

基于 HPLF 的行人再识别^①



杨 戈^{1,2}, 叶杰强¹

¹(北京师范大学珠海分校 智能多媒体技术重点实验室, 珠海 519087)

²(北京大学深圳研究生院 深圳物联网智能感知技术工程实验室, 深圳 518055)

通讯作者: 杨 戈, E-mail: yangge@pkusz.edu.cn

摘 要: 为了更好的挖掘局部特征, 提升行人再识别的精度, 本文提出了一种利用水平池化提取局部特征的 HPLF (Horizontal Pooling for Local Feature) 算法, 在 ResNet-50 网络中对输入的联合数据集进行预处理, 提取特征, 对 ResNet-50 网络生成的特征图进行水平切割, 通过分割的特征图计算两两特征之间的距离, 再用难样本三元组损失 (Triplet loss with Hard example mining, TriHard loss) 来作为局部特征损失函数训练, 通过特征图计算全局距离, 通过难样本三元组损失来训练, 将这两个损失函数加上一个 Softmax 交叉熵损失函数, 联合起来作为总的损失函数进行参数修正. 实验结果表明: 在 Market1501 数据集中, mAP (mean Average Precision), Rank-1, Rank-5, Rank-10 等性能指标上, HPLF 算法比其他算法有 3% 左右的提升.

关键词: 深度学习; 计算机视觉; 行人再识别; 卷积神经网络; 生成对抗网络

引用格式: 杨戈, 叶杰强. 基于 HPLF 的行人再识别. 计算机系统应用, 2021, 30(3): 227-233. <http://www.c-s-a.org.cn/1003-3254/7841.html>

Pedestrian Re-Identification Based on HPLF

YANG Ge^{1,2}, YE Jie-Qiang¹

¹(Key Laboratory of Intelligent Multimedia Technology, Beijing Normal University, Zhuhai, Zhuhai 519087, China)

²(Engineering Lab on Intelligent Perception for Internet of Things (ELIP), Shenzhen Graduate School, Peking University, Shenzhen 518055, China)

Abstract: Pedestrian re-identification generally considered as a sub-problem of image retrieval. Due to the distance between the camera and the pedestrian, the definition of the pedestrian photo is generally fuzzy, and the camera's view angle of pedestrians is fixed, so it is not enough to recognize pedestrians by faces. In order to better mine strong local features and improve the accuracy of pedestrian re-identification, this study proposes an algorithm, namely Horizontal Pooling for Local Feature (HPLF). We preprocess the input joint data set in ResNet-50 network, extract features, and horizontally cut the feature map generated by ResNet-50 network, with which we calculate the distance between every two features. Triple loss with hard example mining (TriHard loss) is used for training as a local feature loss function. The global distance is calculated according to the feature map and trained through TriHard loss. The two loss functions plus a Softmax cross entropy loss function are combined as the total loss function for parameter correction. The experimental results show that HPLF's performances of mean Average Precision (mAP), Rank-1, Rank-5, and Rank-10 in the Market1501 data set are about 3% higher than those of other algorithms.

Key words: deep learning; computer vision; pedestrian re-identification; Convolutional Neural Network (CNN); Generative Adversarial Network (GAN)

① 基金项目: 国家自然科学基金 (61272364); 广东高校省级重大科研项目 (2018KTSCX288, 2019KZDXM015, 2020ZDZX3058); 广东省学科建设专项资金 (2013WYXM0122); 智能多媒体技术重点实验室 (201762005)

Foundation item: National Natural Science Foundation of China (61272364); Provincial Major Science and Technology Research Program of Higher Education of Guangdong Province (2018KTSCX288, 2019KZDXM015, 2020ZDZX3058); Special Fund for Discipline Construction of Guangdong Province (2013WYXM0122); Key Laboratory of Intelligent Multimedia Technology (201762005)

收稿时间: 2020-07-22; 修改时间: 2020-08-13, 2020-08-19, 2020-08-25; 采用时间: 2020-09-01; csa 在线出版时间: 2021-03-03

1 引言

行人再识别 (person Re-Identification, ReID) 是指在不同背景的非重叠摄像头视域下, 确认在不同场景和不同时间出现的行人是否为同一个人, 在视频监控中, 当出现一个目标行人时, 行人再识别会判断这个人是否在另一个地方 (时间) 被另一个摄像机观察过. 作为一个新兴的方向, 也是计算机视觉领域近年来研究热点之一, 近几年越来越受到关注. 这一方向的出现可以归结为: (1) 公众安全需求的增加; (2) 遍布于公园、大学校园和街道等地区的摄像机网络. 这两种原因都使得仅仅依靠蛮力的人为标注来准确和有效地识别或者追踪目标行人是非常昂贵的. 从计算机视觉的角度来看, 行人再识别中最具挑战性的问题是如何在密集的外观变化下, 如照明、姿态和视点等, 正确匹配同一人的两幅图像, 其具有重要的科学价值和应用前景.

近几年, 深度学习 (Deep Learning, DL) 的急速发展, 卷积神经网络 (Convolutional Neural Networks, CNN) 成功进入人们的生产生活中, 使用深度学习与卷积神经网络的解决行人再识别的问题是最有效的一种方式之一^[1]. 按照训练损失将卷积神经网络分为度量学习与表征学习, 按照是否具有局部特征分为局部特征与全局特征, 根据数据集的不同可以分为基于单帧和基于序列. 一种生成对抗网络 (Generative Adversarial Nets, GAN) 的方法利用生成新数据来解决行人再识别的难点问题.

1.1 基于度量学习的 ReID 方法

ReID 作为一个图像检索 (Text-Based Image Retrieval, TBIR) 的子问题, 目的是找出两张照片的相似度. 度量学习 (metric learning) 是普遍用来解决 TBIR 的方法之一. 也就是说: 如果两张照片的是同一行人 ID, 那么它们的相似度就大于别的照片. 于是提出映射概念, 由原始域映射图片至特征域, 定义距离度量函数对两个特征向量间距离计算, 再根据网络最小化度量损失获取最优映射 $f(x)$, 让 ID 相同的行人正样本通过损失函数使其具有较短的余弦距离或者欧式距离, 不同 ID 的行人负样本通过损失函数使其具有更长的余弦距离或者欧式距离, 该 $f(x)$ 映射为经过训练获得的深度卷积网络.

(1) 对比损失 (contrastive loss)

通常对比损失用来训练孪生网络 (Siamese Network)^[2-4], 该网络的输入一般为一对 (两张) 照片, 例如 a 和 b , 这对照片可以是正样本对 (同一行人), 也可以

为负样本对 (不同行人).

(2) 三元组损失 (triplet loss)

三元组损失是近几年被大量运用于度量学习中^[5-8], 很多度量学习的损失函数都是基于三元组损失的形式, 该损失的特点是: 需要每次输入 3 张图片做为训练. 包括: 锚点图片, 正样品图片和负样本图片.

锚点图片 a 和正样品图片 p 是来自同一行人 ID, 锚点图片 A 和图片 N 来自不同行人 ID.

(3) 四元组损失 (quadruplet loss)

四元组损失在三元组损失的基础上一个改进^[9]. 四元组损失需要 4 张照片, 包括: 锚点图片, 正样品图片及两张负样本图片.

文献 [9] 作者认为三元组损失考虑了正样本和负样本之间的相对距离, 但没有考虑正负样本之间的绝对距离. 为此提出了四元组损失. 在神经网络中采用四元组损失可以学习到更明确的特征信息.

1.2 基于特征学习的 ReID 方法

行人再识别另一个基于损失函数的方法就是表征学习 (representation learning), 表征学习极其常见于卷积神经网络. 行人再识别的目的是获取两张不同图间的相似程度, 然而表征学习方法并未在训练网络直接应用时将图片间的相似度考虑在内, 而是看作分类问题或者验证问题来看待. 其中分类问题是指将相同的行人 ID 聚集为一类^[10], 使用 ID 或者属性分类, 而验证问题是在行人重识别中是一个二分问题, 直接让网络判断输入的两张行人照片是否为一个 ID.

分类方法的特点: CNN 网络中提取特征, 提取到的特征经过池化层, 经过一个 FC 层. 该 FC 的维度就是网络中的 ID 数, 将提取到的特征运用分类损失 (identification loss) 函数进行网络训练. 因为训练时候的 FC 层是训练集的维度, 测试时, 训练集和测试集的 ID 一般不相同, 也就是说训练集的全部 FC 层在训练完后全部要丢弃. 分类网络中经常采用一个属性损失 (attribute loss)^[11], 用一些行人属性来辅助行人特征, 能增强网络的性能. 分类损失在 ReID 问题中也被称为 ID 损失, 所以该网络也被称作 IDE 网络. 行人再识别中最主要的一项基准 (baseline) 为 IDE 网络.

这类方法就是将 ReID 问题看为是一类二分问题, 旨在让网络直接判别输入网络的行人 ID 两者是否是同一行人^[11]. 每次输入一对照片, 进入孪生网络进行特征提取. 融合孪生网络中每个网络的特征信息计算一

个二分类损失(验证损失),训练时可以和ID损失一起使用.测试时输入两张图片,网络直接判断两张照片是否属于同一行人.

1.3 基于局部特征与全局特征的 ReID 方法

全局特征是指每一张行人图片的全局信息进行一个特征抽取,这个全局特征没有任何的空间信息.一般的卷积网络提取的都是全局特征,然而行人数据集复杂性逐步上升,采用全局特征难以达到当前需求.目前研究的热点问题为提取复杂度较高的特征,局部特征为确定某个网络或者网络自动对设计者注重的局部区域关注.常用的方法有切块,关键点定位等方法.

图片切块是提取局部特征方法主要为水平切块.文献[12]是这方面的内容.每一块图像经过CNN网络得到特征,局部特征按顺序输入到LSTM(Long Short-Term Memory)网络,自动表达为图像最终的特征.利用对比损失训练网络水平切块是比较早期的工作,目前已经很少使用.改进的水平切块算法PCB^[12]和Aligned ReID^[13]是现在的主流方法.将图像进行水平方向的等分,每一个水平切块通过水平池化提取一个特征, Gate Siamese 和 Aligned ReID 通过设计规则融合所有的局部特征计算距离就, PCB、ICNN、SCPNet 对每一个局部特征计算一个 ReID 损失, 直接将局部特征拼接起来, 联合局部特征和全局特征往往能够得到更好的结果.

图像语义分割是一种极精细的像素级别 part 信息, 图像分割分为粗粒度的行人前景分割和细粒度的肢体语义分割, 分割结果通常作为图像预处理的 Mask 或者 feature map 中的 attention 相乘, 目前基于分割的方法没有取得特别广泛的应用. 行人的局部特征被越来越多的论文来验证是极其有效的, 一定程度上解决了行人姿势不同的问题. 融合全局特征和局部特征的方法在 Person-ReID 领域越来越流行. 如何很效率和低消耗的提取局部特征是现在的研究热点方向之一.

序列重识别也常被称为视频重识别(Video ReID), 是指利用一段连续的行人图片序列进行行人再识别任务, 或者说是个序列搜索序列的问题.

有以下特点: (1) 姿态变化丰富. (2) 遮挡现象普遍. (3) 总有几帧质量好, 也有几帧质量差. (4) 需要考虑如何融合各帧的信息.

1.4 基于 GAN 的 ReID 方法

生成对抗网络(Generative Adversarial Nets, GAN)

在最近几年呈现快速发展状态^[14], 最主要的应用为生成图片. 深度学习方法在实际应用的前提条件为训练数据量庞大, 然而当前行人再识别数据集数量、规模上相对较少, 采用 GAN 进行识别行人任务发展成为热门, 传统 GAN 网络是以随机方式生成图片, 再后期推出的 DualGAN、CycleGAN 及 DiscoGAN 等可以转换图片风格^[15], 推动在行人再识别领域广泛运用 GAN. 网络中有生成器(Generator, G)和鉴别器(Discriminator, D). 有两个数据域分别为 X, Y.

生成器 G 用来生成数据, 生成的数据被鉴别器 D 识别是否为真实数据, 两者存在一种矛盾关系. 生成器 G 使用 X 域中的数据生成的图片越来越接近真实数据. 鉴别器 D 的鉴别能力越来越强, 越来越能够把真实与伪造图片分开来. 最后生成器 G 生成的照片已经接近真实图片, 鉴别器 D 分不清真实图片与伪造图片. 到这时, 生成器 G 和鉴别器 D 达到了动态平衡的状态, 我们就可以使用改 GAN 来生成图片, 用来扩充数据集.

训练完成后, 一般不使用鉴别器, 而只使用生成器来生成我们想要的图片, CycleGAN 一个非常实用的地方就是输入的两张图片可以是任意的两张图片, 也就是非成对, 而且 CycleGAN 可以使两个域的图像风格进行转换.

目前传统行人再识别方法缺少综合考虑多帧信息以及多特征融合思想, 因此本文综合考虑全局特征和局部特征, 提出了一种利用水平池化提取局部特征的 HPLF(Horizontal Pooling for Local Feature)算法, 在 ResNet-50 网络中对输入的联合数据集进行预处理, 提取特征, 对特征图进行水平切割, 计算两两特征之间的距离, 再用难样本三元组损失来作为局部特征损失函数训练, 通过特征图计算全局距离, 提高了行人再识别精度.

2 行人再识别网络的设计

2.1 网络结构

本文的方法是: 将数据集的数据输入到 ResNet-50 网络中进行预处理, 将提取到的特征进入一个池化层进行池化操作, 得到特征图. 得到的特征图分别进入一个局部网络分支, 改网络将特征图进行水平池化操作, 将得到的特征图进行切割, 将每个特征图划分为高为 1 的特征图, 之后用距离函数计算它们之间的距离. 在另一个分支中, 进行一个传统的度量学习操作, 计算完距离后, 两个分支分别进行一个难样本三元组的构造工作,

然后用三元组损失进行网络训练,流程图如图1所示.

2.2 HPLF 方法

HPLF (Horizontal Pooling for Local Feature) 方法主要涵盖一个水平池化跟难样本挖掘的方法.

水平池化层: 在经过一个 ResNet-50 的输出后会得

到一个 $(N \times C \times H \times W)$ 的向量, 其中 N 是 batchsize 的大小, C 是照片的像素大小, H 为照片的高度, W 为照片的宽度. HPLF 方法用一个 $1 \times W$ 的窗口去剪裁该特征图, 最后会得到 H 个 $N \times C \times 1 \times W$ 的局部特征图. 用这个局部特征图去训练网络, 如图2.

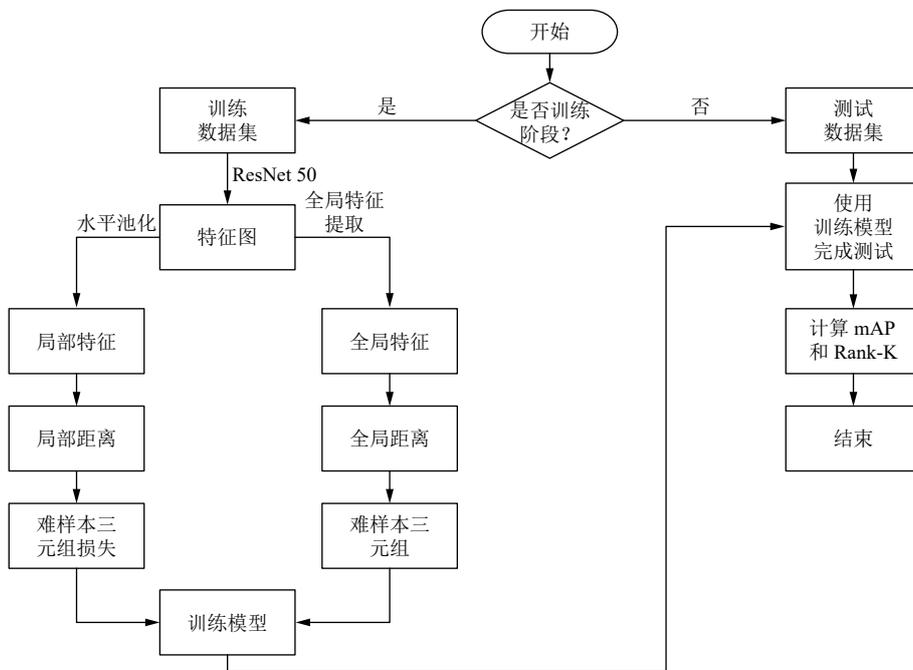


图1 流程图

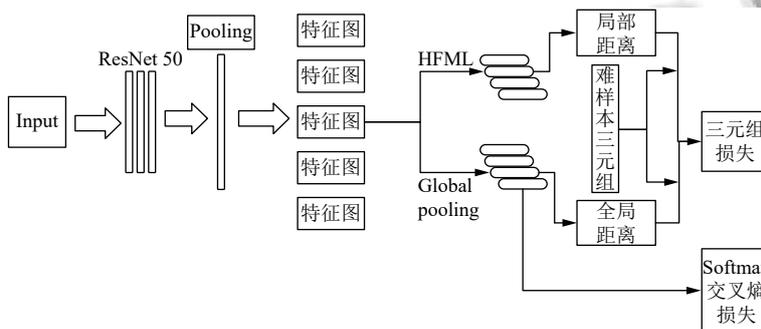


图2 网络结构设计

难样本挖掘: 传统的度量学习方法是随机采样组成元祖样本, 但是这样采集的样本通常是一些非常容易识别的样本, 不利于训练表达能力强的 ReID 网络. 为了使网络更好的识别行人, 利用难样本训练网络, 让网络注意到更多的细节特征, 利用难本来训练网络的方法是很有效的方法. 本文训练时, 会在一个 batchsize(size) 中找到一个最难的样本与一个最难的

负本来作为一个三元组来训练网络.

2.3 损失函数

本文构建的网络中应用了2种损失: Softmax 交叉熵损失 (Softmax Cross Entropy Loss) 函数和难样本三元组损失 (TriHard loss) 函数.

(1) Softmax 交叉熵损失: 在神经网络中, 最后一个输出层的个数与分类任务的标注数相等. 假设最后的

输出层个数为 N , 那么对于神经网络可以有一个 N 维度的数组作为他的输出结果. 在理想的状态下, 假设某一个样本属于 M , 那么这个该样本对于 M 的输出值应该为 1, 对于其他的输出结果应该为 0. 例如这种情况 $[0, 1, 0, 0, \dots, 0, 0]$, 这是最期望的结果, 但一般实际输出与这不相符. 例如 $[0.02, 0.01, 0.7, 0.05, \dots, 0.01]$, 这就是 Softmax 所做的, 将样本的输出值, 做了加权和非线性处理之后, 使其变成了一个概率值, 概率和为 1.

假设 $y_1, y_2, y_3, \dots, y_n$ 是某神经网络的输出, 那么经过 Softmax 处理之后, 如式 (1) 所示:

$$y' = \text{softmax}(y_i) = \frac{e^{y_i}}{\sum_{j=1}^n e^{y_j}} \quad (1)$$

其中, y_i 网络的第 i 个输出, e 是输出的期望值. n 是代表网络有 n 个输出.

交叉熵 (Cross Entropy, CE) 所表达的是输出与期望输出的距离, 如果输出与期望输出之间越接近, 如式 (2) 所示:

$$H(p, q) = - \sum_x p(x) \log q(x) \quad (2)$$

其中, q 为输出, 而 p 为期望输出.

(2) 难样本三元组损失函数

在三元组损失的基础上经过改进后形成难样本三元组损失, TriHard loss 的核心思想是: 每一次训练批量挑选 P 个行人, 将他们的 ID 作为属性. 每个 ID 行人, 随机挑选他们训练集中的照片, 所以每一个批量一共有 $P \times K$ 张图片. 在 $P \times K$ 张图片中, 选取对于网络最困难分辨的照片, 组成三元组, 三元组需要每次输入 3 张图片做为训练, 3 张图片分别被命名为:

- ① 锚点图片 (anchor, a);
- ② 正样品图片 (positive, p);
- ③ 负样本图片 (negative, n).

图片集 A 是锚点图片 (anchor) ID 相同的行人图片集合, B 为与锚点图片 (anchor) ID 不同的图片集合. 如式 (3) 所示:

$$L_{th} = \frac{1}{P \times K} \sum_{a \in \text{batch}} \left(\max_{p \in A} d_{a,p} - \min_{n \in B} d_{a,n} + \alpha \right)_+ \quad (3)$$

其中, batch 为输入批量, max 为最难的正样本对, min 为最难的负样本对. P 为输入的 ID 数, A, B 是输入的两个不同 ID. K 为输入每个 ID 数有几张照片, a 为

锚点照片, p 为正样本照片, n 为负样本照片, d 代表距离. α 是人为设置的参数. 相对于传统的三元组损失, TriHard loss 考虑了极端样本的信息, 而且一般有自适应权重计算, 即对于正样本来说, 越接近的样本, 权重越大. 对于负样本来说, 距离越大, 权重越小.

3 实验结果与分析

3.1 实验环境

本文所使用的的参数如表 1 所示.

表 1 实验环境表

参数	取值
显卡	NVIDIA GTX970
CPU	Inter Core i7 处理器
操作系统	Windows10
开发语言	Python
集成开发环境	Pycharm
实验框架	PyTorch
显卡驱动	CUDA、CuDNN

3.2 训练过程

训练过程以数据集 Market1501^[16], 本数据集来自清华大学校园中 6 个不同区域由摄像头拍摄采集的图像信息, 该训练集中有 751 个有标记 (ID) 的行人图像, 包含 12936 张图片, 训练时将图片像素转为 256×128 , 设置 margin(损失函数的阈值)=0.3, 设置 batch-size(批量)=32, 置 num-instance(输入行人 ID 的个数)=4, 每次输入 32 张照片, 分别来自 4 个 ID 行人, 每个行人 8 张照片. 设置迭代次数 (epoch)=150, stepsize (步长) 设置为 50, 其他参数默认. 在 HPLF 网络中图片经过 ResNet 网络, 对输出结果最后用 Triplet loss 和 Softmax 的联合损失函数来约束网络进行学习. 设定 CNN 的学习率为 0.0002, 并且 150 个 epoch 进行一次学习率递减为之前的 1/10, 进行训练. 采用梯度下降算法来修正网络的参数, 最后绘制训练和校验曲线图. 测试时, 在最后的卷积层提取图像的特征, 对欧式距离求解后排列测试集中的图像或者指定图像, 并且在测试文件中, 进行 mAP, Rank-1, Rank-5, Rank-10, Rank-20 的测试. 当 epoch=150 次以后 loss 函数基本达到最小值.

3.3 实验结果

本文测试所使用的指标为 mAP 以及 Rank-1、Rank-5、Rank-10. 表 2 和表 3 是 Market1501 在本文算法下获得测评结果, 与文献 [17] 中描述的方法对比,

精准性大大提高. 本文算法在单人检索有着 90.6% 的 Rank-1 指标和 77.7 的 mAP 指标. 但是在某些方面数据集的提升不太明显, 主要考虑是图片中目标人体被遮挡.

表 2 本文方法获得性能 (%)

指标	本文方法
Rank-1	90.6
Rank-5	96.4
Rank-10	97.7
Rank-20	98.8

本文实验结果与其他方法对比如表 3 所示.

表 3 Market1501 数据集结果对比 (%)

方法	mAP	Rank-1
MBC ^[18]	26.1	45.5
LSTM ^[3]	35.3	61.6
GAN ^[19]	55.9	79.3
Gate- ReID ^[2]	65.88	39.55
HA-CNN ^[20]	75.7	91.2
本文的方法	77.7	90.2

本文还做了自身对比实验, 在不使用本文的 HFML 算法的情况下得到的训练结果如表 4 所示.

表 4 不使用 HFML 算法获得性能 (%)

指标	本文方法
Rank-1	86.2
Rank-5	95
Rank-10	97.1
Rank-20	98.2
mAP	70.8

与表 2 比较, 可以看出使用本文的 HFML 算法的情况下 mAP 指标高了 6.9%, Rank-1 指标提升 4.4%, Rank-5 指标提升 1.4%, Rank-10 指标提升 0.6%, Rank-20 指标提升 0.6%. 可以看出本文的 HFML 算法可以有效的提取特征, 可以提升精度.

4 总结

本文提出了一种利用水平池化提取局部特征的 HPLF 算法, 在特征图进行水平切割, 将分割的特征图通过难样本三元组损失来作为局部特征损失函数, 同时原先特征图也将通过难样本三元组损失和 Softmax 损失函数来作为全局特征损失函数, 将 3 个损失函数作为联合损失函数进行参数修正. 未来考虑复杂场景下的行人再识别.

参考文献

- 林景栋, 吴欣怡, 柴毅, 等. 卷积神经网络结构优化综述. 自动化学报, 2020, 46(1): 24–37.
- Varior RR, Haloi M, Wang G. Gated Siamese convolutional neural network architecture for human re-identification. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands. 2016. 791–808.
- Varior RR, Shuai B, Lu JW, *et al.* A Siamese long short-term memory architecture for human re-identification. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands. 2016. 135–153.
- Wang YC, Chen ZZ, Wu F, *et al.* Person re-identification with cascaded pairwise convolutions. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 1470–1478.
- Cheng D, Gong YH, Zhou SP, *et al.* Person re-identification by multi-channel parts-based CNN with improved triplet loss function. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 1335–1344.
- Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification. arXiv preprint arXiv: 1703.07737, 2017.
- Liu H, Feng JS, Qi MB, *et al.* End-to-end comparative attention networks for person re-identification. IEEE Transactions on Image Processing, 2017, 26(7): 3492–3506. [doi: 10.1109/TIP.2017.2700762]
- Ristani E, Tomasi C. Features for multi-target multi-camera tracking and re-identification. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 6036–6046.
- Chen WH, Chen XT, Zhang JG, *et al.* Beyond triplet loss: A deep quadruplet network for person re-identification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 1320–1329.
- Chen HR, Wang YW, Shi YM, *et al.* Deep transfer learning for person re-identification. Proceedings of 2018 IEEE Fourth International Conference on Multimedia Big Data. Xi'an, China. 2018. 1–5.
- Lin YT, Zheng L, Zheng ZD, *et al.* Improving person re-identification by attribute and identity learning. Pattern Recognition, 2019, 95: 151–161. [doi: 10.1016/j.patcog.2019.06.006]

- 12 Zhang X, Luo H, Fan X, *et al.* AlignedReID: Surpassing human-level performance in person re-identification. arXiv: 1711.08184, 2017.
- 13 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, Germany. 2018. 501–518.
- 14 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.
- 15 Zhu JY, Park T, Isola P, *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy. 2017. 2242–2251.
- 16 Zheng L, Shen LY, Tian L, *et al.* Scalable person re-identification: A benchmark. Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1116–1124.
- 17 Ristani E, Solera F, Zou R, *et al.* Performance measures and a data set for multi-target, multi-camera tracking. Proceedings of the European Conference on Computer Vision. Amsterdam, the Netherlands. 2016. 17–35.
- 18 Ustinova E, Ganin Y, Lempitsky V. Multi-region bilinear convolutional neural networks for person re-identification. Proceedings of 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance. Lecce, Italy. 2017. 1–6.
- 19 Zheng ZD, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 3774–3782.
- 20 Li W, Zhu X, Gong S. Harmonious attention network for person re-identification. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 2285–2294. [doi: [10.1109/CVPR.2018.00243](https://doi.org/10.1109/CVPR.2018.00243)]