

# 基于多尺度批量特征丢弃网络的行人重识别研究

张德祥<sup>1,2\*</sup>, 袁培成<sup>1\*\*</sup>, 王俊<sup>1\*\*\*</sup>

<sup>1</sup>安徽大学电气工程与自动化学院, 安徽 合肥 230601;

<sup>2</sup>安徽三联学院电子电气工程学院, 安徽 合肥 230601

**摘要** 针对行人重识别存在的遮挡和姿势变化问题和目前网络识别率低的缺陷, 提出了不同空间维度的多分支行人重识别网络模型。首先利用 IBN-Net50-a 为基础骨干网络提取特征; 然后对最后两层卷积层融合批量特征丢弃方法, 以增强局部区域专注特征学习; 最后拼接不同维度的特征, 获得更多浅层、深层的有用信息。在网络训练时, 采用三元组损失和标签平滑损失联合策略训练。使用三个常用的基准数据集 Market1501、DukeMTMC-reID、CUHK03 进行实验验证, 并按照主流策略划分数据集。实验结果表明, 所提方法的特征泛化能力较好, 其中在 Market1501 数据集的 Rank-1 和平均准确率 (mAP) 分别达到 95.3% 和 86.8%; 在 DukeMTMC-reID 数据集的 Rank-1 和 mAP 分别达到 88.5% 和 75.9%; 在 CUHK03 数据集的 Rank-1 和 mAP 分别达到 80.9% 和 77.8%。

**关键词** 机器视觉; 行人重识别; 特征丢弃; 特征学习; 三元组损失; 标签平滑损失

中图分类号 TP751.1

文献标志码 A

DOI: 10.3788/LOP202259.1215009

## Person Reidentification Based on Multiscale Batch Feature-Discarding Network

Zhang Dexiang<sup>1,2\*</sup>, Yuan Peicheng<sup>1\*\*</sup>, Wang Jun<sup>1\*\*\*</sup>

<sup>1</sup>School of Electrical Engineering and Automation, Anhui University, Hefei 230601, Anhui, China;

<sup>2</sup>School of Electronic and Electrical Engineering, Anhui Sanlian University, Hefei 230601, Anhui, China

**Abstract** To address the problems of occlusion and postural change in pedestrian reidentification and the low identification rate of current networks, a multibranch pedestrian reidentification network model with different spatial dimensions is proposed. First, IBN-Net50-a serves as the backbone network to extract features. Second, the last two convolution layers are fused with the batch feature-discarding method to enhance the local focal feature learning. Finally, features with different dimensions can be combined to obtain additional valuable information at both the shallow and deep levels. The combined strategy of triplet loss and label smoothing loss is adopted in network training. Further, three commonly used benchmark datasets, Market-1501, DukeMTMC-reID, and CUHK03, are used for experimental verification; the datasets are divided based on the mainstream strategy. Experimental results show that the proposed method achieves an effective feature generalization capability. On the Market1501, DukeMTMC-reID, and CUHK03 datasets, Rank-1 and mean average precision (mAP) values of 95.3% and 86.8%, 88.5% and 75.9%, and 80.9% and 77.8%, respectively, are achieved.

**Key words** machine vision; pedestrian re-identification; feature discarding; feature learning; triplet loss; label smoothing loss

收稿日期: 2021-07-05; 修回日期: 2021-08-10; 录用日期: 2021-08-17

基金项目: 科技部重点研发计划(2018YFB0504604)

通信作者: \*dzxyzd@126.com; \*\*18714921192@163.com; \*\*\*15205659550@163.com

## 1 引言

行人重识别(Person ReID),旨在通过多个不重叠的相机检索感兴趣的人。行人重识别在智能视频监控、智慧安防和人机交互中有着很多的应用,如犯罪嫌疑人搜索、跨境追踪和轨迹分析等,并逐渐成为维护公共安全和社会稳定的重要手段<sup>[1]</sup>。因此行人重识别的课题研究意义重大,同时也伴随着挑战。

在深度学习技术出现之前,早期的行人重识别研究主要集中于如何手工设计更好的视觉特征和如何学习更好的相似性度量<sup>[2]</sup>。如今得益于深度学习,端对端学习策略成为了行人重识别主要研究方法,为了提取能够完整表述行人的鲁棒特征,卷积神经网络(CNN)成为了其中的代表性网络,但是只能关注人体主体部分。为了解决这个缺陷,文献[3]提出了一个双流网络,分别生成外观和身体部位特征图,然后将两个特征映射融合,得到了一个部件对齐特征,其中图像匹配相似度相当于相应体部局部外观相似性的聚集。该方法聚焦于解决人体零件错位问题。手工裁剪的行人边界框具有较少的错误,针对行人重识别中行人检测误差引起的空间错位,文献[4]提出了结合一阶和二阶空间信息的行人重识别网络,通过建模将局部特征集成到依赖模型中,来获取全局特征表示。近年来,注意力机制已经成为一个重要的概念,并在不同领域得到了充分研究,行人重识别领域也大放光彩,例如文献[1]、文献[5]、文献[6]所研究的内容。文献[7]提出了注意力模型HA-CNN,用以学习细节特征,目的是学习一套和谐注意力,用于全局和局部特征表示,最大限度地提高对特征的识别能力和简单框架间的兼容性。局部特征可以关注更多的细节,应对遮挡、姿态变化等问题时会有更好的表现<sup>[6]</sup>。

单独的局部特征也有局限性,局部位置的对齐是能发挥作用的前提,很多方法采用人体姿势估计方法保证位置的对齐,但这种方式并没有利用行人自身的内部信息。文献[8]提出一个特征切块PCB算法,PCB算法以ResNet50为基础网络,将全局平均池化后的特征图水平切分成6份,然后使用6个分类器的交叉熵的和作为损失函数,最后将各局部特征融合,以描述行人。为了消除行人图像中不相关的部分分散注意力,将更多注意力放在行人主体部分,文献[9]设计了端对端网络(MGN),通过缩小

区域并作为分类任务来学习局部特征,并且结合全局特征,获得强的行人特征表示,取得了不错的效果。文献[10]提出卷积层正则化方法(DropBlock),通过特征映射连续区域内的随机丢弃单元,提高了模型的准确性和鲁棒性。文献[11]提出批处理删除块(BDB)网络,该网络随机丢弃一个批量特征图的某一部分,覆盖输入特征图的语义部分,用一种强迫的方式获得局部特征。文献[12]提出IBN-Net50-a网络架构,该架构能极大提升网络的域自适应能力,对行人重识别的识别率提升有明显作用,但是单纯的IBN-Net50-a网络并不能解决行人的遮挡和姿势变化问题。

为了解决行人重识别存在的遮挡和姿势变化问题以及目前网络识别率低的缺陷,本文提出以IBN-Net50-a为骨干网的改进网络模型,对最后两层卷积层融合批量特征丢弃方法,以增强局部区域对特征的学习。采取多尺度特征融合策略,对不同的分支输出的特征图采用不同的池化,提升网络的鲁棒性、学习效率及识别率。

## 2 行人重识别网络模型

### 2.1 IBN-Net50-a 网络结构

IBN-Net50-a是2018年提出的一个针对提升域自适应能力的网络结构,主体结构如图1所示。

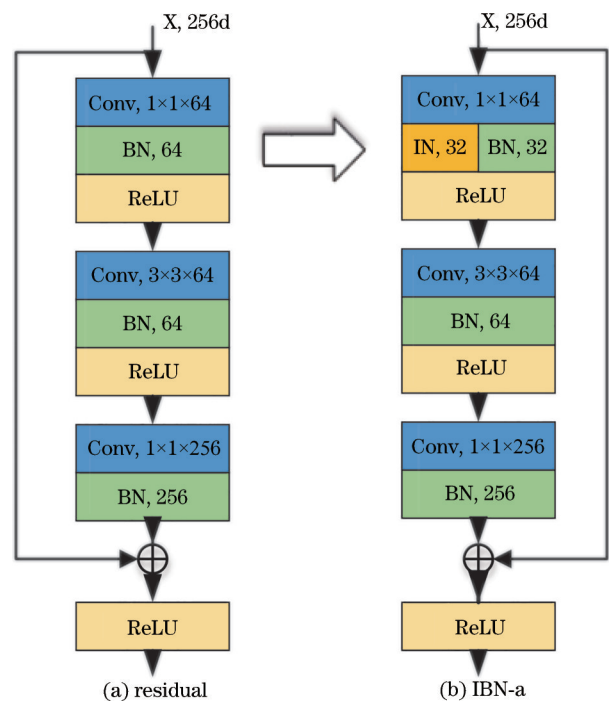


图1 IBN模块结构

Fig. 1 Structure of IBN module

IBN 是 IN(instance normalization)和 BN(batch normalization)的精心组合。IN 提取的鲁棒特征是不随一些外观变化而变化的,如颜色、风格、虚拟/真实等,而 BN 则是保留与内容相关的信息。它们都有各自的局限性,BN 使用一个小批量均值和方差规范每个特征通道,IN 则使用单个样本统计信息,显而易见,BN 在加快训练速度方面有巨大优势。IBN 合理地组合使用 IN 和 BN,同时提升了学习能力和泛化能力。

IBN-Net50-a 只在 ResNet50 前面 3 个 Group (Conv2\_x~Conv\_4-x)中加入 IN,其他 Group 不变。对于残差块,如图 1(a)所示,将第一个卷积输出的前面一半通道用 IN,后面一半通道用 BN,如图 1(b)所示。IN 的添加和 BN 的精心组合使 IBN-Net50-a 有更好的泛化能力和更健壮的认识能力。

### 2.2 批量特征丢弃方法

给定主干网络从一批输入图像中计算出的特征图  $T$ ,批量特征丢弃层随机丢弃相同的特征图  $T$  区域,丢弃区域内的所有单位均为零。对特征图进行处理的过程如图 2 所示。

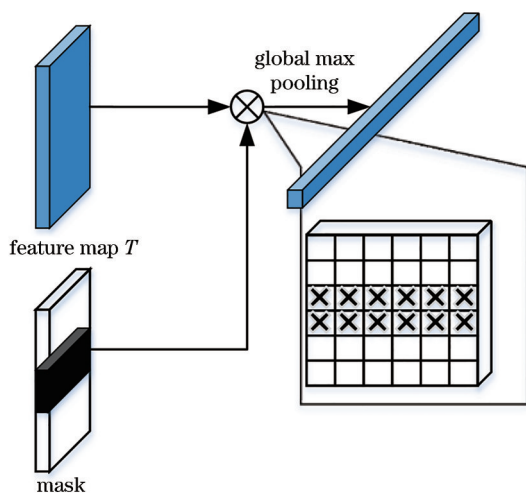


图 2 批量特征丢弃过程

Fig. 2 Process of batch feature discarding

丢弃区域的高度和宽度因任务而异。一般来说,矩形区域为丢弃区域,行人重识别中整个宽度为矩形宽部分。与 DropBlock 不同,批量特征丢弃层在训练中不需要改变丢弃概率超参数,如图 3 所示,其中 FD 表示 feature dropping。

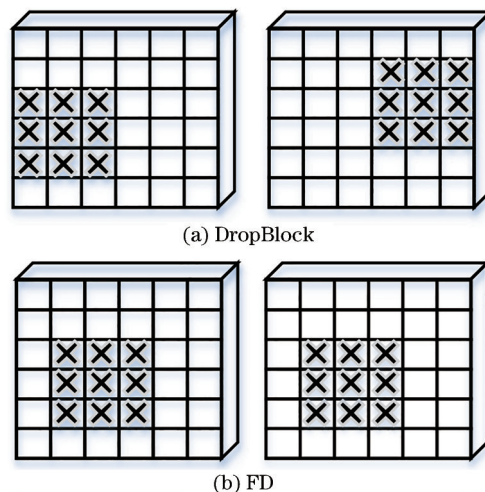


图 3 同一 batch 不同丢弃方法比较

Fig. 3 Comparison of different discarding methods with the same batch

### 2.3 改进 IBN-Net50-a 行人重识别网络结构

IBN-Net50-a 核心是为了提升网络的域自适应能力,行人重识别是跨摄像头检索问题,IBN-Net50-a 对行人重识别的识别率提升有明显作用,但是单纯的 IBN-Net50-a 网络并不能解决存在的遮挡和姿势变化问题。

针对这些问题,提出了不同空间维度的多尺度行人重识别网络模型,以 IBN-Net50-a 为基础骨干网络,对最后两层卷积层融合批量特征丢弃方法,以增强局部区域对特征的专注学习。为了提升网络的鲁棒性和批量特征丢弃方法的有效性,对融合批量特征丢弃方法采用全局最大池化,对其他采用全局平均池化,分别采用卷积降维处理,提高网络的学习效率。改进后的行人重识别网络结构如图 4 所示,主要有骨干网络、卷积层、FD 模块、池化层及全连接层。

对 IBN-Net50-a 进行如下修改,整体网络由 3 个部分组成。

第 1 部分,主要负责从模型不同深度提取不同尺度的特征图 ( $f_1$ 、 $f_2$ 、 $f_3$ ),  $f_1$  的尺寸为  $48 \times 16 \times 512$ ,  $f_2$  的尺寸为  $24 \times 8 \times 1024$ ,最后一个卷积的步长设为 1,  $f_3$  的尺寸为  $24 \times 8 \times 2048$ 。

第 2 部分,对提取到的特征图进行处理,分为 4 个部分,对  $f_2$ 、 $f_3$  分别融合批量特征丢弃方法,如图 2 所示,采用全局最大池化层得到特征向量  $F_2$  维

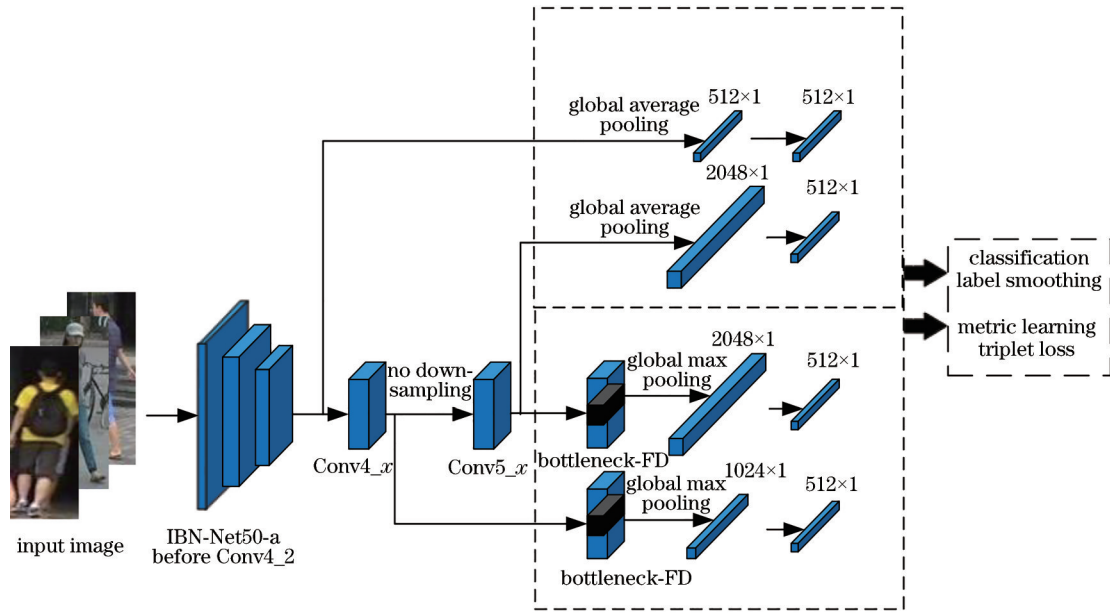


图 4 基于多尺度批量特征丢弃网络的行人重识别网络结构

Fig. 4 Structure of person re-identification network based on multi-scale batch feature discarding network

数为 1024, F3 维数为 2048, 强迫网络获得局部细节特征。此外为了帮助这些融合了 FD 模块的分支进行训练, 对 f3 特征图不做其他处理, 采用全局平均池化层得到特征向量 FF3 维数为 2048。最后对  $1 \times 1$  卷积核进行降维, 维数进一步减少到 512, 对归一化特征进行拼接, 这样是为了获得最后的包含不同层次信息的特征。

第 3 部分, 对处理后的特征向量进行分类和度量学习, 采用标签平滑损失<sup>[13]</sup>和三元组损失<sup>[14]</sup>的联合损失函数进行, 不同尺度特征图池化层之后的特征用三元组损失函数, 最后全连接层用标签平滑损失函数。测试推理时, 则将所有的输出特征拼接起来, 然后通过计算欧氏距离的方式进行排序。

### 3 改进的损失函数

#### 3.1 三元组损失函数

三元组损失函数一开始是为了人脸识别提出的, 度量学习也被广泛应用于行人重识别领域。行人重识别任务中, 三元组损失优化的目的是针对目标样本和正样本以及负样本之间的距离, 使相同行人(类内)距离更近, 不同行人(类间)距离更远, 如图 5 所示。

一个三元组可以被描述为  $(a, p, n)$ , 则三元组损失的思想用欧氏距离形式化表示为

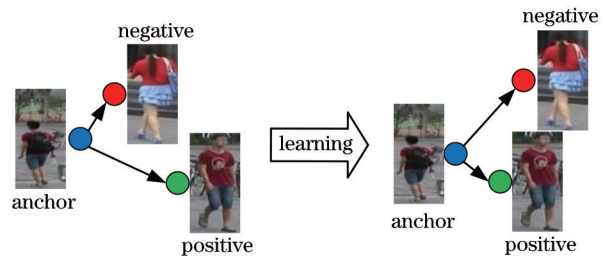


图 5 三元组损失学习过程

Fig. 5 Learning process of triplet loss

$$L_{\text{triplet}} = \sum_i^P \sum_a^K \max\{[d(a, p) - d(a, n) + d_{\text{margin}}], 0\}, \quad (1)$$

式中:  $P$  为行人 ID 数;  $K$  为每个行人图片数量;  $d_{\text{margin}}$  表示一个强制间隔。

#### 3.2 标签平滑损失函数

交叉熵损失函数是分类任务常用的损失函数, 表达式为

$$L_{\text{Cross-Entropy}} = \sum_{i=1}^N -q_i \log_2 p_i, \quad \begin{cases} q_i = 0, n \neq i \\ q_i = 1, n = i \end{cases} \quad (2)$$

式中:  $N$  为数据集行人 ID 数;  $n$  是行人标签;  $p_i$  是网络预测该行人属于标签为  $i$  行人的概率。交叉熵损失函数假设行人标签全部正确, 一些行人可能出现错误标签, 容易造成训练过拟合的现象, 为避免出现这种情况, 使用标签平滑 (label smoothing), 希望网络能接受少量的错误标签<sup>[15]</sup>。分类损失函数使用标签平滑损失函数:

$$L_{\text{Cross-Entropy-ls}} = \sum_{i=1}^N -q_i \log_2 p_i, \begin{cases} q_i = \frac{\varepsilon}{N}, & n \neq i \\ q_i = 1 - \frac{N-1}{N} \varepsilon, & n = i \end{cases}, (3)$$

式中:  $\varepsilon$  为错误率, 则  $1-\varepsilon$  为真实标签, 设置  $\varepsilon=0.1$ 。

### 3.3 联合损失函数

为了提取更强泛化能力的特征, 使用三元组损失函数和标签平滑损失函数优化网络。因为损失函数不同, 最后网络学习提取的特征信息会有较大差别。

行人重识别任务既有度量又有分类, 因此三元组损失和标签平滑损失被应用到本文。但是两个损失函数的机理不同, 三元组损失函数是减小相同行人之间的欧氏距离, 增大不同行人之间的欧氏距离, 标签平滑损失函数是度量两个概率分布之间的不同, 是为了分类任务。则它们同时作用一个目标时会发生损失函数不下降问题。

故用三元组损失函数对特征图池化层之后的特征进行优化, 最后全连接层用标签平滑损失函数优化。故总的损失函数为

$$L_{\text{total}} = \sum_i (L_{\text{triplet}} + L_{\text{Cross-Entropy-ls}}). (4)$$

## 4 实验结果分析

### 4.1 数据集与评价指标

使用四个数据集 Market1501、DukeMTMC-reID、CUHK03、Occluded-Duke 进行实验分析<sup>[16-19]</sup>。数据集的统计信息如表 1 所示, 其中 ID 为行人身份数量, Image 为数据集行人图片数量, Camera 为摄像头数量, Label 为行人边界框获取方式, Hand 表示边界框手动裁剪, Auto 表示边界框自动检测或跟踪。

表 1 数据集的统计数据

Table 1 Statistics for datasets

Dataset	Detail information				
	ID	Image	Camera	Label	Year
Market1501	1501	32668	6	Hand&.Auto	2015
DukeMTMC-reID	1404	36411	8	Hand&.Auto	2017
CUHK03	1467	14097	2	Hand&.Auto	2014
Occluded-Duke	1221	33279	8	Hand&.Auto	2019

一般来说, 深度学习依赖于更多的训练样本, 体现在图片数量和 ID 的增加, 但是同时也增加了重识别困难度。而在实际场景应用中, 相机的数量会大量增加, 这对模型的鲁棒性提出了额外的挑战。因此选择合适的数据集, 获得一个有效且鲁棒的方法至关重要。为了模拟真实场景检测或跟踪错误情况, 分别采用边界框手动裁剪的 CUHK03-Labeled 数据集和自动检测的 CUHK03-Detected 数

据集进行实验结果分析。

为了评估所提方法的有效性, 采用 cumulated matching characteristic (CMC) 曲线和平均准确率 (mAP) 行人重识别领域常用的两个评价指标<sup>[20]</sup>。

### 4.2 实验环境及参数设置

实验硬件环境和软件环境: Centos7 操作系统, 显卡 Tesla V100-PCIE-16GB, python3.6, pytorch1.6。

Backbone 是采用标准的 IBN-Net50-a 作为特征提取网络的基线, 为了节省训练时间, 初始化预训练模型是在 ImageNet 上训练好的 IBN-Net50-a, 输入图像统一调整为  $384 \times 128$ , 训练数据增强方式采用随机水平翻转和 Cutout<sup>[21]</sup> 方法。在批量特征丢弃方法中, 本实验将擦除高度比  $r_h$  设置为 0.3, 将擦除宽度比  $r_w$  设置为 1.0。

$P$  为每个 epoch 中行人的个数,  $K$  为每个 epoch 中每个行人的图片数量, 设置  $P=16, K=4$ , 共进行 400 个 epochs 训练。优化器为 Adam, 权值衰减 (weight decay) 设置为  $5 \times 10^{-4}$ 。学习率初始设置为  $1 \times 10^{-4}$ , 采用了线性 warm-up 策略, 在第 50 个 epoch 时学习率线性增加至  $1 \times 10^{-3}$ , 然后在第 200 个 epoch 时学习率衰减至  $1 \times 10^{-4}$ , 在第 300 个 epoch 之后学习率进一步衰减至  $1 \times 10^{-5}$ 。

### 4.3 实验结果分析

#### 4.3.1 IBN 和多尺度批量特征丢弃模块对实验结果的影响

从表 2 看出, 将基础网络更换为带 IBN 的 ResNet50 深度网络和将批量特征丢弃模块应用到基础的 ResNet50 深度网络中, 精度都有相应的提升。其中 IBN 表示将基础网络 ResNet50 更换成 IBN-Net50-a 深度网络, FD 表示在网络中嵌入批量特征丢弃模块。

对于 Market1501 数据集, 添加批量特征丢弃模块的基础网络的行人重识别精度有一定提升, 而将基础网络 ResNet50 更换成 IBN-Net50-a 深度网络后重识别精度也有提升, 其中 Rank-1 准确率提升了 1.3 个百分点, mAP 提升了 2.4 个百分点。通过将基础网络更换成 IBN-Net50-a 深度网络和嵌入批量特征丢弃模块, 效果也有较大提升, 其中 Rank-1 准确率提升了 1.5 个百分点, mAP 提升了 2.8 个百分点。对于 DukeMTMC-reID 数据集, 通过将基础网络更换成 IBN-Net50-a 深度网络和嵌入批量特征丢弃模块, 可以得出 Rank1 准确率提升了 1.8 个百分点, mAP 提升了 2.8 个百分点。

表 2 不同 ResNet50 得到的实验结果  
Table 2 Experimental results of different ResNet50

Unit: %

Method	Market1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
ResNet50 (without FD and IBN)	93.8	84.0	86.7	73.1
ResNet50 (with IBN)	95.1	86.1	88.1	75.2
ResNet50 (with FD)	94.0	84.4	87.1	74.0
ResNet50 (with FD and IBN)	95.3	86.8	88.5	75.9

4.3.2 参数  $\epsilon$  对实验结果的影响

式(3)中参数  $\epsilon$  用来解决过拟合问题,为了分析  $\epsilon$  对模型性能的影响,选取  $\epsilon=0.05, 0.1, 0.15, 0.2, 0.25, 0.3$ , 分别在 Market1501 和 DukeMTMC-reID 数据集上进行实验。

表 3 和图 6 给出了在 Market1501 和 DukeMTMC-reID 数据集上的实验结果。从表 3 可以看出 Market1501 上  $\epsilon=0.1$  时结果最好, DukeMTMC-reID 上  $\epsilon=0.1, 0.15$  时结果相似。 $\epsilon$  越大,由式(3)可知收敛时的最终损失值也越大,代表给定错误标签量越大,这会使得模型变得不自信,降低识别的准确率。由图 6 可以看出,  $\epsilon$  取值对 Rank-1 的影响要小于对 mAP 的影响,这是因为在模型优化到一定

程度后,最容易匹配到的首位行人已经全部被找到。图 6 结果也表明了参数  $\epsilon$  对模型的影响在两个数据集上的一致性。因此正确选择参数  $\epsilon$  的值对模型性能的提升至关重要。

表 3 不同  $\epsilon$  下的实验结果

Table 3 Experimental results under different  $\epsilon$  Unit: %

$\epsilon$	Market1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
0.05	95.1	86.2	88.2	74.7
0.1	95.3	86.8	88.5	75.9
0.15	94.7	85.6	88.4	75.5
0.2	94.6	84.9	88.1	75.2
0.25	94.1	84.6	87.8	74.3
0.3	93.8	84.5	87.9	73.9

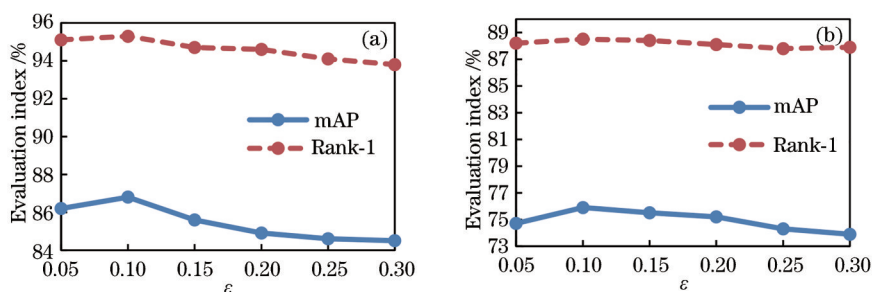


图 6 不同  $\epsilon$  下的实验结果。(a) Market1501 数据集下的实验结果。(b) DukeMTMC-reID 数据集下的实验结果

Fig. 6 Experimental results under different  $\epsilon$ . (a) Experimental results under Market1501 dataset; (b) experimental results under DukeMTMC-reID dataset

4.3.3 数据增强方法对实验结果的影响

为了验证数据增强方法对网络的影响效果,如 Cutout 和 Random Erasing<sup>[22]</sup>, 将它们的随机擦除概率均设为 0.5, 比较两者的不同之处。Cutout 和 Random Erasing 是同一年提出的两种相似的数据增强方法。Random Erasing 一旦进行擦除,则必须在图像上擦除事先设置好的矩形框区域,而 Cutout 一旦进行擦除,则事先设置好的矩形区域不一定全在图像上,两者相比,Cutout 的擦除力度较小,保存了更多的细节信息。

Cutout 和 Random Erasing 的实验结果如表 4 所

示,分别在 Market1501 和 DukeMTMC-reID 数据集上进行实验。其中 RE 表示 Random Erasing, Cut 表示 Cutout,可以看出 Random Erasing 和 Cutout 两种数据增强方法都对网络有提升作用,但是 Cutout 方法明显优于 Random Erasing。

4.3.4 训练技巧对实验结果的影响

为了验证训练技巧对网络的提升效果,依次在相同基础网络上添加训练技巧,而不改变任何训练设置。消融实验的结果显示每个技巧性能得到了提高。所提方法的基础网络在 Market1501 数据集上达到 91.2% 的 Rank-1 准确率和 76.9% 的 mAP,在

表 4 Cut 和 RE 的实验结果  
Table 4 Experimental results of Cut and RE unit: %

Method	Market1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
ResNet50(without FD and IBN)	92.8	82.1	85.1	69.8
ResNet50(without FD and IBN)+RE	93.1	82.6	85.7	71.5
ResNet50(without FD and IBN)+Cut	93.8	84.0	86.7	73.1
ResNet50(with FD and IBN)	94.0	84.4	87.7	73.9
ResNet50(with FD and IBN)+RE	94.3	85.0	87.9	75.4
ResNet50(with FD and IBN)+Cut	95.3	86.8	88.5	75.9

DukeMTMC-reID数据集上达到83.5%的Rank-1准确率和65.1%的mAP。然后,将warm-up策略、Cutout数据增强、标签平滑(LS)、步长为1的卷积层Conv4\_1逐一添加到模型训练过程中。从表5可以看出,warm-up策略和LS应用到网络中后精度提升较为

明显。最后,这些技巧使基础网络在Market1501上获得了95.3%的Rank-1准确率和86.8%的mAP。在DukeMTMC-reID上,Rank-1准确率达到88.5%,mAP达到75.9%。表5的实验结果表明,这些训练技巧将基础网络的mAP提高了10个百分点左右。

表 5 不同训练技巧下的实验结果  
Table 5 Experimental results under different training tricks unit: %

Method	Market1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
Baseline	91.2	76.9	83.5	65.1
+warmup	92.7	82.6	86.0	72.3
+Cutout	93.9	83.9	86.3	73.5
+LS	94.5	85.2	87.6	74.8
+stride=1	95.3	86.8	88.5	75.9

#### 4.3.5 与其他模型比较

在Market1501、DukeMTMC-reID、CUHK03三个基准数据集上的实验结果如表6、7所示。选取的

其他模型具有一定代表性,包括IDE<sup>[23]</sup>、SVDNet<sup>[24]</sup>、HA-CNN<sup>[7]</sup>、SVDNet+Era<sup>[22]</sup>、IANet<sup>[25]</sup>、PCB<sup>[8]</sup>、MGN<sup>[9]</sup>先进算法。

表 6 在Market1501和DukeMTMC-reID上不同方法的实验结果比较  
Table 6 Comparison of experimental results of different methods on Market1501 and DukeMTMC-reID unit: %

Method	Market1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
IDE	72.5	46.0	67.7	47.1
SVDNet	82.3	62.1	76.7	56.8
HA-CNN	91.2	75.7	80.5	63.8
SVDNet+Era	87.1	71.3	79.3	62.4
IANet	94.4	83.1	87.1	73.4
PCB	92.4	77.3	81.9	65.3
PCB+RPP	93.8	81.6	83.3	69.2
MGN	95.7	86.9	88.7	78.4
Ours	95.3	86.8	88.5	75.9
Ours+Re-ranking	95.4	93.1	90.4	87.5

从表6、7的实验结果可以看出,所提行人重识别方法在Market1501数据集上达到了95.3%的Rank-1准确率和86.8%的mAP,表现出了良好的性能。

采用Re-ranking方法<sup>[26]</sup>后,Rank-1准确率和mAP进一步提升至95.4%和93.1%。在DukeMTMC-reID数据集上,Rank-1准确率达到88.5%,mAP达到

表 7 在 CUHK03 上不同方法的实验结果比较  
Table 7 Comparison of experimental results of different methods on CUHK03 unit: %

Method	CUHK03-Labeled		CUHK03-Detected	
	Rank-1	mAP	Rank-1	mAP
IDE	22.2	21.0	21.3	19.7
SVDNet			41.5	37.3
HA-CNN	44.4	41.0	41.7	38.6
SVDNet+Era	49.4	45.0	48.7	37.2
PCB			61.3	54.2
PCB+RPP			62.8	56.7
MGN	68.0	67.4	66.8	66.0
Ours	80.9	77.8	77.9	74.9
Ours+Re-ranking	86.5	88.1	84.4	85.9

75.9%，采用 Re-ranking 方法后，Rank-1 准确率和 mAP 进一步提升至 90.4% 和 87.5%。同样地，针对 CUHK03-Labeled 和 CUHK03-Detected 数据集进行了测试，实验结果表明，所提方法在该数据集上依旧取得了最好的效果，尤其是 CUHK03-Detected 是难度较大的一个数据集。对于 Market1501 和 DukeMTMC-reID 数据集，所提方法取得与 MGN 方法相近的性能，超过了其他大多数方法。综上表明，所提方法确实可以提高行人重识别的准确度。

为了更好地评价所提方法在应对遮挡问题时的识别效果，在遮挡数据集 Occluded-Duke 上进行了有效实验验证，数据集是从 DukeMTMC-reID 数据集挑选出来通过保留遮挡图像，并过滤一些重合图像得到的集合。它包含了 15618 张训练图片，17661 张 gallery 图片和 2210 张 query 遮挡图片。选取具有代表性的行人重识别方法 Part-Aligned<sup>[27]</sup> 和 PCB、具有关键点信息的 SFR<sup>[28]</sup>、针对遮挡行人重识别的 PGFA<sup>[19]</sup> 和 HOREID<sup>[29]</sup>。实验结果如表 8 所示。PCB 和具有关键点信息的 SFR 达到了大约 40% 的 Rank-1 正确率；而对于针对遮挡问题的重识别方法 PGFA 和 HOREID，在遮挡数据集上有明显

表 8 Occluded-Duke 上不同方法实验结果比较

Table 8 Comparison of experimental results of different methods on Occluded-Duke unit: %

Method	Rank-1	mAP
Part-Aligned	28.8	20.2
PCB	42.6	33.7
SFR	42.3	32
PGFA	51.4	37.3
HOREID	55.1	43.8
Ours	58.7	46.4
Ours+Re-ranking	61.7	61.2

的改进，例如 HOREID 在 Occluded-Duke 数据集上达到了 55.1% 的 Rank-1 和 43.8% 的 mAP。所提方法在 Occluded-Duke 数据集上达到了 58.7% 的 Rank-1 和 46.4% 的 mAP，表现出了所提方法在遮挡问题方面的有效性。

为了更直观展示所提方法的重识别效果，在 Market1501 和 DukeMTMC-reID 数据集上都进行了部分结果可视化，如图 7 和图 8 所示，图中显示了待搜索图像的 Rank1~Rank10 结果，搜索正确的边界框为绿色，搜索错误的边界框为红色。

可视化示例选取了具有遮挡、姿势变化的样本。对于 DukeMTMC-reID 数据集中给出的样本，被行人、车辆和障碍物严重遮挡，所提方法通过批量特征丢弃方法迫使网络不仅提取全局特征，也提取出更细节的局部特征，如背包、发型等，从而正确地匹配出同一个人。可视化结果表明，所提方法对行人重识别存在的遮挡和姿势变化问题以及目前网络识别率低的缺陷适应力较好。

## 5 结 论

提出了不同空间维度的多分支行人重识别网络模型。利用 IBN-Net50-a 为基础骨干网络提取特征；然后对最后两层卷积层融合批量特征丢弃方法，最后拼接不同维度的特征。与其他方法相比，所提方法强化了局部特征和全局特征的互补性，拼接不同维度特征，又突出低层有用信息，使得提取的特征具有较好的鲁棒性，能够有更好的识别率。三元组损失和标签平滑损失联合策略更有效地约束类内距离，增大类间距离，避免过拟合。在 Market1501、DukeMTMC-reID、CUHK03 数据集上进行大量实验，与其他方法比较，所提方法的行人



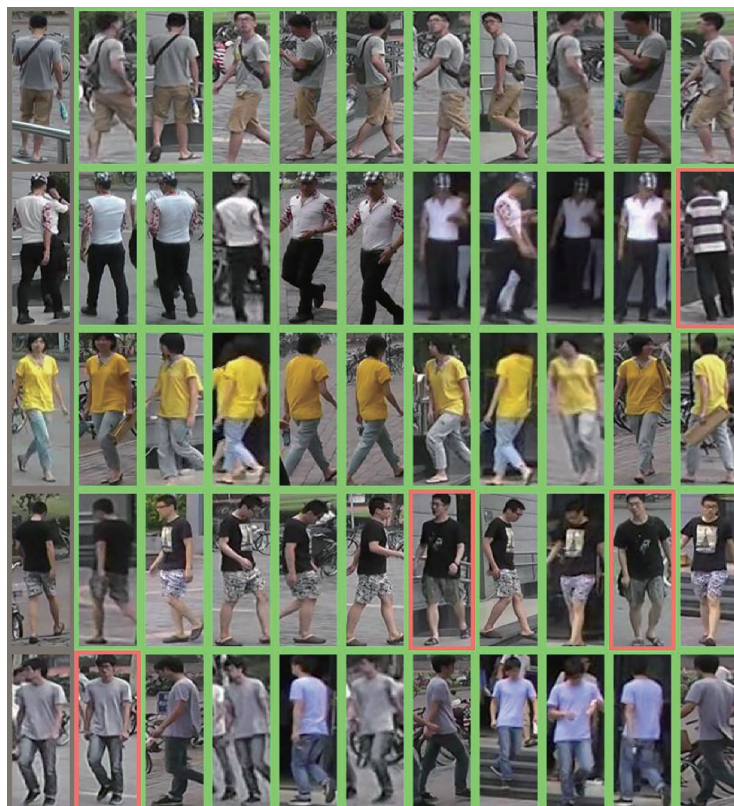


图 7 Market1501数据集可视化结果  
Fig. 7 Visualization results of Market1501

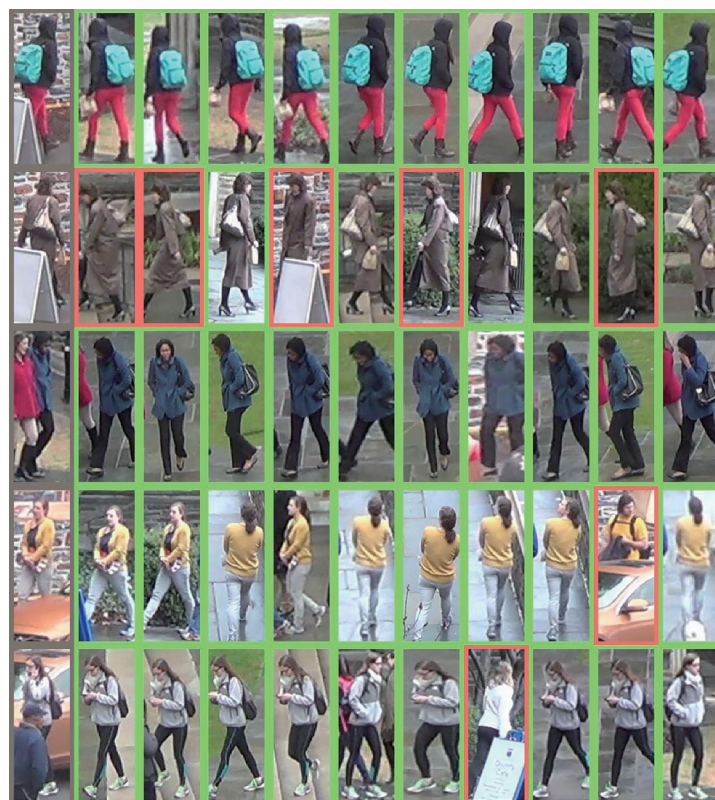


图 8 DukeMTMC-reID数据集可视化结果  
Fig. 8 Visualization results of DukeMTMC-reID

重识别性能有明显提升,在针对遮挡姿势变化问题时也有较好的识别能力。接下来的工作是简化模型复杂度,在保证行人重识别精度的前提下提高模型的计算效率。

### 参 考 文 献

- [1] 张正一, 丁建伟, 魏慧雯, 等. 基于注意力机制的多级特征级联的行人重识别方法[J]. 激光与光电子学进展, 2021, 58(22): 2215003.  
Zhang Z Y, Ding J W, Wei H W, et al. Cascaded multi-level features learning for attention based person re-identification[J]. Laser & Optoelectronics Progress, 2021, 58(22): 2215003.
- [2] 刘紫燕, 万培佩. 基于注意力机制的行人重识别特征提取方法[J]. 计算机应用, 2020, 40(3): 672-676.  
Liu Z Y, Wan P P. Pedestrian re-identification feature extraction method based on attention mechanism[J]. Journal of Computer Applications, 2020, 40(3): 672-676.
- [3] 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.
- [4] 刘莎, 党建武, 王松, 等. 结合一阶和二阶空间信息的行人重识别[J]. 激光与光电子学进展, 2021, 58(2): 0215005.  
Liu S, Dang J W, Wang S, et al. Person re-identification based on first-order and second-order spatial information[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0215005.
- [5] 邬可, 张宝华, 吕晓琪, 等. 基于压缩激励残差网络与特征融合的行人重识别[J]. 激光与光电子学进展, 2020, 57(18): 181007.  
Wu K, Zhang B H, Lü X Q, et al. Person re-identification based on squeeze and excitation residual neural network and feature fusion[J]. Laser & Optoelectronics Progress, 2020, 57(18): 181007.
- [6] 李聪, 蒋敏, 孔军. 基于多尺度注意力机制的多分支行人重识别算法[J]. 激光与光电子学进展, 2020, 57(20): 201001.  
Li C, Jiang M, Kong J. Multi-branch person re-identification based on multi-scale attention[J]. Laser & Optoelectronics Progress, 2020, 57(20): 201001.
- [7] 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 Press, 2018: 2285-2294.
- [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] Wang G S, Yuan Y F, Chen X, et al. Learning discriminative features with multiple granularities for person re-identification[C]//Proceedings of the 26th ACM international conference on Multimedia, October 22-26, 2018, Seoul, Republic of Korea. New York: ACM Press, 2018: 274-282.
- [10] Ghiasi G, Lin T Y, Le Q V. DropBlock: a regularization method for convolutional networks [C]//Proceedings of the 32nd Conference on Neural Information Processing Systems (NIPS), December 2-8, 2018, Montreal, Canada. New York: Curran Associates, 2018: 10727-10737.
- [11] Dai Z Z, Chen M Q, Gu X D, et al. Batch DropBlock network for person re-identification and beyond[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 3690-3700.
- [12] Pan X G, Luo P, Shi J P, et al. Two at once: enhancing learning and generalization capacities via IBN-net[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11208: 484-500.
- [13] Zheng Z D, Zheng L, Yang Y. A discriminatively learned CNN embedding for person reidentification [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2018, 14(1): 1-20.
- [14] Schroff F, Kalenichenko D, Philbin J. FaceNet: a unified embedding for face recognition and clustering [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 815-823.
- [15] 熊炜, 杨荻椿, 熊子婕, 等. 基于全局特征拼接的行人重识别算法研究[J]. 计算机应用研究, 2021, 38(1): 316-320.  
Xiong W, Yang D C, Xiong Z J, et al. Person re-

- identification algorithm based on global feature stitching[J]. *Application Research of Computers*, 2021, 38(1): 316-320.
- [16] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1116-1124.
- [17] Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking[M]//Hua G, Jégou H. *Computer vision-ECCV 2016 workshops. Lecture notes in computer science*. Cham: Springer, 2016, 9914: 17-35.
- [18] Li W, Zhao R, Xiao T, et al. DeepReID: deep filter pairing neural network for person re-identification [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 152-159.
- [19] Miao J X, Wu Y, Liu P, et al. Pose-guided feature alignment for occluded person re-identification[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 542-551.
- [20] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 2818-2826.
- [21] DeVries T, Taylor G W. Improved regularization of convolutional neural networks with cutout[EB/OL]. (2017-08-15) [2021-02-05]. <https://arxiv.org/abs/1708.04552>.
- [22] Zhong Z, Zheng L, Kang G L, et al. Random erasing data augmentation[EB/OL]. (2017-08-16) [2021-02-06]. <https://arxiv.org/abs/1708.04896>.
- [23] Zheng L, Yang Y, Hauptmann A G. Person re-identification: past, present and future[EB/OL]. (2016-10-10) [2021-02-01]. <https://arxiv.org/abs/1610.02984>.
- [24] Sun Y F, Zheng L, Deng W J, et al. SVDNet for pedestrian retrieval[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 3820-3828.
- [25] Hou R B, Ma B P, Chang H, et al. Interaction-and-aggregation network for person re-identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 9309-9318.
- [26] 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, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 3652-3661.
- [27] Zhao L M, Li X, Zhuang Y T, et al. Deeply-learned part-aligned representations for person re-identification [C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 3239-3248.
- [28] He L X, Sun Z N, Zhu Y H, et al. Recognizing partial biometric patterns[EB/OL]. (2018-10-17) [2021-02-04]. <https://arxiv.org/abs/1810.07399>.
- [29] Wang G A, Yang S, Liu H Y, et al. High-order information matters: learning relation and topology for occluded person re-identification[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 6448-6457.