



JOURNAL OF IMAGE AND GRAPHICS

主办: 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

中国图象图形学报

2021
06
VOL.26

ISSN1006-8961
CN11-3758/TB



中国图象图形学学会成立三十周年

图像图形学 发展 年度报告



第26卷第6期（总第302期）

2021年6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

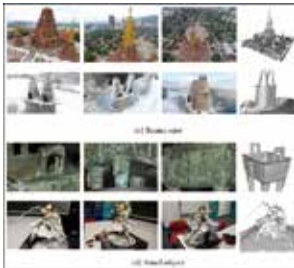
国内邮发代号 82-831

国内定价 60.00元

序言	王耀南
----------	-----



生物特征识别学科发展报告
(第1254页)



大规模室外图像3维重建技术
研究进展(第1429页)



虚实融合场景中的深度感知
研究综述(第1503页)

图像处理与通信技术

视频处理与压缩技术

贾川民, 马海川, 杨文瀚, 任文琦, 潘金山, 刘东, 刘家琰, 马思伟	1179
---	------

面向体验质量的多媒体计算通信

陶晓明, 杨铀, 徐迈, 段一平, 黄丹蓝, 刘文予	1201
----------------------------------	------

数字媒体取证技术综述

李晓龙, 俞能海, 张新鹏, 张卫明, 李斌, 卢伟, 王伟, 刘晓龙	1216
---	------

面向智慧城市的交通视频结构化分析前沿进展

赵耀, 田永鸿, 党建武, 付树军, 王恒友, 万军, 安高云, 杜卓然, 廖理心, 韦世奎	1227
--	------

生物特征识别学科发展报告

孙哲南, 赫然, 王亮, 阚美娜, 冯建江, 郑方, 郑伟诗, 左旺孟, 康文雄, 邓伟洪, 张杰, 韩璇, 山世光, 王云龙, 茹一伟, 朱宇豪, 刘云帆, 何勇	1254
--	------

自然场景文本检测与识别的深度学习方法

刘崇宇, 陈晓雪, 罗灿杰, 金连文, 薛洋, 刘禹良	1330
-----------------------------------	------

基于深度学习的跨模态检索综述

尹奇跃, 黄岩, 张俊格, 吴书, 王亮	1368
----------------------------	------

三维视觉和图形技术

三维视觉前沿进展

龙霄潇, 程新景, 朱昊, 张朋举, 刘浩敏, 李俊, 郑林涛, 胡庆拥, 刘浩, 曹汛, 杨睿刚, 吴毅红, 章国锋, 刘烨斌, 徐凯, 郭裕兰, 陈宝权	1389
--	------

大规模室外图像3维重建技术研究进展

颜深, 张茂军, 樊亚春, 谭小慧, 刘煜, 彭杨, 刘宇翔	1429
--------------------------------------	------

视觉传感成像技术与数据处理进展

王程, 陈峰, 汶德胜, 雷浩, 宋宗玺, 赵航芳	1450
---------------------------------	------

视觉—惯性导航定位技术研究进展

司书斌, 赵大伟, 徐婉莹, 张勇刚, 戴斌	1470
------------------------------	------

三维视觉测量技术及应用进展

张宗华, 刘巍, 刘国栋, 宋丽梅, 屈玉福, 李旭东, 魏振忠	1483
--	------

虚实融合场景中的深度感知研究综述

平佳敏, 刘越, 翁冬冬	1503
--------------------	------

可微绘制技术研究进展

许威威, 周漾, 吴鸿智, 过洁	1521
------------------------	------

沉浸式立体显示技术在临床医学领域中的应用

邵永航, 石俊生	1536
----------------	------

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Overview of biometrics research
(P1254)



Progress in the large-scale
outdoor image 3D reconstruction
(P1429)



Review of depth perception in
virtual and real fusion environment
(P1503)

Image Processing & Communication Technology

Video processing and compression technologies

Jia Chuanmin, Ma Haichuan, Yang Wenhan, Ren Wenqi, Pan Jinshan, Liu Dong, Liu Jiaying, Ma Siwei 1179

Multimedia computing communications

Tao Xiaoming, Yang You, Xu Mai, Duan Yiping, Huang Danlan, Liu Wenyu 1201

Overview of digital media forensics technology

Li Xiaolong, Yu Nenghai, Zhang Xinpeng, Zhang Weiming, Li Bin, Lu Wei, Wang Wei, Liu Xiaolong 1216

Frontiers of transportation video structural analysis in the smart city

Zhao Yao, Tian Yonghong, Dang Jianwu, Fu Shujun, Wang Hengyou, Wan Jun, An Gaoyun, Du Zhuoran, Liao Lixin, Wei Shikui 1227

Overview of biometrics research

Sun Zhenan, He Ran, Wang Liang, Kan Meina, Feng Jianjiang, Zheng Fang, Zheng Weishi, Zuo Wangmeng, Kang Wenxiong, Deng Weihong, Zhang Jie, Han Hu, Shan Shiguang, Wang Yunlong, Ru Yiwei, Zhu Yuhao, Liu Yunfan, He Yong 1254

Deep learning methods for scene text detection and recognition

Liu Chongyu, Chen Xiaoxue, Luo Canjie, Jin Lianwen, Xue Yang, Liu Yuliang 1330

3D Vision & Graphics Technology

Survey on deep learning based cross-modal retrieval

Yin Qiyue, Huang Yan, Zhang Junge, Wu Shu, Wang Liang 1368

Recent progress in 3D vision

Long Xiaoxiao, Cheng Xinjing, Zhu Hao, Zhang Pengju, Liu Haomin, Li Jun, Zheng Lintao, Hu Qingyong, Liu Hao, Cao Xun, Yang Ruigang, Wu Yihong, Zhang Guofeng, Liu Yebin, Xu Kai, Guo Yulan, Chen Baoquan 1389

Progress in the large-scale outdoor image 3D reconstruction

Yan Shen, Zhang Maojun, Fan Yachun, Tan Xiaohui, Liu Yu, Peng Yang, Liu Yuxiang 1429

Review on imaging and data processing of visual sensing

Wang Cheng, Chen Feng, Wen Desheng, Lei Hao, Song Zongxi, Zhao Hangfang 1450

Review on visual-inertial navigation and positioning technology

Si Shubin, Zhao Dawei, Xu Wanying, Zhang Yonggang, Dai Bin 1470

Overview of the development and application of 3D vision measurement technology

Zhang Zonghua, Liu Wei, Liu Guodong, Song Limei, Qu Yufu, Li Xudong, Wei Zhenzhong 1483

Review of depth perception in virtual and real fusion environment

Ping Jiamin, Liu Yue, Weng Dongdong 1503

Differential rendering: a survey

Xu Weiwei, Zhou Yang, Wu Hongzhi, Guo Jie 1521

Application of immersive 3D imaging technology in the clinic medical field

Tai Yonghang, Shi Junsheng 1536

中图法分类号: TP301.6 文献标识码: A 文章编号: 1006-8961(2021)06-1227-27

论文引用格式: Zhao Y, Tian Y H, Dang J W, Fu S J, Wang H Y, Wan J, An G Y, Du Z R, Liao L X and Wei S K. 2021. Frontiers of transportation video structural analysis in the smart city. Journal of Image and Graphics, 26(06): 1227-1253 (赵耀, 田永鸿, 党建武, 付树军, 王恒友, 万军, 安高云, 杜卓然, 廖理心, 韦世奎. 2021. 面向智慧城市的交通视频结构化分析前沿进展. 中国图象图形学报, 26(06): 1227-1253) [DOI: 10.11834/jig.210035]

面向智慧城市的交通视频结构化分析前沿进展

赵耀¹, 田永鸿², 党建武³, 付树军⁴, 王恒友⁵, 万军⁶,
安高云¹, 杜卓然¹, 廖理心¹, 韦世奎^{1*}

1. 北京交通大学信息科学研究所, 北京 100044; 2. 北京大学信息科学技术学院, 北京 100871;
3. 兰州交通大学电子与信息工程学院, 兰州 730070; 4. 山东大学数学学院, 济南 250100;
5. 北京建筑大学理学院, 北京 102616; 6. 中国科学院自动化研究所模式识别国家重点实验室, 北京 100190

摘要: 随着智慧城市建设的不断深入,大量的传感器设备铺置在城市公路和轨道等交通场景,为多维度全方位感知城市交通状态构建了广泛的感知网络,产生了海量的交通视频数据。海量交通视频数据是城市管理的数据宝藏,理解与分析这些数据是智慧城市建设的核心。面对高度冗余的交通视频数据,如何高效准确地挖掘和提取结构化信息,实现对重点目标(如人、车、物)的快速检测、识别与检索,是交通视频处理的核心问题——交通视频结构化分析。交通视频结构化分析包括车辆视频结构化分析、人员结构化分析及其行为分析。其中,车辆结构化作为一个复杂的多步骤任务,主要由车辆的检测、车辆的属性(车牌、车型和颜色)识别以及车辆的检索和重识别等子任务构成。人脸结构化和行人结构化是交通视频中行人结构化智能分析中的两个重要研究方向,主要分析人脸或者行人的一些表现属性。行人行为分析是指对行人在复杂交通环境下做出的动作进行识别和预测。本文从交通视频中的车辆、行人及其行为分析等方面,阐述交通视频结构化分析领域的研究热点及前沿进展,汇总比较国内外的相关成果,并对交通视频结构化分析领域的研究进行总结分析与展望。

关键词: 交通视频; 车辆结构化分析; 行人结构化分析; 行为结构化分析; 车辆检测; 车辆属性识别; 车辆检索; 人脸结构化分析

Frontiers of transportation video structural analysis in the smart city

Zhao Yao¹, Tian Yonghong², Dang Jianwu³, Fu Shujun⁴, Wang Hengyou⁵, Wan Jun⁶,
An Gaoyun¹, Du Zhuoran¹, Liao Lixin¹, Wei Shikui^{1*}

1. Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China; 2. School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China; 3. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China; 4. School of Mathematics, Shandong University, Jinan 250100, China; 5. School of Science, Beijing University of Civil Engineering and Architecture, Beijing 102616, China; 6. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Abstract: As the construction of smart cities continues to deepen, our country gradually builds multidimensional and omnidirectional sensor systems in roads, railways, and urban rails and other ground transportation fields to build strong data sup-

收稿日期: 2021-01-19; 修回日期: 2021-02-04; 预印本日期: 2021-02-11

* 通信作者: 韦世奎 shkwei@bjtu.edu.cn

port for smart transportation. Faced with all-weather traffic data collected by sensors, analyzing the data by relying solely on human resources is no longer possible. Therefore, studying the structural analysis technology of traffic video and establishing a safe, flexible, and efficient intelligent transportation system has significant social benefits and application value. Traffic video structural analysis is the core technology in smart transportation. It aims to use artificial intelligence algorithms to parse unstructured traffic video data into structured semantic information that is easy for workers and computers to understand and provide basic technical support for subsequent related tasks. The structural analysis of traffic video is a key technology for smart city construction. It can help the police in quickly locating criminal vehicles and travel routes, greatly improve the police's efficiency in solving crimes, and maintain city safety; it can also automatically identify illegal vehicles and types of violation, constrains people to abide by the traffic order, and realize a smooth urban traffic environment. With the advent of the 5G internet of things era, ultrahigh network bandwidth and transmission speed further improves the quality and efficiency of vehicle video transmission. Efficiently and accurately conducting traffic video structure analysis will be the focus of research in the next few years. Traffic video structural analysis includes vehicle video structural analysis, personnel structural analysis, and behavior analysis. Among them, as a complex, multistep task, vehicle structuring is mainly composed of three subtasks, namely, vehicle detection, vehicle attributes (license plate, type, and color) recognition, and vehicle retrieval and reidentification. Human face structuring and pedestrian structuring are two important research directions in the intelligent analysis of traffic videos. They mainly analyze some apparent attributes of human faces or pedestrians, such as age, gender, mask, backpack, clothing color, and length. Pedestrian behavior analysis refers to the identification and prediction of pedestrian actions. For example, the speed at which pedestrians currently head and in which direction, whether they are answering calls, and whether they have to cross the road. For the task of vehicle structure analysis, first, the object detection technology must be used to quickly and accurately locate the vehicle. Second, on the basis of positioning the vehicle, it fully excavates the visual characteristics of the vehicle, realizes the identification of the inherent attributes of vehicle, and generates structured tags about the vehicle. Finally, on the basis of structured tags, the retrieval technology and reidentification technology are further combined to realize the retrieval and reidentification of a specific vehicle in the massive video data. Personnel structural analysis and behavior analysis can detect and identify pedestrians in traffic videos and conduct structured data extraction and behavior analysis of detected personnel. In the analysis of personnel structure, a person is extracted as a descriptive individual. In terms of face structure, it includes accurate facial positioning, facial feature extraction, and facial feature comparison. In terms of pedestrian structure, it includes gender, age, and age of the person. Various descriptive information includes height, hair accessories, clothing, carrying items, and walking patterns. Pedestrian behavior analysis is carried out on the basis of personnel structure analysis. Behavior analysis refers to the recognition, comprehension, and prediction of pedestrian actions. In the area of big data processing and analysis of traffic video, research on vehicle structuring started earlier and related technologies have also developed rapidly, but it can still be remarkably developed. The premise of vehicle structuring is vehicle detection, which is affected by the shooting scene and the moving speed of the vehicle. Accurately locating the vehicle in the case of low light and the fast vehicle speed is still a problem to be solved. Many types of vehicles are found in the market, and the differences between models of similar brands are small. License plate recognition has become more important. In complex and changeable scenes, the generalization and accuracy of the positioning and recognition algorithm should be further improved. The extensive deployment of traffic monitoring equipment realizes all-weather monitoring of relevant road systems and further increases the difficulty of vehicle retrieval and reidentification tasks. Rapid retrieval or reidentification of target vehicles in complex and changeable scenes is crucial. It requires continuous investment and a much innovative research by scientific researchers. The need for structured pedestrian analysis has gradually emerged with further improvement of urban management. Pedestrian structuring mainly analyzes some apparent attributes of faces or pedestrians, such as age, gender, and clothing style, and provides more detailed data support for subsequent related tasks. Pedestrian structured analysis technology has also ushered in a period of rapid development with the development of deep learning. However, the structured analysis of pedestrians for specific scenarios, such as accurately identifying the age and gender of a person in an unconstrained environment, implementing the deployment of high-precision models in terminal systems with limited resources, and integrating multimodal information to further improve the accuracy of pedestrian attribute recognition, needs further research. Pedestrian behav-

ior analysis is a more advanced task in traffic video big data processing and analysis. It is more challenging due to factors, such as shooting scenes, moving cameras, viewing angles, and lighting changes. Judging from the behavior recognition effect of the mainstream neural network architecture, the current model does not achieve the desired effect on the large-scale behavior data set Kinetic because the existing model still fails to fully learn and model the behavioral timing relationship. In the field of behavior recognition, future research can still focus on recognition models for designing long-time-dependent network architectures, adapting large-scale data sets, and achieving lightweight behavior. With the development of Internet of Things and 5G technologies, the promotion of new technologies has also played an important role in the structural analysis of traffic video. To be equipped with IoT devices has become an inevitable trend for modern cars. Vehicles can be connected to basic transportation facilities (vehicle to infrastructure, V2I) or to surrounding vehicles (vehicle-to-vehicle, V2V). The development of these technologies depends on the common progress of vehicle video structuring and internet of things technology. With the global popularity of 5G technology, rapid transmission of high-quality video data has become a reality. Extracting structured information more efficiently from traffic videos, such as vehicle information, pedestrian information, and behavior prediction, has become more urgent. Researchers should study on improving the performance of related algorithms, should design more efficient hardware systems, and build more efficient traffic video structured analysis systems through software and hardware collaborations. We discuss the related work on traffic video structural analysis in detail from three aspects, as follows: vehicle, personnel, and behavior analysis. Moreover, we summarize these research works and provide some reasonable directions for future work.

Key words: traffic video; vehicle structural analysis; personnel structural analysis; behavior structural analysis; vehicle detection; vehicle attribute recognition; vehicle retrieval; human face structural analysis

0 引言

随着智慧城市建设的不断深入,我国在公路、铁路和城市轨道等地面交通领域逐步搭建起多维度、全方位的传感器系统,为智慧交通构建了强大的数据支撑。面对传感器采集的全天候交通数据,单纯依赖人力已经无法完成数据解析。因此,研究交通视频结构化分析技术,建立安全、灵活和高效的智慧交通系统具有重大的社会效益和应用价值。交通视频结构化分析是智慧交通中的核心技术,旨在利用人工智能算法将非结构化的交通视频数据解析成便于工作人员和计算机理解的结构化语义信息,为后续相关任务提供基本技术支撑。

交通视频结构化是理解与分析海量交通视频数据的核心,是未来实现智慧城市、智慧交通的关键。当前,国内外厂商已经投入大量人力和财力研发视频结构化技术,并将其应用到相应产品中。国际上,美国 Iveda 公司的交通视频结构化服务可以高效检测交通视频中人、车及车牌号码等,通过对各路段交通流量的预测分析来加强交通管理。加拿大的 IRD 公司搭建了 HTMS (highway traffic management systems) 的高速交通管理系统,使用了自动车辆分类和

计数方法、故障检测方法等多项技术,实现了交通流量的高效监控。在国内,海康威视公司发布了车辆视频结构化分析系统,能够解析出车辆的型号、颜色、类型及车牌号等。宇视科技公司提出了视频结构化分布式系统端到端解决方案,实现了数据采集、处理分析与预测的集成。交通视频结构化作为智慧城市建设的关键技术,不仅可以帮助警方快速锁定犯罪分子的车辆和行进路线,大幅提升警方的破案效率,维护城市安全,而且可以自动识别违章车辆以及违章的类型,约束人们遵守交通秩序,实现通畅的城市交通环境。随着 5G 物联网时代的来临,超高的网络带宽和传输速度会进一步提升车辆视频传输的质量和效率,如何高效准确地进行交通视频结构化分析将是未来几年的研究重点。

交通视频结构化分析包括车辆视频结构化分析、人员结构化分析及其行为分析,如图 1 所示。其中,车辆结构化作为一个复杂的、多步骤的任务,主要由车辆的检测、车辆的属性(车牌、车型和颜色)识别以及车辆的检索和重识别等 3 个子任务构成。人脸结构化和行人结构化是交通视频中行人结构化智能分析中的两个重要研究方向,主要分析人脸或者行人的一些表观属性,比如年龄、性别、口罩、背包、衣着颜色和长短等。行人行为分析是指对行人

的动作进行识别以及预测。比如,行人当前以多大的速度向哪个方向前进,是否在接听电话,是否要过马路等。

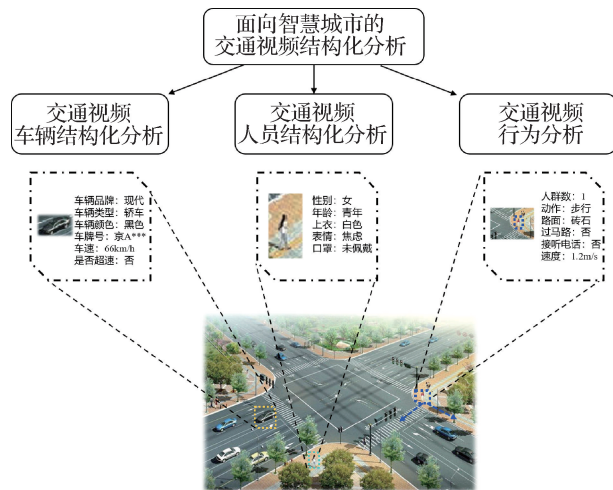


图1 交通视频结构化分析示意图

Fig. 1 The example of transportation video structural analysis

针对车辆结构化分析任务,首先要采用目标检测技术快速准确定位到车辆;然后在定位车辆的基础上充分挖掘车辆的视觉特征,实现对车辆固有属性的识别,生成关于车辆的结构化标签;最后在结构化标签的基础上,进一步结合检索技术和重识别技术,实现特定车辆在海量视频数据中的检索定位和重识别。人员结构化分析及其行为分析,主要对交通视频中的行人进行检测与识别,并对检测到的人员进行结构化数据提取与行为分析。在人员结构化分析中,人作为一个可描述的个体提取出来,人脸结构化包括面部精确定位、面部特征提取和面部特征比对等;行人结构化包括人员的性别、年龄、身高、发饰、衣着、物品携带和步履形态等多种描述信息。在人员结构化分析的基础上,进行行人行为分析。行为分析是指对行人的行为动作进行识别、理解及预测。

本文重点围绕交通视频结构化中的车辆结构化分析、人员结构化分析及行人行为分析,阐述交通视频结构化分析领域的研究热点及前沿进展,并对该领域的研究进行总结与展望。

1 国际研究现状

1.1 交通视频车辆结构化

交通视频车辆结构化分析包括车辆检测、车辆

属性识别和车辆检索与重识别。

1.1.1 车辆检测

在交通视频大数据处理中,车辆的检测是车辆结构化的首要任务,是提取车辆结构化信息的关键。尽管车辆检测技术发展迅速,并取得了大量研究成果,然而仍然面临很多挑战。比如,现实拍摄场景中光照、天气和摄像头位置等物理因素的变化,造成交通视频数据信息丢失等情况,或者出现特殊罕见场景,进而导致车辆检测性能的降低。另外,现实应用场景常常是计算资源受限的终端系统,既要保证算法精度又要保证算法速度,实现准确性和实效性的平衡。基于深度学习模型的车辆检测方法根据是否需要模型预先生成区域候选框,分为基于二阶段和基于单阶段两种车辆检测方法。

基于二阶段的车辆检测方法包括候选区域生成和车辆目标精确分类定位两个阶段。第1阶段利用传统图像分割算法或候选区域生成网络生成车辆目标的候选区域;第2阶段利用检测网络对候选区域进行微调,预测候选区域的类别以及位置信息。基于二阶段的方法主要有 R-CNN (regions with convolutional neural network features) (Girshick 等,2014)、SPP-Net (spatial pyramid pooling net) (He 等,2015)、Fast R-CNN (Girshick,2015)、Faster R-CNN (Ren 等,2015)、R-FCN (region-based fully convolutional networks) (Dai 等,2016)、Mask R-CNN (He 等,2017)、CBNet (composite backbone network) (Liu 等,2020c)、DETR (detection transformer) (Carion 等,2020)等。虽然基于二阶段的车辆检测方法的检测准确率很高,但检测速度较慢。

相比于二阶段的检测方法,基于单阶段的车辆检测方法,省略了候选区域生成阶段,将检测问题归结为回归问题,同时对物体的位置坐标和类别标签进行预测。基于单阶段的方法主要有 OverFeat (Sermanet 等,2013)、YOLO (you only look once) (Redmon 等,2016)、SSD (single shot multibox detector) (Liu 等,2016b)、RetinaNet (Lin 等,2017)等。虽然基于单阶段方法的运行速度相比二阶段方法有了很大提升,但是检测精度会有所下降。

基于深度学习的经典目标检测框架的基础,提出了一些改进的车辆检测方法。Kim 等人 (2016) 和 Ashraf 等人 (2017) 分别利用单阶段的 SSD 算法以及二阶段的 R-CNN 算法检测视频中的

车辆,取得了较好效果。Chadwick 等人(2019)将雷达数据与交通视频数据相结合,实现了远距离的小目标车辆检测。Gao 等人(2020)提出时间—空间预处理模型对车辆视频预处理,通过利用多

邻域阈值对前景像素点进行分类,提升了车辆检测性能。

基于深度学习模型的车辆检测方法的主要研究进展如图2所示。

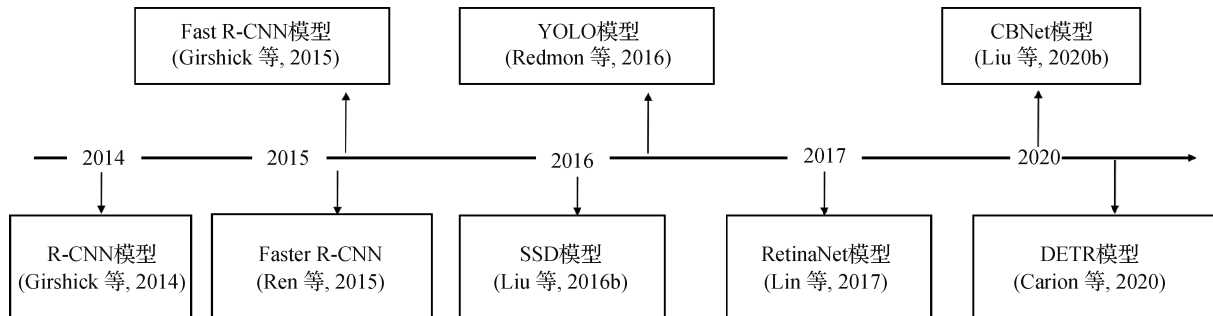


图2 车辆检测研究进展

Fig. 2 The recent advances in vehicle detection

1.1.2 车辆属性识别

车辆的属性信息主要包括车牌信息、车辆型号和车辆颜色。车牌信息是车辆的唯一标识符号,如何在复杂多样的环境中准确识别车牌是公安和交通系统急需解决的问题。车辆型号可以直接反映车辆的加速度、最大时速和刹车等性能,准确识别车型对于智慧交通系统有着重要意义。车辆颜色是车辆最显著的特征,如何在阴暗天气以及强光环境下准确识别车辆颜色是研究的重要领域。

车牌识别主要由车牌检测、车牌字符分割和车牌字符识别3项技术构成,前两项技术是字符识别的基础。

车牌检测用于定位车牌在图像中的位置,是车牌识别任务中必不可少的步骤。由于图像在获取过程中存在视角、形状、颜色、格式及照明条件等不确定因素,因此车牌识别是一项富有挑战性的任务。Ullah 和 Lee(2016)基于形态逻辑属性的方法定位车牌位置,能够有效定位不同大小和结构的英文车牌。Omran 和 Jarallah(2017)提出将光学字符识别(optical character recognition, OCR)技术与模板映射和相关度计算相结合的方法,能够自动解析出车牌。Rana 和 Dahiya(2017)定义了多种车牌检测方法,并与关联成分分析和欧氏距离变换相结合,提高了检测模型的精度。

在准确检测车牌位置的基础上,大量研究集中于字符分割。Fernandes 等人(2013)利用 K-means 方法对车牌字符进行分割,然后利用关联成分标签

分析算法识别关联像素区域,将合适的像素点分在一组以便有效挖掘车牌中的每一个字符。Khare 等人(2015)开发了新的锐度依赖模型分割车牌图像中的字符,在图像模糊的情况下可以较好地分割出字符。Ingole 和 Gundre(2017)提出基于字符特征的车牌预测模型,利用字符垂直和水平方向的特征对字符进行分割,适用于背景不复杂情况的车牌字符分割。Radchenko 等人(2017)利用关联成分分析分割车牌中的字符,取得了不错效果。

在车牌定位和有效字符分割的基础上,提出了通过挖掘字符特征进行字符特征提取和识别的方法。Pustokhina 等人(2020)提出 OKM-CNN (optimal K-means with convolutional neural network)模型,使用 IBA(improved bernsen algorithm)算法和关联成分分析模型对车牌定位,再用卷积神经网络识别车牌字符,取得了很好的效果。Raza 等人(2020)提出一个能够识别多国车辆车牌的自适应框架,无需任何有关国家车牌的先验知识即可识别车牌。Yousif 等人(2020)基于中性集合优化和遗传算法对车牌进行识别,首先通过边缘检测和形态学操作对车牌定位,然后使用遗传算法提取图像的重要特征并对中性集进行优化。中性集的利用减少了车牌图像的不确定性。Yoo 和 Han(2020)通过两个并行的时间匹配层和顶点优化的方法,显著提高了视频中车牌检测的准确性。

车型是描述车辆的重要属性,但是车型识别面临着诸多挑战,比如相同品牌不同型号的车辆在外观上具有很高的相似度、恶劣天气影响等。Clady

等人(2008)提出多类别车辆类型识别系统,采用车辆轮廓点作为相似度量指标,利用最近邻算法判断车辆的类型。Sarfraz 等人(2009)结合深度神经网络和支持向量机对公交车、卡车、面包车和小轿车等4个类别进行分类,性能较其他机器学习算法获得了较大提升。Awang 和 Azmi(2017)提出基于稀疏核及跳层连接的卷积神经网络,在不同天气条件下可以有效识别车辆的类型。Shvai 等人(2018)提出一种优化集成分类器对车辆类型进行识别,使用卷积神经网络和基于梯度提升的分类器将预测的类别概率连续值与两个光学传感器的离散类别标签进行融合,显著提升了识别精度。Soon 等人(2020)提出 PCN(principal component analysis convolutional network)方法,采用无监督学习方法 PCA(principal component analysis)生成卷积核,减少了训练成本并加强了特征提取的鲁棒性,在保证算法速度的同时,保持了较高的识别精度。

车辆颜色信息是车辆属性中最直观的特征信息,但是不同的光照和天气条件对颜色判断会有不

同程度的影响,为了解决这些问题,对车辆颜色识别展开了大量研究。Son 等人(2007)首先将车辆图像映射到 HSV(hue, saturation, value)颜色空间,然后用网格化的核对支持向量机进行改进,提升了颜色分类的准确性。Park 和 Kim(2008)提出利用主成分分析对车辆图像的颜色直方图降维,用支持向量机对颜色类别进行分类,实现了车辆颜色信息的高效检测。随后又进一步将3维颜色空间映射到2维平面并使用朴素贝叶斯分类器将颜色直方图转换成类别直方图,更加准确地识别出车辆颜色(Park 和 Kim,2012)。Aarathi 等人(2017)使用 dark channel prior 方法对图像进行除雾,再用卷积神经网络对除雾后的图像进行颜色分类,研究发现除雾后模型可以更有效地识别车辆的颜色。Kim 等人(2018)使用 Harris 角点检测算法生成代表颜色区域的概率图,随机从概率图中选择点集生成图像,然后输入到卷积神经网络中。

近年来车辆属性识别技术的研究进展如图3所示。

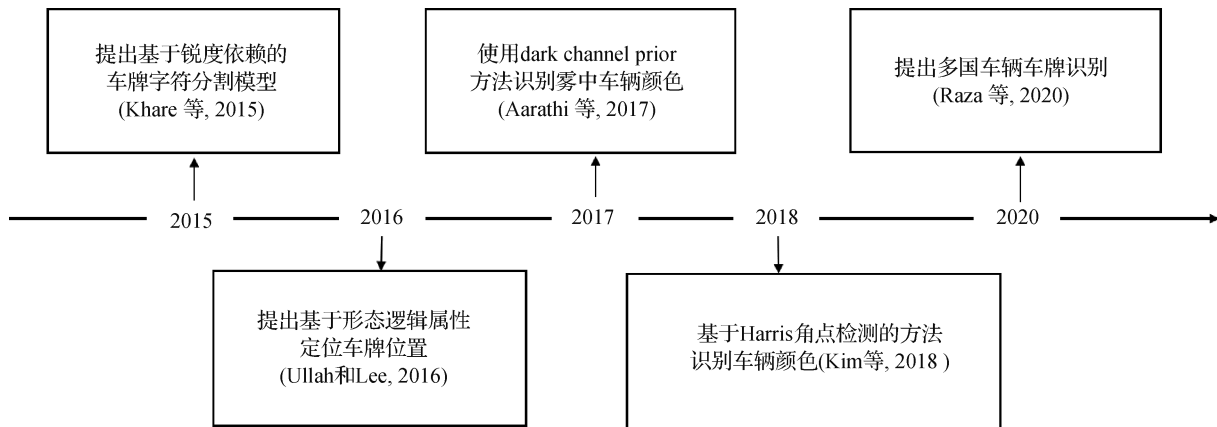


图3 车辆属性识别研究进展

Fig. 3 The recent advances in the recognition of vehicle attributes

1.1.3 车辆检索与重识别

车辆检索技术在交通系统中具有重要作用。车辆检索需要根据车牌号、车辆型号等具体车辆信息以及相关结构化特征在车辆数据库中检索出符合条件的目标车辆。随着图像检索技术日益成熟,基于监控视频的车辆检索工作引起了大量关注(Swain 和 Ballard,1991)。面对交通监控视频的图像分辨率低、目标不易分割以及车辆姿势光照的变化问题,Brown(2010)提出基于实例和颜色相关图的车辆检索模型系统,避免了严格的颜色分类局限性,提高了

检测准确性,并将任务聚焦于基于警报的车辆检索,减少了在大型数据库上的检索时间花费。鉴于传统方法检索车辆要求有固定的方向和颜色,Chen 等人(2009)提出基于特征颜色和多实例学习的车辆检索模型,采用多实例学习的方式从不同图像中学习具体车辆的视觉属性,同时为了缓解多实例学习对数据集的正负样本对比性过高问题,将目标特征颜色技术应用到学习过程中。Lin 等人(2013)提出采用3D建模方法提升基于内容检索的准确度,将定向金字塔直方图与加权雅可比系统相结合提取特征

信息,实现对车辆的快速匹配检索。在上述研究的基础上,Liu 等人(2014b)提取并整合多个监测设备中的车辆信息,并基于车辆 ID(identity document)信息、空间时序信息和交通道路信息,构建了路径排序算法估计车辆的行驶轨迹。由于现实情况中交通监控数据集包含大量环境图像和车辆,使得车辆检索的时间成本增大。为了缓解此问题,Pei 等人(2014)提出基于词袋模型的快速车辆检索模型,通过 hessian 仿射变换和 SIFT(scale invariant feature transform)算子在图像中提取特征并转化为向量作为视觉词,采用词频—逆文本频率(term frequency—inverse document frequency, TF-IDF)为每个词赋予权重,构建车辆特征词袋,以此快速匹配检索出目标车辆图像。

随着卷积神经网络在目标检测、分割等领域的广泛应用,涌现出大量基于深度学习的车辆检索方法。Liu 等人(2016a)提出一种深度相对距离学习方法(deep relative distance learning, DRDL),引入一个分叉的深度卷积网络将初始车辆图像映射到欧氏空间,直接通过欧氏距离测量车辆之间的相似度。针对传统方法精度和识别率低的问题,张子龙等人(2014)提出一种改进的加速鲁棒特征提取算法(speeded up robust features, SURF)用于车辆视频检测。Liang 等人(2017)提出一种新型的监督深度哈希方法处理大规模实例级车辆搜索。Tao 和 Lu(2018)提出一种基于多特征融合的车辆检索方法,将色相、饱和度和灰度图像的直方图,颜色布局描述符、感知哈希汉明距离和尺度不变特征变换(SIFT)关键点进行匹配。Cheong 等人(2018)将时空立方体引入较小的搜索块中解决检索时间的爆炸式增长问题。

车辆重识别是车辆检索任务中的一个重要子任务,指在特定范围内的交通监控场景下,判断非重叠区域内拍摄的车辆图像是否属于同一辆车的车辆检索问题(刘凯等,2020)。理论上车辆重识别可以简化为车牌识别任务,但是在实际的交通监控场景中,监控视角多样性、天气和光照变化复杂性以及监控设备分辨率层次不齐等原因,使得车牌像素区域模糊或不完整,甚至视频中不存在车牌信息,给通过车牌识别寻找目标车辆的任务带来了极大挑战。在交通监管和刑侦安防工作中,采用车辆本身结构外观信息对其再识别具有重要的研究意义和应用价值。

不同于车辆属性识别等判别任务,车辆重识别的本质是近邻重复图像匹配与检索的问题,任务是通过给定的车辆图像,根据外观和辅助信息(如型号、获取时间地点信息)检索到在其他视角下的该车辆图像或视频。经典的车辆重识别流程包括车辆图像或视频的获取、车辆检测、特征提取与表达、相似性度量计算和检测结果的展示(刘凯等,2020)。自20世纪90年代以来,交通道路上安装了大量监控设备,基于视频或图像的车辆检测、跟踪在交通监管和刑事侦测中越来越重要,车辆重识别技术也得到了更多的关注。按照实现方式的不同,车辆重识别技术大致分为基于传感器的方法、基于手工设计特征的方法和基于深度学习的方法。

基于传感器的车辆重识别算法是最早提出的一类识别检索算法,使用各种类型的硬件传感探测器(如红外线、超声波和毫米波等)提取车辆的特征信息,然后进行相似性匹配识别车辆的基本信息。Sanchez 等人(2011)和 Charbonnier 等人(2012)都采用无线磁感应器收集车辆的不同特征,并提取感应器中的时间信息,将二者结合训练高斯极大似然分类器,实现对车辆的匹配和识别。Kwong 等人(2009)在此基础上进行优化,实现了对多路口车辆行驶时间的实时预测。Jeng 和 Chu(2013)利用感应线圈获取车辆的特征信息,并通过插补样条数据以及改进高占用车道的车辆时间窗大小提升车辆重识别的性能。Ndoye 等人(2009)不仅利用感应线圈提供数据完成了车辆重识别任务,而且还可以估计车辆行驶时间。Ali 等人(2013)通过整合多个感应线圈的数据,构建了集车辆重识别和车辆异常变道监控为一体的综合应用系统。借助一些新的传感技术,车辆重识别的算法得到进一步升级,如基于无线射频识别的传感器模型(Prinsloo 和 Malekian, 2016),可以在高速公路收费站进行安装和监控。虽然基于传感器的车辆重识别技术获取信息多样,但是需要预先在道路上部署许多硬件设备,成本巨大,并且容易受到客观环境影响。

基于手工特征的方法采用人工设计算子,通过提取车辆的颜色和纹理等各类特征对车辆进行识别。Ferencz 等人(2005)针对车辆类间变化小,类内又受到光照和姿势位置影响等问题,提取车辆位置和边缘对比度,分别用类间和类内车辆特征集训练和约束分类器,获得了优异性能。为了解决小数据

量样本不利于归纳泛化的问题, Son 等人(2007)构建了一个大型数据集, 采用各类算子提取多种特征向量表达车辆信息, 设计了一个实时的车辆重识别匹配系统, 并验证了该模型在拥挤复杂场景下车辆检索的有效性。局部最大频次法(local maximal occurrence, LOMO)最初用于行人识别任务, Jeng 和 Chu(2013) 将其引入到车辆识别工作中, 利用 LOMO 计算得到的特征嵌入进行车辆相似度匹配和检索, 实现车辆的重识别。虽然特定手工特征算子能捕捉关键的车辆属性信息, 但是 2 维车辆视频或图像提供的信息终究是有限的, 对车辆进行 3D 建模可以收集车辆全方位信息。Woesler(2003) 结合车辆 3 维模型信息和顶面颜色信息, 校正了潮湿街道上的阴影和光反射等问题, 构建了较高精度和实时应用的车辆重识别模型。Zapletal 和 Herout(2016) 基于 3D 边界框技术将车辆多面图像进行拼接, 得到高信噪比的图像, 再采用颜色和方向梯度直方图算子对车辆进行线性分类和重识别。虽然基于手工特征的方法不再依赖传感器设备, 并取得了较好的识别性能, 但是需要巨大的调参工作量, 无法应对复杂多变的交通视频场景。

基于深度学习的车辆重识别方法与传统的机器学习算法不同。深度卷积神经网络(convolutional

neural network, CNN)通过目标函数自动学习特征表征, 在车辆重识别任务中取得了良好性能。车辆重识别作为近年的研究热点, 相关数据集是首要的研究基础。Kanacı 等人(2018) 针对目前存在的数据集都是高质量和细粒度的图像且经过大量人工测试与调整, 构建了更真实、多分辨率和多场景的车辆数据集, 旨在提升现有算法在实际应用中的鲁棒性和泛化能力, 并提出了多尺度车辆表征框架, 在该数据集取得了较好性能。Marín-Reyes 等人(2018) 提出基于度量学习的车辆重识别模型, 采用三元组约束训练网络, 能为相同标签的样本分配较高的相似性, 同时又使不同标签的样本在特征空间内保持距离。Bashir 等人(2018) 提出一种无监督的算法解决车辆重识别问题, 通过训练一个渐进式基础网络, 将深度学习模型获得的表示向未标记的数据集转移。随后又提出了通用的 Vr-proud 模型框架(Bashir 等, 2019) 将特征提取的 CNN 结构与自定进度的渐进式无监督学习相结合, 既加快了学习算法的收敛, 又提高了识别检索的准确率。Brown 等人(2020) 提出基于 smooth-AP 平滑损失函数的车辆重识别算法, 实现了大规模车辆检索任务的端到端学习, 大幅提升了性能。

车辆检索和重识别的关键技术发展过程如图 4 所示。

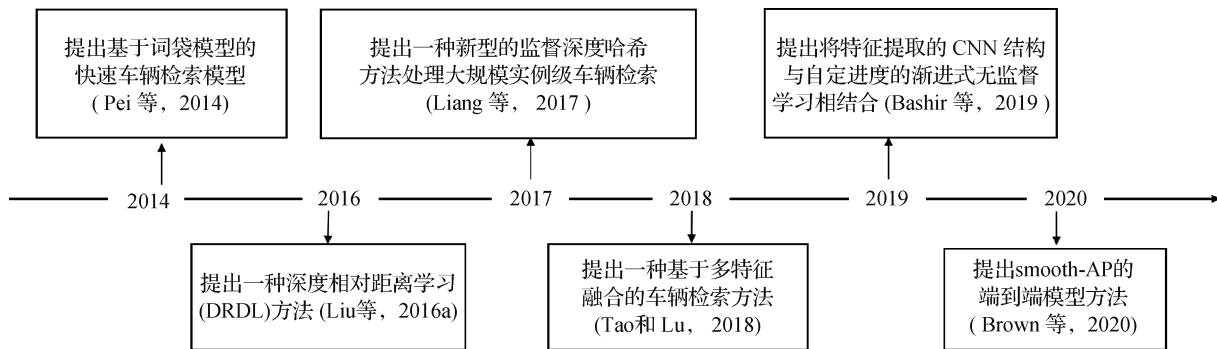


图4 车辆检索和重识别研究进展

Fig. 4 The recent advances in the retrieval and re-identification of vehicle

1.2 交通视频人员结构化

人员结构化分析是交通视频大数据处理的重要组成部分。目前, 人脸结构化和行人结构化是交通视频人员结构化分析中的两个重要研究方向, 重要功能是分析人脸或者行人的一些表观属性, 比如年龄、性别、口罩、背包、衣着颜色和长短等。

1.2.1 人脸结构化分析

人脸结构化分析可细分为基于注意力机制、基

于级联式结构、基于多任务学习、基于辅助信息的人脸结构化分析, 以及人脸结构化分析中的数据扩充与均衡。

1) 基于注意力机制的人脸结构化分析。由于注意力模块具有结构功能明晰、计算资源消耗少和网络收敛速度快的优点, 基于自注意力机制的人脸结构化分析方法受到越来越多的关注。Hand 等人(2018) 提出时间一致性约束和运动注意力机制两

种方法提高视频中属性响应的一致性和准确性,这两种方法都适用于弱监督学习,即在一个序列中只需要一个帧的属性标签(称为锚帧)。在时间一致性约束中,将非锚帧的网络响应移动到锚帧对每个序列的响应上,从而使属性预测更加稳定和准确。在运动注意力机制中,使用锚定和非锚定视频帧之间的运动作为注意力机制,丢弃非锚定视频帧中未发生运动部分的信息。两种方法都取得了较好效果。在极具挑战性的局部遮挡的人脸表情识别技术中注意力机制也得到运用。Ding 等人(2020)提出一种以局部标记为导向的注意力分支,用于从遮挡区域中发现和消除被遮挡区域中的被遮挡特征,使其不再参与识别任务。首先生成一个注意力图指示一个特定的面部区域是否被遮挡,并引导模型关注非遮挡区域。其次利用一个面部区域分支将特征映射划分为不重叠的面部块,并使每个块独立预测表情。这就产生了更多样化和更有区别的特征,使表情识别系统能够在面部局部遮挡的情况下得以恢复。非约束环境下准确识别年龄和性别是非常困难的,主要是由于在分辨率上的强变化性。受生物学和注意力机制在视觉问题和细粒度识别方面的启发,Rodríguez 等人(2017)提出一种新的前馈注意力机制,能够发现一个给定的面孔上信息最丰富和最可靠的部分,从而提高年龄和性别分类精度。

2) 基于级联式的人脸结构化分析。基于级联式的网络架构逐渐成为机器学习领域一个重要的研究方向。级联式学习是一种多学习器级联的学习方法,可以有效提高机器学习的泛化能力,并在许多应用领域展现了良好性能。同样,基于级联式的网络架构也逐渐扩展到人脸识别领域,并且明显提升了系统的识别性能。Ding 等人(2018)提出一种可以实现特定属性面部区域的定位和属性分类而无需对齐的级联网络架构,首先设计一个弱监督的面部区域定位网络自动检测特定属性的区域(或部位),然后通过区域切换层和属性关系层分别构造多个基于局部的网络和基于整个图像的网络组合在一起,以进行最终的属性分类。

3) 基于多任务学习的人脸结构化分析。多任务学习的目的是通过在多个预测任务之间适当地共享相关信息提高预测任务的泛化性能。目前已经提出了许多基于多任务的面部属性估计的方法。Huang 等人(2020)提出一种新的基于人脸子区域的

年龄估计框架,框架中的每个子网都从人的面部区域获取两个图像作为输入,一个是全局人脸区域,另一个是重要的子区域,然后基于多数投票方法将来自不同子区域的预测进行组合,以达到最佳预测效果。在具有挑战性的年龄估计问题上,多任务深度学习作为与主要任务相关的辅助任务(如性别认知)也得到了较好应用。Yoo 等人(2018)提出一种条件多任务学习方法,在结构上将年龄变量分解为深度神经网络中性别条件年龄概率,同时针对缺乏具有离散年龄值的精确训练标签数据,提出一种从弱监督分类标签中增加精确标签数量的标签扩展方法,最后在两个公开数据集上验证了所提方法的通用性。

4) 基于辅助信息的人脸结构化分析。深度学习是由大数据驱动的,然而在很多任务上,其性能往往受限于数据集的规模。此时可以借助外部辅助信息,即从有限的数据中挖掘尽可能多的潜在特征信息辅助网络识别。Kalayeh 等人(2017)提出使用语义分割改进面部属性预测,核心思想是利用许多面部属性描述人脸局部属性。一个属性在一幅图像中出现的概率在空间上并不是均匀的。该方法利用语义分割获得的定位线索,将属性预测的注意力引导到不同属性出现的区域,最终除了识别之外,还能够对属性进行定位。Hand 和 Chellappa(2017)提出可以通过在所有属性中共享最低层信息、共享相关属性的层次结构和在多任务深度卷积神经网络(multi-task CNN, MCNN)上建立一个辅助网络等3种方式使用 MCNN,充分利用属性关系改善每个属性的最终分类。He 等人(2018)提取一种包含面部局部信息和面部纹理信息的抽象的人脸图像,该图像是由生成对抗网络辅助生成,基于此图像,建立了基于原始人脸图像和人脸提取图像的双路径人脸属性识别网络。在提取局部化图像时,改进了人脸特征的识别,特别是小图像的特征识别。在人脸表情识别领域,Zhang 等人(2018)设计了一个有效的多任务网络,能够学习丰富的辅助属性,如性别、年龄和头部姿势,而不仅仅是面部表情数据。传统的监督训练需要数据集附有完整的标签(例如所有样本必须标注性别、年龄和表情),该方法通过一种新的属性传播方式放松了这种要求,尽管不同的数据集分布不同,但是这种方法允许利用异构属性源之间固有的潜在联系。在单幅图像的年龄估计问题上,Savov

等人(2019)提出一个深度学习架构,利用3维人脸重建获得一个稳健的年龄估计,借助表情、位姿、光照、反射率和几何感知等辅助信息从单一2维图像重建3维人脸,学习了有效的表征信息,最后将3维人脸重建网络与基于外观的年龄估计网络相结合,将人脸重建特征与视觉特征共同学习,达到了很好的效果。

5)人脸结构化分析中的数据均衡。人脸属性数据往往呈现高度倾斜的类分布,即大部分数据属于几个多数类,而少数类只包含少量实例。为了缓解这一问题,Huang等人(2016)进行了广泛而系统的实验验证经典方案(如类重抽样或成本敏感训

练)在类不平衡数据表示学习中的有效性,发现通过强制建立一个深度网络保持集群间和类别间的边界,可以学习到更具代表性的深度表征信息。Hand等人(2018)为深度网络引入一种新颖的选择性学习方法,根据每个标签的期望分布自适应地平衡每个批次中的数据,使网络能够学习更强大的属性模型。Dong等人(2019)构建了一个基于批量递增的少数(稀疏采样)类矫正深度学习模型,在模型训练过程中对多数(频繁采样)类进行难样本的挖掘,同时通过在迭代的批处理学习过程中发现少数类的稀疏采样边界来最小化多数类的主导效应。

人脸结构化分析的研究进展如图5所示。

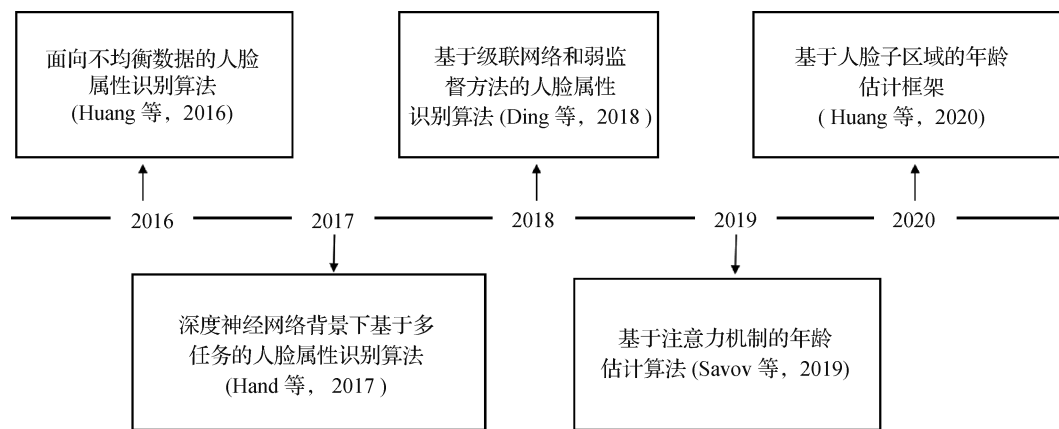


图5 人脸结构化分析研究进展

Fig. 5 The recent advances in the structural analysis of human face

1.2.2 行人结构化分析

行人结构化分析可以细分为基于区域、基于注意力机制、基于相关性探索以及基于辅助信息的行人结构化分析。

1)基于区域的行人结构化分析。随着卷积神经网络在计算机视觉任务上的出色表现,利用图像的全局和局部信息进行行人属性识别的研究受到广泛关注。Abdulnabi等人(2015)提出采用多任务卷积神经网络(multi-task CNN, MTCNN)解决行人属性识别问题,改变了过往忽略行人属性之间关系的情况,其中每一个深度神经网络将预测一个二分类属性,并生成特定于属性的特征表示,以此获得局部信息。值得一提的是,多任务学习允许模型在不同的属性类别之间共享视觉信息,同时分组信息促进同一组内的属性共享特征统计数据,并阻止不同组中的属性分享信息以增强性能。其他研究也采用了相似的思想,不再局限于仅关注图像整体,而是通过联合

局部和全局信息获得更准确的属性识别性能。Gkioxari等人(2015)通过研究局部信息对人体属性识别的重要性,提出基于全局和局部的动作识别和属性分类方法,抓取一系列姿势下的人体部分,在动作和属性分类任务上都取得了当时最好的结果。Park和Zhu(2015)提出基于属性语义的组合模型,在统一的组合层次结构中联合表示检测对象的局部及其语义属性,通过同时执行部分定位和属性识别将模型应用于人体解析任务,以提高属性预测任务的性能。Fabbri等人(2017)提出生成对抗模型应对视频监控场景中常见的低分辨率和遮挡问题,模型包含3个网络,其一专注于对给定的全身图像进行人体属性分类,另外两个生成网络致力于在低分辨率的情况下提高图像的分辨率和重建被遮挡图像中缺失的人体部分。

2)基于注意力机制的行人结构化分析。行人属性识别作为一项具有挑战性的任务,具有大姿态变化、复杂背景和多种摄像机视角等众多干扰因素,

为了选择重要的和可区分区域或像素来对抗变化,大量研究将注意力机制引入行人属性识别算法中,引导网络将其资源集中在图像相关的空间部分。Li 等人(2018b)在研究中发现,以往的再识别任务往往假设模型具有良好对齐的行人检测框图像作为输入,并依据受限的选择机制校准失调的图像。而实际任务中,当任意对齐的行人检测框中存在较大姿态变化的人像时,这种处理方式往往达不到理想的效果。因此,提出了致力于优化行人在失调图像中重新辨识的注意力机制,该模型可最大限度地利用不同类型的注意力信息,在任意的行人检测框中,皆可同时学习软像素级别的注意和硬区域级别的注意,同时优化特征表示,以更好地利用注意选择和特征识别之间的相关互补信息。为了解决行人属性识别中存在的多标签性质、大的底层类不平衡和缺乏空间注释等困难,Sarafianos 等人(2018)提出深度不平衡属性分类的行人属性识别方法,从网络的多个阶段中提取视觉注意力掩膜并在评分级别上将信息聚合,从而使得模型学习到更多的判别特征表示,并且针对该类任务中常见的类不平衡情况,通过分配焦点损失的加权方差处理不平衡的类,使模型得以持续高效工作。针对注意力掩膜导致的属性预测在后续迭代轮次中的高方差现象,引入注意力损失函数,采用惩罚机制降低来自注意力掩膜的影响,最终实现了当时的最佳效果。

3) 基于相关性探索的行人结构化分析。随着行人属性识别研究的逐步深入,研究者不再满足于低级的属性特征,尝试通过探索属性之间的相关性以提升模型性能。Diba 等人(2016)提出一种卷积神经网络挖掘中级(mid-level)图像补丁,并以特征

学习和补丁聚类的迭代对补丁进行更新,补丁可对相当小和不显眼的对象和特征进行检测,以解决识别人类行为和确定人类属性这两个需要细粒度分类的任务。Wang 等人(2017b)提出联合递归方法学习行人属性级的顺序相关性和图像级的上下文信息,通过发现行人图像在端对端编解码递归网络中的顺序依赖,改进了在质量较差的小训练数据下的属性识别性能。Yang 等人(2020)提出了多层特征嵌入模型(hierarchical feature network, HFN),通过结合属性以及 ID 信息,实现细粒度的特征嵌入。

4) 基于辅助信息的行人结构化分析。行人属性识别任务的性能往往因为没有足够的信息可供提取而不甚理想。研究者开始将目光投向其他方法,而不局限在属性本身,借助辅助信息即从多个维度挖掘尽可能多的潜在特征信息以提升模型的识别性能。Zhang 等人(2014)提出一种采用卷积神经网络和利用局部信息的人体属性分类网络——深度属性建模的姿态对齐网络(pose aligned networks for deep attribute modeling, PANDA),通过卷积神经网络提取特征,解决了过往基于区域方法只能使用底层特征的问题,并将提取的特征进行融合,最终使用线性支持向量机作为分类器,产生属性分类器。模型可从较小的数据集中学习到区分能力比较强的形态归一化特征,并在当时取得了有竞争力的性能。Sarfranz 等人(2017)提出的视觉敏感行人属性模型联合预测了行人的粗略姿态,并学习用于特定视图的多标签属性预测,结果显示除身体部分、图像中的属性空间上下文或一般场景上下文之外,粗略的姿态信息也是一个简单但高度相关的线索,可用于可靠的属性推断。

行人结构化分析的研究进展如图 6 所示。

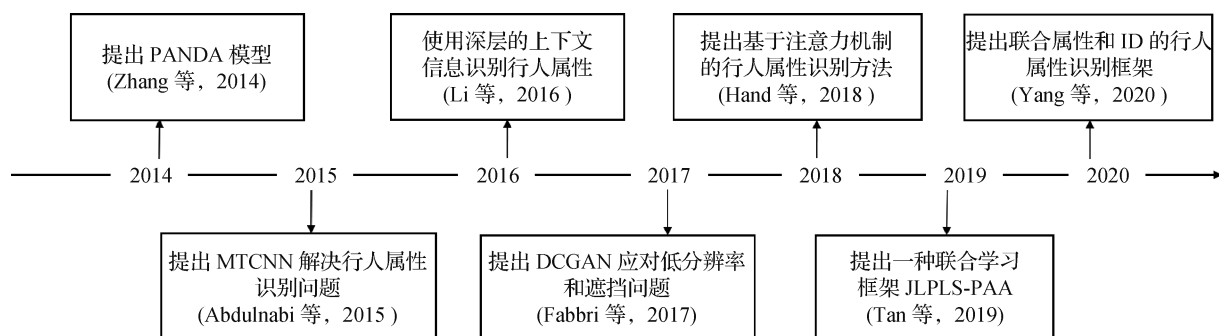


图 6 行人结构化分析研究进展

Fig. 6 The recent advances in the structural analysis of pedestrians

1.3 交通视频行人行为分析

行人行为分析任务是指对一段时间内的行人行为进行特征提取并进行识别、理解以及预测。传统方法通常将特征的提取以及特征的识别作为两个独立的过程,二者串行实现对行人行为的理解及预测。而基于深度学习的相关方法将二者统一在一个端到端的学习框架下,在对行为的空间特征、时序关系进行建模的同时,完成行为的识别、理解以及预测。

1.3.1 传统手工特征提取方法

几何特征是最早人体行为识别的特征。Fujiyoshi 等人(2004)通过星状图构建了人体的四肢和头部 5 个顶点的人体姿态,以其与重心构成的矢量作为特征。行为的运动信息是传统方法经常考虑的特征。Bobick 和 Davis(2001)抽取运动能量图(Bobick 和 Davis,1996)和运动历史图(Weinland 等,2006)的表征行为。基于 Harris 空域特征点,Schuldt 等人(2004)提出了 3 维时空兴趣点,并统计其周围像素的直方图,以此作为描述动作的特征。由于该方法检测到的稳定兴趣点数量少,Dollar 等人(2005)提出在时间和空间维度上进行 Gabor 滤波,这样随局部邻域块的尺寸大小的变化检测到的兴趣点数目亦会改变。Chaudhry 等人(2009)在选定方向上将光流场整理成运动矢量,对其归一化后用作描述行为的特征。

基于方向梯度直方图(histogram of oriented gradient, HOG)特征,Klaeser 等人(2008)设计了 3 维形式的 HOG 特征,即 HOG3D。该特征由规则多面体组成,且在多尺度下可进行快速密度采样。Rapanzikos 等人(2009)通过低通和高通滤波选择感兴趣的时空点。Knopp 等人(2010)提出了 3 维形式的 SURF 特征,即 3DSURF,每个单元包含全部 Haar-wavelet 特征。Rohrbach 等人(2012)证实对时空描述子密集采样可以提高人体行为的分类效果。Kataoka 等人(2014)基于行为 HOG 描述的共生性提出一种密集轨迹描述,取得了较好的识别效果。Packer 等人(2012)在识别阶段采用 latent structural 支持向量机分类器。

1.3.2 基于深度学习的特征提取方法

现有研究表明,深度网络(Hinton, 2007; Schmidhuber, 2015)在样本足够多的时候更适合行为识别。Chéron 等人(2015)采用卷积神经网络建

模并捕获单帧数据和光流数据的运动信息。Varol 等人(2018)将 3D 卷积应用于固定窗口的视频块内。Karpathy 等人(2014)采用多分辨率 CNN 提取行为特征,其输入为两组独立数据流,即低分辨率数据流和原始分辨率数据流。

为进一步建模人体行为的时序关系,Veeriah 等人(2015)提出一种微分递归神经网络,强调信息的变化增益是由相继帧间的显著运动引起的。Ng 等人(2015)使用(long short-term memory)进行行为的时序关系建模,在 UCF101 数据集上的识别率达到 82.6%。Donahue 等人(2015)提出一种长时递归卷积神经网络(long-term recurrent convolutional networks, LRCN),首先采用 CNN 提取单帧图像的特征,然后顺序输入长短期记忆网络(LSTM)以获取人体行为在空间和时间维度上的联合表征,在 UCF101 数据集上的识别率为 82.92%。

上述模型主要考虑对时间特征进行建模。与之不同,Stewart 等人(2016)利用 LSTM 检测交互对象,即在每帧图像中利用边界框间的上下文关系顺序输出交互对象的边界框,但只能检测单个物体。Ni 等人(2016)提出一种检测精细化算法,首先检测场景中的易检测目标,然后利用上下文信息精细化检测其他目标甚至难以检测的目标,在厨房行为数据集 MPII 上取得了 58.9% 的 AP(average precision)精度。Singh 等人(2016)提出一种多源双向 RNN(recurrent neural network)模型,利用 CNN 分别提取全帧以及显著性区域的运动和表观特征,采用双向 LSTM 将多源 CNN 的输出连接起来,在 MPII 数据集及其自建的购物行为数据集 MERL(Mitsubishi Electric Research Laboratories) Shopping Dataset 上的 mAP 精度分别为 41.2% 和 80.3%。

Carreira 和 Zisserman(2017)收集了用于行为识别的大型 kinetics 数据集,在此基础上,提出一种将 2 维卷积网络扩展成 3 维卷积的思想,通过 3 维卷积网络建模行为间的时序关系。Tran 等人(2018)提出一种新的卷积形式建模行为的时空信息,提出(2 + 1)维的结构更好地处理时空信息。受人类视觉神经系统的启发,Feichtenhofer 等人(2019)提出采用双支路视频信息处理方法,将低帧率输入到 slow 网络捕获空间语义信息,将高帧率输入 fast 网络捕获运动信息。Ryoo 等人(2020)针对不同的子网络和输入模态(例如 RGB 和光流)及时间分辨

率,提出一种可学习架构,可提供通用方法学习不同模态特征表示输入间的连接,并针对目标任务进行优化,在用于行为分析及视频理解时,能通过

视频中的表现信息和行为的时序线索更好地生成特征。

行人行为识别的国外研究进展如图7所示。

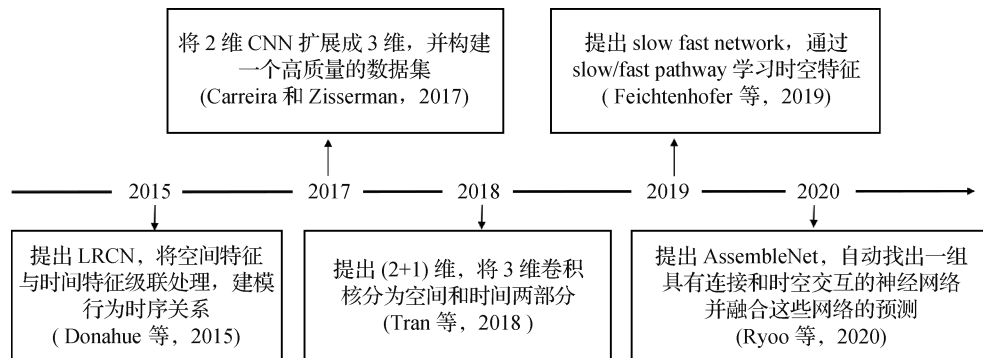


图7 行人行为识别的研究进展

Fig. 7 The recent advances in the structural analysis of human behaviors

2 国内研究进展

2.1 交通视频车辆结构化

2.1.1 车辆检测

相比于传统的车辆检测算法,基于卷积神经网络的方法具有高精度和高鲁棒性的特点,得到了广泛认可。目前,国内研究主要集中在设计精巧的卷积神经网络模型来提升车辆检测的速度。Chu 等人(2018)将不同车辆的属性信息进行共享,利用邻域关系对相邻检测框进行投票以提取更加准确的车辆特征。Tang 等人(2017)首先利用视频跟踪算法对车辆视频片段进行标注并进行数据增强,再使用单阶段的YOLO算法实现对车辆的实时检测。Dai 等人(2019)提出基于两阶段回归的层级物体检测模型 HybridNet,在第1阶段和第2阶段都采用回归模型,并在中间加入过渡阶段用来将第1阶段产生的预选框映射到高分辨率的特征图上,为第2阶段的决策优化提供更优的特征,比二阶段的检测方法速度更快,精度也有所提升,对车辆检测更为有效。

2.1.2 车辆属性识别

1) 车牌识别。国内对车牌识别的研究主要集中在基于深度神经网络的方法。Wang 和 Tu(2020)提出一个框架,将全卷积网络与广泛的学习系统相结合,得到的网络通过融合多尺度和层次的特征来进行随机定位的目标检测。讨论了深度学习在车牌识别中的应用,深入讨论了车牌偏斜、图像噪声以及

车牌模糊3个难点工作,并分析了当前车牌检测算法和字符识别算法的优缺点。Zhang 等人(2020a)提出一种基于孪生神经网络的车牌识别算法,可在大量交通流视频中有效识别出车牌。Zhang 等人(2020b)进一步提出基于时空注意力机制的图像帧级别特征提取方法,联合车牌识别和图像质量评分,得到了较好的效果。

2) 车型识别。与车牌识别类似,国内车型识别的研究聚焦于神经网络相关的方法。Jia 等人(2019)提出一种基于改进胶囊网络的车辆图像识别算法,使用期望最大化算法进行迭代训练。Chen 等人(2019)提出一种基于反馈增强的多分支卷积神经网络,应用于智能交通和无人驾驶两种实际场景,并创建 MVVTR (multi-view vehicle type recognition) 数据集解决多视图粗粒度问题。Guan 等人(2019)提出将迁移学习和 Radon-CDT 技术相结合的方法,将最佳传输理论应用于2维信号的模式识别,提高了实际情况下的识别精度。

3) 车辆颜色识别。车辆颜色识别是车辆结构化信息中的重要组成部分,在车牌等信息缺失情况下,可作为一种重要的补充信息。Hu 等人(2015)将空间金字塔策略与深度学习框架相结合,引入空间特征信息,进一步提高了识别精度。Chen 等人(2014)提出一种新颖的卷积核利用方式,首先将3维彩色图像投影到2维界面中,随后利用贝叶斯朴素分类器将颜色直方图转换成基于类的特征,同时提出一种基于隐式选择的方法用于识别感兴趣区

域,并进行预处理以克服图像质量下降的影响。最后通过具有不同权重的子区域来选择车辆图像中的感兴趣区域,提升识别精度。

2.1.3 车辆检索与重识别

1) 车辆检索。在车辆检索领域, Tian 等人(2020)认为在车辆检索任务中负样本的数量远高于正样本的数量,应充分挖掘负样本中的信息,并基于该观点提出一种基于结构感知的四重损失函数,通过全局信息和局部信息相结合的方式提高了预测精度。Shi 等人(2019)以汽车前挡风玻璃颜色作为重要特征,通过提取挡风玻璃的颜色直方图特征进行检索。Su 等人(2016)提出一种基于感兴趣区域颜色和类型匹配的车辆检索方法,利用纹理特征对车辆的挡风玻璃和车牌进行定位,在定位获取的感兴趣区域进行特征提取,并通过特征之间的距离度量确定检索的结果。Pei 等人(2015)将车辆图像作为检索源,以深度网络框架作为检测器用于检测图像的特点和筛选描述符,最后以图像之间的余弦相似度来度量相似性。该算法可整体部署于智能手机上,在保证算法精度的同时,具有良好的实时性。Zhu 等人(2017)提出基于特殊度的局部描述子特征选择模型,首先提取图像的局部描述特征,然后根据特征的梯度和空间梯度能量分布定义特征的度,再结合特征的显著性和先验信息自适应选择特征,最后采用基于紧凑描述符的视觉搜索法将梯度变化应用到特征选择中,提高了车辆检索的准确性和效率。Wei 等人(2020)提出一种改进的基于年检标签区域特征的快速仿射匹配彩色图像检索方法,实现了对车辆的精确检索。该模型分别采用 GoogLeNet、OverFeat、VGGNet (Visual Geometry Group network) 和 ResNet50 卷积神经网络框架提取年检标签特征信息,再将其融合。同时,为了解决真实监控场景中的光照变化问题,该模型引入区域色彩常数和色饱和度特征。在不同数据集上的实验都展示了该算法的有效性。

2) 车辆重识别。随着计算机视觉算法在交通视频领域的广泛应用,车辆重识别领域也取得了新的突破。He 等人(2019)提出一种基于车辆部分正则化特征鉴别的识别算法,通过将车辆局部特征与全局特征相结合,提高了识别算法对细微差异的感知能力。Zhou 和 Shao (2018) 提出视图感知与多视图推理模型解决车辆重识别中的多角度差异问题,

通过视点感知注意力模型,获取多重单视点车辆图像核心区域,结合对抗性训练架构实现有效的多视点特征推理,最终完成单视点特征到全局多视点特征的转化。Chu 等人(2019)从人对车辆的观察行为出发,提出一种新颖的视觉感知度量学习方法,即对相同的对比视角与相对的视角特征进行组合学习,从而解决车辆重识别中观察视角不统一的问题。Li 等人(2017)认为在同型号不同车辆的识别问题上,车窗与车脸区域存在较大差异,因此选择以上区域进行特征提取并融合,并根据融合特征之间的距离度量进行分类识别。Li 等人(2017)认为单一的车辆特征识别算法难以应对不同场景下的车辆再识别任务,提出基于车辆多维特征融合的算法模型,综合车辆颜色、纹理和高级语义特征,以实现识别的最佳性能。另外, Cui 等人(2017)对车辆颜色、车辆模型和挡风玻璃上标志贴等多重信息分别进行网络特征提取,再进行特征融合,认为多维特征的融合将进一步提高算法的准确性和鲁棒性。

2.2 交通视频行人结构化分析

2.2.1 人脸结构化分析

1) 基于注意力机制的人脸结构化分析。近年来,国内在人脸结构化分析方面开展了注意力机制的研究。Zhang 等人(2019)将注意力机制运用到单幅人脸图像的人脸年龄估计中,提出一种基于注意力长短期记忆网络的野外细粒度年龄估计方法,将多级残差网络模型与 LSTM 单元相结合,提取年龄敏感区域的局部特征,并结合全局特征,有效提高了年龄估计精度。Pei 等人(2019)将注意力机制运用到年龄估计算法中以提高预测准确率,提出一个用于年龄估计的端到端结构——空间索引注意力模型,能够从原始的面部表情视频中同时了解年龄和表情,不仅使模型专注于信息框架和面部区域来提高性能,而且提供了空间面部区域和时间框架与年龄估计任务之间的可解释的对应关系。

2) 基于级联式的人脸结构化分析。基于级联式的方法在年龄估计领域也有应用。针对非约束环境下的人脸脸部的复杂变化, Liu 等人(2015)提出一种新的属性预测深度学习框架,级联了两个 CNN,即 LNet 和 ANet,与属性标签一起进行微调,但是经过不同的预训练。LNet 用于定位人脸具体位置,即定位头部和肩部,在此基础上进一步定位人脸区域;ANet 用于提取人脸属性信息,并将提取的

特征输入到线性支持向量机中进行属性预测。该框架不仅大幅领先于最新技术,而且揭示了学习人脸表征方面的事实。Wan 等人(2018)提出5种基于卷积神经网络的年龄估计级联结构框架,所有框架均由辅助的人口信息学习和指导,且每个级联结构框架包含在一个父网络和几个子网络中。例如一个应用的框架是一个利用性别信息训练的性别分类器,然后分别用男性样本和女性样本训练两个子网络。此外,利用高斯过程回归的级联结构框架提取特征,进一步提高了年龄估计的性能。Xie 和 Pun (2020)提出分解的思想,将基于排序思想的年龄估计任务分解成多个二进制分类任务,该方法具有深度有序集成和两阶段聚合策略的特点。

3) 基于多任务学习的人脸结构化分析。多任务学习的目的是通过在多个预测任务之间适当地共享相关信息来提高预测任务的泛化性能。在深度神经网络背景下,这种想法通常是通过手工设计的网络架构来实现的。Wang 等人(2017b)首次提出一种多任务深度神经网络,用于联合学习面部识别和面部属性预测任务。然而这种可能的多任务深层架构的空间是组合式的,往往很难通过人工探索这个空间来设计最终的网路架构。Lu 等人(2017)提出一种用于设计紧凑型多任务深度学习架构的自动方法,从浅层网络开始,然后在训练过程中以贪婪的方式动态扩展,通过迭代地执行此操作,创建了树状的深层体系结构。在该体系结构上,相似的任务驻留在同一分支中,直到位于顶层为止。该方法生成的模型快速紧凑,能够达到较好效果。在年龄估计领域,考虑到人脸老化过程的复杂性以及数据采集和标记的难度,人脸图像的年龄估计具有很大的挑战性。Han 等人(2018)提出一种基于深度多任务学习(deep multi-task learning, DMTL)的方法来联合估计一个单一的人脸图像的多个异构属性。具体来讲,使用卷积神经网络(CNNs)处理属性的相关性和异质性,该网络学习所有属性的共享特征和异构属性的分类特征。同时,引入无约束人脸数据库 LFW+。LFW+ 是公共人脸数据集 LFW (labeled faces in the wild) 的扩展,包括具有异质性的人口统计属性(年龄、性别和种族)。Liu 等人(2020a)提出一种简单有效的多任务学习(multi-task learning, MTL)网络 CR-MT (classification and regression-multi-task),将分类和回归结合起来用于年龄估计,其中分

类充当回归的辅助任务。通过共享两个任务的信息表示学习,MTL 可以提高年龄回归任务的泛化性能。

4) 基于辅助信息的人脸结构化分析。由于在年龄估计中每幅图像都是由多个个体标记的,所以导致了个体年龄的不确定性。这种不确定性使得该年龄估计任务不同于一般的年龄估计任务。Yang 等人(2015)提出一种使用深度卷积神经网络并借助分布式损失函数的方法进行年龄估计,使用分布作为训练任务,可以利用人工标记产生的不确定性,学习比使用年龄作为目标更好的模型。Hu 等人(2017)提出一种新的学习方案,利用 Kullback-Leibler 散度嵌入年龄差信息,使网络自适应地将熵损失和交叉熵损失应用于每幅图像,使数据分布呈现单峰形状。通过这两种损失函数的有效组合设计,可以驱动神经网络从年龄差异信息逐渐学习到年龄。

5) 人脸结构化分析中的数据均衡。由于人脸老化过程的复杂性以及数据采集和标记的难度,数据集的缺乏对人脸图像的年龄估计提出很大的挑战性。Gao 等人(2017)提出一种深度标签分布学习方法,有效利用了特征学习和分类器学习中的标签模糊性,有助于在训练集很小的情况下防止网络过度拟合。Sun 等人(2019)提出一种基于相似感知的深度对抗学习的面部年龄估计方法。受限于有限的训练样本,以往的工作很可能导致次优化,而该模型根据现有的训练样本寻求一些无法观察且合理的困难样本,这通常会增强所学特征描述符对年龄的区分性。该方法利用重构的难样本空间中的平滑老化模式进行可靠的年龄估计,取得了优越的性能。考虑到整体回归模型是针对不平衡训练数据的,而分治法则高度依赖于年龄分类的效果,年龄分类由于存在交叉年龄相关性,通常会产生边界效应。针对小样本人脸数据集,疏影等人(2020)提出一种融合自监督学习和对抗生成网络的模型,有效扩大了人脸样本训练数据量,显著提升了人脸属性识别性能。Liu 等人(2020a)提出一种结合分类和回归进行年龄估计的简单有效的多任务学习网络,其中分类作为回归的辅助任务。通过对年龄回归任务的信息表示学习,可以提高年龄回归任务的泛化性能。并且借助于分类中的良好年龄数据分区,可以更准确地拟合异构或不平衡的年龄数据。

2.2.2 行人结构化分析

1) 基于区域的行人结构化分析。从图像区域

出发对行人属性进行识别的方法在国内也有研究。Li 等人(2015)提出深度单属性识别和深度联合属性识别两个深度学习模型解决行人属性识别,一方面应用深度单属性识别模型将每个属性都视为一个独立的组件,针对每个属性训练一个二分类网络进行识别。另一方面,为利用属性之间的关系,将行人属性识别视为一个多标签分类任务,应用深度联合属性识别模型来联合识别多种属性。在此模型中,一个属性的识别会从其他属性中获得增益。在流行的行人属性数据集的实验验证了该方法在当时的领先性。Liu 等人(2018)提出一种新型的本地化引导网络,根据预先提取的信息和属性位置之间的相关性,为局部特性分配了特定于属性的权重,整个模型既有局部特征对每个特定属性的自动学习,也有全局特征的交互,将全局信息和局部信息相结合,共同进行行人属性识别。此外,刘弋锋等人(2019)提出基于多任务学习的行人属性识别方法,不仅增强了模型的泛化能力,同时提升了行人属性的识别性能。

2) 基于注意力机制的行人结构化分析。基于注意力的机制能引导网络将资源集中于输入图像属性相关的信息,这一特性在国内引起了广泛关注。Liu 等人(2017)提出一个基于注意力机制的深度神经网络,从多方向将多级注意映射给不同的特征层,捕获从低级到语义级的多级注意力,探索注意力特征的多尺度选择性,使最终特征表示更丰富化。Tan 等人(2019)提出解析注意、标签注意和空间注意联合的学习框架,通过多任务的方式结合 3 种注意力机制进行学习,提取互补和相关的特征用于行人属性分析,效果显著。Wu 等人(2020)提出新型的注意力机制用于行人属性识别,通过由粗到细的注意力机制改进属性定位,在粗步骤中构建一个自掩模块大致区分和减少干扰,在细步骤中应用一个蒙面注意分支进一步消除无关区域,使特征学习更加准确。

3) 基于相关性探索的行人结构化分析。属性的相关性作为提升行人属性识别性能的重要因素,一直在识别领域中广泛使用。Li 等人(2016)所提方法与先前的差异主要在于使用了深层次化的上下文关系,使模型可从所有姿态检测中选择最具属性描述的人体部分与整个身体相结合作为姿态归一化的深度表示,并得到一种基于自适应选择人体部分的深度模型,可用于无约束图像的人体属性识别。针对先前研究只考虑属性级特征嵌入情况而在复杂

环境下鲁棒性较差的问题,Yang 等人(2020)提出分层属性嵌入框架,通过结合属性和身份信息学习细粒度的特征嵌入,使具有相同属性和相同身份的样本都会更紧密地收集,以此限制属性难样本对特征嵌入的影响。Tan 等人(2020)提出应用于行人属性识别的属性和上下文联合学习框架,包含属性相关性模块(attribute relation module, ARM)和上下文相关性模块(contextual relation module, CRM),其中 ARM 构造了具有特定属性特征的属性图,并进一步利用图卷积网络(graph convolutional network, GCN)探讨多属性之间的相关性,CRM 提出了图形投影方案,将 2 维特征映射投影到来自不同图像区域的一组节点中,并进一步利用 GCN 探讨区域间的上下文信息。该方法在 3 个属性数据集上验证了有效性和通用性。

4) 基于辅助信息的行人结构化分析。针对行人属性识别,现有方法经常使用端到端的多标签深度神经网络,然而行人身体结构知识尚未被妥善利用。因此,Li 等人(2018a)提出姿态引导的深度模型,将行人身体的结构知识引入到任务之中,首先从预先训练的姿态估计模型中提取姿态知识,再自适应地定位具有高级信息的身體部分提取特征,最终将基于局部的结果和基于全局的结果相融合,以进行最终的属性预测。Zhao 等人(2018)发现需要进行识别的属性倾向于语义或视觉空间相关性,提出分组递归学习的行人属性识别新方法,将属性按相关性进行分组,通过挖掘属性组中的语义和空间相关性逐组预测属性,再将这些语义组和检测身体区域得到的特征输入递归神经网络,利用组内互斥和组间相关性提高行人识别的性能,大量结果表明,基于行人属性数据集,模型在当时获得了最先进的结果。Tan 等人(2019)提出的应用于行人属性识别的解析注意、标签注意和空间注意联合学习框架将行人解析为不同的语义区域,以提供位置线索指示像素级的特定身体区域进行分割,再通过注意力模块将分离的特征整合,将拆分和聚合方式构建的解析注意机制作为辅助信息,以提升行人属性识别性能。

2.3 交通视频行人行为分析

国内对交通视频行人行为分析的研究逐步增多,研究的角度包括基于传统手工特征提取算法和基于深度学习特征提取算法。

2.3.1 传统手工特征提取方法

从人体深度图中,Yang 和 Tian(2014)采集了关

节点的3维坐标,以此形成人体轮廓并对人体行为的特征进行行为识别。但以传统几何特征为基础的行为识别算法受限于人体几何形状的建模,同时人体是典型的非刚体结构,很难用简单数学模型描述人体几何特征在行为发生期间的变化过程。运动信息即行为的时序关系特征,近年来开始应用于行为识别。Wang等人(2011)比较了几种常见的局部特征描述子,包括HOG3D、HOG/HOF(histogram of oriented gradient/histogram of optical flow)和extended SURF,实验结论是整合梯度与光流信息的描述子的行为识别效果较好,其中HOG3D的识别效果最好。自此,时空兴趣点的特征提取方法也用于行为识别,但在面对分类问题时,不同行为的传统特征间的差异性很小,影响分类效果。为解决这一问题,Lei等人(2012)融合了RGB及深度信息进行行为识别,在目标检测及行为识别过程中结合了时空兴趣点及词袋模型。传统运动特征提取算法在建模行为空间时,其类间边界变窄甚至出现交集,影响分类效果。为解决这一问题,Ni等人(2014)指出行为分析中融合目标物与运动信息的上下文关系可以提高识别性能,采用条件随机场对提取的上下文特征进行融合。在其基础上,Zhou等人(2015)利用局部特征描述子HOG/HOF以及光流轨迹信息对人体行为中的目标物及运动的上下文关系进行了建模。

2.3.2 基于深度学习的特征提取方法

与深度学习相关的行为识别算法主要包括基于卷积神经网络的算法、基于限制玻尔兹曼机的算法和基于递归神经网络的算法。Chen等人(2009)提出一种用于行为识别的空间—时间深度信念网络,使用限制玻尔兹曼机算法联合建模了空间抽取层和时间抽取层,从而提取适合行为识别的时空不变特征,在KTH数据集上获得了91.13%的识别率。同时考虑行为的空间和时间维度的特征提取,在传统CNN的基础上,Ji等人(2013)提出用于行为识别的3DCNN模型,整体结构为一种双流识别网络结构,包括静态帧数据流和帧间动态数据流两部分,均采用CNN进行特征提取,最后使用支持向量机进行识别。Liu等人(2014a)提出一种基于高斯过程的概率框架处理带有显式不确定性模型的回归问题。高斯过程应用于行为识别的一个主要挑战是在推理过程中需要对一个大的协方差矩阵进行反演,另外从流形的角度,数据空间的内在结构只受局部邻域的

约束,而距离较远的数据关系可忽略。因此,通过提出的L1范数正则结构和局部近似协方差权重更新方法设计了新的高斯过程协方差矩阵,该方法对数据噪声具有鲁棒性、自动稀疏性和自适应邻域性。Wang等人(2015)结合人工设计特征和two-stream深度特征,提出一种新的行为描述特征,将预训练的two-stream网络作为固定的特征提取器,并使用时空规范化以保证每个通道的数值范围近似一致,使用通道规范化以保证每个时空位置的描述向量的数值范围近似一致,之后用轨迹池化算法和Fisher向量构建特征,最后用支持向量机分类。由于计算量和GPU(graphics processing unit)内存的限制,以往的行为识别算法通常只能处理固定的64~120帧的序列,无法建模整个视频行为,为克服这一不足,Wang等人(2016)提出一种新的网络结构,使用稀疏采样方案在较长的行为序列上提取短片段,首先将视频划分为固定数量的分段,然后从每个片段中随机采样短片段,最终集成采样的行为片段信息,从而实现了建模整个行为视频的长时时序依赖关系的目的。毕晓君和冯雪赞(2018)将卷积神经网络与受限玻尔兹曼机相结合,有效挖掘了人体行为特征,实现了更准确的人体行为识别。马钰锡等人(2019)提出一种高效的行为识别框架,首先使用YOLO算法实时检测行人的位置,然后采用CNN提取行人特征,最后使用LSTM模型对行人行为进行预测。针对行为样本中数据依赖关系难以捕捉的问题,Wang等人(2018)提出一种非局部的non-local层,可以很好地捕捉到较远位置的像素点之间的依赖关系。考虑到模型对行为的时序信息的建模不够充分,而且由2维卷积到3维卷积的转换不可避免地会引入额外计算量,Lin等人(2019)提出一种基于时间维度特征移位的方法,降低了不同帧之间特征融合联合建模的计算开销。Zheng等人(2020)提出一种全局和局部知识敏感的注意力模型,包括统计注意力和可学习注意力两种类型注意机制,在HMDB51(human metabolome database)和UCF101(University of Central Florida)数据集上取得了国际领先的行为识别性能。

3 国内外研究进展比较

近年来,国内外在交通视频结构化分析的各个

方向展开了大量的探索和研究。针对国内外的研究现状进行比较,有助于该领域研究发展。

在车辆检测领域,国内起步较晚,国外早期多是利用传统手工特征算法对车辆检测进行研究。在深度学习兴起后,国内对车辆检测的研究越来越多,提出了大量车辆检测模型,国内的交通系统也建立了完善的车辆自动检测系统。当前,车辆视频检测与检索主要面临两个挑战:1)交通视频受现实场景中光照变化、车辆间遮挡、车辆阴影和摄像头抖动等因素影响,导致图像的语义信息丢失。2)现有数据集没有包含特殊检测场景,导致车辆检测准确性降低。

在车辆属性识别领域,车牌识别技术相比于国外,国内的研究不多,主要原因在于国内车辆的车牌规格、尺寸和颜色相对统一,对于复杂场景下车牌识别的需求量不大。国外对车牌的识别研究相对较多,对车牌位置检测、车牌字符分割、车牌字符识别都有很多的研究工作。随着国内车辆的普及以及自动驾驶的出现,也需要对车牌识别做进一步的研究,例如解决车牌偏斜、图像噪声以及车牌模糊等问题。在车型信息识别方面,国外的研究重点是如何简单地快速挖掘关于车型的车型特征,而国内研究更加侧重于如何提升车型识别准确率。在车辆颜色信息识别方面,国外的研究主要集中在如何利用数字图像处理的方法准确提取颜色特征,国内的大部分研究是利用深度学习算法探索如何在复杂环境下更准确地识别车辆颜色。

在车辆检索领域,主要围绕着检索速度与精度的平衡展开,国内外在传统的基于内容、颜色、语义信息和卷积神经网络的方法上都有大量研究。从近10年的论文数量来看,国内对车辆检索的研究更多,针对车辆检索速度和精度尝试了各种方法,均取得了不错的效果。此外,作为车辆检索的子任务,车辆重识别领域吸引了大量关注。早期国外的研究主要是基于传感器和基于手工特征的方法实现车辆重识别。在深度学习兴起后,国内对车辆重识别的研究较为深入,先后发布了很多不同场景下的车辆重识别视频数据集,并且在车辆特殊外观敏感度、视角变化鲁棒性、光照变化以及遮挡问题上作出了很大贡献。然而,如何解决现有研究方法泛化性不足的缺点,是未来研究的潜在方向。

在人脸结构化分析方面的工作较多,相应的解决问题的方法和思路也较为全面。国内外主要从采

用注意力机制提取更有效的特征、采用级联网络逐步实现精准分析、采用多任务结构提升人脸结构化分析效率、采用辅助信息帮助人脸结构化分析和处理人脸结构化分析中的数据非均衡问题等方面开展研究工作,都取得了很多优秀的研究成果。如 Rothe 等人(2018)提出一个深度期望年龄估计模型(deep expectation, DEX),将期望提纯模块融入年龄估计,取得了很好的效果;Liu 等人(2015)提出级联式的人脸属性多任务分析网络和人脸多属性分析的数据库 CelebA,在国内外相关研究中产生了较为深远的影响。在人脸结构化分析方面,国内和国外的研究工作相得益彰,相互补充,共同推进了该领域发展。

在行人结构化分析上,国内外主要从基于注意力机制学习更有效的特征、探索属性间或图像区域间的相关性、基于区域特征提升行人属性识别性能以及采用关键点信息和行人分割信息等辅助行人属性识别等方面展开研究。相较而言,国内的相关工作更多更全面。如针对相关性探索的工作,国外主要有 Wang 等人(2017c)采用 LSTM 探索图像部件序列间的相关性,而国内不仅有探索图像部件间的相关性以及属性之间的相关性的工作,还有采用基于注意力的模型结构同时探索图像和属性语义之间序列相关性的工作。总体而言,在行人结构化分析上,国内的相关研究机构更多,相应研究成果也稍多一些,解决问题的思路也更为发散。

4 发展趋势与展望

在交通视频大数据处理分析中,关于车辆结构化的研究开始较早,相关技术发展较快,但仍然有很大发展空间。车辆结构化的前提是车辆检测,受拍摄场景和车辆移动速度的影响,如何在光线暗淡以及车辆速度较快的情况下准确定位车辆仍然是一个待解决的难题。市面上车辆种类繁多,相似品牌车型之间差异较小,车牌识别变得更为重要,在复杂多变的场景中,车牌定位和识别算法的泛化性和准确性有待进一步提升。交通监控设备的广泛部署实现了对相关道路系统的全天候监控,也进一步加大了车辆检索与重识别任务的难度,如何在复杂多变的场景、海量目标群体中快速检索或者重识别目标车辆仍需要科研人员持续的投入和更多创新性的研究。随着城市管理水平的进一步提升,行人结构化

分析的需求正在逐渐显现。行人结构化主要解析人脸或者行人的一些表现属性,例如年龄、性别、衣着样式等,为后续相关任务提供更详实的数据支撑。行人结构化分析技术伴随着深度学习的发展也迎来了快速发展期,然而针对特定场景的行人结构化分析需进一步研究。例如,如何在非约束环境下准确识别人员年龄和性别,如何在资源有限的终端系统中实现高精度模型的部署,如何融合多模态信息进一步提升行人属性识别准确率等。行人行为分析是交通视频大数据处理分析中更为高阶的任务,受拍摄场景、移动摄像头、视角和光照变化等因素的影响而更具有挑战性。尽管行为识别技术在人工智能相关的众多领域扮演着越来越重要的角色,但仍然面临很大挑战。从主流神经网络架构的行为识别效果来看,目前的模型并没有在大型行为数据集 Kinetics 上达到应有的效果,原因是现有模型未能充分学习和建模行为时序关系。针对行为识别这一领域,今后的研究仍可关注如何设计可有效学习到行为的时序关系乃至长时序依赖的网络架构、如何适应大规模数据集和实现行为识别模型的轻量化等。另外,现有数据集如 UCF101 或 Kinetics400,主要包含发生在粗粒度和有限场景中的动作,在这些数据集上学习的模型会过度拟合行为无关的上下文。为深入研究细粒度行为识别,目前也出现了若干细粒度行为识别数据集,可用于检验模型对细粒度动作信息建模的能力。最后,由于昂贵的样本标注成本以及人类行为属性类别的开放性和复杂性,知识迁移也是行为识别领域未来的一个主要研究方向,具体涉及零样本、少样本、弱监督以及自监督行为识别等细分研究方向。

随着物联网技术和 5G 技术的发展,新技术的推动同样在交通视频结构化分析研究中发挥了重要作用。现代汽车搭载物联网设备已成为必然趋势,车辆可以与基础交通设施相连(vehicle to infrastructure, V2I),也可以与周围车辆相连(vehicle-to-vehicle, V2V),这两项技术的发展依赖于车辆视频结构化与物联网技术的共同进步。随着 5G 技术在全球普及,高质量视频数据的快速传输已经成为现实,如何更高效地实现车辆信息、行人信息及其行为预测等交通视频结构化信息提取变得更为迫切,不仅需要研究如何提升相关算法性能,同时还需要设计更高效的硬件系统,通过软硬件协同方式构建更高效

的交通视频结构化分析系统。

参考文献 (References)

- Aarathi K S and Abraham A. 2017. Vehicle color recognition using deep learning for hazy images//Proceedings of 2017 IEEE International Conference on Inventive Communication and Computational Technologies. Coimbatore, India; IEEE; 335-339 [DOI: 10.1109/iciict.2017.7975215]
- Abdulnabi A H, Wang G, Lu J W and Jia K. 2015. Multi-task CNN model for attribute prediction. IEEE Transactions on Multimedia, 17(11): 1949-1959 [DOI: 10.1109/tmm.2015.2477680]
- Ali S S M, George B and Vanajaksh L. 2013. Multiple inductive loop detectors for intelligent transportation systems applications: ramp metering, vehicle re-identification and lane change monitoring systems//2013 IEEE Symposium on Computers and Informatics. Langkawi, Malaysia; IEEE; 176-180 [DOI: 10.1109/isci.2013.6612398]
- Ashraf K, Wu B C, Iandola F N, Moskewicz M W and Keutzer K. 2017. Shallow networks for high-accuracy road object-detection//Proceedings of the 3rd International Conference on Vehicle Technology and Intelligent Transport Systems. Porto, Portugal; SciTePress; 33-40 [DOI: 10.5220/0006214900330040]
- Awang S and Azmi N M A N. 2017. Sparse-filtered convolutional neural networks with layer-skipping (SF-CNNLS) for intra-class variation of vehicle type recognition. Deep Learning for Image Processing Applications, 31: 194-217 [DOI: 10.3233/978-1-61499-822-8-194]
- Bashir R M S, Shahzad M and Fraz M M. 2018. DUPL-VR: deep unsupervised progressive learning for vehicle re-identification//Proceedings of the 13th International Symposium on Visual Computing. Las Vegas, USA; Springer; 286-295 [DOI: 10.1007/978-3-030-03801-4_26]
- Bashir R M S, Shahzad M and Fraz M M. 2019. VR-PROUD: vehicle re-identification using PROgressive unsupervised deep architecture. Pattern Recognition, 90: 52-65 [DOI: 10.1016/j.patcog.2019.01.008]
- Bi X J and Feng X Y. 2018. Human action recognition based on improved depth learning model C-GRBM. Journal of Harbin Engineering University, 39(1): 156-162 (毕晓君, 冯雪赞. 2018. 基于改进深度学习模型 C-GRBM 的人体行为识别. 哈尔滨工程大学学报, 39(1): 156-162) [DOI: 10.11990/jheu.201612051]
- Bobick A and Davis J. 1996. An appearance-based representation of action//Proceedings of the 13th International Conference on Pattern Recognition. Vienna, Austria; IEEE; 307-312 [DOI: 10.1109/icpr.1996.546039]
- Bobick A F and Davis J W. 2001. The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis

- and Machine Intelligence, 23 (3): 257-267 [DOI: 10.1109/34.910878]
- Brown A, Xie W D, Kalogeiton V and Zisserman A. 2020. Smooth-AP: smoothing the path towards large-scale image retrieval//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK; Springer: 677-694 [DOI: 10.1007/978-3-030-58545-7_39]
- Brown L M. 2010. Example-based color vehicle retrieval for surveillance//Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance. Boston, USA; IEEE: 91-96 [DOI: 10.1109/avss.2010.59]
- Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End-to-end object detection with transformers [EB/OL]. [2020-05-28]. <https://arxiv.org/pdf/2005.12872.pdf>
- Carreira J and Zisserman A. 2017. Quo vadis, action recognition? a new model and the kinetics dataset//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 4724-4733 [DOI: 10.1109/cvpr.2017.502]
- Chadwick S, Maddern W and Newman P. 2019. Distant vehicle detection using radar and vision//Proceedings of 2019 International Conference on Robotics and Automation. Montreal, Canada; IEEE: 8311-8317 [DOI: 10.1109/icra.2019.8794312]
- Charbonnier S, Pitton A C and Vassilev A. 2012. Vehicle re-identification with a single magnetic sensor//Proceedings of 2012 IEEE International Instrumentation and Measurement Technology Conference. Graz, Austria; IEEE: 380-385 [DOI: 10.1109/i2mtc.2012.6229117]
- Chaudhry R, Ravichandran A, Hager G and Vidal R. 2009. Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions//Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA; IEEE: 1932-1939 [DOI: 10.1109/cvpr.2009.5206821]
- Chen P, Bai X and Liu W Y. 2014. Vehicle color recognition on urban road by feature context. IEEE Transactions on Intelligent Transportation Systems, 15 (5): 2340-2346 [DOI: 10.1109/TITS.2014.2308897]
- Chen S Y, Hsieh J W, Wu J C and Chen Y S. 2009. Vehicle retrieval using eigen color and multiple instance learning//Proceedings of the 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing. Kyoto, Japan; IEEE: 657-660 [DOI: 10.1109/i2mh-ssp.2009.304]
- Chen Z B, Ying C L, Lin C Y, Liu S and Li W P. 2019. Multi-view vehicle type recognition with feedback-enhancement multi-branch CNNs. IEEE Transactions on Circuits and Systems for Video Technology, 29 (9): 2590-2599 [DOI: 10.1109/TCSVT.2017.2737460]
- Cheong C W, Lim R W S, See J, Wong L K, Tan I K T and Aris A. 2018. Vehicle semantics extraction and retrieval for long-term carpark video surveillance//Proceedings of the 24th International Conference on Multimedia Modeling. Bangkok, Thailand; Springer: 315-326 [DOI: 10.1007/978-3-319-73600-6_27]
- Chéron G, Laptev I and Schmid C. 2015. P-CNN: pose-based CNN features for action recognition//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile; IEEE: 3218-3226 [DOI: 10.1109/iccv.2015.368]
- Chu R H, Sun Y F, Li Y D, Liu Z, Zhang C and Wei Y C. 2019. Vehicle re-identification with viewpoint-aware metric learning//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South); IEEE: 8282-8291 [DOI: 10.1109/iccv.2019.00837]
- Chu W Q, Liu Y, Shen C, Cai D and Hua X S. 2018. Multi-task vehicle detection with region-of-interest voting. IEEE Transactions on Image Processing, 27 (1): 432-441 [DOI: 10.1109/TIP.2017.2762591]
- Clady X, Negri P, Milgram M and Poulencard R. 2008. Multi-class vehicle type recognition system//Proceedings of the 3rd IAPR Workshop on Artificial Neural Networks in Pattern Recognition. Paris, France; Springer: 228-239 [DOI: 10.1007/978-3-540-69939-2_22]
- Cui C, Sang N, Gao C X and Zou L. 2017. Vehicle re-identification by fusing multiple deep neural networks//Proceedings of the 7th International Conference on Image Processing Theory, Tools and Applications. Montreal, Canada; IEEE: 1-6 [DOI: 10.1109/IPTA.2017.8310090]
- Dai J F, Li Y, He K M and Sun J. 2016. R-FCN: object detection via region-based fully convolutional networks. [EB/OL]. [2020-12-19]. <https://arxiv.org/pdf/1605.06409.pdf>
- Dai X R. 2019. HybridNet: a fast vehicle detection system for autonomous driving. Signal Processing: Image Communication, 70: 79-88 [DOI: 10.1016/j.image.2018.09.002]
- Diba A, Pazandeh A M, Pirsiavash H and Van Gool L. 2016. Deep-CAMP: deep convolutional action and attribute mid-level patterns//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE: 3557-3565 [DOI: 10.1109/cvpr.2016.387]
- Ding H, Zhou H, Zhou S K and Chellappa R. 2018. A deep cascade network for unaligned face attribute classification//Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans, USA; AAAI Press: 6789-6796
- Ding H, Zhou P and Chellappa R. 2020. Occlusion-adaptive deep network for robust facial expression recognition//Proceedings of 2020 IEEE International Joint Conference on Biometrics. Houston, USA. IEEE: 1-9 [DOI: 10.1109/IJCB48548.2020.9304923]
- Dollar P, Rabaud V, Cottrell G and Belongie S. 2005. Behavior recognition via sparse spatio-temporal features//Proceedings of 2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. Beijing, China; IEEE: 65-72 [DOI: 10.1109/vspets.2005.1570899]
- Donahue J, Hendricks L A, Guadarrama S, Rohrbach M, Venugopalan

- S, Darrell T and Saenko K. 2015. Long-term recurrent convolutional networks for visual recognition and description//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA; IEEE: 2625-2634 [DOI: 10.1109/CVPR.2015.7298878]
- Dong Q, Gong S G and Zhu X T. 2019. Imbalanced deep learning by minority class incremental rectification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(6): 1367-1381 [DOI: 10.1109/tpami.2018.2832629]
- Fabbri M, Calderara S and Cucchiara R. 2017. Generative adversarial models for people attribute recognition in surveillance//Proceedings of the 14th IEEE International Conference on Advanced Video and Signal based Surveillance. Lecce, Italy; IEEE: 1-6 [DOI: 10.1109/avss.2017.8078521]
- Feichtenhofer C, Fan H Q, Malik J and He K M. 2019. Slowfast networks for video recognition//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South); IEEE: 6202-6211 [DOI: 10.1109/iccv.2019.00630]
- Ferencz A, Learned-Miller E G and Malik J. 2005. Building a classification cascade for visual identification from one example//Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV'05) Volume 1. Beijing, China; IEEE: 286-293 [DOI: 10.1109/iccv.2005.52]
- Fernandes G, Zaccaron A M, Rodrigues J J P C and Proença M L. 2013. Digital signature to help network management using principal component analysis and K-means clustering//Proceedings of 2013 IEEE International Conference on Communications. Budapest, Hungary; IEEE: 2519-2523 [DOI: 10.1109/icc.2013.6654912]
- Fujiyoshi H, Lipton A J and Kanade T. 2004. Real-time human motion analysis by image skeletonization. *IEICE Transactions on Information and Systems*, 87-D(1): 113-120
- Gao B B, Xing C, Xie C W, Wu J X and Geng X. 2017. Deep label distribution learning with label ambiguity. *IEEE Transactions on Image Processing*, 26(6): 2825-2838 [DOI: 10.1109/tip.2017.2689998]
- Gao X, Szep J, Satam P, Hariri S, Ram S and Rodríguez J J. 2020. Spatio-temporal processing for automatic vehicle detection in wide-area aerial video. *IEEE Access*, 8: 199562-199572 [DOI: 10.1109/access.2020.3033466]
- Girshick R, Donahue J, Darrell T and Malik J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA; IEEE: 580-587 [DOI: 10.1109/cvpr.2014.81]
- Girshick R. 2015. Fast R-CNN//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile; IEEE: 1440-1448 [DOI: 10.1109/iccv.2015.169]
- Gkioxari G, Girshick R and Malik J. 2015. Actions and attributes from wholes and parts//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile; IEEE: 2470-2478 [DOI: 10.1109/iccv.2015.284]
- Guan S J, Liao B, Du Y and Yin X Y. 2019. Vehicle type recognition based on Radon-CDT hybrid transfer learning//Proceedings of the 10th IEEE International Conference on Software Engineering and Service Science. Beijing, China; IEEE: 1-4 [DOI: 10.1109/ICSESS47205.2019.9040687]
- Han H, Jain A K, Wang F, Shan S G and Chen X L. 2018. Heterogeneous face attribute estimation: a deep multi-task learning approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(11): 2597-2609 [DOI: 10.1109/tpami.2017.2738004]
- Hand E M and Chellappa R. 2017. Attributes for improved attributes: a multi-task network utilizing implicit and explicit relationships for facial attribute classification//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA; AAAI Press: 4068-4074
- Hand E M, Castillo C D and Chellappa R. 2018. Predicting facial attributes in video using temporal coherence and motion-attention//Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe, USA; IEEE: 84-92 [DOI: 10.1109/wacv.2018.00017]
- He B, Li J, Zhao Y and Tian Y H. 2019. Part-regularized near-duplicate vehicle re-identification//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE: 3997-4005 [DOI: 10.1109/cvpr.2019.00412]
- He K K, Fu Y W, Zhang W H, Wang C J, Jiang Y G, Huang F Y and Xue X Y. 2018. Harnessing synthesized abstraction images to improve facial attribute recognition//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden; IJCAI: 733-740 [DOI: 10.24963/ijcai.2018/102]
- He K M, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy; IEEE: 2961-2969 [DOI: 10.1109/ICCV.2017.322]
- He K M, Zhang X Y, Ren S Q and Sun J. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9): 1904-1916 [DOI: 10.1109/TPAMI.2015.2389824]
- Hinton G E. 2007. Learning multiple layers of representation. *Trends in Cognitive Sciences*, 11(10): 428-434 [DOI: 10.1016/j.tics.2007.09.004]
- Hu C P, Bai X, Qi L, Chen P, Xue G J and Mei L. 2015. Vehicle color recognition with spatial pyramid deep learning. *IEEE Transactions on Intelligent Transportation Systems*, 16(5): 2925-2934 [DOI: 10.1109/TITS.2015.2430892]
- Hu Z Z, Wen Y G, Wang J F, Wang M, Hong R C and Yan S C. 2017. Facial age estimation with age difference. *IEEE Transactions on Image Processing*, 26(7): 3087-3097 [DOI: 10.1109/tip.2016.2633868]

- Huang C, Li Y N, Loy C C and Tang X O. 2020. Deep imbalanced learning for face recognition and attribute prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11): 2781-2794 [DOI: 10.1109/tpami.2019.2914680]
- Huang C, Li Y N, Loy C C and Tang X O. 2016. Learning deep representation for imbalanced classification//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA; IEEE: 5375-5384 [DOI: 10.1109/cvpr.2016.580]
- Ingole S K and Gundre S B. 2017. Characters feature based Indian vehicle license plate detection and recognition//*Proceedings of 2017 International Conference on Intelligent Computing and Control*. Coimbatore, India; IEEE: 1-5 [DOI: 10.1109/i2c2.2017.8321953]
- Jeng S T and Chu L Y. 2013. Vehicle reidentification with the inductive loop signature technology. *Journal of the Eastern Asia Society for Transportation Studies*, 10: 1896-1915 [DOI: 10.11175/easts.10.1896]
- Ji S W, Xu W, Yang M and Yu K. 2013. 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1): 221-231 [DOI: 10.1109/TPAMI.2012.59]
- Jia L, Miao H X, Qi B S and Wang J P. 2019. Vehicle type recognition based on improved capsule network//*Proceedings of 2019 IEEE International Conference on Consumer Electronics-Taiwan*. Yilan, China; IEEE: 1-2 [DOI: 10.1109/ICCE-TW46550.2019.8991734]
- Kalayeh M M, Gong B Q and Shah M. 2017. Improving facial attribute prediction using semantic segmentation//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA; IEEE: 6942-6950 [DOI: 10.1109/cvpr.2017.450]
- Kanacı A, Zhu X T and Gong S G. 2018. Vehicle re-identification in context//*Proceedings of the 40th German Conference on Pattern Recognition*. Stuttgart, Germany; Springer: 377-390 [DOI: 10.1007/978-3-030-12939-2_26]
- Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R and Li F F. 2014. Large-scale video classification with convolutional neural networks//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA; IEEE: 1725-1732 [DOI: 10.1109/cvpr.2014.223]
- Kataoka H, Hashimoto K, Iwata K, Satoh Y, Navab N, Ilic S and Aoki Y. 2014. Extended co-occurrence HOG with dense trajectories for fine-grained activity recognition//*Proceedings of the 12th Asian Conference on Computer Vision*. Singapore, Singapore; Springer: 336-349 [DOI: 10.1007/978-3-319-16814-2_22]
- Khare V, Shivakumara P, Raveendran P, Meng L K and Woon H H. 2015. A new sharpness based approach for character segmentation in license plate images//*Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition*. Kuala Lumpur, Malaysia; IEEE: 544-548 [DOI: 10.1109/acpr.2015.7486562]
- Kim H, Lee Y, Yim B, Park E and Kim H. 2016. On-road object detection using deep neural network//*Proceedings of 2016 IEEE International Conference on Consumer Electronics-Asia*. Seoul, Korea (South); IEEE: 1-4 [DOI: 10.1109/ICCE-Asia.2016.7804765]
- Kim K J, Kim P K, Lim K T, Chung Y S, Song Y J, Lee S I and Choi D H. 2018. Vehicle color recognition via representative color region extraction and convolutional neural network//*Proceedings of 2018 Tenth International Conference on Ubiquitous and Future Networks*. Prague, Czech Republic; IEEE: 89-94 [DOI: 10.1109/icufn.2018.8436710]
- Klaeser A, Marszałek M and Schmid C. 2008. A spatio-temporal descriptor based on 3D-gradients//*Proceedings of the 19th British Machine Vision Conference*. Leeds, UK; BMVA: #99 [DOI: 10.5244/c.22.99]
- Knopp J, Prasad M, Willems G, Timofte R and Van Gool L. 2010. Hough transform and 3D SURF for robust three dimensional classification//*Proceedings of the 11th European Conference on Computer Vision*. Heraklion, Crete, Greece; Springer: 589-602 [DOI: 10.1007/978-3-642-15567-3_43]
- Kwong K, Kavalier R, Rajagopal R and Varaiya P. 2009. Arterial travel time estimation based on vehicle re-identification using wireless magnetic sensors. *Transportation Research Part C: Emerging Technologies*, 17(6): 586-606 [DOI: 10.1016/j.trc.2009.04.003]
- Lei J N, Ren X F and Fox D. 2012. Fine-grained kitchen activity recognition using RGB-D//*Proceedings of the ACM Conference on Ubiquitous Computing*. Pittsburgh, USA; ACM: 208-211 [DOI: 10.1145/2370216.2370248]
- Li D W, Chen X T and Huang K Q. 2015. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios//*Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition*. Kuala Lumpur, Malaysia; IEEE: 111-115 [DOI: 10.1109/acpr.2015.7486476]
- Li D W, Chen X T, Zhang Z and Huang K Q. 2018a. Pose guided deep model for pedestrian attribute recognition in surveillance scenarios//*Proceedings of 2018 IEEE International Conference on Multimedia and Expo*. San Diego, USA; IEEE: 1-6 [DOI: 10.1109/icme.2018.8486604]
- Li W, Zhu X T and Gong S G. 2018b. Harmonious attention network for person re-identification//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA; IEEE: 2285-2294 [DOI: 10.1109/CVPR.2018.00243]
- Li X Y, Yuan M X, Jiang Q Y and Li G M. 2017. VRID-1: a basic vehicle re-identification dataset for similar vehicles//*Proceedings of the 20th IEEE International Conference on Intelligent Transportation Systems*. Yokohama, Japan; IEEE: 1-8 [DOI: 10.1109/itsc.2017.8317817]
- Li Y, Huang C and Loy C C. 2016. Human attribute recognition by deep hierarchical contexts//*Proceedings of the European Conference on*

- Computer Vision. Amsterdam, the Netherlands. Springer; 684-700.
- Liang D W, Yan K, Wang Y W, Zeng W, Yuan Q S, Bao X G and Tian Y H. 2017. Deep hashing with multi-task learning for large-scale instance-level vehicle search//Proceedings of 2017 IEEE International Conference on Multimedia and Expo Workshops. Hong Kong, China; IEEE; 192-197 [DOI: 10.1109/icmew.2017.8026274]
- Lin J, Gan C and Han S. 2019. TSM: temporal shift module for efficient video understanding//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South); IEEE; 7083-7093 [DOI: 10.1109/ICCV.2019.00718]
- Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2017. Focal loss for dense object detection//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy; IEEE; 2980-2988 [DOI: 10.1109/iccv.2017.324]
- Lin Y L, Tsai M K, Hsu W H and Chen C W. 2013. Investigating 3-D model and part information for improving content-based vehicle retrieval. IEEE Transactions on Circuits and Systems for Video Technology, 23 (3): 401-413 [DOI: 10.1109/tcsvt.2012.2204936]
- Liu H Y, Tian Y H, Wang Y W, Pang L and Huang T J. 2016a. Deep relative distance learning: tell the difference between similar vehicles//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE; 2167-2175 [DOI: 10.1109/cvpr.2016.238]
- Liu K, Li Y D and Lin W P. 2020. A survey on vehicle re-identification. Chinese Journal of Intelligent Science and Technology, 2(1): 10-25 (刘凯, 李煜东, 林伟鹏. 2020. 车辆再识别技术综述. 智能科学与技术学报, 2(1): 10-25)
- Liu L, Shao L, Zheng F and Li X L. 2014a. Realistic action recognition via sparsely-constructed Gaussian processes. Pattern Recognition, 47(12): 3819-3827 [DOI: 10.1016/j.patcog.2014.07.006]
- Liu N, Zhang F and Duan F Q. 2020a. Facial age estimation using a multi-task network combining classification and regression. IEEE Access, 8: 92441-92451 [DOI: 10.1109/access.2020.2994322]
- Liu P, Liu X and Yan J. 2018. Localization guided learning for pedestrian attribute recognition. [EB/OL]. [2020-12-19]. <https://arxiv.org/pdf/1808.09102.pdf>
- Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C F and Berg A C. 2016b. SSD: single shot MultiBox detector//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands; Springer; 21-37 [DOI: 10.1007/978-3-319-46448-0_2]
- Liu X C, Ma H D, Fu H Y and Zhou M. 2014b. Vehicle retrieval and trajectory inference in urban traffic surveillance scene//Proceedings of the International Conference on Distributed Smart Cameras. Venezia Mestre, Italy; ACM; 1-6 [DOI: 10.1145/2659021.2659055]
- Liu X, Zhao H and Tian M. 2017. Hydraplus-net: Attentive deep features for pedestrian analysis//Proceedings of the 16th IEEE International Conference on Computer Vision. Venice, Italy; IEEE; 350-359 [DOI: 10.1109/iccv.2017.46]
- Liu Y D, Wang Y T, Wang S W, Liang T T, Zhao Q J, Tang Z and Ling H B. 2020b. CBNet: a novel composite backbone network architecture for object detection//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA; AAAI Press; 11653-11660 [DOI: 10.1609/aaai.v34i07.6834]
- Liu Y F, Li M, Qiu D, Wang W S, Xu Z X and Song C. 2019. Multi-task squeeze-and-excitation networks for pedestrian attributes recognition. Science Technology and Engineering, 19 (24): 237-241 (刘弋锋, 李勖, 邱迪, 王文杉, 许忠雄, 宋超. 2019. 基于多任务压缩激发网络的行人属性识别. 科学技术与工程, 19(24): 237-241) [DOI: 10.3969/j.issn.1671-1815.2019.24.037]
- Liu Z W, Luo P, Wang X G and Tang X O. 2015. Deep learning face attributes in the wild//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile; IEEE; 3730-3738 [DOI: 10.1109/iccv.2015.425]
- Lu Y X, Kumar A, Zhai S F, Cheng Y, Javidi T and Feris R. 2017. Fully-adaptive feature sharing in multi-task networks with applications in person attribute classification//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE; 5334-5343 [DOI: 10.1109/cvpr.2017.126]
- Ma Y X, Tan L, Dong X and Yu C C. 2019. Action recognition for intelligent monitoring. Journal of Image and Graphics, 24(2): 282-290 (马钰锡, 谭励, 董旭, 于重重. 2019. 面向智能监控的行为识别. 中国图象图形学报, 24(2): 282-290) [DOI: 10.11834/jig.180392]
- Marín-Reyes P A, Bergamini L, Lorenzo-Navarro J, Palazzi A, Calderara S and Cucchiara R. 2018. Unsupervised vehicle re-identification using triplet networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, USA; IEEE; 166-171 [DOI: 10.1109/CVPRW.2018.00030]
- Ndoye M, Totten V, Krogmeier J V and Bullock D M. 2009. A signal processing framework for vehicle re-identification and travel time estimation//Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems. St. Louis, USA; IEEE; 1-6 [DOI: 10.1109/itsc.2009.5309763]
- Ng J Y H, Hausknecht M, Vijayanarasimhan S, Vinyals O, Monga R and Toderici G. 2015. Beyond short snippets: deep networks for video classification//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA; IEEE; 4694-4702 [DOI: 10.1109/cvpr.2015.7299101]
- Ni B B, Paramathayalan V R and Moulin P. 2014. Multiple granularity analysis for fine-grained action detection//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA; IEEE; 756-763 [DOI: 10.1109/cvpr.2014.102]
- Ni B B, Yang X K and Gao S H. 2016. Progressively parsing interactional objects for fine grained action detection//Proceedings of 2016

- IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE: 1020-1028 [DOI: 10.1109/cvpr.2016.116]
- Omran S S and Jarallah J A. 2017. Iraqi car license plate recognition using OCR//Proceedings of 2017 Annual Conference on New Trends in Information and Communications Technology Applications. Baghdad, Iraq; IEEE: 298-303 [DOI: 10.1109/NTICT.2017.7976127]
- Packer B, Saenko K and Koller D. 2012. A combined pose, object, and feature model for action understanding//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA; IEEE: 1378-1385 [DOI: 10.1109/cvpr.2012.6247824]
- Park S and Zhu S C. 2015. Attributed grammars for joint estimation of human attributes, part and pose//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile; IEEE: 2372-2380 [DOI: 10.1109/iccv.2015.273]
- Park S M and Kim K J. 2012. Color recognition with compact color features. *International Journal of Communication Systems*, 25 (6): 749-762 [DOI: 10.1002/dac.1229]
- Park S M and Kim K J. 2008. PCA-SVM based vehicle color recognition. *The KIPS Transactions: Part B*, 15B (4): 285-292 [DOI: 10.3745/kipstb.2008.15-b.4.285]
- Pei W J, Dibeklioglu H, Baltrušaitis T and Tax D M J. 2019. Attended end-to-end architecture for age estimation from facial expression videos. *IEEE Transactions on Image Processing*, 29: 1972-1984 [DOI: 10.1109/tip.2019.2948288]
- Pei W, An Z J, Zhu Y Y, Jia X, Zuo X W and Wang F S. 2014. A rapid vehicle recognition and retrieval system//Proceedings of the 2nd International Conference on Systems and Informatics. Shanghai, China; IEEE: 748-753 [DOI: 10.1109/icsai.2014.7009384]
- Pei W, Jia X, Huang X G, Lu M Y, Fu Y L and Zhu Y Y. 2015. A vehicle retrieval and recognition system based on smart phones//Proceedings of the 8th International Congress on Image and Signal Processing. Shenyang, China; IEEE: 527-532 [DOI: 10.1109/CISP.2015.7407936]
- Prinsloo J and Malekian R. 2016. Accurate vehicle location system using RFID, an internet of things approach. *Sensors*, 16 (6): #825 [DOI: 10.3390/s16060825]
- Pustokhina I V, Pustokhin D A, Rodrigues J J P C, Gupta D, Khanna A, Shankar K, Seo C and Joshi G P. 2020. Automatic vehicle license plate recognition using optimal K-means with convolutional neural network for intelligent transportation systems. *IEEE Access*, 8: 92907-92917 [DOI: 10.1109/access.2020.2993008]
- Radchenko A, Zarovsky R and Kazymyr V. 2017. Method of segmentation and recognition of Ukrainian license plates//Proceedings of 2017 IEEE International Young Scientists Forum on Applied Physics and Engineering. Lviv, Ukraine; IEEE: 62-65 [DOI: 10.1109/ysf.2017.8126593]
- Rana N and Dahiya P K. 2017. Localization techniques in ANPR systems; a-state-of-art. *International Journal of Advanced Research in Computer Science and Software Engineering*, 7 (5): 682-686 [DOI: 10.23956/ijarsse/sv7i5/0338]
- Rapantzikos K, Avrithis Y and Kollias S. 2009. Dense saliency-based spatiotemporal feature points for action recognition//Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA; IEEE: 1454-1461 [DOI: 10.1109/cvpr.2009.5206525]
- Raza M A, Qi C, Asif M R and Khan M A. 2020. An adaptive approach for multi-national vehicle license plate recognition using multi-level deep features and foreground polarity detection model. *Applied Sciences*, 10(6): 2165 [DOI: 10.3390/app10062165]
- Redmon J, Divvala S, Girshick R and Farhadi A. 2016. You only look once: unified, real-time object detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE: 779-788 [DOI: 10.1109/CVPR.2016.91]
- Ren S Q, He K M, Girshick R and Sun J. 2015. Faster R-CNN: towards real-time object detection with region proposal networks//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Quebec, Canada; MIT: 91-99
- Rodríguez P, Cucurull G and Gonfaus J M. 2017. Age and gender recognition in the wild with deep attention. *Pattern Recognition*, 72: 563-571 [DOI: 10.1016/j.patcog.2017.06.028]
- Rohrbach M, Amin S, Andriluka M and Schiele B. 2012. A database for fine grained activity detection of cooking activities//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA; IEEE: 1194-1201 [DOI: 10.1109/cvpr.2012.6247801]
- Rothe R, Timofte R and Van Gool L. 2018. Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2/4): 144-157 [DOI: 10.1007/s11263-016-0940-3]
- Ryoo M S, Piergiovanni A J, Tan M X and Angelova A. 2020. Assemblenet: searching for multi-stream neural connectivity in video architectures [EB/OL]. [2020-05-27]. <https://arxiv.org/pdf/1905.13209.pdf>
- Sanchez R O, Flores C, Horowitz R, Rajagopal R and Varaiya P. 2011. Vehicle re-identification using wireless magnetic sensors; algorithm revision, modifications and performance analysis//Proceedings of 2011 IEEE International Conference on Vehicular Electronics and Safety. Beijing, China; IEEE: 226-231 [DOI: 10.1109/icves.2011.5983819]
- Sarafianos N, Xu X and Kakadiaris I A. 2018. Deep imbalanced attribute classification using visual attention aggregation//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany; IEEE: 680-697 [DOI: 10.1007/978-3-030-01252-6_42]
- Sarfraz M S, Saeed A, Khan M H and Riaz Z. 2009. Bayesian prior models for vehicle make and model recognition//Proceedings of the 7th International Conference on Frontiers of Information Technology.

- Abbottabad, Pakistan: ACM; 1-6 [DOI: 10.1145/1838002.1838041]
- Sarfraz M S, Schumann A, Wang Y and Stiefelhagen R. 2017. Deep view-sensitive pedestrian attribute inference in an end-to-end model//Proceedings of the British Machine Vision Conference. London, UK: BMVA Press; 4-7 [DOI: 10.5244/c.31.134]
- Savov N, Ngo M L, Karaoglu S, Dibeklioglu H and Gevers T. 2019. Pose and expression robust age estimation via 3D face reconstruction from a single image//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop. Seoul, Korea (South); IEEE; 1270-1278 [DOI: 10.1109/iccvw.2019.00160]
- Schmidhuber J. 2015. Deep learning in neural networks: an overview. *Neural Networks*, 61; 85-117 [DOI: 10.1016/j.neunet.2014.09.003]
- Schuldt C, Laptev I and Caputo B. 2004. Recognizing human actions; a local SVM approach//Proceedings of the 17th International Conference on Pattern Recognition. Cambridge, UK; IEEE; 32-36 [DOI: 10.1109/icpr.2004.1334462]
- Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R and LeCun Y. 2013. Overfeat: integrated recognition, localization and detection using convolutional networks [EB/OL]. [2020-12-16]. <https://arxiv.org/pdf/1312.6229.pdf>
- Shi X, Zhao C H, Mao Y B and Chen A W. 2019. Vehicle retrieval using features of vehicle front windshields//Proceedings of the 19th COTA International Conference of Transportation Professionals. Nanjing, China; ASCE; 2295-2305 [DOI: 10.1061/9780784482292.200]
- Shu Y, Mao L B, Chen S and Yan Y. 2020. Self-supervised learning and generative adversarial network-based facial attribute recognition with small sample size training. *Journal of Image and Graphics*, 25(11); 2391-2403 (疏颖, 毛龙彪, 陈思, 严严. 2020. 结合自监督学习和生成对抗网络的小样本人脸属性识别. *中国图象图形学报*, 25(11); 2391-2403) [DOI: 10.11834/jig.200334]
- Shvai N, Meicler A, Hasnat A, Machover E, Maarek P, Loquet S and Nakib A. 2018. Optimal ensemble classifiers based classification for automatic vehicle type recognition//Proceedings of 2018 IEEE Congress on Evolutionary Computation. Rio de Janeiro, Brazil; IEEE; 1-8 [DOI: 10.1109/cec.2018.8477979]
- Singh B, Marks T K, Jones M, Tuzel O and Shao M. 2016. A multi-stream bi-directional recurrent neural network for fine-grained action detection//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE; 1961-1970 [DOI: 10.1109/cvpr.2016.216]
- Son J W, Park S B and Kim K J. 2007. A convolution kernel method for color recognition//Proceedings of the 6th International Conference on Advanced Language Processing and Web Information Technology. Luoyang, China; IEEE; 242-247 [DOI: 10.1109/alpit.2007.28]
- Soon F C, Khaw H Y, Chuah J H and Kanesan J. 2020. Semisupervised PCA convolutional network for vehicle type classification. *IEEE Transactions on Vehicular Technology*, 69(8); 8267-8277 [DOI: 10.1109/tvt.2020.3000306]
- Stewart R, Andriluka M and Ng A Y. 2016. End-to-end people detection in crowded scenes//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE; 2325-2333 [DOI: 10.1109/cvpr.2016.255]
- Su Y Q, Cuan B N, Zhang X J and Liu Y H. 2016. Rapid vehicle retrieval using a cascade of interest regions//Proceedings of the 17th Pacific Rim Conference on Multimedia. Xi'an, China; Springer; 126-135 [DOI: 10.1007/978-3-319-48896-7_13]
- Sun P H, Liu H, Wang X, Yu Z H and Wu S P. 2019. Similarity-aware deep adversarial learning for facial age estimation//Proceedings of 2019 IEEE International Conference on Multimedia and Expo. Shanghai, China; IEEE; 260-265 [DOI: 10.1109/icme.2019.00053]
- Swain M J and Ballard D H. 1991. Color indexing. *International Journal of Computer Vision*, 7(1); 11-32 [DOI: 10.1007/bf00130487]
- Tan Z C, Yang Y, Wan J, Guo G D and Li S Z. 2020. Relation-aware pedestrian attribute recognition with graph convolutional networks//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA; AAAI Press; 12055-12062 [DOI: 10.1609/aaai.v34i07.6883]
- Tan Z, Yang Y and Wan J. 2019. Attention-based pedestrian attribute analysis. *IEEE Transactions on Image Processing*, 28(12); 6126-6140 [DOI: 10.3390/s20030811]
- Tang T Y, Deng Z P, Zhou S L, Lei L and Zou H X. 2017. Fast vehicle detection in UAV images//Proceedings of 2017 International Workshop on Remote Sensing with Intelligent Processing. Shanghai, China; IEEE; 1-5 [DOI: 10.1109/RSIP.2017.7958795]
- Tao H J and Lu X B. 2018. Smoky vehicle detection based on multi-feature fusion and ensemble neural networks. *Multimedia Tools and Applications*, 77(24); 32153-32177. [DOI: 10.1007/s11042-018-6248-2]
- Tian Y, Chen T, Cheng G H, Yu S H, Li X, Li J Y and Yang B L. 2020. Global context assisted structure-aware vehicle retrieval. *IEEE Transactions on Intelligent Transportation Systems*, 2020; 1-10 [DOI: 10.1109/TITS.2020.3009000]
- Tran D, Wang H, Torresani L, Ray J, LeCun Y and Paluri M. 2018. A closer look at spatiotemporal convolutions for action recognition//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE; 6450-6459 [DOI: 10.1109/cvpr.2018.00675]
- Ullah I and Lee H J. 2016. An approach of locating Korean vehicle license plate based on mathematical morphology and geometrical features//Proceedings of 2016 International Conference on Computational Science and Computational Intelligence. Las Vegas, USA; IEEE; 836-840 [DOI: 10.1109/CSCI.2016.0162]
- Varol G, Laptev I and Schmid C. 2018. Long-term temporal convolutions for action recognition. *IEEE Transactions on Pattern Analysis and*

- Machine Intelligence, 40(6): 1510-1517 [DOI: 10.1109/tpami.2017.2712608]
- Veeriah V, Zhuang N F and Qi G J. 2015. Differential recurrent neural networks for action recognition//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 4041-4049 [DOI: 10.1109/iccv.2015.460]
- Wan J, Tan Z C, Lei Z, Guo G D and Li S Z. 2018. Auxiliary demographic information assisted age estimation with cascaded structure. IEEE Transactions on Cybernetics, 48(9): 2531-2541 [DOI: 10.1109/tyb.2017.2741998]
- Wang H, Kläser A, Schmid C and Liu C L. 2011. Action recognition by dense trajectories//Proceedings of 2011 Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE: 3169-3176 [DOI: 10.1109/cvpr.2011.5995407]
- Wang J, Zhu X and Gong S. 2017c. Attribute recognition by joint recurrent learning of context and correlation//Proceedings of the 16th IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 531-540.
- Wang L M, Qiao Y and Tang X O. 2015. Action recognition with trajectory-pooled deep-convolutional descriptors//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 4305-4314 [DOI: 10.1109/cvpr.2015.7299059]
- Wang L M, Xiong Y J, Wang Z, Qiao Y, Lin D H, Tang X O and Van Gool L. 2016. Temporal segment networks: towards good practices for deep action recognition//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 20-36 [DOI: 10.1007/978-3-319-46484-8_2]
- Wang W H and Tu J Y. 2020. Research on license plate recognition algorithms based on deep learning in complex environment. IEEE Access, 8: 91661-91675 [DOI: 10.1109/ACCESS.2020.2994287]
- Wang X L, Girshick R, Gupta A and He K M. 2018. Non-local neural networks//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7794-7803
- Wang Z X, He K K, Fu Y W, Feng R, Jiang Y G and Xue X Y. 2017b. Multi-task deep neural network for joint face recognition and facial attribute prediction//Proceedings of 2017 International Conference on Multimedia Retrieval. Bucharest, Romania: ACM: 365-374 [DOI: 10.1145/3078971.3078973]
- Wei J, Wang Y, Liu F, Lin Q L and Wang N. 2020. Colour fast-match for precise vehicle retrieval. The Journal of Engineering, 2020(4): 132-139 [DOI: 10.1049/joe.2019.0882]
- Weinland D, Ronfard R and Boyer E. 2006. Free viewpoint action recognition using motion history volumes. Computer Vision and Image Understanding, 104(2/3): 249-257 [DOI: 10.1016/j.cviu.2006.07.013]
- Woesler R. 2003. Fast extraction of traffic parameters and reidentification of vehicles from video data//Proceedings of 2003 IEEE International Conference on Intelligent Transportation Systems. Shanghai, China: IEEE: 774-778 [DOI: 10.1109/itsc.2003.1252055]
- Wu J J, Liu H, Jiang J G, Qi M B, Ren B, Li X H and Wang Y S. 2020. Person attribute recognition by sequence contextual relation learning. IEEE Transactions on Circuits and Systems for Video Technology, 30(10): 3398-3412 [DOI: 10.1109/tesvt.2020.2982962]
- Xie J C and Pun C M. 2020. Deep and ordinal ensemble learning for human age estimation from facial images. IEEE Transactions on Information Forensics and Security, 15: 2361-2374 [DOI: 10.1109/tifs.2020.2965298]
- Yang J, Fan J R, Wang Y R, Wang Y G, Gan W B, Liu L and Wu W. 2020. Hierarchical feature embedding for attribute recognition//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 13055-13064 [DOI: 10.1109/cvpr42600.2020.01307]
- Yang X D and Tian Y L. 2014. Effective 3D action recognition using EigenJoints. Journal of Visual Communication and Image Representation, 25(1): 2-11 [DOI: 10.1016/j.jvcir.2013.03.001]
- Yang X, Gao B B, Xing C, Huo Z W, Wei X S, Zhou Y, Wu J X and Geng X. 2015. Deep label distribution learning for apparent age estimation//Proceedings of 2015 IEEE International Conference on Computer Vision Workshop. Santiago, Chile: IEEE: 102-108 [DOI: 10.1109/iccvw.2015.53]
- Yoo B, Kwak Y, Kim Y, Choi C and Kim J. 2018. Deep facial age estimation using conditional multitask learning with weak label expansion. IEEE Signal Processing Letters, 25(6): 808-812 [DOI: 10.1109/lsp.2018.2822241]
- Yoo S B and Han M. 2020. Temporal matching prior network for vehicle license plate detection and recognition in videos. ETRI Journal, 42(3): 411-419 [DOI: 10.4218/etrij.2019-0245]
- Yousif B B, Ata M M, Fawzy N and Obaya M. 2020. Toward an optimized neutrosophic K-means with genetic algorithm for automatic vehicle license plate recognition (ONKM-AVLPR). IEEE Access, 8: 49285-49312 [DOI: 10.1109/access.2020.2979185]
- Zapletal D and Herout A. 2016. Vehicle re-identification for automatic video traffic surveillance//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Las Vegas, USA: IEEE: 25-31 [DOI: 10.1109/cvprw.2016.195]
- Zhang C, Wang Q and Li X L. 2020a. EQ-LPR: efficient quality-aware license plate recognition//Proceedings of 2020 IEEE International Conference on Image Processing. Abu Dhabi, United Arab Emirates: IEEE: 653-657 [DOI: 10.1109/ICIP40778.2020.9191206]
- Zhang C, Wang Q and Li X L. 2020b. IQ-STAN: image quality guided spatio-temporal attention network for license plate recognition//Proceedings of 2020 IEEE International Conference on Acoustics, Speech, and Signal Processing. Barcelona, Spain: IEEE: 2268-2272 [DOI: 10.1109/ICASSP40776.2020.9053966]

- Zhang K, Liu N and Yuan X. 2019. Fine-grained age estimation in the wild with attention LSTM networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(9): 3140-3152 [DOI: 10.1109/TCSVT.2019.2936410]
- Zhang N, Paluri M, Ranzato M A, Darrell T and Bourdev L. 2014. PANDA: pose aligned networks for deep attribute modeling//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA; IEEE: 1637-1644 [DOI: 10.1109/cvpr.2014.212]
- Zhang Z L, Xue J, Qiao H H and Zhi Y F. 2014. The vehicle retrieval methods of traffic video based on improved SURF algorithm. *Journal of Northwestern Polytechnical University*, 32(2): 297-302 (张子龙, 薛静, 乔鸿海, 智永锋. 2014. 基于改进 SURF 算法的交通视频车辆检索方法研究. *西北工业大学学报*, 32(2): 297-302) [DOI: 10.3969/j.issn.1000-2758.2014.02.026]
- Zhang Z P, Luo P, Loy C C and Tang X O. 2018. From facial expression recognition to interpersonal relation prediction. *International Journal of Computer Vision*, 126(5): 550-569 [DOI: 10.1007/s11263-017-1055-1]
- Zhao X, Sang L F, Ding G G, Guo Y C and Jin X M. 2018. Grouping attribute recognition for pedestrian with joint recurrent learning//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm, Sweden: IJCAI: 3177-3183 [DOI: 10.24963/ijcai.2018/441]
- Zheng Z X, An G Y, Wu D P and Ruan Q Q. 2020. Global and local knowledge-aware attention network for action recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1): 334-347 [DOI: 10.1109/tnnls.2020.2978613]
- Zhou Y and Shao L. 2018. Viewpoint-aware attentive multi-view inference for vehicle re-identification//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA; IEEE: 6489-6498 [DOI: 10.1109/cvpr.2018.00679]
- Zhou Y, Ni B B, Hong R C, Wang M and Tian Q. 2015. Interaction part mining: a mid-level approach for fine-grained action recognition//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA; IEEE: 3323-3331 [DOI: 10.1109/cvpr.2015.7298953]
- Zhu C, Jia H Z, Lu T, Tao L, Song J W, Xiang G Q, Li Y and Xie X D. 2017. Adaptive feature selection based on local descriptor distinctive degree for vehicle retrieval application//*Proceedings of 2017 IEEE International Conference on Consumer Electronics*. Las Vegas, USA; IEEE: 66-69 [DOI: 10.1109/icce.2017.7889232]

作者简介



赵耀, 1967年生, 男, 教授, 主要研究方向为图像编码、数字水印、跨媒体内容分析与内容理解、多媒体信息处理。

E-mail: yzhao@bjtu.edu.cn



韦世奎, 通信作者, 男, 教授, 主要研究方向为跨媒体智能、计算机视觉、机器学习、智能交通。

E-mail: shkwei@bjtu.edu.cn

田永鸿, 男, 教授, 主要研究方向为视频大数据分析 with 处理、视觉神经信息编码与识别、机器学习与计算机视觉。

E-mail: yhtian@pku.edu.cn

党建武, 男, 教授, 主要研究方向为交通信息工程及控制、智能信息处理。E-mail: dangjw@mail.lzjtu.cn

付树军, 男, 教授, 主要研究方向为图像处理 and 计算机视觉、医学图像分析、大数据计算。E-mail: shujunfu@163.com

王恒友, 男, 副教授, 主要研究方向为图像处理、计算机视觉图像分析、大数据分析。E-mail: wanghengyou@bucea.edu.cn

万军, 男, 副研究员, 主要研究方向为计算机视觉、智能视频分析与理解、视频结构化分析。E-mail: jun.wan@ia.ac.cn

安高云, 男, 教授, 主要研究方向为人脸识别、活体检测、行为分析。E-mail: gyan@bjtu.edu.cn

杜卓然, 男, 博士研究生, 主要研究方向为图像处理、计算机视觉。E-mail: 20112007@bjtu.edu.cn

廖理心, 男, 博士研究生, 主要研究方向为计算机视觉、神经网络增强。E-mail: liaolixin@163.com