

频率感知判别特征学习的跨模态行人重识别方法

程德强¹, 牛凯凯¹, 郝大帅¹, 宋天舒¹, 寇旗旗²

(1. 中国矿业大学信息与控制工程学院, 江苏 徐州 221116; 2. 中国矿业大学计算机科学与技术学院/人工智能学院, 江苏 徐州 221116)

摘要: 针对现有 VI-ReID 频域方法对频带调制权重缺少跨模态一致性与身份可分性联合监督的问题, 本文提出一种频率感知判别特征学习的跨模态行人重识别方法。首先, 设计相位引导频谱调制模块, 利用可见光与红外幅度统计构建共享参考幅度, 在相位结构引导下自适应调节幅度响应, 并聚合低频结构与高频细节信息, 在减小模态间幅度差异的同时保持身份相关的结构与细节信息; 其次, 构建自适应频带特性建模模块, 联合幅度与相位描述估计各基础频带内模态特定、模态共享和身份判别映射的相对权重, 并依据权重分布调节频谱响应, 实现不同频带信息的差异化建模; 最后, 提出频带特性引导损失, 根据频带的跨模态一致性和身份可分性构造软特性权重目标, 对特性权重估计进行监督, 引导网络学习面向跨模态身份匹配的特性权重分布。在 3 个公开数据集上与当前主流方法进行对比, 结果表明, 所提方法在 SYSU-MM01 数据集全搜索模式下的 R-1 和 mAP 分别达到 78.39% 和 74.36%, 验证了所提方法的有效性。

关键词: 行人重识别; 跨模态; 频率感知; 频域建模

中图分类号: TP391

文献标志码: A

Frequency-aware discriminative feature learning for cross-modality person re-identification

CHENG Deqiang¹, NIU Kaikai¹, HAO Dashuai¹, SONG Tianshu¹, KOU Qiqi²

1. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

2. School of Computer Science and Technology/School of Artificial Intelligence, China University of Mining and Technology, Xuzhou 221116, China

Abstract: To address the lack of joint supervision of frequency-band modulation weights based on cross-modal consistency and identity separability in existing frequency-domain methods for visible-infrared person re-identification (VI-ReID), this paper proposes a frequency-aware discriminative feature learning method for cross-modality person re-identification. First, a Phase-Guided Spectral Modulation (PGSM) module is designed to construct a shared reference amplitude from visible and infrared amplitude statistics. It adaptively adjusts amplitude responses under the guidance of phase structure and aggregates low-frequency structural information with high-frequency details, reducing inter-modality amplitude discrepancies while preserving identity-related structure and details. Second, an Adaptive Frequency-Band Characteristic Modeling (AFCM) module is developed to jointly exploit amplitude and phase descriptors to estimate the relative weights of modality-specific, modality-shared, and identity-discriminative mappings within each frequency band. Spectral responses are then adjusted according to the resulting weight distributions, enabling differentiated modeling across frequency bands. Finally, a Frequency-Band Characteristic-Guided Loss (FBGL) is introduced to construct

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

通信作者: 寇旗旗, kouqiqi@cumt.edu.cn

基金项目: 国家重点研发计划“政府间国际科技创新合作”重点专项(No.2024YFE0112000); 国家自然科学基金面上项目(No.52674225); 深地科学与工程云龙湖实验室资助项目(No.104024005)

Foundation Items: National Key R&D Program of China: Intergovernmental International Science and Technology Innovation Cooperation(No.2024YFE0112000), The National Natural Science Foundation of China (No.52674225), Yunlong Lake Laboratory of Deep Underground Science and Engineering Project (No.104024005)

soft characteristic-weight targets based on the cross-modal consistency and identity separability of each frequency band. These targets supervise characteristic-weight estimation, guiding the network to learn weight distributions for cross-modality identity matching. Comparisons with mainstream methods on three public datasets demonstrate that the proposed method achieves a R-1 accuracy of 78.39% and a mAP of 74.36% under the all-search setting on SYSU-MM01, validating its effectiveness.

Key words: person re-identification, cross-modality, frequency-aware learning, frequency-domain modeling

0 引言

行人重识别^[1] (Person Re-identification, ReID) 旨在从非重叠摄像机视角中检索具有相同身份的行人目标^[2], 是智能视频监控和公共安全等领域的重要技术。传统 ReID 主要面向可见光图像, 依赖颜色、纹理、衣着外观及局部结构等视觉信息进行身份匹配。然而在夜间、弱光及复杂照明环境下, 可见光成像质量明显下降^[3], 而红外摄像机在低照度条件下仍能够获取较稳定的人体轮廓与结构信息, 因此可见光-红外行人重识别 (Visible-Infrared Person Re-identification, VI-ReID) 逐渐成为全天时行人检索的重要研究方向。由于可见光与红外图像具有不同的成像机理^[4], 同一身份在两种模式下存在显著的颜色、亮度、纹理及光谱响应差异, 同时还受到姿态和背景变化等因素影响, 导致跨模态类内差异明显增大。因此, 如何在减小模态差异的同时保持身份判别信息是 VI-ReID 面临的核心问题。

针对上述问题, 现有研究主要从图像生成^[5]、特征对齐^[6-7]和辅助模态构造^[8]等方面展开。Qian 等^[9]通过中间模态生成减小可见光与红外图像的外观差异; Shi 等^[10]通过自模仿学习减小模态内差异, 再利用互蒸馏学习实现跨模态特征对齐。进一步地, 部分方法通过局部特征学习、语义对齐和模态特定信息挖掘提升身份判别能力^[11]。例如, Yu 等^[12]通过构建辅助模态联合建模模态共享与模态特定表征, Ren 等^[13]进一步从模态特定特征中挖掘身份相关判别信息, 以避免跨模态对齐过程中有用身份线索被过度抑制。Xu 等^[14]结合空间-通道注意力与门控融合增强身份特征, 并利用互惠邻域中的距离与秩次信息进行特征动态校准, 改善跨模态特征分布。上述方法能够有效缓解模态差异并增强特征判别性, 但主要围绕空间特征及嵌入空间展开, 对可见光与红外模态差异在频率域中的表征特性利用仍相对有限。

近年来, 频域建模^[15]为可见光-红外跨模态表

征学习提供了新的研究视角。Li 等^[16]提出 FDMNet, 通过实例自适应幅度过滤和相位保持归一化, 在图像与特征层面减弱模态相关幅度差异; Zhang 等^[17]提出 FDNM, 通过显著性引导的相位增强与幅度细微信息挖掘学习跨模态判别表征; Gu 等^[18]提出 MFENet, 利用多频率分解与高低频调制增强不同频率成分的表达。Yu 等^[19]提出 WFLNet, 通过小波频率特征学习与跨尺度高低频信息交互增强结构和细节表征; Jin 等^[20]提出 FDSIFA, 结合频域与空间信息进行特征聚合, 增强不同尺度特征的语义关联; Zhong 等^[21]结合语义可靠性与不确定性约束改善跨模态表征。上述方法通过频谱过滤、频率成分增强及特征融合, 有效提升了跨模态特征的表达能力, 但对基于同身份跨模态差异与异身份可分性的频带调制权重显式监督, 仍缺少充分研究。同一频带可同时包含模态相关响应和身份判别线索, 其跨模态一致性与身份可分性共同影响该频带的权重分配。因此, 本文提出一种频率感知判别特征学习的跨模态行人重识别方法, 在各基础频带内构建模态特定、模态共享和身份判别三类映射, 通过幅度与相位联合描述预测各映射的相对权重, 并利用同身份跨模态中心距离与异身份中心间隔构造软目标引导权重学习, 实现跨模态身份匹配的频带自适应调制。具体而言, 构建相位引导频谱调制模块 (Phase-Guided Spectral Modulation, PGSM), 利用跨模态幅度统计与相位结构调节频谱响应, 并融合低、高频信息保持结构与细节; 设计自适应频带特性建模模块 (Adaptive Frequency-Band Characteristic Modeling, AFCM), 联合建模频带的幅度响应与相位结构, 自适应调节不同频带的频谱响应; 提出频带特性引导损失 (Frequency-Band Characteristic-Guided Loss, FBGL), 利用跨模态一致性和身份可分性监督特性权重, 约束频带特性建模。本文主要贡献如下:

1) 提出一种频率感知判别特征学习的跨模态行人重识别方法。从频谱响应调制和频带特性建模

两方面开展频域特征学习,将频谱响应调制、频带映射组合与权重监督相结合,联合建模频带的跨模态响应差异与身份判别作用。

2) 设计相位引导频谱调制模块 (PGSM)。构建基于跨模态幅度统计的共享参考幅度,联合幅度偏差与相位结构进行频谱调制,减小模态差异并保持结构信息;进一步融合低频结构与高频细节,增强身份相关信息的表征能力。

3) 构建自适应频带特性建模模块 (AFCM)。针对不同频带的表征差异,通过幅度与相位频带描述及双向信息交互构建联合频谱表征,并利用模态特定、模态共享和身份判别特性权重加权相应映射响应,提升频带差异化表征能力。

4) 提出频带特性引导损失 (FBGL)。利用同身份跨模态频带中心距离与不同身份频带中心间距生成特性监督信号,对 AFCM 估计的特性权重进行约束,使频带特性建模与跨模态身份匹配目标建立联系,增强频域特征的判别性。

1 本文方法

1.1 整体网络结构

整体网络结构如图 1 所示,所提方法采用 ResNet-50^[22]作为主干网络,并采用双流结构分别处理可见光与红外图像。考虑到两种模态在低层外观信息上存在较大差异,可见光分支和红外分支分别利用独立的 Layer0 提取模态特定的浅层特征;随后将两类特征输入参数共享的 Layer1~Layer4,逐步学习跨模态共享的高层语义表征。为进一步利用可见光与红外特征在频率域中的差异信息,在共享主干的中间层引入 PGSM 和 AFCM。具体而言,分别在 Layer2 和 Layer3 之后引入 PGSM,通过跨模态

幅度统计与相位结构对频谱响应进行自适应调节,并结合低、高频信息增强结构与局部细节表征;在第二个 PGSM 之后进一步引入 AFCM,根据不同基础频带的频域表征计算模态特定、模态共享和身份判别特性权重,并依据特性权重分布进行频谱调制。经 AFCM 增强后的特征继续送入 Layer4 进行高层语义特征提取。经过 Layer4 后,采用广义均值池化对空间特征进行全局聚合,并通过批归一化层获得最终的全局特征。最后,通过身份损失、跨模态双中心约束损失和频带特性引导损失联合优化网络。

1.2 相位引导频谱调制模块

可见光与红外图像的成像差异表现为不同的频谱响应,其中幅度谱反映各频率分量的响应强度,相位信息与轮廓及空间结构密切相关。FDMNet 通过幅度过滤及相位保持归一化减弱模态差异,FDNM 结合相位增强与幅度细微信息挖掘学习判别表征。本文从幅度调整的统计参考与调制条件出发,设计相位引导频谱调制模块 (PGSM)。该模块分别聚合两种模态的幅度统计,经等权融合与指数滑动平均构建共享参考幅度;根据当前样本与共享参考的关系生成候选幅度,并联合幅度相对偏差与相位结构估计门控权重,自适应融合原始幅度与候选幅度。共享参考确定幅度调整的统计基准,样本偏差与相位结构共同决定调整程度,随后通过低、高频特征融合增强结构与细节表达。模块结构如图 2 所示,实现过程如下:

给定输入特征 $X \in R^{C \times H \times W}$, 其中 C 、 H 和 W 分别表示特征通道数、高度和宽度。首先对 X 进行二维傅里叶变换,并通过频谱中心化操作将零频分量移至频谱中心,得到复数频谱 Z ,将其分解为幅

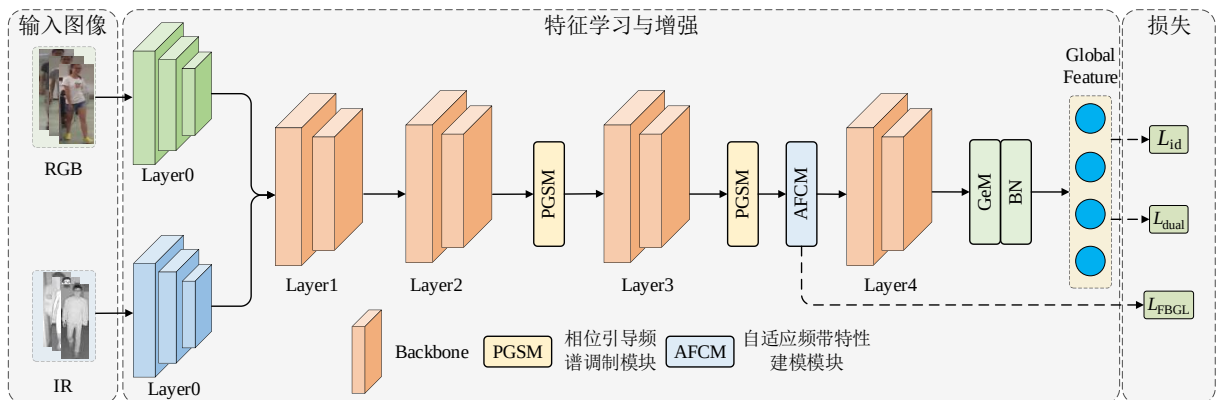


图1 整体网络结构图



其中, $\odot$ 表示逐元素乘法; ε 为保证数值稳定性的微小常数。设第 c 个通道的重标定比例为 $r_c = \frac{\mu_{s,c}}{(\mu_{g,c} + \varepsilon)}$, 则频率位置 k 处的相对幅度偏差 $\Delta_c^A(k)$

可表示为:

$$\Delta_c^A(k) = |1 - r_c| \frac{A_c(k)}{A_c(k) + \varepsilon} \quad (5)$$

其中 k 表示频率位置。对于幅度明显大于 ε 的频率响应, 该偏差近似为 $|1 - r_c|$, 因此幅度偏差主要提供通道级调整信息。为结合不同频率位置的结构特征确定调制程度, 进一步引入式 (1) 中的相位单位向量。分别对幅度偏差和相位结构进行轻量编码, 得到幅度偏移描述 D^A 和相位结构描述 D^P 并联合两类信息估计频谱调制权重 G :

$$\begin{cases} D^A = \varphi_{1 \times 1}^{2d}(\Delta^A) \\ D^P = \varphi_{1 \times 1}^{2d}(U) \\ G = \sigma[\varphi_{1 \times 1}^{2d}(D^A // D^P)] \end{cases} \quad (6)$$

其中, $\varphi_{1 \times 1}^{2d}(\cdot)$ 表示二维 1×1 卷积操作, $//$ 表示特征拼接操作, $\sigma(\cdot)$ 表示 Sigmoid 激活函数。随后将 A 与 A^c 按照 G 进行自适应融合, 得到调制后的幅度谱 $\tilde{A}$ 。随后结合原有相位结构重构复数频谱 $\tilde{Z}$, 并通过二维逆傅里叶变换恢复到空间域得到频谱调制后的空间特征 X^P , 计算过程如下:

$$\begin{cases} \tilde{A} = A + G \odot (A^c - A) \\ \tilde{Z} = \tilde{A} \odot (U^R + jU^I) \\ X^P = X + \alpha \left\{ \text{Re}[\mathcal{F}^{-1}(\tilde{Z}) - X] \right\} \end{cases} \quad (7)$$

其中, $\tilde{A}$ 和 $\tilde{Z}$ 分别表示调制后的幅度谱和复数频谱, $\mathcal{F}^{-1}(\cdot)$ 表示恢复原始频率排列后进行二维逆傅里叶变换, $\text{Re}(\cdot)$ 表示取实部, j 为虚数单位, α 为可学习残差系数。上式通过逆傅里叶变换后的实部投影实现实值重构。为说明其频域含义, 在恢复原始频率排列后, 记 k 与 $\bar{k}$ 为一对共轭频率位置, 取实部等价于对调制频谱进行共轭对称化, 其对应的有效频谱及重构关系为:

$$\begin{cases} \tilde{Z}_H(k) = \frac{\tilde{Z}(k) + \tilde{Z}^*(\bar{k})}{2} \\ \mathcal{F}^{-1}(\tilde{Z}_H) = \text{Re}[\mathcal{F}^{-1}(\tilde{Z})] \end{cases} \quad (8)$$

$$\begin{cases} \mathcal{L}(X) = \text{ReLU} \left[\text{BN} \left(\varphi_{\text{DWC}3 \times 3}^{2d}(\varphi_{1 \times 1}^{2d}(X)) \right) \right] \\ \mathcal{H}(X^H) = \varphi_{1 \times 1}^{2d} \left[\sum_{l=1}^4 g \left(\mathcal{L}^l(\varphi_{1 \times 1}^{2d}(X^H)) \right) \odot \mathcal{L}^l(\varphi_{1 \times 1}^{2d}(X^H)) \right] \end{cases} \quad (11)$$

其中, $\mathcal{H}(\cdot)$ 表示由不同感受野卷积构成的多尺度细节增强函数, $\varphi_{3 \times 3}^{\text{DWC}}(\cdot)$ 表示二维 3×3 深度卷积操

其中, $\tilde{Z}_H$ 为等价重构频谱, 满足共轭对称性, 其逆变换为实值; $(\cdot)^*$ 表示复共轭。对于 PGSM, 原始幅度及其通道级重标定结果在共轭频率对上分别相等, 且调制过程沿用原相位, 上述等价频谱对应的有效门控为:

$$G_c^{\text{eff}}(k) = \frac{G_c(k) + G_c(\bar{k})}{2} \quad (9)$$

其中, c 为通道索引, $G_c(k)$ 为第 c 个通道在频率位置 k 处的原始门控权重, $G_c^{\text{eff}}(k)$ 为实部投影所对应的有效门控权重, 其在共轭频率对上取值相等。由此, PGSM 的实部重构等价于使用共轭频率对上的平均门控调节幅度, 同时沿用原相位结构, 实现共轭对称有效频谱的实值重构。通过频谱调制权重自适应融合原始幅度与候选幅度, 使当前相位结构引导幅度响应调整, 并通过残差形式保持特征传递的稳定性。经过上述频谱调制后, 特征中仍包含不同尺度的频率响应。为进一步整合整体结构与局部细节信息, 以中心化频谱的零频位置为中心, 根据归一化频率距离构造低频掩码 M_L , $M_H = 1 - M_L$ 为高频掩码, 将调制后的频谱 $\tilde{Z}$ 分解为低频 $Z_L = \tilde{Z} \odot M_L$, 高频 $Z_H = \tilde{Z} \odot M_H$ 两部分。进一步通过逆傅里叶变换分别获得低频结构特征 F^L 和细节特征 F^H 。低、高频分支均取逆傅里叶变换的实部作为空间特征。由于实值掩码在共轭频率对上取值相等, 该过程等价于先对调制频谱进行共轭对称化, 再按掩码分解与重构。其中, 高频分量采用 4 个不同深度的轻量卷积分支进行多尺度特征提取, 进一步聚合不同感受野下的局部细节信息, 并通过共享通道门控对各尺度响应进行自适应加权, 得到高频细节特征 F^H 。计算公式如下:

$$\begin{cases} F^L = \text{Re}[\mathcal{F}^{-1}(Z_L)] \\ F^H = \mathcal{H}[X^H], X^H = \text{Re}[\mathcal{F}^{-1}(Z_H)] \end{cases} \quad (10)$$

作, $\mathcal{L}^l(\cdot)$ 表示连续堆叠 l 个轻量卷积单元, $g(\cdot)$ 表示各尺度共享的通道门控函数, $\text{BN}(\cdot)$ 表示对

卷积输出做批量归一化, $\text{ReLU}(\cdot)$ 表示修正线性单元激活函数。随后根据调制后的空间特征、幅度

偏移描述和相位结构描述估计两类频率特征的融合权重, PGSM 最终输出特征 Y^{PGSM} 表示为:

$$\begin{cases} [G^L, G^H] = \sigma_s \{ \text{MLP} [\text{GAP}(X^P) // \text{GAP}(D^A) // \text{GAP}(D^P)] \} \\ Y^{\text{PGSM}} = X^P + \eta \text{BN}(G^L \odot F^L + G^H \odot F^H) \end{cases} \quad (12)$$

其中, σ_s 表示 Softmax 函数, $\text{MLP}(\cdot)$ 表示由两层全连接层及非线性激活组成的多层感知机, 用于对拼接后的全局描述进行通道映射, G^L 和 G^H 分别为低、高频分支的融合权重, $\text{BN}(\cdot)$ 表示批归一化操作, $\text{GAP}(\cdot)$ 表示全局平均池化操作, η 为可学习残差系数。空间特征与频谱描述共同参与两类频率分量的权重估计, 使结构响应与局部细节能够根据当前输入进行自适应聚合。综上, PGSM 首先从输入特征中提取幅度和相位结构, 并基于两种模态的幅度统计建立共享参考; 在此基础上, 通过候选幅度与原始幅度之间的响应偏差刻画频谱变化, 再结合相位结构估计频谱调制权重, 完成幅度响应的自适应调节。最后, 通过低、高频特征聚合进一步补充不同尺度的结构与细节信息, 得到 PGSM 增强特征并送入后续网络进行细粒度频带表征学习。

1.3 自适应频带特性建模模块

经过频谱调制和主干网络特征提取后, 不同频带在跨模态响应差异和身份区分能力方面呈现不同特征。其中, 模态特定性与模态共享性描述频带的跨模态响应关系, 身份判别性描述频带对身份区分的作用。同一频带可以同时包含模态相关响应和身份判别线索, 因此, 本文通过三类映射的加权组合联合建模这些特性。MFENet 通过多频率分解与高低频调制学习频率表征, WFLNet 通过小波频率特征学习增强结构与细节信息, FDSIFA 结合频域信息与空间注意力增强特征响应。本文进一步以每个

基础频带内部的映射组合为建模对象, 设计自适应频带特性建模模块, 结构如图 3 所示。AFCM 首先依据频率位置划分基础频带, 分别提取幅度与相位描述, 并通过双向信息交互构建联合频带表征; 随后预测模态特定、模态共享和身份判别三类映射的归一化权重, 对各映射生成的调制响应进行加权组合。基础频带确定频谱的处理区域, 三类权重确定该区域内不同映射的相对贡献, 使同一频带的调制方式能够随输入表征变化。FBGL 进一步利用频带中心的距离关系监督权重学习, 使映射组合与跨模态一致性、身份可分性建立联系。具体过程如下:

对于 AFCM 的输入特征 $F \in R^{C \times H \times W}$, 首先进行二维傅里叶变换和频谱中心化得到频域表示 $S = \mathcal{F}(F)$, 并提取其幅度谱 $B = |S|$ 。相位结构按式 (1) 相同的单位复向量形式由 S 构造, 将其实、虚分量统一记为二维相位向量 $V = [V^R, V^I]$, 按照归一化频率距离将中心化频谱等间隔划分为 K 个由低至高的基础频带, 对应频带掩码记为 $\{M_k\}_{k=1}^K$, 第 k 个频带的幅度响应 B_k 和相位响应 V_k 计算公式为:

$$\begin{cases} B_k = B \odot M_k \\ V_k = V \odot M_k \end{cases} \quad (13)$$

将整体频谱通过频带划分为多个基础频率区域, 为后续分析各频带的幅度响应与相位结构提供细粒度表示。获得频带响应后, 对各频带进行频带内统计。幅度分支对 B_k 在对应频带区域进行平均

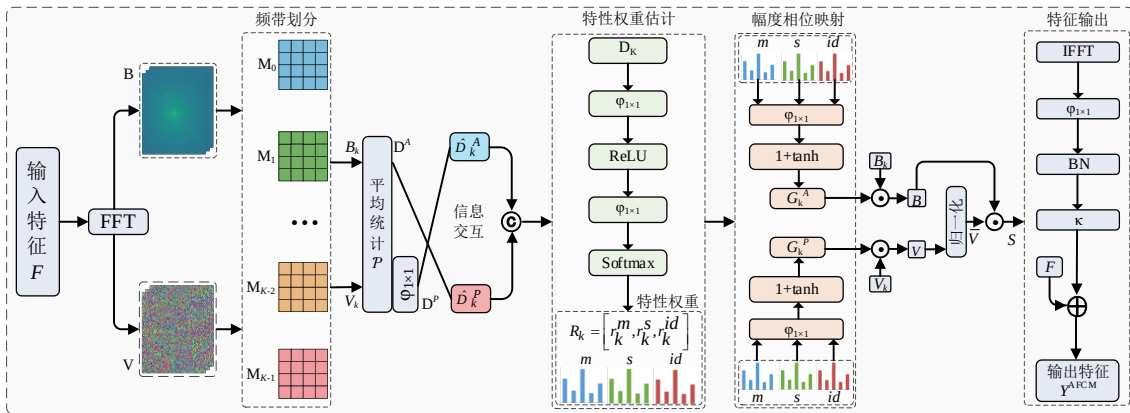


图3 AFCM 模块结构图

聚合得到幅度频带描述 D_k^A ; 对于相位分支分别统计实、虚分量, 采用核为1的一维卷积进行通道投影得到与幅度描述维度一致的相位频带描述 D_k^P :

$$\begin{cases} D_k^A = P_k(B_k) \\ D_k^P = \varphi_{1 \times 1}[P_k(V_k)] \end{cases} \quad (14)$$

$$P_k(Q_k) = \frac{\sum_{h,w} Q_k(:, h, w)}{\sum_{h,w} M_k(h, w) + \varepsilon} \quad (15)$$

其中, $\varphi_{1 \times 1}(\cdot)$ 表示核大小为1的一维卷积操作, $P_k(\cdot)$ 表示对频带区域内的特征响应进行平均统计, Q_k 表示第 k 个频带上的待统计特征响应, 由此, 各频带的响应被统一表示为幅度强度和相位结构两个互补维度。幅度与相位描述分别表征频带的响应强度和结构分布。为在特性权重估计前建立二者的关联, 本文通过双向信息交互, 使幅度描述融合相位结构信息, 同时使相位描述融合幅度响应信息, 得到增强后的幅度描述与相位描述:

$$\begin{cases} \hat{D}_k^A = D_k^A \odot [1 + \zeta \sigma(\varphi_{1 \times 1}(D_k^P))] \\ \hat{D}_k^P = D_k^P \odot [1 + \zeta \sigma(\varphi_{1 \times 1}(D_k^A))] \end{cases} \quad (16)$$

其中, $\square$ 为可学习跨分支交互系数。双向交互使两类描述在保留各自信息的基础上引入另一分支的表征, 随后通过拼接构成联合频带特征 $D_k = \hat{D}_k^A \parallel \hat{D}_k^P$, 为模态特定、模态共享和身份判别三类映射的权重估计提供依据。进一步采用由两层核大小为1的一维卷积和 ReLU 激活组成的特性权重估计网络, 对联合描述进行映射, 并通过 Softmax 得到第 k 个频带的特性权重向量 R_k :

$$R_k = \sigma_s \left\{ \varphi_{1 \times 1} \left[\text{ReLU} \left(\varphi_{1 \times 1}(D_k) \right) \right] \right\} = [r_k^m, r_k^s, r_k^{id}] \quad (17)$$

其中, $r_k^m + r_k^s + r_k^{id} = 1$, r_k^m 、 r_k^s 和 r_k^{id} 分别表示模态特定、模态共享和身份判别特性权重。三类特性权重表示对应映射在当前频带调制中的相对参与程度。Softmax 在同一频带的三个特性分支之间进行归一化, 使各分支以非负且和为1的权重共同参与调制。该约束使分支权重相互关联, 一个分支权重增大时, 其余分支的权重之和相应减小, 适用于不同映射之间的相对贡献分配。结合式 (18), 各分支通过独立的可学习参数生成调制响应, 并依据归一化权重进行加权组合, 使三类映射按照各自的相

对贡献共同参与当前频带的调制。获得频带特性权重分布后, 针对不同特性分别建立幅度和相位响应映射。对于特性分支 $\rho \in \{m, s, id\}$, 分别通过卷积映射生成调制响应。结合特性权重计算第 k 个频带的幅度调制系数 G_k^A 和相位调制系数 G_k^P :

$$\begin{cases} G_k^A = \sum_{\rho \in \{m, s, id\}} r_k^\rho \left[1 + \tanh \left(\varphi_{1 \times 1, \rho}^A(D_k) \right) \right] \\ G_k^P = \sum_{\rho \in \{m, s, id\}} r_k^\rho \left[1 + \tanh \left(\varphi_{1 \times 1, \rho}^P(D_k) \right) \right] \end{cases} \quad (18)$$

其中, $\varphi_{1 \times 1, \rho}^A(\cdot)$ 和 $\varphi_{1 \times 1, \rho}^P(\cdot)$ 分别表示特性分支 ρ 对应的核大小为1的一维卷积映射, 不同特性分支采用独立参数, $G_k^P = [G_k^{P,R}, G_k^{P,I}]$ 表示第 k 个频带的二维相位调制系数, 其两个分量分别作用于相位向量的实部与虚部, $\tanh(\cdot)$ 为双曲正切函数, 将残差调制响应限制在有限范围内。进一步根据 G_k^A 和 G_k^P 对各频带的幅度响应和相位响应进行加权聚合, 得到重构幅度谱 $\tilde{B}$ 和相位向量 $\tilde{V}$ 。

$$\begin{cases} \tilde{B} = \sum_{k=1}^K G_k^A \odot B_k \\ \tilde{V} = \sum_{k=1}^K G_k^P \odot V_k \end{cases} \quad (19)$$

由于各频带的相位响应经过独立调制, 对调制后的相位向量进行单位归一化, 恢复单位相位表示, 并根据实、虚分量重构复数频谱 $\tilde{S}$:

$$\begin{cases} \bar{V} = \frac{\tilde{V}}{\|\tilde{V}\|_2 + \varepsilon} \\ \tilde{S} = \tilde{B} \odot (\bar{V}^R + j\bar{V}^I) \end{cases} \quad (20)$$

其中, $\bar{V}$ 为单位归一化后的相位向量, $\|\cdot\|_2$ 表示相位实、虚分量上的二范数。最后对重构频谱进行二维逆傅里叶变换, 并利用卷积和批归一化完成输出映射得到 AFCM 输出为:

$$Y^{\text{AFCM}} = F + \kappa \text{BN} \left[\varphi_{1 \times 1}^{2d} \left(\mathcal{F}^{-1}(\tilde{S}) \right) \right] \quad (21)$$

其中, κ 为可学习残差系数。

1.4 损失函数

本文使用身份损失^[23]、跨模态双中心约束损失和频带特性引导损失对网络进行联合优化。其中, 身份损失从类别监督层面增强行人身份特征的判别性, 跨模态双中心约束损失进一步减小同一身份在可见光与红外模态间的特征分布差异; 频带特

性引导损失则结合 AFCM 输出的频带特性权重, 对三类特性权重进行监督。

身份损失: 为增强不同身份特征的类间可分性, 本文采用交叉熵损失监督网络学习身份判别特征。身份损失 L_{id} 计算过程为:

$$L_{id} = -\frac{1}{N_c} \sum_{i=1}^{N_c} \log(p(y_i|x_i)) \quad (22)$$

其中, N_c 表示参与身份分类的样本数量。 x_i 为行人图像的输入, y_i 为其对应的身份标签, $p(y_i|x_i)$ 为输入图像和其对应的身份标签经过 Softmax 函数分类后识别为 y_i 的预测概率。

跨模态双中心约束损失: 由于可见光图像和红外图像来源于不同成像机制, 同一身份在两种模态下存在较大的特征分布差异。仅依赖身份分类损失容易使模型关注模态特定的判别线索, 难以充分保证跨模态特征的一致性。因此, 本文引入跨模态双中心约束损失, 通过另一模态的同身份特征中心对当前模态样本进行约束。

设 f_i^v 和 f_i^r 分别表示第 i 个样本的可见光特征和红外特征, Ω_y^v 和 Ω_y^r 分别表示当前批次中属于身份 y 的可见光样本集合和红外样本集合, 则身份 y 在两种模态下的特征中心 c_y^q 定义为:

$$c_y^q = \frac{1}{|\Omega_y^q|} \sum_{i \in \Omega_y^q} f_i^q \quad (23)$$

其中, $|\cdot|$ 表示集合中样本数量。为了实现跨模态中心约束, 本文将可见光特征拉向同一身份的红外中心, 同时将红外特征拉向同一身份的可见光中心。跨模态双中心约束损失 L_{dual} 定义为:

$$L_{dual} = \frac{1}{2N} \left(\sum_{i=1}^{N_v} \|f_i^v - c_{y_i}^r\|_2 + \sum_{i=1}^{N_r} \|f_i^r - c_{y_i}^v\|_2 \right) \quad (24)$$

其中, N_v 和 N_r 分别表示当前批次中可见光样本和红外样本的数量, $N = N_v + N_r$, $c_{y_i}^v$ 与 $c_{y_i}^r$ 分别表示与第 i 个样本身份标签 y_i 对应的可见光和红外特征中心, $\|\cdot\|_2$ 表示欧氏距离。通过双向中心约束, 使同一身份的可见光与红外特征在公共嵌入空间中更加紧凑, 从而缓解模态差异。

频带特性引导损失: AFCM 根据各基础频带的幅度与相位描述预测三类映射的相对权重。为使权重分配与跨模态身份匹配目标建立联系, 本文设计频带特性引导损失。FDSL^[21]通过频域双子空间学习改善跨模态表征, 本文将监督进一步落实到

AFCM 的映射组合权重: 利用同身份跨模态频带中心距离与异身份中心间隔构造特性得分, 经标准化和 Softmax 转换为软权重目标, 监督 AFCM 的权重预测。跨模态距离提供模态差异与一致性的引导, 身份间隔提供身份可分性的引导, 两者共同约束三类映射的相对参与程度。由 1.3 节可知, 第 i 个样本在第 k 个频带上的幅度增强描述和相位增强描述分别为 $\hat{D}_{i,k}^A$ 和 $\hat{D}_{i,k}^P$, 将二者进行拼接并归一化得到联合频带描述 $\bar{D}_{i,k}$, 进一步基于样本索引集合 Ω_y^q 分别聚合同一身份在两种模态下的频带响应, 得到模态频带中心 $c_{y,k}^q$ 及跨模态中心 $c_{y,k}$:

$$\begin{cases} \bar{D}_{i,k} = \text{Norm}(\hat{D}_{i,k}^A // \hat{D}_{i,k}^P) \\ c_{y,k}^q = \frac{1}{|\Omega_y^q|} \sum_{i \in \Omega_y^q} \bar{D}_{i,k}^q, c_{y,k} = \frac{1}{2} (c_{y,k}^v + c_{y,k}^r) \end{cases} \quad (25)$$

其中, $\text{Norm}(\cdot)$ 表示 L_2 归一化操作, $\bar{D}_{i,k}^q$ 表示模态 q 中第 i 个样本在第 k 个频带上的归一化联合频带描述。为从模态差异、跨模态一致性和身份可分性三个角度构造频带特性权重的引导信号, 首先以同一身份在可见光与红外模态下的频带中心距离 $d_{y,k}^{cm}$ 衡量跨模态差异。较大的跨模态距离表明该频带具有较明显的模态特定性, 而较小的距离则表明其具有较高的跨模态一致性。进一步以跨模态频带中心与最近不同身份频带中心之间的距离 $d_{y,k}^{inter}$ 衡量类间分离程度, 并结合该频带的同身份跨模态中心距离 $d_{y,k}^{cm}$ 构造身份判别得分:

$$\begin{cases} d_{y,k}^{cm} = \|c_{y,k}^v - c_{y,k}^r\|_2 \\ d_{y,k}^{inter} = \min_{z \neq y} \|c_{y,k} - c_{z,k}\|_2 \\ s_{y,k}^m = d_{y,k}^{cm} s_{y,k}^s = -d_{y,k}^{cm} \\ s_{y,k}^{id} = d_{y,k}^{inter} - d_{y,k}^{cm} \end{cases} \quad (26)$$

其中, z 表示当前批次中不同于身份 y 的其他身份。较大的 $s_{y,k}^{id}$ 表明该频带在保持较低跨模态差异的同时具有较强的身份区分能力。根据上述三类特性分支得分, 进一步构造身份 y 在第 k 个频带上的软特性权重目标 $T_{y,k}$, 其计算公式为:

$$\begin{aligned} T_{y,k} &= \text{sg} \left\{ \sigma_s \left[\mathcal{N}(s_{y,k}^m), \mathcal{N}(s_{y,k}^s), \mathcal{N}(s_{y,k}^{id}) \right] \right\} \\ &= [t_{y,k}^m, t_{y,k}^s, t_{y,k}^{id}], (t_{y,k}^m + t_{y,k}^s + t_{y,k}^{id} = 1) \end{aligned} \quad (27)$$

其中, $\mathcal{N}(\cdot)$ 表示批内标准化操作, 对当前批次中同一特性分支的所有频带得分分别进行零均值、单

位方差标准化, $\text{sg}(\cdot)$ 表示停止梯度操作; $t_{y,k}^m$ 、 $t_{y,k}^s$ 和 $t_{y,k}^{id}$ 分别表示身份 y 在第 k 个频带上的对应模态特定、模态共享和身份判别特性的软目标概率。频带特性软目标由跨模态距离和身份间分离关系构造, 彼此存在耦合。模态特定得分与模态共享得分采用同一跨模态距离的相反方向, 分别提供模态差异倾向与跨模态一致性倾向的权重引导。身份判别得分则结合异身份中心间隔与同身份跨模态距离, 提供身份可分性方面的引导。上述得分作为特性的代理指标, 共同确定三类映射的目标权重。经标准化和 Softmax 归一化后, 模态特定与模态共享目标的相对比例由跨模态距离得分决定, 两者在整体分配中的份额同时受到身份判别得分的影响。因此, FBGL 利用相关距离指标监督三类映射的相对组合, 使频带调制与跨模态身份匹配目标建立联系。

由 1.3 节可知, 对于第 i 个身份标签为 y_i 的样本, AFCM 在第 k 个频带上预测的特性权重为 $R_{i,k} = [r_{i,k}^m, r_{i,k}^s, r_{i,k}^{id}]$, 以身份 y_i 对应的 $T_{y_i,k}$ 对频带特性权重进行约束, 则 L_{FBGL} 计算公式为:

$$L_{\text{FBGL}} = -\frac{1}{NK} \sum_{i=1}^N \sum_{k=1}^K \sum_{\rho} \lambda_{\rho} t_{y_i,k}^{\rho} \log(r_{i,k}^{\rho} + \varepsilon) \quad (28)$$

其中, N 表示批次中的样本数量, $\rho \in \{m, s, id\}$ 表示模态特定、模态共享和身份判别三类频带特性; $t_{y_i,k}^{\rho}$ 表示第 i 个样本对应身份 y_i 在第 k 个频带上的特性分支 ρ 的软目标概率, $r_{i,k}^{\rho}$ 表示 AFCM 估计的对应特性权重; λ_{ρ} 为不同特性分支约束的平衡系数, 分别对应 λ_m 、 λ_s 和 λ_{id} , ε 为保证对数计算数值稳定性的微小常数。该损失将频带的跨模态响应差异和身份类间分离特性引入特性权重学习, 使 AFCM 获得与跨模态身份匹配目标相一致的频带表征, 并通过特性权重分布调制进一步作用于频域特征学习。综上所述, 本文方法的损失可表示为:

$$L_{\text{total}} = L_{\text{id}} + \lambda_{\text{dual}} L_{\text{dual}} + L_{\text{FBGL}} \quad (29)$$

其中, λ_{dual} 为跨模态双中心约束损失系数。

2 实验结果与分析

2.1 数据集和评价指标

本文采用三个目前主流的跨模态行人重识别公开数据集 SYSU-MM01^[24]、RegDB^[25] 和 LLCM^[26] 进行实验。SYSU-MM01 包含 491 个身份的 287628 张可见光图像和 15792 张红外图像, 包括 296 个训

练身份、99 个验证身份和 96 个测试身份划分, 三者身份互不重叠, 采用 All-search 和 Indoor-search 两种评估模式。RegDB 包含 412 个身份共 8240 张图像, 训练集与测试集各包含 206 个身份, 分别采用 Visible-to-Infrared 和 Infrared-to-Visible 两种检索模式。LLCM 面向低光照跨模态场景, 共包含 1064 个身份的 46767 张图像, 其中 713 个身份用于训练、351 个身份用于测试, 并采用双向跨模态检索协议。本文采用累积匹配特性中的 R-k 准确率和平均精度均值作为评价指标。各数据集均按标准协议评估, 其中, SYSU-MM01 采用 All-search 与 Indoor-search 单样本检索设置, 固定训练模型和红外查询集, 随机生成 10 组可见光检索库, 并报告 10 次测试的平均结果。RegDB 采用 10 组训练 / 测试身份划分, 各划分分别训练模型, 并报告双向检索的平均结果。每个方向以一种模态的全部测试图像作为查询集, 另一模态的全部测试图像作为检索库。LLCM 采用双向单样本检索设置, 查询模态的全部测试图像构成查询集, 图库模态中为每个身份的每个摄像头随机抽取一张图像, 重复抽样 10 次并报告平均结果。

2.2 实验参数

本文实验配置为 RTX 4090、Intel Xeon(R) Platinum 8352V CPU, 在 Ubuntu 20.04 LTS 的系统上采用 Python 3.8 的 PyTorch 框架。采用双流 ResNet-50 作为基础网络。除复杂度比较实验外, 输入图像均调整为 384×192 像素 (高×宽), 训练时可见光与红外输入分别采用一组增强流程^[27], 包括填充、随机裁剪、随机水平翻转和随机擦除, 并采用均值和标准差进行通道归一化。每批次随机采样 6 个身份, 每个身份包含 4 张可见光图像和 4 张红外图像, 批量大小为 48。采用 SGD 优化器, 动量为 0.9, 权重衰减为 5×10^{-4} , 初始学习率为 0.01; 前 10 轮采用预热策略, 在第 25、70 和 90 轮分别将学习率衰减至此前的 0.1 倍, 共训练 150 轮。本文通过 SYSU-MM01 独立验证集确定频带数量及损失权重。其中, 频带数量设为 5, 跨模态双中心约束损失权重为 0.95, FBGL 三项监督系数分别为 0.10、0.15 和 0.10。参数固定后, 将训练集与验证集合并, 使用 395 个身份训练模型, 并在 96 个测试身份上进行性能评价。

2.3 与现有方法对比

为评价所提方法的跨模态检索性能, 本文在 SYSU-MM01、RegDB 和 LLCM 三个数据集上, 与空间特征学习、模态对齐及频域建模等代表性方法进行比较。频域方法包括 FDMNet^[16]、FDNM^[17]、MFENet^[18]、WFLNet^[19] 等。SYSU-MM01 和 RegDB 上的结果见表 1, LLCM 上的结果见表 2。表中加粗和下划线分别表示所列结果中的最优值和次优值; †表示本文复现结果, 保留原网络结构、损失及学习率策略及训练轮数, 数据增强、采样方式和优化器等条件统一采用第 2.2 节设置, 并按标准协议评估。其余对比方法采用相应文献报告的结果。

在 SYSU-MM01 的 All-search 单样本设置下, 本文方法的 R-1 和 mAP 分别达到 78.39% 和 74.36%, 均为表 1 所列结果中的最高值。与 FDNM 相比, 两项指标分别提高 1.72% 和 0.41%; 与 MFENet 相比, 分别提高 2.33% 和 1.42%。在 Indoor-search 设置下, 本文方法取得 84.52% 的 R-1 和 86.71% 的 mAP, 其中 R-1 为表中最高值, 较 FDNM 和 MFENet 分别提高 0.32% 和 1.23%; mAP 较 MFENet 提高 0.84%, FDNM 则取得更高的 87.15%。结果表明, 本文方法在全搜索设置下兼顾首位匹配准确率与整体排序质量。

在 RegDB 的 Visible-to-Infrared 和 Infrared-to-Visible 方向上, 本文方法的 R-1/mAP 分别为 93.08%/86.49% 和 91.21%/84.36%。在 Visible-to-Infrared 方向, 较 FDNM 分别提高 0.22% 和 0.45%, 较 MFENet 分别提高 1.46% 和 0.66%, 其中 R-1 为表中次高值。在 Infrared-to-Visible 方向, FDNM 和 MFENet 取得更高的 R-1 和 mAP。FDMNet 在该数据集上取得表中最高的双向 R-1 及 Visible-to-Infrared 方向 mAP, MFENet 取得 Infrared-to-Visible 方向最高 mAP。

在低光照数据集 LLCM 上, 本文方法在 Visible-to-Infrared 和 Infrared-to-Visible 方向的 R-1/mAP 分别达到 64.24%/65.92% 和 57.31%/63.17%, 取得表 2 所列结果中的双向最高 R-1 及 Infrared-to-Visible 方向最高 mAP。与 FDNM 相比, 两个方向的 R-1 分别提高 1.88% 和 2.88%, mAP 分别提高 1.23% 和 2.64%; 与 MFENet 相比, 双向 R-1 分别提高 0.51% 和 0.40%, Infrared-to-Visible 方向 mAP 提高

0.23%。Visible-to-Infrared 方向的最高 mAP 为 MCST 的 68.30%, MFENet 为 67.76%, 高于本文的 65.92%。本文方法在两个检索方向上的首位匹配准确率均有所提升, 体现了所提频域特征学习方法在低光照跨模态匹配中的有效性。

2.4 消融实验

2.4.1 不同模块消融对比

为验证所提各模块的有效性, 本文在 SYSU-MM01 数据集 All-search 单样本检索设置下进行消融实验, 结果如表 3 所示。基础网络的 R-1 和 mAP 分别为 68.06% 和 64.23%。引入跨模态双中心约束损失 L_{dual} 后, 两项指标分别提高 1.42% 和 0.93%, 说明该约束有助于提升跨模态匹配性能。加入 PGSM 后, R-1 和 mAP 分别达到 72.31% 和 68.58%, 较索引 2 提高 2.83% 和 3.42%。PGSM 利用跨模态幅度统计建立共享参考, 结合幅度偏差与相位结构调节频谱响应, 并融合低频结构与高频细节, 为跨模态匹配提供有效的特征表示。加入 AFCM 后, 两项指标分别达到 73.05% 和 70.35%, 较索引 2 提高 3.57% 和 5.19%。AFCM 通过幅度与相位描述的双向交互, 估计各基础频带中模态特定、模态共享和身份判别映射的相对权重, 实现不同映射响应的自适应组合。两种模块联合使用时, R-1 和 mAP 进一步达到 74.82% 和 71.31%, 说明共享参考下的频谱调制与频带内映射加权具有较好的互补性。在 AFCM 的基础上加入 FBGL, R-1 和 mAP 达到 76.94% 和 73.42%, 分别提高 3.89% 和 3.07%; 在 PGSM 与 AFCM 联合结构中加入 FBGL, 两项指标达到 78.39% 和 74.36%, 分别提高 3.57% 和 3.05%。FBGL 利用同身份跨模态中心距离与异身份中心间隔构造软目标, 引导 AFCM 学习与跨模态一致性和身份可分性相联系的映射权重。上述结果表明, 对特性权重施加监督有助于提升频带自适应调制的检索性能。

2.4.2 不同频带数量对比

为分析 AFCM 中基础频带数量对频带特性建模性能的影响, 本文在 SYSU-MM01 数据集 All-Search 模式下设置不同的 K 值进行参数敏感性实验, 结果如表 4 所示。对于不同的频带数量, AFCM 仅改变频谱的基础频带划分粒度, 当 $K=3$ 时, 模型的 R-1 和 mAP 分别为 77.54% 和 73.60%。当 $K=4$ 时, R-1 和 mAP 分别提升至 77.88% 和

表 1 在 SYSU-MM01 和 RegDB 数据集上与现有方法的结果对比

方法	年份	SYSU-MM01						RegDB					
		All-Search			Indoor-Search			Visible to Infrared			Infrared to Visible		
		R-1	R-10	R-20	mAP	R-1	R-10	R-20	mAP	R-1	R-10	R-20	mAP
AGW ^[28]	TPAM21	47.50	84.39	92.14	47.65	54.17	91.14	95.98	62.97	70.05	86.21	91.55	66.37
CAJ ^[29]	ICCV21	69.88	95.71	98.46	66.89	76.26	97.88	99.49	80.37	85.03	-	-	79.14
DART ^[30]	CVPR22	68.72	96.39	98.96	66.29	72.52	97.84	99.46	78.17	83.60	-	-	75.67
DCLNet ^[31]	MM22	70.79	-	-	65.18	73.51	-	-	76.80	81.20	-	-	74.30
DSCNet ^[32]	TIFS23	73.89	96.27	98.84	69.47	79.35	98.32	99.77	82.65	85.39	-	-	77.30
PMT ^[33]	AAAI23	67.53	95.36	98.64	64.98	71.66	96.73	99.25	76.52	84.83	-	-	76.55
MTMFE ^[34]	PR23	69.47	96.42	99.11	66.41	71.72	97.19	98.97	76.38	85.04	94.38	97.22	82.52
MCBD ^[35]	TIFS23	71.63	94.98	98.37	67.28	79.44	98.32	99.91	79.85	90.10	97.09	98.35	82.73
CSL ^[36]	ICME24	72.50	-	-	68.00	80.20	-	-	82.90	86.20	-	-	77.90
MBCNet ^[37]	TCSS24	72.25	95.66	98.68	68.62	77.89	97.99	99.61	81.68	88.28	96.55	98.21	84.67
ACD ^[38]	TIFS24	74.44	-	-	71.17	78.98	-	-	82.75	84.71	-	-	83.28
RACA ^[39]	SR25	70.84	-	-	67.42	74.35	-	-	78.21	88.69	-	-	81.95
AGPI ^[25]	TIFS25	72.23	97.04	-	70.58	83.45	98.62	-	84.25	89.03	98.19	-	83.89
CSCL ^[7]	TMM25	75.72	97.14	99.18	72.08	80.80	98.14	99.42	83.56	92.16	-	-	84.27
TSKD ^[10]	PR26	76.60	97.10	99.10	73.00	82.70	98.90	99.80	85.30	91.10	-	-	81.70
MCST ^[11]	PR26	<u>77.10</u>	-	-	73.90	82.40	-	-	<u>86.90</u>	81.90	92.90	96.50	78.60
ProtoHPE ^[40]	ACM23	71.92	96.19	97.98	70.59	77.81	98.64	99.59	81.31	88.74	-	-	83.72
FDMNet ^[16]	Arxiv24	75.99	<u>97.63</u>	-	70.71	80.92	98.88	-	82.64	95.92	99.01	-	89.26
FDNN ^{†[17]}	TIFS25	76.67	97.91	98.99	<u>73.95</u>	<u>84.20</u>	<u>99.09</u>	99.34	87.15	92.86	97.43	98.65	86.04
FDSL ^[21]	OLT26	72.00	92.10	99.10	70.20	77.80	98.80	99.80	81.40	90.50	97.50	98.50	85.10
FDSIFA ^[20]	AROC26	75.09	96.68	98.93	71.35	81.01	98.62	99.74	84.09	90.63	97.62	98.83	84.67
WFLNet ^[19]	TCE26	76.68	97.45	99.32	72.26	80.56	98.87	<u>99.82</u>	83.57	92.12	-	-	<u>87.47</u>
MFENet ^{†[18]}	TCSVT26	76.06	97.19	99.05	72.94	83.29	98.13	99.75	85.87	91.62	97.07	<u>98.98</u>	85.83
Ours	-	78.39	97.41	<u>99.23</u>	74.36	84.52	99.18	<u>99.82</u>	86.71	<u>93.08</u>	<u>98.51</u>	99.06	86.49
										91.21	97.52	99.10	84.36

表2 在LLCM数据集上与现有方法的结果对比

方法	年份	Visible to Infrared		Infrared to Visible	
		R-1	mAP	R-1	mAP
AGW ^[28]	TPAMI21	51.50	55.30	43.60	51.80
CAJ ^[29]	ICCV21	56.50	59.80	48.80	56.60
MMN ^[41]	ACM21	59.90	62.70	52.50	58.90
DART ^[30]	CVPR22	60.40	63.20	52.20	59.80
DEEN ^[26]	CVPR23	62.50	65.80	54.90	62.90
MBCNet ^[37]	TCSS24	60.30	55.20	55.40	62.40
AGPI ^[25]	TIFS25	61.78	65.08	56.47	62.79
MCST ^[11]	PR26	62.30	68.30	55.20	62.50
PFMN ^[42]	ICIC25	62.90	66.20	55.60	<u>63.10</u>
FDNM ^[17]	TIFS25	62.36	64.69	54.43	60.53
WFLNet ^[19]	TCE26	61.82	63.88	51.93	58.22
MFENet ^[18]	TCSVT26	<u>63.73</u>	<u>67.76</u>	<u>56.91</u>	62.94
Ours	-	64.24	65.92	57.31	63.17

表3 SYSU-MM01 数据集消融实验结果

索引	设置					All-Search		
	PGSM	AFCM	L_{dual}	L_{FBGL}	R-1	R-10	R-20	mAP
1					68.06	94.68	97.99	64.23
2			√		69.48	95.46	98.35	65.16
3	√		√		72.31	96.32	98.50	68.58
4		√	√		73.05	96.41	98.77	70.35
5	√	√	√		74.82	96.97	99.04	71.31
6		√	√	√	76.94	97.26	99.15	73.42
7	√	√	√	√	78.39	97.41	99.23	74.36

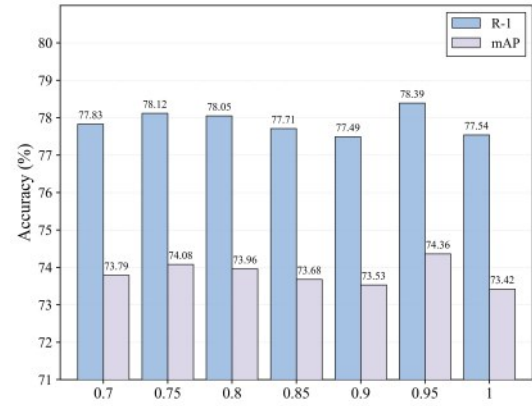
73.98%，表明适当增加基础频带数量能够提供更细粒度的频率描述，有利于 AFCM 进一步区分不同频带中的频谱响应差异。当 $K=5$ 时，模型 R-1 和 mAP 分别达到 78.39% 和 74.36%，说明该频带划分能够较好地支持不同频带的特性权重估计与频谱调制。当 K 进一步增加至 6 和 7 时，R-1 分别下降至 77.98% 和 77.86%，mAP 分别为 74.04% 和 74.01%。随着频带划分进一步细化，单个频带所覆盖的频率

区域减小，其统计描述所包含的信息相应减少，过细的频带划分可能增加特性权重估计和频带调制的学习难度，导致模型性能没有提升。

表4 不同频带数量对比结果

K	SYSU-MM01			
	R-1	R-10	R-20	mAP
3	77.54	96.90	99.03	73.60
4	77.88	97.33	99.13	73.98
5	78.39	97.41	99.23	74.36
6	77.98	97.38	99.16	74.04
7	77.86	97.35	99.14	74.01

2.4.3 损失参数消融对比

图4 不同 λ_{dual} 在SYSU-MM01数据集上实验结果

为分析跨模态双中心约束损失权重 λ_{dual} 对模型性能的影响，本文在SYSU-MM01数据集All-Search模式下设置不同的 λ_{dual} 进行参数敏感性实验，结果如图4所示。随着 λ_{dual} 增大，模型性能呈现一定波动，当 $\lambda_{\text{dual}} = 0.95$ 时R-1和mAP分别达到78.39%和74.36%。当权重较小时，跨模态中心约束作用相对有限，难以充分减小同身份可见光与红外特征间的分布差异；而当 λ_{dual} 增大至1.0时，R-1和mAP下降至77.54%和73.42%，说明过强的跨模态对齐约束可能影响身份判别信息的学习。

为进一步分析FBGL中不同损失权重对模型性能的影响，本文分别对模态特定、模态共享和身份判别特性的平衡系数 λ_m 、 λ_s 和 λ_{id} 进行参数敏感性实验，结果如表5所示。当三项权重均设为0时，模型R-1和mAP分别为74.82%和71.31%。引入FBGL的软特性权重监督后，各组设置均取得不同

表5 不同损失权重在 SYSU-MM01 对比结果

索引	设置			SYSU-MM01			
	λ_m	λ_s	λ_{id}	R-1	R-10	R-20	mAP
1	0	0	0	74.82	96.97	99.04	71.31
2	0.1	0.15	0.1	78.39	97.41	99.23	74.36
3	0.1	0.1	0.1	78.24	97.42	99.20	74.24
4	0.05	0.075	0.05	76.47	96.56	98.48	73.25
5	0.2	0.3	0.2	77.91	97.19	99.08	73.90

程度的性能提升，表明根据跨模态一致性和身份可分性构造的软特性权重目标能够有效引导 AFCM 的特性权重学习。

当 $\lambda_m = 0.10, \lambda_{id} = 0.10$ 时， $\lambda_s = 0.10$ 与 $\lambda_s = 0.15$ 两种设置的性能接近，后者的平均 R-1 和 mAP 分别高出约 0.15% 和 0.12%。针对两个固定模型在相同 10 组检索库上的测试结果进行双侧配对 t 检验，R-1 平均差值为 0.147%，95% 置信区间为 $[-0.015, 0.309]$ ， $p=0.0703$ ；mAP 平均差值为 0.117%，95% 置信区间为 $[0.053, 0.181]$ ， $p=0.0026$ 。该统计反映固定模型在检索库重复抽样条件下的差异。

表6 不同权重配置在随机种子下的性能比较

方法	种子	R-1	mAP
(0.10,0.10,0.10)	0	77.78	73.65
	1	78.24	74.24
	2	78.05	74.11
	总	78.02±0.23	74.00±0.31
(0.10,0.15,0.10)	0	77.92	73.90
	1	78.39	74.36
	2	78.26	74.22
	总	78.19±0.24	74.16±0.24

为进一步分析训练随机性的影响，本文分别采用 0、1、2 三个随机种子对上述两组配置独立训练，其余训练设置保持一致。每个模型均使用相同的 10 组可见光检索库进行评估，先计算各模型的测试平均值，再统计三个训练种子之间的均值与样本标准差，结果如表 6 所示。(0.10,0.15,0.10)在三个训练种子下的 R-1 和 mAP 高于(0.10,0.10,0.10)配置。以相同种子的结果配对，R-1 平均差值为 0.167%，95% 置信区间为 $[0.073, 0.261]$ ，双侧配对 t 检验 $p=0.0168$ ；mAP 平均差值为 0.160%，95%

置信区间为 $[-0.034, 0.354]$ ， $p=0.0711$ 。上述结果表明适当增加模态共享监督权重在本次重复训练中取得方向一致的小幅改善，其中 R-1 差异达到 0.05 显著性水平，mAP 差异未达到该水平。考虑到训练种子数量有限，本文将该结果作为所选权重配置补充证据。

2.4.4 PGSM 内部门控机制消融对比

表7 SYSU-MM01 上 PGSM 门控机制对比结果

方法	R-1	mAP
不加入 PGSM	69.48	65.16
通道幅度重标定	70.94	67.01
门控不使用相位输入	71.13	67.25
PGSM	72.31	68.58

为分析 PGSM 中通道幅度调整与相位引导门控的作用，本文比较不加入 PGSM、通道幅度重标定、门控无相位输入和完整 PGSM 四种配置，通道重标定配置对幅度偏差沿频率维求均值并广播，同时将相位输入置零；无相位配置保留逐位置幅度偏差，仅将相位输入置零；完整 PGSM 联合使用幅度偏差与相位结构估计门控权重。结果如表 7 所示，完整 PGSM 取得 72.31% 的 R-1 和 68.58% 的 mAP，较不加入 PGSM 分别提高 2.83% 和 3.42%。与通道幅度重标定相比，完整 PGSM 分别提高 1.37% 和 1.57%，表明在通道级调整的基础上，结合频率位置的结构信息确定调制程度能够进一步改善检索性能。与门控无相位输入配置相比，完整 PGSM 分别提高 1.18% 和 1.33%，进一步支持相位输入参与门控估计的作用。上述结果表明，共享参考下的幅度调整与相位结构引导相结合，能够为跨模态身份匹配提供更有效的频谱调制。

2.4.5 AFCM 内部双向信息交互机制消融对比

表8 SYSU-MM01 上 AFCM 幅度与相位交互对比结果

方法	R-1	mAP
无跨分支交互	71.36	67.84
幅度→相位	72.07	68.39
相位→幅度	72.33	68.82
AFCM	73.05	70.35

为分析 AFCM 内部幅度与相位双向交互的作用, 本文在 SYSU-MM01 数据集 All-search 单样本检索设置下比较四种交互配置, 结果如表 8 所示。各组以表 3 索引 4 为基础, 不引入 PGSM 和 FBGL, 保留幅度与相位双分支、后续拼接及频谱调制结构, 仅改变跨分支交互方向。无跨分支交互时, R-1 和 mAP 分别为 71.36% 和 67.84%。引入幅度-相位或相位-幅度单向交互后, 两项指标均有所提高, 说明在后续拼接之前建立两类描述之间的关联有助于频带表征学习。完整双向交互取得 73.05% 的 R-1 和 70.35% 的 mAP, 较无交互配置分别提高 1.69% 和 2.51%; 较幅度-相位单向交互分别提高 0.98% 和 1.96%, 较相位-幅度单向交互分别提高 0.72% 和 1.53%。双向交互使幅度响应与相位结构在描述更新阶段相互补充, 为后续特性权重估计提供联合信息。实验结果验证在频带特性建模中采用双向交互的有效性。

2.4.6 FBGL 内部监督项消融对比

表9 SYSU-MM01 上 FBGL 内部监督项消融

设置			All-Search	
模态特定监督项	模态共享监督项	身份判别监督项	R-1	mAP
×	×	×	74.82	71.31
√	√	×	76.67	72.93
×	×	√	77.04	73.25
×	√	√	77.78	73.69
√	×	√	77.60	73.42
√	√	√	78.39	74.36

为分析 FBGL 中各监督项的性能贡献, 本文在 SYSU-MM01 数据集 All-search 单样本检索设置下进行监督项组合消融, 结果如表 9 所示。各组采用

相同的网络结构、基础训练损失和训练测试设置, 通过将相应监督项系数置零实现移除, 其余监督项沿用完整模型的系数, 三类软目标的构造与归一化方式保持一致。三类软目标共享距离统计量, 预测权重通过 Softmax 联合归一化, 各监督项共同作用于映射权重的学习。本文通过逐项移除与部分组合比较, 分析各监督项在不同组合中的性能贡献。无额外频带监督时, R-1 和 mAP 分别为 74.82% 和 71.31%; 同时采用模态特定与模态共享监督项后, 两项指标提升至 76.67% 和 72.93%, 表明跨模态距离关系能够为特性权重学习提供有效引导。仅保留身份判别监督项时, 两项指标达到 77.04% 和 73.25%。该监督项通过身份判别得分同时引入异身份中心间隔和同身份跨模态距离, 使权重学习兼顾类间分离与跨模态一致性。三类监督项共同使用时, R-1 和 mAP 达到 78.39% 和 74.36%。在其余两项保留的条件下, 移除模态特定监督项, 两项指标分别下降 0.61% 和 0.67%; 移除模态共享监督项分别下降 0.79% 和 0.94%; 移除身份判别监督项分别下降 1.72% 和 1.43%。模态特定与模态共享监督项分别从跨模态差异和一致性角度引导映射权重, 身份判别监督项进一步引入类间分离信息。上述结果表明, 各监督项在完整组合中均具有正向性能贡献, 联合监督取得了所比较设置中的最佳性能。

2.4.7 不同频带监督及权重建模方式的性能比较

为进一步验证 FBGL 相较于直接距离约束的作用, 本文在 SYSU-MM01 数据集 All-search 单样本检索设置下比较无额外频带监督、直接距离约束和 FBGL 三种设置, 结果如表 10 所示。各组保持网络结构、基础训练损失及训练测试设置一致。直接距离约束参考 Liu 等^[43]提出的异质中心三元组损失, 在各频带中以同身份跨模态中心构成正样本对, 以最近异身份中心构成负样本对, 通过间隔排序约束促进同身份中心接近和异身份中心分离, 三元组间隔与附加损失权重分别设为 0.3 和 1.0。直接距离约束取得 77.01% 的 R-1 和 73.14% 的 mAP, 较无额外频带监督分别提高 2.19% 和 1.83%; 采用 FBGL 后, 两项指标达到 78.39% 和 74.36%, 较直接距离约束进一步提高 1.38% 和 1.22%。两种监督方式均利用频带中心的距离关系: 直接距离约束作用于中心间的相对位置, FBGL 则将跨模态差异和身份可分性转换为三类特性映射的软权重目标, 通过监督

AFCM的权重预测调节各映射对频谱响应的贡献。该设计将距离关系与频带自适应调制联系起来,使不同频带能够依据其表征特性调整映射权重。实验结果表明,在所比较设置下,采用距离关系引导特性权重分配能够获得更好的检索性能。

表 10 不同频带监督及权重建模方式的性能比较

方法	R-1	mAP
无额外频带监督	74.82	71.31
直接距离约束	77.01	73.14
独立 Sigmoid	77.68	73.51
归一化 Sigmoid	78.02	74.06
本文方法	78.39	74.36

为进一步分析特性权重归一化方式的影响,本文比较独立 Sigmoid、归一化 Sigmoid 和 Softmax 三种方案,结果如表 10 所示。独立 Sigmoid 对各分支分别生成门控权重;归一化 Sigmoid 将各分支输出除以其总和;Softmax 依据分支得分进行联合归一化。各组保持网络结构、频带数量、模块位置及训练测试设置一致,并采用与权重输出形式相适配的监督方式。其中,独立 Sigmoid 采用逐分支二元交叉熵损失,以式(27)构造的软目标进行监督。由于该方案同时调整了权重生成方式与监督形式,因此该比较反映权重建模及配套监督方案的整体效果。在 SYSU-MM01 数据集 All-search 单样本设置下,三种方法的 R-1/mAP 分别为 77.68%/73.51%、78.02%/74.06% 和 78.39%/74.36%。Softmax 较独立 Sigmoid 分别提高 0.71% 和 0.85%,较归一化 Sigmoid 分别提高 0.37% 和 0.30%。在所比较的方案中,采用归一化权重组合映射响应取得了较高的检索性能,本文 Softmax 方法表现最优。

2.4.8 PGSM 嵌入位置与数量对比实验

为验证 PGSM 嵌入位置与数量的合理性,本文以表 3 索引 2 为基准,分别比较 Layer1~Layer4 后的单位置嵌入,以及 Layer1 与 Layer2、Layer2 与 Layer3、Layer3 与 Layer4 后的双位置嵌入,结果如表 11 所示。在单位置嵌入配置中,PGSM 置于 Layer3 后取得最高的 R-1 和 mAP,分别为 71.30% 和 67.68%。在 Layer2 与 Layer3 后分别嵌入 PGSM 时,四项指标均取得所比较配置中的最高值,R-1 和 mAP 达到 72.31% 和 68.58%。结果表明,PGSM

的作用与嵌入阶段有关,在 Layer2 与 Layer3 进行两阶段频谱调制能够获得更好的检索性能,故本文将 PGSM 置于 Layer2 与 Layer3 后。

表 11 PGSM 不同嵌入位置与数量对比结果

Method	SYSU-MM01			
	R-1	R-10	R-20	mAP
PGSM after Layer 1	70.11	94.82	97.58	66.04
PGSM after Layer 2	71.19	95.17	97.85	67.26
PGSM after Layer 3	71.30	95.33	97.90	67.68
PGSM after Layer 4	69.82	94.69	97.43	65.71
PGSM after Layer 1+2	71.74	95.46	98.00	67.99
PGSM after Layer 2+3	72.31	96.32	98.50	68.58
PGSM after Layer 3+4	71.43	95.22	97.79	67.80

2.4.9 AFCM 嵌入位置对比实验

表 12 AFCM 不同嵌入位置对比结果

Method	SYSU-MM01			
	R-1	R-10	R-20	mAP
AFCM after Layer 1	75.96	97.04	98.78	72.21
AFCM after Layer 2	77.58	97.29	99.02	73.33
AFCM after Layer 3	78.39	97.41	99.23	74.36
AFCM after Layer 4	77.05	97.14	98.96	73.02

为分析 AFCM 嵌入位置对模型性能的影响,本文将其分别置于 ResNet-50 的 Layer1~Layer4 之后进行实验,结果如表 12 所示。当 AFCM 位于 Layer3 之后时,R-1 和 mAP 分别达到 78.39% 和 74.36%,取得最优结果。具体而言,浅层特征以纹理和局部结构信息为主,语义判别能力相对有限,不利于 AFCM 根据频带表征估计不同基础频带的特性权重分布。随着网络层次加深,特征逐步获得更丰富的语义信息,Layer3 特征兼具较丰富的结构信息和较强的语义表达,有利于频带幅度-相位信息交互及三类特性权重的动态估计。当 AFCM 置于 Layer4 之后时,R-1 和 mAP 下降至 77.05% 和 73.02%,说明过深层特征的语义抽象可能削弱部分细粒度频率信息,同时 AFCM 增强后的特征缺少后续主干网络的进一步提取,性能出现下降。因此本文最终将 AFCM 设置于 Layer3 后。

2.5 复杂度分析

为分析模型的计算复杂度, 本文比较不同方法的浮点运算量 (FLOPs) 和参数量 (Params), 结果如表 13 所示。复杂度统计统一采用 144×288 (宽 $\times$ 高) 输入, 本文方法与基准模型均在该尺寸下重新训练并测试。采用统一 Python 脚本, 根据网络层配置、特征图尺寸及算子运算规则计算理论运算量, 乘加运算按 $1\text{MAC}=2\text{FLOPs}$ 计数。FLOPs 累计可见光与红外各一张图像经过对应模态入口及共享主干的前向开销, 并计入 FFT/IFFT 理论运算量估算, 以及 BN、激活、池化和频谱逐元素运算等开销。参数量统计完整网络中的两种模态入口、共享主干及新增模块, 共享参数仅计一次。

表 13 SYSU-MM01 数据集上复杂度对比结果

方法	FLOPs/G	Params/M	R-1	mAP
Base	18.23	23.52	67.96	64.06
CAJ ^[29]	31.09	30.8	69.88	66.89
MMN ^[41]	41.38	39.6	70.60	66.90
PIC ^[44]	41.38	37.4	57.50	55.10
CSL ^[36]	36.12	35.8	72.50	68.00
DEEN ^[26]	44.16	41.2	74.70	71.80
Ours	24.68	31.7	77.82	73.91

基准模型的计算量和参数量分别为 18.23GFLOPs 和 23.52M, 完整方法分别为 24.68GFLOPs 和 31.70M。在该输入尺寸下, 本文方法的 R-1 和 mAP 分别为 77.82% 和 73.91%, 较 192×384 设置分别降低 0.57% 和 0.45%。与 DEEN 相比, 本文方法的 FLOPs 和参数量分别降低 44.11% 和 23.06%, R-1 和 mAP 分别提高 3.12% 和 2.11%, 体现了所比较设置下识别性能与计算开销的平衡。此外参考 WFLNet, 在相同硬件与测试设置下, 统计包含数据加载的连续 10 次推理测试平均耗时。基准模型与本文方法分别为 40.3s 和 48.1s。

2.6 随机性分析

为分析训练随机性对实验结果的影响, 本文对表 3 索引 3、4、6 及完整方法分别采用 0、1、2 三个随机种子独立训练, 保持各配置的超参数、模块位置及其余训练设置一致。每次训练所得模型均在 SYSU-MM01 数据集 All-search 单样本设置下, 使

表 14 SYSU-MM01 上不同训练随机种子的性能

方法	种子	R-1	mAP
Base+PGSM	0	72.48	68.62
	1	72.31	68.58
	2	71.97	68.19
	总	72.25 $\pm$ 0.26	68.46 $\pm$ 0.24
Base+AFCM	0	72.64	70.53
	1	73.05	70.35
	2	72.76	69.98
	总	72.82 $\pm$ 0.21	70.29 $\pm$ 0.28
Base+AFCM+FBGL	0	76.63	73.64
	1	76.94	73.42
	2	76.50	72.88
	总	76.69 $\pm$ 0.23	73.31 $\pm$ 0.39
完整方法	0	77.92	73.90
	1	78.39	74.36
	2	78.26	74.22
	总	78.19 $\pm$ 0.24	74.16 $\pm$ 0.24

用相同的 10 组可见光检索库进行评估, 计算 10 次测试的平均结果, 并统计三个训练种子之间的均值与样本标准差, 结果如表 14。完整方法的两项指标样本标准差均为 0.24%, 平均性能高于所比较的消融配置。上述结果表明在本次三个随机种子的实验范围内, 完整方法表现出较小的训练波动, 模块组合带来的性能改善在重复训练中得到进一步验证。

2.7 可视化分析

为分析不同频带的特性权重分配及其与跨模态一致性、身份可分性的联系, 本文在 SYSU-MM01 测试集上对 AFCM 预测的模态特定、模态共享和身份判别权重进行可视化与关联分析, 结果如图 5 所示。图 5 (a) 分别展示随机选取的 100 张 RGB 图像和 100 张 IR 图像的平均权重, B1~B5 表示由低频到高频的 5 个基础频带; 图 5 (b) 展示随机选取的 10 个测试身份的三类特性权重, 横轴为频带序号, 纵轴为测试身份。单元格颜色表示对应权重的大小, 由浅黄绿色向深蓝色变化表示权重逐渐增大。三类权重的取值范围为 [0, 1], 同一身份、同一频带上的三类权重之和为 1。由图 5 (a) 可见, 不同频带对三类映射具有不同的权重分配。在 RGB 图像中, B1 的模态共享权重为 0.427, B3 的

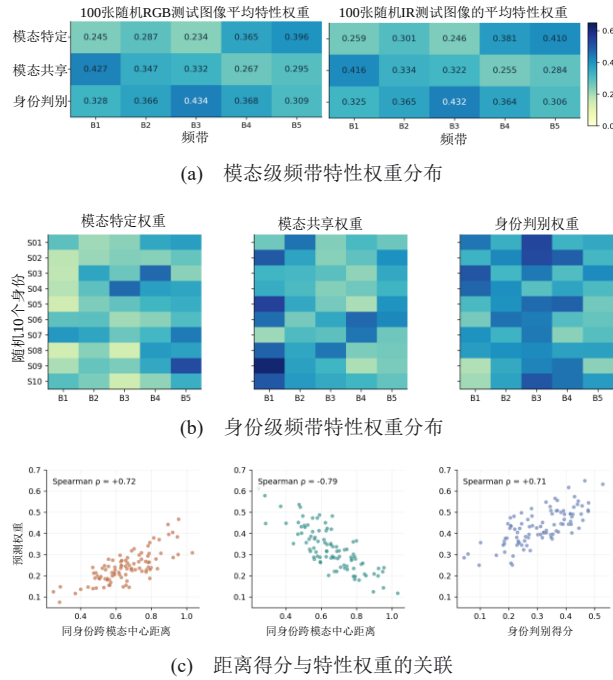
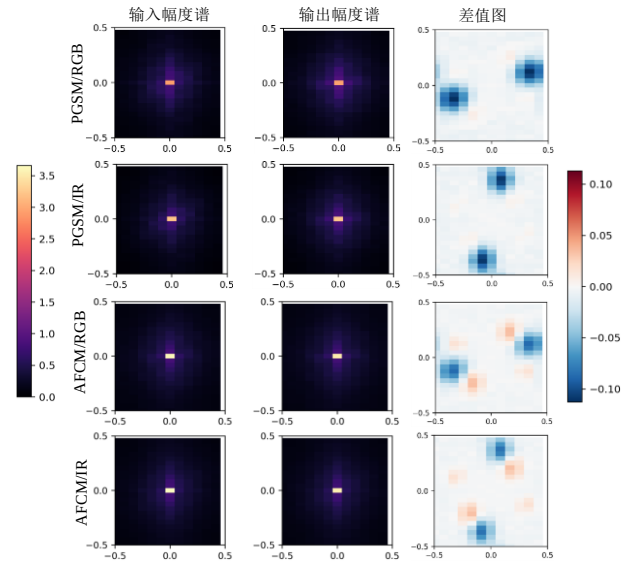


图5 SYSU-MM01上频带特性权重分布与距离得分关联可视化图

身份判别权重为0.434，B5的模态特定权重为0.396，均为对应频带中的最高权重；IR图像呈现相近趋势，分别为0.416、0.432和0.410。表明在所展示的样本中，低频B1更偏重模态共享映射，中间频带B3更偏重身份判别映射，高频B5更偏重模态特定映射。两种模态的主要分配趋势相近，具体权重有所差异，体现了频带调制在可见光与红外输入下的共同趋势及响应差异。图5(b)进一步显示，同一频带的权重随身份发生变化，例如S04与S09的模态特定权重分别在B3和B5上较高，说明AFCM能够结合输入表征调整各频带中三类映射的相对权重。

从频带表征角度看，式(25)中的频带中心由该频带的幅度与相位联合描述经归一化、身份内聚合得到，因此，同身份跨模态中心距离能够从中心表征层面刻画该频带对模态变化的响应差异。式(26)以最近异身份中心距离与同身份跨模态中心距离之差构造身份判别得分，该得分越大，表示该频带表征的类间分离相对于跨模态偏移具有更大的距离间隔。据此，将各身份在5个频带上的距离指标与对应预测权重汇总至身份层面，以全部96个测试身份计算Spearman相关系数，结果如图5(c)所示，每个散点对应一个测试身份。模态特定权重与同身份跨模态中心距离呈正相关，相关系数为

0.72；模态共享权重与该距离呈负相关，相关系数为-0.79。这表明，跨模态中心差异较大的身份倾向于获得较高的模态特定权重，跨模态中心更加接近的身份倾向于获得较高的模态共享权重。身份判别权重与身份判别得分呈正相关，相关系数为0.71，说明异身份中心间隔相对于同身份跨模态距离越大，身份判别映射的权重通常越高。上述结果支持权重分配与所设计代理指标之间的一致性，并为三类映射权重的含义提供了实验依据。



为分析频域模块对频谱响应的调节作用，图6展示了SYSU-MM01测试样本经PGSM和AFCM处理前后的特征幅度谱及差值图。输入与输出幅度谱采用相同色标，响应主要集中于中心低频区域，局部变化相对于整体幅度范围较小，因此两者整体形态接近。差值图采用以零为中心的独立色标，红色和蓝色分别表示幅度响应增大和减小，接近白色表示变化较小。PGSM的变化主要表现为局部频率响应减弱，且在可见光与红外特征中具有不同分布；AFCM则在不同频率位置呈现响应增强与减弱。上述现象表明，在所展示样本中，两种模块保持了幅度谱的主要分布，并对局部频率响应进行了差异化调节。结合表3的消融结果，分别引入PGSM和AFCM后，R-1与mAP均有所提高，可进一步说明频谱调制的有效性。

为直观分析模型的跨模态特征分布，本文采用t-SNE^[45]对跨模态特征进行可视化，结果如图7所

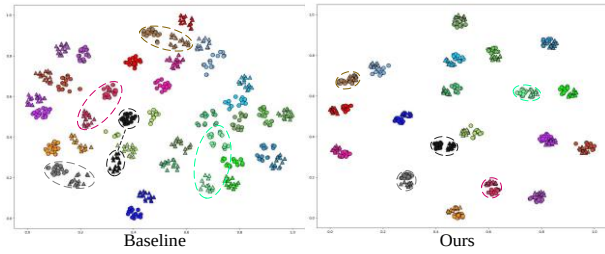


图7 t-SNE特征分布可视化图

示。相比基准模型，本文方法提取的同身份可见光与红外特征聚集更加紧密，不同身份间的分离程度更加明显，表明所提方法能够改善跨模态特征的类内紧凑性和类间判别性。本文统计了同身份（Intra）与不同身份（Inter）跨模态样本对的余弦距离分布，如图8所示。 δ_1 和 δ_2 分别表示基准模型和本文方法的类内、类间距离分布峰值间隔。由图可知 $\delta_2 > \delta_1$ ，表明所提方法能够有效增强同身份可见光与红外特征的一致性，同时保持不同身份之间的可分性。

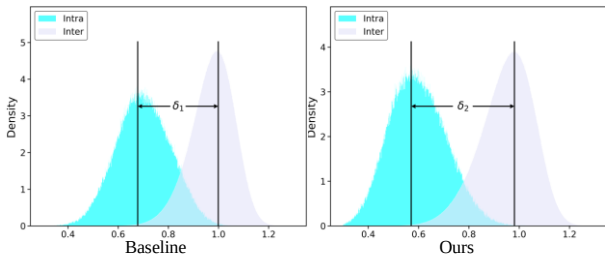


图8 类内与类间距离分布图

为进一步验证模型的跨模态检索能力，本文对SYSU-MM01数据集上的R-10检索结果进行可视化，如图9所示。其中绿色框和红色框分别表示正确匹配和错误匹配，相较于基准模型，本文方法能

够将更多同身份跨模态样本排列在检索序列前部，并明显减少外观相似行人造成的误匹配。尤其在视角变化、姿态差异及可见光—红外外观差异较大的情况下，所提方法仍能够保持较稳定的身份匹配结果，说明其能够有效挖掘身份判别线索，提高复杂条件下的检索准确性和排序质量。

为分析模型对身份相关区域的响应特性，本文对基准模型与所提方法的特征响应进行可视化^[46]。如图10所示，基准模型的响应区域相对分散，部分高响应区域出现在背景或局部非判别区域，对行人主体的关注不够稳定；相比之下，本文方法的响应更加集中于行人躯干、服饰纹理及人体轮廓等身份相关区域，并在可见光与红外模态下呈现较为一致的身份判别区域响应，表明所提方法能够减少背景及模态差异带来的干扰，增强对结构信息和局部身份细节的响应，从而获得更具判别性的特征表示。

3 结束语

本文针对VI-ReID中不同频带信息表征差异刻画不足的问题，提出一种频率感知判别特征学习的跨模态行人重识别方法。该方法利用相位结构辅助幅度响应调节，并自适应估计不同频带在模态特定、模态共享和身份判别映射上的相对权重，使各频带能够依据其响应特性参与跨模态特征学习；同时，通过频带级特性监督进一步约束频带特性建模，在减小模态差异的同时保持身份判别信息。在3个公开数据集上进行的实验证明了所提方法的有效性。后续将研究自适应频带划分与跨场景频域建模，以提升复杂光照和场景变化条件下的泛化能力。



图9 排序结果可视化图

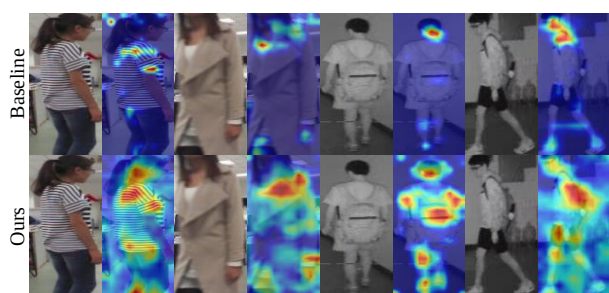


图10 特征响应热力图

程德强 (1979-), 男, 河南洛阳人, 博士, 中国矿业



大学教授、博士生导师, 主要研究方向为图像智能检测与模式识别、图像处理与视频编码。

牛凯凯 (2002-), 男, 安徽亳州人, 中国矿业大学硕



士, 主要研究方向为行人重识别。

郝大帅 (1998-), 男, 江苏徐州人, 中国矿业大学博



士研究生, 主要研究方向为行人重识别。

宋天舒 (1994-), 男, 博士, 中国矿业大学副教授,



主要研究方向为图像质量评价、人工智能+物理、表面生长动力学。

寇旗旗 (1988-), 男, 河南襄城人, 博士, 中国矿业大学副教授、硕士生导师, 主要研究方向为图像处理/AI 视频分析、智能检测与模式识别、图像增强与复原。



参考文献:

- [1] 程德强,姬广凯,张皓翔,等. 基于多粒度融合和跨尺度感知的跨模态行人重识别[J]. 通信学报, 2025, 46(1): 108-123.
CHENG D Q, JI G K, ZHANG H X, et al. Cross-modality person re-identification based on multi-granularity fusion and cross-scale perception[J]. Journal on Communications, 2025,46(1):108-123.
- [2] 励志勇,姜伟,刘浩杰. 可见光-红外跨模态行人重识别研究综述[J]. 电子学报, 2025, 53(12): 4811-4832.
LI Z Y, JIANG W, LIU H J. A Survey on Visible-infrared Cross-modality Person Re-identification[J]. Acta Electronica Sinica, 2025, 53(12): 4811-4832.
- [3] 姜迪,赖惠成,汪烈军. 基于跨模态融合与双曲图注意力机制的视频异常检测[J]. 通信学报, 2025, 46(6): 136-152.
JIANG D, LAI H C, WANG L J. Video anomaly detection via cross-modal fusion and hyperbolic graph attention mechanism[J]. Journal on Communications, 2025, 46(6): 136-152.
- [4] 邸敬,李涵,石淑慧,等. 动态时序扩散和多层级细化的红外与可见光图像融合网络[J]. 通信学报, 2025, 46(10): 233-246.
DI J, LI H, SHI S H, et al. Dynamic temporal diffusion and multi-level refinement network for infrared and visible image fusion[J]. Journal on Communications, 2025, 46(10): 233-246.
- [5] ALEHDAGHI M, JOSI A, CRUZ R M O, et al. Adaptive generation of privileged intermediate information for visible-infrared person re-identification[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 3400-3413.
- [6] 张红颖,樊世钰,罗谦,等. 结合视觉文本匹配和图嵌入的可见光-红外行人重识别[J]. 电子与信息学报, 2024, 46(9): 3662-3671.
ZHANG H Y, FAN S Y, LUO Q, et al. Visible-infrared person re-identification combining visual-textual matching and graph embedding [J]. Journal of Electronics & Information Technology, 2024, 46(9): 3662-3671.
- [7] LIU M, ZHANG Z, BIAN Y, et al. Cross-modality semantic consistency learning for visible-infrared person re-identification[J]. IEEE Transactions on Multimedia, 2025, 27: 568-580.
- [8] YUAN C H, JIANG B, CHEN X P, et al. Dual-auxiliary modalities and multistage semantic information fusion network for visible-infrared person re-identification[J]. Digital Signal Processing, 2025, 166: 105392.
- [9] QIAN Z H, LIN Y T, DU B. Visible - infrared person re-identification via patch-mixed cross-modality learning[J]. Pattern Recognition, 2025, 157: 110873.
- [10] SHI J M, YIN X B, ZHANG D M, et al. Two-stage knowledge distillation for visible-infrared person re-identification[J]. Pattern Recognition, 2026, 169: 111850.
- [11] XIONG J H, XIE X H, FENG Q Y, et al. CLIP-powered modality centering with spiral training for visible-infrared person re-identification [J]. Pattern Recognition, 2026, 177: 113333.

- [12] YU H, CHENG X, PENG W, et al. Modality Unifying Network for Visible-Infrared Person Re-Identification[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2023: 11151-11161.
- [13] REN K J, ZHANG L. Implicit discriminative knowledge learning for visible-infrared person re-identification[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2024: 393-402.
- [14] 许成杰,王蓉,毕艺瀚,等.特征自适应增强与动态校准的可见光-红外行人重识别方法[J/OL].北京航空航天大学学报,1-20[2026-09-21].
https://doi.org/10.13700/j.bh.1001-5965.2026.0166.
XU C J, WANG R, BI Y H, et al. Visible-infrared person re-identification via feature adaptive enhancement and dynamic calibration[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 1-20[2026-09-21]. https://doi.org/10.13700/j.bh.1001-5965.2026.0166.
- [15] CHEN X, JIN S, ZHAO L, et al. Mask-guided frequency feature fusion for visible - infrared remote sensing object detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1-15.
- [16] LI Y L, ZHANG T Z, ZHANG Y D. Frequency domain modality-invariant feature learning for visible-infrared person re-identification [EB/OL]. [2026-09-05]. https://arxiv.org/abs/2401.01839.
- [17] ZHANG Y K, WANG H Z, LU Y, et al. Frequency domain nuances mining for visible-infrared person re-identification[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 5411-5424.
- [18] GU H Y, YANG X G, LU R T, et al. Discovering multi-frequency embedding for visible-infrared person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2026, 36(2): 1766-1780.
- [19] YU T, CHENG D Q, JIANG H, et al. Wavelet-based frequency feature learning for visible-infrared person re-identification[J]. IEEE Transactions on Consumer Electronics, 2026, 72(1): 65-79.
- [20] 金静,朱传斌,翟凤文. 频域空间信息驱动的特征聚合跨模态行人重识别方法[J]. 计算机应用研究, 2026, 43(01): 298-304.
JIN J, ZHU C B, ZHAI F W. Feature aggregation cross-modality person re-identification method driven by frequency domain spatial information[J]. Application Research of Computers, 2026, 43(01): 298-304.
- [21] ZHONG J Q, QIAN H M, DING P. Frequency-aware dual-subspace representation learning for robust visible-infrared cross-modal matching[J]. Optics & Laser Technology, 2026, 204: 116182.
- [22] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 770-778.
- [23] LUO H, GU Y Z, LIAO X Y, et al. Bag of tricks and a strong baseline for deep person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2019: 1487-1495.
- [24] WU A C, 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. Los Alamitos: IEEE Computer Society Press, 2017: 5390-5399.
- [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] ZHANG Y K, WANG H Z. Diverse embedding expansion network and low-light cross-modality benchmark for visible-infrared person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2023: 2153-2162.
- [27] HUA X C, CHENG K, LU H, et al. MSCMNet: multi-scale semantic correlation mining for visible-infrared person re-identification[J]. Pattern Recognition, 2025, 159: 111090.
- [28] YE M, SHEN J B, LIN G J, et al. Deep learning for person re-identification: a survey and outlook[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 2872-2893.
- [29] YE M, RUAN W J, DU B, et al. Channel augmented joint learning for visible-infrared recognition[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 13567-13576.
- [30] YANG M X, HUANG Z Y, HU P, et al. Learning with twin noisy labels for visible-infrared person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 14308-14317.
- [31] SUN H Z, LIU J, ZHANG Z Z, et al. Not all pixels are matched: dense contrastive learning for cross-modality person re-identification[C]//Proceedings of the 30th ACM International Conference on Multimedia. New York: ACM Press, 2022: 5333-5341.
- [32] ZHANG Y Y, KANG Y H, ZHAO S Y, et al. Dual-semantic consistency learning for visible-infrared person re-identification[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 1554-1565.
- [33] LU H, ZOU X Z, ZHANG P P. Learning progressive modality-shared transformers for effective visible-infrared person re-identification[C]//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2023: 1835-1843.
- [34] HUANG N C, LIU J N, LUO Y J, et al. Exploring modality-shared appearance features and modality-invariant relation features for cross-modality person re-identification[J]. Pattern Recognition, 2023, 135: 109145.
- [35] LU Z F, LIN R H, HU H F. Modality and camera factors bi-disentanglement for NIR-VIS object re-identification[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 1989-2004.
- [36] NIE J H, LIN S, KOT A C. Color space learning for cross-color person re-identification[C]//Proceedings of the IEEE International Conference on Multimedia and Expo (ICME). Los Alamitos: IEEE Computer Society Press, 2024: 1-6.
- [37] LIU H J, LUO H, PENG X T, et al. Modality bias calibration network via information disentanglement for visible-infrared person reidentification[J]. IEEE Transactions on Computational Social Systems, 2024, 11(5): 6925-6938.
- [38] PAN H H, PEI W J, LI X, et al. Unified conditional image generation for visible-infrared person re-identification[J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 9026-9038.
- [39] GUO Y W, ZHANG W H, JIAO L C, et al. Visible-infrared person re-identification with region-based augmentation and cross modality attention[J]. Scientific Reports, 2025, 15(1): 18225.
- [40] ZHANG G W, ZHANG Y F, TAN Z C. Protohpe: Prototype-guided high-frequency patch enhancement for visible-infrared person re-identification[C]//Proceedings of the 31st ACM international confer-

- ence on multimedia. 2023: 944-954.
- [41] ZHANG Y K, YAN Y, LU Y, et al. Towards a unified middle modality learning for visible-infrared person re-identification[C] //Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM Press, 2021: 788-796.
- [42] WU Y Q, DING Y K, ZHANG L Y. Frequency-Enhanced Part Feature Mining and Cross-Modality Alignment for Visible-Infrared Person Re-Identification[C]//International Conference on Intelligent Computing. 2025: 500-511.
- [43] LIU H, TAN X, ZHOU X. Parameter sharing exploration and hetero-center triplet loss for visible-thermal person re-identification[J]. IEEE Transactions on Multimedia, 2021, 23: 4414-4425.
- [44] ZHENG X T, CHEN X M, LU X Q. Visible-infrared person re-identification via partially interactive collaboration[J]. IEEE Transactions on Image Processing, 2022, 31: 6951-6963.
- [45] VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(86): 2579-2605.
- [46] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2): 336-359.