

引文格式:

江伟忠, 余海铭, 陈世峰. 基于特征点相关性的行人重识别方法 [J]. 集成技术, 2020, 9(4): 68-78.

Jiang WZ, Yu HM, Chen SF. Person re-identification method based on correlation between features [J]. Journal of Integration Technology, 2020, 9(4): 68-78.

基于特征点相关性的行人重识别方法

江伟忠^{1,2} 余海铭³ 陈世峰¹

¹(中国科学院深圳先进技术研究院 深圳 518055)

²(中国科学院大学深圳先进技术学院 深圳 518055)

³(爱奇艺公司 北京 100010)

摘 要 目前行人重识别算法面临的主要问题包括背景过多、行人区域缺失及图片视角差异等。基于行人区域中显著性特征之间存在着强相关性及行人区域与背景区域特征之间存在着弱相关性两方面的观察, 该研究提出一种基于特征点相关性的行人重识别方法。其中, 通过采用一种基于视觉不变性与弱检测的上下文信息处理模块, 即 CIP (Contextual Information Processing) 模块实现该方法。由于具有强相关性的特征可能分布在不同的粒度之间, 所以嵌入 CIP 模块的多粒度行人重识别方法可以融合粒度之间的相关性信息。实验中, 通过以第一配准率 (Rank-1) 和平均准确率为评价指标, 使用单数据集测试、跨数据集测试两种方法, 在 4 个大型的行人重识别数据集上进行了充分的测试实验, 并利用两个可视化的方法——弱检测效果与行人区域中显著特征点的相关性效果, 对 CIP 模块的有效性进行验证。实验结果显示, 目前主流的多粒度网络通过嵌入 CIP 模块, 性能均有明显的提升。

关键词 行人重识别; 特征相关性; 全局信息

中图分类号 TG 156 文献标志码 A doi: 10.12146/j.issn.2095-3135.20200519001

Person Re-identification Method Based on Correlation Between Features

JIANG Weizhong^{1,2} YU Haiming³ CHEN Shifeng¹

¹(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

²(Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518055, China)

³(iQIYI Inc., Beijing 100010, China)

Abstract Person re-identification suffers from some problems such as confusion with excessive background, incomplete targets, and different viewing point etc. There are two basic observations for most person re-

收稿日期: 2020-05-19 修回日期: 2020-06-15

基金项目: 广东省重点领域研发计划项目 (2019B010155003); 国家自然科学基金项目 (U1713203); 深圳市孔雀技术创新项目 (KQJSCX20180330170238897); 中科院科研仪器设备研制项目 (YJKYYQ20190028)

作者简介: 江伟忠, 硕士研究生, 研究方向为计算机视觉、行人重识别; 余海铭, 硕士, 研究方向为计算机视觉、深度学习; 陈世峰(通讯作者), 博士研究生导师, 研究方向为计算机视觉、图像处理, E-mail: shifeng.chen@siat.ac.cn.

identification applications, i.e., strong correlation exists between the discriminative features, and weak correlation exist between feature points of the pedestrian areas and the background areas. Based on such observations, a person re-identification method based on features correlation is proposed in this paper. The CIP (contextual information processing) module based on viewpoint invariance and soft-detection is applied to realize the proposed method. Since strong correlative features distribute at different granularities, the multi-granularity based person re-identification methods can describe relationship between granularities by embedding the CIP module. The experiments are implemented on four large-scale person re-identification data sets. Both single-domain and cross-domain tests are used in the experiments. The Rank-1 and mean average error criterion are used as the evaluation indicators. As the experiment shows, the proposed method enhances the identification performance of several mainstream multi-granularity methods by CIP module.

Keywords person re-identification; correlation between features; contextual information

1 引言

在智能安防、智能监控和疑犯追踪等领域中,首要工作是判断某个行人是否出现在不同摄像机的视域中。然而目前普遍使用人工进行判断的方法不仅需要耗费大量的人力、物力,而且准确率较低^[1]。随着社会需求逐渐增加以及计算机视觉技术日渐成熟,行人重识别技术已成为当前研究的一个热点方向。其中,行人重识别是一种旨在识别不同摄像机视域中是否出现某个特定目标行人的技术^[2]。

传统的行人重识别主要关注设计手工视觉特征与度量相似性两个问题。随着深度学习的发展,分类、检测、识别和检索等计算机视觉子领域都取得了巨大的突破^[3]。将深度学习应用于行人重识别领域也成为了一种必然的趋势。基于深度学习的方法,研究人员对行人重识别进行了大量、更深入的研究^[4]。与传统方法相比,基于深度学习的行人重识别方法可以自动提取特征,更加高效和精准^[4]。然而,行人重识别还存在着诸多的挑战,如遮挡、模糊、拍摄视角不同和检测算法性能不佳等^[2]。这些问题都可能导致行人背景过多或行人区域缺失。针对行人背景过多以及

背景干扰的问题, Song 等^[5]采用二进制分割掩模去除了行人背景的干扰,并以完整图像、身体和背景三个部分的特征连接三元组损失函数强制模型学习到对杂乱背景鲁棒的特征。Xu 等^[6]创建了基于姿态引导下的区域注意力机制方法,以此减少背景噪声的干扰。行人姿态多变也是行人重识别面临的极大挑战,因此, Zheng 等^[7]通过构造 PoseBox,将姿态不变嵌入法作为行人描述子,消除了行人姿态多变对行人重识别的影响。若仅仅利用行人重识别的全局特征,行人重识别模型无法捕捉到足够的差异性特征,而引入局部特征是提高行人重识别性能的一种思路^[8]。Kalayeh 等^[9]利用语义分割的方法引导模型从行人身体不同区域提取了局部特征。目前,对行人进行分块并搭建一个多粒度网络使模型更加关注局部区域是最为常见的思路。Sun 等^[10]提出一种基于分块的卷积基准网络(Part-based Convolutional Baseline, PCB),通过将行人均等分为六分块引导网络更加关注局部粒度信息。Huang 等^[11]提出一种基于分块对齐池化(Part Aligned Pooling, PAP)方法的多粒度网络,提升了模型在跨数据集测试的性能。Su 等^[12]利用一种姿态估计将行人分块并以特征权重网络给不同

区域赋予不同的权重。Wang 等^[13]也提出一种将判别信息与各粒度相结合的多粒度网络 (Multiple Granularity Network, MGN)。

但目前性能最好的行人重识别方法依然存在不少问题。例如,若引入姿态估计以及分割的方法,则会引入额外数据集的先验;若不借助姿态估计或分割等方法,行人重识别算法的性能则会受到行人图片中背景的干扰^[5]。同时,出现部分行人区域缺失的问题也没能得到较好的解决。另外,大部分多粒度网络在不引入额外数据集的情况下,最常用的分块方法是先横向均等分块,再对每一块所提取的特征进行对比,如头部跟头部对比。但在图片中存在背景过多或部分行人区域缺失时,将会出现头部与背景对比的情况,这导致了行人重识别性能急剧下降。并且,多粒度模型也忽视了粒度特征之间的相关性。为解决上述问题,研究像素点或者特征之间的相关性对行人重识别具有极大的意义。其中,在行人重识别网络模型中引入背景区域特征与行人区域特征之间的弱相关性信息,可实现行人的弱检测效果,这避免了背景区域的干扰。在行人重识别网络模型中引入行人区域特征点之间的强相关性信息,则可实现特征点缺失时图像信息相对完整,这将有效地解决行人区域部分缺失或遮挡等问题。此外,如果特征点相关性在不同视角下具有不变性,那么将对解决行人姿态多变有所帮助。

在图像相邻邻域的相关性研究中,特征相关性普遍存在于文本与图像中,表示一个或多个特征与其他特征之间存在一定的依赖关系。在文本分析领域中,Deerwester 等^[14]在 1990 年提出了经典的浅层语义分析 (Latent Semantic Analysis, LSA) 方法,通过使用词-文档矩阵来描述词与文档之间的依赖关系。此后 Hofmann^[15]提出概率潜在语义分析 (Probabilistic LSA) 模型对 LSA 进行改进,利用概率生成模型对文本进行话题分析。由于相邻领域存在的相关性也普遍存在于图像

中,故概率潜在语义分析也被大量应用于图像分类、检测、检索等领域^[16-17]。图像像素存在一定相关性最早体现在灰度共生矩阵中,它描述了相邻像素之间的灰度存在较强依赖性的现象^[18]。这种相关性不仅存在于相邻邻域,在图像的不同区域也可以存在较强相似性,如行人的两只鞋子区域之间、两只手所在区域之间、两条腿所在区域之间都应该存在较强的相关性。

本文的整体设想及分析部分借鉴了 Wang 等^[19]提出来的非局部神经网络 (Non-local Neural Networks)。其中,Non-local 网络结构可以模块化特征图中任意两个特征点之间的关系。本文借鉴 Non-local 网络的思路设计了 CIP 模块,并将 CIP 模块组成的网络称为 Contextual Information Processing Network (CIPN)。CIPN 能够捕获视角不变性的特征点关系并具有弱检测功能。在嵌入 CIPN 后,行人重识别模型将更加关注行人区域以及粒度之间的关系,使得基于分块的行人重识别模型的性能得到显著的提升。本文在行人重识别的 4 个数据集上进行了充分的测试实验,并使用两个评估指标验证了该模块的有效性。同时,本文通过可视化工作呈现了弱检测效果及显著特征点的相关性效果,验证了该模块的有效性机理。

2 研究方法

2.1 基于深度学习的行人重识别网络结构

本文的行人重识别网络结构基本可以分为特征提取网络、粒度全连接网络、能够捕获视角不变性的特征点关系并具有弱检测功能的 CIPN 三部分,具体如图 1 所示。其中,特征提取网络为用于提取行人基础特征的基础网络,如 VGG^[20]、ResNet^[21]、DenseNet^[22]等。考虑到目前大部分方法普遍使用 ResNet50 作为特征提取网络,本文在图 1 中也使用 ResNet50 作为特征

提取网络图示。

粒度全连接网络主要包括分块、全局最大池化、降维全连接层和分类全连接层。按照分块方式, 粒度全连接网络可分为单粒度和多粒度全连接网络。单粒度全连接网络是对特征提取网络输出的整个特征进行全局最大池化操作, 从而得到一个全局的单粒度行人特征向量。而多粒度全连接网络则是对特征提取网络输出的特征先进行分块, 再对每一块进行全局最大池化操作, 从而得到多个局部的粒度行人特征向量, 如图 1 的粒度全连接网络为三分块粒度全连接网络。目前大部分基于分块的行人重识别网络粒度数量从六个到十几个不等。一般在不引进新数据集的情况下, 分块方法普遍使用横向均等分块。降维全连接层可以对全局最大池化操作所得到的初步行人特征向量进行降维, 得到降维后的粒度行人特征向量。特征降维后, 在各分支后面分别接上分类全连接层, 即可组成完整的粒度全连接网络。其中, 降维全连接层的维度与具体模型相关, 分类全连接层的维度与数据集类别数相关。

CIPN 连接在 ResNet50 中的 layer 3 后面。本

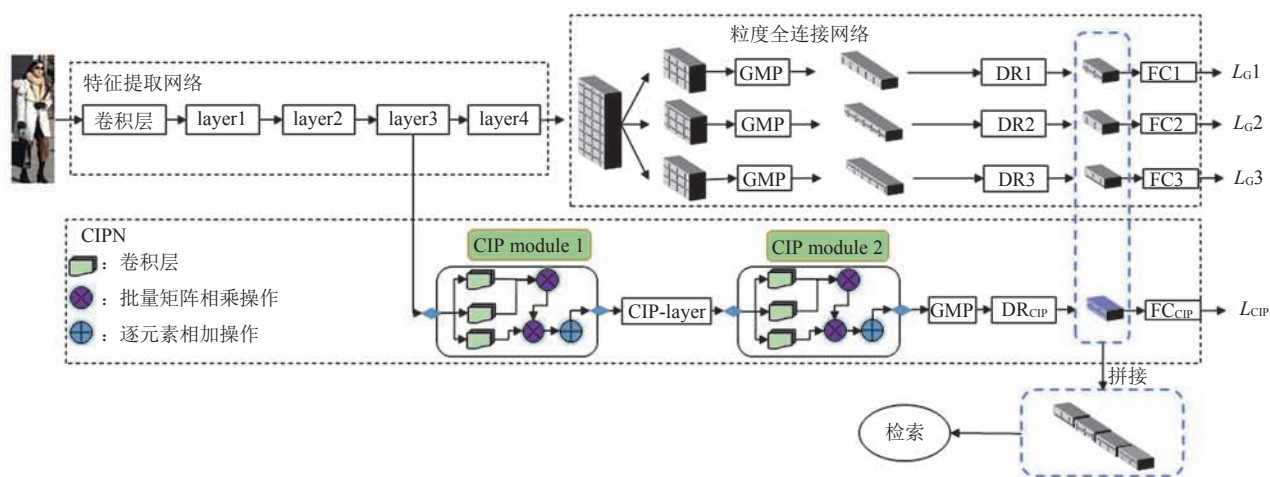
文的 CIPN 由两个 CIP 模块(图 1 分别表示为 CIP module 1 以及 CIP module 2)和连接在中间的残差 layer(图 1 表示为 CIP-layer)组成。CPlayer 的结构与 ResNet50 的第 4 个卷积块堆叠层一样, 但卷积核的步长为 1。ResNet50 的 layer 3 输出的浅层特征先输入到第 1 个 CIP 模块, 再经过 1 个 CIP-layer 以及第 2 个 CIP 模块, 最后得到具有视角不变性与弱检测效果的初步行人特征。CIPN 后半部分连接的网络与粒度全连接网络一致: 先将初步行人特征输入 GMP 进行全局最大值池化操作, 然后连接一个降维全连接层得到降维后的 CIP 行人特征向量, 最后再连接一个分类全连接层。

2.2 CIP 模块的数学表示与模块实现

在特征图中, 任意两个点之间都存在一定程度的相关性。这种相关性可以借鉴 Non-local 的数学表达式:

$$y_i = \frac{1}{C(x)} \sum_{v_j} f(x_i, x_j) \cdot g(x_j) \quad (1)$$

其中, x 为输入的特征, y 为输出的特征, 两者的大小一致; $f(x_i, x_j)$ 为特征图位置 i 与位置 j 的相



注: GMP 表示全局最大池化 (Global Max Pooling); DR 表示降维全连接层 (Dimension Reduction Full Connect Layer); FC 表示分类全连接层 (Full Connect Layer); L_G 表示在粒度全连接网络后面各分支的损失函数; L_{CIP} 表示在 CIPN 后面的交叉熵损失函数

图 1 嵌入 CIP 模块的 ResNet50

Fig. 1 The ResNet50 with CIP modules

关性, $f(x_i, x_j)$ 值越大, 表示两个位置的相关性越强; $g(x_j)$ 表示位置 j 处的输入信号; $C(x)$ 为归一化因子。

CIP 模块的网络结构如图 2 所示。CIP 模块是对公式(1)进行模块化封装的网络结构, 用于学习任意两点之间的相关性。模块输入的特征图为 $X \in \mathbb{R}^{C \times W \times H}$, 其中 $\mathbb{R}^{C \times W \times H}$ 表示大小为 $C \times W \times H$ 的维度空间。首先采用 3 个 1×1 的卷积核 (q, k, v) 对 X 进行卷积, 并获得三组卷积后的特征图, 分别命名为 $Q \in \mathbb{R}^{\frac{C}{4} \times W \times H}$ 、 $K \in \mathbb{R}^{\frac{C}{4} \times W \times H}$ 、 $V \in \mathbb{R}^{C \times W \times H}$ 。其中, 卷积核 q 和 k 主要用于降低通道数, 减少计算量。接着分别对 Q 进行数据重构得到 $Q' \in \mathbb{R}^{HW \times \frac{C}{4}}$, 对 K 进行数据重构与转置操作得到 $K'^T \in \mathbb{R}^{\frac{C}{4} \times HW}$ ($N=W \times H$)。然后, 对 Q' 与 K'^T 进行矩阵相乘操作得到相似度矩阵 S 。最后, 将相似性矩阵 S 输入 Softmax 函数进行归一化可得 S' ($S' \in \mathbb{R}^{N \times N}$)。其中, S' 代表着特征图中每一个位置与其他所有位置之间归一化后的相关性。位置 i 与 j 的关系可以用矩阵 S' 的第 i 和第 j 列的交点值 $s'_{i,j}$ 来表示。 S' 的具体公式如下:

$$S' = \text{softmax}(K'^T \times Q') \quad (2)$$

对于 V 先进行数据重构得到 $V' \in \mathbb{R}^{C \times N}$ ($N=W \times H$), 用 T ($T \in \mathbb{R}^{C \times N}$) 表示 V' 与 S' 进行矩阵相乘所得矩阵:

$$T = V' \times S' \quad (3)$$

与 Non-local 网络的思路有所不同, CIP 模块中 v 没有对输入 X 的通道进行降维, 所以不需要对 T 进行升维, 这样整个网络构造相对 Non-local 网络可以减少一个升维卷积。因此, 只需要对 T 进行数据重构操作即可得到 T' ($T' \in \mathbb{R}^{C \times H \times W}$)。

CIP 模块的输出值 Y 由输入的特征 X 、 T' 和自适应调节因子 γ 组成, 具体如公式(4)所示:

$$Y = X \oplus \gamma T' \quad (4)$$

其中, \oplus 代表着对应元素相加。由公式(4)可以看出, Y 的任意一个元素值都可以看成是 X 各个元素值的显性加权组合。因此 CIP 模块学习到点与点之间的依赖性, 并将这种信息融合到行人特征中, 从而捕获了全文信息。

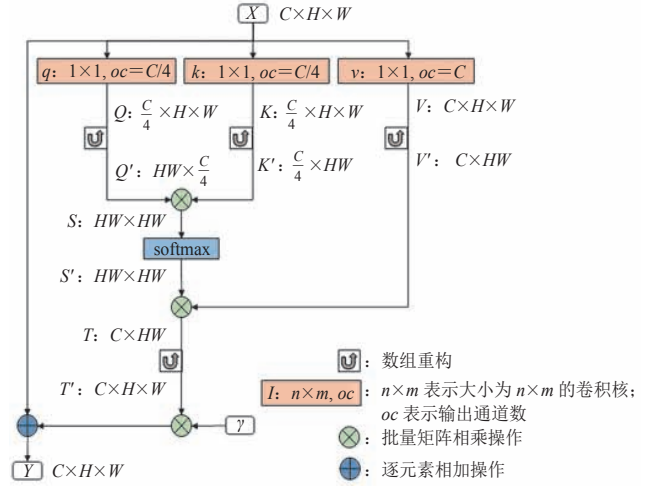


图 2 CIP 模块的图示

Fig. 2 Illustration of the CIP

2.3 损失函数与相似性度量

在粒度全连接网络后面各分支接上各自的损失函数(Granularity Loss), 表示为 L_G 。由于本文主要关注 CIPN 网络, 所以对粒度全连接网络后面各分支所使用的损失函数 L_G 不作更多的阐述。为了学习到更具有判别力的特征, 并使得整个训练简单高效, 本文在 CIPN 后面接上交叉熵损失函数, 表示为 L_{CIP} 。在行人重识别网络中, 交叉熵损失函数通过 Softmax 函数输出行人类别的预测值概率, 并以该预测概率与指示变量进行损失评估, 具体如公式(5)所示:

$$L_{CIP} = - \sum_{c=1}^N y(c) \log[p(c)] \quad (5)$$

其中, $c \in \{1, 2, \dots, N\}$ 表示行人类别数量; $p(c)$ 为所输入的图片属于 c 类的预测概率; $y(c)$ 属于指示变量。若所输入的图片属于 c 类, 则 $y(c)=1$; 若所输入的图片不属于 c 类, 则 $y(c)=0$ 。

总损失函数设置为 L_{CIP} 与各粒度损失函数 L_G

的总和, 具体如公式(6)如所示:

$$L_{\text{all}} = L_{\text{CIP}} + \sum_{i=1}^N L_{G_i} \quad (6)$$

在训练阶段, 通过不断最小化损失函数 L_{CIP} 训练 CIPN 分支, 使得 CIPN 能更好地学习到点与点的相关性。在测试阶段, 先将该分支的 CIP 行人特征向量与粒度全连接层所得到的粒度行人特征向量进行拼接, 得到最后的行人特征向量, 然后采用欧式距离进行相似度度量, 具体如图 1 所示。

3 实验

3.1 数据集与评估方法

为对本文提出的方法进行分析与评估, 选择 4 个常用的行人重识别的数据集 (Market-1501^[23]、CUHK03^[24]、DukeMTMC-ReID^[25]和 MSMT17^[26]) 进行实验, 各数据集的基本情况如表 1 所示。其中, MSMT17 是目前行人 ID 数量最多、图片数量最多、拍摄场景最为复杂和挑战性最大的数据集。数据集一般分为训练集和测试集, 测试集中包含查询图片 (Query) 与候选图片集 (Gallery)。

目前大部分行人重识别的相关研究采用两个评价标准: 第一配准率 (Rank-1) 和平均准确率 (Mean Average Precision, mAP)。其中, 第一配准率主要用于模型对简单样本配准性能的评估, 而平均准确率则兼顾了模型对困难样本配准性能的评估。

3.2 实验参数设置及实验过程

本文实验基于 EANet 的开源代码^[11], 使用 Pytorch 深度学习框架实现, 并在 NVIDIA GTX 1080Ti GPU 硬件上运行。在数据集处理方面, 所有的设置都与基础模型一致。如在 PAP^[11]与 PAP+CP 实验中, 仅采用垂直翻转进行数据增强, 所有的图片均被转化为宽 128 像素且高 256 像素。在训练参数设置方面, 本文使用随机梯度下降法作为优化算法, 冲量 (Momentum) 设置为 0.9, 批大小设置为 32, 权重衰减系数设置为 0.000 5, 总的训练周期 (Epoch) 数量为 60 个。在训练过程中, 先将预训练好的 ResNet50 (已在 ImageNet 上进行预训的模型) 参数固定, 然后采用 0.2 的学习率对新增加的层进行 1 个训练周期的微调。微调结束后, 开放网络所有的层, 对模型的所有参数训练 60 个训练周期。此时新增加的层的起始学习率为 0.2, 其他层的起始学习率为 0.1, 每 25 个训练周期的学习率下降为原来的 1/10。所有的损失函数均采用交叉熵损失函数。训练总损失为所有分支损失之和, 且每个分支的损失函数权重一致, 均设置为 1。在测试过程中, 批大小仍然为 32, 采用欧氏距离作为相似性度量方法。具体的实验流程如图 3 所示。

4 实验结果与分析

4.1 模块的有效性验证分析

为验证 CIP 模块的性能, 本文分别在 ResNet50、PCB^[10]、EANet^[11]和 MGN^[13]四个基

表 1 常用的行人重识别数据集

Table 1 Typical person re-identification datasets

数据集	场景	检测器	摄像头数量(个)	图片数量(张)	行人数量(人)
Market-1501 ^[23]	室外	DPM	6	32 668	1 501
CUHK03 ^[24]	室内	DPM, 人工标注	2	28 192	1 467
DukeMTMC-ReID ^[25]	室外	人工标注	8	36 411	1 812
MSMT17 ^[27]	室外, 室内	Faster RCNN	15	126 441	4 101

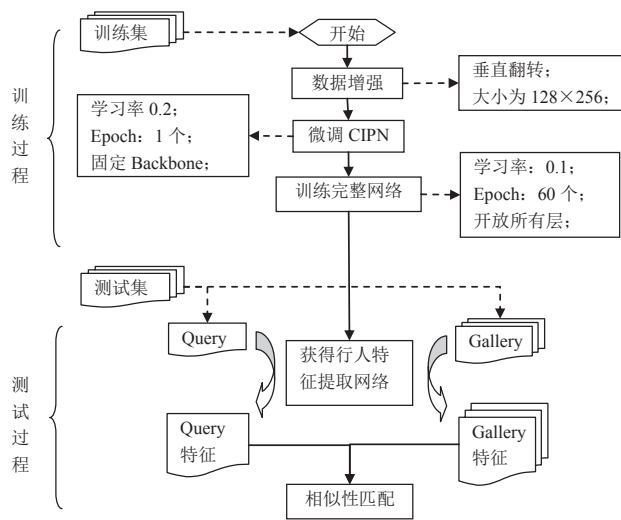


图3 实验流程

Fig. 3 Experiment flow chart

础模型中选择合适的模型加上 CIP 模块进行实验。其中, PCB、EANet 和 MGN 均为目前性能较好的多粒度网络。接入方式采用 CIPN 分支接入, 具体如图 1 所示。CIPN 分支提取的特征为 1 024 维的向量, 其他基础分支的维度因模型而异, 但需保持和基础模型一致。最终将所有分支得到的行人特征向量拼接后进行相似度匹配。本文分别在单数据集和跨数据集对 CIP 模型的有效性进行验证。

在单数据集测试中, 本文在 ResNet50、

PCB、PAP^[11]和 MGN 四个基础模型加上 CIP 模块, 并在 4 个数据集进行训练与测试。其中, 数据集 MSMT17 最为复杂, 更具有代表性。从表 2 可知, 加上 CIP 模块后, 所有的基础网络在 4 个数据集的 Rank-1 与 mAP 两个评价指标上都有明显的提升。CIP 模块对基础网络 ResNet50 的性能提升最为明显, 其中在 CUHK03 数据集上 mAP 提升了 7.3%, 在 MSMT17 数据集上 mAP 提升了 5.2%, 甚至在两项评价指标上已经接近多粒度网络 PCB 的性能。增加 CIP 模块对多粒度网络 PCB 的精度同样具有非常明显的提升, 其中在 MSMT17 数据集上 mAP 提升了 5.0%。然而, 从表 2 可看出, 随着模型分支的逐渐增多, CIP 模块对网络精度的提升作用越来越微弱。主要是因为增加的 CIP 分支随着网络分支总数的增加, CIPN 分支产生的特征对总特征的占比会越来越小。但即使对于 MGN 这种具备 3 个分支、8 个粒度的大型网络, MGN+CIP 模型在 4 个数据集上的各项评估指标也有小幅度的性能改善。

在跨数据集测试中, 本文在 ResNet50、PCB 和 PAP 三个基础模型加上 CIP 模块, 并在两个数据集进行训练与测试。训练测试结果如表 3、表 4 所示。从表 3 和表 4 可知, 加上 CIP 模块的 3 个网络在跨数据集的性能均有显著的提升, 表明

表 2 单数据集测试

Table 2 Single-domain tests

粒度类型	基准模型	数据集测试性能 (%)							
		MSMT17		Market-1501		CUHK03		DukeMTMC-ReID	
		Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
单粒度模型	ResNet50	70.7	42.0	88.1	73.1	45.1	40.7	79.4	61.6
	ResNet50+CIP	74.7	47.2	91.3	78.8	52.1	48.0	83.2	68.2
多粒度模型	PCB	74.0	47.7	93.2	81.1	65.2	60.0	86.3	72.7
	PCB+CIP	78.1	52.7	93.9	83.3	66.9	61.7	86.7	73.5
	PAP	79.2	52.9	94.4	84.5	72.0	66.2	86.1	73.3
	PAP+CIP	81.1	55.6	94.7	85.8	70.8	66.2	87.3	74.9
	MGN [*]	93.3	67.1	94.9	87.9	78.3	75.3	89.1	79.9
	MGN [*] +CIP	93.4	67.9	95.0	88.5	80.5	77.1	89.7	80.6

注: MGN^{*} 表示此处的 MGN 为本文所复现的模型

表 3 跨域数据集测试 1

Table 3 Cross-domain tests 1

粒度 类型	基准模型	数据集测试性能 (%)											
		MS→M		MS→C		MS→D		M→C		M→D		M→MS	
		R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP
单粒度	ResNet50	51.3	25.1	12.9	12.0	56.0	35.2	10.2	9.0	40.2	22.0	11.9	3.6
	ResNet50+CIP	55.7	27.8	15.3	14.1	60.7	38.9	12.2	10.2	42.6	23.8	13.8	4.5
多粒度	PCB	58.9	30.6	14.3	13.2	58.3	38.2	8.9	7.8	42.9	23.8	12.5	3.7
	PCB+CIP	61.5	32.7	15.0	14.0	62.8	42.3	11.1	9.0	45.3	26.7	14.7	4.6
	PAP	63.7	35.3	16.0	15.2	63.5	43.6	11.4	9.9	46.4	27.9	15.6	5.3
	PAP+CIP	64.9	36.3	17.6	16.1	66.5	45.8	12.8	11.3	49.5	30.8	18.4	6.2

注: $X \rightarrow Y$ 表示在数据集 X 进行训练, 在数据集 Y 进行测试; MS 代表数据集 MSMT17; M 代表数据集 Market-1501; C 代表数据集 CUHK03; D 代表数据集 DukeMTMC-ReID; R-1 代表 Rank-1; 同表 4

表 4 跨域数据集测试 2

Table 4 Cross-domain tests 2

粒度 类型	基准模型	数据集测试性能 (%)											
		C→M		C→D		C→MS		D→M		D→C		D→MS	
		R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP	R-1	mAP
单粒度	ResNet50	42.6	18.9	29.0	14.0	11.6	3.0	46.7	20.1	6.8	6.3	17.2	5.0
	ResNet50+CIP	50.3	24.7	39.1	20.3	15.0	4.3	50.3	22.8	8.5	6.9	20.0	6.1
多粒度	PCB	52.1	26.5	29.2	15.2	7.7	2.1	56.5	27.7	8.4	6.9	17.2	5.1
	PCB+CIP	53.5	28.1	34.0	18.5	10.7	3.0	59.3	29.6	8.3	7.3	21.0	6.4
	PAP	55.5	30.0	34.0	17.9	11.1	3.2	59.5	30.6	9.7	8.0	19.1	6.3
	PAP+CIP	58.5	32.5	37.1	19.9	13.0	3.9	62.3	33.4	12.1	9.6	25.0	8.5

该方法对跨数据集行人重识别也能起到一定的作用。在单数据集测试中, CIP 模块对模型精度的提升效果随着模型分支的增加而减弱, 但在跨数据集测试中这种现象有所减缓。以 $M \rightarrow D$ 数据集测试为例, CIP 模块对 ResNet50、PCB 和 PAP 在 Rank-1 指标的提升分别是 2.4%、2.4%、3.1%, 在 mAP 指标的提升分别是 1.8%、2.9%、2.9%。这表明通过增加 CIP 模块, 模型更加关注区域与区域之间的关系, 并在一定程度上提升了模型的泛化能力。通过对比单数据集测试与跨数据集测试两个实验可看出, CIP 模块对多粒度网络模型的性能提升在跨数据集中的表现更为明显。

表 5 为将本文方法与其他深度学习的行人重

识别方法在 4 个数据集的结果进行比较。从表 5 可看出, 本文复现的 $MGN^* + CIP$ 模块性能总体优于其他的行人重识别方法。虽然相对于 MGN^* 仅有小幅度的提升, 但由于 MGN^* 本身是一个具有 3 个分支及 8 个粒度, 针对各条分支结合了三元组损失函数和交叉熵损失函数进行复杂调试的模型, 所以提升相对困难。尽管如此, CIP 模块还是能在一定程度增强了该模型的性能。而且在存在大量行人背景区域过多的数据集 CUHK03 上, Rank-1 和 mAP 指标分别提升了 2.2%、1.8%。这说明了 CIP 模块从弱检测效果上对 MGN 的精度具有更大的提升效应。

4.2 模块的可视化验证分析

为深入探究 CIP 模块能够起作用的机理, 本

表 5 与表现最佳方法的比较

Table 5 Comparison with recent state-of-the-art methods

模型类型	数据集测试性能 (%)							
	MSMT17		Market-1501		CUHK03		DukeMTMC-ReID	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
SFT+DS ^[27]	73.6	47.6	93.4	82.7	/	/	86.9	73.2
MANCS ^[28]	/	/	93.1	82.3	65.5	60.5	84.9	71.8
PCB ^[10]	/	/	93.2	81.1	63.7	57.5	83.3	69.2
PAP ^[11]	79.2	52.9	94.4	84.5	72.0	66.2	86.1	73.3
MGN ^[13]	/	/	95.7	86.9	66.8	66.0	88.7	78.4
MGN*	93.3	67.1	94.9	87.9	78.3	75.3	89.1	79.9
MGN*+CIP	93.4	67.9	95.0	88.5	80.5	77.1	89.7	80.6

注: MGN* 表示此处的 MGN 为本文所复现的模型; “/” 表示没有做该数据集相应指标测试

文在 ResNet50 上进行了可视化探究的实验。其中, 实验所采用的照片均来自 MSMT17 数据集。

本文在 CIPN 第 2 个 CIP 模块(即图 1 中的 CIP module 2)输出的特征上使用类激活映射(Class Activation Mapping)^[29], 对 CIPN 分支更加关注的区域进行可视化, 结果如图 4 所示。从图 4 可以明显观察到, 模型更加关注行人所在区域。这是由于模型通过学习点与点之间的相关性, 而背景区域与行人区域的差异相对较大, 从而使得模型更为关注行人区域, 达到了某种程度的弱检测效果。

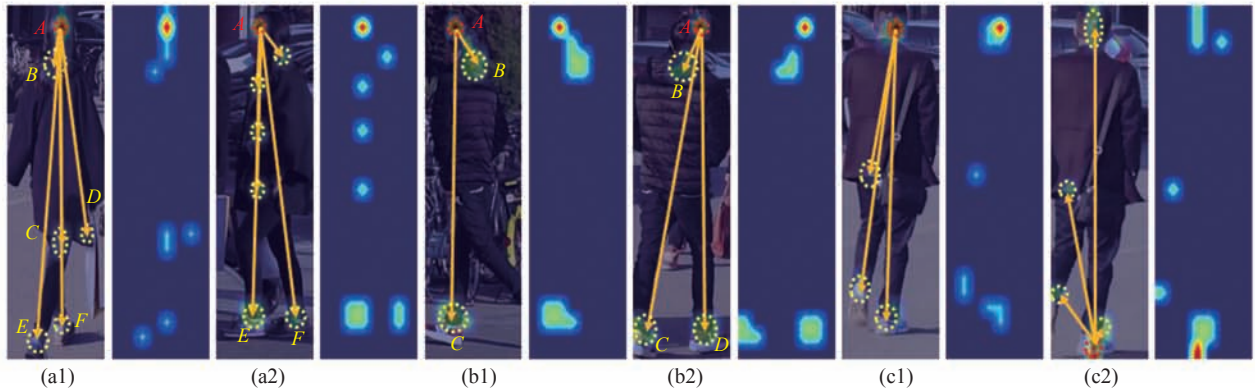
此外, 本文还对特征点之间的关系进行了可视化, 结果如图 5 所示。对于经过 Softmax 函数归一化后的矩阵 $\mathbf{S}' \in \mathbf{R}^{N \times N}$ 上的任意一点 $s'_{i,j}$

(代表着特征图点 i 与点 j 之间的相关性)而言, 为了证明这种相关性的存在, 本文找出矩阵 \mathbf{S} 中最大值对应点 $s'_{i,j} = \max(\mathbf{S})$, 并将与该点相关性最大的 10 个点输出(图 2)。具体操作为, 输出最大点所在列 j' (图 5 标注为虚线红色圈)中排序前 10 的值的对应点(图 5 中标注为虚线黄色圈, 由于部分点互相靠近画在同一个虚线黄色圈, 所以每个行人图片并没有 10 个虚线黄色圈)。图 5(a1)与图 5(a2)为不同视角的同一个人, 可以看出, 在不同视角下显著性特征 C 与 D 虽然丢失了, 但显著性特征 A 、 E 、 F 仍然属于强相关性。对于图 5(b1)与图 5(b2)来说, 虽然图 5(b1)中部分行人区域缺失(少了一只脚), 但特征 A 、 B 、 C 三者的强相关性仍然存在。这表明, 这种相关性



图 4 用类激活映射对 CIP 模块输出的特征图进行可视化

Fig. 4 Visualized feature maps output by the CIP modules



注: 红色虚线圈为受到关注的重点特征; 黄色圈代表较为相关的区域

图 5 图片块之间的相关性进行可视化

Fig. 5 Visualization of the correlations between image patches

不会随着视觉的变化或因遮挡而丢失。这将对摄像机拍摄角度不同的行人重识别非常有效, 也能较好地解决遮挡问题, 是 CIP 模块对各个模型精度提升的关键所在。

5 结 论

本文提出一种适用于行人重识别的 CIP 模块及 CIP 的嵌入方法 CIPN。CIP 模块主要通过关注特征图中点与点之间的依赖来学习视角不变性及区分行人背景的特征。在 4 个数据集上进行的充分实验结果表明, 通过嵌入 CIP 模块, 大部分多粒度行人重识别方法都得到了显著的提升。其中, 通过整合 CIP 模块到 MGN 上, 本文在 4 个数据集上均获得了最好的结果。此外, CIP 模块具有较好的兼容性, 能够很方便地嵌入到任何行人重识别网络。同时, 本文也通过可视化的方法验证了 CIP 模块的有效机理。

参 考 文 献

- [1] 王亦民. 面向监控视频的行人重识别技术研究 [D]. 武汉: 武汉大学, 2014.
- [2] 宋婉茹, 赵晴晴, 陈昌红, 等. 行人重识别研究综述 [J]. 智能系统学报, 2017, 12(6): 770-780.
- [3] 孙志军, 薛磊, 许阳明, 等. 深度学习研究综述 [J]. 计算机应用研究, 2012, 29(8): 2806-2810.
- [4] Yi D, Lei Z, Liao S, et al. Deep metric learning for person re-identification [C] // 2014 22nd International Conference on Pattern Recognition, 2014: 34-39.
- [5] Song CF, Huang Y, Ouyang WL, et al. Mask-guided contrastive attention model for person re-identification [C] // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 1179-1188.
- [6] Xu J, Zhao R, Zhu F, et al. Attention-aware compositional network for person re-identification [C] // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 2119-2128.
- [7] Zheng L, Huang YJ, Lu HC, et al. Pose-invariant embedding for deep person re-identification [J]. IEEE Transactions on Image Processing, 2019, 28(9): 4500-4509.
- [8] 李天霖. 融合局部特征的行人重识别神经网络模型 [D]. 西安: 西安电子科技大学, 2019.
- [9] Kalayeh MM, Basaran E, Gökmen M, et al. Human semantic parsing for person re-identification [C] // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 1062-1071.
- [10] Sun YF, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling

- (and a strong convolutional baseline) [C] // European Conference on Computer Vision (ECCV), 2018: 480-496.
- [11] Huang HJ, Yang WJ, Chen XT, et al. EANet: enhancing alignment for cross-domain person re-identification [Z/OL]. arXiv:1812.11369, 2018.
- [12] Su C, Li JN, Zhang SL, et al. Pose-driven deep convolutional model for person re-identification [C] // The IEEE International Conference on Computer Vision (ICCV), 2017: 3960-3969.
- [13] Wang GS, Yuan YF, Chen X, et al. Learning discriminative features with multiple granularities for person re-identification [C] // Proceedings of the 26th ACM International Conference on Multimedia, 2018: 274-282.
- [14] Deerwester S, Dumais ST, Furnas GW, et al. Indexing by latent semantic analysis [J]. *Journal of the American Society for Information Science*, 1990, 41(6): 391-407.
- [15] Hofmann T. Unsupervised learning by probabilistic latent semantic analysis [J]. *Machine Learning*, 2001, 42(1-2): 177-196.
- [16] Monay F, Gatica-Perez D. PLSA-based image auto-annotation: constraining the latent space [C] // Proceedings of the 12th Annual ACM International Conference on Multimedia, 2004: 348-351.
- [17] Lienhart R, Rombert S, Hörster E. Multilayer pLSA for multimodal image retrieval [C] // Proceedings of the ACM International Conference on Image and Video Retrieval, 2009: 9.
- [18] 高程程, 惠晓威. 基于灰度共生矩阵的纹理特征提取 [J]. *计算机系统应用*, 2010, 19(6): 195-198.
- [19] Wang XL, Girshick R, Gupta A, et al. Non-local neural networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7794-7803.
- [20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [Z/OL]. arXiv Preprint arXiv:1409.1556, 2014.
- [21] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition [C] // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770-778.
- [22] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
- [23] Liang Z, Li SY, Lu T, et al. Scalable person re-identification: a benchmark [C] // The IEEE International Conference on Computer Vision (ICCV), 2015: 1116-1124.
- [24] Li W, Zhao R, Xiao T, et al. DeepReID: deep filter pairing neural network for person re-identification [C] // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: 152-159.
- [25] Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking [C] // European Conference on Computer Vision (ECCV), 2016: 17-35.
- [26] Wei L, Zhang S, Gao W, et al. Person transfer GAN to bridge domain gap for person re-identification [C] // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 79-88.
- [27] Luo CC, Chen YT, Wang NY, et al. Spectral feature transformation for person re-identification [C] // The IEEE International Conference on Computer Vision (ICCV), 2019: 4976-4985.
- [28] Wang C, Zhang Q, Huang C, et al. Manacs: a multi-task attentional network with curriculum sampling for person re-identification [C] // The European Conference on Computer Vision (ECCV), 2018: 365-381.
- [29] Zhou BL, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization [C] // The IEEE International Computer Vision and Pattern Recognition (CVPR), 2016: 2921-2929.