

吕泽, 王进, 张琳钰, 等. 基于特征去噪的行人再识别非局部防御方法[J]. 智能计算机与应用, 2025, 15(10): 24-28. DOI: 10.20169/j.issn.2095-2163.251004

# 基于特征去噪的行人再识别非局部防御方法

吕泽, 王进, 张琳钰, 顾翔, 万杰

(南通大学 信息科学技术学院, 江苏 南通 226019)

**摘要:** 针对行人再识别系统的鲁棒性问题, 提出了基于特征去噪的行人再识别非局部防御方法 NFD。首先, 在预处理阶段对图像 RGB 三个通道分别进行通道级随机擦除, 在每个通道中使用随机擦除, 实现图像在不同通道上的数据随机性; 其次, 设计非局部均值、卷积与残差连接构建的特征去噪块, 捕获长距离的依赖关系信息, 并可以与任何现有的卷积神经网络相结合; 最后, 使用难样本三元组损失函数, 并联合标签平滑正则化的交叉熵损失函数监督网络训练。实验结果显示, 在受攻击状态下, 该防御方法能将主流行人再识别系统的平均精度值(mAP)从 2.6% 提升至 82.1%。

**关键词:** 行人再识别; 对抗防御; 随机擦除; 特征去噪; 鲁棒性

中图分类号: TP391.4

文献标志码: A

文章编号: 2095-2163(2025)10-0024-05

## Non-local defense for pedestrian re-identification with feature denoising

LÜ Ze, WANG Jin, ZHANG Linyu, GU Xiang, WAN Jie

(School of Information Science and Technology, Nantong University, Nantong 226019, Jiangsu, China)

**Abstract:** To deal with the issue of robustness of pedestrian re-identification systems, a defense method named NFD (Non-local Feature Denoising) is proposed. Firstly, in the preprocessing stage, RGB channel-level random erasure is applied separately to each of the RGB channels of the image. This use of random erasure in each channel achieves data randomness across different channels. Secondly, a feature denoising block is designed, constructed with non-local means, convolution, and residual connections. This block captures long-range dependency information and can be combined with any existing convolutional neural network. Lastly, a joint hard-sample triplet loss function is employed, combined with the cross-entropy loss function regulated by label smoothing. The experimental results show that under attack, the NFD can increase the mean of Average Precision (mAP) of mainstream pedestrian re-identification systems from 2.6% to 82.1%.

**Key words:** pedestrian re-identification; adversarial defense; random erasing; feature denoising; robustness

## 0 引言

视频监控领域对智能化的要求越来越高, 行人再识别(Pedestrian Re-Identification, Re-ID)已经成为计算机视觉领域的热点研究课题<sup>[1]</sup>。行人再识别研究是从多个不重叠区域的拍摄设备所捕获的行人图片库中识别出指定的某个行人图像, 现已在寻找嫌犯和失踪人士、跨摄像头人员追踪和行为研究等场景中<sup>[2-3]</sup>得到广泛运用。

自 2016 年后, 行人再识别逐渐取得了一系列显著的成果, 但是其在各种应用中表现出的脆弱性也随即引起了学界关注<sup>[4-5]</sup>。假如犯罪分子戴着帽子

或眼镜、以及背着挎包, 恶意利用这种对抗攻击来逃跑或欺骗监控系统的搜查, 将会对社会构成重大威胁<sup>[6-7]</sup>。因此, 研究行人再识别系统中对抗防御方法<sup>[8]</sup>至关重要。

## 1 相关工作

### 1.1 行人再识别

Re-ID 方法的研究早期构建不同的深度模型和损失函数来提高识别精度, 后来逐渐转向身体局部特征的对齐, 近年来注意力机制<sup>[9-10]</sup>开始在领域兴起。尽管 Re-ID 系统识别精度在不断提高, 但当面对对抗图案、甚至只是不易被察觉的噪声扰动, 其精

**基金项目:** 国家自然科学基金(62002179); 南通市基础科学研究项目(JC22022061)。

**作者简介:** 吕泽(1998—), 男, 硕士研究生, 主要研究方向: 计算机视觉。

**通信作者:** 王进(1981—), 男, 博士, 副教授, 主要研究方向: 人工智能。Email: wj@ntu.edu.cn。

**收稿日期:** 2024-01-08

哈尔滨工业大学主办 ◆ 学术研究与应用

度都会有较大下降<sup>[11-12]</sup>。

## 1.2 对抗防御

对抗攻击带来的扰动会在网络的特征图中产生巨大的噪声,因此 Liao 等学者<sup>[13]</sup>提出了高层表征引导去噪器(High-Level representation guided denoiser, HGD),使用目标模型输出的重建损失替换像素级损失函数,帮助减少噪声并保留图像的语义信息。2022 年, Gong 等学者<sup>[14]</sup>提出了一种联合防御方法,采用对抗训练、像素掩蔽和颜色反转等多种防御策略。

然而,基于对抗训练的防御方法会降低行人再识别模型的识别率,并且可能会被优化的对抗攻击方法规避。为了解决以上问题,本文提出了一种非局部特征去噪防御方法 NFD,借助通道级随机擦除提高训练泛化能力,实现行人再识别系统的对抗防御。

## 2 本文方法 NFD

### 2.1 总体框架

网络模型结构如图 1 所示。在 ImageNet 数据集上预先训练好的 ResNet50 模型作为主干网络,对该网络进行了修改,去掉了网络最后的空间下采样部分、全局平均池化层和全连接层,使得网络生成更大尺寸的特征图。

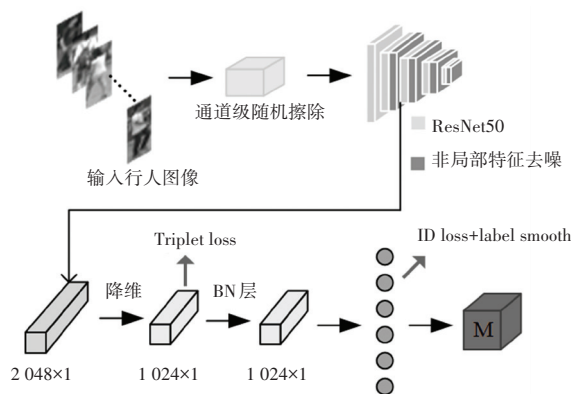


图 1 NFD 网络模型结构

Fig. 1 NFD network model structure

### 2.2 通道级随机擦除

为了提高卷积神经网络的泛化能力,本文采用通道级随机擦除(Channel-level Random Erasing, CRE),其基本思想是给定一个预定义的擦除概率,但与普通随机擦除不同,训练过程中在不同通道(R、G、B)图像的任意位置上选择一个随机大小的矩形区域,然后用所有 3 个通道的随机值替换该矩形区域内的像素值,从而模拟不确定性遮挡带来的对抗噪声,使训练更接近实际场景中的复杂情况。

随机擦除方法的实现过程如下:

考虑一张输入图像  $I$ , 有一定的概率  $P$  被选中进行擦除,则剩下的概率  $1 - P$  保持不变。记随机擦除的图像长为  $H$ , 宽为  $W$ , 则其面积  $S$  大小为:

$$S = W \times H \quad (1)$$

对于随机擦除的区域  $I_e$ , 随机初始化其面积为  $S_e$ , 首先要确定擦除区域的面积比例, 满足以下条件:

$$S_l < \frac{S_e}{S} < S_h \quad (2)$$

其中,  $S_l$  表示擦除区域的面积占整个矩形区域的面积的最小比例,  $S_h$  表示最大比例, 作为超参数设定。此外, 擦除区域的长宽比随机初始化为  $r_e$ , 满足以下条件:

$$r_1 < r_e < r_2 \quad (3)$$

其中,  $r_1$  表示长宽比的最小值,  $r_2$  表示长宽比的最大值。计算擦除区域的长和宽公式具体如下:

$$H_e = \sqrt{S_e \times r_e} \quad (4)$$

$$W_e = \sqrt{S_e / r_e} \quad (5)$$

其中,  $S_e$  表示擦除区域的矩形面积;  $r_e$  表示擦除区域的矩形长宽比;  $H_e$  表示擦除区域的矩形长;  $W_e$  表示擦除区域的矩形宽。首先在图像中随机选择一个点  $P = (x_e, y_e)$  作为擦除区域的左上角顶点, 需要满足如下公式:

$$x_e + W_e \leq W \quad (6)$$

$$y_e + H_e \leq H \quad (7)$$

则将  $(x_e, y_e, x_e + W_e, y_e + H_e)$  作为要擦除的矩形区域, 并用范围  $[0, 255]$  中的随机值来填充这个区域内的像素值。如果不符合条件, 则重新选择点  $P$ , 并重复以上的过程, 直到找到符合条件的矩形区域为止。

在每个通道中, 将选定的擦除区域  $S_e$  中的每个像素分配给一个特定的预定值  $\alpha$ 。  $\alpha$  是由每个通道的平均值计算得来的, 代表相应的通道索引。

### 2.3 非局部特征去噪

非局部操作块的设计旨在捕获长距离的依赖关系信息, 并可以与任何现有的卷积神经网络相结合。 Wang 等学者<sup>[15]</sup>展示了非局部操作块对于视频分类、目标检测和分割以及姿态估计任务的重要性。 在所有任务中, 即使添加一个非局部操作块进行网络改进, 准确率都能比基线提高约 1%。

非局部特征去噪模块 NFD 的结构如图 2 所示, 包括非局部操作块、 $1 \times 1$  卷积和残差连接。

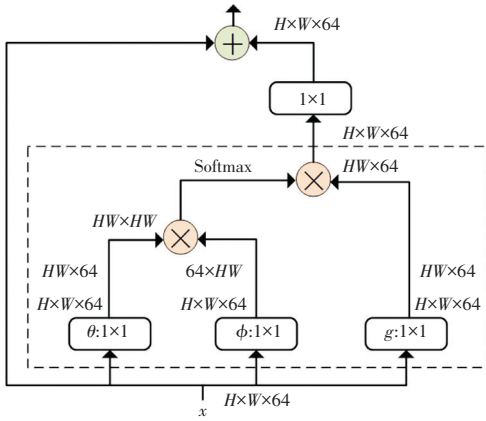


图2 非局部特征去噪块

Fig. 2 Non-local feature denoising block

示意图以64通道为例,标记了特征张量的形状,展示出相应的转置过程,其中 $H$ 和 $W$ 表示特征图的高度和宽度,阴影矩形部分表示非局部操作块在式(1)中的实现。采用基于Softmax的高斯版本进行特征提取和去噪,具体实现是将输入先经过 $1 \times 1$ 卷积层处理,然后进行点积计算得到相似性特征矩阵,相似性特征矩阵经过Softmax函数得到相似性权值,再通过计算得出特征图中的自相关性来去除噪声。将经过去噪的特征表示送入 $1 \times 1$ 卷积层做进一步处理,最后通过残差连接将该卷积层的输出添加到块的输入中,提高模型的性能。

非局部操作块通过采集特征图对所有空间位置的特征进行加权平均,计算出输入特征图 $x$ 的去噪特征图 $y$ 。深度神经网络中通用非局部均值滤波运算方法可表示为:

$$y_i = \frac{1}{C(x)} \sum_{j \in \mathcal{L}} f(x_i, x_j) g(x_j) \quad (8)$$

其中, $x$ 表示输入的行人图像的特征图; $y$ 表示输出的与 $x$ 大小一样的特征图; $x_i$ 对应于输入特征图上位置 $i$ 的一个空间、时间或时空坐标; $\mathcal{L}$ 表示与 $x_i$ 相关的所有可能位置。在计算输出特征图中位置 $i$ 的值 $y_i$ 时,需要考虑 $y_i$ 与所有可能位置 $x_i$ 之间的相似度关系,这些相似度通过加权函数 $f(x_i, x_j)$ 计算得出,反映了位置 $i$ 和 $j$ 之间的联系。在计算特征图在 $j$ 位置的表示时,使用一元函数 $g(x_j)$ 来对特征进行处理。最终的输出特征图 $y$ 通过响应因子 $C(x)$ 进行标准化处理。

加权函数采用嵌入式高斯函数形式,加权函数表示为:

$$f(x_i, x_j) = e^{\frac{1}{d} \theta(x_i)^T \phi(x_j)} \quad (9)$$

其中, $\theta(x_i)$ 和 $\phi(x_j)$ 分别表示对 $x_i$ 和 $x_j$ 的嵌

入函数(由2个 $1 \times 1$ 卷积得到), $d$ 表示通道数。归一化函数的定义公式为:

$$C = \sum_{j \in \mathcal{L}} f(x_i, x_j) \quad (10)$$

NFD在ResNet50中添加4个非局部特征去噪块,分别添加在Res2、Res3、Res4和Res5的最后一个残差块之后。为了保持在嵌入空间中的一致性,该模型采用了特征降维和批处理归一化(Batch Normalization, BN)技术。

损失函数采用联合训练的方式,使用难样本三元组损失函数和标签平滑正则化的交叉熵损失函数来训练行人再识别模型。

难样本三元组损失函数是在包含 $P$ 个不同行人的批次样本中,每个行人选取 $K$ 张图片,然后将其组成一个大小为 $P \times K$ 的数据集。接着,对于每次训练,从这个数据集中以每个图片作为锚点图片,选择该图片的类内距离最远和类间距离最近的2张图片组成困难三元组,并计算损失函数。研究给出的数学公式如下:

$$L_{\text{tri-hard}} = \frac{1}{P \times K} \sum_{A \in \text{batch}} (\max d_{A, \text{Pos}} - \min d_{A, \text{Neg}} + \alpha)_+ \quad (11)$$

其中, $A$ 表示锚点样本;Pos表示正样本;Neg表示负样本; $\alpha$ 表示阈值参数; $d_{A, \text{Pos}}$ 表示锚点到正样本的距离; $d_{A, \text{Neg}}$ 表示锚点到负样本的距离。

交叉熵损失函数是行人再识别任务中最常用的损失函数,用于衡量模型预测结果与真实标签之间的差异,可由如下公式计算求出:

$$L_{\text{cross-entropy}} = \sum_{i=1}^z -q_i \log(p_i) \quad (12)$$

$$q_i = \begin{cases} 0, & y \neq i \\ 1, & y = i \end{cases} \quad (13)$$

其中, $y$ 表示行人真实标签; $p_i$ 表示输出预测身份概率值。

为了进一步提高模型的鲁棒性,避免直接使用交叉熵损失函数可能会导致过拟合的现象,引入了标签平滑方法,即为网络分配少量错误的标签,计算过程可表示为:

$$q_i = \begin{cases} 1 - \frac{N-1}{N} \varepsilon, & y = i \\ \frac{\varepsilon}{N}, & y \neq i \end{cases} \quad (14)$$

其中, $N$ 表示行人ID数量; $\varepsilon$ 表示错误率,一般取0.1,此时可以得到引入标签平滑的交叉熵损失函数 $L_{\text{lsr}}$ 。

最终,构建的联合损失函数  $L_{\text{sum}}$  计算公式如下:

$$L_{\text{sum}} = L_{\text{tri\_hard}} + L_{\text{lsr}} \quad (15)$$

### 3 实验

选用行人再识别领域中主流且识别精度较高的模型 StrongReID<sup>[16]</sup> 和 AlignedReID++<sup>[17]</sup> 做防御实验。模型 StrongReID 是在只使用全局特征情况下添加了多个识别技巧的强力基线模型。

#### 3.1 实验数据集

本文效仿以往文献防御方法的实验内容,只在数据集 Market-1501 上给出了详细数据。

#### 3.2 评价指标

评价指标使用的是 Re-ID 领域常用的主流评价指标 Rank- $n$  以及 mAP。其中,Rank- $n$  表示匹配的结果中最靠前的  $n$  张图像中有正确结果的概率。mAP 表示平均精度值,是指由精准率和召回率所绘制的 PR 曲线下方的面积。Rank- $n$  及 mAP 值越高,说明匹配的结果越高,即该 Re-ID 模型性能越好。

#### 3.3 实验设置

实验假设是在白盒攻击下进行,攻击者能够访问到训练数据和目标模型,由 Wang 等学者<sup>[18]</sup> 在 2020 年 CVPR 会议上提出来的深度误排序攻击方法 DMR 对 Market1501 模拟噪声攻击,生成独立的对抗性样本数据集。由于一些论文代码和细节未公开,因此采纳防御方法 RE<sup>[19]</sup>、GGPR<sup>[20]</sup>、DSN<sup>[21]</sup> 的相关数据进行对比实验。实验训练基于 Ubuntu18.04 操作系统,深度学习框架基于 PyTorch1.6 版本。

#### 3.4 实验分析

本文方法与现有的防御方法在模型 StrongReID 和 AlignedReID++ 上防御的对比结果见表 1、表 2。

表 1 模型 StrongReID 上的对比结果

Table 1 Comparative results on the StrongReID model %				
模型	mAP	Rank-1	Rank-5	Rank-10
StrongReID	88.3	93.3	96.3	97.7
DMR attack	2.6	0.8	3.1	5.2
RE	71.9	77.0	85.4	87.0
GGPR	75.8	77.4	-	-
DSN	80.0	82.7	88.9	89.2
NFD(本文)	82.1	90.3	91.2	91.8

由表 1 可知,原始模型 StrongReID 在受到攻击之前,在数据集 Market-1501 上识别精度较高,但在受到 DMR 攻击之后出现显著下降、降到 2.6%,Rank-1 命中率直接下降到 0.8%。当采用除了本文以外的最好的 DSN 防御模型后,mAP 精度为

80.0%,Rank-1 命中率为 82.7%,表现出良好的防御能力。采用本文提出的非局部防御方法后 mAP 精度为 82.1%,Rank-1 命中率为 90.3%,比原始模型被攻击后分别提升了 79.5%、89.5%,并且比上述几种防御模型表现更好。比较后可以看出本文方法在 StrongReID 模型上防御 DMR 攻击的有效性。

模型 AlignedReID++ 上的对比结果见表 2。由于 Gong 等学者<sup>[20]</sup> 未能给出在 AlignedReID++ 模型上的防御数据,所以不再与 GGPR 防御模型进行比较。原始模型 AlignedReID++ 在受到攻击之前识别率较高,但在受到 DMR 攻击之后波动明显下降到 1.8%,Rank-1 命中率下降到 1.0%。采用除本文研究以外的最好的 DSN 防御模型后,mAP 精度为 68.0%,Rank-1 命中率为 66.7%。采用本文方法后 mAP 精度为 76.0%,Rank-1 命中率为 89.4%,比原始模型被攻击后分别提升了 74.2%、88.4%。综合比较可以看出本文方法在 AlignedReID++ 模型上能对 DMR 攻击产生有效防御。

表 2 模型 AlignedReID++ 上的对比结果

Table 2 Comparative results on the AlignedReID++ model %				
模型	mAP	Rank-1	Rank-5	Rank-10
AlignedReID++	87.6	91.1	95.5	96.6
DMR attack	1.8	1.0	3.1	5.2
RE	63.6	67.6	75.9	77.7
DSN	68.0	66.7	80.0	85.7
NFD(本文)	76.0	89.4	91.3	93.1

为了进一步加强各个模块防御能力的说服力以及验证各模块的通用性,采用由 DMR 攻击生成的对抗数据集分别在 StrongReID 和 AlignedReID++ 这 2 个模型上进行消融实验。+CRE 表示添加了通道级随机擦除模块后的实验结果;+NFD 表示添加了非局部特征去噪模块后的实验结果,实验结果见表 3。

表 3 消融实验对比结果

Table 3 Comparative results of ablation experiments %					
模型	改进后	mAP	Rank-1	Rank-5	Rank-10
Strong-ReID	DMR attack	2.6	0.8	3.1	5.2
	+CRE	71.9	87.0	88.3	90.1
	+NFD	54.6	71.0	76.4	81.6
	+CRE+NFD	82.1	90.3	91.2	91.8
Aligned-ReID	DMR attack	1.8	1.0	3.1	5.2
	+CRE	73.6	87.6	89.0	90.8
	+NFD	64.6	78.7	89.5	92.3
	+CRE+NFD	76.0	89.4	91.3	93.1

从消融实验结果可以看出,原始模型被 DMR



攻击后,mAP 精度下降至 2.6%、1.8%,Rank-1 命中率下降至 0.8%、1.0%。在单独添加通道级随机擦除模块后,识别性能都有提升,同时加上 2 个防御模块后,识别准确度大幅度提升,比单独使用任何一种的效果更好,这就表明了通道级随机擦除与非局部特征去噪的组合是合适且有效的,达到了更好的防御能力。

## 4 结束语

针对噪声攻击提出了一种基于特征去噪的行人再识别非局部防御方法 NFD,引入通道级随机擦除方法进行数据增强,使网络可以处理行人被遮挡等特殊问题,同时增强网络的泛化能力。在卷积神经网络中添加 4 层非局部特征去噪块,利用特征图中全局范围内不同位置的相似性信息,捕获噪声图像更大范围内的特征,对重要特征赋予更高权重,从而抑制无用特征,最终的特征图经过残差连接实现去噪效果。实验结果表明了本文所提出防御方法的有效性,说明了特征去噪块在提高卷积网络的对抗鲁棒性方面的潜力。

## 参考文献

- [1] 罗浩,姜伟,范星,等. 基于深度学习的行人重识别研究进展[J]. 自动化学报, 2019, 45(11): 2032-2049.
- [2] SZEGEDY C, ZAREMBA W, SUTSKEVER I, et al. Intriguing properties of neural networks[J]. arXiv preprint arXiv, 1312.6199, 2013.
- [3] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explaining and harnessing adversarial examples[J]. arXiv preprint arXiv, 1412.0572, 2014.
- [4] ZENG Xiaohui, LIU Chenxi, WANG Y S, et al. Adversarial attacks beyond the image space[C]// Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 4302-4311.
- [5] ZHAO Zhengyu, LIU Zhuoran, MARTHA L. Towards large yet imperceptible adversarial image perturbations with perceptual color distance[C]// Proceedings of 2020 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 1039-1048.
- [6] BAI Song, LI Yingwei, ZHOU Yuyin, et al. Adversarial metric attack and defense for person re-identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(6): 2119-2126.
- [7] WANG Xueping, LI Shasha, LIU Min, et al. Multi-expert adversarial attack detection in person re-identification using context inconsistency[C]// Proceedings of 2021 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2021: 15097-15107.
- [8] KANNAN H, KURAKIN A, GOODFELLOW I J. Adversarial

- logit pairing [EB/OL]. (2023-04-25). <https://arxiv.org/pdf/1803.06373.pdf>.
- [9] LI Dangwei, CHEN Xiaotang, ZHANG Zhang, et al. Learning deep context-aware features over body and latent parts for person re-identification[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 384-393.
- [10] ZHANG Zhong, ZHANG Haijia, LIU Shuang. Person re-identification using heterogeneous local graph attention networks[C]// Proceedings of 2021 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 12136-12145.
- [11] WANG Zhibo, ZHENG Siyan, SONG Mengkai, et al. advpattern: Physical - world attacks on deep person re-identification via adversarially transformable patterns[C]// Proceedings of 2019 IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2019: 8341-8350.
- [12] BOUNIOT Q, AUDIGIER R, LOESCH A. Vulnerability of person re-identification models to metric adversarial attacks[C]// Proceedings of 2020 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2020: 3450-3459.
- [13] LIAO Fangzhou, LIANG Ming, DONG Yinpeng, et al. Defense against adversarial attacks using high-level representation guided denoiser[C]// Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 1778-1787.
- [14] GONG Yunpeng, HUANG Liqing, CHEN Lifei. Person re-identification method based on color attack and joint defence[C]// Proceedings of 2022 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 4313-4322.
- [15] WANG Xiaolong, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]// Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 7794-7803.
- [16] LUO Hao, GU Youzhi, LIAO Xingyu, et al. Bag of tricks and a strong baseline for deep person re-identification[C]// Proceedings of 2020 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, NJ: IEEE, 2019: 4321-4329.
- [17] LUO Hao, JIANG Wei, ZHANG Xuan, et al. AlignedReID++: Dynamically matching local information for person re-identification[J]. Pattern Recognition, 2019, 94: 53-61.
- [18] WANG Hongjun, WANG Guangrun, LI Ya, et al. Transferable, controllable, and inconspicuous adversarial attacks on person re-identification with deep mis-ranking[C]// Proceedings of 2020 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 339-348.
- [19] ZHONG Zhun, ZHENG Liang, KANG Guoliang, et al. Random erasing data augmentation[J]. arXiv preprint arXiv, 1708.04896, 2017.
- [20] GONG Yunpeng, ZENG Zhiyong, CHEN Liwen, et al. A person re-identification data augmentation method with adversarial defense effect[J]. arXiv preprint arXiv, 2101.08783, 2021.
- [21] 王进, 张荣. 行人再识别系统中无感噪声攻击的防御方法[J]. 计算机应用研究, 2022, 39(7): 2172-2177.