

基于增强型注意力残差模块的遥感图像全色锐化方法研究

萧宝玮

(山西大学 现代教育技术学院, 太原 030006)

摘要: 为了解决现有深度学习全色锐化模型结构臃肿、计算开销大的问题, 研究并设计了一种高效、轻量化的新型全色锐化网络; 该方法的核心是一种新设计的增强型注意力残差模块 (EARM), 采用了门控融合机制以智能整合多光谱与全色特征, 并利用双重注意力机制增强关键信息表达, 从而在控制模型复杂度的前提下提升了特征提取能力; 为验证方法有效性, 在 WorldView-II、WorldView-III 及 GaoFen-2 等公开数据集上进行了实验, 实验结果显示, 在光谱保真度和空间细节还原方面, 均取得显著提升, 同时主观视觉质量上也展现出更清晰的边缘和更自然的光谱连续性, 此外, 该网络架构在保持较低参数数量和计算负担的同时, 实现了全色锐化精度与推理效率的有效平衡, 为高分辨率遥感图像的实际工程化应用提供了一个兼具先进性与实用性的解决方案。

关键词: 全色锐化; 深度学习; 图像融合; EARM; 注意力机制

Research on Pansharpening Method for Remote Sensing Images Based on Enhanced Attention Residual Module

XIAO Baowei

(School of Modern Education Technology, Shanxi University, Taiyuan 030006, China)

Abstract: To solve the characteristics of overstuffed structure and high computational cost in existing deep learning panchromatic sharpening models, a new efficient and lightweight network is studied and designed. Its core is a newly designed enhanced attention residual module (EARM), which adopts a gated fusion mechanism to intelligently integrate multispectral and panchromatic features, and uses a dual attention mechanism to enhance the expression of key information, thereby improving the feature extraction ability while controlling the complexity of the model. To verify the effectiveness of the method, experiments are carried out on WorldView-II, WorldView-III, GaoFen-2 and other public datasets. Experimental results show that this method has significantly improved the spectral fidelity and spatial detail restoration while exhibiting clearer edges and more natural spectral continuity in the subjective visual quality. Furthermore, the network architecture achieves an effective balance between panchromatic sharpening accuracy and reasoning efficiency while maintaining a lower parameter and computational burden, and provides an advanced and practical solution in the practical engineering application of high-resolution remote sensing images.

Keywords: pansharpening; deep learning; image fusion; EARM; attention mechanism

0 引言

全色锐化 (Pansharpening) 是遥感图像处理中的一项基本图像融合任务^[1], 其核心目标是将卫星传感器捕获的两种不同类型的图像进行有效融合^[2]。具体而言, 它旨在将一张具有高空间分辨率但缺乏光谱信息的全色 (PAN, panchromatic) 图像, 与一张具有丰富光谱信息但空间分辨率较低的多光谱 (MS, multispec-

tral) 图像相结合, 最终生成一张同时具备高空间分辨率和丰富光谱信息的高分辨率多光谱 (HRMS, high-resolution multispectral) 图像。由于全色锐化的本质是提升多光谱图像的空间分辨率, 这与计算机视觉中的单图像超分辨率 (SR, Super-Resolution) 任务在目标上具有高度相似性, 因此许多 SR 领域的先进思想与模型, 如 EDSR^[3], 也为全色锐化技术的发展提供了宝贵

收稿日期: 2025-05-18; 修回日期: 2025-07-17。

作者简介: 萧宝玮 (1966-), 男, 硕士, 实验师。

引用格式: 萧宝玮. 基于增强型注意力残差模块的遥感图像全色锐化方法研究[J]. 计算机测量与控制, 2025, 33(11): 336-342.

的借鉴。

1 全色锐化技术

1.1 传统全色锐化方法

全色锐化的传统方法主要包括成分替换法 (CS, component substitution) 和多分辨率分析法 (MRA, multi-resolution analysis)。CS 方法, 如主成分分析 (PCA, principal component analysis)^[4] 和 IHS (Intensity Hue Saturation) 变换^[5], 通过将 MS 图像变换到另一空间, 用 PAN 图像替换其中某个分量后进行逆变换。MRA^[6-9] 方法, 如拉普拉斯金字塔^[5], 则是在不同频率尺度上对图像进行分解和融合。这些传统方法^[10-11] 的优势在于实现简单、计算速度快, 但其性能上限较低, 往往会引入明显的光谱失真或空间伪影, 难以满足高精度应用的需求。

1.2 基于深度学习的全色锐化方法

随着深度学习技术的兴起, 基于卷积神经网络 (CNN, convolutional neural network) 的方法极大地提升了全色锐化的性能。早期的探索性工作, 如 PNN^[12], 验证了 CNN 在学习端到端映射方面的可行性, 但其简单的三层结构限制了性能。随后的研究趋势转向构建更深、更复杂的网络架构, 例如 PanNet^[9] 和 MSDCNN^[13] 便引入了残差连接^[14] 与多尺度特征来优化融合效果。这一阶段虽然提升了指标, 但也开启了模型复杂度日益增加的趋势。为了进一步追求重建质量, 研究者们探索了多样化的技术路径: 一方面, 尝试引入生成对抗网络 (GAN, generative adversarial network)^[15-16], 但 GAN 的随机性与确定性重建任务存在固有矛盾; 另一方面, 借助可逆神经网络 (INN, invertible neural networks)^[17] 的结构特性来追求信息保真度, 但这带来了模型参数激增和训练不稳定的高昂代价。此外, 还有方法如 PanFlow^[18] 将任务分解为多个步骤, 虽取得了优异性能, 却牺牲了模型的推理效率。

1.3 注意力机制在全色锐化上的应用

上述方法虽然各有优势, 但也普遍存在一些弊病, 如模型过于臃肿、推理速度慢、训练不稳定等。这促使我们反思: 是否必须通过不断叠加复杂模块来提升性能? 在图像超分辨率领域, EDSR^[3] 的成功给出了一个不同的答案。它通过对当时先进的 SRResNet 网络进行“减法”设计, 移除了在 SR 任务中被证明为非必要的批量归一化 (BN, batch normal) 层, 反而获得了训练更稳定、性能更强、模型更高效的全面提升。受此启发, 本文并非直接对 EDSR 网络进行简单的修改套用, 而是借鉴其“大道至简”的设计哲学, 即通过精炼模块设计而非盲目叠加复杂结构来提升性能。我们针对全色锐化任务的双输入、多模态特性, 设计了一个全新的核

心模块——增强型注意力残差模块 (EARM, enhanced attention residual module), 旨在构建一个性能卓越, 同时没有显著负面效果和高昂代价的网络。

本文的贡献主要在于: 我们提出了一种新颖高效的全色锐化网络架构, 其核心是本文针对该任务专门设计的增强型注意力残差模块 (EARM)。该模块巧妙地融合了门控机制与双重注意力, 能够智能且高效地处理多模态特征。基于此, 我们的方法在不引入参数量激增、训练困难等负面代价的前提下, 实现了对全色锐化性能的显著提升, 在模型效率与性能之间达到了更优的平衡, 为该领域提供了兼具高性能与实用性的解决方案。

2 基于增强型注意力残差网络的方法设计

2.1 整体网络架构

本文提出的网络整体架构如图 1 所示, 其核心思想是构建一个高效的双分支特征提取与融合框架, 并利用深度残差模块进行精细化重建。

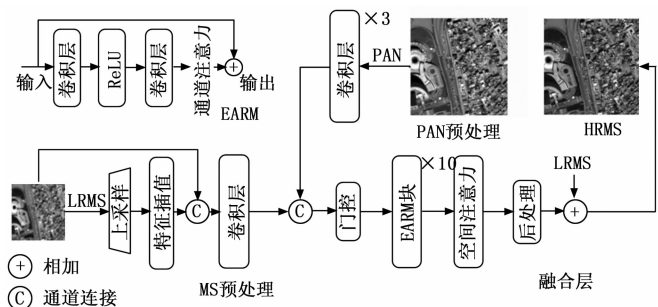


图 1 本文方法的框架流程图

本文提出的网络整体架构如图 1 所示, 它遵循一个设计精巧的双分支输入、深度特征融合与全局残差重建的流程。给定低分辨率多光谱图像 $I_{LRMS} \in R^{H \times W \times C}$ 和全色图像 $I_{PAN} \in R^{4H \times 4W \times 1}$ 作为输入, 其中 H, W 为低分图像的空间维度, C 为光谱通道数。

网络处理流程首先对 I_{LRMS} 进行一次双三次插值上采样, 使其空间分辨率与 I_{PAN} 对齐, 得到上采样多光谱图像 $I_{b_ms} \in R^{4H \times 4W \times C}$ 。随后, I_{b_ms} 与 I_{PAN} 分别进入独立的浅层特征提取模块:

$$F_{ms} = H_{head}(I_{b_ms}) \quad (1)$$

$$F_{pan} = H_{pan_proj}(I_{PAN}) \quad (2)$$

其中: H_{head} 和 H_{pan_proj} 均为一个 3×3 的卷积层, 分别将 I_{b_ms} 和 I_{PAN} 映射到多维特征空间, 得到初始的多光谱特征 $F_{ms} \in R^{4H \times 4W \times C_{feat}}$ 和全色特征 $F_{pan} \in R^{4H \times 4W \times C_{feat}}$ 。 C_{feat} 为网络的基础特征通道数。

接着, 这两个特征流被送入一个门控融合模块进行首次交互, 生成融合特征 F_{fused} 。该特征随后进入由 N 个堆叠的增强型注意力残差模块 (EARM) 构成的网络

主体,进行深度非线性特征变换:

$$F_{\text{deep}} = H_{\text{EARM}_i} \{ \cdots [H_{\text{EARM}_i}(F_{\text{fused}})] \cdots \} \quad (3)$$

经过主体网络处理后,我们引入一个空间注意力模块 H_{SA} 来进一步提炼空间细节。最后,通过一个重建模块 H_{recon} 将深度特征映射回图像空间,并与 I_{b_ms} 进行全局残差相加,得到最终的高分辨率输出 I_{HRMS} :

$$I_{\text{HRMS}} = H_{\text{recon}}[H_{\text{SA}}(F_{\text{deep}})] + I_{b_s} \quad (4)$$

这种全局残差学习的结构使网络能够专注于学习高频细节残差,极大地稳定了训练过程并提升了性能。

2.2 增强型注意力残差模块 (EARM)

EARM 是我们为全色锐化任务专门设计的核心组件,它在 EDSR 的基础残差块上进行了功能扩展,集成了多样的注意力机制与先进的融合策略。

多图像特征融合: 与单图像超分不同,全色锐化需要有效融合来自 MS 和 PAN 两种模态的信息。在我们的架构中,我们并非简单地在输入端拼接原始图像,而是在特征层面进行融合。具体来说,上采样后的 MS 图像 (b_ms) 和 PAN 图像 (x_pan) 首先通过各自独立的 3×3 卷积层 (head 和 pan_proj) 被映射到高维特征空间。这一步骤使得网络可以自适应地学习两种模态的初始特征表示,为后续的深度融合奠定基础通道与空间注意力机制为了让网络能够聚焦于信息量更丰富的特征区域,我们引入了通道注意力和空间注意力机制^[19]。通道注意力 (CA, channel attention) 旨在建模不同特征通道之间的依赖关系。对于一个输入特征图 $X \in R^{H \times W \times C}$,我们首先通过自适应全局平均池化 (Adaptive Average Pooling) 操作 $F_{\text{avg_pool}}$ 来聚合每个通道的空间信息,得到一个通道描述符 $z \in R^{1 \times 1 \times C}$ 。随后,该描述符通过一个由两个卷积层构成的 MLP (多层感知机) 来学习通道间的非线性关系,并由 Sigmoid 函数 σ 生成最终的通道权重 $s \in R^{1 \times 1 \times C}$ 。整个过程可表示为:

$$s = \sigma\{H_{\text{MLP}}[F_{\text{avg_ool}}(X)]\} = \sigma[W_2 \delta(W_1 z)] \quad (5)$$

其中: W_1 和 W_2 是 MLP 的权重, δ 是 ReLU 激活函数。最终,输出特征 $\hat{X}$ 通过将权重 s 与输入特征 X 进行逐通道相乘得到: $\hat{X} = s \cdot X$ 。

空间注意力 (Spatial Attention): 该模块则专注于定位信息最丰富的空间区域。我们将输入特征图 X 在通道维度上分别进行平均池化和最大池化,得到两个二维的空间描述符 $X_{\text{avg}} \in R^{H \times W \times 1}$ 和 $X_{\text{max}} \in R^{H \times W \times 1}$ 。将这两个描述符拼接 (Concatenate) 后,通过一个 7×7 的卷积层 $H_{\text{conv}7 \times 7}$ 和 Sigmoid 函数,生成空间注意力图 $M_s \in R^{H \times W \times 1}$:

$$M_s = \sigma\{H_{\text{conv}7 \times 7} \{ [F_{\text{avg_ool}}(X); F_{\text{max_ool}}(X)] \} \} \quad (6)$$

最终输出 $\hat{X}$ 由该注意力图与输入特征进行逐元素相乘得到: $\hat{X} = M_s \cdot X$ 。

2.3 门控融合模块

为了实现 MS 与 PAN 特征的智能融合,我们设计了一个高效的门控融合模块。不同于简单的相加或拼接,该模块能够动态地控制 PAN 特征的注入强度。我们将 MS 特征 F_{ms} 和 PAN 特征 F_{pan} 在通道维度上进行拼接,然后送入一个门控网络 H_{gate} 。该网络由两个 1×1 卷积层和一个 Sigmoid 激活函数构成,其输出是一个与特征维度相同的门控信号 $g \in R^{4H \times 4W \times C_{\text{low}}}$,其值域在 (0, 1) 之间:

$$g = \sigma\{H_{\text{gate}}[(F_{ms}; F_{pan})]\} \quad (7)$$

这个门控信号 g 随后被用于调制 PAN 特征,最终的融合特征 F_{fused} 通过以下方式计算得出:

$$F_{\text{fused}} = F_{ms} + F_{pan} \odot g \quad (8)$$

其中: $\odot$ 表示逐元素相乘。这种机制允许网络根据局部图像内容,自适应地决定 PAN 空间细节的注入程度,从而在有效提升清晰度的同时,最大程度地保留原始的光谱信息,避免光谱扭曲。

2.4 训练目标

为了从不同维度优化重建图像的质量,我们采用了由 L_1 损失、SSIM 损失和 Focal 损失构成的复合损失函数。

L_1 损失 (或称为 Charbonnier 损失^[20]) 通过计算生成图像 O 与真实图像 G 之间的平均绝对误差来衡量像素级的保真度,它对异常值不敏感,有助于生成更清晰的结果。其定义如下:

$$L_1 = \frac{1}{N} \sum_{i=1}^N \sqrt{(O_i - G_i)^2 + \epsilon^2} \quad (9)$$

其中: N 是像素总数, ϵ 是一个用于保证数值稳定性的极小常数 (例如 $1e-6$)。

结构相似性 (SSIM) 损失^[21]: 从亮度、对比度和结构 3 个方面衡量图像的相似性,比 L_1 损失更符合人类的视觉感知。它有助于保持图像的高频结构细节。损失函数定义为 [SSIM 公式见 (15)]:

$$L_{\text{SSIM}} = 1 - \text{SSIM}(O, G) \quad (10)$$

Focal 损失: 为了让模型在训练过程中更加关注那些难以重建的“硬”像素 (即误差较大的像素),我们借鉴了分类任务中的 Focal Loss 思想,并将其应用于回归任务。其公式如下:

$$L_{\text{Focal}} = \frac{1}{N} \sum_{i=1}^N d_i \cdot \frac{(255 \cdot d_i)^{r_1}}{255} \quad (11)$$

其中: $d_i = |O_i - G_i|$ 是单个像素的绝对误差, r_1 是一个调节因子。该损失函数通过对误差较大的像素赋予更高的权重,从而引导模型集中力量攻克重建难点。

最终,我们的总损失函数是以上三者的加权和,根据实验验证,我们将权重分别设置为 1.5、4.0 和 1.5:

$$L_{\text{total}} = 1.5L_1 + 4.0L_{\text{SSIM}} + 1.5L_{\text{Focal}} \quad (12)$$

3 实验

3.1 实现目的

为了验证模型的性能和可靠性，我们采用了 3 个在全色锐化领域被广泛使用的公开遥感数据集：WorldView-2 (WV2)，WorldView-3 (WV3)，和 GaoFen-2 (GF2)。为了提升模型的泛化能力并简化部署流程，我们采用了一种“all-in-one”的联合训练策略。具体而言，我们将 3 个数据集的训练集进行合并，让模型在包含所有数据域的混合数据上进行端到端的训练，共计 500 个轮次。训练完成后，我们使用这一个统一的模型，分别在 WV2，WV3 和 GF2 各自独立的测试集上进行性能评估。

3.2 实验方法

实验环境：本文所有实验均在 PyTorch 深度学习框架下完成，并使用单张 NVIDIA RTX 4090 GPU 进行模型训练与测试。我们采用 ADAM^[22] 优化器进行模型参数的优化，其中学习率初始值设置为 5×10^{-4} ，动量参数 β_1 和 β_2 分别设置为 0.9 和 0.999。为了在训练后期更精细地调整模型权重，我们采用了余弦退火学习率策略，在总共 500 个训练轮次 (epoch) 中将学习率平滑衰减至一个较小值。

训练方式：在训练阶段，我们从数据集中裁剪图像块 (patch) 进行训练。输入的低分辨率多光谱图像块 (MS) 尺寸为 $32 \times 32 \times C$ (C 为光谱通道数)，对应的高分辨率全色图像块 (PAN) 尺寸为 $128 \times 128 \times 1$ 。这意味着我们的网络处理的是 4 倍空间分辨率提升任务。网络内部会将 MS 图像上采样至 $128 \times 128 \times C$ ，并与 PAN 特征融合，最终输出尺寸为 $128 \times 128 \times C$ 的高分辨率多光谱图像。训练时的批处理大小 (batch size) 设置为 16。

具体参数配置：网络基础特征通道数为 36，核心的 EARM 模块数量为 10 个，残差块内的缩放因子设置为 0.85，并启用了我们设计的全部注意力及门控融合机制。为了确保对比的公平性，所有方法的计算复杂度 (FLOPs) 均在统一的 128×128 输出分辨率下进行测试。特别地，对于像 PanFlow 这样的多步生成方法，我们将其所有步骤的 FLOPs 进行累加作为其总计算量，以真实反映其推理开销。

评价指标：为了从空间和光谱两个维度全面、客观地评估生成图像的质量，采用了 3 种业界公认的评价指标：峰值信噪比 (PSNR)、结构相似性 (SSIM) 和光谱角映射 (SAM)^[23]。其中，PSNR 和 SSIM 主要评估空间结构的保真度，其值越高代表图像质量越好；SAM 则用于衡量光谱信息的失真程度，其值越低代表光谱保真度越高。

峰值信噪比 (PSNR)：PSNR 通过计算重建图像 I_{HRMS} 与真实图像 I_{GT} 之间的均方误差 (MSE) 来衡量图像失真。其定义如下：

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right) \quad (13)$$

$$\text{MSE} = \frac{1}{H \times W \times C} \sum_{i=1}^H \sum_{j=1}^W \sum_{k=1}^C [I_{\text{GT}}(i, j, k) - I_{\text{HRMS}}(i, j, k)]^2 \quad (14)$$

其中： MAX_I 是图像像素值的最大可能值 (例如对于 8 位图像是 255)。

结构相似性 (SSIM)：SSIM 从亮度 (l)、对比度 (c) 和结构 (s) 3 个方面衡量图像相似性，更符合人类视觉感知。对于图像块 x 和 y ，其定义为：

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (15)$$

其中： μ 和 σ^2 分别代表均值和方差， σ_{xy} 是协方差， C_1 和 C_2 是用于避免分母为零的稳定常数。最终的 SSIM 值是所有图像块计算结果的平均值。

光谱角映射 (SAM)：SAM 通过计算重建图像和真实图像在每个像素位置的光谱向量之间的夹角来衡量光谱失真。对于像素 i 的两个光谱向量 v_i 和 $\hat{v}_i$ ，其光谱角为：

$$\text{SAM}(v_i, \hat{v}_i) = \arccos \left(\frac{v_i \cdot \hat{v}_i}{\|v_i\| \cdot \|\hat{v}_i\|} \right) \quad (16)$$

最终的 SAM 值是整幅图像所有像素光谱角的平均值。

3.3 实验结果

定量比较：

将本文提出的方法与多种经典及先进的全色锐化方法进行了定量比较，包括 PNN^[12]、PanNet^[24]、MSDC-NN^[13]、PanFlow^[18]、PSCINN^[17] 和 Pan-Mamba^[25]。所有方法在相同的实验环境下进行训练和测试，结果如表 1 所示。

在 GaoFen-2 数据集上，我们方法的优势尤为显著。其 PSNR 值达到了 47.98 dB，SSIM 值达到了 0.987 8，SAM 值低至 0.010 3，这三项核心指标均大幅领先于所有对比方法。例如，相较于性能同样不俗的 PanFlow 模型，我们的方法在 PSNR 上高出 1.62 dB，这在高质量重建领域是一个巨大的提升。这表明我们的网络在处理该卫星数据时，无论在空间结构保真度还是光谱一致性上都达到了顶尖水平。

在 WorldView-II 数据集上，我们的方法同样展现出极强的竞争力。PSNR 和 SSIM 指标分别为 42.23 dB 和 0.971 7，均位列第一。相较于参数量更大的 Pan-Mamba (488.8 k)，我们的模型 (271 k) 在参数量减少约 45% 的情况下，依然在所有指标上实现了超越，证明了我们架构设计的优越性。

表 1 在 WorldView-II, WorldView-III, 以及 GaoFen-2 数据集上的实验结果

方法	参数量/k	FLOPs/G	WorldView-II			WorldView-III			GaoFen-2		
			PSNR	SSIM	SAM	PSNR	SSIM	SAM	PSNR	SSIM	SAM
PNN	68.9	2.26	39.82	0.954	0.028 2	29.49	0.900 5	0.086 1	43.14	0.966 7	0.017 8
PanNet	80.3	2.63	38.98	0.946 8	0.030 1	29.12	0.892 7	0.093 5	43.26	0.966 8	0.017 6
MSDCNN	239	7.83	40.31	0.958	0.026 7	29.63	0.903 3	0.083 3	43.21	0.967 1	0.017 6
PanFlow	87.3	11.44	41.11	0.964 5	0.024 3	30.04	0.910 6	0.079 9	46.36	0.982 5	0.012 5
PSCINN	3 321.5	108.84	35.60	0.896 7	0.033 6	22.61	0.553 8	0.111 5	42.69	0.961 6	0.018 1
Pan-Mamba	488.8	16.02	41.39	0.966 3	0.023 6	30.17	0.917 4	0.077 9	46.98	0.972 5	0.016 4
本文方法	271	8.88	42.23	0.971 7	0.021 4	30.61	0.924 8	0.074 6	47.98	0.987 8	0.010 3

在更具挑战性的 WorldView-III 数据集上，由于其图像特性，所有模型的性能指标普遍有所下降，但这更能考验模型的鲁棒性。在此数据集上，我们的方法 PSNR 达到了 30.61 dB，SSIM 达到了 0.924 8，再次超越了所有对手，显示了模型在复杂场景下的稳定重建能力。

在模型效率方面，我们的方法同样表现出色。与参数量高达 3 321.5 k、计算量高达 108.84 G FLOPs 的 PSCINN 相比，我们的模型在性能上实现了碾压式的超越，而参数量和计算量仅为它的 8% 和 8.2%。即使与 Pan-Mamba 相比，我们的模型也以更少的参数（271 k vs 488.8 k）和更低的计算量（8.88 G vs 16.02 G FLOPs）取得了更优异的成绩。综上所述，定量结果有力地证明了本文方法在达到 SOTA 性能的同时，兼顾了模型的轻量化与高效性。

定性比较：
除了客观的数值指标，我们还进行了视觉效果的主观比较。图 2、图 3 和图 4 分别展示了在 GaoFen-2、WorldView-II 和 WorldView-III 数据集上，不同方法的锐化结果对比。

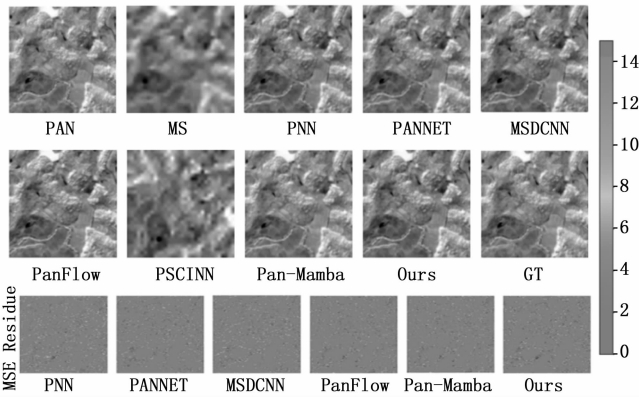


图 2 GaoFen-2 数据集上的视觉对比

通过细致观察可以发现，传统方法如 PNN 和 Pan-Net 的生成结果普遍存在细节模糊、边缘不清晰的问题，并且伴有明显的光谱失真，导致图像整体观感较差。一些较新的方法虽然在空间细节的锐度上有所提

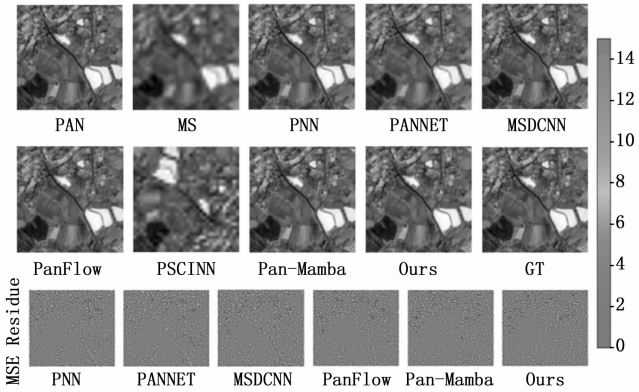


图 3 WorldView-II 数据集上的视觉对比

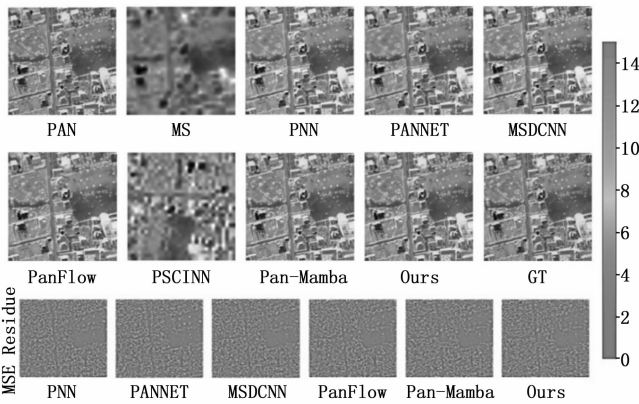


图 4 WorldView-III 数据集上的视觉对比

升，但可能引入不自然的伪影或纹理（如 Pan-Mamba 在某些区域的过度锐化）。相比之下，本文方法生成的图像在视觉上最为清晰、自然且真实。

在图 3 所示的 WorldView-II 数据集中，场景包含了密集的建筑物和道路。本文方法生成的图像中，建筑物的轮廓分明，屋顶的纹理细节清晰可见，道路边缘平滑锐利。反观其他方法，或多或少存在边缘模糊或颜色渗透的问题。下方的均方误差（MSE）残差图也验证了这一点：我们的方法对应的残差图（Ours）颜色最深且分布最均匀，表明其与真实图像（GT）的像素级差异最小，而其他方法的残差图则呈现出更亮、更结构

化的误差模式。

在图 4 所示的 WorldView-III 数据集中，这是一个具有复杂城市地貌的挑战性任务，包含了大量微小的房屋和不规则的道路网。在这种场景下，我们的方法展现了强大的细节刻画能力，能够准确地重建出微小房屋的结构，而不会像其他方法那样出现结构粘连或形态失真问题。

在图 2 所示的 GaoFen-2 数据集中，场景主要为包含植被与裸土的郊野环境，这对算法的光谱保真度提出了极高要求。我们的方法在增强地物边缘清晰度的同时，极好地保持了植被和土地的原始色彩与色调层次，其颜色表现最接近真实图像（GT），避免了许多方法中常见的色偏或“水洗感”问题。MSE 残差图再次证明，我们的方法在光谱和空间两个维度上的重建误差均为最低。

综上所述，无论是在结构复杂的城市环境，还是在光谱信息丰富的自然场景，定性比较结果均一致表明，本文方法能够生成具有更高保真度、更丰富细节和更少伪影的高质量全色锐化图像。

消融实验：

为了验证我们模型中各个创新组件的有效性，我们设计了一系列消融实验。我们以一个基础模型为起点，逐步引入或替换关键模块，包括增强模块（指代注意力机制、门控融合等设计）和我们提出的复合损失函数（All Loss）。实验在 WorldView-II 数据集上进行，结果如表 2 所示。

表 2 消融实验测试指标对比 (WorldView-II)							
消融网络	引入增强模块	增加模型大小	Loss	PSNR	SSIM	SAM	参数量 /k
(a)	否	否	L1	35.197 3	0.882 7	0.035 3	171.2
(b)	是	否	L1	39.241 2	0.951 6	0.029 2	271.0
(c)	否	是	L1	39.923 8	0.955 7	0.027 7	557.6
(d)	否	否	All	41.970 2	0.970 1	0.022 3	171.2
(e)	是	否	All	42.230 3	0.971 7	0.021 4	271.0

从表 2 中可以看出：

(a) vs (d)：在基础网络上，仅将 L1 损失替换为复合损失（All），PSNR 就从 35.197 3 dB 大幅提升至 41.970 2 dB，证明了我们设计的复合损失函数的巨大威力。

(a) vs (b)：在仅使用 L1 损失的情况下，引入我们的增强模块，模型的参数量有所增加，性能也得到提升，证明了增强模块在特征提取和融合上的有效性。

(d) vs (e)：在使用复合损失的基础上，再引入增强模块，PSNR 进一步提升至 42.230 3 dB，达到了最佳性能。

(c) vs (e)：对比模型大小，我们的最终模型（e，

271 k 参数）在参数量远小于模型（c，557.6 k 参数）的情况下，性能反而更优，这说明我们设计的轻量化增强模块比盲目增大模型尺寸更有效。

综合来看，消融实验清晰地表明，我们提出的增强模块和复合损失函数对模型的最终性能都有着至关重要的贡献，并且两者之间存在协同增益效应。

4 结束语

针对当前全色锐化领域模型日趋复杂、计算开销巨大的问题，本文受 EDSR “减法”设计哲学的启发，提出了一种新颖高效的全色锐化网络架构。该架构的核心是本文全新设计的增强型注意力残差模块（EARM），它通过集成注意力机制与门控融合，在保持轻量化的同时实现了强大的特征提取能力。结合精心设计的复合损失函数，在多个公开数据集上的实验结果表明，所提方法在定量指标与视觉质量上均优于多种主流模型，消融研究也证实了 EARM 模块中各组件的有效性。本文的研究证明，遵循高效精炼的模块设计原则，而非盲目堆叠复杂结构，是实现全色锐化性能突破的一条更优路径，为未来遥感图像处理领域提供了兼具高性能与实用性的解决方案^[26-31]。

参考文献：

[1] WALD L. Data fusion: definitions and architectures; fusion of images of different spatial resolutions [M]. Paris: Presses des MINES, 2002.

[2] DDENG L J, VIVONE G, PAOLETTI M E, et al. Machine learning in pansharpening: a benchmark, from shallow to deep networks [J]. IEEE Geoscience and Remote Sensing Magazine, 2022, 10 (3): 279 - 315.

[3] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image Super-Resolution [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2017: 136 - 144.

[4] CHAVEZ P, KWARTENG A. Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis [J]. Photogrammetric Engineering and Remote Sensing, 1989, 55 (3): 339 - 348.

[5] CARPER W, LILLESAND T, KIEFER R, et al. The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data [J]. Photogrammetric Engineering and Remote Sensing, 1990, 56 (4): 459 - 467.

[6] GILLESPIE A R, KAHLE A B, WALKER R E. Color enhancement of highly correlated images. II. Channel ratio and “chromaticity” transformation techniques [J]. Remote Sensing of Environment, 1987, 22: 343 - 365.

- [7] KING R L, WANG J. A wavelet based algorithm for pan sharpening landsat 7 imagery [C] //Proceedings of the IEEE 2001 International Geoscience and Remote Sensing Symposium (IGARSS 2001), IEEE, 2001, 2: 849 – 851.
- [8] LABEN C A, BROWER B V. Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening [P]. US Patent 6, 011, 875, 2000.
- [9] LIU J G. Smoothing filter-based intensity modulation: a spectral preserve image fusion technique for improving spatial details [J]. International Journal of Remote Sensing, 2000, 21 (18): 3461 – 3472.
- [10] YOKOYA N, YAIRI T, IWASAKI A. Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion [J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 50 (2): 528 – 537.
- [11] PALSSON F, SVEINSSON J R, ULFARSSON M O. A new pansharpening algorithm based on total variation [J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11 (1): 318 – 322.
- [12] MASI G, COZZOLINO D, VERDOLIVA L, et al. Pan-sharpening by convolutional neural networks [J]. Remote Sensing, 2016, 8 (7): 594.
- [13] YUAN Q, WEI Y, MENG X, et al. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11 (3): 978 – 989.
- [14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 – 778.
- [15] MA J, YU W, CHEN C, et al. Pan-GAN: an unsupervised pan-sharpening method for remote sensing image fusion [J]. Information Fusion, 2020, 62: 110 – 120.
- [16] ZHOU H, LIU Q, WENG D, et al. Unsupervised cycle-consistent generative adversarial networks for pan sharpening [J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1 – 14.
- [17] WANG J, LU T, HUANG X, et al. Pan-sharpening via conditional invertible neural network [J]. Information Fusion, 2024, 101: 101980.
- [18] YANG G, CAO X, XIAO W, et al. PanFlowNet: a flow-based deep network for pan-sharpening [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023: 16857 – 16867.
- [19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 30: 5998 – 6008.
- [20] CHARBONNIER P, BLANC-FERAUD L, AUBERT G, et al. Two deterministic half-quadratic regularization algorithms for computed imaging [C] //Proceedings of the 1st International Conference on Image Processing. IEEE, 1994, 2: 168 – 172.
- [21] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity [J]. IEEE Transactions on Image Processing, 2004, 13 (4): 600 – 612.
- [22] KINGMA D P, BA J. Adam: a method for stochastic optimization [C] //Proceedings of the 3rd International Conference on Learning Representations (ICLR), 2015.
- [23] ALPARONE L, AIAZZI B, BARONTI S, et al. Multispectral and panchromatic data fusion assessment without reference [J]. Photogrammetric Engineering and Remote Sensing, 2008, 74 (2): 193 – 200.
- [24] YANG J, FU X, HU Y, et al. PanNet: a deep network architecture for pan-sharpening [C] //Proceedings of the IEEE International Conference on Computer Vision, 2017: 5449 – 5457.
- [25] HE X, CAO K, ZHANG J, et al. Pan-mamba: effective pan-sharpening with state space model [J]. Information Fusion, 2025, 115: 102779.
- [26] LI X, HU T, CAO K, et al. Pan-Sharpener via Causal-Aware feature distribution calibration [J]. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1 – 14.
- [27] LIAO W, HUANG X, VAN COILLIE F, et al. Two-stage fusion of thermal hyperspectral and visible RGB image by PCA and guided filter [C] //Proceedings of the 2015 7th Workshop on Hyperspectral Image and Signal Processing (WHISPERS), IEEE, 2015: 1 – 4.
- [28] SCHOWENGERDT R A. Reconstruction of multispatial, multispectral image data using spatial frequency content [J]. Photogrammetric Engineering and Remote Sensing, 1980, 46 (10): 1325 – 1334.
- [29] YUHAS R H, GOETZ A F, BOARDMAN J W. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm [C] //Summaries of the Third Annual JPL Airborne Geoscience Workshop, Volume 1: AVIRIS Workshop, JPL, 1992.
- [30] ZHANG J, CAO K, YAN K, et al. Frequency decoupled domain-irrelevant feature learning for pan-sharpening [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024.
- [31] ZHONG Y, WU X, DENG L J, et al. SSiDiff: Spatial-spectral integrated diffusion model for remote sensing pan-sharpening [J]. Advances in Neural Information Processing Systems, 2025, 37: 77962 – 77986.