光电技术及应用

DOI：10.16818/j.issn1001－5868.20251114002

基于双主干网络的分心驾驶行为检测方法

代少升[1]，何帅兵[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： 针对现有分心驾驶行为检测算法检测精度低、计算量大等问题，文章提出一种基于YOLOv11n改进的双主干网络（YOLO-DCC）。首先，该算法使用两条不同的轻量化主干替换YOLOv11n的原始主干，在保持轻量化的同时，增强了模型的特征提取能力。随后，引入跨注意力融合模块（CAFM）在通道与空间维度上实现跨分支特征的动态交互与加权融合，进一步提升了关键特征的表达能力。最后，在颈部网络中引入上下文引导融合模块（CGFM），通过聚合局部、周边与全局上下文信息，优化了多尺度特征的融合效果，提升了模型的检测精度。在StateFarm数据集上进行实验，结果表明所提算法将*mAP*@0.5提升到了95.7%，相比于YOLOv11n算法检测精度提高了3.1%，计算量减少了7.94%，能够有效检测驾驶员的分心驾驶行为。

关键词： YOLOv11n；双主干网络；特征融合；分心检测

中图分类号： TP391.41　**文章编号：** 1001－5868(2026)01－0172－08

Dual-Backbone Network-Based Method for Distracted Driving Behavior Detection

DAI Shaosheng[1], HE Shuaibing[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: To overcome the limitations of low detection accuracy and high computational load in existing distracted driving detection algorithms, in this study, an improved model based on the YOLOv11n architecture—namely, YOLO-DCC—is proposed. First, the algorithm replaces the original YOLOv11n backbone with two distinct lightweight backbones, which enhance its feature extraction capabilities while maintaining a lightweight architecture. Subsequently, the cross-attention fusion modules (CAFMs) are employed to achieve dynamic interaction and weighted fusion of cross-branch features along both channel and spatial dimensions, and thus enhance the representation of key features. Finally, context guided fusion modules (CGFMs) are introduced into the neck network to optimize the fusion of multiscale features via the integration of local, surrounding, and global contextual information, which improve the detection accuracy of the model. Experiments performed on the StateFarm dataset show that the proposed algorithm achieves a mean average precision (*mAP*) of 95.7% at an IoU threshold of 0.5 while reducing the computational load by 7.94%. This *mAP* represents a 3.1% improvement over that of the baseline YOLOv11n model, demonstrating the effectiveness of the proposed model in detecting various distracted driving behaviors.

Key words: YOLOv11n; dual-backbone network; feature fusion; distracted driving detection

收稿日期：2025－11－14.

*通信作者：余自安

0　引言

近年来，我国交通事业发展迅速，汽车保有量逐年递增，交通事故频发。据统计，仅在2023年，全国就发生了约25万起交通事故，造成了大量的财产损失和人员伤亡[1]。研究表明，分心驾驶是导致交通事故的主要原因之一[2]。因此，设计一种精准且高效的分心驾驶行为检测算法，对于保障道路安全、减少交通事故具有重要意义。

目前，分心驾驶行为检测方法可以分为三类：基于生理信号的检测方法、基于车辆信息的检测方法和基于计算机视觉的检测方法。其中，基于生理信号的检测方法依赖侵入式传感器采集驾驶员的肌电、心电和脑电等生理信号[3-5]。然而，这些传感器不仅价格昂贵，且佩戴时可能引起驾驶员的不适，影响驾驶体验。基于车辆信息的检测方法通过监测车辆速度、加速度和方向盘转动角度等信息，间接推测驾驶员的分心行为[6]。但这种方法容易受到驾驶员个体驾驶风格的影响，且当车辆状态出现显著变化时往往已错失最佳预警时机，难以防范事故的发生。相比之下，基于计算机视觉的检测方法能够在不干扰驾驶员正常驾驶的情况下，实时监测驾驶员的行为。因此，本文主要研究基于计算机视觉的分心驾驶行为检测方法。

早期，基于计算机视觉的分心驾驶行为检测方法主要依赖传统机器学习技术，如决策树、随机森林和支持向量机等。例如，Berri等[7]提出一种基于SVM的模型，用于识别驾驶过程中使用手机的行为，在自建数据集上取得了91.57%的准确率。Zhao等[8]采用轮廓波变换（Contourlet Transform）提取特征，并结合随机森林分类器，实现了对正常驾驶、换挡、饮食和打电话四种行为的有效分类，准确率均超过88%。尽管这些方法在特定场景下表现良好，但泛化能力较差，难以适应复杂多变的真实驾驶环境。

近年来，随着深度学习在多个领域取得巨大成功，越来越多的研究者将其应用于分心驾驶行为检测任务。例如，Ezzouhri等[9]通过深度学习模型预先提取驾驶员身体部位，再进行分心检测，在自建数据集上达到了96%的准确率。Du等[10]对YOLOv8模型进行优化，通过引入GhostC2f模块和SimAM注意力机制，在降低模型计算量和大小的同时提升了检测精度。Zhang等[11]设计了一种基于Swin Transformer与高区分度特征学习策略的模型，在AUC-DDD数据集上达到了95.66%的准确率。此外，为融合不同网络架构的优势，高尚兵等[12]设计了一种CNN与ViT并行的双分支模型，并利用双向特征交互模块实现了局部与全局特征的有效融合，在自建数据集上取得了97.18%的准确率。

尽管深度学习在分心驾驶行为检测方面取得了显著进展，但现有方法仍存在检测精度低、模型体积庞大、计算复杂度高等问题。为此，本文基于YOLOv11n[13]设计了一种分心驾驶行为检测网络YOLO-DCC（YOLO Dual-Backbone with CAFM and CGFM）。该网络采用两条独立的轻量化主干共同提取特征，并在其中一条主干中嵌入了边缘信息增强模块（Edge Information Enhancement Module，EIEM），以突出边缘细节、减少特征冗余。随后，引入跨注意力融合模块（Cross-Attention Fusion Module，CAFM）[14]替代传统的特征拼接操作，实现跨分支特征的动态交互与加权融合，充分挖掘并整合两个主干提取的信息。最后，借鉴CGNet[15]中的思想，设计了上下文引导融合模块（Context Guided Fusion Module，CGFM），并将其集成到颈部网络中，以提升多尺度特征的融合效果。实验结果表明，YOLO-DCC在保持轻量化的同时，提升了分心驾驶行为的检测精度。

1　YOLOv11网络

YOLOv11是Ultralytics团队在YOLOv8基础上开发的最新一代目标检测模型。其网络结构如图1所示，主要包括主干网络（Backbone），颈部网络（Neck）和检测头（Head）三部分。主干网络通过一系列卷积层从输入图像中提取多尺度特征，颈部网络随后对这些特征进行融合，最后检测头利用这些融合后的特征进行边界框回归和类别预测。

YOLOv11的一项重要改进是引入C3k2模块，用以取代YOLOv8中的C2f模块。C3k2模块使用两个小卷积核替代原来的大卷积核，在保持模型性能的同时加快了计算速度。在主干网络部分，YOLOv11保留了先前版本的SPPF模块，并在其后新增了一个带空间注意力机制的C2PSA模块，使模型能够更精准地聚焦图像中的关键区域。此外，YOLOv11将深度可分离卷积（DWConv）应用于检

测头的分类分支中，降低了模型的计算开销。

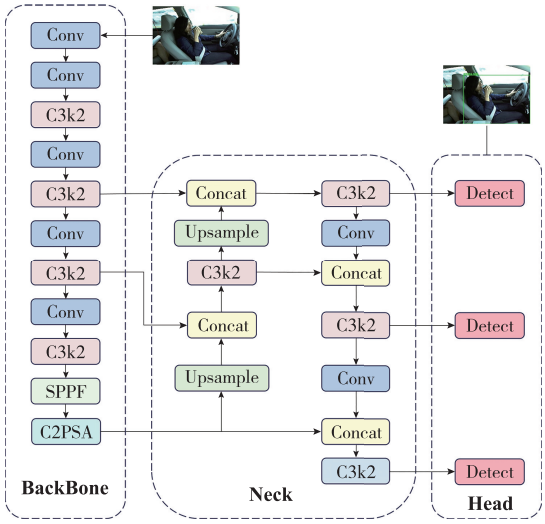


图1 YOLOv11网络结构图

YOLOv11提供了从轻量级到高性能的多种变体(n,s,m,l,x)，以满足不同应用场景的需求。本文选择其中参数量和计算量最小的YOLOv11n作为基线模型，以适配资源受限的车载嵌入式设备。

2 YOLO-DCC网络

如图2所示，本文构建的YOLO-DCC模型采用异构的双主干网络，以提取差异化的特征。为保持模型的轻量化，两个主干均沿用了YOLOv11n的轻量化设计，并将其主干通道数进行了适当缩减。在将特征传入颈部网络前，模型通过跨注意力融合模块(CAFM)筛选并整合两条主干提取的特征，以提升关键特征的表达能力。随后，在颈部网络中，YOLO-DCC引入上下文引导融合模块(CGFM)来更有效地融合不同尺度的特征。最后，考虑到驾驶员通常占据画面较大区域，属于中、大尺度目标，模型移除了P3小目标检测头，降低了模型的计算复杂度，提升了模型的整体运行效率。

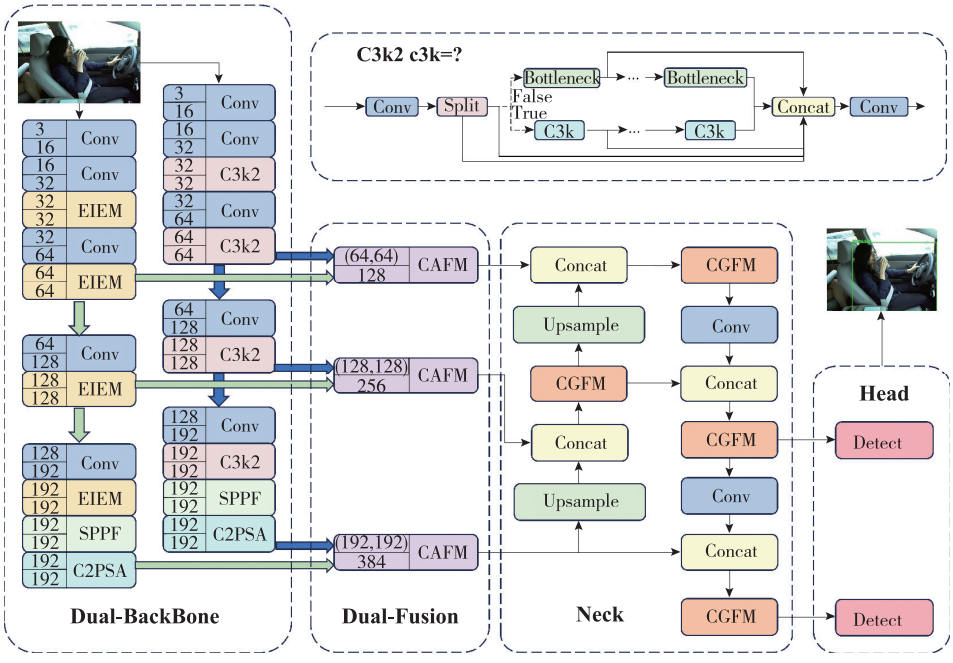


图2 YOLO-DCC网络结构图

2.1 边缘信息增强模块

为了增强网络对图像边缘信息的感知能力，本文在其中一个主干中集成了边缘信息增强模块(EIEM)，其整体结构见图3。通过这一设计，两个主干能够分别聚焦于不同的特征，从而减少信息冗余，得到更全面、更具判别力的特征。

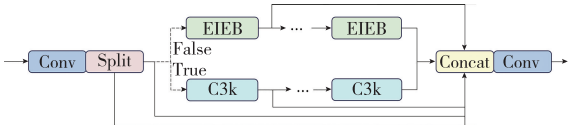


图3 EIEM结构示意图

EIEM由多个边缘信息增强块(EIEB)构成。如图4所示，每个EIEB通过两条并行分支协同工作，以分别提取图像的边缘与空间信息。其中，SobelConv分支利用公式(1)对输入图像 F_{in} 进行卷积操作，以近似获得图像在水平和垂直方向上的梯度值 G_x 和 G_y。接着，通过公式(2)计算得到图像中

每个像素处的边缘强度 G，从而突出图像的高频边缘轮廓。

$$G_x=\begin{bmatrix}-1 & 0 & +1\\ -2 & 0 & +2\\ -1 & 0 & +1\end{bmatrix}*F_{\text{in}},\ G_y=\begin{bmatrix}-1 & -2 & -1\\ 0 & 0 & 0\\ +1 & +2 & +1\end{bmatrix}*F_{\text{in}} \tag{1}$$

$$G=|G_x|+|G_y| \tag{2}$$

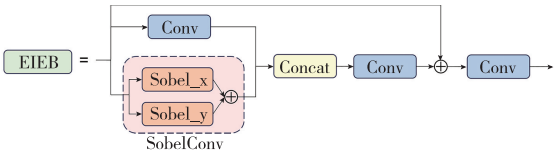


图 4　EIEB 结构示意图

除了边缘信息，图像中的空间信息也同样重要，因此 EIEB 通过一个额外的 Conv 分支来保留重要的空间信息。EIEB 将两条分支的输出在通道维度上进行拼接并通过后续的卷积操作进行融合，最终生成兼具边缘细节和空间信息的特征。此外，EIEB 还引入了残差连接分支，将原始输入信息直接传递给后续层，有效防止了梯度消失问题。

2.2　跨注意力融合模块

在双主干网络中，简单的拼接或相加操作难以有效融合双主干提取的差异化特征。为此，引入跨注意力融合模块（CAFM），其借鉴了 CBAM[16] 与 Cross Attention[17] 等经典注意力机制，通过在通道与空间维度上建立跨分支的动态依赖关系，实现了一种选择性的特征融合，从而提升了关键特征的表达能力。

如图 5 所示，CAFM 的整体结构由两个核心组件串联组成：通道交叉注意力模块和空间注意力融合模块。首先，对于输入特征图 $F^c\in R^{C\times H\times W}$，分别应用全局平均池化与全局最大池化来生成两个通道特征向量。然后，将这两个通道特征向量通过一个共享的两层 MLP 网络进行处理。最后，将 MLP 输出的两个特征向量相加后，再通过 Sigmoid 激活函数便可得到最终的通道注意力权重 M^c。其计算式为：

$$M^c=\sigma\left[\text{MLP}(F^c_{\text{avgpool}})+\text{MLP}(F^c_{\text{maxpool}})\right] \tag{3}$$

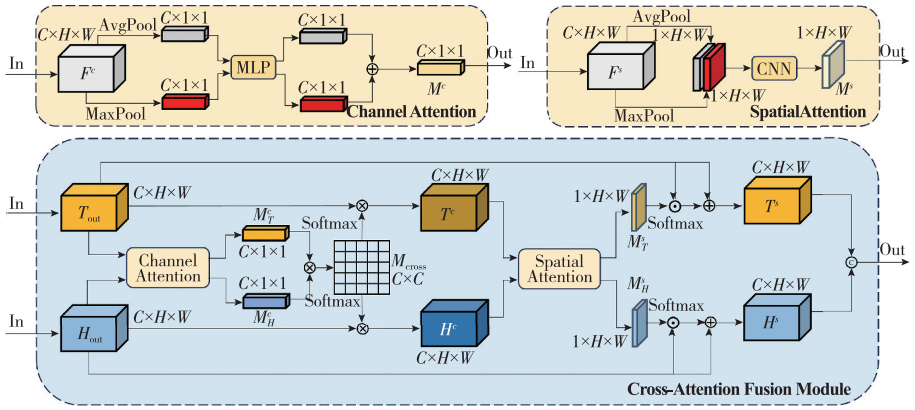


图 5　CAFM 结构示意图

通过上述方法，分别计算出两条主干输入特征图 T_{out} 和 H_{out} 各自的通道注意力权重 M_T^c 和 M_H^c，并将它们相乘得到一个形状为 $C\times C$ 的跨分支通道注意力矩阵 $M_{\text{cross}}=M_T^c(M_H^c)^{\text{T}}$。最后，利用该交叉矩阵对原始特征图进行增强，如公式（4）所示，从而实现跨分支的通道信息交互。

$$\begin{aligned}T^c&=\text{Softmax}(M_{\text{cross}})T_{\text{out}}\\ H^c&=\text{Softmax}(M_{\text{cross}}^{\text{T}})H_{\text{out}}\end{aligned} \tag{4}$$

在完成跨分支的通道信息交互后，模块进一步引入空间注意力机制，以聚焦于空间上的关键区域。对于输入特征图 $F^s\in R^{C\times H\times W}$，先对其在通道维度上分别做最大池化和平均池化，得到两个形状为 $1\times H\times W$ 的空间特征图，并将它们在通道维度上拼接。然后，使用一个共享的卷积层（卷积核大小为 $k\times k$，$k=3$）对拼接后的空间特征图进行卷积运算，生成最终的空间注意力权重 M^s。其计算式为：

$$M^s=f^{k\times k}\left\{[\text{AvgPool}(F^s)\oplus\text{MaxPool}(F^s)]\right\} \tag{5}$$

最后，如公式（6）所示，将学习到的空间注意力权重 M_T^s 和 M_H^s 施加于原始输入特征图 T_{out} 和 H_{out}，并将加权后的特征图与原始输入特征图相加，得到最终融合后的特征图。

$$T^s = \mathrm{Softmax}(M_T^s) \odot T_{\mathrm{out}} + T_{\mathrm{out}}$$
$$H^s = \mathrm{Softmax}(M_H^s) \odot H_{\mathrm{out}} + H_{\mathrm{out}} \quad (6)$$

2.3 上下文引导融合模块

为了更有效地融合双主干网络提取的多尺度特征，引入CGNet中的上下文引导模块(Context Guided Block，CG Block)，并用其替换了C3k2模块中的Bottleneck结构，从而构建了一个全新的上下文引导融合模块(CGFM)。

CG Block的设计灵感来源于人类视觉系统。如图6(a)所示，若人类视觉系统只关注黄色区域，很难准确区分驾驶员的分心驾驶行为。在图6(b)中，当视觉系统同时获取到黄色区域及其周围的上下文(红色区域)时，能更容易区分驾驶员的分心驾驶行为。在图6(c)中，若视觉系统在关注黄色区域及其周边上下文(红色区域)的同时，还能够捕捉到整幅场景的全局上下文(紫色区域)，则能够以更高的置信度对驾驶员的分心驾驶行为进行分类。基于以上分析，通过融合局部特征、周围上下文和全局上下文，网络能够从不同层次的上下文中提取出更多有价值的特征。如图6(d)所示，CG Block主要由四个模块组成：局部特征提取器、周边上下文提取器、联合特征提取器和全局上下文提取器。CG Block主要包括以下两个步骤：

第一步，分别采用局部特征提取器和周边上下文提取器学习局部特征和对应的周边上下文信息。局部特征提取器由一个3×3的标准卷积层构成，用于从黄色区域中提取细粒度的局部特征；而周边上下文提取器由一个3×3的空洞卷积层构成，凭借较大的感受野高效地捕获周边上下文信息。随后，联合特征提取器将局部特征提取器与周边上下文提取器的输出在通道维度上进行拼接，并通过批量归一化和PReLU激活后，得到联合特征。

第二步，通过全局上下文提取器提取全局上下文信息，以进一步增强联合特征。受SENet[18]启发，全局上下文信息被视为一个通道加权向量，用于在通道维度对联合特征进行精细调整，从而增强有用特征、抑制无效信息。在实际操作中，全局上下文提取器先通过全局平均池化汇聚全局信息，再经多层感知机提取全局上下文信息，最后将提取到的全局上下文信息作为通道注意力权重应用于联合特征，得到最终的输出特征。

此外，CG Block还引入全局残差学习，将原始输入特征与全局上下文提取器的输出相加。这一设计有助于网络学习更复杂的特征并改善训练期间梯度的反向传播效果。

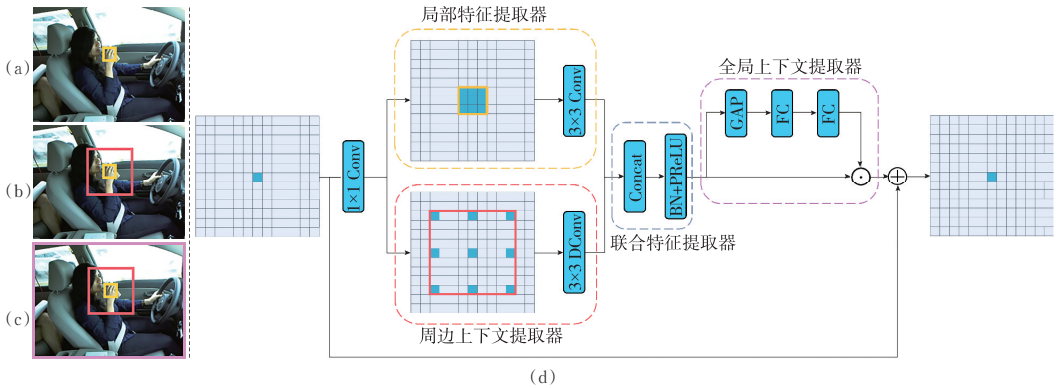


图6 CG Block结构示意图

3 实验结果与分析

3.1 数据集介绍

为了评估所提方法的性能，采用StateFarm数据集，该数据集包含26名不同驾驶员的驾驶记录，共计22 424张分辨率为640×480的彩色图像。如图7所示，StateFarm数据集将驾驶行为划分为10个类别，其中一个类别表示安全驾驶，其余九个类别对应常见的分心驾驶行为(如发短信、打电话等)，每类图像约2 000张，确保了数据的丰富性与多样性(具体数量见表1)。为了评估模型的泛化能力，将数据集按照驾驶员身份随机划分为训练集和验证集。训练集由23名驾驶员的图像组成，验证集则由另外3名驾驶员的图像组成。



（a）安全驾驶　（b）右手发短信　（c）右手打电话　（d）左手发短信　（e）左手打电话

（f）操作收音机　（g）喝水　（h）伸手到后面　（i）整理妆容　（j）与乘客交谈

图 7　StateFarm 数据集行为分类示意图

表 1　StateFarm 数据集详细信息

类别	名称	图片数量
a	安全驾驶	2 489
b	右手发短信	2 267
c	右手打电话	2 317
d	左手发短信	2 346
e	左手打电话	2 326
f	操作收音机	2 312
g	喝水	2 325
h	伸手到后面	2 002
i	整理妆容	1 911
j	与乘客交谈	2 129

3.2　实验环境与评价指标

本文实验环境如表 2 所示。实验的超参数设置如下：初始学习率为 0.01，训练总轮数为 100，动量为 0.937，权重衰减系数为 0.000 5，批大小为 16。

表 2　实验环境

参数	配置
操作系统	Windows 11 专业版
CPU	AMD Ryzen 5 5600X（32 GB）
GPU	NVIDIA GeForce RTX 3080（12 GB）
Python 版本	3.10.16
Pytorch 版本	2.7.0
Cuda 版本	12.8

通过参数量和计算量来评估模型的计算复杂度和资源开销。此外，采用精确率（P）、召回率（R）和平均精度均值（mAP）来衡量模型的检测精度。具体来说，精确率是所有预测为正样本的结果中，预测正确的比例；召回率是指在所有真实正样本中，成功被模型检出的比例；mAP 作为综合性指标，能够全面地衡量模型在所有类别上的整体检测性能。这些指标的计算公式如下：

$$\begin{cases} P = TP/(TP + FP) \\ R = TP/(TP + FN) \\ mAP = \sum AP/N \end{cases} \tag{7}$$

3.3　对比实验

为了全面评估改进后的 YOLO-DCC 模型，在 StateFarm 数据集上将其与当前主流的目标检测模型进行对比，实验结果如表 3 所示。结果表明，YOLO-DCC 在精确率和召回率两项指标上表现最佳，同时 $mAP@0.5$ 达到了 95.7%，优于同系列的 YOLOv8n 等模型。尽管 RTDETR-r18 的 $mAP@0.5$ 略高于 YOLO-DCC，但其 60.4 *GFLOPs* 的计算量对部署设备提出了更高的算力要求。虽然，RTMDet-tiny 的精确率达到了 93.2%，但其召回率偏低，存在漏报风险。相较于其他模型，Faster R-CNN 整体表现较差，且由于其双阶段结构计算量十分庞大。综合来看，YOLO-DCC 在保持较高检测精度的同时，参数量仅 3.03 M，计算量也降至 5.8 *GFLOPs*，更适合车载边缘设备的部署。

表 3　对比实验结果

网络模型	P/%	R/%	$mAP@0.5$/%	*Params*/M	*GFLOPs*
Faster-RCNN-r50	86.9	85.7	91.3	41.39	71.7
RTDETR-r18	92.5	92.4	96.4	19.10	60.4
RTMDet-tiny	93.2	88.7	91.8	4.87	8.0
YOLOv5n	89.1	86.2	89.0	1.77	4.2
YOLOv8n	93.6	90.3	92.8	3.01	8.1
YOLOv11n	92.3	89.9	92.6	2.58	6.3
YOLOv12n	91.4	88.0	91.7	2.56	6.3
本文	94.8	94.6	95.7	3.03	5.8

3.4　消融实验

为验证模型各项改进的有效性，以 YOLOv11n 为基准模型，在 StateFarm 数据集上进行了一系列

消融实验。

3.4.1 双主干网络设计改进结果

首先，将通道压缩后的YOLOv11n主干简单复制形成YOLO-C3k2双主干网络。由表4可知，YOLO-C3k2的*mAP*@0.5相较于YOLOv11n下降了0.4%，检测效果不理想。在此基础上，本文在YOLO-C3k2的主干1中嵌入边缘信息增强模块，构成YOLO-D-P3模型。实验结果显示，YOLO-D-P3的*mAP*@0.5和召回率分别比YOLOv11n提升了1.5%和1.1%，验证了该模块在减少特征冗余、强化边缘细节方面的有效性。最后考虑到驾驶员通常占据画面较大区域，移除了YOLO-D-P3的小目标检测头，得到YOLO-D模型。去掉小目标检测头后，YOLO-D计算量下降了1.1 *GFLOPs*，精确率略降而*mAP*@0.5基本持平，整体性能未受显著影响。

表4 单主干与双主干网络检测效果对比

模型	*P*/%	*R*/%	*mAP*@0.5/%	*Params*/M	*GFLOPs*
YOLOv11n	92.3	89.9	92.6	2.58	6.3
YOLO-C3k2	91.4	89.1	92.2	2.97	7.0
YOLO-D-P3	92.7	91.0	94.1	2.94	7.0
YOLO-D	92.3	91.3	94.2	2.90	5.9

如图8所示，在主干1中引入边缘信息增强模块后，主干1能够有效地关注到图像中的边缘细节。主干2虽然能够精确地关注到驾驶员的手部和面部区域，但其纹理细节丢失较为严重。通过结合两种不同的特征提取结构，YOLO-D有效缓解了双主干网络的特征冗余问题，增强了主干间的互补性，提升了模型对分心驾驶行为的检测性能。



（a）主干1 （b）主干2

图8 双主干网络特征图可视化

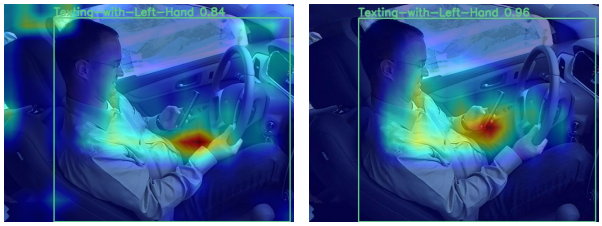
3.4.2 双主干特征融合优化结果

本文在YOLO-D模型的基础上，采用跨注意力融合模块替代了简单的拼接操作去融合两条主干提取的特征，构建了YOLO-DC模型。根据表5可知，跨注意力融合模块提升了模型的各项评价指标，其中精确率提升最为明显，较YOLO-D模型提高了0.9%。

表5 双主干特征融合优化结果

模型	*P*/%	*R*/%	*mAP*@0.5/%	*Params*/M	*GFLOPs*
YOLOv11n	92.3	89.9	92.6	2.58	6.3
YOLO-D	92.3	91.3	94.2	2.90	5.9
YOLO-DC	93.2	92.0	94.6	3.09	6.1

此外，为更直观地分析该模块对模型关注区域的影响，利用Grad-CAM[19]对改进前后的模型关注区域进行了可视化分析，结果如图9所示。可以看出，YOLO-D模型的注意力分布较为分散，不仅关注驾驶员，还受到仪表盘、座椅等背景区域的干扰；而引入跨注意力融合模块后，YOLO-DC模型的注意力更加集中于驾驶员本体以及关键动作区域，有效减少了背景的干扰。



（a）YOLO-D （b）YOLO-DC

图9 Grad-CAM可视化结果对比

3.4.3 颈部网络优化结果

为验证上下文引导融合模块的有效性，在YOLO-DC的颈部网络中集成该模块，形成了YOLO-DCC模型。由表6可知，在与YOLO-DC模型参数量和计算量相近的情况下，YOLO-DCC的*mAP*@0.5、精确率和召回率分别增长了1.1%，1.6%和2.6%。该结果表明，上下文引导融合模块通过联合建模局部特征、周边上下文和全局上下文信息，能够更充分地融合深层与浅层特征，为分心驾驶行为的精确识别提供更丰富的上下文支持。

表6 颈部网络优化结果

模型	*P*/%	*R*/%	*mAP*@0.5/%	*Params*/M	*GFLOPs*
YOLOv11n	92.3	89.9	92.6	2.58	6.3
YOLO-DC	93.2	92.0	94.6	3.09	6.1
YOLO-DCC	94.8	94.6	95.7	3.03	5.8

4 总结

本文提出一种基于双主干网络的分心驾驶行为检测模型YOLO-DCC。首先，该模型采用两条不同结构的轻量化主干提取图像中的差异化特征，有效缓解了双主干网络中常见的特征冗余问题；同

时，引入跨注意力融合模块，在通道和空间维度实现双主干特征的深度交互与加权融合，增强了关键特征的表达能力；此外，在颈部网络中集成上下文引导融合模块，优化多尺度特征的融合效果，进一步提升模型对分心驾驶行为的感知能力。在StateFarm数据集上的实验结果表明，YOLO-DCC在精确率、召回率和*mAP*@0.5方面较原始YOLOv11n模型分别提升了2.5%，4.7%和3.1%，在保持轻量化的同时提高了分心驾驶行为的检测精度。

参考文献：

[1] Debsi A, Ling G, Al-Mahbashi M, et al. Driver distraction and fatigue detection in images using ME-YOLOv8 algorithm [J]. IET Intelligent Transport Systems, 2024, 18(10): 1910-1930.

[2] Sajid Hasan A, Jalayer M, Heitmann E, et al. Distracted driving crashes: a review on data collection, analysis, and crash prevention methods [J]. Transportation Research Record: Journal of the Transportation Research Board, 2022, 2676(8): 423-434.

[3] Fan Y, Gu F, Wang J, et al. SafeDriving: an effective abnormal driving behavior detection system based on EMG signals[J]. IEEE Internet of Things Journal, 2022, 9(14): 12338-12350.

[4] Deshmukh S V, Dehzangi O. Characterization and identification of driver distraction during naturalistic driving: an analysis of ECG dynamics [C]//Advances in Body Area Networks I, 2019: 1-13.

[5] Li G, Yan W, Li S, et al. A temporal - spatial deep learning approach for driver distraction detection based on EEG signals [J]. IEEE Transactions on Automation Science and Engineering, 2022, 19(4): 2665-2677.

[6] Zhang W, Zhang H. Research on distracted driving identification of truck drivers based on simulated driving experiment [J]. IOP Conference Series: Earth and Environmental Science, 2021, 638(1): 012039.

[7] Berri R A, Silva A G, Parpinelli R S, et al. A pattern recognition system for detecting use of mobile phones while driving [C]//2014 International Conference on Computer Vision Theory and Applications (VISAPP), 2015: 411-418.

[8] Zhao C H, Zhang B L, He J, et al. Recognition of driving postures by contourlet transform and random forests[J]. IET Intelligent Transport Systems, 2012, 6(2): 161-168.

[9] Ezzouhri A, Charouh Z, Ghogho M, et al. Robust deep learning-based driver distraction detection and classification[J]. IEEE Access, 2021, 9: 168080-168092.

[10] Du Y, Liu X, Yi Y, et al. Optimizing road safety: advancements in lightweight YOLOv8 models and GhostC2f design for real-time distracted driving detection[J]. Sensors, 2023, 23(21): 8844.

[11] Zhang Z, Yang L, Lv C. Highly discriminative driver distraction detection method based on swin transformer[J]. Vehicles, 2024, 6(1): 140-156.

[12] 高尚兵，张莹莹，王　腾，等. 基于BiViTNet的轻量级驾驶员分心行为检测方法[J]. 重庆交通大学学报(自然科学版)，2024，43(2)：57-64.
Gao Shangbing, Zhang Yingying, Wang Teng, et al. A lightweight driver distraction behavior detection method based on BiViTNet[J]. Journal of Chongqing Jiaotong University (Natural Sciences), 2024, 43(2): 57-64.

[13] Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements[J]. arXiv preprint, arXiv: 2410.17725, 2024. .

[14] Zhou H, Luo F, Zhuang H, et al. Attention multihop graph and multiscale convolutional fusion network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 5508614.

[15] Wu T, Tang S, Zhang R, et al. CGNet: a light-weight context guided network for semantic segmentation[J]. IEEE Transactions on Image Processing, 2021, 30: 1169-1179.

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

[17] Hou R, Chang H, Ma B, et al. Cross attention network for few-shot classification [J]. arXiv preprint, arXiv: 1910.07677,2019.

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

[19] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks *via* gradient-based localization [J]. International Journal of Computer Vision, 2020, 128(2): 336-359.

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嵌入式系统的设计与开发；

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。