姿态引导与多粒度特征融合的行人再识别

张良[1,2], 车进[1,2*]

[1]宁夏大学物理与电子电气工程学院, 宁夏 银川 750021;
[2]宁夏大学沙漠信息智能感知重点实验室, 宁夏 银川 750021

摘要 行人再识别系统中,检索到的行人图像会出现较大的姿态差异、复杂的视角变化以及检测框中行人图像不对齐等问题,为此,提出一种可以直接使用人体关键点信息进行行人图像对齐,并在此对齐基础上提取多粒度特征的重识别算法。首先使用姿态预估模型定位人体骨架关键点信息,并根据提取的骨架关键点直接进行行人图像对齐,然后对行人图像提取多粒度特征。评估阶段使用姿态信息结合多粒度特征进行相似度匹配。仅使用身份(ID)损失函数在 Market1501、CUHK03、DukeMTMC-reID 3 个公开数据集上进行实验。结果表明,所提算法具有一定的优势。

关键词 机器视觉; 深度学习; 人体关键点; 特征融合
中图分类号 TP392 **文献标识码** A **doi**: 10.3788/LOP56.201501

Posture-Guided and Multi-Granularity Feature Fusion for Person Reidentification

Zhang Liang[1,2], Che Jin[1,2*]

[1] *School of Physics and Electronic-Electrical Engineering, Ningxia University, Yinchuan, Ningxia* 750021, *China*;
[2] *Key Laboratory of Intelligent Sensing for Desert Information, Ningxia University, Yinchuan, Ningxia* 750021, *China*

Abstract In the person reidentification system, the retrieved person image will have large posture differences, complex changes in perspectives, and misalignment of person images in the detection frame. In order to solve these problems, a reidentification algorithm is proposed, which can directly use the key point information of the human body for person image alignment and extract multi-granularity features based on this alignment. First, the posture prediction model is used to locate the key points of the human skeleton, and the person image is directly aligned according to the extracted skeleton key points, and then the multi-granularity features are extracted from the person image. The evaluation phase uses posture information combined with multi-granularity features for similarity matching. The experiment is carried out only using the identity(ID) loss function on the three public datasets of Market1501, CUHK03, and DukeMTMC-reID. The results show that the proposed algorithm has certain advantages.

Key words machine vision; deep learning; human body key point; feature fusion
OCIS codes 150.0155; 110.2970; 150.1135

1 引 言

行人再识别[1](Person ReID)通常可以看作是一个图像检索问题,即给定一张目标图像,在跨摄像头、跨时段拍摄的图像数据库中找出该目标的所有图像。由于拍摄角度、光照、姿势、视角、图像分辨率、相机设置、遮挡和背景杂波的影响,可能会导致同一行人的不同图像有显著不同,因此行人再识别在计算机视觉中仍然是一项具有挑战性的任务。

行人再识别可以分为传统算法和深度学习两部分。传统算法主要通过人工设计特征并提取,然后通过相似性度量函数进行相似度度量。作为传统学

收稿日期: 2019-03-20; **修回日期**: 2019-04-07; **录用日期**: 2019-04-26
基金项目: 国家自然科学基金(61861037)
* **E-mail**: koalache@126.com

习算法的基准，Liao 等[1]提出一种局部最大发生(LOMO)特征和交叉视觉二次判别分析(XQDA)度量算法，能较好地解决摄像机视角变化的问题。Gray 等[2]对人物图像水平分割并提取颜色和纹理特征，提出基于高斯混合模型拟合的区域特征提取方法，克服了自然遮挡和位姿的变化。传统算法中人工提取特征的算法依然具有局限性，识别率比较低。随着深度学习[3]的兴起，涌现出许多基于深度学习的优秀算法，这些算法不仅取得了较高的识别率，而且泛化能力也得到进一步的提高。Zheng 等[4]基于特征空间的分布以及样本点之间的关系提出将验证损失和分类损失结合的算法，但是该算法没有考虑到图像对齐以及行人图像的局部特征。Zhao 等[5]利用人体关键点作为先验知识，将人体分割成几个固定的刚体结构，加强局部特征的学习，但是没有考虑人物图像以及背景特征的完整性，导致识别的精度不高。徐龙壮等[6]提出一种多尺度卷积特征融合的行人再识别算法，包含全局和多尺度的局部特征，但是没有考虑到图像的对齐。陈兵等[7]提出通过加强对样本的监督信息的利用和提升对样本间距离关系的约束，通过构造难分的负样本对和引入距离门限来改进网络。Sun 等[8]提出一种对特征图水平划分，融合多粒度特征完成行人再识别任务的方法，但是没有考虑到行人图像的不对齐问题。Zhang 等[9]提出基于最短路径(SP)距离的自动对齐模型，在不需要额外信息的情况下可以自动对齐局部特征。Liu 等[10]利用空间信息网络和运动信息网络提取全局特征和光流图特征，并融合运动信息的序列图像特征提高行人再识别的准确度。随着深度学习算法的发展，数据集样本不足的问题越来越明显。为了解决数据扩充问题，Ma 等[11]提出一个将前景、背景和姿态信息考虑在内的多分支的重构网络，利用高斯噪声加强生成对抗训练，融合多分支网络的特征，将生成的图像加入到行人再识别的数据库中。Liu 等[12]提出一种基于姿态迁移的 ReID 框架，通过引入 Pose 样本库，借助生成对抗网络(GAN)网络进行多姿态标签样本生成，用于辅助训练，并在传统的 GAN 网络的基础上，提出一个引导的子网络，使生成更加满足 ReID 的样本。

本文针对行人图像的不对齐问题，提出一种姿态引导与多粒度特征融合的行人再识别算法。在不增大样本的基础上利用位姿信息实现人物图像对齐，对图像提取多粒度特征并融合。研究本模型在 Market1501、CUHK03、DukeMTMC-reID 数据集上不同尺度下的表现效果。实验结果表明，本算法在这 3 个数据集上均能取得很好的效果。

2 网络架构

2.1 网络框架

网络框架如图 1 所示。将数据集中的行人图像输入姿态预估(PAF)网络完成对人体姿态信息的采集，经过该网络的图像将携带人体关键点信息。其次，将带有姿态信息的人物图像输入到多粒度特征提取网络(MGTN)，根据姿态网络定位的人体关键点将人物图像划分为几个水平区域，实现不同人物图像的对应部件对齐。根据姿态信息完成多粒度特征的提取与相似度匹配，预测行人的身份(ID)信息，最终实现行人再识别任务。

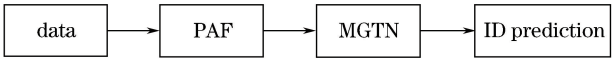

图 1 网络架构

Fig. 1 Network architecture

2.2 PAF 网络架构

精确定位是实现特征提取与匹配的基础，考虑到人物图像的不对齐问题，如果直接将两幅图像进行相似度匹配，可能会产生较大的误差，而人体关键点可以实现跨图像之间的对齐。因此迁移 PAF 网络[13]用于数据集的预处理，PAF 网络将人体分为 18 个部分，每部分的计算符合高斯分布，根据高斯分布图定位人体的 18 个关键点，网络架构如图 2 所示。

利用 VGG-19 模型的前 10 层对图像进行微调处理，得到一组特征映射 $\boldsymbol{F}$，特征 $\boldsymbol{F}$ 通过一个连续的 $T\in\{1,2,\cdots,t\}$ 阶段网络，其中 t 表示网络的总阶段数。网络的每个阶段包含两个分支，上层分支用来预测身体关键点的置信度映射图集 $\boldsymbol{S}=(\boldsymbol{S}^{(1)},\boldsymbol{S}^{(2)},\cdots,\boldsymbol{S}^{(j)},\cdots,\boldsymbol{S}^{(J)})$，其中 $\boldsymbol{S}^{(j)}$ 为第 j 个阶段的置信度输出，J 为上层分支的总阶段数；下层分支用来预测局部关联字段 $\boldsymbol{L}=(\boldsymbol{L}^{(1)},\boldsymbol{L}^{(2)},\cdots,\boldsymbol{L}^{(c)},\cdots,\boldsymbol{L}^{(C)})$，其中，$\boldsymbol{L}^{(c)}$ 为第 c 个阶段的关联向量，C 为下层分支的总阶段数。网络的第一阶段网络产生一组置信度 $\boldsymbol{S}^{(1)}=\rho^{(1)}(\boldsymbol{F})$ 和 $\boldsymbol{L}^{(1)}=\phi^{(1)}(\boldsymbol{F})$，其中 ρ 为置信度的函数，ϕ 为关联向量的函数。在后续的每个阶段，将前一阶段中两个分支的预测置信度与原始图像的特征 $\boldsymbol{F}$ 融合得到 $\boldsymbol{S}^{(t)}$，$\boldsymbol{L}^{(t)}$，分别表示为

$$\boldsymbol{S}^{(t)}=\rho^{(t)}(\boldsymbol{F},\boldsymbol{S}^{(t-1)},\boldsymbol{L}^{(t-1)}),\forall t\geqslant 2, \quad (1)$$

$$\boldsymbol{L}^{(t)}=\phi^{(t)}(\boldsymbol{F},\boldsymbol{S}^{(t-1)},\boldsymbol{L}^{(t-1)}),\forall t\geqslant 2。 \quad (2)$$

对 $\boldsymbol{S}^{(t-1)}$，$\boldsymbol{L}^{(t-1)}$，$\boldsymbol{F}$ 反复迭代直至收敛，网络的损失包括置信度损失和关联字段损失，可以表示为

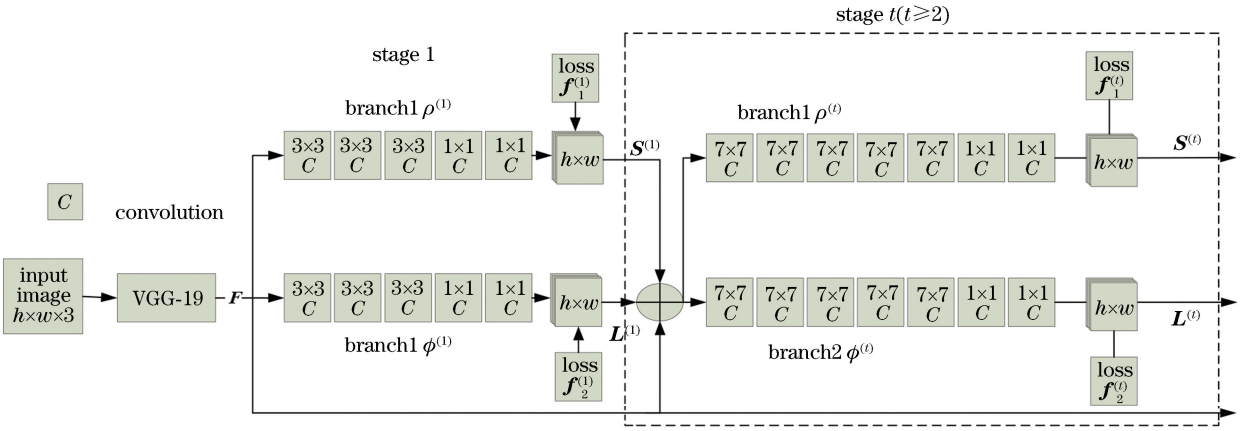

图 2 PAF 网络架构

Fig. 2 PAF network architecture

$$f=\sum_{t=1}^{T}(f_S^{(t)}+f_L^{(t)}), \tag{3}$$

式中，$f_S^{(t)}$，$f_L^{(t)}$ 为置信度与关联字段在 t 阶段的损失，T 为总的阶段数(总部位数)。

置信度和关联字段的损失分别为

$$f_S^{(t)}=\sum_{j=1}^{J}\sum_{p}\boldsymbol{W}(\boldsymbol{p})\cdot\|\boldsymbol{S}_j^{(t)}(\boldsymbol{p})-\boldsymbol{S}_j^{*}(\boldsymbol{p})\|_2^2, \tag{4}$$

$$f_L^{(t)}=\sum_{c=1}^{C}\sum_{p}\boldsymbol{W}(\boldsymbol{p})\cdot\|\boldsymbol{L}_c^{(t)}(\boldsymbol{p})-\boldsymbol{L}_c^{*}(\boldsymbol{p})\|_2^2, \tag{5}$$

式中，$\boldsymbol{p}$ 为图像中任意像素点，$\boldsymbol{S}_j^{(t)}(\boldsymbol{p})$为在 t 阶段的置信度响应，$\boldsymbol{L}_c^{(t)}(\boldsymbol{p})$为在 t 阶段的关联向量。

采用 L_2 损失消除预测值和真实值之间的误差。如果关键点未被准确标注，会导致 $\boldsymbol{S}_j^{*}(\boldsymbol{p})$和 $\boldsymbol{L}_c^{*}(\boldsymbol{p})$为零。因此引入 $\boldsymbol{W}(\boldsymbol{p})=0$ 解决关键点未被准确标注的问题，即 $\boldsymbol{p}$ 点为关键点但却未被标注。

$\boldsymbol{S}_j^{*}(\boldsymbol{p})$表示人体上第 j 个部位的分布置信度，部位 j 中 $\boldsymbol{p}\in R^2$(R 表示图像中所有像素的集合)的置信度响应表示为

$$\boldsymbol{S}_j^{*}(\boldsymbol{p})=\exp\left(-\frac{\|\boldsymbol{p}-\boldsymbol{x}_j\|_2^2}{\sigma^2}\right), \tag{6}$$

式中，σ 控制着传播速度，$\boldsymbol{x}_j$ 为第 j 个部分的关键点坐标。

置信度图是在像素级别上操作且符合高斯分布，因此高斯响应最大值的坐标即关键点坐标，可以表示为

$$\boldsymbol{P}_i=[x_i,y_i]=\underset{x\in[1,x],y\in[1,y]}{\operatorname{argmax}}\boldsymbol{S}_j^{*}(\boldsymbol{p}), \tag{7}$$

式中，x_i，y_i 为第 i 关键点的横纵坐标，x，y 为图像最大的宽高像素值。

根据置信度 $\boldsymbol{S}$，可以完成人体关键点的定位，而 $\boldsymbol{L}_C$ 的作用是将关键点以特定的方式连接，由文献[13]可得。

$\boldsymbol{L}_c^{*}(\boldsymbol{p})$为点 $\boldsymbol{p}$ 是否在人体部位 c 的响应，表示为

$$\boldsymbol{L}_c^{*}(\boldsymbol{p})=\begin{cases}\boldsymbol{v},\boldsymbol{p}\in c\\ \boldsymbol{0},\text{ otherwise}\end{cases}。 \tag{8}$$

$\boldsymbol{v}$ 由 c 上的两个点 $\boldsymbol{x}_{j2}$，$\boldsymbol{x}_{j1}$确定，表示为

$$\boldsymbol{v}=(\boldsymbol{x}_{j2}-\boldsymbol{x}_{j1})/\|\boldsymbol{x}_{j2}-\boldsymbol{x}_{j1}\|_2。 \tag{9}$$

点 $\boldsymbol{p}$ 的阈值条件可以表示为

$$0\leqslant\boldsymbol{v}\cdot(\boldsymbol{p}-\boldsymbol{x}_{j1})\leqslant l_c \text{ 和 } |\boldsymbol{v}_{\perp}\cdot(\boldsymbol{p}-\boldsymbol{x}_{j1})|\leqslant\sigma_l, \tag{10}$$

式中，$\boldsymbol{v}_{\perp}$为垂直于 $\boldsymbol{v}$ 的特征向量，σ_l 为部位 c 宽度的像素距离，长度的像素距离为 $l_c=\|\boldsymbol{x}_{j2}-\boldsymbol{x}_{j1}\|_2$。

行人图像经过 PAF 网络效果如图 3 所示。

2.3 MGTN 架构

考虑到网络的竞争性能以及相对简洁的架构，

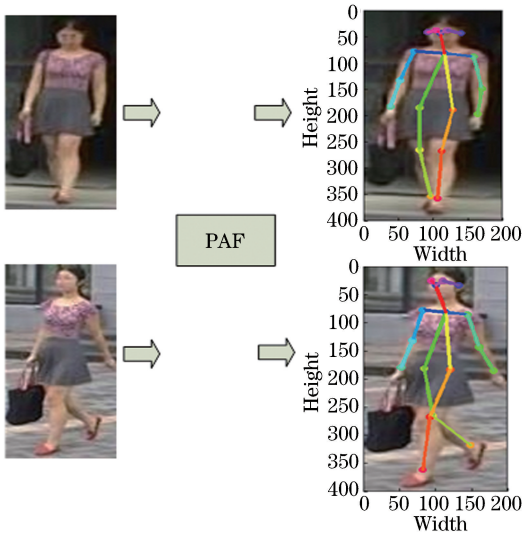

图 3 PAF 效果图

Fig. 3 PAF rendering

使用 ResNet50[14] 作为基准网络，并对基准网络做细微修改。保留全均值池化层(GAP)之前的网络结构，之后的网络层均被抛弃，网络架构如图 4 所示。

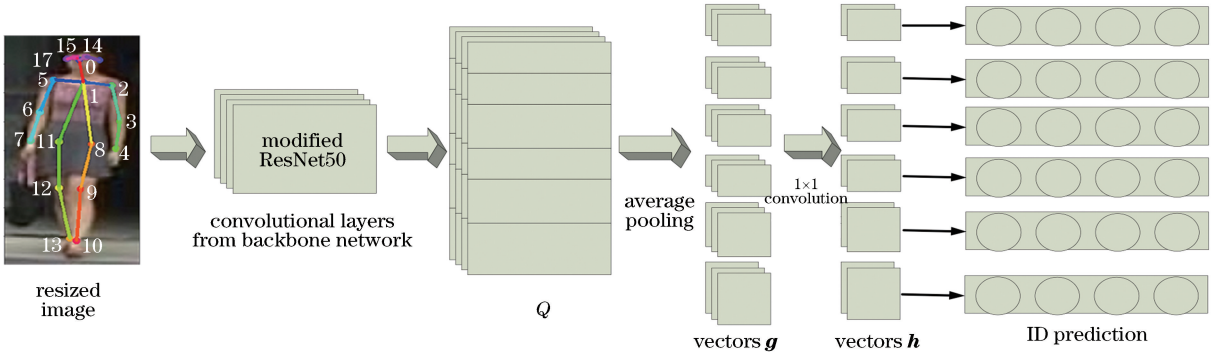

图 4 MGTN 架构

Fig. 4 MGTN architecture

一幅行人图像经过基准网络后得到特征图 Q。根据姿态网络定位的关键点将特征图 Q 划分为 M 个水平区域[8]，实现网络的多粒度特征提取。将该水平区域内所有的列向量(沿通道的激活向量)池化为单个部件级别的列向量 $\boldsymbol{g}_i(i=1,\cdots,m)$，得到 M 个 2048 维的特征向量。对特征向量 $\boldsymbol{g}$ 降维得到 256 维的特征向量 $\boldsymbol{h}$。将特征向量 $\boldsymbol{h}$ 输入到一个由全连接层和 softmax 层组成的分类子网络，用于预测输入行人的 ID，可以表示为

$$\overline{Y}_i=\mathrm{softmax}(\boldsymbol{h})_i=\frac{\exp(\boldsymbol{h}_i)}{\sum_{n=1}^{N}\exp(\boldsymbol{h}_n)},\tag{11}$$

式中，$\boldsymbol{h}_i$ 为第 i 个水平区域的特征向量，N 表示总类别数。水平区域均采用交叉熵损失作为对应区域的目标损失函数，可以表示为

$$l_{\mathrm{loss}_{\boldsymbol{h}_i}}=\sum_i[Y_i\cdot\ln(\overline{Y}_i)+(1-Y_i)\cdot\ln(1-\overline{Y}_i)],\tag{12}$$

式中，Y_i 为非 i 类的概率。

训练过程中，网络是最小化 p 个水平区域的交叉熵损失总和，可以表示为

$$L_{\mathrm{Loss}}=\sum_1^p l_{\mathrm{loss}_{\boldsymbol{h}_i}}。\tag{13}$$

在测试阶段，融合多粒度特征得到最终的特征描述子 $G=\{\boldsymbol{g}_1,\boldsymbol{g}_2,\cdots,\boldsymbol{g}_m\}$ 和 $H=\{\boldsymbol{h}_1,\boldsymbol{h}_2,\cdots,\boldsymbol{h}_m\}$，用来预测行人 ID。在实验中观察到使用 G 的精度要高于使用 H 的精度，这恰恰说明更多的细节特征被利用，但是耗时长。

2.4 参数设置

1) 输入图像的宽高比为 1∶3，输入图像分辨率为 128×384；

2) Q 划分太多的水平区域会损害特征的辨识能力。经过多次实验将网络的优化参数设置为 $M=4$ 或 $M=6$，如图 5 所示，图中的数字表示关键点信息；

3) 描述子 G 设置为 $p\times2048$ 维，描述子 H 设置为 $p\times256$ 维。

3 实验结果

行人再识别算法在公开的数据集 Market1501、CUHK03、DukeMTMC-reID 上进行实验验证，取得不错的效果，表 1 列出了数据集的详细信息。

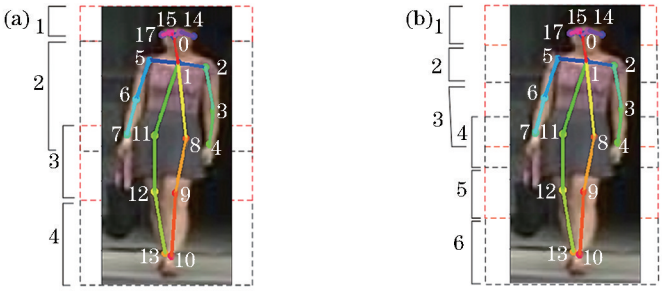

图 5 Part 分割图。(a) 4 parts；(b) 6 parts

Fig. 5 Diagram of part segmentation. (a) 4 parts; (b) 6 parts

表 1 数据集的详细信息

Table 1 Details of the datasets

Dataset	ID	Train ID	Test ID	Bbox	Bbox /ID	Cam
Market1501	1501	751	750	32668	21.8	6
CUHK03-NP	1467	767	700	14097	9.6	2
DukeMTMC-reID	1404	702	702	36411	25.9	8

3.1 数据集

Market1501：该数据集包含 1501 个行人的 32668 幅带标签的边界框。每个身份的图像最多由 6 台摄像机拍摄。根据数据集的设置，数据集被分为训练集和测试集，训练集包含 751 个行人的 12936 个裁剪图像，测试集包含 750 个行人的 19732 个裁剪图像，边界框直接由可变形零件模型(DPM)[15]检测而不是使用手绘的边界框，这更接近于真实的场景，在测试中，使用 3368 张包含 750 个行人的手绘图像作为查询集来识别测试集上的正确行人身份。对于每个查询图像，旨在从 19732 个候选人图像中检索出正确匹配的图像，所有的实验均在 single-shot 模式下进行。

CUHK03：该数据集包含 1467 个行人的 14709 张裁剪图像。每个行人由两个相机所拍摄，且每个行人在每个相机中平均有 4.8 张图片。数据集提供了两种边界框，分别是手动标注的边界框和 DPM 检测到的边界框。利用 DPM 检测到的边界框对模型进行评估，更加接近于现实情况。不同于以往的数据集划分为 1367 个行人的训练集和 100 个行人的测试集。本文采用新提出的数据集划分方法[16]，即为 CUHK03-NP，该数据集划分 767 个行人的训练集和 700 个行人的测试集，这样设置训练集的样本数目很少，对模型来说具有很大的挑战性。重复 20 次，取平均值。

DukeMTMC-reID：该数据集包含 1404 个行人的 36411 张图像，由 8 个摄像头采集数据信息，1404个行人出现在大于2个摄像头视角下，有408个行人仅出现在 1 个摄像头视角下，随机选取 702 个行人图像作为训练集，剩余的 702 个行人图像作为测试集。

3.2 评价指标

使用累积匹配特征曲线(CMC)和平均精度均值(mAP)两个指标来衡量模型的性能。

3.3 实验数据

设置 $M=4,6$，采用描述子 H 在 Market1501 上的实验数据如表 2 和表 3 所示，Rank-1 和 mAP 如图 6 所示。表 2～7 中的 w、h 分别代表图像的宽度与高度。

表 2 $M=4$ 和描述子 H 在 Market1501 的实验数据

Table 2 Experimental data of $M=4$ and descriptor H in Market1501

$w\times h$	Rank-1 /%	Rank-5 /%	Rank-10 /%	mAP /%
64×192	84.4	93.4	95.5	66.3
96×288	87.4	95.1	96.9	71.3
128×384	89.0	95.8	97.6	73.6
160×480	90.3	96.5	97.6	74.6
192×576	90.2	96.6	97.8	74.3

表 3 $M=6$ 和描述子 H 在 Market1501 的实验数据

Table 1 Experimental data of $M=6$ and descriptor H in Market1501

$w\times h$	Rank-1 /%	Rank-5 /%	Rank-10 /%	mAP /%
64×192	85.3	93.9	96.3	68.8
96×288	88.4	95.2	97.0	72.9
128×384	89.7	95.5	96.9	73.8
160×480	91.0	96.1	97.7	74.8
192×576	90.3	96.4	97.6	74.1

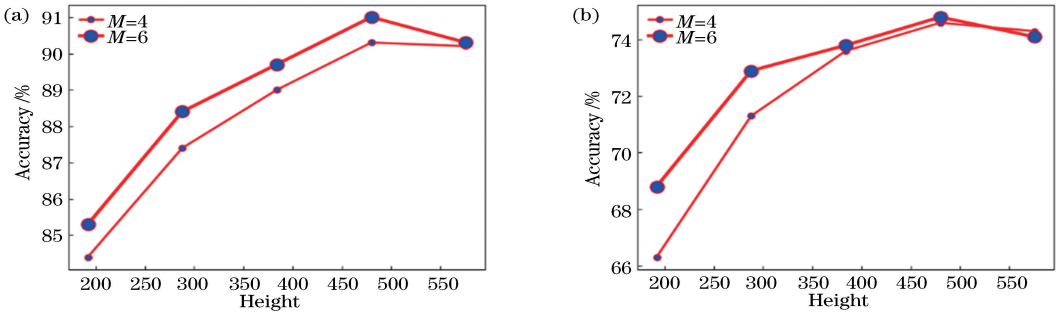

图 6 描述子 H 在 $M=4,6$ 不同分辨率的效果图(Market1501)。(a) Rank-1；(b) mAP

Fig. 6 Effect of descriptor H with different resolutions at $M=4,6$ (Market1501). (a) Rank-1; (b) mAP

从图 6 可直观地看出，采用 160×480 分辨率的 Rank-1，mAP 稍高于 128×384，但是耗时长。加入学习率衰减，动量平均且采用描述子 G 在 $M=6$ 时，128×384 分辨率的表现效果要高于 160×480。实验数据如表 4 所示，Rank-1 和 mAP 如图 7 所示。

图 6 和图 7 可以看出，采用描述子 G 比 H 的效果要好，也就是说更多的细节特征被利用。因此在 CUHK03-NP 上仅仅利用描述子 G 来进行实验数据分析，如表 5 和表 6 所示，Rank-1 和 mAP 如图 8 所示。

表 4 $M=6$ 和描述子 G 在 Market1501 实验数据

Table 4 Experimental data of $M=6$ and descriptor G in Market1501

$w\times h$	Rank-1 /%	Rank-5 /%	Rank-10 /%	mAP /%
64×192	87.7	94.9	96.4	69.7
96×288	91.5	96.6	97.6	75.8
128×384	93.2	97.1	98.1	78.3
160×480	92.4	97.0	97.8	77.7
192×576	92.5	97.3	98.4	77.2

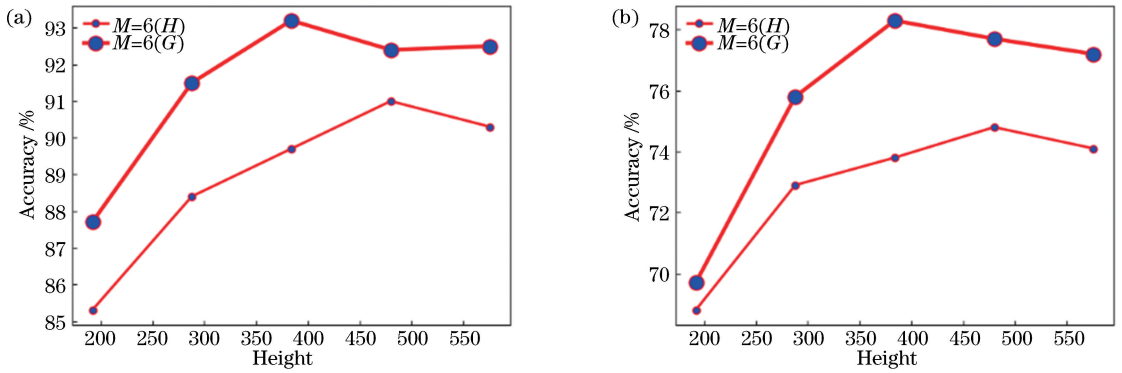

图 7 H 和 G 在 $M=6$ 的对比效果图(Market1501)。(a) Rank-1；(b) mAP

Fig. 7 Comparison of H and G at $M=6$ (Market1501). (a) Rank-1; (b) mAP

表 5 $M=4$ 和描述子 G 在 CUHK03-NP 的实验数据

Table 5 Experimental data of $M=4$ and descriptor G in CUHK03-NP

$w\times h$	Rank-1 /%	Rank-5 /%	Rank-10 /%	mAP /%
64×192	47.9	68.4	76.6	47.5
96×288	52.8	73.1	81.6	52.1
128×384	59.3	78.4	83.9	57.2
160×480	60.4	77.7	84.6	58.1
192×576	51.3	72.2	79.6	50.5

表 6 $M=6$ 和描述子 G 在 CUHK03-NP 的实验数据

Table 6 Experimental data of $M=6$ and descriptor G in CUHK03-NP

$w\times h$	Rank-1 /%	Rank-5 /%	Rank-10 /%	mAP /%
64×192	47.8	68.8	77.5	47.1
96×288	55.3	74.5	82.3	54.5
128×384	59.7	77.1	84.3	57.7
160×480	56.7	75.1	82.4	54.3
192×576	56.4	73.9	82.5	54.1

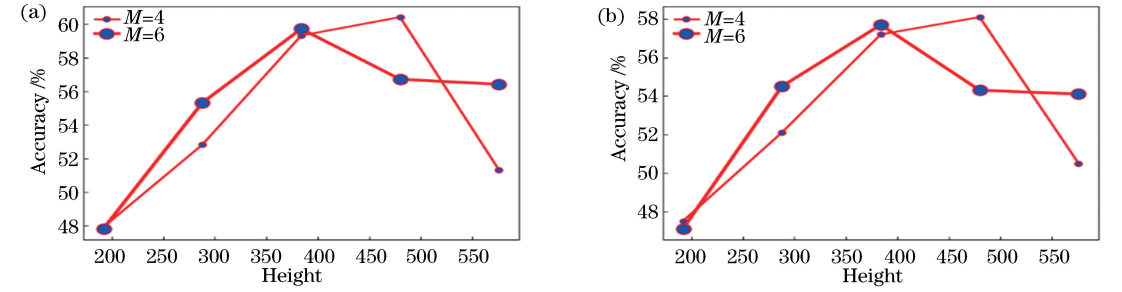

图 8 描述子 G 在 $M=4,6$ 下不同分辨率的效果图(CUHK03-NP)。(a) Rank-1；(b) mAP

Fig. 8 Effect of descriptor G with different resolutions at $M=4,6$ (CUHK03-NP). (a) Rank-1; (b) mAP

由图 6 和图 7 可分析，在 Market1501 数据集上采用分辨率为 128×384，描述子 G 且 $M=6$ 的效果最好，图 8 可以看出在 CUHK03-NP 数据集上的数据显示，在 $M=4$ 且分辨率为 160×480 的效果要稍微强于 $M=6$ 分辨率为 128×384 的效果，但是耗时长。这一点恰恰说明数据集之间存在差异性，也说明该模型的泛化能力强。所以在 DukeMTMC-reID 数据集上仅仅采用描述子 G 且 $M=6$ 条件下的实验数据，如表 7 所示，Rank-1 和 mAP 如图 9 所示。

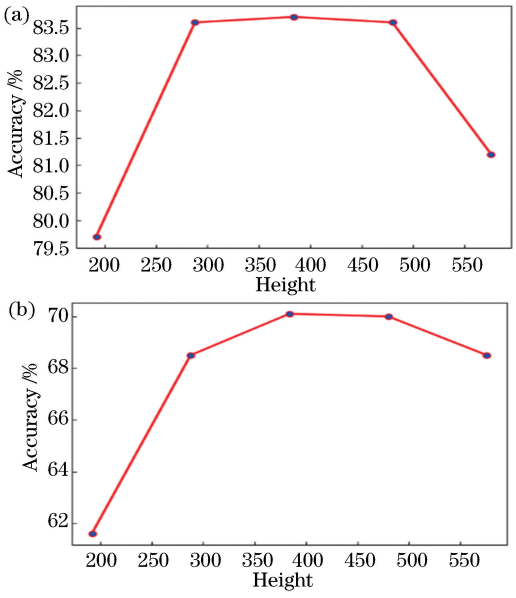

图 9 描述子 G 在 $M=6$ 下不同分辨率的效果图(DukeMTMC-reID)。(a) Rank-1;(b) mAP

Fig. 9 Effect of descriptor G with different resolutions at $M=6$ (DukeMTMC-reID). (a) Rank-1; (b) mAP

表 7 $M=6$ 和描述子 G 在 DukeMTMC-reID 的实验数据

Table 7 Experimental data of $M=6$ and descriptor G in DukeMTMC-reID

$w\times h$	Rank-1/%	Rank-5/%	Rank-10/%	mAP/%
64×192	79.7	88.5	91.1	61.6
96×288	83.6	92.0	93.6	68.5
128×384	83.7	92.1	94.1	70.1
160×480	83.6	91.7	93.7	70.0
192×576	81.2	90.6	92.7	68.5

3.4 实验结果

本文提出的网络模型与现有的网络模型进行比较,结果如表 8~10 所示。

表 8 Market1501 数据集中的方法比较

Table 8 Comparison ofmethods in Market1501 dataset

Method	Rank-1 /%	mAP /%
BoW+Kissme[17]	20.76	44.42
LOMO+XQDA[1]	72.58	51.96
SVDNet[18]	82.3	62.1
PAN [19]	82.8	63.4
AACN[20]	85.90	66.87
AOS[21]	86.5	70.4
PSE[22]	87.7	69.0
Part-aligned[23]	88.8	74.5
HA-CNN[24]	91.2	75.7
PCB[8]	92.4	77.3
Ours(H)	91.0	74.8
Ours(G)	93.2	78.3

表 9 CUHK03-NP 数据集中的方法比较

Table 9 Comparison ofmethods in CUHK03-NP dataset

Method	Rank-1 /%	mAP /%
BoW+Kissme[17]	6.4	6.4
LOMO+XQDA[1]	12.8	11.5
PAN[19]	36.3	34.0
MutiScale[25]	40.7	37.0
SVDNet[18]	41.5	37.3
HA-CNN[24]	41.7	38.6
AOS[21]	47.1	43.3
MLFN[26]	52.8	47.8
DaRe[27]	55.1	51.3
PCB[8]	61.3	54.2
Ours(G)	60.4	58.1

表 10 DukeMTMC-reID 数据集中的方法比较

Table 10 Comparison of methods in DukeMTMC-reID dataset

Method	Rank-1 /%	mAP /%
BoW+Kissme[17]	25.13	12.17
LOMO+XQDA[1]	30.75	17.04
PAN[19]	71.59	51.51
SVDNet[18]	76.7	56.8
AACN[20]	76.84	59.25
PSE[22]	79.8	62.0
HA-CNN[24]	80.5	63.8
MLFN[26]	81.2	62.8
PCB[8]	83.3	69.2
Part-aligned[23]	84.4	69.3
Ours(G)	83.7	70.1

由表 8~10 可以看出,在 CUHK03 和 DukeMTMC-reID 的 Rank-1 稍微低于 PCB 和 Part-aligned,但是 mAP 的指标要高于这两种算法。在行人再识别系统中,指标 mAP 要比 Rank-1 更具有说服力。本文算法仅使用 ID 损失便取得不错的效果,具有一定的优势。

4 结 论

现有的很多工作都是直接对图像进行特征提取与相似度匹配,这就造成特征的不对齐匹配,使得属于同一行人图像的特征相似度很低。本算法充分考虑人物图像的不对齐问题,利用更加精确的人体姿态信息实现人物图像对齐与多粒度特征提取,强化局部细节特征进而得到更强的判别特征。结合图像姿态信息与多粒度特征实现两幅图像的相似度匹配。实验证明,该模型在 Market1501、CUHK03、DukeMTMC-reID 数据集上表现不错。可见人体姿态信息可以优化网络并可进一步提高行人再识别的

识别精度。

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] Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[M]// Forsyth D, Torr P, Zisserman A. Computer vision-ECCV 2008. Lecture notes in computer science. Berlin, Heidelberg: Springer, 2008, 5302: 262-275.

[3] Xu L, Zhao H T, Sun S Y. Monocular infrared image depth estimation based on deep convolutional neural networks[J]. Acta Optica Sinica, 2016, 36(7): 0715002.
许路, 赵海涛, 孙韶媛. 基于深层卷积神经网络的单目红外图像深度估计[J]. 光学学报, 2016, 36(7): 0715002.

[4] Zheng Z D, Zheng L, Yang Y. A discriminatively learned CNN embedding for person re-identification [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2018, 14(1): 13.

[5] 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. New York: IEEE, 2017: 907-915.

[6] Xu L Z, Peng L. Reidentification based on multiscale convolutional feature fusion [J]. Laser & Optoelectronics Progress, 2019, 56(14): 141504.
徐龙壮, 彭力. 基于多尺度卷积特征融合的行人重识别[J]. 激光与光电子学进展, 2019, 56(14): 141504.

[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] 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, 11208: 501-518.

[9] Zhang X, Luo H, Fan X, *et al*. AlignedReID: surpassing human-level performance in person re-identification[J/OL]. (2017-11-22)[2019-03-05]. https://arxiv.org/abs/1711.08184.

[10] Liu H, Jie Z Q, Jayashree K, *et al*. Video-based person re-identification with accumulative motion context[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(10): 2788-2802.

[11] Ma L Q, Sun Q R, Georgoulis S, *et al*. Disentangled person image generation [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 99-108.

[12] Liu J X, Ni B B, Yan Y C, *et al*. Pose transferrable person re-identification [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 4099-4108.

[13] Cao Z, Simon T, Wei S E, *et al*. Realtime multi-person 2D pose estimation using part affinity fields[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI. New York: IEEE, 2017: 1302-1310.

[14] 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.

[15] Lin L, Wang X L, Yang W, *et al*. Discriminatively trained and-or graph models for object shape detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(5): 959-972.

[16] Zhong Z, Zheng L, Cao D L, *et al*. Re-ranking person re-identification with k-reciprocal encoding[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI. New York: IEEE, 2017: 3652-3661.

[17] 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.

[18] 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.

[19] Zheng Z D, Zheng L, Yang Y. Pedestrian alignment network for large-scale person re-identification [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018.

[20] Xu J, Zhao R, Zhu F, *et al*. Attention-aware compositional network for person re-identification[C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 2119-2128.

[21] Huang H J, Li D W, Zhang Z, *et al*. Adversarially occluded samples for person re-identification [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 5098-5107.

[22] Sarfraz M S, Schumann A, Eberle A, *et al*. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 420-429.

[23] Suh Y, Wang J D, Tang S Y, *et al*. Part-aligned bilinear representations for person re-identification [M] // Ferrari V, Hebert M, Sminchisescu C, *et al*. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11218: 418-437.

[24] Li W, Zhu X T, Gong S G. Harmonious attention network for person re-identification[C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 2285-2294.

[25] 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.

[26] Chang X B, Hospedales T M, Xiang T. Multi-level factorisation net for person re-identification [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 2109-2118.

[27] 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.