DOI:10.16136/j.joel.2023.08.0548

损坏图像下基于风格归一化与全局注意力的行 人重识别

熊 炜^{1,2,3*},刘 粤¹,许婷婷¹,孙 鹏¹,赵 迪¹,李利荣^{1,2}

(1. 湖北工业大学 电气与电子工程学院, 湖北 武汉 430068; 2. 襄阳湖北工业大学产业研究院, 湖北 襄阳 441003; 3. 美国南卡罗来纳大学 计算机科学与工程系, 南卡罗来纳 哥伦比亚 29201)

摘要:针对当前网络难以应对各种损坏类型的行人图像与易丢失跨维信息的问题,提出了一种损坏图像下基于风格归一化与全局注意力的行人重识别(pedestrain re-identification, ReID)方法。该方法通过平滑极大单元的风格归一化与恢复(smooth maximum unit-style normalization and restitution, SM-SNR)模块中的实例规范化(instance normalization, IN)过滤掉域中的风格变化,同时平滑极大单元(smooth maximum unit, SMU)能使该模块更充分地从删除的信息中提取行人相关特征并将其恢复至网络中,缓解损坏图像带来的风格差异。此外,全局注意力机制(global attention mechanism, GAM)通过关注通道与空间之间的相互作用,以捕获3个维度上的显著行人特征,减少跨维信息的丢失,最终使本模型在面对行人损坏图像时的识别能力得到有效提高,且保留了在干净数据集上的竞争力。实验结果表明,本算法在损坏测试集上的各项指标与目前主流算法对比具有显著的优越性。其中,本模型与2021年的CIL模型使用CUHK03数据集比较的结果为:在Corrupted Eval上, R-1、mAP 和 mINP 分别提高了15.18%、15.75%与11.65%;在Clean Eval上, R-1与 mINP 仅降低了0.24%、0.75%,mAP 提升了0.25%。

关键词:行人重识别(ReID);行人损坏图像;平滑极大单元的风格归一化与恢复(SM-SNR);全局注意力机制(GAM)

中图分类号: TP183 文献标识码: A 文章编号: 1005-0086(2023)08-0833-09

Pedestrian re-identification based on style normalization and global attention in corrupted images

XIONG Wei^{1,2,3*}, LIU Yue¹, XU Tingting¹, SUN Peng¹, ZHAO Di¹, LI Lirong^{1,2}

(1, School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan, Hubei 430068, China; 2. Xiangyang Industrial Research Institute, Hubei University of Technology, Xiangyang, Hubei 441003, China; 3. Department of Computer Science and Engineering, University of South Carolina, Columbia, South Carolina 29201, USA)

Abstract: Aiming at the problem that the current network is difficult to deal with various corrupted pedestrian images and easily loses cross-dimensional information, a pedestrian re-identification (ReID) method based on style normalization and global attention is proposed for corrupted images. The method filters out style changes in the domain by smooth maximum unit-style normalization and restitution (SM-SNR) module in the instance normalization (IN), and at the same time smooth maximum unit (SMU) enables the module to more fully extract pedestrian-related features from the deleted information and restore them to the network, so as to alleviate the style difference caused by corrupted images. In addition, the global attention mechanism (GAM) captures the salient pedestrian features in three dimensions by focusing on the interaction between the channel and the space, reducing the loss of cross-dimensional in-

^{*} E-mail: xw@mail. hbut. edu. cn

formation. Finally, the recognition ability of the model in recognizing pedestrian corrupted images is effectively improved, and the competitiveness on clean datasets is retained. The experimental results show that the indicators of the algorithm on the corrupted test set has significant advantages compared with the current mainstream algorithms. Among these algorithms, the result of comparison with the 2021 CIL model using the CUHK 0 3 dataset is that : On Corrupted Eval , R- 1 , mAP and mINP increase by 15. 18 % , 15. 75 % and 11. 65 % respectively ; on Clean Eval , R- 1 and mINP only decrease by 0. 24 % , 0. 75 % , and mAP increased by 0. 25 %.

Key words: pedestrian re-identification (ReID); pedestrian corrupted image; smooth maximum unit-style normalization and restitution (SM-SNR) module; global attention mechanism (GAM)

0 引 言

行人重识别(pedestrian re-identification, ReID)旨在跨摄像机、时间与地点检索特定行人。因深度学习的飞速发展,计算机视觉领域取得了质的飞跃[1]。随之基于深度学习的 ReID 在视频监控、智能安防和无人超市等方面得以广泛应用[2],并成为了一个热门的研究方向。

研究人员在现实中采集的行人图像却易被噪 声、模糊、天气和数字等因素所损坏,且当前 ReID 模型在评估时,仅考虑了在干净(常规)数据集下 的性能,却忽略了在各式各样损坏情况下的行人 图像。因此,2019年,文献[3]建立了 ImageNet-C 和 ImageNet-P 两个新基准,用于评估并提高模型 在损坏图像下的性能,为 ReID 的研究带来了新思 路。为进一步提高损坏图像下的鲁棒性,2020年, 文献[4]提出了一种数据增强技术 AugMix,通过 混合多个增强图像,在 ImageNet-C 上的性能获得 了显著提升。为全面地研究损坏行人图像下模型 的识别能力,2021年,文献[5]首次使用干净数据 集通过 ImageNet-C 算法并额外引入了一种新的 损坏类型一"雨",生成了覆盖真实生活中常见的 20 种损坏类型的行人损坏数据集: Market1501-C、 CUHK03-C 和 MSMT17-C,以此用于评估模型的 鲁棒性。该文献通过对随机擦除、BNNeck 和 ID Loss 进行改进,建立了一个针对损坏图像的 ReID 新基线 CIL。虽然该模型性能相比之前的主流模 型在损坏图像下的性能得到了大幅提升,但是整 体指标仍很低,相比干净数据集下的指标存在相 当大的差距,如在 Market1501 数据集的损坏测试 集上, mINP 指标仅为 1.76%, 其与干净测试集下 的差值达到了 56.14%。同时,激活函数在 ReID 网络的训练中起着核心作用[6],而 ReLU(linear rectification function)非线性激活函数在负空间会 直接将值置为 0,造成导数为 0,因此出现参数不 更新的情况,使神经元在训练期间死亡。于是 2013年, Leaky ReLU函数被提出,通过给定一个

非常小的系数,使在负空间的导数接近于零而不等于零,以此让参数得到更新。2017年,谷歌提出了 Swish 非线性激活函数^[7],用于改善 ReLU 的上述问题,但 Leaky ReLU、Swish 在实际使用中不太稳定,导致在面对复杂的损坏类型图像时,效果不佳。

近年来,越来越多的注意力机制也被运用于 ReID以提升模型性能,如 WANG等[8]通过编码 一解码器残差注意力模块以增加对行人特征的提 取能力;WOO等[9]使用通道注意力与空间注意力 机制依次进行操作,且将 BAM 并行运行,可从通 道与空间两个维度提取显著行人的注意力特征。 但由于信息的减少与维度分离,使模块忽略了通 道、空间宽度和空间高度之间的交互作用,丢失了 跨维信息。

为解决上述问题,本文在 CIL 网络基础上,提出了一种损坏图像下基于风格归一化与全局注意力的 ReID 方法。本方法使用风格归一化和恢复(style normalization and restitution, SNR)模块^[10],过滤掉各种损坏图像中的干扰,并从过滤信息中捕获出行人相关特征恢复至网络中,同时在该模块中加入平滑极大单元(smooth maximum unit, SMU)代替 ReLU,解决梯度消失问题,使模块能够稳定地提取行人特征。最后,利用全局注意力机制(globe attention mechanism, GAM)^[11]减少信息弥散以放大"全局"跨维度的交互作用,使本模型在面对损坏但不丢失原始语义信息的行人图像时,能够更充分地捕捉到相对原始图像不变的行人表征。

1 损坏图像下的 ReID 模型

1.1 基于风格归一化与全局注意力的 ReID 网络框架

本文针对图像损坏情形提出了一种基于风格归一化与全局注意力的 ReID 网络框架,该网络结构简单,未引入额外的局部支路,如图 1 所示。

本网络框架以 ResNet50 作为骨干网,并在其网

络的 $4 \wedge \text{Layer}$ 层之后添加使用了平滑极大单元的风格归一化和恢复(smooth maximum unit-style normalization and restitution, SM-SNR) 模块,将特征通过实例规范化(instance normalization, IN)来弱化损坏图像与干净图像之间的风格差异,同时恢复在进行归一化操作时所丢失的行人相关信息。其次,在最后一个 SM-SNR 模块后加入 GAM,并以该 SM-SNR 模块的输出做为输入,通过减少全局信息的缺失和提高全局特征的交互,达到提高网络识别

性能的目的。之后删除 ResNet50 尾部的池化层和全连接层,将最后一个 block 的步长由 2 改为 1,通过增大 GAM 输出特征的尺寸,使得细粒度特征更加丰富;并将该特征通过广义均值池化(generalized mean pooling,GeM)得到大小为 $192\times2\,048\times1\times1$ 的全局特征 F。因为损坏图像与干净图像的统计信息(均值方差)有着较大差异,若用 BNNeck 后的特征做推断会降低模型的鲁棒性,因此本网络使用全局特征 F用于 triplet loss 进行度量学习。最后将 F 经过 BN-

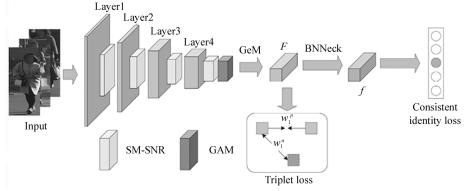


图 1 基于风格归一化与全局注意力的行人重识别网络框架

Fig. 1 A network framework for pedestrian re-identification based on style normalization and global attention

Neck 归一化处理得到大小为 192×2 048 的特征 f,用于一致性 ID Loss (consistent identity loss, CID loss)。

1.2 平滑极大单元的风格归一化和恢复(SM-SNR) 模块

由于现实世界中的损坏类型众多且不可预测,使得全监督 ReID 方法通常因域的差异而导致鲁棒性能较差。为了解决该问题,本文引入了 SNR 模块,并将该模块中平均池化后的 ReLU 层替换成 SMU,构成本文所提出的 SM-SNR 模块。在消除图像风格差异的同时确保行人 ID 特征不丢失,并解决传统激活函数的缺点,在模型面对风格多变的目标域时,提升其辨别能力。

1.2.1 风格归一化和恢复(SNR)模块

为使网络能够应对损坏图像复杂且多变的风

格,本文使用了 SNR 模块。通过 IN 消除当前域的 风格,以减少行人样本之间的差异性,从而增强网络 的鲁棒性^[12]。然而, IN 不可避免地会删除一些鉴别信息,导致网络的鉴别能力减弱^[13,14],因此本模块可从删除的 IN 信息中提取具有特定任务的鉴别性特征,并将其进行还原。

该模块以特征图 $F \in R^{C \times H \times W}$ 作为输入, $F^+ \in R^{C \times H \times W}$ 作为输出, $C \setminus H$ 和 W 分别表示为特征图的通道数、高度和宽度,如图 2 所示。首先在风格归一化阶段:通过 IN 来减少输入特征域的差异性。

$$\widetilde{F} = IN(F) = \lambda \left(\frac{F - \theta(F)}{\omega(F)} \right) + \alpha,$$
 (1)

式中, $\theta(\cdot)$ 和 $\omega(\cdot)$ 表示每个通道和样本在空间维度上独立计算的平均值与标准偏差, $\lambda,\alpha \in R^c$ 是从数据中学习到的参数:接着是风格恢复阶段,虽然

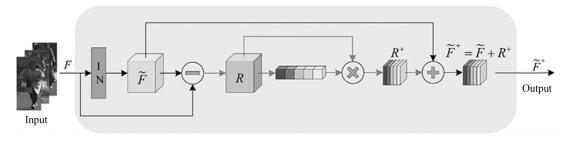


图 2 风格归一化和恢复模块结构图

Fig. 2 The structure diagram of the style normalization and restitution module

IN 减少了风格差异,提高了模型的鲁棒性,但是数学运算具有确定性与任务无关性,它不可避免地会丢失一些具有辨别性的行人信息。于是通过从残差特征R中提取与ID相关的特征,并将其恢复至网络中:

$$R = F - \widetilde{F} \,\, . \tag{2}$$

该残差特征 R 表示原始输入特征 F 和风格归一化特征 F 之间的差异,通过学习通道注意力向量 $a = [a_1, a_2, \cdots, a_C] \subset R^C$:

$$R^{+}$$
 (:,:, k) = $a_{k}R$ (:,:, k), (3)
式中, R (:,:, k) ∈ $R^{H\times W}$ 为特征图 R 的第 k 个通道, $k = 1, 2, \cdots, C$ 。通道注意力向量 a 能够自适应地提取 ID 相关特征以达到还原于网络的目的,其操作类似于 $SE^{[15]}$ 通道注意力,表达式如下:

$$a = \varphi(R) = \gamma(T_1 \varepsilon(T_2 pool(R)))_{\circ}$$
 (4

其由一个全局平均池化 pool 和两个其后分别连接 SMU ε (•) 和 sigmoid 激活函数 γ (•) 的 FC 全连接层 $T_1 \in R^{(\frac{c}{r}) \times c}$ 、 $T_2 \in R^{c \times (\frac{c}{r})}$ 构成。为了减少参数量,本文将降维比值 r 设置为 16。最后将 a 捕捉到的鉴别性特征 R^+ 添加至风格归一化后特征 F 后,得到还原特征:

$$\widetilde{F}^+ = R^+ + \widetilde{F} \ . \tag{5}$$

本文通过 SNR 模块充分利用 IN 来减少目标域中损坏图像风格与源域中干净数据集的风格差异,并进一步从原始信息的残差和归一化信息中恢复所丢失并具有鉴别性的特征,从而大大提升本网络在损坏图像下识别候选行人的能力。

1.2.2 平滑极大单元(SMU)

激活函数在保障深度神经网络训练中的有效性 发挥着极其重要的作用,其中 ReLU 因结构简单而 成为神经网络中最常见的选择。但是其存在一个明 显的缺点,即 ReLU 很"脆弱",会导致网络多达 50% 的神经元在训练过程中"坏死",使神经元无法再被 任何数据激活。为让 SNR 模块能够更加充分地捕 捉残差特征中具有辨别力的信息,本文将该模块中 全局平均池化后的 ReLU 非线性激活函数使用 SMU 进行替换。

该平滑激活函数通过 Maximum 函数的平滑近似来实现,其 Maximum 函数表达式如下^[6]:

$$\max(x_1, x_2) = \frac{(x_1 + x_2) + |x_1 - x_2|}{2}.$$
 (6)

使用 $x \operatorname{erf}(\mu x)$ 作为平滑函数,以达到对 Maximum 函数近似的目的。其中 $\operatorname{erf}(x)$ 是高斯误差函

数,定义如下:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^{2}} dt.$$
 (7)

 $x \operatorname{erf}(\mu x)$ 将式(6)中的 |x| 进行替换,从而得到 Maximum 函数的平滑近似公式[6]:

$$f(x_1, x_2; \mu) = \frac{(x_1 + x_2) + (x_1 - x_2) \operatorname{erf}(\mu(x_1 - x_2))}{2}.$$
 (8)

当 $x_1 = x$, $x_2 = \eta x$ 时,代人式(8)中,得到 SMU 的表达式^[6]:

$$f(x, \eta x; \mu) = \frac{(1+\eta)x + (1-\eta)x \operatorname{erf}(\mu(1-\eta)x)}{2}.$$
 (9)

其对x的求导式如下:

$$\frac{\mathrm{d}}{\mathrm{d}x}f(x,\eta x;\mu) = \frac{1}{2} \Big[(1+\mu) + (1-\mu) \Big]$$

$$\mathrm{erf}(\mu(1-\eta)x) + \frac{2}{\pi}\mu(1-\eta)^2 x \mathrm{e}^{-\mu(1-\eta)x^2} \Big],$$
(10)

最后通过反向传播技术更新激活函数的参数 $\eta \ \mu$, 且 $\eta = \mu$ 的梯度可从如下两个公式中计算得到:

$$\frac{\sqrt[3]{3}}{\sqrt[3]{\eta}} = \frac{x - x \operatorname{erf}(\mu(1 - \eta)x) - \mu(1 - \eta)x^{2} \operatorname{e}^{-(\mu(1 - \eta)x)^{2}}}{2},$$
(11)

$$\frac{\partial f}{\partial \mu} = \frac{(1-\eta)^2 x^2 e^{-(\mu(1-\eta)x)^2}}{2},$$
 (12)

式中,η、μ是超参数或可训练的参数。

当 $\eta = 0.25$, $\mu = 1.0$ 时,式(10)的函数图和 ReLU 函数图如图 3 所示。因 SMU 在输入 $x \le 0$ 时 值不为 0,避免了训练时参数不更新,解决了梯度消失的问题。因此,将 SMU激活函数作用于 SNR模

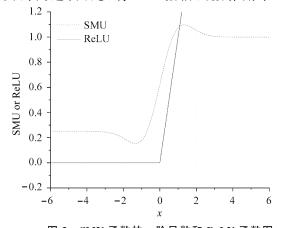


图 3 SMU 函数的一阶导数和 ReLU 函数图

Fig. 3 The first order derivatives graph of SMU function and the ReLU function graph

块中时,可更好地进行非线性映射,提升整个模块的 表达能力。

1.3 全局注意力机制(GAM)

GAM 与以往注意力不同之处在于:其可以通过专注于通道和空间方面的信息来提升"全局"跨维度的交互作用。该注意力采用序贯的通道-空间注意力机制,如图 4 所示,设第 4 个 SM-SNR 模块后的输出特征为 $F_1 \in R^{\text{CXHXW}}$,以此做为 GAM 的输入, F_2 为处于中间状态时的特征, F_3 为最终输出特征,定义如下:

$$F_2 = M_C(F_1) \otimes F_1, \qquad (13)$$

$$F_3 = M_S(F_2) \otimes F_2, \tag{14}$$

式中, M_c 是通道注意力图, M_s 是空间注意力图, \otimes 则表示按元素进行乘法操作。在通道注意力中,使

用三维排列在 3 个维度上保留信息;再通过一个多层感知器 (multilayer perception, MLP)来放大跨维通道-空间的依赖性,而 MLP是一种与 BAM 类似的编码-解码结构,其降维比值为r;将 F_1 与 $M_C(F_1)$ 按元素进行乘法操作后得到了 F_2 。其次,为捕获空间信息,空间注意力以 F_2 为输入,使用了两个 7×7 Conv进行空间信息融合;同时,因最大池化减少了维度信息的运用,则通过删除最大池化以此保留更多的行人相关特征;最后再将 F_2 与 $M_S(F_2)$ 按元素进行乘法操作后得到了具有多维度信息的特征 F_3 。

本文将 GAM 运用于 ResNet50 第 4 个 Layer 层之后,放大了最后一个 SNR 模块输出特征跨维度的交互作用,使网络能够在3个维度上捕捉行人显著

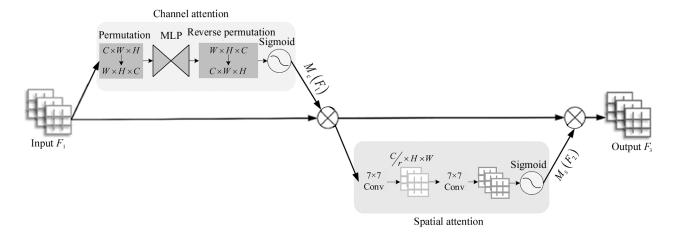


图 4 全局注意力机制结构图

Fig. 4 Structure diagram of global attention mechanism

特征,提高模型的辨别力和鲁棒性。

1.4 损失函数

因当前 ID Loss^[16]仅能计算每张图像的单个增广样本的损失,则本文使用了对 ID Loss 改进后的一致性 CID loss。通过利用原始样本 x_o 与增广样本的后验分布之间的 JS(Jenson-Shannon)散度,使模型对各式各样的数据(损坏图像)响应更加平滑,如下式所示^[4]:

$$L_{\text{cid}} (p_{o}; p_{a1}; p_{a2}) = \frac{1}{3} \text{KL} [p_{o} \parallel M] + \text{KL} [p_{a1} \parallel M] + \text{KL} [p_{a2} \parallel M], \qquad (15)$$

式中, $M = \frac{1}{3}(p_0 + p_{a1} + p_{a2})$, $p_o = p(y|x_o)$, p_{a1} = $p(y|x_{a1})$, $p_{a2} = p(y|x_2)$ 。而 x_o 为带有标签 y的一个原始输入图像, p_o 表示用 softmax 函数将 x_o 识别为类别 y 的预测概率, p_{a1} 与 p_{a2} 同理。KL 表示 KL 散度函数,式(15)通过 3 个不同 KL 散度之和取 均值,来表达 x_o 与增广样本 x_{a1} 、 x_{a2} 的后验分布之间的 JS 散度。

本文同时使用三元组损失(triplet loss)^[17]作为度量学习损失函数,首先任意选取一个图像 A、正样本 P 和负样本 N,设定正负样本相似度的阈值为 β ,则 $d_{A,P}$ 和 $d_{A,N}$ 分别表示样本 A 与其正、负样本的相似度。三元组损失的表达式如下:

$$L_{t} = (d_{A,P} - d_{A,N} + \beta)_{+}$$
, (16)
式中, (•)₊ 表示取绝对值。

为了更好地发挥损坏图像数据增强的作用并且加快模型的收敛速度,本文将 CID loss 和 triplet loss 进行联合训练,其总损失表达式如下:

$$L_{\text{total}} = L_{\text{cid}} + L_{t} \,\, \circ \tag{17}$$

2 实验结果分析

2.1 损坏数据集与评价指标

本文选用 Market1501 数据集、CUHK03 数据集和 MSMT17 大规模数据集对网络进行训练,分别包含 1501、1467 和4101个行人,共计 32668、14097 和126441 张行人图像。因国外行人隐私问题,文中未使用 DukeMTMC-ReID 数据集。同时本文选用了文献[4]提出的损坏数据集 Market1501-C,CUHK-03-C和 MSMT17-C对模型进行评估,例如,Market1501-C是由算法对 Market1501 测试集人为进行破坏生成的行人损坏图像。上述损坏图像类型分为噪声、模糊、天气和数字4个大类,详细信息如表1所示。每种损坏有5个不同程度的级别,从而组成了100种不同程度、不同类型的行人损坏图像。同时,为了更加充分地评估模型的鲁棒性,从干净测试集中的每个图像中随机选择一种损坏类型和一种程度

的损坏来创建以下 3 种损坏测试集: Corrupted query and gallery (corrupted eval), query 和 gallery 都是损坏类型的图像库; Corrupted query, 损坏查询样本; Corrupted gallery, 损坏图像库。

为评估本模型性能,共采用3种ReLD评价指标。如CMC-k(cumulative matching characteristics),又称 rank-k; mAP (mean average precision); mINP (mean inverse negative penalty)。本文通过 mINP 来衡量模型在最难正确匹配项上的检索能力,表达式如下:

$$mINP = \frac{1}{n} \sum_{i=1}^{n} (1 - NP_{i}) = \frac{1}{n} \sum_{i=1}^{n} \frac{R_{i}^{\text{hard}} - |Z_{i}|}{R_{i}^{\text{hard}}} = \frac{1}{n} \sum_{i=1}^{n} \frac{|Z_{i}|}{R_{i}^{\text{hard}}},$$
(18)

式中, R_i^{hard} 为最难匹配项的 rank 值, $|Z_i|$ 是查询第i 个图片正确匹配样本的总数值。

表 1 损坏类型

Tab. 1 Corruption class

Corruption class	Corruption type
Noise	Gaussian, Shot, Impulse, Speckle
Blur	Defocus, Frosted Glass, Motion, Zoom, Gaussian
Weather	Snow, Frost, Fog, Brightness, Spatter, Rain
Digital	Contrast, Elastic, Pixelate, JPEG Compression, Saturate

2.2 实验设置

实验采用 Pytorch=1.8.0 的深度学习框架,操作系统为 Ubuntu20.04。硬件配置为: GeForce GTX Titan GPU, Intel(R) Core(TM) i7-10700F CPU,32 GB 显存,16 GB 内存。使用了两种新颖的局部数据增强方法:自增补混合算法(self patch mixing)和软随机擦除(soft random erasing),以帮助模型挖掘更多样性的细粒度信息,提升模型面对损坏图像时的鲁棒性。设置超参数 $\eta=0.25$,并将 μ 视为可训练参数,并对 SMU 初始化为 1.0;本文对 Market1501(M-1501)、CUHK03 数据集进行 10 次评估,对 MSMT17 大规模数据集进行 3 次评估,最终结果取均值。在训练与测试时,batch size 分别设置为 64 和 128,训练 150 个 epoch。

2.3 实验分析

为验证 SM-SNR 模块中的 IN 在网络中所发挥的关键作用,本文分别使用 BN 层与 IBN 层替换 SM-SNR 模块中的 IN 层,在 M-1501 和 CUHK03

数据集上进行了实验与评估,实验结果如表 2 所示。 从实验数据分析可知,在两个数据集上,虽然使用 IBN 层的网络在干净测试集上的性能整体上有微小 的优势;但是在 3 个损坏测试集上,3 种网络的性能: IN 最优,IBN 次之,BN 垫底,且两两之间相差较大。 证明了 IN 相较于 BN 与 IBN 能够更好地消除域的 风格差异,从而使 SM-SNR 模块过滤出与行人无关 的干扰,学习到损坏图像中行人的显著特征。

此外,为证明本文在 SM-SNR 模块中使用 SMU 激活函数的合理性,该实验分别使用 ReLU、Leak-ReLU(L-ReLU)和 SMU 3 种激活函数作为 SNR 模块中平均池化后的激活函数,并在 M-1501 和 CU-HK03 数据集上进行了实验,结果如表 3 所示。实验发现,使用 SMU 代替原始的 ReLU 激活函数后,在干净数据集上各指标至少提升一个百分点,在各损坏测试集上也取得了最佳的整体性能。验证了SMU通过减少神经元在训练过程中"坏死"的现象,使 SNR 模块更充分地还原了所丢失的行人特征。

表 2 BN、IN 和 IBN 3 种归一化层对网络影响的实验结果

Tab. 2 The experimental results of the influence of BN, IN and IBN on the network

Dataset	Method	Clean Eval			Corrupted Eval			Corrupted query			Corrupted gallery		
		R-1	mAP	mINP	R-1	mAP	mINP	R-1	mAP	mINP	R-1	mAP	mINP
M-1501	BN	93.12	83.72	57.01	56.53	29.54	1.94	65.35	55.06	31.34	86.19	50.07	3.54
	IN	93.45	84.62	58.80	74. 32	50.14	11.64	79.63	66.61	38.80	89.73	65.12	16.49
	IBN	93.56	84. 83	58.23	67.79	41.02	5.23	73.81	62.67	36.50	88.45	59.01	9.05
	BN	66.96	64.85	52.83	23.58	15.66	4.56	35.22	36.01	25.49	46.13	32.41	9.38
CUHK03	IN	67.05	65.41	53.12	38. 14	32.08	15.90	46. 43	45.97	35.32	54. 14	43.89	23. 23
	IBN	67.35	65.23	53. 24	30.25	25.32	9.11	41.89	42.01	32.23	52.18	38.95	16.08

表 3 SMU 激活函数的验证实验

Tab. 3 Verification experiment of SMU activation function

Dataset	Method	Clean Eval			Corrupted Eval			Corrupted query			Corrupted gallery		
		R-1	mAP	mINP	R-1	mAP	mINP	R-1	mAP	mINP	R-1	mAP	mINP
M-1501	ReLU	92.15	83.56	57.55	73.21	49.23	11.71	78.68	65.88	37.62	88.01	64.03	15.11
	L-ReLU	92.64	83.51	57.97	74.18	50.02	11.55	78.84	65.36	37.89	88.14	64.37	15.21
	SMU	93.45	84.62	58.80	74.32	50.14	11.64	79.63	66.61	38.80	89.73	65.12	16.49
	ReLU	65.90	64.12	52.02	37.28	31.17	15.01	45.76	45.31	34.69	54. 55	43.04	22.56
CHUK03	L-ReLU	65.21	64.80	51.98	37.09	31.21	15.24	45.81	44.93	34.55	54.10	43.18	22.53
	SMU	67.05	65.41	53.12	38. 14	32.08	15.90	46. 43	45. 97	35.32	54.14	43.89	23. 23

为验证本文所使用的 3 个模块对网络性能的影响,此消融实验以 ResNet50 和 GeM 池化作为 Baseline,然后使用 SNR、SMU 和 GAM 3 个模块在 CU-HK03 数据集上进行,其实验结果与消融实验中mAP 数据的可视化见表 4 和图 5。分析数据可得出,在 Baseline 基础上使用 SNR 模块后,模型在干净

数据集上的性能有所降低,但是在损坏数据上的识别能力得到了大幅度提升;再分别使用 SMU 激活函数和 GAM 后,模型在干净数据集上的性能随之提升,同时各损坏测试集上的指标也存在一定的提高。由此验证了本算法加入的 3 个模块可让网络更有效地识别不同域中不变的行人显著特征。

表 4 不同模块的消融实验

Tab. 4 Ablation experiment of different modules

Method	(Clean Eva	ıl	Corrupted Eval			Coı	rrupted q	uery	Corrupted gallery			
	R-1	mAP	mINP	R-1	mAP	mINP	R-1	mAP	mINP	R-1	mAP	mINP	
Baseline	66.48	63.92	52.41	23.93	16.05	4.39	34.04	33.85	25.28	46.93	31.23	9.02	
+SNR	64.95	63.12	50.54	36.41	30.12	14.16	44.82	43.69	33.84	52.64	41.78	22.13	
+smu	65.91	64.18	51.67	36.89	30.53	14.24	44.71	43.84	33.97	52.72	42.26	22.18	
+GAM	67.05	65.41	53.12	38.14	32.08	15.90	46.43	45.97	35.32	54.14	43.89	23.23	

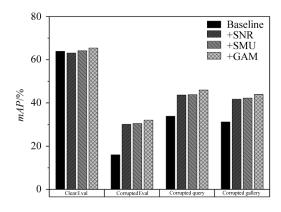


图 5 消融实验数据可视化

Fig. 5 Data visualization of ablation experiment

2.4 对比实验

为证明所提出的算法在面对行人图像损坏情形时的优越性,将与BoT^[18]、AGW^[19]、SBS^[20]和CIL^[4]4个SOTA模型在所述的3个数据集上进行对比实验,数据如表5所示。通过实验数据对比可得到,本模型在Clean Eval上性能表现并不出色,但与以上模型差距不大。在损坏测试集上,本模型取得的各项指标相比BoT、AGW和SBS3个常规的模型有相当大的突破;对比2021年提出的行人损坏图像新基准CIL模型,本模型的性能也取得了很大的提升。例如,在CUHK03数据集的Corrupted Eval上,本模型的R-1、mAP和mINP相较于CIL模型的分别

提高了 15. 18%、15. 75%与 11. 65%; 在 Clean Eval 上,本模型的R-1和mINP相较于CIL模型仅降低了0. 24%、0. 75%, mAP提升了0. 25%。由此证明

本模型在面对损坏行人图像时,对比其他模型具有 明显的优势,同时在干净数据集上也具有一定的竞 争力。

表 5 本算法与主流模型的性能对比

Tab. 5 Performance comparison between this algorithm and mainstream models

Dataset	SOTA	Clean Eval			Corrupted Eval			Corrupted query			Corrupted gallery		
		R-1	mAP	mINP	R-1	mAP	mINP	R-1	mAP	mINP	R-1	mAP	mINP
M-1501	ВоТ	93.38	85.06	59.30	27.05	8.42	0.20	31.92	26.89	14.56	76.78	26.82	0.39
	AGW	94.00	86.51	64.03	31.90	12.13	0.35	35.25	31.75	19.44	80.45	33.38	0.67
	SBS	95.90	88.33	60.03	34.13	11.54	0.29	42.06	35.33	18.47	83.11	32.65	0.53
	CIL	93.38	84.04	57.90	55.57	28.03	1.76	62.29	52.53	29.99	85.52	48.95	3.45
	Ours	93.45	84.62	58.80	74. 32	50.14	11.64	79.63	66.61	38. 80	89.73	65. 12	16.49
	AGW	64.64	62.25	49.97	5.90	3.45	0.46	16.26	17.20	12.69	33.43	19.40	2.89
CUHK03	CIL	67. 29	65.16	53.87	22.96	16.33	4.25	34.03	34.62	26.61	46.81	31.81	9.07
	Ours	67.05	65.41	53.12	38. 14	32.08	15.90	46. 43	45.97	35. 32	54. 14	43.89	23. 23
	ВоТ	73.53	48.34	9.91	20.20	5.28	0.07	25.92	15.78	2.75	59.06	16.10	0.09
	AGW	75.21	51.84	12.38	22.77	6.53	0.08	28.06	18.42	3.82	61.45	18.08	0.15
MSMT17	SBS	82.02	56.62	10.26	28.77	7.89	0.05	36.68	22.71	3.23	70.65	21.16	0.12
	CIL	76.10	52.40	12.45	39.79	15.33	0.32	45.51	29.08	5.84	68.31	27.99	0.50
	Ours	76.80	52.10	11.97	51.16	23.36	1, 11	55.85	35. 02	6. 54	70.50	34. 85	1.75

2.5 损坏图像与结果可视化

本文对损坏严重性为一级程度的雪、斑点、散焦和像素4种模拟的损坏类别进行了可视化处理,同时对GAM层输出的特征进行了热力图、映射图可视化,如图6所示,其中图(a)从左至右分别为行人原图、雪损坏类别行人图、损坏时行人特征热力图和损坏时行人

特征映射图。从热力图和映射图可以观察到,对于雪、斑点和像素 3 类损坏,本模型提取的特征图响应区域仍较为精准,进一步验证了本模型出色的鲁棒性能。但在面对散焦类别的损坏时,效果不太理想,是因为该损坏图像丢失了部分原始的语义信息,导致模型无法充分地学习到相对原始行人图像不变的表征。

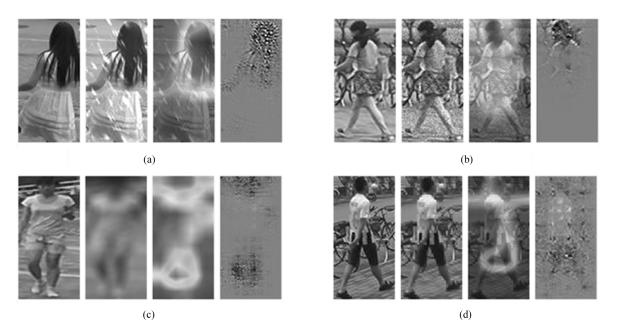


图 6 损坏图像结果可视化:(a) 雪;(b) 斑点;(c)散焦;(d) 像素

Fig. 6 Result visualization of corrupted images: (a) Snow; (b) Speckle; (b) Defocus; (b) Pixelate

3 结 论

本文提出了一种基于风格归一化与全局注意力 的 ReID 方法,解决了网络面对复杂且风格多样的损 坏图像时,难以提取显著行人特征以及缺失"全局" 跨维度信息的问题。本网络框架以 ResNet50 为主 干网,使用 SM-SNR 模块,减少风格变化带来的干 扰,并充分地恢复 IN 所丢失的行人特征。同时利用 GAM 放大通道和空间宽度、高度之间跨维度的交互 作用,以在"全局"中捕获更多具有鉴别性的行人特 征。各项实验分析证明,本文所提出的网络对比当 前先进方法,在损坏图像下,能更好地学习到相对原 始数据不变的行人表征,使得R-1、mAP 和mINP得到明显增长,但依旧存在很大的提升空间,且在干 净图像下的识别性能略有降低。所以如何进一步提 高模型面对现实世界中各式损坏行人图像的识别能 力,并保证在干净数据集下的性能仍具优越性,乃是 未来探索的重点。

参考文献:

- [1] XIONG W,YUE L,ZHOU L,et al. Pedestrian re-identification algorithm based on multi-level feature fusion [J].

 Journal of Optoelectronics Laser, 2021, 32 (8): 872-878.
 - 熊炜,乐玲,周蕾,等.基于多层级特征融合的行人重识别算法[J].光电子·激光,2021,32(8):872-878.
- [2] LUO H, JIANG W, FAN X, et al. A survey on deep learning based person re-identification [J]. Acta Automatica Sinica, 2019, 45(11); 2032-2049.

 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.
- [3] HENDRYCKS D.DIETTERICH T. Benchmarking neural network robustness to common corruptions and perturbations[EB/OL]. (2019-03-28)[2022-07-28]. https://arxiv.org/abs/1903.12261.
- [4] HENDRYCKS D,MU N,CUBUK E D,et al. Augmix: a simple data processing method to improve robustness and uncertainty [EB/OL]. (2020-02-17) [2022-07-28]. https://arxiv.org/abs/1912.02781v2.
- [5] CHEN M, WANG Z, ZHENG F. Benchmarks for corruption invariant person re-identification [EB/OL]. (2021-11-01) [2022-07-28]. https://arxiv.org/abs/2111.00880v1.
- [6] BISWAS K,KUMAR S,BANERJEE S,et al. SMU;Smooth activation function for deep networks using smoothing maximum technique [EB/OL]. (2022-04-11) [2022-07-28]. https://arxiv.org/abs/2111.04682v2.
- [7] PRAJIT R, BARRET Z, QUOC V. Searching for activation functions[EB/OL]. (2017-10-16)[2022-07-28]. https://arxiv.org/abs/1710.05941v2.
- [8] WANG F, JIANG M, QIAN C, et al. Residual attention network for image classification [C]//IEEE Conference on Computer Vision and Pattern Recognition, July 21-26,

- 2018, Honolulu, HI, USA. New York: IEEE, 2017: 3156- 3164
- [9] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]// European Conference on Computer Vision, September 8-14, 2018, Munich, Germany. Berlin: Springer, 2018: 3-19.
- [10] JIN X.LAN C.ZENG W.et al. Style normalization and restitution for generalizable person re-identification [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition, September 13-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 3143-3152.
- [11] LIU Y, SHAO Z, HOFFMANN N. Global attention mechanism: retain information to enhance channel-spatial interactions [EB/OL]. (2021-12-10) [2022-07-28]. https://arxiv.org/abs/2112.05561.
- [12] ZHOU K,YANG Y,CAVALLARO A,et al. Omni-scale feature learning for person re-identification[C]//IEEE/CVF International Conference on Computer Vision, June 15-20, 2019,Long Beach,CA,USA. New York: IEEE, 2019: 3702-3712.
- [13] HUANG X, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization [C]//IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 1501-1510
- [14] PAN X,LUO P,SHI J,et al. Two at once: enhang learning and generalization capacities via IBN-Net[C]//European Conference on Computer Vision, September 8-14, 2018, Munich, Germany. Berlin: Springer, 2018: 464-479.
- [15] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//IEEE Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 7132-7141.
- [16] ZHENG Z, ZHENG L, YANG Y. A discriminatively learned CNN embedding for person re-identification [EB/OL]. (2017-02-03) [2022-07-28]. https://arxiv.org/abs/1611.05666.
- [17] HERMANS A, BEYERL, LEIBE B. In defense of the triplet loss for person re-identification [EB/OL]. (2017-03-22) [2022-07-28]. https://arxiv.org/abs/1703.07737.
- [18] LUO H, GU Y, LIAO X, et al. Bag of tricks and a strong baseline for deep person re-identification [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, June 15-20, 2019, Long Beach, CA, USA. New York: IEEE, 2019:8514-8522.
- [19] YE M, SHEN J, LIN G, et al. Deep learning for person reidentification: a survey and outlook [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44 (6): 2872-2893.
- [20] HE L,LIAO X,LIU W, et al. FastReID: a pytorch toolbox for real-world person re-identification [EB/OL]. (2020-06-04)[2022-07-28].https://arxiv.org/abs/2006.02631. 2006.02631.

作者简介:

熊 炜 (1976一),男,博士,副教授,硕士生导师,主要从事数字图像 处理和计算机视觉方面的研究.