中图法分类号:TP751.2;TP391　　**文献标识码**:A　　**文章编号**:1006-8961(2026)11-3863-13

论文引用格式:Xue J Z, Yu F M and Liu Y. 2026. Hyperspectral and multispectral image fusion via sparse reconstruction and cross-guidance. Journal of Image and Graphics, 31(11):3863-3875(薛吉则, 俞凤梅, 刘颖. 2026. 稀疏重建与交叉引导的高光谱与多光谱图像融合. 中国图象图形学报, 31(11):3863-3875)[DOI:10.11834/jig.250618]

稀疏重建与交叉引导的高光谱与多光谱图像融合

薛吉则*,俞凤梅,刘颖

西安邮电大学信息与通信工程学院,西安　710100

摘　要:**目的**　针对现有高光谱与多光谱图像融合算法光谱重建策略不足和模态融合机制缺失的问题,提出一种稀疏重建与交叉引导的高光谱多光谱图像融合算法(hyperspectral and multispectral image fusion via sparse reconstruction and cross-guidance, HMIF-SRNet)。**方法**　该算法采用端到端的融合架构,通过预处理、特征提取、核心融合和特征重建4个模块实现高质量融合。为了光谱信息的高效利用和精确保真以及双模态信息的充分互补,设计了两个核心模块:光谱稀疏重建单元(spectral reconstruction unit,SRU)和交叉引导单元(cross-reference unit,CRU)。具体而言,SRU模块采用双重统计门控策略和动态稀疏重建机制,自适应优化光谱维度的信息表达,有效降低光谱冗余;CRU模块通过交叉引导机制实现高光谱和多光谱特征的互补融合,增强了跨模态信息的交互效果。两个核心模块通过串行级联实现递进式的特征优化。**结果**　HMIF-SRNet在峰值信噪比(peak signal-to-noise ratio,PSNR)、相对全局无量纲误差(relative dimensionless global error in synthesis,ERGAS)和通用图像质量指数(universal image quality index,UIQI)3项指标上取得最优结果;在深度学习类对比方法中SAM(spectral angle mapper)指标最优,传统变换方法SAM数值更低,但整体图像重建性能较差。其中,在CAVE(cave multispectral image database)数据集上PSNR相比次优方法提升了4.41 dB,SAM减少了32.0%,在真实场景下展现出卓越的鲁棒性。**结论**　通过轻量化网络架构设计,该算法在保持优异融合性能的同时有效降低了计算复杂度,验证了动态稀疏重建策略与交叉引导融合机制的有效性。

关键词:高光谱图像(HSI);多光谱图像(MSI);图像融合;全局特征;稀疏注意力

Hyperspectral and multispectral image fusion via sparse reconstruction and cross-guidance

Xue Jize*, Yu Fengmei, Liu Ying

School of Information and Communication Engineering, Xi'an University of Posts and Telecommunications, Xi'an　710100, China

Abstract: **Objective**　High-resolution hyperspectral imaging is critically important across diverse application domains, including remote sensing monitoring, precision agriculture, mineral exploration, and environmental assessment. However, current imaging sensors face an inherent trade-off between spatial and spectral resolution due to physical constraints and manufacturing limitations. Specifically, increasing the number of spectral sampling channels reduces the light energy received by individual detectors, thereby decreasing the signal-to-noise ratio and limiting improvements in spatial sampling density. Conversely, enhancing spatial resolution requires reducing the number of spectral channels to maintain sufficient

收稿日期:2025-12-19;**修回日期**:2026-02-12;**预印本日期**:2026-02-19

*通信作者:薛吉则　Jize.Xue@xupt.edu.cn

基金项目:国家自然科学基金项目(62301423);陕西省教育厅科学研究专项项目(23JK0671)

Supported by: National Natural Science Foundation of China(62301423); Scientific Research Special Project of Shaanxi Provincial Department of Education(23JK0671)

signal strength. Therefore, existing sensors can typically provide either high-spatial-resolution multispectral images (3–10 bands) or high-spectral-resolution images with low spatial resolution (dozens to hundreds of bands), making it difficult to simultaneously achieve both high spatial and high spectral resolution. **Method** To address the limitations of existing hyperspectral and multispectral image fusion algorithms, particularly the insufficient spectral reconstruction strategies and the lack of effective modality fusion mechanisms, this paper proposes HMIF-SRNet (hyperspectral and multispectral image fusion with sparse reconstruction networks), a novel fusion framework based on sparse reconstruction networks. The proposed method adopts an end-to-end architecture that achieves high-quality fusion through four carefully designed modules: preprocessing, feature extraction, core fusion, and feature reconstruction. The preprocessing module standardizes and aligns the two modalities via bicubic interpolation upsampling and spectral response function mapping, ensuring consistency in both spatial and spectral dimensions. The feature extraction module employs a dual-branch parallel architecture to process hyperspectral images (HSI) and multispectral images (MSI) separately. The HSI branch captures joint spectral-spatial features using multi-layer three-dimensional convolutions, while the MSI branch focuses on extracting spatial details through two-dimensional convolutions with residual connections, thereby maintaining gradient flow and accelerating network convergence. To ensure efficient utilization and precise preservation of spectral information, as well as effective complementarity between modalities, two key modules are designed: the spectral reconstruction unit (SRU) and the cross-reference unit (CRU). The SRU addresses the lack of effective sparse reconstruction strategies in the spectral domain by introducing a dual statistical gating strategy combined with a dynamic sparse reconstruction mechanism. It first performs global statistical analysis on fused features using both global average pooling and global maximum pooling to capture average response intensity and peak response characteristics. These complementary statistics provide rich information for subsequent weight learning. A multi-layer perceptron (MLP) is then used to learn adaptive spectral channel weights, enabling dynamic adjustment of each channel's contribution. Finally, a spatial gating mechanism is incorporated to form a dual-gating strategy that jointly regulates spectral channels and spatial positions, ensuring both efficient utilization and high-fidelity preservation of spectral information. The CRU addresses the absence of explicit guidance mechanisms in cross-modal fusion by introducing a cross-guidance strategy for complementary feature integration. The module takes three inputs: the original F_hsi and F_msi from the feature extraction stage (used for global context modeling), and the optimized features F_stage1 generated by the SRU. Global context features are first computed via global average pooling, producing channel-wise descriptors that reflect the statistical properties of each modality. These descriptors are then passed through a two-layer linear transformation with nonlinear activation to generate cross-guidance weights, which encode the relative importance of each modality across channels. Using these weights, the CRU performs channel-wise weighted fusion of SRU-optimized features and downsampled MSI features, dynamically adapting the fusion strategy to different scenes and achieving an optimal balance between spectral fidelity and spatial detail. The SRU and CRU modules are cascaded to enable progressive feature refinement. First, the SRU performs dynamic sparse reconstruction in the spectral domain using a frequency decomposition strategy that separates features into high-frequency and low-frequency components. The high-frequency branch preserves fine spectral details through spatial attention and dynamic sparse selection, while the low-frequency branch captures the main spectral structure via global max pooling. These branches are then dynamically fused and downsampled to produce features with enhanced spectral representation. This stage primarily reduces spectral redundancy and selects the most discriminative features. Subsequently, the CRU performs cross-modal fusion by generating channel-wise adaptive weights based on global context and applying them to combine SRU-optimized features with MSI features. The final output integrates the spectral fidelity provided by SRU with the spatial detail contributed by MSI. **Result** Comparative experiments with 10 mainstream fusion methods on the CAVE and Harvard standard datasets demonstrate that HMIF-SRNet achieves superior performance in terms of PSNR, ERGAS and UIQI. For the spectral angle mapper(SAM) metric, the proposed method obtains the best result among deep learning-based fusion approaches, with traditional transformation methods yielding lower SAM values yet poor overall reconstruction quality. On the cave multispectral image database (CAVE) dataset, PSNR improves by 4.41 dB (a relative improvement of 9.4%) compared to the second-best method SSRNet, reaching 51.124 4 dB. ERGAS decreases by 9.5% to 0.472 6, SAM is reduced by 32.0% to 2.014 3, and UIQI increases to 0.983 2. On the Harvard dataset, which represents more complex real-world scenarios, HMIF-SRNet

maintains superior performance, with ERGAS decreasing by 4.44% compared to the second-best method SSFCNN and UIQI reaching 0.942 2. Ablation studies further confirm the effectiveness of each core module, demonstrating that both SRU and CRU contribute significantly to performance improvements across all metrics. **Conclusion** Through a lightweight network design, the proposed method effectively reduces computational complexity while maintaining excellent fusion performance. Visualization results show that HMIF-SRNet produces fusion images that are closest to the ground truth, with minimal error distribution, demonstrating strong robustness in real-world scenarios. Overall, the results validate the effectiveness of the dynamic sparse reconstruction strategy and the cross-guidance fusion mechanism, confirming that the proposed method outperforms existing techniques in spectral fidelity, spatial detail preservation, and overall fusion quality.

Key words: hyperspectral image(HSI); multispectral image(MSI); image fusion; global feature; sparse attention

0 引 言

受成像传感器的物理原理和制造工艺约束，当前高光谱成像系统面临着空间分辨率与光谱分辨率之间的固有权衡。具体而言，增加光谱采样通道数会导致单个探测器接收的光能减少，进而降低信噪比并限制空间采样密度的提升；反之，提高空间分辨率则需要减少光谱通道数以保证足够的信号强度。因此，现有传感器只能提供高空间分辨率的多光谱图像（通常3～10个波段）或高光谱分辨率的低空间分辨率图像（数十至数百个波段），难以同时满足高空间和高光谱分辨率的需求。

然而，在遥感监测、精准农业、矿产勘探以及环境评估等众多应用领域中，同时具备高空间分辨率和高光谱分辨率的图像数据至关重要。高空间分辨率能够提供精细的地物几何结构和纹理信息，而高光谱分辨率则能够揭示物质的光谱特征和化学组成，两者结合才能实现对目标场景的全面精准分析。为了获取这种理想的图像数据，目前主要存在两种技术路线。

第1种路线是改进硬件传感器设计，通过优化光学系统、提升探测器性能等方式突破物理限制。然而，受限于光学衍射极限、信号噪声比以及高精度制造工艺的成本约束，这一路线在短期内难以实现突破性进展，且所需的研发和制造成本极其高昂，限制了其在实际应用中的推广。

第2种路线是采用计算成像的融合策略，通过同时采集同一场景的低空间分辨率高光谱图像（low resolution hyperspectral image，LR-HSI）和高空间分辨率多光谱图像（high resolution multispectral image，HR-MSI），利用算法重建出高空间分辨率高光谱图像（high resolution hyperspectral image，HR-HSI）。这种软硬件协同的方式不仅成本较低、实施灵活，更重要的是能够充分利用两种模态数据的互补特性：LR-HSI提供丰富的光谱先验信息指导空间细节重建，HR-MSI的高分辨率空间结构则约束光谱恢复过程，从而实现光谱保真度与空间精度的双重保障。因此，基于图像融合的计算成像方法已成为当前研究和应用的主流方向。

针对上述高光谱与多光谱图像融合问题，学术界已开展了大量研究工作。传统的基于模型的方法采用手动设置的先验，如总变分（He等，2016）、非局部相似（He等，2019）、低秩性质（Chang等，2017，2020）和稀疏性（Wei等，2017）正则化等技术来探索空间和光谱相关性，并且获得很好的性能。然而，基于模型的HSI复原方法的有效性取决于人工制作先验的准确性。此外，在实际的HSI复原应用中，这些方法通常很难达到实时处理效果，并且在各种退化场景中的泛化能力有限。深度学习方法（Cao等，2022；Fu等，2022；Zhang等，2022；余典 等，2023）能够利用数据优势去拟合较优的解，已成功应用于HSI复原，取得了较优的效果。然而，它们是可解释性有限的黑盒，忽略了HSI的固有特征，如光谱相似度、空间局部相似度等。部分工作（Dong等，2016；Kanatsoulis等，2018）试图利用潜在的光谱特性获得更好的结果。近年来，基于注意力机制的方法（Vaswani等，2017）以及各种改进（Zhong等，2022；杨勇等，2025）在图像融合领域都获得了较好的性能。该技术能够提取全局特征，对HR-HSI重建是有效的。对于LR-HSI的超分辨率问题，一些工作（Chen等，2024；Liu等，2024）采用Spatial Transformer提取全局特征。然而，现有方法在处理高光谱数据时仍存在两个关键问题：一是在光谱维度上缺乏有效的稀疏

重建策略，导致冗余光谱信息的干扰；二是在跨模态融合时缺乏显式的引导机制，难以充分挖掘HSI和MSI的互补特性。这些局限性影响了融合图像的质量，同时带来了较高的计算成本。

为了解决上述问题，综合考虑模型驱动和数据驱动的优势，本文设计了一种稀疏重建与交叉引导的高光谱多光谱图像融合算法，采用这种方式，不仅很好地保留了源图像的固有特性，并对互补的信息进行融合；同时，先验信息不是通过手工设计，而是通过深度学习的数据驱动优势来探索先验知识。

本文的主要贡献包括：1）针对现有方法在光谱维度缺乏有效稀疏重建策略的问题，提出了光谱稀疏重建单元（spectral reconstruction unit，SRU）。该单元通过动态稀疏重建机制和双重统计门控策略，实现光谱维度的自适应优化和高效表达，有效降低光谱冗余并提升光谱保真度。在CAVE多光谱图像数据集（cave multispectral image database，CAVE）上，相比次优方法光谱角映射（spectral angle mapper，SAM）指标降低32.0%；2）针对跨模态融合缺乏显式引导机制的问题，设计了交叉引导单元（cross-reference unit，CRU）。该单元通过双向注意力机制实现高光谱与多光谱特征的深度互补融合，显著增强跨模态信息交互效果。在Harvard数据集上通用图像质量指数（universal image quality index，UIQI）指标达到0.942 2，有效平衡了空间细节与光谱一致性；3）构建了完整的端到端融合框架，将SRU和CRU通过串行级联实现递进式特征优化。在CAVE和Harvard两个标准数据集上与10种主流方法的对比实验表明，所提方法在峰值信噪比（peak signal-to-noise ratio，PSNR）、相对全局无量纲误差（relative dimensionless global error in synthesis，ERGAS）和UIQI共3指标上均取得最优结果；在深度学习类对比方法中，SAM指标最优，传统变换方法SAM数值更优，但整体重建性能较差。其中CAVE数据集PSNR相比次优方法提升4.41 dB，验证了动态稀疏重建策略与交叉引导融合机制的有效性。

1 提出的方法

现有高光谱与多光谱图像融合方法存在两个关键问题：一是在光谱维度上缺乏有效的稀疏重建策略，导致冗余光谱信息的干扰，影响光谱保真度；二是在跨模态融合时缺乏显式的引导机制，难以充分挖掘HSI和MSI的互补特性，导致空间细节与光谱一致性难以平衡。针对上述问题，本文提出一种稀疏重建与交叉引导的高光谱多光谱图像融合算法（hyperspectral and multispectral image fusion via sparse reconstruction and cross-guidance，HMIF-SRNet）。该算法采用端到端的融合架构，通过预处理、特征提取、核心融合和特征重建4个模块实现LR-HSI与HR-MSI的高质量融合。其中，核心融合模块包含两个创新单元：SRU通过动态稀疏重建机制自适应优化光谱维度信息表达，有效降低光谱冗余；CRU通过交叉引导机制实现高光谱与多光谱特征的互补融合，增强跨模态信息交互效果。两个核心模块通过串行级联实现递进式的特征优化。本节依次介绍模型整体架构、光谱重建单元（SRU）、交叉引导单元（CRU）、模块协同机制以及特征重建模块的设计细节。

1.1 模型架构

HMIF-SRNet的架构如图1所示。整个网络采用分阶段处理策略，依次通过预处理、特征提取、核心融合和特征重建4个模块实现高质量的HSI-MSI融合。

设输入的低分辨率高光谱图像为$\boldsymbol{X}_{\mathrm{LR}} \in \mathbf{R}^{h \times w \times c}$，高分辨率多光谱图像为$\boldsymbol{X}_{\mathrm{MS}} \in \mathbf{R}^{H \times W \times b}$，其中$h$、$w$分别为低分辨率图像的高度和宽度，$H$、$W$为高分辨率图像的高度和宽度，$c$为高光谱通道数，$b$为多光谱通道数，下采样因子$r = H/h = W/w$。目标是重建高分辨率高光谱图像$\boldsymbol{Y} \in \mathbf{R}^{H \times W \times c}$。

预处理阶段对输入的两种模态数据进行标准化和尺度对齐。首先，通过双三次插值将$\boldsymbol{X}_{\mathrm{LR}}$上采样至与$\boldsymbol{X}_{\mathrm{MS}}$相同的空间分辨率。具体为

$$\boldsymbol{X}_{\mathrm{up}} = \mathcal{U}\left(\boldsymbol{X}_{\mathrm{LR}}\right) \tag{1}$$

式中，$\mathcal{U}(\cdot)$表示双三次插值上采样操作，$\boldsymbol{X}_{\mathrm{up}} \in \mathbf{R}^{H \times W \times c}$为上采样后的高光谱图像。

然后，采用光谱响应函数建立两种模态间的光谱对应关系。具体为

$$\boldsymbol{X}_{\mathrm{sim}} = \boldsymbol{X}_{\mathrm{MS}} \times \boldsymbol{R} \tag{2}$$

式中，$\boldsymbol{R} \in \mathbf{R}^{b \times c}$为从多光谱到高光谱的光谱响应矩阵，$\boldsymbol{X}_{\mathrm{sim}} \in \mathbf{R}^{H \times W \times c}$为模拟的高光谱图像，确保两种模态数据在空间和光谱维度上的兼容性。

特征提取模块采用双分支并行架构分别处理

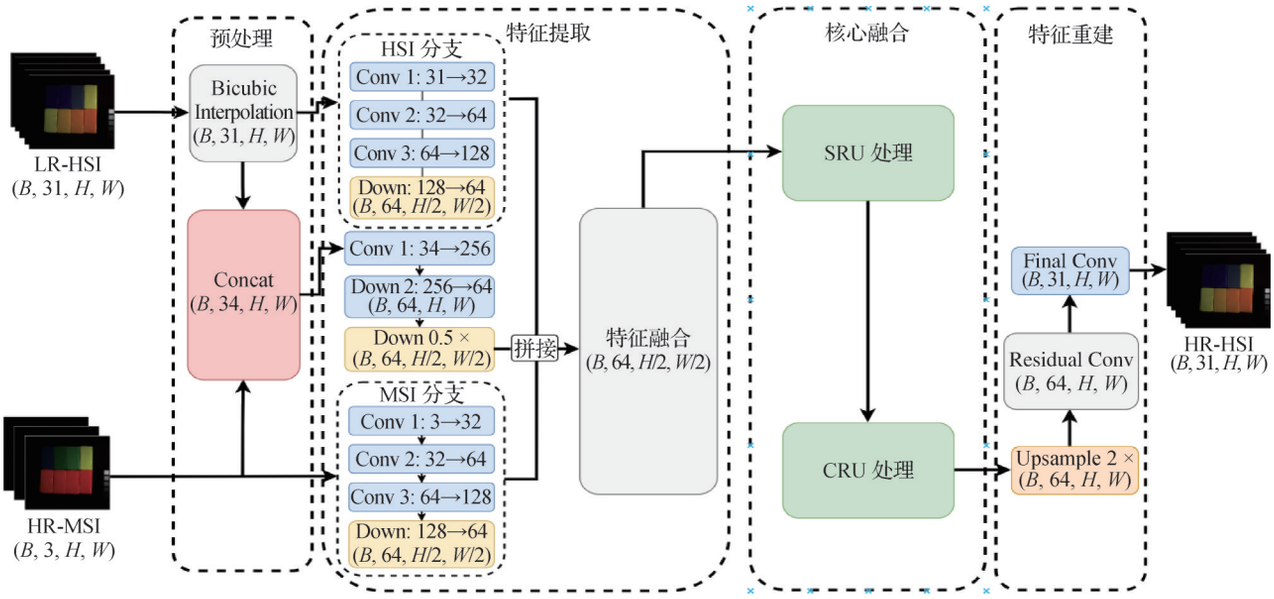


图 1　HMIF-SRNet 网络结构图

Fig. 1　HMIF-SRNet network structure diagram

HSI 和 MSI 数据。HSI 分支通过多层三维卷积提取光谱—空间联合特征。具体为

$$\boldsymbol{F}_{\mathrm{hsi}}^{l}=\phi\left(f_{\mathrm{3D}}\left(\boldsymbol{F}_{\mathrm{hsi}}^{l-1};\theta_{\mathrm{3D}}^{l}\right)+\boldsymbol{F}_{\mathrm{hsi}}^{l-1}\right) \tag{3}$$

MSI 分支专注于空间细节特征提取，具体为

$$\boldsymbol{F}_{\mathrm{msi}}^{l}=\phi\left(f_{\mathrm{2D}}\left(\boldsymbol{F}_{\mathrm{msi}}^{l-1};\theta_{\mathrm{2D}}^{l}\right)+\boldsymbol{F}_{\mathrm{msi}}^{l-1}\right) \tag{4}$$

式中，$f_{\mathrm{3D}}(\cdot;\theta_{\mathrm{3D}})$ 和 $f_{\mathrm{2D}}(\cdot;\theta_{\mathrm{2D}})$ 分别表示三维卷积和二维卷积操作，$\boldsymbol{\theta}$ 表示卷积核参数，$\phi(\cdot)$ 为 ReLU（rectified linear unit）激活函数，$l\in\{1,2,\cdots,L\}$ 为网络层索引，$\boldsymbol{F}_{\mathrm{hsi}}^{0}=\boldsymbol{X}_{\mathrm{up}}$，$\boldsymbol{F}_{\mathrm{msi}}^{0}=\boldsymbol{X}_{\mathrm{MS}}$。残差连接设计用于保持梯度流动并加速网络收敛。

每个分支的输出特征经过批归一化和维度对齐处理。具体为

$$\boldsymbol{F}_{\mathrm{hsi}}=\mathcal{A}\left(\mathcal{BN}\left(\boldsymbol{F}_{\mathrm{hsi}}^{L}\right)\right) \tag{5}$$

$$\boldsymbol{F}_{\mathrm{msi}}=\mathcal{A}\left(\mathcal{BN}\left(\boldsymbol{F}_{\mathrm{msi}}^{L}\right)\right) \tag{6}$$

式中，$\mathcal{BN}(\cdot)$ 表示批归一化操作，$\mathcal{A}(\cdot)$ 表示特征维度对齐操作，确保 $\boldsymbol{F}_{\mathrm{hsi}}$ 和 $\boldsymbol{F}_{\mathrm{msi}}$ 具有相同的空间和通道维度，以便后续融合。

初始融合特征通过简单拼接得到。具体为

$$\boldsymbol{F}_{\mathrm{merged}}=\left[\boldsymbol{F}_{\mathrm{hsi}};\boldsymbol{F}_{\mathrm{msi}}\right] \tag{7}$$

式中，$[\cdot;\cdot]$ 表示沿通道维度的拼接操作。

核心融合模块由两个核心组件构成：光谱重建单元（SRU）和交叉引导单元（CRU）。需要说明的是，图 1 展示的是网络的抽象处理流程，其中“SRU 处理”和“CRU 处理”代表核心融合阶段的主要功能模块。在具体实现中，SRU 通过动态稀疏重建机制优化光谱表示，CRU 通过交叉引导机制融合双模态特征，两个模块协同工作实现递进式特征优化，详细结构分别如图 2 和图 3 所示。

1.2　光谱重建单元（SRU）

SRU 采用动态稀疏重建策略，结构如图 2 所示。该单元的核心思想是利用光谱维度的内在稀疏特性，动态选择最具信息量的光谱通道进行重建。

SRU 首先对融合特征进行全局统计分析，通过全局平均池化和最大池化提取光谱通道的统计特征。具体为

$$\boldsymbol{g}_{\mathrm{avg}}=f_{\mathrm{GAP}}\left(\boldsymbol{F}_{\mathrm{merged}}\right) \tag{8}$$

$$\boldsymbol{g}_{\mathrm{max}}=f_{\mathrm{GMP}}\left(\boldsymbol{F}_{\mathrm{merged}}\right) \tag{9}$$

式中，$f_{\mathrm{GAP}}(\cdot)$ 和 $f_{\mathrm{GMP}}(\cdot)$ 分别表示全局平均池化和全局最大池化操作，$\boldsymbol{g}_{\mathrm{avg}}$、$\boldsymbol{g}_{\mathrm{max}}\in\mathbf{R}^{1\times1\times C}$，$C$ 为 $\boldsymbol{F}_{\mathrm{merged}}$ 的通道数。这种双重统计策略能够同时捕获光谱通道的平均响应强度和峰值响应特征，为后续的权重学习提供丰富的统计信息。

然后，通过多层感知机（multi-layer perceptron，MLP）学习光谱通道权重。具体为

$$\boldsymbol{W}_{\mathrm{spec}}=\sigma\left(f_{\mathrm{MLP}}\left(\left[\boldsymbol{g}_{\mathrm{avg}};\boldsymbol{g}_{\mathrm{max}}\right];\Theta\right)\right) \tag{10}$$

式中，$\sigma(\cdot)$ 为 sigmoid 激活函数，$f_{\mathrm{MLP}}(\cdot;\Theta)$ 表示多层感知机网络，Θ 为 MLP 的参数，$[\cdot;\cdot]$ 表示特征拼接，$\boldsymbol{W}_{\mathrm{spec}}\in\mathbf{R}^{1\times1\times C}$ 为输出的光谱通道权重矩阵。该

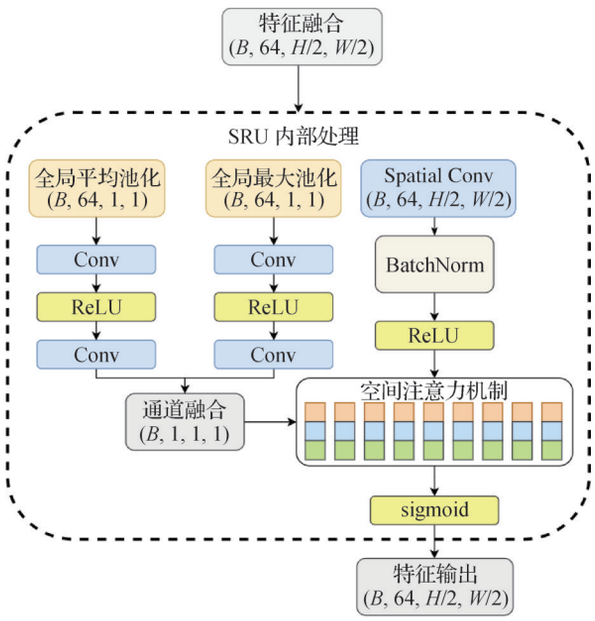


图2　光谱重建单元(SRU)

Fig. 2　Spectral reconstruction unit (SRU)

权重矩阵能够自适应地调节不同光谱通道的贡献程度,实现动态稀疏重建。

最后,结合空间门控机制实现自适应光谱重建,具体为

$$\boldsymbol{F}_{\mathrm{sru}}=\boldsymbol{F}_{\mathrm{merged}}\odot\boldsymbol{W}_{\mathrm{spec}}\odot\sigma\left(f_{\mathrm{conv}}\left(\boldsymbol{F}_{\mathrm{merged}};\theta_s\right)\right) \tag{11}$$

式中,$\odot$表示逐元素乘法操作,$f_{\mathrm{conv}}(\cdot;\theta_s)$表示卷积操作,用于生成空间门控图,$\theta_s$为卷积核参数。该式通过双重门控机制实现精细化的光谱重建:第1重门控$\boldsymbol{W}_{\mathrm{spec}}$控制光谱通道级别的权重分配;第2重门控$\sigma\left(f_{\mathrm{conv}}\left(\boldsymbol{F}_{\mathrm{merged}};\theta_s\right)\right)$实现空间位置级别的自适应调节。两个门控作用于原始融合特征$\boldsymbol{F}_{\mathrm{merged}}$,实现通道和空间双维度的自适应加权。这种多层次的门控策略确保了光谱信息的高效利用和精确保真。

1.3　交叉引导单元(CRU)

CRU通过交叉引导机制实现HSI和MSI特征的互补融合,结构如图3所示。该单元的设计理念是充分挖掘两种模态间的互补特性。如图3所示,CRU接收3个输入:来自特征提取模块的原始$\boldsymbol{F}_{\mathrm{hsi}}$和$\boldsymbol{F}_{\mathrm{msi}}\in\mathbf{R}^{B\times64\times H/2\times W/2}$(用于计算全局上下文),以及来自SRU的优化特征$\boldsymbol{F}_{\mathrm{stage1}}\in\mathbf{R}^{B\times64\times H/2\times W/2}$(用于后续融合)。

首先计算全局上下文特征,分别从原始HSI和MSI特征中提取全局描述符。具体为

$$\boldsymbol{c}_{\mathrm{hsi}}=f_{\mathrm{GAP}}\left(\boldsymbol{F}_{\mathrm{hsi}}\right) \tag{12}$$

$$\boldsymbol{c}_{\mathrm{msi}}=f_{\mathrm{GAP}}\left(\boldsymbol{F}_{\mathrm{msi}}\right) \tag{13}$$

式中,$\boldsymbol{c}_{\mathrm{hsi}}$、$\boldsymbol{c}_{\mathrm{msi}}\in\mathbf{R}^{B\times d\times1\times1}$,$d=64$为特征通道数。通过全局平均池化操作将空间维度($H\times W$)压缩为($1\times1$),得到每个通道的全局表示。这种全局上下文特征能够反映整个特征图的统计特性,为后续的交叉引导权重计算提供全局视角的信息。

通过两层线性变换和非线性激活生成交叉引导权重。具体为

$$\boldsymbol{W}_{\mathrm{cross}}=\sigma\left(f_{L2}\left(\phi\left(f_{L1}\left(\left[\boldsymbol{c}_{\mathrm{hsi}};\boldsymbol{c}_{\mathrm{msi}}\right];\boldsymbol{\Theta}_1\right)\right);\boldsymbol{\Theta}_2\right)\right) \tag{14}$$

式中,$f_{L1}(\cdot;\boldsymbol{\Theta}_1)$和$f_{L2}(\cdot;\boldsymbol{\Theta}_2)$分别表示第1层和第2层线性变换,$\boldsymbol{\Theta}_1$和$\boldsymbol{\Theta}_2$为对应的权重参数,$\phi(\cdot)$为ReLU激活函数,$\sigma(\cdot)$为sigmoid激活函数。该结构首先将拼接的全局特征$\left[\boldsymbol{c}_{\mathrm{hsi}};\boldsymbol{c}_{\mathrm{msi}}\right]\in\mathbf{R}^{B\times128\times1\times1}$通过特征融合降维至$\mathbf{R}^{B\times64\times1\times1}$,中间使用ReLU激活函数增加非线性表达能力,最后通过sigmoid函数将输出限制在[0,1]范围内,形成归一化的逐通道权重系数$\boldsymbol{W}_{\mathrm{cross}}\in\mathbf{R}^{B\times64\times1\times1}$。该权重系数反映了HSI和MSI两种模态在每个通道上的相对重要性。

在模块协同工作时,CRU利用从$\boldsymbol{F}_{\mathrm{hsi}}$和$\boldsymbol{F}_{\mathrm{msi}}$学习到的交叉引导权重$\boldsymbol{W}_{\mathrm{cross}}$,结合SRU输出的$\boldsymbol{F}_{\mathrm{stage1}}$和下采样后的原始MSI特征进行融合。具体为

$$\boldsymbol{F}_{\mathrm{cru}}=\boldsymbol{W}_{\mathrm{cross}}\odot\boldsymbol{F}_{\mathrm{stage1}}+\left(1-\boldsymbol{W}_{\mathrm{cross}}\right)\odot\boldsymbol{F}_{\mathrm{msi}} \tag{15}$$

此融合操作对经过SRU优化的特征$\boldsymbol{F}_{\mathrm{stage1}}$和原始MSI特征进行逐通道加权融合,权重由HSI和MSI的全局上下文自适应学习得到。

通过这种设计,CRU能够根据不同场景动态调整每个通道的融合策略,确保融合结果在光谱保真度和空间细节方面的最优平衡。最终输出$\boldsymbol{F}_{\mathrm{cru}}\in\mathbf{R}^{B\times64\times H/2\times W/2}$,如图3所示。

1.4　模块协同机制

两个核心模块通过精心设计的串行级联结构实现协同优化,形成递进式的特征增强链路。具体为

$$\boldsymbol{F}_{\mathrm{stage1}}=f_{\mathrm{SRU}}\left(\boldsymbol{F}_{\mathrm{merged}}\right) \tag{16}$$

$$\boldsymbol{F}_{\mathrm{stage2}}=f_{\mathrm{CRU}}\left(\boldsymbol{F}_{\mathrm{stage1}},\boldsymbol{F}_{\mathrm{hsi}},\boldsymbol{F}_{\mathrm{msi}}\right) \tag{17}$$

式中,$f_{\mathrm{SRU}}(\cdot)$和$f_{\mathrm{CRU}}(\cdot)$分别表示光谱重建单元和交叉引导单元的操作。

首先,通过SRU对初始融合特征$\boldsymbol{F}_{\mathrm{merged}}\in\mathbf{R}^{B\times64\times H/2\times W/2}$进行光谱域的动态稀疏重建。如图2所示,SRU采用分频处理策略,将特征分为高频和低频两个分支:高频分支通过空间注意力机制

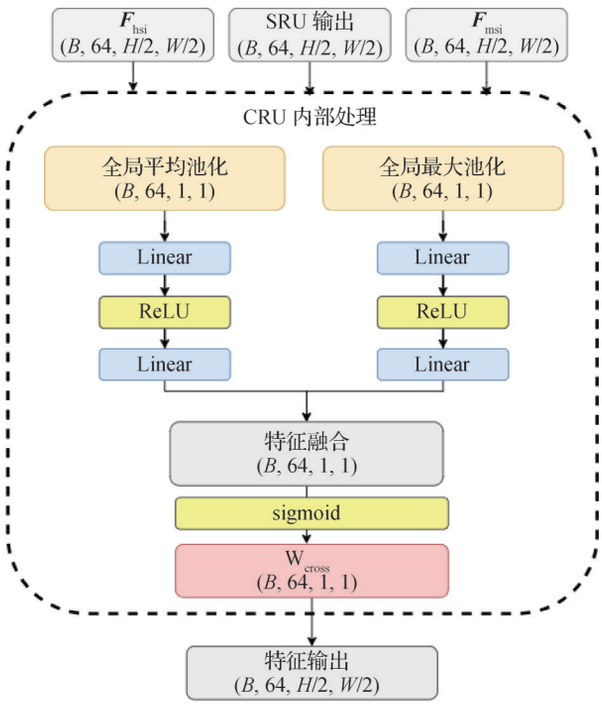


图3　交叉引导单元(CRU)结构

Fig. 3　Cross guidance unit (CRU) structure

和动态稀疏选择保留精细光谱信息；低频分支通过全局最大池化提取主要光谱结构。两个分支的输出经过动态融合和下采样，生成具有优化光谱表示的 $\boldsymbol{F}_{\text{stage1}} \in \mathbf{R}^{B\times 64\times H/2\times W/2}$。该阶段主要解决光谱通道间的冗余问题，通过稀疏重建机制自适应选择最具判别力的光谱特征，为后续处理提供高质量的光谱基础。

然后，CRU 接收 3 个输入实现跨模态的交叉引导融合：1）SRU 的输出特征 $\boldsymbol{F}_{\text{stage1}}$（已优化的光谱信息）；2）原始 HSI 特征 $\boldsymbol{F}_{\text{hsi}} \in \mathbf{R}^{B\times 64\times H/2\times W/2}$（保留的细节信息）；3）原始 MSI 特征 $\boldsymbol{F}_{\text{msi}} \in \mathbf{R}^{B\times 64\times H/2\times W/2}$（保留的空间信息）。

如图 3 所示，CRU 首先从 $\boldsymbol{F}_{\text{hsi}}$ 和 $\boldsymbol{F}_{\text{msi}}$ 中提取全局上下文，通过两层 MLP 生成逐通道的自适应权重 $\boldsymbol{W}_{\text{cross}} \in \mathbf{R}^{B\times 64\times 1\times 1}$。该权重反映了 HSI 和 MSI 两种模态在每个通道上的相对重要性。随后，CRU 利用学习到的权重对 $\boldsymbol{F}_{\text{stage1}}$ 和下采样后的 $\boldsymbol{F}_{\text{msi}}$ 进行加权融合，具体为

$$\boldsymbol{F}_{\text{stage2}} = \boldsymbol{W}_{\text{cross}} \odot \boldsymbol{F}_{\text{stage1}} + \left(1 - \boldsymbol{W}_{\text{cross}}\right) \odot \boldsymbol{F}_{\text{msi}} \tag{18}$$

最终，输出 $\boldsymbol{F}_{\text{stage2}} \in \mathbf{R}^{B\times 64\times H/2\times W/2}$，兼具 SRU 优化后的光谱。这种串行级联设计实现了两个层次的优化：1）光谱层面：SRU 通过稀疏重建消除冗余，提升光谱表达效率；2）模态层面：CRU 通过交叉引导融合双模态互补信息，增强特征表达能力。

两个模块的输出最终通过特征重建模块上采样至原始分辨率，完成端到端的融合过程。保证重建结果能够保留原始 MSI 的空间细节与光谱保真度。

1.5　特征重建模块

特征重建模块作为网络的最终输出层，负责将核心融合模块输出的高级语义特征转换为目标分辨率的高光谱图像。该模块采用渐进式重建策略，通过残差连接确保重建质量。

首先，进行特征融合与精炼，具体为

$$\boldsymbol{F}_{\text{recon}} = f_{\text{conv}}\left(\boldsymbol{F}_{\text{stage2}} + \boldsymbol{F}_{\text{res}}; \theta_r\right) \tag{19}$$

式中，$f_{\text{conv}}(\cdot; \theta_r)$ 表示卷积操作，θ_r 为卷积核参数，$\boldsymbol{F}_{\text{res}}$ 为来自浅层网络的残差特征，用于将网络浅层的细节信息与深层的语义信息相结合，防止重建过程中细节信息的丢失。

最终，通过 1 × 1 卷积生成高分辨率高光谱图像，具体为

$$\boldsymbol{Y} = f_{1\times 1}\left(\boldsymbol{F}_{\text{recon}}; \theta_{\text{out}}\right) \tag{20}$$

式中，$f_{1\times 1}(\cdot; \theta_{\text{out}})$ 表示 1 × 1 卷积操作，θ_{out} 为卷积核参数，$\boldsymbol{Y} \in \mathbf{R}^{H\times W\times c}$ 为最终重建的高分辨率高光谱图像。1 × 1 卷积核起到通道维度变换的作用，将特征通道数映射到目标光谱通道数 c，同时保持空间维度不变。这种轻量级的映射方式既保证了计算效率，又确保了光谱维度的精确重建。

1.6　损失函数

本文采用 L1 损失函数（也称为平均绝对误差（mean absolute error，MAE））训练高光谱图像超分辨率重建网络。L1 损失函数的计算方式为 $L = \frac{1}{N\times C\times H\times W}\sum_{i=1}^{N}\sum_{c=1}^{C}\sum_{h=1}^{H}\sum_{w=1}^{W}\left|y_{i,c,h,w} - \hat{y}_{i,c,h,w}\right|$ 式中，N 表示批次大小，C 表示光谱通道数，H 和 W 分别表示图像的高度和宽度，$y_{i,c,h,w}$ 表示真实高分辨率高光谱图像在位置 (i,c,h,w) 处的像素值，$\hat{y}_{i,c,h,w}$ 表示网络预测的重建图像在相应位置的像素值。

2　实　验

2.1　数据集

本文基于 CAVE 数据集和 Harvard 高光谱图像数据集评估 HMIF-SRNet 模型的融合性能。CAVE 数据集是高光谱图像融合领域的标准基准，包含

32个不同场景的高光谱图像，涵盖日常物体和材质，具有31个光谱波段（400～700 nm，光谱分辨率10 nm）和512×512像素空间分辨率。Harvard超光谱图像数据集是由美国哈佛大学计算机视觉实验室构建的真实场景高光谱图像数据库，包含27个不同场景的高光谱图像，采用人工或混合照明条件获取，具有31个光谱波段（420～720 nm，光谱分辨率10 nm）和1 392×1 040像素空间分辨率。

为了模拟实际应用场景中的数据退化情况，本文采用Wald协议构建训练和测试数据。具体而言，将原始高分辨率高光谱图像作为参考真值（ground truth），通过高斯模糊和下采样操作生成低分辨率高光谱图像，为了保证实验的一致性，两个数据集均采用相同的下采样因子$r=8$，确保训练和测试条件的统一，即CAVE数据集的LR-HSI分辨率为64×64×31，Harvard数据集的LR-HSI分辨率为174×130×31。高分辨率多光谱图像则通过将高光谱图像与CIE 1931标准色彩匹配函数进行卷积得到3通道RGB图像，保持与原始图像相同的空间分辨率。这种数据生成方式遵循了高光谱图像融合领域的标准实验协议，确保了实验结果的可比性和可重复性。

2.2 比较方法

为了全面验证所提方法的有效性，本文选取了涵盖传统方法和深度学习方法的10种代表性算法进行对比实验。具体包括：主成分分析（principal component analysis，PCA）（Turk和Pentland，1991）、强度色调饱和度融（intensity-hue-saturation fusion，IHS）（Starovoitov等，2007）、耦合非负矩阵分解（coupled non-negative matrix factorization，CNMF）（Yokoya等，2012）、空间卷积网络（SpatNet）（Wu等，2024）、光谱卷积网络（spectral convolutional neural network，SpecCNN）（Wang等，2020）、光谱残差网络（spectral residual network，SpecRNet）（Han等，2018）、光间融合卷积网络（spectral-spatial fusion convolutional neural network，SSFCNN）（Zhang等，2021）、光谱空间残差网络（spectral-spatial residual network，SSRNet）（Jia等，2023）、多尺度密集卷积网络（multi-scale dense convolutional neural network，MSDCNN）（Chen等，2021）和残差时频网络（residual time-frequency network，ResTFNet）。这些方法代表了当前高光谱与多光谱图像融合领域的先进技术，从不同角度探索了深度特征提取和模态融合策略。通过与上述涵盖不同技术路线和发展阶段的方法进行系统对比，可以客观评估HMIF-SRNet在融合质量、计算效率和泛化能力等方面的综合性能

2.3 评估指标

图像质量评估是衡量融合算法性能的核心指标，用于量化融合图像在光谱一致性、空间细节保持以及与真实图像相似性方面的表现，为了定量评估所提方法HMIF-SRNet及其与其他比较方法的性能，选择了峰值信噪比（PSNR），光谱角映射（SAM），相对全局无量纲误差（ERGAS），通用图像质量指数（UIQI）4个广泛使用的评估指标。

在这些指标中，ERGAS和SAM的值越低，结果质量越高；而PSNR和UIQI的值越高，结果质量越好。这些指标共同构成了一种多维度的评估框架，用于全面比较所提方法与现有方法的性能优劣。

2.4 模型配置

为保证公平比较，所有深度学习模型采用相同的训练配置，具体参数设置如下：传统方法中，PCA主成分数设为10，该参数通过保留累积方差贡献率达95%以上的主成分确定，既能捕获主要光谱信息，又避免过多冗余分量；CNMF端元数设为8，该值参考Yokoya等人（2012）的研究，适用于一般场景的物质分解，能够平衡解混精度和计算效率；IHS采用标准RGB到IHS色彩空间变换，无需额外参数调整。深度学习方法中，SSRNet采用8个3D残差块，特征通道数64，该配置来自Jia等人（2023）研究中的最优设置，在多个数据集上表现稳定；SpatNet采用6层2D卷积、特征通道数32，该配置参考Wu等人（2024）针对空间特征提取的推荐配置；SpecCNN使用卷积核大小为3的1D卷积，特征通道数64，该参数设置参考Wang等人（2020）的研究，适合光谱维度的特征学习；SpecRNet包含6个光谱残差块，特征通道数32，遵循Han等人（2018）的网络设计；SSFCNN采用多分支结构，光谱分支和空间分支各4层，特征通道数分别为64和32，该配置来自Zhang等人（2021）的消融实验最优组合；MSDCNN采用3个尺度分支，每个分支4层卷积，特征通道数64，该配置参考Chen等人（2021）的多尺度设计；ResTFNe包含6个残差块和2个Transformer层，特征维度128，该配置平衡了计算成本和融合性能。消融实验方法中，Baseline采用4层基础卷积层（3×3卷积核，特征通道数64），无SRU和CRU模块，用于验证核心模块的

有效性。

HMIF-SRNet参数设置和训练参数配置如表1和表2所示。

表1　HMIF-SRNet参数设置

Table 1　HMIF-SRNet parameter settings

参数	数值
输入光谱通道数	31(对应CAVE和Harvard数据集)
空间分辨率/像素	CAVE:64×64(LR-HSI), 512×512(HR-MSI) Harvard:174×130(LR-HSI), 1 392×1 040(HR-MSI)
下采样因子	$r=8$
多光谱通道数	3(RGB通道)
核心模块配置	SRU、CRU两个模块串行级联

表2　训练参数配置

Table 2　Training parameter configuration

训练参数	数值
优化器	Adam优化器,学习率设置为0.001
训练轮数	50个epoch,确保模型充分收敛
损失函数	均方误差(MSE)损失
权重保存	每10个epoch保存一次模型权重
随机种子	设置为0,确保实验可重现

2.5　实验结果

2.5.1　CAVE数据集

为了验证所提方法在真实数据上的有效性,本文在CAVE数据集上进行了全面的定量和定性评估实验。表3列出了HMIF-SRNet与10种对比方法在4个关键评估指标上的性能对比。从表3可以看出,HMIF-SRNet在PSNR、ERGAS、UIQI 3项指标上取得最优结果;SAM指标上,传统PCA、IHS、CNMF方法数值更低,但这类方法空间重建效果差,PSNR与UIQI指标显著落后。相比于深度学习对比方法中的次优模型SSRNet,本文方法SAM减少了32.0%,PSNR提升了4.41 dB(相对提升9.4%),ERGAS降低了9.5%,UIQI提升至0.983 2。这些显著的性能提升充分验证了动态稀疏重建策略和交叉引导融合机制的有效性,表明所提方法在光谱保真度和空间细节保持方面均优于现有技术。

图4和图5的可视化对比结果不仅在定性层面验证了定量指标分析的结论,更从直观的视觉角度

表3　多模型在CAVE数据集的实验数据对比

Table 3　Comparison of experimental data from multiple models on CAVE dataset

方法	PSNR/dB	ERGAS	SAM	UIQI
PCA	20.698 1	119.913 8	0.431 7	0.468 0
IHS	33.442 7	31.288 2	**0.208 1**	0.965 5
CNMF	31.947 7	36.996 5	0.406 8	0.955 1
SpecCNN	42.813 6	0.954 7	6.842 1	0.925 4
SpecRNet	42.941 4	0.853 6	5.314 6	0.934 3
SSFCNN	45.716 9	0.641 7	4.144 7	0.942 8
SSRNet	46.718 5	0.522 4	2.962 4	0.963 9
MSDCNN	43.785 6	1.097 3	7.212 2	0.865 2
ResTFNet	41.457 1	1.142 2	8.419 6	0.914 7
HMIF-SRNet(本文)	**51.124 4**	**0.472 6**	2.014 3	**0.983 2**

注:加粗字体表示各列最优结果。

展现了不同技术路线在实际应用中的效果差异。传统方法由于依赖简单的统计变换或矩阵分解,无法有效处理高光谱多光谱数据中复杂的非线性关系,导致在颜色保真度、空间细节保持和整体融合质量方面都存在明显不足。深度学习方法通过强大的特征学习能力和端到端优化策略显著改善了融合效果,但早期方法在网络架构设计和损失函数优化方面还存在局限性。HMIF-SRNet通过引入SRU模块的双重统计门控策略,自适应地优化光谱通道权重分配,抑制冗余光谱信息;CRU模块通过交叉引导机制实现HSI和MSI特征的自适应融合,充分挖掘两种模态的互补特性。两大核心模块协同作用,在可视化效果与数值精度上均实现显著提升,验证了动态稀疏重建策略和交叉引导融合机制的有效性。该模型能够降低光谱冗余、增强光谱—空间特征一致性,证明本文所提技术创新在理论先进性与实际应用效果上具备明显突破,为高光谱多光谱图像融合领域的深入研究与工程落地提供可靠技术方案。可视化对比结果可直观体现不同方法在该融合任务中的性能差异,通过3组代表性样本的融合图像与误差映射图综合分析,能够清晰展现从传统方法、深度学习方法到新型智能化方法的技术演进轨迹。

2.5.2　Harvard数据集

综合Harvard数据集分析表明,真实场景对算法鲁棒性提出更高要求。传统方法性能全面下降暴露

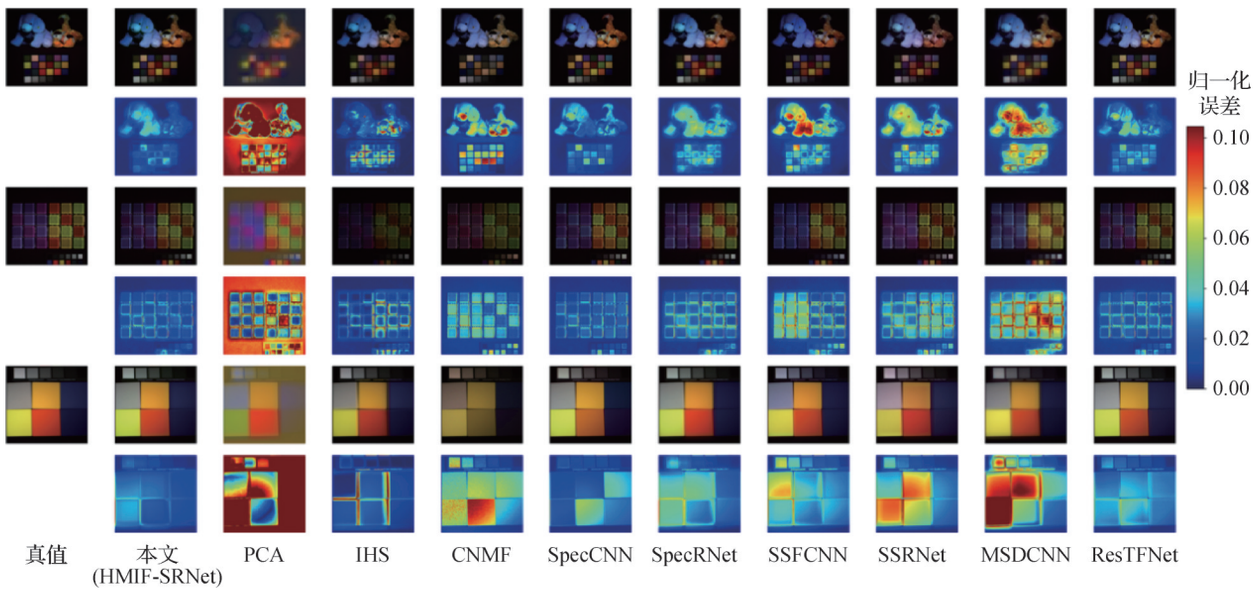


图4　模型检测结果对比(CAVE数据集)

Fig. 4　Comparison of model detection results(CAVE dataset)

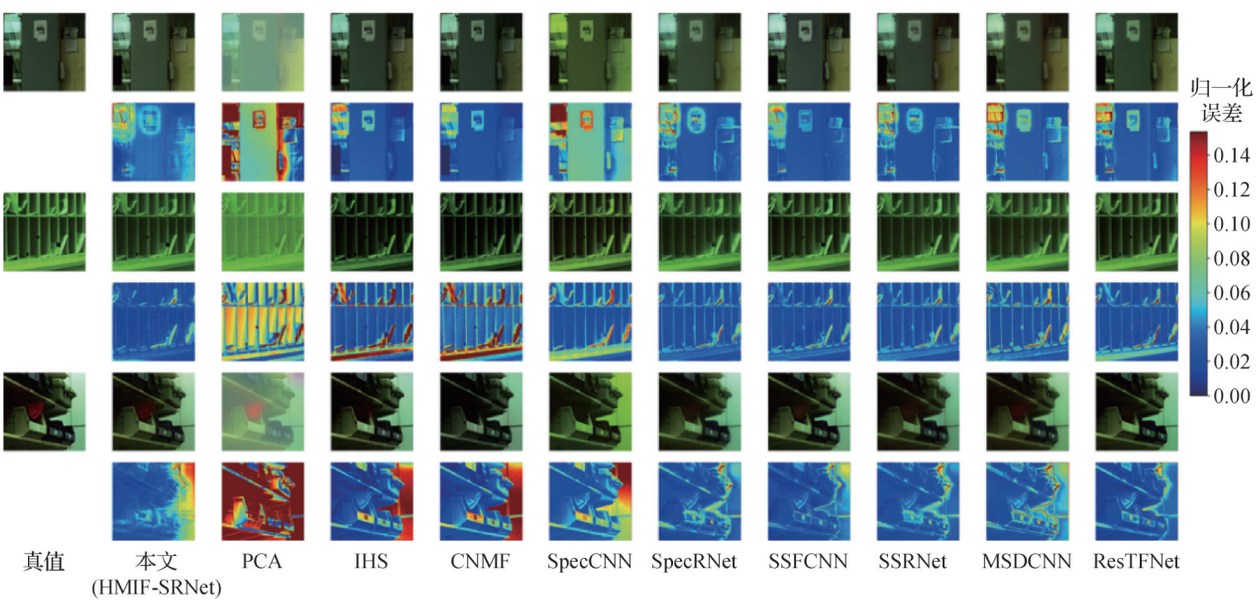


图5　模型检测结果对比(Harvard数据集)

Fig. 5　Comparison of model detection results(Harvard dataset)

其根本局限,而深度学习方法通过技术演进持续优化,如表4所示,HMIF-SRNet在PSNR、ERGAS和UIQI 3项指标上均达到最优;SAM指标上,传统PCA、IHS、CNMF方法数值更低,但这类方法PSNR显著偏低、UIQI明显落后,整体重建质量较差。在深度学习类对比方法中,本文方法SAM指标最优,相比于次优方法SSFCNN,ERGAS降低了4.44%,UIQI提升至0.942 2。这一突破验证了动态稀疏重建策略在真实场景中的技术价值,为高光谱多光谱融合的实际应用提供了可靠方案。

2.6　消融实验

2.6.1　CAVE数据集

为了深入验证HMIF-SRNet中核心模块的有效性,本节设计了系统性的消融实验。如表5所示,消融实验结果清晰地展现了各个模块组合对融合性能的贡献程度。从实验结果可以观察到,基础网络(Baseline)在所有指标上的表现均为最低,PSNR为50.127 3 dB,ERGAS为0.508 2,SAM为2.146 2,UIQI为0.981 1,这验证了核心创新模块的必要性。当引入光谱重建单元(SRU)和交叉引导单元(CRU)

表 4　多模型在 Harvard 数据集的实验数据对比

Table 4　Comparison of experimental data from multiple models on Harvard dataset

方法	PSNR/dB	ERGAS	SAM	UIQI
PCA	16.620 4	69.740 2	0.185 5	0.397 5
IHS	18.786 4	52.692 7	0.158 9	0.849 6
CNMF	16.984 9	120.718 4	0.290 7	0.788 2
SpecCNN	20.468 8	78.733 2	0.319 3	0.834 8
SpecRNet	24.853 7	40.517 8	0.197 4	0.920 4
SSFCNN	25.400 2	37.844 6	0.185 1	0.926 6
SSRNet	24.890 8	40.057 6	0.202 6	0.918 5
MSDCNN	22.712 7	41.317 4	0.219 9	0.731 2
ResTFNet	20.759 0	49.174 3	0.300 7	0.883 0
HMIF-SRNet(本文)	**25.729 6**	**36.165 5**	**0.152 6**	**0.942 2**

注：加粗字体表示各列最优结果。

表 5　在 CAVE 数据集的消融实验数据对比

Table 5　Comparison of ablation experimental data on CAVE dataset

模型	PSNR/dB	ERGAS	SAM	UIQI
Baseline	50.127 3	0.508 2	2.146 2	0.981 1
HMIF-SRNet(本文)	**51.124 4**	**0.472 6**	**2.014 3**	**0.983 2**

注：加粗字体表示各列最优结果。

后，HMIF-SRNet 模型在所有指标上均有显著提升，PSNR 提升至 51. 124 4 dB，ERGAS 降低至 0. 472 6，SAM 减少至 2. 014 3，UIQI 提升至 0. 983 2。这表明 SRU 的动态稀疏重建策略和 CRU 的交叉引导机制有效提升了光谱保真度和双模态信息融合质量。

2. 6. 2　Harvard 数据集

为了深入验证 HMIF-SRNet 在 Harvard 数据集上各个核心模块的有效性，设计了系统性的消融实验。表 6 所示为消融实验结果，从实验结果可以观察到，基础网络（Baseline）在 Harvard 数据集上的表现为 PSNR 为 26. 312 8 dB，ERGAS 为 35. 082 8，SAM 为 0. 191 3，UIQI 为 0. 933 4，这些指标反映了基础架构在处理 Harvard 数据集复杂光谱特性时的起始性能水平。当引入光谱重建单元（SRU）和交叉引导单元（CRU）后，HMIF-SRNet 模型在光谱角度映射和图像质量指标上均有显著提升。值得注意的是，虽然 PSNR 略有下降（从 26. 31 dB 降至 25. 73 dB），但这种权衡换来了光谱保真度的显著提升：ERGAS 改善至 36. 17，SAM 显著减少至 0. 15（降幅 20. 2%），UIQI 提升至 0. 94。这表明 HMIF-SRNet 在 Harvard 数据集上更加注重光谱信息的精确重建，这对于高光谱图像的实际应用（如物质识别、光谱分析）具有更重要的意义。PSNR 的轻微下降主要源于网络在平衡空间和光谱保真度时，优先保证了光谱维度的准确性。这表明 SRU 的动态稀疏重建策略和 CRU 的交叉引导机制在 Harvard 数据集上有效提升了光谱保真度，特别是在光谱角度一致性方面表现出色，验证了这两个模块在处理高光谱数据时的专业性和针对性。

表 6　在 Harvard 数据集的消融实验数据对比

Table 6　Comparison of ablation experimental data on Harvard dataset

模型	PSNR/dB	ERGAS	SAM	UIQI
Baseline	26.312 8	**35.082 8**	0.191 3	0.933 4
HMIF-SRNet（本文）	**25.729 6**	36.165 5	**0.152 6**	**0.942 2**

注：加粗字体表示各列最优结果。

图 6 和图 7 展示了消融实验的可视化对比结果，通过 3 个不同场景样本的融合图像和误差映射图，

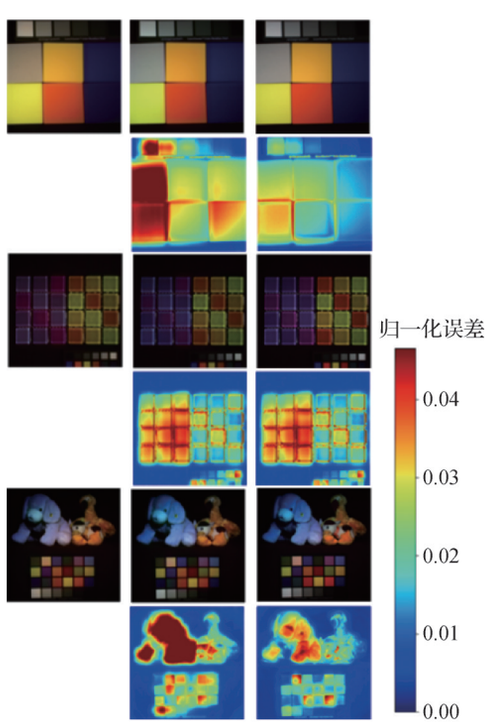


图 6　消融检测结果对比（CAVE 数据集）

Fig. 6　Comparison of ablation detection results (CAVE dataset)

直观地验证了各个核心模块的有效性。从可视化结果可以清晰观察到不同模块组合对融合质量的影响程度。

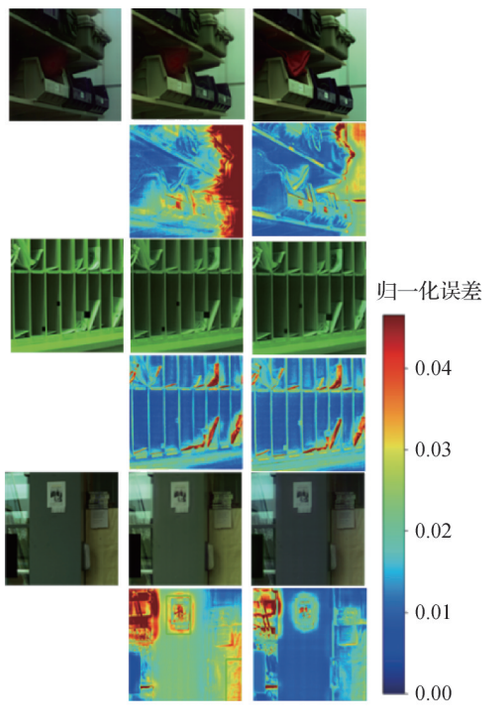


图7　消融检测结果对比(Harvard 数据集)

Fig. 7　Comparison of ablation detection results (Harvard dataset)

3　结　论

本文提出的稀疏重建与交叉引导的高光谱多光谱图像融合算法——HMIF-SRNet,有效解决了高光谱图像与多光谱图像融合中的多个核心挑战。通过系统性设计的端到端融合架构,HMIF-SRNet包含预处理、特征提取、核心融合(SRU和CRU)和特征重建4个阶段,其中SRU和CRU两个核心创新模块在光谱重建和模态融合方面克服了现有技术的局限性,显著提升了融合图像的空间分辨率与光谱保真度的平衡能力。首先,光谱重建单元(SRU)通过动态稀疏优化策略,自适应地增强了关键光谱通道的表达能力,抑制了冗余信息,实现在光谱维度上的高效重建;其次,交叉引导单元(CRU)采用双向注意力机制,深入挖掘HSI与MSI模态之间的互补性,有效缓解了空间细节与光谱一致性之间的矛盾。实验结果表明,HMIF-SRNet在多个公开数据集上获得了最优性能,尤其是在真实场景(如Harvard数据集)中展现出了卓越的鲁棒性。核心模块的协同作用不仅突破了PSNR、ERGAS、SAM等关键指标,提升了算法的实用性。

本文研究为后续进一步融入时序信息与多模态协同处理的工作奠定了坚实的算法基础,并为遥感图像融合技术的发展提供了重要参考。未来研究可以在此基础上,探讨如何将时序特征和更多类型的遥感数据相结合,进一步提升融合结果的综合性能。

参考文献(References)

Cao X Y, Fu X Y, Xu C and Meng D Y. 2022. Deep spatial-spectral global reasoning network for hyperspectral image denoising. IEEE Transactions on Geoscience and Remote Sensing, 60: #5504714 [DOI: 10.1109/TGRS.2021.3069241]

Chang Y, Yan L X, Zhao X L, Fang H Z, Zhang Z J and Zhong S. 2020. Weighted low-rank tensor recovery for hyperspectral image restoration. IEEE Transactions on Cybernetics, 50(11): 4558-4572 [DOI: 10.1109/TCYB.2020.2983102]

Chang Y, Yan L X and Zhong S. 2017. Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5901-5909 [DOI: 10.1109/CVPR.2017.625]

Chen L H, Vivone G, Qin J Y, Chanussot J and Yang X M. 2024. Spectral-spatial transformer for hyperspectral image sharpening. IEEE Transactions on Neural Networks and Learning Systems, 35(11): 16733-16747 [DOI: 10.1109/TNNLS.2023.3297319]

Chen N, Sui L C, Zhang B, He H J, Gao K, Li Y D, et al. 2021. Fusion of hyperspectral-multispectral images joining spatial-spectral dual-dictionary and structured sparse low-rank representation. International Journal of Applied Earth Observation and Geoinformation, 104: #102570 [DOI: 10.1016/j.jag.2021.102570]

Dong W S, Fu F Z, Shi G M, Cao X, Wu J J, Li G Y, et al. 2016. Hyperspectral image super-resolution via non-negative structured sparse representation. IEEE Transactions on Image Processing, 25(5): 2337-2352 [DOI: 10.1109/TIP.2016.2542360]

Fu Y, Zhang T, Zheng Y Q, Zhang D B and Huang H. 2022. Joint camera spectral response selection and hyperspectral image recovery. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(1): 256-272 [DOI: 10.1109/TPAMI.2020.3009999]

Han X H, Shi B X and Zheng Y Q. 2018. SSF-CNN: spatial and spectral fusion with CNN for hyperspectral image super-resolution//Proceedings of the 25th IEEE International Conference on Image Processing (ICIP). Athens, Greece: IEEE: 2506-2510 [DOI: 10.

1109/ICIP.2018.8451142]

He W, Yao Q M, Li C, Yokoya N and Zhao Q B. 2019. Non-local meets global: an integrated paradigm for hyperspectral denoising//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019). Long Beach, USA: IEEE: 6861-6870 [DOI: 10.1109/CVPR.2019.00703]

He W, Zhang H Y, Zhang L P and Shen H F. 2016. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. IEEE Transactions on Geoscience and Remote Sensing, 54(1): 178-188 [DOI: 10.1109/TGRS.2015.2452812]

Jia S, Min Z C and Fu X Y. 2023. Multiscale spatial-spectral transformer network for hyperspectral and multispectral image fusion. Information Fusion, 96: 117-129 [DOI: 10.1016/j. inffus. 2023. 03.011]

Kanatsoulis C I, Fu X, Sidiropoulos N D and Ma W K. 2018. Hyperspectral super-resolution: a coupled tensor factorization approach. IEEE Transactions on Signal Processing, 66(24): 6503-6517 [DOI: 10.1109/TSP.2018.2876362]

Liu H B, Feng C G, Dian R W and Li S T. 2024. SSTF-Unet: spatial-spectral transformer-based U-Net for high-resolution hyperspectral image acquisition. IEEE Transactions on Neural Networks and Learning Systems, 35(12): 18222-18236 [DOI: 10.1109/TNNLS. 2023.3313202]

Starovoitov V, Makarau A, Zakharov I and Dovnar D. 2007. Multispectral image enhancement based on fusion and super-resolution//Proceedings of the 15th European Signal Processing Conference. Poznan, Poland: IEEE: 2174-2178

Turk M A and Pentland A. 1991. Face recognition using eigenfaces//Proceedings of 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Maui, USA: IEEE: 586-591 [DOI: 10.1109/CVPR.1991.139758]

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS 2017). Long Beach, USA: MIT Press: 6000-6010 [DOI: 10.5555/3295222.3295349]

Wang Q, Li Q and Li X L. 2020. Spatial-spectral residual network for hyperspectral image super-resolution[EB/OL]. [2025-12-19]. https://arxiv.org/pdf/2001.04609.pdf

Wei W, Zhang L, Tian C N, Plaza A and Zhang Y N. 2017. Structured sparse coding-based hyperspectral imagery denoising with intracluster filtering. IEEE Transactions on Geoscience and Remote Sensing, 55(12): 6860-6876 [DOI: 10.1109/TGRS.2017.2735488]

Wu T Y, Li H N, Kumari S and Chen C M. 2024. A spectral convolutional neural network model based on adaptive Fick's law for hyperspectral image classification. Computers, Materials & Continua, 79(1): 19-46 [DOI: 10.32604/cmc.2024.048347]

Yang Y, Wang X Z, Liu X, Huang S Y, Liu Z Y and Wang S Z. 2025. Dual-branch U-Net with hybrid attention for hyperspectral pansharpening. Journal of Image and Graphics, 30(4): 989-1002 (杨勇, 王晓争, 刘轩, 黄淑英, 刘紫阳, 王书昭. 2025. 面向高光谱全色锐化的混合注意力双分支U型网络. 中国图象图形学报, 30(4): 989-1002) [DOI: 10.11834/jig.240410]

Yokoya N, Yairi T and Iwasaki A. 2012. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. IEEE Transactions on Geoscience and Remote Sensing, 50(2): 528-537 [DOI: 10.1109/TGRS.2011.2161320]

Yu D, Li K, Zhang W, Li D D, Tian X and Jiang H. 2023. Deep network-interpreted multispectral image fusion in remote sensing. Journal of Image and Graphics, 28(1): 290-304 (余典, 李坤, 张玮, 李对对, 田昕, 江昊. 2023. 可解译深度网络的多光谱遥感图像融合. 中国图象图形学报, 28(1): 290-304) [DOI: 10. 11834/jig.220575]

Zhang T, Liang Z Y and Fu Y. 2022. Joint spatial-spectral pattern optimization and hyperspectral image reconstruction. IEEE Journal of Selected Topics in Signal Processing, 16(4): 636-648 [DOI: 10. 1109/JSTSP.2022.3179806]

Zhang X T, Huang W, Wang Q and Li X L. 2021. SSR-NET: spatial-spectral reconstruction network for hyperspectral and multispectral image fusion. IEEE Transactions on Geoscience and Remote Sensing, 59(7): 5953-5965 [DOI: 10.1109/TGRS.2020.3018732]

Zhong Z L, Li Y, Ma L F, Li J and Zheng W S. 2022. Spectral-spatial transformer network for hyperspectral image classification: a factorized architecture search framework. IEEE Transactions on Geoscience and Remote Sensing, 60: #5514715 [DOI: 10.1109/TGRS. 2021.3115699]

作者简介

薛吉则，男，副教授，主要研究方向为高光谱图像处理、张量/矩阵建模和深度学习。E-mail：jize.xue@xupt.edu.cn

俞凤梅，女，硕士研究生，主要研究方向为高光谱与多光谱图像融合。E-mail：yufm1999@163.com

刘颖，女，教授，主要研究方向为图像处理与模式识别及其在公共安全领域的应用。E-mail：ly_yolanda@sina.com