

DOI: 10.11830/ISSN.1000-5013.202001010



深度可分离卷积网络的 驾驶状态识别算法

马迎杰, 王佳斌, 郑力新, 朱新龙

(华侨大学 工学院, 福建 泉州 362021)

摘要: 针对嵌入式设备内存小及多分类准确率低等导致驾驶员检测问题,提出经过深度可分离卷积网络改进而成的,快速下采样网络(fast downsampling network, MF-Net)驾驶状态识别系统. 即将快速下采样策略应用于深度可分离卷积网络,在 12 层内执行 32 倍下采样,以有效降低计算成本、增加信息容量并实现性能改进. 实验结果表明:与 VGG(visual geometry group)和 ResNet 50 等其他卷积神经网络(CNN)模型相比,所提出的 MF-Net 模型深度可分离卷积大大减少参数量,快速下采样方案的运用增加了网络的信息容量,不仅模型较小且在驾驶员状态分类方面能够表现出更好的性能. 同时,信息容量的增加可以对更多信息进行编码,加深对图像内容的理解,有利于之后的嵌入式系统移植.

关键词: 驾驶状态; 状态特征检测; 深度学习; 深度卷积; 逐点卷积

中图分类号: TP 391 **文献标志码:** A **文章编号:** 1000-5013(2021)02-0259-09

Driving State Recognition Algorithm Based on Deep Separable Convolutional Network

MA Yingjie, WANG Jiabin,
ZHENG Lixin, ZHU Xinlong

(College of Engineering, Huaqiao University, Quanzhou 362021, China)

Abstract: Aiming at the problem of driver detection caused by the small memory of embedded devices and the low accuracy of multi-classification, a fast fast down sampling network (MF-Net) driving state recognition system improved by deep separable convolutional network is proposed, the key idea is applying a fast down-sampling strategy to deep separable convolutional networks, which performs 32-fold downsampling within 12 layers to reduce computational cost seffectively, increase information capacity, and achieve performance improvements. The experimental results show that: compared with other convolutional neural network (CNN) models such as VGG (visual geometry group) and ResNet 50, MF-Net model has deep separable convolutions that reduce the amount of parameters greatly and the application of fast down sampling increase the information capacity of the network, the model not only smaller but also show better performance in the classification of driver status, at the same time, the increase in information capacity can encode more information and deepen the understanding of the image content, which is beneficial to transplantation of embedded systemsin the future.

收稿日期: 2020-01-09

通信作者: 王佳斌(1974-),男,副教授,主要从事物联网、云计算、大数据和智能仪器的研究. E-mail: fatwang@hqu.edu.cn.

基金项目: 国家自然科学基金青年科学基金资助项目(61505059);福建省厦门市科技局产学研协同创新资助项目(3502Z20173046);华侨大学研究生科研创新能力培育计划资助项目(17013084002)

Keywords: driving state; state feature detection; deep learning; depthwise convolution; pointwise convolution

根据国际标准化组织(ISO)的定义,分心驾驶是指驾驶时注意力指向与正常驾驶不相关的活动,从而导致驾驶操作能力下降的一种现象.美国运输部表示,分心驾驶是导致危险发生的最大因素.另外,根据美国国家公路交通安全管理局(NHTSA)的数据,2015 年,共有 35 092 人死于美国道路的交通事故,较 2014 年死亡人数上升了 7.2%,其中分心驾驶需要为 39.1 万名伤者和 3 477 名死者负责^[1].

早期,研究人员运用生理数据^[2-4]和车辆控制数据^[5]来进行疲劳检测. Zhao 等^[6]提出了一种使用随机森林(RF)分类器的特征提取方法. Berri 等^[7]提出了一种基于支持向量机(SVM)的模型,对于视觉系统获得 91.57%的成功率. Craye 等^[8]提出基于有源传感器 Kinect 和计算机视觉工具,但是他们的解决方案的数据来源于室内的模拟,与真实驾驶环境还是有很大不同. Le 等^[9]设计了一个快速 RCNN (regions with convolutional neural network features)模型来检测驾驶员是否正在使用手机,以及驾驶员的手是否在方向盘上. Ruiz 等^[10]通过在 300W-LP 数据集上训练多损耗卷积神经网络来提供一种优雅而稳健的方法来确定姿势.

现有的驾驶人状态检测方法大都基于某一单项指标,在限定条件下能够达到一定的精度. 然而,在实际复杂多变的行车环境下,其准确性和可靠性上还存在问题,难以达到预期要求,且大多数模型不易移植到嵌入式系统中. 基于此,本文提出基于快速下采样网络结构(MF-Net)的驾驶状态识别系统.

2 实验数据与处理

2.1 数据处理

实验数据集来源于 Kaggle“国家农场分心驾驶员检测”竞赛,包括具有 22 400 张标记的样本训练集(这些样本在各类之间平均分配),以及 79 700 张未标记的测试样本. 考虑到摄像头的画面、位置、角度差异、司机调整座椅和方向盘位置,以及训练图像大小只有约 22 KB,因此希望从训练集中综合获取更多图像,以确保模型不会因神经网络具有数百万个参数而学到不必要的特征,从而产生过拟合^[11].

为了防止这种原因导致的过拟合,在训练数据之前,要对训练数据做数据增强. 在图像数据中,数据增强的常用方法包括:对图片进行角度偏移、左右偏移、上下偏移、随机放大,以及缩小、水平翻转等. 如果将上述方法组合排列起来,对图片进行放大后再进行角度偏移,将会产生大量图像. 在此使用 Keras 中的图片生成器 Image Data Generator()来创建更多图像,其中垂直平移范围(height_shift_range)=0.5,水平平移范围(width_shift_range)=0.5,随机缩放范围(zoom_range)=0.5,图像旋转角度范围(rotation_range)=40.

图 1 是某一位驾驶员身体姿势状态数据. 其识别状态中,C0 为安全类,C1~C9 为非安全类. 除了简单的对图像进行几何上的形变,还可以对图片的饱和度、亮度色彩进行小范围的幂次缩放或者乘法缩放等数学操作. 另外还可以在破坏图像质量的前提下,对图像的每个像素进行整体加减法等操作.

对于此数据,“图像增强”还具有其他一些优势. 有时来自两个不同类别的图像之间的差异可能非常细微,如图 2 所示. 在这种情况下,通过不同角度对同一幅图像进行多次查看会有所帮助. 如果不仔细观



(a) C0(正常驾驶)



(b) C1(右手打字)



(c) C2(右手打电话)



图 1 驾驶员身体姿势识别状态数据

Fig. 1 Driver body posture recognition status data

察图 2 的两张对比图像,几乎认为他们属于同一类别. 但是,图 2(a)的图像是 C2(右手打电话)类别,而图 2(b)图像是 C8(头发和化妆)类别. 通过数据增强,网络可以更好的发现其特征,以利于分类.



图 2 两幅不同类别的图像对比

Fig. 2 Comparison of two images of different categories

2.2 特征可视化

卷积神经网络(CNN)采用类激活映射(class activation mapping)和类特定显着图(class-specific saliency map)寻找目标及改变视频中的注意力. 即使用一种 CNN 中的全局平均池(GAP)生成类激活

映射的技术^[12],用于特定类别的类激活图表示 CNN 用于识别该类别的区别图像区域.类激活映射过程,如图 3 所示.

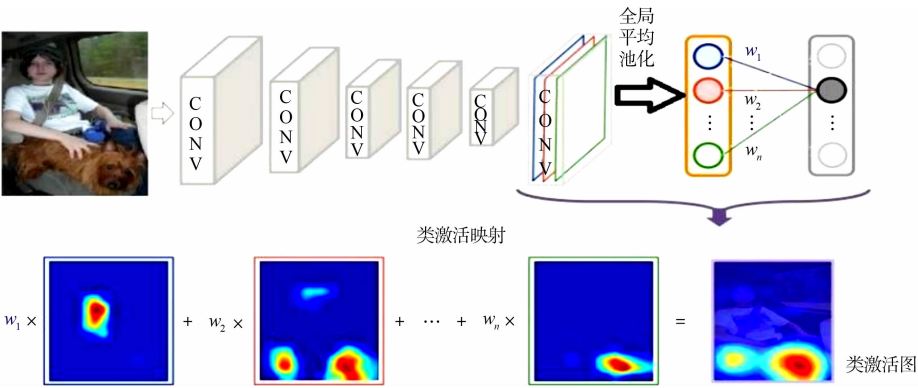


图 3 类激活映射过程

Fig. 3 Class activation mapping process

类激活图可用于对 CNN 的预测决策进行透视,而 CNN 由图像的不同语义区域触发,用于不同的预测.此外,网络的深层功能可以用于通用本地化,使用最新训练的 SVM 的权重来生成类激活映射,然后获得类特定的显著映射.随着神经网络训练次数的演化,重要的特征点对应的权重,通过一层层传递,会比较大;反之,不能作为特征的点,就会几乎没有权重.对于 CAM 图而言,权重大的点会从红到蓝分布,颜色越红说明这个地方越是特征部位.图 4 为 CAM 类激活图.

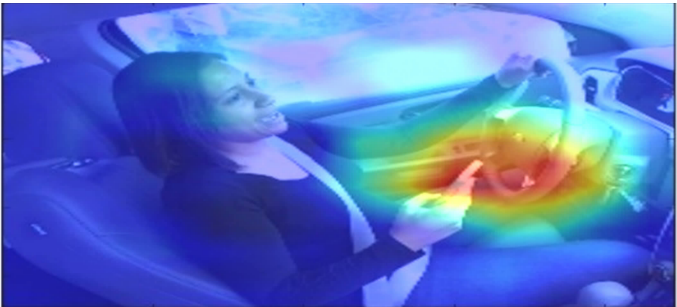


图 4 CAM 类激活图

Fig. 4 CAM activation diagram

3 基于深度可分离卷积网络的驾驶状态识别算法

3.1 总体框架

采用卷积神经网络(CNN),把图片的部分相邻像素变成长宽更小,高度更深的单元,从而提炼出局部的特征;然后一层一层将特征往后映射,并提取出最终的特征. MF-Net 利用深度可分离卷积^[13]作为构建块,在一定程度上可降低网络的复杂度.深度可分离卷积实质上是标准卷积分成了两步:1)深度卷积,即对每个输入通道单独使用一个卷积核处理;2)逐点卷积,即 1×1 卷积用于将深度卷积的输出组合起来,并且深度卷积和逐点卷积的输入与输出都是相同的.

为更好地理解,在此举例说明.首先输入的是 2 维数据,然后进行 3×3 卷积并输出 3 维数据.标准卷积参数和深度可分离卷积参数示意图,分别如图 5,6 所示.标准卷积参数数量为 $2 \times 3 \times 3 \times 3 = 54$,而深度可分离卷积参数数量为 $2 \times 3 \times 3 \oplus 2 \times 1 \times 1 \times 3 = 24$.将两个参数数量进行对比,可得 $24 \div 54 \approx 0.44$,可以看到深度可分离卷积的参数数量是正常卷积参数数量的一半,甚至更少,只不过在输入与输出维度相差不大的情况下,效果不那么明显.

当计算能力非常小时,信息容量在网络性能中起着更重要的作用.常规地,减少信道的数量可使紧凑的网络架构适应一定的复杂性.在采用慢速下采样方案的情况下,网络变得太狭窄而无法编码获得足够的信息,从而导致严重的性能下降.因此拟采用快速下采样策略,可以在有限的计算能力下实现了较高精度和较高效率.

快速下采样策略是指,在现代的 CNN 采用分层架构中,同一阶段内各层的空间尺寸保持相同,而下阶段的尺寸则通过下采样来减小.鉴于计算能力有限,紧凑型网络的特征表示能力和信息容量均受到限制.不同的下采样策略为紧凑型网络提供了详细的功能表示和大量的信息容量之间的折衷.在

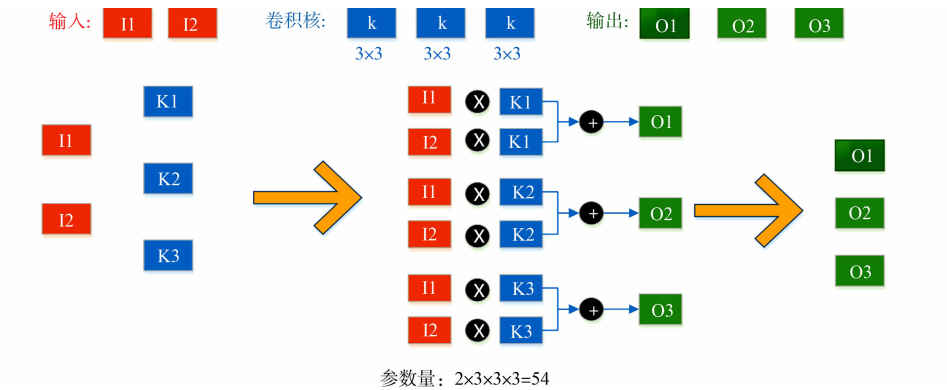


图 5 标准卷积参数示意图

Fig. 5 Schematic diagram of standard convolution parameters

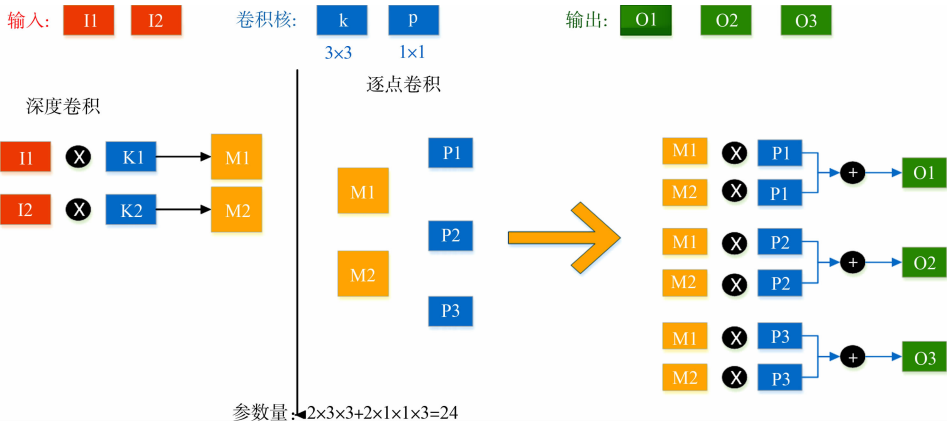
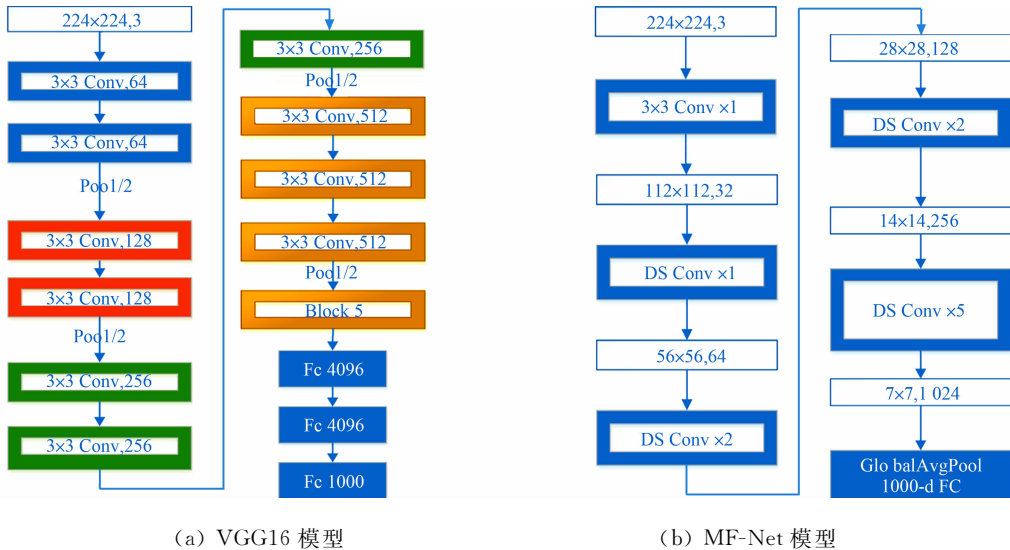


图 6 深度可分离卷积参数示意图

Fig. 6 Schematic diagram of depth separable convolution parameters

低速下采样策略中,下采样是在网络的后续层中执行的,故更多的层具有较大的空间尺寸;相反,在网络的开始以快速的下采样策略执行下采样,这将大大降低计算成本.因此,在固定的计算预算下,慢速下采样策略倾向于生成更详细的功能,而快速下采样策略则可以增加通道数量并允许更多信息进行编码.

不同 CNN 模型的结构对比示意,如图 7 所示.从图 7 可知:VGG16 模型的输入图像是 $224\text{ px}\times$



(a) VGG16 模型

(b) MF-Net 模型

图 7 不同 CNN 模型的结构对比示意

Fig. 7 Comparison of structure of different CNN models

$224\text{ px}\times 3$,结构中通道数翻倍,由 $64\rightarrow 128\rightarrow 256\rightarrow 512$ 后保持不变,不再翻倍,而高和宽变均减半,即 $224\rightarrow 112\rightarrow 56\rightarrow 28\rightarrow 14\rightarrow 7$;MF-Net 模型采用更快的下采样策略并利用更多的渠道,从而扩大了信息

容量并提高了性能. DS Conv:深度方向可分离卷积. 每个深度可分离卷积都由两层组成.

基于以上见解,决定在 MF-Net 的体系结构中采用快速下采样策略,并通过在网络开始时以较大步幅连续应用深度可分离卷积来实现,同时将特征提取过程推迟到最小分辨率. 这里不使用最大池化,因为发现它不能提高性能,反而会引入额外的计算. 所提出的 MF-Net 接受尺寸为 $224\text{ px}\times 224\text{ px}$ 的图像,并在前 2 层内执行 4 倍下采样,其后在 12 层内执行 32 倍下采样. 更具体地说,这 12 个层由 1 个标准卷积层,5 个深度可分离卷积(每个都有一个深度卷积层和一个点向卷积层)和 1 个深度卷积层组成.

MF-Net 采用 24 层的堆叠架构,包括 1 个标准卷积层,11 个深度可分离卷积和 1 个全连接层,如表 1 所示. 表 1 中:“/2”表示该层的步幅为 2;DWConv 表示深度卷积. 由表 1 可知:在建议的体系结构中可以利用更多的层和更多的通道. 这里利用 6 个深度可分离卷积来提高生成特征的表示能力,前 5 个深度可分离卷积的输出通道是 512 个,而最后一个是 1 024 个. 信道数量的增加有助于获取更大的信息容量,这对于在有限的计算资源下的网络性能至关重要.

表 1 MF-Net 的网络架构图
Tab.1 MF-Net network architecture diagram

输出大小/px	层	输出大小/px	层
224×224	图像参数		
112×112	3×3 Conv,32,/2	14×14	3×3 DWConv,256 1×1 Conv,256
	3×3 DWConv,32,/2		3×3 DWConv,256,/2
56×56	1×1 Conv,64		1×1 Conv,512
	3×3 DWConv,64,/2		3×3 DWConv,512(4 个卷积组)
28×28	1×1 Conv,128	7×7	1×1 Conv,512(4 个卷积组)
	3×3 DWConv,128		3×3 DWConv,512
	1×1 Conv,128		1×1 Conv,1 024
14×14	3×3 DWConv,128,/2		Global Average Pooling
	1×1 Conv,256	1×1	1000-d FC

与此同时,引入微调策略来训练框架. 结果表明,该方法取得良好的效果. 为提高训练速度,减轻深层神经网络过度参数化,防止过拟合,经过试验在 BatchNormalization 之后将 Dropout 比率设置为 0.5,以减少复杂的共适应关系,从而创造一个稀疏的网络架构. 另外,采用迁移学习以减少训练时间.

3.2 损失函数

在继续构建模型时,重要的是选择正确的度量标准来衡量其性能. 准确性是第一个指标,但不是分类问题的唯一指标. 准确性仅考虑预测的正确性,即预测的标签是否与真实标签相同. 为此在神经网络最后一层中,用损失函数来评价网络模型输出的预测值 $\hat{Y}=f(X)$ 与真实值 Y 之间的差异. 对数损失(与交叉熵有关)用于衡量分类模型的性能,其中预测输入为 0 到 1 之间的概率值. 深度学习模型的目标是最小化该值,损失值越小,网络模型性能越好. 理想模型的对数损失为 0,且随着预测概率与实际标签的偏离而增加. 这里使用 $LOSS(Y,\hat{Y})$ 来表示损失函数,计划采用 Categorical_Crossentropy 的计算原理来使用 LOSS,其数学表达式为

$$\lg \text{LOSS} = -\frac{1}{N} \cdot \sum_{i=1}^N \sum_{j=1}^M y_{i,j} \cdot \lg P_{i,j}. \tag{1}$$

而常用的二分分类模型一般采用的是 Binary_Crossentropy 的方法,其数学表达式为

$$\lg \text{LOSS} = -\frac{1}{n} \cdot \sum_{i=1}^n [y_i \cdot \lg(\hat{y}_i) + (1 - y_i) \cdot \lg(1 - \hat{y}_i)]. \tag{2}$$

面对的问题是 10 个分类,而不仅仅是 2 个分类,因此应采用 Categorical_Crossentropy 的方法,而不该采用 Binary_Crossentropy 的方法.

3.3 优化器选择

在机器学习算法中,有时候需要对原始模型构建损失函数,然后通过优化算法对损失函数进行优化,以便寻找到最优的参数,使得损失函数的值最小. Adam 具有如下 4 个特点:1) 结合了 Adagrad 善于处理稀疏梯度和 RMSprop 善于处理非平稳目标的优点;2) 对内存需求较小;3) 为不同的参数计算

不同的自适应学习率;4) 同样适用于大多非凸优化,即适用于大数据集和高维空间. 由于 Adam 的这些特点,因此选择 Adam 作为网络优化器. 在此选择自适应矩估计(Adam)来做优化器,其本质上是带有动量项的 RMSprop,而 RMSprop 优化器可以算作 Adadelata 的一个特例. 其中: $\rho=0.5 \times E\left|g^2\right|_t=\rho \times E\left|g^2\right|_{t-1}+(1-\rho) \times g_t^2$,就变为了求梯度平方和的平均数.

如果再求根的话,就变成了 RMS(均方根),则有 $\text{RMS}\left|g\right|_t=\sqrt{E+\epsilon}$. 此时,这个 RMS 就可以作为学习率\eta 的一个约束,即

$$\Delta x_t=-\frac{\eta}{\text{RMS}\left|g\right|_t} \times g_t$$

(3)

RMSprop 有两个特点:1) RMSprop 依然依赖于全局学习率;2) RMSprop 是 Adagrad 的一种发展和 Adadelata 的变体,效果趋于二者之间,适合处理非平稳目标.

Adam 方法将惯性保持和环境感知两个优点集于一身. 一方面,Adam 记录梯度的一阶矩,即过往梯度与当前梯度的平均,这体现了惯性保持;另一方面,Adam 还记录梯度的二阶矩,即过往梯度平方与当前梯度平方的平均. 这类似 AdaGrad 方法,体现了环境感知能力,为不同参数产生自适应的学习率. 一阶矩和二阶矩采用类似于滑动窗口内求平均的思想进行融合,即当前梯度和近一段时间内梯度的平均值,时间久远的梯度对当前平均值的贡献呈指数衰减. 具体来说,它利用梯度的一阶矩估计和二阶矩估计动态调整每个参数的学习率. Adam 的优点主要在于经过偏置校正后,每一次迭代学习率都有个确定范围,使得参数比较平稳,其计算式为

$$\left. \begin{aligned} m_t &= \mu \times m_{t-1} + (1-\mu) \times g_t, & n_t &= v \times n_{t-1} + (1-v) \times g_t^2, \\ \hat{m}_t &= \frac{m_t}{1-\mu^t}, & \hat{n}_t &= \frac{n_t}{1-v^t}, & \Delta \theta_t &= -\frac{\hat{m}_t}{\sqrt{\hat{n}+\epsilon}} \times \eta. \end{aligned} \right\}$$

(4)

4 实验结果与分析

实验环境包括硬件设备和软件配置,测试所用的计算机配置为:GPU Tesla K80, RAM 25.51GB. 软件环境为:Linux 内核版本 4.14.137+;Python3.6;tensorflow 1.15.0;keras 2.2.5. 将文中所提出的算法与其他算法进行比较,结果如表 2 所示. 表 2 中: A_T, A_V 分别为训练准确率和验证准确率; L_T, L_V 分别为训练损失和验证损失; M 为参数量; N 为模型大小;modified VGG16 模型是参考文献[14]的相关研究进行实验得到的.

表 2 不同算法的实验结果比较

Tab. 2 Comparison of experimental results of different algorithms

模型	$A_T/\%$	L_T	$A_V/\%$	L_V	$M/\times 10^6$	N/MB	$A_T/\%$	
							第 5 次	第 10 次
VGG16	43.92	1.703 9	66.71	1.501 8	16.82	128	40.21	43.92
modified VGG16	87.72	0.365 3	79.32	0.809 5	14.96	114	63.09	78.54
ResNet 50	89.05	0.321 0	83.13	0.732 7	24.59	188	75.76	82.87
Xception	90.42	0.274 6	83.72	0.600 1	24.54	187	61.26	76.98
MF-Net	91.07	0.278 2	84.53	0.491 5	3.24	25	80.20	86.58

由表 2 可知:MF-Net 模型在相同的计算能力下比其他几种模型都有一定的改进. MF-Net 模型验证准确率比 ResNet 50 模型提升了 1.4%,验证损失函数值提升 0.241 2;MF-Net 模型验证准确率比 Xception 模型提升了 0.81%,验证损失函数值提升了 0.108 6. 将这些改进归因于 MF-Net 中快速下采样策略的有效性. 最初的 Xception 模型采用后期慢速采样策略,更多的层具有相对较大的特征图,从而计算量也更大;而为保持计算效率,Xception 模型相对较窄,这就限制了信息容量. MF-Net 采用更快的下采样策略,这样可以利用更多的信道并减轻信息容量的下降.

由表 2 可知:MF-Net 模型的参数量最少,为 3.24×10^6 个,这也意味着训练出来的模型最小,为 25 MB,而其他的训练模型都大于 100 MB,需要的存储容量大,不利于部署. 如存储 ResNet 50 模型的权重值文件大小为 188 MB,模型过大不利于移植到嵌入式系统中. Xception 模型训练时间过长,调参难

度较大. 对于移动或嵌入式设备,内存和缓存资源是有限的,复杂的计算图可能导致频繁的内存/高速缓存切换,从而降低实际推断速度. 通过分析表 2 的迭代次数与准确率间的关系发现:MF-Net 网络可以在相同迭代次数情况下可以达到更高的准确率. 另外在实验过程中,ResNet 50 模型和 Xception 模型在迭代将近 40 次时才达到表 2 中的实验结果,而 MF-Net 模型只迭代了 19 次,这也就是说 MF-Net 模型可以降低计算成本. 当前的 MF-Net 深度学习网络通过构建非循环计算图来实现神经网络的推理,继承了 VGG16 的简单体系结构,结合了深度可分离卷积,并采用了前期快速下采样策略,使得 MF-Net 对工程实施非常友好,并在移动物理设备上比较高效.

F1-score 是分类问题的一个衡量指标,它是精确率和召回率的调和平均数,最大为 1,最小为 0,一些多分类问题常常将 F1-score 作为最终测评的方法. 将文中提出的算法与其他算法的 F1-score 值进行比较,结果如表 3 所示. 由表 3 可知:文中所提算法在 F1-score 均值比较中表现最好,其值为 0.87.

表 3 不同算法的 F1-score 值比较
Tab. 3 Comparison of F1-score values of different algorithms

模型	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	平均值
VGG16	0.80	0.95	0.87	0.90	0.70	0.96	0.73	0.87	0.46	0.79	0.82
modified VGG16	0.90	0.90	0.98	0.80	0.80	0.88	0.89	0.81	0.60	0.75	0.84
ResNet 50	0.80	0.92	0.96	0.90	0.80	0.92	0.89	0.86	0.58	0.67	0.85
Xception	0.70	0.96	0.95	0.90	0.80	0.87	0.83	0.81	0.46	0.64	0.81
Ours	0.80	0.96	0.93	0.90	0.80	0.97	0.87	0.91	0.64	0.79	0.87

5 结束语

文中提出利用深度可分离卷积作为网络结构 MF-Net 构建块,并结合快速下采样策略构建驾驶状态识别系统. 与 VGG 16,ResNet 50 等神经网络相比,该模型的深度可分离卷积大大减少了参数量,快速下采样方案的运用增加了网络的信息容量. 这种设计有助于性能的提升且所需存储容量小,便于今后嵌入式系统的移植. 实验结果表明,文中所提的 MF-Net 网络模型在驾驶员状态识别任务上具有有效性与优越性.

参考文献:

[1] TAMAS V, MATIES V, NNEBE S E, *et al.* Real-time distracted drivers detection using deep learning[J]. American Journal of Artificial Intelligence, 2019, 3(1): 1-8.

[2] 吴绍斌, 高利, 王刘安. 基于脑电信号的驾驶疲劳检测研究[J]. 北京理工大学学报, 2009, 29(12): 1072-1075. DOI: 10.15918/j.tbi.t1001-0645. 2009. 12. 013.

[3] JAP B T, LAL S, FISCHER P, *et al.* Using EEG spectral components to assess algorithms for detecting fatigue[J]. Expert Systems with Applications, 2009, 36(2): 2352-2359. DOI: 10.1016/j.eswa. 2007. 12. 043.

[4] LI Mingai, ZHANG Cheng, YANG Jinfu. An EEG-based method for detecting drowsy driving state[C]// Seventh International Conference on Fuzzy Systems and Knowledge Discovery. Yantai: IEEE Press, 2010: 2164-2167. DOI: 10.1109/FSKD. 2010. 5569757.

[5] 王玉海, 宋健, 李兴坤. 基于模糊推理的驾驶员意图识别研究[J]. 公路交通科技, 2005, 22(12): 116-121, 142. DOI: 10.3969/j.issn.1002-0268. 2005. 12. 030.

[6] ZHAO C H, ZHANG Bailing, HE Jie, *et al.* Recognition of driving postures by contourlet transform and random forests[J]. IET Intelligent Transport Systems, 2012, 6(2): 161-168. DOI: 10.1049/iet-its. 2011. 0116

[7] BERRI R A, SILVA A G, PARPINELLI R S, *et al.* A pattern recognition system for detecting use of mobile phones while driving[C]// International Conference on Computer Vision Theory and Applications (VISAPP). Lisbon: IEEE Press, 2014: 411-418. DOI: 10.5220/0004684504110418.

[8] CRAYE C, KARRAY F. Driver distraction detection and recognition using RGB-D sensor[J/OL]. eprint arXiv: 1502.00250, 2015: 1-11. [2015-02-01]. <http://de.arxiv.org/pdf/1502.00250>.

[9] LE T H N, ZHENG Yutong, ZHU Chenchen, *et al.* Multiple scale faster-rcnn approach to driver's cell-phone usage and hands on steering wheel detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern

- Recognition Workshops. Las Vegas; IEEE Press, 2016: 46-53. DOI: 10. 1109/CVPRW. 2016. 13.
- [10] RUIZ N, CHONG E, REHG J M. Fine-grained head pose estimation without keypoints[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City; IEEE Press, 2018: 2074-2083. DOI: 10. 1109/CVPRW. 2018. 00281.
- [11] SIMARD P Y, STEINKRAUS D, PLATT J C. Best practices for convolutional neural networks applied to visual document analysis[C]// Seventh International Conference on Document Analysis and Recognition. Edinburgh; IEEE Press, 2003. DOI: 10. 1109/ICDAR. 2003. 1227801.
- [12] LIN Ming, CHEN Qiang, YAN Shuicheng. Network in network[EB/OL]. (2013-12-16) [2014-03-04]. <https://arxiv.org/abs/1312.4400>.
- [13] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico; IEEE Press, 2017: 1251-1258. DOI: 10. 1109/CVPR. 2017. 195.
- [14] 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, Salt Lake City; IEEE Press, 2018. DOI: 10. 1109/CVPRW. 2018. 00150.
- [15] SANG Haifeng, WANG Chuanzheng, HE Dakuo, *et al.* Multi-information flow CNN and attribute-aided reranking for person reidentification[J/OL]. Computational Intelligence and Neuroscience, 2019; 1-12. <http://downloads.hindawi.com/journals/cin/2019/7028107.pdf>. DOI: 10. 1155/2019/7028107.
- [16] LI Xiaoguang, LAM K M, QIU G P, *et al.* Examplebased image super-resolution with class-specific predictors[J]. Journal of Visual Communication and Image Representation, 2009, 20(5): 312-322. DOI: 10. 1016/j. jvcir. 2009. 03. 008.
- [17] NIE Weizhi, WANG Kun, WANG Hongtao, *et al.* The assessment of 3D model representation for retrieval with CNN-RNN networks[J]. Multimedia Tools and Applications, 2019, 78(1): 16979-16994. DOI: 10. 1007/s11042-018-7102-2.
- [18] JIJI C V, CHAUDHURI S. Single-frame image super-resolution through contourlet learning[J/OL]. EURASIP Journal on Advances in Signal Processing, 2006. <https://asp-eurasipjournals.springeropen.com/track/pdf/10.1155/ASP/2006/73767>. DOI: 10. 1155/asp/2006/73767.
- [19] YILDIRIM O, BALOGLU U B, TAN R S, *et al.* A new approach for arrhythmia classification using deep coded features and LSTM networks[J]. Computer Methods and Programs in Biomedicine, 2019, 176: 121-133. DOI: 10. 1016/j. cmpb. 2019. 05. 004.
- [20] 李键红, 吴亚榕, 吕巨建. 基于自相似性与多任务高斯过程回归的单帧图像超分辨率重建[J]. 光学精密工程, 2018, 26(11): 2814-2826. DOI: 10. 3788/OPE. 20182611. 2814.
- [21] RAJEEV R, SAMTH J A, KARTHIKEYAN N K. An intelligent recurrent neural network with long short-term memory (LSTM) BASED batch normalization for medical image denoising[J/OL]. Journal of Medical Systems, 2019, 43(8); [2019-06-15]. DOI: 10. 1007/s10916-019-1371-9.
- [22] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. arXiv: 1409. 1556, 2015. (2014-09-04) [2015-04-10]. <http://www.arxiv.org/pdf/1409.1556.pdf>.
- [23] SHI Wenzhe, CABALLERO J, HUSZÁR F, *et al.* Real-time single image and video super-resolution using an efficient subpixel convolutional neural network[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas; IEEE Press, 2016. DOI: 10. 1109/CVPR. 2016. 207.
- [24] OTSU N. A threshold selection method from gray-level histograms[J]. IEEE Transactions on Systems, Man and Cybernetics, 1979, 9(1): 62-66. DOI: 10. 1109/TSMC. 1979. 4310076
- [25] XU Xiaohong, WU Zhihui, CHEN Yu, *et al.* Plant root spatial distribution measurements based on the hough transformation[J]. Neurocomputing, 2014, 145: 209-220. DOI: 10. 1016/j. neucom. 2014. 05. 041.

(责任编辑: 黄仲一 英文审校: 吴逢铁)