

基于灾难性遗忘及组合叠加擦除的跨模态行人重识别预训练方法

孙 锐^{1,2}, 谢瑞瑞^{1,2}, 张 磊^{1,2}, 张旭东^{1,2}, 高 隽^{1,2}

(1. 合肥工业大学计算机与信息学院, 安徽合肥 230601; 2. 工业安全与应急技术安徽省重点实验室, 安徽合肥 230009)

摘 要: 面向构建 24 小时全时段视频监控系统的需要, 基于可见光与近红外的跨模态行人重识别受到工业界与学术界的广泛关注. 然而, 目前大部分跨模态行人重识别任务都试图利用在 ImageNet 上预训练的模型来提前学习模态内共性特征, 但 ImageNet 与跨模态行人数据模态差异较大, 且预训练过程中将颜色信息作为判别特征之一, 导致预训练中学习到的共性特征并不适用于无色彩红外图像的信息表示. 本文提出了一种基于灾难性遗忘及组合叠加擦除的自监督跨模态行人重识别预训练方法, 首先利用提出的灾难性遗忘评分来对预训练数据进行筛选, 旨在减小预训练数据与后续任务数据存在的域间差距, 进一步减少模型训练时间. 其次, 针对传统跨模态识别中的关键区分性特征提取, 本文设计了一种强通道数据增强策略, 通过对 R、G、B 三通道的通道级擦除与组合, 生成了颜色迥异的多类型样本, 有利于促使模型关注于纹理信息而非颜色信息. 最后基于本文提出的跨模态数据筛选指标以及通道增强策略, 构建了跨模态任务的自监督学习框架. 实验结果表明, 本文提出的预训练方法所训练的 ResNet50 网络在迁移到众多跨模态行人重识别方法时优于目前主流自监督预训练方法, 其中在经典方法 AGW 的基础上 Rank1 与 mAP 分别提高了 8.02% 与 5.81%.

关键词: 自监督; 行人重识别; 跨模态; 预训练; 灾难性遗忘; 组合叠加擦除

基金项目: 国家自然科学基金面上项目 (No.61876057); 安徽省自然科学基金 (No.2208085MF158); 安徽省重点研发计划-科技强警专项项目 (No.202004d07020012)

中图分类号: TP391.4

文献标识码: A

文章编号: 0372-2112(2023)10-2925-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20221190

Cross-Modal Pedestrian Re-identification Pre-training Method Based on Catastrophic Forgetting and Combination Superimposed Erasure

SUN Rui^{1,2}, XIE Rui-rui^{1,2}, ZHANG Lei^{1,2}, ZHANG Xu-dong^{1,2}, GAO Jun^{1,2}

(1. School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, Anhui 230601, China;

2. Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei, Anhui 230009, China)

Abstract: To meet the need of building a 24-hour full-time video surveillance system, cross-modal pedestrian recognition based on visible light and near-infrared is widely concerned by industry and academia. However, most of the current cross-modal pedestrian recognition tasks attempt to use pre-trained models on ImageNet to learn modal commonalities in advance, but there are large modal differences between ImageNet and cross-modal pedestrian data, in the pre-training process, the color information is taken as one of the distinguishing features, which leads to the common features learned in the pre-training is not suitable for the information representation of the colorless infrared image. This paper proposes a self-supervised cross-modal pedestrian recognition pre-training method based on catastrophic forgetting and combined superposition erasure. Firstly, the pre-training data are filtered by using the proposed catastrophic forgetting score, the aim is to reduce the domain gap between the pre-training data and the follow-up task data, and further reduce the training time of the model. Secondly, aiming at the key distinguishing feature extraction in traditional cross-modal identification, this paper designs a strong channel data enhancement strategy by erasing and combining the R, G and B channels at the channel level, multi-type samples with different colors are generated, which makes the model focus on texture information instead of color information. Finally, a self-supervised learning framework for cross-modal tasks is constructed based on the cross-modal data

filtering index and channel enhancement strategy. The experimental results show that the ResNet50 network trained by the proposed pre-training method is superior to the current self-supervised pre-training methods when migrating to a large number of cross-modal pedestrian recognition methods, on the basis of AGW, Rank1 and mAP were increased by 8.02% and 5.81% respectively.

Key words: self-supervised; pedestrian re-identification; cross-modality; pre-training; catastrophic forgetting; strong channel combination

Foundation Item(s): National Natural Science Foundation of China (No.61876057); National Natural Science Foundation of Anhui (No.2208085MF158); Key Research and Development Plan of Anhui Province (No.202004d07020012)

1 引言

重识别(Re-Identification, ReID)是一种跨场景跨摄像机的图像检索任务,该任务旨在多个无重叠视觉区域的摄像头下,检索在其中一个摄像头中出现的目标行人是否在其他摄像头中出现.可见光相机难以识别夜间及光照不足情况下的行人.红外摄像机在夜间或较差的照明条件下仍能捕捉行人的外观特征,可见光与红外相结合的双模式摄像机广泛使用为构建24小时全天时行人重识别系统提供了条件,同时也衍生出一种新型的跨模态再识别任务^[1],即可见光-近红外跨模态行人重识别(RGB-NIR cross-modality person re-identification).

从现有机器学习研究来看^[2,3],通过在大规模数据集上进行预训练来获取基础特征后,再迁移到目标任务的主干网络(如ResNet^[4],ViT^[5])上微调已经成为了计算机视觉任务成功的范式.基于深度学习的行人重识别模型的性能很大程度上依赖于标记训练图像的质量与数量^[6],而跨模态行人重识别任务的数据收集和标注是十分困难的.面对大规模标注数据集的缺乏,在ImageNet^[7]上进行预训练再将预训练后的主干网络迁移到目标数据集^[8-11]上微调成为了跨模态行人重识别任务成功的关键.

本文探索了从单模态数据中挖掘与跨模态数据相近的数据信息的预训练方法,并从数据增强的角度缓解了RGB与IR的模态差异.首先,本文提出了跨模态灾难性遗忘评分(Cross-modality Catastrophic Forgetting Score, CCFS),用来评估预训练数据与目标任务数据的相似程度,从而获得与下游视觉任务更为接近的预训练数据.其次,本文设计了一种新的数据增强策略-强通道组合与通道叠加擦除,旨在从纯RGB图像中学习颜色无关的特征.最后,本文结合灾难性遗忘指数与通道组合叠加擦除策略,构建了针对跨模态行人重识别任务的自监督预训练框架,通过在AlignGAN, AGW, DDAG, CAJ四个方法上广泛的实验验证了该预训练框架的有效性.

2 相关工作

自监督学习^[12](Self-Supervised Learning, SSL)常用

于在大量的无标签数据中学习对下游任务有帮助的判别特征. He^[13,14]与Chen^[15]等人提出的对比学习方法(Momentum Contrast, MoCo)及简易对比框架(Simple framework for Contrastive Learning of visual Representations, SimCLR),将样本的一对数据增强视为正样本对,将其他的所有样本视为负样本对来进行预训练.与传统对比学习方法相反,Ge^[16]等人提出了一种无需负样本的新自监督学习范式(Bootstrap Your Own Latent, BYOL),在不使用任何负样本的情况下, BYOL通过预测的方式完成训练也表现出良好的性能. MoCo, SimCLR, BYOL等一系列基于对比学习的自监督方法,利用大量的无监督数据进行预训练,达到了有监督预训练的性能.几乎在所有的下游视觉任务中,自监督的预训练方法都能获得很好的性能增益.

为了缩小预训练数据与行人重识别任务数据集的域间差距, Deng^[17]等人构建了一个大规模的无标签行人数据集“LUPerson”,并第一次尝试利用自监督学习方法来预训练模型并迁移到行人重识别任务当中.在此基础上Luo^[18]等人提出了灾难性遗忘评分(Catastrophic Forgetting Score, CFS)用来评估预训练数据和微调数据之间的差距,进一步减小了数据集之间的域间差距. LUPerson以及CFS证明,通过减小预训练数据和目标任务数据之间的域差距可以有效的提高预训练的速度与识别率.此外, Zhu^[19]与Wan^[20]等人提出了特定于行人重识别的预训练方法,与传统对比学习方式不同,其通过恢复打乱的图像块的顺序来学习图像中的潜在特征,提高了局部特征的可辨别性.在本文中,我们从构建对比学习框架与缓解数据差异性的角度出发,来增强预训练模型的鲁棒性.

3 基于自监督学习的跨模态行人重识别预训练方法

本文提出基于灾难性遗忘与组合叠加擦除的跨模态行人重识别预训练方法.总体分为两个阶段,第一阶段我们使用提出的跨模态灾难性遗忘评分计算大型数据集LUPerson与SYSU-MM01近红外图像的特征相似度,以此对LUPerson进行筛选,获取其高质量子集LU-

Person*. 第二阶段利用提出的通道组合叠加擦除方法生成正负样本,构建自监督对比学习框架完成无标签的预训练. 算法总体流程图如图 1 所示,网络框架如图 2 所示.

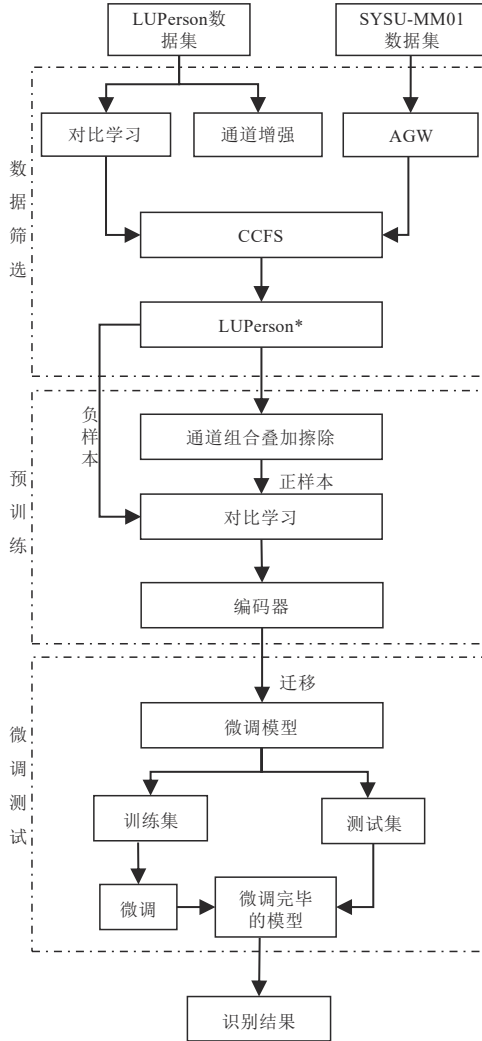


图 1 本文方法总体流程图

3.1 跨模态灾难性遗忘评分

灾难性遗忘^[21,22]是指在模型迁移过程中,先前学习到的共性特征发生遗忘的现象,为了减小预训练模型迁移过程中的灾难性遗忘程度,更好地评估 RGB 单模态数据集与 RGB-IR 双模态数据集之间的域差距,本文提出了跨模态灾难性遗忘评分(Cross-modality Catastrophic Forgetting Score, CCFS)作为单模态数据集与跨模态数据集域差距的评价指标,如式 1 所示:

$$c_{s,h}^i = \frac{\langle \theta_s(x_h^i), \theta_t(x_s^i) \rangle}{\|\theta_s(x_h^i)\| \|\theta_t(x_s^i)\|} \quad (1)$$

其中, θ_s 为预训练模型, θ_t 为目标域模型,本文中 θ_s 为在 LUPerson 上使用 MoCoV2 训练的 ResNet50 网络, θ_t 为在

SYSU-MM01^[1]上使用 AGW^[2]模型微调训练的 ResNet50 网络. x_s 为 LUPerson 中的行人图片, x_h 为 x_s 经过通道增强策略^[23]的数据增广图片, $c_{s,h}^i$ 越大代表数据之间的域差距越小,遗忘程度也就越低.

3.1.1 分析与对比

本节给出了关于跨模态灾难性遗忘指数 CCFS 的理论推导及其与灾难性遗忘指数 CFS 的对比分析以通过此说明 CCFS 作为预训练数据筛选指标的有效性.

目标任务数据集为 $D_t = (X_t, Y_t)$, 其中, $X_t = \{x_t^1, x_t^2, x_t^3, \dots, x_t^n\}$, Y_t 为 X_t 身份标签, 同理描述大规模无监督预训练数据为 $D_s = \{x_s^1, x_s^2, x_s^3, \dots, x_s^m\}$. 我们希望找到 $D_s' \in D_s$, 且 D_s' 中数据应尽量与 D_s 中数据接近, 即寻找 D_s 的高量子集.

分析: 为了计算预训练数据与目标任务数据之间的相似程度, 本文从数据的特征相似度出发, 希望缩小的是目标域模型分别处理预训练数据和目标数据时的特征映射差距, 当特征映射差距较小时则表明数据之间的域间差距越小, 即

$$\min \|\theta_t(x_s^i) - \theta_t(x_t^i)\| \quad (2)$$

根据三角不等式有

$$\begin{aligned} \|\theta_t(x_s^i) - \theta_t(x_t^i)\| &= \|\theta_t(x_s^i) - \theta_s(x_h^i) + \theta_s(x_h^i) - \theta_t(x_t^i)\| \\ &\leq d_{s,h,t} + \|\theta_s(x_h^i) - \theta_t(x_t^i)\| \end{aligned} \quad (3)$$

将 $d_{s,h,t} = \|\theta_t(x_s^i) - \theta_s(x_h^i)\|$ 作为筛选指标, $\|\theta_s(x_h^i) - \theta_t(x_t^i)\|$ 为误差项. $d_{s,h,t}$ 越小则代表数据之间的域间差距越小. 同时 $d_{s,h,t}$ 可以与 CCFS 相互转化, 如式 (4) 所示, 在筛选过程中将数据按照 $c_{s,h}^i$ 大小从大到小排序, 最终取排名前 100 万张数据组成预训练数据.

$$\begin{aligned} \|\theta_t(x_s^i) - \theta_s(x_h^i)\|^2 &= \|\theta_t(x_s^i)\|^2 + \|\theta_s(x_h^i)\|^2 - 2\langle \theta_t(x_s^i), \theta_s(x_h^i) \rangle \\ &= 2 - 2c_{s,h}^i \end{aligned} \quad (4)$$

对比: CFS 与 CCFS 的出发点一致, 差异在于不等式放缩时选取的中间量不同. CCFS, CFS 放缩结果分别如式 (3) 和式 (5) 所示. CCFS 将经过通道增强的数据作为放缩中间量, 通过通道增强策略将原本 RGB 模态的数据转换成与 IR 模态相近的数据, 改变了筛选指标与误差项的数据模态构成.

$$\begin{aligned} \|\theta_t(x_s^i) - \theta_t(x_t^i)\| &= \|\theta_t(x_s^i) - \theta_s(x_s^i) + \theta_s(x_s^i) - \theta_t(x_t^i)\| \\ &\leq d_{s,t} + \|\theta_s(x_s^i) - \theta_t(x_t^i)\| \end{aligned} \quad (5)$$

CCFS 的误差项为 $h_{s,h,t} = \|\theta_s(x_h^i) - \theta_t(x_t^i)\|$, CFS 的误差项为 $h_{s,t} = \|\theta_t(x_t^i) - \theta_s(x_s^i)\|$. CFS 将 $d_{s,t} = \|\theta_t(x_s^i) - \theta_s(x_s^i)\|$ 作为筛选指标, 在面对多模态数据集时 (SYSU-

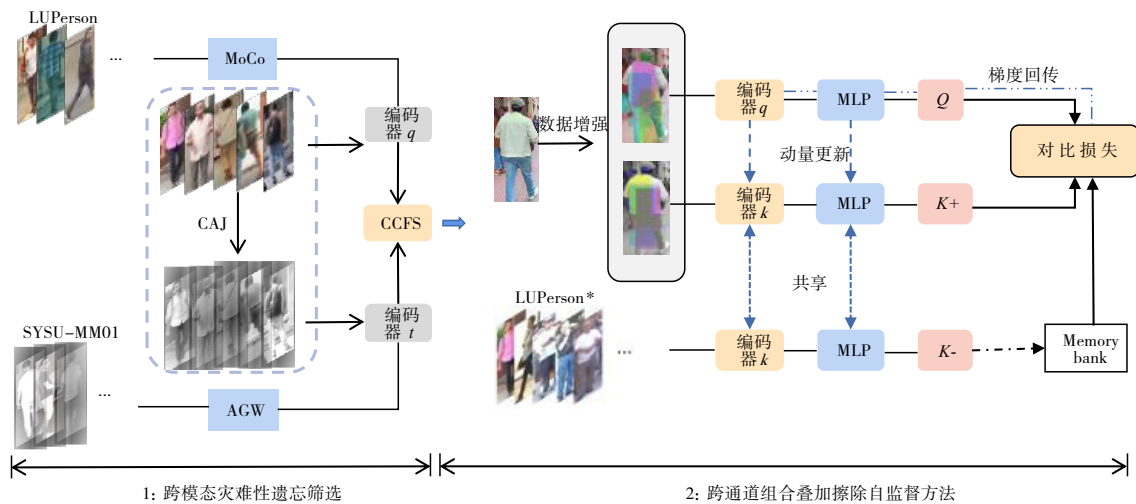


图2 本文方法网络框图

MM01, RegDB), 其误差项 $h_{s,t}$ 中的数据 x_s 与 x_t 由原本的不同模态数据 (RGB-IR) 变成了跨模态数据 (RGB-IR), 这就导致 $h_{s,t}$ 往往会偏大, 从而减弱了筛选指标 $d_{s,t}$ 的影响. 这种由跨模态数据所带来的误差导致 CFS 并不适用于度量 RGB 单模态数据集与 RGB-IR 双模态数据集^[24]之间的域间差距. 具体对比如图3所示.

从图3可以看出, 与CFS相比, CCFS的筛选指标由两个跨模态的数据经过编码器得到, 而误差项则是通过相似模态数据得到, 数据构成刚好与CFS相反, 从理论上来说, 模态相似数据得到的特征差距也越相似. 因此, 在筛选指标 $d_{s,t}$ 增大, 误差项 $h_{s,t}$ 缩小的条件下, CCFS在面对跨模态的数据集时可以更有效的筛选出与下游任务接近的数据.

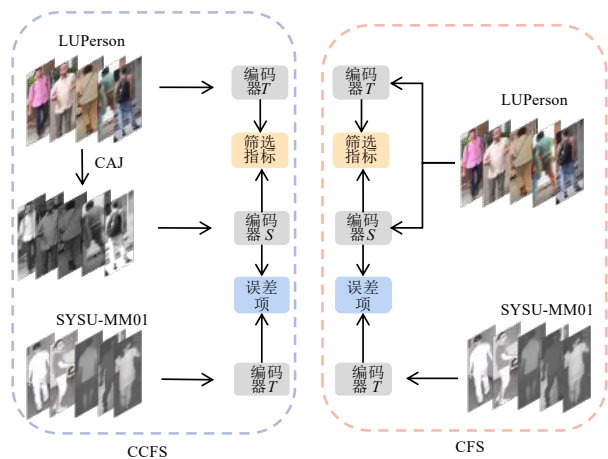


图3 CCFS与CFS对比图示

3.2 自监督对比学习框架

本文采用对比学习框架进行跨模态行人重识别的自监督预训练, 总体框架如图1右半部分所示. 给定一个输入图像 x_t , 对该图像执行两种不同的数据增强策略

(T_1, T_2), 生成两个增强图像 (x_1, x_2), 将 x_1 输入编码器 E_q 得到特征 q , 将 x_2 输入另一个编码器 E_k 得到特征 k^+ . 此外, 训练过程中当前Batch中的其他图片经过编码器 E_k 可以得到一系列特征 $\{k_j^-\}$. 得到特征之后对比学习就类似于一个字典查询任务, 给定字典中的一个查询信息 q 和字典内容 $\{k^+, k_1^-, k_2^-, \dots, k_j^-\}$, 训练过程希望 q 可以查询到其对应正样本 k^+ , 对比学习实质上鼓励 q 在收敛过程中不断接近正样本 k^+ , 远离负样本 $\{k_j^-\}$. 表述如下:

$$\text{score}(E_q(q), E_k(k^+)) \gg \text{score}(E_q(q), E_k(k^-)) \quad (6)$$

score 函数是一个度量两个特征之间相似性的指标. 该任务中正样本是来自于同一张图片的数据增强, 而负样本则是数据集的所有其他图片, 整个过程无须利用人工标签, 是一个自监督的学习过程. 在形式上, 对比损失^[25]函数定义为

$$L = -\log \frac{\exp(q \cdot k^+ / \tau)}{\exp(q \cdot k^+ / \tau) + \sum_{i=0}^{K-1} \exp(q \cdot k_i^- / \tau)} \quad (7)$$

对比损失可以看作是一个大型的分损失, 负样本的规模越大^[26]对训练过程是有益的, 因为丰富的样本可以有效的提高模型的泛化能力, 同时增加了难样本的数量, 帮助挖掘深层信息. 但真实训练过程中, 由于预训练数据量巨大, 这就导致负样本的数量同样巨大, 而在对比学习这个字典查询任务当中字典内容是由GPU所储存的, 受限与GPU内存大小, 字典的大小往往不能够设置得很大. 因此本文跟随 MoCo^[13]的设置, 构建一个队列的结构来保持一个大型的字典内容, 随着训练的进程不断更新字典. 同时, 更新字典的过程中, 编码器也会更新, 而更新前后编码器本身会产生差异, 从而导致正负样本被差异较大的编码器提取特征, 产生的特征中不仅包含正负样本本身的语义特征差

距,还包含了编码器自身不同导致的特征偏差,因此我们跟随 MoCo 构建了一个动量编码器,来保证正样本与负样本都来自于同样或类似的编码器. 具体来说,在梯度回传更新过程中编码器 E_q 率先得到更新,之后 E_k 根据 E_q 的参数进行动量更新, E_q 和 E_k 具有相同的网络结构,但是其参数略有不同.

$$E_k = mE_k + (1 - m)E_q \quad (8)$$

表 1 不同动量 m 时的 Rank 1 精度 单位: %

Momentum m	0.9	0.99	0.999	0.999 9
Rank-1	51.94	54.17	55.52	55.34

m 为动量参数,从式(8)可以看出, m 越接近 1,动量更新的程度越低, E_k 与 E_q 就越相似. 如表 1 所示,我们在 AGW 模型的全搜索模式下测试了 m 在 0.9~0.999 9 之间取值时的 Rank-1 识别率,在 m 取 0.999 时,模型的识别率最高,本文中 m 取 0.999.

3.3 数据增强策略

数据增强在自监督对比学习中极为重要,通过不同的数据增强方法对训练样本模拟不同的类内变化是自监督学习成功的关键^[27],例如:遮挡等问题可以用随机擦除^[28]模拟,姿势改变可以用水平翻转与裁剪模拟. 为了更好地模拟近红外数据的特点,增加模型对颜色信息的鲁棒性,我们需要一种新的针对颜色信息的数据增强方法.

3.3.1 强通道组合

在传统的单模态行人重识别任务当中,颜色信息^[29]以及细粒度信息^[11]往往是模型需要学习用来辨别身份的高级语义特征. 相反,在跨模态行人重识别任务当中,由于红外图像只有形状和结构等^[30]信息,这导致在跨模态行人重识别任务中颜色信息不能成为有效的身份鉴别特征. 另外,跨模态行人重识别领域公共数据集稀少,目前针对于行人识别任务的主流预训练数据集是 ImageNet 与 LUPerson,尽管 LUPerson 在一定程度上减轻了与行人识别任务数据的域差距,但 LUPerson 与 ImageNet 模态一致,仍然是 RGB 单一模态的数据集,面对跨模态的下游任务,预训练过程难免会学习与模态相关的特征,而这些特征^[30](如:颜色信息)在可见光-近红外识别任务中是无意义甚至会影响鉴别的. 因此本文提出了一种强通道组合增强策略(Strong Channel Combination, SCC),通过这种增强方式将原本单一的 RGB 图像转换成风格不同的多模态图像或多风格图像,使模型能够更多地关注于行人的细粒度结构信息.

图像的颜色信息由红(Red, R)、绿(Green, G)、蓝(Blue, B)三个颜色通道的数值变化及其相互之间的叠加组成,为构建同一图像的颜色多样化样本,本方法的主要思想是将图像 R、G、B 三个通道分离,再对三个通

道进行复制和排列组合,组合结果可以包含三个或者两个重复通道,根据排列组合公式生成结果共有 27 种,具体包括 R-R-R、R-R-G、G-R-B 等. 每个通道中的像素的分布都可能发生变化,同时可能伴随着某一颜色通道的消失,这种增强方式极大地扩大了图像的颜色复杂度.

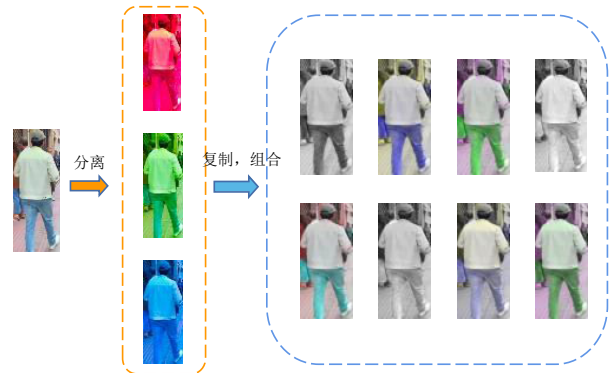


图 4 由强通道组合得到的颜色多样化样本

在训练过程中,同一张图像的颜色信息是多变的,但是其语义信息是不变的,促使模型在学习过程中学习颜色及模态不变特征. 另外,强通道交换增强作为一个轻量级数据增强方法,可以与其他基本增强方法无缝衔接,如:水平翻转、随机擦除、随机裁剪等. 通过与现存其他增强方法组合可以生成风格迥异的大量图像,有利于提高模型的泛化能力. 加强过程可被表述为

$$x_{h^i}^{v^i} = (x_{C^j}^i, x_{C^k}^i, x_{C^l}^i), \quad C^j = \text{Random}(R, G, B) \quad (9)$$

3.3.2 通道叠加擦除

随机擦除^[28]方法作为一种轻量级的数据增强方法,已经在各种视觉任务上得到了应用^[31]并取得了良好的实验结果,其主要思想是给定一个擦除概率 P ,根据概率判断是否在原图随机选择一个矩形区域,将该区域的像素替换为随机值. 经过数据增强的图片被不同程度的擦除,在行人重识别任务中可用作模拟行人遮挡现象,如 DcMOT^[32]采用随机像素块模拟复杂行人场景中的低分辨率遮挡情况,有效的降低了过拟合的风险并提高模型面对遮挡的鲁棒性. 此外 CAJ^[23]与其所提出通道增强相结合,设计了一种通道级随机擦除策略 CRE(Channel-level Random Erasing)来丰富训练样本的多样性. 上述方法的基本思想是在训练图像中随机选择一个矩形区域,并将所有三个通道的像素值替换为随机值,从而模拟不确定的遮挡情况. 尽管 DcMOT 与 CAJ 均采用了擦除策略丰富了样本的多样性,但其所实现的仍然是纹理层面的硬擦除效果. 为了进一步丰富训练样本的颜色多样性,提高模型泛化能力,本文结合通道组合策略提出了一种新的通道叠加擦除(Channel Overlay Erasure, COE)策略,其实现的是颜色

以及纹理层面的双层擦除效果。

具体来说,首先将原 RGB 图像 $x_{r,g,b}^{v,i}$ 分离成三个单通道色彩图像 x_r^i, x_g^i, x_b^i 之后分别对三个通道图像分别进行随机擦除,再将随机擦除的图片按照通道组合策略进行复制和组合,实验阶段选择从大规模数据集 LUPerson 中所有图像计算得到的 R、G、B 通道的平均值 $[0.352\ 5, 0.310\ 6, 0.314\ 0]$,作为每个通道的擦除值,如擦除的是 R 通道,则被擦除区域的像素统一为 0.352 5. 具体结果如图 5 所示。

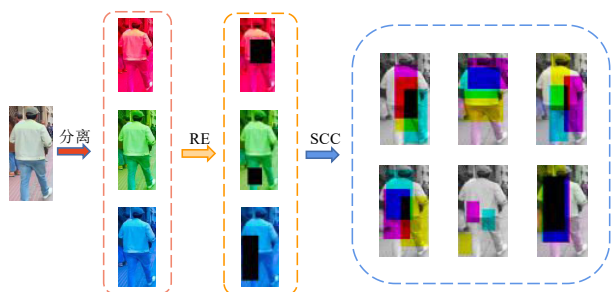


图 5 由通道叠加擦除得到的颜色多样化样本

RE 为随机擦除,SCC 为强通道组合,为了使结果更明显,擦除区域使用 0 像素代替。

通道擦除策略与通道交换策略相结合,生成了大量风格的多模态图像,处理过程中伴随着擦除区域的随机选取与叠加,对颜色层面进行了叠加擦除处理,同时当 R、G、B 通道擦除区域有重叠部分时还保留了纹理方面的硬擦除。作为一种轻量级的数据增广方法,通道擦除策略进一步深化了图像的颜色变换。

4 实验结果与分析

4.1 实验数据集

本文实验共包含预训练与迁移下游任务测试两个阶段。预训练阶段,选取从大规模无标签数据集 LUPerson^[17]中筛选的 100 万张图片作为预训练数据。测试阶段,为了验证本文预训练方法的有效性,我们在两个跨模态行人重识别主流数据集 RegDB^[24]和 SYSU-MM01^[1]上评估本文的预训练方法。

LUPerson 是从 Youtube 上依照“城市名+街景”格式搜索获取的 7 万多个视频,再通过 YOLO-V5^[33]来检测和裁剪这些视频获取的行人图片。该数据集共包括超过 20 万行人身份的 46 260 个场景的 4 180 243 张无标签行人数据。实验过程中利用 CCFS 筛选出与下游任务数据较为接近的前 100 万张行人图片作为预训练数据。

RegDB 是由双摄像机系统(近红外-可见光)所采集的小型数据集,一共包含一台可见光摄像机和一台热成像摄像机。该数据集一共包含 412 个身份的 8 024 张图片,平均每个身份有 10 张可见光图像和 10 张红外图像。遵循文献^[34]的评估协议,测试阶段选取的训练集与

原任务保持一致,一般来说随机选取 206 个身份用于训练,另外 206 个身份用于测试。测试阶段共包含两种检索模式,根据检索与被检索图像的不同,分为红外检索模式(Thermal to Visible)与可见光检索模式(Visible to Thermal)。最终结果为 10 次实验平均结果。

SYSU-MM01 作为目前跨模态行人重识别主流的数据集,由 4 个可见光摄像机和 2 个近红外摄像机拍摄采集而成。SYSU-MM01 一共包含 491 个行人身份,每个行人最少由两个不同的摄像机拍摄,一共有 287 628 张可见光 RGB 图像和 15 792 张近红外图像。与 RegDB 类似,测试阶段选取的训练集与原任务保持一致。具体来说,选取 395 个行人身份图像作为训练集,剩余 96 个行人身份图像作为测试集。随机选取 301 幅可见光图像作为检索图像,3 803 幅红外图像作为被检索图像。根据标准评估协议^[1]与拍摄环境不同,SYSU-MM01 包含 All-Search 和 Indoor-Search 两种检索模式。其中,Indoor-Search 模式由室内可见光摄像机 1,2 拍摄的图片作为被检索图像,红外摄像机 3,6 拍摄的图片作为检索图像。在 All-Search 模式下所有的可见光图片作为被检索图像,同理所有的近红外图片作为检索图像。为保证实验的准确性,在 Single-Shot 设置下进行实验,取十次实验平均结果。



图 6 LUPerson、SYSU-MM01、RegDB 数据集的示例图像

4.2 评估指标与实验参数设置

评估指标:为公平起见,实验部分的评价指标与原任务一致,具体包括累积匹配特性(Cumulative Matching Characteristics, CMC)、平均精度均值(mean Average Precision, mAP)。

参数设置:我们使用 Pytorch 训练框架,训练阶段所有图片大小调整至 256×128 。本文选择从 LUPerson 计算的像素平均值与方差 $[0.352\ 5, 0.310\ 6, 0.314\ 0]$, $[0.266\ 0, 0.252\ 2, 0.250\ 5]$ 来进行图像归一化。除了强通道交换增强与通道叠加擦除外,本文还使用了随机裁剪、水平翻转、灰度化等作为辅助增强方法。训练过程使用两个 RTX 3090 GPU 训练 300 个 epoch,初始学习率为 0.007 5,学习率使用余弦衰减^[35],优化器采用 Adam^[36]。迁移至下游任务时,我们将 ResNet50 的全连接层以及池化层删除,同时保持其他默认设置。

4.3 迁移到主流方法的结果

将经过预训练的 ResNet50 迁移到目前 4 个先进的跨模态行人重识别方法中进行测试,具体包括: AGW^[2]、CAJ^[23]、DDAG^[37]、AlignGAN^[38]. 对比预训练方法包括有监督和无监督两种,有监督方法为在 ImageNet 上进行分类任务训练. 无监督方法分别为

BYOL、SimCLR、MoCoV2 以及本文方法. 微调阶段使用所测试方法的默认设置来微调预训练模型. 其中在 ImageNet 上有监督预训练的主干网络是所有测试任务的默认方法. 为保持公平,所有无监督预训练方法使用的预训练数据均为通过 CCFS 筛选的 LUPerson 数据集.

表 2 各方法在 SYSU-MM01 上的识别结果

单位: %

方法	数据集与预训练方法	All-Search				Indoor-Search			
		Rank 1	Rank 10	Rank 20	mAP	Rank 1	Rank 10	Rank 20	mAP
AlignGAN	IN + Sup	42.40	85.00	93.70	40.70	45.90	87.60	94.40	54.30
	LUP* + BYOL	40.50	80.87	89.41	39.81	43.28	86.23	91.43	53.25
	LUP* + SimCLR	38.02	78.18	86.22	37.28	41.87	83.29	89.47	50.71
	LUP* + MoCoV2	42.13	85.83	93.37	41.81	45.16	88.05	94.19	54.89
	LUP* + 本文方法	43.28	86.54	94.49	43.25	47.70	89.42	95.06	56.63
AGW	IN + Sup	47.50	84.39	92.14	47.65	54.17	91.14	95.98	63.97
	LUP* + BYOL	51.14	88.47	95.24	50.14	55.68	91.16	95.36	65.14
	LUP* + SimCLR	48.27	85.36	92.14	48.28	54.69	90.54	94.32	63.84
	LUP* + MoCoV2	53.74	89.23	95.48	52.26	56.21	91.47	96.04	66.87
	LUP* + 本文方法	55.52	90.86	96.32	53.46	57.24	92.14	96.48	68.36
DDAG	IN + Sup	54.75	90.39	95.81	53.02	61.02	94.06	98.41	67.98
	LUP* + BYOL	49.71	85.20	89.00	48.28	53.43	89.11	92.75	60.67
	LUP* + SimCLR	45.19	80.15	85.11	44.34	49.42	84.60	90.11	58.13
	LUP* + MoCoV2	50.42	85.86	90.67	48.14	54.04	89.14	92.96	61.23
	LUP* + 本文方法	52.61	87.14	92.38	50.06	57.47	91.17	97.76	64.45
CAJ	IN + Sup	69.88	95.71	98.46	66.89	76.26	97.88	99.49	80.37
	LUP* + BYOL	65.56	91.32	93.98	61.58	72.78	92.66	93.97	74.82
	LUP* + SimCLR	63.17	90.24	93.78	61.37	71.25	91.39	92.48	74.36
	LUP* + MoCoV2	66.52	92.47	95.45	64.48	74.47	95.36	96.45	76.17
	LUP* + 本文方法	68.74	94.77	97.06	65.47	75.79	96.48	98.24	78.17

各方法在 SYSU-MM01 数据集上对比结果如表 2 所示. IN 代表 ImageNet 数据集, LUP* 代表 CCFS 筛选的 LUPerson 子集, Sup(SuperVised) 代表有监督训练. 从表 2 可以看出本文提出的针对于跨模态预训练方法明显优于其他主流自监督方法. 相比于无监督方法中表现最好的预训练框架 MoCoV2, 本文方法在目前最优识别方法 CAJ 的基础上在 Rank-1 和 mAP 这两个主要指标上分别提升了 2.22% 和 1.18%. 同时, 在部分方法中, 本文预训练方法甚至超过了经过有监督预训练模型的迁移识别率, 如在典型方法 AGW、AlignGAN 的基础上, 与有监督预训练(IN + Sup) 相比 Rank-1 与 mAP 分别提升了 8.02%, 5.81% 与 1.28%, 2.55%.

表 3 显示了在 RegDB 上的微调结果, 可以看出, 在 RegDB 上我们的预训练方法同样优于其他全部的无监督方法. 在 Visible to Thermal 模式下, 本文方法预训练的模型与 BYOL 相比在 AGW 的基础上将 Rank-1 与 mAP 分别提高了 3.18% 与 1.53%. 在 Thermal to Visible

模式下本文方法同样将 AGW 的 Rank-1 提高了 3.87%, mAP 提高了 2.55%.

本文对不同自监督方法的算法速度进行了测试, 测试结果如表 4 所示, 与其他自监督方法相比本文方法耗时略长, 但整体时间与其他算法仍处于同一量级.

上述实验表明, 与众多自监督方法相比, 本文所提出的基于灾难性遗忘及组合叠加擦除的跨模态行人重识别预训练方法在性能上都具有一定的优势. 在迁移到众多任务的实验结果也证明, 本文所训练的模型有效的提高了跨模态行人重识别任务的识别率, 同时通过多种模式的识别实验表明本文的模型具有一定的鲁棒性.

4.4 消融实验

我们首先使用经过各种筛选方式筛选的预训练数据进行了广泛的实验, 以评估本文提出的跨模态灾难性遗忘指数的有效性. 表 5 展示了使用不同方法筛选的 LUPerson 子集所预训练的模型迁移到 AGW 上后, 再

表3 各方法在RegDB上的识别结果

单位:%

方法	数据集与预训练方法	Visible to Thermal				Thermal to Visible			
		Rank 1	Rank 10	Rank 20	mAP	Rank 1	Rank 10	Rank 20	mAP
AlignGAN	IN + Sup	56.30	—	—	53.40	57.90	—	—	53.60
	LUP* + BYOL	55.84	86.68	91.71	52.13	56.09	85.37	92.34	52.19
	LUP* + SimCLR	53.31	83.37	88.23	48.02	53.42	84.40	88.11	48.80
	LUP* + MoCoV2	56.16	86.54	92.88	53.57	57.93	86.95	93.93	53.03
	LUP* + 本文方法	58.15	88.21	94.10	55.36	59.37	88.76	94.69	54.42
AGW	IN + Sup	70.05	86.21	91.55	66.37	70.49	87.12	91.84	65.90
	LUP* + BYOL	70.17	86.82	90.35	66.61	70.37	84.90	88.59	63.22
	LUP* + SimCLR	67.49	84.75	89.30	63.35	68.18	82.57	87.13	61.67
	LUP* + MoCoV2	70.42	86.42	91.14	65.14	70.73	85.47	89.14	63.36
	LUP* + 本文方法	73.35	87.09	91.21	68.14	74.24	88.14	92.48	65.77
DDAG	IN + Sup	69.34	86.19	91.49	63.46	68.06	85.15	90.31	61.80
	LUP* + BYOL	66.14	81.19	85.12	61.52	65.19	81.28	86.51	59.21
	LUP* + SimCLR	63.90	79.26	84.26	58.30	62.89	79.93	84.83	57.93
	LUP* + MoCoV2	66.46	82.24	86.41	61.14	66.53	82.01	87.24	60.67
	LUP* + 本文方法	67.61	84.14	88.45	62.47	67.17	84.08	89.14	62.87
CAJ	IN + Sup	85.03	94.49	97.54	79.14	84.75	95.33	97.51	77.82
	LUP* + BYOL	80.39	89.87	91.69	73.45	80.15	88.96	91.90	71.98
	LUP* + SimCLR	78.68	86.20	88.25	71.42	77.34	85.81	87.74	69.50
	LUP* + MoCoV2	81.56	90.41	92.75	74.29	80.78	89.04	91.45	73.46
	LUP* + 本文方法	83.64	92.64	95.03	76.47	82.14	91.97	95.12	75.96

表4 算法速度测试结果

单位:h

预训练方法	10-epoch
SimCLR	8.1
MoCoV2	9.0
本文方法	9.3

在SYSU-MM01数据集上微调 and 测试的识别准确率。

表中Random为随机采样,Sampled为以4为采样频率平均采样,CFS为灾难性遗忘评分,CCFS为本文提出的跨模态灾难性遗忘评分.四种方式的筛选数据与ImageNet-1k保持一致,筛选得出的数据均为100万张。

表5 采用不同方法对LUPerson进行筛选的测试结果 单位:%

筛选方式	All-Search		Indoor-Search	
	Rank 1	mAP	Rank 1	mAP
Random	52.18	52.40	52.04	65.01
Sampled	52.04	52.47	51.98	64.97
CFS	54.81	53.17	54.04	66.70
CCFS(ours)	55.52	53.46	57.24	68.36

为了验证强通道交换增强与通道叠加擦除增强的有效性,我们分别在预训练过程以及微调过程针对不同的数据增强进行了测试.表6展示了使用不同数据增强方法预训练的模型迁移到AGW方法上的测试结果.其中Default为默认设置,具体包括随机裁剪、随机灰度化、水平翻转等,RE(Random Erasing)代表随机擦

除策略。

表6 采用不同数据增强方法的预训练实验结果 单位:%

数据增强方式	All-Search		Indoor-Search	
	Rank 1	mAP	Rank 1	mAP
Default	53.74	52.26	56.21	66.87
RE	53.93	52.74	56.43	66.72
SCC	54.96	53.01	56.89	67.36
SCC + RE	55.07	53.14	57.04	67.43
SCC + COE	55.52	53.46	57.24	68.36
SCC + COE + RE	53.14	51.89	55.43	65.07

从表6可以看出在默认设置下增加强通道组合与通道叠加增强策略之后,性能得到了明显的提升,在Rank-1和mAP上分别提升了1.78%和1.2%.同时,可以发现在默认设置下加入随机擦除策略(RE)后性能也有略微提升,但当随机擦除(RE)与叠加擦除(COE)共同使用时,性能出现了明显的下降,这是因为叠加擦除保留了在纹理层面的硬擦除,当随机擦除与叠加擦除同时存在时,相当于扩大了硬擦除的范围,而大范围的擦除可能会增加大量的困难样本,影响训练过程的收敛。

在微调阶段,我们在AGW、DDAG两种方法上评估本文提出的增强策略,具体结果如表7所示.表7结果表明,在原方法基础上加入我们的数据增强策略可以

显著提高模型性能,在 AGW 上 Rank-1 甚至得到了超过 15% 的提升. 强通道组合与叠加擦除作为一种轻量级的数据增强方法,可以无缝衔接到已存的跨模态行人重识别方法当中,并有效的提高其识别率.

表 7 采用不同数据增强方法的微调实验结果 单位:%

方法	All-Search		Indoor-Search	
	Rank 1	mAP	Rank 1	mAP
AGW	47.50	47.65	54.17	63.97
AGW + Ours	64.34	59.50	70.97	77.64
	↑16.84	↑11.85	↑16.80	↑13.67
DDAG	54.75	53.02	61.02	67.98
DDAG + Ours	60.87	58.16	69.09	73.72
	↑6.12	↑5.14	↑8.07	↑5.74

4.5 超参数分析

在对比损失当中,温度超参数 τ 决定了对比损失对困难负样本的关注程度^[25]. τ 值过小,则模型对困难负样本的关注程度就越低,而 τ 过大则会降低正负样本之间的鉴别性. 因此合适的温度系数 τ 对提高模型的学习能力是极其重要的. 本节对不同的温度系数进行了讨论,实验数据集为 SYSU-MM01. 实验结果如图 7 所示.

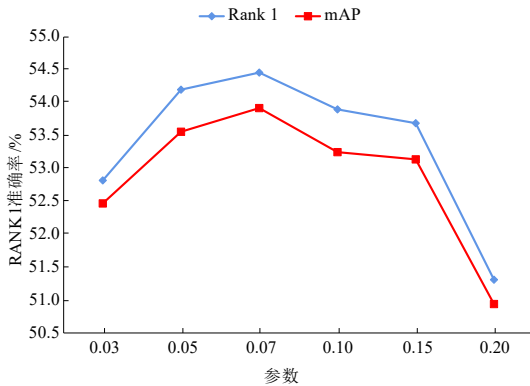


图 7 超参数 τ 与 Rank-1 和 mAP 的关系

从 0 到 0.07 时,随着 τ 的增加,Rank-1 和 mAP 参数也在提升,当 τ 大于 0.07 时,模型性能开始出现下降的趋势,在 RegDB 上也展现出了类似的结果,因此本文取 $\tau=0.07$.

5 结论

本文提出了面向于跨模态行人重识别任务的预训练方法. 首先通过跨模态灾难性遗忘指数对预训练数据进行筛选,该指数通过计算图片之间的特征相似度,有效的减少了 LUPerson 与 SYSU-MM01 的域间差距. 其次,本文利用了提出的强通道组合与通道叠加擦除方法构建了自监督预训练模型,通过这种强增强方式提高了模型对颜色的鲁棒性,促使模型关注于行人图像

的细粒度信息. 在两个跨模态行人数据集 SYSU-MM01, RegDB 上的迁移实验结果表明,本文所提出的预训练方法明显优于目前主流的自监督方法.

参考文献

- [1] WU A C, ZHENG W S, YU H X, et al. RGB-infrared cross-modality person re-identification[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 5390-5399.
- [2] YE M, SHEN J B, LIN G J, et al. Deep learning for person re-identification: A survey and outlook[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 2872-2893.
- [3] 孙锐, 赵争晖, 杨梓, 等. 跨模态异构行人再识别的研究进展[J]. 模式识别与人工智能, 2020, 33(12): 1066-1082. SUN R, ZHAO Z H, YANG Z, et al. A survey on cross-modality heterogeneous person re-identification[J]. Pattern Recognition and Artificial Intelligence, 2020, 33(12): 1066-1082. (in Chinese)
- [4] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 770-778.
- [5] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[EB/OL]. (2021-10-22) [2022-10-02]. <https://arxiv.org/abs/2010.11929>.
- [6] FU C Y, HU Y B, WU X, et al. CM-NAS: Cross-modality neural architecture search for visible-infrared person re-identification[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2022: 11803-11812.
- [7] DENG J, DONG W, SOCHER R, et al. ImageNet: A large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009: 248-255.
- [8] HE K M, GIRSHICK R, DOLLAR P. Rethinking ImageNet pre-training[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2020: 4917-4926.
- [9] ZHENG L, SHEN L Y, TIAN L, et al. Scalable person re-identification: A benchmark[C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2016: 1116-1124.
- [10] WEI L H, ZHANG S L, 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. Piscataway: IEEE, 2018: 79-88.
- [11] 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. New York: ACM, 2018: 274-282.
- [12] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[EB/OL]. (2013-06-16)[2022-10-02]. <https://arxiv.org/abs/1301.3781>.
- [13] HE K M, FAN H Q, WU Y X, et al. Momentum contrast for unsupervised visual representation learning[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 9726-9735.
- [14] CHEN X, FAN H, GIRSHICK R, et al. Improved baselines with momentum contrastive learning[EB/OL]. (2021-03-09)[2022-10-02]. <https://arxiv.org/abs/2003.04297>.
- [15] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//Proceedings of the International Conference on Machine Learning. Shangri-La: JMLR, 2020: 1597-1607.
- [16] GRILLI J B, STRUB F, ALTCHÉ F, et al. Bootstrap your own latent-a new approach to self-supervised learning[J]. *Advances in Neural Information Processing Systems*, 2020, 33: 21271-21284.
- [17] FU D P, CHEN D D, BAO J M, et al. Unsupervised pre-training for person re-identification[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 14745-14754.
- [18] LUO H, WANG P, XU Y, et al. Self-supervised pre-training for transformer-based person re-identification[EB/OL]. (2021-11-23)[2022-10-02]. <https://arxiv.org/abs/2111.12084>.
- [19] ZHU K, GUO H, YAN T, et al. Part-aware self-supervised pre-training for person re-identification[EB/OL]. (2022-03-08)[2022-10-02]. <https://arxiv.org/abs/2203.03931>.
- [20] WAN L, JING Q, SUN Z, et al. Self-supervised modality-aware multiple granularity pre-training for rgb-infrared person re-identification[EB/OL]. (2021-12-12) [2022-10-02]. <https://arxiv.org/abs/2112.06147>.
- [21] RAMASESH V V, DYER E, RAGHU M. Anatomy of catastrophic forgetting: Hidden representations and task semantics[EB/OL]. (2020-07-14)[2022-10-02] <https://arxiv.org/abs/2007.07400>.
- [22] THOMPSON B, GWINNUP J, KHAYRALLAH H, et al. Overcoming catastrophic forgetting during domain adaptation of neural machine translation[C]//Proceedings of the 2019 Conference of the North. Stroudsburg: Association for Computational Linguistics, 2019: 2062-2068.
- [23] YE M, RUAN W J, DU B, et al. Channel augmented joint learning for visible-infrared recognition[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2022: 13547-13556.
- [24] NGUYEN D, HONG H, KIM K, et al. Person recognition system based on a combination of body images from visible light and thermal cameras[J]. *Sensors*, 2017, 17 (3): 605.
- [25] WANG F, LIU H P. Understanding the behaviour of contrastive loss[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 2495-2504.
- [26] YE H C-H, HONG C-Y, HSU Y-C, et al. Decoupled contrastive learning[J]. (2022-10-13) [2022-10-02]. <https://arxiv.org/abs/2110.06848>.
- [27] ZHANG J Y, GE Y X, GU X Q, et al. Self-supervised pre-training on the target domain for cross-domain person re-identification[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 4268-4276.
- [28] ZHONG Z, ZHENG L, KANG G, et al. Random erasing data augmentation[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence and the Thirty-Second Conference on Innovative Applications of Artificial Intelligence and the Tenth Symposium on Educational Advances in Artificial Intelligence. New York: AAAI Press, 2020: 13001-13008.
- [29] ZHONG Z, ZHENG L, LUO Z M, et al. Invariance matters: Exemplar memory for domain adaptive person re-identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 598-607.
- [30] JÜNGLING K KAI, ARENS M. Local feature based person reidentification in infrared image sequences[C]//2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance. Piscataway: IEEE, 2010: 448-455.
- [31] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. *Journal of Big Data*, 2019, 6(1): 1-48.
- [32] DENG K Y, ZHANG C X, CHEN Z, et al. Jointing recurrent across-channel and spatial attention for multi-object

tracking with block-erasing data augmentation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(8): 4054-4069.

- [33] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 779-788.
- [34] HAO Y, WANG N, LI J, et al. HSME: Hypersphere manifold embedding for visible thermal person re-identification[C]//Proceedings of the 33rd AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu: AAAI Press, 2019: 8385-8392.
- [35] LOSHCHILOV I, HUTTER F. SGDR: Stochastic gradient descent with warm restarts[EB/OL]. (2016-08-13) [2022-10-02]. <https://arxiv.org/abs/1608.03983>.
- [36] KINGMA D P, BA J. Adam: A method for stochastic optimization[EB/OL]. (2014-12-22)[2022-10-02]. <https://arxiv.org/abs/1412.6980>.
- [37] YE M, SHEN J B, CRANDALL D J, et al. Dynamic dual-attentive aggregation learning for visible-infrared person re-identification[C]//Computer Vision - ECCV 2020. Cham: Springer International Publishing, 2020: 229-247.
- [38] WANG G A, ZHANG T Z, CHENG J, et al. RGB-infrared cross-modality person re-identification via joint pixel and feature alignment[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2020: 3622-3631.



张 磊 男, 1997年出生于安徽省. 现为合肥工业大学计算机与信息学院在读硕士研究生. 主要研究方向为图像信息处理、计算机视觉.

E-mail: 2020171121@mail.hfut.edu.cn



张旭东 男, 1966年出生于安徽省. 现为合肥工业大学计算机与信息学院教授. 主要研究方向为智能信息处理、机器视觉.

E-mail: xudong@hfut.edu.cn



高 隼 男, 1963年出生于安徽省, 中国科学技术大学博士. 现为合肥工业大学图像信息处理实验室教授. 主要研究方向为图像信息处理、智能信息处理.

E-mail: gaojun@hfut.edu.cn

作者简介



孙 锐 男, 1976年出生于安徽省. 现为合肥工业大学计算机与信息学院教授. 主要研究方向为计算机视觉、机器学习. 中国电子学会会员编号: E190005402S.

E-mail: sunrui@hfut.edu.cn



谢瑞瑞(通讯作者) 男, 1999年出生于安徽省. 现为合肥工业大学计算机与信息学院硕士研究生. 主要研究方向为智能信息处理、行人重识别.

E-mail: 2021171133@mail.hfut.edu.cn