

导盲机器人多尺度轻量检测算法优化研究

王世刚¹ 方欣怡¹

WANG Shigang FANG Xinyi

摘要

为提升盲人和视障人士 (BVIP) 在复杂的城市交通环境中的自主导航能力, 文章提出一种面向导盲机器人导向信息多尺度轻量检测的算法。首先, 引入 MultiSEAM 和 LSKA 模块 (local spatial kernel attention), 并结合 DcovN 机制 (dynamic covariance normalization) 和自适应空间卷积注意力机制, 以增强模型的鲁棒性和检测精度; 其次, 为便于算法部署于导盲机器人, 采用轻量级的 MobileNetV3 网络替换 YOLOv5s 的 Backbone 主干特征提取网络; 最后, 引入 CA 注意力机制和 EIoU 损失函数以补偿精度损失。实验结果表明, 改进后的 YOLOv5s 算法在模型参数量上降低 20.8%, 在 F_1 得分、mAP@0.5 上分别提高 0.19 和 12.5%, 对比同系列 SOTA 模型如 YOLOv10、YOLOv11 等均显示出较强的鲁棒性。

关键词

多尺度目标检测; YOLOv5s; 导盲机器人; 深度学习; 轻量化

doi: 10.3969/j.issn.1672-9528.2025.09.046

0 引言

据数据统计, 全球约有 22 亿视力障碍者^[1], 其中, 失明人数约 4 300 万; 我国盲人数量超过 1 700 万, 每年新增的盲人数量高达 45 万。而目前 BVIP 出行面临着许多挑战, 例如盲道维护不足、交通信号缺失等问题, 严重影响了其日常生活的自主性, 而导盲犬训练成本高且数量有限^[2]。因此, 开发高效经济的导盲机器人系统成为一个重要的研究方向。

1. 广西科技大学自动化学院 广西柳州 545616

[基金项目] 广西科技基地和人才专项 (AD22080004)

近年来, 集成先进传感器如 LiDAR 和摄像头的导盲机器人逐渐兴起, 但在复杂城市环境中, 如何精准地识别导向目标 (如盲道、斑马线、红绿灯等) 仍然是技术难题^[3], 因此, 研究者们提出了多种算法优化方案。例如, 黄宁霞等人^[4]提出优化的 Faster RCNN 算法, 能够提升小型障碍物检测精度; 贾存南等人^[5]采用边缘计算的 YOLOv3 算法, 可以实现盲道环境实时检测; Sugashini 等人^[6]提出 SABO-YOLO 模型, 能够协助盲人检测障碍物。

尽管诸多研究取得了一定进展, 但在实际应用中仍然面临着许多挑战。特别是面对户外复杂导盲场景以及多尺度目

[10] 孙全, 叶秀清, 顾伟康. 一种新的基于证据理论的合成公式 [J]. 电子学报, 2000(8):117-119.

[11] FEI L G, HU Y, XIAO F Y, et al. A modified TOPSIS method based on D numbers and its applications in human resources selection[J]. Mathematical problems in engineering, 2016, 2016(1): 6145196.

[12] GENG J B, WEI Z J. A new method for constructing the basic probability assignment function in D-S evidence theory[J]. Journal of physics: conference series. IOP publishing, 2021, 1746(1): 012038.

[13] 张欢, 陆见光, 唐向红. 面向冲突证据的改进 D-S 证据理论算法 [J]. 北京航空航天大学学报, 2020, 46(3):616-623.

[14] 朱智宇, 宋亚飞. 证据理论中基于区间概率的不确定测度 [J]. 计算机应用, 2021, 41(S2):25-30.

[15] 邓勇, 施文康, 朱振福. 一种有效处理冲突证据的组合方法 [J]. 红外与毫米波学报, 2004(1):27-32.

【作者简介】

侯中华 (1999—), 男, 四川蓬溪人, 硕士研究生, 研究方向: 智能检测与控制技术。

石则云 (1994—), 男, 湖北十堰人, 博士研究生, 副教授, 研究方向: 非线性与量子光学。

黄海波 (1974—), 男, 湖北十堰人, 博士研究生, 教授, 研究方向: 汽车电子与智能控制。

吕子涵 (2001—), 男, 湖北应城人, 硕士研究生, 研究方向: 汽车电子与感应加热电源。

(收稿日期: 2025-04-13 修回日期: 2025-09-08)

标检测难度大、数据集类别不够全面、算法难以实现在嵌入式设备等移动平台上实现高效部署，同时，轻量化处理后的神经网络可能出现检测性能下降等问题。针对这些挑战，本文提出一种改进 YOLOv5s 算法，提升导盲机器人多尺度检测能力，进行轻量化改进以减少计算资源需求，提高检测效率。

1 YOLOv5s 算法简介

YOLOv5s 是基于 YOLOv5 架构，优化了速度和延迟而产生的一种高度优化的实时目标检测深度神经网络，作为 YOLOv5 系列中最小且速度最快的版本，YOLOv5s 适用于计算资源有限的设备，如移动设备或边缘计算环境。因此本文选择 YOLOv5s 算法作为导盲机器人目标检测算法并进行改进。

2 改进 YOLOv5s 的检测算法网络设计

在导盲机器人导向信息检测任务中，对于多尺度、小物体检测，固定的卷积核无法灵活地适应不同尺寸的物体，尤其是在复杂背景和目标尺度变化比较大时，传统卷积操作虽然能提取局部特征，但其感受野固定，致使卷积神经网络难以有效捕捉长距离空间依赖。因此，本文提出改进 YOLOv5s 提高导盲机器人对多尺度导向信息检测的能力。

具体思路是将 MultiSEAM 和 LSKA 两个模块结合应用，以更好地解决目标检测中的多尺度特征提取和空间感知问题。此外，引入 MobileNetV3 结构实现轻量化，并引入 CA 模块和 EIoU 损失函数以弥补轻量化导致的精度损失，进一步提高检测准确性。改进后的网络结构如图 1 所示。

2.1 MultiSEAM 模块

导盲机器人在城市交通环境中进行导向信息目标检测容易受到复杂背景的影响，为了更加精确地捕捉多尺度特征，

并增强对复杂背景和小物体的鲁棒性，引入 MultiSEAM 模块，在多尺度下捕捉并融合 Backbone 中提取的含深层语义特征的全局信息，为后续空间特征融合做准备。

如图 2 所示，MultiSEAM 包括 3 个不同的通道与空间混合模块（channel and spatial mixing module, CSMM），采用不同尺度的卷积核并行处理输入的特征信息，从而捕捉多尺度的特征信息。CSMM 将特征信息输出后进行全局平均池化，再经过两层全连接层融合通道信息，此过程引入了动态协方差归一化机制（dynamic covariance normalization, DcovN），通过计算特征图的协方差矩阵来动态调整每个通道之间的关系，这使得网络能够在归一化过程中更加关注特征之间的依赖性，从而提升了模型对复杂模式的表达能力。最后通过通道扩展（Channel exp）操作整合多尺度特征。

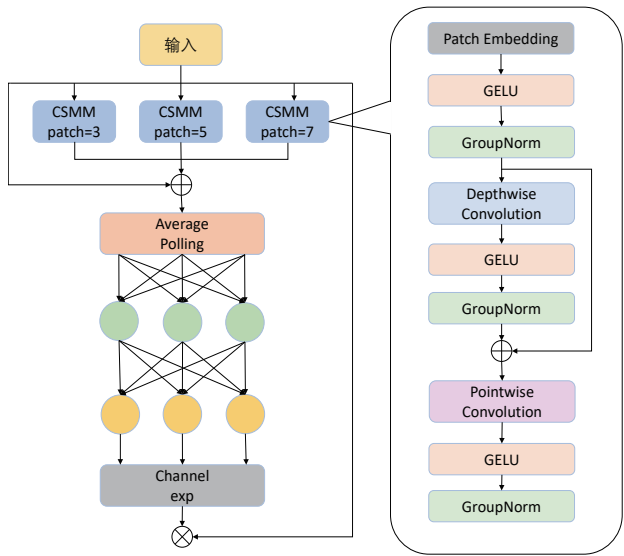


图 2 MultiSEAM 组成结构

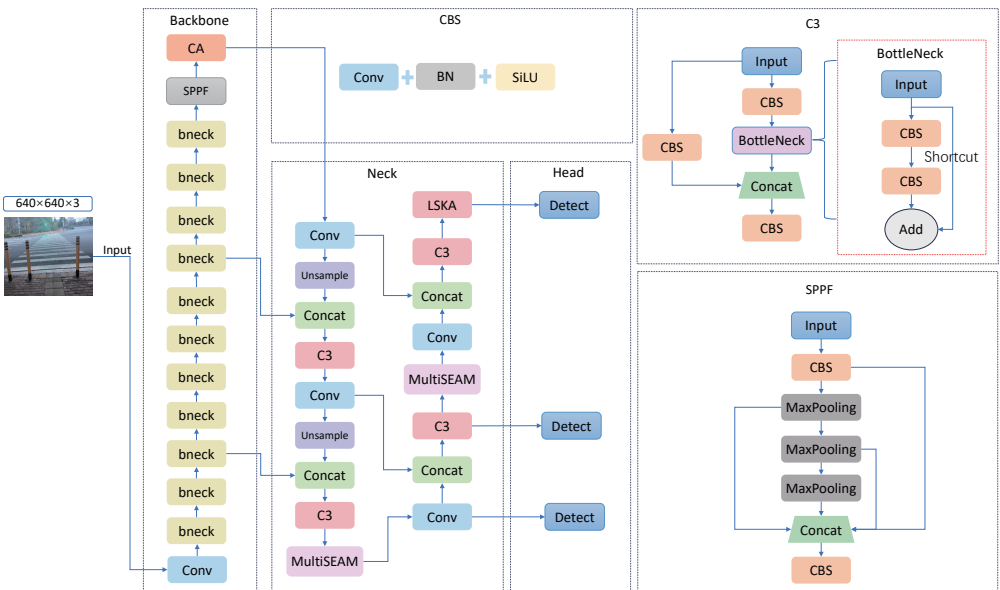


图 1 改进 YOLOv5s 导盲机器人多尺度轻量检测模型框架

2.2 LSKA 模块

导盲机器人导向信息背景的复杂性以及干扰因素增加了检测难度，引入大型可分离核注意力机制（local spatial kernel attention, LSKA），优化模型的特征提取和表达能力。

LSKA 模块凭借大范围可分离卷积核与空间扩张卷积的联合运用，解决了传统卷积对全局空间结构建模能力不足的问题，进一步增强了 YOLOv5 对局部细节和远距离空间依赖的感知能力。如

图3所示,该模块将大核卷积分解为深度卷积、深度扩张卷积和 1×1 卷积,将LKA (large kernel attention) 中的深度卷积层的2D卷积核分解为串联的水平和垂直的1D卷积核,这种分解方法使得可以直接在注意力模块中使用带有大卷积核的深度卷积层,而不需要额外的模块。LSKA 注意力机制在处理大型可分离核时显著减小了计算复杂度,可以有效提升目标检测模型的性能。

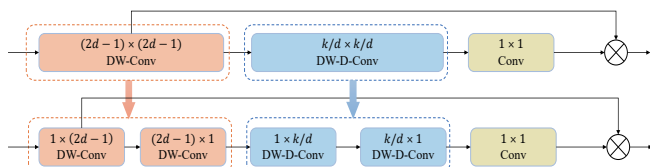


图3 LSKA 注意力结构

2.3 MobileNetV3 轻量化网络

MobileNetV3 的整体架构基本延续了 MobileNetV2 的设计理念,其主干网络由1个 3×3 卷积层和11个连续的bneck结构组成^[7]。引入轻量级的深度可分离卷积和残差块等结构,并在全连接层中增加了注意力机制模块,该网络在最后一步的平均池化层之前提前引入了一个HardSwish激活函数,并去除传统的最后卷积层,以减小模型的体积并降低计算成本,确保较高的检测准确性。

图4为深度可分离卷积分解为逐深度卷积与逐点卷积两个步骤的过程^[8]。

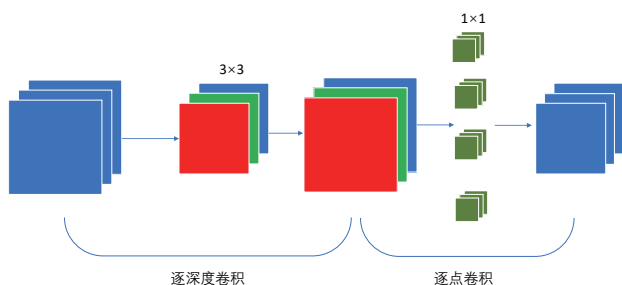


图4 深度可分离卷积

在逐深度卷积阶段,每个输入通道均分配一个独立的卷积核,这些卷积核分别在对应通道执行卷积操作。此阶段生成的输出特征图,其通道数与输入通道数相同,主要目的是提取空间特征。然后用 1×1 大小的卷积核对逐深度卷积的输出结果进行通道间的线性组合,以此提取通道特征。通过将传统的标准卷积分解为两个独立步骤,不仅显著减少计算量和参数量,还能有效保持强大的特征提取能力。

在 MobileNetV3 网络中,融入 SE 模块,该模块由压缩 (Squeeze) 和激励 (Excitation) 两部分组成。首先通过全局平均池化操作,将每个通道的特征图压缩为一个全局平均值,从而提取全局空间信息。然后利用两个全连接层学习通道间的相关性,并生成每个通道的加权系数。通过第一个

全连接层对通道进行压缩,然后用 ReLU 激活函数增加非线性表达能力,再通过第二个全连接层还原通道维度,使用 hardsigmoid 激活函数生成加权系数。如图5所示,虽然这一机制在一定程度上增加了参数量,但显著提高了模型在图像分类和目标检测任务中的性能,整体上提升了精度。

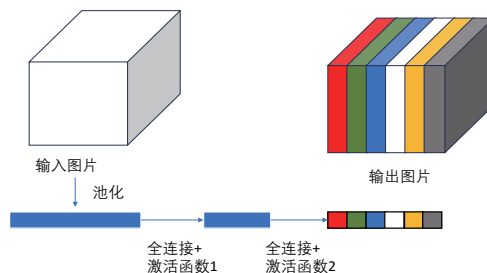


图5 SE 模块

2.4 CA 模块

由于 MobileNetV3 网络参数量小,可能导致模型精度降低和 mAP 值损失。为提升模型性能,本研究将着重提高精度,平衡轻量化结构带来的精度下降。为减少改进模型下采样过程中的信息丢失,提出用 CA (Coordinate Attention) 模块^[9]替换网络中原 MobileNetV3 的第一层卷积,以捕获更丰富的图像信息。CA 模块旨在不增加计算量的情况下增强特征表达能力。图6为 CA 注意力机制结构。

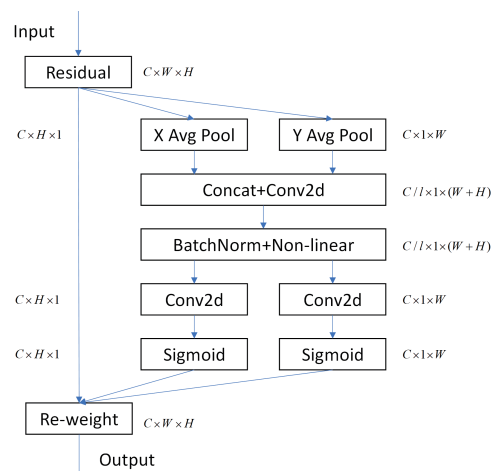


图6 CA 注意力机制结构

2.5 EIoU 损失函数

针对目标检测任务,传统的YOLOv5s算法采用CIoU作为边界框回归的损失函数,CIoU在DIoU基础上,再引入长宽比相关的惩罚项,以此提升定位的准确性,但CIoU计算复杂且成本高。

为优化计算效率,本文采用EIoU作为边界框回归损失函数。在CIoU的惩罚项基础上,EIoU分别针对预测框与真实框的宽度和高度计算损失,并融入了Focal机制以聚焦于高质量的锚框,从而更精确地评估框的尺寸差异^[10]。这一方法

不仅增强了模型对目标框尺寸变化的敏感度，而且保持了计算的简洁与高效，有效解决了 CIoU 存在的问题。EIoU 的计算公式为：

$$L_{\text{EIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b^{\text{gt}})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{\text{gt}})}{(w^c)^2} + \frac{\rho^2(h, h^{\text{gt}})}{(h^c)^2} \quad (1)$$

式中： w^c 和 h^c 表示预测框和真实框的最小外接框的宽和高。

3 实验与分析

3.1 数据集

针对 BVIP 出行场景中导向信息数据稀缺的问题，本文对实际出行场景进行了详细调查，分析了 9 个关键导向信息类别：盲道、斑马线、红灯、绿灯、树木、障碍物、汽车、行人和台阶。为此，采集了 24 段涵盖多种场景的视频，按帧率裁剪得到 1 886 张图像，并通过数据增强技术扩充至 4 380 张，以丰富数据集。数据集如图 7 所示。



图 7 实验数据集

3.2 实验设置与评价指标

所有实验均在 Windows 10 操作系统环境下进行，硬件平台为 NVIDIA GeForce RTX 4090 显卡，CUDA 版本为 12.2。语言环境为 Python 3.11，利用 PyTorch 深度学习框架实现。网络训练采用 SGD 优化算法，以最小化损失函数并优化模型权重。输入图像尺寸为 640 px×640 px，批量大小为 64，训练周期为 150，初始学习率为 0.01。

本文算法改进方向是提升复杂背景下多尺度目标检测精度，以及为移动设备实现做准备，需要提升模型检测性能、压缩网络模型尺寸，降低参数量和计算量，提高计算速度。因此算法性能评估主要以模型参数数量、计算复杂度、 F_1 -score，以及平均精度均值（mAP）等作为评价指标。

3.3 消融实验

本研究对 YOLOv5s 网络从 4 方面进行改进：一是将 MultiSEAM 和 LSKA 两个模块结合应用；二是将 YOLOv5s 的 Backbone 特征提取网络替换为轻量级神经网络模型 MobileNetV3；三是融入注意力机制 CA 模块；四是采用 EIoU 对损失函数进行优化。为验证这 4 种改进方法的有效性，基于原网络结构，分别将四种方法进行组合设置了 15 组实验，实验数据如表 1 所示。

表 1 消融实验结果

S/N	改进策略				F_1	mAP@0.5 /%
	MultiSEAM + LSKA	MobileNetV3	CA	EIoU		
1					0.44	51.2
2	√				0.58	59.6
3		√			0.47	49.2
4			√		0.46	57.4
5				√	0.45	57.2
6	√	√			0.56	55.4
7	√		√		0.58	60.7
8	√			√	0.59	61.9
9		√	√		0.45	55.0
10		√		√	0.46	54.3
11			√	√	0.49	57.7
12	√	√	√		0.57	60.4
13	√	√		√	0.57	64.2
14		√	√	√	0.54	56.9
15	√	√	√	√	0.63	63.7

实验数据显示，加入 MultiSEAM 和 LSKA 模块后，模型的 mAP@0.5% 提高了 8.4%，显著提升了检测精度。单独使用 MobileNetV3 进行轻量化导致 mAP 下降 2%，表明其结构可能降低精度。添加 CA 注意力模块和 EIoU 损失函数分别使 mAP 提高了 6.2% 和 6%，CA 模块增强了特征提取，EIoU 优化了定位精度。当 4 种改进结合使用时， F_1 因子提高了 0.19，mAP 提升了 12.5%，大幅提高了整体检测性能。

为验证各类损失函数的实验效果，将 EIoU 与 CIoU、SIoU、GIoU 以及 Focal-EIoU 损失函数进行对比实验分析，实验结果如表 2 所示。

表 2 损失函数对比实验结果

损失函数	F_1	mAP@0.5/%
CIoU	0.42	51.2
GIoU	0.44	51.9
SIoU	0.43	51.7
Focal-EIoU	0.45	57.1
EIoU	0.45	57.2

实验结果表明，与其他损失函数相比，EIoU 在整体检测性能上更具优势， F_1 因子为 0.45，mAP@0.5% 提升至 57.2%，证明了其在目标检测中的优越性和鲁棒性。

3.4 对比实验与分析

为了深入评估改进模型在目标检测任务中的性能表现，分别采用 YOLOv5s、YOLOv7、YOLOv7-tiny、YOLOv8、YOLOv9 以及 YOLO 系列中 SOTA 模型如 YOLOv10、YOLOv11 等与本文改进 YOLOv5s 进行对比实验，对比实验

的迭代次数与实验环境均相同。选取模型的网络层数、参数量、精度、计算复杂度（GFLOPs）、 F_1 值、mAP@0.5 作为各模型的评价指标，实验结果如表 3 所示。

表 3 YOLO 系列算法对比实验

Model	网络层数	参数量	P	复杂度/GFLOPs	F_1	mAP@0.5/%
YOLOv5s	270	7 043 902	0.508	16.0	0.44	51.2
YOLOv7	415	37 239 708	0.532	105.3	0.47	51.9
YOLOv7-tiny	263	6 036 636	0.513	13.2	0.47	51.6
YOLOv8n	225	3 012 603	0.541	8.2	0.49	52.9
YOLOv9s	486	7 170 571	0.555	26.4	0.52	53.8
YOLOv9t	486	1 972 539	0.592	7.7	0.53	57.6
YOLOv10n	285	2 697 926	0.574	6.7	0.57	58.9
YOLOv10s	293	8 041 926	0.598	21.6	0.58	58.7
YOLOv11n	238	2 583 907	0.617	6.5	0.61	58.9
YOLOv11s	238	9 416 283	0.622	9.4	0.61	61.2
改进 YOLOv5s	320	5 576 692	0.652	6.2	0.63	63.7

改进后的 YOLOv5s 模型在检测精度上显著优于 YOLOv10n 和 YOLOv11n，以及原始的 YOLOv5s。具体来说，改进模型的 P 值、 F_1 值和 mAP@0.5% 分别提升了 13.5%、10.5%、8.1% 和 5.6%、3.2%、8.1%，以及 14.4%、19% 和 12.5%。这些提升得益于 MultiSEAM 和 LSKA 模块的结合应用，增强了目标检测性能。同时，改进 YOLOv5s 在参数量和复杂度上分别减少了 20.8% 和 61.25%，实现了轻量化，提高了处理复杂目标检测任务时的鲁棒性。

使用改进 YOLOv5s 对导向信息检测效果如图 8 所示。



图 8 改进模型检测效果图

能够体现改进 YOLOv5s 在复杂干扰背景下对导向信息多尺度检测的能力，能够很好地处理小目标和被遮挡目标的检测问题，减少漏检的情况，适用于将改进后的模型部署在移动的导盲机器人中。

4 结论

本文提出一种面向导盲机器人导向信息多尺度轻量检测的算法。以提升 BVIP 的自主出行能力。构建了包含盲道、斑马线、红绿灯等 9 个类别的数据集，并从 4 个方面改进 YOLOv5s：结合 MultiSEAM 和 LSKA 模块、采用轻量级主

干网络、引入注意力机制和优化损失函数。改进后，模型在多尺度检测、复杂背景抑制和局部细节提取方面表现优异，同时实现轻量化。

然而，目前仍面临一些挑战，一是缺乏完善且多样化的数据集，影响模型泛化能力；二是摄像头镜头可能产生光斑或鬼影，影响检测效果。后续需进一步优化数据集和提升模型鲁棒性。

参考文献：

[1] WHO. Blindness and vision impairment [EB/OL]. [2024-08-10]. <https://www.who.int/zh/news-room/fact-sheets/detail/blindness-and-visual-impairment>.

[2] HONG B, LIN Z X, CHEN X, et al. Development and application of key technologies for guide dog robot: a systematic literature review[J]. Robotics and autonomous systems, 2022, 154: 104104.

[3] 武墨晗, 荣学文, 范永. 导盲机器人研究现状综述 [J]. 计算机工程与应用, 2020, 56(14): 1-13.

[4] 黄宁霞, 张荣芬, 刘宇红. 改进深度学习框架 Faster RCNN 的人行道障碍物目标检测 [J]. 机械设计与研究, 2021, 37(2): 7-12.

[5] 贾存南, 刘新春, 李昀骏, 等. 基于边缘计算和 YOLOv3 的盲道识别方法 [J]. 现代信息科技, 2023, 7(21): 102-105.

[6] SUGASHINI T, BALAKRISHNAN G. YOLO glass: video-based smart object detection using squeeze and attention YOLO network[J]. Signal, image and video processing, 2024, 18: 2105-2115.

[7] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3[EB/OL]. (2019-11-20)[2025-02-22]. <https://doi.org/10.48550/arXiv.1905.02244>.

[8] 李明, 何志奇, 党青霞, 等. 面向户外导盲场景的道路目标检测算法 [J]. 计算机工程与应用, 2025, 61(9): 242-254.

[9] CHENG W J, ZENG P F, HAO Y P. Lightweight YOLOv7 for bushing surface defects detection[J/OL]. Journal of real-time image processing, 2025[2025-06-10]. <https://link.springer.com/article/10.1007/s11554-025-01630-0>.

[10] 鲁玲, 李明良, 熊威, 等. 基于改进的 YOLOv8 变电设备红外图像检测 [J/OL]. 红外技术, 1-7[2025-05-14]. <http://kns.cnki.net/kcms/detail/53.1053.TN.20240508.1504.002.html>.

【作者简介】

王世刚（1970—），男，辽宁建平人，硕士，教授、硕士研究生导师，研究方向：机器人技术、智能控制与智能自动化、逆向工程、3D 打印等。

方欣怡（2000—），女，湖北荆门人，硕士研究生，研究方向：模式识别与智能系统。

（收稿日期：2025-02-19 修回日期：2025-08-29）