

文章编号 1004-924X(2026)10-1624-14

# 光照权重调配机制引导的双流逐层交互式 红外与可见光图像融合网络

韩玉兰\*, 吴桐, 黄海茹

(哈尔滨理工大学 测控技术与通信工程学院, 黑龙江 哈尔滨 150080)

**摘要:** 图像融合可整合不同模态图像的一致与互补特征, 有效弥补单一模态图像细节不足的问题。针对现有方法特征提取与融合相互独立, 光照信息利用不足、复杂光照场景下融合效果不佳等问题, 本文提出光照权重调配机制引导的双流逐层交互式红外与可见光图像融合网络。该网络采用逐层交互策略交替完成特征提取与跨模态融合。通过光照先验分类网络将全局光照信息嵌入双流网络中, 缓解光照不均影响。在特征提取阶段融入阶梯式扩张策略与轻量级全局上下文信息块, 实现多尺度长距离语义关联。同时利用差分特征与并行双注意力机制强化模态互补信息筛选与增强, 并依据光照动态调整融合权重, 避免单一模态主导。实验结果表明, 该方法在多个数据集上指标优于主流方法, 在 MSRS 数据集上, MI 和  $Q^{A/B/F}$  比次优算法分别提高了 3.76% 和 7.86%。在弱光、浓雾等复杂光照场景下的目标轮廓与细节保留效果更优。由此表明本文方法有效平衡底层空间信息与高层语义信息, 显著提升复杂光照场景下图像融合的精度与视觉效果。

**关键词:** 图像融合; 光照权重; 特征交互; 多尺度; 差分信息

中图分类号: TP391 文献标识码: A

doi: 10.37188/OPE.20263410.1624 CSTR: 32169.14.OPE.20263410.1624

## Dual-stream hierarchical interactive fusion network guided by illumination weight allocation

HAN Yulan\*, WU Tong, HUANG Hairu

(College of Measurement and Control Technology and Communication Engineering,  
Harbin University of Science and Technology, Harbin 150080, China)

\* Corresponding author, E-mail: hanyulan@hrbust.edu.cn

**Abstract:** Image fusion integrates the consistent and complementary features of multi-modal images to effectively compensate for insufficient details in single-modal images. Aiming at the problems that existing methods separated feature extraction and fusion, underutilized illumination information, and achieved unsatisfactory fusion results in complex illumination scenes, this paper proposed a dual-stream hierarchical interactive infrared and visible image fusion network guided by illumination weight allocation. The network adopted a layer-wise interactive strategy to alternately perform feature extraction and cross-modal fusion. A global illumination prior classification network was used to embed illumination information into the dual-stream network, so as to alleviate the impact of uneven illumination. The step-wise dilation strategy and

收稿日期: 2026-02-03; 修订日期: 2026-04-08.

基金项目: 黑龙江省“优秀青年教师基础研究支持计划”重点项目 (No. YQJH2024064)

lightweight global context block were integrated into feature extraction to achieve multi-scale and long-range semantic correlation. Differential features and parallel dual-attention mechanisms were adopted to enhance the screening and enhancement of modal complementary information, and fusion weights were dynamically adjusted according to illumination to avoid single-modal dominance. Experimental results show that the proposed method outperforms state-of-the-art methods on multiple datasets. On the MSRS datasets, MI and  $Q^{AB/F}$  are improved by 3.76% and 7.86%, respectively, compared with the suboptimal algorithm. It achieves better performance in preserving target contours and details in complex scenes such as low light and dense fog. The proposed method effectively balances low-level spatial information and high-level semantic information, significantly improving the accuracy and visual effect of image fusion in complex illumination scenarios.

**Key words:** image fusion; illumination weight; feature interaction; multi-scale; differential information

## 1 引言

多模态图像融合技术可整合模态的一致与互补特征,生成具有更全场景表征的融合图像,有效弥补单一模态图像细节不足<sup>[1]</sup>。红外图像抗干扰能力强,可精准凸显热目标,但分辨率低、对比度差。可见光图像分辨率高,能有效捕获色彩、纹理及边缘等结构信息,却易受复杂环境因素影响<sup>[2]</sup>。红外与可见光图像融合可实现信息互补,完善场景综合表达能力,为军事侦察、安防监控等领域提供有力支撑<sup>[3]</sup>。

传统红外与可见光图像融合方法依赖数学变换与人工规则,特征挖掘不充分,难以适配复杂实际场景。深度学习方法具备强大的特征建模能力,成为研究主流。深度学习融合方法包含生成对抗网络、自编码器网络及卷积神经网络三类架构。Ma等人<sup>[4]</sup>通过对抗博弈平衡双模态信息分布,但训练稳定性较差。Li等人<sup>[5]</sup>提出基于残差密集连接的DenseFuse网络。Li等人<sup>[6]</sup>利用跨模态交叉注意力增强特征互补性,但自编码器对称结构灵活性较差,仍依赖人工先验约束。后续研究逐渐转向卷积神经网络,Tang等人<sup>[7]</sup>构建SeAFusion网络,引入语义信息强化关键目标特征。Zhao等人<sup>[8]</sup>采用动态交互融合策略缓解模态语义差异问题。Xu等人<sup>[9]</sup>构建无监督模型U2Fusion,实现通用跨任务融合。针对统一融合建模的缺陷,Hu等人<sup>[10]</sup>提出TITA实现跨任务间共享互补信息的挖掘。Tang等人<sup>[11]</sup>通过C2RF的模态共性分解策略区分共享与私有特征,解决

配准与融合任务的内在矛盾。此外,随着Transformer模型的发展,CNN(Convolutional Neural Network)与Transformer结合的融合架构<sup>[12-13]</sup>弥补了卷积长距离建模的短板,但模型参数量与计算开销显著增加。

上述方法取得显著进展,但多数模型忽略光照变化影响,易造成可见光弱光细节丢失、红外目标过度平滑等问题。PIAfusion<sup>[14]</sup>引入光照感知损失约束网络训练。Tang等人<sup>[15]</sup>通过低光增强优化图像光照效果。Chen等人<sup>[16]</sup>提出循环反馈联合增强和融合框架,实现弱光增强与融合的双向引导。光照相关约束思路虽已在部分视觉任务中有所应用,但现有模型光照相关策略仅简单应用于损失函数或预处理环节,缺乏与融合全过程的深度协同,未设计面向融合任务定制化的动态权重调节机制,复杂光照场景泛化能力不足。同时现有模型的特征提取与融合相互独立,难以协同兼顾底层空间信息与高层语义特征,对模态一致与互补信息缺乏有效约束。针对以上问题,本文提出光照权重调配机制引导的双流逐层交互式红外与可见光图像融合网络(Dual-Stream Hierarchical Interactive Fusion Network Guided by Illumination Weight Allocation, DH-IFusion)。主要工作如下:

(1) 设计光照先验引导的逐层交互式融合结构。采用逐层交互策略交替完成特征提取与跨模态特征融合,兼顾底层空间与高层语义信息。通过光照先验分类网络(Illumination Prior Classification Network, IPC-Net)建立光照状态与双模

态特征的关联,改善光照不均衡导致的信息丢失;

(2) 提出多路径协同上下文多尺度聚合模块(Multi-path Collaborative Contextual Multi-scale Aggregation, MC<sup>2</sup>MA)。基于阶梯式扩张策略展开正方形卷积窗口,通过可变尺度的水平与垂直卷积构建多尺度感受野,结合轻量级全局上下文块(Global Context Block, GCB<sup>[17]</sup>),增强长距离语义关联,实现跨尺度、跨层次特征的有效表征;

(3) 提出跨模态互补信息自适应感知融合模块(Cross-modal Complementary Information Adaptive Perception and Fusion, C<sup>2</sup>IAPF)。利用模态差分建模互补特征,结合双注意力筛选特征,并依据光照动态调控融合权重,均衡双模态信息,适配复杂光照场景下的融合需求。

## 2 网络模型

DHIFusion网络整体框架如图1所示,主要由光照先验分类网络(IPC-Net)、逐层交互式特征提取融合网络及特征重构网络三部分组成。IPC-Net建模场景光照分布,生成光照权重以动态调整融合策略。逐层交互式特征提取融合网络依托多路径协同上下文多尺度聚合模块(MC<sup>2</sup>MA)实现双模态特征的多尺度深度提取,结合跨模态互补信息自适应感知融合模块(C<sup>2</sup>IAPF)筛选并增强模态间互补特征,实现跨

层级、跨模态信息的深度融合。最后特征重构网络将融合后的深度特征映射为最终融合图像。特征重构网络由4个3×3卷积和1个1×1卷积构成的5个卷积层,结构简洁高效。3×3卷积采用反射填充策略,对特征图边缘像素进行镜像扩展,避免边界伪影。

### 2.1 光照先验分类网络

场景光照变化会影响各模态的信息分布,光照良好时可见光图像承载主要信息,反之红外图像有效信息更丰富。为此,本文设计IPC-Net,用于估计光照信息,以光照分类置信度表征各模态信息丰富度,具体结构如图1所示。IPC-Net的特征提取采用二级结构,将2个3×3卷积层与坐标注意力块(Coordinate Attention Block, CoordAtt Block)<sup>[18]</sup>结合,通过跳跃连接实现跨层级信息融合,充分利用光照在不同尺度下的表现差异。CoordAtt Block通过正交空间轴分解完成特征双向编码,生成兼具全局关联性与像素级坐标精度的加权特征图,精准定位光照敏感区域。浅层CoordAtt Block捕捉阴影边界等局部光照细节,深层CoordAtt Block建模全局光照分布,全面覆盖像素级到语义级的光照表征。平均池化层整合区域光照信息,两个全连接层分别完成特征降维与分类映射,经Softmax归一化得到场景光照条件的分类置信度,并将其作为全局光照先验用于指导融合权重分配。IPC-Net具体过程见式(1)和式(2):

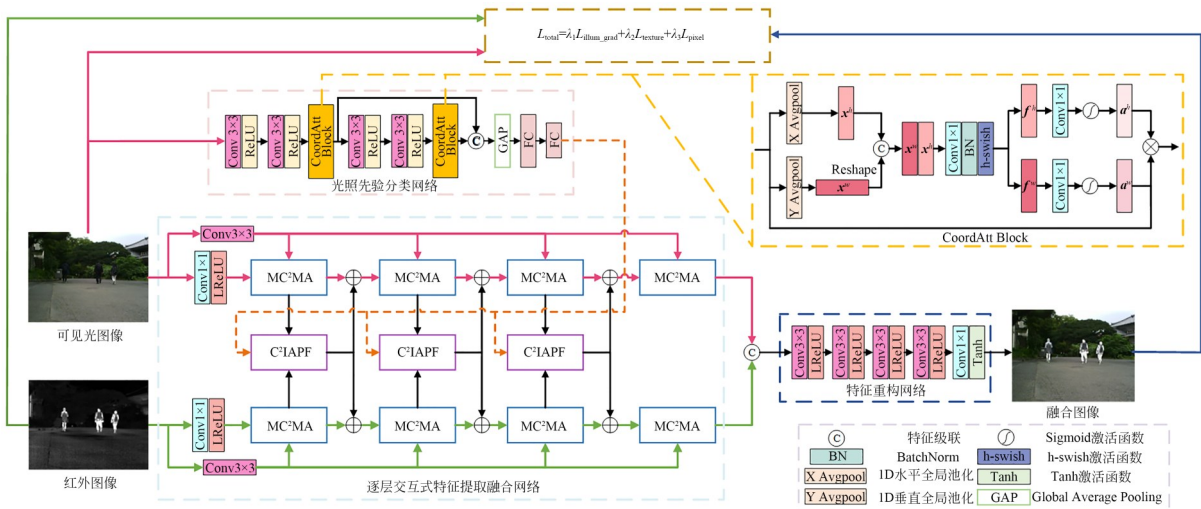


图1 DHIFusion总体架构

Fig. 1 Overall architecture of DHIFusion

$$(P_g, P_b) = \text{Softmax} \left( \text{FC} \left( \text{GAP} \left( f \left( \text{CC}(I_{vi}), \text{CC}(\text{CC}(I_{vi})) \right) \right) \right) \right), \quad (1)$$

$$W_{ir} = \frac{P_b}{P_g + P_b}, W_{vi} = \frac{P_g}{P_g + P_b}, \quad (2)$$

其中: $I_{vi}$ 表示可见光图像, $\text{CC}(\cdot)$ 表示卷积与Co-ordAtt Block组合操作, $f(\cdot)$ 表示特征级联操作, $\text{GAP}(\cdot)$ 和 $\text{FC}(\cdot)$ 分别表示全局平均池化和全连接, $\text{Softmax}(\cdot)$ 表示Softmax函数, $P_g, P_b$ 表示场景光照条件的分类置信度, $W_{ir}, W_{vi}$ 分别表示基于光照先验的红外与可见光模态融合贡献权重。

IPC-Net为预训练网络,本文在MSRS数据集中选取427张光照条件良好的白天图像和376张低照度夜间图像作为训练样本。采用one-hot编码标注全局光照标签,并裁剪为的局部图像块后用于训练。该标注策略可适配同场景局部光照变化,提升模型对复杂光照的适应能力,增强泛化性与鲁棒性。

## 2.2 双流逐层交互式特征提取融合网络

DHIFusion的特征提取融合结构如图1所示。各模态通过 $1 \times 1$ 卷积来初始化低层特征,

由MC<sup>2</sup>MA模块和C<sup>2</sup>IAPF模块进行深层特征处理,采用逐层交互式结构,借助光照先验信息指导模态间互补信息和一致性特征的融合传递,在凸显差异特征的同时累计互补信息,促进下一层特征提取。

### 2.2.1 MC<sup>2</sup>MA 模块

常规卷积虽能捕获图像局部相关性,但长距离依赖建模不足,且正方形卷积窗口限制特征捕获灵活性。对此,本文提出MC<sup>2</sup>MA,如图2所示,由两个串行级联的并行瓶颈金字塔上下文特征聚合模块(Parallel Bottleneck-Pyramid Context Aggregator, PBPCA)构成残差组,用于提取不同深度特征图,同时嵌入单模态浅层基础特征,兼顾浅层特征传递与跨层级信息交互。

PBPCA通过双路径构建跨尺度特征:第1条路径以 $3 \times 3$ 卷积结合Bottleneck结构挖掘深层语义信息,并拼接浅层细节提升特征复用率;第2

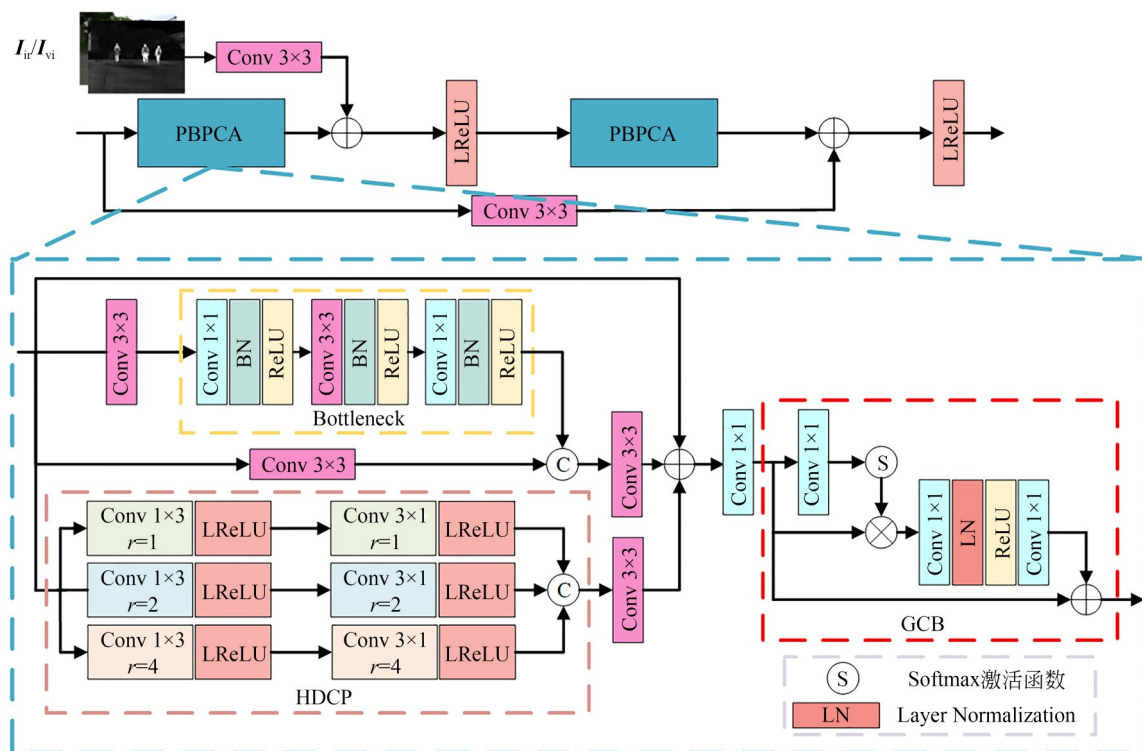


图2 MC<sup>2</sup>MA 模块

Fig. 2 MC<sup>2</sup>MA module

条路径通过混合膨胀卷积金字塔(Hybrid Dilated Convolution Pyramid, HDCP)展开正方形卷积窗口,基于阶梯式扩张策略,利用三组可变膨胀率的水平与垂直卷积核分别构建多尺度感受野,并引入残差连接。最后轻量级GCB<sup>[17]</sup>模块借助特征加权与通道注意力图的相关性捕获多尺度长距离上下文信息,增强特征语义关联性。MC<sup>2</sup>MA的具体过程见式(3):

$$F_{mc^2ma} = LRe\left(C_{3 \times 3}(F_a) + PP\left(LRe\left(C_{3 \times 3}(I_a) + PP(F_a)\right)\right)\right), \quad (3)$$

其中: $F_a$ 表示输入特征, $I_a$ 表示原始图像, $LRe(\cdot)$ 表示Leaky ReLU激活函数, $C_{3 \times 3}(\cdot)$ 表示 $3 \times 3$ 卷积操作, $PP(\cdot)$ 表示PBPCA模块功能, $F_{mc^2ma}$ 表示MC<sup>2</sup>MA模块的输出特征。

对于输入特征 $F_a$ ,PBPCA的具体过程为:

$$F_{step}^r = LRe\left(C_{3 \times 1}^r\left(LRe\left(C_{1 \times 3}^r(F_a)\right)\right)\right), \quad (4)$$

$$F_{hdcp} = f\left(F_{step}^1, F_{step}^2, F_{step}^4\right), \quad (5)$$

$$F_{bottle} = f\left(Bottleneck\left(C_{3 \times 3}(F_a)\right), C_{3 \times 3}(F_a)\right), \quad (6)$$

$$F_{pbpca} = GCB\left(C_{1 \times 1}\left(F_a + C_{3 \times 3}(F_{bottle}) + C_{3 \times 3}(F_{hdcp})\right)\right), \quad (7)$$

其中: $F_{step}^r$ 表示输入 $F_a$ 经过可变膨胀率 $r(r=$

1, 2, 4)提取到的特征, $C_{1 \times 3}^r(\cdot)$ , $C_{3 \times 1}^r(\cdot)$ 分别表示扩张率为 $r$ ,内核大小为 $1 \times 3$ 和 $3 \times 1$ 的卷积操作, $F_{bottle}$ 表示Bottleneck的输出, $F_{hdcp}$ 表示HDCP的输出, $F_{pbpca}$ 为PBPCA的输出, $Bottleneck(\cdot)$ 表示瓶颈结构运算, $C_{1 \times 1}(\cdot)$ 表示 $1 \times 1$ 卷积操作。 $GCB(\cdot)$ 表示GCB的功能。

## 2.2.2 C<sup>2</sup>IAPF 模块

传统融合中,加法操作易造成信息混叠,弱化互补特征,通道级联则易引入冗余且难以区分模态独有特性。对此,本文提出C<sup>2</sup>IAPF,如图3所示。通过减法计算模态间差分特征,过滤一致信息,聚焦互补信息,引入双向差分关系及光照自适应权重调整。C<sup>2</sup>IAPF含两个残差分支,首先提取差分浅层特征,再输入交叉互补双注意力模块(Cross Complementary Dual Attention Block, CCDAB):空间注意力(Spatial Attention Module, SAM)和通道注意力(Channel Attention Module, CAM)并行处理,抑制与显著性无关的冗余特征。经Softmax生成可学习权重向量,结合光照先验权重动态调整融合,最后将增强后的差异特征与源模态特征相加,实现互补信息强化融合。

以可见光特征为例,C<sup>2</sup>IAPF的具体过程为:

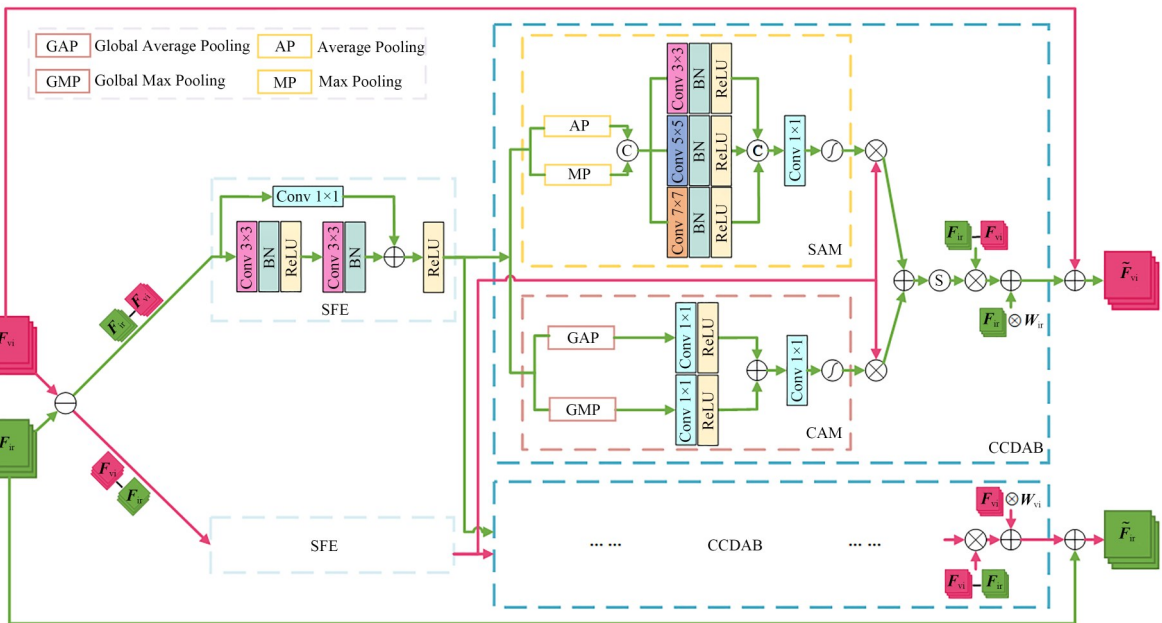


图3 C<sup>2</sup>IAPF 模块

Fig. 3 C<sup>2</sup>IAPF module

$$F_{\text{sab}} = \text{SAM}(SFE(F_{\text{ir}} - F_{\text{vi}})) \otimes SFE(F_{\text{vi}} - F_{\text{ir}}), \quad (8)$$

$$F_{\text{cab}} = \text{CAM}(SFE(F_{\text{ir}} - F_{\text{vi}})) \otimes SFE(F_{\text{vi}} - F_{\text{ir}}), \quad (9)$$

$$\tilde{F}_{\text{vi}} = F_{\text{vi}} + F_{\text{ir}} \otimes W_{\text{ir}} + \text{Softmax}(F_{\text{sam}} + F_{\text{cam}}) \otimes (F_{\text{ir}} - F_{\text{vi}}), \quad (10)$$

其中:  $F_{\text{ir}}$  和  $F_{\text{vi}}$  表示红外与可见光输入特征,  $SFE(\cdot)$  表示浅层特征提取,  $\text{SAM}(\cdot)$ ,  $\text{CAM}(\cdot)$  表示空间和通道注意力操作,  $\otimes$  表示逐元素相乘,  $F_{\text{sab}}, F_{\text{cab}}$  表示注意力处理后的特征,  $\tilde{F}_{\text{vi}}$  表示输出的可见光特征。

### 2.3 损失函数

#### 2.3.1 加权梯度一致性光照损失

结合 Retinex 理论的光照-反射分解思想, 可见光图像梯度映射包含光照分布与细节信息, 红外图像梯度对应热目标轮廓。本文利用光照先

$$L_{\text{texture}} = \frac{1}{HW} \left\| \nabla I_{\text{fused}} - \max(|\nabla I_{\text{ir}}|, |\nabla I_{\text{vi}}|) \times \left[ \frac{\nabla I_{\text{ir}} + \nabla I_{\text{vi}}}{|\nabla I_{\text{ir}} + \nabla I_{\text{vi}}| + \epsilon} \right] \right\|_1, \quad (12)$$

其中:  $|\cdot|$  表示绝对值操作, 用于计算梯度强度以表征纹理显著度,  $[\cdot]$  表示四舍五入取整操作, 用于离散化纹理方向信息,  $\max(\cdot)$  表示逐像素取最大值,  $\epsilon$  表示极小常数, 为了避免分母为 0。

#### 2.3.3 像素损失

通过构建融合双模态优势特征的伪参考图像  $I_{\text{pseudo}}$  作为监督基准, 利用 L1 损失约束融合图像与伪参考图像的像素差异, 保障融合结果在全局亮度与色彩分布均衡, 避免局部优化导致的视觉失衡。像素损失计算过程见式(13)和式(14):

$$I_{\text{pseudo}} = \text{RGB}\left(a \cdot \rho \max(f(I_{\text{vi}}^Y, I_{\text{ir}})) + (1-a) \cdot (W_{\text{vi}} \cdot I_{\text{vi}}^Y + W_{\text{ir}} \cdot I_{\text{ir}})\right), \quad (13)$$

$$L_{\text{pixel}} = \frac{1}{HW} \left\| I_{\text{pseudo}} - I_{\text{fused}} \right\|_1, \quad (14)$$

其中:  $\rho \max(\cdot)$  表示通道级逐像素取最大值操作,  $I_{\text{vi}}^Y$  表示可见光图像 Y 通道,  $a$  为平衡系数, 此处设置  $a=0.7$ ,  $\text{RGB}(\cdot)$  表示转换为 RGB 色彩空间的操作。

综合以上, 网络总损失  $L_{\text{total}}$  计算过程见式(15):

$$L_{\text{total}} = \lambda_1 L_{\text{illum\_grad}} + \lambda_2 L_{\text{texture}} + \lambda_3 L_{\text{pixel}}, \quad (15)$$

其中,  $\lambda_i (i=1, 2, 3)$  表示控制各项损失之间权重

信息构建加权梯度一致性光照损失函数, 计算过程见式(11)。通过动态光照权重调节模态梯度贡献度, 使融合图像梯度与源图像光照敏感区域相协调, 兼顾红外目标轮廓与可见光光照相关细节。

$$L_{\text{illum\_grad}} = W_{\text{vi}} \cdot \frac{1}{HW} \left\| \nabla I_{\text{fused}} - \nabla I_{\text{vi}} \right\|_1 + W_{\text{ir}} \cdot \frac{1}{HW} \left\| \nabla I_{\text{fused}} - \nabla I_{\text{ir}} \right\|_1, \quad (11)$$

其中:  $H, W$  表示输入图像的高度和宽度;  $\|\cdot\|_1$  表示 L1 范数;  $\nabla$  表示 Sobel 梯度算子;  $I_{\text{fused}}, I_{\text{vi}}, I_{\text{ir}}$  分别表示融合图像、可见光图像、红外图像。

#### 2.3.2 纹理损失

为了确保各模态源图像中的关键细节得到充分保留, 并使融合图像的纹理特征汇聚各模态源图像中最显著的纹理特性, 本文引入纹理损失函数, 计算过程见式(12):

平衡的参数。参考 DATFuse<sup>[12]</sup>, PIAFusion<sup>[13]</sup>, DIVFusion<sup>[15]</sup> 等经典融合模型及参数实验, 此处设置  $\lambda_1 = 15, \lambda_2 = 50, \lambda_3 = 7$ 。

## 3 实验与分析

### 3.1 实验数据集和评价指标

本文采用多光谱道路场景数据集 MSRS<sup>[14]</sup> 开展模型训练, 该数据集共含 1 444 对配准的红外和可见光图像, 统一分辨率为  $640 \times 480$ 。其中, 1 083 对图像经过分块裁剪得到 26 113 对  $64 \times 64$  的局部图像块作为模型训练样本, 361 对独立测试集用于评估测试。另外, 本文在公开数据集 TNO<sup>[19]</sup> 中选取 42 对图像进行测试分析, 该数据集包含高精度配准多波段图像, 涵盖夜间军事、安防类应用场景。为了进一步丰富复杂光照条件下的实验验证场景, 本文在自制数据集中采用 41 对图像进行测试分析, 该数据集由 FLIR E8 热成像仪采集, 包含 41 对配准的可见光与红外图像, 分辨率统一为  $1 132 \times 848$ , 采集场景涵盖道路、建筑入口、校园等典型城市公共区域, 覆盖不同时间段的复杂光照条件。

为验证算法有效性, 本文选取互信息 (Mutu-

al Information, MI)、像素级特征互信息 (Feature-based Mutual Information at Pixel Level, FMI<sub>p</sub>)、信息熵 (Entropy, EN)、平均梯度 (Average Gradient, AG)、空间频率 (Spatial Frequency, SF) 和基于边缘保持度的指标 (Quality Assessment of Blended Features,  $Q^{AB/F}$ ) 作为评价指标, 上述指标均为正向指标。

### 3.2 实验环境及参数设置

IPC-Net 为预训练网络, 该网络训练时的 Batchsize 设置为 32、训练 100 个 Epoch; 融合网络 Batchsize 设置为 16、训练 50 个 Epoch。模型使用 Adam 优化器来更新参数, 初始学习率设置为  $10^{-4}$ 。训练硬件环境为 NVIDIA GTX 3060 GPU, 软件环境基于 Python 3.7, CUDA Toolkit 11.3 搭建, 网络模型在 Pytorch 框架下完成训练。模型推理阶段加载训练完成的最优权重, 不更新参数, 仅执行前向传播生成最终融合图像。

### 3.3 对比实验与结果分析

为验证本文方法的有效性, 本节在 3.1 节所

述的三个数据集上开展测试实验, 并与当前主流图像融合方法进行对比, 从主观视觉效果与客观定量指标两个维度完成性能分析。实验中所有对比方法均采用原论文公开的预训练模型, 在统一测试集与评价代码下完成推理与指标计算, 确保实验公平性。

#### 3.3.1 客观定量指标分析

表 1 为不同数据集上各评价指标的平均值, 其中黑色加粗字体为最优值, 下划线为次优值。实验结果表明, DHIFusion 在 MSRS 数据集上有五项指标取得最优, 相较于次优算法, MI 提升 3.76%, EN 提升 0.19%, AG 提升 7.76%,  $Q^{AB/F}$  提升 7.86%, FMI<sub>p</sub> 提升 0.18%; 在 TNO 数据集上有五项指标取得最优, 相较于次优算法, MI 提升 4.09%、EN 提升 0.66%、AG 提升 0.45%、 $Q^{AB/F}$  提升 3.60%、FMI<sub>p</sub> 提升 1.31%; 在自制数据集上, 六项指标均取得最优值, 相较于次优算法, MI 提升 4.99%、EN 提升 0.49%、AG 提升 1.03%、SF 提升 6.46%、 $Q^{AB/F}$  提升 3.22%,

表 1 不同方法在不同数据集上的对比

Tab. 1 Performance comparison of different methods across multiple datasets

Datasets	Algorithm	Evaluation metric					
		MI	EN	AG	SF	$Q^{AB/F}$	FMI <sub>p</sub>
MSRS	SeAFusion <sup>[7]</sup>	4.037 6	6.651 3	3.696 6	11.106 1	<u>0.674 5</u>	0.917 9
	PIAFusion <sup>[14]</sup>	<u>4.582 2</u>	6.570 9	3.759 8	11.490 7	0.642 3	0.920 5
	DATFuse <sup>[12]</sup>	3.896 5	6.479 5	3.573 9	10.926 9	0.640 3	0.912 3
	SwinFusion <sup>[13]</sup>	4.546 9	6.621 4	3.565 9	11.087 9	0.654 2	0.920 7
	DIVFusion <sup>[15]</sup>	2.394 1	<u>7.116 6</u>	<u>4.690 5</u>	<u>12.598 2</u>	0.340 5	0.879 5
	CrossFuse <sup>[6]</sup>	3.112 2	6.496 5	3.008 7	9.611 5	0.560 3	0.924 8
	C2RF <sup>[11]</sup>	1.825 9	6.339 2	3.236 0	10.712 4	0.221 6	0.883 8
	TITA <sup>[10]</sup>	4.161 2	6.565 1	3.459 4	10.905 9	0.666 9	<u>0.930 2</u>
	FreeFusion <sup>[8]</sup>	2.593 1	6.872 7	4.520 9	<b>12.788 9</b>	0.555 4	0.913 4
Ours	<b>4.754 5</b>	<b>7.130 2</b>	<b>5.054 6</b>	12.256 7	<b>0.727 5</b>	<b>0.931 9</b>	
TNO	SeAFusion <sup>[7]</sup>	2.838 2	7.132 8	4.980 3	12.252 6	0.487 9	0.898 4
	PIAFusion <sup>[14]</sup>	3.246 3	6.893 9	4.426 8	11.397 9	0.517 3	0.908 3
	DATFuse <sup>[12]</sup>	3.132 3	6.453 1	3.560 1	9.605 7	0.499 7	0.877 1
	SwinFusion <sup>[13]</sup>	3.342 1	6.890 9	4.211 2	10.722 3	0.521 4	0.902 1
	DIVFusion <sup>[15]</sup>	2.222 4	<u>7.193 3</u>	5.260 1	13.462 5	0.311 7	0.853 4
	CrossFuse <sup>[6]</sup>	2.963 7	6.892 4	3.737 3	9.887 2	0.420 9	0.898 0
	C2RF <sup>[11]</sup>	2.478 4	6.871 8	3.698 2	9.328 9	0.381 0	0.883 2
	TITA <sup>[10]</sup>	<u>3.411 6</u>	6.863 5	3.859 7	10.275 0	<u>0.544 5</u>	<u>0.910 7</u>
	FreeFusion <sup>[8]</sup>	2.019 0	6.915 1	<u>5.385 9</u>	<b>14.102 3</b>	0.438 2	0.879 4
Ours	<b>3.551 0</b>	<b>7.240 9</b>	<b>5.409 9</b>	<u>13.761 2</u>	<b>0.564 1</b>	<b>0.922 6</b>	

续表 1 不同方法在不同数据集上的对比

Tab. 1 Performance comparison of different methods across multiple datasets

Datasets	Algorithm	Evaluation metric					
		MI	EN	AG	SF	$Q^{A/B/F}$	FMI_p
Self-made	SeAFusion <sup>[7]</sup>	4.478 6	7.258 8	1.570 6	4.477 7	0.551 1	0.912 5
	PIAFusion <sup>[14]</sup>	4.499 6	7.165 0	1.816 3	5.134 2	0.532 5	0.918 3
	DATFuse <sup>[12]</sup>	3.865 7	6.602 7	1.136 6	3.680 0	0.410 2	0.910 7
	SwinFusion <sup>[13]</sup>	4.582 1	7.156 4	1.602 2	4.797 0	0.573 0	0.917 7
	DIVFusion <sup>[15]</sup>	2.883 2	7.255 9	<u>2.137 1</u>	<u>6.396 4</u>	0.402 9	0.891 4
	CrossFuse <sup>[6]</sup>	3.150 8	6.729 6	1.254 5	3.589 2	0.448 0	0.936 5
	C2RF <sup>[11]</sup>	3.317 7	7.246 0	1.570 4	4.746 1	0.368 7	0.922 8
	TITA <sup>[10]</sup>	<u>4.806 1</u>	7.293 3	1.454 0	4.381 7	<u>0.574 5</u>	<u>0.937 1</u>
	FreeFusion <sup>[8]</sup>	3.097 2	<u>7.305 6</u>	2.122 5	5.858 1	0.574 2	0.931 1
	Ours	<b>5.045 7</b>	<b>7.341 3</b>	<b>2.159 1</b>	<b>6.809 6</b>	<b>0.593 0</b>	<b>0.938 8</b>

FMI\_p 提升 0.18%, 综合表明本文算法融合性能优势显著。从图像融合物理机理分析, 空间频率 SF 主要反映图像高频纹理振荡强度与边缘杂乱程度。FreeFusion, DIVFusion 等模型未引入光照自适应约束与伪影抑制模块, 融合过程直接保留大量冗余高频噪声、边缘毛刺及光照光晕信息, 依靠无效高频杂散信息虚高 SF 数值。与之相比, 本文方法通过 C<sup>2</sup>IAPF 模块针对性强化热目标与结构细节相关的有效高频特征, 更注重融合图像的结构一致性, 同时抑制了无效高频噪声, 虽 SF 略低于以上两种方法, 但主观视觉质量更优。

### 3.3.2 主观视觉分析

为进一步验证本文方法在过渡光照条件下的合理性, 本文在三种测试数据集中分别选取傍晚局部暗区、浓雾光照衰减、夜间局部强光等场景开展可视化对比分析, 部分融合结果如图 4~图 6 所示, 其中差异显著区域以红色、绿色方框标注(彩图见期刊电子版)。

图 4 为含有局部暗区的白天常规光照场景融合结果对比。CrossFuse 和 C2RF 的融合结果过度偏向可见光, 导致绿色框内树干目标特征丢失。DATFuse, CrossFuse, C2RF 及 FreeFusion 方法的融合结果人物腿部轮廓虚化, 细节辨识度低。DIVFusion 方法生成的图像中人物边界、树干纹理因噪点覆盖丢失清晰边缘。SeAFusion, PIAFusion, DATFuse, SwinFusion 及 TITA 方法的融合图像街道区域过度平滑, 在不同程度上忽

略了红外图像信息。相较之下, 本文方法有效平衡可见光纹理细节与红外热特征, 人物边缘锐利, 完整保留了局部暗区树干纹理及轮廓, 在全局场景理解与局部细节呈现上均表现更优。

如图 5 所示为白天浓雾光照不均场景的融合结果对比。从图中可知该场景中浓雾导致可见光的光照散射, 雾区细节完全丢失。对比结果显示 PIAFusion, DATFuse, SwinFusion 及 CrossFuse 方法的融合图像显著弱化红外模态表达, 浓雾覆盖区域内集装箱的纹理基本丢失。此外, DATFuse 方法的融合结果未保留绿色框内山脉边界与纹理特征。DIVFusion 方法的融合结果在无雾区与雾区过渡生硬, 画面层次感割裂, 视觉效果较差。C2RF, TITA 及 FreeFusion 等方法的融合图像对天空烟雾区域存在显著的特征湮灭缺陷, 相较之下, 本文方法针对光照衰减的雾区, 通过动态提升红外模态贡献权重, 保留了雾中人物、山脉、集装箱的清晰边缘及纹理信息, 兼顾了雾体的空间弥散结构与灰度信息, 整体画面协调、层次感自然。

如图 6 所示为含有局部亮区的夜间极端弱光场景下的融合结果对比。DATFuse 和 CrossFuse 方法的融合结果显著弱化了人物特征。PIAFusion, DIVFusion 及 CrossFuse 方法的融合结果抑制了窗口区域的信息表达。SeAFusion 方法生成的图像存在严重线条伪影, C2RF 方法生成的图像整体线条扭曲。SwinFusion, TITA 及 FreeFusion 方法的融合结果丢失了窗格大部

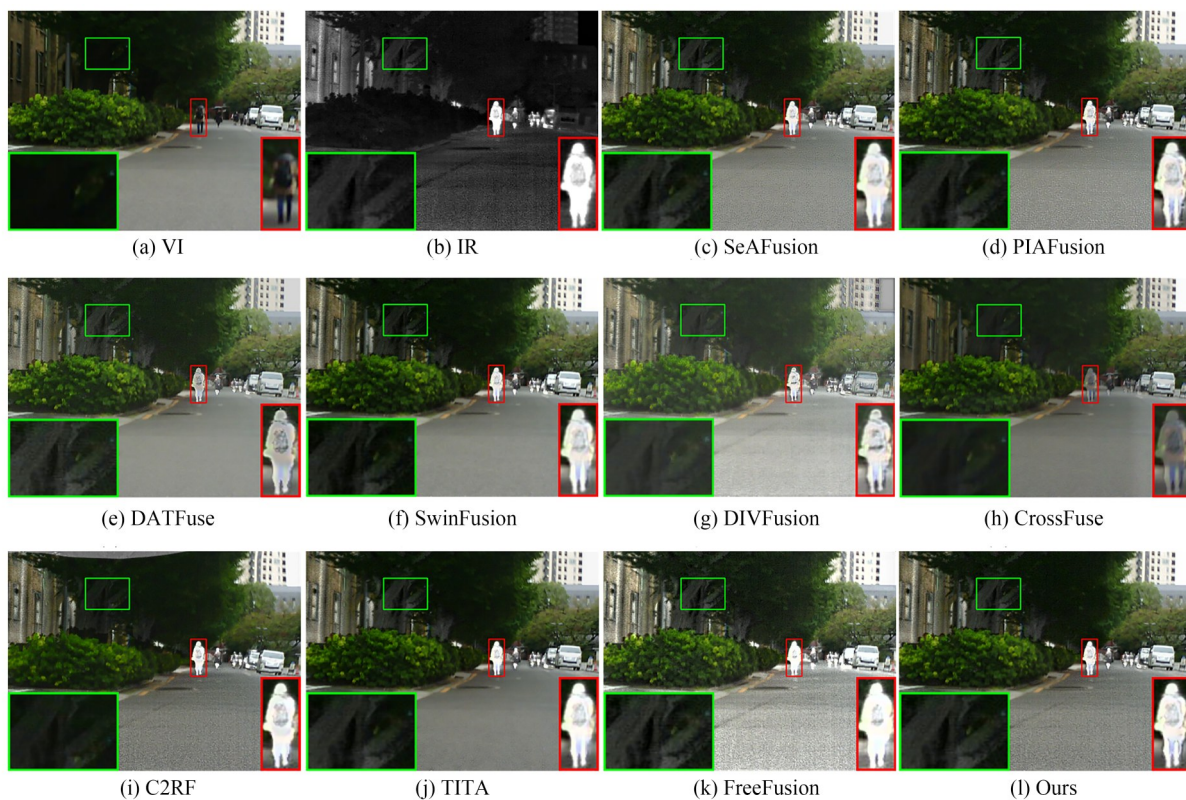


图 4 含有局部暗区的白天场景的图像视觉对比

Fig. 4 Visual contrast of daytime scene images with local dark areas

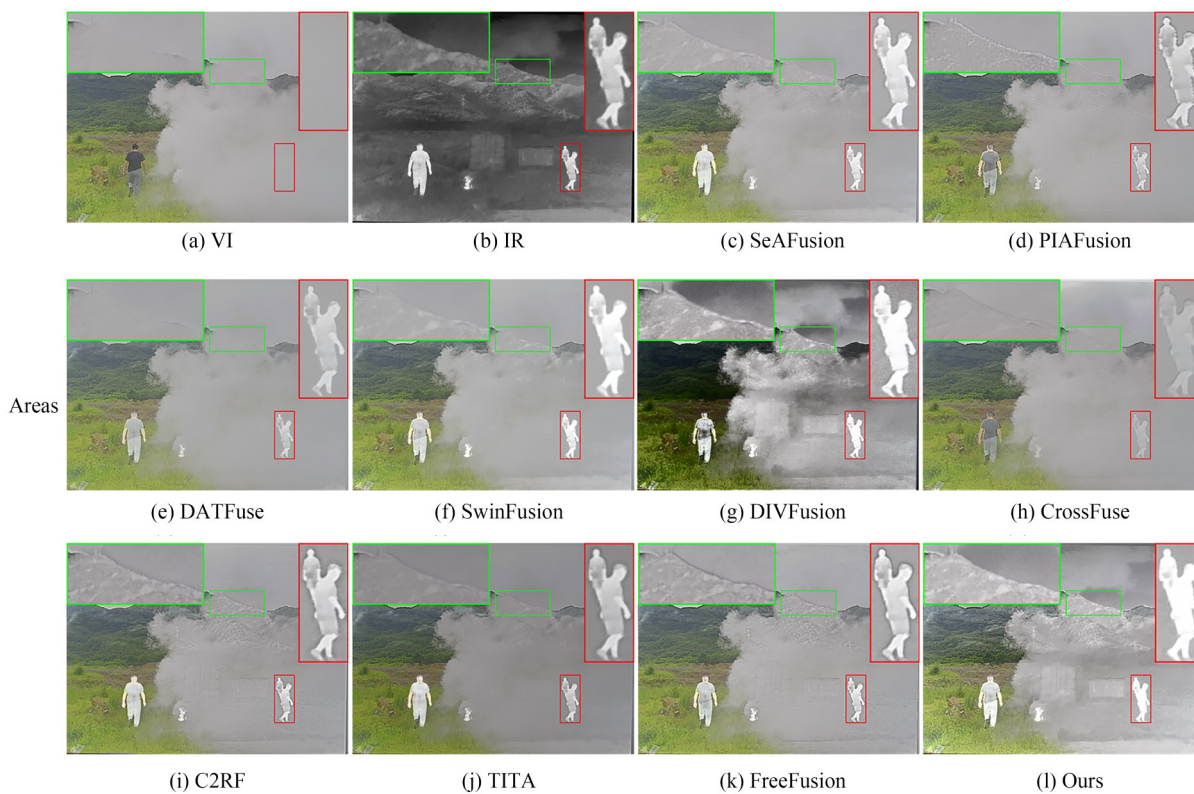


图 5 白天浓雾下光照不均场景的图像视觉对比

Fig. 5 Visual contrast of images in illumination-uniformity scenes under heavy daytime fog

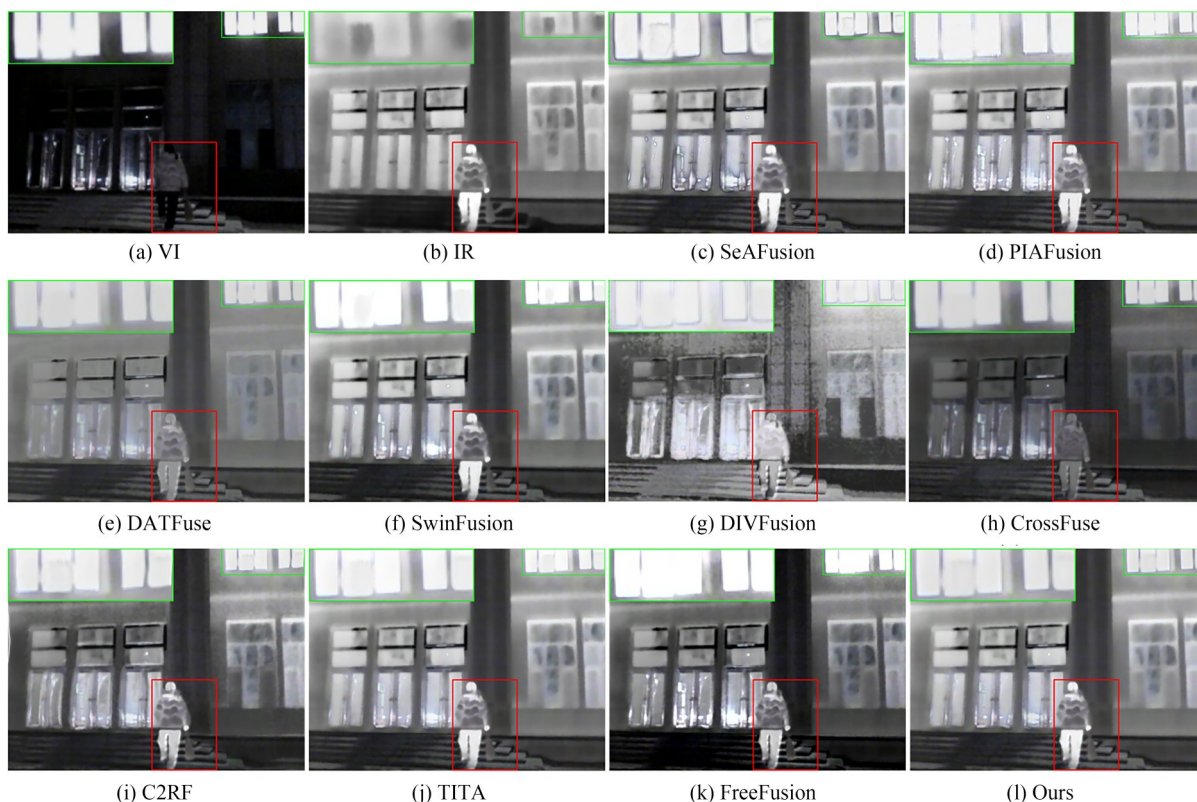


图 6 含有局部亮区的夜间弱光场景的图像视觉对比

Fig. 6 Visual contrast of images in low-light night scenes with local bright areas

分线条结构。综合对比可知,本文方法的融合图像实现亮暗区域均衡适配,完整保留弱光下人物特征与亮区窗格细节,整体场景还原度最佳。

### 3.4 消融实验

#### 3.4.1 IPC-Net 中 CoordAtt Block 嵌入位置分析

针对光照分布的动态变化差异,本文在 IPC-Net 中分层嵌入坐标注意力机制。为进一步验证 CoordAtt Block 的有效性及其嵌入位置对融合网络性能的影响,本节在 MSRS 数据集

上进行消融实验,实验结果如表 2 所示,其中:第 2 行浅层嵌入表示在第二层卷积之后嵌入;第 3 行深层嵌入表示在第四层卷积之后嵌入;第 4 行夹卷积嵌入表示每一层卷积之后均嵌入。实验结果表明,本文的分层嵌入方式取得最佳表现。

#### 3.4.2 加权梯度一致性光照损失分析

本节在 MSRS 数据集上对加权梯度一致性光照损失  $L_{illum\_grad}$  开展消融实验,实验结果见表 3。实验设置两组对照:第 1 组将损失函数中的动

表 2 CoordAtt Block 嵌入位置对应的六项指标均值

Tab. 2 Mean values of six indicators for each CoordAtt Block embedding position

嵌入方式	MI	EN	AG	SF	$Q^{AB/F}$	FMI_p
无	4.438 2	6.412 5	3.521 7	11.496 7	0.667 1	0.928 5
浅层	4.492 6	6.583 7	3.762 5	11.634 5	0.679 3	0.929 2
深层	4.437 5	6.389 2	3.486 3	11.523 2	0.642 7	0.913 8
夹卷积	4.443 8	6.534 3	3.805 6	11.570 0	0.687 2	0.930 3
分层	4.754 5	7.130 2	5.054 6	12.256 7	0.727 5	0.931 9

表 3 不同损失函数对应的六项指标均值

Tab. 3 Mean values of six metrics corresponding to different loss functions

Loss	MI	EN	AG	SF	$Q^{AB/F}$	FMI_p
1	4.437 1	6.521 8	3.715 4	11.637 2	0.648 2	0.916 4
2	4.469 5	6.683 5	3.892 7	11.722 3	0.666 7	0.922 6
3	<b>4.754 5</b>	<b>7.130 2</b>	<b>5.054 6</b>	<b>12.256 7</b>	<b>0.727 5</b>	<b>0.931 9</b>

态光照权重固定为 0.5;第 2 组将损失函数中的梯度差替换为原始图像像素强度差,其余设置保持不变;第三组为  $L_{illum\_grad}$ 。结果表明,该损失通过物理引导的梯度解耦与动态权重调控,有效规避了固定权重导致的模态信息平均化、局部特征丢失问题,在 MI, EN, FMI\_p 指标上表现最优;同时其梯度一致性约束可强化高频细节保留,保障目标轮廓完整与锐度,在 AG, SF,  $Q^{AB/F}$  指标上亦获最优结果。

### 3.4.3 模型模块分析

本节在 MSRS 数据集做消融实验,分别对 IPC-Net, MC<sup>2</sup>MA 和 C<sup>2</sup>IAPF 的有效性进行分析,实验结果如表 4 所示,其中 C<sup>2</sup>IAPF\* 表示未引入光照权重的融合模块,√表示使用相应模块,×表示未使用。实验结果表明,同时集成 IPC-Net, MC<sup>2</sup>MA 和 C<sup>2</sup>IAPF 时,指标取得最优,整体性能最佳。

部分融合结果如图 7 所示,由图可知,仅

表 4 含有不同模块对应的六项指标均值

Tab. 4 Mean values of six metrics corresponding to different modules

Group	IPC-Net	C <sup>2</sup> IAPF*	MC <sup>2</sup> MA	C <sup>2</sup> IAPF	MI	EN	AG	SF	$Q^{AB/F}$	FMI_p
1	√	×	×	×	4.411 5	6.572 7	3.814 6	11.575 9	0.689 9	0.930 4
2	×	√	×	×	4.445 7	6.457 2	3.691 7	11.301 1	0.671 9	0.930 2
3	×	×	√	×	4.635 2	6.618 7	3.832 6	11.703 4	0.676 5	0.930 5
4	√	√	×	×	4.502 6	6.588 9	3.821 4	11.557 2	0.688 5	0.929 5
5	√	×	×	√	4.578 2	6.754 9	3.872 5	11.653 7	0.691 2	0.930 0
6	√	√	√	×	4.688 7	6.934 6	3.984 6	11.697 9	0.694 5	0.930 1
7	√	×	√	√	<b>4.754 5</b>	<b>7.130 2</b>	<b>5.054 6</b>	<b>12.256 7</b>	<b>0.727 5</b>	<b>0.931 9</b>

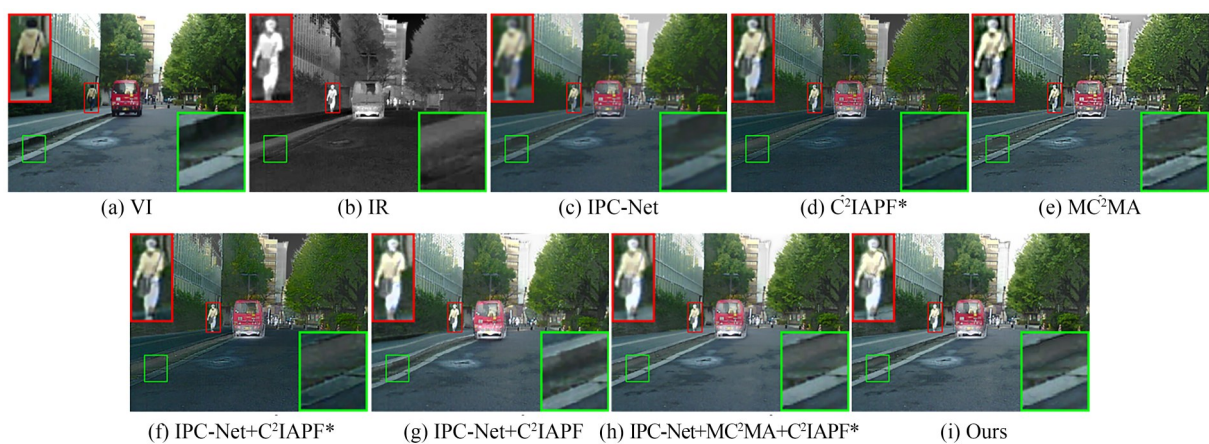


图 7 各个模块消融研究的视觉质量比较

Fig. 7 Comparison of visual quality of ablation studies in individual modules

有 IPC-Net 时图像的色彩理解较为正确,但丢失大量纹理细节(彩图见期刊电子版)。C<sup>2</sup>IAPF\*做基础融合时,整体融合效果有所提升,但人物和白色实线处的边界引入大量噪点。使用 MC<sup>2</sup>MA 的结果保留了较多纹理信息,但针对光影的处理较差,树的边缘出现锯齿效应。从图 7(f)和(g)可知,光照权重引导融合过程的画面明暗层次自然。本文方法对细节处理最佳,人物边界清晰、白色实线内水泥路段处特征也实现完整表达,其综合表现最优。

### 3.5 模型效率分析

本节在统一硬件平台下,对可在相同推理环境运行的代表性融合方法开展效率对比,统计模型参数量(Parameters, Params)、浮点运算量(Floating Point Operations, FLOPs)及单张图像平均推理时间(Average Inference Time, AIT),实验结果如表 5 所示,其中黑色加粗字体表示最优,下划线表示次优。本文方法推理速度处于优势水平,计算开销整体合理,虽然在 FLOPs 指标尚未达最优,但实现了融合精度与效率的良好平衡,具备较好的实际部署潜力。

表 5 不同方法在两个数据集上的 Params, AIT 及 FLOPs

Tab. 5 Params, AIT and FLOPs of different methods on two datasets

Methods	Parameter	MSRS(640×480)		TNO(768×576)	
	Params/M	AIT/s	FLOPs/G	AIT/s	FLOPs/G
SwinFusion <sup>[13]</sup>	<b>0.973 7</b>	2.495 7	<u>298.737 6</u>	2.352 9	<u>430.182 1</u>
CrossFuse <sup>[6]</sup>	23.310 0	0.496 5	<b>133.708 5</b>	0.508 5	<b>194.279 1</b>
C2RF <sup>[11]</sup>	15.662 0	0.747 9	773.984 6	0.619 1	1 114.537 8
FreeFusion <sup>[8]</sup>	5.669 8	<u>0.244 8</u>	451.654 7	<u>0.253 9</u>	650.354 3
TITA <sup>[10]</sup>	1.386 9	2.990 9	377.843 2	2.963 1	544.093 6
Ours	<u>1.213 6</u>	<b>0.122 5</b>	353.658 0	<b>0.251 8</b>	494.427 5

## 4 结 论

本文提出光照权重调配机制引导的双流逐层交互式红外与可见光图像融合网络。该网络通过逐层交替提取与跨模态融合实现层级特征深度交互,借助阶梯式扩张与全局上下文块捕获多尺度长距离语义关联;嵌入坐标注意力的光照先验分类网络将光照信息集成到双流网络,缓解光照不平衡;利用差分特征与并行双注意力机制强化模态互补特征选择,动态调整融合权重以避免单一模态主导。多个公开数据集上的实验结果及主客观分析表明,本文方法得到的融合图像在边缘区分度、语义一致性以及细节保留等方面表现更优,针对局部暗区、浓雾、夜间弱光等复杂场景下的目标轮廓与细节保留效果更为突出。

与次优算法相比,在 MSRS 测试集上 MI, AG 分别提升 3.76%, 7.76%; TNO 测试集上 MI, Q<sup>AB/F</sup> 分别提升 4.09%, 3.60%; 在自制数据集上 SF 提升 6.46%。在后续研究中,将继续优化本文所提方法,减少模型参数以提升融合效率。此外,针对多维特征的交互与矫正问题也将进一步探索。

### 作者贡献声明:

韩玉兰:论文构思指导、对论文质量、内容逻辑结构及语言表达进行审核优化,确保学术规范性;

吴桐:提出核心研究思路与方案,完成论文撰写,负责实验的整体设计及数据处理与分析;

黄海茹:完成论文辅助工作及发表前基础准备。

## 参考文献:

- [1] 吴一非, 杨瑞, 吕其深, 等. 红外与可见光图像融合: 统计分析, 深度学习方法和未来展望[J]. 激光与光电子学进展, 2024, 61(14): 32-50.  
WU Y F, YANG R, LÜ Q S, *et al.* Infrared and visible image fusion: statistical analysis, deep learning approaches and future prospects[J]. *Laser & Optoelectronics Progress*, 2024, 61(14): 32-50. (in Chinese)
- [2] ZHANG X C, DEMIRIS Y. Visible and infrared image fusion using deep learning[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(8): 10535-10554.
- [3] 李晓玲, 陈后金, 李艳凤, 等. 多重关系感知的红外与可见光图像融合网络[J]. 电子与信息学报, 2024, 46(5): 2217-2227.  
LI X L, CHEN H J, LI Y F, *et al.* Infrared and visible image fusion network with multi-relation perception [J]. *Journal of Electronics & Information Technology*, 2024, 46(5): 2217-2227. (in Chinese)
- [4] MA J Y, YU W, LIANG P W, *et al.* Fusion-GAN: a generative adversarial network for infrared and visible image fusion [J]. *Information Fusion*, 2019, 48: 11-26.
- [5] LI H, WU X J. DenseFuse: a fusion approach to infrared and visible images[J]. *IEEE Transactions on Image Processing*, 2019, 28(5): 2614-2623.
- [6] LI H, WU X J. CrossFuse: a novel cross attention mechanism based infrared and visible image fusion approach [J]. *Information Fusion*, 2024, 103: 102147.
- [7] TANG L F, YUAN J T, MA J Y. Image fusion in the loop of high-level vision tasks: a semantic-aware real-time infrared and visible image fusion network [J]. *Information Fusion*, 2022, 82: 28-42.
- [8] ZHAO W D, CUI H S, WANG H P, *et al.* FreeFusion: infrared and visible image fusion via cross reconstruction learning [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025, 47(9): 8040-8056.
- [9] XU H, MA J Y, JIANG J J, *et al.* U2Fusion: a unified unsupervised image fusion network [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(1): 502-518.
- [10] HU X Y, JIANG J J, WANG C Y, *et al.* Balancing task-invariant interaction and task-specific adaptation for unified image fusion [C]. 2025 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 19-25, 2025, Honolulu, HI, USA. IEEE, 2026: 11262-11272.
- [11] TANG L F, YAN Q L, XIANG X Y, *et al.* C2RF: bridging multi-modal image registration and fusion via commonality mining and contrastive learning[J]. *International Journal of Computer Vision*, 2025, 133(8): 5262-5280.
- [12] TANG W, HE F Z, LIU Y, *et al.* DATFuse: infrared and visible image fusion via dual attention transformer [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(7): 3159-3172.
- [13] MA J Y, TANG L F, FAN F, *et al.* SwinFusion: cross-domain long-range learning for general image fusion via swin transformer [J]. *IEEE/CAA Journal of Automatica Sinica*, 2022, 9(7): 1200-1217.
- [14] TANG L F, YUAN J T, ZHANG H, *et al.* PIA-Fusion: a progressive infrared and visible image fusion network based on illumination aware [J]. *Information Fusion*, 2022, 83: 79-92.
- [15] TANG L F, XIANG X Y, ZHANG H, *et al.* DIFusion: Darkness-free infrared and visible image fusion [J]. *Information Fusion*, 2023, 91: 477-493.
- [16] CHEN J, YANG L L, LIU W, *et al.* LENFusion: a joint low-light enhancement and fusion network for nighttime infrared and visible image fusion [J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 5018715.
- [17] CHEN X X, XU S W, HU S H, *et al.* MGFA: a multi-scale global feature autoencoder to fuse infrared and visible images[J]. *Signal Processing: Image Communication*, 2024, 128: 117168.
- [18] LI X L, LI Y F, CHEN H J, *et al.* CCAFusion: cross-modal coordinate attention network for infrared and visible image fusion [J]. *IEEE Transac-*

*tions on Circuits and Systems for Video Technology*, 2024, 34(2): 866-881.

[19] TOET A. The TNO multiband image data collection[J]. *Data in Brief*, 2017, 15: 249-251.

作者简介:



韩玉兰(1984—),女,黑龙江大庆人,讲师,硕士生导师,2007年、2010年于黑龙江大学分别获得学士、硕士学位,2019年于哈尔滨工业大学获得博士学位主要从事机器学习、图像处理、计算机视觉、图像重构。E-mail: hanyulan@hrbust.edu.cn

作者简介:



吴桐(2001—),女,黑龙江哈尔滨人,硕士研究生,2023年于齐齐哈尔大学获得学士学位,现就读于哈尔滨理工大学,主要从事基于深度学习的红外与可见光图像融合。E-mail: 1685914181@qq.com