 46(2): 95-97, 102.

[10] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision* (*ICCV*), Santiago, 7-13 December 2015, 1440-1448. https://doi.org/10.1109/iccv.2015.169