

基于 EfficientNet 的异常驾驶行为识别方法

刘兰华¹ 蓝凡立² 梁国际¹

1. 南宁学院人工智能学院, 南宁, 530200
2. 广西民族大学数学与物理学院, 南宁, 530005

摘要 针对异常驾驶行为监控识别场景, 提出基于高效卷积网络 EfficientNet 的图片增强异常驾驶行为识别方法和预处理模型的迁移学习异常驾驶行为识别方法。首先加载图片时通过旋转、剪切、平移等多种方式进行图片增强, 接着使用在大规模图片数据库 ImageNet 上进行预训练的模型获取模型特征权重, 然后自定义预处理模型的全连接层使模型在异常驾驶行为数据集上的效率和效果更好。实验数据证明了提出的算法对异常驾驶行为识别准确率高达 99%, 是可行和有效的。

关键字 驾驶行为识别, 深度学习, 预处理模型, EfficientNet

Identification Method of Abnormal Driving Behavior Based On EfficientNet

Liu LanHua

School of Artificial Intelligence
Nanjing University
Nanning 530200, China;
lhliu5@iflytek.com

Lan Fanli

School of Mathematics and Physics
Guangxi University for Nationalities
Nanning 530005, China;
Lanfanli_gxun@163.com

Liang Guoji

School of Artificial Intelligence
Nanjing University
Nanning 530200, China;
gjliang@iflytek.com

Abstract—Aiming at the abnormal driving behavior monitoring and recognition scenario, this paper proposes a transfer learning abnormal driving behavior recognition method based on the picture enhancement abnormal driving behavior recognition method and pre-processing model of the EfficientNet. First, the picture is enhanced by rotating, cutting, panning, etc. when loading the picture, and then the model pre-trained on the large-scale picture database ImageNet is used to obtain the model feature weights, and then the full connection layer of the custom preprocessing model makes the model more efficient and effective on the abnormal driving behavior dataset. Experimental results show that the accuracy of the proposed algorithm is as high as 99%, and the feasibility and effectiveness is proved.

Key words—Driving behavior recognition, Deep learning, Pre-processing models, EfficientNet

1 引言

根据国家统计局显示,截至 2021 年,我国私人汽车拥有量为 26246 万辆,每年汽车发生交通事故数(起)为 15 万左右,其中大约有 4 万人死于交通事故。由于交通事故的发生不仅造成大量人员伤亡,给无数家庭带来不幸,而且严重影响着经济发展和社会稳定,已引起了各级政府的高度重视和关注。但是目前的交通探头仅可识别车辆前排人员安全带佩戴情况,这单一的功能已不能满足日益复杂的路况场景。而通过交通警察来约束司机的异常驾驶行为效果有限,还会造成极大的人力资源浪费。随着深度学习的迭代与发展,基于神经网络人工智能的行为识别技术,让智能识别判断异常驾驶行为有了从科幻走进了现实的可能。因此,开展驾驶行为识别算法的研究,具有较强的现实意义与应用价值。

罗冠泰^[1]提出基于卷积神经网络的分心驾驶行为识别,基于 ResNet18 的多标签分心驾驶行为识别方

法以及基于迁移学习的多分类分心驾驶行为识别方法,该方法在 State Farm 数据集上达到了 93.5% 的准确率。但是该方法的训练时间较长,而且随着 ResNet 网络深度增加时所产生的 Degradation 问题^[2],即准确率会先上升然后达到饱和,再持续增加深度则会导致准确率下降的情况。而当前最强的卷积网络 EfficientNet,在网络变大时模型效果提升明显,训练时间更快而且能够把精度上限进一步提升。

李此君等^{[3][4]}研究一种高效的异常驾驶行为正确识别分类的识别方法,提出了一种基于协方差流形和基于二分类思想的多类 LogitBoost 分类器的异常驾驶行为识别方法。该方法通过提取图像的纹理、颜色和梯度方向等基础特征,克服了基于单一特征识别驾驶行为的不足。然后,利用协方差流形进行多特征融合,实现消除特征冗余、降低由于不同特征数值差异过大而可能给图像处理及识别带来的影响的目的。该方法能大幅地提高多分类的正确率,针对相同检测目标的

正确识别率可达 98% 以上, 针对不同目标的正确识别率可达 81.08%, 有效提高了驾驶行为识别的效果。

王肖^[1]针对异常行为在驾驶员身上的表现区域, 研究设计了驾驶员眼部、头部和上半身三个关注区域的异常驾驶行为识别模型; 在建立异常行为识别模型的基础上提出基于横向与纵向估计相结合的异常行为分级预警模型; 为了避免重复检测相同的信息, 采用关键帧捕捉方法检测关键信息。提出将强化学习思想作为关键帧捕捉方法的策略, 使关键帧捕捉具备了根据驾驶员状态变化而切换策略的自适应能力。

宇泓儒^[1]针对现有异常变速行为识别方法存在的动态性不足、适应性差等问题, 综合运用以公交车辆历史轨迹数据为核心的公交运行数据, 提出了速度-加速度联合的公交驾驶员异常变速行为离散阈值确定方法, 和动态阈值拟合方法。然后在此基础上构造公交驾驶员在不同情形下的风险评价函数, 建立以公交运行综合安全风险最低为目标函数, 设计了适用于排班迭代优化的解初始化方法和基于退火算法框架的模型求解方法。

本文针对异常驾驶行为监控识别场景, 提出基于高效卷积网络 EfficientNet 的图片增强异常驾驶行为识别方法和预处理模型的迁移学习异常驾驶行为识别方法。并通过仿真实验, 说明所提出的算法对异常驾驶行为为识别的可行性和有效性。

2 EfficientNet 高效卷积网络

EfficientNet 源自 Google Brain 的论文 EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. 从标题也可以看出, 这篇论文最主要的创新点是 Model Scaling. 论文提出了 compound scaling, 混合缩放, 把网络缩放的三种方式: 深度、宽度、分辨率, 组合起来按照一定规则缩放, 从而提高网络的效果^[3]. EfficientNet 在网络变大时效果提升明显, 把精度上限进一步提升, 成为了当前最强网络。EfficientNet-B7 在 ImageNet 上获得了最先进的 84.4% 的 top-1 精度和 97.1% 的 top-5 精度, 比之前最好的卷积网络 (GPipe, Top-1: 84.3%, Top-5: 97.0%) 大小缩小 8.4 倍、速度提升 6.1 倍。图 1 是 EfficientNet 与其他网络的对比, 可以看到 EfficientNet 饱和值高, 并且到达速度快。

2.1 EfficientNet 算法原理

通常增加卷积网络参数可以获得更好的精度, 有的卷积网络会通过增加网络的 width 即增加卷积核的个数 (增加特征矩阵的 channels) 来提升网络的性能, 如 2 所示, 左边为 baseline 卷积网络, 右图为增加 width 使得网络变大。

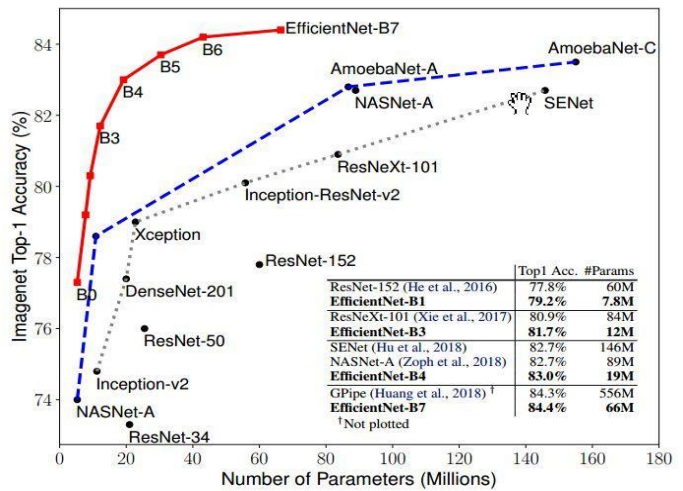


图 1 EfficientNet 与其他网络的对比效果

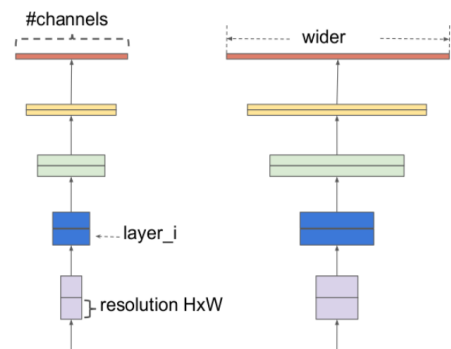


图 2 增加网络的宽度 width 以放大网络

增加网络的宽度 width 能够获得更高细粒度的特征并且也更容易训练, 但对于 width 很大而深度较浅的网络往往很难学习到更深层次的特征。

有的卷积网络会通过增加网络的深度即使用更多的层结构来提升网络的性能, 如图 3 所示。

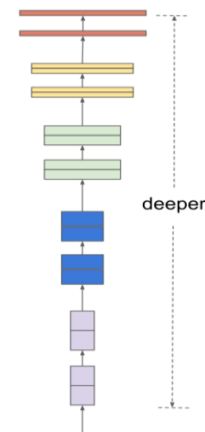


图 3 增加网络的深度 depth 以放大网络

增加网络的深度 depth 能够得到更加丰富、复杂的特征并且能够很好的应用到其它任务中。但网络的深

度过深会面临梯度消失，训练困难的问题。

有的卷积网络会通过增加输入网络的分辨率来提升网络的性能,如图4所示。

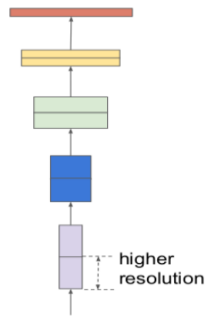


图 4 增加网络的分辨率resolution以放大网络

增加输入网络的图像分辨率resolution能够潜在获得更高细粒度的特征模板，但对于非常高的输入分辨率，准确率的增益也会减小。并且大分辨率图像会增加计算量。

直观上来讲，这三种缩放方式并不独立。对于

分辨率高的图像，应该用更深的网络，因为需要更大的感受野，同时也应该增加网络宽度来获得更细粒度的特征。

之前增加网络参数都是单独放大这三种方式中的一种，并没有同时调整，也没有调整方式的研究。EfficientNet使用了compound scaling方法，统一缩放网络深度、宽度和分辨率，如图5所示。

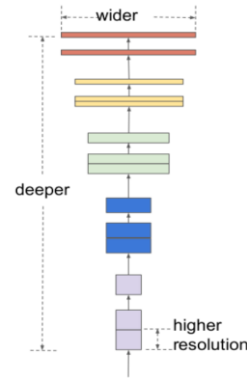


图 5 EfficientNet通过compound scaling以放大网络

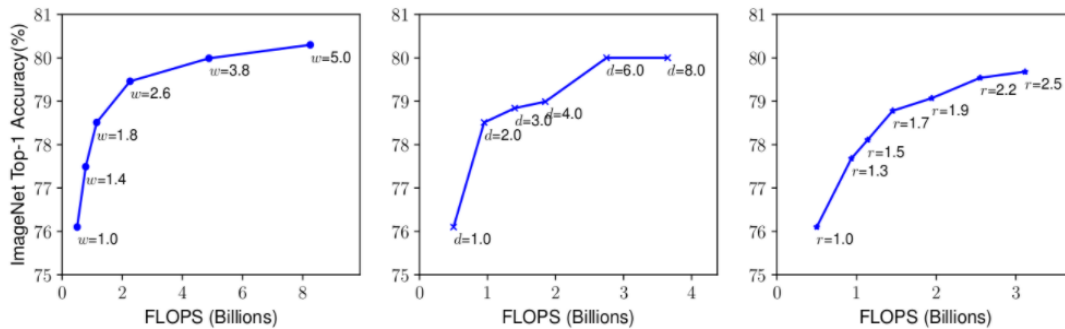


图 6 EfficientNetB-0单独增加width、depth以及resolution效果

2.2 卷积网络优化效果对比

单独对卷积网络的宽度width、深度depth以及分辨率resolution加强优化的效果非常有限，图7展示了在基准EfficientNetB-0上分别增加width（左图）、depth（中图）以及resolution（右图）后得到的统计结果。通过图6可以看出大概在Accuracy达到80%时就趋于饱和了，可见三个维度中任一维度的放大都可以带来精度的提升，但是随着倍率的越来越大，提升越来越小。

如果采用不同的depth, resolution组合，然后不断改变EfficientNetB-0网络的width就得到了如图7所示的4条曲线，通过分析可以发现在相同的FLOPs下，同时增加width、depth以及resolution的效果最好。

由图7可知EfficientNet网络要想达到最有效果，就需要的求得在有限的资源里面如何确定平衡这网络的深度depth，卷积通道的宽度width，还有输入图像的分辨率resolution。

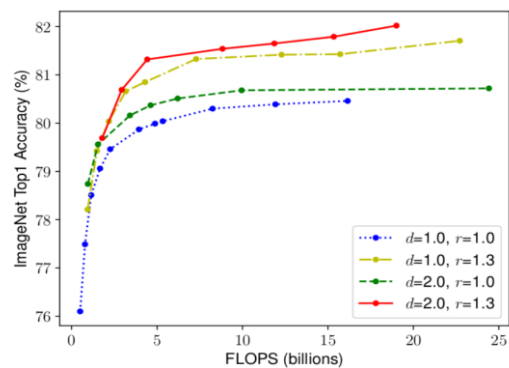


图 7 EfficientNetB-0同时增加width、depth以及resolution效果

2.3 EfficientNet 参数选择

为了能够统一 width, depth, resolution 的参数，EfficientNet 使用了混合缩放方法 (compound scaling method) ，利用一个混合因子 ϕ 去统一缩放这三者，

具体的计算公式如下:

$$\begin{aligned} \text{depth: } d &= \alpha^\phi \\ \text{width: } w &= \beta^\phi \\ \text{resolution: } r &= \gamma^\phi \\ \text{s.t. } \alpha \cdot \beta^2 \cdot \gamma^2 &\approx 2 \\ \alpha \geq 1, \beta \geq 1, \gamma &\geq 1 \end{aligned} \quad (1)$$

其中, α 、 β 、 γ 是常数, 可以确定由一个小网格搜索。直观地说, ϕ 是一个用户指定的系数, 它控制了有多少更多的资源可以用于模型缩放, 而 α 、 β 、 γ 分别指定了将这些额外的资源分配给网络宽度、深度和分辨率。

接下来, 先将复合系数 ϕ 固定为 1, 先假设有两倍以上计算资源可以用, 然后对 α 、 β 、 γ 进行网络搜索。对于 EfficientNet-B0 网络, 在约束条件下如下时

$$\alpha \times \beta^2 \times \gamma^2 \approx 2$$

其中, α 、 β 、 γ 分别取 1.2, 1.1 和 1.15 时效果最佳。然后固定 α 、 β 、 γ , 通过复合调整公式对基线网络进行扩展, 得到 B1 到 B7 网络。于是就有了图 1, EfficientNet 在 ImageNet 上的效果碾压, 而且模型规模比此前的 GPipe 小了 8.4 倍, 详细数据如表 1 所示, EfficientNet 使用了 compound scaling 后, 效果非常显著, 在不同参数量和计算量都取得了多倍的提升。

表 1 EfficientNet 与当前主流网络的性能参数对比

Model	Top-1 Acc.	Top-5 Acc.	#Params	Ratio-to-EfficientNet	#FLOPS	Ratio-to-EfficientNet
EfficientNet-B0	76.3%	93.2%	5.3M	1x	0.39B	1x
ResNet-50 (He et al., 2016)	76.0%	93.0%	26M	4.9x	4.1B	11x
DenseNet-169 (Huang et al., 2017)	76.2%	93.2%	14M	2.6x	3.5B	8.9x
EfficientNet-B1	78.8%	94.4%	7.8M	1x	0.70B	1x
ResNet-152 (He et al., 2016)	77.8%	93.8%	60M	7.6x	11B	16x
DenseNet-264 (Huang et al., 2017)	77.9%	93.9%	34M	4.3x	6.0B	8.6x
Inception-v3 (Szegedy et al., 2016)	78.8%	94.4%	24M	3.0x	5.7B	8.1x
Xception (Chollet, 2017)	79.0%	94.5%	23M	3.0x	8.4B	12x
EfficientNet-B2	79.8%	94.9%	9.2M	1x	1.0B	1x
Inception-v4 (Szegedy et al., 2017)	80.0%	95.0%	48M	5.2x	13B	13x
Inception-resnet-v2 (Szegedy et al., 2017)	80.1%	95.1%	56M	6.1x	13B	13x
EfficientNet-B3	81.1%	95.5%	12M	1x	1.8B	1x
ResNeXt-101 (Xie et al., 2017)	80.9%	95.6%	84M	7.0x	32B	18x
PolyNet (Zhang et al., 2017)	81.3%	95.8%	92M	7.7x	35B	19x
EfficientNet-B4	82.6%	96.3%	19M	1x	4.2B	1x
SENet (Hu et al., 2018)	82.7%	96.2%	146M	7.7x	42B	10x
NASNet-A (Zoph et al., 2018)	82.7%	96.2%	89M	4.7x	24B	5.7x
AmoebaNet-A (Real et al., 2019)	82.8%	96.1%	87M	4.6x	23B	5.5x
PNASNet (Liu et al., 2018)	82.9%	96.2%	86M	4.5x	23B	6.0x
EfficientNet-B5	83.3%	96.7%	30M	1x	9.9B	1x
AmoebaNet-C (Cubuk et al., 2019)	83.5%	96.5%	155M	5.2x	41B	4.1x
EfficientNet-B6	84.0%	96.9%	43M	1x	19B	1x
EfficientNet-B7	84.4%	97.1%	66M	1x	37B	1x
GPipe (Huang et al., 2018)	84.3%	97.0%	557M	8.4x	-	-

We omit ensemble and multi-crop models (Hu et al., 2018), or models pretrained on 3.5B Instagram images (Mahajan et al., 2018).

2.4 EfficientNet 网络结构

表2是EfficientNet-B0的网络框架, 其中B1-B7就是在B0的基础上修改Resolution, Channels以及Layers。从表中可以看出, 网络总共分成了9个Stage, 第一个Stage就是一个卷积核大小为3x3步距为2的普通卷积层(包含BN和激活函数Swish); Stage2~Stage8都是在重复堆叠MBConv结构, 其中最后一列的Layers表示该Stage重复MBConv结构多少次; Stage9由一个普通的1x1的卷积层(包含BN和激活函数Swish)一个平均池化层和一个全连接层组成。表格中每个MBConv后会跟一个数字1或6, 这里的1或6就是倍率因子n即MBConv中第一个1x1的卷积层会将输入特征矩阵的channels扩充为n倍, 其中k3x3或k5x5表示MBConv中Depthwise Conv所采用的卷积核大小。Channels表示通过该Stage后输出特征矩阵的Channels。

MBConv 其实就是 MobileNetV3 网络中的 InvertedResidualBlock, 但也有些许区别。一个是采用的激活函数不一样 (EfficientNet的MBConv中使用的都是Swish激活函数), 另一个是在每个MBConv中都加入了SE (Squeeze-and-Excitation) 模块。MBConv结构如图8所示, 主要由一个1x1的普通卷积(升维作用, 包含BN和Swish), 一个kxk的Depthwise Conv卷积(包含BN和Swish) k的具体值可看EfficientNet-B0的网络框架主要有3x3和5x5两种情况, 一个SE模块, 一个1x1的普通卷积(降维作用, 包含BN), 一个Droupout层构成。其中, SE模块如图9所示, 由一个全局平均池化, 两个全连接层组成。第一个全连接层的节点个数是输入该MBConv特征矩阵channels的1/4, 且使用Swish激活函数。第二个全连接层的节点个数等于Depthwise Conv层输出的特征矩阵channels, 且使用Sigmoid激活函数。

表 2 EfficientNet-B0网络框架

Stage i	Operator \mathcal{F}_i	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels \hat{C}_i	#Layers \hat{L}_i
1	Conv3x3	224 × 224	32	1
2	MBCov1, k3x3	112 × 112	16	1
3	MBCov6, k3x3	112 × 112	24	2
4	MBCov6, k5x5	56 × 56	40	2
5	MBCov6, k3x3	28 × 28	80	3
6	MBCov6, k5x5	14 × 14	112	3
7	MBCov6, k5x5	14 × 14	192	4
8	MBCov6, k3x3	7 × 7	320	1
9	Conv1x1 & Pooling & FC	7 × 7	1280	1

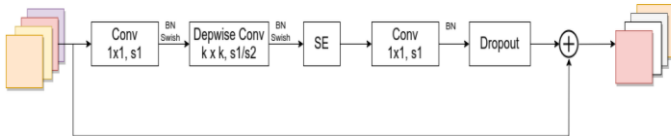


图 8 MBConv结构

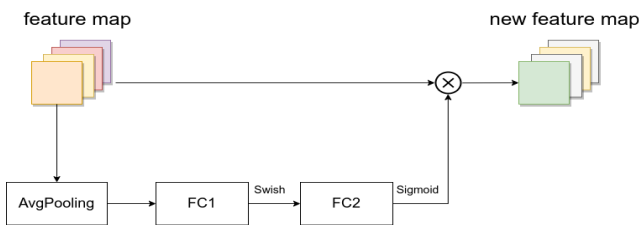


图 9 SE模块结构

根据图8、图9给出的结构，我们就可以搭建出EfficientNetB0网络。其他版本的详细参数如表3所示。

表 3 EfficientNetB0-B7详细参数

Model	input_size	width_coefficient	depth_coefficient	drop_connect_rate	dropout_rate
EfficientNetB0	224x224	1.0	1.0	0.2	0.2
EfficientNetB1	240x240	1.0	1.1	0.2	0.2
EfficientNetB2	260x260	1.1	1.2	0.2	0.3
EfficientNetB3	300x300	1.2	1.4	0.2	0.3
EfficientNetB4	380x380	1.4	1.8	0.2	0.4
EfficientNetB5	456x456	1.6	2.2	0.2	0.4
EfficientNetB6	528x528	1.8	2.6	0.2	0.5
EfficientNetB7	600x600	2.0	3.1	0.2	0.5

3 实验及结果分析

3.1 数据集的准备

本文采用凯格勒平台公开的驾驶行为识别竞赛数据集[4]进行实验,该数据集包含 10 种驾驶行为, 第一种为安全驾驶, 剩余 9 种为异常驾驶行为, 分别为右手玩手机、右手打电话、左手玩手机、左手打电话、调收音机、喝东西、向后伸手、整理头发和与乘客交谈。数据集 22424 张图像数据,图片大小都为 640 × 480。为验证模型训练后真实的驾驶行为识别能力,将

数据集的 20%单独划分为验证集和测试集。数据集分布情况如表 4 所示。

表 4 驾驶行为数据集分布情况

图片类别编号	图片类别名称	图片数目
c0	安全驾驶	2489
c1	右手玩手机	2267
c2	右手打电话	2317
c3	左手玩手机	2346
c4	左手打电话	2326
c5	调收音机	2312
c6	喝东西	2325
c7	向后伸手	2002
c8	整理头发	1911
c9	与乘客交谈	2129
总计		22424

3.2 数据预处理

原始图片的大小、位置、像素等信息不一定能满足模型的要求, 而且图片如果比较多, 一次全部加载需要耗费的内存和CPU资源也比较多, 为了优化图片的加载速度和获取图片数据的更多隐藏特征, 本文在EfficientNet模型训练之前通过ImageDataGenerator图片生成器在batch中进行了旋转、垂直和水平平移、剪切变换、曝光率增强、翻转等多种图片增强方式, 同时还打乱数据加载顺序, 对图片数据进行归一化和one-hot编码, 转换成满足EfficientNet模型要求的数据集格式, 达到增强模型泛化的效果。

图片增强预处理后, 模型对于每个类别图片的关键特征像素的所在位置可以通过热力图直观的呈现出来, 9种异常驾驶行为的关键特征热力图如图10所示。

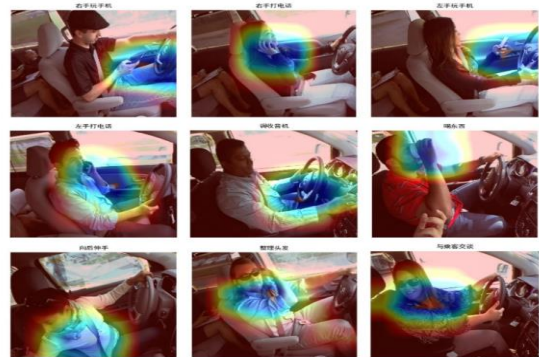


图 10 异常驾驶行为的关键特征热力图

3.3 模型训练与测试

迁移学习在计算机视觉领域中是一种很流行的方法[5], 它通过运用已有的知识来学习新的知识, 找到已有知识和新知识之间的相似性, 建立精确的模型, 耗时更短。为保留最真实的图片像素, 本次选择与图片大小最相近的EfficientNetB5模型, 并且使用在大规模

模图片数据库ImageNet上进行预训练过的模型,该数据集一共有20000多类标注好的数据集,共包含超过1400万张图片。在构造EfficientNetB5模型时通过指定模型初始化的权重检查点为imagenet,以获得预训练卷积网络的通用权重,预训练模型只用于特征提取,所以不包含最上层的全连接层,本文使用10个驾驶行为类别做为全连接层Dense的输出。EfficientNetB5的卷积基有28534010个参数,非常多,在编译和训练模型之前,先冻结卷积基确保训练过程中其权重保持不变。为了在训练期间可以使用数据增强,本文对预训练conv_base模型进行扩展,并将其当作一个卷积层加入Sequential模型进行编译以确保修改的卷积网络生效。本文编译之后的EfficientNetB5模型参数如图11所示。

```

Model: "sequential"
-----
Layer (type)                Output Shape              Param #
-----
efficientnet-b5 (Model)      (None, 15, 15, 2048)     28513520
-----
global_average_pooling2d (G1 (None, 2048)
-----
dense (Dense)                (None, 10)                20490
-----
Total params: 28,534,010
Trainable params: 3,436,682
Non-trainable params: 25,097,328

```

图 11 EfficientNetB5 模型编译后的参数

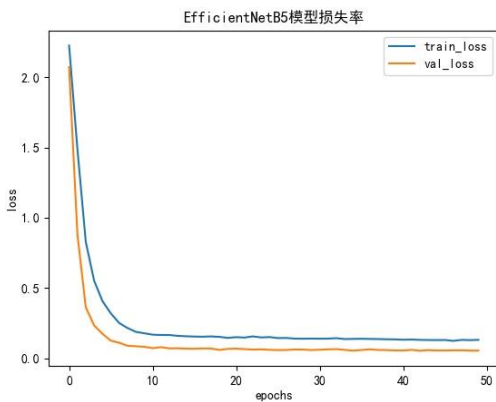


图 12 EfficientNetB5模型损失率

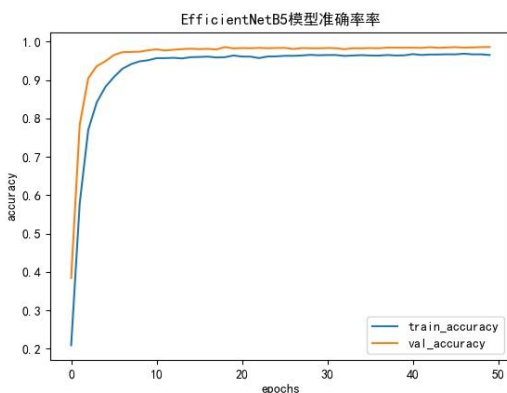


图 13 EfficientNetB5模型准确率

因为本地内存资源、CPU和显卡性能有限, EfficientNetB5模型每次只读取64张图片进行训练,考虑到学习率对卷积网络的模型收敛效果的影响比较大^[6],设置了学习率区间为[0.00001,0.0005],学习率指数衰减因子为0.6,学习率会随着模型的50次epochs训练迭代进行递增和递减变化以获得最优的模型效果。模型的训练结果表明,模型训练时间较短,64张图片仅训练633秒,在训练集上面的准确率为96.52%,在测试集上的准确率为98.59%,模型效果比较理想,未出现过拟合现象。模型Loss损失精值曲线和Acc精度值曲线如图12和图13所示,由两个图可以看出模型收敛效果速度快。

4 结束语

本文提出了基于当前最强的卷积网络 EfficientNet 的高效卷积网络识别异常驾驶行为的方法,模型高效准确;同时还采用迁移学习,并通过图片增强和增加神经网络模型神经元数量,提高模型识别准确率,减少模型的大量图片数据上的训练时间。通过最终的实验数据证明本文设计模型在异常驾驶行为识别有更好的识别性能,并且具备更好的性能和效率,下一步将考虑将算法应用到实时道路监控视频中,并集成语音提醒功能,及时提醒纠正驾驶员的危险驾驶行为,让交通生活更智能更安全。

参考文献

- [1] 翟羽. 基于卷积神经网络的人脸识别系统设计与实现[D]. 西安: 西安电子科技大学, 2019
- [2] C. S. Wickramasinghe C S, D. L. Marino and M. Manic, "ResNet Autoencoders for Unsupervised Feature Learning From High-Dimensional Data: Deep Models Resistant to Performance Degradation," in IEEE Access, vol. 9, pp. 40511-40520, 2021.
- [3] 李此君,刘云鹏.基于协方差流形的异常驾驶行为识别方法[J].太赫兹科学与电子信息学报. 2018,16(02):323-329
- [4] 李此君,刘云鹏.基于协方差流形和 LogitBoost 的异常驾驶行为识别方法[J].激光与光电子学进展. 2018,55(11):338-345
- [5] 王肖.典型异常驾驶行为识别与预警方法研究[D].重庆: 重庆邮电大学, 2020
- [6] 宇泓儒.基于公交运行大数据的异常驾驶行为识别与排班优化方法研究[D].扬州: 扬州大学, 2022
- [7] 张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. 计算机学报, 2019, 42(3): 453-482.
- [8] Anthony G,William C,et al.State Farm Distracted Driver Detection[EB/OL].<https://www.kaggle.com/c/state-farm-distracted-driver-detection>,2016
- [9] 庄福振, 罗平, 何清. 迁移学习研究进展[J]. 软件学报, 2015, 26(1): 26-39.
- [10] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.