基于多粒度融合和跨尺度感知的跨模态行人重识别

程德强[1]，姬广凯[1]，张皓翔[1]，江鹤[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，江苏 徐州 221116）

摘　要：提出一种基于多粒度融合和跨尺度感知的跨模态行人重识别网络，该网络能够有效提取行人图像特征并减少图像间的模态差异。首先，提出多尺度特征融合注意力机制并设计一种多粒度非局部融合框架，有效融合不同模态和不同尺度的图像特征；其次，提出一种跨尺度特征信息感知策略，该策略可有效降低因视角变化、行人背景变化等产生的无关噪声对行人判别的影响；最后，针对行人图像特征信息不足，设计并行空洞卷积残差模块，获取更为丰富的行人特征信息。将所提方法在2个标准公共数据集与当前先进的跨模态行人重识别方法比较。实验结果表明，所提方法在SYSU-MM01数据集的全搜索模式下的R-1和平均精度（mAP）分别达到75.9%和73.3%，在RegDB数据集的可见光到红外的搜索（VIS to IR）模式下的Rank-1和mAP分别达到93.7%和89.3%，优于所对比的方法，充分证实了所提方法的有效性。

关键词：行人重识别；跨模态；特征融合；跨尺度信息
中图分类号：TP391
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2025019

Cross-modality person re-identification based on multi-granularity fusion and cross-scale perception

CHENG Deqiang[1], JI Guangkai[1], ZHANG Haoxiang[1], JIANG He[1], KOU Qiqi[2]
1. School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China
2. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China

Abstract: A cross-modality person re-identification network based on multi-granularity fusion and cross-scale perception was proposed, which could effectively extract person image features and reduce the modality discrepancies between images. Firstly, a multi-scale feature fusion attention mechanism was proposed, and a multi-granularity non-local fusion framework was designed to effectively integrate image features from different modalities and scales. Secondly, a cross-scale feature information perception strategy was proposed, which could effectively reduce the influence of irrelevant noise caused by the change of perspective and person background on person discrimination. Finally, in view of the lack of person image feature information, a parallel dilated convolution residual module was designed to obtain more abundant person feature information. The proposed method was compared with current state-of-the-art cross-modal person re-identification algorithms on two standard public datasets. Experimental results show that the Rank-1 and mAP of the proposed method reach 75.9% and 73.3%, respectively, in the all search mode of the SYSU-MM01 dataset, and 93.7% and 89.3% in the VIS to IR retrieval mode of the RegDB dataset, respectively, which is better than the compared methods, which fully confirms the effectiveness of the proposed method.

Keywords: person re-identification, cross-modality, feature fusion, cross-scale information

收稿日期：2024–09–02；**修回日期：**2024–12–11
通信作者：寇旗旗，kouqiqi@cumt.edu.cn
基金项目：国家自然科学基金资助项目（No.52204117, No.52304182）；济宁市重点研发计划基金资助项目（No.2023KJHZ007）
Foundation Items：The National Natural Science Foundation of China (No.52204117, No.52304182), The Key Research and Development Program of Jining City (No.2023KJHZ007)

0　引言

行人重识别是当今计算机视觉领域的热点研究内容[1]之一，目的为在不同监控图像和视频中进行行人检索[2]。该技术被广泛应用于视频监控、智能安保、智慧寻人[3]等领域。传统的行人重识别方法致力于检索在白天[4]捕获的可见光图像，这些方法在黑暗环境下不能有效地检索到行人信息，因此跨模态行人重识别方法被提出，匹配可见光图像和红外光图像的同一行人。

粒度大小的具体含义为图像特征处于不同阶段的细化程度，粒度越细，对图片特征的关注程度越具体，如图 1 所示，图像划分的条带越多，表达其分区的粒度越细，模型能捕捉到的细节越多。行人重识别模型在早期训练阶段倾向于从全局图像比较，学习粗粒度和容易的身份信息，很难直接区分不同行人身份和细粒度信息，其整体难度为如何高效融合不同粗细粒度的行人特征信息以及如何区分在细粒度特征上不同的行人特征对象，本文由此引入多粒度思想设计网络架构，通过关注图像中的粗粒度和细粒度之间的相互关系，捕捉和检索图像中不同层次的特征和信息。

图 1　在行人重识别领域的粗细粒度表示

从近年来跨模态行人重识别的研究方法来看，在基于度量学习的方法中，Gao 等[5]设计非局部注意力块，Wang 等[6]设计局部特征之间的关系增强模型，Hu 等[7]设计解耦与模态不变表示（DMIR, decoupling and modality-invariant representation）损失函数，都倾向于直接将特征映射到公共特征空间，未能充分融合多粒度的行人特征信息，忽视多粒度特征的重要性及在不同粒度下特征融合的潜力，忽略分布在不同层上的粒度信息和不同模态间的具有判别性的自有特征。本文从充分融合不同粒度行人特征角度出发，设计全局信息和各粒度局部信息结合的端到端特征学习策略，并通过设计级联层级结构促进从粗到细的分层细粒度特征学习。

在基于特征对齐的方法中，Ye 等[8]提出分层跨模态匹配模型，Lu 等[9]分离各模态特有信息，Hao 等[10]设计网络最大化跨模态间的相似性，Wu 等[11]加入模态缓解模块和模态对齐模块增强鉴别特征，由于模态差异和色彩信息缺少、图像清晰度较低等不利因素，上述方法很难有效将不同模态的图像直接映射到同一个特征空间，仅聚焦于捕获全局粗粒度或局部细粒度特征，对行人特征的全面学习存在明显不足，本文从充分关注行人特征角度出发，设计分支网络关注各粒度的行人共享特征和重要特征，同时注重多级特征上下文信息间的传递和交互，借助粗粒度特征引导网络精准高效识别细粒度特征信息。

在基于模态互转的方法中，Wang 等[12-13]直接将红外光图像转换为对应可见光图像，Wang 等[14]将跨模态图像配对，Ye 等[15]将生成的可见光图像的灰度图作为辅助模态，Fan 等[16]提出基于生成对抗的模态互转方法，Wang 等[17]使用模态编码器和属性编码器以交换编码生成新的图像，Zhang 等[18]构建生成对抗网络和预训练师生模型，Choi 等[19]提出分层跨模态解纠缠方法，由于缺乏可见光-红外光图像对，此类方法生成的跨模态图像常伴随许多噪声，且极易破坏原始图像的结构信息，图像特征丢失现象严重，本文从充分利用特定粒度的行人特征角度出发，最大程度保留图像原有结构信息，并通过空洞卷积扩大感受野增强网络对行人特征信息的提取能力，针对有限行人图像生成丰富且多样化的特征嵌入。

综上所述，近年来相关工作中，受可见光和红外光图像间模态差异、各粒度特征等未被充分挖掘等因素影响，现有模型未能有效利用并融合不同粒度的行人特征信息，且本文分别针对其具体问题提出解决方案，进而提出一种基于多粒度融合和跨尺度感知的跨模态行人重识别网络。

本文主要贡献总结如下。

1) 提出一种基于多粒度融合和跨尺度感知的跨模态行人重识别网络，通过加入多粒度非局部融合框架和跨尺度特征信息感知策略，有效融合不同模态间的行人特征，提取行人图像信息。

2) 提出多粒度非局部融合框架，有效融合不同模态和尺度的行人特征，促进从粗到细的分层粒度特征学习，探索具有判别性的身份表示。

3) 设计跨尺度特征信息感知策略，关注行人跨尺度特征信息，有效减缓行人周边复杂背景等噪声的影响。

4) 设计并行空洞卷积残差模块，通过扩大感受野方式增强网络对行人特征信息的提取能力，生成有效的多样化的嵌入，挖掘潜在行人特征。

1　本文方法

1.1　整体网络结构

针对行人图像中不同模态差异较大、行人背景变化干扰过多且特征信息不足，本文提出基于多粒度融合和跨尺度感知的跨模态行人重识别网络，整体网络框架如图2所示。该网络以ResNet-50网络作为主干网络并采用双流网络结构，网络由可见光路径和红外路径组成，其中浅层卷积模块Layer0参数不共享，深层卷积模块Layer1~Layer4参数共享，网络由提取特征的主干网络、多粒度非局部融合框架和跨尺度特征信息感知策略共同组成，分别用来提取行人主体特征、融合不同模态和不同尺度的行人特征和关注行人跨尺度特征信息。

识别网络中，输入图片通过由ResNet-50网络构成的主干网络学习行人不同模态和不同尺度的细粒度特征，通过跨尺度特征信息感知策略获取行人跨尺度特征信息，并输入并行空洞卷积残差模块，以生成更多嵌入。

此外，针对不同模态和不同尺度的图像特征经过主干网络的前3层后，均输入多粒度非局部融合框架，其中在经过主干网络的Layer0后，该融合框架充分融合不同模态的行人特征，在经过主干网络的Layer1和Layer2后，该融合框架有效提取上下文信息。

1.2　多粒度特征融合框架

1.2.1　多尺度特征融合注意力机制

当前注意力机制被证明在行人重识别领域能够取得较好效果，Xu等[20]提出多样化局部注意力的网络。Woo等[21]在通道和空间维度依次操作，但未考虑2种注意力的结合，易丢失跨尺度特征信息。Jia等[22]在遮挡行人重识别过程中引入注意力，而针对跨模态行人重识别领域应用较少。

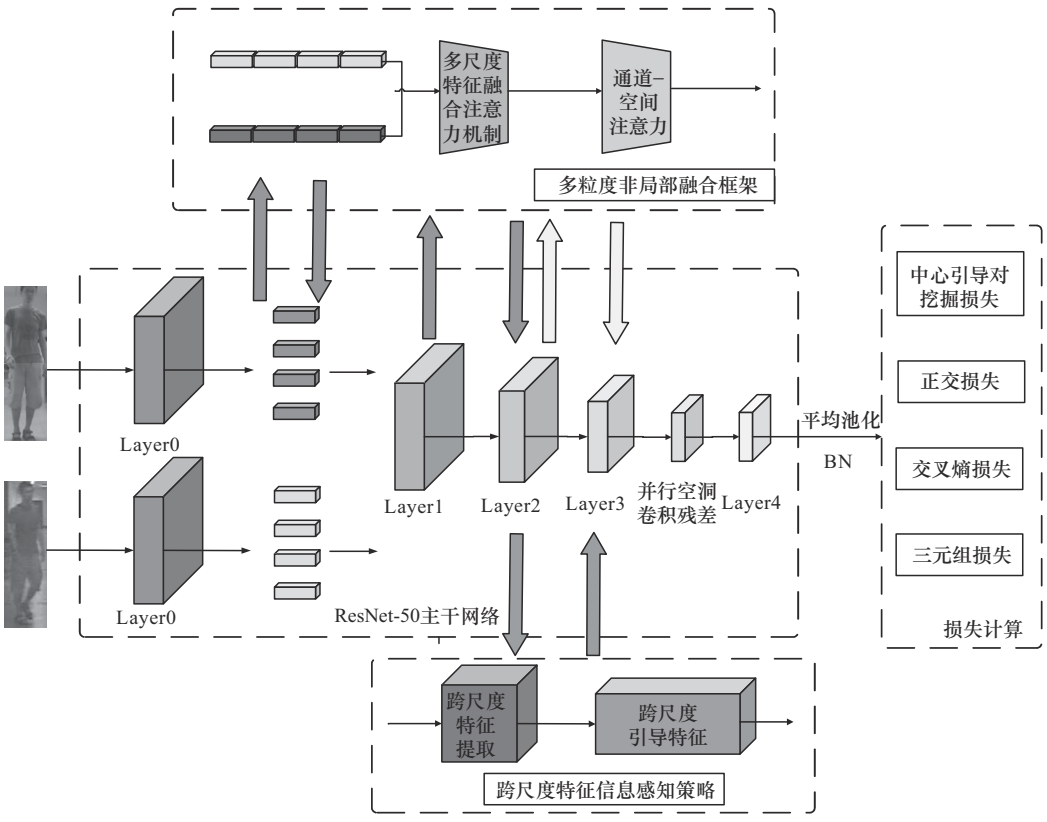


图2　整体网络框架

为获取图像中的关键特征信息，减少冗余信息干扰的同时保留多粒度的图像特征信息，本文提出多尺度特征融合注意力机制并加入多粒度非局部融合框架，通过可学习的参数和动态矩阵的混合，为每组子模块赋予不同权重，增强对判别特征的学习能力的同时抑制对重要特征的过度关注，并通过指定通道数缩减突出特征图中重要信息，同时利用池化和局部卷积操作，捕获长距离相互作用，凸显信息丰富区域，从而增强网络多尺度感知能力，融合多级网络特征。

多尺度特征融合注意力机制如图3所示，在前向传播过程中，输入特征$\boldsymbol{x}_i$为三维张量，$\varphi^1(\cdot)$代表对其进行1×1卷积操作，通过3个1×1卷积层后线性映射为$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$，其计算过程可描述为

$$\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}=\varphi_{\boldsymbol{Q}}^{1}(\boldsymbol{x}_i),\varphi_{\boldsymbol{K}}^{1}(\boldsymbol{x}_i),\varphi_{\boldsymbol{V}}^{1}(\boldsymbol{x}_i) \tag{1}$$

通过对$\boldsymbol{K}$区域输出层激活，得到概率级分布特征$\boldsymbol{K}_{\text{soft}}$，对$\boldsymbol{V}$区域求解均值，得到区域级分布特征$\boldsymbol{V}_{\text{avg}}$，通过$\boldsymbol{K}_{\text{soft}}$与$\boldsymbol{V}_{\text{avg}}$的矩阵乘法得到区域权重矩阵，其计算过程可描述为

$$\boldsymbol{K}_{\text{soft}}=\text{soft}(\boldsymbol{K}) \tag{2}$$

$$\boldsymbol{V}_{\text{avg}}=\text{avg}(\boldsymbol{V}) \tag{3}$$

$$\boldsymbol{M}_{\text{RW}}=\boldsymbol{K}_{\text{soft}}\otimes\boldsymbol{V}_{\text{avg}} \tag{4}$$

其中，soft为softmax激活函数，用于构造输出的概率分布，avg为平均池化（average pooling）函数，用于求解区域均值，⊗为矩阵乘法操作，矩阵$\boldsymbol{M}_{\text{RW}}$中的不同区域大小的值表示不同区域的重要程度。通过sigmoid激活函数进行输入元素映射获取$\boldsymbol{Q}$中的重要区域并馈送至7×7卷积层，对关键信息加权，提取更加丰富的特征信息，调整局部特征学习范围，消除无关信息干扰，其计算过程可描述为

$$\boldsymbol{Q}_{\text{sig}}=\varphi_{7\times7}(\text{sig}(\boldsymbol{Q})) \tag{5}$$

其中，sig为sigmoid激活函数操作，$\varphi_{7\times7}(\cdot)$表示7×7卷积操作

矩阵$\boldsymbol{M}_{\text{RW}}$与$\boldsymbol{Q}_{\text{sig}}$点乘得到表示特定区域的注意力特征图$\boldsymbol{M}_{\text{SAA}}$，其计算过程可描述为

$$\boldsymbol{M}_{\text{SAA}}=\boldsymbol{M}_{\text{RW}}\boldsymbol{Q}_{\text{sig}} \tag{6}$$

最后通过跳跃连接，将注意力特征图$\boldsymbol{M}_{\text{SAA}}$叠加到原始特征$\boldsymbol{x}_i$之上，特征叠加操作用⊕表示，得到最终输出$\boldsymbol{y}_{\text{out}}$，其计算过程可描述为

$$\boldsymbol{y}_{\text{out}}=\boldsymbol{x}_i\oplus\boldsymbol{M}_{\text{SAA}} \tag{7}$$

1.2.2　多粒度非局部融合框架

在跨模态行人重识别任务中，高效融合不同粗细粒度的行人特征信息，能够使网络模型关注到行人图像各个粒度的信息和全局信息，增强对图像特征信息的利用率和有效性，如图4所示，粗粒度帮助人们识别行人目标并区分数据集中其他无关干扰项，细粒度帮助人们通过鞋帽、姿势、表情等不同特征辨别从属于不同标签的具有细微差别的不同行人。

受多粒度[23]思想影响，本文提出多粒度非局部融合框架，以多尺度特征融合注意力机制为主要构成，并联合通道-空间注意力嵌入主干网络的不同层，增强空间和通道双重特征，该融合框架通过端到端特征学习策略，借助级联融合方式，不仅考虑网络深层次与浅层次信息，同时抽取每层中不同粒度的图像特征，主要实现如下。

1) 现有跨模态行人重识别方法在处理不同模态间行人特征时，往往忽视了特征的多粒度特性，本文特别考虑跨模态行人重识别中的跨模态性，针对不同模态特征进行融合，通过多尺度特征融合注意力机制调整特征提取的自适应权重，为不同模态的差异性特征赋予较大的权重，为其共有特征赋予较小的权重，从而将可见光图像和红外图像的行人

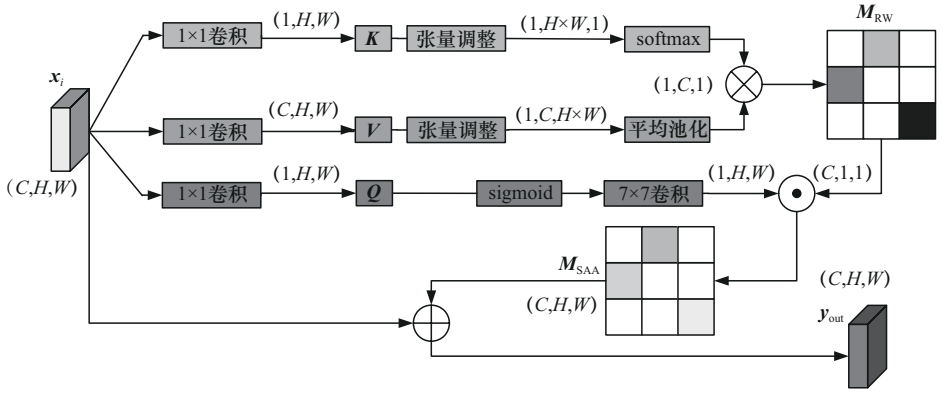
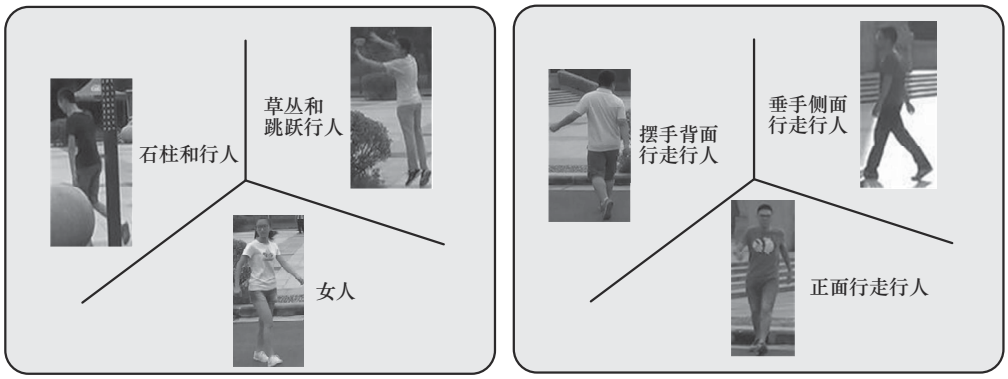


图3　多尺度特征融合注意力机制



(a) 粗粒度识别　　(b) 细粒度识别

图4　粗细粒度行人图像分析

特征充分融合，得到具有各模态判别性特征的跨模态融合特征并促使网络弥补模态间的特征差异，其具体实现过程如图5所示。2种模态特征均先依次经过1×1卷积、批归一化（BN）层、ReLU激活函数和最大池化操作，保留最重要特征并提高网络的表达能力，然后经过多尺度特征融合注意力机制和特征提取层后，通过通道注意力并与通过空间注意力的特征充分融合得到跨模态融合特征。

2) 针对不同尺度特征进行融合，利用多尺度特征融合注意力机制联合不同粒度的低维特征和高维特征，并通过权重的学习聚焦重要的行人身份信息，聚合来自不同尺度的特征以挖掘不同的通道和空间特征表示，从粗粒度层到细粒度层逐层进行特征提取，有效融合不同粒度的特征信息，引导浅层信息和深层信息充分融合，学习图像中不同粒度层之间的信息以及每个粒度层之间的信息，并生成有效的模态共享特征，减少卷积过程中有效特征信息的丢失，其具体实现如图6所示。该框架从不同粒

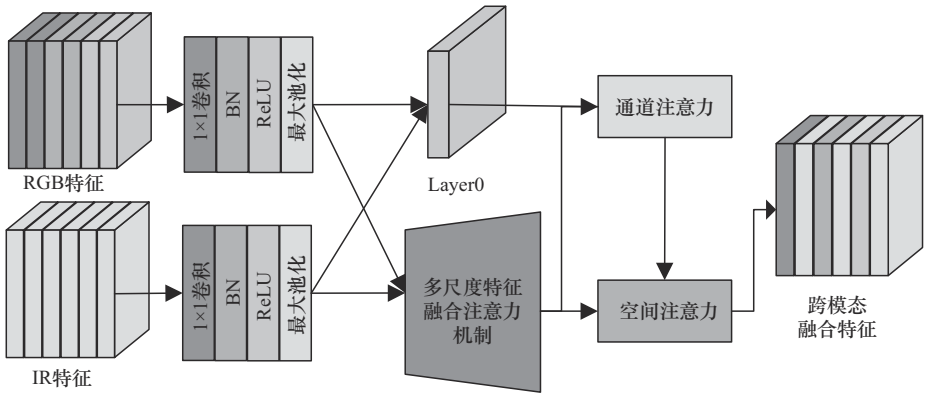


图5　针对不同模态的多粒度非局部融合框架

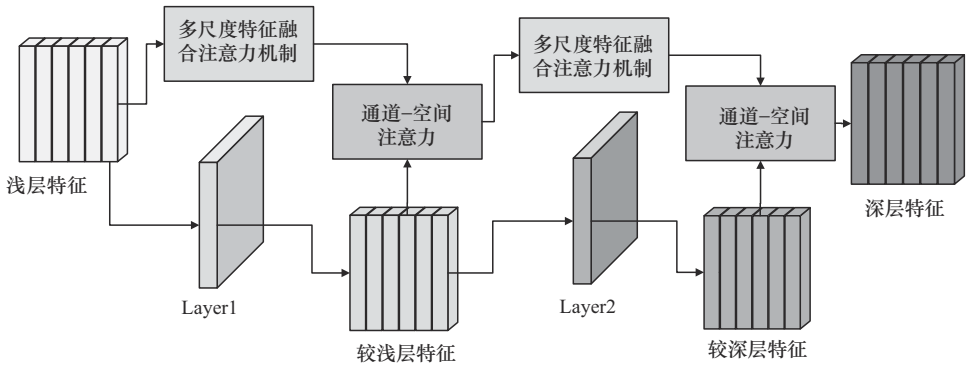


图6　针对不同尺度的多粒度非局部融合框架

度特征图提取信息，与特征金字塔结构相比具有更为简化和高效的模型架构。

1.3　跨尺度特征感知策略

1.3.1　跨尺度特征提取模块

为抑制行人背景干扰，深入挖掘行人与周边复杂背景的显著性区别，突出图像中关键区域，强化行人信息与周边背景信息的非关联程度，提出一个跨尺度特征提取模块，引导网络模型高效关注行人图像的跨尺度特征信息，如图7所示。

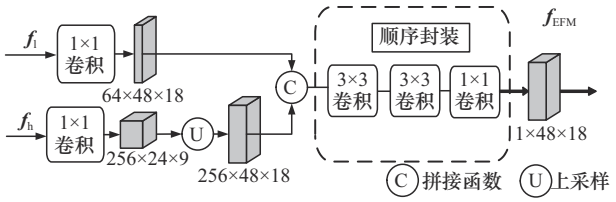


图7　跨尺度特征提取模块

在该模块中，从包含局部细节的低级特征$\boldsymbol{f}_{\mathrm{l}}$和包含边界位置信息的高级特征$\boldsymbol{f}_{\mathrm{h}}$中挖掘与行人对象相关的跨尺度语义信息，具体过程如下：首先，使用2个1×1卷积层分别将低级特征$\boldsymbol{f}_{\mathrm{l}}$和高级特征$\boldsymbol{f}_{\mathrm{h}}$的输出通道数更改为64和256，然后对高级特征$\boldsymbol{f}_{\mathrm{h}}$执行上采样操作，获取与低级特征$\boldsymbol{f}_{\mathrm{l}}$相同大小的特征，最后，低级特征$\boldsymbol{f}_{\mathrm{l}}$和高级特征$\boldsymbol{f}_{\mathrm{h}}$拼接后依次通过2个3×3卷积层和1个1×1卷积层，获得输出通道为1的跨尺度语义特征$\boldsymbol{f}_{\mathrm{out}}$。其计算过程可描述为

$$\boldsymbol{f}_{\mathrm{out}}=F_{\mathrm{Seq}}\{F_{\mathrm{cat}}(\varphi_{1\times1}^{1}(\boldsymbol{f}_{\mathrm{l}}),F_{\mathrm{U}}(\varphi_{1\times1}^{1}(\boldsymbol{f}_{\mathrm{h}})))\} \tag{8}$$

其中，$\varphi_{1\times1}^{1}(\cdot)$代表对其进行1×1卷积操作，$F_{\mathrm{U}}(\cdot)$代表执行上采样操作，$F_{\mathrm{cat}}(\cdot)$代表拼接操作，$F_{\mathrm{Seq}}(\cdot)$代表依次执行2次3×3卷积操作和1次1×1卷积操作。

1.3.2　跨尺度引导特征模块

为有效关注主体特征和共享特征，引导网络借助粗粒度特征关注有效的细粒度特征信息，提出一种跨尺度引导特征模块，注重多级特征上下文信息间的传递和交互，将含有局部细节的低级特征和位置语义信息的高级特征结合以探索出行人对象中相关的跨尺度特征，高效利用粗粒度特征深化细粒度特征学习，如图8所示。

在该模块中，通过跨尺度特征挖提取块后得到的跨尺度信息特征$\boldsymbol{f}_{\mathrm{EFM}}$与包含边界位置信息的高级特征$\boldsymbol{f}_{\mathrm{h}}$跨通道交互，同时关注其通道特征以探索通道间的关键线索。

对于得到中间特征$\boldsymbol{f}_i$，其具体过程如下：对跨尺度信息特征$\boldsymbol{f}_{\mathrm{EFM}}$执行下采样操作，获取与高级特征$\boldsymbol{f}_{\mathrm{h}}$相同大小的特征并进行元素乘法，与$\boldsymbol{f}_{\mathrm{h}}$求和后通过3×3卷积层得到中间特征$\boldsymbol{f}_i$。其计算过程可描述为

$$\boldsymbol{f}_i=\varphi_{3\times3}^{1}(\boldsymbol{f}_{\mathrm{h}}\oplus(\boldsymbol{f}_{\mathrm{h}}\otimes F_{\mathrm{D}}(\boldsymbol{f}_{\mathrm{EFM}}))) \tag{9}$$

其中，$\varphi_{3\times3}^{1}(\cdot)$表示3×3卷积操作，下同，$F_D(\cdot)$表示下采样操作，$\oplus$表示元素求和，$\otimes$表示元素乘法。

对于得到最终跨尺度语义特征$\boldsymbol{f}_{\mathrm{out}}$，其具体过程如下：对中间特征$\boldsymbol{f}_i$进行全局平均池化求解通道在空间维度上平均值，采用1×1卷积操作进行维度调整，进行sigmoid函数激活操作分配重要信息权重，与中间特征$\boldsymbol{f}_i$进行元素乘法得到最终输出$\boldsymbol{f}_{\mathrm{out}}$。其计算过程可描述为

$$\boldsymbol{f}_{\mathrm{out}}=F_{\mathrm{sig}}(\varphi_{1\times1}^{1}(F_{\mathrm{avg}}(\boldsymbol{f}_i)))\otimes\boldsymbol{f}_i \tag{10}$$

其中，F_{avg}表示全局平均池化，F_{sig}表示sigmoid函数激活。

1.3.3　跨尺度特征信息感知策略

跨尺度特征信息感知策略主要由跨尺度特征提取模块和跨尺度引导特征模块构成，该策略能够动态强化网络针对多模态的推理能力，通过潜在空间中重构不同粒度的图像特征来增强鲁棒的特征学习。

该结构针对行人图像信息做出更具针对性的特征提取，能够有效关注图像主体特征，通过在不同

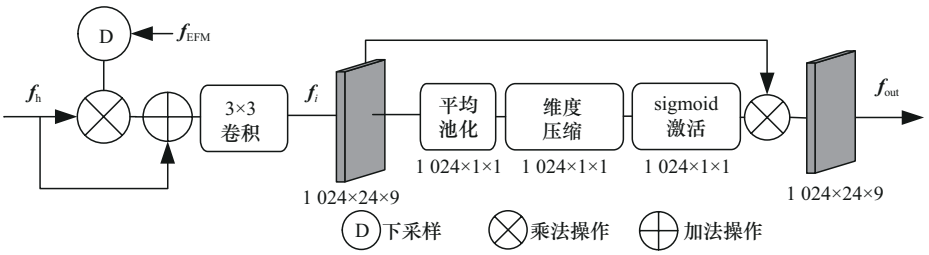


图8　跨尺度引导特征模块

尺度上对特征信息进行提取，关注共享特征信息的同时，引导全局特征等粗粒度特征信息关注有效的细粒度特征信息，有效降低行人背景变化等噪声的干扰。

如图9所示，其中包含局部细节的低级特征$\boldsymbol{f}_{\mathrm{l}}$和包含边界位置信息的高级特征$\boldsymbol{f}_{\mathrm{h}}$经由跨尺度特征提取模块挖掘出与行人对象相关的跨尺度语义信息，并与包含边界位置信息的高级特征$\boldsymbol{f}_{\mathrm{h}}$经由跨尺度引导特征模块得到包含行人对象相关跨尺度特征的最终输出$\boldsymbol{f}_{\mathrm{out}}$。

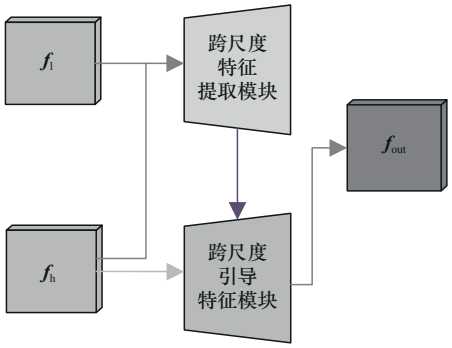


图9　跨尺度特征信息感知策略

1.4　并行空洞卷积残差模块

空洞卷积被广泛用于语义分割[24]、目标检测[25]、图像增强[26]等任务中，并在这些相关领域中取得显著性能提升，成为当前较为有效的操作之一。

针对红外光图像特征信息不足，提出一种并行空洞卷积残差模块，该模块针对有限行人图像，生成有效的多样化的嵌入，获取更为丰富的行人特征信息。在使用不同膨胀率的基础上，加入残差连接和3×3卷积，通过加入残差连接减缓网络层数加深对网络性能的影响，减少卷积层上的信息损失，有效避免梯度消失的问题，通过加入3×3卷积缓解使用空洞卷积层导致提取的特征信息不连续的问题，此外，每个卷积层后均加入激活函数层。

并行空洞卷积残差模块如图10所示，输入$\boldsymbol{f}_i$通过3个膨胀率分别为1、2、3的空洞卷积并进行输出层激活，从不同的感受野去捕捉不同尺度的语义信息，得到不同输出特征图$\boldsymbol{F}_1,\boldsymbol{F}_2,\boldsymbol{F}_3$，其计算过程可描述为

$$\boldsymbol{F}_1,\boldsymbol{F}_2,\boldsymbol{F}_3 = \mathrm{sig}(\varphi^1_{3\times3}(\boldsymbol{f}_i)),\mathrm{sig}(\varphi^2_{3\times3}(\boldsymbol{f}_i)),\mathrm{sig}(\varphi^3_{3\times3}(\boldsymbol{f}_i)) \tag{11}$$

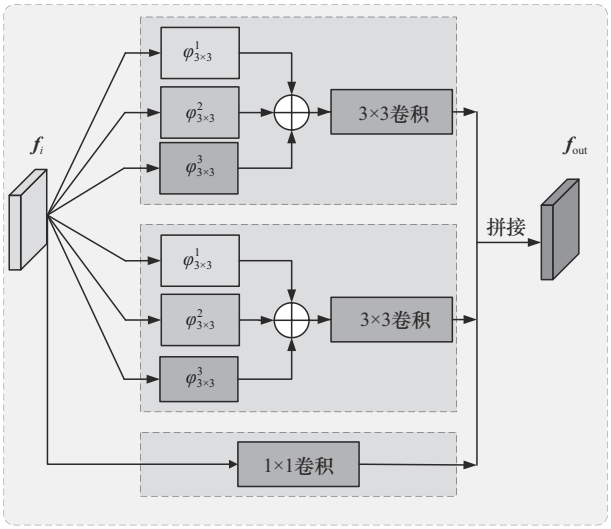


图10　并行空洞卷积残差模块

通过对$\boldsymbol{F}_1,\boldsymbol{F}_2,\boldsymbol{F}_3$求和并馈送到3×3卷积层得到与输入相同维度的嵌入特征$\boldsymbol{F}'$，其计算式为

$$\boldsymbol{F}' = \varphi^1_{3\times3}(\boldsymbol{F}_1 + \boldsymbol{F}_2 + \boldsymbol{F}_3) \tag{12}$$

对于输出$\boldsymbol{F}'$，与经过1×1卷积后的残差连接进行合并拼接得到最终输出$\boldsymbol{f}_{\mathrm{out}}$，其计算式为

$$\boldsymbol{f}_{\mathrm{out}} = F_{\mathrm{cat}}(\varphi^1_{1\times1}(\boldsymbol{f}_i),\boldsymbol{F}',\boldsymbol{F}') \tag{13}$$

1.5　损失函数

本文使用中心引导对挖掘损失函数、正交损失函数、交叉熵损失函数和三元组损失函数约束。

为有效保证并行空洞卷积残差模块生成的嵌入信息尽可能多样化，并减少可见光图像和红外光图像之间的模态差异，设计中心引导对挖掘损失，具体通过拉远生成嵌入和原始嵌入之间的距离；拉近可见光图像生成嵌入和红外光图像原始嵌入之间的距离；拉近红外图像生成嵌入和可见光图像原始嵌入之间的距离；保证类内距离小于类间距离，约束生成的嵌入，α设置为0.2，其计算过程可描述为

$$L(c_v,c_n,c^i_{v+}) = D(c^j_n,c^{ij}_{v+}) - D(c^j_v,c^{ij}_{v+}) - D(c^j_v,c^k_v) + \alpha \tag{14}$$

其中，c_v、c_n分别表示来自可见光模态和红外光模态的原始类嵌入中心，c^i_{v+}、c^i_{n+}分别表示来自可见光模态和红外光模态生成的类嵌入中心，i、j表示不同行人身份。同理，对于由红外光生成的类嵌入中心的控制，其计算过程可描述为

$$L(c_v,c_n,c^i_{n+}) = D(c^j_v,c^{ij}_{n+}) - D(c^j_n,c^{ij}_{n+}) - D(c^j_n,c^k_n) + \alpha \tag{15}$$

最终的中心引导对挖掘损失函数定义为

$$L_{\text{cpm}} = L(c_v, c_n, c_{v+}^i) + L(c_v, c_n, c_{n+}^i) \tag{16}$$

为保证不同分支生成的嵌入学习不同的信息特征表示，强制要求其在特征空间中保持正交性，设计正交损失函数，其计算过程可描述为

$$L_{\text{ort}} = \sum_{m=1}^{i-1} \sum_{n=m+1}^{i} ((f_+^m)^{\mathrm{T}} f_+^n) \tag{17}$$

其中，f_+^m和f_+^n分别是由原始嵌入生成的第m个和第n个嵌入，T表示对该嵌入进行转置操作。

为最小化模型输出的概率分布与真实概率分布之间的差异，提高身份信息的鉴别能力，防止模型过拟合，设计交叉熵损失函数，在训练过程中使用该函数用于行人检索过程中的身份识别，把跨模态行人重识别任务当作图像分类任务。其计算过程可描述为

$$L_{\text{id}} = -\frac{1}{n} \sum_{i=1}^{n} \log(p(y_i|x_i)) \tag{18}$$

其中，n代表训练的样本数，x_i代表行人图像的输入，y_i代表其对应的身份标签，$p(y_i|x_i)$表示输入图像和其对应的身份标签经过softmax函数分类后被识别为y_i的预测概率。

为优化统一身份和不同身份之间的三元组关系，设计三元组损失函数，其核心思想是保证同一个行人图像的特征距离小于不同行人间的特征距离，把跨模态行人重识别任务当作图像检索任务。其计算过程可描述为

$$L_{\text{tri}} = \sum_{i=1}^{P} \sum_{i=1}^{P} \left[\max_{p=1,\cdots,K} (F(x_i^a) - F\|x_i^p\|_2^2) - \min_{n=1,\cdots,K} (F(x_i^a) - F\|x_i^n\|_2^2) \right] + \varphi \tag{19}$$

其中，x_i^a、x_i^p和x_i^n分别代表锚样本、正样本和负样本，x_i^a和x_i^p属于同一行人的ID，x_i^a和x_i^n则分别使用不同行人的ID；$F(\cdot)$表示特征提取函数，$\|\cdot\|_2^2$表示求解正、负样本之间欧氏距离的过程，最困难的正样本是与锚样本属于同一类别且距离最远的样本，最困难的负样本是与锚样本属于不同类别且距离最近的样本，P表示单批次包含的行人数，K表示每个行人相应的图片数目，φ表示边缘参数，设置为0.3。

本文方法通过最小化4个损失函数之和共同优化网络，比例系数u和v分别确定为0.8和0.1。

$$L_{\text{total}} = L_{\text{id}} + L_{\text{tri}} + uL_{\text{cpm}} + vL_{\text{ort}} \tag{20}$$

2　实验验证与分析

2.1　实验参数

在训练阶段，输入图片的尺寸统一调整为384×144，训练迭代次数设置为150次，通过最小化总的损失函数的值，优化网络进而约束训练，随着模型训练轮次的增加，损失函数的值逼近于0，模型最终收敛。

由于RegDB数据集行人图片较少，本文使用ResNet-50结构中前3个阶段训练，在Layer3前加入并行空洞卷积残差模块，在Layer2和Layer3间应用跨尺度特征信息感知策略。

2.2　数据集和评价指标

SYSU-MM01数据集[27]共包含491个行人，可见光图像共计287 628张，红外光图像共计15 792张，并包含2种测试模式：全搜索模式（All-search）和室内搜索模式（Indoor-search）。RegDB数据集[28]通过一个可见光相机和一个红外相机共同收集所得。该数据集共包含412个行人，并包含2种不同的测试模式：可见光到红外的搜索（VIS to IR）和红外到可见光的搜索（IR to VIS）。

实验采用R-k和平均精度（mAP, mean average precision）评估模型性能，其中R-k表示每张检测的行人图像匹配时的前k个结果是否存在正确结果的概率，mAP用以评估检索的行人正样本在匹配结果中的分布情况，当行人正样本在匹配结果中越靠前，mAP越大。

2.3　与现有方法对比

为了验证本文方法的有效性，将所提方法和近年来的主流方法在SYSU-MM01、RegDB这2个公开数据集上进行对比，包括SMCL[29]、CAJ[30]、MPANet[11]、CM-EMD[31]、SGIEL[32]、DEEN[33]等基于特征对齐的方法，LbA[34]、NFS[35]、DART[36]等涉及图像噪声解决的方法和Hi-CMD[37]、CM-NAS[38]、FMCNet[39]等丰富图像特征信息的方法。

本文方法与现有方法在SYSU-MM01和RegDB数据集上的结果对比分别如表1和表2所示，其中，加粗和加下划线数据分别表示排名最优和次优结果。由表1可知，本文方法在SYSU-MM01数据集中All-search模式下，R-1、R-10、R-20和mAP分别达到了75.9%、97.7%、99.5%和73.3%，相较于

表1 本文方法与现有方法在SYSU-MM01数据集上的结果对比

方法	All-search				Indoor-search			
	R-1	R-10	R-20	mAP	R-1	R-10	R-20	mAP
Hi-CMD[37]	34.9%	77.6%	—	35.9%	—	—	—	—
JSIA-ReID[14]	38.1%	80.7%	89.9%	36.9%	43.8%	86.2%	94.2%	52.9%
X-Modality[40]	49.9%	89.8%	96.0%	50.7%	—	—	—	—
DDAG[41]	54.8%	90.4%	95.8%	53.0%	61.0%	94.1%	98.4%	68.0%
LbA[34]	55.4%	—	—	54.1%	58.5%	—	—	66.3%
NFS[35]	56.9%	91.3%	96.5%	55.5%	62.8%	96.5%	99.1%	69.8%
CM-NAS[38]	60.8%	92.1%	96.8%	58.9%	68.0%	94.8%	97.9%	52.4%
MCLNet[10]	65.8%	93.3%	97.1%	62.0%	72.6%	97.0%	99.2%	76.6%
FMCNet[39]	66.3%	—	—	62.5%	68.2%	—	—	74.1%
SMCL[29]	67.4%	92.9%	96.8%	61.8%	68.8%	96.6%	98.8%	75.6%
DART[36]	68.7%	96.4%	99.0%	66.3%	72.5%	97.8%	99.5%	78.2%
CAJ[30]	69.9%	95.7%	98.5%	66.9%	76.3%	97.9%	99.5%	80.4%
MPANet[11]	70.6%	96.2%	98.8%	68.2%	76.7%	98.2%	99.6%	81.0%
MMN[42]	70.6%	96.2%	99.0%	66.9%	76.2%	97.2%	99.3%	79.6%
DCLNet[43]	70.8%	—	—	65.3%	73.5%	—	—	76.8%
MAUM[44]	71.7%	—	—	68.8%	77.0%	—	—	81.9%
CM-EMD[31]	73.4%	—	—	68.6%	80.5%	—	—	82.7%
SGIEL[32]	75.2%	96.9%	97.3%	70.1%	78.4%	97.5%	98.9%	81.2%
DEEN[34]	74.7%	97.6%	99.2%	71.8%	80.3%	99.0%	99.8%	83.3%
本文方法	**75.9%**	**97.7%**	**99.5%**	**73.3%**	**83.4%**	98.9%	**99.8%**	**85.8%**

DEEN方法，R-1、mAP分别提升了1.2%、1.5%，Indoor-search模式下R-1、R-10、R-20和mAP分别达到了83.4%、98.9%、99.8%和85.8%，相较于DEEN方法，R-1、mAP分别提升了3.1%、2.5%，充分证实方法有效性，并能够针对近年来限制相关方法进一步提升的几个关键问题进行针对性解决。

由表2可知，本文方法在RegDB数据集中VIS to IR时，R-1、R-10、R-20和mAP分别达到了93.7%、98.1%、99.3%和89.3%，相较于DEEN方法，R-1、mAP相应提升了2.6%、4.2%，IR to VIS时R-1、R-10、R-20和mAP分别达到了93.6%、98.7%、99.5%和87.3%，相较于DEEN方法，R-1、mAP分别提升了4.1%、3.9%，在多个指标上均达到领先水平。

为充分评估本文方法在实际场景中受低光照影响后的模型性能，将其与现有方法在LLCM数据集[33]中进行对比，实验结果如表3所示，本文方法取得较为优异的结果，充分说明在实际应用场景也能取得出色性能。

2.4 消融实验

2.4.1 不同模块消融

为证明本文方法不同模型的有效性，本节分别在SYSU-MM01和RegDB这2个数据集上进行实验验证，结果如表4和表5所示。其中，CFIP、PDCRM和MGNF分别代表跨尺度特征信息感知策略、并行空洞卷积残差模块以及多粒度非局部融合框架。

由表4和表5分析可知，加入CFIP在R-1、R-10、R-20和mAP上的指标均有提高，说明跨尺度特征信息感知策略能够有效增强具有多重语义的行人跨尺度特征表示，挖掘有区分性的粒度特征，PDCRM在R-1和mAP指标上较之基准模型有了较大提升，反映出并行空洞卷积残差模块在生成多样化嵌

表 2　　**本文方法与现有方法在 RegDB 数据集上的结果对比**

方法	VIS to IR				IR to VIS			
	R-1	R-10	R-20	mAP	R-1	R-10	R-20	mAP
Hi-CMD[37]	70.9%	86.4%	—	66.0%	—	—	—	—
JSIA-ReID[14]	48.1%	—	—	48.9%	48.5%	—	—	49.3%
X-Modality[40]	62.2%	83.1%	91.7%	60.2%	—	—	—	—
DDAG[41]	69.3%	86.2%	91.5%	63.5%	68.1%	85.2%	90.3%	61.8%
LbA[34]	74.2%	—	—	67.6%	67.5%	—	—	72.4%
NFS[35]	80.5%	91.6%	95.1%	72.1%	78.0%	90.5%	93.6%	69.8%
CM-NAS[38]	82.8%	95.1%	97.7%	79.3%	81.7%	94.1%	96.9%	77.6%
MCLNet[10]	80.3%	92.7%	96.0%	73.1%	75.9%	90.9%	94.6%	69.5%
FMCNet[39]	89.1%	—	—	84.4%	88.4%	—	—	83.9%
SMCL[29]	83.9%	—	—	79.8%	83.1%	—	—	78.6%
DART[36]	83.6%	—	—	75.7%	82.0%	—	—	73.8%
CAJ[30]	85.0%	95.5%	97.5%	79.1%	84.8%	95.3%	97.5%	77.8%
MPANet[11]	82.8%	—	—	80.7%	83.7%	—	—	80.9%
MMN[42]	91.6%	97.7%	98.9%	84.1%	87.5%	96.0%	98.1%	80.5%
DCLNet[43]	81.2%	—	—	74.3%	78.0%	—	—	70.6%
MAUM[44]	87.9%	—	—	85.1%	87.0%	—	—	84.3%
CM-EMD[31]	94.4%	—	—	88.2%	92.8%	—	—	86.9%
SGIEL[32]	92.2%	—	—	86.6%	91.1%	—	—	85.2%
DEEN[33]	91.1%	97.8%	98.9%	85.1%	89.5%	96.8%	98.4%	83.4%
本文方法	93.7%	**98.1%**	**99.3%**	**89.3%**	**93.6%**	**98.7%**	**99.5%**	**87.3%**

表 3　　**本文方法与现有方法在 LLCM 数据集上的结果对比**

方法	IR to VIS				VIS to IR			
	R-1	R-10	R-20	mAP	R-1	R-10	R-20	mAP
DDAG[41]	42.4%	72.7%	80.6%	49.0%	51.4%	81.5%	88.3%	38.8%
LbA[34]	42.8%	77.4%	86.1%	51.0%	54.8%	85.1%	91.6%	40.8%
AGW[8]	46.4%	77.8%	85.2%	54.8%	63.7%	88.7%	92.8%	47.2%
CAJ[30]	49.9%	78.9%	85.8%	56.4%	63.7%	88.0%	92.4%	47.7%
MMN[42]	50.1%	79.8%	87.3%	56.7%	64.0%	88.7%	93.1%	48.5%
MRCN[45]	51.3%	80.1%	87.2%	57.7%	65.3%	88.1%	93.1%	49.5%
DART[36]	53.0%	80.8%	87.1%	59.3%	65.3%	89.4%	93.3%	51.1%
DEEN[33]	55.5%	83.9%	90.0%	62.1%	69.2%	91.0%	95.1%	55.5%
本文方法	**56.7%**	**84.6%**	**90.6%**	**63.4%**	**70.5%**	**91.7%**	**95.4%**	**56.7%**

入方面起到了较大作用，MGNF 在各个指标的测试中均有相应提高，说明粒度框架结构实现了先进性能，多粒度非局部融合框架能有效关注信息丰富区域，融合多级网络特征。综上所述，本文方法在关注行人跨模态和跨尺度特征信息、挖掘行人潜在特征及提取上下文信息方面取得了一定的改进效果。

表4 不同模块性能在SYSU-MM01数据集上验证结果

模块			All-search				Indoor-search			
CFIP	PDCRM	MGNF	R-1	R-10	R20	mAP	R-1	R-10	R-20	mAP
×	×	×	74.70%	97.60%	99.20%	71.80%	80.30%	99.00%	99.80%	83.30%
√	×	×	74.74%	97.68%	99.46%	72.37%	83.30%	97.71%	99.91%	86.03%
×	√	×	75.11%	97.24%	99.24%	72.31%	82.75%	98.75%	99.55%	85.33%
×	×	√	75.44%	97.51%	99.37%	73.14%	82.73%	97.93%	99.70%	85.10%
√	√	×	75.21%	97.70%	99.50%	72.45%	83.35%	97.85%	99.61%	86.04%
√	×	√	75.68%	97.56%	99.54%	72.42%	83.39%	98.98%	99.62%	85.45%
×	√	√	75.25%	97.85%	99.38%	73.27%	82.76%	98.85%	99.65%	85.60%
√	√	√	75.87%	97.74%	99.53%	73.34%	83.41%	98.92%	99.83%	85.84%

表5 不同模块性能在RegDB数据集上验证结果

模块			VIS to IR				IR to VIS			
CFIP	PDCRM	MGNF	R-1	R-10	R20	mAP	R-1	R-10	R-20	mAP
×	×	×	91.10%	97.80%	98.90%	85.10%	89.50%	96.80%	98.40%	83.40%
√	×	×	91.41%	97.41%	98.83%	87.02%	88.84%	96.97%	98.51%	85.43%
×	√	×	91.79%	97.53%	99.00%	86.83%	91.25%	97.46%	99.08%	86.56%
×	×	√	91.58%	99.04%	99.53%	89.29%	92.28%	98.37%	99.13%	86.19%
√	√	×	92.01%	97.58%	99.01%	89.97%	92.62%	97.90%	99.01%	87.08%
√	×	√	91.68%	99.14%	99.63%	89.30%	93.20%	98.47%	99.23%	87.20%
×	√	√	91.82%	97.78%	98.85%	89.20%	92.66%	97.17%	99.44%	86.79%
√	√	√	93.71%	98.14%	99.34%	89.33%	93.63%	98.73%	98.74%	87.34%

2.4.2 并行空洞卷积残差模块嵌入位置的研究

为验证并行空洞卷积残差模块嵌入ResNet-50主干网络中的哪一阶段取得最佳效果，设置如表6所示的实验。由表6可知，在Layer3之后嵌入并行空洞卷积残差模块后，模型取得最佳效果。

2.4.3 跨尺度特征提取模块应用位置的研究

为验证跨尺度特征提取模块应用至ResNet-50主干网络中的哪两层间取得最佳效果，设置如表7所示的实验。由表7可知，在Layer2~Layer3加入该模块后取得最佳效果，由此得出跨尺度特征信息感知策略能够提升关键特征感知能力，得到更具判别性和全面的提取特征。

2.4.4 联合损失函数系数选择

为评估联合损失函数系数u、v对于模型训练过程的影响，设置如表8和表9所示的对比实验。实验结果证明，当u=0.8、v=0.1时，实验取得最佳效果，考虑该系数搭配在数据集上表现较好，最后将u=0.8、v=0.1确定为联合损失函数的系数。

表6 并行空洞卷积残差模块插入位置的研究

模块插入位置	All-search				Indoor-search			
	R-1	R-10	R20	mAP	R-1	R-10	R-20	mAP
在Layer1之后	61.71%	93.89%	98.06%	56.57%	67.38%	96.47%	99.05%	72.61%
在Layer2之后	66.64%	94.99%	98.45%	61.44%	72.91%	97.71%	99.18%	77.51%
在Layer3之后	70.56%	96.26%	99.01%	66.22%	75.02%	98.04%	99.33%	79.08%
在Layer4之后	65.54%	94.42%	98.12%	61.82%	72.43%	96.97%	99.22%	77.13%

表 7　跨尺度特征提取模块应用位置的研究

模块应用位置	All-search				Indoor-search			
	R-1	R-10	R20	mAP	R-1	R-10	R-20	mAP
Layer0	74.70%	97.60%	99.20%	71.80%	80.30%	99.00%	99.80%	83.30%
Layer1~Layer2	70.32%	96.36%	98.95%	67.73%	79.57%	98.91%	99.80%	82.94%
Layer2~Layer3	75.90%	97.65%	99.51%	73.29%	83.37%	99.77%	99.76%	85.79%
Layer1~Layer3	75.14%	96.58%	99.26%	72.17%	83.35%	99.09%	99.86%	85.78%

表 8　不同系数 u 在 SYSU-MM01 数据集上的结果对比

u	All-search				Indoor-search			
	R-1	R-10	R-20	mAP	R-1	R-10	R-20	mAP
0	71.03%	91.21%	98.66%	68.70%	78.22%	98.37%	99.66%	81.80%
0.4	74.06%	97.20%	99.26%	72.00%	81.06%	98.92%	99.72%	84.00%
0.6	74.11%	96.77%	99.10%	72.65%	81.29%	98.86%	99.74%	84.36%
0.8	75.90%	97.65%	99.51%	73.29%	83.37%	99.77%	99.76%	85.79%
1.0	73.53%	96.88%	99.36%	70.99%	82.79%	98.19%	99.61%	84.73%

表 9　不同系数 v 在 SYSU-MM01 数据集上的结果对比

v	All-search				Indoor-search			
	R-1	R-10	R-20	mAP	R-1	R-10	R-20	mAP
0	74.80%	97.38%	99.34%	71.63%	81.27%	98.05%	98.80%	84.94%
0.05	74.90%	97.16%	99.21%	72.22%	82.49%	98.22%	98.95%	85.26%
0.10	75.90%	97.65%	99.51%	73.29%	83.37%	99.77%	99.76%	85.79%
0.15	75.12%	97.33%	99.37%	72.67%	83.93%	99.29%	99.65%	86.04%
0.20	74.49%	97.00%	99.23%	72.05%	82.60%	99.10%	99.93%	85.35%

2.4.5　时间复杂度对比

此外，为进一步证明本文方法以较低成本实现，进行浮点运算次数（FLOP, floating point operation）实验[46]并与其他方法进行对比，实验结果如表 10 所示。

表 10　本文方法与现有方法在 SYSU-MM01 数据集上的时间复杂度比较

方法	FLOP	R-1	mAP
CAJ[30]	31.09%	69.88%	66.89%
MMG[45]	41.38%	70.60%	66.90%
PIC[47]	41.38%	57.50%	55.10%
MMD[48]	20.69%	66.75%	62.25%
MCLNet[10]	20.72%	65.40%	61.98%
MCALNet[49]	20.69%	71.16%	68.01%
本文方法	**14.11%**	**75.90%**	**73.30%**

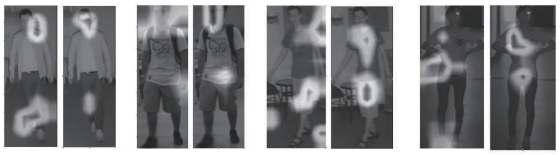
结果表明，本文方法实现了最低的 FLOP，拥有较低时间复杂度的同时，能够拥有最佳性能指标。

2.5　结果可视化

2.5.1　跨尺度特征信息感知策略有效性验证

为直观地展现出本文所提取出的跨尺度特征信息感知策略有效性，随机检索 4 个待查询行人，并通过梯度加权类激活映射（Grad-Cam, gradient-weighted class activation mapping）方法[50]绘制热力图。

图 11 为使用跨尺度特征信息感知策略有效性的对比，其中，图 11(a)~图 11(d)左侧为使用结构前，右侧为使用结构后。由实验结果可知，经过跨尺度特征信息感知策略后，待检索行人图像的人脸、胸口、腿部等身体部位标注为高亮，建筑、路面等周边背景未被明显关注，证明经跨尺度特征信息感知策略后，行人与周边背景的界限被明显区分，表现出对周边复杂背景较强的抗干扰能力。

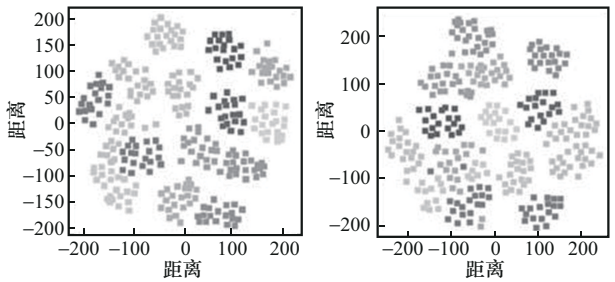


(a) 第一组行人 (b) 第二组行人 (c) 第三组行人 (d) 第四组行人

图11　跨尺度特征信息感知策略有效性的对比

2.5.2　模型特征分布直观验证

图12为本文方法与基准方法在特征分布方面的对比。通过对比可发现，本文方法能够根据不同颜色的特征形成了不同的簇，且相对基准方法更为清晰分明，说明本文方法能够有效聚合和区分同一行人的特征嵌入。



(a) 本文方法　(b) 基准方法

图12　本文方法与基准方法在特征分布方面的对比

2.5.3　本文方法行人重识别效果

为直观展现本文方法的行人重识别效果，随机检索4组待查询行人，并与基准方法的行人检索结果进行对比，如图13所示。

由图13可以看出，本文方法在4组检索结果中均优于原有基准方法。对于第一组行人，待检索行人在有正面轮廓且上衣有明显特征的前提下，在黑暗条件下检索时的基准方法，第1、4、6、9位序检索出错，而本文方法只有第7位序检索是不准确的，错误率更低；对于第二组行人，待检索行人仅有侧面轮廓且人脸特征部分被遮挡，基准方法在第4位序检索就出错，而本文方法检索出的10个位序全部正确。对于第三组行人，待检索的行人仅有背面轮廓且大半上身被书包所遮挡，基准方法在第7、8、10位序检索出错，而本文方法仅在第8、10位序检索出错，出错位序相对靠后。对于第四组行人，检索部分背面及侧面的行人图像，基准方法在第2、4、6、8位序检索出错。通过以上4组检索结果对比，在周边复杂背景干扰和行人部分特征被遮挡的情况下，本文方法相较于基准方法均表现出了更优的性能，充分证明本文方法的性能。

图13　本文方法与基准方法行人检索结果对比

3　结束语

本文提出了一种基于多粒度融合和跨尺度感知的跨模态行人重识别网络。多粒度非局部融合框架通过有效融合多粒度的网络特征，从而有效提取行人图像特征。跨尺度特征信息感知策略通过增强行人跨尺度特征表示，有效减少模态差异，并行空洞卷积残差模块在生成多样化嵌入方面起到了较大作用。在SYSU-MM01数据集、RegDB数据集中的实验结果表明，本文方法可有效解决跨模态行人重识别问题。后续将进一步探索跨模态在匹配时如何建模更丰富的特征从而缓解模态差异，并考虑如何进一步在无监督模型中减少模态差异。

参考文献：

[1] 罗浩, 姜伟, 范星, 等. 基于深度学习的行人重识别研究进展[J]. 自动化学报, 2019, 45(11): 2032-2049.
LUO H, JIANG W, FAN X, et al. A survey on deep learning based person re-identification[J]. Acta Automatica Sinica, 2019, 45(11): 2032-2049.

[2] ZOU C, CHEN Z Q, CUI Z C, et al. Discrepant and multi-instance proxies for unsupervised person re-identification[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 11024-11034.

[3] 程德强, 黄绩, 寇旗旗, 等. 基于相关性得分的伪标签优化行人重识别[J]. 控制与决策, 2024, 39(8): 2579-2587.
CHENG D Q, HUANG J, KOU Q Q, et al. Person re-identification with pseudo label refinement based on correlation score[J]. Control and Decision, 2024, 39(8): 2579-2587.

[4] 寇旗旗, 黄绩, 程德强, 等. 基于语义融合的域内相似性分组行人重识别[J]. 通信学报, 2022, 43(7): 153-162.
KOU Q Q, HUANG J, CHENG D Q, et al. Person re-identification with intra-domain similarity grouping based on semantic fusion[J]. Journal on Communications, 2022, 43(7): 153-162.

[5] GAO G W, SHAO H, WU F, et al. Leaning compact and representative features for cross-modality person re-identification[J]. World Wide Web, 2022, 25(4): 1649-1666.

[6] WANG C D, ZHANG C, FENG Y J, et al. Learning visible thermal person re-identification via spatial dependence and dual-constraint loss[J]. Entropy, 2022, 24(4): 443.

[7] HU W P, LIU B H, ZENG H T, et al. Adversarial decoupling and modality-invariant representation learning for visible-infrared person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(8): 5095-5109.

[8] YE M, LAN X Y, LI J W, et al. Hierarchical discriminative learning for visible thermal person re-identification[C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. Palo Alto: AAAI Press, 2018: 7501-7508.

[9] LU Y, WU Y, LIU B, et al. Cross-modality person re-identification with shared-specific feature transfer[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 4610-4617.

[10] HAO X, ZHAO S Y, YE M, et al. Cross-modality person re-identification via modality confusion and center aggregation[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 16403-16412.

[11] WU Q, DAI P Y, CHEN J, et al. Discover cross-modality nuances for visible-infrared person re-identification[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 4328-4337. .

[12] 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 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2019: 3622-3631.

[13] WANG Z X, WANG Z, ZHENG Y Q, et al. Learning to reduce dual-level discrepancy for infrared-visible person re-identification[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 618-626.

[14] WANG G A, ZHANG T Z, YANG Y, et al. Cross-modality paired-images generation for RGB-infrared person re-identification[J]. Neural Networks, 2020, 34(7): 12144-12151.

[15] 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, 2020, 16: 728-739.

[16] FAN X, JIANG W, LUO H, et al. Modality-transfer generative adversarial network and dual-level unified latent representation for visible thermal Person re-identification[J]. The Visual Computer, 2022, 38(1): 279-294.

[17] WANG Z J, LIU L, ZHANG H X. Dual-path image pair joint discrimination for visible – infrared person re-identification[J]. Journal of Visual Communication and Image Representation, 2022, 85: 103512.

[18] ZHANG Z Y, JIANG S, HUANG C, et al. RGB-IR cross-modality person ReID based on teacher-student GAN model[J]. Pattern Recognition Letters, 2021, 150: 155-161.

[19] CHOI S, LEE S M, KIM Y, et al. Hi-CMD: hierarchical cross-modality disentanglement for visible-infrared person re-identification[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 10254-10263.

[20] 徐胜军, 刘求缘, 史亚, 等. 基于多样化局部注意力网络的行人重识别[J]. 电子与信息学报, 2022, 44(1): 211-220.
XU S J, LIU Q Y, SHI Y, et al. Person re-identification based on diversified local attention network[J]. Journal of Electronics & Information Technology, 2022, 44(1): 211-220.

[21] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block atten-

tion module[C]//2018 Proceedings of the European conference on computer vision. Berlin: Springer, 2018: 3-19.

[22] JIA M X, SUN Y F, ZHAI Y P, et al. Semi-attention partition for occluded person re-identification[C]//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2023: 998-1006.

[23] YANG K W, YANG J W, TIAN X M. Learning multi-granularity features from multi-granularity regions for person re-identification[J]. Neurocomputing, 2021, 432: 206-215.

[24] 王新兴, 蓝凯, 李伟. 注意力空洞卷积无人机遥感多目标检测方法研究[J]. 测绘与空间地理信息, 2023, 46(12): 136-139.
WANG X X, LAN K, LI W. Research on multi-target detection method of UAV remote sensing based on attention hole convolution[J]. Geomatics and Spatial Information Technology, 2023, 46(12): 136-139.

[25] 陈清江, 顾媛. 基于多尺度深度可分离卷积的低照度图像增强算法[J]. 计算机工程与科学, 2023, 45(10): 1830-1837.
CHEN Q J, GU Y. A low-light image enhancement algorithm based on multi-scale depthwise separable convolution[J]. Computer Engineering and Science, 2023, 45(10): 1830-1837.

[26] 李庆, 王宏健, 李本银, 等. 基于改进的 IIE-SegNet 的快速图像语义分割方法[J]. 哈尔滨工程大学学报, 2024, 45(2): 314-323.
LI Q, WANG H J, LI B Y, et al. Fast image semantic segmentation method based on improved IIE-SegNet[J]. Journal of Harbin Engineering University, 2024, 45(2): 314-323.

[27] WU A C, ZHENG W S, YU H X, et al. RGB-infrared cross-modality person re-identification[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 5390-5399.

[28] NGUYEN D, HONG H, KIM K, et al. Person recognition system based on a combination of body images from visible light and thermal cameras[J]. Sensors, 2017, 17(3): 605-634.

[29] WEI Z Y, YANG X, WANG N N, et al. Syncretic modality collaborative learning for visible infrared person re-identification[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 225-234.

[30] YE M, RUAN W J, DU B, et al. Channel augmented joint learning for visible-infrared recognition[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 13547-13556.

[31] LING Y G, ZHONG Z, LUO Z M, et al. Cross-modality earth mover's distance for visible thermal person re-identification[C]//Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence. Palo Alto: AAAI Press, 2023: 1631-1639.

[32] FENG J W, WU A C, ZHENG W S. Shape-erased feature learning for visible-infrared person re-identification[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 22752-22761.

[33] 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 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 2153-2162.

[34] PARK H, LEE S, LEE J, et al. Learning by aligning: visible-infrared person re-identification using cross-modal correspondences[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 12026-12035.

[35] CHEN Y, WAN L, LI Z H, et al. Neural feature search for RGB-infrared person re-identification[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 587-597.

[36] 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 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 14288-14297.

[37] CHOI S, LEE S M, KIM Y, et al. Hi-CMD: hierarchical cross-modality disentanglement for visible-infrared person re-identification[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 10257-10266.

[38] FU C Y, HU Y B, WU X, et al. CM-NAS: cross-modality neural architecture search for visible-infrared person re-identification[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2021: 11803-11812.

[39] ZHANG Q, LAI C Z, LIU J N, et al. FMCNet: feature-level modality compensation for visible-infrared person re-identification[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 7339-7348.

[40] 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 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 4610-4617.

[41] YE M, SHEN J B, CRANDALL D J, et al. Dynamic dual-attentive aggregation learning for visible-infrared person re-identification[C]//Proceedings of the 16th European Conference. Berlin: Springer, 2020: 229-247.

[42] 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.

[43] 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.

[44] LIU J L, SUN Y F, ZHU F, et al. Learning memory-augmented unidirectional metrics for cross-modality person re-identification[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 19344-19353.

[45] ZHANG Y K, YAN Y, LI J, et al. MRCN: a novel modality restitution and compensation network for visible-infrared person re-identification[C]// Proceedings of the 37th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2023: 3498-3506.

[46] MOLCHANOV P, TYREE S, KARRAS T, et al. Pruning convolutional neural networks for resource efficient inference[J]. arXiv Preprint, arXiv: 1611.06440, 2016.

[47] 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.

[48] JAMBIGI C, RAWAL R, CHAKRABORTY A. Mmd-reid: a simple but effective solution for visible-thermal person rei[J]. arXiv Preprint, arXiv: 2111.05059, 2021.

[49] ZHANG H D, CHENG S L, DU A Y. Multi-stage auxiliary learning for visible-infrared person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(11): 12032-12047.

[50] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]// Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 618-626.

[作者简介]

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，主要研究方向为机器视觉与模式识别、图像智能检测与信息处理等。

姬广凯（1999– ），男，山东泰安人，中国矿业大学硕士生，主要研究方向为跨模态行人重识别。

张皓翔（1994– ），男，江苏徐州人，中国矿业大学博士生，主要研究方向为目标检测、图像检索等。

江鹤（1990– ），男，江苏徐州人，博士，中国矿业大学讲师，主要研究方向为图像增强与修复、图像检测、图像识别等。

寇旗旗（1988– ），男，河南襄城人，博士，中国矿业大学副教授，主要研究方向为图像增强与复原、智能检测与模式识别。