

基于改进 YOLOv4 算法的疲劳驾驶检测^{*}

李昭慧 张玮良

(西安科技大学 通信工程学院 西安 710600)

摘 要: 提出一种基于深度学习的改进 YOLOv4 算法的疲劳驾驶检测。首先通过使用迁移学习将 VOC 数据集已训练好权重作为预权重进行训练,其次在训练中将框架中的特征金字塔 SPP 结构前后增加卷积,提高框架对深层特征的提取,再引入空洞卷积技术在不损失图片信息情况下增加卷积输出感受野,增加图片位置信息的获取能力。实验结果表明,改善后的 YOLOv4 算法在测试中 mAP 值达到 97.29%,相较原 YOLOv4 算法提高了 1.98%,其中对眼部部位的检测提高了 6%。并在检测中加入时延判定,避免其他行为影响结果,使误判概率大大减小。

关键词: 深度学习;YOLOv4;疲劳驾驶检测;迁移学习;空洞卷积

中图分类号: TP391.4 **文献标识码:** A **国家标准学科分类代码:** 510.4

Fatigue driving detection based on improved YOLOv4 algorithm

Li Zhaohui Zhang Weiliang

(School of Communication Engineering, Xi'an University of Science and Technology, Xi'an 710600, China)

Abstract: This paper propose an improved YOLOv4 algorithm for detecting fatigue driving. First, under the transfer learning, the weight of VOC dataset has been trained as pre weight for training. Then in the training, convolution is added before and after the SPP structure of feature pyramid in the frame to improve the extraction of deep features. And introduce the dilated convolution to increase convolution output receptive field and the ability to obtain image location information. Experimental results show that the mAP value of the improved YOLOv4 algorithm during the test is 97.29%, 1.98% higher than raw YOLOv4 algorithm, the detection of eye parts increased by 6%. Add the frame delay to the detection, for avoiding other behaviors affecting the results and reducing the probability of misjudgment.

Keywords: deep learning;YOLOv4;fatigue driving detection;transfer learning;dilated convolution

0 引 言

随着经济飞速发展,至 2020 年 6 月底我国汽车数量高达 2.7 亿辆,并且全国以每年 2 000 万辆的速度增长。与此同时我国交通事故数量也居高不下,而每年在众多交通事故中有超过 20% 是因疲劳驾驶引发的事故。疲劳驾驶已然成为行车事故的主要原因之一。因此,疲劳驾驶的实时检测与反馈成为保护驾驶员生命财产安全的必要手段。本文将对疲劳驾驶检测部分进行分析。

目前,国内外的疲劳驾驶的检测方法主要通过对目标特征进行实时检测,根据检测数据进行分析与结论。例如,文献[1]根据监控驾驶员生理颈腰部肌和脑电信号来判别疲劳驾驶,该方法通过生理特征的信号参数建立模型来辨别疲劳,综合正确率达 95% 以上,但是复杂的检测布置和

对驾驶员的入侵性使其较难广泛应用;文献[2]提出了通过车载传感器采集车辆运动特征样本和轨迹来检测疲劳驾驶,但不能保证检测实时性。因此,本文将通过驾驶员面部特征作为目标进行疲劳驾驶检测。

驾驶员疲劳驾驶的表现较突出的特征就是驾驶员的面部特征,如眼睛、嘴巴。例如文献[3]通过伪 3D 卷积神经网络和注意力机制结合检测驾驶员面部特征,但该方法的网络结构过于繁琐,搭建较为困难。近年来,基于卷积神经网络的深度学习技术随着计算机软件的发展而广泛应用^[4]。相比传统图像识别检测技术,基于深度学习的图像识别与检测在精密度和实时性都超过传统图像识别检测。文献[5]通过深度卷积神经网络提取驾驶员嘴巴与眼睛进行疲劳预警,但是平均准确率为 91.9%,准度略低。由此,本文将使用准确度较高、框架结构较简洁的算法进行改进。

收稿日期:2021-05-12

^{*} 基金项目:国家自然科学基金(61705178)项目资助

现阶段较主流的深度学习检测方法分为单阶段目标检测器和两阶段目标检测器。其中单阶段目标检测器有 YOLO(you only look once)^[6]、YOLOv2^[7]、YOLOv3^[8]、SSD(single shot multibox detector)^[9]等,二阶段目标检测器有 R-CNN^[10]、Fast R-CNN^[11]等。一阶相比两阶速度更快,能达到实时性的要求,但精确度略低,影响检测结果。随着单阶段算法 YOLO 的不断优化,新算法 YOLOv4^[12]相比其他算法显著提高了精确度,实现了精确度和速度的完美兼顾。

因此,为了快速便捷地检测疲劳驾驶,本文将在 YOLOv4 算法的基础上进行改进,加深了主干特征提取网络 CSPDarknet53 的输出端和特征金字塔 SPP 模块输出端特征层的处理^[13],提高深层特征的提取,并引入了空洞卷积(dilated convolution)^[14],在不降低空间分辨率的情况下扩大了感受野,从而更好检测目标。使用 VOC2007 数据集已经训练好的权重进行预训练,保证初始权值不破坏的情况下加速训练。在用于驾驶员疲劳驾驶检测研究时,首

先收集不同的人疲劳时的神态图片做数据集,使用 labeling 标注数据,分类为面部、眼部和嘴部 3 类,标注文件为 xml 格式,然后将每个对应的图片位置和其真实框位置转为 txt 格式文件存入训练文件中。训练结束后将结果与其他单一改进方式进行比较,对比差异,总结方法。

1 YOLOv4 算法

1.1 算法框架

单阶段目标检测算法 YOLO 系列随着近年来迭代与优化,新的 YOLOv4 算法已成为最强的实施目标检测的模型之一。它的主干部分在沿用 YOLOv3 的基础上加入了跨阶段部分链接(cross-stage-partial-connection)结构,避免学习到重复信息,也保证了学习深度;YOLOv4 框架还加入了空间金字塔池化结构(spatial pyramid pooling, SPP)、路径聚合模块(PANet),头部沿用 YOLOv3 的结构。网络框架如图 1 所示。

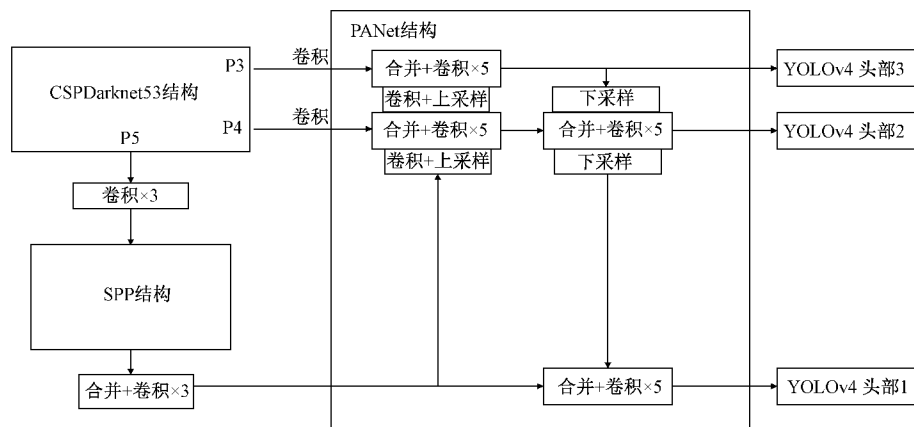


图 1 YOLOv4 网络结构

在原算法中图片首先经过主干卷积部分,将二维卷积、批次标准化运算和 Mish 激活函数作为卷积块定义,然后通过数量不同的滤波器(filter)进行卷积运算。每个卷积块中使用 1×1 卷积对通道数进行整合, 3×3 卷积提取特征,并使用残差结构以保持精度和降低运算。之后对主干最后一个特征层卷积进行 3 次卷积进入特征金字塔 SPP 结构,在 SPP 结构经过卷积核大小为 1×1 、 5×5 、 9×9 、 13×13 的最大池化处理输出,将池化后的结果合并再经过 3 次卷积与主干网 P3、P4 两处传入 PANet 网络。它在下采样结束后返回在进行上采样,并通过横向连接获取同级下采样信息。最后 YOLO 头部用于输出网络预测结果,由于是基于 VOC 数据集训练,因为每个边界框由 6 个数字表示,且该数据有 20 个类别,YOLOv4 特征层上每个点存在 3 个先验框,因此维度为 $3 \times 25 = 75$,输出层的 shape 为 $(13, 13, 75)$ 、 $(26, 26, 75)$ 、 $(52, 52, 75)$ 。

1.2 激活函数

在多层神经网络中,神经元上层输出和下层输入之间由激活函数(activation functions)相连接,它的存在是为了

解决神经网络中线性传递的问题,因此引入非线性函数作为激活函数可使网络不再是线性组合,而可以逼近任意函数。在 YOLOv4 模型中,CSPDarknet53 模块使用了 Mish 激活函数,如式(1)所示。其他模块使用了 Leaky ReLU 激活函数,如式(2)所示。函数图如图 2 所示。

$$\text{Mish} = x * \tanh(\ln(1 + e)) \quad (1)$$

$$f(x) = \max(ax, x) \quad (2)$$

图 2 中可看出 Mish 函数在负半轴不是完全截断,允许较小的梯度流入,保障信息的流动性。Leaky ReLU 函数对比 ReLU 函数在负半轴输入有一定坡度,解决了 ReLU 函数进入负区间神经元无效,梯度为 0 的问题。

2 改进的 YOLOv4 算法

2.1 网络结构

YOLOv4 模型训练的图片使用 480×640 像素图片大小。在改进算法中,虽然使用的图片分辨率较默认值高,但对图像深层特征信息没有帮助,因此需要多次卷积来提取深层的特征信息。故将主干网络输出的卷积层数和在

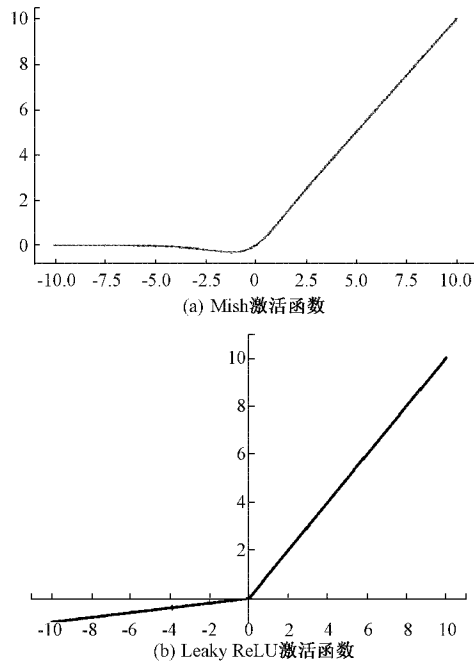


图 2 激活函数

SPP 空间金字塔结构上下两侧的卷积层数由原来的卷积核为 $1 \times 1, 3 \times 3, 1 \times 1$ 三次卷积变为卷积核为 $1 \times 1, 3 \times 3, 1 \times 1, 3 \times 3$ 四次,以加深特征捕获,提高检测效果。同时在 PANet 结构中的 P3、P4 头部输出端前侧和从特征金字塔提取后的 P5 加入了空洞卷积,提高深层感受野,防止丢失图片位置信息,如图 3 所示。

2.2 空洞卷积

空洞卷积(dilated convolution),又叫扩张卷积,在语义分割和目标检测等任务中有着广泛的运用。相比标准卷积来说,空洞卷积是通过在卷积核中填充空洞的方式来增加感受野。在常规卷积时图像通过卷积再进入池化层,降低了图像尺寸,后进行上采样恢复图片原大小进行预测,这样丢失了部内部数据结构,损失掉一些内部信息,对分辨率较低的图片损坏尤为严重。而空洞卷积则将卷积核权重间隔开来,间隔处的权重为 0,不对卷积核做改变,使其可以接受更广的信息,以不损耗图片分辨率为条件增大了感受野,强化了对多尺度上下层信息的捕捉,尤其对例如闭合的眼睛等模糊目标的检测中有明显提高,如图 4 所示。

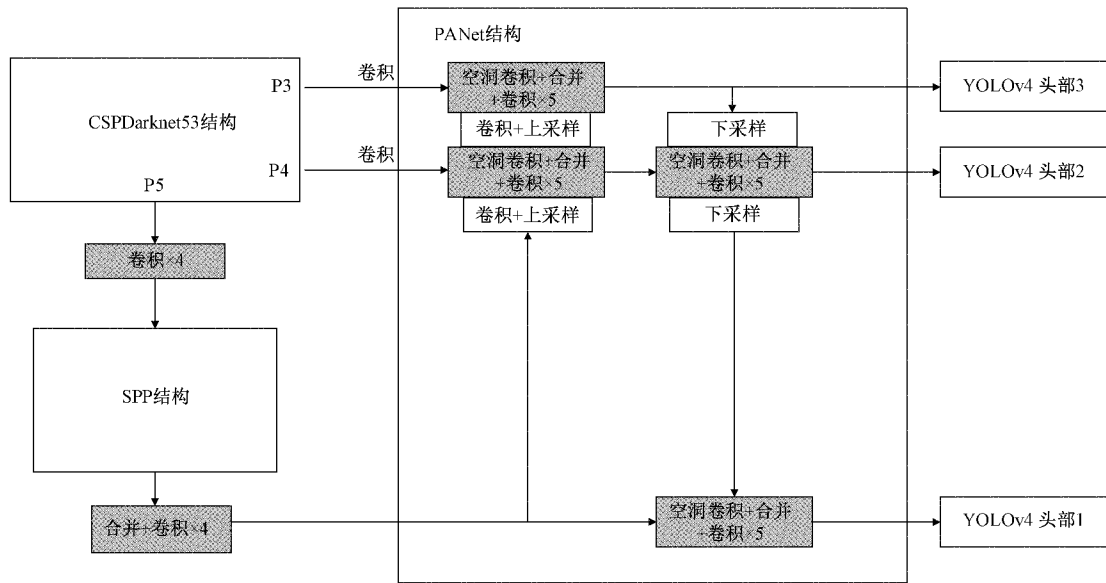


图 3 改进的 YOLOv4 框架

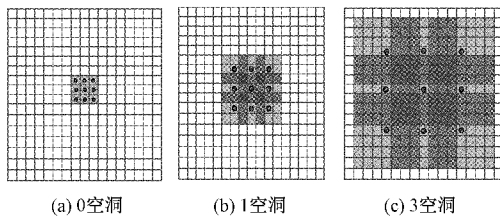


图 4 空洞卷积图解

其中,图 4(a)为正常的 3×3 的卷积核,图 4(b)中卷积核依旧为 3×3 ,但是空洞为 1,所以卷积核的额感受野为 7×7 ,

但是实际权重只有 9 个点,其余权重均为 0。图 4(c)同理是 3×3 卷积核,空洞为 3,遍历图片使用 15×15 的 patch,感受野为 15×15 。因此,空洞卷积的感受野大小 RF 为:

$$RF = k + (k - 1) * (r - 1) \quad (3)$$

式中: $k * k$ 为标准卷积核(kernel size)的大小, r 为扩张率(dilated rate),堆叠不同大小的空洞卷积也可以增加感受野大小:

$$RF = RF_1 + RF_2 \quad (4)$$

在本文中,卷积核使用的空洞数为 1,原 3×3 的卷积核感受野变为 7×7 。

2.3 卷积层

在原始 YOLO 算法中, SPP 结构上下的 3 次卷积核为 1×1 、 3×3 、 1×1 , 滤波器数量为 512、1 024、512, 即先使用 1×1 的卷积核遍历图片, 完成信息整合, 在使用 3×3 的卷积核升维, 加深特征提取, 然后再降低维度还原。但是由于行车中车内环境多变, 驾驶员眼部和嘴部经常因光线等问题无法有效检测, 因此需要较多的滤波器进行特征提取, 因此本文在 3 次卷积后加上滤波器数量为 1 024 的 3×3 卷积, 以加深对眼部和嘴部的特征提取。

3 实验结果与分析

3.1 实验平台

本实验使用的计算机为 Windows10 64 位操作系统, 配置为: 中央处理器: Intel(R) Core(TM) i5-10200H CPU @ 2.40 GHz; 内存 (RAM): 8G; 图像处理器: NVIDIA GeForce RTX2060。

3.2 数据采集

本文的数据集来自网络图片和下线所采集图片, 图片尺寸统一修改为 480×640 , 分类为脸部 (face)、眼睛 (eyes)、嘴巴 (mouth) 共 3 类, 共 310 张图片。其中将图片中人的面部标记为 face, 眯眼闭眼部分标记为 eyes, 打哈欠部分标记为 mouth, 所有图片标签均通过 labelingm 软件自行标注, 生成保存图片路径和标签位置信息的 xml 文件并

用于训练。由于采集图片数量有限, 故使用迁移学习技术, 将 VOC 数据集权重作为初始训练权重, 采取 Mosaic 图片增强技术来丰富数据集, 以训练集: 测试集为 9:1 的比例进行划分, 训练集 279 张, 测试集 31 张。3 种标签类型如图 5 所示。

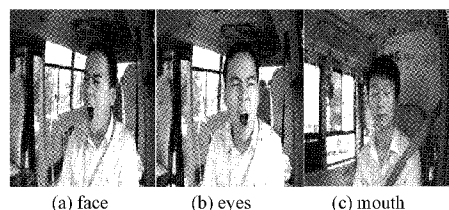


图5 标签类型

1) 迁移学习

假设设计了一个已经训练了一个图像识别的神经网络, x 代表一张图片, y 代表任意物体的标签, 如果想将这个已经学习过的神经网络提取出来并迁移到另一个不同的任务中, 需要做的是取出这个神经网络的最后一层并初始化这一层的相关权重, 然后为最后一层神经网络创建一个新的随机初始化的权重, 即把新数据集 x' 、 y' 设定为放射影像和想要的预测结果, 然后新的数据集上训练这个神经网络, 这就是迁移学习 (transfer learning), 如图 6 所示。

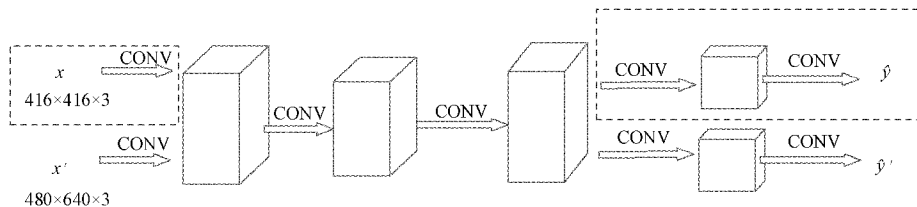


图6 迁移学习框架

VOC 数据集含有 20 个类, 不包括疲劳驾驶类, 因此本文使用其已经训练好的权重作为预训练权重, 再将预训练过的模型作用于疲劳驾驶数据集进行模型微调。具体为先将原训练框架前 200 层进行冻结, 防止初期权重被破坏并加快训练速度, 当训练后期损失函数无明显下降后再开启冻结层数进一步训练, 从而达到训练自己模型的目的。

2) Mosaic 技术

Mosaic 数据增强技术借鉴了 CutMix 数据增强方法, 将 4 张图片拼接成一张图片, 特点是极大丰富了检测数据集, 且随机缩放图片会增添很多较小目标, 提升了网络算法的鲁棒性, 提高模型泛化能力, 同时在批量标准化 (BN) 计算时会出现 4 张图片的数据, 从而更精确地提取图片信息。其具体做法为: 首先从训练数据集中随机抽取 4 张图片, 然后分别进行翻转、裁剪、缩放等步骤, 将他们分别放至 4 角, 再将他们拼接成一张图片。本文使用的部分数据集如图 7、8 所示。

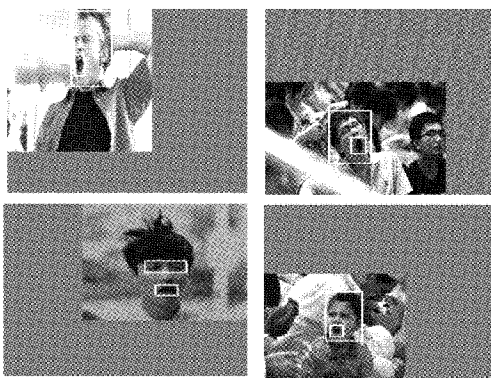


图7 随机选择 4 张图片处理

3.3 疲劳识别

通过卡内基梅隆研究所的实验论证, 得到了度量瞌睡得物理量: PERCLOS (percentage of eyelid closure over the pupil over time)^[15], 即定义为 1~1.5 min 之内眼睛闭合比例达到 70%~80% 时, 就认为发生了瞌睡。公式如下:

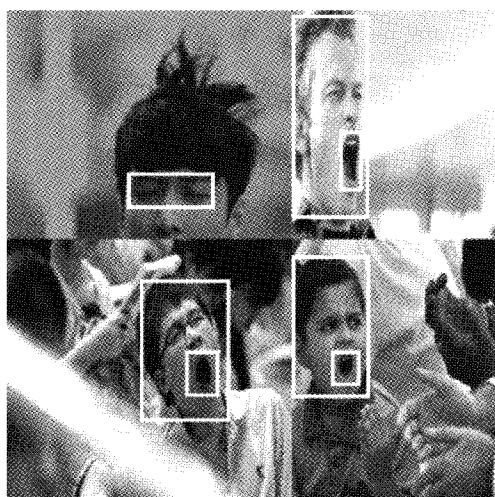


图 8 Mosaic 数据增强

$$PERCLOS = \frac{\text{眼睛闭合帧数}}{\text{检测时间段总帧数}} \times 100\% \quad (5)$$

在疲劳驾驶检测中,由于需要实时检测,及时进行疲劳判定,故不可能将测试时间设置为 1 min 或更长,根据研究可得,人一般眨眼需要 0.3~0.5 s,为去除眨眼对检测的误判,因此本文将检测时间设为 0.6 s,即闭眼时间大于 1 s 时对眼睛进行画框标注,对于判定眼睛开合程度,本文所选数据集均满足瞳孔面积超过 80% 被眼睑遮盖,同时可根据嘴部开合状态判断驾驶员是否疲劳,嘴部判定也设置 0.6 s 延迟,避免将驾驶员唱歌说话等行为进行误判。因此可以准确判定是否为疲劳驾驶。

3.4 数据分析

1) 评判标准

本文使用 AP(average precision)值作为模型的评判标准,AP 即利用不同的准确度(prorecision)和召回率(recall)组合而成的曲线下的面积。运算公式如下:式(6)为准确度计算公式,式(7)为召回率计算公式。

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

式中:TP 表示被分类器认为是正样本并确实为正样本的例子;FP 表示被分类器认为是正样本但实际上是负样本的例子。TP 和 FP 作为分母说明准确度在主观上评判预测的效果好坏。FN 表示被分类器认为是负样本但实际上是正样本的例子。TP 和 FN 作为分母说明召回率在客观程度评判预测效果好坏。

2) YOLOv4 不同改进方法对比

为测试本文在 YOLOv4 改进的算法与框架和其他改进方法的优劣效果,根据控制变量法设定了 3 组对照试验。即实验 1 使用原 YOLOv4 框架与算法,不做任何改变;实验 2 只将特征金字塔 SPP 模块输入输出端的卷积层

数改为 4 层;实验 3 在原 v4 框架上增加空洞卷积,对比两种方式对测试结果的不同影响,实验 4 为综合实验 2 和实验 3 的改进进行训练,即为本文改进方法。图 9 为 4 组实验 3 种标签的 AP 值对比。

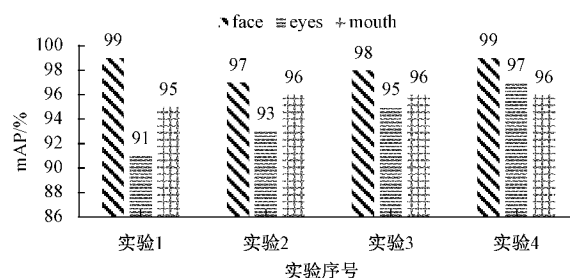


图 9 YOLOv4 不同改进方法 AP 值对比

由图 9 可知,相比实验 1 和实验 2,增加 SPP 前后的卷积层后眼部和嘴部 AP 值有所上升,原因为增加卷积层数后,随着卷积的深入,对眼部和嘴部特征提取加深,但是训练的图片中人脸图像各异,卷积层数加深就导致人脸检测 AP 值有所降低。对比实验 1 和实验 4,增加空洞卷积后其对眼睛和嘴部的检测略优于原始框架,原因为眼睛和嘴巴相较脸部区域小,使用空洞卷积增加感受野可以使卷积核有效提取他们较为整体的特征,但是卷积层数有限,无法更深层次地提取到语义特征信息。实验 4 相比实验 2 和实验 3,综合了两者的优势,随着卷积的深入,使用空洞卷积增大感受野,使深层也能提取较准确的位置信息,兼顾了语义信息和位置信息,因此 AP 值高于其他 3 组实验。

表 1 所示为 4 种实验的 mAP 值对比,在原始框架下增加卷积层后 mAP 值略有下降,降低了 0.13%,原始框架加入空洞卷积后 mAP 值增加 1.23%,将前两者组合后,所得到的 mAP 值相较最初 YOLOv4 算法提高了 1.98%。

表 1 不同改进方法的 mAP 值比较

实验序号	算法模型	平均精密度/%
实验 1	原始算法	95.31
实验 2	增加卷积层数	95.18
实验 3	添加空洞卷积	96.54
实验 4	改进后算法	97.29

4 结 论

本文提出了一种改进 YOLOv4 算法的驾驶员疲劳检测方法。通过增加 SPP 结构前后卷积层数加大疲劳特征提取能力;在 PANet 模块增加空洞卷积加大感受野,使得如眯眼、闭眼等小目标更容易被检测;使用迁移学习通过冻结框架前 200 层来防止初期权值被个别图片破坏,并增加训练速度。试验结果相较原始 YOLOv4 算法,眯眼闭眼检测效果提高显著,可以更准确快速捕捉到驾驶员疲劳驾驶。

同时,在驾驶员戴墨镜或戴口罩时,疲劳特征被遮挡,无法有效检测。因此本文的算法仍有局限。本文后续工作将结合车载传感器或驾驶员生理参数进行辅助检测,增加多种环境下检测鲁棒性。

参考文献

- [1] 王琳,化成城,姜鑫,等.基于颈腰部肌电及脑电信号的疲劳驾驶检测[J].东北大学学报(自然科学版),2018,39(1):102-107.
- [2] 陈志军.基于运动特征的车辆危险行为识别方法研究[D].武汉:武汉理工大学,2016.
- [3] 庄员,戚湧.伪3D卷积神经网络与注意力机制结合的疲劳驾驶检测[J].中国图象图形学报,2021,26(1):143-153.
- [4] TAO X, HOU W, XU D. A review of surface defect detection methods based on deep learning [J/OL]. Acta Automatica Sinica, 2020; 1-19[2020-07-20].
- [5] 戴诗琪,曾智勇.基于深度学习的疲劳驾驶检测算法[J].计算机系统应用,2018,27(7):113-120.
- [6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real time object detection [C]. Computer Vision and Pattern Recognition, 2017; 6517-6525.
- [7] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017; 6517-6525.
- [8] REDMON J, FARHADI A. Yolov3: An incremental improvement [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018, ArXiv: 1804.0276.
- [9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. Proceedings of European Conference on Computer Vision. Springer, 2016; 21-37.
- [10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision & Pattern Recognition, IEEE Computer Society, 2014; 580-587.
- [11] GIRSHICK R. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2015; 1440-1448.
- [12] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2020, ArXiv: 2004.10934v1.
- [13] 李彬,江诚,吴静,等.改进 YOLOv4 算法的航空发动机部件表面缺陷检测[J/OL].激光与光电子学进展, 2021;1-17[2021-04-07].
- [14] 化嫣然,张卓,龙赛,等.基于改进 YOLO 算法的遥感图像目标检测[J].电子测量技术,2020,43(24):87-92.
- [15] 吴康华.基于 PERCLOS 的驾驶疲劳检测系统设计[D].杭州:浙江大学,2008.

作者简介

李昭慧,讲师,硕士生导师,主要研究方向为光衍射特性计算与反演、图像处理。

E-mail: lizhaohui080821@126.com

张玮良,硕士研究生,主要研究方向为图像处理。

E-mail: 269066979@qq.com