

基于深度学习的可见光与红外图像融合方法研究

蒋庆泽^a, 胡晓阳^b

(沈阳理工大学 a. 自动化与电气工程学院, b. 装备工程学院, 沈阳 110159)

摘要: 为提高可见光与红外图像融合的性能, 以 SwinFusion 模型作为主干网络进行改进研究。首先, 将一种表示学习引导的红外与可见光图像融合网络(LRRNet)中基于低秩和稀疏表示的模块引入模型的浅层与深层特征提取之间, 以增强多尺度特征的表达能力; 其次, 将 SwinFuse 中基于 L1 范数的融合策略引入主干网络的特征融合部分, 以优化多模态信息的整合。仿真实验结果表明改进后的模型能准确捕捉各自模态的特征分布, 验证了所提方法的有效性。

关键词: 可见光与红外图像融合; 深度学习; 特征融合

中图分类号: TP391.4

文献标志码: A DOI:10.3969/j.issn.1003-1251.2026.04.005

Research on Visible-Infrared Image Fusion Method Based on Deep Learning

JIANG Qingze, HU Xiaoyang

(Shenyang Ligong University, Shenyang 110159, China)

Abstract: To improve the performance of infrared and visible image fusion, this study conducts an enhancement of the SwinFusion model as the backbone network. Firstly, a module based on low-rank and sparse representation from the novel representation learning guided fusion network (LRRNet) is introduced between the shallow and deep feature extraction stages, aiming to enhance the expression capability of multi-scale features. Secondly, the L1-norm-based fusion strategy from SwinFuse is incorporated into the feature fusion part of the backbone network to optimize the integration of multi-modal information. Simulation results demonstrate that the improved model can accurately capture the feature distributions of each modality, validating the effectiveness of the proposed method.

Key words: visible-infrared fusion; deep learning; feature fusion

随着深度学习技术的飞速发展, 可见光与红外图像融合技术已成为实现多模态信息互补、提升图像质量的重要研究方向。现有方法大多基于卷积神经网络(CNN)以及残差网络(ResNet)等, 并结合注意力机制实现对图像细节和全局结构的有效提取。

文献[1]提出了 RFN-Nest 端到端残差融合网络, 通过两阶段训练策略和细节保持损失, 大幅提升了融合图像的细节表现能力。文献[2]提出了

基于改进小波变换的方法对红外与可见光图像进行融合, 小波变换的整体模型对红外与可见光双传感器获取的目标高频部分进行局部能量准则融合, 对低频部分进行系数绝对值最大规则法融合, 能够对多种传感器的图像进行信息互补处理, 且能分析出多种复杂环境条件对融合结果的影响。文献[3]提出了一种生成对抗网络(GAN)的红外与可见光图像融合方法, 利用语义分割网络将红外和可见光图像分割为目标图像和背景图像, 使

融合后的目标图像更突出,背景信息更丰富。文献[4]提出了一种结合非下采样剪切波变换(NSST)、ResNet与N-RGAN模型。通过分解源图像为高频和低频子带,分别进行特征融合与自适应加权优化,解决了传统方法中复杂场景适用性低、细节丢失及对比度不足的问题。文献[5]提出了PEFusion,利用多尺度曝光校正模块,优化曝光差异下的图像质量,增强了显著目标表达。

文献[6]基于Transformer架构提出了STF-Net自监督Transformer融合框架,进一步增强了模态间上下文关系建模能力。为提升模型可解释性,文献[7]通过可解释性分析方法DDFImage,对融合模型的内部机制进行了深入挖掘,并验证了局部解释算法对融合性能的促进作用。此外,为了抑制噪声和伪影,文献[8]提出了基于双边最小二乘滤波的混合模型DBLSF,结合结构张量自适应融合策略,在TNO数据集上实现了优异的边缘保护与对比度增强效果;文献[9]提出的MSFF网络则引入3D深度可分离卷积与中心像素注意力机制,兼顾了光谱与空间特征提取。

然而,上述方法在融合过程中仍存在局部特征整合不足、全局信息捕捉不够以及不同模态信息协同不足等问题,导致融合后图像在细节保留与结构一致性方面仍有待提升。为此,本文提出一种基于深度学习的可见光红外融合方法,在SwinFusion模型^[10]的基础上进行改进。首先,将LRRNet^[11]中基于低秩和稀疏表示的模块引入到模型的浅层和深层特征提取之间。该模块通过自动分解输入图像为反映全局结构信息的低秩部分与捕获局部细节的稀疏部分,有效提升特征提取的精度与鲁棒性;其次,在特征融合阶段引入SwinFuse^[12]中基于L1范数的融合策略,通过对全局特征矩阵分别从行和列两个维度计算L1范数,动态自适应地为不同模态特征分配权重,从而实现更均衡且细腻的特征融合。

1 SwinFusion 模型原理

SwinFusion提出了一种基于Swin Transformer的通用图像融合框架,通过联合CNN与Transformer的特征提取、注意力引导的跨域融合以及统一的多任务损失函数,实现了对多模态与多场景图像的高效融合。模型先采用两层卷积网络对输入图像进行局部浅层特征编码,再通过多层Swin Transformer捕捉全局长程依赖;在此基础上,交替堆叠域内多头自注意力与域间多头交叉注意力模块,既能挖掘同域内部的全局信息,又能实现不同图像域之间的特征互补;随后,通过若干Swin Transformer对融合特征进行深层重构,并以三层卷积网络生成融合图像。为兼顾结构一致性、纹理细节与强度分布,设计了包含结构相似性损失、基于Sobel梯度的纹理损失以及针对不同任务采用最大值或均值聚合的强度损失的统一目标函数。

2 基于深度学习的可见光与红外融合的主干网络结构

基于深度学习的可见光与红外融合的主干网络结构如图1所示,图中MSA为多头自注意力机制,FFN为前馈神经网络,LN为层归一化,MCA为多头交叉注意力机制。在SwinFusion框架基础上,将LRRNet中基于低秩与稀疏表示的协同特征分解模块嵌入到模型的浅层特征提取与深层语义抽象之间。低秩表征模块通过矩阵分解提取跨模态特征的全局结构信息,抑制噪声干扰并增强特征鲁棒性;稀疏表征模块则聚焦局部细节特征的精细化建模,强化可见光纹理与红外热辐射特征的有效解耦与表达。二者的联合建模实现了模态间特征的有效解耦与表达。

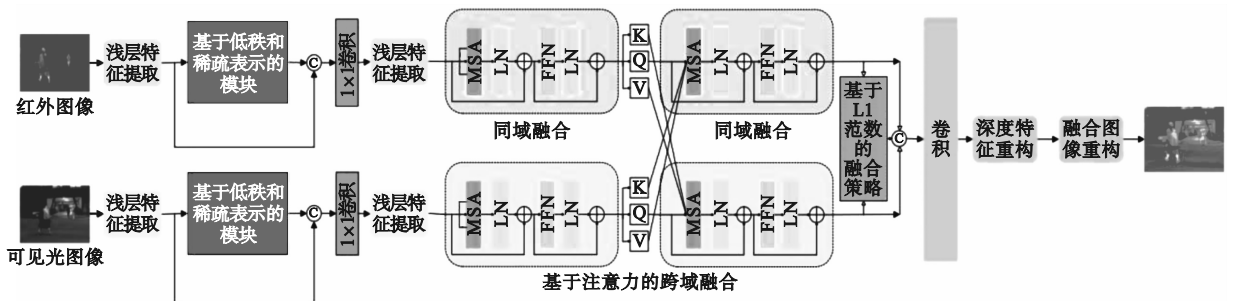


图1 基于深度学习的可见光与红外融合的主干网络

Fig. 1 A deep learning-based backbone network for visible-infrared image fusion

此外,在特征融合阶段引入基于 L1 范数的自适应融合策略,通过可学习的稀疏约束动态分配多模态特征权重,优化特征选择过程,解决传统融合方法中信息冗余与特征覆盖不足的问题。整体主干网络通过低秩-稀疏联合建模与自适应融合机制,实现了跨模态特征的结构化表达与高效整合。

2.1 基于低秩和稀疏表示的模块

本文引用了 LRRNet 中提出的一种基于低秩表示(LRR)和稀疏表示(SR)的图像分解模型。该模型同时采用核范数提取输入图像的“基底”部分,并通过 L1 范数提取“显著”部分,其原始形式为

$$\begin{cases} \min_{D_1, D_2, L, S} \|L\|_* + \lambda \|S\|_1 \\ \text{s. t. } X = D_1 L + D_2 S \end{cases} \quad (1)$$

式中: X 为一个 m 行 n 列的矩阵,每列都是一个图像块; L 为低秩系数矩阵; S 为稀疏系数矩阵; λ 为正则化参数;字典 D_1 、 D_2 分别将两矩阵映射到基底与显著特征空间; $\|\cdot\|_*$ 与 $\|\cdot\|_1$ 分别表示核范数与 L1 范数。

在满足条件 $X = D_1 L + D_2 S$ 时,可得

$$\min_{D_1, D_2, L, S} \|X - D_1 L - D_2 S\|_F^2 + \lambda_1 \|L\|_* + \lambda_2 \|S\|_1 + \lambda_3 (\|D_1\|_F^2 + \|D_2\|_F^2) \quad (2)$$

式中: λ_1 、 λ_2 、 λ_3 为正则化参数; $\|\cdot\|_F$ 表示 Frobenius 范数。借助核范数与 Frobenius 范数的等价映射得

$$\begin{cases} \|L\|_* = \min_{A, B} \frac{1}{2} \|A\|_F^2 + \frac{1}{2} \|B\|_F^2 \\ \text{s. t. } AB = L \end{cases} \quad (3)$$

结合公式(2)和公式(3)可得

$$\min_{D_1, D_2, A, B, S} \|X - D_1 AB - D_2 S\|_F^2 + \frac{\lambda_1}{2} (\|A\|_F^2 + \|B\|_F^2) + \lambda_2 \|S\|_1 + \lambda_3 (\|D_1\|_F^2 + \|D_2\|_F^2) \quad (4)$$

进一步令 $D = (D_1, D_2)$, $Z = (AB; S)$, $\phi(D) =$

$$\lambda_3 (\|D_1\|_F^2 + \|D_2\|_F^2), \psi(Z) = \frac{\lambda_1}{2} (\|A\|_F^2 + \|B\|_F^2) + \lambda_2 \|S\|_1, \text{即可归纳为}$$

$$\min_{D, Z} \frac{1}{2} \|X - DZ\|_F^2 + \phi(D) + \psi(Z) \quad (5)$$

在该模型中,输出 Z 同时包含低秩系数矩阵 AB 与稀疏系数矩阵 S 。

学习型迭代收缩阈值算法(LISTA)主要用于求解稀疏编码问题,公式(5)可通过此算法求解。LISTA 的一次迭代步骤可表示为

$$Z_t = h_\theta(\mu D^T X + (\hat{\lambda}_3 I - \mu D^T D) Z_{t-1}) \quad (6)$$

式中: h_θ 表示软阈值算子; θ 为收缩输入矩阵的阈值,本框架中 θ 是一个可学习参数; t 为实际迭代

次数; μ 为步长。

基于 LISTA 的计算流程如表 1 所示,其中 P_l 和 P_s 分别表示输入 X 的低秩部分和稀疏部分。该算法的首要目标是学习低秩系数矩阵 L 和稀疏系数矩阵 S 。在字典 D_1 和 D_2 固定的情况下, P_l 和 P_s 可通过 $D_1 L$ 和 $D_2 S$ 计算得到。

表 1 基于 LISTA 的算法

Table 1 Algorithm based on LISTA

输入: $X, D_1, D_2, \lambda_1, \lambda_2, \lambda_3$ 。输出: P_l, P_s 。
定义 1: $D = (D_1, D_2)$, $\hat{\lambda}_3 = 1 - \lambda_3$ 。
定义 2: $H = \lambda_3 I - \mu D^T D$, $W_c = \mu D^T$, $\theta = \mu \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$
1. 初始化 $B = W_c X$, $Z_0 = h_\theta(B)$ 。
2. for $t = 1$ to T do
3. $Z'_t = B + HZ_{t-1}$
4. $Z_t = h_\theta(Z'_t)$
5. end for
6. $Z = Z_T$, 分割 $Z = (AB; S)$, 并令 $L = AB$ 。
7. 输出 $P_l = D_1 L$ 和 $P_s = D_2 S$ 。

表 1 中算法的迭代过程可表达为

$$Z_t = h_\theta(W_c(X - DZ_{t-1}) + \hat{\lambda}_3 Z_{t-1}) \quad (7)$$

用卷积层替代矩阵乘法,上式可表达为

$$Z_t = h_\theta(C_2 * (X - C_1 * Z_{t-1}) + \hat{\lambda}_3 Z_{t-1}) \quad (8)$$

式中: $*$ 表示卷积操作; C_1 和 C_2 为可训练变量(卷积层)。基于该公式的图像分解架构称为学习低秩表示(LLRR)模块。LLRR 模块的架构如图 2 所示,包含两个卷积层和两条短连接。此处 $h_\theta(X)$ 定义为

$$h_\theta(X) = \text{sign}(X) \cdot \max(|X| - \theta, 0) \quad (9)$$

式中 X 表示输入数据。

特征 Z 由堆叠的 LLRR 模块输出得到。 Z 可拆分为 AB 和 S 两部分,即 $Z = (AB; S)$ 。 X 的低秩部分 $P_l = C_1 * L$, 稀疏部分 $P_s = C_s * S$ (可通过设置适当的卷积层 C_l 和 C_s 计算)。

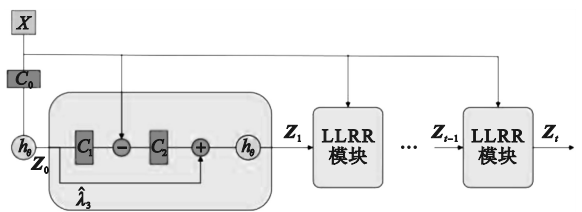


图 2 基于低秩和稀疏表示的模块

Fig. 2 A module based on low-rank and sparse representation

2.2 基于 L1 范数的融合策略

本文的改进模型引入了 SwinFuse 中提出的

一种面向序列矩阵的多维度自适应融合机制,其核心思想是通过 L1 范数度量跨模态特征的显著性差异,在行、列双向量空间动态分配权重。

选择 L1 范数而非 L2 范数主要有两点核心原因:一是 L1 范数定义为向量元素绝对值之和,优化时倾向于生成稀疏解。对于红外与可见光图像融合而言,这一特性能够突出亮度目标或纹理边缘等显著特征,并抑制非显著区域,从而实现有针对性的动态权重分配;二是 L2 范数定义为元素平方和,优化结果通常为全局平滑解,容易将各模态的重要信号平均化,削弱两种图像之间的互补优势。相比之下,L1 范数更能保持模态间的差异性和互补性,提高融合后图像的细节还原度和视觉效果。

图3展示了基于L1范数的融合策略,给定红外与可见光全局特征矩阵 $\Phi_{GF}^{ir} \in \mathbf{R}^{M \times N \times C}$, $\Phi_{GF}^{vis} \in \mathbf{R}^{M \times N \times C}$,其中 M 和 N 分别为特征图的高度和宽度, C 为通道数。融合过程分为以下几个阶段。

1) 行向量显著性度量:对特征矩阵第 i 行向量计算 L1 范数,通过 Softmax 生成模态间竞争权重。

$$\varphi_{row}^{ir}(i) = \frac{\exp(\|\Phi_{GF}^{ir}(i)\|_1)}{\exp(\|\Phi_{GF}^{ir}(i)\|_1) + \exp(\|\Phi_{GF}^{vis}(i)\|_1)} \quad (10)$$

$$\varphi_{row}^{vis}(i) = 1 - \varphi_{row}^{ir}(i) \quad (11)$$

式中 $\varphi_{row}^{ir}(i)$ 和 $\varphi_{row}^{vis}(i)$ 分别为第 i 行红外和可见光特征的权重。

2) 列向量互补性评估:对特征矩阵第 j 列向量执行相同操作。

$$\varphi_{col}^{ir}(j) = \frac{\exp(\|\Phi_{GF}^{ir}(j)\|_1)}{(\exp(\|\Phi_{GF}^{ir}(j)\|_1) + \exp(\|\Phi_{GF}^{vis}(j)\|_1))} \quad (12)$$

$$\varphi_{col}^{vis}(j) = 1 - \varphi_{col}^{ir}(j) \quad (13)$$

式中 $\varphi_{col}^{ir}(j)$ 和 $\varphi_{col}^{vis}(j)$ 分别为第 j 列红外和可见光特征的权重。

3) 行权重融合:按行维度加权聚合特征。

$$\Phi_F^{row}(i,j) = \Phi_{GF}^{ir}(i,j) \cdot \varphi_{row}^{ir}(i) + \Phi_{GF}^{vis}(i,j) \cdot \varphi_{row}^{vis}(i) \quad (14)$$

式中 $\Phi_F^{row}(i,j)$ 为行融合结果。

4) 列权重融合:按列维度补充细节信息

$$\Phi_F^{col}(i,j) = \Phi_{GF}^{ir}(i,j) \cdot \varphi_{col}^{ir}(j) + \Phi_{GF}^{vis}(i,j) \cdot \varphi_{col}^{vis}(j) \quad (15)$$

式中 $\Phi_F^{col}(i,j)$ 为列融合结果。

5) 跨维度特征增强:通过元素加操作融合双维度结果。

$$\Phi_F(i,j) = \Phi_F^{row}(i,j) + \Phi_F^{col}(i,j) \quad (16)$$

式中 $\Phi_F(i,j)$ 为最终融合特征,通过行、列双维度增强。

在 SwinFuse 中,该融合策略仅在测试阶段保留,而在训练阶段被移除,在本文的改进模型中,将该融合策略与低秩稀疏模块联合训练,通过核范数约束增强一致性。与 SwinFuse 相比,本文的改进模型通过特征提取阶段的特征解耦,在保留 SwinFuse 双维度 L1 范数优势的同时,强化了目标显著性与细节保真度。

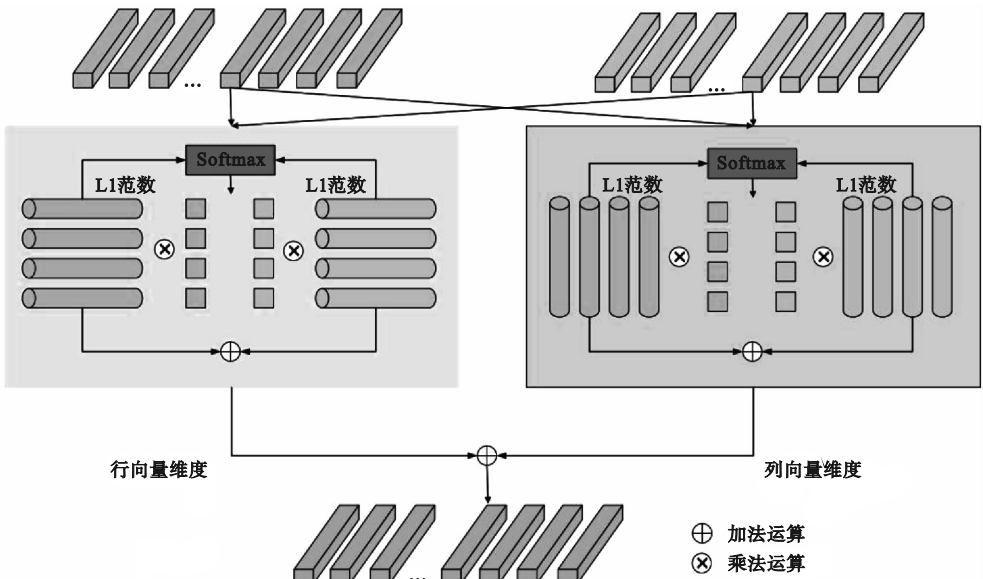


图3 基于L1范数的融合策略

Fig.3 Fusion strategy based on L1-norm

3 实验与分析

3.1 数据集

本文研究采用 DroneVehicle 数据集^[13]。该数据集由无人机采集的 56 878 张图像组成,其中一半是 RGB 图像,其余为红外图像,适用于可见光图像和红外图像融合的目标检测研究。为确保图像融合质量,本研究在实验阶段采用了 5 041 对对齐的可见光-红外图像,包括了日间与夜间的场景。其中,4 016 对作为训练集,剩余的 1 025 对作为测试集,所有图像的尺寸统一调整为 640×512 像素。

3.2 实验设置

在图像融合实验中,为了提升训练效率,训练集中的图像被裁剪为 64×64 的大小,批大小设为 4,Transformer 的窗口大小设为 8,初始学习率为 2×10^{-3} ,采用自适应矩阵估计优化器(Adam Optimizer)进行参数优化,训练到 8 000 代时终止。

为扩充训练数据集,将图像裁剪步长设为 24,所得到的每个补丁的大小统一为 128×128 ,结果生成了 6 921 对用于训练的图像补丁。由于红外图像中显著区域仅占很小部分,为平衡显著区域与背景区域的损失,本实验中 α 参数设置为 7。测试时,源图像直接输入融合网络,无需裁剪。

实验的软件环境包括 Windows 11 操作系统、Python 3.6.13、Pytorch 1.10.1 深度学习框架和 Pyc-harm 开发平台;硬件环境包括 16 GB 内存、Intel i7-9750H CPU 和 NVIDIA Geforce GTX 1660 Ti GPU。

3.3 实验结果分析

本文选用 7 种客观评价指标:EN(信息熵)用于衡量图像信息复杂度;CC(相关系数)和 MI(互信息)用于评估图像间的相关性;Qabf 用于量化融合图像质量;PSNR(信噪比)与 VIF(视觉保真度)用于衡量重建质量;SD(标准差)用于反映图像对比度或噪声水平。

3.3.1 消融实验

本节逐步将基于 L1 范数的融合策略和基于低秩和稀疏表示的模块引入 SwinFusion 主干网络中,并验证其有效性,探讨两个模块在优化特征提取与融合策略方面是否显著提升模型的预测精度与鲁棒性。

在 DroneVehicle 数据集上开展的消融实验结果如表 2 所示。未引入任何模块时,原模型在 EN(7.143)、CC(0.731)、Qabf(0.521)等指标上表现中等,PSNR(63.821)、MI(2.933)、VIF(0.550)与 SD(46.158)指标表明其信息保留与对比度能力有限。

仅引入 L1 范数融合策略,EN 提升至 7.289, Qabf 增至 0.551, VIF 显著提高至 0.621,验证了动态加权策略对特征显著性的捕捉能力;CC 略降至 0.727,SD 降至 45.513,说明单一策略可能削弱部分结构相关性。

仅引入低秩-稀疏模块,PSNR 跃升至 65.599, MI 提升至 3.112, VIF 达 0.607,表明该模块通过分解全局结构与局部细节,显著增强了特征鲁棒性;但 SD 进一步降至 45.174,反映单一模块对对比度优化的局限性。

联合引入双模块的协同作用下,EN(7.306)、CC(0.734)、Qabf(0.554)、SD(46.232)均达最优值,SD 回升至 46.232,且 PSNR(64.569)保持较高水平。尽管 PSNR 和 MI 较仅引入 LLRR 模块时分别下降 1.03 和 0.078,但 Qabf 提升 0.01、VIF 提升 0.005,证明双模块在结构保真与噪声抑制间实现了更优的平衡。

总之,L1 范数策略侧重动态特征加权,提升细节显著性与视觉感知质量;LLRR 模块通过特征解耦,优化全局信息保留与抗噪能力;二者协同弥补单一模块的不足,最终在 7 项指标中 4 项达到最优,验证了改进方案的系统性优势,证明了所提改进方案在提升特征提取与融合能力方面的有效性。

表 2 消融实验结果

Table 2 Results of ablation experiment

L1 范数	LLRR	EN	CC	Qabf	PSNR/dB	MI	VIF	SD
×	×	7.143	0.731	0.521	63.821	2.933	0.550	46.158
√	×	7.289	0.727	0.551	64.593	3.029	0.621	45.513
×	√	7.284	0.729	0.544	65.599	3.112	0.607	45.174
√	√	7.306	0.734	0.554	64.569	3.034	0.612	46.232

3.3.2 对比实验

在对比实验中,本文选取了八种具有代表性的先进图像融合方法作为对比算法。AUIF-Net^[14]通过自适应权重分配实现多模态特征动态整合;Dual-branch^[15]采用双分支架构结合交叉注意力机制增强跨模态特征交互;TGFuse^[16]结合了Transformer与GAN框架,利用通道/空间变换器建模全局依赖;DIDFuse^[17]基于深度图像分解范式,通过位置注意力协调低频背景与高频细节;DenseFuse^[18]借助密集连接促进特征复用,结合双融合策略优化特征表达;RFN-Nest采用嵌套残差结构与两阶段训练策略提升重建保真度;Nest-Fuse^[19]设计嵌套编解码器与双阶段注意力机制实现多尺度特征自适应加权;SwinFusion则构建基于Swin Transformer的通用框架,通过域内/跨域注意力单元统一多源信息融合。为验证本文所提出算法的有效性,基于前文提及的7项核心指标,在DroneVehicle数据集上与以上八种主流融合算法进行对比实验。实验结果如表3所示。为保证实验的公平性,所有算法均在相同分辨率(640×512)的图像输入条件下进行评估。

从表3可以看出,传统算法AUIF-Net在各项

指标中均表现不佳,尤其在Qabf指标上仅为0.218,难以有效保留图像的结构与边缘信息。本文所提算法在EN、Qabf、MI和VIF等4项指标上达到最优,并且在CC和PSNR上的性能指标也较好。EN(7.306)与MI(3.034)均达到最高值,验证了低秩-稀疏模块对全局信息与多模态特征互补性的高效保留能力。SwinFusion整体性能较为均衡,相比之下,本文所提算法在Qabf(0.554)与VIF(0.612)方面优于Swinfusion;尽管PSNR(64.569)略低于DenseFuse(65.613),但其Qabf与VIF的显著提升证明所提算法在噪声抑制与细节保留间实现了更优平衡。此外,CC(0.734)并不明显低于最优值(0.794),进一步佐证了融合结果与源图像的高度相关性。值得注意的是,DIDFuse在SD指标上达到了72.356,明显优于其他算法,体现了该算法对比度和信息分布的均匀性,但本文所提算法在与其他算法的对比中该项指标仍高于平均水平。总之,所提算法在4项指标(EN、Qabf、MI、VIF)中取得最优,其余指标亦处于领先水平,充分验证了低秩-稀疏特征解耦与动态加权融合策略为复杂场景下的多模态图像融合提供了更鲁棒的解决方案。

表3 对比实验结果

Table 3 Comparative experimental results

算法	EN	CC	Qabf	PSNR/dB	MI	VIF	SD
AUIF-Net	4.919	0.658	0.218	55.591	2.117	0.251	46.162
Dual-branch	7.242	0.680	0.416	64.145	2.682	0.460	49.406
TGFuse	7.236	0.723	0.461	64.405	2.224	0.474	45.527
DIDFuse	7.238	0.743	0.341	60.281	2.508	0.432	72.356
DenseFuse	6.802	0.767	0.184	65.613	2.271	0.339	33.518
RFN-Nest	7.135	0.794	0.415	65.395	2.449	0.455	42.634
NestFuse	7.271	0.755	0.459	64.516	3.032	0.534	46.752
SwinFusion	7.143	0.731	0.521	63.821	2.933	0.550	46.158
本文算法	7.306	0.734	0.554	64.569	3.034	0.612	46.232

3.3.3 可视化对比

为了进一步检验本文提出算法的融合质量,将本文算法与对比实验中选取的算法在DroneVehicle数据集中选取其中一对图像的测试结果进行可视化对比,结果如图4所示。从图4可以看出,部分算法在处理夜间有照明条件下的复杂交通环境图像时表现不佳,如AUIF-Net算法的融合结果曝光过于严重;DIDFuse的融合结果中光线亮度过高,而车辆细节丢失严重;Dual-branch在复杂环

境中引入大量的噪声与伪影,导致车辆细节与车道实线部分融合效果不佳;在DenseFuse中融合图像的清晰度下降。其余几种算法的融合结果表现较为均衡,但在RFN-Nest、TGFuse中仍然存在轻微的因特征覆盖不足,导致红外目标(如车辆)的亮度信息被可见光背景弱化的情况。

相比之下,本文所提算法表现更优。在夜间复杂交通场景中,融合结果能增强红外目标的显著性,同时保持可见光背景的自然过渡,例如实现

灯光照明区域与阴影区域之间的平滑衔接;在强光场景中,融合结果能提供更均衡的亮度分布,有效抑制高光溢出,从而避免高亮区域(如车灯)的信息丢失,保留车灯周围的细节信息;在动态场景中,本文算法的融合结果更干净,减少了模糊与伪影;在复杂环境中,融合结果能同时保留红外穿透性与可见光场景语义,图像中红色客车的细节与色彩还原度更高。本文提出的融合算法通过引入低秩与稀疏特征解耦模块以及 L1 范数动态加权策略,在可视化对比中展现出显著优势,能够同时保留可见光图像的细节纹理与红外图像的热目标显著性,并通过全局对比度优化与噪声抑制,生成结构一致、伪影极少的融合结果,在复杂光照与动态场景下表现尤为突出。

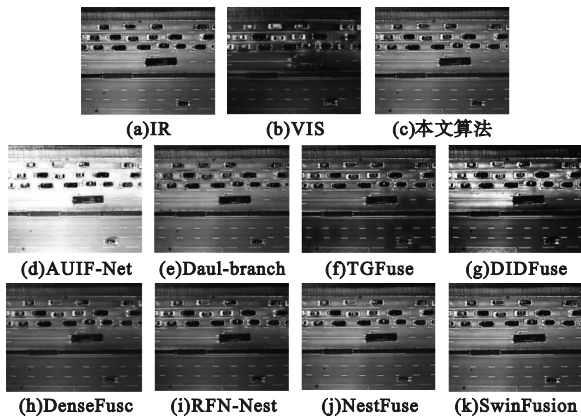


图 4 可视化对比

Fig. 4 Visual comparisons

4 结论

本研究提出了一种基于深度学习的可见光与红外图像融合方法,并在 SwinFusion 主干网络基础上,通过引入基于低秩和稀疏表示的模块以及基于 L1 范数的融合策略,实现了多模态特征的高效提取与自适应融合。实验结果表明,改进后的模型在 DroneVehicle 数据集上明显改善了图像的结构一致性、信息熵和视觉信息保真度等指标,相较于现有的主流方法,表现出更优的融合效果和较高的鲁棒性。通过实验验证此算法在自动驾驶等实际应用场景中展现出较大潜力。未来将进一步关注模型结构的优化、训练数据的扩展以及在更多实际应用场景中的验证,以期不断推动多模态图像处理技术的进步。

参考文献 (References):

[1] Li H, Wu X J, Kittler J. RFN-Nest: an end-to-end residual fu-

sion network for infrared and visible images[J]. Information Fusion, 2021, 73: 72 - 86.

- [2] 孙学蕾, 高宏伟. 改进小波变换的红外与可见光融合方法研究[J]. 沈阳理工大学学报, 2021, 40(3): 19 - 23, 28.
Sun X L, Gao H W. Study on infrared and visible light fusion methods to improve wavelet transformation[J]. Journal of Shenyang Ligong University, 2021, 40(3): 19 - 23, 28. (in Chinese)
- [3] 刘砚菊, 崔洁, 宋建辉, 等. GAN 网络的红外与可见光图像融合方法研究[J]. 沈阳理工大学学报, 2022, 41(3): 14 - 19, 26.
Liu Y J, Cui J, Song J H, et al. Research on infrared and visible images fusion method of GAN network[J]. Journal of Shenyang Ligong University, 2022, 41(3): 14 - 19, 26. (in Chinese)
- [4] 沈瑜, 梁丽, 王海龙, 等. 基于 N-RGAN 模型的红外与可见光图像融合[J]. 红外技术, 2023, 45(9): 897 - 906.
Shen Y, Liang L, Wang H L, et al. Infrared and visible image fusion based on N-RGAN model[J]. Infrared Technology, 2023, 45(9): 897 - 906. (in Chinese)
- [5] Liang Z Y, Yu M X, Sun Y C, et al. PEFusion: an infrared and visible light fusion method for enhanced representation of prominent targets and exposure errors[J]. Infrared Physics & Technology, 2024, 141: 105496.
- [6] Liu Q, Pi J T, Gao P, et al. STFNet: self-supervised transformer for infrared and visible image fusion[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2024, 8(2): 1513 - 1526.
- [7] Yuan B, Sun H Y, Guo Y J, et al. Explainable analysis of infrared and visible light image fusion based on deep learning[J]. Scientific Reports, 2025, 15(1): 2223.
- [8] Lu Q, Han Z D, Hu L K, et al. An infrared and visible image fusion algorithm method based on a dual bilateral least squares hybrid filter[J]. Electronics, 2023, 12(10): 2292.
- [9] Gong G, Wang X P, Zhang J H, et al. MSFF: a multi-scale feature fusion convolutional neural network for hyperspectral image classification[J]. Electronics, 2025, 14(4): 797.
- [10] 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.
- [11] Li H, Xu T Y, Wu X J, et al. LRRNet: a novel representation learning guided fusion network for infrared and visible images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(9): 11040 - 11052.
- [12] Wang Z S, Chen Y L, Shao W Y, et al. SwinFuse: a residual Swin Transformer fusion network for infrared and visible images[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1 - 12.
- [13] Sun Y M, Cao B, Zhu P F, et al. Drone-based RGB-infrared cross-modality vehicle detection via uncertainty-aware learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(10): 6700 - 6713.
- [14] Zhao Z X, Xu S, Zhang J S, et al. Efficient and model-based infrared and visible image fusion via algorithm unrolling[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(3): 1186 - 1196.
- [15] Fu Y, Wu X J. A Dual-branch network for infrared and visible image fusion[C]//2020 25th International Conference on Pattern Recognition (ICPR). Milan, Italy: IEEE, 2021: 10675 - 10680.

(下转第 58 页)

- [15] Zhang L Y, Wu X, Liu Z H, et al. ESD-YOLOv8: an efficient solar cell fault detection model based on YOLOv8 [J]. IEEE Access, 2024, 12: 138801 – 138815.
- [16] Liu Z G, Sun B S, Bi K Y. Optimization of YOLOv7 based on PConv, SE attention and wise-IoU [J]. International Journal of Computational Intelligence and Applications, 2024, 23: 2350033.
- [17] Hughes D P, Salathé M. An open access repository of images on plant health to enable the development of mobile disease diagnostics [PP/OL]. arXiv (2016 – 04 – 12) [2025 – 08 – 15]. <https://doi.org/10.48550/arXiv.1511.08060>.
- [18] Universitas Atma Jaya. Tomato leaf disease dataset [DS/OL]. Roboflow Universe (2023 – 04 – 01) [2025 – 08 – 15]. <https://universe.roboflow.com/universitas-atma-jaya/tomato-leaf-disease-rxcft>.
- [19] Zhu X K, Lyu S C, Wang X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal, QC, Canada; IEEE, 2021: 2778 – 2788.
- [20] Hussain M. YOLOv1 to v8: unveiling each variant; a comprehensive review of YOLO [J]. IEEE Access, 2024, 12: 42816 – 42833.
- [21] Murat A A, Kiran M S. A comprehensive review on YOLO versions for object detection [J]. Engineering Science and Technology, an International Journal, 2025, 70: 102161.
- [22] Ali M L, Zhang Z. The YOLO framework: a comprehensive review of evolution, applications, and benchmarks in object detection [J]. Computers, 2024, 13 (12): 336.
- [23] Sapkota R, Flores-Calero M, Qureshi R, et al. YOLO advances to its genesis: a decadal and comprehensive review of the You Only Look Once (YOLO) series [J]. Artificial Intelligence Review, 2025, 58 (9): 274.

(责任编辑:宋颖韬)

(上接第41页)

- [16] Rao D Y, Xu T Y, Wu X J. TGFuse: An infrared and visible image fusion approach based on transformer and generative adversarial network [J]. IEEE Transactions on Image Processing, 2023, PP (99): 1.
- [17] Zhao Z X, Xu S, Zhang C X, et al. DIDFuse: deep image decomposition for infrared and visible image fusion [C]//2020 29th International Joint Conference on Artificial Intelligence (IJCAI). Yokohama, Japan; IJCAI Organization, 2020: 970 –

976.

- [18] 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.
- [19] Li H, Wu X J, Durrani T. NestFuse: an infrared and visible image fusion architecture based on nest connection and spatial/channel attention models [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69 (12): 9645 – 9656.

(责任编辑:和晓军)

(上接第49页)

- [7] Awad N H, Ali M Z, Suganthan P N. Ensemble sinusoidal differential covariance matrix adaptation with Euclidean neighborhood for solving CEC2017 benchmark problems [C]//2017 IEEE Congress on Evolutionary Computation (CEC). Donostia, Spain; IEEE, 2017: 372 – 379.
- [8] Zhang G, Liu H C, Zhang J B, et al. Wind power prediction based on variational mode decomposition multi-frequency combinations [J]. Journal of Modern Power Systems and Clean Energy, 2019, 7 (2): 281 – 288.
- [9] Van Houdt G, Mosquera C, Nápoles G. A review on the long short-term memory model [J]. Artificial Intelligence Review, 2020, 53 (8): 5929 – 5955.
- [10] Zhong C T, Li G, Meng Z, et al. Starfish optimization algorithm (SFOA): a bio-inspired metaheuristic algorithm for global optimization compared with 100 optimizers [J]. Neural Computing and Applications, 2025, 37 (5): 3641 – 3683.
- [11] Oueslati R, Ouertani M W, Amdouni A, et al. MS-FSOA-LightGBM: multi-strategy starfish optimization algorithm with LightGBM for software defect prediction [J]. Procedia Computer Science, 2025, 270: 4054 – 4063.
- [12] Liu S L, Jin Z K, Lin H T, et al. An improve crested porcupine algorithm for UAV delivery path planning in challenging environments [J]. Scientific Reports, 2024, 14: 20445.

- [13] Abdollahpour A, Rouhi A, Pira E. An improved gazelle optimization algorithm using dynamic opposition-based learning and chaotic mapping combination for solving optimization problems [J]. The Journal of Supercomputing, 2024, 80 (9): 12813 – 12843.
- [14] 尚凯凯. 融合扰动策略的自适应哈里斯鹰优化算法 [J]. 计算机与数字工程, 2025, 53 (2): 338 – 346.
- Shang K K. Adaptive Harris Hawks optimization algorithm based on disturbance strategies [J]. Computer and Digital Engineering, 2025, 53 (2): 338 – 346. (in Chinese)
- [15] Mirjalili S, Lewis A. The whale optimization algorithm [J]. Advances in Engineering Software, 2016, 95: 51 – 67.
- [16] Hashim F A, Hussain K, Houssein E H, et al. Archimedes optimization algorithm: a new metaheuristic algorithm for solving optimization problems [J]. Applied Intelligence, 2021, 51 (3): 1531 – 1551.
- [17] Amiri M H, Mehrabi Hashjin N, Montazeri M, et al. Hippopotamus optimization algorithm: a novel nature-inspired optimization algorithm [J]. Scientific Reports, 2024, 14: 5032.
- [18] Guan Y M, Shi Y L, Wang G, et al. Electricity behaviors anomaly detection based on multi-feature fusion and contrastive learning [J]. Information Systems, 2025, 127: 102457.

(责任编辑:和晓军)