

# 基于 LSTM 的无监督域自适应行人重识别<sup>①</sup>



胡卓晶, 王 敏

(河海大学 计算机与信息学院, 南京 210024)

通讯作者: 胡卓晶, E-mail: hzj15503689046@163.com

**摘 要:** 本文提出一种基于无监督域自适应的行人重识别方法. 给定有标签的源域训练集和无标签的目标域训练集, 探索如何提升行人重识别模型在目标域测试集上的泛化能力. 以此为目的, 在模型的训练过程中, 将源域和目标域训练集同时输入到模型中进行训练, 提取全局特征的同时, 提取局部特征进行行人图像描述以学到更加细粒度的特征. 提出将长短时记忆网络 (LSTM) 以端到端的方式应用于行人的建模, 将其视为从头到脚的身体部分序列. 本文方法主要分为两个步骤: (1) 利用 StarGAN 对无标签目标域图片进行数据增强; (2) 源域和目标域数据集同时输入到全局分支和基于 LSTM 的局部分支共同训练. 在 Market-1501 和 DukeMTMC-reID 数据集上, 本文提出的模型都取得了较好的性能, 充分体现了其有效性.

**关键词:** 域自适应; 长短时记忆网络; 无监督

引用格式: 胡卓晶, 王敏. 基于 LSTM 的无监督域自适应行人重识别. 计算机系统应用, 2021, 30(2): 182-187. <http://www.c-s-a.org.cn/1003-3254/7780.html>

## LSTM-Based Unsupervised Domain Adaptive Person Re-Identification

HU Zhuo-Jing, WANG Min

(College of Computer and Information, Hohai University, Nanjing 210024, China)

**Abstract:** In this study, we propose a method of unsupervised domain adaptive person re-identification. Given a labeled source-domain training set and an unlabeled target-domain training set, we explore how to improve the generalization ability of the person re-identification model on the target-domain test set. For this purpose, during the training of the model, the source-domain and target-domain training sets are simultaneously input into the model for training. While extracting global features, we extract local features to describe the person images and learn more fine-grained features. Furthermore, we apply Long Short-Term Memory (LSTM) for the modeling of a person in an end-to-end manner, treating the person as a sequence of body parts from the head to feet. Specifically, the method in this paper mainly includes two steps: (1) StarGAN is adopted to enhance the data of unlabeled target domain images; (2) the data sets of source domain and target domain are input into global branch and LSTM-based local branch at the same time for joint training. Finally, on the Market-1501 and DukeMTMC-reID data sets, the proposed model has achieved sound performance, which fully reflects its effectiveness.

**Key words:** domain adaptive; Long Short-Term Memory (LSTM); unsupervised

## 1 引言

行人重识别 (person re-identification) 又称行人再识别, 被广泛认为是图像检索的子问题, 其目标是给定

一张监控行人图像, 跨设备检索该行人, 即确认不同摄像头在不同时刻拍到的是否为同一行人. 如何提取行人特征以及如何进行相似度度量就是行人重识别需要

<sup>①</sup> 收稿时间: 2020-06-10; 修改时间: 2020-07-10, 2020-07-21; 采用时间: 2020-07-27; csa 在线出版时间: 2021-01-27

解决的核心问题。行人重识别与行人检测、行人跟踪技术相结合,在公共安防的刑侦工作中以及图像检索等场景中有很高的应用价值。但由于图像拍摄的时间、地点随机,且光线、角度、行人姿态不同,再加上行人容易受到检测精度、遮挡等因素的影响,不同摄像头下造成行人外观的巨大变化,因此行人重识别技术仍面临着重大的挑战。

早期的行人重识别研究集中于如何手工设计好的视觉特征以及如何学习更好的相似性度量。随着深度学习的发展,鉴于其可以自动提取图像特征并学习好的相似性度量,研究者们致力于利用深度学习技术来研究行人重识别。起初研究者们主要关注的是用网络提取单帧图片的全局特征,根据损失的不同分为表征学习和度量学习两大类,前者将行人重识别问题看作分类问题或验证问题:利用行人ID或属性等作为标签来训练模型或输入一对行人图片,网络通过学习特征表示来判断两张图片是否属于同一行人;后者旨在通过网络学习出两张图片的相似度。全局特征的学习遇到性能瓶颈后,研究者引入局部特征和序列特征进行研究,并作为全局特征的重要补充。Sun等<sup>[1]</sup>在PCB的基础上,通过测量余弦距离,为各分区的离群值进行重新分区进而细化模型。Wang等<sup>[2]</sup>将图像均匀地水平划分为若干块,并在各局部分支中改变分区的数量,从而获得具有多个粒度的全局和局部特征表示。目前行人重识别在有监督学习方面已取得了很好的成绩。但有监督学习的一个最大限制是为数据贴标签的成本过高,在如今数据爆炸的时代下将行人重识别技术应用于实际的可能性微乎其微。为解决上述难题,研究者们开始逐渐尝试利用无标签数据进行训练,越来越多基于半监督和无监督的行人重识别方法开始出现。Wu等<sup>[3]</sup>首先使用每个身份的一个有标签的轨迹来初始化模型,然后使用该模型特征表示的识别能力来为未标记的轨迹赋标签,提出一种循序渐进的抽样策略,逐步增加伪标签候选项的数目,以取代现有的静态抽样策略。Deng等<sup>[4]</sup>以无监督的方式将有标签的图像从源域转换到目标域,然后用转换后的图像以有监督的方式训练行人重识别模型。Fu等<sup>[5]</sup>在源域上进行模型的预训练,然后利用无标签样本中(包含全局和局部)潜在的相似性从不同视角构建多个聚类并为其分配伪标签,分组和细化迭代进行。从研究者的实验来看,迁移学习的优势十分明显,可以充分利用有限的有标签数据

集,将有标签数据与无标签数据相结合共同训练网络,从而更好地解决行人重识别问题。

在实际应用中,对大规模数据集贴标签成本过高,且各摄像头所拍摄图像的风格差异较大。在本文中,利用有标签的源域数据集和经过风格转换的无标签目标域数据集同时对模型进行训练,这种设置在解决实际问题时更有意义。本文提出的网络架构包含全局分支和局部分支,局部分支利用LSTM实现,以更好地利用行人图像的各局部信息生成更加鲁棒的行人特征表示。通过这种方式,可以更好地优化模型,提升模型在目标域测试集上的泛化能力。下面将具体介绍本文所提出的行人重识别方法。

## 2 基于LSTM的无监督域自适应行人重识别

在行人重识别领域,对于无监督域自适应问题,有标签的源域数据集 $\{X_s, Y_s\}$ 包含 $N_s$ 张行人图片,每张图片 $x_s$ 对应一个标签 $y_s$ ,其中 $y_s \in \{1, 2, \dots, M_s\}$ ,并且 $M_s$ 是有标签的源域数据集中行人ID的数量。同样,在无标签的目标域中有 $N_t$ 张无标签的目标域图片 $\{X_t\}$ ,其中每张目标域图片 $x_t$ 的身份是未知的。本文的目的就是利用有标签的源域图片和无标签的目标域图片来提升模型在目标域测试集上的泛化能力。

### 2.1 网络架构

图1展示了本文模型的网络架构图。采用在ImageNet上预训练的ResNet50作为主干网络,其在一些行人重识别的研究中已取得了不错的性能。与原版本的不同之处在于我们移除了最后的全局平均池化层和1000维全连接层,并添加了两个独立的分支,分别学习全局特征表示和局部特征表示。第一个分支是全局分支,学习行人的全局特征表示,第一个全连接层的输出是2048维,命名为FC-2048,第二个全连接层的输出维度为源域ID数量,命名为FC-#ID。基于整张行人图片学习特征表示关注的更多的是整体的信息,包含体型等高维语义信息。然而在很多情况下,人体的局部比如头部、上半身、下半身含有更多具有判别力的信息,一些基于身体部位来学习特征表示的方法也证明其可以提升行人重识别的效果。因此学习行人的局部特征表示可以作为全局特征表示的一个强有力的补充。大多数基于局部的方法会将各身体部位严格划分后各自输入到完全独立的分支中,但这种方式忽略了各部位之间的空间连通性,会损失一部分各关联区域有判别力

的信息,因此在本文提出的局部分支中我们将特征向量水平划分为3个部分,再使用双向LSTM将各个区域连接起来就像是一个从头到脚的序列,这样可以增强各部位之间的连通性.局部分支中通道维度是2048维,双向LSTM的隐层单元数设为256,后续全连接层的命名方式同前一个分支一样,两个分支共享POOL-5

之前的部分.这样通过结合全局特征和基于LSTM的局部特征能够增强行人重识别模型所学特征的判别力.

本文提出的网络架构中各分支均使用两个损失函数来学习,一个是用于分类的交叉熵损失,增强模型的判别能力;一个是进行相似度学习的三元组损失,用于增强模型的相机不变性和域连通性.

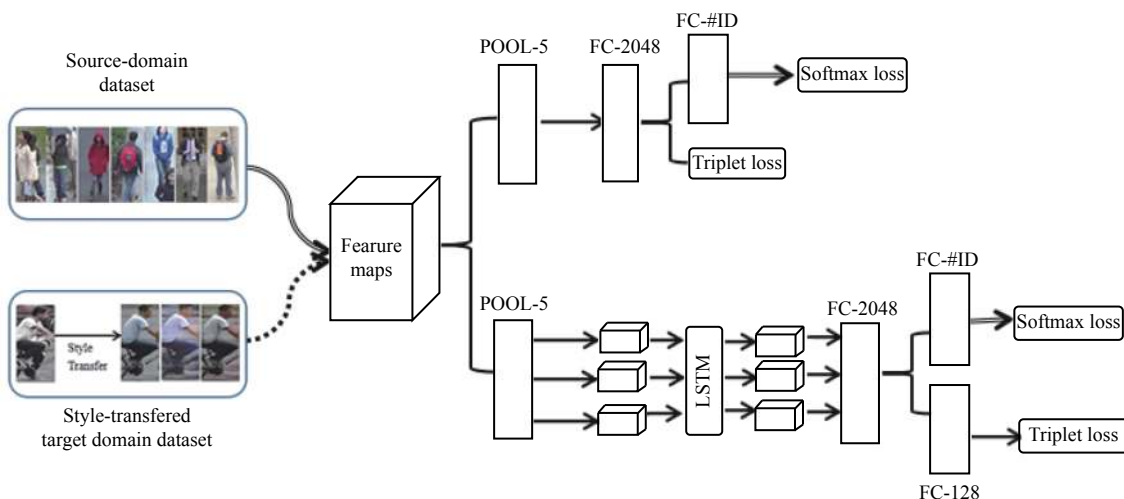


图1 网络架构图

网络架构图中的断点箭头为目标域训练集所经过的分支,空心箭头为源域训练集经过的分支,实心箭头为两个训练集共同经过的部分.

### 2.2 相机风格转换

在行人重识别测试阶段,由摄像头造成的图片风格差异是一个关键的影响因素.为使目标域图片不受相机风格转换的影响,我们使用无标签的目标域图片和该行人在其他相机中的对应图片进行相机一致性学习.我们采用StarGAN<sup>[6]</sup>构建的相机风格转换模型对目标域训练集进行风格迁移,这是因为StarGAN允许采用单个模型来训练多相机之间的图片-图片转换.使用学到的StarGAN模型,对于目标域训练集第 $j$  ( $j \in 1, 2, \dots, C$ )个相机拍摄的真实图片 $X_t^j$ ,我们可以生成 $C$ 张伪造的图片 $x_{r^*,1}, x_{r^*,2}, \dots, x_{r^*,C}$ ,所生成的图片都或多或少保留了行人的身份信息,但是整体风格分别与相机 $1, 2, \dots, C$ 类似.为在目标训练集中学习相机不变的行人特征映射,我们将原图与对应的生成图片视作同一类,其余图片视作不同类.

### 2.3 基于LSTM的局部特征表示学习

PCB<sup>[1]</sup>、MGN<sup>[2]</sup>等证实采用局部特征进行行人图

像描述可以学到更加细粒度的特征,因而提升行人重识别的性能.大多数基于部分的方法将行人身体部位严格划分,划分后得的各部分被输入到独立的分支来学习对应的局部特征.但部分独立学习的过程忽略了各部位之间的空间连通性,而这在行人重识别中对学习有判别力且鲁棒的特征映射是相当有用的.假如直接将行人水平划分为3部分,各行人姿态不同,可能在某些判别力强的部位进行分割,如衣服上的logo分割后各部分单独学习,无法学习完整的有判别力的特征,从而降低模型的判别能力.我们注意到行人自上而下可以划分为一个从头到脚的序列,即使在不同图片中各部分不会稳定在某一位置,所有的部分可以以一种序列的方式得益于身体结构的先验知识.LSTM单元架构图如图2所示.LSTM单元之间的循环连接能够生成依赖历史输入的特征.更重要的是,受益于内部门机制,LSTM可以控制信息从当前状态流入下一状态.基于上述分析,我们采用LSTM来为行人重识别建模身体序列.

LSTM单元结构:包含一个细胞模块 $c_t$ 和3个门,分别是输入门 $i_t$ 、输出门 $o_t$ 以及遗忘门 $f_t$ ,在 $t$ 时刻,LSTM将第 $t$ 个特征切片 $x_t$ 和前一个隐层的状态 $h_{t-1}$ 作为输入,并且预测一个特征向量 $o_t$ .



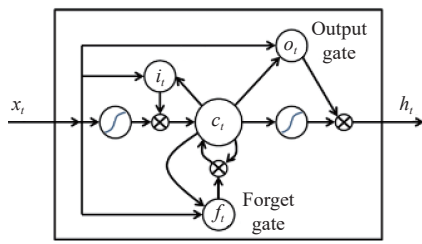


图2 LSTM单元架构图

## 2.4 损失函数

如2.1节所述,不同分支所学到的行人特征是互补的,因此我们联合训练整个网络学习具有判别力的全局特征和局部特征来预测行人身份。我们提出的模型不仅关注特征表示,还关注特征学习。给定有标签的训练集,一个有效的策略是为行人重识别学习ID判别映射(IDE),利用交叉熵损失将训练过程转换为分类问题,该策略利用所学的深层特征来区分不同的行人ID。交叉熵损失公式如下:

$$L_{\text{cross}} = -\frac{1}{n_s} \sum_{i=1}^{n_s} \log p_i(y) \quad (1)$$

其中,  $n_s$  为一个训练批中有标签的训练图片数量,  $p_i(y)$  指的是输入图片属于真实类别  $y$  的可能性。

但行人重识别的需求是将待检索图片与图库图片进行配对,而分类任务并不能直接学习行人重识别所要求的相似度。此外,基于IDE的模型在有标签数据集上能得到很好的性能,但迁移到一个新的数据集时,性能就会大打折扣,研究表明,利用三元组损失训练的距离排序分支可以学习图片的相似度。在整个训练过程中将交叉熵损失与三元组损失联合训练也是行人重识别框架中的一种传统操作。这样,判别分支和应用三元组损失的距离排序分支可以互补。接下来介绍我们提出的模型的训练策略。

如图1所示,有标签的源域图片和无标签的目标域图片同时输入到网络中,在全局分支,有标签的源域图片利用交叉熵损失和三元组损失来训练,无标签的目标域数据集行人ID数未知,因此仅利用三元组损失进行优化训练。三元组损失函数如下:

$$L_{\text{triplet}} = -\sum_{i=1}^P \sum_{a=1}^K \left[ \alpha + \max_{p=1, \dots, K} \|f_a^{(i)} - f_p^{(i)}\|_2 - \min_{\substack{n=1, \dots, K \\ j=i}} \|f_a^{(i)} - f_n^{(j)}\|_2 \right]_+ \quad (2)$$

式(2)中,随机选择训练样本的  $P$  个行人ID,以及各

ID对应的  $K$  个样本,  $f_a^{(i)}$ 、 $f_p^{(i)}$ 、 $f_n^{(i)}$  分别为 anchor、positive 以及 negative 的特征表示,上标中的  $i$ 、 $j$  分别表示的是行人ID,  $\alpha$  指的是三元组损失中的参数 margin。通过 Triplet Loss 的学习后使得 positive 元和 anchor 元之间的距离最小,而和 negative 元之间距离最大。其中 anchor 为训练数据集中随机选取的一个样本, positive 为与 anchor 属于同一类的样本,而 negative 则为与 anchor 不同类的样本。

同全局分支一样,在基于局部的分支中,有标签的源域图片同样利用交叉熵损失和三元组损失来训练,无标签的目标域图片利用三元组损失训练。

在行人重识别任务中,不同的域包含完全不同的类别或身份,因此一张源域图片和一张目标域图片自然构成一对负训练样本对。以此为先验条件,我们提出通过将源域和经过风格转换的目标域图片视为负样本对来为系统习得域连通性。给一张源域图片,我们使用源域标签构造一对正样本对,然后选择一张经过风格转换的目标域图片形成负样本对。因此给定有标签的源域图片和无标签的目标域图片,域连通性学习的损失函数可以定义为:

$$L_D = L_T \left( \{X_s^i\}_{i=1}^{n_s} \cup \{X_t^i\}_{i=1}^{n_t} \cup \{X_r^i\}_{i=1}^{n_r} \right) \quad (3)$$

最后,在一个训练批中的总的损失函数可以描述为:

$$Loss = L_{\text{cross}} + \theta L_D \quad (4)$$

其中,  $\theta$  是平衡交叉熵损失和域连通性损失的权重因子。

## 2.5 最大池化和平均池化

平均池化考虑特定部分的整个区域,因此,平均池化所生成的特征表示的判别能力很容易受到不相关背景模式的影响。例如,行人的某个分区判别能力很强,但由于周围有背景,此时全局平均池化所得到的该部分与周围背景的平均值,因此削弱该部分的判别能力。相反,全局最大池只保留局部视图的最大响应值。我们认为这两种池化策略在从全局和局部视图生成特性表示方面是互补的。因此,我们在模型中联合这两种池化策略,以融合并发挥这两种策略的优势。

## 3 实验分析

### 3.1 实验数据与评估标准

本文在行人重识别的两个常用行人数据集上进行训练与评估,分别为: Market-1501<sup>[7]</sup> 和 DukeMTMC-ReID<sup>[8,9]</sup>。其中 Market-1501 包含 6 个摄像头下的 1501 个

行人的 32668 张图片, 为方便评估, 751 个行人的 12936 张图片用于训练, 剩余的 750 个行人的 10732 张图片作为评估数据集. DukeMTMC-ReID 数据集包含 8 个摄像头下的 1404 个行人的 36411 张图片, 类似于 Market-1501 的划分, 该数据集包含 702 个行人的 16522 张训练图片和剩余 702 个行人的 2228 张待查找图片以及 17661 张图片形成的图库. 本文在各数据集上使用传统的平均准确度 mAP 和 rank- $n$  对实验结果进行评估.

### 3.2 实验环境及参数配置

本实验选取在 ImageNet 上预训练的 ResNet50 作为主干网络, 移除最后的全连接层和全局平均池化层. 通道数设为 2048, 双向 LSTM 中的隐层单元数设为 256, 三元组特征维度为 128. 我们的模型基于 PyTorch 框架实现, 使用两个 NVIDIA TITAN X GPU 进行训练.

本实验中采用通用的数据增强策略, 在训练时首先将所有图片大小调整为  $256 \times 128$ ; 然后随机裁剪每一张调整后的图像, 尺寸在区间  $[0.64, 1.0]$  内, 长宽比为  $[2, 3]$ ; 再将裁剪后的图片大小调整为  $256 \times 128$ , 应用概率为 0.5 的随机水平翻转. 在测试阶段, 仅将输入图片大小调整为  $256 \times 128$ . 模型的训练过程总共 30 个 epoch, 使用 Adam 优化器进行训练. 训练阶段学习率初始化为  $3e-4$ . 三元组损失中的边缘参数被置为 0.5, 参数  $\theta$  置为 0.3, dropout 率为 50%. 测试时, 提取 POOL-5 层的输出作为图片的特征表示, 并采用欧氏距离来计算待查找图片和数据库图片之间的相似度.

### 3.3 实验比较

将我们的方法与最先进的无监督学习方法进行比较. 表 1 给出了以 Market-1501/Duke 为源数据集, 以 Duke/Market-1501 为目标数据集时的比较. 其中 LOMO<sup>[10]</sup> 和 Bow<sup>[7]</sup> 是人工提取特征的方法, CAMEL<sup>[11]</sup> 是无监督学习方法, PTGAN<sup>[12]</sup>、SPGAN<sup>[4]</sup>、SPGAN+LMP<sup>[4]</sup>、TJ-AIDL<sup>[13]</sup> 和 HHL<sup>[14]</sup> 是无监督域自适应方法. 两种人工提取的特征直接应用于测试集而不需要任何训练, 但很明显, 这两种方法的性能都较差. 很明显, 在目标域数据集上进行训练时, 无监督方法的性能总是优于人工提取特征. 与无监督域自适应方法相比, 本文提出的方法性能更好. 具体来说, 在 Market-1501 上测试, 我们的结果高于所有竞争方法, rank-1 准确率 = 65.8%, mAP = 35.2%. 例如, 与最近发表的 HHL 方法<sup>[14]</sup> 相比, 我们的结果在 rank-1 精度上提高了 3.6%, 在 mAP 上提高了 3.8%. 在 DukeMTMC-reID 上进行测

试, 我们的方法获得 rank-1 准确率 = 48.1%, mAP = 28.7%, 也优于之前的方法. 验证了我们方法的有效性.

表 1 与当前先进算法性能比较

方法	Duke->Market				Market->Duke			
	mAP	R-1	R-5	R-10	mAP	R-1	R-5	R-10
LOMO <sup>[10]</sup>	8.0	27.2	41.6	49.1	4.8	12.3	21.3	26.6
Bow <sup>[7]</sup>	14.8	35.8	52.4	60.3	8.3	17.1	28.8	34.9
PTGAN <sup>[12]</sup>	-	38.6	-	66.1	-	27.4	-	50.7
SPGAN <sup>[4]</sup>	22.8	51.5	70.1	76.8	22.3	41.1	56.6	63.0
CAMEL <sup>[11]</sup>	26.3	54.5	-	-	-	-	-	-
SPGAN+LMP <sup>[4]</sup>	26.7	57.7	75.8	82.4	26.2	46.4	<b>62.3</b>	<b>68.0</b>
TJ-AIDL <sup>[13]</sup>	26.5	58.2	74.8	81.1	23.0	44.3	59.6	65.0
HHL <sup>[14]</sup>	31.4	62.2	78.8	84.0	27.2	46.9	61.0	66.7
本文	<b>35.2</b>	<b>65.8</b>	<b>82.5</b>	<b>87.4</b>	<b>28.7</b>	<b>48.1</b>	60.7	64.8

## 4 结论与展望

本文提出了一种新的无监督域自适应方法来解决无标签的行人重识别问题, 本次实验考虑不同数据集及各相机间拍摄风格的差异, 充分利用现有的有标签的数据来辅助无标签数据集联合训练网络, 在学习相机不变性和域联通性的同时, 利用 LSTM 来对行人进行建模, 提取细粒度特征的同时增强了各局部区域之间的连通性, 全局特征与局部特征相结合以学习更加鲁棒的行人特征表示, 进一步提升模型在无标签目标域的判别力. 在数据集 Market-1501 和 DukeMTMC-ReID 上的实验结果表明本文提出的方法效果良好. 但对于行人重识别问题的实际应用, 尤其是在半监督和无监督方面, 仍面临着巨大挑战, 未来有很多工作值得去做.

### 参考文献

- 1 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.
- 2 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, Republic of Korea. 2019. 274-282.
- 3 Wu Y, Lin YT, Dong XY, *et al.* Exploit the unknown gradually: One-shot video-based person re-identification by stepwise learning. Proceedings of 2018 IEEE/CVF

- Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 5177–5186.
- 4 Deng WJ, Zheng L, Ye QX, *et al.* Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 994–1003.
  - 5 Fu Y, Wei YC, Wang GS, *et al.* Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Republic of Korea. 2019. 6111–6120.
  - 6 Choi Y, Choi M, Kim M, *et al.* StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 8789–8797.
  - 7 Zheng L, Shen LY, Tian L, *et al.* Scalable person re-identification: A benchmark. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1116–1124.
  - 8 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.
  - 9 Ristani E, Solera F, Zou R, *et al.* Performance measures and a data set for multi-target, multi-camera tracking. Proceedings of European Conference on Computer Vision. Amsterdam, the Netherland. 2016. 17–35.
  - 10 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, MA, USA. 2015. 2197–2206.
  - 11 Yu HX, Wu AC, Zheng WS. Cross-view asymmetric metric learning for unsupervised person re-identification. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 994–1002.
  - 12 Wei LH, Zhang SL, Gao W, *et al.* Person transfer GAN to bridge domain gap for person re-identification. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 79–88.
  - 13 Wang JY, Zhu XT, Gong SG, *et al.* Transferable joint attribute-identity deep learning for unsupervised person re-identification. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 2275–2284.
  - 14 Zhong Z, Zheng L, Li SZ, *et al.* Generalizing a person retrieval model hetero- and homogeneously. Proceedings of the 15th European Conference on Computer Vision. Munich, Germany. 2018. 176–192.