

DOI:10.19651/j.cnki.emt.2417408

基于改进 YOLOv8 的电梯内电动车检测算法

沙彦佑¹ 左官芳²

(1. 南京信息工程大学电子与信息工程学院 南京 210044; 2. 无锡学院电子信息工程学院 无锡 214105)

摘要: 针对 YOLOv8 算法在复杂场景下电梯内电动车检测精度的问题,提出了一种以 YOLOv8n 算法为基础改进的算法。首先,对于主干网络,将 C2f 模块与通用倒置瓶颈结构融合形成新的 C2f_ UIB 模块来优化计算效率,降低参数数量的同时提高全局信息捕获能力,同时在主干网络中添加空间和通道协同注意力模块 SCSA,提高特征提取能力和模型的鲁棒性。其次,用改进后的重参数化广义特征金字塔网络 DSRepGFPN 对颈部网络进行重构,增强跨尺度特征融合能力,提升多尺度目标的检测效果并减小模型的计算复杂度。最后将原有的损失函数 CIOU 替换为 MPDIU,提高目标框的定位精度,特别是在光照变化和目标遮挡场景中表现出更强的定位与识别能力。实验结果表明,在电梯内电动车数据集中,相较于 YOLOv8n,本文所改进的 YOLOv8-UAR 算法在 mAP50 上提高了 2.5%,在 mAP50-95 上提高了 1.8%,同时检测速度达到 94 fps,方便部署在边缘设备上,且更符合电动车进电梯检测的实际应用要求。

关键词: YOLOv8; 倒置瓶颈结构; 注意力机制; 重参数化特征金字塔; MPDIU; 电动车检测

中图分类号: TP391.4; TN98 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Electric vehicle detection algorithm in elevators based on the improved YOLOv8

Sha Yanyou¹ Zuo Guanfang²

(1. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China;

2. School of Electronic Information Engineering, Wuxi University, Wuxi 214105, China)

Abstract: An improved algorithm based on the YOLOv8n model is proposed to address the challenge of detection accuracy for electric vehicles entering elevators in complex scenarios. First, the backbone network integrates the C2f module with the universal inverted bottleneck structure to form a new C2f_ UIB module, optimizing computational efficiency and reducing parameters while improving global information capture capability. Additionally, a spatial-channel synergistic attention (SCSA) module is added to the backbone to enhance feature extraction and model robustness. Second, the neck network is reconstructed using an improved re-parameterized generalized feature pyramid network (DSRepGFPN), which enhances cross-scale feature fusion, improves multi-scale object detection performance, and reduces computational complexity. Finally, the original CIOU loss function is replaced with MPDIU, improving bounding box localization accuracy, especially in scenarios with lighting variations and object occlusions. Experimental results on an elevator electric vehicle dataset show that the proposed YOLOv8-UAR algorithm achieves a 2.5% improvement in mAP50 and a 1.8% improvement in mAP50-95 compared to YOLOv8n, with a detection speed of 94 fps. This makes it suitable for deployment on edge devices, aligning better with practical requirements for electric vehicle elevator detection.

Keywords: YOLOv8; inverted bottleneck structure; attention mechanism; re-parameterized feature pyramid; MPDIU; electric vehicle detection

0 引言

我国是电动车生产和消费大国,根据工信部统计数据,目前我国电动车保有量已达 3.5 亿辆。在中高层小区中,许多人将电动车推进电梯上楼进行充电,这种行为极易引

发火灾,根据国家消防救援局统计显示,2023 年全国共接报电动自行车火灾 2.1 万起^[1]。目前,大多数小区采用人工实时监控方式来防止电动车进入电梯,耗费巨大的人力和精力且容易造成目标遗漏。因此,对电动车进电梯的有效检测成为重点,基于深度学习的目标检测方法为此提供

了有效的解决方案^[2]。

目前深度学习目标检测的算法不断发展,主要分为两阶段算法和单阶段算法。两阶段算法主要以 Girshick 等^[3]提出的区域卷积神经网络(region convolutional neural networks, R-CNN)算法,以及随后不断发展出的快速区域卷积神经网络(fast region-based convolutional neural network, Fast R-CNN)^[4]和更快区域卷积神经网络(faster region-based convolutional neural network, Faster R-CNN)^[5]算法为主。单阶段算法主要包括单阶段多框检测器^[6](single shot multibox detector, SSD)系列算法和 YOLO(you only look once)系列算法^[7-13]。章曙光等^[14]提出用 MobileNet(mobile neural network)网络改进 SSD 前端的 VGG16(visual geometry group 16-layer)网络,得到 MobileNet-SSD 网络用于检测电梯内电动车。杨献瑜^[15]提出用 GhostNet(ghost neural network)替换 YOLOv4 中的主干特征提取网络 CSPDarkNet53(cross stage partial darknet 53),同时在 GhostNet 和 PAnet(path aggregation network)之间引入 CBAM(convolutional block attention module)注意力机制来检测电梯内电动车。Wang 等^[16]用 GhostNet 来改进 YOLOv5s 的主干网络,同时将损失函数改为 CIOU(complete intersection over union)损失来检测电梯内电动车。路成龙等^[17]结合 AUGMIX(augmented mixing)图像增强技术,同时用变形卷积 DCNv3(deformable convolutional network version 3)和动态稀疏注意力机制改进 YOLOv8,虽然一定程度上增加电梯内电动车的检测精度,但是模型在面对大面积遮挡情况是效果较差且模型较大。Wang 等^[18]提出引入 SCConv(switchable convolution)改造主干网络的卷积,同时结合软池化和多尺度卷积来增强 YOLOv8 检测电梯内场景的性能,但是同样导致模型的体积较大。尽管上述研究通过引入轻量化网络、注意力机制及改进损失函数等方法提升目标检测精度,但这些方法普遍存在模型复杂度过高和复杂场景适应性差的问题。部分方法采用复杂网模块,导致参数量显著增加,难以部署于资源受限的边缘设备;同时,在遮挡、亮度突变及反光等复杂场景下,检测精度下降明显,影响实际应用效果。

为了提升 YOLOv8 算法在电梯内复杂场景中的检测精度,尤其是在遮挡、亮度突然变化以及金属和玻璃反光等情况下,同时为了方便边缘设备的部署,本研究提出一种基于改进 YOLOv8n 算法的电动车进电梯检测算法。包括如下贡献:

1) 利用 MobileNetv4^[19]中的通用倒置瓶颈结构(unified inverted bottleneck, UIB)设计 C2f_UIB 模块,利用深度可分离卷积与通道动态调整机制来降低参数量同时提高全局信息捕获能力。

2) 在主干网络中增加空间和通道协同注意力模块(spatial and channel synergistic attention, SCSA)^[20]来增强

在亮度突然变化和遮挡时的特征的提取能力。

3) 利用改进后的重参数化广义特征金字塔网络(reparameterized generalized feature pyramid network, RepGFPN)^[21],引入 Shuffle 操作与多分支融合策略来重构颈部网络,增强跨尺度特征融合能力,提高模型的检测效果。

4) 用基于最小点距离的交并比损失函数(minimum point distance intersection over union, MPDIOU)^[22]替换 CIOU,提高光照变化和目標遮挡的情况下提升了定位与识别能力。

1 YOLOv8 算法简述

YOLOv8 的网络结构继承并发展了 YOLO 系列的设计理念,主要由 Backbone、Neck、Head 3 部分组成^[23]。

Backbone 的作用是从输入图像中提取多尺度的特征图,通过残差结构、注意力机制等方式来提升对细节和全局信息的捕捉能力。Neck 部分负责融合不同尺度的特征图。这一步是为了更好地捕捉不同尺度物体的信息,使网络能够更有效地处理多尺度目标检测任务。Head 部分主要负责对特征图进行目标分类和边界框回归。YOLOv8 的网络结构如图 1 所示^[24]。

2 改进 YOLOv8 算法

为了提高对电梯内的电动车检测精度和模型大小问题,本研究提出一种改进的 YOLOv8n 的模型,首先,利用 MobileNetv4 中的通用倒置瓶颈结构 UIB 设计 C2f_UIB 模块,减少模型参数量的同时提升全局信息的捕获能力;其次,在主干网络中引入空间和通道协同注意力模块 SCSA,增强亮度变化场景下的特征提取能力;然后,通过改进的重参数化广义特征金字塔网络 DSRepGFPN 重构颈部网络,增强跨尺度特征融合能力,提高模型的检测效果;最后,用基于最小点距离的交并比损失函数 MPDIOU 替换 CIOU,有效提升了光照变化和目標遮挡情况下的目标定位与识别能力。改进后的 YOLOv8n 网络结构如图 2 所示。

2.1 C2f_UIB 模块设计

YOLOv8 中 C2f(concatenate and fuse)模块通道特征提取机制相对简单,无法动态关注重要的通道特征,导致对目标的检测能力不足,且在面对光照突然变化和遮挡场景时,无法充分捕捉全局信息,同时 C2f 模块的参数量和计算复杂度相对较高,不利于在资源受限的边缘设备上部署。考虑到以上这些情况,利用 MobileNetv4 中提出的 UIB 结构设计一种新的 C2f_UIB 结构, C2f_UIB 的倒置瓶颈结构强化了通道特征提取,能够动态调整通道间的权重,更好地捕捉关键特征,适应遮挡及光照变化场景。同时 C2f_UIB 中引入了扩展卷积,能够扩大感受野,提高全局信息捕获能力。此外 C2f_UIB 保持高效特征提取能力的同时显著减少了参数量和计算复杂度,更适合低功耗边缘设备上部署。通用倒置瓶颈结构如图 3 所示。

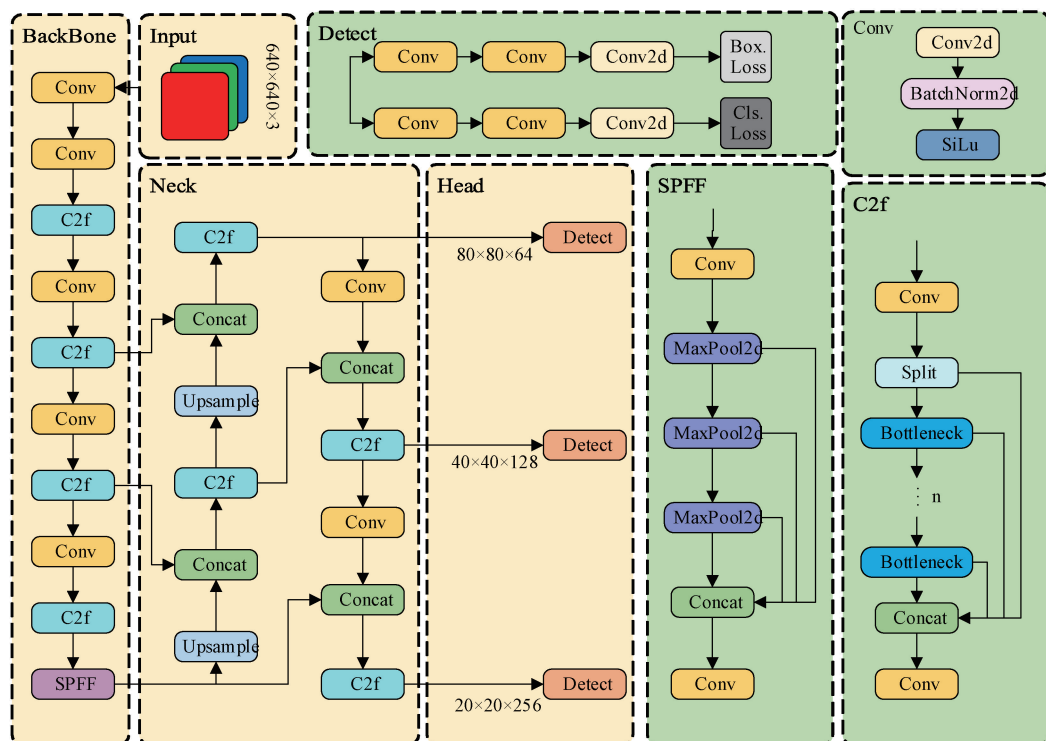


图1 YOLOv8网络结构

Fig. 1 YOLOv8 network architecture

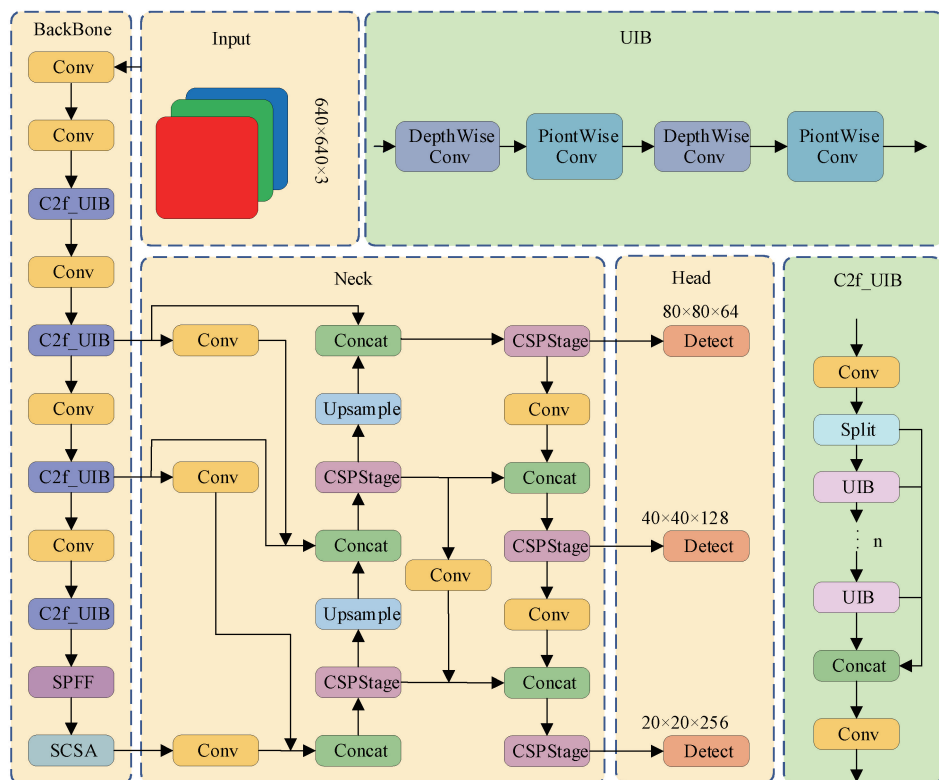


图2 改进的YOLOv8网络结构

Fig. 2 Improved YOLOv8 network architecture

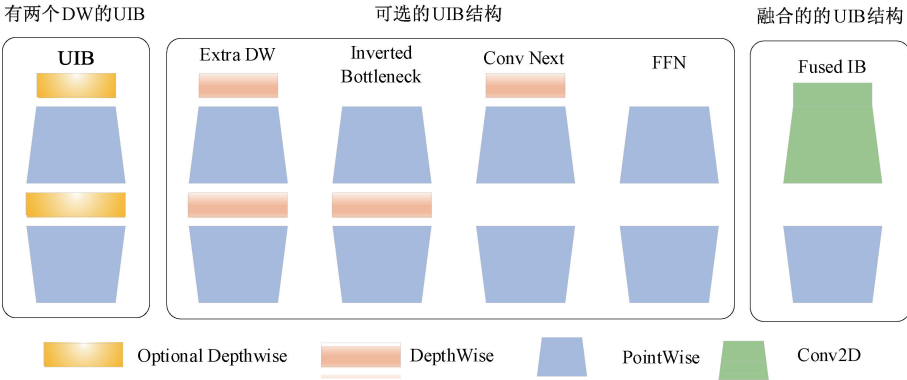


图 3 通用倒置瓶颈结构
Fig. 3 Universal inverted bottleneck

2.2 SCSA 注意力机制

由于电梯内空间狭小,且大多只有一个摄像头且位置固定不变,这就会大大增加目标被遮挡的概率,特别是当电梯内反光严重时,会导致检测目标的特征与背景环境相似,增加检测的难度。为了提高模型在这种情况下检测精度,在主干网络中添加空间和通道协同注意力 SCSA。SCSA 通过结合可共享多语义空间注意力 SMSA (shareable multi-semantic spatial attention)和渐进式通道

自注意力 PCSA (progressive channel-wise self-attention),实现了空间和通道特征的协同增强。具体而言,SMSA 通过捕获局部和全局特征,有效提取复杂场景中目标的空间信息,特别是在光照变化、金属反光以及玻璃反射等情况下,显著提升了模型对目标区域的关注能力。PCSA 则通过动态调整通道权重,强化关键目标的特征表达能力,同时减少冗余信息的干扰,增强了模型的鲁棒性。空间和通道协同注意力机制结构如图 4 所示。

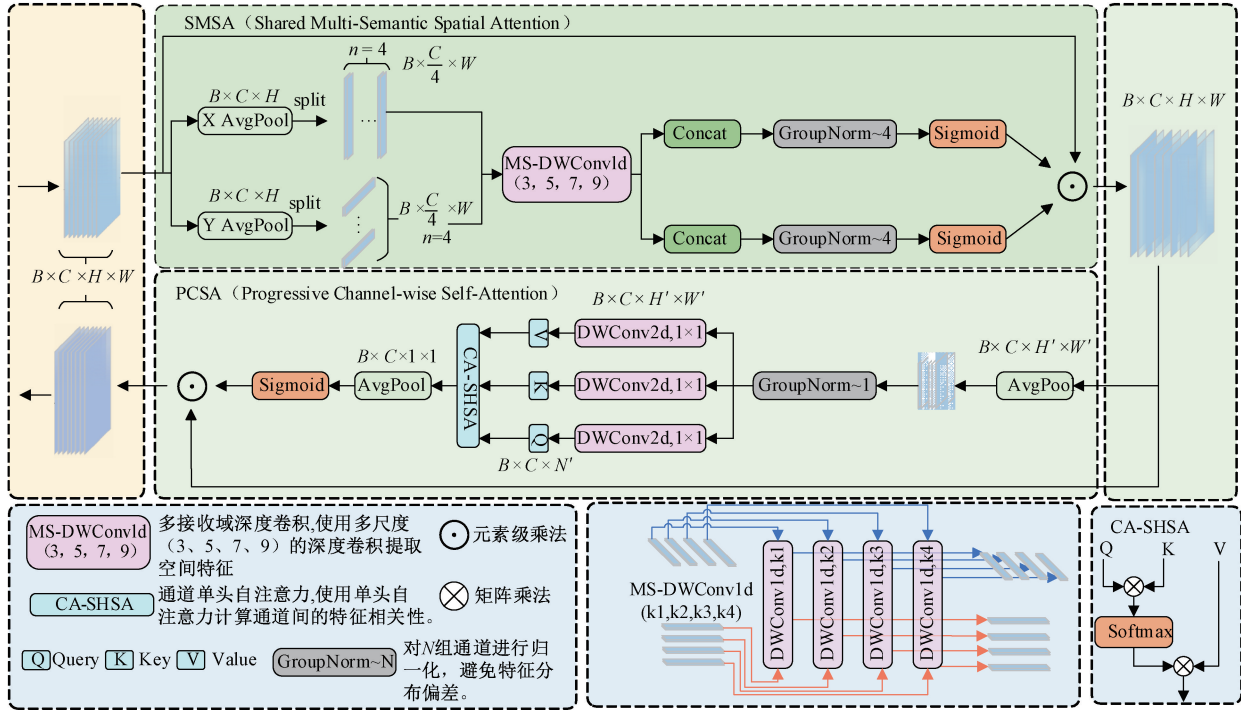


图 4 空间和通道协同注意力机制结构

Fig. 4 Spatial and channel synergistic attention architecture

2.3 DSRepGFPN 颈部网络设计

原 YOLOv8 的颈部网络对目标特征的细粒度表达能力有限,尤其在光照变化、目标遮挡和背景复杂的情况下,会出现漏检或误检。且原 YOLOv8 的颈部网络通过单一

路径和简单的上采用与下采样融合不同尺度的特征,缺乏灵活性。因此用重参数化广义特征金字塔网络 RepGFPN 来重构颈部网络,RepGFPN 通过对不同尺度的特征使用不同的通道数来高效融合这些尺度的特征,在训练和推理

阶段使用不同的 Rep, 减少推理时的计算量, 且 Rep 模块能在结构中引入更大的感受野, 通过多层 3×3 卷积捕获全局上下文信息, 使得跨尺度特征融合更高效。

为了解决原 RepGFPN 结构存在特征的利用效率低和跨通道信息交互不足的问题, 对 RepGFPN 中的 CSPStage (cross stage partial stage) 结构进行改进, 在第 1 次 Concat 操作后额外增加一个分支, 通过 1×1 卷积和深度可分离卷积 (depthwise convolution, DWConv) 来提取特征与另外两个之路进行 Concat 操作。在第 2 次 Concat 后增加一个 Shuffle

操作, 通过重新排列通道来增强不同通道之间的交互, 且可以有效缓解特征之间的冗余, 提高计算效率和特征利用率, 进一步优化了特征表达。改进后的 DSRepGFPN 网络能够更好的提升了模型对多尺度目标的检测效果, 尤其在复杂背景场景中表现出色。同时能够充分利用多尺度特征进行信息对齐, 减少漏检率, 提升定位精度, 在亮度变化和目标遮挡场景中具有显著优势。重参数化广义特征金字塔网络 RepGFPN 结构如图 5 所示。原始的 CSPStage 如图 6 所示, 改进后的 CSPStage 如图 7 所示。

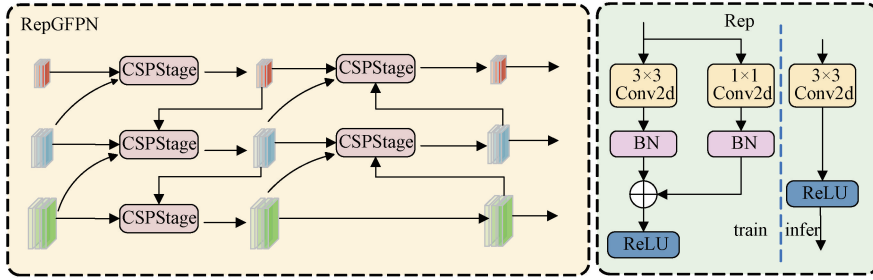


图 5 重参数化广义特征金字塔网络结构

Fig. 5 Reparameterized generalized feature pyramid network architecture

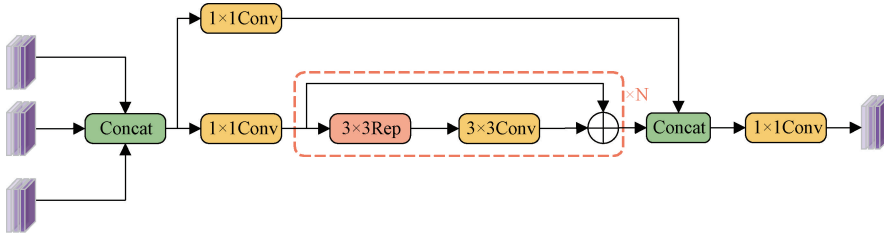


图 6 原跨阶段部分网络结构

Fig. 6 Original cross stage partial stage network

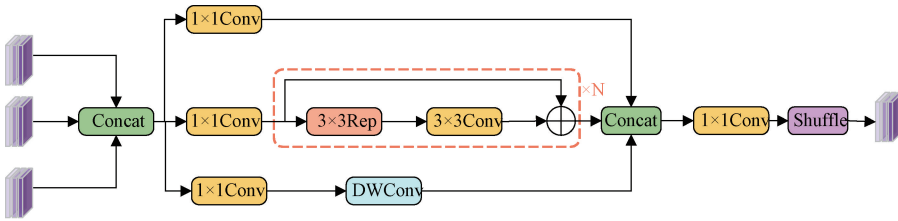


图 7 改进后的跨阶段部分网络结构

Fig. 7 Improved cross stage partial stage network

2.4 损失函数改进

YOLOv8 使用 CIoU 作为损失函数, 通过计算真实框和预测框之间的重叠与非重叠面积、两个中心点之间的距离、预测框与真实框的长和宽的比例来计算损失, 公式为:

$$\mathcal{L}_{CIoU} = 1 - IOU + \frac{\rho^2(B_{gt}, B_{pred})}{C^2} + \alpha\nu \quad (1)$$

$$IOU = \frac{B_{gt} \cap B_{pred}}{B_{gt} \cup B_{pred}} \quad (2)$$

$$\nu = \frac{4}{\pi^2} (\arctan \frac{w_{gt}}{h_{gt}} - \arctan \frac{w_{pred}}{h_{pred}})^2 \quad (3)$$

$$\alpha = \frac{\nu}{(1 - IOU) + \nu} \quad (4)$$

其中, B_{gt} , B_{pred} 分别表示真实框和预测框, IOU 是预测框和真实框的交集面积与并集面积的比值, $\rho^2(B_{gt}, B_{pred})$ 表示预测框中心点与真实框中心点之间的欧氏距离, ν 是宽高比差异项, α 是宽高比惩罚系数。在大多数情况下 CIoU 已经能很好的计算损失, 但是当遇到预测框和真实框的中心点一样, 且预测框与真实框的长宽比一致但是值不一致时就会导致损失函数不会充分表达他们之间的损失关系, 这会影响模型的精度和收敛速度。为了充分

考虑边框回归的几何特性,使用 MPDIoU 来替换 CIoU。MPDIoU 公式如下:

$$\mathcal{L}_{\text{MPDIoU}} = 1 - \text{IoU} + \frac{d_1^2}{h^2 + w^2} + \frac{d_2^2}{h^2 + w^2} \quad (5)$$

$$d_1^2 = (x_1^{\text{pred}} + x_1^{\text{gt}})^2 + (y_1^{\text{pred}} + y_1^{\text{gt}})^2 \quad (6)$$

$$d_2^2 = (x_2^{\text{pred}} + x_2^{\text{gt}})^2 + (y_2^{\text{pred}} + y_2^{\text{gt}})^2 \quad (7)$$

其中, $(x_1^{\text{pred}} + y_1^{\text{pred}})$ 表示预测框的左上角, $(x_2^{\text{pred}} + y_2^{\text{pred}})$ 表示预测框的左下角, 同理 $(x_1^{\text{gt}} + y_1^{\text{gt}})$ 和 $(x_2^{\text{gt}} + y_2^{\text{gt}})$ 分别表示真实框的左上角和左下角。

MPDIoU 通过最小化预测框与真实框关键点的欧氏距离, 同时综合考虑边界对齐和框内对齐, 提升了回归任务对框边界的精确性。在复杂目标场景中, MPDIoU 能够显著提升框的几何对齐能力, 边界拟合更精确, 同时能够减少对交并比的依赖, 在遮挡或亮度变化的情况下仍能对目标框进行有效的调整。因此, 使用 MPDIoU 替换掉原来的 CIoU。MPDIoU 如图 8 所示。

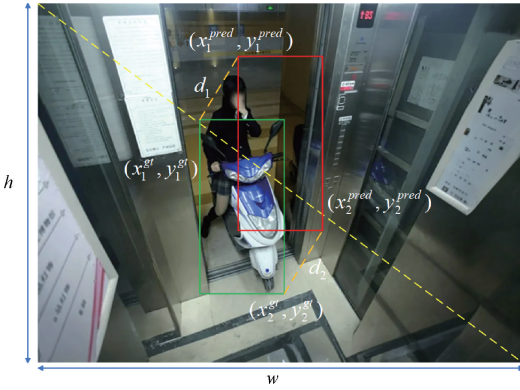


图 8 MPDIoU 示意图

Fig. 8 MPDIoU diagram

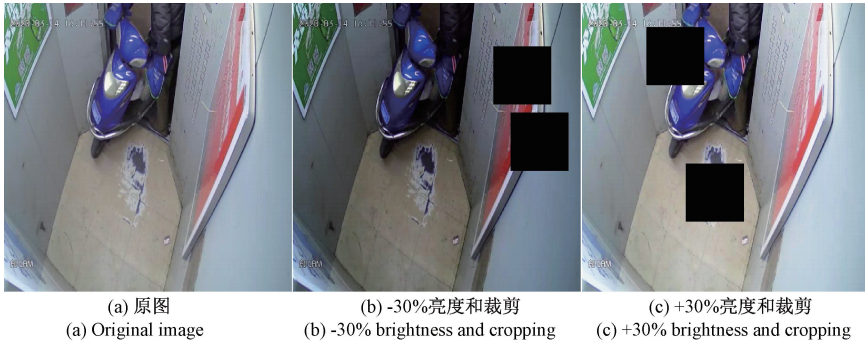


图 9 数据集增强结果

Fig. 9 Dataset augmentation results

和算法占用内存大小(Params)也被用作评价标准。计算公式分别见式(8)~(11)所示。

$$P = \frac{TP}{TP + FP} \times 100\% \quad (8)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (9)$$

3 实验结果分析

3.1 实验环境和参数

本研究所使用的实验的环境为 ubuntu20.01、python3.8、pytorch 为 2.0.0、GPU 为 RTX3090(24 GB)、CPU 为 14 vCPU Intel(R) Xeon(R) Platinum 8362 CPU @ 2.80 GHz、CUDA 为 11.8。训练参数设置为批量大小为 32、动量参数为 0.937、权重衰减率为 0.000 5、初始学习率为 0.01、优化器为 SGD, 其他未列出来的参数均与 YOLOv8 算法的基本配置相同, 训练 150 次。

3.2 数据集

因网上相关的高质量数据集资料较少, 选择自建数据集, 通过网上爬取、视频截图、现实拍摄等方式搜集原始图片 2 275 张, 为了防止将自行车误判为电动车, 所以本数据集增加一个自行车的类别, 共 3 个类别: 电动车、自行车、人。为了应对电梯开关门时可能会出现亮度差异, 满足实时检测的需求, 对数据集进行 -30% 和 +30% 的亮度增强处理, 同时为了更好的适应被遮挡情况下的检测, 对数据集做 2 个 20% 大小的裁剪处理, 增加模型的鲁棒性。最终数据集扩充至 5 461 张图片, 并按照 7:2:1 的比例划分为训练集、验证集和测试集。处理后的图片如图 9 所示, 图 9(a)为原图, 图 9(b)为 -30% 亮度和 2 个 20% 大小的裁剪处理后的图, 图 9(c)为 +30% 亮度和 2 个 20% 大小的裁剪处理后的图。

3.3 评价指标与结果

本文采用以下评价指标: 准确率(precision, P)、召回率(recall, R)、平均精度(average precision, AP)、平均精度均值(mean average precision, mAP)。同时, 计算复杂度(FLOPs)

$$AP = \int_0^1 P(R) dR \times 100\% \quad (10)$$

$$mAP = \frac{\sum_{i=1}^k AP}{k} \quad (11)$$

其中, TP(true positive)表示算法正确预测为正例的

样本数;FP(false positive)表示算法错误预测为正例的样本数;FN(false negative)表示算法错误预测为负例的样本数;TN(true negative)表示算法正确预测为负例的样本。

3.4 不同注意力对比实验

为了进一步验证本文增加的注意力机制的有效性,与其他多种注意力机制进行对比,如SE(squeeze-and-excitation)^[25]、GGA(global group attention)^[26]、LSKA(large-scale kernel attention)^[27]、MSDA(multi-scale dynamic attention)^[28]、DAT(dynamic attention transformer)^[29]、MLCA(multi-level channel attention)^[30]。通过与其他注意力模块的对比,明确展示了SCSA注意力机制在该场景下的优越性。实验结果如表1所示。从表1中可以得知,添加了SCSA注意力机制的YOLO-v8与基线YOLOv8n相比,在仅增加0.2GFLOPs的情况下,mAP50提高1.3%,mAP50-95提高0.9%。SCSA注意力机制与其他注意力机制相比,mAP是提高最多的,虽然与MSDA注意力在mAP指标上相等,但是在Params和FLOPs两个指标上少于MSDA,更能够方便的部署在边缘设备上,更符合本文改进模型的实际应用。

表1 不同注意力对比实验

Table 1 Comparison of different attention mechanisms				
模型	mAP50 /%	mAP50-95 /%	Params /10 ⁶	FLOPs /G
YOLOv8	83.8	64.3	3.0	8.1
+SE	83.9	64.4	3.0	8.2
+GGA	84.2	64.7	3.0	8.4
+LSKA	84.1	64.9	3.0	8.3
+MSDA	85.1	65.2	3.3	8.8
+DAT	84.8	64.8	3.3	8.4
+MLCA	84.3	64.7	3.1	8.4
+SCSA	85.1	65.2	3.0	8.3

3.5 不同IoU对比实验

为了验证本文使用的MPDIoU损失函数的有效性,本

文还与其他损失函数如IoU(intersection over union)、DIoU(distance intersection over union)、GIoU(generalized intersection over union)、CIoU、EIoU(enhanced intersection over union)进行对比^[31],实验结果如表2所示。从表2中可以得知,使用MPDIoU改进后的YOLOv8的精度与其他IoU相比更高,与使用CIoU的原YOLOv8相比,在mAP50上提升0.7%,在mAP50-95上提升0.4%,说明在该场景下使用MPDIoU能够更好的预测边框损失,使模型预测更加精准。

表2 不同IoU对比实验

Table 2 Comparison of different IoU metrics		
模型	mAP50/%	mAP50-95/%
YOLOv8n+IoU	83.0	63.9
YOLOv8n+DIoU	83.5	64.1
YOLOv8n+GIoU	83.3	64.0
YOLOv8n+CIoU	83.8	64.3
YOLOv8n+EIoU	84.2	64.4
YOLOv8n+MPDIoU	84.5	64.7

3.6 消融实验

为了验证本研究算法改进的有效性及各模块改进的效果,进行了消融实验,实验平台和参数与3.1节一致,实验结果如表3所示。从表中可以看出,使用UIB倒瓶颈结构改进后的C2f能够在不降低精度的同时降低模型10%的参数量和7.4%的浮点数,并且FPS(frames per second)能够达到110,有效提高了检测速度。增加了SCSA注意力机制后,能够更好的处理遮挡情况,mAP50和mAP50-95分别提升1.3%和0.9%。在使用DSRepGFPN重构颈部网络后,尽管参数量和浮点运算数多于YOLOv8n,但是mAP50和mAP50-95分别提升1.8%和1.4%,说明在面对亮度变化和遮挡时能够达到更高的检测效果。使用MPDIoU作为损失函数后,对边框损失预测更加精确,mAP50和mAP50-95分别提升0.7%和0.4%。综合使用C2f_UIB模块、SCSA注意力机制、RepGFPN结构和

表3 改进消融实验对比表

Table 3 Comparison of improved ablation experiments								
C2f_UIB	SCSA	DSRepGFPN	MPDIoU	mAP50/%	mAP50-95/%	Params/10 ⁶	FLOPs/G	FPS/fps
—	—	—	—	83.8	64.3	3.0	8.1	96
√	—	—	—	83.9	64.3	2.7	7.5	110
—	√	—	—	85.1	65.2	3.0	8.3	92
—	—	√	—	85.6	65.7	3.2	8.6	87
—	—	—	√	84.5	64.7	3.0	8.1	95
√	√	—	—	85.4	65.5	2.8	7.8	106
√	√	√	—	85.8	65.8	3.0	8.2	96
√	√	√	√	86.3	66.1	3.0	8.2	94

MPDIoU 损失函数,在训练 150 轮后,在仅增加 0.1 GFLOPs 的情况下,mAP50 和 mAP50-95 分别提升 2.5% 和 1.8%,mAP50 如图 10(a)所示,mAP50-95 如图 10(b)。

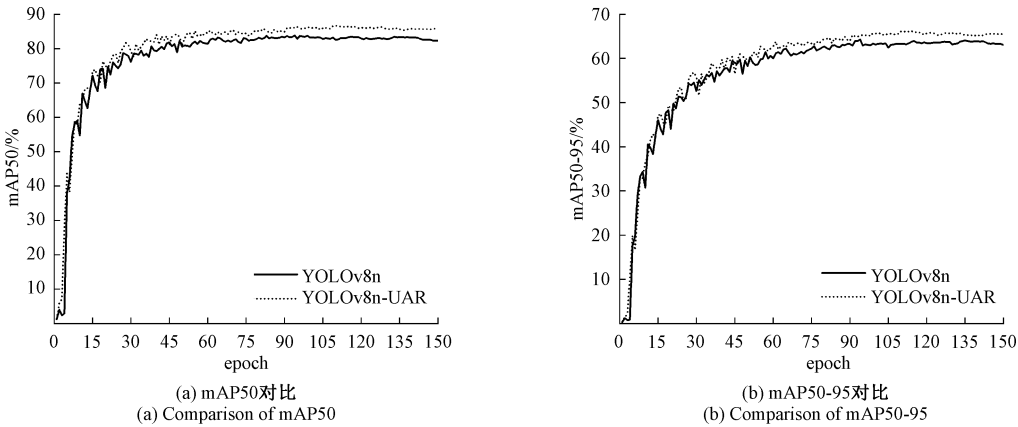


图 10 模型改进前后 mAP 曲线对比
Fig. 10 Comparison of mAP curves before and after model improvement

3.7 对比实验

为了进一步验证本文所改进 YOLOv8 模型的性能,与多种模型进行比较,结果如表 4 所示。

可以看出,本研究提出的 YOLOv8-UAR 算法与传统算法如 YOLOv5s、YOLOv6、YOLOv7tiny、YOLOv8n 和 YOLOv8s 相比,YOLOv8-UAR 在 mAP50 上分别提高 7.7%、7.2%、7%、2.5%、2.9%,在 mAP50-95 上分别提高 5.3%、4.2%、3.7%、1.8%、2%。与改进后的 YOLO 系列算法(文献[15-18])在本文数据集上的实验对比,YOLOv8-UAR 在 mAP50 上分别提高 9.8%、7.9%、0.7%、0.2%,在 mAP50-95 上分别提高 7.5%、6.5%、0.8%、0.3%。这些性能提升得益于本文在模型设计中的多项改进,包括轻量化的 C2f_UIB 模块,通过通道动态分配和深度可分离卷积减少计算冗余并优化特征表达;SCSA 注意力机制实现空间与通道特征的协同增强,在遮挡、亮度变化及反光等复杂场景下有效提升了检测鲁棒性;以及改进的 DSRepGFPN 结构,增强了多尺度特征融

合能力,在复杂背景下取得了显著的检测性能提升,且 YOLOv8-UAR 参数量并没有增加,相比其他的改进算法更容易的部署在边缘设备上。

可视化检测结果对比图 11 所示,当电梯开关门导致亮度快速变化时,如图 11(a)和(b)所示,虽然 YOLOv8 能够检测并识别电动车,但出现了误判,且其精度相比 YOLOv8-UAR 较低。这是因为在电梯开关门时,电梯内狭小空间的亮度突然变化较大,增加了算法对目标识别的难度。相较而言,YOLOv8-UAR 具备更强的物体识别和定位能力,在这种情况下不会发生误判,且精度更高。当电梯内亮度稳定,但电动车部分被遮挡或其外形与周围环境颜色相似时,如图 11(c)和(d)所示,YOLOv8n 漏检了图中的电动车。这是因为部分遮挡物与周围环境或物体相似,或者物体被大面积遮挡,导致其原有部分特征消失,从而增加了检测和识别的难度。相比之下,YOLOv8-UAR 具有更强的特征提取能力,能够在物体被遮挡时准确识别出该物体。

表 4 不同算法实验对比结果

Table 4 Comparison results of different algorithms in experiments					
模型	mAP50/%	mAP50-95/%	Params/ 10^6	FLOPs/G	FPS/fps
文献[15]	76.5	58.6	11.0	22.3	58
YOLOv5s	78.6	60.8	2.5	7.2	112
文献[16]	78.4	59.6	2.8	8.4	90
YOLOv6	79.1	61.9	4.2	11.9	68
YOLOv7tiny	79.3	62.4	6.3	13.2	76
YOLOv8n	83.8	64.3	3.0	8.1	96
YOLOv8s	83.4	64.1	11.1	28.7	52
文献[17]	85.6	65.3	3.6	9.4	82
文献[18]	86.1	65.8	3.2	8.2	90
YOLOv8-UAR(本文)	86.3	66.1	3.0	8.2	94

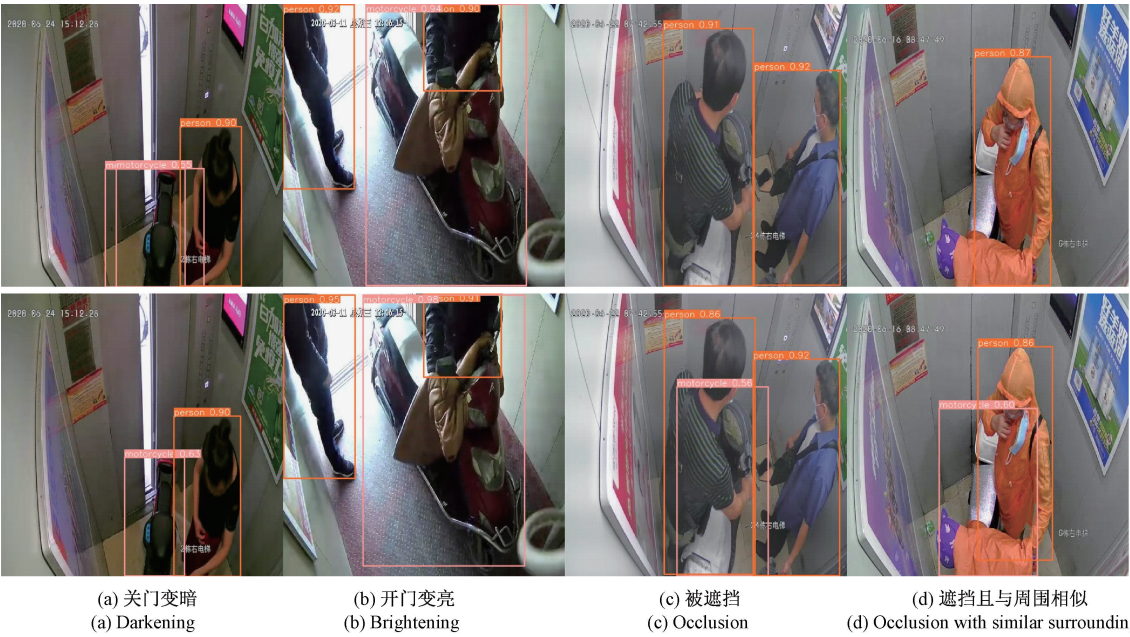


图 11 检测结果对比

Fig. 11 Comparison of detection results

3.8 YOLOV8-UAR 模型泛化性验证

为了进一步验证 YOLOV8-UAR 模型的泛化性能,本研究选择在公开数据集 MobilityDevices 数据集 Motorcar 数据集和 COCO2017 数据集上验证本研究算法。Mobility-Devices 数据集是在黑夜、雨雾天等更加复杂环境下拍摄的图片,有 15 546 张图片,类别有 bicycle、car、bus、motorcycle、scooter、van 等多达 13 个车类别。Motorcar 数据集为多种场景下拍摄的图片,类别仅有电动车,共 26 452 张图片。此外,将 COCO2017 数据集进行筛选,只选择人和车相关的类别图片^[32]。实验环境与参数配置与 3.1 节保持一致,实验结果如表 5 所示。从表 4 中可以看出,在 MobilityDevices 数据集上 YOLOv8-UAR 的 mAP50、mAP50-95 分别提升 1.2%、0.7%,在 Motorcar 数据集上,YOLOv8-UAR 的 mAP50、mAP50-95 分别提升 3.2%、1.5%。在 COCO2017 车和人的相关类别上,YOLOv8-UAR 的 mAP50、mAP50-950 分别提升 2.3%、1.3%。从实验结果来看,本研究的 YOLOv8-UAR 算法在 Mobility-Devices 数据集上,面对雨雾天、黑夜等场景下检测效果相对差一些,原因可能是本研究没有针对雨雾天、黑夜等场景进行加强,但是本研究算法在指标上仍然比 YOLOv8n 算法表现的更好。此外,本研究改进的 YOLOv8-UAR 算法在 Motorcar 数据集和 COCO2017 数据集上比 YOLOv8n 表现的更加优秀。同时还与其他改进 YOLOv8 用于检测电动车的算法(文献[17-18])在这 3 个数据集上进行对比,结果显示,本研究改进的算法指标均是最优。进一步验证了本研究算法的检测能力和泛化性能,同时参数量没有增加,更符合实际的检测需求。

表 5 YOLOV8-UAR 在其他数据集上的表现

Table 5 Performance of YOLOV8-UAR on other datasets

数据集	算法	mAP50/%	mAP50-95/%
Mobility Devices	YOLOv8n	63.1	46.8
	文献[17]	62.8	46.2
	文献[18]	63.6	47.0
	YOLOv8-UAR	64.3	47.5
Motorcar	YOLOv8n	86.3	65.2
	文献[17]	87.3	65.8
	文献[18]	88.6	66.2
	YOLOv8-UAR	89.5	66.7
COCO2017	YOLOv8n	66.4	48.5
	文献[17]	67.2	48.9
	文献[18]	67.6	49.2
	YOLOv8-UAR	68.7	49.8

4 结 论

本研究提出一种基于改进 YOLOv8 的电动车进电梯检测算法,提高了检测性能和计算效率。首先,为了提高在电梯多种场景下的检测准确性,本研究对原有数据集进行曝光和裁剪处理,以增强算法的鲁棒性。然后,利用 MobileNetv4 中的通用倒置瓶颈结构设计 C2f_UIB 模块来降低参数量同时提高全局信息捕获能力。再在主干网络中增加空间和通道协同注意力机制 SCSA 来增强在亮度突然变化和遮挡时的特征的提取能力。其次,利用改进后的重参数化广义特征金字塔网络 DSRepGFPN 来重构颈

部网络,增强跨尺度特征融合能力,提高模型的检测效果。最后,用基于最小点距离的交并比损失函数 MPDIOW 替换 CIOU,提高光照变化和目标遮挡的情况下提升了定位与识别能力。最终试验结果表明,本文提出的 YOLOv8-UAR 的电动车进电梯检测中比原来 YOLOv8n 模型在 mAP50、mAP50-95 分别提升 2.5%、1.8%,且参数量没有增加,方便部署在边缘实时设备上,更加符合实际应用的需求。

参考文献

- [1] 孔德洪. 电动车火灾频发惨痛教训何时休? [J]. 防灾博览, 2024(2): 43-47.
KONG D Q. When will the painful lessons of frequent electric vehicle fires stop? [J]. Disaster Prevention and Mitigation, 2024(2): 43-47.
- [2] 陈琳, 陈传敬, 张鹏. 基于 YOLOv4-tiny 的电梯内电动车检测模型[J]. 价值工程, 2022, 41(12): 117-119.
CHEN L, CHEN CH J, ZHANG P. Electric vehicle detection model in elevators based on YOLOv4-tiny[J]. Value Engineering, 2022, 41(12): 117-119.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [4] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [5] REN SH Q, HE K M, 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, 2016, 39(6): 1137-1149.
- [6] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV 2016: 14th European Conference, Springer International Publishing, 2016: 21-37.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 779-788.
- [8] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [9] FARHADI A, REDMON J. YOLOv3: An incremental improvement[C]. Computer vision and pattern recognition. Berlin/Heidelberg, Germany: Springer, 2018, 1804: 1-6.
- [10] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. ArXiv preprint arXiv:2004.10934, 2020.
- [11] ZHU X K, LYU SH C, WANG X, et al. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]. IEEE/CVF International Conference on Computer Vision, 2021: 2778-2788.
- [12] LI C Y, LI L L, JIANG H L, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. ArXiv preprint arXiv:2209.02976, 2022.
- [13] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.
- [14] 章曙光, 张文韬, 刘洋, 等. 改进 MobileNet-SSD 的电梯内电动车识别方法[J]. 机械设计与制造, 2024(9): 340-345.
ZHANG SH G, ZHANG W T, LIU Y, et al. Improved MobileNet-SSD-based method for electric vehicle recognition in elevators[J]. Machinery Design & Manufacture, 2024(9): 340-345.
- [15] 杨献瑜. 基于改进 YOLOv4 的电梯内电动车检测算法[J]. 计算机时代, 2023(10): 54-58.
YANG X Y. Elevator electric vehicle detection algorithm based on improved YOLOv4[J]. Computer Era, 2023(10): 54-58.
- [16] WANG ZH H, HU CH P, LI J. Electric bicycle detection in elevator car based on YOLOv5[C]. 3rd International Conference on Artificial Intelligence, Automation, and High-Performance Computing (AIAHPC 2023). SPIE, 2023, 12717: 125-133.
- [17] 路成龙, 冯月贵, 庆光蔚. 基于改进 YOLOv8 的电梯内电动车识别方法研究[J]. 机械制造与自动化, 2024, 53(4): 219-223, 229.
LU CH L, FENG Y G, QING G W. Research on elevator electric vehicle recognition method based on improved YOLOv8[J]. Mechanical Manufacturing and Automation, 2024, 53(4): 219-223, 229.
- [18] WANG ZH H, CHEN J Y, YU P, et al. SC-YOLOv8 Network with soft-pooling and attention for elevator passenger detection[J]. Applied Sciences, 2024, 14(8): 3321.
- [19] QIN D F, LEICHNER C, DELAKIS M, et al. MobileNetV4: Universal models for the mobile ecosystem[C]. European Conference on Computer Vision. Springer, Cham, 2025: 78-96.
- [20] SI Y ZH, XU H Y, ZHU X ZH, et al. SCSA:

- Exploring the synergistic effects between spatial and channel attention [J]. ArXiv preprint arXiv:2407.05128, 2024.
- [21] XU X ZH, JIANG Y Q, CHEN W H, et al. Damo-YOLO: A report on real-time object detection design[J]. ArXiv preprint arXiv:2211.15444, 2022.
- [22] MA S L, XU Y. MPDIoU: A loss for efficient and accurate bounding box regression[J]. ArXiv preprint arXiv:2307.07662, 2023.
- [23] 周建新, 张媛, 贾梓涵, 等. 改进 YOLOv8 的路面缺陷检测算法 [J]. 电子测量技术, 2024, 47 (19): 146-154.
- ZHOU J X, ZHANG Y, JIA Z H, et al. Improved YOLOv8-based pavement defect detection algorithm [J]. Electronic Measurement Technology, 2024, 47(19): 146-154.
- [24] 张瑶, 陈姚节. 改进 YOLOv8 的水面小目标检测算法[J]. 计算机系统应用, 2024, 33(4): 152-161.
- ZHANG Y, CHEN Y J. Improved YOLOv8 small object detection algorithm for water surface [J]. Computer Systems Applications, 2024, 33 (4): 152-161.
- [25] 王军, 伍毅, 陈正超. 基于 SMT-YOLOv8 的 PCB 缺陷检测研究 [J]. 电子测量技术, 2024, 47 (11): 131-137.
- WANG J, WU Y, CHEN ZH CH. Research on PCB defect detection based on SMT-YOLOv8 [J]. Electronic Measurement Technology, 2024, 47(11): 131-137.
- [26] LIU X Y, PENG H W, ZHENG N X, et al. EfficientVit: Memory efficient vision transformer with cascaded group attention[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 14420-14430.
- [27] LAU K W, PO L M, REHMAN Y A U. Large separable kernel attention: Rethinking the large kernel attention design in CNN [J]. Expert Systems with Applications, 2024, 236: 121352.
- [28] JIAO J Y, TANG Y M, LIN K Y, et al. Dilateformer: Multi-scale dilated transformer for visual recognition [J]. IEEE Transactions on Multimedia, 2023, 25: 8906-8919.
- [29] XIA Z F, PAN X R, SONG S J, et al. Vision transformer with deformable attention [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 4794-4803.
- [30] WAND H, LU R SH, SHEN S Y, et al. Mixed local channel attention for object detection[J]. Engineering Applications of Artificial Intelligence, 2023, 123: 106442.
- [31] 霍旭, 盖绍彦, 洪濡, 等. 基于多阶段关联的多目标跟踪算法[J]. 仪器仪表学报, 2023, 44(11): 205-214.
- HUO X, GAI SH Y, HONG R, et al. Multi-target tracking algorithm based on multi-stage association[J]. Chinese Journal of Scientific Instrument, 2023, 44(11): 205-214.
- [32] 王娟敏, 皮建勇, 黄昆, 等. 面向复杂场景的多尺度行人和车辆检测算法 [J/OL]. 现代电子技术, 1-15 [2024-11-19]. <http://kns.cnki.net/kcms/detail/61.1224.tn.20240902.1325.002.html>.
- WANG J M, PI J Y, HUANG K, et al. Multi-scale pedestrian and vehicle detection algorithm for complex scenarios[J/OL]. Modern Electronic Technology, 1-15 [2024-11-19]. <http://kns.cnki.net/kcms/detail/61.1224.tn.20240902.1325.002.html>.

作者简介

沙彦佑, 硕士研究生, 主要研究方向为嵌入式、深度学习。
E-mail: 1819735834@qq.com

左官芳(通信作者), 硕士研究生, 副教授, 主要研究方向为嵌入式设计、深度学习。
E-mail: zgf@cw Xu.edu.cn