

面向智能驾驶的广义零样本驾驶行为识别方法

许祖荫, 孟宪佳, 解家豪
(西北大学 计算机学院, 710127, 西安)

摘要: 针对智能驾驶行为识别依赖预定义类别、不可见行为缺少真实训练样本以及复杂生成模型难以边缘部署的问题, 面向广义零样本学习任务, 提出结合语义条件生成与潜在谱校准的驾驶行为识别方法(Diff-HAR)。首先, 采用卷积时序编码器获得统一潜在表示, 以类别语义为条件通过扩散模型生成不可见类特征; 其次, 利用主子空间保持的潜在谱校准和统计扰动调整生成特征结构与分布; 最后, 将其与真实可见类特征联合用于轻量分类器训练, 并通过偏置校准实现统一识别, 最终形成云端离线生成训练—边缘在线轻量推理的解耦架构。实验结果表明, 在 DriverActivity 数据集中, Diff-HAR 方法的不可见类准确率和调和平均值分别为 39.0% 和 44.1%, 较条件变分自编码器(CVAE)分别高 10.1%和 7.7%, 与带梯度惩罚的 Wasserstein 生成对抗网络(WGAN-GP)的对应指标 39.6%和 43.9%总体相当; 在 PAMAP2 数据集上, 两项指标分别达到 94.7%和 72.3%, 较 FM-IoT 分别高 41.0%和 10.2%; ESP32-S3 轻量分类头回放测试中, 模型大小为 0.24 MB, 平均推理延迟为 9.379 ms。研究表明, Diff-HAR 方法可在不可见类无真实训练样本的条件下兼顾可见类与不可见类识别, 并验证了云端生成训练—边缘轻量分类在资源受限智能驾驶节点上的可行性。

关键词: 智能驾驶; 广义零样本学习; 驾驶行为识别; 潜在谱校准

中图分类号: TP391 **文献标志码:** A

DOI: **文章编号:**

A Generalized Zero-Shot Driving Behavior Recognition Method for Intelligent Driving

XU Zuyin, MENG Xianjia, XIE Jiahao

(College of Computer Science, Northwest University, Xi'an 710127, China)

Abstract: To address the dependence of intelligent-driving behavior recognition on predefined classes, the lack of real training samples for unseen behaviors, and the difficulty of deploying complex generative models at the edge, a driving behavior recognition method combining semantic-conditioned generation and latent spectral calibration (Diff-HAR) is proposed for generalized zero-shot learning. First, a convolutional temporal encoder is used to obtain a unified latent representation, and unseen-class features are generated by a diffusion model conditioned on class semantics. Second, principal-subspace-preserving latent spectral calibration and statistical perturbation are used to adjust the structure and distribution of the generated features. Finally, the generated features are combined with real seen-class features to train a lightweight classifier, and bias calibration enables unified recognition, forming a decoupled architecture of cloud-based offline generation and training and edge-based online lightweight inference. Experimental results show that, on the DriverActivity dataset, Diff-HAR achieves an unseen-class accuracy of 39.0% and a harmonic mean of 44.1%, exceeding the conditional variational autoencoder (CVAE) by 10.1 and 7.7 percentage points, respectively, and is overall comparable to the Wasserstein generative adversarial network with gradient penalty (WGAN-GP), whose corresponding metrics are 39.6% and 43.9%. On the PAMAP2 dataset, the two metrics reach 94.7% and 72.3%, respectively, outperforming FM-IoT by 41.0 and 10.2 percentage points. In the ESP32-S3 replay test of the lightweight classification head, the model size is 0.24 MB and the average inference latency is 9.379 ms. The results show that Diff-HAR balances seen- and unseen-class recognition without real training samples for unseen classes and demonstrate the feasibility of cloud-based generation and training with lightweight edge classification on resource-constrained intelligent-driving nodes.

收稿日期: 2026-06-12。 作者简介: 许祖荫(2002—), 男, 硕士生; 孟宪佳(通信作者), 男, 教授, 博士生导师。
基金项目: 国家自然科学基金资助项目(62276211); 交通运输大数据与人工智能教育部重点实验室(北京交通大学)开放基金资助项目(BATLAB202404)。
网络出版时间: 网络出版地址:

Key words: intelligent driving; generalized zero-shot learning; driving behavior recognition; latent spectral calibration

智能网联汽车的感知与决策正向信息融合、车路协同、车路云协同及端到端驾驶等方向发展^[1-4]。除道路环境感知外,对驾驶员及乘员等与车辆运行密切相关的驾驶行为进行识别,是安全预警、人机协同和智能交互的重要基础。视觉、毫米波雷达与惯性测量单元等传感方式为驾驶行为感知提供了多样化的数据来源^[5-9]。然而,现有驾驶行为识别模型通常依赖预先设定的类别集合和充分标注样本,当部署阶段出现训练数据未覆盖的驾驶行为类别时,封闭类别模型难以直接完成识别。

智能驾驶场景中的行为类别覆盖不足并非仅由数据分布不均导致。随着智能座舱功能、人机交互与应用需求演化,部署阶段可能出现训练集合未覆盖的新型、低频或高风险难采行为。与通常仍有少量样本的长尾类别不同,本文关注训练阶段完全缺少目标行为真实传感样本的零样本识别,即解决预定义类别集合与实际部署需求之间的不一致,而非一般类别不均衡或域偏移。

零样本学习利用类别属性或文本语义,在无目标类别真实训练样本时迁移知识;广义零样本学习要求在统一类别空间同时识别可见类与不可见类^[10-11]。本文设定已知待识别类别名称和语义描述,但训练阶段无其真实传感样本,以类别语义建立知识桥梁;该任务用于模拟驾驶行为类别扩展,不涉及动态类别发现、未知类拒识或持续学习。生成式广义零样本识别仍面临三点:语义与时序潜在特征存在表征鸿沟;合成特征可能产生类间混淆、潜在结构偏移和可见类偏置^[12-13];语义编码、扩散采样和结构校准开销较高,难以直接运行于资源受限节点。

智能驾驶系统包含高性能车载计算平台和分布式感知节点等计算层级。前者承担复杂感知、融合与决策;与传感器直连的智能座舱分布式感知节点、车载物联网节点等持续采集数据并执行初步行为识别,受算力、存储和功耗约束,不适合运行复杂生成模型。因此,本文将语义编码、扩散生成、潜在谱校准和分类器训练置于云端离线阶段,边缘端仅执行轻量特征处理、分类与偏置校准,以云边解耦服务资源受限节点。

针对上述问题,本文提出 Diff-HAR,主要贡献如下:(1)针对不可见类合成特征的潜在结构偏移,提出主子空间保持的潜在谱校准方法,在不重新训练生成器、不改变主奇异向量方向的条件下,以非扩张方式收缩高阶谱能量;(2)构建云端离线生成训练一边缘在线轻量推理解耦流程,云端完成复杂生成与训练,资源受限节点执行时序特征提取、MLP 分类和偏置校准。对比、消融、潜在谱分析及 ESP32-S3 分类头回放实验分别验证零样本识别效果、结构校准机制与轻量分类链路可执行性。

1 智能驾驶行为识别与零样本学习研究进展

面向智能驾驶的驾驶行为感知主要包括驾驶员操作与分心行为识别、乘员交互及座舱动作理解,常采用视觉、毫米波雷达、惯性传感和多源融合等技术路线。视觉方法能够利用姿态、骨架和局部上下文刻画细粒度动作,毫米波雷达具有非接触与隐私友好特点,惯性测量单元适于连续运动感知,多源混合模态则可增强复杂座舱环境下的表征能力^[5-9]。现有智能座舱与驾驶行为研究已在预定义类别的监督识别方面取得进展,但训练类别集合通常在部署前固定,对训练阶段未覆盖驾驶行为的直接识别能力仍然有限。

零样本与广义零样本方法利用类别属性或文本语义,将可见类知识迁移至训练阶段缺少真实样本的不可见类,主要技术路线包括语义嵌入映射和生成式特征增强^[10-14]。面向传感行为的研究进一步引入词向量、预训练语言模型或跨模态对齐,以传递类别名称及动作描述中的知识^[15-20]。其中,嵌入式方法直接学习传感特征与语义空间之间的映射,生成式方法则依据类别语义合成不可见类特征,并与真实可见类特征共同训练统一分类器。相较传统零样本设定,广义零样本识别还需处理真实可见类与合成不可见类之间的决策偏置。

综上,现有研究仍存在以下三方面不足:智能驾驶行为模型多基于封闭类别监督学习,难以直接服务行为类别扩展;零样本行为方法更重视样本级语义对齐或特征生成,对批量合成潜在特征的整体奇异值谱结构关注不足;语义编码、生成采样与结构校准的计算开销也不适合直接部署于资源受限节点。扩散模型已为时序与惯性数据生成提供基础^[21-22]。在已有条件扩散生成框架基础上,本文进一步关注生成后批量潜在特征的奇异值谱结构,并通过云端离线生成训练与边缘在线轻量推理解耦处理边缘执行约束。

2 智能驾驶行为零样本识别任务定义与云边系统架构

2.1 问题边界与核心概念

本文所称“驾驶行为”是指智能驾驶相关场景中可由时序传感数据表征的驾驶员及乘员行为,不指车辆轨迹、驾驶策略或控制动作。本文将“零样本识别”作为具体学习任务,将“驾驶行为类别扩展”作为该任务服务的应用需求;“开放环境”用于描述部署背景,“长尾、低频、高风险难采和新型交互”用于解释不可见类训练样本缺失的可能来源。本文区分原始

信号频谱与潜在特征奇异值谱: 前者描述传感序列的频率成分, 后者描述批量潜在特征的主方向能量分布, 本文校准仅作用于后者。“潜在谱塌陷”表示生成特征谱过度集中或高阶能量异常衰减; “语义—潜在结构鸿沟”表示类别语义与连续运动潜在表征之间的差异。

2.2 输入数据与类别语义表征

输入样本以滑动时间窗表示, 来源包括可穿戴惯性测量单元 (IMU) 以及用于人体运动感知的毫米波雷达。对于单个传感窗口, 本文记为, 其中, 表示时间窗长度, 表示传感通道数。

惯性测量单元由多轴加速度和角速度通道构成; 毫米波点云经统计聚合形成时序矩阵。各模态分别映射至统一潜在空间, 无需同步采集或预先配准。

设完整的驾驶行为类别集合为 $Y = Y_s \cup Y_u$, 其

中 Y_s 表示可见类集合, 即训练阶段具有真实标注传感样本的行为类别; Y_u 表示不可见类集合, 即训练阶段没有真实传感样本、但类别名称及语义描述可用的目标行为类别, 且 $Y_s \cap Y_u = \emptyset$ 。不可见类样本缺失可由新型交互、低频发生、高风险难采或训练类别覆盖不足等因素造成, 不将不可见类与长尾类等同。每个类别均关联一个语义嵌入 $y \in Y$ 。本文采用预训练句子语义编码模型 SBERT^[15] 对行为类别的自然语言描述进行编码, 得到类别级语义嵌入 a_y 。类别语义描述主要刻画行为的运动特征、身体部位参与方式和潜在风险属性, 例如“周期性头部运动”“上身快速前倾”“手部接近头部区域”等。该语义表征用于在可见类与不可见类之间建立知识迁移桥梁, 是后续不可见类特征生成的条件输入。在此基础上, Diff-HAR 的总体架构如图 1 所示。

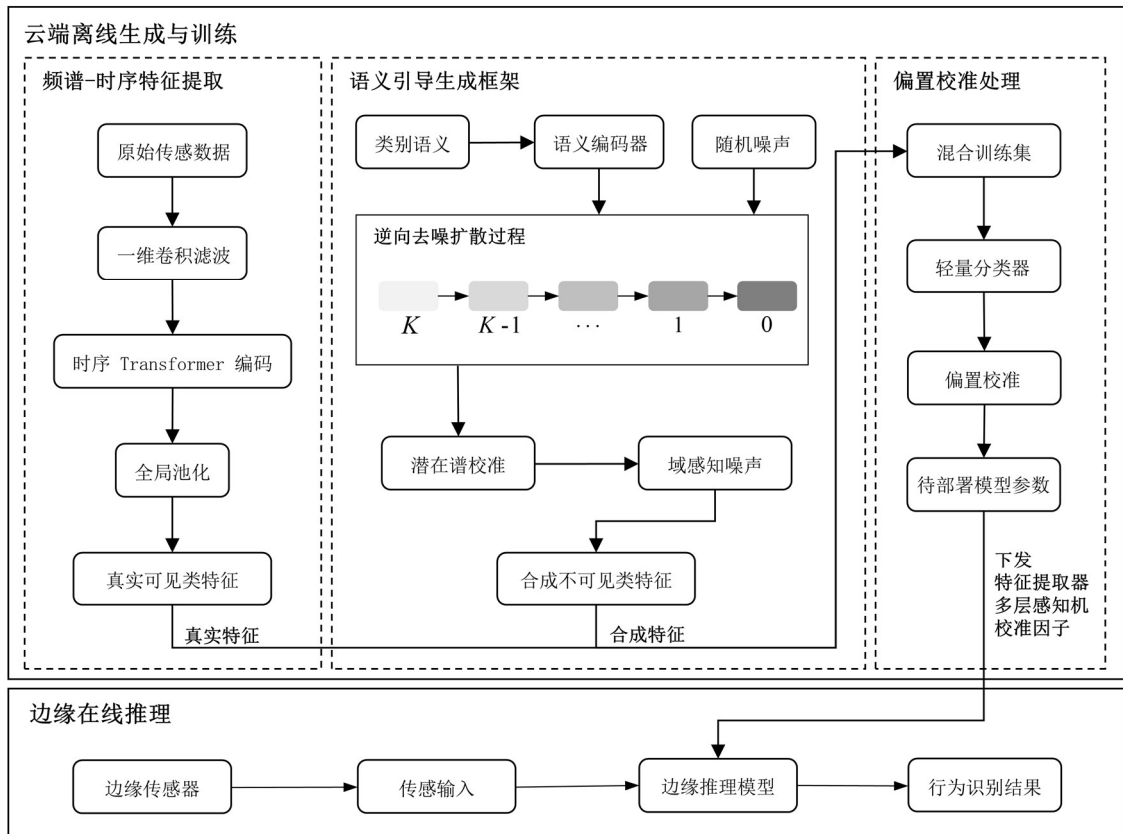


图 1 Diff-HAR 云端解耦总体架构与系统流程

Fig. 1 Overall cloud-edge decoupled architecture and system workflow of Diff-HAR

2.3 广义零样本行为识别任务

本文采用固定广义零样本协议评价智能驾驶行为零样本识别: 训练阶段仅提供可见类真实样本和全部类别语义, 测试阶段在可见类与不可见类的统一类别空间内进行预测。Diff-HAR 将传感窗口映射

至潜在空间, 依据不可见类语义生成潜在特征, 并与真实可见类特征共同训练轻量分类器。该固定协议用于模拟类别集合扩展需求, 不涉及动态类别发现、未知类拒识或持续学习。

2.4 云边解耦系统架构

本文所称“边缘端”具体指智能座舱分布式感知节点、车载物联网节点及与传感器连接的资源受限计算设备，是整车计算体系中承担持续数据采集与轻量行为识别的节点。高性能车载计算平台负责复杂感知、融合与决策；本文关注云边解耦如何将生成式广义零样本识别能力服务于上述资源受限节点。在该架构中，云端完成时序特征编码器训练、SBERT 语义编码、条件扩散训练与采样、潜在谱校准、统计扰动及分类器训练；边缘端执行时序特征提取、轻量分类与偏置校准，不执行语义编码、扩散生成、潜在谱校准或统计扰动。本文板端实验仅评价“96 维潜在特征回放—MLP 分类—偏置校准”链路。

3 Diff-HAR 广义零样本行为识别方法

根据第 2.3 节定义的固定广义零样本任务，Diff-HAR 依次完成时序传感特征表征、不可见类语义条件生成、批量潜在特征结构校准以及可见—不可见类统一分类。条件扩散作为云端离线特征生成器，将类别语义转化为不可见类潜在特征；主子空间保持的潜在谱校准用于抑制合成特征的异常高阶能量；统计扰动和偏置校准分别改善类内覆盖与两类决策平衡；云边解耦则将复杂生成过程与边缘在线识别隔离。

3.1 时序传感特征编码与统一表征

采用卷积—时序组合编码器获得 96 维统一潜在表示，该编码器作为后续语义条件生成和潜在谱校准的特征输入模块。毫米波点云按距离、速度及空间分布统计均值、标准差和极差，形成 15 维定长表示：

$$x_i^{\text{PASA}} = [\mu_i, \sigma_i, r_i, \dots, \mu_5, \sigma_5, r_5]^T \in \mathbf{R}^{15} \quad (1)$$

式中： μ_i 、 σ_i 和 r_i 分别表示第 i 个统计维度的均值、标准差和极差。各帧 PASA 特征沿时间组成窗口 $X_{\text{PASA}} = [x_1^{\text{PASA}}, \dots, x_L^{\text{PASA}}]^T \in \mathbf{R}^{15}$ 。记时序特征提取器为 $E(\cdot)$ ，则输入窗口 X 经编码得到 96 维潜在表示 $z_0 = E(X) \in \mathbf{R}^{96}$ ；毫米波输入满足 $X = X_{\text{PASA}}$ 。

3.2 不可见类语义条件生成

为把类别语义转化为可训练的不可见类潜在特征，本文采用已有条件去噪扩散框架^[21]在潜在空间进行云端离线采样。前向扩散通过逐步向 z_0 注入高斯噪声，使其随扩散时间步增加而接近标准高斯分布。训练时，噪声预测网络以带噪特征、时间步和类别语义为输入，优化目标为

$$L_{\text{diff}} = \mathbb{E}_{z_0, a_y, k, \varepsilon} [\|\varepsilon - \varepsilon_\phi(z_k, k, a_y)\|_2^2] \quad (2)$$

式中： k 为随机采样的扩散时间步， K 为扩散总

步数； ε 为标准高斯噪声，带噪潜在特征对应第 k 步状态，类别语义嵌入作为条件输入。该目标使噪声预测网络学习在不同噪声水平下恢复类别条件潜在分布。

训练时随机置空类别语义，使同一网络学习条件与无条件去噪，并采用无分类器引导 (CFG)^[23] 进行采样；采样时按式 (3) 组合两类预测：

$$\hat{\varepsilon} = (1 + \omega)\varepsilon_\phi(z_k, k, a_y) - \omega\varepsilon_\phi(z_k, k, \emptyset) \quad (3)$$

式中： $\omega \geq 0$ 为 CFG 引导强度，两项分别表示条件噪声预测和无条件噪声预测。增大 ω 可强化目标语义响应，但过大会降低样本多样性，因此 ω 仅在验证集上选择。

3.3 主子空间保持的潜在谱校准机制

类别语义能够约束生成方向，却不直接约束一批合成特征在潜在空间各方向上的能量分配。若能量过度集中于少数方向，合成样本易出现结构塌陷和类内覆盖不足；若高阶方向能量异常，又可能把生成噪声带入分类边界。因此，本文在批量潜在特征层面校准奇异值谱，而非处理单条传感信号的频率成分，也不把谱一致性等同于原始信号物理一致性。

从约束对象看，现有生成特征优化通常在生成器训练或逐样本表示层面引入判别、重构和语义约束^[12-13]；本文则在生成完成后直接处理批量特征矩阵的奇异值谱，不增加样本级距离目标，也不改变生成器参数。两者的区别不在于是否使用特征后处理，而在于本文只约束批量表示的能量分配，并显式保持主子空间。

设一批生成特征组成 $B \times d$ 矩阵，其中 B 为批大小， d 为潜在维度。对该矩阵进行奇异值分解：

$$\hat{Z} = U\Sigma V^T, \Sigma = \text{diag}[s_1, s_2, \dots, s_q] \quad (4)$$

式中： $U \in \mathbf{R}^{B \times q}$ 和 $V \in \mathbf{R}^{d \times q}$ 分别为左、右奇异向量矩阵， $\Sigma \in \mathbf{R}^{q \times q}$ 为奇异值对角矩阵，其中奇异值按降序排列； $q = \text{rank}(\hat{Z})$ 为矩阵秩。本文所称“谱”特指批量潜在特征的奇异值谱，用于刻画主方向能量分布，不等同于原始传感信号的傅里叶频谱。

为保留主导表示子空间并抑制异常高阶扰动，设 τ 为谱截断阈值， $\rho > 0$ 为高阶谱衰减率，定义逐阶缩放系数和对角掩码矩阵：

$$m_i = \begin{cases} 1, & 1 \leq i \leq \tau, \\ \exp[-\rho(i - \tau)], & \tau < i \leq q, \end{cases} \quad (5)$$

$$M = \text{diag}[m_1, m_2, \dots, m_q]$$

式中： m_i 为第 i 阶奇异值的缩放系数；

$M \in \mathbf{R}^{q \times q}$ 为由全部缩放系数组成的对角掩码矩阵。谱截断阈值满足 $1 \leq \tau \leq q$ ；阈值内的缩放系数保持

为 1, 阈值后按阶数距离指数衰减。据此, 校准后的批量生成特征矩阵为

$$\hat{Z}_{\text{cal}} = U(\Sigma \square M)V^T \quad (6)$$

式中: $\square$ 表示矩阵对应元素相乘; $\hat{Z}_{\text{cal}} \in \mathbf{R}^{B \times d}$ 为潜在谱校准后的批量合成特征矩阵。式 (4) 的输入矩阵和式 (6) 的输出矩阵维度均为“批大小 $\times$ 潜在维度”, 该操作不增加可训练参数。

由于缩放系数 $(0,1]$, 该变换保持阈值内主奇异向量及奇异值不变, 仅衰减高阶奇异值, 因此校准后的 Frobenius 范数不增。该性质仅描述潜在特征结构约束, 不代表原始传感信号得到物理重建。

校准后的样本仍可能在部分维度过度集中, 因此依据可见类潜在特征的类内波动对 $\hat{Z}_{\text{cal}} \in \mathbf{R}^{B \times d}$ 施加按维度缩放的统计扰动, 增强矩阵为

$$\hat{Z}_{\text{aug}} = \hat{Z}_{\text{cal}} + \eta \Xi \text{diag}[\bar{\sigma}_{\text{intra}}] \quad (7)$$

式中: $\hat{Z}_{\text{aug}} \in \mathbf{R}^{B \times d}$ 为增强矩阵, $\Xi \in \mathbf{R}^{B \times d}$ 为标准高斯矩阵, $\eta \geq 0$ 为扰动尺度, $\bar{\sigma}_{\text{intra}} \in \mathbf{R}^d$ 为可见类训练特征估计的共享类内标准差向量。该扰动仅扩展潜在特征覆盖, 不依赖不可见类真实样本, 也不表示物理噪声。

3.4 可见—不可见类别偏置校准

分类器易因真实可见类与合成不可见类的分布差异而偏向可见类, 因此采用混合训练和推理偏置校准。

真实可见类潜在特征与合成不可见类潜在特征共同构成分类器训练样本:

$$\mathbf{D}_{\text{mix}} = \mathbf{D}_s \cup \hat{\mathbf{D}}_u \quad (8)$$

式中: 真实可见类潜在特征集合为 $\mathbf{D}_s$, 合成不可见类集合为 $\hat{\mathbf{D}}_u$, 混合训练集由二者的并集构成。

本文在混合训练集上训练轻量多层感知机分类器, 并采用交叉熵目标:

$$L_{\text{cls}} = -\frac{1}{|\mathbf{D}_{\text{mix}}|} \sum_{(z_i, y_i) \in \mathbf{D}_{\text{mix}}} \log p_{\psi}(y_i | z_i) \quad (9)$$

式中: ψ 为分类器参数; 类别不均衡时采用加权采样。

设分类器对类别 y 的未校准得分为输入, 则统一类别空间中的预测为

$$\hat{y} = \text{argmax}_{y \in Y_s \cup Y_u} [f_y(z) - \gamma I(y \in Y_s)] \quad (10)$$

式中: 指示函数判断候选类别是否属于可见类, γ 为可见类得分惩罚系数, 由可见类内部构造的伪不可见类验证任务确定。

3.5 云端生成训练—边缘轻量推理解耦流程

Diff-HAR 训练与部署流程如图 2 所示。时序特征提取器训练、SBERT 语义编码、条件扩散训练与采样、潜在谱校准、统计扰动和分类器训练均在云端离线完成; 目标部署模块包括时序特征提取器、轻量 MLP 分类器及偏置校准参数, 边缘在线推理不执行语义编码、生成采样、潜在谱校准或统计扰动。受 ESP32-S3 运行库对注意力算子的支持限制, 本文板端实验仅验证 96 维潜在特征回放—MLP 分类—偏置校准链路, 尚未完成原始传感输入到行为输出的端到端板端实测。

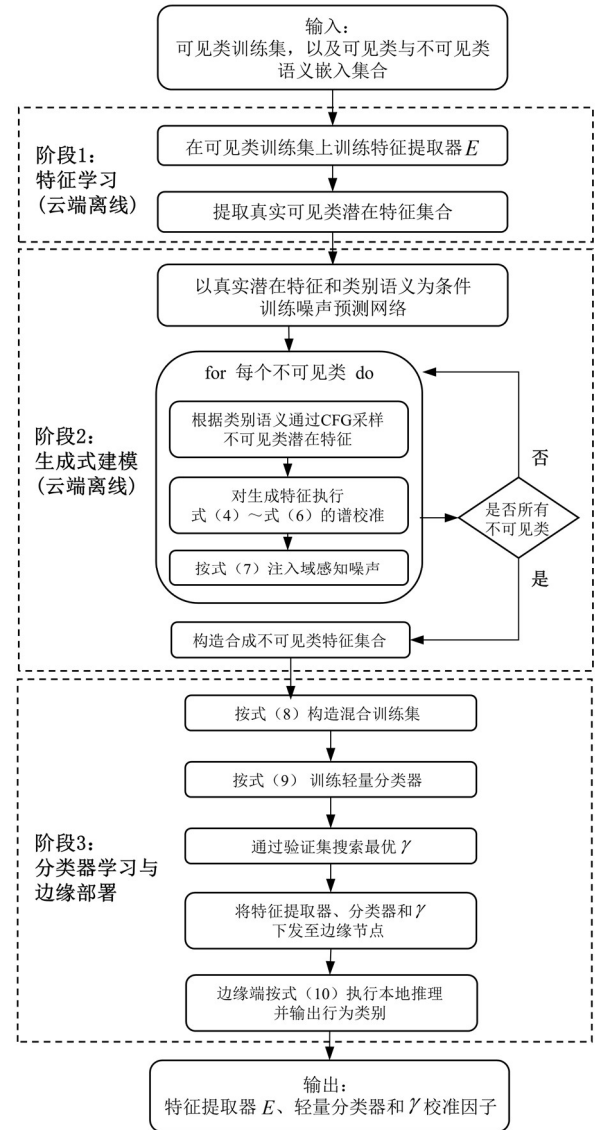


图 2 Diff-HAR 训练与部署算法流程
Fig. 2 Training and inference pipeline of Diff-HAR

4 实验设计与结果分析

4.1 实验设置

本文以 DriverActivity 数据集为智能驾驶主基准,

MM-Fi 数据集补充无线感知验证, PAMAP2 数据集和 USC-HAD 数据集补充惯性模态验证^[24-27]; 这些公开数据均不等同于真实车辆闭环测试。

DriverActivity 数据集含 11 类驾驶行为^[27]。本文固定 8 类可见行为(等待、安全驾驶、换挡、喝水、整理头发、与乘客交谈、查看手机、侧身取物)和 3 类不可见行为(查看后视镜、接电话、点头); 训练仅使用可见类真实样本, 不可见类只提供语义描述。

对比方法包括嵌入式、生成式及语言语义驱动方法^[11-14,16-20], 协议不一致者仅构建固定 8×3 模拟基线。

评价指标为可见类准确率 A_s 、不可见类准确率 A_u 、调和平均值 H 和可见—不可见准确率曲线下面积 A ; H 定义为

$$H = \frac{2A_s A_u}{A_s + A_u} \quad (11)$$

式中: A_s 与 A_u 为两类平均准确率, H 越高表明识别越均衡。

曲线下面积 A 通过不同校准强度下准确率组合点的梯形积分计算:

$$A = \sum_{j=1}^{J-1} \frac{A_u^{(j)} + A_u^{(j+1)}}{2} |A_s^{(j+1)} - A_s^{(j)}| \quad (12)$$

式中: $A_s^{(j)}$ 、 $A_u^{(j)}$ 为第 j 个校准点的两类准确率, J 为校准点数, $|A_s^{(j+1)} - A_s^{(j)}|$ 为相邻点间隔; A 越高表示整体折中越稳定。

实验使用 PyTorch 和 TensorFlow, 云端平台为

NVIDIA RTX 4060, 边缘平台为 ESP32-S3; 生成相关过程均在云端离线执行。

DriverActivity 数据集毫米波点云首先经过 PASA 统计聚合, 形成 15 维窗口级时序输入特征, 随后输入 Diff-HAR 特征提取模块。可见类样本按 8×1×1 划分, 并在可见类内部构造伪不可见类验证任务用于超参数选择; 真实不可见类样本仅用于最终测试, 不参与训练、验证及参数选择。

特征提取器含 2 层一维卷积和 4 层 Transformer 编码器, 潜在维度为 96。扩散模型训练 300 轮, 每个不可见类生成 500 个特征; 无分类器引导强度 $\omega = 3.0$, 统计扰动尺度 $\eta = 0.3$, 主导奇异值保留维数为 15。

轻量多层感知机输入维度为 96、隐藏层维度为 128, 训练 100 轮; 相关超参数均由上述验证任务确定, 真实不可见类测试集不参与参数选择。实验采用 42、123 和 2024 三个随机种子并报告均值和标准差。端到端时序编码器因运行库不支持注意力算子, 未进行板端实测, 因此, 本文边缘实验仅针对轻量分类头回放路径展开, 不涉及完整时序编码器模型部署。

4.2 方法性能对比

4.2.1 同协议主对比结果

表 1 列出同协议或具有明确复现口径的结果, 表 2 补充近期方法在统一协议下的近似对照结果。广义零样本识别的关键不是获得较高的单一可见类准确率, 而是在不可见类没有真实训练样本的条件下避免任一侧识别能力塌陷, 因此主要考察调和平均值及可见类—不可见类权衡曲线下面积。

表 1 同协议下与现有模型在 GZSL 任务上的性能对比

Table 1 Performance comparison with existing models under the same GZSL protocol

数据集	模型	可见类准确率/%	不可见类准确率/%	调和平均值/%
PAMAP2 数据集	ALE ^[11]	70.10±3.90	12.10±1.90	15.50±3.60
	DCN	42.20±0.90	33.10±0.20	36.70±0.30
	BERT ^[16]	74.70±0.00	49.90±0.70	59.30±0.00
	f-CLSWGAN ^[12]	92.40±0.20	27.80±1.10	41.70±1.50
	FREE ^[13]	87.70±0.30	37.20±0.20	52.10±0.20
	FM-IoT ^[17]	74.60±0.10	53.70±0.40	62.10±0.20
	本文模型 (Diff-HAR)	58.50±1.80	94.70±0.20	72.30±1.50
MM-Fi 数据集 (毫米波)	ALE ^[11]	86.50±0.10	0.01±0.00	2.00±0.00
	DCN	67.00±1.30	30.20±1.30	40.30±0.90
	BERT ^[16]	71.80±0.00	36.90±0.60	48.30±0.50
	f-CLSWGAN ^[12]	77.20±0.30	29.70±0.50	42.30±0.40
	FREE ^[13]	87.70±0.10	25.30±0.80	38.30±1.10
	FM-IoT ^[17]	73.30±0.00	40.40±0.50	51.70±0.30
	本文模型 (Diff-HAR)	62.20±3.20	47.90±8.20	53.70±5.10

数据集	模型	可见类准确率/%	不可见类准确率/%	调和平均值/%
USC-HAD 数据集	ALE ^[11]	92.50±0.00	0.60±0.00	1.10±0.00
	DCN	56.60±3.20	37.10±1.50	43.30±1.10
	BERT ^[16]	74.90±0.10	41.60±1.30	52.20±0.70
	f-CLSWGAN ^[12]	81.30±0.50	29.20±3.50	39.50±4.90
	FREE ^[13]	90.90±0.10	14.00±0.60	23.20±1.60
	FM-IoT ^[17]	73.10±0.50	54.80±1.80	61.10±0.70
	本文模型 (Diff-HAR)	57.09±0.53	66.65±7.12	61.33±2.89
DriverActivity 数据集	FM-IoT ^[17]	50.10±41.80	13.00±16.70	5.45±2.91
	WGAN-GP	49.50±3.10	39.60±2.20	43.90±2.10
	CVAE	49.20±2.80	28.90±1.40	36.40±1.60
	本文模型 (Diff-HAR)	51.00±2.10	39.00±3.10	44.10±2.10

注: DriverActivity 数据集上的 FM-IoT 结果为基于公开描述和统一输入协议得到的对照实验结果; PAMAP2、USC-HAD 和 MM-Fi 的可见/不可见类别划分、数据预处理及评价协议均沿用文献[17], 各方法采用相同固定划分; DCN、WGAN-GP 和 CVAE 基线均按文献[17]公开的模型设置和评价协议构建; 近期方法的近似对照结果见表 2; 加粗数据表示最优结果。

Diff-HAR 在 DriverActivity、PAMAP2、USC-HAD 和 MM-Fi 数据集上的调和平均值分别为 44.1%±2.1%、72.3%±1.5%、61.33%±2.89%和 53.7%±5.1%; DriverActivity 上的曲线下面积为 0.313±0.036。DriverActivity 上的 44.1%不应解读为接近封闭集识别上限, 也不能据此宣称相对 WGAN-GP 具有显著优势, 两者调和平均值分别为 44.1%±2.1%和 43.9%±2.1%。其工程价值在于: 在点云统计聚合可能损失细粒度信息、3 个不可见行为完全没有真实训练样本的固定 8:3 协议下, Diff-HAR 仍同时取得 51.0%±2.1%的可见类准确率和 39.0%±3.1%的不可见类准确率, 未出现单侧识别能力塌陷。结合曲线下面积和跨模态结果, 该结果可作为行为类别扩展可行性的证据, 但不代表量产座舱精度。

4.2.2 统一协议下的近期方法近似对照

表 2 在统一的 DriverActivity 数据集预处理和固定 8:3 协议下构建 GZSL-EM、HARGPT、ZeroHAR 和 IMUZero 的近似对照结果^[14,18-20], 不代表原作者代码的严格复现或官方结果。在该统一近似设置下, 单纯语义映射或提示增强难以同时保持两类准确率;

语言驱动的跨模态对齐相对均衡, 但其调和平均值仍低于本文方法。

表 2 近期模型在 DriverActivity 数据集固定 8:3 GZSL 协议下的近似对照结果

Table 2 Approximate comparison under the fixed 8/3 GZSL protocol on the DriverActivity dataset			
模型	可见类准确率/%	不可见类准确率/%	调和平均值/%
GZSL-EM ^[14]	39.26	24.00	29.79
HARGPT ^[18]	23.21	32.12	26.95
ZeroHAR ^[19]	22.91	32.76	26.96
IMUZero ^[20]	37.67	30.66	33.81

4.3 消融实验

表 3 和表 4 给出两数据集的代表性算法消融结果, 表 1 主结果为多次运行的均值和标准差。消融仅用于评价语义条件生成、潜在谱校准、统计扰动和偏置校准等算法机制; 云边解耦属于系统执行架构, 其可运行性在 4.6 节单独实测。

表 3 PAMAP2 数据集上关键组件的消融实验

Table 3 Ablation study of key components on PAMAP2 dataset					
配置	可见类准确率/%	不可见类准确率/%	调和平均值/%	最优偏置校准因子	曲线下面积
去除无分类器引导	37.77	95.17	54.07	15	0.584
去除潜在谱校准	57.64	95.17	71.79	27	0.728
去除统计扰动	53.05	68.62	59.20	25	0.636
去除偏置校准	98.19	5.45	10.32	0	-
线性分类器	56.16	56.30	56.23	7	0.610
完整模型 (Diff-HAR)	60.43	93.74	73.48	23	0.728

表 4 DriverActivity 数据集上关键组件的消融实验

Table 4 Ablation study of key components on DriverActivity dataset					
配置	可见类准确率/%	不可见类准确率/%	调和平均值/%	最优偏置校准因子	曲线下面积
去除无分类器引导	32.00	40.85	33.15	10	0.242

配置	可见类准确率/%	不可见类准确率/%	调和平均值/%	最优偏置校准因子	曲线下面积
去除潜在谱校准	48.83	40.85	44.01	22	0.302
去除统计扰动	44.95	29.44	36.30	20	0.264
去除偏置校准	87.35	0.27	0.53	0	-
线性分类器	47.58	24.15	34.47	2	0.253
完整模型 (Diff-HAR)	51.20	40.23	45.05	19	0.302

注：去除偏置校准时固定 $\gamma = 0$ ，不计算基于校准扫描的曲线下面积。

消融结果体现了各机制的作用层次：语义条件生成负责构造不可见类样本；去除潜在谱校准后，PAMAP2 和 DriverActivity 上的调和平均值分别下降 1.69 和 1.04 个百分点，说明其作用是提供增量式结构校准，而非单独决定全部性能；统计扰动用于改善覆盖。去除偏置校准后，两数据集的不可见类准确率分别降至 5.45% 和 0.27%，表明决策校准主要负责两类平衡。云边解耦流程的运行开销与部署结果见 4.6 节。

4.4 潜在谱校准与关键参数分析

除特别说明外，图 3~图 7 均采用 PAMAP2 数据集上的代表性运行结果。图 3~图 5 给出生成引导、决策校准和统计扰动等辅助参数的选择结果；图 6 结合相应消融结果分析潜在谱校准机制。

4.4.1 无分类器引导强度分析

图 3 显示， $\omega = 3$ 附近取得最佳识别折中；过弱引导不能充分利用语义，过强引导则降低样本多样性。

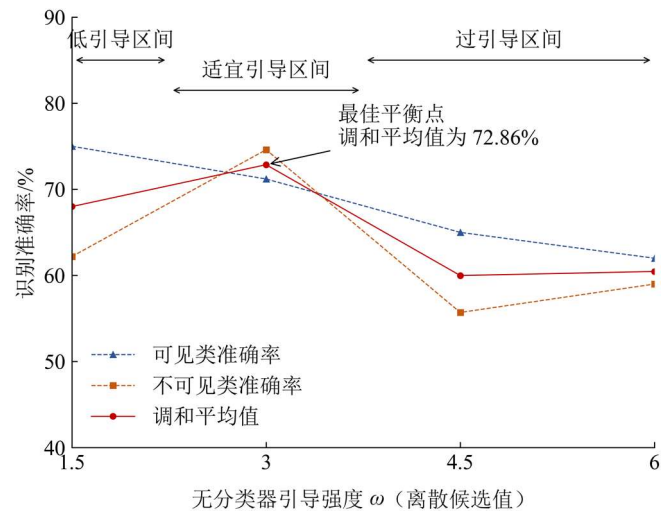


图 3 无分类器引导强度对调和平均值的影响
Fig. 3 Effect of classifier-free guidance scale on H-mean

4.4.2 偏置校准因子敏感性分析

图 4 表明 $\gamma = 23$ 时调和平均值达到峰值，据此确定默认偏置校准参数。

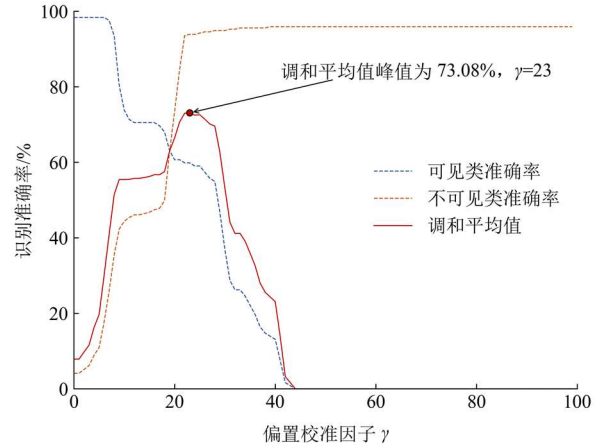


图 4 偏置校准因子的敏感性分析
Fig. 4 Sensitivity analysis of the bias-calibration factor

4.4.3 统计扰动尺度敏感性分析

图 5 显示 $\eta = 0.3$ 时调和平均值最高，说明适度统计扰动有利于兼顾覆盖范围和语义保持。

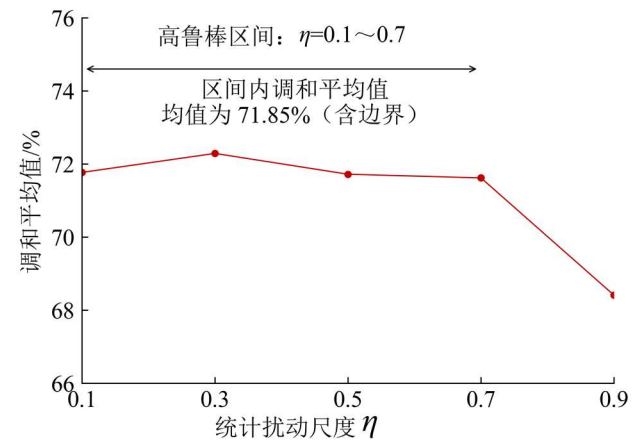


图 5 统计扰动尺度对调和平均值的影响
Fig. 5 Effect of statistical perturbation scale on H-mean

4.4.4 潜在谱校准机制有效性验证

图 6 比较真实特征、Diff-HAR 和生成对抗网络基线的潜在特征奇异值谱。校准后的生成特征在保持主要谱方向的同时，其高阶奇异值衰减趋势更接近真实特征，与式 (5)—式 (6) 的非扩张设计一致。结合去除谱校准后的调和平均值下降，可认为该机制改善了生成特征的批量结构一致性；该证据支持结构校准机制解释，但不能证明原始传感信号已被物理重建。该结果仅说明潜在特征谱的一致性改善，不等同于重建原始惯性信号、微多普勒图或点云结

构, 并且真实不可见类特征仅用于训练完成后的可视化分析, 不参与模型训练、超参数选择或偏置校准。

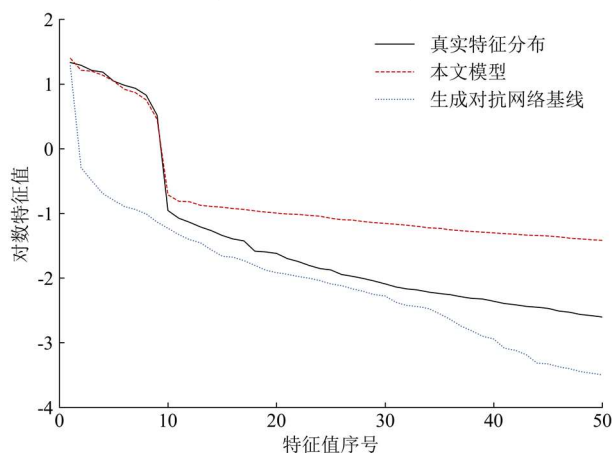


图 6 不同生成方法的潜在谱一致性分析

Fig. 6 Latent spectral consistency across generation methods

4.5 补充加性扰动检验

图 7 显示, 调和平均值随高斯加性噪声增强而下降: 噪声强度不超过 0.8 时仍不低于 70.3%, 增至 1.0 时降至 67.7%。该结果仅提供随机扰动下的初步稳定性证据, 不能替代真实车载扰动测试。

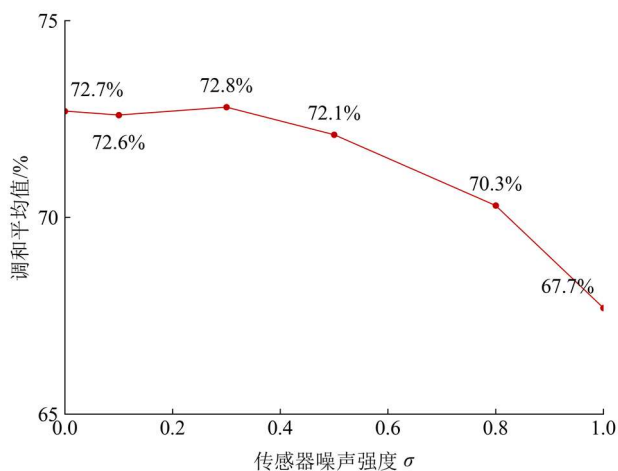


图 7 加性噪声强度对调和平均值的影响

Fig. 7 Effect of additive-noise intensity on the harmonic mean

4.6 边缘轻量分类链路实测

本文选择算力较低、存储受限的 ESP32-S3 作为资源受限边缘节点代表, 以严格资源条件检验云边解耦后轻量分类链路的可执行性。表 5 显示, “96 维潜在特征回放—MLP 分类—偏置校准”链路完成板端实测, 模型大小为 0.24 MB, 平均推理延迟为 9.379 ms。该实验输入为离线获得的 96 维潜在特征, 仅评价轻量分类链路在有限资源条件下的运行能力, 不包含原始传感输入或完整 Diff-HAR 端到端部署。

表 5 边缘轻量分类链路回放实测结果

Table 5 Edge Replay Results of the Lightweight Classification Path

指标	FM-IoT 参考模型	Diff-HAR 端到端导出	轻量分类头板端回放
模型大小 /MB	224.02	1.59	0.24
运行口径	部署失败 (内存不足)	仅导出	96 维多层感知机分类头
平均推理延迟/ms	部署失败 (内存不足)	(运行库不支持)	9.379
平均总延迟/ms	部署失败 (内存不足)	(运行库不支持)	11.598
单次推理能耗/mJ	不适用	(运行库不支持)	7.41
单次链路能耗/mJ	不适用	(运行库不支持)	9.16

注: 未测试项分别源于内存不足或运行库不支持 BATCH_MATMUL; 可运行结果为 96 维多层感知机分类头回放, 能耗按约 790 mW 活动功耗估算。

5 结论

本文针对训练阶段未见智能驾驶行为缺少真实样本的问题, 提出 Diff-HAR, 以语义条件生成补充不可见类潜在特征, 并通过主子空间保持的潜在谱校准缓解结构偏移。DriverActivity 上的调和平均值为 44.1%, 消融与谱分析支持校准的增量式改善。ESP32-S3 完成“96 维潜在特征回放—MLP 分类—偏置校准”测试, 验证了资源受限节点上轻量分类链路的可执行性。

当前实验主要验证资源受限节点上的轻量分类能力。后续将进一步探索完整模型在不同计算平台上的部署优化, 并结合真实车辆数据流与扰动条件开展验证。

参考文献:

- [1] 黄昭彦, 杨烁, 吴建华, 等. 基于信息融合的智能网联汽车安全交互决策[J]. 自动化学报, 2025, 51(9): 1883-1898.
Huang Zhaoyan, Yang Shuo, Wu Jianhua, et al. Safety interactive decision-making for intelligent connected vehicles based on information fusion[J]. Acta Automatica Sinica, 2025, 51(9): 1883-1898.
- [2] 张新钰, 卢毅果, 高鑫, 等. 面向智能网联汽车的车路协同感知技术及发展趋势[J]. 自动化学报, 2025, 51(2): 233-248.
Zhang Xinyu, Lu Yiguo, Gao Xin, et al. Vehicle-road collaborative perception technology and development trend for intelligent connected vehicles[J]. Acta Automatica Sinica, 2025, 51(2): 233-248.
- [3] 丁飞, 张楠, 李升波, 等. 智能网联车路云协同系统架构与关键技术研究综述[J]. 自动化学报, 2022, 48(12): 2863-2885.

- Ding Fei, Zhang Nan, Li Shengbo, et al. A survey of architecture and key technologies of intelligent connected vehicle-road-cloud cooperation system[J]. *Acta Automatica Sinica*, 2022, 48(12): 2863-2885.
- [4] 陈妍妍, 田大新, 林椿晒, 等. 端到端自动驾驶系统研究综述[J]. *中国图象图形学报*, 2024, 29(11): 3216-3237.
- Chen Yanyan, Tian Daxin, Lin Chunmian, et al. Survey of end-to-end autonomous driving systems[J]. *Journal of Image and Graphics*, 2024, 29(11): 3216-3237.
- [5] 董连飞, 马志雄, 朱西产. 基于车载毫米波雷达动态手势识别网络[J]. *北京理工大学学报*, 2023, 43(5): 493-498.
- Dong Lianfei, Ma Zhixiong, Zhu Xichan. Dynamic gesture recognition network based on vehicular millimeter wave radar[J]. *Transactions of Beijing Institute of Technology*, 2023, 43(5): 493-498.
- [6] 朱煜, 赵江坤, 王逸宁, 等. 基于深度学习的人体行为识别算法综述[J]. *自动化学报*, 2016, 42(6): 848-857.
- Zhu Yu, Zhao Jiangkun, Wang Yining, et al. A review of human action recognition based on deep learning[J]. *Acta Automatica Sinica*, 2016, 42(6): 848-857.
- [7] 赵荣峰, 卢宝莉, 唐小江, 等. 面向智能座舱的多源混合模态数据集及层次化融合分类方法[J]. *智能系统学报*, 2026, 21(1): 83-94.
- Zhao Rongfeng, Lu Baoli, Tang Xiaojiao, et al. Multi-source hybrid-modality dataset and hierarchical fusion classification method for intelligent cockpits[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(1): 83-94.
- [8] 胡宏宇, 黎烨宸, 张争光, 等. 基于多尺度骨架图和局部视觉上下文融合的驾驶员行为识别方法[J]. *汽车工程*, 2024, 46(1): 1-8.
- Hu Hongyu, Li Yechen, Zhang Zhengguang, et al. Driver behavior recognition based on multi-scale skeleton graph and local visual context method[J]. *Automotive Engineering*, 2024, 46(1): 1-8.
- [9] 李少凡, 高尚兵, 张莹莹. 用于驾驶员分心行为识别的姿态引导实例感知学习[J]. *中国图象图形学报*, 2023, 28(11): 3550-3561.
- Li Shaofan, Gao Shangbing, Zhang Yingying. Pose-guided instance-aware learning for driver distraction recognition[J]. *Journal of Image and Graphics*, 2023, 28(11): 3550-3561.
- [10] 吕露露, 黄毅, 高君宇, 等. 多模态零样本人体动作识别[J]. *中国图象图形学报*, 2021, 26(7): 1658-1667.
- Lyu Lulu, Huang Yi, Gao Junyu, et al. Multimodal-based zero-shot human action recognition[J]. *Journal of Image and Graphics*, 2021, 26(7): 1658-1667.
- [11] Akata Z, Perronnin F, Harchaoui Z, et al. Label-embedding for image classification[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(7): 1425-1438.
- [12] Xian Y, Lorenz T, Schiele B, et al. Feature generating networks for zero-shot learning[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 5542-5551.
- [13] Chen S, Wang W, Xia B, et al. FREE: Feature refinement for generalized zero-shot learning[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021: 122-131.
- [14] Wang W, Li Q. Generalized zero-shot activity recognition with embedding-based method[J]. *ACM Transactions on Sensor Networks*, 2023, 19(3): 1-25.
- [15] Reimers N, Gurevych I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks[C]//*Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. 2019: 3982-3992.
- [16] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//*Proceedings of NAACL-HLT*. 2019: 4171-4186.
- [17] Xue D, Fan X, Chen T, et al. Leveraging foundation models for zero-shot IoT sensing[C]//*Proceedings of the 27th European Conference on Artificial Intelligence*. Amsterdam: IOS Press, 2024: 3805-3812.
- [18] Ji S, Zheng X, Wu C. HARGPT: Are LLMs zero-shot human activity recognizers?[C]//*2024 IEEE International Workshop on Foundation Models for Cyber-Physical Systems & Internet of Things (FMSys)*. 2024: 38-43.
- [19] Chowdhury R R, Kapila R, Panse A, et al. ZeroHAR: Sensor context augments zero-shot wearable action recognition[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. 2025, 39(15): 16046-16054.
- [20] Su J, Ge F, Wen Z, et al. IMUZero: Zero-shot human activity recognition by language-based cross modality fusion[J]. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2025, 9(4): 1-28.
- [21] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models[C]//*Advances in Neural Information Processing Systems*. 2020, 33: 6840-6851.
- [22] Oppel H, Munz M. A diffusion model for inertial based time series generation on scarce data availability to improve human activity recognition[J]. *Scientific Reports*, 2025, 15: 16841.
- [23] Ho J, Salimans T. Classifier-free diffusion guidance[EB/OL]. arXiv:2207.12598, 2022[2026-06-24].
- [24] Reiss A, Stricker D. Introducing a new benchmarked dataset for activity monitoring[C]//*Proceedings of the 16th International Symposium on Wearable Computers*. 2012: 108-109.
- [25] Zhang M, Sawchuk A A. USC-HAD: A daily activity dataset for ubiquitous activity recognition using wearable sensors[C]//*Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. 2012: 1036-1043.
- [26] Yang J, Huang H, Zhou Y, et al. MM-Fi: Multi-modal non-intrusive 4D human dataset for versatile wireless sensing[C]//*Advances in Neural Information Processing Systems*. 2023, 36.
- [27] Li G H, Chiang H C, Li Y C, et al. A driver activity dataset with multiple RGB-D cameras and mmWave radars[C]//*Proceedings of the 15th ACM Multimedia Systems Conference*. 2024: 360-366.

(编辑 刘杨 陶晴 刘懿莹)