

DOI:10.19651/j.cnki.emt.2211607

基于3DCNN的驾驶员细微动作识别*

秦康 张小俊 张明路 杨亚昆
(河北工业大学机械工程学院 天津 300000)

摘要: 针对驾驶员相似的背景下的细微动作的动作识别,提出了一种基于X3D卷积神经网络X3D-M-GC-AE。通过引入轻量级的自注意力网络GCNet,提高对时间和空间关键特征的关注度,不引入参数数量的情况下,提高检测精度;设计了一种运动增强模块,使网络对时序上的运动信息更加敏感;引入知识蒸馏,将X3D-XL作为教师网络,X3D-M-GC-AE作为学生网络,可以使用较少的参数数量和计算量,达到可以实车应用的程度。实验结果表明教师网络测试精度最高可以达到75.56%,学生网络最高可以达到71.13%,该框架在车载硬件设备要求较低的情况下能够实现较高精度的检测效果。

关键词: 驾驶员动作识别;时间序列;X3D;注意力机制;知识蒸馏;深度学习

中图分类号: TP391.4;U463.6 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Driver subtle action recognition based on 3DCNN

Qin Kang Zhang Xiaojun Zhang Minglu Yang Yakun
(College of Mechanical Engineering, Hebei University of Technology, Tianjin 300000, China)

Abstract: Aiming at the action recognition of subtle actions in the similar background of drivers, X3D-M-GC-AE based on X3D network is proposed. By introducing the lightweight self-attention network GCnet, the attention to key features in time and space is improved, and the detection accuracy is improved without increasing parameter quantities. Action enhancement block is designed to make the network more sensitive to the action information in time series. Introducing knowledge distillation, taking X3D-XL as the teacher network and X3D-M-GC-AE as the student network, so that X3D-M-GC-AE can be used in real vehicles with less parameters and calculations. The experimental results show that the maximum test accuracy of teacher network can reach 75.56%, and that of student network can reach 71.13%. This framework can achieve high-precision detection results in the case of low requirements for vehicle hardware equipment.

Keywords: driver action recognition; time series; X3D; attention mechanism; knowledge distillation; deep learning

0 引言

随着经济和科学技术的发展,交通体系变得愈发发达,尤其是在公路交通方面,在运输、载人方面起到了中流砥柱的作用;但是随着公路交通的繁荣,交通安全问题也愈发严重,对驾驶员造成了严重的生命财产损失。通过多年的调查研究,在驾驶员、道路和车辆组成的交通系统中,驾驶员一般处于弱势。

据统计,国内因为驾驶员失误造成的事故占有所有事故的80%~90%^[1-2],调研报告《2018-2020中国青年注意力分散与交通事故白皮书》显示,曾经在驾驶中有“使用通信工具”行为的中国青年在接受调查的人数中占比高达八成以上^[3]。居高不下的事故率给国家和人民带了巨大的损

失,所以驾驶员注意力监测方法的研究成为迫在眉睫的需求。

驾驶员分心行为通过监测车辆信息判断,采集模拟驾驶员饮水时的汽车车速、加速度和油门踏板行程参数与正常驾驶进行比较,建立饮水的车辆信息模型,来判断驾驶员是否在进行分心行为^[4];有通过计算机视觉技术来监测,通过肤色信息判断手部和唇部的关系,进而判断是否进行打电话行为^[5];有通过使用 Faster-RCNN 识别驾驶员手的判断驾驶员是处于安全驾驶还是处于打电话状态^[6];有使用 YOLO(you only look once)算法,通过识别出手机和烟来判断驾驶员的吸烟和打电话的行为^[7];有使用 OpenPose 识别人体骨骼关键点来判断人体动作,但是由于驾驶员处

收稿日期:2022-10-06

* 基金项目:天津市新一代人工智能科技重大专项(18ZXZNGX00230)资助

于车内,车内的摄像头只能看到驾驶员的上半身,无法很好地适用于驾驶员的动作识别^[8];有使用卷积神经网络(convolutional neural network, CNN)+长短期记忆网络(long short-term memory, LSTM)手势肌电识别,虽然保存了时间信息^[9],但是没有为测量肌电信号,会与人直接接触,并不适合应用在驾驶员动作识别中;有使用 2DCNN+Transformer 算法进行人体识别,但是无法很好地识别相似的驾驶员动作且 Transformer 部分的计算量较大,并不能很好的应用在汽车级别的运算平台上^[10];有将同一静态图片复制多次输入到 3DCNN 中判断驾驶员是否进行了分心行为^[11],但是这些图片并没有时间的连续性,只是单独的静态图片;有使用 M(2-1)D 算法将 3DCNN 拆分成空间上的 2D 和时间上的 1D 有效的减少计算量,证明了 3DCNN 在驾驶员动作识别的可应用性^[12]。这些算法有的忽略了驾驶员动作在时间上的连续性,只是针对单一静态图片判断,无法了解到动作的前因后果;有的缺少了对背景和相似动作的有效区分。

文章使用的 X3D-M-AE-GC 算法,将驾驶员整个动作共 16 帧图片进行输入,在时间序列上判断驾驶员的行为,增强了时间信息的利用率,引入时间和通道注意力机制和动作增强模块,使得模型专注于驾驶员细微的动作,提高了驾驶员动作识别精度,引入知识蒸馏,降低了模型的参数量和计算量,提高了运算速度,实现了车载硬件设备要求较低的情况下能够实现较高精度的检测效果。

1 X3D 网络概述

X3D^[13] (extend 3D) 网络,如图 1 所示,其核心思想是在平衡计算量和准确率的前提下,只沿时间维度进行扩展并不一定会比扩张其他维度模型效果好,尤其是在追求更小的计算量的情况下,沿着其他维度扩展可能准确率更高。作者尝试从不同的维度对 2D 卷积进行扩展,使其适用于 3D 时空数据处理,扩展的维度包括时间维度大小 γ_t 、采样帧率 γ_T 、输入分辨率大小 γ_s 、卷积核的数量 γ_w 、时间维度的卷积设置 γ_b 以及网络的深度 γ_d 。最终确定了 X3D-S、X3D-M、X3D-L、X3D-XL 等平衡高精度与低计算量的网络结构。

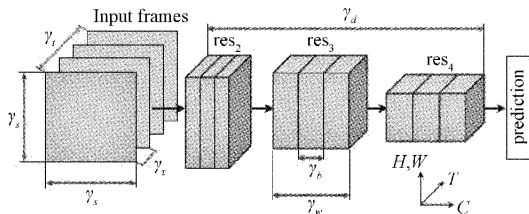


图 1 X3D 网络结构

驾驶员的动作也属于人类日常行为,可以使用 X3D 对其识别,但是驾驶员动作发生在汽车环境中,动作部位和动作幅度受到了限制,存在一些微小的和相似的动作;并且由

于汽车计算资源的限制,光流数据无法应用在汽车中,这些都对驾驶员动作识别产生了一定的影响。如何精确的关注到运动部位、精确的区分不同运动的差异和合理利用汽车的计算资源成为了识别驾驶员动作的阻力。

为解决这些问题,在使用较为轻量级网络 X3D-M 的同时,对网络加以改进,设计出一种适合驾驶员动作监测的网络结构。文章将以 X3D-M 网络为基础,提出一个对细微动作有高关注度、高分度且能实车部署的优秀的算法。

2 X3D-M-AE-GC 网络模型

修改后的 X3D-M-AE-GC 网络如图 2 所示,为了部署在实车上,选择 X3D 算法中的轻量级的 M 模型,具体算法结构如表 1 所示。在 Layer2、Layer3 和 Layer4 前引入 3D 版本的全局上下文网络(global context block, GC block),单独一个网络不仅可以提高对时空的关注度,而且可以提高对关键通道的敏感度,提高算法在时间、空间和通道高方面对驾驶员动作的注意力;针对驾驶员相似的背景下的微妙动作,在 Layer2、Layer3 和 Layer4 后加入运动增强模块(action enhancement block, AE block),将时间间隔的两个图片帧相减,之后在进行卷积操作,增强运动前景信息,抑制静态背景场景信息。表 1 为改进后的网络结构。

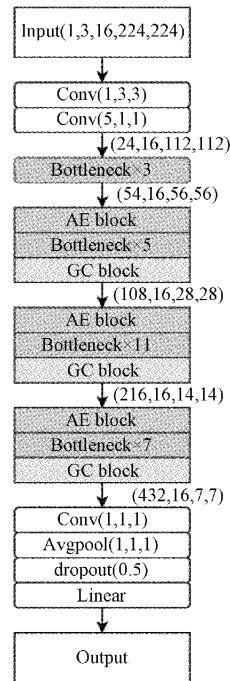


图 2 X3D-M-AE-GC 网络结构

2.1 轻量级注意力机制 GCNet

GCNet^[14] 如图 3 所示,结合非本地网络(non-local block, NLnet)和压缩-激发网络(squeeze excitation, SENet)各自的优点,即能够像 NLNet 一样有效的对全局上下文建模,又能够像 SENet 一样调整通道依赖。为了满足动作识

表1 X3D-M-GC-AE网络结构

网络结构	卷积核	输出尺寸
Conv_s	Stride5,12	16×224×224
Conv_t	1×3 ² ,3×1,24	16×112×112
Layer1	$\begin{bmatrix} 1 \times 1^2, 54 \\ 3 \times 3^2, 54 \\ 1 \times 1^2, 24 \end{bmatrix} \times 3$	16×56×56
Layer2	AE+ $\begin{bmatrix} 1 \times 1^2, 108 \\ 3 \times 3^2, 108 \\ 1 \times 1^2, 48 \end{bmatrix} \times 5$ +GC	16×28×28
Layer3	AE+ $\begin{bmatrix} 1 \times 1^2, 216 \\ 3 \times 3^2, 216 \\ 1 \times 1^2, 96 \end{bmatrix} \times 11$ +GC	16×14×14
Layer4	AE+ $\begin{bmatrix} 1 \times 1^2, 432 \\ 3 \times 3^2, 432 \\ 1 \times 1^2, 192 \end{bmatrix} \times 7$ +GC	16×7×7
Conv5	1×1 ² , 432	16×7×7
pool5	16×7×7	1×1×1
fc1	1×1 ² , 2 048	1×1×1
fc2	1×1 ² , # classes	1×1×1

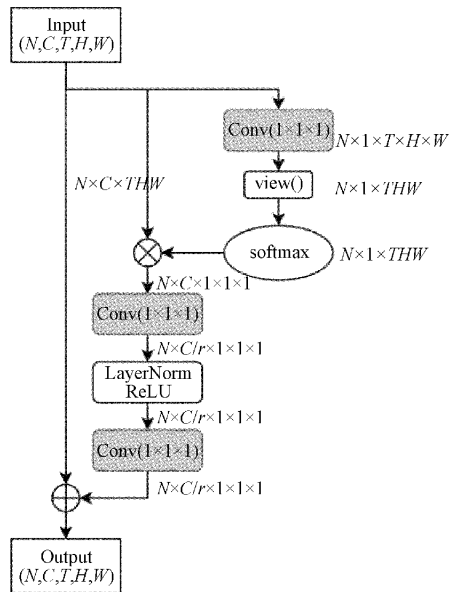


图3 GCNet网络结构

别的需求,文章将GCNet修改为3D模式,以合理接收3D卷积传下来的数据,增加对驾驶员动作的时间和空间的注意力,又可以提高有效通道的所占的比例,增强对驾驶员细微动作的区分;GCNet通过计算一个全局的注意力特征图来简化non-local block,并且对所有位置共享这个全局注意力特征图,这使得GCNet具有更轻量级的网络结构,降低算力需求,更好的实车检测驾驶员动作。

2.2 运动增强模块 AE block

驾驶员背景相似且动作细微对识别造成的挑战。与日常动作不同,前景的运动特征和背景的环境特征是人类识别行为的两条基本线索,但是在单一的驾驶员场景中,相似的背景对驾驶员的动作理解帮助很小,在文章中静止部分和运动部分不需要得到相同的关注度,所以引入运动增强模块,运动增强有利于提高对运动信息敏感度,将网络的重点转移到运动的部分。

为了对细微动作进行良好的识别,设计了运动增强模块AE block,如图4所示,主要通过相同时间间隔的图片帧作差,然后经过卷积操作提取出帧与帧之间的运动信息。对驾驶员动作识别来讲,使模型能够发现并增强帧与帧之间的运动信息是有益的。

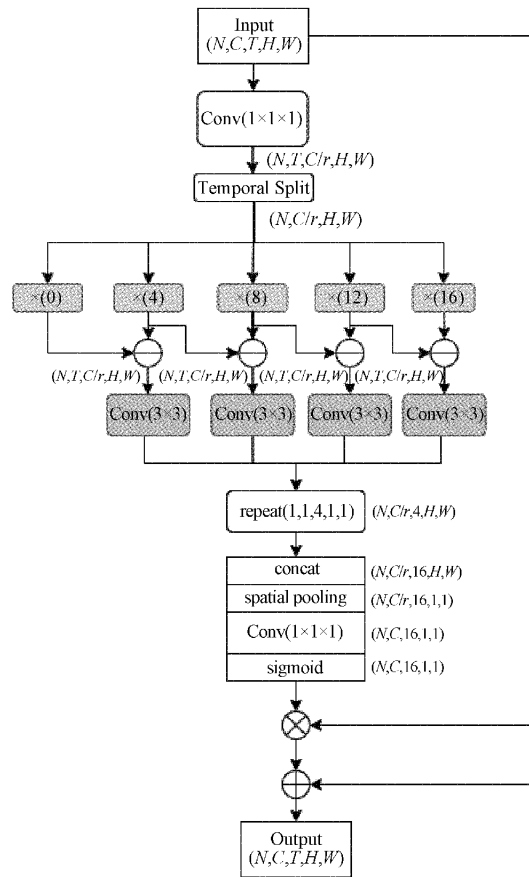


图4 运动增强模块

具体操作步骤如下:首先将输入的 (N, C, T, H, W) 通过一个 $1 \times 1 \times 1$ 的三维卷积层来减少特征通道,减少计算量;将输入的时间序列每隔 m 帧取出一个($m=4$),然后依顺序相减,得到的结果即可视为这个时间间隔中的运动部分;然后在将相减都得结果通过 $\text{Conv}3 \times 3$ 提取特征,对运动特征进行进一步的提取;将提取后的特征在时间维度重复4份,将所有的特征在时间维度拼接起来,保证时间维度和输入相同,然后利用全局平均池层来总结空间信息;利用另一个 1×1 二维卷积层将运动特征的通道维数扩展到原

始通道维数,并利用 sigmoid 函数得到运动注意权重;为了不忽略静止部分信息使用残差结构与初始输入相加融合,即突出了运动部分信息,又对静止的背景有所保留。

3 知识蒸馏

文章引入知识蒸馏模块^[15] (knowledge distillation, KD),如式(1)所示,知识蒸馏是一种在繁琐的模型中提炼知识并将其压缩为单个模型的方法,以便可以将轻量级网络算法部署到实际应用中。知识蒸馏是指将繁重的网络模型(教师)的学习结果教授到较小的网络模型中(学生)。

$$L = (1 - \alpha)loss(y, p) + \alpha loss(q, p) \quad (1)$$

如图 5 所示,教师产生的输出被用作训练学生的“软目标”如式(2)所示。将参数量较大的 X3D-XL 网络视作教师网络,将优化后的 X3D-M-GC-AE 的网络视为学生网络。其中, α 为两个损失函数所占比重,取 $\alpha = 0.3$; $loss(y, p)$ 为教师网络输出的预测结果损失, $loss(q, p)$ 为学生网络输出的预测结果的损失。 T 为温度,取 $T = 3$; T 值越大, softmax 函数输出越平滑,不仅可以说明该动作属于哪一类,而且可以说明他和其他动作类别的相似程度。

$$q_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \quad (2)$$

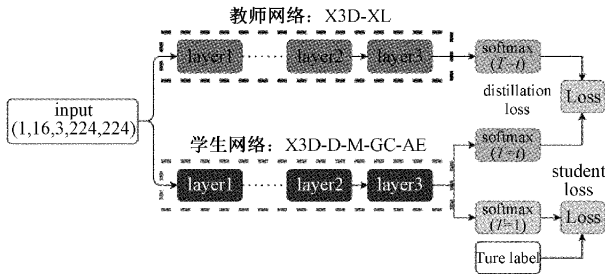


图 5 知识蒸馏网络结构

4 数据集

MDAD^[16]数据集包括真实驾驶场景,其中年龄在 19~41 岁之间的 50 名参与者(38 名男性和 12 名女性)被要求在自然驾驶环境中驾驶大众 Polo 4。要求每位驾驶员在不同的路段下执行一项安全驾驶和 15 项常见的次要任务,共 800 段视频,每段视频 2~5 s 不等。

MDAD 数据集使用安装在驾驶员前面仪表盘的摄像头所采集的数据集,分辨率为 640×480 pixels 的彩色图像;驾驶员动作分类如表 2 所示,具体动作如图 6 所示。

使用 ffmpeg 工具分割 RGB 视频为图片帧,如图 7 所示部分打电话图片帧,1 s 32 帧图片的制作,每个视频 2~5 s 不等,可以获得 64~128 张不等的视频帧,输入网络的 16 张图片为等差采样,公差计算由式(3)获得:

$$q = \frac{len}{16} \quad (3)$$

式中: q 为公差、 len 为每个视频分割出的帧数。

表 2 驾驶员行为分类

行为编号	类别
A1	安全驾驶
A2	梳头或者化妆
A3	调整收音机
A4	GPS 操作
A5	用右手发短信
A6	用左手发短信
A7	用右手打电话
A8	用左手打电话
A9	拍照
A10	与乘客交谈
A11	唱歌或者跳舞
A12	打瞌睡
A13	用右手喝水
A14	用左手喝水
A15	向后拿东西
A16	吸烟

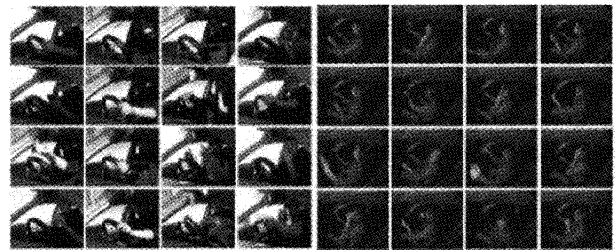


图 6 驾驶员动作数据集



图 7 喝水动作分解

5 结果分析

文章实验在 NVIDIA GeForce RTX 3070 上进行,辅助环境为 CUDA 11.1 和 cuDNN 8.0.5,使用编程语言 Python3.8,集成开发环境为 Pycharm。

测试结果如图 8 所示,算法可以很好的区分喝水、打电话等动作,对于动作幅度较小且动作部位相同的唱歌、说话动作识别效果有待提高。

5.1 可视化

为了解网络对驾驶员动作的关注点在哪里,通过 Grad-CAM^[17]绘制出热力图来分析网络的关注点在哪里。Grad-CAM 将最后一层包含感兴趣目标的语义信息的输出进行反向传播,得到反向传播特征层的梯度信息,然后将这些梯度进行全局平均池化并通过 ReLU 函数传递,以减少

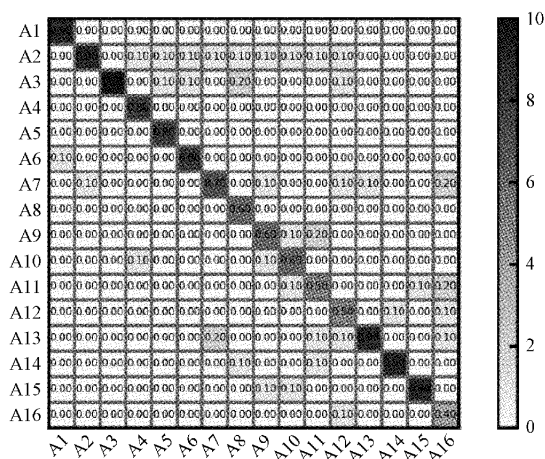


图 8 混淆矩阵

来自其他类的不需要的激活的影响。产生与最终卷积层形状相同的一般激活区域的低维热图。

文章对 Grad-CAM 网络进行修改, 以实现 3D 卷积的可视化, 如图 9(a)~(p)所示。将得到的热度图分别放置在 16 帧图片上; 这里等距截取了驾驶员动作的 4 帧图片, 图中颜色越红就说明此位置为网络关注的地方, 可以看到拿出算法大多关注驾驶员的头部和手部这些运动部分, 打电话, 喝水等动作区分较为明显, 但是说话、唱歌和打哈气等动作区分效果较差。

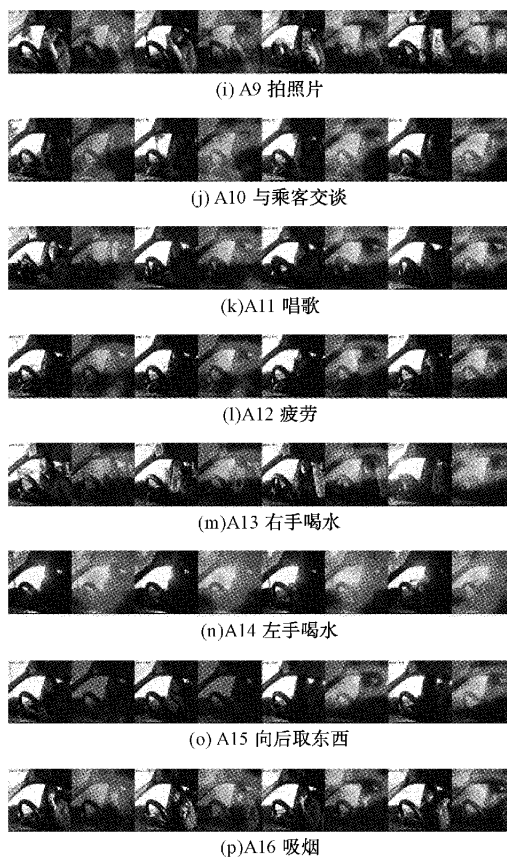
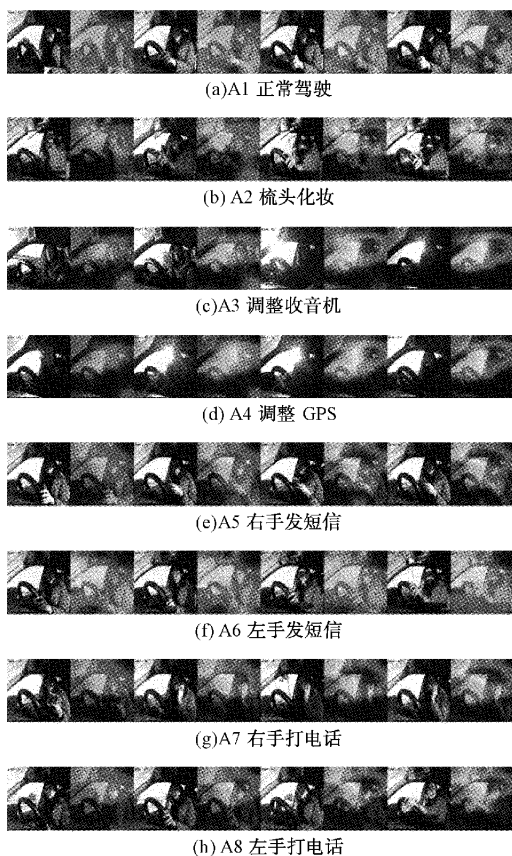


图 9 算法可视化

5.2 性能评价指标

图 10 为 X3D-M-GC-AE 在数据集上的损失函数(loss function)变化趋势图, 实线为训练损失, 虚线是测试损失。训练集的处理尺寸(batch size)设置为 4, 迭代次数(iteration number)设置为 1 000 次。从图中可以看出, 在开始的 100 次迭代中损失下降的很快, 后续趋于稳定。

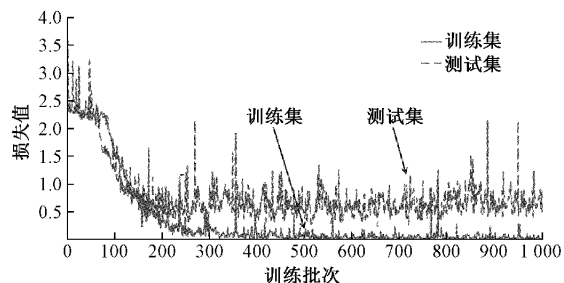


图 10 训练损失和验证损失

图 11 为 X3D-GC-AE 在数据集上的精度变化趋势图, 实线为训练精度, 虚线为测试精度。和损失函数图相同, 前 100 次迭代后训练精度快速变化, 后续趋于稳定。训练精度趋于 98%, 测试精度最高可以达到 71.13%。

选取准确度均值作为精度评价指标, 模型尺寸和 FLOPs(floating point of operations)作为速度评价指标,

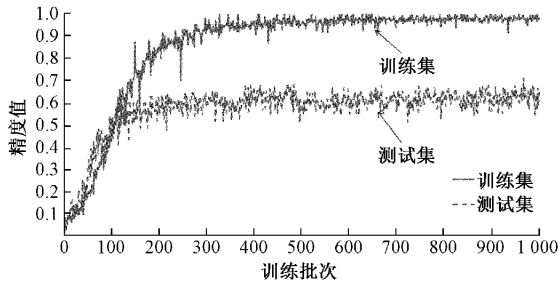


图 11 训练精度和验证精度

表 3 为在相同数据集,不同帧数输入下,不同算法的精度、模型尺寸和 FLOPs 的对比。

表 3 算法性能对比

算法名称	帧数	精度/ %	FLOPs/ G	模型尺寸/ MB
I3D ^[18]	32 帧	73.13	12.30	48.29
Slowfast ^[19]	32 帧×4 帧	70.25	33.58	131.68
X3D-L	16 帧	70.25	5.35	21.81
X3D-XL(教师)	16 帧	75.75	10.40	41.44
X3D-M	16 帧	68.88	3.01	12.06
X3D-M+GC+AE(学生)	16 帧	71.38	3.06	12.27

通过多次消融实验如表 4 所示,在相同 16 帧的情况下,原版 X3D-M 的精度为 68.88%,对安全驾驶、化妆、调整收音机、左右手喝水和向后取东西都有很好的识别效果,可以取得良好的时序信息,即使处于运动状态,没有达到最后稳定状态也有很好的识别效果,但是对不同姿势的使用手机状况识别效果较差;单独加入 GC block 之后,X3D-M-GC 提升到了 69.50%,整体效果都有些许提升;单独引进 AE block 之后,X3D-M-AE 精度提升到了 70.13%,对左右手打电话喝水这种有一个明显的运动轨迹的动作识别有所提高;将两个模块都引入之后,X3D-M-GC-AE 精度提高到了 71.38%准确度。而他们的 FLOPs 几乎没有变化,仅增长 0.05 G,模型大小从 12.06 M 增长到了 12.27 M。而表 3 中 I3D 和 Slowfast 算法虽然有较高的精度,但是他们的参数量是巨大的,且输入帧数也是 X3D 算法的 2 倍,不适合用于驾驶员行为识别。

表 4 消融实验

算法名称	GC	AE	精度/ %	FLOPs/ G	模型 尺寸/MB
X3D-M			68.88	3.01	12.06
X3D-M+GC	✓		69.50	3.02	12.12
X3D-M+AE		✓	70.13	3.04	12.21
X3D-M+GC+AE	✓	✓	71.38	3.06	12.27
X3D-XL+GC+AE			Front(RGB)		95.00%

表 5 为相同 32 帧输入帧数条件下不同算法之间的对比,为了保证条件相同,本算法将 AE 模块修改为 8 帧为一个间隔,且将训练类修改为与 DADCNet 相同的 9 类,保留了正常驾驶、调整收音机、左手发短信、右手发短信、右手打电话、左手打电话、与乘客交谈、右手喝水和向后取东西这 9 类动作。

表 5 相同帧数不同算法的精度计较

算法	描述	精度/%
DC-EPCC	Front(D)	90.87
CL-MOD	Front+Top(DIR)	95.99
DAD	Front+Top(DIR)	96.73
DADCNet	Front+Top(D)	96.60
X3D-M+GC+AE	Front(RGB)	91.11

5.3 实车部署

智能汽车的算法主要部署在 NVIDIA Jetson TX2 嵌入式开发板上,开发板的系统是基于 Linux 的 Ubuntu18.04, Linux 是一款免费的操作系统,它由众多微内核组成,其源代码完全开源,其次,它具有非常强大的网络功能,支持所有的因特网协议,并且 Linux 系统工具链完整,简单操作可以配置出合适的开发环境,系统稳定可靠,具有较强的移植性,加速深度学习等视觉算法的推理,是极好的移动部署终端,部署情况如图 12 所示,鼠标、键盘和摄像头通过 USB 与 TX2 连接,显示器通过 HDMI 与 TX2 连接,通过电源接口直接供电,当识别到分心行为时,蜂鸣器发出声音预警。

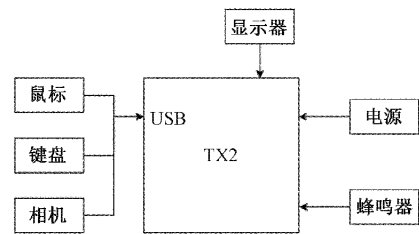


图 12 系统框图

如图 13 所示,将 TX2 和摄像头组成的平台放在实车

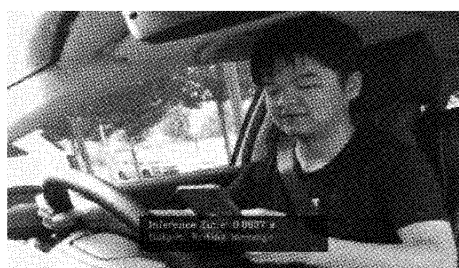


图 13 实车平台搭建

环境中进行测试,使用 Tensor RT 对 X3D-M-AE-GC 算法进行优化,提高他在移动设备上的运行速度,最终可以达到 1 s 测试 5~10 次,可以完成实车部署任务,效果如图 14(a)、(b)所示,可以准确的识别出吸烟和发短信动作。



(a) 吸烟



(b) 发短信

图 14 驾驶员动作识别图

6 结 论

针对驾驶员相似的背景下的微妙动作的动作识别,提出了一种 X3D-M-GC-AE 算法。使用 3D 卷积提取整个时序上动作信息;通过引入 3D 形式的轻量级的自注意力网络 GCNet,提高对时间和空间关键特征的关注度;设计了一种运动增强模块,使网络对时序上的运动信息更加敏感;引入知识蒸馏,可以使用较少的参数量和计算量的网络算法,达到可以实车应用的程度。使用 Tensor RT 对模型进行加速处理,完成在 Jetson TX2 平台上的实车部署。

参考文献

- [1] BUCSUHÁZY K, MATUCHOVÁ E, ZŮVALA R, et al. Human factors contributing to the road traffic accident occurrence [J]. *Transportation Research Procedia*, 2020, 45: 555-561.
- [2] HU L, BAO X, LIN M, et al. Research on risky driving behavior evaluation model based on CIDAS real data[J]. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 2021, 235: 095440702098597.
- [3] 林斐. 八成中国青年曾驾车打电话[N]. *IT 时报*, 2021-07-16, DOI:10.28404/n.cnki.nitsd.2021.000333.
- [4] 冯建国, 钱大琳, 罗艺, 等. 基于 SVM 的危货车驾驶员饮水分心判别模型[J]. *公路交通科技*, 2017, 34(S2): 16-22.
- [5] 王丹. 基于机器视觉的驾驶员打电话行为检测[D]. 北京:北京理工大学, 2015.
- [6] LE T, ZHENG Y, ZHU C, et al. Multiple scale faster-RCNN approach to driver's cell-phone usage and hands on steering wheel detection [C]. *Computer Vision & Pattern Recognition Workshops*, 2016.
- [7] 熊群芳, 林军, 岳伟, 等. 基于深度学习的驾驶员打电话行为检测方法[J]. *控制与信息技术*, 2019(6): 53-56, 62.
- [8] 李一凡, 袁龙健, 王瑞. 基于 OpenPose 改进的轻量化人体动作识别模型[J]. *电子测量技术*, 2022, 45(1): 89-95.
- [9] 陈思佳, 罗志增. 基于长短时记忆和卷积神经网络的手势肌电识别研究[J]. *仪器仪表学报*, 2021, 42(2): 162-170.
- [10] 朱相华, 智敏, 殷雁君. 基于 2D CNN 和 Transformer 的人体动作识别[J]. *电子测量技术*, 2022, 45(15): 123-129.
- [11] NGUYEN D L, PUTRO M D, JO K H. Distracted driver recognizer with simple and efficient convolutional neural network for real-time system[C]. *21st International Conference on Control, Automation and Systems, ICCAS 2021*, 2021: 371-375.
- [12] DUAN H, ZHANG C, LYU Z. Driving action recognition based on 3D convolution [C]. *2021 17th International Conference on Computational Intelligence and Security(CIS)*, 2021: 94-98.
- [13] FEICHTENHOFER C. X3D: Expanding architectures for efficient video recognition [C]. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)*, 2020: 200-210.
- [14] CAO Y, XU J, LIN S, et al. GCNet: Non-local networks meet squeeze-excitation networks and beyond [J]. *ArXiv Preprint*, 2019, ArXiv:1904.11492.
- [15] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network [J]. *Computer Science*, 2015, 14(7): 38-39.
- [16] JEGHAM I, KHALIFA A B, ALOUANI I, et al. MDAD: A multimodal and multiview in-vehicle driver action dataset [C]. *International Conference on Computer Analysis of Images and Patterns CAIP 2019: Computer Analysis of Images and Patterns*, 2019.
- [17] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization [C]. *IEEE International Conference on Computer Vision*, 2017.

- [18] CARREIRA J, ZISSERMAN A. Quo vadis, action recognition? A new model and the kinetics dataset[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6299-6308.
- [19] FEICHTENHOFER C, FAN H, MALIK J, et al. SlowFast networks for video recognition[C]. 2019 IEEE/CVF International Conference on Computer Vision(ICCV), 2019.

作者简介

秦康, 硕士研究生, 主要研究方向为智能汽车。

E-mail:kang_qin211@163.com。

张小俊, 博士, 教授, 博士生导师, 主要研究方向为汽车电子控制技术。

张明路, 博士, 教授, 博士生导师, 主要研究方向为特种机器人机构与应用技术研究。

杨亚昆, 硕士研究生, 主要研究方向为智能汽车。