

基于多层级特征的行人重识别

刘可文^{1,2}, 房攀攀^{1,2}, 熊红霞^{3*}, 刘朝阳⁴, 马圆^{1,2}, 李小军^{1,2}, 陈亚雷^{1,2}

¹ 武汉理工大学信息工程学院, 湖北 武汉 430070;

² 武汉理工大学宽带无线通信和传感器网络湖北省重点实验室, 湖北 武汉 430070;

³ 武汉理工大学土木工程与建筑学院, 湖北 武汉 430070;

⁴ 中国科学院武汉物理与数学研究所波谱与原子分子物理国家重点实验室, 湖北 武汉 430071

摘要 针对现有的行人重识别方法提取行人特征过程中存在因信息缺失导致鲁棒性和判别力较差的问题, 提出了一种基于残差神经网络提取行人图像多层级特征的方法。该方法在训练阶段使用残差网络分别在 4 个卷积残差模块之后提取阶段特征, 以此来弥补信息丢失, 使用三元组损失函数对每个特征向量进行监督训练。在相似性度量阶段, 针对 4 个特征向量分别计算特征相似度, 使用映射函数进行求和, 并对求和结果进行相似度匹配。将该方法在 Market-1501 和 DukeMTMC-ReID 数据集上进行仿真, 首中准确率 (Rank-1) 分别达到了 91.7% 和 84.9%, 平均准确率 (mAP) 分别达到了 86.8% 和 80.7%。结果表明所提方法提取的多层级特征具有较好的鲁棒性和判别力, 提高了行人重识别的准确度。

关键词 机器视觉; 行人重识别; 残差网络; 多层级特征; 相似性度量

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP57.081503

Person Re-Identification Based on Multi-Layer Feature

Liu Kewen^{1,2}, Fang Panpan^{1,2}, Xiong Hongxia^{3*}, Liu Chaoyang⁴,

Ma Yuan^{1,2}, Li Xiaojun^{1,2}, Chen Yalei^{1,2}

¹ School of Information Engineering, Wuhan University of Technology, Wuhan, Hubei 430070, China;

² Hubei Key Laboratory of Broadband Wireless Communication and Sensor Networks,
Wuhan University of Technology, Wuhan, Hubei 430070, China;

³ School of Civil Engineering and Architecture, Wuhan University of Technology, Wuhan, Hubei 430070, China;

⁴ State Key Laboratory of Magnetic Resonance and Atomic Molecular Physics, Wuhan Institute of
Physics and Mathematics, Chinese Academy of Sciences, Wuhan, Hubei 430071, China

Abstract To address the issue that existing person re-identification (Re-ID) algorithms have low robustness and discriminative capability when extracting pedestrian features with information loss, a novel Re-ID algorithm based on residual neural network is proposed for extracting multi-layer features of pedestrian images. During training phase, the residual network is used to extract the phase features after the four convolutional residual modules, to compensate for the information loss. And the triple loss function is used to supervise training of each feature vector. During the similarity measurement phase, the feature similarity is calculated according to the four feature vectors, the similarity of each stage is calculated by the summation of mapping function, and then the result of the summation is used to perform similarity matching. During the experiment, we validate the proposed algorithm on the Market-1501 and DukeMTMC-ReID datasets. The accuracy (Rank-1) of our algorithm reaches 91.7% and 84.9% and mean average precision (mAP) reaches 86.8% and 80.7%, respectively. Experimental results show that the multi-layer features extracted by our algorithm have considerable robustness and discriminative capability, which improves the accuracy of Re-ID.

Key words machine vision; person re-identification; residual network; multi-layer feature; similarity measurement

收稿日期: 2019-06-26; 修回日期: 2019-07-31; 录用日期: 2019-09-20

基金项目: 国家重点研发计划(2018YFC0115000)

* E-mail: xionghongxia@whut.edu.cn

1 引言

行人重识别^[1]技术主要解决跨摄像头、跨场景下行人的识别与检索。行人重识别是许多监控和安防应用中的主要任务^[2],姿态差异、视角变化、光照和遮挡等问题^[3]使行人图像变化巨大,因此学习到更有效的特征成为行人重识别的研究热点。

传统的行人重识别方法利用的是手工设计的图像特征^[4-5],然而手工特征描述能力有限,难以适应复杂场景下大数据集任务。随着深度学习的发展,利用神经网络模型自动学习复杂特征^[6-7],极大地提升了行人重识别的性能。文献[8-9]采用卷积神经网络(CNN)从行人图像中直接学习特征表达,采用马氏距离等通用度量方式来计算两张行人图像之间的距离,这种模型在CUHK03、Market-1501等相对较大的数据集上被证明非常有效,相比于传统的方法行人重识别性能有了很大提升。但是这些方法基于全局特征,会忽略一些局部信息,随着行人数据集越来越复杂,仅使用全局特征并不能达到要求,因此为了进一步挖掘细粒度特征,提取局部特征成为了研究热点。常见的提取局部特征的思路包括图片分块以及骨架关键点定位等。文献[10]将图片从上到下均分为六等份,分割好的若干图像块按照顺序被送入长短时记忆网络,最后得到融合了所有图像块信息的局部特征;该方法虽然获得了信息丰富的局部特征,但是缺点在于该方法要求图片对齐,若两幅图片A、B存在误对齐问题,就很可能出现将图片A的头部特征和图片B的上半身特征进行对比的情况,使得模型判断出错。文献[11]根据人体姿态估计模型提取的14个关键点将人体区域划分为7个部位,将7个部位图片和原始图片输入到同一个CNN网络,最终得到一个全局特征和多个局部特征融合的特征。文献[12]利用提取的关键点将图片划分成头部、上半身和下半身三部分,融合全图和三个局部图片提取到的特征。文献[11-12]利用骨架关键点模型进行图像分块,解决了行人误对齐的问题,但是该过程需要引入额外的姿态估计模型,严重影响特征提取的效率,增加系统开销。文献[13]将图片水平划分为6个区域,对每个区域分别进行监督,得到判别力很强的模型,提取有辨别力的局部特征,同时又引入RPP(refined part pooling)部件对齐池化,高效地实现了行人图像对齐。

上述方法均是通过提取丰富的特征来提高准确

率,主要关注全局特征或局部特征,是目前的主流方法。根据CNN的性质,不同层次的特征旨在编码不同层次的信息。高层特征更关注语义信息,而低层特征包含更多细节信息^[14]。上述算法基本采用高层特征查询行人图像,可能会导致信息丢失,影响行人重识别的准确度。

为解决特征提取过程中信息丢失问题,本文提出了一种提取多层次特征的行人重识别方法。该方法利用残差卷积网络,在网络的4个卷积残差模块之后提取阶段特征,分别利用三元组损失函数进行监督训练,在相似性度量阶段,分别针对4个特征向量计算相似度,并通过映射函数求和的方式得到最终的相似性分数,以此作为排序依据。对本文方法在目前使用最广泛的数据集Market-1501^[15]和DukeMTMC-ReID^[16]上进行了有效性验证,仿真结果表明本文算法提高了行人重识别的准确度。

2 基本原理

2.1 网络结构

本文方法采用残差网络结构Resnet50^[17]作为基础网络。残差网络通过跨层连接,构造了残差模块(图1),通过将输入信息绕道传到输出模块,保护了信息的完整性,同时保证网络性能不会随着网络层数的加深而变差,解决了深度网络的退化问题。Resnet50包含了1个独立卷积层、1个最大池化层、4个不同的卷积残差模块、1个平均池化层和1个全连接层。图1中 x 表示输入, $F(x)$ 表示残差函数。

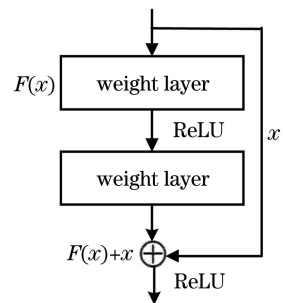


图1 残差模块

Fig. 1 Residual module

本文采用的网络结构如图2所示,在Resnet50的基础上进行如下改进:1)移去了平均池化层和全连接层;2)将每个卷积残差模块的输出作为阶段特征,并将其输入到一个深度监督网络(DSN)。每个DSN网络的结构包括:全局平均池化层

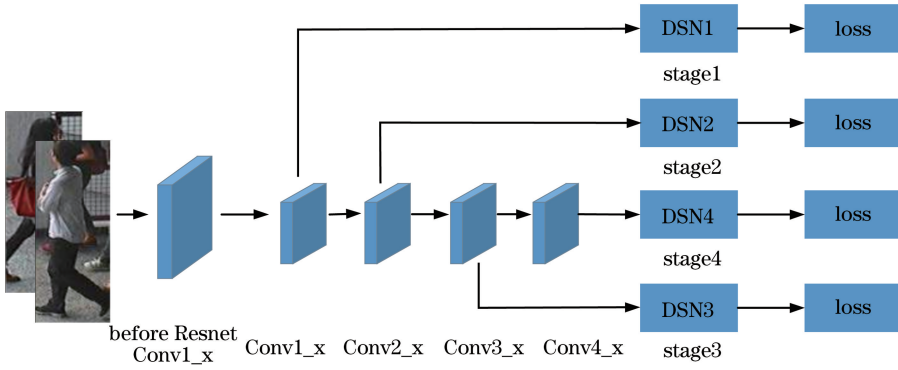


图2 网络结构

Fig. 2 Network structure

(GAP), $d_1 \times d_2$ 维的全连接层 (FC1), Batch Normalize 层, ReLU 层, 全连接层 FC2。其中 FC1 中 d_1 为 Resnet 的激励信号维度, d_2 为特征向量的维度, FC2 为 128 维。同时采用三元组损失函数进行网络优化, 网络的其他参数和 Resnet50 保持一致。

2.2 基于残差网络的多层级特征提取

本文提出了一种基于残差网络的多层级特征提取方法, 以提取不同网络层次的特征, 包含底层的细粒度特征和高层的语义特征, 使得对行人的特征描述更加丰富。根据 CNN 的性质, 随着网络层数的增加, 提取到的特征从不同类型的边缘、亮度等特征到纹理特征再过渡到抽象的语义特征。因此利用深层 CNN 提取特征, 会忽视图像的细节纹理信息, 影响行人重识别的精度。如图 2 所示, 本文方法提取出 Resnet50 4 个阶段的特征向量, 为了保持特征维度的一致性, 该特征向量需经过全局平均池化层和全连接层的处理。然后利用损失函数分别对 4 个阶段特征进行监督训练, 优化网络。

2.3 相似性度量

假设两张输入图片 I_1 和 I_2 的特征向量为 \mathbf{x}_1 和 \mathbf{x}_2 , 为了描述特征空间中特征向量的相似度, 需要定义一个距离度量函数, 同时, 为了实现端到端训练, 度量函数应尽可能连续可导, 因此选择余弦距离作为度量函数:

$$d_{I_1, I_2} = 1 - \frac{\mathbf{x}_1 \cdot \mathbf{x}_2}{\|\mathbf{x}_1\|_2 \|\mathbf{x}_2\|_2}, \quad (1)$$

式中: d_{I_1, I_2} 表示两个特征之间的余弦距离, 其取值范围为 $[0, 1]$ 。

2.4 多层特征融合

相似性度量计算式为

$$d = \sum_{s=1}^4 d_s, \quad (2)$$

式中: d_s 表示两张图像第 s 阶段特征的相似度; d 表示各阶段特征相似度总和。但是多层特征相似性直接求和的方式会导致受高层特征相似性的影响, 低层特征相似性不能在区分细粒度方面起到重要作用。为了解决这个问题, 采用文献[18]提出的映射函数, 更多的利用低层特征来测量图像之间的细粒度相似性:

$$d = \sum_{s=1}^4 f_s(d_s), \quad (3)$$

式中: f_s 为映射函数, 可表示为

$$f_s(d_s) = \begin{cases} t_s + (1 - t_s) \left(\frac{d_s - t_s}{1 - t_s} \right)^p, & d_s > t_s \\ d_s, & d_s \leq t_s \end{cases}, \quad (4)$$

式中: t_s 表示第 s 阶段特征相似性的阈值; p 的取值范围是 $(0, 1)$ 。

f_s 的函数图像如图 3 所示, 显示了映射函数与固定阈值 t_s 和不同 p 的关系。该映射函数旨在降低大于阈值 t_s 的相似性之间的差异, 范围 $[t_s, 1]$ 中的高层特征相似度将被削弱, 因此可以更好地区分低层特征的细节信息。在仿真中, 高层特征使用较低的阈值, 满足 $t_1 > t_2 > t_3 > t_4$, 其中最低层的阈值 t_1 设置为 1。指数 p 控制弱化程度, 特别地, 当 p 设置为 0 时, 本文方法将会变为使用低层相似性重新排序高层排序结果。当 p 设置为 1 或者 t_s 设置为 1 时, 本文方法转换为各层特征直接级联。该映射函数是挖掘不同层特征互补强度的关键。

2.5 损失函数

三元组损失可以有效地解决类间相似、类内差异问题, 因此被广泛应用在行人重识别领域。本文使用文献[19]提出的基于难样本挖掘的三元组损失, 能够有效提高网络的泛化能力。三元组损失函数定义式为

$$l = \sum_{p=1}^P \sum_{k=1}^K \ln \{ 1 + \exp \{ \max_{a=1,2,\dots,K} D[\phi(x_p^k), \phi(x_p^a)] - \min_{\substack{q=1,2,\dots,P \\ b=1,2,\dots,K \\ q \neq p}} D[\phi(x_p^k), \phi(x_q^b)] \} \}, \quad (5)$$

式中: P 为类别数目; K 为每个类别的图像数目; $\phi(x_p^k)$ 表示 p 类中的第 k 张图像的特征; $D(\cdot, \cdot)$ 表示两个特征向量之间的余弦距离。

整体网络的损失函数采用简单的加权求和方法进行求解:

$$l_{\text{all}} = \sum_{s=1}^4 \omega_s l_s, \quad (6)$$

式中: l_s 表示第 s 阶段特征的损失函数; ω_s 表示第 s 阶段损失函数的权重, 该权重是调试出来的参数; l_{all} 表示 4 个阶段特征的损失函数之和。

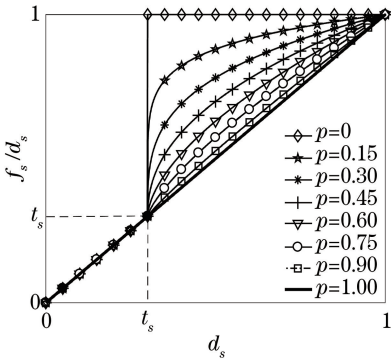


图 3 映射函数的图像

Fig. 3 Graph of mapping function

3 仿真结果分析与讨论

3.1 数据集及评价指标

为了验证本文方法的有效性, 在 Market1501 和 DukeMTMC-ReID 两个数据集上进行仿真。Market-1501 数据集包含从 6 个不同的摄像头拍摄到的 1501 个行人, 共 32668 张行人图像[由 DPM (deformable parts model)^[20]检测得到], 其中: 训练集包含 751 个行人, 共 12936 张图像; 测试集包含 750 个行人, 共 19732 张图像。DukeMTMC-ReID 数据集包含从 8 个不同的摄像头拍摄到的 1404 个行人, 共 34183 张图像, 其中: 训练集包含 702 个行人, 共 16522 张图像; 测试集包含 702 个行人, 共 17661 张图像。这两个数据集包含了姿态变化、遮挡、光照变化等情形, 比较符合实际应用。

目前行人重识别方法普遍采用首中准确率 (Rank-1) 和平均匹配度 (mAP) 两种评价指标评估方法的性能。

3.2 仿真环境及参数设置

本文采用的硬件和软件仿真开发环境如表 1 所示。在训练过程中, 骨干网络采用在 ImageNet 上预训练过的 Resnet50 模型, 权重初始化已完成, 以减少训练时间。行人图片被调整为分辨率为 256 pixel × 128 pixel 的输入, 各层特征维度分别为 64, 128, 256, 512。训练批次大小为 72, 包含 18 个不同行人, 每个行人的图片数为 4。学习率初始值为 0.01, 迭代次数为 60000 次, 在 30000 次迭代后学习率呈指数衰减, 衰减系数为 0.01。优化器选择的是带动量系数的 Adam, 动量系数为 0.9。训练损失函数为 4 个阶段损失函数加权之和。

表 1 仿真平台参数

Table 1 Simulation platform parameters

Configuration	Parameter
Operating system	Windows 7, 64 bit
RAM (random access memory)	8G
CPU (central processing unit)	Intel Pentium G3260 3.3 GHz
GPU (graphics processing unit)	NVIDIA GTX 1070Ti
Software platform	Pytorch 1.1

3.3 单层特征仿真结果分析

为了研究各层特征对识别性能的影响, 本文单独使用各阶段特征, 计算其在 Market-1501 上 Rank-1 的结果, 如表 2 所示, 从表中数据可以看出, 第 3 阶段特征 Rank-1 要高于第 4 阶段, 原因是第 4 阶段的特征属于高层特征, 更多地描述语义信息, 导致细节信息丢失。

表 3 中方法 1 表示只使用第 1 阶段特征得到的结果, 方法 2 表示利用前 2 个阶段特征和 (2) 式得到的结果, 方法 3 表示利用前 3 个阶段特征和 (2) 式得到的结果, 方法 4 表示利用前 4 个阶段特征和 (2) 式得到的结果, 方法 5 表示将 4 个阶段特征直接连接得到融合特征的结果。可以发现, 前 4 种方法的准确率逐步提升, 证明了高层特征和低层特征具有互补性, 减少了信息丢失。方法 5 使用融合特征的结果达到了方法 3 的水平, 但是低于方法 4, 说明融合特征十分有效, 但是由于高层特征相似性的影响, 低层特征相似性在细节差异上不起重要作用。

表 2 不同阶段的仿真结果

Table 2 Simulation results of different stages

Stage	Rank-1 / %	mAP / %
Stage 1	61.0	69.9
Stage 2	81.8	71.3
Stage 3	89.1	74.9
Stage 4	88.7	73.2

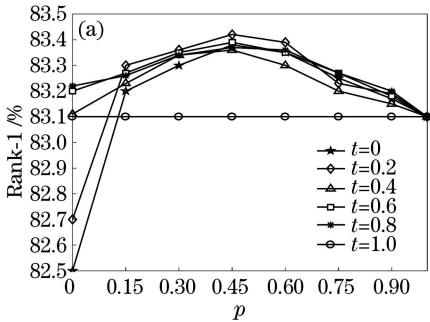
表 3 不同方法下的仿真结果

Table 3 Simulation results of different methods

Method No.	Rank-1 / %	mAP / %
1	61.0	69.9
2	83.4	72.6
3	89.7	77.9
4	90.4	84.6
5	89.8	83.7

3.4 多层特征融合仿真结果分析

根据(4)式,首先设置固定阈值 t_s 和指数 p 。为了便于解释,选择第 3 阶段特征和第 4 阶段特征来分析阈值 t_s 和指数 p 的影响,如上所述,最低层



的阈值设置为 1,因此直接使用第 3 阶段特征的相似性,在这种情况下,映射函数只对第 4 阶段特征的相似性起作用。特别地,当第 4 阶段特征的阈值设置为 1 时,曲线展示的是特征级联的效果。图 4 展示了在数据集 Market-1501 和 DukeMTMC-ReID 上不同 t 和 p 时 Rank-1 的趋势。图中 t 表示 t_s 的取值,通过实验确定最佳阈值 t_s 。

通过图 4 可以发现,采用映射函数适当削弱高层特征的相似性是有效的。虽然不同的设置会导致改进效果产生差异,但大多数设置对两个数据集的结果是有提升的。同时,实现最佳改进的参数 t 和 p 在两个数据集中是相似的,最佳阈值 t_s 设置为 0.2 并且 p 为 0.45。

针对 4 个阶段特征,同样设置最低层特征中 $t_1=1$,第 2 阶段特征中 $t_2=0.4$,由于上述已测试第 3 阶段和第 4 阶段的特征,因此将两者组合后设置 $t_3=0.45$, p 均为 0.45,在数据集 Market-1501 和 DukeMTMC-ReID 上测试的结果如表 4 所示。

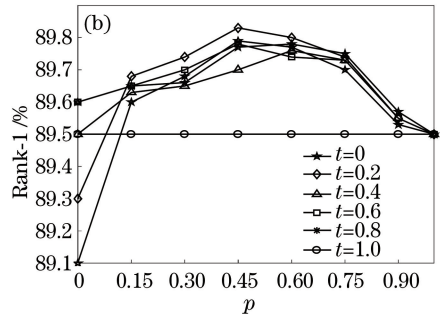


图 4 数据集 Market-1501 和 DukeMTMC-ReID Rank-1 结果。(a) Market-1501; (b) DukeMTMC-ReID
Fig. 4 Rank-1 results of Market-1501 and DukeMTMC-ReID. (a) Market-1501; (b) DukeMTMC-ReID

表 4 多层特征融合仿真结果

Table 4 Simulation results of multi-layer feature fusion

Method	Market-1501		DukeMTMC-ReID	
	Rank-1	mAP	Rank-1	mAP
Method 4	90.4	84.6	84.6	80.2
Method 5	89.8	83.7	84.4	80.0
Propoed method	91.7	86.8	84.9	80.7

表 4 中方法 4 和 5 指的是 3.3 节中的方法 4 和 5,可以发现本文采用映射函数参与相似性计算是有效的,相比于直接使用各层特征或者多层特征级联的方式,本文方法的性能是略有提升的,这表明了高层特征和低层特征具有互补关系,同时利用高层特征的语义信息和低层特征的细节纹理信息,并使用映射函数可以减弱高层特征的影响,提高行人重识别性能。

3.5 测试运行时间

将所提算法的运行速度与 GLAD^[12]、ThriNet^[19]、DaRe^[21]等主流算法进行对比,主要测试在 Market-1501 测试集上运行所花费的时间。Market-1501 测试集包含 750 个行人,共 19732 张图片。为保证对比的公平性,本文在相同的硬件条件下对上述几种主流算法进行了复现,仿真的硬件环境参数如表 1 所示。本文设定所有算法的网络参数设置与原论文相同,均为最优,并且进行充分训练,利用训练好的模型在测试集上进行测试。表 5 给出了各算法的测试时间以及单张图片的匹配时间。

Test time 表示待查询图片与图像检索库中所有图片匹配花费的时间,Per match 表示待查询图片与检索库中一张图片匹配花费的时间。从表 5 中数据可以发现,所提算法的运行速度比大部分算法

表5 算法运行时间对比结果

Table 5 Comparative results of running time of different methods

Method	Time /s	
	Test time	Per match(19732)
GLAD ^[12]	368	0.0186
ThriNet ^[19]	343	0.0173
DaRe ^[21]	315	0.0159
Proposed method	328	0.0166

快,但是比 DaRe 算法略慢,牺牲少量的运行时间提高了性能。在实际监控条件下,行人图像检索库是随着行人在监控摄像头中出现的时间慢慢增加的,因此在匹配的时候只需要考虑新增加的行人图片,本文方法只匹配 1 张图片所用时间仅为 0.0166 s,足以满足现实监控需求。

3.6 与主流方法的比较

为了验证所提方法的有效性,将本文方法与目前主流的 5 种行人重识别方法进行对比,这几类方法均是基于 CNN 提取特征,对照方法的网络参数均设置为最优,仿真结果如表 6 所示。从表 6 中数据可以看出,本文方法在 Market-1501 和 DukeMTMC-ReID 两个数据集上的 Rank-1 和 mAP 均超过了对比方法,这表明本文方法的有效性。

表6 本文方法与主流方法的性能比较

Table 6 Performance comparison between proposed algorithm and mainstream methods %

Method	Market-1501		DukeMTMC-ReID	
	Rank-1	mAP	Rank-1	mAP
SVDNet ^[22]	82.3	62.1	76.7	56.8
MultiScale ^[23]	88.9	73.1	79.2	60.6
GLAD ^[12]	89.9	73.9	—	—
ThriNet ^[19]	84.9	69.1	—	—
DaRe ^[21]	90.8	85.9	84.4	79.6
Proposedmethod	91.7	86.8	84.9	80.7

4 结 论

提出了一种利用残差网络提取多层次特征的行人重识别方法,首先根据 CNN 的性质并利用残差网络 4 个残差模块提取 4 个阶段特征,其次利用映射函数求和在相似性度量阶段对各个特征相似度进行相似度匹配。本文方法强化了各层次特征之间的互补性,突出了低层特征的细节信息,解决了信息丢失问题。仿真结果表明,本文方法在数据集 Market-1501 和 DukeMTMC-ReID 上的性能与目前主流方法相比有一定提升。

参 考 文 献

- [1] Liao S C, Hu Y, Zhu X Y, et al. Person re-identification by Local Maximal Occurrence representation and metric learning [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 2197-2206.
- [2] Sun Y. Research on person reidentification technology based on deep feature in surveillance video [D]. Xiamen: Xiamen University, 2017.
孙妍. 基于深度特征的监控视频下的行人重识别技术研究[D]. 厦门: 厦门大学, 2017.
- [3] Zhu X B, Che J. Person re-identification algorithm based on feature fusion and subspace learning [J]. Laser & Optoelectronics Progress, 2019, 56(2): 021503.
朱小波, 车进. 基于特征融合与子空间学习的行人重识别算法[J]. 激光与光电子学进展, 2019, 56(2): 021503.
- [4] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C] // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), June 20-25, 2005, San Diego, CA, USA. New York: IEEE, 2005: 8588935.
- [5] Sun X W, Xu Q S, Cai Y, et al. Sea sky line detection based on edge phase encoding in complicated background[J]. Acta Optica Sinica, 2017, 37(11): 1110002.
孙熊伟, 徐青山, 蔡熠, 等. 基于边缘相位编码的复杂背景下海天检测[J]. 光学学报, 2017, 37(11): 1110002.
- [6] Cheng D, Gong Y H, Zhou S P, et al. Person re-identification by multi-channel parts-based CNN with improved triplet loss function [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 1335-1344.
- [7] Chen B, Zha Y F, Li Y Q, et al. Person re-identification based on convolutional neural network discriminative feature learning [J]. Acta Optica Sinica, 2018, 38(7): 0720001.
陈兵, 查宇飞, 李运强, 等. 基于卷积神经网络判别特征学习的行人重识别[J]. 光学学报, 2018, 38(7): 0720001.
- [8] Wu L, Shen C, Hengel A. PersonNet: person re-identification with deep convolutional neural networks

- [J/OL]. (2016-06-20)[2019-06-25]. <https://arxiv.xilesou.top/abs/1601.07255>.
- [9] Ahmed E, Jones M, Marks T K. An improved deep learning architecture for person re-identification[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 3908-3916.
- [10] Varior R R, Shuai B, Lu J W, et al. A Siamese long short-term memory architecture for human re-identification[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9911: 135-153.
- [11] Zhao H Y, Tian M Q, Sun S Y, et al. Spindle Net: person re-identification with human body region guided feature decomposition and fusion[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 907-915.
- [12] Wei L, Zhang S, Yao H, et al. Glad: global-local-alignment descriptor for pedestrian retrieval [C] // Proceedings of the 25th ACM International Conference on Multimedia, October 23-27, 2017, Mountain View, California, USA. New York: ACM, 2017: 420-428.
- [13] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[M]// Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 112008: 501-518.
- [14] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[J]. Computer, 2013, 8689: 818-833.
- [15] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1116-1124.
- [16] Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking[M]// Hua G, Jégou H. Computer vision-ECCV 2016 Workshops. Lecture notes in computer science. Cham: Springer, 2016, 9914: 17-35.
- [17] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [18] Yu W, Yang K Y, Yao H X, et al. Exploiting the complementary strengths of multi-layer CNN features for image retrieval[J]. Neurocomputing, 2017, 237: 235-241.
- [19] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification [J/OL]. (2017-11-21)[2019-06-25]. <https://arxiv.xilesou.top/abs/1703.07737>.
- [20] Felzenszwalb F, Girshick B, McAllester D, et al. Object detection with discriminatively trained part-based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32 (9): 1627-1645.
- [21] Wang Y, Wang L Q, You Y R, et al. Resource aware person re-identification across multiple resolutions [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 8042-8051.
- [22] Sun Y F, Zheng L, Deng W J, et al. SVDNet for pedestrian retrieval [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 3820-3828.
- [23] Chen Y B, Zhu X T, Gong S G. Person re-identification by deep learning multi-scale representations [C] // 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 2590-2600.