

基于 Faster R-CNN 的交通警察目标检测研究

柳胜超¹, 王夏黎¹, 王丽红¹, 柳秋萍²

(1. 长安大学 信息工程学院, 陕西 西安 710061;

2. 陕西航天动力高科技股份有限公司, 陕西 西安 710077)

摘要:针对自动驾驶技术的高速发展与日渐复杂的交通系统网络缺少能够准确检测与定位交通警察的方法,将基于深度学习对在复杂环境中交通警察的准确检测与定位进行研究,该研究一方面对于有效保障交通警察的人身安全和提高通行效率具有积极的促进作用,另一方面为后续的交通警察手势的识别提供重要的检测基础。但是在实际的应用当中交通警察一般处于复杂的场景当中,比如干扰人群、强光环境、复杂的天气,都会对交通警察的准确检测与定位造成影响。因此,该文首先通过对交通警察图像去噪和 Gamma 曲线校正的预处理方式来增强图像中的关键特征信息并排除噪声对图像的影响,然后分别对基于 SSD 网络模型的方法、基于 HOG 特征提取和 SVM 分类的方法、基于 Faster R-CNN 网络模型的方法进行实验,然后对三种方法进行对比,得出了基于 Faster R-CNN 网络模型的检测速度为 61.523 ms,检测准确率为 98.75%,均高于其他两种方法。

关键词:目标检测与定位;深度学习;图像预处理;SSD 网络;Faster R-CNN 网络

中图分类号:TP391.4

文献标识码:A

文章编号:1673-629X(2021)04-0080-06

doi:10.3969/j.issn.1673-629X.2021.04.014

Research on Target Detection of Traffic Police Based on Faster R-CNN

LIU Sheng-chao¹, WANG Xia-li¹, WANG Li-hong¹, LIU Qiu-ping²

(1. School of Information Engineering, Chang'an University, Xi'an 710061, China;

2. Shaanxi Aerospace Power High-tech Co., Ltd., Xi'an 710077, China)

Abstract: In view of the rapid development of autonomous driving technology and the increasingly complex transportation system network, there is a lack of methods that can accurately detect and locate traffic police. We will study the accurate detection and location of traffic police in complex environments based on deep learning. Effectively protecting the traffic police's personal safety and improving traffic efficiency have a positive role in promoting, on the other hand, it provides an important detection basis for the subsequent traffic police's gesture recognition. However, in actual applications, the traffic police are generally in complex scenes, such as interfering crowds, strong light environments and complex weather, which will affect the accurate detection and positioning of the traffic police. Therefore, we firstly preprocess the traffic police images to enhance the key feature information in the image and eliminate the influence of noise on the image, and then separately experiment on the method based on SSD network model, the method based on HOG feature extraction and SVM classification and the method based on Faster R-CNN network model. Comparing the three methods, it is concluded that the detection speed based on the Faster R-CNN network model is 61.523 ms, and the detection accuracy rate is 98.75%, which is higher than the other two methods.

Key words: target detection and localization; deep learning; image preprocessing; SSD network; faster R-CNN network

0 引言

伴随着计算机视觉的飞速发展,目标检测算法在自动驾驶和辅助驾驶领域也取得了重大突破,而在交通领域目标检测算法主要应用于对交通警察、信号灯、行人和道路标线等的检测。交通警察作为交通网络的

重要一员,能够有效疏导交通防止交通拥堵,对城市交通网络更好更快的发展有积极促进作用。因此对交通警察的检测算法的研究对于保护交通警察人身安全和保障交通网络健康发展具有重要意义。

传统的目标检测算法有很多,包括基于 HOG 和

收稿日期:2020-06-24

修回日期:2020-10-27

基金项目:国家自然科学基金(51678061)

作者简介:柳胜超(1993-),男,硕士研究生,研究方向为图形图像处理、基于深度学习的交通警察目标检测;王夏黎,副教授,研究方向为图像处理与智能交通领域的应用技术研究。

SVM 的跟踪^[1]、学习和检测的算法,但是该方法容易受到光照、遮挡和尺度变化等因素的影响,并且提取的特征单一,鲁棒性不强。如今深度学习的快速发展也极大地推动了目标检测算法的研究,基于卷积神经网络的目标检测算法也成为当前主流的方法,其中主要包括两种,一种是基于回归思想的 One-stage 类别的目标检测算法,另一种是基于候选框思想的 Two-stage 类别的目标检测算法。目前常用的典型的 One-stage 目标检测网络包括 YOLOV1、YOLOV2^[2]、YOLOV3^[3]和 SSD^[4]等直接回归目标物体的类别和位置。Two-stage 目标检测网络包括 Faster R-CNN^[5]、Fast R-CNN^[6]和 RCNN^[7-8]等,该类型的网络主要分两步,一步是由算法生成一系列作为样本的候选框,另一步是由卷积神经网络进行样本分类^[9]。两种类别的方法各有优缺点,该文分别对基于 SSD 网络和基于 Faster R-CNN 网络的交通警察的检测进行实验,实验结果最终表明基于 Faster R-CNN 网络可以有效提高交通警察检测的精确率和检测速度。

1 交通警察采集图像的预处理

1.1 图像去噪

数字图像在传输和数值化过程中会受到诸多因素的影响造成图像质量下降,对后续图像的处理和目标的检测带来困难。比如图像噪声和成像设备的影响,因此需要对图像进行去噪处理来提高图像质量,增加信噪比,更好地突出需要的某些信息和特征。该文将分别采用中值滤波和小波变换去噪^[10]的方法对图像进行去噪,实验结果表明基于小波变换去噪的方法满足了后续交通警察的目标检测要求。

(1)中值滤波。

中值滤波是由 Tukey 在 1977 年提出的一种非线性平滑滤波器^[11],基本原理是该像素点的灰度值由该像素点邻域灰度值的中值来取代。令 $f_{x,y} = \text{Med}(x, y) \in E^2$ 表示图像各个点的灰度值。则滤波窗口为 A 的二维中值滤波表达式为:

$$f_{x,y} = \text{Med}_A(x_{i,j}) = \text{Med}\{x_{i+r,j+s}(r+s)\} \in A(i,j) \in E^2 \quad (1)$$

在图像处理中,中值滤波可以在较完整保留图像部分边缘信息的前提下有效去除如椒盐噪声、脉冲噪声的影响^[12]。但是当待处理图像细节部分较多时,中值滤波去噪会使得图像缺失部分关键信息。

(2)小波变换去噪。

基于小波变换 (wavelet transform, WT) 的图像去噪方法结合特征提取和低通滤波的综合作用,小波降噪常根据阈值的选择来进行降噪处理。一般步骤分为图像对数变换、小波分解变换和低频系数阈值处理。阈值函数一般分为软阈值函数和硬阈值函数,表达式分别为式(3)和式(4)。小波变换的表达式为:

$$\text{WT}(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) * \Psi\left(\frac{t-\tau}{a}\right) dt \quad (2)$$

其中,只有两个变量平移 a 和平移量 τ , a 控制图像而 $\psi(t)$ 是给定的函数。

$$\chi(a, b) = \begin{cases} a + b & a \leq b \\ 0 & |a| \leq b \\ a - b & a \geq b \end{cases} \quad (3)$$

$$\chi(a, b) = \begin{cases} a & |a| \geq b \\ 0 & |a| < b \end{cases} \quad (4)$$

其中, a 是变量, b 是阈值。对于阈值的选择应满足式(5)。

$$b = w_n \sqrt{2 \ln N} \quad (5)$$

其中, w_n 表示噪声标准方差, N 表示信号的长度。

小波变换具有多分辨率、多尺度和稳定的信号的特性,能够有效获得待处理图像细节处的信息和轮廓信息。该文通过对同一包含交通警察目标图像做小波变换去噪和中值滤波去噪来消除图像噪声的预处理,得到的实验结果如图 1 所示。对比实验结果可以看出,小波变换去噪的方法能有效保留交通警察图像的细节信息和边缘特征,符合交通警察目标检测的需要。所以该文采用基于小波变换去噪的方法对图像进行去噪处理。



图 1 图像去噪结果

1.2 基于 Gamma 曲线校正的图像增强

Gamma^[13] 曲线是一种特殊的色调曲线,当 Gamma 值等于 1 时,如图 2 中灰色虚线,输入和输出的密度相同;当 Gamma 值大于 1 时,如图 2 中 gamma2 曲线,输出图像结果出现亮化;当 Gamma 值小于 1 时,如图 2 中 gamma1 曲线,输出图像结果出现暗化。在计算机系统中,由于硬件设备如显卡和显示器的原因会导致最后输出的图像存在亮度上的差异,而 Gamma 曲线校正就是通过一定的方法来校正图像的这种偏差的方法。实际情况下,当用于 Gamma 校正的值大于 1 时,将会使得图像的高光部分被压缩而暗调部分被扩展;当 Gamma 校正的值小于 1 时,会使得图像的暗调部分被压缩而高光部分被扩展,Gamma 校正一般用于平滑的扩展暗调的细节。Gamma 校正的曲线如图 2 所示。当已经知道 Gamma 值时可以通过式(6)的方法进行 Gamma 校正。

$$\Psi = \lambda\mu^\gamma \tag{6}$$

其中, μ 为图像的像素值, λ, γ 为常数。当 $\gamma > 1$ 时图像的亮度会降低,图像变暗。当 $\gamma < 1$ 时,图像的亮度增加。

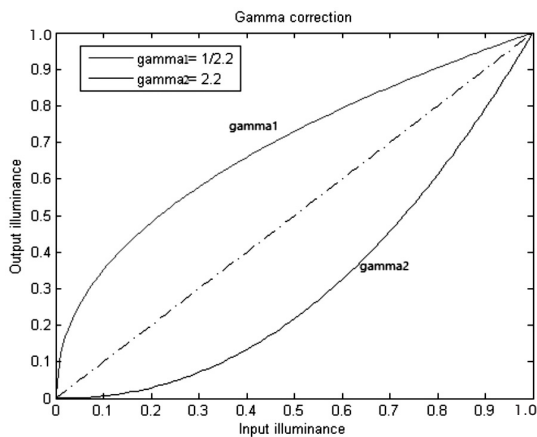


图 2 Gamma 曲线

通过对小波变换去噪后的图像进行 Gamma 校正得到增强后的图像实验结果,如图 3 所示,可以看出 Gamma 曲线校正的结果满足该文待检测图像的需要,在图像未失真的前提下有效增强了图像的细节信息。



(a)小波变换效果 (b)Gamma 效果

图 3 图像增强

2 基于深度学习的交通警察检测

Faster R-CNN 可以看成基于“区域生成网络 RPN+ Fast R-CNN”的系统,而 Fast R-CNN 中的 Search 方法是由区域生成网络代替的。通过区域生成网络(RPN)来提取候选区域,有效地将 RPN 网络和 Fast R-CNN 网络统一到一个网络共享卷积层,并且解决了 R-CNN 和 Fast R-CNN 特征提取中存在的冗余问题。

2.1 R-CNN 模型

由 Girshick 等人提出的区域卷积神经网络 (region-based convolutional neural network, R-CNN)^[14] 根据传统的目标检测思路,R-CNN 算法将目标检测分为提取框内图像特征提取、对图像进行特征分类、非极大值抑制的步骤实现目标检测,在特征提取阶段利用深度学习的方法提取特征^[15]。采用 Selective Search 方法得到 2 000 个左右的候选区域,在固定大小后输入到卷积神经网络进行特征提取,利用多个 SVM 对特征进行分类,最后校准候选区域的位置。

2.2 Fast R-CNN 模型

Fast R-CNN 模型^[16]通过整张图像归一化后直接送入深度网络,随后送入提取的候选区域,避免了前几层特征不必要的重复计算,优化了 R-CNN 模型训练、测试速度慢和训练空间大的问题。具体的实现过程如图 4 所示。相较于 R-CNN 模型,一方面 Fast R-CNN 模型增加了池化层 (region of interest, ROI),可以把任意大小的输入降维成设定固定尺度特征向量,另一方面,边框回归和分类可以同时进行,共享卷积层并相互促进。

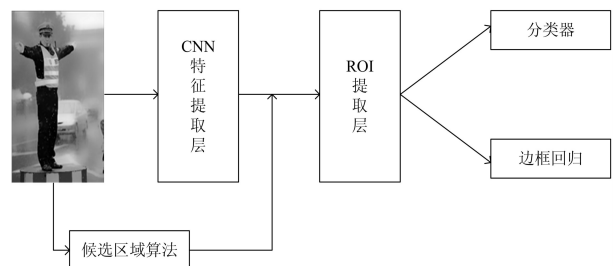


图 4 Fast R-CNN 网络结构

2.3 Faster R-CNN 网络检测流程

Faster R-CNN 网络模型将候选区域获取、深度信息提取与目标检测识别融入到同一卷积神经网络模型中,实现了端到端的目标检测与训练^[17-20]。主要步骤为:

- (1) 卷积层特征提取:通过一系列的激化、卷积和池化层的组合,对输入图像进行特征的提取。
- (2) RPN 获取候选区域:通过 RPN 的滑动窗口生成候选框,设置候选框的尺寸和比例可以看作一个尺度 51 * 39 的 256 通道图像,对于该图像的每一个位

置,共组合为9个可能的候选窗口;三种面积三种比例,这些候选窗口称为锚框(anchors boxes)。通过利用边框回归修正 anchors boxes 获得更加准确的候选框,根据非极大值抑制算法(NMS)进行输出区域筛选,得到 M 个不同大小和比例的目标候选框输入到 ROI 池化层。

(3)ROI 池化层获取区域建议特征:利用 RPN 得到 M 个不同比例的候选框和卷积层提取到的特征得到固定大小的特征向量^[21]。

(4)分类和回归:利用 Softmax 和 Proposals 判断目标候选区域的具体类别,利用 Bounding-box 回归的思想预测目标物体的准确位置^[22]。

2.4 损失函数

Faster R-CNN 损失主要包括两部分,分别为 RPN 损失和 Fast RCNN 损失,计算公式如式(7)所示,并且两者都包括分类损失和回归损失。

$$L(\{p\}, \{t_i\}) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(t_i, t_i^*) \quad (7)$$

其中, i 表示候选区域索引; p_i 表示 anchor 预测为目标的概率; $p_i^* \in \{0,1\}$ 表示目标和非目标。

分类损失表达式为:

$$L_{\text{cls}}(p_i, p_i^*) = -\log[p_i p_i^* + (1 - p_i^*)(1 - p_i)] \quad (8)$$

其中, t 表示 Faster R-CNN 阶段的预测偏移量; t^* 表示 Faster R-CNN 阶段前景实际偏移量。

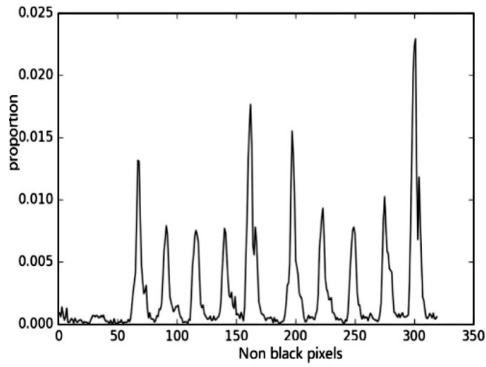
回归损失表达式为:

$$L_{\text{reg}}(t_i, t_i^*) = \sum_{i \in (x,y,w,h)} \text{smooth}_{L_1}(t_i - t_i^*) \quad (9)$$

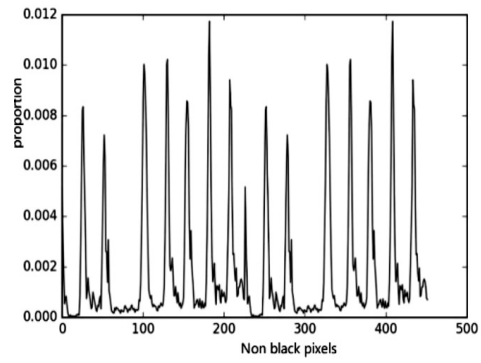
其中,平滑 L_1 损失函数 $\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$ 。

2.5 交通警察的定位

交通警察在指挥交通时一般身体在局部运动状态,因此利用帧间差分法进一步对视频帧中的目标进行检测,对于提高目标检测的精确率有较为理想的帮助。对运动目标具体检测的方法是:帧间差分法通过差分运算对时间上相邻 2-3 帧图像中的像素点做减法运算,当灰度差的绝对值超过一定阈值时,判断为所检测的运动目标^[23-24]。通过统计各个交通警察局部运动非黑像素阈值范围,以此来进一步实现对目标区域中指挥交通警察的定位。图 5(a)是对交通警察变道信号非黑像素占比的统计,图 5(b)是对左转待转非黑像素占比。



(a)变道非黑像素占比



(b)左转非黑像素占比

图5 非黑像素占比

3 实验与分析

3.1 实验环境

仿真实验环境:处理器 AMD Ryzen 5 3600 6-Core Processor 3.59 Hz;GPU 为 NVIDIA GeForce RTX2060

SUPER,编程语言为 python;操作系统为 Ubuntu16.04;深度学习框架采用 Tensorflow^[25]。

3.2 实验参数

对整个网络训练的参数设置如表 1 所示。

表1 实验参数

参数	值
Momentum	0.9
Batch size	0.1
学习率	0.001
权重衰减	0.0005
迭代次数	300000

3.3 网络对比训练

分别使用 Faster R-CNN 网络和 SSD 网络对测试集中的 11 228 张图片进行实验测试,得到的部分检测效果图如图 6 所示。

图 6 中,(a)、(c)表示 Faster R-CNN 网络检测的效果图,(b)、(d)表示 SSD 网络检测的效果图。通过

对比可以发现,SSD 网络对交通警察的检测率均低于 Faster R-CNN 网络的检测率,并且对比图 (c) 和 (d) 可以发现 SSD 网络存在漏检的问题,而 Faster R-CNN 网络则成功地检测出了图像中存在的第三位交通警察。

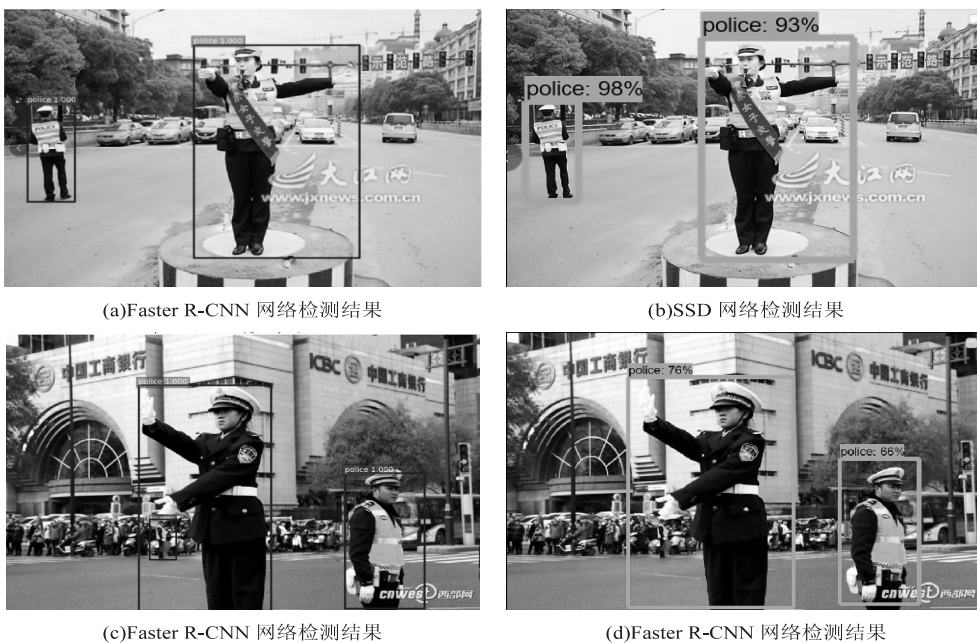


图 6 检测效果

通过网络对比实验,分别对两种基于深度学习的方法和传统的基于 HOG 和 SVM 的方法进行检测,实验结果如表 2 所示。通过表 2 可以直观地看出,Faster

R-CNN 网络的检测速度为 61.523 ms,准确率为 98.75%,相较于 SSD 网络和传统的方法都具有明显优势。

表 2 实验结果对比

检测方法	准确率/%	识别速度/ms
Faster R-CNN	98.75	61.523
SSD	91.61	81.252
文献[1]	89.65	101.352

通过实验同时得出了两种方法的损失值曲线,从图 7 中可以看出损失值呈下降的趋势,Faster R-CNN 网络的损失值在 20 000 次左右已经下降到 0.3 以下,最终在 30 000 次迭代停止时损失值为 0.112 5,表明

Faster R-CNN 模型训练效果较好。SSD 模型最终在 30 000 次迭代时停止学习,损失值在 0.204 5 左右,从曲线中可以看出 SSD 模型的波动较大,训练效果无法满足检测的要求。

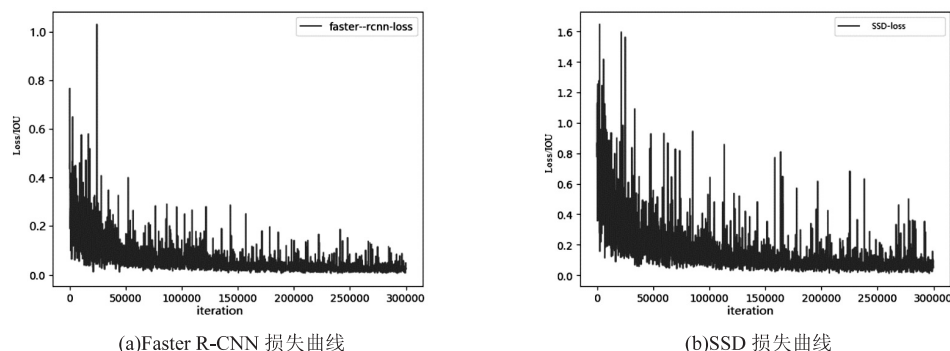


图 7 损失曲线

综上所述,对于交通警察的检测 Faster R-CNN 模型检测的成功率和速度均高于 SSD 模型, Faster R-CNN 模型能够精确地定位到复杂背景下的交通警察。

4 结束语

该文提出一种基于 Faster R-CNN 模型与帧间差分法结合的方法来实现交通警察的检测,通过对比 SSD 模型和 Faster R-CNN 模型的实验数据结果可以看到,虽然两种方法均能检测出目标,但 Faster R-CNN 模型的检测准确率达到 98.5%,检测速度仅为 50 ms,检测的召回率达到 98.2%,均比 SSD 模型有较高的提升。该方法能够有效解决目前交通警察存在的漏检、误检和检测效果不佳的问题,并且能够有效排除光照和遮挡物的影响,提高对交通视频帧中交通警察检测的精确性和鲁棒性。

参考文献:

- [1] 翟社平,李 炆,马蒙雨,等. 基于 LBP 和 SVM 的人脸检测[J]. 计算机技术与发展,2017,27(9):44-47.
- [2] 李云鹏,侯凌燕,王 超. 基于 YOLOv2 的复杂场景下车辆目标检测[J]. 电视技术,2018,42(5):100-106.
- [3] 徐 融,邱晓晖. 一种改进的 YOLO V3 目标检测方法[J]. 计算机技术与发展,2020,30(7):30-33.
- [4] 赵建国,曹朝辉,梁 杰. 卷积神经网络 SSD 的道路目标检测[J]. 机械设计与制造,2020(6):181-184.
- [5] 邝先验,李洪伟,杨 柳. 基于改进 Faster R-CNN 算法的两轮车视频检测[J]. 现代电子技术,2020,43(9):129-134.
- [6] 蒋宇豪. 基于深度学习的视频预测及视频行人检测方法研究[D]. 济南:山东科技大学,2017.
- [7] 原 蕾. 基于非负稀疏矩阵分解的人脸表情识别[D]. 哈尔滨:哈尔滨工程大学,2013.
- [8] 夏 军. 机器人在复杂条件下人脸检测与识别算法研究[D]. 兰州:西北师范大学,2017.
- [9] 吕石磊,卢思华,李 震,等. 基于改进 YOLOv3-LITE 轻量级神经网络的柑橘识别方法[J]. 农业工程学报,2019,35(17):205-214.
- [10] 段晓杰,张绍成,曲大鹏,等. 基于混沌和小波变换系数的数字图像水印算法[J]. 计算机技术与发展,2015,25(10):34-37.
- [11] 李 建,赵玮烨,李积宪. 复合噪声图像去噪[J]. 兰州交通大学学报,2007,26(4):116-118.
- [12] 江 帆,田 青. 基于结构化稀疏投影的多视图特征提取框架[J]. 计算机技术与发展,2019,29(3):6-11.
- [13] 丁 毅,李玉惠,李 勃. 基于图像不同亮度区域特征的 Gamma 矫正方法[J]. 计算机技术与发展,2016,26(6):37-39.
- [14] SAKOE H, CHIBA S. Dynamic programming algorithm optimization for spoken word recognition[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1978, 26(1):43-49.
- [15] 林永春. 载人潜水器水下摄像信息识别技术研究[D]. 青岛:青岛科技大学,2019.
- [16] 周云成,许童羽,邓寒冰,等. 基于双卷积链 Fast R-CNN 的番茄关键器官识别方法[J]. 沈阳农业大学学报,2018,49(1):65-74.
- [17] 陆兴华,蔡 韬. 基于 CNN 的安防监控步态特征提取研究[J]. 计算机技术与发展,2019,29(11):123-127.
- [18] 王婷婷. 改进的 Faster R-CNN 在目标参数测量方面的研究与应用[D]. 哈尔滨:哈尔滨工程大学,2018.
- [19] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//European conference on computer vision. Amsterdam, The Netherlands: Springer, 2016:21-37.
- [20] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. Montreal, Quebec, Canada: NIPS, 2015:91-99.
- [21] EVERINGHAM M, GOOL L V, WILLIAMS C K, et al. The pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2):303-338.
- [22] WANG J, CHEN K, YANG S, et al. Region proposal by guided anchoring[C]//Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). Long Beach, CA, USA: IEEE, 2019:2965-2974.
- [23] 赵嘉兴,王夏黎,王丽红,等. 多尺度密集时序卷积网络的单幅图像去雨方法[J]. 计算机技术与发展,2020,30(5):115-120.
- [24] 李虎月,郝鹏飞,廖云霞,等. 车牌检测与识别系统设计[J]. 计算机技术与发展,2020,30(7):150-153.
- [25] 邢艳芳,段红秀,何光威. TensorFlow 在图像识别系统中的应用[J]. 计算机技术与发展,2019,29(5):192-196.