

结合双非局部注意力感知的 SAR 和光学图像金字塔细节融合网络^①



朱佳佳^{1,2}, 杨学志^{1,2}, 梁宏博^{1,2}, 杨翔宇^{1,2}

¹(合肥工业大学 计算机与信息学院, 合肥 230601)

²(工业安全与应急技术安徽省重点实验室, 合肥 230601)

通信作者: 杨学志, E-mail: xzyang@hfut.edu.cn

摘要: 合成孔径雷达 (synthetic aperture radar, SAR) 与光学图像融合旨在利用卫星传感器的成像互补性, 生成更全面的地貌信息. 然而, 由于各单一卫星传感器数据分布的异质性和成像物理机制的差异, 现有网络模型在融合过程中往往存在成像精度低的问题. 为了解决上述问题, 本文提出 DNAP-Fusion, 一种新的结合双非局部注意力感知的 SAR 和光学图像金字塔细节融合网络 (dual non-local-aware-based pyramid fusion net). 该方法利用双非局部注意力模块, 在空间尺度逐渐减小的多级图像金字塔中提取 SAR 图像的结构信息和光学图像的纹理细节. 然后在空间和通道维度上融合它们的互补特征. 然后, 通过图像重构将融合特征注入上采样光学图像中, 得到最终的融合结果. 此外, 在网络训练之前, 采用图像封装决策来增强同一场景中 SAR 和光学图像中目标之间的共性关系. 定性和定量的实验结果表明, 提出的方法优于现有融合方法, 其中客观评价指标中的相关系数 (correlation coefficient, CC) 为 0.990 6, 峰值信噪比 (peak signal to noise ratio, PSNR) 为 32.156 0 dB. 此外, 所提方法有效地融合了 SAR 图像和可见光图像的互补特征, 为提高遥感图像融合的精度和有效性提供了一种有价值的思路和方法.

关键词: 合成孔径雷达图像; 光学图像; 图像融合; 双非局部注意力; 深度学习

引用格式: 朱佳佳, 杨学志, 梁宏博, 杨翔宇. 结合双非局部注意力感知的 SAR 和光学图像金字塔细节融合网络. 计算机系统应用, 2024, 33(8): 155-165. <http://www.c-s-a.org.cn/1003-3254/9521.html>

Pyramid Detail Fusion Network for SAR and Optical Image Based on Dual Non-local Attention Perception

ZHU Jia-Jia^{1,2}, YANG Xue-Zhi^{1,2}, LIANG Hong-Bo^{1,2}, YANG Xiang-Yu^{1,2}

¹(School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230601, China)

²(Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei 230601, China)

Abstract: Synthetic aperture radar (SAR) and optical image fusion aim to leverage the imaging complementarity of satellite sensors for generating more comprehensive geomorphological information. However, existing network models often exhibit low imaging accuracy during the fusion process due to the heterogeneity in data distribution of each single satellite sensor and differences in imaging physical mechanisms. This study proposes the DNAP-Fusion, a novel SAR and optical image fusion network that incorporates dual non-local attention perception. The proposed method utilizes a dual non-local perceptual attention module to extract structural information from SAR images and texture details from optical images within a multi-level image pyramid with a gradually decreasing spatial scale. It then fuses their complementary features in both spatial and channel dimensions. Subsequently, the fused features are injected into the upsampled optical image through image reconstruction, resulting in the final fusion outcome. Additionally, before network training, image

① 基金项目: 安徽省重点研究与开发计划 (202004a07020030); 中央高校基本科研业务费专项资金 (JZ2021HGTD0111); 安徽省自然科学基金 (2108085MF 233)

收稿时间: 2023-12-15; 修改时间: 2024-01-17; 采用时间: 2024-01-23; csa 在线出版时间: 2024-06-28

CNKI 网络首发时间: 2024-07-01

encapsulation decisions are employed to enhance the commonality between objects in SAR and optical images within the same scene. Qualitative and quantitative experimental results demonstrate that the proposed method outperforms state-of-the-art (SOTA) multisensor fusion methods. The correlation coefficient (CC) in the objective evaluation indices is 0.990 6, and the peak signal to noise ratio (PSNR) is 32.156 0 dB. Moreover, the proposed method effectively fuses the complementary features of SAR and optical images, offering a valuable idea and method for enhancing the accuracy and effectiveness of remote sensing image fusion.

Key words: synthetic aperture radar (SAR) image; optical image; image fusion; dual non-local attention; deep learning

1 引言

合成孔径雷达 (synthetic aperture radar, SAR) 成像系统在全天时和全天候条件下能够提供丰富的空间结构信息和高空间分辨率图像, 在土地利用、船舶探测和灾害监测等领域具有独特的优势. 然而, 随着地球观测技术的开展, 单一数据源的遥感场景解译已经无法满足当前的需求. 为了提供更全面的空间信息和更丰富的地物内容, 多源协同遥感技术成为解决地物目标缺失问题的有效手段. 因此, 结合 SAR 和光学图像的融合技术可以克服单个图像的局限性, 充分利用两者互补的特性, 在目标检测、图像分割以及灾害评估等领域具有重要的意义.

针对光学和 SAR 图像融合的问题, 研究者提出了多种不同的融合方法, 目前主要借鉴多光谱 (multi-spectral, MS) 图像和全色 (panchromatic, PAN) 图像融合的思想, 即泛锐化融合方法. 传统图像融合方法可归纳为以下 3 类: 多分辨率分析法^[1-4]、基于稀疏表示的方法^[5]、混合方法^[6,7]. 1) 多分辨率分析法 (multi-resolution analysis, MRA) 通常利用小波 (wavelet) 变换、拉普拉斯金字塔等从 SAR 图像中提取空间结构信息. 常用方法包括稀疏表示^[1,2]、Shearlet 变换法和非下采样轮廓波变换 (non subsampled contourlet transform, NSCT)^[3,4]. 2) 基于稀疏表示的方法^[5]一般从多个重叠的图块中训练出一个完备字典, 然后根据字典提出相应的稀疏分解和重构生成结果. 3) 混合办法综合了成分替换法 (component substitution, CS) 和 MRA 的优势, 以获得更好的融合性能, 例如, Shah 等人^[6]提出将主成分分析 (principal component analysis, PCA) 与 NSCT 相结合, NSCT 在保留空间细节信息的同时, 与 PCA 相结合可以有效地保持图像的光谱信息. 这些方法在像素级进行图像融合, 采用线性模型模拟融合过程, 存在一定的局限性, 难以在提高空间质量和保持光谱质量之间达

到最佳平衡. 因此, 为了克服这些限制, 基于深度学习的方法也被提出并得到广泛应用.

近年来, 在遥感图像融合领域中深度学习 (deep learning, DL) 的作用愈发重要. 卷积神经网络 (convolutional neural network, CNN) 凭借其强大的特征提取和重构能力^[8-11], 在图像融合领域得到了广泛应用. 2016 年, Masi 等人^[8]受到卷积神经网络在单幅图像超分辨率方面的启发, 提出了一个具有 3 层 CNN 结构的开创性方法. 此后, Yang 等人^[9]提出了双分支 CNN 融合架构, 通过输入多源图像来学习细节信息, 并通过低频内容和原始图像生成高通域, 从而改善融合图像的边缘和细节, Liu 等人^[10]利用两流浅深卷积网络从 PAN 图像中提取中层和高层空间特征的空间细节, Jin 等人^[12]引入了拉普拉斯金字塔网络结构, 将全色和多光谱图像进行多尺度分解, 并将金字塔分解的结果逐步集成到 CNN 中, 吴俊华等人^[13]在 PanNet 的基础上, 针对颜色损失和结构畸变问题提出了双分支网络结构. 值得一提的是, 当前 DL 技术已经主导了 PAN 与 MS、SAR 与 MS 图像融合的研究, 但针对 SAR 和光学图像融合的研究相对较少. 在 SAR 和光学图像的融合方面, Ye 等人^[14]采用无监督的结构-纹理分解策略进行特征融合, Zhang 等人^[15]先分别提取光学图像和 SAR 图像的纹理特征, 然后将其进行分层叠加融合, 以提高图像融合的提取精度. 然而, SAR 和光学图像融合的关键挑战在于如何有效地整合两者的互补特征, 以获得全面的特征表达, 而不是只专注于单一成分的增强或改进; 此外, 这些方法通常依赖卷积运算来提取局部感知域的特征, 这可能导致在融合图像中丢失一些重要的全局上下文信息. 另外, 这些方法通常仅关注原始尺度的处理, 未能充分利用更细粒度的网络结构.

综上所述, 针对 SAR 图像和光学图像的互补特征, 本文提出了一种结合双非局部注意力感知的 SAR 与

光学图像融合方法. 该方法结合多级金字塔结构, 并引入 SAR 和光学图像封装决策, 有效保留不同尺度下的显著特征并提取丰富的结构和纹理信息. 之后引入双非局部注意力模型作为融合策略, 将基于 SAR 的结构信息与基于光学的纹理信息进行融合, 建立通道和空间位置上的长程依赖关系, 使网络更加关注图像中的典型目标和细节信息, 从而显著提升了 SAR 和光学图像融合的效果. 本文主要贡献总结如下.

(1) 基于多级金字塔结构设计空间上下文感知模块, 从每个尺度中提取相应的 SAR 图像空间结构高频信息和光学图像纹理信息高频信息, 并跨尺度融合所有的尺度特征.

(2) 引入新的图像协同封装决策以加强 SAR 强度与光学图像各个波段之间的关联关系, 进一步增强特

征耦合度.

(3) 提出了一种双非局部感知的融合策略, 实现对通道和空间位置上的有效互补特征融合与重构. 由 Sentinel-1 与 Landsat-8 获取的 SAR 与光学图像进行一系列实验表明所提方法具有一定的优越性.

2 研究方法

2.1 整体网络结构

本文受 PanNet^[9]和 LPPN^[12]启发, 设计遵循编解码结构的结合双非局部注意力感知的 SAR 和光学图像融合网络 (SAR and optical image fusion network with dual nonlocal attention perception, DNAP-Fusion), 该网络由空间上下文感知编码器、双非局部注意力特征融合模块和融合重建解码器组成, 如图 1 所示.

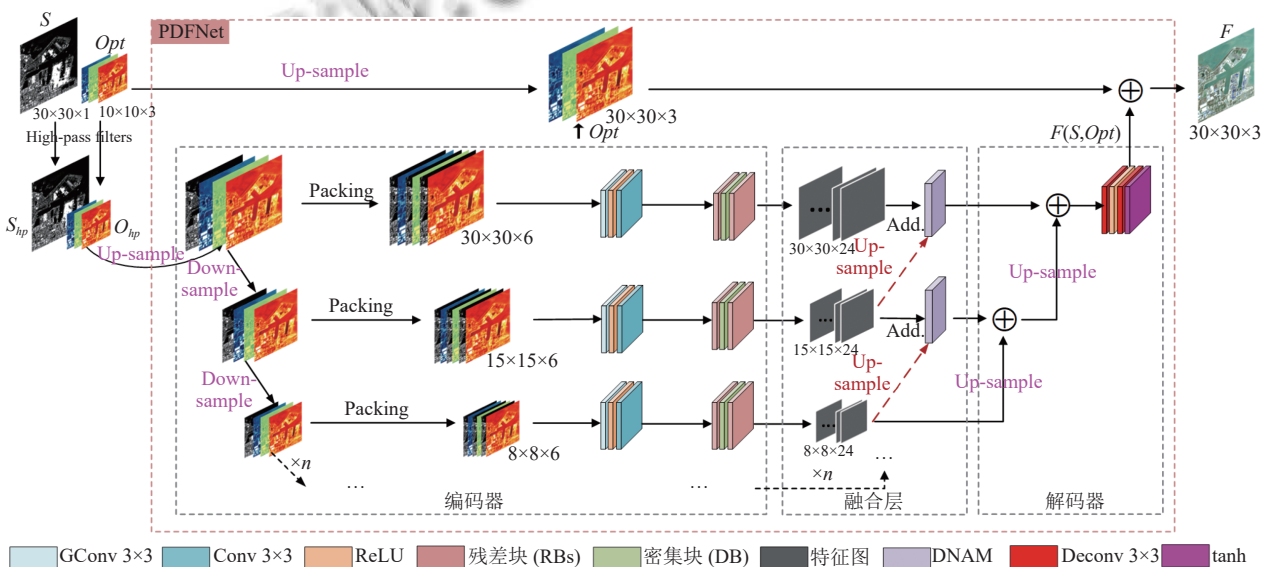


图 1 整体网络结构

空间上下文感知编码器模块主要由分组卷积、整流线性单元 (ReLU) 激活和残差-密集-残差块组成, 它具备提取更深层次语义信息的能力. 用于融合多尺度信息的双非局部注意力特征融合模块 (DNAM) 是 DNAP-Fusion 的关键部分, 其使用两个 DNAM 块, 在不同尺度上融合来自 SAR 图像的结构信息和光学图像的纹理特征 (即编码器的输出). 最后在解码器中合并和重构图像信息, 生成最终融合图像 F . 此外, 为了增强 SAR 图像强度信息与光学图像每个光谱波段之间的关联性, 我们采用了图像封装决策来建立它们之间的映射关系.

2.2 空间上下文感知特征提取模块

2.2.1 封装决策

在不同感受野下, 基于 CNN 的方法所提取的空间上下文存在差异性^[15]. 在此启发下, 本文提出了一种在图像域中的方法, 该方法利用显著的卷积归纳偏置参数从多层金字塔中充分提取 SAR 的结构信息和光学图像的纹理信息. 与传统的深度学习方法不同, 本文采用了一种新的输入格式, 当设计处理每个图像金字塔的子网络时, 将 SAR 图像和光学图像的一个波段封装在一个处理单元中, 并沿着通道维度排列所有这些单元来构造子网络的输入 (式 (1)).

$$\hat{S} = [[S; Opt_1]; [S; Opt_2]; \dots; [S; Opt_i]], i \in [1, d] \quad (1)$$

其中, S 表示来自 SAR 图像的强度值, Opt_i 表示来自第 i 波段的辐射亮度值, d 代表波段, $[\cdot]$ 代表沿通道方向做每个光谱波段的特征合并。

这种输入格式的设计是考虑到不同光谱波段描述地物特性时存在差异, 以实现每个光谱波段的相对独立处理, 可以更好地利用 SAR 和光学图像之间的互补信息。

2.2.2 基于封装决策的空间上下文感知

给定输入 SAR 图像 S 和对应光学图像 Opt , 通过高通滤波器得到 S_{hp} 和 O_{hp} , 为了使光学图像的空间分辨率与 SAR 图像的空间分辨率保持一致, 以便更准确地将 SAR 图像的纹理特征与光学图像的光谱信息相结合, 首先将光学图像的高频信息 (O_{hp}) 经过 3 倍上采样得到 \tilde{O} , 使其与 SAR 图像 (S_{hp}) 具有相同的尺寸。然后, 我们构建一个具有 P 级 ($P = 1, \dots, n$) 的图像金字塔 (P 的最优值在第 2.3.2 节中讨论), 对于每个金字塔层, 将空间尺寸减小 2^P 因子的输入 $[S^{(P)}, \tilde{O}_1^{(P)}, \dots, \tilde{O}_3^{(P)}]$, 通过将 $S^{(P)}$ 重复 3 次得到 $[S^{(P)}; \tilde{O}_1^{(P)}, \dots, S^{(P)}; \tilde{O}_3^{(P)}]$, 以便提取更多特定的空间细节信息。

为了增强 SAR 和光学图像之间的局部联系, 我们

提出了特征聚合块, 该块由带有 B 滤波器组的深度卷积 (即分组卷积)、标准卷积 (具有 32 个 3×3 大小的滤波器, 没有滤波器组) 和线性激活函数 (ReLU) 组成。特征聚合块的公式形式为:

$$Z'_l = BN(\sigma\{ConvDepthwise(Z_{l-1})\}) + Z_{l-1} \quad (2)$$

$$Z_{l+1} = BN(\sigma\{Conv(Z'_l)\}) \quad (3)$$

其中, $ConvDepthwise(\cdot)$ 表示深度卷积 (即分组卷积); $Conv(\cdot)$ 表示标准的 3×3 卷积; $\sigma(\cdot)$ 和 $BN(\cdot)$ 表示每个卷积后面都有一个激活和一个激活后的批归一化 (Batch-Norm)。

提取的特征被传递到残差-密集-残差块 (简称为 RGDB 块) 中, 以提供图像边缘信息^[16]。如图 2 所示, RGDB 块由连续的 3 个部分组成: 第 1 部分和第 3 部分是由两个残差块 (RBs) 组成的残差组, 称为 RG; 第 2 部分是密集块 (dense block), 称为 DB。每个残差块由标准卷积 (具有 32 个 3×3 大小的滤波器, 没有滤波器组) 和线性激活函数 (ReLU) 组成。DB 由 3 个 3×3 卷积层、ReLU 和 1 个 1×1 的卷积层组成, 通过在前层和后层之间创建并行连接来进行跨层信息交换。RB 和 DB 具有强大的特征提取能力, 可以进一步利用两类图像中可用的信息, 提取更具代表性的特征。

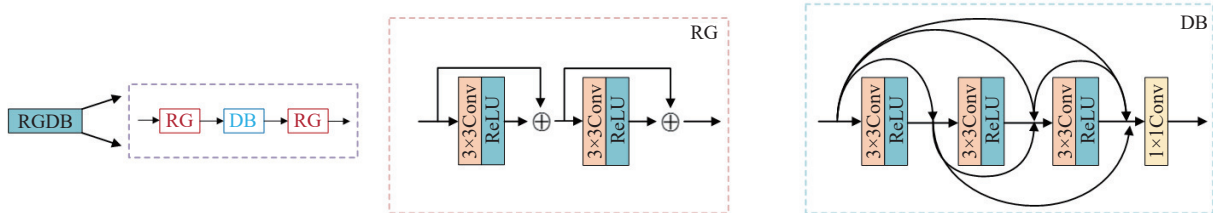


图2 RGDB 块及其组成部分的体系结构

2.3 双非局部注意力特征融合模块

受非局部神经网络^[17]的启发, 设计关于非局部的空间和通道注意力模块作为融合策略。双非局部注意力模型的相应细节如图 3 所示, 该模块利用跨尺度融合, 将高层与低层特征分别连接到特征融合模块的两个输入端, 用于将每个金字塔层提取的特性互补信息进行融合并获得全局上下文信息^[18]。

关于多尺度特征图 Φ_1 和 Φ_2 , 采纳双非局部注意力模型来得到各自的空间和通道注意力图。然后通过加权融合策略得到双重融合的注意力图, 分别称为 Φ_F^S 和 Φ_F^C 。最后, 我们提出了一个融合规则来生成最终的融

合注意力图 Φ_F , 其定义如下:

$$\Phi_F(i, j) = \frac{(\Phi_F^S(i, j) + \Phi_F^C(i, j))}{\sigma} \quad (4)$$

其中, i, j 代表图像特征图的坐标, i 和 j 的取值在图像特征图长、宽的范围; σ 表示参与计算的注意力特征图数量。

2.3.1 空间非局部注意力模型

如图 3 所示 (右上方箭头所指模块), 对于特征图 $\Phi \in R^{C \times H \times W}$, 首先将其转置为 $\Phi^X \in R^{H \times W \times C}$, 然后使用 8×8 池化操作, 并对其重塑生成两个特征图 $\Phi^Y, \Phi^Z \in R^{C \times HW/n^2}$, 然后对 X 和 Y 进行矩阵乘法, 并用 Softmax

来计算空间注意力图 $S^S \in R^{HW \times HW/n^2}$, 公式如下:

$$S^S(x, y) = \frac{E(\Phi^X(x, c) \times \Phi^Y(x, c))}{\sum_{y=1}^{HW/n^2} E(\Phi^X(x, c) \times \Phi^Y(x, c))} \quad (5)$$

其中, $S^S(x, y)$ 表示第 x 个位置对第 y 个位置的影响, $x \in [1, HW], y \in [1, HW/n^2], c \in [1, C]$. 然后对 S^S 和 Φ^Z 进行矩阵乘法, 并对其进行重塑, 得到空间注意力图 $\Phi^S \in R^{C \times H \times W}$, 由式 (6) 表示:

$$\Phi^S = R\left(\sum_{y=1}^{HW/n^2} S^S(x, y) \times (\Phi^Z(c, y))^T\right) + \Phi \quad (6)$$

其中, $R(\cdot)$ 和 $(\cdot)^T$ 分别代表重塑和转置. 在此基础上, 将多尺度特征图 Φ_1 和 Φ_2 输入到空间非局部注意力模型, 得到空间注意力图 Φ_1^S 和 Φ_2^S , 其中, 加权空间注意力系数由以下方法计算得出:

$$\alpha^S(i, j) = \frac{\Phi_1^S(i, j)}{\Phi_1^S(i, j) + \Phi_2^S(i, j)} \quad (7)$$

$$\beta^S(i, j) = \frac{\Phi_2^S(i, j)}{\Phi_1^S(i, j) + \Phi_2^S(i, j)} \quad (8)$$

最后, 融合后的空间注意力图 $\Phi_F^S(i, j)$ 通过以下公式计算:

$$\Phi_F^S(i, j) = \alpha^S(i, j) \times \Phi_1(i, j) + \beta^S(i, j) \times \Phi_2(i, j) \quad (9)$$

2.3.2 通道非局部注意力模型

如图 3 所示 (右下方箭头所指模块), 首先将特征图 $\Phi \in R^{C \times H \times W}$ 重塑为 $\Phi \in R^{C \times HW}$, 然后对其进行矩阵乘法和转置, 最后使用 Softmax 计算得到通道注意力图 $S^C \in R^{C \times C}$, 其定义如下:

$$S^C(c, c) = \frac{E(\Phi(c, x) \times (\Phi(c, x))^T)}{\sum_{c=1}^C E(\Phi(c, x) \times (\Phi(c, x))^T)} \quad (10)$$

其中, $S^C(c, c)$ 表示第 c 通道对其他 c 通道的影响. 之后, 对 S^C 和 Φ 进行矩阵乘法, 以获得通道注意力图 $\Phi^C \in R^{C \times H \times W}$, 其公式如下:

$$\Phi^C = R\left(\sum_{c=1}^C S^C(c, c) \times \Phi(c, x)\right) + \Phi \quad (11)$$

多尺度特征图 Φ_1 和 Φ_2 的通道注意力图通过通道注意力模型获得, 称为 Φ_1^C 和 Φ_2^C , 其中, 加权信道注意力系数由以下方法计算得出:

$$\alpha^C(i, j) = \frac{\Phi_1^C(i, j)}{\Phi_1^C(i, j) + \Phi_2^C(i, j)} \quad (12)$$

$$\beta^C(i, j) = \frac{\Phi_2^C(i, j)}{\Phi_1^C(i, j) + \Phi_2^C(i, j)} \quad (13)$$

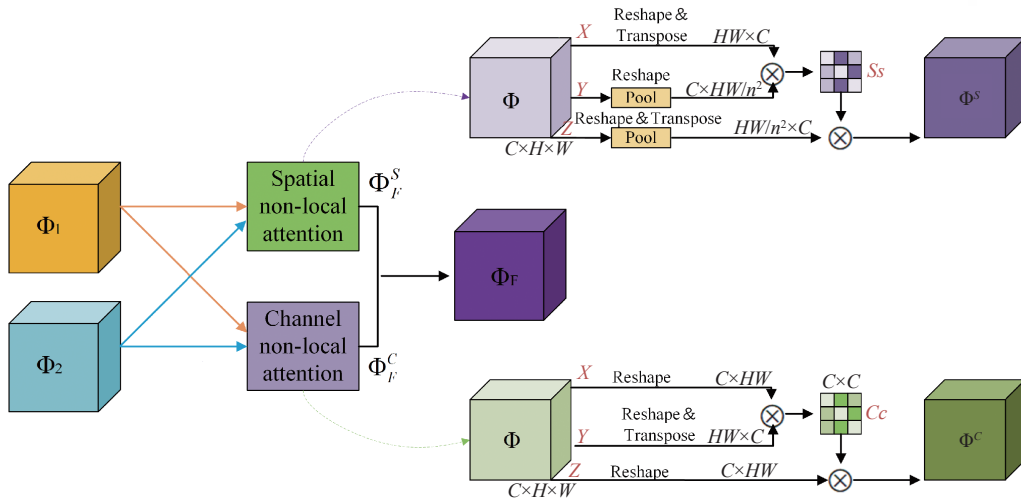


图 3 双非局部注意力模型的融合过程

最后, 融合后的通道注意力图 $\Phi_F^C(i, j)$ 通过以下公式计算:

$$\phi_F^C(i, j) = \alpha^C(i, j) \times \phi_1(i, j) + \beta^C(i, j) \times \phi_2(i, j) \quad (14)$$

2.4 解码器模块

解码器网络由两个与 ReLU 激活层相连的卷积层

组成, 用于重构融合结果. 考虑到输入被归一化到 $[0, 1]$ 的范围, 本文采用 tanh 激活层来确保输出结果保持在相同的空间范围内. 参考文献 PanNet^[9]和细节注入深度神经网络^[19]的方法, 我们采用了从输入到输出的跳跃连接方式, 将融合的特征与上采样后的可见光图像

进行叠加, 以实现可见光图像的光谱信息 ($\uparrow Opt$) 和重构图像的结构纹理信息 $F(S, Opt)$ 同时注入融合图像 F 中 (式 (15)).

$$F = \uparrow Opt \oplus F(S, Opt) \quad (15)$$

2.5 损失函数

参考双分支嵌套融合网络^[20], 我们网络架构使采用的损失函数 L_{total} (式 (16)) 由光谱损失函数 $L_{spectral}$ (式 (17)) 和细节损失函数 L_{detail} (式 (18)) 组成.

$L_{spectral}$ 为融合结果 F 和参照图像 GT 的 L2 范数, L_{detail} 为重构图像的结构纹理信息 $F(S, Opt)$ 和 SAR 图像信息 S_{hp} 之间的 L2 范数.

$$L_{total} = L_{spectral} + \lambda L_{detail} \quad (16)$$

$$L_{spectral} = \frac{1}{N} \sum_{i=1}^N \|F^{(i)} - GT^{(i)}\|_2^2 \quad (17)$$

$$L_{detail} = \frac{1}{N} \sum_{i=1}^N \|F_{hp}^{(i)} - S_{hp}^{(i)}\|_2^2 \quad (18)$$

其中, 参数 λ 为了平衡光谱损失函数和细节损失函数; N 为每批训练图像对的数量, $GT^{(i)}$ 为第 i 个参照的原始可见光图像; $S_{hp}^{(i)}$ 表示第 i 个 SAR 图像的高频信息; $F^{(i)}$ 为第 i 对最终融合后的图像; $F_{hp}^{(i)}$ 表示第 i 对融合重

构图像的结构和纹理信息.

3 实验结果与分析

3.1 实验区域及数据

为避免单一地物对实验结果的影响, 实验选取通州湾和南通长江流域的场景作为分析数据集 (图 4、图 5). 第 1 组数据是南通市通州湾图像, 包含道路、草地、水体等多种类型的地物; 第 2 组数据为南通芦泾港口及临近长江流域图像, 包括大量房屋、森林、船舶等地物. 由于在实际应用中缺乏理想的标准, 因此添加 Sentinel-1 原始光学图像数据 (空间分辨率为 10 m) 作为参考图像.

本文使用 Sentinel-1 和 Landsat-8 卫星获取的 SAR 和光学图像作为实验数据集. Sentinel-1 卫星提供空间分辨率为 10 m 的 VV/VH 强度 SAR 图像, Landsat-8 卫星提供空间分辨率为 30 m 的光学图像. 本文提取区域大小为 900×900 的 SAR 图像的 VV 通道和区域大小为 300×300 的蓝、绿、红 3 个波段光学图像进行配对观测. 本文按照 Wald 协议^[21]对图像进行预处理: SAR 图像预处理包括轨道矫正、辐射定标、相干斑滤波和地形校正, 主要使用 SNAP 进行处理. 光学图像预处理包括辐射定标和大气校正, 主要使用 ENVI 进行处理. 然后将两者进行图像配准.



图 4 通州湾实验数据



图 5 南通市长江流域实验数据

3.2 实验设置

实验采用 Intel(R) Xeon(R) CPU E5-2620 v4 @ 2.10 GHz 处理器, 显卡为 Quadro M2000, CUDA 版本为 8.0, CUDNN 版本为 6.0.21. 本文网络结构主要在 TensorFlow 和 Python 3.6 的框架下搭建, 学习率、动量衰减因子、批大小和总迭代次数分别设置为 0.000 1、0.99、100 和 30 000. 使用 Adam 优化器在 200 个周期内对所有网络进行优化. 我们将配对数据集随机裁剪为 10 000 组图像用于模型训练, 并将其分割为 8 000/2 000 用于训练/验证, 大小分别为 90×90 和 30×30 , 未裁剪的整幅图像作为输入进行测试. 在数据预处理阶段, 经过 3 倍上采样得到 SAR 和光学图像的图像对, 然后将 GT ($30 \times 30 \times 3$)、SAR ($30 \times 30 \times 1$)、Opt ($10 \times 10 \times 3$) 这 3 组数据作为网络训练的输入, 网络训练时长约为 4.2 h.

3.3 消融实验

3.3.1 参数 λ 的消融实验

在本文中, 针对融合网络的损失函数设计参数进行了详细评估, 以确定最优参数值式 (16). 我们选择了一系列参数值 (例如 $\lambda=0.1$ 、0.5、1、5、10、100), 并分析了它们在训练和验证阶段的损失函数随着迭代次数增加的变化情况 (见图 6、图 7). 从实验结果可以看出, 随着 λ 的减小, 总损失也减小, 这表明参数的选择对网络性能具有重要影响. 在进一步的分析中, 图像光谱损失在 $\lambda=0.5$ 、10 和 100 时取得的数值较小, 空间细节损失值在 $\lambda=0.5$ 和 5 时较小. 鉴于这一发现, 我们意识到需要在光谱损失和空间细节损失之间进行平衡, 以确保融合性能的最佳表现. 因此, 我们选择了参数值为 0.5, 这样可以更好地兼顾两个训练阶段的光谱与细节损失.

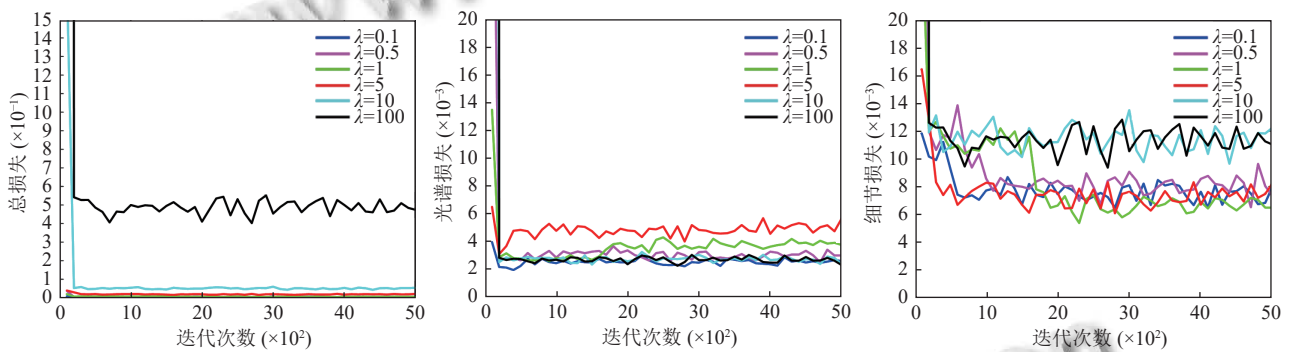


图 6 训练阶段损失函数曲线

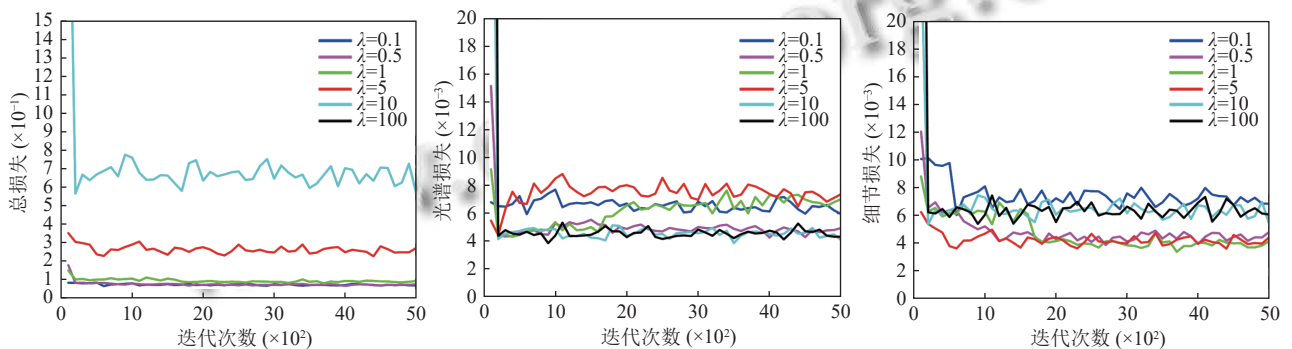


图 7 验证阶段损失函数曲线

3.3.2 金字塔层数 P 的消融实验

在其他参数保持不变的情况下, 我们选择最优 P 值以平衡整体性能和计算量. P 值从 1 变化到 3, 并对第 1 组南通湾数据集进行了实验. 针对光谱信息的有效提取, 我们使用了 4 种定量性能指标来评价融合结果的光谱质量, 其中 CC 代表相关系数, 用于评估融

合图像与原始图像之间的相似性; QAVE 代表平均波段通用图像质量指数, 用来评价多波段图像质量的指数; SAM 代表光谱角度映射器, 用来度量两个光谱向量之间的角度, 反映了它们在光谱空间中的相似性; ERGAS 代表相对平均光谱误差, 用于评估图像整体光谱性能的指标. SAM 和 ERGAS 的理想值为 0, 而

QAVE 和 CC 的理想值为 1. 每个测试的平均运行时间和定量性能等指标列于表 1 中, 加粗字体为每列最优值. 在可承受的网络参数数量下, 3 个金字塔层产生了令人满意的性能. 我们可以合理地推测, 逐步增加金字塔层数到 4 层或更高, 可以使更多的信息参与训练过程, 从而实现更精确的图像融合. 然而, 考虑到计算资源数量和时间的增加等其他因素, 我们需要权衡选择.

表 1 不同金字塔层数对南通湾测试集的消融结果

P	CC↑	QAVE↑	SAM↓	ERGAS↓	Time (s)
1	0.9851	0.7254	0.3698	1.5657	0.2123
2	0.9899	0.7421	0.2103	1.0021	0.9985
3	0.9906	0.7445	0.0879	0.9471	2.7182

3.3.3 融合网络的消融实验

在融合层中, 提出了双非局部注意力模型来建模局部特征的长程依赖性, 为验证双非局部注意力特征融合模块在网络结构中的作用, 我们比较了 3 种验证模型 (表 2), 即 3 种融合策略: 加法、只保留空间非局部注意力和只保留通道非局部注意力. 为了全面评估融合图像在空间和光谱质量上的整体效果, 我们选择了 6 种评估指标来验证, 其中熵 (EN) 表示图像的信息量, 互信息 (MI) 度量了融合后图像与原始图像之间的一致性, 峰值信噪比 (PSNR) 用于量化图像质量, 多尺度结构相似性指数度量 (MS-SSIM) 是结构相似性指数 (SSIM) 的多尺度版本, 用于评估图像在不同尺度上的结构相似性. 除 ERGAS 外, 其他评估指标越高, 表明融合性能越好, 粗体表示最优值, 下划线表示次优值. 在空间非局部注意力和通道非局部注意力方面, 通道得到了 EN 和 MI 的最优值, 说明通道在保持亮度方面具有较强的建模能力, 但不可避免地减弱了一些有用的空间信息和细节信息, 空间注意力模型倾向于保留丰富的纹理细节, 可以弥补通道注意力模型的不足. 从多指标评价的角度来看, 本文算法的融合性能优于其他 3 种融合策略, 说明双非局部注意力模型是有效的, 使用双非局部注意力特征融合模块能融合图像更多的多尺度深度特征, 提高融合性能.

表 2 不同特征融合层的评价指标

融合层	EN↑	MI↑	CC↑	PSNR (dB)↑	MS-SSIM↑	ERGAS↓
Addition	7.1246	3.5331	0.9801	<u>32.7862</u>	0.9299	<u>0.9321</u>
Channel	7.3152	3.5479	<u>0.9822</u>	31.5621	0.8448	0.9384
Spatial	7.1594	3.5394	0.9796	33.0125	<u>0.9116</u>	0.9514
Ours	<u>7.2277</u>	<u>3.5465</u>	0.9906	32.1560	0.8873	0.9471

3.4 不同融合方法比较

为验证本文算法在融合 SAR 和光学图像互补特性上的有效性, 将本文算法与传统算法 IHS、NSCT^[4]、IHS_NSST_SR^[22]、Wavelet、深度学习算法 RSIFNN^[23]、PanNet^[9]、RFNNet^[24]、SOSTF^[14]及双分支卷积神经网络 (TCNN)^[13]进行比较, 对比算法的参数全部按照原始论文中的参数设置.

3.4.1 定性评价

图 8 和图 9 显示了不同算法的融合结果, 红色框部分突出显示了融合图像的一些关键区域, 将红框中的区域进行上采样, 放到图像的左下角 (图 8) 和右上角 (图 9), 以便进行更明显的比较. 与其他方法相比, 图 8(j) 本文方法实现了细粒度级目标增强, 并充分提取了局部和非局部的特征信息, 得到了最清晰的轮廓、最丰富的纹理和最少的颜色误差. 特别是在有桥梁的图像中, 蓝色虚线方框表明本文方法能够很好保持一些小目标的形状和结构, 而其他方法容易产生一些边缘失真和纹理模糊的现象. 相比之下, 虽然其他方法在处理结构相对简单的水体和山地植被图像时可以达到一定的性能, 但对于其他类型的图像, 它们都存在不同方面的缺陷. 其中, 图 8(d) Wavelet 算法和图 8(e) RSIFNN 算法的融合结果纹理锯齿状明显且无法生成清晰的边界, 这在红色放大框中表现得很明显. 在图 8(a) IHS 算法和图 8(b) NSCT 算法融合的图像中空间细节明显增强, 但存在严重的颜色失真问题, 特别是在耕地和水体部分. 图 8(f) PanNet、图 8(g) RFNNet 和图 8(h) SOSTF 的融合结果在大多数类型上都表现得比较好, 但在清晰度和细节方面都不如本文方法, 从建筑区域上可以看出细节不够清晰、边缘比较模糊. 最后, 图 8(c) IHS_NSST_SR 和图 8(i) TCNN 在空间细节和光谱保持上都取得了较好的效果, 但在细粒度级目标上仍不理想, 存在地物边缘失真现象. 第 2 组南通市长江流域数据集的融合结果 (图 9) 与上述评论基本一致. 总之, 本文所提出的方法在融合所有类别的图像上都有较好的效果, 并且融合的结果在视觉上最接近真实图像.

3.4.2 定量评价

本文使用 5 种定量性能指标: 相关系数 (CC)、平均波段通用图像质量指数 (QAVE)、峰值信噪比 (PSNR)、结构相似性 (SSIM)、光谱角映射器 (SAM) 和相对平均光谱误差 (ERGAS). SAM 和 ERGAS 的理想值为 0, 而 CC、QAVE、PSNR 和 SSIM 的理想值为 1.

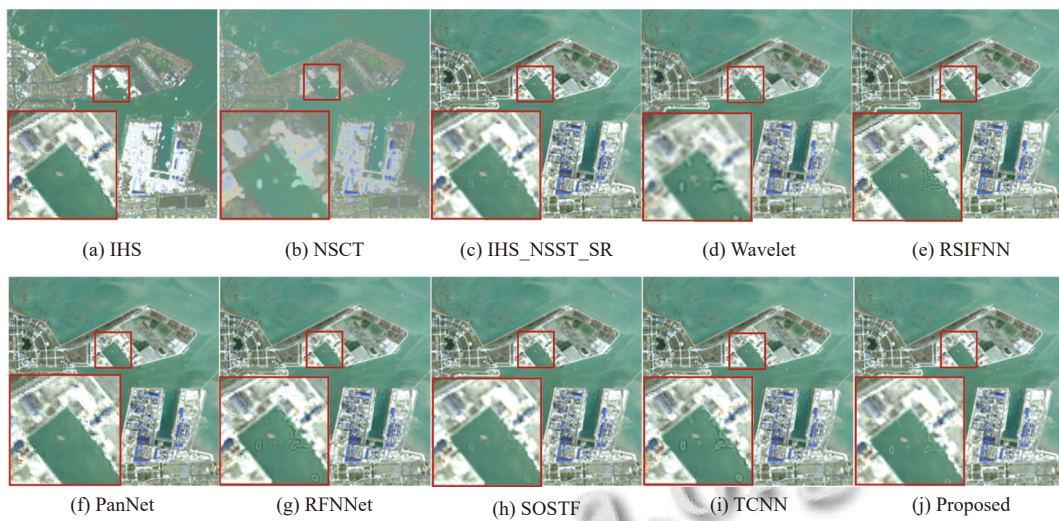


图8 通州湾数据集的不同算法融合结果

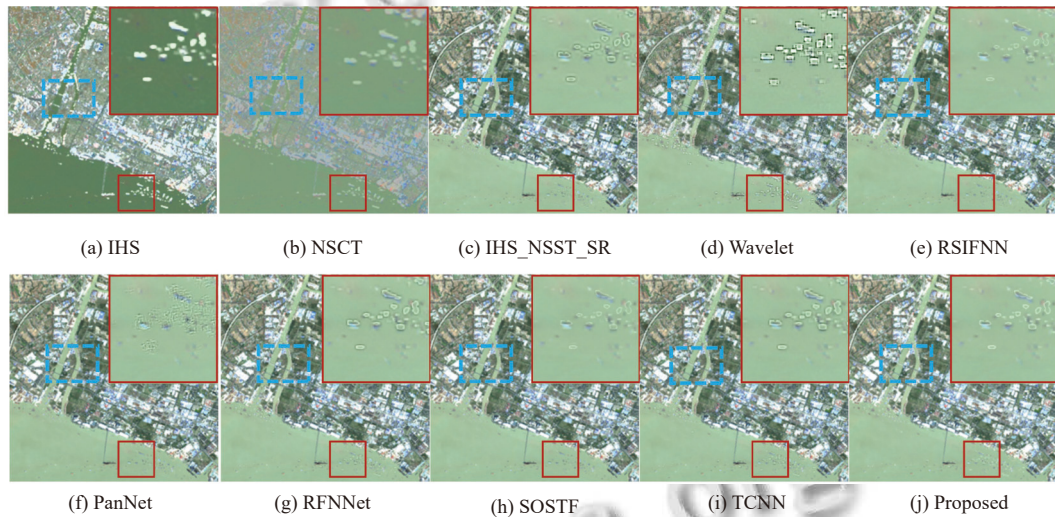


图9 南通市长江流域数据集的不同算法融合结果

表3和表4显示了使用所有测试图像进行定量评估的平均结果。用粗体标记的值代表所有方法中的最佳值,下划线表示次优值。结果表明,本文方法在空间结构和光谱质量上均取得较好结果。就CC、PSNR和ERGAS而言,我们的方法明显优于其他方法。特别是对于PSNR,我们的方法获得了最高的值32.1560 dB,这表明我们融合图像的空间细节质量很好,与参考图像的像素差异最小。从CC和QAVE上看,IHS和NSCT算法结果较差,说明两者在融合SAR和光学图像的过程中会造成光谱失真;实验结果显示,Wavelet算法、RSIFNN算法都过多融入了SAR图像的强度信息,这导致了图像边缘的失真现象;相反地,IHS_NSST_SR算法和双分支算法在保持光谱信息方面表现较好,尽

管如此,它们仍然存在一些需要进一步改进的空间;深度学习方法(TCNN、RFNNet和PanNet)大都取得了比传统方法更高的性能,但仍然不能超过本文提出的方法;本文方法的融合结果光谱质量较好。通过PSNR和SSIM的评估,我们可以看出在IHS和NSCT算法中,融合结果的表现较差,表明存在明显的细节丢失。而在Wavelet算法中,融合结果有所改善,但地物边缘出现了明显的锯齿状现象;IHS_NSST_SR和TCNN都取得较好的结果,融合结果明显增强了SAR图像的结构信息;RSIFNN、PanNet、RFNNet和SOSTF均取得了较好的性能,说明深度学习方法在图像融合中能明显增强空间结构信息;表4最优的SSIM值和表3次优的SSIM值均表明本文融合结果与真实图像具有最

大的结构特征相似度. 从 SAM 和 ERGAS 上看, 本文方法图像融合的整体效果最好, 较低的 SAM 值表明本文融合结果与真实图像的光谱最相似, 属于同类地物的可能性最大; 综合实验结果显示, IHS、NSCT 和 Wavelet 算法在提取 SAR 图像的结构信息和光学图像的纹理信息方面的表现都较为不理想, 与上述主观评

价结果一致; RSIFNN、PanNet 和 RFN-Nest 算法有明显的提升, 解决了传统算法中纹理损失问题; IHS_NSST_SR、TCNN 和 SOSTF 算法有效提升了图像融合的整体效果. 综上所述, 本文算法得到的融合图像更完整地保持原始图像的互补特性, 增强 SAR 图像的结构信息, 同时包含光学图像的大部分纹理信息.

表3 不同算法通州湾数据集融合结果比较

算法	CC↑	QAVE↑	PSNR (dB)↑	SSIM↑	SAM↓	ERGAS↓	Time (s)
IHS	0.2769	0.7104	14.1272	0.9596	1.4125	8.5903	0.0410
NSCT	0.2992	0.7248	15.8234	0.9689	1.0781	7.0499	236.7816
IHS_NSST_SR	<u>0.9891</u>	0.7299	30.9868	0.9901	<u>0.1542</u>	1.4253	121.6978
Wavelet	0.8717	0.6960	21.3440	0.9782	0.9623	3.7541	1.0547
RSIFNN	0.9745	0.7208	28.3314	0.9799	0.3849	1.6705	9.3323
PanNet	0.9819	0.7212	29.7012	0.9836	0.4156	1.4425	3.0707
RFNNet	0.9812	0.7306	29.9687	0.9864	0.2765	1.3870	6.9980
SOSTF	0.9833	0.7277	29.7717	0.9983	0.3248	<u>1.0995</u>	<u>0.7247</u>
TCNN	0.9842	<u>0.7321</u>	<u>31.5219</u>	0.9971	0.2626	1.3020	0.7990
Proposed	0.9906	0.7445	32.1560	<u>0.9981</u>	0.0879	0.9471	2.7182

表4 不同算法南通市长江流域数据集融合结果比较

算法	CC↑	QAVE↑	PSNR (dB)↑	SSIM↑	SAM↓	ERGAS↓	Time (s)
IHS	0.0462	0.6894	12.6452	0.9612	1.7772	9.7507	0.0397
NSCT	0.1294	0.7169	15.2224	0.9678	1.1759	7.2372	240.8844
IHS_NSST_SR	<u>0.9806</u>	0.7221	29.2484	0.9899	<u>0.2089</u>	1.6175	123.2611
Wavelet	0.8054	0.6684	19.0315	0.9675	1.3809	4.6973	1.0547
RSIFNN	0.9669	0.7207	26.9991	0.9736	0.5173	1.8574	8.9370
PanNet	0.9773	0.7189	29.7012	0.9789	0.5549	1.5727	3.0233
RFNNet	0.9821	0.7308	<u>29.9684</u>	0.9796	0.2765	<u>1.3872</u>	7.0173
SOSTF	0.9720	0.7249	27.8952	0.9978	0.4650	1.7085	0.7710
TCNN	0.9747	<u>0.7431</u>	28.2542	0.9893	0.3811	1.4430	0.7122
Proposed	0.9887	0.7437	30.9815	<u>0.9912</u>	0.1205	1.1583	<u>0.6957</u>

4 结论

为了更加有效地融合 SAR 和光学图像的互补特征, 本文提出一种结合双非局部注意力感知的 SAR 和光学图像融合网络, 该网络由空间上下文感知编码器、双非局部注意力特征融合模块以及融合重建解码器组成. 通过与传统算法 IHS、NSCT、IHS_NSST_SR、Wavelet、深度学习算法 RSIFNN、PanNet、RFNNet、TCNN 及 SOSTF 算法进行比较, 实验结果在定性和定量评价上均表现出色, 结论如下.

该方法充分保留了不同尺度下的有效特征, 通过空间上下文感知模块构建多级金字塔结构, 用于提取 SAR 图像的空间结构信息和光学图像的纹理信息, 并在不同尺度之间进行特征融合; 通过图像封装决策, 将 SAR 和光学图像的每个通道进行封装重排以便于进行相对独立的处理, 增强 SAR 图像的结构信息与光学图

像的每个光谱波段之间的共性关系, 进一步获得高精度融合结果; 引入双非局部注意力模块作为融合策略, 有效提取 SAR 图像的结构信息和光学图像的纹理信息, 实现了特征的融合和重建, 极大地提升 SAR 和光学图像融合的有效性. 未来研究将尝试使用其他可替代的注意力模型, 例如 GCNET^[25], 以进一步降低计算复杂度, 或者尝试将 CNN 同 Transformer 相结合, 改进和优化算法以提高融合性能.

参考文献

- 1 金益如, 杨学志, 董张玉, 等. 一种 NSST 与稀疏表示相结合的遥感图像融合算法. 地理与地理信息科学, 2016, 32(2): 60–66. [doi: 10.3969/j.issn.1672-0504.2016.02.012]
- 2 邵建豪, 潘斌, 赵珊珊, 等. 基于 Shearlet 变换的 SAR 与多光谱遥感影像融合. 武汉大学学报 (信息科学版), 2017, 42(4): 468–474.

- 3 童涛, 杨帆, 谭海峰, 等. 基于 NSCT 变换的多传感器图像融合算法. 地理与地理信息科学, 2013, 29(2): 22–25, 59.
- 4 易维, 曾湧, 原征. 基于 NSCT 变换的高分三号 SAR 与光学图像融合. 光学学报, 2018, 38(11): 1110002.
- 5 Zhang Q, Liu Y, Blum RS, *et al.* Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review. *Information Fusion*, 2018, 40: 57–75. [doi: [10.1016/j.inffus.2017.05.006](https://doi.org/10.1016/j.inffus.2017.05.006)]
- 6 Shah VP, Younan NH, King RL. An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets. *IEEE Transactions on Geoscience and Remote Sensing*, 2008, 46(5): 1323–1335. [doi: [10.1109/TGRS.2008.916211](https://doi.org/10.1109/TGRS.2008.916211)]
- 7 Yang Y, Zhang YM, Huang SY, *et al.* Infrared and visible image fusion using visual saliency sparse representation and detail injection model. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 5001715. [doi: [10.1109/TIM.2020.3011766](https://doi.org/10.1109/TIM.2020.3011766)]
- 8 Masi G, Cozzolino D, Verdoliva L, *et al.* Pansharpening by convolutional neural networks. *Remote Sensing*, 2016, 8(7): 594. [doi: [10.3390/rs8070594](https://doi.org/10.3390/rs8070594)]
- 9 Yang JF, Fu XY, Hu YW, *et al.* PanNet: A deep network architecture for pan-sharpening. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 1753–1761. [doi: [10.1109/ICCV.2017.193](https://doi.org/10.1109/ICCV.2017.193)]
- 10 Liu L, Wang J, Zhang EL, *et al.* Shallow-deep convolutional network and spectral-discrimination-based detail injection for multispectral imagery pan-sharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, 13: 1772–1783. [doi: [10.1109/JSTARS.2020.2981695](https://doi.org/10.1109/JSTARS.2020.2981695)]
- 11 Chen YX, Bruzzone L. Self-supervised SAR-optical data fusion of Sentinel-1/-2 images. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 5406011. [doi: [10.1109/TGRS.2021.3128072](https://doi.org/10.1109/TGRS.2021.3128072)]
- 12 Jin C, Deng LJ, Huang TZ, *et al.* Laplacian pyramid networks: A new approach for multispectral pansharpening. *Information Fusion*, 2022, 78: 158–170. [doi: [10.1016/j.inffus.2021.09.002](https://doi.org/10.1016/j.inffus.2021.09.002)]
- 13 吴俊华, 杨学志, 方帅, 等. 基于双分支卷积神经网络的 SAR 与多光谱图像融合实验. 地理与地理信息科学, 2021, 37(2): 22–30. [doi: [10.3969/j.issn.1672-0504.2021.02.004](https://doi.org/10.3969/j.issn.1672-0504.2021.02.004)]
- 14 Ye YX, Liu WC, Zhou L, *et al.* An unsupervised SAR and optical image fusion network based on structure-texture decomposition. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 4028305.
- 15 Zhang YZ, Zhang HS, Lin H. Improving the impervious surface estimation with combined use of optical and SAR remote sensing images. *Remote Sensing of Environment*, 2014, 141: 155–167. [doi: [10.1016/j.rse.2013.10.028](https://doi.org/10.1016/j.rse.2013.10.028)]
- 16 Fang FM, Li JC, Yuan YT, *et al.* Multilevel edge features guided network for image denoising. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(9): 3956–3970. [doi: [10.1109/TNNLS.2020.3016321](https://doi.org/10.1109/TNNLS.2020.3016321)]
- 17 Wang XL, Girshick R, Gupta A, *et al.* Non-local neural networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7794–7803.
- 18 范艺华, 董张玉, 杨学志. 结合上下文编码与特征融合的 SAR 图像分割. 中国图象图形学报, 2022, 27(8): 2527–2536. [doi: [10.11834/j.issn.1006-8961.2022.8.zgtxxb-a2022.08018](https://doi.org/10.11834/j.issn.1006-8961.2022.8.zgtxxb-a2022.08018)]
- 19 Deng LJ, Vivone G, Jin C, *et al.* Detail injection-based deep convolutional neural networks for pansharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(8): 6995–7010. [doi: [10.1109/TGRS.2020.3031366](https://doi.org/10.1109/TGRS.2020.3031366)]
- 20 董张玉, 许道礼, 张晋, 等. 基于双分支多尺度残差融合嵌套的 SAR 和多光谱图像融合架构与实验. 地理与地理信息科学, 2023, 39(1): 23–30. [doi: [10.3969/j.issn.1672-0504.2023.01.004](https://doi.org/10.3969/j.issn.1672-0504.2023.01.004)]
- 21 Wald L, Ranchin T, Mangolini M. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogrammetric Engineering and Remote Sensing*, 1997, 63(6): 691–699.
- 22 盛佳佳, 杨学志, 董张玉, 等. 基于 NSST-IHS 变换稀疏表示的 SAR 与可见光图像融合. 图学学报, 2018, 39(2): 201–208.
- 23 Shao ZF, Cai JJ. Remote sensing image fusion with deep convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, 11(5): 1656–1669. [doi: [10.1109/JSTARS.2018.2805923](https://doi.org/10.1109/JSTARS.2018.2805923)]
- 24 Li H, Wu XJ, Kittler J. RFN-Nest: An end-to-end residual fusion network for infrared and visible images. *Information Fusion*, 2021, 73: 72–86. [doi: [10.1016/j.inffus.2021.02.023](https://doi.org/10.1016/j.inffus.2021.02.023)]
- 25 Cao Y, Xu JR, Lin S, *et al.* Global context networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(6): 6881–6895. [doi: [10.1109/TPAMI.2020.3047209](https://doi.org/10.1109/TPAMI.2020.3047209)]

(校对责编: 孙君艳)