

中图法分类号: TP751; TP391 文献标识码: A 文章编号: 1006-8961(2026)03-0944-14

论文引用格式: Yang H Y, Yang Y, Huang S Y, Liu Z Y and Zhang L. 2026. Dual-domain information fusion network for hyperspectral image super-resolution. Journal of Image and Graphics, 31(3):0944-0957(杨怀远, 杨勇, 黄淑英, 刘紫阳, 张龙. 2026. 面向高光谱图像超分辨率的双域信息融合网络. 中国图象图形学报, 31(3):0944-0957)[DOI:10.11834/jig.250325]

面向高光谱图像超分辨率的双域信息融合网络

杨怀远¹, 杨勇^{2,3*}, 黄淑英⁴, 刘紫阳⁵, 张龙¹

1. 天津工业大学电子与信息工程学院, 天津 300387; 2. 天津工业大学计算机科学与技术学院, 天津 300387;
3. 沧州市天津工业大学研究院, 沧州 061000; 4. 天津工业大学软件学院, 天津 300387;
5. 天津工业大学控制科学与工程学院, 天津 300387

摘要: 目的 高光谱图像超分辨率(hyperspectral image super-resolution, HISR)旨在融合高空间分辨率的多光谱图像(high-resolution multispectral image, HR-MSI)与低空间分辨率的高光谱图像(low-resolution hyperspectral image, LR-HSI),以生成具有高空间与光谱分辨率的高光谱图像(high-resolution hyperspectral image, HR-HSI)。现有方法在融合过程中往往忽略HR-MSI与LR-HSI之间在模态和结构上的差异,融合结果存在光谱扭曲与空间细节模糊等现象。针对上述问题,提出一种面向高光谱图像超分辨率的双域信息融合网络(dual-domain information fusion network, DDIF-Net),实现频域与空间域信息的联合建模与互补增强。**方法** 构建一个频率-空间特征融合模块(frequency-spatial feature fusion module, FSFFM),在频率域中通过频率特征注入模块(frequency-feature injection block, FIB)实现HR-MSI的幅值信息注入,并结合相位信息保持结构一致性;在空间域中利用空间特征增强模块(spatial feature enhancement block, SFEB)对局部与长程依赖进行建模。此外,提出多尺度特征融合模块(multi-scale feature fusion module, MFFM),进一步增强图像细节恢复能力。DDIF-Net在保持光谱一致性的同时,提升了空间结构的清晰度。**结果** 在Pavia Center数据集中,相比性能第2的方法,峰值信噪比(peak signal-to-noise ratio, PSNR)提升了0.23 dB,光谱角制图(spectral angle mapper, SAM)降低了0.24;Botswana数据集中,PSNR提升了0.41 dB,SAM降低了0.12;在Chikusei数据集中,PSNR提升了0.37 dB,SAM降低了0.01。**结论** 实验结果表明,所提出的DDIF-Net结构能更好地融合空间-光谱信息,显著提升高光谱图像超分辨率融合图像的质量。

关键词: 图像融合;遥感;深度学习;频率域信息;图像超分辨率

Dual-domain information fusion network for hyperspectral image super-resolution

Yang Huaiyuan¹, Yang Yong^{2,3*}, Huang Shuying⁴, Liu Ziyang⁵, Zhang Long¹

1. School of Electronics and Information Engineering, Tiangong University, Tianjin 300387, China; 2. School of Computer Science and Technology, Tiangong University, Tianjin 300387, China; 3. Cangzhou Institute of Tiangong University, Cangzhou 061000, China; 4. School of Software, Tiangong University, Tianjin 300387, China; 5. School of Control Science and Engineering, Tiangong University, Tianjin 300387, China

收稿日期:2025-07-21;修回日期:2025-09-02;预印本日期:2025-09-09

* 通信作者: 杨勇 yangyong@tiangong.edu.cn

基金项目: 国家自然科学基金项目(62072218);天津市自然科学基金项目(24JCZDJC00130,25JCZDJC00540);河北省自然科学基金项目(F2025110006,F2025110010);沧州市天津工业大学研究院项目(TGCIYY-Z-0303)

Supported by: National Natural Science Foundation of China(62072218); Tianjing Municipal Natural Science Foundation(24JCZDJC00130, 25JCZDJC00540); Natural Science Foundation of Hebei Province, China(F2025110006, F2025110010); Project of Cangzhou Institute of Tiangong University(TGCIYY-Z-0303)

Abstract: Objective Hyperspectral image super-resolution (HISR) has become a crucial task in the field of remote sensing, aiming to reconstruct high-resolution hyperspectral images (HR-HSI) that possess fine spatial details and rich spectral information. Given the inherent limitations of remote sensing imaging sensors, directly acquiring images with simultaneously high spatial and spectral resolutions remains technically challenging. Typically, only high-resolution multispectral images (HR-MSI) with limited spectral bands or low-resolution hyperspectral images (LR-HSI) with abundant spectral information are obtainable. However, relying solely on either data source is insufficient to meet the demands of practical applications such as land cover classification, urban analysis, and environmental monitoring. Therefore, fusing HR-MSI and LR-HSI to generate high-quality HR-HSI is a promising direction. Although existing methods have made some progress, most current deep learning-based HISR approaches primarily perform feature fusion in the spatial domain. Such methods often overlook the inherent modality differences between HR-MSI and LR-HSI, leading to spectral distortion and spatial structure blurring in the reconstructed images. Although some recent studies have introduced frequency-domain information, existing frequency-domain fusion methods typically treat frequency features as auxiliary enhancements without deeply modeling their internal structures. To overcome these limitations, this paper proposes a novel HISR network that jointly models the interactions between HR-MSI and LR-HSI across both spatial and frequency domains. The proposed method aims to mitigate modality differences, enhance feature complementarity, and effectively utilize multiscale and long-range dependency features to improve the accuracy, fidelity, and robustness of the reconstructed images. **Method** To address the limitations of existing HISR techniques, we propose a novel dual-domain information fusion network (DDIF-Net), which simultaneously models spatial and frequency information in a unified architecture. The network first extracts multiscale features from HR-MSI through a convolutional encoder path. Meanwhile, LR-HSI is upsampled to the resolution of HR-MSI and combined with it to produce shallow fused features. These features are fed into a U-shaped dual-domain fusion branch composed of a series of frequency-spatial feature fusion modules (FSFFMs). Each FSFFM integrates a frequency-feature injection block, which decomposes input features into amplitude and phase components via Fourier transform. The amplitude components, representing spectral intensity, are enhanced through injection and modulation mechanisms, while phase components, containing structural information, are fused and refined. The inverse Fourier transform then reconstructs the frequency-enhanced feature maps. Each FSFFM is followed by a spatial feature enhancement block, which leverages attention mechanisms to model long-range dependencies and reinforce spatial coherence, to further enhance the spatial structure. In the decoder path, a multiscale feature fusion module (MFFM) aggregates features from multiple scales and progressively refines them to reconstruct the HR-HSI. The network is trained using an L1 loss function, ensuring spectral accuracy and spatial consistency. The architecture is fully convolutional and optimized using the adaptive moment estimation optimizer with a two-stage learning rate decay strategy. **Result** We evaluate DDIF-Net on three benchmark hyperspectral remote sensing datasets: Pavia Center, Botswana, and Chikusei. Compared with eight representative deep learning-based models (DARN, HSRNet, HSRNet, HyperPNN, SSFCNN, SSRNet, HyperDSNet, HyperRefiner, and PPFNet), our method consistently achieves superior performance. On the Pavia Center dataset, DDIF-Net improves the peak signal-to-noise ratio (PSNR) by 0.23 dB and reduces the spectral angle mapper (SAM) by 0.24 compared with the second-best method. On the Botswana dataset, our approach gains a PSNR improvement of 0.41 dB and reduces SAM by 0.12. On the Chikusei dataset, we observe a PSNR improvement of 0.37 dB and a SAM reduction of 0.01. Visual inspection of reconstructed images further confirms our method's ability to preserve fine-grained texture and maintain spectral integrity, particularly in urban regions with complex spatial layouts. Moreover, we conduct ablation studies to demonstrate the contribution of each module within DDIF-Net. Removing the frequency-domain module or the multiscale fusion strategy results in significant performance drops, indicating their crucial role in the network's effectiveness. Finally, classification experiments using the k-means algorithm in the Environment for Visualizing Images software show that the fused HR-HSI generated by DDIF-Net leads to the most accurate classification results across all methods, further validating the quality of its spatial-spectral fusion. **Conclusion** This paper presents DDIF-Net, a dual-domain fusion network designed for hyperspectral image super-resolution, which integrates spatial and frequency information through a novel architecture composed of FSFFM and MFFM modules. The proposed network explicitly models amplitude and phase characteristics to bridge the modality gap between LR-HSI and HR-MSI, while enhancing spatial details via long-range attention mechanisms. Exten-

sive experiments on multiple datasets confirm that DDIF-Net outperforms state-of-the-art methods in objective metrics and visual quality, achieving significant gains in PSNR, SAM, and classification accuracy. The ablation studies validate the importance of dual-domain modeling and multiscale integration, highlighting the robustness and generalizability of our approach. In future work, we aim to optimize the network's computational efficiency and extend it to larger-scale satellite datasets, such as Sentinel-2 or Gaofen series, to further explore its applicability in real-world remote sensing scenarios.

Key words: image fusion; remote sensing; deep learning; frequency domain information; image super-resolution

0 引言

在遥感领域,由于硬件限制,不同传感器仅能分别捕获含有稀疏光谱信息的高空间分辨率的多光谱图像(high-resolution multispectral image, HR-MSI)和低空间分辨率的高光谱图像(low-resolution hyper spectral image, LR-HSI),单独一种类型的图像由于光谱信息或者空间信息的缺乏不能满足实际应用的需求。基于融合的高光谱图像超分辨率(hyperspectral image super-resolution, HISR)技术旨在融合这两类图像的互补信息,生成既具有高空间分辨率又含丰富光谱信息的 HR-HSI (high-spatial-resolution hyper-spectral image)。基于融合的 HISR 方法主要依赖于基于优化的模型重建 HR-HSI 图像,但这类方法通常伴随着高计算成本和复杂的参数设置问题(Zhang 等, 2022)。

传统的 HISR 方法可大致归为 3 类: PAN 锐化 (PAN-sharpening) 方法、基于贝叶斯推理的方法和基于分解的方法。PAN 锐化方法是一种传统的图像融合方法,通过将多光谱数据与相应的高分辨率 PAN 图像融合来增强 MSI 的空间分辨率。PAN 锐化方法可以大致分为 4 类: 成分替换 (component substitution, CS) (Restaino 等, 2021)、多分辨率分析 (multi resolution analysis, MRA) (Vivone 等, 2014)、变分方法 (Huang 等, 2018; Kawakami 等, 2011) 和混合方法 (Liao 等, 2015; Licciardi 等, 2014; 黄渝萍和李伟生, 2023; 于钺 等, 2010)。由于 PAN 锐化被视为高光谱—多光谱融合问题的特例, 因此一些研究将现有的 PAN 锐化方法推广到高光谱—多光谱融合任务中 (Tseng 等, 2001; Zhang 和 He, 2007)。Zhuo 等人 (2022) 构建了一种高光谱全色锐化网络架构 (hyper-DSNet), 利用多尺度卷积模块和光谱注意力模块, 充分保留空间细节和光谱保真度。Zhou 等人 (2023) 通过使用双注意力块和精细注意力块级联,

以完成特征的提取和融合, 有效地保留高光谱图像的边缘光谱信息。然而, 其性能在很大程度上取决于光谱重采样方法, 这导致了提高所有光谱波段图像空间分辨率的困难。Zhang 等人 (2009) 首次提出基于贝叶斯融合的方法, 这种方法的融合过程在小波域中进行, 称为小波 MAP (maximum a posteriori), 通过对 HSI 的加性噪声成像模型进行假设, 使用插值作为先验来避免对空间退化算子的估计, 并在小波域对增强图像进行近似, 实现对高分辨率图像的最优估计。Zeng 等人 (2021) 将 HSI 划分为局部重叠的三维小块, 在局部范围内引入低秩张量恢复方法, 有效分离出具有低秩结构的干净图像块与复杂噪声, 去除 HIS 中多种类型的噪声, 达到 HIS 超分辨率效果。虽然传统方法已经在 HISR 领域取得一定成果, 但这些方法仍存在许多缺陷。例如, PAN 锐化方法容易产生光谱失真 (Peng 等, 2023); 基于贝叶斯推理的方法在处理不同高光谱数据结构时具有较弱的灵活性 (Wei 等, 2015); 基于矩阵分解的技术将三维 HSI 和 MSI 重构为矩阵, 因此面临学习空间和光谱之间关系的挑战 (Yokoya 和 Iwasaki, 2014)。虽然基于张量的方法能够保持输入图像三维结构, 但它们需要更多的内存和更强大的硬件计算能力 (Yuan 等, 2020)。

卷积神经网络 (convolutional neural network, CNN) 的出现允许将融合问题看做是一个高度非线性映射。CNN 将 LR-HSI 和 HR-MSI 作为输入, 自动学习 HSI 与 MSI 两种图像之间的端到端映射, 求解得到最优的 HR-HIS (He 等, 2019; Zhang 等, 2018; Han 等, 2018; Zhang 等, 2021; Yu 等, 2020)。Yang 等人 (2018) 使用双分支 CNN 从 LR-HSI 和 HR-MSI 中分别提取光谱和空间特征, 然后通过卷积层和全连接层将特征信息进行融合, 使卷积神经网络在 HISR 领域中取得了巨大成功。Dian 等人 (2018) 提出一种深度 HSI 锐化方法, 通过基于深度卷积神经网络 (deep convolutional neural network, DCNN) 的残差学

习对图像先验进行自动学习,得到了高质量的融合结果。Liu 等人(2020)将浅—深卷积网络和基于光谱判别的细节注入模型结合起来,显著提高了PAN锐化性能。Hu 等人(2022a)利用注意力机制和 pixelShuffle 模块从上采样的 HSI 中提取信息,进一步提升了特征提取效果。然而,这些方法仅通过卷积操作对局部特征进行学习,由于卷积核感知域大小的限制,基于 CNN 的方法在捕捉长距离依赖关系和自相似先验方面存在局限性。

近年来,由于 Transformer 架构(Zheng 等, 2020; 胡帅 等, 2024; Li 等, 2023; 李焱 等, 2025; Zhang 等, 2025a)在计算机视觉任务中的成功应用,研究人员开始尝试通过建立 HSI 和 MSI 之间的长距离依赖关系提高 HISR 的性能。Hu 等人(2022b)针对 HISR 任务设计了一种基于 Transformer 的架构,对特征内部的全局关系进行探索和利用。Zhang 等人(2023)提出一种语言感知领域泛化网络 LDGnet (language-aware domain generalization network),网络采用双流架构,分别提取视觉和语言特征,并通过监督对比学习实现视觉与语言特征在该语义空间中的对齐。Chen 等人(2024)通过探索光谱和空间的长程依赖性提出光谱—空间 Transformer (spectral-spatial transformer, SST),证明了 Transformer 在高光谱—多光谱融合中的潜力。然而,这些方法专注于学习长距离特征依赖关系,并未充分考虑两种不同图像数据之间的特征差异。

目前已有工作中,小波变换、多分辨率傅里叶变换、频谱注意力等方法广泛应用于图像超任务中。Zhao 等人(2025)提出了 CFFormer (cross-fusion transformer framework)超分辨率框架,通过设计交叉特征通道注意模块 (cross feature channel attention, CFCA) 优化频域中的特征,同时还集成了全局注意力块以增强全局信息捕获,从而提高特征利用率。Zhang 等人(2025a)设计了一个跨频增强空间—频谱融合网络 CESFusion (cross-frequency enhanced spatial-spectral fusion network),网络通过傅里叶变换进行全局特征分析,有效地集成和增强融合图像的频域信息,实现空间域和频域之间的充分交互。但是这些方法大多直接将频域特征作为补充信息与空间特征进行拼接或通道加权,缺乏针对频域内特征内部结构(如相位和幅值)的有效表达与利用。

基于上述分析,本文提出一种面向高光谱图像

超分辨率的双域信息融合网络 (dual-domain information fusion network, DDIF-Net),该网络通过构建空—频双域融合分支在空间域和频率域来实现对 HR-MSI 和 LR-HSI 不同尺度特征的融合。其中,一个由频率特征注入块 (frequency-feature injection block, FIB) 和空间特征增强块 (spatial feature enhancement block, SFEB) 构成的频率—空间特征融合模块 (frequency-spatial feature fusion module, FSFFM) 被设计用于促进空间域和频率域之间的互补学习,充分探索和利用了双域中的信息,提升了高频细节的重建能力,增强了低频结构的稳定性,实现了高质量的融合图像重建。另外,通过构建多尺度特征融合模块 (multi-scale feature fusion module, MFFM) 在最终的融合特征重建阶段提出多尺度特征逐步融合策略,有效地集成来自不同尺度的信息,从而进一步提升特征融合效率和重建 HR 图像的质量。

本研究的贡献如下:1)针对遥感 HSI 超分辨率重建任务构建了遥感 HISR 网络 DDIF-Net,该网络将来自 HR-MSI 的不同尺度的细节特征注入到 LR-HSI 特征中,重建出具有丰富结构信息的 HR-HSI。2)设计一个具有编解码结构的空—频双域融合分支,通过构建 FSFFM 实现频率域和空间域的特征融合,更好地表示图像的潜在特征,提升模型特征的提取和融合能力。3)设计一个多尺度特征融合模块 (MFFM),有效实现对空—频双域分支重建的不同尺度特征的融合,以重建精细的高分辨空间的融合特征。

1 方 法

1.1 整体框架概述

为更好地将 HR-MSI 中的空间细节信息注入到 LR-HSI 中,本文提出 DDIF-Net,结构如图 1 所示。该网络通过学习 HR-MSI 和 LR-HSI 多尺度特征的上下文信息并对其进行融合来实现 HR-HSI 特征的重建。DDIF-Net 由一个多尺度 MS 特征提取分支、一个多尺度特征融合分支,以及一个多尺度融合模块构成。

1.1.1 MS 特征提取分支

该分支由两个 3×3 的卷积层和一系列的降采样和上采样操作构成,以提取 HR-MSI 不同尺度的特征。具体实现可表示为

$$F_{\text{MSI}}^1 = \text{Conv}_{3 \times 3}(\text{Conv}_{3 \times 3}(I_{\text{HM}})) \quad (1)$$

$$F_{\text{MSI}}^i = \text{Conv}_{2 \times 2}^2(F_{\text{MSI}}^{i-1}), \quad i = 2, 3 \quad (2)$$

$$F_{\text{MSI}}^i = \text{Conv}_{2 \times 2}^2(F_{\text{MSI}}^{i-1}), \quad i = 4, 5 \quad (3)$$

式中, $I_{\text{HM}} \in \mathbb{R}^{C \times H \times W}$ 代表输入的 HR-MSI, C 表示 MSI 的通道数。 $\text{Conv}_{3 \times 3}(\cdot)$ 表示 3×3 的卷积操作。 $\text{Conv}_{2 \times 2}^2(\cdot)$ 表示步长为 2 的 2×2 卷积, 以实现特征图的下采样, 并应用深度可分离卷积层增加特征图的通道数量。 $\text{Conv}_{2 \times 2}^2(\cdot)$ 表示步长为 2 的 2×2 转置卷积层, 实现特征图的上采样和通道压缩。 $F_{\text{MSI}}^i \in \mathbb{R}^{S \times H \times W}$

表示不同尺度的特征。网络中不同尺度特征图大小分别为 $H \times W \times S$, $H/2 \times W/2 \times 2S$ 和 $H/4 \times W/4 \times 4S$ (S 代表输入特征图的通道数, H 和 W 分别代表输入特征图的高度和宽度)。MS 特征提取分支得到的不同尺度特征图中, F_{MSI}^1 和 F_{MSI}^5 的尺寸为 $H \times W \times S$, F_{MSI}^2 和 F_{MSI}^4 的尺寸为 $H/2 \times W/2 \times 2S$, F_{MSI}^3 的尺寸为 $H/4 \times W/4 \times 4S$ 。此外, 为了防止信息丢失和解决梯度消失问题, 特征 F_{MSI}^1 和 F_{MSI}^2 分别与 F_{MSI}^5 和 F_{MSI}^4 进行相加整合, 使网络能够更好地利用 HR-MSI 的特征。

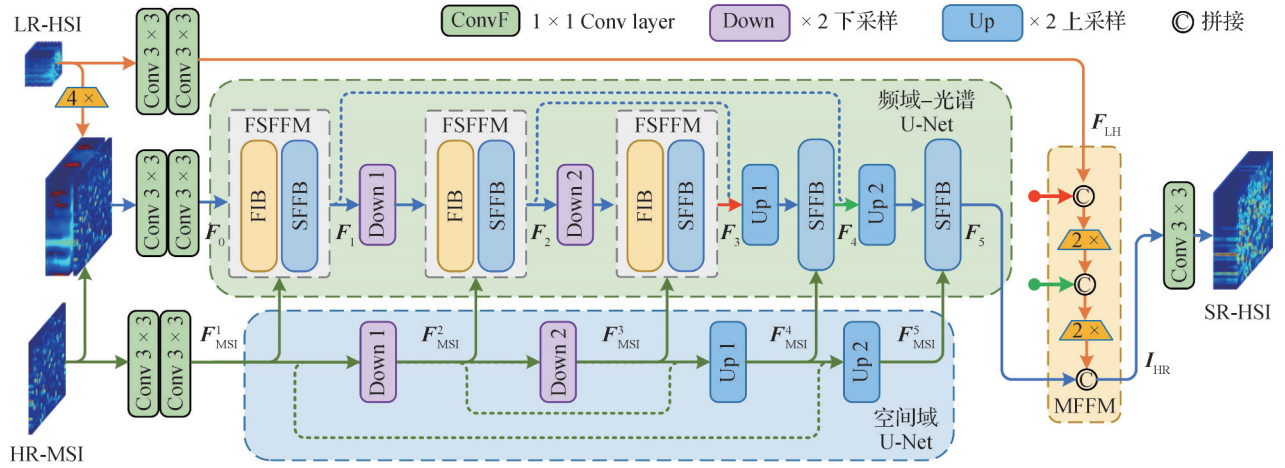


图1 DDIF-Net 整体结构

Fig. 1 Overall architecture of DDIF-Net

1.1.2 空-频双域融合分支

该分支是一个由编码阶段和解码阶段构成的 U 型网络。编码阶段的每个尺度层采用一个 FSFFM 实现频率域和空间域中 HR-MSI 与 LR-HSI 的不同尺度特征的融合。解码阶段中每个尺度层包含一个 SFEB, 其利用 HR-MSI 特征对融合特征进行增强, 以实现不同尺度融合特征的重建。具体操作如下:

在编码阶段, 对 LR-HSI 执行 PixelShuffle 操作, 将 LR-HSI 上采样 4 倍, 使其空间尺寸与 HR-MSI 匹配。然后, 上采样的 LR-HSI 与 HR-MSI 进行级联, 并送入两个 3×3 卷积层以获得两种图像的混合浅层特征 F_0 。再将浅层特征 F_0 与来自 HR-MSI 的特征 F_{MSI}^1 作为第 1 尺度层中 FSFFM 的输入实现当前尺度下两种特征的融合。相继地, 第 2 尺度层和第 3 尺度层执行与第 1 尺度层中相同的操作。编码阶段的具体可表示为

$$F_0 = \text{Conv}_{3 \times 3}(\text{Conv}_{3 \times 3}(PS(I_{\text{LH}}), I_{\text{HM}})) \quad (4)$$

$$F_1 = \text{FSFFM}(F_0, F_{\text{MSI}}^1) \quad (5)$$

$$F_i = \text{FSFFM}(D_n(F_{i-1}), F_{\text{MSI}}^i), \quad i = 2, 3 \quad (6)$$

式中, $I_{\text{LH}} \in \mathbb{R}^{C \times H \times W}$ 代表输入的 LR-HIS, $PS(\cdot)$ 表示 PixelShuffle, $\text{FSFFM}(\cdot)$ 表示频率-空间特征融合。 $F_i \in \mathbb{R}^{C' \times H \times W}$ 表示多尺度特征融合分支中第 i 个尺度层的输出特征, C' 表示特征图的通道数。

在解码阶段的每个尺度层中, 一个 SFEB 被构建, 其利用对应尺度的 HR-MSI 特征增强每个尺度的融合特征。另外, 为了减少特征的丢失, 编码阶段的特征被传递到解码阶段中的对应尺度层以实现特征的补充和增强, 从而重建不同尺度的融合特征。解码阶段的具体操作可表示为

$$F_i = \text{FEB}(Up(F_{i-1}), F_{\text{MSI}}^i) + F_{5-i+1}, \quad i = 4, 5 \quad (7)$$

式中, $Up(\cdot)$ 表示上采样, $\text{FEB}(\cdot)$ 表示特征增强块。

1.1.3 多尺度特征融合模块(MFFM)

为了避免光谱特征的失真, 本文构建一个多尺度融合模块, 将来自 LR-HSI 的浅层特征与来自多尺度融合分支的不同尺度特征实现渐进式的特征整合, 以获得精细化的融合特征。这里的上采样操作

采用了步长为2的 2×2 转置卷积层,以实现特征图尺寸的放大以及特征的整合。最后的融合特征被送入一个 3×3 卷积层,以获得重建的HR图像。具体可表示为

$$\mathbf{F}_{\text{LH}} = \text{Conv}_{3 \times 3}(\text{Conv}_{3 \times 3}(\mathbf{I}_{\text{LH}})) \quad (8)$$

$$\mathbf{I}_{\text{HR}} = \text{Cat}\left(\text{ConvT}_{2 \times 2}^T\left(\text{Cat}\left(\text{ConvT}_{2 \times 2}^T\left(\text{Cat}(\mathbf{F}_{\text{LH}}, \mathbf{F}_3)\right), \mathbf{F}_4\right), \mathbf{F}_5\right)\right) \quad (9)$$

$$\hat{\mathbf{I}}_{\text{HR}} = \text{Conv}_{3 \times 3}(\mathbf{I}_{\text{HR}}) \quad (10)$$

式中, $\mathbf{F}_{\text{LH}}$ 表示从LR-HSI中提取的浅层特征,

$\hat{\mathbf{I}}_{\text{HR}} \in \mathbf{R}^{C \times H \times W}$ 表示最终的融合结果。U2Net通过实现多尺度的空间—光谱特征的有效融合,极大地增强了模型的超分辨率重建性能,并能获得具有空间和光谱双保真的融合结果。

1.2 频率—空间特征融合模块(FSFFM)

现有的基于深度学习HISR方法大都是直接在空间域执行LR-HSI和HR-MSI特征的融合,由于两种图像特征存在模态差异,容易导致重建的HR-HSI出现空间和光谱信息的扭曲。为克服这一问题,本文构建了一个频率—空间特征融合模块,该模块由一个FIB和一个SFEB构成,结构如图2所示。

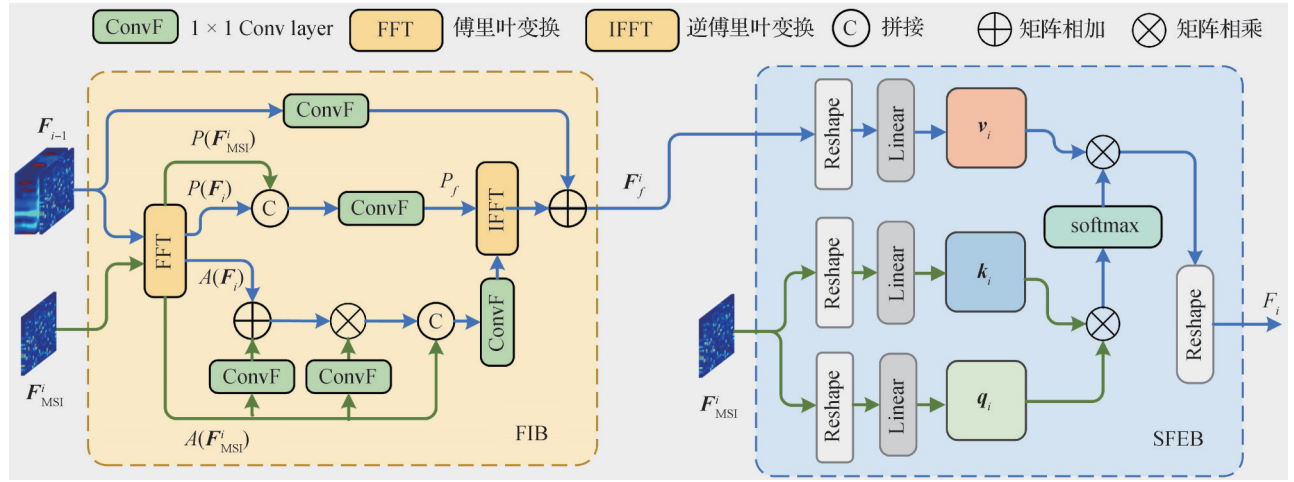


图2 FSFFM内部结构(FIB结构,SFEB结构)

Fig. 2 Internal structure of FSFFM(FIB structure,SFEB structure)

1.2.1 频率特征注入块(FIB)

在FIB中,为了减弱LR-HSI和HR-MSI特征间的模态差异性,傅里叶变换首先用于将输入的融合特征与HR-MSI特征进行特征分解,获得两种特征的幅值分量 $\{A(\mathbf{F}_i), A(\mathbf{F}_{\text{MSI}}^i)\}$ 和相位分量 $\{P(\mathbf{F}_i), P(\mathbf{F}_{\text{MSI}}^i)\}$,后续操作则实现对两种分量特征的融合。该分解操作可表示为

$$\{A(\mathbf{F}_i), P(\mathbf{F}_i)\} = \text{FFT}(D_n(\mathbf{F}_{i-1})) \quad (11)$$

$$\{A(\mathbf{F}_{\text{MSI}}^i), P(\mathbf{F}_{\text{MSI}}^i)\} = \text{FFT}(\mathbf{F}_{\text{MSI}}^i) \quad (12)$$

式中, $\text{FFT}(\cdot)$ 表示傅里叶变换。

相位分量主要表征图像的高频信息,包含图像的细节和结构特征,而幅值分量则主要表征图像的低频信息,包含图像的光谱信息。由于两种图像的模态特征差异性主要体现在光谱特征上,因此对于两种图像特征的相位分量,本文直接将它们进行合

并,并通过一个 1×1 的卷积操作来实现两种分量的整合。而对于两种图像特征的幅值分量,为了更好地融合光谱特征,首先将HR-MSI的幅值分量通过一个 1×1 的卷积,并通过一个加法操作将其与LR-HSI特征的幅值分量进行融合;然后将HR-MSI的幅值分量与融合的幅值进行相乘,以进一步实现特征的增强;增强后的幅值分量 $\mathbf{EA}(\mathbf{F}_i)$ 与HR-MSI的幅值分量进行合并,并通过一个 1×1 的卷积获得融合的幅值分量 $\mathbf{A}_f$ 。上述过程可表达为

$$\mathbf{P}_f = \text{Conv}_{1 \times 1}(\text{Cat}(P(\mathbf{F}_i), P(\mathbf{F}_{\text{MSI}}^i))) \quad (13)$$

$$\mathbf{EA}(\mathbf{F}_i) = (A(\mathbf{F}_i) + \text{Conv}_{1 \times 1}(A(\mathbf{F}_{\text{MSI}}^i))) \odot A(\mathbf{F}_{\text{MSI}}^i) \quad (14)$$

$$\mathbf{A}_f = \text{Conv}_{1 \times 1}(\text{Cat}(\mathbf{EA}(\mathbf{F}_i), A(\mathbf{F}_{\text{MSI}}^i))) \quad (15)$$

融合后的幅值分量和相位分量通过逆傅里叶变换得到频率域的融合特征,其与输入的融合特征相加,则获得FIB输出的融合特征 $\mathbf{F}_f$ 。

$$\mathbf{F}_f^i = \text{IFFT}(\mathbf{A}_f, \mathbf{P}_f) + \mathbf{D}_n(\mathbf{F}_{i-1}) \quad (16)$$

式中, $\text{IFFT}(\cdot)$ 表示逆傅里叶变换。

1.2.2 空间特征增强块(SFEB)

为了获得具有丰富纹理信息的不同尺度融合特征,构建了一个特征增强块,结构如图2所示。该模块通过建立当前尺度特征 $\mathbf{F}_{\text{MSI}}^i$ 的长距离依赖关系,从而实现融合特征 $\mathbf{F}_f^i$ 的增强。在该结构中,本文采用单头注意力机制以减少模型对计算资源的需求,并加快模型的运算速度。首先,对来自 FIB 的融合特征 $\mathbf{F}_f^i$ 以及同一尺度的特征 $\mathbf{F}_{\text{MSI}}^i$ 执行矩阵维度的改变操作,再利用3个线性映射层得到3个特征向量 $\mathbf{q}_i$ 、 $\mathbf{k}_i$ 和 $\mathbf{v}_i$ 。其定义为

$$\mathbf{q}_i = \text{Linear}(\text{Reshape}(\mathbf{F}_{\text{MSI}}^i)) \quad (17)$$

$$\mathbf{k}_i = \text{Linear}(\text{Reshape}(\mathbf{F}_{\text{MSI}}^i)) \quad (18)$$

$$\mathbf{v}_i = \text{Linear}(\text{Reshape}(\mathbf{F}_f^i)) \quad (19)$$

式中, $\text{Reshape}(\cdot)$ 表示改变矩阵维度, $\text{Linear}(\cdot)$ 表示线性映射。

然后,计算特征向量 $\mathbf{q}_i$ 和 $\mathbf{k}_i$ 的高频特征自相关矩阵,并将其用于增强向量 $\mathbf{v}_i$ 的结构信息。首先,从特征向量 $\mathbf{q}_i$ 和 $\mathbf{k}_i$ 中减去均值信息获得高频特征;其次,高频特征矩阵被执行矩阵乘法得到自相关矩阵 $\mathbf{S}_{\text{self}}$;同时,自相关矩阵通过一个 softmax 层获得注意力权重系数,该系数与向量 $\mathbf{v}_i$ 相乘得到增强特征向量;最后,增强特征向量被转换回高维特征空间得到第 i 个尺度层中 FSFFM 的输出。具体操作可表示为

$$\mathbf{S}_{\text{self}} = \text{MatMul}((\mathbf{q}_i - \text{mean}(\mathbf{q}_i)), (\mathbf{k}_i - \text{mean}(\mathbf{k}_i))^T) \quad (20)$$

$$\mathbf{F}_i = \text{Reshape}(\text{MatMul}(\text{softmax}(\mathbf{S}_{\text{self}}), \mathbf{v}_i)) \quad (21)$$

式中, $\text{MatMul}(\cdot)$ 表示矩阵乘法, $\text{mean}(\cdot)$ 表示计算矩阵平均值, $\mathbf{S}_{\text{self}}$ 表示学习得到的 HR-MSI 自相关矩阵。

1.3 损失函数

使用 L1 损失函数对模型进行训练,具体为

$$L = \frac{1}{M} \sum_{m=1}^M \|\hat{\mathbf{I}}_{\text{HR}}^m - \hat{\mathbf{I}}_{\text{HR}}\|_1 \quad (22)$$

式中, $\hat{\mathbf{I}}_{\text{HR}}^m \in \mathbf{R}^{C \times H \times W}$ 表示真实的 HR 图像, M 代表训练样本的总数。

2 实验

本节内容分为4个主要部分:首先对实验中的

参数设置以及训练环境做出详细阐述;然后在大量遥感 HSI 上对本文提出的方法及一些近期的先进方法进行测试,通过对比主观视觉效果及客观性能评价指标,验证所提出方法的有效性和优势;接着通过设计所提出模型的不同变体对各模块进行消融研究,分析模型架构的合理性及各个组成模块的重要性;最后通过遥感分类任务实验验证网络的下游应用价值。

2.1 参数设置及数据集和评价指标

为证明 DDIF-Net 的有效性,本文在3个公开的高光谱数据集 Pavia Center、Botswana 和 Chikusei 上对网络的效果进行了实验验证。实验采用6种主要的性能评价指标:SCC(spectral cross correlation)、SAM(spectral angle mapper)、ERGAS(relative dimensionless global error of synthesis)、UIQI(universal image quality index)、SSIM(structural similarity index measure)和 PSNR(peak signal-to-noise ratio)。所有指标在 HSI 的所有波段上进行计算,并取平均值以得到最终结果。

实验数据集处理遵循 Wald 协议,将原始数据集处理为4倍下采样的 LR-MSI 和对应的 HR-MSI,从中选取80%的图像作为训练集,剩余20%作为测试集。为确保比较的公平性,所有基于深度学习的方法均在 Linux 系统上经由 PyTorch 平台重新训练。实验使用 NVIDIA GeForce GTX 3090 GPU,采用 Adam 优化器,经过10 000个训练周期达到最优性能,其中初始学习率设置为 1×10^{-4} ,在8 000个周期后以 1×10^{-5} 的因子进行衰减。

2.2 实验结果与分析

实验部分将 DDIF-Net 与8种深度学习方法 SSFCNN(spatial and spectral fusion with CNN)(Han 等, 2018)、HyperPNN(hyperspectral pansharpening via spectrally predictive CNN)(He 等, 2019)、DARN(deep hyperspectral prior and dual-attention residual network)(Zheng 等, 2020)、SSRNet(spatial-spectral reconstruction network)(Zhang 等, 2021)、HSRNet(hyperspectral image super-resolution via deep spatio-spectral attention convolutional neural networks)(Hu 等, 2022b)、HyperDSNet(deep-shallow fusion network for hyperspectral pansharpening)(Zhuo 等, 2022)、HyperRefiner(the refined hyperspectral pansharpening network)(Zhou 等, 2023)和 FPFNet(feature pyra-

mid fusion network) (Dong 等, 2025) 进行了比较。图 3 展示了在 Pavia Center、Botswana 和 Chikusei 3 个数据集上基于深度学习的方法所得到的重建结果。

为符合人类视觉习惯, 本文选取相应的 RGB 通道合成重建结果的伪彩色图像。为更直观地比较各方法的重建质量, 本文计算重建结果与 GT-HSI 之间的绝对平均误差 (mean absolute error, MAE) 图, 并选取局部区域进行放大, 放置在相应的结果下方用于细节对比。虽然伪彩色图像间的差异不显著, 但通过比较放大的 MAE 图, 可以明显看出在处理复杂场景 (如城市建筑) 时, DDIF-Net 的 MAE 图几乎没有显

示原始图像的明显轮廓信息, 这意味着 DDIF-Net 能够更准确地恢复这些复杂场景的细节信息, 反映了 DDIF-Net 在细粒度纹理和粗粒度结构恢复方面的显著优势。此外, 为评估各对比方法的光谱保真能力, 图 4 展示了各方法的光谱拟合曲线。结果显示, 本文提出的方法实现了最小的光谱差异, 表明其具备优异的光谱保真性。

除视觉质量对比外, 从定量角度对所提出的方法进行验证, 表 1 列出了各方法在测试集上的平均质量指标。HyperRefiner 是一个非常新颖的方法, 但其在 LR-HSI 光谱特征提取中采用自编码器进行压缩和降维, 导致光谱信息存在部分丢失, 并且由于模

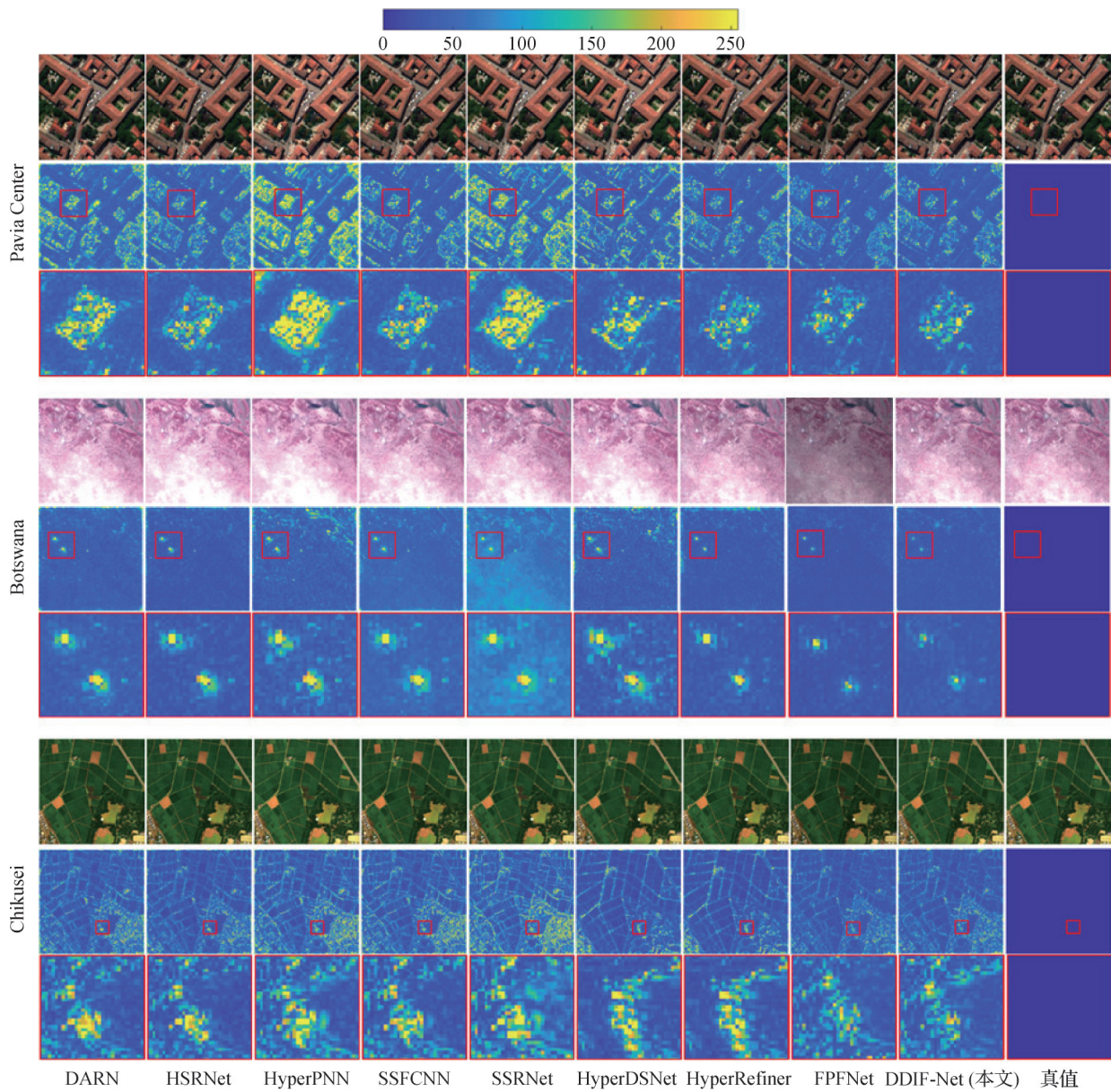


图 3 Pavia Center、Botswana 和 Chikusei 数据集上主观对比
Fig. 3 Subjective comparison on Pavia Center, Botswana and Chikusei datasets

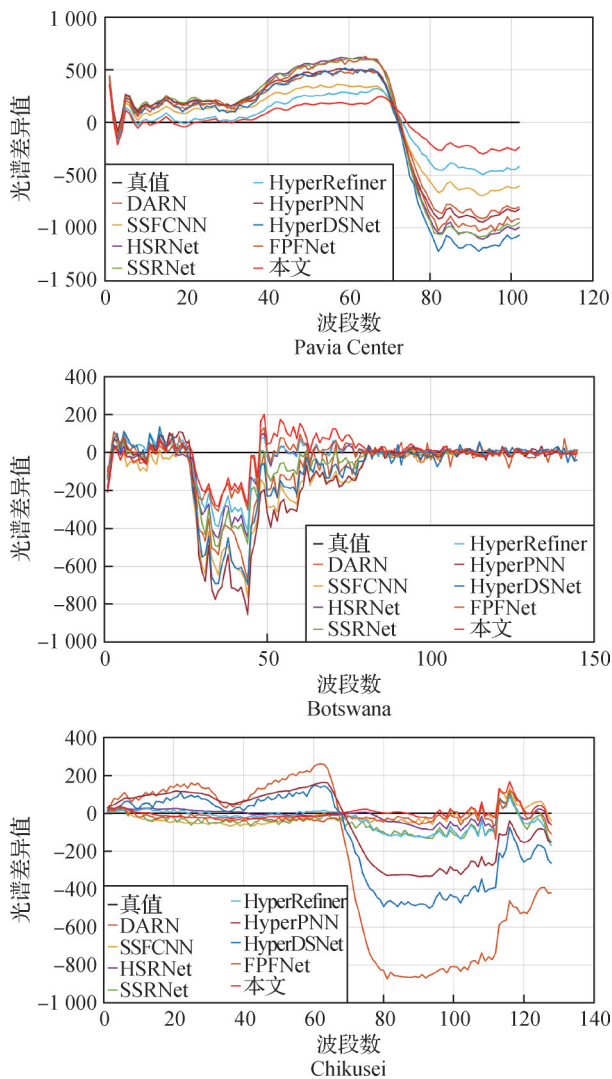


图4 不同方法在Pavia Center、Botswana和Chikusei数据集上随机位置光谱差异值曲线结果

Fig. 4 The spectrum difference value at random pixels for different methods on Pavia Center, Botswana and Chikusei datasets

型结构的复杂性,其运行时间较长。HSRNet展示了较优的性能,并且优于同样应用了注意力机制的HyperDSNet和DARN。这可能是因为前者的网络架构更加简洁合理,避免了大量卷积运算,有效减少了重建过程中的信息损耗。从表中数据可知,在Pavia Center数据集上,DDIF-Net在所有评价指标上均排名第1;在Botswana数据集上,除ERGAS指标略逊于FPFNet外,其他所有指标均领先;在Chikusei数据集上,除SCC指标仅低于最优指标0.003外,其他所有指标均领先。此外,本文提出的方法的运行时间显著缩短,实现了性能与效率的更好平衡。

以上定性和定量分析表明,本文提出的DDIF-

Net在提供更高质量、更少伪影以及在视觉上更为准确的HSI超分重建结果方面,明显优于现有方法。

2.3 消融实验分析

本文提出的DDIF-Net包含一个并行双U型网络架构及两个创新模块:FSFFM和MFFM。为深入探究网络内各组件对整体模型性能的贡献,本文设计了3组消融实验,并在Botswana数据集的24个降分辨率HSI样本上进行了验证。实验1(V1)维持双U形网络架构,但用串联操作代替了FSFFM,以评估FSFFM在融合频率域与空间域信息上的有效性;实验2(V2)移除了FSFFM中的频率特征注入模块FIB,旨在验证高频特征注入及双域互补学习策略的必要性;实验3(V3)省略了多尺度特征融合模块,只采用了简单的Concat和卷积操作来结合4倍上采样的LR-HSI特征与主网络输出的特征,意在检验多级融合策略在特征恢复中的作用;实验4(V4)使用3×3卷积特征提取模块代替SFEB进行空间特征的提取。实验结果详见表2。

从表2可以看出,即使在移除FSFFM的情况下,DDIF-Net架构也能够保持出色的性能。这一成就主要归功于网络结构的设计,这种结构使得网络可以充分提取和融合来自不同图像的特征信息。此外,实验结果还表明,移除FIB后,模型的所有评估指标性能均有所下降,从而验证了在频率域进行高频信息注入对于提升模型整体性能的重要性。同样,多级特征融合策略对于本网络而言是不可或缺的,而采用单一尺度融合明显降低了模型的表现。综上所述,当DDIF-Net的所有组件协同工作时,模型展现了最优的性能,确认了所提出的各组件都在提高最终重建质量中发挥了关键作用。

为了进一步验证模型的有效性,本文给出了基于深度学习的对比方法在Chikusei数据集上测得的模型参数量和计算复杂度(每秒浮点运算次数(floating point operations per second,FLOPs)),如表3所示。

虽然本文模型相比HyperPNN、HyperRefiner和HyperDSNet具有更多的参数量,但结合表1中的性能比较,可以看出本文模型融合结果优于这些方法。另外,与FPFNet方法相比,本文在使用较少参数和较低计算复杂度的情况下取得了更好的结果。综合来看,本文提出的DDIF-Net具有优良的HSR性能。

表 1 Pavia Center、Botswana 和 Chikusei 数据集上的平均定量指标结果
Table 1 Average quantitative results on Pavia Center, Botswana and Chikusei datasets

数据集	方法	PSNR/dB ↑	SAM ↓	ERGAS ↓	UIQI ↑	SSIM ↑	SCC ↑	运行时间/ms
Pavia Center	DARN	38.01	5.13	2.80	0.962	0.954	0.975	1.675
	HSRNet	38.49	4.49	2.60	0.967	0.963	0.983	7.011
	HyperPNN	35.29	5.99	3.66	0.941	0.924	0.952	1.786
	SSFCNN	40.42	4.33	2.40	0.968	0.964	0.982	2.101
	SSRNet	37.31	5.43	3.17	0.951	0.935	0.958	1.634
	HyperDSNet	37.17	5.15	2.89	0.962	0.954	0.982	1.862
	HyperRefiner	38.98	4.32	2.32	0.969	0.963	0.985	9.135
	FPFNet	<u>42.04</u>	<u>4.17</u>	<u>2.17</u>	<u>0.971</u>	<u>0.969</u>	<u>0.986</u>	5.183
	DDIF-Net (本文)	42.27	3.93	2.11	0.973	0.972	0.987	1.505
Botswana	DARN	42.72	1.57	2.30	0.923	0.971	0.992	1.675
	HSRNet	44.77	<u>1.27</u>	1.45	0.959	<u>0.980</u>	<u>0.995</u>	7.011
	HyperPNN	43.08	2.05	1.60	0.933	0.959	0.981	1.786
	SSFCNN	42.39	1.86	7.81	0.870	0.918	0.994	2.101
	SSRNet	40.49	2.02	2.80	0.765	0.824	0.993	1.634
	HyperDSNet	44.34	1.70	1.45	0.942	0.961	0.985	1.862
	HyperRefiner	44.79	1.44	1.86	0.948	0.976	<u>0.995</u>	9.135
	FPFNet	<u>45.75</u>	1.29	1.37	<u>0.956</u>	0.969	0.996	5.183
	DDIF-Net (本文)	46.16	1.17	<u>1.42</u>	0.959	0.984	0.996	1.129
Chikusei	DARN	43.87	1.88	2.47	0.957	0.965	0.965	4.086
	HSRNet	45.12	<u>1.69</u>	<u>2.25</u>	0.965	<u>0.974</u>	0.974	13.569
	HyperPNN	41.57	2.04	5.64	0.926	0.927	0.953	6.041
	SSFCNN	42.26	2.04	2.77	0.950	0.958	0.953	3.893
	SSRNet	43.03	2.20	3.11	0.940	0.946	0.935	3.219
	HyperDSNet	41.91	2.42	3.65	0.945	0.972	<u>0.977</u>	4.109
	HyperRefiner	40.21	1.57	2.72	0.961	0.963	0.974	12.275
	FPFNet	<u>46.36</u>	1.79	<u>2.25</u>	<u>0.962</u>	<u>0.974</u>	0.978	8.674
	DDIF-Net (本文)	46.73	1.57	2.23	0.965	0.975	0.975	6.472

注:加粗、下划线字体表示在各数据集上各指标的最优、次优结果。↑表示值越大越好,↓表示值越小越好。

表 2 Botswana 数据集上消融研究
Table 2 Ablation study on Botswana dataset

方法	SCC ↑	SAM ↓	ERGAS ↓	UIQI ↑	SSIM ↑	PSNR/dB ↑
V1	0.995	1.41	2.06	0.936	0.978	43.86
V2	0.995	1.36	1.63	0.946	0.981	44.84
V3	0.995	1.39	1.80	0.945	0.979	44.35
V4	0.992	1.32	1.89	0.944	0.979	45.46
DDIF-Net(本文)	0.996	1.17	1.42	0.959	0.984	46.16

注:加粗字体表示各列最优结果。↑表示值越大越好,↓表示值越小越好。

表3 基于深度学习方法在Chikusei数据集上的参数量和计算复杂度比较

方法	Params/M	FLOPs/G
HyperPNN	0.52	17.72
HyperDSNet	0.53	43.78
HyperRefiner	19.32	170.07
FPFNet	22.34	349.99
DDIF-Net (本文)	17.98	227.46

2.4 分类实验

为了进一步验证 DDIF-Net 在高层次任务中的有效性,本文在融合结果上进行了目标分类实验。使用 ENVI 软件中的 K-means 聚类算法对不同的融合结果进行分类,分类的类别数设置为 4。分类结果如图 5 和表 4 所示,展示了不同方法在

Pavia Center 数据集上的总体准确率(overall accuracy, OA)和卡帕(Kappa)系数。从图 5 可以看出,融合图像的分类结果与真实图像(ground truth, GT)的分类结果最为接近;表 4 显示, DDIF-Net 所得的分类精度高于其他对比方法,表明了本文的融合结果在空间与光谱保真度方面更接近 GT。

表4 Pavia Center数据集上分类结果的客观对比
Table 4 Objective comparison of classification results on Pavia Center dataset

方法	OA ↑/%	Kappa ↑
SSFCNN	94.15	0.920
HyperDSNet	94.60	0.926
HyperRefiner	94.54	0.926
FPFNet	94.67	0.927
DDIF-Net (本文)	95.07	0.933

注:加粗字体表示各列最优结果。↑表示值越大越好。

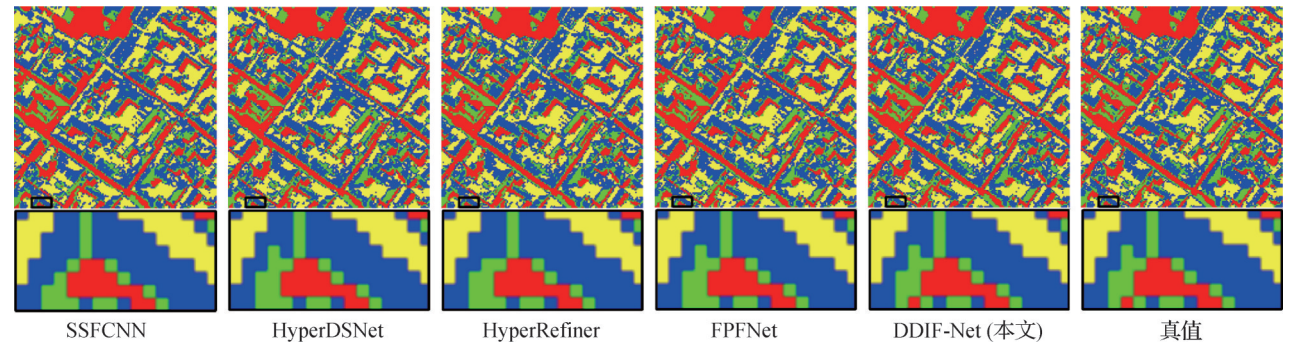


图5 对比方法融合图像在Pavia Center数据集上的分类结果

Fig. 5 Classification results of fused images from comparison methods on Pavia Center dataset

3 结 论

考虑到 HR-MSI 与 LR-HSI 特征间的模态差异性,本文提出一种基于频域—空间域互补学习策略的空间—光谱特征融合网络 DDIF-Net,用于 HISR 任务。网络通过构建一个 MS 特征提取分支实现 HR-MSI 不同尺度特征的提取,并通过构建一个空—频双域特征融合分支在空间域和频率域实现特征的融合和重建。其中,在双域特征融合分支中,构建一个 FIB 将图像特征变换到傅里叶域来减弱两种图像特征的模态差异。针对来自 FIB 的融合特征,构建出

SFEB,通过注意力机制引导,实现融合特征的结构信息的增强。另外,MFFM 被构建以整合不同尺度的融合特征,获得高质量的重建结果。实验结果表明,本文提出的方法在多个数据集上的性能均超越现有的一些 HISR 方法,展现了该模型在特征学习、信息整合和泛化能力方面的优势。

本文所提出的 DDIF-Net 在 Pavia Center、Botswana 和 Chikusei 3 个广泛使用的地面遥感高光谱数据集上取得了优异的重建效果,验证了方法在精细场景下的有效性。但是,由于目前 DDIF-Net 的参数量和计算量较高,在面对资源受限的平台上部署时可能带来实时性挑战。另一方面,本文使用高

质量数据集进行训练和测试,在极端光照、强噪声或大面积遮挡等复杂场景下,模型在特征提取与匹配过程中可能难以维持与训练阶段相同的判别能力,导致跨模态特征融合过程中容易出现信息错配、权重分配失衡等问题。未来工作将重点优化模型结构以降低计算开销,并在更大尺度的多源卫星遥感数据(如 Sentinel-2 高分系列)上进一步验证方法的通用性和工程实用性,提升模型在多源、多场景高光谱数据中的泛化能力与实用性。

参考文献(References)

- 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]
- Dian R, Li S T, Guo A J, and Fang L Y. 2018. Deep hyperspectral image sharpening. *IEEE Transactions on Neural Networks and Learning Systems*, 29(11): 5345-5355 [DOI: 10.1109/TNNLS.2018.2798162]
- Dong W Q, Yang Y H, Qu J H, Li Y S, Yang Y F and Jia X P. 2025. Feature pyramid fusion network for hyperspectral pansharpening. *IEEE Transactions on Neural Networks and Learning Systems*, 36(1): 1555-1567 [DOI: 10.1109/TNNLS.2023.3325887]
- 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. Athens, Greece; IEEE: 2506-2510 [DOI: 10.1109/ICIP.2018.8451142]
- He L, Zhu J W, Li J, Plaza A, Chanussot J and Li B. 2019. HyperPNN: hyperspectral pansharpening via spectrally predictive convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(8): 3092-3100 [DOI: 10.1109/JSTARS.2019.2917584]
- Hu J F, Huang T Z, Deng L J, Dou H X, Hong D F and Vivone G. 2022a. Fusformer: a transformer-based fusion network for hyperspectral image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, 19: #6012305 [DOI: 10.1109/LGRS.2022.3194257]
- Hu J F, Huang T Z, Deng L J, Jiang T X, Vivone G and Chanussot J. 2022b. Hyperspectral image super-resolution via deep spatio-spectral attention convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12): 7251-7265 [DOI: 10.1109/TNNLS.2021.3084682]
- Hu S, Gao F, Gong Z R, Tao S E, Shanguan X Y and Dong J Y. 2024. Parallel channel shuffling and Transformer-based denoising for hyperspectral images. *Journal of Image and Graphics*, 29(7): 2063-2074 (胡帅, 高峰, 龚卓然, 陶盛恩, 上官心语, 董军宇. 2024. 基于 Transformer 和通道混合并行卷积的高光谱图像去噪. *中国图象图形学报*, 29(7): 2063-2074 [DOI: 10.11834/jig.230381])
- Huang Y P and Li W S. 2023. A review of medical image fusion methods. *Journal of Image and Graphics*, 28(1): 118-143 (黄渝萍, 李伟生. 2023. 医学图像融合方法综述. *中国图象图形学报*, 28(1): 118-143 [DOI: 10.11834/jig.220603])
- Huang Z H, Chen Q, Chen Q H and Liu X G. 2018. Variational pansharpening for hyperspectral imagery constrained by spectral shape and Gram-Schmidt transformation. *Sensors*, 18(12): #4330 [DOI: 10.3390/s18124330]
- Kawakami R, Matsushita Y, Wright J, Ben-Ezra M, Tai Y W and Ikeuchi K. 2011. High-resolution hyperspectral imaging via matrix factorization//Proceedings of 2011 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Colorado Springs, CO, USA: IEEE: 2329-2336 [DOI: 10.1109/CVPR.2011.5995457]
- Li M Y, Fu Y and Zhang Y L. 2023. Spatial-spectral transformer for hyperspectral image denoising//Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington, USA: AAAI: 1368-1376 [DOI: 10.1609/aaai.v37i1.25221]
- Li Y, Dong S H, Zhang J W, Zhao R and Zheng Y H. 2025. Cross-scale transformer image super-resolution reconstruction with fusion channel attention. *Journal of Image and Graphics*, 30(3): 784-797 (李焱, 董仕豪, 张家伟, 赵茹, 郑钰辉. 2025. 融合通道注意力的跨尺度 Transformer 图像超分辨率重建. *中国图象图形学报*, 30(3): 784-797 [DOI: 10.11834/jig.240279])
- Liao W Z, Huang X, Van Coillie F, Gautama S, Pižurica A, Philips W, et al. 2015. Processing of multiresolution thermal hyperspectral and digital color data: outcome of the 2014 IEEE GRSS data fusion contest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6): 2984-2996 [DOI: 10.1109/JSTARS.2015.2420582]
- Licciardi G, Veganzones M A, Simões M, Bioucás J and Chanussot J. 2014. Super-resolution of hyperspectral images using local spectral unmixing//Proceedings of the 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing. Lausanne, Switzerland: IEEE: 1-4 [DOI: 10.1109/WHISPERS.2014.8077569]
- Liu L, Wang J, Zhang E L, Li B, Zhu X, Zhang Y Q, et al. 2020. Shallow-deep convolutional network and spectral-discrimination-based detail injection for multispectral imagery pan-sharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 1772-1783 [DOI: 10.1109/JSTARS.2020.2981695]
- Peng S R, Guo C H, Wu X and Deng L J. 2023. U2Net: a general framework with spatial-spectral-integrated double U-Net for image fusion//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada: ACM: 3219-3227 [DOI: 10.1145/3581783.3612084]

- Restaino R, Vivone G, Addesso P and Chanussot J. 2021. Hyperspectral sharpening approaches using satellite multiplatform data. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1): 578-596 [DOI: 10.1109/TGRS.2020.3000267]
- Tseng D C, Chen Y L and Liu M S C. 2001. Wavelet-based multispectral image fusion//Scanning the Present and Resolving the Future. *Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium*. Sydney, Australia: IEEE: 1956-1958 [DOI: 10.1109/IGARSS.2001.977128]
- Vivone G, Restaino R, Dalla Mura M, Licciardi G and Chanussot J. 2014. Contrast and error-based fusion schemes for multispectral image pansharpening. *IEEE Geoscience and Remote Sensing Letters*, 11(5): 930-934 [DOI: 10.1109/LGRS.2013.2281996]
- Wei Q, Bioucas-Dias J, Dobigeon N and Tournet J Y. 2015. Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7): 3658-3668 [DOI: 10.1109/TGRS.2014.2381272]
- Yang J X, Zhao Y Q and Chan J C W. 2018. Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network. *Remote Sensing*, 10(5): #800 [DOI: 10.3390/rs10050800]
- Yokoya N and Iwasaki A. 2014. Effect of unmixing-based hyperspectral super-resolution on target detection//*Proceedings of the 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*. Lausanne, Switzerland: IEEE: 1-4 [DOI: 10.1109/WHISPERS.2014.8077602]
- Yu C Y, Han R, Song M P, Liu C Y and Chang C I. 2020. A simplified 2D-3D CNN architecture for hyperspectral image classification based on spatial-spectral fusion. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 2485-2501 [DOI: 10.1109/JSTARS.2020.2983224]
- Yu Y, Gu H and Sun W D. 2010. Spectral unmixing based image restoration for optical remote sensing images covered by thin cloud. *Journal of Image and Graphics*, 15(11): #1670 (于钺, 顾华, 孙卫东. 2010. 基于混合像元分解的薄云下光学遥感图像恢复方法. *中国图象图形学报*, 15(11): #1670) [DOI: 10.11834/jig.20101104]
- Yuan Y, Wang Q and Li X L. 2020. Hyperspectral and multispectral image fusion using non-convex relaxation low rank and total variation regularization//*Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*. Waikoloa, USA: IEEE: 2683-2686 [DOI: 10.1109/IGARSS39084.2020.9323227]
- Zeng H J, Xie X Z, Cui H J, Yin H P and Ning J F. 2021. Hyperspectral image restoration via global L1-2 spatial-spectral total variation regularized local low-rank tensor recovery. *IEEE Transactions on Geoscience and Remote Sensing*, 59(4): 3309-3325 [DOI: 10.1109/TGRS.2020.3007945]
- Zhang B, Wu Y F, Zhao B Y, Chanussot J, Hong D F, Yao J, et al. 2022. Progress and challenges in intelligent remote sensing satellite systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 1814-1822 [DOI: 10.1109/JSTARS.2022.3148139]
- Zhang H Z, Yang Y H, Li C Y, Lu Y J, Zhang G D and Chen S Y. 2025a. CESFusion: cross-frequency enhanced spatial-spectral fusion network for hyperspectral and multispectral image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 63: #5511815 [DOI: 10.1109/TGRS.2025.3563847]
- Zhang M M, Li W and Du Q. 2018. Diverse region-based CNN for hyperspectral image classification. *IEEE Transactions on Image Processing*, 27(6): 2623-2634 [DOI: 10.1109/TIP.2018.2809606]
- 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]
- Zhang Y F, De Backer S and Scheunders P. 2009. Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 47(11): 3834-3843 [DOI: 10.1109/TGRS.2009.2017737]
- Zhang Y F and He M Y. 2007. Multi-spectral and hyperspectral image fusion using 3-D wavelet transform. *Journal of Electronics (China)*, 24(2): 218-224 [DOI: 10.1007/s11767-005-0232-5]
- Zhang Y X, Li W, Jia W, Zhang M M, Tao R and Liang S L. 2025b. Cross-domain hyperspectral image classification based on bi-directional domain adaptation. *IEEE Transactions on Circuits and Systems for Video Technology*, 1-12 [DOI: 10.1109/TCSVT.2025.3586282]
- Zhang Y X, Zhang M M, Li W, Wang S and Tao R. 2023. Language-aware domain generalization network for cross-scene hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61: #5501312 [DOI: 10.1109/TGRS.2022.3233885]
- Zhao J Q, Zhang M, Zhou Z H, Wang Z X, Lang F K, Shi H T, et al. 2025. CFFormer: a cross-fusion transformer framework for the semantic segmentation of multisource remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 63: #4401117 [DOI: 10.1109/TGRS.2024.3507274]
- Zheng Y X, Li J J, Li Y S, Guo J, Wu X Y and Chanussot J. 2020. Hyperspectral pansharpening using deep prior and dual attention residual network. *IEEE Transactions on Geoscience and Remote Sensing*, 58(11): 8059-8076 [DOI: 10.1109/TGRS.2020.2986313]
- Zhou B, Zhang X F, Chen X, Ren M and Feng Z Y. 2023. HyperRefiner: a refined hyperspectral pansharpening network based on the autoencoder and self-attention. *International Journal of Digital Earth*, 16(1): 3268-3294 [DOI: 10.1080/IJDE.2023.2246944]
- Zhuo Y W, Zhang T J, Hu J F, Dou H X, Huang T Z and Deng L J. 2022. A deep-shallow fusion network with multidetail extractor and spectral attention for hyperspectral pansharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*

ing, 15: 7539-7555 [DOI: 10.1109/JSTARS.2022.3202866]

作者简介

杨怀远,男,博士研究生,主要研究方向为遥感图像处理与机器学习。E-mail:2320080124@tiangong.edu.cn
杨勇,通信作者,男,教授,主要研究方向为遥感图像处理、计算机视觉与机器学习。E-mail:yangyong@tiangong.edu.cn

黄淑英,女,教授,主要研究方向为遥感图像处理、计算机视觉与机器学习。E-mail:huangshuying@tiangong.edu.cn
刘紫阳,男,博士研究生,主要研究方向为遥感图像处理与机器学习。E-mail:qduliuzy@163.com
张龙,男,博士研究生,主要研究方向为遥感图像处理与机器学习。E-mail:redmac@126.com