

基于通道注意力 YOLOV5s 的驾驶行为识别研究

罗国荣, 戚金凤

(广州科技职业技术大学 自动化工程学院, 广州 510550)

摘要: 设计了一种集成通道注意力机制的 YOLOV5s 检测网络的驾驶行为识别方法, 用以实时检测并识别驾驶员在驾驶室內的驾驶行为, 从而有利于纠正驾驶员的不良驾驶行为, 减少交通事故发生的概率; 建立了驾驶室内驾驶员手部动作的图像数据集; 在 YOLOV5s 网络结构中引入通道注意力机制, 通过对比实验、消融实验研究了通道注意力模块嵌入 YOLOV5s 中的较佳作用位置、配置数量的影响及其检测识别性能效果; 论证了带通道注意力的改进 YOLOV5s 可保留信息量大的特征、抑制不相关的特征, 模型参数量和复杂度降低, 从而加快检测速度; 测试结果显示, 较原 YOLOV5s 网络, 改进的 YOLOV5s 在平均精确度和召回率上相当, 而检测速度提升了 26.08%, 该方法能够较好地满足驾驶员手部动作的实时监控需求。

关键词: 行为识别; 目标检测; 驾驶行为; YOLOV5s; 通道注意力机制

Study on Identification of Driving Behavior Based on YOLOV5s with Channel Attention

LUO Guorong, QI Jinfeng

(School of Automation Engineering, Guangzhou Vocational University of Science and Technology, Guangzhou 510550, China)

Abstract: A driving behavior identification method of YOLOV5s detection network with integrated channel attention mechanism is designed to detect and identify driver's driving behavior in cab in real time, so as to correct driver's bad driving behavior and reduce the probability of traffic accidents. The image data set of the hand movements of drivers in cab is established, the channel attention mechanism is introduced into the YOLOV5s network structure, the comparative experiments and ablation experiments are used to study the effects of optical location and number of configurations for embedding channel attention modules into YOLOV5s network structure, as well as detection and recognition performance, prove the features of large information for the improved YOLOV5s with channel attention and suppress irrelevant features, and reduce the number and complexity of model parameters to accelerate the detection speed. The test results show that compared with original YOLOV5s network, the improved YOLOV5s is comparable in the average accuracy and recall rate, while the detection speed is increased by 26.08%. This method can better meet the real-time monitoring requirements of drivers' hand movements.

Keywords: behavior recognition; target detection; driving behavior; YOLOV5s; channel attention mechanism

0 引言

随着我国经济的快速发展, 以及汽车电动化、智能化的技术革新, 人们对汽车的购买需求显著增加, 我国的汽车保有量也随之增加, 但伴随的交通事故却没有明显减少, 究其原因, 发现驾驶人员不良驾驶行为引起的交通事故占 95% 以上^[1]。因此, 为了有效减少交通事故的发生, 研究驾驶行为, 辅助规范驾驶行为, 对减少由不良驾驶行为引起的交通事故, 具有重要的意义。

文献 [2] 为挖掘汽车变道这一重要因素, 研究汽车的速度和加速度, 提取 7 个相关的驾驶行为特征参数, 利用 K-means 聚类算法, 对驾驶行为进行分类评价, 取得较好的效果。文献 [3] 通过汽车 OBD 诊断仪采集汽车车速、

发动机转速等行车数据, 综合利用聚类法、主成分和因子法进行驾驶行为分析, 得出了危险型、一般型和谨慎型三种驾驶行为。文献 [4] 首先利用车联网技术采集 30 辆车的行驶数据, 对这些数据进行清洗、分段、筛选, 利用熵权一主成分分析法建立行车安全评价模型, 分析其行驶安全状况, 最终得出行驶强度指标。文献 [5] 陶红兴搭建系统硬件和软件完成车载 OBD 信息以及 ADAS 数据的信息采集, 分析车辆的状态和行驶轨迹特征, 融合机器学习算法实现了三种驾驶行为辨识系统。文献 [6] 黄俊提出了一种信息融合方法, 该方法采集了驾驶员脸部特征信息和车辆行驶信息, 通过计算车道中心线与图像中心点间距离的变化率来判断车辆行驶状态, 并与驾驶员脸部特征信息进行

收稿日期: 2023-08-05; 修回日期: 2023-08-25。

基金项目: 2022 年度广东省普通高校特色创新类项目(2022KTSCX192)。

作者简介: 罗国荣(1981-), 男, 硕士研究生, 讲师。

戚金凤(1985-), 女, 硕士, 副教授。

引用格式: 罗国荣, 戚金凤. 基于通道注意力 YOLOV5s 的驾驶行为识别研究[J]. 计算机测量与控制, 2023, 31(10): 273-278, 284.

信息融合, 实验结果表明, 该方法可以较为准确地检测出驾驶员当前的行为状态。文献 [7] 为了分析不同人格对驾驶员驾驶行为的影响, 从驾驶员的“生理和心理”维度采集相关数据, 并利用 K 均值聚类算法对这些数据进行统计聚类, 得到五种人格的关键指标阈值, 并在此基础上建立驾驶能力评估模型, 该模型能够有效量化评估驾驶人在不同情境下的驾驶行为。文献 [8] 为了识别驾驶员的风险驾驶行为, 从驾驶经验与技能、驾驶员性格和驾驶态度、能力暂时性缺失、社会心理因素等影响因素进行分析, 获得了良好的效果。文献 [9] 提出了一个集成平台, 该平台收集、存储、处理、分析车辆的不同数据流, 应用深度学习算法与聚类技术对来自不同车辆的数据进行分析和处理, 取得一定的效果。文献 [10] 从驾驶员—车辆—环境 (DVE) 系统中获取大量的汽车行驶数据, 通过机器学习 (ML) 模型处理, 建立了一个影响驾驶员行为的多个维度的解释框架, 具有一定的现实意义。文献 [11] 提出了一种车辆加速度预测模型, 该模型提取车辆的相对距离、相对速度和加速度作为特征变量来描述驾驶行为, 通过机器学习方法来分析、预测驾驶行为, 结果表明, 该模型能预测驾驶员的驾驶行为。文献 [12] 应用汽车跟踪场景, 将脑电图的心理特征与行为反应联系起来。构建一个伪影成分池的驾驶行为模型, 为驾驶员的驾驶行为分析提供了一种思路。文献 [13] 研发一个基于差分全球导航卫星系统 (DGNSS) 模块的车载终端平台, 利用车辆实时跟踪技术, 进行驾驶员行为分析, 结果表明, 该平台能够自动、准确地提取出驾驶行为特征。文献 [14] 提出了一种基于深度学习的时间序列建模方法的能量感知驱动模式分析系统。对能量感知的纵向加减速行为和横向变道行为进行了统计分析, 对小型汽车驾驶行为的个性化评估具有一定参考意义。文献 [15] 为了减少渣土车驾驶员行车安全事故, 通过采集汽车的行车数据和北斗定位数据, 利用 K-means 聚类算法分析、识别驾驶倾向, 并建立模型评估渣土车驾驶员的驾驶行为, 取得一定效果。文献 [16] 为实时监测驾驶员的驾驶行为, 设计出一种智能安全驾驶监测系统, 主要是通过检测驾驶员酒精浓度, 从而成功识别酒精驾驶的危險驾驶行为。文献 [17] 从 OBD 模拟器上获取汽车的行车数据, 采用模糊综合评价法分别从行车里程、超速、急加速、急转弯等方面构建驾驶评价指标, 取得不错的效果。文献 [18] 提出一种驾驶行为评估模型, 该模型通过模拟驾驶实验获取相关行车数据, 利用信息熵和随机森林算法分析归纳驾驶行为, 实验结果表明, 该模型的驾驶行为风险总体辨识精度达到 80%。文献 [19] 为优化服务区入口匝道减速设施, 通过模拟驾驶实验, 采集汽车车速、加速度和刹车踏板开度等数据, 分析减速设施对驾驶员驾驶行为的影响, 成功预测了驾驶员驾驶行为的变化规律。文献 [20] 为了研究标线亮度对驾驶员驾驶行为的影响, 通过模型驾驶实验, 采集车速、横向位移和方向盘转角等数据, 并据此拟合恰当的数学模型, 实验表明, 该数学模型能在

不同的夜间标线亮度预测驾驶员的驾驶行为。

上述学者都是通过车载自诊断系统 (OBD) 或车联网技术获取车载传感器的相关数据, 或是通过卫星定位系统 (GPS) 获取汽车的位置及速度数据, 通过统计、分析这些数据, 间接预测驾驶员的驾驶行为, 具有一定的研究价值, 由于驾驶行为的预测是通过分析汽车行驶数据间接得出, 存在时间滞后的缺点, 因此本文通过研究驾驶员驾驶动作的方式, 直接采集驾驶室内驾驶员玩手机、喝水等不良驾驶行为的视频数据, 并将采集到的视频数据转换为图像数据集, 利用深度学习算法研究驾驶员的日常驾驶行为, 挖掘日常驾驶图像数据之间的内在联系, 形成更有价值的信息, 以帮助驾驶员纠正不良的驾驶行为, 从而减少交通事故的发生。

1 数据集

实验数据采集选择普通行车记录仪作为数据采集设备, 从副驾驶室向驾驶室的角度拍驾驶员的室内驾驶视频, 再进行取帧保存, 选取 974 张较为清晰图片, 针对驾驶员玩手机、喝水两种不良的驾驶行为进行了数据标注, 并对图片应用旋转、平移、缩放、添加噪声、裁剪等方式进行 Mosaic 扩充。扩充后数据为 1 374 张, 如图 1 所示。



图 1 数据图像增强效果

2 YOLOV5s 网络结构及特点

YOLOV5s 网络结构如图 2 所示, 整个网络结构可分为输入网络 (input)、骨干网络 (backbone)、特征整合网络 (neck) 和预测网络 (output) 四部分。

输入网络进行图像数据增强、锚框的自适应计算、图片的自适应缩放三个操作: 1) 图像数据增强通过对每一幅图像数据进行旋转、缩放、翻转、色彩变换等来扩大数据集。此外, 还延用了 YOLOV4 的 Mosaic 图像数据增强, 其显著的优点是丰富了检测物体的背景, 同时在标准 BN 计算时, 经过 Mosaic 图像数据增强的图像相当于 4 张图像的数据量, 有效加快了训练速度; 2) 锚框的自适应计算是 YOLOV5s 网络模型特有的操作, 有一组应用于 COCO 数据集的预设初始锚框, 尺寸分别为 (10, 13), (16, 30), (33, 23), (30, 61), (62, 45), (59, 119), (116, 90), (156, 198), (373, 326); 前三个是针对大特征图的锚框, 用于检测尽可能多的小目标信息, 后三个是针对小特征图的锚框, 用于检测大目标信息; 3) 图像的自适应缩放采用 letterbox 自适应图像缩放技术, 该技术能够在图像缩放填

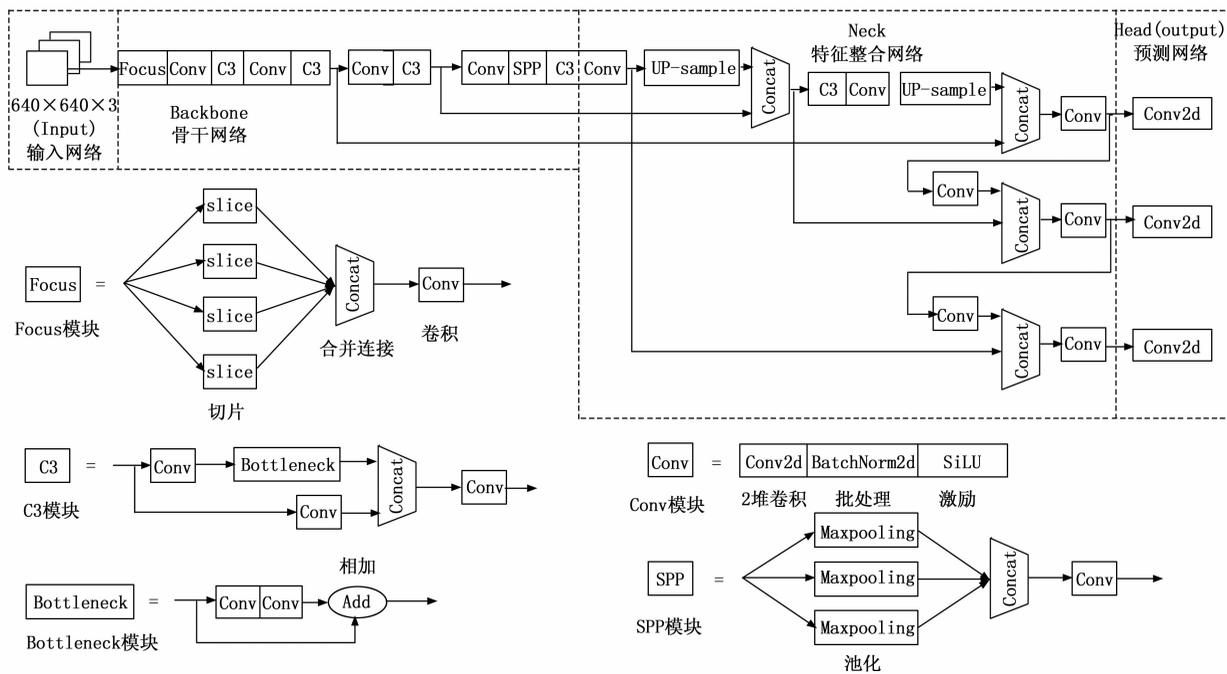


图 2 YOLOV5s 网络结构

充后, 有效减少填充黑边的信息冗余量, 在一定程度上提高算法的推理速度。

骨干网络 (backbone) 是由 1 个 Focus 网络、4 个 CBL 网络、2 个 CSPNET 网络和 1 个 SPP 网络组成^[17-18]。Focus 网络的主要作用是减少网络深度、模型参数以及模型计算量 (FLOPs), 在保证不影响 mAP 性能的情况下加快网络前向传播和反向传播的速度。CSPNET 网络又称跨阶段局部网络, 其作用有三点: 1) 解决因网络在优化中梯度信息重复引起推理计算量大的问题, 在降低计算量的同时保证了准确率; 2) 由于该网络将计算量均匀地分配到每一层, 这样可以将因网络瓶颈结构造成的部分闲置计算单元充分地利用起来, 从而提升了每一个计算单元的利用率, 降低了计算瓶颈; 3) 在特征金字塔生成过程中采用跨通道池 (cross-channel pooling) 来压缩特征图, 从而降低内存的使用率。SPP (spatial pyramid pooling) 层又称空间金字塔池化层, 它能将任意尺寸大小的特征图转换成固定大小的特征向量, 该层在完成这个转换任务的同时, 不仅有效避免了图像缩放带来的图像失真问题, 还解决了卷积神经网络重复提取特征的问题, 提高了网络的运行速度。

特征整合网络是由特征金字塔网络 (FPN) 与金字塔注意力网络 (PAN) 相结合组成的网络。其目的是将深层网络表征强语义的特性和浅层网络表征强定位的特性充分地利用起来, 更好地表达图像中的目标和位置尺寸, 为一阶段的网络预测提供坚实的基础。其网络结构如图 3 所示。

预测网络在输出层直接对目标框进行回归操作, 以确定图像中目标的类别和位置^[19]。具体操作如下: 1) 通过骨干网络 (backbone) 和特征整合网络 (neck) 将一幅输入图像分成 $S \times K$ 个网格, 而预测网络则需要对每个网格预测,

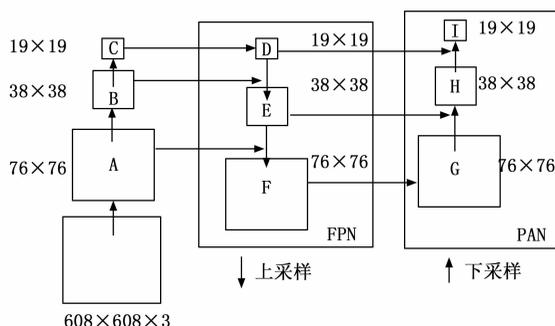


图 3 Neck 网络结构

若图像中的某个目标落在某个网格, 该网格就负责预测这个目标。因此, 在 $S \times K$ 个网格中, 每个网格负责预测 3 个锚框 (anchor) 和 1 个类别, 而每个锚框负责该框的 4 个位置信息 (x, y, w, h) 和 1 个置信度信息 (confidence), 这样一幅图像就预测出 $S \times K \times 3$ 个锚框; 2) 筛选锚框, 去除冗余的锚框, 保留包含目标的锚框。首先计算锚框的类别置信度分值, 通过设定阈值去除可能性低的锚框。

3 通道注意力机制的引入

注意力机制类似于人类对外界事物的观察机制, 人类观察事物时, 首先会观察事物中较为吸引人的某些重要局部区域, 对这些局部区域投入更多的注意力, 以获得更多的细节信息, 然后再将其它局部区域联合起来组成一个整体观感。

空间注意力和通道注意力是注意力机制的两种类型。通道注意力利用卷积特征通道之间的相互关系, 从通道方面使网络校准有用的特征响应, 从而抑制不相关的特征, 有选择

地强调信息量大的特征。本文主要创新点是将通道注意力机制集成进 YOLOV5s 中，即在网络中引入 ECABL 模块，如图 4 所示。可见，ECABL 模块可分为卷积、挤压 (squeeze)、激励 (excitation) 和尺寸变换 (scale) 四个操作。

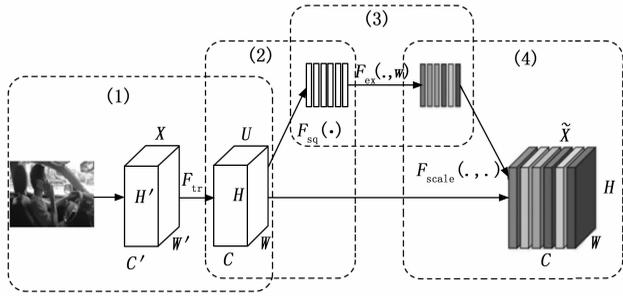


图 4 ECABL 模块网络结构

1) 卷积操作。假设卷积结果 $U = [u_1, u_2, u_3, \dots, u_c]$ ，卷积核 $V = [v_1, v_2, v_3, \dots, v_c]$ ，输入 $X = [x_1, x_2, x_3, \dots, x_c]$ ，卷积操作符号为 $*$ ，则卷积操作 F_{tr} 可表示为：

$$u_c = F_{tr}(x) = v_c * X = \sum_{s=1}^C v_c^s * x^s \quad (1)$$

式中， u_c 为第 C 通道的卷积结果； v_c 为第 C 通道的卷积核。

2) 挤压操作。其公式为：

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (2)$$

式中， F_{sq} 为挤压操作； H, W 分别为特征图的高和宽； $u_c(i, j)$ 为第 C 个通道的第 (i, j) 个元素； z_c 为采用全局平均池化 (global pooling) 将高宽为 (H, W) 的第 C 个通道的特征图挤压成一个通道权重。

3) 激励操作。其公式为：

$$s = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \delta(W_1 z)) \quad (3)$$

式中， F_{ex} 为激励操作； $\sigma(\cdot)$ 为 sigmoid 激活函数；共进行了两次的全连接与 sigmoid 激活操作。

4) 尺寸变换操作。其公式为：

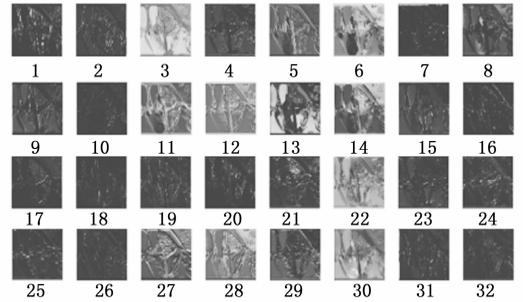
$$\tilde{x} = F_{scale}(u_c, s_c) = s_c u_c \quad (4)$$

式中， F_{scale} 为尺寸变换的操作； s_c 为激励操作的结果， u_c 为卷积操作的结果。可以看出，尺寸变换就是将上述两种操作结果进行通道相乘操作。

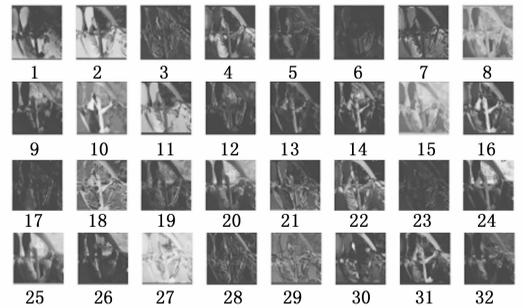
综上，通道注意力模块首先进行卷积操作，得到相应通道的特征图，每一个通道的特征图都不一样；再通过挤压操作，得到每一个通道特征图的通道权重；接着通过激励操作过滤通道权重较低特征图，保留通道权重较高特征图，抑制与目标位置形状不相关的特征；最后进行特征图的尺寸变换，以适应后续神经网络的输入要求。

为了对比普通卷积与引入通道注意力的效果，将一张图片分别进行卷积操作和引入通道注意力，比较两者的特征图，图 5 (a) 为卷积操作后的 32 通道的特征图，(a) 中序号为 3、5、6、11、14、22、27、28、30 等 9 张特征图计算得到通道权重较高，强调目标的位置和形状信息，而图中其它序号的特征图计算得到通道权重较低，位置和形状

信息不明显；图 5 (b) 为引入通道注意力机制后的 32 通道特征图，(b) 只有序号为 5、6、17、23 等 4 张特征图显示的位置和形状信息不明显，其它大部分特征图都突出了目标的形状和位置。可见，在浅层网络中引入通道注意力机制后，保留了与目标位置形状相关度大的特征图，同时删减了部分通道权重低的特征图，从而减少了计算量，加快了训练和检测速度。同理，在深层神经网络中，ECABL 网络模块通过以上 4 个步骤的操作，也能将强调语义的通道特征图保留下来，删减与语义不相关的通道特征图。



(a) 在浅层网络中经过卷积操作的 32 通道特征图



(b) 引入通道注意力模块得到的 32 通道特征图

图 5 32 个通道特征图

4 实验与结果分析

实验环境使用 win10 操作系统，软件环境平台为 anaconda+pytorch。在硬件配置上 CPU 使用英特尔酷睿 i9-7900X；GPU 为英伟达 GTX 1080Ti 11G 显存。

4.1 评价标准

模型的评价指标很多，包括精准度 (precision)、召回率 (recall)、mAP (mean average precision) 等。识别结果包括真阳性 (TP, true positive)、真阴性 (TN, true negative)、假阳性 FP (false positive)、假阴性 (FN, false negative) 四种情况。精准度 (precision) 是针对最后预测的结果，即指一个分类器预测出来的正类占所有真正正负类的比率。其计算公式为：

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (5)$$

召回率 (recall) 是衡量一个分类器能否找出所有的正类能力，即是在所有正类的样本中，分类器能预测出多少正类样本。其计算公式为：

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

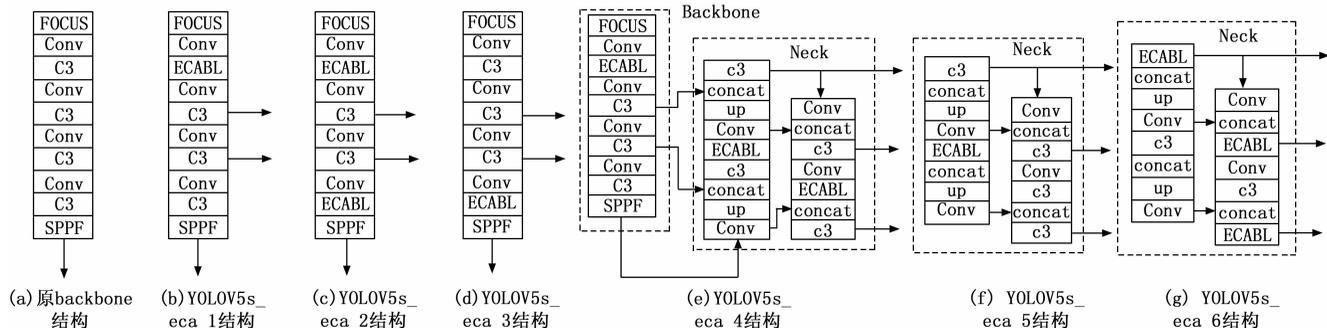


图 6 ECABL 嵌入 YOLOV5s 不同位置的各种结构

对于多标签样本的分类还可以用 mAP 值来衡量检测网络的性能, mAP 值也是综合精准度和召回率这两个指标的一个评估值。其计算公式为:

$$mAP = \frac{\sum_{k=1}^N P(k) \Delta R(k)}{C} \quad (7)$$

式中, N 为测集中的样本个数; $P(k)$ 为同时识别 k 个样本时精准率的大小; $\Delta R(k)$ 为检测样本个数从 $k-1$ 个变为 k 个时召回率的变化情况; C 为多分类检测任务类别的个数。

4.2 对比实验

在卷积神经网络中嵌入通道注意力机制, 可以使得网络自适应地保留重要的信息和忽略一些无关信息, 从而提升网络总体表现。本文在 YOLOV5s 中的不同位置去掉某一 C3 网络, 再嵌入数量不等的通道注意力模块。如图 6 所示, 分别在 YOLOV5s 的 backbone 和 neck 网络中的不同位置嵌入数量不等的通道注意力 (ECABL) 模块, 实验结果如表 1 所列。观察表 1 可知:

表 1 带 ECABL 模块的 YOLOV5s 检测识别性能

模型	Precision /%	Recall /%	mAP50 /%	Parameter / 10^6	GFLOPs
原 YOLOV5s	97.1	97.5	99.1	7.056 607	16.3
YOLOV5s_eca_1	92.5	99.9	99.4	6.932 583	15.3
YOLOV5s_eca_2	96.0	96.9	98.7	5.472 760	12.9
YOLOV5s_eca_3	98.1	97.6	99.4	5.834 340	14.9
YOLOV5s_eca_4	94.7	96.9	99.0	6.668 460	13.8
YOLOV5s_eca_5	97.1	99.9	99.5	6.687 076	14.8
YOLOV5s_eca_6	96.6	95.0	99.3	5.598 318	13.3

1) 在 YOLOV5s 网络中嵌入 2 个或 3 个 ECABL 模块, 即 YOLOV5s_eca_2、YOLOV5s_eca_4 和 YOLOV5s_eca_6 结构, 虽然模型参数量和复杂度有所减少, 但 $mAP50$ 、精度率和召回率都有所下降, 说明嵌入的模块数量并不是越多越好。这可能是因为, 频繁地强调通道注意力, 导致过多地抑制了一些关联小的通道特征, 使得检测识别性能略有下降。

2) 在 YOLOV5s 的 backbone 网络中嵌入 1 个 ECABL 模块, 即 YOLOV5s_eca_1 和 YOLOV5s_eca_3 结构,

$mAP50$ 均为 99.4%, 相比原 YOLOV5s 网络提升了 0.3 个百分点; 召回率分别为 99.9% 和 97.6%, 较 YOLOV5s 网络的 97.5%, 略高于原 YOLOV5s 网络; 而模型参数量分别减小了 124.024×103 和 $1\ 222.267 \times 103$, 模型复杂度 (GFLOPs) 分别减少了 1 和 1.4 (约 6.3% 和 8.5%)。在 YOLOV5s 的 neck 网络中嵌入 1 个 ECABL 模块, 即 YOLOV5s_eca_5 结构, 也能获得较好的效果, $mAP50$ 也为 99.5%, 召回率优于原 YOLOV5s 网络, 精度率相当, 模型参数量减少 369.529×103 , 模型复杂度 (GFLOPs) 减少了 1.5 (约 9.2%)。这说明通道注意力机制起了作用, 恰当地抑制了不相关的通道特征, 强调了有用的通道特征, 从而减少了模型参数量和模型复杂度, 在保证检测识别性能的同时提升了检测速度。

3) 可以看出, YOLOV5s_eca_5 结构性能最优, YOLOV5s_eca_3 结构次之; 从位置上看, 两者都是嵌入到 YOLOV5s 网络的中部, 对网络的影响更好。究其原因, 在网络的中部, 已提取了一定数量代表位置轮廓的浅层特征和代表语义的深层特征, 此时利用通道注意力机制删除一些不相关的通道特征, 可减少模型的参数量, 同时也能保证模型原来的性能, 如果嵌入的位置太前或太后, 则通道注意力机制发挥的作用就不明显。

4.3 消融实验

为了验证所嵌入的 ECABL 模块对 YOLOV5s 模型的改进是否有效, 进行消融实验。根据表 1 采用综合性能最好 YOLOV5s_eca_5 结构, 分三种方案进行: 1) 没有嵌入 ECABL 模块, 保留原 YOLOV5s 的结构; 2) 在 Neck 网络中的第一个 C3 网络处嵌入 ECABL 模块, 称之为 YOLO_Neck_add; 3) 去掉 Neck 网络中的第一个 C3 网络, 再在此位置嵌入 ECABL 模块, 即用 ECABL 模块代替 C3 网络, 称为 YOLO_Neck_replace, 得到的结果如表 2 所列。

表 2 消融实验结果

模型	Precision /%	Recall /%	mAP50 /%	Parameter / 10^6	GFLOPs
原 YOLOV5s	97.1	97.5	99.1	7.056 607	16.3
YOLO_Neck_add	97.5	98.5	99.4	7.053 668	16.0
YOLO_Neck_replace	97.1	99.9	99.5	6.687 076	14.8

从表 2 中可以看出, 方案 2) 直接在原 YOLOV5s 网络中嵌入 ECABL 模块, 检测识别性能有所提升, 参数数量和模型复杂度也有所下降。方案 3) 得出的结果, mAP 性能比原网络提升了 0.4 个百分点, 召回率提升了 2.4 个百分点, 精确率不变, 并且参数数量和模型复杂度有效减小, 说明嵌入的通道注意力网络起了一定作用, 所提出的改进 YOLOV5s 结构有效可行。

图 7 为 YOLOV5s_eca_5 结构和原 YOLOV5s 网络结构在训练过程中损失曲线的变化情况。改进的 YOLOV5s_eca_5 结构在原 YOLOV5s 网络结构的左下方, 这说明改进的网络结构收敛速度更快; 并且在第 180 个至第 195 个 epochs 时, 两者的损失率曲线重叠, 说明两者在性能上相当, 随后 YOLOV5s_eca_5 结构损失曲线一直处于原网络的下方, 这进一步表明改进的网络结构有效。

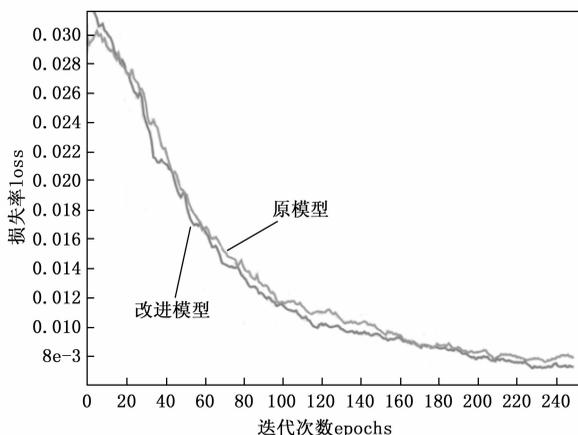


图 7 损失率变化曲线

4.4 检测结果

为了进一步展示改进的 YOLOV5s_eca_5 结构网络的检测效果, 在测试集中随机抽取 30 张图像进行检测实验, 分三个测试组: 1) 第一组为 10 张仅有玩手机动作的图像; 2) 第二组为 10 张仅喝水动作的图像; 3) 第三组为同时含有玩手机和喝水动作的图像。其中 2 张图像实验结果如图 8 所示, 图 8 (a) 为 YOLOV5s 检测结果, 图 8 (b) 为改进的 YOLOV5s 检测结果。在检测第一张图像时, 两者都能检测到玩手机动作, 其中改进 YOLOV5s 网络的置信度比原 YOLOV5s 网络高 0.5 个百分点; 检测第二张图像时, 改进 YOLOV5s 网络的置信度比原 YOLOV5s 网络高 0.2 个百分点。这说明改进的 YOLOV5s 网络与原 YOLOV5s 网络的检测性能相差不多, 也就是改进的 YOLOV5s 网络继承了原网络的检测性能。

在检测速度方面, 在预处理、推理和极大值抑制 (NMS) 三个步骤上统计计算耗时, 三组图像的检测结果如表 3 所示。可以看出: 1) 主要耗时体现在推理步骤上; 2) 改进的 YOLOV5s 网络模型的检测总耗时比原 YOLOV5s 网络快 10.69 ms; 单张图片的平均耗时分别为 1.38 ms 和 1.02 ms, 即检测速度提升了约 $(1.38 - 1.02) / 1.41 = 26.08\%$ 。



(a) 原 YOLOV5s 检测结果



(b) 改进的 YOLOV5s 网络检测结果

图 8 改进的 YOLOV5s 与原 YOLOV5s 网络检测结果对比

表 3 改进的 YOLOV5s 网络与原 YOLOV5s 网络检测时间结果对比

模型	组别	预处理时间/ms	推理/ms	NMS/ms	耗时/ms	总耗时/ms	平均耗时/ms
原 YOLOV5s	①玩手机	0.20	11.80	2.00	14.00	41.4	1.38
	②喝水	0.40	13.7	0.70	14.8		
	③玩手机喝水	0.50	11.4	0.70	12.60		
改进的 YOLOV5s	①玩手机	0.01	9.90	1.10	11.01	30.71	1.02
	②喝水	0.30	8.60	0.60	9.5		
	③玩手机喝水	0.60	8.80	0.80	10.20		

5 结束语

1) 改进的 YOLOV5s 方法是通过将不同数量的通道注意力 ECABL 模块嵌入到 YOLOV5s 原网络中的不同位置 (即 backbone 或 neck) 来实现的。对比实验、消融实验研究结果表明: ①ECABL 模块嵌入到 YOLOV5s 网络的中部效果更好, 且配置数量并非越多越好; ②改进的 YOLOV5s 可保留信息量大的特征、抑制不相关的特征, 模型参数数量和复杂度均有所降低因此检测速度更快。

2) 检测结果显示, 较原 YOLOV5s 网络, 改进的 YOLOV5s 在目标检测识别性能上相当, 而检测速度提升了 26.08%, 能够更好地满足驾驶员手部动作的实时监控需求。

参考文献:

[1] 郑恒杰, 熊 昕, 张 上, 等. 基于车联网数据挖掘的驾驶员行为分析 [J]. 信息通信, 2019 (8): 57-60.
 [2] 李储岩. 基于 K-means 算法的驾驶行为分析研究 [D]. 哈尔滨: 哈尔滨工程大学, 2020.
 [3] 夏 兵. 基于 OBD 信息的高速公路驾驶特征研究 [D]. 重庆: 重庆理工大学, 2018.
 [4] 张晓博, 陈 熙. 基于车联网的营运车辆驾驶行为分析 [J]. 企业科技与发展, 2020, 463 (5): 61-63.

(下转第 284 页)