计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(5):298−303 [doi: 10.15888/j.cnki.csa.008477]

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

基于全局特征改进的行人重识别①

张晓涵

(中国石油大学 (华东) 计算机科学与技术学院, 青岛 266580)
通信作者: 张晓涵, E-mail: s19070014@s.upc.edu.cn

摘　要: 由于行人重识别面临姿态变化、遮挡干扰、光照差异等挑战, 因此提取判别力强的行人特征至关重要. 本文提出一种在全局特征基础上进行改进的行人重识别方法, 首先, 设计多重感受野融合模块充分获取行人上下文信息, 提升全局特征辨别力; 其次, 采用 GeM 池化获取细粒度特征; 最后, 构建多分支网络, 融合网络不同深度的特征预测行人身份. 本文方法在 Market1501 和 DukeMTMC-ReID 两大数据集上的 mAP 指标分别达到 83.8% 和 74.9%. 实验结果表明, 本文方法有效改进了基于全局特征的模型, 提升了行人重识别的识别准确率.

关键词: 行人重识别; 全局特征; 感受野; GeM 池化; 特征融合; 深度学习

引用格式: 张晓涵.基于全局特征改进的行人重识别.计算机系统应用,2022,31(5):298–303. http://www.c-s-a.org.cn/1003-3254/8477.html

Improved Person Re-identification Based on Global Feature

ZHANG Xiao-Han

(College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China)

Abstract: Person re-identification faces challenges such as posture change, occlusion interference, and illumination difference, and thus it is very important to extract pedestrian features with strong discriminability. In this paper, an improved person re-identification method based on global features is proposed. Firstly, a multi-receptive field fusion module is designed to fully obtain pedestrian context information and improve the global feature discriminability. Secondly, generalized mean (GeM) pooling is used to obtain fine-grained features. Finally, a multi-branch network is constructed, and the features of different depths of the network are fused to predict the identity of pedestrians. The mAP indexes of this method on Market1501 and DukeMTMC-ReID are 83.8% and 74.9%, respectively. The experimental results show that the proposed method can effectively improve the model based on global features and raise the recognition accuracy of person re-identification.

Key words: person re-identification; global feature; receptive field; generalized mean (GeM) pooling; feature fusion; deep learning

行人重识别 (person re-identification) 也称行人再识别, 近年来引起学术界与工业界的广泛关注, 成为一个研究热点. 行人重识别旨在检索跨摄像头下的某一目标行人, 该技术可以与人脸识别、行人检测等相结合, 促进嫌犯追踪、走失救助等智慧安防领域以及无人超市等智慧商业领域的发展. 然而, 在真实的场景下, 不同摄像头的同一行人由于受到光照、姿态、遮挡、分辨率等各种因素的影响, 往往呈现很大的外观差异, 这给行人重识别的研究与应用带来诸多挑战[1]. 因此, 如何提取更具判别力的行人特征, 并采用高效的相似性度量方法以减小类内差距, 增大类间差距成为行人重识别的关键问题.

① 收稿时间: 2021-08-04; 修改时间: 2021-08-31; 采用时间: 2021-09-09; csa 在线出版时间: 2022-04-11

传统的行人重识别由特征提取与相似性度量两个子任务组成, 首先手工设计颜色、纹理、形状等低级特征对行人进行表示, 然后借助度量学习算法对特征距离进行约束, 无法实现端到端, 费时费力且精度普遍低下. 随着深度学习的兴起, 行人重识别将特征提取与相似性度量整合到一个统一的框架中. 通过卷积神经网络提取行人的高层特征, 同时设计度量损失函数控制类内类间距离, 大大提升了行人重识别的性能. 2016 年, Zheng 等[2] 提出 IDE 模型, 把行人重识别看做一个分类任务, 成为后续很多工作的基线模型. 近年, 为了提升行人重识别在数据集上的精度, 大量工作采用结合行人局部特征的方法. Zhao 等[3] 提出的 Spindle Net, 由姿态估计模型获得人体的若干关键点, 产生 7 个子区域, 然后分区域进行特征提取与融合. Kalayeh 等[4] 提出了 SPReID, 为了获取局部特征, 将人体分割模型引入行人重识别中, 得到前景和 4 个不同身体区域的掩膜. Sun 等[5] 提出了经典的 PCB 网络, 把行人图像水平均分为 6 块, 得到 6 个局部特征向量, 之后每个特征向量都经过降维和全连接层后送入分类器中, 进行单独预测. 基于此工作, Wang 等[6] 结合了全局特征设计了 MGN 网络, 通过将行人图像分别水平均分为 2 块和 3 块, 得到了不同粒度的行人局部特征. 这些方法虽然性能表现更好, 但无疑增加了网络的复杂程度. 基于姿态估计的方法和基于语义分割的方法需要引入额外的标注信息, 很大程度上依赖于预训练模型的性能. 基于水平切块的方法通常包含多个支路, 且粗暴的划分容易造成相应部件语义不对齐问题. 而基于全局特征的行人重识别网络结构简单, 近年来一直被忽视, 具有很大的研究意义. 相似性度量方面, 通过将行人特征映射到欧几里得空间, 最小化度量损失使得正样本间特征距离减小, 负样本间特征距离增大. 行人重识别中的度量损失主要有对比损失[7]、三元组损失[8]、四元组损失[9] 等, 其中使用最广泛的是三元组损失. 在三元组损失的基础上, Hermans 等[10] 提出了难样本采样三元组损失, 进一步提升模型的泛化性.

本文以 ResNet50 为骨干网络, 在特征提取层面做出 3 点改进, 最大程度上利用行人的全局特征: (1) 设计一种多重感受野融合模块, 采用不同大小的卷积核获取不同感受野的行人信息; (2) 采用 GeM 池化代替被普遍使用的全局平均池化获取细粒度特征; (3) 分别从 ResNet50 的 Conv4_x 和 Conv5_x 层进行采样, 得到两个通道数不同的特征图, 各送入一个分支, 两个分支得到的特征均使用分类损失与难样本采样三元组损失联合训练. 本文方法在 Market1501 数据集与 DukeMTMC-reID 数据集上验证, 实验结果表明, 本文方法具有较好的表现, 甚至优于一些基于局部特征的方法.

1 本文方法

此前基于全局特征的行人重识别方法整体流程大致如图 1 所示, 网络提取特征后, 采用全局平均池化 (GAP) 获得全局特征向量, 之后经过全连接层获得低维的输出特征. 这类方法结构简单但精度普遍较低, 在此基础上, 本文对基于行人全局特征的方法进行了改进. 本文整体的网络结构如图 2 所示, 使用在 ImageNet 上预训练过的 ResNet50 作为骨干网络, 该网络在行人重识别中被普遍使用. 移除网络最后的平均池化层和全连接层, 把最后一个卷积层的步长由 2 设为 1 以获取分辨率更高的特征图, 这就使得 Conv4_x 与 Conv5_x 采样的特征图具有相同的尺寸. 之后是两个独立的分支, 第一个分支是从 Conv5_x 得到的特征图, 经过本文设计的多重感受野融合模块, 之后进行 GeM 池化, 得到 2 048 维的特征向量, 使用难样本采样三元组损失约束. 该特征向量经过一个全连接层得到 512 维的输出特征, 使用交叉熵损失约束. 第二个分支是从 Conv4_x 得到通道数为 1 024 的特征图, 后续结构与第一分支保持相同. 在测试阶段, 将两个分支经过 GeM 池化得到的特征向量进行融合得到 3 072 维的向量对行人进行检索.

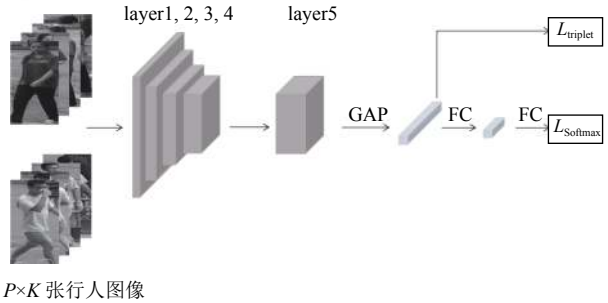

图 1 基于行人全局特征的网络结构

1.1 多重感受野融合模块

如图 1 所示, 基于全局特征的行人重识别方法在提取特征之后, 通常直接进行池化操作获取行人的全局特征向量. 但是, 在行人重识别中存在行人尺度变化

的问题, 而对于尺度不同的的目标来说, 不同大小的感受野将会产生不同的效果. 如果不考虑此问题, 直接将网络提取到的特征输入到后续的池化层中, 将在一定程度上降低行人重识别的识别精度. 为了适应行人的尺度变化, 本文在网络提取特征后先经过本文设计的多重感受野融合模块, 该模块通过在不同分支设置不同的感受野大小, 最后进行融合, 有效利用目标上下文信息. 空洞卷积[11] 最初源于语义分割任务, 不需要增加参数量便可实现扩大感受野的目的. 如图 3 所示, 多重感受野融合模块共包含 3 个分支, 对输入的特征图 X 分别进行卷积操作. 3 个分支均选取 3×3 卷积核, 但空洞率分别为 1、2、3, 得到特征图 F1、F2、F3. 为了更有效地利用来自不同分支的特征, 关注更重要的信息, 将 F1、F2、F3 分别经过一个通道注意力模块, 这里的通道注意力模块源于卷积注意力模块[12], 该模块结构如图 4 所示. 在通道注意力模块中, 输入的特征图同时采用最大池化与平均池化得到两个一维的向量, 之后被送进权重共享的多层感知机中, 将输出进行逐元素的相加后经过 Sigmoid 激活即可得到对应的注意力权重. 将 3 个分支得到的注意力权重系数分别于特征图 F1、F2、F3 相乘, 得到通道加权后的特征 F1′、F2′、F3′. 最后, 将 F1′、F2′、F3′进行融合, 即可得到最终的输出特征 X'. 多重感受野融合模块可以有效聚合不同感受野的特征, 使行人重识别性能进一步得到提升.

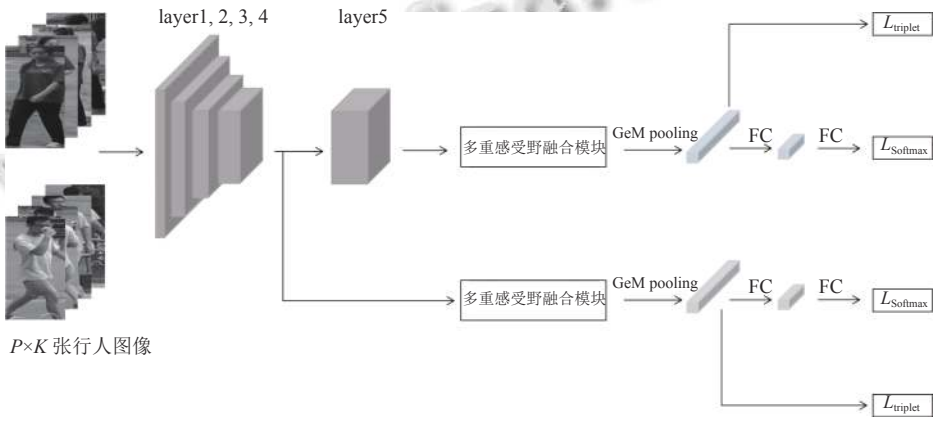

图 2 本文网络结构

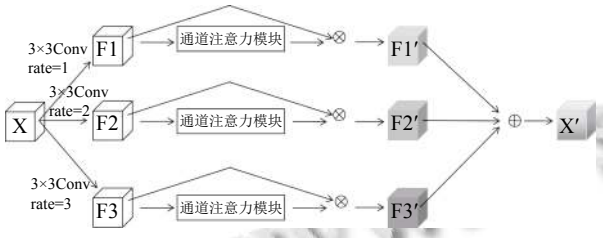

图 3 多重感受野融合模块

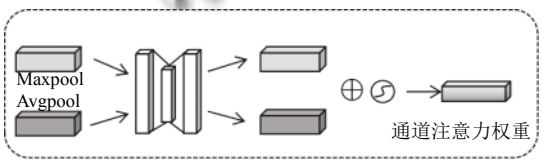

图 4 通道注意力模块

1.2 GeM 池化

在行人重识别网络中, 提取全局特征之后通常采用全局平均池化, 如图 1 所示. 全局平均池化关注的是图像整体的信息, 很容易受到背景和遮挡的干扰. 因此, 在本文中, 采用 generalized-mean (GeM) 池化[13], 该池化方法已在图像检索任务中取得了显著成果. 相比于传统的全局平均池化, GeM 池化包含了可学习的参数捕获细粒度信息. 公式如下:

$$f = [f_1 \cdots f_k \cdots f_K]^{\mathrm{T}}, f_k = \left(\frac{1}{|X_k|} \sum_{x \in X_k} x^{p_k} \right)^{\frac{1}{p_k}} \tag{1}$$

其中, X 为池化层的输入, f 为池化层的输出. p_k 是一个超参数, 在反向传播的过程中学习. 当 $p_k = 1$ 时, GeM 池化等价于全局平均池化; 当 $p_k \to \infty$ 时, GeM 池化等价于全局最大池化.

1.3 多层特征融合

目前, 大部分行人重识别工作仅利用网络最后一个卷积层提取到的高层语义特征图, 如图 1 所示. 高层语义特征能有效获取行人的显著信息, 但会忽视部分细节信息, 而这些细节信息对行人身份的判别同样有效. 因此, 为弥补缺失的细节信息, 本文同时采样低层

特征, 进行多层特征融合. 具体来说, 分别采样 ResNet50 的 Conv5_x 的高层显著特征与 Conv4_x 低层细节特征, 得到两个特征图, 通道数分别为 2 048 和 1 024. 如图 2 所示, 两个特征图形成两个独立的分支 G1 和 G2, 两分支采用相同的结构, 都经过多重感受野融合模块-GeM 池化-全连接层. 两分支都使用交叉熵损失与难样本采样三元组损失训练, 在测试时, 将两个分支的特征向量进行融合, 通过整合网络不同深度的信息进行预测. 多层特征融合的操作, 结合了行人的高层显著信息与低层细节信息, 使得特征更具判别力.

1.4 损失函数

本文联合交叉熵损失与难样本采样三元组损失预测行人的身份. 交叉熵损失被广泛应用于图像分类任务中, 通过最小化真实概率分布与预测概率分布之间的差异对网络进行优化. 在行人重识别中, 每张行人图像对应一个身份标签, 因此可以转化为一个分类问题.使用身份标签作为监督信号, 类别数即为训练集中行人身份数. 对于第 i 个学习到的特征 f_i, 交叉熵损失表示如下:

$$L_{\text{Softmax}} = -\sum_{i=1}^{N}\log\frac{e^{W_{yi}^{\mathrm{T}}f_i}}{\sum_{k=1}^{C}e^{W_k^{\mathrm{T}}f_i}} \tag{2}$$

其中, N 表示一个批次中的图像数量, W_k 表示对应类别的权重向量, C 表示训练集中行人的类别数. 在本文实验中, 把 G1 分支学习到的 512 维特征与 G2 分支学习到的 512 维特征计算交叉熵损失. 交叉熵损失函数使得行人图像接近所属的类别, 实现分类的效果.

此外, 行人重识别本质是一个图像检索问题, 这就需要引入一个度量学习函数使网络学习两张图像之间的相似度. 对于行人重识别来说, 就是使得相同行人不同图像之间比不同行人不同图像之间更为相似. 三元组损失经常被用于行人重识别任务中, 但由于行人图像存在大量外观相似的负样本对, 如果采样简单易区分的样本进行训练, 不利于网络学习到更有用的特征从而提升模型的可泛化性. 因此, 本文中对 G1 分支的 2 048 维特征与 G2 分支的 1 024 维特征使用难样本采样三元组损失, 公式如下:

$$L_{\text{triplet}} = -\sum_{i=1}^{P}\sum_{a=1}^{K}\left[\alpha+\max_{p=1,\cdots,K}\|f_a^{(i)}-f_p^{(i)}\|_2-\min_{\substack{n=1,\cdots,K\\ j=1,\cdots,P\\ j\neq i}}\|f_a^{(i)}-f_n^{(j)}\|_2\right]_+ \tag{3}$$

在一个训练批次中, 选取 P 个身份的行人, 每个行人采样 K 张图像. $f_a^{(i)}$、$f_p^{(i)}$、$f_n^{(j)}$分别为采样得到的锚图像、正样本图像、负样本图像的的特征, α是一个超参数. 在这里由困难正负样本与锚图像组成三元组, 即身份标签相同但距离最远的样本与身份标签不同但距离最近的样本.

2 实验过程

2.1 实验数据集

本文在行人重识别最常用的 Market-1501 数据集和 DukeMTMC-ReID 数据集上进行评估与分析. Market-1501 数据集采集自 5 个高分辨率摄像头和 1 个低分辨率摄像头, 共包含 1 501 个行人的 32 668 张图像. 其中训练集由 751 个行人的 12 936 张图像组成, 测试集由其他 750 个行人的 3 368 张 query 图像与 19 732 张 gallery 图像组成. DukeMTMC-ReID 数据集取自 DukeMTMC 数据集, 该数据集由 8 个高分辨率摄像头拍摄的 1 404 个行人的 36 411 张图像组成. 随机采样 702 个行人的 16 522 张图像组成训练集, 剩余 702 个行人的 2 228 张 query 图像和 17 661 张 gallery 图像组成测试集.

2.2 评估标准

本文实验使用行人重识别普遍使用的平均准确度 mAP (mean average precision) 和首位命中率 Rank-1 作为评估标准.

2.3 实验设置

本文实验基于 PyTorch 框架, 实验环境为 NVIDIA Tesla P100. 骨干网络采用在 ImageNet 上预训练的 ResNet50, 移除最后的全连接层与平均池化层, 并将最后一块的卷积步长设为 1. 实验所采用的行人图像尺寸为 256×128, 通过随机擦除、随机水平翻转、随机裁剪进行数据增强. 设置一个训练批次为 32, 一个批次选取的行人类别数 P=4, 一个行人采样的图片数 K=8. 超参数α设置为 0.3, p_k设置为 3. 采用 SGD 优化器进行梯度更新, weight decay 设为 5E–4, momentum 设为 0.9. 共训练 80 个 epoch, 初始学习率为 0.05, 训练 40 个 epoch 后学习率衰减至原来的 1/10.

3 实验分析

3.1 消融实验

为了分析本文各模块的有效性, 分别在 Market-1501 数据集和 DukeMTMC-ReID 数据集进行消融实

验, 具体结果如表 1 和表 2 所示. 其中, w/o MFF 表示去掉多重感受野融合模块, w/o GeM 表示去掉 GeM 池化采用全局平均池化, w/o MSF 表示去掉多尺度特征融合只保留 G1 分支的结果.

表 1 Market-1501 数据集上去掉各模块的结果 (%)

方法	Rank-1	mAP
本文	93.6	83.8
w/o MFF	93.3	82.6
w/o GeM	92.8	82.2
w/o MSF	93.0	81.6

表 2 DukeMTMC-ReID 数据集上去掉各模块的结果 (%)

方法	Rank-1	mAP
本文	85.8	74.9
w/o MFF	85.4	73.9
w/o GeM	84.6	72.1
w/o MSF	84.6	71.0

从实验结果可以看出, 本文各模块都起到了一定的作用. 去掉多重感受野融合模块, Market-1501 数据集和 DukeMTMC-ReID 数据集的 mAP 分别为 82.6% 和 73.9%, 添加多重感受野融合模块之后, 两个数据集上的 mAP 分别提升了 1.2% 和 1.0%. 这证明了该模块可以有效获取整合不同感受野的特征, 减轻行人图像尺度变化带来的负面影响. 为了证明本文中 GeM 池化更有优势, 将网络中的 GeM 池化替换为一般的全局平均池化. 结果表明, 采用 GeM 池化后, 在 Market-1501 数据集上和 DukeMTMC-ReID 数据集上的 mAP 分别提升了 1.6% 和 2.8%, Rank-1 分别提升了 0.8% 和 1.2%. 同时, 相比于去掉 G2 分支, 只保留 G1 分支训练和测试, 本文方法在 Market-1501 数据集上 mAP 提升了 2.2%, Rank-1 提升了 0.6%, 在 DukeMTMC-ReID 数据集上 mAP 提升了 3.9%, Rank-1 提升了 1.2%. 结果表明, 融合多层信息可以提升网络性能.

3.2 对比实验

表 3 和表 4 分别给出了本文方法在 Market-1501 数据集和 DukeMTMC-ReID 数据集上与其他行人重识别方法的对比, 包括传统方法 (LOMO+XQDA[14]、BoW+kissme[15]) 和基于深度学习的方法 (IDE[2]、SVDNet[16]、HA-CNN[17]、PCB+RPP[5]、Mancs[18]、IANet[19]). 从表中可以看出, 传统方法在数据集上的表现较差. 此外, 相比于近年来基于深度学习的方法, 本文方法均有显著提升. 具体来说, 本文方法在 Market-1501 数据集上的 mAP 指标达到 83.8%, 在引入重排序算法[20] 后, mAP 达到了 92.6%. 同时, 在 DukeMTMC-ReID 数据集上的 mAP 指标达到 74.9%, 通过重排序后, 精度进一步提升, 其中 mAP 达到了 88.6%. 本文方法具有竞争性, 虽然是基于简单的全局特征进行改进, 效果却优于部分基于局部特征的方法, 如 PCB+RPP[5].

表 3 Market-1501 数据集上与其他方法的对比 (%)

方法	Rank-1	mAP
LOMO+XQDA[14]	43.8	22.0
BoW+kissme[15]	43.2	22.0
IDE[2]	72.5	46.0
SVDNet[16]	82.3	62.1
HA-CNN[17]	91.2	75.7
PCB+RPP[5]	93.8	81.6
Mancs[18]	93.1	82.3
IANet[19]	94.4	83.1
本文	93.6	83.8
本文(+重排序)	94.4	92.6

表 4 DukeMTMC-ReID 数据集上与其他方法的对比 (%)

方法	Rank-1	mAP
LOMO+XQDA[14]	30.8	17.0
BoW+kissme[15]	25.1	12.2
IDE[2]	65.2	44.9
SVDNet[16]	76.7	56.8
HA-CNN[17]	80.5	63.8
PCB+RPP[5]	83.3	69.2
Mancs[18]	84.9	71.8
IANet[19]	87.1	73.4
本文	85.8	74.9
本文(+重排序)	90.9	88.6

将检索结果可视化, 如图 5 所示, 实线框的检索结果表示该图像匹配正确, 虚线框的检索结果表示该图像匹配错误. 从匹配结果来看, 本文方法取得了较高的检索准确率. 但一些外观十分相似的不同行人很难区分, 虽然难样本采样三元组损失一定程度上减轻了该问题, 但还是会造成少数样本误检测.

图 5 可视化结果

4 结论与展望

本文提出了一种基于全局特征的行人重识别方法. 为适用行人尺度变化, 设计多重感受野融合模块, 采用空洞率不同的卷积核进行卷积操作并使用注意力模块获取关键特征, 然后将各分支的特征进行融合. 此外, 将全局平均池化替换为更有效的 GeM 池化. 采样网络不同深度的特征送入不同分支, 测试时将多层特征融合用来预测. 在公开数据集上的实验表明, 本文方法具有较好的性能. 未来研究中将考虑行人重识别的实用性, 尤其是在无监督与跨模态方面提升检索准确率.

参考文献

1 罗浩, 姜伟, 范星, 等. 基于深度学习的行人重识别研究进展. 自动化学报, 2019, 45(11): 2032–2049.

2 Zheng L, Yang Y, Hauptmann AG. Person re-identification: Past, present and future. arXiv: 1610.02984, 2016.

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

4 Kalayeh MM, Basaran E, Gökmen M, *et al*. Human semantic parsing for person re-identification. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 1062−1071.

5 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.

6 Wang GS, Yuan YF, Chen X, *et al*. Learning discriminative features with multiple granularities for person re-identification. Proceedings of the 26th ACM International Conference on Multimedia. Seoul: ACM, 2018. 274–282.

7 Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping. Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2006. 1735–1742.

8 Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 815–823.

9 Chen WH, Chen XT, Zhang JG, *et al*. Beyond triplet loss: A deep quadruplet network for person re-identification. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 1320–1329.

10 Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification. arXiv: 1703.07737, 2017.

11 Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv: 1511.07122, 2015.

12 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.

13 Radenović F, Tolias G, Chum O. Fine-tuning CNN image retrieval with no human annotation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(7): 1655–1668. [doi: 10.1109/TPAMI.2018.2846566]

14 Liao SC, Hu Y, Zhu XY, *et al*. Person re-identification by local maximal occurrence representation and metric learning. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 2197–2206.

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

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

17 Li W, Zhu XT, Gong SG. Harmonious attention network for person re-identification. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 2285–2294.

18 Wang C, Zhang Q, Huang C, *et al*. Mancs: A multi-task attentional network with curriculum sampling for person re-identification. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 384–400.

19 Hou RB, Ma BP, Chang H, *et al*. Interaction-and-aggregation network for person re-identification. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 9309–9318.

20 Zhong Z, Zheng L, Cao DL, *et al*. Re-ranking person re-identification with k-reciprocal encoding. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 3652–3661.