同行专家业内评价意见书编号：20250861026

附件1

浙江工程师学院（浙江大学工程师学院）
同行专家业内评价意见书

姓名：何琦

学号：22260183

申报工程师职称专业类别（领域）：交通运输

浙江工程师学院（浙江大学工程师学院）制

2025年05月26日

填表说明

一、本报告中相关的技术或数据如涉及知识产权保护、军工项目保密等内容，请作脱密处理。

二、请用宋体小四字号撰写本报告，可另行附页或增加页数，A4纸双面打印。

三、表中所涉及的签名都必须用蓝、黑色墨水笔，亲笔签名或签字章，不可以打印代替。

四、同行专家业内评价意见书编号由工程师学院填写，编号规则为：年份4位＋申报工程师职称专业类别(领域)4位+流水号3位，共11位。

一、个人申报

（一）基本情况【围绕《浙江工程师学院（浙江大学工程师学院）工程类专业学位研究生工程师职称评审参考指标》，结合该专业类别(领域)工程师职称评审相关标准，举例说明】

1.对本专业基础理论知识和专业技术知识掌握情况(不少于200字)

作为交通运输工程专业研究生，我在系统化课程学习与科研实践中，构建了扎实的理论知识体系，并在智能交通系统、交通流建模、交通仿真与优化等领域积累了专业技术能力，具体掌握情况如下：
1.交通工程原理
系统掌握交通流理论核心模型（Intelligent Driver Model
智能模型、Newell跟驰模型）、通行能力计算方法，能够运用泊松分布、排队论等数学工具解析交叉口车流特征。完成《交通工程学》《工程师创新创业思维》等课程的学习（最好成绩96/100），构建了从微观车辆动力学到宏观路网运行分析的完整认知框架。
2.智能交通系统理论
深入研究车路协同（Vehicle to Everything,V2X）通信协议、自动驾驶感知-
决策链技术架构（多传感器标定、轨迹预测算法），在《交通运输工程科学与技术前沿》课程中完成基于深度学习的异常驾驶行为识别任务（模型准确率提升6.8%）。
3.交通规划方法论
熟练掌握四阶段法（出行生成-分布-方式划分-
分配）、Logit模型应用，并且以杭州市居民出行即服务（Mobility as a
Service，MaaS）为基础，构建了基于可解释机器学习模型的 MaaS
方案选择行为分析框架。结果显示，年龄、是否持有公交月卡、出行模式、感知风险及使用意向为影响 MaaS 方案选择的关键因素。具体而言，老年用户（55
岁及以上）、私家车用户倾向于不订阅方案（维持现状）；年轻用户（18~34
岁）、私家车＆出租车/网约车用户倾向于订阅周方案；持有公交月卡用户、出行模式含有公交的用户（电动自行车＆公交用户、公交＆地铁用户、公交用户）倾向于订阅月方案。

2.工程实践的经历(不少于200字)

1.交通仿真与数据分析
精通VISSIM（二次开发COM接口）、SUMO（TraCI控制模块）仿真工具，主导完成重型货车碰撞预警仿真项目：该方案通过非线性最小二乘法标定异质性驾驶行为参数，使跟驰轨迹中90%以上的轨迹拟合误差控制在1.5m以内。采用自适应核密度估计耦合蒙特卡洛采样的场景生成技术，使KL散度稳定于0.18-
0.30，场景覆盖度较GMM方法提升51.2%。基于SUMO平台构建动态交通流环境，设计四类典型工况测试表明：A-
TTC算法在本研究所有测试场景中避撞成功率最高（>99%），但其偏保守的预警策略涉及，在车辆低速行驶时触发频率较高，可能干扰驾驶员正常操作；Transformer模型因训练样本不足导致慢减速场景失败率近2%，但其动态的预警时机使车辆在整个制动过程保持安全性的同时也兼顾驾驶舒适性，有助于提高用户接受度。仿真结果验证了自适应预警算法在常规工况中的优越性，为智能驾驶系统迭代提供了标准化测试范式，成果发表于EI会议（第一作者）。
2.交通大数据挖掘
在深圳市车米云图企业实习期间，针对单模态数据对驾驶行为表征能力有限的问题，主导完成了多模态注意力神经网络（MMANN）的构建。该模型集成三大特征提取模块：1）CNN-

ViT模块采用局部Transformer层提取驾驶员面部微表情特征；2）MCViViT模块通过管道采样构建时空序列；3）AT-BiGRU模块利用双向门控单元捕捉操纵行为时序特征。此外，通过改进的OvT多模态注意力机制，模型在测试集上实现82.4%的准确率，较最优单模态模型提升8.6%。消融试验表明多模态融合的优势性，其中驾驶环境特征对识别准确率提升贡献了5.82%，驾驶员面部图片贡献了4.12%，成果发表于交通领域高质量国际会议（第一作者），并作讲台汇报。

3.在实际工作中综合运用所学知识解决复杂工程问题的案例（不少于1000字）

实践案例

——基于多模态数据融合的自动驾驶虚拟测试平台开发

一、项目背景与问题定义

在深圳市车迷云图科技有限公司实习期间，我主导了驾驶员异常行为预警系统的研发工作。该模块需解决行业两大痛点：

1）传统预警系统对复杂驾驶行为适应性差：基于规则引擎的阈值判定方法在驾驶员分心（如使用手机）、疲劳（PERCLOS≥0.4）等场景下误报率高达28%，且无法识别新型风险行为（如车机系统过度交互）；

2）多模态数据利用率低：现有方案仅依赖方向盘转角、车速等单一信号，未有效融合面部表情（眼睑闭合度、头部姿态）、语音交互（指令响应延迟）等跨模态特征，导致关键风险漏检率超过15%。

二、理论知识与技术方案

1. 异常行为建模理论框架（对应课程：交通心理学、人因工程学）

风险特征库构建:

基于NHTSA自然驾驶数据库与自采的2,300小时卡车驾驶视频，提取三级风险标签：

一级风险：持续分心（≥5秒视线偏离）

二级风险：疲劳累积（心率变异系数HRV<30ms且眨眼频率<10次/分钟）

三级风险：应激反应（方向盘转角突变率＞3°/s^2并伴随急刹）

对驾驶风险进行准确识别有助于预测驾驶员在危险发生时是否能够迅速反应，并采取正确的风险规避措施，其可以被表述为一个时间序列分类问题，即给定时长为 l
的多模态时间序列 X^l，驾驶员的异质性驾驶行为可分类为：分心组（ y)
^_j=0，正常组(y) ^_j=1，及压力组(y) ^_j=2。公式如下：

y ^_j= 〖arg max┬(y ^_j ∈{0,1,2})〗fo 〖p(y_j |X_j^l 〗) ,j=1,2,⋯,N

X^l=[〖 x〗_1,x_2,⋯,x_l]

其中， N 表示样本大小，x 表示每个时间步的特征。

在自然驾驶数据集中时间序列信息由前置摄像头获取的视频〖 M〗^l
和车内摄像头获取的驾驶员面部图像 G^l、车载传感器获取的车辆运动学数据〖 K〗^l
共同组成，表示为X^l=[M^l,G^l,K^l]。具体而言，x=[m,g,k]，其中 k=[v,a,Δs,Δv]
，包括重型货车的速度v、加速度a，以及与前车的相对距离Δs和相对速度Δv。

2.混合驱动建模：

融合数据驱动（多模态时序特征挖掘）与知识驱动（驾驶行为风险等级规则），构建动态风险评估模型。

受Transformer模型及驾驶人注意力机制的启发，本研究以三种模态数据特征提取模型为基础，提出一种基于多模态信息融合的异质性驾驶行为识别模型（Multimodal Attention Neural Network, MMANN）。

首先，MMANN将每个模态数据传递给各自的特征提取器：MCViViT模型用于视频；CNN-ViT模型用于图像；AT-BiGRU模型用于车辆状态数据。然后，提取的特征向量通过OvT多模态注意力层进行融合。最后，通过多层感知机（Multilayer Perceptron，MLP）输出异质性驾驶行为的分类结果。

（1）驾驶员面部图像特征提取

Transformer模型已成为自然语言处理的首选，尤其是与自注意力机制结合使用。Transformer处理具有时序特征的视觉数据时，由于高维特性，常常面临模型的复杂性和时间信息利用不足问题。针对此挑战，本研究基于多尺度视觉Transformer框架构建CNN-ViT模型，进行驾驶员面部特征提取，其结构如图2所示：W_q 、W_k 分别为Query权重和Key权重，v_{pos} 为注意力位置中心编码， r_{ij} 为相对位置编码，σ为sigmoid函数，ω是门控参数。

在CNN-ViT模型中，给定一幅输入图像，记作$G \in R^{H*W*C}$，首先将原始图片重塑为一系列图像块，记作$G_{flat} \in R^{N*P^2*C}$，其中 H,W,C,P 分别代表原始多通道图像的高度、宽度、通道数和图像块大小，$N=(H*W)/P^2$ 代表图像块数量。每个图像块转换为低维向量表征，通过位置编码将图像的全局位置信息引入模型中。然后，将编码向量送入局部Transformer层，提取的局部特征向量通过标准Transformer层，生成驾驶员状态特征表征向量Lat^G。其中，标准Transformer层包括多头自注意力模块（Multi-head Self-Attention，MSA）和多层感知机模块。在MSA模块中，每个自注意力头 h 执行以下操作：

$$MSA_h(G)=A^h GW^h$$

其中，W^h 是权重矩阵，A^h是注意力矩阵。

此外，在局部Transformer层使用位置自注意力层（Positional Self-Attention，PSA）替代自注意力层，以实现局部感知，即A^h 被替换为A_{IJ}^h，该注意力矩阵表示为

$$A_{ij}^h=softmax({Q_i^h K_j^h}^T+{V_{pos}^h}^T r_{ij})$$

每个注意力头h 使用一个可学习的注意力中心位置编码$V_{pos}^h \in R^{D_{pos}}$，并且相对位置编码 $r_{ij} \in R^{D_{pos}}$ 表示图像块i与j之间的相对位置，其大小仅取决于像素 i 与 j 之间的距离。当PSA层包含h个注意力头且相对位置编码维度 $D_{pos} \geq 3$时，PSA可被视为卷积核大小为 $\sqrt{h} \times \sqrt{h}$ 的卷积层，优化后的该网络层更关注图像的局部特征。

（2）驾驶环境视频特征提取

视频数据信息丰富，可以提取交通标识、车道线、交通参与者、天气和环境等信息。由于视频数据会产生大量时空图像块序列，必须考虑长时间范围内图像序列间的关联性和模型的效率问题。基于视觉Transformer架构，本研究使用一种改进的多通道视频神经网络MCViViT模型提取驾驶环境特征。给定一个视频输入，记作$M \in R^{T*H*W*C}$，采用管道采样的方法对图像块编码，即利用管道$t \in R^{t*h*w}$ 进行采样，可获得$n_t*n_h*n_w$ 个图像块序列 $\tilde{m} \in R^{n_t*n_h*n_w*C}$，其中$n_t=T/t$，$n_w=W/w$，$n_t=T/t$。然后，将位置编码和类别编码添加到每个序列$\tilde{m}$ 中。

视觉Transformer采用全局机制将不同位置联系起来，可以同时对空间和时间信息进行建模。而与视觉Transformer均匀采样方法相比，改进的MCViViT模型构建的序列可以高效融合时空信息。MCViViT的空间编码器和时间编码器结构如图3所示。首先，空间编码器对同一时间

索引的图像序列间的交互信息进行提取，获得每一时间索引的表征$h_i \in R^d$。然后，将全部空间表征向量拼接并送入时间编码器以聚合来自不同时间索引的表示，对应于时空信息的后期融合。时间编码器的输出即为驾驶环境特征提取的表征向量Lat^M。MCViViT模型中浅层的局部Transformer层以高空间分辨率模拟低层视觉信息，并在空间粗糙的深层表示复杂的高层特征。

（3）驾驶操纵行为特征提取

车辆动力学数据（如速度、加速度、横摆角等）适合构建为多变量时间序列，包含各种驾驶行为特征。循环神经网络在语音识别和语言翻译方面取得了巨大成功。为增强时间序列数据的学习能力，两种常用有效变体分别为：长短期记忆网络和门控循环单元网络。考虑到驾驶操纵行为数据的特征，本研究采用双向门控循环单元来捕捉时间序列的双向依赖性，并使用注意力机制来突出重要时间步。该模型内部信息传输和网络结构如图4所示。

驾驶操纵数据作为注意力层的输入，为双向门控循环单元层（BiGRU）的输入分配权重。权重越大，表示该时间步的数据越重要，其计算公式为

$$\begin{cases} w_t = softmax(k_t h_t + b_t) \\ S = \sum_{t=1}^{T_l} w_t h_t \end{cases}$$

其中，h_t 是某一时间点的双向门控循环单元网络的隐藏层输出，b、k分别是偏置和权重，T_l 是驾驶行为序列的总长度，S 是注意力层的最终表示向量。

BiGRU从前、后两个方向获取特征向量。门控循环单元计算由注意力层传递过来的特征向量，并输出固定维度的向量作为驾驶操纵行为特征的表征向量Lat^K。在门控循环单元内部，计算过程表示为

$$\begin{cases} r_t = \sigma(K_r * [h_{t-1}, s_t] + b_r) \\ z_t = \sigma(K_z * [h_{t-1}, s_t] + b_z) \\ \tilde{h}_t = \tanh(K_{\tilde{h}} * [r_t * h_{t-1}, s_t] + b_{\tilde{h}}) \\ h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \end{cases}$$

其中，tanh 是双曲正切函数，s_t 是门控循环单元层的输入向量，h_t 是门控循环单元的输出向量，r_t 是重置门状态向量，z_t 是更新门状态向量。

（4）多模态信息融合与异质性驾驶行为识别

目前车辆动力学数据多依赖自注意力或交叉注意力进行数据特征融合，但扩展性较差，难以应用于超过两种模态的复杂数据集。针对此问题，本文采用改进的OvT注意机制[83]对不同模态数据进行有效融合，并在此基础上理解每种模态的相对重要性，模型结构如图5所示。

通过各特征提取器获得的表征向量 $m_i \in Lat^M, Lat^G, Lat^K$，分别计算各模态 m_i 与其他模态的组合表示之间的相似性。计算公式可表示为

$$score(m_i) = m_i^T K \left(\sum_{i \neq j}^{k} m_j\right) / (k-1)$$

$$m_{att_i} = softmax(score(m_i) * m_i)$$

其中，m_{att_i} 是融合其他模态信息后的表征向量，k是模态数量，K 是权重矩阵。多模态融合特征向量输入到多层感知机中，输出每个类别的置信值，以形成最终识别结果。然后，将批训练样本识别的置信度与真值进行比较，以计算损失并进行反向传播。训练过程结束后，模型根据最高置信度分数判定驾驶员认知状态类别。

三、技术创新与工程成果

1）动态风险分级预警算法：

基于异质性驾驶行为与风险场景的耦合关系，提出两种互补的自适应预警策略，突破传统预警算法局限：

a.动态预警阈值优化算法：融合驾驶员认知状态、车辆运动状态以及安全避险时间，构建预警阈值动态调整模型，实现阈值随驾驶行为自适应的预警触发机制。
b.多因子融合的预警算法：基于机器学习构建融合驾驶员异质性行为、预警反应时间及交通环境状态的安全距离预测模型，设计安全距离预测-
预警时机调节的联合优化策略，平衡紧急制动与误报抑制的冲突需求。
为开发适用于配备驾驶辅助系统的碰撞预警算法，采用车米云图公司的重型货车自然驾驶数据进行算法测试与验证。图6展示了不同类型的预警算法在驾驶员状态差异下的预警效果，Mazda模型（橙色虚线）趋近于保守，因此预警距离远远大于车间距，在真实交通环境中容易发生误警情况。固定时间阈值模型（TTC=2.7s，黄色虚线）的预警距离波动较小，且该模型无法根据驾驶员异质性动态调整预警时刻，算法有效性较低。基于机器学习方法（粉色虚线）与基于阈值优化的自适应预警策略（蓝色虚线），相较于传统碰撞预警算法，更趋近人类驾驶员特性。这两种算法通过融合驾驶员状态、道路环境特征等，实现预警时刻动态调整，表明所提出的两种自适应算法在驾驶员异质性建模方面的优势，有利于增强驾驶员对预警系统的信任度。

通过上述分析，可以采用具体的预警反馈指标来分析各预警算法的有效性。其中，以0.1g作为判断驾驶员制动行为的阈值，较高的阈值可以捕获更多的低风险样本，这将提供一个更为完整的驾驶员风险偏好视图。如表1所示，可将FCW样本分为“安全”和“威胁”两类事件， N 为事件数量。

表 1　碰撞预警算法评估指标定义

	有预警信号	无预警信号
威胁事件	N_1	N_2
安全事件	N_3	N_4

采用的性能指标包括准确率、精确度、召回率、特异度和误报率，其中误报率在本项目中定义为所有预警事件中安全事件所占的比例，仿照定义漏警率为威胁发生时未预警事件所占的比例。将六个指标分别表示为ϕ_1 $(=(N_1+N_4)/(N_1+N_2+N_3+N_4)*100\%)$，$\phi_2$ $(=N_1/(N_1+N_3)*100\%)$，ϕ_3 $(=N_1/(N_1+N_2)*100\%)$，ϕ_4 $(=N_4/(N_3+N_4)*100\%)$，ϕ_5 $(=N_3/(N_1+N_3)*100\%)$，ϕ_6 $(=N_2/(N_1+N_2)*100\%)$ 。

表 1　碰撞预警算法有效性评估结果

模型	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
Transformer模型	81.78%	90.62%	84.23%	81.60%	9.38%	15.77%
A-TTC模型	74.31%	74.84%	98.81%	3.13%	25.16%	1.19%
TCN模型	68.71%	84.89%	70.37%	63.89%	15.11%	29.63%
TTC=2.7s	70.75%	74.47%	92.35%	8.68%	25.53%	7.65%
Mazda模型	72.04%	74.66%	98.57%	2.78%	25.34%	1.33%
Honda模型	70.49%	75.02%	90.44%	12.5%	24.98%	29.51%
Berkeley模型	70.67%	75.69%	89.25%	16.67%	24.32%	10.75%

评估结果表明，基于Transformer的自适应预警算法在预警任务中展现出显著的综合性能优势，其各项指标均优于经典预警模型，尤其在准确率、精确度与特异度等关键性能指标上表现突出。如表1所示，Transformer模型的总体准确率（ϕ_1=81.78%）显著高于次优的基于阈值优化模型（ϕ_1=74.31%）及其他对比模型（均值ϕ_1=71.45%），这一差距在统计学上具有显著意义（$p<0.01$）。在精确度方面，Transformer较次优模型提升15.78%，表明其具有更强的正类样本判别能力，有效降低了误警风险。值得注意的是，基于阈值优化模型虽然实现了极高召回率（ϕ_3=98.81%），但其特异度（ϕ_4=3.13%）显著偏离合理区间，导致误报率（ϕ_5=25.16%）较高。这种高敏感性与低特异性的矛盾特征揭示了该模型存在过拟

合风险，其预警机制可能过度依赖所采集的自然驾驶数据样本中的特定模式，在实际道路场景中易产生频繁的虚假预警，导致驾驶员产生适应性疲劳，反而降低安全效用。相较之下，Transformer模型在保持较高召回率（ϕ_3=84.23%）的同时，实现了最优特异度（ϕ_4=81.60%），其均衡性指标（F1_score=0.83）较基于阈值优化模型（F1_score=0.79）提升了5.06%，充分验证了注意力机制在驾驶行为数据时空特征提取中的优越性。

从系统鲁棒性角度分析，Transformer的误报率（φ_5=9.38%）与漏报率（φ_6=15.77%）均处于较低水平，较其他经典碰撞预警算法模型降低约43%和27%。这种特性源于其自注意力机制对复杂道路交互关系的深度建模能力，通过动态权重分配有效捕捉长程依赖关系，解决了普通神经网络模型（如基于TCN的自适应预警模型）因卷积核尺度限制导致的时序特征丢失问题。而基于阈值优化的自适应预警模型虽通过参数调优实现了次优的准确率，但其结构固有的归纳偏置限制了特征空间的探索能力，在应对突发驾驶场景时易产生系统性偏差。

综上所述，基于Transformer的自适应预警模型凭借其多头部注意力机制与位置编码技术，在碰撞预警任务中实现了预测精度与泛化能力的双重突破，其性能优势具有统计学显著性和工程应用价值。基于阈值优化的自适应预警模型作为次优方案，可作为特定场景下的补充方案，如作为极端工况下安全性能的底线保障，但在综合性能指标上与基于深度学习的方法仍存在系统性差距。

2）商业落地与行业认可：

a.系统已用于杭州、广州、南昌等在途重型货车驾驶员的行为监测，异常驾驶行为识别准确率提升至89.5%，节约近60%需人工进行处理的异常行为事件，节省保险成本超100万元/年；

b.获2024年深圳市创新应用大赛三等奖，入选《智能网联汽车创新技术目录》。

3)实时推理能力及模型轻量化：

a.模型轻量化：

采用知识蒸馏技术，将模型参数量从128M降84M，推理延迟从180ms优化至93ms（NVIDIA GeForce RTX 4070实测）。

b.边缘-云协同架构：本地设备处理实时预警信号（＜100ms），云端平台异步更新驾驶员风险画像（日均驾驶行为分析报告生成：包括疲劳、分心预警，超速行为等统计次数，以及月度驾驶员危险等级排行表）。

<table>
<tr><td colspan="6">（二）取得的业绩（代表作）【限填3项，须提交证明原件（包括发表的论文、出版的著作、专利证书、获奖证书、科技项目立项文件或合同、企业证明等）供核实，并提供复印件一份】</td></tr>
<tr><td colspan="6">1.
公开成果代表作【论文发表、专利成果、软件著作权、标准规范与行业工法制定、著作编写、科技成果获奖、学位论文等】</td></tr>
<tr><td>成果名称</td><td>成果类别
[含论文、授权专利（含发明专利申请）、软件著作权、标准、工法、著作、获奖、学位论文等]</td><td>发表时间/授权或申请时间等</td><td>刊物名称/专利授权或申请号等</td><td>本人排名/总人数</td><td>备注</td></tr>
<tr><td>Truck Drivers’ Cognitive Feature Learning with Multimodal Information Fusion</td><td>会议论文</td><td>2024年09月28日</td><td>2025 Transportation Research Board</td><td>1/6</td><td></td></tr>
<tr><td>驾驶行为异常识别：研究进展、方法与数据驱动实例</td><td>会议论文</td><td>2024年04月07日</td><td>2024世界交通运输大会</td><td>1/4</td><td></td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="6">2.其他代表作【主持或参与的课题研究项目、科技成果应用转化推广、企业技术难题解决方案、自主研发设计的产品或样机、技术报告、设计图纸、软课题研究报告、可行性研究报告、规划设计方案、施工或调试报告、工程实验、技术培训教材、推动行业发展中发挥的作用及取得的经济社会效益等】</td></tr>
</table>

（三）在校期间课程、专业实践训练及学位论文相关情况	
课程成绩情况	按课程学分核算的平均成绩： 87 分
专业实践训练时间及考核情况(具有三年及以上工作经历的不作要求)	累计时间： 1.1 年（要求1年及以上） 考核成绩： 84 分
本人承诺	
个人声明：本人上述所填资料均为真实有效，如有虚假，愿承担一切责任，特此声明！ **申报人签名：** 何[illegible]	

22260183

二、日常表现考核评价及申报材料审核公示结果

日常表现考核评价	非定向生由德育导师考核评价、定向生由所在工作单位考核评价： ☑优秀 □良好 □合格 □不合格 德育导师/定向生所在工作单位分管领导签字（公章）： 年 月 日
申报材料审核公示	根据评审条件，工程师学院已对申报人员进行材料审核（学位课程成绩、专业实践训练时间及考核、学位论文、代表作等情况），并将符合要求的申报材料在学院网站公示不少于5个工作日，具体公示结果如下： □通过 □不通过（具体原因： ） 工程师学院教学管理办公室审核签字（公章）： 年 月 日

中共浙江大学工程师学院委员会

浙 江 大 学 研 究 生 院

攻读硕士学位研究生成绩表

学号：22260183	姓名：何琦	性别：女	学院：工程师学院	专业：交通运输	学制：2.5年
毕业时最低应获：24.0学分		已获得：26.0学分		入学年月：2022-09	毕业年月：
学位证书号：		毕业证书号：			授予学位：

学习时间	课程名称	备注	学分	成绩	课程性质	学习时间	课程名称	备注	学分	成绩	课程性质
2022-2023学年秋季学期	工程技术创新前沿		1.5	78	专业学位课	2022-2023学年春季学期	自然辩证法概论		1.0	91	公共学位课
2022-2023学年秋冬学期	工程伦理		2.0	95	公共学位课	2022-2023学年春季学期	研究生英语基础技能		1.0	83	公共学位课
2022-2023学年秋冬学期	研究生论文写作指导		1.0	93	专业学位课	2022-2023学年夏季学期	交通运输工程科学与技术前沿		2.0	通过	跨专业课
2022-2023学年秋冬学期	数据分析的概率统计基础		3.0	81	专业选修课	2022-2023学年春夏学期	高阶工程认知实践		3.0	88	专业学位课
2022-2023学年冬季学期	新时代中国特色社会主义理论与实践		2.0	91	公共学位课	2022-2023学年夏季学期	智能交通系统与实践应用		2.0	85	专业学位课
2022-2023学年冬季学期	产业技术发展前沿		1.5	88	专业学位课	2022-2023学年夏季学期	工程师创新创业思维		2.0	96	专业选修课
2022-2023学年秋冬学期	研究生英语		2.0	84	公共学位课		硕士生读书报告		2.0	通过	

说明：1. 研究生课程按三种方法计分：百分制，两级制（通过、不通过），五级制（优、良、中、及格、不及格）。

2. 备注中“*”表示重修课程。

学院成绩校核章：

成绩校核人：张梦依

打印日期：2025-06-03

浙江大学研究生院（60）成绩校核章

收录邮件

TRB Paper - TRBAM-25-03387 - Review Results

Xiaobing Li 发送给 何琦

发起会议

2024-09-29 04:01:34

You are being carbon copied ("cc:'d") on an e-mail "To" "Jingqiu Guo" guojingqiu@hotmail.com
CC: mouyidbin.islam@gmail.com, xiaobing@uky.edu, Qi He" 22260183@zju.edu.cn, "Yibing Wang" wangyibing@zju.edu.cn, "Xuewen Yao" homan.yao@che-mi.com.cn

Ref.: Ms. No. TRBAM-25-03387
Truck Drivers' Cognitive Feature Learning with Multimodal Information Fusion
Transportation Research Board

Dear Qi He,

Your paper, number TRBAM-25-03387, "Truck Drivers' Cognitive Feature Learning with Multimodal Information Fusion", was peer reviewed by a TRB standing committee. Based on the review results, the committee is pleased to recommend your paper for presentation at the TRB Annual Meeting and for further assessment by the TRR Editorial Board. The review comments can be found below.

For publication review, the corresponding author will receive an email with additional guidance by October 31, 2024. If you do not receive this email, please reach out to trr@nas.edu for assistance.

For the Annual Meeting, TRB will notify you in early November regarding where and when your paper will be presented. Your paper will be posted in the TRB 2025 online program and made available to our sponsors and attendees. If you do not want the full paper posted, you will have the ability to upload a Short Format Research Summary version via MyTRB by November 20. Full papers and research summaries will not be indexed in the TRID database.

Thank you for submitting your paper for Presentation and Publication. Yours was one of more than 6,000 papers received this year. Approximately half of these papers will be presented at the Transportation Research Board 2025 Annual Meeting.

论文全文

Truck Drivers' Cognitive Feature Learning with Multimodal Information Fusion

Keywords: Truck Driving Behavior, Cognitive State Recognition, Naturalistic Driving, Multimodal Information Fusion, Transformer

Qi He
Polytechnic Institute
Zhejiang University, Hangzhou, China, and 310058
Email: heqi0921@zju.edu.cn

Yibing Wang
Professor
College of Civil Engineering and Architecture
Zhejiang University, Hangzhou, China, 310058
Email: wangyibing@zju.edu.cn

Qingyan (Ken), Yang
Senior System Engineer, AECOM
Novi, Michigan, USA 48377
Email: kyang@aecom.com

Xuewen Yao
Senior Research Fellow
Shenzhen Chemiyuntu Digital Technology Co., Ltd., Shenzhen, China,518063
Email: homan.yao@che-mi.com.cn

Jingqiu Guo, Corresponding Author
Associate Professor
College of Transportation Engineering
Tongji University, Shanghai, China, 201804
Email: guojingqiu@hotmail.com

Mark Stevenson
Professor
The University of Melbourne
Email: mark.stevenson@unimelb.edu.au

Words Count:7,536 words + 4 table(250 words per table)=8,536 words

Submitted [March,22,2025]

ABSTRACT

Real-time assessment of drivers' cognitive states is critical for improving road safety, especially in freight transport, where long-haul truck drivers frequently encounter prolonged fatigue and diverse traffic interactions. Current Advanced Driver Assistance Systems (ADAS) primarily cater to passenger vehicles, overlooking the distinctive challenges inherent to heavy trucks, such as delayed braking responses, extended sensor fusion latencies, and drivers' reliance on roadway familiarity. To address this gap, this paper proposes a Multimodal Attention Neural Network (MMANN) framework that integrates three asynchronous data streams: vehicle kinematics, driver facial states captured via low-frequency imaging, and driving environment videos. The proposed model utilizes interpretable attention mechanisms to fuse multimodal data, facilitating the classification of cognitive states into low-activity (characterized by distraction or fatigue), normal-activity, or high-activity (indicative of stress or aggressive driving). Trained on an extensive 180-day naturalistic dataset, MMANN achieves an impressive recognition accuracy of 82.4% - a notable improvement of 8.6% over single-modal baselines. This research pioneers the development of adaptive cognitive models, specifically tailored to the unique operational patterns and environmental constraints of truck driving.

Keywords: Truck Driving Behavior, Cognitive State Recognition, Naturalistic Driving, Multimodal Information Fusion, Transformer

INTRODUCTION

Advanced Driving Assistance Systems (ADAS), such as adaptive cruise control, forward collision warning (FCW), lane departure warning (LDW), speeding warning, and automated emergency braking, are designed to enhance safety by either alerting drivers to potential hazards or by taking over certain driving maneuvers *(1)*. These features are important for the successful deployment of autonomous vehicles in the future *(2)*. While ADAS can assist with driving tasks, until fully mature autonomous driving is achieved, drivers need to be cautious in case any of these systems fail to respond or disengage *(3,4)*. Therefore, real-time monitoring of driver cognitive states can significantly enhance driving safety through proactive risk identification, particularly for commercial vehicle operators and long-haul truck drivers engaged in prolonged transportation missions *(5)*.

The effectiveness of some ADAS, such as FCW and LDW systems, has been evaluated in various studies *(6,7)*. Some efforts have also been made to explore the key factors influencing drivers' acceptance of advanced features. Previous studies indicated that partially autonomous driving systems often result in passive fatigue, reduced vigilance and insufficient cognitive load, all of which negatively impact drivers' performance in critical situations *(8)*. Moreover, drivers' attention can be diverted when using conditionally automated driving systems, as they may engage in non-driving-related tasks, thereby reducing situational awareness *(9)*. A few studies used self-reported surveys and physiological indicators to investigate changes in drivers' cognitive status during partially and conditionally autonomous driving *(10,11)*. Prior studies utilized self-assessment tools such as NASA-TLX *(12)* for cognitive state evaluation and MMSE *(10)* for workload measurement. Some were conducted by providing drivers with a test vehicle to capture and evaluate driving behavior *(13)*. While these studies provide insights on driver interaction of advanced features, they are limited to selected scenarios, and drivers are susceptible to memory biases, which reduces their ability to accurately estimate their own cognitive state. In general, the behavioral differences between drivers operating vehicles equipped with ADAS versus those without ADAS have not been thoroughly investigated through naturalistic driving studies.

To avoid the limitations of the above subjective measurements, psychological, physiological, and behavioral datasets including eye movements, galvanic skin response, heart rate, head rotation, and vehicle states were widely employed *(14)*. Changes such as in pupil movement, facial features, and vehicle acceleration can be recorded by sensors and other sources to infer drivers' cognitive state *(15,16)*. Some studies also analyzed patterns of brain electrical activity measured using electroencephalography (EEG) to estimate drivers' levels of stress, attention, and alertness *(17)*. Each individual measure has its advantages and limitations. Drivers' cognitive state (e.g., attention, fatigue) can affect takeover performance *(18)*. Therefore, it is important to accurately identify drivers' cognitive state in partially autonomous vehicles and incorporate it into the design of the fallback procedure *(19)*.

Traditionally, vision, text, audio, and kinematic data have been examined within distinct domains *(20, 21)*. However, this fragmented analytical approach is insufficient for understanding the intricacies of driver behavior. Drivers engage with traffic through a multi-sensory modality, where diverse information streams are processed and interpreted by specialized brain regions to form a complex yet cohesive sensory loop *(5)*. This biologically-driven sensory integration requires the development of computational frameworks capable of emulating the brain's adaptive fusion hierarchy.

According to the relevant literature review, several critical limitations emerge: 1) Previous studies have predominantly relied on physiological data, e.g., EEG, eye tracking, interface pressure, and heart rate. However, collecting such data in naturalistic driving scenarios is both invasive and technically challenging. 2) The naturalistic driving datasets currently available suffer from a small participant pool, which may lead to a statistical over-representation problem. 3) Single-modal data architectures lack the capacity to detect drivers' complex cognitive-load state with sufficient accuracy and timeliness. Furthermore, current ADAS research predominantly focuses on standardized scenarios derived from passenger vehicle operations, overlooking the unique challenges inherent to commercial truck driving. The prolonged operational cycles of freight transportation induce cumulative cognitive fatigue that existing passenger-oriented models fail to quantify. Compounded by the complex interactions in mixed traffic environments (e.g., vulnerable road user detection during urban freight distribution), these factors necessitate the development of specialized cognitive state assessment frameworks.

To address the research gap in commercial vehicle safety, this study focuses on exploring multimodal driving data feature extraction, fusion, and cognitive state recognition. Initially, the study conceptualizes driver activity as a complex interplay of various cognitive structures, each influenced by factors related to the driver's state (e.g., maintaining situational awareness during autonomous driving) and driving-related factors (e.g., consuming cognitive resources to perform driving tasks) *(22,23)*. Building on this conceptual foundation, the framework systematically integrates three asynchronous data streams of vehicle dynamics, driving environment information, and driver facial features obtained from non-invasive devices. Meanwhile, the multimodal data fusion method is used to estimate the driver’s cognitive profile, which was often estimated by physiological and/or psychological indicators in previous studies. Moreover, this study explores naturalistic driving data of heavy trucks when driving with FCW and overspeed warning in urban, rural, and freeway scenarios under varying traffic flow, lighting and weather conditions. Thereby, the proposed framework bridges the knowledge gap between passenger-centric ADAS research and the safety demands of commercial freight transport.

DATASET

The naturalistic driving dataset utilized in this study specifically focuses on heavy trucks equipped with Advanced Driver Assistance Systems (ADAS) operating in Guangzhou Province. During the data acquisition phase, comprehensive vehicular operational data, including geolocation tracking information, real-time vehicular operational metrics, safety alert logs, forward-facing visual recordings, and driver-facing camera feeds, were systematically transmitted to the cloud infrastructure.

The ADAS system in these vehicles incorporates multiple safety features, including Forward Collision Warning (FCW), lane-departure warning, and real-time monitoring of the driver's facial status. The facial monitoring system provides alerts for fatigue, distraction, and abnormal behaviors such as smoking or phone usage. Given the heterogeneous nature of multimodal driving data, FCW events were selected as the primary observational samples in this study due to their unique integration of three critical components: driver states, vehicle operational parameters, and roadway situational contexts.

The cloud platform records alarm events with a set of data streams, including the alarm time (T_W), vehicle driving status data, environmental data, the driver's facial image at the time of alarm, and vehicle front video (as illustrated in FIGURE 1). The data streams were recorded at varying frequencies: 1fps for facial images captured via infrared cameras, 10Hz for vehicle dynamics through radar signals, and 10fps for driving environment video. Each modal dataset encompasses driving data from 5 seconds before the warning to 5 seconds after the warning initiation time, providing a comprehensive temporal context for analysis.

FIGURE 1 Multimodal data sample. (SV: subject vehicle ; FV: front vehicle)

9,453 original warning data segments from 569 heavy trucks were collected over a 180-day period from January 1 to June 30, 2023. TABLE 1 shows a set of sample data. After data preprocessing, 3,519 valid FCW events were extracted (μ = 6.2 instances per vehicle). Considering the potential impact of driver states (e.g., fatigue, distraction, stress), this study conducted a comparative analysis of the trigger time between FCW and other warnings. Based on this analysis, a preliminary classification framework was established, comprising three distinct categories of FCW warning:

1) Distraction group: system detected ≥ 2 fatigue or distraction warnings within 300 seconds prior to FCW activation.

2) Normal group: no significant fatigue, distraction, or abrupt acceleration/deceleration warnings were detected before or after FCW activation, with driver behavior parameters remaining within normal operating thresholds.

3) Driver stress group: temporal overlap ($\Delta t \leq 3s$) between FCW activation and abrupt acceleration/deceleration warnings, indicating concurrent aggressive driving behaviors during FCW events.

TABLE 1 Main characteristics of the dataset

Time Index	Driver ID	Warning Label	Speed(km/h)	Relative Distance(m)	Front Vehicle Speed(km/h)	Heading(°)
1	77	0	66.4	11.7	67.2	112
2	77	0	66.6	11.3	65.8	113
3	77	0	66.8	10.9	65.4	112
···	···	···	···	···	···	···
51	77	1	63.8	7.8	61.6	122
52	77	1	62.4	7.4	60.3	124
53	77	1	61.7	6.5	59.8	125
···	···	···	···	···	···	···

Reaction time was calculated through comparative analysis of the acceleration and relative velocity (speed differential between subject truck and leading vehicle) curves *(24)*. **FIGURE 2** shows a sample of the curves of subject truck acceleration and relative velocity over time. For each extremum point (local minima/maxima) of the relative velocity R_1, there is a corresponding extremum point (local minima/maxima) of the acceleration R_2. The temporal peaks correspond to warning initiation and driver response, and the time between them is the reaction time T_r.

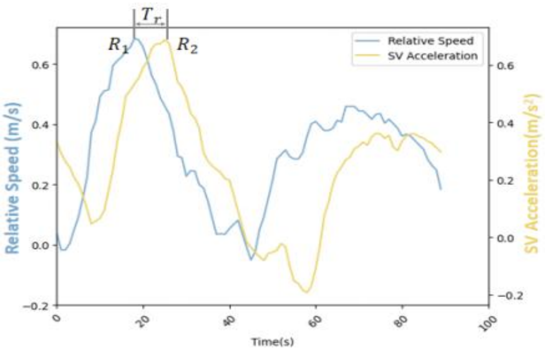


FIGURE 2 Reaction time determination

METHODOLOGY

In this work, a transformer-based framework 'Multimodal Attention Neural Network (MMANN)' is proposed, which consists of a feature extraction module designed to capture driving behavior, drivers' states and environment, along with a fusion module to integrate multimodal data. This section begins by outlining the problem definition and modeling approach.

A. Problem Definition

Driver activity was applied to characterize the complex cognitive structures of drivers during FCW events. A primary objective of driver activity estimation is to assess whether the driver can respond promptly and execute appropriate risk avoidance measures when encountering hazardous situations. Consequently, the driver activity recognition task is framed as a time series classification problem, with the driver activity level serving as the designated label: $\hat{y}_j \in \{0,1,2\}$,where 0 = low-activity level, 1 = normal-activity level, 2 = high-activity level, given the multimodal temporal sequences X^l of size l.As this classification, low-activity means fatigue driving or distraction driving state, for example, the driver's attention resources devoted to driving tasks are reduced due to fatigue or non-driving related tasks; normal-activity means normal driving state, at this time, the driver's attention is concentrated and the environmental pressure is appropriate; high-activity means aggressive driving or stressful driving state, for example, drivers of ' road rage ' tend to ignore some environmental information to overdrive. Therefore, the problem is to generate the corresponding driver activity level $\hat{y}_j$ based on temporal series information as

$$\hat{y}_j = arg \max_{\hat{y}_j \in \{0,1,2\}} p(y_j|X_j^l) \quad ,j = 1,2,\ldots,N \tag{1}$$

$$X^l = [\ x_1, x_2, \ldots, x_l\ ] \tag{2}$$

where

j is the j^{th} sample,

N is sample size,

x represents the feature at each time step.

Multimodal temporal information consists of front camera video M^l and driver's facial images G^l obtained by camera sensor, vehicle kinematics K^l obtained by on-board sensors, denoted by $X^l = [M^l, G^l, K^l]$, and $x = [m, g, k]$,respectively. Specifically, $k = [v, a, \Delta s, \Delta v]$ represents the velocities/accelerations of the truck, and the relative distance/velocity with respect to the front one, respectively.

B. Model Construction

The MMANN model, as shown in **FIGURE 3**, presents a multimodal fusion architecture. First, MMANN passes each modal data to its dedicated feature extractors: 1) Multi-channel Video Vision Transformer (MCViViT) model for videos, 2) Convolutional Neural Network-Vision Transformer (CNN-ViT) model for images, and 3) an Attention-BiGRU (AT-BiGRU) for driving kinematics. Then, the features extracted from the three models were fused using One-to-Two (OvT) attention layers. At the same time, this layer learns attention weights to decide the importance ratio for each modality. Last, the Multilayer Perceptron (MLP) outputs a multi-classification result of driver's activity level.

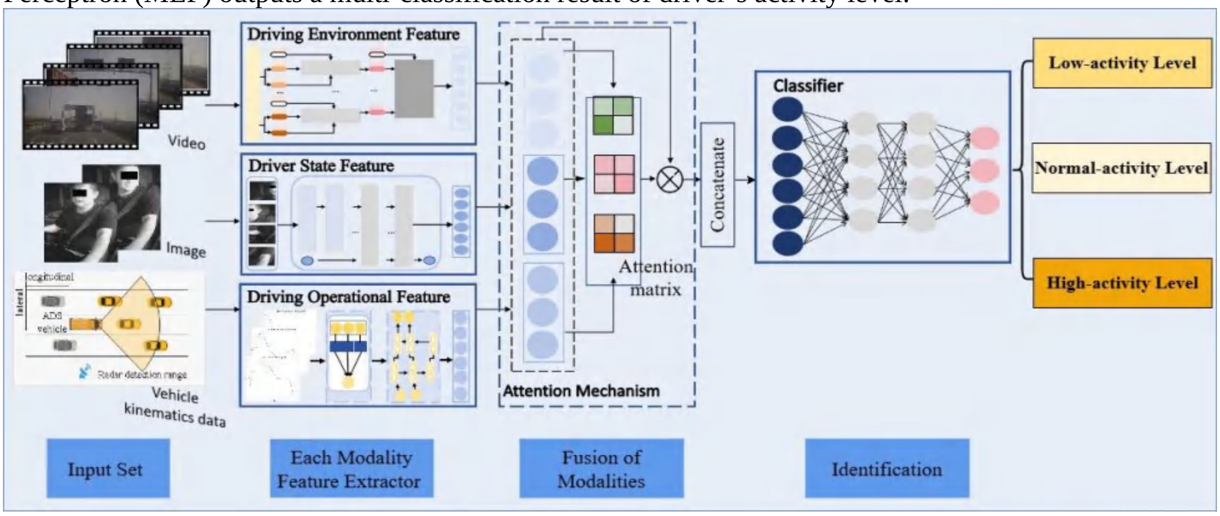


FIGURE 3 Overview of the MMANN model

Driver State Feature Extraction

In recent years, Transformer model has become the first choice for natural language processing (NLP) tasks, especially working together with the self-attention mechanism. Vision Transformer (ViT) applies to image processing, which involves dividing image into blocks, and then processing these linearly arranged block sequences as the input. In this context, the concept of image blocks is similar to the concept of tokens in NLP tasks. **FIGURE 4** illustrates the structure of an improved ViT module for feature extraction of multi-channel images of driver's state.

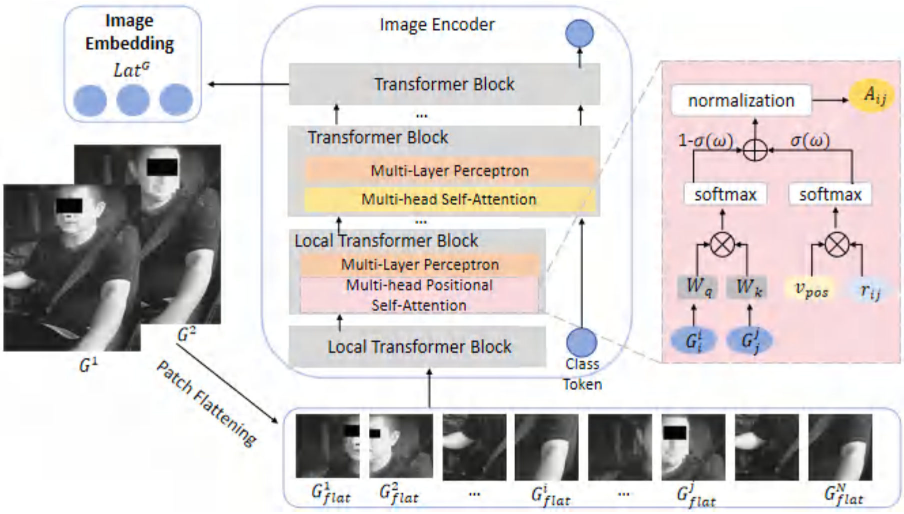


FIGURE 4 Workflow of the CNN-ViT model (W_q: query weights; W_k: key weights; v_{pos}: attention center and span embeddings; r_{ij}: relative position encodings; σ:sigmoid function; ω: gating parameter)

In the CNN-ViT Model, given an input image denoted as $G\epsilon R^{H*W*C}$, it is reshaped into a sequence of flattened patches, denoted as $G_{flat}\epsilon R^{N*P^2*C}$, where H, W, C and P represent the height, width, number of channels, and the size of the image block respectively, and $N = \frac{H*W}{P^2}$ represents the number of image blocks. Subsequently, each block is linearly embedded into a low-dimensional feature space, and an additional position embedding is added to the patch embedding. The embedding process can be expressed as follows:

$$E_0 = L_{pos} + \left[G_{flat}^1 L; G_{flat}^2 L; \dots ; G_{flat}^i L; \dots ; G_{flat}^N L\right] \tag{3}$$

where

G_{flat}^i is the i^{th} patch,

L is the linear embedding function,

L_{pos} is the position embedding function,

E_0 is the output of the embedding operations.

Then E_0 is transmitted to Local Transformer (LT) layer and the extracted local features E_l are sent to the Standard Transformer (ST) layer, the extra class embedding is added, denoted as $E_{local} = [G_{class}; E_l]$, and E_{class} represents embedding class label. Driver state feature representation Lat^G is generated through ST layer.

LT and ST contain a sequential array of Transformer blocks, and the total number are D_l and D_s, respectively. Each ST block contains a Multi-head Self-Attention (MSA) block, followed by a MLP block:

$$E'_d = MSA\big(LN(O_{d-1})\big) + E_{d-1} \;, l = 1,2, \dots , D \tag{4}$$

$$E_d = MLP\big(LN(O'_d)\big) + E'_d \;, l = 1,2, \dots , D \tag{5}$$

where

LN represents the normalization function,
E_d and E'_d correspond to the output of MLP and MSA blocks, respectively.

In the MSA layer, each self-attention head h performs the following operations:

$$MSA_h(G) = A^h G W^h \tag{6}$$

where
W^h is the value matrix,
A^h is attention matrix.

In LT block the MSA with positional self-attention (PSA) is used to replace the self-attention layer to achieve local perception, that is, A^h is replaced by A_{IJ}^h, and could be defined as:

$$A_{ij}^h = softmax(Q_i^h K_j^{h^T} + V_{pos}^{h}{}^T r_{ij}) \tag{7}$$

Each attention head h uses a trainable embedding $V_{pos}^h \in R^{D_{pos}}$, and the relative position encoding $r_{ij} \in R^{D_{pos}}$ represents the relative position between the patch i and j. Its size depends only on the distance between the pixels i and j . When the PSA layer contains h attention heads and the relative position coding dimension $D_{pos} \geq 3$, the PSA could be regarded as a convolutional layer with a convolution kernel size of $\sqrt{h} \times \sqrt{h}$, and the network pays more attention to local features.

Driving Environment Feature Extraction

The input video generates a large number of spatio-temporal tokens, and the context of the long-range annotation sequence and the efficiency of the model must be considered. Based on ViT, this paper expounds the structure of an improved multi-channel ViViT module, which is used to extract driving environment features. The tubelet-embedding method is used to map the video M into a sequence of tokens $\tilde{m}$. Given a video input, denoted by $M \in R^{T*H*W*C}$, defining $t \in R^{t*h*w}$ of the tube t , then $n_t * n_h * n_w$ sequences $\tilde{m}$ can be obtained, $\tilde{m} \in R^{n_t*n_h*n_w*C}$, $n_t = \frac{T}{t}$, $n_w = \frac{W}{w}$, $n_t = \frac{T}{t}$. Then, position embedding and class embedding are added to each sequence token $\tilde{m}$, and the input tokens $m \in R^{N*d}$ to the following Transformer is acquired. Compared with the uniform sampling method in ViT, sequence constructed in this way can better fuse spatio-temporal information.

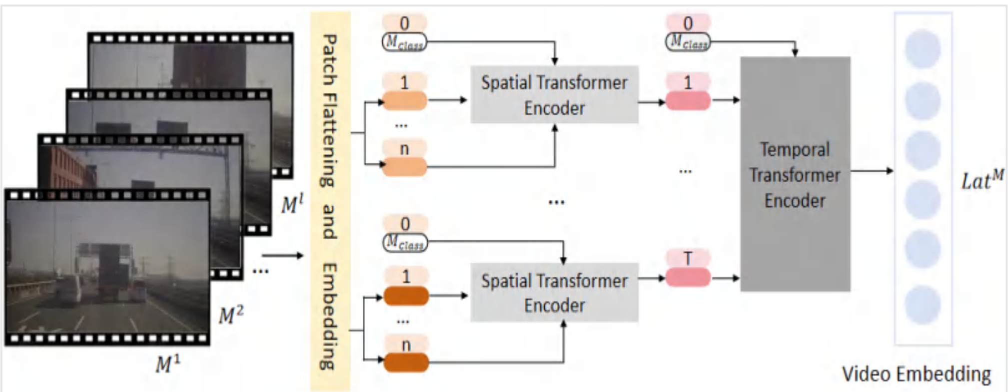


FIGURE 5 Workflow of MCViViT model

FIGURE 5 presents Transformer-based spatial encoder and temporal encoder structures. First, the spatial encoder models the interaction between tokens extracted by the same time index, a representation $h_i \in R^d$ for each temporal index is obtained after L_s layers. Then h_i are concatenated into $H \in R^{n_t*d}$, and forwarded through the temporal encoder consisting of L_t layers to aggregate the representations from different temporal indices, and corresponds to the late fusion of spatial-temporal information. The output token of the temporal encoder is the feature representation of the driving environment feature extraction Lat^M.

Driving Operational Feature Extraction

Vehicle dynamics data is often constructed as multivariate time sequences that contain various driving behavior characteristics for optimal utility. RNNs have achieved great success in speech recognition and language translation. To enhance the learning ability of sequence data, researchers proposed two effective variants: LSTM and GRU. Considering the characteristics of driving operation data, this study employs BiGRU to capture the two-way dependence of time sequences, and uses the attention layer to highlight important time indicators. The internal information transmission and network structure are shown in **FIGURE 6**.

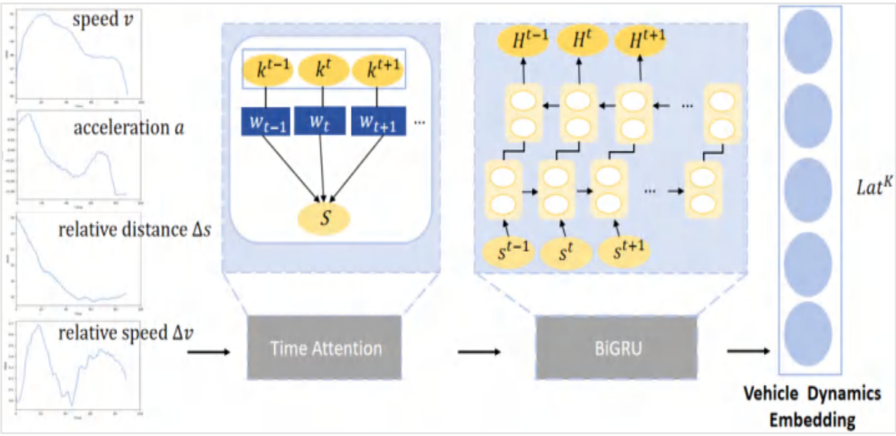


FIGURE 6 Workflow of AT-BiGRU model

Driving operational data is the input of the attention layer and assigns weights to the input of the BiGRU layer. The greater the weight, the more important the data at that time step, which means 'attention' is focused here, and will affect the final result of the entire problem. The attention calculation formula is:

$$\begin{cases} w_t = softmax\ (k_t h_t + b_t) \\ S = \sum_{t=1}^{T_l} w_t\ h_t \end{cases} \tag{8}$$

where

h_t is the hidden layer output of BiGRU network at a certain time,

b and k are bias and weight respectively,

T_l is the total length of driving behavior sequence,

S is the final representation vector of attention layer.

Bi-GRU obtain feature vectors from both back and forward directions. GRU calculates the feature vector transmitted by attention layer and outputs a vector with a fixed dimension. Inside the GRU unit, there are four parts of calculation:

$$\begin{cases} r_t = \sigma\left(W_r * \left[h_{t-1},\ \ s_t\right] + b_r\right) \\ z_t = \sigma\left(W_z * \left[h_{t-1},\ \ s_t\right] + b_z\right) \\ \tilde{h}_t = tanh\ (W_{\tilde{h}} * [r_t * h_{t-1} s c_t] + b_{\tilde{h}}) \\ h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \end{cases} \tag{9}$$

where

σ is the sigmoid function,

$tanh$ is the hyperbolic tangent function,

s_t is the input vector of the GRU layer,

h_t is the output vector of the GRU unit,

r_t is the reset gate state vector,

z_t is the update gate state vector,

W and b correspond to the weight vector and the bias vector respectively.

Since the input is the whole time series, the characteristics of a certain moment depend on both the previous information and the future information at that moment. The calculation formula is as follows:

$$Lat_t^K = [h_t^1, h_t^2] \tag{10}$$

where

Lat_t^K is the output of BiGRU layer at time t,

h_t^1 is the hidden state of forward propagation,

h_t^2 is the hidden state of backward propagation.

Fusion of Modalities

Many multimodal models rely on self-attention or cross-attention for effective data integration, but do not scale well for applications with more modalities. Hence, an improved OvT attention mechanism is utilized *(25)* to understand the relative importance of each modal (in **FIGURE 7**).

Given an embedding mode m_i obtained from a dedicated encoder, $m_i \epsilon Lat^M, Lat^G, Lat^K$, calculate the similarity between this mode and the combined representation of all other modes. The calculation formula is as follows:

$$score(m_i) = m_i^T W \frac{\sum_{i \neq j}^{k} m_j}{k-1} \tag{11}$$

$$m_{att_i} = softmax(score(m_i) * m_i) \tag{12}$$

where

m_{att_i} is the feature representation of m_i after fusing other modal information,

W is weight matrix can be learned,

k is number of modes, and $i \in 1,2, \ldots, k$.

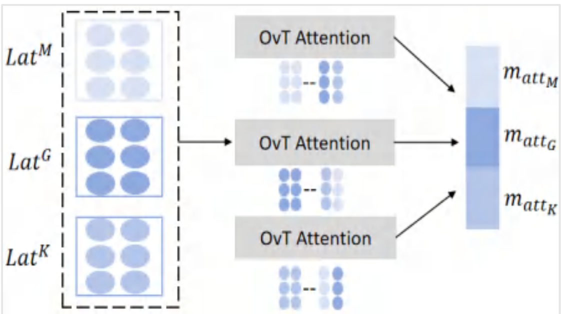


FIGURE 7 Workflow of multimodal fusion model

Output Classifier Layer

As in FIGURE 3, the multi-modal fusion feature representation m_{att_i} is input into MLP, and confidence value of each class is output to form final prediction $\hat{y}_t = 0$: low-activity level of the truck driver, $\hat{y}_t = 1$: normal-activity level, $\hat{y}_t = 2$: high-activity level. Then, the confidence $\hat{Y}$ of the batch training prediction is compared with the true label to calculate the loss and back propagation. Once trained the model predicts classes based on the highest confidence score.

C. Model Training and Evaluation

Training was performed on a Nvidia RTX 3070 GPU. The dataset was divided using stratified random sampling implemented via Python's scikit-learn library, and class distribution was preserved through the 'stratify' parameter. Through data shuffling and two-stage splitting, samples from each driver activity level were allocated to training (70%), validation (10%), and test (20%) sets. Adam optimizer was selected for training with a maximum of 150 epochs and a batch size of 24, a learning rate (lr) of 1×10^{-5} and a learning rate scheduler named CosineAnnealingLR. CrossEntropy, an advanced loss function for the multi-classification problem, was also applied with label smoothing *(26,27)* as loss metric. Label smoothing can reduce the harm caused by overfitting and overconfidence in model training, i.e., alleviates the troubles caused by wrong labels on model learning. The formula of Cross-Entropy is as follows:

$$\mathcal{L}_C = -\sum y_i \log(\mathrm{softmax}(x_i)) \tag{13}$$

where

x is the output of the model,

y is the true label with the form of one-hot.

When using Label Smoothing, the formula can be changed as follow:

$$\mathcal{L}_{LSC} = -\sum q_i \log(\mathrm{softmax}(x_i)) \tag{14}$$

where

$$q_i = \begin{cases} 1-\omega\ , & y_i = 1 \\ \frac{1}{N-1}\omega & , otherwise \end{cases},$$

N is number of classes,

ω is hyperparameter and tends to be a small numeral.

In this model, $N = 3$, $\omega = 0.1$.

For operation feature extraction, each of one attention layer and two BiGRU layers has a hidden dimension of 32. All features are uniformly resampled to 10 Hz and normalized to the same scale. For driver state feature extraction, the two raw images of each driver are unified into 256×256 to avoid data misalignment during model training. For driving environment feature extraction, the videos are sampled at 10fps. The hyperparameters selected for the CNN-ViT model and the MCViViT model are as follows: 1) CNN-ViT: the number of patches is 16, the last dimension of the output tensor after the CNN-VIT module is 64, and the number of LT and ST blocks is 2 and 2, respectively. 2) MCViViT: the last dimension of the output tensor after the ViViT module is 128, the number of ViViT Transformer blocks is 4, and 3) the number of heads in the MSA layer is 3.

To evaluate the proposed model's performance, metrics such as accuracy, precision, recall, and F1 score were used. Each metric can be calculated by the following formula. Among them, TP, TN, FP and FN refer to the number of true positives, true negatives, false positives and false negatives, respectively. For multi-classification, macro-averaging rules are required for Precision, Recall, and F1 calculations.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{15}$$

$$Precision = \frac{TP}{TP+FP} \tag{16}$$

$$Recall = \frac{TP}{TP+FN} \tag{17}$$

$$F1 = \frac{2\times Precision \times Recall}{Precision + Recall} \tag{18}$$

RESULTS

In this section, the driving operational feature, driver's state features and driving environment feature extractors of FCW events in trucks are evaluated. Different evaluation metrics, inference time, and ablation study are employed to analyze the model performance.

A. Statistical Analysis

A total of 3,519 FCW records were screened over a 180-day period. Several variables such as road type, front vehicle type and traffic flow density were applied. The classification of traffic flow states is primarily utilized based on the quantity and velocity of preceding vehicles detected by the forward-mounted radar in naturalistic driving vehicles *(28)*. Specifically, the low-density is defined as the number of preceding vehicles fewer than 3 and the speed greater than 60km/h; the medium-density is that the number of preceding vehicles is 3-5 and the speed is 30 km/h-60 km/h; the high-density is that the number of preceding vehicles is more than 5 and the speed is less than 30 km/h.

As shown in **TABLE 2**, the selected heavy trucks are mostly driving on highways or freeways, where drivers are more likely to show distraction, and on urban or rural roads, drivers are more likely to drive aggressively. The alert time of the distracted group mostly occurs in the early morning or night (after 22:00 hours, the group exhibited the highest standard deviation (SD)). Most of the vehicles in front of the distracted group are also trucks. At this point, the driver's front vision may be obscured, so the driver remains conservative while driving. When the vehicle in front of the driver is a passenger car, less-obstructed vision can induce drivers' reliance on his/her driving habits and ignore warning signals. The absolute values of yaw angle and SD in the distracted group are the smallest, indicating that the driver's operation of the steering wheel is weakened at this time, and it is easy to ignore the warning signals at this time, while the driver's stress group may maneuver more frequently to try to change lanes, resulting in a larger steering wheel angle. Based on the above statistics, the following conclusion could be drawn: 1) Drivers are more likely to be distracted when traffic density is low and vehicle speed is high. Drivers often avoid frequent acceleration and deceleration maneuvers by maintaining a longer distance from the vehicle in front, and the driver has a longer reaction time. 2) When driving on urban or rural roads with complex traffic conditions, drivers react more quickly and adjust their speed more frequently.

TABLE 2 Statistics of FCW record data characteristics

Variable		**Distraction**		**Normal**		**Driver stress**	
		Avg[1]	**SD**[2]	**Avg**	**SD**	**Avg**	**SD**
Warning time(0-23h)		8.50	6.77	13.69	4.59	13.86	5.36
Reaction time(s)		2.14	0.49	1.16	0.74	0.86	0.37
Speed(km/h)		57.34	14.25	54.27	14.76	51.65	12.73
Acceleration(m/s^2)		-0.11	0.18	-0.13	0.23	-0.21	0.16
Yaw angle(rad/s)		-0.20	1.98	0.06	2.28	-0.25	2.52
Relative distance(m)		11.63	4.50	9.64	5.86	8.77	4.35
Relative speed(m/s)		0.34	0.33	0.38	0.46	0.56	0.26
Variable	**Type**	**Count**	**Pro**[3] **(%)**	**Count**	**Pro (%)**	**Count**	**Pro (%)**
Road type	1(Highway)	312	52.0%	915	46.9%	423	43.8%
	2(Urban road)	207	34.5%	585	29.9%	348	36.0%
	3(Rural road)	81	13.5%	453	23.2%	195	20.2%
Front vehicle type	1(Passenger car)	206	34.3%	955	48.9%	570	59.0%
	2(Truck)	394	65.7%	998	51.1%	396	41.0%
Traffic density	1(Low)	321	53.5%	795	40.7%	165	17.1%
	2(Middle)	213	35.5%	792	40.6%	348	36.0%
	3(High)	66	11.0%	366	18.7%	453	46.9%

[1]Avg: average; [2]SD: standard deviation ;[3]Pro: proportion

B. Driver activity level

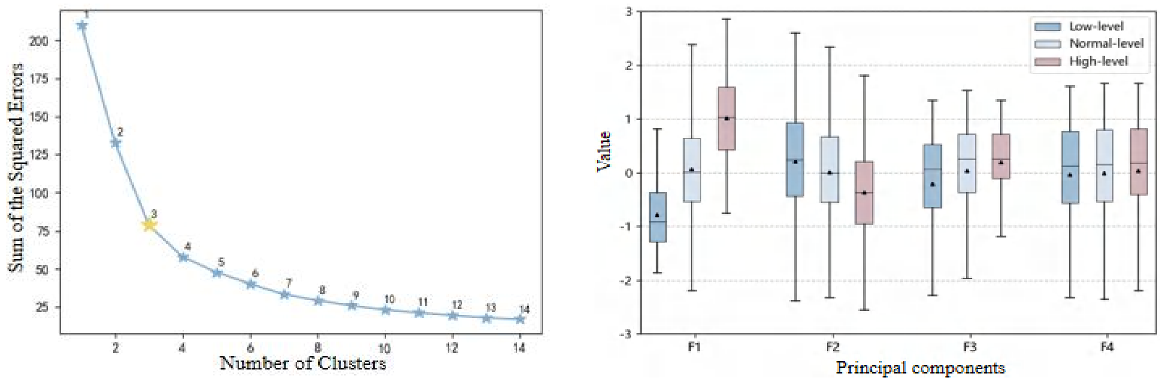
In the preliminary analysis, eight variables (e.g., reaction time, average speed/acceleration) were associated with the driving activity level. Principal component analysis (PCA) was used to extract features from highly correlated variables, and k-means++ was used to cluster the extracted principal components. First, KMO test and Bartlett sphere test were carried out to prove that conditions of PCA. As shown in **TABLE 3**, the first four principal components explain 83.4 % of the total variance, indicating that these components can represent most of the original data information and play a role in dimensionality reduction. Therefore, four principal components $Com_{i \in 1,2,3,4}$ were extracted as F_1, F_2, F_3, F_4, respectively. The coefficients of each index can be solved according to the principal component score and the component matrix coefficient, and the principal component coefficient matrix was obtained. From the perspective of the coefficient values, F_1 is mainly positively correlated with the average/maximum acceleration. F_2 is positively correlated with spacing, negatively correlated with yaw angle. F_3 is positively correlated with the road type and reaction time, and negatively with the vehicle speed. F_4 is positively correlated with traffic density and acceleration. In order to make a more accurate assessment of driving activity, F_1, F_2, F_3, F_4 obtained by the above PCA are selected as indicators, and K-means++ is utilized for clustering. The elbow method helps to determine the optimal number of clusters (**FIGURE 8(a)**).

Based on the elbow curve, driver activity could be divided into three levels: low-activity, normal-activity, and high-activity. As shown in the statistical box plot of each class (**FIGURE 8(b)**), F_1 and F_3 value distributions in cluster1 are the lowest, however, mean of F_2 is highest with this cluster. Drivers tend to maintain a long front gap to prevent sudden violent deceleration, and driver's control behavior of the steering wheel is weakened, indicating that the driver is usually in a driving environment with fewer surrounding vehicles, and has a longer reaction time at this time, so this cluster1 could be seen as the 'low-activity group'. Similarly, cluster2 and cluster3 can be regarded as 'normal-activity level' and 'high-activity level', respectively. The statistical values in the box plot provide a categorical indication that the actual velocity, acceleration, and other values in the warning segment will be dynamic within and between classes. On the basis of identifying driver activity labels, the proposed MMANN model can be trained to achieve efficient recognition of different driver activity levels through multi-modal input features.

TABLE 3 Driving activity variance interpretation and component matrix

Component	Initial eigenvalue			Principal component score				Primitive variable
	Eig[1]	Pro[2]	Cum[3]	F1	F2	F3	F4	
Com1	2.283	0.295	0.295	-0.171	0.140	0.504	0.147	Road type
Com2	1.806	0.226	0.521	0.276	0.529	0.143	0.115	Relative distance
Com3	1.429	0.177	0.698	0.244	-0.382	0.137	0.104	Mean yaw angle
Com4	1.066	0.136	0.834	0.671	-0.279	-0.084	0.334	Mean acceleration
Com5	0.502	0.059	0.893	0.787	-0.348	-0.035	0.411	Max acceleration
Com6	0.408	0.048	0.941	0.241	0.098	-0.458	-0.156	Mean speed
Com7	0.334	0.040	0.981	-0.224	0.175	-0.217	0.607	Traffic density
Com8	0.155	0.019	1	0.211	-0.096	0.281	-0.488	Reaction time

[1]Eigenvalue ;[2]Proportion;[3]Cumulation



(a) (b)

FIGURE 8 Driver activity level classification (a) elbow rule (b) classification results (values are normalized).

C. Model Performance

FIGURE 9 depicts the loss and accuracy curves for both the training and validation sets across epochs. The model's performance on the training set continues to improve, while the validation set performance plateaus after approximately 60 epochs. To mitigate the risk of overfitting and enhance the model's generalization capability, an early stopping mechanism is implemented. This mechanism halts training when the validation error increases consecutively for 10 epochs. **FIGURE 10** presents a confusion matrix based on the test set, which provides insights into the model's predictive accuracy for different activity levels. The precision scores for low, high, and normal activity levels are 0.867, 0.838, and 0.796, respectively. The model excels in differentiating between the distracted state and the stress-driven state. Consequently, the model can effectively identify the abnormal driving states of drivers.

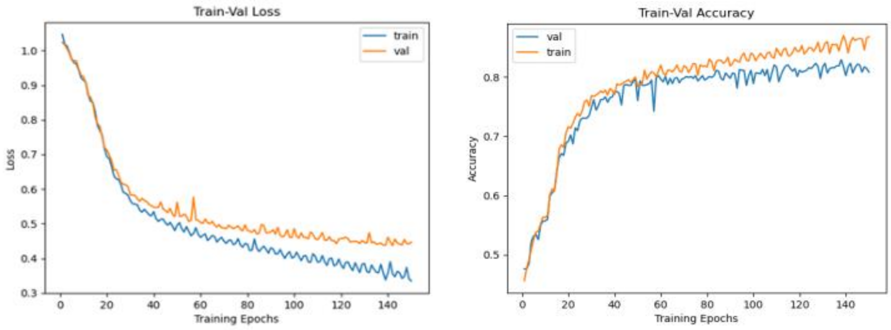


FIGURE 9 MMANN model training and validation loss and accuracy variation over epochs

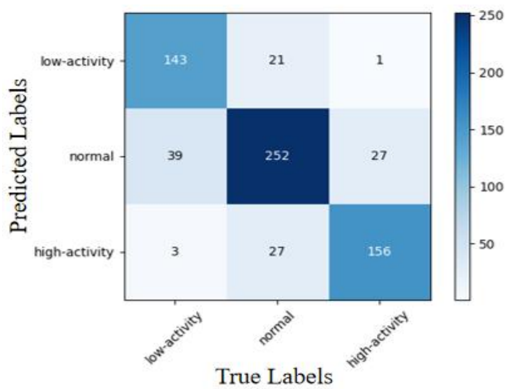


FIGURE 10 Confusion matrix of MMANN model on testing set

In real-time applications, computer vision systems in ADAS applications require at least 10 FPS (frames per second) *(29)*. To evaluate performance, the timing of each scene was measured, simulating the continuous processing of the real world on an extended sequence without interruption to reduce inaccuracies and overhead due to timing functions and data transfers. The average time per frame and FPS are as shown in **FIGURE 11**. In terms of scalability, these models are designed to process video clips of the driving environment 5 seconds before a warning is issued to perform driver activity estimates. In deployment, the model will use a fixed window corresponding to the length of the test video to ensure real-time performance.

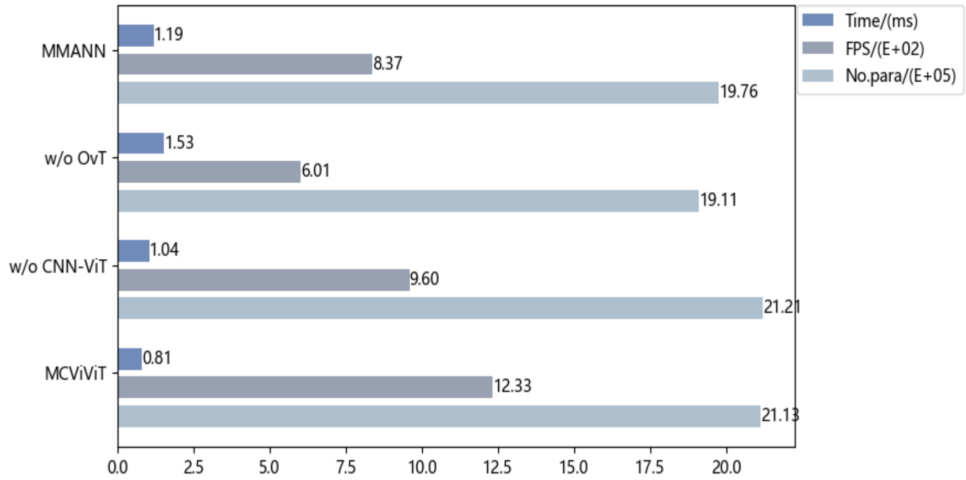


FIGURE 11 Model real-time operation. (Time: average time to process one frame; MMNN: model without OvT attention layer; w/o: without.)

D. Model Comparison

A group of learning-based models used for sequence data such as DTW-KNN *(30)*, LSTM *(31)*, TCN *(32)*, and FCN *(33)*, is employed for comparison. The driving behavior matrix, which consists of the speed, acceleration and spacing of the main vehicle in the 5s before warning, is used as the input of each model. **FIGURE 12** shows a performance comparison on various single and multimodal models in terms of Accuracy, Precision, Recall and F1 Score.

The multimodal model demonstrates superior performance compared to single-modal models. The results indicate that the proposed MMANN model surpasses other baseline models across all evaluation metrics. The driver's facial image effectively captures the distribution of the driver's synchronous attention, while the driving environment video indirectly reflects the driver's cognitive load at any given moment. The MMANN model validates the effectiveness and feasibility of multimodal data fusion for identifying driver activities.

DTW-KNN exhibits the least performance, highlighting the advantages of machine learning in handling complex and high-dimensional time series data. Among the models, FCN and LSTM-FCN yield promising results, indicating that convolutional networks have potential in time series data analysis and modeling. This underscores the capability of the MMANN model to leverage multimodal inputs for robust driver activity recognition.

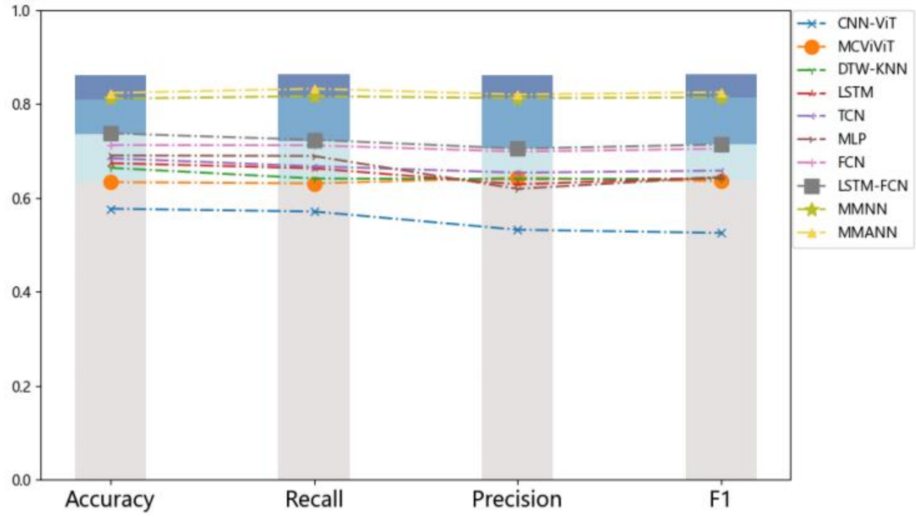


FIGURE 12 Performance of multimodal model and benchmark models.

E. Ablation Study

An ablation study involves removing certain components of the network to better understand its behavior and causality. In this experiment, vehicle dynamics were taken as the baseline data, and the driver's state feature extraction module, driving environment feature extraction module and multimodal attention module were tested in ablation to determine the significance of specific network components. Without the inter-modal OvT Attention module, the feature representations extracted from these three modalities are directly concatenated and fed into the MLP for classification. This implies that the importance of the features extracted from each modality is considered uniform (the impact on driver state recognition). The results for these 7 models are shown in **TABLE 4**.

Overall, the MMANN model outperforms the six ablation models in terms of accuracy, precision, recall, and F1 score, confirming the necessity of multi-source data fusion. The results indicate that the driving environment and the driver's state characteristics are effective supplement to the driver's activity recognition. In a single-mode, the driver's operational data excels in recognizing the driver's activities, while the recognition reflected in video features and images of the driver's face is often less optimal. Multimodal fusion significantly improves the accuracy of driver activity recognition.

Table 4 Results of Ablation Study

Model	Accuracy	Recall	Precision	F1-score
MMANN	**0.8236**	**0.8326**	**0.8203**	**0.8253**
w/o OvT Attention	0.8117	0.8166	0.8127	0.8144
w/o ConViT	0.7957	0.7925	0.7928	0.7926
w/o MCViViT	0.7787	0.7796	0.7896	0.7836
AT-BiGRU	0.7375	0.7238	0.7051	0.7138
MCViViT	0.6336	0.6309	0.6430	0.6357
ConViT	0.5769	0.5710	0.5322	0.5253

(w/o: without)

Given that driver activity is a complex physiological phenomenon, constructing a robust driver activity recognition model using unimodal data is challenging. Facial images mainly reflect the general level of concentration during the driving process, but do not provide insights into the specific processing of the driver's driving environment information, or confirm whether the driver pays enough attention. Although the front video of the driving environment offers a clear understanding of the surroundings, it does not reveal the driver's internal state while processing this information. In addition, driving operation data is objective, but multi-dimensional data is necessary for accurate identification.

The fusion of information from different modalities can provide valuable information for comprehensively evaluating driver activities, indicating that multimodal fusion better simulates the unconscious behaviors and cognitive states of drivers in various driving scenarios. This multimodal approach enriches the data representation, leading to more accurate and reliable activity recognition.

DISCUSSION

The 'black-box' nature of deep learning models leads to drivers distrusting their ability to evaluate driving cognitive states. To explore the interpretability of the model, this paper attempts to visualize and analyze different feature extraction modules. For example, for the feature extraction module of driving manipulation behavior, the data outcomes of the attention mechanism are visualized.

As shown in **FIGURE 13**, the longitudinal axis represents the average attention value, showcasing the manipulation behavior feature vector of drivers across different driving activity levels on a single time step. Temporal attention distribution of the input features is uneven, exhibiting a clear bimodal phenomenon, where the attention weight at the end of the evaluation time window is not the highest. Moreover, the closer the bimodal distribution of attention weight is to the boundary of the evaluation window, the less redundant timing information is introduced by the model, indicating that a 5-second time window length is more reasonable. From the perspective of the change rule of driving state, the driver's complex cognitive state will be evaluated by multiple perceptions of the external traffic environment, rather than a single perception.

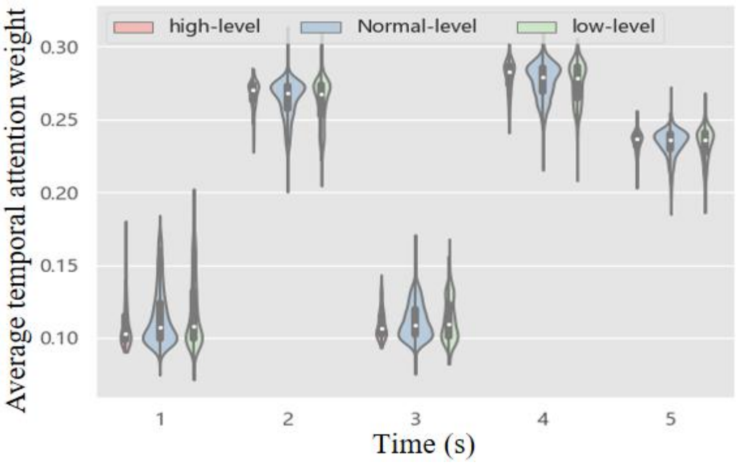


FIGURE 13 Temporal Attention Distribution of Driving Manipulation Feature

For the driver's facial features and driving environment feature extraction module, Gradient-weighted Class Activation Mapping (Grad-CAM) *(34)* is used, which can use the average gradient of the target layer to highlight important areas, and achieve visualization of the deep learning network recognition process through heatmaps.

FIGURES 14 and **15** display the feature map of the last convolutional layer before the pooling layer of each feature extraction module. Different modules focus on different types of significant information, including the driver's face, the type and quantity of the front car, the lane where the car is located and the ground signs on the lane. This focus aligns closely with intuitive human judgment, demonstrating that the model possesses the ability to learn valuable information.

For example, when evaluating the front video of a high activity group driving, the area of concern is larger and evolves over time. Initially, the model focuses on scattered vehicle information. As the subject vehicle approaches the front vehicle, the model increasingly focuses on the distance to the front vehicle. This change in focus is consistent with human driving, where vehicles are controlled according to the type of the front vehicle and the distance between them while following. This analysis underscores the model's capability to adapt its attention in a manner that reflects human-like cognitive processes.

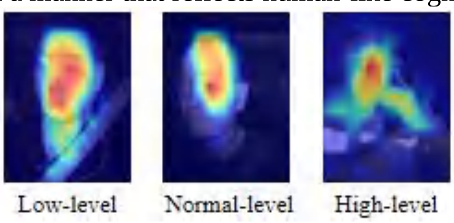


FIGURE 14 Visualization of Driver Facial Feature

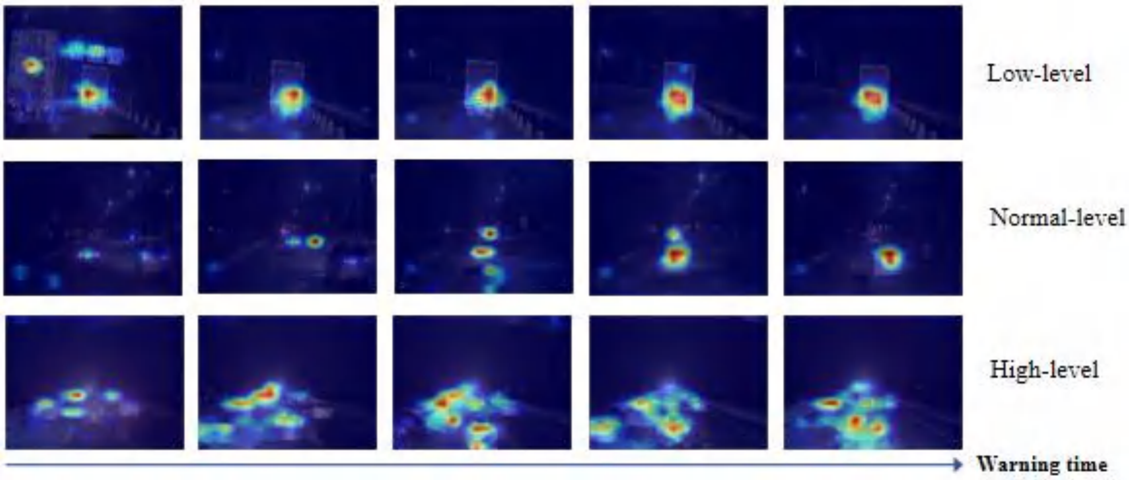


FIGURE 15 Visualization of Driving Environment Feature

CONCLUSIONS

The objective of this paper is to devise a multimodal feature learning method for the real-time analysis of drivers' cognitive states. This study aims to identify whether drivers in intelligent vehicles exhibit cognitive deficit or heightened concentration, due to stronger cognitive ability or a lack of trust in ADAS. To this end, a novel driver cognitive state estimation framework is developed through multimodal fusion (MMANN). The effectiveness of this framework is rigorously validated using a naturalistic heavy truck driving dataset, which spans 180 days and encompasses a wide array of road types and traffic conditions.

Through observational analysis of truck drivers during Forward Collision Warning (FCW) events, the cognitive state of drivers can be categorized into three distinct levels: low activity, normal activity, and high activity. Drivers exhibiting low activity frequently experience monotony and distraction, demonstrating reduced sensitivity to external stimuli and exhibiting longer reaction times (Mean = 2.14s, SD = 0.49). Conversely, highly active drivers are susceptible to stress or overconfidence, often neglecting certain environmental cues and engaging in aggressive driving behaviors. Comparative analysis with the benchmark models and ablation experiments reveals that reliance on unimodal data is insufficient for constructing a robust driver activity evaluation framework. The integration of diverse modal information significantly enhances recognition accuracy, furnishes more comprehensive insights for evaluating driver activities, and underscores the efficacy of multimodal learning approaches.

It is foreseeable that multimodal fusion can effectively simulate the unconscious behavior and cognitive states of the driver in various driving scenarios. Additionally, employing a multi-level attention mechanism allows for the highlighting of crucial moments and key module feature information while reducing the computational complexity of the whole network. These findings have important implications for enhancing driver safety and improving ADAS performance.

Recommended future work could explore different fusion schemes, as well as the evolutionary mechanisms of driver activity and its multimodal characteristics. This could involve investigating alternative methods for integrating multimodal data, understanding how driver behaviors and cognitive states evolve over time, and identifying new multimodal features that could further improve the accuracy and reliability of driver state recognition systems. Such research could pave the way for even more advanced and effective ADAS technologies, contributing to safer and more efficient driving experiences.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ACKNOWLEDGMENTS

This work was partially supported by the National Natural Science Foundation of China (Grant No. 52172306,52272315).

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: Qi He: Methodology, Data analysis, Writing up. Yibing Wang, Ken Yang: Conceptualization, Methodology. Xuewen Yao: Data preparation, Validation. Jingqiu Guo: Conceptualization, Methodology, Writing up. Mark Stevenson: Writing up.

REFERENCES

1. Hussain, R., and S. Zeadally. Autonomous Cars: Research Results, Issues, and Future Challenges. *IEEE Communications Surveys & Tutorials*, Vol. 21, No. 2, 2019, pp. 1275-1313.
2. Hu, z., C. Lv, P. Hang, C. Huang and Y. Xing. Data-Driven Estimation of Driver Attention Using Calibration-Free Eye Gaze and Scene Features. *IEEE Transactions on Industrial Electronics*, Vol. 69, No. 2, 2022, pp. 1800-1808.
3. Department of Transportation. *Ensuring American Leadership in Automated Vehicle Technologies: Automated Vehicles 4.0*. 3. Department of Transportation, U.S. Accessed on March 28,2022.
4. SAE International. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Society of Automotive Engineers. Accessed on April 30, 2021.
5. Yang, H., J. Wu, Z. Hu and C. Lv. Real-Time Driver Cognitive Workload Recognition: Attention-Enabled Learning with Multimodal Information Fusion. *IEEE Transactions on Industrial Electronics*, Vol. 71, No. 5, 2024, pp. 4999-5009.
6. Yue, L., A. A. Mohamed, Y. Wu, J. Ugan, and C. Yuan. Effects of Forward Collision Warning Technology in Different Pre-crash Scenarios. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol.76, 2021, pp. 336-352.
7. Gaspar, J., and T. Brown. Matters of State: Examining the Effectiveness of Lane Departure Warnings as a Function of Driver Distraction. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 71, 2020, pp. 1-7.
8. Çapalar, J., and C. O. Monreal. Hypovigilance in Limited Self-driving Automation: Peripheral Visual Stimulus for a Balanced Level of Automation and Cognitive Workload. *20th International Conference on Intelligent Transportation Systems (ITSC)*, Yokohama, IEEE, Japan, 2017.
9. Liu, R., B. Reimer, S. Song, B. Mehler, E. Solovey. Unsupervised FNIRS Feature Extraction with CAE and ESN Autoencoder for Driver Cognitive Load Classification. *Journal of Neural Engineering*, Vol. 18, No. 3, 2021.
10. Chen, B., X. Zhao, Y. Li, and X. Liu. Exploring the Associations of Demographics and Scale Measures with Cognitive Driving Behavior among Older Drivers in China. *Accident Analysis & Prevention*, Vol. 200, 2024.
11. Lee, J., and J. H. Yang. Analysis of Driver's EEG Given Take-Over Alarm in SAE Level 3 Automated Driving in a Simulated Environment. *International Journal of Automotive Technology*, Vol. 21, 2020, pp. 719–728.
12. Li, N., Y. Lu, and K. Li. Comparison of Subjective Mental Workload Evaluation and Heart Rate Data of UAV Operation. *Journal of Safety Science and Technology*, Vol. 17, No. 4, 2021, pp. 35-41.
13. Tanshi, F., and D. Söffker. Modeling of Takeover Variables with Respect to Driver Situation Awareness and Workload for Intelligent Driver Assistance. *2019 IEEE Intelligent Vehicles Symposium (IVS)*, Paris, IEEE, France, 2019.
14. Wang, A., J. Wang, W. Shi, and D. He. Cognitive Workload Estimation in Conditionally Automated Vehicles Using Transformer Networks Based on Physiological Signals. *Transportation Research Record*, Vol. 2678, No.12, 2024, pp. 1183-1196.
15. Azman, A., M. F. A. Abdullah, S. Yogarayan, S. F. A. Razak, H. Azman, K. S. Muthu, and H. Suhaila. Measuring Driver Cognitive Distraction through Lips and Eyebrows. *International Journal of Electrical and Computer Engineering*, Vol. 12, No. 1, 2022, pp. 756–769.

16. He, D., Z. Wang, E. B. Khalil, B. Donmez, G. Qiao, and S. Kumar. Classification of Driver Cognitive Load: Exploring the Benefits of Fusing Eye-Tracking and Physiological Measures. *Transportation Research Record*, Vo. 2676, No. 10, 2022, pp. 670-681.
17. Kumar, S., D. He, G. Qiao, and B. Donmez. Classification of Driver Cognitive Load based on Physiological Data: Exploring Recurrent Neural Networks. *2022 International Conference on Advanced Robotics and Mechatronics (ICARM)*, Guilin, IEEE, China, 2022.
18. He, D., B. Donmez, C. C. Liu, and K. N. Plataniotis. High Cognitive Load Assessment in Drivers Through Wireless Electroencephalography and the Validation of a Modified N-Back Task. *IEEE Transactions on Human-Machine Systems*, Vol. 49, No. 49, 2019, pp. 362-371.
19. Winter, J. C. F., R. Happee, M. H. Martens, and N. A. Stanton. Effects of Adaptive Cruise Control and Highly Automated Driving on Workload and Situation Awareness: A Review of the Empirical Evidence. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 27, 2014, pp. 196–217.
20. Fan, P., J. Guo, Y. Wang, and W. S. Jasper. A Hybrid Deep Learning Approach for Driver Anomalous Lane Changing Identification. *Accident Analysis & Prevention*, Vol. 171, No. 106661, 2022.
21. Fan, C., Y. Peng, S. Peng, H. Zhang, Y. Wu, and S. Kwong. Detection of Train Driver Fatigue and Distraction Based on Forehead EEG: A Time-Series Ensemble Learning Method. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, No. 8, 2022, pp. 13559-13569.
22. Ma S., W. Zhang, Z. Yang, C. Kang, and H. Li. Promote or Inhibit: An Inverted U-shaped Effect of Workload on Driver Takeover Performance. *Traffic Injury Prevention*, Vol. 21, No. 7, 2020, pp. 482-487.
23. Agrawal, S., and S. Peeta. Evaluating the Impacts of Driver's Pre-warning Cognitive State on Takeover Performance under Conditional Automation. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 83, 2021, pp. 80-98.
24. Gurusinghe, G. S., T. Nakatsuji, Y. Azuta, P. Ranjitkar, and Y. Tanaboriboon. Multiple Car-Following Data with Real-Time Kinematic Global Positioning System. *Transportation Research Record*, Vol. 1802, No. 1, 2002, pp. 166-180.
25. Golovanevsky, M., E. Schiller, A. Nair, R. Singh, and C. Eickhoff. One-Versus-Others Attention: Scalable Multimodal Integration for Clinical Data. *arXiv Preprint arXiv:2307.05435*.
26. Szegedy, C., V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, IEEE, USA, 2016.
27. Muller, R., S. Kornblith, and G. E. Hinton. When does label smoothing help? *Advances in neural information processing systems*, Vol. 32, 2019, pp. 4694 – 4703.
28. Wang X., M. Yang, and D. Hurwitz. Analysis of Cut-in Behavior Based on Naturalistic Driving Data. *Accident Analysis & Prevention*, Vol. 124, 2019, pp. 127-137.
29. Arnold, E., O. Y. Al-Jarrah, M. Dianati, S. Fallah, D. Oxtoby, and A. Mouzakitis. A Survey on 3D Object Detection Methods for Autonomous Driving Applications. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, No. 10, 2019, pp. 3782–3795.
30. Oehmcke, S., O. Zielinski, and O. Kramer. KNN Ensembles with Penalized DTW for Multivariate Time Series Imputation. *2016 International Joint Conference on Neural Networks*, Vancouver, BC, IEEE, Canada, 2016.

31. Wijnands, J., J. Thompson, G. D. P. A. Aschwanden, and M. Stevenson. Identifying Behavioural Change among Drivers using Long Short-Term Memory Recurrent Neural Networks. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 53, 2018, pp. 34-49.
32. Bai, S., J. Z. Kolter, and V. Koltun. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv Preprint arXiv:1803.01271*.
33. Wang, Z., W. Yan, and T. Oates. Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline. *2017 International joint conference on neural networks (IJCNN)*. Anchorage, AK, IEEE, USA, 2017.
34. Ramprasaath, R. S., M. Cogswell, and A. Das, R. Vedantam,D. Parikh, and D. Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, IEEE, Italy, 2017.

收录信息

世 界 交 通 运 输 大 会

WORLD TRANSPORT CONVENTION

2024世界交通运输大会摘要集录用证明

何琦、樊鹏程、王亦兵、郭静秋 先生/女士：

您的论文“驾驶行为异常识别：研究进展、方法与数据驱动实例”，经WTC学部委员会专家评审，已被大会录用并收录至大会摘要集。

特此证明。

世界交通运输大会执委会

2024 年 4 月 30 日

论文全文

驾驶行为异常识别：研究进展、方法与数据驱动实例

何琦[1]，樊鹏程[2]，王亦兵[1]，郭静秋[3]

（1.浙江大学 智慧交通研究所，杭州，310058；

2.同济大学 道路与交通工程教育部重点实验室，上海 201804 ）

摘要：驾驶行为异常识别包含如何定义异常驾驶，精准识别异常行为，及合理解释识别结果与相应行为机理三个层次。本文首先围绕异常行为预警这一核心目的，对异常识别与预测的研究进展进行梳理，重点关注基于驾驶员意图和状态的异常行为识别、自主驾驶环境下多源数据驱动的概率统计、神经网络、深度学习异常识别及预测方法等，分析各种技术路径的优缺点及面临的挑战和发展方向。同时，本文提出一种基于神经网络-深度学习的复合无监督学习模型，并聚焦车辆换道事件，从车辆换道特征指标中挖掘显性和隐性异常行为进行模式识别。基于轨迹大数据换道样本开展实例分析，结果表明不同驾驶员驾驶状态异常率分布及车辆角速度、车头间距等特征呈现异质性。此外，无监督属性的注意力机制复合模型无需先验知识和阈值，具备良好鲁棒性和自学习能力，为自主驾驶环境下准确快速识别异常驾驶行为研究提供新思路。

关键词： 交通工程；智车系统；驾驶状态辨识；异常识别；注意力机制

中图分类号： **文献标志码：**

Anomalous Driving State Recognition: State-of-the-art overview, methods, and case studies

Abstract: Anomalous driving behavior recognition includes three levels: definition, identification, and explanations. This paper reviews the research progress of anomalous behavior recognition and prediction, with the focus on recognition methods based on driver intentions and states, probability statistics driven by multi-source data in autonomous driving environments, neural networks, deep learning anomaly recognition and prediction methods, and analyzes the advantages or disadvantages. Then the study proposes a hybrid unsupervised deep learning model on lane changing events, mining explicit and implicit abnormal behavior patterns. Based on trajectory datasets, experimental analysis was conducted, and the results showed heterogeneity in the distribution of anomalous driving states among drivers, as well as characteristics such as vehicle angular velocity and headway. In addition, the hybrid unsupervised attention mechanism model does not require prior knowledge and thresholds, and shows good robustness and self-learning ability, providing ideas for accurately and efficiently identifying anomalous driving behavior in autonomous driving environments.

Key words: transportation engineering; state recognition; anomaly; attention mechanism;

1. 异常驾驶行为识别研究

驾驶行为是驾驶员在多变交通环境中，结合自身状态对车辆进行操纵的表征，具有复杂性和不确定性。目前，智车系统如何精准高效识别异常驾驶行为仍是一项极具挑战的工作[1]。本节主要从行为驱动和数据驱动的角度对异常行为识别方法进行综述，讨论异常驾驶行为识别研究存在的主要挑战及发展方向。

1.1 行为驱动的异常驾驶行为识别

1.1.1 驾驶员意图识别研究

早期研究主要从驾驶员生理和心理指标分析驾驶意图和驾驶状态，如 Bonchek-Dokow[2]基于心理学理论构建的识别模型分为意图检测和意图预测：前者是基于对意图的度量，后者是将不完整的动作行为延

收稿日期：
基金项目：

伸到其预期目标的能力[14]。驾驶人意图可理解为控制车辆未来运动过程的规划[3]，从而能够辨别一个动作行为是否有潜在意图，如关注换道[4]、超车[5]、制动[6]等行为的驾驶意图检测。基于对意图的度量，意图预测能对驾驶员的潜在驾驶意图进行判别和分析，且能精细判别行驶环境[15-16]。

学者基于人机交互过程，通过人类姿态、生理信号、交互作用力等指标推断人类意图，并尝试将意图信号输出给规划和控制策略[17-19]。如使用参数化心率变异率（HRV, Heart Rate Variability)分析方法表征驾驶员当前压力水平，并利用道路正动能、平曲线半径等特征来推算驾驶员压力水平如何演变[24]。此外，驾驶员社会性特征被认为是驾驶员有效感知环境和驾驶安全的关键因素[25-26]，而对驾驶人的异常状态识别也受到较高关注，如疲劳[7]、分心[8]、激进[9]等状态类型。一项基于驾驶焦虑的研究结果表明，焦虑受试者的路怒症程度更高，并伴随高水平的驾驶回避现象和消极驾驶行为[27]。不同驾驶风格类型和驾驶状态驾驶员的警觉性和操控能力都存在一定差异[10-13]。但是这类实验往往存在数据采集成本高、特征维度单一、检测设备对自然驾驶状态具有干扰性等不足。

1.1.2 智车环境下意图识别研究

近年来学者分别针对智联驾驶环境下驾驶策略层意图、战术层意图、操作层意图开展研究（见图 1)。策略层可以理解为行程路径的总体规划，持续时间较长,且提供实时天气、道路环境信息，对安全驾驶和合理路径规划至关重要[24]。战术层需根据周围驾驶环境做出短期行为决策，常见行为包括换道、超车、转向、制动等[25,26,27]。操作层可理解为驾驶意图的体现过程，研究主要集中于驾驶风格辨别和异常行为识别，如 Chen 等[28]基于驾驶员异质性提出了一种图形化建模方法来直接描述驾驶员行为特征；Sun 等[29]基于网联车真实数据，利用指数加权移动平均算法与累积和图(Cumulative Sum charts,CUSUM)提出了在线监测驾驶员状态方案以及时提供分心驾驶警告。

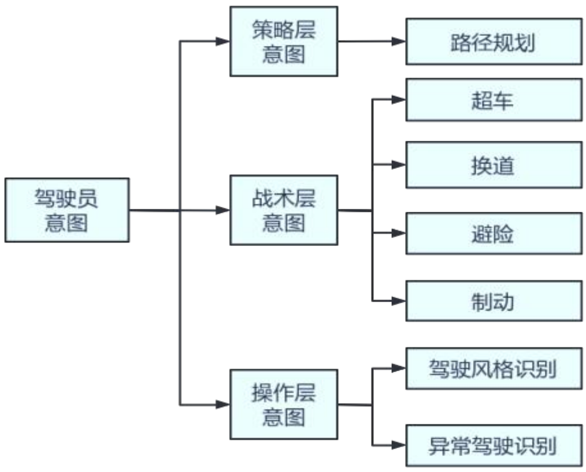


图 1 驾驶员意图分类[30]

Fig.1 Driver’s Intention Classifications

服务于自主驾驶的意图识别系统需要多个模块（交通与路权感知模块、载具运动测量模块、行为和生理信号模块等）协同工作。当前，大多数研究主要采用机器学习方法来进行驾驶员意图识别[31,32,33]，算法主要包括规则模型、深度学习、半监督学习和在线学习、生成模型等。监督学习利用标注数据集挖掘交互耦合关系、驾驶风格等隐式信息，其中，生成式模型相较判别式模型对多意图推理任务更为有效[34,35,36]。在此基础上，Yang 等[37]提出了主车驾驶员意图推断框架，其将环境感知视为意图刺激，驾驶员行为和车辆运动是对刺激的响应。此外，学者指出意图识别模型性能评价应以模型精度和预测时域为主要指标。在数据源方面，目前多采用非网联环境下的自感知数据作为输入，仅有少量研究利用网联数据进行意图识别探索[38,39]，网联环境如何影响驾驶意图仍需研究，且缺乏基于多源多维异构信息开发的意图识别模型。

1.2 数据驱动的异常驾驶行为识别

智能网联和传感技术的迅速发展支持数据驱动异常驾驶行为识别研究的深入发展。本节主要分概率统计、神经网络、深度学习三类算法概述。

1.2.1 概率统计方法

研究者主要采用的统计分析方法见表 1。此外，通过融合不同方法能提高识别精度，如 Zhang 等[40]基于博弈论建立了 GMM - HMM 车辆行为识别模型；Li 等[41]提出了一种结合 HMM 和 BF 算法来识别驾驶员换道意图，其在 HMM 的当前和历史输出的基础上使用 BF 输出最终驾驶行为分类，较单纯 HMM 算法有了显著提高。

表 1 驾驶行为研究常用算法介绍

Tab.1 An Introduction of Driving Behavior Algorithms

常用方法	主要功能	特点	相关文献
隐马尔可夫（Hidden Markov Model,HMM)	预测	状态条件分布只依赖于当前状况，需进行一定理想化假设	[42,122,123,124,125]
高斯混合模型(Gaussian Mixture Model,GMM)	分类	算法速度快，但易受数据结构影响	[40,44,133]
贝叶斯滤波(Bayesian Filtering,BF)	分类	可适应不同类型非线性系统，简单高效	[43,126,127,128]
卡尔曼滤波(Kalman Filtering，KF)	预测	线性无偏，但依赖原始数据质量	[45,129,130]
差分自回归移动平均（Autoregressive Integrated Moving Average Model,ARIMA）	预测	无需借助外生变量，模型复杂度低	[131,132,133,134]

异常轨迹通常是指在距离度量方面(如路线和行程时间)与数据库中其他轨迹显著不同的部分或全部轨迹[46]。一般利用聚类或频繁模式挖掘方法，如果轨迹不能归在任何簇内或者不频繁，则具备较大概率被视为异常轨迹[47]。因此，识别异常轨迹的关键技术在于轨迹相似性度量与聚类。

针对车辆轨迹时间序列属性产生了一系列相似度描述方式。Faloutsos 等[48]阐述了一种在时序数据上进行相似性搜索的通用框架，之后 SVD(Singular Value Decomposition) [49]、DWT(Discrete Wavelet Transformation)[50] 、切比雪夫多项式[51]等时间序列降维方法被先后提出，这些方法都能有效避免误判。在轨迹异常检测方面，Sanjay 等[53]提出了两步式挖掘和 L1 优化来以数据推断道路交通中出现异常的根本原因。Mao 等[54]根据特征差异，划分了两种异常轨迹类型（局部异常轨迹片段和异常移动对象），并结合一种新数据结构（Ally Fragment)来有效地检测连续轨迹流的两类异常。

目前,在轨迹数据异常检测方面已经开展了深入研究,但由于轨迹数据的稀疏性、时变演化性和偏态分布等特征，面向智车系统轨迹数据实时异常检测的研究相对较少[72,73]。Bu 等[72]利用轨迹局部连续性特征在轨迹流上构建局部簇，并通过高效剪枝策略来实时检测异常轨迹，实现了数量级性能提升。此外，发现正常轨迹和高效的在线检测是两个主要挑战，解决思路之一是利用机器学习算法来捕获轨迹中包含的复杂序列信息，并通过高效轨迹生成来支持轨迹实时检测[74]。

1.2.2 神经网络

利用激光雷达、摄像头、高精度 GPS 等智能设备采集高精度行驶数据，再结合神经网络相关算法对进

行数据挖掘和模式分析。智能车辆传感器可获取车辆运行状态，如激光雷达、摄像头、CAN 总线及嵌入在智能设备中的传感器。基于智能手机数据，Bergasa 等[55]开发了一款检测驾驶员疏忽驾驶行为并给予相应反馈的应用程序；Ma 等[57]提出了一种长短期记忆-残差算法来实时检测异常驾驶行为，并采集 4 种典型异常驾驶动作（急加速、急制动、急转弯和急变道）的车辆数据对算法有效性进行验证。但是，智能设备采集频率较低且没有足够的驾驶环境感知能力，设备放置姿态对信息精确程度有显著影响。

时空环境下，综合使用多类型智能传感器对一个或多个移动对象（车辆，轮船，人等）采集所获得的数据信息称为轨迹大数据[58]，该类数据包含了时间戳、地理坐标、对象类型、速度等信息，给人们带来了海量信息，同时也为驾驶行为研究提供了新的着力点。通常，异常轨迹可视为异常驾驶行为的一致性有效外显，利用轨迹数据来识别异常驾驶行为的关键在于：异常轨迹数据的识别和轨迹预测。异常轨迹识别能为之后的轨迹预测圈定数据范围，而轨迹预测能补全轨迹，提高识别精度，两者之间相辅相成[47]。

在异常轨迹识别方面，常通过聚类算法来发现异常驾驶事件，主流研究方法见表 2。在此研究基础上，一些算法的变体也先后被提出用于挖掘轨迹序列特征，实现异常驾驶行为检测[59,60]。Mirco 和 Dino[61]在 OPTICS 算法基础上将对象间距的空间概念推广到轨迹之间距离的时空概念，采用了一种基于时间聚焦（T-OPTICS）轨迹聚类方法，旨在利用时间维度的内在特性来提高轨迹聚类质量。在轨迹预测方面，当前研究方法大多基于马尔科夫链或其他改良模型[62,63,64]，或是基于历史轨迹的查询补充，如文献[65]扩展了马尔科夫链(Mobility Markov Chain，MMC)模型，研究人员使用前 n 个位置数据提出了一种位置预测算法 n – MMC，在 3 个真实数据集上的效率评估表明，当 n = 2 时，对下一个位置的预测精度在 70% ~95%之间。

表 2 不同类型的聚类算法

Tab.2 Different Types of Clustering Algorithms

分类	常见算法	主要特点	相关文献
基于密度	DBSCAN、OPTICS	在给定范围中包含的数据点	[61,73,74,75]
基于距离	R-Tree、TRAOD	与大多数轨迹具有较远距离的轨迹视为异常	[66,67,68,69]
基于分类	K-Means、K-Medoids、EM	将对象集合划分为聚类簇	[70,71,72]
基于网格	STING、CLIQUE	将空间划分为有限数量单元	[76,77,78,79]

在实际交通系统中，异常驾驶行为不常发生，且普遍低采样频率，导致收集到的数据十分稀疏。为解决此问题，Jianming 等[80]融合了多个局部卷积层对轨迹数据进行增强补偿，其提出将轨迹从不同角度建模为二维图像的预测算法 T – CONV，并用多层卷积神经网络提取多尺度二维轨迹特征实现精确预测，结果表明轨迹起始点和终点位置附近对目的地的预测有明显影响。Chiang 等[81]在时间约束移动图上提出了一个基于可达性的模型(Reachability-based prediction model on Time-constrained Mobility Graph，RTMG)来预测候选位置。研究者设计了一种自适应时间探索方法来提取时间上接近给定查询时间的有效支持轨迹，且根据 RTMG 选取可达概率最大的候选位置作为预测结果，大量真实数据验证了 RTMG 的有效性和高效性。

1.2.3 深度学习

由于驾驶行为具有不确定性、动态性和多交互性质，传统特征提取方式不能全面地获取信息，而深度学习由人工神经网络演变而来，对行驶数据特征的学习更加准确，泛化能力强，弥补了概率统计方法识别精度不高的缺点。

在深度学习中，作为基础网络结构的多层感知机（Multi-Layer Perception，MLP）被广泛应用于驾

驶行为研究领域中，文献[82,83,84]分别利用 MLP 实现交通环境的感知、交通流量的预测以及三维目标检测。在混合交通条件下，可利用MLP量化分析行人行为特征，并将其融合到驾驶决策中以根据不同行人特征进行差异化决策，提升道路通行能力和通行效率[85]。

相较于 MLP,堆叠编码器（Stacked Autoencoder，SAE）能够降低数据维度并提取潜在的数据特征，且对含噪数据保持较高的准确性和鲁棒性[86]。Guo 等[87]提出了一个混合无监督深度学习模型来研究危险驾驶行为，其结合了自动编码器和自组织映射（Self-Organizing Map，SOM）两种降维技术，对较高风险相关的各种驾驶模式进行分类和聚类。实验结果表明，通过自动编码器反向传播对非线性多模态降维有效，且模型学习到的驾驶行为特征和聚类结果是可解释的。

当前，深度学习模型发展进化呈现出以下趋势：模型深度增加[88]，结构越来越复杂[89]，模型中孤立路径[90]的数量也大幅增加。一些常用深度学习方法被运用到驾驶行为识别和预测当中:递归神经网络（Recurrent Neural Network,RNN)在处理时间序列上具有优势，能有效识别异常驾驶行为[57,91]；卷积神经网络(Convolutional Neural Network, CNN)能很好地处理高维数据，一些经典模型如 DenseNet[92]、GoogleNet、ResNet[93]由此结构演变而来。受 DenseNet 启发，Huang 等[88]引入了三种深度学习融合模型，并将模型宽度显著增加来全面描述时间和空间潜在信息，首次实现了基于视频的异常驾驶行为检测。

车联网交通背景环境下，视距范围内的不同交通主体行为逻辑存在显著差异，低时延、高精度的驾驶行为识别和预测算法会间接影响交通系统的安全性、稳定性和舒适性。主流研究方法总结见表 3，物理学模型方法主要考虑载具底层运动信息，后两者方法则主要考虑上层运动信息和历史运动数据。其中数据驱动的轨迹预测方法如强化学习、迁移学习等，可实现长时域内的车辆行为预测[94]，逐渐成为主流技术，如 Uber 的 Lane RCNN、Google 的 Vector net、Waymo 的 TNT、Aptive 的 Covernet、MIT 的 Social LSTM 等[95,96,97,98,99]。此外，基于混合模型的驾驶行为识别方法通过融合不同方法的优点可以提高识别精度和拓展预测时域，如文献[100,101]分别将基于规则的方法与机器学习融合、将分子动力学理论与深度强化学习融合来研究车辆换道决策问题。

表 3 驾驶行为识别和预测研究方法

Tab.3 Methods of Driving Behavior Recognition and Prediction

在真实驾驶环境中，交互对象行为不确定性、环境感知不确定性、车辆模型及其行车环境的不确定性都会直接影响识别与预测算法的有效性，进而降低模型精度。随着载运系统自主化水平和网联程度逐步提升，意图识别和行为预测系统要求更高的可靠性、可解释性、实时性，以下三个方向值得重点突破：

研究分类	常用方法	相关文献	优点	缺点
物理学模型	车辆动力学	[103,104]	精度较高	需进行一定的理想化假设
行为驱动	意图检测 意图预测	[5,6,7,9,10] [16,17,18,19]	结构清晰，方法多元化	针对单一的意图识别场景
数据驱动	概率统计模型 神经网络 深度学习 混合模型	[105,106,107,108] [59,60,110,111] [88,89,90,112] [87,100,101,102]	模型可解释性强，应用广泛 自适应能力好 泛化能力强 精度高，实时性强	参数设置复杂，易受结构影响 收敛速度较慢 可解释性较弱 模型较复杂，训练时间长

1）海量行为数据的不确定性与数据在时间和空间两个维度上的演变模式紧密相关，传统的不确定性度量无法胜任。因此，探究可以同时给出预测结果和不确定性度量的时空序列预测算法值得研究。

2）基于深度学习的行为预测算法虽然可以从感知信息中学习到数据蕴含模式，但缺乏因果推断能力，影响了其在数据分布偏移、干预情况下的态势预测能力。将因果科学与深度学习结合也是推进多模感知数据预测模型落地应用的重要方向。

3）探索在跨域应用场景下将预测模型与实时混行交通环境对接，实现流式模型训练是提高实时预测能力和可拓展能力的重要课题。

在此背景下，本文有三个主要目标：1）通过进行简明而深入的概述，以确定各种技术路径的优缺点，并为未来的研究制定可行的路线图；2）设计一种无监督个性化异常换道行为识别模型，避免主观设定异常阈值和模型训练数据标签的问题。综合使用非参数核密度估计（Kernel Density Estimation，KDE）和注意力机制，深入分析异常换道行为和正常换道行为的时空特征差异，增进深度学习模型的可解释性。

2 研究方法

基于时空轨迹数据的异常行为识别与轨迹预测通过挖掘目标历史位置信息和行为习惯，计算目标未来的位置信息和行为动态并加以对比，可以预警潜在的风险要素、提升交通态势的感知能力。传统方法难于从复杂历史数据中学习到轨迹结构特征和驾驶行为特征。本文基于深度学习复合模型融合不同方法的优点，并引入注意力机制提升特征显著性，构建一种驾驶行为异常识别个性化模型。

虽然驾驶状态呈多样性，为了增强模型泛化性，本文将驾驶状态集设定为{正常，异常}，转化为无监督二分类问题进行建模。参考相关研究[2,12,13]及前序研究，如表 1 所示选择 8 项换道特征 $\xi\,[v, \omega_Z, a_x, a_y, ,d_o ,d_t ,v_{ro} ,v_{rt}]$作为时序输入。

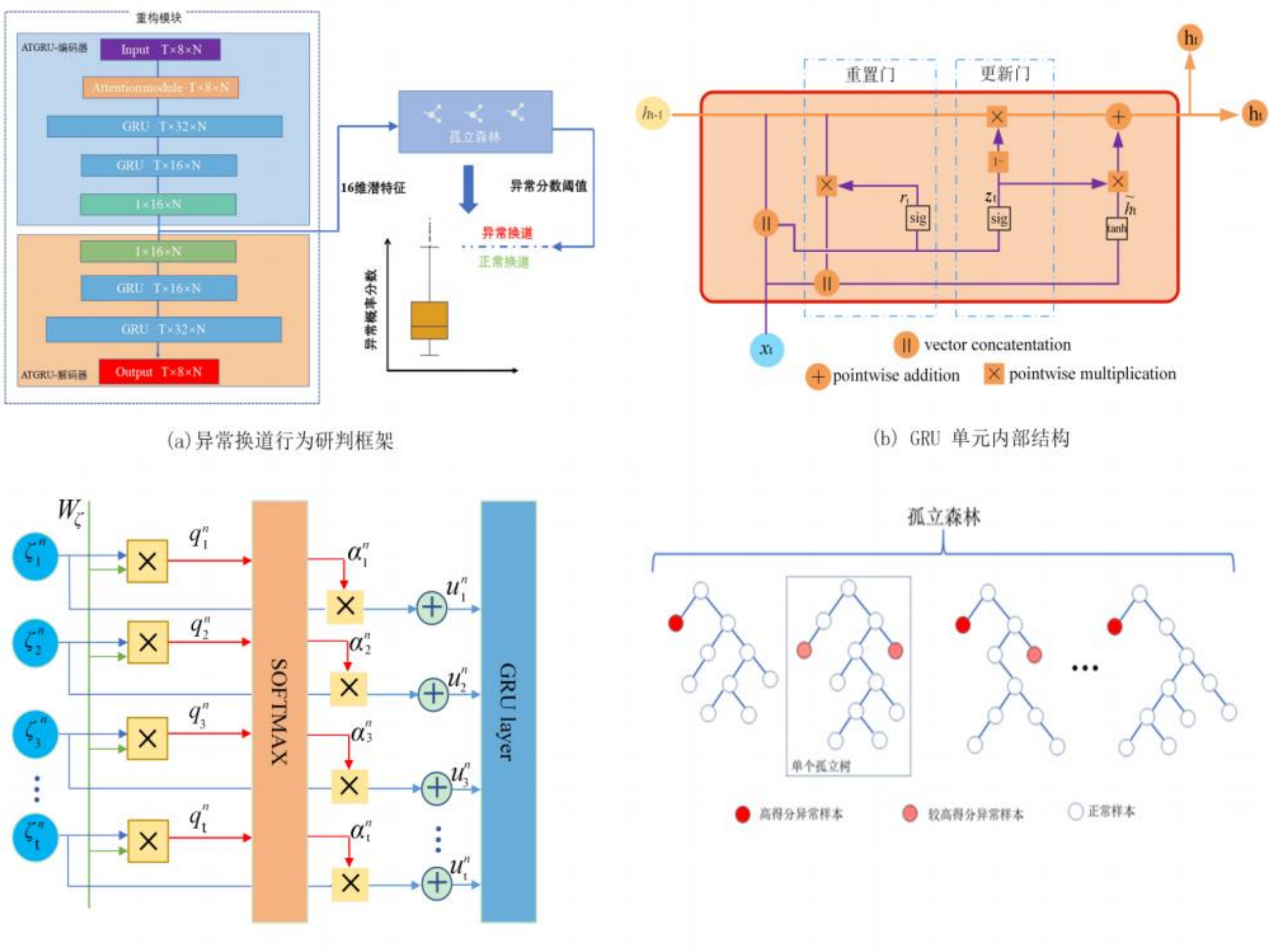
表 1 输入特征

Tab.1 Parameter Descriptions

符号	类型	相关参数解释	单位
v	自身特征	本车速度	m • s^{1}
ω_z	自身特征	本车角速度 (向左: 负; 向右: 正值)	degree • s^{-1}
a_x	自身特征	本车纵向加速度	g • m • s^{2}
a_y	自身特征	本车横向加速度	g • m • s^{2}
d_o	交互型	与初始车道上前车形成的车头间距	m
d_t	交互型	与目标车道上前车形成的车头间距	m
v_{ro}	交互型	与初始车道上前车形成的速度差	m • s^{1}
v_{rt}	交互型	与目标车道上前车形成的速度差	m • s^{1}

2.1 注意力机制方法

针对个体驾驶员换道行为构建一种基于神经网络-深度学习的复合无监督学习模型，包括换道行为异常概率计算模型、异常概率量化指标及阈值划分准则（如图 2）。门控循环单元组成的时序自动编码机（Attention-based Gated Recurrent Unit Autoencoder（AT-GRU））数据重构模块以单个驾驶员在换道时间窗内的自身轨迹特征和周边车辆的交互信息作为输入，叠加基于孤立森林（IForest）的异常行为识别模块共同工作。其中，AT-GRU 从时序换道数据重构过程中提炼原始输入的低维潜特征表达与嵌入式系统特征；IForest 则利用该潜特征输入进行递归划分训练；基于个性化的换道异常概率分数和修正的 Tukey’s test 准则选择合理的分数阈值进行异常行为研判。不同于常见的集计类模型学习群体特征，此复合模型关注不同驾驶员的驾驶状态空间和反应特性偏好异质性，并可在小方差标准下进一步降低拟合偏差，有效避免出现偏差-方差窘境。



(a)异常换道行为研判框架　　(b) GRU 单元内部结构

(c)注意力模块机制图示　　(d) 孤立森林结构

图 2 个性化异常换道行为研判框架

Fig. 2 The Framework for Anomalous Lane-changing Behaviors

2.1.1 重构模块

AT-GRU 自动编码机是一种无监督全对称的 seq2seq 神经网络，各神经网络层的输入或输出张量维度如图 2 所示。T 为时间步长，N 为单个驾驶员换道行为输入样本量。训练阶段的编码器把输入换道行为时序数据非线性映射到低维空间，同时解码器则将低维潜特征尽可能无损地恢复到原维度。利用梯度反向传播算法实现最小化重构误差，一旦达到收敛，编码机能够捕捉到换道特征向量的低维表征。AT-GRU 综合轨迹结构特征和驾驶行为特征学习，并可衍生相应行为风格（如驾驶员换道风格画像）。模型表达式如下：

$$min\sum_{t=1}^{k}\sum_{i=1}^{m}\left\|xt^{i},g_{\theta}(f_{\phi}(xt^{i}))\right\| \tag{1}$$

式中：f_φ 和 g_θ 分别代表编码器和解码器的复杂映射函数，$\|\ \|$ 是均方误差损失函数，t 是时间步，k 为换道序列长度，m 为数据维度。

作为循环神经网络的变体，GRU 网络层具备强大的信息挖掘和自学习能力，更有效地学习换道行为时序数据内在关联依赖；从而有效避免梯度爆炸和梯度弥散。GRU 内部结构为两个相互协调控制信息流的门：重置门（reset gate）与更新门(update gate)；前者控制前一时间步状态信息的忽视程度，后者控制前一时间步状态信息引入当前状态的比重。表达式如下：

$$
\begin{cases}
r_t = \sigma(W_r * [h_{t-1}, x_t] + b_r) \\
z_t = \sigma(Wz * [h_{t-1}, x_t] + b_z) \\
\tilde{h}_t = tanh(W_{\tilde{h}}[r_t \bullet h_{t-1}, x_t] + b_{\tilde{h}}) \\
h_t = (1 - z_t) \bullet h_{t-1} + z_t \bullet \tilde{h}_t
\end{cases}
\tag{2}
$$

式中：σ 是 sigmoid 函数，x_t 是单元的输入向量，h_t 是对应的输出隐层状态向量，r_t 是重置门状态向量，z_t 是更新门状态向量，*tanh* 为双曲正切激活函数，*w*,*b* 分别对应 GRU 自学习训练的权重向量和偏置向量。

2.1.2 注意力机制

为优化 GRU 自动编码机学习能力，本文引入近年在自然语言处理领域快速发展的软时序注意力机制(soft attention)[21]。注意力机制增强了深度学习方法的可解释性[22,23]，其机理类似于视觉感知规律，换道过程中周边车辆产生不同的影响效力，驾驶员基于历史经验及实时视觉感知，能够忽略无关信息而关注重点信息，不断调整自身换道决策。

采用的软注意力机制计算分三个步骤：首先利用单层网络进行相似性计算，见下式。

$$
{q_t}^i = W_\xi {\xi_t}^i + b_\xi \tag{3}
$$

其中 ${q_t}^i$ 是在第 t 个时间步上分配给第 i 个输入变量的原始权重分数，W_ξ 和 b_ξ 分别是梯度下降中网络学习到的权重项和偏置项。

$$
{\alpha_t}^i = \frac{exp({q_t}^i)}{\sum_{i=1}^m exp({q_t}^i)}, \sum_{i=1}^m {\alpha_t}^i = 1 \tag{4}
$$

式（4）对原始的注意力 q_t 向量中所有元素进行 *softmax* 归一化操作，使其处于[0,1]权重分数区间，得到任一输入第 i 维向量在时间步为 t 上的注意力权重 , m 为输入序列维度。最终注意力模块输出新的语义向量 u_t（由输入向量与注意力分数作乘积）传递至后续 GRU 层。

$$
{u_t}^i = \sum_m {\alpha_t}^i {\xi_t}^i \tag{5}
$$

可微分的软注意力可以通过神经网络算出梯度并且前向传播和后向反馈来学习得到注意力的权重。在 AT-GRU 自动编码机的输入层之后（而非在内部隐藏 GRU 层）加入注意力机制模块，利用时序注意力机制输出、输入数据的注意力权重分布，表征输入特征时间维重要性分布。该模型不易忽视输入数据内部隐藏的重要时序信息，支持编码机为相关性更高的时序向量优先分配权重值，有助提升编码机学习精度和效率。同时，输出的潜特征能够有效表征个体驾驶员的原始换道行为输入特征，成为先验潜在异常检测指标。且每一个 AT-GRU 模型仅接收单个驾驶员换道数据集，以个性化的方式形成自适应的特征指标，解决指标泛化性，显著抵抗干扰信息并提升针对个体的可解释能力。

2.1.3 异常识别模块

IForest 模型接收 AT-GRU 自动编码机的潜特征输入 Z 并进行训练，输出不同换道行为样本的异常概率分值。相较于单分类支持向量机或局部离群因子法(LOF)，孤立森林有线性时间复杂度和高精准度的优势[24]，面对高维海量数据和冗余特征仍能保持良好性能；可以避免变分编码机这类生成式深度神经网络[25]进行异常检测时面临地参数优化训练困难等问题。

IForest 将连续数据中的异常样本定义为数量稀少且孤立的离群点，通过递归构造二叉树结构的大量孤立树完成异常检测。首先，在输入训练集 Z 内进行小数目随机采样。然后，随机划分换道行为潜特征，并在某一特征 q 的值域内随机选定阈值 p 进行孤立树的构造，直至树深度达到最大深度或者仅存在单个叶子节点。训练结束后通过计算样本 Z_i 在所有孤立树上

的平均路径长度 $E(h(Z_i))$获得该样本异常概率分值，见式（6-7）。监测数据最终所处节点越接近孤立树的根节点，其异常概率被视为更高。

$$s(Z_i, N) = 2^{-\frac{E(h(Z_i))}{c(N)}} \tag{6}$$

$$c(N) = 2H(N-1) - \frac{2N-2}{N} \tag{7}$$

式中：c(N)是由样本数量 N 确定的二叉树搜索平均路径长度；$H(x)$ 是调和级数函数，在 n 较大时可用 ln(n)+0.57721（欧拉常数）代替。

2.1.4 异常换道研判准则

个性化行为异常识别很难通过经验性的阈值设定判断异常。同时，考虑驾驶员行为异质性，不同驾驶员具有不同的异常换道分数概率阈值。参考 Tukey's test 法则，本准则从个体驾驶员的换道数据特征出发（即无需假设数据分布概率），具有相对客观稳定的判识能力。异常换道行为准则如式（8）：

$$s(Z_i, n) \geq s(Z, n)_{25} + \frac{1}{\theta} \bullet (s(Z, n)_{75} - s(Z, n)_{25}) \tag{8}$$

式中：$s(Z,n)_{25}$ 和 $s(Z,n)_{75}$ 分别为驾驶员换道行为样本异常分数分布 25 分位值与 75 分位值，$s(Z_i,n)$是第 i 个换道行为样本的异常分数值。θ 为异常灵敏因数，表征异常换道行为识别灵敏度。该因数越小，单个驾驶员所检测的异常换道行为数量越少，即对异常换道行为识别要求越严格。根据驾驶员驾龄和事故记录等因素可以优化设置，本文采用 θ=2/3。在孤立森林模型运行完成后，模型按该准则输出当前换道行为数据的异常状态标签。

2.2 LSTM 方法

第二种尝试是基于长短期记忆网络的无监督框架，包括重构、识别和可视化三个模块，如图 3 所示，模型细节详见文献[fan2022]。重构模块基于递归卷积-自动编码器(RC-AE)从高维轨迹数据集中提取时空特征。编码器包括一个 LSTM 层、1D -CNN 层和几个最大池化层。解码器模块由 1D -CNN 层、上采样层和 LSTM 层组成。识别模块使用基于 Pauta 准则的重建损失分析以及单类支持向量机对提取的潜在特征空间进行分析。可视化模块采用 t 分布随机邻域算法（t-SNE）嵌入将异常数据样本投影到二维空间中，提升解释性。最后，通过个性化的灰色关联系数分析变道事件的时间异常，以表征单个驾驶员的稳健相似性。

另外，基于 Stacked Convolution Autoencoder (SCAE)的重构算法与 RC-AE 有类似的对称网络，但将一维卷积层替换为二维卷积层（2D -CNN）。几种方法各有优缺点和适用性，需要结合具体数据集及研究目标评价。

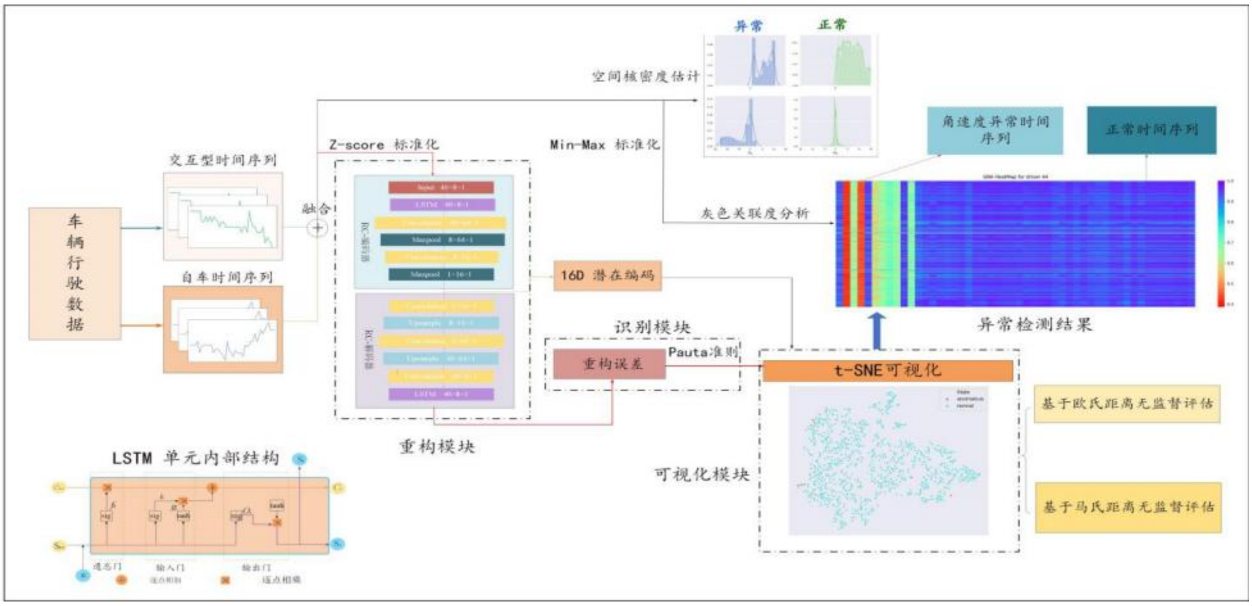


图 3 基于LSTM的研究框架[26]

Fig.3 Framework Based on LSTM

3 实证研究

3.1 数据准备

上海自然驾驶研究项目 SH-NDS（Shanghai Naturalistic Driving Study）是由同济大学、通用汽车中国公司、美国弗吉尼亚理工三方合作的中国自然驾驶项目。参与的驾驶员总计 60 位，年龄分布在 35-50 岁之间，参加试验之前已行驶总里程均大于 2 万公里，日常行驶平均里程不低于 40km[26]。5 辆试验车辆均配备了 SHRP 2 Next Gen 数据采集系统，包含全球定位系统、多普勒雷达、车道偏移系统、前后向高清摄像头等。每个驾驶员在自然状态下驾驶车辆，系统以无干扰的方式全程记录驾驶信息，如雷达传感的车辆运动信息、经纬度坐标、驾驶员的头部手部动作、周围车流视频等，数据采集频率为 10Hz。

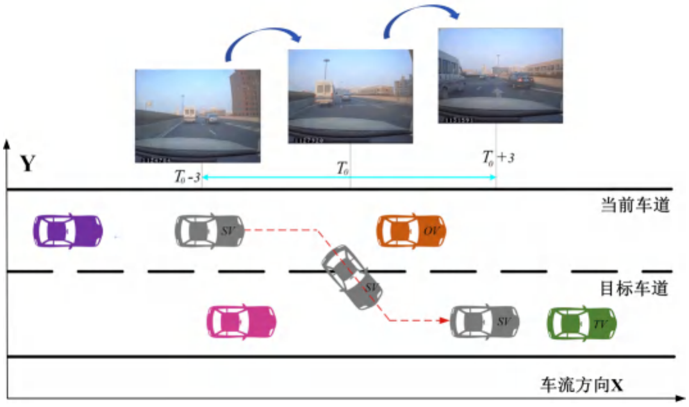


图 4 换道过程示例

Fig.4 A Lane-changing Event Sample

如图所示，*SV* 代表产生换道行为的自然驾驶车辆，*OV* 为当前车道的前车，*TV* 是目标车道的前车。换道样本提取原则为：以车道中心线偏移与方向盘转角指标进行识别单次换道行为，车辆跨过车道线时定义为换道中间点 T_0；从 T_0 双向各遍历 3*s* 数据，*T* 为时窗长度，依

次利用试验车与车辆纵横向距离指标进行匹配。为消除数据噪声，利用指数平均法进行滤波处理，以线性插值法填补缺失值。经预处理后累计提取了多位驾驶员在 3 年中累计共 6,658 次出行数据，保留了有效换道行为样本量为 58,015。为了避免各特征值间的数量级差异，在输入至 AT-GRU 自动编码机前对连续型变量的原始数据进行了 *Z*-score 标准化处理。此外为构建个性化异常换道研判模型，针对每个 AT-GRU 自动编码机模型训练输入的换道行为仅源自个体驾驶员数据。

重构模块采用 python3.6 语言并基于 Keras 2.3.1 深度学习框架实现，基于 0.001 学习率的 Adam 优化器，训练损失函数为均方误差，编码器和解码器分别有 2 层 GRU 层堆叠。训练集与测试集的比例为 85:15，训练迭代批次上界为 128，自动编码机内部各层间的激活函数为 ReLu。作为识别模块的孤立森林则利用 Scikit-Learn 框架实现，孤立树数量取 100，设定算法自动选取最大采样数。

3.2 结果与分析

由于篇幅有限，本文随机选取 6 位驾驶员进行详细分析（驾驶员编号为 6、9、26、38、43、54）。换道行为样本数介于 91 到 2465。表 2 给出了复合模型训练完成后对 6 位驾驶员异常换道统计分析，并基于式（6）计算不同驾驶员换道异常发生率。如换道记录数最低的 38 号（2.2%），而其异常换道概率阈值最高（0.6267），对应其更稳定的换道行为风格。图 7 对 6 位驾驶员的换道异常概率分数分布进行了可视化。

表 2 4.0s 秒时窗基准工况异常识别结果

Tab. 2 Anomalous Analysis Results Using 4.0sTimeWindow Length

驾驶员	换道事件数量	异常换道事件数	%	异常概率阈值
6	500	41	8.20	0.5620
9	2465	175	7.10	0.5333
26	1163	59	5.07	0.5615
38	91	2	2.20	0.6267
43	940	51	5.43	0.5670
54	1742	95	5.45	0.5355

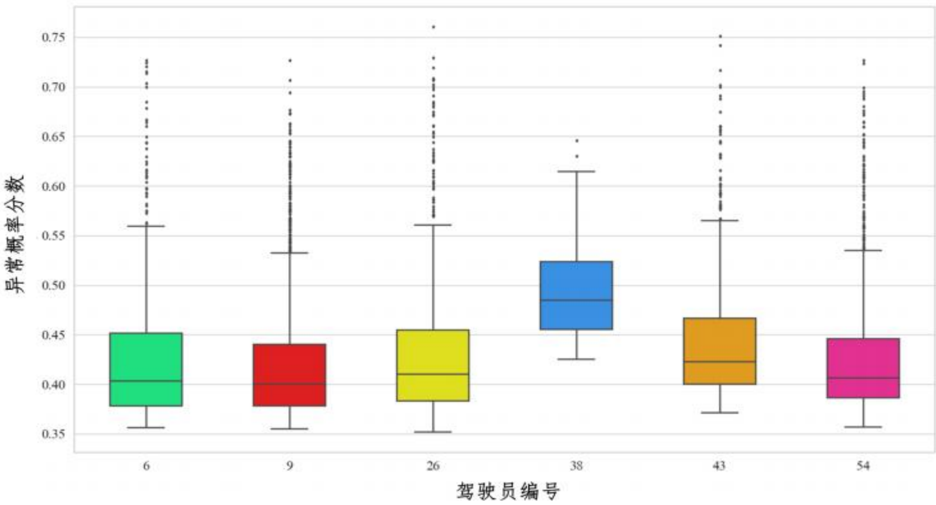


图 7. 4.0s 时窗下驾驶员异常分数阈值分布

Fig. 7 The Anomaly Scores Distribution of 6 Drivers

3.2.1 基于注意力机制的换道行为时窗

换道行为时窗长度是重要超参数之一，考虑路况、交通流、换道紧迫度、心理等要素，需要通过敏感性分析确定参数合理值。长时窗可能引入冗余时序信息，抑制 AT-GRU 的行为表征学习；短时窗则不易捕捉到数据时序内在依赖性，遗失重要特征信息。本文分别采用 2.0s、6.0s 时窗与基准工况 4.0s 进行比对分析，见表 3-4。总体而言，异常换道行为数量和异常分数阈值上没有显著波动，进一步验证了 AT-GRU 编码机和孤立森林 IForest、Tukey's test 组合使用，能对高维数据进行抽象挖掘，刻画个体驾驶员行为谱，同时弱化异常换道行为数据对模型自学习的影响，从而生成兼具个性化和稳定性的异常研判阈值分数。

表 3 2.0s 和 6.0s 对比工况下的异常换道结果敏感性分析（4.0s 为基准）

Tab. 3 Anomalous LC Results Sensitivity Analysis

编号	2.0s			6.0s		
	异常换道数	变化趋势	变化值	异常换道数	变化趋势	变化值
6	45	↑	4(0.8%)	43	↑	2(0.4%)
9	115	↓	60(-2.4%)	140	↓	35(-1.4%)
26	58	↓	1(-0.1%)	85	↑	26(2.3%)
38	4	↑	2(2.2%)	2	→	0(0.0%)
43	46	↓	5(-0.5%)	49	↓	2(-0.2%)
54	68	↓	27(-1.5%)	101	↑	6(0.3%)

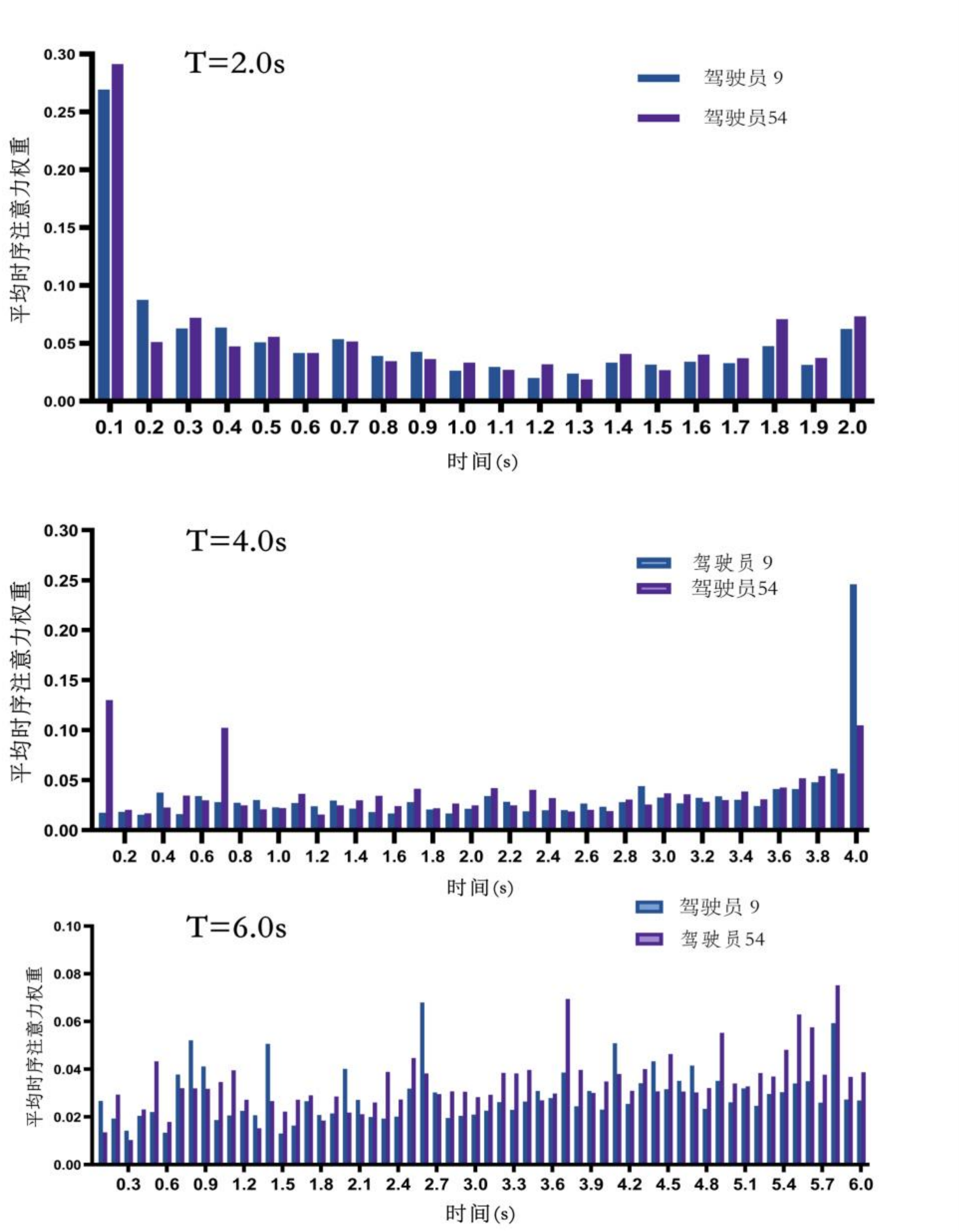
表 4 2.0s 和 6.0s 的对比工况下的异常分数阈值敏感性分析

Tab. 4 Anomalous Score Threshold Sensitivity Analysis

编号	2.0s			6.0s		
	异常换道数	变化趋势	变化值	异常换道数	变化趋势	变化值
6	0.5361	↓	-0.0259	0.5443	↓	-0.0177
9	0.5658	↑	0.0325	0.5634	↑	0.0301
26	0.5594	↓	-0.0429	0.5381	↑	0.0234
38	0.6262	↓	-0.0005	0.6087	↓	-0.0180
43	0.5550	↓	-0.0120	0.5722	↑	0.0052
54	0.5476	↑	0.0121	0.5451	↑	0.0096

图 8. 9 号和 54 号驾驶员在三类时窗下的平均注意力权重分布（颜色看的旧旧的）
Fig 8 The Average Temporal Attention Values by Drivers 9 and 54

图 8 针对软时序注意力机制数据结果进行可视化。纵轴为个体换道数据重构学习后单个

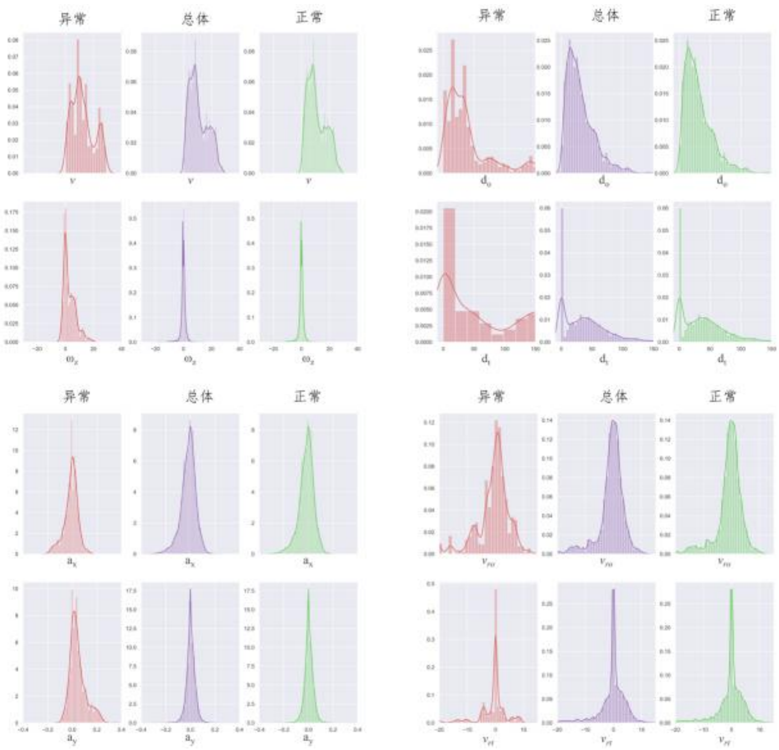


时间步上的行为向量平均注意力值，可见不同驾驶员的注意力权重分布时序差异：如使用 2.0s 时窗长度，换道行动点前后时刻的注意力权重分数均较低，换道起始点附近时间步的注意力权重分数要远高于中间部分，介于 0.25-0.3 之间，其次是换道结束点。在 4.0s 工况下，失衡情况有所缓解。9 号驾驶员更多的注意力权重分布在换道时窗末期，最终时间步的注意力权重分数可达 0.23，这与 54 号驾驶员明显不同。此外，换道行为时窗与换道持续时间 T_{lc} 高度相关，Yang 等[20]推荐 T_{lc} 样本均值 3.81s。综合考虑本研究采用 4.0s 时窗进行异常换道行为检测。

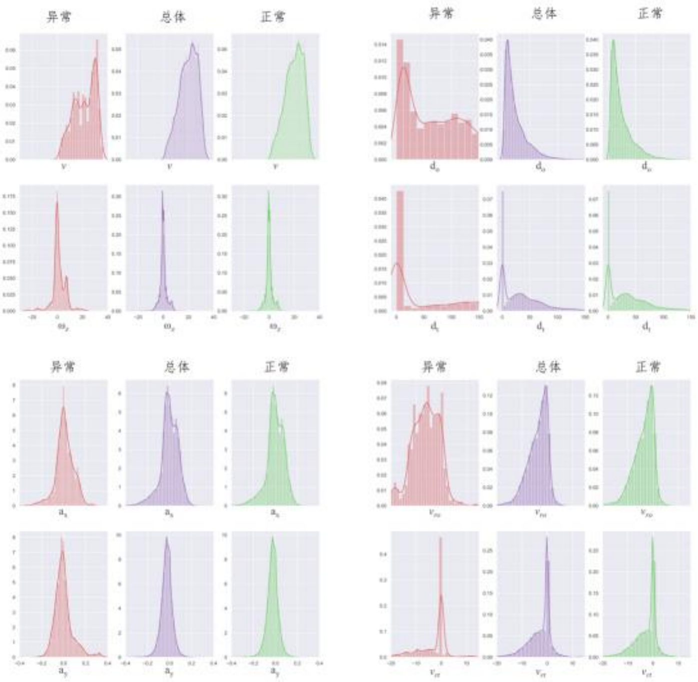
3.2.2 异常换道行为特征分析

为了提升模型可解释性，本文借助无需有关数据分布先验知识的非参数核密度估计（KDE）获取 8 个输入特征上异常组、正常组和总体组的概率密度分布展开分析。设定核函数为高斯核函数，优化积分均方误差为优化目标对最优窗宽进行自动搜寻，其结果可从概率分布上揭示 AT-GRU 自编码机和孤立森林共同学习获得的规律。

仍以 9 号和 54 号驾驶员为例，如图 10 所示，两位驾驶员的异常组概率密度分布在多数变量上不同，且在车辆角速度（ω_Z）、本车到前方两车的车头间距 d_o、与当前车道前车的速度差 v_{ro} 等变量上，异常组和正常组呈现明显差异，异常组的变量分布范围也更为分散。具体而言，9 号驾驶员正常组换道角速度的概率密度分布接近于高斯分布（均值为 0），而其异常组角速度概率分布为正偏态分布（均值>0），这表明其异常换道行为中向右换道的条件概率要高于向左侧换道，且 ω_Z 达到 10~20degree/s 的换道行为可能是紧急变道行为。而两位驾驶员换道行为异常组和正常组的纵、横向加速度指标概率密度分布之间没有显著差异，故角速度序列信息相较于加速度序列更有可能是异常换道行为研判的主要特征。54 号驾驶员在 4.0s 工况下，异常组的车头间距变量 d_o 概率密度分布为拖尾左偏分布，与正常组差异显著，这表明该驾驶员执行异常换道操作时倾向与前车保持较大距离，具备明显的风险规避特征。



a) T=4.0s Driver 9



b) T=4.0s Driver 54
图 9 核密度估计结果
Fig 9. The Results of Kernel Density Estimation

4 结束语

本文提出了一种基于注意力机制的时序 GRU 自动编码机和孤立森林复合无监督学习方法，用于个性化驾驶员异常换道行为研判。GRU 自动编码机（AT-GRU）负责重构高维时序换道特征数据，并在训练结束后捕捉相应的低维特征表达；在此基础上，孤立森林（IForest）作为异常识别模块主体单元结合修正的 Tukey’s test 准则输出无先验知识下的异常识别结果。基于上海自然驾驶实验多位驾驶员长期积累的海量换道数据库开展了多个体案例验证。结果显示，不同驾驶员的换道行为异常率差异明显，分布于 2.2~8.2%水平之间，且异常换道行为群组与正常换道行为群组之间的车辆角速度、与周围车辆纵向距离等特征变量呈现较显著的概率分布差异。不同于以往研究，本文给出了换道行为研究时域 2s、4s、6s 的敏感性分析结果，并基于注意力机制结果开展了相应的个性化异常驾驶行为分析。本文提出的个性化异常识别复合模型可以应用于其他类型的驾驶行为，为在复杂多变的交通场景下开展端到端异常行为主动安全防御提供了新思路。下一步研究工作将尝试融合传感数据与视频图像等多源数据，分析异常换道行为发生的时空特征，并探究异常致因；基于自然驾驶数据深度挖掘，揭示我国驾驶员的驾驶行为异质性，并构建典型驾驶风格系列画像。

参考文献：

[1] ELAMRANI ABOU ELASSAD Z, MOUSANNIF H, AL MOATASSIME H,et al. The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review[J/OL]. Engineering Applications of Artificial Intelligence, 2020, 87: 103312.

[2] Bonchek-Dokow E.Cognitive Modeling of Human Intention Recognition[D].RamatGan:Bar Ilan University,2011.

[3] 付锐,张海伦,刘文晓,张洪加.驾驶人意图识别综述[J].长安大学学报(自然科学版),2022,42(01):33-60.FU Rui, ZhANG Hai-lun, LIU Wen-xiao,et al. Review on driver intention recognition [J].Journal of Chang 'an University (Natural Science Edition),2022,42(01):33-60.

[4] 张兰芳,陈程,张佳妍,等.基于自然驾驶数据的高速公路出口换道决策模型[J].同济大学学报(自然科学版),2018,46(3):318.ZHANG Lanfang, CHEN Cheng, ZHANG Jiayan, et al. Modeling Lane-Changing Behavior in Freeway Off-Ramp Areas Using Naturalistic Driving Data[J].Journal of Tongji University(Natural Science) ,2018,46(3):318.

[5] 麻婷婷,涂孝军,朱伟达.基于支持向量机的驾驶员超车意图识别[J].上海工程技术大学学报,2016,30(03):203-208.MA Tingting,TU Xiao-jun,ZHU Wei-da.Driver’s overtaking intention recognition based on support vector machine [J].Journal of Shanghai University of Engineering Science,2016,30(03):203-208.

[6] Hernάndez Luis G, Martinez M O , Ferrάndez José M, et al. EEG-Based Detection of Braking Intention Under Different Car Driving Conditions[J]. Frontiers in Neuroinformatics, 2018, 12:29.

[7] Civik E , Yuzgec U . Deep Learning Based Continuous Real-Time Driver Fatigue Detection for Embedded System[C]// Signal Processing and Communications Applications Conference. IEEE, 2020.

[8] 华强. 混行网联环境下驾驶人分心识别及风险补偿方法研究[D].吉林大学,2022.

[9] Love S , Kannis-Dymand L , Davey J , et al. Metacognition, rumination and road rage: An examination of driver anger progression and expression in Australia[J]. Transportation research, Part F. Traffic psychology and behaviour, 2022(Jan.):84.

[10] 王磊,杨晨煊,路巧珍,朱彤.信息刺激条件下驾驶人警觉性改善规律[J].中国公路学报,2018,31(04):52-58.WANG Lei, YANG Chenxuan, LU Qiaozhen, ZHU Tong.Effect of information stimulus on driver’s vigilance improvement[J].China Journal of Highway and Transport,2018,31(04):52-58.

[11] BOON-GIIN LEE, BOON-LENG LEE, WAN-YOUNG CHUNG. Smartwatch-based driver alertness monitoring with wearable motion and physiological sensor[C/OL]//2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Milan: IEEE, 2015: 6126-6129[2023-03-17].

[12] KARKOUCH A, MOUSANNIF H, AL MOATASSIME H. CADS: A Connected Assistant for Driving Safe[J/OL]. Procedia Computer Science, 2018, 127: 353-359.

[13] 罗裕祥,潘雨帆.基于信息融合的驾驶警觉度识别方法研究[J].综合运输,2020,42(10):71-77.LUO Yu-xiang, PAN Yu-fan. Recognition method of driving alertness based on the fusion of information[J].China Transportation Review, 2020,42 (10) : 71-77.

[14] West C K , Gibson J J . The Senses Considered as Perceptual Systems[J]. Journal of Aesthetic Education, 1969.

[15] 范佳琦,何钢磊,张羽翔,王玉海.考虑车辆间交互作用的驾驶意图预测方法[J].同济大学学报(自然科学版),2021,49(S1):155-161.FAN Jia-qi, HE Gang-lei, ZHANG Yu-xiang, WANG Yu-hai..A driving intention

prediction method considering vehicle interaction[J].Journal of Tongji University (Natural Science), 2021,49 (S1) : 155-161.
[16] 陈勐. 轨迹预测与意图挖掘问题研究[D].山东大学,2016.
[17] Kulic D , Croft E . Estimating Intent for Human-Robot Interaction[J]. Advanced Robotics, 2003.
[18] Durdu A ,İ Erkmen, Erkmen A M.Estimating and reshaping human intention via human-robot interaction[J]. Turkish Journal of Electrical Engineering & Computer Sciences, 2016, 24:88-104.
[19] Noohi E , Efran M . Estimating Human Intention During a Human–Robot Cooperative Task Based on the Internal ForceInternal force Model[J]. Springer International Publishing, 2017.
[20] Munoz-Organero M , Corcoba-Magana V. Predicting Upcoming Values of Stress While Driving[J]. IEEE Transactions on Intelligent Transportation Systems, 2017:1-10.
[21] Gazder U , Assi K J . Determining Driver Perceptions about Distractions and Modeling Their Effects on Driving Behavior at Different Age Groups[J]. Journal of Traffic and Transportation Engineering (English Edition), 2021(3).
[22] 焦方源,李薇.路网状态和驾驶心理驱动的风险规避研究[J].软件工程,2021,24(06):13-16+5.JIAO Fangyuan, LI Wei. Research on risk avoidance driven by road network state and driving psychology [J]. Software Engineering, 2021,24 (06) : 13-16 + 5.
[23] Taylor J E . The extent and characteristics of driving anxiety[J]. Transportation Research Part F: Traffic Psychology and Behaviour, 2018, 58(OCT.):70-79.
[24] Khan M N , Ahmed M M . Trajectory-level fog detection based on in-vehicle video camera with TensorFlow deep learning utilizing SHRP2 naturalistic driving data - ScienceDirect[J]. Accident Analysis & Prevention, 142.
[25] 张洪加,郭应时,高松,刘卓凡.网联与非网联环境下驾驶人换道意图识别研究[J/OL].中国公路学报:1-18[2023-04-03].ZHANG Hong-jia, GUO Ying-shi, GAO Song, LIU Zhuo-fan. Research on driver lane-changing intention recognition in connected and non-connected environments [J / OL]. China Journal of Highway and Transport : 1-18 [2023-04-03].
[26] Fan P, Guo J, Wang Y. A hybrid deep learning approach for driver anomalous lane changing identification[J]. Accident Analysis & Prevention,Vol 171,2022.
[27] Wang S , Zhao X , Yu Q , et al. Identification of Driver Braking Intention Based on Long Short-Term Memory (LSTM) Network[J]. IEEE Access, 2020, 8:180422-180432.
[28] Chen C , Zhao X , Zhang Y , et al. A graphical modeling method for individual driving behavior and its application in driving safety analysis using GPS data[J]. Transportation Research Part F Traffic Psychology and Behaviour, 2019, 63:118-134.
[29] SUN W, AGUIRRE M, JIN J (Judy),et al. Online distraction detection for naturalistic driving dataset using kinematic motion models and a multiple model algorithm[J/OL]. Transportation Research Part C: Emerging Technologies, 2021, 130: 103317.
[30] Michon, John A. critical View of Driver Behavior Models: What do we Know, what should we do?, 1985.
[31] Song R . Driver intention prediction using model-added Bayesian network[J]. Proceedings of the Institution of Mechanical Engineers Part D Journal of Automobile Engineering, 2020:095440702096896.
[32] 郭应时,黄涛.基于注意力机制的多模态自动驾驶行为决策模型[J].中国公路学报,2022,35(09):141-156.GUO Ying-shi, HUANG Tao.Multimodal autonomous driving behavior prediction model based on attention

mechanism [J]. China Journal of Highway and Transport, 2022, 35 (09) : 141-156.
[33] 徐文翔,王俊骅,傅挺.基于注意力机制及分层网络的危险驾驶行为预测方法[J].同济大学学报(自然科学版),2022,50(05):722-730.XU Wenxiang, WANG Junhua, FU Ting.Aggressive driving behavior prediction method based on attention mechanism and hierarchical network [J]. Journal of Tongji University (Natural Science), 2022,50 (05) : 722-730.
[34] Zhang M , Fu R , Morris D D , et al. A Framework for Turning Behavior Classification at Intersections Using 3D LIDAR[J]. IEEE Transactions on Vehicular Technology, 2019, 68(99):7431-7442.
[35] 邓院昌,史晨军.基于贝叶斯结构方程模型的疲劳驾驶行为意图研究[J].安全与环境学报,2019,19(02):520-526.DENG Yuanchang, SHI Chenjun. Psycho-intentional analysis for the factors leading to fatigue driving based on the Bayesian-SEM[J]. Journal of Safety and Environment, 2019,19 (02) : 520-526.
[36] 刘兴亮,朱西产,李霖等. 基于支持向量机的驾驶意图识别[C].LIU Xing-liang,ZHU Xi-chan,LI Lin,et al.Driver intention recognition based on support vector machine[C].Proceedings of the 12th International Forum of Automotive Traffic Safety.2015:157-163.
[37] Xing Y , Lv C , Wang H , et al. Driver Lane Change Intention Inference for Intelligent Vehicles: Framework, Survey, and Challenges[J]. IEEE Transactions on Vehicular Technology, 2019, 68(5):4377-4390.
[38] SUN W, AGUIRRE M, JIN J (Judy),et al. Online distraction detection for naturalistic driving dataset using kinematic motion models and a multiple model algorithm[J/OL]. Transportation Research Part C: Emerging Technologies, 2021, 130: 103317.
[39] GUO H, KEYVAN-EKBATANI M, XIE K. Lane change detection and prediction using real-world connected vehicle data[J/OL]. Transportation Research Part C: Emerging Technologies, 2022, 142: 103785.
[40] Zhang S, Zhi Y , He R , et al. Research on Traffic Vehicle Behavior Prediction Method Based on Game Theory and HMM[J]. IEEE Access, 2020, PP(99):1-1.
[41] Li K , Wang X , Xu Y, et al. Lane changing intention recognition based on speech recognition models[J]. Transportation Research Part C Emerging Technologies, 2016, 69(aug.):497-514.
[42] Xiong X , Chen L , Liang J . A New Framework of Vehicle Collision Prediction by Combining SVM and HMM[J]. IEEE Transactions on Intelligent Transportation Systems, 2018:1-12.
[43] 王新胜,卞震.基于贝叶斯模型的驾驶行为识别与预测[J].通信学报,2018,39(03):108-117.WANG Xinsheng, BIAN Zhen. Driving behavior recognition and prediction based on Bayesian model [J].Iournal of Communications, 2018,39 (03) : 108-117.
[44] Dalsnes B R , Hexeberg S , Flaten A L , et al. The Neighbor Course Distribution Method with Gaussian Mixture Models for AIS-based Vessel Trajectory Prediction[J]. 2018.
[45] Barth A , Franke U . Where will the oncoming vehicle be the next second?[C]// IEEE. IEEE, 2008.
[46] Zheng, Yu. Trajectory Data Mining[J]. Acm Transactions on Intelligent Systems & Technology, 2015, 6(3):1-41.
[47] 仇功达,何明,杨杰,曹玉婷,孙继红.异常轨迹数据预警与预测关键技术综述[J].系统仿真学报,2017,29(11):2608-2617.QIU Gong-da, HE Ming, YANG Jie, CAO Yu-ting, SUN Ji-hong.Key technologies of precaution and prediction of abnormal spatial-temporal trajectory:a review of recent advances [J]. Journal of System Simulation, 2017,29 (11) : 2608-2617.

[48] Faloutsos C , Ranganathan M , Manolopoulos Y . Fast subsequence matching in time-series databases[J]. ACM SIGMOD Record, 1994.
[49] Korn F , Jagadish H V , Faloutsos C . Efficiently supporting ad hoc queries in large datasets of time sequences[J]. ACM SIGMOD Record, 1997.
[50] Chan K , Fu W . Efficient time series matching by wavelets[C]// Proceedings 15th International Conference on Data Engineering (Cat. No.99CB36337). IEEE, 1999.
[51] Cai Y , Ng R . Indexing Spatio-Temporal Trajectories with Chebyshev Polynomials[C]// Acm Sigmod International Conference on Management of Data. ACM, 2004:599.
[52] 孙影,王铁.基于时间序列的目标车辆轨迹预测算法[J].汽车实用技术,2020(06):31-33.SUN Ying, WANG Tie. Target vehicle trajectory prediction algorithm based on time series [J].Automobile Applied Technology, 2020 (06) : 31-33.
[53] Chawla S , Zheng Y , Hu J . Inferring the Root Cause in Road Traffic Anomalies. 2012.
[54] Mao J , Tao W , Jin C , et al. Feature Grouping-Based Outlier Detection Upon Streaming Trajectories[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, PP(99):1-1.
[55] Bu Y , Lei C , Fu W C , et al. Efficient anomaly monitoring over moving object trajectory streams[J]. ACM, 2009.
[56] Yu Y , Cao L , Rundensteiner E A , et al. Detecting Moving Object Outliers In Massive-Scale Trajectory Streams[J]. ACM, 2014:422-431.
[57] Liu Y , Zhao K , Cong G , et al. Online Anomalous Trajectory Detection with Deep Generative Sequence Modeling[C]// 2020 IEEE 36th International Conference on Data Engineering (ICDE). IEEE, 2020.
[58] 毛嘉莉,金澈清,章志刚,周傲英.轨迹大数据异常检测:研究进展及系统框架[J].软件学报,2017,28(01):17-34.Mao Jia-li, JIN Che-qing, ZHANG Zhi-gang, ZHOU Ao-ying. Anomaly detection for trajectory big data : advancements and framework [J].Journal of Software, 2017,28 (01) : 17-34.
[59] 惠飞,彭娜,景首才,周琪,贾硕.基于凝聚层次的驾驶行为聚类与异常检测方法[J].计算机工程,2018,44(12):196-201. HUI Fei, PENG Na, JING Shoucai, ZHOU Qi, JIA Shuo.Driving behavior clustering and abnormal detection method based on agglomerative hierarchy [J]. Computer Engineering, 2018,44 (12) : 196-201.
[60] Palma A T, Bogorny V, Kuijpers B , et al. A clustering-based approach for discovering interesting places in trajectories[J]. DBLP, 2008.
[61] Nanni M , Pedreschi D. Time-focused clustering of trajectories of moving objects[J]. Journal of Intelligent Information Systems 27 (2006): 267-289.
[62] Asahara, Akinori, et al. Pedestrian-Movement Prediction Based on Mixed Markov-Chain Model, ACM, 2011.
[63] Ashbrook D , Starner T . Learning Significant Locations and Predicting User Movement With GPS[C]// Wearable Computers, 2002. (ISWC 2002). Proceedings. Sixth International Symposium on. 2002.
[64] 施雯,刘艳娟,赵彬.基于车辆-行人轨迹预测的人车冲突预警策略研究[J].计算机与数字工程,2022,50(07):1480-1484+1603. SHI Wen, LIU Yanjuan, ZHAO Bin. Research on warning strategy of human-vehicle conflict based on vehicle-pedestrian trajectory prediction [J].Computer and Digital Engineering, 2022,50 (07) : 1480-1484 + 1603.

[65] Gambs S , Killijian M O, DPCM Nú?Ez. Next place prediction using mobility Markov chains[C]// Workshop on Measurement. ACM, 2012:1-6.
[66] Mao J , Tao W , Jin C , et al. Feature Grouping-Based Outlier Detection Upon Streaming Trajectories[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, PP(99):1-1.
[67] Knorr E M , Ng R , Tucakov V . Distance-based outliers: algorithms and applications[J]. The VLDB Journal — The International Journal on Very Large Data Bases, 2000.
[68] Lee J G, Han J,Li X.Trajectory Outlier Detection: A Partition-and-Detect Framework[C]// IEEE International Conference on Data Engineering. IEEE, 2008.
[69] 刘良旭,乔少杰,刘宾,乐嘉锦,唐常杰.基于R-Tree的高效异常轨迹检测算法.软件学报,2009,20(9):2426-2435.LIU Liangxu, QIAO Shaojie, LIU Bin, LE Jiajin, TANG Changjie. Efficient trajectory outlier detection algorithm based on R-Tree. Journal of Software, 2009,20 (9) : 2426-2435
[70] Li X , Han J , Kim S , et al. Roam: Rule- and motif-based anomaly detection in massive moving object data sets[C]// Proceedings of the Seventh SIAM International Conference on Data Mining, April 26-28, 2007, Minneapolis, Minnesota, USA. DBLP, 2007.
[71] 孙吉贵,刘杰,赵连宇.聚类算法研究[J].软件学报,2008(01):48-61.SUN Jigui, LIU Jie, ZHAO Lianyu. Clustering algorithms Research [J]. Journal of Software, 2008 (01) : 48-61.
[72] Dempster A P . Maximum likelihood from incomplete data via the EM algorithm[J]. Journal of the Royal Statistical Society, 1977, 39.
[73] Hinneburg A , Keim D A . An efficient approach to clustering in large multimedia databases with noise[C]// National Conferences on Aritificial Intelligence. AAAI Press, 1998.
[74] 何明,仇功达,周波,柳强,曹玉婷.基于改进密度聚类与模式信息挖掘的异常轨迹识别方法[J].通信学报,2017,38(12):21-33.HE Ming,QIU Gongda,ZHOU Bo,LIU Qiang, CAO Yuting. Abnormal trajectory detection method based on enhanced density clustering and abnormal information mining [J].Journal of Communications, 2017,38 (12) : 21-33.
[75] Ge Y , Xiong H , Zhou Z , et al. Top-Eye: top-k evolving trajectory outlier detection.[C]// ACM. ACM, 2010.
[76] Chen C , Zhang D , Castro P S , et al. Real-Time Detection of Anomalous Taxi Trajectories from GPS Traces[C]// Mobile and Ubiquitous Systems: Computing, Networking, and Services. 8th International ICST Conference, MobiQuitous 2011, Copenhagen, Denmark, December 6-9, 2011, Revised Selected Papers. 2012.
[77] Pang L X , Chawla S , Liu W, et al. On mining anomalous patterns in road traffic streams[C]// Advanced Data Mining and Applications. Springer, Berlin, Heidelberg, 2011.
[78] 贾惠迪. 基于网格区域交通状态的城市路网行程时间预测研究[D].北京交通大学,2021.
[79] Zhang D , Nan L , Zhou Z H, et al. iBAT: Detecting anomalous taxi trajectories from GPS traces[C]// UbiComp 2011: Ubiquitous Computing, 13th International Conference, UbiComp 2011, Beijing, China, September 17-21, 2011, Proceedings. 2011.
[80] LV J, LI Q, SUN Q, et al. T-CONV: A Convolutional Neural Network for Multi-scale Taxi Trajectory Prediction[C/OL]//2018 IEEE International Conference on Big Data and Smart Computing (BigComp). Shanghai: IEEE, 2018: 82-89[2023-03-18].
[81] Chiang M F , Lin Y H , Peng W C , et al. Inferring distant-time location in low-sampling-rate trajectories[C]//

Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013.
[82] HENNING M, MULLER J, GIES F, et al. Situation-Aware Environment Perception Using a Multi-Layer Attention Map[J/OL]. IEEE Transactions on Intelligent Vehicles, 2023, 8(1): 481-491.
[83] Akiyama T , Inokuchi H . Long term estimation of traffic demand on urban expressway by neural networks[C]// 2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS). IEEE, 2014.
[84] 陆慧敏,杨朔.基于深度神经网络的自动驾驶场景三维目标检测算法[J].北京工业大学学报,2022,48(06):589-597.LU Huimin, YANG Shuo. 3D object detection algorithm based on deep neural network for autonomous driving[J].Journal of Beijing University of Technology, 2022,48 (06) : 589-597.
[85] WANG Y, HUANG H, ZHANG B, et al. A differentiated decision‐making algorithm for automated vehicles based on pedestrian feature estimation[J/OL]. IET Intelligent Transport Systems, 2023: itr2.12335.
[86] GUO H, KEYVAN-EKBATANI M, XIE K. Lane change detection and prediction using real-world connected vehicle data[J/OL]. Transportation Research Part C: Emerging Technologies, 2022, 142: 103785.
[87] GUO J, LIU Y, ZHANG L, et al. Driving Behaviour Style Study with a Hybrid Deep Learning Framework Based on GPS Data[J/OL]. Sustainability, 2018, 10(7): 2351.
[88] HUANG W, LIU X, LUO M, et al. Video-Based Abnormal Driving Behavior Detection via Deep Learning Fusions[J/OL]. IEEE Access, 2019, 7: 64571-64582.
[89] Zagoruyko S , Komodakis N . Wide Residual Networks[J]. 2016.
[90] Xie S , Girshick R , P Dollár, et al. Aggregated Residual Transformations for Deep Neural Networks[J]. IEEE, 2016.
[91] 杜绎如,马印怀,吴建波,惠飞,阮仕峰,郭星.基于LSTM-att的车辆异常驾驶行为识别[J].计算机系统应用,2022,31(05):165-173. DU Yi-ru, MA Yin-huai, WU Jian-bo, HUI Fei, RUAN Shi-feng, GUO Xing. Recognition of vehicle abnormal driving behaviors based on LSTM-att [J]. Computer system application, 2022,31 (05) : 165-173.
[92] Huang G, Liu Z, Laurens V,et al. Densely Connected Convolutional Networks[J]. IEEE Computer Society, 2016.
[93] He, Kaiming, X. Zhang, Shaoqing Ren and Jian Sun. Deep Residual Learning for Image Recognition.2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015): 770-778.
[94] YUAN M,SHAN J,MI K.Deep Reinforcement Learning Based Game-Theoretic Decision-Making for Autonomous Vehicles[J/OL]. IEEE Robotics and Automation Letters, 2022, 7(2): 818-825.
[95] Zeng W , Liang M ,Liao R , et al. LaneRCNN: Distributed Representations for Graph-Centric Motion Forecasting[J]. 2021.
[96] Gao J , Sun C , Zhao H , et al. VectorNet: Encoding HD Maps and Agent Dynamics from Vectorized Representation[C]// 2020.
[97] Zhao H , Gao J , Lan T , et al. TNT: Target-driveN Trajectory Prediction[J]. 2020.
[98] Phan-Minh T , Grigore E C , Boulton F A , et al. CoverNet: Multimodal Behavior Prediction using Trajectory Sets[J]. 2019.
[99] Alahi A , Goel K , Ramanathan V, et al. Social LSTM: Human Trajectory Prediction in Crowded Spaces[C]//

2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.
[100] 贾寒冰,刘鹏,张雷等.基于规则与机器学习融合的换道决策建模方法研究[J].机械工程学报,2022,58(04):212-221.JIA Hanbing, LIU Peng, ZHANG Lei,et al. Lane-change decision model development by combining rules abstract and machine learning technique [J].Journal of Mechanical Engineering, 2022,58 (04) : 212-221.
[101] 赵建东,贺晓宇,余智鑫等.多网联范围下的智能网联车换道决策组合模型研究[J].交通运输系统工程与信息,2023,23(01):77-85.ZHAO Jiandong, HE Xiaoyu, YU Zhixin, et al. A combination model for connected and autonomous vehicles lane-changing decision-making under multi connectivity range[J].Journal of Transportation System Engineering and Information Technology, 2023,23 (01) : 77-85.
[102] YUAN M, SHAN J, MI K. Deep Reinforcement Learning Based Game-Theoretic Decision-Making for Autonomous Vehicles[J/OL]. IEEE Robotics and Automation Letters, 2022, 7(2): 818-825.
[103] Du Y ,Wang Y, Chan CY . Autonomous lane-change controller via mixed logical dynamical[C]// IEEE International Conference on Intelligent Transportation Systems. IEEE, 2014.
[104] Tehrani H , Do Q H, Egawa M,et al.General Behavior and Motion Model for Automated Lane Change[J]. IEEE, 2015.
[105] Li K, Wang X,Xu Y , et al. Lane changing intention recognition based on speech recognition models[J]. Transportation Research Part C Emerging Technologies, 2016, 69(aug.):497-514.
[106] Zheng,Yu. Trajectory Data Mining[J]. Acm Transactions on Intelligent Systems & Technology, 2015, 6(3):1-41.
[107] Bu Y , Lei C , Fu WC, et al. Efficient anomaly monitoring over moving object trajectory streams[J]. ACM, 2009.
[108] Yu Y ,Cao L, Rundensteiner E A , et al. Detecting Moving Object Outliers In Massive-Scale Trajectory Streams[J]. ACM, 2014:422-431.
[109] Liu Y,Zhao K,Cong G, et al. Online Anomalous Trajectory Detection with Deep Generative Sequence Modeling[C]// 2020 IEEE 36th International Conference on Data Engineering (ICDE). IEEE, 2020.
[110] Gambs S , Killijian M O, DPCM Nú?Ez. Next place prediction using mobility Markov chains[C]// Workshop on Measurement. ACM, 2012:1-6.
[111] Chiang MF,Lin YH,Peng W C,et al. Inferring distant-time location in low-sampling-rate trajectories[C]// Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2013.
[112] MA Y, XIE Z,CHEN S, et al. Real-time detection of abnormal driving behavior based on long short-term memory network and regression residuals[J/OL]. Transportation Research Part C: Emerging Technologies, 2023, 146: 103983.
[113] YI D.,SU J.,LIU C.,et al. A machine learning based personalized system for driving state recognition[J]. Transportation Research Part C: Emerging Technologies, 2019, 105:241.
[114] 季学武,费聪,何祥坤,等.基于LSTM网络的驾驶意图识别及车辆轨迹预测[J].中国公路学报,2019,32(6):34.JI Xuewu, FEI Cong, HE Xiangkun, et al. Intention Recognition and Trajectory Prediction for Vehicles Using LSTM Network[J]. China Journal of Highway and Transport,2019,32(6):34.

[115] 韩皓,谢天.基于注意力Seq2Seq网络的高速公路交织区车辆变道轨迹预测[J].中国公路学报,2020,33(6):106.HAN Hao, XIE Tian. Lane Change Trajectory Prediction of Vehicles in Highway Interweaving Area Using Seq2Seq-attention N[J]. China Journal of Highway and Transport,2020,33(6):106.

[116] BAHDANAU D, CHO K, BENGIO A. Neural Machine Translation by Jointly Learning to Align and Translate, 2015.arXiv preprint:1409.0473.

[117] ZHANG Jun，WU Zhongcheng，LI Fang，et al. Attention-Based Convolutional and Recurrent Neural Networks for Driving Behavior Recognition Using Smartphone Sensor Data[J]. IEEE Access, 2019,7:148031.

[118] Do L N N, Vu H L, Vo B Q, et al. An effective spatial-temporal attention based neural network for traffic flow prediction[J]. Transportation Research Part C: Emerging Technologies, 2019,108: 12.

[119] LIU F T, TING K, ZHOU Z. Isolation Forest. Presented at in ICDM 2008[C]//IEEE, Eighth IEEE International Conference on Data Mining. Pisa: IEEE,2009:1.

[120] Boquet G, Morell A, Serrano J, et al. A variational autoencoder solution for road traffic forecasting systems: Missing data imputation, dimension reduction, model selection and anomaly detection[J]. Transportation Research Part C: Emerging Technologies, 2020,115: 102622.

[121] Yang M, Wang X, Quddus M. Examining lane change gap acceptance, duration and impact using naturalistic driving data[J]. Transportation Research Part C: Emerging Technologies, 2019, 104: 317.

[122] Asahara, Akinori, et al. Pedestrian-Movement Prediction Based on Mixed Markov-Chain Model, ACM, 2011.

[123] Ashbrook D , Starner T . Learning Significant Locations and Predicting User Movement With GPS[C]// Wearable Computers, 2002. (ISWC 2002). Proceedings. Sixth International Symposium on. 2002.

[124] 施雯,刘艳娟,赵彬.基于车辆-行人轨迹预测的人车冲突预警策略研究[J].计算机与数字工程,2022,50(07):1480-1484+1603. SHI Wen, LIU Yanjuan, ZHAO Bin. Research on warning strategy of human-vehicle conflict based on vehicle-pedestrian trajectory prediction [J].Computer and Digital Engineering, 2022,50 (07) : 1480-1484 + 1603.

[125] Gambs S , Killijian M O, DPCM Nú?Ez. Next place prediction using mobility Markov chains[C]// Workshop on Measurement. ACM, 2012:1-6.

[126] 李绍兵,胡晓辉,洪鹏.基于贝叶斯博弈的自动驾驶车辆换道预测模型[J/OL].LI Shaobing, HU Xiaohui, HONG Peng. Lane-changing prediction model of autonomous vehicle based on Bayesian game.Journal of Measurement Science and Instrumentation:1-14[2023-04-17].

[127] Hou Y , Edara P , Sun C . Situation assessment and decision making for lane change assistance using ensemble learning methods[J]. Expert Systems with Applications, 2015, 42(8):3875-3882.

[128] Ulbrich S , Maurer M. Towards Tactical Lane Change Behavior Planning for Automated Vehicles[C]// 2015 IEEE 18th International Conference on Intelligent Transportation Systems - (ITSC 2015). IEEE, 2015.

[129] 乔少杰,韩楠,朱新文,舒红平,郑皎凌,元昌安.基于卡尔曼滤波的动态轨迹预测算法[J].电子学报,2018,46(02):418-423.QIAO Shaojie, HAN Nan, ZHU Xinwen, SHU Hongping, ZHENG Jiaoling, YUAN Changan.A dynamic trajectory prediction algorithm based on Kalman filter[J].Acta Electronica Sinica, 2018,46 (02) : 418-423.

[130] Xiong Z , Ren S , Liu C , et al. Chapter 29 A Carrier Tracking Algorithm of Kalman Filter Based on

Combined Maneuvering Target Model[C]// Chinese Intelligent Systems Conference. 2015.

[131] Cartlidge J , Gong S, Bai R , et al. Spatio-temporal Prediction of Shopping Behaviours Using Taxi Trajectory Data[C]// 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA). IEEE, 2017.

[132] Jeung H , Liu Q ,Shen H T , et al. A Hybrid Prediction Model for Moving Objects[C]// IEEE International Conference on Data Engineering. IEEE, 2008.

[133] 高建,毛莺池,李志涛.基于高斯混合-时间序列模型的轨迹预测[J].计算机应用,2019,39(08):2261-2270.GAO Jian, MAO Yingchi, LI Zhitao. Trajectory prediction based on Gauss mixture time series model [J].Journal of Computer applications, 2019,39 (08) : 2261-2270.

[134] Zheng Y , Liu F , Hsieh H P. U-Air: when urban air quality inference meets big data[C]// Acm Sigkdd International Conference on Knowledge Discovery & Data Mining. ACM, 2013.