

文章编号: 2095-2163(2021)12-0122-05

中图分类号: TP391.3

文献标志码: A

# 基于局部特征融合的行人重识别方法

涂园园, 贺松, 姚绍华

(贵州大学 大数据与信息工程学院, 贵阳 550025)

**摘要:** 随着智能化时代的来临, 社会智能监控设备全面覆盖, 行人重识别成为极具挑战的研究课题。行人重识别的核心在于设计深度学习网络对图像提取强判别力的特征, 本文针对局部特征的重要性提出一种局部特征融合的行人重识别模型。将一个卷积块接出2条独立支路, 分别接入局部特征提取模块和批特征擦除模块, 最后和全局特征提取支路进行特征融合获得高细粒度的特性, 并采用联合损失函数训练网络。在 Market1501 数据集上验证所提方法具有有效性,  $mAP$  达到 84.12%,  $Rank - 1$  达到 95.06%。

**关键词:** 局部特征; 批特征擦除; 深度学习; 损失函数

## Pedestrian re-identification method based on local feature fusion

TU Yuanyuan, HE Song, YAO Shaohua

(School of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

**【Abstract】** With the advent of the intelligent era and the full coverage of social intelligent monitoring equipment, pedestrian re-recognition has become a very challenging research topic. The core of pedestrian re-recognition is to design a deep learning network to extract strong discriminant features from the image. Aiming at the importance of local features, this paper proposes a pedestrian re-recognition model based on local feature fusion. A convolution block is connected to two independent branches, which are respectively connected to the local feature extraction module and the batch feature erasure module. Finally, the features are fused with the global feature extraction branch to obtain the characteristics of high fine granularity, and the joint loss function is used to train the network. The effectiveness of the proposed method is verified on market1501 dataset, with  $mAP$  of 84.12% and  $rank - 1$  of 95.06%.

**【Key words】** local features; batch feature erasure; deep learning; loss function

## 0 引言

随着互联网技术的不断进步, 行人重识别在视频监控、智能商业、智能安防等领域具有广泛的应用价值和研究意义。行人重识别是指在已知来源与无重叠摄像机拍摄到的图片或视频的资料库中识别出已知行人的技术, 即一个摄像头(a)拍摄到某行人的图像或视频, 利用计算机视觉技术在摄像头(b)拍摄的图像或视频中准确找到该行人。在实际监控环境中, 由于光照、摄像头视角、行人姿态、遮挡及背景杂乱等复杂干扰, 出现不同身份的行人图片很相似, 而同一身份的行人图片却差异大的问题, 如何对行人图片提取鲁棒性更强的特征是当前行人重识别研究亟需解决的问题。

行人重识别技术流程主要有两个步骤: 第一步, 利用特征表达方法从目标图像中提取鲁棒的行人图像特征; 第二步, 根据提取到的图像特征对行人图像

进行相似度度量, 行人图像对经过该距离度量函数能够准确反映出无重叠视域摄像头下的行人身份差异<sup>[1]</sup>。行人重识别研究主要方法有基于传统方法和基于深度学习方法, 传统的行人重识别方法是从基于特征设计和基于距离度量学习两方面研究。基于特征设计的方法是通过设计鲁棒的、隐含的特征行人外观模型, 提取图像任务当中的底层特征, 比如颜色、纹理、光建点等特征, 来减少光照条件、视角变化、姿态不同、遮挡等因素对图像对的影响, 而且对身份异同具有识别能力。2006年, Gheissari 等人首次提出利用 H-S 直方图和边缘直方图提取行人特征<sup>[2]</sup>; 2008年, Douglas Gray 等人针对摄像头视角不同, 提出在不同的摄像头中运用特征学习的方式来挑选一组最优的特征, 进而提高行人图像特征的整体鲁棒性<sup>[3]</sup>; 2015年, Liao 等人针对视觉变化问题, 提出一种局部最大发生率 (Local Maximal Occurrence, LOMO) 的特征表达算法, 首先在水平条

**作者简介:** 涂园园(1995-), 女, 硕士研究生, 主要研究方向: 数字图像处理与行人重识别; 姚绍华(1996-), 女, 硕士研究生, 主要研究方向: 机器视觉与图像处理; 贺松(1974-), 男, 硕士, 副教授, 主要研究方向: 图像处理与医疗大数据。

**通讯作者:** 贺松 Email: 1465433166@qq.com

收稿日期: 2021-09-01

带上提取 HSV 颜色直方图特征和 SILTP (Scale Invariant Local Ternary Pattern) 纹理直方图特征;其次,将水平条带上各个特征维度的最大出现频率特征作为该条带特征;最后,级联条带特征作为行人图像特征表达,取得了较好的识别效果<sup>[4]</sup>。而基于度量学习的方法是通过学习找到一个距离度量标准,用来计算图像对之间的距离或者相似度,使计算结果能够更准确地反映行人身份异同,研究方向侧重于欧式距离、马氏距离为核心思想提出的各种度量方法。但是传统的识别方法识别性能不佳,在海量、行人类别多的大数据集中无法使用。基于深度学习的方法与传统的行人重识别不同,其把特征提取和度量学习两个模块联合在一个网络中训练。2013年,LeNet-5 的发明者 Lecun 基于 AlexNet 提出了 OverFeat,提出了学习边框的概念,引起了界内学者的研究热潮<sup>[5]</sup>;2014年,Yi 等人基于 Siamese 网络模型,把行人分割成 3 个重叠的局部块,每个局部块用两个卷积层和一个全连接层获得特征向量,再计算余弦相似度<sup>[6]</sup>;2015年,郑良等人提出第一个行人重识别的大数据集 Market1501 后,深度学习的性能逐年大幅度提高<sup>[7]</sup>。

由于学习全局特征模型不够全面表征行人,对于衣着、外观等全局特征相似的目标,局部细节信息往往是区分不同行人的关键。本文提出一种多支路的全局和局部特征学习模型,ResNeXt50 作为基础的骨干网络,采用两个分支发掘高质量的局部特征,第一个支路采取传统的水平分割得到局部特征,每个局部特征以加权求和的方式进行融和;第二个支路采用批特征擦除 (Batch Feature Model) 模型。度量距离时使用联合交叉熵损失函数和三元组损失函数,增加网络性能。本文所提网络模型在 Market1501 数据集上进行了实验,结果表明所提取的特征具有良好的表征能力和泛化能力。

## 1 基于局部特征融合的行人重识别网络模型

本文提出的网络模型结构如图 1 所示,主要有 ResNet-50 基本主干网络、全局特征提取分支模块、局部特征提取模块和批特性擦除模块组成,将 3 个模块的特征进行融合,最后采用交叉熵损失函数与三元组联合训练。ResNet-50 由 6 个小模块组成,Conv1 模块为卷积层,Conv2-Conv5 为残差模块,每个残差块包含多个卷积层、BN 层和激活函数<sup>[8]</sup>。本文对 ResNet-50 网络进行修改:把 Conv4 模块引出分成局部特征提取模块 ( $z_1$ ) 和批特性擦除模块

( $z_3$ ),之前的网络结构和参数保持不变,把 Conv5 层输出的特征经过全局平均池、降维得到全局特征  $f_2$ ,该分支 ( $z_2$ ) 用于监督局部特征提取模块和特性擦除模块学习。经过 3 个分支融合得到最终的行人特征,运用交叉熵损失函数与三元组联合训练。

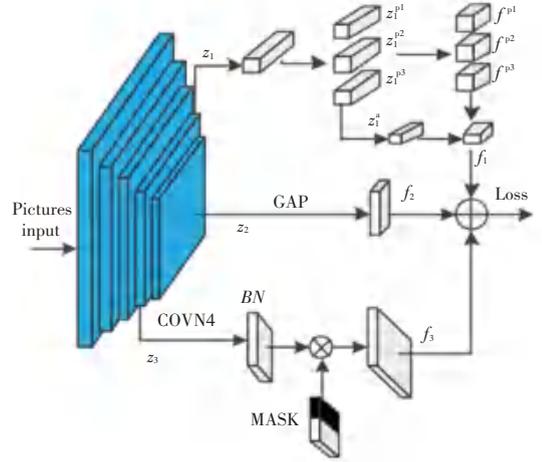


图 1 网络模型图

Fig. 1 Network model diagram

### 1.1 局部特征提取模块

如图 1 所示,由于 Conv4 所提取到的特征语义较高,对其水平分割得到 3 个局部特征提取支流 ( $Z_1^{p1}, Z_1^{p2}, Z_1^{p3}$ ),经过全局平均池化得到 3 个局部特征描述子 ( $f^{p1}, f^{p2}, f^{p3}$ ),由于图片存在不对齐现象,因此通过局部特征融合更能表达支流重要性,通过对 3 个支流经过  $3 \times 1$  卷积核、实例正侧化 IN (Instance Normalization) 和激活函数,得到局部特征权重向量  $Z_1^a$ ,对 3 个局部特征加权求和,得到融合后的特征  $f_1$ 。

### 1.2 批特性擦除模块

批特征擦除网络 (Batch Feature Erasing, BFE) 是由阿里智能实验室 Zuozhuo Dai 等人在 2018 年的 CVPR 会议中提出。批特征擦除网络对同一批次行人图像的相同区域进行随机擦除,从而迫使网络对剩余的细粒度特征学习,学习到更具有判别力的局部特征。BEF 网络结构有两个特征提取分支,Global Branch 和 Feature Erasing Branch<sup>[9]</sup>。Global Branch 提供了全局特征表示,同时对 Feature Erasing Branch 的训练进行监督,本文同样增加一条全局特征提取支流,对特征擦除分支和局部特征提取分支进行监督。根据设置擦除的矩形高度比  $h_e$  和宽度比  $w_e$ ,计算擦除矩形区域的高  $H_{BFE}$  和宽  $W_{BFE}$ ,式(1)所示:

$$\begin{aligned} H_{BFE} &= H \times h_e \\ W_{BFE} &= W \times w_e \end{aligned} \quad (1)$$

其中,  $H$  和  $W$  分别表示特征图的高和宽。

擦除后学习到具有判别力的局部特征。最后, 通过平均池化得到特征  $f_3$ 。

### 1.3 损失函数

本文的网络模型采用交叉熵损失函数 (Cross Entropy Loss) 与三元组损失函数 (Triplet Loss) 联合训练。交叉熵损失函数表示真实概率与预测概率分布之间的差异, 交叉熵的值越小, 说明模型预测效果越好, 常用于分类问题。模型训练过程中采用交叉熵损失函数, 不同的行人身份作为不同的类别标签<sup>[10]</sup>。对于一个含有  $M$  个样本训练的数据集  $D = \{(x_1, y_2), \dots, (x_i, y_i), \dots, (x_M, y_M)\}$  来说,  $x_i$  表示样本集中第  $i$  个样本,  $y_i$  表示该样本的标签, 如式(2)所示:

$$L_{CE} = -\frac{1}{M} \sum_{i=1}^M (y_i \cdot \log(p_i)) \quad (2)$$

其中,  $p_i$  表示第  $i$  个样本被分类为第  $j$  类标签的概率。

三元组损失函数可以发掘样本之间的相对相似度, 从而使正样本对之间的距离减小, 负样本对之间的距离增大。输入三幅图片到网络中, 正样本、负样本和锚样本, 其中正样本与锚样本为同一身份组成正样本对, 负样本与锚样本为不同身份组成负样本对。网络经过训练后使得同一身份的行人图片在特征空间的距离比不同身份的行人图片小<sup>[11]</sup>, 如公式(3)所示:

$$L_{triplet} = \max(d(a, p) - d(a, n) + m, 0) \quad (3)$$

其中,  $d(a, p)$  为锚样本和正样本之间的距离;  $d(a, n)$  为锚样本与负样本之间的距离;  $m$  为设定的阈值。

因此, 本文采用联合损失函数训练, 对网络进行更好的优化, 保证模型学习到具有判别力的特征, 如式(4)所示:

$$L_{loss} = L_{triplet} + L_{CE} \quad (4)$$

## 2 实验与分析

### 2.1 实验环境与参数设置

本实验环境为 64 位 windows10 操作系统, 2080Ti 显卡, 内存 64GB, 基于 python3.8 的深度学习框架 pytorch1.6 编程。初始化行人图片大小为  $258 \times 128$ , 利用 ImageNet 预训练得到初始权重, 批量大小为 32 张, 60 次迭代, 使用 ADAM 优化器更新梯度参数, 初始学习率为  $3.5 \times 10^{-4}$ 。

### 2.2 数据集与评价指标

Market1501 数据集是用 6 个不同的摄像头拍摄 1 501 个行人的 32 668 张图像, 其中训练集有 751 个行人身份, 共 12 936 张图片, 测试集有 750 个行人身份, 共 19 732 张图片。实验中采用行人重识别中的首位匹配率 Rank-1 和平均准确率 (mean Average Precision, mAP) 作为评价指标。

### 2.3 实验分析

#### 2.3.1 网络结构的消融

本文在 Market-1501 数据集上, 设置损失函数为交叉熵损失函数和三元组损失函数, 不同网络结构组合下的性能结果, 见表 1。首先, 对每个网络结构进行实验; 其次, 两个结构融合; 最后, 进行 3 个网络结构的实验。通过实验可知, 单个模块相比与融合模块精确程度低了 6 个百分点左右, 说明不同的模块放在同一个网络, 相比与原来的模块性能得到提升, 反映了不同模块融合具有互补的作用。在单模块实验中, 批特性擦除模块识别效果最好, 在 2 分支网络实验中批特征擦除模块与局部模块组合的网络结构最好, 可能是批特性擦除有助于网络学习注意力和弱特征区域的原因。最后 3 个模块同时实验, 在 Market-1501 数据集上 mAP 达到了 84.12%, Rank-1 达到了 95.16%, 表明 3 个模块在网络上相互监督学习, 提升网络性能。

表 1 不同模块在 Market-1501 数据集上的性能比较

Tab. 1 Performance comparison of different modules on Market-1501 dataset

网络结构	mAP / %	Rank - 1 / %
局部特征提取模块	75.58	86.28
全局学习模块	71.60	83.64
批特性擦除模块	77.50	88.24
局部特征提取模块+全局学习模块	79.19	91.36
局部特征提取模块+批特性擦除模块	82.32	93.61
全局学习模块+批特性擦除模块	81.68	92.52
局部特征提取模块+局部特征提取模块+批特性擦除模块	84.12	95.06

#### 2.3.2 损失函数的消融

本文在 Market-1501 数据集上, 设置在 3 个分支融合的网络结构下, 使用不同的网络损失函数得到网络性能结果, 见表 2。使用损失函数为交叉熵损失函数和三元组损失函数联合训练网络比单独使用三元组损失函数在 Market-1501 数据集上的 mAP 和 Rank-1 分别高出了 2.68% 和 1.29%, 可见本文使用联合损失函数一定程度上提高了模型的性能。

表 2 不同损失函数在 Market-1501 数据集上的性能比较

Tab. 2 Performance comparison of different loss functions on market-1501 dataset

损失函数	$mAP / \%$	$Rank - 1 / \%$
交叉熵损失函数	75.32	90.16
三元组损失函数	78.58	91.06
交叉熵损失函数+三元组损失函数	81.26	92.35

### 2.3.3 与其他方法比较

本文模型在 Market-1501 上与 5 个主流模型的性能比较, 见表 3。本文模型在 Market-1501 上  $mAP$  和  $Rank - 1$  分别达 84.12% 和 95.06%, 相比 Tri-Net,  $mAP$  和  $Rank - 1$  分别提升了 14.98% 和 10.14%, 与 BEF 相比, 本文在  $Rank - 1$  上高出了 0.66%, 与目前 DG-Net 优质模型相比,  $Rank - 1$  提高了 0.26%。

表 3 Market-1501 数据集上各方法比较

Tab. 3 Comparison of methods on market-1501 dataset

方法	$mAP / \%$	$Rank - 1 / \%$
Tri-Net	69.14	84.92
PCB	77.4	92.3
PCB+RPB	81.6	93.8
BEF	85.0	94.4
DG-Net	86.0	94.8
本文模型	84.12	95.06

### 2.3.4 可视化分析

本文模型在 Market-1501 上行人图像检索结果如图 2 所示, 每张 query 有 10 张按准确率递减排序的查询结果, 绿色数字为正确结果, 红色数字为错误查询结果。可以看出与 query 具有状态一致性的图像匹配更准确; 与 query 角度不同的图像通过局部特征提取也能查询出来; 对背景复杂, 图像噪声大的图像匹配效果基本准确, 验证本文所提模型具有较好的效果。



图 2 图像检索结果

Fig. 2 Image retrieval results

## 3 结束语

本文采用 ResNet50 作为基础模型, 通过 3 个分支, 即融合局部特征提取分支、全局特征提取分支和批特征擦除分支进行局部特征融合。提取具有鉴别性的行人特征, 并加入交叉熵损失函数和三元组损失函数联合训练模型。通过在 Market-1501 实验, 表明局部提取具有有效性, 得到细粒度特征, 有利于提升网络对行人判别能力, 提高精确度。

## 参考文献

- [1] 林兰. 行人重识别若干问题的研究[D]. 成都: 电子科技大学, 2019.
- [2] CHEISSARI N, SEBASTIAN T B, HARTLEY R, et al. Person

reidentification using spatiotemporal appearance[C]//Proc of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006: 1528-1535.

- [3] GRAY D, TAO H. Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features[C]// Computer Vision-ECCV, Berlin, 2008: 262-275.
- [4] LIAO S C, ZHU Y X Y, et al. Person re-identification by Local Maximal Occurrence representation and metric learning [C]// IEEE Conference on Computer Vision and Pattern Recognition, Boston, 2015: 2197-2206.
- [5] SERMANET P, EIGEN D, ZHANG X, et al. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks[J]. Eprint Arxiv, 2013.
- [6] YI D, LEI Z, LIAO S C, et al. Deep Metric Learning for Person Re-identification [C]//International Conference on Pattern Recognition, Stockholm, 2014: 34-39.

(下转第 132 页)