

文章编号: 2095-2163(2023)09-0153-05

中图分类号: TP391

文献标志码: A

基于改进 Faster-R-CNN 塔式起重机驾驶人员行为监测研究

李亚伟¹, 陈文铿¹, 林鸿强¹, 张旭生¹, 陈子健¹, 林进浔², 陈国栋¹

(1 福州大学 物理与信息工程学院, 福州 350108; 2 福建数博讯信息科技有限公司, 福州 350002)

摘要: 考虑到塔吊驾驶环境的特殊性,为减少塔吊驾驶人员不规范的驾驶行为,降低塔吊事故的发生率,本文结合手部检测的塔吊驾驶人员行为规范监测方法,改进 Faster R-CNN 算法模型,融合了剪枝、通道注意力机制等算法,提出了 CF-R-CNN 模型。根据手部和被检测物体的预测框交并比阈值,判断驾驶人员是否存在违规行为。改进后,网络的 $F1$ 值相比原网络只降低了 2.1%,但 FPS 提高了 23.0%,并与 FRC-Tiny 和 Cut-YOLOv3 算法进行了对比。实验结果证明,该网络在性能上有一定的提升,达到了实时性检测的要求,可在移动端进行部署。

关键词: 驾驶人员检测; 深度学习; 通道注意力; Faster R-CNN; 目标检测

Research on behavior monitoring of tower crane drivers based on Faster-R-CNN

LI Yawei¹, CHEN Wenkeng¹, LIN Hongqiang¹, ZHANG Xusheng¹, CHEN Zijian¹, LIN Jinxun², CHEN Guodong¹

(1 Department of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China;

2 Fuzhou Shuboxun Software and Development Co., Ltd, Fuzhou 350002, China)

【Abstract】 The prevention of tower crane accidents in the construction industry has always been a difficult problem. Considering the particularity of tower crane driving environment, reducing the non-standard driving behavior of tower crane drivers and the incidence of tower crane accidents, this paper puts forward the behavior monitoring method of tower crane drivers combined with hand detection, improves the Faster R-CNN algorithm model, integrates the pruning, channel attention mechanism, and other algorithms, and puts forward the CF-R-CNN model. The $F1$ value of the improved network is only 2.1% lower than that of the original network, but the FPS is increased by 23.0%. Compared with FRC-Tiny and Cut-YOLOv3 algorithms, the experimental results show that the network has a certain improvement in performance, meets the requirements of real-time detection, and can be deployed at the mobile terminal.

【Key words】 driver detection; deep learning; channel attention mechanism; Faster R-CNN; target detection

0 引言

塔式起重机(又名“塔吊”),是建筑工地上常用的起重设备,纵观国内塔吊租赁行业,每年塔吊安全事故都有发生。据有关部门统计,目前国内塔吊重大安全事故率约为 0.3%,多数情况都是由驾驶人员的违规操作造成的,且一旦发生事故,驾驶人员的生还率极低。

目前,对于驾驶人员的行为规范监测研究中,基

于面部识别的驾驶员疲劳危险驾驶检测系统^[1-2]针对眨眼和打哈欠时的人脸特征点进行建模、提取和训练,通过疲劳阈值来判断是否疲劳驾驶,但疲劳阈值的设定存在因人而异,泛化能力不够的问题;基于双向长短记忆网络^[3-4]的异常驾驶行为的检测方法,通过构建前向 LSTM(Long Short-Term Memory)和后向 LSTM 双向网络来记忆过去时刻的车辆运动状态(横纵向加速度和转向角等),再将结果输入到全连接层中回归分类异常驾驶行为(急刹车和急拐

基金项目: 福建省科技计划项目(2021H0013); 福建省科技型中小企业创新资金项目(2021C0019); 福州市科技计划项目(2022-XG-001)。

作者简介: 李亚伟(1999-),男,硕士研究生,主要研究方向:计算机图形图像处理;陈国栋(1979-),男,博士,副研究员,主要研究方向:计算机图形图像处理、深度学习、计算机仿真技术。

通讯作者: 陈国栋 Email: cgd@fzu.edu.cn

收稿日期: 2022-09-11

弯等),但此方法并不适用于塔吊驾驶人员检测;多任务级联卷积神经网络 MTCNN^[5] 结合人脸识别方法,同时对手机和烟头进行检测,通过级联结构达到较高的人脸检测对齐速度,构建其关联性,来判断驾驶人员是否存在违规行为,但若驾驶人员使用手机时离脸部较远会出现漏判情况;双线性注意力网络^[6] 在 RNN 的基础上引入了长短时记忆单元,拥有了记忆功能,解决了 RNN 的梯度消失问题,构建了多重损失网络的结构,能对多种分心行为进行分类,但每个 LSTM 的神经元都连接多个全连接层,对设备的算力、内存要求高,而且耗时,较难在移动端实施实时性部署;彭豪等人^[7] 提出基于改进的 Faster R-CNN 的 FRC-Tiny 目标检测模型,并与郑福建等人^[8] 提出基于剪枝的 Cut-YOLOv3 算法模型,通过融合算法,在原生算法性能上得到进一步提高。

针对上述研究存在的问题,本文在此基础上提出了结合手部检测的塔吊驾驶人员行为规范监测方法,在 Faster R-CNN^[9-10] 网络的基础上,融合剪枝、注意力机制等算法,对模型的效率和精度进行了优化,通过手部和被测物体的 *IOU* 对比,更准确判断驾驶人员的违规行为,具备效率高、精度高和鲁棒性强等特点,并与 FRC-Tiny 检测模型和 Cut-YOLOv3^[11] 模型进行对比。实验结果表明,该模型在行为监测上表现更优。

1 CF-R-CNN 网络模型

1.1 整体网络结构

Faster R-CNN 是两阶段的检测算法,引入 Anchor 先验框,利用 RPN 快速生成候选框,预测每一个 Anchor 属于前景与背景的概率,生成区域建议框并计算真实物体相对于 Anchor 的偏移量。通过 RoI Pooling 层后,特征信息池化到统一的维度上,再输入到全连接层进行分类和回归,兼顾了目标检测的精度和效率。本文在此网络的基础上,融入了剪枝处理和通道注意力模块,在精度略有下降的同时,提高了模型的效率。基于改进的 Faster R-CNN 主干网络如图 1 所示。在此将网络模型分为两条支路,输入的数据分别经过两个 ResNet101 模块。第一部分是 ResNet101-FPN 层-RPN 层,利用 FPN 层的双向特征金字塔来提取全局特征;第二部分是 ResNet101 网络+注意力机制,用于生成局部特征,结合第一部分的全局特征进行上下文信息融合,通过 FC 层进行 softmax 回归。

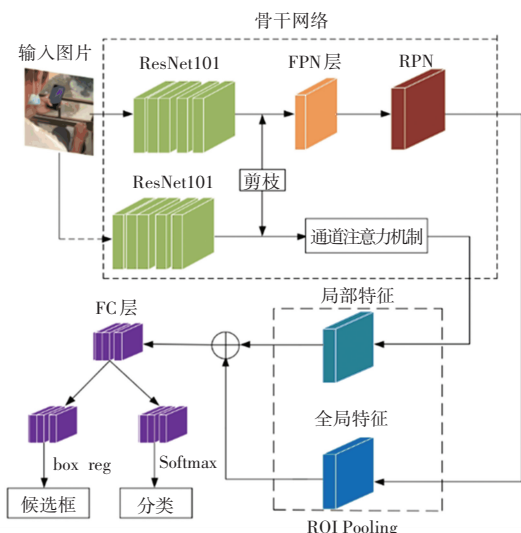


图1 主干网络 CF-R-CNN 模型图

Fig. 1 Backbone network model diagram of CF-R-CNN

1.2 模型剪枝处理

由于移动端的算力水平普遍偏低,若在移动端部署,需对网络进行轻量化处理,故对 Faster R-CNN 网络进行剪枝处理。在 Faster R-CNN 模型中,网络计算量集中在 ResNet101 模块上。文献[12-13]中提出的 ResNet101 网络,在目标检测中表现优异。其内部的残差块使用了跳接方式,缓解了神经网络深度加深时带来的梯度消失问题,保留 ResNet101 的残差跳接部分,对残差块的层数进行优化。但 ResNet101 网络在训练数据时,会将数据扩展到多维度上,有些维度上的权值较小,对特征不敏感,可考虑将其剪除达到网络“瘦身”的目的。具体的处理过程如图 2 所示。在 ResNet101 中的第二和第三个大残差块中,将其中的小残差块进行整合。如:输入 $56 \times 56 \times 256$ 维度的特征图,经过三层卷积之后又恢复成 $56 \times 56 \times 256$ 的特征图,再与卷积前的特征图跳接,经过激活层后,再通过三层卷积层最后输出 $28 \times 28 \times 512$ 的特征图。针对这些反复提取特征的残差块,本文先将输入信息输入给卷积核大小为 3×3 的卷积层,而不是原网络的 1×1 卷积核,增大了感受野,然后再把特征信息输入到 1×1 的卷积核中,对输入进行降维,再重复一次这样的操作,只是把第三层的步长 S 设为 2,进行上采样,最后将输入与输出进行融合后送入激活函数。跳接部分不采用输入直接与输出跳接,而是将输入进行上采样后,再与输出进行跳接,从而减少网络参数的计算量。

1.3 通道注意力机制

在获取被处理的图像时,摄像头的拍摄角度和人体遮挡等意外情况,使得检测准确率不理

想^[13-14]。因此在网络层中引入了注意力机制,加强对细节信息的感知能力。输入信息在被喂给网络的同时会生成多维特征图,注意力机制可以在成百上千的特征点中聚焦更为关键的信息,增强来自全局的感受野^[15]。通道注意力机制主要分为特征提取、计算得分和权重分配3部分。输入模块经过两层的3*3卷积提取,假设一个神经层中的输入序列 $X = [x_1, x_2, x_3, \dots,$

$x_n]$,首先通过线性变换得到3组向量序列:

$$Q = W_Q X \in R^{d_3 \times N} \quad (1)$$

$$K = W_K X \in R^{d_3 \times N} \quad (2)$$

$$V = W_V X \in R^{d_2 \times N} \quad (3)$$

其中, Q, K, V 分别为查询向量序列、键向量序列和值向量序列; W_Q, W_K, W_V 是可以学习的参数矩阵,输出后的维度保持不变。

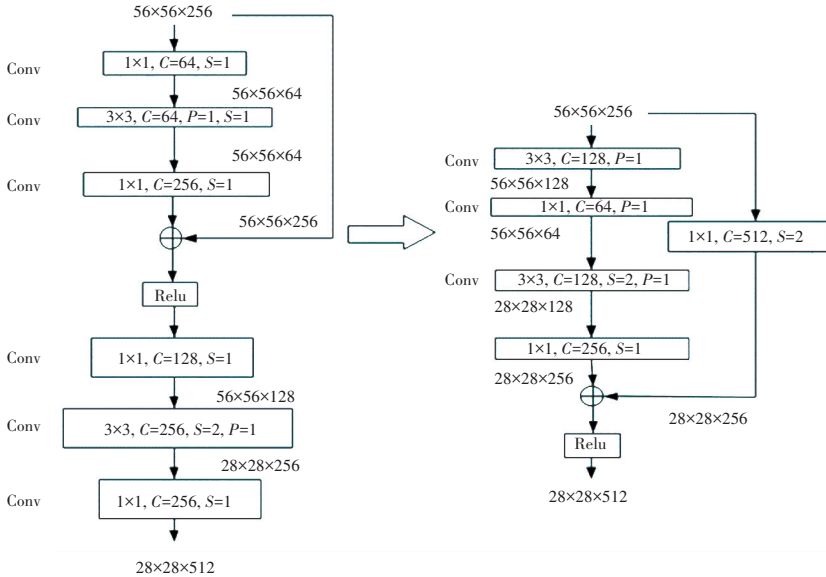


图2 ResNet 剪枝流程

Fig. 2 Pruning process of ResNet

通过全局平均池化的方式压缩通道上的特征信息,再经过打分模型公式:

$$s(x_i, q) = x_i^T q \quad (4)$$

计算注意力得分,利用 softmax 对注意力得分转换到 0~1 后,计算注意力分布:

$$\alpha = \text{softmax}(s_i) = \frac{\exp(s_i)}{\sum_{j=1}^N \exp(s_j)} \quad (5)$$

最后,采用加权求和的方式对输入信息进行汇总,得到注意力值:

$$\text{Attention}((K, V), Q) = \sum_{i=1}^N \alpha_i v_i \quad (6)$$

注意力值可以帮助模型在训练时重新分配最佳的权重,通道注意力模块在训练中能够学习通道之间的相关性,抵抗噪声和遮挡无关干扰。在模块中采用了卷积核的卷积层,仅引入少量参数就可提高模型的精度^[16-17]。

1.4 回归结果处理

图片信息输入到网络后,经过最后的 FC 层,回归结果有 4 类,即检测到手机、烟头和手,或者是三者都未检测到。为了提高模型的鲁棒性,本文采用

交并比算法判定检测结果。当检测到手机或烟头时,会进一步判断是否检测到了手部,通过手部和手机或烟头的预测框 IOU 占比计算,判断驾驶人员是否在玩手机或吸烟。这里的 IOU 为手部和手机或烟头的预测框的交并比,其计算公式如下:

$$IOU = \frac{A \cap B}{A \cup B} \quad (7)$$

IOU 数值介于 0~1 之间,反映了两个集合之间的相似度。如图 3 所示,其中的 A 代表手部的预测框, B 代表手机或烟头的预测框,分子是两者的交集,分母是两者的并集。通过设定 IOU 的阈值来判断手部和手机烟头是否存在接触,增加了预测的精确度。

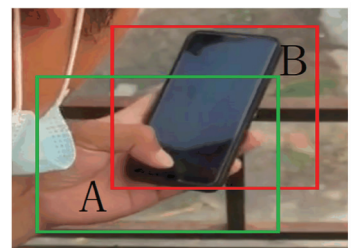


图3 IOU 占比图

Fig. 3 Proportion of IOU

2 实验结果与分析

2.1 数据集处理和训练参数设定

训练前将图片尺寸变换成 244×244 大小,并对数据进行归一化处理。为了提高泛化性能,本文对数据进行了增强处理:

- (1) 对现有数据进行旋转、平移、缩放和填充;
- (2) 在图像中加入椒盐噪声和高斯白噪声,扩充数据集。

训练过程中,训练集和验证集按照 8 : 2 分配。实验训练环境为: Ubuntu18.04 + NVIDIA GeForce RTX 3080 + AMD 锐龙 R7 - 5800 + Pytorch1.8 + python3.7;模型训练 70 轮,学习率为 0.001,每 20 轮降一半;根据图 4 中不同 *IOU* 阈值的精确度对比,阈值设置从 0.3 增大到 0.9 时,分类精度曲线先增后减,对于玩手机行为的分类精度在 0.5 时达到最高,抽样行为的分类精度则在 0.6 时达到峰值。

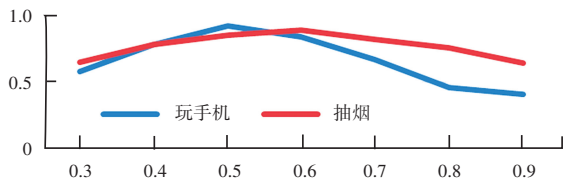


图 4 不同的 *IOU* 阈值下的分类精度

Fig. 4 Classification accuracy under different *IOU* thresholds

2.2 评价指标

评价指标可以衡量有一个模型的好坏程度,传统的指标是精确度(P)和召回率(R)。其中, P 的含义是真正例在所有预测为正例中的比例, R 的含义是真正例在所有样本正例中的比例。计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (8)$$

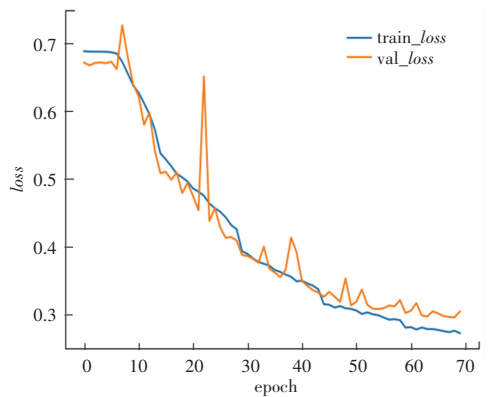
$$R = \frac{TP}{TP + FN} \quad (9)$$

因为 P 和 R 本身就是一对冲突的度量,为了兼顾 P 和 R ,采用 $F1$ 分数作为指标。 $F1$ 可理解为 P 和 R 的加权平均值, $F1$ 分数被归一化到 0 ~ 1 之间,1 为最佳分数,0 则为最低分数。 $F1$ 表达式如下:

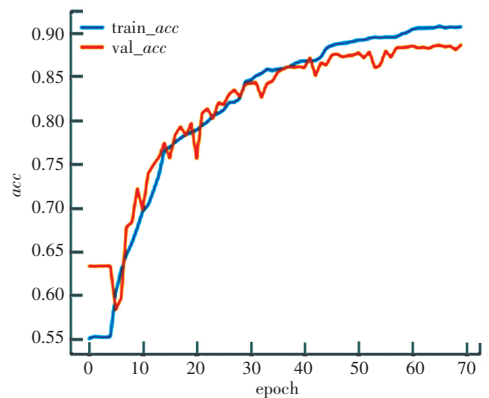
$$F1 = 2 \times \frac{P \times R}{P + R} \quad (10)$$

实时性指标则采用 FPS ,其含义为每秒传输帧数。 FPS 越高,则代表实时性越好。

由图 5 可知,到达 70 轮时,实验训练和验证的 $loss$ 都达到了底部,趋于稳定,训练和验证的 acc 达到了最高值。



(a) $Loss$



(b) acc

图 5 实验的 $loss$ 和 acc 图

Fig. 5 $Loss$ and acc diagram of experiment

在 ResNet 中的第二个和第三个大残差块中,将其中的小残差块进行整合,输入特征图经过多层卷积,先将输入信息送入大小为 3×3 的卷积核,再输入到 1×1 的卷积核中,对输入进行降维,经过一次重复操作,设置第三层步长 S ,进行上采样,最后将输入与输出融合后送入激活函数。具体的处理过程如图 2 所示。

2.3 实验结果分析

为了更直观体现改进后网络 and 原网络的区别,本文逐一添加各个模块进行对比实验,结果见表 1、表 2。

表 1 逐一加入各个模块下的性能对比

Tab. 1 Performance comparison under each module added one by one

网络	$F1$	FPS (Frame/s)
Faster-R-CNN	0.915	16.38
Faster-R-CNN+剪枝	0.872	22.86
Faster-R-CNN+通道注意力	0.923	16.12
CF-R-CNN	0.897	20.16

由表 1 可知,原生 Faster-R-CNN 网络在剪枝处理后的网络在 FPS 上提升 39.5%,但精度降低了

4.9%;添加通道注意力机制后,*FPS* 数值上略微下降,*F1* 提高了 0.8%;将上述算法融合改进后,网络的 *F1* 值相比原网络只降低了 2.1%,但 *FPS* 提高 23%,兼顾了精度和实时性的要求。因此,本文提出的模型在精度和实时性上优于 Cut-YOLOv3 和 FRC-Tiny 模型。

表 2 与引言其他(不同)模型的性能对比

Tab. 2 Performance comparison with other (different) models in the introduction

网络	<i>F1</i>	<i>FPS</i> (Frame/s)
Cut-YOLOv3	0.866	19.38
FRC-Tiny	0.906	18.45
CF-R-CNN	0.897	20.16

表 2 中,CF-R-CNN 模型的 *FPS* 值为 20.16,*F1* 数值为 0.897,对比 Cut-YOLOv3 模型,*FPS* 和 *F1* 数值分别提高了 3.6% 和 4.0%;对比 FRC-Tiny 模型,*F1* 数值下降了 1.0%,但 *FPS* 提高了 9.3%。

如图 6 所示,Faster R-CNN 方法能够正确的识别出数据集中的手机、烟头和手,并在图片中标记。



图 6 检测结果样例

Fig. 6 Example of test results

3 结束语

本文针对现有的驾驶人员行为监测方法不适用性、效率低和效果差等情况,提出了一种改进的 Faster R-CNN 方法。结合手部检测的方式,对 Faster-R-CNN 的 ResNet101 模块进行了剪枝处理,并融合注意力机制等方法。改进后网络的 *F1* 值相比原网络仅降低了 2.1%,但 *FPS* 相比提高了 23.0%,满足在移动端部署的要求;同时使用了交并比阈值的方法,提高了检测的精度。

虽然本文的实验研究对塔式起重机驾驶人员行为起到督促和监测的作用,对维护工地安全方面有一定的贡献,也存在着不足,如在极端情况下会发生漏判等问题。后续可考虑对手部进行 3D 建模等处理,增加手部和手机烟头的特征关联性,提升检测效果。

参考文献

- [1] 黄志超,赵红梅,陈奕迅,等. 基于面部识别的驾驶员疲劳危险驾驶检测系统[J]. 机电工程技术,2021,50(12):143-146.
- [2] 张越. 基于深度学习的抽烟与打手机行为检测算法研究[D]. 成都:电子科技大学,2021.
- [3] 郭静. 基于深度学习的异常驾驶行为检测识别[D]. 西安:长安大学,2021.
- [4] 梁宇宁,王绍华,金向明,等. 基于区域划分与标准时间的手部异常行为检测[J]. 北京航空航天大学学报,2021,47(10):1969-1979.
- [5] 黄向康. 基于深度学习的分心驾驶行为检测研究[D]. 重庆:重庆邮电大学,2021.
- [6] 陆一宁. 基于双线性注意力网络的分心驾驶图像分类算法研究[D]. 长春:吉林大学,2021.
- [7] 彭豪,李晓明. 基于改进 Faster R-CNN 的小目标检测模型[J]. 电子测量技术,2021,44(24):122-127.
- [8] 郑福建,汪彦,左应祥,等. 基于剪枝 YOLOv3 算法的高处作业人员安全带佩戴检测研究[J]. 信息系统工程,2021(10):117-119.
- [9] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
- [10] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [11] 冯号,黄朝兵,文元桥. 基于改进 YOLOv3 的遥感图像小目标检测[J]. 计算机应用,2022,42(12):3723-3732.
- [12] 施冬梅,肖锋. 基于改进长短时记忆网络的驾驶行为检测方法研究[J]. 汽车工程,2021,43(8):1203-1209,1262.
- [13] 吴萌萌,张泽斌,宋尧哲,等. 基于自适应特征增强的小目标检测网络[J]. 激光与光电子学进展,2023,60(6):65-72.
- [14] 郑伟成,李学伟,刘宏哲,等. 基于深度学习的疲劳驾驶检测算法[J]. 计算机工程,2020,46(7):21-29.
- [15] 汪兴伟. 基于视频的危險驾驶行为检测方法研究[D]. 南京:南京邮电大学,2021.
- [16] 龚圣斌,王少杰,侯亮,等. 基于 Faster-RCNN 算法的轻量化改进及其在沙滩废弃物检测中的应用[J]. 厦门大学学报(自然科学版),2022,61(2):253-261.
- [17] 霍光,林大为,刘元宁,等. 基于多尺度特征和注意力机制的轻量级虹膜分割模型[J/OL]. 吉林大学学报(工学版):1-12 [2022-04-12].