激光与光电子学进展 先进成像

基于互预测学习的细粒度跨模态行人重识别

李爽[1,2]，李华锋[1,2]，李凡[1,2*]

[1]昆明理工大学信息工程与自动化学院，云南 昆明 650500；

[2]云南省人工智能重点实验室，云南 昆明 650500

摘要 目前，有监督行人重识别方法着重关注单一模态（可见光）的行人检索问题。然而，在 24 h 的监控系统中，除可见光图像外，还存在大量的红外图像（这类图像缺少颜色和纹理信息）。因此，跨模态的行人检索方法可有效提升行人重识别技术的实用性。针对当前跨模态行人重识别方法存在忽视不同模态下独有判别性特征而导致的模型性能受限问题，提出了一种跨模态身份互预测学习和细粒度特征学习的跨模态行人重识别方法。该方法通过对模态专有身份分类器的设计，提升了模态内专有特征的判别性和鲁棒性，并通过构建交叉学习机制，促使网络将不同模态下的专有特征转化为模态不变特征，有效利用了模态特有判别性信息。此外，细粒度特征学习进一步从局部和全局两方面增强了网络特征表示的判别性。所提方法在公开数据集 SYSU-MM01 和 RegDB 上与同类方法相比，其结果优势明显，证明了所提方法的优越性。

关键词 行人重识别；跨模态；互预测；细粒度特征

中图分类号 TP391.4 **文献标志码** A **DOI**：10.3788/LOP202259.1010010

Fine-Grained Cross-Modality Person Re-Identification Based on Mutual Prediction Learning

Li Shuang[1,2], Li Huafeng[1,2], Li Fan[1,2*]

[1]*Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming* 650500, *Yunnan, China*;

[2]*Yunnan Key Laboratory of Artificial Intelligence, Kunming* 650500, *Yunnan, China*

Abstract At present, the supervised person re-identification methods focus on the problem of single modality (visible image). However, in addition to visible images, there are a large number of infrared images which lack color and texture information in the 24-hour surveillance system. Therefore, the cross-modality pedestrian retrieval method can effectively improve the practicability of person re-identification technology. The current cross-modality person re-identification methods ignore the unique discriminant features from different modalities, which leads to the performance limitation. This paper proposes a cross-modality person re-identification method based on cross-modality identity mutual prediction learning and fine-grained feature learning. A modal specific identity classifier is designed to improve the discrimination and robustness of modal specific features. A cross learning mechanism is constructed to promote the network to transform the specific features of different modal into modal invariant features, so as to make effective use of the modal specific discriminant information. In addition, fine-grained feature learning further enhances the discrimination of network feature representation from both local and global aspects.

收稿日期：2021-07-08；**修回日期**：2021-08-22；**录用日期**：2021-09-23

基金项目：国家自然科学基金（61966021）、云南省重大科技专项计划项目（202002AD080001）、云南省基础研究计划项目（202101AT070136）

通信作者：*478263823@qq.com

Comparisons with the state-of-the-art methods on open datasets SYSU-MM01 and RegDB show the advantages of the proposed method.

Key words person re-identification; cross-modality; mutual prediction; fine-grained features

1 引 言

行人重识别任务[1-4]旨在通过多个非重叠的相机视角进行特定行人图像的检索。由于该任务在智能视频监控上具有较高的实用价值，因此受到了计算机视觉研究人员的广泛关注。目前的24 h监控摄像头在白天为可见光拍摄模式，夜晚自动转为红外拍摄模式。拍摄原理的不同，使得可见光模态图像和红外模态图像存在差异：相较于可见光图像，红外模态下的图像缺少纹理以及颜色信息；而相较于红外图像，可见光图像缺少红外图像所具备的热信息。因此，模态差异给跨模态行人重识别带来了挑战。此外，传统可见光图像行人重识别所遇到的挑战，如视角差异、光照差异、分辨率差异、姿态差异，也同样给跨模态行人重识别带来了挑战。所以，跨模态行人重识别任务主要面临两个挑战：1)模态差异带来的图像判别性信息不匹配问题，如颜色在可见光行人重识别任务中是重要的判别性信息，但在跨模态任务中，这是一个冗余信息，因为在红外模态下无法找到相匹配的颜色信息；2)模态内变化带来的挑战，如果完全忽视了模态内的判别性特征学习，即使能够消除模态差异，也无法进行正确的跨模态的行人检索。

现有的跨模态行人重识别主要包括模态共享特征学习和图像迁移两类方法。基于图像迁移的方法通常借助已有的对抗生成网络进行像素级别的图像迁移，通过将可见光图像风格转变为红外图像风格来达到将跨模态行人重识别任务转化为单模态行人重识别任务的目的，从而实现跨模态行人重识别任务的性能提升。但这类算法在训练识别模型时需要使用额外的模型对训练数据进行特定的预处理，这极大地影响了模型在实际场景中的应用效率。此外，这类算法的性能极其依赖图像迁移算法的能力，因此图像风格迁移的图像质量直接决定了最终的识别性能。从已有的工作来看，图像风格迁移生成的图像质量与真实图像还有较大的差距，这限制了跨模态行人重识别性能。

模态共享特征学习方法往往倾向于通过消除模态的特有特征（如颜色、热信息）来学习模态的公共特征，通常此类方法采用双流浅层网络、共享深层网络和模态共享分类器来将不同模态的特征映射到潜在的公共空间，从而学习模态共有知识。从传统行人重识别来看，模态的特有特征（如颜色等）是重要的判别性特征，因此消除模态的特有特征实际上抛弃了一些重要的判别性特征。这导致在消除模态差异的同时，模态内特征的判别性也被削弱，进而限制了跨模态行人重识别的性能。

因此，除了模态共有信息以外，模态特有的判别性信息对于跨模态任务也很重要。如果能够充分利用这些模态特有的判别性信息，就能极大提升跨模态行人重识别的性能。为此，本文提出了基于互预测学习的细粒度跨模态行人重识别方法，其从细粒度特征学习和有效利用模态特有判别性信息两方面进行行人判别性特征的学习。用不同模态的数据分别训练各自模态下的身份分类器以及编码器，该过程能够促使编码器从给定图像中编码出具有模态内强判别性的特征。另外，为有效利用模态特有的判别性信息，使用交叉分类机制迫使编码器学习从输入图像中编码出其他模态下的判别性特征。这一设计迫使编码器不断地从单一模态下挖掘其他模态的特有判别性特征，不仅能够消除模态间的差异而且能够最大可能地保留模态特有判别性特征，同时实现不同模态下的特征在身份层面的对齐。此外，为了更加全面地学习行人判别性特征，对全局特征图和局部特征图分别进行池化得到粗、细两种粒度的判别性特征并进行有效的结合，最后作为行人的外貌描述。所提方法在两个公开数据集SYSU-MM01和RegDB取得了有竞争力的性能，充分证明了其有效性和优越性。

2 相关工作

2.1 可见光模态下的行人重识别

已存在的行人重识别大多聚焦于可见光模态下的行人重识别，这类方法主要解决不连续摄像头下的行人图像匹配问题。针对行人重识别数据集的不足（规模较小、不符合实际等），Zheng等[5]提出了一个大规模可见光模态下的数据集Market-1501，并提供一个基准分类方法。然而这类常规的分类模型仅仅

考虑类间的关系，并没有考虑样本之间的关系。为此，Chen等[6]在分类模型的基础上增加一个验证损失来判断两个图像是否是同一行人。为增强行人特征表示的判别性，Cheng等[7]在三元组损失的基础上提出了一个改进的三元组损失，相比于原有的三元组损失仅仅要求类内距离小于类间距离，该损失要求类内距离也要小于一个设定的边缘参数。这使得同类特征更相似从而有效地提升了特征判别性。考虑到仅仅依赖单尺度的行人重识别方法可能会丢失其他尺度下的潜在有用信息，从而导致识别性能不高，Qian等[8]提出了多尺度深度学习网络，该网络能够在不同尺度下学习深度判别性特征表示，并自动确定最适合匹配的尺度。Li等[9]设计了一个多尺度注意力机制来挖掘图像中的显著信息和次显著信息，从而提升特征判别性。考虑到行人重识别数据量有限的问题，Zheng等[10]提出通过GAN网络（Generative adversarial network）来进行训练集数据的扩充，从而通过更多的数据参与训练以提高模型的识别性能。由于早期的行人重识别主要学习全局判别性特征，忽略了局部重要特征，Sun等[11]将行人特征图切分成局部特征图，并通过部分精炼池化策略将各局部特征图的极端值进行重新分配以获得更有效的细粒度特征，从而提升识别性能。为解决行人不同图片语义信息不对齐、局部遮挡等造成识别性能受限的问题，Zhao等[12]提出了Spindle网络，该网络通过人体关键点引导来进行行人局部特征的提取，并将这些局部特征和全局特征融合得到语义对齐的判别性特征表示。为应对行人姿态变化给行人重识别带来的干扰，Zheng等[13]通过姿态估计生成PoseBox来引导行人和标准姿态对齐从而获得姿态不变特征。目前，有监督的可见光模态行人重识别已经取得了极好的成绩，然而大的模态差异导致这些方法无法直接应用于跨模态行人重识别任务。

2.2 基于度量学习的跨模态行人重识别

为了学习模态共享信息，Wu等[14]将相同模态相似性作为约束提出了焦点模态感知相似性保护损失（Focal modality-aware similarity-preserving loss）。为解决不同相机环境和姿态导致的模态间和模态内的差异问题，Ye等[15]提出双向中心约束Top-Ranking损失，该损失同时考虑了模态间和模态内的变化。此外，在特征度量上，该损失用样本与类中心计算相似度替换了样本与样本之间计算相似度。同样地，为了缓解模态间和模态内的差异，Liu等[16]基于三元组损失提出了双模态三元组损失（Dual-modality triplet loss）来引导模型学习模态间和模态内的判别性特征。Lin等[17]提出了五元组损失（Hard pentaplet loss）来提升交叉模态下的特征判别性。Zhu等[18]提出了异中心损失（Hetero-center loss），具体来说，该损失通过约束类内不同模态间的类中心相似性来达到缩小模态差异，进而学习模态不变特征的目的。Hao等[19]针对分类损失和度量损失结合训练忽略了分类子空间和特征嵌入子空间的相关性的问题，提出了通过Sphere Softmax学习一个超球体流行嵌入，并通过互惠排序损失（Reciprocal ranking loss）和Sphere Loss对模态间和模态内变化进行约束。这类方法通常倾向于提取模态共享特征，忽略了模态特有的判别性特征，因此导致识别性能受限。

2.3 基于图像风格迁移的跨模态行人重识别

基于图像风格迁移的方法通常借助GAN网络将不同模态的图像风格进行统一，从而将跨模态行人重识别任务转化为单模态行人重识别任务。Kniaz等[20]第一次使用GAN网络将可见光图像风格由可见光转化为红外图像风格，然后在红外模态下进行单模态行人重识别任务的学习。然而，由于GAN网络是独立训练的，基于GAN网络的两阶段算法不利于跨模态行人重识别。为此，Wang等[21]提出了一个端到端对齐生成对抗网络（AlignGAN），该网络联合像素级对齐和特征级对齐两种策略对图像风格迁移和特征学习两个任务进行统一优化，从而缓解模态间和模态内的差异。为了同时解决跨模态行人重识别任务中的外观差异和模态差异，Wang等[22]提出了端到端的双路差异减少方法（D2RL），该方法首先通过双向循环GAN网络将可见光模态图像和红外模态图像迁移到多光谱图像，从而实现模态风格统一，然后通过特征级网络训练来减少行人外观差异。Choi等[23]提出了层次跨模态解纠缠（Hi-CMD）方法来处理模态间和模态内的差异，该方法通过生成不同光照和不同姿态变化的图像来学习可见光图像和红外图像之间的公共判别性特征。这类基于风格迁移的方法能够取得较为优异的识别性能，然而这类算法需要额外模型参与训练，这极大地影响了模型在实际场景中的应用效率。且从目前的图像风格迁移结果来看，其图像生成质量与真实图像还存在较大的差距，这也是图像风格迁移模型性能受限的原因之一。

3　方　　法

3.1　问题定义

在跨模态行人重识别任务中，数据集包含可见光图像和红外图像。将可见光和红外图像训练集数据分别定义为 $V=\{(x_{v,i},y_{v,i})|_{i=1}^{n_v}\}$、$T=\{(x_{t,i},y_{t,i})|_{i=1}^{n_t}\}$，其中 $x_{v,i}$、$x_{t,i}$ 分别表示可见光数据和红外数据的第 i 张图像，$x_{v,i}\in\mathbb{R}^{H\times W\times 3}$，$x_{t,i}\in\mathbb{R}^{H\times W\times 3}$，$H$、$W$ 分别表示图像的长和宽，$y_{v,i}$ 表示 $x_{v,i}$ 对应的身份标签，$y_{t,i}$ 表示 $x_{t,i}$ 对应的身份标签，$y_{v,i}\in\{1,2,3,\cdots,K\}$，$y_{t,i}\in\{1,2,3,\cdots,K\}$，$K$ 表示训练集中身份总数，n_v、n_t 分别表示数据集中可见光图像数据量和红外图像数据量，因此整个数据集训练集的图像数量为 $n=n_v+n_t$。在训练阶段，通过 V 和 T 对编码器 E 进行训练；在测试阶段，给定一张红外（或可见光）图像作为查询图像，可见光（或红外）模态下的图像作为被查询对象。通过余弦距离进行查询图像和被查询图像相似度的度量，根据相似度从大到小给出与查询图像相似的图像序列。

针对该任务，提出了基于互预测学习的细粒度跨模态行人重识别方法，主要分为双流网络（DSN，Dual stream network）、细粒度特征学习（FGFL，Fine grained feature learning）和跨模态身份互预测学习（CMIMPL，Cross-modality identity mutual predictive learning）三个模块。

3.2　双流网络

本文所提出的方法以双流网络为基准（baseline）方法，本节通过双流网络结构和双流网络损失对其进行介绍。

3.2.1　双流网络结构

在跨模态行人重识别任务中，编码器往往采用双流网络进行模态共享特征的提取，双流网络由特征提取子网络和特征嵌入子网络两部分组成。如图 1 中 DSN 模块所示，特征提取子网络由两个权重不同的 Resnet 浅层网络构成（包括 Conv1、BN1、ReLU 以及 Maxpool 层），特征嵌入子网络由 Resnet 的深层网络组成（Conv2～5 层）。该设计的目的是让特征提取子网络能够针对不同模态进行图像特征提取，进而促使特征嵌入子网络进行跨模态共享特征的提取。不同模态的图像信息若能够被共享权重的特征嵌入子网络进行有效编码，则表明特定模态下的特征提取子网络能够一定程度地消除模态差异。

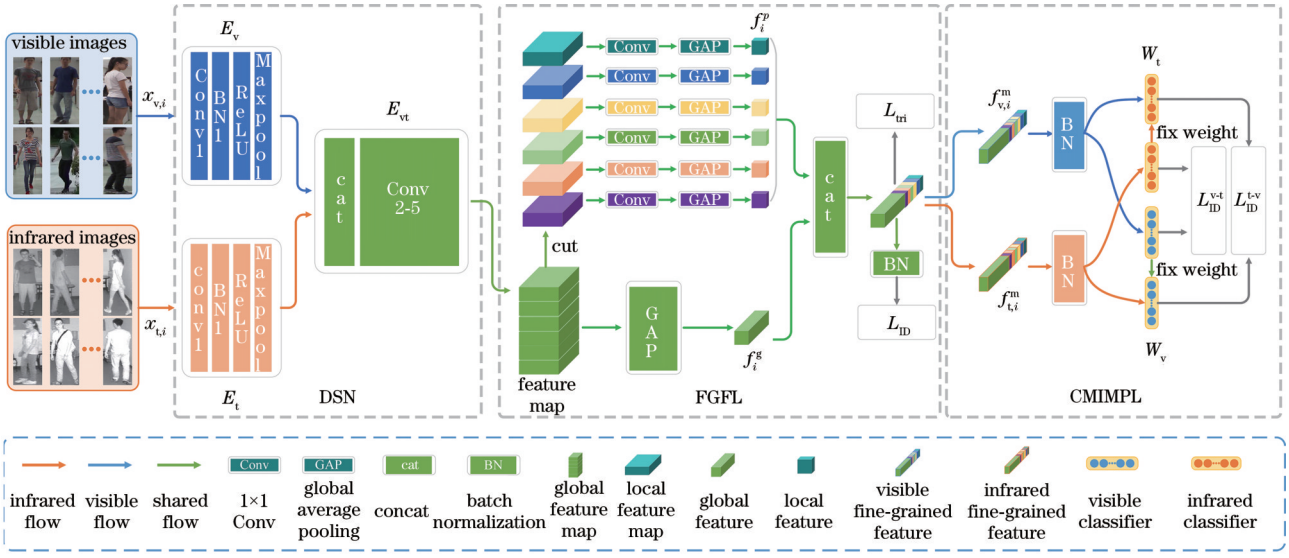

图 1　互预测学习的细粒度跨模态网络结构图

Fig. 1　Diagram of fine-grained cross-modality network for mutual prediction learning

3.2.2　双流网络损失

令可见光特征提取子网络为 E_v，红外特征提取子网络为 E_t，特征嵌入子网络为 E_{vt}。因此，该网络提取的可见光图像特征和红外图像特征可以表示为 $\boldsymbol{f}_{v,i}=\mathrm{GAP}\left(E_{vt}\left(E_v\left(x_{v,i}\right)\right)\right)$ 和 $\boldsymbol{f}_{t,i}=\mathrm{GAP}\left(E_{vt}\left(E_t\left(x_{t,i}\right)\right)\right)$，其中 GAP 表示全局平均池化。为了使该网络能够提取到具有身份判别性的模态共享特征，本文采用交叉熵损失和三元组损失对特征 $\boldsymbol{f}_{v,i}$ 和 $\boldsymbol{f}_{t,i}$ 进行约束。交叉熵损失可以表示为

$$L_{ID}=-\frac{1}{n_b}\sum_{i=1}^{n_b}\boldsymbol{q}_{v,i}\log\left(W_{id}\left(\boldsymbol{f}_{v,i}\right)\right)+\boldsymbol{q}_{t,i}\log\left(W_{id}\left(\boldsymbol{f}_{t,i}\right)\right),\quad(1)$$

式中：W_{id}表示身份分类器；n_{b}表示批处理图像数量（Batch size）；$\boldsymbol{q}_{\mathrm{v},i}\in\mathbb{R}^{K\times 1}$和$\boldsymbol{q}_{\mathrm{t},i}\in\mathbb{R}^{K\times 1}$均为onehot向量，只有第$y_{\mathrm{v},i}$个和第$y_{\mathrm{t},i}$个元素为1。

除了使用交叉熵损失对网络进行优化以外，还通过三元组损失约束模态内和模态间的相同身份特征具有高相似度，不同身份具有低相似度。具体优化公式为

$$L_{\mathrm{tri}}=-\frac{1}{n_{2\mathrm{b}}}\sum_{i=1}^{n_{2\mathrm{b}}}\left[m+\|\boldsymbol{f}_i-\boldsymbol{f}_i^{\mathrm{p}}\|_2-\|\boldsymbol{f}_i-\boldsymbol{f}_i^{\mathrm{n}}\|_2\right]_+,\tag{2}$$

式中：L_{tri}表示同时对模态间和模态内样本进行约束，因此$n_{2\mathrm{b}}=2n_{\mathrm{b}}$，即$n_b$个红外样本和$n_{\mathrm{b}}$个可见光样本均参与该损失的计算。$\boldsymbol{f}_i$表示$n_{2\mathrm{b}}$个样本中的一个，$\boldsymbol{f}_i^{\mathrm{p}}$表示$\boldsymbol{f}_i$对应的难正样本，$\boldsymbol{f}_i^{\mathrm{n}}$表示$\boldsymbol{f}_i$对应的难负样本，$m$设置为0.3。

3.3　细粒度特征学习

在基准方法中，通过GAP对整体特征图进行池化，将其转变为特征向量，用于三元组损失和交叉熵损失的计算以及推理阶段进行特征相似度计算。然而，用GAP操作对整体特征图进行池化是粗糙的，这是因为该操作是通过计算特征图所有像素值的平均值实现的，这导致网络在关注全局判别性信息的同时忽略了局部重要判别性信息。为了使网络关注的模态不变特征更加全面从而增强模型特征表示的身份判别性，从全局和局部两方面出发进行特征图池化使得网络能够同时关注全局特征和局部特征。

如图1中的FGFL所示，为了获得细粒度的特征，将特征图从上往下进行均匀切分，得到6个局部特征图，然后对局部特征图进行GAP得到局部特征向量。因此第p个局部特征$\boldsymbol{f}_{\mathrm{v},i}^{p}$和$\boldsymbol{f}_{\mathrm{t},i}^{p}$表示为

$$\boldsymbol{f}_{\mathrm{v},i}^{p}=\mathrm{GAP}\left(\mathrm{ReLU}\left(\mathrm{Conv}_{1\times 1}^{p}\left(\mathrm{cut}_p\left(E_{\mathrm{vt}}\left(E_{\mathrm{v}}\left(x_{\mathrm{v},i}\right)\right)\right)\right)\right)\right),\tag{3}$$

$$\boldsymbol{f}_{\mathrm{t},i}^{p}=\mathrm{GAP}\left(\mathrm{ReLU}\left(\mathrm{Conv}_{1\times 1}^{p}\left(\mathrm{cut}_p\left(E_{\mathrm{vt}}\left(E_{\mathrm{t}}\left(x_{\mathrm{t},i}\right)\right)\right)\right)\right)\right),\tag{4}$$

式中：$p=0,1,2,3,4,5$；$\mathrm{Conv}_{1\times 1}^{p}$表示$\boldsymbol{f}_{\mathrm{v},i}^{p}$和$\boldsymbol{f}_{\mathrm{t},i}^{p}$对应的$1\times 1$卷积操作，其输入通道为2048，输出通道为256，目的是将局部特征维度减少以节省计算量；ReLU表示Rectified Linear Units激活函数。

对于全局特征图，直接对全局特征图进行GAP，得到

$$\boldsymbol{f}_{\mathrm{v},i}^{\mathrm{g}}=\mathrm{GAP}\left(E_{\mathrm{vt}}\left(E_{\mathrm{v}}\left(x_{\mathrm{v},i}\right)\right)\right),\tag{5}$$

$$\boldsymbol{f}_{\mathrm{t},i}^{\mathrm{g}}=\mathrm{GAP}\left(E_{\mathrm{vt}}\left(E_{\mathrm{t}}\left(x_{\mathrm{t},i}\right)\right)\right),\tag{6}$$

式中，上标g表示全局的含义。为了从局部和全局两方面来增强特征的判别性，将全局和局部特征进行通道上的拼接得到兼顾全局和局部信息的特征向量。即

$$\boldsymbol{f}_{\mathrm{v},i}^{\mathrm{m}}=\mathrm{Cat}\left(\boldsymbol{f}_{\mathrm{v},i}^{\mathrm{g}},\boldsymbol{f}_{\mathrm{v},i}^{0},\boldsymbol{f}_{\mathrm{v},i}^{1},\boldsymbol{f}_{\mathrm{v},i}^{2},\boldsymbol{f}_{\mathrm{v},i}^{3},\boldsymbol{f}_{\mathrm{v},i}^{4},\boldsymbol{f}_{\mathrm{v},i}^{5}\right),\tag{7}$$

$$\boldsymbol{f}_{\mathrm{t},i}^{\mathrm{m}}=\mathrm{Cat}\left(\boldsymbol{f}_{\mathrm{t},i}^{\mathrm{g}},\boldsymbol{f}_{\mathrm{t},i}^{0},\boldsymbol{f}_{\mathrm{t},i}^{1},\boldsymbol{f}_{\mathrm{t},i}^{2},\boldsymbol{f}_{\mathrm{t},i}^{3},\boldsymbol{f}_{\mathrm{t},i}^{4},\boldsymbol{f}_{\mathrm{t},i}^{5}\right),\tag{8}$$

式中，上标m表示兼顾全局和局部信息的含义。同样地，为了使上述细粒度特征能够具有一定的模态不变性和身份判别性，本文使用交叉熵损失和三元组损失约束$\boldsymbol{f}_{\mathrm{v},i}^{\mathrm{m}}$和$\boldsymbol{f}_{\mathrm{t},i}^{\mathrm{m}}$。

$$L_{\mathrm{ID}}=-\frac{1}{n_{\mathrm{b}}}\sum_{n_{\mathrm{b}}=1}^{n_{\mathrm{b}}}\boldsymbol{q}_{\mathrm{v},i}\log\left(W_{\mathrm{id}}\left(\boldsymbol{f}_{\mathrm{v},i}^{\mathrm{m}}\right)\right)+\boldsymbol{q}_{\mathrm{t},i}\log\left(W_{\mathrm{id}}\left(\boldsymbol{f}_{\mathrm{t},i}^{\mathrm{m}}\right)\right),\tag{9}$$

$$L_{\mathrm{tri}}=-\frac{1}{n_{2\mathrm{b}}}\sum_{i=1}^{n_{2\mathrm{b}}}\left[m+\|\boldsymbol{f}_i^{\mathrm{m}}-\boldsymbol{f}_i^{\mathrm{m,p}}\|_2-\|\boldsymbol{f}_i^{\mathrm{m}}-\boldsymbol{f}_i^{\mathrm{m,n}}\|_2\right]_+,\tag{10}$$

式中，$\boldsymbol{f}_i^{\mathrm{m,p}}$表示$\boldsymbol{f}_i^{\mathrm{m}}$对应的难正样本，$\boldsymbol{f}_i^{\mathrm{m,n}}$表示$\boldsymbol{f}_i^{\mathrm{m}}$对应的难负样本。

目前，大多数跨模态行人重识别方法仅关注行人全局特征。与该类方法相比，本文提出的细粒度特征学习模块不仅能够关注全局判别性特征还能够进行局部细粒度特征的学习，有效地增强了行人外貌特征表示的判别性，提升了模型的性能。

3.4　跨模态身份互预测学习

细粒度特征学习模块通过对全局和局部特征的挖掘，提升了模型挖掘模态共享特征的能力，使模型特征表示具有一定的模态不变性。然而这忽视了模态内的判别性信息，使得特征表示在单模态内不具备强判别性。同时，这也导致特征表示在模态间的判别性不足，最终限制了模型的性能。

为解决上述问题，提出了跨模态身份互预测学习，如图2所示。具体地，通过可见光模态身份分类器W_{v}和红外模态身份分类器W_{t}来提升编码器E_{vt}、E_{v}和E_{t}提取具有模态内强判别性特征的能力。同时，这也能够促使W_{v}和W_{t}进行单模态下类原型的学习。该过程通过$L_{\mathrm{ID}}^{\mathrm{v\text{-}t}}$损失对编码器$E_{\mathrm{vt}}$、$E_{\mathrm{v}}$、$E_{\mathrm{t}}$以及$W_{\mathrm{v}}$、$W_{\mathrm{t}}$进行训练实现：

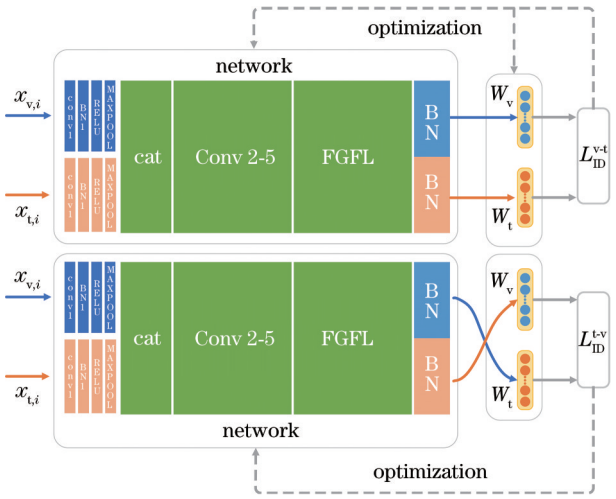

图 2　跨模态身份互预测学习

Fig. 2　Cross-modality identity mutual prediction learning

$$L_{\mathrm{ID}}^{\mathrm{v\text{-}t}}=-\frac{1}{n_{\mathrm{b}}}\sum_{n_{\mathrm{b}}=1}^{n_{\mathrm{b}}}\boldsymbol{q}_{\mathrm{v},i}\log\left(W_{\mathrm{v}}\left(\boldsymbol{f}_{\mathrm{v},i}^{\mathrm{m}}\right)\right)+\boldsymbol{q}_{\mathrm{t},i}\log\left(W_{\mathrm{t}}\left(\boldsymbol{f}_{\mathrm{t},i}^{\mathrm{m}}\right)\right),\tag{11}$$

通过 $L_{\mathrm{ID}}^{\mathrm{v\text{-}t}}$ 的约束，$\boldsymbol{f}_{\mathrm{v},i}^{\mathrm{m}}$ 和 $\boldsymbol{f}_{\mathrm{t},i}^{\mathrm{m}}$ 在模态内的判别性得到了保证，然而强模态内判别性也会使得模型提取的不同模态特征具有较大的模态差异，从而导致跨模态特征相似度匹配的性能下降。这是因为上述学习仅仅迫使模型对单模态内判别性信息关注，忽略了模态间的特征差异。若能够将不同模态下的强模态内判别性特征映射到同一潜在子空间从而建立起联系，将能够使得模型的特征表示同时具备强模态内判别性和模态不变性。为实现上述目的，在通过式(11)更新 W_{v} 和 W_{t} 的权重后，采用交叉训练机制促使权重固定的 W_{v} 来引导 $\boldsymbol{f}_{\mathrm{t},i}^{\mathrm{m}}$ 将红外模态特有的判别特征转化为可见光模态下的特征从而建立起与可见光模态特征下的联系。反之，通过权重固定的 W_{t} 来引导 $\boldsymbol{f}_{\mathrm{v},i}^{\mathrm{m}}$ 进行模态特有判别性特征的转化。具体为

$$L_{\mathrm{ID}}^{\mathrm{t\text{-}v}}=-\frac{1}{n_{\mathrm{b}}}\sum_{n_{\mathrm{b}}=1}^{n_{\mathrm{b}}}\boldsymbol{q}_{\mathrm{v},i}\log\left(W_{\mathrm{t}}\left(\boldsymbol{f}_{\mathrm{v},i}^{\mathrm{m}}\right)\right)+\boldsymbol{q}_{\mathrm{t},i}\log\left(W_{\mathrm{v}}\left(\boldsymbol{f}_{\mathrm{t},i}^{\mathrm{m}}\right)\right),\tag{12}$$

该模块首先通过模态专有分类器的训练保证了模态内特征的判别性，然后通过模态专有分类器来构建交叉学习机制，从而实现网络在不同模态下的互预测学习。该学习机制不仅能够增强模态内特征表示的判别性，还能够充分利用单模态下的特征来增强模态间的识别性能。相比于仅关注不同模态图像下共享特征的方法，该模块能够利用不同模态图像的更多判别性信息来提高模型性能。

3.5　网络结构和优化

3.5.1　网络结构

使用 Resnet-50 作为基准网络，其中第一个卷积层(Conv1)作为特征提取子网络，后面的卷积层(Conv2～5)作为特征嵌入子网络。W_{v} 和 W_{t} 的结构相同，均采用输入通道为 2048，输出通道为 K 的全连接层。W_{id} 采用输入通道为 3584，输出通道为 K 的全连接层。

3.5.2　优　　化

在训练阶段，本文的整体优化函数为

$$L_{\mathrm{full}}=L_{\mathrm{ID}}+L_{\mathrm{tri}}+\alpha L_{\mathrm{ID}}^{\mathrm{v\text{-}t}}+\beta L_{\mathrm{ID}}^{\mathrm{t\text{-}v}}。\tag{13}$$

这里的 $\alpha,\beta\geqslant 0$ 作为权重去平衡相关损失项的重要性。所有损失均参与双流网络(E_{v}、E_{t} 和 E_{vt})的优化，$L_{\mathrm{ID}}^{\mathrm{v\text{-}t}}$ 损失除了优化双流网络以外，还对分类器 W_{v} 和 W_{t} 进行优化。网络一共迭代 80 次。在 20 代之前，仅仅使用 L_{ID} 和 L_{tri} 对网络进行优化使网络具备提取细粒度模态共有特征的能力。20 代之后，$L_{\mathrm{ID}}^{\mathrm{v\text{-}t}}$ 或 $L_{\mathrm{ID}}^{\mathrm{t\text{-}v}}$ 交替(间隔 5 代)参与网络优化促使网络具备将模态特有判别性特征转变为模态不变特征的能力。

4　实　　验

4.1　数据集及评估协议

4.1.1　数据集

SYSU-MM01[24] 是一个大规模的跨模态(红外-可见光)行人重识别基准数据集。如表 1 所示，该数据集由 6 个摄像头采集获得，其中 3 个摄像头对应室内场景，另外 3 个摄像头用于采集户外场景数据，不同场景下的数据也给该数据集增加了一定的难度。该数据包含 491 个行人的 287628 张可见光图像和 15792 张红外图像。其中，训练集包含 395 个行人的 19659 张可见光图像和 12792 张红外图像。测试集一共有 96 个行人，其中查询集(query set)有 3803 张红外图像，301 张可见光图像作为图库集(gallery set)。该数据集有 2 个评估模式，分别是

表 1　SYSU-MM01 数据集采集环境以及采集设备

Table 1　SYSU-MM01 dataset collection environment and collection equipment

Camera	Location	(In/Out)door	Camera type	Device
1	room1	indoor	rgb	Kinect V1
2	room2	indoor	rgb	Kinect V1
3	room2	indoor	ir	—
4	gate	outdoor	rgb	—
5	garden	outdoor	rgb	—
6	passage	outdoor	ir	—

indoor-search 和 all-search。在 all-search 模式下，红外摄像头 3 和 6 作为查询集，可见光摄像头 1、2、4、5 作为图库集。相较于 all-search 模式，indoor-search 模式在图库集中移除了可见光摄像头 4、5，因此 indoor-search 模式的测试难度比 all-search 模式低。

RegDB[25]的数据通过 2 个摄像头（可见光摄像头型号为 Logitech C600，红外摄像头型号为 Tau2 camera）采集获得，共有 412 个行人的 8240 张图像，训练集和测试集分别包含 206 个行人。其中，每个行人有 10 张可见光图像和 10 张红外图像。该数据集包含 2 种评估模式：可见光查红外（visible-thermal）和红外查可见光（thermal-visible）。根据文献[25]，该数据集包含 10 种不同的训练集测试集划分方式，因此取 10 次实验结果的平均指标作为提出方法在该数据集上的最终性能。

4.1.2 评估协议

根据已有跨模态行人重识别方法的标准评估协议，查询集和图库集使用不同模态下的数据，并采用 CMC（Cumulative matching characteristics）和 mAP（Mean average precision）来进行性能的测试。

4.2 实验细节

在训练阶段，图像尺寸被统一到 288 pixel×144 pixel。与文献[26]相似，笔者通过随机裁剪、随机翻转来实现数据增强。实验中，batch size 设置为 32，整个网络采用 SGD（Stochastic gradient descent）优化器，权重衰减设为 0.0005，学习率为 0.1。根据文献[27]，在 0～10 代通过 warm-up 策略[28]线性调节学习率。学习率在 20 代和 50 代按 1/10 减少。对于 SYSU-MM01 数据集，超参数 α 和 β 分别设置为 12.2 和 0.4。对于 RegDB 数据集，α 和 β 设置为 35 和 4。所有实验均在单张 2080TI GPU 上，基于 Pytorch 框架实现。

4.3 方法比较

本节将提出方法与近年该研究领域最好的跨模态行人重识别方法进行了比较，主要包括 Zero-Pad[24]（Zero-padding）、cmGAN[29]（Cross-modality generative adversarial network）、HCML[30]（Hierarchical cross-modality metric learning）、HSME[19]（Hypersphere manifold embedding）、D2RL[22]（Dual-level discrepancy reduction learning）、MAC[31]（Modality-aware collaborative）、AliGAN[21]（Alignment generative adversarial network）、HPILN[17]（Hard pentaplet and identity loss network）、DFE[32]（Dual-alignment feature embedding）、Hi-CMD[23]（Hierarchical cross-modality disentanglement）、EDFL[16]（Enhancing the discriminative feature learning）、CDP[33]（Cross-spectrum dual-subspace pairing）、eBDTR[15]（Bi-directional center-constrained top-ranking）、XIV[34]（X-infrared visible）、expAT[35]（Bidirectional exponential angular triplet）、MSR[36]（Modality-specific representations）、JSIA[37]（Joint set-level and instance-level alignment）、CMSP[14]（Cross-modality similarity preservation）、DFLA[38]（Deep feature learning with attributes）、AGW[26]（Attention generalized mean pooling with weighted triplet loss）、cm-SSFT[39]（Cross-modality shared specific feature transfer）、HAT[40]（Homogeneous augmented tri-modal）。RegDB 数据集和 SYSU-MM01 数据集的结果分别展示在表 2 和表 3。如表 2 所示，提出方法在 RegDB 数据集的 visible-thermal 和 thermal-visible 模式下均能够超过次优方法 cm-SSFT。具体来看，在 visible-thermal 模式上，提出方法 Rank1 超出了 15.34%，mAP 超出了 5.55%，在 thermal-visible 模式上，Rank1 和 mAP 分别超出了 13.82% 和 4.63%。对于 SYSU-MM01 数据集，在 all-search 模式中，提出方法的 Rank1 和 mAP 比次优的 AGW 分别高出 8.98% 和 4.26%，在 indoor-search 模式中，提出方法的 Rank1/mAP 比次优方法 AGW 高 5.52%/2.89%。AGW 与提出方法均采用了双流网络的结构，该方法在双流网络的基础上添加了非局部自注意力（Non-local attention）和 GEM（Generalized-mean pooling）池化层。非局部自注意力能够促使网络关注到更多的模态间共有判别性信息，GEM 池化层相较于普通池化层提取到更多判别性全局特征。但该方法忽视了模态内的判别性信息以及细粒度判别性特征，这是提出方法在性能上超过 AGW 的主要原因。相比于 AGW 这一类仅关注模态间共有判别性信息的方法，提出方法能够将细粒度的模态特有特征和共有特征一起映射到统一的潜在空间，因此能够在不削弱模态内特征判别性的基础上减少模态间的差异，进而提升特征表示的判别性。

4.4 消融实验

提出方法主要包括细粒度特征学习和跨模态身份互预测学习 2 个模块，为验证其有效性，对每个模块进行分析以确保每个模块的有效性，如表 4 所示。在 SYSU-MM01 和 RegDB 上的消融实验分别采用 all-search 模式和 visible-thermal（trial=1）模式。

表 2 在 RegDB 数据集上的对比实验

Table 2 Comparative experiments on RegDB dataset unit: %

Method	Visible-thermal				Thermal-visible			
	r=1	r=10	r=20	mAP	r=1	r=10	r=20	mAP
Zero-Pad[24]	17.75	34.21	44.35	18.90	16.63	34.68	44.25	17.82
HCML[30]	24.44	47.53	56.78	20.80	21.70	45.02	55.58	22.24
HSME[19]	50.85	73.36	81.66	47.00	50.15	72.40	81.07	46.16
D2RL[22]	43.40	66.10	76.30	44.10	–	–	–	–
MAC[31]	36.43	62.36	71.63	37.03	36.20	61.68	70.99	39.23
AliGAN[21]	57.90	–	–	53.60	56.30	–	–	53.40
DFE[32]	70.13	86.32	91.96	69.14	–	–	–	–
eBDTR[15]	34.62	58.96	68.72	33.46	34.21	58.74	68.64	32.49
MSR[36]	48.43	70.32	79.95	48.67	–	–	–	–
JSIA[37]	48.50	–	–	49.30	48.10	–	–	48.90
EDFL[16]	52.58	72.10	81.47	52.98	51.89	72.09	81.04	52.13
XIV[34]	62.21	83.13	91.72	60.18	–	–	–	–
CDP[33]	65.00	83.50	89.60	62.70	65.3	84.5	91.0	62.1
expAT[35]	66.48	–	–	67.31	67.45	–	–	66.51
CMSP[14]	65.07	83.71	–	64.50	–	–	–	–
Hi-CMD[23]	70.93	86.39	–	66.04	–	–	–	–
HAT[40]	71.83	87.16	92.16	67.56	70.02	86.45	91.61	66.30
cm-SSFT[39]	72.30	–	–	72.90	71.00	–	–	71.70
AGW[26]	70.05	–	–	66.37	–	–	–	–
Ours	**87.64**	**95.61**	**97.6**	**78.45**	**84.82**	**94.64**	**97.03**	**76.33**

表 3 在 SYSU-MM01 数据集上的对比实验

Table 3 Comparative experiments on SYSU-MM01 dataset unit: %

Method	All-search				Indoor-search			
	r=1	r=10	r=20	mAP	r=1	r=10	r=20	mAP
Zero-Pad[27]	14.80	54.12	71.33	15.95	20.58	68.38	85.79	26.92
cmGAN[29]	26.97	67.51	80.56	27.80	31.63	77.23	89.18	42.19
HCML[30]	14.32	53.16	69.17	16.16	24.52	73.25	86.73	30.08
HSME[19]	20.68	62.74	77.95	23.12	–	–	–	–
D2RL[22]	28.90	70.60	82.40	29.20	–	–	–	–
MAC[31]	33.26	79.04	90.09	36.22	36.43	62.36	71.63	37.03
AliGAN[21]	42.40	85.00	93.70	40.70	45.90	87.60	94.40	54.30
HPILN[17]	41.36	84.78	94.51	42.95	45.77	91.82	98.46	56.52
DFE[32]	48.71	88.86	95.27	48.59	52.25	89.86	95.85	59.68
Hi-CMD[23]	34.94	77.58	–	35.94	–	–	–	–
EDFL[16]	36.94	85.42	93.22	40.77	–	–	–	–
CDP[33]	38.00	82.30	91.70	38.40	–	–	–	–
expAT[35]	38.57	76.64	86.39	38.61	–	–	–	–
XIV[34]	49.92	89.79	95.96	50.73	–	–	–	–
eBDTR[15]	27.82	67.34	81.34	28.42	32.46	77.42	89.62	42.46
MSR[36]	37.35	83.40	93.34	38.11	39.64	89.29	97.66	50.88
JSIA[37]	38.10	80.70	89.90	36.90	43.80	86.20	94.20	52.90
CMSP[14]	43.56	86.25	–	44.98	48.62	89.50	–	57.50
DFLA[38]	47.14	87.93	94.45	47.08	48.03	88.13	95.14	56.84
AGW[26]	47.50	–	–	47.65	54.17	–	–	62.97
Ours	**56.48**	**88.25**	**94.43**	**51.91**	**59.69**	**92.80**	**97.55**	**65.86**

表 4　消融实验

Table 4　Ablation experiment

Ablation study setting	SYSU-MM01(all-search)				RegDB(visible-thermal)			
	r=1	r=10	r=20	mAP	r=1	r=10	r=20	mAP
Baseline	48.17	82.2	89.32	45.39	67.62	85.63	91.41	62.74
Baseline+FGFL	51.22	85.59	93.40	49.43	73.69	86.46	91.21	65.81
Baseline+FGFL+CMIMPL	56.48	88.25	94.43	51.91	89.37	95.44	96.89	80.30

4.4.1　细粒度特征学习的有效性

为验证细粒度特征学习模块的有效性，用细粒度特征学习模块替换了基准方法中对应的部分，并用相同的交叉熵损失和三元组损失对模型进行优化。如表 4 所示，与基准方法比较，Rank1/mAP 在 SYSU-MM01（RegDB）上超出了 3.05%/4.04%（6.07%/3.07%）。这表明细粒度特征学习模块不仅能够促使网络关注全局判别性特征，也能够关注局部的重要特征。

4.4.2　跨模态身份互预测学习的有效性

跨模态身份互预测学习模块的有效性验证通过在细粒度特征学习的基础上添加跨模态身份互预测学习模块实现。从表 4 可以观察到，与只添加细粒度特征学习模块的性能相比，在 RegDB 和 SYSU-MM01 上，Rank1（mAP）分别从 73.69%（65.81%）提升到了 89.37%（80.30%），从 51.22%（49.43%）提升到了 56.48%（51.91%）。这是因为跨模态互预测学习能够使网络提取的模态不变特征中不仅包含模态共有特征，还含有重要的模态特有判别性特征。此外，这也表明细粒度特征学习和跨模态互预测学习具备互补性，两者的结合使网络能够关注到更多的模态不变的关键信息并对其进行编码。

4.5　参数分析

本文中涉及两个超参数 α 和 β。在超参数作用的分析中，通过固定一个参数，来分析另一个参数对实验性能的影响。为了获得每个数据集上的最佳的超参数，分别对 SYSU-MM01 和 RegDB 两个数据集进行超参数分析。

4.5.1　超参数 α

在式（13）中，超参数 α 主要起调节 $L_{\mathrm{ID}}^{\mathrm{v\text{-}t}}$ 的作用。该损失项主要用来保证双流网络（E_{v}、E_{t} 和 E_{vt}）能够提取模态内的判别性特征，并为每个模态训练模态专有身份分类器（W_{v} 和 W_{t}）。图 3（a）和图 3（b）中展示

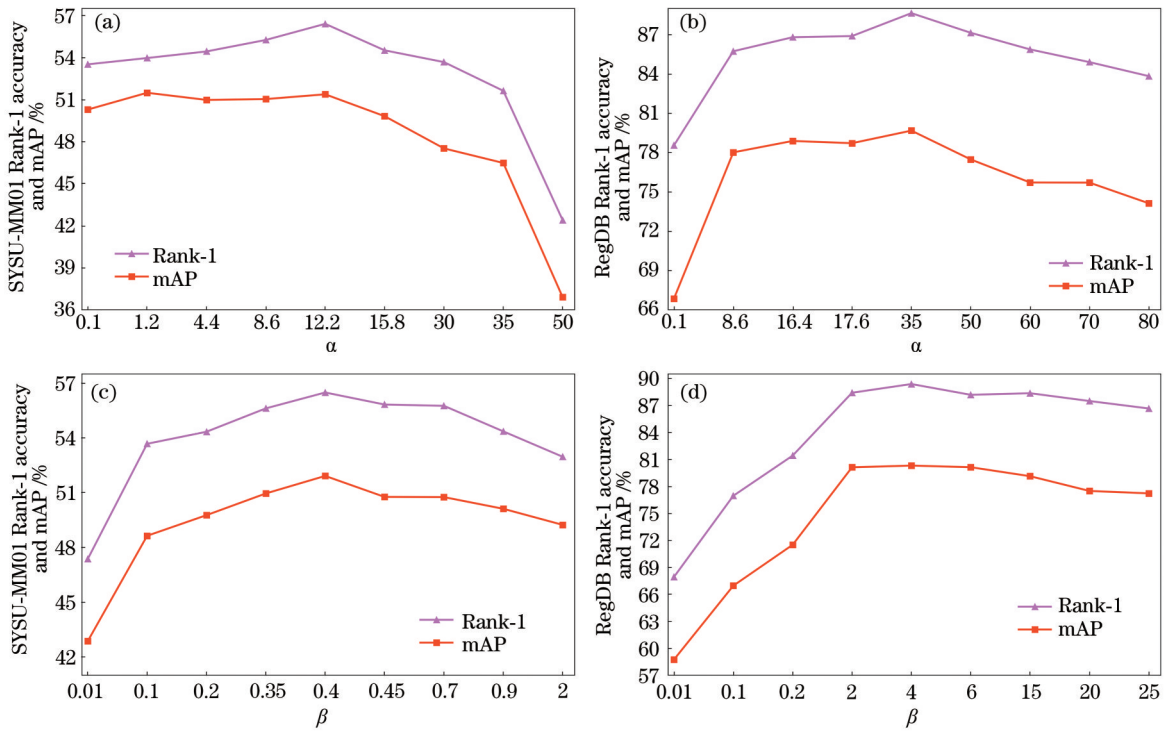

图 3　超参数 α 和 β 的有效性分析。（a）（b）α 的有效性分析；（c）（d）β 的有效性分析

Fig. 3　Effect analysis on hyperparameters α and β. (a)(b) Effect analysis of α; (c)(d) effect analysis of β

了在SYSU-MM01和RegDB数据集中，α为不同值时对Rank-1和mAP的影响。对于数据集SYSU-MM01，当$\alpha\in[0.1, 12.2]$时，本文算法在两个任务上的Rank-1和mAP识别精度有整体提升之势；当$\alpha\in[12.2, 2]$时，两个任务上的Rank-1和mAP出现下降，因此$\alpha=12.2$是最佳选择。对于数据集RegDB，当$\alpha=35$时性能达到最优，因此α应当设为35。

4.5.2 超参数β

超参数β起到调整式(13)中损失$L_{\mathrm{ID}}^{\mathrm{t\text{-}v}}$作用。该损失项是通过$W_{\mathrm{v}}$和$W_{\mathrm{t}}$来引导双流网络将模态特有特征转化为模态不变特征。通过固定超参数α，进行β的参数分析。在SYSU-MM01数据集和RegDB数据集上，β取不同值时的Rank-1和mAP变化如图3(c)和图3(d)所示，由此可以看出β分别为0.4和4时，所提方法获得最优的性能，因此将β设为1是合理的。

4.6 鲁棒性和适用性分析

对于RegDB数据集，主要通过visible-thermal和thermal-visible模式对模型性能进行评估，前者通过可见光图像检索红外图像，后者用红外图像检索可见光图像。两种模式下的实验结果一定程度下验证了本文方法的鲁棒性和有效性。

由4.1节可知，RegDB数据集仅由2个摄像头所采集的图像且相同身份的行人在不同模态下的姿态变化不大，这导致了该数据集的检索难度偏低。为进一步验证本文方法在复杂环境下的性能，在更具挑战性的SYSU-MM01数据集上进行了实验。SYSU-MM01数据集具有indoor-search和all-search两种模式。在indoor-search协议下，摄像头3(room2)和摄像头6(passage)采集的图像作为查询集，摄像头1(room1)和摄像头2(room2)采集的图像作为图库集，因此该模式能够对室内环境下跨模态行人检索的性能进行评估。相对于indoor-search模式，all-search模式更具挑战性。该模式在indoor-search的基础上，增加摄像头4(gate)和摄像头5(garden)的数据至图库集，将模型性能评估的难度从室内提升到户外。

从4.3节的实验结果来看，本文方法在两个数据集上的实验结果均取得了较好的成绩，具有较强的鲁棒性和适用性。

5 结 论

提出了一种新颖的跨模态行人重识别方法。该方法主要由细粒度特征学习和跨模态身份互预测学习两部分组成。其中，细粒度特征学习模块通过对全局信息和局部信息的关注促进网络进行不同粒度下的模态共有判别性特征的挖掘；跨模态身份互预测学习模块通过增强模态内特征的判别性并将其转化为模态不变的判别性特征，有效地消除了不同模态间的差异，促进了识别性能的大幅提升。此外，本文方法不需要进行像素级的图像风格迁移，因此具有更强的现实意义。在两个公共基准数据集上的实验验证了本文方法在跨模态行人重识别任务上的有效性以及相对于同类算法的优越性。

参 考 文 献

[1] 刘可文, 房攀攀, 熊红霞, 等. 基于多层级特征的行人重识别[J]. 激光与光电子学进展, 2020, 57(8): 081503.
Liu K W, Fang P P, Xiong H X, et al. Person re-identification based on multi-layer feature[J]. Laser & Optoelectronics Progress, 2020, 57(8): 081503.

[2] 刘莎, 党建武, 王松, 等. 结合一阶和二阶空间信息的行人重识别[J]. 激光与光电子学进展, 2021, 58(2): 0215005.
Liu S, Dang J W, Wang S, et al. Person re-identification based on first-order and second-order spatial information[J]. Laser & Optoelectronics Progress, 2021, 58(2): 0215005.

[3] 张涛, 易争明, 李璇, 等. 一种基于全局特征的行人重识别改进算法[J]. 激光与光电子学进展, 2020, 57(24): 241503.
Zhang T, Yi Z M, Li X, et al. Improved algorithm for person re-identification based on global features [J]. Laser & Optoelectronics Progress, 2020, 57(24): 241503.

[4] 邬可, 张宝华, 吕晓琪, 等. 基于压缩激励残差网络与特征融合的行人重识别[J]. 激光与光电子学进展, 2020, 57(18): 181007.
Wu K, Zhang B H, Lü X Q, et al. Person re-identification based on squeeze and excitation residual neural network and feature fusion[J]. Laser & Optoelectronics Progress, 2020, 57(18): 181007.

[5] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1116-1124.

[6] Chen H R, Wang Y W, Shi Y M, et al. Deep

transfer learning for person re-identification[C]//2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM), September 13-16, 2018, Xi'an, China. New York: IEEE Press, 2018: 18168307.

[7] Cheng D, Gong Y H, Zhou S P, et al. Person re-identification by multi-channel parts-based CNN with improved triplet loss function[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1335-1344.

[8] Qian X L, Fu Y W, Jiang Y G, et al. Multi-scale deep learning architectures for person re-identification [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 5409-5418.

[9] 李聪, 蒋敏, 孔军. 基于多尺度注意力机制的多分支行人重识别算法[J]. 激光与光电子学进展, 2020, 57(20): 201001.
Li C, Jiang M, Kong J. Multi-branch person re-identification based on multi-scale attention[J]. Laser & Optoelectronics Progress, 2020, 57(20): 201001.

[10] Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline *in vitro*[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 3774-3782.

[11] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11208: 501-518.

[12] Zhao H Y, Tian M Q, Sun S Y, et al. Spindle net: person re-identification with human body region guided feature decomposition and fusion[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 907-915.

[13] Zheng L, Huang Y J, Lu H C, et al. Pose-invariant embedding for deep person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(9): 4500-4509.

[14] Wu A C, Zheng W S, Gong S G, et al. RGB-IR person re-identification by cross-modality similarity preservation[J]. International Journal of Computer Vision, 2020, 128(6): 1765-1785.

[15] Ye M, Lan X, Wang Z, 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.

[16] Liu H J, Cheng J, Wang W, et al. Enhancing the discriminative feature learning for visible-thermal cross-modality person re-identification[J]. Neurocomputing, 2020, 398: 11-19.

[17] Lin J W, Li H. HPILN: a feature learning framework for cross-modality person re-identification [EB/OL]. (2019-06-07) [2021-05-06]. https://arxiv.org/abs/1906.03142.

[18] Zhu Y X, Yang Z, Wang L, et al. Hetero-center loss for cross-modality person re-identification[J]. Neurocomputing, 2020, 386: 97-109.

[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, January 27-February 1, 2019, Honolulu, Hawaii, USA. Menlo Park: AAAI Press, 2019: 8385-8392.

[20] Kniaz V V, Knyaz V A, Hladůvka J, et al. ThermalGAN: multimodal color-to-thermal image translation for person re-identification in multispectral dataset[M]//Leal-Taixé L, Roth S. Computer vision-ECCV 2018 workshops. Lecture notes in computer science. Cham: Springer, 2019, 11134: 606-624.

[21] 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, October 27-November 2, 2019, Seoul, South Korea. New York: IEEE Press, 2019: 3622-3631.

[22] Wang Z X, Wang Z, Zheng Y Q, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 618-626.

[23] Choi S, Lee S, Kim Y, et al. Hi-CMD: hierarchical cross-modality disentanglement for visible-infrared person re-identification[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10254-10263.

[24] Wu A, Zheng W S, Yu H X, et al. RGB-infrared cross-modality person re-identification[C]//Proceedings

of the IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 5380-5389.

[25] 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[J]. Sensors, 2017, 17(3): 605.

[26] Ye M, Shen J, Lin G, et al. Deep Learning for Person Re-identification: A Survey and Outlook[EB/OL]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, (2021-01-06)[2021-05-05]. https://arxiv.org/abs/2001.04193.

[27] Luo H, Gu Y Z, Liao X Y, et al. Bag of tricks and a strong baseline for deep person re-identification[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 16-17, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 1487-1495.

[28] Fan X, Jiang W, Luo H, et al. SphereReID: deep hypersphere manifold embedding for person re-identification[J]. Journal of Visual Communication and Image Representation, 2019, 60: 51-58.

[29] Dai P, Ji R, Wang H, et al. Cross-modality person re-identification with generative adversarial training [C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, July 13-19, 2018, Stockholm, Sweden. New York: IJCAI, 2018: 677-683.

[30] 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, February 2-7, 2018, New Orlean, Louisiana, USA. Menlo Park: AAAI Press, 2018: 7501-7508.

[31] Ye M, Lan X Y, Leng Q M. Modality-aware collaborative learning for visible thermal person re-identification[C]//Proceedings of the 27th ACM International Conference on Multimedia, October 21-25, 2019, Nice, France. New York: ACM, 2019: 347-355.

[32] Hao Y, Wang N N, Gao X B, et al. Dual-alignment feature embedding for cross-modality person re-identification[C]//Proceedings of the 27th ACM International Conference on Multimedia, October 21-25, 2019, Nice, France. New York: ACM, 2019: 57-65.

[33] Fan X, Luo H, Zhang C, et al. Cross-spectrum dual-subspace pairing for RGB-infrared cross-modality person re-identification[EB/OL]. (2020-02-29)[2021-05-05]. https://arxiv.org/abs/2003.00213.

[34] Li D G, Wei X, Hong X P, et al. Infrared-visible cross-modal person re-identification with an X modality[C]//Proceedings of the AAAI Conference on Artificial Intelligence, February 7-12, 2020, New York, NY, USA. Menlo Park: AAAI Press, 2020, 34(4): 4610-4617.

[35] Ye H, Liu H, Meng F, et al. Bi-directional exponential angular triplet loss for Rgb-infrared person re-identification[J]. IEEE Transactions on Image Processing, 2020, 30: 1583-1595.

[36] Feng Z X, Lai J H, Xie X H. Learning modality-specific representations for visible-infrared person re-identification[J]. IEEE Transactions on Image Processing, 2020, 29: 579-590.

[37] Wang G A, Zhang T Z, Yang Y, et al. Cross-modality paired-images generation for RGB-infrared person re-identification[C]//Proceedings of the AAAI Conference on Artificial Intelligence, February 7-12, 2020, New York, NY, USA. Menlo Park: AAAI Press, 2020: 12144-12151.

[38] Zhang S K, Chen C H, Song W R, et al. Deep feature learning with attributes for cross-modality person re-identification[J]. Journal of Electronic Imaging, 2020, 29(3): 033017.

[39] Lu Y, Wu Y, Liu B, et al. Cross-modality person re-identification with shared-specific feature transfer [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 13376-13386.

[40] Ye M, Shen J B, Shao L. Visible-infrared person re-identification via homogeneous augmented tri-modal learning[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 728-739.