一种双分支特征交互融合的高效红外图像彩色化方法

陈宇[1]，詹伟达[1]，蒋一纯[1]，朱德鹏[1]，韩登[1,2]
（1. 长春理工大学电子信息工程学院，130022，长春；2. 吉林省知行物联网研究院有限公司，130117，长春）

摘要：针对现有的红外图像彩色化方法在全局特征捕获和计算复杂度方面存在显著局限性的问题，提出了一种双分支特征交互融合的高效红外图像彩色化方法。设计双分支编码器，通过局部特征提取分支获取局部空间上下文信息，确保细粒度特征的捕获，并通过全局特征提取分支获取全局特征，满足对长程依赖的需求。设计交互融合模块，对两个分支提取到的特征进行有效整合，显著增强了模型的整体性能。在解码器部分提出上下文聚合模块，进一步优化多尺度语义特征的聚合能力，改善了彩色化结果的边缘清晰度和细节表现力。在KAIST和FLIR数据集上进行广泛实验验证，结果表明：与现有方法相比，所提方法在两个数据集上均具有更高的彩色化质量，峰值信噪比分别达到28.645、30.459 dB，结构相似度达到0.507、0.725，均优于对比方法，且有效性和先进性也得到了验证。研究结果可为提升红外图像的可读性与可解释性以及提高夜视与恶劣环境下的观测能力提供参考。

关键词：红外图像彩色化；细粒度特征；长程依赖；交互融合；上下文聚合
中图分类号：TN911.73；TP391 **文献标志码**：A
DOI：10.7652/xjtuxb202508020 **文章编号**：0253-987X(2025)08-0211-12

An Efficient Infrared Image Colorization Method Based on Dual-Branch Feature Interaction Fusion

CHEN Yu[1], ZHAN Weida[1], JIANG Yichun[1], ZHU Depeng[1], HAN Deng[1,2]
(1. Electronic and Information Engineering, Changchun University of Science and Technology, Changchun 130022, China; 2. Jilin Province Zhixing IoT Research Institute Co., Ltd., Changchun 130117, China)

Abstract: To address the significant limitations of existing infrared image colorization methods in global feature capture and computational complexity, an efficient infrared image colorization method based on dual-branch feature interaction fusion is proposed. A dual-branch encoder is designed, where the local feature extraction branch captures local spatial context information to ensure fine-grained feature acquisition, while the global feature extraction branch obtains global features to meet long-range dependency requirements. An interaction fusion module is developed to effectively integrate features extracted from both branches, significantly enhancing the model's overall performance. In the decoder part, a context aggregation module is proposed to further optimize multi-scale semantic feature aggregation, improving edge clarity and detail representation in the colorization results. Extensive experimental validation on the KAIST and FLIR datasets demonstrates

收稿日期：2025-02-21。 作者简介：陈宇(1998—)，男，博士生；詹伟达(通信作者)，男，教授，博士生导师。 基金项目：吉林省重点研发计划资助项目(20240302029GX)；吉林省发展和改革委员会创新能力建设专项资助项目(2024C021-8)。
网络出版时间：2025-03-19 网络出版地址：https://link.cnki.net/urlid/61.1069.T.20250319.1429.002

that compared to existing methods, the proposed approach achieves superior colorization quality on both datasets, with peak signal-to-noise ratios reaching 28.645 dB and 30.459 dB, and structural similarity scores of 0.507 and 0.725, respectively, outperforming all comparative methods. The effectiveness and advancement of the method are thus verified. The research findings provide valuable references for enhancing the readability and interpretability of infrared images, as well as improving observation capabilities in night vision and harsh environments.

Keywords: infrared image colorization; fine-grained feature; long-range dependency; interactive fusion; context aggregation

红外图像因其能够捕捉人眼无法感知的热辐射信息,在弱光、烟雾和恶劣天气等复杂环境中,尤其在目标检测、识别与追踪等方面展现了独特的优势[1]。然而,由于红外图像缺乏丰富的颜色、细节和纹理信息,使得其在视觉表现力和直观感知上存在较大局限,难以为实际工程应用提供充分的场景理解与分析。为了提升红外图像的可视化效果,红外图像彩色化技术[2-3]通过引入颜色信息,显著增强了图像的细节解析和视觉呈现,使得红外图像更加真实、易于理解,特别是在军事侦察、自动驾驶和安防监控等应用领域,彩色化图像能够有效提高目标的可辨识度和场景理解能力。

近年来,基于卷积神经网络(CNN)的红外图像彩色化方法得到了广泛研究,尤其是以 UNet[4] 作为主干网络,用于学习像素之间的映射关系。Aswatha 等[5]提出了一种双重编解码网络,通过在不同深度层次进行尺度变化,有效减少了图像中的伪影。Isola 等[6]将图像彩色化视为图像到图像的映射问题,使用生成对抗网络[7]来增强彩色化效果。Kuang 等[8]设计了复合损失函数来约束生成的图像,提高了红外图像彩色化的质量。Liao 等[9]利用多重混合跳跃连接,通过设计特征提取模块和注意力机制,进一步提升彩色化效果。虽然这类方法[10-11]在局部特征提取方面表现出色,但在捕捉全局特征和长程依赖关系方面存在明显的局限性。

视觉 Transformer(ViT)[12]的引入为红外图像彩色化任务提供了新的思路。ViT 能够有效学习图像中的长程依赖关系,弥补卷积操作在捕获全局信息时的不足。Torbunov 等[13]将 ViT 与 UNet 相结合,实现了多尺度局部空间特征的提取,并在全局信息和局部信息之间取得平衡。He 等[14]利用 CNN 编码器提取局部空间上下文信息,同时利用 ViT 编码器捕捉长程依赖关系,显著提升了彩色化性能。然而,尽管基于 ViT 的彩色化方法在全局建模方面展现出卓越的性能,但自注意力机制的计算复杂度随着图像尺寸增加呈二次增长,导致计算资源消耗较大。

目前,状态空间模型(SSM)[15]引起了许多研究者的关注。由于 Mamba 模型通过其高效的选择性扫描机制显著增强了 SSM 的性能,不仅使得模型能够有效捕获序列数据中的长程依赖关系,而且保证了其计算复杂度与输入大小呈线性关系。Liu 等[16]首次将 Mamba 引入视觉领域,并在多个视觉任务中取得了显著效果。当前,研究者们已将 Mamba 应用于图像分类、目标检测和图像分割等任务,如 RSMamba[17]、MambaIR[18] 和 TransMamba[19] 等。特别是,Zhai 等[20]首次将 Mamba 应用于红外图像彩色化,设计了结合深度卷积和注意力机制增强图像边界的分辨能力,并减少了序列模型中潜在的混淆。然而,由于无法捕获局部细粒度特征,直接应用 Mamba 会导致彩色化效果受限。

为了解决上述问题,本文提出了一种双分支特征交互融合的高效红外图像彩色化方法,通过提升细粒度特征提取与长程依赖建模能力,进而优化彩色化结果的边界清晰度和细节表现力。本文的主要贡献如下:①设计了双分支编码器,采用局部特征提取分支(LFEB)捕捉局部空间上下文信息,全局特征提取分支(GFEB)提取全局信息和长程依赖关系,以弥补细粒度特征与全局一致性的不足;②提出了交互融合模块(IFM),将一个分支的特征映射作为另一个分支的互补信息源,使模型能够获取更为全面的信息,提升彩色化性能;③设计了上下文聚合模块(CAM)以替代传统解码器块,增强了多尺度信息的利用能力,进一步优化了红外图像彩色化的细节与边界清晰度。

1 本文算法的网络设计

1.1 总体结构

如图 1 所示,本文算法包含双分支编码器、IFM 和 CAM。其中,双分支编码器由 LFEB 和 GFEB

构成，LFEB 提取图像的局部空间特征，能够有效识别车辆边缘和道路纹理等细节，使得彩色化结果在局部细节上更为丰富；GFEB 擅长捕获全局信息和长程依赖关系，能够整合远处建筑和街道背景等信息，提升彩色化图像的全局一致性。随后，IFM 将双分支编码器的输出特征映射交互融合，增强了特征表达的完整性。最后，CAM 对融合后的特征进行上采样处理，聚合多尺度上下文信息，生成边界清晰、细节丰富的彩色化图像。

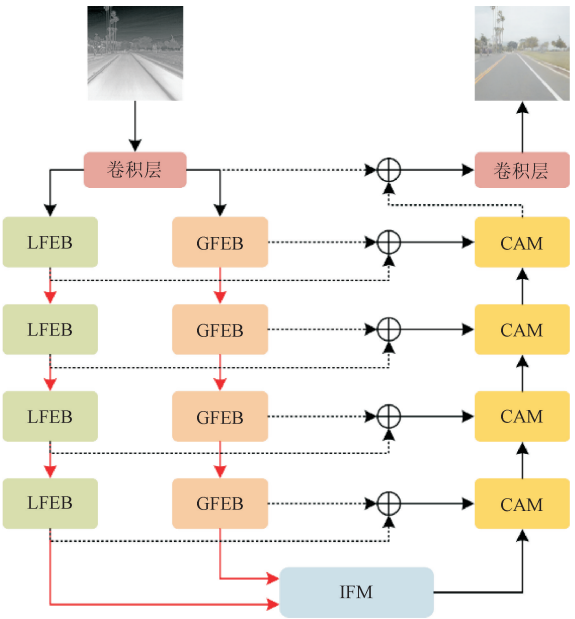


图 1　本文算法的结构

Fig. 1　Structural diagram of the proposed algorithm

1.2　LFEB 设计

为了提升 CNN 的多尺度特征提取能力，本文设计了 LFEB 增强特征表示的精度和丰富性，如图 2 所示。LFEB 通过多层次的残差连接实现了特征的多尺度建模，借助不同数量和不同尺度的卷积核，使得网络能够在多个感受野上同时提取特征信息。具体而言，输入特征映射首先通过尺寸为 1×1 的卷积进行通道变换，生成的新特征映射进一步划分为 s 个特征子集，每个子集具有相同的空间分辨率和 $1/s$ 个通道数。接着，除第一个特征子集外的每一个子集都经过一个尺寸同为 3×3 的卷积核（$X_1 \sim X_4$）处理，并与后续子集相加，形成级联关系，逐步融合为多个尺度特征（$Y_1 \sim Y_4$）。随后，通过通道维度上的连接操作，将所有子集特征映射重新组合，并采用尺寸为 1×1 的卷积统一处理。最后，与原始输入特征连接以生成最终输出特征。LFEB 的表达式如下

$$y_i = \begin{cases} x_i, & i = 1 \\ f^i_{\mathrm{Conv}}(x_i), & i = 2 \\ f^i_{\mathrm{Conv}}(x_i + y_{i-1}), & 2 < i \leqslant s \end{cases} \tag{1}$$

式中：x_i 为特征子集；$f^i_{\mathrm{Conv}}(\cdot)$ 为尺寸为 3×3 的卷积；y_i 为 $f^i_{\mathrm{Conv}}(\cdot)$ 的输出。

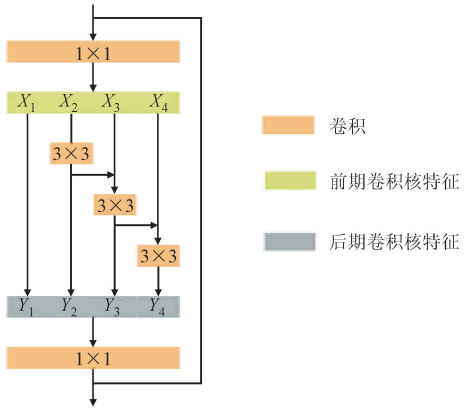


图 2　LFEB 结构

Fig. 2　Structure of LFEB

值得注意的是，LFEB 能够细致捕获物体边缘及纹理区域的多尺度特征，边缘清晰度和细节丰富度质量较高，显著增强了编码器在多尺度特征提取上的表现，提供了更加丰富且精细的特征表示。

1.3　GFEB 设计

由于 CNN 的感受野较小，现有的红外图像彩色化方法难以有效捕获图像的全局上下文信息；而基于 ViT 的方法虽扩展了感受野，但计算效率方面的限制使其在红外图像彩色化任务中存在一定挑战。相比之下，Mamba 以线性复杂度捕捉长程依赖关系，展现出更高效的生成任务性能。然而，原始的 Mamba 将图像彩色化为一维序列默认处理方式，忽略了图像中的局部上下文信息，对红外图像彩色化任务的性能有所影响。

针对上述问题，本文提出采用 GFEB 解决原始 Mamba 在图像边界处混淆的问题，具体过程如图 3(a)所示。在状态空间模型的扫描序列中设计了可学习的填充标记，同时插入在不具有直接空间关联的标记之间，增强了图像边界的区分度和模型对边界的敏感性。通过这种方式，GFEB 能够更加准确地识别和处理图像的边界信息，减少序列模型中因边界混淆带来的误差。GFEB 不仅有效提高了对图像边界的敏感性，同时减少了边界信息的模糊性和潜在混淆现象，显著提升了红外图像彩色化性能。GFEB 的数学公式如下

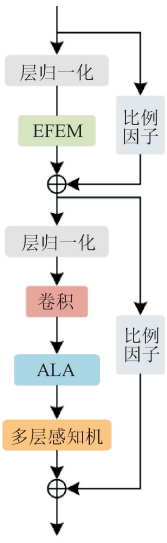
$$\begin{cases} f_{x_1} = f_\xi(f_\psi(f_\delta(f_{\mathrm{DW}}(f_\eta(x_{\mathrm{in}}))))) \\ f_{x_2} = f_\delta(f_\eta(x_{\mathrm{in}})) \end{cases} \tag{2}$$

式中：x_{in} 为输入特征图；f_{x_1} 和 f_{x_2} 为中间特征图；$f_\eta(\cdot)$ 为线性层；$f_{\mathrm{DW}}(\cdot)$ 为深度卷积；$f_\delta(\cdot)$ 为 SiLU 激活函数；$f_\psi(\cdot)$ 为扫描策略；$f_\xi(\cdot)$ 为层归一化。

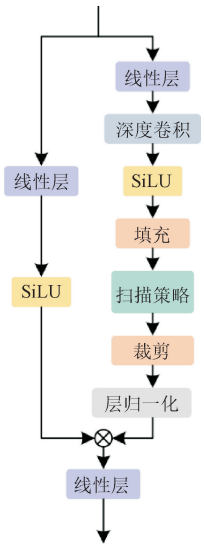
本文在 GFEB 中提出高效特征提取模块(EFEM)，以捕捉图像中的细节和边缘特征，保留了丰富的局部上下文信息，如图 3(b)所示。EFEM 使得模型在捕获边缘和细节方面表现优越，有效提升了红外图像彩色化中图像细节的保真度。在 EFEM 中引入自适应线性注意力(ALA)机制[21]保证 GFEB 在全局范围内建立像素间的关联，使得模型能够有效编码长程依赖关系。如图 3(c)所示，ALA 基于像素间的相关性计算，对图像的全局信息进行编码，实现了对全局上下文信息的精准获取，大幅提升了处理红外图像彩色化任务时的全局信息捕捉能力。此外，ALA 引入一组额外的标记特征，从键向量 $\boldsymbol{K}$ 和值向量 $\boldsymbol{V}$ 中聚合信息，同时在查询向量 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 之间建立间接联系，保留了全局上下文建模能力。由于标记特征的数量远小于 $\boldsymbol{Q}$ 的数量，因此实现了线性的计算复杂度。EFEM 的数学表达式如下

$$\begin{cases} f_{x_3} = f_\eta(f_{x_1} \odot f_{x_2}) + sf_{x_{\mathrm{in}}} \\ f_{x_4} = f_\varphi(f_{\bar{\omega}}(f_{\mathrm{Conv}}(f_\xi(f_{x_3})))) \\ f_{\mathrm{EFEM}}^{\mathrm{out}} = f_{x_4} + sf_{x_3} \end{cases} \tag{3}$$

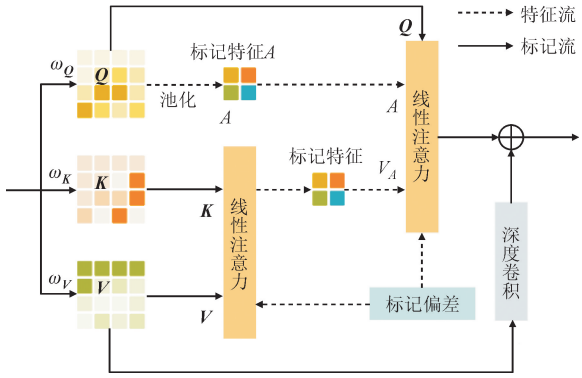
式中：$f_{x_{\mathrm{in}}}$ 为输入特征图；f_{x_3} 和 f_{x_4} 为中间特征图；$f_{\mathrm{EFEM}}^{\mathrm{out}}$ 为 EFEM 的输出；$\odot$ 为 Hadamard 乘积；$f_\varphi(\cdot)$ 为多层感知器；$f_{\bar{\omega}}(\cdot)$ 为 ALA 中的标记特征；s 为可学习比例因子。



(a)GFEB 的内部结构



(b)EFEM 的内部结构



(c)ALA 的内部结构

ω_Q、ω_K、ω_V— 向量 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 的权重；V_A— 标记特征 A 的权重。

图 3 GFEB 结构

Fig. 3 Structure of GFEB

1.4 IFM 设计

双分支编码器生成不同的特征映射，通过信息交互，形成富含全面信息的特征图。然而，双分支特征在语义层次和空间分布上存在差异，直接融合容易导致特征不对齐，影响彩色化结果的精确性。此外，在高语义相关性区域下，双分支特征存在信息冗余或冲突问题，也降低了网络对颜色和细节的恢复能力。

为了解决上述问题，本文提出了 IFM，其结构如图 4 所示。IFM 利用一个分支的特征映射作为另一个分支的互补信息源，通过互信息处理和减去补充信息的方法，确保输入和输出维度一致，使模型能够获取更为全面的信息，提升彩色化性能。IFM 接收来自双分支编码器的特征图，分别记为 $\boldsymbol{F}^i \in \mathbf{R}^{H\times W\times C}$(来自 GFEB)和 $\boldsymbol{G}^i \in \mathbf{R}^{H\times W\times C}$(来自 LFEB)，其中 H、W、C 分别为特征图的高度、宽度和通道数。为了实施特征交互融合，采用全局平均池化(GAP)操作，从每个特征图中提取全局信息。通过特征向量的拼接操作，实现两个分支之间的信息互补。将 $\boldsymbol{F}^i$ 和 $\boldsymbol{G}^i$ 的特征向量进行拼接，得到特征图 $\boldsymbol{G}_1^i \in \mathbf{R}^{(H\times W+1)\times C}$。同样地，将 $\boldsymbol{G}^i$ 的特征向量与 $\boldsymbol{F}^i$ 的特征图进行拼接，得到另一个特征图 $\boldsymbol{F}_1^i \in \mathbf{R}^{(H\times W+1)\times C}$。随后，将 $\boldsymbol{G}_1^i$ 和 $\boldsymbol{F}_1^i$ 通过卷积层进行通道间的信息交互，以实现特征的深度融合。在完成特征交互融合后，通过重塑操作去除辅助信息，并将图像恢复到原始尺寸，以便后续的重建操作。IFM 的表达式如下

$$\begin{cases} \boldsymbol{F}_1^i = f_{\mathrm{Cat}}[f_\eta(f_{\mathrm{GAP}}(\boldsymbol{G}^i)), \boldsymbol{F}^i] \\ \boldsymbol{G}_1^i = f_{\mathrm{Cat}}[f_\eta(f_{\mathrm{GAP}}(\boldsymbol{F}^i)), \boldsymbol{G}^i] \\ \boldsymbol{F}_2^i = f_{\mathrm{Conv}}(\boldsymbol{F}_1^i) \\ \boldsymbol{G}_2^i = f_{\mathrm{Conv}}(\boldsymbol{G}_1^i) \\ \boldsymbol{Z}_{\mathrm{out}}^i = f_{\mathrm{SE}}(f_{\mathrm{Conv}}(f_{\mathrm{Cat}}[\boldsymbol{F}_2^i, \boldsymbol{G}_2^i])) \end{cases} \tag{4}$$

式中：$\boldsymbol{F}_2^i$ 和 $\boldsymbol{G}_2^i$ 为中间特征图；$\boldsymbol{Z}_{\mathrm{out}}^i$ 为 IFM 的输出；

$f_{Cat}[\cdot]$ 为拼接操作；$f_{GAP}(\cdot)$ 为全局平均池化操作；$f_{SE}(\cdot)$ 为 SE 注意力机制。

值得注意的是，IFM 的设计不仅考虑到了特征融合的效率和准确性，还充分利用了卷积在处理图像特征时的强大能力。通过特征向量的拼接和卷积的通道间信息交互，IFM 成功地实现了双分支编码器之间信息的有效互补和融合，生成了包含更全面语义信息的特征表示。

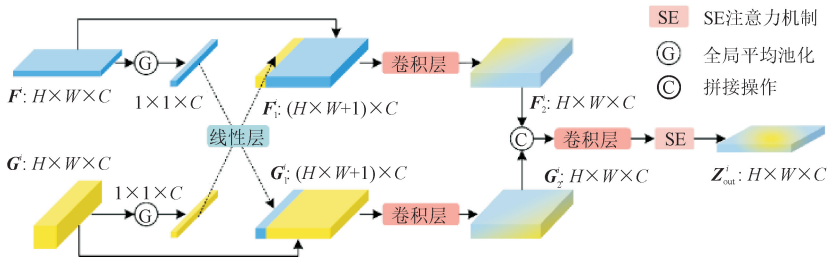
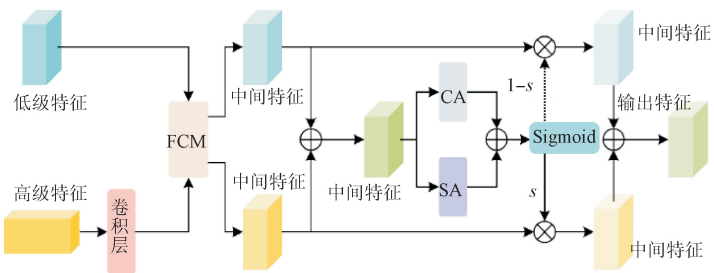


图 4　IFM 结构

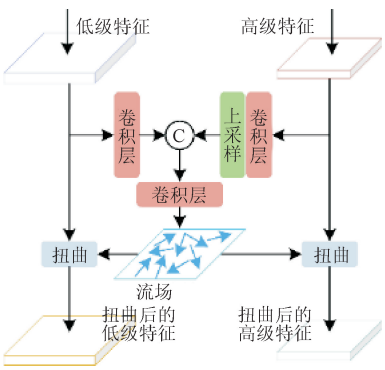
Fig. 4　Structure of IFM

1.5　CAM 设计

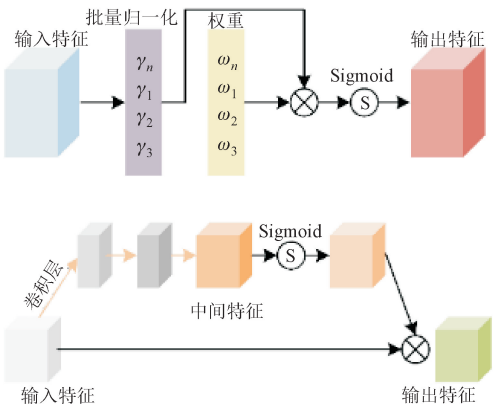
现有彩色化方法通常直接将解码层与相应尺度的编码层进行融合，但未能充分挖掘不同编码层的信息，且忽视了空间扭曲的问题，导致彩色化图像的边缘细节和信息不够完整，影响彩色化质量。因此，如何在解码层减少特征信息的损失，并维持空间一致性，是设计网络的关键问题。为了解决上述问题，本文设计了 CAM 实现多层级特征的有效整合，如图 5 所示。



(a)CAM 的内部结构



(b)特征矫正模块(FCM)的内部结构



(c)通道注意力(CA)机制和空间注意力(SA)机制的内部结构

图 5　CAM 结构

Fig. 5　Structure of CAM

在红外图像彩色化任务中，低层特征具有丰富的空间细节，但语义信息较少；高层特征包含大量语义信息，但空间细节不足。CAM 通过将低层特征的空间细节与高层特征的语义信息有效结合，实现了特征间的互补融合，在保持空间信息完整性的同时增强了特征的语义表达能力。此外，在 CAM 中

设计了特征矫正模块(FCM),通过对特征图进行尺度调整和精细校正,确保不同层级特征在融合过程中的空间一致性。FCM 对特征进行处理以提高融合的协调性,优化空间细节的表达与重建,其数学表达式如下

$$\begin{cases}(\Delta x,\Delta y)=f_{\mathrm{Conv}}(f_{\mathrm{Cat}}[f_{\mathrm{Conv}}(x_{\mathrm{l}}),f_{\mathrm{UP}}(f_{\mathrm{Conv}}(x_{\mathrm{h}}))])\\ h'=\min(\max(0,h+\Delta y),h)\\ w'=\min(\max(0,w+\Delta x),w)\\ f_{\mathrm{FCM}}^{\mathrm{out}}(h,w)=F^{\mathrm{c}}(h',w')\end{cases} \tag{5}$$

式中:x_{l} 为低级特征;x_{h} 为高级特征;(h,w) 和 (h',w') 分别为输入特征图、扭曲后特征图的高和宽;$f_{\mathrm{UP}}(\cdot)$ 为上采样操作;$F^{\mathrm{c}}(h',w')$ 为原始特征在 (h',w') 处的特征值;Δx、Δy 为偏移量;min、max 分别为取最小值和最大值。

CAM 设计了通道注意力(CA)和空间注意力(SA),进一步提高特征融合的精确性。CA 通过计算每个特征通道的权重,使模型能有效突出关键信息区域,强化含有重要上下文信息的通道,减少冗余信息干扰。SA 通过捕捉特征图中的空间依赖关系,帮助模型在特征空间上关注关键区域,提升彩色化图像质量。通过 CAM 的多层级、多维度的特征融合和校准策略,模型在红外图像彩色化任务中可以实现边界清晰、语义丰富的特征提取和增强,在复杂城市场景中能够有效捕捉细节,确保彩色化结果的清晰性和准确性。CAM 的数学表达式如下

$$\begin{cases}(x_{\mathrm{l1}},x_{\mathrm{h1}})=f_{\mathrm{FCM}}^{\mathrm{out}}(x_{\mathrm{l}},f_{\mathrm{Conv}}(x_{\mathrm{h}}))\\ x_{\mathrm{f}}=x_{\mathrm{l1}}+x_{\mathrm{h1}}\\ x_{\mathrm{f1}}=f_{\mathrm{CA}}(x_{\mathrm{f}})+f_{\mathrm{SA}}(x_{\mathrm{f}})\\ s=f_{\sigma}(x_{\mathrm{f1}})\\ f_{\mathrm{CAM}}^{\mathrm{out}}=x_{\mathrm{l1}}(1-s)+x_{\mathrm{h1}}s\end{cases} \tag{6}$$

式中:$(x_{\mathrm{l}},x_{\mathrm{h}})$、$(x_{\mathrm{l1}},x_{\mathrm{h1}})$ 分别为输入和中间输出的低级特征、高级特征;$f_{\mathrm{CAM}}^{\mathrm{out}}$ 为 CAM 的输出特征;$f_{\sigma}(\cdot)$ 为 Sigmoid 函数;$f_{\mathrm{CA}}(\cdot)$ 为通道注意力函数;$f_{\mathrm{SA}}(\cdot)$ 为空间注意力函数。

1.6 损失函数

为了提高生成彩色化图像与可见光图像之间的相似度,本文算法采用 4 种损失函数来优化生成器的性能。首先,引入对抗损失 $\mathcal{L}_{\mathrm{adv}}$[6,8],使生成图像更加真实细腻,增强其细节表现。其次,采用像素损失 $\mathcal{L}_{\mathrm{pixel}}$[9,22],以保证生成的彩色化图像在色度和亮度上与可见光图像的差异最小,并更好地匹配真实色彩。接着,为了减少主观感知上的差异,采用感知损失 $\mathcal{L}_{\mathrm{per}}$[2,14] 通过高层特征约束来优化生成图像与可见光图像的视觉一致性。最后,引入结构相似性损失 $\mathcal{L}_{\mathrm{ssim}}$[2,9],进一步减小生成图像与可见光图像在结构上的偏差,定量衡量并提升两者的整体相似性。将上述 4 种损失函数相组合,有助于生成具有更高视觉保真度和准确性的彩色化图像。4 种损失函数的数学表达式如下

$$\begin{cases}\mathcal{L}_{\mathrm{adv}}=E_{\boldsymbol{x}\sim P_{\mathrm{data}(\boldsymbol{x})}}[\log(1-D(\hat{\boldsymbol{y}},\boldsymbol{x}))]\\ \mathcal{L}_{\mathrm{pixel}}=\dfrac{1}{HW}\displaystyle\sum_{i=1}^{H}\sum_{j=1}^{W}\|\boldsymbol{y}_{i,j}-\hat{\boldsymbol{y}}_{i,j}\|_1\\ \mathcal{L}_{\mathrm{per}}=\dfrac{1}{HW}\displaystyle\sum_{i=1}^{H}\sum_{j=1}^{W}\|\boldsymbol{\varphi}_{n,i,j}(\boldsymbol{y})-\boldsymbol{\varphi}_{n,i,j}(\hat{\boldsymbol{y}})\|_1\\ \mathcal{L}_{\mathrm{ssim}}=1-\dfrac{(2\mu_x\mu_y+C_1)+(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)+(\sigma_x^2\sigma_y^2+C_2)}\end{cases} \tag{7}$$

式中:$\boldsymbol{x}$ 为输入的红外图像;$\boldsymbol{y}$、$\hat{\boldsymbol{y}}$ 分别为可见光图像和生成的彩色化图像;$P_{\mathrm{data}(\boldsymbol{x})}$ 为可见光数据集的数据分布;$D(\cdot)$ 为鉴别器;$\boldsymbol{\varphi}_n(\cdot)$ 为 VGG-16 网络中第 n 层特征图;μ、σ^2 分别为均值和方差;E 为期望;σ_{xy} 为图像的协方差;C_1 和 C_2 为防止分母为 0 的常数。

本文算法的损失函数定义为

$$\mathcal{L}_{\mathrm{total}}=\lambda_{\mathrm{adv}}\mathcal{L}_{\mathrm{adv}}+\mathcal{L}_{\mathrm{pixel}}+\mathcal{L}_{\mathrm{per}}+\lambda_{\mathrm{ssim}}\mathcal{L}_{\mathrm{ssim}} \tag{8}$$

式中:λ_{adv} 为对抗损失的权重;λ_{ssim} 为结构相似性损失的权重。

结合现有研究[2,9,22]以及实验分析,本文对损失函数中不同损失项的权重进行了调整。在实验过程中,对权重 λ_{adv} 和 λ_{ssim} 进行了多组测试,以平衡对抗损失与其他损失项的量级,并确保彩色化结果的质量和稳定性。最终,结合实验结果和现有先进方法[2,9,22]的权重设置,确定 $\lambda_{\mathrm{adv}}=0.03$ 和 $\lambda_{\mathrm{ssim}}=1.25$ 为最优权重配置,以实现最佳的彩色化效果。

2 实验结果与分析

2.1 数据集及实验设置

KAIST 数据集[23]包含 95 328 对可见光和红外图像,图像尺寸为 640×512 像素。该数据集覆盖了日间和夜间多个时间段,主要采集于校园、街道和乡村等常规交通场景,内容丰富且多样化。FLIR 数据集[24]包含 9 711 张红外图像和 9 233 张可见光图像,包括行人、汽车和建筑物等。由于计算资源的限制,本文将彩色化任务生成的所有图像的分辨率限制为 256×256 像素。

所有实验均在 Ubuntu 22.04.1 LTS 系统上运行,使用 NVIDIA Tesla V100 GPU 16 GB 加速算法,并基于深度学习框架 Torch 2.3.1 和 CUDA

11.8 实现模型训练。采用 Adam 优化器，设置参数 $\beta_1=0.5$，$\beta_2=0.999$。在训练过程中，采用 Lambda 学习率衰减策略，初始学习率为 2×10^{-4}，训练总周期为 200 轮。在训练周期过半后，学习率线性衰减至 0。生成器和鉴别器的第一卷积层通道数均设置为 32，批次设置为 4。

本文算法的训练过程如下。首先设定红外图像集 X 和可见光图像集 Y，并定义生成器的迭代次数 n_{iter}、批量大小 m 以及训练轮数 n_{epoch}。优化目标包括生成器 G 和鉴别器 D，其中生成器参数 θ_G 和鉴别器参数 θ_D 进行随机初始化，同时定义生成器损失 $\mathcal{L}_{\text{gen}}$ 和鉴别器损失 $\mathcal{L}_{\text{dis}}$。

在每个训练轮次 i（共 n_{epoch} 轮）中，执行以下步骤。在每次迭代 j（共 n_{iter} 次）中，从红外图像集 X 中采样 m 个样本 x_1，x_2，…，x_m，并从可见光图像集 Y 中采样 m 个对应样本 y_1，y_2，…，y_m。然后，利用生成器 G 生成相应的伪彩色图像 $G(x_1)$，$G(x_2)$，…，$G(x_m)$。基于这些生成数据计算鉴别器损失 $\mathcal{L}_{\text{dis}}$，并更新鉴别器参数 θ_D。在完成所有迭代后，计算生成器损失 $\mathcal{L}_{\text{gen}}$ 并更新生成器参数 θ_G。重复上述过程直至完成所有训练轮次。

2.2　与现有先进算法的对比实验

为了验证本文算法的优越性和先进性，将其彩色化结果与 Pix2pix[6]、TICC-GAN[8]、E2GAN[24]、UVCGAN[13]、LKAT-GAN[14]、VM-UNet[25] 和 ColorMamba[20] 这 7 种先进算法进行比较。其中：Pix2pix、TICC-GAN、E2GAN 是典型的 CNN 方法；UVCGAN 采用 ViT 架构，能够评估自注意力机制对彩色化任务的影响；LKAT-GAN 结合 CNN 和 ViT，代表了混合架构的优势；VM-UNet 和 ColorMamba 采用 Mamba 结构，能够验证 Mamba 在序列建模能力上的有效性。通过对比，能够全面分析不同网络架构在红外图像彩色化任务中的适用性，为本文算法的优势与不足提供系统性的参考。

为了确保公平性，采用各算法的源代码和超参数对模型重新进行训练。所有对比方法均在相同的硬件与软件平台上，采用相同的 GPU、深度学习框架版本和优化器配置完成，确保实验条件一致。为了定量评估本文算法的有效性，采用多项指标对彩色化结果进行质量评估[26]，包括峰值信噪比 P_{SNR}、结构相似性 S_{SIM}、Fréchet 初始距离 F_{FID} 和学习感知图像块相似度 L_{LPI}，这些指标在计算机视觉领域被广泛使用，用于综合衡量彩色化质量，计算式分别如下

$$\begin{cases} P_{\text{SNR}} = 10\lg\left(\dfrac{M_{\text{I}}^2}{N}\right) \\ S_{\text{SIM}} = \dfrac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \\ L_{\text{LPI}} = \displaystyle\sum_L \frac{1}{H_L W_L}\sum_{h=1}^{H_L}\sum_{w=1}^{W_L} \| \boldsymbol{\omega}_L \odot (\hat{\boldsymbol{y}}_{hw,L} - \boldsymbol{y}_{hw,L}) \|_2^2 \\ F_{\text{FID}} = \| \boldsymbol{\zeta}_y - \boldsymbol{\zeta}_{\hat{y}} \|_2^2 + \text{tr}(\boldsymbol{\Sigma}_y + \boldsymbol{\Sigma}_{\hat{y}} - 2(\boldsymbol{\Sigma}_y\boldsymbol{\Sigma}_{\hat{y}})^{1/2}) \end{cases} \tag{9}$$

式中：M_{I} 为图像的最大像素；N 为均方误差；$\boldsymbol{\omega}_L$ 为加权的权重向量；$\boldsymbol{y}_{hw,L}$ 表示图像在网络的第 L 层（通常是深度特征图）；$\boldsymbol{\zeta}$ 为图像在特征空间中的均值向量；$\boldsymbol{\Sigma}$ 为斜方差矩阵；tr 为矩阵的迹。

2.2.1　对比实验的定量分析

表 1 展示了不同方法在 KAIST 数据集上的性能对比。其中，“↑”表示指标越高越好，“↓”表示指标越低越好。可以看出，本文算法在各项评估指标上均优于其他方法，其中 P_{SNR} 和 S_{SIM} 的性能表现最佳。此外，F_{FID} 和 L_{LPI} 性能表现也较好。TICC-GAN 的 F_{FID} 结果最优。然而，进一步的定性分析表明，TICC-GAN 生成的图像在彩色化方面存在严重失真，这表明 F_{FID} 主要是在自然图像上进行预训练，由于红外图像彩色化任务的目标域与自然图像存在一定差异，导致 F_{FID} 对该任务的敏感性降低，影响对模型性能的准确衡量。本文算法在细节和边缘保持方面具有优势，保留了部分红外图像的纹理模式和细节信息，与自然图像存在部分差异，因此导致 F_{FID} 较高。相较之下，本文算法在各项指标上均表现出了优秀的彩色化性能。综上，本文算法中的 LFEB 和 GFEB 在局部细节和全局一致性方面表现优异，能够生成边缘清晰、细节丰富的彩色化图像。

表 1　不同方法在 KAIST 数据集上的性能对比

Table 1　Performance comparison of different methods on the KAIST dataset

算法	P_{SNR}/dB ↑	S_{SIM} ↑	F_{FID} ↓	L_{LPI} ↓
Pix2pix	28.412	0.427	81.476	0.274 0
TICC-GAN	28.456	0.475	**63.386**	0.267 2
E2GAN	28.501	0.436	71.677	0.272 1
UVCGAN	28.503	0.413	73.590	0.278 5
LKAT-GAN	28.419	0.458	82.183	**0.251 9**
VM-UNet	28.503	0.494	95.854	0.274 6
ColorMamba	28.594	0.486	82.475	0.255 9
本文	**28.645**	**0.507**	73.005	0.253 4

注：黑体数据为最优值。

表 2 展示了不同方法在 FLIR 数据集上的性能对比。可以看出，本文算法在大部分指标上均优于对比方法，其中 S_{SIM}、F_{FID} 和 L_{LPI} 的性能尤佳。结果表明，本文算法在与可见光的相似性和整体图像质量方面具有显著优势，生成的彩色图像在细节表现和全局一致性方面得到了显著优化。此外，本文算法的 P_{SNR} 性能也较好，进一步验证了其在提升视觉质量方面的有效性。这些量化结果充分表明，本文算法能够有效聚合多尺度上下文信息，增强特征表达的完整性，在 KAIST 数据集和 FLIR 数据集的复杂城市场景中实现了出色的彩色化性能。

表 2 不同方法在 FLIR 数据集上的性能对比

Table 2 Performance comparison of different methods on the FLIR dataset

算法	P_{SNR}/dB ↑	S_{SIM} ↑	F_{FID} ↓	L_{LPI} ↓
Pix2pix	29.857	0.623	61.843	0.134 7
TICC-GAN	29.992	0.648	53.508	0.115 7
E2GAN	30.105	0.658	47.782	0.109 3
UVCGAN	28.867	0.569	56.395	0.159 8
LKAT-GAN	**30.463**	**0.725**	51.762	0.087 5
VM-UNet	29.662	0.687	53.417	0.112 2
ColorMamba	30.148	0.719	47.932	0.102 5
本文	30.459	**0.725**	**46.791**	**0.085 8**

注：黑体数据为最优值。

表 3 展示了不同方法在计算复杂度方面的对比结果。可以看出，本文算法与 E2GAN 算法相比，参数量减少了 10.71×10^6，与 Pix2pix 算法相比，GPU 内存占用降低了 0.62×10^9。这些数据表明，本文算法在有效提升彩色化性能的同时，大幅度减少了计算资源的消耗。这得益于双分支结构对特征提取的高效分工，以及上下文聚合模块在解码阶段进一步优化了多尺度特征的利用效率，有效降低了模型的冗余计算。

表 3 不同方法在计算复杂度方面的性能比较

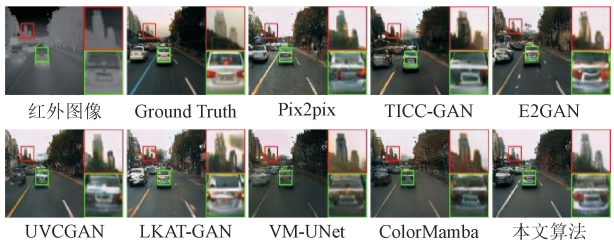
Table 3 Performance comparison of different methods in computational complexity

算法	机制	参数数量/10^6 ↓	GPU 内存/10^9 ↓	推理速度/s ↓
Pix2pix	CNN	54.42	3.57	0.28
TICC-GAN	CNN	45.85	3.70	0.26
E2GAN	Conv	27.41	4.40	0.35
UVCGAN	CNN	67.59	11.47	0.46
LKAT-GAN	CNN+ViT	74.19	4.81	0.39
VM-UNet	Mamba	34.62	4.86	0.17
ColorMamba	Mamba	89.08	4.97	0.21
本文	CNN+Mamba	**16.70**	**2.95**	**0.16**

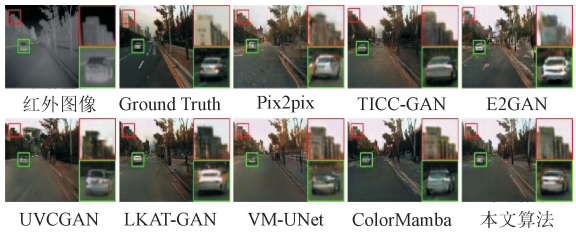
注：黑体数据为最优值。

2.2.2 对比实验定性分析

图 6 和图 7 展示了本文算法与其他对比算法在 KAIST 和 FLIR 数据集上的彩色化效果对比。



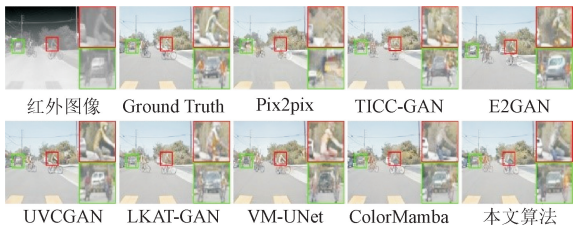
(a)场景 1



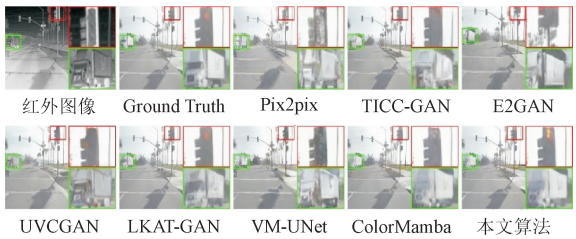
(b)场景 2

图 6 不同方法在 KAIST 数据集上的彩色化结果对比

Fig. 6 Comparison of colorization results by different methods on the KAIST dataset



(a)场景 1



(b)场景 2

图 7 不同方法在 FLIR 数据集上的彩色化结果对比

Fig. 7 Comparison of colorization results by different methods on the FLIR dataset

由图 6 和图 7 可见，Pix2pix 由于采用单一的卷积结构，无法充分捕捉图像的复杂特征，导致生成的彩色化结果不准确。TICC-GAN 和 E2GAN 在 Pix2pix 基础上增加了网络的深度和宽度，提升了彩色化效果，但也显著增加了计算复杂度和模型参数

量。UVCGAN 完全依赖于 Transformer 结构捕获图像的全局信息，缺少细致的局部信息提取，导致生成的彩色化图像在边缘细节（如车辆边缘）上表现不足。LKAT-GAN 结合了卷积和 Transformer 结构，通过长程依赖和多尺度信息捕捉，生成了较为准确的彩色化结果，但在语义信息的细致处理方面仍存在不足。VM-UNet 和 ColorMamba 虽然采用 Mamba 结构降低了计算复杂度，但在生成小目标的边缘细节方面仍有欠缺，特别是对精细纹理的还原不够准确。相比之下，本文算法采用了 LFEB 和 GFEB 捕获局部信息和全局信息。两者通过 IFM 实现信息的有效互补和深度融合，提升了彩色化图像在细节和整体一致性上的表现。同时，通过 CAM 进一步增强了多尺度上下文语义特征的聚合能力，使生成的彩色化结果在边界清晰度和细节表现力方面显著优于其他方法。本文算法在建筑物、车辆等细节的复原上表现尤为出色，呈现出更为真实的色彩分布和纹理细节。

2.3　消融实验

为了全面评估本文算法中各组成部分的有效性，在 FLIR 数据集上从编码器、融合模块和解码器 3 个方面开展消融实验。所有实验均在相同的数据集和实验环境下进行，以确保结果的公平性。

2.3.1　编码器分析

为了全面验证双分支编码器的有效性，本文设计了消融实验，将双分支编码器分别替换为 UNet 结构[4]（记为 Baseline）或单一分支进行性能对比。如表 4 所示，当编码器被替换为 UNet 结构时，性能显著下降，表明在局部细节和全局一致性方面存在局限性。若仅保留 LFEB 或 GFEB，虽然能在局部或全局特征提取上有所表现，但彩色化性能仍低于双分支编码器，这充分表明了两分支协同作用的重要性。

表 4　不同编码器的性能评估

Table 4　Performance evaluation of different encoders

算法	P_{SNR}/dB ↑	S_{SIM} ↑	F_{FID} ↓	L_{LPI} ↓
Baseline	29.036	0.663	77.860	0.164 9
LFEB	29.089	0.662	85.358	0.156 6
GFEB	29.377	0.688	68.027	0.144 2
本文	**30.459**	**0.725**	**46.791**	**0.085 8**

注：黑体数据为最优值。

图 8 给出了不同编码器在 FLIR 数据集上的彩色化结果对比，可以看到，双分支编码器生成的彩色化图像在边缘清晰度和纹理表现上更加自然且连贯，能够精确还原车辆的整体外观及细节；而单分支方法生成的图像中，车辆外观出现破散现象，细节缺失，边缘模糊，进一步暴露了单一分支在特征提取上的局限性。这些结果表明，双分支编码器通过 LFEB 和 GFEB 的协同作用，在局部细节复原和全局一致性增强上具有显著优势，能够生成质量更高的彩色化图像。

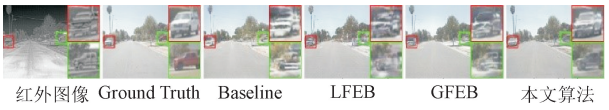


图 8　不同编码器在 FLIR 数据集上的彩色化结果对比

Fig. 8　Comparison of colorization results by different encoders on the FLIR dataset

2.3.2　融合模块分析

为了验证所提出 IFM 的有效性，本文设计消融实验开展性能分析，与常用的融合策略 Concat 和 Add 进行对比。表 5 给出了不同融合模块的性能评估结果。由表可知，相较于 Concat 和 Add，本文所提 IFM 呈现出显著的性能优势，其中 P_{SNR} 和 S_{SIM} 分别提升了 0.673 dB、0.019，F_{FID} 和 L_{LPI} 分别降低了 3.133、0.026 5。定量结果表明，本文提出的 IFM 通过更加深入的特征信息交互，有效优化了特征信息融合的质量。

表 5　不同融合模块的性能评估

Table 5　Performance evaluation of different fusion modules

算法	P_{SNR}/dB ↑	S_{SIM} ↑	F_{FID} ↓	L_{LPI} ↓
Add	29.205	0.678	92.210	0.1846
Concat	29.786	0.706	49.924	0.1123
本文	**30.459**	**0.725**	**46.791**	**0.085 8**

注：黑体数据为最优值。

图 9 展示了不同融合模块在 FLIR 数据集上的彩色化结果对比。进一步观察到，简单的 Add 策略生成的图像细节严重缺失，无法完整还原目标外观；而 Concat 尽管保留了一定信息，但由于缺乏有效的信息交互机制，生成的图像在细节还原上仍显不足。相比之下，本文提出的 IFM 通过融合过程中全局平均池化、互信息处理以及卷积层的深度交互，克服了简单连接或相加策略中出现的特征冗余与信息丢失问题，显著提升了彩色化任务的图像质量和细节还原能力。

图 9 不同融合模块在 FLIR 数据集上的彩色化结果对比

Fig. 9 Comparison of colorization results by different fusion modules on the FLIR dataset

2.3.3 解码器分析

尽管采用了双分支编码器，但由于 IFM 的存在，本文算法仅包含由 CAM 组成的解码器。为了评估 CAM 在解码器中的关键作用，将 CAM 替换为 AGM[27] 和 AFM[28]，保持其余实验条件一致。如表 6 所示，使用 CAM 后，解码器的性能在各项指标上均优于 AGM 和 AFM，展示了最佳的彩色化效果。

表 6 不同解码器的性能评估

Table 6 Performance evaluation of different decoders

算法	P_{SNR}/dB ↑	S_{SIM} ↑	F_{FID} ↓	L_{LPI} ↓
AGM	29.512	0.683	62.945	0.126 3
AFM	29.358	0.695	61.823	0.143 3
本文	**30.459**	**0.725**	**46.791**	**0.085 8**

注：黑体数据为最优值。

图 10 给出了不同解码器在 FLIR 数据集上的彩色化结果对比，可以看出，CAM 在解码过程中通过多层次特征的深度融合和精准校准，显著增强了模型的语义信息捕获和细节表达能力。这种设计不仅提升了彩色化图像的边缘清晰度，还保证了全局颜色一致性，显著改善了图像的整体清晰度和准确性。

图 10 不同解码器在 FLIR 数据集上的彩色化结果对比

Fig. 10 Comparison of colorization results by different decoders on the FLIR dataset

3 结 论

围绕红外图像彩色化中的细粒度特征提取与长程依赖关系建模问题，开展了系统的理论分析与实验研究。针对现有彩色化方法的不足，分析了其在局部细节保留和全局一致性建模方面的局限性，提出了一种双分支特征交互融合的高效红外图像彩色化方法，充分结合 CNN 与 Mamba 的优势，提升了彩色化图像的质量，得到如下主要结论。

(1)提出了 LFEB 和 GFEB 相结合的双分支编码器结构，其中 LFEB 通过卷积操作提取局部信息，确保细粒度特征的精准获取，GFEB 基于 Mamba 建模长程依赖关系，实现更全面的特征表达。

(2)提出了 IFM 模块，使双分支编码器提取特征能够互相补充，强化不同特征层之间的信息交互，提高了局部细节和全局一致性的综合建模能力。

(3)设计 CAM 模块，增强了多尺度语义特征的聚合能力，有效优化了彩色化结果的边缘清晰度和细节表现，提高了红外图像的视觉可辨识性。

(4)在 KAIST 和 FLIR 数据集上开展实验，结果表明，本文算法在定量评估和定性分析方面均优于现有先进算法，得到的峰值信噪比分别达到 28.645、30.459 dB，结构相似性分别达到 0.507、0.725，能够更精准地恢复红外图像中的真实场景与细节。

尽管本文算法在计算资源利用效率上已有优化，但在嵌入式和移动设备上的高效部署策略仍需进一步研究，以提升实时性和计算适应性。

参考文献：

[1] 林再平，罗伊杭，李博扬，等. 基于梯度可感知通道注意力模块的红外小目标检测前去噪网络 [J]. 红外与毫米波学报，2024，43(2)：254-260.
LIN Zaiping，LUO Yihang，LI Boyang，et al. Gradient-aware channel attention network for infrared small target image denoising before detection [J]. Journal of Infrared and Millimeter Waves，2024，43(2)：254-260.

[2] CHEN Yu，ZHAN Weida，JIANG Yichun，et al. A feature refinement and adaptive generative adversarial network for thermal infrared image colorization [J]. Neural Networks，2024，173：106184.

[3] CHEN Yu，ZHAN Weida，JIANG Yichun，et al. Contrastive learning with feature fusion for unpaired thermal infrared image colorization [J]. Optics and Lasers in Engineering，2023，170：107745.

[4] RONNEBERGER O，FISCHER P，BROX T. U-Net：convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham，Germany：Springer International Publishing，2015：234-241.

[5] ASWATHA S M，MALLADI S P K，MUKHERJEE J. An encoder-decoder based deep architecture for visible to near infrared image transformation [C]//Proceedings of the Twelfth Indian Conference on Computer Vision，Graphics and Image Processing. New York，

USA：ACM，2021：29.

[6] ISOLA P，ZHU Junyan，ZHOU Tinghui，et al. Image-to-image translation with conditional adversarial networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway，NJ，USA：IEEE，2017：5967-5976.

[7] GOODFELLOW I J，POUGET-ABADIE J，MIRZA M，et al. Generative adversarial nets [C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge，MA，USA：MIT Press，2014：2672-2680.

[8] KUANG Xiaodong，ZHU Jianfei，SUI Xiubao，et al. Thermal infrared colorization via conditional generative adversarial network [J]. Infrared Physics & Technology，2020，107：103338.

[9] LIAO Hangying，JIANG Qian，JIN Xin，et al. MU-GAN：thermal infrared image colorization using mixed-skipping UNet and generative adversarial network [J]. IEEE Transactions on Intelligent Vehicles，2023，8(4)：2954-2969.

[10] 朴燕，康继元. RAUGAN：基于循环生成对抗网络的红外图像彩色化方法 [J/OL]. 吉林大学学报(工学版)．(2024-04-28)[2024-06-20]. https://doi.org/10.13229/j.cnki.jdxbgxb.20240012.
PIAO Yan，KANG Jiyuan. RAUGAN：infrared image colorization method based on cycle generative adversarial networks [J/OL]. Journal of Jilin University(Engineering and Technology Edition)．(2024-04-28)[2024-06-20]. https://doi.org/10.13229/j.cnki.jdxbgxb.20240012.

[11] 赵伟强，王洋，祝勇，等. 基于颜色预测和语义感知双分支结构的近红外图像彩色化算法 [J]. 光学技术，2023，49(3)：371-378.
ZHAO Weiqiang，WANG Yang，ZHU Yong，et al. Colorization algorithm for near infrared images based on dual branching structure of color prediction and semantic perception [J]. Optical Technique，2023，49(3)：371-378.

[12] DOSOVITSKIY A，BEYER L，KOLESNIKOV A，et al. An image is worth 16×16 words：transformers for image recognition at scale [C]//Proceedings of the International Conference on Learning Representations. New York，USA：ICLR，2021：1-22.

[13] TORBUNOV D，HUANG Yi，YU Haiwang，et al. UVCGAN：UNet vision transformer cycle-consistent GAN for unpaired image-to-image translation [C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway，NJ，USA：IEEE，2023：702-712.

[14] HE Youwei，JIN Xin，JIANG Qian，et al. LKAT-GAN：a GAN for thermal infrared image colorization based on large kernel and attentionUNet-transformer [J]. IEEE Transactions on Consumer Electronics，2023，69(3)：478-489.

[15] GU A，DAO T. Mamba：linear-time sequence modeling with selective state spaces [EB/OL]．(2024-05-31)[2024-05-31]. https://arxiv.org/abs/2312.00752.

[16] LIU Yue，TIAN Yunjie，ZHAO Yuzhong，et al. Vmamba：visual state space model [C]//Proceedings of the Advances in Neural Information Processing Systems. San Diego：NIPS Press，2024：103031-103063.

[17] CHEN Keyan，CHEN Bowen，LIU Chenyang，et al. RSMamba：remote sensing image classification with state space model [J]. IEEE Geoscience and Remote Sensing Letters，2024，21：1-5.

[18] GUO Hang，LI Jinmin，DAI Tao，et al. MambaIR：a simple baseline for image restoration with state-space model [C]//Computer Vision-ECCV 2024. Cham：Springer Nature Switzerland，2025：222-241.

[19] SUN Shangquan，REN Wenqi，ZHOU Juxiang，et al. A hybrid transformer-mamba network for single image deraining [EB/OL]．(2024-08-31)[2024-08-31]. https://arxiv.org/abs/2409.00410.

[20] ZHAI Huiyu，JINGuang，YANG Xingxing，et al. ColorMamba：towards high-quality NIR-to-RGB spectral translation with mamba [C]//Proceedings of the 16th Asian Conference on Machine Learning. Chia Laguna Resort，Sardinia，Italy：PMLR，2025：765-780.

[21] HAN Dongchen，YE Tianzhu，HAN Yizeng，et al. Agent attention：on the integration of softmax and linear attention [C]//Computer Vision-ECCV 2024. Cham：Springer Nature Switzerland，2025：124-140.

[22] CHEN Yu，ZHAN Weida，JIANG Yichun，et al. DDGAN：dense residual module and dual-stream attention-guided generative adversarial network for colorizing near-infrared images [J]. Infrared Physics & Technology，2023，133：104822.

[23] HWANG S，PARK J，KIM N，et al. Multispectral pedestrian detection：benchmark dataset and baseline [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway，NJ，USA：IEEE，2015：1037-1045.

[24] CHEN Yu，ZHAN Weida，JIANG Yichun，et al. Exploring efficient and effective generative adversarial network for thermal infrared image colorization [J]. Complex & Intelligent Systems，2023，9(6)：7015-7036.

[25] RUAN Jiacheng, LI Jincheng, XIANG Suncheng. VM-UNet: vision mamba uNet for medical image segmentation [EB/OL]. (2024-11-08)[2024-11-08]. https://arxiv.org/abs/2402.02491.

[26] 吴洪伍，盖绍彦，达飞鹏. 采用感受野优化与渐进特征融合的图像超分辨率算法 [J]. 西安交通大学学报，2025，59(1)：136-147.
WU Hongwu, GE Shaoyan, DA Feipeng. Image super resolution algorithm based on receptive field optimization and progressive feature fusion [J]. Journal of Xi'an Jiaotong University, 2025, 59(1): 136-147.

[27] YADAV N K, SINGH S K, DUBEY S R. MobileAR-GAN: MobileNet-based efficient attentive recurrent generative adversarial network for infrared-to-visual transformations [J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-9.

[28] HUANG Junjian, REN Hao, LIU Shulin, et al. Real-time attentive dilated U-Net for extremely dark image enhancement [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2024, 20(8): 231.

（编辑 李慧敏）