

基于联合注意力的渐进式遥感图像融合

程月馈, 刘晓文

(中国人民公安大学 信息网络安全学院, 北京 100038)

摘要: 为发挥遥感图像在国防军事、公共安全、环境监测等领域的重要作用, 如何融合已配准的高分辨率全色图像与低分辨率多光谱图像的互补信息成为当前研究的重点。尽管近年来全色锐化方法已取得较大进步, 但大多数方法仍受到以下限制: 一方面, 利用 Wald 协议退化生成不同尺寸图像时会造成信息损失; 另一方面, 受到网络结构和单一注意力的限制, 无法同时利用全局和局部特征。为解决以上问题, 本文提出了基于联合注意力的渐进式网络 (Pan-sharpening based on multi-attention progressive network), 称为 MAPNet。在该网络中, 首先采用多阶段训练以减小尺寸变化带来的光谱和细节损失。其次设计联合注意力模块, 将自注意力、空间注意力和通道注意力结合, 实现对全局特征和局部特征、空间特征和通道特征的多模态分析, 进一步提高 MAPNet 对纹理细节的保留能力。在高分二号卫星上进行大量对比实验和消融实验, 定性和定量结果表明, 本文方法融合效果优于其他 10 种方法, 能够改善光谱失真和细节纹理丢失等问题。

关键词: 遥感图像融合; 深度学习; 联合注意力机制; 多光谱图像; 全色图像

中图分类号: TP391 **文献标识码:** A **DOI:** 10.3969/j.issn.1001-5078.2024.05.023

Pan-sharpening based on multi-attention progressive network

CHENG Yue-kui, LIU Xiao-wen

(Department of Information and Cyber Security, People's Public Security University of China, Beijing 100038, China)

Abstract: Due to various physical and technical limitations, the radiation energy received by different sensors and the amount of data collected vary, and a single sensor cannot simultaneously obtain high spatial and spectral images. Therefore, it is necessary to develop an ideal application-oriented technique for generating multi-spectral image with high spatial resolution. The pan-sharpening method fuses the low spatial resolution multispectral image with the high spatial resolution panchromatic image to obtain a hyperspectral image with rich spatial spectral information. Although significant progress has been made in pan-sharpening methods in recent years, most methods still have two limitations: firstly, limited by network structure and single attention mechanism, global and local features cannot be used simultaneously, resulting in loss of spatial information; secondly, using the Wald protocol to obtain high-resolution multispectral images leads to loss of spectral and detail information. To address these problems, this paper proposes a pan-sharpening framework MAPNet based on multiple attention progressive network. In order to extract more important information, we fully utilize the feature information contained in panchromatic and multispectral images to reduce the interference of redundant information. The low resolution and full resolution phases are closely linked using a progressive pattern. MAPNet trains the ability to extract global information, spectral information and gradient information to reduce the loss of spectrum and detail due to size changes. The multi-attention module combines self-attention, spatial atten-

基金项目: 国家重点研发计划项目 (No. xxxxxx31400) 资助。

作者简介: 程月馈 (1999 -), 女, 硕士研究生, 主要研究方向为遥感图像融合技术。E-mail: chengyuekui@sina.com

收稿日期: 2023-08-21; **修订日期:** 2023-09-26

tion and channel attention to achieve multi-modal analysis of global features, local features, spatial features and channel features, thereby further improving MAPNet's ability to retain texture details. The algorithm proposed in this paper is compared with the existing traditional methods BT-H, C-MTF-GLP-CBD, GS, BDS, PRACS and deep learning methods MUCNN, MDCUN, Band-Aware, PNN and TFNet on the GF-2 dataset. Additionally, this paper records the performance of models with different stages and structures. Objective measurements include RMSE, RASE, SAM, ER-GAS, QAVE, SSIM, FSIM, QNR, D_s , D_λ . By combining subjective visual assessment with objective evaluation, the results indicate that MAPNet fusion images retain more spectral and detail information.

Keywords: remote sensing image fusion; deep learning; multi-attention; multispectral image; panchromatic image

1 引言

由于 IKONO、SPOT、QuickBird、WorldView3 等卫星的快速发展,遥感影像因实时性强、全球性等优点,在各类对地观测如灾害监测^[1]、资源管理、环境保护等任务中彰显出其应用的广泛性和必要性。但受到多种物理和技术限制,不同传感器接收的辐射能和采集数据量不同,单个传感器无法同时获得高空间分辨率和高光谱分辨率图像^[2]。因此,有必要将低空间分辨率高光谱分辨率的多光谱图像(Multispectral Image, MS)与高空间分辨率低光谱分辨率的全色图像(Panchromatic Image, PAN)进行融合,得到具有丰富空间和光谱信息的高光谱图像。

在过去的几十年里,大量学者致力于设计和优化全色锐化算法,其中代表性方法主要包括四大类^[3]:组件替换法(Component Substitution, CS)、多尺度分解法(Multiresolution Analysis, MRA)^[4]、变分模型优化法(Variational model optimization, VO)^[5]以及深度学习法(Deep Learning, DL)^[6]。

传统遥感图像融合方法由于变换策略不当和求解条件不合理,往往存在光谱和空间畸变。相比之下,DL 具有强大的非线性拟合和特征提取能力,全色锐化效果更好。为此研究者们将深度学习引入图像全色锐化领域,克服了传统方法需要手动选择参数和模型的局限性。然而,依靠单一网络结构和注意力机制不能完全提取图像特征,低分辨率尺寸融合无法充分利用全分辨率下的空间和光谱信息,导致光谱与空间信息损失,融合效果不理想。为解决以上问题,本文提出基于联合注意力机制的渐进式网络,该模型的主要贡献概括如下:

(1)为平衡不同分辨率的融合性能,设计渐进式融合网络,旨在利用不同尺度图像信息逐步优化网络,减少因 Wald 协议^[7]对光谱向量一致性和各波段图像一致性理想化度量造成的信息损失。

(2)为提高网络对局部和全局特征的表征能

力,提出通道注意力、空间注意力和自注意力组成的联合注意力模块,密切联系局部和全局特征,避免单一注意力导致的特征信息保留不完全。

(3)利用高分二号(GF-2)卫星的低分辨率和全分辨率数据集对该方法与其他 10 种对比方法进行测试,分别从主观视觉和客观定量两方面对融合结果的空间结构与光谱保真度进行质量评估,证明本方法全色锐化效果的优越性。

本文的其余部分结构如下。第一节介绍全色锐化的相关工作,第二节详细描述所提方法,包括框架概述、联合注意力机制和损失函数。第三节提供了详细的实验参数设置,并通过与几种经典方法进行比较来验证所提方法的优越性。

2 相关工作

遥感图像是通过遥感技术获取的展现地球表面特征和变化的图像,应用较为广泛。在图像分类和目标检测等实际任务中,单个传感器难以满足图像在空间和光谱域均具有高分辨率的要求。为此光学对地观测卫星搭载了两种光学传感器,获取特性不同但互补的多模态数据,其中全色传感器仅获取单波段的高空间分辨率图像,多光谱传感器获取多波段的低空间分辨率图像。全色锐化技术可将 PAN 图像和 MS 图像优势互补,获得高空间分辨率的多光谱图像。

2.1 传统方法

CS 方法的主要思想是将 N 波段图像投影到另一个向量空间,并假设投影变换能够将 MS 图像的空间结构和光谱信息分离至不同的投影分量,然后将 PAN 图像替换为从 MS 图像中提取的分量。该方法空间细节信息保真度高、速度快、易于实现。MRA 方法先利用金字塔或小波变换函数将 MS 和 PAN 图像分解成多个尺度,再将 PAN 图像的空间信息注入到不同尺度的 MS 图像中。与 CS 方法相比, MRA 方法在锐化 MS 图像时带来的光谱畸变较小,

但融合图像中可能存在局部空间伪影。VO 方法需要设置各种先验和约束条件的参数,融合效果较好,但普适性较差。综上,相较于深度学习方法,传统方法计算效率和稳定性虽高,但处理复杂问题和非线性问题时仍存在局限。

2.2 深度学习

多层神经网络的训练和优化可以从原始图像中提取语义信息和空间结构,这些高级的特征有利于捕捉图像的融合规则,从而实现了对图像的理解和分析。深度学习技术在计算机视觉任务中表现优异,越来越多的研究人员尝试使用深度学习模型来解决全色锐化问题。

2.2.1 有监督学习

有监督学习方法使用端到端网络来实现全色锐化过程,典型的例子包括 PNN (CNN-based Pan-sharpening)、TFNet (Two-stream Fusion Network)、MUCNN (Multiscale U-shaped Convolution Neural Network)、Band-Aware (PAN-Guided Band-Aware Multispectral Feature Enhancement Network) 等网络。Masi^[6]等在超分辨率结构的基础上,将几个典型的遥感非线性辐射指数图作为输入,利用三层卷积网络 PNN 来实现全色锐化。Liu^[8]等设计了将 PAN 和 MS 图像在特征域融合的双流融合网络 TFNet。Wang Y 等^[9]提出利用多尺度 U 型卷积神经网络 MUCNN 以充分利用图像的多光谱信息。Zhou M^[10]等人设计了 Band-Aware 网络,利用全色特征进行波段感知,选择性地增强多光谱每个谱带的信息。

2.2.2 无监督学习

如何获得参考图像是有监督学习不可避免的问题。Wald 协议提出将原始 MS 和 PAN 图像依据光学成像原理执行下采样和模糊退化过程,比例按照 PAN 图像和 MS 图像的空间分辨率之比,仿真生成低分辨率多光谱和全色图像,并利用原始数据作为伪地面真实多光谱图像检验网络的融合效果。为降低退化策略造成的光谱和细节损失,研究者们提出利用无监督网络完成全色锐化任务。例如 Zhou C^[11]等人基于自动编码器和感知损失,提出了不需要退化步骤进行训练的 PercepPan (Perceptual Pan-sharpening)。

无监督方法可以保留多光谱图像丰富的光谱信息和全色图像的纹理信息,但由于缺乏先验知识的指导,融合效果受多种因素影响。为此 Zhang H^[12]

等设计了有监督和无监督结合的网络 P2Sharpen (Progressive Pan-sharpening Network),该网络跨越了多光谱数据和全色数据之间的模态差异,建立了从 MS 到 PAN 图像的精确映射函数,并有效地平衡了低分辨率和全分辨率两个尺度上的性能。该方法光域信息保存较好,但未关注长距离依赖特征,部分融合结果空间细节模糊。

总体来看,尽管深度学习方法 (Deep Learning, DL) 比传统方法全色锐化表现好,但现有的 DL 方法缺乏所谓的高分辨率多光谱参考图像,大多采用退化生成的低分辨率图像作为输入,将原始 MS 图像作为伪地面真值来指导融合过程,存在无法完全利用高空间分辨率细节信息,特征的长距离依赖和短距离依赖把握不全面的问题。为进一步优化算法,本文提出跨模态提取全色和多光谱的特征信息,采用联合注意力机制提高网络的表征能力,利用不同尺寸图像补充空间和光谱信息。

3 本文方法

本文所提出的模型 MAPNet 是利用渐进式策略训练的联合注意力网络。渐进式策略中低分辨率图像训练网络的全局信息提取能力和光谱信息提取能力,全分辨率图像训练网络的梯度信息提取能力。联合注意力模块包括通道注意力、空间注意力、自注意力和损失函数,使得模型能够更好地理解图像的语义信息。

3.1 网络框架

针对低分辨率和全分辨率之间性能失衡的问题,本文提出渐进式学习策略如图 1。其中高度为 M ,宽度为 N 的全分辨率 PAN 图像表示为 $P_h \in R^{M \times N \times 1}$;波段数、高度和宽度为 B 、 m 、 n 的可观测多光谱图像表示为 $MS_h \in R^{m \times n \times B}$ 。 $P_L \in R^{\frac{M}{r} \times \frac{N}{r} \times 1}$ 和 $MS_L \in R^{\frac{m}{r} \times \frac{n}{r} \times B}$ 均为利用 Wald 协议^[8]生成的低分辨率图像, $\uparrow r$ 和 $\downarrow r$ 分别表示比例为 $r = \frac{M}{m} = \frac{N}{n} = 4$ 的上、下采样。区别于端到端的网络训练方式,渐进式一方面指将第一阶段输出作为第二阶段输入;另一方面意味着利用全分辨率和低分辨率数据集训练 MAPNet,并共享和优化同一组参数 Θ ,表示为:

$$F_L = F(MS_L \uparrow, P_L, \Theta) \quad (1)$$

$$F_h = F(F_L \uparrow, P_h, \Theta) \quad (2)$$

式(1)、(2)分别表示第一、二阶段全色锐化, F_L 、 F_h 分别表示低分辨率和全分辨率融合结果。

量转换为查询向量 q 、键向量 k 和值向量 v ，再转化成三个不同的矩阵 Q 、 K 和 V ，接下来利用公式 $S = Q \cdot K^T$ 计算输入序列中每个元素与其他元素之间的关联性，再用 $S_n = \frac{s}{\sqrt{d_k}}$ 将相似度矩阵进行归一化操作得到权重矩阵，通过 $P = \text{softmax}(S_n)$ 函数将权重转换为概率，最后将概率矩阵与输入序列进行加权 $Z = V \cdot P$ 得到对应位置元素的上下文信息。通过这种方式，基于全局考量的关键像素区域将获得更高的权重，从而在后续模型中被更多地考虑和利

用。这个过程可以统一为单独的函数：

$$M_{S_n} = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V \quad (5)$$

自注意力机制需要理解序列中不同元素之间的相对位置关系，因此需要位置编码。然而位置编码数量固定，无法适应不同长度的输入序列，使得自注意力机制无法满足全色锐化对普适性的要求。为克服该问题，本文选择将经过特征提取模块的全色、多光谱特征图 (fea_ms、fea_pan) 及联合注意力特征图 (fea_com) 输入自注意力网络，如图 3。

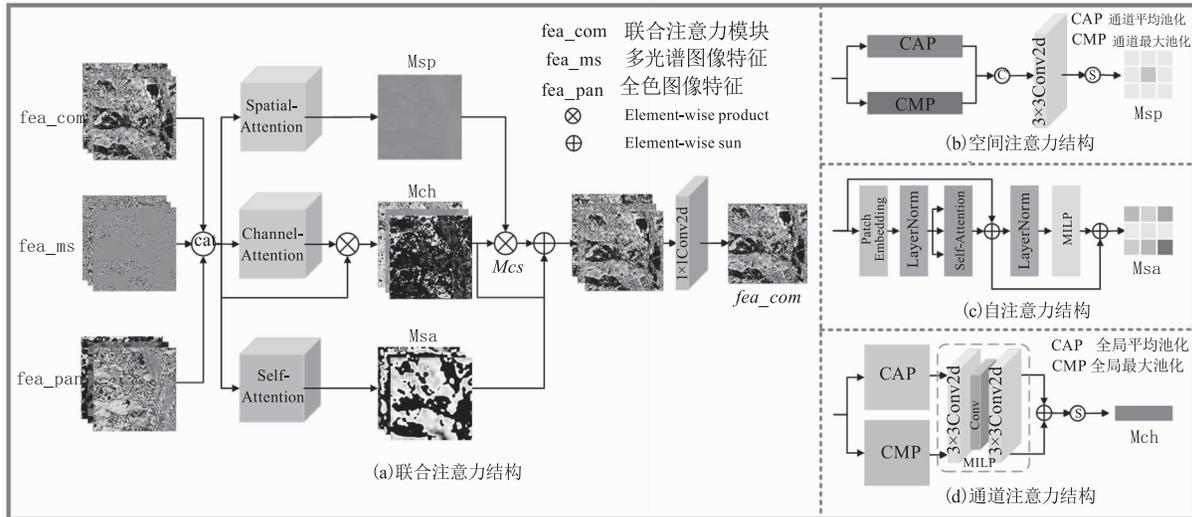


图 3 联合注意力模块

Fig. 3 Multi-attention Block

3.2.2 通道注意力

通道注意模块如图 3(d)，首先采用全局平均池化和全局最大池化层对空间维度上的特征图进行压缩。其次将压缩后的特征映射 F_{avg}^c 和 F_{max}^c 发送到包含一个隐藏层的多层感知器 (Multilayer Perceptron, MLP) 中。最后将 MLP 的输出逐像素相加后通过 sigmoid 激活函数，其数学表达式为：

$$\begin{aligned} M_{\text{ch}}(F) &= \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \\ &= \sigma(W_1(W_0(F_{\text{avg}}^c)) + W_1(W_0(F_{\text{max}}^c))) \end{aligned} \quad (6)$$

其中， σ 表示 sigmoid 函数， $W_0 \in R^{C/r \times C}$ ， $W_1 \in R^{C \times C/r}$ 。

3.2.3 空间注意力

空间注意力模块的目的是突出显示包含关键信息的区域如图 3(b)。首先沿通道维度压缩特征映射得到 F_{avg}^s 和 F_{max}^s ，然后将特征图 F_{avg}^s 和 F_{max}^s 拼接后进行卷积核大小为 7 的卷积操作，最后通过 sig-

moid 函数计算出空间注意图，其数学表达式如下：

$$\begin{aligned} M_{\text{sp}}(F) &= \sigma(f^{7 \times 7}([\text{AvgPool}(F); \\ &\text{MaxPool}(F)])) = \sigma(f^{7 \times 7}([F_{\text{avg}}^s; F_{\text{max}}^s])) \end{aligned} \quad (7)$$

该模块利用空间注意力、通道注意力关注局部特征，自注意力关注全局特征，令神经网络更加关注包含重要信息的目标区域，捕获 PAN 和 MS 模态的冗余和补充信息，从而优化全色锐化性能，提高目标检测的整体精度。

3.3 损失函数

如 3.1 节所述，MAPNet 的训练分全分辨率和低分辨率两个阶段进行，因此整体网络结构的损失分为两部分，表示为：

$$L = L_{\text{LR}} + L_{\text{HR}} \quad (8)$$

低分辨率阶段损失表示为 L_{LR} ，由两项内容约束：

$$L_{\text{LR}} = \alpha L_{\text{LR}}^1 + \beta L_{\text{LR}}^2 \quad (9)$$

其中， α, β 是平衡损失的权重； L_{LR}^1 是确保在空间退

化的前提下,还原分辨率融合图像具有与原始多光谱图像相同的光谱分布; L_{LR}^2 是受高分辨率多光谱图像约束的损失,公式为:

$$L_{LR}^1 = \frac{1}{N} \sum_{n=1}^N \|F_L^{(n)} \downarrow - MS_L^{(n)}\|_F^2 \quad (10)$$

$$L_{LR}^2 = \frac{1}{N} \sum_{n=1}^N \|F_L^{(n)} - MS_h^{(n)}\|_F^2 \quad (11)$$

若仅在低分辨率尺寸下依赖上述约束,全色锐化模型依然无法摆脱因尺寸变化产生的损失。因此,本文进一步引入了全分辨率阶段的损失函数来调节性能:

$$L_{HR} = \frac{1}{N} \sum_{n=1}^N \|F_h^{(n)} \downarrow - MS_h^{(n)}\|_F^2 \quad (12)$$

在充分约束的指导下,MAPNet 可以产生具有精确光谱分布和丰富空间信息的融合图像。

4 实验结果与分析

为证明 MAPNet 的有效性,本章首先针对渐进式策略和联合注意力模块进行消融实验,其次将本文方法与 5 种传统方法和 5 种深度学习方法进行对比实验。鉴于全色锐化在实际应用中需要通过不同要素(如植被、水体、土壤等)对地物“身份”进行识别,融合质量评估对融合结果的空间结构清晰度与光谱保真度均有较高要求,为此本文采用视觉分析和定量分析评估本文模型的融合质量。

4.1 数据集

实验采用 GF-2 卫星捕获的 MS 和 PAN 图像对,经过配准等一系列处理后根据 Wald^[8] 协议建立的数据集,包含两组 11511 对训练图像块和 541 对测试图像块,多光谱图像大小为 64×64 、 256×256 ,全色图像尺寸则对应为 256×256 、 1024×1024 。

4.2 参数设置及实现细节

网络的实现基于 Pytorch 深度学习框架,并在 NVIDIA A40 GPU 上迭代训练 100 次。为优化模型参数,使用 Adam 优化器并将学习率设置为 3×10^{-4} 。考虑到渐进式网络两阶段输入的图像尺寸差异较大,全分辨率阶段所需内存远大于低分辨率阶段,将批量大小设置为 1。

4.3 评价指标

为评估低分辨率数据集融合性能,本文选择七种常用的参考质量指标:均方根误差(RMSE)^[15]和相对平均光谱误差(RASE)^[16]、光谱角映射器

(SAM)^[17]、相对无量纲综合全局误差(ERGAS)^[18]、频带平均通用图像质量指数(QAVE)^[19]、结构相似性指数度量(SSIM)^[20]和特征相似性指数度量(FSIM)^[21]。值得一提的是,QAVE、SSIM 和 FSIM 的值越大表示质量越高,而 ERGAS、RMSE、RASE 和 SAM 的值越小表示性能越好。

全分辨率图像测试在卫星接收图像的原始分辨率下进行,因此本文使用三种常用的无参考指标光谱失真指数(D_λ)、空间失真指数(D_s)以及无参考质量评价指标(QNR)^[22]来评估全分辨率融合性能。 D_λ 衡量融合图像的光谱信息失真程度, D_s 衡量融合图像的空间信息丢失程度,其值越接近于 0 表示光谱失真或空间信息丢失越小,融合结果越好。

4.4 消融实验

在基于联合注意力机制的渐进式融合网络设计过程中,融合性能受到渐进式阶段次数 L 和联合注意力模块等初始化设置的直接影响。为此,本节针对以上重要因素设计了联合注意力机制单阶段(Att_1L)、三阶段(Att_3L)融合网络,无注意力机制两阶段(noAtt_2L)融合网络等消融实验。图 4、图 5 分别为低分辨率和全分辨率数据集消融实验视觉对比图,直观展示本文算法在细节补充和光谱保存的优越性能;表 1 为低分辨率和全分辨率数据集消融实验定量结果,可以看出本文提出的网络结构在 ERGAS 等七种客观指标上最优,SSIM 等三种客观指标次优。结合消融实验定量和定性结果,可以认为当渐进式网络阶段数为 2 且结合注意力机制时网络性能最佳。

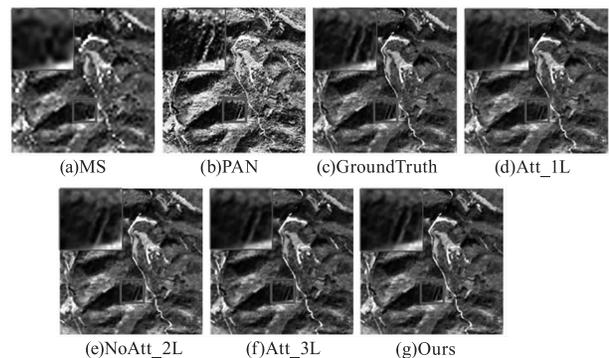


图 4 低分辨率数据集消融实验视觉比较 (选用红、绿、蓝 3 个波段显示)

Fig. 4 Visual comparison of ablation experiments from reduced-resolution datasets (select RGB for demonstration)

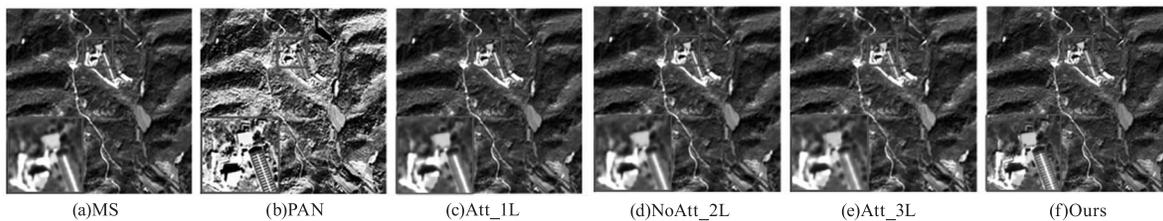


图 5 全分辨率数据集消融实验视觉比窃(选用红、绿、蓝 3 个波段显示)

Fig. 5 Visual comparison of ablation experiments from full-resolution datasets (select RGB for demonstration)

表 1 消融实验定量比较

Tab. 1 Quantitative comparison of ablation experiments

方法	低分辨率数据集							全分辨率数据集		
	ERGAS ↓	RMSE ↓	RASE ↓	SAM ↓	QAVE ↓	SSIM ↑	FSIM ↑	QNR ↑	D_A ↓	D_s ↓
Att_1L	<u>1.58 ± 0.03</u>	<u>1.60 ± 0.02</u>	<u>6.04 ± 0.32</u>	<u>1.89 ± 0.02</u>	0.9723	0.9821	0.9839	0.9184	0.0334	<u>0.0499</u>
NoAtt_2L	2.07 ± 0.08	2.01 ± 0.03	7.58 ± 0.81	2.04 ± 0.02	0.9695	0.9748	0.978	0.8849	0.0098	0.1061
Att_3L	3.06 ± 0.21	3.38 ± 0.54	11.07 ± 2.47	3.08 ± 0.13	0.9212	0.9482	0.9571	0.8537	0.0469	0.1037
Ours	1.53 ± 0.02	1.54 ± 0.02	5.78 ± 0.24	1.84 ± 0.03	0.9744	<u>0.9813</u>	<u>0.9825</u>	0.9330	<u>0.0266</u>	0.0346

4.5 对比实验结果与分析

将传统方法 PRACS^[23]、BDS^[24]、BT-H^[25]、C-MTF-GLP-CBD^[26]、GS^[22]，DL 方法 PNN^[6]、TF-Net^[8]、MUCNN^[9]、Band-Aware^[10]和 MDCUN^[27]等方法参数调整为最佳,在高分二号(GF-2)卫星相同的训练和测试样本上与本文方法进行对比。

4.5.1 低分辨率图像实验结果与分析

图 6 为 GF-2 卫星捕获的某山区图像融合结果和局部放大图。图 6(a)为上采样 4 倍后尺寸为 256 × 256 的多光谱图像。图 6(b)为 256 × 256 的全色图像。BT-H 变换的融合图像如图 6(c),出现光谱失真,梯田区域由黄绿变成棕红色。C-MTF-GLP-CBD 算法的融合结果不仅道路边缘存在伪影,而且光谱发生明显畸变,梯田区域显示为灰绿色。BDS 算法保留了较多光谱信息,但是放大区域中存在部分模糊。观察 GS 算法融合结果图 6(f)可以发现,道路部分出现伪影和不连续现象。PRACS 算法的融合结果如图 6(g),该算法在保留原始图像主要信息的基础上提高了细节信息质量。图 6(c) ~ (g)为传统算法的融合结果,以上方法受到提出策略等限制,无法呈现出细节纹理和光谱信息完整的融合结果。PNN 算法是深度学习首次在遥感图像融合领域取得的成果,其结果如图 6(h)。该算法不仅保留了多光谱图像的光谱信息,而且加入了全色图像的部分细节信息。TFNet 网络模型相比 PNN 算法细节信息保留能力有较大提高,其结果如图 6(i)。MUCNN 算法在多光谱的基础

上,保留了更多的空间细节信息,但道路部分出现了细节扭曲的情况如图 6(j)。MDCUN 算法融合效果如图 6(k),放大区域中道路仍存在边缘模糊。图 6(l)为 Band-Aware 算法的融合结果,在细节信息提升和光谱信息保留方面较图 6(c) ~ (k)均有较大进步,但仍存在一些白色噪声。本文提出的算法融合结果为图 6(m),与 GroundTruth 相比,光谱信息保持较好,房屋轮廓、道路边缘细节比较清晰,纹理细节质量得到了较好的提升。

图 7 为 GF-2 卫星捕获的某农耕地区图像融合结果和局部放大图。BT-H 融合结果中田地边缘出现伪影。BDS 算法部分边缘模糊。PRACS 方法边缘信息保留不完整。PNN 受到单一尺度网络的限制,特征学习能力较差,如图 7(h)中心区域田畦边缘不易区分,下方田地光谱失真由紫色变为灰色。图 7(i)(j)光谱信息保留较好,但均存在区域模糊、边缘信息扭曲的问题。参考 GroundTruth,本文算法结果在补充空间细节信息与保持光谱信息等方面比上述算法有较好的改善。

表 2 给出了 GF-2 卫星图像低分辨率数据集的 10 种对比方法的客观评价指标评估结果,其中最优值和次优值分别用粗体和下划线突出显示。PRACS 算法在传统方法中表现最佳,BT-H 算法、C-MTF-GLP-CBD 算法与 GS 算法各类指标数值相近,三者融合结果客观类似。BDS 算法各项指标相对较差。与其他深度学习方法相比,TFNet 在各项客观

指标上表现较好,说明利用残差网络在全色和多光谱特征域融合可以减少纹理和光谱损失。本文提出的融合算法 MAPNet 在 ERGAS、RMSE、RASE、

QAVE、SSIM 等指标结果最好,SAM、FSIM 指标仅次于 TFNet。综上,MAPNet 在图像的细节纹理补充和光谱信息保留的效果较优于参考算法。

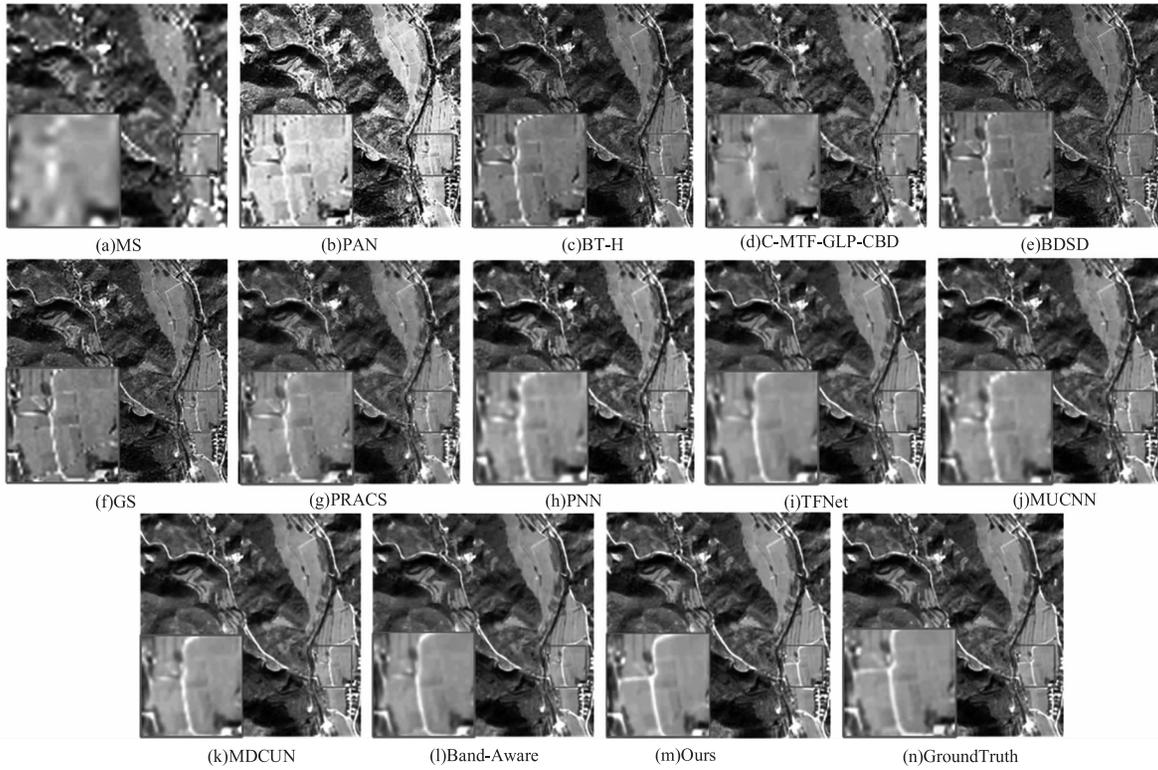


图6 低分辨率数据集融合图像视觉比较(选用红、绿、蓝3个波段显示)

Fig. 6 Visual comparison of fused images from reduced-resolution datasets (select RCB for demonstration)

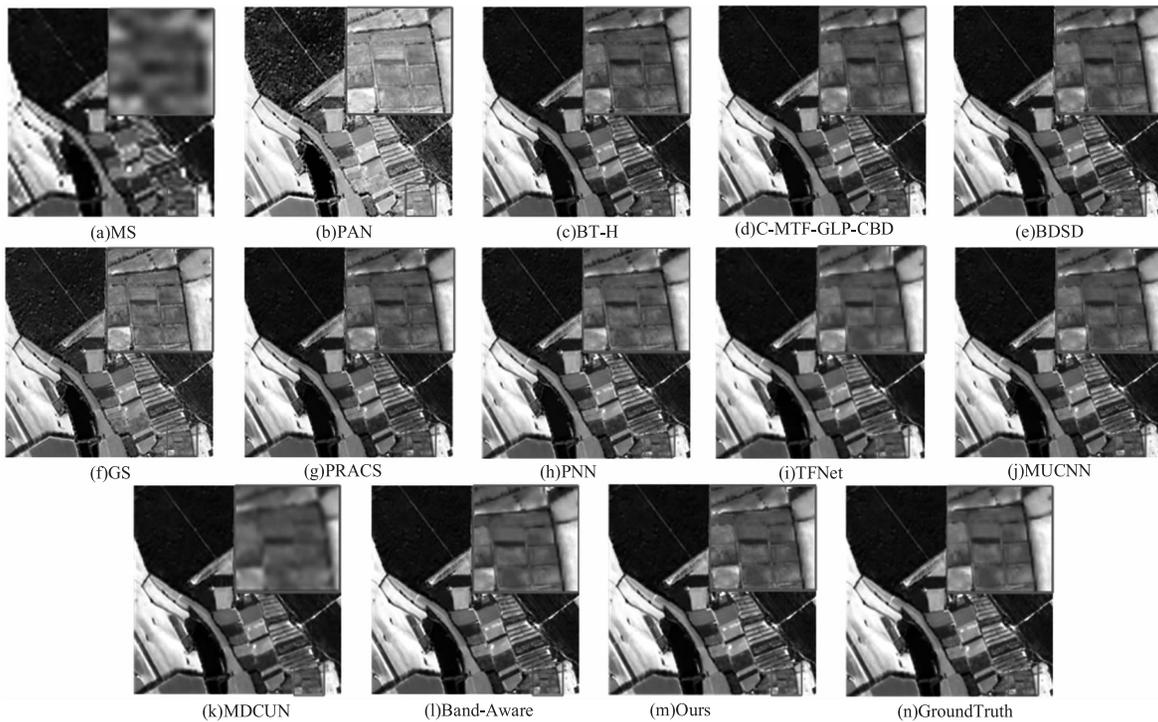


图7 低分辨率数据集融合图像视觉比较(选用红、绿、蓝3个波段显示)

Fig. 7 Visual comparison of fused images from reduced-resolution datasets (select RGB for demonstration)

表2 低分辨率数据集不同方法的融合图像定量比较

Tab.2 Quantitative comparison of fused results of different methods in reduced-resolution datasets

方法	ERGAS ↓	RMSE ↓	RASE ↓	SAM ↓	QAVE ↑	SSIM ↑	FSIM ↑
BT-H	4.61 ± 0.60	4.98 ± 0.81	18.69 ± 8.93	2.48 ± 0.04	0.9237	0.8431	0.916
C-MTF-GLP-CBD	4.42 ± 0.44	4.56 ± 0.47	17.12 ± 5.27	2.84 ± 0.13	0.9324	0.8531	0.9154
BDSB	5.35 ± 0.40	5.59 ± 0.68	20.93 ± 5.19	3.45 ± 0.14	0.9114	0.7993	0.8917
GS	4.83 ± 0.49	4.96 ± 0.68	18.57 ± 5.98	2.74 ± 0.09	0.9372	0.8356	0.9092
PRACS	3.60 ± 0.29	3.76 ± 0.47	14.07 ± 3.98	2.47 ± 0.03	0.9421	0.8935	0.9382
PNN	2.74 ± 0.29	3.18 ± 0.96	10.37 ± 4.05	2.81 ± 0.13	0.9315	0.9422	0.9524
TFNet	<u>1.62 ± 0.04</u>	<u>1.86 ± 0.14</u>	<u>6.11 ± 0.56</u>	1.81 ± 0.03	<u>0.9673</u>	<u>0.9779</u>	0.9828
MUCNN	2.13 ± 0.10	2.46 ± 0.31	8.07 ± 1.332	2.28 ± 0.07	0.948	0.9626	0.9703
MDCUN	1.83 ± 0.09	2.10 ± 0.26	6.86 ± 1.12	2.00 ± 0.05	0.9583	0.9734	0.9784
Band-Aware	1.69 ± 0.05	1.94 ± 0.16	6.34 ± 0.67	1.86 ± 0.04	0.9646	0.9768	0.9816
Ours	1.53 ± 0.02	1.54 ± 0.02	5.78 ± 0.24	<u>1.84 ± 0.03</u>	0.9744	0.9813	<u>0.9825</u>

4.5.2 全分辨率图像实验结果与分析

观察图8(c)~(g)的融合结果,发现传统方法存在光谱畸变、不同程度模糊等问题。PNN与MDCUN方法放大区域中房屋颜色信息损失较为严重,但道路边缘信息保留较好,如图8(h)、(k)。此外,其他DL的方法(如TFNet、MUCNN和Band-Aware)光谱信息虽保存较好,但边缘细节模糊,部分区域产

生伪影。相比较来说,DL方法融合效果略优于传统方法。而本文所提出方法无论是光谱信息还是频谱信息都保存较好,山丘的沟壑、树林、道路的边界信息被清晰地保留下来。综上所述,与其他方法相比,本文所提方法保留了PAN图像的空间结构,空间细节补充效果最好,并可以保留MS图像的光谱信息,避免光谱失真。

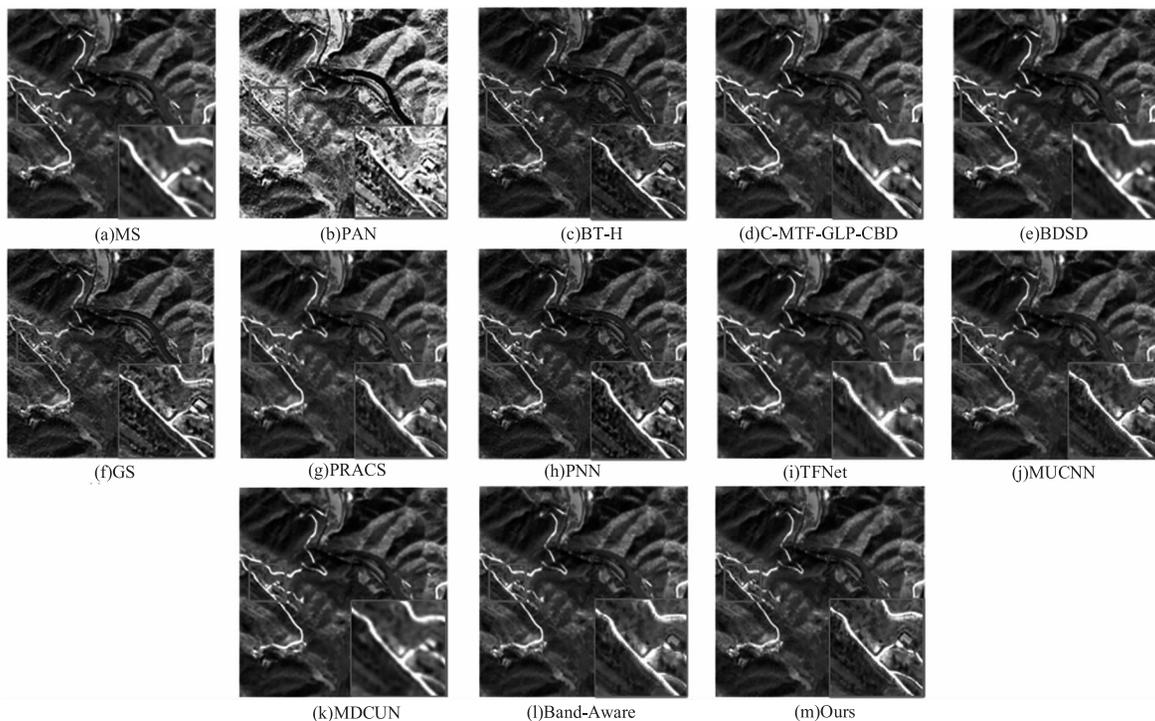


图8 全分辨率数据集融合图像视觉比较(选用红、绿、蓝3个波段显示)

Fig.8 Visual comparison of fused images from full-resolution datasets (select RGB for demonstration)

表3展示了全分辨率数据集下融合结果的定量结果。本文方法在QNR和 D_s 两个指标表现最佳,

D_A 表现次佳,在一定程度上证明本文方法的有效性。

表3 全分辨率数据集不同方法的
融合图像定量比较

Tab.3 Quantitative comparison of fused results
of different methods in full-resolution datasets

方法	QNR \uparrow	$D_\lambda \downarrow$	$D_s \downarrow$
BT-H	0.791	0.0357	0.1799
C-MTF-GLP-CBD	0.9212	0.0296	0.0509
BDS	0.9192	0.0284	0.0539
GS	0.7392	0.0526	0.2195
PRACS	0.8246	0.0312	0.1488
PNN	0.8048	0.0583	0.1453
TFNet	<u>0.9326</u>	0.0231	0.0388
MUCNN	0.9028	0.0356	0.0639
MDCUN	0.9243	0.0551	0.2430
Band-Aware	0.9322	0.0306	<u>0.038</u>
Ours	0.933	<u>0.0266</u>	0.0346

5 结论

本文利用联合注意力机制的渐进式全色锐化网络即 MAPNet,通过联合注意力模块提高了图像关键信息的相关性和完整度,在低分辨率阶段和全分辨率阶段构建全色锐化结果与多光谱图像的一致性约束,减小了尺寸变化带来的光谱和细节损失。实验结果表明,本文提出的全色锐化算法获取的融合图像弥补了 Wald 协议造成的信息损失,提取了高相关性特征,保留了丰富的光谱信息和空间细节信息,具有较好的视觉效果和客观指标。后续可利用不同的卫星数据开展进一步的验证工作。其次,需要设计出更优的融合网络,在优化融合性能的同时提高网络的运行效率。

参考文献:

- [1] Wang Lin, Li Xun, Bao Yunxuan, et al. Research progress of remote sensing application on transportation meteorological disasters[J]. Remote Sensing for Land and Resources, 2018, 30(4): 1-7. (in Chinese)
王琳,李迅,包云轩,等.遥感技术在交通气象灾害监测中的应用进展[J].国土资源遥感,2018,30(4):1-7.
- [2] Xū Jia, Guan Zequn, He Xiufeng, et al. Novel method for merging panchromatic and multi-spectral images based on sensor spectral response[J]. Journal of Remote Sensing, 2009, 13(1): 97-102. (in Chinese)
徐佳,关泽群,何秀凤,等.基于传感器光谱特性的全

色与多光谱图像融合[J].遥感学报,2009,13(1): 97-102.

- [3] Vivone G, Dalla Mura M, Garzelli A, et al. A new benchmark based on recent advances in multispectral pansharpening: revisiting pansharpening with classical and emerging pansharpening methods[J]. IEEE Geoscience and Remote Sensing Magazine, 2020, 9(1): 53-81.
- [4] Otazu X, Gonzalez-Audicana M, Fors O, et al. Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods[J]. IEEE Transactions on Geoscience and Remote Sensing, 2005, 43(10): 2376-2385.
- [5] Ballester C, Caselles V, Igual L, et al. A variational model for P+XS image fusion[J]. International Journal of Computer Vision, 2006, 69(1): 43-58.
- [6] Masi G, Cozzolino D, Verdoliva L, et al. Pansharpening by convolutional neural networks[J]. Remote Sensing, 2016, 8(7): 594-615.
- [7] Wald L, Ranchin T, Mangolini M. Fusion of satellite images of different spatial resolutions: assessing the quality of resulting images[J]. Photogrammetric Engineering and Remote Sensing, 1997, 63(6): 691-699.
- [8] Liu X, Liu Q, Wang Y. Remote sensing image fusion based on two-stream fusion network[J]. Information Fusion, 2020, 55: 1-15.
- [9] WANG Y, DENG L J, ZHANG T J, et al. SSconv: explicit spectral-to-spatial convolution for pansharpening[C]// Proceedings of the 29th ACM International Conference on Multimedia. 2021: 4472-4480.
- [10] Zhou M, Yan K, Fu X, et al. PAN-guided band-aware multi-spectral feature enhancement for pan-sharpening[J]. IEEE Transactions on Computational Imaging, 2023, 9: 238-249.
- [11] Zhou C, Zhang J, Liu J, et al. Percep pan: towards unsupervised pan-sharpening based on perceptual loss[J]. Remote Sensing, 2020, 12(14): 2318.
- [12] Zhang H, Wang H, Tian X, et al. P2Sharpen: a progressive pansharpening network with deep spectral transformation[J]. Information Fusion, 2023, 91: 103-122.
- [13] Meng X, Wang N, Shao F, et al. Vision transformer for pansharpening[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-11.
- [14] Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]// Proceedings of the European Conference on Computer Vision

- (ECCV). 2018;286 – 301.
- [15] Chai T, Draxler R R. Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature [J]. *Geoscientific Model Development*, 2014, 7(3):1247 – 1250.
- [16] Choi M. A new intensity-hue-saturation fusion approach to image fusion with a tradeoff parameter [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2006, 44(6):1672 – 1682.
- [17] Wald L. Data fusion: definitions and architectures: fusion of images of different spatial resolutions [M]. *Presses des MINES*, 2002.
- [18] Alparone L, Wald L, Chanussot J, et al. Comparison of pansharpening algorithms: outcome of the 2006 GRS-S data-fusion contest [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2007, 45(10):3012 – 3021.
- [19] Pushparaj J, Hegde A V. Evaluation of pan-sharpening methods for spatial and spectral quality [J]. *Applied Geomatics*, 2017, 9:1 – 12.
- [20] 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.
- [21] Zhang L, Zhang L, Mou X, et al. FSIM: a feature similarity index for image quality assessment [J]. *IEEE Transactions on Image Processing*, 2011, 20(8):2378 – 2386.
- [22] Vivone G, Alparone L, Chanussot J, et al. A critical comparison among pansharpening algorithms [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 53(5):2565 – 2586.
- [23] Choi J, Yu K, Kim Y. A new adaptive component-substitution-based satellite image fusion by using partial replacement [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2010, 49(1):295 – 309.
- [24] Garzelli A. Pansharpening of multispectral images based on nonlocal parameter optimization [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 53(4):2096 – 2107.
- [25] 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(3):343 – 365.
- [26] Restaino R, Dalla Mura M, Vivone G, et al. Context-adaptive pansharpening based on image segmentation [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 55(2):753 – 766.
- [27] Yang G, Zhou M, Yan K, et al. Memory-augmented deep conditional unfolding network for pan-sharpening [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022:1788 – 1797.