Computer Science and Application 计算机科学与应用, 2021, 11(5), 1301-1314
Published Online May 2021 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2021.115132

Hans汉斯

多维注意力与部件关注的无监督行人重识别

麻可可，薛丽霞，汪荣贵，杨　娟

合肥工业大学计算机与信息学院，安徽　合肥
Email: m17185380928@163.com

收稿日期：2021年4月17日；录用日期：2021年5月11日；发布日期：2021年5月18日

摘　要

行人重识别的目的是通过将人物的探测图像与图像库中的所有图像进行比较，从而在图像库中找到感兴趣的人。大多数的行人重识别算法都是在一些小的带标签的数据集上进行监督训练，直接将这些训练好的模型部署到真实世界的大型摄像机网络中可能会由于拟合不足而导致性能低下。因此，有必要在没有明确监督的情况下，自主地对模型进行训练。因此本文提出了一个多维注意力网络和部件关注网络联合学习的无监督行人重识别方法。首先多维注意力网络对行人图像复杂的高阶统计信息进行建模和利用，其次使用部件关注网络关注不同的部件，最后是一系列的损失函数来引导部件关注网络学习未标记数据集上的部件特征。在Market-1501和DukeMTMC-reID两个数据集上的实验结果表明，本文提出的方法有效并取得了显著的效果。

关键词

行人重识别，多维注意力网络，部件关注网络

Unsupervised Person Re-Identification Based on Multi-Dimensional Attention and Part Focus Network

Keke Ma, Lixia Xue, Ronggui Wang, Juan Yang

School of Computer and Information, Hefei University of Technology, Hefei Anhui
Email: m17185380928@163.com

Received: Apr. 17th, 2021; accepted: May 11th, 2021; published: May 18th, 2021

文章引用: 麻可可, 薛丽霞, 汪荣贵, 杨娟. 多维注意力与部件关注的无监督行人重识别[J]. 计算机科学与应用, 2021, 11(5): 1301-1314. DOI: 10.12677/csa.2021.115132

Abstract

Person re-identification (Re-ID) aims at finding a person of interest in the image gallery by comparing the probe image of this person with all the gallery images. Most of the Re-ID algorithms conduct supervised training in some small labeled datasets, so directly deploying these trained models to the real-world large camera networks may lead to a poor performance due to underfitting. Therefore, it is necessary to train models without explicit supervision in an autonomous manner, and propose an unsupervised Re-ID method based on Multi-dimensional Attention Network (MDAN) and Part Focus Network (PFN). MDAN can model and utilize the complex higher-order statistics information in attention mechanism, so as to capture the subtle differences among pedestrians and to produce the discriminative attention proposals. Then there is a PFN, which is deployed into an improved spatial transform network (STN) so that each branch can focus on different parts of the pedestrian. We evaluate the proposed method on two public datasets, including Market-1501 and DukeMTMC-reID. Extensive experimental results show that the proposed method is effective and achieves impressive results.

Keywords

Person Re-Identification, Multi-Dimensional Attention Network, Part Focus Network

1. 引言

行人重识别(person re-identification, Re-ID)由于在视频监控、人机交互领域的重要应用，近年来受到了越来越多的关注。在视频监控中，由于相机分辨率和拍摄角度的缘故，通常无法得到高质量的人脸图片，此时人脸识别[1] [2] [3]失效。行人重识别就成了一个非常重要的替代技术，它广泛被认为是一个图像检索的子任务，给定目标人物的图像，行人重识别的目标是找到不同相机或同一相机在不同时间捕捉到的同一人的其他图像。

最近几年，越来越多的研究者尝试将行人重识别的研究与深度学习[4] [5] [6]结合在一起，并取得了很好的效果。现有的行人重识别的工作大部分聚焦于监督学习[7]-[14]，它们假设可以为每一对相机视图提供大量手动标记的匹配对，来学习该相机相对优化的特征表示或匹配度量函数。然而，这种规模的手动标签不仅在现实世界中收集起来非常困难，而且在许多情况下是不可行。例如可能没有足够的训练人员在每对相机视图中重新出现。这限制了它在真实应用场景中的扩展性和可用性。

针对上述问题，一种通用的解决方案是设计无监督模型[15]-[26]。虽然一些无监督行人重识别算法已经被提出，但是与监督学习方法相比，它们的识别效果较弱。一个主要的原因是，如果没有跨视图的标记数据，则无监督方法由于不同的视角、背景和照明而缺乏跨视图下相同身份视觉特征变化所需要的必要知识。对此，我们提出了一个多维注意力模块来解决行人图片中的视角、背景等的干扰，该模块可以去除图片中杂乱无章的背景噪声，从而提取出具有鲁棒性的嵌入特征。

此外，在行人重识别中一些基于部件生成的方法[12] [14]被提出来并超过了基于全局特征的方法。例如，Rahul 等人[12]提出了基于图像切块的方法，将图像块按顺序送到一个 LSTM 网络中，最后进行

特征融合。Wei 等人[14]提出了一种全局－局部对齐的特征描述子，来解决行人姿态变化的问题。虽然基于部件匹配效果超过了基于全局特征的方法，但是我们也发现了部件匹配存在的问题。那就是部件无法进行很好地对齐，从而影响了效果。为此，我们提出了部件关注模块来解决部件对齐的问题。该模块可以使部件能够很好地进行匹配，从而提取出具有鲁棒性的局部特征。

基于以上分析，本文提出了一个多维注意力网络和部件关注网络联合学习的无监督行人重识别方法。首先通过一个多维注意力模块来解决行人图片中的背景噪声的干扰；然后是一个部件关注网络，将它部署到一个改进的 STN 后，使得每个分支可以关注人体不同的部件，包括头部特征、上半身特征和下半身特征。

分别在两个已知的行人重识别的数据集上对本文提出的方法进行了评估，包括 Market-151 和 DukeMTMC-reID。结果表明本文提出的方法取得了非常好的效果。

2. 相关工作

在本节中，简要叙述行人重识别之前的工作，包括监督的行人重识别、无监督的行人重识别和行人重识别中的注意力机制。

2.1. 监督行人重识别

监督学习方法是解决行人重识别问题最常用的方法，包括基于表征的学习[7] [8]、基于度量学习[9] [10] [11]、基于局部特征学习[12] [13] [14]的行人重识别，它们主要得益于卷积神经网络的快速发展。

表征学习的方法是一种非常常见的行人重识别方法，它主要得益于卷积神经网络的快速发展。由于卷积神经网络可以自动从原始图像数据中根据任务需求自动提取表征特征，所以一些研究者把行人重识别问题看作分类或验证问题。例如，Zheng 等人[7]提出使用两种模型来学习行人描述符，分别是分类模型和验证模型。其中分类模型对输入图片进行预测，根据预测的 ID 来计算分类误差损失。验证模型融合两张图片的特征，判断两张图片是否为同一身份。Lin 等人[8]提出了基于属性和身份学习的方法，提出的模型不仅会学习到行人的 ID 特征，还可以学习到行人的属性特征，这大大增强了模型的泛化能力。

度量学习旨在通过网络学习出两张图片的相似度。在行人重识别问题上，具体表现为同一个人的不同图片相似度大于不同人图片的相似度。例如，Rahul 等人[9]提出了一个基于 Siamese 网络的方法，将一对图片输入到孪生网络中，使用交叉熵损失判断两张图片的相似度，使得正样本对之间的距离逐渐变小，负样本对之间的距离逐渐变大。Cheng 等人[10]提出了基于三元组损失的网络，需要输入三张图片(一对正样本对，一对负样本对)，使用三元组损失拉近正样本对的距离，推开负样本对的距离。Alexander 等人[11]提出了一个难样本采样三元组损失，主要解决的是三元组损失采样抽取出来的都是简单易区分样本对的问题。

早期行人重识别的研究大家只关注全局特征，就是用整张图片得到一个特征向量进行图像检索。但是仅仅使用全局特征达不到理想的效果，所以一些研究者开始关注局部特征。例如，Rahul 等人[12]提出了基于图像切块的方法，将图片垂直切割为若干块，按顺序送到一个 LSTM 网络中，最后进行特征融合。为了解决图像切块无法对齐问题，Zheng 等人[13]提出了基于姿态估计与关节点定位的方法，首先使用姿态估计模型估计出行人的 14 个关键点，然后使用仿射变换使得相同的关键点对齐，这些关键点可以将人体分为不同的部位。Wei 等人[14]提出了一种全局－局部对齐的特征描述子，来解决行人姿态变化的问题。

2.2. 无监督行人重识别

近年来一些研究者提出了具有手工特征的非监督方法[15] [16] [17] [18] [19]。然而，与监督学习方法相比，它们的再识别性能较差。例如，Zheng 等人[15]提出了一种无监督的 BOW 描述符，受图像搜索系统的启发，将每个行人图片表示成视觉词汇直方图，并支持全局快速匹配。Liao 等人[17]提出一个特征提取方法 LOMO，主要着眼于光照和视角问题，基于 HSV 颜色直方图和 XQDA 度量学习。Tetsu 等人[19]提出了一种基于像素特征的层次分布描述符，通过分层高斯分布描述图像中的局部区域。

由于卷积神经网络在监督行人重识别领域中取得了无可比拟的性能，因此一些研究者在无监督行人重识别中也使用了深度学习的方法，并取得了一定的效果。目前基于深度学习的无监督方法主要分为基于聚类的无监督行人重识别和基于跨域的无监督行人重识别。

聚类分析[20] [21] [22]是一种长期存在的无监督深度学习方法，近几年，基于聚类的无监督行人重识别被提出。例如，Fan 等人[20]提出了基于聚类和微调的无监督学习方法，采用 K-means 聚类进行标签估计，逐步选取可靠的图像，并利用这些图像对深度神经网络进行微调，学习识别特征。Yu 等人[21]提出了无监督非对称距离度量学习的方法，基于非对称的 K-means 聚类来实现试图不变性。Lin 等人[22]提出了一种自底向上的聚类框架，该框架根据预定义的标准将聚类分层组合，基于一个非常简单的最小距离准则和一个集群大小正则化项。然而，聚类得到的图像的伪标签可能是有噪声的，因为它可能会将相同的标签分配给具有不同身份的相似图片，使得区分相似的人更加困难。

最近，一些研究者提出了跨域迁移学习的无监督行人重识别[23] [24] [25] [26]的方法，主要利用带有标签的数据集来提高模型在目标数据集上的性能。例如，Peng 等人[23]提出了无监督跨数据集的迁移学习，通过字典学习机制将人的外观视图不变表示从带有标签的源数据集转移到无标签的目标数据集中，并获得了更好的性能。Zhong 等人[24]引入了摄像机样式转换方法来处理多个视图中的图像样式变化，并学习了一个摄像机不变的描述子空间。Wang 等人[25]首先利用源域上的属性进行训练，然后学习身份和属性的联合特征表示。

Gao 等人[26]首先提出了“相机感知”域适应，以减少源域和目标域之间的差异，然后利用目标域中每个摄像机的时间连续性来创建有区别的信息。

2.3. 行人重识别中的注意力机制

最近几年，注意力机制在计算机视觉[27]领域取得了巨大的成功，例如目标检测[28]、图像分割[29]和姿态估计[30]。对于行人重识别[31]-[39]，也存在着许多基于注意力机制的方法。

这些方法的共同策略是将注意力机制整合到深层模型中，以解决行人重识别的定位和不对齐问题。Li 等人[34]提出了一种协调注意力模型(harmonious attention, HA)，该模型从全身图像中定位身体部位，同时学习多尺度特征图。Song 等人[31]提出了 MGCAM 移除背景、视点和姿态等因素对行人的影响，使网络更关注于前景区域图像。Wang 等人[32]提出了一种基于新的无参数空间注意力，将特征图上激活之间的空间关系引入到模型中。Cheng 等人[35]提出了 ABD-Net，该模型将注意力模块和多样性正则化融入整个网络，其中注意力机制结合了通道和空间信息，避免注意力机制过度集中于前景，正则化可以增强隐藏激活和权重的多样性。Zhao 等人[37]提出了一种基于部件映射检测器的部件对齐表示方法，用于每个预定义的身体部件。Si 等人[39]提出了一种基于类间和类内注意模块的双重注意匹配网络，用于捕获视频序列的上下文信息。

3. 方法

在本节中，详细介绍了改进的无监督行人重识别学习框架，该框架如图 1 所示，它是基于 ResNet-50

的卷积神经网络。

首先，是一个多维注意力模块，将 ResNet-50 分解为两个部分，即 P1 (from conv1 to layer2)与 P2 (from layer3 to GAP)。P1 用于将给定的图像从原始像素空间编码为中层特征空间，P2 用于将注意信息编码为高层特征空间，可以对数据进行分类。在 P1 与 P2 之间放置不同维度的注意力模块，生成多样化的注意力特征，强化所学知识的丰富性。在 Sec.3.1 中详细介绍。

其次，使用一个改进的 STN 形成三个分支，在每个分支后加入一个部件关注模块，并使用带有标签的数据集来预训练部件关注模块，使得每个分支可以关注人体不同的部件，包括头部特征，上半身特征和下半身特征。在本节中的 Sec.3.2 详细介绍。

最后是损失函数模块，首先使用了一个存储器来存储部件的更新特征，然后是一系列的损失函数来引导部件关注网络学习未标记数据集上的部件特征。在本节的 Sec.3.3 中详细介绍。

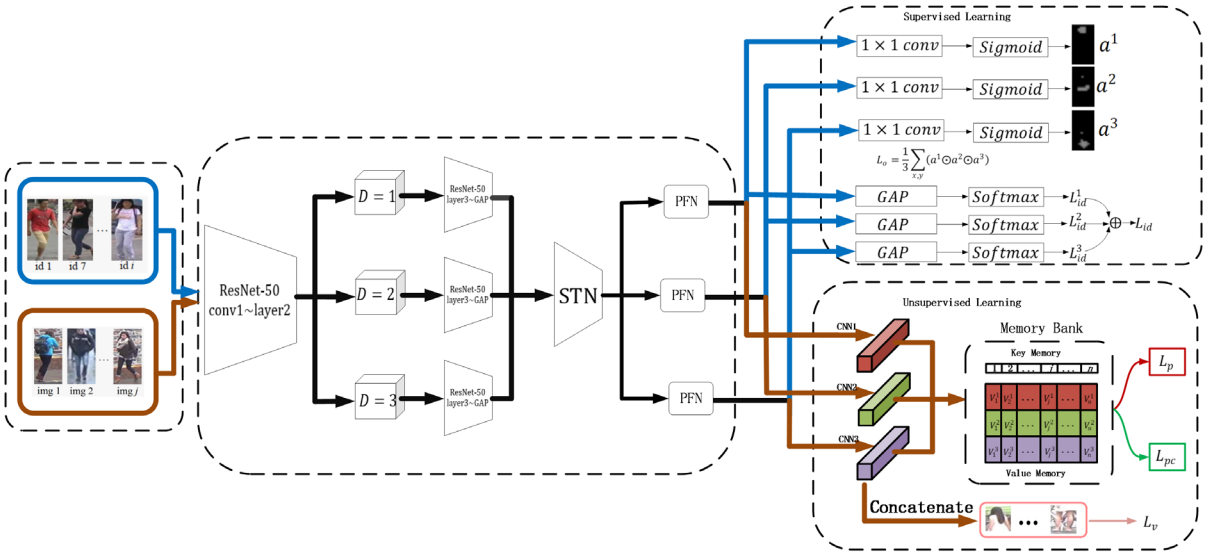

Figure 1. The framework of the proposed approach
图 1. 网络结构图

3.1. 多维注意力模块

注意力机制在行人重识别中变得更有吸引力，因为它能够将可用资源分配给输入信号中信息最丰富的部分，同时我们可以得到注意力方法的一般情况。具体的说，对于给定的三维张量 X，我们可以得到一个卷积激活输出。其中 $X \in R^{C\times H\times W}$，$C$、$H$、$W$ 分别表示通道数，高度和宽度。我们的目标是将卷积激活输出进行重新加权，将此过程表述为：

$$y = F(X) \odot X \tag{1}$$

其中 $y \in R^{C\times H\times W}$ 是注意力模块的激活输出，$\odot$ 表示两个矩阵对应元素相乘，$F(X)$ 是一个权重项，每个元素的值在区间 $[0,1]$ 之间。

然而，目前这些常用的注意力方法要么是粗糙的，要么是一维的，如图 2 中 $D=1$所示，仅限于挖掘简单而粗糙的信息。为此我们提出了高维注意力模块，如图 2 所示。以 $D=3$为例，它有三个分支，对于第 r 个分支，我们使用 r 组通道数相同的 1 × 1 的卷积核 $\{U_s^r\}_{s=1,\cdots,r}$ 生成通道为 D_r 的一组特征图 $\{Z_s^r\}_{s=1,\cdots,r}$。操作如下：

$$Z_s^r = W_s^r * X \tag{2}$$

其中 $W_s^r \in \mathbb{R}^{1\times1\times C\times D_r}$ 为卷积核的权重，X 为输入的一个三维张量。

将这组特征图结合在一起可得第 r 个分支的特征图为：

$$Z^r = Z_r^1 \odot \cdots \odot Z_r^r \tag{3}$$

然后再使用一组 1×1 的卷积核 α^r 来生成通道数为 C 的特征图，可以得到：

$$F(X) = \text{sigmoid}\left(\sum_{r=1}^{3} \alpha^{rT} \delta\left(z^r\right)\right) \tag{4}$$

其中 δ 代表 ReLU 激活函数，$F(X)$ 作为一个权重项，使用 sigmoid 激活函数将其中的每个元素值映射到 $[0,1]$ 的区间，将 $F(X)$ 代入公式(1)就可以得到高维注意力激活特征。

最后，将不同维度的注意力模块组合在一起，共同学习更具鲁棒性的特征。

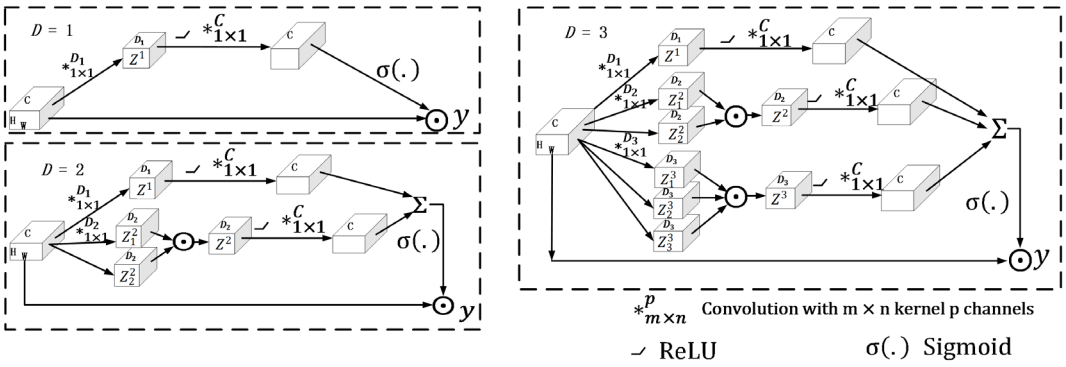

Figure 2. Illustration of Multi-Dimensional Attention Network (MDAN)
图 2. 多维注意力模块

3.2. 部件关注模块

我们首先使用一个改进的 STN 形成三个分支，在每个分支后加入一个部件关注模块，使得每个分支可以关注人体不同的部件。该模块由两部分组成，如图 3 所示。对于第一个分支，它关注部件的空间信息，对于第二个分支，它关注着部件的通道信息。

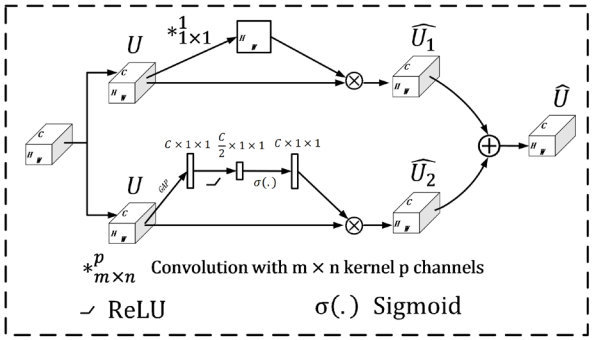

Figure 3. Illustration of Part Focus Network (PFN)
图 3. 部件关注网络

对于第一个分支，假设输入层的特征图为 U，令 $U = \left[u^{1,1}, u^{1,2}, \cdots, u^{i,j}, \cdots, u^{H,W}\right]$，$u^{i,j} \in \mathbb{R}^{1\times1\times C}$ 对应着

(i,j)的空间位置，其中$i\in\{1,2,\cdots,H\}$，$j\in\{1,2,\cdots,W\}$。通过通道数为 1 的 1×1 卷积实现空间压缩操作，如下：$q=W_{sq}*U$，其中$W_{sq}\in\mathbb{R}^{1\times1\times C\times1}$为权重，生成一个映射张量$q\in\mathbb{R}^{H\times W}$。映射的每一个$q_{i,j}$表示空间位置$(i,j)$的所有通道 C 的线性组合，再经过 sigmod 对其进行归一化操作到$[0,1]$。操作如下：

$$\hat{U}_1=\left[\sigma\left(q_{1,1}\right)u^{1,1},\cdots,\sigma\left(q_{i,j}\right)u^{i,j},\cdots,\sigma\left(q_{H,W}\right)u^{H,W}\right] \tag{5}$$

每一个$\sigma\left(q_{i,j}\right)$的值代表着特征图中空间位置坐标$(i,j)$的重要性。

对于第二个分支，将输入层特征图$U=[u_1,u_2,\cdots,u_C]$看作是信道$u_i\in\mathbb{R}^{H\times W}$的组合，经过全局池化层之后得到向量$z\in\mathbb{R}^{1\times1\times C}$，每个位置 k 的值为：

$$z_k=\frac{1}{H\times W}\sum_{i}^{H}\sum_{j}^{W}u_k(i,j) \tag{6}$$

然后经过两次全连接层，过程如下：

$$\hat{z}=W_1\left(\delta\left(W_2z\right)\right) \tag{7}$$

其中$W_1\in\mathbb{R}^{C\times\frac{C}{2}}$，$W_2\in\mathbb{R}^{\frac{C}{2}\times C}$，$W_1$，$W_2$分别为全连接层的权重，$\delta(\cdot)$代表着 RelU 激活操作。再经过 sigmod 对$\hat{z}$进行归一化操作到$[0,1]$。操作如下：

$$\hat{U}_2=\left[\sigma\left(\hat{z}_1\right)u_1,\sigma\left(\hat{z}_2\right)u_2,\cdots,\sigma\left(\hat{z}_i\right)u_i,\cdots,\sigma\left(\hat{z}_C\right)u_C\right] \tag{8}$$

$\sigma\left(\hat{z}_i\right)$代表的信息就是第 i 个通道u_i的重要性程度。

最后，将上述的两个模块结合从而得到此时的部件特征图$\hat{U}=\hat{U}_1+\hat{U}_2$。

为了更好地学习到行人的部件特征，我们使用带有标签的数据集 MSMT17 [43]对部件关注网络进行预训练。对于每一张标签为 t 的训练图像 I，其中 t 为目标类的索引。对于每个分支 i，可以得到部件特征图，即，$M^i\in\mathbb{R}^{h\times w\times c}$，为了使部件关注模块能够更好的关注我们感兴趣的部件，我们在每一个分支后添加 1 × 1 卷积和一个 Sigmoid 函数将值映射到(0,1)来获得每个部件的掩码$a^i\in\mathbb{R}^{h\times w}$。

$$a^i(x,y)=\frac{1}{1+\exp\left(-\left(W_{sq}*M^i\right)\right)} \tag{9}$$

其中 W 为 1 × 1 卷积的权重。由于目标是引导不同的分支来关注不同的身体区域，因此不同分支a^i的非零区域应该是不重叠的。因此，我们提出了一个损失函数来对重叠区域进行惩罚，定义如下：

$$L_o=\frac{1}{N}\sum_{x,y}\left(a^1\odot a^2\odot\cdots\odot a^N\right) \tag{10}$$

其中$\odot$表示逐元素乘法，N 表示分支数。

对部件特征M^i进行全局平均池化后，可以得到部件的类分数S^i，通过 Softmax 函数进一步归一化为概率分布$y^i\in\mathbb{R}^C$，因此第 i 个分支的分类损失被计算为预测概率y^i与真实值之间的交叉熵。

$$L_{id}^i=-\log\left(y_t^i\right) \tag{11}$$

其中 t 是目标类的索引，所有分支的损失相加构成了识别损失，$L_{id}=\sum_{i=1}^{N}L_{id}^i$。

最终的损失函数是L_{id}和L_o的加权和。

$$L_{total}=L_{id}+\alpha L_o \tag{12}$$

其中α是权重，并且在后面的实验中我们令$\alpha=1$。

3.3. 损失函数

3.3.1. 特征存储

在此我们使用存储器来存储部件的最新特征，存储器是一个键值对结构(key-value, K-V)。K 用来存储每个图像的索引。V 用来存储每一张图片的部件特征，包括三部分，分别用来存储行人的头部特征，上半身特征和下半身特征。如图所示，每个部件特征的存储特征可以表示为：$V^m=\left\{v_j^m\right\}_{j=1}^N$ 其中 $m\in\{1,2,3\}$ 分别代表着三个部件特征，N 表示训练图片的个数。我们通过以下公式对存储体进行更新：

$$v_{j,t}^m=\begin{cases}(1-l)\times v_{j,t-1}^m+l\times x_{j,t}^m, & t>0\\ x_{j,t}^m, & t=0\end{cases} \tag{13}$$

其中 t 是训练轮回数(epoch)，$t=0$ 表示存储器的初始状态。l 是 $v_{j,t}^m$ 的学习率，$x_{j,t}^m$ 是第 t 个轮回时第 j 张图片的第 m 个部件的特征。

3.3.2. 基于部件的损失

我们提出了一种基于部件的损失函数(PBL)，通过将相似部件的特征拉在一起，将不同的部件推开，从而引导部件提取网络学习未标记数据集上的部件特征。我们可以使用式 13 对存储器进行不断的更新，在此过程中，我们计算每个 x 与每个 v 的差异。详细的说，对于每一个 x_i^m，通过计算 x_i^m 与存储体 $V^m=\left\{v_j^m\right\}_{j=1}^N$ 之间的 l_2 距离从而获得 x_i^m 的 k 个最近的集合 K_i^m，为了拉近特征相似的部件，推远特征不相似的部件，我们提出了一个基于部件的损失函数，通过下式计算 PBL：

$$L_p^m=-\log\frac{\sum_{v_j^m\in K_i^m}\mathrm{e}^{-\frac{s}{2}\left\|x_i^m-v_j^m\right\|_2^2}}{\sum_{j=1,j\neq i}^N\mathrm{e}^{-\frac{s}{2}\left\|x_i^m-v_j^m\right\|_2^2}} \tag{14}$$

其中 s 是比例数，最小化 L_p^m 会促使模型将与 x_i^m 相似的部件 K_i^m 拉近，同时在特征空间中将与 x_i^m 不相似的部件 $\left\{v_j^m\mid v_j^m\notin K_i^m\right\}$ 推离。

为了对所有的部件进行约束，我们提出了基于部件约束的损失(BC)，它包含三部分，头部与上半身的约束，上半身与下半身的约束及整体的约束。这些约束就是将人体的部件进行融合，从而得到部件的约束特征。此时的样本约束特征可以表示为 $\left\{x_i^n\right\}$，存储体可以表示为 $\left\{v_j^n\right\}_{j=1}^N$。可以得到基于部件约束的损失函数为：

$$L_{pc}^n=-\log\frac{\sum_{v_j\in K_i}\mathrm{e}^{-\frac{s}{2}\left\|x_i^n-v_j^n\right\|_2^2}}{\sum_{j=1,j\neq i}^N\mathrm{e}^{-\frac{s}{2}\left\|x_i^n-v_j^n\right\|_2^2}} \tag{15}$$

其中 s 是比例数，$n=1$ 表示头部与上半身的约束，$n=2$ 表示上半身与下半身的约束，$n=3$ 表示整体的约束。

3.3.3. 基于图像级的损失

为了进一步挖掘图像间的潜在信息并最小化类内差异同时最大化类间差异，我们提出了一个三元组损失对全局特征进一步的约束。在我们的实验中，我们定义了一系列随机变换来生成代理正样本，包括图像的裁剪、缩放、旋转、亮度、对比度和饱和度。然后为每个真实样本生成一个代理正样本。

同时我们通过循环排序挖掘到负样本对，如图 4 所示：给定一小批量样本特征 $\{x_i\}_{i=1}^{B}$，每个样本 x_i 的排序结果可以基于成对相似度度量来得到。我们使用 l_2 距离来度量成对的相似性，从而得到图像 x_i 的排序列表 N_i。然后我们按顺序遍历排序列表 N_i。对于每个候选样本 $x_j \in N_i$，我们使用相同的方法来计算排序列表。最后，如果 x_i 不是 x_j 的 top-r 最近邻，那么我们认为 x_j 很可能是 x_i 的负样本。此外，由于难分的负样本对可以更有效地学习判别特征，因此我们只考虑第一个符合上述条件的候选样本 x_j。我们把这个负的候选项表示成负样本 n_i。在这里我们给定三元组损失的定义：

$$L_v = \max\left\{\left\|x_i - p_i\right\|_2 - \left\|x_i - n_i\right\|_2 + m, 0\right\} \tag{16}$$

其中 m 代表着三元组损失的间隔(margin)，p_i 代表正样本，n_i 代表负样本。

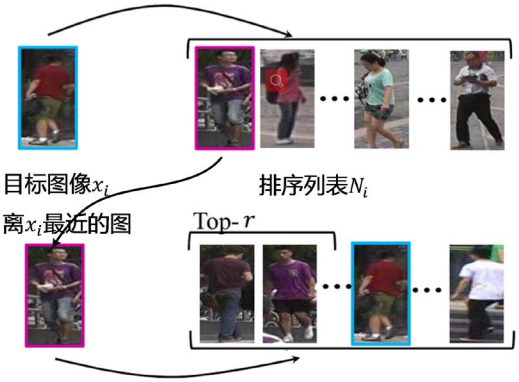

Figure 4. Illustration of the cyclic ranking
图 4. 循环排序模块

3.3.4. 最终的损失

因此，我们模型中的每一幅图像的总体损失函数可以表示为：

$$L = L_v + \lambda \frac{1}{M}\sum_{i=1}^{M} L_p^m + \mu \frac{1}{N}\sum_{i=1}^{N} L_{pc}^n \tag{17}$$

其中 λ 是部件损失的权重系数，μ 是部件约束损失的权重系数。

4. 实验

在这一部分中，主要是对所提出的方法进行评估。首先在第 4.1 节中介绍使用的数据集与评估指标，其次在 4.2 节中的介绍了实验细节。然后在 4.3 节中，进行消融实验来验证组件的有效性。最后将本文提出的方法与其他最先进的算法进行比较。

4.1. 数据集与评估指标

Market-1501 [40]是一个大型的行人重识别数据集，它包含 1501 个 id 的 32,688 张图像。这些图片是在校园里被 6 台摄像机拍了下来。数据集分为两部分：其中 751 个 id 的 12,936 张图像用于训练，750 个 id 的 19,732 张图像用于测试。

DukeMTMC-reID [41]是行人跟踪数据集 DukeMTMC [42]的子集，它包含了从 8 个相机中收集到的 36,411 张 1812 个 id 的图片。与 Market-1501 数据集的划分类似，数据集包含来自 702 个 ID 的 16,522 个训练图像样本，来自其他 1,110 个 ID 的 2228 个查询图像样本和 17,661 个待匹配的图像库样本。

MSMT17 [43]是目前最大的行人重识别数据集，它包含了 15 个摄像头中 4101 个 id 的 126,441 张

图像。与 Market-1501 和 DukeMTMC-reID 类似，数据集被分为两部分：用于训练的 1041 个 id 的 32,621 张图像，用于查询的 3060 个 id 的 82,161 张图像和 11,659 个待匹配的图像库样本。

Evaluation Metrics：本文使用 rank1、rank5、rank10 和平均精度(mAP)来评估这三个数据集。

4.2. 实验细节

实验基于 Linux 环境下的开源 Pytorch 框架[36]，硬件基础为 NVIDIA GeForce RTX 2080GPU。使用 ResNet-50 [44]作为基本 CNN 模型，并在 ImageNet 数据集[45]上进行预训练，首先将 ResNet-50 [44]分解成两部分 P1 和 P2，在 P1 和 P2 之间放置不同维数的注意力模块，生成不同维度的注意图，每个注意力模块的输出特征的维数为 256。然后将不同的注意力模块特征相融合并使用一个改进的空间变换网络(STN)形成三个分支，在每个分支后加入一个部件关注网络，并使用带有标签的数据集 MSMT17 [43]对部件关注网络进行预训练，使之关注人体的不同部件，即生成 3 个 256 维的部件特征(头部，上半身，下半身)。对于生成的每个部件，使用特征存储器来进行部件特征存储更新，更新率l设置为 0.1。同时将 DukeMTMC-reID [41]数据集上的比例系数 s 设置为 10，Market-1501 [40]上的设置为 30。在对未标记的数据集进行训练时，随机采样图像，并将其大小调整为 384 × 128。每个 mini-batch 由 8 个真实样本组成，另外还有 8 个代理正样本用于计算三元组损失。使用 SGD [32]作为优化算法，初始的学习率设置为 0.0001，每 20 个轮回数(epoch)衰减 0.1。在未标记的数据集上对模型进行了 60 个轮回数(epoch)的训练。测试时，将同一幅图像的部件特征拼接在一起，计算出两两距离。

4.3. 消融实验

为了对本文提出的方法性能进行验证，在 Market-1501 [40]和 DukeMTMC-reID [41]数据集上进行了大量的消融实验，来分析每个组件的有效性。实验结果如表所示。

注意力维度个数对实验的影响：从表 1 可以看出，维度个数为 4 实验效果最好。

部件个数对实验的影响：从表 2 可以看出，部件个数对实验有一定的影响。多次实验结果表明，部件个数为 3 时，效果最好。

不同模块对实验的影响：从表 3 可以看出，“MDAN”的结果比“基本网络”的结果要好；“PFN”的结果比“基本网络”的结果要好，“MDAN”与“PFN”联合学习的效果最好。

Table 1. The influence of the number of attention dimensions on the experiment (%)
表 1. 注意力维度个数对实验的影响

维度个数	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
2	70.9	40.7	70.2	50.1
3	72.1	41.5	71.8	52.0
4	73.2	42.3	73.5	53.8
5	71.5	40.9	71.1	51.8

Table 2. The influence of different parts number on experiment (%)
表 2. 不同部件个数对实验的影响

部件个数	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
2	70.3	39.7	69.2	49.3
3	72.1	41.9	72.6	52.1
4	71.4	40.3	71.1	50.6

Table 3. The influence of different modules (%)
表 3. 不同模块对在两个数据集上对实验的影响

模块	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
基本网络	64.3	36.5	66.9	45.7
MDAN	73.2	42.3	73.5	53.8
PFN	72.1	41.9	72.6	52.1
MDAN + PFN	76.2	45.1	75.6	55.9

4.4. 对比实验

将本文提出的模型与其他先进的无监督行人重识别模型进行了比较，包括：1) 基于手工制作的模型 LOMO [17]、BOW [40]和 UDML [24]；2) 基于伪标签学习的模型 PUL [20]、DBC [21]和 BUC [22]；3) 基于无监督域自适应模型 TJ-AIDL [25]、HHL [26]、UCDA-CCE [46]、SSL-CCE [47]和 ECN [48]。

其中表 4 是 Market-1501 数据集的比较结果，表 5 是 DukeMTMC-reID 数据集的比较结果。可以看到，本文提出的方法在两个数据集上都明显优于其他方法。与手工制作的基于特征表示的模型比较时，性能差距最为显著。其主要原因是这些早期的作品大多基于启发式设计，因此无法学习到最优的判别特征；与基于伪标签学习模型的比较时，本文模型明显优于基于伪标签学习的无监督行人重识别模型。一个关键的原因是基于伪标签的方法可能会将不同身份的相似图片分成相同的伪标签，在本文中，即使将不同身份的某一部件拉近，仍然有图像的其他部件进行补充；与基于无监督域自适应的行人重识别模型相比，所提出的模型具有明显的优势，一个关键的原因是源域图像与目标域图像之间的差距大于图像部分之间的差距，所以基于图像的特征学习模型很难转移到目标域。

Table 4. Comparison to the state-of-art unsupervised results in the Market-1501 dataset
表 4. 不同算法在 Market-1501 数据集上的性能比较

方法	Rank-1	Rank-5	Rank-10	mAP
LOMO [17]	27.2	41.6	49.1	8.0
BOW [40]	35.8	52.4	60.3	14.8
UMDL [24]	34.5	52.6	59.6	12.4
PUL [20]	45.5	60.7	66.7	20.5
BUC [22]	66.2	79.6	84.5	38.3
DBC [21]	69.2	83.0	87.8	41.3
HHL [26]	62.2	78.8	84.0	31.4
TJ-AIDL [25]	58.2	74.8	81.1	26.5
UCDA-CCE [46]	64.3	-	-	34.5
SSL-CCE [47]	71.7	83.8	87.4	37.8
ECN [48]	75.1	87.6	91.6	43.0
Ours	76.2	89.1	93.4	45.1

Table 5. Comparison to the state-of-art unsupervised results in the DukeMTMC-reID dataset
表 5. 不同算法在 DukeMTMC-reID 数据集上的性能比较

方法	Rank-1	Rank-5	Rank-10	mAP
LOMO [17]	12.3	21.3	26.6	4.8
BOW [40]	17.1	28.8	34.9	8.3

Continued

UMDL [24]	18.5	31.4	37.6	7.3
PUL [20]	45.5	60.7	66.7	20.5
BUC [22]	47.4	62.6	68.4	27.5
DBC [21]	51.5	64.6	70.1	30.0
HHL [26]	46.9	61.0	66.7	27.2
TJ-AIDL [25]	44.3	59.6	65.0	23.0
UCDA-CCE [46]	55.4	-	-	36.7
SSL-CCE [47]	52.5	63.5	68.9	28.6
ECN [48]	63.3	75.8	80.4	40.4
Ours	75.6	84.4	90.7	55.9

5. 结论

在本文中，首先提出了多维注意力网络，对注意机制中复杂的高阶统计信息进行建模和利用，从而捕捉行人之间的细微差异，产生有区别的注意建议。其次，使用一个改进的 STN 形成三个分支，在每个分支后加入一个部件关注网络，并使用带有标签的数据集来预训练部件关注网络，使得每个分支可以关注人体不同的部件。最后是损失函数，使用了一个存储器来存储部件的更新特征，并通过一系列损失函数来引导部件关注网络学习未标记数据集上的部件特征。大量的实验验证了该方法的有效性。然而目前的行人重识别数据集基本上维持着几万张图片几千个 ID 的水平上，这大大阻碍了基于深度学习的行人重识别的研究。因此，在下一步的研究中，可以使用 GAN 来增加数据量的规模从而提高模型泛化的能力。

参考文献

[1] Kemelmacher-Shlizerman, I., Seitz, S.M., Miller, D. and Brossard, E. (2016) The Megaface Benchmark: 1 Million Faces for Recognition at Scale. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 4873-4882. https://doi.org/10.1109/CVPR.2016.527

[2] Liao, S., Lei, Z., Yi, D. and Li, S.Z. (2014) A Benchmark Study of Large-Scale Unconstrained Face Recognition. *IEEE International Joint Conference on Biometrics*, Clearwater, 29 September-2 Octoner 2014, 1-8. https://doi.org/10.1109/BTAS.2014.6996301

[3] Taigman, Y., Yang, M., Ranzato, M. and Wolf, L. (2014) DeepFace: Closing the Gap to Human-Level Performance in Face Verification. *Proceedings of the* 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 1701-1708. https://doi.org/10.1109/CVPR.2014.220

[4] LeCun, Y., Bengio, Y. and Hinton, G.E. (2015) Deep Learning. *Nature*, **521**, 436-444. https://doi.org/10.1038/nature14539

[5] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) Imagenet Classification with Deep Convolutional Neural Networks. 25*th International Conference on Neural Information Processing Systems*, Stateline, December 2012, 1097-1105.

[6] Schmidhuber, J. (2015) Deep Learning in Neural Networks: An Overview. *Neural Networks*, **61**, 85-117. https://doi.org/10.1016/j.neunet.2014.09.003

[7] Zheng, Z., Zheng, L. and Yang, Y. (2017) A Discriminatively Learned CNN Embedding for Person Re-Identification. *ACM Transactions on Multimedia Computing, Communications, and Applications*, **14**, Article No. 13. https://doi.org/10.1145/3159171

[8] Lin, Y., Zheng, L., Zheng, Z., Wu, Y. and Yang, Y. (2016) Improving Person Re-Identification by Attribute and Identity Learning. *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, March 2017, 20-28.

[9] Rahul, V., Rama, Mrinal, H. and Gang, W. (2016) Gated Siamese Convolutional Neural Network Architecture for Human Re-Identification. *European Conference on Computer Vision*, 8-16 October, Amsterdam, 791-808. https://doi.org/10.1007/978-3-319-46484-8_48

[10] Cheng, D., Gong, Y., Zhou, S., Wang, J. and Zheng, N. (2016) Person Re-Identification by Multichannel Parts-Based

CNN with Improved Triplet Loss Function. *Proceedings of the* 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 1335-1344. https://doi.org/10.1109/CVPR.2016.149

[11] Hermans, A., Beyer, L. and Leibe, B. (2017) In Defense of the Triplet Loss for Person Re-Identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, March 2017, 5767-5782.

[12] Varior, R.R., Shuai, B., Lu, J., Xu, D. and Wang, G. (2016) A Siamese Long Short-Term Memory Architecture for Human Re-Identification. *European Conference on Computer Vision*, Amsterdam, 8-16 October, 135-153. https://doi.org/10.1007/978-3-319-46478-7_9

[13] Zheng, L., Huang, Y., Lu, H. and Yang, Y. (2017) Pose-Invariant Embedding for Deep Person Re-Identification. *IEEE Transactions on Image Processing*, **28**, 4500-4509. https://doi.org/10.1109/TIP.2019.2910414

[14] Wei, L., Zhang, S., Yao, H., Gao, W. and Tian, Q. (2017) GLAD: Global-Local-Alignment Descriptor for Scalable Person Re-Identification. *IEEE Transactions on Multimedia*, **21**, 986-999. https://doi.org/10.1109/TMM.2018.2870522

[15] Farenzena, M., Bazzani, L., Perina, A., Murino, V. and Cristani, M. (2010) Person Re-Identification by Symmetry-Driven Accumulation of Local Features. 2010 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, 13-18 June 2010, 2360-2367. https://doi.org/10.1109/CVPR.2010.5539926

[16] Chen, D., Yuan, Z., Chen, B. and Zheng, N. (2016) Similarity Learning with Spatial Constraints for Person Re-Identification. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 1268-1277. https://doi.org/10.1109/CVPR.2016.142

[17] Liao, S., Hu, Y., Zhu, X. and Li, S.Z. (2015) Person Re-Identification by Local Maximal Occurrence Representation and Metric Learning. 2015 *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 2197-2206. https://doi.org/10.1109/CVPR.2015.7298832

[18] Zhao, R., Ouyang, W. and Wang, X. (2013) Unsupervised Salience Learning for Person Re-Identification. 2013 *IEEE Conference on Computer Vision and Pattern Recognition*, Portland, 23-28 June 2013, 3586-3593. https://doi.org/10.1109/CVPR.2013.460

[19] Wang, H., Gong, S. and Xiang, T. (2014) Unsupervised Learning of Generative Topic Saliency for Person Re-Identification. *Proceedings of* 2014 *British Machine Vision Conference*, Nottingham, 1-5 September 2014, 1-11.

[20] Fan, H., Zheng, L., Yan, C. and Yang, Y. (2018) Unsupervised Person Reidentification: Clustering and Fine-Tuning. *ACM Transactions on Multimedia Computing, Communications, and Applications*, **14**, Article No. 83. https://doi.org/10.1145/3243316

[21] Ding, G., Khan, S., Tang, Z. and Zhang, J. (2019) Towards Better Validity: Dispersion Based Clustering for Unsupervised Person Re-Identification. *IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, June 2019, 1485-1494.

[22] Lin, Y., Dong, X., Zheng, L., Yan, Y. and Yang, Y. (2019) A Bottom-up Clustering Approach to Unsupervised Person Re-Identification. *AAAI Conference on Artificial Intelligence*, Honolulu, 27 January-1 February 2019, 8738-8745.

[23] Chen, H., Wang, Y., Shi, Y., Yan, K., Geng, M., Tian, Y., *et al*. (2018) Deep Transfer Learning for Person Re-Identification. *IEEE 4th International Conference on Multimedia Big Data*, Xi'an, 13-16 September 2018, 1-5. https://doi.org/10.1109/BigMM.2018.8499067

[24] Peng, P., Xiang, T., Wang, Y., Pontil, M., Gong, S., Huang, T. and Tian, Y. (2016) Unsupervised Cross-Dataset Transfer Learning for Person Re-Identification. *Proceedings of the* 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 1306-1315. https://doi.org/10.1109/CVPR.2016.146

[25] Wang, J., Zhu, X., Gong, S. and Li, W. (2018) Transferable Joint Attribute-Identity Deep Learning for Unsupervised Person Re-Identification. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 2275-2284. https://doi.org/10.1109/CVPR.2018.00242

[26] Zhong, Z., Zheng, L., Li, S. and Yang, Y. (2018) Generalizing A Person Retrieval Model Hetero- and Homogeneously. *Proceedings of the* 2018 *European Conference on Computer Vision*, Munich, 8-14 September 2018, 172-188.

[27] Li, G. and Yu, Y. (2016) Visual Saliency Detection Based on Multiscale Deep CNN Features. *IEEE Transactions on Image Processing*, **25**, 5012-5024. https://doi.org/10.1109/TIP.2016.2602079

[28] Chen, L., Zhang, H., Xiao, J., Nie, L., Shao, J., Liu, W. and Chua, T.-S. (2017) SCA-CNN: Spatial and Channel-Wise Attention in Convolutional Networks for Image Captioning. 2017 *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 6298-6306. https://doi.org/10.1109/CVPR.2017.667

[29] Chen, L.-C., Yang, Y., Wang, J., Xu, W. and Yuille, A.L (2016) Attention to Scale: Scale-Aware Semantic Image Segmentation. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 3640-3649. https://doi.org/10.1109/CVPR.2016.396

[30] Liu, H., Feng, J., Qi, M., Jiang, J. and Yan, S. (2016) End-to-End Comparative Attention Networks for Person Re-Identification. *IEEE Transactions on Image Processing*, **26**, 3492-3506. https://doi.org/10.1109/TIP.2017.2700762

[31] Ba, J., Mnih, V. and Kavukcuoglu, K. (2014) Multiple Object Recognition with Visual Attention. arXiv:1412.7755.

[32] Chu, X., Yang, W., Ouyang, W., Ma, C., Yuille, A.L. and Wang, X. (2017) Multi-Context Attention for Human Pose Estimation. *Proceedings of the* 2017 *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 1831-1840. https://doi.org/10.1109/CVPR.2017.601

[33] Si, J., Zhang, H., Li, C.-G., Kuen, J., Kong, X., Kot, A. and Wang, G. (2018) Dual Attention Matching Network for Context-Aware Feature Sequence Based Person Re-Identification. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 5363-5372. https://doi.org/10.1109/CVPR.2018.00562

[34] Li, W., Zhu, X. and Gong, S. (2018) Harmonious Attention Network for Person Re-Identification. *Proceedings of the* 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 2285-2294. https://doi.org/10.1109/CVPR.2018.00243

[35] Xu, J., Zhao, R., Zhu, F., Wang, H. and Ouyang, W. (2018) Attention-Aware Compositional Network for Person-nre-Identification. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 2119-2128. https://doi.org/10.1109/CVPR.2018.00226

[36] Chang, X., Hospedales, T. and Xiang, T. (2018) Multi-Level Factorisation Net for Person Re-Identification. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 2109-2118. https://doi.org/10.1109/CVPR.2018.00225

[37] Xu, J., Zhao, R., Zhu, F., Wang, H. and Ouyang, W. (2018) Attention-Aware Compositional Network for Person-nre-Identification. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 2119-2128. https://doi.org/10.1109/CVPR.2018.00226

[38] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141. https://doi.org/10.1109/CVPR.2018.00745

[39] Xiao, T., Li, S., Wang, B., Lin, L. and Wang, X. (2017) Joint Detection and Identification Feature Learning for Person Search. 2017 *IEEE International Conference on Computer*, Honolulu, 21-26 July 2017, 3376-3385. https://doi.org/10.1109/CVPR.2017.360

[40] Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J. and Tian, Q. (2015) Scalable Person Re-Identification: A Benchmark. 2015 *IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1116-1124. https://doi.org/10.1109/ICCV.2015.133

[41] Zheng, Z., Zheng, L. and Yang, Y. (2017) Unlabeled Samples Generated by GAN Improve the Person Re-Identification Baseline *in Vitro*. 2017 *IEEE International Conference on Computer*, Venice, 22-29 October 2017, 3774-3782. https://doi.org/10.1109/ICCV.2017.405

[42] Ristani, E., Solera, F., Zou, R.S., Cucchiara, R. and Tomasi, C. (2016) Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking. 2016 *ECCV Workshop on Benchmarking Multi-Target Tracking*, Amsterdam, 8-16 October 2016, 17-35. https://doi.org/10.1007/978-3-319-48881-3_2

[43] Wei, L., Zhang, S., Gao, W. and Tian, Q. (2018) Person Transfer GAN to Bridge Domain Gap for Person Re-Identification. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 79-88. https://doi.org/10.1109/CVPR.2018.00016

[44] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[45] Deng, J., Dong, W., Socher, R., Li, J., Li, K. and Li, F. (2009) ImageNet: A Large-Scale Hierarchical Imagedatabase. 2009 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Miami, 20-25 June 2009, 248-255. https://doi.org/10.1109/CVPR.2009.5206848

[46] Qi, L., Wang, L., Huo, J., Zhou, L., Shi, Y. and Gao, Y. (2019) A Novel Unsupervised Camera-aware Domain Adaptation Framework for Person Re-Identification. 2019 *IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 8079-8088. https://doi.org/10.1109/ICCV.2019.00817

[47] Lin, Y., Xie, L., Wu, Y., Yan, C. and Tian, Q. (2020) Unsupervised Person Re-Identification via Softened Similarity Learning. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognitionn*, Seattle, 13-19 June 2020, 3387-3396. https://doi.org/10.1109/CVPR42600.2020.00345

[48] Zhong, Z., Zheng, L., Luo, Z., Li, S. and Yang, Y. (2019) Invariance Matters: Exemplar Memory for Domain Adaptive Person Re-Identification. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 598-607. https://doi.org/10.1109/CVPR.2019.00069