

基于多层小波深度聚合网络的高光谱 图像超分辨率方法

方健¹, 杨劲翔^{1,2}, 肖亮^{1,2*}

(1. 南京理工大学计算机科学与工程学院, 江苏南京 210094; 2. 江苏省光谱成像与智能感知重点实验室, 江苏南京 210094)

摘要: 利用低空间分辨率高光谱(Low Resolution HyperSpectral Image, LR-HSI)和高空间分辨率多光谱图像(High Resolution MultiSpectral Image, HR-MSI)的有机结合,实现高光谱空间分辨率增强,是当前高光谱图像处理的热点问题. 目前,深度学习已成为高光谱-多光谱图像融合超分辨率的代表性方法,然而如何有效挖掘两者的互补空谱信息,实现空间结构和细节注入,在提升高光谱图像空间分辨率的同时保持高保真光谱信息,依然存在诸多挑战. 本文提出了一种多层小波深度聚合网络(Multilevel Wavelet-Deep Aggregation Network, MW-DAN). 该网络有机结合非抽取小波(UnDecimated Wavelet Transform, UDWT)分解和深度残差网络,建立双分支互补信息融合网络,提升图像重建性能. 其中,通过深度残差网络中引入跳层汇聚连接,设计信息聚合型结构,并对多光谱图像进行UDWT方向子带分解,逐层注入到网络中间隐层,增强了方向子带结构的细节注入和光谱保真能力. 整个网络通过LR-HSI, HR-MSI和HR-HSI(High Resolution HyperSpectral Image)端对端训练,能够学习性能优越的空-谱融合的超分辨率非线性映射. 大量仿真数据集和真实数据集上的大量融合实验表明,本文提出的方法在客观评价指标、光谱保持和视觉效果上优于目前主流的深度学习方法.

关键词: 高光谱图像; 图像融合; 深度学习; 非抽取小波变换; 深度残差聚合模块

基金项目: 国家自然科学基金(No.61871226, No.61571230, No.62001226); 江苏省重点研发计划(No.BE2018727); 江苏省自然科学基金(No.BK20200465); 中央高校基本科研业务费专项资金(No.30920021134)

中图分类号: TP751 **文献标识码:** A **文章编号:** 0372-2112(2024)01-0201-16

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220800

MW-DAN: Multilevel Wavelet-Deep Aggregation Network for Hyperspectral Image Super-Resolution

FANG Jian¹, YANG Jing-xiang^{1,2}, XIAO Liang^{1,2*}

(1. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, China;

2. Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing, Jiangsu 210094, China)

Abstract: The utilization of the synergistic fusion of low resolution hyperspectral image (LR-HSI) and high resolution multispectral image (HR-MSI) for the purpose of achieving enhanced hyperspectral spatial resolution has emerged as a prominent and actively pursued research area within the domain of hyperspectral image processing. At the present time, deep learning has become an efficient tool for HSI-MSI fusion. Despite the potential of deep learning, there are still some challenging, such as how to effectively mine the complementary information of HSI and MSI, how to inject the spatial structure and detail of MSI into HSI, and how to maintain the spectral fidelity of HSI. This study proposes a multilevel wavelet-deep aggregation network (MW-DAN). It has dual branches, which combine undecimated wavelet transform (UDWT) with deep residual network to promote the image reconstruction. Particularly, the UDWT directional subband decomposition of MSIs is performed by introducing jumper aggregation connections in the deep residual network to design an information aggregation type structure, and injected into the middle hidden layer of the network layer by layer to enhance the detail injection and spectral fidelity of the directional subband structure. The entire network are trained from LR-HSI, HR-MSI and

HR-HSI in an end-to-end fashion. It could learn the spatial-spectral fusion nonlinear mapping with superior performance. Experimental results on simulation and real datasets show that the proposed method is superior to the state-of-the-art fusion methods in terms of objective evaluation index, spectral fidelity and visual performance.

Key words: hyperspectral image; image fusion; deep learning; undecimated wavelet transform; deep residual aggregation module

Foundation Item(s): National Natural Science Foundation of China (No.61871226, No.61571230, No.62001226); Key R&D Plan of Jiangsu Province (No.BE2018727); Jiangsu Natural Science Foundation (No.BK20200465); Special Fund for Basic Research Business Expenses of Central Universities (No.30920021134)

1 引言

高光谱相机在可见光-近红外、短波红外甚至中红外和热红外波段范围内成像,具有纳米(nm)级光谱分辨率,其获取的数据是包含丰富空间、辐射和光谱信息的“图谱合一”数据立方体.它被广泛应用于对地观测、矿产资源调查、环境监测和军事侦察等领域^[1-3].然而,由于现有成像传感器的限制,很难获取同时具有高空间和高光谱分辨率的图像.由于该问题具有欠定性和互补信息融合的挑战,高光谱图像超分辨是遥感领域的前瞻问题.高光谱图像超分辨分为单幅高光谱图像超分辨和基于辅助图像融合的高光谱图像超分辨.本文聚焦高光谱与多光谱图像融合问题,即将相同场景的高空间分辨率多光谱图像(High Resolution MultiSpectral Image, HR-MSI)和低空间分辨率高光谱图像(Low Resolution HyperSpectral Image, LR-HSI)进行融合来实现超分辨率增强^[4].

早期的方法主要基于细节注入框架实现,包括经典的成分替代(Component Substitution, CS)、多分辨率分析(Multi-Resolution Analysis, MRA)等.其中,CS方法包括亮度-色调-饱和度(Intensity-Hue-Saturation, IHS)^[5]、主成分分析(Principal Component Analysis, PCA)^[6]和自适应Gram-Schmidt(GS Adaptive, GSA)^[7,8]等.这些方法往往通过将LR-HSI插值到HR-MSI尺度,之后进行特定投影变换,在投影空间进行成分替代或者细节注入,因此可以统一为投影成分替代方法.这种框架通常服从文献[9]中提出的“空间结构细节注入增强分辨率”,法语中简称为ARSIS(Amélioration de la Résolution Spatiale par Injection de Structures)的概念.多分辨率分析(MultiResolution Analysis, MRA)方法将图像分解至不同的尺度空间,实现尺度自适应空谱细节注入和特征的融合.另一个符合ARSIS的理念的框架是MRA方法.例如, Selva等人^[10]通过广义拉普拉斯金字塔提取HR-MSI细节信息注入到LR-HSI中来获得HR-HSI.文献[11]提出一种基于多分辨率小波分解的图像融合方法,该方法将HR-MSI分解的高频子带图像

注入到上采样后的LR-HSI数据中得到融合图像.根据不同的小波MRA框架,研究者提出了不同改进方法^[12,13],包括严格采样的离散小波(2D-DWT, 3D-DWT)和非抽取离散小波变换(UnDecimated Wavelet Transform, UDWT)方案.虽然CS和MRA的融合方法能有效将MSI的空间细节注入到HSI图像,但是容易引起较为严重的光谱失真.

模型优化方法是高光谱图像融合超分辨的另一大类方法,其基本思路是根据高光谱成像的退化过程,利用HSI和MSI的数据约束关系设计数据保真项,并结合图像先验建立优化模型进行求解.代表性方法包括基于贝叶斯、基于矩阵分解和基于张量表示的方法.基于贝叶斯的方法^[14-17]通过最大后验概率对融合过程进行贝叶斯推断模型.例如,文献[14]提出了非参数贝叶斯融合方法,该方法首先推断了场景中物质光谱的概率分布及其比例,之后利用这些分布计算高分辨率图像的稀疏编码.其关键是利用Beta过程学习贝叶斯字典,然后通过贝叶斯稀疏编码重建HSI图像.而文献[17]则将贝叶斯估计和非抽取小波相结合,提出一种针对噪声图像更鲁棒的HSI-MSI融合方法.基于矩阵分解的方法^[18-23]通过优化来估计光谱基和系数.文献[18]提出矩阵分解高光谱融合方法,该方法首先利用解混算法求解高光谱的光谱基,然后将求解的光谱基和多光谱图像联合求解高光谱图像.董伟生等人^[22]提出一种基于块坐标下降优化技术的高效非负字典学习算法,该算法通过迭代更新原子,并利用学习到的稀疏编码之间的空间相关性提高融合精度.基于张量表示方法^[24-28]利用了HSI和MSI同时包含空间和光谱信息的三维立方体结构的特性,该类方法利用了不同的张量表示方法.文献[24]提出将超分辨率问题重新表述为3种模态的字典和对应的核心张量的估计,并将HSI的非局部空间自相似性引入张量分解.文献[25]将超分辨率问题表述为张量链秩正则化优化问题,该方法首先利用非局部相似性原理,将HR-HSI(High Resolution HyperSpectral Image), HR-MSI和LR-HSI的相似块构成4-D张量,然后对这些4-D张量施加低秩张量链约束.基于模型的方法,充分利用稀疏性、低秩性和全局相似

性等图像先验,具有良好的解释性,但是这些先验很难刻画高光谱图像复杂的相关性和非线性特征,融合性能有限.

由于神经网络卓越的非线性特征学习和复杂特征表达能力,深度学习在计算机视觉和模式识别中取得巨大成功,因此在高光谱图像融合领域也受到广泛关注.早期的深度学习融合方法建立在端到端的网络形式下,利用不同的卷积变形形式和网络连接策略,学习从 LR-HSI 和 HR-MSI 到 HR-HSI 的超分辨映射关系^[29-32].例如,文献[31]利用残差学习、空洞卷积和多通道卷积联合构建网络学习深度特征来重建 HSI 图像.文献[32]使用双分支神经网络分别提取 LR-HSI 和 HR-MSI 图像的深度特征,最后通过全连接层进行特征融合重构 HR-HSI.模型驱动的深度学习方法是目前研究的新思路.这类方法往往结合正向过程和深度学习先验,提升深度融合网络的重建性能,并增强网络的可解释性.例如,文献[33]提出一种深度高光谱锐化方法,该方法利用深度卷积残差网络学习图像先验信息,依据退化约束关系建立空-谱保真的最小化模型进行求解.文献[34]提出深度盲源图像融合网络,该网络学习空间和光谱退化过程,并克服了融合算法中必须已知点扩散函数和光谱响应函数的缺陷.文献[35]将深度图像先验整合到融合模型中,提出了一种无监督网络来融合 HSI 图像.为了提高深度融合网络的可解释性,文献[36]提出一种 HSI 和 MSI 融合变分网络(VaFuNet).该网络基于变分融合计算模型的迭代过程,建立模型超参数和网络参数可学习的隐式迭代网络,并引入空谱注意力机制提升细节注入和光谱保真能力.针对有监督深度学习需要大量的数据样本训练网络模型的情况,无监督深度学习融合方法受到广泛关注.例如,文献[37]提出无监督编码器-解码器架构来解决高光谱图像超分辨问题.虽然基于深度学习的 HSI-MSI 融合方法取得了一定进展,但是由于高光谱图像复杂的退化过程和复杂的空谱联合结构,如何进一步设计新型网络架构,并联合多尺度多方向的空间细节注入机制,仍然是值得研究的方向.

目前,端到端的深度融合网络参数规模大,在不同方向高频细节注入和信息融合学习机制等方面存在不足,且网络参数规模较大.本文利用基于 MRA 方法提取多光谱图像空间细节信息,来增强 HSI 图像的多尺度空间细节信息,提出一种多层小波深度聚合网络(Multilevel Wavelet-Deep Aggregation Network, MW-DAN),如图 1 所示.其中, C_n 是尺度 n 下的低频子带图像; W_n^1 ,

W_n^2 和 W_n^3 分别对应尺度 n 的水平、垂直和对角 3 个方向的高频子带图像; Conv1*1, Conv3*3 和 Conv5*5 分别表示空间大小为 1×1, 3×3 和 5×5 的卷积. MW-DAN 由空间细节抽取模块(Spatial Detail Extraction module, SDE)和深度残差聚合模块(Deep Residual Aggregation module, DRA)组成.空间细节抽取模块使用非抽取小波变换(Undecimated Wavelet Transform, UWT)将 HR-MSI 图像分解为低频子带图像和高频子带图像,然后对 LR-HSI 上采样至 HR-MSI 同样的空间大小,接着将低通子带图像和上采样的 HSI 拼接作为网络的输入,最后将不同尺度的高频子带图像依次注入到 DRA.在 DRA 模块中,为了使残差信息不随网络的加深而丢失,使用跳层聚合的方式将信息传递到后面的网络层.经过 n 次 DRA 模块得到融合图像.为了适应非抽取小波提取 HR-MSI 的低频子带图像和高频子带图像,本文使用 L1 范数作为损失函数.大量实验结果表明,本文提出的方法融合的图像相比于目前主流先进算法,拥有更加接近参考图像的细节信息和更少的光谱失真.

本文的主要贡献有:

(1)提出一种多层小波深度聚合网络,该网络采取双分支端对端学习架构,并引入通过小波方向子带逐层注入空间高频信息并减少网络的参数规模,提升高光谱与多光谱图像融合的重建性能;

(2)区别于其他深度残差网络的跳层连接和稠密连接,在深度残差网络中引入跳层聚合连接模块,缩短了各层接受到损失函数反向传播的路径总长,有效抑制梯度消失,并提升了特征复用和信息聚合能力,实现了空间细节和光谱信息实现高效互补融合,提升了光谱保真重建能力;

(3)通过在 CAVE 和 Harvard 数据集上进行大量融合实验和消融实验,利用客观指标、光谱保持和视觉效果等手段进行综合评价,证明本文方法优于目前主流的深度学习基准方法.

2 本文方法

基于多分辨率分析的融合方法可以有效地将 HR-MSI 空间细节注入到上采样后的 LR-HSI,但融合质量较差.基于深度学习的方法利用神经网络学习深度特征来融合 HSI 和 MSI.实验证明,深度学习可以很好地提升融合效果,然而现有的神经网络无法充分挖掘多光谱数据的空间细节信息.本文利用多分辨率分析方法提取多光谱的空间细节注入深度特征来融合 HSI 图像,不仅增强了空间细节信息,还较好地保留了光谱信息.

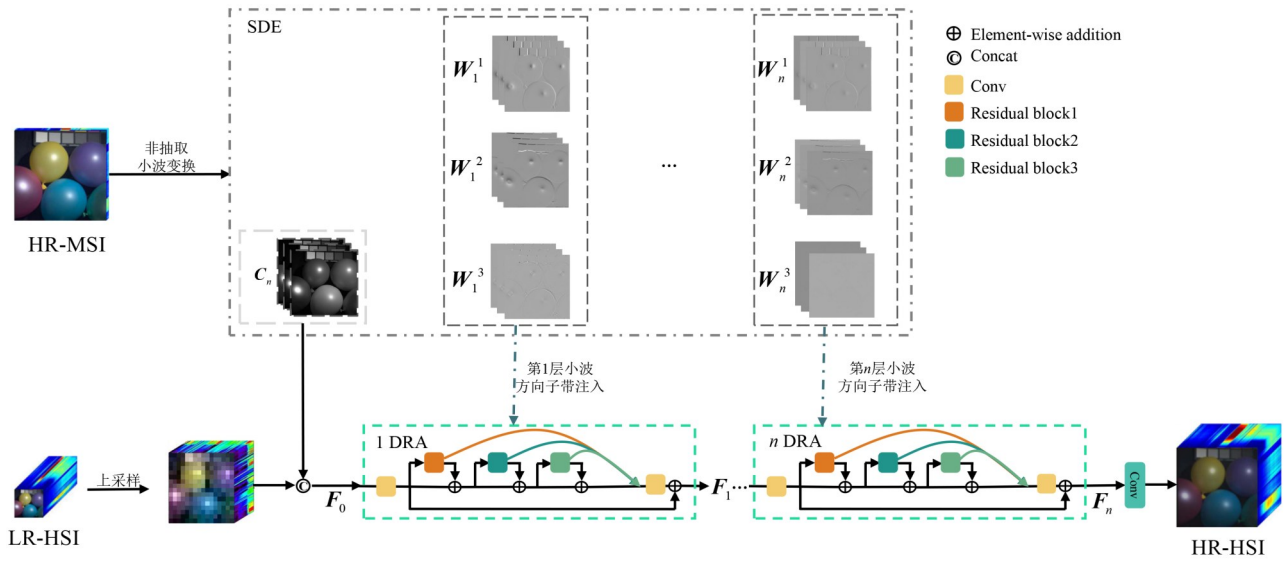


图1 多层小波深度聚合网络

2.1 问题描述

在传统的高光谱MRA方法的细节注入框架,通常提取多光谱图像的细节注入到上采样后的低分辨率高光谱图像.考虑到利用非抽取小波变换方法提取高频细节,此过程可看成多光谱图像与其低频子带图像的残差作为高频细节,然后将提取的细节注入到低分辨率的高光谱图像中来增强高光谱图像空间分辨率.多光谱图像的低频子带图像表示为

$$\mathbf{ML} = \mathbf{M} * h_{LP} \quad (1)$$

其中,*表示卷积运算; \mathbf{M} 表示多光谱图像; h_{LP} 表示非抽取小波中的低通滤波器.整个细节注入框架为

$$\bar{H}_n = \hat{H}_n + g_n(M_i - \mathbf{ML}_i) \quad (2)$$

其中, \bar{H}_n 和 \hat{H}_n 分别表示融合后的高光谱和上采样后的高光谱第 n 个波段; g_n 表示第 n 个波段的注入增益; M_i 表示与 \hat{H}_n 对应的多光谱图像第 i 个波段.

随着深度学习的兴起,研究者提出一系列深度学习的高光谱超分辨率方法.为了解决深度网络隐藏层过多时梯度消失的问题,文献[38]提出残差网络.通常残差网络是多个残差块以堆叠形式作为网络结构,如图2(a)所示.残差网络为解决高光谱融合问题提供了有效途径.文献[39]利用HR-HSI与HR-MSI相似的空间结构信息,在残差网络中输入HR-MSI引导特征学习,来融合HR-HSI.文献[40]通过增强残差密集网络联合循环神经网络提取空谱特征,利用这些空谱特征重构图像,密集残差块如图2(b)所示.

残差块以堆叠形式作为网络结构时,浅层的残差信息会随着网络的加深逐渐对网络的影响越来越弱,因此文献[41]提出残差聚合模块.该模块将连续的几个残差块和深度特征进行融合,使前面更多的残差信

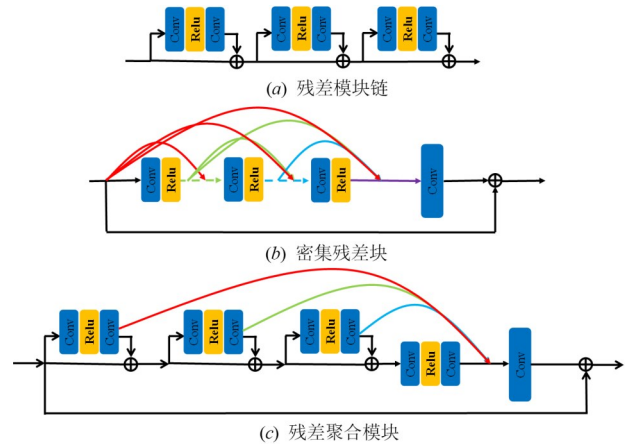


图2 目前存在的残差块

息能够传递到后面的网络层,如图2(c)所示.残差聚合模块相比密集残差块,拥有更少的模型参数量并且提升了性能.因此本文将残差聚合网络结构作为基础模块.

常见的高光谱图像超分辨网络框架,首先将高光谱图像上采样或反卷积至HR-MSI相同空间大小,然后将MSI和上采样后的HSI图像在通道维拼接作为网络的输入,最后通过卷积神经网络获得HR-HSI.此过程表示为

$$\mathbf{X} = f_{\Theta_1}(\text{Concat}(\mathbf{Y}, H_{up}(\mathbf{Z}))) \quad (3)$$

其中, $f_{\Theta_1}(\cdot)$ 表示神经网络; Θ_1 表示网络参数; $H_{up}(\cdot)$ 表示上采样操作; $\text{Concat}(\cdot)$ 表示拼接操作. $\mathbf{X} \in \mathbb{R}^{W \times H \times B}$, $\mathbf{Y} \in \mathbb{R}^{W \times H \times b}$ 和 $\mathbf{Z} \in \mathbb{R}^{w \times h \times B}$ ($w \ll W, h \ll H, b \ll B$)分别表示融合的HR-HSI、输入的HR-MSI和LR-HSI; h, w 和 B 分别表示LR-HSI的高度、宽度和光谱波段数; H, W 和 b 分别表示HR-MSI的高度、宽度和光谱波段数.

本文通过非抽取小波方法将HR-MSI图像分解为

低频子带图像和 3 个方向的高频子带图像. 由于低频子带图像中含有丰富的光谱信息, 因此将其作为网络的输入. 非抽取小波提取的高频子带图像在训练的过程中注入到网络. 整个过程表示为

$$X = f_{\Theta_1}(\text{Concat}(Y_{LL}, H_{up}(Z)), Y_{LH}, Y_{HL}, Y_{HH}) \quad (4)$$

其中, $f_{\Theta_1}(\cdot)$ 表示神经网络; Θ_1 表示网络参数; Y_{LL} 表示 HR-MSI 分解的低频子带图像; Y_{LH} , Y_{HL} 和 Y_{HH} 分别表示 HR-MSI 分解的水平、垂直和对角 3 个方向的高频子带图像.

2.2 多层小波深度聚合网络

本节主要由 4 个部分组成: 空间细节提取模块、深度残差聚合模块、网络整体结构和损失函数. 本文提出的网络结构, 即多层小波深度聚合网络, 如图 1 所示.

2.2.1 空间细节提取模块

基于抽取的正交和双正交离散小波在信号处理和图像压缩方面取得了非常成功的应用, 但是它们却缺少平移不变性, 当修改小波系数来重建图像时往往会大量出现大量虚假失真信息^[42].

对于融合和重构^[43]的许多应用, 研究者提出 UWT, 它通过保留滤波器组的重建, 去掉正交变换的抽取步骤, 并提供快速的二进算法, 可以增强平移不变性, 并提升图像重建性能^[43]. 因此, 本文采用非抽取小波^[44]来提取空间细节. 由于没有经过下采样操作, 因此每个小波方向子带的大小等于输入图像的大小, 其中, 低通滤波器用 h 表示, 高通滤波器用 g 表示. 具体实现时, 采取图像复原性能优异的 \acute{a} trous 小波, 其滤波器组分别为 $h = [1, 4, 6, 4, 1]/16$, $g = [-1, -4, 10, -4, -1]/16$, $\tilde{h} = \tilde{g} = [0, 0, 1, 0, 0]$, 并利用多孔算法^[45] (多孔算法是由 M. Shen 于 1992 年提出的一种利用 Mallat 算法结构计算小波变换的快速算法) 提取多光谱图像的空间细节信息来实现.

第 d 个尺度下, 行和列均为低通滤波器的低频子带图像为

$$C_d = \bar{h}^{(d)} \bar{h}^{(d)}(C_{d-1}) \quad (5)$$

3 个方向的高频子带图像为

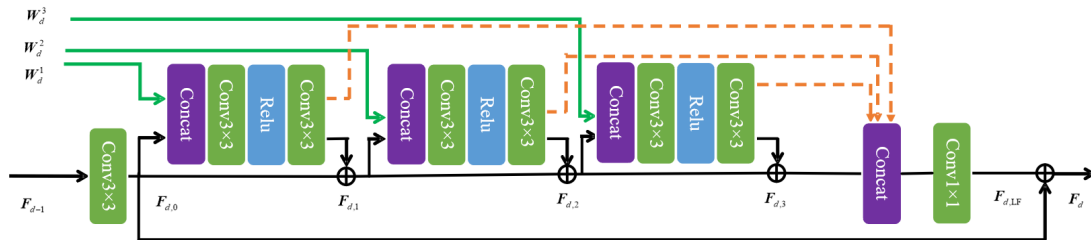


图 4 第 d 层深度残差聚合模块

本文利用前置的 3×3 卷积层提取特征, 该操作有利于更好地进行残差学习, 即

$$W_d^1 = \bar{g}^{(d)} \bar{h}^{(d)}(C_{d-1}) \quad (6)$$

$$W_d^2 = \bar{h}^{(d)} \bar{g}^{(d)}(C_{d-1}) \quad (7)$$

$$W_d^3 = \bar{g}^{(d)} \bar{g}^{(d)}(C_{d-1}) \quad (8)$$

其中, $hg(c)$ 表示图像 c 与可分离滤波器 hg 的卷积; C_{d-1} 表示第 $d-1$ 尺度下的低通子带图像; W_d^1 , W_d^2 和 W_d^3 分别表示第 d 个尺度下的 Y_{LH} , Y_{HL} 和 Y_{HH} . 多光谱图像 Y 是第 1 个尺度小波的输入, 即 $C_0 = Y$. 由于 \acute{a} trous 小波具有平移不变性, 因此将所有尺度小波方向子带图像和最终的低频子带图像简单相加就可以得到重构图像, 即

$$C_0 = C_n + \sum_{d=1}^n \sum_{j=1}^3 W_d^j \quad (9)$$

2.2.2 深度残差聚合模块

为了实现残差特征和小波子带特征的有机融合, 本文提出深度残差聚合模块. 该模块包括基本残差块、超分辨残差块和深度残差聚合模块.

(1) 基本残差块以及超分辨残差块. 残差网络^[46]是由一系列残差块组成的, 如图 3(a) 所示. 一个残差块表示为

$$x_{i+1} = F(x_i, w) + x_i \quad (10)$$

其中, x_i 和 x_{i+1} 表示网络层的输入和输出. 函数 $F(x_i, w)$ 表示学习的残差映射. 该网络结构在图像分类和超分辨中得到广泛应用. 然而, 在应用到超分辨等图像重建问题上, 由于需要保持图像幅度范围, 因此无需批归一化 (Batch Normalization, BN) 操作^[47], 如图 3(b) 所示.

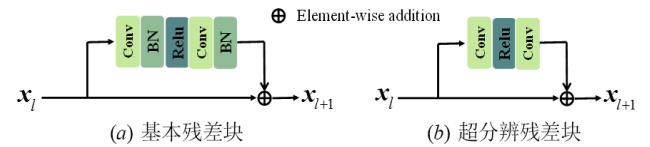


图 3 残差块

(2) 深度残差聚合模块: 该模块将非抽取小波方向子带加入到残差网络来增强融合 HSI 的空间细节, 如图 4 所示.

$$F_{d,0} = \text{Conv}_{3 \times 3}(F_{d-1}) \quad (11)$$

其中, $1 \leq d \leq n$, $\text{Conv}_{3 \times 3}$ 表示空间大小为 3×3 , 特征通道

数为 64 的卷积. 利用超分辨率残差块的基础模块 (Basic Module, BM) 学习拼接后的特征图, 表示为

$$\mathbf{F}_{\text{BM}} = \text{Con v}_{3 \times 3}(\text{Relu}(\text{Con v}_{3 \times 3}(\mathbf{F}_{\text{in}}))) \quad (12)$$

其中, \mathbf{F}_{in} 表示输入的特征图, Relu 表示 Relu 激活函数.

第 d 个 DRA 的第 c ($1 \leq c \leq 3$) 个残差特征为

$$\mathbf{RF}_{d,c} = \mathbf{F}_{\text{BM}}(\text{Concat}(\mathbf{F}_{d,c-1}, \mathbf{W}_d^{c-1})) \quad (13)$$

其中, \mathbf{W}_d^{c-1} 表示第 d 尺度下分解的第 $c-1$ 个高频子带图像.

第 d 个 DRA 的第 c 个残差输出表示为

$$\mathbf{F}_{d,c} = \mathbf{RF}_{d,c} + \mathbf{F}_{d,c-1} \quad (14)$$

为了将高频子带图像和深度特征融合的信息传递到后面的网络层, 这里将前面 3 个提取的特征通过跳跃连接方式与 $\mathbf{F}_{d,3}$ 拼接. 由于第 $d-1$ 个 DRA 的特征映射以串联方式直接引入第 d 个 DRA, 因此必须减少特征数量. 受到 MemNet^[48] 的启发, 引入一个空间大小 1×1 的卷积层来自适应地控制输出信息. 此操作表述为

$$\mathbf{F}_{d,\text{LF}} = \text{Con v}_{1 \times 1}(\text{Concat}(\mathbf{F}_{d,3}, \mathbf{RF}_{d,2}, \mathbf{RF}_{d,1}, \mathbf{RF}_{d,0})) \quad (15)$$

其中, $\text{Con v}_{1 \times 1}$ 表示空间大小为 1×1 , 特征通道数为 64 的卷积.

设 \mathbf{F}_{d-1} 和 \mathbf{F}_d 分别表示第 d 个 DRA 的输入和输出, 将 \mathbf{F}_{d-1} 经过卷积和 $\mathbf{F}_{d,\text{LF}}$ 相加, 得到 \mathbf{F}_d , 表示为

$$\mathbf{F}_d = \text{Con v}_{3 \times 3}(\mathbf{F}_{d-1}) + \mathbf{F}_{d,\text{LF}} \quad (16)$$

2.2.3 网络整体结构

整个网络的流程如图 1 所示. 首先将 HR-MSI 图像通过空间细节提取模块提取其低频子带图像和 3 个方向的高频子带图像, 然后将低频子带图像和经过双线性插值上采样的 LR-HSI 拼接作为网络的输入, 表示为

$$\mathbf{F}_0 = \text{Concat}(\mathbf{Y}_{\text{LL}}, H_{\text{up}}(\mathbf{Z})) \quad (17)$$

接着将同一尺度的高频子带图像注入 DRA 模块, 经过 n 个 DRA 模块, 得到输出的结果 \mathbf{F}_n (经验设置为 $n=2$). 此过程表示为

$$\begin{aligned} \mathbf{F}_n &= H_{D,n}(\mathbf{F}_{n-1}, \mathbf{W}_n^1, \mathbf{W}_n^2, \mathbf{W}_n^3) \\ &= H_{D,n}(H_{D,n-1}(\cdots(H_{D,1}(\mathbf{F}_0, \mathbf{W}_1^1, \mathbf{W}_1^2, \mathbf{W}_1^3))\cdots), \mathbf{W}_n^1, \mathbf{W}_n^2, \mathbf{W}_n^3) \end{aligned} \quad (18)$$

其中, $H_{D,n}$ 表示第 n 个 DRA 的操作. \mathbf{F}_n 通过一个卷积和 Relu 激活函数, 最终得到融合图像 HR-HSI \mathbf{X} , 表示为

$$\mathbf{X} = \text{Relu}(\text{Con v}_{5 \times 5}(\mathbf{F}_n)) \quad (19)$$

其中, $\text{Con v}_{5 \times 5}$ 表示空间大小为 5×5 , 通道数为 B 的卷积.

2.2.4 损失函数

MW-DAN 使用平均绝对误差 (Mean Absolute Error, MAE), 即 L1 范数, 作为损失函数. 本文利用非抽取小波变换将 HR-MSI 分解为低频子带图像和 3 个方向的高频子带图像, 而小波系数具有稀疏性. L1 范数可以刻画数值的稀疏性, 并且有锐化重建图像边缘的效果. 因此

本文使用 L1 范数作为损失函数进行网络训练. 网络的损失函数表示为

$$L(\Theta) = \frac{1}{B} \sum_{i=1}^B \|f_{\Theta}(Y_i, Z_i) - X_i\|_1 \quad (20)$$

其中, f_{Θ} 为网络模型; Θ 表示网络参数; X_i , Y_i 和 Z_i 分别表示与输入 LR-HSI 配对的 HR-HSI, HR-MSI 和 LR-HSI 的第 i 个波段.

3 实验结果与分析

3.1 实验设置

(1) 实现细节. 在网络训练过程中, 使用 $\beta_1=0.9$ 和 $\beta_2=0.999$ 的 ADAM^[49] 优化器, 学习率为 $\gamma=0.0001$, 迭代次数为 1000, 批量大小设置为 32. 所有实验均在 Window10 上进行, 使用 Intel(R) Core(TM) i7-9700 CPU 和 NVIDIA 2080TI GPU.

(2) 比较算法. 本文将 MW-DAN 与 8 种目前主流的 HSI 超分辨率方法进行比较, 包括: 5 种传统方法, 即耦合非负矩阵分解 (Coupled Nonnegative Matrix Factorization, CNMF)^[19]、HySure^[16]、CSU^[20]、非负结构稀疏表示 (Non-negative Structured Sparse Representation, NSSR)^[22] 和基于张量链秩表示的方法 (Learning a low Tensor-Train Rank representation, LTTR)^[25]; 5 种深度学习方法, 即空谱融合的 CNN (Spatial and Spectral Fusion with CNN, SSF-CNN) 高光谱图像超分辨^[30]、深度高光谱图像 Sharpening¹ (Deep Hyperspectral Image Sharpening, DHSIS)^[33]、深度盲源高光谱图像融合² (Deep Blind hyperspectral Image Fusion, DBIN)^[34]、基于 CNN 去噪的方法 (CNN-Fus)^[50]; 1 种新颖模型驱动的深度卷积网络 (MoG-DCN)^[51] 方法. CNMF^[19] 利用 HSI 的端元矩阵和 MSI 的丰度矩阵来重建 HR-HSI. HySure^[16] 利用低秩先验来重构 HR-HSI. CSU^[20] 通过考虑光谱解混约束来规范 HSI 超分辨率问题. NSSR^[22] 侧重于将潜在的 HR-HSI 的稀疏表示作为超分辨率的先验信息. LTTR^[25] 学习非局部相似 HR-HSI 立方体的空间、光谱和非局部模型之间的相似性. SSF-CNN^[30] 利用卷积神经网络学习空谱融合的映射. 在 DHSIS^[33] 中, 潜在的 HR-HSI 的先验是由一个深度神经网络利用大量数据来学习. 在 DBIN^[34] 和 MoG-DCN^[51] 中, 利用神经网络来学习包括空谱依赖的图像先验. CNN-Fus^[50] 是一种利用子空间表示和卷积神经网络去噪的算法.

为了公平起见, 在所有方法中都使用相同的数据预处理, 比较方法中基于深度学习的方法使用作者提供的代码进行训练, 并在相同训练数据上使用建议的参数及相同协议, 用于评估所有方法的实验结果.

3.2 实验数据集

为了评估本文方法的有效性, 在两个公开数据集

上进行仿真实验,即哥伦比亚计算机视觉实验室数据集(CAVE)和哈佛数据集(Harvard). CAVE 数据集由 32 个室内高光图像组成,空间大小为 512×512 , 31 个波段,波段范围为 400~700 nm,光谱分辨率为 10 nm. Harvard 数据集总共 50 个,包括室内和室外场景的高光谱图像. 高光图像的空间大小为 1040×1392 , 31 个波段,波段范围 420~720 nm,光谱分辨率为 10 nm. 在实验中,为了方便采集训练数据块,本文从图像左上方开始截取 1024×1024 空间大小作为训练和测试集. CAVE 和 Harvard 数据集都被广泛应用于高光超分辨文献中. 在仿真实验中,空间下采样因子为 8. 图 5 中展示了 CAVE 和 Harvard 数据集的一些 HR-RGB 图像.

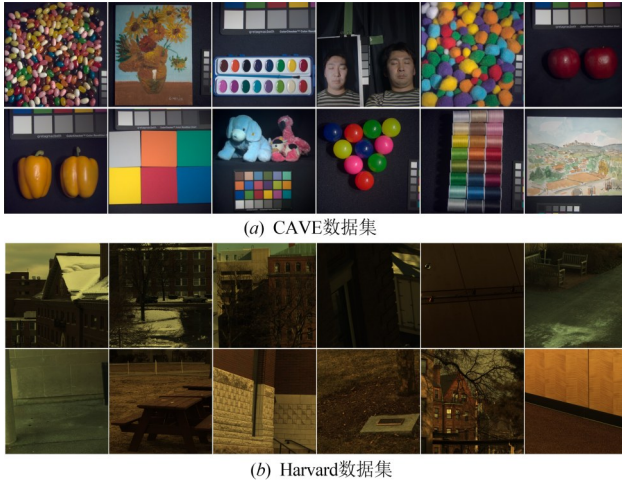


图 5 CAVE 和 Harvard 数据集的 HR-RGB 图像

参考图像来自两个数据集的高光谱图像. LR-HSI 是通过应用 8×8 高斯滤波器(均值为 0, 标准差为 2), 并通过对参考图像的每个波段在垂直和水平方向上每 8 个像素进行下采样来实现的,即抽取因子为 8×8 . 模拟同一场景的 HR-MSI (RGB 图像), 对 \mathbf{X} 进行下采样. 光谱下采样矩阵 \mathbf{S} 采用 Nikon D700 相机的响应函数. CAVE 数据集的前 20 个 HSI 用于训练, 后 12 个 HSI 用于测试. 对于 Harvard 数据集, 前 30 个 HSI 用于训练, 后 20 个 HSI 用于测试. 本文采用取块的方式获得训练样本, 每个 LR-HSI, HR-MSI 和 HR-HSI 训练数据块的大小分别为 $4 \times 4 \times 31$, $32 \times 32 \times 3$ 和 $32 \times 32 \times 31$.

3.3 量化指标

本文使用 6 个客观评价指标, 即平均峰值信噪比、光谱角映射、平均结构相似性指标、相对维数无关的全局综合误差、均方根误差和通用图像质量指数, 来评估重建的高光谱图像质量. 下面将详细描述这些指标.

平均峰值信噪比 (Mean Peak Signal-to-Noise Ratio, MPSNR) 表示所有波段峰值信噪比的平均, 定义为

$$\text{MPSNR}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{B} \sum_{i=1}^B 10 \cdot \log_{10} \left(\frac{\max(X^i)^2}{\|X^i - \hat{X}^i\|_2^2 / (H \times W)} \right) \quad (21)$$

其中, 第 i 个波段的 $\hat{\mathbf{X}}$ 和 \mathbf{X} 分别用 \hat{X}^i 和 X^i 表示. MPSNR 度量了融合图像和参考图像的相似性, 它的数值越大意味着融合图像和参考图像越相似.

光谱角映射 (Spectral Angle Mapper, SAM) 表示融合图像的像素和参考图像的像素在整个空间范围内的平均, 定义为

$$\text{SAM}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W \arccos \frac{\mathbf{X}(i,j)^T \hat{\mathbf{X}}(i,j)}{\|\mathbf{X}(i,j)\|_2 \|\hat{\mathbf{X}}(i,j)\|_2} \quad (22)$$

其中, $\hat{\mathbf{X}}(i,j)$ 和 $\mathbf{X}(i,j)$ 分别表示 $\hat{\mathbf{X}}$ 和 \mathbf{X} 在像素点 (i,j) 的光谱向量. 光谱角映射是以度为单位和估计光谱质量的指标, SAM 的数值越小, 说明融合图像的光谱畸变越少.

平均结构相似性 (Mean Structural SIMilarity, MSSIM) 定义为

$$\text{MSSIM}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{1}{B} \sum_{i=1}^B \frac{(2\mu_{X^i} \mu_{\hat{X}^i} + C_1)(2\sigma_{X^i \hat{X}^i} + C_2)}{(\mu_{X^i}^2 + \mu_{\hat{X}^i}^2 + C_1)(\sigma_{X^i}^2 + \sigma_{\hat{X}^i}^2 + C_2)} \quad (23)$$

其中, μ_{X^i} 和 $\mu_{\hat{X}^i}$ 分别表示 X^i 和 \hat{X}^i 的均值平方, σ_{X^i} 和 $\sigma_{\hat{X}^i}$ 分别表示 \hat{X}^i 和 X^i 的标准差, $\sigma_{X^i \hat{X}^i}$ 表示 \hat{X}^i 和 X^i 的协方差, $C_1 = (k_1 L)^2$ 和 $C_2 = (k_2 L)^2$ 是用来维持稳定的常数, L 是像素值的动态范围, $k_1 = 0.01$, $k_2 = 0.03$. MSSIM 表示融合图像的空间结构完整程度的指标, 它的数值越大, 说明融合图像的空间结构保留得越完整.

相对维数无关的全局综合误差 (Erreur Relative Global Adimensionnelle Synthèse, ERGAS) 定义为

$$\text{ERGAS}(\mathbf{X}, \hat{\mathbf{X}}) = \frac{100}{d} \sqrt{\frac{1}{B} \sum_{i=1}^B \frac{\text{MSE}(X^i, \hat{X}^i)}{u_{\hat{X}^i}^2}} \quad (24)$$

其中, d 是空间下采样因子, $\text{MSE}(X^i, \hat{X}^i)$ 是 X^i 和 \hat{X}^i 的均方误差. ERGAS 反映了融合图像与参考图像的误差大小, 越低的 ERGAS 数值意味着融合图像与参考图像误差越小.

均方根误差 (Root Mean Square Error, RMSE) 描述融合图像和参考图像间的接近程度, 其数值越小, 表示融合图像与参考图像越接近, 即融合效果越好. RMSE 定义为

$$\text{RMSE}(\mathbf{X}, \hat{\mathbf{X}}) = \sqrt{\frac{1}{B} \sum_{i=1}^B (X^i - \hat{X}^i)^2} \quad (25)$$

通用图像质量指数 (Universal Image Quality Index, UIQI) 表示评价融合图像与参考图像的结构失真程度, 其理想数值为 1. 数值越接近于 1, 说明图像融合质量越好. UIQI 定义为

$$UIQI = \frac{\sigma_{\hat{X}X}}{\sigma_{\hat{X}}\sigma_X} \times \frac{2\mu_{\hat{X}}\mu_X}{\mu_{\hat{X}}^2 + \mu_X^2} \times \frac{2\sigma_{\hat{X}}\sigma_X}{\sigma_{\hat{X}}^2 + \sigma_X^2} \quad (26)$$

其中, $\sigma_{\hat{X}X}$ 表示 \hat{X} 和 X 的协方差, $\sigma_{\hat{X}}$ 和 σ_X 分别表示 \hat{X} 和 X 的标准差.

3.4 实验结果

本文给出了不同方法的评估结果, 如表 1 和表 2 所

示. 对于 CAVE 和 Harvard 数据集, 首先分别在 10 和 20 个测试集上评估不同的融合方法, 然后计算评估指标的均值. 从表中可以看出, MW-DAN 结果中的 MPSNR, UIQI 和 MSSIM 指标最高, RMSE, ERGAS 和 SAM 指标最低. 这意味着 MW-DAN 方法重构的 HSI 具有更完整的空间结构和更少的光谱失真.

表 1 CAVE 数据集上测试方法的平均定量结果

方法	MPSNR	RMSE	ERGAS	SAM	UIQI	MSSIM
CNMF ^[19]	34.302 7	5.472 32	2.600 7	7.892 0	0.771 1	0.938 8
HySure ^[16]	34.782 2	5.323 2	2.418 1	11.545 1	0.804 3	0.910 7
CSU ^[20]	35.688 9	4.769 42	2.206 5	7.878 7	0.798 7	0.953 2
NSSR ^[22]	40.320 5	2.941 5	1.260 9	5.041 6	0.901 9	0.980 9
LTTR ^[25]	46.223 2	1.544 5	0.694 3	3.888 0	0.918 5	0.990 3
SSF-CNN ^[30]	42.566 3	2.186 97	0.970 8	5.046 0	0.901 3	0.986 6
DHSIS ^[33]	46.297 7	1.465 35	0.664 1	3.845 3	0.924 3	0.990 5
DBIN ^[34]	45.783 2	1.523 3	0.678 2	3.603 5	0.926 1	0.992 6
CNN-Fus ^[50]	44.578 9	1.899 4	0.868 9	5.424 1	0.873 5	0.985 0
MoG-DCN ^[51]	46.316 7	1.421 7	0.645 0	3.596 7	0.927 8	0.992 4
MW-DAN	46.864 4	1.388 9	0.617 2	3.393 3	0.933 1	0.993 1
最好的值	$+\infty$	0	0	0	1	1

表 2 Harvard 数据集上测试方法的平均定量结果

方法	MPSNR	RMSE	ERGAS	SAM	UIQI	MSSIM
CNMF ^[19]	42.136 9	2.502 6	1.087 4	2.812 0	0.861 3	0.971 6
HySure ^[16]	40.749 3	2.960 2	1.316 6	3.467 6	0.865 3	0.969 4
CSU ^[20]	40.370 9	3.153 7	1.374 1	3.261 3	0.858 1	0.972 0
NSSR ^[22]	45.850 8	1.552 4	0.911 1	2.357 3	0.894 5	0.981 8
LTTR ^[25]	45.969 2	1.502 4	0.887 5	2.306 1	0.891 7	0.982 2
SSF-CNN ^[30]	45.067 1	1.502 4	0.909 9	2.502 3	0.892 2	0.981 6
DHSIS ^[33]	45.519 7	1.618 8	0.916 7	2.442 2	0.890 6	0.980 9
DBIN ^[34]	45.673 1	1.596 0	0.832 1	2.360 0	0.893 8	0.982 3
CNN-Fus ^[50]	44.832 2	1.780 3	1.015 2	2.825 0	0.869 7	0.976 8
MoG-DCN ^[51]	45.928 3	1.544 1	0.816 6	2.404 9	0.894 6	0.982 5
MW-DAN	46.195 5	1.511 9	0.785 7	2.272 2	0.895 7	0.982 6
最好的值	$+\infty$	0	0	0	1	1

由于从图 6 和图 7 中观察到融合的 HR-HSI 彼此接近, 因此为了从视觉上突出它们的差异, 本文在图 8 和图 9 中给出均方误差图像的可视化热力图, 其中蓝色表示误差小, 红色表示误差大.

从图 8 中可以看出, CNMF 和 HySure 的融合结果有大面积重建误差. CSU 方法的融合图像在方框上面有部分重建误差很大. LTTR, SSF-Net, DHSIS 和 CNN-Fus, 这些方法融合的图像在方框边缘上和方框下面的

字有很大的重建误差. NSSR, DBIN 和 MoG-DCN 的融合图像只在方框上有少部分重建误差. 本文方法融合的图像拥有最少的重建误差. 参考图像与融合图像每个波段 MPSNR 和 MSSIM 展示在图 10(a) 和图 11(a) 中. 从图中可以看出, 本文方法融合的图像在每个波段上指标都最高.

从图 9 中看出, CNMF, HySure 和 CSU 方法融合的图像在地上的雪和枝条上有较大的重建误差. NSSR,

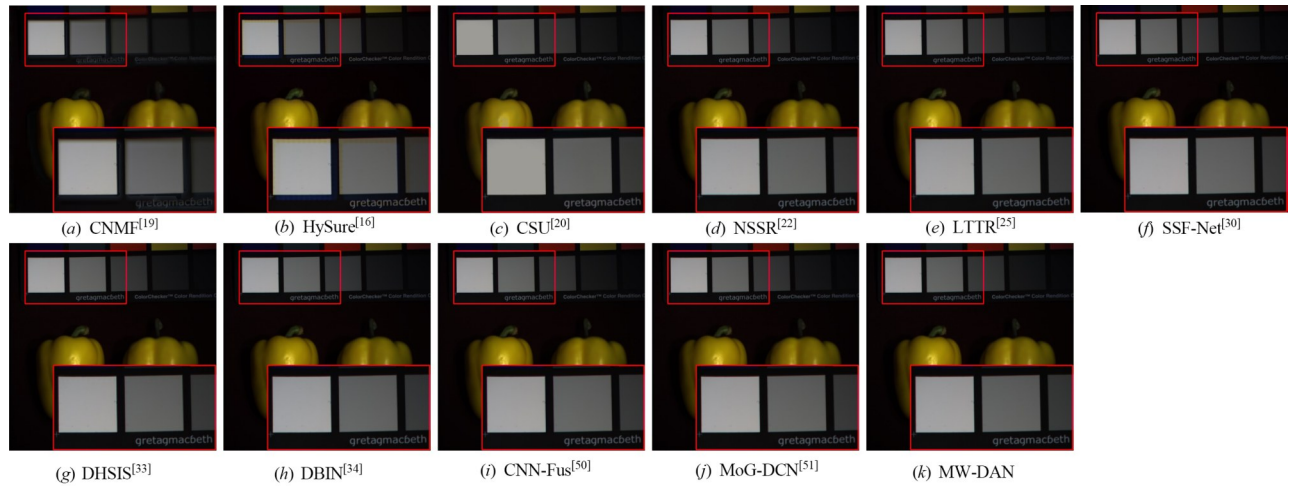


图6 Real and fake apples的融合图像伪彩色合成图(波段 30, 20, 10)

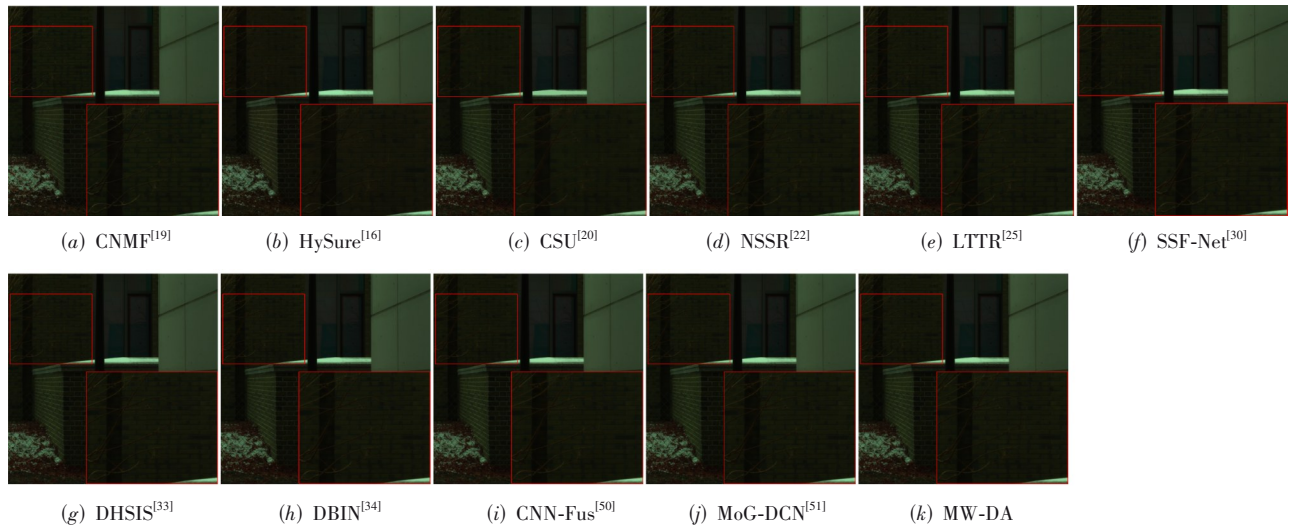


图7 Imgc1融合图像的伪彩色合成图(波段 30, 20, 10)

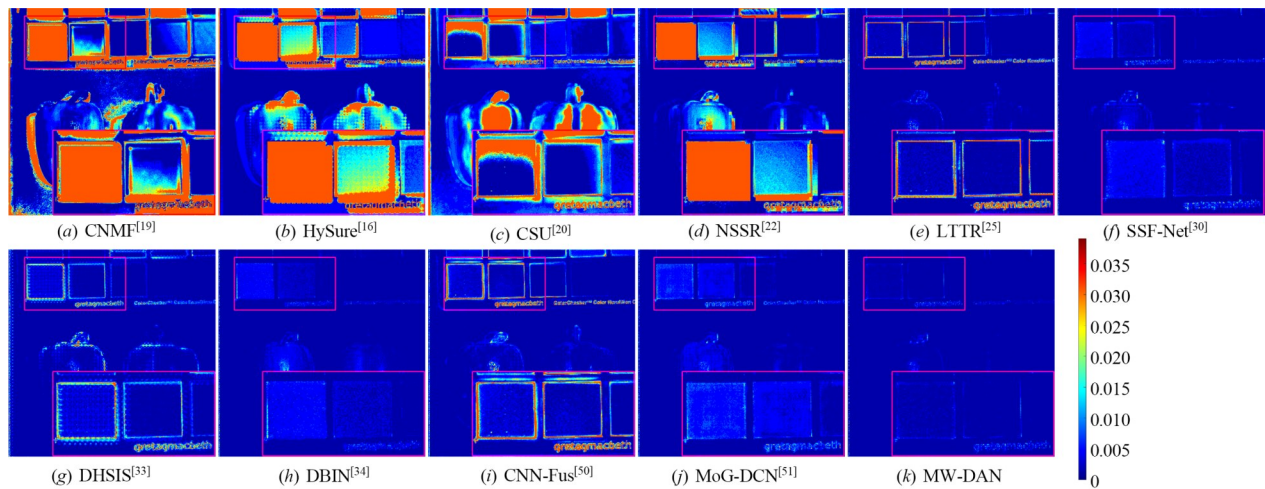


图8 Real and fake apples的融合图像比较方法的均方误差图像

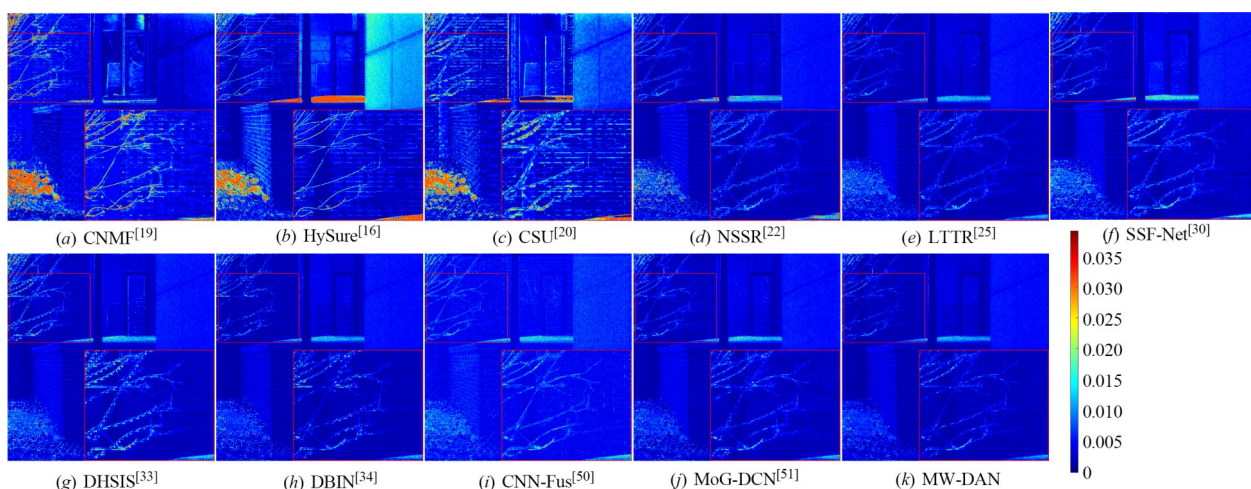
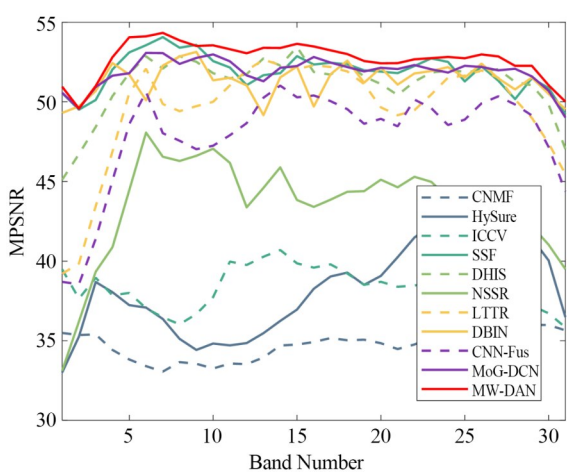


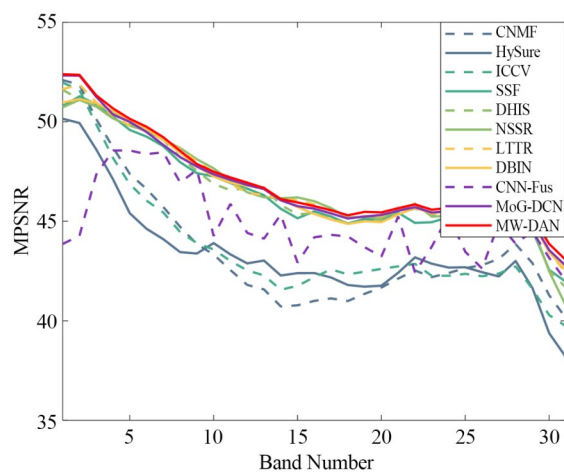
图9 Imge1融合图像比较方法的相应均方差图像

LTTR, SSF-Net, DHSIS, DBIN, CNN-Fus 和 MoG-DCN, 这些方法的融合结果只在枝条上有较大的重建误差. 本文方法融合的图像具有最小的重建误差. 参考图像与

融合图像每个波段段的 MPSNR 和 MSSIM 如图 10(b) 和图 11(b)所示. 从图中可以看出, 本文方法融合的图片在每个波段上的指标都超过了其他方法.

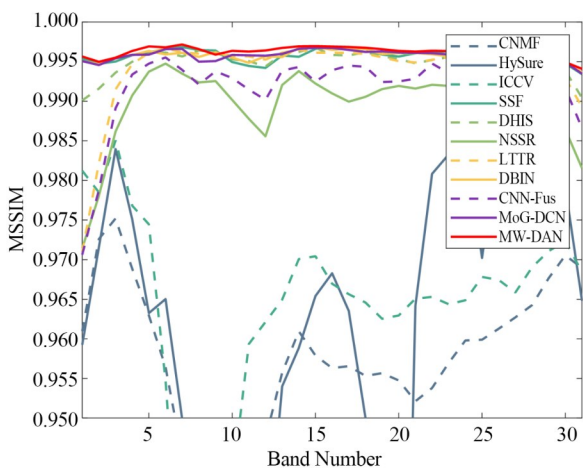


(a) CAVE 数据集的 real and fake apples

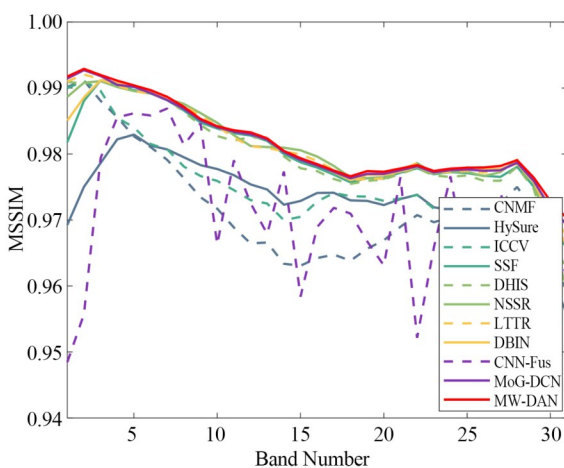


(b) Harvard 数据集的 imge1

图 10 各个波段段的 MPSNR 曲线比较



(a) CAVE 数据集的 real and fake apples



(b) Harvard 数据集的 imge1

图 11 各个波段段的 MSSIM 曲线比较

为了进一步比较不同融合结果的光谱失真,本文使用光谱角映射来度量光谱重建的好坏,并在图 12 和

图 13 中展示光谱角映射函数计算结果的可视化热力图. 图中蓝色表示光谱畸变少,红色表示光谱严重畸变.

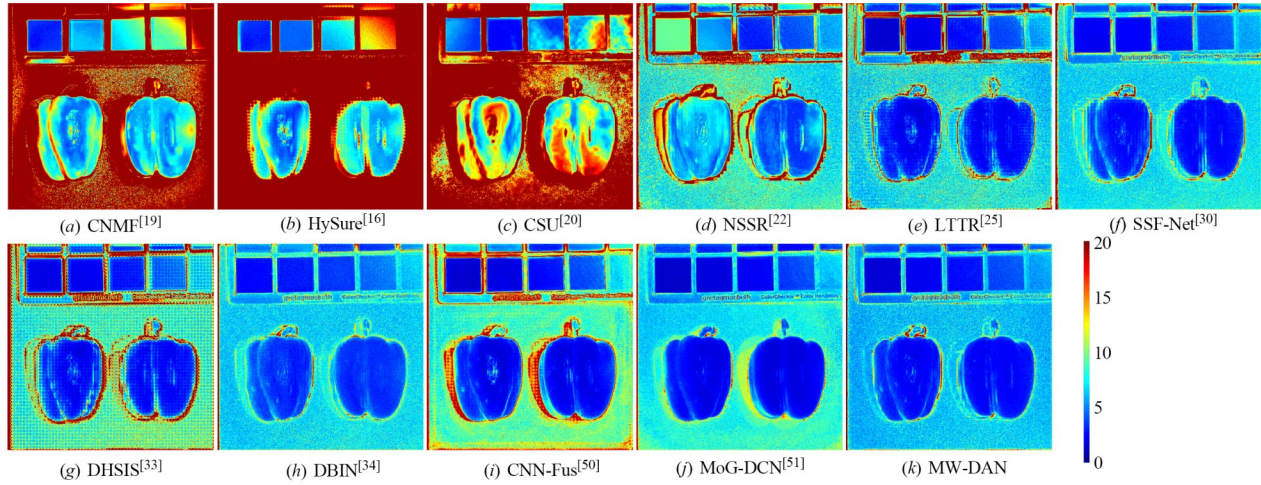


图 12 Real and fake apples 的融合图像比较方法的光谱角映射图像

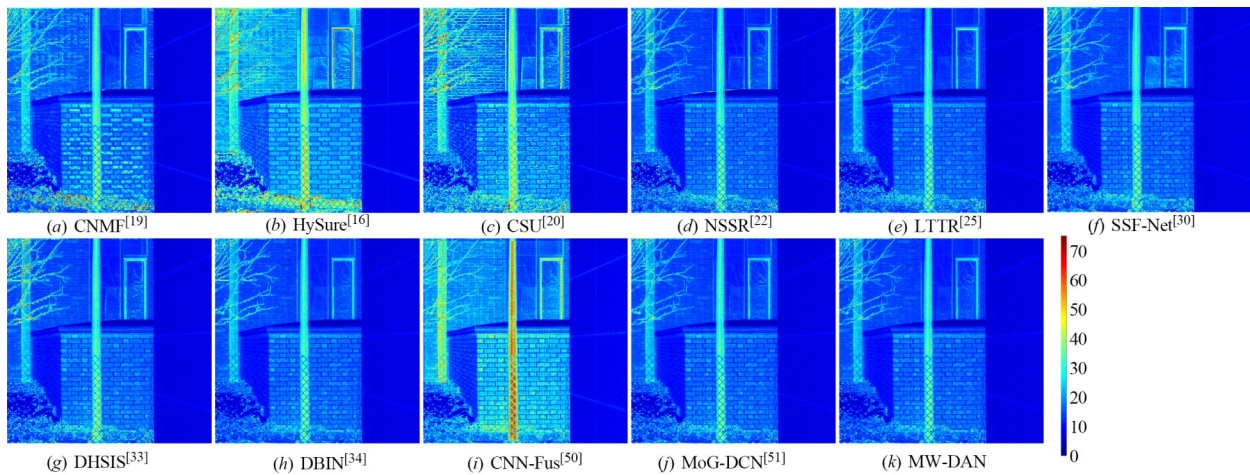


图 13 Imgc1 的融合图像比较方法光谱角映射图像

从图 12 中可以看出, CNMF, HySure 和 CSU 方法融合的结果在背景、方框和文字上存在严重的光谱畸变. NSSR, LTTR, SSF-Net, DHSIS 和 CNN-Fus, 这些方法在方框的边缘、苹果的边缘和文字上有明显的光谱畸变. DBIN 和 MoG-DCN 方法只在方框的边缘和苹果边缘有少量的光谱畸变. 本文方法融合的图像拥有最少的光谱畸变. 图 14(a) 给出了参考图像和融合结果的 SAM 值, 并按照像素升序的方式给出各像素 SAM 的曲线. 从图中可以看出, 本文方法融合的图像光谱畸变少于其他方法.

从图 13 中可以看出, CNMF, HySure, CSU 和 CNN-Fus, 这些方法的融合结果在柱子和枝条上存在严重光谱畸变. NSSR, LTTR, SSF-Net, DHSIS, DBIN 和 MoG-DCN, 这些方法融合的图像只在枝条上有明显的光谱

畸变. 本文方法融合的图像具有最少的光谱畸变. 图 14(b) 给出了参考图像和融合结果的 SAM 值, 并按照像素升序的方式给出各像素 SAM 的曲线. 从图中可以看出, 本文方法融合的图像光谱畸变最少.

3.5 参数分析与消融实验

本文全面分析 CAVE 数据集上的 MW-DAN 模型, 并进行了广泛的消融实验.

(1) DRA 嵌入方案. 在本文的框架中, 输入 HR-MSI 的空间细节信息逐步注入网络, 以重建 HR-HSI. 为了验证这种渐进方式的优势, 本文用不同数量的模块训练 MW-DAN. 表 3 中加粗数据表示最优结果, 从表 3 可以看出, 随着模块数的增加, 重建质量逐渐提高, 增长率越来越小. 实验表明, 2 个深度聚合模块即能达到较好的融合性能. 更深的网络层能够提取更高级特征,

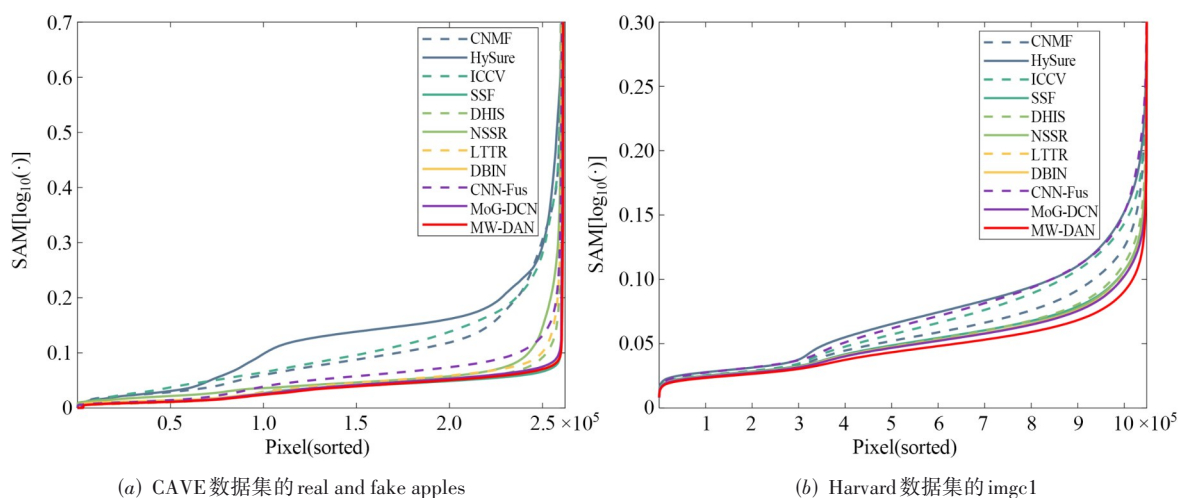


图 14 像素排序的 SAM 曲线比较

表 3 不同 DRA 模块数量网络的比较

模块的数量	MPSNR	RMSE	ERGAS	SAM	UIQI	MSSIM
$n=1$	46.804 0	1.422 9	0.628 9	3.508 3	0.931 8	0.992 4
$n=2$	46.864 4	1.388 9	0.617 2	3.393 3	0.933 1	0.993 1
$n=3$	46.739 4	1.413 7	0.632 1	3.382 1	0.932 4	0.993 0
$n=4$	46.439 4	1.444 2	0.646 1	3.540 5	0.928 7	0.992 5
$n=5$	46.479 2	1.449 1	0.636 5	3.457 6	0.931 4	0.992 9

同时需要更深的解码层完成特征融合与重建. 但当样本量不足时, 较深的网络容易过拟合, 进而导致性能退化. 由于实验条件有限, 目前还不足以实施相关实验. 在未来的工作中, 我们将扩大的样本量.

(2) 损失函数的选择. 本文使用 L1 范数作为损失函数, 也研究了其他的损失函数, 例如 L2 范数. 实验结果如表 4 所示(加粗数据表示最优结果), 它验证了前文的分析, L1 范数相比 L2 范数更有效.

表 4 损失函数对重建质量的研究

损失函数	MPSNR	RMSE	ERGAS	SAM	UIQI	MSSIM
L1	46.864 4	1.388 9	0.617 2	3.393 3	0.933 1	0.993 1
L2	46.391 5	1.448 2	0.646 6	3.652 7	0.927 5	0.992 3

(3) 小波类型的选择. 本文使用 \acute{a} trous 小波作为高频细节提取的工具, 也研究了其他类型的小波来提取高频细节, 例如 Haar 小波. 实验结果如表 5 所示(加粗数据表示最优结果). 实验结果表明, \acute{a} trous 小波相比 Haar 小波更有效.

(4) 跳层聚合连接模块的消融. 本文使用跳层聚合连接模块, 实现空间细节和光谱信息实现互补聚合, 因此本文对比了没有跳层聚合连接模块的实验. 实验结果如表 6 所示(加粗数据表示最优结果). 实验结果表

表 5 \acute{a} trous 与 Haar 小波变换的对比

小波变换的类型	MPSNR	RMSE	ERGAS	SAM	UIQI	MSSIM
MW-DAN- \acute{a} trous	46.864 4	1.388 9	0.617 2	3.393 3	0.933 1	0.993 1
MW-DAN-Haar	46.433 3	1.467 8	0.651 1	3.481 4	0.931 0	0.992 8

表 6 跳层聚合连接模块的消融

有无跳层连接模块	MPSNR	RMSE	ERGAS	SAM	UIQI	MSSIM
有	46.864 4	1.388 9	0.617 2	3.393 3	0.933 1	0.993 1
没有	46.128 3	1.487 7	0.659 3	3.640 8	0.928 6	0.992 1

明, 本文使用的跳层聚合连接模块对融合性能有明显的提升.

(5) 网络参数规模分析. 从表 7 可以看出, SSF-CNN 与本文方法相比参数量很少, 但是它与本文提出的网络的性能相去甚远. DHSIS 和 CNN-Fus 方法的参数量与本文的方法即使很接近, 然而这些方法的神经网络的作用是提取深度先验, 并不是端到端的网络. 本文方法在取得很好的融合效果情况下, 参数量相比最新的端到端的网络 DBIN 和 MoG-DCN 的参数量有大幅度的降低. 因此, 本文将小波方向子带逐层注入空间高频信息, 减少了网络的参数规模.

表 7 网络参数规模分析

CNMF	HySure	CSU	NSSR	LTTR	SSF-CNN	DHSIS	DBIN	CNN-Fus	MoG-DCN	MW-DAN
—	—	—	—	—	0.07 M	0.56 M	3.40 M	0.52 M	7.07 M	0.59 M

3.6 真实数据融合结果

为了证明本文提出的网络在真实数据集上也具有优越性,下面将使用真实数据集进行对比实验. LR-HSI 利用 Eo-1 卫星的 Hyperion 传感器获取. HR-MSI 利用 Sentinel-2 卫星采集. 高光谱图像 Hyperion 的光谱范围为 400~2 500 nm,包括 242 个波段,空间分辨率为 30 m. 高光谱图像在除去水汽波段和噪声波段之后,保留了 89 个波段. 多光谱图像 S2 总共有 13 个波段,从中选择 490 nm, 560 nm, 665 nm 和 842 nm 这 4 个波段作为融合的 HR-MSI,空间分辨率为 10 m.

本文目标是将 30 m HSI 数据和 10 m MSI 数据融合获得 10 m HSI. 由于本文使用的网络是有监督学习的,因此需要 10 m HSI 数据作为参考图像,然而真实场景中没有 10 m HSI 数据. 本文使用文献[52, 53]中的策略,将 30 m HSI 和 10 m MSI 通过下采样变成 90 m HSI 和 30 m MSI,原始的 30 m HSI 作为参考图像参与训练. 在测试阶段,使用原始的 30 m HSI 和 10 m MSI 获得 10 m HSI 数据. 实验中,Hyperion HSI 大小为 2350 × 990 像素, S2 MSI 数据大小为 7050 × 2970 像素,首先取 200 × 200 的 HSI 和 600 × 600 的 MSI 数据作为测试集,剩余图像部分作为训练数据集,测试图像如图 15 所示. 本文使用取块的方式采集训练数据集,每个 LR-HSI, HR-MSI 和 HR-HSI 图像块的尺寸为 4 × 4 × 89, 12 × 12 × 4 和 12 × 12 × 89.



(a) 多光谱图像的伪彩色图像(波段 3, 2, 1) (b) 高光谱图像的伪彩色图像(波段 20, 5, 3)

图 15 Hyperion-sentinel 数据

从图 16 中可以看出, CNMF 方法生成的融合图像中不仅含有大量的噪声,而且在河流部分有大量的光谱失真. HySure 方法生成的图像中不仅绿色部分发生了光谱失真,而且大部分白色区域也已经消失,其他区域也存在着光谱失真. CSU 方法生成的融合图像中绿色部分畸变成灰色了. NSSR 方法生成的图像中边缘有模糊,白色区域丢失. LTTR 方法融合的图片中有条状噪声. SSF-Net 方法融合的图片虽然保留了大量的原始信息,但是图像边缘存在模糊. DHSIS 和 CNN-Fus 方法融合的结果中不但白色区域发生了光谱失真,而且图像中有条状的噪声. DBIN 方法融合的图片中存在大量噪声. MoG-DCN 方法融合的图片中存在少量的噪声.

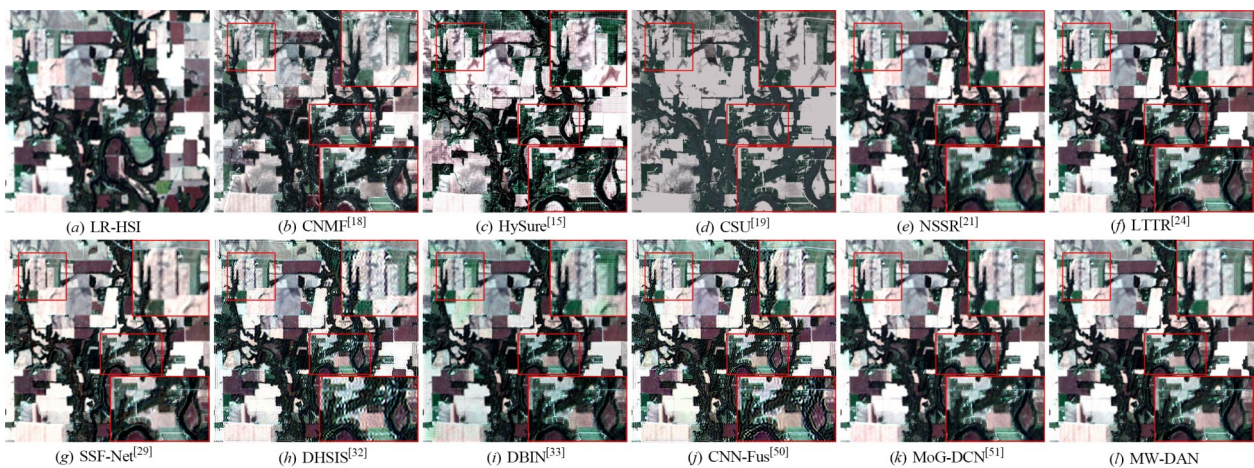


图 16 Hyperion-sentinel 融合结果的伪彩色合成图((波段 20, 5, 3);分辨率 10;尺寸 600 × 600)

Hyperion 和 S2 数据集采集的时间相隔一个月,因此图上的地物会发生变化,这是导致融合结果发生光谱失真的一个因素. 相比其他方法,本文的 MW-DAN 方法得出的融合结果拥有较少的光谱失真.

由于在真实场景中没有参考图像,因此需要参考图像来评价融合质量的方法无法使用. 文献[54]提出学习多变量高斯分布模型(distribution of the Learned MultiVariate Gaussian model, LMVG)评价方

法. 该方法首先利用大量其他场景的高光谱图像作为训练集, 然后在训练集图像和重构图像上采用多变量高斯分布模型提取特征, 最后将训练集和重构图像对应的特征间的距离作为高光谱图像质量畸变度量. LMVG 方法可以很好地评价重建图像光谱

和空间恢复的好坏, 因此本文采用该方法来评价 Hyperion-sentinel 融合结果. LMVG 方法得出的分数越低, 表明图像畸变越少、恢复图像质量越好. 从表 8 展示的结果可以看出, 本文方法具有最小的 LMVG 畸变度量值.

表 8 Hyperion-sentinel 数据集上测试方法的平均 LMVG 畸变值

方法	CNMF	HySure	CSU	SSF-CNN	DHSIS	NSSR	LTTR	DBIN	CNN-Fus	MoG-DCN	MW-DAN
LMVG	14.114 8	14.953 2	14.123 9	14.120 7	16.864 0	17.466 0	18.152 2	14.601 8	14.698 4	14.251 8	13.554 8

从以上分析可以得出, MW-DAN 网络在真实数据集上拥有较好的空间和光谱重建能力.

4 结论

本文针对深度学习在多光谱图像空间细节信息上还可进一步挖掘的问题, 提出一种多层小波深度聚合网络, 称为 MW-DAN. MW-DAN 由空间细节抽取模块和深度残差聚合模块组成. 空间细节抽取模块对 HR-MSI 图像进行非抽取小波变换, 变换后的图像分为低频子带图像和高频子带图像, 然后将低频子带图像和 LR-HSI 作为网络的输入. 深度残差聚合模块将输入的深度残差特征和提取的高频子带图像进行融合. 为了使残差信息不随网络的加深而丢失, 本文使用跳层聚合的方式将信息传递到后面的网络层. 在仿真和真实数据集上的融合结果表明, 本文方法优于目前最新的方法.

本文利用非抽小波提取多光谱图像的空间细节信息注入到高光谱图像中. 虽然这种方法可以有效地提升高光谱图像的空间细节, 但是这种直接将预处理的小波细节特征加入到融合网络中的方法缺乏一定的灵活性. 因此在未来的工作中, 将考虑进一步利用深度神经网络与小波细节特征相结合来提取多光谱图像的多尺度深度细节信息, 并注入到融合网络, 以进一步提高高光谱图像融合的性能.

参考文献

[1] 潘宗序, 禹晶, 肖创柏, 等. 基于光谱相似性的高光谱图像超分辨率算法[J]. 自动化学报, 2014, 40(12): 2797-2807.
 PAN Z X, YU J, XIAO C B, et al. Spectral similarity-based super resolution for hyperspectral images[J]. Acta Automatica Sinica, 2014, 40(12): 2797-2807. (in Chinese)

[2] 徐金环, 沈煜, 刘鹏飞, 等. 联合核稀疏多元逻辑回归和 TV-L1 错误剔除的高光谱图像分类算法[J]. 电子学报, 2018, 46(1): 175-184.
 XU J H, SHEN Y, LIU P F, et al. Hyperspectral image classification combining kernel sparse multinomial logistic regression and TV-L1 error rejection[J]. Acta Electronica Sinica, 2018, 46(1): 175-184. (in Chinese)

[3] 张号逵, 李映, 姜晔楠. 深度学习在高光谱图像分类领域的研究现状与展望[J]. 自动化学报, 2018, 44(6): 961-977.
 ZHANG H K, LI Y, JIANG Y N. Deep learning for hyperspectral imagery classification: The state of the art and prospects[J]. Acta Automatica Sinica, 2018, 44(6): 961-977. (in Chinese)

[4] YOKOYA N, GROHNfeldt C, CHANUSSOT J. Hyperspectral and multispectral data fusion: A comparative review of the recent literature[J]. IEEE Geoscience and Remote Sensing Magazine, 2017, 5(2): 29-56.

[5] CARPER W R, LILLESAND T M, KIEFER R. The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data[J]. Photogrammetric Engineering and Remote Sensing, 1990, 56: 457-467.

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

[7] AIAZZI B, BARONTI S, SELVA M. Improving component substitution pansharpening through multivariate regression of MS+Pan data[J]. IEEE Transactions on Geoscience and Remote Sensing, 2007, 45(10): 3230-3239.

[8] CHEN Z, PU H Y, WANG B, et al. Fusion of hyperspectral and multispectral images: A novel framework based on generalization of pan-sharpening methods[J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11(8): 1418-1422.

[9] RANCHIN T, WALD L. Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation[J]. Photogrammetric Engineering & Remote Sensing, 2000, 66: 49-61.

[10] SELVA M, AIAZZI B, BUTERA F, et al. Hyper-sharpening: A first approach on SIM-GA data[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(6): 3008-3024.

[11] NUNEZ J, OTAZU X, FORS O, et al. Multiresolution-based image fusion with additive wavelet decomposition [J]. IEEE Transactions on Geoscience and Remote Sensing, 1999, 37(3): 1204-1211.

[12] GOMEZ R B, JAZAERI A, KAFATOS M. Wavelet-based hyperspectral and multispectral image fusion[C]//

- Proceeding SPIE 4383, Geo-Spatial Image and Data Exploitation II. Orlando: SPIE, 2001, 4383: 36-42.
- [13] ZHANG Y F, HE M Y. Multi-spectral and hyperspectral image fusion using 3-D wavelet transform[J]. Journal of Electronics (China), 2007, 24(2): 218-224.
- [14] AKHTAR N, SHAFAIT F, MIAN A. Bayesian sparse representation for hyperspectral image super resolution [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 3631-3640.
- [15] ZHANG Y F, DE BACKER S, SCHEUNDERS P. Noise-resistant wavelet-based Bayesian fusion of multispectral and hyperspectral images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2009, 47(11): 3834-3843.
- [16] SIMÕES M, BIOUCAS-DIAS J, ALMEIDA L B, et al. A convex formulation for hyperspectral image super-resolution via subspace-based regularization[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(6): 3373-3388.
- [17] BUNGERT L, COOMES D A, EHRHARDT M J, et al. Blind image fusion for hyperspectral imaging with the directional total variation[J]. Inverse Problems, 2018, 34(4): 044003.
- [18] KAWAKAMI R, MATSUSHITA Y, WRIGHT J, et al. High-resolution hyperspectral imaging via matrix factorization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2011: 2329-2336.
- [19] 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.
- [20] LANARAS C, BALTSAVIAS E, SCHINDLER K. Hyperspectral super-resolution by coupled spectral unmixing [C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2016: 3586-3594.
- [21] KWON H, TAI Y W. RGB-guided hyperspectral image upsampling[C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2016: 307-315.
- [22] DONG W S, FU F Z, SHI G M, et al. Hyperspectral image super-resolution via non-negative structured sparse representation[J]. IEEE Transactions on Image Processing, 2016, 25(5): 2337-2352.
- [23] LIU J J, WU Z B, XIAO L, et al. A truncated matrix decomposition for hyperspectral image super-resolution[J]. IEEE Transactions on Image Processing, 2020, 29: 8028-8042.
- [24] DIAN R W, FANG L Y, LI S T. Hyperspectral image super-resolution via non-local sparse tensor factorization [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 3862-3871.
- [25] DIAN R W, LI S T, FANG L Y. Learning a low tensor-train rank representation for hyperspectral image super-resolution[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(9): 2672-2683.
- [26] XU Y, WU Z B, CHANUSSOT J, et al. Nonlocal patch tensor sparse representation for hyperspectral image super-resolution[J]. IEEE Transactions on Image Processing, 2019, 28(6): 3034-3047.
- [27] XU Y, WU Z B, CHANUSSOT J, et al. Nonlocal coupled tensor CP decomposition for hyperspectral and multispectral image fusion[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(1): 348-362.
- [28] HE W, CHEN Y, YOKOYA N, et al. Hyperspectral super-resolution via coupled tensor ring factorization[J]. Pattern Recognition, 2022, 122: 108280.
- [29] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [30] HAN X H, SHI B X, ZHENG Y Q. SSF-CNN: Spatial and spectral fusion with CNN for hyperspectral image super-resolution[C]//2018 25th IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2018: 2506-2510.
- [31] CHANG Y, YAN L X, FANG H Z, et al. HSI-DeNet: Hyperspectral image restoration via convolutional neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(2): 667-682.
- [32] YANG J X, ZHAO Y Q, CHAN J. Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network[J]. Remote Sensing, 2018, 10(5): 800.
- [33] DIAN R W, LI S T, GUO A J, et al. Deep hyperspectral image sharpening[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(11): 5345-5355.
- [34] WANG W, ZENG W H, HUANG Y, et al. Deep blind hyperspectral image fusion[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2020: 4149-4158.
- [35] ZHANG L, NIE J T, WEI W, et al. Deep blind hyperspectral image super-resolution[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(6): 2388-2400.
- [36] YANG J X, XIAO L, ZHAO Y Q, et al. Variational regularization network with attentive deep prior for hyperspectral-multispectral image fusion[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-17.
- [37] QU Y, QI H R, KWAN C. Unsupervised sparse dirichlet-net for hyperspectral image super-resolution[C]//2018

- IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 2511-2520.
- [38] 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.
- [39] HAN X H, CHEN Y W. Deep residual network of spectral and spatial fusion for hyperspectral image super-resolution[C]//2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM). Piscataway: IEEE, 2019: 266-270.
- [40] XIAO J J, YUAN Q Q, LI J, et al. Enhanced residual dense network joint with GRUS for multispectral and hyperspectral image fusion[C]//2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. Piscataway: IEEE, 2021: 2448-2451.
- [41] LIU J, ZHANG W J, TANG Y T, et al. Residual feature aggregation network for image super-resolution[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 2356-2365.
- [42] 王相海, 赵晓阳, 毕晓昀, 等. 小波域多角度轮廓模板变分模型的单幅图像超分辨率重建[J]. 电子学报, 2018, 46(9): 2256-2262.
WANG X H, ZHAO X Y, BI X Y, et al. Single image super-resolution reconstruction approach based on multi-angle contour templates variational calculus model in wavelet domain[J]. Acta Electronica Sinica, 2018, 46(9): 2256-2262. (in Chinese)
- [43] STARCK J L, FADILI J, MURTAGH F. The undecimated wavelet decomposition and its reconstruction[J]. IEEE Transactions on Image Processing, 2007, 16(2): 297-309.
- [44] 王相海, 赵晓阳, 王鑫莹, 等. 非抽取小波边缘学习深度残差网络的单幅图像超分辨率重建[J]. 电子学报, 2022, 50(7): 1753-1765.
WANG X H, ZHAO X Y, WANG X Y, et al. Single image super-resolution reconstruction using deep residual networks with non-decimated wavelet edge learning[J]. Acta Electronica Sinica, 2022, 50(7): 1753-1765. (in Chinese)
- [45] SHENSA M J. The discrete wavelet transform: Wedding the a trous and Mallat algorithms[C]//IEEE Transactions on Signal Processing. Piscataway: IEEE, 2002: 2464-2482.
- [46] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[J]. IEEE Transactions on Signal Processing, 1992, 61(3): 211-231.
- [47] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2017: 1132-1140.
- [48] TAI Y, YANG J, LIU X, et al. Memnet: A persistent memory network for image restoration[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 4549-4557.
- [49] KINGMA D P, BA J. Adam: A method for stochastic optimization[EB/OL]. (2014-12-22) [2022-07-08]. <https://arxiv.org/abs/1412.6980>.
- [50] DIAN R W, LI S T, KANG X D. Regularizing hyperspectral and multispectral image fusion by CNN denoiser[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(3): 1124-1135.
- [51] DONG W S, ZHOU C, WU F F, et al. Model-guided deep hyperspectral image super-resolution[J]. IEEE Transactions on Image Processing, 2021, 30: 5754-5768.
- [52] XIE Q, ZHOU M H, ZHAO Q, et al. MHF-net: An interpretable deep network for multispectral and hyperspectral image fusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(3): 1457-1473.
- [53] SCARPA G, VITALE S, COZZOLINO D. Target-adaptive CNN-based pansharpening[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(9): 5443-5457.
- [54] YANG J X, ZHAO Y Q, YI C, et al. No-reference hyperspectral image quality assessment via quality-sensitive features learning[J]. Remote Sensing, 2017, 9(4): 305.

作者简介



方健男, 1994年11月生, 安徽安庆人. 南京理工大学计算机科学与工程学院硕博连读生. 主要研究方向为遥感图像处理.
E-mail: fangjian@njust.edu.cn



杨劲翔男, 1988年3月生, 安徽滁州人. 南京理工大学计算机科学与工程学院讲师. 主要研究方向为遥感图像处理、机器学习与模式识别.
E-mail: yang123jx@njust.edu.cn



肖亮男, 1972年2月生, 湖南宁乡人. 南京理工大学计算机科学与工程学院教授、博士生导师, 江苏省“光谱成像与智能感知”重点实验室和教育部“高维信息智能感知与系统”重点实验室副主任. 主要研究方向为图像处理与计算机视觉、机器学习与模式识别.
E-mail: xiaoliang@mail.njust.edu.cn