基于时空共现模式的视觉行人再识别

钱锦浩 [1] 宋展仁 [1] 郭春超 [1] 赖剑煌 [1, 2, 3] 谢晓华 [1, 2, 3]

摘 要 基于视频图像的视觉行人再识别是指利用计算机视觉技术关联非重叠域摄像头网络下的相同行人, 在视频安防和商业客流分析中具有重要应用. 目前视觉行人再识别技术已经取得了相当不错的进展, 但依旧面临很多挑战, 比如摄像机的拍摄视角不同、遮挡现象和光照变化等所导致的行人表观变化和匹配不准确问题. 为了克服单纯视觉匹配困难问题, 本文提出一种结合行人表观特征跟行人时空共现模式的行人再识别方法. 所提方法利用目标行人的邻域行人分布信息来辅助行人相似度计算, 有效地利用时空上下文信息来加强视觉行人再识别. 在行人再识别两个权威公开数据集 Market-1501 和 DukeMTMC-ReID 上的实验验证了所提方法的有效性.

关键词 行人再识别, 深度学习, 时空共现模式, 行人邻域

引用格式 钱锦浩, 宋展仁, 郭春超, 赖剑煌, 谢晓华. 基于时空共现模式的视觉行人再识别. 自动化学报, 2022, **48**(2): 408−417

DOI 10.16383/j.aas.c200897

Visual Person Re-identification Based on Spatial and Temporal Co-occurrence Patterns

QIAN Jin-Hao[1] SONG Zhan-Ren[1] GUO Chun-Chao[1] LAI Jian-Huang[1, 2, 3] XIE Xiao-Hua[1, 2, 3]

Abstract Person re-identification technology plays important roles in video surveillance of security and customer analysis of business, which is to associate the same person under the non-overlapping camera network. At present, the technology of person re-identification has made great progress, however, it still faces many challenges, such as the appearance changes and inaccurate matching of pedestrians caused by different camera viewpoints, occlusion, and illumination changes. In this paper, a method combining the appearance features with the spatial and temporal co-occurrence pattern is proposed. The proposed method strengthens the computation of the similarity between pedestrian images by using the association of surrounding pedestrians of the target pedestrian. The proposed method effectively utilizes spatiotemporal context information to enhance visual person re-identification. Experiments on two public data sets, namely Market-1501 and DukeMTMC-ReID, verify the effectiveness of the proposed method.

Key words Person re-identification, deep learning, co-occurrence pattern, person neighbourhood

Citation Qian Jin-Hao, Song Zhan-Ren, Guo Chun-Chao, Lai Jian-Huang, Xie Xiao-Hua. Visual person re-identification based on spatial and temporal co-occurrence patterns. *Acta Automatica Sinica*, 2022, **48**(2): 408−417

目前, 随着 “智慧城市”和 “平安城市”等项目建设, 众多公共场所均部署了大量的监控摄像头, 形成了庞大的监控摄像头网络. 对这些摄像头的内容进行关联分析显得越来越重要, 这也是计算机视觉领域当前研究热点之一. 行人再识别 (Person re-identification) 技术旨在判断跨摄像头视域下的多个行人图像是否来自同一行人[1]. 行人再识别技术能够进一步应用于跨摄像头下的目标追踪、目标路径分析以及目标搜索等问题. 该技术实现了监控视频的智能关联分析, 其在智慧城市、公共安全、商业客流分析、城市安防和视频图像大数据处理等方面扮演极其重要的角色, 具备非常广泛的应用场景.

目前基于视频图像的行人再识别领域的研究工作主要分为两个类别, 分别是基于表征的方法以及基于度量学习的方法. 这两类方法分别旨在寻找识别性强的特征表达与学习特征间相似度度量, 使得相同身份的行人之间的相似度较大, 相异身份的行人之间的相似度较小. 随着深度学习技术的发展, 以上两类方法逐渐达成紧密结合. 然而, 这两类方

收稿日期 2020-10-26 录用日期 2021-04-16
Manuscript received October 26, 2020; accepted April 16, 2021
国家自然科学基金 (62072482, 62076258), 广东省信息安全技术重点实验室开放课题基金 (2017B030314131) ,公安部科技强警基础工作专项项目 (2019GABJC39) 资助
Supported by National Natural Science Foundation of China (62072482, 62076258), the Opening Project of GuangDong Province Key Laboratory of Information Security Technology (2017B030314131) and the Key Laboratory of Video and Image Intelligent Analysis and Applicaiton Technology, Ministry of Public Security, China (2019GABJC39)
本文责任编委 薛建儒
Recommended by Associate Editor XUE Jian-Ru
1. 中山大学计算机学院 广州 510006 2. 视频图像智能分析与应用技术公安部重点实验室 广州 510006 3. 机器智能与先进计算教育部重点实验室 广州 510006
1. School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006 2. Key Laboratory of Video and Image Intelligent Analysis and Applicaiton Technology, Ministry of Public Security, China, Guangzhou 510006 3. Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, Guangzhou 510006

法的研究重点均聚焦于行人的表观视觉信息. 由于现实场景中的目标行人姿态变化多端, 加之环境遮挡物的影响、拍摄角度和距离的改变以及光照的变化, 监控摄像头拍摄的行人视频图像会呈现较大变化, 这无疑为单纯依靠视觉匹配的行人再识别带来巨大挑战.

针对单纯视觉识别的不足, 研究人员开始应用各种上下文信息用于补充视觉匹配, 比如视频图像采集的时空信息[2−4]、人群辅助[5−7] 等. 其中, 人群辅助方法主要基于这种观察: 在实际人流中经常存在相对稳定的小群体, 这种群体也许是互相认识的同伴, 也许是由于某些特殊原因形成相同时空轨迹的陌生人小群体 (譬如在火车站相同班次到站的人群). 这种相对稳定小群体对特定行人的再识别具有积极的辅助作用.

根据上面分析, 本文将人群定义为一个时间窗口内从同一摄像头下经过的行人集合. 基于此定义, 本文提出了一种结合表观特征与行人时空共现模式的行人再识别方法. 所提方法把现实中行人之间的时空联系看作是一种共同出现的模式状态作为上下文信息来辅助行人相似度的计算. 本文在行人再识别两个权威的公开数据集 Market-1501[8] 和 DukeMTMC-ReID[9] 上对该方法的有效性进行了实验验证.

1 相关工作

视觉行人再识别技术狭义上包括对行人的特征表达以及行人的跨摄像头匹配. 因此行人再识别相关的绝大部分研究主要侧重于两个方面: 行人视觉特征表达和度量学习. 本节简要介绍这两类研究成果, 同时介绍采用人群辅助行人再识别的相关方法.

1.1 行人特征表达

行人图像的特征表达方法逐渐从手工设计特征向深度学习特征过渡. 手工设计特征主要有颜色特征、纹理特征、属性特征、形状和关键点特征等. 其中颜色特征[10] 是描述行人最为简单、直观的特征, 主要包含颜色名称和基于统计特性的颜色直方图. 颜色名称特征使用具体的颜色名称作为特征, 物理意义明确, 表达简洁高效. 行人图像的纹理特征是描述行人表面性质的统计特征, 具有较强的抗噪性质与旋转不变性. 行人再识别中常用的纹理特征有 Gabor 特征[11]、局部二值模式特征[12] (Local binary pattern, LBP) 和 Schmid 特征[13] 等. 形状和关键点特征是通过图像的形状特征与关键点的信息来描述局部的特征, 主要包含方向梯度直方图特征[14] (Histogram of oriented gradient, HOG)、尺度不变特征变换特征[15] (Scale invariant feature transform, SIFT) 和加速鲁棒特征[16] (Speeded up robust features, SURF). 描述行人图像的属性特征[17−19] 是更加接近人类的认知方式, 其作为辅助信息提升了行人再识别算法模型的泛化能力. 属性特征一般包括生物属性、附属物品属性和服装属性特征等, 如行人图像的发型、性别、着装、是否携带背包等.

得益于深度卷积神经网络的发展, 深度学习成为行人视觉特征学习的基准算法. 深度学习特征是行人图像的多层抽象语义特征, 对行人图像有着更加精确的表达. 近年来行人再识别在新任务和新方法上都取得了不错的进展, 例如 Chen 等[20] 提出使用分类子网络与验证子网络相结合进行训练网络; Jing 等[21] 提出了一种半耦合低秩判别字典学习方法 (Semi-coupled low-rank discriminant dictionary learning, SLD2L) 用于超分辨率行人再识别; Ma 等[22] 提出一种不对称的视频内投影半耦合字典对学习方法 (Semi-coupled dictionary pair learning, SDPL) 用以解决彩色到灰色视频行人再识别问题; Zhu 等[23] 提出了一种基于视频行人再识别的视频内和视频间同步远程学习方法 (Simultaneously learning intra-video and inter-video distance metrics, SI2DL); Zhang 等[24] 提出多尺度时空注意力 (Multi-scale spatial-temporal attention, MSTA) 模型, 着重于在空间和时间两方面挖掘每帧局部区域对整个视频表示的重要性; Wu 等[25] 利用位姿对齐连接和特征亲和连接构造自适应结构感知邻接图, 并通过图神经网络学习高判别性特征; Wang 等[26] 利用时间差信息提出一种既挖掘视觉语义信息又挖掘时空信息的双流时空行人再识别框架. 与已有大多数方法不同, 本文不侧重于表观或者时序运动特征的提取, 而关注利用目标行人的邻域行人分布信息来辅助行人相似度计算.

1.2 行人度量学习

度量学习即通过找到一种度量行人特征间相似度的准则, 使得相同身份的行人之间的相似度较大, 相异身份的行人之间的相似度较小. 基于度量学习的方法在损失函数上体现为相同身份的行人图像对之间的距离小于不同身份行人图像对之间的距离. 常用的度量学习损失函数主要有对比损失[27]、三元组损失[1, 28−29]、困难样本三元组损失[30−31]、四元组损失[32] 以及边界挖掘损失[33] 等.

1.3 人群辅助方法

除了表征学习和度量学习, 部分研究者尝试利

用人群作为辅助信息来增强行人再识别的准确率. 在文献 [5] 中, 作者提出运用行人群体匹配来辅助个体匹配, 着重研究了在同一图像上的人群的视觉描述子. 然而在实际中, 由于摄像头视域的限制, 特定人群的行人未必会同时出现在同一视频帧上, 可能随着时间依次出现于监控画面. 因此, 本文考虑的人群范围比文献 [5] 中的人群范围更广. 除此之外, 有研究者提出将人群作为一个弱标注信息并应用于弱监督学习行人再识别方法中[6−7]. 上述方法都将人群定义局限于共同出现在同一视频帧的行人组合, 且重点关注目标行人与人群的从属关系. 本文所提方法所定义人群允许跨视频帧出现, 且重点关注目标行人与人群中其他个体成员的时空关系, 具有更广泛的应用背景和更精确的时空特征刻画.

2 方法

2.1 方法动机

在一些特定的场景, 比如车站、校道、街道等, 行人在行走过程中的路线存在一致性. 在这样情况之下, 特定行人身边会形成相对稳定的人群, 本文称之为行人邻域. 通常情况下, 同一行人的邻域具有相对稳定的时空分布结构, 不同行人的邻域则存在着一定的差异性. 自然而然地, 相对稳定的行人邻域会对特定行人关联匹配起到一定的辅助作用. 基于此, 本文提出基于行人邻域的行人时空共现模式方法来辅助视觉行人再识别. 给定两个待匹配行人的图像, 其相似度取决于视觉表观特征与时空共现模式. 若两者之间的行人邻域共现模式越相似, 则两者越可能具有相同的身份. 图 1 是结合行人时空共现模式与视觉表观的行人匹配示意图. 实际上, Zheng 等[5] 已经意识到人群对行人识别的辅助作用, 并最早提出了人群视觉描述子以及匹配方法. 该方法把人群限制在同一视频帧内出现. 此外, 有研究人员提出将人群作为监督信息从而发展出弱监督学习的行人再识别方法[6−7]. 上述方法将人群定义为共同出现在同一视频帧内的行人. 然而在实际中, 人群是移动的, 同人群的不同个体未必会同时出现在同一个视频帧, 但是会出现在同一个时间窗口内的多帧照片, 因此本文提出用特定时间窗口内的行人集合来定义人群更具合理性. 下文介绍相关技术细节.

2.2 行人视觉表观特征提取

行人视觉特征提取可由任何一种行人表征方法实现. 本文采用一种基于行人全局特征的开源模型 SphereReID[34] 作为基准网络模型, 这是目前被最广泛应用的行人特征提取基准网络模型之一, 其中骨干网络使用残差网络 (ResNet50)[35]. 在训练时采用三元组损失函数约束, 并采用学习率启动策略[36]、随机擦除数据增强[37]、标签平滑[38]、移除最后一层下采样层[39]、对全局特征进行批归一化[40] 等策略. 相关实验细节将在第 3.2 节实验设置中详细说明.

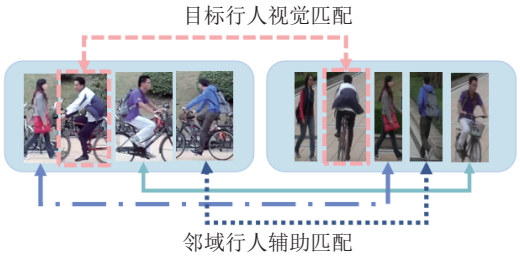


图 1 行人时空共现模式辅助视觉匹配示意图 (每个圆角矩形框代表一个摄像头视域, 虚线框指定目标行人, 其他行人表示目标行人在相应视域内的邻域)

Fig. 1 Illustration of spatiotemporal co-occurrence pattern aided pedestrian matching (Each rounded rectangle box represents a camera field. The dotted box specifies the target pedestrian, and other pedestrians indicate the target pedestrian′s neighborhood in the corresponding view field)

2.3 行人时空共现模式建模

行人时空共现模式是行人匹配中视觉表观特征的补充信息, 其需要确定目标行人的行人邻域集合, 并对该邻域进行匹配. 本节主要介绍行人时空共现模式的建模过程. 其中第 2.3.1 节中介绍行人邻域的确定方式, 第 2.3.2 节中介绍基于邻域行人匹配的详细过程.

算法 1. 行人邻域图像匹配流程

输入. 给定行人邻域 $Q=\{q_1,\cdots,q_n\}$ 和 $G=\{g_1,\cdots,g_m\}$, 相似度阈值 θ, 表观相似度度量函数 S_{app}, G 中与 q_i 的最相似行人的相似度 s_{i-max}

输出. 配图像对的相似度集合 $S=\{s_1,\cdots,s_k\}$

```
begin 初始化 S = {}
for q_i in Q do
    s_{i-max} ← 0
    for g_j in G do
        if S_app(q_i, g_j) > s_{i-max} then
            s_{i-max} = S_app(q_i, g_j)
        end if
    end for
    if s_{i-max} ≥ θ then
        S = S ∪ s_{i-max}
    end if
```

12) **end for**

13) 返回匹配图像对的相似度集合 S

14) **end begin**

2.3.1 行人邻域的确定

本文考虑的目标行人的邻域为一个指定时间窗口内 (以目标行人出现时间为时间窗口的中点), 从同一摄像头下经过的行人集合, 人群中的不同行人可以跨视频帧出现. 本文把一个人群中行人之间的时空联系看作一种行人共同出现的模式状态. 具体实验中, 使用时间戳信息确定目标行人的行人邻域. 通过预先设定时间差阈值, 检索行人库中与目标行人图像的时间戳差值小于设定阈值的所有行人图像, 并把返回的行人图像组成目标行人的邻域. 显然, 根据这种方式提取的行人邻域中很有可能包含目标行人的多帧图像. 为了专注于对邻域行人的分析, 需要删除邻域内与目标行人具有相同身份的行人图像, 但保留其他身份行人的多帧图像. 多帧图像可以提供同一个人的丰富视觉信息, 更加有利于跨摄像头下的人群时空关联分析. 具体地, 首先计算目标行人与邻域内所有行人图像的相似度. 其次, 剔除邻域内与目标图像的相似度分数大于预定阈值的行人图像.

实际上, 我们也可以考虑使用非极大值抑制方法对行人邻域中相同身份的行人进行图像去重, 即对邻域内每个身份行人 (包括目标行人) 只选取有代表性的一张图像进行保留. 本文后面提供的实验结果将表明使用非极大值抑制方法效果并不如前面介绍的处理方法. 因此, 本文在确定行人邻域上保留邻域行人中相同身份行人的多帧图像.

2.3.2 基于邻域的行人匹配

本节讨论如何基于邻域度量两个行人 (如 q 和 g) 之间的相似度. 用 $Q=\{q_1,\cdots,q_n\}$ 表示 q 的行人邻域; 用 $G=\{g_1,\cdots,g_m\}$ 表示 g 的行人邻域. 首先我们讨论如何计算邻域 Q 和 G 间的相似度. 一种自然的想法就是倘若两个行人邻域内拥有相同身份的行人越多, 则这两个行人邻域越相似.

行人邻域图像匹配伪代码在算法 1 中给出. 对于 Q 中的每一个 q_i 与 G 中的每一个 g_i , 首先经过表观特征提取网络获取每张行人图像的表观特征表示, 再利用相似度度量函数计算得到它们之间的表观特征相似度 $S_{app}(q_i,g_i)$. 对于每一个 q_i , 记录 G 中与 q_i 的最相似行人的相似度 s_{i-max} , 若该相似度大于给定的相似度阈值 θ , 则加入匹配图像对的相似度集合 S .

经过如上处理, 可以从 Q 和 G 之间比较返回具备相同身份的图像对. 值得注意的是, 这种匹配结果可能出现 Q 中多个 q_i 与 G 中相同的 g_i 形成匹配, 这主要是由于保留了邻域中同一个行人多帧图像所造成的. 但是上述问题并不会对邻域间的匹配造成困扰, 因为一对多的匹配本质上是相同身份行人的多次匹配. 假设这种配对的图像有 k 对, 记他们之间的相似度为 $S=\{s_1,\cdots,s_k\}$, S 的均值 $S_{enh}=\frac{1}{k}\sum_{i=1}^{k}s_i$, 则可以作为两个邻域 G 和 Q 之间的相似度度量. 对相似度集合 S 求均值的主要目的是为了平衡相同身份行人的多次匹配问题.

另记 q 和 g 的表观相似度为 $S_{app}(q,g)$, 则 q 和 g 的最终相似度 $S_{fin}(q,g)$ 由 S_{app} 和 S_{enh} 加权获得, 即

$$S_{fin}(q,g)=S_{app}(q,g)+\lambda\cdot S_{enh}(q,g) \tag{1}$$

其中 λ 为加权系数.

3 实验结果与分析

我们在行人再识别的权威数据集 Market-1501[8] 和 DukeMTMC-ReID[9] 上对所提方法的性能进行评估, 包括与其他主流方法的对比、消融实验以及模型参数敏感度分析.

3.1 实验数据集和评价指标

Market-1501[8] 数据集包含 1 501 个行人在 6 个摄像机下拍摄的 32 668 张行人图像. 其中, 训练集包含 751 个不同身份行人的 12 936 张图像; 测试集由查询行人库和模板行人库两部分组成, 包含 750 个不同行人共计 19 732 张图像. 对 750 个行人, 在每个摄像机下随机选择 1 张图像组成查询行人库. 一共有 3 368 张行人图像, 其余的则作为模板库. 每张行人图像由可变形部件模型 (Deformable parts model, DPM)[41] 检测得到行人矩形框.

DukeMTMC-ReID[9] 数据集由 8 个摄像机记录而成, 其包含出现在 2 个以上摄像机的 1 404 个不同行人, 以及仅仅在 1 个摄像机出现的 408 个行人 (干扰者) 共计 36 411 张图像. 训练集包含 702 个行人共计 16 522 张图像, 测试集由剩下 702 个行人组成. 查询行人库由在测试集中的每个行人在每个摄像机下选取 1 张图像组成, 共计 2 228 张查询图像; 测试集中余下的行人图像以及 408 个干扰行人的图像共同组成测试的模板行人库, 共 17 661 张行人图像. Market-1 501 和 DukeMTMC-ReID 数据集中的每张图像都包含了自身的身份信息、摄像机的 ID 和视频序列编号时间戳信息.

本节实验使用累积匹配曲线 (Cumulative

match characteristic, CMC) 和平均精度均值 (mean average precision, mAP) 对本文中涉及的行人再识别模型的性能进行量化评价. 其中 CMC 反映检索精度, mAP 反映召回率. 本文以 rank-1 的得分来代表 CMC 曲线, 其中 rank-1 是检索结果中首位候选的准确率. mAP 是所有查询平均精度的平均值, 其中每个查询的平均精度 (Average precision, AP) 是根据其精度召回曲线计算.

3.2 实验设置

本文实验基于被广泛使用的开放源码 OpenReID[1], 采用 SphereReID[34] 作为表观特征提取算法, 使用在 ImageNet[42] 上预训练的 ResNet50 模型作为基础网络, 并将全连接层的维度改为数据集中的行人身份总数. 在训练阶段, 每个批训练样本包含 64 张行人图像, 其中每个行人 4 张图像, 共 16 个行人. 每个行人图像统一裁剪为 256×128 的分辨率, 并以 0.5 概率水平翻转进行样本增广.

基准网络训练过程如下: 每张图像经过基准网络模型可得到分辨率为 16×8 的全局特征图; 在空间维度上, 对全局特征图进行平均池化可得到行人图像的特征向量表示. 根据特征向量计算三元组损失; 而后对特征向量进行批归一化处理再计算身份损失. 算法模型使用自适应矩估计 (Adaptive moment estimation, ADAM) 优化器进行优化, 一共进行 100 轮迭代优化. 优化器的初始学习率设置为 0.00035, 在第 40、70 轮迭代分别降低为原本学习率的 0.1 倍. 此外, 为了验证本文方法在不同表征能力的基准网络的泛化性能, 我们通过采用不同的训练策略来产生两种基准网络进行实验, 即通过采用行人再识别领域先进的训练策略来增强基准网络的表征能力. 这些训练策略包括随机擦除数据增强[37]、标签平滑[38]、移除最后一层下采样层[39]、对全局特征进行批归一化[40].

3.3 实验结果

3.3.1 与主流行人再识别算法的比较

本节将本文所提方法与当前主流的行人再识别方法进行性能实验比较. 参与对比的方法涵盖了手工特征和学习特征, 部分采用到行人姿态估计、行人掩模分割、注意力机制、生成对抗网络等最先进技术. 其中基于手工特征的算法模型有词袋模型[8] (Bags of words and keep-it-simple-and-straightforward metric, BoW+kissme)、核局部费希尔判别分类器[36] (Kernel local Fisher discriminant classifier, KLFDA)、Null space[43] 和加权近似秩分量分析[44] (Weighted approximate rank component analysis, WARCA); 基于姿态估计的算法包括全局局部对齐描述子[45] (Global-local-alignment descriptor, GLAD)、姿势不变嵌入向量[46] (Pose-invariant embedding, PIE) 和姿势敏感嵌入向量[47] (Pose-sensetive embedding, PSE); 基于掩模的算法有语义解析行人再识别[48] (Semantic parsing person re-identification, SPReID) 和基于掩膜的行人再识别[49] (MaskReID); 基于局部特征学习的算法包括 AlignedReID[50]、时空平行网络[51] (Spatial-channel parallelism network, SCPNet)、基于分部卷积基线模型[40] (Part-based convolutional baseline, PCB)、Pyramid[52] 和 Batch dropblock[53]; 基于注意力机制的算法有多任务注意力机制循环采样网络[54] (Multi-task attentional network with curriculum sampling, MANCS)、双注意力机制匹配网络[55] (Dual attention matching network, DuATM) 和和谐注意力机制网络[56] (Harmonious attention network, HA-CNN); 基于生成对抗网络 (Generative adversarial network, GAN) 的模型有 Camstyle[57] 和姿态标准化生成对抗网络[58] (Pose-normalized generative adversarial network, PN-GAN); 基于全局学习特征的算法包括多目标多摄像机追踪与再识别[59] (Multi-target multi-camera tracking and re-identification, MTMCReID), 矩阵分解网络[60] (SVDNet), 视角不变行人再识别[61] (Viewpoint invariant pedestrian recognition, IDE) 和对比注意力机制网络[1] (Comparative attention networks, CAN).

表 1 展示了不同算法在 Market-1501 和 DukeMTMC-ReID 数据集上的实验结果. 在数据集 Market-1501 中, 本文方法取得了 96.2 % 的 rank-1 准确率以及 89.2 % 的 mAP; 在数据集 DukeMTMC-ReID 中, 本文方法取得 89.2 % 的 rank-1 准确率及 80.1 % 的 mAP. 本文提出的方法比现有主流的行人再识别算法具有较大的性能提升, 表明行人时空共现模式的方法充分挖掘了行人的上下文特征, 有效提高了行人再识别的准确性. 此外, 本文方法只使用了全局特征而没有利用局部特征, 姿态估计和掩模等额外信息, 但本文方法的准确率却能超越上述方法, 表明行人时空共现模式方法是除了视觉表观特征以外强有力的辅助方法.

3.3.2 行人时空共现模式消融实验

为验证行人时空共现模式对行人视觉特征在行人再识别上的辅助作用, 我们采用两种基准网络进

[1] 下载地址: https://github.com/Cysu/open-reid

表 1 本文方法与主流算法在 Market-1501、DukeMTMC-ReID 数据集上实验结果比较 (%)
Table 1 Comparison with state-of-the-arts on Market-1501 and DukeMTMC-ReID data sets (%)

类型	算法	Market-1501		DukeMTMC-ReID	
		rank-1	mAP	rank-1	mAP
基于手工特征	BoW+kissme[8]	44.4	20.8	25.1	12.2
	KLFDA[36]	46.5	—	—	—
	Null space[43]	55.4	29.9	—	—
	WARCA[44]	45.2	—	—	—
基于姿态估计	GLAD[45]	89.9	73.9	—	—
	PIE[46]	87.7	69	79.8	62
	PSE[47]	78.7	56	—	—
基于掩模	SPReID[48]	92.5	81.3	84.4	71
	MaskReID[49]	90	75.3	78.8	61.9
基于局部特征	AlignedReID[50]	90.6	77.7	81.2	67.4
	SCPNet[51]	91.2	75.2	80.3	62.6
	PCB[40]	93.8	81.6	83.3	69.2
	Pyramid[52]	95.7	88.2	89	79
	Batch dropblock[53]	94.5	85	88.7	75.8
基于注意力机制	MANCS[54]	93.1	82.3	84.9	71.8
	DuATM[55]	91.4	76.6	81.2	62.3
	HA-CNN[56]	91.2	75.7	80.5	63.8
基于GAN	Camstyle[57]	88.1	68.7	75.3	53.5
	PN-GAN[58]	89.4	72.6	73.6	53.2
基于全局特征	IDE[61]	79.5	59.9	—	—
	SVDNet[60]	82.3	62.1	76.7	56.8
	CAN[1]	84.9	69.1	—	—
	MTMCReID[59]	89.5	75.7	79.8	63.4
	本文方法	**96.2**	**89.2**	**89.2**	**80.1**

表 2 用不同基准网络模型在数据集 Market-1501 和 DukeMTMC-ReID 上的消融实验 (%)
Table 2 Ablation experiment for proposed method on Market-1501 and DukeMTMC-ReID data set on different baseline network models (%)

算法模型	Market-1501		DukeMTMC-ReID	
	rank-1	mAP	rank-1	mAP
基准网络模型	86.7	71.7	76.4	60.9
基准网络模型+时空共现模式	91.3	76.1	79.4	64.2
基准网络模型(*)	94.4	85.4	86.6	75.5
基准网络模型(*)+时空共现模式方法	96.2	89.2	89.2	80.1

行了消融实验. 两种基准网络采用的是相同的基干网络, 但是采用的训练技巧不同. 表 2 给出了消融实验结果, 其中 “基准网络模型 (*)” 表示在训练网络的时候使用了近年来行人再识别中采用的先进训练策略, 包括随机擦除数据增强[37]、标签平滑[38]、移除最后一层下采样层[39]、对全局特征进行批归一化[40]. “基准网络模型” 则表示没有采用这些策略.

表 2 实验结果表明采用了行人时空共现模式进行行人匹配辅助之后, 与基准视觉特征模型相比, 所提方法在 Market-1501 的 rank-1 准确率上升了 4.6 % (从 86.7 % 升至 91.3 %), mAP 上升了 4.4 % (从 71.7 % 升至 76.1 %). 在 DukeMTMC-ReID 数据集上, rank-1 准确率上升了 3.0 % (从 76.4 % 升至 79.4 %), mAP 上升了 3.3 % (从 60.9 % 升至 64.2 %). 由此可见, 行人时空共现模式方法对增强行人再识别起到积极的作用. 采用了先进的训练策略后, 提

升照样很明显, 在 Market-1501 的 rank-1 准确率上升了 1.8 % (从 94.4 % 升至 96.2 %), mAP 上升了 3.8 % (从 85.4 % 升至 89.2 %). 在 DukeMTMC-ReID 数据集上, rank-1 准确率上升了 2.6 % (从 86.6 % 升至 89.2 %), mAP 上升了 4.6 % (从 75.5 % 升至 80.1 %). 综合表 2 的结果, 行人时空共现模式方法在表征能力强弱不等的基准网络中均能带来稳定的提升, 说明了本文方法具备良好的泛化性能.

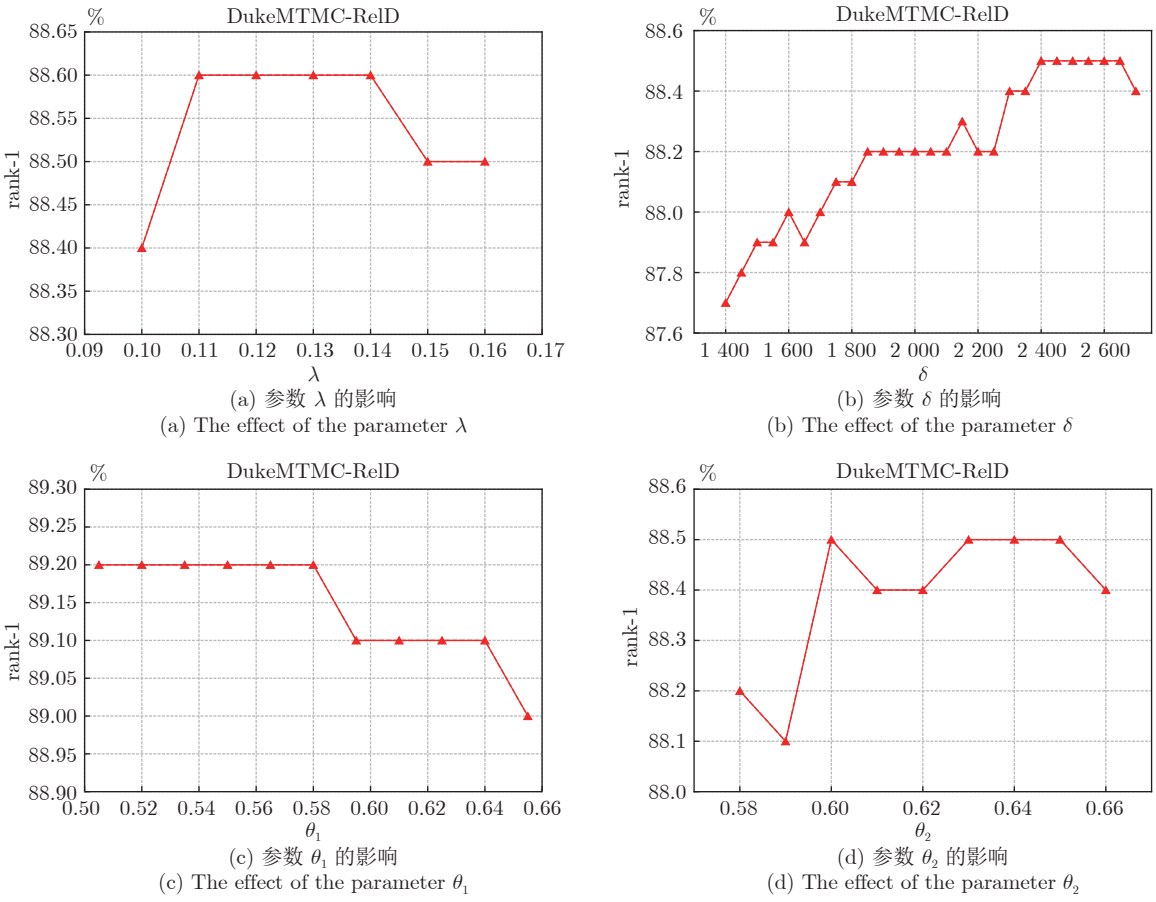
3.3.3 模型参数敏感性分析

为了探究行人时空共现模式方法中重要参数的影响, 本论文针对所提方法涉及的 4 个重要参数在数据集 DukeMTMC-ReID 上进行了敏感性分析实验. 其中, 行人邻域的时间差阈值参数 δ 的范围从 1400 帧到 2 700 帧; 行人邻域后处理的相似度阈值 θ_1 以及行人邻域匹配的相似度阈值 θ_2 的变化范围从 0.5 到 0.65; 衡量加强相似度分数重要程度的比例系数 λ 参数范围从 0.1 到 0.16. 实验结果由图 2 可知, 4 个参数都对模型的性能产生影响, 其中时间差阈值参数 δ 的影响最为显著. 由于时间差阈值 δ 会直接影响行人邻域的范围, 因此当时间差阈值 δ 取值过小, 行人邻域没有足够的上下文信息辅助目标行人的匹配; 随着时间差阈值 δ 增大至一定范围内, 行人时空共现模式方法的优势得以体现. 实验结果表明, 当参数在合理的范围内, 本文方法对于参数的选择不敏感. 本文提出的行人再识别算法模型的参数配置为 $\delta = 2\,500$, $\theta_1 = 0.55$, $\theta_2 = 0.6$, $\lambda = 0.13$.

3.3.4 行人邻域图像去重策略探究

在第 2.3.1 节, 我们讨论了对行人邻域进行后处理的方法, 其中主要有非极大值抑制方法以及仅仅剔除与目标行人具有相同身份的行人图像两种方法. 本节通过实验比较两种策略的优劣性, 实验结果如表 3 所示.

在 Market-1501 和 DukeMTMC-ReID 上的实验结果表明, 本文使用的方法在 rank-1 准确率、mAP 等各项指标上都超过非极大值抑制方法, 其中 DukeMTMC-ReID 上 rank-1 准确率提升了 1.3 % (从 87.9 % 到 89.2 %), mAP 提升了 1 % (从 79.1 % 到 80.1 %). 实验充分证明了, 保留邻域中同一行人



(a) 参数 λ 的影响
(a) The effect of the parameter λ

(b) 参数 δ 的影响
(b) The effect of the parameter δ

(c) 参数 θ_1 的影响
(c) The effect of the parameter θ_1

(d) 参数 θ_2 的影响
(d) The effect of the parameter θ_2

图 2 超参数对模型性能的影响, 纵坐标为 rank-1 准确率

Fig. 2 Influence of hyper-parameters on model performance (rank-1 accuracy)

表 3 不同行人邻域后处理策略在 Market-1501 和 DukeMTMC-ReID 数据集性能比较
Table 3 Comparison of different post-processing strategies for pedestrian neighborhood on Market-1501 and DukeMTMC-ReID datasets

后处理策略	mAP	rank-1	rank-5	rank-10
		Market-1501		
非极大值抑制	88.8	96.0	98.9	99.4
行人共现模式方法	89.2	96.2	99.1	99.5
		DukeMTMC-ReID		
非极大值抑制	79.1	87.9	95	96.9
行人共现模式方法	80.1	89.2	95.4	97.3

的多帧图像可以提供更丰富的视觉信息以用于行人匹配. 因此, 本文采用保留邻域中同一行人的多帧图像的处理方法.

4 结论

在某些公共场合, 行人在行走过程中偶尔会在一段持续时间内处于某个特定小群体, 这为行人匹配提供了一种特殊的上下文信息, 可以用于加强行人再识别. 基于此, 本文提出一种结合行人表观特征跟行人时空共现模式的行人再识别算法. 在行人再识别两个权威公开数据集 Market-1501 和 DukeMTMC-ReID 上的实验验证了所提算法的有效性. 未来可以继续将行人时空共现模式应用于行人再识别无监督或弱监督学习方法上.

References

1 Liu H, Feng J S, Qi M B, Jiang J G, Yan S C. End-to-end comparative attention networks for person re-identification. *IEEE Transactions on Image Processing*, 2017, **26**(7): 3492−3506

2 Chen G Y, Lu J W, Yang M, Zhou J. Spatial-temporal attention-aware learning for video-based person re-identification. *IEEE Transactions on Image Processing*, **28**(9): 4192−4205

3 Lv J M, Chen W H, Li Q, Yang C. Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 7948−7956

4 Cho Y J, Kim S A, Park J H, Lee K, Yoon K J. Joint person re-identification and camera network topology inference in multiple cameras. *Computer Vision and Image Understanding*, 2019, **180**: 34−46

5 Zheng W S, Gong A G, Xiang T. Associating groups of people. In: Proceedings of the 2009 British Machine Vision Conference. London, UK: BMVA, 2009. 23.1−23.11

6 Meng J K, Wu S, Zheng W S. Weakly supervised person re-identification. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 760−769

7 Wang G R, Wang G C, Zhang X J, Lai J H, Yu Z T, Lin L. Weakly supervised person Re-ID: Differentiable graphical learning and a new benchmark. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, **32**(5): 2142−2156

8 Zheng L, Shen L Y, Tian L, Wang S J, Wang J D, Tian Q. Scalable person re-identification: A benchmark. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 1116−1124

9 Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person Re-identification baseline in vitro. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 3774−3782

10 Yang Y, Yang J M, Yan J J, Liao S C, Yi D, Li S Z. Salient color names for person re-identification. In: Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014. 536−551

11 Fogel I, Sagi D. Gabor filters as texture discriminator. *Biological Cybernetics*, 1989, **61**(2): 103−113

12 Ojala T, Pietikainen M, Harwood I. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 1996, **29**(1): 51−59

13 Schmid C. Constructing models for content-based image retrieval. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, USA: IEEE, 2001. 39−45

14 Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, USA: IEEE, 2005. 886−893

15 Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, **60**(2): 91−110

16 Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 2008, **110**(3): 346−359

17 Lin Y T, Zheng L, Zheng Z D, Wu Y, Hu Z L, Yan C G, Yang Y. Improving person re-identification by attribute and identity learning. *Pattern Recognition*, 2019, **95**: 151−161

18 Li D W, Chen X T, Huang K Q. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios. In: Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition (ACPR). Kuala Lumpur, Malaysia: IEEE, 2015. 111−115

19 Deng Y B, Luo P, Loy C C, Tang X O. Pedestrian attribute recognition at far distance. In: Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, USA: ACM, 2014. 789−792

20 Chen H R, Wang Y W, Shi Y M, Yan K, Geng M Y, Tian Y H, et al. Deep transfer learning for person re-identification. In: Proceedings of the 4th IEEE Fourth International Conference on Multimedia Big Data (BigMM). Xi′an, China: IEEE, 2018. 1−5

21 Jing X Y, Zhu X K, Wu F, You X G, Liu Q L, Yue D, et al. Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 695−704

22 Ma F, Jing X Y, Zhu X, Tang Z M, Peng Z P. True-color and grayscale video person re-identification. *IEEE Transactions on Information Forensics and Security*, 2020, **15**: 115−129

23 Zhu X K, Jing X Y, Wu F, Feng H. Video-based person re-identification by simultaneously learning intra-video and inter-video distance metrics. In: Proceedings of the 25th International Joint

Conference on Artificial Intelligence. New York, USA: ACM, 2016. 3552−3558

24 Zhang W, He X Y, Yu X D, Lu W Z, Zha Z J, Tian Q. A multi-scale spatial-temporal attention model for person re-identification in videos. *IEEE Transactions on Image Processing*, 2020, **29**: 3365−3373

25 Wu Y M, El Farouk Bourahla O, Li X, W u, F, Tian Q, Zhou X. Adaptive graph representation learning for video person re-identification. *IEEE Transactions on Image Processing*, 2020, **29**: 8821−8830

26 Wang G C, Lai J H, Huang P G, Xie X H. Spatial-temporal person re-identification. In: Proceedings of the 2019 AAAI Conference on Artificial Intelligence. Hawaii, USA: AAAI, 2019. 8933−8940

27 Varior R R, Haloi M, Wang G. Gated siamese convolutional neural network architecture for human re-identification. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 791−808

28 Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 815−823

29 Cheng D, Gong Y H, Zhou S P, Wang J J, Zheng N N. Person re-identification by multi-channel parts-based CNN with improved triplet loss function. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 1335−1344

30 Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification [online] available: https://arxiv.org/abs/1703.07737, April 16, 2021

31 Zhu X K, Jing X Y, Zhang F, Zhang X Y, You X G, Cui X. Distance learning by mining hard and easy negative samples for person re-identification. *Pattern Recognition*, 2019, **95**: 211−222

32 Chen W H, Chen X T, Zhang J G, Huang K Q. Beyond triplet loss: A deep quadruplet network for person re-identification. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE, 2017. 1320−1329

33 Xiao Q Q, Luo H, Zhang C. Margin sample mining loss: A deep learning based method for person re-identification [online] available: https://arxiv.org/abs/1710.00478, April 16, 2021

34 Fan X, Jiang W, Luo H, Fei M J. SphereReID: Deep hypersphere manifold embedding for person re-identification. *Journal of Visual Communication and Image Representation*, 2019, **60**: 51−58

35 He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 770−778

36 Karanam S, Gou M R, Wu Z Y, Rates-Borras A, Camps O, Radke R J. A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, **41**(3): 523−536

37 Zhong Z, Zheng L, Kang G L, Li S Z, Yang Y. Random erasing data augmentation. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence, AAAI 2020, The 32nd Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The 10th AAAI Symposium on Educational Advances in Artificial Intelligence. New York, USA: AAAI, 2020. 13001−13008

38 Zheng Z D, Zheng L, Yang Y. A discriminatively learned CNN embedding for person reidentification. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2018, **14**(1): 13

39 Luo H, Jiang W, Gu Y Z, Liu F X, Liao X Y, Lai S Q, et al. A strong baseline and batch normalization neck for deep person re-identification. *IEEE Transactions on Multimedia*, 2020, **22**(10): 2597−2609

40 Sun Y F, Zheng L, Yang Y, Tian Q, Wang S J. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018. 501−518

41 Felzenszwalb P F, Girshick R B, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(9): 1627−1645

42 Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S A, et al. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 2015, **115**(3): 211−252

43 Zhang L, Xiang T, Gong S G. Learning a discriminative null space for person re-identification. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA: IEEE, 2016. 1239−1248

44 Jose C, Fleuret F. Scalable metric learning via weighted approximate rank component analysis. In: Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 875−890

45 Wei L H, Zhang S L, Yao H T, Gao W, Tian Q. GLAD: Global-local-alignment descriptor for pedestrian retrieval. In: Proceedings of the 25th ACM International Conference on Multimedia. Mountain View, USA: ACM, 2017. 420−428

46 Zheng L, Huang Y J, Lu H C, Yang Y. Pose-invariant embedding for deep person re-identification. *IEEE Transactions on Image Processing*, 2019, **28**(9): 4500−4509

47 Sarfraz M S, Schumann A, Eberle A, Stiefelhagen R. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 420−429

48 Kalayeh M M, Basaran E, Gokmen E, Kamasak M E, Shah M. Human semantic parsing for person re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 1062−1071

49 Qi L, Huo J, Wang L, Shi Y H, Gao Y. A mask based deep ranking neural network for person retrieval. In: Proceedings of the 2019 IEEE International Conference on Multimedia and Expo (ICME). Shanghai, China: IEEE, 2019. 496−501

50 Zhang X, Luo H, Fan X, Xiang W L, Sun Y X, Xiao Q Q, et al. AlignedReID: Surpassing human-level performance in person re-identification [online] available: https://arxiv.org/abs/1711.08184, April 16, 2021

51 Fan X, Luo H, Zhang X, He L X, Zhang C, Jiang W. SCPNet: Spatial-channel parallelism network for joint holistic and partial person re-identification. In: Proceedings of the 14th Asian Conference on Computer Vision. Perth, Australia: Springer, 2018. 19−34

52 Zheng F, Sun X, Jiang X Y, Guo X W, Yu Z Q, Huang F Y. Pyramidal person re-identification via multi-loss dynamic training. In: Proceedings of of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE, 2019. 8514−8522

53 Dai Z Z, Chen M Q, Gu X D, Zhu S Y, Tan P. Batch DropBlock network for person re-identification and beyond. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, 2019. 3690−3700

54 Wang C, Zhang Q, Huang C, Liu W Y, Wang X G. Mancs: A multi-task attentional network with curriculum sampling for person re-identification. In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018. 384−400

55 Si J L, Zhang H G, Li C G, Kuen J, Kong X F, Kot A C, et al. Dual attention matching network for context-aware feature sequence based person re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 5363−5372

56 Li W, Zhu X T, Gong S G. Harmonious attention network for person re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 2285−2294

57 Zhong Z, Zheng L, Zheng Z D, Li S Z, Yang Y. CamStyle: A

novel data augmentation method for person re-identification. *IEEE Transactions on Image Processing*, 2019, **28**(3): 1176−1190

58 Qian X L, Fu Y W, Xiang T, Wang W X, Qiu J, Wu Y, et al. Pose-normalized image generation for person re-identification. In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018. 661−678

59 Ristani E, Tomasi C. Features for multi-target multi-camera tracking and re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 6036−6046

60 Sun Y F, Zheng L, Deng W J, Wang S J. SVDNet for pedestrian retrieval. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, 2017. 3820−3828

61 Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Proceedings of the 10th European Conference on Computer Vision. Marseille, France: Springer, 2008. 262−275

钱锦浩 中山大学计算机学院硕士研究生. 主要研究方向为行人再识别.
E-mail: qianjh6@mail2.sysu.edu.cn
(**QIAN Jin-Hao** Master student at the School of Computer Science and Engineering, Sun Yat-sen University. His main research interest is person re-identification.)

宋展仁 中山大学计算机学院硕士研究生. 主要研究方向为行人再识别.
E-mail: songzr3@mail2.sysu.edu.cn
(**SONG Zhan-Ren** Master student at the School of Computer Science and Engineering, Sun Yat-sen University. His main research interest is person re-identification.)

郭春超 中山大学计算机学院博士研究生. 主要研究方向为行人再识, 光学字符识别, 广告内容素材理解.
E-mail: chunchaoguo@gmail.com
(**GUO Chun-Chao** Ph. D. candidate at the School of Computer Science and Engineering, Sun Yat-sen University. His research interest covers person re-identification, optical character recognition and advertising content material understanding.)

赖剑煌 中山大学教授. 主要研究方向为计算机视觉与模式识别.
E-mail: stsljh@mail.sysu.edu.cn
(**LAI Jian-Huang** Professor at Sun Yat-Sen University. His research interest covers computer vision and pattern recognition.)

谢晓华 中山大学计算机学院副教授. 主要研究方向为计算机视觉与模式识别. 本文通信作者.
E-mail: xiexiaoh6@mail.sysu.edu.cn
(**XIE Xiao-Hua** Associate professor at the School of Computer Science and Engineering, Sun Yat-sen University. His research interest covers computer vision and pattern recognition. Corresponding author of this paper.)