

基于双通道多尺度特征提取和注意力的 SAR 与多光谱图像融合

种法亭^{1,2}, 董张玉^{1,2,3}, 杨学志^{2,3}, 曾庆旺^{1,2}

(1. 合肥工业大学 计算机与信息学院, 安徽 合肥 230031; 2. 工业安全与应急技术安徽省重点实验室, 安徽 合肥 230031;
3. 智能互联系统安徽省实验室, 安徽 合肥 230031)

摘要: 图像融合的根本任务是提取图像特征, 由于合成孔径雷达 (Synthetic Aperture Radar, SAR) 图像和多光谱 (Multi Spectral, MS) 图像存在通道差异, 针对现有算法难以充分提取和利用 SAR 图像的高频细节信息和多光谱图像的低频光谱信息, 融合图像存在细节丢失和光谱失真问题。本文提出了一种基于双通道多尺度特征提取和混合注意力的图像融合算法。首先采用双通道网络提取 SAR 和多光谱图像的多尺度高频细节特征和低频光谱特征, 并连续使用不同空洞率的扩张卷积扩大感受野。然后将提取的特征映射到混合注意力模块中进行特征增强, 再将这些增强特征与上采样的多光谱图像叠加。同时构建了基于光谱角度距离的损失函数, 可以进一步缓解细节丢失和光谱失真。最后通过解码网络重建图像, 得到高分辨率的融合图像。实验结果表明, 本文算法达到了领先水平, 并且融合图像在细节和光谱上保持了较好的平衡。

关键词: SAR 图像融合; 扩张卷积; 多尺度提取; 残差网络; 注意力机制

中图分类号: TP391 文献标识码: A 文章编号: 1001-8891(2024)01-0061-13

SAR and Multispectral Image Fusion Based on Dual-channel Multi-scale Feature Extraction and Attention

CHONG Fating^{1,2}, DONG Zhangyu^{1,2,3}, YANG Xuezhi^{2,3}, ZENG Qingwang^{1,2}

(1. College of Computer and Information, Hefei University of Technology, Hefei 230031, China;
2. Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei 230031, China;
3. Anhui Province Laboratory of Intelligent Interconnection System, Hefei 230031, China)

Abstract: The fundamental task of image fusion is to extract image features. Because of the channel differences between synthetic aperture radar (SAR) images and multispectral (MS) images, existing algorithms have difficulty in fully extracting and utilizing the high-frequency detail information of SAR images and low-frequency spectral information of multispectral images, and the fused images have problems with detail loss and spectral distortion. In this study, an image fusion algorithm based on dual-channel multiscale feature extraction and hybrid attention is proposed. First, a dual-channel network is used to extract multi-scale high-frequency detail features and low-frequency spectral features of SAR and multispectral images, and successively expand the perceptual field using dilated convolution with different void rates. The extracted features are then mapped to the hybrid attention module for feature enhancement, and these enhanced features are superimposed on the upsampled multispectral images. A loss function based on the spectral angular distance was also constructed, which could further alleviate the problems of detail loss and spectral distortion. Finally, the image is reconstructed using a decoding network to obtain a high-resolution fused image. The experimental results show that the proposed algorithm achieves the best performance and that the fused image maintains a good balance of details and spectra.

Key words: SAR image fusion, dilated convolution, multiscale extraction, residual network, attention mechanism

收稿日期: 2023-03-08; 修订日期: 2023-05-18.

作者简介: 种法亭 (1997-), 男, 硕士研究生。主要研究方向: 遥感图像融合。Email: fatingchong@163.com。

通信作者: 董张玉 (1986-), 男, 副教授, 硕士生导师。主要研究方向: 遥感图像处理、数字信号处理。Email: dzyhfut@hfut.edu.cn。

基金项目: 安徽省重点研究与开发计划项目 (202004a07020030); 安徽省自然科学基金项目 (2108085MF233); 中央高校基本科研业务费专项 (JZ2021HGTB0111)。

0 引言

合成孔径雷达是一种微波有源主动成像雷达，具有全天时、全天候、多极化、多角度、地表穿透力强等特点，使得 SAR 卫星广泛地应用于农业、林业、海洋、灾害监测等领域。

SAR 卫星能够穿透云层以及地表覆盖的自然植被，可以详尽地绘制地形地貌，获得高分辨率且细节清晰的地表图像，MS 传感器可以提供地表不同波段的光谱信息，但 MS 图像缺少细节和纹理信息。因此将来自不同传感器的信息进行互补整合，得到融合图像。融合图像可以保留 MS 图像的光谱信息和 SAR 图像的细节信息，同时具有高分辨率。

传统的图像融合方法包含以下 3 类。①主成分替换法：主要包括的方法有（Intensity Hue Saturation, HIS）变换^[1]、主成分分析（Principal Component Analysis, PCA）变换^[2]、施密特变换（Gram Schmid, GS）以及 Brovey 变换^[3]等，该类算法直接融合源图像的像素特征。此类算法简单易实现，但需要较大的计算量，是其它融合算法的基础。融合后的图像包含显著的 SAR 图像细节信息，但由于两种图像的光谱信息和细节信息存在较大差异，融合后的图像会产生较严重的光谱失真；②多尺度变换法：主要包括的方法有拉普拉斯金字塔多尺度变换法^[4]、小波变换法^[5]、非下采样剪切波变换^[6]和非下采样轮廓波变换^[7]。该类算法可以在较好保持光谱特征的同时，加入 SAR 图像的细节信息，它在一定程度上缓解光谱失真，但是会造成较严重空间扭曲；③混合法：首先采用主成分分析法分解图像，得到图像的特征信息，然后采用多尺度分析法融合图像的特征信息，最后采用逆变换得到融合图像。虽然该类算法能够更好保留源图像的细节和光谱信息，但随着场景的复杂度提高，该算法求解过程缓慢，且无法保证精度。

基于深度学习的算法：相较于传统算法，卷积神经网络有着强大的特征提取和表达能力，广泛应用于图像融合领域。例如，Masi 等设计了一种基于深度学习的遥感图像融合算法（CNN-based Pan-sharpening Neural Network, PNN），在融合效果上，相较于传统算法有较大提升，由于两种图像存在通道差异，算法难以在高分辨率全色图像和低分辨率高光谱图像之间建立联系^[8]。Wei 等提出了一种基于残差网络的遥感图像融合算法 DRPNN（Deep Residual Pan-Sharpening Neural Network）^[9]，可以解决训练时出现的梯度消失现象，提升模型的鲁棒性，但采用单一的卷积核，限制了网络的特征提取能力。Yang 等为提取更加丰富的图像特征，提出了基

于 CNN 的双通道图像锐化网络（PanNet）^[10]，使用双通道网络提取多光谱图像和全色图像的光谱特征和细节特征，并采用残差来改善特征连接。但在特征提取时，每个通道的特征被同等对待，不能专注于更有用的特征。吴俊华等^[11]使用的双分支网络结构极大程度保留了 SAR 和 MS 图像的细节特征和光谱特征，但是不能充分利用原始图像的上下文信息，导致感受野受限。Xu 等人设计了一种无监督的图像融合网络 U2Fusion^[12]，可以在无参考图像的情况下建立统一的融合模型，但模型无法分辨聚焦区域和散焦区域。Liu Q. 等提出一种基于注意力机制的遥感图像融合算法 HANet^[13]，注意力能够充分利用特征图的通道和空间之间的关系，增强关键特征并抑制光谱通道和空间位置中干扰特征，但融合图像难以在光谱和细节上实现平衡。许道礼^[14]等提出一种基于双分支多尺度残差融合嵌套的 SAR 和多光谱图像融合算法，该算法使用多尺度残差融合网络，增强多尺度的表达能力，并采用嵌套连接解码器融合不同尺度的特征，但算法未关注不同注意力机制对图像融合结果的影响，融合效果有待进一步提升。

基于以上问题，在前人研究的基础上，本文提出了一种基于双通道多尺度特征提取和注意力的 SAR 与多光谱图像融合算法，网络使用多种不同尺度的卷积提取特征，可以有效构建 SAR 和 MS 图像之间的关系；加入扩张卷积关注全局上下文信息，扩大感受野的同时不增加计算量；使用混合注意力增强网络的特征提取能力，提取隐藏在复杂背景中的 SAR 图像特征，并抑制无关特征；使用 SAD（Spectral Angle Distance, SAD）函数^[15]构建损失，进一步平衡融合图像在光谱和细节之间的关系。

1 原理及方法

1.1 模型建立

SAR 图像和 MS 图像可以分别表示为 $F \in R^{m \times n \times 1}$ ，
 $M \in R^{\frac{m \times n \times b}{c}}$ ， $m \times n$ 表示 SAR 图像的空间分辨率， b 表示 MS 图像的波段数， b 一般取值为 3， c 表示 SAR 图像和 MS 图像的空间分辨率之比，而图像融合就是把低分辨率的 MS 图像和高分辨率的 SAR 图像融合，生成高分辨率的融合图像 $F \in R^{m \times n \times b}$ 。

1.2 网络框架

本文在 PanNet^[10]和双分支^[11]网络架构的基础上，结合多尺度特征提取、扩张卷积、注意力机制和残差网络，提出了一种基于双通道多尺度特征提取和注意力的 SAR 与多光谱图像融合算法，算法使用细节保

持和光谱保持两个支路，在细节和光谱上同时提升。网络遵循“特征提取-特征融合-解码器”结构，并含有跳线连接，算法的网络框架如图1所示。

1.3 双通道模块

双通道模块包含细节保持模块（Detail Preservation Block, DPB）和光谱保持模块（Spectral Preservation Block, SPB）两部分。

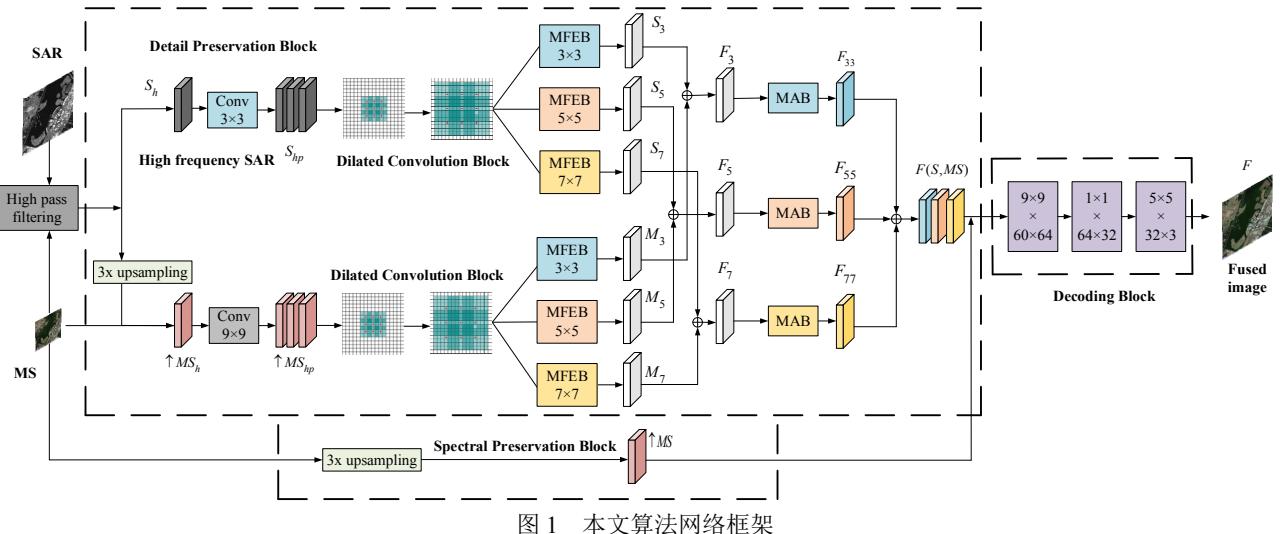


Fig.1 Algorithmic network framework

“较小尺度的特征对较小尺寸的卷积做出反应，较粗糙的结构倾向于被较大尺寸的卷积提取”^[16]，本文采用不同尺寸的卷积核分别提取 SAR 和 MS 图像特征。主要步骤如下：①为了保持原始 SAR 图像和 MS 图像结构，首先将预处理后 SAR 图像和 MS 图像通过高通滤波器提取高频分量。②然后采用小尺度 3×3 卷积核提取细节信息丰富的 SAR 图像，采用较大尺度 9×9 卷积核提取分辨率较低的 MS 图像，得到两种图像浅层特征。③其次将得到的浅层图像特征通过扩张卷积扩大其感受野，再分别使用 3 种不同尺度的 MFEB 模块，提取图像的深层特征，将提取的深层特征分别映射到 3 组混合注意力机制模块。④最后将上一步输出的特征按照同尺度叠加，得到细节保持模块的输出 $F(S, MS)$ 。详细运算流程如下：

1) 浅层特征提取。 S_h 表示原始 SAR 图像通过高通滤波器后的图像， $\uparrow MS_h$ 表示原始 MS 图像通过高通滤波器后并 3 倍上采样后的图像。 S_{hp} 表示用 3×3 的卷积提取的经过高通滤波器后的 SAR 图像的浅层特征， $\uparrow MS_{hp}$ 表示用 9×9 的卷积提取上采样后并通过高通滤波器的 MS 图像的浅层特征。浅层特征提取如公式(1)和(2)所示。

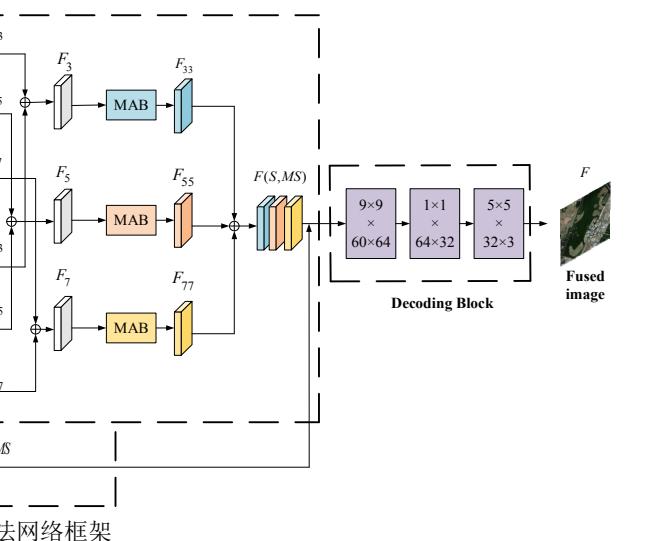
$$S_{hp} = \varphi_{3 \times 3}(S_h) \quad (1)$$

$$\uparrow MS_{hp} = \varphi_{9 \times 9}(\uparrow MS_h) \quad (2)$$

2) 深层特征提取。将浅层特征依次通过扩张卷积

1.3.1 细节保持模块

由图 1 可以看出，细节保持模块由 2 组扩张卷积模块（Dilated Convolution Block, DCB）、6 组多尺度特征提取模块（Multi scale feature extraction block, MFEB）和 3 组混合注意力模块（Mixed Attention Block, MAB）构成。



模块（DCB）和多尺度特征提取模块（MFEB），在 MFEB 中使用大小为 3×3 、 5×5 和 7×7 的 3 种尺度卷积核，提取 SAR 和 MS 图像深层特征，深层特征用 S_i 和 M_i 表示，其中 φ 表示深层特征提取， $i=3,5,7$ 。如公式(3)和(4)所示。

$$S_i = \varphi_{i \times i}(S_{hp}) \quad (3)$$

$$M_i = \varphi_{i \times i}(\uparrow MS_{hp}) \quad (4)$$

图 2 分别展示了 3 种尺度大小为 3×3 、 5×5 和 7×7 的多尺度特征提取模块。



图 2 3 种不同尺度的 MFEB

Fig 2 Three different scales of MFEB

3) 特征重构。将上一步提取的 SAR 图像深层特征 S_i 和 MS 图像深层特征 M_i ，按照同尺度进行像素叠加，此操作可以将 SAR 图像中的高分辨率细节特征融入 MS 图像中， F_i 表示按照同尺度像素叠加的特征，过程如公式(5)所示。然后将 3 组叠加后的特征 F_i 通过注意力机制模块 MAB 进行特征重构，得到重构后的多尺度特征 F_{ii} ， ϕ 代表注意力模块的特征重构过程。如公式(6)所示。

$$F_i = S_i + M_i \quad (5)$$

$$F_{ii} = \phi(F_i) \quad (6)$$

4) 特征融合。最后将 3 种不同尺度的重构特征

F_{ii} 进行通道叠加, 得到最终的细节保持模块输出 $F(S, MS)$ 。过程如公式(7)所示:

$$F(S, MS) = \sum_{i=3,5,7} F_{ii} \quad (7)$$

1.3.2 光谱保持模块

在光谱保持模块中, 采用跳线连接细节保持模块的输出和 3 倍上采样后的 MS 图像, 即 MS 图像的光谱信息传递到融合图像中, 得到了 SAR 与 MS 图像最终的融合结果 F 。光谱特征融合的过程如图 3 所示。

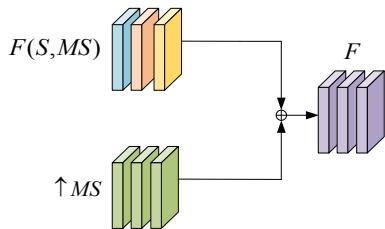


图 3 细节特征与光谱特征叠加流程图

Fig.3 Flow chart of superposition of detailed features and spectral features

图 3 中 $F(S, MS)$ 表示细节保持模块的输出, $\uparrow MS$ 表示按照双三次插值法进行 3 倍上采样后的 MS 图像。

1.4 扩张卷积模块

在对图像提取的过程中, 可以通过增加卷积核的大小和网络深度, 来扩大感受野获取更丰富的上下文信息, 但是以上操作会增加网络参数和网络的复杂度, 不利于网络的训练。扩张卷积可以增加感受野, 能够让卷积核更好地捕捉到输入图像的特征信息, 提高网络对图像的理解能力, 并且扩张卷积不会改变输入和输出的大小, 可以无缝地集成到现有网络架构中^[17]。扩张卷积中含有可以修改的“空洞率”(Dilation Rate, DR), 使用不同的空洞率可以获得不同的感受野, 即在卷积核中插入空洞来实现这种操作。扩张卷积虽然可在不损失特征图尺寸的情况下增大感受野, 但卷积核具有间隔, 会造成卷积中心点不连续, 在叠加相同或成倍数的空洞率的扩张卷积层时尤为严重, 为避免这一问题, 将空洞率分布设为“锯齿状”, 依次为 1、2、3, 如图 4(c)所示, 这样卷积中心的分布便不再有遗漏。图 4 展示了不同分布方式的扩张卷积感受野。

标准卷积可以看作空洞率为 1 的扩张卷积, 其感受野的计算公式如公式(8)和(9)所示。

$$p' = p + (p-1) \times (d-1) \quad (8)$$

$$R^{p+1} = R^p + (p'-1) \times \prod_{p=1}^k S_p \quad (9)$$

式中: p 为扩张卷积的卷积核大小; d 是空洞率, $d-1$ 代表卷积核点的间隔数量; p' 是其等价卷积核大小;

S_p 表示第 p 层卷积的步长(stride); R^{p+1} 表示当前层的感受野; R^p 表示上一层感受野。结合图 4 和公式(8)(9)可知, 扩张卷积可以保持与标准卷积一致的分辨率并在不增加额外参数的情况下扩大感受野。

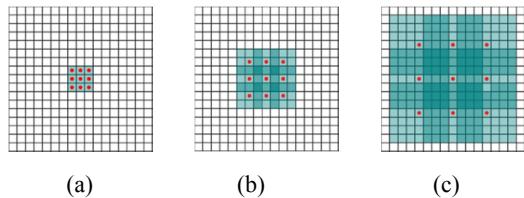


图 4 不同分布方式的扩张卷积感受野对比。(a) 3×3 的标准卷积感受野; (b) 3×3 的扩张卷积感受野 (空洞率=2); (c) 锯齿状分布扩张卷积感受野

Fig.4 Comparison of dilated convolution receptive field with different distribution modes (a)The receptive field of 3×3 standard convolutions; (b) The receptive field of 3×3 dilated convolution (dilation ratio=2); (c) Serrated distribution dilated convolution receptive field

1.5 混合注意力模块 (MAB)

在网络中, 所有输入的特征都被平等地看待, 但实际上一些特征对结果的贡献更加重要, 因此引入注意力机制可以让模型在学习过程中自动地关注那些更重要的特征, 提高模型的精度^[18]。但在构建更深的网络时, 通过逐层的非线性变换提取输入数据的高层次特征, 这种逐层的变换可能会导致梯度在反向传播过程中逐渐变小, 从而使训练过程变得非常困难, 甚至无法收敛。在 ResNet^[19]网络中, 通过引入残差连接, 可以让梯度更容易地在网络中传递, 从而避免梯度消失和梯度爆炸问题, 使训练变得收敛。

基于文献[18]、[19]本文设计一种混合注意力模块(Mixed Attention Block, MAB)。混合注意力模块由残差连接(Residual Network, ResNet)、通道注意力模块(Channel Attention Block, CAB)和空间注意力模块(Spatial Attention Block, SAB)组合而成, MAB 不仅可以在通道层面提高运行效率, 还可以增强目标特征和抑制无关特征, 例如增强 SAR 图像中的纹理信息和 MS 图像中的光谱信息, 抑制 SAR 图像中的噪声信息和 MS 图像中的粗糙纹理。MAB 的结构如图 5 所示。

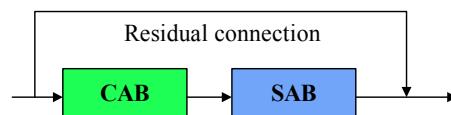


图 5 MAB 结构

Fig.5 The structure of MAB

1.5.1 通道注意力模块 (CAB)

对于传统的通道注意力 SENet 来说, Wang 等认

为捕获所有通道间的依赖关系是低效率且不必要的，并且一维卷积具有良好的跨通道特征获取能力^[20]。设计的通道注意力模型如图6所示。

我们首先将提取的特征做全局平均池化（Global Average Pooling, GAP）处理，对不同通道进行权值重标定，调整通道间的依赖关系，为了降低模型的复杂度并且更好地捕获通道间的特征，相较于传统的SENet，我们加入了卷积核大小为*i*的一维卷积Conv1D，卷积核的大小*i*可以根据输入通道的维度*C*和权重*a*确定，*a*一般取值为2， $|Q|_{\text{odd}}$ 表示最接近*Q*的奇数，对于一维卷积，其大小由公式(10)确定。假设通道注意力的输入 $X = [X_1, X_2 \dots X_C]$ ， X_C 表示一个大小为*H*×*W*，通道为*C*的特征图，通过Sigmoid激活函数对一维卷积的输出做非线性化处理得到加权系数 W_C ，最后将输入 X_C 与加权系数 W_C 相乘得到输出 F_C ，计算过程如公式(11)和(12)所示。

$$i = \left\lceil \frac{\log_2(C)}{a} \right\rceil_{\text{odd}} \quad (10)$$

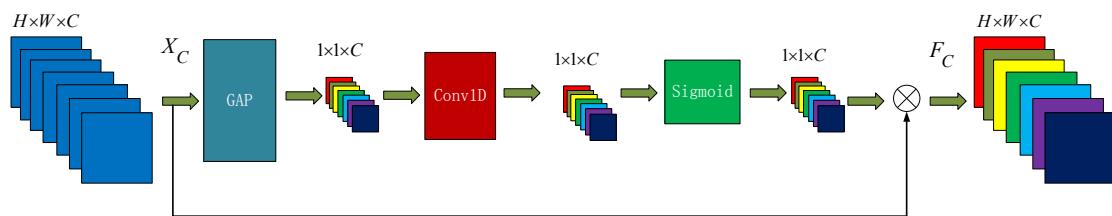


图6 通道注意力模块

Fig.6 Channel attention block

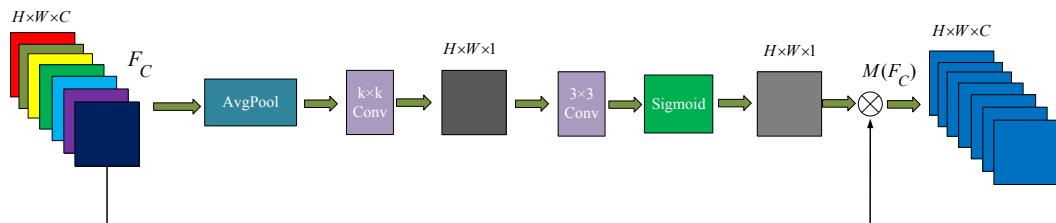


图7 空间注意力模块

Fig.7 Spatial attention block

CAB的输入 F_C ，通过SAB后其大小与尺寸保持不变，即：

$$F_{\text{avg}} = \text{Avgpool}(F_C) \quad (13)$$

$$M(F_C) = \sigma f^{k \times k} (f^{3 \times 3}(F_{\text{avg}})) F_C \quad (14)$$

式中： σ 代表sigmoid函数； $f^{k \times k}$ 和 $f^{3 \times 3}$ 分别表示使用*k*×*k*和3×3的卷积核运算，最后与CAB的输入 F_C 相乘得到最终的输出 $M(F_C)$ 。

1.6 解码模块

在解码模块中，本文参照一种端到端的网络模型

$$W_C = \sigma[\text{Conv1D}(\text{GAP}(X_C))] \quad (11)$$

$$F_C = W_C \times X_C \quad (12)$$

式中：GAP(X_C)表示对 X_C 做全局平均池化处理； σ 表示使用Sigmoid函数激活。

1.5.2 空间注意力模块（SAB）

因为SAR图像经过预处理后仍存在一些噪声和无关特征，我们采用了空间注意力模块SAB让网络更加专注于更多有用信息的特征，例如，SAR图像的高频纹理和边缘轮廓特征，并且SAB可以加强网络对SAR图像噪声的抑制能力，从而获得出色的细节保持效果。受文献[21]启发，为了关注空间信息和减少计算参数，在SAB中，本文采用一个卷积核大小自适应变化的卷积层进行空间信息融合，其中通道压缩比*r*设置为6。卷积核大小*k*随着多尺度卷积的大小变化而变化： $k=0.5(i+1)$ ， $i=3,5,7$ ，并且由于最大池化操作减少了信息的传递，与传统的空间注意力相对比，本文删除了最大池化，更加有利于信息的流动，且结构更轻量。所构建的空间注意力模块如图7所示。

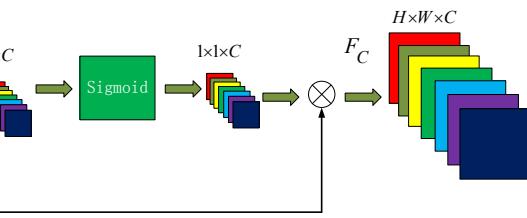


图7 空间注意力模块

Fig.7 Spatial attention block

SRCNN^[22]（Super Resolution Convolutional Neural Network）设计解码模块，将融合后的特征进行图像超分辨率。SRCNN模型由3个大小分别为9×9、1×1和5×5的卷积层组成，每个卷积层都包含许多滤波器，学习低级特征（如边缘和纹理）到高级特征的映射，然后将这些高级特征重建为高分辨率图像。在本文的解码模块中，第一层使用9×9的卷积层，将特征由128通道降至64通道；第二层使用1×1的卷积层，将特征由64通道降至32通道；最后一层使用5×5的卷积层将通道维度降至3维，获得清晰的高分辨率

融合图像。SRCNN的结构如图1中的解码模块所示。

1.7 损失函数

损失函数(Loss Function)是影响图像融合的重要因素，本文采用 L^i 损失函数对网络进行优化， L^i 损失函数由损失函数 L_1^i 和损失函数 L_2^i 构成。 L_1^i 中的 $|\Phi|$ 表示 Φ 的范数，它表示生成的融合图像与参考图像之间的平均绝对误差。 Φ 由公式(15)确定，损失函数 L_1^i 如公式(16)所示：

$$\Phi = \Phi(X_{\text{SAR}}^i, X_{\text{MS}}^i; \Theta) - X_{\text{HMS}}^i \quad (15)$$

$$L_1^i = \frac{1}{N} \sum_{i=1}^N |\Phi| \quad (16)$$

式中： X_{SAR}^i 和 X_{MS}^i 分别表示第*i*张SAR和第*i*张MS图像； X_{HMS}^i 表示第*i*张参考的MS图像(参考图像为Landsat 8MS图像)， Θ 表示网络参数。

光谱角度距离(Spectral Angle Distance, SAD)^[15]是用于计算两个光谱向量之间相似性的指标。在遥感图像处理中，每个像素都可以表示为一个光谱向量，该向量包含不同波段的反射率或辐射值。SAD使用光谱向量之间的夹角来度量它们之间的相似性，两个光谱向量之间的夹角越小，则它们的SAD值越小，表示它们越相似。本文使用光谱角度距离SAD构建损失函数 L_2^i ，假设有两个光谱向量 x 和 y ，它们的长度相等，它们之间的SAD可以计算为它们之间的夹角的反余弦值，即：

$$J(x, y) = \arccos \left(\frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2} \right) \quad (17)$$

式中： $\|x\|_2$ 和 $\|y\|_2$ 分别表示向量 x 和向量 y 的长度。SAD的值通常在 $0\sim 0.5\pi$ 之间，值越小表示两个向量越相似。

损失函数 L_2^i 表示测量MS图像和融合图像之间的差异， X_F^i 表示第*i*张融合图像，损失函数 L_2^i 可由公式(18)表示：

$$L_2^i = \frac{1}{N} \sum_{i=1}^N J(X_{\text{MS}}^i, X_F^i) \quad (18)$$

最终的损失函数 L^i 可定义为 L_1^i 和 λ 倍的 L_2^i 相加，经过实验测试，将 L_2^i 的系数 λ 设为0.1，融合的图像可以在光谱和细节保持较好的平衡。计算过程如公式(19)所示：

$$L^i = L_1^i + \lambda L_2^i \quad (19)$$

2 实验与分析

2.1 实验设置

试验区选取的图像位置为天津市郊区，地物信息

丰富。研究数据包括Sentinel-1W GRD级SAR图像(10m分辨率)和Landsat 8 MS图像(30m分辨率)，分别从阿拉斯加卫星数据中心(Alaska Satellite Facility, ASF)和美国Earth Explorer网站下载。考虑到时序变化问题，两种图像的成像时间差控制在1周以内，Sentinel-1W GRD数据采集时间为2020年4月17日，Landsat 8数据采集时间为2020年4月22日。Sentinel-1 SAR数据选取其中VH极化方式成像的图像，Landsat 8 MS数据选取30m分辨率波段中的2、3、4(蓝、绿、红)波段，将3个波段叠加形成30m分辨率RGB图像。

选取两组地物信息丰富且感兴趣的区域，第一组实验数据包含船只、河道与港口等明显的地物信息。第二组实验数据包含耕地、住宅区、河道与厂房等。两组实验数据如图8和图9所示。



(a)Landsat 8 MS 图像

(b)Sentinel-1B SAR 图像

图8 第1组实验数据

Fig.8 The first set experiment data



(a)Landsat 8 MS 图像

(b)Sentinel-1B SAR 图像

图9 第2组实验数据

Fig.9 The second set experiment data

本文实验环境为64位Windows 10操作系统，搭载RTX 2060显卡，在TensorFlow和Python3.7的框架下搭建的卷积神经网络系统。采用Adam算法自适应优化网络模型。初始学习率设置为0.001，有利于加速收敛过程，每迭代3000次时，学习率下降一半，能够减少反复震荡，提高训练精度，模型全部迭代次数为30000次。

从预处理后的SAR和MS图像裁剪出了12000

组图像用于模型训练，大小分别为 180×180 和 60×60 ，按照9:1的比例，10800组图像作为训练集，1200组图像作为验证集，训练集总大小为4.5GB。最后裁剪2组大小 900×900 的SAR图像和 300×300 的MS图像验证实验。将原始的MS图像设为Ground Truth，然后将裁剪好的SAR图像、MS图像和Ground Truth打包送入训练网络。

2.2 评价指标

在图像融合实验部分，本文选择相关系数(Correlation Coefficient, CC)、峰值信噪比(Peak Signal to Noise Ratio, PSNR)、光谱映射角度(Spectral Angle Mapping, SAM)^[23]、空间相关系数(Spatial Correlation Coefficient, sCC)^[24]、通用图像质量指标(Universal Image Quality Index, UIQI)^[25]这5种评价指标用于评估融合结果。本文还采用了QNR(the Quality with No Reference Index)^[26]以及QNR的光谱分量 D_λ 和空间分量 D_S ，这3种不需要参考图像的客观指标来客观评价融合图像的质量。 $QNR=(1-D_\lambda)^\alpha(1-D_S)^\beta$ ，其中 α 和 β 的取值为1。

2.3 算法对比实验

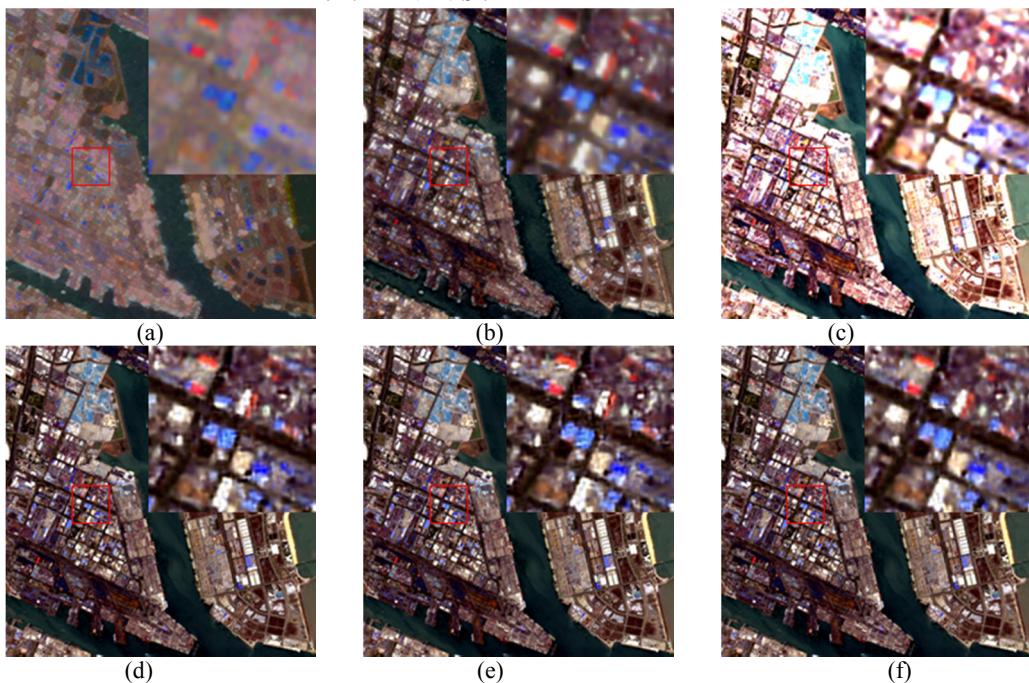
为了验证本算法在图像纹理细节以及光谱保持上的提升，评估网络性能，本文选取了9种对比实验算法。其中包含传统算法：(Principal Component Analysis) PCA^[2]、NSCT-SM-PCNN^[27]，以及基于深度

学习的算法：(Remote Sensing Image Fusion Network) RSIFNN^[28]、(Deep Residual Pan-Sharpening Neural Network) DRPNN^[9]、(Pan-sharpening Network) PanNet^[10]、(Double Branch) 双分支^[11]、(Unified Unsupervised Fusion) U2Fusion^[12]、(Hybrid Attention Network) HANet^[13]和(Double-branch Multiscale Residual-fusion Nested-connections Net) DMRN-Net^[14]。对比实验中的参数设置全部按照原始论文中的参数进行设置。

从主观视觉评价和客观定量评价两个方面来评估分析不同算法的融合效果。主观视觉评价是根据融合图像和参考图像光谱特征和细节特征的相似程度，来判断融合算法的优劣。客观定量评价是根据评价函数，从空间域和光谱域定量评价算法优劣。在实验结果中，主观视觉评价是依据融合后图像的光谱信息与原始的MS图像的光谱信息对比，如果光谱信息越接近，则光谱保持效果越好，细节信息与SAR图像对比，越接近效果越好。

2.3.1 主观视觉评价

在Sentinel-1B和Landsat 8的数据上选取两组数据做实验。实验结果如图10、图11所示。客观评价指标的结果如表1、2、3和4所示，最优的结果已加粗。



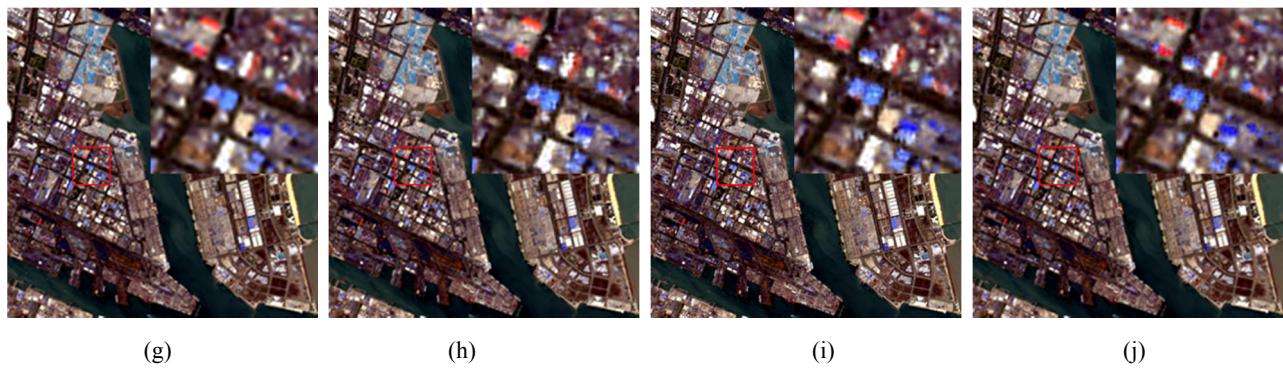


图 10 第 1 组对比实验融合结果。(a) PCA; (b) NSCT_SM_PCNN; (c) RSIFNN; (d) DRPNN; (e) PanNet; (f) 双分支; (g) U2Fusion; (h) HANet; (i) DMRN-Net; (j) 本文算法

Fig.10 Fusion results of the first set of comparative experiments. (a) PCA; (b) NSCT_SM_PCNN; (c) RSIFNN; (d) DRPNN; (e) PanNet; (f) Double branch; (g) U2Fusion; (h) HANet; (i) DMRN-Net; (j) Proposed algorithm

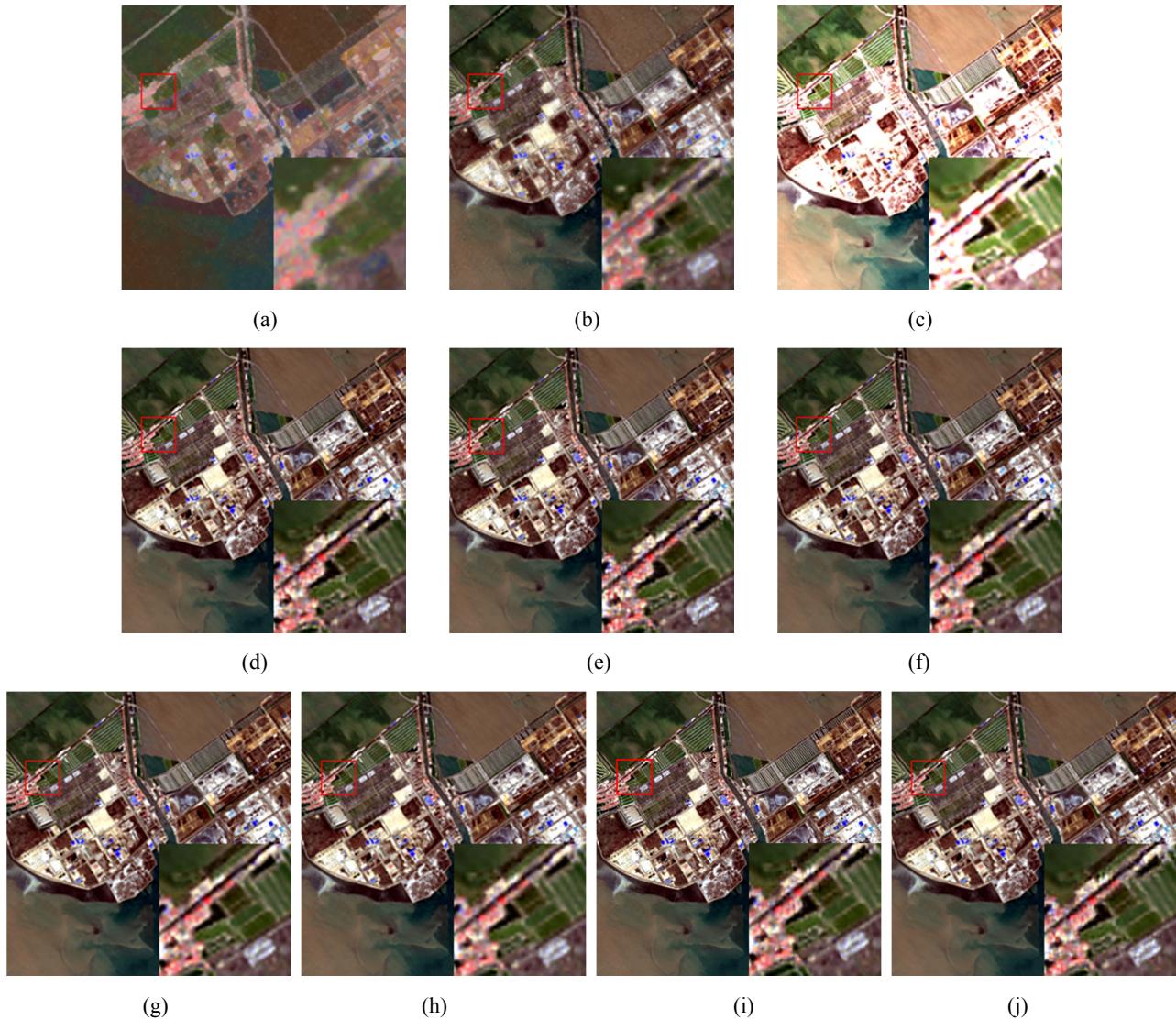


图 11 第 2 组对比实验融合结果。(a) PCA; (b) NSCT_SM_PCNN; (c) RSIFNN; (d) DRPNN; (e) PanNet; (f) 双分支; (g) U2Fusion; (h) HANet; (i) DMRN-Net; (j) 本文算法

Fig.11 Fusion results of the second set of comparative experiments. (a) PCA; (b) NSCT_SM_PCNN; (c) RSIFNN; (d) DRPNN; (e) PanNet; (f) Double branch; (g) U2Fusion; (h) HANet; (i) DMRN-Net (j) Proposed algorithm

从第一组实验结果图 10 来看, 传统算法 PCA 虽然在 MS 图像中加入了 SAR 图像的细节信息, 但是光谱失真现象十分严重, 得到的图像局部变蓝, 并且非常模糊。传统算法 NSCT_SM_PCNN 在光谱上有一一定程度提升, 但与参考图像对比, 空间分辨率较低, 如图 10(a)和(b)中的蓝色房顶, 整体的清晰度较低。从总体来看, 深度学习算法在空间细节提升和光谱保持方面优于传统算法。在放大的区域中, RSIFNN (图 10(c)) 融合算法显著增强了 MS 图像的细节信息, 但由于加入过多的 SAR 图像信息, 出现了十分严重的光谱失真, 在建筑物中体现最为明显。DRPNN、PanNet、双分支和本文算法, 整体上都保持了多光谱图像的光谱特征, 同时也改善了图像的细节信息。但是 DRPNN 算法整体上出现了振铃效应, 在放大部分出现伪影。PanNet 算法融合图像的轮廓和边缘保持较好, 但是在光谱保持方面弱于本文算法。双分支算法光谱保持较好, 但加入的 SAR 图像信息较少, 导致局部细节不明显。U2Fusion 算法在光谱保持方面弱于本文算法, 且图像放大后较为模糊, 源于算法网络不能充分提取图像特征。HANet 算法能够很好地保持光谱特征, 但是在放大区域可以看到红色屋顶粘连到一起, 并且光谱保持能力较差, 不利于后期的图像识别 (如图 10(h))。DMRN-Net 算法 (如图 10(i)) 从局部区域来看, 目标的细节信息没有本文清晰, 但结构和轮廓优于其它对比算法。本文算法融合后的图像既不会因为 SAR 图像信息加入过多, 造成光谱失真, 也不会在边缘出现伪影, 所提出的方法在光谱保持和细节提升方面取得了平衡, 从主观评价来看本文算法优于其它算法。第二组实验图 11 效果与第一组实验大致相同。

2.3.2 客观定量评价

对于参考的评价指标, 当 CC、PSNR、sCC 和 UIQI 较大而 SAM 较小时, 图像融合算法更好。对于非参

考评价指标, 当 D_λ 和 D_S 较小而 QNR 的值较大时, 图像融合的效果更好。从表 1~表 4 可以看出, 本文算法在 CC、PSNR、sCC 和 QNR 取得了最大值, 而在 SAM、 D_λ 和 D_S 取得了最小值, 以上这些指标取得最优值。

从光谱评价指标 CC、UIQI 和 SAM 来看, 传统算法 PCA 表现的融合效果最差, 相关系数仅为 0.4373; RSIFNN 也因为加入了过多的 SAR 图像信息导致光谱扭曲, 与参考图像相关性仅为 87.34%, 低于传统算法 NSCT_SM_PCNN。DRPNN 和 PanNet 融合算法在光谱上有明显的改善; 传统算法 NSCT_SM_PCNN 与 PCA 对比, 各种指标有明显提高, 但仍有提升空间。从空间细节评价指标 PSNR 和 sCC 来看, PCA 和 RSIFNN 算法对比于其它算法表现较差, 是因为这两种融合算法中存在严重的光谱畸变和色彩失真, 这与主观评价结果一致; PanNet 和双分支算法较于以上两种算法在光谱和细节上有明显提升; U2Fusion 和 HANet 在 CC 上相较于其它算法有较大提升, 但是在指标 PSNR 和 SAM 上远不如本文算法, 源于在特征提取上没有将光谱和细节分开。从光谱评价指标 SAM 来看, DMRN-Net 与本文算法有较大差异, 因此在光谱保持上弱于本文算法, 同时从空间细节评价指标 PSNR 来看, 本文算法与 DMRN-Net 相比也有明显提升, 因此本文算法在细节保持上也取得不错效果。本文构建了基于光谱角度距离的损失函数, 因此在评价指标 SAM 上相对于其它算法有较大提升。在表 1 中, 本文算法与最优的对比算法 DMRN-Net 相比, 从指标 CC、PSNR、sCC、UIQI 来看, 分别高了 0.22%、4.76%、0.64% 和 0.87%, 从指标 SAM 来看, 降低了 0.72%, 因此本文算法在 5 种评价指标 CC、PSNR、SAM、sCC、UIQI 上取得最优值, 说明本文算法在光谱保持和细节提升上均取得较好的效果。

表 1 第 1 组不同算法实验数据

Table 1 The first set of experimental data of different algorithms

Algorithms	CC↑	PSNR↑	SAM↓	sCC↑	UIQI↑	Time/s↓
PCA	0.4373	17.2903	8.3840	0.5512	0.6587	0.0643
NSCT_SM_PCNN	0.9318	28.4791	4.0874	0.7985	0.8091	86.7521
RSIFNN	0.8734	20.5217	8.0587	0.7785	0.8127	0.8032
DRPNN	0.9421	28.3581	3.8354	0.7821	0.8548	1.5214
PanNet	0.9721	30.5549	2.1245	0.7743	0.8611	0.7749
双分支	0.9784	29.8364	1.9743	0.8019	0.8546	0.8544
U2Fusion	0.9816	30.8673	1.7374	0.8251	0.8852	0.7965
HANet	0.9894	31.1665	1.6885	0.8378	0.9045	0.8247
DMRN-Net	0.9931	31.7482	1.6401	0.8414	0.9136	0.8472
Proposed	0.9953	33.2601	1.5946	0.8468	0.9244	0.7826

表2 第2组不同算法实验数据

Table 2 The second set of experimental data of different algorithms

Algorithms	CC↑	PSNR↑	SAM↓	sCC↑	UIQI↑	Time/s↓
PCA	0.3679	16.8396	7.8233	0.4315	0.6218	0.0658
NSCT_SM_PCNN	0.9142	27.7532	4.3214	0.7783	0.8324	80.6325
RSIFNN	0.8612	18.3357	7.8521	0.7652	0.8052	0.7935
DRPNN	0.9217	26.2145	3.7412	0.7624	0.8375	1.3254
PanNet	0.9654	30.1842	2.2156	0.7839	0.8501	0.7683
双分支	0.9751	29.2644	2.0546	0.8121	0.8478	0.8774
U2Fusion	0.9807	29.8485	1.7681	0.8335	0.8813	0.8157
HANet	0.9887	30.8674	1.6997	0.8364	0.9082	0.8344
DMRN-Net	0.9916	31.3149	1.6648	0.8374	0.9123	0.8548
Proposed	0.9947	32.8233	1.6073	0.8387	0.9221	0.7978

从算法运行时间来看, PCA 方法最简单, 所用的时间最短, NSCT_SM_PCNN 所用的时间最长。RSIFNN、DRPNN、PanNet、双分支、U2Fusion、HANet 和本文算法采用的是基于深度学习的方法, 使用的测试集相同。本文算法的训练时间为 7 h 左右, 测试时间为 0.78 s, DMRN-Net 算法测试时间为 0.84 s, 该算法采用嵌套连接的融合方式能够实现特征复用和缓解梯度消失, 但由于嵌套连接需要进行通道叠加, 网络需要加深且频繁读取内存, 导致运行效率低于本文算法。本文算法运行时间远小于传统的多尺度分解法的 80.6 s, 更适合处理海量的遥感数据。

综上所述, 本文所提出的图像融合算法在保持光谱特征的基础上, 有效改善了 MS 图像的空间细节, 在视觉上更接近 MS 图像, 并且所用的时间相对较少。

表3 第1组无参考客观实验数据

Table 3 No reference objective experimental data for the first set

Algorithms	D _A ↓	D _S ↓	QNR↑
PCA	0.1587	0.1697	0.6985
NSCT_SM_PCNN	0.1042	0.1127	0.7948
RSIFNN	0.0932	0.1088	0.8081
DRPNN	0.0898	0.0927	0.8258
PanNet	0.0723	0.0764	0.8568
双分支	0.0714	0.0743	0.8596
U2Fusion	0.0756	0.0785	0.8518
HANet	0.0693	0.0681	0.8673
DMRN-Net	0.0583	0.0635	0.8819
Proposed	0.0574	0.0617	0.8844

2.4 消融实验

为了验证本文所构建的模块会对实验产生怎样的影响, 本小节设计了 3 种不同的组合架构来验证算法的有效性, 分别是: (a)加入多尺度特征提取模块和混合注意力模块。(b)加入多尺度特征提取模块、混合注意力模块和扩张卷积模块。(c)加入多尺度特征提取模块、混合注意力模块、扩张卷积模块和损失函数。方法(c)也就是本文所提出的算法, 可以用下列组合表示(a)MFEB+MAB; (b)MFEB+MAB+DC; (c)MFEB+MAB+DC+loss。选择使用上述的两组 SAR 和 MS 图像, 分别按照上述组合架构依次做实验。实验结果如图 12 和图 13 所示。

表4 第2组无参考客观实验数据

Table 4 No reference objective experimental data for the second set

Algorithms	D _A ↓	D _S ↓	QNR↑
PCA	0.1658	0.1754	0.6879
NSCT_SM_PCNN	0.1325	0.1388	0.7471
RSIFNN	0.0974	0.1243	0.7904
DRPNN	0.0951	0.0934	0.8204
PanNet	0.0758	0.0803	0.8400
双分支	0.0743	0.0785	0.8530
U2Fusion	0.0769	0.0797	0.8495
HANet	0.0708	0.0757	0.8589
DMRN-Net	0.0621	0.0686	0.8736
Proposed	0.0594	0.0621	0.8822

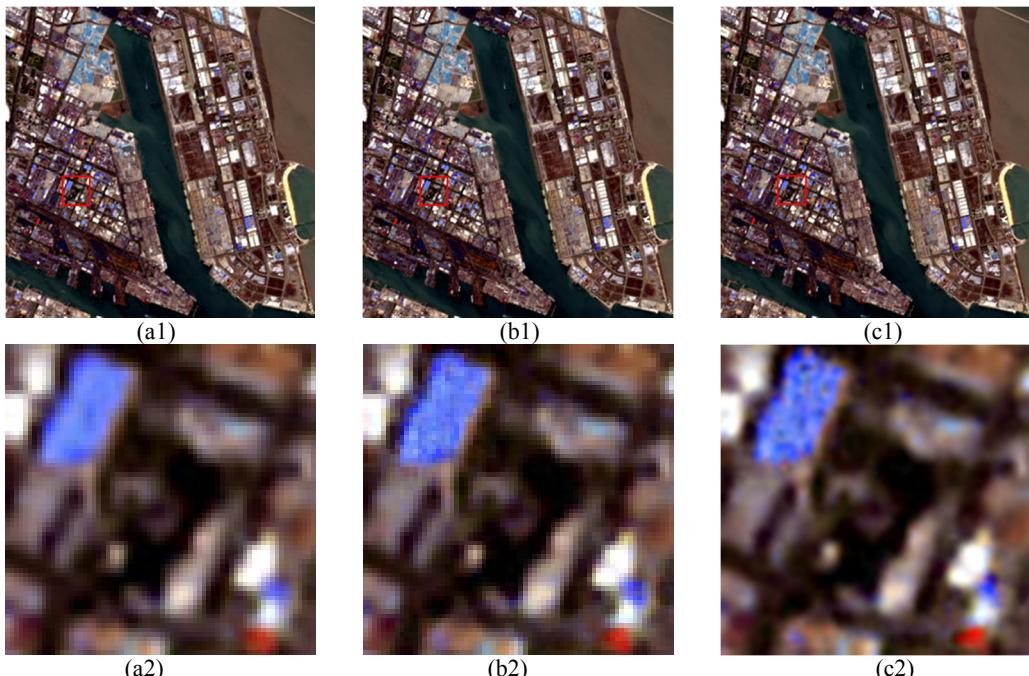


图 12 第 1 组消融实验结果。(a1)MFEB+MAB; (b1)MFEB+MAB+DC; (c1)MFEB+MAB+DC+loss; (a2)、(b2)和(c2)分别为它们的区域放大图

Fig.12 The first set of ablation experiment results. (a1) MFEB+MAB;(b1) MFEB+MAB+DC;(c1)MFEB+MAB+DC+loss;(a2), (b2) and (c2) are their regional enlarged views respectively

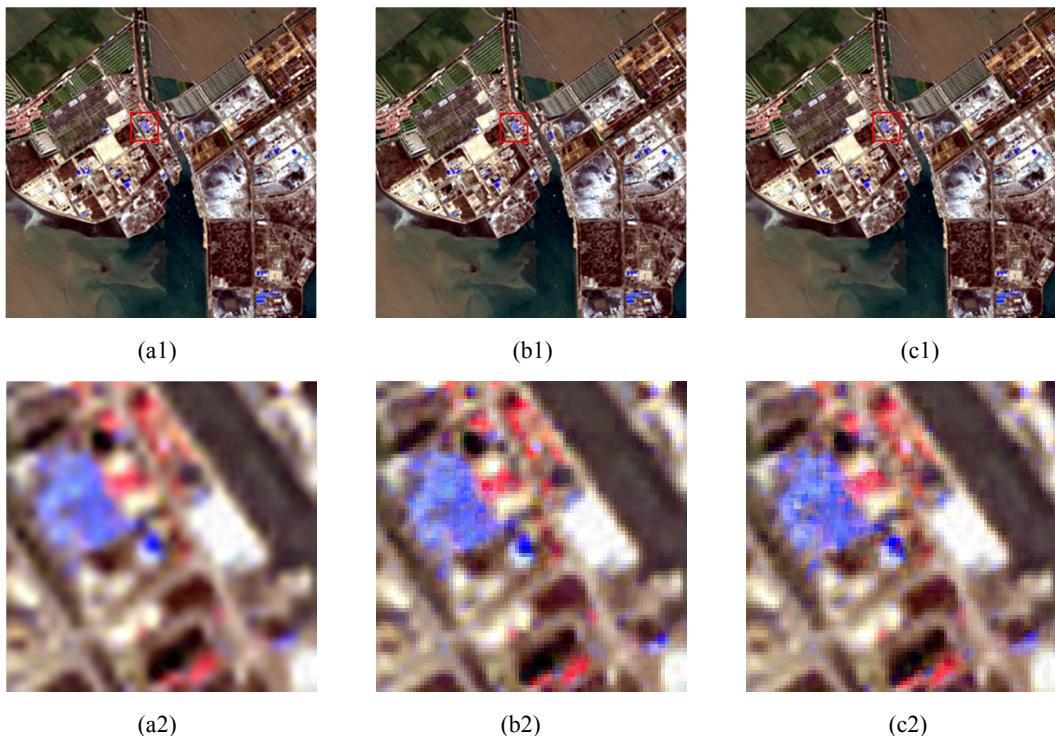


图 13 第 2 组消融实验结果。(a1)MFEB+MAB; (b1)MFEB+MAB+DC; (c1)MFEB+MAB+DC+loss; (a2)、(b2)和(c2)分别为它们的区域放大图

Fig.13 The second set of ablation experiment results. (a1) MFEB+MAB; (b1) MFEB+MAB+DC; (c1)MFEB+MAB+DC+loss; (a2), (b2) and (c2) are their regional enlarged views respectively

通过表 5 和表 6 来看, 对比 3 种构架方式, 可以看出方法(c) MFEB+MAB+DC+loss (本文算法) 的实验效果最好, 方法(a) MFEB+MAB 的效果最差。

从表 5 中可以看出, 方法(c)在 5 个评价指标 CC、PSNR、SAM、sCC 和 UIQI 上均取得最优值, 分别为 0.9947、32.8233、1.6073、0.8387 和 0.9921。从图 12

和图13中(a1)、(b1)、(c1)局部区域图,可以看出融合图像的细节逐渐清晰,纹理逐渐丰富,可以看出本文所提出的模块,对本章算法起到较为明显的提升。

表5 第1组消融实验数据

Table 5 The first set of ablation experiment data

Methods	CC↑	PSNR↑	SAM↓	sCC↑	UIQI↑
a	0.9913	32.4855	1.6343	0.8394	0.9167
b	0.9934	32.8586	1.6257	0.8434	0.9175
c	0.9953	33.2601	1.5946	0.8468	0.9244

表6 第2组消融实验数据

Table 6 The second set of ablation experiment data

Methods	CC↑	PSNR↑	SAM↓	sCC↑	UIQI↑
a	0.9906	31.8541	1.6371	0.8342	0.9134
b	0.9927	32.4387	1.6288	0.8366	0.9161
c	0.9947	32.8233	1.6073	0.8387	0.9221

3 总结与展望

本文提出一种基于双通道混合注意力的SAR和MS图像融合算法。将双通道多尺度特征提取模块和混合注意力用于图像的特征提取。最后采用Sentinel-1B SAR图像和Landsat 8 MS图像(SAR与MS图像分辨率之比为3)验证实验,具体结论如下:

1) 本文设计的双通道多尺度模块有着不俗的特征提取能力,用来分别提取SAR图像的细节特征和MS图像的光谱特征,可以有效改善细节丢失和光谱失真现象,同时在光谱和细节上保持了较好的平衡。

2) 本文构建的混合注意力模块,可以有效加强特征提取能力,并且通过对不同部分的输入特征进行加权,使网络更加关注有用信息,提高特征的表示能力。

3) 构建基于SAD的损失函数,通过度量融合图像与原始MS图像之间的光谱角距,进一步缓解融合图像的光谱失真和细节丢失现象。

因为缺少相关的数据源,本文算法只在Sentinel-1B和Landsat 8图像上验证实验,未来可考虑在分辨率更高的遥感图像上做进一步研究,例如,高分3号SAR图像和高分1号MS图像,获得更加清晰的融合图像,提高算法的适用性。

参考文献:

- [1] TU T M, HUANG P S, HUNG C L, et al. A fast intensity-hue-saturation fusion technique with spectral adjustment for IKONOS imagery[J]. *IEEE Geoscience and Remote Sensing Letters*, 2004, 1(4): 309-312.
- [2] Pal S K, Majumdar T J, Bhattacharya A K. ERS-2 SAR and IRS-1C LISS III data fusion: A PCA approach to improve remote sensing based geological interpretation[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2007, 61(5): 281-297.
- [3] TU T M, LEE Y C, CHANG C P, et al. Adjustable intensity-hue-saturation and Brovey transform fusion technique for IKONOS/QuickBird imagery[J]. *Optical Engineering*, 2005, 44(11): 116201.
- [4] Burt P J, Adelson E H. *The Laplacian Pyramid as a Compact Image Code*[M]. Readings in Computer Vision. Morgan Kaufmann, 1987: 671-679.
- [5] Ranchin T, Wald L. The wavelet transform for the analysis of remotely sensed images[J]. *International Journal of Remote Sensing*, 1993, 14(3): 615-619.
- [6] GUO K, Labate D, Lim W Q. Edge analysis and identification using the continuous shearlet transform[J]. *Applied and Computational Harmonic Analysis*, 2009, 27(1): 24-46.
- [7] DA Cunha A L, ZHOU J, DO M N. The nonsubsampled contourlet transform: theory, design, and applications[J]. *IEEE Transactions on Image Processing*, 2006, 15(10): 3089-3101.
- [8] Masi G, Cozzolino D, Verdoliva L, et al. Pansharpening by convolutional neural networks[J]. *Remote Sensing*, 2016, 8(7): 594.
- [9] WEI Y, YUAN Q, SHEN H, et al. Boosting the accuracy of multispectral image pansharpening by learning a deep residual network[J]. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(10): 1795-1799.
- [10] YANG J, FU X, HU Y, et al. PanNet: A deep network architecture for pansharpening[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 5449-5457.
- [11] 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[J]. 地理与地理信息科学, 2021, 37(2): 22-30.
- [12] WU J H, YANG X Z, FANG S, et al. SAR and multispectral image fusion experiment based on dual branch convolutional neural network [J]. *Geography and Geo-information Science*, 2021, 37(2): 22-30.
- [13] XU H, MA J, JIANG J, et al. U2Fusion: A unified unsupervised image fusion network[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 44(1): 502-518.
- [14] LIU Q, HAN L, TAN R, et al. Hybrid attention based residual network for pansharpening[J]. *Remote Sensing*, 2021, 13(10): 1962.
- [15] 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[J]. 地理与地理信息科学, 2023, 39(1): 23-30.
- [16] DONG Z Y, XU D L, ZHANG J, et al. Architecture and experiments of SAR and multispectral image fusion based on double-branch multiscale residual-fusion nesting[J]. *Geography and Geo-information Science*, 2023, 39(1): 23-30.
- [17] MIN A, GUO Z, LI H, et al. JMnet: Joint metric neural network for hyperspectral unmixing[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 1-12.
- [18] 郭彭浩. 基于卷积神经网络和贝叶斯理论的遥感图像Pansharpening算法研究[D]. 南京: 南京信息工程大学, 2021.
- [19] GUO P H. Research on Pansharpening Algorithm of Remote Sensing Image Based on Convolution Neural Network and Bayesian Theory[D]. Nanjing: Nanjing University of Information Engineering, 2021.

- [17] 申兴成, 杨学志, 董张玉, 等. 结合扩张卷积的残差网络 SAR 图像去噪[J]. 测绘科学, 2021, 46(12): 106-114.
- SHEN X C, YANG X Z, DONG Z Y, et al. Residual network combined with dilated convolution for SAR image denoising[J]. *Science of Surveying and Mapping*, 2021, 46(12): 106-114.
- [18] 黄玲琳, 李强, 路锦正, 等. 基于多尺度和注意力模型的红外与可见光图像融合[J]. 红外技术, 2023, 45(2): 143-149.
- HUANG L L, LI Q, LU J Z, et al. Infrared and visible image fusion based on multi-scale and attention model[J]. *Infrared Technology*, 2023, 45(2): 143-149.
- [19] 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.
- [20] WANG Q, WU B, ZHU P, et al. Supplementary material for 'ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 13-19.
- [21] LIU Y, SHAO Z, Hoffmann N. Global attention mechanism: retain information to enhance channel-spatial interactions[J/OL]. *arXiv preprint arXiv*, 2021, <https://arxiv.org/abs/2112.05561>.
- [22] DONG C, LOY C C, HE K, et al. Image super-resolution using deep convolutional networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(2): 295-307.
- [23] LI P, LEE S H, HSU H Y, et al. Nonlinear fusion of multispectral citrus fruit image data with information contents[J]. *Sensors*, 2017, 17(1): 142.
- [24] ZHOU J, Civco D L, Silander J A. A wavelet transform method to merge Landsat TM and SPOT panchromatic data[J]. *International Journal of Remote Sensing*, 1998, 19(4): 743-757.
- [25] WANG Z, Bovik A C. A universal image quality index[J]. *IEEE Signal Processing Letters*, 2002, 9(3): 81-84.
- [26] CHENG J, LIU H, LIU T, et al. Remote sensing image fusion via wavelet transform and sparse representation[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 104: 158-173.
- [27] WANG X L, CHEN C X. Image fusion for synthetic aperture radar and multispectral images based on sub-band-modulated non-subsampled contourlet transform and pulse coupled neural network methods[J]. *The Imaging Science Journal*, 2016, 64(2): 87-93.
- [28] SHAO Z, CAI J. Remote sensing image fusion with deep convolutional neural network[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2018, 11(5): 1656-1669.