一种上下文感知特征增强的行人重识别方法

刘志刚[1,2]， 王 晴[1]， 李 琪[1]， 杜 娟[1]， 刘苗苗[1,2]
（1. 东北石油大学 计算机与信息技术学院，黑龙江 大庆 163318；
2. 黑龙江省石油大数据与智能分析重点实验室，黑龙江 大庆 163318）

摘 要：针对行人重识别中背景噪声干扰问题，提出了上下文感知特征增强的行人重识别方法。首先，在样本层面设计了自适应随机背景擦除策略，采用距离空间变换和优先级方法，减少背景噪声的引入。其次，在特征层面设计了上下文感知多分支空洞卷积结构，扩大模型感受野，深入挖掘行人与背景的空间关系，有效提取上下文全局和局部细粒度特征。最后，利用正交损失函数优化行人与背景的正交关系，抑制背景噪声干扰，提取除最显著特征之外的一些判别性特征，提高模型的准确性。仿真实验表明：该模型性能优于目前大部分算法，在 DukeMTMC-reID 数据集上，R_1 达到 92.6%，α_{mAP}达到 83.2%，验证了方法的有效性。

关键词：行人重识别；上下文感知；背景抑制；特征增强；正交损失
DOI：10.15938/j.jhust.2026.03.002
中图分类号：TP391.41 **文献标志码**：A **文章编号**：1007-2683(2026)03-0013-11

Context-Aware Feature Enhancement Method for Person Re-Identification

LIU Zhigang[1,2]， WANG Qing[1]， LI Qi[1]， DU Juan[1]， LIU Miaomiao[1,2]
（1. School of Computer and Information Technology，Northeast Petroleum University，Daqing 163318，China；
2. Heilongjiang Petroleum Big Data and Intelligent Analysis Key Laboratory，Daqing 163318，China）

Abstract：To address the issue of background clutter in person re-identification (ReID)，this paper proposes a context-aware feature enhancement method. Firstly，an adaptive random background erasure strategy is designed at the sample level，employing distance transform algorithms and priority-based methods to reduce the introduction of background clutter. Secondly，at the feature level，a context-aware multi-branch dilated convolutional structure is designed to expand the model's receptive field，deeply explore the spatial relationship between the person and the background，and effectively extract both global context and local fine-grained features. Finally，the orthogonal loss function is applied to optimize the orthogonal relationship between person and the background，suppress background clutter，and extract additional discriminative features beyond the most salient ones，thereby improving the model's accuracy. Extensive experimental results demonstrate the effectiveness of each component and the proposed method achieves state-of-the-art results. Specifically，on the DukeMTMC-reID dataset，Rank-1 reaches 92.6% and mAP reaches 83.2%，verifying the method's effectiveness.

Keywords：person re-identification；context-aware；background suppression；feature enhancement；orthogonal loss

收稿日期：2024-07-29
基金项目：国家自然科学基金(42002138)；黑龙江省自然科学基金(LH2020F003)；河北省自然科学基金(D2023107002)；黑龙江省博士后科研启动基金(216230006)；东北石油大学领域软件工程与石油大数据团队项目(2022TSTD-03).
作者简介：王 晴(1998—)，女，硕士研究生；
李 琪(1996—)，女，硕士研究生.
通信作者：刘志刚(1979—)，男，博士，教授，E-mail：ZhigangLiu@nepu.edu.cn.

0 引 言

行人重识别(person re-identification, ReID)利用计算机视觉技术来检索图像或视频序列中是否存在特定的行人,被广泛应用于刑事侦查、智能安防和视频监控等多个领域,是当前研究热点。尽管深度学习推动了 ReID 性能的显著提升,但在实际场景中,由于光照变化、遮挡、图像分辨率低及复杂背景等因素的影响,ReID 技术仍面临着较大挑战。

早期 ReID 方法通常从整张图像中提取行人特征,但也引入了大量背景噪声,干扰识别的准确度。为此,TIAN 等[1]尝试完全分离行人与背景,只提取行人特征,但性能反而下降,证实了背景噪声影响模型性能。SONG 等[2]通过实验完全去除背景,发现这会破坏图像的结构信息。同时,分割还会导致行人轮廓附近的 RGB 梯度急剧变化,使得图像不平滑,进而影响模型的识别能力。其次,有学者则尝试通过姿态估计和关键点检测来专注于行人特征,以减少背景干扰。ZHENG 等[3]和 PANG 等[4]定位人体部位,重点提取行人特征。然而,这些方法存在局限性,当图像质量差或部分行人被遮挡时,关键点检测不准确,导致识别错误;且当行人姿态或外观变化时,固定关键点检测难以自适应,提取的特征仍包含噪声。

为进一步增强对行人特征的关注,研究者们尝试将语义分割或注意力机制与 ReID 方法相结合。YANG 等[5]和 GONG 等[6]通过引入注意力机制来提升模型对关键特征的捕捉能力。此外,LIU 等[7]设计前景与背景分支,借助目标注意力损失相互促进两分支学习。CAI 等[8]提出 MMGA,通过身体掩码引导训练,以增强模型对行人区域的关注。但是此类方法存在局限性:一方面,注意力机制虽然能够突出重要特征,但忽略了特征之间的相互关系和上下文信息,缺乏对全局信息的综合考量。另一方面,分割掩码的质量会直接影响模型性能,如果掩码生成不准确,可能会引入额外的噪声,干扰模型的学习过程,从而影响最终的识别效果。

综上分析,传统方法在处理背景噪声干扰问题时,忽略了行人图像中前景与背景之间的内在关系。WANG 等[9]指出,图像中的每个像素点不是孤立存在的,都与其周围像素存在关联,即“上下文信息”。对于行人图像,部分背景信息可以为行人识别提供重要的线索。因此,简单去除全部背景并不可取,保留部分背景辅助行人识别非常必要。魏明军等[10]提出前景与背景交互融合网络,验证了部分背景信息对识别性能的提升作用。基于此,本文提出可以平衡无关背景和有用上下文信息的自适应背景擦除的数据增强策略。

此外,全局特征反映了行人的整体属性,对图像的尺度和视角变化具有不变性。目前,传统卷积神经网络(convolutional neural networks, CNN)感受野较小,在获取远距离依赖的全局上下文特征中存在局限性。CNN 池化操作虽能扩大感受野,但会降低特征图分辨率,丢失重要细节,影响行人关键特征的匹配。而空洞卷积[11]能够在不牺牲分辨率的同时扩大模型感受野,李晓楠等[12]提出 Res-DU-Net 网络,即通过空洞多尺度特征提取减少下采样来提升分辨率。因此,本文设计了上下文感知的多分支空洞卷积结构,扩大模型感受野捕捉行人与背景的丰富上下文关系,提升行人识别能力。

鉴于以上分析,为抑制背景噪声干扰并提高全局上下文信息的感知能力,本文提出了上下文感知特征增强的行人重识别网络(context-aware feature enhancement for person re-identification networks, CA-FE-Net)。该网络从样本预处理和特征提取两方面深入挖掘行人与背景之间的空间关系,在抑制背景噪声的同时提取上下文语义的全局和细粒度特征,增强特征表达能力。本文的主要工作如下:

1)在样本预处理阶段,提出自适应背景随机擦除模块,结合距离空间变换和优先级评分策略,实现自适应背景擦除,减少背景噪声的引入。

2)在特征提取阶段,设计了上下文感知的多分支空洞卷积结构,在不牺牲分辨率的前提下扩大模型感受野,挖掘行人与其周围背景之间的上下文语义关系,增强模型对行人的感知能力,提取判别性的行人特征。

3)提出伪对抗学习的背景抑制策略,采用正交损失函数[13]监督前景与背景向量之间的正交性学习,抑制背景噪声。学习除显著性特征外的一些次显著性辨别特征,增加行人特征的多样性。

1 本文方法

上下文感知特征增强的行人重识别网络(CA-FE-Net)结构如图 1 所示。该网络主要由 3 个模块组成:自适应随机背景擦除(adaptive random background erasing augmentation, RBEA)、上下文感知的多分支空洞卷积结构(context-aware multi-branch di-

lated convolutional structure, CMDC)及伪对抗学习的背景抑制模块(pseudo-adversarial background suppression module, PAB)。首先,在数据预处理阶段,CAFE-Net 对原始图像进行动态处理,生成 3 种不同类型的训练样本,包括随机擦除背景的全局图像、仅包含前景的图像及仅包含背景的图像作为多样化的输入源。其次,CAFE-Net 采用 ResNet50 作为主干网络,包含 4 个残差块(Conv1 ~ Conv4),并将 Conv1 和 Conv2 中 3 × 3 卷积替换为所设计的 CMDC 结构,以增强模型对行人上下文信息的感知能力。而后,在 Conv4 后引入 PAB 模块,监督前景与背景特征的正交性学习,增加行人特征的多样性,同时进一步抑制噪声干扰。最后,CAFE-Net 采用多个损失函数进行联合训练,以提升模型识别性能。

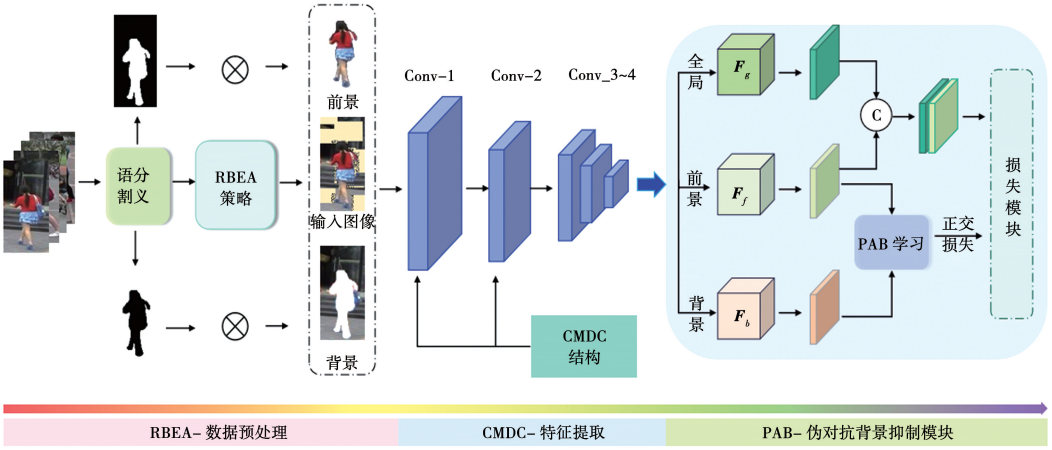


图 1　CAFE-Net 网络结构

Fig. 1　The structure of CAFE-Net

1.1　自适应随机背景擦除 RBEA

本文在数据预处理阶段,设计了自适应随机背景擦除策略 RBEA,在保留上下文背景信息的同时,有效抑制无关背景噪声的干扰。与传统的随机擦除 REA[14]方法相比,RBEA 避免了 REA 可能误擦行人关键特征的问题,二者对比如图 2 所示。



(a) 随机擦除 REA



(b) 随机背景擦除 RBEA

图 2　随机擦除与随机背景擦除对比图

Fig. 2　Comparison chart of random erasure and random background erasure

1.1.1　无关背景与上下文信息分析

研究表明,行人图像中的上下文信息与无关背景之间并无绝对清晰的界限,而本文从空间相对距离出发量化无关背景与上下文信息的关系。通常情况下,距离行人轮廓较近,且与行人外观或行为具有直接或潜在关联的背景区域更可能被视为上下文信息,可帮助模型更好地理解行人位置、大小、形状及其与其他对象的关系,辅助行人识别。相反,与行人位置距离较远或在远景中与行人身份或行为无关的物体,则更可能被视为无关背景噪声,容易对行人识别造成干扰。本文设计的 RBEA 策略,可根据行人轮廓与背景区域的相对距离确定擦除的位置,具体擦除过程见 1.1.2 节。

1.1.2　RBEA 方法

Step1 定位背景区域。通过已训练的 DeepLabv3 + [13]模型对行人图像 I 进行分割处理,生成对应的二值掩码 M,用于区分背景与前景区域。

Step2 背景候选区域筛选。对背景区域中的随机位置选取多个矩形块 I_e^* 作为筛选的目标。

Step3 擦除决策。为不过度擦除背景信息,适当保留行人上下文背景信息,设定阈值 μ,根据背景像素在图像中的比重 R 决定是否进行擦除。当 $R > \mu$ 时,实施擦除操作;否则,图像保持不变,如式(1) ~ (2)所示。

$$R=\frac{N_b}{N_t} \tag{1}$$

$$I_u=\begin{cases} I*(1-M_{I_e^*}), & R>\mu \\ I, & R\leqslant\mu \end{cases} \tag{2}$$

式中：N_b 为背景像素个数；N_t 为图像总像素个数；I_u 为已擦除更新后的行人图像；$M_{I_e^*}$ 为选定擦除的 I_e^* 的二值掩码；I 为原图像；$*$ 为逐元素相乘。

Step4 确定行人轮廓。首先，对图像前景区域的像素点进行分析，分析其水平、垂直及对角线方向的 8 邻域内是否包含背景像素点。如果前景像素点 8 邻域内至少存在 1 个背景像素点，则该前景像素点被视为边界点，构成行人轮廓集合 P，具体公式如下：

$$P=\begin{cases} (x,y)\mid M(x,y)=1\wedge \\ \exists(x',y')\in N_8(x,y)\wedge M(x',y')=0 \end{cases} \tag{3}$$

式中：(x,y) 为图像像素坐标；$M(x,y)=1$ 表示该像素为前景像素；$\wedge$ 为逻辑与，表示左右两边条件需同时成立；$N_8(x,y)$ 为以 (x,y) 为中心的 8 邻域像素集合；$\exists$ 为存在量词，表示至少存在 1 个；$M(x',y')=0$ 表示 (x',y') 为背景像素。

Step5 距离计算。为高效计算目标背景块到行人轮廓的距离，本文选用距离变换算法[15]进行计算。算法核心思想是通过计算图像中像素到特定目标区域的距离，将图像像素值转换为距离矩阵。本文则是计算背景中的像素点与行人轮廓像素点之间的距离，并选取最小值作为该像素点到行人轮廓的距离 $D(x,y)$。在此基础上，背景块 I_e 与行人的距离是通过计算 I_e 内所有像素点到行人轮廓集合的最小距离的平均值得到的，即 $T_p(I_e)$，具体公式如下：

$$D(x,y)=\min_{(x',y')\in p}\sqrt{(x-x')^2+(y-y')^2} \tag{4}$$

$$T_p(I_e)=\frac{1}{|I_e|}\sum_{(x,y)\in I_e}D(x,y) \tag{5}$$

式中：(x,y) 为 I_e 内的某个像素点；(x',y') 为行人轮廓集合 P 中的像素点坐标；I_e 为待擦除背景块；$|I_e|$ 为 I_e 内像素点的个数。

Step6 优先级筛选。为实现自适应擦除，引入了基于背景块与行人轮廓距离的优先级筛选策略。首先，将所有待处理背景块按其与行人轮廓的平均距离 $T_p(I_e)$ 降序排列，距离越远的背景块擦除优先级越高。随后，按此顺序依次擦除背景块，并更新剩余背景像素占图像总像素的比例，当该比例小于阈值 μ 时终止操作，实现擦除强度的控制。

Step7 擦除区域填充。为维持图像的视觉连贯性，本文采用平滑模糊填充擦除区域，缓解由擦除引起的边缘梯度变化问题。

1.2 上下文感知的多分支空洞卷积结构 CMDC

在 ReID 任务中，上下文信息作为视觉语义理解的关键辅助要素，能够提高特征表达的判别性。然而，标准 3×3 卷积的感受野较小，难以在早期阶段捕获长距离依赖信息，导致行人特征的上下文信息不足，难以捕捉复杂的行人空间位置信息，影响全局特征的判别能力。为此，本文设计了上下文感知的多尺度空洞卷积模块 CMDC，如图 3 所示。具体而言，CMDC 利用不同空洞率的卷积核扩大模型感受野，从而捕捉上下文多尺度信息，使模型在浅层空间中有效提取全局行人信息，提升模型对行人的感知能力。

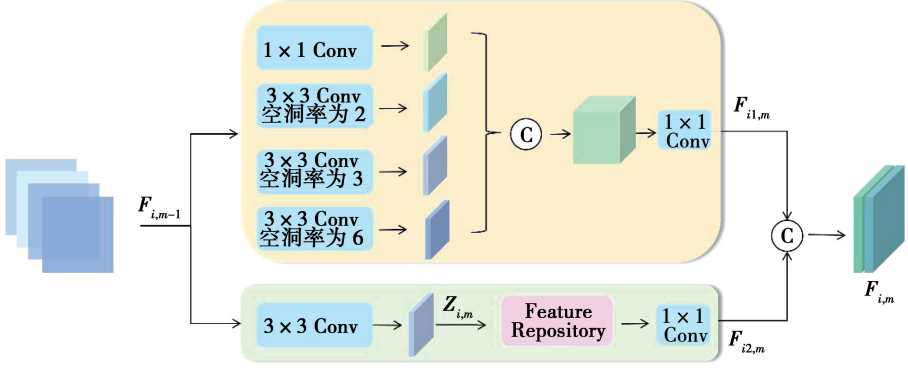


图 3 CMDC 卷积结构

Fig. 3 The convolution structure of CMDC

CMDC 包括 2 个分支，分别为多尺度空洞卷积分支与标准卷积分支。其中，多尺度空洞卷积分支通过采用不同空洞率的卷积核，实现了对输入特征图的多尺度上下文信息捕获。该分支由多个不同空洞率的 3×3 卷积和 1 个 1×1 卷积串联组成，该分支扩大模型的感受野，使其能够捕获更广泛的行人

上下文信息。此外,1×1 卷积促进了特征图各通道间的信息交互,进一步增强特征的表达能力。经过多分支卷积处理后,通过通道融合操作,获得多尺度信息的特征表示 $\boldsymbol{F}_{i1,m}$。标准卷积分支则用于提取行人的局部特征和轮廓细节。该分支主要由 1 个 3×3卷积层和 1 个 1×1 卷积层组成,负责捕获行人图像中的细粒度特征 $\boldsymbol{F}_{i2,m}$,弥补空洞卷积的不足。为增强特征的表征能力,将 2 个分支输出的特征融合获得 $\boldsymbol{F}_{i,m}$,既含有丰富的上下文全局又有局部细粒度的特征,其公式定义如下:

$$\boldsymbol{F}_{i1,m} = \mathrm{Conv}_{1\times1}(\mathrm{Concat}(\mathrm{DConv}(\boldsymbol{F}_{i,m-1}, k_i, d_i))) \tag{6}$$

$$\boldsymbol{Z}_{i,m} = \mathrm{Conv}_{3\times3}(\boldsymbol{F}_{i,m-1}) \tag{7}$$

$$\boldsymbol{F}_{i2,m} = \mathrm{Conv}_{1\times1}(\boldsymbol{Z}_{i,m}) \tag{8}$$

$$\boldsymbol{F}_{i,m} = \mathrm{Concat}(\boldsymbol{F}_{i1,m}, \boldsymbol{F}_{i2,m}) \tag{9}$$

式中:$\boldsymbol{F}_{i,m-1}$为第 $m-1$ 层的第 i 个特征;DConv 为多分支的卷积结构,其 3 个参数分别为输入特征、卷积核尺寸 k 及空洞率 d 的大小,k 和 d 范围为[0,1,2,3,4,5,6]。其次,传统残差结构跳跃连接通常是上一层输出的结果,而在 CMAD 结构中,为减少特征冗余并缓解梯度消失问题,将标准卷积分支的中间特征图 $\boldsymbol{Z}_{i,m}$作为跳跃连接。残差结构如图 4 所示,具体残差公式为

$$\boldsymbol{F}_{i,m+1} = \delta(\boldsymbol{F}_{i,m}) + \boldsymbol{Z}_{i,m} \tag{10}$$

式中:$\delta(\cdot)$为非线性变换的组合,包括归一化(batch normalization, BN)、激活 ReLU 及 Conv 等操作。

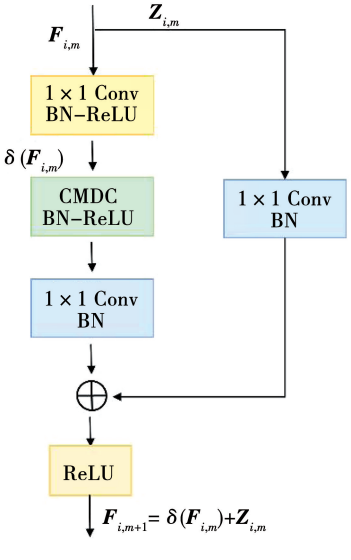


图 4　残差结构

Fig. 4　The structure of residual

1.3　伪对抗学习的背景抑制模块 PAB

为进一步增强行人全局特征的表征能力并抑制背景噪声的影响,本文设计了 PAB 模块。其核心思想是通过增强前景与背景特征向量之间的正交性,减少特征冗余,使模型能够在抑制背景噪声干扰的同时,更好地学习次显著性行人识别特征,从而提取差异化行人特征,提升特征表达能力,结构如图 5 所示。

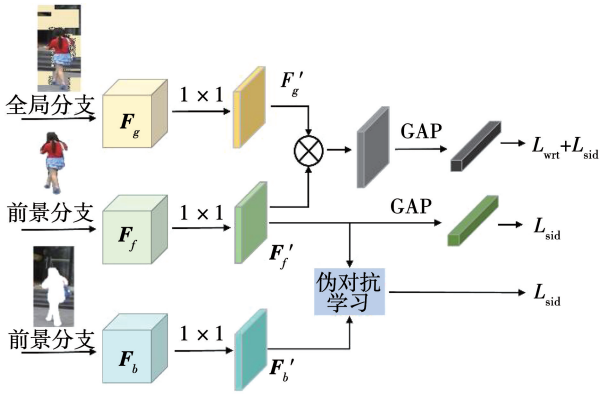


图 5　PAB 结构

Fig. 5　The structure of PAB

1.3.1　正交性

在信息论中,正交用于描述向量或函数间的特定关系[16]。对于 2 个向量 $\boldsymbol{a}$ 和 $\boldsymbol{b}$,若满足运算 $\boldsymbol{a}\cdot\boldsymbol{b}=0$,则称 2 个向量正交。从数学的角度来看,内积可表示为

$$\boldsymbol{a}\cdot\boldsymbol{b} = |\boldsymbol{a}||\boldsymbol{b}|\cos\theta \tag{11}$$

式中:θ 为 $\boldsymbol{a}$ 和 $\boldsymbol{b}$ 之间的夹角。当 2 个向量正交时,$\cos\theta=0$,表示这 2 个向量在各维度之间不存在线性相关性,所携带的信息是相互独立的。因此,在行人重识别的特征空间中,可通过计算 $\cos\theta$ 来衡量前景与背景特征向量之间的正交程度,再通过优化正交损失函数来优化模型的参数。设行人特征向量为 $\boldsymbol{f}_i$ 和 $\boldsymbol{f}_j$,具体计算公式如下:

$$\cos\theta_{ij} = \frac{\boldsymbol{f}_i\cdot\boldsymbol{f}_j}{|\boldsymbol{f}_i||\boldsymbol{f}_j|} \tag{12}$$

1.3.2　PAB

PAB 模块由全局、前景和背景 3 个分支构成。首先,该模块通过特征提取分别获得全局特征 $\boldsymbol{F}_g$、前景特征 $\boldsymbol{F}_f$ 和背景特征 $\boldsymbol{F}_b$,其中 $\boldsymbol{F}\in\mathbb{R}^{C\times H\times W}$,$H$、$W$ 和 C 分别表示特征的高度、宽度和通道数。为使得维度一致,对每个分支进行 1×1 卷积调整特征维度,生成新特征图 $\boldsymbol{F}_g'$、$\boldsymbol{F}_f'$ 和 $\boldsymbol{F}_b'$。随后,将 $\boldsymbol{F}_f'$ 与 $\boldsymbol{F}_b'$ 组成“伪对抗特征对”,构造正交损失函数。通过约束 $\boldsymbol{F}_f'$ 与 $\boldsymbol{F}_b'$ 之间的相似度,增强前景和背景特征向量之间的正交性学习,使模型学习前景行人主体特

征,更好地学习到具有判别性的行人特征,抑制背景噪声,提升模型在复杂背景下的识别准确率。此外,为确保全局特征能够充分利用全局信息,增强行人判别性信息,采用全局平均池化(global average pooling,GAP)提取全局特征向量 $\boldsymbol{\gamma}$,并通过加权正则化方式对三元组损失和标签平滑交叉熵监督优化。另外,前景 $\boldsymbol{F}_f^{'}$ 经过 GAP 之后得到特征向量 $\boldsymbol{\tau}$,公式可定义为

$$\boldsymbol{\gamma} = \mathrm{GAP}(\boldsymbol{F}_g^{'} \otimes \boldsymbol{F}_f^{'}) \tag{13}$$

$$\boldsymbol{\tau} = \mathrm{GAP}(\boldsymbol{F}_f^{'}) \tag{14}$$

式中:$\otimes$为逐元素相乘。特征 $\boldsymbol{F}_f^{'}$ 和特征 $\boldsymbol{F}_b^{'}$ 经过伪对抗学习处理之后获得前景特征向量 $\boldsymbol{f}_i$ 和背景特征向量 $\boldsymbol{b}_i$。

其次,在本文中,正交损失函数基于前景和背景特征向量间的余弦相似度关系判断正交性。余弦相似度是离散的,直接作为损失函数求导困难,不利于模型训练。因此,本文选用指数函数 $y = \mathrm{e}^x$ 来计算离散点的导数。正交损失函数 L_{o} 具体如下:

$$L_{\mathrm{or}} = \exp\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\boldsymbol{f}_i \cdot \boldsymbol{b}_i}{|\boldsymbol{f}_i||\boldsymbol{b}_i|}\right) - 1 \tag{15}$$

$$L_{\mathrm{o}} = \begin{cases} 0, & L_{\mathrm{or}} \leqslant \theta \\ L_{\mathrm{or}} - \theta, & L_{\mathrm{or}} > \theta \end{cases} \tag{16}$$

式中:N 为批次样本总数;$\boldsymbol{f}_i$ 和 $\boldsymbol{b}_i$ 分别为第 i 个样本的前景和背景特征向量;$\boldsymbol{f}_i \cdot \boldsymbol{b}_i$ 为 $\boldsymbol{f}_i$ 和 $\boldsymbol{b}_i$ 的内积;$|\cdot|$为特征向量的模长;θ 为预设的阈值,用于控制正交性的优化。L_{or}损失函数是单调递增的,当 $\boldsymbol{f}_i \cdot \boldsymbol{b}_i = 0$ 时,前景特征向量与背景特征向量完全正交。理论上,两者应该相互独立没有相似性,此时损失函数最小。然而,在实际场景中,完全正交可能会忽略背景中的有用信息。因此,通过设置阈值 θ,在减少噪声干扰的同时挖掘有助于行人识别的背景信息,优化特征质量,提升模型准确率。

1.4 损失函数

本文采用加权正则化三元组损失 L_{wrt}[17]、交叉熵标签平滑损失 L_{sid}[18] 和正交损失 L_{o} 联合训练。首先加权正则化三元组损失公式如下:

$$L_{\mathrm{wrt}} = \frac{1}{N}\sum_{i=1}^{N}\log\left(\sum_{ij} w_{ij,p} d_{ij,p} - \sum_{ij} w_{ik,n} d_{ik,n}\right) \tag{17}$$

$$w_{ij,p} = \frac{\exp(d_{ij,p})}{\sum\limits_{d_{ij,p} \in p_i} \exp(d_{ij,p})} \tag{18}$$

$$w_{ik,n} = \frac{\exp(d_{ik,n})}{\sum\limits_{d_{ik,n} \in n_i} \exp(d_{ik,n})} \tag{19}$$

式中:(i,j,k)为样本 x_i 的 1 个三元组;对于 x_i,p 和 n 分别为对应的正负样本集;$d_{ij,p}$与 $d_{ik,n}$为正负样本对的距离;d_{ij}为两样本之间的欧氏距离;$w_{ij,p}$,$w_{ik,n}$为正负样本对的权值。模型在训练过程中平衡不同难易程度的样本对,有效避免特定困难样本的过拟合问题。

在交叉熵损失函数中引入标签平滑,加入平滑因子 ε,避免过度自信导致的模型泛化性低,具体公式如下:

$$\varphi = \begin{cases} 1 - \varepsilon + \dfrac{\varepsilon}{C-1}, & i = y \\ \dfrac{\varepsilon}{C-1}, & i \neq y \end{cases} \tag{20}$$

$$L_{\mathrm{sid}} = -\frac{1}{N}\left(\sum_{i=1}^{N}\varphi \log P(y_i|\boldsymbol{\gamma}_i) + \sum_{i=1}^{N}\varphi \log P(y_i|\boldsymbol{\tau}_i)\right) \tag{21}$$

式中:y_i 为真实值;ε 为平滑参数(取 0.1);N 为批次训练样本的总数;C 为训练的样本数。该损失通过优化全局和前景特征的交叉熵损失,增强行人身份判别能力。

为提高模型性能,3 种损失函数联合优化,总训练损失为

$$L_{\mathrm{tot}} = \lambda_{\mathrm{sid}} L_{\mathrm{sid}} + \lambda_{\mathrm{wrt}} L_{\mathrm{wrt}} + \lambda_{\mathrm{a}} L_{\mathrm{o}} \tag{22}$$

式中:λ_{sid}、λ_{wrt}和 λ_{a} 分别为上述 3 种损失函数对应的权重因子,用于平衡损失之间的影响。

2 实验结果与分析

2.1 数据集与评价指标

为验证本研究所提出方法的有效性,本文在行人重识别领域的 3 个通用标准数据集上进行训练:Market-1501[19]、DukeMTMC-reID[20] 和 MSMT17[21],具体数据集信息如表 1 所示。

表 1 ReID 数据集信息表

Tab. 1 ReID dataset information sheet

数据集	年份	行人数	图像数	相机数	标注
Market1501	2015	1 501	32 668	6	DPM
DukeMTMC-reID	2017	1 812	36 441	6	人工
MSMT17	2018	4 101	126 441	15	Faster-RCNN

为评估模型性能,本文使用 ReID 领域的标准评价指标:Rank-1(R_1)和 mAP(α_{mAP})。R_1 表示查询图像检索结果中首位为同一行人的概率,而 α_{mAP}通过计算所有测试图像平均精度的平均值来衡量检索结果中正确图像的排名情况。

2.2　实验设置

本文实验环境配置信息如表 2 所示。

表 2　实验环境配置

Tab. 2　Experimental environment configuration

名称	配置信息
操作系统	Window11
开发语言	Python 3.9.0
框架	Pytorch 2.1.0 + cuda 12.1
CPU	Intel(R)Core(TM) i9-13900HX
GPU	GeForce RTX 4070Ti(12G)

实验开发框架主要基于 PyTorch 完成,行人掩码利用预训练的 DeepLab V3 + 模型生成,且将图像尺寸统一调整为 384 × 128,随机背景擦除阈值 μ 初始化为 0.1。采用 ResNet-50 网络,并将其前 2 个残差块(Conv1 和 Conv2)中的 3 × 3 卷积替换为 CMDC 结构。其次,在模型的最后去掉全连接层,采用多分支伪对抗背景抑制模块监督学习,该模块的阈值 θ 依据经验设置为 0.1 后经训练不断调整优化。模型总训练轮次为 120,初始学习率设置为0.000 35,并采用周期性学习率衰减策略,即每经过 40 个轮次,学习率减少到原来的 1/10,直至模型收敛。

2.3　与现有方法比较

将本文模型与主流方法在 Market1501、DukeMTMC-reID 和 MSMT17 数据集上比较,结果如表 3 所示。由表可见,在 Market1501 数据集上,本文方法的 mAP 和 Rank-1 分别达到 90.5% 和 96.8%,优于其他模型。在 MSMT17 数据集上,mAP 明显高于其他 ReID 经典方法,达到 70.2%。模型效果较好的原因是:在复杂环境下,现有注意力方法过于关注与行人身份无关的背景或辨别性较低的特征,导致特征表达冗余度较高,而本文模型不仅有效抑制了无关背景干扰,同时又充分挖掘了行人上下文之间有辅助作用的背景信息。其次,传统特征提取方法基于 CNN 模型,感受野较小,在不同尺度上的全局上下文信息获取能力有限,导致模型对行人外观变化的适应性较差。而 CMDC 模块不仅扩大模型感受野,提高模型对不同尺度的鲁棒性,还减少浅层细粒度特征的丢失。PAB 则是有效抑制背景噪声的干扰,增强模型对前景特征的关注。同时,增强对行人次显著性判别特征的提取能力,从而提高模型的判别与泛化能力。

表 3　与其他方法的比较结果(%)

Tab. 3　Comparison of results with different methods

Model	Market1501		DukeMTMC-reID		MSMT17	
	R_1	α_{mAP}	R_1	α_{mAP}	R_1	α_{mAP}
MGCAM (CVPR)[7]	83.79	74.33	—	—	—	—
MultiScale (ICCV)[22]	88.9	73.1	79.2	60.6	—	—
PCB + RPP(ECCV)[23]	93.8	81.6	83.3	69.2	—	—
MSCN(PRL)[24]	94.2	86.5	88.10	77.9	—	—
OSNet(ICCV)[25]	94.8	84.9	88.6	73.5	78.7	52.9
MMGA (CVPR)[8]	95.0	87.2	89.5	78.1	—	—
TransReID(ICCV)[26]	95.2	89.5	90.7	82.6	86.2	69.4
FPR(ICCV)[27]	95.42	86.58	88.64	78.42	—	—
Pyramid (CVPR)[28]	95.7	88.2	89.0	79.0	—	—
UniHCP(CVPR)[29]	—	90.3	—	—	—	67.3
ALDER(TIP)[30]	95.6	88.9	89.9	78.9	82.5	59.1
DFLN (TCSVT)[4]	95.9	89.8	91.3	81.8	83.6	64.5
CAFE-Net (本文方法)	96.8	90.5	2.6	83.2	79.3	70.2

2.4　消融实验

目前,ReID 方法中较好的基线模型为 ReID-Strong-Baseline[31](Baseline-S),本文在此基线的基础之上设计了消融实验,以验证所提出的 RBEA、CMDC 和 PAB 模块的有效性。首先,对 Baseline-S 进行简化移除随机擦除和中心损失。其次,设计 REA 与 RBEA 对比消融实验,以评估 RBEA 对模型性能的影响,实验结果如表 4 所示,表中如 + RBEA 表示 Baseline-S + RBEA,省略了 Baseline-S,其他模块同理。

表 4 消融实验对比结果(%)

Tab. 4 Ablation experiment comparison results

数据集	Market-1501		DukeMTMC-reID		MSMT17	
	R_1	α_{mAP}	R_1	α_{mAP}	R_1	α_{mAP}
Baseline-S	90.4	78.1	81.2	67.5	62.8	45.6
+ REA	92.8	82.6	83.8	72.6	64.2	47.9
+ RBEA	94.2	84.7	85.4	76.8	69.4	54.3
+ CMDC	94.5	84.9	85.9	76.5	72.7	57.5
+ PAB	93.4	85.2	86.6	75.3	67.2	58.6
+ RBEA + CMDC	95.9	88.4	88.1	80.5	73.1	65.7
+ RBEA + CMDC + PAB	96.8	90.5	92.6	83.2	79.3	70.2

在 Market-1501 数据集中,相较于加入 REA,单独加入 RBEA 的 Rank-1 和 mAP 分别提升了 1.4 和 2.1 个百分点,表明 RBEA 能有效去除无关背景噪声干扰,同时又不破坏行人上下文信息。CMDC 模块通过多分支扩张卷积结构扩大模型感受野(扩张率分别为 2、3、6,对应感受野从 5×5 至 15×15)实现了多尺度特征融合,不仅能够捕捉不同尺度的上下文特征,还在浅层网络中保持了较高的空间分辨率,为后续特征提取提供细节信息。PAB 则是在 CMDC 基础上进一步采用前景背景特征对抗式学习策略,挖掘行人潜在特征,有效抑制背景噪声干扰,增强行人特征的多样性和表达能力。综上,当模型结构为 Baseline-S + RBEA + CMDC + PAB 时,模型性能最佳,在 Market-1501 数据集上,Rank-1 达到 96.8%,mAP 达到 90.5%,充分验证了提出模块的有效性。为进一步观察模型性能,将 Baseline-S 和 CAFE-Net 在 Market-1501 上 Rank-1 随训练轮次的变化曲线进行展示,如图 6 所示。

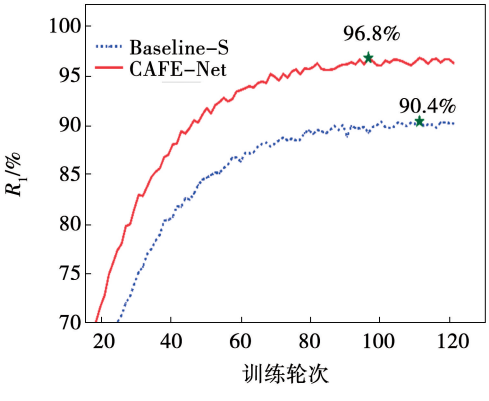


图 6 Baselie-S 与 CAFE-Net 的对比

Fig. 6 Comparison of Baseline-S and CAFE-Net

2.5 参数分析

为评估不同尺度特征对行人重识别任务的贡献,鉴于行人的特性,模型感受野不宜太大,能够覆盖整个行人即可。因此,本文选择了几组较小的空洞率进行实验,实验结果如表 5 所示。

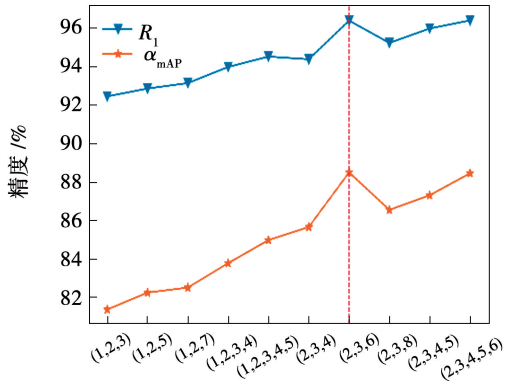
表 5 不同空洞率组合的实验结果(%)

Tab. 5 Experimental results of different void fraction combinations

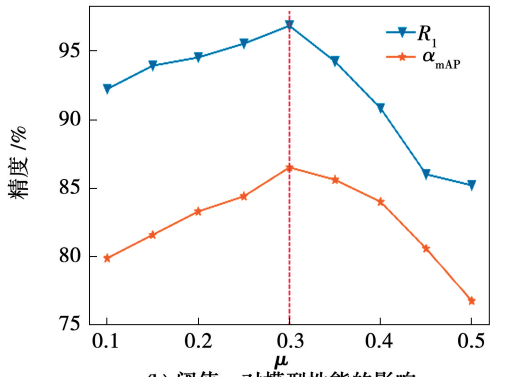
R = (r_1,r_2,r_3,r_4)	Market1501	
	R_1	α_{mAP}
(1,2,3)	92.43	81.35
(1,2,5)	92.48	82.23
(1,2,7)	92.74	82.68
(1,2,3,4)	93.96	83.73
(1,2,3,4,5)	94.51	84.96
(2,3,4)	94.36	85.64
(2,3,6)	96.37	88.48
(2,3,8)	95.21	86.53
(2,3,4,5)	95.96	87.29
(2,3,4,5,6)	96.38	88.43

结果显示,空洞率组合为(2,3,6)、(2,3,4,5)和(2,3,4,5,6)的模型在 mAP 和 Rank-1 指标上表现优异,验证了多分支结构能够在各分支相同层级上捕捉不同范围的上下文信息,增强了多尺度上下文特征的提取,使模型在浅层空间中更全面地获取行人全局信息。综合考虑性能提升与计算效率,本文采用了 3 分支的(2,3,6)组合,既提高了模型性能,又保持了合理的计算成本。图 7(a)为不同空洞率组合下模型性能的精度变化曲线。

参数 μ 对模型性能的影响如图 7(b)所示。当 μ 值设为 0.3 时,模型性能最佳,因为此时模型既擦除了对行人干扰的无关背景,同时又很好地保留了对行人识别有益的上下文信息。其次,当 μ 取值较小时,抑制背景力度较强,几乎屏蔽所有背景信息,使得对行人识别有帮助的上下文信息也被移除,模型性能下降,类似于引言提到的完全去掉背景。相反,当 μ 取值较大时,背景抑制效果不充分,噪声依然较多,对行人识别产生负面影响。所以,选择合适的随机擦除参数对模型性能有重要的影响。



(a) 空洞率组合对模型性能的影响



(b) 阈值 μ 对模型性能的影响

图 7　空洞率组合和阈值 μ 对模型性能的影响

Fig. 7　The influence of void fraction combinations and threshold μ on model performance

2.6　可视化分析

为方便观察背景与行人轮廓的相对距离关系,本文对部分样本进行了可视化展示,如图 8 所示。其中第 1、2、3 行分别为原始图像、行人轮廓及距离变换算法结果的热力图(颜色越深距离行人轮廓越近)。其次,为方便观察,将数据预处理后的随机背景擦除结果进行可视化,如图 9 所示,随机背景擦除部分用暗黄色表示。在实际训练中,擦除背景使用平均背景像素值填充。

为更直观地展示模型性能,本文采用 Grad-CAM[32]进行可视化,如图 10 所示。第 1 行为原图,第 2、3 行分别为 Baseline-S 和 CAFE-Net 的热力图。颜色越亮表示模型的关注度越高,由图可见,传统 Baseline-S 主要关注行人的显著区域,而 CAFE-Net 则覆盖整个行人,关注到除显著性特征之外的其他多样性次显著性特征,激活整个行人区域,证明了本文方法的有效性。



图 8　行人轮廓与距离变换可视化图

Fig. 8　Visualization of pedestrian contour and distance transformation

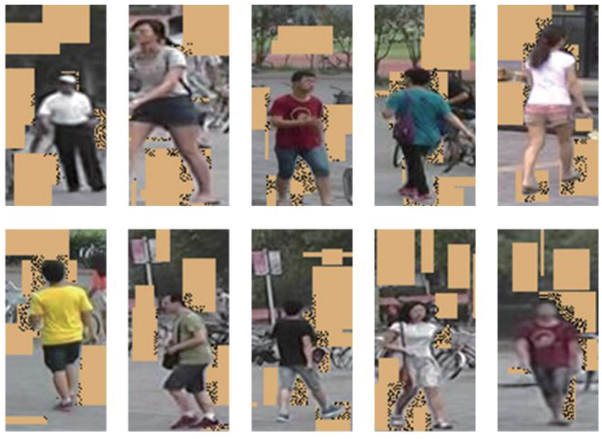


图 9　随机背景擦除结果图

Fig. 9　RBEA results

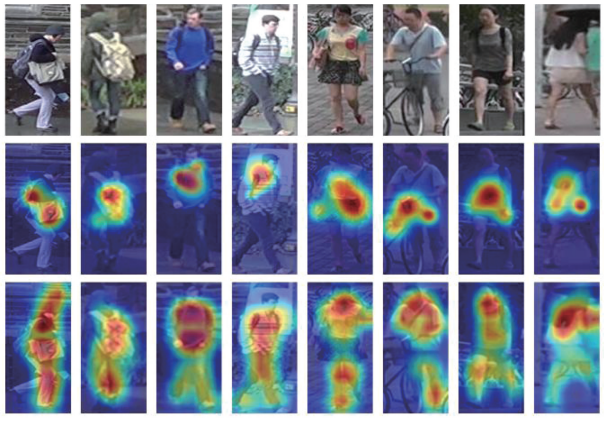


图 10　热力图可视化结果

Fig. 10　Heatmap visualization results

为方便观察,本文展示了检索识别对比图,如图 11 所示。其中第 1 部分为本文方法 CAFE-Net 的识别结果,第 2 部分为 Baseline-S 网络的结果。绿色圆球表示匹配正确,红色三角表示匹配错误。由图可见,当图像中有复杂背景时,本文方法依然能够识别并成功匹配,进一步验证了本文方法的有效性。



图 11 检测结果对比图

Fig. 11 Comparison figures of test results

3 结 语

本文针对行人重识别中的背景噪声干扰问题,提出了联合多分支空洞卷积与伪对抗学习策略的上下文感知特征增强的行人重识别方法。首先,通过自适应背景随机擦除数据增强策略,在去除无关背景噪声干扰的同时保留行人上下文结构信息。其次,设计了上下文感知的多分支空洞卷积结构,通过不同空洞率的卷积增大模型感受野,使模型在浅层空间中尽早感知完整行人,提取上下文多尺度和全局及局部的多粒度特征,增强特征的表达能力。最后,通过伪对抗学习的背景抑制策略,利用正交损失函数监督学习前景与背景特征间的正交独立性,提取除最显著特征之外的次显著判别性行人特征,抑制背景噪声干扰,提高模型的准确率。实验结果表明:该方法在多个标准数据集上显著提升了模型性能,并增强了模型对复杂多变背景的适应性和鲁棒性。

参 考 文 献:

[1] TIAN Maoqubg, YI Shuai, LI Hongsheng, et al. Eliminating Background-bias for Robust Person Re-identification[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 5794.

[2] SONG C, HUANG Y, OUYANG W, et al. Mask-Guided Contrastive Attention Model for Person Re-identification [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 1179.

[3] ZHENG Liang, HUANG Yujia, LU Huchuan, et al. Pose-invariant Embedding for Deep Person Re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(9): 4500.

[4] PANG Y, ZHANG H, ZHU L, et al. Feature Generation Based on Relation Learning and Image Partition for Occluded Person Re-identification [J]. Journal of Visual Communication and Image Representation, 2023, 91: 103772.

[5] YANG S, LIU W, YU Y, et al. Diverse Feature Learning Network with Attention Suppression and Part Level Background Suppression for Person Re-Identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(1): 283.

[6] GONG Y, WANG R, YANG J, et al. Person Re-identification with Global-Local Background_bias Net[J]. Journal of Visual Communication and Image Representation, 2021, 74: 102961.

[7] LIU Y, ZHOU W, LIU J, et al. An End-to-end Foreground-aware Network for Person Re-identification [J]. IEEE Transactions on Image Processing, 2021, 30: 2060.

[8] CAI H, WANG Z, CHENG J. Multi-Scale Body-Part Mask Guided Attention for Person Re-identification[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 1555.

[9] WANG F, WANG Y, LI D, et al. Enhancing CTR Prediction with Context-aware Feature Representation Learning[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022: 343.

[10] 魏明军, 刘铭, 刘亚志, 等. 前景与背景交互融合网络用于伪装目标检测[J]. 哈尔滨理工大学学报, 2025, 30(2): 53.
WEI Mingjun, LIU Ming, LIU Yazhi, et al. Foreground and Background Interaction Fusion Network for Camouflaged Object Detection[J]. Journal of Harbin University of Science and Technology, 2025, 30(2): 53.

[11] ZHANG Y, DENG W. Deep Learning Pavement Crack

Detection based on Atrous Convolution and Deep Supervision[C]//2022 14th Int ernational Conference on Measuring Technology and Mechatronics Automation (ICMTMA), 2022: 588.

[12] 李晓楠, 朱朦, 任洪娥, 等. 融合多分支与多粒度特征的东北虎重识别[J]. 哈尔滨理工大学学报, 2023, 28(6): 95.
LI Xiaonan, ZHU Meng, REN Honge, et al. Re-identification of Amur Tigers by Combining Multi-branch and Multi-granularity Features[J]. Journal of Harbin University of Science and Technology, 2023, 28(6): 95.

[13] YANG S, DENG W, WANG M, et al. Orthogonality Loss: Learning Discriminative Representations for Face Recognition[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(6): 2301.

[14] ZHONG Z, ZHENG L, KANG G, et al. Random Erasing Data Augmentation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 13001.

[15] WANG Y, WEI X, LIU F, et al. Deep Distance Transform for Tubular Structure Segmentation in CT Scans [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 3832.

[16] 万伟, 崔宏环, 尹文峰, 等. 基于正交设计的泥岩膨胀试验及其影响因素研究[J]. 森林工程, 2024, 40(5): 172.
WAN Wei, CUI Honghuan, YIN Wenfeng, et al. Mudstone Swelling Test and Its Influencing Factors Based on Orthogonal Design [J]. Forest Engineering, 2024, 40(5): 172.

[17] ZHU J, WU H, ZHAO Q, et al. Visible-infrared Person Re-identification Using High Utilization Mismatch Amending Triplet Loss [J]. Image and Vision Computing, 2023, 138: 104797.

[18] MURUGESAN B, LIU B, GALDRAN A, et al. Calibrating Segmentation Networks with Margin-based Label Smoothing [J]. Medical Image Analysis, 2023, 87: 102826.

[19] ZHENG L, SHEN L Y, TIAN L, et al. Scalable Person Re-identification: A Benchmark[C]//2015 IEEE International Conference on Computer Vision (ICCV), 2015: 1116.

[20] RISTANI E, SOLERA F, ZOU R, et al. Performance Measures and a Data Set for Multi-target, Multi-camera Tracking[C]//European Conference on Computer Vision, 2016: 17.

[21] WEI L, ZHANG S, GAO W, et al. Person Transfer GAN to Bridge Domain Gap for Person Re-identification[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 79.

[22] CHEN Y, ZHU X, GONG S. Person Re-identification by Deep Learning Multi-scale Representations [C]//2017 IEEE International Conference on Computer Vision (ICCV), 2017: 2590.

[23] SUN Y F, ZHENG L, YANG Y, et al. Beyond Part Models: Person Retrieval with Refined Part Pooling [C]//European Conference on Computer Vision (ECCV), 2018:480.

[24] LI S, LIU X, ZHAO Y, et al. Person Re-identification Based on Multi-scale Constraint Network [J]. Pattern Recognition Letters, 2020, 138: 403.

[25] ZHOU K, YANG Y, CAVALLARO A, et al. Omni-Scale Feature Learning for Person Re-Identification[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 3701.

[26] HE S, LUO H, WANG P, et al. TransReID: Transformer-based Object Re-Identification[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 14993.

[27] HE L, WANG Y, LIU W, et al. Foreground-aware Pyramid Reconstruction for Alignment-free Occluded Person Re-identification [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 8449.

[28] ZHENG F, DENG C, SUN X, et al. Pyramidal Person Re-IDentification via Multi-Loss Dynamic Training[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 8506.

[29] CI Y, WANG Y, CHEN M, et al. UniHCP: A Unified Model for Human-Centric Perceptions[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 17840.

[30] ZHANG Q, LAI J, FENG Z, et al. Seeing Like a Human: Asynchronous Learning with Dynamic Progressive Refinement for Person Re-Identification[J]. IEEE Transactions on Image Processing, 2022, 31: 352.

[31] LUO H, GU Y, LIAO X, et al. Bag of Tricks and a Strong Baseline for Deep Person Re-Identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 1487.

[32] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization[C]//2017 IEEE International Conference on Computer Vision (ICCV), 2017: 618.

(编辑:温泽宇)