

基于深度学习的驾驶员分心行为检测方法^{*}

赵栓峰 李乐平 王茂权 李小雨 谢乐坤

(西安科技大学机械工程学院 西安 710054)

摘要: 驾驶员分心行为检测对于开发以驾驶员为中心的人车协同驾驶系统具有至关重要的意义。针对现有基于卷积神经网络的驾驶员分心行为检测模型缺乏全局特征提取能力、泛化性能较弱以及忽视了驾驶场景中不同区域的重要性,构建一种基于深度学习的驾驶员分心行为检测模型,实现对驾驶员分心行为的准确检测。首先,开发了基于 HorNet 的残差结构,通过高阶空间交互来增强特征表示能力;其次,受人类注意力机制以及现有注意力机制的启发,设计一种自适应加权注意策略来提取与驾驶行为最相关的特征;然后,在现有的分类数据集上训练本文模型,并使用先验知识作为初始权值来改善训练结果,进而提高模型的泛化能力;最后,对驾驶行为特征进行可视化,以提高人们对于本文模型的信任度。实验结果表明,本文模型可以准确地检测驾驶员分心行为,在准确性和可靠性方面明显优于现有方法。

关键词: 驾驶员行为;先验知识;深度学习;视觉可解释性

中图分类号: TN0; TP391 文献标识码: A 国家标准学科分类代码: 520.6

Driver distraction behavior detection method based on deep learning

Zhao Shuanfeng Li Leping Wang Maoquan Li Xiaoyu Xie Lekun

(College of Mechanical Engineering, Xi'an University of Science and Technology, Xi'an 710054, China)

Abstract: Driver distraction behaviour detection is of crucial significance for the development of driver-centered human-vehicle co-driving systems. Aiming at the existing convolutional neural network-based driver distraction detection models that lack global feature extraction capability, have weak generalisation performance and neglect of the importance of different regions in the driving scene, a driver distraction detection model based on deep learning is constructed to achieve accurate prediction of driver distraction behaviour. First, a residual structure based on HorNet is developed to enhance the feature representation capability through higher-order spatial interactions; second, inspired by the human attention mechanism and the existing attention mechanisms, an adaptive weighted attention strategy is designed to extract the features most relevant to the driving behaviour; and then, the model in this paper is trained on the existing categorical dataset, and the a priori knowledge is used as the initial weights to improve the training results which in turn improves the generalisation ability of the model; finally, the driving behaviour features are visualised to improve the trust in this paper's model. The experimental results show that the model in this paper can accurately detect driver distraction behaviour, which is significantly better than existing methods in terms of accuracy, and reliability.

Keywords: driver behavior; prior knowledge; deep learning; visual interpretability

0 引言

近年来,由于人工智能、物联网、传感技术和信息通讯技术取得了令人瞩目的进展,自动驾驶技术取得了巨大的突破。目前,自动驾驶已经完全实现了 L2 级(部分自动辅助驾驶)功能,而 L3 级(有条件自动驾驶)的智能网联汽车即将实现规模化应用。尽管完全自动驾驶被认为具有广阔

的前景,但现阶段的自动驾驶技术对车辆的控制仍然不够成熟。此外,由于各种限制,包括现有的决策模型在陌生的非结构化驾驶环境中难以实现商业级别的稳健性,以及交通立法和伦理问题,完全自动驾驶在商业应用中尚未广泛实施^[1]。因此,以人为中心的人车协同驾驶系统是未来几年的发展趋势。此外,依据美国国家公路交通安全管理局发布的统计数据^[2],在 2019 年,由驾驶员分心行为引发的

交通事故数量达到了 3 142 起。鉴于以上分析,实时、准确地检测驾驶员分心行为对于人车协同驾驶系统至关重要。驾驶员分心行为检测可以预测潜在的不安全操作,协助人车协同驾驶系统做出相应决策(语音提醒或启动制动),减少交通事故发生的风险。因此,开发一种能够准确、及时地检测驾驶员分心行为的方法,对于减少交通事故、保障行车安全具有重大的现实意义和价值。

许多学者对驾驶员分心行为检测进行了广泛且深入的研究,取得了重大进展。根据数据和传感器的不同,驾驶员分心行为检测方法可以分为 3 类:基于驾驶人生理信号^[3-4](如眼动数据、心动数据和脑动数据)的方法,基于车辆信息^[5](如车辆速度、方向盘转角和加速度)的方法,以及基于驾驶员图像的方法。基于驾驶人生理信号的方法依赖于侵入式生理传感器。然而,大部分生理传感器不仅价格昂贵,而且会对驾驶人的正常驾驶产生影响,难以在实际应用中发挥作用。基于车辆信息的方法也存在一定的局限性。例如,传感器的硬件故障和外部干扰可能导致采集的信息不准确,进而影响系统决策的准确性。更重要的是,一旦车辆速度、方向盘转角和加速度等信息发生显著变化,通常意味着交通事故已经发生。相比之下,基于驾驶员图像的方法是非侵入性的,不会影响驾驶人的正常驾驶,也不会像基于车辆信息的方法那样存在滞后性。因此,本文主要研究基于图像的驾驶员分心行为检测方法。该方法只需一个单目摄像机,通过实时监测驾驶人的头部、面部、手部等部位来检测驾驶员是否分心。

在 21 世纪初期,传统的机器学习分类算法被广泛用于驾驶员分心行为检测,包括决策树、随机森林、支持向量机等。例如,Zhao 等^[6]结合了同态滤波、皮肤分割和轮廓波变换作为特征提取工具。然后使用随机森林、线性感知器和多层次感知机等分类器对司机的行为进行分类,包括正常驾驶、饮食、使用手机、和换挡操作。刘明周等^[7]提出了一种基于尺度不变特征变换和反向传播神经网络的检测方法,用于检测驾驶员行为。然而,上述机器学习方法需要专家知识来进行特征提取,不能很好地处理大量样本^[8]。此外,传统的机器学习分类算法可以在特定场景中准确地检测各种驾驶员行为,但在复杂环境下的泛化能力较差。因此,探索更为高效、具有更强泛化能力的驾驶员分心行为检测方法具有重要的研究价值和实际意义。

由于计算机视觉的发展以及深度学习理论的成熟^[9-10],基于卷积神经网络(convolutional neural networks,CNN)的驾驶员分心行为检测方法正成为最新的研究热点。与传统的机器学习方法相比,基于 CNN 的方法在准确性、检测速度和泛化能力方面具有更好的性能。例如,Tran 等^[11]分别使用了 VGG-16、AlexNet、GoogLeNet 和 ResNet 4 种 CNN 模型来检测驾驶分心行为,实验结果表明 GoogLeNet 模型取得了最好的检测性能。尹智帅等^[12]提出了一种基于 2D/3D 人体姿态估计的分心驾驶行为检

测方法,通过将驾驶姿态估计网络与 ResNet 分类网络融合,建立了基于驾驶人姿态估计的分心驾驶行为检测模型,显著提高了驾驶行为检测的准确性。然而,文献[11-12]主要采用传统监督学习方法进行训练,这需要大量的标记数据进行训练,大大增加了标记成本和训练成本。此外,有监督学习模型对标记数据集的依赖性降低了它们的泛化能力。针对文献[11-12]存在的问题,Huang 等^[13]提出了一种基于人脑巩固学习的驾驶行为识别模型。Liu 等^[14]提出了一种基于知识蒸馏的框架,通过将教师网络中的知识迁移到学生网络,实现了预测准确率的显著提升。然而,上述方法主要采用传统的 CNN 模型进行特征提取。CNN 模型擅长从图像中提取局部特征,对全局特征信息的捕捉能力有限。

此外,上述文献[11-14]都是利用 CNN 模型对整个驾驶场景进行特征提取而忽视了不同区域特征对驾驶员分心行为检测的重要性。例如:当驾驶员在打电话时,模型应该重点关注手机所在的区域,并抑制其他区域的干扰。基于上述文献存在的优点和问题,本文提出了一种基于深度学习的驾驶员分心行为检测模型,以实现对驾驶员分心行为的准确预测。为了增强模型对全局特征信息的捕捉能力,提出了一种基于 HorNet 的残差结构,通过高阶空间交互来增强模型特征表示能力。为了提取与驾驶行为相关的特征,设计一种自适应加权注意策略,用以提取驾驶场景中与驾驶行为相关的重要特征,并抑制复杂背景的干扰。为了减少对大量标记数据的依赖性,并增强模型的泛化性能,本文开发了一种基于先验知识的训练方法。该方法在现有的分类数据集上训练本文模型以获得先验知识,并将其作为初始权值来改进训练结果。此外,本文通过对注意区域的可视化直观展示模型在预测时所关注的区域,这有助于增强预测模型的可解释性,进而提高人们对于驾驶行为预测模型的可信度。

1 基于深度学习的驾驶员分心行为检测模型架构

通过对现有驾驶员分心检测模型的调研得出:CNN 模型对全局特征信息的捕捉能力有限,且监督学习模型依赖大量的标记数据进行训练;现有端到端的驾驶员分心检测模型对整个驾驶场景进行特征提取而忽视了不同区域特征对驾驶员分心行为检测的重要性。基于上述问题,受人类注意力机制的启发,提出一种基于深度学习的驾驶员分心行为检测模型,其详细结构如图 1 所示。本文所提出驾驶员分心行为检测模型的核心结构包括:基于 HorNet 的残差结构、自适应加权注意策略和基于先验知识的训练方法。

1.1 基于 HorNet 的残差结构

在驾驶员分心行为检测领域,传统的基于卷积神经网络的驾驶员分心行为检测模型在特征提取方面效率较低,对整体图像信息的捕捉能力有限。卷积神经网络只堆叠卷积和池化运算。因此,它主要关注局部特征的提取,而检测驾驶员

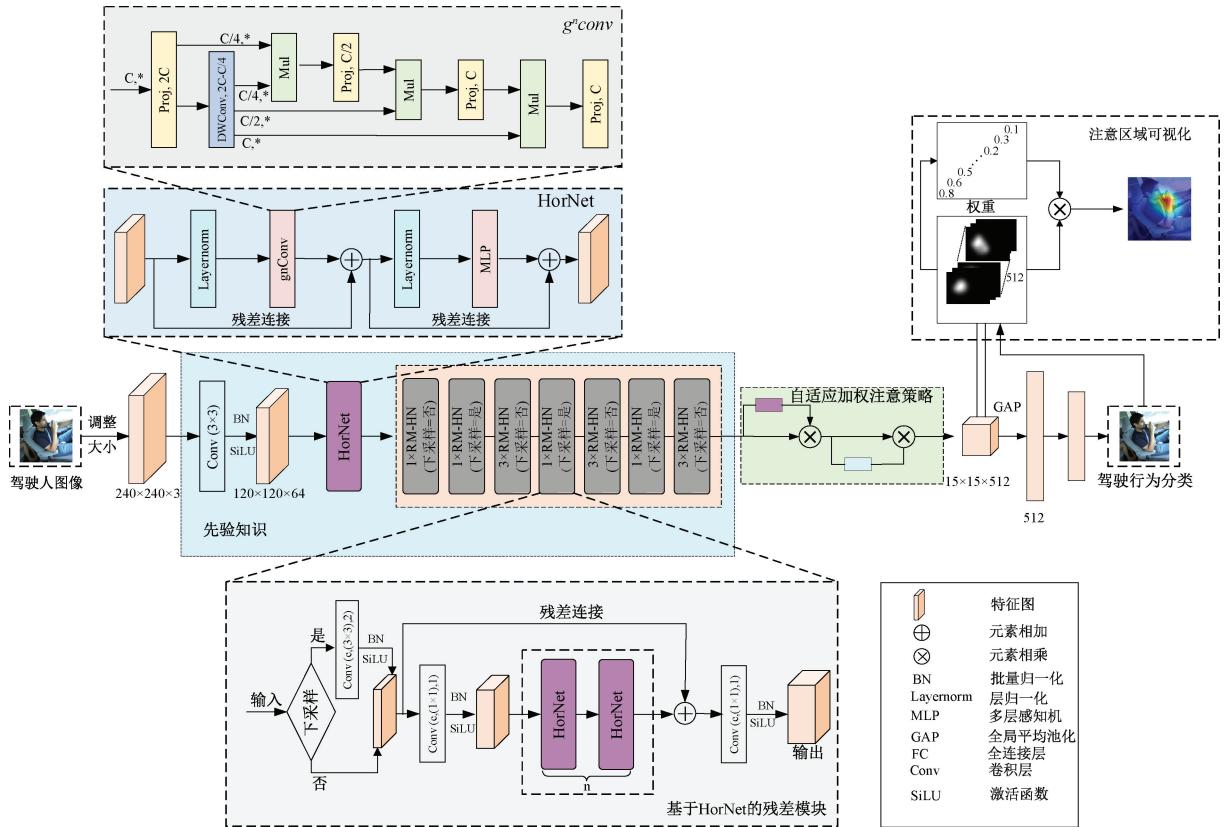


图 1 驾驶员分心行为检测模型整体框架

行为的任务需要更好地掌握全局信息。为了解决这一问题,提出了一种基于 HorNet 的残差结构(residual module based on HorNet, RM-HN)来捕获高级语义特征。如图 1 下部分所示, RM-HN 是残差模块^[15]的一种变体。其中, HorNet^[16]使用递归门控卷积(gⁿconv)学习图像的远距离依赖关系并获取高阶特征,递归门控卷积的架构如图 1 上部分所示。

假定递归门控卷积的二维输入特征为 $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, 可以得到一组投影特征 p_0 和 $\{q_k\}_{k=0}^{n-1}$, 计算公式如下:

$$[p_0^{HW \times C_0}, q_0^{HW \times C_0}, \dots, q_{n-1}^{HW \times C_{n-1}}] = \varphi_{in}(\mathbf{X}) \quad (1)$$

其中, $\varphi_{in}(\cdot)$ 表示投影运算。然后, 门控卷积由下式递归执行:

$$p_{k+1} = f_k(q_k) \odot g_k(p_k) / \alpha \quad (2)$$

其中, α 表示缩放参数, $f_k(\cdot)$ 表示深度卷积操作, $g_k(\cdot)$ 表示用于调整通道数的线性投影层。

$$g_k = \begin{cases} \text{Identity}, & k = 0 \\ \text{Linear}(C_{k-1}, C_k), & 1 \leq k \leq n-1 \end{cases} \quad (3)$$

最后, 通过对最后一个递归层进行特征投影得到输出。在残差结构中, 采用残差连接保留原始特征信息; 批量归一化(batch normalization, BN)层用于加快网络的训练和收敛的速度, 并防止梯度消失; SiLU 激活函数通过非线性变换增强模型的表达能力。基于 HorNet 的残差结构的架构如图 1 下部分所示, 其过程如下: 是否进行下采样处理 \rightarrow Conv + BN + SiLU \rightarrow HorNet \rightarrow Conv + BN + SiLU。

1.2 自适应加权注意策略

驾驶员图像中包含许多与驾驶行为无关的特征, 例如背景信息。人类倾向于优先处理与任务相关的细节信息, 而减弱其他无关的信息, 这被称为人类的视觉注意力机制。受人类的视觉注意力机制以及现有注意力机制^[17-18]的启发, 开发了自适应加权注意策略, 进而提取驾驶场景中与驾驶行为相关的重要特征, 抑制复杂背景的干扰。自适应加权注意策略的实现细节如图 2 所示。该策略由通道注意模块和空间注意模块组成。通道注意模块和空间注意模块都具有丰富的全局池化信息。通道注意模块对不同通道的特征进行权重分配, 有助于提升模型的表征能力。空间注意模块可以突出驾驶员所在区域的位置特征, 忽略无关紧要的区域, 从而减少背景噪声的干扰。在通道注意模块中, 利用全局平均池化层(global average pooling, GAP)提取输入特征的全局信息。然后, 通过一维卷积操作对全局信息进行特征编码, 输出特征向量 \mathbf{Z} 。最后, 通过激活函数对 \mathbf{Z} 进行归一化得到每个通道的权重 \mathbf{X} 。计算过程如下:

$$\mathbf{Z} = \text{Conv1d}(\text{GAP}(\mathbf{E})) \quad (4)$$

$$\mathbf{X} = \text{Sigmoid}(\mathbf{Z}) \quad (5)$$

空间注意模块的工作流程类似于通道注意模块。全局池化层用于提取特征图的全局信息, 并沿通道维度进行连接。然后进行卷积操作, 得到融合后的特征图 \mathbf{K} 。最

后,利用激活函数对 \mathbf{K} 进行归一化处理,输出空间注意特征 \mathbf{L} 。计算过程如下:

$$\mathbf{K} = \text{conv2d}(\text{Concat}(\text{Avgpool}(\mathbf{F}), \text{Maxpool}(\mathbf{F}))) \quad (6)$$

$$\mathbf{L} = \text{Sigmoid}(\mathbf{K}) \quad (7)$$

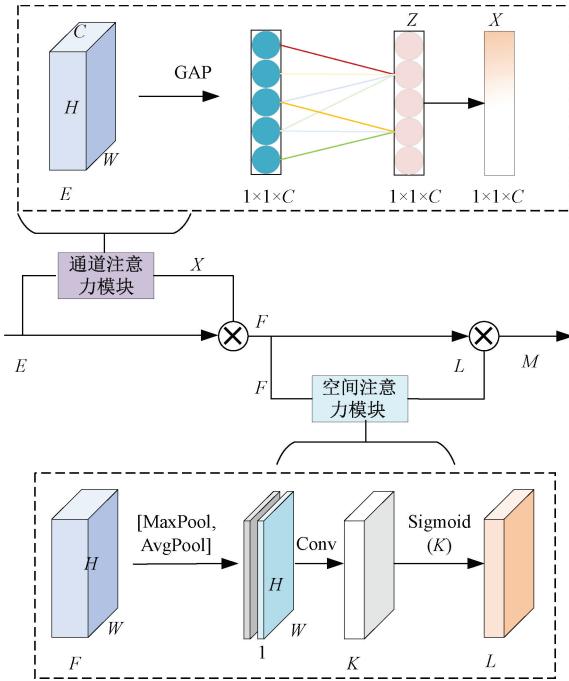


图 2 自适应加权注意策略

1.3 基于先验知识的训练方法

传统的有监督学习模型的权值是随机初始化的。因此,有监督学习模型依赖大量高质量的驾驶员图像进行训练,这无疑增加了训练成本。在现实驾驶场景中,收集并标记数据集需要大量人力和资源。更为重要的是,如果仅依赖有限数量的高质量图像进行训练,会阻碍有效特征的提取,进而导致模型的检测精度低、泛化能力弱。人类可

以将在一个领域学到的知识应用到另一个相关领域,以提高学习效率和效果。例如,学习小提琴可以帮助更快地学习钢琴,学习识别狗可以帮助识别猫。受人类跨领域知识迁移能力的启发,本文开发了一种基于先验知识的训练方法,以防止训练过程中模型的过度拟合,并提高其泛化性能。与传统的有监督学习不同,基于先验知识的学习方法使用源任务的信息作为学习目标任务的基础。值得注意的是,源任务和目标任务之间必须存在一定的相关性或共同特征。如果两个任务之间缺乏相关性,迁移学习可能无法带来预期的性能提升^[19]。在本研究中,目标任务是驾驶行为的检测,这是一个分类问题。因此,源任务也应该是分类任务。基于先验知识的训练方法的过程如下:

1) 确定源任务。为了更好地训练基于先验知识的驾驶行为检测模型,考虑到源任务和目标任务之间的相关性,本研究选择 CIFAR-10 分类任务作为源任务。这是因为该任务与驾驶行为检测都属于同一类问题。另外,CIFAR-10 数据集具有多样和丰富的数据。

2) 获取先验知识。在确定源任务后,将 CIFAR-10 数据集输入到本文模型中,以获得先验知识来准确地对驾驶行为进行分类。

2 实验与结果分析

本文所提出的基于深度学习的驾驶员分心行为检测模型在配置为 GeForce RTX4080、Intel(R) Core(TM) i7-13700KF 的电脑上开展实验。为充分验证本文所提出驾驶员分心行为检测模型的检测性能,本文基于 2.1 节所述的 StateFarm 数据集、2.2 节所述的模型超参数和 2.3 节所述的评价指标开展对比试验、消融试验。

2.1 实验数据

用于测试所提出的模型的实验数据包括 StateFarm 数据集。如图 3 所示,StateFarm 数据集包括 10 个驾驶员行为类别,其中包括正常驾驶、聊天、操控中控、喝水等。



图 3 StateFarm 数据集分心驾驶行为分类

2.2 模型超参数设置

为了清晰描述本文所提出模型的结构细节和训练细节,本节将详细阐述与模型结构相关的超参数以及模型训

练时使用的超参数。为了方便陈述与模型结构相关的超参数,将基于 HorNet 的残差结构和自适应加权注意策略 (adaptive weighted attention strategy, AWAS) 定义为

RM-HN(c, n, s)、AWAS。其中 c 为输出特征图的通道数; n 表示 RM-HN 中 HorNet 的数量; s 为步长。根据图 1 的模型结构得出与驾驶行为预测模型结构相关的超参数如表 1 所示。模型训练时使用的超参数包括损失函数、学习率、优化器、批大小和训练轮次大小。采用交叉熵损失作为损失函数, 初始学习率设置为 0.001, 优化器为随机梯度下降优化器, 批大小为 64, 训练轮次为 100。

表 1 与模型结构相关超参数

输入	操作	c	n	s	下采样
$240 \times 240 \times 3$	Conv 3×3	64		2	
$120 \times 120 \times 64$	HorNet				
$120 \times 120 \times 64$	RM-HN	64	1		否
$120 \times 120 \times 64$	RM-HN	128	1		是
$60 \times 60 \times 128$	RM-HN	128	3		否
$60 \times 60 \times 128$	RM-HN	256	1		是
$30 \times 30 \times 256$	RM-HN	256	3		否
$30 \times 30 \times 256$	RM-HN	512	1		是
$15 \times 15 \times 512$	RM-HN	512	3		否
$15 \times 15 \times 512$	AWAS				
$15 \times 15 \times 512$	GAP				
512	FC				
		输出: 10			

2.3 模型评价指标

评价模型性能的指标包括精确率(precision)、召回率(recall)和准确率(accuracy)。相关公式的计算如下:

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$Recall = \frac{TP}{TP + FN} \quad (9)$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (10)$$

其中, TP 表示正确预测的样本的数量, FN 表示将阳

性类别预测为阴性类别的样本数目, FP 表示阴性类别预测为阳性类别的样本数量, TN 表示将阴性类别预测为阴性类别的样本数目。驾驶员分心行为检测属于多分类问题, 因此本文选用精确率、召回率和准确率的平均值来评价分类性能, 即 mP 、 mR 和 $mAcc$ 。

为了提高模型的可信度, 还采用可视化方法对模型的注意区域进行评估。

2.4 对比实验

为验证本文模型的优越性, 3 个经典的模型(InceptionV3^[20]、DenseNet^[21] 和 ResNet-101^[22])被用来在 StateFarm 数据集上与本文所提出的模型进行对比。 InceptionV3、DenseNet 和 ResNet-101 是深度学习领域公认的经典网络结构, 它们在驾驶员分心行为检测任务中取得了卓越的性能。这些模型各自采用了不同的网络结构。 InceptionV3 通过多尺度特征提取和因子分解减少了参数数量, 提高了模型的效率。 DenseNet 通过特征重用机制增强了模型的表达能力。 ResNet-101 使用残差连接来降低参数优化的难度, 可以解决深层网络中的退化问题。然而, 这些模型主要关注局部特征提取, 对于全局特征的捕捉能力有限。此外, 这些模型都是利用 CNN 对整个驾驶场景进行特征提取, 缺乏针对特定任务(如驾驶员分心行为检测)的注意力机制。与之相比, 本文提出的模型引入了基于 HorNet 的残差结构和自适应加权注意策略, 这不仅增强了模型对全局特征的捕捉能力, 而且能够更加精准地提取与驾驶行为相关的特征, 从而提高了模型的准确性和可靠性。对比实验结果如表 2~3 和图 4~5 所示。

表 2 不同模型性能对比

模型	mP	mR	$mAcc$
本文模型	98.40	98.40	98.40
DenseNet	95.56	95.55	96.05
InceptionV3	96.55	97.05	97.00
ResNet-101	97.50	97.50	97.50

表 3 不同模型精确率

模型	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	mP
本文模型	97.03	99.98	99.49	99.01	98.52	97.97	98.99	96.02	99.49	97.50	98.40
DenseNet	93.50	99.98	97.02	94.23	98.26	97.99	99.99	96.53	89.25	88.79	95.56
InceptionV3	90.82	99.97	98.98	96.55	98.00	97.46	99.99	95.10	94.12	94.50	96.55
ResNet-101	97.03	95.00	97.98	98.02	99.98	97.47	97.98	96.57	97.47	97.50	97.50

如表 2 所示, 本文模型具有最高的 $mAcc$ 、 mP 、 mR , 相比于 DenseNet、InceptionV3 和 ResNet-101, $mAcc$ 分别提高了 2.35%、1.4% 和 0.9%。如表 3 所示, 本文模型在测试集上的表现更好, 相比于 DenseNet、InceptionV3 和 ResNet-101, mP 分别提高了 2.84%、1.85% 和 0.9%。不同模型的混淆矩阵如图 4 所示。本文采用 Grad-CAM^[23]

绘制热力图, 展示模型关注的区域。各模型在预测时关注的区域如图 5 所示。由图 5 可知, 尽管这些区域的位置有所不同, 但所有模型均能准确检测出驾驶员的 10 种行为。与其他 CNN 模型相比, 该模型能更有效地提取驾驶行为特征, 且关注的区域更加准确, 与人类判断驾驶行为时关注的区域一致。例如, 当驾驶员与乘客聊天时, 注意区域

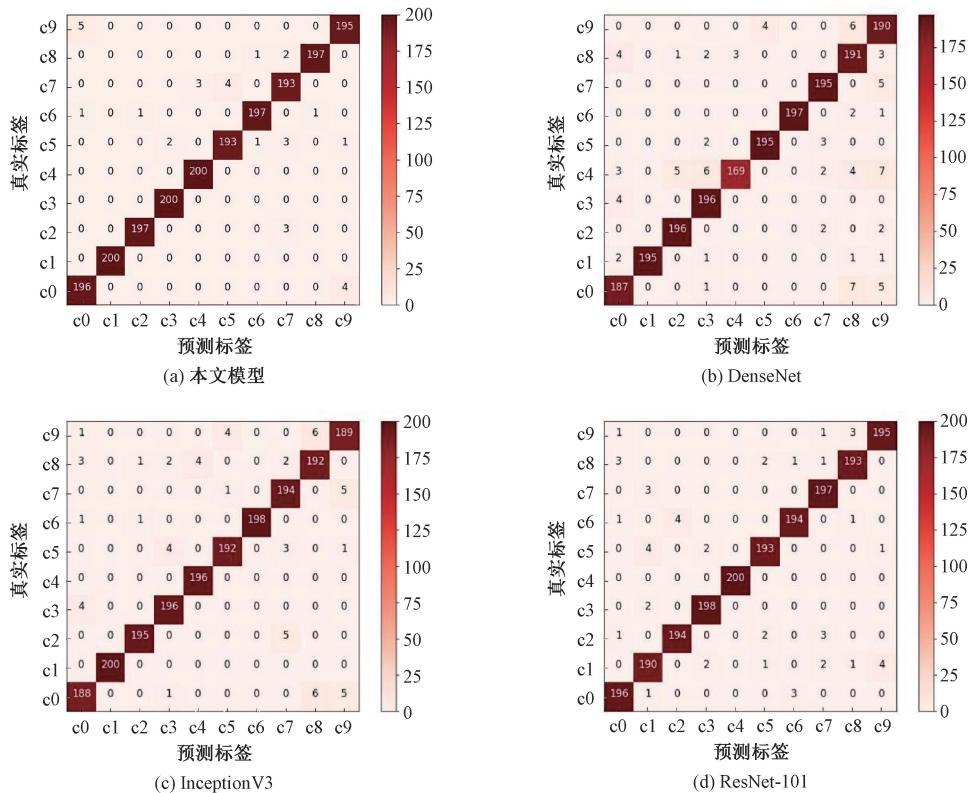


图 4 不同模型的混淆矩阵

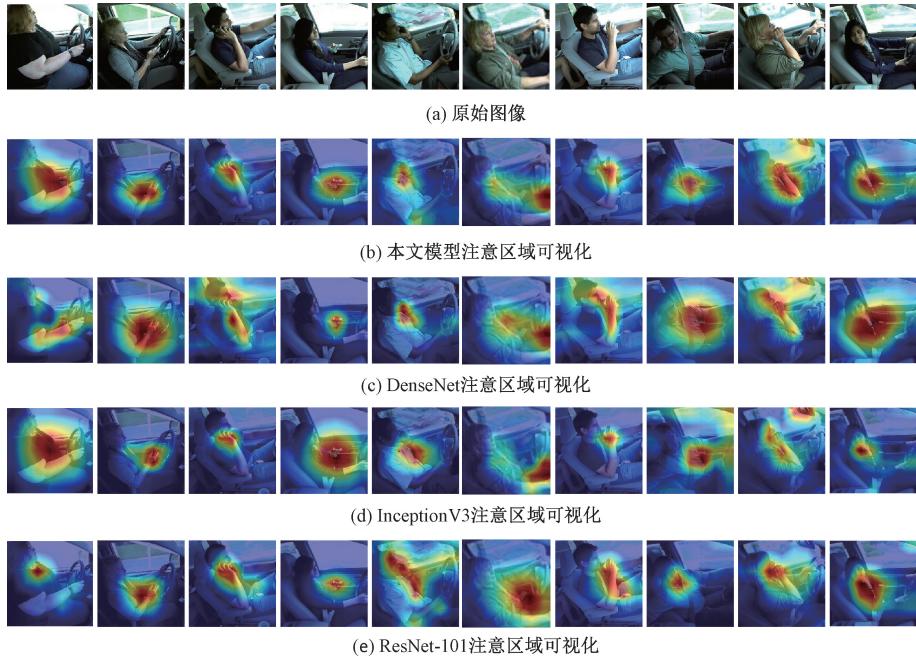


图 5 不同模型的可视化结果

集中在驾驶员的颈部;当驾驶员整理妆容时,注意区域集中在驾驶员的脸部和化妆镜。其他 CNN 模型也能有效提取驾驶行为特征,但所关注的区域相对分散。

2.5 消融试验

本文的模型具有以下优点:引入 HorNet 以提高增强模型的特征表示能力;使用自适应加权注意策略来提取与

驾驶行为相关的特征,消除无关信息的干扰;采用基于先验知识的训练方法,以改善训练效果。为验证本文所提出的分心驾驶行为检测模型中不同组件对分心驾驶行为预测的重要性,本文开展消融实验,以检验不同组件对模型性能的贡献。为了确保实验的公平性,所有参数均设置一致。消融实验设计如下:

模型 1: 移除所提出的分心驾驶行为检测模型中自适应加权注意策略;模型 2: 提出的分心驾驶行为检测模型采用传统有监督学习进行训练;模型 3: 移除所提出的分心驾驶

行为检测模型中 HorNet 结构;模型 4: 本文所提出的分心驾驶行为检测模型。消融实验结果如表 4~5 和图 6~7 所示。

表 4 不同模型性能对比

模型	mP	mR	$mAcc$
本文模型	98.40	98.40	98.40
模型 1	97.31	97.30	97.30
模型 2	97.15	96.95	96.95
模型 3	95.62	95.50	95.50

表 5 不同模型的精确率

模型	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	mP
本文模型	97.02	99.99	99.49	99.01	98.52	97.97	98.99	96.02	99.49	97.50	98.40
模型 1	97.01	94.97	97.99	98.02	99.99	97.99	97.97	98.01	94.12	96.98	97.31
模型 2	95.57	98.00	98.47	97.03	99.00	96.97	99.46	95.56	96.46	95.00	97.15
模型 3	93.69	97.95	95.57	94.69	98.84	97.50	99.48	96.08	93.17	89.25	95.62

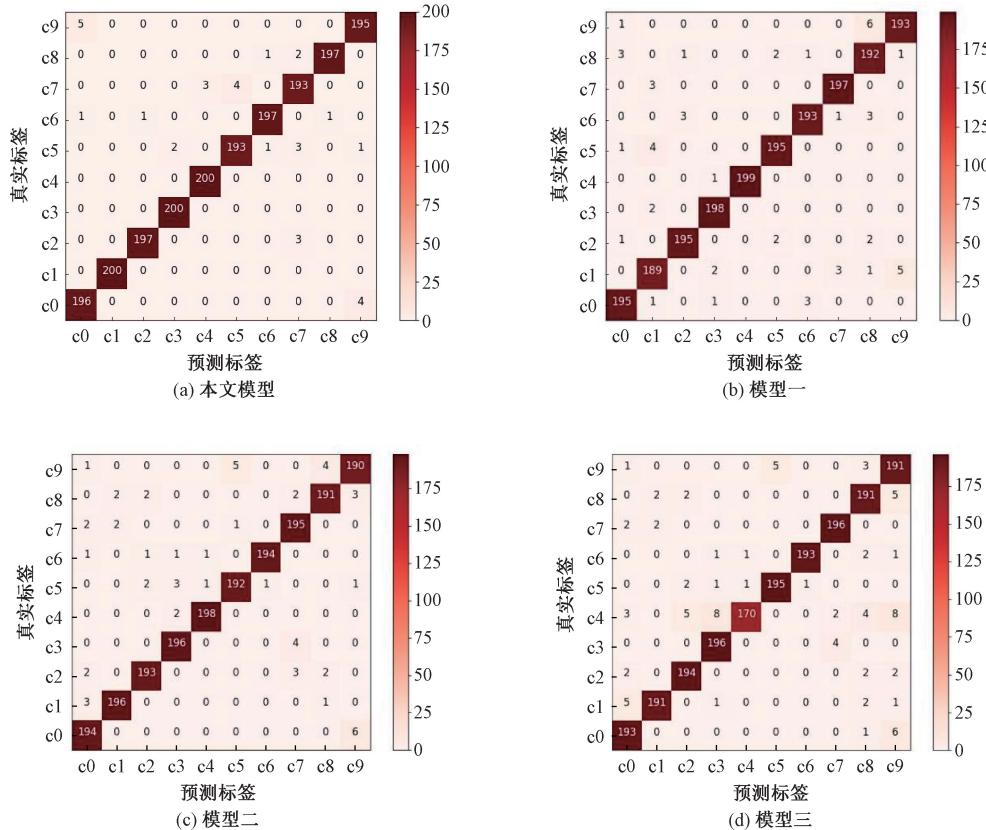


图 6 不同模型的混淆矩阵

如表 4 所示,本文模型的 mP 、 mR 和 $mAcc$ 显著高于消融实验中的其他模型,表明 HorNet、自适应加权注意策略和基于先验知识的训练方法显著提高了模型的检测性能。具体而言,相比于模型 1、模型 2 和模型 3, $mAcc$ 分别提高了 1.1%、1.75% 和 2.9%。消融实验中各模型的各类别精确率和平均精确率如表 5 所示,由表 5 可知,本文模型

的平均精确率高于其他模型。消融实验中各模型的混淆矩阵如图 6 所示。消融实验中各模型在预测时关注的区域如图 7 所示。由图 7 可知,本文模型的注意区域明显优于消融试验中的其他模型。模型 1 和模型 2 关注的区域较为准确,但是存在注意区域分散的现象。模型 3 存在注意区域错误的情况。例如,如图 7(e)所示,当驾驶人在操控

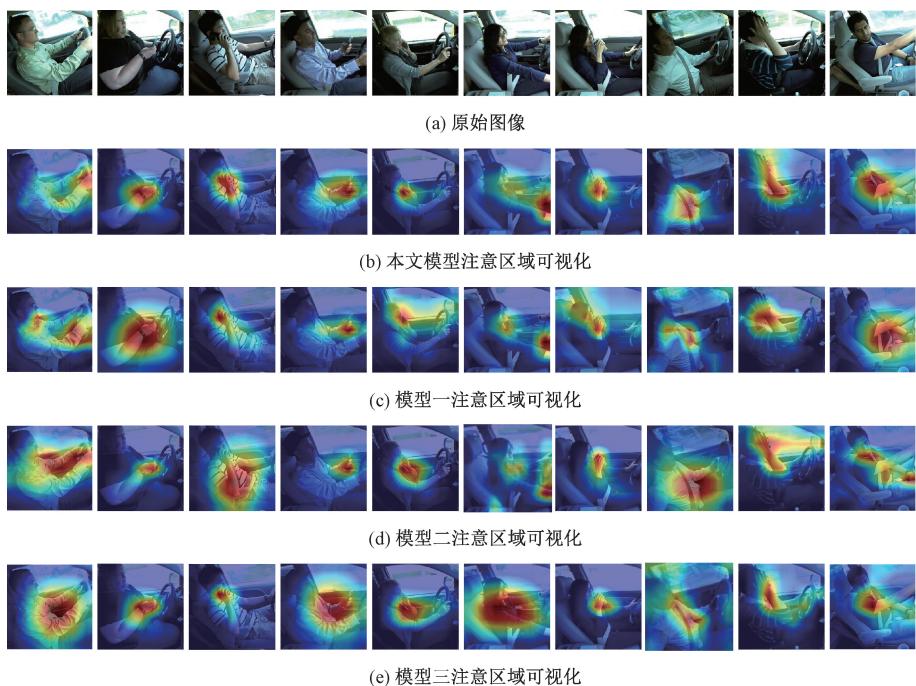


图 7 不同模型注意区域可视化

中控时,模型 3 关注的区域并非驾驶人的手部。综上所述,HorNet、自适应加权注意策略和基于先验知识的训练方法可以提高模型的检测性能。

3 结 论

本文主要针对驾驶员分心行为检测进行研究,构建一种基于深度学习的驾驶员分心行为检测模型。为了增强模型对全局特征信息的捕捉能力,提出了一种基于 HorNet 的残差结构,通过高阶空间交互来增强模型特征表示能力。为了提取与驾驶行为相关的特征,设计一种自适应加权注意策略,用以提取驾驶场景中与驾驶行为相关的重要特征,并抑制复杂背景的干扰。在现有的分类数据集上训练本文模型以获得先验知识,并将其作为初始权值来提高模型的泛化能力。此外,本文通过对注意区域的可视化直观展示模型在预测时所关注的区域,以提高人们对于模型的可信度。实验结果表明,本文模型可以准确地检测驾驶员分心行为。后续工作将继续研究强光或弱光环境下的驾驶员分心行为检测。

参考文献

- [1] HU Y, QU T, LIU J, et al. Human-machine cooperative control of intelligent vehicle: Recent developments and future perspectives [J]. *Acta Automatica Sinica*, 2019, 45(7): 1261-1280.
- [2] ZHANG Y, LI T, LI C, et al. A novel driver distraction behavior detection method based on Self-Supervised learning with masked image modeling[J]. *IEEE Internet of Things Journal*, DOI:10.1088/1757-899X/782/2/022012.
- [3] ZUO X, ZHANG C, CONG F, et al. Driver distraction detection using bidirectional long short-term network based on multiscale entropy of EEG[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(10): 19309-19322.
- [4] CHAI R, NAIK G R, NGUYEN T N, et al. Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system[J]. *IEEE Journal of Biomedical and Health Informatics*, 2016, 21(3): 715-724.
- [5] WANG X, XU R, ZHANG S, et al. Driver distraction detection based on vehicle dynamics using naturalistic driving data[J]. *Transportation Research Part C: emerging Technologies*, 2022, 136: 103561.
- [6] ZHAO C H, ZHANG B L, HE J, et al. Recognition of driving postures by contourlet transform and random forests [J]. *IET Intelligent Transport Systems*, 2012, 6(2): 161-168.
- [7] 刘明周,蒋倩男,扈静.基于面部几何特征及手部运动特征的驾驶员疲劳检测[J].*机械工程学报*,2019,55(2):18-26.
- [8] MCDONALD A D, FERRIS T K, WIENER T A. Classification of driver distraction: A comprehensive analysis of feature generation, machine learning, and input measures[J]. *Human Factors*, 2020, 62 (6): 1019-1035.
- [9] 叶彦斐,程立,侯翔一.基于改进 YOLO v5 的钢轨内

- 部伤损 B 显图像识别与分类[J]. 国外电子测量技术, 2023, 42(12): 70-76.
- [10] 罗仁泽, 唐祥, 余泓, 等. 基于改进 Faster RCNN 的射线图像焊缝缺陷检测方法[J]. 电子测量技术, 2023, 46(22): 160-168.
- [11] TRAN D, MANH D H, SHENG W, et al. Real-time detection of distracted driving based on deep learning[J]. IET Intelligent Transport Systems, 2018, 12(10): 1210-1219.
- [12] 尹智帅, 钟恕, 聂琳真, 等. 基于人体姿态估计的分心驾驶行为检测[J]. 中国公路学报, 2022, 35(6): 312-323.
- [13] HUANG T, FU R, CHEN Y. Deep driver behavior detection model based on human brain consolidated learning for shared autonomy systems [J]. Measurement, 2021, 179: 109463.
- [14] LIU D, YAMASAKI T, WANG Y, et al. Toward extremely lightweight distracted driver recognition with distillation-based neural architecture search and knowledge transfer [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 24 (1): 764-777.
- [15] HE K, ZHANG X, REN S, et al. Identity mappings in deep residual networks [C]. Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part IV 14. Springer International Publishing, 2016: 630-645.
- [16] RAO Y, ZHAO W, TANG Y, et al. Hornet: Efficient high-order spatial interactions with recursive gated convolutions[J]. Advances in Neural Information Processing Systems, 2022, 35: 10353-10366.
- [17] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 3-19.
- [18] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11534-11542.
- [19] ZHUANG F, QI Z, DUAN K, et al. A comprehensive survey on transfer learning [J]. Proceedings of the IEEE, 2020, 109(1): 43-76.
- [20] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.
- [21] HUANG G, LIU Z, VAN D M L, et al. Densely connected convolutional networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
- [22] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [23] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 618-626.

作者简介

赵栓峰, 博士, 教授, 主要研究方向为计算机视觉、目标检测、交通安全等。

E-mail: zsf@xust.edu.cn

李乐平(通信作者), 硕士研究生, 主要研究方向为计算机视觉、图像处理。

E-mail: 1693522484@qq.com