

文章编号: 2095-2163(2022)09-0017-10

中图分类号: TP391.4

文献标志码: A

# 基于全局和局部特征的跨模态行人再识别方法

袁瑞超<sup>1</sup>, 胡晓光<sup>1,2</sup>, 杨世欣<sup>1</sup>

(1 中国人民公安大学 信息网络安全学院, 北京 102600; 2 中国人民公安大学 侦查学院, 北京 102623)

**摘要:** 跨模态行人再识别是对同一行人的可见光图像和红外图像之间进行匹配和识别, 相对于单模态行人再识别的难度进一步加大。本文采用循环生成对抗网络(Cycle GAN)转换和扩充数据集, 尽可能减少行人姿态变化带来的影响; 在 ResNet50 网络的基础上引入全局特征对比模块和局部特征模块, 减少图像噪声和行人遮挡带来的影响; 将交叉熵损失和改进的三元组损失以比例加和的形式作为多损失联合函数, 对网络监督训练。实验结果表明, 该方法的平均精确度均值和前十位命中率都达到了较高的水平, 优于当前多数方法。

**关键词:** 跨模态行人再识别; 循环生成对抗网络; 全局特征对比模块; 多损失联合函数

## Cross-modal pedestrians re-identification method based on global and local features

YUAN Ruichao<sup>1</sup>, HU Xiaoguang<sup>1,2</sup>, YANG Shixin<sup>1</sup>

(1 School of Information and Cyber Security, People's Public Security University of China, Beijing 102600, China;

2 School of Investigation, People's Public Security University of China, Beijing 102623, China)

**[Abstract]** Cross-modal pedestrians re-identification is to match and identify the visible light image and infrared image of the same pedestrian, which is more difficult than single-modal pedestrian re-identification. In this paper, Cycle GAN is used to transform and expand the dataset to minimize the impact of pedestrian posture changes. Based on the ResNet50 network, a global feature comparison module and a local feature module are introduced to reduce images noise and the impact of pedestrians occlusion. The cross-entropy loss and the improved triplet loss in the form of proportional summation are used as a multi-loss joint function to supervise the training of the network. The experimental results show that the average precision and the top ten hit rate of the proposed method have reached a high level, which is better than most of the current methods.

**[Key words]** cross-modal pedestrians re-identification; cycle-consistent adversarial networks; global feature comparison module; multi-loss joint function

## 0 引言

当前行人再识别方法主要是基于可见光图像和视频的行人再识别方法, 着重解决可见光条件下目标人物的背景、光照、姿态等问题。然而, 如果目标人物白天活动较少, 夜间活动较多, 仅依靠采集的可见光图像对目标人物再识别, 难以应对目前犯罪嫌疑人昼伏夜出的问题。较为发达的城市除了大范围安装彩色(RGB)摄像机外, 同时还会安装红外(IR)摄像机, 不仅在白天可以采集到 RGB 图像, 还会在夜晚采集到 IR 图像。因此, 基于 RGB 和 IR 图像的跨模态行人再识别技术对于目标人物轨迹特征的捕捉非常重要。

在 ILSVRC12 的比赛中, Krizhevsky 等人<sup>[1]</sup>设计

了识别精度较高的大型卷积神经网络, 赢得了比赛胜利。自此, 以卷积神经网络为代表的深度学习方法开始流行。应用在行人再识别领域的深度学习方法的特点是通过各种卷积神经网络获得丰富的行人特征, 设计各种各样的损失函数来训练网络。Zhang 等人<sup>[2]</sup>提出了端到端的局部特征联合(Aligned)方法, 采用难采样三元组损失监督训练, 达到再识别目的。Park 等人<sup>[3]</sup>提出了相关关系网络模型, 将三元组损失和交叉熵损失联合, 作为最终的损失函数, 多阶段对网络监督训练。Hermans 等人<sup>[4]</sup>改进传统的三元组损失函数, 在行人类间差距大于类内距离的基础上, 使类内距离要小于某个设定阈值, 来对网络模型监督训练, 获得了更好的识别效果。张涛等人<sup>[5]</sup>提出了基于全局特征的三元组损失改进方法,

**基金项目:** 上海市现场物证重点实验室开放课题基金(2020XCWZK05); 中国人民公安大学新型犯罪研究专项(2021XXFZ010); 中国人民公安大学公共安全行为科学实验室开放课题(2021SYS03); 中国人民公安大学 2021 年度拔尖创新人才培养项目(2021yjyky013)。

**作者简介:** 袁瑞超(1996-), 男, 硕士研究生, 主要研究方向: 行人再识别、智能刑侦技术; 胡晓光(1980-), 男, 博士, 讲师, 硕士生导师, 主要研究方向: 人工智能、计算机视觉; 杨世欣(1996-), 男, 硕士研究生, 主要研究方向: 行人再识别、智能刑侦技术研究。

收稿日期: 2022-04-16

在传统三元组损失的基础上引入类内距离,增大类间距离的同时,缩小类内距离,该模型达到了较好的识别效果。

近年来,基于 RGB-IR 图像的跨模态行人再识别方法成为了广泛关注点,用 RGB 和 IR 摄像机拍摄的行人图像,对网络进行训练,提高同一行人 RGB 和 IR 图像的匹配准确率。目前广泛使用的方法有 2 种。一种是图像生成方法,多采用生成对抗网络,将混合 RGB 和 IR 的多模态图像统一转换为单一模态的 RGB 或 IR 图像;另一种是神经网络学习分类的方法,对行人的 RGB 和 IR 图像同时进行特征提取,获得具有鉴别力的特征,使用合理的损失函数对网络进行监督训练。对此拟展开阐释分述如下。

(1) 采用图像生成的方法。Dai 等人<sup>[6]</sup>采用 GAN 网络,从 RGB 和 IR 图像数据集中学习出具有判别力的特征,用判别器进行判断,最终提高判别器的识别能力。Wang 等人<sup>[7]</sup>提出了像级子网络,将 RGB 图像转换为 IR 图像,一定程度上扩充了数据集,在此基础上训练带有特征嵌入的子网络,从而减小模态差异,达到再识别目的。Wang 等人<sup>[8]</sup>利用 GAN 网络将 RGB 图像转换为 IR 图像,对真实的 IR 图像和虚假的 IR 图像进行判别验证,达到再识别目的。

(2) 采用神经网络学习分类的方法。赵立昌等人<sup>[9]</sup>提出了灰度能量差异性的红外图像与可见光图像融合方法,突出了目标信息和对比度。Wu 等人<sup>[10]</sup>提出了基于深度零填充的方法,采集了 RGB-IR 图像数据集 SUSU-MM01,解决了再识别的研究问题。Ye 等人<sup>[11]</sup>提出了双向中心约束的 top-ranking (eBDTR) 算法,使用了一对网络来分别提取 RGB 和 IR 图像特征,并采用 SoftMax 损失和双向约束排序损失 (Bi-directional ranking loss),对网络进行监督训练。杜鹏等人<sup>[12]</sup>提出自注意力模态融合的方法,利用 Cycle GAN 生成跨模态图像,使用自注意力模块区分原始图像和生成图像,使用模态融合模块融合特征,使用 SoftMax 损失监督训练,达到再识别目的。

本文提出了基于全局和局部特征的方法。首先,采用 Cycle GAN 网络对数据集进行转换和扩充,降低行人姿态变化带来的影响;在 ResNet50 网络的基础上,设计全局特征对比模块和局部特征模块,降低图像噪声带来的影响和行人图像的遮挡问题带来的影响;最后,将交叉熵损失和改进的三元组损失以比例加和的形式作为多损失联合函数,对网络进行监督训练。采用 RGB 图像和 IR 图像混合、RGB 图像转换为 IR 图像、IR 图像转换为 RGB 图像三种实

验思路,做了相关实验。

## 1 相关理论介绍

### 1.1 Cycle GAN 网络原理

GAN 网络通过构造生成器和判别器,使 2 种不同风格的图像相互学习,相互博弈,不断地提高其生成能力和鉴别能力,最终将一种风格的图像生成另外一种风格的图像,其生成效果非常逼真。虽然 GAN 网络的变体模型较多,Cycle GAN 网络采用卷积和反卷积策略,不要求以成对的方式输入图像,生成图像的质量较好,故本文采用 Cycle GAN 网络实现 RGB 图像和 IR 图像之间的转换和扩充。

Cycle GAN 网络实现 RGB 图像和 IR 图像转换和扩充的原理是:RGB 图像经过生成器  $G(AB)$ ,生成 IR 图像,和原有的 IR 图像经过判别器  $D(A)$ ,如原有的 IR 图像被判别为真 (Real),生成的 IR 图像被判别为假 (Fake),Fake-IR 图像再一次经过生成器  $G(BA)$  生成 Cycle-RGB 图像,Cycle-RGB 图像和原始的 RGB 图像几乎一致;同理可知,IR 图像也会转换为 Fake-RGB 图像和 Cycle-IR 图像。Fake-RGB 图像和 Fake-IR 图像几乎可以以假乱真。Cycle GAN 的对 RGB 图像和 IR 图像的转换和扩充示意图如图 1 所示。

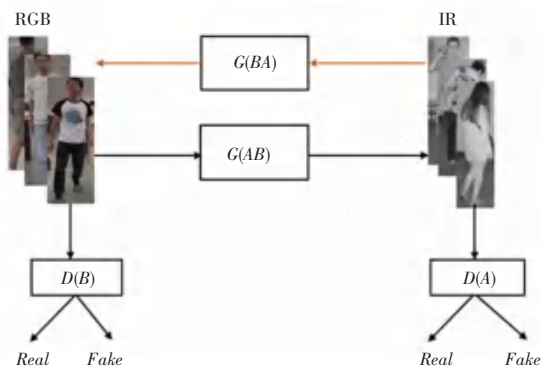


图1 Cycle GAN 网络对 RGB 图像和 IR 图像转换和扩充的示意图  
Fig. 1 Schematic diagram of the conversion and augmentation of RGB images and IR images by Cycle GAN network

### 1.2 全局特征对比模块和局部特征模块

使用 Cycle GAN 网络扩充数据集后,采用双线性插值算法,将图像分辨率统一为高 384 像素,宽 128 像素。使用扩充后的数据集训练 ResNet50 网络,得到大小为  $h \times w \times c$  的整体特征图,将整体特征图输入全局特征对比模块,并将整体特征图在  $h \times w$  的层面上沿着  $h$  划分出 6 个相等局部特征区域,获得局部特征图,将局部特征图输入局部特征模块,如图 2 所示。

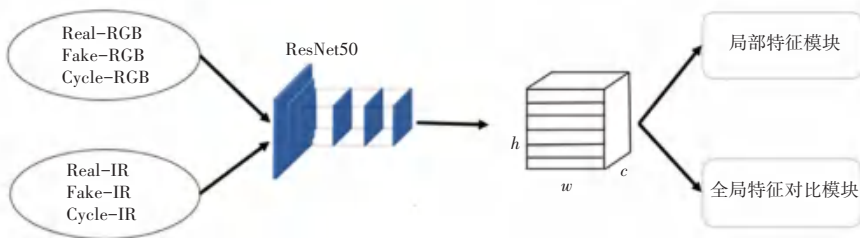


图 2 图像通过 ResNet50 网络后输入全局特征对比模块和局部特征模块的示意图

Fig. 2 The flow chart of global feature comparison module and local feature module after the image passes through the ResNet50 network

全局特征对比模块的原理: 扩充后的数据集通过 ResNet50 网络后得到特征图, 将其作为整体, 分别进行全局最大池化和全局平均池化, 得到通道数为  $c$  的特征向量  $P_a$  和  $P_m$ ; 采取  $P_a - P_m$  方法, 使用  $1 \times 1$  的卷积核进行卷积, 批处理归一化,  $ReLU$  函数

激活, 得到  $P_n$ ; 最终将特征图的通道  $c$  降低。全局特征对比模块的原理示意如图 3 所示。图 3 中,  $h, w, c$  分别表示特征图的高、宽、通道数,  $A_P, A_M$  分别表示平均池化和最大池化。

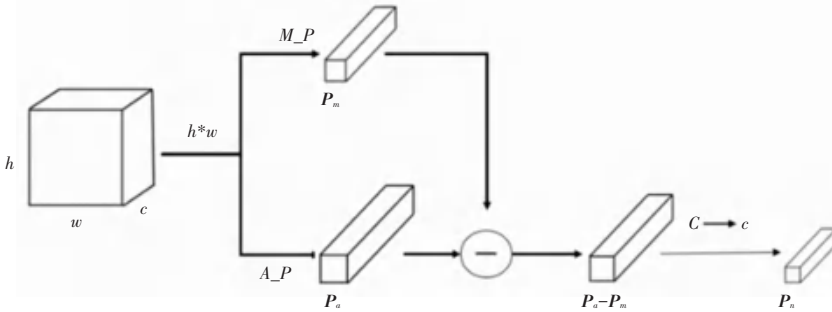


图 3 全局特征对比模块原理示意图

Fig. 3 Schematic diagram of the principle of global feature comparison module

噪声在数字图像中体现为白点, 卷积运算的原理是点乘、相加, 噪声同样参与卷积运算, 其影响同样存在于特征图和特征向量当中。平均池化和最大池化过程中图像噪声影响示意如图 4 所示。研究假设图像经过 2 个不同的卷积核的卷积运算形成  $M_1$  和  $M_2$ , 假设  $M_1$  的  $x_3, x_4$  有噪声的成分, 且在数值上大于  $x_1, x_2$ ,  $M_1$  经过最大池化 (Max Pooling) 所形成的特征值  $y_1$  也包含噪声成分; 平均池化 (Average Pooling) 不同于最大池化, 是将  $M_2$  的所有值加和平均, 含有噪声的  $x_6, x_7$  参与了平均池化运算, 形成的特征值  $y_2$  也会有噪声成分。将这 2 个同时含有噪声成分的特征值  $y_2$  和  $y_1$  相减, 在一定程度上能够降低噪声。因此, 全局特征对比模块中,  $P_a - P_m$  的方法在一定程度上能够降低图像噪声带来的影响。

局部特征模块如图 5 所示, 这里将整体特征图在  $h \times w$  层面上沿着  $h$  方向划分为 6 个大小相等的区域, 选择其中一个区域全局最大池化, 得到特征向量  $L_i$ ; 其余的 5 个区域按顺序组合, 形成另外一部分特征图, 将这一部分特征图全局最大池化, 得到特征向量  $R_i$ 。一方面将  $L_i$  和  $R_i$  分别使用  $1 \times 1$  的卷积核进行卷积、批处理归一化和  $ReLU$  函数激活操作, 得到  $P_i$  和  $Q_i$ ; 另一方面将  $L_i$  和  $R_i$  拼接, 得到特征向量  $T_i$ , 对  $T_i$  使用  $1 \times 1$  的卷积核卷积、批处理归一化和  $ReLU$  函数激活操作, 得到  $H_i$ 。由于将特征图划分为 6 个区域, 每次选择其中的一个区域, 其余的 5 个区域顺序组合, 这样选择局部特征的方式共有 6 种, 即  $P_i, Q_i, T_i$  分别有 6 种, 记  $P_i = \{P_1, P_2, P_3, P_4, P_5, P_6\}, Q_i = \{Q_1, Q_2, Q_3, Q_4, Q_5, Q_6\}, T_i = \{T_1, T_2, T_3, T_4, T_5, T_6\}$ 。

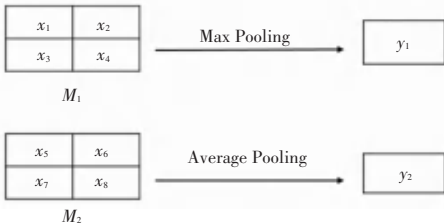


图 4 平均池化和最大池化过程中图像噪声影响的示意图

Fig. 4 Schematic illustration of the effect of images noise during Average Pooling and Max Pooling

由于将特征图均匀分割为 6 份, 选取其中的一份进行运算得到  $P_i$ 。假设行人图像的腿脚区域被遮挡, 而  $P_i$  是头部区域运算得到的, 这样通过运算得到的  $P_i$  就降低了遮挡带来的影响, 局部特征模块在一定程度上能够降低遮挡对识别模型的影响。

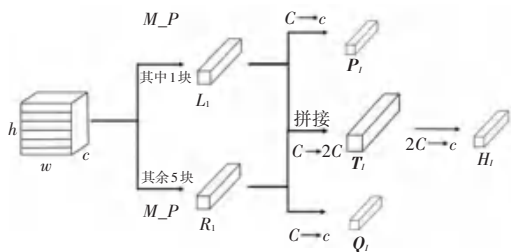


图5 局部特征模块示意图

Fig. 5 Schematic diagram of global feature comparison module

### 1.3 多损失联合

交叉熵损失经常用于行人再识别中的多分类问题。传统的三元组损失在于寻找原始样本、正样本、负样本,通过网络不断学习,使原始样本和正样本之间距离不断地缩小,和负样本之间的距离尽可能地增大。难采样的三元组损失对其改进,所选择的正样本和原始样本之间的距离尽可能远,负样本和原始样本之间距离尽可能近。本文参考难采样三元组损失,对三元组损失进行修改:假设所选择的原始样本是行人的 RGB 图像,正样本选取同一行人的 IR 图像,正样本的背景与行人的动作姿态和原始样本的差异尽可能大;负样本选取不同行人的 RGB 图像,且和原始样本的 RGB 图像在背景和行人姿态方面尽可能相似。改进的三元组损失如图 6 所示。图 6 中,  $A_{-RGB}$  表示原始样本的 RGB 图像,  $N_{-RGB}$  表示和原始样本的 RGB 图像尽可能相似的负样本图像,  $P_{-IR}$  表示和原始样本 RGB 图像之间的距离尽可能远的正样本 IR 图像;反之,若原始样本选取为 IR 图像,其正样本和负样本的选取策略相似。

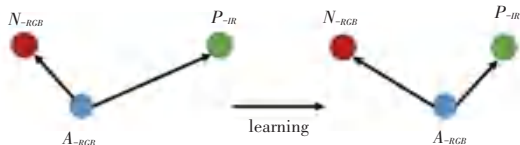


图6 改进三元组损失示意图

Fig. 6 Schematic diagram of improving triple loss

将交叉熵损失和改进三元组损失以比例加和的形式作为多损失联合函数,对此可表示为:

$$L = L_{IT} + kL_C \quad (1)$$

其中,  $L_C$  表示交叉熵损失;  $L_{IT}$  表示改进的三元组损失;参数  $k$  取正整数。

## 2 基于全局和局部特征的跨模态行人再识别方法的设计

根据以上原理,本文设计了基于全局和局部特征的跨模态行人再识别的网络结构,如图 7 所示。由图 7 可知,首先利用数据集训练 Cycle GAN 网络,分别将所有的 RGB 图像转换为 IR 图像,所有的 IR 图像转换为 RGB 图像。将同一行人生成的图像和原始的图像视作拥有同样的行人身份,对图像进行镜像、裁剪、颜色变换操作,以进一步扩充数据集。接着使用扩充后的数据集训练 ResNet50 网络,得到整体特征图,一方面将整体特征图输入全局特征对比模块;另一方面将整体特征图输入局部特征模块。最后,使用多损失联合函数对网络进行监督训练。

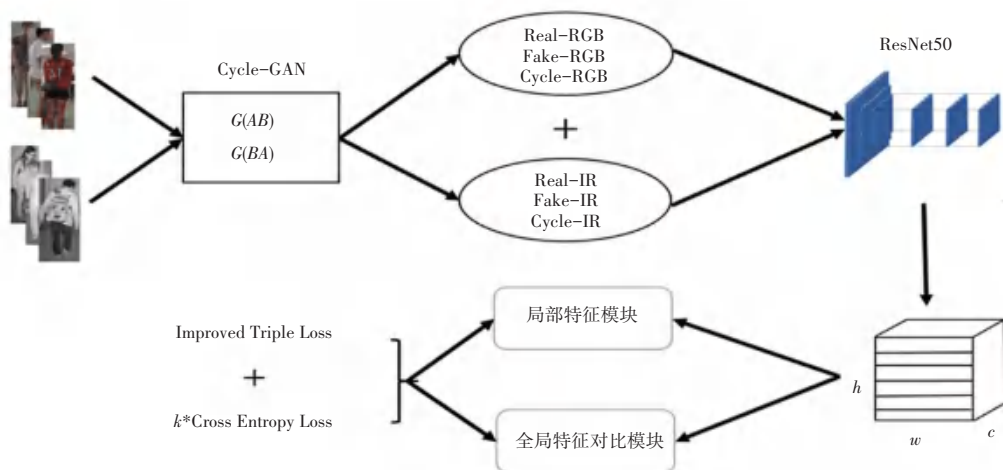


图7 基于全局和局部特征的跨模态行人再识别方法的网络结构图

Fig. 7 Network architecture diagram of cross-modal pedestrians re-identification method based on global and local features

Cycle GAN 网络实现 RGB 图像和 IR 图像之间的转换和扩充后,将同一行人生成的图像和原始的图像视作拥有相同的行人身份,并进行镜像、裁剪、颜色变换操作,以进一步扩充数据集。将所有图像的分辨率统一调整为  $128 \times 384$ ,将生成的图像和原始的图像一同输入到 ResNet50 网络中进行训练,得到通道数为 2 048 的整体特征图。一方面,将整体特征图输入全局特征对比模块,在得到通道数为 2 048 的特征向量  $P_a$  和  $P_m$  后,采取  $P_a - P_m$  方法,使用 256 个  $1 \times 1$  的卷积核进行卷积,批处理归一化,  $ReLU$  函数激活,得到  $P_n$ ,最终将特征图的通道降低到 256;另一方面,将整体特征图输入局部特征模块,在得到通道数为 2048 的特征向量  $L_l$  和  $R_l$  后,使用 256 个  $1 \times 1$  的卷积核进行卷积、批处理归一化和  $ReLU$  函数激活操作,得到  $P_l$  和  $Q_l$ ,将特征图的通道数降低到 256;将  $L_l$  和  $R_l$  拼接,得到  $T_l$ ,对  $T_l$  使用 256 个  $1 \times 1$  的卷积核进行卷积、批处理归一化和  $ReLU$  函数激活操作,得到  $H_l$ ,将特征图的通道数降低到 256;最后,使用多损失联合函数对网络监督训练,其中参数取正整数。

### 3 实验仿真

本文实验均在 PyTorch 框架下运行。 $batch\ size$  设定为 64; $epoch$  设定为 200;初始学习率设定为 0.001,从第 50 个  $epoch$  开始进行衰减,衰减系数为 0.1;多损失联合函数的参数  $k$  取正整数。

评估指标选择常见的累计匹配特性 (CMC) 曲线的  $top - 1$ 、 $top - 10$ 、 $top - 20$  和平均准确率均值 ( $mAP$ )。指标越高,表示模型的识别效果越好。

#### 3.1 数据集

(1)数据集 SYSU-MM01。分为室内模式和室外模式,共包含 287 628 张 RGB 图像,15 792 张 IR 图像,共计 491 个行人。本文实验选取其中的 395 个行人的图像用于训练和验证,剩余的 96 个行人的图像用于测试,并对其多次划分,减小模型训练结果的误差,部分图像示例如图 8 所示。

(2)数据集 RegDB。包含 8 240 张行人图像,412 个行人,每个行人 10 张 RGB 图像、10 张 IR 图像。本文实验对数据集多次划分,一半用于训练和验证,另一半用于测试,减小模型训练结果的误差,部分示例图像如图 9 所示。

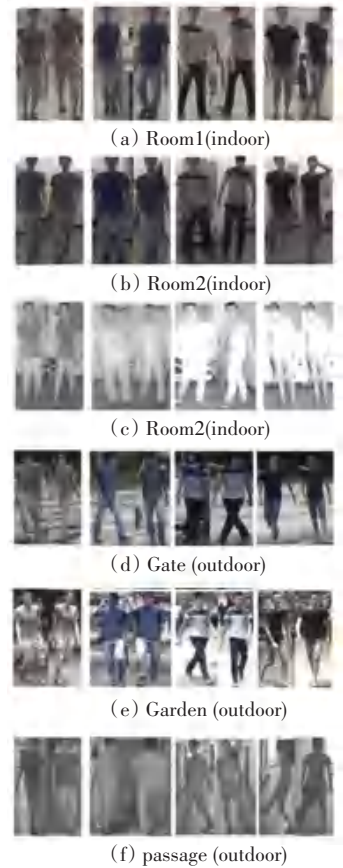


图 8 数据集 SYSU-MM01 部分图像示例图

Fig. 8 Example images of part of SYSU-MM01 dataset



图 9 数据集 RegDB 的部分图像示例图

Fig. 9 Example images of part of RegDB dataset

### 3.2 实验结果与分析

#### 3.2.1 Cycle GAN 网络对不同模态图像的转换结果

以 Cycle GAN 网络训练 RegDB 数据集为例,损失曲线如图 10 所示。图 10 中,  $D_A$  的曲线表示判别器  $D(A)$  的损失;  $G_A$  的曲线表示生成器  $G(AB)$  的损失;  $cycle_A$  的曲线表示  $Real - A$  图像经过生成器  $G(AB)$  生成  $Fake - B$  图像后,由  $Fake - B$  再经过生成器  $G(BA)$  生成的  $A$  图像,和原始的  $Real - A$  图像相比较的损失;  $D_B$  的曲线表示判别器  $D(B)$  的损失;  $G_B$  的曲线表示生成器  $G(BA)$  的损失;  $cycle_B$  的曲线表

示  $Real - B$  图像经过生成器  $G(BA)$  生成  $Fake - A$  图像,由  $Fake - A$  图像经过生成器  $G(AB)$  生成的  $B$  图像,和原始的  $Real - B$  图像相比较的损失;  $idt_A$  的曲线表示将  $Real - B$  图像输入生成器  $G(AB)$  生成的图像和原始的  $Real - B$  图像相比较的损失,主要用于判断生成器  $G(AB)$  对输入的  $Real - B$  图像能否判断为其自身;  $idt_B$  的曲线表示将  $Real - A$  图像输入生成器  $G(BA)$  生成的图像和原始的  $Real - A$  图形相比较的损失,主要用于判断生成器  $G(BA)$  对输入的  $Real - A$  图像能否判断为其自身。

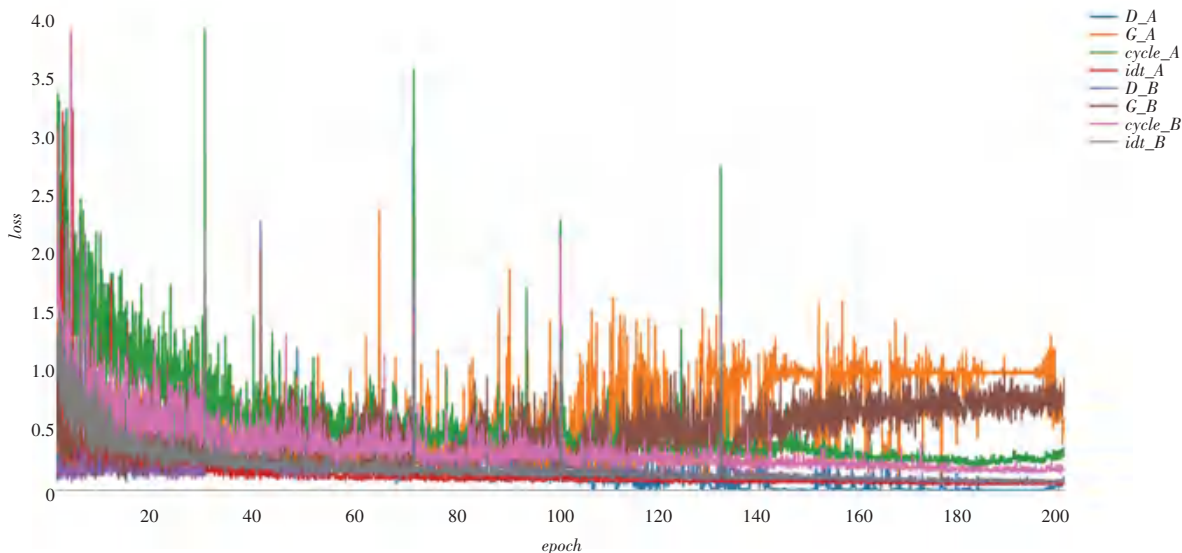


图 10 Cycle GAN 训练 RegDB 数据集损失曲线图

Fig. 10 Loss curve of Cycle GAN training RegDB dataset

在选取生成器对数据集转换之前,综合考虑每一组模型各个损失,使其尽可能小。本次实验选取  $epoch$  为 75 的一组模型,其中,判别器  $D(A)$  的损失为 0.232;生成器  $G(AB)$  的损失为 0.418;  $Real - A$  图像经过生成器  $G(AB)$  生成  $Fake - B$  图像后,由  $Fake - B$  再经过生成器  $G(AB)$  生成的  $A$  图像,和原始的  $Real - A$  图像相比较的损失为 0.540;  $Real - B$  图像输入生成器  $G(AB)$  生成的图像和原始  $Real - B$  图像相比较的损失为 0.123;判别器  $D(B)$  的损失为 0.248,生成器  $G(BA)$  的损失为 0.256;  $Real - B$  图像经过生成器  $G(BA)$  生成  $Fake - A$  图像,由  $Fake - A$  图像经过生成器  $G(AB)$  生成的  $B$  图像,和原始的  $Real - B$  图像相比较的损失为 0.229;  $Real - A$  图像输入生成器  $G(BA)$  生成的图像和原始的  $Real - A$  图像相比较的损失为 0.217。

数据集 RegDB 经过 Cycle GAN 网络转换后的

部分图像示意如图 11 所示。图 11 中,第一行人物图像的  $Fake - IR$  图像相比于  $Real - IR$  图像,轮廓和动作更加清晰,  $Cycle - RGB$  图像、 $Cycle - IR$  图像与原始的图像几乎没有大的变化;第二行人物图像的  $Fake - IR$  图像相比  $Real - IR$  图像,衣着和背包显得更加清晰,  $Cycle - RGB$  图像、 $Cycle - IR$  图像与原始图像几乎一致;第三行人物图像的  $Fake - IR$  图像相比于  $Real - IR$  图像,轮廓更为清晰。  $Real - IR$  图像转换为  $Fake - RGB$  图像,在图像背景、人物衣着等有所变化,肢体动作等几乎保持原样,这类情况可以看作同一行人穿着不同颜色的衣服,在不同的光照条件下被摄像头捕捉到。因此,数据集通过 Cycle GAN 网络的转换和扩充,一方面增加了图像的数据量,另一方面增加了图像的轮廓等细节,这有助于降低跨模态行人再识别中行人的姿态变化带来的影响。

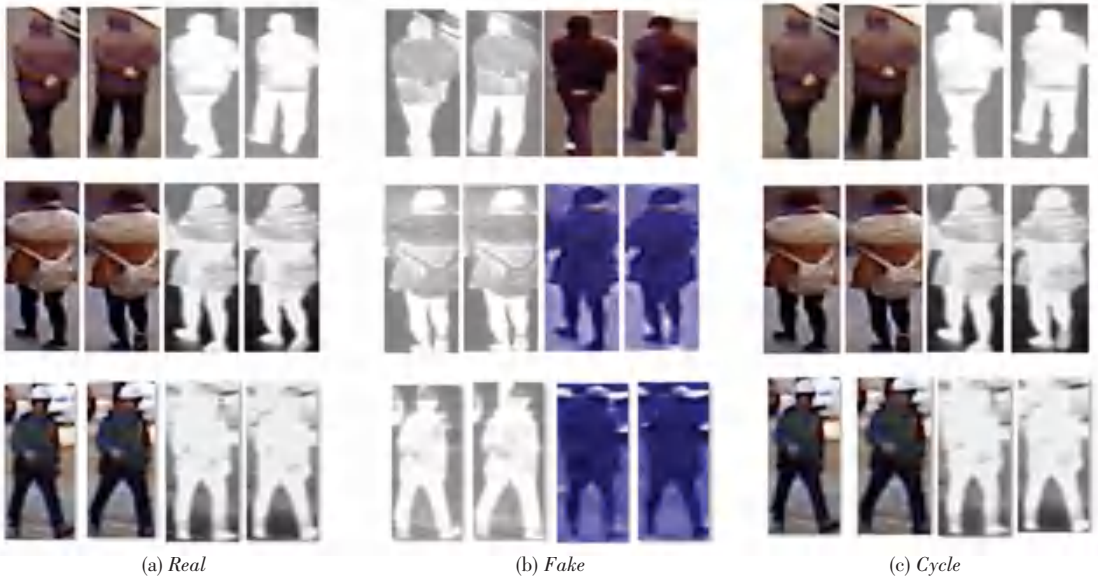


图 11 数据集 RegDB 经过 Cycle GAN 网络转换后的部分图像示意图

Fig. 11 Schematic diagram of some images of RegDB dataset transformed by the Cycle GAN

3.2.2 在 SYSU-MM01 数据集上的识别结果与分析  
 在数据集 SYSU-MM01 上识别测试有单击和多击两种模式。其中,单击模式选取所有的 IR 图像作为 *query*,并随机选取对应行人的一张 RGB 图像组成 *gallery*;在多击模式中,*query* 的选取和单击模式一致,*gallery* 是每个人选取 10 张对应的 RGB 图像。室内场景下环境条件较为理想,图像质量较高;

室外场景下行人图像的光照变化较大,存在遮挡现象,再识别难度较大。相比之下,单击模式更有实际意义,本次测试不论在室内模式、还是全搜索模式均采用单击模式。因该数据集存在部分行人缺失 RGB 图像或者 IR 图像的现象,因此研究中采用 RGB 图像和 IR 图像混合的方式进行实验。本文方法(ALL)的识别精度和其他方法的对比见表 1。

表 1 在 SYSU-MM01 数据集上单击模式的识别精度和其他方法对比

Tab. 1 Comparison of the recognition accuracy between single-click patterns and other methods on SYSU-MM01 dataset %

方法	室内搜索				全搜索			
	<i>top - 1</i>	<i>top - 10</i>	<i>top - 20</i>	<i>mAP</i>	<i>top - 1</i>	<i>top - 10</i>	<i>top - 20</i>	<i>mAP</i>
Zero-pad <sup>[10]</sup>	20.58	68.38	85.79	26.92	14.80	54.12	71.33	15.95
TONE <sup>[13]</sup>	20.82	68.86	84.46	26.38	12.52	50.72	68.60	14.42
HCML <sup>[13]</sup>	24.52	73.25	86.73	30.08	14.32	53.16	69.17	16.16
BDTR <sup>[11]</sup>	31.92	77.18	89.28	41.86	27.32	66.96	81.07	27.32
cm GAN <sup>[6]</sup>	31.63	77.23	89.18	42.19	26.97	67.51	80.56	27.80
MSR <sup>[14]</sup>	39.64	89.29	97.66	50.88	37.35	83.40	93.34	38.11
Align GAN <sup>[8]</sup>	45.94	87.63	94.42	54.31	42.42	85.03	93.71	40.72
HPILN <sup>[15]</sup>	45.77	91.82	98.46	56.52	41.36	84.78	94.31	42.95
自注意力模态融合 <sup>[12]</sup>	38.09	81.68	90.61	47.86	33.31	74.51	85.79	39.71
本文方法(ALL)	39.67	86.47	92.43	51.67	<b>43.82</b>	<b>86.37</b>	<b>95.43</b>	<b>45.13</b>

在全搜索场景下的单击模式,本文方法利用 Cycle GAN 网络对数据集 SYSU-MM01 扩充和转换,将生成的 RGB 图像、IR 图像和原始的图像一同

输入到 ResNet50 网络中进行训练;将特征图分别输入到全局特征对比模块和局部特征模块,使用参数 *k* 为 2 的多损失联合函数进行监督训练,模型效果

达到最好。本文方法(ALL)和 Align GAN<sup>[8]</sup>方法相比  $top - 1$  指标提高了 1.4%,  $top - 10$  指标提高了 1.34%; 和 HPILN<sup>[15]</sup>方法相比  $top - 20$  指标提高了 1.12%,  $mAP$  值提高了 2.18%。在 SYSU-MM01 数据集上测试时部分图像的认识结果如图 12 所示。



(a) query (b) gallery

图 12 在 SYSU-MM01 数据集上测试时部分图像的认识结果图

Fig. 12 Recognition results of some images when tested on SYSU-MM01 dataset

### 3.2.3 在 RegDB 数据集上的识别结果与分析

在数据集 RegDB 上测试时,测试集中所有的 IR 图像组成 *query*, 所有对应的 RGB 图像组成 *gallery*, 识别精度和其他方法的对比见表 2。

由表 2 可知,本文方法(ALL)使用参数  $k$  为 3 的多损失联合函数进行监督训练,模型精度最高。本文方法(ALL)的  $top - 10$  指标达到了 89.03%, 比 MACE<sup>[18]</sup>方法提高了 0.63%,  $mAP$  值达到 69.26%, 比 DFE<sup>[19]</sup>方法提高了 0.12%。

本文方法(IR)和本文方法(RGB)是分别利用 Cycle GAN 网络将数据集 RegDB 的 RGB 图像转换为 IR 图像、IR 图像转换为 RGB 图像,将原始的图像和转换后的图像一同输入到 ResNet50 网络中

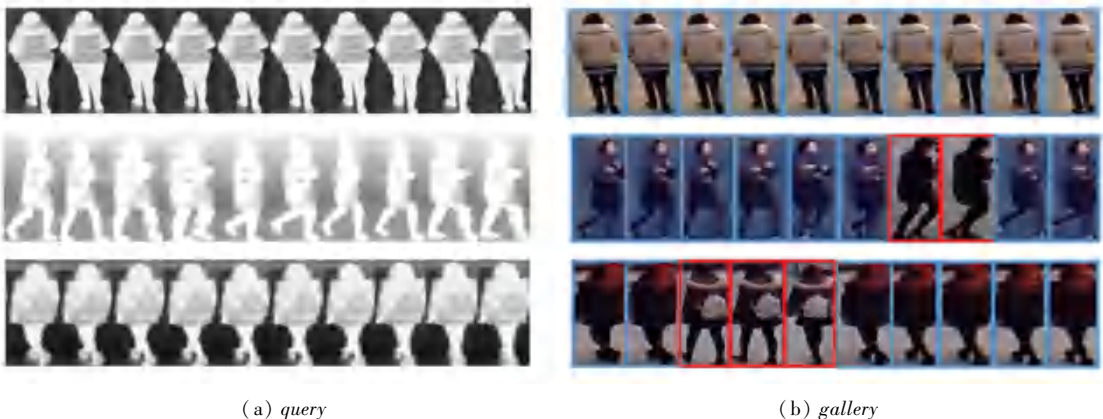
训练,将特征图分别输入到全局特征对比模块和局部特征模块,并使用参数  $k$  为 2 的多损失联合函数监督训练,模型识别效果较好。 $mAP$  指标分别达到了 65.28% 和 62.96%,  $top - 1$  指标分别达到 67.16% 和 66.67%,  $top - 10$  指标分别达到了 86.47% 和 81.67%, 接近 MACE<sup>[18]</sup>、DFE<sup>[19]</sup>方法,  $top - 20$  指标分别达到 91.06% 和 90.23%, 与本文方法(ALL)较为接近。

表 2 在 RegDB 数据集上的识别精度和其他方法的对比

Tab. 2 Comparison of recognition accuracy with other methods on RegDB dataset %

方法	RGB - IR			
	$top - 1$	$top - 10$	$top - 20$	$mAP$
Zero-pad <sup>[10]</sup>	17.74	34.21	44.35	18.90
HCML <sup>[13]</sup>	24.44	47.53	56.78	20.08
BDTR <sup>[11]</sup>	33.56	58.61	67.43	32.76
HSME <sup>[16]</sup>	50.85	73.36	81.66	47.00
JSIA <sup>[17]</sup>	48.52	—	—	49.32
MSR <sup>[14]</sup>	48.43	70.32	79.95	48.67
MACE <sup>[18]</sup>	72.37	88.40	93.59	69.09
DFE <sup>[19]</sup>	82.54	86.32	91.96	69.14
本文方法(IR)	66.67	81.67	90.23	62.96
本文方法(RGB)	67.16	86.47	91.06	65.28
本文方法(ALL)	74.18	<b>89.03</b>	<b>92.67</b>	<b>69.26</b>

在 RegDB 数据集上测试时部分图像的认识结果如图 13 所示。图 13 中,图 13(b) 中的蓝色框部分的图像表示识别正确的图像,红色框部分的图像表示识别错误的图像。从识别错误的图像看出,识别错误的 RGB 图像在轮廓和姿势等方面都比较接近 IR 图像。



(a) query

(b) gallery

图 13 在 RegDB 数据集上测试时部分图像的认识结果图

Fig. 13 Plot of recognition results for some images when tested on RegDB dataset



3.2.4 在 RegDB 数据集上的消融实验结果与分析  
数据集 RegDB 的图像分布比较均匀,没有极低质量的图像,每个行人的图像之间较为相似,姿态变

化较小。因此,在数据集 RegDB 进行消融实验,识别精度和其他方法对比见表3。

表3 在 RegDB 数据集上消融实验的结果  
Tab. 3 Results of ablation experiments on RegDB dataset

Cycle GAN	全局特征对比模块和 局部特征模块	损失函数	$top - 1$	$top - 10$	$top - 20$	$mAP$
0	0	$L = L_{IT} + 3L_C$	17.73	34.24	44.32	18.94
0	P	$L = L_C$	47.32	65.31	75.10	49.23
0	P	$L = L_{IT}$	51.41	75.52	84.20	50.33
0	P	$L = L_{IT} + L_C$	49.23	58.61	72.20	40.32
0	P	$L = L_{IT} + 2L_C$	53.21	72.83	82.04	52.33
0	P	$L = L_{IT} + 3L_C$	64.53	82.43	87.93	63.62
P(IR)	P	$L = L_{IT} + 2L_C$	66.73	81.63	90.20	62.92
P(RGB)	P	$L = L_{IT} + 2L_C$	67.23	86.42	91.03	65.24
P(ALL)	P	$L = L_{IT} + 3L_C$	74.22	89.03	92.62	69.24

在消融实验当中,使用原始数据集 RegDB 训练 ResNet50 网络得到特征图,将特征图输入全局特征对比模块和局部特征模块,采用参数  $k$  为 3 的多损失联合函数进行监督训练,与单独使用 ResNet50 网络,采用参数  $k$  为 3 的多损失联合函数进行监督训练的方法相比, $mAP$ 、 $top - 1$ 、 $top - 10$  和  $top - 20$  指标都有大幅度的提高,证明了全局特征对比模块和局部特征模块有助于提高模型的精度。

使用原始数据集 RegDB 训练 ResNet50 网络得到特征图,将特征图输入全局特征对比模块和局部特征模块,采用参数  $k$  为 3 的多损失联合函数进行监督训练。和使用单个的交叉熵损失或改进的三元组损失监督训练的方法相比较, $top - 1$  指标分别从 47.32%、51.41% 提高到 64.53%; $top - 10$  指标分别从 65.31%、75.52% 提高到 82.43%; $top - 20$  指标分别从 75.10%、84.20% 提高到 87.93%; $mAP$  值则分别从 49.23%、50.33% 提高到 63.62%,证明了本文的参数  $k$  为 3 的多损失联合函数有助于提高模型的精度。

本文方法(IR)和本文方法(RGB)是使用 Cycle GAN 网络将数据集分别转换为 IR 图像和 RGB 图像,输入到 ResNet50 网络得到特征图,将特征图输入全局特征对比模块和局部特征模块,采用参数  $k$  为 2 的多损失联合函数进行监督训练,和不使用 Cycle GAN 网络对数据集转换的方法相比较, $top - 1$  指标从 53.21% 分别提高到 66.73% 和 67.23%; $top - 10$  指标从 72.83% 分别提高到 81.63% 和 86.42%;

$top - 20$  指标从 82.04% 分别提高到 90.20% 和 91.03%,接近本文方法(ALL)的 92.62%; $mAP$  值从 52.33% 分别提高到 62.92% 和 65.24%,证明本文使用 Cycle GAN 网络将 RegDB 数据集统一转换为 RGB 图像或 IR 图像的方法同样对模型的识别精度有一定程度的提高。

本文方法(ALL)和不使用 Cycle GAN 网络对数据集扩充和转换的方法相比较, $top - 1$  指标从 64.53% 提高到 74.22%; $top - 10$  指标从 82.43% 提高到 89.03%; $top - 20$  指标从 87.93% 提高到 92.62%, $mAP$  指标从 63.62% 提高到 69.24%,证明了采用 Cycle GAN 网络扩充和转换的数据集,对于网络模型的识别精度有较大程度的提高,同时也验证了本文方法(ALL)的有效性。

## 4 结束语

本文提出了基于全局和局部特征的跨模态行人重识别方法。使用 Cycle GAN 网络对数据集扩充和转换,来降低同一行人的 RGB-IR 图像的姿态变化带来的影响;全局特征对比模块和局部特征模块分别降低图像噪声和行人遮挡现象带来的影响;将改进的三元组损失和交叉熵损失以比例加和的形式作为多损失联合函数,进行监督训练;探索多损失联合函数参数  $k$  取不同的值,对模型精度的影响;分别采取 RGB 图像和 IR 图像联合实验、RGB 图像转换为 IR 图像实验、IR 图像转为 RGB 图像实验。在当前的 SYSU-MM01、RegDB 数据集上与当前多数方法

相比,识别准确率有一定程度的提升。通过消融实验论证了 Cycle GAN 网络不论是数据集的扩充,还是将其统一转换为 RGB 图像或 IR 图像,对于网络模型的识别精度都有一定程度的提升。但是,Cycle GAN 网络在 IR 图像转换为 RGB 图像时,转换的图像有一定的失真,噪声也较大;单流网络架构对于 RGB-IR 图像的跨模态行人再识别存在一定的局限性。因此,生成对抗网络对于 RGB-IR 图像数据集的扩充和转换、图像噪声的抑制、更好的网络模型及损失函数的探索可能是未来 RGB-IR 图像跨模态行人再识别研究的热点。

## 参考文献

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C]//Advances in Neural Information Processing Systems. Lake Tahoe, Nevada, United States;NIPS Foundation, 2012, 28:1097-1105.
- [2] ZHANG X, LUO H, FAN X, et al. Alignedredid: Surpassing human-level performance in person re-identification[J]. arXiv preprint arXiv:1711.08184, 2017.
- [3] PARK H, HAM B. Relation network for person re-identification [C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA; AAAI, 2020, 34(07): 11839-11847.
- [4] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification [J]. arXiv preprint arXiv: 1703.07737, 2017.
- [5] 张涛,易争明,李璇,等.一种基于全局特征的行人再识别改进算法[J].激光与光电子学进展,2020,57(24):324-330.
- [6] DAI P, JI R, WANG H, et al. Cross-modality person re-identification with generative adversarial training [C]//International Joint Conference on Artificial Intelligence (IJCAI). Stockholm, Sweden; IBM Research,2018;677-683.
- [7] WANG Zhixiang, WANG Zheng, ZHENG Yinqiang, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA;IEEE,2019; 618-626.
- [8] WANG Guan'an, ZHANG Tianzhu, CHENG Jian, et al. Rgb-infrared cross-modality person re-identification via joint pixel and feature alignment [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South);IEEE, 2019; 3623-3632.
- [9] 赵立昌,张宝辉,吴杰,等.基于灰度能量差异性的红外与可见光图像融合[J].红外技术,2020,42(08):775-782.
- [10] WU Ancong, ZHENG Weishi, YU Hongxing, et al. RGB-infrared cross-modality person re-identification[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy;IEEE, 2017; 5380-5389.
- [11] YE Mang, LAN Xiangyuan, WANG Zheng, et al. Bi-directional center-constrained top-ranking for visible thermal person re-identification[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 407-419.
- [12] 杜鹏,宋永红,张鑫瑶.基于自注意力模态融合网络的跨模态行人再识别方法研究[J].自动化学报,2022,48(06):1457-1468.
- [13] YE M, LAN X, LI J, et al. Hierarchical discriminative learning for visible thermal person re-identification [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Louisiana, USA; AAAI,2018;7501-7508.
- [14] FENG Zhanxiang, LAI Jianhuang, XIE Xiaohua. Learning modality-specific representations for visible-infrared person re-identification[J]. IEEE Transactions on Image Processing, 2019, 29: 579-590.
- [15] LIN Jianwu, LI Hao. HPILN: A feature learning framework for cross-modality person re-identification[J]. arXiv preprint arXiv: 1906.03142, 2019.
- [16] HAO Yi, WANG Nannan, Li Jie, et al. HSME: Hypersphere manifold embedding for visible thermal person re-identification [J]. Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 8385-8392.
- [17] WANG Guan'an, YANG Yang, ZHANG Tianzhu, et al. Cross-modality paired-images generation for RGB-infrared person re-identification[J]. Neural Networks,2020,128;294-304.
- [18] YE Mang, LAN Xiangyuan, LENG Qingming, et al. Cross-modality person re-identification via modality-aware collaborative ensemble learning [J]. IEEE Transactions on Image Processing, 2020, 29: 9387-9399.
- [19] HAO Yi, WANG Nannan, GAO Xinbo, et al. Dual-alignment feature embedding for cross-modality person re-identification [C]//Proceedings of the 27<sup>th</sup> ACM International Conference on Multimedia. Nice, France;ACM,2019; 57-65.
- [13] IFTIKHAR S, SHAHID S, HASSAN M U, et al. Assessment and prediction of restless leg syndrome (RLS) in patients with diabetes mellitus type II through artificial intelligence (AI) [J]. Pakistan Journal of Pharmaceutical Sciences, 2020, 33(5(Sp)): 2399-2403.
- [14] HAIDEGGER T. Autonomy for surgical robots: Concepts and paradigms [J]. IEEE Transactions on Medical Robotics and Bionics,2019,1(2):65-76.
- [15] 陶洋,邓行,杨飞跃,等.基于DTW距离度量的层次聚类算法[J].计算机工程与设计,2019,40(01):116-121.
- [16] 黄嘉欣.改进Apriori算法挖掘妊娠期糖尿病患者数据[J].现代计算机,2021,27(25):14-19.
- [17] 关小明,张意茗,范晓东.单孔腹腔镜技术的发展及展望[J].山东大学学报(医学版),2019,57(12):5-9.
- [18] 黄陈,陈庆奎,吴路平,等.协同感知腹腔镜手术恒压支持系统及其实现方法;中国,CN111243725A[P].2020-06-05.

(上接第16页)