

基于改进的 Faster R-CNN 的古建筑地砖缺陷检测

陈利^{1,2}, 刘艳艳^{1,2}

(1.南开大学 光电子薄膜器件与技术天津市重点实验室, 天津 300350;

2.南开大学 薄膜光电子技术教育部工程研究中心, 天津 300350)

摘要: 缺陷检测对于古建筑的保护和修缮具有重要的意义, 传统的地砖缺陷检测通过目视检查, 存在受人力影响大、耗时长等限制。基于深度学习的良好应用前景, 建立故宫地砖缺陷的数据集, 提出改进型 Faster R-CNN 的网络。首先, 构建可变形卷积, 通过网络学习并提取地砖中的缺陷特征; 然后, 将特征图输入区域生成网络中生成候选区域框, 将生成的特征图和候选区域框进行池化操作; 最后, 输出缺陷检测结果。在故宫地砖图片数据集的测试下, 改进后的模型平均准确率均值到达 92.49%, 与 Faster R-CNN 模型相比提高了 2.99%, 更适用于地砖缺陷检测。

关键词: 缺陷检测; Faster R-CNN; 可变形卷积

中图分类号: TN03; TP181

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.200555

中文引用格式: 陈利, 刘艳艳. 基于改进的 Faster R-CNN 的古建筑地砖缺陷检测[J]. 电子技术应用, 2021, 47(1): 31–35.

英文引用格式: Chen Li, Liu Yanyan. Defects detection of floor tiles of ancient buildings based on Faster R-CNN[J]. Application of Electronic Technique, 2021, 47(1): 31–35.

Defects detection of floor tiles of ancient buildings based on Faster R-CNN

Chen Li^{1,2}, Liu Yanyan^{1,2}

(1.Key Laboratory for Photoelectronic Thin Film Devices and Technology of Tianjin, Nankai University, Tianjin 300350, China;

2.Engineering Research Center of Thin Film Optoelectronics Technology, Ministry of Education, Nankai University, Tianjin 300350, China)

Abstract: Defect detection is of great significance for the protection and repair of ancient buildings. The traditional floor tile defect detection has been subject to visual inspection, which has limitations due to human influence and time-consuming. Based on the good application prospects of deep learning, this paper builds a data set of imperfections in the Forbidden City, and proposes an improved Faster R-CNN. Firstly, the deformable convolution was constructed, and the defect features in the floor tile were learned and extracted through the network. Then, the feature graph was input into region proposal network to generate the candidate region box, and the generated feature graph and candidate region box was pooled. Finally, the defect detection results were output. Under the test of the image data set of floor tiles of the Forbidden City, the mean accuracy of the improved model reached 92.49%, which was 2.99% higher than the Faster R-CNN model and more suitable for the floor tile defect detection.

Key words: defect detection; Faster R-CNN; deformable convolution

0 引言

随着科技高速发展, 对文物的保护和修缮越来越得到重视。传统对文物的缺陷检测主要依靠人力进行目视检查, 但容易受到天气、时间等原因影响。地砖缺陷具有形状不规则、背景噪声系数大等特征, 目前大多数缺陷检测算法都是根据应用场景不同进行手工提取缺陷特征, 直接或者通过机器学习算法进行分类^[1]。这种有监督机器学习存在一定局限性, 受图片中缺陷类别数目、特征形状等因素影响, 人为提取特征需要具有很强的专业性, 检测结果不好, 鲁棒性差, 所以不能很好地适用于对地砖缺陷检测。

随着计算机视觉不断发展, 深度学习作为计算机视

觉的分支, 越来越受到人们重视, 目标检测是深度学习的广泛应用之一。近些年来, 目标检测取得了很大突破。目标检测主要分为两类: 一类是基于候选框的 R-CNN (Region Convolutional Neural Network) 系列算法, 如 R-CNN、Fast R-CNN (Fast Region Convolutional Neural Network)、Faster R-CNN (Faster Region Convolutional Neural Network)^[2], 它们是生成候选框后进行分类和位置回归; 另一类是 YOLO (You Only Look Once)^[3]、SSD (Single Shot MultiBox Detector), 从回归角度出发, 直接在图像中回归出目标边框和位置, 这类算法仅使用一个卷积神经网络。第一类方法准确度高, 速度慢; 第二类算法速度快, 可以到达实时检测, 但是准确性低。

目前基于深度学习的目标检测算法很多,应用在目标检测的效果也很突出^[4-6],但是现有算法在缺陷检测中并不能很好地体现出来^[7-9],尤其是在地砖缺陷检测中,现有目标检测算法受限于地砖特征的多样性以及纹理特性。为此,本文提出了一种基于改进型 Faster R-CNN 网络用于检测地砖缺陷。对卷积核中每个采样点位置都增加了一个偏移变量,通过这些变量,卷积核就可以在当前位置附近随意采样,而不再局限于之前的规则格点,形状多变的感受野丰富了语义信息,从而提高检测精度^[3]。

1 基本原理

1.1 可变形卷积

在地砖缺陷数据集中,缺陷具有大小不一、形状不同等特点,而 Faster R-CNN 中卷积网络模块采用固定 3×3 卷积核,所有激活层的感受野具有一致性,但是由于不同位置可能对应着不同尺度的物体,因此很有必要对尺度或者感受野进行自适应调整,进而更精确地定位。地砖缺陷这类非刚性目标,在图片中形状、大小不同,固定的几何结构已经无法提取更准确的特征。为此,本文引入可变形卷积模块,将标准卷积核替换成可变形卷积核^[10]。

可变形卷积示意图如图 1 所示,图 1(a)是标准的 3×3 卷积操作,图 1(b)是加入偏置量之后的采样点的变化,图 1(c)是特殊的各种尺度变换,当出现图 1(d)所示情况时,则卷积具有各向异性、旋转变换等特点。

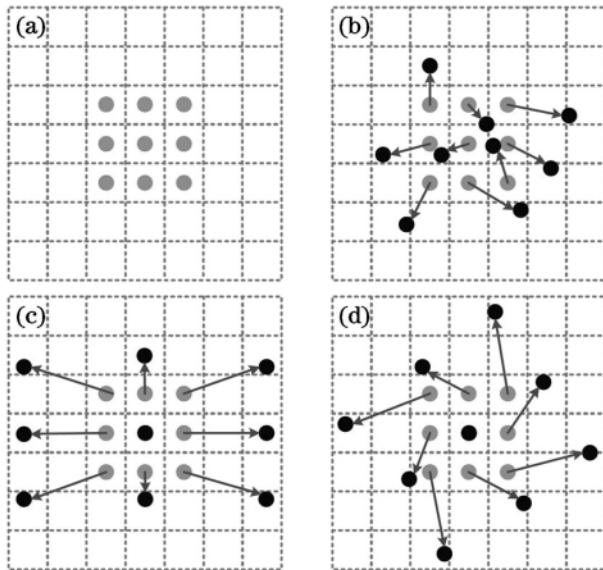


图 1 可变形卷积采样点变化

在可变形卷积网络中,卷积和池化操作是二维,在同一个通道上进行,卷积操作在输入的特征图 X 上使用规则网格 R 采样,然后进行加权运算, R 定义了感受的大小和扩张。对于在输出的特征图上的每个点的位置 P_0 ,通过下式计算:

$$Y(p_0) = \sum_{p_n \in R} W(p_n) \cdot X(p_0 + p_n + \Delta p_n) \quad (1)$$

其中, Δp_n 为偏置量,是一个分数; W 是采样点权重; p_n 是标准采样位置。现在,采样位置变成了不固定位置,由于 Δp_n 通常是小数,而非整数坐标是无法在图像这种离散整数数据上使用,因此通过双线性插值法进行实现,该方法通过寻找距离这个对应坐标最近的 4 个像素点来计算该点值。可变形卷积网络结构如图 2 所示,在输入特征图中,原始卷积网络通过滑动窗口直接输出特征图,加入可变形卷积后,原来的卷积网络分为两路,上面一路引入一个额外卷积层学习偏移量,输出一个 $H \times W \times 2N$ 的偏移量, $2N$ 有 x, y 两个方向的偏移,得到的滑动窗口不是 3×3 ,而是经过平移后的窗口,特征图和偏移作为可变形卷积输入^[11]。加入可变形卷积可以使感受野自适应目标的尺度变化,打破了原始卷积网络采用固定感受野的限制,对特征信息的融合更加灵活,以少量复杂模型的代价提高了整体网络对形变的学习能力。

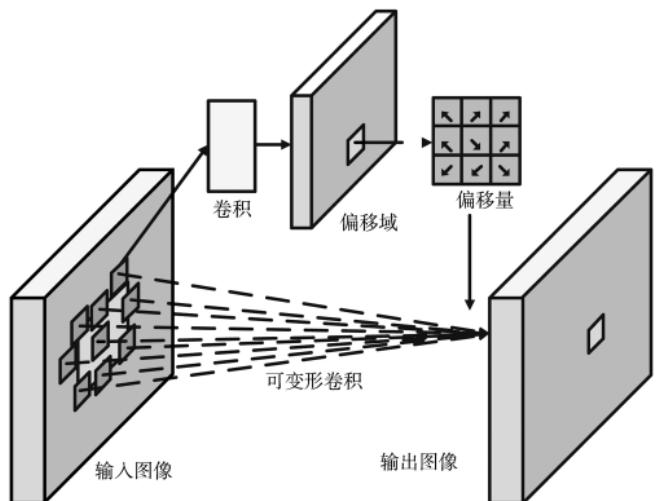


图 2 可变形卷积的实现过程

1.2 区域生成网络

经典的检测方法生成检测框非常耗时,例如 R-CNN 和 Fast R-CNN 使用选择性搜索方法生成检测框,而 Faster R-CNN 摒弃了原始滑动窗口和选择性搜索,直接使用区域候选网络生成检测框,极大提升了检测框的生成速度。

区域生成网络结构图如图 3 所示。在得到特征图后,经过一次 3×3 卷积,得到新的 $n-d$ 特征图,这样做可以将特征信息集中,然后通过分类层和回归层获取目标的分类和位置。其中,按照不同的长宽比、不同面积的锚框在滑动窗口处得到 k 个候选区域,每个候选区域和分类准确性得到前景概率和背景概率对应图片上的概率得分。由于每个候选区域和分类准确性的差异性,从候选区域通过平移缩放得到分类准确性需要 4 个平移缩放参数。候选区域作为区域生成网络的输出由两个任

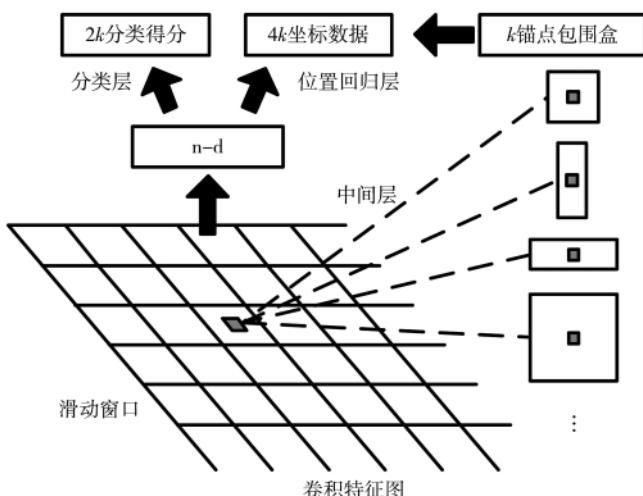


图 3 区域生成网络

务判定:(1)通过 Softmax 判定候选区域属于前景类还是背景类;(2)通过边界回归预测候选框的偏移量,来降低定位的误差。

区域生成网络训练时要把区域生成网络分类和回归的损失都加进来一起训练。损失公式如下:

$$L(\{p_i\}, \{u_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (2)$$

其中, N_{cls} 是一个 batch 的大小 256, $L_{cls}(p_i, p_i^*)$ 是前景和背景的对数损失, p_i 是锚框预测为目标的概率, p_i^* 是前景的标签, N_{reg} 是锚框的总数, λ 是两种损失的平衡比例, t_i 是候选区域的标签, t_i^* 是训练时每一个锚框与分类准确值间的偏差。

1.3 感兴趣区域池化层

感兴趣区域池化层负责收集候选区域,并得出候选区域特征图,送入到后续全连接层网络。该层有两个输入,一个是原始特征图,另一个是区域生成网络输出大小不同的候选框。区域生成网络生成候选区域是通过对锚框进行边框回归操作得到,所以候选区域大小形状各不相同,而传统的卷积网络训练好后输入图像尺寸必须是固定值,所以要进行感兴趣区域池化操作。感兴趣区域池化层通过对输入的特征图进行两次量化得到固定尺寸的输出。第一次是对输入特征图进行取整操作,第二次是将特征图分为 49 个相同大小的区域,对于每一个小区域,使用最大池化的操作从中选取最大的像素点作为输出,这样,就达到输出固定尺寸特征图目的了。

1.4 改进型 Faster R-CNN 网络

以 ResNet101(Residual Network, ResNet)为骨干网络,网络模型如图 4 所示。输入图片经过一个 18 层和 34 层的简单网络,即将所有层进行简单的叠加,然后经过同样层数的残差网络,被调整为 512×512 的像素。在利用

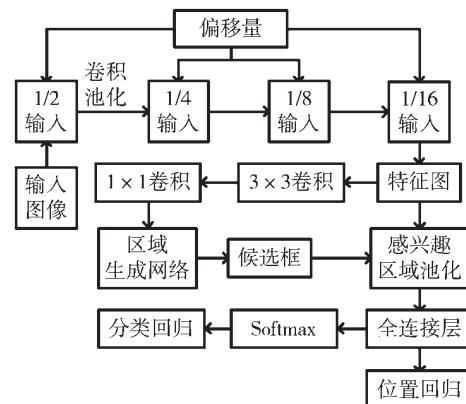


图 4 改进的 Faster R-CNN 网络

卷积网络提取特征的同时引入可变形卷积增加后者减小卷积核的大小,使卷积核自适应调整为缺陷特征的大小,灵活调整感受野,获得不同感受野的信息将其融合。获得的特征图经过 3×3 卷积后输入进区域生成网络得到候选框,获得缺陷的大致位置,继续训练,获得更准确的位置。利用前面获得的精确位置,从特征图中得出位置回归和分类的目标。

2 实验分析结果

2.1 样本采集

实验使用的数据集来自于现场采集,图像分辨率约 5 000 像素×3 000 像素。样本采集首先通过矩形框选出缺陷具体位置。采集的样本数量充足,背景丰富,避免了类别不平衡造成检测精度的下降。数据集如图 5 所示。地砖缺陷分为 4 类,包括磨损(abrasion)、崩落(collapse)、开裂(crack)、风化(weathering)。本文按照 PASCAL VOC 的标注方式建立了一个故宫地砖缺陷的数据集,用 LabelImg 工具对地砖数据集进行标注,标签包括 perfect、abrasion、collapse、crack、weathering。数据集包括 314 张图片,其中训练集 245 张,测试集 62 张。

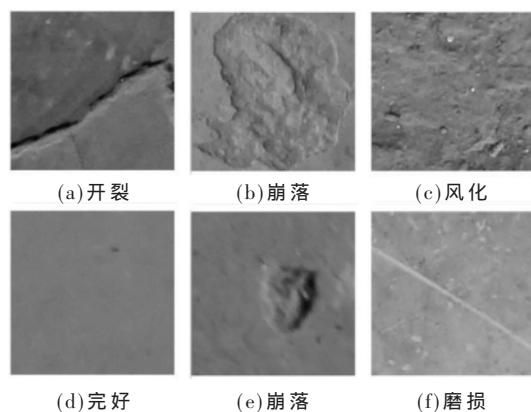


图 5 故宫缺陷样本

2.2 训练过程

本文在 Ubuntu16.04 系统下采用 GeForce GTX1080 TI GPU 加速,内存为 15 GB,算法在 Mxnet 框架下

采用 Python2.7 编写实现。

实验分为 3 个阶段, 第一阶段采用 VGG16(Visual Geometry Group, VGG)分类模型和 Faster R-CNN 网络训练数据, 参数设定为: 学习率为 0.001, 初始动量为 0.9, 权重衰减为 0.0001, 共迭代 50 000 次。VGG16 包括 13 个卷积层、3 个全连接层和 5 个池化层。VGG16 使用了 3 个 3×3 卷积核来代替 7×7 卷积核, 使用了 2 个 3×3 卷积核代替 5×5 卷积核, 在保证具有相同感受野的条件下, 增加了网络的深度, 提升了网络的效果。

第二阶段采用 ResNet101 分类模型和 Faster R-CNN 网络训练数据, 参数设定同第一阶段一样。首先构建了一个 18 层和一个 34 层的简单网络, 然后构建了一个 18 层和一个 34 层残差网络, 这两个网络的参数量、计算量相同, 并且与之前 VGG-16 相比, 计算量要小很多。

第三阶段采用 ResNet101 网络和改进的 Faster R-CNN 网络训练。参数设定为: 学习率设置为 0.0005, 初始动量为 0.9, 初始权重为 0.0005, 共迭代 200 轮。

2.3 实验结果

为评估算法在砖块缺陷检测中的有效性, 采用平均准确率均值(mean Average Precision, mAP)来作为衡量模型性能指标。mAP 是指所有类别目标的平均准确率(Average Precision, AP)的均值, 它可以作为目标检测实际度量标准。

本文对比了不同网络模型对地砖缺陷检测的精度, 如表 1 所示。

实验结果表明, 本文改进的网络在测试集上取得了最高精度, 在测试集上的平均准确率均值为 92.49%, 相比原始 Faster R-CNN 网络提高了 2.99%。可以看出对特征的提取是提高地砖缺陷检测准确率的重要性。从表 1 可以得出, 除了崩落这类缺陷检测精度略低于 ResNet101 分类模型的 Faster R-CNN 网络, 其他缺陷检测结果均表现了优良的结果。尤其是在检测完好地砖的结果中, 表现尤为突出。

本文网络训练过程中的损失(Loss)和检测精度(Accuracy)随迭代轮数(Epoch)变化曲线图如图 6 所示。

从图 6 中发现, 当训练达到 150 轮之后, 损失曲线和检测精度曲线趋于平稳。训练完成后, 对测试集进行测试, 检测的部分结果如图 7 所示。

3 结论

本文针对地砖缺陷目标尺度多变的复杂情况, 搭建了基于改进的 Faster R-CNN 对故宫地砖缺陷检测网

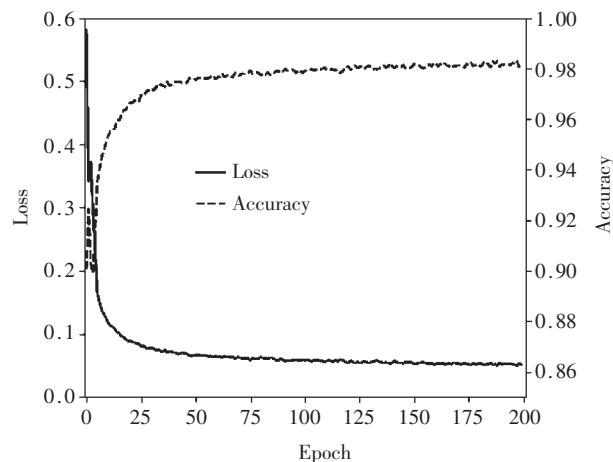


图 6 本文训练的损失和检测精度曲线

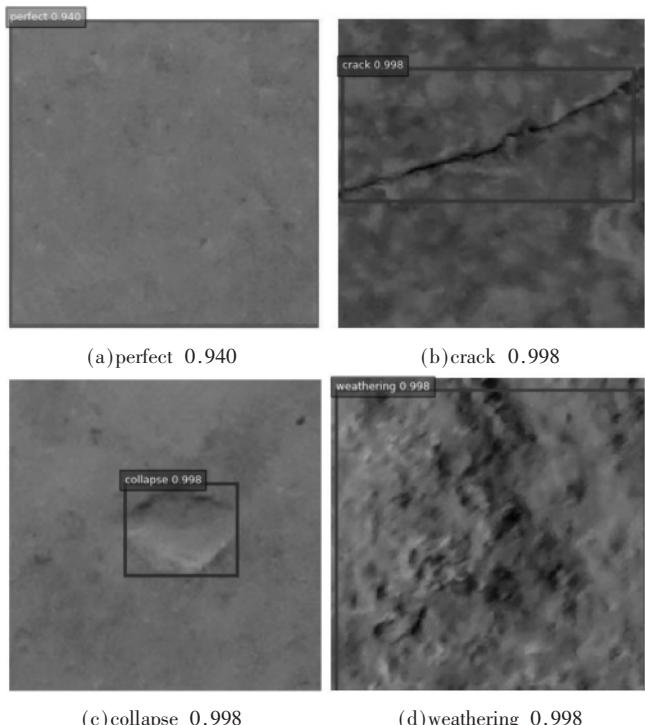


图 7 故宫地砖缺陷检测结果

络。加入可变形卷积网络, 对卷积核和池化核做增加偏移量操作, 脱离原始正方形的卷积核, 可以把卷积核更多聚集在缺陷的轮廓上, 提升对不规则目标的表现结果。实验表明, 相比于 VGG-16 和 ResNet101 分类模型与 Faster R-CNN 结合的网络, 本文网络显著提高了地砖缺陷像素的分类精度, 为地砖缺陷检测领域提供了

表 1 不同网络结构的检测性能

网络模型	AP/%					mAP/%
	perfect	abrasion	collapse	crack	weathering	
VGG16+Faster R-CNN	97.40	80.39	69.71	100.00	100.00	89.95
ResNet101+Faster R-CNN	88.24	90.91	81.44	100.00	95.96	91.31
本文网络	100.00	91.08	71.39	100.00	100.00	92.49

新的参考。

参考文献

- [1] 蔡明军, 聂高乾, 周辉, 等. 基于机器视觉的苹果缺陷识别自动分拣包装系统[J]. 自动化应用, 2018(8): 86–88, 96.
- [2] Ren Shaoqing, He Kaiming, ROSS G, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137–1149.
- [3] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.
- [4] 李策, 张栋, 杜少毅, 等. 一种迁移学习和可变形卷积深度学习的蝴蝶检测算法[J]. 自动化学报, 2019, 45(9): 1772–1782.
- [5] 王静波, 孟令军. 卷积神经网络人脸检测算法[J]. 电子技术应用, 2020, 46(1): 34–38.
- [6] 张润生, 贺超, 况朝青. 基于深度神经网络的伪装人脸识别[J]. 电子技术应用, 2020, 46(5): 27–30.
- [7] 付磊, 任德均, 胡云起, 等. 基于 ResNet 网络的医用塑瓶制造缺陷检测方法[J]. 计算机与现代化, 2020(4): 104–

108.

- [8] 张玉燕, 李永保, 温银堂, 等. 基于 Faster R-CNN 的金属点阵结构缺陷识别方法[J]. 兵工学报, 2019, 40(11): 2329–2335.
- [9] 郭梦浩, 徐红伟. 基于 Faster RCNN 的红外热图像热斑缺陷检测研究[J]. 计算机系统应用, 2019, 28(11): 265–270.
- [10] DAI J, QI H, XIONG Y, et al. Deformable convolution networks[C]. Proceeding of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu. IEEE, 2017: 764–773.
- [11] 汪梓艺, 苏育挺, 刘艳艳, 等. 一种改进 DeeplabV3 网络的烟雾分割算法[J]. 西安电子科技大学学报, 2019, 46(6): 52–59.

(收稿日期: 2020-06-25)

作者简介:

陈利(1994-), 女, 硕士, 主要研究方向: 人工智能图像处理。

刘艳艳(1978-), 通信作者, 女, 博士, 高级实验师, 主要研究方向: 集成电路设计、数字信号处理与模式识别, E-mail: lyytianjin@nankai.edu.cn。

(上接第 30 页)

- 汽车与配件, 2020(7): 68–71.
- [7] 王琰, 马欣, 郭维明. 智能驾驶汽车发展对汽车保险行业的影响[J]. 时代汽车, 2019(20): 103–104.
 - [8] 李梦静, 吉桂林. 视频行人重识别研究进展[J]. 南京师大学报(自然科学版), 2020, 43(2): 1–11.
 - [9] 刘辉, 余世干, 李林国. 数字视频行人检测算法综述[J]. 西安文理学院学报(自然科学版), 2020, 23(2): 58–61.
 - [10] 何松华, 章阳. 基于快速检测和 AdaBoost 的车辆检测[J].

计算机工程与设计, 2020(1): 203–207.

- [11] 黎達, 孙建伟, 孙德生. 一种基于后拉摄像头图像 ADAS 分析的倒车安全系统: CN110712591A[P]. 2020-01-21.
- [12] 赵津, 杜志彬, 张庆余, 等. 中国乘用车 ADAS 市场发展趋势浅析[J]. 时代汽车, 2019(2): 16–19.

(收稿日期: 2020-06-10)

作者简介:

孙德生(1980-), 男, 硕士研究生, 工程师, 主要研究方向: 嵌入式系统、智能电子产品设计和开发。



欢迎订阅, 2021年度 《信息技术与网络安全》

中国科技期刊数据库来源期刊

主管单位: 中国电子信息产业集团有限公司 主办单位: 华北计算机系统工程研究所 (中国电子信息产业集团有限公司第六研究所)

月刊 定价: 26 元 / 期

编辑部电话: (010) 66608908 66608981

电 话 订 阅: (010) 82306084

邮 局 订 阅: 邮发代号 82-417



版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所