

基于视线估计的驾驶员分心状态检测方法

杨思翰 何 青 杨世琦

(长沙理工大学人工智能学院 长沙 415000)

摘要: 随着智能驾驶技术的发展,驾驶员注意力监测已成为行车安全保障的重要研究方向。传统分心检测依赖离散行为分类,难以刻画注意力的连续变化过程,同时忽视驾驶员在生理结构及注视习惯方面的个体差异。为此,提出了一种基于视线估计的分心状态检测方法。该方法通过计算视线欧拉角相对于安全注视区域的偏离程度,实现分心状态的连续识别。构建了 ResNeXt-Transformer 混合网络,融合局部细节与全局依赖关系,并引入特征增强模块以强化任务相关表征,从而提升视线特征提取能力。同时,提出多层次个性化校准机制,在特征层与输出层分别进行偏置补偿,实现模型在少样本条件下的快速个体化适配。实验结果表明,本方法在 MPIIGaze 数据集上跨用户视线估计任务中的平均角度误差降至 2.94° ,显著优于现有方法。结合可视化与实际驾驶场景验证,证明该方法具有良好的可解释性与应用潜力。

关键词: 视线估计;分心驾驶检测;深度学习;个性化

中图分类号: TP391;TN919.8 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Gaze estimation-based driver distraction detection method

Yang Sihan He Qing Yang Shiqi

(School of Artificial Intelligence, Changsha University of Science and Technology, Changsha 415000, China)

Abstract: With the rapid development of intelligent driving technology, driver attention monitoring has become a crucial research focus for ensuring road safety. Traditional distraction detection methods rely on discrete behavior classification, which fails to capture the continuous variation of driver attention and often neglects individual differences in physiological structures and gaze habits. To address these limitations, this paper proposes a gaze estimation-based distraction detection method that continuously identifies distraction states by evaluating the deviation of gaze Euler angles from a predefined safe attention region. The proposed model adopts a ResNeXt-Transformer hybrid architecture that integrates local detail extraction with global dependency modeling, and incorporates a feature enhancement module to strengthen task-relevant representations and improve gaze feature extraction. Furthermore, a multi-level personalized calibration mechanism performs bias compensation at both the feature and output layers, enabling rapid individual adaptation under few-shot conditions. Experimental results on the MPIIGaze dataset demonstrate that the proposed method significantly outperforms existing approaches in cross-user gaze estimation, reducing the average angular error to 2.94° . Visualization and validation in real-world driving scenarios further confirm that the method exhibits strong interpretability and high practical applicability.

Keywords: gaze estimation;distraction driving detection;deep learning;personalization

0 引言

近年来,随着汽车保有量的快速增长,交通安全形势愈发严峻。大量统计研究表明,驾驶员注意力分散是导致交通事故的核心因素之一。美国国家公路交通管理局 2023 年公布的最新数据显示,2021 年因分心驾驶导致的交通事故共造成 3 522 人死亡,相较 2020 年增加了 12.1%^[1]。当

驾驶员视线脱离道路时,反应时间显著延长,碰撞风险随之急剧上升。因此,高效、准确地监测驾驶员注意状态对于智能驾驶与主动安全系统具有关键意义。

现有分心检测方法主要可分为 3 类^[2]:基于生理信号、基于车辆行为数据以及基于面部特征。基于生理信号(如脑电图、心电图等)可直接表征认知状态,但其设备昂贵且易干扰驾驶,实际部署成本高。基于车辆行为数据(如方向

盘角度、汽车加速度等)受司机驾驶习惯和外界环境影响较大,难以实时决策,泛化性较弱。相比之下,基于面部视觉特征的方法因其非接触、易部署以及与驾驶行为之间的直接关联性,近年来成为主流发展方向之一。

随着计算机技术的飞速发展,深度学习方法只需通过低成本、易于部署的摄像头实时采集驾驶员的面部表情、头部姿态以及车内外环境等信息,再通过卷积神经网络(convolutional neural networks, CNN)等进行特征提取与高层语义建模,从而实现对驾驶员分心状态的自动识别与分类。冯笑等^[3]通过浅层 CNN 提取单通道脑电特征,引入自适应特征重校准机制增强特征判别性,并结合 LSTM 建模时序信息,从而在跨被试场景下实现了高效且具有神经生理解释性的疲劳状态识别。赵桂峰等^[4]提出了一种基于深度学习的驾驶员分心行为检测方法,利用高阶空间交互网络(high-order spatial interaction network, HorNet)的残差结构和自适应加权注意策略来提高分心行为识别的精度。殷旭鹏等^[5]在 YOLOv10 框架基础上通过对网络结构进行轻量化改进,并结合注意力机制与多尺度特征融合策略,有效提升了检测精度与实时性能。Subbulakshmi 等^[6]通过集成多种 CNN 模型构建分心驾驶检测系统,充分利用各模型的互补优势,实现高精度分类性能。

然而, CNN 模型固有的局部感受野特性,使其在捕捉视线特征,建模全局空间依赖关系方面存在显著瓶颈^[7]。另外,上述方法通常将分心状态定义为驾驶员执行特定动作(如打电话、玩手机或转头交流等)并将其定义为多类别动作识别任务。然而基于行为分类的方法难以准确反映驾驶注意力的连续变化过程,实际驾驶场景中,短暂且细微的视线偏移更为常见,却常因未被定义为特定危险行为而漏检,交通场景的复杂性也对驾驶员状态感知系统提出了更高要求。许雨梅等^[8]针对自动驾驶场景中三维目标检测在远距离目标稀疏及小目标感知能力不足的问题,提出了一种基于生成式上采样的三维目标检测方法 PUDet,通过特征层的补全与增强提升环境感知的可靠性。但其研究重点集中于车外信息建模,对驾驶员是否感知并关注关键目标尚缺乏直接刻画。相比之下,视线作为注意力焦点最直接生理特征,可实现分心程度与空间偏离量的定量关联,具备更高的可解释性和泛化能力。

近年来,基于外观式视线估计取得快速发展^[9-10]。Chen 等^[11]提出一种低复杂度的外观基础视线估计校准方法,基于视线分解的思想,能够在较小校准数据集上实现高精度估计。Guan 等^[12]提出 MCGaze 方法,整合了头部、面部和眼部特征的时空交互,通过综合全局和局部特征,优化了视线估计精度。Jin 等^[13]提出了一种考虑眼球反向旋转的 Kappa 角回归方法,通过结合真实与合成眼图进行校准,提升了视线估计的精度与稳定性。值得注意的是,已有研究开始尝试在智能车辆系统中引入驾驶员个体特性。张冰波等^[14]在再生制动控制中综合考虑车辆行驶工况、驾驶

员特性及道路信息,通过对驾驶风格进行建模,实现了控制策略对不同驾驶员的自适应优化,体现了将驾驶员差异纳入智能车辆决策与控制过程的研究趋势。尽管如此,当前大多数驾驶员分心状态检测方法还是普遍忽略了不同驾驶员之间存在的天然个体差异,如眼球结构、面部特征、注视习惯等,统一模型难以兼顾所有个体的特征分布,导致精准度难以进一步提高^[15-18]。

基于上述文献存在的问题,本文提出了一种基于视线估计的驾驶员分心检测模型,通过设定驾驶员视线偏离前方安全区域,在无需具体行为标签的情况下,实现对驾驶注意力连续变化的感知。为了增强模型对视线特征信息的捕捉能力,提出 ResNeXt(residual networks with next)与 Transformer 编码器混合结构设计,融合眼部纹理细节与面部几何结构,再通过注意力机制对视线特征进行增强,从而使模型在复杂注视姿态变化下能保持预测稳定性。为减少驾驶员个体生理差异对视线估计的影响,提出少样本个性化适配策略,通过个性化校准学习不同个体的特定参数,利用少量数据快速修正用户特定偏差,实现跨用户迁移性能提升。

1 基于视线估计的模型方法

1.1 总体框架

在视线估计任务中,驾驶员眼部的纹理变化与头部姿态的空间变化共同影响视线方向。卷积网络擅长捕捉局部纹理特征,而自注意力机制能够捕获全局依赖关系。为此,本文将 ResNeXt 与 Transformer 编码器结合,首先利用 ResNeXt 提取眼部区域纹理特征,随后通过视线特征增强模块(gaze feature enhancement module, GFEM)对视线敏感维度进行强化,最后利用 Transformer 编码器聚合全局上下文信息,从而增强模型对面部几何差异与头部姿态变化的鲁棒性。在此基础上,本文提出多层次个性化校准机制,包括特征层的个性化特征调制模块(personalized feature modulation, PFM)与输出层的个体偏差对齐模块(individual bias alignment, IBA),分别用于减少外观差异引起的特征漂移与补偿视线几何偏移。同时,引入一致性约束策略约束多层次个性化分支在特征空间与输出空间的预测一致性。该结构协同实现跨用户泛化能力与个性化精度的统一优化。模型的详细结构如图 1 所示。

1.2 ResNext 网络结构

ResNet 残差网络通过引入残差连接有效缓解了深层网络中的梯度消失问题,并能高效地提取局部视觉特征。ResNeXt 网络在 ResNet 的基础上进一步引入了分组卷积(group convolution)与基数(cardinality)的概念,其中一个典型残差块如图 2 所示。

ResNeXt 网络将输入特征划分为多个并行分支进行处理,每个分支执行具有相同拓扑结构的卷积变换,然后将各分支的输出进行聚合。与 ResNet 网络相比,这种“分组聚合”的结构设计在保持网络拓扑一致的前提下,能够更有

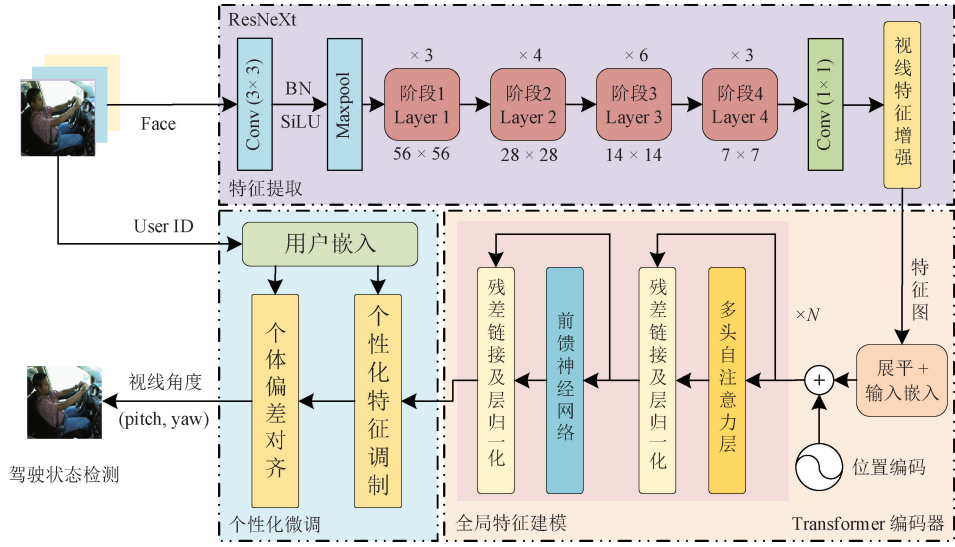


图 1 驾驶员分心状态检测模型整体框架

Fig. 1 Overall framework of driver distraction state detection model

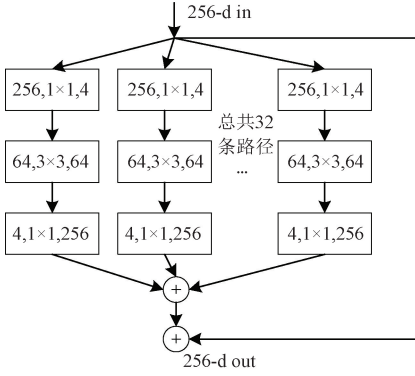


图 2 ResNeXt 网络基础残差块

Fig. 2 ResNeXt network basic residual block

效捕捉眼部和面部区域的多尺度特征,并增强网络对不同姿态下的视线方向变化的适应能力,同时简化了超参数设计与网络扩展过程。其聚合变换可表示为:

$$y = x + \sum_{i=1}^G T_i(x) \quad (1)$$

其中, $T_i(\cdot)$ 表示具有相同拓扑结构的并行支路, $G=32$, 表示分支数。该结构提高了对眼部肌肉纹理、瞳孔显现等局部变化的感知力,为后续全局建模提供信噪比更高的表达。

输入图像经过 ResNeXt 主干网络后得到尺寸为 $F \in \mathbf{R}^{7 \times 7 \times 512}$ 的空间特征图。为降低 Transformer 的计算复杂度,使用 1×1 卷积将通道压缩至 32 维,形成更紧凑的特征表达 $F' \in \mathbf{R}^{7 \times 7 \times 32}$, 实现了特征维度与 Transformer 输入用户嵌入向量维度的对齐。

1.3 视线特征增强模块

Transformer 的全局敏感性会引入许多与视线估计无关的结构信息。为提升视线敏感维度权重,本文在编码器

之后引入 GFEM,通过归一化、非线性激活与两层 1×1 卷积构成的残差结构,用于强化瞳孔显现和眼睑轮廓等关键区域。设输入特征为 $x \in \mathbf{R}^{B \times C \times H \times W}$, 首先对其进行归一化与非线性变换:

$$\hat{x} = \text{ReLU}(\text{BN}(x)) \quad (2)$$

随后利用两层 1×1 卷积实现通道映射与非线性特征变换:

$$z = \text{ReLU}(W_1 * \hat{x}) \quad (3)$$

$$h = W_2 * z \quad (4)$$

其中, W_1, W_2 表示 1×1 卷积核, $*$ 表示卷积操作,最终通过残差连接获得精化后的输出为:

$$\tilde{x} = x + h \quad (5)$$

通过残差连接结构将增强特征 h 加回到原特征中,能抑制光照变化和人脸结构差异带来的冗余响应,从而提高视线方向回归的稳定性与泛化能力。

1.4 Transformer 编码器

在视线估计任务中,不同面部区域对视线方向的贡献并不均等。如瞳孔与眼部区域高度相关,而脸颊、额头及背景多为干扰信息。传统卷积网络因其固定的局部感受野,难以自适应地区分这些区域的重要性。此外视线估计是回归任务而非序列生成任务,只需要从输入图像中提取关键特征并捕捉全局关系,因此本文采用 Transformer 编码器对 ResNeXt 与 GFEM 提取的特征序列进行全局语义聚合。该编码器利用多头自注意力机制 (multi-head attention, MHA) 建模眼部与面部区域间的空间依赖关系,并自适应地为各区域分配差异化权重,从而强化对视线敏感的关键特征。Transformer 编码器结构如图 3 所示。

增强后的输入特征 $\tilde{x}$ 首先被展平为长度为 $H \times W$ 的

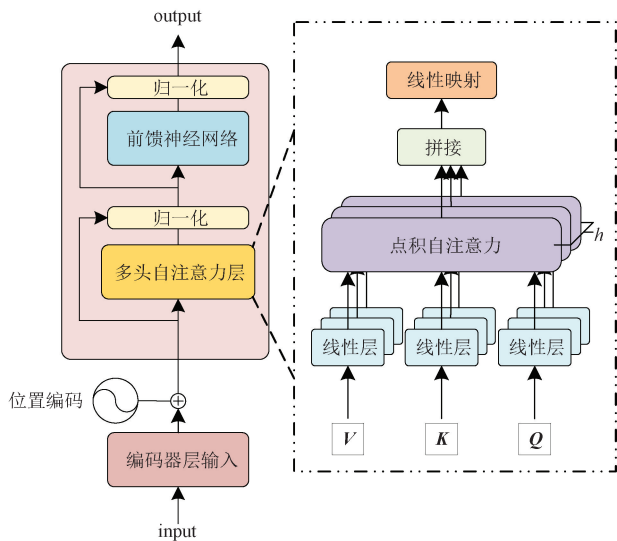


图 3 Transformer 编码器结构

Fig. 3 Transformer encoder structure

token 序列,并在前端加入一个全局可学习变量[CLS] token,通过位置编码保留卷积特征的空间结构。各 token 经线性变换得到查询(Q)、键(K)和值(V)向量。模型通过学习为输入序列中的每个特征 token 生成一组注意力权重 $\{a_{ij}\}$ 。这些权重动态地决定了在聚合全局信息时,其他每个区域对当前区域的重要性,以此鲁棒地融合局部细节与全局几何结构,并通过[CLS]token 汇聚成一个信息丰富的全局视线表征。其核心计算为:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \quad (6)$$

其中, d_k 为查询特征维度。多层 Transformer 堆叠使模型能够自动强化对视线敏感区域(如瞳孔、眼裂边缘)的关注,并通过[CLS]token 汇聚最终的全局表征,实现稳健的视线方向回归。

1.5 多层次个性化校准

驾驶员在面部外观(如结构、肤色、眼形)和注视习惯上存在显著的个体差异,这严重影响了通用视线估计模型的预测精度。为此,本文提出多层次个性化校准机制,由特征层的 PFM 与输出层的 IBA 组成,并结合一致性正则化约束构建联合优化过程。PFM、IBA 与一致性约束这 3 个部分,分别从特征表示、输出偏置与预测稳定性 3 个方向进行补偿,从而提升跨用户泛化能力与个体专属性能。

1) 个性化特征调制

不同驾驶员在外观层面的差异(如眼球形态、面部结构等)会造成特征分布偏移。同样的注视方向,不同用户提取到的视觉特征不同,导致模型无法直接对齐。本文在 Transformer 编码器隐空间引入 PFM 模块,通过融合用户嵌入向量与全局特征,将其送入一个轻量级的两层感知机,以学习用户相关的特征残差偏移量,从而实现个体差异建模:

$$\mathbf{h}' = \mathbf{h} + \mathbf{W}_2(\sigma(\mathbf{W}_1[\mathbf{h}, \mathbf{z}])) \quad (7)$$

其中, $\mathbf{h} \in \mathbf{R}^C$ 为全局特征表示, $\mathbf{z} \in \mathbf{R}^C$ 为用户嵌入向量, $\mathbf{W}_1 \in \mathbf{R}^{(C+d) \times C}$ 、 $\mathbf{W}_2 \in \mathbf{R}^{C \times C}$ 为可学习权重, $\sigma(\cdot)$ 为非线性激活函数。

为了保证训练稳定性,最后一层线性映射 $\mathbf{W}_2$ 被初始化为 0,使得模块初始时近似恒等映射,随后在微调过程中逐渐学习用户特定残差。

2) 个体偏差对齐

不同驾驶员即使注视同一点,其头部姿态、眼球结构不同,导致预测值存在系统性的几何偏差。为此,本文进一步提出 IBA 模块,在输出空间对模型预测结果进行用户相关仿射校准:

$$\mathbf{g}' = \mathbf{A}_u \mathbf{g} + \mathbf{b}_u \quad (8)$$

其中, $\mathbf{g} = [\text{yaw}, \text{pitch}]$ 为未校准的预测视线, $\mathbf{A}_u \in \mathbf{R}^{2 \times 2}$ 、 $\mathbf{b}_u \in \mathbf{R}^2$ 为用户相关校准参数。为了避免过度修正,训练过程中对 $\mathbf{A}_u$ 施加正则化约束,使其接近单位矩阵。

3) 特征一致性约束

同一驾驶员在不同姿态、光照、摄像角度下,特征表示可能不稳定,即使注视方向相同,模型提取的中间特征也可能发生漂移,进而影响预测的稳定性。为此,本文在特征层引入一致性约束机制,使同一用户在视线相似的情况下保持潜在特征表示一致。

模型输出中间层特征表示为 $\mathbf{r} \in \mathbf{R}^C$ 。定义样本对集合为:

$$P = \{(i, j) \mid \text{id}(i) = \text{id}(j), \angle(\dot{\mathbf{y}}_i, \dot{\mathbf{y}}_j) < \tau\} \quad (9)$$

其中, τ 为阈值,用于筛选视线方向相近的样本对。则一致性损失定义为:

$$L = L_{\text{reg}} + \lambda \left[\frac{1}{|P|} \sum_{(i,j) \in P} (1 - \cos(\mathbf{r}_i, \mathbf{r}_j)) \right] \quad (10)$$

特征余弦相似度定义为:

$$\cos(\mathbf{r}_i, \mathbf{r}_j) = \frac{\mathbf{r}_i \cdot \mathbf{r}_j}{\|\mathbf{r}_i\| \|\mathbf{r}_j\|} \quad (11)$$

其中, L_{reg} 为视线回归损失, λ 为一致性约束的权重,通过最大化同一用户在不同外观条件下特征表征的一致性,可显著减少因外观干扰因素导致的特征漂移。

2 实验与结果分析

本文视线估计模型在搭载 NVIDIA GeForce RTX 4060、Intel Core i9-13900KF 及 32 GB 内存的主机上进行实验,基于 PyTorch 实现。使用 AdamW 优化器,初始学习率设为 5×10^{-4} ,批次大小 16,总训练 80 轮次,前 5 轮采用线性学习率预热(warmup),之后每 40 轮衰减 50%。损失函数为 L1 回归损失。

2.1 数据预处理

本文选用 MPIIGaze 公开数据集作为实验数据来源,该数据集包含 15 名受试者,共计约 45 000 张人脸图像。为提升模型训练效率与特征提取质量,本文对原始数据进

行了系统的预处理。首先,依据人脸关键点信息对图像进行裁剪与几何对齐,精确提取人脸与双眼区域;其中右眼图像通过水平翻转操作进行对称化处理,并相应修正其视线标签。随后,对所有图像执行归一化与直方图均衡化操作,以减少光照变化对模型的影响。最后,将图像尺寸统一调整为 224×224 ,以适配模型输入需求。

此外,为验证分心检测可行性,使用 DMD 驾驶员数据集进行分心状态判断对比实验以及部分可视化展示。该数据集采集自真实驾驶舱及模拟舱环境,更具实际驾驶场景代表性。为验证模型在分心检测任务上的可行性,从该数据集中筛选出包含“发短信”、“打电话”、“取东西”等典型分心行为的图像作为“分心”类,并选取正常驾驶图像作为“专心”类,构建一个二分类数据集进行对比实验。两数据集使用目的在任务上保持独立。

2.2 实验设置及模型评估指标

为全面评估模型的视线估计能力,实验设计采用留一人交叉验证(leave one person out, LOPO)策略,在每一折中,将一名受试者的全部数据作为测试集,其余所有受试者的数据按 9:1 的比例划分为训练集和验证集。除基准模型对比外,本文还开展了消融实验以验证各模块的有效性。在个性化微调实验中,从测试用户的数据中随机抽取 k 个样本构成校准集用于模型微调,该用户的其余数据作为测试集。校准集与测试集互斥,确保评估公正。所有实验均固定随机种子并重复 5 次取平均值,以保证结果可靠性。

本文采用视线估计领域中最常用且公认的角度误差进行评估,用于衡量预测视线向量与真实视线向量之间的夹角大小。设真实视线方向为 $\mathbf{g} \in \mathbf{R}^3$, 预测方向为 $\mathbf{g}' \in \mathbf{R}^3$, 则角度误差的计算公式为:

$$L_{\text{angular}} = \arccos\left(\frac{\mathbf{g} \cdot \mathbf{g}'}{\|\mathbf{g}\| \|\mathbf{g}'\|}\right) \quad (12)$$

角度误差能够直观反映模型对视线方向的估计精度,误差越小,明模型预测的视线方向与真实方向越接近,从而在驾驶员状态识别任务中具备更高的可靠性。

本文提出的驾驶员分心状态检测方法与其他方法对比时,采用准确率(accuracy, Acc)作为模型的评价指标,其计算公式如下:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \quad (13)$$

其中,TP 指实际类别为正类,且被分类器正确预测为正类的样本实例数量;FN 指实际类别为正类,但被分类器错误预测为负类的样本实例数量;FP 指实际类别为负类,但被分类器错误预测为正类的样本实例数量;TN 指实际类别为负类,且被分类器正确预测为负类的样本实例数量。

2.3 对比实验

为全面评估本文所提出的视线估计模型性能,选取目前具有代表性的 3 种方法进行对比,包括 PureGaze^[19]、

GazeTR^[20]与 DGE-GM^[21],在 MPIIGaze 数据集上采用相同设置进行训练与测试。PureGaze 通过抑制光照与身份等与视线无关的干扰因素提升跨域泛化能力;GazeTR 首次采用 ResNet 与 Transformer 混合编码结构以增强局部与全局语义建模;DGE-GM 在卷积主干中引入特征增强模块并结合混合损失以提升特征表达能力与鲁棒性。然而,上述方法均没有考虑不同用户之间固有的生理差异以及注视习惯差异,在进行跨用户测试时仍然会导致误差逐渐累积。相比之下,本文提出的基于 ResNeXt-Transformer 混合主干,引入 GFEM、PFM 与 IBA 实现特征精化与个体偏差补偿,使模型在未进行个性化校准时已具备较高跨用户泛化性能,并在小样本微调后取得进一步提升。各模型方法性能对比结果如表 1 所示。

表 1 不同模型性能对比

Table 1 Performance comparison of different models (%)

模型	角度误差
puregaze	5.20
GazeTR	4.00
DGE-GM	3.76
本文模型(微调前)	4.27
本文模型(微调后)	2.94

由表 1 可知,本文方法在跨用户视线估计任务中取得了最低的角度误差,表现最优。在未使用任何个性化数据的情况下,本文模型的角度误差为 4.27° ,较 PureGaze (5.20°)降低了 17.9%。这一结果验证了本文提出的 ResNeXt-Transformer 混合编码结构及 GFEM 模块,能够有效学习对用户身份不敏感、对视线方向敏感的特征表示,从而提升了跨用户的泛化能力。进一步加入 PFM 和 IBA 进行个性化微调后,相较于 PureGaze、GazeTR 和 DGE-GM,本方法在个性化小样本校准后角度误差分别降低了 43.5%、26.5%和 21.8%,实现了精度显著提升。

2.4 个性化小样本实验

为进一步验证本文方法在少量样本条件下的适应能力,开展小样本个性化微调实验。对于每一位新受试者,本文选取不同微调样本数量作为校准数据,通过启用 PFM 与 IBA 模块在预训练模型上进行快速微调,在剩余测试样本上进行性能评估,对训练步数进行归一化处理,以保证不同样本规模下收敛过程的可比性。并与 FAZE^[22]小样本模型进行对比。实验结果如图 4、5 所示。

由图 4、5 可知,未进行个性化微调时的平均角度误差为 4.27° 。当 $k=1$ 时平均角度误差已较未微调状态降低 13.3%,说明模型在极少样本条件下即可有效学习用户偏差。随着样本数量增加,性能持续提升并快速逼近平衡区间。例如,当 $k=10$ 时误差降至 2.94° ,相较基线下降

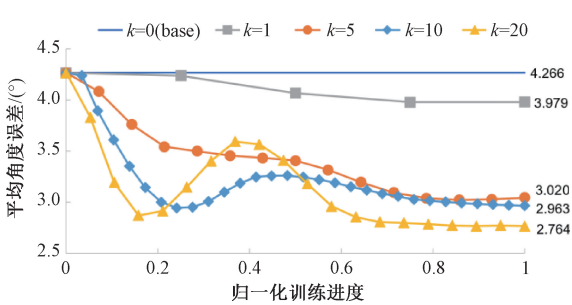


图 4 不同校准样本模型收敛速率变化曲线

Fig. 4 Model convergence rate curve of different correction sample

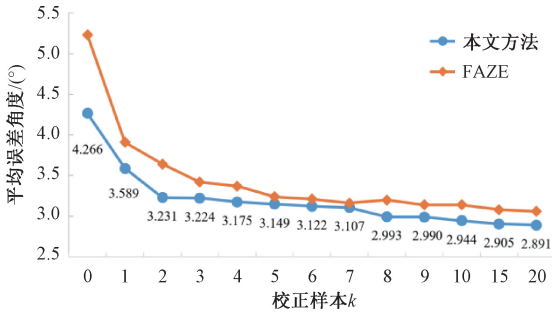


图 5 不同模型误差变化曲线

Fig. 5 Error variation curves of different models

31%，当 $k=20$ 时进一步降至 2.76° ，提升幅度收敛，表明模型已接近性能饱和。整体而言，模型性能随 k 增加而快速提升，并呈现边际收益递减的趋势，并在不同校准样本下均优于现有方法。在成本受限的实际应用场景中，该特性能够显著减少个性化标注开销，同时保证较高的预测精度，体现出该方法良好的实用性与可扩展性。

2.5 消融实验

为系统性地评估各模块对视线估计精度的贡献，本文对所提出的 GFEM、PFM 与 IBA 模块进行消融实验，所有实验均在 MPIIGaze 数据集上采用相同训练策略与数据划分进行，并以跨用户测试误差作为衡量指标，结果取多次实验平均值。实验结果如表 2 所示。

表 2 不同结构性能对比

Table 2 Comparison results of different structural properties

(°)

ResNet	ResNeXt	Transformer	GFEM	PFM	IBA	角度误差
✓		✓				5.64
	✓	✓				5.54
	✓	✓	✓			5.24
	✓	✓		✓		4.55
	✓	✓			✓	3.63
	✓	✓	✓	✓		4.27
	✓	✓	✓		✓	2.99
	✓	✓		✓	✓	3.02
	✓	✓	✓	✓	✓	2.94

由表 2 可知，将主干由 ResNet 更换为 ResNeXt 后，角度误差由 5.64° 降至 5.52° ，说明分组卷积结构能够提取更丰富的局部细节特征，增强了特征表达能力。启用 GFEM 后，误差降至 5.24° ，说明 GFEM 能够增强与视线相关的特征表达并抑制冗余结构信息，有效改善跨用户拟合性能。在此基础上引入 PFM，误差进一步降至 4.26° ，表明 PFM 能够融合用户嵌入实现初步偏差补偿，使模型获得更高泛化能力。单独启用 IBA 时，能够将误差降至 3.10° ，表明个体偏置学习在适应新用户方面具有更直接作用。所有模块联合启用时误差进一步降低至 2.94° ，实现了从全局结构增强到个体差异补偿的性能协同提升。实验结果表明，各模块均有效提升了估计性能，其中 IBA 在用户偏差校准方面作用最为显著，而 GFEM 与 PFM 通过提升全局表征能力和提供先验偏置支持，为个性化阶段提供了更优基础。

同时为直观展示本文所提出视线检测模型的预测效果，数据集部分样本的可视化结果如图 6 所示。

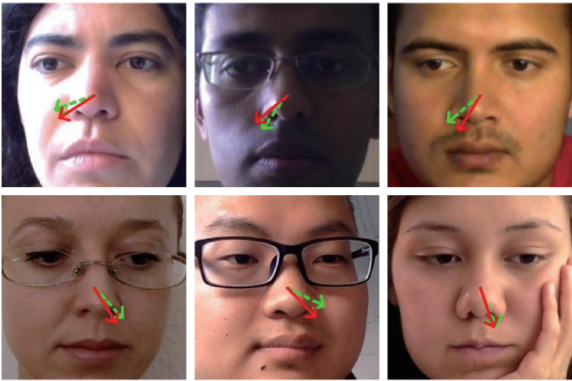


图 6 结果可视化图像

Fig. 6 Results visual images

其中,绿色虚线箭头表示真实视线方向,红色实线箭头表示模型预测结果。由图 6 可知,模型在不同光照、种族及注视方向下预测方向和真实方向均吻合良好,证明所提出模型面对真实世界复杂变化时具有良好的鲁棒性与泛化能力。

2.6 分心检测应用验证

为验证本文视线估计模型在下驾驶员分心检测中的应用效果,基于预测的视线角度构建了一个简单的阈值判定策略,并对预测结果进行可视化验证。以驾驶画面中前方道路区域为注意参考域,当视线方向偏离该区域超过设定阈值时认为存在分心风险。依据驾驶员注意力监测的相关研究,阈值设定为 $|\text{yaw}|=20^\circ$ 、 $|\text{pitch}|=10^\circ$,持续时间为 2 s^[23]。判断逻辑如图 7 所示。

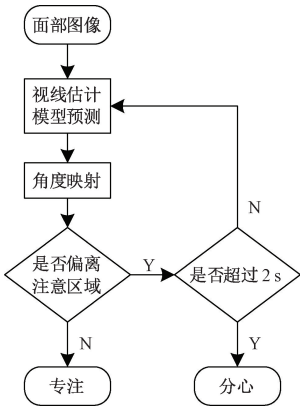


图 7 驾驶员分心状态判断流程

Fig. 7 Driver distraction state judgment process

同时为验证本文方法在驾驶员分心检测领域的实用性,本文选取 YOLOv7^[24]、RT-DETR^[25] (real-time detection transformer)以及 HorNet^[26]等目前主流方法在 DMD 驾驶员数据集上进行对比验证。YOLOv7 通过高效的网络架构优化和模块设计,在保持高帧率的同时,显著提升了复杂场景下的检测性能。RT-DETR 通过高效的混合编码器和自适应匹配策略,消除了传统检测方法中的非

极大值抑制后处理步骤,在保证高精度的同时大幅提升了推理速度。HorNet 是通过高阶空间交互机制,在不引入额外计算开销的情况下,显著提升了模型对复杂空间关系的建模能力。需要特别说明的是,现有主流方法是直接对驾驶员行为图像进行分类,属于端到端的图像分类任务;而本文方法是先进行视线方向回归,再根据规则判断分心,属于回归+决策的两阶段任务。为在同一评估标准下进行公平对比,将连续的视频片段按上述阈值规则转化为“专注”与“分心”的二分类标签,并以此计算所有方法的准确率。对比实验结果如表 3 所示。

表 3 不同模型分心检测对比

Table 3 Comparison of distraction detection of different models (%)

模型	Acc
YOLOv7	83.00
ResNet50	86.91
RT-DETR	88.32
HorNet	89.07
本文模型	88.64

由表 3 所示,本文模型的准确率高干大部分目前现有分心状态检测方法,相较于 YOLOv7、ResNet50 和 RT-DETR,Acc 分别高干 6.8%、2% 和 0.4%。证明了本文方法在驾驶员分心检测领域有较好的实用性及发展潜力。另外本文精确度略低于 HorNet 方法。通过对实验数据进行分析,发现有部分驾驶员进行如喝水或打电话动作时,即使处在分心驾驶状态,目光依旧保持注视前方,导致模型判断为专注驾驶;但同时部分驾驶员因外部干扰导致视线偏离驾驶方向,本文方法成功捕捉并判断为分心驾驶,而 HorNet 方法则未能识别此情况,侧面印证了本文方法的可解释性和泛用性。

本文方法在真实驾驶场景中的视线预测可视化效果如图 8 所示,其中箭头表示驾驶员视线方向,绿色实线箭

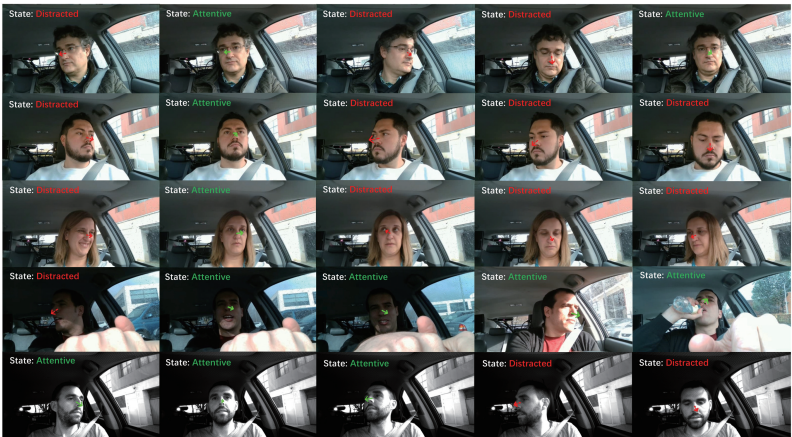


图 8 驾驶员分心状态检测可视化

Fig. 8 Visualization of driver distraction state detection

头表示正常驾驶,红色虚线箭头则表示分心驾驶。眼镜反射对基于外观的注视估计方法构成额外挑战,为此本文额外选用戴眼镜驾驶员的数据进行验证,并引入部分存在面部遮挡的数据以检验方法的通用性。此外,还采用不同光照条件下的驾驶员数据,评估本文方法在极端光照场景中的性能,包括使用夜间红外摄像头采集的实际行驶视频片段。综合来看,在上述多种复杂条件下,本文方法所预测的视线方向与真实视线轨迹吻合良好,能够准确判别驾驶员注意力是否集中于道路且具有良好的鲁棒性。这表明本文提出的基于视线估计的分心检测方法不仅具备良好的方向预测精度,也在实际驾驶场景中展现出较强的应用潜力。

综上所述,本文通过系统的对比实验、消融实验、小样本验证以及多场景可视化,全面验证了所提出模型的有效性 with 泛化能力。此外,本文打破了对特定行为标签或显式分类器的依赖,创新性地提出了一种基于连续视线回归的检测范式。该范式依据物理意义明确的视线方向,即可实现高精度、可解释的分心状态识别,为驾驶员状态监测乃至更广泛的注意力感知交互领域提供了更具可解释性与泛化能力的解决方案。

3 结 论

本文提出了一种基于视线检测的驾驶员分心状态检测方法,以连续视线角度偏移替代离散行为分类,实现了对驾驶员注意力变化的更高敏感性与可解释性。所设计的 ResNeXt-Transformer 混合网络结合 GFEM,在局部细节与全局依赖建模方面取得平衡,具备更强的视线特征提取能力,并在跨用户预测任务中表现出优于现有方法的性能。同时,本文提出的多层次个性化校准机制能够在少量个体样本条件下实现快速适配与误差校准,显著提升模型在真实驾驶场景下的稳定性与可部署性。未来研究将针对复杂环境因素(光照极端变化、遮挡等)继续增强鲁棒性,并融合头部姿态、生理行为信号等多源信息,以构建更全面的驾驶员注意力表征,以实现更加全面且安全可靠的驾驶员状态监测。

参考文献

- [1] Banerjee A, Pathivada B K, Haleem K, et al. Developing distraction-related safety performance functions at interchange ramp terminals in Kentucky[J]. Journal of Safety Research, 2025, 93: 66-78.
- [2] 陈小强,熊烨,王英,等. 面部多特征融合的列车司机疲劳检测方法研究[J]. 铁道学报, 2021, 43(12): 70-78.
Chen Xiaoqiang, Xiong Ye, Wang Ying, et al. Research on train driver fatigue detection method based on facial multi-feature fusion[J]. Journal of the China Railway Society, 2021, 43(12): 70-78.
- [3] 冯笑,代少升,黄炼. 基于可解释深度学习的单通道脑电跨被试疲劳驾驶检测[J]. 仪器仪表学报, 2023, 44(5): 140-149.
Feng Xiao, Dai Shaosheng, Huang Lian. Cross-subject driver fatigue detection from single-channel EEG with an interpretable deep learning model[J]. Chinese Journal of Scientific Instrument, 2023, 44(5): 140-149.
- [4] 赵栓峰,李乐平,王茂权,等. 基于深度学习的驾驶员分心行为检测方法[J]. 电子测量技术, 2024, 47(9): 145-153.
Zhao Shuanfeng, Li Leping, Wang Maoquan, et al. Driver distraction behavior detection method based on deep learning[J]. Electronic Measurement Technology, 2024, 47(9): 145-153.
- [5] 殷旭鹏,赵兴强,王雄飞,等. 基于改进 YOLOv10 的驾驶疲劳检测算法[J]. 电子测量与仪器学报, 2025, 39(10): 41-51.
Yin Xupeng, Zhao Xingqiang, Wang Xiongfei, et al. Driving fatigue detection algorithm based on improved YOLOv10[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(10): 41-51.
- [6] Subbulakshmi S, Arathy A K, Madhu P. Ensemble model for driver distraction detection[C]//4th IEEE Global Conference for Advancement in Technology (GCAT), 2023: 1-7.
- [7] Li W L, Huang J, Huang Y Z, et al. SWAM: Driver distraction recognition based on attention mechanism[C]//IEEE Smartworld, Ubiquitous Intelligence & Computing, Scalable Computing & Communications, Digital Twin, Privacy Computing, Metaverse, Autonomous & Trusted Vehicles, 2022: 2032-2039.
- [8] 许丽梅,周治国,周学华. PUDet: 基于生成上采样网络的 3D 目标检测方法[J]. 仪器仪表学报, 2025, 46(8): 228-243.
Xu Limei, Zhou Zhiguo, Zhou Xuehua. PUDet: A 3D object detection method based on generative upsampling network[J]. Chinese Journal of Scientific Instrument, 2025, 46(8): 228-243.
- [9] Cheng Y H, Wang H F, Bao Y W, et al. Appearance-based gaze estimation with deep learning: A review and benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 7509-7528.
- [10] Ghosh S, Dhall A, Hayat M, et al. Automatic gaze analysis: A survey of deep learning based approaches[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 46(1): 61-84.

- [11] Chen Z K, Shi B E. Towards high performance low complexity calibration in appearance based gaze estimation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 1174-1188.
- [12] Guan Y R, Chen Z G, Zeng W Z, et al. End-to-end video gaze estimation via capturing head-face-eye spatial-temporal interaction context[J]. IEEE Signal Processing Letters, 2023, 30: 1687-1691.
- [13] Jin S W, Dai J, Nguyen T. Kappa angle regression with ocular counter-rolling awareness for gaze estimation[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 2659-2668.
- [14] 张冰战, 边博乾, 杨梓恒, 等. 考虑工况、驾驶员与道路信息的再生制动策略[J]. 电子测量与仪器学报, 2025, 39(4): 181-192.
Zhang Bingzhan, Bian Boqian, Yang Ziheng, et al. Regenerative braking strategy considering working conditions, driver and road information[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(4): 181-192.
- [15] Albazzaz H F M, Khayeat A R H. A review on distracted driver detection using machine learning techniques [C]//2025 IEEE 15th Symposium on Computer Applications & Industrial Electronics (ISCAIE), 2025: 445-450.
- [16] Aponso A, Speelman C, Johnstone M. A multidisciplinary review of machine learning for driver distraction prediction: unveiling challenges[C]//2025 10th International Conference on Information and Network Technologies (ICINT), 2025: 30-35.
- [17] 程鸣, 严运兵. 驾驶员视线区域自适应半监督学习标定方法[J]. 电子测量与仪器学报, 2023, 37(12): 136-142.
Cheng Ming, Yan Yunbing. Driver's line-of-sight region adaptive semi-supervised learning calibration method[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(12): 136-142.
- [18] Sharma P K, Chakraborty P. A review of driver gaze estimation and application in gaze behavior understanding [J]. Engineering Applications of Artificial Intelligence, 2024, 133: 108117.
- [19] Cheng Y H, Bao Y W, Lu F. PureGaze: Purifying gaze feature for generalizable gaze estimation [C]//AAAI Conference on Artificial Intelligence, 2022, 36(1): 436-443.
- [20] Cheng Y H, Lu F. Gaze estimation using transformer[C]//2022 26th International Conference on Pattern Recognition (ICPR), 2022: 3341-3347.
- [21] Li T G, Zhang Y Z, Li Q Q. Appearance-based driver 3D gaze estimation using GRM and mixed loss strategies[J]. IEEE Internet of Things Journal, 2024, 11(23): 38410-38424.
- [22] Park S, Mello S D, Molchanov P, et al. Few-shot adaptive gaze estimation [C]//IEEE/CVF International Conference on Computer Vision, 2019: 9368-9377.
- [23] 夏小均, 黄毅, 陈德兵, 等. 基于注视行为的高速分心驾驶识别方法[J]. 机电工程技术, 2025, 54(6): 129-134, 149.
Xia Xiaojun, Huang Yi, Chen Debing, et al. High-speed distracted driving recognition method based on gaze behavior[J]. Mechanical & Electrical Engineering Technology, 2025, 54(6): 129-134, 149.
- [24] 李富, 徐凯, 朱灵龙, 等. 基于上下文语义联合 YOLOv7 的分心驾驶检测算法[J]. 国外电子测量技术, 2024, 43(3): 121-128.
Li Fu, Xu Kai, Zhu Linglong, et al. Distracted driving detection algorithm based on context semantics combined with YOLOv7 [J]. Foreign Electronic Measurement Technology, 2024, 43(3): 121-128.
- [25] Zhao Y A, Lyu W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.
- [26] Rao Y M, Zhao W L, Tang Y S, et al. HorNet: Efficient high-order spatial interactions with recursive gated convolutions[J]. Advances in Neural Information Processing Systems, 2022, 35: 10353-10366.

作者简介

杨思翰, 硕士研究生, 主要研究方向为计算机视觉、多源数据融合。

E-mail: 1410256060@qq.com

何青(通信作者), 博士, 副教授, 主要研究方向为智能控制、现代控制理论以及嵌入式系统。

E-mail: 254739027@qq.com