中图法分类号:TP751　　**文献标识码**: A　　**文章编号**: 1006-8961(XXXX)XX-0001-15

论文引用格式: Wang Baixiang, Huo Hongtao, Zheng Bowen, Li Zhiqian. Hierarchical Clustering Guided and Global Context Enhanced Pan-sharpening Network[J/OL]. Journal of Image and Graphics, XXXX:1-15. DOI: 10.11834/jig.260115. (王柏翔, 霍宏涛, 郑博文, 李志倩. 层次聚类引导与全局上下文增强的全色锐化网络[J/OL]. 中国图象图形学报,XXXX:1-15. DOI: 10.11834/jig.260115.)[DOI:10.11834/jig.260115]

层次聚类引导与全局上下文增强的全色锐化网络

王柏翔,霍宏涛,郑博文,李志倩

中国人民公安大学,北京　100038

摘　要: **目的**　全色锐化技术旨在充分利用高分辨率全色影像的空间细节与低分辨率多光谱影像的光谱信息,生成兼具高清纹理细节与准确光谱的融合图像。传统方法受局部感受野限制,难以对地物结构进行差异化建模,导致复杂场景下融合结果出现光谱畸变、细节模糊等问题。本文借助层次化聚类方法,将地物先验知识运用到特征提取过程,并通过全局上下文增强机制,实现了空间细节注入与光谱保真的协同优化。**方法**　提出了一种层次聚类与全局上下文增强的全色锐化网络(HCPNet)。网络首先通过层次聚类估计的簇数初始化K-means,以得到同质区域先验,从而引导差异化卷积与特征路由;随后再引入全局上下文增强块(EFT Block),利用自注意力建模长程依赖,以强化大尺度结构一致性。训练方面,联合使用聚类一致性、光谱角与重建损失进行约束,确保网络训练的稳定性。**结果**　在降分辨率评估中,Quick Bird数据集上SAM为0.114、ERGAS为3.673、Q_{avg}为0.9919,相较次优方法,SAM与ERGAS分别提高3.4%和4.0%,Q_{avg}提高1.6%;在GaoFen-2数据集上SAM为0.023、ERGAS为0.633,相较次优方法,SAM与ERGAS分别提高8.0%和1.2%。定性实验结果表明,融合图像道路与建筑边缘更清晰,边界细节更连贯,有效抑制了纹理增强可能引发的局部色彩偏移。**结论**　HCPNet通过将区域感知融合的聚类先验与高效的基于注意力的全局上下文增强模块,兼顾空间细节增强与光谱保真,适用于复杂地物场景的高保真全色锐化。

关键词: 全色锐化;层次聚类;全局上下文;自注意力;区域感知融合;光谱保真

Hierarchical Clustering Guided and Global Context Enhanced Pan-sharpening Network

Wang Baixiang, Huo Hongtao, Zheng Bowen, Li Zhiqian

People's Public Security University of China, Beijing 100038, China

Abstract: **Objective** Pan-sharpening aims to fuse a high-resolution panchromatic (PAN) image with a low-resolution multispectral (LRMS) image to produce a high-resolution multispectral (HRMS) product. The key difficulty lies in injecting spatial textures from PAN without violating the spectral relationships among multispectral bands. Maintaining this balance becomes more challenging in heterogeneous scenes where land covers vary within a small neighborhood and structural edges cross semantic boundaries. Many deep learning approaches adopt convolutional neural networks (CNNs) and learn a global mapping from the concatenated inputs, but the locality of convolution and the lack of explicit region priors often lead to over-sharpening, ringing, and color shifts, especially around boundaries and in shadow and vegetation areas. In addition, sensor-dependent factors such as blur, noise, and band responses make it hard to transfer a fixed sharpening pattern across scenes. Hence, incorporating both region-aware priors and long-range context is important for stable spatio-spectral fusion.

收稿日期:2026-03-04;**修回日期**:2026-04-29

基金项目:高分辨率地球观测计划项目(项目编号:01-Y30F05-9001-20/22, GFZX0404130307)

Supported by:The Program of High Resolution Earth Observation under Grant

Methods We present a hierarchical clustering-guided pan-sharpening network with global context enhancement, referred to as HCPNet. The network follows an encoder-fusion-decoder design and predicts an HRMS output by adding a learned high-frequency residual to the LRMS image upsampled to the PAN resolution. Skip connections pass shallow textures to the decoder, and residual learning stabilizes optimization and reduces the risk of over-sharpening. The guidance mask used by the proposed method is computed directly from LRMS spectra and does not require extra annotations. Two complementary ingredients are introduced. First, region-aware fusion is driven by a hierarchical clustering prior. Instead of treating all pixels equally, we construct an explicit homogeneous-region guidance mask from the LRMS spectra. To obtain reliable partitions in complex scenes, we adopt a hybrid strategy that combines agglomerative hierarchical clustering (agglomerative nesting, AGNES) and K-means. The hierarchical stage builds a dendrogram using spectral angle distance and Ward linkage, and an adaptive cluster number is selected according to clustering validity (Davies-Bouldin index) and the change rate across cut heights. The resulting cluster centroids are then used to initialize K-means, producing pixel-wise cluster labels that are reshaped into the guidance mask. This prior is injected into the backbone to route features and to realize differential processing: pixels within a cluster share convolutional responses to preserve spectral consistency in homogeneous areas, while inter-cluster interactions are handled through residual fusion to avoid block artifacts and to prevent the propagation of sharpening errors across boundaries. In the shallow encoder, a Hybrid-H block leverages relatively stable low-level features to generate a reliable partition and an index map; in deeper layers, a Hybrid-D block updates routing with lower overhead to adapt to feature distribution shifts as depth increases. A patch-centroid representation for cluster prototypes and a lightweight dynamic adjustment mechanism are used to improve stability across different scene complexities. In our implementation, the deep routing uses 32 clusters with a small filtering threshold of 0.005. Second, global context enhancement is achieved via an Efficient Feature Transformer (EFT) block. To overcome the limited receptive field of convolution, EFT augments local features with long-range dependencies using spatial-reduction multi-head self-attention. Convolutional features are projected to query, key, and value embeddings; attention weights are computed with scaled dot-product similarity and normalized by softmax; and the attended feature is fused back through residual connections. Spatial reduction is applied when computing keys and values, which lowers complexity while preserving coarse global structures. The resulting context feature helps propagate cues across distant regions, improves large-scale structural coherence, and reduces boundary artifacts caused by purely local sharpening. In the experiments, a lightweight configuration is used, including a single attention head, a small feed-forward expansion, and dropout regularization, to keep the additional cost moderate. The network is optimized with a multi-constraint objective. The total loss is a weighted sum of three terms: a spectral-angle loss encouraging the predicted and reference spectra to have similar directions, a clustering-consistency loss aligning feature distributions with cluster prototypes (including a symmetric Kullback-Leibler divergence term and a compactness term), and a reconstruction loss penalizing pixel-wise absolute errors. These terms jointly supervise spectral shape, region coherence, and fidelity of the reconstructed details. During training, images are normalized to the sensor radiometric range, and reduced-resolution samples are generated following the Wald protocol. **Results** Extensive experiments are conducted on the Quick Bird and GaoFen-2 datasets. Under the reduced-resolution Wald protocol, where an HRMS reference is available, we report spectral angle mapper (SAM), ERGAS (Erreur Relative Globale Adimensionnelle de Synthèse), and mean universal image quality index averaged over bands. HCPNet consistently improves these metrics across scenes and sensors. On Quick Bird, it increases SAM and ERGAS by 3.4% and 4.0% and increases Q_{avg} by 1.6% compared with the second-best baseline; on GaoFen-2, it increases SAM and ERGAS by 8.0% and 1.2% . Across both datasets, improvements are observed in urban and rural scenes, and the method avoids noticeable spectral shifts in vegetation and shadow regions. Qualitatively, the proposed method produces sharp yet clean edges and maintains consistent colors in homogeneous regions; it also suppresses ringing and halo effects near high-contrast boundaries, where many learning-based methods tend to over-inject high-frequency components. Ablation studies verify the complementary roles of the two key components: removing the clustering guidance increases boundary artifacts and degrades spectral stability, while removing EFT mainly harms cross-region consistency and large-scale structural coherence. Because clustering is performed on low-dimensional spectra and EFT uses spatial reduction, the overall computational overhead remains manageable in practice. **Conclusion** By combining an explicit hierarchical clustering prior for region-aware fusion with an efficient

attention-based global context enhancement module, HCPNet improves pan-sharpening quality in heterogeneous remote-sensing scenes, delivering clearer boundaries with lower spectral distortion.

Key words: pan-sharpening; hierarchical clustering; global context; self-attention; region-aware fusion; spectral preservation

论文引用格式:[DOI:10.11834/jig.260115]

0 引　言

高分辨率多光谱遥感影像在资源调查、环境监测、精准农业及城市规划等领域发挥着不可替代的作用(Li等,2017)。然而,受成像硬件限制,单一传感器难以同时获得高空间分辨率与高光谱分辨率的图像。目前的对地观测卫星通常提供两类数据:一是全色(panchromatic,PAN)图像,其空间分辨率高但仅含单波段信息;二是多光谱(multispectral,MS)图像,光谱信息丰富但空间分辨率较低。为解决单一传感器获取信息不全的问题,全色锐化(pan-sharpening)技术应运而生。

全色锐化技术的目的是将全色图像的空间细节注入多光谱图像中,融合生成兼具高空间分辨率与高光谱保真度的高质量影像(Kaur等,2021)。传统的全色锐化方法主要分为成分替换(component substitution,CS)和多分辨率分析(multiresolution analysis,MRA)两大类(Vivone等,2015)。CS类方法将多光谱图像投影至新的特征空间,利用全色图像替换其中的空间分量来实现融合(Aiazzi等,2007)。此类方法计算效率高且空间细节增强效果显著,但该方法假设“全色图像是多光谱各波段的线性组合”,这在复杂地物场景下往往难以成立,极易引发光谱失真。相比之下,MRA类方法采用多尺度分解策略提取全色图像的高频信息并注入多光谱图像中(Garzelli和Nencini,2009)。尽管此类方法在光谱保真度方面表现较好,但往往存在高频细节注入不足的问题,容易在融合图像的边缘处产生吉布斯效应(Gibbs phenomenon)。因此,如何在有效提升空间分辨率的同时,最大程度地保持原始光谱特征,仍是传统方法面临的核心难题。

近年来,基于深度学习的全色锐化算法凭借其强大的非线性特征拟合能力,逐步取代了传统方法(Masi等,2016)。在深度学习早期,研究人员主要利用卷积神经网络(convolutional neural networks,CNN)构建端到端的映射模型,例如Masi等人(2016)提出的三层卷积结构,直接学习低分辨率多光谱图像到高分辨率图像的非线性映射,为数据驱动方法奠定了基础。为进一步提升高频细节恢复能力,残差学习机制被引入以构建更深层网络,有效缓解了梯度消失问题并增强了特征表达能力(Wei等,2017)。与此同时,一些方法转向高通滤波域进行网络训练,使网络更专注于恢复图像的高频纹理残差(Yang等,2017)。

随着深度学习的发展,网络结构由单流逐步发展为多分支架构,双流融合网络分别提取全色图像的空间特征与多光谱图像的光谱特征,再在深层进行特征交互以降低光谱失真(Liu等,2020)。为进一步利用网络各层特征,Jin等人(2022b)提出了全深度特征融合策略,充分挖掘不同层级的信息以提升重建质量。尽管CNN在局部特征提取方面表现优异,但其感受野有限,难以进行跨区域上下文关联。为此,Huang等人(2025)提出小波辅助多频注意力网络,结合频域分析以增强对多频特征的感知能力。为更精细地平衡空间与光谱信息,该研究还设计了一种通用的双层加权机制,实现了权重自适应优化(Huang等,2025a)。针对多光谱图像特性,Hou等人(2025)引入双边自适应演化Transformer架构,以提升模型对复杂地物细节的建模能力。进一步的,研究人员设计采用了全局自注意力机制,以解决网络长程依赖不足的问题(李妙宇和付莹,2023)。另一方面,为处理Transformer计算复杂度随序列长度二次增长的问题,He等人(2025)提出了基于状态空间模型(state space model,SSM)的Pan-Mamba架构,该架构通过引入的态势感知机制,不仅保持了其捕捉全局依赖的能力,同时将计算的复杂度降低,为全色锐化任务提供了更加高效的解决方案。

从地物异质性角度看,Jin等人(2022a)提出的LAGConv通过内容自适应卷积核与全局谐波偏置增强局部—全局表征,Duan等人(2024)提出的CAN-Conv在非局部卷积框架中引入内容自适应权重以提升跨区域交互能力。相关的模型驱动深度网络也

通过引入退化模型或记忆机制提升可解释性与稳定性(Yan等,2022;Tian等,2023)。然而,上述方法多依赖隐式注意力来实现不同类别地物的“区域自适应”,这样的方法缺乏显式、可控的解释,忽略了全色多光谱图像中含有的语义先验信息,同时在平滑区域的光谱保持与纹理区域的细节重建之间仍难以自适应切换。

综上所述,尽管现有深度学习方法已取得显著进展,但在处理复杂遥感场景时仍面临深层次挑战。首先,局部细节注入与全局结构一致性难以协同,现有方法往往强调局部纹理增强而对跨区域结构关联建模不足;与此同时,当前主流方法大多采用全局统一的处理策略,将整幅影像视为均匀同质的整体,从而忽视地物异质性的影响。这种未加区分的处理方式使得算法难以在兼顾平滑区域光谱一致性的同时,有效实现纹理区域的细节重建。

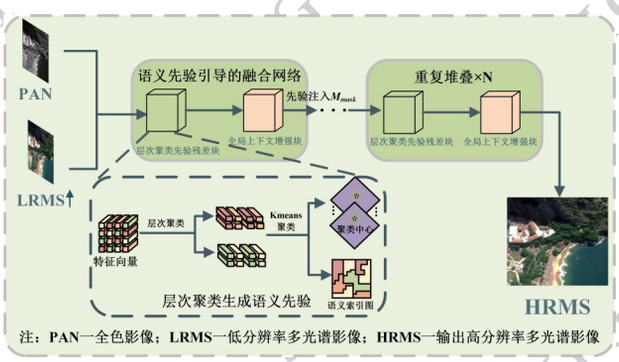


图1 HCPNet方法技术流程图

Fig. 1 Technical workflow of the proposed HCPNet

针对复杂场景中地物异质性带来的区域差异化融合需求,以及局部细节增强与全局结构一致性难以同时兼顾的问题,如图一所示,本文提出层次聚类引导的全色锐化网络。首先对输入多光谱进行层次聚类,并结合K-means生成同质区域引导掩膜;随后将PAN、LMS_{up}与掩膜M_{mask}拼接输入主干网络,在层次聚类先验残差块和全局上下文增强块的多尺度交互中注入细节并保持结构连续性,并引入多约束损失联合抑制光谱失真与细节伪影,最后以残差方式重建信息并得到高分辨率多光谱影像。总的来说,本文工作主要如下:

1)提出基于层次聚类与K-means的混合语义先验生成策略。该策略首先估计簇数并给出初始聚类中心,进而生成同质区域掩膜并初始化卷积网络参数,实现了特定地物属性的自适应处理。

2)设计了全局上下文增强块(efficient feature transformer block,EFT Block)。该模块结合卷积神经网络和注意力机制,实现了跨区域上下文信息交互。

3)构建了多约束协同优化框架。结合混合聚类先验网络,设计了聚类一致性损失、光谱角损失与重建损失的复合损失函数,在增强空间细节的同时有效抑制了光谱失真与细节伪影。

1 本文方法

本文中,记低分辨率多光谱图像为LRMS,并将其通过双三次插值上采样到与全色图像同分辨率的结果记为LMS_{up},其中S代表光谱波段数。LMS_{up}包含较完整的光谱信息但空间纹理相对不足;记全色图像为PAN,其具备较高空间分辨率但仅包含单波段强度信息。两者之间存在固定的空间分辨率比率r,即$H=r\times h$,$W=r\times w$,其中(H,W)与(h,w)分别为高、低分辨率下的空间尺寸。本文旨在学习端到端的非线性映射函数f,从观测图像对(PAN,LMS_{up})中重构出高分辨率多光谱图像$\boldsymbol{Y}$。该任务的核心在于:在尽可能保持原始光谱特性的同时,将全色影像的高频空间细节有效注入多光谱影像,并尽量避免由细节增强引入的光谱偏移与结构伪影。

如图2所示,HCPNet采用分阶段的编码—解码式框架。网络集成三个核心模块:其一,层次聚类先验生成模块。该模块从多光谱像素的光谱分布出发挖掘潜在同质区域结构,生成引导掩膜M_{mask}作为显式先验,从而提升网络对异质地物的差异化表征能力;其二,全局上下文增强模块。该模块以EFT块为核心,在卷积局部表征基础上引入高效自注意力建模,显式捕捉跨区域长程依赖,增强大尺度结构一致性并抑制伪影;其三,高分辨率图像重建模块。通过多级跳跃连接与残差学习策略,该策略从融合特征中逐级恢复空间细节。在训练阶段,网络在联合损失函数的指导下进行参数优化,该损失同时考虑光谱一致性、空间结构保持与聚类一致性约束。通过结合卷积神经网络的局部纹理刻画能力与EFT块的全局上下文建模能力,HCPNet能够在深层特征空间维持光谱向量一致性,并对边缘与细长结构实现更稳健的锐化。

1.1 层次聚类引导模块

在遥感图像中，复杂场景的地物光谱与纹理存在显著差异，若对整幅图像采用共享卷积，易在同质区域出现色偏，并在边界处引入伪影。为获得稳定且可解释的区域先验，本文采用层次聚类引导的混合聚类策略：先由层次聚类提供多尺度簇结构以估计簇数并提取初始中心，再以该初始化运行K-means生成同质区域引导掩膜，用于后续差异化融合。该过程可分为以下三步骤：

1）光谱样本构建：将每个像素的S维光谱向量记为$\boldsymbol{x}_i(i=1,2,\cdots,N)$，其中$H$、$W$分别为图像高度与宽度，$N=H\times W$为像素总数。

2）树状图构建：采用聚合层次聚类（agglomerative nesting，AGNES）对$\{\boldsymbol{x}_i\}$进行聚合，同时采用光谱角聚类（spectral angle distance，SAD）衡量样本间距离，并使用Ward连接准则生成树状图T。

3）最优簇数与初始中心确定：在树状图的不同截断层级下计算聚类有效性指标指数（Davies-Bouldin index，DBI），并结合簇间距离变化率自适应选择最优簇数K^*，其定义为

$$K^* = \underset{K}{argmax}\ \frac{\Delta DBI(K)}{\Delta K} \tag{1}$$

式中，K^*为最优簇数；K为候选簇数；DBI(K)为戴维森—堡丁指数，用于评价对应簇数下的聚类有效性；$\frac{\Delta DBI(K)}{\Delta K}$表示相邻簇数下DBI的变化率。

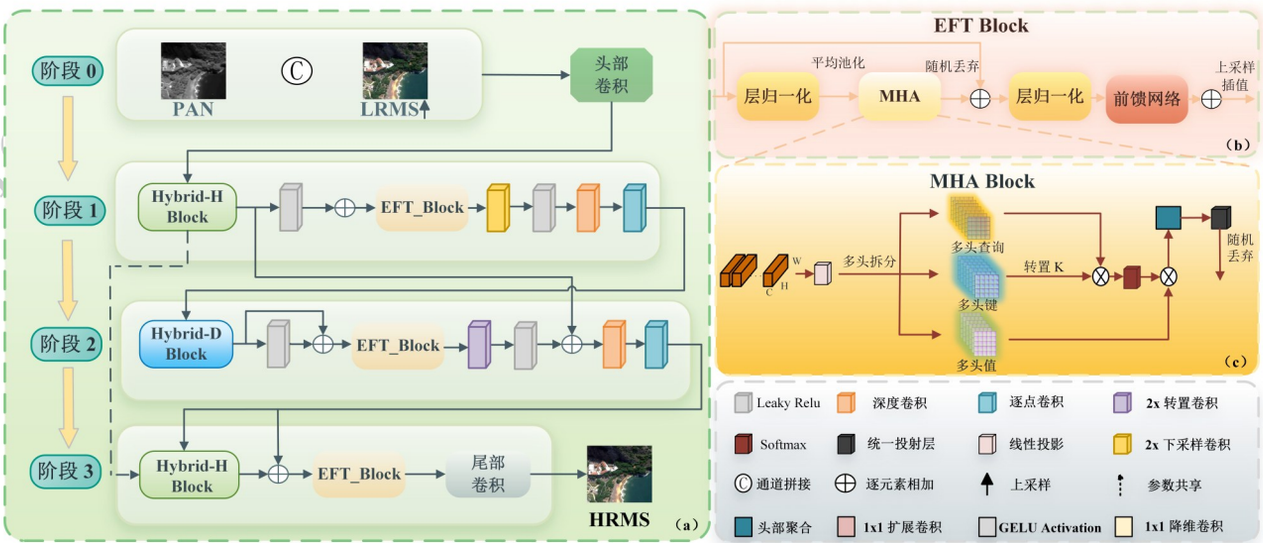
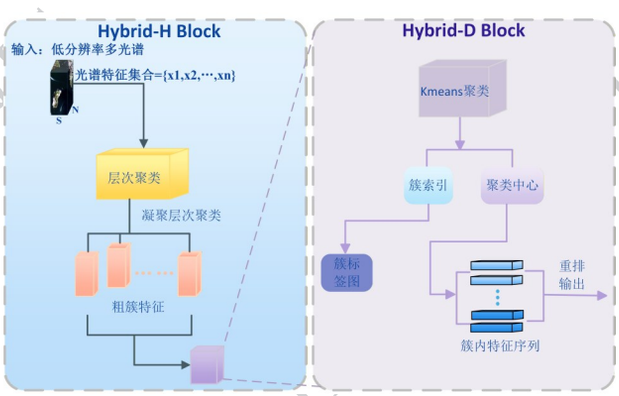


图2　HCPNet与全局上下文增强模块示意图。(a)HCPNet整体网络架构；(b)EFT块；(c)多头自注意力子模块。

Fig. 2　Schematic diagram of HCPNet and Global Context Enhancement Module. (a) The overall network architecture of HCPNet; (b) EFT block; (c) Multi-head self-attention sub-module.



(Hybrid-H与Hybrid-D)结构

Hybrid-H and Hybrid-D blocks.

图3　层次聚类引导的混合残差块

Fig. 3　Structure of the hierarchical-clustering-guided

在得到最优簇数后，从树状图T对应的K^*个簇提取质心作为K-means的初始聚类中心集合。随后以最优簇数与初始中心为初始化运行K-means，得到每个像素的类别标签mi，并将标签序列按图像空间尺寸重排得到同质区域引导掩膜，整个过程可用公式表达如下：

$$C_0 = \{c_1,c_2,\cdots,c_{K^*}\} = Centroid(T,K^*) \tag{2}$$

$$M_{mask} = Reshape([m_1,m_2,\cdots,m_N],(H,W)) \tag{3}$$

式中，$\boldsymbol{C}_0$为初始中心集合；$\boldsymbol{c}_k$为第k个簇的中心向量；T为层次聚类生成的树状图；Centroid(·)表示在给

定切割层次下提取各簇质心的算子。m_i为第i个像素的聚类标签，取值范围为$\{1,2,\cdots,K^*\}$；Reshape(·)表示将长度为N的一维标签序列按$H\times W$重排为二维掩膜；H、W分别为图像高度与宽度。

为便于展示聚类先验在特征增强中的作用，图3给出了混合残差块的结构及其浅深层差异化处理流程。在本文中，深层动态路由的簇数固定为32，过滤阈值取0.005。

1.2 主干网络架构与全局上下文增强模块

主干网络采用分阶段的编码—解码式框架，并通过跨阶段跳跃连接实现多尺度特征融合与重建。输入端将全色影像、上采样多光谱影像与引导掩膜在通道维拼接为输入张量：

$$X = Concat(PAN, LMS_{up}, M_{mask}) \tag{4}$$

式中，$\boldsymbol{X}$为网络输入张量；PAN为全色影像；LMS_{up}为双三次上采样后的多光谱影像；M_{mask}为同质区域引导掩膜；Concat(·)表示通道维拼接算子。

编码阶段包含两个混合交互阶段：Stage 1处于浅层，特征仍包含较多辐射与纹理细节，采用Hybrid-H块在聚类先验引导下进行局部细节增强并产生索引图用于路由；Stage 2位于瓶颈层，采用Hybrid-D块根据深层特征分布动态调整路由，以减少重复初始化带来的开销。解码阶段利用跳跃连接融合浅层细节与深层语义，尾部重建模块预测高频残差，并与上采样多光谱影像LMS_{up}相加得到输出$\boldsymbol{Y}$。

为突破卷积的局部感受野限制并增强跨区域信息交互，本文基于Transformer的全局上下文智增强模块，该模块在保持卷积对局部纹理敏感性的同时，引入自注意力机制显式建模长程依赖，从而保持道路、河流等细长结构的连续性，并提升大面积同质区域整体色彩的稳定性。

EFT块采用预归一化结构，先对输入特征进行层归一化得到$\boldsymbol{F}_n$，再通过线性投影生成查询、键和值矩阵：

$$Q = F_n W_Q,\quad K = F_n W_K,\quad V = F_n W_V \tag{5}$$

式中，$\boldsymbol{F}_n$为层归一化后的输入特征；$\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$分别为查询、键和值矩阵；$\boldsymbol{W}_Q$、$\boldsymbol{W}_K$、$\boldsymbol{W}_V$为对应的可学习线性投影矩阵。

在此基础上，EFT块采用缩放点积自注意力对全局上下文进行聚合：

$$F_a = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

式中，$\boldsymbol{F}_a$为注意力聚合后的全局特征；softmax(·)为归一化函数；$\boldsymbol{K}^{\mathrm{T}}$表示转置；$d_k$为键向量维度，用于缩放点积相似度以稳定训练。

值得注意的是，由于遥感影像空间尺寸较大，直接在全分辨率上计算自注意力会带来较高的计算与显存开销。因此，本文在EFT块内部对空间分辨率进行降采样，在较低分辨率上建立全局关联后再回注入原分辨率特征，以在可控开销下获得跨区域上下文信息。与此同时，局部卷积分支提供高频纹理与边缘细节，全局注意力分支补充跨区域结构关系，二者通过残差方式融合，从而同时兼顾细节锐化与光谱保真。

在网络整体流程中，聚类引导模块提供同质区域掩膜M_{mask}用于路由与差异化卷积，EFT块进一步在更大范围内整合跨区域信息。两者相互补充，同质区域的色调一致性更容易保持，边缘区域的振铃、晕影等伪影也会相对减少。

1.3 多约束损失函数

训练阶段采用联合损失，兼顾像素域重建、光谱形状保持与区域一致性约束。总损失定义为：

$$L = \lambda_s L_s + \lambda_c L_c + \lambda_1 L_1 \tag{7}$$

式中，L为总损失；L_s、L_c和L_1分别为光谱角损失、聚类一致性损失与重建损失；λ_s、λ_c和λ_1为对应权重系数。本文取λ_s=0.05，λ_c=0.02，λ_1=1.0。

其中聚类一致性损失由分布对齐项与紧致性项组成：

$$L_c = \lambda_{KL} L_{KL} + \lambda_p L_p \tag{8}$$

式中，L_c为聚类一致性损失；L_{KL}为分布对齐项；L_p为紧致性项；λ_{KL}与λ_p为对应权重。

分布对齐项采用对称KL散度刻画预测分布与目标分布的一致性：

$$L_{KL} = \frac{1}{2}\left[D_{KL}(p \,\|\, q) + D_{KL}(q \,\|\, p)\right] \tag{9}$$

式中，$\mathrm{D}_{\mathrm{KL}}(\cdot \,\|\, \cdot)$为KL散度算子；$\boldsymbol{p}$与$\boldsymbol{q}$分别为网络预测分布与由聚类先验得到的目标分布。

紧致性项通过最小化特征点到对应聚类中心的加权距离，鼓励同质区域在特征空间内更紧凑：

$$L_p = \frac{1}{BN}\sum_{b,n,k} p_{b,n,k}\, d^2\left(f_{b,n}, c_k\right) \tag{10}$$

式中，B为批次大小；N为参与计算的特征点数；$p_{b,n,k}$

为第b个样本第n个特征点属于第k类的预测概率；$\boldsymbol{f}_{b,n}$为对应特征向量；$\boldsymbol{c}_k$为第k个聚类中心；$\mathrm{d}^2(\cdot,\cdot)$为平方欧氏距离。

重建损失采用平均绝对误差：

$$L_1 = \frac{1}{CHW}\sum_{c=1}^{C}\sum_{h=1}^{H}\sum_{w=1}^{W}\left|Y_{c,h,w} - T_{c,h,w}\right| \tag{11}$$

式中，C为波段数；H与W分别为图像高度与宽度；$\boldsymbol{Y}_{c,h,w}$与$\boldsymbol{T}_{c,h,w}$分别为输出与真值在位置(h,w)、波段c处的像素值。

光谱角损失以光谱向量夹角约束输出与真值的光谱形状一致性：

$$L_s = \frac{1}{N}\sum_{i=1}^{N}\arccos\left(\frac{<y_i,t_i>}{\|y_i\|_2\|t_i\|_2+\varepsilon}\right) \tag{12}$$

式中，N为像素数；$\boldsymbol{y}_i$与$\boldsymbol{t}_i$分别为像素i处的输出与真值光谱向量；$<\cdot,\cdot>$表示内积，$\|\cdot\|_2$为二范数；ε为防止分母为零的微小常数。

2 实验与结果分析

2.1 实验设置

本文实验采用Pan Collection公开基准中的Quick Bird与GaoFen-2数据集，并沿用其训练集、验证集与测试集划分开展实验，以保证结果可复现。本文未对数据集进行额外筛选或重划分。

为了全面评估算法性能，本文在降分辨率协议即Wald协议下计算全参考指标：光谱角映射(spectral angle mapper，SAM)、无量纲全局相对误差(erreur relative globale adimensionnelle de synthèse，ERGAS)以及平均通用图像质量指数(mean universal image quality index，Q_{avg})。三个指标分别用于衡量光谱向量方向差异、整体相对误差、多波段结构一致性。

本文两套数据集均为4波段，因此Q_{avg}按4个波段的UQI取平均；其中ERGAS的分辨率比例参数取r等于1/4(r为高分辨率像元尺寸与低分辨率像元尺寸之比，与Wald协议的尺度关系一致)。在全分辨率(无真值参考)协议下，采用光谱失真(spectral distortion，D_λ)与空间失真(spatial distortion，D_s)，并以无参考综合质量指数(hybrid quality with no reference，HQNR)作为空谱一致性的综合评价指标。所有指标在将影像按传感器量化范围归一化到[0，1]后计算(QB：2047，GF2：1023)。为验证本文方法的有效性并覆盖不同技术路线，本文选取2014—2025年间2种传统方法与9种近年来代表性深度学习方法进行对比。传统方法包括TV(total variation)(Palsson等，2014)与FSRIC(full scale regression-based injection coefficients)(Vivone等，2018)；深度学习方法包括DiCNN(detail-injection convolutional neural network)(Deng等，2021)、FusionNet(two-stream fusion network)(Liu等，2020)、MMNet(memory-augmented model-driven network)(Yan等，2022)、LAGConv(local-context adaptive convolution kernels with global harmonic bias)(Jin等，2022a)、LGPConv(learnable Gaussian perturbation convolution)(Zhao等，2023)、HMPNet(interpretable model-driven deep network for hyperspectral，multispectral，and panchromatic image fusion)(Tian等，2023)、CANConv(content-adaptiv e non-local convolution)(Duan等，2024)、ADWM(adaptive dual-level weighting mechanism)(Huang等，2025a)和Pan-Mamba(pan-sharpening with state space model)(He等，2025)。对比实验优先采用作者开源实现或公开复现版本。

为验证本文方法的有效性并覆盖不同技术路线，本文选取2014—2025年间2种传统方法与9种近年来代表性深度学习方法进行对比。传统方法包括TV(total variation)(Palsson等，2014)与FSRIC(full

scale regression-based injection coefficients)(Vivone等，2018)；深度学习方法包括DiCNN(detail-injection convolutional neural network)(Deng等，2021)、FusionNet(two-stream fusion network)(Liu等，2020)、MMNet(memory-augmented model-driven network)(Yan等，2022)、LAGConv(local-context adaptive convolution kernels with global harmonic bias)(Jin等，2022a)、LGPConv(learnable Gaussian perturbation convolution)(Zhao等，2023)、HMPNet(interpretable model-driven deep network for hyperspectral，multispectral，and panchromatic image fusion)(Tian等，2023)、CANConv(content-adaptive non-local convolution)(Duan等，2024)、ADWM(adaptive dual-level weighting mechanism)(Huang等，2025a)和Pan-Mamba(pan-sharpening with state space model)(He等，2025)。对比实验优先采用作者开源实现或公开复现版本。

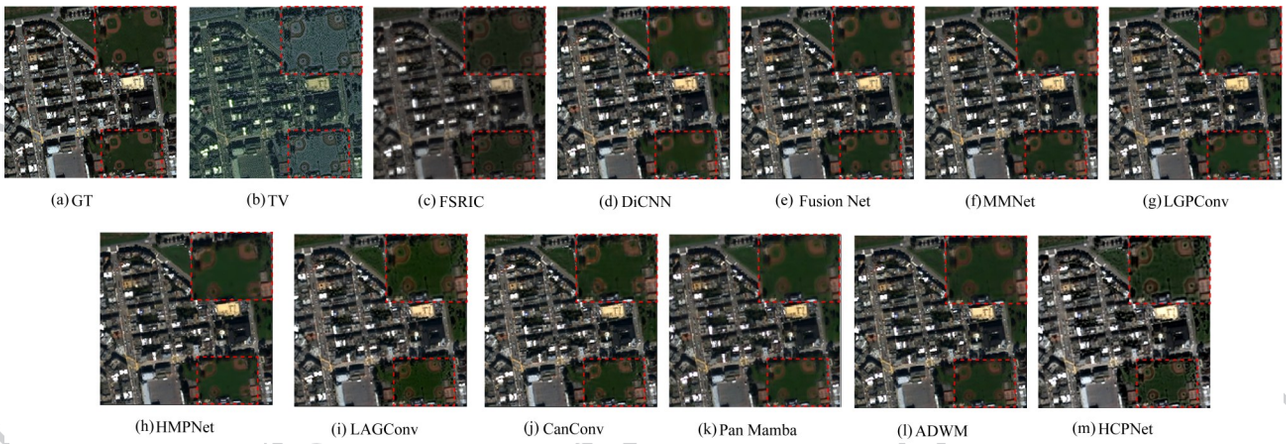


图4　Quick Bird数据集的低分辨率视觉对比

Fig. 4　Reduced-resolution visual comparison on Quick Bird dataset ((a) GT; (b) TV; (c) FSRIC; (d) DiCNN; (e) FusionNet; (f) MMNet; (g) LGPConv; (h) HMPNet; (i) LAGConv; (j) CANConv; (k) Pan-Mamba; (l) ADWM; (m) HCPNet)

在超参数设计方面，网络的基础通道数设置为32。混合聚类残差块中预设基础簇数为32，过滤阈值为0.005，并采用块级质心(patch-centroid)形式表示聚类中心，以提升聚类与路由稳定性；同时启用动态簇数量调整机制以适配不同场景地物分布。EFT Block采用单头自注意力机制(头数设置为1)，前馈网络扩展倍率为1，随机丢弃(Dropout)概率设置为0.1。

在实现细节方面，所有实验均基于PyTorch 2.0深度学习框架搭建。实验采用AdamW优化器，初始学习率为$5e^{-6}$，并使用余弦退火策略衰减至$1e^{-6}$；批量大小(batch size)为16，总训练轮次(epoch)为500，并固定随机种子为10。

2.2　定性结果分析

为验证模型在不同传感器与评估协议下的有效性，本文在Quick Bird与GaoFen-2两个数据集上给出全分辨率与降分辨率场景的定性对比结果。定性评价主要关注两点：空间细节是否真实可信，边缘与纹理是否清晰且无明显光晕、振铃等伪影；光谱一致性是否稳定，不同地物区域的色调是否自然、是否存在颜色漂移。为降低主观判断的不确定性，文中结合相应定量指标对视觉差异进行解释，使现象与数据能够相互印证。

2.2.1　降分辨率定性结果分析

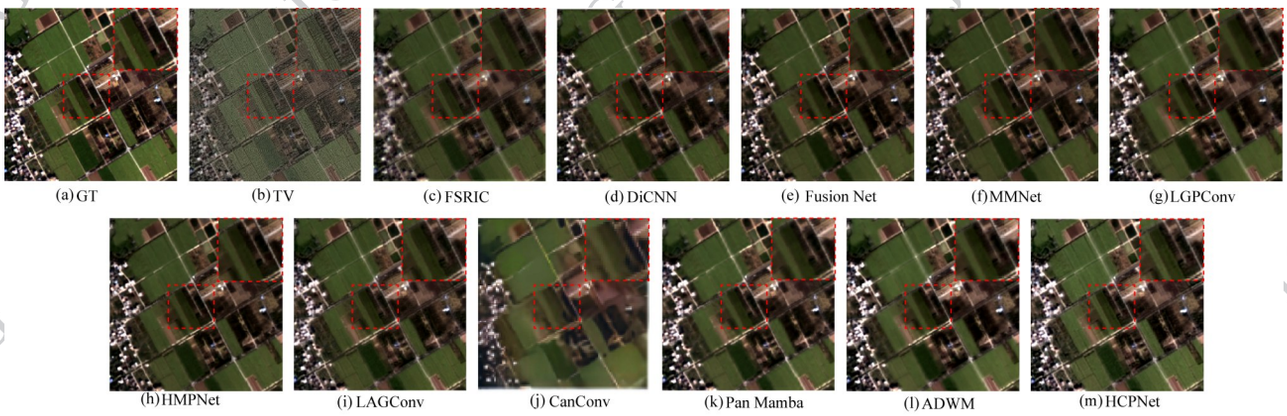
如图4和图5所示，本文给出了Quick Bird与GaoFen-2数据集在降分辨率场景下的视觉对比结果，用于观察不同方法在道路与建筑边缘的清晰度、纹理连续性以及同质区域是否出现块状伪影或色偏。实验样本按Wald协议构造，并以GT(原始MS)作为参考对照。

对比结果可以看到，传统注入类方法在道路或建筑边缘处易出现边缘过渡变宽、细节纹理不连续等现象；部分深度学习方法在屋顶纹理、道路标线等高频区域虽更锐利，但容易伴随振铃、光晕或同质区域的假纹理，并在植被和阴影等区域出现轻微色偏。相比之下，本文方法在建筑轮廓、道路边界处边缘更干净、纹理更连续，同时在水体和草地等同质区域更少出现块状伪影与彩色噪点，整体色调更接近GT(降分辨率)或LRMS基调(全分辨率)。

2.2.2　全分辨率定性结果分析

在全分辨率测试中，由于缺少GT真值参考，本文以PAN的空间纹理与LRMS的光谱信息作为对照依据，从图6与图7可视化结果可以看出，TV、FSRIC等传统优化方法通常能维持相对更稳定的整体色调，但细节恢复能力有限，局部纹理对比度不足，导致道路边界与建筑轮廓的锐利程度受限。部分深度学习方法能够显著增强锐度，但在高频注入过强时容易产生边缘伪影，同时在植被或阴影区域可能出现色调漂移与跨波段不一致。综合Quick Bird与GaoFen-2两组代表性全分辨率样例可以发现，本文方法在细节清晰度与色彩自然性上取得了较好的结果：在建筑屋顶与道路结构处，纹理融合结果连续且边界干净；在同质区域如水体、裸地与建筑外轮廓场景中，本文方法能够在不过度锐化的前提下维持更稳定的色彩与更连续的结构细节。如图7所示，在与DiCNN方法的比较中，其融合结果具有较明显的

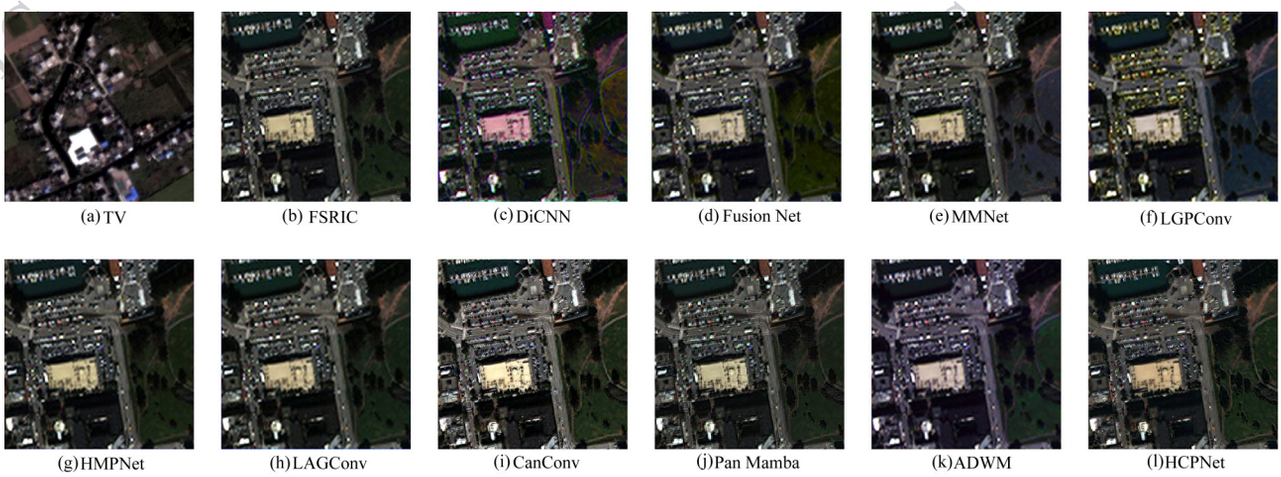
彩色伪影，ADWM方法存在偏紫或偏暗的色调；与CanConv、Pan-Mamba等方法相比，本文方法虽然具有较低的亮度，但最终融合图像中含有更少的伪影和更稳定的色彩，呈现出更自然的观感。



（f）MMNet；（g）LGPConv；（h）HMPNet；（i）LAGConv；（j）CANConv；（k）Pan-Mamba；（l）ADWM；（m）HCPNet）

图5　GaoFen-2数据集的低分辨率视觉对比

Fig. 5　Reduced-resolution visual comparison on GaoFen-2 dataset（（a）GT；（b）TV；（c）FSRIC；（d）DiCNN；（e）FusionNet；



（b）FSRIC；（c）DiCNN；（d）FusionNet；（e）MMNet；（f）LGPConv；（g）HMPNet；（h）LAGConv；（i）CANConv；（j）Pan-Mamba；（k）ADWM；（l）HCPNet）

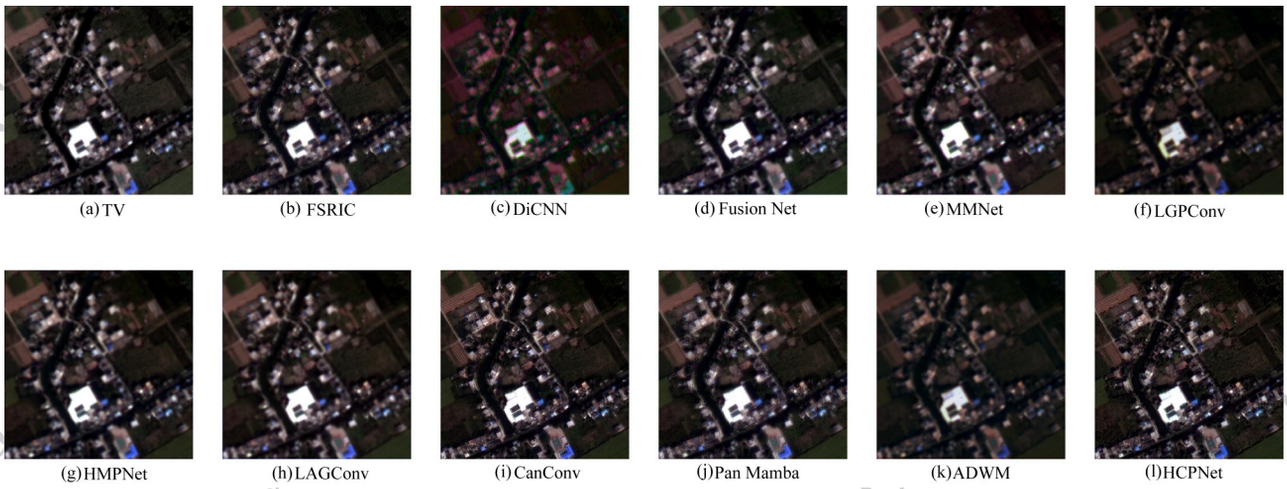
图6　Quick Bird数据集的全分辨率视觉对比

Fig. 6　A qualitative result comparison chart of representative methods on the Quick Bird full-resolution dataset（（a）TV；

2.3　定量结果分析

本研究在Quick Bird和GaoFen-2数据集上对所提方法（HCPNet）进行了系统的定量评估，具体结果如表1和表2所示。两套数据均同时给出降分辨率与全分辨率两种数据集下的统计结果，其中降分辨率统计有可对比的理想参考图像GT，可计算全参考指标SAM、ERGAS、Q_{avg}；全分辨率统计无参考图像，强调D_λ、D_s与HQNR等无参考指标对空谱失真的刻画。对比方法覆盖传统变分优化模型以及2019至2025年的代表性深度学习模型，包括TV、FSRIC以及Pan-Mamba等先进架构，能够较全面地检验方法在不同传感器特性与不同地物分布下的性能和稳定性。

在降分辨率指标方面，HCPNet在Quick Bird数据集上体现出更突出的光谱保真优势。SAM指标反映光谱向量方向差异，而ERGAS指标反映全局相对误差。所提方法的SAM和ERGAS显著低于对比方法，这表明模型在细节注入过程中能够有效抑制跨波段偏移。此外，综合质量指标Q_{avg}能够衡量融合结果多波段整体结构的一致性，所提方法在该指标上仅轻微次于CANConv，但CANConv在SAM和ERGAS上表现不佳，表明HCPNet在增强融合结果



(b) FSRIC; (c) DiCNN; (d) FusionNet; (e) MMNet; (f) LGPConv; (g) HMPNet; (h) LAGConv; (i) CANConv; (j) Pan-Mamba; (k) ADWM; (l) HCPNet)

图7　GaoFen-2数据集的全分辨率视觉对比

Fig. 7　A qualitative result comparison chart of representative methods on the GaoFen-2 full-resolution dataset ((a) TV;

表1　在Quick Bird数据集上进行的结果基准测试，采用20个低分辨率样本和20个全分辨率样本进行评估。

Table 1　The benchmark results on the Quick Bird dataset were evaluated using 20 low-resolution samples and 20 full-resolution samples.

方法	降分辨率指标			全分辨率指标		
	SAM ↓	ERGAS ↓	Q_{avg} ↑	D_λ ↓	D_s ↓	HQNR ↑
TV(Palsson等,2014)	0.349	11.23	0.7226	<u>0.0012</u>	0.1677	0.8313
FSRIC(Vivone等,2018)	0.304	9.900	0.7929	**0.0006**	0.1366	0.8628
DiCNN(Lin等,2019)	0.174	5.603	0.9447	0.0026	**0.0665**	**0.9311**
FusionNet(Deng等,2021)	0.134	4.318	0.9668	0.0028	0.1002	0.8973
MMNet(Yan等,2022)	0.126	4.044	0.9706	0.0032	0.0919	0.9052
LGPConv(Zhao等,2023)	0.130	4.163	0.9691	0.0028	0.1234	0.8741
HMPNet(Tian等,2023)	0.124	3.998	0.9718	0.0031	0.0990	0.8983
LAGConv(Jin等,2022a)	0.119	3.828	0.9737	0.0030	0.1678	0.8297
CANConv(Duan等,2024)	0.122	3.926	0.9726	0.0028	0.0968	0.9007
ADWM(Huang等,2025)	<u>0.118</u>	<u>3.824</u>	<u>0.9762</u>	0.0032	0.1993	0.7981
Pan-Mamba(He等,2025)	0.146	4.812	0.9619	0.0033	0.0995	0.8975
HCPNet(本文)	**0.114**	**3.673**	**0.9919**	0.0091	<u>0.0853</u>	<u>0.9063</u>

注：加粗字体表示各列结果最优值，下划线字体表示次优值。↑表示值越高越好，↓表示值越低越好，图中指标均为无量纲指标(无单位)。

局部细节的同时，能够进一步维持波段间相关性，使其在颜色与纹理的协同一致性上更接近参考图像。

究其原因，HCPNet所采用的层次聚类引导能够为同质区域提供更明确的区域先验，使网络在区域内部更倾向于保持原始MS图像的光谱分布，同时在边缘与细节区域进行有控制的高频补偿。多约束损失函数从光谱角度、像素重建与特征一致性三个层面共同约束训练过程，使得模型在保持空间连续性的同时，避免将PAN图像的高频信息简单叠加到所有波段而导致的色彩失真。同时，所提出的EFT Block通过上下文建模进一步提升了全局结构的一致性，增强细节保持能力，从而在指标上获得更稳定

的综合提升。

GaoFen-2数据集相较于Quick Bird包含更复杂的异质地物分布与更显著的尺度变化，对融合方法的空间细节注入与光谱保持的平衡性提出更加严格的要求。如表2所示，HCPNet在GaoFen-2数据集上的SAM和ERGAS指标上保持显著优势。这表明融合后的结果在光谱向量与参考目标的偏离程度更小，整体相对误差的控制更加稳定。进一步说明了所提方法在纹理细节复杂的场景下具有更好的跨数据集鲁棒性。

在全分辨率指标方面，本文方法在无真值参考约束下依然表现出较好的空谱平衡能力。以表1为例，D_λ和D_s分别衡量光谱信息和空间信息的失真，HCPNet在两个指标上的表现均优于其他对比方法，说明方法在光谱与空间两个维度均实现了更强的失真抑制。综合质量HQNR指标由D_λ与D_s共同决定，两者同时改善会带来更高的HQNR值。与一些方法倾向于通过强化高频注入来降低D_s却引起D_λ上升不同，HCPNet能在细节增强与光谱保持之间保持同步优化，其关键在于EFT Block对全局上下文关联增强的能力能够缓解全局上下文结构失配，而聚类先验与光谱约束共同限制了局部色彩偏移。综合降分辨率与全分辨率两种协议的结果可以看出，HCPNet在多项关键指标上均呈现一致优势，表明所提框架能够在细节增强与光谱保持之间实现更稳定的协同优化，并具有面向真实应用场景的可靠性。

2.4 消融实验与分析

为验证所提模块的有效性和多约束损失函数中各项损失的贡献，本文设计了两组消融实验，并在两个数据集上进行了分别验证。其中，多约束损失

表2 在GaoFen-2数据集上进行的结果基准测试，采用20个低分辨率样本和20个全分辨率样本进行评估。

Table 2 The benchmark results on the GaoFen-2 dataset were evaluated using 20 low-resolution samples and 20 full-resolution samples.

方法	降分辨率指标			全分辨率指标		
	SAM ↓	ERGAS ↓	Q_{avg} ↑	D_λ ↓	D_s ↓	HQNR ↑
TV（Palsson等，2014）	0.082	2.583	0.9926	0.0021	0.0491	0.9501
FSRIC（Vivone等，2018）	0.068	1.832	0.9963	0.0018	0.0526	0.9465
DiCNN（Lin等，2019）	0.044	1.195	0.9964	0.0018	0.0495	0.9486
FusionNet（Deng等，2021）	0.038	1.026	0.9963	0.0020	0.0496	0.9485
MMNet（Yan等，2022）	0.028	0.907	0.9964	0.0035	0.0646	0.9322
LGPConv（Zhao等，2023）	0.036	0.980	0.9967	0.0018	0.0506	0.9477
HMPNet（Tian等，2023）	0.028	0.753	0.9199	<u>0.0017</u>	0.0506	0.9478
LAGConv（Jin等，2022a）	0.027	0.733	0.9965	0.0019	0.0503	0.9479
CANConv（Duan等，2024）	0.027	0.729	<u>0.9971</u>	0.0020	<u>0.0436</u>	**0.9545**
ADWM（Huang等，2025）	<u>0.025</u>	<u>0.641</u>	0.9964	0.0018	0.0504	0.9479
Pan-Mamba（He等，2025）	0.033	0.899	0.9963	0.0023	**0.0435**	<u>0.9543</u>
HCPNet（本文）	**0.023**	**0.633**	**0.9997**	**0.0017**	0.0505	0.9480

注：加粗字体表示各列结果最优值，下划线字体表示次优值。↑表示值越高越好，↓表示值越低越好。图中指标均为无量纲指标（无单位）。

1）光谱角度映射损失的有效性验证：作为消融实验的基准，首先测试仅使用光谱角度映射损失L_s进行训练的模型性能。如表3所示，L_s能够迫使模型较好地维持重建图像的光谱分布，确保其与参考图像在色调上的一致性，这验证了L_s在维护物理属性方面的基础作用。定性对比结果如图8和图9所示，若缺失空间域的强约束，模型往往难以准确恢复图像的几何结构，特别是在地物边缘和纹理丰富区域，仅依赖光谱约束会导致高频信息无法有效注入，从而使重建图像呈现出边缘模糊和纹理过度平滑的现象。

2）像素级重建损失的空间细节提升：为提高基

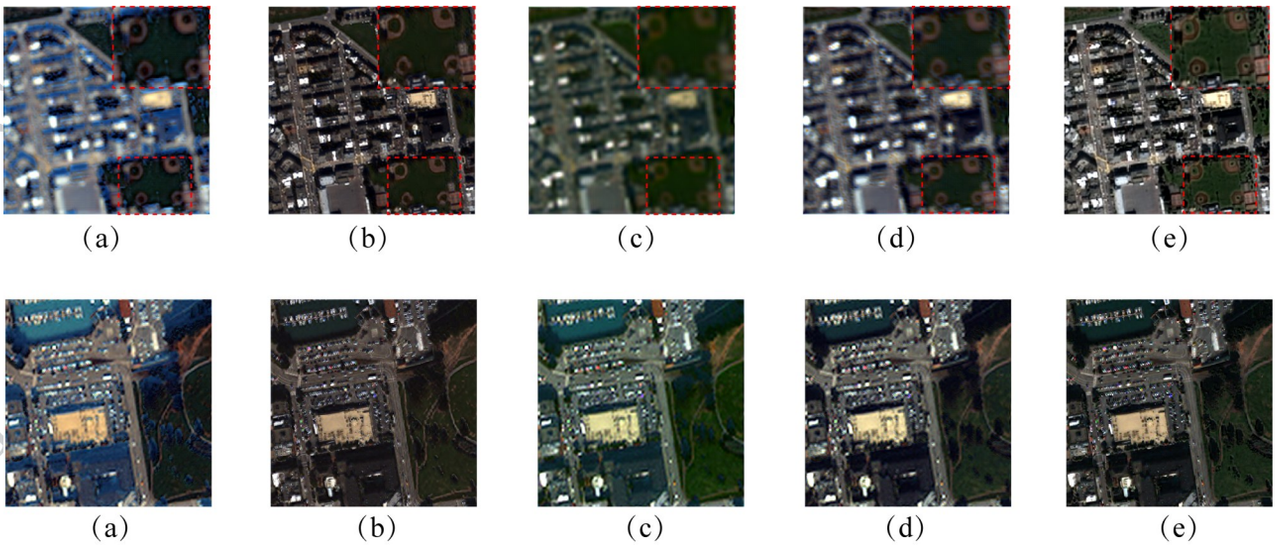


图8　Quick Bird数据集消融实验定性结果((a)去除重建损失L_1与聚类一致性损失L_c;(b)去除聚类一致性损失L_c;(c)去除混合聚类语义先验;(d)去除EFT块;(e)HCPNet。)上行为降分辨率结果,下行为全分辨率结果

Fig. 8　Qualitative ablation results on Quick Bird. (a) w/o L_1 and L_c; (b) w/o L_c; (c) w/o hybrid clustering prior; (d) w/o EFT block; (e) HCPNet. Top: reduced-resolution; bottom: full-resolution

准模型对空间结构的还原度,在多约束损失中引入了像素级重建损失L_1。指标对比如表4所示,该约束显著提升了融合图像的空间分辨率。定性对比如图8和图9所示,在保持原有光谱优势的前提下,若移除L_1,模型将难以捕捉高频纹理信息。该消融实验证实,强度一致性约束对于提升视觉清晰度、抑制空间模糊具有不可替代的作用。

3)聚类一致性损失的全局优化机制:在上述损失函数基础上,本文进一步引入聚类一致性损失L_c以构建完整的优化框架。实验结果表明,该配置在

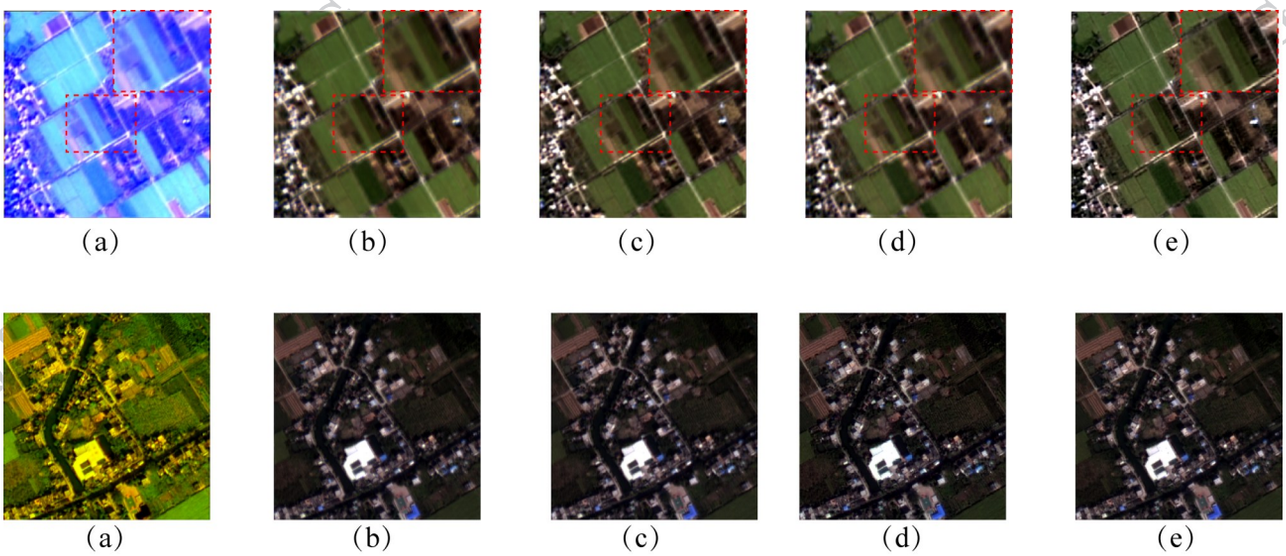


图9　Gaofen-2数据集消融实验定性结果((a)去除重建损失L_1与聚类一致性损失L_c;(b)去除聚类一致性损失L_c;(c)去除混合聚类语义先验;(d)去除EFT块;(e)HCPNet。)上行为降分辨率结果,下行为全分辨率结果

Fig. 9　Qualitative ablation results on Gaofen-2 (a) w/o L_1 and L_c; (b) w/o L_c; (c) w/o hybrid clustering prior; (d) w/o EFT block; (e) HCPNet. Top: reduced-resolution; bottom: full-resolution

光谱保真度和空间细节之间取得了最佳平衡。通过在特征空间施加分布对齐与类内聚合约束,L_c显著改善了同质区域的语义一致性。基于图8和图9的定性结果对比分析说明,若缺乏该聚类引导策

略，模型在处理复杂地物边界时容易产生伪影，且难以保证连续区域的平滑度，证明了层次聚类引导在提升全局融合质量方面的必要性。

2.4.2　网络架构组件消融实验

1）层次聚类与传统K-mean聚类策略对比：为了验证层次聚类策略在特征聚合中的优势，将混合聚类模块替换为传统的K-means算法进行对比。实验结果表明，K-means算法依赖于固定的簇中心，难以捕捉数据的多尺度结构。若采用K-means替代层次聚类，模型在异质地物交界区域的重建质量将受到限制，且极易受初始聚类中心选择的影响。相比之下，本文采用的层次聚类策略能够自适应地确定最佳类别数，并通过树状图分析提供更合理的初始聚类中心，从而更有效地引导网络学习语义一致性特征表示。

2）EFT Block对全局跨区域上下文增强的贡献：本文采用EFT Block捕捉像素空间的全局上下文关联信息，为了评估其有效性，实验将EFT Block替换为常规卷积模块（CNN）进行对比分析。结果表明，常规CNN缺乏建立跨区域上下文关联的能力。若移除EFT Block，模型在处理城市建筑群和自然地貌过渡区等大范围空间结构时，将面临明显的局限性，导致重建结果出现结构不一致和细节丢失，这验证了EFT Block及其内嵌的自注意力机制在建立全局上下文联系、保持地物连续性和结构完整性方面的关键作用。

2.4.3　模型综合性能分析

综合各项指标与可视化结果，HCPNet在空谱平衡方面表现较为稳定。在光谱保真度方面，得益于层次聚类先验提供的精准语义引导与光谱角度映射损失的严格约束，模型重建的光谱特征与参考图像保持了极高的一致性。特别是在区分不同地物类别时，模型能够准确还原地物色彩，有效避免了常见的光谱偏差。

表3　在Quick Bird数据集上进行损失函数消融实验，采用20个低分辨率样本和20个全分辨率样本进行评估

Table 3　Ablation study of the loss function on the Quick Bird dataset, using 20 low-resolution and 20 full-resolution samples for evaluation

实验设置	降分辨率指标			全分辨率指标		
	SAM ↓	ERGAS ↓	Q_{avg} ↑	D_λ ↓	D_s ↓	HQNR ↑
w/o L_1& L_c	0.349	11.61	0.798	0.0109	0.0959	0.8963
w/o L_c	0.140	4.980	0.954	0.0124	0.0962	0.8781
HCPNet	**0.114**	**3.673**	**0.992**	**0.0091**	**0.0853**	**0.9063**

注：加粗字体表示各列结果最优值。↑表示值越高越好，↓表示值越低越好。图中指标均为无量纲指标（无单位）。

在空间细节表现上，相较于传统方法在同质区域易产生的块状伪影现象，HCPNet的输出更加平滑且连续，证明了EFT Block成功捕捉了宏观空间结构，而聚类一致性损失专注于局部区域的约束，两者的协同作用确保了图像在保留宏观结构完整性的同时，实现了纹理细节的精确复原。降分辨率指标的对比结果进一步印证了上述分析。如表1所示，在Quick Bird数据集上，本方法的SAM和ERGAS值显著优于所有对比方法；同时，Q_{avg}值优于Pan-Mamba等近期代表性模型。这表明层次聚类引导策略能够生成精准的地物分区先验，有效指导网络对光谱信息进行精细化保持。这种基于语义先验的差异化处理机制，本质上克服了传统卷积网络在处理光谱差异显著区域时易发的“光谱失真”问题，在保证空间连续性的同时维持了高度的光谱一致性。如表2所示，在GaoFen-2数据集上，面对复杂的异质地物分布，本方法的SAM和ERGAS分别达到0.023和

表4　在GaoFen-2数据集上进行模块消融实验，采用20个低分辨率样本和20个全分辨率样本进行评估

Table 4　Module ablation study on the GaoFen-2 dataset, using 20 low-resolution and 20 full-resolution samples for evaluation

实验设置	降分辨率指标			全分辨率指标		
	SAM ↓	ERGAS ↓	Q_{avg} ↑	D_λ ↓	D_s ↓	HQNR ↑
w/o Hybrid Clustering	0.074	1.203	0.9972	0.0057	0.0569	0.9240
w/o EFT Block	0.034	0.988	0.9984	0.0044	0.0690	0.9245
HCPNet	**0.023**	**0.633**	**0.9998**	**0.0017**	**0.0505**	**0.9480**

注：加粗字体表示各列结果最优值。↑表示值越高越好，↓表示值越低越好。图中指标均为无量纲指标（无单位）。

0.633，延续了领先优势，说明模型具有跨传感器与跨地物尺度的强大鲁棒性。

最后，全分辨率场景下的实验表现进一步证实了完整模型的泛化优势。在缺乏理想参考图像约束的情况下，HCPNet依然能够保持卓越的空谱平衡性。EFT Block通过对全局上下文区域增强，有效降低了空间几何畸变，而层次聚类引导则在特征层面提供了强语义约束，从而大幅抑制了光谱失真的发生。使得最终生成的融合图像在保留清晰锐利纹理细节的同时，呈现出自然真实的光谱特性，实现了主观视觉感知与客观评价指标的高度统一。

3 结 论

本文针对当前深度学习全色锐化方法在跨区域上下文关联和语义信息利用方面的局限性，提出一种层次聚类引导的全色锐化网络（HCPNet）。该方法创新性地设计了混合聚类策略，使聚类构建的同质地物分区图作为先验知识与特征深度融合，通过自适应的同质区域划分解决语义信息利用不足的问题；同时，设计了全局上下文增强块（EFT Block），采用空间降采样自注意力计算策略增强跨区域上下文信息交互，缓解传统卷积感受野受限问题；此外，构建多约束损失函数体系，从光谱、像素及特征维度协同优化重建质量。实验表明，本文方法在光谱特性保持、空间细节重建及同质区域一致性方面均展现出显著优势，有效减少了伪影现象，在视觉效果和数值表现上均优于现有主流模型。然而，目前的全局上下文增强模块仍存在结构性冗余，且尚未充分挖掘信号的频域物理特性。因此，后续研究将探索基于频域物理属性的去冗余机制，力求通过多域协同降低复杂度并提升跨域泛化性。

参考文献（References）

Aiazzi B, Baronti S and Selva M. 2007. Improving component substitution pansharpening through multivariate regression of M-S+Pan data. IEEE Transactions on Geoscience and Remote Sensing, 45(10): 3230-3239 [DOI: 10.1109/TGRS.2007.901007]

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]

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]

Duan Y, Wu X, Deng H and Deng L J. 2024. Content-adaptive non-local convolution for remote sensing pansharpening//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 27738-27747 [DOI: 10.1109/CVPR52733.2024.02650]

Ghassemian H. 2016. A review of remote sensing image fusion methods. Information Fusion, 32: 75-89 [DOI: 10.1016/j.inffus.2016.03.003]

Gu A, Goel K and Ré C. 2022. Efficiently modeling long sequences with structured state spaces//International Conference on Learning epresentations. Virtual: OpenReview.net [DOI: 10.48550/arXiv.2111.00396]

He X, Cao K, Yan K, Li R, Xie C, Zhang J, et al. 2025. Pan-Mamba: Effective pan-sharpening with state space model. Information Fusion, 115: 102779 [DOI: 10.1016/j.inffus.2024.102779]

Hou J, Chen X, Wu C, Zhou M, Li J and Hong D. 2025. Bilateral Adaptive Evolution Transformer for Multispectral Image Fusion. IEEE Transactions on Geoscience and Remote Sensing, 63: 5400612 [DOI: 10.1109/TGRS.2024.3494562]

Huang J, Chen H, Ren J and Deng L J. 2025a. A General Adaptive Dual-level Weighting Mechanism for Remote Sensing Pansharpening//Proceedings of the Computer Vision and Pattern Recognition Conference, 7447-7456 [DOI: 10.1109/CVPR52734.2025.00698]

Huang J, Huang R, Xu J, Peng S, Duan Y and Deng L J. 2025b. Wavelet-Assisted Multi-Frequency Attention Network for Pansharpening//Proceedings of the AAAI Conference on Artificial Intelligence, 39(4): 3662-3670 [DOI: 10.1609/aaai.v39i4.32381]

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 AAAI Conference on Artificial Intelligence, 36(1): 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]

Kaur G, Saini K S, Singh D and Kaur M. 2021. A comprehensive study on computational pansharpening techniques for remote sensing images. Archives of Computational Methods in Engineering, 28: 3155-3185 [DOI: 10.1007/s11831-021-09565-y]

Li H, Jing L and Tang Y. 2017. Assessment of pansharpening methods applied to WorldView-2 imagery fusion. Sensors, 17(1): 89 [DOI: 10.3390/s17010089]

Li M Y and Fu Y. 2023. Joint self-attention Transformer for multispectral and hyperspectral image fusion. Journal of Image and Graphics, 28(12): 3922-3934 (李妙宇, 付莹. 2023. 用于多光谱和高光谱图

像融合的联合自注意力Transformer. 中国图象图形学报，28(12)：3922-3934)［DOI：10.11834/jig.220954］

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］

Lin H，Rao Y，Li J，Chanussot J，Plaza A，Zhu J，et al. 2019. Pansharpening via Detail Injection Based Convolutional Neural Networks. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing，12(4)：1188-1204［DOI：10.1109/JSTARS.2019.2898574］

Palsson F，Sveinsson J R and Ulfarsson M O. 2014. A New Pansharpening Algorithm Based on Total Variation. IEEE Geoscience and Remote Sensing Letters，11(1)：318-322［DOI：10.1109/LGRS.2013.2257669］

Tian X，Li K，Zhang W，Wang Z Y and Ma J Y. 2023. Interpretable model-driven deep network for hyperspectral，multispectral，and panchromatic image fusion. IEEE Transactions on Neural Networks and Learning Systems，35(10)：14382-14395［DOI：10.1109/TNNLS.2023.3278928］

Vivone G，Alparone L，Chanussot J，Dalla Mura M，Garzelli A，Licciardi G A，et al. 2015. A critical comparison among pansharpening algorithms. IEEE Transactions on Geoscience and Remote Sensing，53(5)：2565-2586［DOI：10.1109/TGRS.2014.2361734］

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］

Wang W H，Xie E Z，Li X，Fan D P，Song K T，Liang D，et al. 2021. Pyramid vision transformer：A versatile backbone for dense prediction without convolutions//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal，QC，Canada：IEEE：568-578［DOI：10.1109/ICCV48922.2021.00061］

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］

Yan K Y，Zhou M，Zhang L and Xie C J. 2022. Memory-augmented model-driven network for pansharpening//European Conference on Computer Vision. Cham：Springer Nature Switzerland，13679：306-322［DOI：10.1007/978-3-031-19800-7_18］

Yang J F，Fu X Y，Hu Y，Huang Y，Ding X H and Paisley J. 2017. PanNet：A deep network architecture for pan-sharpening//Proceedings of the IEEE International Conference on Computer Vision. Venice，Italy：IEEE：5449-5457［DOI：10.1109/ICCV.2017.214］

Yuan Q Q，Wei Y C，Meng X C，Shen H F and Zhang L P. 2018. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing，11(3)：978-989［DOI：10.1109/JSTARS.2018.2794888］

Zhao C Y，Zhang T J，Ran R，Chen Z X and Deng L J. 2023. LGP-Conv：Learnable Gaussian Perturbation Convolution for Lightweight Pansharpening//Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. Macao，China：International Joint Conferences on Artificial Intelligence Organization：4647-4655［DOI：10.24963/ijcai.2023/517］

Zhang T，Wang B F，Fu Y，Liu S R，Ye J C，Shan P H，et al. 2024. Deep learning-based spectral image super-resolution：a survey. Journal of Image and Graphics，29(8)：2113-2136（张涛，王彬洋，付莹，刘松荣，叶吉超，单培红，颜成钢. 2024. 基于深度学习的光谱图像超分辨率综述. 中国图象图形学报，29(8)：2113-2136)［DOI：10.11834/jig.230747］

作者简介

王柏翔，男，硕士研究生，主要研究方向为遥感图像融合全色锐化算法研究。E-mail：；wangluyangqwe@outlook.com

霍宏涛，通信作者，男，教授，主要研究方向为遥感与地理信息系统应用。E-mail：huohongtao@ppsuc.edu.cn

郑博文，男，硕士研究生，主要研究方向为遥感目标检测。E-mail：2022211499@stu.ppsuc.edu

李志倩，女，硕士研究生，主要研究方向为遥感图像全色锐化算法研究。E-mail：lizhiqian010728@163.com