

基于图神经网络的行人重识别方法

郝志峰^{1,2}, 苏伟根², 许柏炎², 温雯², 蔡瑞初²

¹佛山科学技术学院数学与大数据学院, 广东 佛山

²广东工业大学计算机学院, 广东 广州

Email: weigensu@gmail.com

收稿日期: 2021年3月22日; 录用日期: 2021年4月16日; 发布日期: 2021年4月23日

摘要

基于深度学习的行人重识别方法面临的挑战在于如何解决由于行人姿态多变、背景混杂、摄像头视角差异大和部分遮挡等情况引起的行人不对齐的问题。如何提取细粒度的、具有强判别性的特征成为解决问题的关键, 因此本文提出了一种新型的基于图神经网络的行人重识别方法, 其中包括: 1) 结合人体语义解析结果提取细粒度局部特征, 构建部位关系图, 并通过图神经网络学习到细粒度的图表示。2) 通过图度量学习方法联合优化学习网络, 学习强判别性的特征表示。本文提出的方法在主流评估数据集上与行人重识别前沿方法进行实验比较, 结果表明了方法的有效性。

关键词

行人重识别, 图度量学习, 人体语义解析, 局部对齐

Person Re-Identification Method Based on Graph Neural Network

Zhifeng Hao^{1,2}, Weigen Su², Boyan Xu², Wen Wen, Ruichu Cai²

¹School of Mathematics and Big Data, Foshan University, Foshan Guangdong

²School of Computer Science, Guangdong University of Technology, Guangzhou Guangdong

Email: weigensu@gmail.com

Received: Mar. 22nd, 2021; accepted: Apr. 16th, 2021; published: Apr. 23rd, 2021

Abstract

The challenge for the deep learning-based person re-identification method is how to solve the problem of pedestrian misalignment caused by pedestrian posture change, mixed background,

large camera viewing angle difference, and partial occlusion. How to extract fine-grained and highly discriminative features has become the key to solving the problem. Therefore, a new person re-identification method based on graph neural network is proposed, which includes: 1) Combine the human semantic parsing model to locate the position map and learn fine-grained local features as graph representation through the graph neural network. 2) Through the proposed graph metric learning method to jointly optimize the learning network to learn more discriminative feature representation. Experiments were performed on mainstream evaluation datasets, and the results showed the effectiveness of the method.

Keywords

Person Re-Identification, Graph Metric Learning, Human Semantic Parsing, Part Alignment

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

行人重识别(Person Re-Identification, Person Re-ID), 是指在跨摄像头的情况下, 在多张行人图像中找出匹配目标人物的图像[1], 广泛应用于安防系统和智能商业等领域。行人重识别实现思路为计算查询图像与图库中所有图像的相似度, 根据相似度从大到小排列得到重识别结果[2]。目前, 行人重识别亟需解决的问题有: 行人姿态多变、背景变化大、拍摄角度差异大、存在局部遮挡等。解决上述难题的关键在于特征提取和相似性度量。

传统行人重识别任务研究主要集中在人工设计特征提取器, 比如文献[3]提出的尺度不变特征变换(Scale-Invariant Feature Transform, SIFT), 文献[4]提出的局部二值模式(Local Binary Patterns, LBP)以及文献[5]提出的局部最大特征值(Local Maximal Occurrence, LOMO)。但这些方法提取到的特征描述能力有限, 难以应用于复杂情况下大数据量的行人重识别任务。随着深度学习的兴起, 卷积神经网络(Convolution Neural Network, CNN)凭借出色的特征提取能力, 取代了传统的人工特征提取器。研究者利用卷积神经网络提取行人整图图像的全局特征, 并通过不同的局部划分策略, 如图像切块[6]、通过姿态语义信息分块[7]通过注意力机制分块[8]、和通过人体语义解析信息分块[9]等, 进行局部特征提取, 最后将全局特征和局部特征拼接得到行人的特征表示, 并通过计算行人特征表示之间的欧式距离作为行人图像的相似度。

基于深度卷积网络的局部特征提取方法虽然可以提取到有判别性的特征, 在行人重识别任务上取得了不错的效果, 但仍存在两个问题导致局部特征的判别性被削弱。第一, 由于局部划分的粒度比较大, 获取到的局部区域包含行人信息和背景噪声信息。比如 PCB [6]使用垂直方向均匀切块的方法, 每个局部块都包含了背景噪声信息。SpindleNet [7]根据语义将人体姿态中身体部位的关节点分成五组, 对应身体的五个部位, 将每组关节点在行人图像上围起来的区域作为局部特征提取的区域。这种方式提取到的局部图像同样也包含了背景噪声信息。第二, 忽视了提取到的局部特征之间的关联关系。文献[6] [7] [8] [9]等局部特征提取方法, 直接将提取到的局部特征拼接, 然后计算拼接后的局部特征之间的相似度, 这种方式导致了模型在辨别两个穿着相似的行人时, 由于缺少了额外的位置辅助信息, 从而很难进行精确匹配。

针对上述问题, 本文提出了一种基于人体语义解析和图神经网络的行人重识别模型。本文的主要贡

献分为三个部分：

1) 本文引入人体语义解析网络预先提取行人身体局部区域，用于提取细粒度的局部特征，获取到行人身份的敏感信息，减少背景噪声信息带来的负面影响。

2) 本文引入图卷积神经网络，用于提取局部特征之间的关联关系，增强局部特征的判别性。

3) 本文将行人局部特征构造成图，引入全局对级相似度损失函数进行模型训练，进一步增强局部特征的判别性。

本文的结构如下：第二章介绍行人重识别的相关工作，第三章介绍提出的方法，第四章展示实验，第五章总结全文。

2. 相关工作

2.1. 行人重识别

行人重识别旨在实现跨摄像头检索目标人物，主要分为特征提取和度量学习两部分。

随着深度学习技术的发展，CNN 已经广泛应用于计算机视觉任务中的图像特征提取。通过 CNN 直接提取整张图像特征的方式简单高效，但提取到的特征很容易忽略行人身份的敏感信息，从而导致全局特征无法很好的区分类间相似度高或类内差异大的行人样本。因此，研究者们提出了基于局部特征的行人重识别方法，通过定位行人的身体部位，提取对应部位的身份敏感信息。Sun [6]等人提出了 PCB 模型，将行人图像水平切割分块，通过骨干网络提取到每块的特征，拼接成局部特征。Zhao [7]等人提出的 SpindleNet 引入了行人姿态信息，通过姿态关键点提取到 7 个身体区域，并从身体区域中提取局部特征。Li [8]等人提出了 HA-CNN 模型，引入了多条局部分支，每条分支分别关注图片的一块局部区域的特征，最终将各条分支学到的特征进行融合得到行人的局部特征。文献[9]中的 SPReID 模型引入行人人体语义解析网络提取行人的身体区域的掩膜图像，与原图经过骨干网络得到的特征图结合，得到行人身体区域的局部特征。众多基于局部特征的行人重识别方法证实了局部特征的有效性。

度量学习是广泛应用于图像检索领域的一种方法，旨在通过网络学习出两张图片的相似度。在行人重识别任务上表现为同一行人的不同图像的相似度要大于不同行人的图像的相似度。Varior [10]等人使用了孪生网络，同时输入两张行人图像，采用对比损失(Contrastive Loss)训练模型，使得相同行人图像(正样本)之间的距离逐步缩小，不同行人图像(负样本)之间的距离逐步增大。文献[11]首先构造包括一对正样本和一对负样本的三元组，分别命名为目标图片(Anchor)，正样本图片(Positive)和负样本图片(Negative)。然后采用三元组损失函数训练网络，拉近了正样本之间的距离，增大了负样本之间的距离。但三元组损失没有考虑到正样本的绝对距离，因此文献[12]提出改进三元组损失，在损失函数上增加正样本绝对距离公式项约束。Chen [13]等人进一步考虑了正负样本对的绝对距离，提出了四元组损失(Quadruplet Loss)。

2.2. 图卷积神经网络

图卷积网络(Graph Convolutional Network, GCN)是一种能对图数据进行深度学习的方法。它是对卷积神经网络在图域的扩展，能同时对结点特征信息与结构信息进行端到端的学习。文献[14]首次将图卷积用于半监督结点分类任务并取得了成功。之后图卷积神经网络广泛应用于计算机视觉领域，如动作识别[15]、多标签图像识别[16]、3D 几何重构[17]等。在行人重识别领域，很多方法将行人图像样本建图，使用 GCN 来更新行人图像的特征表示。比如，Shen [18]等人将查询集和候选集的图片两两组成对，并通过 GCN 来学习查询集图片与候选集图片之间的关联关系。Ye [19]等人为每个摄像机中的图像样本构造一个图，然后利用图匹配方案进行跨摄像机的标签关联等。

3. 本文方法

本节主要介绍基于图神经网络的行人重识别模型的具体细节。如图 1 所示, 网络模型分为全局表示学习分支和图表示学习分支, 其中全局分支使用交叉熵损失函数和三元组损失函数对模型进行训练, 学习到行人图像的全局表示。图表示学习分支先通过人体语义解析的结果得到行人图像的局部特征, 然后将每块局部特征作为图结点, 将局部块语义连接作为边构成图, 通过图卷积网络更新局部特征, 再通过交叉熵损失函数和图度量学习模块共同训练网络, 增强提取到的特征的表达能

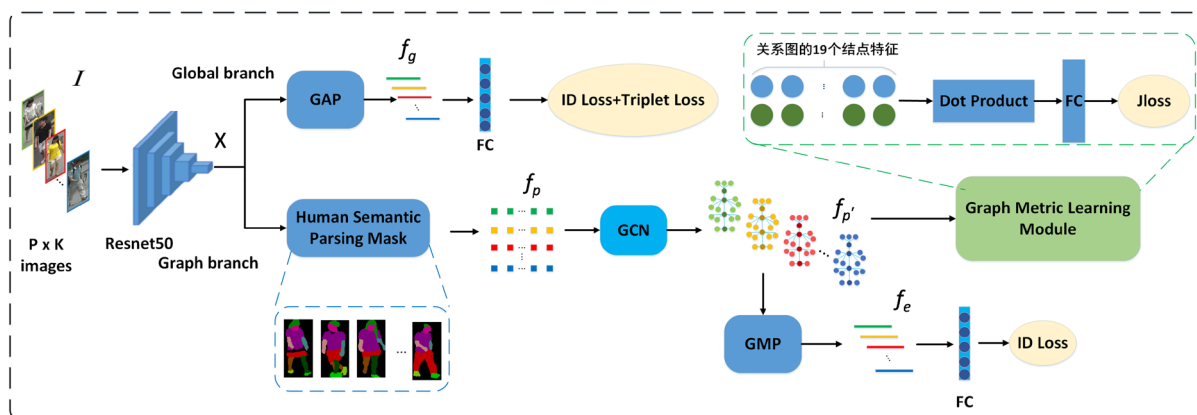


Figure 1. Human semantic parsing and graph metric learning fused model

图 1. 人体语义解析与图度量学习融合模型

3.1. 全局表示学习

由于残差神经网络[20]在行人重识别任务取得了很好的效果, 因此本文选取 Resnet-50 骨干网络来提取输入图片相应的特征图。为了增大特征图的感受野, 本文将 Resnet-50 网络最后一层卷积层的步长设置为 1, 使得大小为 256×128 的原始图片经过骨干网络后输出的特征图大小为 8×4 , 然后对特征图进行全局平均池化操作(Global Average Pooling, GAP)得到形状为 $C \times H \times W$ 的特征, 其中 C 为通道数, H 为高度, W 为宽度。再依次经过一层全连接层(Fully Connected, FC)进行降维、一层批处理归一化层(Batch Normalization, BN)进行归一化, 得到全局表示。全局分支使用交叉熵损失函数和困难样本三元组损失函数进行联合训练, 本文参考了文献[21]的做法, 将经过 BN 层前的特征用于计算三元组损失, BN 层后的特征用于计算交叉熵损失。

3.2. 基于人体语义解析的图表示学习

3.2.1. 局部特征提取

现有的局部特征提取方法[6]证明了局部特征的有效性, 但 these 方法提取到的局部区域是粗粒度的。为了精准定位行人身体部位的区域, 本文使用了在人体语义解析数据集(Look into Person, LIP [22])上预训练好的 SCHP [23]模型, 通过该模型提取到行人图像的 19 个像素级人体部位标注, 包括头部、手臂、左腿、上半身、下半身等, 形成掩膜图像, 如图 2 所示。将骨干网络提取到的特征图和二值化的掩膜图像对应位置相乘, 再将所得结果进行全局池化, 从而提取人体不同身体部位区域的特征作为局部特征。

3.2.2. 局部结构图的构建

现有的行人重识别方法提取到局部特征后, 往往将各块局部特征单独考虑, 通过简单的拼接或者最大池化操作进行融合, 忽略了局部特征之间的关系。本文将提取到的 19 个身体部位的特征作为图的结点,

并以 19 个结点的自然连接作为边，构成局部特征关系图 G ，如图 3 所示，每个柱形表示该部位的特征。获取到关系图 G 后，通过两层图卷积组成的图卷积网络进行结点间信息的传播，从而对关系图中的结点进行更新，得到融入结构信息的局部特征 f_p ，最后通过最大池化融合局部特征得到整图的表示 f_e 。图卷积函数的公式表达如式(1)所示，其中 H^L 表示第 L 层图卷积后的特征， \hat{A} 表示归一化后的邻接矩阵， W^{L-1} 是模型学习到的变换矩阵， h 表示非线性激活函数。

$$H^L = h(\hat{A}H^{L-1}W^{L-1}) \quad (1)$$



Figure 2. Human semantic parsing result of some query images in Market-1501

图 2. Market-1501 上部分检索图像的人体语义解析图

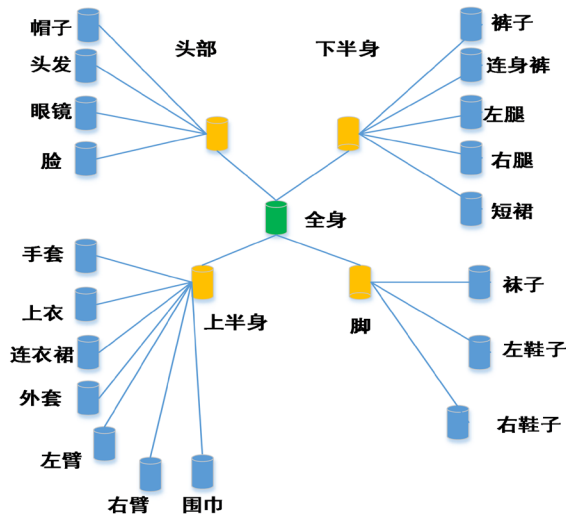


Figure 3. Relation graph of local features

图 3. 局部特征关系图

3.2.3. 图度量模块

图度量学习旨在训练模型学习到满足正类图在特征空间距离更近，负类图在特征表示空间距离更远的图表示。常用的图度量有图嵌入(Graph Embedding)、模式计数(Motif Counting)和图编辑距离(Graph Edit Distance)等。本文的身体局部连接关系是固定的，因此不适用模式计数。而图编辑距离的方式需要预先计算编辑代价，不够高效。图嵌入的方式只考虑整图全局的相似度，容易丢失图结构信息，因此本文引

入了一种图度量学习方法，通过对比两个图对应结点之间的相似性，并赋予不同的权重来获得整图的相似性。具体实现模块见图 1 Graph Metric Learning Module。首先通过一层点积层计算两个图的对应结点的相似性，再经过一层全连接层学习各个结点的重要程度，得到整图的相似性，并通过全局成对相似度损失函数对正负样本的相似性分布做约束。

全局成对相似度损失函数由文献[24]提出，旨在避免过度采样和欠采样，基于相同类别和不同类别的特征相似性呈独立分布的假设，最小化正类相似度和负类相似性分布的方差，最小化正类相似性分布的均值，并最大化负类相似性分布的均值，在学习局部图片特征的任务上效果远超传统的损失函数。该损失函数的表达式如下：

$$J = (\sigma^{2+} + \sigma^{2-}) + \lambda \max(0, m - (\mu^+ - \mu^-)) \quad (2)$$

$$\mu^+ = \sum_{i=1}^N g(f_{p_i}, f_{p_i^+}) / N \quad (3)$$

$$\mu^- = \sum_{i=1}^N g(f_{p_i}, f_{p_i^-}) / N \quad (4)$$

$$\sigma^{2+} = \sum_{i=1}^N (g(f_{p_i}, f_{p_i^+}) - \mu^+)^2 / N \quad (5)$$

$$\sigma^{2-} = \sum_{i=1}^N (g(f_{p_i}, f_{p_i^-}) - \mu^-)^2 / N \quad (6)$$

3.3. 损失函数

3.3.1. 标签平滑交叉熵损失函数

行人重识别通常被视作分类任务，并用交叉熵损失函数训练行人重识别网络。由于行人重识别中训练集和测试集中的行人是不同的，用传统的交叉熵损失函数训练得到的网络容易对训练集的行人标签过拟合，为了避免这种情况，我们借鉴了文献[25]的做法，对标签进行了平滑处理，其中 K 表示行人标签数， ε 表示阈值，实验中设置为 0.1，如式(7)，(8)所示。

$$Y_c = \begin{cases} 1 - \frac{K-1}{K} \varepsilon, & c = Y \\ \varepsilon / K, & \text{其他} \end{cases} \quad (7)$$

$$L_{\text{xcent}} = - \sum_{c=1}^K Y_c \log(p_c) \quad (8)$$

3.3.2. 困难样本三元组损失函数

行人重识别任务也被视为聚类任务，目标是直接将相同行人的若干图片映射到高维空间，形成聚类效应。行人重识别任务中常用三元组损失函数进行模训练。考虑到大多数三元组是简单样本，不利于模型的训练，因此借鉴了 Alex [26]等人提出的困难样本三元组损失函数，对于每个目标图片 anchor 在一个训练批次内选择欧式距离最远的正样本图片 positive 和欧式距离最近的负样本图片 negative 来训练网络，增强网络的泛化能力，从而使网络学习到更好的表征。难样本三元组损失函数如式(9)所示，其中 α 表示目标样本与正负样本的间隔参数，设置为 0.3， $d_{a,p}$ ， $d_{a,n}$ 分别表示目标样本和正样本的欧氏距离，目标样本与负样本的欧氏距离。

$$L_{\text{tri}} = (\max d_{a,p} - \min d_{a,n} + \alpha, 0)_+ \quad (9)$$

3.3.3. 联合损失函数

本文提出的行人重识别网络通过标签平滑交叉熵损失函数、三元组损失函数和全局成对相似度损失函数进行训练,共同约束特征,其中全局分支使用标签平滑交叉熵损失函数 L_{xcent_g} 和三元组损失函数 L_{tri_g} 进行训练,局部分支使用标签平滑交叉熵损失函数 L_{xcent_l} 和全局成对相似度损失函数 J 进行训练。关于全局成对相似度损失函数已在 3.2.3 节中详细阐述。因此,本文的损失函数如式(10)所示。

$$L_{\text{total}} = L_{\text{xcent}_g} + L_{\text{tri}_g} + L_{\text{xcent}_l} + J \quad (10)$$

4. 实验与分析

4.1. 数据集

本文所提出的算法在行人重识别领域的三大主流数据集(Market1501、DukeMTMC-reID、CUHK03)进行了广泛实验,并通过 mAP、Rank-1、Rank-5、Rank-10 评价指标作出对比评估,我们的实验结果并未进行重排序操作。

Market1501 数据集中,训练集图片数为 12936,包括 751 个人;测试集为 19732 张图像,包括查询集 3368 张图像。DukeMTMC-reID 数据集在 2017 年公开,训练集中 702 人,16522 张图像;测试集中 702 人,19889 张图像,包括查询集 2228 张图像。CUHK03 数据集在 2014 年公开,本文实验中采用了新版的测试协议,将数据集分为包含 767 个行人的训练集和包含 700 个行人的测试集。表 1 列出了每个数据集的详细信息。

Table 1. Details of the datasets

表 1. 数据集的细节

数据集	行人标签数	摄像头数	图片数	采集场景
Market-1501	1501	6	32668	清华大学
DukeMTMC-reID	1404	8	36411	杜克大学
CUHK03	1467	10	14097	香港中英文大学

4.2. 评估标准

评估行人重识别模型的性能时,需要获取相关实验的模型测评指标。针对不同的研究任务,评价指标也不尽相同。在行人重识别任务中,常用首位匹配率(Rank-1)和均值平均精度(mAP)这两个评价指标来描述模型性能的优劣。对应两个指标越高,说明模型对行人重识别的效果就越好。

首位匹配率(Rank-1)是在对应的查询集(Query Set)中的每张行人图像,通过行人重识别算法在候选集中选出候选序列,对应行人的图像排在候选序列第一位的数量占整个查询集图像数量的比例。假定查询集 $P = \{p_1, p_2, \dots, p_m\}$, 候选集 $G = \{g_1, g_2, \dots, g_n\}$, m, n 分别表示图像张数。 P 和 G 之间计算相似度得到相似度矩阵 S ,将 S 中的每行按数值从大到小的顺序排列, S 中的数值越大,表示两张图像的相似度越高,定义 s_1 为 S 中第一列中对应的来自查询集和候选集组成的图像对是同一行人身份的数量,则首位匹配率可表示成: $\text{Rank1} = s_1/m$ 。Rank-5, Rank-10 表示候选序列中前 5 张图片和前十张图片中目标任务的命中率,可由同样的计算方法得出。

均值平均精度(mAP)是衡量模型总体好坏的指标,由模型的召回率和精确率进行计算得出。对于查询集中的每个查询图像,定义精确度 P 为候选列表中前 k 个图像中正确匹配的图像数目 n 与 k 的比值,可表示成: $P = n/k$ 。而图像样本对应的召回率 R 为在候选列表的前 k 个图像正确匹配的数量 n 与行人在

候选集中的图像数量 m 的比值，即： $R = k/m$ 。

当 k 取不同值时，可以在坐标系上得到多个 P 值的点和 R 值的点，用曲线分别依次连接各个 P 值点和各个 R 值点可以得到 PR 图。该样本的平均精度(AP)即为 PR 图中曲线围成的面积，而 mAP 为所有查询图像的 AP 值的均值。mAP 作为常用的评估指标，可以更加综合和客观的评价模型性能。

4.3. 消融实验

为了进一步验证提取到的局部特征的有效性，本文以全局分支组成的模型 Model_G 作为 Baseline，并采用标签平滑交叉熵损失函数和三元组损失函数进行训练。基于相同 Baseline 在 Market1501 和 DukeMTMC-reID 数据集做了多组消融实验，实验结果如表 2 所示。

从结果中我们发现，同时考虑全局特征和局部特征的 Model_G_L 模型，在 Market1501 数据集上的表现为 rank-1，mAP 比 Baseline 提升了 0.3% 和 1.3%，在 DukeMTMC-reID 数据集表现为 rank-1 提升了 1.3%，mAP 提升了 0.6%。证明了局部特征能够补足全局特征遗漏的行人细节信息。

在进一步考虑局部特征关系，引入图卷积网络后，模型 Model_G_L_GCN 在两个数据集上 rank-1，mAP 分别再提升 0.4%，0.5% 和 0.2%，1.8%。证明了引入局部关系信息，使用图卷积得到行人的图表示，可以更好的实现行人重识别任务。

为了验证图度量学习模块的有效性，训练局部分支时，去掉了 triplet loss，引入图度量学习模块组成 Model_G_L_GCN_Jloss 模型，在 rank-1 和 mAP 上进一步提升了 1.0%，0.4% 和 0.5%，1.0%。证明了图度量学习模块中的全局成对相似度损失函数在图表示学习任务上的效果优于三元组损失函数。图度量学习模块考虑了各个结点之间的相似性，并通过全连接层学习各个结点的重要性，比三元组损失单独考虑整体相似度更能学到鲁棒性的特征。

Table 2. Person re-ID evaluation value on Market-1501 and DukeMTMC-reID
表 2. 在 Market-1501 和 DukeMTMC-ReID 数据集上的实验结果

模型	Market-1501				DukeMTMC-reID			
	Rank-1	Rank-5	Rank10	mAP	Rank-1	Rank-5	Rank-10	mAP
Model_G	93.8	97.8	98.3	84.5	85.5	92.9	94.7	72.7
Model_G_L	94.1	98.0	98.6	85.8	86.8	93.1	94.8	73.3
Model_G_L_GCN	794.5	98.1	98.7	86.3	87.0	93.5	95.2	75.1
Model_G_L_GCN_Jloss	95.5	98.3	98.9	86.7	87.5	93.7	95.4	76.1

4.4. 与当前主流方法对比

为了验证本文中所提算法的优越性，本文与近几年的行人重识别主流算法分别在 Market1501、DukeMTMC-reID、CUHK03-NP 三个数据集上进行了比较。由表 3 可知，在 Market1501、DukeMTMC-reID 数据集上，本文提出的算法评价指标 Rank-1、Rank-5、Rank-10、mAP 与众多对比模型相比均达到了较高水平。对比 MGN 和 PSRL 模型，本文在 Market1501 和 DukeMTMC 的 rank-1 和 mAP 指标略低，这是因为 MGN 利用了多层粒度的特征图提取局部特征，能适应背景环境变化大的情况。PSRL 模型使用了特征拼接模块，进一步融合局部特征，并训练模型学习得到局部结构之间的连接关系。从表 4 可知，MGN 和 PSRL 在行人存在严重遮挡的情况下提取到的特征不够鲁棒。CUHK03 数据集的行人姿态变化大，遮挡问题严重，行人重识别难度比其他两个数据集要大。而本文提出的算法在该数据集上的效果远远超过上述两个模型，证明了本文通过图度量学习训练获取到的图表示更具鲁棒性。

Table 3. Comparison with mainstream algorithm evaluation indicators on Market-1501 and DukeMTMC-reID
表 3. 在 Market-1501 和 DukeMTMC-reID 上与主流算评价指标比较

模型	Market-1501				DukeMTMC-reID			
	Rank-1	Rank-5	Rank10	mAP	Rank-1	Rank-5	Rank-10	mAP
SpindleNet [7]	76.9	98.1	94.6	-	-	-	-	-
PSE [27]	87.7	91.5	96.8	69.0	79.8	89.7	92.2	62.0
PGR [28]	93.8	90.8	-	77.2	83.6	91.7	-	65.9
MaskReID [29]	90.0	97.7	-	75.3	78.8	-	-	61.9
AlignedReID [30]	91.8	97.6	-	79.3	-	-	-	-
PCB [6]	93.8	-	98.5	81.6	83.3	90.5	92.5	69.2
BFE [31]	94.2	97.1	-	84.3	86.8	-	-	72.1
MGN [32]	95.7	97.5	-	86.9	88.7	-	-	78.4
DuATM [33]	91.4	-	-	76.6	81.8	90.2	-	64.6
HA-CNN [8]	91.2	-	-	75.7	80.5	-	-	63.8
Manes [34]	93.1	97.1	-	82.3	84.9	-	-	71.8
PSRL [35]	95.9	-	-	87.2	88.1	93.9	95.7	76.7
Our approach	95.5	98.3	98.9	86.7	87.5	93.7	95.4	76.1

Table 4. Comparison with mainstream algorithm evaluation indicators on CUHK03-NP
表 4. 在 CUHK03-NP 上与主流算评价指标比较

模型	Detected		Labeled	
	Rank-1	mAP	Rank-1	mAP
HA-CNN [8]	41.7	38.6	44.4	41.0
PCB [6]	62.8	56.7	-	-
Manes [34]	65.5	60.5	65.5	60.5
MGN [32]	68.0	67.4	68.0	67.4
BFE [31]	73.6	69.7	73.6	71.7
PSRL [35]	67.6	65.8	68.7	68.2
Our approach	76.3	70.8	76.4	72.5

5. 结论

本文从由行人姿态多变、背景混杂、摄像头视角差异大和部分遮挡导致的行人不对齐这一问题着手，提出了一种基于图神经网络的行人重识别算法，使用人体语义解析得到行人身体部位标注，提取细粒度局部特征构建部位关系图，再通过图卷积网络融入局部结构信息，并用图度量学习得到更具鲁棒性的图表示。最后在 3 个数据集上的实验结果表明，本文所提模型相比目前的主流算法重识别率更优。未来的工作是简化网络模型复杂度的情况下，寻求效果更优的重识别率的方法，并研究更为有效的跨数据集识别方法，实现跨域行人重识别任务。

基金项目

国家自然科学基金资助项目(61876043, 61976052)。

参考文献

- [1] 罗浩, 姜伟, 范星, 等. 基于深度学习的行人重识别研究进展[J]. 自动化学报, 2019, 45(11): 2032-2049.
- [2] Zheng, L., Shen, L., Tian, L., *et al.* (2015) Scalable Person Re-Identification: A Benchmark. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1116-1124. <https://doi.org/10.1109/ICCV.2015.133>
- [3] Lindeberg, T. (2012) Scale Invariant Feature Transform. *Scholarpedia*, 7, 10491. <https://doi.org/10.4249/scholarpedia.10491>
- [4] Ahonen, T., Hadid, A. and Pietikäinen, M. (2004) Face Recognition with Local Binary Patterns. In: Pajdla, T. and Matas, J., Eds., *European Conference on Computer Vision*, Springer, Berlin, Heidelberg, 469-481. https://doi.org/10.1007/978-3-540-24670-1_36
- [5] Liao, S., Hu, Y., Zhu, X. and Li, S.Z. (2015) Person Re-Identification by Local Maximal Occurrence Representation and Metric Learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 2197-2206. <https://doi.org/10.1109/CVPR.2015.7298832>
- [6] Sun, Y., Zheng, L., Yang, Y., *et al.* (2018) Beyond Part Models: Person Retrieval with Refined Part Pooling (and a Strong Convolutional Baseline). *Proceedings of the European Conference on Computer Vision (ECCV)*, 480-496. https://doi.org/10.1007/978-3-030-01225-0_30
- [7] Zhao, H., Tian, M., Sun, S., *et al.* (2017) Spindle Net: Person Re-Identification with Human Body Region Guided Feature Decomposition and Fusion. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 1077-1085. <https://doi.org/10.1109/CVPR.2017.103>
- [8] Li, W., Zhu, X. and Gong, S. (2018) Harmonious Attention Network for Person Re-Identification. *Proceedings of the 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>
- [9] Kalayeh, M.M., Basaran, E., Gökmen, M., *et al.* (2018) Human Semantic Parsing for Person Re-Identification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 1062-1071. <https://doi.org/10.1109/CVPR.2018.00117>
- [10] Varior, R.R., Haloi, M. and Wang, G. (2016) Gated Siamese Convolutional Neural Network Architecture for Human Re-Identification. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *European Conference on Computer Vision*, Springer, Cham, 791-808. https://doi.org/10.1007/978-3-319-46484-8_48
- [11] Hermans, A., Beyer, L. and Leibe, B. (2017) In Defense of the Triplet Loss for Person Re-Identification. arXiv preprint arXiv:1703.07737.
- [12] Cheng, D., Gong, Y., Zhou, S., *et al.* (2016) 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, 27-30 June 2016, 1335-1344. <https://doi.org/10.1109/CVPR.2016.149>
- [13] Chen, W., Chen, X., Zhang, J. Huang, K. (2017) Beyond Triplet Loss: A Deep Quadruplet Network for Person Re-Identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 403-412. <https://doi.org/10.1109/CVPR.2017.145>
- [14] Kipf, T.N. and Welling, M. (2016) Semi-Supervised Classification with Graph Convolutional Networks. arXiv preprint arXiv:1609.02907.
- [15] Caramalau, R., Bhattarai, B. and Kim, T.K. (2020) Sequential Graph Convolutional Network for Active Learning. arXiv preprint arXiv:2006.10219.
- [16] Chen, Z.M., Wei, X.S., Wang, P. and Guo, Y. (2019) Multi-Label Image Recognition with Graph Convolutional Networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 5177-5186. <https://doi.org/10.1109/CVPR.2019.00532>
- [17] Lei, H., Akhtar, N. and Mian, A. (2020) Spherical Kernel for Efficient Graph Convolution on 3D Point Clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1. <https://doi.org/10.1109/TPAMI.2020.2983410>
- [18] Shen, Y., Li, H., Yi, S., *et al.* (2018) Person Re-Identification with Deep Similarity-Guided Graph Neural Network. *Proceedings of the European Conference on Computer Vision (ECCV)*, 486-504. https://doi.org/10.1007/978-3-030-01267-0_30
- [19] Ye, M., Li, J., Ma, A.J., *et al.* (2019) Dynamic Graph Co-Matching for Unsupervised Video-Based Person Re-Identi-

- fication. *IEEE Transactions on Image Processing*, **28**, 2976-2990. <https://doi.org/10.1109/TIP.2019.2893066>
- [20] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [21] Luo, H., Jiang, W., Gu, Y., Liu, F., et al. (2019) A Strong Baseline and Batch Normalization Neck for Deep Person Re-Identification. *IEEE Transactions on Multimedia*, **22**, 2597-2609. <https://doi.org/10.1109/TMM.2019.2958756>
- [22] Gong, K., Liang, X., Zhang, D., et al. (2017) Look into Person: Self-Supervised Structure-Sensitive Learning and a New Benchmark for Human Parsing. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 932-940. <https://doi.org/10.1109/CVPR.2017.715>
- [23] Li, P., Xu, Y., Wei, Y. and Yang, Y. (2020) Self-Correction for Human Parsing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1**. <https://doi.org/10.1109/TPAMI.2020.3048039>
- [24] Kumar, B.G.V., Carneiro, G. and Reid, I. (2016) Learning Local Image Descriptors with Deep Siamese and Triplet Convolutional Networks by Minimising Global Loss Functions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 5385-5394.
- [25] Szegedy, C., Vanhoucke, V., Ioffe, S., et al. (2016) Rethinking the Inception Architecture for Computer Vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 2818-2826. <https://doi.org/10.1109/CVPR.2016.308>
- [26] Alex, D., Sami, Z., Banerjee, S. and Panda, S. (2018) Cluster Loss for Person Re-Identification. *Proceedings of the 11th Indian Conference on Computer Vision, Graphics and Image Processing*, December 2018, Article No. 43. <https://doi.org/10.1145/3293353.3293396>
- [27] Saquib Sarfraz, M., Schumann, A., Eberle, A. and Stiefelhagen, R. (2018) A Pose-Sensitive Embedding for Person Re-Identification with Expanded cross Neighborhood Re-Ranking. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 420-429. <https://doi.org/10.1109/CVPR.2018.00051>
- [28] Li, J., Zhang, S., Tian, Q., et al. (2019) Pose-Guided Representation Learning for Person Re-Identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **1**. <https://doi.org/10.1109/TPAMI.2019.2929036>
- [29] Song, C., Huang, Y., Ouyang, W., et al. (2018) Mask-Guided Contrastive Attention Model for Person Re-Identification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 1179-1188. <https://doi.org/10.1109/CVPR.2018.00129>
- [30] Zhang, X., Luo, H., Fan, X., et al. (2017) Alignedreid: Surpassing Human-Level Performance in Person Re-Identification. arXiv preprint arXiv:1711.08184.
- [31] Dai, Z., Chen, M., Gu, X., et al. (2019) Batch DropBlock Network for Person Re-Identification and Beyond. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 3691-3701. <https://doi.org/10.1109/ICCV.2019.00379>
- [32] Wang, G., Yuan, Y., Chen, X., et al. (2018) Learning Discriminative Features with Multiple Granularities for Person Re-Identification. *Proceedings of the 26th ACM international conference on Multimedia*, October 2018, 274-282. <https://doi.org/10.1145/3240508.3240552>
- [33] Si, J., Zhang, H., Li, C.G., et al. (2018) Dual Attention Matching Network for Context-Aware Feature Sequence Based Person Re-Identification. *Proceedings of the 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] Wang, C., Zhang, Q., Huang, C., et al. (2018) Mancs: A Multi-Task Attentional Network with Curriculum Sampling for Person Re-Identification. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Proceedings of the European Conference on Computer Vision (ECCV)*, Springer, Cham, 365-381. https://doi.org/10.1007/978-3-030-01225-0_23
- [35] Li, Y., Yao, H., Zhang, T. and Xu, C. (2020) Part-Based Structured Representation Learning for Person Re-Identification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, **16**, 1-22. <https://doi.org/10.1145/3412384>