

# 基于特征融合网络的行人重识别<sup>①</sup>



种衍杰, 方 琰, 沙 涛

(南京理工大学 电子工程与光电技术学院, 南京 210094)

通讯作者: 沙 涛, E-mail: [shaye@mail.njust.edu.cn](mailto:shaye@mail.njust.edu.cn)

**摘 要:** 行人重识别旨在大规模的分布式监控系统中进行行人目标匹配, 紧凑且具有鲁棒性的特征表达对其至关重要, 为此, 本文提出了一种基于特征融合网络的特征提取方法. 首先, 利用 STEL 算法增强了 LOMO 特征对背景噪声的抗噪性能, 利用 KPCA 算法降低维度以便于后续融合. 随后, 本文探索了手工特征和 CNN 特征的互补性, 将改进 LOMO 特征融入至卷积神经网络之中, 得到了区分度更高的融合特征. 在 VIPeR 和 CUHK01 数据集上的测试结果表明, 本文融合特征的区分度明显高于单一特征和级联特征, Rank-1 较级联特征分别提高了 3.73% 和 2.36%.

**关键词:** 行人重识别; 卷积神经网络; LOMO 特征; 特征融合网络

引用格式: 种衍杰, 方琰, 沙涛. 基于特征融合网络的行人重识别. 计算机系统应用, 2019, 28(1): 127-133. <http://www.c-s-a.org.cn/1003-3254/6731.html>

## Person Re-Identification by Feature Fusion Network

CHONG Yan-Jie, FANG Yan, SHA Tao

(School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

**Abstract:** Person re-identification aims at pedestrian target matching under distributed monitoring systems. Compact and robust feature is critical to it. For this reason, this study proposes a feature extraction method based on feature fusion network. Firstly, the STEL algorithm is used to enhance the immunity of LOMO feature to background noise, and the KPCA algorithm is used to reduce dimension. Subsequently, we explore the complementarity between manual features and Convolutional Neural Network (CNN) features, and integrate the improved LOMO feature into the CNN to obtain a fusion feature with better performance. Experiments on two datasets (VIPeR and CUHK01) validate the effectiveness of our proposal, the Rank-1 of fusion feature is 3.73% and 2.36% higher than the cascade feature, respectively.

**Key words:** person re-identification; Convolutional Neural Network (CNN); LOMO; feature fusion network

自 2006 年, 行人重识别的概念第一次在 CVPR 会议上被提出后<sup>[1]</sup>, 相关研究越来越受到学者们的关注. 行人重识别旨在大规模的视频监控系统中进行行人目标匹配, 即确认不同摄像头在不同时间、不同地点所记录到的行人目标是否为同一个人, 进而达到跟踪或是搜索的目的<sup>[2]</sup>. 由于光照、视角、姿态等方面的差异, 导致同一行人在不同画面中可能呈现截然不同的外观; 加之存在图像分辨率低下、遮挡等问题, 使得行人重识别成为一个颇具挑战的课题.

行人重识别中两个关键问题: 特征表示和度量学习<sup>[3]</sup>. 一个有效的特征应当紧凑且对诸如视角、光照等变化具有鲁棒性; 而度量学习则通过学习获得一个新的距离度量空间, 使得相同行人目标之间的距离更近, 而不同目标之间的距离更远. 许多学者沿着这两个方向做了许多贡献. Gray 等<sup>[4]</sup>提出了局部特征集 (Ensemble of Localized Features, ELF), 融合了颜色、纹理多种特征, 依据分类效果对特征赋予相应的权重. Farenzena 等<sup>[5]</sup>提出了局部特征对称性驱动累积

① 收稿时间: 2018-07-12; 修改时间: 2018-08-09; 采用时间: 2018-08-16; csa 在线出版时间: 2018-12-26

(Symmetry-Driven Accumulation of Local Features, SDALF), 从 STEL(STel Component Analysis)<sup>[6]</sup>模型提取的行人轮廓中寻找垂直对称轴, 依据对称轴对像素加权后提取颜色和纹理特征. Liao 等<sup>[3]</sup>提出了局部最大事件 (Local Maximal Occurrence representation, LOMO), 联合了 HSV 颜色直方图和 SILTP (Scale Invariant Local Ternary Patter)<sup>[7]</sup>纹理特征描述子, 并使用最大池化的方法获得了更加稳定的特征表达, 同时, 还提出了跨视角二次判别 (cross-view Quadratic Discriminant Analysis, XQDA) 的度量学习方法. Wu 等<sup>[8]</sup>提出了特征融合网络 (Feature Fusion Net, FFN), 尝试将手工剪裁特征和卷积神经网络提取的 CNN (Convolutional Neural Network) 特征进行了有效的融合.

本文聚焦于行人重识别中的特征表达, 首先研究了原始 LOMO 特征的提取流程, 接着, 利用 STEL 算法改善了原始 LOMO 特征对背景噪声的抗噪性能, 利用 KPCA (Kernel Principal Component Analysis)<sup>[9]</sup>算法降低维度以便于后续融合, 在特征维度更低的情况下, 本文改进 LOMO 特征取得了较原始 LOMO 特征更加优异的识别准确率. 随后, 本文探索了手工特征和 CNN 特征的互补性, 经由全连接层将改进 LOMO 特征融入卷积神经网络之中, 反向传播使得改进 LOMO 特征能够优化网络参数的学习, 得到了区分度更高的融合特征. 在 VIPeR 和 CUHK01 数据集上的测试结果表明, 本文融合特征的区分度明显高于单一特征和级联特征, Rank-1 较级联特征分别提高了 3.73% 和 2.36%.

## 1 改进 LOMO 特征

### 1.1 原始 LOMO 特征提取方法

由于行人图像分辨率较低, 诸如人脸等其它生物特征难以被捕捉到, 所以, 行人重识别主要依据行人的表观特征, 例如颜色、纹理等特征. Liao 提出的 LOMO 特征, 先将行人图像水平划分, 然后提取 HSV 颜色直方图和 SILTP<sup>[7]</sup>纹理特征, 并使用最大池化的方法获得了更加稳定的表达. 图 1 展示了 LOMO 特征的提取流程.

#### 1.1.1 处理光照变化

颜色是描述行人图像的一个重要特征, 然而, 由于光照条件、相机设置以及阴影等因素, 同一个行人在不同相机下所呈现的颜色可能相距甚远<sup>[10]</sup>. 为此, LOMO 特征首先使用了多尺度 Retinex 算法<sup>[11]</sup>预处理行人图片. Retinex 算法能够很好地近似人类视觉感知, 并加

强了阴影区域的细节表现. 在 LOMO 特征使用的多尺度 Retinex 算法中, 小尺度 Retinex 用于动态范围压缩, 大尺度 Retinex 用于色调再现. 图 2 展示了行人图片经 Retinex 处理前后的对比, 其中 (a) 为原始图片, (b) 为经 Retinex 处理后的图片, (a)(b) 中同一列为同一行人图像. 可以看出, 处理后的行人图片在光照和颜色上有着更好的一致性.

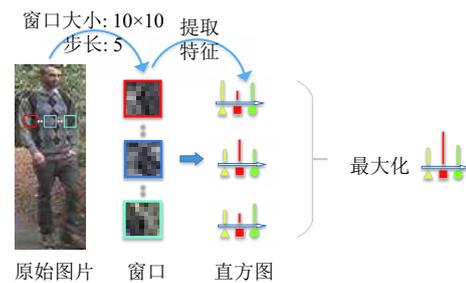


图 1 LOMO 特征的提取流程



图 2 行人图片经 Retinex 处理前后对比<sup>[3]</sup>

#### 1.1.2 处理视角变化

考虑到行人在不同相机下往往呈现出不同的视角, LOMO 特征使用了滑动窗口来描述图像的局部细节. 具体而言, LOMO 特征使用了尺寸为  $10 \times 10$  的窗口, 在经裁剪的行人图像上以 5 像素步长滑动. 在每个窗口内提取  $8 \times 8 \times 8$ -bin HSV 颜色直方图和  $SILTP_{4,3}^{0,3}$ 、 $SILTP_{4,3}^{0,3}$  两尺度纹理特征. 为了使产生的直方图对视角变化具有鲁棒性, LOMO 特征扫描同一水平位置的所有窗口, 提取每种种子特征的最大值组成一个新的特征表达.

为了进一步考虑多尺度信息, LOMO 特征构建了三尺度图像金字塔, 对原始图像进行下采样后, 重复上述特征提取流程.

最后级联所有特征, 得到了长度为  $(8 \times 8 \times 8 \text{HSV} + 3^4 \times 2 \text{SILTP}) \times (24 + 11 + 5) = 26\ 960$  维的 LOMO 特征.

### 1.2 改善抗噪性能

LOMO 特征使用滑动窗口来描述行人图像的局部

细节,并通过最大化水平窗口特征值的方式来处理视角变化.该方法简单有效,但却没有很好地处理背景噪声的影响.在我们的实验中,LOMO特征在不同的数据集下表现得不够稳定也验证了这一点.受到SDALF算法<sup>[5]</sup>利用对称信息处理视角变化的启发,本文使用行人垂直对称信息和加权函数来消除背景噪声的干扰.

首先,利用STEL<sup>[6]</sup>将行人从背景中提取出来,然后由行人轮廓确定垂直对称轴,依据对称轴对滑动窗

口赋予不同的权值:距离对称轴较近的窗口更有可能是前景(行人),赋予较大权重;而对远离对称轴的窗口赋予较小权值.这样,所提取的LOMO特征更加专注于前景而减小了背景噪声带来的影响.权值函数选用高斯函数:

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (1)$$

其中, $\mu$ 取行人垂直对称轴的坐标值, $\sigma$ 取图像宽度值的0.3.图3展示了窗口加权LOMO特征的提取过程.

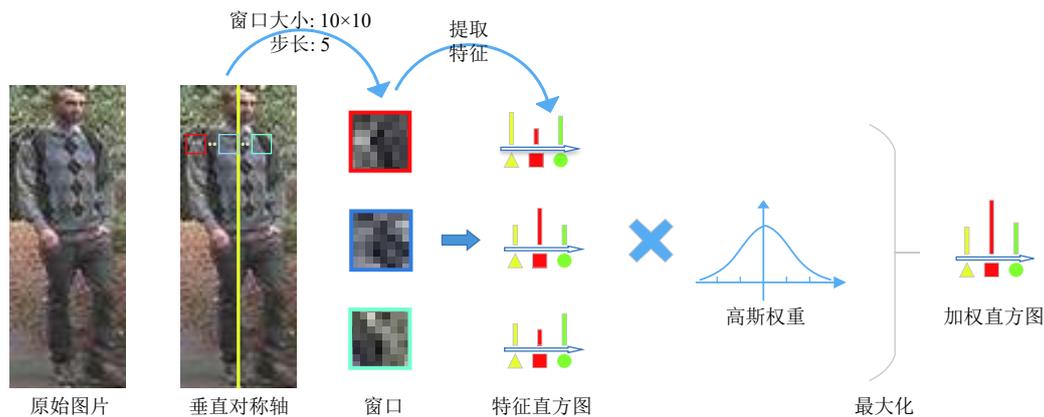


图3 窗口加权LOMO特征提取流程

具体而言,在行人对称轴确定以后,使用了尺寸为 $10 \times 10$ 的滑动窗口,在行人图像上以5像素步长滑动.在每个窗口内提取 $8 \times 8 \times 8$ -bin HSV颜色直方图和 $SILTP_{4,3}^{0.3}$ 、 $SILTP_{4,3}^{0.3}$ 两尺度纹理特征.扫描同一水平位置所有窗口,使用高斯函数对窗口特征值进行加权,最后选取最大特征值作为该水平位置的LOMO特征描述符.

### 1.3 降低特征维度

LOMO特征在每个水平位置提取 $8 \times 8 \times 8$ 的HSV颜色特征以及 $3^4 \times 2$ 的SILTP纹理特征,对于原始尺寸为 $128 \times 48$ 的图像,在滑动窗口大小为 $10 \times 10$ ,步长为5的情况下,共有24个水平条,两次 $2 \times 2$ 平均池化后分别有11和5个水平条,故LOMO特征维度为 $(8 \times 8 \times 8 \text{HSV} + 3^4 \times 2 \text{SILTP}) \times (24 + 11 + 5) = 26\,960$ 维.这样的数据维度甚至数倍于原始图像,不仅增加了后续特征融合、度量学习的复杂度,也使得存储空间大大增加.因此,本文在保证一定特征精度的情况下,对其进行降维(特征抽取)操作.

首先,使用上述流程提取窗口加权LOMO特征,将LOMO特征矩阵变换为 $674 \times X$ 维的矩阵, $X$ 为图像分割的水平条数.在降维算法上选取了KPCA算法,该算法应

用了Mercer核来扩展PCA(Principal Component Analysis),由输入空间经非线性映射得到高维空间,然后在高维空间计算主成分<sup>[12]</sup>.图4展示了应用KPCA算法对LOMO特征进行降维的性能曲线,测试图库为VIPeR cam\_a图集,共632张图片.当累积方差贡献率为0.9时,平均所需主成分数目仅为17.20,即平均使用11 591维的数据即可表示原始26 960维LOMO特征近90%的信息量.

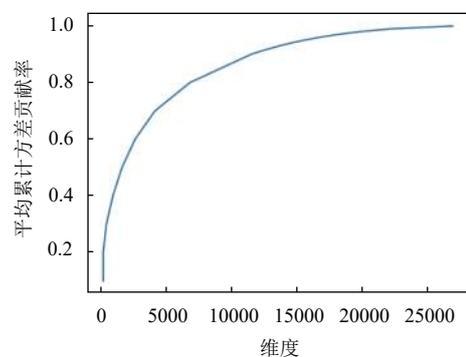


图4 KPCA对LOMO特征降维的性能曲线

结合图4,我们将LOMO特征矩阵降维至原始维度的50%,此时平均累积方差贡献率为0.91.

## 2 特征融合网络

### 2.1 网络架构

近来,专为行人重识别设计的、传统的手工多特征融合<sup>[3,5,13]</sup>在性能上遇到了瓶颈.随着神经网络在图像识别领域的流行,越来越多的研究聚焦于基于深度学习的行人重识别.然而,深度学习的主要瓶颈是缺乏训练数据,大多数的重识别数据集对每个行人只提供两张图片,例如 VIPeR.所以,仅采用深度学习的行人重识别算法在某些数据集上的效果并不显著.

受到 FFN<sup>[8]</sup>的启发,本文提出了一种基于特征融合网络的图像特征提取方式,该网络融合了 CNN 特征和

改进 LOMO 特征,将其映射至一统一的特征空间.在反向传播的过程中,手工剪裁的改进 LOMO 特征能够优化神经网络的参数学习.这样,我们得到了比单个特征或者级联特征更有区分度的融合特征.如图 5 所示,特征融合网络由两部分组成,第一部分使用普通的卷积神经网络从输入图像中提取特征,是对图像颜色、纹理、形态等特征的综合学习;第二部分使用改进 LOMO 来提取手工特征.借助全连接层使得第二部分能够优化第一部分的学习过程,两个特征最终融合在一起形成一个更加充分的特征表达.图 6 展示了在训练好的融合网络中进行行人目标匹配的流程,系统输出了匹配度较高的前 20 个结果.

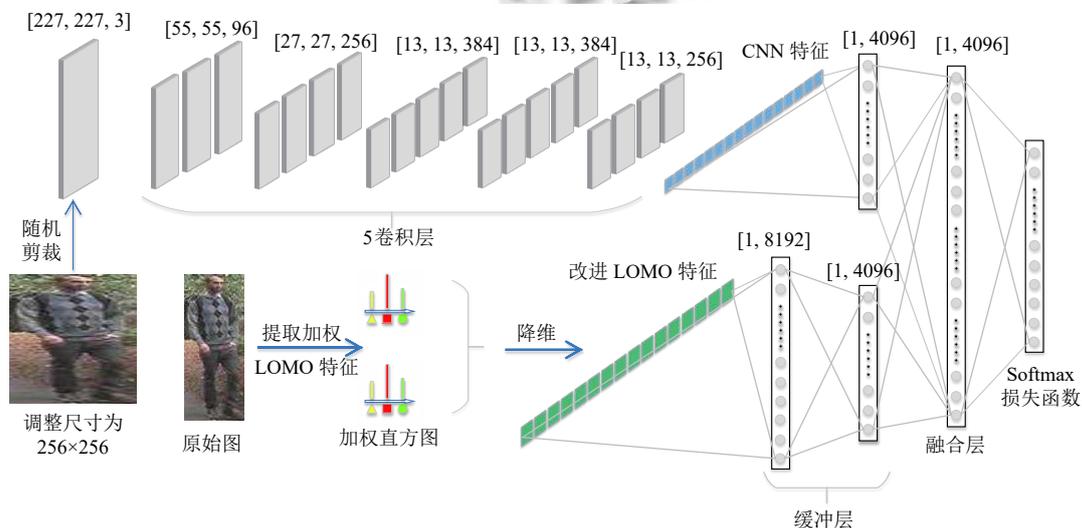


图 5 特征融合网络架构图

### 2.2 CNN 特征的提取

特征融合网络首先使用 5 个卷积层来提取图像的 CNN 特征,每个卷积层后使用 ReLU 函数作为激活函数,池化层使用最大池化以更好地保留纹理特征,忽略背景噪声,最终得到了 4096 维的 CNN 特征.卷积神经网络具有权值共享的特性,有着更少的连接和参数,因而更容易被训练<sup>[14]</sup>.

鉴于目前流行的行人重识别数据集行人数量在 1000 个左右,图片总数大多小于 10 000 张,较小的训练集容易使得神经网络出现过拟合现象.为此,本文首先采取图像变化和水平反转来增强数据<sup>[14]</sup>.具体地,首先将图像尺寸调整至  $256 \times 256$ ,而后从  $256 \times 256$  图像上随机剪裁大小为  $227 \times 227$  的图像块,将它们及其水平

反转送入神经网络进行学习,这样训练样本总数增大了 1682 倍.另外,为了加快网络的收敛速度,在训练集的图像上进行了减均值操作.

### 2.3 缓冲层和融合层

由于 LOMO 特征和 CNN 特征差异巨大,为了使融合得以成功,在 CNN 特征和 LOMO 特征融合之前分别由一到两个缓冲层 (Buffer Layer) 进行缓冲. CNN 特征的缓冲层由一个全连接层组成, LOMO 特征的缓冲层由两个输出维度分别为 8192 和 4096 的全连接层组成.为了使融合更加充分, LOMO 特征事先进行了降维操作,然后再通过两级维度递减的全连接层进行缓冲.

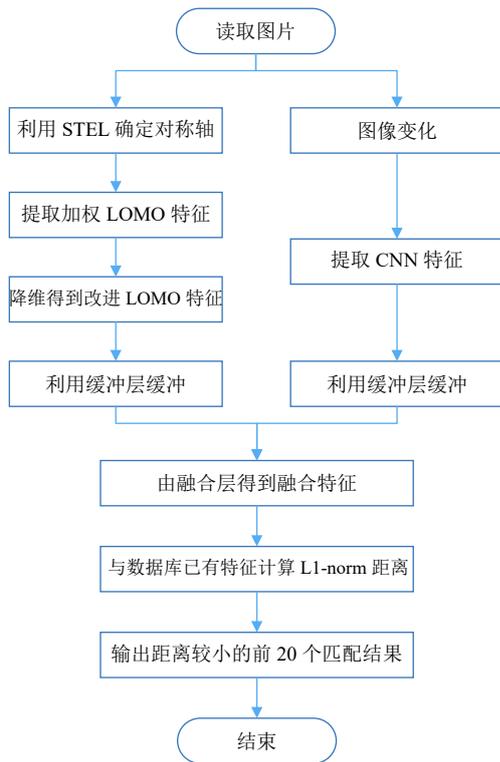


图 6 融合特征网络行人匹配流程图

缓冲层之后的融合层 (fusion layer) 输出了维度为 4096 维的融合特征, 该层利用了全连接层来增强对 LOMO 特征和 CNN 特征的自适应能力. 另外, 缓冲层和融合层使用了概率为 0.5 的失活函数, 降低了特定神经元之间的互相依赖, 从而迫使其学习更加鲁棒的特征<sup>[14]</sup>.

假定缓冲层的输出分别为  $B_{LOMO}$  和  $B_{CNN}$ , 则最终融合层输出的融合特征可以表示为:

$$Z_{Fusion} = h(W_{Fusion}^T [B_{LOMO}, B_{CNN}] + b_{Fusion}) \quad (2)$$

其中,  $W_{Fusion}^T$  为连接权重,  $b_{Fusion}$  为阈值,  $h(\cdot)$  表示 ReLU 激活函数. 则反向传播时, 第一次第  $l$  层的参数更新为:

$$W_1^{(l)} = W^{(l)} - \alpha \left[ \left( \frac{1}{m} \Delta W^{(l)} \right) + \lambda W^{(l)} \right] \quad (3)$$

$$b_1^{(l)} = b^{(l)} - \alpha \left( \frac{1}{m} \Delta b^{(l)} \right) \quad (4)$$

其中,  $\alpha$  为学习率,  $\lambda$  为动量. 损失函数层使用 Softmax Loss 函数, 每个类别的预测概率为:

$$p(y = j|x; \theta) = \frac{e^{\theta_j^T x}}{\sum_{k=1}^n e^{\theta_k^T x}} \quad (5)$$

每个样本在所有输出节点的损失函数之和为:

$$J_{softmax}(\theta) = - \left[ \sum_{k=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_{\theta}(x^{(i)})) \right] \quad (6)$$

损失函数的导数为:

$$\nabla_{\theta^{(k)}} J(\theta) = - \sum_{i=1}^m [x^{(i)} (\{y^{(i)} = k\} - p(y^{(i)} = k|x^{(i)}; \theta))] \quad (7)$$

利用链式求导法则逐层更新整个网络参数.

### 3 实验与分析

#### 3.1 实验规则

分别在 VIPeR<sup>[15]</sup> 和 CUHK01<sup>[16]</sup> 数据集上进行测试, 在每个数据集测试中随机选取一半样本作为训练数据集, 另一半作为测试数据集, 重复 10 次, 计算平均 Rank- $i$  并绘制 CMS (Cumulative Matching Characteristic) 曲线. 在分类器上选择了最原始的非监督分类算法 L1-norm, L1-norm 距离又称曼哈顿距离, 能够表示特征最原始的区分度. 如图 6 所示, 从融合层获得融合特征后, 与已有特征计算 L1-norm 距离, 获得了匹配度较高的前 20 个结果.

在特征融合网络的训练中, 使用分批随机梯度下降法以 20 张图片为单位, 前向传导至最后的 Softmax 损失层, 计算样本损失的平均值, 并以此作为反向传播的损失值. 初始学习率  $\gamma_{initial} = 0.001$ , 更新策略:

$$\gamma_{new} = \gamma_{initial} \times (1 + \text{gamma} \times \text{iter})^{-power} \quad (8)$$

其中,  $\text{gamma}$  取 0.0001,  $\text{power}$  取 0.75,  $\text{iter}$  为迭代次数.

在每次实验中, 分别对比原始 LOMO 特征、本文改进 LOMO 特征、本文提取的 CNN 特征、改进 LOMO 特征和 CNN 特征的级联特征以及融合特征, gBiCov<sup>[17]</sup> 以及 ELF16<sup>[8]</sup> 作为额外对比项.

#### 3.2 实验结果

##### 3.2.1 在 VIPeR 上实验

VIPeR 数据集是行人重识别领域最著名的数据集, 包含 632 个行人共 1264 张图像. VIPeR 数据集的特点是光照和视角的多样性. 图 7 以及表 1 展示了 VIPeR 数据集上的测试结果.

##### 3.2.2 在 CUHK01 上实验

CUHK01 数据集是由香港中文大学于 2012 年发

布的行人重识别基准测试数据集, 包含了 2 个摄像头、每个摄像头 971 个 ID、每个 ID 2 张图片, 共 1264 张图片. CUHK01 采集自行人天桥, 图片整体亮度较低. 图 8 以及表 2 展示了 CUHK01 数据集上的测试结果.

表 1 各种特征在 VIPeR 数据集、L1-norm 下的性能比较

Rank-i	i=1	i=5	i=10	i=20
本文融合特征	19.76	34.57	43.59	53.87
本文改进 LOMO+CNN 特征	16.03	30.01	39.58	48.27
ELF16 特征	12.37	26.72	34.14	44.56
本文改进 LOMO 特征	11.70	23.41	32.43	42.87
原始 LOMO 特征	9.80	20.43	28.45	39.61
gBiCov 特征	9.43	19.37	27.17	35.45
本文 CNN 特征	5.61	13.05	18.01	26.05

表 2 各种特征在 CUHK01 数据集、L1-norm 下的性能比较

Rank-i	i=1	i=5	i=10	i=20
本文融合特征	19.47	34.12	42.78	52.34
本文改进 LOMO+CNN 特征	17.11	29.58	35.36	43.91
本文改进 LOMO 特征	12.07	24.25	32.06	41.03
原始 LOMO 特征	10.24	22.78	28.85	36.12
gBiCov 特征	7.25	14.93	19.01	24.35
本文 CNN 特征	6.47	15.01	19.07	25.86
ELF16 特征	5.37	12.67	17.47	23.57

### 3.2.3 运行时间

表 3 评估了本文所对比的行人重识别特征的提取时间, 测试图集为 VIPeR cam\_a, 取其提取 632 张图像特征的平均时间. 测试机器 CPU 为 Intel i5-5200U, 内存 8 GB, 显卡为 NVIDIA GTX950M.

表 3 各种特征的提取时间、输出维度比较

特征	提取时间 (s)	输出维度
本文融合特征	0.31+0.18=0.49	4096
本文 CNN 特征	0.18	4096
gBiCov 特征	13.60	5940
ELF16 特征	0.57	8064
本文改进 LOMO 特征	0.31	13 480
原始 LOMO 特征	0.22	26 960

可以看到, 融合网络的特征提取时间甚至快于某些手动特征. 虽然融合网络的训练需要花费大量时间, 但在训练完的网络中运行一次前向传播只需 0.49 s, 而且, 其中大部分时间 (0.31 s) 花费在改进 LOMO 特征的提取上. 相较于其它特征, 融合特征的维度更小, 在后续的度量学习中也有着更大的优势.

### 3.3 结果分析

图 7 和图 8 展示了在 L1-norm 下, 本文融合特征与对比特征分别在 VIPeR 和 CUHK01 数据集上的表

现. L1-norm 衡量了各种特征最直观的区分度, 结果表明, 融合特征的区分度大于其它特征.

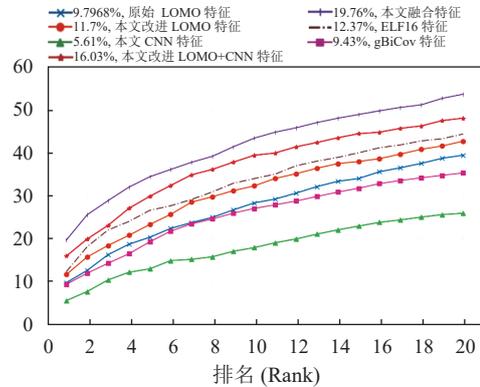


图 7 VIPeR 数据集、L1-norm 下各种特征的 CMS 曲线

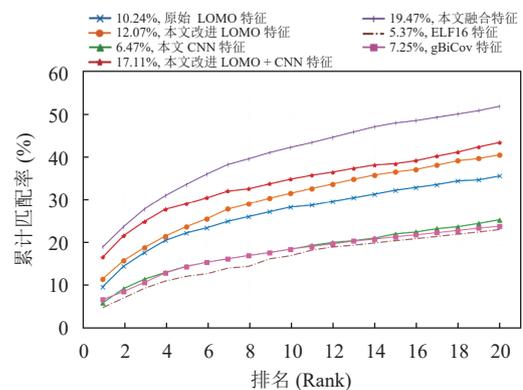


图 8 CUHK01 数据集、L1-norm 下各种的特征 CMS 曲线

本文改进 LOMO 特征和 CNN 特征的级联特征准确率大于改进 LOMO 特征、CNN 特征两个单独特征, 说明了手工特征和 CNN 特征在一定程度上是可以互补的. 然而简单的级联并不是最优的融合方式, 在 VIPeR 和 CUHK01 数据集上, 融合特征 Rank-1 为 19.76% 和 19.47%, 较级联特征的 16.03% 和 17.11%, 分别提高 3.73% 和 2.36%. 融合特征的准确率大于级联特征的主要原因在于:

- (1) 经由缓冲层和融合层使得改进 LOMO 特征参与了神经网络的训练过程, 在反向传播时, 提取 CNN 特征的卷积核 (Filters) 学习到了改进 LOMO 特征的特性.
- (2) 缓冲层和融合层由全连接层组成, 能够更好的融合两者的互补特性.

另外, 本文改进 LOMO 特征由于更加专注于行人, 减小了背景噪声的影响, 在特征维度更小的情况下, 表现出比原始 LOMO 特征更加优异的性能, 在 VIPeR

和 CUHK01 数据集上分别较原始 LOMO 特征提高了 1.9% 和 1.83%。

#### 4 总结

本文首先利用 STEL 算法增强了 LOMO 特征对背景噪声的抗噪性能,为了便于后续融合,利用 KPCA 进行了降维,在特征维度更低的情况下取得了较原始 LOMO 特征更加优异的性能。随后,本文探索了手工特征和 CNN 特征的互补性,将改进 LOMO 特征融入到了卷积神经网络之中,得到了区分度更高的融合特征。在两个颇具挑战的数据集上,本文融合特征表现出比单独或者级联特征更加优异的性能,Rank-1 较级联特征分别提高了 3.73% 和 2.36%。

#### 参考文献

- 1 Gheissari N, Sebastian TB, Hartley R. Person reidentification using spatiotemporal appearance. Proceedings of 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, NY, USA. 2006. 1528–1535.
- 2 Zheng L, Yang Y, Hauptmann AG. Person re-identification: Past, present and future. arXiv preprint arXiv: 1610.02984, 2016.
- 3 Liao SC, Hu Y, Zhu XY, *et al.* Person re-identification by local maximal occurrence representation and metric learning. Proceedings of 2005 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 2197–2206.
- 4 Gray D, Brennan S, Tao H. Evaluating appearance models for recognition, reacquisition, and tracking. Proceedings of the 10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS). Rio de Janeiro. 2007, 3: 1–7.
- 5 Farenzena M, Bazzani L, Perina A, *et al.* Person re-identification by symmetry-driven accumulation of local features. Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA. 2010. 2360–2367.
- 6 Andriluka M, Roth S, Schiele B. Pictorial structures revisited: People detection and articulated pose Estimation. Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA. 2009. 1014–1021.
- 7 Liao SC, Zhao GY, Kellokumpu V, *et al.* Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes. Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA. 2010. 1301–1306.
- 8 Wu SX, Chen YC, Li X, *et al.* An enhanced deep feature representation for person re-identification. Proceedings of 2016 IEEE Winter Conference on Applications of Computer Vision. Lake Placid, NY, USA. 2016. 1–8.
- 9 Schölkopf B, Smola A, Müller KR. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation, 1998, 10(5): 1299–1319. [doi: [10.1162/089976698300017467](https://doi.org/10.1162/089976698300017467)]
- 10 章登义, 王骞, 朱波, 等. 利用人体部位特征重要性进行行人再识别. 武汉大学学报·信息科学版, 2017, 42(1): 84–90.
- 11 Jobson DJ, Rahman Z, Woodell GA. Properties and performance of a center/surround retinex. IEEE Transactions on Image Processing, 1997, 6(3): 451–462. [doi: [10.1109/83.557356](https://doi.org/10.1109/83.557356)]
- 12 张田昊. 数据降维算法研究及其应用[博士学位论文]. 上海: 上海交通大学, 2008.
- 13 Yang Y, Yang JM, Yan JJ, *et al.* Salient color names for person re-identification. European Conference on Computer Vision. Zurich, Switzerland. 2014. 536–551.
- 14 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 1097–1105.
- 15 Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features. European Conference on Computer Vision. Marseille, France. 2008. 262–275.
- 16 Li W, Zhao R, Wang XG. Human reidentification with transferred metric learning. Asian Conference on Computer Vision. Daejeon, Korea. 2012. 31–44.
- 17 Ma BP, Su Y, Jurie F. Covariance descriptor based on bio-inspired features for person re-identification and face verification. Image and Vision Computing, 2014, 32(6–7): 379–390.