光电技术及应用

DOI：10.16818/j.issn1001－5868.20251009004

基于改进YOLO11n的大车盲区人员检测方法

代少升[1]，宋国庆[1]，余自安[2*]

（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： 针对大车盲区中行人和骑行者目标检测存在遮挡、检测精度低、实时性差等问题，文章提出一种基于改进YOLO11n网络的盲区人员检测方法。首先，在主干网络中引入感受野增强注意力卷积模块（RFCBAMConv），改进C3k2模块，以提升对遮挡目标的感知能力；其次，引入内容引导注意力特征融合结构（CGAFusion），引导网络关注关键区域特征，增强多尺度语义融合效果；最后，采用多尺度特征共享检测头（MFSD），通过多分支共享卷积核降低模型参数量与计算负担。实验在自建大车盲区数据集上进行，结果表明所提方法在mAP@0.5上分别提升约2.5%，召回率显著高于YOLOv5n等同类轻量模型，同时保持仅2.87 M参数量和8.2 GFLOPs的低计算复杂度，验证了其在精度、效率与部署适应性上的优势。该方法适用于资源受限的边缘计算设备，在智能交通安全领域具有广阔的应用前景。

关键词： 人员检测；YOLO11n；注意力机制；特征融合

中图分类号： TP391　**文章编号：** 1001－5868(2025)06－1071－08

Personnel Detection in Large Vehicle Blind Spots Using Improved YOLO11n

DAI Shaosheng[1], SONG Guoqing[1], YU Zian[2]

(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: This paper presents a blind spot detection method based on an improved YOLO11n network that addresses the challenges of pedestrian and cyclist detection in large vehicle blind spots, including occlusion, low detection accuracy, and poor real-time performance. First, the receptive-field augmented attention convolutional module (RFCBAMConv) is introduced into the backbone network to improve the C3k2 module and thus enhance the perception of occluded objects. Second, the content-guided attention feature fusion architecture (CGAFusion) is incorporated to ensure that the network focuses on key regional features, enhancing multiscale semantic fusion. Finally, a multiscale feature sharing detection (MFSD) head is employed, which ensures that convolution kernels are shared across multiple branches and thus reduces the number of model parameters and computational burden. Experiments on a self-constructed large vehicle blind spot dataset demonstrate that the proposed method achieves approximately 2.5% improvement in mAP@0.5 and significantly higher recall compared with similar lightweight models such as YOLOv5n, while maintaining low computational complexity (only 2.87 M parameters and 8.2 GFLOPs). These results demonstrate its advantages in accuracy, efficiency, and deployment adaptability. Therefore, the proposed method is suitable for resource-constrained edge computing devices and has broad application prospects in the field of intelligent transportation safety.

Key words: personnel detection; YOLO11n; attention mechanism; feature fusion

收稿日期：2025－10－09.

*通信作者：余自安

0 引言

随着我国经济的持续发展，交通运输业迅速壮大，大型车辆凭借其运载能力强、运输效率高的优势，在公路运输中发挥着重要作用[1]。但由于车身庞大、结构复杂，大型车辆普遍存在明显的视野盲区，尤其在转弯或倒车过程中更为突出。同时，道路人员通常缺乏对盲区危险性的认知，一旦进入内轮差或倒车盲区，极易引发交通事故。因此，如何精准检测盲区内人员并及时预警，已成为智能交通安全领域亟待解决的关键研究问题。

近年来，基于计算机视觉的目标检测算法在车辆盲区检测中得到广泛应用。Masalov等[2]将HOG特征与LBP特征相结合，用于提取骑行者在不同视角下的外观特征，并采用AdaBoost分类器进行训练，在测试集上取得了88.12%的平均准确率。但该方法本质上是分类模型，缺乏位置信息，难以满足实时预警的精确性需求。Wang等[3]针对行人和骑行者检测中频繁出现的误检和漏检问题，在Fast R-CNN[4]框架下提出多项改进策略，包括困难样本挖掘、多目标候选区域输入以及多层次特征融合等方法，有效提升了检测性能。但由于两阶段算法模型复杂、参数量和计算量大，难以兼顾实时性要求。韦昊[5]则在YOLOv3算法的基础上引入图像增强与网络剪枝技术，用于大型货车盲区障碍物的检测，有效压缩了模型体积并增强了其在夜间环境下的检测能力。

需要指出的是，复杂盲区场景中人员通常位于车头、车尾或车侧等不可视区域，在图像中呈现小尺寸、畸变显著、特征不明显等特征，同时还常被车辆结构或其他障碍物部分遮挡，给检测带来极大挑战。此外，行人和骑行者等弱势交通参与者虽目标种类有限，但其行为具有高度不确定性，如突然穿越道路、贴近车辆行驶等，进一步提升了检测系统对准确性和响应速度的要求。

综上分析，采用单阶段目标检测器更能满足大车盲区人员检测对实时性与精度的双重需求。因此，本文选择YOLO系列中对遮挡小目标检测性能表现优越的YOLO11n网络[6-10]，并在其基础上进行改进。在主干网络中引入感受野注意力(RFCBAM)机制以增强特征提取能力，并结合多尺度特征融合思想，设计基于内容引导注意力(CGA)的特征融合模块。同时，考虑到检测类别较少，检测头各分支间有较强的相关性，特引入MFSD以降低模型参数量和计算负担。

在由自行采集的盲区图像、nuImages右后方视角子集以及Tsinghua Daimler数据集组成的自建数据集中，改进后的YOLO11n网络在保持较低计算复杂度的同时，显著提升了检测精度，降低了漏检与误检率，验证了其在大车盲区人员检测任务中的有效性与实用性。

1 YOLO11网络

YOLO11的网络结构如图1所示，整体由三部分组成：主干网络(Backbone)、颈部网络(Neck)和检测头(Head)。

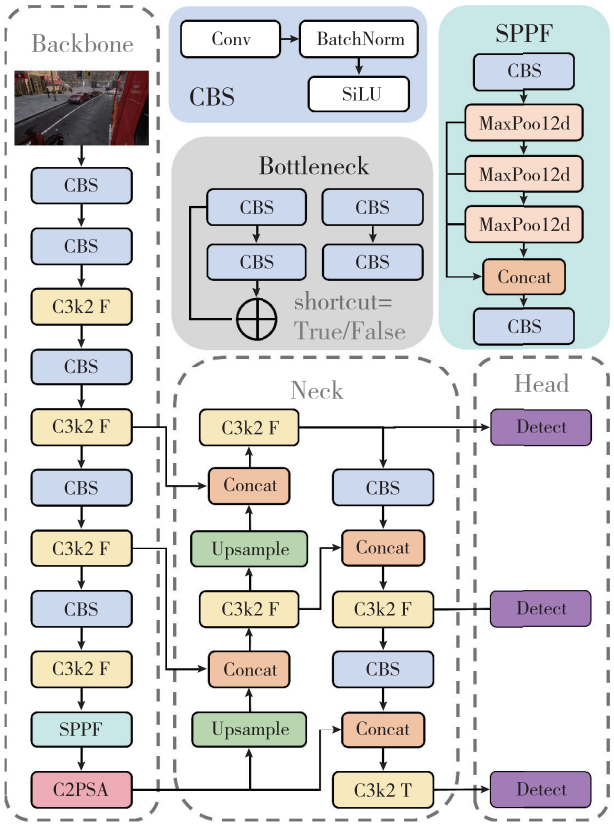


图1 YOLO11网络结构图

主干网络负责提取输入图像的多尺度特征信息，主要由CBS，C3k2，SPPF和C2PSA模块组成。CBS模块通过卷积、归一化和激活函数实现特征下采样。C3k2模块在YOLOv8的C2f基础上改进，通过参数C3k控制Bottleneck的类型：当C3k设置为False时，模块结构退化为轻量化的C2f形式；当设置为True时，则启用C3k模块，在浅层网络中通常禁用C3k以减少计算开销，而在深层网络中启用以增强特征提取能力。

位于空间金字塔池化模块SPPF之后的C2PSA模块对C2f进行扩展改进，融合了逐点空间注意力（PSA）机制，以增强网络对全局上下文信息的感知。PSA模块采用多头自注意力机制，可并行处理多组特征通道，提高特征表达的丰富性和网络性能。该模块还集成了前馈网络FFN，将特征映射到更高维空间，有助于建模复杂的非线性关系，增强模型对细粒度目标及边缘信息的感知能力。

颈部网络仍采用PAN-FPN结构，通过自底向上的信息流补充自顶向下FPN结构中缺失的定位信息，从而更好地融合浅层的位置信息与深层的语义特征，提升特征传递的完整性与检测精度。

检测头部分采用解耦式结构，设置独立的分支分别用于类别预测与位置回归。其中，类别预测分支中额外引入了深度可分离卷积操作，显著降低了模型的参数量和计算开销。

2　改进的YOLO11n网络

针对盲区场景中小目标尺寸、遮挡严重和实时性要求高的问题，本文在YOLO11n基础上引入了三项改进，改进后的结构如图2所示。

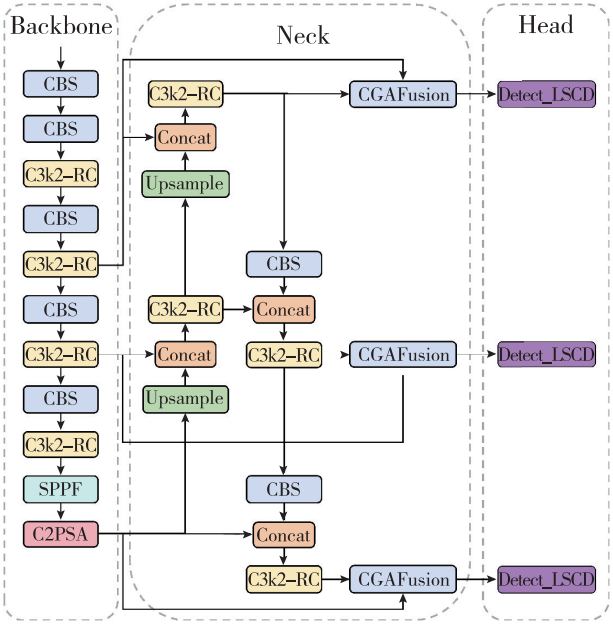


图2　改进的YOLO11n网络

首先，在C3k2模块中引入感受野注意力卷积（RFAConv）和卷积注意力（CBAM）相结合的感受野增强注意力卷积（RFCBAMConv），改进了C3k2模块，提高了模型对遮挡区域的感知能力。

其次，在特征融合阶段引入了内容引导注意力特征融合模块（CGAFusion）。该模块结合多粒度特征之间的上下文交互信息，引导网络关注目标区域的关键特征。通过并行通道注意力与空间注意力机制，CGAFusion能够自适应调整不同尺度特征的重要性，增强语义与细节之间的融合效果，提升对多尺度目标的识别能力。

最后，引入了共卷积检测头（MFSD）。该结构采用共享卷积核实现多层特征图的统一编码，减少了冗余参数的同时，提升了推理效率。

2.1　感受野增强注意力特征提取模块

为提升模型对遮挡区域的感知能力，改进了C3k2模块中的残差瓶颈网络，将其中的第二个卷积替换为RFCBAMConv，形成了C3k2-RC模块。该改进不仅保留了CBAM的轻量高效特性，同时结合感受野空间特征建模，使注意力机制具备更强的空间建模能力。

2.1.1　卷积块注意力机制

CBAM[11]是一种轻量级且高效的注意力机制，核心思想是通过串行残差连接通道注意力模块与空间注意力模块，引导模型更加关注关键信息区域，从而提升特征表达能力。其结构如图3所示。

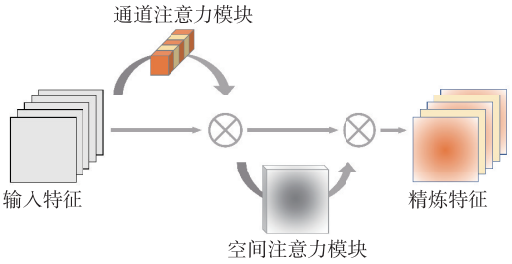


图3　CBAM结构图

然而，CBAM中采用的标准卷积操作存在一定局限。该操作使用固定大小的卷积核在整张特征图上滑动，且所有位置共享同一组参数。虽然这种参数共享机制能显著降低模型复杂度，但也导致卷积核在面对复杂背景、遮挡区域或小目标时难以捕捉不同位置的语义差异，从而限制了模型的空间建模能力。

2.1.2　感受野增强注意力卷积

感受野空间特征能更好地表达不同位置的特征差异，其核心思想是通过分组卷积对输入特征图进行高效空间特征提取，然后按照感受野滑动窗口的形状进行特征重排，将原始的$C\times H\times W$特征图转换为显式空间结构，即$C\times KH\times KW$，使网络能够对每个滑动窗口内的空间感受野进行独立建模

与加权，从而实现更细粒度的注意力控制。具体步骤为：

（1）通过深度卷积生成展开后的局部感受野特征：

$$F_g = \mathrm{ReLU}\left\{\mathrm{BN}\left[\mathrm{Conv}_{dw}(X)\right]\right\} \quad (1)$$

式中，Conv_{dw}表示分组数等于通道数的深度卷积，输入特征X维度为$C\times H\times W$，输出特征F_g维度为$(C\cdot K^2)\times H\times W$。

（2）再通过重排展开为二维特征图，形状为$C\times KH\times KW$。

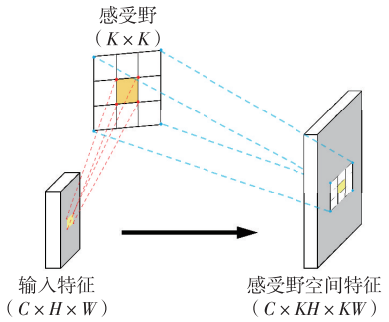


图4 感受野空间特征

基于上述思想，将感受野空间特征引入CBAM注意力机制中，得到感受野增强注意力卷积模块RFCBAMConv[12]，如图5所示。

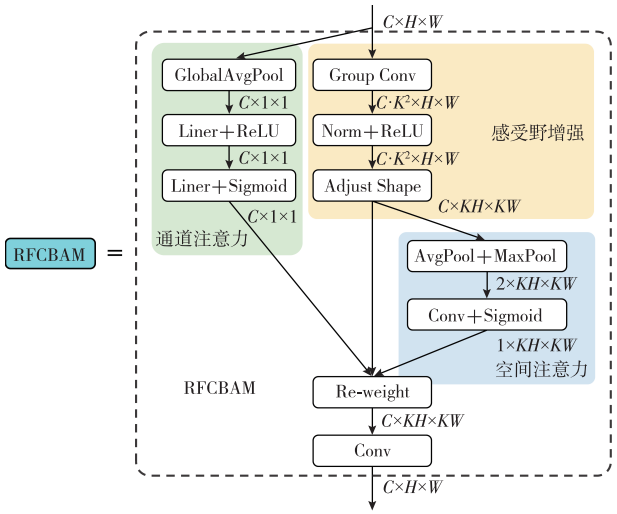


图5 RFCBAM感受野注意力卷积模块

具体而言，RFCBAMConv首先构建感受野空间特征图F_g，随后引入通道与空间双重注意力机制（为了减少计算开销，RFCBAM采用SE注意力[13]取代RFCBAM中的CAM）。其中SE通道注意力的计算公式为：

$$M_c(x)=\sigma\left\{\mathrm{MLP}_2\left\{\mathrm{ReLU}\left\{\mathrm{MLP}_1\left[\mathrm{GAP}(x)\right]\right\}\right\}\right\} \quad (2)$$

式中，σ表示Sigmoid函数，GAP为全局池化操作，MLP_1是将通道数从C压缩到C/r的全连接层，MLP_2是将通道数从C/r恢复到C的全连接层。空间注意力机制的原理如下：

$$M_s(x)=\sigma\left\{\mathrm{Conv}_{k\times k}\left\{\left[\mathrm{AvgPool}(x);\ \mathrm{MaxPool}(x)\right]\right\}\right\} \quad (3)$$

式中，[;]表示通道维度拼接，$\mathrm{Conv}_{k\times k}$为$k\times k$卷积操作。

最后，将生成的注意力权重分别作用于F_g的通道和空间维度，再通过卷积操作得到最终的输出特征图，从而强化网络对重要特征区域的聚焦能力：

$$F_{\mathrm{out}}(x)=\mathrm{Conv}_{k\times k}(F_g\odot M_c\odot M_s) \quad (4)$$

式中，$\odot$表示逐元素相乘。

2.2 内容引导注意力特征融合模块

CGAFusion[14]是一种融合空间注意力、通道注意力与像素级注意力的多模态特征融合机制，在多源特征融合过程中充分挖掘关键语义与结构信息，从而提升特征表示的准确性与鲁棒性。该机制的核心思想是在融合不同特征图的基础上，借助多层次注意力引导策略，对通道、空间及像素维度的重要性进行动态建模与调整。

如图6所示，CGA模块的具体流程如下：首先，将输入的两个特征图X和Y进行逐元素加和，得到初步融合特征图Z；然后分别对Z进行通道注意力和空间注意力建模，用于评估各通道与空间位置的重要性分布；接着将两者相加得到联合响应图，并

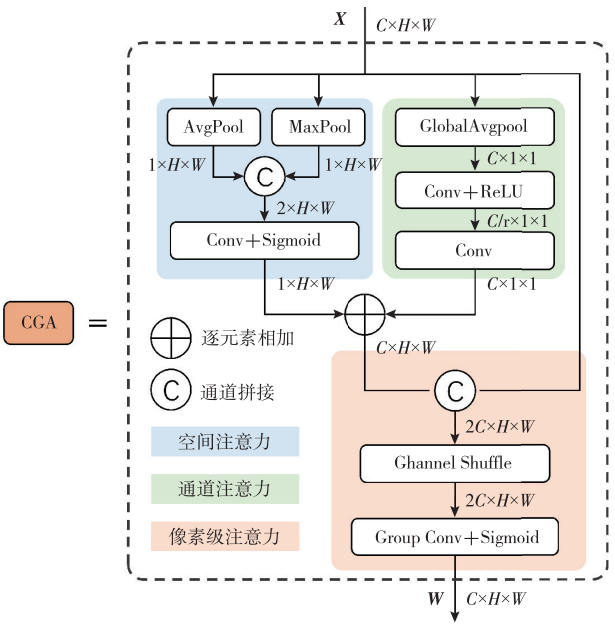


图6 内容引导注意力机制

输入像素级注意力模块，生成用于精细调控的权重图 W。

最后，融合操作通过加权组合方式完成，即：

$$F_{out}=Z+WX+(1-W)Y \tag{5}$$

$$Z=X+Y \tag{6}$$

融合结果还将通过一个1×1卷积进一步整合信息，生成最终输出特征图。

将CGAFusion机制引入YOLO11n模型的检测头前，特征通过P3，P4和P5多尺度分辨率路径被逐级提取，并通过上采样和特征拼接操作进行多尺度融合，在最终融合的特征到达检测头前，通过引入如图7所示的CGAFusion模块，分别再与最初P3，P4，P5层的特征进行内容引导的自适应特征融合，以此生成更为丰富的多尺度特征表示。该过程使不同分辨率的特征有效融合，以提高小目标和细节信息的保留能力。

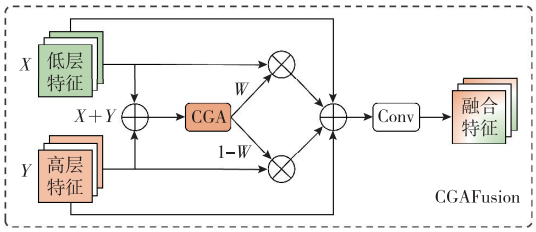


图7　内容引导注意力特征融合模块

2.3　多尺度特征共享检测头

为兼顾目标检测任务中的检测精度与计算效率，引入了MFSD结构用于替代YOLO系列模型中传统的多分支独立计算方式。MFSD通过替换多个尺度的CBS（Conv-BatchNorm-SiLU）模块为统一的CGS（Conv-GroupNorm-SiLU）模块，即共享多尺度特征间卷积路径，实现了特征编码过程中的跨层信息压缩与冗余控制，有效降低参数数量与推理复杂度。其网络结构如图8所示。

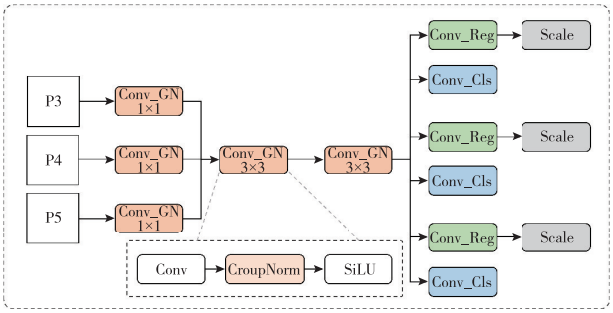


图8　多尺度特征共享检测头

在CBS模块中，BatchNorm通过对每个batch的均值与方差进行标准化来稳定训练。而在大车盲区检测中，由于模型部署在车载边缘设备或工业相机系统中，通常批次大小（batch size）很小（甚至为1），BN的均值方差估计极不稳定，导致模型性能下降。

使用GroupNorm对单个样本内的通道分组做归一化，而不依赖batch内多样本统计，可以在推理阶段，无需切换到推理模式或缓存均值方差，直接处理任意大小输入样本，过程响应快、误差小。

MFSD模块充分发挥了GroupNorm与共享卷积编码的优势，能够在有效保持特征表达能力的同时，显著降低计算开销与模型复杂度。

3　实验结果及分析

3.1　数据集构建

本文数据集由自行采集的盲区图像、nuImages数据集中右后方视角子集以及Tsinghua Daimler[15]数据集共同组成，共7 131张图像。

自行采集的盲区图像针对大车盲区场景进行定向采集，真实反映了盲区人员的环境特征与成像条件。nuImages作为大规模自动驾驶数据集，其右后方视角子集涵盖了大型车辆盲区的典型观察角度，包含多样化的城市交通场景及不同天气和光照条件，有效提升模型的泛化能力。考虑到骑行者样本较少，特引入Tsinghua Daimler数据集，该数据集采集自车载摄像头，涵盖复杂的骑行者行为及多样化姿态，更加契合盲区人员检测的需求。

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pedestrian，cyclist共2类。训练集、验证集和测试集按7∶1∶2的比例随机分配。

3.2　实验环境及评价指标

所使用的实验环境参数如表1所示，迭代次数epoch设置为200，batch size设置为32，优化器采用SGD，初始学习率为0.01，动量为0.937，并启用Mosaic数据增强。

表1　实验环境参数

参数	配置
操作系统	Debian GNU/Linux 12
CPU	Intel Core i5-13600KF（32GB）
GPU	GeForce RTX 4070Super（12GB）
Python版本	3.10.14
Pytorch版本	2.2.2
Cuda版本	12.6

评价指标采用精确率(P)、召回率(R),均值平均精度mAP,模型参数量和每秒10亿次的浮点运算数GFLOPs。其中P,R和mAP的计算公式如下:

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

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

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

$$mAP=\frac{1}{N}\sum_{i=1}^{N}AP(i) \tag{10}$$

式中,N表示检测的类别总数;TP,FP和FN分别表示真阳性(True Positive)、假阳性(False Positive)和假阴性(False Negative)的数量。P和R是在不同置信度阈值下计算得到的精确率和召回率;AP为平均精度,其定义为精确率-召回率曲线下的面积,反映了模型在该类别上的检测性能;mAP是所有类别AP的均值,用于综合评估模型整体的检测效果。

在不同的交并比(IoU)阈值下,mAP的取值会有所不同,常见的指标包括mAP@0.5(IoU大于等于0.5时的mAP)和mAP@0.5:0.95(IoU阈值从0.5到0.95以0.05步长取平均)。

3.3 对比实验

为全面评估所提方法在盲区人员检测任务中的实际性能,选取了当前具有代表性的目标检测算法进行对比实验,包括传统的两阶段检测器(如Faster R-CNN[16])、主流的一阶段检测器(如SSD512[17],RT-DETR-1[18]),以及多个版本的轻量化YOLO系列模型(如YOLOv5n,YOLOv8n等)。实验结果如表2所示,在保持轻量化设计的前提下,本文方法在各项指标中均表现优异。

表2 对比实验结果

网络模型	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%	参数量/M	*GFLOPs*
Faster R-CNN	75.40	60.51	70.70	43.09	41.35	207.9
SSD512	72.72	55.21	60.92	33.32	23.88	87.7
RT-DETR-l	66.87	59.52	65.31	39.23	31.99	103.4
YOLOv5n	77.65	59.61	66.14	38.20	1.76	4.1
YOLOv7-tiny	69.74	56.41	62.70	37.44	6.01	13.1
YOLOv8n	81.99	61.92	70.44	45.16	3.01	8.1
YOLOv9-t	82.91	57.19	67.42	43.85	2.62	10.7
YOLOv10n	79.73	61.90	70.47	44.36	2.27	6.5
YOLO11n(baseline)	82.50	62.20	70.92	45.59	2.58	6.3
YOLOv12n	82.21	58.67	68.58	43.93	2.51	5.8
YOLOv13n	80.48	59.08	67.91	42.52	2.45	6.2
本文	84.76	63.30	73.39	47.64	2.87	8.2

在检测精度与召回率方面,本文方法均取得最优性能,尤其在召回率方面,相较于YOLOv5n,YOLOv7-tiny等早期轻量模型具有显著提升,有效降低了漏检率与误检率,表明该结构能够有效缓解盲区场景中的检测难点,对遮挡目标及小目标具备更强的感知能力。

在mAP@0.5指标上,本文方法相比YOLOv8n和YOLO11n均实现了约2.5%的提升。而在更加严格的mAP@0.5:0.95指标下,本文方法达到47.64%,在所有轻量级模型中排名第一,比未改进的YOLO11n高出2.05%,充分体现出其在多尺度适应性与边界框定位精度方面的优势。

在模型复杂度方面,本文方法的参数量仅为2.87 M,显著低于Faster R-CNN,SSD512和RT-DETR等大型模型;其GFLOPs仅为8.2,与YOLOv8n相当,但在检测精度方面实现了全面超越,展示了良好的效率与性能平衡能力。

综上所述,本文方法兼具高检测精度与低计算开销的优势,在盲区人员检测等复杂场景中具有广泛的实际应用潜力,尤其适合对部署资源敏感的边缘计算与嵌入式系统。

3.4 消融实验

为验证各模块在检测性能中的贡献,本文设计了8组消融实验,分别引入C3k2-RC,CGAFusion与MFSD三个模块进行组合测试,实验结果如表3所示。

可以看出,基线模型(实验1)在未引入任何模块的情况下,mAP@0.5和mAP@0.5:0.95分别为70.92%和45.59%。引入C3k2-RC模块(实验2)后,mAP@0.5提升至72.18%,说明其增强了骨干

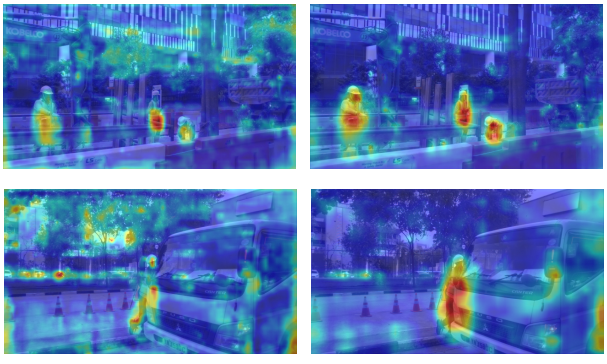
表 3　消融实验结果

实验	模块	*P*/%	*R*/%	mAP@0.5/%	mAP@0.5:0.95/%	参数量/M	*GFLOPs*
1	未引入	82.50	62.20	70.92	45.59	2.58	6.3
2	C3k2-RC	82.44	62.83	72.18	46.45	2.71	6.9
3	CGAFusion	81.24	62.18	71.28	45.72	2.97	8.7
4	MFSD	84.01	61.91	71.42	45.45	2.42	5.6
5	C3k2-RC+CGAFusion	82.17	64.16	72.42	47.04	3.11	9.3
6	C3k2-RC+MFSD	84.61	62.86	72.64	47.54	2.69	7.6
7	CGAFusion+MFSD	82.29	62.28	71.60	46.36	2.73	7.5
8	C3k2-RC+CGAFusion+MFSD	84.76	63.30	73.39	47.64	2.87	8.2

特征的表达能力。引入CGAFusion模块(实验3)也带来了小幅提升,验证了其在上下文信息融合中的有效性;引入MFSD模块(实验4)虽提升有限,但在轻量前提下保持了较优的召回率。进一步将模块组合后,实验5和实验6分别将mAP@0.5:0.95提升至47.04%和47.54%,表明模块间具有良好的互补性。

最终,实验8同时引入全部三个模块,取得最优检测性能,mAP@0.5提升至73.39%,mAP@0.5:0.95提升至47.64%,相较基线模型分别提高2.47%和2.05%,参数量和计算量仅略有增加,验证了所提模块在提升精度的同时保持了较高的效率。

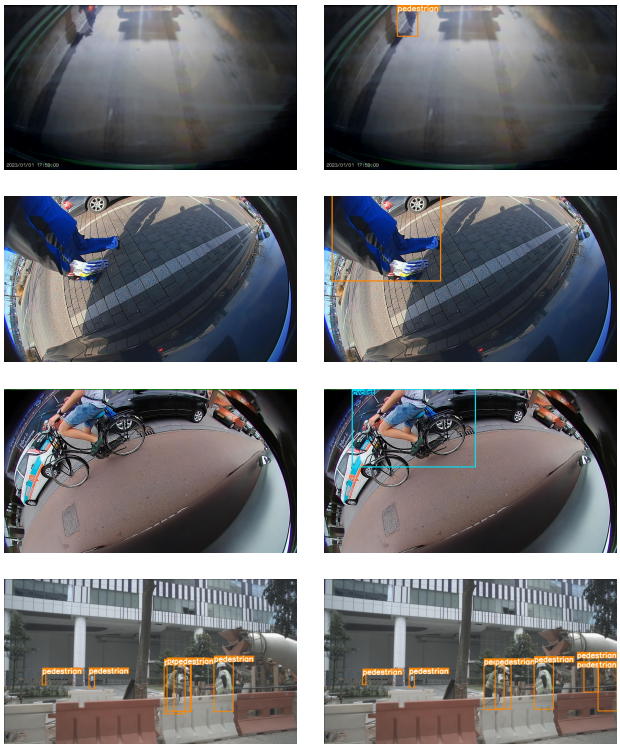
为直观展示改进模块在特征表达与目标感知方面的提升效果,采用Grad-CAM绘制了模型改进前后在多个盲区场景下的检测头输入特征热力图,如图9所示。可以看出,改进模型在各类场景中能够生成更为集中的目标响应区域,尤其在复杂背景或遮挡较严重的情况下,对目标的关注能力明显提升。



(a) YOLO11n　　(b) 改进YOLO11n

图9　热力图对比

从图10中可以看出,基准模型在一些复杂场景下存在漏检、误检及边界框定位不准确的问题。在前三个场景中,基准模型未能检测出被遮挡的行人和骑行者,而改进后的模型能够准确识别几乎所有目标。在第四个场景中,由于背景复杂加上工人受到遮挡较多,基准模型未能全部识别出。



(a) YOLO11n　　(b) 改进YOLO11n

图10　部分预测结果对比

总体而言,改进模型在多目标检测、遮挡场景下的识别、边缘目标感知以及小目标检测等方面均优于基准模型,充分验证了所引入模块在特征增强、注意力机制引导和检测头设计方面的有效性。

4　结论

本文围绕大车盲区中行人与骑行者目标检测任务,提出了一种基于改进YOLO11n的轻量化检测模型,以应对盲区场景存在遮挡、检测精度低、实时性差等挑战。通过引入RFCBAMConv感受野注

意力模块增强对遮挡目标的感知能力，使用CGAFusion模块实现多尺度语义融合，并通过MFSD检测头以减少冗余计算，模型在保持低计算复杂度的同时，实现了精度与实时性的平衡。实验结果表明，该模型在小目标与遮挡目标检测方面表现出较强鲁棒性，具有在智能交通安全领域的广泛应用前景。未来工作将进一步探索模型在多传感器融合与实际车载环境中的应用，以增强其在复杂交通场景下的适应性。

参考文献：

[1] 交通运输部. 2024年交通运输行业发展统计公报[N]. 中国交通报, 2025-06-12(002).
Ministry of Transportation and Communications. Statistical Bulletin on Transportation Industry Development in 2024[N]. China Transportation News, 2025-06-12(002).

[2] Masalov A, Ota J, Corbet H, et al. CyDet: improving camera-based cyclist recognition accuracy with known cycling jersey patterns [C]//2018 IEEE Intelligent Vehicles Symposium (IV), 2018: 2143-2149.

[3] Wang K, Zhou W. Pedestrian and cyclist detection based on deep neural network fast R-CNN[J]. International Journal of Advanced Robotic Systems, 2019, 16 (2): 1729881419829651.

[4] Girshick R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision, 2015: 1440-1448.

[5] 韦 昊. 基于深度学习的货车盲区障碍目标检测算法研究[D]. 长沙: 湖南大学, 2020: 26-41.
Wei Hao. Research on the detection algorithm of obstacle in van blind spot based on deep learning[D]. Changsha: Hunan University, 2020: 26-41.

[6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[7] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint, 2018: arXiv:1804.02767.

[8] Wang C Y, Yeh I H, Mark Liao H Y. YOLOv9: learning what you want toLearn using programmable gradient information[C]// Computer Vision - ECCV 2024, 2025: 1-21.

[9] Chen H, Chen K, Ding G, et al. YOLOv10: real-time end-to-end object detection[C]//Advances in Neural Information Processing Systems 37. Neural Information Processing Systems Foundation, 2024: 107984-108011.

[10] Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements[J]. arXiv preprint, 2024: arXiv: 2410.17725.

[11] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision-ECCV 2018, 2018: 3-19.

[12] Zhang X, Liu C, Yang D, et al. RFAConv: Innovating spatial attention and standard convolutional operation [J]. arXiv preprint, 2023 arXiv:2304.03198.

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

[14] Chen Z, He Z, Lu Z M. DEA-net: single image dehazing based on detail-enhanced convolution and content-guided attention[J]. IEEE Transactions on Image Processing, 2024, 33: 1002-1015.

[15] Li X, Flohr F, Yang Y, et al. A new benchmark for vision-based cyclist detection[C]//2016 IEEE Intelligent Vehicles Symposium, 2016: 1028-1033.

[16] Ren S, He K, 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, 2017, 39(6): 1137-1149.

[17] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector [C]//Computer Vision-ECCV 2016, 2016: 21-37.

[18] Zhao Y, Lv W, Xu S, et al. DETRs beat YOLOs on real-time object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.

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；

余自安(1974—)，男，河南省光山县人，高级工程师，主要从事机械设计及理论研究。