基于骨架特征的驾驶分心行为识别方法

王海玮[1],蔡耿基[2],张 为[3],杨敬姗[2],游 峰[2]

(1.广东交通职业技术学院 智慧交通工程学院,广州 510650;2.华南理工大学 土木与交通学院,广州 510641;
3.车路一体智能交通全国重点实验室,北京 100088)

摘 要:针对现有视觉方法在复杂驾驶环境中易受光照和背景干扰的问题,提出了一种基于骨架特征的驾驶分心行为识别方法。通过OpenPose姿态估计算法提取驾驶员13个关键骨架关节点,构建多维度动作识别特征描述符,包括骨架结构向量、骨架关节矢量角向量和骨架模比向量,并采用特征优选策略筛选出7维显著特征。在此基础上,改进深度残差网络,以增强模型的泛化能力。实验在公开驾驶行为数据集上进行,结果表明:本文方法在经人体语义分割的数据集上准确率达到97.33%。此外,真实场景验证表明,本文方法在复杂光照、背景和视角变化下仍能保持良好的识别稳定性。本文为车载监控系统和智能交通安全预警提供了高效、鲁棒的技术支持,具有重要的实际应用价值。

关键词:交通安全;驾驶分心行为;骨架特征;人体姿态估计;特征工程

中图分类号:U471.3 **文献标志码**:A **文章编号**:1671-5497(2026)06-1548-16

DOI:10.13229/j.cnki.jdxbgxb.20250607

Skeleton-feature-based method for identifying driving distraction behaviors

WANG Hai-wei[1],CAI Geng-ji[2],ZHANG Wei[3],YANG Jing-shan[2],YOU Feng[2]

(1. *School of Smart Transportation Engineering*, *Guangdong Communication Polytechnic*, *Guangzhou* 510650, *China*;
2. *School of Civil Engineering and Transportation*, *South China University of Technology*, *Guangzhou* 510641, *China*;
3. *State Key Lab of Intelligent Transportation System*, *Beijing* 100088, *China*)

Abstract: Aiming at the problem that the existing visual methods are vulnerable to the interference of illumination and background in complex driving environments, a method for identifying driving distraction behaviors based on skeleton features is proposed. Thirteen key skeleton joint points of drivers are extracted through OpenPose, and multi-dimensional Action Description Features (ADFs), including skeleton structure vectors, skeleton joint vector angle vectors, and skeleton modulus ratio vectors, are constructed. A feature-selection strategy is adopted to screen out 7-dimensional significant features. On this basis, a deep residual network is improved to enhance the model's generalization ability. Experiments are carried out

收稿日期:2025-07-06.

基金项目:车路一体智能交通全国重点实验室课题项目(2024-B011);广东省自然科学基金项目(2024A1515012115);广东省普通高校特色创新类项目(2022KTSCX256);中国高校产学研创新基金项目(2021ZYB04003).

作者简介:王海玮(1983-),女,副教授,博士.研究方向:智慧交通,车路协同.Email:hwwang17@gdcp.edu.cn

通信作者:游峰(1977-),男,副教授,博士.研究方向:交通安全与控制,车路协同.E-mail: youfeng@scut.edu.cn

on public driving behavior datasets. The results show that the accuracy of the proposed method on the dataset with human semantic segmentation reaches 97.33%. Furthermore, real-world scenario validation demonstrates that the proposed method maintains good recognition stability under complex variations in illumination, background, and viewpoint. This research provides an efficient and robust technical support for vehicle monitoring systems and intelligent traffic safety early warning, and has important practical application value.

Key words: traffic safety; driving distraction behavior; skeleton features; human pose estimation; feature engineering

0　引　言

近年来，交通运输行业的迅猛发展使得道路交通安全问题日益凸显。相关统计数据显示，超过 90% 的道路交通事故归因于驾驶员行为因素[1]，而驾驶过程中的分心行为则是其中最主要的诱因。现行法律法规对分心驾驶的约束机制存在滞后性，难以针对个体驾驶特性实现实时预警。因此，探究驾驶员正常驾驶与分心驾驶的特征差异，提取表征驾驶员驾驶状态的数理特征，构建实时、稳定的驾驶员分心监测系统，对分心行为进行预警，将成为降低相关交通事故发生率的有效途径。

随着深度学习在包括分心驾驶行为检测在内的各个领域都取得了巨大的成功，神经网络模型被广泛应用于驾驶员行为识别任务。目前的研究主要围绕提升模型的检测精度、实时性与鲁棒性展开，具有代表性的方法之一则聚焦于改进目标检测或图像分类框架，以增强对复杂驾驶场景的适应能力。Zhang 等[2]基于端到端的深度无监督多模态融合网络（UMMFN）检测驾驶员的分心行为，通过计算 UMMFN 的重建误差来做出精细的检测决策；陈仁祥等[3]提出基于改进 YOLOv5 的驾驶员分心驾驶检测方法，引入 Ghost 模块，采用线性变换代替部分常规卷积进行特征提取，并从认知分心和视觉分心两个方面检测驾驶员分心驾驶行为；Baheti 等[4]提出了 mobileVGG 架构，基于深度可分离卷积检测和分类驾驶员分心；赵霞等[5]提出基于深度卷积和 Tokens 降维的 ViT 模型用于驾驶员驾驶分心行为实时检测；Liu 等[6]提出了一种基于通道扩展和注意力机制的 CEAM-YOLOv7 模型用于驾驶员分心行为检测，该模型扩大了全局维度交互特征，使网络能提取关键特征，并显著简化了卷积计算，有利于提高模型检测速度；陈军等[7]设计了一种级联卷积神经网络检测框架，由第一级分心行为预筛选卷积网络和第二级分心行为精确检测卷积网络两个全卷积网络级联构成。这些方法大多依托成熟的检测架构，通过轻量化设计或结构优化提升效率与精度，但其性能高度依赖图像质量，在光照变化或背景复杂条件下易出现误检。

另一类相关研究则致力于多源特征融合，以提升模型在噪声环境下的鲁棒性。Ping 等[8]提出了一种基于时空双线深度学习网络（TSD-DLN）和因果与或图（C-AOG）的分心行为识别方法，该方法在处理连续帧检测分心驾驶行为时具有优势，且在不完全观测场景下表现出准确的连续识别能力；Alkinani 等[9]提出了一种混合分心驾驶检测方案，首先获取 HOG 特征，然后利用 4 个深度 CNN 模型提取 DCNN 特征，在级联层将 DCNN 提取的特征与 HOG 提取的特征进行融合，实现了对 10 类分心驾驶行为进行分类。此类方法注重结合传统手工特征与深度学习特征，在数据不完整或噪声较多的情况下表现较优，但特征融合策略的设计复杂度较高，且仍依赖高质量视觉输入。

此外，部分研究开始探索更高层次的行为语义建模，尝试从人体姿态或注意力机制中挖掘关键动作信息。尹智帅等[10]提出一种适用于驾驶舱环境的驾驶特征提取方法，将 2D 人体姿态估计网络 Simple Baseline 和 3D 姿态估计网络 DensePose 应用于分心驾驶行为检测；贺宜等[11]将注意力机制引入轻量型 MobileViT 网络，通过选择有效的网络主干、注意力模块 CA、网络嵌入位置从而设计出分类模型 MobileViT-CA，有效提升分类网络的性能；张斌等[12]提出了一种基于类间特征距离优化的分心驾驶行为识别模型训练策略，通过扩大不同类别样本在特征空间中的分布间

距,强制模型学习更具判别性的特征表示,从而实现分类性能的整体优化。这类方法试图利用人体结构或注意力机制捕捉行为本质,增强了模型对行为的理解能力,特别是姿态估计的引入为摆脱对图像质量的依赖提供了新思路,但目前多数研究仍将其作为辅助模块,未能充分发挥其独立表征潜力。

综上所述,尽管现有基于视觉的驾驶分心行为识别方法在模型结构优化、多模态融合等方面取得了显著进展,但仍存在若干关键问题亟待解决。一方面,大多数方法依赖原始图像的像素信息,其性能易受驾驶舱内光照变化、遮挡及复杂背景等干扰因素影响;另一方面,当前研究多集中于网络架构的改进,而对驾驶员自身动作结构特征的挖掘尚不充分。这在一定程度上限制了模型对分心行为本质的理解能力,也制约了识别精度的进一步提升。因此,如何构建一种对环境干扰具有较强鲁棒性、且能充分刻画驾驶员行为动态特征的识别方法,成为当前亟需突破的技术瓶颈。

为解决上述问题,本文提出了一种基于骨架特征的驾驶分心行为检测方法。该方法通过以下创新点实现性能提升:优化骨架关键点提取流程,筛选与驾驶行为强相关的关节点;构建融合空间结构与动态变化的多维度特征描述符;改进网络架构以增强骨架特征的分类能力。该技术方案可为智能车载监控系统提供高鲁棒性的解决方案,对预防分心驾驶引发的交通事故具有重要实践价值。

1 基于骨架特征的驾驶分心行为识别模型

驾驶分心行为多种多样,其识别方法的有效性与所关注的行为类别密切相关。为聚焦核心问题并确保研究的深度,本文需明确界定研究范围。现有研究表明[13-16],与乘客交谈、操作手机、进食饮水、吸烟、操作车载设备等行为是导致驾驶员注意力分散的最常见且高风险的行为。本研究选择“与乘客交谈”、“接打电话”和“饮食”作为三类典型分心行为,其依据如下:

普遍性与高风险性:这3类行为在实际驾驶中发生频率高,且涉及驾驶员头部、手部等关键部位的显著姿态改变,与正常驾驶状态存在可量化的差异,是分心驾驶研究中的经典范例。

动作特征具有代表性:接打电话和饮食行为都涉及手部向面部的移动,但其运动姿态角度存在细微差异;与乘客交谈则可能表现为头部转动和上半身姿势调整。这3类行为的组合,能有效检验骨架特征对不同模式动作的区分能力。

研究的可扩展性:本文构建的基于骨架特征的驾驶分心行为识别框架,主要捕捉驾驶员上半身的核心动态。一旦在上述典型行为上验证了方法的有效性,本框架可扩展至识别其他具有显著姿态变化的分心行为(如操作中控台、抽烟等)。

因此,本文将聚焦于“3+1”类驾驶行为模式,即3类典型分心行为(与乘客交谈、接打电话、饮食)与正常驾驶行为的对比分析。

1.1 模型整体架构

基于驾驶员骨架的分心行为识别任务,本质上属于监督学习框架下的多分类问题。具体流程为:首先利用驾驶分心行为数据集,通过姿态估计算法提取驾驶员的骨架关键点坐标及其拓扑关系,这些特征数据为行为识别提供重要依据;随后训练分类模型,使其能对输入图像中的驾驶行为进行自动标注,输出包括“与乘客交谈”、“打电话”、“饮食”或“正常驾驶”等类别标签。

针对以上问题描述,本文建立的基于骨架特征的驾驶分心行为识别网络结构如图1所示。本文提出的模型架构由3个模块组成:骨架特征提取模块、数据处理与特征向量构建模块和动作识别模块。骨架特征提取模块提取驾驶员骨架特征数据,其次通过数据处理与特征向量构建模块进行人工特征构建,得到骨架动作识别特征描述符,最后通过动作识别模块对驾驶分心行为进行分类。各模块具体功能如下:

(1)骨架特征提取模块采用OpenPose姿态估计网络精确定位驾驶员身体各关节点的空间坐标。每张输入图像对应输出一组包含关节点位置及其拓扑连接关系的骨架特征数据。

(2)数据处理与特征向量构建模块接收原始骨架坐标数据,通过计算结构向量、关节角度向量和模比向量等特征量,构建初始动作描述符。进一步通过特征优选算法筛选出最具判别力的特征组合,形成用于表征驾驶员上半身姿态的最终特征向量。

(3)动作识别模块以优化后的特征描述符作为输入,基于改进的ResNet50网络架构进行模型

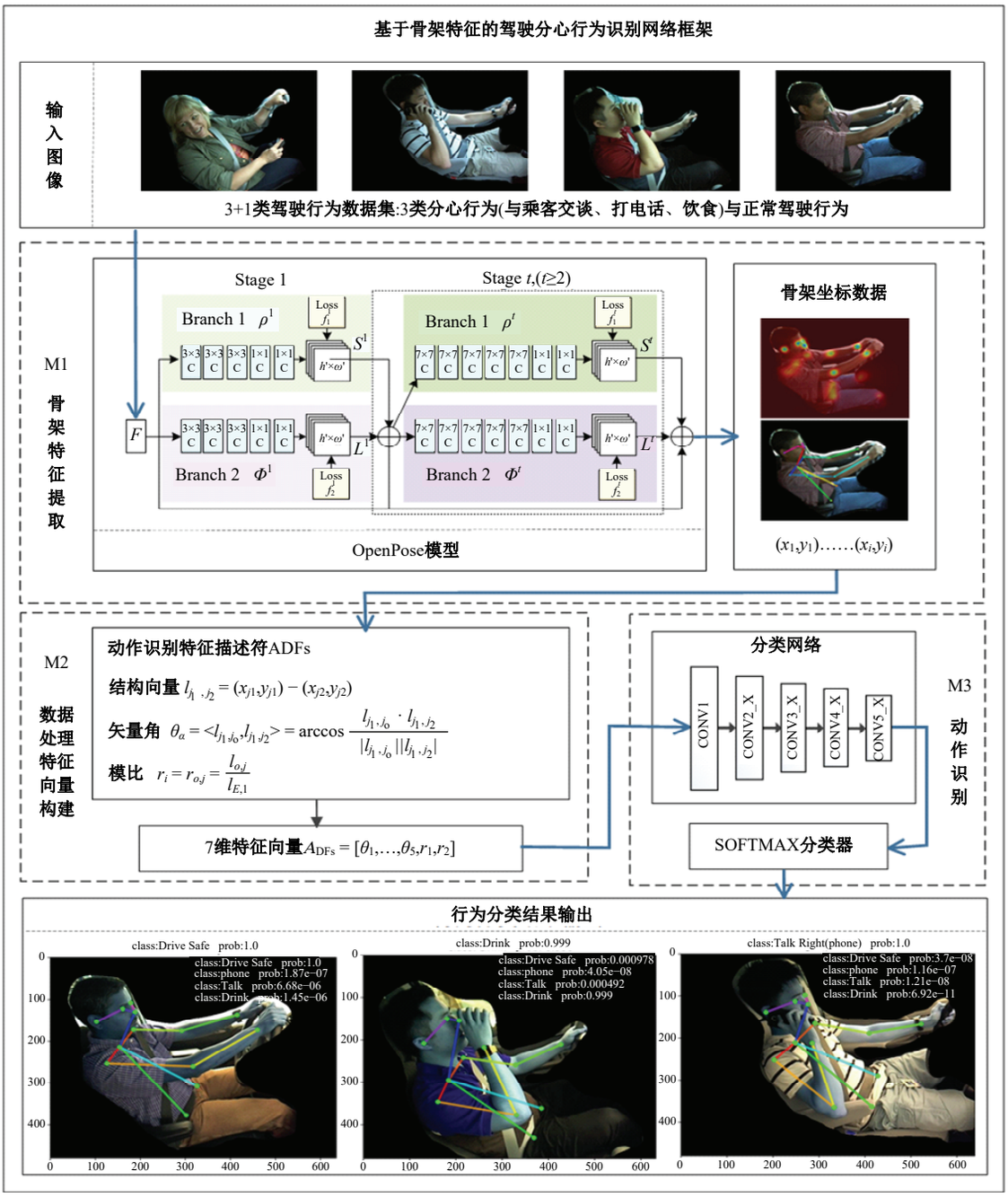


图 1　基于骨架特征的驾驶分心行为识别网络模型架构

Fig. 1　Network model architecture of driving distraction behavior recognition based on skeleton features

训练。训练过程中，分类器通过不断调整网络参数来学习不同分心行为对应的骨架特征模式。经过优化的模型能实现端到端的驾驶行为识别：从输入图像中提取骨架特征、构建特征描述符、输出对应的行为分类结果。

1.2　基于 OpenPose 的驾驶员骨架特征提取

本文选用以 OpenPose 为算法基础，构建驾驶员骨架特征的提取方法。OpenPose 算法[17]是一个实时进行姿态估计的框架，无需额外的预处理或后处理步骤。其特点在于计算速度快、精度高，已成功应用于交互设计、运动科学、增强现实以及健康监护等多个领域。

为充分表征驾驶行为的关键信息，如驾驶员头部、肩膀，手腕等，本文参考 COCO18 人体关键点模型，对 18 个人体骨架关节点进行编号，见表 1。分析发现，编号为 9、10、12、13 的关键点位于驾驶员腿部区域，与本研究所关注的驾驶分心行为关联性较低；同时，编号 17 的左耳关键点在现

表1 COCO 18人体关键点模型编号

Table 1 COCO 18 human key point model number

编号	关节点	编号	关节点	编号	关节点
0	鼻子	6	左肘部	12	左膝盖
1	肩部中心	7	左手腕	13	左脚踝
2	右肩膀	8	右臀部	14	右眼
3	右肘部	9	右膝盖	15	左眼
4	右手腕	10	右脚踝	16	右耳
5	左肩膀	11	左臀部	17	左耳

有采集视角下难以稳定获取。因此,本文在COCO18人体关键点模型的基础上,剔除以上5个相关度较弱的骨架关节点,保留13个相关度较强的骨架关节点。由此减少与驾驶行为相关度较弱的骨架关节点干扰,同时提高特征提取效率。

OpenPose的网络架构如图2所示。该模型以驾驶场景图像作为输入,首先通过预训练的卷积神经网络提取特征,随后生成驾驶员各关节点的置信图(即关键点热力图),并基于置信图确定各关节点的精确像素坐标,最终将这些坐标映射回原始图像。OpenPose的处理流程分为两个并行分支:Branch 1(顶部分支)负责预测置信图,用于定位人体各部位在图像中的可能位置;Branch 2(底部分支)负责预测亲和场(PAF),用于表征不同关节点之间的连接关系。两个分支均采用多阶段(Stage 1至Stage t)的迭代优化机制,每个阶段结合前一阶段的输出和原始特征层,逐步细化预测结果。

将本文选取的驾驶员骨架关节点在原图中进行可视化,可视化结果如图3所示,其中图3(a)(c)(e)(g)为驾驶员各关节点置信图,图3(b)(d)(f)(h)为驾驶员上半身姿态骨架可视化。进而基于所提取驾驶员上半身姿态,构建骨架特征向量,对驾驶员分心行为进行表征。

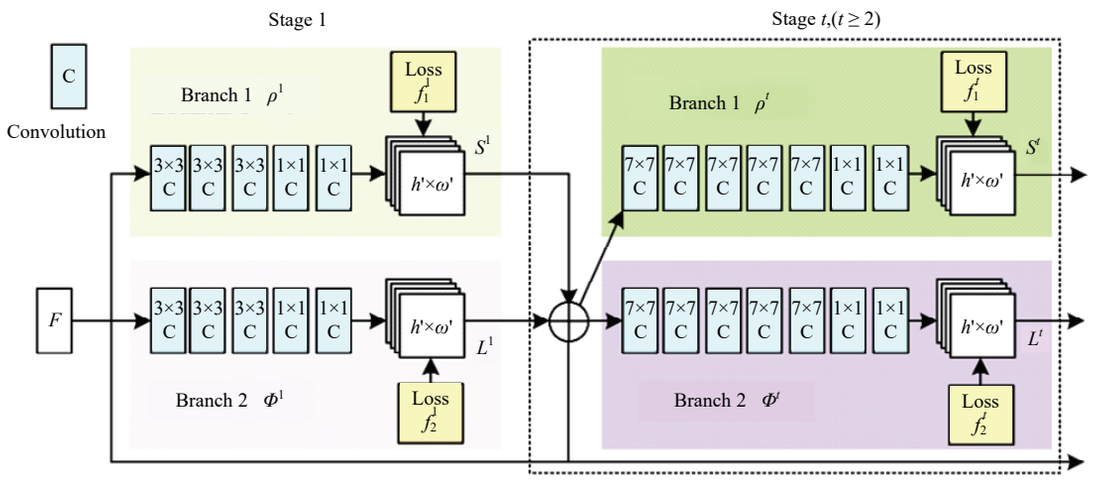


图2 OpenPose网络结构

Fig. 2 OpenPose network architecture

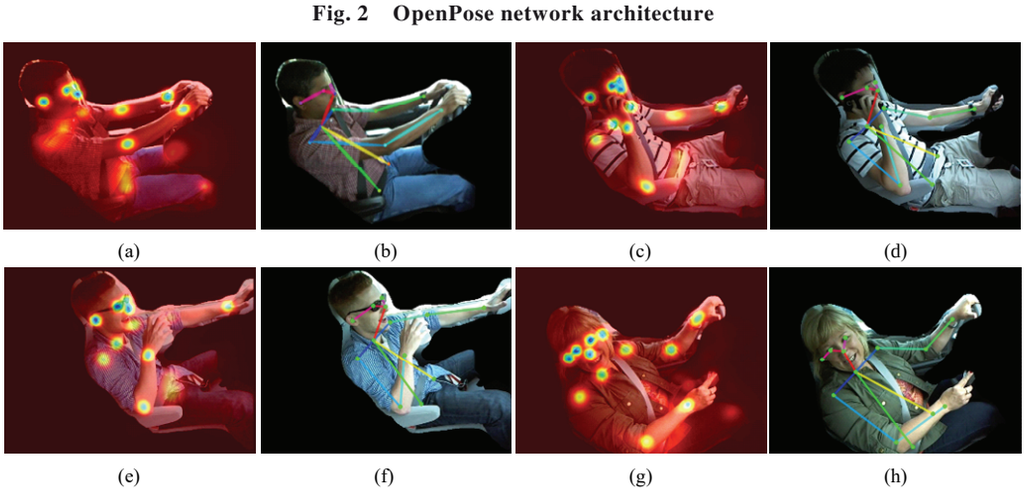


图3 四类驾驶行为的姿态可视化

Fig. 3 Pose visualization of four classes of driving behaviors

1.3　驾驶员骨架特征向量构建

直接使用13个骨架关节点数据进行建模可能会降低模型的泛化性能。为此，本文结合人体姿态分析，设计了基于驾驶员骨架的特征向量，以此获取更具判别力的动作识别特征描述符（Action description features，ADFs）。文中驾驶员骨架的特征向量由骨架结构向量、骨架关节矢量角向量和骨架模比向量3个部分组成。

1.3.1　骨架结构向量

两个关节点间的结构向量表征为其空间方向与距离关系。在驾驶员骨架特征识别中，每个关节点对应身体的特定部位（如肩部、肘部、膝盖等）。各关节点的坐标位置及其相对关系共同构成驾驶员的姿态和运动状态。通过计算关节点间的结构向量，可深入解析驾驶过程中驾驶员的身体动作和姿态变化。以驾驶员肩部与肘部的结构向量为例，该向量能体现手臂的伸展程度及方向。

结构向量的计算核心在于对同一帧图像中两个关节点的坐标进行差值运算，从而确定二者的相对位置关系。具体表达式如下：

$$l_{j_1,j_2}=(x_{j_1},y_{j_1})-(x_{j_2},y_{j_2}) \tag{1}$$

式中：(x_{j_1},y_{j_1})和(x_{j_2},y_{j_2})是关节点j_1，j_2的坐标，l_{j_1,j_2}是由j_1，j_2关节点构成的结构向量。

针对驾驶时相对稳定的关节点（1号肩部中心、8号右臀、11号左臀），本文定义了两个辅助点用于特征构建：点E（关节点8与11的中点）和点O（由关节点1、8、11构成三角形的重心）。如图4所示（彩图参见电子版，下同），本文共构建19个骨架结构向量，用于后续角度与模比计算，黑色实线表示基于人体实际骨架连接的结构向量，蓝色虚线表示为辅助特征构建而定义的非骨架结构向量。

1.3.2　骨架关节矢量角向量

在结构向量的基础上，进一步设计骨架关节的矢量角向量，以更完整地表征驾驶员骨架的动态特性。骨架关节矢量角向量表征两个结构向量之间的角度值，能反映身体部位间的相对转动关系，进而捕捉驾驶过程中的细微动作差异。

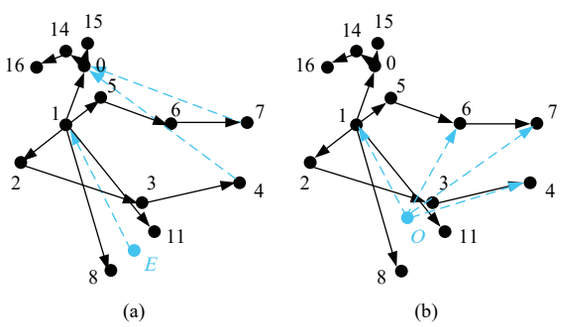


图4　骨架结构向量

Fig. 4　skeleton structure vectors

骨架关节矢量角向量由余弦定律求取：

$$\theta_\alpha=\left\langle l_{j_1,j_0},l_{j_1,j_2}\right\rangle=\arccos\frac{l_{j_1,j_0}\cdot l_{j_1,j_2}}{\left|l_{j_1,j_0}\right|\left|l_{j_1,j_2}\right|} \tag{2}$$

式中：θ_α为向量l_{j_1,j_0}、l_{j_1,j_2}之间的角度。

本文初步构建了13个骨架关节矢量角向量，如图5所示。另外，以6号（左肘）关节点为例，θ_6是图中上臂与前臂之间的角度，可用于测量前臂相对于上臂的摆动角度。θ_1和θ_2在图中分别表示左肘关节、左手腕关节相对于人体脊柱和O点的角度关系。

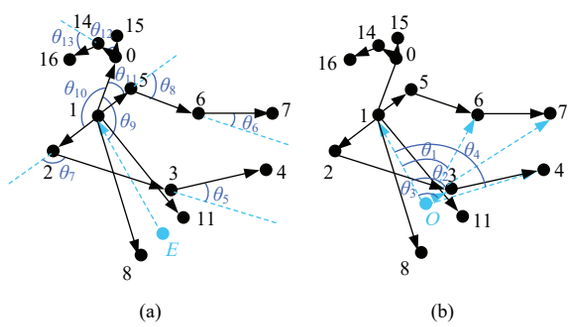


图5　骨架关节矢量角向量

Fig. 5　Skeleton joint vector angle vectors

1.3.3　骨架模比向量

为降低驾驶员个体体型差异对动作识别的影响，本文采用骨架模比向量（两个结构向量的比值）作为相对距离度量指标，而非直接使用关节绝对距离。该向量能反映驾驶过程中关节点相对位置的变化特征，有效降低个体体型因素的干扰。需要强调的是，基于骨架的分析方法核心是分析关节点间的相对位置、角度和运动模式，从本质上对驾驶员的性别、身高、体型等静态生理特征具有高度鲁棒性。本文进一步通过引入骨架模比向量这一相对度量，将个体体型差异的影响降至最低，从而确保了所构建的特征向量主要反映与驾驶行为相关的动态姿态变化，而非静态的生理差异。

本文选取关节点8（右臀）和11（左臀）的中点E，与关节点1（肩部中心）构成的结构向量$l_{E,1}$作为基向量。这两个关节点在驾驶过程中稳定性较高，能有效表征不同驾驶员的体型特征。因此，以该基准向量为参照，可更准确地量化骨架关节的动态变化。骨架模比向量表达式如下：

$$r_{O,i}=\frac{l_{O,i}}{l_{E,1}} \tag{3}$$

$$r_{j,0}=\frac{l_{j,0}}{l_{E,1}} \tag{4}$$

式中:$r_{O,i}$为辅助点O到各骨架关节点距离与基准向量$l_{E,1}$长度的比值;$r_{j,0}$为各骨架关节点到鼻部关节点距离与基准向量$l_{E,1}$长度的比值

本文总共构建了8个骨架模比向量,见表2。

表2 骨架模比向量编号

Table 2 skeleton modulus ratio vectors

缩写	骨架模比向量	缩写	骨架模比向量
r_1	$r_{O,2}$	r_5	$r_{O,6}$
r_2	$r_{O,3}$	r_6	$r_{O,7}$
r_3	$r_{O,4}$	r_7	$r_{4,0}$
r_4	$r_{O,5}$	r_8	$r_{7,0}$

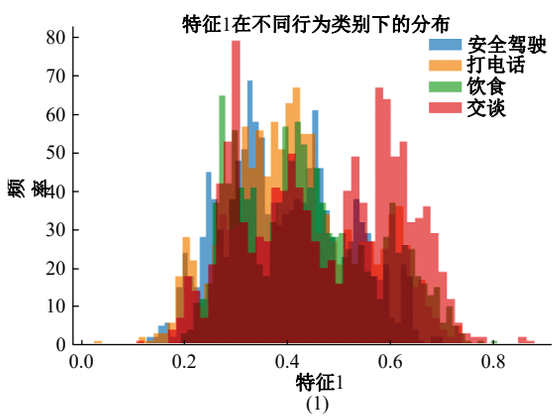
本文初步确定人体结构的13个骨架关节矢量角向量和8个骨架模比向量作为驾驶员动作特征,共21个ADFs,如下所示:

$$\mathrm{ADF}s=[\theta_1,\cdots,\theta_{13},r_1,\cdots,r_8] \tag{5}$$

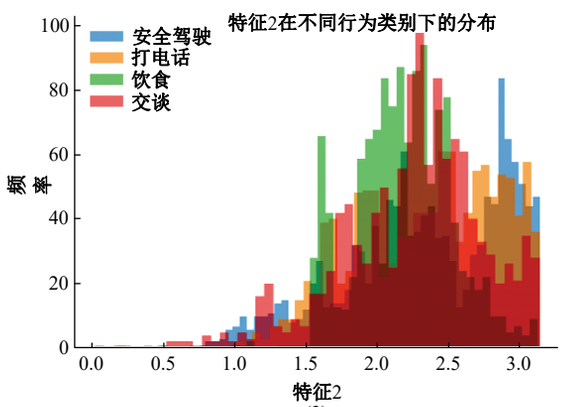
1.4 驾驶员骨架动作特征ADF优选

驾驶员骨架动作特征向量的数据分布图如图6所示,该特征向量共计21维,其分别是13个骨架关节矢量角向量和8个骨架模比向量。

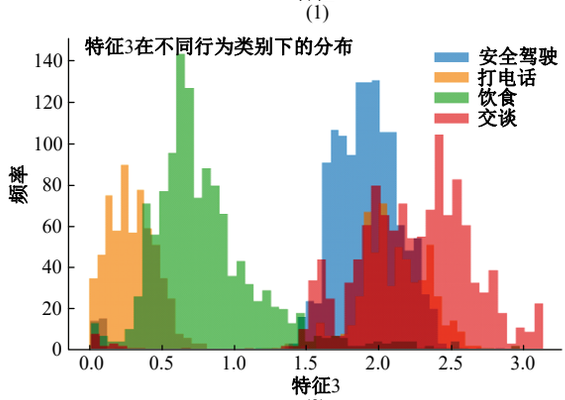
分析发现,部分驾驶员骨架动作特征向量难以有效区分不同驾驶行为,某些特征存在数据分布重叠现象,导致行为识别区分度不足。因此,本文将特征分为两类:显著区分特征(图6(g))和非显著区分特征(图6(a))。



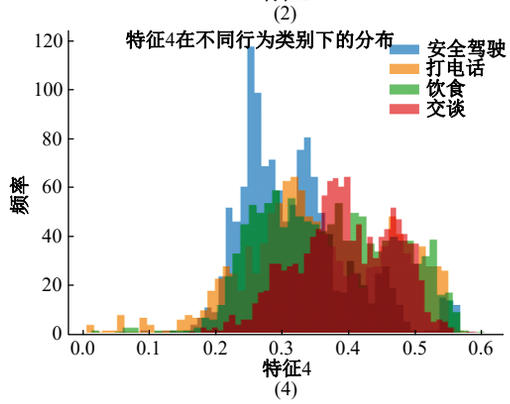
(1)



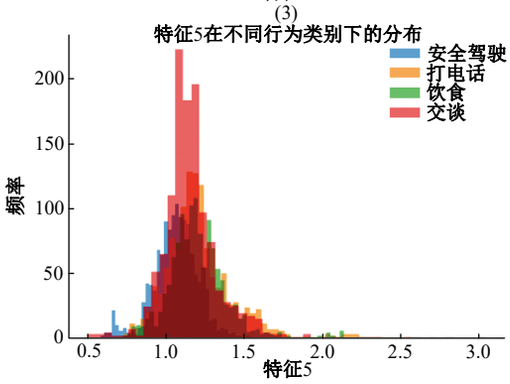
(2)



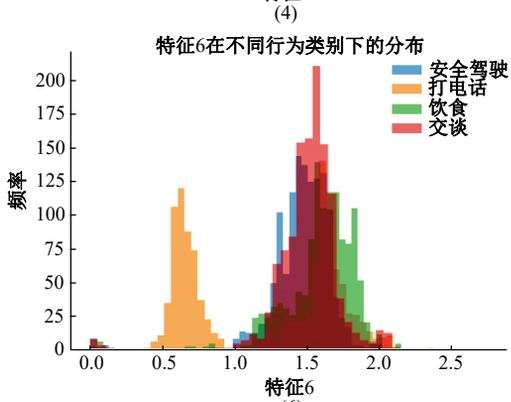
(3)



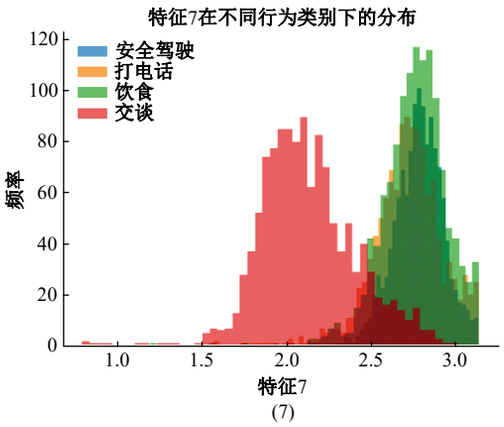
(4)



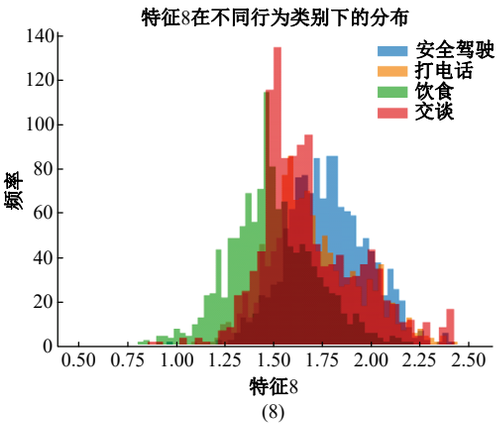
(5)



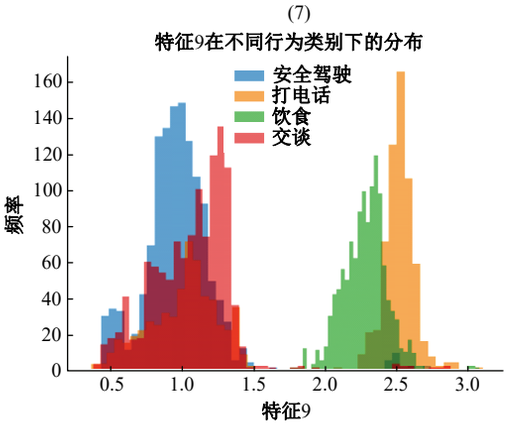
(6)



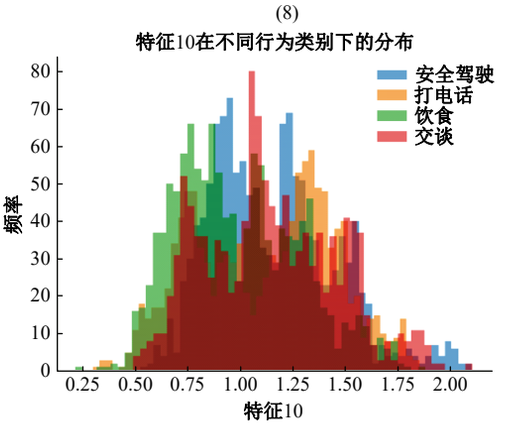
(7)



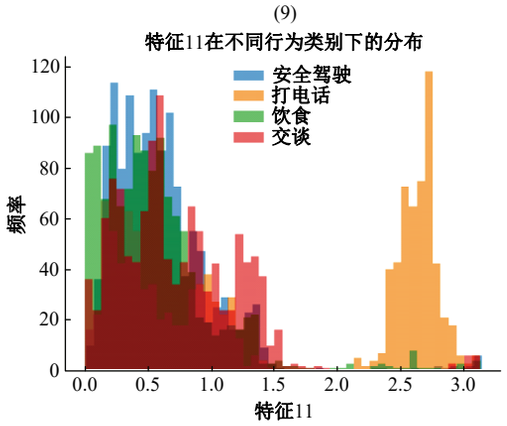
(8)



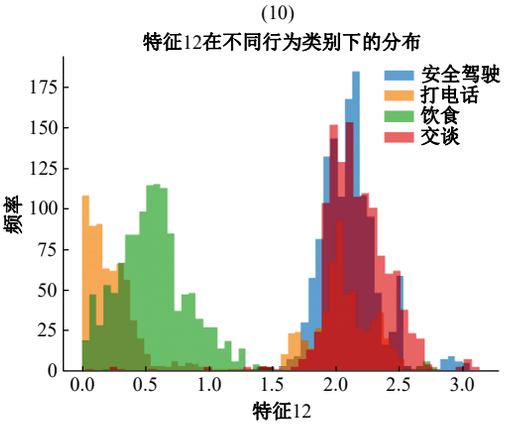
(9)



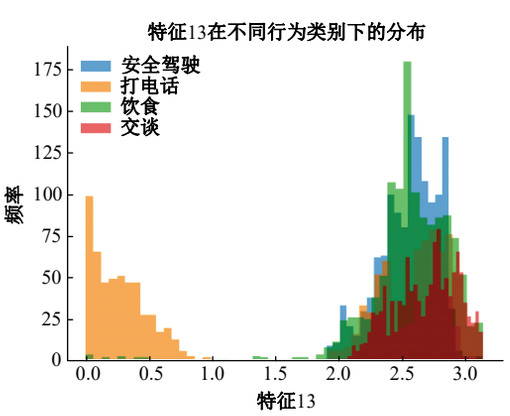
(10)



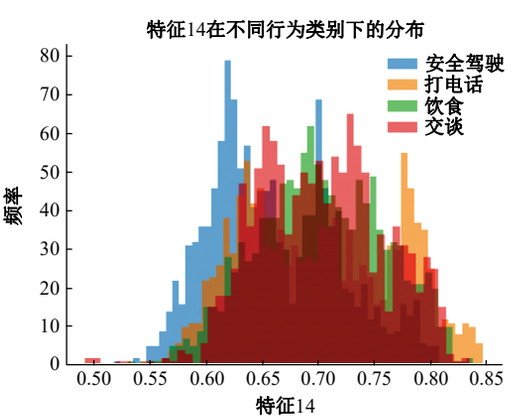
(11)



(12)



(13)



(14)

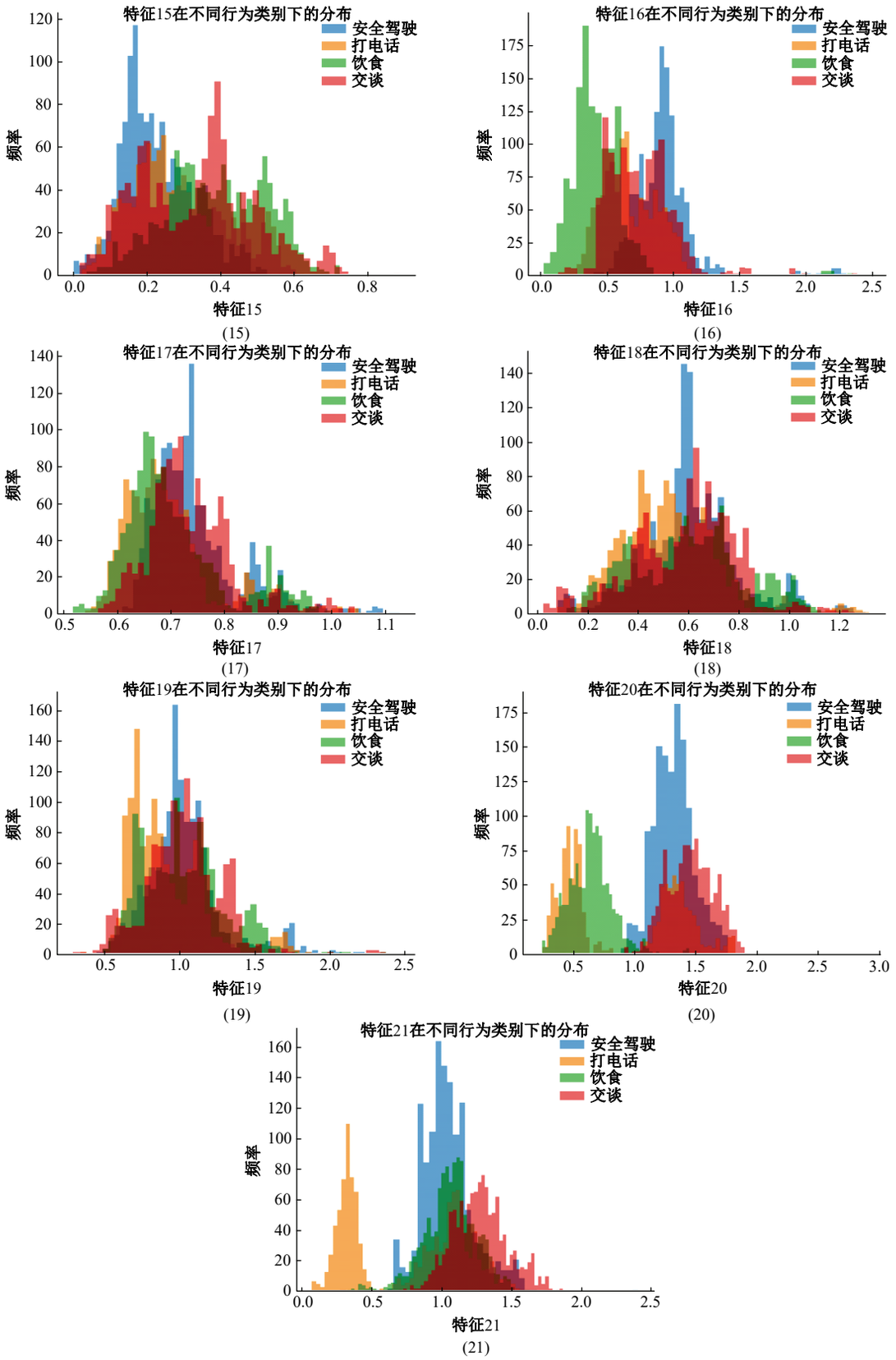


图 6 驾驶员骨架特征向量数据分布

Fig. 6 Data distribution of driver skeleton feature vectors

为提升模型效率与鲁棒性，研究筛选出7个具有显著区分度的特征向量：特征编号3、6、7、9、11、20、21。最终选定5个骨架关节矢量角向量和2个骨架模比向量组成7维ADFs特征向量。

$$\mathrm{ADF}s=[\theta_1,\cdots,\theta_5,r_1,r_2] \tag{6}$$

1.5　基于ResNet50的驾驶分心行为分类网络

为提高网络模型的训练效率和性能，加速收敛速度，本文构建基于ResNet50的驾驶分心行为分类模型。ResNet(Residual Network)[18]通过引入残差模块，添加了“捷径连接”或“跳过连接”，允许网络训练更深结构，而不会引发梯度消失或梯度爆炸。另外，ResNet还采用了全局平均池化(Global average pooling)替代全连接层，减少了参数数量，降低了过拟合的风险，且使网络对输入图像的尺寸变化更加鲁棒。

为提高模型的泛化性，本文在Resnet50模型的末尾处添加了Dropout和Dense层。Dropout层通过在训练过程中随机省略一些神经元，打破了神经元之间的相互依赖关系，使得模型更加健壮，并降低过拟合的风险。Dense层则通过采用L2正则化技术来减小权重的大小，进一步防止模型过度拟合训练集。

结合本文研究情况，即应用OpenPose提取驾驶员骨架特征，并根据骨架特征对驾驶行为予以分类与辨识，故选用交叉熵损失函数。交叉熵损失函数可以用来衡量模型预测输出概率P与真实标签分布Y之间的差异，其表达式如下：

$$Loss(Y,P)=-\frac{1}{N}\sum_{i=0}^{N-1}\sum_{k=0}^{K-1}y_{ik}\log(p_{ik}) \tag{7}$$

式中：N为样本总数；K为类别总数；i为样本索引($i=0,1,\cdots,N-1$)；k为类别索引($k=0,1,\cdots,K-1$)；y_{ik}为第i个样本在第k类的真实标签；p_{ik}为模型预测的第i个样本属于第k类的概率。

2　实验与分析

2.1　数据集

2.1.1　StateFarm数据集

State farm distracted driver detection是由Kaggle竞赛平台发布的驾驶行为数据集。该数据集收录了车辆行驶过程中驾驶员各类分心行为的图像，采集自26名不同体型、肤色的受试者，总计包含22 424张图像。如图7所示，数据集涵盖9类典型分心行为及正常驾驶行为的图像样本。该数据集仅提供驾驶行为分类标签，未包含人体语义标注信息，因此通常适用于端到端驾驶行为分析任务。

2.1.2　AUC数据集

AUC Distracted Driver Detection是功能相似的驾驶行为数据集，数据规模为32 718张图像，同样包含9类分心行为与正常驾驶的图像，如图8所示。该数据集同样仅提供行为分类标签，故同样仅适用于端到端驾驶行为分析。

2.2　软硬件平台

本文所述模型基于Python实现，利用PyTorch构建深度学习框架，具体软硬件参数如表3所示。

2.3　评价指标

驾驶分心行为识别本质上是多分类问题。本文采用多分类问题中常用的评价指标准确率(Accuracy)Acc来评价驾驶分心行为检测性能，其表达式如下：

$$Acc(\hat{y},y)=\frac{1}{N}\sum_{i=0}^{N-1}\mathbb{I}(\hat{y}_i=y_i) \tag{8}$$

式中：$\hat{y}_i$为第i个样本的预测类别；y_i为真实类别；$\mathbb{I}$为指示函数，当括号内条件成立时取值为1，否则为0；N为样本总数。

2.4　模型训练评估

训练过程分为两个阶段：第一阶段通过OpenPose从分心驾驶数据集中提取人体骨架姿态数据，构建ResNet分类网络的训练集；第二阶段训练ResNet分类网络：模型权重采用ImageNet预训练参数，在Ubuntu 18.04系统下使用2块RTX 3080 GPU并行训练。关键训练参数为：每块GPU的BatchSize为32，总训练轮次30轮，初始学习率0.01，并在第10轮和第20轮进行衰减。

通过实验评估不同Dropout取值(0、0.1、0.2、0.3、0.4、0.5)对识别性能的影响，最终确定最优Dropout超参数。实验结果如图9所示，Dropout值为0.4时模型取得最高测试准确率(95.67%)。因此，本文的驾驶分心行为识别模型最终采用Dropout=0.4的改进ResNet50作为分类网络。

图10展示了Dropout值设为0.4时模型训练与验证准确率随训练轮次的变化趋势。从收敛过



图7 StateFarm数据集部分图片

Fig. 7 Some images of the StateFarm dataset



图8 AUC数据集部分图片

Fig. 8 Some images of the AUC dataset

表3 实验软硬件参数

Table 3 Software and hardware parameters of the experiment

	平台	参数
硬件	服务器GPU	RTX3080*2(显存10G)
	本地GPU	RTX1070ti*1(显存8G)
软件+	服务器操作系统	Ubuntu18.04
	本地操作系统	Windows10
	开发语言	Python 3.8.5
	深度学习框架	Pytorch 1.10.0 - GPU
	图像处理库	OpenCV 4.4.0
	CUDA版本	10.2
	CUDNN版本	7.6.5

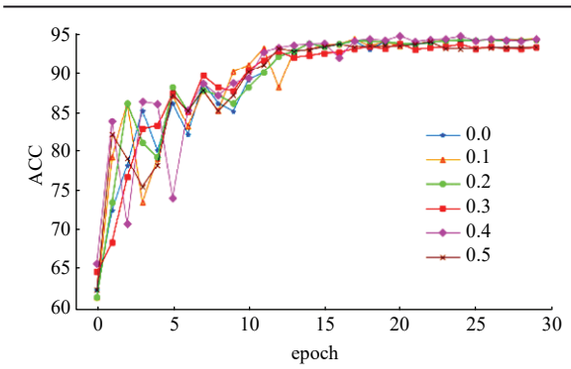


图9 采用不同Dropout参数试验结果

Fig. 9 Results of trials using different Dropout parameters

程看,训练前期,训练集与验证集准确率均快速提升,体现模型对骨架特征的快速学习能力;训练后期,曲线趋于平稳,训练与验证准确率稳定在95%左右。整体而言,网络在30轮内实现高效收敛,且训练与验证准确率差距始终较小,证明本文模型对骨架特征兼具拟合能力与稳定性,为驾驶分心行为精准识别提供了可靠支撑。

图11为基于骨架特征的驾驶分心行为识别模型的混淆矩阵,其中蓝色对角线区域代表行为预测正确的情况,对角线颜色越深表明识别精度越高。矩阵中c_0、c_1、c_2、c_3分别对应正常驾驶、打电话、与乘客交谈、饮食四类驾驶行为。从图中数值分布可见:c_0(正常驾驶)对角线数值达0.97,c_1(打电话)为0.95,c_2(与乘客交谈)为0.96,c_3(饮食)为0.94,四类行为的预测正确率均稳定在94%以上;非对角线区域数值普遍低于0.05(如c_0行仅0.03误判至c_1),说明不同驾驶行为间的混淆程度极低。整体上,深色主导的对角线与极小的非对角线数值,直观验证了模型对各类驾驶行为的分类稳定性与类别区分能力。

2.5 消融实验

消融实验通过增减模型组件明确各模块对识别性能的独立贡献。本节针对"骨架特征提取-数据处理与特征构建-动作识别"三模块设计消融实验,在原始数据集与经人体语义分割预处理的数据集上进行对比实验,以量化模块各自的贡献与

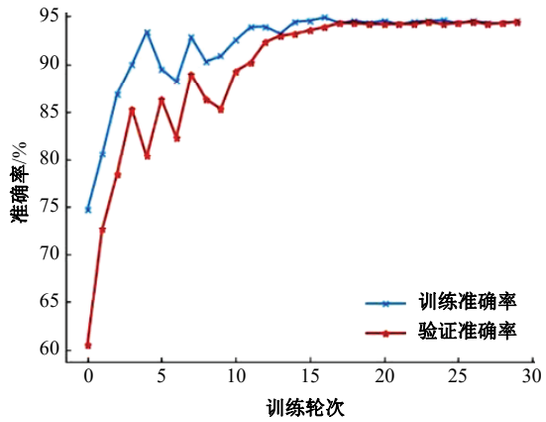


图 10　Dropout 值为 0.4 的 ResNet50 网络训练结果

Fig. 10　Training result of ResNet50 network with Dropout value of 0.4

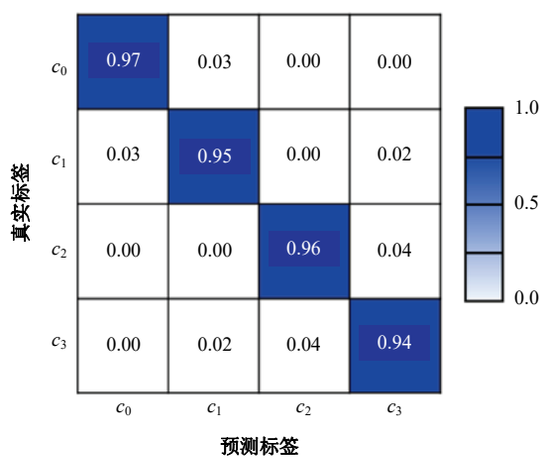


图 11　基于骨架特征的驾驶分心行为识别模型的混淆矩阵

Fig. 11　Confusion matrix of skeleton-feature-based distracted driving behavior recognition model

模块间的协同效应，如表 4 所示。

实验(1)在未做人体语义分割的原始数据集上单独测试动作识别模块，精度为 75.84%。该实验用于验证动作识别模块的“基础识别能力”——在含背景噪声、人体信息未分离的原始数据中，模块对分心行为的直接判别性能。

实验(2)在经人体语义分割的数据集上测试同一动作识别模块，精度提升至 93.59%。性能提升表明：动作识别模块对“去噪后数据”(人体语义分割减少背景干扰)的适应性更强，侧面体现模块对纯净姿态信息的识别潜力。

实验(3)组合“骨架特征提取模块＋动作识别模块”(不含数据处理与特征构建模块)，测试精度达 95.95%。该实验直接验证骨架提取模块的核心价值：通过 OpenPose 提取的姿态特征(关节位置、连接关系等)，为动作识别提供了比图像端到端识别更具区分性的关键信息，成为性能突破的核心支撑。

表 4 消融实验结果

Table 4　Results of ablation experiments

实验编号	模块组合	描　述	*Acc*/%
(1)	模块三*	原始数据集；直接进行动作识别	75.84
(2)	模块三	人体语义分割数据集；直接进行动作识别	93.59
(3)	模块一＋模块三	人体语义分割数据集；经骨架特征提取进行动作识别	95.95
(4)	模块一＋模块二＋模块三*	原始数据集；经骨架特征提取、数据处理与特征构建进行动作识别	91.24
(5)	模块一＋模块二＋模块三	人体语义分割数据集；经骨架特征提取、数据处理与特征构建进行动作识别	97.33

注：模块一为骨架特征提取模块，模块二为数据处理与特征向量构建模块，模块三为动作识别模块，*为未经人体语义分割的驾驶分心行为数据集。

实验(4)测试“骨架提取＋数据处理与特征构建＋动作识别”全模块在原始数据集的性能，精度为 91.24%。虽略低于实验(3)，但体现全模块协同在复杂环境(原始数据含噪声)下的泛化能力——数据处理模块对原始骨架特征的优化，与动作识别模块的分类能力结合，使系统在未预处理数据中仍保持高鲁棒性。

实验(5)全模块在经语义分割数据集上测试，精度达 97.33%。该结果验证了三模块协同的极致性能：骨架提取提供精准姿态特征、数据处理模块强化特征判别性、动作识别模块实现精准分类，三者协同使系统在优质特征与精准分类的共同作用下达到最优表现。

通过消融实验可明确三模块的技术价值：骨架特征提取模块突破图像端到端识别的局限性，为动作识别提供精准姿态特征(如关节空间关系)，是性能提升的基础；数据处理与特征构建模块对原始骨架特征进行判别性增强(如结构向量、角度向量、模比向量的构建与优选)，使特征更适配后续分类任务；动作识别模块基于优化后的特征实现细粒度分类，完成从姿态特征到分心行为标签的映射。三模块的协同设计，是本方法在复杂驾驶场景中保持高鲁棒性与高精度的核心逻辑。

2.6　真实场景验证

为评估本文方法在真实驾驶环境中的适用性

和鲁棒性,采集了多段不同光照条件、不同背景、不同角度下的驾驶员行为视频。将未经处理的视频直接输入至训练好的模型进行驾驶员行为识别。图12~图15展示了模型在真实场景视频中的部分识别结果,可以看出,尽管存在复杂的光照变化、背景干扰及角度变化,模型仍能稳定地提取

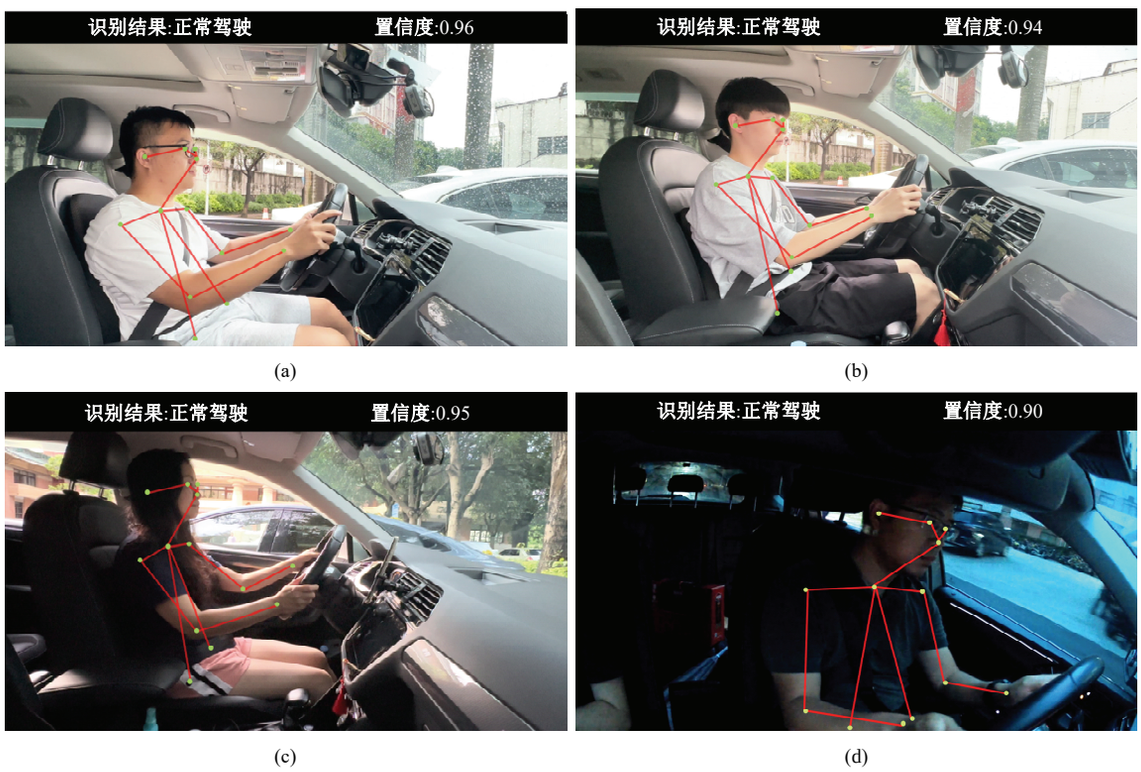


图12 识别结果:正常驾驶

Fig. 12 Identification result: normal driving



图13 识别结果:打电话

Fig. 13 Identification result: phone

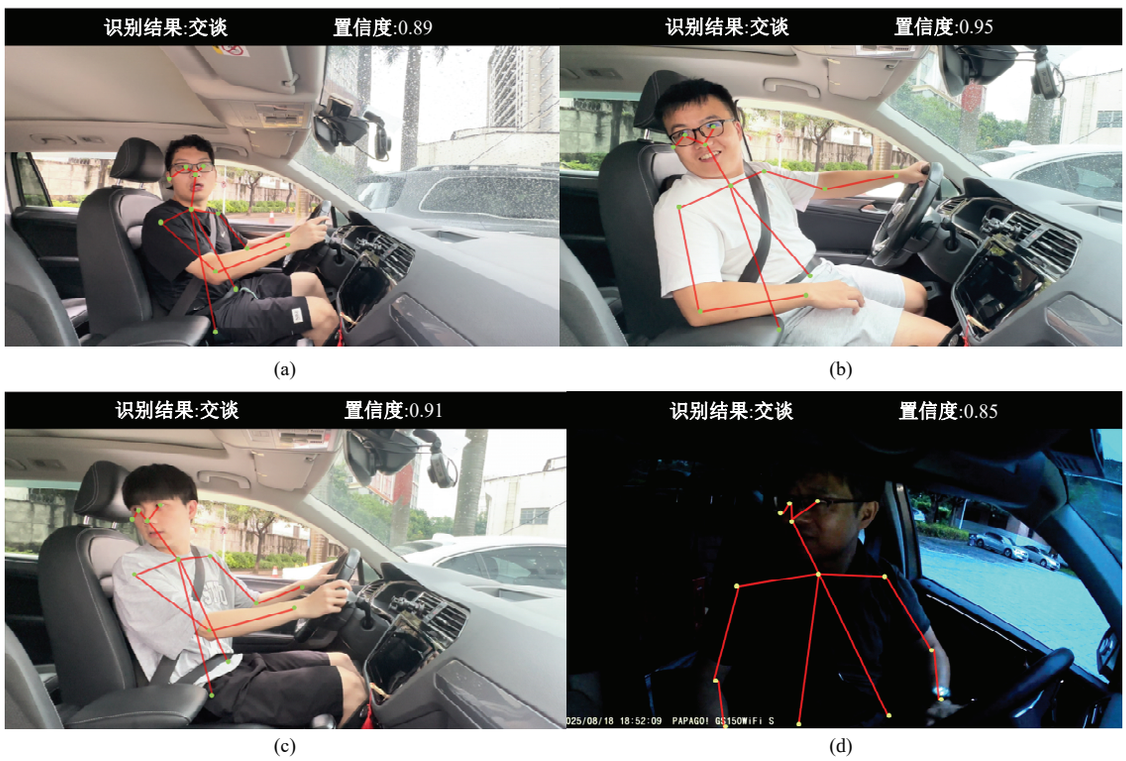


图 14　识别结果：交谈

Fig. 14　Identification result: talk

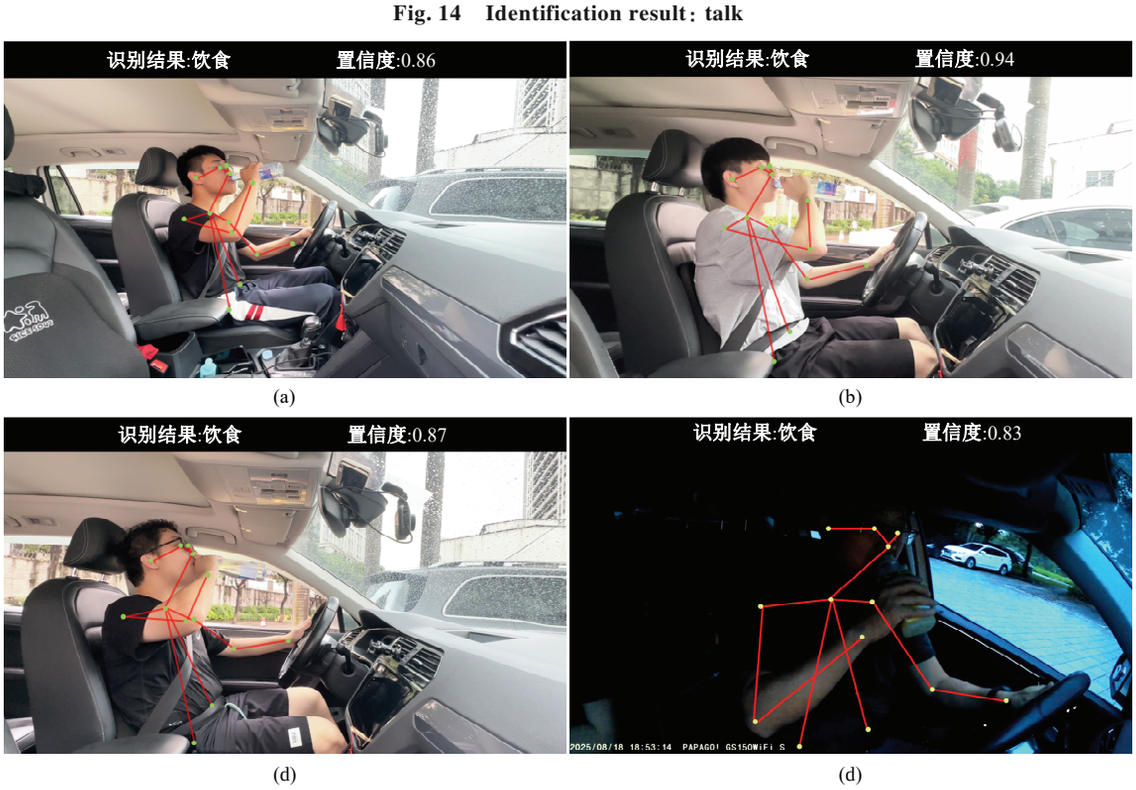


图 15　识别结果：饮食

Fig. 15　Identification result: drink or eat

出驾驶员的骨架特征，并准确地输出相应的驾驶行为类别标签。这表明，本文基于骨架特征的识别方法对实际场景中的光照变化、背景干扰与角度变化具有较强的鲁棒性。同时，尽管本研究基于二维骨架进行分析，但在实际验证中，模型对驾驶员在三维空间中的典型分心动作表现出良好的

识别能力。二维骨架特征足以捕捉本研究限定的分心驾驶动作在关键关节点上的核心动态模式,从而实现有效判别。

验证结果表明,本文提出的基于骨架特征的驾驶分心行为识别方法在脱离受控的公开数据集环境后,依然具备良好的鲁棒性和实用性,能有效应对真实驾驶场景中的光照、背景和角度变化。这验证了本方法在实际车载系统中的应用潜力。

3 结束语

本文针对驾驶分心行为识别中传统视觉方法易受光照和背景干扰的局限性,提出了一种基于骨架特征的深度学习方法,通过创新性的特征提取与模型优化策略,显著提升了复杂驾驶环境下的识别精度与鲁棒性。具体而言,基于OpenPose算法提取驾驶员上半身13个关键骨架关节点,构建了融合骨架结构向量、关节矢量角向量和模比向量的多维度特征描述符,并通过特征优选策略筛选出7维显著区分性特征,有效降低了冗余并增强动作表征能力。在模型设计上,改进ResNet50网络结构,集成Dropout层和Dense层抑制过拟合,显著提升了模型的泛化性能。实验在StateFarm和AUC公开数据集上验证了方法的有效性:经人体语义分割预处理后,完整模型在数据集上准确率达97.33%;消融实验进一步证明了骨架特征提取、数据处理与特征构建、动作识别三模块的协同作用。最后,通过真实采集的视频验证了本研究方法在实际应用中的鲁棒性与实用性。

本文的创新性在于:通过骨架特征进一步克服了传统驾驶分心行为识别方法对光照和背景的敏感性;提出的多维度特征描述符与优选策略增强了动作区分能力;提出的驾驶分心行为识别完整架构为实时高精度识别提供了可行方案。需要指出的是,当前方法主要依赖显性姿态变化进行识别,对于与乘客交谈等无显著肢体动作的行为,其骨架特征与正常驾驶相似,存在识别局限。未来工作将聚焦于轻量化网络部署、实时性优化,并探索融合面部表情等多模态信息,以提升对更广泛的驾驶分心行为的识别能力。研究成果将为车载监控系统与智能交通安全预警提供可靠的技术支持。

参考文献:

[1] 张丽霞, 刘涛, 潘福全, 等. 驾驶员因素对道路交通事故指标的影响分析[J]. 中国安全科学学报, 2014, 24(5): 79-84.
Zhang Li-xia, Liu Tao, Pan Fu-quan, et al. Analysis of effects of driver factors on road traffic accident indexes[J]. China Safety Science Journal, 2014, 24(5): 79-84.

[2] Zhang Y, Chen Y, Gao C. Deep unsupervised multimodal fusion network for detecting driver distraction[J]. Neurocomputing, 2021, 421: 26-38.

[3] 陈仁祥, 胡超超, 胡小林, 等. 基于改进YOLOv5的驾驶员分心驾驶检测[J]. 吉林大学学报: 工学版, 2024, 54(4): 959-968.
Chen Ren-xiang, Hu Chao-chao, Hu Xiao-lin, et al. Driver distracted driving detection based on improved YOLOv5[J]. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(4): 959-968.

[4] Baheti B, Talbar S, Gajre S. Towards computationally efficient and realtime distracted driver detection with MobileVGG network[J]. IEEE Transactions on Intelligent Vehicles, 2020, 5(4): 565-574.

[5] 赵霞, 李朝, 付锐, 等. 基于深度卷积-Tokens降维优化视觉Transformer的分心驾驶行为实时检测[J]. 汽车工程, 2023, 45(6): 974-988.
Zhao Xia, Li Zhao, Fu Rui, et al. Real-time detection of distracted driving behavior based on deep convolution-tokens dimensionality reduction optimized visual Transformer model[J]. Automotive Engineering, 2023, 45(6): 974-988.

[6] Liu S, Wang Y, Yu Q, et al. CEAM-YOLOv7: improved YOLOv7 based on channel expansion and attention mechanism for driver distraction behavior detection[J]. IEEE Access, 2022, 10: 129116-129124.

[7] 陈军, 张黎, 周博, 等. 基于级联卷积神经网络的驾驶员分心驾驶行为检测[J]. 科学技术与工程, 2020, 20(14): 5702-5708.
Chen Jun, Zhang Li, Zhou Bo, et al. Driver distracted driving behavior detection based on cascaded convolutional neural network[J]. Science Technology and Engineering, 2020, 20(14): 5702-5708.

[8] Ping P, Huang C, Ding W, et al. Distracted driving detection based on the fusion of deep learning and causal reasoning[J]. Information Fusion, 2023, 89: 121-142.

[9] Alkinani M H, Khan W Z, Arshad Q, et al. HSDDD: a hybrid scheme for the detection of distracted driving through fusion of deep learning and handcrafted features[J]. Sensors, 2022, 22(5): s22051864.

[10] 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.
Yin Zhi-shuai, Zhong Shu, Nie Lin-zhen, et al. Distracted driving behavior detection based on human pose estimation[J]. China Journal of Highway and Transport, 2022, 35(6): 312-323.

[11] 贺宜, 鲁曼可, 高嵩, 等. 基于MobileViT-CA模型的营运车辆驾驶人分心行为检测[J]. 中国公路学报, 2024, 37(1): 194-204.
He Yi, Lu Man-ke, Gao Song, et al. Distracted behavior detection of commercial vehicle drivers based on the MobileViT-CA model[J]. China Journal of Highway and Transport, 2024, 37(1): 194-204.

[12] 张斌, 付俊怡, 夏金祥. 基于类间距优化的分心驾驶行为识别模型训练方法[J]. 汽车工程, 2022, 44(2): 225-232.
Zhang Bin, Fu Jun-yi, Xia Jin-xiang. A metric space optimized method for driver distraction recognition model training[J]. Automotive Engineering, 2022, 44(2): 225-232.

[13] Sullman M J M. An observational study of driver distraction in England[J]. Transportation Research Part F: Traffic Psychology and Behaviour, 2012, 15(3): 272-278.

[14] Sullman M J M, Prat F, Tasci D K. A roadside study of observable driver distractions[J]. Traffic Injury Prevention, 2015, 16(6): 552-557.

[15] Prat F, Planes M, Gras M E, et al. An observational study of driving distractions on urban roads in Spain [J]. Accident Analysis and Prevention, 2015, 74: 8-16.

[16] Kountouriotis G K, Spyridakos P, Carsten O M J, et al. Identifying cognitive distraction using steering wheel reversal rates[J]. Accident Analysis and Prevention, 2016, 96: 39-45.

[17] Cao Z, Hidalgo G, Simon T, et al. OpenPose: realtime multi-person 2D pose estimation using part affinity fields[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(1): 172-186.

[18] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, 2016: 770-778.