计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(4):268−273 [doi: 10.15888/j.cnki.csa.009035]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

姿态驱动的局部特征对齐的行人重识别①

王　琦, 刘志刚, 王　淼, 赵宜珺

(东北石油大学 计算机与信息技术学院, 大庆 163318)
通信作者: 刘志刚, E-mail: dqpiLzg@163.com

摘　要: 针对行人重识别研究中的遮挡问题, 本文提出了一种姿态驱动的局部特征对齐的行人重识别方法. 网络主要包括姿态编码器和行人部件对齐模块. 其中, 姿态编码器通过重构姿态估计热力图抑制遮挡区域骨骼关键点置信度, 引导网络提取行人可见部位的特征. 行人部件对齐模块依据姿态编码器输出的关键点置信图, 提取行人局部特征进行特征对齐, 降低非行人特征的干扰. 在遮挡、半身数据集上的仿真实验表明, 该方法获得了优于其他对比网络的结果.

关键词: 行人重识别; 姿态估计; 全局特征; 局部特征; 部件对齐; 图像检索; 特征提取

引用格式: 王琦,刘志刚,王淼,赵宜珺.姿态驱动的局部特征对齐的行人重识别.计算机系统应用,2023,32(4):268–273. http://www.c-s-a.org.cn/1003-3254/9035.html

Pose-driven Person Re-identification with Local Feature Alignment

WANG Qi, LIU Zhi-Gang, WANG Miao, ZHAO Yi-Jun

(School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China)

Abstract: To address the occlusion problem in person re-identification, this study presents a person re-identification method based on pose-driven local feature alignment. The network mainly consists of a pose encoder (PE) and a human part alignment module (HPAM). Specifically, the PE restrains the confidence of the key points on the bones in obscured areas by reconstructing the pose estimation heatmap to guide the network to extract the features of the person's visible parts. The HPAM extracts the person's local features according to the confidence map of the key points output by the PE for feature alignment, which further reduces the interference of non-person features. The simulation and experiments on occlusion datasets and half-body datasets show that the proposed method delivers better results than those produced by other networks under comparison.

Key words: person re-identification; pose estimation; global features; local features; part alignment; image retrieval; feature extraction

行人重识别 (person re-identification, Re-ID) 指跨摄像头, 跨场景在特定视频或者图像集中检索同一行人或者区别不同行人[1]. 随着智能视频监控日益凸显的应用需求, Re-ID 的研究已成为计算机视觉领域的热点问题[2−4]. 在监控场景中, 行人个体的相互遮挡、行人受背景的遮挡是普遍存在的现象, 增加了 Re-ID 的难度. 关键问题在于, 遮挡会在行人表征过程中引入过多的噪声, 降低特征的表示能力. 近年来, 尽管深度学习在 Re-ID 应用研究中的不断深入, 但同样面临该问题的挑战. 相较于一般 Re-ID, 遮挡 Re-ID 需要处理图像中由遮挡物造成的噪声信息. 因此直接对整张图像进行特征提取, 往往难以学习到鲁棒的识别特征.

目前针对遮挡 Re-ID 的多数研究方法都关注于遮挡区域的抑制. Sun 等人[5] 提出了一种可见部件感知模

① 基金项目: 黑龙江省自然科学基金 (LH2020F003); 黑龙江省高等教育教学改革项目 (SJGY20210109)
收稿时间: 2022-08-30; 修改时间: 2022-09-27; 采用时间: 2022-10-21; csa 在线出版时间: 2023-02-10
CNKI 网络首发时间: 2023-02-13

型 (visibility-aware part model, VPM), 先将人体划分为 3 个部分, 然后通过自监督学习方式确定人体部件可见或不可见, 网络只需关注不同遮挡图像中共同可见的部件特征. 但 VPM 的区域划分较为粗粒度, 不够细致. 而文献 [6–9] 通过局部区域定位的方式引导网络关注行人非遮挡区域, 起到了过滤遮挡噪声的作用, 从而降低遮挡对 Re-ID 的干扰. Zhao 等人[6] 使用多阶段特征融合的方式提取不同部位的特征, 再按像素层比较特征值, 保留最大值的特征, 达到去除遮挡噪声的目的. Su 等人[7] 根据人体姿态估计模型输出的 14 个关键点把行人分成 6 区域, 通过赋予行人每个特征区域一个不同的权重, 起到抑制遮挡噪声作用. Miao 等人[9] 提出了 (pose-guided feature alignment, PGFA) 方法, 通过预先定义关节点置信度的阈值, 来确定未遮挡的人体区域, 然后结合人体区域的全局特征和水平切块的局部特征, 增强网络处理遮挡问题的能力. Gao 等人[8] 提出的 (pose-guided visible part matching, PVPM) 网络则是通过生成伪标签的形式单独训练一个人体部件可见度网络, 引导网络关注未遮挡区域. 但这些方法存在过分依赖姿态估计模型的问题. 它们将行人特征的提取以及后续构建其他行人身份匹配的规则都完全建立在姿态估计模型输出的准确性上. 当姿态估计网络不准确时, 会导致 Re-ID 的性能出现严重下降, 增添了网络的不稳定性.

上述遮挡 Re-ID 方法都试图先定位遮挡物位置, 然后提取非遮挡部分的行人特征, 最后将可见行人部分特征与完整行人图像特征进行对齐与身份匹配. 于是, 针对 Re-ID 中的遮挡问题, 本文提出了一种姿态驱动的特征重构网络 (pose-driven feature alignment network, PDFA), 该网络包括姿态编码器 (pose encoder, PE)、行人部件对齐模块 (human parts alignment module, HPAM) 两个部分组成. 其中, 姿态修正编码器 PE 对人体姿态估计模型 (human pose estimation network, HPEN) 输出的关键点的置信度进行动态调整, 抑制区域遮挡导致的关键点错误; 最后, 在全身 (holistic) 数据集、半身 (partial) 数据集和遮挡 (occluded) 数据集进行了仿真实验. 实验结果表明所提方法行人重识别效果优于其他对比算法.

1 姿态驱动的行人重识别方法

本网络主要包含姿态编码器和姿态对齐两个模块, 如图 1. 姿态编码模块的作用是判别行人是否存在遮挡情况, 并通过骨骼关键点置信度, 引导特征提取网络关注可见部分的行人特征; 姿态对齐模块首先根据姿态修正模块的置信图, 对行人各个部位的特征进行精确定位, 然后提取不同区域的局部特征.

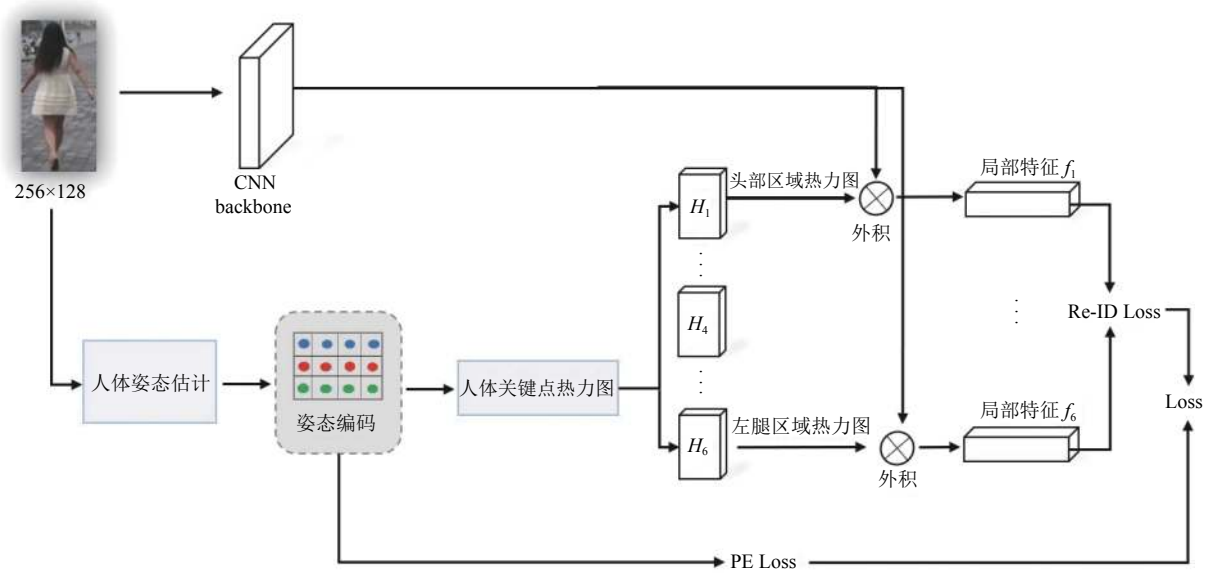

图 1 网络框架

1.1 姿态编码器

解决遮挡问题的关键在于提取行人特征, 减少遮挡区域特征的干扰. 通过观察经典的 HPEN (如 OpenPose[10] 等) 在遮挡与半身 Re-ID 数据集上的表现, 说明 HPEN

无法很好地解决遮挡状况下人体关键点的定位问题. 若在上述问题没有得到很好处置的情况下, Re-ID 网络完全依赖于 HPEN 的输出, 则必然会出现不理想的 Re-ID 结果. 为了使网络在遇到遮挡情况时引导特征提取网络关注行人特征, 本文提出了姿态编码模块, 包括编码器 (encoder) 与双边增益函数 (bilateral optimization function, BOF), 如图 2 所示.

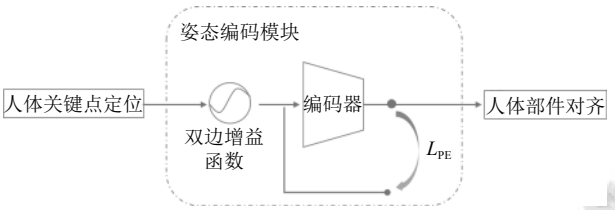

图 2 姿态编码器

由于 HPEN 不参加训练, 其网络参数是恒定的, 当遮挡造成关键点热力图错误时, Re-ID 结果也会出现偏差. 但 Re-ID 网络无法通过改变 HPEN 的参数来抑制噪声. 于是本文提出了编码器. 当样本存在遮挡区域时, 编码器可以通过梯度的反向传播精准的微调 (fine-tune) 编码器, 减少被遮挡人体部位的置信度, 使得 Re-ID 网络降低对遮挡区域特征的响应程度, 从而达到抑制非行人特征, 关注行人特征的目的.

具体说, 本文在 HPEN 后连接了一个 CNN 结构, 对 HPEN 输出的关键点热力图进行自适应调整. 首先, 输入样本通过 HPEN 输出关键点热力图Y_{HPEN}, 接着将Y_{HPEN}输入 BOF. BOF 根据式 (1) 将Y_{HPEN}中每个通道上的每个像素点优化后输出Y'_{HPEN}. 最后, PE 根据损失函数L_{PE}对Y'_{HPEN}的每个关键点通道进行置信值调整, 最后输出修正后的关键点热力图Y_{PE}, 如式 (2) 和式 (3):

$$y=\begin{cases}1, & y\geqslant 1\\ \mathrm{e}^{x^2}-1, & \text{others}\end{cases} \tag{1}$$

$$\delta^i_{\mathrm{HPEN}}=\max(y^i_{\mathrm{HPEN}}) \tag{2}$$

$$\begin{aligned}L_{\mathrm{PE}}&=MSE(Y'_{\mathrm{HPEN}},Y_{\mathrm{PE}})\\&=\frac{1}{n}\sum_{i=1}^{n}\delta^i_{\mathrm{HPEN}}(y'^i_{\mathrm{HPEN}}-y^i_{\mathrm{PE}})^2\end{aligned} \tag{3}$$

其中, L_{PE}代表 PE 模块的损失, i为 HPEN 输出的第i个通道. n=17 表示 HPEN 的输出通道数. δ^i_{HPEN}表示第i个通道的置信因子. $MSE(Y'_{\mathrm{HPEN}},Y_{\mathrm{PE}})$代表双边增益函数输出 Y'_{HPEN}与编码器输出Y_{PE}之间的均方误差损失; y^i_{HPEN} ($y^i_{\mathrm{HPEN}}\in Y_{\mathrm{HPEN}}$) 表示 HPEN 输出的第$i$个关键点置信图, 其中每个通道输出一个骨骼关键点, 该置信图经编码器修正后记为y^i_{PE} ($y^i_{\mathrm{PE}}\in Y_{\mathrm{PE}}$).

1.2 人体姿态部件对齐

遮挡 Re-ID 的本质是遮挡后的行人图像与原始行人图像之间的匹配问题. 由于遮挡导致行人特征不完整, 在进行特征比较时, 易出现特征错位, 进而增大了 Re-ID 的难度. 为此, 本文在 PE 后设计了姿态部件对齐模块, 获得更多细致的行人特征信息.

此模块对图像中的人体部位进行定位, 抑制遮挡噪声, 以获得有效的行人特征. 首先, 此模块根据 HPEN 输出的 17 个骨骼关键点热力图划分为 6 局部区域热力图 H_1–H_6, 如图 3 所示. 接着, 将局部区域热力图和人体全局特征图 进行外积操作, 然后将外积后获得的不同人体部位的局部特征图通过全连接层 FC 映射成 2 048 维向量f_i (一个H_i 产生一个f_i, 共 6 个), 最后分别对不同区域的局部特征向量进行交叉熵损失, 达到对齐的目的, 增强网络对行人特征的响应能力. 具体过程如式 (4).

$$\begin{cases}f_i=(y_{\mathrm{Backbone}})\otimes(y^i_{\mathrm{PE}}\cdot\delta^i_{\mathrm{HPEN}})\\ L_{\mathrm{cross}}=-\sum_{i=0}^{n}p(x_i)\log(q(x_i))\end{cases} \tag{4}$$

其中, y_{Backbone}代表特征提取网络输出的特征图; L_{cross}为交叉熵损失, n表示一个批次样本图片总数, $p(x_i)$表示真实的概率分布, $q(x_i)$表示预测的概率分布.

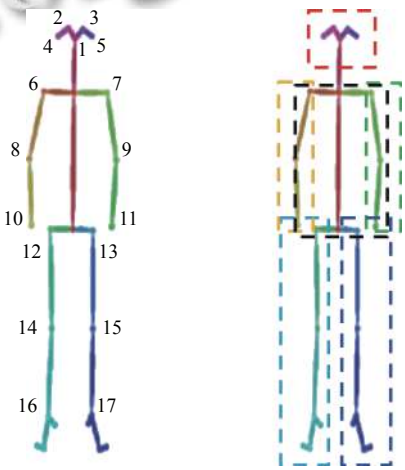

图 3 人体部件划分示意图

1.3 损失函数

为了使新提出的 Re-ID 模型拥有出色的行人特征

提取能力, 采用多任务学习的策略. 同时训练多个网络分支, 包括 2 个主要部分: 特征提取模块与姿态编码器模块. 于是总的损失函数为式 (5):

$$L = L_{\text{cross}} + L_{\text{PE}} \tag{5}$$

其中, L_{cross}为局部特征的交叉熵损失, L_{PE}为姿态编码器损失.

2 实验

2.1 实验环境

硬件环境: CPU Intel Xeon(R) E5-2640, 内存 8 GB, 显卡 Nvidia GTX2070 Super; 软件环境: 操作系统为 64 位 Ubuntu 16.04.7, 基于 Python 3.7.9 的深度学习框架 PyTorch 1.8.1 完成程序编程.

2.2 实验数据与评估指标

为了验证本文所提方法的有效性, 分别使用 7 种公开的数据集进行实验, 包括 3 种全身数据集 (Market-1501[3]、DukeMTMC-ReID[11]), 两种半身数据集 (Partial-ReID[12]、Partial-iLIDS[13]) 和两种遮挡数据集 (Occluded-Duke[9]、Occluded-ReID[14]). 本文使用了平均精确均值 (mean average precision, *mAP*) 和首位准确率 (Rank-1) 作为评估指标. Rank-1 代表第 1 张图像与目标图像是同一行人 ID 的准确度. *mAP* 如式 (6), i表示检索图像的序号, m表示与目标图像匹配图像的个数. $p(i)$表示序号为i图像在所有图像中的比例; 当$g(i)$=1 时表示i号图像与目标图像匹配, 否则$g(i)$=0; AP_i表示第i类的平均准确度, C表示类别的个数. 并且所有的实验表现都是在 single-shot 评价模式下获得的, single-shot 是指 gallery 中每个人的图像为一张.

$$\begin{cases} AP_i = \dfrac{1}{m}\displaystyle\sum_{i=1}^{n}(p(i)\cdot g(i)) \\ mAP = \displaystyle\sum_{i=1}^{n}\dfrac{AP_i}{C} \end{cases} \tag{6}$$

2.3 实现细节

本文主干网络采用 ResNet50[15] 作为特征提取网络, 并移除其 global average pooling 层与 classifier 层. 使用 BYOL[16] 算法分别在 Market1501 和 DukeMTMC-ReID 对特征提取网络进行无监督预训练. 本实验中的姿态估计模型是 HR-Net[17], 它在 COCO 数据集[18] 上被预训练. 训练过程中, 输入网络的行人样本尺寸为 256×128 (h×w), 模型训练迭代 400 个周期. 通过对比不同训练集批次的实验效果, 如图 4, 本文将 batch 设为 64. 初始化学习率被设置为 3.5×10^{-5}, 然后迭代到 50, 150 和 350 周期时分别衰减 0.1 倍, 选择 Adam 作为模型参数优化器.

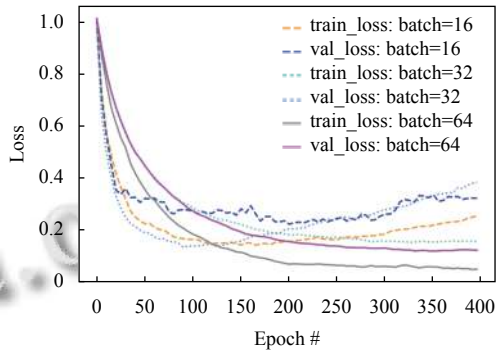

图 4 不同 batch 参数下 train_loss 和 val_loss 的实验结果

2.4 PE 的有效性

PE 主要有两个组件: 编码器与双边增益函数. 首先, 进行消融试验确认每个组件都发挥了作用, 结果如表 1. 实验表明只添加 BOF 时, 在单域条件下 *mAP* 和 Rank-1 最大分别增加了 1.7% 与 1.6%; M→D 的 Rank-1 提升了 3.9%, D→M 的 *mAP* 却下降了 2.6%, 说明 BOF 具有修正姿态的作用, 但存在矫正盲区. 此时完全依赖姿态信息会让模型变得脆弱, 网络完全有机会学习到非行人的“虚假”特征, 使模型缺乏泛化性, 阻碍算法提高 Re-ID 的能力; 当添加编码器时, M→M 与 D→D 的 *mAP* 与 Rank-1 分别得到 7% 和 10% 左右的大幅提升; 当编码器与 BOF 联合训练时, 模型无论是单域 (训练与测试样本同源) 还是跨域 (训练与测试样本非同源) 的性能都有极大提升, 且大于分别加入编码器与 BOF 的总和. 这说明 PE 有效增强了模型获取真正行人分类特征的能力, 验证了 PE 设计的意义性.

最后, 引入两个数据集: 半身 (partial) 数据集和遮挡 (occluded) 数据集. 为了更加全面地展现 PE 设计的合理性, 使用 3 组网络进行对比: 组别 1 (†), 同样使用骨骼关键点作为辅助的网络; 组别 2 (‡), 使用了切片对齐的网络; 组别 3 (※), 针对遮挡和部分行人设计的网络. 最终, 本文提出的方法在 Partial-ReID、Partial-iLIDS、Occluded-Duke 和 Occluded-ReID 数据集上分别取得了 Rank-1 77.2%、72.4%、56.7% 与 71.4% 的良好表现, 见表 2. 需要特别指出的是依据表 2 中组别

1 和 2 的结果证明: 在行人信息部分可见的错综场景下, 本模型取得良好 Re-ID 效果的主要原因不是使用了姿态估计或部件对齐方法. 更有力佐证了姿态修正方法的有效性.

表 1 PDFA 主要组件消融实验效果 (%)

Encoder	BOF	deta (δ)	M→M		D→D		M→D		D→M	
			Rank-1	*mAP*	Rank-1	*mAP*	Rank-1	*mAP*	Rank-1	*mAP*
×	×	—	86.1	67.9	76.1	55.6	36.6	18.3	47.5	21.2
×	√	—	87.7	68.5	76.4	57.3	40.7	22.2	46.6	18.6
√	×	×	92.1	77.4	83.7	67.0	41.0	22.4	51.6	22.8
√	√	×	93.2	78.5	84.9	68.7	42.1	25.3	52.0	22.2
√	√	√	93.6	80.7	85.7	70.3	45.4	26.2	53.9	25.8

表 2 半身与遮挡数据集准确率对比结果 (%)

方法	Partial-ReID		Partial-iLIDS		Occluded-Duke		Occluded-ReID	
	Rank-1	*mAP*	Rank-1	*mAP*	Rank-1	*mAP*	Rank-1	*mAP*
SpindleNet[6]†	67.0	—	66.3	—	—	—	—	—
Part-Aligned[19]†	55.9	24.6	55.2	63.7	36.3	20.0	25.1	18.6
PGFA[9]†	69.0	61.5	69.1	—	51.4	37.3	—	—
PVPM[8]†	75.3	71.4	—	—	47.0	37.7	70.4	61.2
SVDNet[20]‡	42.3	41.4	68.1	76.1	—	—	32.3	30.2
AlignedReID[21]‡	58.0	56.1	70.6	77.2	39.5	31.4	47.9	45.1
PCB[22]‡	56.3	50.5	68.9	73.0	49.6	36.8	46.7	40.3
AMC+SWM[12]※	36.0	—	49.6	—	—	—	31.2	27.3
VPM[5]※	67.7	—	65.5	—	—	—	—	—
STNReID[23]※	66.7	—	54.6	—	—	—	—	—
PDFA	77.2	71.6	72.4	77.6	56.7	45.0	71.4	63.6

3 结论

本文的目的是让网络提取更具鲁棒性的判别特征. 于是, 提出姿态驱动的局部特征对齐的行人重识别模型. 首先, 通过重构姿态估计模型输出的热力图来获得高阶的人体姿态信息起到定位行人可见区域的作用, 减弱噪声的干扰, 避免非行人特征给模型带来的低鲁棒性. 其次, 参照其他优秀算法在模型中使用局部特征对齐的方法, 进一步提升网络在遮挡情况下对行人特征的匹配能力. 最终, 在众多的数据集上进行的大量实验证明了本文提出的方法的有效性.

参考文献

1 Karanam S, Gou MR, Wu ZY, *et al*. 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. [doi: 10.1109/TPAMI.2018.2807450]

2 Ye M, Shen JB, Lin GJ, *et al*. Deep learning for person re-identification: A survey and outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 2872–2893. [doi: 10.1109/TPAMI.2021.3054775]

3 Zheng L, Shen LY, Tian L, *et al*. Scalable person re-identification: A benchmark. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 1116–1124.

4 Zheng ZD, Zheng L, Yang Y. Pedestrian alignment network for large-scale person re-identification. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(10): 3037–3045. [doi: 10.1109/TCSVT.2018.2873599]

5 Sun YF, Xu Q, Li YL, *et al*. Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 393–402.

6 Zhao HY, Tian MQ, Sun SY, *et al*. SpindleNet: Person re-identification with human body region guided feature decomposition and fusion. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 907–915.

7 Su C, Li JN, Zhang SL, *et al*. Pose-driven deep convolutional

model for person re-identification. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 3980–3989.

8 Gao S, Wang JY, Lu HC, *et al*. Pose-guided visible part matching for occluded person ReID. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 11741–11749.

9 Miao JX, Wu Y, Liu P, *et al*. Pose-guided feature alignment for occluded person re-identification. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 542–551.

10 Cao Z, Hidalgo G, Simon T, *et al*. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(1): 172–186. [doi: 10.1109/TPAMI.2019.2929257]

11 Zheng ZD, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 3774–3782.

12 Zheng WS, Li X, Xiang T, *et al*. Partial person re-identification. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 4678–4686.

13 He LX, Liang J, Li HQ, *et al*. Deep spatial feature reconstruction for partial person re-identification: Alignment-free approach. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 7073–7082.

14 Zhuo JX, Chen ZY, Lai JH, *et al*. Occluded person re-identification. Proceedings of the 2018 IEEE International Conference on Multimedia and Expo (ICME). San Diego: IEEE, 2018. 1–6.

15 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.

16 Grill JB, Strub F, Altché F, *et al*. Bootstrap your own latent a new approach to self-supervised learning. Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 21271–21284.

17 Sun K, Xiao B, Liu D, *et al*. Deep high-resolution representation learning for human pose estimation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 5686–5696.

18 Lin TY, Maire M, Belongie S, *et al*. Microsoft COCO: Common objects in context. Proceedings of the 13th European Conference on Computer Vision. Zurich: Springer, 2014. 740–755.

19 Zheng ZD, Yang XD, Yu ZD, *et al*. Joint discriminative and generative learning for person re-identification. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 2133–2142.

20 Sun YF, Zheng L, Deng WJ, *et al*. SVDNet for pedestrian retrieval. Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 3820–3828.

21 Zhang X, Luo H, Fan X, *et al*. AlignedReID: Surpassing human-level performance in person re-identification. arXiv:1711.08184, 2017.

22 Sun YF, Zheng L, Yang Y, *et al*. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 501–518.

23 Luo H, Fan X, Zhang C, *et al*. STNReID: Deep convolutional networks with pairwise spatial transformer networks for partial person re-identification. IEEE Transactions on Multimedia, 2020, 22(11): 2905–2913. [doi: 10.1109/TMM.2020.2965491]

(校对责编: 孙君艳)