

# 基于 Faster R-CNN 的道路裂缝识别\*

李太文, 范昕炜

(中国计量大学 质量与安全工程学院, 浙江 杭州 310000)

**摘要:** 传统的道路裂缝识别方法有基于 R-CNN、SPPnet、HOG+SVM 等多种方法, 但识别精度低、检测速度慢。针对这些缺点, 提出一种基于 Faster R-CNN 的道路裂缝识别方法。首先, 采集道路裂缝图像, 建立 Pascal VOC 数据集; 其次, 基于谷歌开发的 TensorFlow 深度学习框架, 用数据集对 Faster R-CNN 进行训练并分析各项性能参数指标。实验结果表明, 在迭代 20 000 次的情况下, 可将训练损失降到 0.188 5, AP 值达到 0.780 2, 取得了良好效果。

**关键词:** 机器学习; 深度学习; 卷积神经网络; 道路裂缝; Faster R-CNN; 图像识别

中图分类号: TN13

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.191412

中文引用格式: 李太文, 范昕炜. 基于 Faster R-CNN 的道路裂缝识别[J]. 电子技术应用, 2020, 46(7): 53-56, 59.

英文引用格式: Li Taiwen, Fan Xinwei. Road crevice recognition based on Faster R-CNN[J]. Application of Electronic Technique, 2020, 46(7): 53-56, 59.

## Road crevice recognition based on Faster R-CNN

Li Taiwen, Fan Xinwei

(School of Quality and Safety Engineering, China Jiliang University, Hangzhou 310000, China)

**Abstract:** Traditional road crack recognition methods are based on R-CNN, SPPnet, HOG+SVM and other methods, but the recognition accuracy is low and the detection speed is slow. In view of these shortcomings, a road crack recognition method based on Faster R-CNN is proposed. Firstly, road crack images were collected to build Pascal VOC data set. Secondly, the TensorFlow deep learning framework developed based on Google trains the Faster R-CNN with data sets and analyzes various performance parameters. The experimental results show that the training loss can be reduced to 0.188 5 and the AP value can reach 0.780 2 in the case of 20 000 iterations, achieving good results.

**Key words:** machine learning; deep learning; CNN; road cracks; Faster-RCNN; image recognition

### 0 引言

近几十年公路在中国得到蓬勃发展, 保养维护也日益成为一个问题, 需要定期对路面状况进行检查, 以便制定相应的维护策略, 其中重要的一项指标是路面裂缝。若能在裂缝的出现初期就能发现, 并及时跟踪它的发展情况, 那么它的维护费用将大大降低。如何在不影响正常的交通情况下对整段路面进行实时的监测, 成为亟待解决的一大难题。传统的基于人工视觉的识别方法越来越不能适应高速公路发展的要求, 其耗人力、耗时、危险、花费高、效率低, 还影响正常的交通。计算机高性能处理器、大容量存储器以及图像处理技术的快速发展, 使得路面裂缝的实时自动识别与识别技术成为可能。文献[1]提出基于改进 K-means 算法的不均匀光照下道路裂缝识别, 文献[2]对基于数字图像的混凝土道路裂缝识别方法进行了描述, 传统的裂缝目标识别算法有基于 SVM<sup>[3-4]</sup>、HOG<sup>[5]</sup>特征和 DPM<sup>[6]</sup>等多种方法, 但

些方法在识别过程中分多个阶段进行识别, 精度不高且检测速度慢。针对传统的裂缝目标识别方法存在的不足, 本文提出一种基于 Faster-RCNN<sup>[7]</sup>(Faster Region-Convolutional Neural Network)的道路裂缝识别方法, 不仅可以自动提取裂缝特征, 而且在识别精度和检测速度方面也取得了良好的效果。

### 1 数据采集与建立数据集

道路裂缝图像来自道路现场的拍照和网络获取, 包括不同的环境、光照、路面、形状, 一共 370 张, 以 9:1 的比例划分为训练数据集和测试数据集, 部分样本示例如图 1 所示。图像集使用 LabelImg 工具标记图片中的裂缝信息, 因为 Faster R-CNN 使用的是 Pascal VOC 数据集



图 1 道路裂缝样例

\* 基金项目: 国家重点研发计划项目(2017YFF0209702)

格式,所以按照 Pascal VOC 数据集格式制作道路裂缝图像数据集。

2 识别方法

Faster R-CNN 模型结构如图 2 所示,它包含特征提取、候选区域以及识别定位 3 部分。Faster R-CNN 用 RPN(Region Proposal Networks)网络代替 Fast R-CNN 中 SS (Selective Search)方法实现候选框的提取,相比于 RCNN<sup>[8]</sup> (Region-Convolutional Neural Network)和 Fast R-CNN<sup>[9]</sup> (Fast Region-Convolutional Neural Network), 无论是在识别精度还是识别速度上都有了很大的提升。

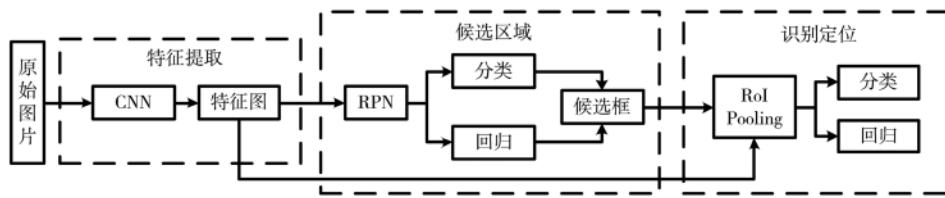


图 2 Faster-RCNN 结构图

2.1 基于 Faster R-CNN 的识别步骤

基于 Faster R-CNN 的识别步骤如下：

- (1)用 CNN(Convolutional Neural Network)<sup>[10-14]</sup> 提取原始图像中的特征生成道路裂缝特征图,该特征图分为两路,被后续的 RPN 层和 RoI Pooling 层所共享；
- (2)训练 RPN 网络,通过 softmax 分类 anchor(锚点)获得 foreground 和 background(识别目标是 foreground),并通过回归 anchor 得到候选框位置,Proposal 层综合 foreground 和候选框生成 proposals；
- (3)RoI Pooling 层输出固定尺寸的 RoI 特征图；
- (4)综合特征图与 RoI 特征图,判别 proposals 所属物体类别,并最终回归道路裂缝识别框的精确位置。

2.2 候选区域

如图 3 所示,可以看到 RPN 网络分为 2 条线,上面一条通过 softmax 分类 anchor 获得 foreground 和 background (识别目标是 foreground, background 是图像背景);下面一条用于计算对于 anchor 的 bounding box regression 偏移量。最后的 Proposal 层则负责综合 foreground anchor 和 bounding box regression 偏移量获取 proposals(候选区域),同时剔除太小和超出边界的 proposals。

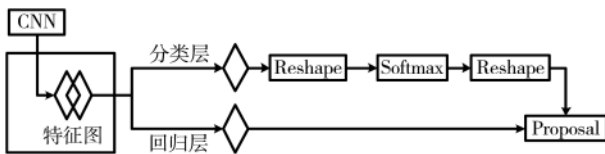


图 3 RPN 网络结构

2.3 识别定位

从图 2 中可以看到, RoI Pooling 层的输入来自特征图和候选框两部分,它输出固定尺寸的 RoI 特征图用作后续全连接层的输入。如图 4 所示,利 RoI Pooling 获得

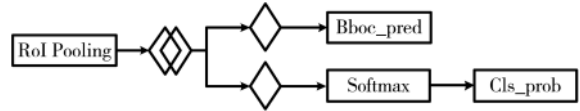


图 4 识别定位网络结构

的 RoI 特征图,通过 full connect 层与 softmax 计算每个 proposal 具体属于哪个物体类别,输出类别(cls\_prob)概率向量;同时再次利用 bounding box regression 获得每个 proposal 的位置偏移量 bbox\_pred,用于回归更加精确的目标识别框。

3 实验仿真及结果分析

实验环境为 Windows 10(64 位)的 Intel COREi5,显卡是 NVIDIA GTX60M,主频 3.7 GHz,内存 8 GB, IDE 是 pycharm,编程语言是 Python。基于谷歌开发的 TensorFlow 深度学习框架,使用建立的 Pascal

VOC 数据集训练 Faster R-CNN 模型。

3.1 训练 RPN 网络

用预训练好的 VGG-16<sup>[15]</sup>深度学习模型提取道路裂缝图像中的特征生成特征图,初始化 RPN 网络。RPN 网络分为分类层和回归层 2 条线,在分类层中,使用 softmax 分类器对 anchor 进行 foreground 和 background 判断,保留识别目标 foreground,丢弃 background;在回归层中,通过调整 anchor 的中心坐标与长宽,拟合出候选框位置,Proposal 网络负责综合 foreground 和候选框位置生成 proposals(候选区域)。在训练过程中,batch\_size 取 1 幅道路裂缝图像,根据 IoU(预测框与真实框交并比值),在该图像中随机采样 128 个 positive anchor 和 128 个 negative anchor。IoU 大于 0.7 的 anchor 标记为 positive anchor,即 foreground;IoU 小于 0.3 的标记为 negative anchor,即 background;0.3 < IoU < 0.7 的 anchor 则不参与训练。对 positive anchor 标签为 1,negative anchor 标签为 0。这些带标签的样本和标注框的坐标被用于 RPN 的有监督的训练。在训练过程中,RPN 网络的总损失函数计算如下:

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

$$L_{cls}(\{p_i\}, \{p_i^*\}) = -\log[p_i p_i^* + (1-p_i)(1-p_i^*)] \quad (2)$$

$$L_{reg}(t_i, t_i^*) = \sum_i \text{smooth}_{L_1}(t_i - t_i^*) \quad (3)$$

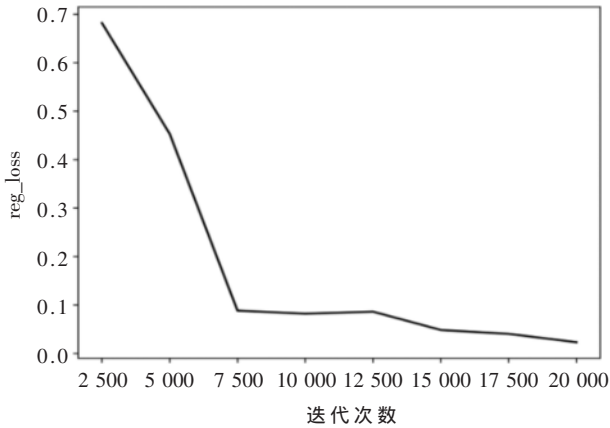
$$\text{smooth}_{L_1}(t_i - t_i^*) = \begin{cases} 0.5(t_i - t_i^*) & |t_i - t_i^*| < 1 \\ |t_i - t_i^*| - 0.5 & \text{其他} \end{cases} \quad (4)$$

$$p_i^* = \begin{cases} 1 & \text{IoU} > 0.7 \\ 0 & \text{IoU} < 0.3 \end{cases} \quad (5)$$

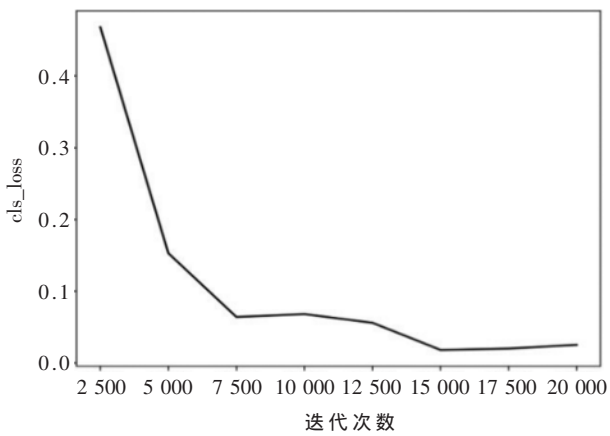
式中,  $i$  表示 anchor 索引,  $p_i$  表示 foreground 概率,  $p_i^*$  表示

anchor 判别值,  $N_{cls}$  为分类数量,  $N_{reg}$  为回归数量,  $t_i$  为预测边界框坐标向量,  $t_i^*$  为 foreground anchor 对应的真实边界框坐标向量,  $\lambda$  为权重参数。

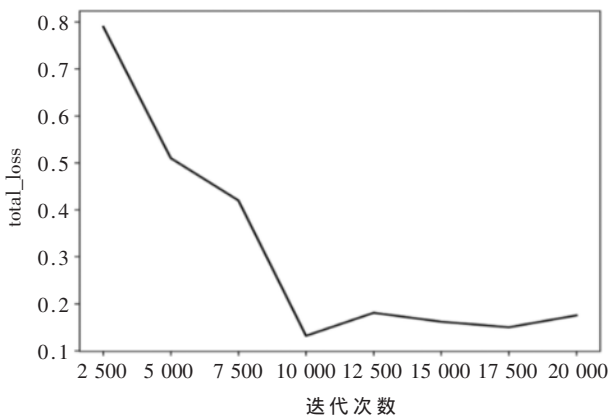
可以看到, 整个损失函数分为分类层损失函数与回归层损失函数 2 部分。图 5(a)、图 5(b)、图 5(c) 所示为在迭代 20 000 次的情况下各个损失的变化曲线,  $x$  轴代表迭代次数,  $y$  轴分别是训练过程中分类损失、回归损失和总损失, 可以看到, 各类损失都是一个很小的值, 模型收敛良好。表 1 展示了训练完成后各类损失结果。



(a) 回归损失



(b) 分类损失



(c) 总损失

图 5 各类损失图

表 1 各类损失结果

性能评估参数	训练 20 000 次结果
分类层损失	0.025 5
回归层损失	0.027 9
总损失	0.188 5

### 3.2 试验评价指标

目标识别中衡量识别精度的指标是 mAP (mean Average Precision), mAP 是多个类别 AP 的平均值。由于只有路面裂缝这一个类别, 因此采用 AP (Average Precision) 作为目标识别的评价指标。AP 体现模型识别效果的优劣程度, 其值越大效果越好, 反之越差。根据 Recall (召回率) 和 Precision (准确率) 绘制一条曲线, AP 就是该曲线下的面积, 即 AP 值是对精确率-召回率曲线进行积分, 积分公式为:

$$P = \frac{T_p}{T_p + F_p} \quad (6)$$

$$R = \frac{T_p}{T_p + F_N} \quad (7)$$

$$AP = \int_0^1 PRdR \quad (8)$$

式中,  $P$  表示精确率,  $R$  表示召回率,  $T_p$  表示被识别的含裂缝的正类样本数,  $F_p$  表示被识别的不含裂缝的负样本数,  $F_N$  表示不被识别的含裂缝正类样本数。

图 6 所示为迭代 20 000 次的情况下 AP 值的变化曲线, 在前 5 000 次 AP 值提高得很快, 随后趋于平稳, 最终的 AP 值是 0.780 2, 平均检测速度为 0.43 s。

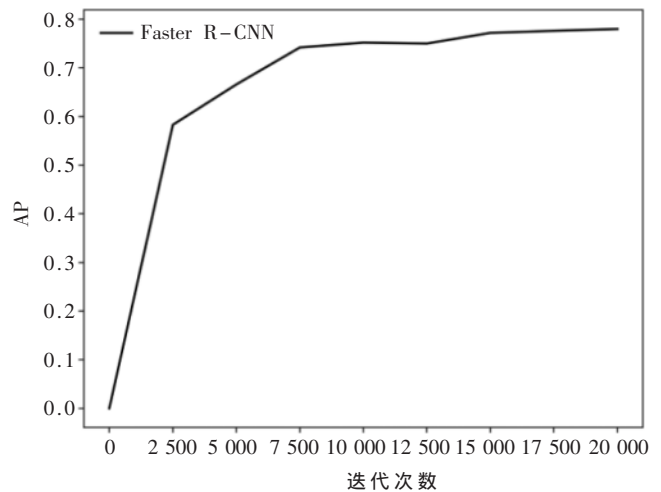


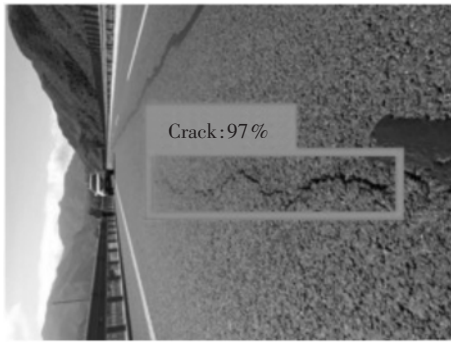
图 6 Faster-RCNN AP 值

### 3.3 测试结果

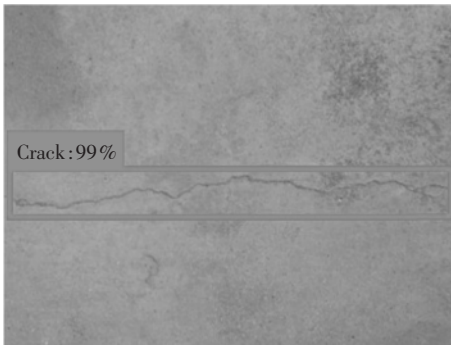
使用训练好的 Faster R-CNN 模型对道路裂缝进行识别, 如图 7 所示, 在不同的光照、不同路面、不同裂缝的图像上测试都取得了很好的识别效果。

### 3.4 不同识别方法比较

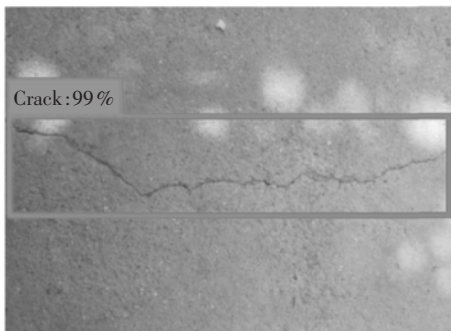
R-CNN、SPPnet<sup>[16]</sup>、Fast R-CNN 都是采用 SS 方法生成候选区域, 为了验证本方法的识别优势, 在相同实验



(a)道路裂缝 1



(b)道路裂缝 2



(c)道路裂缝 3

图 7 识别结果

条件下,输入一张道路裂缝图像,将 Faster R-CNN 与以上 3 种目标识别算法做了比较,如图 8 所示,比较结果如表 2 所示。

通过对比可以看出,Faster R-CNN 无论是在 AP 值还是检测速度方面都要更好,这是因为 R-CNN、SPPnet、Fast R-CNN 都是采用 SS 方法生成候选区域,Faster R-CNN 是用 RPN 网络生成候选区域。

#### 4 结论

经过 20 000 次训练后,最终的总损失降到了 0.188 5, AP 值为 0.780 2,在识别路面裂缝时取得了良好的效果。该方法同样适用于桥梁、隧道、墙体裂缝识别,为其他方面的裂缝检测提供了一种新的方法。下一步的研究工作是继续改善网络结构和参数,提高识别 AP 值。

#### 参考文献

- [1] 王德方,曾卫明,王倪传.基于改进 K-means 算法的不均匀光照下道路裂缝检测[J].计算机应用与软件,2015,32

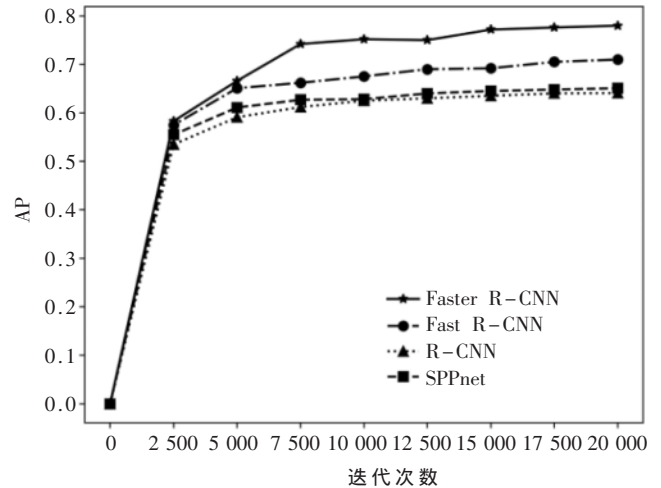


图 8 各类 AP 值比较

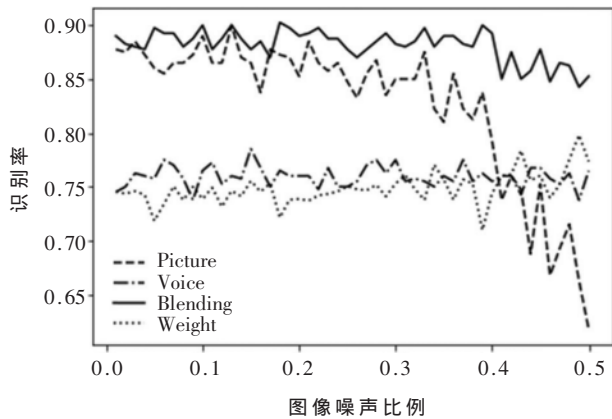
表 2 不同方法对比

目标识别方法	AP 值	检测速度/s
R-CNN	0.651 0	51.27
SPPnet	0.682 3	46.32
Fast R-CNN	0.711 6	0.91
Faster R-CNN	0.780 2	0.43

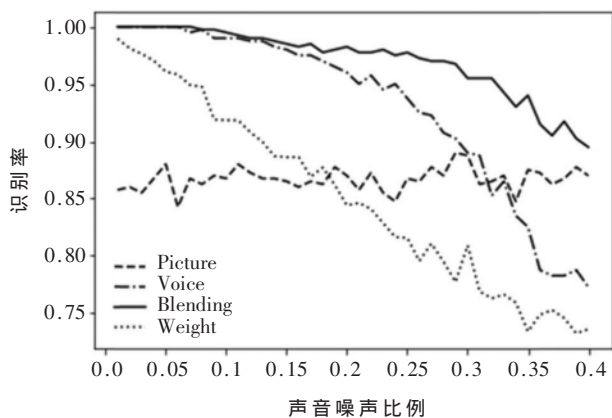
(7): 244-247, 300.

- [2] 蒋文波,罗秋容,张晓华.基于数字图像的混凝土道路裂缝检测方法综述[C].西华大学学报(自然科学版),2018,37(1):75-84.
- [3] 胡涛,胡军,郭杭.基于 SVM 邻域学习的视频目标检测方法[J].现代电子技术,2017,40(14):95-98.
- [4] 熊昊,李伟.基于 SVM 的视频火焰检测算法[J].传感器与微系统,2020,39(1):143-149.
- [5] DARAL N.Histograms of oriented gradients for human detection[C].Proceedings of CVPR, 2005.
- [6] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al.Object detection with discriminatively trained partbased models[J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.
- [7] 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, 2015, 39(6): 1137-1149.
- [8] GIRSHICK R, DONAHUE J, DARRELL T, et al.Rich feature hierarchies for accurate object detection and semantic segmentation[C].Conference on Computer Vision and Pattern Recognition(CVPR).IEEE, 2014: 580-587.
- [9] GIRSHICK R.Fast R-CNN[C].International Conference on Computer Vision(ICCV), IEEE, 2015: 1440-1448.
- [10] LE CUN Y, BENGIO Y.Word-level training of a hand-written word recognizer based on convolutional neural networks[C].Proceedings of the International Conference

(下转第 59 页)



(a)不同图像噪声



(b)不同声音噪声

图4 不同噪声识别率变化图

定性也降低,但仍强于人脸识别系统。

综合图4(a)、图4(b)可知,融合后的识别效果明显要强于任何一种单模态的识别效果,并且具有较高的识别率。

## 5 结论

本文针对在单一模态受噪声影响下识别率下降问

题,提出对语音、图像信息分别进行提取特征值和计算匹配分数,然后采用自适应加权融合算法选取最优匹配权值,得出最优决策的方法,并进行实验验证。经实验证明,高斯混合模型的声音识别稳定性远远强于SVM分类的人脸识别;而且在噪声背景下,采用该匹配层融合的方法远远优于任何一种单模态的识别效果,同时也证明了采用多模态的融合识别将是身份识别的重要发展方向。

## 参考文献

- [1] 孙贵华,陈淑荣.一种改进的Refine多尺度人脸检测方法[J].电子技术应用,2019,45(8):34-39.
- [2] 范叶平,李玉.基于深度集成学习的人脸智能反馈认知方法[J].电子技术应用,2019,45(5):5-8,13.
- [3] 余成波,李彩虹,曾亮.K-means指纹定位的优化算法[J].电子技术应用,2018,44(2):70-74.
- [4] 柳欣,耿佳佳.多生物特征融合发展现状及其展望[J].小型微型计算机系统,2017,38(8):1792-1799.
- [5] 陈倩.多生物特征融合身份识别研究[D].杭州:浙江大学,2007.
- [6] 柯晶晶,周萍.差分和加权Mel倒谱混合参数应用于说话人识别[J].微电子学与计算机,2014,31(9):88-91.
- [7] 周萍,沈昊.基于MFCC与GFCC混合特征参数的说话人识别[J].应用科学学报,2019,37(1):24-32.
- [8] 张庶,李子月,刘玉超.基于PCA与SVM的人脸识别技术[J].指挥与控制学报,2019,5(3):249-253.
- [9] 刘艳丽,王铁建.SVM算法在人脸识别中的应用研究[J].电脑知识与技术,2017,13(18):176-177.

(收稿日期:2019-12-30)

## 作者简介:

李傲梅(1972-),女,博士,教授,主要研究方向:信息与通信工程、电子工程。

胡正豪(1992-),通信作者,男,硕士研究生,主要研究方向:智能控制、模式识别,E-mail:289449169@qq.com。

周川川(1990-),男,硕士研究生,主要研究方向:电子工程、智能控制。

(上接第56页)

on Pattern Recognition.IEEE,1994.

- [11] KRIZHEVSKY A,SUTSKEVER I,HINTON G E.Imagenet classification with deep convolutional neural networks[J].Advances in Neural Information Processing Systems,2012,25(2):1106-1114.
- [12] ZEILER M D,FERGUS R.Visualizing and understanding convolutional networks[J].ECCV,2014,8689(53):818-833.
- [13] 周飞燕,金林鹏,董军.卷积神经网络研究综述[J].计算机学报,2017,40(6):1229-1251.
- [14] SZEGEDY C,Liu Wei,Jia Yangqing,et al.Going deeper with convolutions[C].Conference on Computer Vision and Pattern Recognition(CVPR).IEEE,2015.

- [15] SIMONYAN K,ZISSERMAN A.Very deep convolutional networks for large-scale image recognition[C].LCLR2015,2015.

- [16] He Kaiming,Zhang Xiangyu,Ren Shaoqing,et al.Spatial pyramid pooling in deep convolutional networks for visual recognition[C].ECCV 2014,European Conference on Computer Vision,2014:346-361.

(收稿日期:2019-12-25)

## 作者简介:

李太文(1994-),男,硕士,主要研究方向:深度学习、模式识别。

范昕炜(1978-),男,博士,高级工程师,主要研究方向:机器人、模式识别与智能控制、数据挖掘。

## 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