

doi:10.19562/j.chinasae.qcgc.2020.08.005

基于图卷积网络的多信息融合 驾驶员分心行为检测^{*}

白中浩^{1,2},王轲宇¹,张林伟¹

(1. 湖南大学,汽车车身先进设计制造国家重点实验室,长沙 410082;

2. 福建工程学院,福建省汽车电子与电驱动重点实验室,福州 350118)

[摘要] 为减少由驾驶员分心造成的交通事故,并检测驾驶员在自动驾驶情况下的分心状态以判断驾驶员是否有接管车辆的能力,提出了一种基于图卷积的多信息融合驾驶员分心行为检测方法。通过分析驾驶员分心行为和姿态特征,设计了驾驶员姿态估计图,基于图卷积网络对驾驶员姿态估计图进行特征提取,使用全连接层对所提取特征进行行为分类,同时融合手机等关键物体信息对驾驶员分心行为进行再判断。实验结果表明,本文提出的方法在 SrateFarm 数据集和自制数据集上分别达到了 90% 和 93% 的准确率,检测速度约为 20 帧/s,准确性和实时性均达到检测要求。

关键词: 驾驶员分心;姿态估计;行为识别;图卷积网络

Driver Distraction Behavior Detection with Multi-information Fusion Based on Graph Convolution Networks

Bai Zhonghao^{1,2}, Wang Yunyu¹ & Zhang Linwei¹

1. Hunan University, State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Changsha 410082;

2. Fujian University of Technology, Fujian Key Laboratory of Automotive Electronics and Electric Drive, Fuzhou 350118

[Abstract] In order to reduce the traffic accidents caused by driver's distraction, and detect the distraction state of the driver during automatic driving for judging whether the driver has the ability to take over the vehicle, a multi-information fusion driver distraction behavior detection method based on graph convolution is proposed. By analyzing the distraction behavior and posture features of the driver, the posture estimation graph of the driver is designed. The feature extraction of driver's posture estimation graph is conducted based on graph convolution network, the features extracted are classified by using full connected layer, and the driver's distraction behaviors are judged again by the fusion of key objects like mobile phone. Experiment results show that the method proposed can achieve 90% and 93% accuracy on SrateFarm dataset and the self-made dataset respectively with a detection speed of 20 frames per second, meeting the detection requirements of accuracy and real-time performance.

Keywords: driver distraction; posture estimation; behavior recognition; graph convolution network

前言

近年来,有超过 1/5 的交通事故是由驾驶员分心导致的^[1]。此外,在实现完全自动驾驶(L5 级)之前,依然要求驾驶员保持专注,随时准备在紧急情况

下重新控制转向盘^[2-3]。因此,通过对驾驶员行为进行实时监控,减少驾驶员分心,对于交通安全来说尤为重要。

解决驾驶员驾驶分心状态问题的一种方法是开发驾驶员监测系统,在这样的监测系统中,正确检测驾驶员分心行为至关重要,这也是本文的重点。驾

^{*} 国家自然科学基金(51621004,51475153)和福建工程学院科研创新平台开放基金(KF-X18001)资助。

原稿收到日期为 2019 年 9 月 26 日,修改稿收到日期为 2019 年 11 月 22 日。

通信作者:白中浩,教授,E-mail:baizhonghao@163.com。

驶员分心检测可以检测驾驶员分心行为,并识别造成分心的原因,进而警告驾驶员不要分心或采取预防措施。另一方面,在紧急情况下,有助于判断驾驶员是否具有接管车辆的能力,可通过警告驾驶员停止分心行为,集中注意力投入到车辆驾驶中,减少意外事故的发生。这使驾驶员分心检测系统成为车辆智能化的一个重要组件。此外,分心检测还可用于驾驶员意图分析中,为实现完全自动驾驶(L5级)奠定基础。

近年来,众多学者使用压力传感器、接近传感器、深度相机、单目摄像头等对驾驶员分心行为进行检测^[4-8]。其中,从成本和使用便利性等方面考虑,单目视觉摄像头是目前使用最广泛的传感器。目前,驾驶员分心检测算法大多基于对驾驶员局部身体部位的感知,Seshadri等^[9]通过检测驾驶员头部姿态的变化,使用原始像素和梯度方向直方图(HOG)特征与AdaBoost分类器相结合,对使用手机这类分心行为进行检测,在自制数据集上获得了较好的效果。Abouelnaga等^[10]模仿Kaggle上分心驾驶比赛数据集,制作了类似的图片数据集,通过人脸、手部定位和皮肤分割排除冗余信息,然后基于卷积神经网络和遗传算法对驾驶员分心行为进行分类,取得了84.64%的准确率^[11]。Baheti等^[12]提出了基于改进VGG-16的算法,不仅可检测驾驶员分心行为,还可识别出原因,获得了更高的识别精度。Ou等^[13]通过在ImageNet上预训练的ResNet-50模型进行迁移学习,以解决驾驶员分心行为数据不足的问题,并证明了其在光照变化下具有较强的鲁棒性。

目前,由于数据集的限制,大多算法都是基于图片的驾驶员行为识别,这是对实时检测驾驶员分心行为的一种挑战。首先,本文中模仿Kaggle数据集的分类,从拍摄角度、驾驶员体型、光照、干扰物等方面,制作了驾驶员分心驾驶视频片段数据集。然后,基于人体姿态,通过对驾驶员行为分析,设计了驾驶员姿态估计图。最后,基于图卷积网络(graph convolution network, GCN),提出了一种融合驾驶员骨架信息与关键物体信息的驾驶行为检测算法,在SrateFarm数据集和自制数据集上均取得了较好的效果。

1 图卷积网络

图是一种数据形式,不像图像、视频和信号数据那样的网格化数组结构。一个图的顶点(边)与顶

点(边)之间的邻居数是不同的,且对于两个顶点(边)之间的任何给定连接都不需要几何解释。它可用于表示社交网络、通信网络和蛋白分子网络等,图中的节点表示网络中的个体,连边表示个体之间的连接关系。将深度学习中常用于图像的卷积神经网络应用到数据上,Kpif等^[14]提出了图卷积网络(GCN),它为通过骨架信息来进行动作识别提供了一个崭新的思路。Yan等^[15]第一个将图卷积网络应用于动作识别中,提出了ST-GCN行为分类的方法,并在两个大型动作数据集上表现出极好的效果。Zhang等^[16]设计了图边缘卷积神经网络(GECNN),并融合了图边缘卷积神经网络与图卷积神经网络的结果,在图卷积网络基础上取得了不错的提升。

与二维卷积网络不同,图卷积没有固定空间顺序排列的领域像素,须根据节点周围的领域在图标记过程中定义空间顺序。人体不同关节点的运动大致可分为向心运动、离心运动和静止,因此本文中有关节点 v_{ii} 的领域集 $B(v_{ii})$ 划分为3个子集:第1部分连接空间位置上比该关节点更远离人体颈部的相邻关节点,包含离心运动特征;第2部分连接更靠近颈部相邻关节点,包含向心运动特征;第3部分连接自身关节点,包含静止特征。将1个图分为了3个子图,包含3种不同的运动特征,卷积核也从1个变成了3个,3个卷积核分别表达了不同的特征,使用每个卷积核对每个图分别卷积,最后进行加权平均得到最终的结果。

根据Yan等^[15]的研究,在单帧驾驶员图卷积模型中,将卷积神经网络模型类比到图卷积网络模型中:

$$f_{\text{out}}(v_{ii}) = \sum_{v_{ij} \in B(v_{ii})} \frac{1}{Z_{ii}(v_{ij})} f_{\text{in}}(v_{ij}) \cdot w(l_{ii}(v_{ij})) \quad (1)$$

式中: $l_{ii}(v_{ij}) = (v_{ii}, v_{ij})$ 对应不同子集; $Z_{ii}(v_{ij}) = |\{v_{ik} | l_{ii}(v_{ik}) = l_{ii}(v_{ij})\}|$ 为相应子集的基数。

在空间上,使用GCN来提取相邻关节的局部特征,在时间上,采用时间卷积网络(temporal convolutional network, TCN)来学习时间中关节变化的局部特征。将空间图卷积网络扩展到时间上,使邻域信息包含关节点的时间信息。

$$B(v_{ii}) = \{v_{qj} | d(v_{qj}, v_{ii}) \leq 3, |q-t| \leq \lfloor \Gamma/2 \rfloor\} \quad (2)$$

$$l_{ST}(v_{qj}) = l_{ii}(v_{qj}) + (q-t + \lfloor \Gamma/2 \rfloor) \times 3 \quad (3)$$

式中: Γ 为相关联的帧数范围; q 为相邻某一帧; $l_{ST}(v_{qj})$ 为 v_{ii} 的邻域。

本文采用图卷积神经网络单层最终形式:

$$f_{out}=\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}f_{in}W$$

(4)

式中: $\tilde{A}=\mathbf{A}+\mathbf{I}$ 为添加自连接矩阵, \mathbf{A} 为邻接矩阵,表示图中节点的连接情况的矩阵; \tilde{D} 为度矩阵, $\tilde{D}_{ii}=\sum_j \tilde{A}_{ij}$,表示该节点相连的边数量的矩阵; W 为待训练参数。

2 算法概述

本文的算法是一个分类问题,将驾驶员行为分为正常驾驶行为和分心驾驶行为,在分心驾驶行为中又具体分为玩手机、喝水等 9 种分心行为。该算法的流程主要分为驾驶员姿态估计、物体检测和行为识别,行为识别算法通过图卷积提取驾驶员动作特征,同时融合关键物体信息的方法,获得最终分类结果。本文中的驾驶员行为识别算法整体框架如

- 图 1 所示。
- (1) 首先输入为包含驾驶员的视频片段,通过 Openpose^[17]对视频中驾驶员进行姿态估计,获取每帧图像驾驶员的关节点信息,包含点的位置坐标和置信度。由于不同帧下关节点坐标差异很大,须对驾驶员关节点坐标进行归一化处理,将不同骨架序列缩放至统一的尺度特征。
- (2) 然后根据得到的骨架序列图,按照时空姿态运动规律,构建驾驶员时空图,交替地使用图卷积网络和时间卷积网络,对空间和时间维度进行变换,最后使用平均池化进行下采样,并用全连接层对特征进行分类。
- (3) 同时使用 Yolo-V3^[18]对视频片段中感兴趣物体进行识别,获得物体类别信息和位置信息,通过物体信息对驾驶员行为进行再判断,得到最终分类结果。

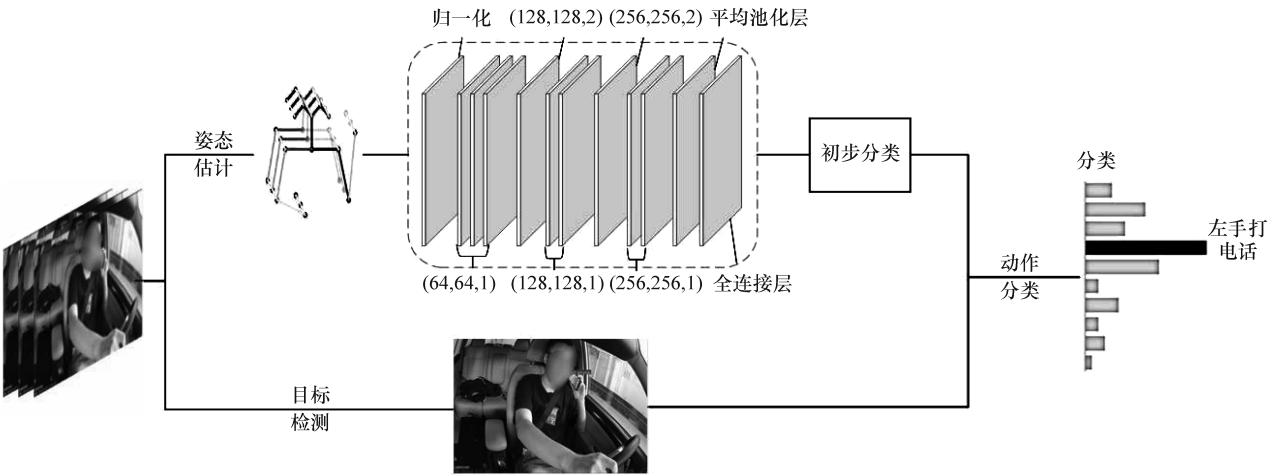


图 1 算法流程图

2.1 驾驶员姿态估计图

动作识别算法主要基于驾驶员的姿态特性。人体的骨架和关节轨迹对于照明变化和场景变化具有较强的鲁棒性,且由于高度精确的深度传感器或姿势估计算法,不容易受图像人数、人员体型、外观和背景干扰的影响。本文中使用的姿态估计的方法是 Openpose, 该方法是美国卡耐基梅隆大学(CMU)基于卷积神经网络和监督学习并以 caffe 为框架开发的开源库,具有极好的鲁棒性和实时性。在 Openpose 中,训练的人体 COCO 模型(图 2 左)有 18 个关节点。

考虑到驾驶员在正常驾驶情况下,摄像头一般只能捕捉到驾驶员上半身图像,且驾驶员动作重要信息大多表现在上半身。同时由于驾驶环境导致

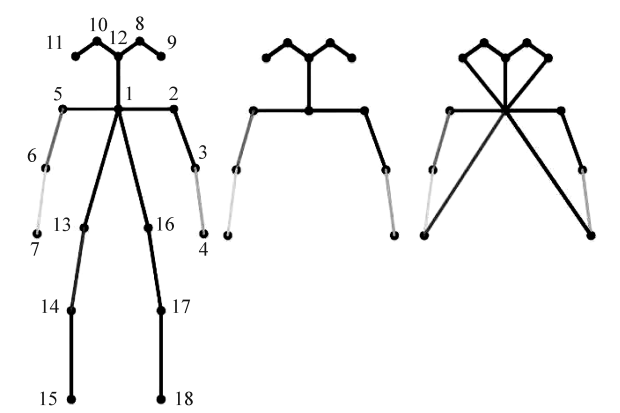


图 2 驾驶员姿态估计图

遮挡,下半身部分关键的检测不可避免地会受到较强的干扰噪声(错误检测或未检测到),从而对驾

驶员的分心行为检测造成巨大的干扰。为消除下半身对肢体捕捉的影响,本文设计的驾驶员姿态估计图只考虑对上半身12个关节点进行姿态估计(图2):①颈部,②右肩,③右肘,④右手腕,⑤左肩,⑥左肘,⑦左手腕,⑧右眼,⑨右耳,⑩左眼,⑪左耳,⑫鼻子。

人类的运动起源于关节,关节连接着两块骨头,而两块骨头的形状、长度和位置决定了人体在实践中如何运动。大多数情况下,驾驶员的动作取决于头部和手部的运动,而有些关节点,如颈部位置是相对固定的,没有明显的坐标变化。当识别一个动作时,应更关注产生动作的主要关节点。相对于次要关节点,主要关节点相对人体质心的运动轨迹会对动作的识别产生很好的效果,同时还能排除次要关节点细微变化产生的噪声干扰。在设计驾驶员的姿态估计图时,不仅基于人体基础骨架的连接,同时添加了如下关节点连接:即颈部分别与右手腕、左手腕、右耳、左耳的连接(图2右)。

综上所述,对驾驶员骨架模型可构建时空图 $G=(V,E)$,其中驾驶员关节点矩阵集合 $V=\{v_{it}|t=1,\dots,T,i=1,\dots,N\}$, T 为帧数, N 为一帧内关节点数,该集合包含了驾驶员在时间和空间上的所有关节点。 E 表示图中的边集,边集分为空间和时间上的边集:空间上,在每一帧中按照本文设计的驾驶员姿态估计图(图2右)构建空间边集 $E_s=\{v_{it}v_{ij}\}$;时间上,不同帧之间同一关节点连接构成驾驶员骨架关系的时序边集 $E_t=\{v_{it}v_{(t+1)i}\}$,代表一个特定的关节随时间推移的运动轨迹。

2.2 图卷积网络结构

图1上半部分展现了基于图卷积的驾驶员分心检测网络结构(Dri-GCN)。通过Openpose提取每帧驾驶员上半身12个关节点坐标 (x,y) 和置信度 S ,用元组 (x,y,S) 表示每个关节点,驾驶员骨架图用包含12个元组的数组表示。由于图卷积网络模型不同节点之间共享权重,因此不同节点上的输入数据比例须保持一致。在实验过程中,由于不同帧之间关节点坐标变化很大,须先将一个关节在不同帧下的位置特征进行归一化。

每个输入序列皆可表示为 (C,T,V) 的张量。其中 C 表示关节点特征数,为3; T 为视频序列长度,取300帧,长度不足300帧的视频序列进行循环填充得到; V 表示估计关节点数量,为12。采用式(4)将所得张量与人体骨架图的邻接矩阵相乘,以进行空间卷积。时间卷积采用大小为 $1\times K_t$ 的卷积核进行

卷积。模型由9层时空图卷积组成,前3层输出通道数设为64,中间3层输出通道数设为128,最后3层输出通道数设为256,其中第4、7层的时间卷积层为池化层。最后将输出的256个通道数的输出进行平均池化,并用全连接层(FCN)对特征进行分类,分类得到每类动作的置信度,根据置信度得到初步分类结果。

2.3 目标融合

通过提取身体骨架信息来识别驾驶员行为,可过滤掉众多冗余信息,但与此同时,许多关键信息也会被过滤掉,这些关键信息对于驾驶员某些动作的识别至关重要。通过“人+物”这种多信息模式的融合(Dri-GCN+obj),可更加精确地对驾驶员行为做出判断,同时可以排除某些驾驶员行为的误判。

使用Yolo-V3对驾驶员场景中的关键目标进行检测。图3为目标检测效果图,该模型在COCO数据集上进行训练,包括33万张图像,150万个对象实例和80个对象类别,Yolo-V3网络模型融合了3个不同的层,在小目标检测有非常好的效果。本文感兴趣的目標主要是与驾驶员动作相关的目标(手机、水杯、瓶子),在识别时,目标分类设为手机、水杯、瓶子和未检测到感兴趣目标4类。考虑到非目标物体的噪声干扰,根据阈值去除可能性较低的目标,采用非极大值抑制方法(NMS)去除冗余目标的影响。

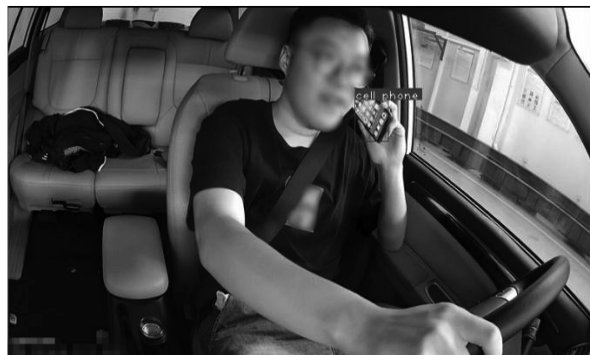


图3 目标检测效果图

将关键目标检测结果与Dri-GCN结果进行融合,对驾驶员行为进行再判断。通过检测目标范围的感兴趣目标,判断驾驶员的某种分心行为,如检测到水杯则判断为喝水,然后与Dri-GCN检测结果对比分析是否一致,如果一致则认为分类结果正确。否则与Dri-GCN第2置信度检测结果对比分析判断是否正确,如果错误则与第3置信度结果对比。如果Dri-GCN前3置信度结果与感兴趣目标

检测结果不符,则认为目标检测发生错误,依然采用 Dri-GCN 检测结果。通过融合目标检测结果与驾驶员姿态信息识别结果,能更准确判断驾驶员分心行为。

3 模型训练测试

为评估本文算法的效果,在 StateFarm 数据集和自制数据集上对 ST-GCN、Dri-GCN 和 Dri-GCN+obj 3 种模型训练后进行测试。

3.1 数据集简介

3.1.1 StateFarm 数据集^[10]

StateFarm 是 Kaggle 上检测驾驶员分心竞赛使用的数据集,同时也是第一个公开可用的驾驶员行为分类数据集。在 StateFarm 中,有来自 26 个人(不同种族的男性和女性的混合),其中每个人有 10 种行为(安全驾驶、用右手发短信、用右手打电话、用左手发短信、用左手打电话、调收音机、喝水、手从后面

拿东西、整理头发和化妆、和乘客说话)共 260 个剪辑(22 424 个图像)。研究人员提出的许多方法都是基于单个图像分类,而本文的实验是基于视频的行为识别,因此将每个剪辑中的图片制作成一个视频,由于图片数量有限,帧率为 5 fps。

3.1.2 自制数据集

受到 StateFarm 的分心驾驶员竞赛的启发,本文设计了一个遵循相同行为的类似视频数据集,如图 4 所示。在其他数据集中,都是将摄像头安装在副驾驶门槛梁上或驾驶员头顶,这会造成副驾驶乘员对驾驶员拍摄产生遮挡或导致拍摄驾驶员不完整。因此,最终决定将摄像头固定在车内后视镜上。本文设计的数据集由 30 位志愿者组成,包含不同身高和体型的驾驶员,在两台不同的车上和不同驾驶场景下进行收集,如光线较强或较暗,同时模拟真实驾驶环境,在拍摄范围内添加了干扰物(后排乘客和旁边行人)。该数据集共 300 个视频剪辑,每个剪辑为 3~10 s,帧率为 30 fps。

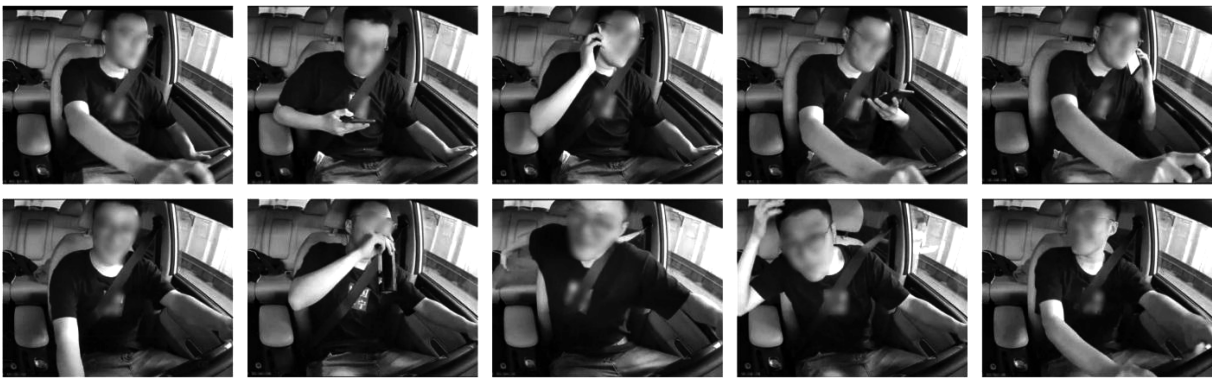


图4 自制数据集驾驶员行为示例

3.2 网络训练

实验平台为 Xeon E5 2620 V4,2.1 GHz,32 GB 内存,GTX 1070Ti 8 GB,实验基于 Ubuntu16.04LTS 系统,配套环境为 CUDA8.0,CUDNN6.0。

训练过程中先将视频分辨率调为 340×256,然后通过 Openpose 将驾驶员姿态信息提取出来,归一化后输入 GCN 网络进行训练,实验参数设置如下:网络模型通过随机梯度下降(SGD)进行优化,初始学习率设为 0.01,动量为 0.9,总共训练 60 个 epoch,第 20、30、40、50epoch 学习率迭代下降 0.1。训练过程中,前 30 个 epoch 迭代损失下降较快,迭代 30 个 epoch 后损失开始缓慢下降,到 60epoch 时基本达到收敛。

4 实验结果分析

表 1 为 3 种模型在 StateFarm 上的测试结果,其中通过改进的 Dri-GCN 达到了 84% 的准确率,比 ST-GCN 提高了 8 个百分点,而通过融合目标信息后,准确率能达到 90%。

表1 驾驶员行为分类在 StateFarm 上的结果 %

方法	准确率
ST-GCN	76
Dri-GCN	84
Dri-GCN+obj	90

将数据集随机分割成 3 种不同的训练/测试样本,训练集和测试集的样本数量比例为 7 : 3。本实验分别对 3 种分割方式进行评估,评价指标为识别准确率,并取 3 次结果的平均值作为最终结果。表 2 示出 3 种模型在 3 种不同分割方式上的识别率。由表看出,本文提出的方法 Dri-GCN 比 ST-GCN 平均提高了 8.14 个百分点,而融合物体信息后比 ST-GCN 平均提高了 15.59 个百分点。

表 2 驾驶员行为分类在自制数据集上的结果 %

方法	Split1	Split2	Split3	平均准确率
ST-GCN	73.33	80.00	78.89	77.41
Dri-GCN	83.33	88.89	84.44	85.55
Dri-GCN+obj	94.44	90.11	94.44	93.00

将在自制数据集上的实验结果用分类混淆矩阵表示,以描述不同种类行为的识别率。矩阵行表示正确的行为,列表示预测的行为,混淆矩阵的主对角线数值表示正确识别率,其他数值为错误识别率。矩阵外的行、列序号 1 ~ 10 分别代表第 3.1 节中所述的 10 种行为。

图 5 ~ 图 7 分别为 ST-GCN、Dri-GCN 和 Dri-GCN+obj 在本文数据集上的混淆矩阵。由图可见,3 种方法在大多数行为方面都取得了不错的效果,但小部分相似度较高的动作会产生较大的识别误差。从图 5 可见,ST-GCN 在正常驾驶时,驾驶员的一些细微动作可能会被识别为其他动作,正常驾驶正确识别率仅为 59.3%。而 29.7% 和 7.3% 的正常驾驶行为分别被误识别为左手发短信和打电话,看反光镜之类的正常驾驶行为也可能被误识别为和乘客说

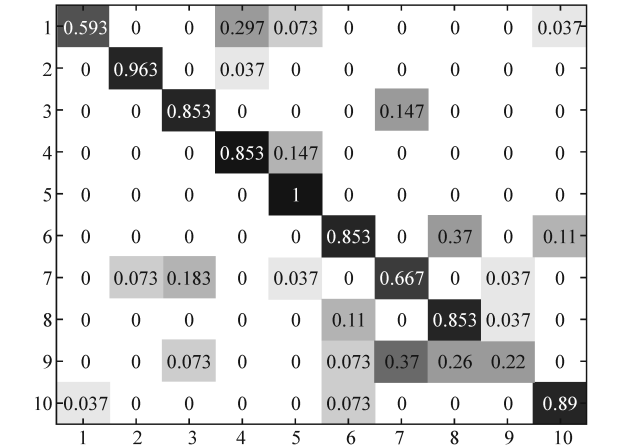


图 5 ST-GCN 实验结果

话。而对于整理头发与化妆这类手上动作和右手打电话、喝水也会产生较大的混淆,其中喝水和整理头发与化妆的准确率仅为 66.7% 和 22%。由图 6 看出,本文设计的 Dri-GCN 考虑到手和头部相对质心的运动,对人体行为感兴趣点精确定位,在区分正常行驶的一些细微动作方面取得了不错的效果,识别率提高了 37 个百分点,整理头发与化妆的识别率提高了 48.7 个百分点。

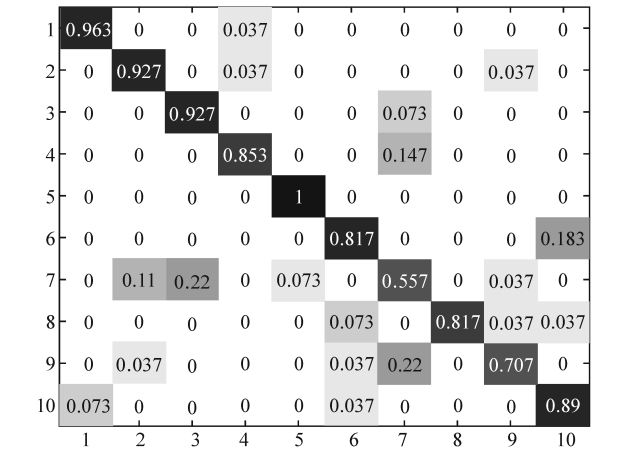


图 6 Dri-GCN 实验结果

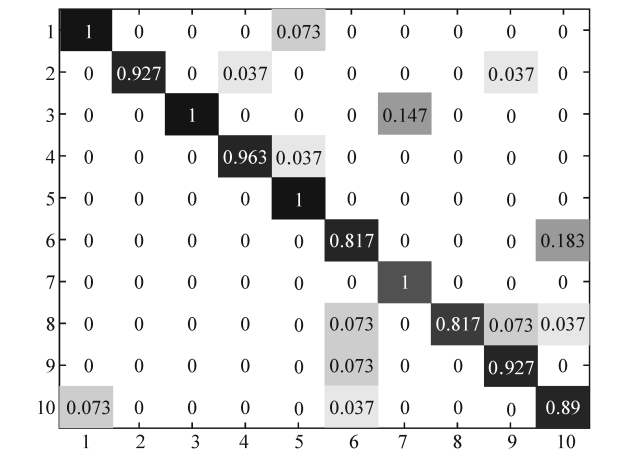


图 7 Dri-GCN+obj 实验结果

ST-GCN 和 Dri-GCN 对喝水动作的识别效果都比较差,单纯从手部轨迹来看,其与打电话具有极高的相似度,易产生混淆。从图 7 看出,在 Dri-GCN 上加目标识别后,通过对范围内水杯的识别,可以非常精确地区分喝水与其他动作,从而大幅度提高喝水动作的识别率,识别率达到 100%。同时,通过对手机和水杯的识别,可使一些误判为玩手机或喝水的动作得到更加准确的判断,如整理头发和化妆的识别率提高到了 92.7%。

综上所述,本文设计方法在检测驾驶员分心,并识别分心原因方面取得了较好的效果,同时检测速度主要取决于驾驶员姿态估计速度,总速度约为 20 fps,基本上能满足实时性检测要求。

5 结论

针对分心驾驶问题,提出一种基于单目视觉的驾驶员分心判别模型,并识别驾驶员分心原因,主要得出如下结论。

(1) 设计主要关注驾驶员头部和双手的姿态估计图,并基于图卷积网络提取姿态估计图在时间与空间上的信息特征,能有效提高驾驶员分心行为的识别准确率。

(2) 确定“人+物”多信息融合的方式对驾驶员分心原因进行再识别,如通过水杯等关键物体识别能提高喝水等相应动作的识别率,并排除其他动作的误识别。

(3) 制作了帧率为 30 fps 的驾驶员分心行为视频数据集,并将本文算法在 5 fps 的 StateFarm 数据集和 30 fps 的自制数据集上进行了验证,平均准确率分别达到了 90% 和 93%,检测速度约为 20 fps,基本满足实时性和准确性的要求。

参考文献

[1] DINGUS T A, KLAUER S G, NEALE V L, et al. The 100-car naturalistic driving study, phase II - results of the 100-car field experiment[R]. Blacksburg, NHTSA, 2006.

[2] 马雷,陈珂,王明露,等. 车路协同环境下驾驶员行为识别方法研究[J]. 汽车工程,2018,40(11):1330-1338.

[3] 刘志强,吴雪刚,倪捷,等. 基于 HMM 和 SVM 级联算法的驾驶意图识别[J]. 汽车工程,2018,40(7):113-119.

[4] DING M, SUZUKI T, OGASAWARA T. Estimation of driver's posture using pressure distribution sensors in driving simulator and on-road experiment[C]. 2017 IEEE International Conference on Cyborg and Bionic Systems (CBS), IEEE,2017:215-220.

[5] 陈骥驰,王宏,王翹秀,等. 基于脑电信号的疲劳驾驶状态研究[J]. 汽车工程,2018,40(5):515-520.

[6] BRAUN A, FRANK S, MAJEWSKI M, et al. CapSeat: capacitive

proximity sensing for automotive activity recognition[C]. Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, ACM, 2015:225-232.

[7] DEMIRDJIAN D, VARRI C. Driver pose estimation with 3D time-of-flight sensor[C]. 2009 IEEE Workshop on Computational Intelligence in Vehicles and Vehicular Systems, IEEE,2009:16-22.

[8] XING Y, LV C, ZHANG Z, et al. Identification and analysis of driver postures for in-vehicle driving activities and secondary tasks recognition[J]. IEEE Transactions on Computational Social Systems, 2017, 5(1):95-108.

[9] SESHADRI K, JUEFEI-XU F, PAL D K, et al. Driver cell phone usage detection on strategic highway research program (shr2) face view videos[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,2015:35-43.

[10] State farm distracted driver detection[G/OL]. <https://www.kaggle.com/c/state-farm-distracted-driver-detection>, 2016.

[11] ABOUELNAGA Y, ERAQI H M, MOUSTAFA M N. Real-time distracted driver posture classification[J/OL]. <https://arxiv.org/abs/1706.09498>, 2017.

[12] BAHETI B, GAJRE S, TALBAR S. Detection of distracted driver using convolutional neural network[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,2018:1032-1038.

[13] OU C, OUALI C, KARRAY F. Transfer learning based strategy for improving driver distraction recognition[C]. International Conference Image Analysis and Recognition. Springer, Cham, 2018:443-452.

[14] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. arXiv Preprint arXiv:1609.02907, 2016.

[15] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]. Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[16] ZHANG X, XU C, TIAN X, et al. Graph edge convolutional neural networks for skeleton based action recognition[J]. arXiv Preprint arXiv:1805.06184, 2018.

[17] CAO Z, SIMON T, WEI S E, et al. Realtime multi-person 2d pose estimation using part affinity fields[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:7291-7299.

[18] REDMON J, FARHADI A. Yolov3: an incremental improvement[J]. arXiv Preprint arXiv:1804.02767, 2018.