光电技术及应用

DOI:10.16818/j.issn1001－5868.20240930001

基于改进YOLOv8n的货车盲区目标检测方法

代少升[1],周　漫[1],余自安[2*],林玥楠[1*],于鑫尧[1]

(1. 重庆邮电大学 通信与信息工程学院,重庆 400065; 2. 昆明云内动力股份有限公司,昆明 650200)

摘　要: 针对货车盲区覆盖范围大、场景背景复杂、目标多尺度变化显著,以及传统检测方法存在漏检率高、特征表征能力不足等问题,提出了一种改进YOLOv8n盲区目标检测算法。该算法通过以下三个层面进行优化:1)在主干网络嵌入混合局部通道注意力模块(MLCA),提升网络局部空间特征提取能力;2)采用尺度序列特征融合模块(SSFF)融合图像多个尺度的深层语义信息,结合三重特征编码模块(TFE)完成局部细节增强;3)采用Inner-CIoU作为边框损失函数,提高边框检测精度。实验结果表明,在自建货车盲区多目标数据集上,改进算法的平均检测精度提升了3.14%,显著降低了复杂场景下的漏检率与误检率。

关键词: 深度学习; 目标检测; 注意力机制; 盲区检测

中图分类号: TP391　**文章编号:** 1001－5868(2025)01－0172－08

Object Detection in the Blind Spot of Truck Based on Improved YOLOv8n

DAI Shaosheng[1], ZHOU Man[1], YU Zian[2], LIN Yuenan[1] YU Xinyao[1]

(1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, CHN; 2. Kunming Yunnei Power Co., Ltd., Kunming 650200, CHN)

Abstract: To address the problems of large truck blind-zone scope, complex background, large variation of target scale, poor effect of existing truck blind spot detection methods, and easy-to-miss recognition, an improved YOLOv8n truck blind zone target-detection algorithm is proposed. A mixed local channel attention module is added to the backbone network to improve the local spatial feature extraction capability of the network. The feature fusion network is improved by the scale sequence feature fusion module to fuse the deep semantic information of multiple scales of the image, the triple feature encoding module was used to capture the local details of the target, and Inner-CIoU is adopted as the bezel loss function to improve border detection accuracy. The experimental results show that on the self-built vehicle - pedestrian dataset, the proposed algorithm has a 3.14% better average detection accuracy than the traditional YOLOv8n algorithm, as well as better target detection in the blind zone of trucks.

Key words: deep learning; object detection; attention mechanism; blind spot detection

0　引言

随着社会经济的快速发展,公路货运量呈现持续增长态势,作为核心运输载体的货车保有量亦随之快速攀升[1]。但随之而来的安全隐患也不容忽视,其中货车盲区范围大的问题尤为突出,及时、准确地检测货车盲区具有重要意义。

实验基于盲区摄像头的检测算法,通过深度学习算法监测货车周围环境,从而有效降低交通事故

收稿日期:2024－09－30.

*通信作者:余自安,林玥楠

的风险。虽然目前也有基于非视觉传感器的方法，如雷达[2]和红外传感器[3]，但这些技术成本高昂、迭代速度较慢。因此，基于盲区摄像头的方法在成本和效率方面具有一定优势。

货车在行驶过程中，由于车身尺寸大而限制了驾驶员视野，而摄像头的合理布置则可有效观察到视野盲区。再者，基于摄像头及计算机视觉技术成本低、算法可迭代，使其在车身环境感知系统中得到广泛应用。文献[4]提出一种基于卷积神经网络的盲区检测方法，通过图像分类的方法对盲区摄像头的画面进行分类。该分类网络在CIFAR10上能够达到95.35%的分类准确率，但其作为盲区检测网络无法获得目标位置信息而难以进行实际应用。

2021年，Worrakantapon等将摄像头安装在车辆侧窗后视镜下，利用机器学习对货车侧方盲区进行行人检测[5]。实验结果表明，实验车辆能基于单核CPU实时检测出在卡车侧方盲区中的行人，但实际道路环境中的检测准确率仍需提高。2023年，文献[6]提出一种基于机器视觉的驾驶员盲区安全预警方法，通过HOG和LBP特征提取并采用串行融合策略，实现了盲区行人检测在自建数据集上达到94.8%的准确率。但其实验场景单一，且没有在复杂场景上进行测试。

货车盲区范围大，其环境感知能力在复杂气象条件、多目标动态遮蔽等干扰场景下呈现显著局限性，在面对该复杂场景时，可能出现检测精度下降，影响盲区检测系统的准确性和实用性。因此，本文基于YOLO[7-9]系列中的YOLOv8n设计了一种具备混合局部通道注意力的目标检测网络，在主干网络中加入混合局部通道注意力模块[10]，结合多尺度特征融合的思想，引入尺度序列特征融合模块和三重特征编码模块[11]改进特征融合网络。采用Inner-CIoU Loss[12]代替CIoU Loss计算锚框损失。在PASCAL VOC 2017中提取车辆行人数据集并添加货车实际运行环境下的正负样本，改进的YOLOv8n网络在此数据集中有更好的目标检测精度，更低的漏检、错检率。

1　YOLOv8n网络介绍及改进

1.1　YOLOv8n

YOLOv8n的网络结构如图1，其结构主要可分为BackBone，Neck和Detect三部分。BackBone为主干网络，负责提取图形特征信息，由CBS，C2f和SPPF模块构成。C2f模块由YOLOv5中的C3模块改进而来，减少了一个卷积模块，增加了BottleNeck模块的跨层连接数。SPPF是针对SPP池化策略的改进，其中SPP采用的是不同尺度的空间金字塔池化，而SPPF则采用逐点最大池化。逐点最大池化对每个特征图中的每个像素进行最大池化操作，因此SPPF的计算量大大减少。SPP只使用一种尺度的池化核，而SPPF则引入了多个不同尺度的池化核。这使得SPPF能够捕获更丰富的空间信息，增强模型的表征能力。

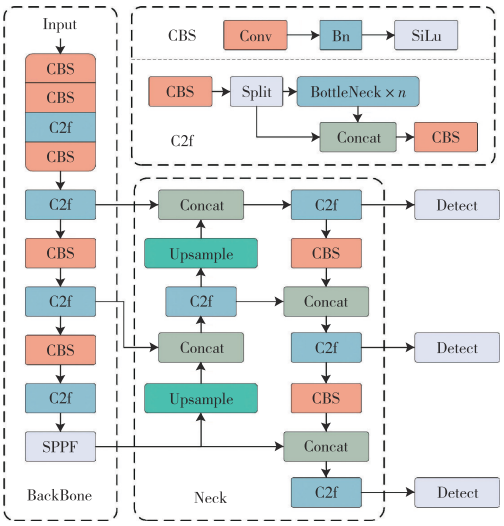


图1　YOLOv8n结构图

1.2　YOLOv8n网络改进

YOLOv8n网络虽然能在目标检测精度和识别速度上有较好的性能，但在货车盲区目标检测场景中检测效果并不理想。因为货车盲区范围较大，其运行过程中目标尺度变化速度快，目标的可分辨性差，在复杂背景情况下容易造成目标与背景混淆。基于此，在YOLOv8n的基础上，设计了一种改进的YOLOv8n网络，改进YOLOv8n网络结构如图2所示。改进主要分为三点：1)在主干网络中的C2f模块中加入混合局部通道注意力(Mixed Local Channel Attention，MLCA)设计C2f_MLCA模块，提升网络局部空间特征提取能力；2)在特征融合网络引入尺度序列特征融合模块(Scale Sequence Feature Fusion，SSFF)和三重特征编码模块(Triple Feature Encoding，TFE)，SSFF模块负责融合来自图像多个尺度的全局或深层语义信息，TFE模块捕获小目标的局部细节。再通过Concat

和C2f模块有效整合不同尺度的特征图；3）使用Inner-CIoU损失函数替换CIoU损失函数，进一步提高边框检测精度。通过上述改进，网络在复杂环境中的目标识别能力得到显著提升，尤其是在快速变化的场景中，有效提升了复杂背景下目标检测的精确率。

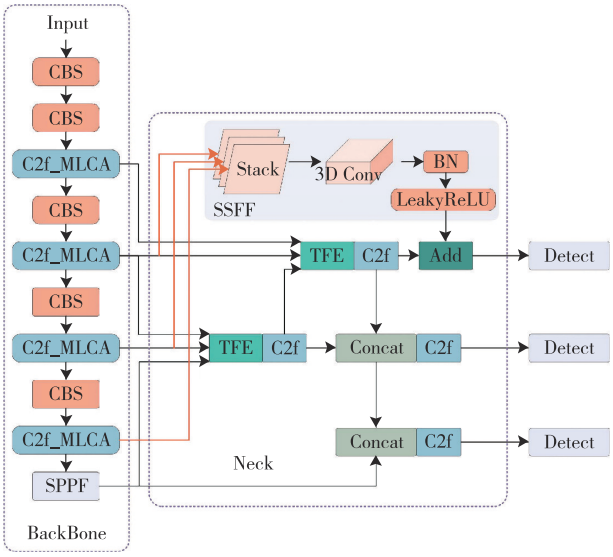


图2 改进的YOLOv8n网络

1.2.1 混合局部通道注意力

YOLOv8n中采用金字塔特征提取作为骨干网络，考虑到盲区背景复杂，在主干网络中添加注意力机制以提高骨干网络的特征提取能力，增强特征表示能力。常用的注意机制如SE[13]和ECA[14]考虑通道间的整体关系，而忽略了每个通道的空间信息。当一个通道的注意力权重较低而通道的一部分权重较高时，这部分重要的特征信息就会丢失。

近年来，许多学者开始研究基于局部特征信息的注意力机制，并通过实验证明，基于局部信息的注意力机制在局部信道信息、局部空间信息、局部节点信息等方面能取得良好的效果。采用的MLCA注意力机制其原理如图3所示。MLCA的输入特征图进行局部评价池化（Local Average Pooling，LAP），随后将特征图分为两个分支，第一个分支在LAP的基础上进行一次全局评价池化（Global Average Pooling，GAP），负责提取全局信息，第二个分支重构为$C\times ks\times ks\times 1\times 1$的向量复杂保存其局部空间信息。利用两个分支将输入转换为一维向量，第一个分支包括全局信息，第二个分支包含局部空间信息。经过一维卷积后，通过反池化恢复两个向量的原始分辨率，然后对信息进行融合，达到混合注意力的目的。

其中，Conv1d为一维卷积，卷积核大小k与信道维数C成正比，表明在捕获局部跨信道交互信息时，仅考虑每个信道与其k个相邻信道之间的关系。k的计算式为：

$$k=\Phi(C)=\left|\frac{\log_2(C)}{\gamma}+\frac{b}{\gamma}\right|_{odd} \tag{1}$$

式中，C表示通道数，b和γ都是超参数，默认值取2，odd表示k只取奇数，如果k为偶数则加1。将MLCA添加至主干网络中C2f模块的BottleNeck中，构建C2f_MLCA模块，如图4所示，以达到最佳特征提取效果。

1.2.2 尺度序列特征融合模块

在货车盲区检测场景中，不同尺度目标的融合能力对于提高检测效果至关重要，当特征金字塔提取时，网络从多个尺度对目标位置进行预测，以兼顾不同尺度的目标。但现有文献中的特征金字塔结构仅采用求和或串联对金字塔特征进行融合，这种方法无法有效利用所有金字塔特征图之间的相关性。本研究引入了一种新颖的SSFF模块，该模块可以更好地将多尺度特征图，即深层特征图的高级语义与具有相同纵横比的浅层特征图的详细信息结合起来。

SSFF模块将特征图P3，P4和P5先与一系列标准差递增的高斯核进行卷积[15]，可表示为：

$$F_\sigma(i,j)=\sum_{\mu}\sum_{\nu}f(i-\mu,j-\nu)\times G_\sigma(\mu,\nu) \tag{2}$$

$$G_\sigma(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right) \tag{3}$$

式中，f表示二维（2D）特征图，F_σ是通过使用2D高斯滤波器对一系列卷积进行平滑生成的，其标准差为σ。

然后将这些不同尺度的特征图水平堆叠，并使用3D卷积来提取它们的尺度序列特征。由于高斯平滑输出特征图具有不同的分辨率，采用最近邻插值方法将所有特征图对齐到与P3相同的分辨率。这是因为高分辨率特征图级P3包含了对小物体的检测，因此SSFF模块是基于P3的分辨率设计的。如图5所示，所提出的SSFF模块由以下组件组成：

（1）使用1×1卷积将P4和P5特征图的通道数更改为256；

（2）使用最近邻插值法将P4和P5大小调整为

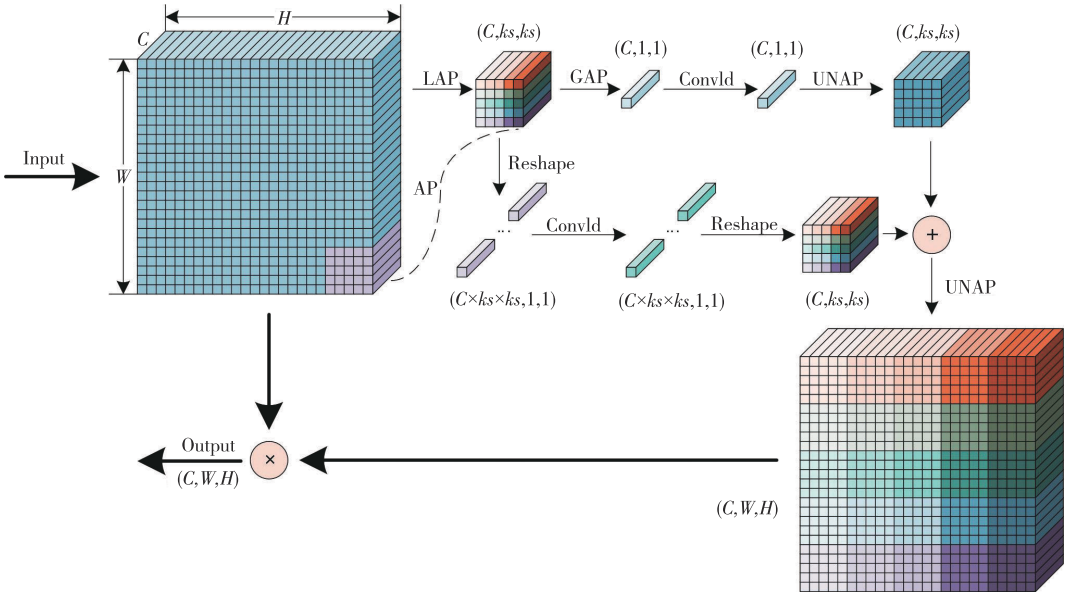


图 3　MLCA 模块原理图

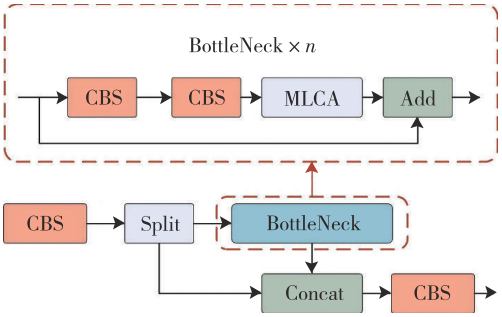


图 4　C2f_MLCA 模块示意图

P3 水平；

（3）将卷积后的 P3，P4 与 P5 的 3D 特征图合并为一个 4D 特征图；

（4）然后将 4D 特征图沿深度维度连接起来，形成 3D 特征图，用于后续卷积；

（5）最后，经过 3D 卷积、3D 批量归一化和 LeakyReLU[16]激活得到输出。

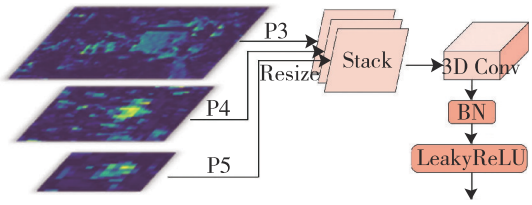


图 5　SSFF 模块示意图

1.2.3　三重特征编码模块

为了能从复杂背景中识别出目标，可以通过放大图像来参考和比较不同比例的形状或外观变化。由于骨干网的不同特征层具有不同的大小，传统的 FPN[17]融合机制只对小尺寸的特征图进行上采样，然后将其拆分或添加到前一层的特征中，而忽略了大尺寸特征层的丰富详细信息。TFE 模块将大、中、小特征拆分，添加大尺寸特征图，进行特征放大以完善详细特征信息。图 6 为 TFE 模块的结构。

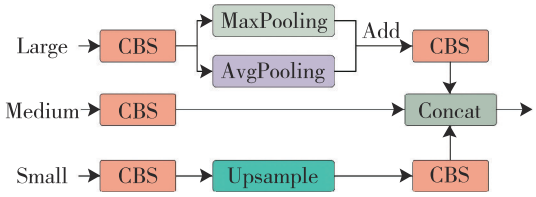


图 6　TFE 模块示意图

在特征编码前，首先调整特征通道的数量，使其与中尺寸特征（Medium）的通道数一致。卷积模块处理大尺寸特征图（Large）后，调整其通道数。然后，采用最大池化和平均池化混合结构进行下采样，有助于降低特征的空间维度，实现平移不变性，增强网络对输入图像空间变化和平移的鲁棒性。对于小尺寸特征图（Small），首先使用卷积模块调整通道数，然后使用最近邻插值方法进行上采样。这有助于保留局部要素并防止小对象要素的信息丢失，因为最近邻插值技术可以通过利用相邻像素来填充特征图，并考虑子像素邻域。最后，对尺寸已经相同的大、中、小尺寸的三个特征图进行一次卷积，然后在通道维度上进行拼接。通过这种多尺度

特征的编码和融合，TFE模块不仅保留了各尺度特征图的关键信息，还增强了网络对不同尺寸目标的感知能力。特别在货车盲区检测场景中，该模块可以有效捕捉小目标的细节，同时保持对大目标的准确检测，从而提升了模型在复杂背景下的鲁棒性和检测精度。

1.2.4 Inner－CIoU损失函数

边界框回归损失函数是目标检测任务的一个重要组成部分，检测器的定位精度很大程度上取决于边界框的回归情况，而边界框回归（Bounding Box Regression，BBR）在目标检测中起着关键的作用。本研究引入Inner-IoU损失的方法，结合CIoU中的形状损失和距离损失得到Inner-CIoU，该损失使用辅助边界框进行计算，拥有更好的检测效果。

在BBR中，IoU损失用来描述预测的边界框与基准（Ground Truth，GT）框之间的匹配程度，保证模型在训练过程中能够学习到目标的位置信息。作为现有主流边界框回归损失函数的重要组成部分，IoU的定义式为：

$$IoU=\frac{\left|B\cap B^{gt}\right|}{\left|B\cup B^{gt}\right|} \tag{4}$$

式中，B和B^{gt}表示预测框和GT框，定义了IoU后，对应的损失函数为：

$$Loss_{IoU}=1-IoU \tag{5}$$

目前主流的IoU损失函数主要有GIoU[18]，CIoU[19]，SIoU[20]等，其主要改进一般为在IoU损失的基础上添加新的损失项来加速收敛。如CIoU，其计算公式为：

$$\begin{cases} Loss_{CIoU}=1-IoU+\dfrac{\rho^2(p,p^{gt})}{\left|c^2\right|}+\dfrac{V^2}{(1-IoU)+V} \\ V=\dfrac{4}{\pi^2}\left(\arctan\dfrac{w^{gt}}{h^{gt}}-\arctan\dfrac{w}{h}\right)^2 \end{cases} \tag{6}$$

CIoU是在IoU损失的基础上增加一个形状损失和一个距离损失。虽然边界框回归损失函数可以通过在IoU损失函数中添加新的几何约束来有效加速收敛并提高检测性能，但没有考虑IoU损失本身的合理性，这在很大程度上决定了检测结果的质量。为弥补这一缺陷，引入Inner-IoU损失。Inner-IoU提出了辅助边界框的方法，其计算公式为：

$$\begin{cases} b_l^{gt}=x_c^{gt}-\dfrac{w^{gt}\times ratio}{2},\ b_r^{gt}=x_c^{gt}+\dfrac{w^{gt}\times ratio}{2} \\ b_t^{gt}=y_c^{gt}-\dfrac{h^{gt}\times ratio}{2},\ b_b^{gt}=y_c^{gt}+\dfrac{h^{gt}\times ratio}{2} \end{cases} \tag{7}$$

式中，$ratio$取值范围为0.5～1.5，对于预测框也应计算其辅助框b_l，b_r，b_t和b_b，且$ratio$取值应相同。则Inner-IoU计算过程如下：

$$\begin{aligned} inter=&(\min(b_r^{gt},b_r)-\max(b_l^{gt},b_l))\times\\ &(\min(b_b^{gt},b_b)-\max(b_t^{gt},b_t)) \end{aligned} \tag{8}$$

$$\begin{aligned} union=&(w^{gt}\times h^{gt})\times(ratio)^2+\\ &(w\times h)\times(ratio)^2-inter \end{aligned} \tag{9}$$

$$IoU_{inner}=\frac{inter}{union} \tag{10}$$

IoU与Inner-IoU仅存在尺度上的差异，因此损失函数的计算方式是相同的，根据CIoU损失函数的定义，可以得出Inner-CIoU损失函数定义如下：

$$Loss_{Inner-CIoU}=Loss_{CIoU}+IoU-IoU_{inner} \tag{11}$$

经多次试验对比，确定$ratio$取值为0.75可达到最佳效果。

2 实验结果及分析

2.1 数据集构建

为保证模型的泛化性能，从PASCAL VOC 2017数据集中截取出行人和车辆的部分，再加入货车盲区摄像头实拍的场景，其中包含了复杂环境下的正负样本。实验数据集共18383张图片，部分加入的样本如图7所示。



（a）左右视角

（b）前后视角

图7 部分数据集样本

数据集标签按照PASCAL VOC数据集标准格式，采用labelImg标注。检测类分别为Person，Bus，Bicycle，Motorbike，Car共5类。训练集、验证集按8:2的比例随机分配。同时启用Mosaic数据增强，使数据集有更好的泛化效果。

2.2 实验环境及评价指标

实验参数如表1所示。

表1 实验环境参数

参数	配置
操作系统	Windows 11
CPU	Intel Core i5－12490f(32GB)
GPU	GeForce RTX 3060Ti(8GB)
Python版本	3.8.19
Pytorch版本	2.0.1
Cuda版本	12.4

设置迭代次数为300轮次，batch-size设置为16，优化器采用SGD，初始学习率为0.01，动量为0.937。采用精确率(Precision，P)、召回率(Recall，R)，均值平均精度(mean Average Precision，mAP)，模型参数量，每秒10亿次的浮点运算数(Giga Floating-point Opera-tions Per Second，GFLOPs)和推理速度进行作为评价指标。其中，P，R和mAP的计算式为：

$$\begin{cases} P = TP/(TP + FP) \\ R = TP/(TP + FN) \\ mAP = \sum AP/N \end{cases} \tag{12}$$

式中，TP，FP和FN分别表示真阳性，假阳性和假阴性的数量。AP为不同置信度阈值下计算查准率和召回率，然后绘制精度-召回率曲线(Precision-Recall Curve，PR)的面积。N表示检测的类别总数。

2.3 对比实验

为验证所提出算法的先进性，使用在2.1节中构建的车辆行人数据集，同当前主流的轻量级目标检测网络进行对比测试，对比实验如表2所示。分析表2可知，所提出的算法的mAP值最高，达到76.70%，同时平均推理速度为2.5 ms处于较好水平。对比次优方法YOLOv5s，虽然后者的mAP接近，但其精确率较低，且$GFLOPs$值高达23.8，对运行设备的性能要求较高；其参数量也仅次于Resnet50-SSD的13.5 M，难以满足嵌入式设备的硬件需求。YOLOv9t同样具有较低的参数量和$GFLOPs$，但实际运行时的平均推理延迟要高于本算法。RTMDet-tiny具有较好的mAP，但经过测试其精确率较低，且推理速度远远低于YOLO系列算法。因此，本算法在mAP、精确率和推理速度等关键指标上均显示出了明显的优势。

表2 不同网络模型实验对比表

网络模型	分辨率	P/%	R/%	mAP@0.5/%	参数量/M	$GFLOPs$	推理速度/ms
Resnet50－SSD	300×300	63.95	73.45	69.91	13.50	15.3	18.7
RTMDet－tiny	640×640	60.97	79.20	74.06	5.53	30.3	21.3
YOLOv5s	640×640	69.91	73.49	76.12	9.11	23.8	3.2
YOLOv6n	640×640	74.68	69.56	74.06	4.23	11.9	2.5
YOLOv7－tiny	640×640	73.35	68.28	73.13	6.01	13.1	2.9
YOLOv9t	640×640	74.02	69.42	74.86	1.97	7.6	3.1
本文	640×640	75.20	69.93	76.70	3.05	8.6	2.5

2.4 消融实验

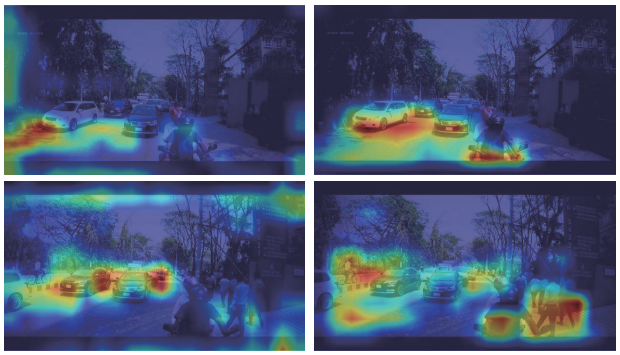
为验证各项改进的有效性，以YOLOv8n模型为基准进行不同改进来做消融实验，训练数据集为自建的车辆行人数据集，表3给出了相关实验结果。

实验结果表明，原始的YOLOv8n精确率为72.86%，召回率为68.23%，mAP值达到73.56%。以此为基础，只在主干部分加入MLCA混合局部注意力模块后，mAP值等明显提升，且推理速度的参数量没有明显变化。输入测试图片，读取SPPF层的热力图如图8所示，对比图8(a)和图8(b)可见在主干网络中混合局部通道注意力模块，使改进的YOLOv8n网络能够更加准确和全面地关注到检测目标的位置。

表中Neck表示特征融合网络的改进，加入后mAP值提升1.02%，精确率提升幅度较大。同时仅增加了少量的参数，在性能影响不大的情况下，改

表3 消融实验对比表

模型	P/%	R/%	mAP@0.5/%	参数量/M	推理速度/ms
YOLOv8n	72.86	68.23	73.56	3.03	2.3
YOLOv8n＋MLCA	73.86	69.49	75.03	3.03	2.3
YOLOv8n＋MLCA＋Neck	77.36	69.53	76.01	3.05	2.5
YOLOv8n＋MLCA＋Neck＋Inner－CIoU	75.20	69.93	76.70	3.05	2.5



（a）YOLOv8n （b）改进的YOLOv8n

图 8 热力图对比

进的特征融合网络有效提高了检测精度。实验结果表明，改进的特征融合网络具有更强的深层特征融合能力，在复杂背景下能更好地获取图像中的目标信息。相对于原始的YOLOv8n在提高准确率的情况下，只增加少量的参数，且总体推理速度仍然保持在2.5 ms。

图9显示了部分结果对比，从中可以看出改进的YOLOv8n在漏检方面有明显改进，同时在各种场景下的预测精度也有所提升。



（a）YOLOv8n （b）改进的 YOLOv8n

图 9 部分预测结果对比

3 总结

本文提出了一种改进YOLOv8n的货车盲区检测方法。采用混合局部通道注意力模块增强主干网络获取图像中的局部空间特征的性能。使用尺度序列特征融合模块和三重特征编码模块设计改进的特征融合网络，对比普通的FPN融合机制，改进的特征融合网络能够有效的融合各尺寸特征图的信息，且具有更好的特征表达能力。实验表明，改进的YOLOv8n网络具有更好的局部空间特征提取能力，能够融合深层和浅层特征信息，提高了检测精度。同时，在不影响模型性能的情况下，引入Inner-IoU的方法，将损失函数改为Inner-CIoU，使模型精度进一步提升。在自建数据集上进行对比实验，验证了相关改进的有效性。所提出方法最终实现了对货车多路盲区检测，对道路交通安全发展具有重要意义。

参考文献：

[1] 交通运输部. 2023年交通运输行业发展统计公报[N]. 中国交通报, 2024-06-18(002).
Ministry of Transportation and Communications. Statistical Bulletin on Transportation Industry Development in 2023[N]. China Transportation News, 2024-06-18(002).

[2] Kim W, Yang H, Kim J. Blind Spot Detection Radar System Design for Safe Driving of Smart Vehicles [J]. Applied Sciences, 2023, 13(10): 6147-6169.

[3] 陈 岚, 赵 竹, 吴 欢, 等. 基于人体红外传感器的大型车辆盲区预警系统设计[J]. 无线互联科技, 2023, 20(6): 74-76.
Chen Lan, Zhao Zhu, Wu Huan, et al. Design of blind zone warning system for large vehicles based on human infrared sensor [J]. Wireless Internet Technology, 2023, 20(6): 74-76.

[4] Zhao Y, Bai L, Lyu Y, et al. Camera-Based Blind Spot Detection with a General Purpose Lightweight Neural Network [J]. Electronics, 2019, 8(2): 233.

[5] Worrakantapon A, Pongsena W, Kerdprasop K, et al. Real-Time Human Detection in a Restricted Area for Safety in Truck Dumper Control System Using Deep Learning[J]. 2021, 10(1): 29-35.

[6] 王 巍, 白 龙, 马欢畅, 等. 基于机器视觉的驾驶员视野盲区安全预警方法研究[J]. 计算机科学, 2023, 50(S1): 498-504.
Wang Wei, Bai Long, Ma Huanchang, et al. Research on machine vision-based safety warning method for driver's visual field blind zone [J]. Computer Science, 2023, 50(S1): 498-504.

[7] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv: 2004.10934, 2020.

[8] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications [J]. arXiv preprint arXiv: 2209.02976, 2022.

[9] 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]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 7464-7475.

[10] Wan D, Lu R, Shen S, et al. Mixed local channel attention for object detection[J]. Engineering Applications of Artificial Intelligence, 2023, 123: 106442.

[11] Kang M, Ting C M, Ting F F, et al. ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation [J]. Image and Vision Computing, 2024, 147: 105057.

[12] Zhang H, Xu C, Zhang S. Inner-IoU: more effective intersection over union loss with auxiliary bounding box[J]. arXiv preprint arXiv:2311.02877, 2023.

[13] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 7132-7141.

[14] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 11534-11542.

[15] Tran D, Bourdev L, Fergus R, et al. Learning Spatiotemporal Features with 3D Convolutional Networks [C]//2015 IEEE International Conference on Computer Vision (ICCV), 2015: 4489-4497.

[16] Xu B, Wang N, Chen T, et al. Empirical evaluation of rectified activations in convolutional network [J]. arXiv preprint arXiv:1505.00853, 2015.

[17] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2117-2125.

[18] Rezatofighi H, Tsoi N, Gwak J, et al. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 658-666.

[19] Zheng Z, Wang P, Ren D, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation [J]. IEEE Transactions on Cybernetics, 2021, 52(8): 8574-8586.

[20] Gevorgyan Z. SIoU loss: More powerful learning for bounding box regression [J]. arXiv preprint arXiv: 2205.12740, 2022.

作者简介：

代少升(1974—)，男，河南省潢川县人，博士，教授，主要从事红外成像系统及SOPC嵌入式系统设计与开发；

林玥楠(2001—)，女，云南省昆明市人，硕士研究生，主要从事红外成像技术的研究。