中图法分类号:TP391;TP18　　**文献标识码**:A　　**文章编号**:1006-8961(2026)11-4119-14

论文引用格式:Huang S S and Liu C X. 2026. Prism pyramid fusion network for detail-preserving pansharpening. Journal of Image and Graphics, 31(11):4119-4132(黄珊珊, 刘春晓. 2026. 面向细节保持全色锐化的棱镜金字塔融合网络. 中国图象图形学报, 31(11):4119-4132)[DOI: 10.11834/jig.250527]

面向细节保持全色锐化的棱镜金字塔融合网络

黄珊珊,刘春晓*

浙江工商大学计算机科学与技术学院,杭州　310018

摘　要:**目的**　为了解决传统全色锐化方法中空间细节恢复不足、光谱保持不稳定及上采样退化等问题,本文提出一种基于棱镜金字塔融合(prism pyramid fusion,PPFusion)的细节保持型全色锐化网络,以实现空间与光谱特征的协同优化。**方法**　所提网络首先对低分辨率多光谱图像(low-resolution multispectral image,LRMS)进行上采样,使其与全色图像(panchromatic image,PAN)在空间尺度上对齐,并将二者堆叠作为网络输入。网络的主干部分由差分增强卷积模块(differentially enhanced convolution module,DEConv)与Restormer模块组成,两者并行提取高频纹理与长程光谱依赖特征,并通过多次特征交互实现逐级细节恢复。随后,融合阶段设计了层次化内容引导的注意力融合模块(hierarchical content-guided attention fusion module,HCGAF),将LRMS、PAN、初步融合结果及前一阶段融合特征进行多尺度内容引导融合,以获得光谱一致且纹理丰富的最终输出。此外,提出了双分支注意力引导的共享上采样模块(dual-branch attention-guided shared upsampling module,DASU)作为正则化约束,上采样模块独立学习从低分辨率到高分辨率域的映射,并在损失函数中设计了主干输出与上采样输出之间的一致性约束,从而稳定训练并提升模型的泛化能力。**结果**　在WorldView-3、高分二号和QuickBird三个数据集上的实验结果表明,本文提出的PPFusion在PSNR(peak signal-to-noise ratio)、SAM(spectral angle mapper)、ERGAS(erreur relative globale adimensionnelle de synthèse)及QNR(quality with no reference)等指标上整体优于现有代表性方法,并在复杂场景下生成纹理更清晰、边缘更自然的融合图像。消融实验进一步验证了DEConv模块在高频增强、Restormer模块在光谱建模、HCGAF模块在层次融合及DASU约束在稳定训练中的关键作用。**结论**　所提出的PPFusion网络通过"差分增强—Restormer重建—层次化融合—上采样约束"四者协同机制,在光谱一致性与空间锐度之间取得了良好平衡,为高分辨率遥感图像的融合与重建提供了一种高效、可扩展的新思路。

关键词:注意力机制;卷积神经网络(CNN);Restormer编码器;特征融合;差分卷积;全色锐化

Prism pyramid fusion network for detail-preserving pansharpening

Huang Shanshan, Liu Chunxiao*

School of Computer Science and Technology, Zhejiang Gongshang University, Hangzhou　310018, China

Abstract: **Objective** Pansharpening aims to generate a high-resolution multispectral (HRMS) image by integrating the fine spatial structure of a high-resolution panchromatic (PAN) image with the rich spectral information of a low-resolution

收稿日期:2025-11-05;**修回日期**:2026-01-30;**预印本日期**:2026-02-06

*通信作者:刘春晓　cxliu@mail.zjgsu.edu.cn

基金项目:国家自然科学基金项目(U25A20440);浙江省自然科学基金项目(LY24F020004);浙江省重点研发计划项目(2023C01039);浙江省大学生科技创新活动计划暨新苗人才计划项目(2026R408A013);国家级大学生创新创业训练计划项目(CX20260448031)

Supported by: National Natural Science Foundation of China (U25A20440); Zhejiang Provincial Natural Science Foundation, China (LY24F020004); Key R&D Program of Zhejiang Province, China (2023C01039); Zhejiang University Students Science and Technology Innovation Activity Plan and New Talent Plan (2026R408A013); National College Students Innovation and Entrepreneurship Training Program (CX20260448031)

multispectral (LRMS) image. Although deep-learning-based approaches have made substantial progress, existing methods still face several key limitations, including insufficient spatial-detail recovery, spectral inconsistency, and degradation during upsampling. These problems arise largely because current frameworks do not model spatial texture, spectral dependency, and scale transition in a unified manner. To address these issues, this paper proposes a detail-preserving pansharpening network based on prism pyramid fusion (PPFusion). The proposed model integrates differential enhancement, Restormer-based long-range dependency modeling, hierarchical attention fusion, and shared upsampling regularization into a single deep-learning architecture. By jointly optimizing spatial and spectral representations, PPFusion aims to achieve a balanced and interpretable fusion process, producing HRMS images with improved fidelity and structural clarity. **Method** The PPFusion framework first upsamples the LRMS image to match the spatial resolution of the PAN image. The upsampled LRMS and PAN images are then concatenated and fed into a dual-path backbone. The first branch uses a differentially enhanced convolution (DEConv) module to capture high-frequency spatial details and edge contours explicitly through directional differential operators. This design strengthens texture gradients and sharp structures that are often lost during fusion. The second branch adopts a Restormer-based transformer module, which uses multi-head self-attention and feed-forward channel aggregation to model long-range spectral dependencies and maintain spectral consistency across channels. The two branches interact through three cross-scale fusion stages, allowing mutual refinement of spatial and spectral representations. This progressive interaction supports hierarchical detail reconstruction and spectral-spatial alignment. After backbone processing, a hierarchical content-guided attention fusion (HCGAF) module performs advanced multi-scale aggregation. HCGAF integrates four complementary inputs—the upsampled LRMS image, the PAN image, the intermediate fusion output, and the fusion features from the previous stage. Within a prism-pyramid structure, multi-level attention maps guide the integration of local fine-grained texture and global spectral semantics, ensuring that the fused image preserves both high-frequency precision and global coherence. To further stabilize training, a dual-branch attention-guided shared upsampling (DASU) module is introduced as a regularization branch rather than part of the main fusion pipeline. The DASU branch independently learns a mapping from LRMS to HRMS, using the PAN image as spatial guidance. During training, a self-consistency loss is imposed between the DASU output and the main-network output, forming a teacher-student-style constraint that regularizes optimization, mitigates overfitting, and improves generalization. The total loss includes three components: an $L1$ spectral-fidelity term between the fused image and the reference HRMS image, a spatial-consistency term derived from DASU supervision, and a smoothness constraint that promotes stable optimization. Together, these losses encourage balanced optimization of spectral accuracy and spatial sharpness. **Result** Experiments were conducted on three benchmark satellite datasets—WorldView-3 (WV3), GaoFen-2 (GF2), and QuickBird (QB)—with twelve state-of-the-art deep-learning pansharpening methods used as baselines. Quantitative results show that PPFusion achieves state-of-the-art performance in spectral angle mapper (SAM) while maintaining competitive PSNR, ERGAS, and SSIM across datasets. Full-resolution evaluation also indicates comparable quality with no reference (QNR) scores, demonstrating robustness under no-reference conditions. Visual inspection shows sharper building edges, more realistic vegetation textures, and more natural color transitions, with fewer spectral shifts and fewer blurred boundaries. Ablation studies confirm the effectiveness of each module: DEConv sharpens edges, Restormer improves spectral-spatial coherence, HCGAF enables effective multi-scale integration, and DASU enhances training stability. **Conclusion** The proposed PPFusion network introduces a new framework that jointly leverages differential convolution, Restormer-based long-range modeling, hierarchical content-guided fusion, and shared upsampling regularization. By combining local-detail reconstruction with global spectral modeling, it effectively balances spatial sharpness and spectral fidelity. Extensive experiments verify that PPFusion achieves state-of-the-art performance in both quantitative accuracy and perceptual quality. The DASU regularization branch further stabilizes training and improves adaptability across different satellite sensors. Future work will extend the prism-pyramid fusion paradigm to cross-sensor and multi-temporal pansharpening and explore the integration of physical imaging priors and generative modeling for more interpretable and adaptive remote-sensing image fusion.

Key words: attention mechanism; convolutional neural network (CNN); Restormer; feature fusion; differential convolution; pansharpening

0 引 言

高分辨率多光谱图像(high-resolution multispectral image,HRMS)在目标检测(Cheng和Han,2016)、场景分类(Nogueira等,2017)及环境监测(Foody,2003)等任务中具有重要价值。然而,受限于卫星数据传输能力,大多数遥感卫星难以同时获取高空间与高光谱分辨率的图像,通常仅能提供低分辨率多光谱图像(low-resolution multispectral image,LRMS)与高分辨率全色图像(panchromatic image,PAN)。由于LRMS光谱信息丰富但空间细节不足,PAN空间精度高却缺乏光谱信息,二者的互补性使得融合生成兼具高空间分辨率和丰富光谱特征的图像成为必要手段。

早期的全色锐化方法相对成熟,主要包括成分替换法、多分辨率分析法以及基于模型的方法三大类。成分替换法是最早广泛使用的策略,例如主成分分析(principal component analysis,PCA)方法(Ghadjati等,2019)和强度—色相—饱和度(intensity-hue-saturation,IHS)方法(Tu等,2001),该类算法计算高效但容易导致光谱失真问题。多分辨率分析法(Liu,2000)采用多尺度滤波策略缓解光谱失真的问题,却容易受到滤波器的影响导致边缘区域产生伪影。基于模型的方法(Vivone等,2018)则将LRMS和PAN视为HRMS的光谱与空间退化版本,尽管理论上具有较强的解释性,但受模型假设与优化复杂度限制,实际应用效果有限。

随着深度学习在计算机视觉的广泛应用,基于深度学习的全色锐化方法逐渐成为研究热点(Deng等,2023)。该类方法能够在保持光谱一致性的同时,有效重建空间细节。Masi等人(2016)最早提出基于3层卷积的全色锐化神经网络(pansharpening neural network,PNN);随后,Wei等人(2017)结合残差网络提出深度残差全色锐化网络。此后,研究者通过更复杂的结构持续提升光谱与空间特征融合的能力,如Jin等人(2022a)的局部上下文自适应卷积网络(local-context adaptive convolution,LAGConv)。为进一步挖掘PAN与LRMS的互补性,多分支结构与注意力机制被广泛采用,如Liu等人(2020)的双流融合网络和Jin等人(2022b)的全深度特征融合网络,有效增强了特征融合能力。然而,这类方法依赖对LRMS的直接上采样,易导致伪影。为此,部分研究探索跨域特征映射策略,如Wang等人(2021)的光谱到空间卷积,实现了光谱与空间特征的直接映射。近年来,为了进一步提升性能,Hou等人(2023)提出了双域建模全色锐化网络(bidomain modeling pansharpening,BiMPan),同时处理局部光谱和全局空间信息;Zhang等人(2024)提出了小波增强信息网络(wavelet-enhanced information network,WINet),利用小波增强高频信息;Duan等人(2024)提出了自适应非局部卷积(content-adaptive non-local convolution,CANConv),引入自相似性以提高空间适应性。近期,Chen等人(2025)提出高保真全色锐化网络(high-fidelity pansharpening network,HFPan),通过CNN-Transformer编码器和双向解码器实现多尺度特征融合,并结合内容引导注意力优化光谱与空间的信息集成,在定性和定量评估中均优于现有方法。

然而,通过对HFPan方法的深入分析,本文发现其仍存在两方面不足:一是依赖直接上采样导致的伪影与空间不连续问题尚未解决;二是多分支与注意力机制虽然能部分缓解特征错配,但缺乏统一的跨模态约束机制,限制了特征对齐效果。针对这些问题,本文在保留HFPan方法优势的基础上提出了一种改进方法——基于棱镜金字塔融合(prism pyramid fusion,PPFusion)的端到端全色锐化网络。该方法构建了"差分增强—Restormer重建—层次化融合—上采样约束"的协同结构,实现了光谱与空间特征的高效互补与一致建模。

本文的主要贡献如下:1)提出了双分支注意力引导的共享上采样模块作为上采样约束分支,通过自监督一致性学习与光谱—空间双重约束,兼顾光谱保真与细节增强。2)设计了层次化内容引导的注意力融合模块,在通道与空间层面实现跨层特征融合,充分整合不同尺度下的光谱与空间信息,增强特征表达能力。3)改进了差分增强卷积模块,结合Restormer模块分别捕获局部高频细节与长程光谱依赖,实现空间细节恢复与光谱保真度的协同提升。

实验结果表明,本文提出的PPFusion在光谱角映射(spectral angle mapper,SAM)指标上具有明显优势,在其他指标上具有较强竞争力,证实了其在全色锐化任务中的有效性与可靠性。

1 本文方法

为了提升全色锐化性能，本文提出了一种面向细节保持的棱镜金字塔融合网络（PPFusion）。该方法通过跨模态引导与层次化建模，在融合过程中兼顾光谱一致性与空间清晰度。整体设计包括主干双分支特征提取、层次化内容引导融合及光谱—空间双重约束优化，实现光谱保持与细节增强的平衡。

1.1 总体网络框架

如图1所示，棱镜金字塔融合网络（PPFusion）采用“编码—融合—重建”的总体架构。输入为LRMS上采样后与PAN堆叠的特征图。主干网络由差分增强卷积分支和改进的Restormer分支（Zamir等，2022）组成，前者强化高频纹理特征，后者建模全局上下文。两条分支在3个阶段进行特征融合，得到初步融合结果。随后，层次化内容引导的注意力融合模块（hierarchical content-guided attention fusion module，HCGAF）对多尺度特征进行最终融合，输出高分辨率多光谱图像。此外，双分支注意力引导的共享上采样模块在训练中对LRMS与PAN进行辅助超分辨率，以特征一致性约束强化空间细节恢复。

总体而言，PPFusion通过“主干双分支—多阶段融合—HCGAF最终融合—共享上采样模块辅助监督”的设计，实现了更高的光谱保持与更优的空间清晰度的统一。

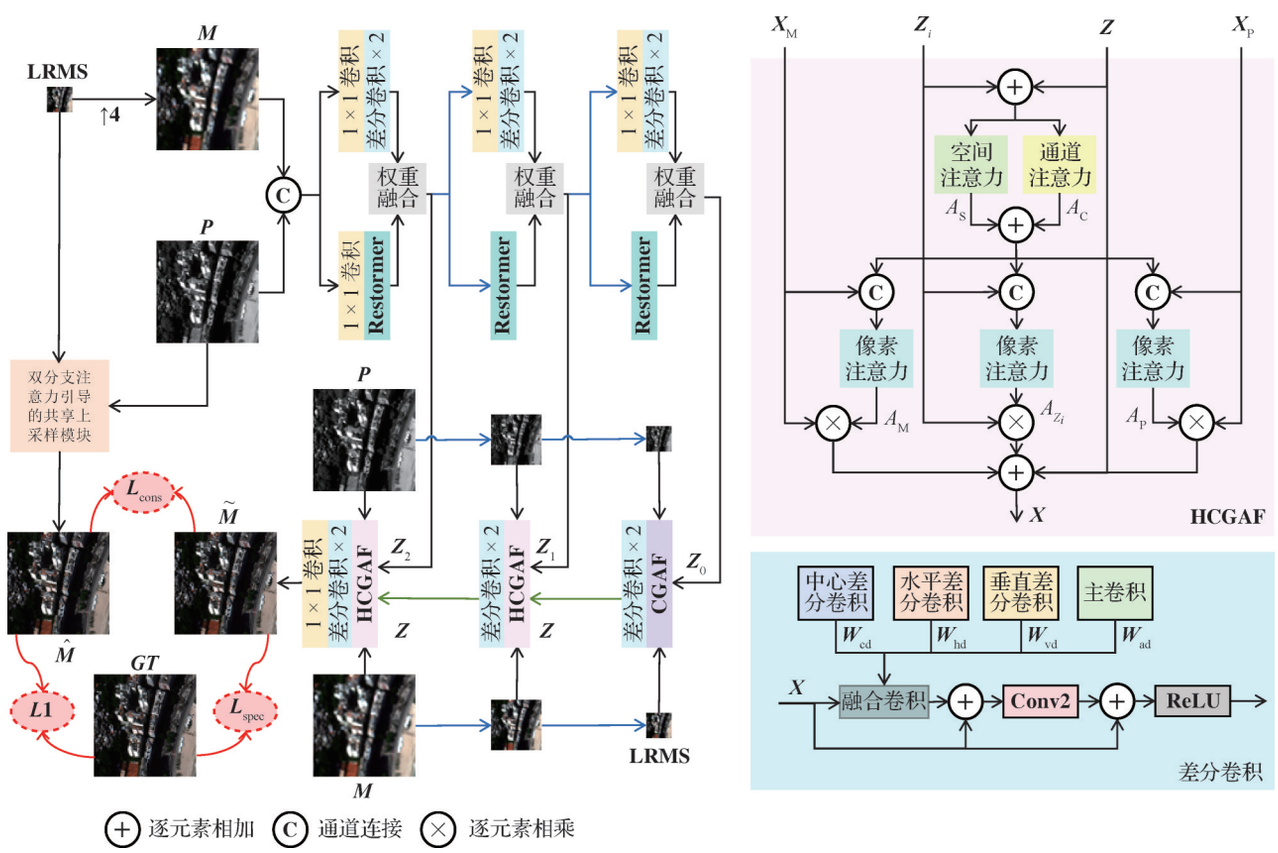


图1 本文算法网络结构图

Fig. 1 The architecture of proposed network

1.2 双分支注意力引导的共享上采样模块DASU

在全色锐化任务中，多光谱图像（multispectral image，MS）与PAN存在显著的分辨率差异。为了使MS充分利用PAN所蕴含的丰富空间结构信息，本文提出了一种双分支注意力引导的共享上采样模块（dual-branch attention-guided shared upsampling module，DASU），其核心思想是以PAN特征作为空间细节的参照，引导MS实现高质量的分辨率恢复，如图2所示。需要强调的是，DASU虽然采用双分支与共享上采样的基础框架，但其设计逻辑与通用超分辨率及现有全色锐化方法存在本质差异：一方面，通用超分辨率中的共享上采样多面向单模态特征的尺

度恢复，未考虑跨模态特征分布差异（Lin 和 Xiao，2022）；另一方面，现有全色锐化的双分支方法多为“空间—光谱”的简单分离，未引入自监督约束与特征解耦机制（Li 等，2025）。具体而言，DASU 由 PAN 分支与 MS 分支组成，并在二者之间共享同一套参数化的上采样模块以保证特征空间的一致性。

在 PAN 分支中，本文首先将高分辨率的 PAN 下采样至与 MS 相同的低分辨率尺度，并通过卷积映射得到紧凑的表征。随后，这一表征经过共享上采样模块恢复至原始尺度，并通过轻量的重建头映射回 PAN 空间。设输入的 PAN 经下采样得到的低分辨率图像表示为 $\boldsymbol{P}_{\mathrm{LR}}$，得到的重建结果为 $\hat{\boldsymbol{P}}$，则该过程表示为

$$\hat{\boldsymbol{P}} = R_{\mathrm{P}}\left(Up\left(\boldsymbol{P}_{\mathrm{LR}}\right)\right) \tag{1}$$

式中，$R_{\mathrm{P}}(\cdot)$为重建头，$Up(\cdot)$表示由转置卷积与像素重排组成的可学习上采样模块，其参数在 PAN 分支与 MS 分支中完全共享。通过在 PAN 分支中引入自监督重建约束，共享上采样模块被显式地引导学习 PAN 中稳定且通用的空间结构特征，从而为 MS 分支提供可靠的空间先验。这种基于 PAN 的自监督约束有助于缓解跨模态特征分布差异带来的冲突，使共享上采样模块更专注于尺度恢复与几何结构建模，而非模态特有信息。

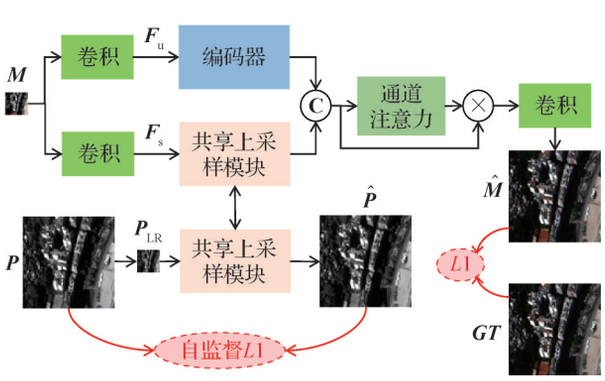


图 2　共享上采样模块

Fig. 2　Shared upsampling module

在 MS 分支中，输入的低分辨率多光谱图像首先被解耦为共享特征与独特特征两部分。共享特征通过与 PAN 分支一致的上采样模块提升分辨率，以确保两种模态在空间尺度上的对齐；独特特征则采用轻量卷积网络建模其光谱专属性，并利用双线性插值调整至目标尺寸。最终，二者在通道维度上融合，并通过通道注意力机制自适应地分配权重，得到高分辨率的 MS 重建结果。设输入的 LRMS 经解耦后得到共享特征 $\boldsymbol{F}_{\mathrm{s}}$ 与独特特征 $\boldsymbol{F}_{\mathrm{u}}$，得到的上采样结果为 $\hat{\boldsymbol{M}}$，其融合过程可表述为

$$\hat{\boldsymbol{M}} = \Phi\left(\alpha \times Up\left(\boldsymbol{F}_{\mathrm{s}}\right) \oplus (1-\alpha) \times B\left(\boldsymbol{F}_{\mathrm{u}}\right)\right) \tag{2}$$

式中，$\oplus$ 表示通道拼接，α 为自适应学习到的融合系数，$B(\cdot)$表示双线性插值，$\Phi(\cdot)$表示基于全局通道统计的通道注意力函数，用于自适应平衡空间增强特征与光谱保持特征之间的贡献。该注意力机制通过显式建模不同特征子空间的重要性，抑制由共享上采样可能引入的光谱干扰，从而在增强空间细节的同时保持多光谱通道间的光谱一致性。这种共享—自监督结构使上采样模块能够在两个模态中共享空间特征表征，从而实现跨模态对齐。

通过上述设计，本文设计的 DASU 模块能够同时保证空间细节的完整性与光谱信息的保真性：PAN 分支提供了高频结构的强约束，MS 分支保持了多光谱通道间的差异性，二者在共享上采样模块与注意力机制的协同作用下实现互补融合。

1.3　层次化内容引导的注意力融合模块 HCGAF

在复杂场景中，MS 与 PAN 之间的差异不仅体现在分辨率上，更体现在不同尺度下空间与光谱信息分布的多样性。单一尺度的上采样策略往往难以兼顾细节恢复与全局一致性。为此，受内容引导注意力融合（content-guided attention fusion module，CGAF）模块（Chen 等，2025）的启发，本文设计了一个层次化内容引导的注意力融合模块（hierarchical-CGAF，HCGAF），以解决多源特征在跨尺度对齐与细节保持上的不充分问题。与传统的 CGAF 模块不同，HCGAF 在融合过程中不仅考虑了当前尺度的特征表示 $\boldsymbol{Z}$，还引入了来自相同层次的上下文特征 $\boldsymbol{Z}_i$，从而在层次化建模的框架下提升融合的表达能力。如图 1 所示，HCGAF 以棱镜金字塔结构为核心，将不同尺度的 MS 与 PAN 特征进行逐级内容引导融合。

具体而言，设低分辨率多光谱特征为 $\boldsymbol{X}_{\mathrm{M}}$，高分辨率全色特征为 $\boldsymbol{X}_{\mathrm{P}}$，当前尺度特征为 $\boldsymbol{Z}$，上一级特征为 $\boldsymbol{Z}_i$。首先，本文构造了层次化的通道注意力权重 A_{C} 与空间注意力权重 A_{S}，具体为

$$A_{\mathrm{C}} = \sigma\left(MLP\left(GAP\left(\boldsymbol{Z} + \boldsymbol{Z}_i\right)\right)\right) \tag{3}$$

$$A_{\mathrm{S}} = \sigma\left(f_{\mathrm{conv}}\left(C\left(\mathrm{avg}\left(\boldsymbol{Z} + \boldsymbol{Z}_i\right), \max\left(\boldsymbol{Z} + \boldsymbol{Z}_i\right)\right)\right)\right) \tag{4}$$

式中，$\sigma(\cdot)$表示sigmoid激活，$GAP(\cdot)$表示全局平均池化，$f_{\text{conv}}(\cdot)$表示卷积映射，$C(\cdot)$表示沿通道拼接。通道注意力与空间注意力通过逐元素相加获得综合权重，具体为

$$A = A_{\text{C}} + A_{\text{S}} \tag{5}$$

在获得层次化引导权重后，进一步设计了像素级注意力，以便在更精细的尺度上调节特征的重要性。像素注意力的形式为

$$A_i = \sigma\left(f_{\text{pa}}\left(\left[\boldsymbol{X}, A\right]\right)\right) \tag{6}$$

式中，$\boldsymbol{X} \in \left\{\boldsymbol{X}_{\text{M}}, \boldsymbol{X}_{\text{P}}, \boldsymbol{Z}_i\right\}$，$[\cdot]$表示特征拼接，$f_{\text{pa}}(\cdot)$表示$7 \times 7$卷积实现的通道压缩，$\left\{A_i\right\}$表示得到的权重。可以看出，HCGAF在CGAF的基础上通过设计跨层次特征$\boldsymbol{Z}_i$，在通道—空间—像素3个层面上实现了联合建模与细粒度调制，使融合过程能够同时兼顾跨尺度的上下文一致性与局部细节的精确恢复。最终得到融合特征，其具体形式为

$$\boldsymbol{X} = \boldsymbol{Z} + A_{\text{M}}\boldsymbol{X}_{\text{M}} + A_{\text{P}}\boldsymbol{X}_{\text{P}} + A_{\text{Z}_i}\boldsymbol{Z}_i \tag{7}$$

总之，本文以金字塔式的层次化建模方法构建的HCGAF不仅提升了特征交互的丰富性，而且使网络能够在高层捕获全局上下文，在低层强化局部细节，从而提升融合质量。

1.4 差分增强卷积模块DEConv

在高分辨率遥感图像的重建过程中，边缘与纹理等高频信息对于空间细节的恢复具有至关重要的作用。为增强网络对局部差分信息的建模能力，本文改进了差分增强卷积（differentially enhanced convolution module，DEConv），其核心思想是将多种方向性的差分算子与标准卷积核在权重层面进行组合，以获得更具判别力的特征表征。

如图1所示，差分增强卷积通过结合水平、垂直与中心差分卷积核，增强网络对高频细节的敏感性。以中心差分卷积（center-difference convolution，CDConv）为例，其权重可表示为

$$\boldsymbol{W}_{\text{cd}}(i, j) = \boldsymbol{W}(i, j) - \frac{1}{k^2 - 1}\sum_{(u,v) \neq (i,j)} \boldsymbol{W}(u, v) \tag{8}$$

式中，$\boldsymbol{W}(i,j)$表示卷积核中心位置的权重，该操作在空间上等价于在中心位置引入负反馈。进一步地，本文设计了一种自适应差分卷积（adaptive difference convolution，ADConv），通过可学习的参数θ对邻域权重进行重排，从而使卷积核能够自适应捕获不同方向与尺度下的细节变化，具体为

$$\boldsymbol{W}_{\text{ad}} = \boldsymbol{W} - \theta \times \pi(\boldsymbol{W}) \tag{9}$$

式中，$\pi(\cdot)$表示对权重矩阵的邻域置换操作。差分增强卷积最终通过在权重层面对上述多个算子进行融合得到，即

$$\boldsymbol{W}_{\text{DE}} = \sum_{i=0}^{4} \alpha_i \boldsymbol{W}_i \tag{10}$$

式中，$\boldsymbol{W}_0 = \boldsymbol{W}$，$\boldsymbol{W}_1 = \boldsymbol{W}_{\text{cd}}$，$\boldsymbol{W}_2 = \boldsymbol{W}_{\text{hd}}$，$\boldsymbol{W}_3 = \boldsymbol{W}_{\text{vd}}$，$\boldsymbol{W}_4 = \boldsymbol{W}_{\text{ad}}$，$\left\{\alpha_i\right\}$为可学习参数，用于在训练过程中自适应地调节不同卷积核的贡献比例。

在实际网络结构中，本文进一步将DEConv融入残差块中构成DEBlock模块，使其输出能够与原始特征形成残差连接，具体为

$$\boldsymbol{X}_{i+1} = \Theta\left(\boldsymbol{W}_{\text{DE}} \times \boldsymbol{X}_i + b\right) + \boldsymbol{X}_i \tag{11}$$

式中，$\boldsymbol{X}_i \in \mathbf{R}^{\text{C} \times \text{H} \times \text{W}}$表示输入特征，$\Theta(\cdot)$表示非线性激活函数。与标准卷积相比，DEConv更能突出边缘、纹理等高频特征，从而有效缓解锐化图像的过平滑现象。

1.5 损失函数

在全色锐化任务中，理想的重建结果不仅要求保持MS的光谱一致性，还需要继承PAN的空间细节。为了同时满足这两方面的需求，本文在训练过程中设计了光谱—空间双重约束损失函数，以同时优化光谱保真度与空间结构恢复能力。

首先，光谱保持约束损失用于确保模型最终输出的融合图像在光谱维度上与真实图像保持一致。由于$L1$损失更有利于保持光谱的一致性，所以本文采用了$L1$范数作为光谱保持约束，以确保重建结果在光谱分布上与参考图像一致。设主干网络输出的重建多光谱图像为$\tilde{\boldsymbol{M}}$，对应的真实图像为$\boldsymbol{GT}$，则光谱保持约束表示为

$$L_{\text{spec}} = \left\| \tilde{\boldsymbol{M}} - \boldsymbol{GT} \right\|_1 \tag{12}$$

其次，为了进一步提升空间结构建模能力，本文在训练中设计了基于共享上采样模块的一致性约束。具体而言，设上采样模块对LRMS生成的重建结果为$\hat{\boldsymbol{M}}$，主干网络的输出为$\tilde{\boldsymbol{M}}$。通过最小化二者的差异，可以将上采样模块学习到的空间细节知识迁移至主网络，具体为

$$L_{\text{cons}} = \left\| \tilde{\boldsymbol{M}} - \hat{\boldsymbol{M}} \right\|_1 \tag{13}$$

这一约束为共享上采样模块提供了正则化信号，从而提升了网络的稳定性与空间细节恢复能力。

最终的损失函数由上述两部分加权组合而

成，即

$$L = L_{\text{spec}} + \lambda L_{\text{cons}} \tag{14}$$

式中，λ为平衡参数，用于调节光谱保持与空间一致性约束的相对重要性。在实验中，本文通过实验选择合适的参数大小，以确保网络在光谱—空间双重约束下实现最优性能。

2 实验与分析

2.1 数据集

为了验证所提出的全色锐化模型的有效性，本文采用3类国际公认的卫星数据集开展实验，分别为WorldView-3（WV3）、高分二号（GF2）及QuickBird（QB）。由于缺少参考图像作为监督，本文使用Wald协议（Wald等，1997）对源图像进行下采样生成训练数据，每个数据集均包含数千组样本，可完整支撑模型的训练、验证与测试任务，保证实验过程的可靠性与合理性。

2.1.1 训练数据集

在数据分辨率方面，本文将训练集中的多光谱图像（MS）、全色图像（PAN）及真实图像（ground trath，GT）的空间分辨率统一设置为64×64像素，而低分辨率多光谱图像（LRMS）的分辨率设定为16×16像素，该尺寸既能够保证模型充分学习图像的细节特征，又可兼顾训练效率。

2.1.2 测试数据集

本文将测试集划分为降尺度测试集与全尺度测试集，以全面评估模型性能。降尺度测试集用于快速验证基础锐化效果，MS、PAN和GT的分辨率均为256×256像素，LRMS的分辨率为64×64像素，在保证验证有效性的同时显著降低计算成本。全尺度测试集更贴近实际遥感场景，用于评估模型的高分辨率重建能力，其中MS与PAN的分辨率提升至512×512像素，LRMS的分辨率为128×128像素，以检验模型在细节恢复与边缘锐化方面的表现，更具应用参考价值。

2.2 实验设置

2.2.1 训练细节

本文中所有深度学习模型均在Linux系统，Python 3.11，PyTorch 2.5框架下训练，并使用了NVIDIA RTX 3090 GPU提供算力。训练时采用Adam优化器（$\beta_1=0.9$，$\beta_2=0.999$）。批次大小为16，初始学习率为0.000 1，整个训练过程共计500轮。

2.2.2 实验指标

为了评估模型在低分辨率数据集上的性能，采用4种常用的评估指标：峰值信噪比（peak signal-to-noise ratio，PSNR）（Horé和Ziou，2010）、光谱角映射（spectral angle mapper，SAM）（Yuhas等，1992）、总体相对误差（erreur relative globaleadimensionnelle de synthèse，ERGAS）（Wald，2002）以及结构相似性指数（structural similarity index measure，SSIM）（Nilsson和Akenine-Möller，2020）。对于全分辨率数据集的评估，使用3类指标来综合衡量重建图像的质量，分别为：无参考质量指标（quality with no reference，QNR）、光谱失真指数（spectral distortion index，D_λ）以及空间失真指数（spatial distortion index，D_s）（Alparone等，2008），以全面反映模型在空间细节恢复和光谱保持方面的表现。

实验选取12种深度学习方法作为对比，包括：PNN（Masi等，2016）、PanNet（a deep network architecture for pan-sharpening）（Yang等，2017）、FusionNet（deep convolutional neural network for fusion）（Deng等，2021）、GPPNN（gradient projection based pan-sharpening neural network）（Xu等，2021）、LAGConv（Jin等，2022a）、BiMPan（Hou，2023）、SSDiff（spatial-spectral integrated diffusion model）（Zhong等，2024）、MSAN（multiscale self-attention network）（Lu等，2025）、ADWM（adaptive dual-level weighting mechanism）（Huang等，2025）、HFPan（Chen等，2025）、ARNet（adaptive rectangular convolution network）（Wang等，2025）以及FAFormer（frequency-analysis-based transformer model）（Meng等，2025）。为了保证比较的公平性，均基于各方法作者公开的代码进行了实现，并在相同数据集对所有方法进行训练与测试，以便对比各模型在全色锐化任务中的性能表现。

2.3 实验结果

2.3.1 降尺度评估

本文方法与对比方法在WV3、GF2和QB数据集的降尺度评估对比如表1所示。实验结果表明，本文方法在3个数据集均取得了具有竞争力的性能。具体而言，本文方法在3个数据集上的SAM指标均优于现有方法，平均下降约0.3°，体现了更出色的光谱保真度；在结构相似性指标SSIM上，除QB数据集外，本文方法优于其他所有对比方法，进一步验

证了本文模型在融合图像的结构一致性与细节纹理保持方面的优势；PSNR指标在WV3、GF2和QB分别达到40.972 3、49.044 4和45.224 4，相比HFPan方法提升明显；ERGAS指标也表现稳定，进一步验证了DASU模块与HCGAF融合机制在细节恢复与光谱一致性方面的有效性。

可视化结果如图3—图5所示。可以看出，本文方法能够更清晰地重建建筑物边界和植被细节，同时保持光谱保真。均方误差（mean squared error，MSE）残差图进一步表明，本文方法的误差分布更为均匀且整体残差较低，充分验证了模型在光谱—空间一致性方面的优越性。

表1　各数据集的降尺度评估对比

Table 1　Reduced-resolution evaluation comparison across datasets

算法	WV3				GF2				QB			
	PSNR↑/dB	SAM↓	ERGAS↓	SSIM↑	PSNR↑	SAM↓	ERGAS↓	SSIM↑	PSNR↑	SAM↓	ERGAS↓	SSIM↑
PNN	38.095 5	2.367 3	4.706 2	0.980 5	43.150 7	0.994 4	1.032 2	0.973 7	43.114 1	2.473 4	4.985 2	0.972 3
PanNet	34.996 8	2.396 3	9.415 3	0.974 7	45.089 4	0.890 6	0.828 5	0.982 2	41.894 1	2.535 2	11.668 6	0.970 4
FusionNet	39.317 7	2.103 0	4.164 1	0.985 9	43.028 8	0.944 7	1.046 7	0.974 5	44.595 8	2.403 7	4.156 9	0.975 7
GPPNN	38.001 7	2.265 5	5.006 1	0.981 5	39.725 4	1.165 3	1.543 6	0.963 5	42.668 0	2.462 5	5.219 2	0.972 4
LAGConv	40.385 5	1.998 7	3.638 7	0.988 0	45.711 7	0.880 1	0.769 2	0.985 6	45.131 4	2.369 9	3.836 7	0.977 2
BiMPan	39.665 3	2.010 7	4.024 5	0.987 2	44.776 3	1.042 5	0.838 5	0.980 2	43.502 2	2.434 8	4.500 4	0.974 2
SSDiff	40.640 2	1.806 9	3.514 1	0.988 9	48.747 7	0.665 3	0.540 8	0.991 8	45.075 3	2.377 9	3.982 7	0.976 8
MSAN	40.675 9	1.799 0	3.522 0	0.988 9	47.120 1	0.785 9	0.657 9	0.989 1	45.174 3	2.358 4	3.886 9	**0.977 8**
ADWM	40.046 3	1.930 6	3.744 6	0.987 5	41.296 6	1.489 9	1.284 5	0.966 2	45.068 0	2.408 7	3.940 2	0.976 4
HFPan	40.121 7	2.303 9	3.782 8	0.988 1	44.506 3	1.205 7	0.899 7	0.989 4	45.184 0	2.473 1	3.844 7	0.977 5
ARNet	39.922 1	1.916 6	3.847 5	0.987 7	47.259 7	0.780 6	0.646 6	0.988 8	**45.257 3**	2.375 4	**3.834 0**	0.977 6
FAFormer	34.216 7	2.809 8	7.416 5	0.958 7	41.671 9	1.243 4	1.243 3	0.968 9	42.513 1	2.526 9	5.426 6	0.970 2
本文（PPFusion）	**40.972 3**	**1.773 0**	**3.417 9**	**0.989 7**	**49.044 4**	**0.656 2**	**0.530 1**	**0.992 6**	45.224 4	**2.347 9**	3.857 2	0.977 7

注：加粗、下划线字体表示各列最优、次优结果。↑表示值越高越好，↓表示值越低越好。

2.3.2　全尺度评估

本文方法与对比方法在WV3、GF2和QB数据集的全尺度评估对比如表2所示。在全尺度实验中，本文提出的PPFusion在QNR及其分量指标D_λ与D_s上整体表现优异，在多个数据集上取得与最佳方法相当或略优的结果。与对比方法相比，本文方法的QNR提升幅度约为0.01～0.02，表明其在无参考场景下仍能较好地平衡光谱与空间信息。

综上所述，本文设计的PPFusion网络在不同类型的场景下均展现出良好的稳健性与泛化能力，进一步验证了所提出的棱镜金字塔结构在提升性能方面的有效性。

2.4　消融实验

2.4.1　各模块的有效性评估

为了验证各模块对整体性能的贡献，本文在WV3数据集上进行了逐步消融实验，结果如表3所示。具体而言，逐步引入不同模块后，本文模型的性能均得到不同程度提升。当仅加入DEConv模块时，PSNR提升约0.23 dB，表明其有效增强了高频纹理信息。继续引入HCGAF模块后，光谱一致性得到保持，体现了跨尺度特征交互的优势。进一步结合DASU模块后，本文模型在3项指标上均达到最优，显著提升了融合质量。

综合分析可知，DEConv模块增强高频纹理特征，并改善空间清晰度；HCGAF模块实现光谱与空间信息的协调融合；DASU模块强化空间细节重建的连续性。三者协同作用，使模型在光谱保真与空间分辨率之间实现最佳平衡。

2.4.2　损失函数影响分析

为了评估一致性约束的作用，本文在固定网络

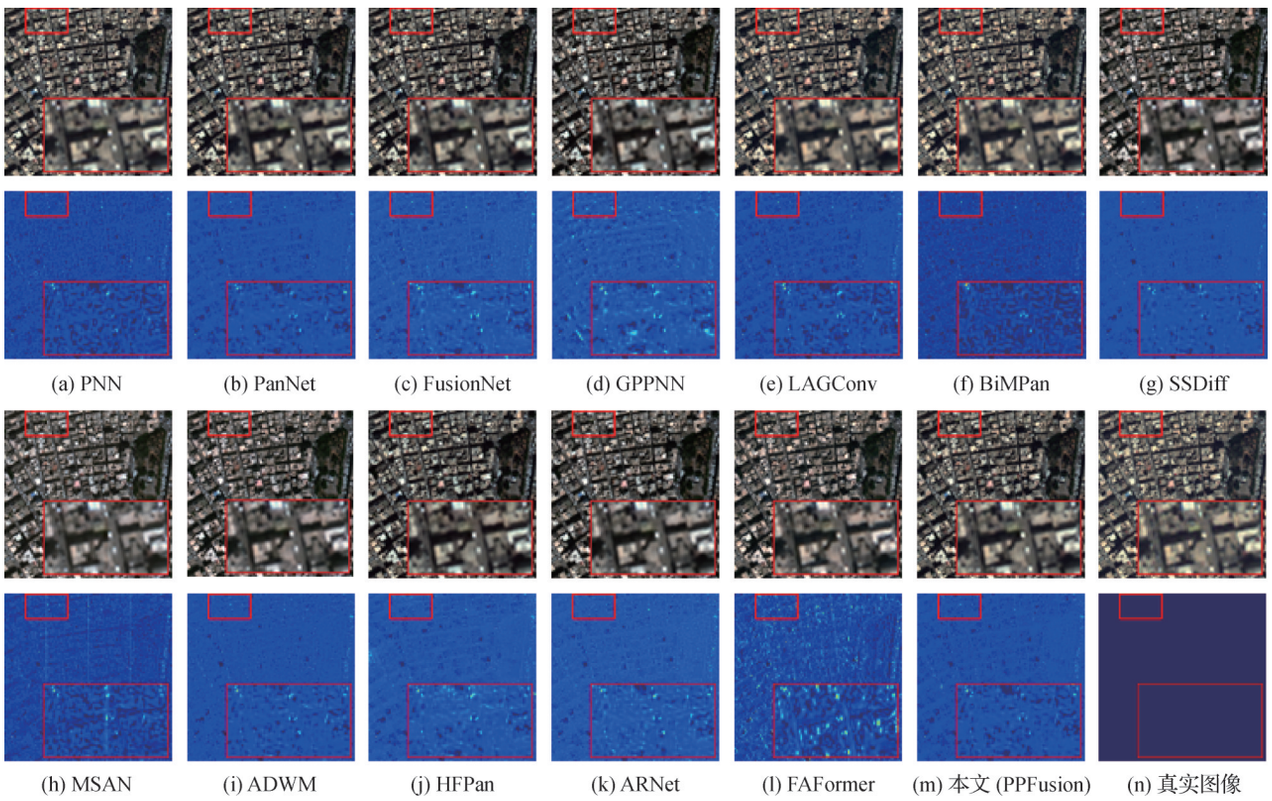


图 3　WV3 可视化结果

Fig. 3　Visualization results of WV3((a)PNN;(b)PanNet;(c)FusionNet;(d)GPPNN;(e)LAGConv;(f)BiMPan;(g)SSDiff;(h)MSAN;(i) ADWM;(j)HFPan;(k)ARNet;(l)FAFormer;(m)ours;(n)GT)

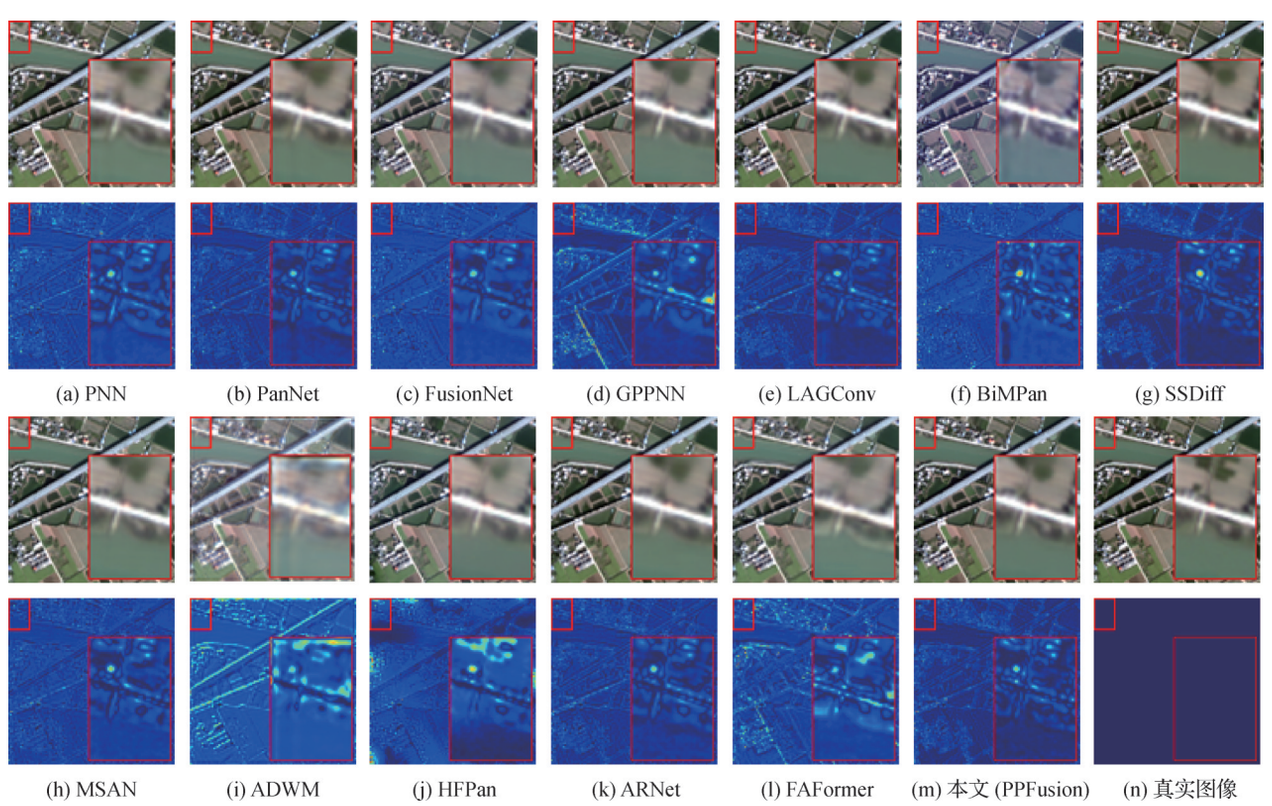


图 4　GF2 可视化结果

Fig. 4　Visualization results of GF2((a)PNN;(b)PanNet;(c)FusionNet;(d)GPPNN;(e)LAGConv;(f)BiMPan;(g)SSDiff;(h)MSAN;(i) ADWM;(j)HFPan;(k)ARNet;(l)FAFormer;(m)ours;(n)GT)

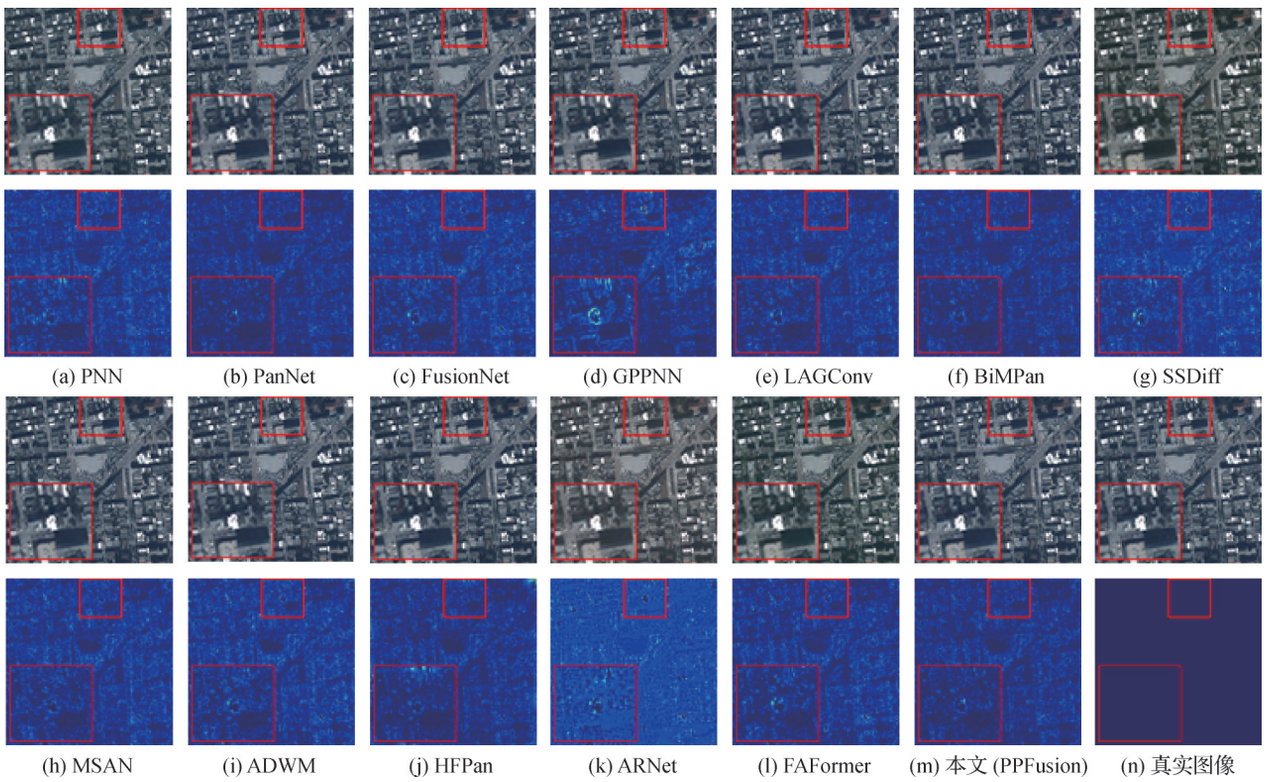


图5　QB可视化结果

Fig. 5　Visualization results of QB((a)PNN;(b)PanNet;(c)FusionNet;(d)GPPNN;(e)LAGConv;(f)BiMPan;(g)SSDiff;(h)MSAN;(i) ADWM;(j)HFPan;(k)ARNet;(l)FAFormer;(m)ours;(n)GT)

表2　各数据集的全尺度指标评估对比

Table 2　Full-resolution evaluation comparison across datasets

算法	WV3			GF2			QB		
	$D_\lambda\downarrow$	$D_s\downarrow$	QNR↑	$D_\lambda\downarrow$	$D_s\downarrow$	QNR↑	$D_\lambda\downarrow$	$D_s\downarrow$	QNR↑
PNN	0.034 8	0.105 8	0.863 0	0.023 5	0.028 1	0.948 9	0.034 3	0.103 7	0.865 6
PanNet	0.045 5	0.134 1	0.826 6	0.022 7	0.032 0	0.946 0	**0.024 1**	0.081 9	0.895 9
FusionNet	0.041 8	0.091 7	0.870 3	0.018 9	0.025 8	0.955 8	0.040 7	0.094 1	0.869 0
GPPNN	0.042 0	0.068 7	0.892 2	0.024 7	**0.024 3**	0.951 6	0.049 9	0.059 3	0.893 8
LAGConv	0.039 5	0.084 6	0.879 3	0.027 0	0.038 7	0.935 4	0.045 9	0.065 0	0.892 1
BiMPan	0.066 8	0.153 1	0.790 3	**0.004 7**	0.025 0	**0.970 4**	0.039 4	0.067 3	0.895 9
SSDiff	0.032 5	0.066 9	0.902 8	0.018 7	0.030 8	0.951 0	0.034 8	0.061 2	0.906 1
MSAN	**0.030 6**	0.110 1	0.862 6	0.019 8	0.031 0	0.949 8	0.035 1	0.053 4	0.913 3
ADWM	0.037 0	0.078 5	0.887 4	0.025 4	0.029 0	0.946 3	0.048 6	0.060 9	0.893 4
HFPan	0.033 3	0.073 8	0.895 4	0.023 0	0.033 7	0.944 1	0.044 4	0.038 3	0.919 0
ARNet	0.045 7	0.064 9	0.892 3	0.026 8	0.035 4	0.938 8	0.049 7	0.061 8	0.891 5
FAFormer	0.048 4	**0.046 4**	**0.907 4**	0.027 1	0.030 5	0.943 2	0.042 3	0.137 2	0.826 3
本文(PPFusion)	0.035 5	0.084 3	0.883 1	0.017 5	0.031 1	0.951 9	0.037 1	**0.029 3**	**0.934 6**

注:加粗字体表示各列最优结果。↑表示值越高越好,↓表示值越低越好。

表 3　消融实验结果

Table 3　The results of ablation experiments

配置	DEConv	HCGAF	DASU	PSNR/dB ↑	SAM ↓	ERGAS ↓
1	×	×	×	40.463 1	2.139 3	3.583 9
2	√	×	×	40.695 5	1.846 2	3.539 6
3	√	√	×	40.838 7	1.786 8	3.476 4
本文	√	√	√	**40.972 3**	**1.773 0**	**3.417 9**

注：加粗字体表示各列最优结果。“√”表示使用对应模块，“×”表示未使用对应模块。↑表示值越高越好，↓表示值越低越好。

结构的前提下，对损失函数中平衡参数 λ 的取值进行了对比实验，结果如表 4 所示。从表 4 可以看出，当 λ 取值过小时（如 0.003），模型对空间细节的约束较弱且空间纹理恢复能力受限；当 λ 过大（如 0.1）时，光谱一致性受到破坏。综合各项指标表现，当 $\lambda = 0.01$ 时，PPFusion 模型性能最优，说明此时在光谱保真与空间一致性之间实现了较好的平衡，也验证了一致性约束设计的合理性与有效性。

表 4　损失函数参数分析

Table 4　Loss function parameter analysis

λ	PSNR/dB ↑	SAM ↓	ERGAS ↓
0.003	40.907 1	**1.752 5**	3.442 2
0.01	**40.972 3**	1.773 0	**3.417 9**
0.03	40.886 5	1.888 6	3.438 2
0.1	40.519 3	1.870 7	3.571 1

注：加粗字体表示各列最优结果。↑表示值越高越好，↓表示值越低越好。

2.4.3　HCGAF 模块输入特征必要性验证

为了验证 HCGAF 模块中 LRMS、PAN、初步融合结果及前一阶段特征的必要性，在 WV3 数据集上设计消融实验：以完整模块为基准，采用控制变量法依次将任一输入替换为同维度全零向量构建消融模型，其余设置保持一致以保障实验公平性。实验结果如表 5 所示，移除任一输入特征后模型性能均出现退化，其中无前一阶段特征时退化最显著；完整 HCGAF 模块在所有指标上均最优。这证明 4 个输入特征均为必要组件，它们从不同维度形成协同作用，保障了本文提出的 HCGAF 模块层次化内容引导的最优融合性能。

2.4.4　HCGAF 与 Transformer 层级交互结构对比实验

为了验证 HCGAF 层次化内容引导融合机制的优势，本节构建 Transformer 层级交互结构作为对比基准。通过设计功能对等的 Transformer 融合模块，在相同实验条件下对比两者的融合性能，明确 HCGAF 在特征交互有效性上的优势所在。实验选用 WV3 数据集，保持与主实验及消融实验一致的训练参数与评估指标，确保对比的公平性。实验结果如表 6 所示，HCGAF 各项核心指标均显著优于 Transformer，其中，PSNR 提升 0.654 8 dB、SAM 降低 0.126 6、ERGAS 降低 0.261 8、SSIM 提升 0.001 2。上述两者的核心差异在于交互机制：Transformer 的全局自注意力未区分特征优先级，难以精准捕捉有效关联；而 HCGAF 的层次化引导机制可按特征属性渐进优化，聚焦关键信息，故在多光谱融合任务中更具优势，能更好地平衡光谱保真与空间细节保留。

表 5　HCGAF 输入必要性实验结果

Table 5　Experimental results of HCGAF input necessity

配置	PSNR/dB ↑	SAM ↓	ERGAS ↓	SSIM ↑
不含 $\boldsymbol{X}_{\mathrm{M}}$	40.457 2	1.909 3	3.574 1	0.988 7
不含 $\boldsymbol{X}_{\mathrm{P}}$	40.741 6	1.828 8	3.502 3	0.989 1
不含 $\boldsymbol{Z}$	40.662 1	1.883 9	3.526 1	0.988 9
不含 $\boldsymbol{Z}_i$	40.391 1	1.861 9	3.642 7	0.988 4
本文	**40.972 3**	**1.773 0**	**3.417 9**	**0.989 7**

注：加粗字体表示各列最优结果。↑表示值越高越好，↓表示值越低越好。

表 6　HCGAF 与 Transformer 的对比

Table 6　Comparison of HCGAF and Transformer

方法	PSNR/dB ↑	SAM ↓	ERGAS ↓	SSIM ↑
Transformer	40.317 5	1.899 6	3.679 7	0.988 5
本文	**40.972 3**	**1.773 0**	**3.417 9**	**0.989 7**

注：加粗字体表示各列最优结果。↑表示值越高越好，↓表示值越低越好。

2.4.5　计算复杂度分析

为了全面评估模型的实用性与部署潜力，本节从参数量（Params）、浮点运算量（floating point operations，FLOPs）及推理时间3个核心维度，对本文提出的PPFusion与主流全色锐化方法进行了定量对比，实验结果如表7所示。由表7可知，本文提出的PPFusion参数量为1 368.90 K，高于LAGConv、ADWM等轻量级模型，但远低于ARNet、FAFormer等大参数量模型，参数量增长源于HCGAF与DASU模块的创新设计。在计算效率上，PPFusion的浮点运算量仅为3.26 G，显著优于BiMPan、FAFormer、HFPan等方法，仅略高于LAGConv。PPFusion的推理时间为22.39 ms，与LAGConv和FusionNet相当，远优于ARNet和FAFormer等大参数量模型，进一步验证了模型的高效性。本文通过DEConv轻量化设计、HCGAF注意力稀疏计算及DASU弱约束策略，实现了高精度—低复杂度的平衡。

表7　计算复杂度实验结果

Table 7　The results of computational complexity

算法	参数量/K	浮点运算量/G	推理时间/ms
PNN	104.36	6.83	2.54
PanNet	79.43	5.22	2.41
FusionNet	506.27	11.86	21.93
GPPNN	119.30	7.82	5.19
LAGConv	151.40	2.07	22.64
BiMPan	407.42	32.34	45.78
SSDiff	1 413.94	3.23	5.50
MSAN	2 621.45	3.93	5.19
ADWM	87.26	5.14	2.87
HFPan	707.10	19.19	16.34
ARNet	15 923.72	15.70	59.34
FAFormer	2 338.37	39.03	118.65
本文（PPFusion）	1 368.90	3.26	22.39

3　结　论

本文针对传统全色锐化方法在空间细节恢复与光谱保真之间难以平衡的问题，提出了一种基于棱镜金字塔融合（PPFusion）的端到端全色锐化网络。本文提出的PPFusion通过差分增强卷积模块（DEConv）与Restormer双分支结构实现多阶段特征融合，利用层次化内容引导的注意力融合模块（HCGAF）增强多尺度特征交互与跨模态信息耦合，同时引入双分支注意力引导的共享上采样模块（DASU）作为辅助监督，以空间先验约束提升高频细节重建能力。与其他方法相比，实验结果表明本文提出的PPFusion网络框架在WV3、GF2和QB等典型遥感数据集上均取得了优异的定量与视觉表现，在光谱一致性与空间清晰度之间实现了良好的平衡，验证了所提方法的有效性与鲁棒性。

然而，模型在复杂地物场景下仍存在一定的细节过平滑问题，主要原因在于场景中高频细节分布复杂，部分精细纹理难以完整保留。此外，当前实验尚未涉及跨传感器遥感数据的适配性验证，后续工作将重点关注轻量化结构设计与跨传感器泛化能力的提升，以进一步推动全色锐化技术在高分辨率遥感影像重建中的应用。

参考文献（References）

Alparone L, Aiazzi B, Baronti S, Garzelli A, Nencini F and Selva M. 2008. Multispectral and panchromatic data fusion assessment without reference. Photogrammetric Engineering and Remote Sensing, 74(2): 193-200 [DOI: 10.14358/PERS.74.2.193]

Chen L H, Song T X, Jian L H, Zhang D, Vivone G and Zhou X C. 2025. High-fidelity pansharpening via trigeminal pyramid decoding of CNN-Transformer encoded features. IEEE Transactions on Geoscience and Remote Sensing, 63: #5404316 [DOI: 10.1109/TGRS.2025.3582642]

Cheng G and Han J W. 2016. A survey on object detection in optical remote sensing images. ISPRS Journal of Photogrammetry and Remote Sensing, 117: 11-28 [DOI: 10.1016/j. isprsjprs. 2016.03.014]

Deng L J, Ran R, Wu X and Zhang T J. 2023. CNN-based remote sensing pan-sharpening: a critical review. Journal of Image and Graphics, 28(1): 57-79 (邓良剑, 冉燃, 吴潇, 张添敬. 2023. 遥感图像全色锐化的卷积神经网络方法研究进展. 中国图象图形学报, 28(1): 57-79) [DOI: 10.11834/jig.220540]

Deng L J, Vivone G, Jin C and Chanussot J. 2021. Detail injection-based deep convolutional neural networks for pansharpening. IEEE Transactions on Geoscience and Remote Sensing, 59(8): 6995-7010 [DOI: 10.1109/TGRS.2020.3031366]

Duan Y L, Wu X, Deng H Y and Deng L J. 2024. Content-adaptive non-local convolution for remote sensing pansharpening//Proceedings of

2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 27738-27747 [DOI: 10.1109/CVPR52733.2024.02620]

Foody G M. 2003. Remote sensing of tropical forest environments: towards the monitoring of environmental resources for sustainable development. International Journal of Remote Sensing, 24(20): 4035-4046 [DOI: 10.1080/01431160310001 03853]

Ghadjati M, Moussaoui A and Boukharouba A. 2019. A novel iterative PCA-based pansharpening method. Remote Sensing Letters, 10(3): 264-273 [DOI: 10.1080/2150704X.2018.1547443]

Horé A and Ziou D. 2010. Image quality metrics: PSNR vs. SSIM//Proceedings of the 20th International Conference on Pattern Recognition. Istanbul, Turkey: IEEE: 2366-2369 [DOI: 10.1109/ICPR.2010.579]

Hou J M, Cao Q, Ran R, Liu C, Li J L and Deng L J. 2023. Bidomain modeling paradigm for pansharpening//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada: Association for Computing Machinery: 347-357 [DOI: 10.1145/3581783.3612188]

Huang J, Chen H R, Ren J X, Peng S R and Deng L J. 2025. A general adaptive dual-level weighting mechanism for remote sensing pansharpening//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 7447-7456 [DOI: 10.1109/CVPR52734.2025.00698]

Jin Z R, Zhang T J, Jiang T X, Vivone G and Deng L J. 2022a. LAGConv: local-context adaptive convolution kernels with global harmonic bias for pansharpening//Proceedings of the 36th AAAI Conference on Artificial Intelligence. [s.l.]: AAAI: 1113-1121 [DOI: 10.1609/aaai.v36i1.19996]

Jin Z R, Zhuo Y W, Zhang T J, Jin X X, Jing S Q and Deng L J. 2022b. Remote sensing pansharpening by full-depth feature fusion. Remote Sensing, 14(3): #466 [DOI: 10.3390/rs14030466]

Li Z C, Gong X Q, Ma A L, He H Q, Lv P Y, Liu X S, et al. 2025. DCCPNet: a dual-branch channel cross-concatenation pansharpening network for satellite remote sensing imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18: 14132-14147 [DOI: 10.1109/JSTARS.2025.3574743]

Lin R R and Xiao N F. 2022. Dual projection fusion for reference-based image super-resolution. Sensors, 22(11): #4119 [DOI: 10.3390/s22114119]

Liu J G. 2000. Smoothing filter-based intensity modulation: a spectral preserve image fusion technique for improving spatial details. International Journal of Remote Sensing, 21(18): 3461-3472 [DOI: 10.1080/014311600750037499]

Liu X Y, Liu Q J and Wang Y H. 2020. Remote sensing image fusion based on two-stream fusion network. Information Fusion, 55: 1-15 [DOI: 10.1016/j.inffus.2019.07.010]

Lu H Y, Yang Y, Huang S Y, Liu R X and Guo H M. 2025. MSAN: multiscale self-attention network for pansharpening. Pattern Recognition, 162: #111441 [DOI: 10.1016/j.patcog.2025.111441]

Masi G, Cozzolino D, Verdoliva L and Scarpa G. 2016. Pansharpening by convolutional neural networks. Remote Sensing, 8(7): #594 [DOI: 10.3390/rs8070594]

Meng Y F, Zhu H, Yi X Y, Hou B, Wang S, Wang Y, et al. 2025. FAFormer: frequency-analysis-based transformer focusing on correlation and specificity for pansharpening. IEEE Transactions on Geoscience and Remote Sensing, 63: #5403413 [DOI: 10.1109/TGRS.2025.3568527]

Nilsson J and Akenine-Möller T. 2020. Understanding SSIM [EB/OL]. [2025-11-05]. https://arxiv.org/pdf/2006.13846.pdf

Nogueira K, Penatti O A B and Dos Santos J A. 2017. Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognition, 61: 539-556 [DOI: 10.1016/j.patcog.2016.07.001]

Tu T M, Su S C, Shyu H C and Huang P S. 2001. A new look at IHS-like image fusion methods. Information Fusion, 2(3): 177-186 [DOI: 10.1016/S1566-2535(01)00036-7]

Vivone G, Restaino R and Chanussot J. 2018. Full scale regression-based injection coefficients for panchromatic sharpening. IEEE Transactions on Image Processing, 27(7): 3418-3431 [DOI: 10.1109/TIP.2018.2819501]

Wald L. 2002. Data fusion. Definitions and architectures — fusion of images of different spatial resolutions. Paris: Les Presses de l'École des Mines

Wald L, Ranchin T and Mangolini M. 1997. Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images. Photogrammetric Engineering and Remote Sensing, 63(6): 691-699

Wang X Y, Zheng Z X, Shao J D, Duan Y L and Deng L J. 2025. Adaptive rectangular convolution for remote sensing pansharpening//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 17872-17881 [DOI: 10.1109/CVPR52734.2025.01665]

Wang Y D, Deng L J, Zhang T J and Wu X. 2021. SSconv: explicit spectral-to-spatial convolution for pansharpening//Proceedings of the 29th ACM International Conference on Multimedia. [s.l.]: ACM: 4472-4480 [DOI: 10.1145/3474085.3475600]

Wei Y C, Yuan Q Q, Shen H F and Zhang L P. 2017. Boosting the accuracy of multispectral image pansharpening by learning a deep residual network. IEEE Geoscience and Remote Sensing Letters, 14(10): 1795-1799 [DOI: 10.1109/LGRS.2017.2736020]

Xu S, Zhang J S, Zhao Z X, Sun K, Liu J M and Zhang C X. 2021. Deep gradient projection networks for pan-sharpening//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 1366-1375 [DOI: 10.1109/CVPR46437.2021.00142]

Yang J F, Fu X Y, Hu Y W, Huang Y, Ding X H and Paisley J. 2017. PanNet: a deep network architecture for pan-sharpening//Proceed-

ings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 1753-1761 [DOI: 10.1109/ICCV.2017.193]

Yuhas R H, Goetz A F H and Boardman J W. 1992. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm [EB/OL]. [2025-11-05]. https://ntrs. nasa. gov/api/citations/19940012238/downloads/19940012238.pdf

Zamir S W, Arora A, Khan S, Hayat M, Khan F S and Yang M H. 2022. Restormer: efficient transformer for high-resolution image restoration//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 5718-5729 [DOI: 10.1109/CVPR52688.2022.00564]

Zhang J, He X H, Yan K Y, Cao K, Li R, Xie C J, et al. 2024. Pan-sharpening with wavelet-enhanced high-frequency information. IEEE Transactions on Geoscience and Remote Sensing, 62: #5402914 [DOI: 10.1109/TGRS.2024.3367165]

Zhong Y, Wu X, Cao Z H, Dou H X and Deng L J. 2024. SSDiff: spatial-spectral integrated diffusion model for remote sensing pan-sharpening//Proceedings of the 38th International Conference on Neural Information Processing Systems. Vancouver, Canada: Curran Associates, Inc.: 77962-77986 [DOI: 10.52202/079017-2478]

作者简介

黄珊珊，女，本科生，主要研究方向为遥感图像全色锐化。
E-mail：2312190503@pop.zjgsu.edu.cn
刘春晓，通信作者，男，副教授，主要研究方向为视觉计算与计算机图形学、深度学习与智能系统、视觉安全与隐私保护。
E-mail：cxliu@mail.zjgsu.edu.cn