

基于多池化融合与背景消除网络的跨数据集行人再识别方法

李艳凤, 张斌, 孙嘉, 陈后金, 朱锦雷

(北京交通大学电子信息工程学院, 北京 100093)

摘 要: 现有跨数据集行人再识别方法一般致力于减小 2 个数据集之间的数据分布差异, 忽略了背景信息对识别性能的影响。针对上述问题, 提出了一种基于多池化融合与背景消除网络的跨数据集行人再识别方法。为了兼顾全局特征和局部特征, 同时实现特征的多细粒度表示, 构建了多池化融合网络。为了使监督网络能提取有用的行人前景特征, 构建了特征级有监督背景消除网络。采用结合行人分类损失及特征激活损失的多任务学习损失函数, 在 3 个公开行人再识别数据集上对方法进行评估, 当 MSMT17 作为训练集时, Market-1501 上的跨数据集识别性能 mAP 为 35.53%, 相比 ResNet50 网络提升了 9.24%; DukeMTMC-reID 上的跨数据集识别性能 mAP 为 41.45%, 相比于 ResNet50 网络提升了 10.72%。与现有方法相比, 所提方法具有更优的跨数据集行人再识别性能。

关键词: 行人再识别; 跨数据集; 背景消除; 多池化融合; 深度学习

中图分类号: TN391

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2020181

Cross-dataset person re-identification method based on multi-pool fusion and background elimination network

LI Yanfeng, ZHANG Bin, SUN Jia, CHEN Houjin, ZHU Jinlei

School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100093, China

Abstract: The existing cross-dataset person re-identification methods were generally aimed at reducing the difference of data distribution between two datasets, which ignored the influence of background information on recognition performance. In order to solve this problem, a cross-dataset person re-ID method based on multi-pool fusion and background elimination network was proposed. To describe both global and local features and implement multiple fine-grained representations, a multi-pool fusion network was constructed. To supervise the network to extract useful foreground features, a feature-level supervised background elimination network was constructed. The final network loss function was defined as a multi-task loss, which combined both person classification loss and feature activation loss. Three person re-ID benchmarks were employed to evaluate the proposed method. Using MSMT17 as the training set, the cross-dataset mAP for Market-1501 was 35.53%, which was 9.24% higher than ResNet50. Using MSMT17 as the training set, the cross-dataset mAP for DukeMTMC-reID was 41.45%, which was 10.72% higher than ResNet50. Compared with existing methods, the proposed method shows better cross-dataset person re-ID performance.

Key words: person re-identification, cross-dataset, background elimination, multi-pool fusion, deep learning

1 引言

行人再识别^[1-2]用于解决非重叠视域中人员匹

配的问题, 是智能视频分析技术的重要组成部分, 可用于追踪犯罪嫌疑人、寻找走失人员等, 具有广阔的应用前景^[3]。近年来, 行人再识别技术获得了

收稿日期: 2020-04-23; 修回日期: 2020-08-05

通信作者: 陈后金, hjchen@bjtu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61872030); 山东省重大科技创新工程基金资助项目 (No.2019TSLH0206)

Foundation Items: The National Natural Science Foundation of China (No.61872030), The Major Science and Technology Innovation Project of Shandong Province (No.2019TSLH0206)

广泛关注，成为计算机视觉领域的研究热点。

早期的行人再识别研究主要依靠传统方法，例如手工设计特征表达^[4-5]和距离度量^[6-7]。得益于深度学习的快速发展，越来越多的研究者使用深度学习来解决行人再识别问题。Yi 等^[8]首次将深度神经网络应用于行人再识别，采用端到端的网络联合进行特征提取与度量学习。近年来，由局部生成行人的细粒度特征方法被广泛研究，并取得了较好的性能。Yao 等^[9]提出部件损失网络（PL-Net, part loss network）以同时最小化分类风险与表示学习风险。Sun 等^[10]提出对特征图进行均匀分割的基于部件的卷积基线（PCB, part-based convolutional baseline）网络框架用于行人再识别，该方法将特征图均分为多个局部特征图分支，每个分支分别与行人 ID 标签计算损失。实现了图像局部特征的提取，但忽略了全局特征表示。Fu 等^[11]提出了水平金字塔匹配（HPM, horizontal pyramid matching）行人再识别方法。该方法同时实现了行人的全局特征和局部特征表示，在局部特征表示中进行多尺度池化，并在池化方式上，将全局特征平均池化（GAP, global average pooling）和全局特征最大池化（GMP, global max pooling）相结合，但该方法中大量的池化支路导致特征向量维度较大，需要将特征压缩后再组合，从而造成信息丢失。

在实际应用场景中，监控视频的摄像头数量是不固定的，并且会增加，因此行人再识别是一个开放性的问题，其模型应具有适应开放环境的能力。为了更加符合实际场景的应用需求，研究人员开始关注跨数据集行人再识别方法。跨数据集行人再识别的任务是将一个数据集（源数据集）上训练好的模型转移到另一个数据集（目标数据集）上进行测试，希望获得较好的识别性能。目前跨数据集行人再识别方法主要有两类，具体如下。

1) 利用迁移学习知识减小源数据集与目标数据集数据分布的差异。Qi 等^[12]关注摄像机之间不同的特征分布，利用迁移学习的思想设计了基于“相机感知”的域自适应方法，以减少 2 个数据集之间的数据差异。Li 等^[13]利用迁移学习中的最大平均差异度量损失使源数据集与目标数据集生成相似的特征分布来缩小 2 个数据集间的差异。Huang 等^[14]使用部件分割约束增强模型的泛化能力，实现域自适应行人再识别。Wang 等^[15]提出迁移联合属性-类别深度学习（TJ-AIDL, transferable joint attribute-identity deep learning）的行人再识别方法。该方法

使用双支路网络，上支路以行人类别信息作为标签进行特征提取，下支路以属性信息作为标签进行特征提取，然后将训练好的模型迁移到目标数据集进行行人识别方法。Lin 等^[16]提出多任务中级特征对齐（MMFA, multi-task mid-level feature alignment）网络实现跨数据集行人再识别。该方法基于源域与目标域共享中层特征空间的假设，使用源域-目标域中层特征对齐正则化项对网络进行优化。Zhong 等^[17]提出异质-同质学习（HHL, hetero-homogeneous learning）的行人再识别方法，通过同质学习实现相机不变性，通过异质学习实现域连通性。

2) 利用生成对抗网络（GAN, generative adversarial network）将源数据集图像转换为目标数据集图像的风格进行训练。Wei 等^[18]提出行人迁移生成对抗网络（PTGAN, person transfer GAN），在保证行人本体前景不变的情况下，将源域图像的背景转换成目标域图像的背景风格进行训练。Deng 等^[19]设计了相似度保持生成对抗网络（SPGAN, similarity preserving GAN），在图像生成过程中添加了自相似性及域不相似性限制。Liu 等^[20]提出自适应迁移网络（ATNet, adaptive transfer network）来实现跨数据集行人再识别。该方法将复杂的跨数据集迁移分为 3 个子问题，然后对每个子问题分别进行风格迁移。

现有跨数据集方法一般致力于减小 2 个数据集之间的数据分布差异，忽略了背景信息对识别性能的影响。Tian 等^[21]研究了图像背景对行人识别性能的影响，在使用仅包含背景的行人图像进行测试时，也获得了 Rank-1 为 5.2% 的正确率。上述研究表明将整张行人图像输入行人再识别网络中进行训练，网络在学习行人特征的同时，也会学习相应的背景特征。为降低背景信息的影响，本文提出了一种基于多池化融合与背景消除网络的跨数据集行人再识别方法，该方法通过结合多池化融合（MPF, multi-pooling fusion）网络与特征级有监督背景消除网络，有效提升跨数据集行人再识别的性能。本文的主要贡献如下：1) 构建了一种多池化融合 MPF 网络结构，能够兼顾全局特征和局部特征，同时实现特征的多细粒度表示；2) 设计了一种特征级有监督背景消除网络，将该网络的特征激活损失函数与行人分类损失函数相结合为多任务学习损失，以监督网络提取有用的行人前景特征。

2 本文方法

2.1 网络结构

本文方法的网络结构如图 1 所示，包括多池化融合网络和特征级有监督背景消除网络。以 ResNet50 网络^[22]作为主干网络，以主干网络得到的特征图为输入，构建多池化融合网络，其包含 4 个分支，其中，一个为全局特征分支，其他 3 个为不同区域的局部分支。全局特征分支采用 2 种不同细粒度的池化，且每种细粒度分别采用 GAP 和 GMP 这 2 种池化方式；其他 3 个局部分支采用一种细粒度的 GAP 和 GMP 池化。输出特征进行连接，得到 6 个分支的特征。每个分支分别连接全连接（FC, fully connection）层用于行人分类，得到的预测结果与行人类别标签进行计算交叉熵（CE, cross entropy）损失。特征级有监督背景消除网络中，首先通过行人分割获得行人掩码，然后对主干网络的特征图进行 Sigmoid 激活，并计算激活结果与行人掩码的均方误差（MSE, mean square error）损失。将 CE 损失与 MSE 损失之和作为本文方案的最终损失，以实现网络模型提取有用的行人前景特征。

2.2 多池化融合网络

ResNet50 网络常被用于行人再识别的主干网络，但其单一的池化结构易造成大量信息丢失。为解决这一问题，提高特征的代表能力，本文构建了

多池化融合网络，其结构如图 1 下半部分所示。

全局特征对行人进行整体性的描述，而局部特征则针对不同局部区域对行人进行描述，因此全局特征和局部特征结合可以提高特征的代表能力。行人图像通过 ResNet50 网络生成的特征图被分为 4 个支路，一个支路为全局特征，3 个支路为不同区域的局部特征。为防止硬划分破坏局部特征之间的连续性，本文采用有重叠划分的方式来得到局部特征图，并将特征图有重叠地划分为上、中、下三部分。全局特征图和 3 个局部特征图如图 2 所示。



(a)全局特征图 (b)局部特征图 1 (c)局部特征图 2 (d)局部特征图 3
图 2 全局特征图和局部特征图

不同的池化方式具有不同的特点，GMP 输出行人图像特征的最大响应点，GAP 有利于提取全局特征信息。本文融合了 GAP 与 GMP，有利于提升特征的代表能力，然后对 4 个支路的特征图分别进行 GAP 和 GMP 处理。为了得到不同细粒度的特征表

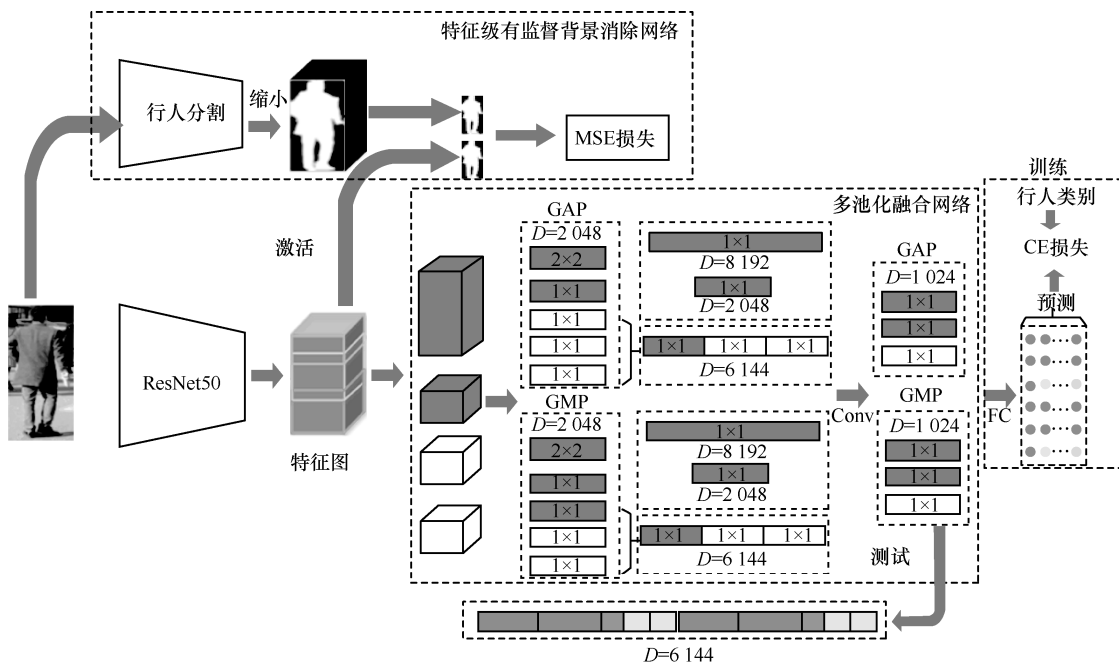


图 1 网络结构

示, 本文对全局特征图进行 2 种不同尺寸的池化, 宽、高、通道的尺寸分别为 $1 \times 1 \times 2048$ 和 $2 \times 2 \times 2048$ 。因此全局特征将生成 2 种池化方式、2 种池化尺寸, 共 4 个特征向量。对于局部特征, 每个局部特征图的池化尺寸为 $1 \times 1 \times 2048$, 分别采用 GAP 和 GMP 这 2 种池化方式, 因此 3 个局部支路共生成 6 个尺寸为 $1 \times 1 \times 2048$ 的特征向量。

将全局特征图池化得到的 2 个尺寸为 $2 \times 2 \times 2048$ 的特征向量分别展开, 得到 2 个尺寸为 $1 \times 1 \times 8192$ 的特征向量。对于 3 个局部特征图, 将每种池化方式下的 $1 \times 1 \times 2048$ 特征向量级联, 形成 2 个新的尺寸为 $1 \times 1 \times 6144$ 的特征向量。经过上述操作后, 共得到 6 个特征向量 (GAP 方式下的 2 个全局特征向量和 1 个局部特征向量, GMP 方式下的 2 个全局特征向量和 1 个局部特征向量)。为降低特征向量的维度, 分别对 6 个特征向量进行 1×1 卷积, 得到 6 个 1024 维的特征向量。训练时, 在 6 个特征向量后分别连接全连接层用于分类, 得到的分类结果分别与行人类别标签计算交叉熵损失, 多分类交叉熵损失函数如式(1)所示。

$$L_{CE} = -\sum_{i=1}^N \log \frac{\exp(\mathbf{W}_{y_i}^T \mathbf{x}_i + b_{y_i})}{\sum_{k=1}^C \exp(\mathbf{W}_k^T \mathbf{x}_i + b_k)} \quad (1)$$

其中, i 表示行人图像索引, \mathbf{x}_i 表示图像 I_i 的特征向量, y_i 表示图像 I_i 的行人 ID 号, C 表示训练集中行人 ID 的总数量, N 表示每个批次读取行人图像的数量, $\{\mathbf{W}, \mathbf{b}\}$ 表示分类层的权重和偏置。

多池化融合网络的损失函数 L_{ID} 为 6 个支路损失函数之和, 如式(2)所示, L_{CE}^j 为第 j 个支路的损失。

$$L_{ID} = \frac{1}{6} \sum_{j=1}^6 L_{CE}^j \quad (2)$$

本文设计的多池化融合网络对全局特征进行不同类型和不同尺寸的池化操作, 使多个特征向量与特征图建立了不同的映射关系, 实现了特征的多细粒度表示。对局部特征进行不同类型的池化操作, 并以全局特征和局部特征对行人进行综合表示, 可有效提升特征的表征能力。

2.3 特征级有监督背景消除网络

为降低背景信息的影响, 本文构建了特征级有监督背景消除网络, 其结构如图 1 上半部分所示。本文方法并不是完全丢弃图像的背景信息, 而是使

网络自动学习适当消除背景, 保留前景。

为了实现“背景消除”, 首先需要对行人图像进行前景分割。本文将全卷积神经网络 (FCN, fully convolutional network)^[23]与宏-微对抗网络 (MMAN, macro-micro adversarial network)^[24]结合进行前景分割。FCN 可以得到低分辨率行人图像的大致分割轮廓, 但分割不精细、对细节信息不敏感; 对于高分辨图像可以得到精确的分割结果, 但对于低分辨率图像易造成误分割, 将二者相结合可以弥补各自的不足。本文首先在行人分割数据集 LIP (look into person)^[25]上对 MMAN 和 FCN 的网络模型进行训练, 然后将训练得到的网络模型分割行人再识别数据集。

为使网络自动学习弱化背景, 将 ResNet50 网络得到的特征图进行 Sigmoid 函数激活, 生成二值激活结果图, 以监督特征提取网络。二值分割结果要求图像最大值接近 1, 最小值接近 0, 但直接使用 Sigmoid 函数对特征进行激活, 其激活结果的最大值和最小值并不接近 1 和 0, 因此需要先对特征图进行线性拉伸。根据 Sigmoid 函数自变量与值域的关系, 将特征图线性拉伸 [5,5] 的取值范围, 则 Sigmoid 函数的激活结果在 (0,1) 范围。特征值的线性拉伸如式(3)所示。

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} (y_{\max} - y_{\min}) + y_{\min} \quad (3)$$

其中, x 表示拉伸前的特征值, y 表示拉伸后的特征值, x_{\max} 和 x_{\min} 表示拉伸前特征值的最大和最小值, y_{\max} 和 y_{\min} 表示拉伸后的最大和最小特征值, 取值分别为 $y_{\max}=5$, $y_{\min}=5$ 。将线性拉伸后的结果作为输入, 经过 Sigmoid 激活得到激活图。计算激活图和行人分割结果的 MSE 损失, 以监督网络更多地提取行人前景特征。MSE 损失 L_{MSE} 计算式为

$$L_{MSE}(\mathbf{M}, \mathbf{M}') = \frac{1}{N} \sum_{i=1}^N (\mathbf{M}_i - \mathbf{M}'_i)^2 \quad (4)$$

其中, \mathbf{M}_i 表示行人图像分割掩码的第 i 个像素点取值, \mathbf{M}'_i 表示激活图的第 i 个像素点取值, N 表示图像像素点的数量。

将 CE 损失与 MSE 损失之和作为本文方法的总体损失, 并将其作为损失函数对网络进行训练, 损失函数 L 如式(5)所示。

$$L = L_{ID} + L_{MSE} \quad (5)$$

3 实验过程

3.1 数据集与评估方法

本文使用 Market-1501^[26]、DukeMTMC-reID^[27]、MSMT17^[18]这 3 个大规模行人再识别数据集评估本文方法的有效性。

Market-1501 数据集包含 6 个摄像头拍摄得到的 32 217 张固定尺寸行人图像, 行人 ID 数量为 1 501 个。751 个行人 ID 用于网络模型训练, 750 个行人 ID 用于测试。

DukeMTMC-ReID 数据集包含 8 个摄像头拍摄得到的 36 411 张多尺寸行人图像, 行人 ID 数量为 1 404 个。训练集包含 702 个行人, 其余行人作为测试集。测试集中有 2 228 张行人图像被选为查询图像, 其余的 17 661 张图像(包括 702 个作为测试的行人图像及 408 个作为干扰的行人图像)作为候选集。

MSMT17 数据集包含 15 个摄像头拍摄得到的 126 441 张多尺寸行人图像, 有效行人 ID 数量为 4 101。训练集包含 1 041 个行人的 32 621 张图像, 测试集包含 3 060 个行人的 93 820 张图像。对于测试集, 11 659 张图像作为查询集, 其余 82 161 张图像作为候选集。

使用累计匹配特性(CMC, cumulative match characteristic)和平均准确率(mAP, mean average precision)作为评价指标。CMC 曲线主要反映模型的准确率, 常以 Rank- n 的形式表现, Rank- n 表示前 n 个匹配结果中正确匹配的比例, 本文使用 Rank-1、Rank-5、Rank-10 作为参数。mAP 表示算法在全部测试数据上的平均性能, 兼顾准确率和召回率。AP 是某个类别所有返回结果的平均准确率, 如式(6)所示。

$$AP = \frac{1}{n} \sum_{i=1}^n \frac{p_i}{l_i} \quad (6)$$

其中, p_i 表示第 i 个正确匹配的目标行人图像在查询序列中的位置, l_i 表示第 i 个正确匹配的目标行人图像在候选集重新排序中的位置, n 表示和查询图像正确匹配的图像个数。mAP 是所有类别 AP 的平均值, 如式(7)所示。

$$mAP = \frac{1}{m} \sum_{j=1}^m AP_j \quad (7)$$

其中, m 表示查询集中查询图像的总个数。

3.2 实验参数设置

实验所使用的硬件平台的 CPU 为 Inter Xeon E5-2620 v4, GPU 为 10G NVIDIA GeForce GTX 1080Ti, 使用的操作系统是 Ubuntu16.04。实验使用的深度学习框架 PyTorch 版本为 1.01, 数据处理包 torchvision 版本为 0.22, Python 版本为 3.6, cudnn 版本为 8.0, cuda 版本为 8.0。

训练时每个批次读取 $N=32$ 张图像, 其中每个行人类别随机读取固定的图像数量。每个 batch 读取 8 个行人类别, 每类行人的图像数量为 4。使用图像随机水平翻转实现数据扩增, 输入图像尺寸归一化为 256×128 。使用随机梯度下降(SGD, stochastic gradient descent)算法作为优化器, 设置初始学习率为 0.03, 网络模型的训练迭代次数为 60。

4 结果分析与讨论

4.1 多池化融合网络性能评价

首先对本文构建的 MPF 网络的性能进行评价。为了深入分析 MPF 网络相比于 ResNet50 网络的性能提升, 分别从兼顾全局特征和局部特征以及构建不同池化方式 2 个角度进行消融实验。以 G-L 网络表示兼顾全局特征和局部特征, 但采用单一池化方式的方法。比较 ResNet50 网络、G-L 网络和 MPF 网络在 3 个数据集上的识别性能, 如图 3 所示。对于每张查询图像, 展示前 15 个识别结果, 其中, “×”表示识别错误的行人, “√”表示识别正确的行人。由图 3 可以看出, 识别精度由高到低分别为 MPF 网络、G-L 网络、ResNet50 网络。

ResNet50 网络、G-L 网络和 MPF 网络的定量实验结果如表 1 所示。在 3 个数据集中, G-L 网络的识别性能均明显优于 ResNet50 网络, Market-1501 数据集上的 mAP 和 Rank-1 分别提升了 7.42% 和 5.57%, DukeMTMC-reID 数据集上的 mAP 和 Rank-1 分别提升了 6.87% 和 4.24%, MSMT17 数据集上的 Rank-1 和 mAP 分别提升了 5.15% 和 3.76%。实验结果表明, 在特征提取时兼顾全局特征和局部特征可以提升行人再识别的性能。相较于 G-L 网络, MPF 网络可进一步提升识别性能, 在 Market-1501 数据集上 Rank-1 和 mAP 分别提升了 3.93% 和 8.17%, 在 DukeMTMC-reID 数据集上 Rank-1 和 mAP 分别提升了 5.04% 和 7.19%, 在 MSMT17 数据集上 Rank-1 和 mAP 分别提升了 6.07% 和 6.27%。因此在特征提取时采

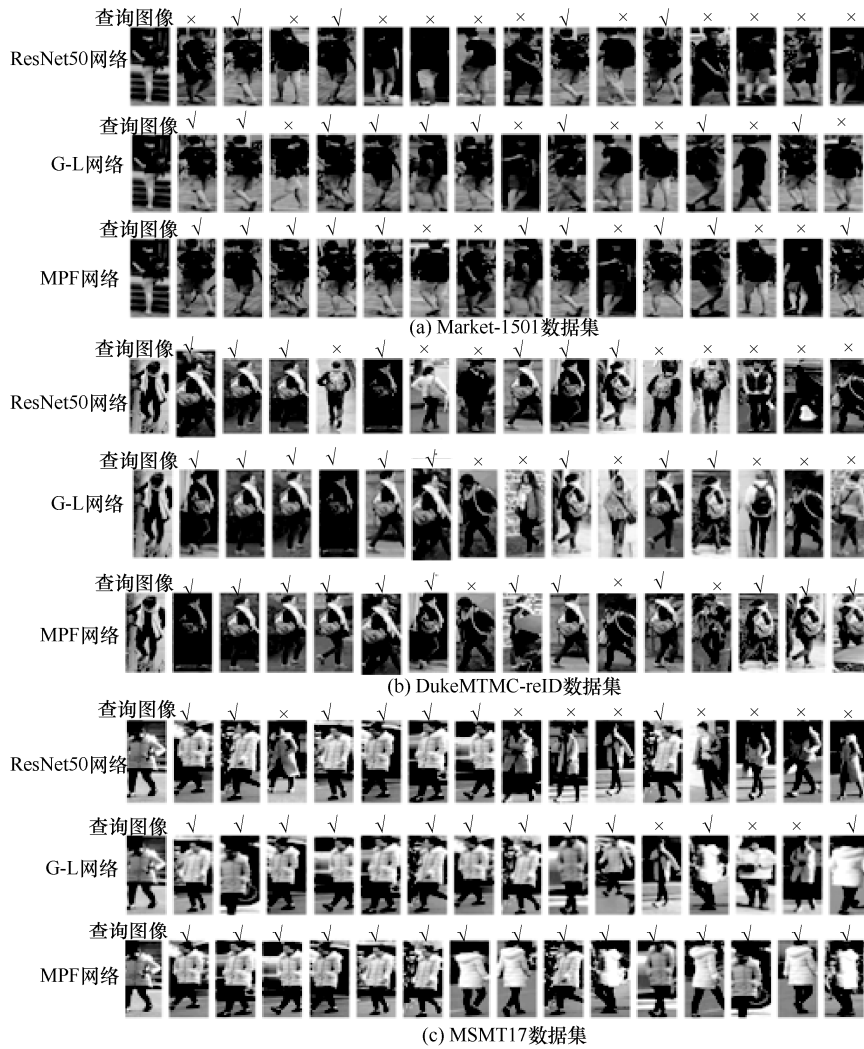


图 3 ResNet50 网络、G-L 网络、MPF 网络在 3 个数据集上的识别结果对比

用不同的池化方式是有效的。MPF 网络结合了全局特征和局部特征且构建了不同的池化方式，从而可以提取更多的有用信息，相比于 ResNet50 网络明显提升了行人再识别性能。

表 1 MPF 网络消融实验

数据集	网络	Rank-1	Rank-5	Rank-10	mAP
Market-1501	ResNet50	82.40%	93.15%	95.50%	60.76%
	G-L	87.97%	95.63%	97.18%	68.18%
	MPF	91.90%	97.01%	98.25%	76.35%
DukeMTMC-reID	ResNet50	74.66%	85.92%	89.79%	54.23%
	G-L	78.90%	88.78%	91.52%	61.10%
	MPF	83.94%	92.11%	94.21%	68.29%
MSMT17	ResNet50	60.85%	74.85%	80.68%	30.67%
	G-L	66.00%	78.97%	83.38%	34.43%
	MPF	72.07%	83.07%	86.50%	40.70%

4.2 行人前景分割性能分析

本节比较 FCN、MMAN 和本文采用的 FCN 与 MMAN 相结合的行人前景分割性能，分割结果示例如图 4 所示。从图 4 可以看出，在低分辨率图像中（图 4 的前两列），MMAN 存在分割不完整的情况。对于行人前景与背景对比度较低的高分辨率图像（图 4 的后两列），FCN 存在分割不完整的情况。将 FCN 与 MMAN 结合，获得了更完整的行人前景。由于行人再识别数据集没有行人前景分割的标准，且本文主要目标是提升跨数据集行人再识别的性能，因此未对行人前景分割进行定量评价。

4.3 背景消除的跨数据集行人再识别性能评价

本文方法的总体损失（total loss）函数为 CE 损失与 MSE 损失之和。MSMT17 数据集下，本文方法在训练过程中的 CE loss、MSE loss 和 total loss 如图 5 所示。由图 5 可知，随着训练过程的进行，CE

loss、MSE loss 和 total loss 逐渐降低。

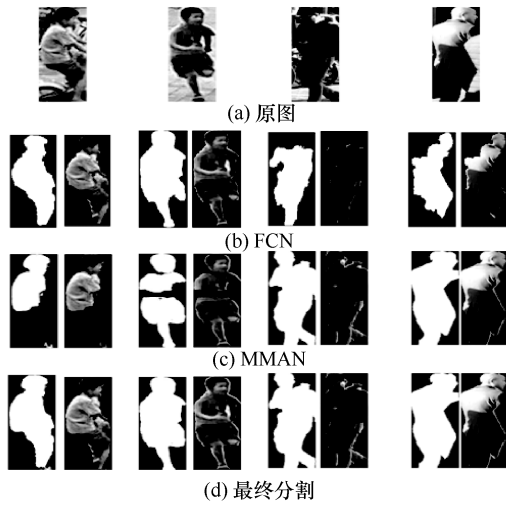


图 4 行人前景分割结果

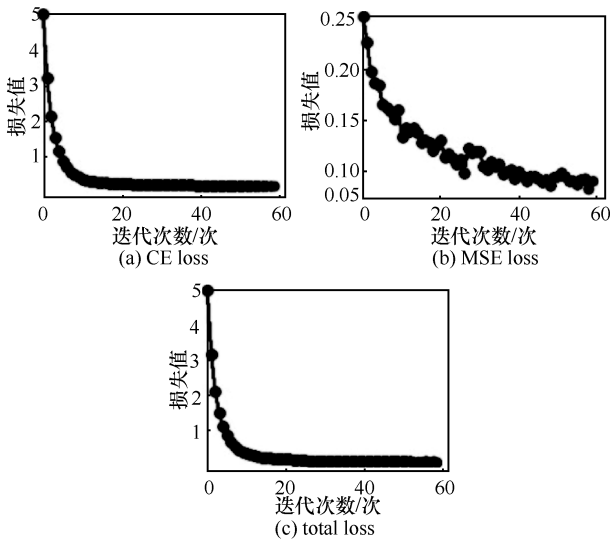


图 5 MSMT17 的训练损失函数曲线

本节首先对特征图激活是否可以引导网络自动学习弱化背景信息进行定性分析。图 6 为融合特征激活损失函数后主干网络得到的特征图。从图 6 可以看出，主干网络提取到的特征图主要关注前景的部分区域，背景区域特征图的取值很低，且不同图像背景信息的弱化是自适应的。



图 6 融合特征激活损失后的特征图

然后对本文方法的跨数据集识别性能进行验证。为了深入分析 MPF 网络和特征级有监督背景消除网络对性能的提升，进行如下消融实验。以 MPF 网络表示仅使用 MPF 网络，以 MPF+背景消除网络表示使用本文方法，将上述方法与 ResNet50 网络进行比较。图 7 和图 8 给出了 MSMT17 数据集作为训练集，测试集分别为 Market-1501 和 DukeMTMC-reID 数据集的跨数据集行人再识别结果。可以看出，MPF+背景消除网络具有更高的识别精度。

定量实验结果如表 2~表 4 所示。可以看出，在 3 个数据集中，MPF 网络的跨数据集识别性能均明显优于 ResNet50 网络，说明 MPF 网络可以提升

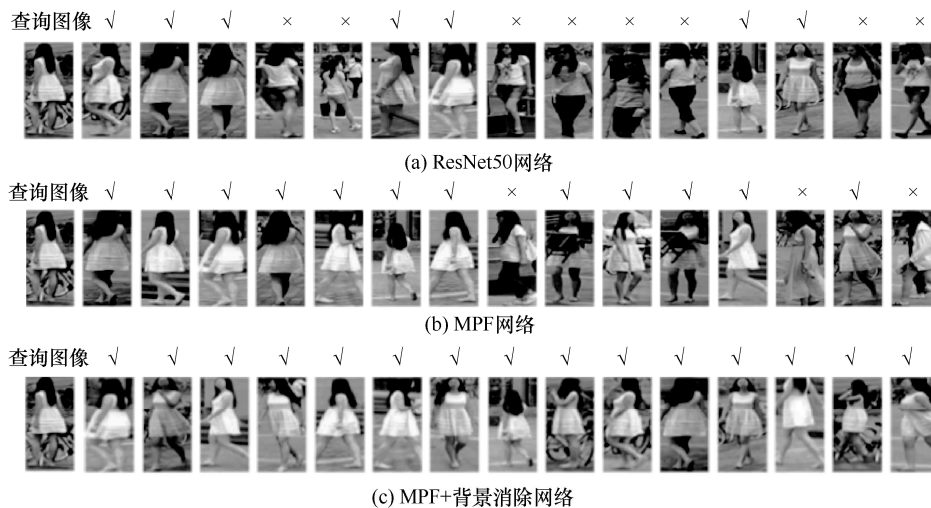


图 7 MSMT17→Market-1501 跨数据集识别结果对比

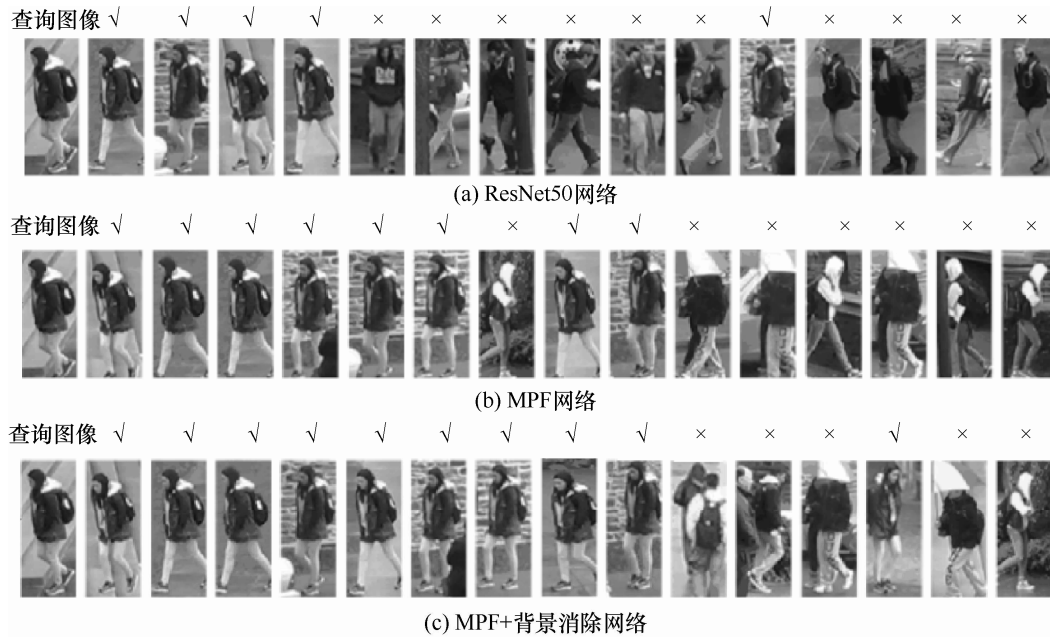


图 8 MSMT17→DukeMTMC-reID 跨数据集识别结果对比

跨数据集行人再识别的性能。相较于 MPF 网络，MPF+背景消除网络可进一步提升行人再识别的性能。当采用 Market-1501 作为测试集，DukeMTMC-reID 和 MSMT17 作为训练集时，相比于 MPF 网络，MPF+背景消除网络的 mAP 和 Rank-1 分别提升了 2.68%/4.37%、3.31%/5.52%。当采用 DukeMTMC-reID 作为测试集，Market-1501 数据集和 MSMT17 作为训练集时，MPF+背景消除网络的 mAP/Rank-1 分别提升了 6.09%/12.52% 和 4.48%/4.82%。当采用 MSMT17 数据集作为测试集，Market-1501 数据集和 DukeMTMC-reID 数据集作为训练集时，MPF+背景消除网络的 mAP/Rank-1 分别提升了 2.89%/ 8.88%和 3.83%/9.30%。实验结果说明，添加特征级有监督背景消除网络后，跨数据集行人再识别的性能得到了提升，表明构建的有监督背景消除网络的有效性。

表 2 Market-1501 的跨数据集再识别结果

训练集	网络	Rank-1	Rank-5	Rank-10	mAP
DukeMTMC-reID	ResNet50	47.60%	65.03%	71.80%	20.45%
	MPF	58.11%	73.97%	79.25%	28.04%
	MPF+背景消除	62.48%	78.12%	84.27%	30.72%
MSMT17	ResNet50	54.93%	71.20%	77.47%	26.29%
	MPF	62.33%	78.51%	84.09%	32.22%
	MPF+背景消除	67.85%	82.52%	86.97%	35.53%

表 3 DukeMTMC-reID 的跨数据集再识别结果

训练集	网络	Rank-1	Rank-5	Rank-10	mAP
Market-1501	ResNet50	32.50%	47.40%	52.52%	16.43%
	MPF	43.05%	57.23%	63.51%	24.64%
	MPF+背景消除	55.57%	68.81%	74.06%	30.73%
MSMT17	ResNet50	56.06%	70.29%	75.58%	34.20%
	MPF	60.09%	74.01%	78.54%	36.97%
	MPF+背景消除	64.91%	77.88	82.10%	41.45%

表 4 MSMT17 的跨数据集再识别结果

训练集	网络	Rank-1	Rank-5	Rank-10	mAP
Market-1501	ResNet50	14.62%	23.82%	28.75%	4.31%
	MPF	23.92%	34.09%	39.91%	7.87%
	MPF+背景消除	32.80%	44.91%	50.45%	10.76%
DukeMTMC-reID	ResNet50	22.43%	34.30%	39.88%	6.95%
	MPF	29.35%	41.59%	47.48%	9.36%
	MPF+背景消除	38.65%	51.20%	56.67%	13.19%

最后对 ResNet50 网络、MPF 网络和 MPF+背景消除网络的训练时间进行比较，如表 5 所示。MPF 网络采用了多池化融合，其训练时间较 ResNet50 网络明显有所增加。MPF+背景消除网络结合了 MPF 网络与有监督背景消除网络，其训练时间进一步增加。

表 5 不同数据集的训练时间

数据集	ResNet50/min	MPF/min	MPF+背景消除/min
Market-1501	66	96	200
DukeMTMC-reID	100	150	325
MSMT17	200	240	560

4.4 与现有跨数据集方法的性能比较

由于 MSMT17 是行人再识别数据集, 目前使用 MSMT17 进行跨数据集实验的方法较少, 因此本文仅使用 Market-1501 数据集和 DukeMTMC-reID 数据集与现有跨数据集方法进行比较。表 6 给出了训练集为 Market-1501 数据集, 测试集为 DukeMTMC-reID 的跨数据集对比结果; 表 7 给出了训练集为 DukeMTMC-reID, 测试集为 Market-1501 的跨数据集对比结果; 由表 6 和表 7 可以看出, 在 DukeMTMC-reID 数据集上测试时, MPF+背景消除网络的 Rank-1 和 mAP 分别达到了 55.57% 和 30.73%, 比现有性能最好的 HHL 方法性能提升了 8.67% 和 3.53%。在 Market-1501 数据集上测试时, MPF+背景消除网络的 Rank-1 和 mAP 分别达到了 62.48% 和 30.72%, 与现有性能最好的 HHL 方法持平。

表 6 Market-1501→DukeMTMC-reID 的跨数据集结果对比

网络	Rank-1	Rank-5	Rank-10	mAP
TJ-AIDL ^[15]	44.3%	59.6%	65.0%	23.0%
MMFA ^[16]	45.3%	59.8%	66.3%	24.7%
HHL ^[17]	46.9%	61.0%	66.7%	27.2%
PT-GAN ^[18]	27.4%	—	50.7%	—
SP-GAN ^[19]	41.1%	56.6%	63.0%	22.3%
ATNet ^[20]	45.1%	59.5%	64.2%	24.9%
MPF+背景消除	55.57%	68.81%	74.06%	30.73%

表 7 DukeMTMC-reID→Market-1501 的跨数据集结果对比

网络	Rank-1	Rank-5	Rank-10	mAP
TJ-AIDL ^[15]	58.2%	74.8%	81.1%	26.5%
MMFA ^[16]	56.7%	75.0%	81.8%	27.4%
HHL ^[17]	62.2%	78.8%	84.0%	31.4%
PT-GAN ^[18]	38.6%	—	66.1%	—
SP-GAN ^[19]	51.5%	70.1%	76.8%	22.8%
ATNet ^[20]	55.7%	73.2%	79.4%	25.6%
MPF+背景消除	62.48%	78.12%	84.27%	30.72%

5 结束语

为减弱背景信息对跨数据集行人再识别性能的影响, 本文提出了结合多池化融合与背景消除网络的跨数据集行人再识别方法。本文构建的多池化融合网络兼顾了全局特征和局部特征, 同时实现了特征的多细粒度表示, 增加了特征向量的复杂性和多样性。构建了特征级有监督背景消除网络, 结合该网络得到的特征激活损失与行人分类损失共同训练识别网络, 减弱了背景信息对识别性能的影响。在 3 个行人再识别数据集 Market-1501、DukeMTMC-reID、MSMT17 上对本文方法进行评估, 实验结果表明, 本文方法能有效提升跨数据集行人再识别性能。

参考文献:

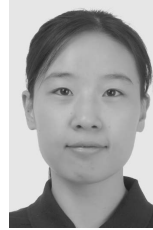
- [1] CHEN K, CHEN Y, HAN C, et al. Hard sample mining makes person re-identification more efficient and accurate[J]. *Nerocomputing*, 2020(382): 259-267.
- [2] SERBETCI A, AKGUL Y S. End-to-end training of CNN ensembles for person re-identification, *Pattern Recognition*[J]. *Pattern Recognition*, 2020, 104: 107319.
- [3] 李幼蛟, 卓力, 张菁, 等. 行人再识别技术综述[J]. *自动化学报*, 2018, 44(9): 1554-1568.
LI Y J, ZHUO L, ZHANG J, et al. Robust resource allocation algorithm for heterogeneous wireless network with SWIPT[J]. *ACTA Automatica Sinica*, 2018, 44(9): 1554-1568.
- [4] GOU M, ZHANG X, RATES-BORRAS A. Person re-identification in appearance impaired scenarios[C]//*Proceedings of British Machine Vision Conference*. Saarland: DBLP, 2016: 1-14.
- [5] MATSUKAWA T, OKABE T, SUZUKI E. Hierarchical gaussian descriptor for person re-identification[C]// *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2016: 1363-1372.
- [6] KOESTINGER M, HIRZER M, WOHLHART P. Large scale metric learning from equivalence constraints[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2012: 2288-2295.
- [7] ZHENG W S, GONG S, XIANG T. Re-identification by relative distance comparison[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2013, 35(3): 653-668.
- [8] YI D, LEI Z, LIAO S, et al. Deep metric learning for person re-identification[C]// *International Conference on Pattern Recognition*. Stockholm: Institute of Electrical and Electronics Engineers Incorporated, 2014: 34-39.
- [9] YAO H, ZHANG S, ZHANG Y, et al. Deep representation learning with part loss for person re-identification[J]. *IEEE Transactions on Image Processing*. 2019, 28(6): 2860-2871.
- [10] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional base-

- line)[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 501-508.
- [11] FU Y, WEI Y, ZHOU Y, et al. Horizontal pyramid matching for person re-identification[J]. arXiv Preprint, arXiv:1804.05275, 2018.
- [12] QI L, WANG L, HUO J, et al. A novel unsupervised camera-aware domain adaptation framework for person re-identification[C]// Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 8079-8088.
- [13] LI Y, LIN C, LIN Y, et al. Cross-dataset person re-identification via unsupervised pose disentanglement and adaptation[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 7918-7928.
- [14] HUANG H, YANG W, CHEN X, et al. EANet: enhancing alignment for cross-domain person re-identification[J]. arXiv Preprint, arXiv: 1812.11369, 2018.
- [15] WANG J, ZHU X, GONG S, et al. Transferable joint attribute-identity deep learning for unsupervised person re-identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 2275-2284.
- [16] LIN S, LI H, LI C, et al. Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification[J]. arXiv Preprint, arXiv:1807.01440, 2018.
- [17] ZHONG Z, ZHENG L, LI S, et al. Generalizing a person retrieval model hetero-and homogeneously[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 176-192.
- [18] WEI L, ZHANG S, GAO W, et al. Person transfer gan to bridge domain gap for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 79-88.
- [19] DENG W, ZHENG L, YE Q, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 994-1003.
- [20] LIU J, ZHA Z, CHEN D, et al. Adaptive transfer network for cross-domain person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 7202-7211.
- [21] TIAN M, YI S, LI H, et al. Eliminating background-bias for robust person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 5794-5803.
- [22] HE K, ZHANG X, REN S. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.
- [23] LONG J, SHELHAMER E, DARRELL T, et al. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 3431-3440.
- [24] LUO Y, ZHENG Z, ZHENG L, et al. Macro-micro adversarial network for human parsing[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 424-440.
- [25] GONG K, LIANG X, ZHANG D, et al. Look into person: self-supervised structure-sensitive learning and a new benchmark for human parsing[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017:

6757-6765.

- [26] ZHENG L, SHEN L, TIAN L. Scalable person re-identification: a benchmark[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 1116-1124.
- [27] ZHENG Z, ZHENG L, YANG Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 3774-3782.

[作者简介]



李艳凤（1988-），女，河北廊坊人，博士，北京交通大学副教授、博士生导师，主要研究方向为图像处理与模式识别。



张斌（1995-），男，山东德州人，北京交通大学硕士生，主要研究方向为图像处理与模式识别。



孙嘉（1995-），女，辽宁丹东人，北京交通大学博士生，主要研究方向为图像处理与模式识别。



陈后金（1965-），男，安徽马鞍山人，博士，北京交通大学教授、博士生导师，主要研究方向为图像处理与模式识别。



朱锦雷（1983-），男，山东曹县人，北京交通大学博士生，主要研究方向为图像处理与模式识别。