

模糊场景下行人与车辆检测算法^①

郑广海, 张海宁, 曲英伟

(大连交通大学 轨道智能工程学院, 大连 116028)

通信作者: 张海宁, E-mail: 2875616904@qq.com



摘要: 针对雾霾、雨雪等恶劣天气下拍摄到的图像退化模糊, 难以进行准确识别与检测的问题, 本文提出模糊场景下行人与车辆检测算法 LiteBlurVisionNet (轻模糊视觉网络). 在主干网络部分使用 global context enhancer 注意力改进轻量级 MobileNetV3 模块, 减少了参数量, 使得模型在雾霾、雨雪等恶劣天气条件下图像处理效率更高. 颈部网络采用更为轻量化的 Ghost 模块和由 Ghost bottleneck 模块改进的 spectral Ghost unit 模块, 能够更有效地捕获全局上下文信息, 提高特征的区分度和表达能力, 有助于减少参数数量和计算复杂度, 从而提高网络处理速度和效率; 预测部分采用 $DIoU$ NMS 基于非极大抑制方法进行最大局部搜索, 去除冗余的检测框, 提高检测算法在模糊场景下的准确性. 实验结果表明, LiteBlurVisionNet 算法模型的参数量比 RTDETR-ResNet50 算法模型下降了 96.8%, 比 YOLOv8n 算法模型下降了 55.5%, LiteBlurVisionNet 算法模型的计算量比 Faster R-CNN 算法模型下降了 99.9%, 比 YOLOv8n 算法模型下降了 57%, LiteBlurVisionNet 算法模型的 $mAP_{0.5}$ 比 IAL-YOLO 算法模型提高了 13.71%, 比 YOLOv5s 算法模型提高了 2.4%, 这意味着模型在存储和计算方面更加高效, 尤其适用于资源受限的环境或移动端设备.

关键词: MobileNetV3; 模糊场景; 人车检测; 轻量化; LiteBlurVisionNet

引用格式: 郑广海, 张海宁, 曲英伟. 模糊场景下行人与车辆检测算法. 计算机系统应用, 2025, 34(2): 145–153. <http://www.c-s-a.org.cn/1003-3254/9778.html>

Pedestrian and Vehicle Detection Algorithm in Blurred Scenarios

ZHENG Guang-Hai, ZHANG Hai-Ning, QU Ying-Wei

(School of Railway Engineering, Dalian Jiaotong University, Dalian 116028, China)

Abstract: Aiming at degraded and blurred images captured under harsh weather conditions such as haze, rain, and snow, which make accurate recognition and detection challenging, this study proposes a pedestrian and vehicle detection algorithm, lightweight blur vision network (LiteBlurVisionNet), for blurred scenes. In the backbone network, the global context enhancer attention-improved lightweight MobileNetV3 module is used, reducing the number of parameters and making the model more efficient in image processing under harsh weather conditions such as haze and rain. The neck network adopts a lighter Ghost module and the spectral ghost unit module improved from the Ghost bottleneck module. These modules can more effectively capture global context information, improve the discrimination and expressive ability of features, help reduce the number of parameters and computational complexity, and thereby improve the network's processing speed and efficiency. In the prediction part, $DIoU$ NMS based on the non-maximum suppression method is used for maximum local search to remove redundant detection boxes and improve the accuracy of the detection algorithm in blurred scenes. Experimental results show that the parameter count of the LiteBlurVisionNet algorithm model is

① 收稿时间: 2024-06-08; 修改时间: 2024-07-10, 2024-07-24; 采用时间: 2024-09-19; csa 在线出版时间: 2024-12-19

CNKI 网络首发时间: 2024-12-20

reduced by 96.8% compared to the RTDETR-ResNet50 algorithm model, and by 55.5% compared to the YOLOv8n algorithm model. The computational load of the LiteBlurVisionNet algorithm model is reduced by 99.9% compared to the Faster R-CNN algorithm model and by 57% compared to the YOLOv8n algorithm model. The *mAP*0.5 of the LiteBlurVisionNet algorithm model is improved by 13.71% compared to the IAL-YOLO algorithm model and by 2.4% compared to the YOLOv5s algorithm model. This means the model is more efficient in terms of storage and computation and is particularly suitable for resource-constrained environments or mobile devices.

Key words: MobileNetV3; blurred scenario; pedestrian and vehicle detection; lightweight; LiteBlurVisionNet

随着人工智能技术的快速发展,深度学习目标检测算法已在遥感^[1]、医学成像^[2]和无人驾驶^[3]等领域广泛应用。这些领域常需要实时反馈,因此对检测算法的精度和速度提出了极高要求。尽管目标检测算法已取得进步,但将模型部署到移动设备时,由于其相对较弱的计算能力,执行速度可能会受到影响。因此,轻量化研究对模型的部署至关重要。

在计算机视觉领域,不良天气条件会削弱摄像头拍摄图像的质量,造成误检和漏检,为解决此问题,本文提出模糊场景下行人与车辆检测算法 LiteBlurVision-Net(轻模糊视觉网络),旨在模糊场景中准确检测行人和车辆,以提高道路交通信息感知的精度,降低交通事故发生率。

目前,基于深度学习的目标检测算法主要可划分为双阶段和单阶段两类。双阶段算法将目标检测分解为提取候选区域和对其进行分类与精确定位两个阶段,Ren 等^[4]提出的 Faster R-CNN 将选择性搜索替换为 RPN 网络,以提高候选区域提取效率,但仍存在计算冗余。Cai 等^[5]提出的 Cascade R-CNN 通过多次修正预测区域以提高准确性。院老虎等^[6]通过将暗通道先验算法与 Faster R-CNN 结合,以提高雾天条件下的性能。尽管这些算法精度较高,但在雾天等复杂条件下仍无法满足实时检测需求,因为模型参数较多,计算时间较长。

单阶段算法结构相对简单,检测速度更快,更适用于复杂的交通场景。代表算法有 YOLO^[7], SSD^[8], Retina-Net^[9]等。目前,通过对 YOLO 系列检测算法进行改进,在复杂天气条件下的检测也取得了一定成果。蒋超等^[10]针对传统行人检测参数量较大和计算复杂度高的问题,通过对 YOLOv5 模型进行替换主干网络、加入注意力机制等方法,获得了一种轻量化的行人检测算法;刘书

刚等^[11]将图像去雾算法加入图像预处理,同时微调 YOLOv4 网络,使雾天场景的行人检测取得了较好的效果;院老虎等^[12]通过改进 YOLOv5 网络的图像增强技术,优化 NMS 非极大值抑制等方法,改进了雾天场景下的车辆检测效果。

然而,这些方法未考虑常规卷积难以捕获不同位置的信息差异,同时不良天气条件下目标特征易受干扰,导致网络难以提取有效特征,从而影响整体检测性能。受 YOLOv5 的启发,本文提出 LiteBlurVisionNet 模型,旨在模糊场景下进行行人和车辆检测,其算法流程如图 1 所示。

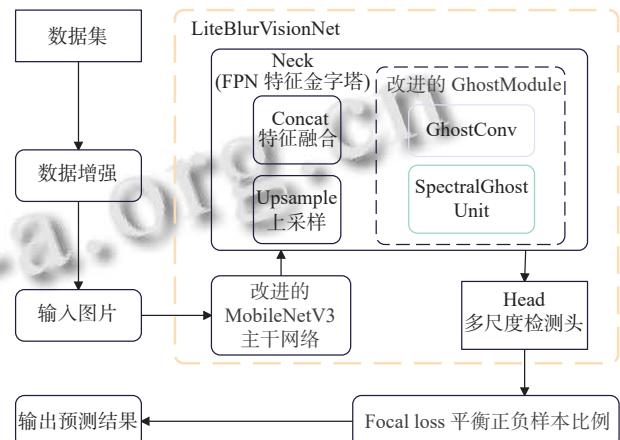


图 1 LiteBlurVisionNet 模型流程图

该框架的具体流程如下。

- (1) 准备好的数据集通过 Mosaic、随机缩放、裁剪、拼接等方式进行数据增强。
- (2) 将数据增强过后的图片输入 LiteBlurVisionNet 模型。
- (3) LiteBlurVisionNet 模型使用改进的 MobileNetV3 作为主干网络提取卷积的特征,即主干特征。
- (4) 把主干特征输入到颈部网络进行特征提取,其

中颈部网络包括 Concat 特征融合、Upsample 上采样以及改进的 Ghost 模块。

(5) 采用多尺度检测头检测特征提取后的内容,从而进行结果预测。

(6) 使用 Focal loss 平衡正负样本比例,计算负样本置信度,并进行梯度更新。

(7) 输出当前图像目标检测结果,即该图像中所有目标的分类准确度和定位框坐标信息。

LiteBlurVisionNet 模型有如下 3 点改进。

(1) 在模型设计方面,主干网络部分引入由 global context enhancer 注意力改进的 MobileNetV3 模块,减少了参数量,使得模型在处理模糊图片时能更高效,本文引入的全局上下文增强器(global context enhancer)模块,结合了 ECA (efficient channel attention)^[13] 和 GMA (global MLP attention) 注意力机制及通道洗牌操作,该模块同时结合通道级别和全局级别的注意力机制来增强网络的特征提取能力。该模块能够有针对性地关注感受野内不同特征的重要性,并通过通道洗牌增加特征的多样性和鲁棒性,进一步提高模型的性能。

(2) 在模型的颈部结构中,采用了轻量级网络 GhostNet^[14] 及由频谱注意力机制(ECA 注意力机制)改进的 Ghost bottleneck 模块^[14],被命名为 spectral Ghost unit 模块。相比于 Ghost bottleneck 模块, spectral Ghost block 模块引入了频谱注意力机制能更有效地捕获全局上下文信息,提高特征的区分度和表达能力,使模型能够更好地理解特征之间的关系。通过频谱注意力机制对特征进行增强和提升,有效地增强模型对特征之间关系的捕获,提升模型的性能,使得模型在相同参数量的情况下获得更好的性能,为各种视觉任务提供更好的性能表现。

(3) 针对雾天条件下目标遮挡导致的识别困难,本文采用 DIoU NMS^[15]作为后处理部分的方法,综合考虑了目标框之间的位置距离和大小差异,从而在处理各种复杂情况下的目标检测任务时具有更强的抑制能力,进一步减少了漏检和错检率。

1 LiteBlurVisionNet 算法原理

为解决在不良天气条件下行人与车辆检测问题,本文提出 LiteBlurVisionNet 模型,结构如图 2 所示,能够降低模型参数量,提高网络计算效率,同时模型的检

测性能也有明显提高。LiteBlurVisionNet 算法模型的特点主要在于主干特征提取网络和颈部网络。

1.1 Global MobileNetV3 主干网络

主干网络采取的是 MobileNetV3 网络, MobileNetV3 模块参数少,计算速度快。在 MobileNetV3-Large 模型中,进一步减少模型的参数量,采用非线性激活函数 *h-swish*,并将第 1 层卷积核的数量修改为 16 个 3×3 卷积。同时,MobileNetV3 模块引入通道注意力机制(SENet)模块^[16],以提升对行人和车辆特征的捕获能力。该模块不仅增强了图像局部信息特征的提取能力,还充分利用了上下文信息,从而提高了行人和车辆特征的提取能力。

然而,MobileNetV3 中的 SE 注意力机制存在一些缺陷。SE 注意力机制主要关注输入特征图中不同通道之间的关系,但未涉及全局信息的捕获。此外,SE 模块仅在通道维度上考虑注意力,无法全面捕获全局信息,这可能影响网络的学习能力。其次,SE 模块中的全局池化和两个全连接层会增加模型的计算复杂度,导致训练和推理时间增加。

因此,合理捕获全局信息、提高训练和推理速度变得至关重要。为了实现该目标,在 Backbone 部分,本文采用了一种新型注意力模块 global context enhancer 来改进 MobileNetV3 网络。该模块综合了 ECA 和 GMA 模块以及通道混洗操作^[17],以有效捕获全局信息。

ECA 模块采用了较为轻量级的设计,通过一维卷积操作有效减少了模型参数和计算复杂度。同时,利用全局平均池化和一维卷积捕获了特征图中通道间的关系,增强了信息交互,提高了训练与推理速度,结构如图 3 所示。

对于给定的输入特征图 $F_1 \in \mathbb{R}^{H \times W \times C}$,首先通过平均池化来提取特征,然后使用卷积核大小为 K 的一维卷积解决降维对通道注意力的负面影响,接着使用激活函数,得到通道权重 $M_C^{\text{avg}}(F_1)$,最终将输入特征图 F_1 与通道权重 $M_C^{\text{avg}}(F_1)$ 逐点相乘,最终得到通道注意力 $M_C(F_1)$ 。公式如下:

$$M_C^{\text{avg}}(F_1) = \sigma(f_1(\text{AvgPool}(F_1))) \quad (1)$$

$$M_C(F_1) = F_1 \otimes M_C^{\text{avg}}(F_1) \quad (2)$$

其中, σ 表示 Sigmoid 函数^[18], f_1 表示一维卷积运算, \otimes 表示逐点相乘。

GMA 模块结合了全局信息聚合、多层次感知机^[19]

处理和激活函数加权等操作,从而提取并加强了输入数据的全局上下文信息。这使得模型能够更全面地理

解输入数据的整体结构和语义信息,提高模型的表达能力。结构如图4所示。

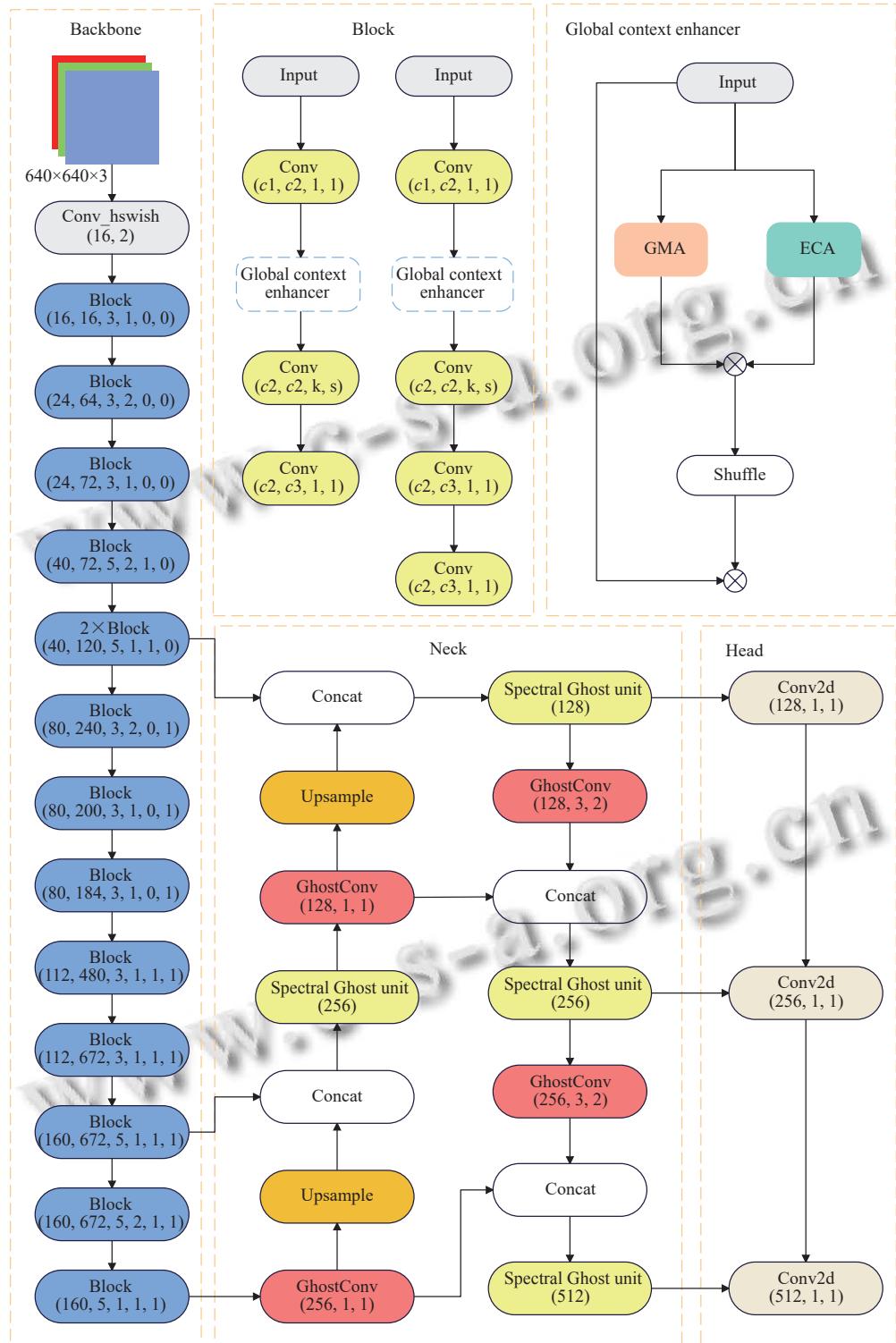


图2 LiteBlurVisionNet 模型结构图

对于给定的输入特征图 $F_2 \in R^{H \times W \times C}$,首先通过全局平均池化来提取输入宽高的特征,得到空间权重

$M_S^{\text{GlobalAvg}}(F_2)$,使得模块能够全面考虑输入特征图的内容,从而更好地理解整体语境和信息。然后通过使用

*ReLU*激活函数^[20]的MLP层的多层感知机处理,从而使模块能够学习到输入特征的复杂非线性关系,提高特征的抽象表达能力。空间注意力 $M_S(F_2)$ 计算公式如下:

$$M_S^{\text{GlobalAvg}}(F_2) = \text{GlobalAvgPool}(F_2) \quad (3)$$

$$M_S(F_2) = \sigma(M_S^{\text{GlobalAvg}}(F_2)) \quad (4)$$

其中, σ 表示 *Sigmoid* 函数, MLP 表示多层感知机。

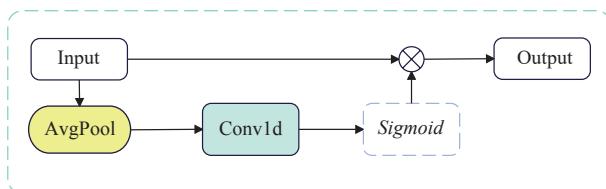


图3 ECA 注意力结构图

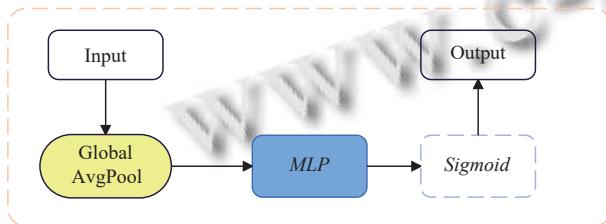


图4 GMA 结构图

通过结合这些操作,构建了 global context enhancer 模块,如图 5 所示。该模块综合利用了 ECA 和 GMA 注意力机制,同时结合了通道级别和全局级别的注意力机制。这样做可以使模型能够更全面地捕获特征之间的关系。此外,引入通道洗牌操作增加了特征的多样性和鲁棒性,有助于模型更好地捕获数据中的关键特征和模式,进一步提高模型的性能,该模块突出了感受野滑块内不同特征的重要性,并优先对感受野空间特征进行排序,有效解决了卷积核参数共享的问题,从而增强了模型的特征表达能力。

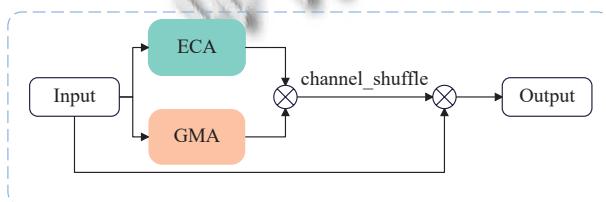


图5 Global context enhancer 结构图

这种结合方式不仅具备灵活性和可调性,而且可以根据任务需求进行参数调节,进一步提升模型的性能和适应性。这样的设计使得模型在特征表示方面更为丰富多样。引入 global context enhancer 模块显著增

强了模型对全局信息的捕获和理解能力,从而使其在数据表达和泛化方面更具优势。

1.2 Spectral Ghost neck net 颈部网络

在本文中, LiteBlurVisionNet 模型采用了轻量级网络 Ghost 以及拥有通道注意力的 Ghost bottleneck 模块来构建颈部网络。Ghost 模块是一种模型压缩方法,其核心思想是将卷积操作分解为两步:首先使用少量的卷积生成部分特征图;然后通过线性操作在这些特征图上生成更多特征图,并将它们拼接在一起。Ghost bottleneck 主要由两个堆叠的 Ghost 模块组成:第 1 个 Ghost 模块用于扩展通道数;而第 2 个 Ghost 模块则用于减少通道数以与快捷路径匹配。接着,使用 shortcut 连接将这两个 Ghost 模块的输入和输出连接起来。因此,颈部网络能够在降低模型参数量和计算复杂度的同时,实现模型的轻量化,并且在不改变输出特征图大小的情况下,减少推理时间并提升性能。Ghost 模块的结构如图 6 所示。

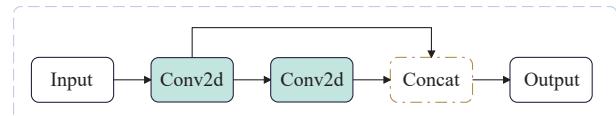


图6 Ghost 结构图

相比于 Ghost bottleneck 模块, spectral Ghost block 模块引入了频谱注意力机制(ECA 注意力机制),能够更有效地捕获全局上下文信息,提高特征的区分度和表达能力。使得模型在处理复杂数据时能够更好地理解特征之间的关系。使模型更专注于关键特征,提高对输入数据的表征能力。使得模型在参数量轻微变化的情况下获得更好的性能,有助于模型在各种视觉任务中取得好的性能表现。与此同时,尽管引入了 ECA 注意力模块,但模型仍然保持了轻量化的设计。结构如图 7 所示。

1.3 DIoUNMS

在传统的 NMS 算法中,检测框的最高置信度得分与其他检测框逐个比较,以计算它们的交并比(*IoU*)。交并比通过将预测框与真实框之间的交集面积除以它们的并集面积得到。如果 *IoU* 值超过预定义的阈值,则认为相应的检测框是冗余的,并将其过滤掉。显然,在传统的 NMS 中, *IoU* 是唯一的考虑因素,也是评估检测成功的唯一标准。

然而,仅依靠 *IoU* 存在局限性,因为它仅考虑两个

检测框的交集和并集区域之间的关系。这可能导致一些情况下无法作为有效的筛选标准。如图 8 所示，红色代表真实框，蓝色和绿色代表检测框。图 8(a) 展示了当两个检测框与真实框距离较远时，由于 IoU 值为 0，这些框未能被 NMS 过滤。图 8(b) 说明了两个检测框虽然大小相同，且与真实框的交集面积相等，但由于 IoU 值相同，它们未被正确区分，导致无法过滤。图 8(c) 则展示了两个大小相同的检测框完全包含在真实框内的情况，此时 IoU 值相等，也导致 NMS 处理不当，难以有效过滤。

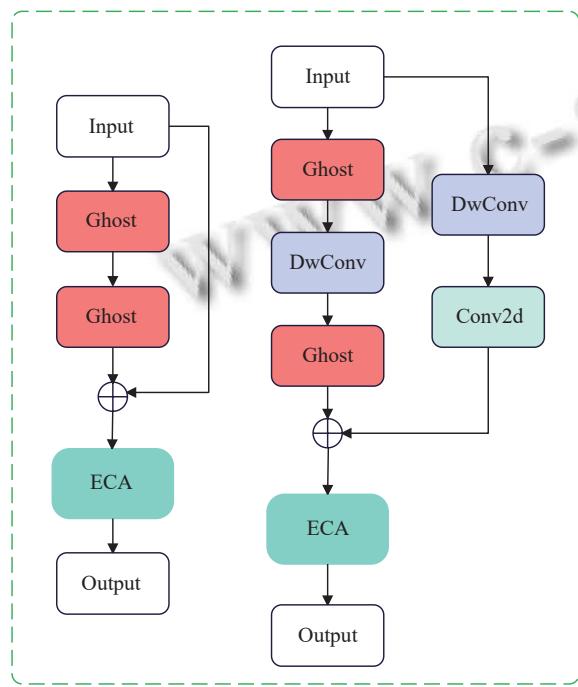


图 7 Spectral Ghost unit 结构图

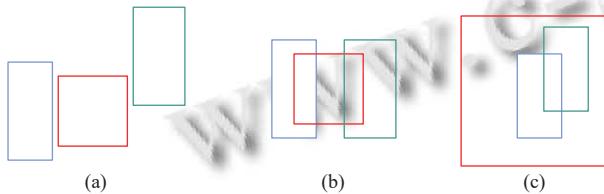
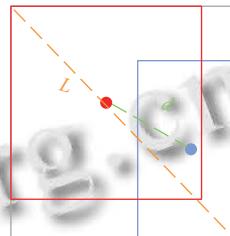


图 8 NMS 示意图

常规的检测算法一般采用的是 NMS 算法，针对 NMS 的上述局限性，提出了一种 $DIoU$ NMS 方法，引入了距离交并比 ($DIoU$) 计算方法。 $DIoU$ NMS 将传统 NMS 中的 IoU 计算替换为 $DIoU$ 。如图 9 所示， $DIoU$ 考虑了两个检测框之间的重叠区域和距离，可以更准确地选择检测框并促进更快的收敛。 $DIoU$ 的表达式如下所示：

$$DIoU = 1 - IoU + \frac{d^2(A, B)}{L^2} \quad (5)$$

其中， d 代表着地面实况框 A 和检测框 B 之间的距离。 L 表示最小外接框的对角线长度，该框包含了检测框和实况框。 IoU 表示交集比联合，它是通过将预测框和实况框之间的交集面积除以它们的并集面积来计算的。

图 9 $DIoU$ 示意图

1.4 K-means anchor 聚类

在目标检测训练中，确定适合数据集的锚框是至关重要的。锚框的选择直接影响到模型对目标位置、大小和类别的学习效果。一般来说，锚框越适应数据集中目标的大小和形状分布，模型就越能够准确地捕捉目标。为了解决不同数据集中目标尺寸和形状的差异，本文采用了 K-means 聚类算法^[21-23]对数据集中的目标进行分析，从而得到了适合数据集特征的初始锚框，即 [23, 17, 39, 30, 61, 47], [43, 100, 106, 68, 76, 186] 和 [133, 106, 201, 145, 141, 314]。通过在模型训练中使用这些聚类出来的锚框，提高了模型在目标检测任务上的性能，减少了推理过程中目标丢失的情况，进一步提高了检测结果的准确性和稳定性。

2 实验结果与分析

实验环境为 Ubuntu 操作系统，模型采用 PyTorch 深度学习框架实现，显卡为 NVIDIA GeForce RTX 3080，显卡运行内存为 10 GB，CPU 为 Intel(R) Xeon(R) Platinum 8255C，内存为 40 GB。模型初始输入图像大小设置为 640×640，模型训练周期 (epoch) 设置为 300，batch 大小设置为 8。初始学习率设置为 0.00918，循环学习率设置为 0.1269，学习率动量设置为 0.98。使用 SGD 作为优化函数，使用余弦退火进行训练，开启数据增强。

2.1 数据集

当前针对人车检测的公开数据集主要关注晴天场景，恶劣天气下的数据量可能较少，缺乏对恶劣天气下

道路场景的充分覆盖。并且一些恶劣天气下的数据可能受到雾霾、雨水或雪花等因素的影响，导致图像质量下降，标注的不准确性，目标物体变得模糊或难以识别。甚至因为标签错误可能会误导模型学习错误的特征，这会影响模型的性能和泛化能力，导致模型在恶劣天气条件下的性能不佳。

若为保证恶劣天气数据的数量而将多个数据集合并，又会缺乏统一的标准来定义数据集和评估指标。这使得不同数据集之间的比较和模型性能评估变得困难。

因此本文的实验采用自制数据集进行训练，通过网络爬虫、实景拍摄等手段进行数据集的获取，数据集包括3000多张道路场景下的人车图像，分别拍摄于雾霾、雨天、雪天等恶劣天气。对收集到的人车图像进行分类整理后，使用LabelImg软件进行人工标注，包含行人、汽车两个类别。部分数据集标注示例如图10所示。



图10 部分数据集标注示例图

2.2 评价参数

评估算法主要涉及两个方面：计算成本和准确性。在计算成本方面，本文主要考虑参数数量（Params）和每秒千兆浮点运算（GFLOPs）。一般来说，较小的Params和GFLOPs表示模型具有较低的计算成本和硬件性能需求。而准确性则主要通过以下参数来衡量：精度、召回率、平均精度（AP）和平均精度均值（mAP）。这些评价参数的具体计算方式如下：

$$P = \frac{tp}{tp + fp} \quad (6)$$

$$R = \frac{tp}{tp + fn} \quad (7)$$

$$AP = \int_0^1 P(R)dR \quad (8)$$

$$mAP = \frac{\sum_{i=1}^K AP_i}{K} \quad (9)$$

其中， P 为精确度， R 为召回率。 tp 表示正确检测到的正

样本数量， fp 表示被检测为正样本但实际为负样本的样本数量， fn 表示误检测为负样本的正样本数量， K 为检测到的目标类别总数。

3 消融实验

将LiteBlurVisionNet模型的Neck部分设置与YOLOv5相同，从而有效对比两个模型主干方面的差异，并将其作为基准模型。

首先，进行骨干网络对比实验。本文将基准模型用作骨干网络，并将其命名为Baseline。实验结果见表1。结果表明，Baseline基准模型的参数量和GFLOPs相较于YOLOv5s降低了74%，同时mAP提升了1.6%。这表明LiteBlurVisionNet模型的主干部分相较于YOLOv5s更加优越，模型更小。

表1 骨干网络消融实验

Model	Params	GFLOPs	P (%)	R (%)	mAP0.5 (%)
YOLOv5s	7015519	15.8	87.1	76.5	86
YOLOv5s+K-means	7015519	15.8	87.9	78.7	86.6
Baseline(串行)	1666804	4.1	84.2	80.3	87
Baseline	1666804	4.1	87.3	79.4	87.6

其次，本文在基准模型基础上，将head部分修改为Ghost模块和spectral Ghost unit模块，对应的对比模型分别命名为Baseline-GSConv和Baseline-spectral Ghost unit，同时使用两种模块被命名为Baseline-spectral Ghost，进行对比实验。实验结果见表2。结果表明，同时应用Ghost模块和spectral Ghost unit模块在修改骨干网络的基础上模型参数数量又降低了20%，并使mAP提升了0.04%。这些实验结果验证了本文所构建的颈部网络更为有效和轻便。

表2 颈部网络消融实验

Model	Params	GFLOPs	P (%)	R (%)	mAP0.5 (%)
Baseline	1666804	4.1	87.3	79.4	87.6
Baseline-Ghost	1570228	3.9	86.9	80.0	87.8
Baseline-SpectralGhost	1339804	3.5	88.1	79.8	88.0

同时，本文进行了全面评估，对LiteBlurVisionNet算法优化进行了逐步消融。相关实验数据如表3所示。

表3 总体消融实验

Model	Params	Model weights	GFLOPs	mAP0.5 (%)
		size (MB)		
YOLOv5s	7015519	13.7	15.8	86
YOLOv5s+K-means	7015519	13.7	15.8	86.6
Baseline	1666804	3.59	4.1	87.6
Baseline-spectral Ghost	1339804	3.03	3.5	88.0
Baseline+DIoU NMS	1339804	2.99	3.5	88.4

最后,为验证 LiteBlurVisionNet 算法的性能,本文将其与当前经典目标检测算法和相同应用的代表性算法进行对比实验,选择的目标检测算法包括 YOLOv7^[24]、YOLOv8、Faster R-CNN、RTDETR^[25], 相同应用的代表性算法为 IA-YOLO^[26] 和 IAL-YOLO^[26], 其中, YOLOv7 模型的深度和宽度调整为与 YOLOv5s 相同, 本文将其命名为 YOLOv7s。同时本文采用了 YOLOv8 中最轻量级的 8n 版本。IA-YOLO 为恶劣天气下目标检测, IAL-YOLO 为低光照的恶劣天气下目标检测。对比实验结果如表 4 和表 5 所示。可以看出, LiteBlurVisionNet 在参数量、计算量和 $mAP_{0.5}$ 方面都表现出了明显的优势。LiteBlurVisionNet 算法模型在参数量(1 339 804)和计算量(3.5 GFLOPs)上远低于其他算法模型, 这意味着模型在存储和计算方面更加高效, 尤其适用于资源受限的环境或移动端设备。同时其 $mAP_{0.5}$ (88.4%)也显著高于其他模型, 包括 IA-YOLO(74.95%)、IAL-YOLO(74.69%)、YOLOv5s(86%)、YOLOv7s(83.4%)和 YOLOv8n(86%)、Faster R-CNN(79.65%)、RTDETR-ResNet50(85.4%)。这表明 LiteBlurVisionNet 算法在轻量化、准确性和性能方面表现出了明显优越性, 适用于需要高精度目标检测的场景。该模型在模糊场景下进行行人与车辆检测任务也展现显著潜力, 为恶劣环境下的目标检测任务提供了更为可靠且高效的解决方案。

表 4 对比消融实验

Model	Params	GFLOPs	P (%)	R (%)	$mAP_{0.5}$ (%)
YOLOv5s	7015519	15.8	87.1	76.5	86
YOLOv7s	9137726	26	81	77.3	83.4
YOLOv8n	3006038	8.1	86.8	76.9	86
Faster R-CNN	28285573	474.1	36.9	87.3	79.65
RTDETR-ResNet50	41938794	125.6	85.2	77.8	85.4
LiteBlurVisionNet	1 339 804	3.5	85.2	80.9	88.4

表 5 相同应用对比消融实验

Models	Params	$mAP_{0.5}$ (%)
LiteBlurVisionNet	1 339 804	88.4
IA-YOLO	61694062	74.95
IAL-YOLO	61693997	74.69

4 结论

为提高模糊条件下人车检测算法的精度和降低模型复杂度, 本文提出模糊场景下行人与车辆检测模型 LiteBlurVisionNet(轻模糊视觉网络)。通过本文的改进, 本文有以下 3 方面的提高。

(1) 在主干网络部分, 引入了一种由全局上下文增强(global context enhancer)注意力模块改进的 Mobile-NetV3 模块, 该模块融合了 ECA 和 GMA 注意力机制, 以增强网络的特征提取能力。该模块能够有针对性地关注感受野内不同特征的重要性, 并通过通道洗牌操作增加了特征的多样性和鲁棒性, 从而进一步提高模型性能。通过对骨干网络的改进, 模型参数量下降了 76%, 计算量下降了 74%, $mAP_{0.5}$ 提高了 1.6%。

(2) 另外, 在模型的颈部结构中, 本文采用了轻量级的 Ghost 模块, 并提出了拥有通道注意力的 Ghost bottleneck 模块, 命名为 spectral Ghost unit 模块。采用频谱注意力机制能更有效地捕获全局上下文信息, 提高特征的区分度和表达能力, 使模型能够更好地理解特征之间的关系。通过频谱注意力机制对特征进行增强和提升, 有效地增强了模型对特征之间关系的捕获, 从而显著提升了模型的性能, 使得模型在相同参数量的情况下能够获得更好的性能。颈部网络在修改骨干网络的基础上参数量下降 20%, 计算量下降 15%, $mAP_{0.5}$ 提高了 0.04%。

(3) 针对不良天气条件下目标遮挡导致的识别困难, 本文采用 $DIoU$ NMS 作为后处理部分的改进方法, 考虑了目标框的位置和大小之间的距离, 从而在处理各种复杂情况下的目标检测时具有更强的抑制能力, 进一步减少了漏检和错检率。通过使用 $DIoU$ NMS, LiteBlurVisionNet 模型, $mAP_{0.5}$ 提高了 0.04%。

这显示了 LiteBlurVisionNet 模型在轻量化、准确性和性能方面都具有明显的优势, 适用于需要高精度目标检测的场景。下一步将继续研究去雾算法对检测算法的影响, 进一步提高检测精度。

参考文献

- Luo S, Yu J, Xi YJ, et al. Aircraft target detection in remote sensing images based on improved YOLOv5. IEEE Access, 2022, 10: 5184–5192. [doi: 10.1109/ACCESS.2022.3140876]
- Wang D, Zhang Y, Zhang KX, et al. FocalMix: Semi-supervised learning for 3D medical image detection. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 3950–3959.
- Liu Z, Cai YF, Wang H, et al. Robust target recognition and tracking of self-driving cars with radar and camera information fusion under severe weather conditions. IEEE

- Transactions on Intelligent Transportation Systems, 2022, 23(7): 6640–6653. [doi: [10.1109/TITS.2021.3059674](https://doi.org/10.1109/TITS.2021.3059674)]
- 4 Ren SQ, He KM, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149. [doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031)]
- 5 Cai ZW, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6154–6162.
- 6 院老虎, 翟柯嘉, 张泽鹏, 等. 基于模拟雾天遥感数据集的飞机目标检测研究. 南京邮电大学学报(自然科学版), 2021, 41(3): 77–84. [doi: [10.14132/j.cnki.1673-5439.2021.03.011](https://doi.org/10.14132/j.cnki.1673-5439.2021.03.011)]
- 7 Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.
- 8 Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam: Springer, 2016. 21–37.
- 9 Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318–327. [doi: [10.1109/TPAMI.2018.2858826](https://doi.org/10.1109/TPAMI.2018.2858826)]
- 10 蒋超, 张豪, 章恩泽, 等. 基于改进YOLOv5s的行人车辆目标检测算法. 扬州大学学报(自然科学版), 2022, 25(6): 45–49.
- 11 刘书刚, 张林坤, 杜昊东, 等. 雾天条件下改进YOLOv4的目标检测. 系统仿真学报, 2023, 35(8): 1681–1691.
- 12 院老虎, 常玉坤, 刘家夫. 基于改进YOLOv5s的雾天场景车辆检测方法. 郑州大学学报(工学版), 2023, 44(3): 35–41.
- 13 Wang QL, Wu BG, Zhu PF, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 11531–11539.
- 14 Han K, Wang YH, Tian Q, et al. GhostNet: More features from cheap operations. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 1577–1586.
- 15 Zheng ZH, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression. Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI, 2020. 12993–13000.
- 16 Hu J, Shen L, Sun G. Squeeze-and-Excitation networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7132–7141.
- 17 Zhang X, Zhou X, Lin M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6848–6856.
- 18 Xu QW, Lin RZ, Yue H, et al. Research on small target detection in driving scenarios based on improved YOLO network. IEEE Access, 2020, 8: 27574–27583. [doi: [10.1109/ACCESS.2020.2966328](https://doi.org/10.1109/ACCESS.2020.2966328)]
- 19 Tolstikhin I O, Houlsby N, Kolesnikov A, et al. MLP-mixer: An all-MLP architecture for vision. Advances in Neural Information Processing Systems, 2021, 34: 24261–24272.
- 20 Agarap A F. Deep learning using rectified linear units (relu). arXiv:1803.08375, 2018.
- 21 Sinaga KP, Yang MS. Unsupervised K-means clustering algorithm. IEEE Access, 2020, 8: 80716–80727. [doi: [10.1109/ACCESS.2020.2988796](https://doi.org/10.1109/ACCESS.2020.2988796)]
- 22 Ghazal TM, Hussain MZ, Said RA, et al. Performances of K-means clustering algorithm with different distance metrics. Intelligent Automation & Soft Computing, 2021, 30(2): 735–742.
- 23 Zhao YP, Zhou XL. K-means clustering algorithm and its improvement research. Journal of Physics: Conference Series, 2021, 1873(1): 012074. [doi: [10.1088/1742-6596/1873/1/012074](https://doi.org/10.1088/1742-6596/1873/1/012074)]
- 24 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 7464–7475.
- 25 Zhao Y, Lv WY, Xu SL, et al. DETRs beat YOLOs on real-time object detection. Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024. 16965–16974.
- 26 Liu WY, Ren GF, Yu RS, et al. Image-adaptive YOLO for object detection in adverse weather conditions. arXiv:2112.08088, 2022.

(校对责编: 王欣欣)