

# 基于局部异质协同双路网络的 跨模态行人重识别

郑爱华<sup>1,2</sup> 曾小强<sup>1</sup> 江波<sup>1,2</sup> 黄岩<sup>3</sup> 汤进<sup>1,2</sup>

**摘要** 针对现有跨模态行人重识别方法忽略行人的局部特征及模态间的相互协同的问题,文中提出基于局部异质协同双路网络的跨模态行人重识别方法.首先,通过双路网络提取不同模态的全局特征进行局部精细化,挖掘行人的结构化局部信息.然后,通过标签和预测信息建立跨模态局部信息之间的关联,进行协同自适应的跨模态融合,使不同模态的特征之间相互补充,获得富有判别力的特征.在 RegDB、SYSU-MM01 跨模态行人重识别数据集上的实验验证文中方法的有效性.

**关键词** 行人重识别, 跨模态, 局部特征, 协同融合, 卷积神经网络

**引用格式** 郑爱华,曾小强,江波,黄岩,汤进.基于局部异质协同双路网络的跨模态行人重识别.模式识别与人工智能,2020,33(10):867-878.

**DOI** 10.16451/j.cnki.issn1003-6059.202010001

中图法分类号 TP 391.41

## Cross-Modal Person Re-identification Based on Local Heterogeneous Collaborative Dual-Path Network

ZHENG Aihua<sup>1,2</sup>, ZENG Xiaoqiang<sup>1</sup>, JIANG Bo<sup>1,2</sup>, HUANG Yan<sup>3</sup>, TANG Jin<sup>1,2</sup>

**ABSTRACT** The coordinating fusion between modalities is ignored in the existing cross-modal person re-identification methods in the learning process. In this paper, a strategy for cross-modal person re-identification (Re-ID) based on local heterogeneous collaborative dual-path network is proposed. Firstly, the global features of each modality are extracted by the dual-path network for local refinement, and the structured local information of pedestrians is mined. Then, the local information of different modalities is correlated with the label and prediction information to achieve cooperative adaptive fusion and learn more discriminative features. The effectiveness of the proposed method is demonstrated through comprehensive

收稿日期:2020-07-14;录用日期:2020-10-29

Manuscript received July 14, 2020;

accepted October 29, 2020

新一代人工智能重大项目(No. 2018AAA0100400)、国家自然科学基金项目(No. 61976002)、安徽省重点研究与开发计划对外科技合作项目(No. 201904b11020037)、安徽省高等学校自然科学研究重点项目(No. KJ2019A0033)、模式识别国家重点实验室开放课题基金项目(No. 201900046)、安徽大学物质科学与信息技术研究院开放课题资助

Supported by Major Project for New Generation of AI (No. 2018AAA0100400), National Natural Science Foundation of China (No. 61976002), Key Project of Research and Development of Anhui Province (No. 201904b11020037), Natural Science Foundation of Anhui Higher Education Institutions of China (No. KJ2019A0033), Open Project of National Laboratory of Pattern Recognition (NLPR) (No. 201900046), and Open Fund

for Discipline Construction of Institute of Physical Science and Information Technology of Anhui University

本文责任编辑 兰旭光

Recommended by Associate Editor LAN Xuguang

1. 安徽大学 计算智能与信号处理教育部重点实验室 合肥 230601
2. 安徽省科学技术厅 安徽省工业图像处理与分析重点实验室 合肥 230039
3. 中国科学院自动化研究所 模式识别国家重点实验室 北京 100190

1. Key Laboratory of Intelligent Computing and Signal Processing, Ministry of Education, Anhui University, Hefei 230601
2. Key Laboratory of Industrial Image Processing and Analysis of Anhui Province, Science and Technology Department of Anhui Province, Hefei 230039
3. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190

experiments on two benchmarks RegDB and SYSU-MM01.

**Key Words** Person Re-identification, Cross-Modal, Local Feature, Collaborative Fusion, Convolutional Neural Network

**Citation** ZHENG A H, ZENG X Q, JIANG B, HUANG Y, TANG J. Cross-Modal Person Re-identification Based on Local Heterogeneous Collaborative Dual-Path Network. *Pattern Recognition and Artificial Intelligence*, 2020, 33(10): 867–878.

给定一幅特定摄像机拍摄的行人图像,行人重识别需要在其它相机拍摄的图像中检索该行人.由于行人重识别在实际的监控场景和智能安全等领域中具有广泛的应用前景<sup>[1-2]</sup>,因此一直受到学术界和工业界的关注.

传统的行人重识别方法利用手工提取的特征,在应对光照、视角变化上取得一定效果<sup>[3-9]</sup>,但距在实际场景中应用仍有差距.近年来,随着深度卷积神经网络的广泛应用,研究者们除提出一系列现实场景下的数据<sup>[10-13]</sup>外,同时提出深度神经网络特征提取方法.基于深度卷积神经网络的行人重识别模型包括:分类模型<sup>[14-21]</sup>、验证模型<sup>[10,22-25]</sup>和深度度量学习模型<sup>[18-21,26-31]</sup>.分类模型将行人看成独立的身份标签,通过多分类任务完成行人分类识别.验证模型判别输入的行人图像对的相似性得分,判断其是否为同一行人.深度度量学习模型使用度量损失函数作为代价函数,引导模型进行训练,如常用的三元组损失函数.上述方法可归纳为单模态(基于可见光图像的)行人重识别.

单模态行人重识别在弱光照和雾霾情况下难以满足现实应用的需求,而在实际监控场景中,很多监控摄像头在光照不足时可自动转换到红外模式,因此基于可见光和红外图像的跨模态行人重识别得到广泛关注<sup>[32-34]</sup>.

跨模态行人重识别是一项具有挑战性的工作,原因在于其不仅要面临单模态行人重识别中行人姿势、相机视角、背景等不同因素带来的问题,同时要解决可见光和红外两种不同模态的异质数据的融合难题.在此之前,研究者提出一系列的解决方案.Wu等<sup>[32]</sup>使用深度零填充网络,针对性地学习特定模态的结构信息.Ye等<sup>[33]</sup>采用基于度量学习的方法,让网络提取的不同模态下的特征表达尽可能相似.

在此基础上,学者们提出双向限定的排序损失,进一步增进不同模态特征之间的相似性<sup>[34-35]</sup>.Dai等<sup>[36]</sup>使用度量学习和生成对抗的策略学习模态之间共有的特征表达.Hao等<sup>[37]</sup>提出超球面流形嵌入模型,解决跨模态行人重识别问题.Wang等<sup>[38]</sup>采用

分步策略,将不同模态的图像空间映射至同一图像空间后,再嵌入特征空间中,学习模态共有的特征表达,实现跨模态行人重识别.Liu等<sup>[39]</sup>将中间网络层的特征融入行人特征,加强行人的特征表达,通过身份损失和双模态的三元组损失引导不同模态的特征融合.Wang等<sup>[40-41]</sup>将生成的内容对齐的跨模态图像对输入网络中学习共享特征,图像内容对齐使学习模态更容易共享特征.Lu等<sup>[42]</sup>分离模态独有信息,同时挖掘模态独有信息和共有信息间的联系,进行模态融合.Zhu等<sup>[43]</sup>提出异质中心损失函数,计算拉近模态间的样本中心特征,促进模态特征间的融合.

综上所述,将不同模态下的异质性数据高效融合并提升不同模态数据之间的相似性是解决跨模态行人重识别问题的一种重要方法.然而,现有的方法仅使用身份损失或三元组损失作为学习目标解决模态融合问题,其中身份损失只引导不同模态特征朝同个类中心聚集,三元组损失仅减小不同模态特征在空间上的距离,两者都只是单方面地从空间层面上拉近不同模态特征的距离,模态信息本身未发生变化,这导致在学习过程中忽视不同模态特征之间的协同交流及不同模态信息之间的互补,限制现有方法的有效性.

在跨模态行人重识别任务中,不同模态的行人信息尽管分布在不同的特征空间中,但依旧享有共同的信息,如身份标签、行人的语义内容等.协同学习<sup>[44-47]</sup>强调各个网络之间通过共有的信息引导进行交流通信,最终达到网络之间信息互补的目的.在跨模态行人重识别任务中,行人图像的可见光信息和红外信息享有共同的身份标签和视觉语义等信息,它们之间可相互交流进行协同学习.

除了面临不同模态数据如何进行融合的难题,跨模态行人重识别还需要解决模态内部行人姿势、相机视角等问题.针对这些问题,基于可见光图像的行人重识别使用特征精细化的方法<sup>[48-50]</sup>,取得较优效果.

基于上述分析,本文提出基于局部异质协同双

路网络的跨模态行人重识别方法,方法包含 3 步: 1) 双路提取. 使用预训练好的网络分别提取行人的可见光图像和红外图像的特征. 2) 特征精细化. 通过特征精细化挖掘不同模态特征的局部信息. 3) 协同融合. 将不同局部的模态特征进行协同融合. 在这一过程中,先将不同模态的局部特征通过共享参数的分类器映射至相同的特征子空间,减小不同模态特征间的距离. 再经过 Softmax 函数,得到各个局部特征的概率向量,通过交叉熵损失引导特征与标签之间建立关联. 进一步协同损失引导网络进行学习,促进每一个局部特征与另一个模态所有局部特征相互交流,在学习过程中达到信息互补的目的. 相比先前的跨模态行人重识别方法,本文方法不仅可学习更精细的模态特征信息,同时协同融合过程可进一步自适应融合跨模态异质性信息. 在现有的 2 个数

据集上的实验验证本文方法的有效性.

## 1 基于局部异质协同双路网络的跨模态行人重识别方法

本文提出基于局部异质协同双路网络的跨模态行人重识别方法,方法整体网络框架如图 1 所示. 首先采用双路网络两个分支提取可见光图像和红外图像特征. 然后将深度网络提取的行人特征进行特征精细化,即划分成多个局部细节特征. 最后进行协同融合,分为 2 步:1) 网络阶段的融合,表现在不同模态特征下相同位置的局部特征输入共享参数的分类器中. 2) 目标阶段的融合,表现在损失函数引导同一身份下不同模态的局部特征相互补充信息.

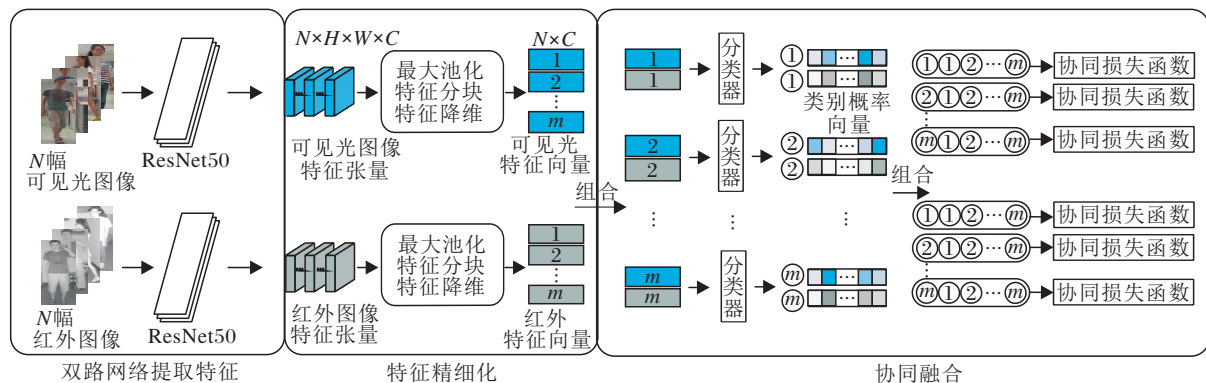


图 1 本文方法网络框图

Fig. 1 Network architecture of the proposed method

### 1.1 双路网络

本文使用预训练好的残差网络 (ResNet50)<sup>[51]</sup> 提取不同模态特征. 同样其它的深度网络也可用于提取特征,如密集连接的卷积网络 (DenseNet)<sup>[52]</sup>、VGG<sup>[53]</sup> 等. 本文使用的 ResNet50 网络<sup>[51]</sup> 和原始的残差网络 (Residual Network, ResNet) 有所区别: ResNet 整体分为 4 个模块,本文使用网络的前 3 个模块提取图像特征. 输出特征的通道数减少为原始网络的 1/2,既可节省计算量,又能避免因数据集过小网络过深而造成的过拟合.

### 1.2 特征精细化

受文献[48]启发,本文将单模态下的特征精细化技术扩展至跨模态行人重识别,在获取更鲁棒的特征的同时,为后续的协同融合提供更精细的特征. 首先,专注于学习行人不同位置的局部特征,再采用全局最大池化 (Global Max Pooling, GMP) 提取行

人局部的精细特征. 与单模态行人提取特征的方法<sup>[48]</sup>类似,特征精细化如图 1 中对应部分所示. 具体过程如下.

1) 特征提取. 将  $N$  对可见光和红外图像通过深度网络得到 2 个大小为  $N \times H \times W \times C$  的特征张量,其中,  $H$ 、 $W$  表示特征张量的高、宽,  $C$  表示通道数.

2) 特征池化. 将 1) 中提取的行人特征张量通过全局最大池化得到具有区分度的特征,特征张量的大小变成  $N \times H \times 1 \times C$ .

3) 特征细化. 将 2) 获取高度为  $H$  的特征平均分割成  $m$  份大小为  $N \times \frac{H}{m} \times 1 \times C$  的特征,不同位置的局部特征包含行人对应部位的局部信息.

### 1.3 协同融合

跨模态行人重识别中的关键问题是可见光图像和红外图像两种异质数据如何进行融合. 受协同学



习<sup>[4]</sup>的启发,本文提出协同融合策略,解决不同模态数据的融合问题.不同于文献[47]中使用多个分类器协同学习同一数据,本文提出的协同融合策略先将不同模态的特征投射至相同的特征空间,然后让一个模态中不同局部的特征向另一个模态的所有特征相互补充信息,使不同模态信息之间的相似度更高,提升重识别的效果.

本文的协同融合过程如图 1 中的对应部分所示.首先将不同模态相同位置的局部特征输入由 2 个全连接层组合的分类器中:第 1 个全连接层单元包含 512 个神经元,在其后添加 BN (Batch Normalization) 层和 ReLU 激活函数,第 2 个全连接层单元的神经元数目和训练样本中的类别数相同.注意到不同位置的局部特征使用不同的分类器.上述过程是网络阶段的融合,目的是为了让不同模态的特征通过一系列转换后映射至相同的特征空间.经过上述过程得到各模态不同位置的局部特征信息对应的类别向量.再进行学习目标阶段的融合,在此阶段学习目标不关注特征属于哪个局部位置,也不关注它们分布在哪个特征空间,只关注局部特征所属的身份,让同个身份下不同模态的局部特征交流融合.具体的学习目标函数的定义过程如下.

网络输出的类别概率向量进行 Softmax 转换,生成对应的类别概率向量,假设类别概率向量为

$$\mathbf{z}^{(r,h)} = \{z_1^{(r,h)}, z_2^{(r,h)}, \dots, z_k^{(r,h)}\},$$

其中,  $k$  为类别数,  $h$  为特征所属的局部位置,  $r$  为可见光模态  $v$  或红外模态  $t$ . 带有参数  $T$  的 Softmax 函数定义如下:

$$\delta_i(\mathbf{z}^{(r,h)}; T) = \frac{\exp\left(\frac{z_i^{(r,h)}}{T}\right)}{\sum_{j=1}^k \exp\left(\frac{z_j^{(r,h)}}{T}\right)},$$

其中,  $T$  值的大小表示概率分布的软化程度,  $T$  值越大,软化程度越高.

再将得到的类别概率向量分两步计算损失. 1) 分组. 一个模态下获得某个局部特征的类别概率向量与另一个模态的所有类别概率向量为的一组. 2) 计算协同损失. 以某一个可见光局部特征最终计算得到的类别概率向量和红外图像得到的所有类别概率向量进行组合为例,协同损失函数如下:

$$L_{v,h} = \beta J_{\text{hard}}(\mathbf{y}, \mathbf{z}^{(v,h)}) + (1 - \beta) J_{\text{soft}}(\mathbf{q}^{(v,h)}, \mathbf{z}^{(v,h)}).$$

其中:  $\mathbf{z}^{(v,h)}$  为当前的可见光图像获得的类别概率向量;  $v$  为可见光;  $\beta$  在  $(0, 1]$  之间,用于平衡  $J_{\text{hard}}^{[47]}$  和  $J_{\text{soft}}^{[47]}$  的比例;  $J_{\text{hard}}$  为常用的交叉熵函数,通过标签

引导网络学习方向,由对应的 One-Hot 编码和标签  $y$  得到,

$$J_{\text{hard}} = - \sum_{i=1}^k q_i \ln(\delta_i(\mathbf{z}^{(v,h)}; 1));$$

$J_{\text{soft}}$  通过拉近跨模态信息的距离学习模态间的协同互补,这得益于学习标签  $\mathbf{q}^{(v,h)}$  来源于另一个模态. 在可见光模态学习时,  $\mathbf{q}^{(v,h)}$  等于其它所有红外局部位置的特征对应的类别概率向量的平均值,

$$\mathbf{q}^{(v,h)} = \delta\left(\frac{1}{m} \sum_{i=1}^m \mathbf{z}^{(t,i)}; T\right),$$

$m$  为总共的局部特征数量,  $t$  为红外.

$\mathbf{q}^{(v,h)}$  综合考虑相同标签下同一模态信息的平均情况. 携带软标签和  $T$  值的损失函数<sup>[47]</sup> 如下:

$$J_{\text{soft}}(\mathbf{q}^{(v,h)}; \mathbf{z}^{(v,h)}) = - \sum_{j=1}^k q_j^{(v,h)} \ln(\delta_j(\mathbf{z}^{(v,h)}; T)).$$

$J_{\text{soft}}$  度量 2 个模态输出向量的距离. 最小化  $J_{\text{soft}}$  意味着将一个模态的平均信息传递至另一个模态的局部信息,同时可正则化整个网络.

同理,根据上述过程得出  $L_{t,h}$ ,最终不同模态的损失函数之和为

$$L_{\text{all}} = \sum_{h=1}^m (L_{v,h} + L_{t,h}),$$

其中,  $L_{v,h}$  为可见光图像的协同损失函数,  $L_{t,h}$  为红外图像的协同损失函数.

## 1.4 与现有工作的区别和联系

首先,受行人局部特征提取方法<sup>[48]</sup>启发,本文将单模态下的特征精细化技术扩展至跨模态行人重识别,在获取更鲁棒的特征的同时,为后续的协同融合提供更精细化的特征. 实验结果也验证通过特征精细化可进一步提升协同融合的效果.

另外,协同学习<sup>[47]</sup>广泛应用于分类任务中,主要思想是利用多个结构一致的分类器学习同一数据的不同知识,通过分类器间的信息协同获得更优的分类效果. 本文借鉴协同学习的思想,通过局部特征间的协同学习和不同模态间的协同学习实现跨模态行人重识别的协同融合. 本文方法充分发挥局部信息和跨模态信息的互补性,提高跨模态行人重识别的鲁棒性.

## 2 实验及结果分析

### 2.1 实验环境

为了验证本文方法的有效性,在跨模态行人重识别数据集 RegDB<sup>[54]</sup> 和 SYSU-MM01<sup>[32]</sup> 上和当前

主流的跨模态行人重识别的方法进行对比。

RegDB 数据集的红外图像由远红外相机拍摄,共有 412 个行人,单个行人包含 10 幅 RGB 图像和 10 幅红外图像.本文采用的评测协议和文献[33]~文献[35]一致:从 412 人中随机挑选 206 人用于训练,另外 206 人用于测试.在测试阶段,将其中一个模态的数据当作被检索图像,另一个模态的数据当作检索图像.最终结果均为 10 次实验的平均结果.

与 RegDB 数据集<sup>[54]</sup>不同,SYSU-MM01 数据集的红外图像由近红外相机拍摄,与 RegDB 数据集的图像风格差异较大.数据集包含来自于 6 个相机采集的 491 个行人图像,包括 4 个 RGB 相机和 2 个红外相机.训练集包含 395 个行人,共 22 258 幅 RGB 图像和 11 909 幅红外图像.测试集包含 96 个行人,其中 3 803 幅红外图像当作被检索图像,随机抽取的 301 幅 RGB 图像当作检索图像.由于随机选择图像当作检索图像的原因,在测试阶段采用 10 次随机实验的平均值作为最终结果.同时,遵循文献[32]的验证协议,本文分别采用单幅图像命中加全局搜索(Single-Shot All-Search)的评估模式评测方法.

为了评价方法的性能,本文使用 Rank1、Rank10、Rank20 和平均精度均值(Mean Average Precision, mAP)作为评估指标.给定一幅被检索的行人图像,Rank 排名反映搜索正确的匹配率,mAP 评估算法的平均检索性能.

在训练期间,批大小设置为 32,取样阶段从训练集中随机各取 1 幅 RGB 图像和 1 幅红外图像,本文仅使用图像随机水平翻转作为图像增强,图像初始大小与文献[48]~文献[50]一致,为  $384 \times 128$ .预训练好的 ResNet50 网络被用作主体框架,优化器使用随机梯度下降算法(Stochastic Gradient Descent, SGD),其中动量(Momentum)等于 0.9.预训练网络的初始学习率为 0.001,分类器的初始学习率为 0.01,经过训练迭代 30 轮以后,所有网络的学习率降至原来的 1/10.在协同融合的过程中,当数据集为 SYSU-MM01 时, $T = 2, \beta = 0.6$ .数据集为 RegDB 时, $T = 2, \beta = 0.5$ .代码的实现使用 pytorch 1.1.0 框架,总共训练 50 轮,训练过程使用 GTX1080Ti 显卡加速.2 个数据集的训练过程都遵从上述设定.

实验中选择的对比方法如下:单流网络算法(One-Stream Network, One-Stream)<sup>[32]</sup>,双流网络算法(Two-Stream Network, Two-Stream)<sup>[32]</sup>,零填充网络算法(Deep Zero-Padding, Zero-Padding)<sup>[32]</sup>,

双流卷积神经网络算法(Two-Stream CNN Network, TONE)<sup>[33]</sup>,分层跨模态度量学习算法(Hierarchical Cross-Modality Metric Learning, HCML)<sup>[33]</sup>,双向跨模态排序算法(Bi-directional Cross-Modality Top-Ranking, BCTR)<sup>[34]</sup>,双向双重限制的排序算法(Bi-directional Dual-Constrained Top-Ranking, BDTR)<sup>[34]</sup>,双向中心约束的排序算法(Bi-directional Center-Constrained Top-Ranking, eBDTR)<sup>[35]</sup>,跨模态生成对抗网络算法(Cross-Modality Generative Adversarial Network, cmGAN)<sup>[36]</sup>,超球面流形嵌入算法(HyperSphere Manifold Embedding, HSME)<sup>[37]</sup>,双重差异减小算法(Dual-Level Discrepancy Reduction Learning, D2RL)<sup>[38]</sup>,增强辨别性特征学习算法(Enhancing Discriminative Feature Learning, EDFL)<sup>[39]</sup>,对齐的生成对抗网络算法(Alignment Generative Adversarial Network, AlignGAN)<sup>[40]</sup>,数据集与实例的联合对齐算法(Joint Set-Level and Instance-Level Alignment ReID, JSIA-ReID)<sup>[41]</sup>,跨模态共享与特异特征变换算法(Cross-Modality Shared-Specific Feature Transfer Algorithm, cm-SSFT)<sup>[41]</sup>,双流局部特征网络算法(Two-Stream Local Feature Network, TSLFN)<sup>[43]</sup>,异质中心损失和双流局部特征网络算法(Two-Stream Local Feature Network + Hetero-Center Loss, TSLFN + HC)<sup>[43]</sup>.

## 2.2 在 RegDB 数据集上的实验结果

在 RegDB 数据集上,由于行人数量和图像数量具有较好的对称性,训练集和测试集可随机分成均等的两份,因此本文在 RegDB 数据集上进行 2 种不同搜索模式下的实验:可见光图像搜索红外图像(Visible to Thermal)模式、红外图像搜索可见光图像(Thermal to Visible)模式.

在 RegDB 数据集上,各方法在 2 种不同的匹配模式下的实验结果如表 1 所示.

由表 1 可看出,相比 cm-SSFT,在 Visible to Thermal 模式中,本文方法的 Rank1 值和 mAP 值达到 68.62% 和 66.27%,提高各 3.22% 和 0.67%.在 Thermal to Visible 模式中,本文方法 Rank1 值和 mAP 值达到 67.65% 和 64.49%,依旧占据优势.在两种模式中,Visible to Thermal 模式的结果在各项指标上略高,原因可能在于三通道的 RGB 图像比单通道的红外图像携带更丰富的信息,因此搜索结果更优.

表 1 各方法在 RegDB 数据集上的实验结果对比

Table 1 Experimental results comparison of different methods on RegDB dataset

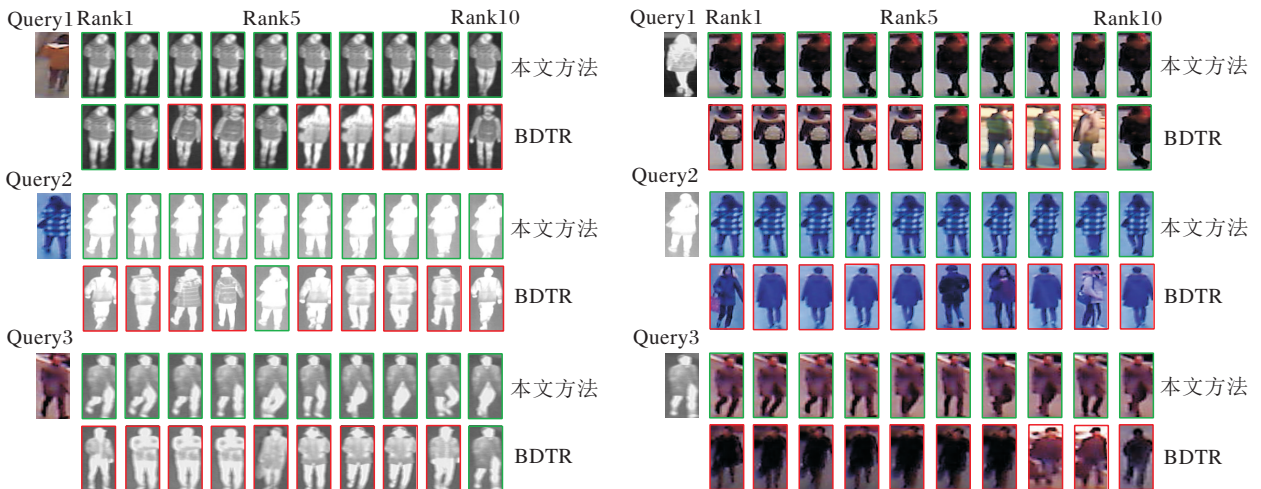
方法	Visible to Thermal				Thermal to Visible				%
	Rank1	Rank10	Rank20	mAP	Rank1	Rank10	Rank20	mAP	
One-Stream	13.11	32.98	42.51	14.02	-	-	-	-	
Two-Stream	12.43	30.36	40.96	13.42	-	-	-	-	
Zero-Padding	17.75	34.21	44.35	18.90	16.63	34.68	44.25	17.82	
TONE	16.87	34.03	44.10	14.92	13.86	30.08	40.05	16.98	
HCML	24.44	47.53	56.78	20.80	21.70	45.02	55.58	22.24	
BCTR	32.67	57.64	66.58	30.99	-	-	-	-	
BDTR	33.47	58.42	67.52	31.83	32.72	57.96	68.86	31.10	
eBDTR	34.62	58.96	68.72	33.46	34.21	58.74	68.64	32.49	
HSME	50.85	73.36	81.66	47.00	50.15	72.40	81.07	46.16	
D2RL	43.40	66.10	76.30	44.10	-	-	-	-	
EDFL	52.58	72.10	81.47	52.98	51.89	72.09	81.04	52.13	
AlignGAN	57.90	-	-	53.60	56.30	-	-	53.40	
JSIA-ReID	48.50	-	-	49.30	48.10	-	-	48.90	
cm-SSFT	65.40	-	-	65.60	63.80	-	-	64.20	
本文方法	68.62	84.64	90.35	66.27	67.65	84.08	89.68	64.49	

本文方法效果最优的原因如下:

1) 本文方法学习不同位置的行人局部信息, 获取不同模态下更精细、更具有区分度的特征, 此种方式在单模态行人重识别任务中被证明是有效的;

2) 本文的协同融合策略考虑不同模态之间的局部信息互补, 不仅在空间层面上缩小模态之间的距离, 同时也让模态信息之间相互转换, 最终达到更优的特征融合效果。

为了更好地展示本文方法的有效性, 将 RegDB 数据集上检索排序结果的前 10 名与 BDTR 进行对比, 结果分别包含 Visible to Thermal 和 Thermal to Visible 两种模式, 具体如图 2 所示, 图中绿色框表示匹配正确, 红色框表示匹配错误。由图可看出, 相比 BDTR, 无论被检索图像是可见光图像还是红外图像, 本文方法在 2 种搜索模式下的匹配成功率都较高。



(a) Visible to thermal

(b) Thermal to visible

图 2 本文方法和 BDTR 在 RegDB 数据集上的重识别结果

Fig. 2 Re-identification results of the proposed method and BDTR on RegDB dataset



当被检索的图像是可见光图像时,本文方法可清晰分辨信息量少的红外图像,如图 2(a) 所示. BDTR 难以完整地识别正确的结果,而本文方法可全部正确搜索不同行人对应不同姿势的红外图像,这验证本文方法可提取鲁棒的特征.

当被检索图像是红外图像时,本文方法依旧可通过较少的行人结构信息检索对应的可见光图像. 如图 2(b) 中 Query 2 所示,本文方法检索富含颜色信息(该行人身穿花纹格子衫)的行人可见光图像,而 BDTR 检索姿势相似而身份不同的行人图像. 这进一步说明本文方法对同一身份的行人在不同模态之间建立的关系更牢固.

### 2.3 在 SYSU-MM01 数据集上的实验结果

SYSU-MM01 数据集为一个大型的跨模态行人重识别数据集,红外图像由近红外相机拍摄,符合大多数的实际场景条件. 与文献[32]的评测协议保持一致,在此数据集上本文使用行人的红外图像检索可见光图像.

各方法在 SYSU-MM01 数据集上对比结果如表 2 所示.

表 2 各方法在 SYSU-MM01 数据集上的实验结果对比

Table 2 Experimental results comparison of different methods on SYSU-MM01 dataset

方法	Rank1	Rank10	mAP
One-Stream	12.04	49.68	13.67
Two-Stream	11.65	47.99	12.85
Zero-Padding	14.80	54.12	15.59
TONE	12.52	50.72	14.42
HCML	14.32	53.16	16.16
BDTR	17.01	55.43	19.66
eBDTR	27.82	67.34	28.42
cmGAN	26.97	67.51	27.80
HSME	20.68	62.74	23.12
D2RL	28.90	70.60	29.20
EDFL	36.94	84.52	40.77
AlignGAN	42.40	85.00	40.70
JSIA-ReID	38.10	80.70	36.90
cm-SSFT	47.7	-	54.10
TSLFN	46.78	86.13	46.13
TSLFN+HC	56.96	91.50	<b>54.95</b>
本文方法	50.12	89.64	50.08
本文方法+HC	<b>57.01</b>	<b>92.96</b>	53.80

相比使用度量学习的方法 BDTR、EDFL,及同时使用度量学习和对抗生成策略的 cmGAN、D2RL、AlignGAN 和 JSIA-ReID,本文方法具有更高的识别性

能,Rank1 值和 mAP 值分别提高 33.11% 和 30.42% (BDTR),13.18%和9.31% (EDFL),23.15% 和22.28% (cmGAN),21.22% 和 20.88% (D2RL),7.72% 和 9.38% (AlignGAN),12.02% 和 13.18% (JSIA-ReID). 同时,本文方法的 Rank1 值比 cm-SSFT 提高 3% 左右,也明显优于 TSLFN.

由于异质中心(Hetero-Center, HC)度量损失的引用,TSLFN + HC 与本文方法相比具有明显优势. 为公平起见,在本文方法上增加 HC 损失函数的结果进行验证. 实验结果如表 2 最后一行所示. 由表可看出,增加 HC 度量损失函数后,本文方法在各个指标上的效果和 TSLFN + HC 相当(Rank1 值略高, mAP 值略低). 结果再次证实协同学习的有效性,同时验证 HC 度量损失的优越性. 本文方法优于大部分方法的原因在于它不仅可获取不同模态下更精细的局部特征,从粒度层面上缩小不同模态特征之间的差异,同时也让不同模态的局部特征之间进行信息互补、相互转换,进一步缩小不同模态之间的差异,所以会有更优的识别效果.

将本文方法在 SYSU-MM01 数据集上的检索结果和 BDTR 进行对比. 在该数据集的测试阶段只包含 Thermal to Visible 模式,随机选取其中的 3 个示例对比,如图 3 所示,图中绿色框表示匹配正确,红色框表示匹配错误.

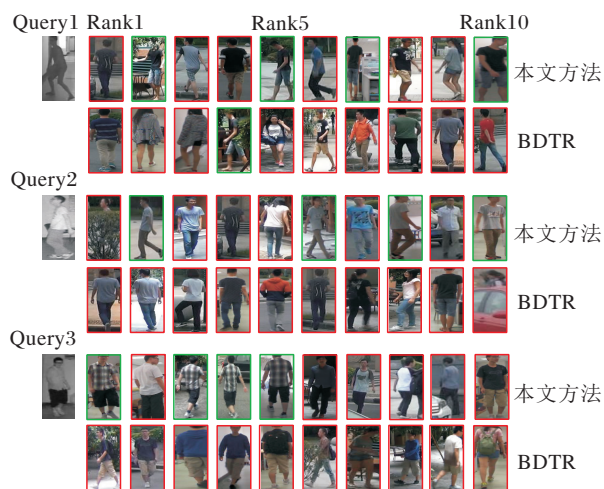


图 3 本文方法和 BDTR 在 SYSU-MM01 数据集上的重识别结果

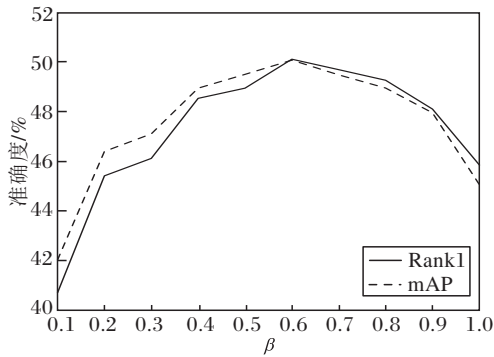
Fig. 3 Re-identification results of the proposed method and BDTR on SYSU-MM01 dataset

由图 3 可看出,整体上 SYSU-MM01 数据集因为行人姿势、光照和背景杂斑等因素影响较大,挑战难度更大. 由于跨模态异质图像中行人外观及跨相机

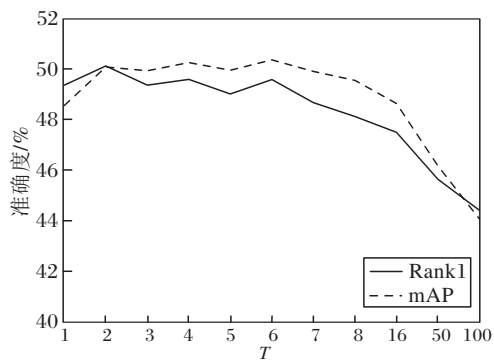
行人姿态的差异, BDTR 很难查询到正确结果, 而本文方法可较好地克服这些挑战, 成功匹配正确图像. 然而, 本文方法在光线较暗(如 Query1 的 Rank1 和 Rank4) 及有遮挡的场景(如 Query2 的 Rank1), 仍无法查询到正确结果. 同时, 对于具有类似姿势的不同行人(如 Query3 的 Rank6 和 Rank10), 本文方法也表现一般. 由此可见, 相比单模态行人重识别任务而言, 跨模态行人重识别存在更严峻的挑战.

## 2.4 超参数分析

为了软化网络输出的概率分布和平衡协同损失与交叉熵损失之间的学习过程, 本文使用超参数  $\beta$  和  $T$  增强最终的学习效果. 其中,  $\beta$  可以平衡损失  $J_{\text{hard}}$  和  $J_{\text{soft}}$ ,  $T$  可以软化输出的概率分布. 在 SYSU-MM01 数据集上重识别检索性能指标 Rank1 和 mAP 随超参数  $\beta$  和  $T$  的变化情况如图 4 所示.



(a)  $\beta$



(b)  $T$

图 4 在 SYSU-MM01 数据集上超参数  $\beta$  和  $T$  对模型性能的影响

Fig. 4 Influence of hyperparameter  $\beta$  and  $T$  on Rank1 and mAP on SYSU-MM01 dataset

在分析  $\beta$  的影响时,  $T = 2$ ; 在分析  $T$  时,  $\beta = 0.6$ . 图 4 结果表明, 当  $\beta = 0.6$  时, 效果最优, 此时 Rank1

值为 50.12%, mAP 值为 50.08%. 当  $T$  值为较小值 (2, 3, 4, 5, 6) 时, Rank1 值和 mAP 值相差在 1% 以内, 说明  $T$  值在小范围内波动对性能的影响不大. 而当  $T$  值大于 6 时, Rank1 值和 mAP 值随着  $T$  值增大而逐渐减小. 本文最终设定  $T = 2$ .

## 2.5 消融实验

为了验证本文提出的协同融合方法的关键部分的有效性, 在 SYSU-MM01 数据集上对不同的网络融合方式、学习目标阶段不同的融合损失函数及不同的局部特征数量进行分析.

首先分析不同的网络融合方式对不同数量的局部特征的影响. 分别得到行人的可见光模态的  $m$  个局部特征后, 局部特征的网络融合策略的选择是本文方法的重点. 因此本文设计 4 种不同的网络融合方式进行验证. 4 种融合方式的具体结构如图 5 所示, 主要区别在于不同模态的特征共享全连接层参数的方式: 方式 1 中所有局部位置的特征共享同一个分类器的参数; 方式 2 中不同模态下相同位置的局部特征共享同一个分类器; 方式 3 中一个模态下各局部的特征与另一个模态所有局部的特征共享分类器; 方式 4 中各个模态下每个局部特征使用单独的分类器.

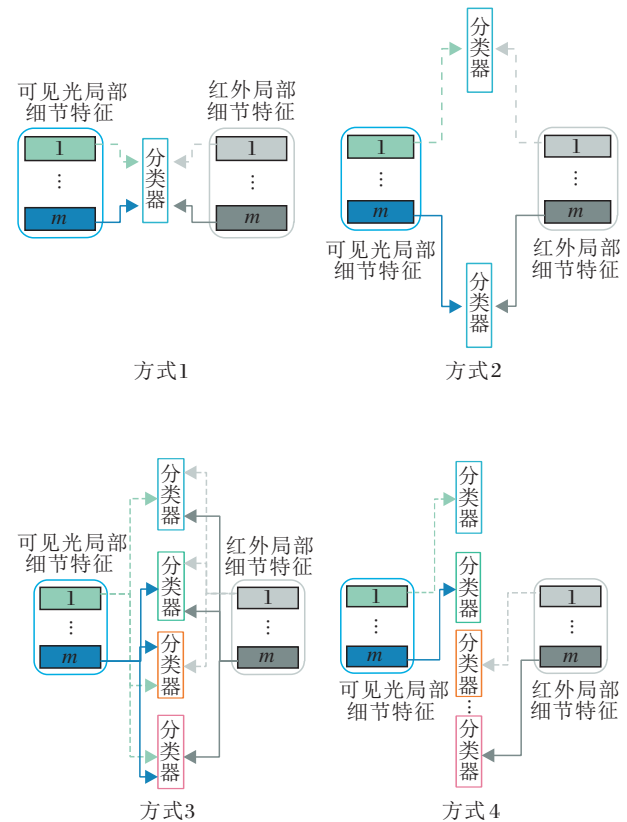


图 5 四种不同的网络融合方式

Fig. 5 Architectures of 4 kinds of network fusion



在实验中,将不同模态下的行人全局特征划分成表示不同位置的局部特征进行实验.图 6 为不同的局部特征数量使用 4 种不同的网络融合方式在 SYSU-MM01 数据集上的 Rank1 值的结果对比.由图可见,第 2 种网络融合方式的效果最佳,原因在于此种方式分类器的职责分工明确,不同的分类器融合对应的跨模态局部特征,因此方式 2 中对齐融合的方式结果最优.第 1 种融合方式效果次佳,原因可能在于一个分类器难以支撑多个局部特征进行融合,因此重识别效果会稍差.第 3 种网络融合方式使得各个全连接层的学习过程中混杂各个局部特征的信息,为其它局部的特征学习过程中引入过多的噪音,因此几乎没有效果.第 4 种融合方式让不同模态的各局部位置的特征在分离的参数空间中学习,达不到融合的目的.

根据上述实验结果,本文采用第 2 种方式作为局部特征的网络融合策略.

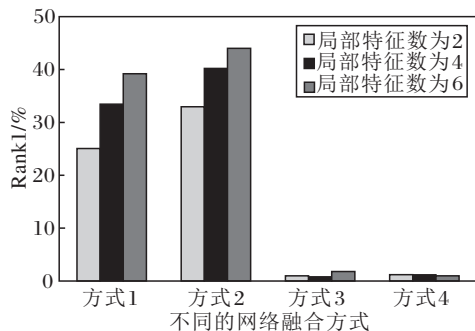


图 6 局部特征数不同时 4 种网络融合方式的 Rank1 值对比

Fig. 6 Rank1 values comparison of different local features using 4 kinds of network fusion

再分析不同的损失融合方式在不同数量的局部特征上的影响.为了验证本文使用的协同融合损失的有效性,将其与通用的交叉熵损失(普通融合)进行对比.两者都使用身份标签进行引导,不同之处在于协同融合的损失考虑不同分类器网络之间进行交互学习,而交叉熵损失只引导单个分类器学习.两种融合损失基于不同数量的局部特征的对比结果如图 7 所示.

由图 7 可看出,总体而言,本文采用的协同融合损失在不同数量的局部特征下均优于通用的交叉熵融合,原因在于,相比交叉熵损失函数,协同损失更好地学习各个局部特征间的信息互补,最终学习的特征更有辨识力.当局部特征的数量在一定范围(普通融合数量为 8,协同融合数量为 6)内,匹配结

果随局部特征数量的增加而提升,说明将特征分成一定数量的局部特征可更有效地挖掘各个模态下的特征信息,从而为后期的协同融合阶段提供更好的特征表达.当局部特征数量(局部特征数量为 1 时,此时只有全局特征)超过一定范围时,无论是通用的交叉熵损失还是协同融合损失,匹配结果趋于相对稳定而后下降的情况.原因在于过度地将特征细化,导致深度网络难以学习更有区分度的特征.同时随着分类器的数目增多,整个网络的训练参数增加使网络训练难度增加.局部特征的数量为 6 时的协同融合效果最佳,因此本文实验的局部特征数量固定为 6.

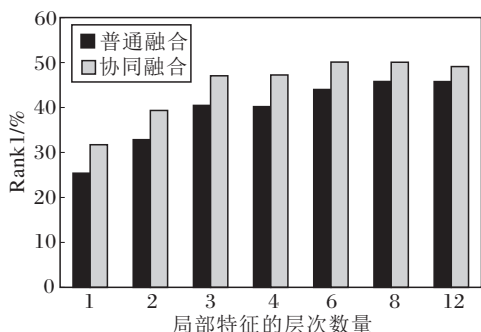


图 7 局部特征的层次数量不同时 2 种融合损失的 Rank1 值对比

Fig. 7 Rank1 values comparison of 2 kinds of fusion with different numbers of local features

### 3 结束语

本文提出基于局部异质协同双路网络的跨模态行人重识别方法,包括特征精细化和协同融合两个过程.特征精细化过程从行人不同位置的局部信息中获取更精细的特征,协同融合过程让不同模态下同一行人中蕴含不同身体局部信息的特征之间互补信息.实验结果显示本文方法效果较优,充分验证本文方法的有效性.在跨模态行人重识别中,提取不同模态的图像特征依旧是研究的重点.凭借使用预训练网络的强大性能,在单模态行人重识别任务中如何提取可见光图像的特征已经被广泛研究,并且在各个数据集上已取得较优结果.红外图像和可见光图像两者在本质上结构不同,如何更有效地提取红外图像特征需要进一步研究.另外,不同模态特征如何融合一直是跨模态匹配的关键问题,找到更优的融合方法应对跨模态行人重识别任务依旧需要进一步研究.

## 参 考 文 献

- [1] LAN X Y, MA A J, YUEN P C, *et al.* Joint Sparse Representation and Robust Feature-Level Fusion for Multi-cue Visual Tracking. *IEEE Transactions on Image Processing*, 2015, 24 (12): 5826–5841.
- [2] CHANG X J, HUANG P Y, SHEN Y D, *et al.* RCAA: Relational Context-Aware Agents for Person Search // *Proc of the European Conference on Computer Vision*. Berlin, Germany: Springer, 2018: 86–102.
- [3] FARENZENA M, BAZZANI L, PERINA A, *et al.* Person Re-identification by Symmetry-Driven Accumulation of Local Features // *Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2010: 2360–2367.
- [4] GRAY D, TAO H. Viewpoint Invariant Pedestrian Recognition with an Ensemble of Localized Features // *Proc of the European Conference on Computer Vision*. Berlin, Germany: Springer, 2008: 262–275.
- [5] KVIATKOVSKY I, ADAM A, RIVLIN E. Color Invariants for Person Re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(7): 1622–1634.
- [6] LIAO S C, HU Y, ZHU X Y, *et al.* Person Re-identification by Local Maximal Occurrence Representation and Metric Learning // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2015: 2197–2206.
- [7] MA B P, SU Y, JURIE F. Local Descriptors Encoded by Fisher Vectors for Person Re-identification // *Proc of the European Conference on Computer Vision*. Berlin, Germany: Springer, 2012: 413–422.
- [8] MATSUKAWA T, OKABE T, SUZUKI E, *et al.* Hierarchical Gaussian Descriptor for Person Re-identification // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2016: 1363–1372.
- [9] ZHAO R, OUYANG W L, WANG X G. Learning Mid-level Filters for Person Re-identification // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2014: 144–151.
- [10] LI W, ZHAO R, XIAO T, *et al.* DeepReID: Deep Filter Pairing Neural Network for Person Re-identification // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2014: 152–159.
- [11] RISTANI E, SOLERA F, ZOU R, *et al.* Performance Measures and a Data Set for Multi-target, Multi-camera Tracking // *Proc of the European Conference on Computer Vision*. Berlin, Germany: Springer, 2016: 17–35.
- [12] ZHENG L, BIE Z, SUN Y F, *et al.* MARS: A Video Benchmark for Large-Scale Person Re-identification // *Proc of the European Conference on Computer Vision*. Berlin, Germany: Springer, 2016: 868–884.
- [13] ZHENG L, SHEN L Y, TIAN L, *et al.* Scalable Person Re-identification: A Benchmark // *Proc of the IEEE International Conference on Computer Vision*. Washington, USA: IEEE, 2015: 1116–1124.
- [14] SUN Y F, ZHENG L, DENG W J, *et al.* SVDNet for Pedestrian Retrieval // *Proc of the IEEE International Conference on Computer Vision*. Washington, USA: IEEE, 2017: 3820–3828.
- [15] XIAO T, LI H S, OUYANG W L, *et al.* Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2016: 1249–1258.
- [16] ZHENG L, YANG Y, HAUPTMANN A G. Person Re-identification: Past, Present and Future [ C/OL ]. [ 2020 – 07 – 12 ]. <https://arxiv.org/pdf/1610.02984.pdf>.
- [17] HOU R B, MA B P, CHANG H, *et al.* Interaction-and-Aggregation Network for Person Re-identification // *Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2019: 9317–9326.
- [18] ZHANG Z Z, LAN C L, ZENG W J, *et al.* Densely Semantically Aligned Person Re-identification // *Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2019: 667–676.
- [19] SUN Y F, XU Q, LI Y L, *et al.* Perceive Where to Focus: Learning Visibility-Aware Part-Level Features for Partial Person Re-identification // *Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2019: 393–402.
- [20] DAI Z Z, CHEN M Q, GU X D, *et al.* Batch DropBlock Network for Person Re-identification and Beyond // *Proc of the IEEE/CVF International Conference on Computer Vision*. Washington, USA: IEEE, 2019: 3690–3700.
- [21] ZHENG F, DEND C, SUN X, *et al.* Pyramidal Person Re-identification via Multi-loss Dynamic Training // *Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2019: 8506–8514.
- [22] VARIOR R R, HALOI M, WANG G. Gated Siamese Convolutional Neural Network Architecture for Human Re-identification // *Proc of the European Conference on Computer Vision*. Berlin, Germany: Springer, 2016: 791–808.
- [23] YI D, LEI Z, LIAO S C, *et al.* Deep Metric Learning for Person Re-identification // *Proc of the 22nd International Conference on Pattern Recognition*. Washington, USA: IEEE, 2014: 34–39.
- [24] GENG M Y, WANG Y W, XIANG T, *et al.* Deep Transfer Learning for Person Re-identification [ C/OL ]. [ 2020–07–12 ]. <https://arxiv.org/pdf/1611.05244.pdf>.
- [25] ZHENG M, KARANAM S, WU Z Y, *et al.* Re-identification with Consistent Attentive Siamese Networks // *Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2019: 5735–5744.
- [26] CHEN W H, CHEN X T, ZHANG J G, *et al.* Beyond Triplet Loss: A Deep Quadruplet Network for Person Re-identification // *Proc of the IEEE Conference on Computer Vision and Pattern Recognition*. Washington, USA: IEEE, 2017: 1320–1329.

- [27] CHENG D, GONG Y H, ZHOU S P, *et al.* Person Re-identification by Multi-channel Parts-Based CNN with Improved Triplet Loss Function // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2016: 1335–1344.
- [28] DING S Y, LIN L, WANG G R, *et al.* Deep Feature Learning with Relative Distance Comparison for Person Re-identification. Pattern Recognition, 2015, 48(10): 2993–3003.
- [29] HERMANS A, BEYER L, LEIBE B. In Defense of the Triplet Loss for Person Re-identification [C/OL]. [2020-07-12]. <https://arxiv.org/pdf/1703.07737.pdf>.
- [30] SHI H L, YANG Y, ZHU X Y, *et al.* Embedding Deep Metric for Person Re-identification: A Study Against Large Variations // Proc of the European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 732–748.
- [31] 蒋松慧, 张荣, 李小宝, 等. 融合直接度量和间接度量的行人再识别. 模式识别与人工智能, 2018, 31(2): 167–174.  
(JIANG H H, ZHANG R, LI X B, *et al.* Pedestrian Re-identification Fusing Direct Metric and Indirect Metric. Pattern Recognition and Artificial Intelligence, 2018, 31(2): 167–174.)
- [32] WU A C, ZHENG W S, YU H X, *et al.* RGB-Infrared Cross-Modality Person Re-identification // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2017: 5390–5399.
- [33] YE M, LAN X Y, LI J W, *et al.* Hierarchical Discriminative Learning for Visible Thermal Person Re-identification // Proc of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2018: 7501–7508.
- [34] YE M, WANG Z, LAN X Y, *et al.* Visible Thermal Person Re-identification via Dual-Constrained Top-Ranking // Proc of the 27th International Joint Conference on Artificial Intelligence. New York, USA: ACM, 2018: 1092–1099.
- [35] YE M, LAN X Y, WANG Z, *et al.* Bi-directional Center-Constrained Top-Ranking for Visible Thermal Person Re-identification. IEEE Transactions on Information Forensics and Security, 2019, 15: 407–419.
- [36] DAI P Y, JI R R, WANG H B, *et al.* Cross-Modality Person Re-identification with Generative Adversarial Training // Proc of the 27th International Joint Conference on Artificial Intelligence. New York, USA: ACM, 2018: 677–683.
- [37] HAO Y, WANG N N, LI J, *et al.* HSME: Hypersphere Manifold Embedding for Visible Thermal Person Re-identification // Proc of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2019: 8385–8392.
- [38] WANG Z X, WANG Z, ZHENG Y Q, *et al.* Learning to Reduce Dual-Level Discrepancy for Infrared-Visible Person Re-identification // Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2019: 618–626.
- [39] LIU H J, CHENG J, WANG W, *et al.* Enhancing the Discriminative Feature Learning for Visible-Thermal Cross-Modality Person Re-identification. Neurocomputing, 2020, 398: 11–19.
- [40] WANG G A, ZHANG T Z, CHENG J, *et al.* RGB-Infrared Cross-Modality Person Re-identification via Joint Pixel and Feature Alignment // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2019: 3623–3632.
- [41] WANG G A, ZHANG T Z, YANG Y, *et al.* Cross-Modality Paired-Images Generation for RGB-Infrared Person Re-identification // Proc of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2020: 12144–12151.
- [42] LU Y, WU Y, LIU B, *et al.* Cross-Modality Person Re-identification with Shared-Specific Feature Transfer // Proc of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2020: 13376–13386.
- [43] ZHU Y X, YANG Z, WANG L, *et al.* Hetero-Center Loss for Cross-Modality Person Re-identification. Neurocomputing, 2020, 386: 97–109.
- [44] HE D, XIA Y C, QIN T, *et al.* Dual Learning for Machine Translation // LEE D D, SUGIYAMA M, LUXBURG U V, *et al.*, eds. Advances in Neural Information Processing Systems 29. Cambridge, USA: The MIT Press, 2016: 820–828.
- [45] BATRA T, PARIKH D. Cooperative Learning with Visual Attributes [C/OL]. [2020-07-12]. <https://arxiv.org/pdf/1705.05512.pdf>.
- [46] LAN X, ZHU X T, GONG S G. Knowledge Distillation by On-the-Fly Native Ensemble // Proc of the 32nd International Conference on Neural Information Processing Systems. Cambridge, USA: The MIT Press, 2018: 7528–7538.
- [47] SONG G C, CHAI W. Collaborative Learning for Deep Neural Networks // Proc of the 32nd International Conference on Neural Information Processing Systems. Cambridge, USA: The MIT Press, 2018: 1837–1846.
- [48] SUN Y F, ZHENG L, YANG Y, *et al.* Beyond Part Models: Person Retrieval with Refined Part Pooling (and a Strong Convolutional Baseline) // Proc of the European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 501–518.
- [49] WANG G S, YUAN Y F, CHEN X, *et al.* Learning Discriminative Features with Multiple Granularities for Person Re-identification // Proc of the 26th ACM International Conference on Multimedia. New York, USA: ACM, 2018: 274–282.
- [50] FU Y, WEI Y C, ZHOU Y Q, *et al.* Horizontal Pyramid Matching for Person Re-identification // Proc of the AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2019: 8295–8302.
- [51] HE K M, ZHANG X Y, REN S Q, *et al.* Deep Residual Learning for Image Recognition // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2016: 770–778.
- [52] HUANG G, LIU Z, VAN DER MAATEN L, *et al.* Densely Connected Convolutional Networks // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2017: 2261–2269.
- [53] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [C/OL]. [2020-07-12]. <https://arxiv.org/pdf/1409.1556.pdf>.



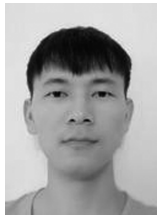
- [54] NGUYEN D T, HONG H G, KIM K W, *et al.* Person Recognition System Based on a Combination of Body Images from Visible Light and Thermal Cameras. *Sensors*, 2017, 17 (3). DOI:10.3390/s17030605.

## 作者简介



郑爱华, 博士, 副教授, 主要研究方向为行人、车辆重识别、音视频计算. E-mail: ahzheng214@ahu.edu.cn.

(ZHENG Aihua, Ph. D., associate professor. Her research interests include person/vehicle re-identification and audio visual computing.)



曾小强, 硕士研究生, 主要研究方向为跨模态行人重识别. E-mail: xqzeng136@foxmail.com.

(ZENG Xiaoqiang, master student. His research interests include cross-modal person re-identification.)



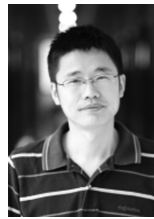
江波 (通讯作者), 博士, 副教授, 主要研究方向为图像匹配、图匹配. E-mail: jiangbo@ahu.edu.cn.

(JIANG Bo (Corresponding author), Ph. D., associate professor. His research interests include image matching and graph matching.)



黄岩, 博士, 副教授, 主要研究方向为机器学习、模式识别. E-mail: yhuang@nlpr.ia.ac.cn.

(HUANG Yan, Ph. D., associate professor. His research interests include machine learning and pattern recognition.)



汤进, 博士, 教授, 主要研究方向为计算机视觉、模式识别. E-mail: tj@ahu.edu.cn.

(TANG Jin, Ph. D., professor. His research interests include computer vision and pattern recognition.)

\*\*\*\*\*

## 第三届中国认知计算与混合智能学术大会 (线上会议) 成功召开

近年来, 认知科学、神经科学以及人工智能等领域迅速发展, 已经成为国际学术研究的热门方向, 并孵化出一系列高新科技产业, 在全球范围内带来了巨大的经济效应。人工智能正在改变人类学习、工作和生活方式, 推动人类社会步入智能化时代, 是引领未来的一项战略性技术。2017年7月, 国务院发布了《新一代人工智能发展规划》, 将人工智能上升为国家战略。

为了认真贯彻落实上述《新一代人工智能发展规划》要求, 推动人工智能相关领域的研究与发展, 为从事人工智能相关研究的学者、工程师、教师和学生提供交流的平台, 由中国自然科学基金委员会、中国自动化学会、中国认知科学学会和西安交通大学共同组织的“第一届和第二届中国认知计算与混合智能学术大会”在西安市成功召开, 共邀请了85位国内外世界知名学者做学术报告, 交流了相关研究领域的最新成果与发展趋势, 取得了很好的效果和影响, 为促进我国新一代人工智能发展做出一定贡献。

鉴于当前国内国际新型冠状病毒肺炎疫情防控形势, 为保证全体参会人员的安全和健康, 今年10月17日至18日举办的“第三届中国认知计算与混合智能学术大会”采用了线上形式召开。大会主席由中国认知科学学会理事长陈霖院士、中国自动化学会理事长郑南宁院士担任。大会共设立了四个论坛, 主要围绕混合智能、知识表达与推理、统计学习与深度学习、计算机视觉、智能机器人、人工智能新型计算架构与芯片、自主智能系统与无人驾驶等主题, 邀请了来自国内外知名大学、研究所和科技公司的25位知名学者和专家做论坛报告。其中, 其中17日上午的“论坛一: 认知计算的基础理论”共6个报告, 17日下午的“论坛二: 混合增强智能与深度学习”共6个报告, 18日上午的“论坛三: 受神经科学启发的计算架构与器件”共6个报告, 18日下午的“论坛四: 先进感知与智能环境”共7个报告。600余人线上聆听了会议。

各位知名学者和专家在报告中展示了研究团队在认知科学、神经科学与人工智能学科等领域交叉融合的最新进展和前沿技术。与会人员进行了深入、热烈地交流和研讨, 共同探索学科交叉领域中的新理论、新方法和新技术, 会议促进了人工智能学科与其他相关科学领域的共同发展。