



中国图形学报

ISSN1006-8961
CN11-3758/TB



第27卷第8期（总第316期）
2022年8月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

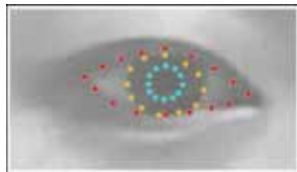
Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

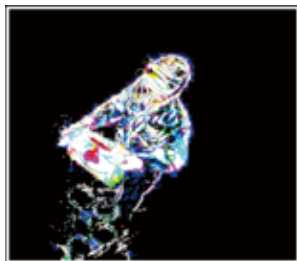
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

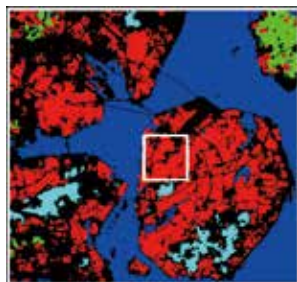
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



自然光普通摄像头的眼部分割及特征点定位数据集ESLD (第2329页)



双视图三维卷积网络的工业装箱行为识别(第2368页)



多源特征自适应融合网络的高分遥感影像语义分割(第2516页)

综述

虚拟现实图像客观质量评价研究进展

周玉, 汪一, 李雷达, 高陈强, 卢兆林 2313

数据集论文

自然光普通摄像头的眼部分割及特征点定位数据集ESLD

张俊杰, 孙光民, 郑鲲, 李煜, 付晓辉, 慈康怡, 申俊杰, 孟凡超, 孔江萍, 张玥 2329

图像处理和编码

双尺度顺序填充的深度图像修复

陈东岳, 朱晓明, 马腾, 宋园园, 贾同 2344

二维码位流长度最小化算法

袁泰凌, 徐昆 2356

图像分析和识别

双视图三维卷积网络的工业装箱行为识别

胡海洋, 潘健, 李忠金 2368

面向航拍图像中工程车辆检测与识别的改进胶囊网络

钟映春, 郑海阳, 张文祥, 王波, 罗志勇 2380

用于骨架行为识别的多维特征融合注意力机制

姜权晏, 吴小俊, 徐天阳 2391

高性能整数倍稀疏网络行为识别研究

臧影, 刘天娇, 赵曙光, 杨东升 2404

面向工业零件分拣系统的低纹理目标检测

闫明, 陶大鹏, 普园媛 2418

融合策略优选和双注意力的单阶段目标检测

戴坤, 许立波, 黄世旸, 李鉴铃 2430

融合视觉词与自注意力机制的视频目标分割

季传俊, 陈亚当, 车洵 2444

基于Gaussian-Hermite矩的旋转运动模糊不变量

郭锐, 贾丽, 郝宏翔, 墨瀚林, 李华 2458

图像理解和计算机视觉

RGB-D语义分割: 深度信息的选择使用

赵经阳, 余昌黔, 桑农 2473

自纠正噪声标签的人脸美丽预测

甘俊英, 吴必诚, 翟懿奎, 何国辉, 麦超云, 白振峰 2487

面向非对称特征注意力和特征融合的太赫兹图像检测

曾文健, 朱艳, 沈韬, 曾凯, 刘英莉 2496

面向纹理平滑的方向性滤波尺度预测模型

林俊彦, 刘春晓, 章金凯, 李泓易 2506

遥感图像处理

多源特征自适应融合网络的高分遥感影像语义分割

张文凯, 刘文杰, 孙显, 许光奎, 付琨 2516

结合上下文编码与特征融合的SAR图像分割

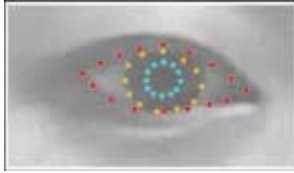
范艺华, 董张玉, 杨学志 2527

一阶全卷积遥感影像倾斜目标检测

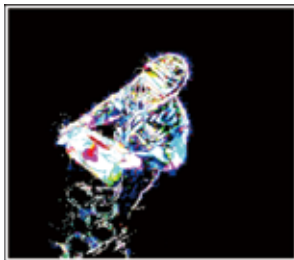
周院, 杨庆庆, 马强, 薛博维, 孔祥楠 2537

CONTENTS

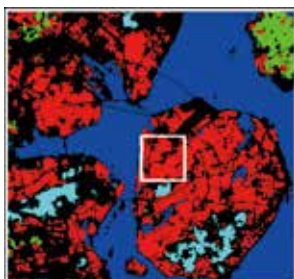
JOURNAL OF IMAGE AND GRAPHICS



ESLD: eyes segment and landmark detection in the wild (P2329)



Dual-view 3D ConvNets based industrial packing action recognition(P2368)



Multi-source features adaptation fusion network for semantic segmentation in high-resolution remote sensing images(P2516)

Review

Research progress in objective quality evaluation of virtual reality images

Zhou Yu, Wang Yi, Li Leida, Gao Chenqiang, Lu Zhaolin 2313

Dataset

ESLD: eyes segment and landmark detection in the wild

Zhang Junjie, Sun Guangmin, Zheng Kun, Li Yu, Fu Xiaohui, Ci Kangyi, Shen Junjie, Meng Fan-
chao, Kong Jiangping, Zhang Yue 2329

Image Processing and Coding

Depth image recovery based on dual-scale sequential optimized filling

Chen Dongyue, Zhu Xiaoming, Ma Teng, Song Yuanyuan, Jia Tong 2344

Minimization of bit stream length of QR codes

Yuan Tailing, Xu Kun 2356

Image Analysis and Recognition

Dual-view 3D ConvNets based industrial packing action recognition

Hu Haiyang, Pan Jian, Li Zhongjin 2368

Improved capsule network method for engineering vehicles detection and recognition in aerial ima-
ges

Zhong Yingchun, Zheng Haiyang, Zhang Wenxiang, Wang Bo, Luo Zhiyong 2380

M2FA: multi-dimensional feature fusion attention mechanism for skeleton-based action recognition

Jiang Quanyan, Wu Xiaojun, Xu Tianyang 2391

Action recognition analysis derived of integer sparse network

Zang Ying, Liu Tianjiao, Zhao Shuguang, Yang Dongsheng 2404

Texture-less object detection method for industrial components picking system

Yan Ming, Tao Dapeng, Pu Yuanyuan 2418

Single stage object detection algorithm based on fusing strategy optimization selection and dual
attention mechanism

Dai Kun, Xu Libo, Huang Shiyang, Li Yunling 2430

Visual words and self-attention mechanism fusion based video object segmentation method

Ji Chuanjun, Chen Yadang, Che Xun 2444

Rotational motion blur invariants based on Gaussian-Hermite moments

Guo Rui, Jia Li, Hao Hongxiang, Mo Hanlin, Li Hua 2458

Image Understanding and Computer Vision

RGB-D semantic segmentation: depth information selection

Zhao Jingyang, Yu Changqian, Sang Nong 2473

Self-correcting noise labels for facial beauty prediction

Gan Junying, Wu Bicheng, Zhai Yikui, He Guohui, Mai Chaoyun, Bai Zhenfeng 2487

Terahertz image detection combining asymmetric feature attention and feature fusion

Zeng Wenjian, Zhu Yan, Shen Tao, Zeng Kai, Liu Yingli 2496

Texture-smoothing-oriented directional filtering scales-predicting model

Lin Junyan, Liu Chunxiao, Zhang Jinkai, Li Hongyi 2506

Remote Sensing Image Processing

Multi-source features adaptation fusion network for semantic segmentation in high-resolution re-
mote sensing images

Zhang Wenkai, Liu Wenjie, Sun Xian, Xu Guangluan, Fu Kun 2516

The integrated contextual encoding and feature fusion SAR images segmentation method

Fan Yihua, Dong Zhangyu, Yang Xuezhi 2527

Improved one-stage fully convolutional network for oblique object detection in remote sensing ima-
gery

Zhou Yuan, Yang Qingqing, Ma Qiang, Xue Bowei, Kong Xiangnan 2537

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2022)08-2487-09

论文引用格式: Gan J Y, Wu B C, Zhai Y K, He G H, Mai C Y and Bai Z F. 2022. Self-correcting noise labels for facial beauty prediction. Journal of Image and Graphics, 27(08): 2487-2495 (甘俊英, 吴必诚, 翟懿奎, 何国辉, 麦超云, 白振峰. 2022. 自纠正噪声标签的人脸美丽预测. 中国图象图形学报, 27(08): 2487-2495) [DOI: 10.11834/jig.210125]

自纠正噪声标签的人脸美丽预测

甘俊英*, 吴必诚, 翟懿奎, 何国辉, 麦超云, 白振峰

五邑大学智能制造学部, 江门 529020

摘要: 目的 人脸美丽预测是研究如何使计算机具有与人类相似的人脸美丽判断或预测能力, 然而利用深度神经网络进行人脸美丽预测存在过度拟合噪声标签样本问题, 从而影响深度神经网络的泛化性。因此, 本文提出一种自纠正噪声标签方法用于人脸美丽预测。方法 该方法包括自训练教师模型机制和重标签再训练机制。自训练教师模型机制以自训练的方式获得教师模型, 帮助学生模型进行干净样本选择和训练, 直至学生模型泛化能力超过教师模型并成为新的教师模型, 并不断重复该过程; 重标签再训练机制通过比较最大预测概率和标签对应预测概率, 从而纠正噪声标签。同时, 利用纠正后的数据反复执行自训练教师模型机制。结果 在大规模人脸美丽数据库 LSFBD (large scale facial beauty database) 和 SCUT-FBP5500 数据库上进行实验。结果表明, 本文方法在人工合成噪声标签的条件下可降低噪声标签的负面影响, 同时在原始 LSFBD 数据库和 SCUT-FBP5500 数据库上分别取得 60.8% 和 75.5% 的准确率, 高于常规方法。结论 在人工合成噪声标签条件下的 LSFBD 和 SCUT-FBP5500 数据库以及原始 LSFBD 和 SCUT-FBP5500 数据库上的实验表明, 所提自纠正噪声标签方法具有选择干净样本学习、充分利用全部数据的特点, 可降低噪声标签的负面影响, 能在一定程度上降低人脸美丽预测中噪声标签的负面影响, 提高预测准确率。

关键词: 深度学习; 噪声标签; 人脸美丽预测; 特征分类; 深度神经网络

Self-correcting noise labels for facial beauty prediction

Gan Junying*, Wu Bicheng, Zhai Yikui, He Guohui, Mai Chaoyun, Bai Zhenfeng

Department of Intelligent Manufacturing, Wuyi University, Jiangmen 529020, China

Abstract: **Objective** Human facial beauty prediction is the research on how to make computers have the ability to judge or predict the beauty of faces similar to humans. However, deep neural networks based facial beauty prediction has challenged the issue of noisy label samples affecting the training of deep neural network models, which thus affects the generalizability of deep neural networks. Noisy labels are mislabeled in the database, which usually affect the training of deep neural network models, thus reduce the generalizability of deep neural networks. To reduce the negative impact of noisy labels on deep neural networks in facial beauty prediction, a self-correcting noisy label method was proposed, which has the features of selection of clean samples for learning and full utilization of all data. **Method** Our method is composed of a self-training teacher model mechanism and a re-labeling retraining mechanism. First, two deep convolutional neural networks (CNNs)

收稿日期: 2021-03-15; 修回日期: 2021-05-21; 预印本日期: 2021-05-28

* 通信作者: 甘俊英 junyinggan@163.com

基金项目: 国家自然科学基金项目(6177010044); 广东省基础与应用基础研究基金项目(2019A1515010716); 广东省普通高校基础研究与应用基础研究重点项目(2018KZDXM073)

Supported by: National Natural Science Foundation of China (6177010044); Basic and Applied Basic Research Fund of Guangdong Province (2019A1515010716); Key Projects of Basic and Applied Basic Research in General Universities of Guangdong Province (2018KZDXM073)

are initialized with the same structure simultaneously, and the network is used as the teacher model with stronger generalization ability, while the other network is used as the student model. The teacher model can be arbitrarily specified during initialization. Second, small batches of training data are fed to the teacher and student models both at the input side together. The student model receives the sample number and finds the corresponding sample and label for back-propagation training until the generalization ability of the student model exceeds that of the teacher model. Then, the student model shares the optimal parameters to the teacher model, i. e., the original student model becomes the new teacher model, where it is called the self-training teacher model mechanism. After several iterations of training, small batches of data are fed into the teacher model with the strongest generalization ability among all previous training epochs, and its prediction probability of each category is calculated. If the maximum output probability predicted by the teacher model for this data is higher than a certain threshold of the corresponding output probability of the label, it is considered that the sample label should be corrected. The self-training teacher model mechanism is then iteratively executed utilizing the corrected data, where the process above is called the relabeling retraining mechanism. Finally, the teacher model is output as the final model. **Result** The ResNet-18 model pre-trained on the ImageNet database is used as the backbone deep neural network, which is regarded as a baseline method with cross entropy as the loss function. The experiments on the large scale facial beauty database (LSFBD) and SCUT-FBP5500 database are divided into two main parts as mentioned below: 1) the first part is performed under synthetic noise label conditions, i. e., 10%, 20%, and 30% of the training data are selected from each class of facial beauty data on the two databases mentioned above, while their labels are randomly changed. The accuracy of the method in this paper exceeds the baseline method by 5.8%, 4.1% and 3.7% on the LSFBD database at noise rates of 30%, 20% and 10%, respectively. The accuracy exceeds the baseline method by 3.1%, 2.8%, and 2.5% on the SCUT-FBP5500 database, respectively. Therefore, it is demonstrated that our method can reduce the negative impact of noisy labels under synthetic noisy label conditions. 2) The second part is carried out on the original LSFBD database and the original SCUT-FBP5500 database, and our method exceeded the prediction accuracy of the baseline method by 2.7% and 1.2% on the original LSFBD database and the original SCUT-FBP5500 database, respectively. Therefore, our demonstrated illustration can reduce the negative impact of noise labels under the original data conditions. **Conclusion** Our proposed method of self-correcting noise labels can reduce the negative impact of noise label in human facial beauty prediction in some extent and improve the prediction accuracy based on the LSFBD and SCUT-FBP5500 databases under synthetic noisy label circumstances, the original LSFBD and SCUT-FBP5500 facial beauty databases, respectively.

Key words: deep learning; noise labels; facial beauty prediction; characteristics classification; deep neural networks

0 引言

噪声标签是指标注错误的标签,通常会降低深度神经网络的泛化性(Zhang等,2017;Arpit等,2017)。为避免噪声标签对深度神经网络的负面影响,针对噪声标签数据库的有监督分类成为一个重要研究方向(Algan和Ulusoy,2021a)。Wang等人(2019)提出对称交叉熵学习的方法,添加反向交叉熵作为损失函数,避免了仅使用交叉熵作为损失函数造成深度网络过度拟合噪声标签样本影响泛化。Xu等人(2019)提出一种新型的信息理论损失函数,这是第一个证明对与实例无关的标签噪声具有鲁棒性的损失函数,无论噪声模式如何,均可直接应用于任何现有的分类神经网络,而无需辅助信息。

Liu等人(2020)提出利用半监督学习技术产生模型输出的目标概率,并设计一个正则化项防止深度神经网络记忆噪声标签。Algan和Ulusoy(2021b)提出在有噪声标签的情况下利用元学习训练深度网络。首先根据元目标寻求最佳的标签分布,即最小化小元数据集的损失;然后利用元软标签生成框架,生成预测的软标签训练网络;最后在训练过程中不断重复上述两个阶段,从而避免噪声标签的负面影响。Wang等人(2019)、Xu等人(2019)、Liu等人(2020)、Algan和Ulusoy(2021b)利用设计鲁棒性损失函数、正则化器或者利用元学习的方法避免噪声标签的负面影响,所提方法没有通过建立噪声模型估计噪声结构。

深度学习模型一开始会记忆简单样本,随着训练次数增加逐渐拟合到困难样本(Arpit等,2017)。

当存在噪声样本时,深度学习网络甚至会记忆这些错误的标签从而影响泛化(Zhang等,2017)。Han等人(2018)、Guo等人(2018)和Li等人(2020)从含有噪声标签的数据库中选择出干净样本,并用其训练深度学习网络,从而隐性地建立噪声模型避免噪声标签的负面影响。Han等人(2018)提出一种互相教学的策略训练一个鲁棒性的深度学习网络。该策略同时训练两个深度卷积网络,训练前期一般拟合的是简单样本,两个网络互相选择损失值小的样本给对方进行训练,从而防止拟合到噪声标签样本。Guo等人(2018)提出在大规模弱监督的网络图像上训练深度神经网络,利用课程学习,先使模型从简单样本开始学习,逐渐进阶到复杂样本和知识,从而提高模型的泛化能力并降低噪声标签的负面影响。Li等人(2020)提出 DivideMix 框架,利用半监督学习和干净样本选择,避免噪声标签负面影响。首先用混合模型对每个样本损失分布进行建模,然后将训练数据动态地划分为具有干净样本的标签集和具有噪声样本的无标签集,接着以半监督的方式在有标签和无标签数据上训练模型。为了避免偏差,同时训练了两个网络,每个网络都使用另一个网络的数据集划分。最后在半监督训练阶段,通过不断对有标签和无标签样本分别进行标签的互相精细和互相猜测来改进动态划分策略。陈庆强等人(2019)提出一种基于数据分布的标签噪声过滤方法。首先对于数据集中的每一个样本,根据其近邻内样本的分布,将其及邻域样本形成的区域划分为高密度区域和低密度区域,然后针对不同的区域采用不同的噪声过滤规则进行过滤。张增辉等人(2021)提出一种基于标签置信度分布的局部概率抽样方法进行标签噪声过滤。首先利用随机森林分类器对样本的标签进行投票,从而获取每个样本的标签置信度;然后根据标签置信度的大小,将样本划分为易识别样本和难识别样本;最后分别采用不同的过滤策略对样本进行过滤。

若只选择干净样本进行学习,未能充分利用全部数据。Shu等人(2019)、Wang等人(2020)、唐轲和郎丛妍(2021)利用样本重要性加权策略,即首先设计一个从训练损失到样本权重的加权函数映射,然后利用该加权函数对不同样本的重要性加权并影响分类器的输出,从而避免深度网络过度拟合噪声标签样本。Shu等人(2019)提出利用具有一个隐藏

层的多层感知器作为加权函数,该加权函数能直接从数据中自适应地学习样本重要性权重。Wang等人(2020)提出了一种自校正网络,通过排名正则化对每个训练样本的重要性进行加权,防止深度网络过度拟合噪声标签的人脸表情图像,并重新标注噪声标签样本,从而提高深度网络泛化能力。为解决行人再识别的噪声标签问题,唐轲和郎丛妍(2021)结合噪声、非噪声数据训练差异化特征,提出一种噪声标签自适应的行人再识别方法,不需要使用额外的验证集以及噪声比例、类型等先验信息,完成对噪声数据的筛选过滤。此外,该方法可以自适应地学习噪声样本权重,进一步降低噪声影响。

无论是 Han 等人(2018)方法中的样本选择,还是 Wang 等人(2020)方法中的样本重要性加权,其有效的共性在于,计算深度神经网络输出损失值时,简单样本对应的损失值一般小于噪声标签样本对应的损失值。该假定的不足在于,困难样本对应的损失值一般大于简单样本对应的损失值。因此,将深度神经网络输出的损失值作为判断噪声标签样本依据,不利于困难样本的学习。判断噪声标签样本能力更强的深度网络可以帮助区分噪声标签样本(Jiang等,2020),且自训练的方式有助于获得噪声标签样本判断能力更强的深度网络(Xie等,2020)。为此,本文提出一种自纠正噪声标签方法,能稳定选择干净样本学习,并重新标注噪声标签样本,从而充分利用全部数据。

人脸美丽预测任务更易受到噪声标签影响。首先,客观评价标准不明确,在进行人工标签标注时容易受标注人员的影响,从而产生标注错误,即容易产生更多噪声标签。其次,受人力、物力等客观因素影响,人脸美丽数据库包含的标签数据样本有限,从而使数据在表达差异性特征能力上存在不足,进而更易受噪声标签的干扰。最后,不同类别人脸美丽等级的人脸图像之间特征相似、抽象且难以提取,故深度神经网络更难以区分噪声标签和真实标签进而降低泛化能力。因此,本文使用自纠正噪声标签方法对人脸美丽进行预测。

本文提出的自纠正噪声标签方法包括自训练教师模型机制和重标签再训练机制。自训练教师模型机制首先利用 Han 等人(2018)方法中的样本选择方法,自训练出噪声标签样本判断能力更强的模型,以此作为教师模型。然后,借助教师模型的噪声标

签样本判断能力帮助学生模型学习,直至训练出噪声标签判断能力更强的学生模型,并将其作为新的教师模型不断重复该过程。最后,重标签再训练机制利用 Wang 等人 (2020) 方法中的重标签机制,重新标注潜在的噪声标签并引入自训练教师模型机制再训练。

本文主要贡献如下:1) 首次提出将人脸美丽预测问题理解成噪声标签问题,并用实验表明了该假设成立。2) 提出一种自纠正噪声标签方法,能一定程度降低噪声标签对人脸美丽深度卷积预测网络的负面影响。3) 在人工合成噪声标签 LSFBD (large scale facial beauty database) 数据库、原始 LSFBD (甘俊英 等,2019) 数据库和人工合成噪声标签 SCUT-FBP5500 数据库、原始 SCUT-FBP5500 数据库 (Liang 等,2018) 上,验证了自纠正噪声标签方法的有效性。对比实验表明,所提方法每个环节皆能降低噪声标签对人脸美丽深度网络的影响。

1 本文方法

为了降低噪声标签对深度神经网络的负面影响,本文提出了基于深度卷积神经网络的自纠正噪声标签方法,包括自训练教师模型机制和重标签再训练机制。图 1 为自纠正噪声标签方法框图。首先,同时初始化两个结构相同的深度卷积神经网络,并将泛化能力较强的网络作为教师模型,另一个网络作为学生模型。初始化时可任意指定教师模型。其次,在输入端同时输入小批次训练数据给教师、学生模型。其中,教师模型自身不进行反向传播训练,仅用于选择干净标签样本序号给学生模型;学生模型接收该样本序号并找到对应样本和标签进行反向传播训练,直至学生模型泛化能力超过教师模型。此时,学生模型共享最优参数给教师模型,即原学生模型成为新的教师模型,称之为自训练教师模型机制。经反复训练后,将小批次数据输入教师模型中,计算其每类预测概率值;将其输入到重标签模块中进行标签纠正。最后,利用纠正后的数据再反复执行自训练教师模型机制,并将教师模型作为最终模型输出。

1.1 自训练教师模型机制

在有噪声标签的数据中进行深度神经网络的训练,一个很重要的思路就是进行干净样本选择。

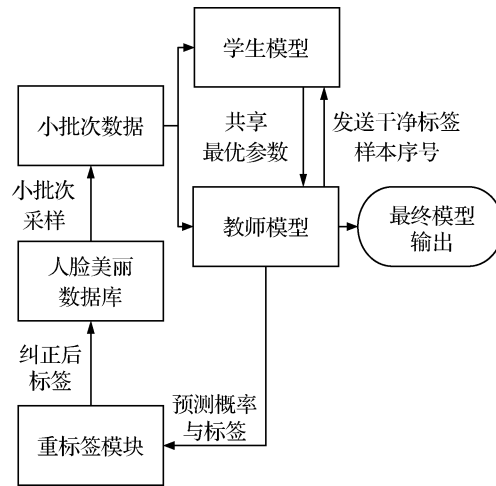


图 1 自纠正噪声标签方法框图

Fig. 1 Block diagram of self-correcting noise labels method

Han 等人 (2018) 在每批次数据完成训练后,直接利用深度神经网络进行干净样本判断与选择。本文将自训练出泛化能力相对好的深度神经网络作为教师模型,并仅利用教师模型进行样本选择,从而避免了每次训练完成后深度神经网络泛化性能波动造成的误判。设 N 幅小批次输入人脸图像及其对应的人脸美丽等级标签表示为 $(\mathbf{x}, \mathbf{y}) = \{(\mathbf{x}_i, y_i), \dots, (\mathbf{x}_N, y_N)\}$, 其中, $\mathbf{x} \in \mathbf{R}^{N \times 3 \times \tau \times v}$ 表示共 N 幅人脸图像, 3 层颜色通道, 每幅人脸图像的分辨率为 $\tau \times v$, $\mathbf{y} \in \mathbf{R}^N$ 为其对应的标签, $\mathbf{x}_i$ 为第 i 幅人脸图像, $y_i \in \{0, \dots, c-1\}$ 为第 i 幅人脸图像的人脸美丽等级标签, c 为人脸美丽等级标签总等级数; $\alpha \in [0, 1]$ 为噪声控制率, 即超参数。

首先,同时初始化两个结构相同的深度卷积神经网络,选定其中一个网络作为教师模型,另一个为学生模型。在训练时,将每个小批次人脸图像同时输入到教师、学生模型中,经正向传播并先计算教师模型损失值 $\mathbf{L}_t = \{l_1^t, l_2^t, \dots, l_N^t\}$ 。接着,将这 N 个损失值从低到高进行排序,并记录其中 $N \times (1 - \alpha)$ 个低损失值对应的样本序号。然后,将样本序号发送给学生模型,学生模型则从自身正向传播计算的损失值 $\mathbf{L}_s = \{l_1^s, l_2^s, \dots, l_N^s\}$ 中,挑选出接收样本序号对应的损失 $\mathbf{L}'_s = \{l_1^{s'}, l_2^{s'}, \dots, l_{N \times (1 - \alpha)}^{s'}\}$ 。最后,对学生模型进行反向传播更新参数,不断循环该过程直至出现验证准确率更高的学生模型,保存其参数作为新的教师模型。图 2 为自训练教师模型机制示意图。

1.2 重标签再训练机制

在自训练教师模型机制中,通过限制噪声标签

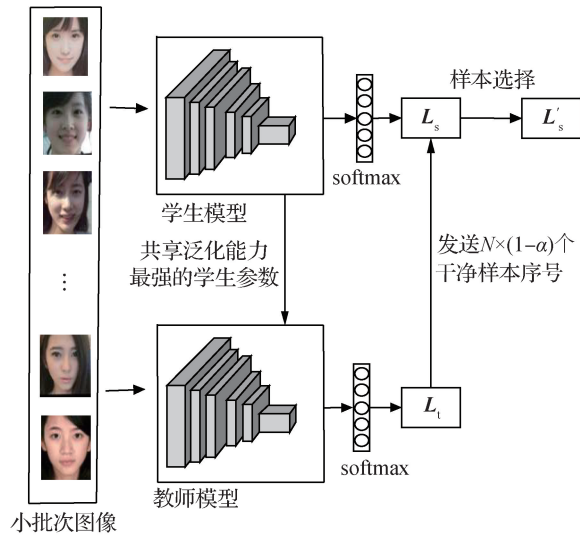


图2 自训练教师模型机制示意图

Fig. 2 Schematic diagram of the mechanism of self-training teacher model

样本的学习可减少噪声标签的影响。由于人脸美丽数据库仍存在标签数据样本不足的问题,为充分利用所有可用数据,本文融合 Wang 等人(2020)方法中的噪声标签纠正方法纠正噪声标签。

如果每幅人脸图像经教师模型预测的最大输出概率高于标签对应输出概率一定阈值,则认为该样本标签应该纠正,即

$$\bar{y} = \begin{cases} y_m & P_m - P_l > \delta \\ y & \text{其他} \end{cases} \quad (1)$$

式中, P_m 为预测输出的最大概率, P_l 为标签对应的预测输出概率, δ 为阈值间隔; y_m 为最大输出概率对应的标签, y 为原标签; $\bar{y}$ 为标签纠正后的伪标签;所有的预测概率皆为经 softmax 分类器输出的概率。

1.3 算法流程

本文提出的自纠正噪声标签算法的具体步骤如下:

1) 设学生和教师模型参数分别为 ω_s 和 ω_t , 学习率为 γ , 训练次数为 T , 纠正前训练总次数为 T_n , 最大训练总次数为 T_m , 小批次次数为 k_m , 小批次样本数为 N , 学生和教师模型验证准确率分别为 A_s 和 A_t , 测试准确率为 A , 学生模型为 ϕ , 教师模型为 ψ , 训练集为 D 。

2) for $T = 1, 2, \dots, T_n - 1$ 执行:

打乱训练集 D 中样本顺序;

for $k = 1, \dots, k_m$ 执行:

从训练集 D 中采样小批次样本 $\bar{D}$;

$\bar{D}_s = \arg \min_{N \times (1-\alpha)} L_t(\psi, \bar{D})$ // 教师模型采样 $N \times (1 - \alpha)$ 个小损失样本;

$\omega_s = \omega_s - \gamma \nabla L_s(\phi, \bar{D}_s)$ // 学生模型通过 $\bar{D}_s$ 更新 ω_s ;

end for。

在验证集上验证学生模型预测准确率, 记为 A_s ;

如果 $A_s > A_t$ 执行:

$\omega_t = \omega_s, A_t = A_s$ 且保存 ω_t, A_t ;

3) end for。

4) 如果 $T = T_n$ 执行:

读取 ω_t 作为教师模型;

for $k = 1, \dots, k_m$ 执行:

从训练集 D 中采样小批次样本 $\bar{D}$;

$P_m, P_l = \psi(\bar{D})$ // 正向传播获得预测概率
利用式(1)纠正噪声标签, 更新训练集为 D' ;

end for。

5) for $T = T_n + 1, T_n + 2, \dots, T_m$ 执行:

打乱训练集 D' 中样本顺序;

重复步骤2)中第3行及以下过程直至步骤

3);

6) end for。

7) 在测试集测试教师模型预测准确率, 记为 A 。

8) 输出 ω_t, A 。

2 实验结果与分析

2.1 实验对象

实验在 GeForce RTX2080ti GPU, Intel i7-9800X CPU 和 62.5 GB 内存的计算机上进行, 使用 PyTorch 框架; 采用 Albumentations 工具进行一系列数据管理和增强操作。首先, 将数据按照 6 : 2 : 2 的比例划分为训练集、验证集和测试集, 同时对训练集图像进行随机亮度、对比度、水平翻转、垂直翻转和随机仿射变换(移动、缩放)。然后, 将在 ImageNet 数据库上预训练过的 ResNet-18(He 等, 2016) 作为骨干神经网络, 使用交叉熵作为损失函数, 并以此作为基准方法。最后, 使用动量为 9×10^{-1} 的随机梯度下降优化器, 配置权重衰减系数为 1×10^{-2} 。使用余弦学习率(He 等, 2019), 并初始化学习率为 1×10^{-2} , 批

次大小设置为 40。

2.1.1 LSFBD 数据库

LSFBD(甘俊英等,2019)数据库由五邑大学项目组建立,共 20 000 幅带标签的人脸图像(男女各 10 000 幅)和 80 000 幅无标签人脸图像,分辨率为 144 × 144 像素,分为 0、1、2、3、4 共 5 个类别,分别对应极不吸引人、不吸引人、平均、吸引人和最吸引人。本文着重于女性吸引力预测,仅使用 10 000 幅女性人脸图像验证方法的有效性。其中,类别 0 有 948 幅、类别 1 有 1 148 幅、类别 2 有 3 846 幅、类别 3 有 2 718 幅、类别 4 有 1 340 幅,如图 3 所示。

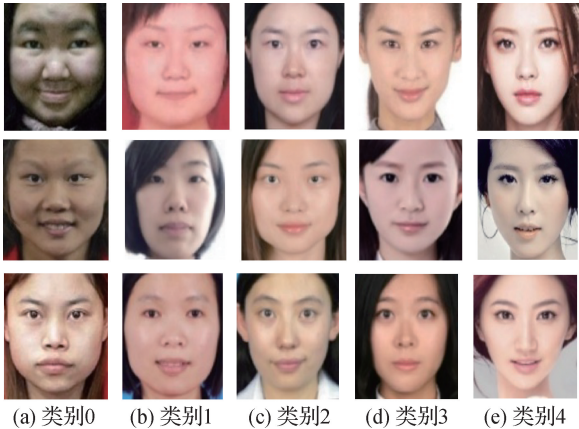


图 3 LSFBD 数据库中人脸图像
Fig. 3 Face images in LSFBD database
((a) type 0; (b) type 1; (c) type 2; (d) type 3; (e) type 4)

表 1 不同方法在人工合成噪声标签数据库不同噪声率的分类准确率对比
Table 1 Comparison of classification accuracy with various noise rates among different methods on synthetic noise label databases

噪声容忍方法	LSFBD 数据库			SCUT-FBP5500 数据库		
	噪声率 10%	噪声率 20%	噪声率 30%	噪声率 10%	噪声率 20%	噪声率 30%
基准(He 等,2016)	56.4	54.3	52.0	72.8	72.0	69.8
co-teaching(Han 等,2018)	58.2	56.6	55.6	73.2	72.7	70.8
self-cure network(Wang 等,2020)	58.9	57.3	56.1	73.9	73.3	71.8
自纠正噪声标签(本文)	60.1	58.4	57.8	75.3	74.8	72.9

注:加粗字体表示各列最优结果。

噪声率增加,本文方法的效果更加明显。Han 等人(2018)和 Wang 等人(2020)的方法皆能超过基准方法,但仍逊色于本文方法。

2.3 基于原始数据库实验

为进一步验证自纠正噪声标签方法的有效性,

2.1.2 SCUT-FBP5500 数据库

SCUT-FBP5500(Liang 等,2018)数据库由华南理工大学建立,共有 5 500 幅正面人脸图像,具有不同属性(男性或女性、亚洲人或白种人、年龄)和不同标签(人脸特征点、范围 1 ~ 5 的美丽分数、美丽分数分布),可实现不同人脸美丽预测范式的计算模型。因为该数据库是由 60 个志愿者对所有图像评分为 5 种人脸美丽等级,故本文将同一幅图像评分相同等级的志愿者人数最多的评分等级作为最终评分等级。

2.2 基于人工合成噪声标签数据库实验

本文分别使用 LSFBD 和 SCUT-FBP5500 数据库 10%、20%、30% 训练集噪声标签比率,探索自纠正噪声标签方法的有效性。从每一类人脸美丽等级中选择 10%、20%、30% 的训练数据,并随机改变它们的标签。表 1 为自纠正噪声标签方法与基准方法(没有考虑噪声标签的传统卷积神经网络训练)和其他方法的比较结果。本文执行人脸美丽预测五分类任务,以分类准确率(classification accuracy, CA)作为评价指标,表 1 中结果皆为 5 次实验结果的平均值。可以看出,在噪声率为 30%、20%、10% 时,本文方法在 LSFBD 数据库上分别超出基准方法 5.8%、4.1%、3.7%;在 SCUT-FBP5500 数据库上分别超出基准方法 3.1%、2.8%、2.5%。显然,随着

在原始 LSFBD 数据库和原始 SCUT-FBP5500 数据库(不进行人工合成噪声标签)上进行实验。假定在构建 LSFBD 数据库和 SCUT-FBP5500 数据库时,会不可避免地产生人为因素造成的噪声标签。虽然在验证集和测试集中仍可能存在少量噪声标

签,但由于训练集数据样本数量大于验证集、测试集数量,即训练数据样本中存在的噪声标签数量大于验证集、测试集中噪声标签数量,故在原始LSFBD数据库和原始SCUT-FBP5500数据库实验中,本文方法仍然取得了最好的实验效果。如表2

所示,本文方法在原始LSFBD数据库和原始SCUT-FBP5500数据库上预测准确率分别超出基准方法2.7%和1.2%。Han等人(2018)和Wang等人(2020)的方法皆能略高于基线方法,但都逊色于本文方法。

表2 不同方法在原始数据库的分类准确率对比

Table 2 Comparison of classification accuracy among different methods on the original databases

噪声容忍方法	原始LSFBD数据库	原始SCUT-FBP5500数据库	/%
基准(He等,2016)	58.1	74.3	
co-teaching(Han等,2018)	58.5	74.6	
self-cure network(Wang等,2020)	60.5	75.0	
自纠正噪声标签(本文)	60.8	75.5	

注:加粗字体表示各列最优结果。

2.4 两种机制对比实验

为验证比较自训练教师模型机制和重标签再训练机制的有效性,在原始LSFBD数据库和具有人工合成30%训练集噪声标签的LSFBD数据库中分别进行实验,实验结果如表3所示。在原始LSFBD数据库上,将表3第1行与第3行比较、第2行与第4行比较,可知使用自训练教师模型机制使分类准确率提升了0.5%和0.8%,平均提升0.7%。同理,在人工合成30%训练集噪声标签LSFBD数据库上,使用自训练教师模型机制使分类准确率平均提升了3.6%。即噪声水平越高,自训练教师模型机制的提升效果越好。因为自训练教师模型机制通过样本选择进行噪声标签限制,达到理想效果时学习的样本皆为干净标签样本。但噪声水平低时噪声标签样本少,受噪声标签负面影响少,故效果提升不明显。在原始LSFBD数据库上,将表3第1行与第2行、第3行与第4行比较,可知使用重标签再训练机制使分类准确率提升了1.9%和2.2%,平均提升2.1%。同理,在人工合成30%训练集噪声标签LSFBD数据库上,使用重标签再训练机制使分类准确率平均提升了1.9%。两种机制对比实验表明,本文方法在每一个环节都具有充分的有效性。

2.5 不同阈值间隔的对比实验

重标签机制中, δ 是预测输出的最大概率 P_m 与标签对应的预测输出概率 P_l 的阈值间隔。固定其他设置,在原始LSFBD数据库中 δ 取不同值时,实验结果如表4所示。从表4可知,当 δ 取值越远离

表3 在原始LSFBD数据库和人工合成30%训练集噪声标签的LSFBD数据库上的比较

Table 3 Comparison on the original LSFBD database and the LSFBD database with artificially synthesized 30% training set noise labels

自训练教师模型机制	重标签再训练机制	分类准确率		/%
		原始LSFBD	合成LSFBD + 30%噪声	
×	×	58.1	52.4	
×	√	60.0	54.6	
√	×	58.6	56.3	
√	√	60.8	57.8	

注:加粗字体表示各列最优结果,×表示没有使用,√表示使用。

表4 不同阈值间隔的分类准确率对比

Table 4 Comparison of classification accuracy with different threshold intervals

不同阈值间隔 δ	分类准确率/%
0.2	59.5
0.3	59.9
0.4	60.4
0.5	60.8
0.6	60.3
0.7	60.0
0.8	59.8

注:加粗字体表示最优结果。

0.5时,分类准确率越低。这是因为 $\delta < 0.5$ 时,会

导致更多错误的重标签操作,从而影响分类准确率。然而 $\delta > 0.5$ 时,没有完全纠正所有噪声标签,从而影响分类准确率。

3 结 论

为了减少人脸美丽预测中噪声标签的负面影响,本文提出一种自纠正噪声标签方法,在样本选择方法的基础上,自训练出噪声标签样本判断能力更强的深度网络,从而帮助区分噪声标签样本。本文方法包括自训练教师模型机制和重标签再训练机制。自训练教师模型机制以自训练的方式得到教师模型,帮助学生模型进行干净样本选择和训练,直至学生模型泛化能力超过教师模型并成为新的教师模型,并不断重复该过程。重标签再训练机制通过比较最大预测概率和标签对应预测概率,从而纠正噪声标签。最后,利用纠正后的数据反复执行自训练教师模型机制训练。在人工合成噪声标签 LSFBD 数据库、人工合成噪声标签 SCUT-FBP5500 数据库和原始 LSFBD 数据库、原始 SCUT-FBP5500 数据库上进行实验,皆证明本文方法能降低噪声标签的负面影响且效果优于基准方法。

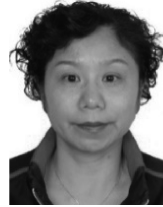
由于数据库中存在类间数据样本数量不平衡的问题,利用自纠正噪声标签方法判断数据样本数量少的类别中存在的噪声标签比较困难,这是由于数据样本数量少的类别表达的差异性特征不够强,因此更难以建立噪声标签样本判断的决策边界,后续将融合代价敏感方法,降低类间不平衡带来的影响,不断优化该方法。

参考文献 (References)

- Algan G and Ulusoy I. 2021a. Image classification with deep learning in the presence of noisy labels: a survey. *Knowledge-Based Systems*, 215: #106771 [DOI: 10.1016/j.knsys.2021.106771]
- Algan G and Ulusoy I. 2021b. Meta soft label generation for noisy labels//*Proceedings of the 25th International Conference on Pattern Recognition*. Milan, Italy: IEEE: 7142-7148 [DOI: 10.1109/icpr48806.2021.9412490]
- Arpit D, Jastrzebski S, Ballas N, Krueger D, Bengio E, Kanwal M S, Maharaj T, Fischer A, Courville A, Bengio Y and Lacoste-Julien S. 2017. A closer look at memorization in deep networks//*Proceedings of the 34th International Conference on Machine Learning*. Sydney, Australia: PMLR: 233-242
- Chen Q Q, Wang W J and Jiang G X. 2019. Label noise filtering based on the data distribution. *Journal of Tsinghua University (Science and Technology)*, 59(4): 262-269 (陈庆强, 王文剑, 姜高霞. 2019. 基于数据分布的标签噪声过滤. 清华大学学报(自然科学版), 59(4): 262-269) [DOI: 10.16511/j.cnki.qhdxxb.2018.26.059]
- Gan J Y, Zhai Y K, Huang Y, Zeng J Y and Jiang K Y. 2019. Research of facial beauty prediction based on deep convolutional features using double activation layer. *Acta Electronica Sinica*, 47(3): 636-642 (甘俊英, 翟懿奎, 黄聿, 曾军英, 姜开永. 2019. 基于双激活层深度卷积特征的人脸美丽预测研究. 电子学报, 47(3): 636-642) [DOI: 10.3969/j.issn.0372-2112.2019.03.017]
- Guo S, Huang W L, Zhang H Z, Zhuang C F, Dong D K, Scott M R and Huang D L. 2018. CurriculumNet: weakly supervised learning from large-scale web images//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer: 139-154 [DOI: 10.1007/978-3-030-01249-6_9]
- Han B, Yao Q M, Yu X R, Niu G, Xu M, Hu W H, Tsang I W and Sugiyama M. 2018. Co-teaching: robust training of deep neural networks with extremely noisy labels//*Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Montréal, Canada: MIT Press: 8536-8546
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/cvpr.2016.90]
- He T, Zhang Z, Zhang H, Zhang Z Y, Xie J Y and Li M. 2019. Bag of tricks for image classification with convolutional neural networks//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 558-567 [DOI: 10.1109/cvpr.2019.00065]
- Jiang L, Huang D, Liu M and Yang W L. 2020. Beyond synthetic noise: deep learning on controlled noisy labels//*Proceedings of the 37th International Conference on Machine Learning*. Virtual: PMLR: 4804-4815
- Li J N, Socher R and Hoi S C H. 2020. DivideMix: learning with noisy labels as semi-supervised learning [EB/OL]. [2021-06-01]. <https://arxiv.org/pdf/2002.07394.pdf>
- Liang L Y, Lin L J, Jin L W, Xie D R and Li M R. 2018. SCUT-FBP5500: a diverse benchmark dataset for multi-paradigm facial beauty prediction//*Proceedings of the 24th International Conference on Pattern Recognition*. Beijing, China: IEEE: 1598-1603 [DOI: 10.1109/icpr.2018.8546038]
- Liu S, Niles-Weed J, Razavian N and Fernandez-Granda C. 2020. Early-learning regularization prevents memorization of noisy labels//*Proceedings of the 34th Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates, Inc.
- Shu J, Xie Q, Yi L X, Zhao Q, Zhou S P, Xu Z B and Meng D Y.

2019. Meta-weight-net: learning an explicit mapping for sample weighting [EB/OL]. [2021-06-01]. <https://arxiv.org/pdf/1902.07379.pdf>
- Tang K and Lang C Y. 2021. Noise label based self-adaptive person reidentification. *Journal of Data Acquisition and Processing*, 36(1): 103-112 (唐轲, 郎丛妍. 2021. 基于噪声标签自适应的行人再识别方法. *数据采集与处理*, 36(1): 103-112) [DOI: 10.16337/j.1004-9037.2021.01.010]
- Wang K, Peng X J, Yang J F, Lu S J and Qiao Y. 2020. Suppressing uncertainties for large-scale facial expression recognition//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 6896-6905 [DOI: 10.1109/cvpr42600.2020.00693]
- Wang Y S, Ma X J, Chen Z Y, Luo Y, Yi J F and Bailey J. 2019. Symmetric cross entropy for robust learning with noisy labels//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 322-330 [DOI: 10.1109/iccv.2019.00041]
- Xie Q Z, Luong M T, Hovy E and Le Q V. 2020. Self-training with noisy student improves ImageNet classification//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 10684-10695 [DOI: 10.1109/cvpr42600.2020.01070]
- Xu Y L, Cao P, Kong Y Q and Wang Y Z. 2019. L_{DMI} : a novel information-theoretic loss function for training deep nets robust to label noise//*Proceedings of the 33rd Conference on Neural Information Processing Systems*. Vancouver, Canada: MIT Press: 6222-6233
- Zhang C Y, Bengio S, Hardt M, Recht B and Vinyals O. 2017. Understanding deep learning requires rethinking generalization [EB/OL]. [2021-06-01]. <https://arxiv.org/pdf/1611.03530.pdf>
- Zhang Z H, Jiang G X and Wang W J. 2021. Label noise filtering method based on local probability sampling. *Journal of Computer Applications*, 41(1): 67-73 (张增辉, 姜高霞, 王文剑. 2021. 基于局部概率抽样的标签噪声过滤方法. *计算机应用*, 41(1): 67-73) [DOI: 10.11772/j.issn.1001-9081.2020060970]

作者简介



甘俊英, 1964 年生, 女, 教授, 主要研究方向为生物特征和模式识别。

E-mail: junyinggan@163.com

吴必诚, 男, 硕士研究生, 主要研究方向为生物特征识别和深度学习。E-mail: wbc1076002292@163.com

翟懿奎, 男, 副教授, 主要研究方向为生物特征识别及合成孔径雷达图像识别。E-mail: yikuizhai@163.com

何国辉, 男, 教授, 主要研究方向为云计算与大数据技术及生物特征身份识别。E-mail: ghhe126@126.com

麦超云, 男, 讲师, 主要研究方向为信号与信息处理、深度学习。E-mail: maichaoyun@foxmail.com

白振峰, 男, 硕士研究生, 主要研究方向为生物特征识别和机器视觉。E-mail: baizhenfeng1212@163.com