



JOURNAL OF IMAGE AND GRAPHICS

主办：中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象学报 图形学报

2020
05
VOL.25

ISSN1006-8961
CN11-3758/TB



智能疫
“小珈”
助力



第25卷第5期（总第289期）

2020年5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 中国科学院遥感与数字地球研究所

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 顾行发

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@radi.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 Institute of Remote Sensing and Digital Earth, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Gu Xingfa

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@radi.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



多相图像分割Vese-Chan模型连续最大流方法(第0926页)



保持区域平滑的3维动画形状高效编辑(第1019页)



树形结构卷积神经网络优化的城区遥感图像语义分割(第1043页)

学者观点

多源空一谱遥感图像融合方法进展与挑战

肖亮, 刘鹏飞, 李恒 0851

综述

中国图像工程: 2019

章毓晋 0864

图像处理和编码

相同编码参数HEVC视频重压缩检测

潘鹏飞, 姚晔, 王慧 0879

密集连接卷积网络图像去模糊

吴迪, 赵洪田, 郑世宝 0890

密集网络图像哈希检索

王亚鸽, 康晓东, 郭军, 李博, 张华丽, 刘汉卿 0900

图像分析和识别

域自适应城市场景语义分割

张桂梅, 潘国峰, 刘建新 0913

多相图像分割Vese-Chan模型连续最大流方法

王洁, 潘振宽, 魏伟波, 徐子森 0926

结合注意力机制和多属性分类的行人再识别

郑鑫, 林兰, 叶茂, 王丽, 贺春林 0936

多分辨率特征注意力融合行人再识别

沈庆, 田畅, 王家宝, 焦珊珊, 杜麟 0946

全卷积网络电线识别方法

刘嘉玮, 李元祥, 龚政, 刘心刚, 周拥军 0956

双核压缩激活神经网络艺术图像分类

杨秀芹, 张华熊 0967

图像理解和计算机视觉

高效检测复杂场景的快速金字塔网络SPNet

李鑫泽, 张轩雄, 陈胜 0977

结合自底向上注意力机制和记忆网络的视觉问答模型

闫茹玉, 刘学亮 0993

统筹图像变换与缝合线生成的无参数影像拼接

高炯笠, 吴军, 刘祺昌, 徐刚 1007

计算机图形学

保持区域平滑的3维动画形状高效编辑

冼楚华, 杨煜, 黄俊贤, 李桂清 1019

医学图像处理

肌骨超声图像特征检测及拼接

颜焕欢, 张培镇, 王伊依, 李璇, 阳维, 王青 1032

遥感图像处理

树形结构卷积神经网络优化的城区遥感图像语义分割

胡伟, 高博川, 黄振航, 李瑞瑞 1043

融合累积变异比和集成超限学习机的高光谱图像分类

尹玉萍, 魏林, 刘万军 1053

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Continuous max-flow method for multiphase segmentation Vese-Chan model(P0926)



Efficient 3D animation shape manipulation with region smoothness preservation(P1019)



Semantic segmentation of urban remote sensing image based on optimized tree structure convolutional neural network (P1043)

Scholar View

Progress and challenges in the fusion of multisource spatial-spectral remote sensing images

Xiao Liang, Liu Pengfei, Li Heng 0851

Review

Image engineering in China: 2019

Zhang Yujin 0864

Image Processing and Coding

Detection of double compression for HEVC videos with the same coding parameters

Pan Pengfei, Yao Ye, Wang Hui 0879

Motion deblurring method based on DenseNets

Wu Di, Zhao Hongtian, Zheng Shibao 0890

Image Hash retrieval with DenseNet

Wang Yage, Kang Xiaodong, Guo Jun, Li bo, Zhang Huali, Liu Hanqing 0900

Image Analysis and Recognition

Domain adaptation for semantic segmentation based on adaption learning rate

Zhang Guimei, Pan Guofeng, Liu Jianxin 0913

Continuous max-flow method for multiphase segmentation Vese-Chan model

Wang Jie, Pan Zhenkuan, Wei Weibo, Xu Zisen 0926

Improving person re-identification by attention and multi-attributes

Zheng Xin, Lin Lan, Ye Mao, Wang Li, He Chunlin 0936

Multi-resolution feature attention fusion method for person re-identification

Shen Qing, Tian Chang, Wang Jiabao, Jiao Shanshan, Du Lin 0946

Power line recognition method via fully convolutional network

Liu Jiawei, Li Yuanxiang, Gong Zheng, Liu Xingang, Zhou Yongjun 0956

Art image classification with double kernel squeeze-and-excitation neural network

Yang Xiuqin, Zhang Huaxiong 0967

Image Understanding and Computer Vision

Snap Pyramid Network: Real-time complex scene detecting system

Li Xinze, Zhang Xuanxiong, Chen Sheng 0977

Visual question answering model based on bottom-up attention and memory network

Yan Ruyu, Liu Xueliang 0993

Stitching of parametric-free images based on coordinated image transformation and seam-line generation

Gao Jiongli, Wu Jun, Liu Qichang, Xu Gang 1007

Computer Graphics

Efficient 3D animation shape manipulation with region smoothness preservation

Xian Chuhua, Yang Yu, Huang Junxian, Li Guiqing 1019

Medical Image Processing

Feature detection algorithm of musculoskeletal ultrasound image and its application of image stitching

Yan Huanhuan, Zhang Peizhen, Wang Yinong, Li Xuan, Yang Wei, Wang Qing 1032

Remote Sensing Image Processing

Semantic segmentation of urban remote sensing image based on optimized tree structure convolutional neural network

Hu Wei, Gao Bochuan, Huang Zhenhang, Li Ruirui 1043

Ensemble extreme learning machine with cumulative variation quotient for hyperspectral image classification

Yin Yuping, Wei Lin, Liu Wanjun 1053

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2020)05-0936-10

论文引用格式: Zheng X, Lin L, Ye M, Wang L, He C L. 2020. Improving person re-identification by attention and multi-attributes. Journal of Image and Graphics, 25(05):0936-0945 [郑鑫, 林兰, 叶茂, 王丽, 贺春林. 2020. 结合注意力机制和多属性分类的行人再识别. 中国图象图形学报, 25(05):0936-0945] [DOI:10.11834/jig.190185]

结合注意力机制和多属性分类的行人再识别

郑鑫¹, 林兰², 叶茂², 王丽¹, 贺春林¹

1. 西华师范大学计算机学院, 南充 637002; 2. 电子科技大学计算机科学与工程学院, 成都 611731

摘要: 目的 在行人再识别中,经常出现由于行人身体部位被遮挡和行人图像对之间不对齐而导致误判的情况。利用人体固有结构的特性,关注具有显著性特征的行人部件,忽略带有干扰信息的其他部件,有利于判断不同摄像头拍摄的行人是否为同一人。因此,提出了基于注意力机制和多属性分类的行人再识别方法。方法 在训练阶段,利用改进的 ResNet50 网络作为基本框架提取特征,随后传递给全局分支和局部分支。在全局分支中,将该特征作为全局特征进行身份和全局属性分类;在局部分支中,按信道展开特征,获取每层响应值最高的点,聚合这些点,分成4个行人部件,计算每个行人部件上显著性特征的权重,并乘以初始特征得到每个部件的总特征。最后将这4个部件的总特征都进行身份和对应属性的分类。在测试阶段,将通过网络提取的部位特征和全局特征串联起来,计算行人间的相似度,从而判断是否为同一人。结果 本文方法引入了 Market-1501_attribute 和 DukeMTMC-attribute 数据集中的属性信息,并在 Market-1501 和 DukeMTMC-reid 数据集上进行测试,其中 rank-1 分别达到 90.67% 和 80.2%, mAP 分别达到 76.65% 和 62.14%;使用 re-ranking 算法后,rank-1 分别达到 92.4% 和 84.15%, mAP 分别达到 87.5% 和 78.41%,相比近年来具有代表性的其他方法,识别率有了极大提升。结论 本文方法通过学习行人属性更快地聚集行人部件的注意力,而注意力机制又能更好地学习行人部位的显著性特征,从而有效解决了行人被遮挡和不对齐的问题,提高了行人再识别的准确率。

关键词: 行人再识别;全局特征;局部特征;行人部件;注意力机制;属性信息

Improving person re-identification by attention and multi-attributes

Zheng Xin¹, Lin Lan², Ye Mao², Wang Li¹, He Chunlin¹

1. School of Computer Science, China West Normal University, Nanchong 637002, China;

2. School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract: **Objective** Person re-identification (ReID) refers to the retrieval of target pedestrians from multiple non-overlapping cameras. This technology can be widely used in various fields. In the security field, police can quickly track and retrieve the location of suspects and easily find missing individuals by using person ReID. In the field of album management, users manage electronic albums according to person ReID. In the field of e-commerce, managers of unmanned supermarkets use person ReID to track user behavior. However, this technology poses a variety of challenges, such as low resolution, light change, background clutter, variable pedestrian action, and occlusion of pedestrian body parts. Numerous methods have been proposed to solve these problems. The traditional methods mainly include two ways, namely, feature design and

收稿日期: 2019-05-15; 修回日期: 2019-10-27; 预印本日期: 2019-11-03

基金项目: 国家重点研发计划项目(2018YFC0831800); 国家自然科学基金项目(61773093, 61731330); 四川省科技厅科技计划项目(17ZDYF3184); 成都市重点研发计划项目(2018-YF08-00039-GX)

Supported by: National Key Research and Development Program of China (2018YFC0831800); National Natural Science Foundation of China (61773093, 61731330)

distance metric. The main idea of feature design is to design a feature representation with strong discriminability and robustness. Thus, an expressive feature can be extracted from a pedestrian image. The distance metric method is used to reduce the distance among similar pedestrian images while increasing the distance among different pedestrian images. In recent years, deep learning has been widely used in the field of computer vision. Given the popularity of deep learning, person ReID based on deep learning has achieved higher recognition rate than traditional methods. The deep learning method mainly uses convolutional neural networks (CNNs) to classify pedestrian images, extract the representation of pedestrian images from the network, and determine whether pedestrian pairs form matches by calculating the similarity between pedestrian pairs. In this work, the proposed method is based on a CNN. We observe that the problems of misjudgment due to the occlusion of pedestrian body parts and misalignment between pedestrian image pairs are frequent. Thus, we attempt to focus on pedestrian parts with distinctive features and ignore other parts with interference information to determine whether pedestrian pairs captured by different cameras form matches. We then propose a method for person ReID using attention and attributes to overcome pedestrian occlusion and misalignment problem in ReID.

Method Our method introduces the attribute information of pedestrians. Attribute information is divided into local and global attributes. The local attributes are composed of the head, upper body, hand, and lower body part attributes. The global attributes include gender and age. The proposed method falls into two stages: training and test stages. In the training stage, the improved ResNet50 network is used as the basic framework to extract features. The training stage is composed of two branches: a global branch and a local branch. In the global branch, features are used as the global feature to classify identity and global attributes. Meanwhile, in the local branch, multiple processes are involved. First, features are extended into multiple channels. Second, we obtain the point with the highest response value per channel. Third, we cluster these points, and the cluster points are divided into four pedestrian parts. Fourth, we obtain the weight of the distinctive feature on each pedestrian part through calculation. Fifth, the weights of these four parts are multiplied by the initial features to obtain the total features of each part. Finally, the total features of these four parts are classified according to the identity and corresponding attributes. In the test stage, the part and global features are extracted through the trained network and concatenated as joint features. The similarity between pedestrian pairs is calculated to determine whether the pedestrian pairs form matches. In this method, the network locates pedestrian parts by using attribute information. At the same time, the attention mechanisms of the network are used to obtain discriminative features of pedestrian parts.

Result The proposed method uses attribute information in the Market-1501_attribute and DukeMTMC-attribute datasets and then evaluates the Market-1501 and DukeMTMC-reid datasets. In the field of person ReID, two evaluation protocols are used to evaluate the performance of the model: cumulative matching characteristics (CMC) curve and mean average precision (mAP). The following steps are performed to draw the CMC curve. First, we calculate the Euclidean distance between the probe image and the gallery images. Second, we sort the ranking order of all gallery images from the image with the shortest distance to the longest distance for each probe image. Finally, the true match percentage of the top m sorted images is calculated and called as rank- m . Specifically, rank-1 is an important indicator of person ReID. The following steps are performed to complete the mAP. First, we compute the average precision, which is the area under the precise-recall curve of each probe image. Next, the mean of average precision of all probe images is calculated. In the Market-1501 and DukeMTMC-reid datasets, the rank-1 accuracies are 90.67% and 80.2%, respectively, whereas the mAP accuracies are 76.65% and 62.14%, respectively. Recently, the re-ranking method has been widely used in the field of person ReID. In the re-ranking method, the rank-1 accuracies are 92.4% and 84.15%, whereas the mAP accuracies are 87.5% and 78.41%. Compared with other state-of-the-art methods, our accuracies are greatly improved.

Conclusion In this study, we propose a method for ReID based on attention and attributes. This proposed method can direct attention toward pedestrian parts by learning the attributes of pedestrians. Then, the attention mechanism can be used to learn the distinctive features of pedestrians. Even if pedestrians are occluded and misaligned, the proposed method can still achieve high accuracy.

Key words: person re-identification; global feature; local feature; pedestrians part; attention mechanism; attribute information

0 引言

行人再识别是指在多个非重叠摄像头中,按照指定的目标行人图像,在不同摄像头中检索出目标行人的技术,可以广泛运用于安防、图像管理、电子商务等领域。使用该技术,警方能够更快地跟踪和检索犯罪嫌疑人位置、更易找到失踪人口;用户可以更便捷地管理电子图像;无人超市的管理者可以追踪顾客的行为。但是使用中存在视频分辨率低、光照变化、背景干扰、行人动作多变、行人被遮挡等问题。为此,学者们提出了许多解决方法,包括传统手工设计方法和基于深度学习的方法。

传统手工设计方法主要分为特征设计和距离度量。特征设计的主要思路是设计具有辨别力且鲁棒性强的特征表达,使得在行人图像上能提取更具表现力的特征,主要是根据行人外观设计特征。Hamdoun 等人(2008)通过图像上积累的兴趣点获取行人的外观特征;Gray 和 Tao(2008)为克服摄像头视觉角度不变问题设计了局部集合特征(ensemble of localized features, ELF);Farenzena 等人(2010)根据行人的显著外观设计了对称驱动累积局部特征(symmetry-driven accumulation of local features, SDALF);Liao 等人(2015)为了克服视觉和光照问题,提出了增强的局部最大发生率(local maximal occurrence, LOMO)特征。距离度量的主要思路是设计一个距离测量方法,使同一行人图像间的距离更近、不同行人图像间的距离更远。Zheng 等人(2013)引入一种相对距离比较模型(relative distance comparison, RDC)使得正确匹配对的距离更近;Ma 等人(2013)提出域转移排序支持向量机(domain transfer ranked support vector machines, DTRSVM)方法用于目标域相机下的行人再识别;齐美彬等人(2018)设计了一种融合多种特征和交替方向乘子优化算法实现行人再识别。

随着深度学习的兴起,深度学习已经广泛运用于计算机视觉领域,而基于深度学习的行人再识别方法比传统手工设计方法取得了更高的识别率,主要包括 Siamese 网络(Nair 和 Hinton, 2010)模型和分类模型。Siamese 网络模型最早用于人脸识别领域,后来应用到行人再识别领域,并取得了不错的效果。Varior 等人(2016)设计了一种基于门控的 Sia-

messe 网络框架,其中引用了 Hadsell 等人(2006)提出的对比损失函数(contrastive loss),其本质是输入两幅图像进行二分类,判定是否为同一人;Schroff 等人(2015)提出了当今广泛运用的三元组损失(triplet loss)函数,与对比损失函数不同,1 个输入的三元组除了正负样本外,还新增 1 个固定照样本,使得同一行人的样本在特征空间聚类,达到行人再识别的目的。目前大量度量学习都是基于三元组损失函数演变而来,Chen 等人(2017)提出的四元组损失(quadruplet loss)比三元组多了 1 幅负样本;Hermans 等人(2017)提出的难样本采样三元组损失(TriHard loss)将三元组的随机抽取正负样本改进为挑取最难正负样本进行训练;Xiao 等人(2017)引入了难样本采样思想提出边界挖掘损失(margin sample mining loss, MSML)方法。这些方法主要用于 Siamese 网络,但也可用于分类网络。还有大部分工作是基于分类网络模型完成的,通过分类模型提取特征,然后进行相似度对比。Yi 等人(2014)提出了深度度量学习(deep metric learning, DML)方法,通过多信道过滤捕捉各种特征;Li 等人(2014)提出了一种过滤配对网络(filter pairing neural network, FPNN)模型,用于处理行人再识别中未对准、光照变换、遮挡和背景杂乱等问题;Ahmed 等人(2015)提出了一个同时学习特征和对应的相似性度量的网络模型;朱福庆等人(2018)设计了一种两路互补对称的卷积神经网络(convolutional neural networks, CNN)结构克服行人的变化。但这些方法都是假设行人图像可以很好地对齐,随着行人姿势的变化,这些方法会在一定程度上显得无效。

一些学者提出了基于注意力机制的深度学习方法进行行人再识别。Zhao 等人(2017)提出一种用于处理身体部位错位问题的模型,通过模拟身体部位提取和特征计算,用端到端方式学习模型参数;Li 等人(2017b)设计多尺度上下文感知网络(multi-scale context-aware network, MSCAN),能够同时学习行人全局和局部特征,通过网络中设计的堆叠多尺度卷积层,可以很好地对齐身体部位。注意力机制也时常用于学习行人的显著性特征,Liu 等人(2017)提出的 HP-net 网络将多层次的注意力图多向分发到不同的要素层,从而学到细粒度特征;Su 等人(2017)提出姿态驱动深度卷积模型(pose-driven deep convolutional, PDC),能从全局图像和不同的

局部图像学习更具表现力的特征表示; Song 等人 (2018) 设计一个掩模引导的对比注意模型 (mask-guided contrastive attention model, MGCAM) 学习身体和背景区域分开的特征; Zhou 等人 (2017) 在注意力机制的基础上引入语义属性进行行人再识别。语义属性是指一段自然语言或多个单词标签, 专门用来描述行人图像的外表特征, 许多工作通过引入语义属性取得了不错的成就。Layne 等人 (2012) 以属性特征为中心描述行人, 将中级语义属性与低级特征协同使用; Khamis 等人 (2015) 将语义集成到模型中, 有效利用属性和外观间的相互作用进行行人间的匹配; Li 等人 (2017b) 提出一个大型人物描述数据集, 包含各种人物图像信息的语言注释; Lin 等人 (2019) 在 Market-1501 (Zheng 等, 2015) 和 DukeMT-

MC-reid (Ristani 等, 2016) 数据集上标记属性标签并予以公开, 同时设计属性行人识别 (attribute-person recognition, APR) 网络, 使得能同时学习行人的身份和属性。

在行人再识别的研究中存在许多挑战, 因为行人对之间的不对齐、行人身体部位被遮挡导致的错误判断经常出现。由于人体的结构特点, 有时通过注意行人部件上的显著性特征判断, 而不关注其他部位的干扰信息, 反而更容易判断不同摄像头拍摄的行人对是否为同一人。本文基于注意力机制和多属性分类的行人再识别方法 (如图 1 所示), 就是通过注意力机制关注具有识别力的部件, 通过引入属性, 可以更快注意这些部位, 提高行人识别的准确率。

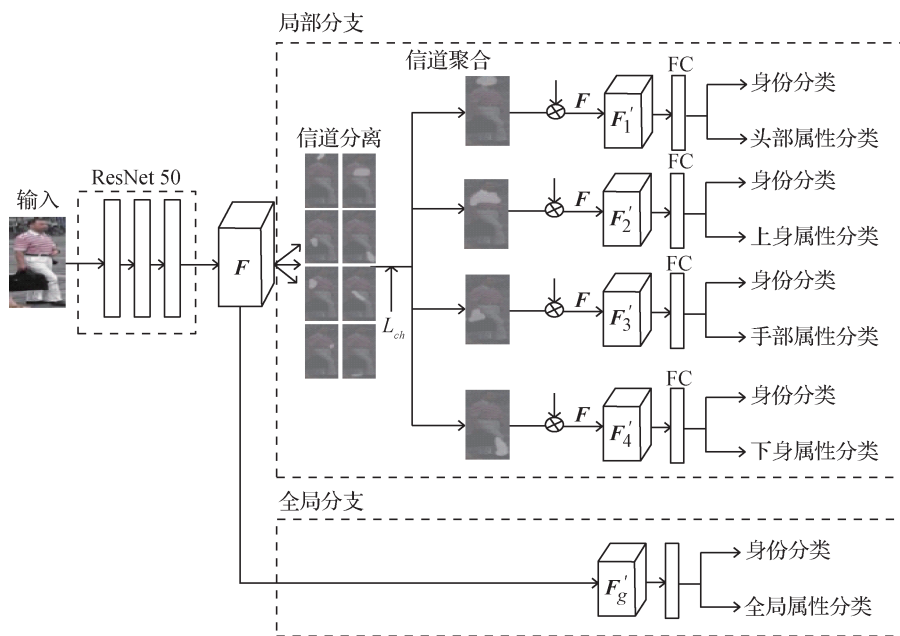


图1 注意力机制和多属性分类模型示意图

Fig. 1 Schematic diagram of attention mechanism and multi-attribute classification model

在训练阶段, 利用 ResNet50 网络作为特征提取的基本骨架, 在 ImageNet 数据集上进行网络预训练, 去掉最后一层全连接层后添加一层池化层, 得到特征向量。然后传入训练数据, 通过 ResNet50 网络获取并展开全局特征张量 F , 得到每层响应值最高的点, 聚合这些点, 将其分为 4 个行人部件。值得一提的是, 为了能端到端训练, 本文使用全连接层拟合这些点属于哪一个行人部件。最后将这 4 个部件的权重向量乘以特征 F 得到每个部件的总特征, 将这 4 个总特征和全局特征进行行人身份标签和属性标

签的分类。

在测试阶段, 首先将测试数据传入模型, 通过训练好的模型提取 4 个部件的局部特征和全局特征并全部级联起来。然后计算不同行人间的最终级联特征的欧氏距离, 得到行人间的相似度并以此判定是否为同一人。

本文的主要贡献是: 1) 提出了一个端到端训练的深度学习模型用于提取行人的全局特征, 并利用注意力机制对齐行人部件, 这种整体与局部联合互补学习的模型极大地提高了行人的识别率; 2) 引入

行人的属性标签,提出了一种属性和行人部件对应的方案,并设计相应的损失函数进行梯度反向传播训练,使模型在学习特征的同时,能学习部件对应的属性,进而更精确地定位行人的部位;3)提出的基于注意力机制和多属性分类的行人再识别方法在 Market-1501 和 DukeMTMC-reid 数据集上进行测试评估,取得了较高的准确率。

1 注意力机制和多属性分类方法

本文提出通过注意力机制和多属性分类进行行人再识别的方法,该方法提取特征的基本网络是改进后的 ResNet50 神经网络,为了得到初始特征 F ,去掉了 ResNet50 的最后一层全连接层,并加上了一层池化层。在取得初始特征后,该方法设计了两个分支:局部分支和全局分支。

1.1 局部分支

在局部分支中,首先将得到的初始特征 F 展开获得多个信道,然后在每层信道上找到响应值最高的点,聚合这些点并分成 4 部分,最后将这 4 部分的权重分别乘以初始特征 F 得到总特征,并据此进行行人身份分类和每个部件的属性分类。

初始特征 F 由 ResNet50 网络得到,假设输入的图像为 X , W 代表网络中的参数, $*$ 代表网络中一系列卷积操作,则特征 F 表示为

$$F = W * X \quad (1)$$

将特征中的信道一一展开,得到 n 个信道,再将 n 个信道中的响应峰值取出,就可以获得坐标位置,具体为 $[x_1, y_1, x_2, y_2, x_3, y_3, \dots, x_n, y_n]$ 。

一般情况下,可以通过均值聚类等算法将展开的信道根据响应峰值聚合起来,但是这样会使模型无法进行端到端训练,为此,使用全连接层进行聚合,过程如下:

1) 假设变量 w_{ij} 表示第 i 个信道是否属于第 j 个身体部位,若属于则为 1,否则为 0。那么所有信道对应的第 j ($j \in \{1, 2, 3, 4\}$) 个身体部位可表示为

$$w_j = [w_{1j}, w_{2j}, w_{3j}, \dots, w_{nj}] \quad (2)$$

2) 假设有 m 个身体部位,就设计 m 个全连接层 (fully connected layer, FC)。通过全连接层进行回归,将第 j 个全连接层的函数表示为 $f_j(\cdot)$, $j \in \{1, 2, 3, \dots, m\}$,则每一个全连接层输入初始特征值,在不同信道上产生的权重向量为

$$c_j = f_j(F) \quad (3)$$

式中, $c_j = [c_1, c_2, c_3, \dots, c_n]$ 。通过式 (3) 拟合式 (2) 初始化全连接层的参数,然后由网络进行端到端训练。为了使每一个信道仅属于一个行人部位,本文使用损失函数进行约束。通过信道聚合,可以获得每一个行人部位的注意力图 Pa_j ,具体为

$$Pa_j = \text{Sigmoid}\left(\sum_{i=1}^n c_i F_i\right) \quad (4)$$

式中, c_i 是一个标量,代表向量 c_j 中的值; F_i 指特征 F 的第 i 层信道; $c_i F_i$ 之间进行的操作是标量与矩阵之间的乘法。将每个属于第 j 个部位的信道相加,使用 Sigmoid 函数计算得到每一个行人的部位注意力图。将局部注意力图 Pa_j 与初始特征 F 中的每一个信道进行对应元素相乘 (element-wise),即可获得每个部位 j 的最后特征 F'_j ,即

$$F'_j = F \cdot Pa_j \quad (5)$$

3) 部位的最后特征 F'_j 经过全连接层后进行行人的身份分类和属性分类。本文将行人的属性分为 4 类,分别对应行人的相应部位,例如 4 个部位分别对应人体的头部、上半身、手部和下半身。

1.2 全局分支

由于网络在局部分支中主要学习部件的显著性特征,容易忽略全局性的相关特征,因此本文设计了全局分支,通过初始特征 F 进行卷积得到全局特征 F'_g 进行行人身份分类和属性分类,从而学习有代表性的全局特征。

在全局分支中,依次设计了卷积层、BN (batch normalization) 层 (Ioffe 和 Szegedy, 2015) 和全连接层。卷积层的卷积核大小是 1,输出的信道 (channels) 数为 512,这样全局分支输出的信道数与每个部件输出的信道数一致,方便在测试阶段进行特征串联。BN 层让网络能学习更好的参数。网络的初始化参数、学习率、权重衰减系数、dropout 系数等参数对训练至关重要,设置 BN 层可以极大简化调整参数的工作。全连接层的设计是为了对行人身份和全局属性进行分类,通过网络预测的标签与真实标签计算相应的损失函数,然后由损失函数进行网络的反向传播,从而使得网络学习到具有代表性的特征。

1.3 损失函数

本文损失函数主要包含分类损失和信道聚合损失,即总体损失函数

$$Loss = L_{cls} + L_{ch} \quad (6)$$

式中, L_{ch} 为信道聚合损失。 L_{cls} 为分类损失, 包括行人身份分类损失 L_{ID} 和对应的属性分类损失 L_{attr} , 为了便于表达, 本文将全局分类和局部的部件分类合并为

$$L_{cls} = \frac{1}{m+1} \sum_{i=1}^{m+1} \left(\lambda L_{ID}(Y_{ID}^*, Y_{ID}) + \frac{1}{M} \sum_{j=1}^M L_{attr}(Y_{attr}^*, Y_{attr}) \right) \quad (7)$$

式中, λ 是平衡行人身份和属性分类损失的调整参数, m 代表局部部件总数, M 代表每个部件对应的属性个数, Y_{ID}^* 和 Y_{attr}^* 分别表示实际的身份和属性标签, Y_{ID} 和 Y_{attr} 分别表示预测的身份和属性。 L_{ID} 和 L_{attr} 使用的是标准的交叉熵损失函数, 定义过程(Lin等, 2019)如下:

对于 k 个输出, 全连接层的输出 $z = [z_1, z_2, z_3, \dots, z_k] \in \mathbf{R}^K$, 每一个预测的输出概率

$$p(k) = \frac{\exp(z_k)}{\sum_{i=1}^K \exp(z_i)} \quad (8)$$

则身份分类和属性分类的交叉熵损失分别为

$$L_{ID}(Y_{ID}^*, Y_{ID}) = - \sum_{k=1}^K \log(p(k)) q(k) \quad (9)$$

$$L_{attr}(Y_{attr}^*, Y_{attr}) = - \sum_{k=1}^{Attr} \log(p(k)) q(k) \quad (10)$$

式中, K 代表行人总数, $Attr$ 代表属性分类中输出个数, 预测正确时 $q(k) = 1$, 否则 $q(k) = 0$ 。

式(6)中的信道聚合损失 L_{ch} 是为了使同一部位的坐标尽量聚集、不同部位尽量疏远, 具体定义为

$$L_{ch} = Dis(\mathbf{Pa}_j) + \alpha Div(\mathbf{Pa}_j) \quad (11)$$

式中, α 是平衡部件内聚合点的距离 $Dis(\mathbf{Pa}_j)$ 与部件间的距离 $Div(\mathbf{Pa}_j)$ 的系数, $Dis(\mathbf{Pa}_j)$ 和 $Div(\mathbf{Pa}_j)$ 的定义(Zheng等, 2017)分别为

$$Dis(\mathbf{Pa}_j) = \sum_{(x,y) \in \mathbf{Pa}_j} pa_j(x,y) [\|x - x_j\|^2 + \|y - y_j\|^2] \quad (12)$$

$$Div(\mathbf{Pa}_j) = \sum_{(x,y) \in \mathbf{Pa}_j} pa_j(x,y) [\max_{k \neq j} pa_k(x,y) - margin] \quad (13)$$

式中, $pa_j(x,y)$ 是式(4)中 $\mathbf{Pa}_j$ 坐标为 (x,y) 的值, 代表点 (x,y) 注意力的权重, x_j 和 y_j 指的是信道峰值坐标, 通过 $Dis(\mathbf{Pa}_j)$ 可使部位内的点能更好聚集。 k 和 j 表示不同注意力图的索引值, $margin$ 表示

边距, 是为了防止损失对噪声过度敏感, 其值是根据实验调参得到的, 实验中设置为 0.02, $Div(\mathbf{Pa}_j)$ 的作用是为了使不同部位间的距离增大。

由于网络设计的是端到端训练, 所以在训练过程中, 通过式(6)的损失函数进行反向传播, 训练出网络中的参数即可。 在最后测试中, 将部件的特征 F'_1 、 F'_2 、 F'_3 、 F'_4 和全局特征 F'_g 串联(concatenate)起来, 进行标准的行人再识别操作(Zheng等, 2015)。

2 实验

2.1 实验数据集

实验使用了4个数据集: Market-1501、DukeMTMC-reid、Market-1501_attribute 和 DukeMTMC-attribute(Lin等, 2019), 后两个数据集是前两个数据集在属性信息上的扩展。 本文通过上述数据集评估了所提方法的有效性。

Market-1501 数据集由6个摄像头拍摄的1501个行人和32668个行人矩形框组成, 其中, 训练集有751个行人, 包括12936幅图像; 测试集有750个行人, 包括19732幅图像。 query 为测试集中750人在6个摄像头中随机挑选出的图像集, 共3368幅图像。

Market-1501_attribute 数据集是对 Market-1501 数据集的行人属性标注后的公开数据集, 为每一个行人标注了27个属性, 包括性别、头发长度、袖子长度、下半身衣服长度、下半身衣服类型、是否戴帽子、是否背双肩包、是否有手提包、年龄、上衣的8种颜色(黑、白、红、紫、黄、灰、蓝、绿)和下身衣服的9种颜色(黑、白、粉、紫、黄、灰、蓝、绿、棕)。

DukeMTMC-reid 包括8个摄像头拍摄的1404个行人的36411幅图像, 其中, 训练集有702个行人, 包括16522幅图像; 测试集有702个行人, 包括17661幅图像。 query 为测试集中702人在8个摄像头中随机选择出的图像集, 共2228幅。

DukeMTMC-attribute 是对 DukeMTMC-reid 数据集的行人属性标注后的公开数据集, 为每一个行人标注了23个属性, 包括性别、上衣长度、是否穿靴子、是否戴帽子、是否背双肩包、是否有包、是否有手提包、鞋子颜色、上衣的8种颜色(黑、白、红、紫、灰、蓝、绿、棕)、下身衣服的7种颜色(黑、白、红、

灰、蓝、绿、棕)。

实验中将上述标注的属性分为局部部位属性和全局属性。局部部位包括头部、上半身、手部、下半身,全局属性指性别和年龄。网络通过对相应部位进行学习,能使得注意力机制更好地关注部位的细节,从而提取更有代表性的特征。

2.2 实验的评估标准

实验基于深度学习框架 MXNet 实现,硬件配置如下:GPU 为 GTX 1080 Ti(显存:11 GB)、CPU 为 Intel(R) Xeon(R) CPU E5-2696 v3 @ 2.30 GHz、内存 62 GB。

在行人再识别领域,普遍通过累积匹配特征(cumulative matching characteristics, CMC)曲线和平均精度均值(mean average precision, mAP)评估模型性能。累积匹配特征指前 K 幅匹配成功的概率,本文采用第 1 幅就匹配成功的概率,记为 rank-1。平均精度均值是计算所有样本的准确率—召回率曲线下的面积的平均值。测试时从 query 中取 1 幅图像与测试集中所有图像匹配,计算相似度并排序。

2.3 实验的结果及分析

基线方法采用 APR(attribute-person recognition)网络(Lin 等, 2019),因为 Market-1501_attribute 和 DukeMTMC-attribute 两个数据集是 Lin 等人(2019)标注并公开的,而且 APR 网络与本文方法一样,都是同时学习了行人身份和属性信息,因此与 APR 网络进行比较,可以验证本文方法的有效性。

在 Market-1501 数据集中,将本文方法与基线方法以及 MSCAN(multi-scale context-aware network)(Li 等, 2017a)、DLPA(deeply-learned part-aligned)(Zhao 等, 2017)、PDC(Su 等, 2017)、JLML(jointly learning multi-loss)(Li 等, 2017c)、TriNet(Hermans 等, 2017)、CamStyle(Zhong 等, 2017)、MobileNet + DML(Zhang 等, 2017)、PSE(pose-sensitive embedding)(Sarfranz 等, 2018)和 AWTL(adaptive weighted triplet loss)(Ristani 和 Tomasi, 2018)等近几年行人再识别领域具有代表性的方法进行对比,其中 MSCAN 也是引用了注意力机制定位行人部位, PSE 方法引入了行人身份标签和姿态信息,这些方法在行人再识别领域都取得了很好成就。对比结果如表 1 所示。

从表 1 可以看出,与基线方法相比,本文方法的准确率取得了极大提升,在单查询中,rank-1 和 mAP 分别提升了 6.37% 和 11.9%;在多查询中,rank-1

表 1 本文方法与其他方法在 Market-1501

数据集上的比较结果

Table 1 Comparison results of our method and other methods on Market-1501 dataset

方法	单查询		多查询	
	rank-1	mAP	rank-1	mAP
APR(基线)	84.29	64.67	—	—
MSCAN	80.3	57.5	86.8	66.7
DLPA	81.0	63.4	—	—
PDC	84.1	63.4	—	—
JLML	85.1	65.5	89.7	74.5
TriNet	84.92	69.14	90.53	76.42
CamStyle	88.12	68.72	—	—
mobileNet + DML	87.73	68.83	91.66	77.14
PSE	87.7	69.0	—	—
AWTL	89.46	75.67	—	—
本文	90.67	76.65	94.0	83.2

注:加粗字体表示最优结果,“—”表示无数据。

和 mAP 分别达到 94% 和 83.2%,说明本文方法在处理行人再识别问题上非常有效。与其他方法相比,rank-1 和 mAP 均占优,说明本文方法具有很好的竞争力。

目前行人再识别领域广泛使用 Zhong 等人(2018)提出的重排序(re-ranking)方法,该方法通过编码特征降低计算量,引入 Jaccard 距离与原始距离做加权处理后进行最终的排序计算,极大地提升了行人再识别的性能。在都引用 re-ranking 技术的情况下,本文方法与其他代表性算法在 Market-1501 数据集上进行了比较,结果如表 2 所示。

表 2 本文方法与其他引用重排序技术的方法在

Market-1501 数据集上的比较结果

Table 2 Comparison results of our method and other methods on Market-1501 dataset by re-ranking

方法	%	
	rank-1	mAP
JLML + re-ranking	88.8	72.9
TriNet + re-ranking	86.67	81.07
CamStyle + re-ranking	89.49	71.55
PSE + re-ranking	90.0	84.0
本文 + re-ranking	92.4	87.5

注:加粗字体表示最优结果。

将表 2 与表 1 对比,可以看出在使用 re-ranking 后,同一方法的指标得到了极大提升。JLML、Tri-Net、CamStyle、PSE 和本文方法的 rank-1 分别提升了 3.7%、1.75%、1.36%、2.3%、1.73%,mAP 分别提升了 7.4%、11.93%、2.83%、15%、10.85%。

从表 2 可以看出,本文方法与有代表性的 PSE 方法相比,rank-1 和 mAP 分别提高了 2.4% 和 3.5%,说明在 re-ranking 方法下,本文方法也具有一定的优越性。

为进一步验证本文方法的性能,与基线方法以及 JLML、SVDNet-ResNet50(Sun 等,2017)、CamStyle 和 PSE 等方法在 DukeMTMC-reid 数据集上进行比较,结果如表 3 所示。

表 3 本文方法与其他方法在 DukeMTMC-reid 数据集上的比较结果
Table 3 Comparison results of our method and other methods on DukeMTMC-reid dataset

方法	rank-1	mAP
APR(基线)	70.69	51.88
JLML	73.3	56.4
SVDNet-ResNet50	76.7	56.8
CamStyle	75.27	53.48
PSE	79.8	62.0
本文	80.2	62.14
本文 + re-ranking	84.15	78.41

注:加粗字体表示最优结果,斜体表示次优结果。

从表 3 可以看出:1)本文方法的 rank-1 和 mAP 分别比基线方法高 9.51% 和 10.25%,在 re-ranking 方法下,rank-1 和 mAP 比其线方法分别增加 3.95% 和 16.27%,说明通过注意力机制和属性的联合训练可以显著提高行人再识别的准确率。2)由于 DukeMTMC-reid 数据集中的图像大多不对齐,且许多图像中的行人被遮挡,因此导致许多方法的性能指标不高,仅 PSE 引入姿态信息进行行人对齐处理,得到的指标较高,而本文方法是通过提出注意力机制和多属性分类的方法解决这些问题,取得了较高的指标。

3 结 论

行人再识别领域一直存在着许多挑战,现有的

方法是直接引用属性进行分类来提取行人特征,但没有充分利用属性信息,并且预设行人图像天然对齐。而在实际场景中,行人不对齐、被遮挡等问题经常出现,对此,本文提出基于注意力机制和多属性分类的行人再识别方法,根据人体结构将属性分成全局属性和局部属性。在全局分支中,通过全局属性学习行人的全局特征;在局部分支中,通过学习局部属性使得行人部件的注意力能更快聚集。注意力机制使得网络能够更好地学习行人部位的显著性特征。最后将全局特征和行人部位的显著性特征进行串联,计算行人间的相似度。与直接利用属性分类的方法相比,本文方法更好地利用了属性信息,并且通过注意力机制学习行人部件的显著性特征,忽略带有干扰信息的部件特征,能在一定程度上处理行人不对齐、被遮挡等问题,从而使得性能指标得到提升。实验表明,与直接使用属性的方法相比,在 Market-1501 和 DukeMTMC-reid 数据集上,rank-1 分别提升了 6.37% 和 9.51%;mAP 分别提升了 11.9% 和 10.25%,说明本文方法在性能上具有一定的优越性。使用 re-ranking 方法后,性能进一步提升,与其他使用 re-ranking 的方法相比仍具有一定的优越性,证明本文方法提取的特征的确具有足够的辨别力。

本文方法使用的是有监督学习,同时引用行人身份和属性信息,在两个数据集上取得了较高的指标,但尚存不足。例如,属性信息由人工标注,代价比较高昂,如何能够使用少量标注甚至无标注数据进行行人再识别是今后的研究方向。目前,有监督行人再识别方法在数据集上取得的效果越来越好,而无监督行人再识别方法还有进一步的研究空间,如何使用半监督或者无监督学习进行行人再识别也是未来的研究方向。

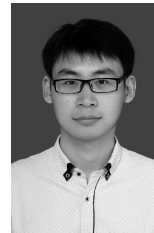
参考文献 (References)

- Ahmed E, Jones M and Marks T K. 2015. An improved deep learning architecture for person re-identification//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE: 3908-3916 [DOI: 10.1109/CVPR.2015.7299016]
- Chen W H, Chen X T, Zhang J G and Huang K Q. 2017. Beyond triplet loss: a deep quadruplet network for person re-identification//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 1320-1329 [DOI: 10.1109/CVPR.

2017. 145]
- Farenzena M, Bazzani L, Perina A, Murino V and Cristani M. 2010. Person re-identification by symmetry-driven accumulation of local features//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE; 2360-2367 [DOI: 10.1109/CVPR.2010.5539926]
- Gray D and Tao H. 2008. Viewpoint invariant pedestrian recognition with an ensemble of localized features//Proceedings of the 10th European Conference on Computer Vision. Marseille, France: Springer; 262-275 [DOI: 10.1007/978-3-540-88682-2_21]
- Hadsell R, Chopra S and LeCun Y. 2006. Dimensionality reduction by learning an invariant mapping//Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE; 1735-1742 [DOI: 10.1109/CVPR.2006.100]
- Hamdoun O, Moutarde F, Stanculescu B and Steux B. 2008. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences//Proceedings of the 2nd ACM/IEEE International Conference on Distributed Smart Cameras. Stanford: IEEE; 1-6 [DOI: 10.1109/ICDSC.2008.4635689]
- Hermans A, Beyer L and Leibe B. 2017. In defense of the triplet loss for person re-identification [EB/OL]. [2019-05-10]. <https://arxiv.org/pdf/1703.07737.pdf>
- Ioffe S and Szegedy C. 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift//Proceedings of the 32nd International Conference on Machine Learning. Lille, France: JMLR; 448-456
- Khamis S, Kuo C H, Singh V K, Shet V D and Davis L S. 2015. Joint learning for attribute-consistent person re-identification//Proceedings of European Conference on Computer Vision. Zurich, Switzerland: Springer; 134-146 [DOI: 10.1007/978-3-319-16199-0_10]
- Layne R, Hospedales T and Gong S G. 2012. Person re-identification by attributes//Proceedings of the British Machine Vision Conference. [s. l.]; BMVA Press; 24. 1-11 [DOI: 10.5244/C.26.24]
- Li D W, Chen X T, Zhang Z and Huang K Q. 2017a. Learning deep context-aware features over body and latent parts for person re-identification//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE; 7398-7407 [DOI: 10.1109/CVPR.2017.782]
- Li S, Xiao T, Li H S, Zhou B L, Yue D Y and Wang X G. 2017b. Person search with natural language description//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE; 5187-5196 [DOI: 10.1109/CVPR.2017.551]
- Li W, Zhao R, Xiao T and Wang X G. 2014. DeepReID: deep filter pairing neural network for person re-identification//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE; 152-159 [DOI: 10.1109/CVPR.2014.27]
- Li W, Zhu X T and Gong S G. 2017c. Person re-identification by deep joint learning of multi-loss classification//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia; [s. n.]; 2194-2200 [DOI: 10.24963/ijcai.2017/305]
- Liao S C, Hu Y, Zhu X Y and Li S Z. 2015. Person re-identification by local maximal occurrence representation and metric learning//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE; 2197-2206 [DOI: 10.1109/CVPR.2015.7298832]
- Lin Y T, Zheng L, Zheng Z D, Wu Y, Hu Z L, Yan C G and Yang Y. 2019. Improving person re-identification by attribute and identity learning. Pattern Recognition, 95: 151-161 [DOI: 10.1016/j.patcog.2019.06.006]
- Liu X H, Zhao H Y, Tian M Q, Sheng L, Shao J, Yi S, Yan J J and Wang X G. 2017. HydraPlus-Net: attentive deep features for pedestrian analysis//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE; 350-359 [DOI: 10.1109/ICCV.2017.46]
- Ma A J, Yuen P C and Li J W. 2013. Domain transfer support vector ranking for person re-identification without target camera label information//Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney: IEEE; 3567-3574 [DOI: 10.1109/ICCV.2013.443]
- Nair V and Hinton G E. 2010. Rectified linear units improve restricted Boltzmann machines//Proceedings of the 27th International Conference on Machine Learning. Haifa: Omnipress; 807-814
- Qi M B, Wang C C, Jiang J G and Li J. 2018. Person re-identification based on multi-feature fusion and alternating direction method of multipliers. Journal of Image and Graphics, 23(6): 827-836 (齐美彬, 王慈淳, 蒋建国, 李佶. 2018. 多特征融合与交替方向乘子法的行人再识别. 中国图象图形学报, 23(6): 827-836) [DOI: 10.11834/jig.170507]
- Ristani E and Tomasi C. 2018. Features for multi-target multi-camera tracking and re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE; 6036-6046 [DOI: 10.1109/CVPR.2018.00632]
- Ristani E, Solera F, Zou R, Cucchiara R and Tomasi C. 2016. Performance measures and a data set for multi-target, multi-camera tracking//Proceedings of European Conference on Computer Vision. Amsterdam, The Netherlands: Springer; 17-35 [DOI: 10.1007/978-3-319-48881-3_2]
- Sarfraz M S, Schumann A, Eberle A and Stiefelhofen R. 2018. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE; 420-429 [DOI: 10.1109/CVPR.2018.00051]
- Schroff F, Kalenichenko D and Philbin J. 2015. FaceNet: a unified embedding for face recognition and clustering//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE; 815-823 [DOI: 10.1109/CVPR.2015.7298682]

- Song C F, Huang Y, Ouyang W L and Wang L. 2018. Mask-guided contrastive attention model for person re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 1179-1188 [DOI: 10.1109/CVPR.2018.00129]
- Su C, Li J N, Zhang S L, Xing J L, Gao W and Tian Q. 2017. Pose-driven deep convolutional model for person re-identification//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE: 3980-3989 [DOI: 10.1109/ICCV.2017.427]
- Sun Y F, Zheng L, Deng W J and Wang S J. 2017. SVDNet for pedestrian retrieval//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE: 3820-3828 [DOI: 10.1109/ICCV.2017.410]
- Varior R R, Haloi M and Wang G. 2016. Gated siamese convolutional neural network architecture for human re-identification//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer: 791-808 [DOI: 10.1007/978-3-319-46484-8_48]
- Xiao Q Q, Luo H and Zhang C. 2017. Margin sample mining loss: a deep learning based method for person re-identification [EB/OL]. [2019-05-07]. <https://arxiv.org/pdf/1710.00478.pdf>
- Yi D, Lei Z, Liao S C and Li S Z. 2014. Deep metric learning for person re-identification//Proceedings of the 22nd International Conference on Pattern Recognition. Stockholm: IEEE: 34-39 [DOI: 10.1109/ICPR.2014.16]
- Zhang Y, Xiang T, Hospedales T M and Lu H C. 2017. Deep mutual learning [EB/OL]. [2019-05-07]. <https://arxiv.org/pdf/1706.00384.pdf>
- Zhao L M, Li X, Wang Y T and Wang J D. 2017. Deeply-learned part-aligned representations for person re-identification//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE: 3239-3248 [DOI: 10.1109/ICCV.2017.349]
- Zheng H L, Fu J L, Mei T and Luo J B. 2017. Learning multi-attention convolutional neural network for fine-grained image recognition//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE: 5219-5227 [DOI: 10.1109/ICCV.2017.557]
- Zheng L, Shen L Y, Tian L, Wang S J and Wang J D. 2015. Scalable person re-identification: a benchmark//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 1116-1124 [DOI: 10.1109/ICCV.2015.133]
- Zheng W S, Gong S G and Xiang T. 2013. Reidentification by relative distance comparison. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(3): 653-668 [DOI: 10.1109/TPAMI.2012.138]
- Zhong Z, Zheng L, Cao D L and Li S Z. 2017. Re-ranking person re-identification with k-reciprocal encoding//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE: 3652-3661 [DOI: 10.1109/CVPR.2017.389]
- Zhong Z, Zheng L, Zheng Z D, Li S Z and Yang Y. 2018. Camera style adaptation for person re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 5157-5166 [DOI: 10.1109/CVPR.2018.00541]
- Zhou T, Chen M H, Yu J and Terzopoulos D. 2017. Attention-based natural language person retrieval//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu: IEEE: 27-34 [DOI: 10.1109/CVPRW.2017.10]
- Zhu F Q, Kong X W, Fu H Y and Tian Q. 2018. Two-stream complementary symmetrical CNN architecture for person re-identification. Journal of Image and Graphics, 23(7): 1052-1060 (朱福庆, 孔祥维, 付海燕, 田奇. 2018. 两路互补对称 CNN 结构的行人再识别. 中国图象图形学报, 23(7): 1052-1060) [DOI: 10.11834/jig.170557]

作者简介



郑鑫, 1993年生, 男, 硕士研究生, 主要研究方向为机器学习和计算机视觉。

E-mail: iceborain@163.com



贺春林, 通信作者, 男, 教授, 主要研究方向为模式识别。

E-mail: 93401318@qq.com

林兰, 女, 博士研究生, 主要研究方向为计算机视觉和机器学习。E-mail: linlan0921@163.com

叶茂, 男, 教授, 博士生导师, 主要研究方向为机器学习与计算机视觉。E-mail: cvlab.uestc@gmail.com

王丽, 女, 硕士研究生, 主要研究方向为计算机视觉。

E-mail: 781477696@qq.com