

YOLO-TKD: 用于自动驾驶的轻量化时序交互信号检测模型

胡振斌¹, 姚肖婷¹, 席治远¹, 王晓^{1,2}

(1. 安徽大学人工智能学院, 安徽 合肥 230601;

2. 安徽大学具身智能研究院, 安徽 合肥 230031)

摘要: 在自动驾驶的感知任务中, 准确理解交通参与者发出的交互信号是实现安全拟人化决策的关键。当前高精度模型难以实现实时交互信号检测, 且普遍缺乏对转向灯闪烁、行人手势等动态时序信号的有效建模, 导致复杂场景下漏检误判频繁发生, 构成安全隐患。为此, 提出一种面向自动驾驶边缘设备的轻量-时序-蒸馏协同优化框架模型, 即轻量化时序知识蒸馏YOLO (YOLO-TKD), 旨在解决交互信号检测中时序建模需求与计算资源受限的矛盾。该模型以YOLOv8n为基线, 首先采用MobileNetV2作为主干网络以压缩模型规模; 继而引入时序卷积网络模块(TCN), 捕捉连续帧之间的动态依赖关系; 最后, 结合知识蒸馏策略(KD), 保持其识别性能。实验结果表明, 本方法的精确率、召回率和平均精度均值分别达到87.5%、86.4%和84.3%, 相较于基线YOLOv8n分别提升了3.2%、2.6%和1.1%。同时, 模型参数量减少26.3%。所提YOLO-TKD在显著降低资源消耗的同时, 有效提升了交互信号检测的精度与时序鲁棒性, 为自动驾驶边缘设备提供了可行的技术路径。

关键词: 交互信号检测; 轻量化模型; 时序卷积网络; 知识蒸馏; YOLOv8

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.2096-6652.ZNKX26020

YOLO-TKD: lightweight temporal interactive signal detection model for autonomous driving

Hu Zhenxiao¹, Yao Xiaoting¹, Xi Zhiyuan¹, Wang Xiao^{1,2}

1. School of Artificial Intelligence, Anhui University, Hefei 230601, China

2. Anhui University Embodied Intelligence Research Institute, Hefei 230031, China

Abstract: In autonomous driving perception, accurate understanding of interaction signals is the key to safe, human-like decisions. Real-time detection of these signals is found difficult to achieve with current high-precision models. Effective modeling of dynamic temporal signals was generally lacking. However, effective modeling of dynamic temporal signals, such as turn signal flashes and pedestrian gestures. As a result, missed detections and false alarms occur frequently in complex scenarios, creating safety risks. To address this, a lightweight-temporal-knowledge distillation collaborative optimization model for autonomous driving edge devices was proposed, named lightweight temporal knowledge-distilled YOLO (YOLO-TKD). The model was designed to resolve the conflict between the need for temporal modeling in interactive signal detection and the constraints of computational resources. Firstly, the YOLOv8n baseline was adopted, with MobileNetV2 utilized as the backbone network to compress the model size. Subsequently, a temporal convolutional network (TCN) module was introduced to capture dynamic dependencies between consecutive frames. Finally, a knowledge distillation (KD) strategy was incorporated to maintain recognition performance. Experimental results on the dataset show that the proposed method achieves precision, recall, and mean average precision of 87.5%, 86.4%, and 84.3%, respectively,

收稿日期: 2026-02-10; 修回日期: 2026-05-25

通信作者: 王晓, xiao.wang@ahu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62522601)

Foundation Item: The National Natural Science Foundation of China (No.62522601)

which represent improvements of 3.2%, 2.6%, and 1.1% over the baseline YOLOv8n. At the same time, the model parameter count is reduced by 26.3%. The proposed YOLO-TKD not only significantly reduces resource consumption but also effectively improves the accuracy and temporal robustness of interaction signal detection, providing a viable technical solution for autonomous driving edge devices.

Key words: interaction signal detection, lightweight model, temporal convolutional network, knowledge distillation, YOLOv8

0 引言

随着全球信息化与数字化进程的加速,汽车在提升居民出行便捷性的同时,也带来了日益严峻的交通拥堵与安全事故^[1]。自动驾驶技术作为应对这些挑战的核心途径,已成为智能交通领域的研究焦点^[2]。其终极目标是构建一个安全、高效、可靠的交通环境,而这高度依赖于系统对周围环境的精确、实时感知^[3]。环境感知作为自动驾驶系统的基石,核心任务在于可靠地识别道路参与者(车辆、行人等)及其所发出的交互信号,从而为后续的决策规划模块提供高质量的输入^[4-5]。

在此背景下,得益于人工神经网络技术的突破,基于深度学习的目标检测算法已取代传统方法,成为环境感知的主流解决方案。这些方法在检测速度、精度和误检控制方面均有显著提升^[6]。现有深度学习检测器主要可分为两阶段检测,如R-CNN(region-based convolutional neural network)^[7]、Mask R-CNN(mask region-based convolutional neural network)^[8],以及单阶段检测,如YOLO(you only look once)系列^[9]、SSD(single shot multiBox detector)^[10]。其中,YOLO系列因其在速度与精度间取得的优异平衡、强大的泛化能力及易于部署的特性,在自动驾驶实时感知任务中得到了广泛应用^[11]。

然而,自动驾驶的感知任务不限于静态目标的检测。在真实场景中,基础YOLO模型对静态车辆和行人的检测精度已达到较高水平,但当面对需要理解行为意图的交互信号(如转向灯闪烁、行人挥手等待)时,其性能会发生显著退化^[12]。这类交互信号本质上是时空耦合的:空间特征定义了信号类型,而时序动态特征揭示了意图状态。现有深度学习检测器大多基于单帧图像进行独立推理,缺乏时间维度的上下文关联。例如,在面对闪烁的转向灯时,单帧模型往往在“亮起”与“熄灭”的连续帧之间产生严重的漏检或状态误判,继而可能引

发严重的交通事故^[13]。因此,自动驾驶系统具备时序融合的交互信号理解能力,是提升决策安全性的重要课题。

针对模型在实际部署中面临的挑战,研究者提出了多种改进方案以降低计算复杂度与参数量。Liang等^[14]提出一种基于边云协同架构的Edge YOLO系统,通过重构卷积神经网络(convolutional neural network, CNN)结构,结合剪枝优化和特征压缩,构建轻量化检测模型,提升多尺度目标预测效率。Carrasco等^[15]提出改进YOLOv5的T-YOLO模型,通过设计多尺度卷积模块(multi scale module, MSM)和引入空间-通道注意力机制(spatial channel attention mechanism, SCAM),提升高空俯视场景下微小车辆检测精度。在行人检测方面,Xu等^[16]提出WOG-YOLO模型,融合灰狼优化算法(grey wolf optimizer, GWO),引入自适应缩放因子和种群集中度控制机制,显著提升对行人、骑车者等小目标的检测精度。Li等^[17]提出基于YOLOv7改进的轻量级遮挡行人检测算法,采用MobileNetV3作为主干网络以减少参数量并提升推理速度,引入高分辨率特征金字塔结构(high-resolution feature pyramid network, HR-FPN)增强对小目标和遮挡行人可见部分的特征提取能力,并在检测头嵌入高效通道与空间注意力机制(efficient channel-spatial attention, ECSA)以抑制冗余检测框。交通信号检测同样至关重要。Jensen等^[18]将YOLO算法应用于LISA信号灯数据集,该数据集的交通信号灯在不同光线和天气条件下拍摄,检测准确率达到90%。余杨等^[19]提出SVDD-YOLOv8方法,通过引入全局注意力机制(global attention mechanism, GAM)、支持向量数据描述(support vector data description, SVDD)重识别模块和高效交并比(efficient intersection over union, EIou)损失函数,显著提升交通信号灯的检测精度与抗干扰能力。尽管上述研究在空间特征优化与模型轻量化方面取得了进展,但这些研究普遍忽视了一个关键维度:交

互信号的时序动态特性。现有方法大多基于单帧图像进行推理,难以捕捉如转向灯闪烁、行人连续手势等关键时序模式。这一局限性使模型在应对动态变化的交互信号时,缺乏时序上下文理解能力,从而导致在自动驾驶真实场景中易出现信号状态误判。

为应对上述时序建模需求与边缘计算资源受限的矛盾,本文在YOLOv8架构基础上,提出一种面向自动驾驶边缘设备的轻量-时序-蒸馏协同优化框架模型(YOLO-TKD)。本文的核心不仅在于特定模块的单点替换,还在于构建了闭环的系统优化架构,主要贡献如下。

(1) 面向边缘部署的轻量设计:采用MobileNetV2替换原始主干,利用倒残差结构在压缩参数的同时保留关键空间细节,显著降低了模型的参数量与计算复杂度,为车载边缘设备的实时推理奠定基础。

(2) 面向社会交互信号的时序融合建模:引入基于膨胀因果卷积的TCN模块^[20],通过多帧特征的拼接与并行卷积,高效捕捉交互信号的周期动态特征,实现时空特征的深度融合。

(3) 面向性能保持的蒸馏联合优化:针对轻量化带来的精度折损,引入通道知识蒸馏策略^[21],在轻量化与时序模块训练同步进行的过程中,引导学生模型学习大模型的深层特征分布,实现精度与速度的最佳平衡。

通过上述系统性改进,YOLO-TKD旨在提供一个在识别精度、时序鲁棒性、模型效率与部署可行性之间取得更优平衡的解决方案,为自动驾驶车辆在复杂真实环境中实现可靠的环境感知与交互理解提供技术支持。

1 改进的YOLO-TKD模型

在自动驾驶的复杂交互场景中,车辆需要实时、准确地识别周围交通参与者的交互信号(如转向灯、手势、刹车灯),这是实现安全决策的关键。人类能够融合时序信息从而理解信号的动态变化过程,而传统深度学习模型往往难以在计算效率与时序建模能力之间取得良好的平衡。针对自动驾驶场景中交互信号的实时性和精确性的需求,本文提出YOLO-TKD模型,该模型旨在解决对时序动态特征捕捉不足的问题。该框架以YOLOv8为基础检测器,首先将其主干网络替换为轻量化结构,

显著降低计算负担,以满足车载平台对实时推理的严格要求;随后引入TCN模块,对连续帧间的交互信号进行动态建模,以有效捕捉诸如转向灯闪烁频率、手势动作连续性等关键时序特征,从而提升模型在遮挡、光线变化等复杂场景下的鲁棒性;此外,本文构建了知识蒸馏机制,以原始完整模型作为教师模型,指导轻量化学生模型学习其输出响应与深层特征分布,从而在显著提升推理速度的同时,最大程度保持识别精度。该体系增强了自动驾驶系统在真实场景下对动态交互信号的感知与理解能力,为车辆决策提供了更可靠、拟人化的视觉感知支撑。

1.1 YOLO-TKD总体结构

改进的YOLO-TKD网络架构如图1所示,主要由两大部分深度耦合而成:模型构建阶段与知识蒸馏训练阶段。在模型构建阶段,首先采用MobileNetV2作为主干网络替换YOLOv8原有的CSP-Darknet53,借助倒残差结构实现空间特征的高效提取;其次,在主干网络与检测头之间嵌入TCN模块,利用膨胀因果卷积有效捕捉交互信号的动态时序依赖,由此构建出轻量化时序架构“YOLO-TKD”。整体框架与后续各微观子模块图紧密嵌套:学生网络中的轻量化主干网络由图2所示的MobileNetV2倒残差结构堆叠而成;TCN模块则具体嵌套了图3、图4、图5详细阐述的因果卷积、膨胀卷积、残差块结构。这种宏观架构与微观模块的层次化关联,共同支撑了模型的轻量化与时序特征提取能力。在知识蒸馏训练阶段(如图1上下分支所示),以预训练的、特征表征能力更强的完整YOLOv8n模型作为“教师模型”,以本文构建的YOLO-TKD作为“学生模型”。训练过程中,学生模型不仅学习真实标注(hard label),更通过通道蒸馏损失函数对齐教师模型深层特征图所产生的软标签预测(soft prediction)。这一联合优化机制使YOLO-TKD在维持极高推理效率的同时,显著弥补了模型压缩导致的空间细节丢失问题。

1.2 轻量化主干网络

在确定轻量化改造方向后,需选取一种在有限计算资源下仍能维持较高特征表征能力的主干网络。社会交互信号(如车辆转向灯、行人手势等)通常呈现目标尺寸小、空间细节重要且易受到复杂交通环境干扰的特性。因此,主干网络必须在实现轻量化的同时,具备有效提取与保留细节特征的

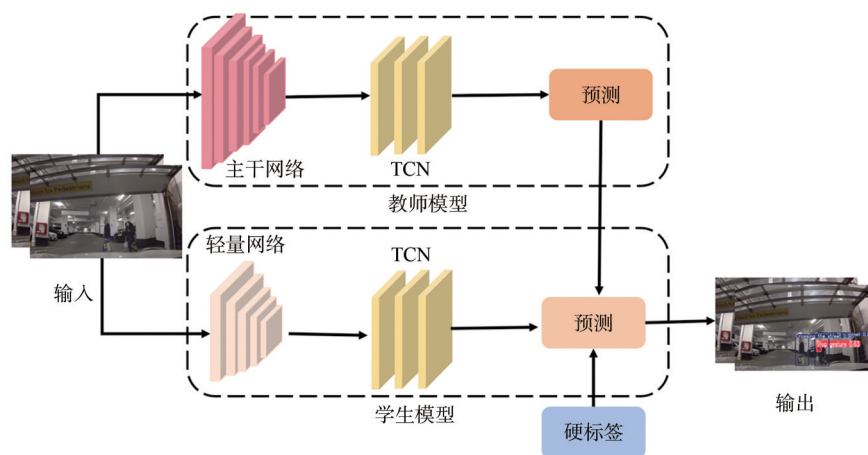


图1 改进的YOLO-TKD网络架构

能力。

选择YOLOv8作为基准检测框架主要基于以下考量：该架构在检测精度与推理效率间实现了最佳平衡，同时其模块化的组件设计便于进行针对性的轻量化改造。特别是可分离的主干网络和检测头架构允许独立优化，为后续的模型压缩和加速提供了基础。

在此基础上，对当前主流轻量化架构进行了系统比较，包括ShuffleNetV2^[22]、MobileNetV3^[23]、MobileNetV2^[24]及MobileNeXt^[25]。实验结果显示，ShuffleNetV2虽计算效率较高，但特征提取能力有限；MobileNetV3在精度方面表现良好，但其结构复杂度较高，不利于后续优化与部署；MobileNeXt作为新型架构，其性能稳定性尚需进一步验证。综合比较多种性能指标后，本文选用MobileNetV2作为替代主干网络，其核心设计基于倒残差结构（in-

verted residual）与线性瓶颈层（linear bottleneck），该架构理念与任务需求高度契合，MobileNetV2倒残差结构如图2所示^[24]。

首先，MobileNetV2采用深度可分离卷积构建基础单元，通过将标准卷积分解为深度卷积与逐点卷积两步操作，显著降低了计算复杂度与参数量，从而为车载边缘设备上的实时推理提供了基础。其次，其倒残差结构遵循“先扩展后压缩”的设计原则，先通过 1×1 卷积提升通道数，再利用 3×3 深度卷积提取空间特征。最后，经线性瓶颈层将通道维度压缩回低维。该机制针对社会交互信号检测中小目标细节容易丢失的问题，通过线性瓶颈层避免了低维空间中的信息退化，增强了对细小交互特征的保留能力。传统残差块在低维空间应用ReLU激活函数往往导致特征信息损失，而倒残差结构在高维空间进行非线性变换后，再通过线性映射降维，显著缓

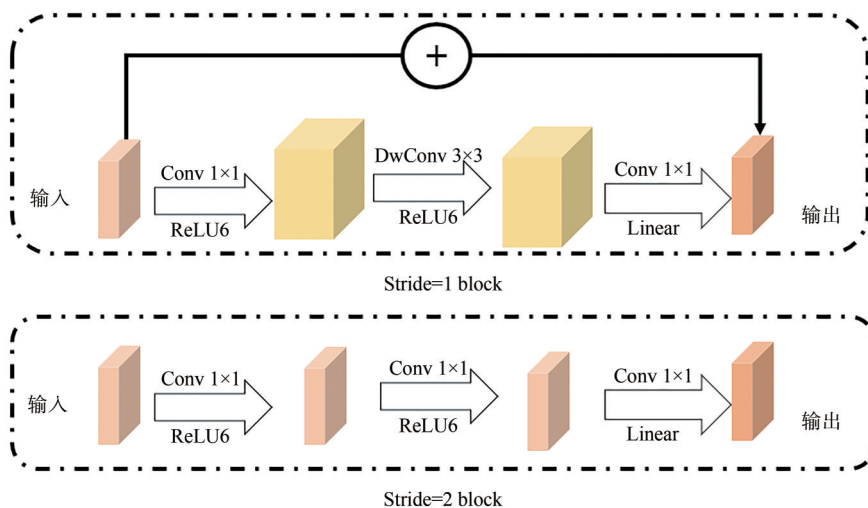


图2 MobileNetV2倒残差结构

解了低维特征的信息退化,有利于网络更稳健地编码如转向灯形态、手势指向等细微而关键的空间特征。此外,其中所用ReLU6激活函数,因其在正区间的有界饱和特性,在低精度移动端推理中具备更佳的数值稳定性,这对计算资源受限且要求输出持续可靠的自动驾驶边缘场景尤其重要。表达式为:

$$y = \text{ReLU6}(x) = \min(\max(x, 0), 6) \quad (1)$$

在确定轻量化改造方向时,本文对比了ShuffleNetV2、MobileNetV3、MobileNetV2及MobileNeXt。实验结果表明,尽管ShuffleNetV2参数量极低,但其通道洗牌机制在处理远处微小的转向灯信号时容易丢失细粒度特征;MobileNetV3虽然引入了注意力机制,但分支结构增加了边缘设备的访存开销。综合比较下,MobileNetV2的倒残差与线性瓶颈结构在“先扩展后压缩”的过程中,能以较低的计算代价显著缓解低维特征的信息退化,对社会交互信号中小目标细节容易丢失的问题具有较好的鲁棒性。

综上所述,MobileNetV2不仅是一种轻量化网络,其倒残差与线性瓶颈结构在实现显著模型压缩的同时,仍能保持优异的细节特征,契合社会交互信号检测对高精度与鲁棒性的要求。因此,本文将作为YOLO-TKD的轻量化主干网络,以在计算效率与检测精度之间达到更优平衡。

1.3 时序卷积网络模块

为实现对交互信号动态时序特性的有效建模,本文引入时序卷积网络模块,将该模块嵌入MobileNetV2主干网络与YOLOv8检测头之间。该模块承担连续帧之间时序特征融合的关键任务。与传统的循环神经网络(recurrent neural network, RNN)及其变体长短期记忆(long short-term memory, LSTM)网络相比,后者在序列建模中面临计算效率瓶颈和梯度传播问题,而时序卷积网络在架构上具有显著优势:RNN与LSTM在序列建模中存在明显的串行计算瓶颈,其当前时间步的计算需依赖前一步的结果,无法充分利用车载GPU/NPU的并行算力,导致推理时延较高;而TCN的全卷积结构支持高度并行计算,彻底避免了序列依赖带来的效率瓶颈,更契合自动驾驶对实时推理的严苛要求。此外,膨胀因果卷积机制的引入,使网络能够在不增加参数量的情况下有效建模长序列中的远距离依赖关系;残差连接的设计则有助于缓解深层网络训练

中的梯度问题。这些特性增强了模型对转向灯闪烁、手势动作等周期性交互信号的实时建模能力。

TCN以MobileNetV2提取的单帧特征作为输入,其维度为 $[C \times H \times W]$,分别表示通道数、高度和宽度。将连续 L 帧的特征图在通道维度拼接,得到维度为 $[L, C \times H \times W]$ 的序列格式数据。随后通过 1×1 卷积将通道数压缩,得到TCN的标准输入,以降低计算量。

由于交互信号检测要求当前时刻的预测仅依赖历史信息,本文采用因果卷积(causal convolution, CC)作为基本运算单元。因果卷积确保任意时间步的输出仅依赖于当前及历史输入,不引入未来信息,满足实时系统的因果性约束,其结构如图3所示。在时间步 t 的输出仅依赖于时间步 t 及 t 之前的输入,符合实时系统的因果性要求。

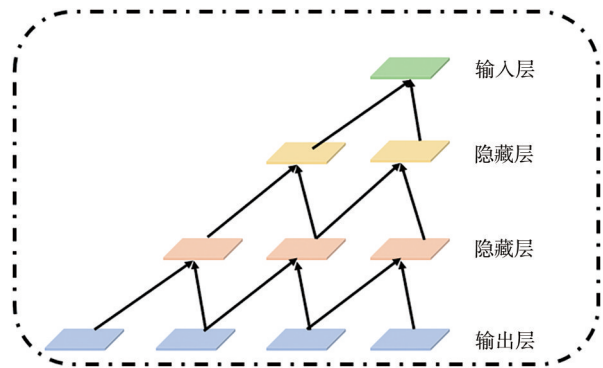


图3 因果卷积结构

为进一步扩大时序感受野以捕捉长程依赖,模块采用了膨胀卷积(dilated convolution)结构,如图4所示。膨胀卷积通过在输入元素间插入间隔来扩大卷积核的感受野,能以较少层数覆盖更长时间范围的上下文信息。计算公式为:

$$F(t) = (XF)(t) = \sum_{j=0}^{k-1} f_j^T X_{t-d(k-1-j)} \quad (2)$$

其中, d 表示膨胀系数。该模块采用4个级联的残差块结构,每个残差块包含两层膨胀卷积层,其膨胀系数按指数规律逐块递增。通过上述设计,时序卷积模块能够在有限的网络深度下构建覆盖足够长时间窗口的感受野。具体而言,该模块仅需4个残差块即可实现对16帧时序上下文的有效建模。这样的设计使模型能够充分捕捉如转向灯闪烁、刹车灯亮灭等交互信号中的动态时序变化模式,从而有效建模常见交互信号的动态特征。

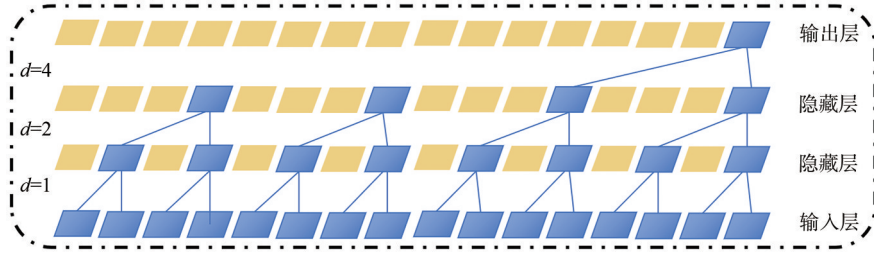


图4 膨胀卷积结构

模块采用多层残差块结构^[24]，如图5所示。残差连接通常由两个部分组成，即恒等映射和残差映射，两者共同实现信息传递与网络学习，并通过跳跃连接将输入直接传递至输出，缓解了梯度消失问题。表达式为：

$$O(x) = \sigma(x + o(x)) \quad (3)$$

$$o(x) = W_2 \sigma(W_1 x + b_1) + b_2 \quad (4)$$

其中， W_1 、 W_2 为权重矩阵， b_1 、 b_2 为偏置项， $\sigma(\cdot)$ 为激活函数。

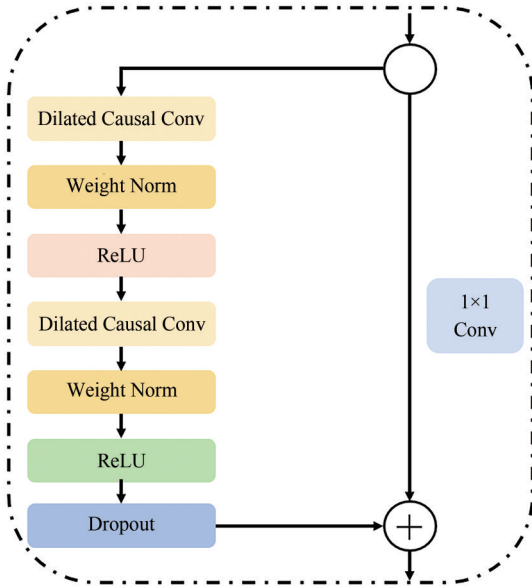


图5 残差块结构

总体而言，TCN继承了传统卷积网络的并行计算优势，能够高效建模局部时序相关性，同时通过膨胀卷积与残差连接的结合，增强了对长程时序动态的捕捉能力。相比于传统时序模型，该模块在计算效率与建模能力之间取得了更优的平衡，能够满足自动驾驶场景下对动态交互信号的实时识别需求。

在参数设计方面，时序窗口长度 L 是平衡计算成本与信息完整性的超参数。基于对PIE (pedes-

trian intention estimation) 与JAAD (joint attention in autonomous driving) 数据集中交互信号的统计分析，在真实交通场景中，常见车辆转向灯闪烁频率为1~2 Hz，完整闪烁周期为0.5~1.0 s，而行人挥手或迈步的典型动作周期通常为1.0~1.5 s。在本系统设定的10 fps视频帧率下，本文将时间窗口设为 $L=8$ 帧。该参数选择确保了时序窗口能够覆盖0.8 s的历史信息，这一时间长度包含至少半个完整的信号动态变化周期，从而为模型提供了具有足够判别力的动态时序特征。同时， $L=8$ 帧的设置避免了因输入序列过长而引入的冗余空间对齐噪声，实现了精度与计算开销的较好平衡。

1.4 基于知识蒸馏的优化

MobileNetV2主干在显著减少通道数和网络深度时，不可避免地会损失细粒度的空间特征与复杂的决策边界，这对小尺度交互信号（如远处的转向灯）的检测影响较大。为实现交互信号检测模型的轻量化与性能保持之间的平衡，需采用有效的知识迁移策略。教师模型与学生模型的特征图在空间上存在强烈的对应关系^[26]，而传统基于输出logit的蒸馏方法难以传递丰富的空间上下文信息；尤其是在自动驾驶与智能交通系统的实际应用中，交通场景的高度复杂性、动态演变特性及严格的安全要求，使模型在压缩过程中面临信息损失与性能下降的双重挑战。为此，本文采用通道知识蒸馏方法(channel-wise knowledge distillation, CWD)^[27]，旨在对学生模型进行监督，使其特征图在通道维度的统计分布与教师模型严格对齐，从而隐式继承大型模型对复杂交通环境特征的提取能力。

定义教师网络和学生网络分别为 T 和 S ，其对应通道上的激活分别记为 y^T 和 y^S 。通过softmax函数和温度参数 t 的配合，将特征激活转换为平滑的概率分布。蒸馏损失函数可表示为：

$$\varphi(\phi(y^T), \phi(y^S)) = \varphi(\phi(y_c^T), \phi(y_c^S)) \quad (5)$$

其中, $\phi(\cdot)$ 用于将激活函数转换为概率分布, 表示为:

$$\phi(y_c) = \frac{\exp\left(\frac{y_{c,i}}{t}\right)}{\sum_{i=1}^{W \cdot H} \exp\left(\frac{y_{c,i}}{t}\right)} \quad (6)$$

其中, $c=1, 2, \dots, C$ 表示通道; i 为空间位置索引; t 为温度超参数 (Temperature), 用于控制教师模型输出概率分布的平滑程度, 较大的 t 值会产生更平滑的分布, 使学生模型不仅学习到最可能的绝对类别, 还能关注到教师模型在判别复杂信号时的不确定性, 这显著提升了轻量化模型的泛化能力; $\phi(\cdot)$ 用于评估教师网络与学生网络分布差异, 通常采用 KL 散度度量对应通道激活分布之间的差异, 并作为蒸馏损失的组成部分。表示为:

$$\varphi(y^T, y^S) = \frac{t^2}{C} \sum_{c=1}^C \sum_{i=1}^{W \cdot H} \phi(y_{c,i}^T) \cdot \log \left[\frac{\phi(y_{c,i}^T)}{\phi(y_{c,i}^S)} \right] \quad (7)$$

该方法在交互信号检测任务中使 YOLO-TKD 模型在参数量显著减少的情况下, 仍能保持较高的检测精度, 为自动驾驶边缘设备上的实时交互信号检测提供了有效的轻量化与性能的平衡方案。

2 实验与结果分析

2.1 实验数据集

实验数据集采用来自车辆前向视角的 PIE 与 JAAD 自动驾驶数据集。PIE 数据集^[28]是一个专注于行人行为预测的公开数据集, 旨在提升自动驾驶系统对行人交互意图与危险行为的识别能力, 共包含超过 30 万帧图像及 1 842 个独立行人样本。JAAD 数据集^[29]广泛用于交通场景中的行人行为分析, 包含 346 个高分辨率视频片段 (时长为 5~15 s)。本文对原始视频数据进行了后处理, 提取包含明确交互意图的视频段, 统一按 10 fps 的帧率抽帧, 提取 2 453 个视频帧用于构建交互信号检测数据集。利用 Labellmg 标注工具, 对视频帧中的车辆转向信号、刹车信号以及行人手势信号进行了标注, 以确保监督学习标签的有效性。将数据集类别划分为车辆信号 (转向信号、刹车信号) 和行人信号 (原地等待、过街、手势等待), 按 8:1:1 的比例划分为训练集、验证集和测试集。

2.2 实验环境

实验环境为: GPU 采用 NVIDIA GeForce RTX

4070 SUPER 12 GB, 操作系统为 64 位 Windows 11, 采用 Anaconda 管理的 Python 3.10.16 和 PyTorch 1.12.1 深度学习框架, CUDA 版本为 12.3, cuDNN 版本为 8.3.2。

模型训练设置如下: epoch 为 150, 批次大小为 16, 使用 Adam 优化器, 初始学习率为 0.01, 权重衰减系数为 0.000 5。在知识蒸馏部分, 设置温度参数 $t=3$, 用于控制软标签的平滑程度。输入图像分辨率统一调整为 640 像素×640 像素。

2.3 实验评估指标

本实验采用精度 (precision, P), 召回率 (recall, R), 平均精度 (average precision, AP)、平均精度均值 (mean average precision, mAP)、参数量和计算量作为评价指标。

精度是衡量预测为正样本中真正样本的比例, 计算式为:

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

其中, TP 表示真阳性样本数, 即正确预测的交互信号数量; FP 表示假阳性样本数, 即错误预测为交互信号的数量; P 表示正确预测交互信号数量 (TP) 占全部预测交互信号数量 (TP+FP) 的比例, 该指标反映了模型识别结果的可靠性, 精度越高, 误检率越低。

召回率是评估实际正样本中被正确预测的比例, 计算式为:

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

其中, FN 表示假阴性样本数, 即未被检测到的真实交互信号数量。

平均精度是基于精度-召回率曲线下面积计算的类别判别性能指标, 反映模型在单一类别上的检测质量, 计算式为:

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

平均精度均值 (mAP) 为所有类别 AP 的平均值, 用于整体评估模型在多类别检测中的综合精度, 计算式为:

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (11)$$

参数量指模型的参数总量, 包括偏置值和权重参数等。

计算量用浮点运算次数 (FLOPs) 表示, 可衡量模型的计算复杂度与计算效率。

推理速度（以fps为单位）是衡量模型能否满足实时检测要求的重要指标。

2.4 与主流模型对比实验

为充分验证本文改进算法的性能优势，本文在PIE与JAAD的混合数据集上开展了对比实验。对比基准不仅包括无时序模块的YOLO系列（YOLOv5^[30]、YOLOv6^[31]、YOLOv7-tiny^[32]、YOLOv8^[33]、YOLOv10^[34]）及轻量级Transformer架构（RT-DETR^[35]），还引入了具备时空建模能力的视频目标检测模型YOWO^[36]。本文方法和其他方法在数据集上的对比见表1。

实验结果表明，相比无时序模块的YOLOv8，引入时序融合的模型在召回率与平均精度上均有明显提升，表明时序信息对减少动态信号漏检具有重要作用。然而，YOWO架构由于存在复杂的3D卷积或串行计算结构，推理速度无法满足自动驾驶的实时决策需求。相比之下，本文YOLO-TKD模型在保持高精度的前提下，参数量仅为 2.24×10^6 ，计算复杂度在所有对比模型中最低。在深度学习模型部署层面，本文模型的理论计算量与参数量均显著低于所有对比模型。这一优势的根本来源在于其主干网络引入了MobileNetV2特有的深度可分离卷积架构，该设计可大幅降低浮点运算次数与内存访问成本。因此，上述数值优势与底层轻量化机制共同构成了本模型具备边缘端部署可行性的核心理论依据。上述在计算复杂度和内存占用上的显著优势，从理论层面充分验证了本模型在算力受限的车载边缘设备上的高效推理潜力，表明了本框架在检测精度与边缘计算效率之间实现了良好平衡，兼具高准确性与硬件部署适应性。

2.5 轻量化主干网络选择实验

为确定适用于交互信号检测任务的主干网络替代方案，本文对多种轻量化模型作为YOLOv8n主

干网络时的性能进行了评估。实验选取了ShuffleNetv2、MobileNetV3、MobileNetV2及MobileNeXt这4种轻量化架构作为候选主干网络，替换原始YOLOv8n的主干部分，并在JAAD数据集下进行比较分析。各轻量化主干网络在测试集上的性能见表2。由表2可以看出，以MobileNetV2为主干的YOLOv8n模型在综合轻量化效果上表现最佳，尽管YOLOv8n-MobileNeXt模型在检测准确率上略具优势，但其计算量与参数量显著高于其他方案，不适合在资源受限场景中部署。综合考虑检测精度、计算效率与模型体积之间的平衡，本文最终选用MobileNetV2作为YOLOv8n的主干网络替代结构，以在保持精度的同时实现较大程度的模型轻量化。

2.6 消融实验

为验证本文各项改进方案在交通场景下的有效性，以原始YOLOv8n为基准模型，在PIE与JAAD的混合数据集上进行了消融实验，结果见表3。其中，方法1为原始YOLOv8n基线模型；方法2仅替换MobileNetV2为主干网络，旨在评估轻量化改造对检测精度的影响；方法3单独引入TCN模块，用于验证时序信息建模的必要性；方法4在轻量化主干基础上融合TCN，考察两者协同效果；方法5结合TCN与KD，用以分析蒸馏策略在无轻量化主干时的性能增益；方法6将轻量化主干与KD相结合，用以验证蒸馏对轻量化模型精度恢复的作用；方法7为本文完整提出的YOLO-TKD模型。基线模型的平均精度为83.2%，参数量为 3.04×10^6 。当对模型进行轻量化处理时，虽然参数量和计算量分别降至 1.71×10^6 和3.1 GFLOPs，但空间特征提取能力的减弱导致mAP出现预期内的下降。为弥补这一损失，在轻量化基础上引入了TCN时序模块。结果显示，得益于时间窗口对连续帧动态特征的有效捕捉，模型对时序信号的漏检率显著下降，mAP

表1 本文方法和其他方法在数据集上的对比

方法	P	R	mAP0.5	参数量	计算量/GFLOPs	推理速度/fps
YOLOv5	79.4%	80.1%	80.5%	8.64×10^6	12.9	45
YOLOv6	81.2%	81.6%	81.4%	4.62×10^6	11.3	52
YOLOv7-tiny	82.7%	83.2%	82.1%	6.07×10^6	13	48
YOLOv8	84.3%	83.8%	83.2%	3.04×10^6	8.1	<u>60</u>
RT-DETR	88.5%	87.6%	84.1%	18.74×10^6	56.4	22
YOLOv10	84.9%	84.6%	83.3%	<u>2.49×10^6</u>	<u>6.2</u>	63
YOWO	86.5%	84.9%	84.5%	<u>15.2×10^6</u>	<u>45.3</u>	27
本文方法	<u>87.5%</u>	<u>86.4%</u>	<u>84.3%</u>	2.24×10^6	3.6	58

表2 各轻量化主干网络在测试集上的性能

方法	P	mAP0.5	参数量	计算量/ GFLOPs
YOLOv8n	84.3%	83.2%	3.04×10^6	8.1
YOLOv8n-ShuffleNetv2	78.2%	79.3%	1.83×10^6	3.7
YOLOv8n-MobileNetV3	79.4%	77.4%	2.77×10^6	3.5
YOLOv8n-MobileNetV2	83.5%	81.2%	1.71×10^6	3.1
YOLOv8n-MobileNeXt	84.1%	82.2%	8.9×10^6	4.85

表3 消融实验

方法	MobileNetV2	TCN	KD	P	mAP0.5	参数量	计算量/ GFLOPs
1				84.3%	83.2%	3.04×10^6	8.1
2	√			83.5%	81.2%	1.71×10^6	3.1
3		√		85.1%	85.2%	3.74×10^6	9.2
4	√	√		86.1%	84.2%	2.24×10^6	3.6
5		√	√	88.4%	85.2%	3.74×10^6	9.2
6	√		√	84.2%	82.4%	1.71×10^6	3.1
7	√	√	√	87.5%	84.3%	2.24×10^6	3.6

提升了3.0%，表明时序信息对空间特征具有补偿作用。

为进一步提升轻量化模型的性能，最终的YOLO-TKD模型在上述基础上引入了CWD知识蒸馏策略。在不增加额外推理负担的前提下，精确率提升了1.4%，mAP最终达到84.3%。总体而言，最终的YOLO-TKD模型较原始基线不仅精确率提升了3.2%，而且在参数量压缩26.3%、计算量降低55.6%的情况下，全面超越了基线模型的感知性能，实现了资源压缩与高精度动态感知的有效平衡。

2.7 可视化分析

为验证本文改进算法在自动驾驶场景中的适用性，以YOLOv8n为基准模型，与YOLO-TKD模型在PIE与JAAD混合数据集上进行对比实验，可视化检测结果对比如图6所示。

实验结果表明，本文YOLO-TKD模型在环境感知任务中表现出良好的综合性能。相较于基准模型，YOLO-TKD显著提高了对多类交通目标（如车辆转向灯、行人信号）的检测精度，能够准确识别复杂交通场景中的关键目标与交互信号。由图6第二行可以看出，当行人做出“停止”等动态手势时，单帧基准模型由于缺乏时间上下文极易漏检，而YOLO-TKD凭借TCN模块对连续动作周期的捕捉，能够稳定输出预测结果。在保持较低计算开销的前提下，YOLO-TKD有效提升了多类别目标的



图6 可视化检测结果对比

检测准确率与鲁棒性，验证了其在自动驾驶环境感知系统中的实用价值。

3 结论与展望

针对自动驾驶场景中交互信号识别面临的时序建模不足、计算资源受限等挑战，本文提出了一种轻量级时序知识蒸馏模型YOLO-TKD。YOLO-TKD模型在PIE与JAAD数据集上的精确率、召回率和平均精度均值指标分别达到87.5%、86.4%和84.3%，比基准模型分别提升了3.2%、2.6%和1.1%。模型参数量减少至 2.24×10^6 ，仅为基准模型的73.7%。这表明YOLO-TKD模型具有更高的检测精度、更少的参数量和更强的鲁棒性。

本文模型具有广泛的应用前景，尤其适用于智能交通^[37]和自动驾驶系统中的关键场景。在车路协同感知系统中，该模型可实现对交通参与者与信号的实时精准感知；在车载嵌入式平台中，可对车辆与行人行为意图进行连续预测^[38]，为自动驾驶决策系统提供可靠依据。

未来的研究将重点围绕以下方向展开：一是在系统集成层面，探索轻量化时序感知模块与自动驾驶大模型的可插拔集成机制^[39]，增强大模型在动态交互场景中的推理效率与行为预测可解释性，推动感知-决策的闭环优化；二是在多模态融合层面，研究视觉、声学等多信息在极端天气^[40-42]下的

协同感知机制, 构建基于跨模态注意力对齐与时序特征增强的稳健感知框架, 提升系统在复杂环境^[43]下的交互信号识别稳定性与适应性。

参考文献:

- [1] Wang X, Jiang R, Li L, et al. Capturing car-following behaviors by deep learning[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2018, 19(3): 910-920.
- [2] Wang X M, Zheng X H, Chen W, et al. Visual human-computer interactions for intelligent vehicles and intelligent transportation systems: the state of the art and future directions[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, 51(1): 253-265.
- [3] Wang Z J, Wu Y, Niu Q Q. Multi-sensor fusion in automated driving: a survey[J]. *IEEE Access*, 2020, 8: 2847-2868.
- [4] 刘泽禹, 张慧, 李涓东, 等. 面向自动驾驶的感知决策一体化综述[J]. *智能科学与技术学报*, 2025, 7(1): 4-20.
Liu Z Y, Zhang H, Li Y D, et al. A review on perception-decision integration for autonomous driving[J]. *Chinese Journal of Intelligent Science and Technology*, 2025, 7(1): 4-20.
- [5] Wang S, Yuan Y, Wang X, et al. An overview of smart contract: architecture, applications, and future trends[C]//*Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV)*. Piscataway: IEEE Press, 2018: 108-113.
- [6] Diwan T, Anirudh G, Tembhurne J V. Object detection using YOLO: challenges, architectural successors, datasets and applications[J]. *Multimedia Tools and Applications*, 2023, 82(6): 9243-9275.
- [7] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//*Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2014: 580-587.
- [8] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV)*. Piscataway: IEEE Press, 2017: 2980-2988.
- [9] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE Press, 2016: 779-788.
- [10] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]//*Computer Vision - ECCV 2016*. Berlin: Springer, 2016: 21-37.
- [11] Zhao Y F, Gao H, Wang S, et al. A novel approach for traffic signal control: a recommendation perspective[J]. *IEEE Intelligent Transportation Systems Magazine*, 2017, 9(3): 127-135.
- [12] 杜泉成, 王尧, 李灵犀, 等. 行人轨迹预测方法关键问题研究: 现状及展望[J]. *智能科学与技术学报*, 2023, 5(2): 143-162.
Du Q C, Wang X, Li L X, et al. Key problems and progress of pedestrian trajectory prediction methods: the state of the art and prospects[J]. *Chinese Journal of Intelligent Science and Technology*, 2023, 5(2): 143-162.
- [13] Du Q C, Wang X, Yin S G, et al. Social force embedded mixed graph convolutional network for multi-class trajectory prediction[J]. *IEEE Transactions on Intelligent Vehicles*, 2024, 9(9): 5571-5580.
- [14] Liang S Y, Wu H, Zhen L, et al. Edge YOLO: real-time intelligent object detection system based on edge-cloud cooperation in autonomous vehicles[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(12): 25345-25360.
- [15] Carrasco D P, Rashwan H A, García M Á, et al. T-YOLO: tiny vehicle detection based on YOLO and multi-scale convolutional neural networks[J]. *IEEE Access*, 2023, 11: 22430-22440.
- [16] Xu L Z, Yan W, Ji J S. The research of a novel WOG-YOLO algorithm for autonomous driving object detection[J]. *Scientific Reports*, 2023, 13: 3699.
- [17] Li C, Wang Y D, Liu X M. An improved YOLOv7 lightweight detection algorithm for obscured pedestrians[J]. *Sensors*, 2023, 23(13): 5912.
- [18] Jensen M B, Nasrollahi K, Moeslund T B. Evaluating state-of-the-art object detector on challenging traffic light data[C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. Piscataway: IEEE Press, 2017: 882-888.
- [19] 余杨, 何刘. 基于YOLOv8和SVDD重识别的交通信号灯检测方法[J]. *汽车实用技术*, 2024, 49(23): 49-54.
Yu Y, He L. Traffic light detection method based on YOLOv8 and re-identification using SVDD[J]. *Automobile Applied Technology*, 2024, 49(23): 49-54.
- [20] Bai S J, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[PP]. V2. (2018-04-19)[2026-02-10]. arXiv:arXiv.1803.01271.
- [21] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[PP]. V1. (2015-03-09)[2026-02-10]. arXiv:arXiv.1503.02531.
- [22] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2018: 6848-6856.
- [23] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]//*Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Piscataway: IEEE Press, 2019: 1314-1324.
- [24] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2018: 4510-4520.
- [25] Zhou D Q, Hou Q B, Chen Y P, et al. Rethinking bottleneck structure for efficient mobile network design[C]//*Computer Vision - ECCV 2020*. Berlin: Springer, 2020: 680-697.
- [26] Tian S S, Li L S, Li W J, et al. A survey on few-shot class-incremental learning[J]. *Neural Networks*, 2024, 169: 307-324.
- [27] Shu C Y, Liu Y F, Gao J F, et al. Channel-wise knowledge distillation for dense prediction[C]//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. Piscataway: IEEE Press, 2021: 5291-5300.
- [28] Rasouli A, Kotseruba I, Kunic T, et al. PIE: a large-scale dataset and models for pedestrian intention estimation and trajectory prediction[C]//*Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Piscataway: IEEE Press, 2019: 6261-6270.
- [29] Rasouli A, Kotseruba I, Tsotsos J K. Are they going to cross? a benchmark dataset and baseline for pedestrian crosswalk behavior[C]//*Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. Piscataway: IEEE Press, 2017: 206-213.
- [30] Zhu X K, Lyu S C, Wang X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[C]//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*. Piscataway: IEEE Press, 2021: 2778-2788.
- [31] Li C Y, Li L L, Jiang H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[PP]. V1. (2022-09-07)[2026-02-10]. arXiv: arXiv.2209.02976.

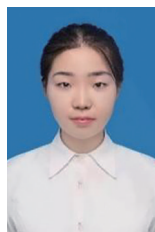
- [32] Wang C Y, Bochkovskiy A, Liao H M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 7464-7475.
- [33] Yaseen M. What is YOLOv8: an in-depth exploration of the internal features of the next-generation object detector[PP]. V1. (2024-08-28) [2026-02-10]. arXiv: arXiv.2408.15857.
- [34] Wang A, Chen H, Liu L H, et al. YOLOv10: real-time end-to-end object detection[PP]. V2. arXiv (2024-10-30) [2026-02-10]. arXiv: arXiv.2405.14458.
- [35] Zhao Y A, Lv W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2024: 16965-16974.
- [36] KÖPÜKLÜ O, Wei X, Rigoll G. You only watch once: a unified CNN architecture for real-time spatiotemporal action localization[PP]. V5. (2021-10-18)[2026-02-10]. arXiv: arXiv.1911.06644.
- [37] Zhu F H, Lv Y S, Chen Y Y, et al. Parallel transportation systems: toward IoT-enabled smart urban traffic control and management[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(10): 4063-4071.
- [38] Du Q C, Xu L L, Wu Q, et al. SafeCrossNet: multi-modal fusion with social-aware for pedestrian crossing intention prediction[J]. Information Fusion, 2026, 126: 103609.
- [39] Cui Y D, Huang S C, Zhong J M, et al. DriveLLM: charting the path toward full autonomous driving with large language models[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(1): 1450-1464.
- [40] Jang D, Lee S, Choi G, et al. Energy-based domain adaptation without intermediate domain dataset for foggy scene segmentation[J]. IEEE Transactions on Image Processing, 2024, 33: 6143-6157.
- [41] Shen Y, Wei M, Wang Y, et al. Rethinking real-world image deraining via an unpaired degradation-conditioned diffusion model[PP]. V4. (2024-05-01)[2026-02-10]. arXiv: arXiv.2301.09430.
- [42] Qin H D, Lu X, Wang L C, et al. Foggy-aware teacher: an unsupervised domain adaptive learning framework for object detection in foggy scenes[J]. IEEE Robotics and Automation Letters, 2025, 10(7): 7508-7515.
- [43] 刘文华, 李滢东, 王涛, 等. 基于高维特征表示的交通场景识别[J]. 智能科学与技术学报, 2019, 1(4): 392-399.
- Liu W H, Li Y D, Wang T, et al. Transportation scene recognition

based on high level feature representation[J]. Chinese Journal of Intelligent Science and Technology, 2019, 1(4): 392-399.

[作者简介]



胡振虢 (2000-), 男, 安徽大学人工智能学院硕士生, 主要研究方向为交互信号识别、计算机视觉、机器学习。



姚肖婷 (1998-), 女, 安徽大学人工智能学院硕士生, 主要研究方向为智能决策、轨迹预测、智能交通系统。



席治远 (1995-), 男, 博士, 安徽大学人工智能学院博士生, 主要研究方向为人工智能、强化学习、无人驾驶决策。



王晓 (1988-), 女, 安徽大学人工智能学院教授, 主要研究方向为社会计算、群体行为建模、无人自主系统及其平行测试。