

# 基于异质信息对齐和重排序的跨模态行人重识别方法

赵铁柱<sup>1</sup>, 梁校伦<sup>1</sup>, 杨秋鸿<sup>2</sup>, 张国斌<sup>3</sup>, 龚莨皓<sup>1</sup>

(1. 东莞理工学院 计算机科学与技术学院, 广东 东莞 523808; 2. 东莞城市学院 人工智能学院, 广东 东莞 523419;

3. 东莞理工学院 电子工程与智能化学院, 广东 东莞 523808)

**摘要:** 可见光图像和红外图像成像原理不同, 面向可见光和红外光的跨模态行人重识别面临较大的跨模态差异, 行人异质信息对齐和挖掘异常困难。为此, 提出基于异质信息对齐和重排序的跨模态行人重识别方法。在异质信息对齐方面, 提出一种新的异质局部信息对齐算法, 通过求取行人异质局部信息距离矩阵的最短路径, 实现同一行人异质局部信息的动态匹配, 解决行人异质信息对齐问题; 在重排序方面, 提出拓展  $k$  近邻重排序算法, 通过动态地拓展查询图像  $k$  近邻异质信息, 实现同一行人异质信息的融合, 解决行人异质信息挖掘困难问题。实验结果表明, 在 SYSU 数据集全场景查询模式上, 所提方法与 AGW 模型结合  $k$  近邻重排序算法相比, 在评价指标 mAP 和 Rank-1 上分别提升了 10.12% 和 8.6%。

**关键词:** 跨模态; 行人重识别; 异质信息对齐; 重排序; 深度学习

**中图分类号:** TN929.5

**文献标志码:** A

## Cross-modal person re-identification method based on heterogeneous information alignment and reranking

ZHAO Tiezhu<sup>1</sup>, LIANG Xiaolun<sup>1</sup>, YANG QiuHong<sup>2</sup>, ZHANG Guobin<sup>3</sup>, GONG Lianghao<sup>1</sup>

(1. School of Computer Science and Technology, Dongguan University of Technology, Dongguan 523808, China;

2. School of Artificial Intelligence, Dongguan City University, Dongguan 523419, China;

3. School of Electrical Engineering & Intelligentization, Dongguan University of Technology, Dongguan 523808, China)

**Abstract:** Cross-modal person re-identification between visible and infrared light images is a challenge due to the differences in imaging principles. The alignment and mining of heterogeneous pedestrian information become difficult. To address this, we proposed a cross-modal person re-identification method based on heterogeneous information alignment and reranking. The proposed method includes a new algorithm for heterogeneous local information alignment, which dynamically matches the same pedestrian heterogeneous local information by obtaining the shortest path of the distance matrix. We also proposed an extended  $k$ -nearest neighbor reranking algorithm, which realizes the same pedestrian heterogeneous information fusion and reduces the difficulty of information mining by dynamically expanding the heterogeneous information of the query image's  $k$  mutual nearest neighbors. The experimental results show that our method improves mAP and Rank-1 evaluation indexes by 10.12 % and 8.6 % respectively on the SYSU dataset compared to the AGW model combined with  $k$  mutual nearest neighbor reranking algorithm.

**Key words:** cross-modal; person re-identification; heterogeneous information alignment; reranking; deep learning

**收稿日期:** 2023-02-01

**基金项目:** 国家自然科学基金项目(61901115); 广东省普通高校重点领域专项(2021ZDZX3007); 东莞城市学院青年教师发展基金项目(2022QJY005Z)

**作者简介:** 赵铁柱(1983—), 男, 湖南双峰人, 副教授, 博士, 主要从事数据挖掘、大数据、分布式计算/存储等方面的研究。

梁校伦(1999—), 男, 广东东莞人, 硕士研究生, 主要从事计算机视觉研究, 本文通信作者。

E-mail: xiaolun\_liang@yeah.net

行人重识别被普遍看作是跨设备下的行人检索问题,是获取被检索行人位置分布信息的技术<sup>[1]</sup>。由于可见光监控设备在夜间弱光照环境下无法拍摄到清晰的图像,夜间监控作用有限,无法满足刑事侦查、智能监控与安保等应用需求<sup>[2]</sup>,光照条件下在可见光模式与红外模式之间自动转换的监控设备逐步被广泛部署。由于可见光和红外光两种模式下采集的图像存在模态上的差异,而且由于光照、视角、人的姿态以及背景的变化,同一个人在不同的监控视频下变化很大,因此,面向可见光和红外光的跨模态行人重识别(以下简称跨模态行人重识别)问题被提出<sup>[3]</sup>。

现有的跨模态行人重识别方法,大多通过单流卷积神经网络或者双流卷积神经网络,聚焦行人在跨模态图像下共享特征的学习<sup>[4-6]</sup>。如 Ye 等<sup>[5]</sup>提出双流网络,该网络由双流卷积网络和共享参数的全连接层组成,分别提取两个模态下行人的特异特征,通过共享全连接层参数,将不同模态下的特征映射到同一个特征空间,以提取同一行人在两个模态下的共享特征。然而,基于行人全局共享特征学习的方法对图像中噪声格外敏感<sup>[7]</sup>。如图 1(a)所示,行人特征受到视角变换、姿态变换、遮挡物等图像噪声影响。一些方法<sup>[8-9]</sup>通过划分行人不同部位特征方式,学习行人在跨模态下共享局部特征信息,以此减少图像噪声对行人特征的影响。如 Ye 等<sup>[8]</sup>引入局部注意力机制,设计了动态双向注意力聚合学习算法,通过将共享特征切块,并与相邻的特征进行局部注意力机制操作,能够让网络提取到有效的局部特征信息,识别特定行人。然而,这种基于行人局部特征学习的方法易受行人姿态变化的影响,且难以处理异质信息之间的差异。如图 1(b)所示,图像匹配过程中存在行人局部特征对齐混淆的问题,即行人红外图像头部的局部特征错误匹配到行人可见光肩部特征。此外,在单模态行人重识别中,重排序算法通过挖掘图像样本的分布信息,能够显著改善行人检索结果,提高模型性能<sup>[10-13]</sup>。如 Zhong 等<sup>[13]</sup>提出  $k$  近邻重排序算法(K-reciprocal Reranking, K-Re),通过挖掘两张相似图像之间存在互为  $k$  近邻信息,计算图像之间的杰卡德距离,融合图像的欧氏距离和杰卡德距离,生成新的排序列表。然而,如图 1(c)所示,由于可见光和红外光之间模态的差异往往大于模态内的差异,重排序算法往往只挖掘到同模态近邻信息,难以挖掘跨模态下行人图像的近邻信息。

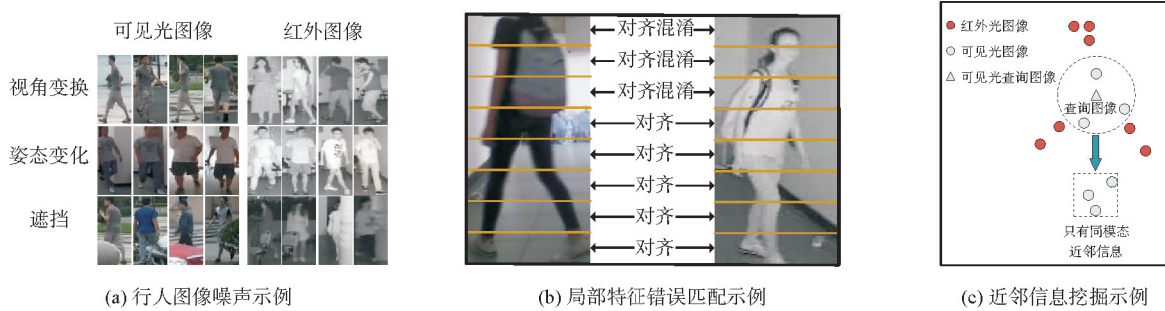


图 1 跨模态行人重识别难点示意图

Fig. 1 Schematic diagram of illustrating the challenges of cross-modal person re-identification

为了解决行人姿态变化导致的行人异质局部信息对齐混淆问题,受文献[13-16]启发,提出异质局部信息对齐算法(heterogeneous local information alignment algorithm,HLAA),通过将行人姿态对齐问题转化成人特征序列对齐问题,解决行人局部特征对齐混乱的问题。同时,为了解决跨模态近邻信息难以挖掘的问题,提出拓展  $k$  近邻重排序算法(extended K-nearest neighbor reranking, EK-Re),通过动态拓展查询图片的近邻信息,提高跨模态行人重识别方法的性能。此外,本研究还采用残差收缩模块(residual shrinkage building unit,RSBU)提取行人特征,减少行人特征受图像噪声的影响,增加行人特征的鲁棒性。

## 1 双流 ResNet50 网络模型

### 1.1 基线模型

本研究使用的基线模型如图 2 所示,包括特征提取和度量学习两部分。在特征提取部分,双流 Res-Net50 作为主干特征提取网络,是通过修改标准的双流 ResNet50 网络得到。相较于标准的双流 ResNet50

网络,在以下几个方面进行了改进:①双流 ResNet50 网络的第 1 个卷积模块是参数独立的,用于提取行人的特异特征,其余卷积模块是参数共享的,用于提取行人不同模态之间的共享特征;②在双流 ResNet50 网络的第 2 个卷积模块的最后两个残差块和第 3 个卷积模块的最后三个残差块中,插入非局部注意机制模块,以获取感受野更大的行人特征;③将双流 ResNet50 网络的第 5 个卷积模块的最后一个下采样层替换成可调参数的广义平均池化(generalized-mean pooling,GEM)层,以提取更加精细的行人特征。其中,广义平均池化的计算过程可以表示为:

$$f_k = \left( \frac{1}{|u_k|} \sum_{x \in u_k} x^{p_k} \right)^{\frac{1}{p_k}}, k = 1, 2, \dots, K. \quad (1)$$

式中: $K$  表示通道数, $u_k$  是每个通道的特征值, $p_k$  是池化层的超参数。 $p_k$  可动态调整特征感受野的大小,当  $p_k$  趋近于无穷大时,广义平均池化退化成最大池化;当  $p_k = 1$  时,广义平均池化退化成平均池化。通过调节  $p_k$  的取值,可以提取具有判别性的细粒度特征。在基线模型的广义平均池化层中, $p_k$  取值为 3.0,实验结果表明,该取值使得基线模型在跨模态行人重识别任务中取得较好的性能。

在度量学习部分,采用身份损失( $L_{id}$ )与加权正则化三元组损失( $L_{wrt}$ )联合监督的方式对基线模型进行约束。其中,身份损失采用交叉熵函数,以行人身份为类别数,引导网络对行人身份进行分类,使网络能够学习到与模态无关的行人身份特征表示。加权正则化三元组损失( $L_{wrt}$ )则以图像三元组形式训练模型。如图 2 所示,加权正则化三元组损失由 3 张图片( $i, j, k$ )组成。其中,图片( $i, j$ )属于同一个行人身份,组成一对正样本对;图片( $i, k$ )属于不同行人身份,组成一对负样本对。通过优化正负样本对之间的距离,引导网络提取出更具有判别性和鲁棒性的行人特征。正负样本对之间距离计算式为:

$$L_{wrt} = \ln [1 + \exp(\omega_i^p d_{ij}^p - \omega_i^n d_{ik}^n)], \quad \omega_i^p = \frac{\exp(d_{ij}^p)}{\sum_{p \in P} \exp(d^p)}, \quad \omega_i^n = \frac{\exp(-d_{ik}^n)}{\sum_{n \in N} \exp(-d^n)}. \quad (2)$$

式中:( $i, j, k$ )表示每个批次内的难样本策略采样的三元组图片, $P$  表示正样本集, $N$  表示负样本集, $d_{ij}^p$  表示最大的正样本对距离, $d_{ik}^n$  表示最小的负样本对距离, $d^p$  表示各个正样本对之间的距离, $d^n$  表示各个负样本对之间的距离。

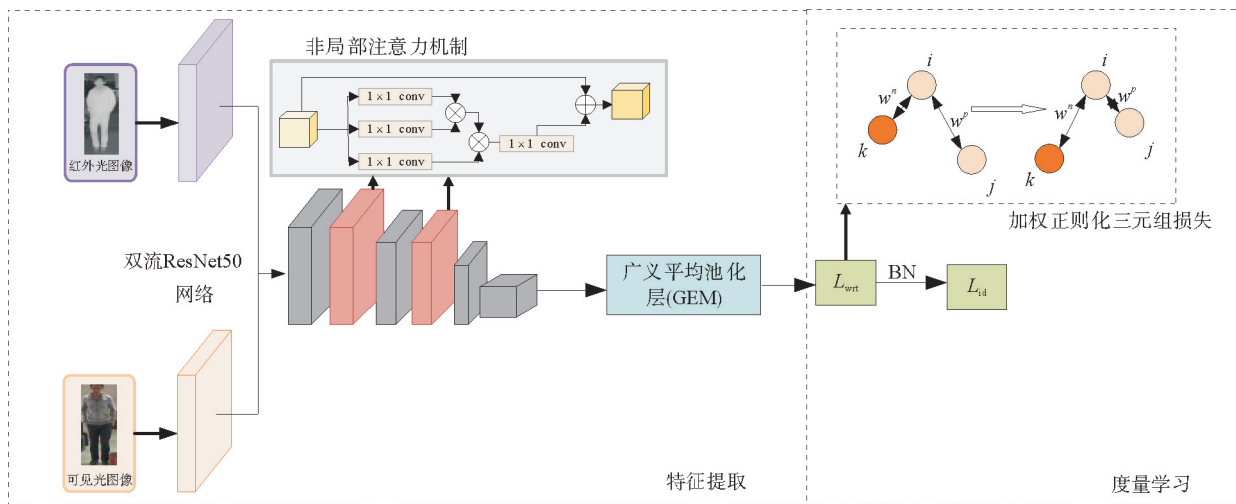


图 2 基线模型示意图

Fig. 2 Schematic diagram of the baseline model

## 1.2 残差收缩模块

现有卷积神经网络中,大多数采用深度残差网络残差块的设计,该设计克服了卷积神经网络加深而产生的学习率变低、无法有效提升准确率的问题。受到残差块启发,在双流 ResNet50 网络残差块中采用残差收缩模块(RSBU),能够减少冗余特征的提取,进一步提高提取特征的判别性。

残差收缩模块的整体结构如图3所示,从本质上讲,残差收缩模块是由残差模块、压缩与激励模块(squeeze-and-excitation module)和软阈值化函数的集成。首先,压缩与激励模块通过 Sigmoid 函数将全连接层的输出值缩放至 0~1 范围内;将缩放后的全连接层输出值乘以每个通道平均池化后的值,以自适应的方式学习到每个通道的软阈值  $\tau$ ;将经过软阈值化的特征值与原本的特征值相加,得到新的特征。

## 2 基于异质信息对齐和重排序的跨模态行人重识别

本研究在基线模型基础上,提出一种基于异质信息对齐和重排序的跨模态行人重识别方法,主要改进包括两个方面:①在度量学习阶段,提出异质局部信息对齐算法对行人姿态信息进行对齐,并在垂直方向和水平方向采用加权正则化三元组损失(记作  $L_{wrt}^v$  和  $L_{wrt}^h$ )对网络进行监督学习;②提出一种拓展  $k$  近邻重排序算法对初始排序列表进行重排序,通过精炼初始排序列表的结果,提高模型识别的准确率。整个跨模态行人重识别方法流程如图4所示。

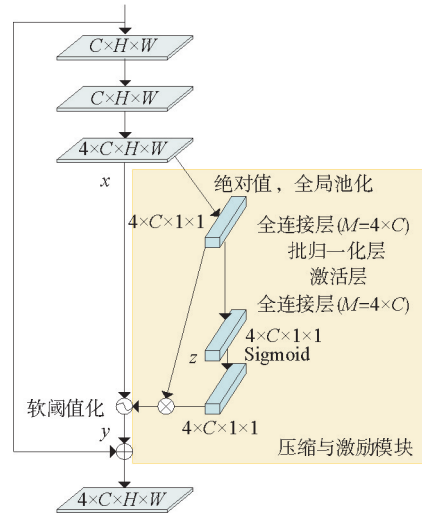


图3 残差收缩模块示意图

Fig. 3 Schematic diagram of the residual shrinkage module

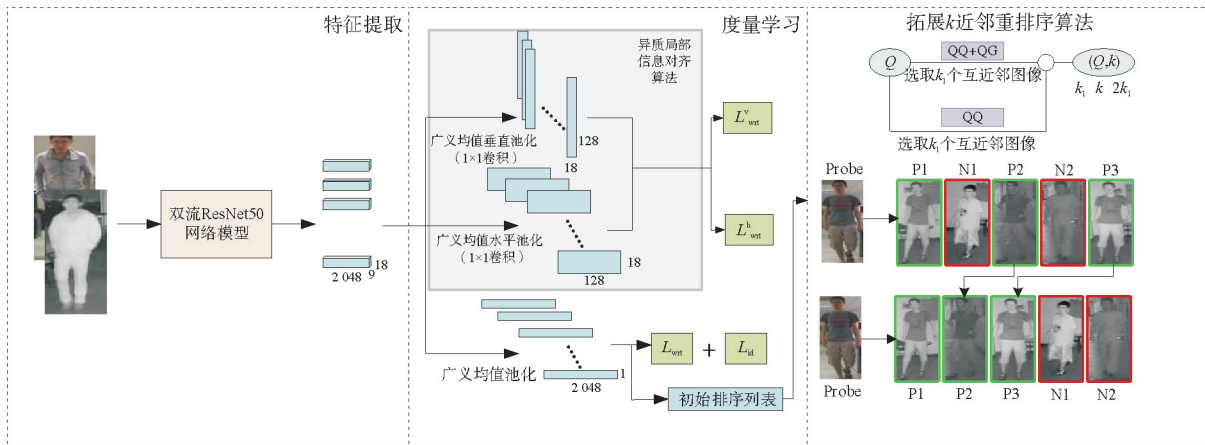


图4 跨模态行人重识别方法流程图

Fig. 4 Flow chart of cross-modal person re-identification method

### 2.1 异质局部信息对齐算法

异质局部信息对齐算法过程如图5所示,给定一张行人可见光图像和一张红外图像:①在水平和垂直方向上,将行人可见光图像和红外图像均匀划分成8份;②在水平和垂直方向上,提取行人可见光图像和红外图像的局部特征,并对局部特征进行广义均值池化和  $1 \times 1$  卷积操作;③在水平和垂直方向上,对行人可见光局部特征和红外局部特征进行相似度度量,形成距离矩阵  $D^k$  和  $D^v$ ;④采用动态规划算法,计算距离矩阵  $D^k$  和  $D^v$  坐标(0,0)到(7,7)最短距离,得出行人可见光图像和红外图像在水平和垂直方向上动态对齐后的相似度。

异质局部信息对齐算法主要对提取到的行人局部特征进行相似度度量。局部特征的相似度度量基于假设:对于同一个行人的两幅图像,第一幅图像行人身体部分的局部特征,更相似于另一幅图像对应身体部分的局部特征。将两幅图像的归一化后局部特征分别记作  $I_A = \{I_{A1}, I_{A2}, \dots, I_{AH}\}$  和  $I_B = \{I_{B1}, I_{B2}, \dots, I_{BH}\}$ , 则两幅图像的相似度计算过程为:首先,计算两幅图像两两局部特征的欧氏距离

$$d_{i,j} = \frac{e^{\|I_{A_i} - I_{B_j}\|_2} - 1}{e^{\|I_{A_i} - I_{B_j}\|_2} + 1}, \quad i, j \in 1, 2, \dots, H. \quad (3)$$

式中:  $d_{i,j}$  为第一幅图像中第  $i$  个垂直/水平部分的局部特征与第二幅图像中第  $j$  个垂直/水平部分的局部特征的距离。两幅图像中任意两个局部特征距离  $d_{i,j}$  组成距离矩阵  $D$ 。然后,采用动态规划算法计算两幅图像的最短距离:

$$S_{i,j} = \begin{cases} d_{i,j}, & i=1, j=1; \\ S_{i-1,j} + d_{i,j}, & i \neq 1, j=1; \\ S_{i,j-1} + d_{i,j}, & i=1, j \neq 1; \\ \min(S_{i-1,j}, S_{i,j-1}) + d_{i,j}, & i \neq 1, j \neq 1. \end{cases} \quad (4)$$

式中:  $S_{i,j}$  表示距离矩阵  $D$  从  $(1,1)$  到  $(i,j)$  的最短路径总距离。具体而言,两幅图像经过异质局部信息对齐算法处理,将每幅图像在水平与垂直方向等距划分为  $H$  个局部特征区域,那么两幅图像之间的相似度为  $S_{H,H}$ 。

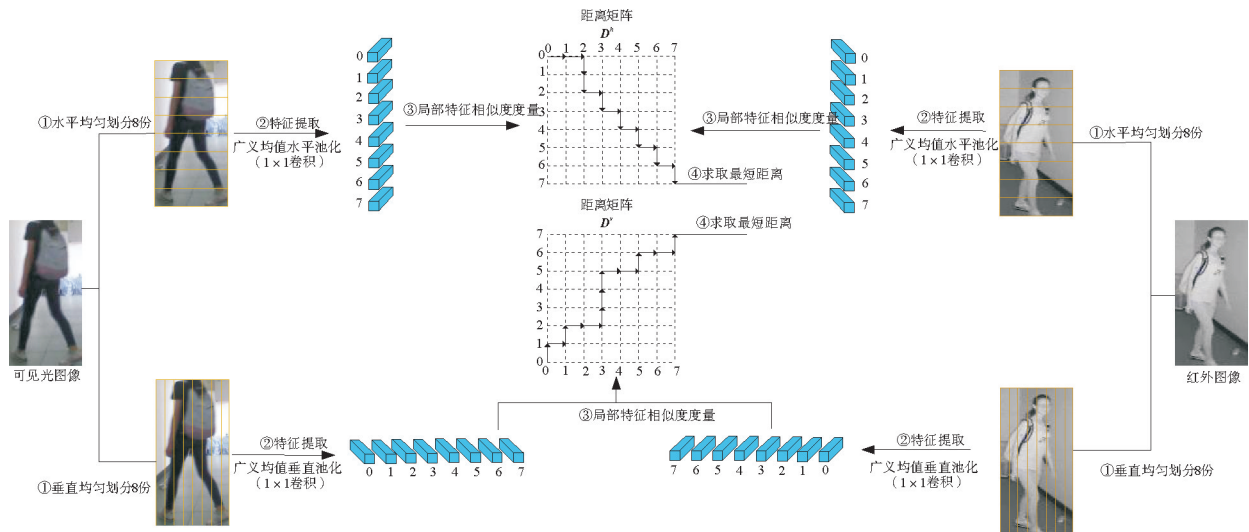
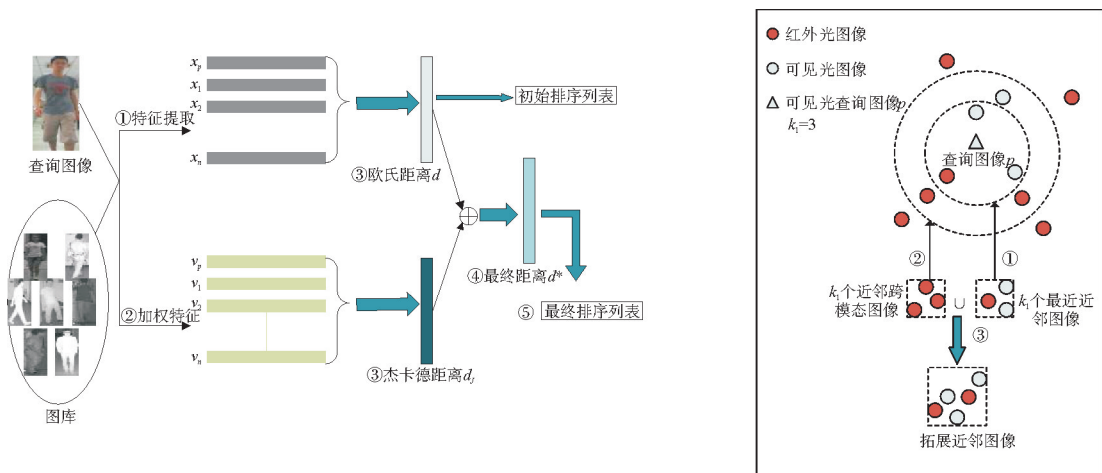


图 5 异质局部信息对齐算法过程示意图

Fig. 5 Schematic diagram of heterogeneous local information alignment algorithm

## 2.2 拓展 $k$ 近邻重排序算法

给定一张查询行人图像和一个行人图像库,拓展  $k$  近邻重排序算法过程如图 6(a) 所示。图 6(a) 中,①



(a) 拓展  $k$  近邻重排序算法过程图

(b)  $R(p, k')$  拓展近邻图像挖掘过程图

图 6 拓展  $k$  近邻重排序算法示意图

Fig. 6 Schematic diagram of the extended  $k$ -nearest neighbor reranking algorithm

为每个行人提取特征  $x$ ; ②为根据距离远近采用高斯核函数加权重度的行人特征  $v$ ; ③为根据  $x$  和  $v$ , 分别计算图库中每个行人图像与查询行人图像的欧氏距离  $d$  和杰卡德距离  $d_j$ ; ④为融合欧氏距离  $d$  和杰卡德距离  $d_j$  得到最终距离  $d^*$ ; ⑤为使用最终距离进行重排序, 生成新的行人图像排序列表。

拓展  $k$  近邻重排序算法基于假设: 如果两张图片 A、B 相似, 那么 B 应该会在 A 的前  $k$  个近邻里面, 反过来, A 也会在 B 的前  $k$  个近邻里面。但如果两张图 C、D 不相似, 即使 C 在 D 的前  $k$  个近邻里面, D 也不会出现在 C 的前  $k$  个近邻里面。基于该假设, 计算所有待查找行人图片和图库中的图片之间的杰卡德距离

$$d_j(p, g_i) = 1 - \frac{|R(p, k') \cap R(g_i, k')|}{|R(p, k') \cup R(g_i, k')|} \quad (5)$$

式中:  $R(p, k')$  表示图片  $p$  在全部图片集合中挖掘  $k_1$  个近邻图像, 以及在跨模态图像集合中挖掘  $k_1$  个近邻图像的并集,  $k_1 \leq k' \leq 2k_1$ 。简而言之,  $R(p, k')$  表示图片  $p$  的拓展近邻图像的集合, 其操作过程如图 6(b) 所示。同理,  $R(g_i, k')$  表示图库中图片  $g_i$  拓展近邻图像的集合。给定一张查询图片  $p$ , 首先, 从最近邻图像中选择前  $k_1$  个图像, 记为  $N_1(p, k_1)$ ; 然后, 从近邻跨模态图像中选择前  $k_1$  个图像, 记为  $N_2(p, k_1)$ ; 最后, 将  $N_1(p, k_1)$  和  $N_2(p, k_1)$  的图像集合取并集, 得到查询图片  $p$  的拓展近邻图像的集合, 即  $R(p, k') = \{N_1(p, k) \cup N_2(p, k)\}$ 。

在通过近邻信息求得行人图像的杰卡德距离之后, 由于初始的检索结果也包含了行人的一定信息, 最终距离融合欧氏距离和杰卡德距离, 进行相似度度量, 其最终距离表示为:

$$d^*(p, g_i) = (1 - \lambda)d_j(p, g_i) + \lambda d(p, g_i) \quad (6)$$

式中:  $\lambda$  控制杰卡德距离的权重,  $\lambda$  越大, 杰卡德距离的权重越小。

### 3 实验

#### 3.1 数据集及评估协议

从跨模态行人重识别领域两个公开的数据集 SYSU-MM01<sup>[4]</sup> 和 RegDB<sup>[17]</sup> 获取数据, 实验验证所提出方法的有效性。SYSU-MM01 数据集来源于 4 个可见光摄像头和 2 个红外摄像头拍摄的图像, 包含 491 个不同身份的行人, 训练集包括 22 258 张可见光图像和 11 909 张红外图像, 测试集包括剩余的 7 108 张可见光图像和 3 803 张红外图像。由于 SYSU-MM01 数据集采集的图像涵盖了室内和室外环境, 因此评测分为全场景查询(All Search)和室内查询(Indoor Search)两种模式。RegDB 数据集中的图像来源于 1 个可见光摄像头和 1 个红外摄像头, 分别采集了 412 个不同身份行人的 10 张可见光图像和红外光图像。评测标准分为可见光图像到红外图像查询(Visible to Thermal)和红外图像到可见光图像查询(Thermal to Visible)两种模式。

在跨模态行人重识别领域的实验评估中, 累积匹配特性(cumulative match characteristic, CMC)曲线、平均精度均值(mean average precision, mAP)和平均逆向样本惩罚(mean inverse negative penalty, mINP)是 3 种广泛采用的性能评估指标。其中, CMC 曲线通过展示不同排名(Rank- $n$ )水平上的识别正确率, 为模型性能提供直观评价。特别是, Rank-1 正确率(简称 R1)因直接反映模型在首次尝试中识别目标的能力而成为首选指标, 这对用户而言至关重要, 因为用户通常最关心模型在第一次尝试的准确性。为此, 本实验采用 R1、mAP 和 mINP 作为评价指标, 评估跨模态行人重识别模型的性能。

#### 3.2 实验设置

实验采用 Pytorch 深度学习框架, 运行机器软硬件配置为 Intel Xeon Silver 4210R CPU(2.40 GHz)、NVIDIA RTX 3080、显存 10 G、内存 80 G、64 位 Ubuntu 18.04 系统。

选取在 ImageNet 上预训练的双流 Resnet50 网络作为骨干网络, 输入图像尺寸统一调整为  $288 \times 144$ , 采用随机裁剪和随机翻转策略增强图像数据。为了统一图像的通道维度, 将红外图像在通道维度上复制 3 次, 与可见光图像一样成为三通道图像。采用随机梯度下降策略作为优化器, 动量参数设置为 0.9, 初始学习率设置为 0.1, 并在前 10 次迭代训练过程中采用学习率预热策略。迭代训练 120 次, 为了防止学习率过大导致振荡, 在迭代训练 20、50、80、100 次时分别设置学习率为 0.01、0.001、0.000 5 和 0.000 1。每次选取

4 个批次,每个批次中随机选择 4 个行人身份,在可见光和红外图像模态上各选取 4 张行人图像,每个批次共计选取 16 张行人可见光图像和 16 张红外图像。

### 3.3 对比实验

选取 SYSU-MM01 和 RegDB 数据集上最新的研究方法进行对比。其中,Zero-Pad<sup>[4]</sup>采用零填充扩展图像通道方法,BDTR<sup>[18]</sup>采用双向特征映射约束方法,HSME<sup>[19]</sup>采用超球流形嵌入方法,AlignGAN<sup>[20]</sup>采用像素和特征对齐方法,cm-SSFT<sup>[21]</sup>是利用特异特征与共享特征关系的方法,DDAG<sup>[8]</sup>是一种基于全局和局部双注意机制的方法,AGW<sup>[14]</sup>是一种新的基线模型,LoLeG-Net<sup>[22]</sup>是采用局部和全局特征融合的方法,K-Re 算法<sup>[13]</sup>是利用图像互惠邻关系的一种重排序算法。

表 1 展示了跨模态行人重识别方法的实验对比结果。表 1 中 AGW\* 为 AGW 模型测得的性能结果,EK-Re 为所提出的拓展  $k$  近邻重排序算法。由表 1 可以看出,本研究提出的跨模态行人重识别方法在评价指标 R1、mAP 和 mINP 均有提升。相对于 AGW\* + EK-Re 方法,所提方法在 SYSU-MM01 数据集的 All Search 模式在 3 个指标上都有明显提升,特别是 mINP 指标提升最明显,表明所提方法能够有效提高模型对于难样本的检索能力。同时,R1 和 mAP 指标在提高效果上表现得相差不大,表明所提方法既注重行人图像首次检索能力,又关注行人图像检索的整体性能。另外,在 RegDB 数据集的 Visible to Thermal 模式实验中,所提方法比 AGW\* + EK-Re 方法在 3 个指标上分别提高 3.45%、3.46% 和 3.9%,但比在 SYSU-MM01 数据集上的性能提升幅度小,这是因为 RegDB 数据集的训练集数量较小,且行人跨模态图片风格相似,导致模型未得到充分的训练。

表 1 在 SYSU-MM01 数据集和 RegDB 数据集的实验结果对比  
Table 1 Comparison of experimental results on SYSU-MM01 dataset and RegDB dataset %

方法	发表年份	SYSU-MM01						RegDB					
		All Search			Indoor Search			Visible to Thermal			Thermal to Visible		
		R1	mAP	mINP	R1	mAP	mINP	R1	mAP	mINP	R1	mAP	mINP
Zero-Pad <sup>[4]</sup>	2017	14.80	15.95	—	20.58	26.92	—	24.44	20.08	—	21.70	22.24	—
BDTR <sup>[18]</sup>	2018	27.32	27.32	—	37.92	41.96	—	33.56	32.76	—	32.92	31.96	—
HSME <sup>[19]</sup>	2019	20.68	23.12	—	—	—	—	50.85	47.00	—	50.15	46.16	—
AlignGAN <sup>[20]</sup>	2019	42.40	40.70	—	45.90	54.30	—	57.90	53.60	—	56.30	53.40	—
cm-SSFT <sup>[21]</sup>	2020	47.70	54.10	—	—	—	—	72.30	72.90	—	71.00	71.70	—
DDAG <sup>[8]</sup>	2020	54.75	53.02	39.62	61.02	67.98	62.61	69.34	63.46	49.24	68.06	61.80	48.62
AGW <sup>[14]</sup>	2021	47.50	47.65	35.30	54.17	62.97	59.23	70.05	66.37	50.19	70.49	65.90	51.24
LoLeG-Net <sup>[22]</sup>	2022	51.40	51.41	38.73	56.63	64.69	60.69	76.58	73.36	62.28	74.50	71.88	59.95
AGW*	—	54.75	52.77	38.80	59.93	67.02	62.97	74.71	68.61	54.48	71.58	66.11	51.51
AGW* + K-Re	—	66.53	62.82	48.61	75.33	78.20	74.46	72.67	74.57	69.15	70.88	71.79	66.94
AGW* + EK-Re	—	67.41	65.60	54.00	75.72	79.04	75.67	80.22	81.99	80.19	78.17	80.28	78.65
本研究方法	2024	<b>75.13</b>	<b>72.96</b>	<b>62.78</b>	<b>80.29</b>	<b>83.76</b>	<b>81.48</b>	<b>83.67</b>	<b>85.45</b>	<b>84.02</b>	<b>81.46</b>	<b>83.57</b>	<b>82.33</b>

### 3.4 消融实验

为了验证各个模块的有效性,在 SYSU-MM01 和 RegDB 数据集上开展消融实验,结果如表 2 所示。其中,Baseline 表示基线模型,RSBU 表示残差收缩模块,HLLA 表示提出的异质局部信息对齐算法,EK-Re 表示提出的拓展  $k$  近邻重排序算法。

由表 2 可见,在 SYSU-MM01 数据集中,单独加入 RSBU 模块和 HLLA 模块以及二者同时加入时,性能都有所提升,这与 RegDB 数据集上的性能结果表现不同。这是因为 SYSU-MM01 数据集的训练数据集较多,且行人在可见光模态和红外光模态视角、姿态等方面都具有差异,使神经网络能够被充分训练,从而提

高模型的性能。

由表2中RegDB实验结果可以看出,单独加入RSBU模块或HLAA模块,算法性能均有所下降,这是因为RegDB数据集中训练集图像数量相对较少,可见光模态和红外光模态下行人图像姿态是一一对齐的,而且行人之间的风格较为相似,训练集图像的这些特征导致神经网络没有得到充分训练。而同时加入RSBU模块和HLAA模块时算法性能有所上升,这是因为提出的HLAA模块能否解决行人局部信息的对齐问题,在很大程度上依赖于特征提取阶段的特征提取效果,而RSBU模块能够减少图像噪声对图像特征提取的影响,使得神经网络能够提取更具判别性的行人特征,在一定程度上有助于减少后续行人局部信息对齐的信息混淆问题,从而提高了算法性能。

表2 在SYSU-MM01数据集和RegDB数据集上的消融实验结果

Table 2 Ablation experiment results on SYSU-MM01 and RegDB dataset

%

方法	SYSU-MM01						RegDB					
	All Search			Indoor Search			Visible to Thermal			Thermal to Visible		
	R1	mAP	mINP	R1	mAP	mINP	R1	mAP	mINP	R1	mAP	mINP
Baseline	54.75	52.77	38.80	59.93	67.02	62.97	74.71	68.61	54.48	71.58	66.11	51.51
+RSBU	54.82	53.09	39.56	60.53	67.48	63.58	73.63	67.40	53.07	70.68	65.06	50.33
+HLAA	54.40	53.35	40.02	60.91	68.18	64.44	70.24	64.47	49.50	70.24	64.45	49.50
+EK-Re	67.41	65.60	54.00	75.72	79.04	75.67	80.22	81.99	80.19	78.17	80.28	78.65
+RSBU+HLAA	56.27	55.36	42.84	63.12	69.64	65.99	78.57	72.25	58.38	75.61	69.72	54.71
+RSBU+HLAA+EK-Re	75.13	72.96	62.78	80.29	83.76	81.48	83.67	85.45	84.02	81.46	83.57	82.33

消融实验结果显示,单独采用EK-Re重排序算法远没有结合RSBU、HLAA和EK-Re重排序算法效果好,这是因为重排序算法的性能增益来源于特征鲁棒性和判别性的提升<sup>[10]</sup>。加入RSBU和HLAA模块有助于神经网络提取到更具鲁棒性和判别性的行人特征,一定程度上提高了EK-Re重排序算法对跨模态行人重识别方法的性能。因此,同时加入RSBU、HLAA和EK-Re重排序算法时,本研究提出的跨模态行人重识别方法最优。

### 3.5 参数量和计算复杂度分析

表3是在SYSU-MM01数据集All Search模式下各个模型的计算开销。其中,参数量表示占用的显存,时间表示训练一个epoch所用时间和测试所用时间的总和。由表3可见,不加重排序算法的模型与基线模型相比,参数量增加12.31 MB,时间增加133 s,两者增加的主要原因是在残差块上引入了RSBU模块和HLAA模块。进一步加上EK-Re重排序算法后,参数量不变,时间增加了23 s,这是因为重排序算法是一种后处理过程且主要运行在CPU上,且提出的k近邻重排序算法时间复杂度为 $O(N\log_2 N)$ 。可以看出,与基线模型相比,本研究方法在参数量和计算复杂度上都有所增加,时间复杂度增加 $O(N\log_2 N)$ 。结合表2,本研究方法比基线模型+EK-Re重排序算法,在R1、mAP和mINP评价指标上分别增加7.72%、7.36%和8.78%。综合考虑模型性能和计算开销,本研究方法能够满足跨模态行人重识别任务的精度要求,同时可以在线下进行行人身份识别和预测。

### 3.6 可视化分析

图7(a)展示了异质局部信息对齐算法的可视化结果,黄线为局部特征划分,虚线为对齐结果。异质局部信息对齐算法中,距离矩阵最短路径采用动态规划求解,从而实现水平和垂直方向行人姿态的依次对齐。可见,异质局部信息对齐算法能够很好地依次对齐行人局部信息。

表3 不同方法的计算开销

Table 3 Computational overhead of the different methods

模型	时间/s	参数量/MB
Baseline	239	71.37
+RSBU+HLAA	372	83.68
+RSBU+HLAA+EK-Re	395	83.68



图 7(b)展示了拓展  $k$  近邻重排序算法的可视化结果,其中红色框表示错误匹配,绿色框表示正确匹配。未经过拓展的  $k$  近邻重排序算法中,前 10 个近邻图像中有 8 张可见光图像和 2 张红外图像,由于红外图像信息不足,导致排序列表中出现错误匹配结果。在初始排序列表基础上,经过拓展  $k$  近邻重排序算法改进,前 10 个查询对象图像中增加了 3 张红外图像,提供了更多的行人红外图像近邻信息,提高了行人身份匹配的准确性。



图 7 可视化结果示意图

Fig. 7 Illustration of visualization results

### 3.7 拓展 $k$ 近邻重排序算法参数分析

采用控制变量法在 SYSU-MM01 数据集上进行参数分析实验,测试拓展  $k$  近邻重排序算法在参数  $k_1$  和  $\lambda$  不同数值时的性能表现,获得参数的最优取值。

图 8 展示了参数  $k_1$  对跨模态行人重识别方法性能的影响。结果显示,随着参数  $k_1$  增大,性能先提高后下降,这是因为更大的  $k_1$  会导致互近邻图像中包含更多错误匹配的行人图像。 $k_1 = 30$  时,两种模式均取得最优的 mAP 和 R1 值。

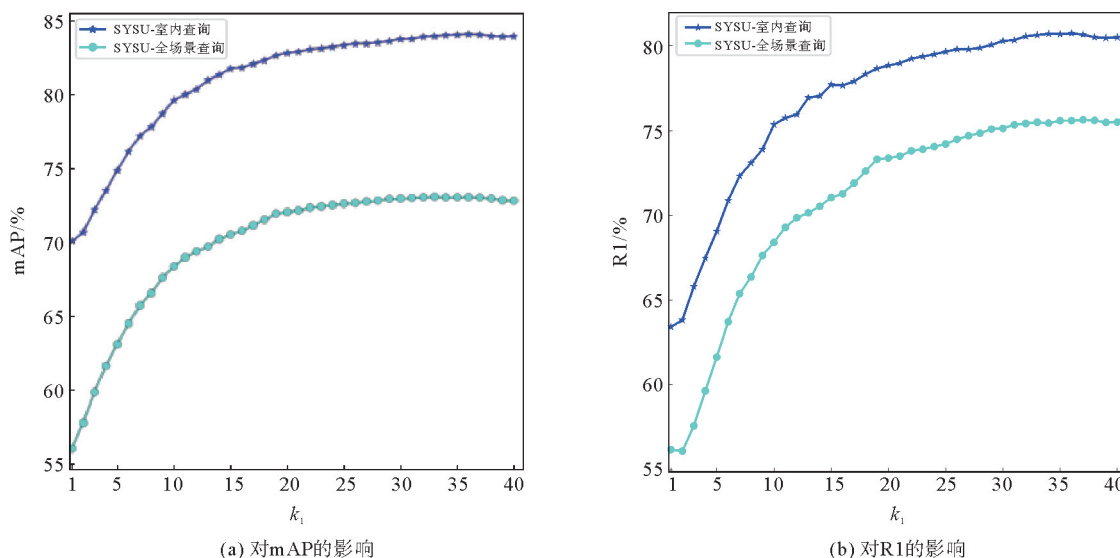


图 8 重排序算法参数  $k_1$  对跨模态行人重识别性能的影响

Fig. 8 Impact of re-ranking algorithm parameters  $k_1$  on cross-modal pedestrian re-identification performance

图 9 展示了参数  $\lambda$  对跨模态行人重识别性能的影响。 $\lambda$  控制欧氏距离和杰卡德距离在距离度量中的比重, $\lambda$  越小,杰卡德距离的权重越大。从图 9 可以看出,随着  $\lambda$  增大,跨模态行人重识别性能略微提高后下

降。这是因为欧氏距离更注重行人首次匹配的正确性,而杰卡德距离注重行人匹配的平均正确性,这两种不同的距离度量方法都对神经网络的学习起到监督作用。 $\lambda=0.1$ 时,两种模式取得最优的 mAP 和 R1 值。

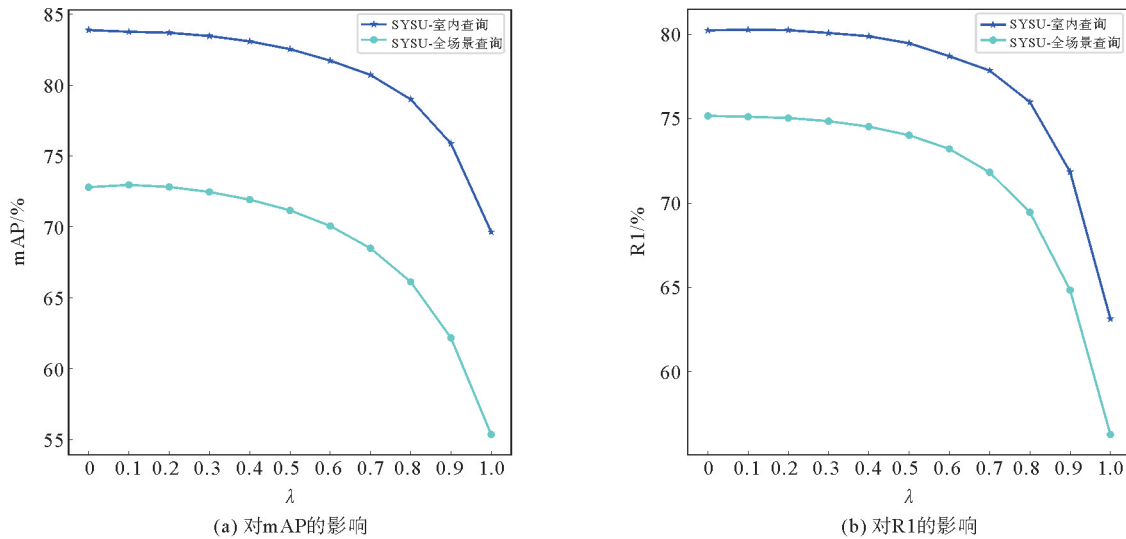


图9 重排序算法参数  $\lambda$  对跨模态行人重识别性能的影响

Fig. 9 Impact of reranking algorithm parameters  $\lambda$  on cross-modal pedestrian re-identification performance

## 4 结论

本研究提出一种基于异质信息对齐和重排序的跨模态行人重识别方法。在跨模态行人重识别模型中应用残差收缩模块减少冗余特征的提取,并提出异质局部信息对齐算法,通过求取行人异质局部信息距离矩阵的最短路径,能够较好地解决行人跨模态局部特征匹配混淆的问题;同时提出拓展  $k$  近邻重排序算法,通过动态地拓展查询图像  $k$  近邻异质信息,进一步提高了模型的性能。所提出方法的有效性在公开的 SYSU-MM01 数据集和 RegDB 数据集上得到很好的验证。下一步的研究将着重解决行人可见光图像和红外图像中存在行人遮挡和分辨率低的问题。

## 参考文献:

- [1] 罗浩,姜伟,范星,等. 基于深度学习的行人重识别研究进展[J]. 自动化学报,2019,45(11):2032-2049.  
LUO Hao,JIANG Wei,FAN Xing,et al. A survey on deep learning based person re-identification[J]. Acta Automatica Sinica,2019,45(11):2032-2049.
- [2] 叶钰,王正,梁超,等. 多源数据行人重识别研究综述[J]. 自动化学报,2020,46(9):1869-1884.  
YE Yu,WANG Zheng,LIANG Chao,et al. A survey on multi-source person re-identification[J]. Acta Automatica Sinica,2020,46(9):1869-1884.
- [3] 朱松豪,吕址函,宋杰. 基于特征学习的双路径红外-可见光行人重识别算法[J]. 山东科技大学学报(自然科学版),2022,41(5):82-90.  
ZHU Songhao,LÜ Zhihan,SONG Jie. Dual-path infrared-visible person re-identification algorithm based on feature learning [J]. Journal of Shandong University of Science and Technology(Natural Science),2022,41(5):82-90.
- [4] WU A C,ZHENG W S,YU H X,et al. RGB-infrared cross-modality person re-identification[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice,2017:5390-5399.
- [5] YE M,ZHENG W,LAN X Y,et al. Visible thermal person re-identification via dual-constrained top-ranking[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm,2018:1092-1099.
- [6] ZHANG S Z,YANG Y F,WANG P,et al. Attend to the difference:Cross-modality person re-identification via contrastive correlation[J]. IEEE Transactions on Image Processing,2019,14(8):8861-8872.

- [7] 李大伟,曾智勇. 基于动态双注意力机制的跨模态行人重识别模型[J]. 计算机应用, 2022, 42(10): 3200-3208.  
LI Dawei, ZENG Zhiyong. Cross-modal person re-identification model based on dynamic dual-attention mechanism[J]. Journal of Computer Applications, 2022, 42(10): 3200-3208.
- [8] YE M, SHEN J B, CRANDALL D J, et al. Dynamic dual-attentive aggregation learning for visible-infrared person re-identification[C]//Proceedings of the European Conference on Computer Vision, Newcastle, 2020: 229-247.
- [9] WU Q, DAI P Y, CHEN J, et al. Discover cross-modality nuances for visible-infrared person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, 2021: 4330-4339.
- [10] 郭若沛. 基于图像的行人再识别重排序算法研究[D]. 北京: 北京邮电大学, 2020: 37-46.  
GUO Ruopei. Research on reranking approaches for image-based person re-identification[D]. Beijing: Beijing University of Posts and Telecommunications, 2020: 37-46.
- [11] SHEN X H, LIN Z, BRANDT J, et al. Object retrieval and localization with spatially-constrained similarity measure and k-NN reranking[C]//Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, Piscataway, 2012: 3013-3020.
- [12] YE M, CHEN J, LENG Q M, et al. Coupled-view based ranking optimization person re-identification[C]//Proceedings of 21st International Conference on Multimedia Modeling, Sydney, 2015: 105-117.
- [13] ZHONG Z, ZHENG L, CAO D L, et al. Re-ranking person re-identification with k-reciprocal encoding[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, 2017: 1318-1327.
- [14] 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, 2021, 99(1): 1-25.
- [15] ZHAO M H, ZHONG S S, FU X Y, et al. Deep residual shrinkage networks for fault diagnosis[J]. IEEE Transactions on Industrial Informatics, 2020, 16(7): 4681-4690.
- [16] LUO H, JIANG W, ZHANG X, et al. Alignedreid++: Dynamically matching local information for person re-identification[J]. Pattern Recognition, 2019, 94(1): 53-61.
- [17] 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): 605-633.
- [18] YE M, LAN X Y, WANG Z, et al. Bi-directional center-constrained top-ranking for visible thermal person re-identification[J]. IEEE Transactions on Information Forensics and Security, 2020, 15(1): 407-419.
- [19] HAO Y, WANG N N, LI J, et al. Hsme: Hypersphere manifold embedding for visible thermal person re-identification[C]//Proceedings of the AAAI Conference on Artificial Intelligence, Hawaii, 2019: 8385-8392.
- [20] WANG G A, ZHANG T Z, CHENG J, 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, 2019: 3623-3632.
- [21] LU Y, WU Y, LIU B, et al. Cross-modality person re-identification with shared-specific feature transfer[C]//Proceedings of the Computer Vision and Pattern Recognition, Seattle, 2020: 13379-13389.
- [22] WANG J, JIANG K W, ZHANG T Q, et al. Visible-infrared person re-identification via global feature constraints led by local features[J]. Electronics, 2022, 11(17): 2645-2660.

(责任编辑:傅 游)