

基于深度展开和双流网络的高光谱图像融合

刘丛, 姚佳浩

(上海理工大学光电信息与计算机工程学院, 上海 200093)

摘要: 针对基于深度学习的高光谱图像融合算法通常堆积多个卷积以学习映射关系、没有充分利用问题的特性以及缺乏可解释性等问题, 提出一种结合深度展开与双流网络的深度网络。首先使用卷积稀疏编码建立融合模型, 该模型将低分辨率高光谱图像(Low-resolution hyperspectral images, LR-HSI)和高分辨率多光谱图像(high-resolution multispectral images, HR-MSI)映射到低维子空间中。在融合模型设计中, 考虑了LR-HSI和HR-MSI的共有信息以及LR-HSI的独有信息, 并将HR-MSI作为辅助信息加入模型中。其次将该融合模型展开为可学习的可解释深度网络。最后, 使用双流网络获取更精确的高分辨率高光谱图像(High-resolution hyperspectral images, HR-HSI)。实验表明, 该网络在高光谱图像融合中可以获得出色的效果。

关键词: 高光谱图像融合; 卷积稀疏编码; 深度展开网络; 双流网络; 深度学习

中图分类号: TP391 **文献标志码:** A

Hyperspectral Image Fusion via Deep Unfolding and Dual-stream Networks

LIU Cong, YAO Jiahao

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Hyperspectral image fusion algorithms based on deep learning typically stack multiple convolutional layers to learn mapping relationships, which suffer from the problems of not fully utilizing the characteristics of the task and lack of interpretability. To address these problems, this paper proposes a deep network combining deep unfolding and dual-stream networks. Firstly, an image fusion model is established using convolutional sparse coding, which maps low-resolution hyperspectral images (LR-HSI) and high-resolution multispectral images (HR-MSI) into a low-dimensional subspace. In the design of the fusion model, we consider the common information of LR-HSI and HR-MSI as well as the unique information of LR-HSI, and add HR-MSI to the model as auxiliary information. Next, the fusion model is unfolded into a learnable interpretable deep network. Finally, the dual-stream network is used to get more accurate high-resolution hyperspectral images (HR-HSI). Experiments prove that the network obtains excellent results in the hyperspectral image fusion task.

Key words: hyperspectral image fusion; convolutional sparse coding; deep unfolding networks; dual-stream networks; deep learning

引言

高光谱图像是一种通过收集同一场景的不同光谱反射信息而形成的高维图像。与灰度图像或RGB彩色图像等低光谱图像相比,高光谱图像含有更为丰富的场景及物质信息,可以更好地对不同地物特征进行识别和检测,因此其广泛地应用于图像识别、图像分类、图像分割、目标检测及光谱分离等众多领域^[1-5]。然而,由于传感器硬件的性能限制,高光谱图像在成像过程中很难同时保证较高的空间分辨率和光谱分辨率。部分研究者为了获取较高的光谱信息,选择牺牲空间分辨率,从而获取一幅低分辨率高光谱图像(Low-resolution hyperspectral images, LR-HSI)。而另外一部分研究者为了获取较高的空间分辨率,选择牺牲光谱信息,从而获取一幅高分辨率多光谱图像(High-resolution multispectral images, HR-MSI)。因此,将HR-MSI和LR-HSI融合以生成期望的高分辨率高光谱图像(High-resolution hyperspectral images, HR-HSI)是一种比较经济可行的方法,受到研究者的广泛关注。

近年来,各种不同的高光谱图像融合算法被相继提出,可将其划分为变换域方法、矩阵分解方法、张量分解方法以及深度学习方法。早期的高光谱图像融合方法以变换域融合为主,该方法将LR-HSI和HR-MSI映射到变换域中融合,然后使用反变换获取HR-HSI,如主成分分析(Principal component analysis, PCA)^[6]或小波变换^[7]。矩阵和张量分解方法通常将LR-HSI和HR-MSI分解为字典和系数的形式,并加入各种正则化模型以挖掘图像的先验信息。文献[8]提出贝叶斯稀疏编码策略,该算法将LR-HSI和HR-MSI分解为贝叶斯字典和稀疏编码以建立高光谱融合模型。文献[9]提出非负结构稀疏表示(Non-negative structured sparse representation, NSSR)模型,该模型首先提出块坐标下降算法求解字典,其次设计基于聚类的模型来寻找同类子图像的相似性。文献[10]使用空间和光谱的双字典保存空间信息和光谱信息,并使用 l_1 范数保证系数的稀疏性。文献[11]将LR-HSI和HR-MSI映射到高维向量空间中,使用Tensor Ring分解完成融合操作。文献[12]提出非局部耦合张量CP分解(Nonlocal coupled tensor CANDECOMP/PARAFAC decomposition, NCTCP)模型,该模型将非局部自相似性加入到融合模型中,使用CP分解对图像进行融合。文献[13]提出低阶张量序列分解(Low tensor-train rank, LTTR)模型,该模型将高光谱图像划分为多个子图像,然后将相似的子图像划分到同一类中以形成一个四维图像,并使用Tensor-Rrain算法分解该四维图像以挖掘核向量的低秩约束。文献[14]提出张量稀疏分解的半盲融合算法,其使用张量Tucker分解将高光谱图像分解为1个核和3个方向的字典,并对核张量使用 l_1 范数保证系数的稀疏性。文献[15]首先使用字典学习对图像分解,并将图像自相似性作为先验信息加入到融合模型中。文献[16]提出了自适应非负稀疏表示(Adaptive nonnegative sparse representation, ANSR)模型,引入了自适应稀疏表示方法来优化非负结构稀疏模型。通过平衡系数的稀疏性和相关性,作者在 l_1 范数和 l_2 范数之间的约束下,自适应地得到更加精确的稀疏系数。文献[17]提出基于图正则化的低秩张量分解融合算法,分别使用空间图和光谱图将原始图像映射到低维空间中以去除冗余信息,并且对不同方向字典加入低秩约束以更好地发现空间和光谱间的相关性。文献[18]使用截断奇异值分解(Truncated singular value decomposition, TSVD)将高光谱图像分解为张量字典和系数,然后对图像分组,每组学习一个紧字典,并使用 l_1 范数约束系数的稀疏性。文献[19]将自适应稀疏表示引入非负结构稀疏模型中,通过平衡系数的稀疏性和相关性,在 l_1 范数和 l_2 范数之间的约束下自适应得到更精确的稀疏系数。尽管上述方法已经取得了良好的融合效果,但大多数先验模型都是人工设计的,无法实现端到端的优化,也无法学习潜在的映射关系导致性能有限。

近年来深度卷积神经网络(Deep convolutional neural network, DCNN)因其强大的学习能力被广泛地应用于图像处理任务中,其在高光谱融合领域也获得了比较好的效果。文献[20]提出了一种基于3D卷积的高光谱融合网络,并使用主成分分析降低时间复杂度。文献[21]提出一种深度渐进零心残差学习高光谱图像融合网络,该网络使用零心残差学习特征信息并获得比较好的融合效果。文献[22]提

出 SSR-Net 网络,该网络通过融合空间差分与光谱差分等建立新的优化函数。文献[23]提出了 UMAG-Net 网络,该网络不需要训练样本并使用编码解码器和两个注意力充分挖掘有效特征。然而,上述深度学习方法类似于“黑箱”机制,忽略了图像原有的观察模型和明显的先验结构,不仅缺乏可解释性,而且对训练数据的变化非常敏感。为了解决该问题,研究者提出深度展开的方法,将基于模型的方法与深度学习方法相结合,避免了基于模型方法的手工先验设置和计算复杂度大的问题,也解决了深度学习缺乏可解释性的问题。文献[24]提出模型引导的深度卷积网络(Model-guided deep convolutional network, MoG-DCN),将模型的迭代过程展开成网络的形式并巧妙地利用降噪网络求解正则项。文献[25]首先使用融合模型求解一个初值,然后将该初值作为深度网络的输入。文献[26]设计了变分融合模型并展开为网络,此外还将非局部自相似性嵌入到网络中。文献[27]提出 CNN-Fus 算法,该算法将融合模型分解为保真项和正则项两个部分,前者使用传统的迭代优化求解,后者可被视为图像去噪问题,使用 CNN 去噪器对其进行求解。文献[28]将融合模型分成 3 个子问题,使用深度网络对先验子问题深度展开。文献[29]提出 MHF-Net 网络,利用低分辨率观测模型和低秩信息构建为融合模型,并将其展开成深度卷积网络。文献[30]提出 CU-Net 网络并将其应用于自然图像融合中,其使用卷积稀疏编码将不同模态的图像分别分解为共有信息以及每个模态的独有信息两部分,在图像融合阶段结合共有信息和独有信息来重建融合图像,并将该模型展开为深度网络的形式,所以具有很好的可解释性。该网络在多模态融合中取得了较好的效果,但其只能对尺寸相同的两个自然融合,当面对高光谱融合问题时,很难直接对不同维度的 LR-HSI 和 HR-MSI 直接融合。文献[31]提出学习的多模式卷积稀疏编码(Learned multimodal convolutional sparse coding, LMCS)模型,该模型将多模态融合问题看作引导图像超分辨率问题,设置含有 l_1 - l_1 范数的超分辨率模型,并将其展开为网络的形式。

为了获取更好的高光谱图像融合效果,本文提出一种基于深度展开和双流网络的图像融合网络(Deep unfolding and dual stream fusion network, DUDS-FusionNet)。该网络分为深度展开先验网络和双流融合网络两部分。在深度展开先验网络中,使用卷积稀疏编码设计融合模型,并使用深度展开原理将该融合模型的求解迭代过程展开为深度网络。在设计卷积稀疏编码融合模型时,首先借鉴 CU-Net 网络的思想,将 LR-HSI 分解成共有信息和独有信息两部分。其次借鉴 LMCS 模型的思想,将 HR-MSI 分解为卷积字典和卷积稀疏表示,并将该卷积稀疏表示作为辅助信息以更好地获取融合图像。再者,针对 CU-Net 网络和 LMCS 模型无法直接融合不同波段的高光谱图像的问题,将三维稀疏字典堆积成可表示高光谱图像的四维稀疏字典。通过该深度展开先验网络,可获得带有先验信息的高光谱融合图像。为了充分挖掘 HR-MSI 的结构信息,将其与深度展开网络得到的先验信息输入到双流融合网络中以获取空间精度更高的高光谱融合图像。实验结果表明,本文所提出的方法在高光谱图像融合任务中获得了出色的效果。

1 相关工作

本节将对本文中用到的技术做简要介绍,包括稀疏编码、卷积稀疏编码、基于辅助设计的卷积稀疏编码以及模型展开技术。

1.1 稀疏编码

1.1.1 传统稀疏编码

给定一个 d 维的输入信号 $\mathbf{y} \in \mathbf{R}^d$, 可将其逼近为一组字典的线性组合形式 $\mathbf{y} \approx \mathbf{D}\boldsymbol{\alpha}$, 其中 $\mathbf{D} \in \mathbf{R}^{d \times m}$ 表示稀疏字典, m 表示字典中原子的个数, $\boldsymbol{\alpha} \in \mathbf{R}^m$ 表示 \mathbf{y} 在字典 \mathbf{D} 上的稀疏表示。为了保证系数的稀疏性,该问题可描述为

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \quad \text{s.t. } \mathbf{y} \approx \mathbf{D}\boldsymbol{\alpha} \quad (1)$$

由于 l_0 范数的最优化问题是一个非凸的并且是非确定性多项式难题,故在实际求解中将其松弛为 l_1 范数,即

$$\min_{\alpha} \|\alpha\|_1 \quad \text{s.t. } \mathbf{y} \approx D\alpha \quad (2)$$

使用拉格朗日乘子法将约束条件转化为带有惩罚项系数的拉格朗日函数求解,即

$$\alpha = \operatorname{argm} \min_{\alpha} \frac{1}{2} \|\mathbf{y} - D\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (3)$$

式中: λ 表示拉格朗日乘子系数; $\|\cdot\|_1$ 表示 l_1 范数。

1.1.2 卷积稀疏编码及其网络展开

由于传统的稀疏编码将图像划分为多个小块且向量化,不可避免地丢失了相邻块和相邻元素之间的结构性。为了避免该问题,文献[32]提出了卷积稀疏编码(Convolutional sparse coding, CSC),该编码是一种全图稀疏编码方式,其使用卷积直接对原始图像进行运算,既不需要分块也不需要向量化,避免了传统稀疏编码丢失相邻块和相邻元素结构性的问题。给定输入图像 $\mathbf{y} \in \mathbf{R}^{n \times n}$,使用卷积稀疏编码可将其分解为卷积字典和稀疏特征,可描述为

$$\min_{\{\alpha_i\}} \frac{1}{2} \left\| \mathbf{y} - \sum_{i=1}^m d_i * \alpha_i \right\|_2^2 + \lambda \sum_{i=1}^m \|\alpha_i\|_1 \quad (4)$$

式中: $d_i \in \mathbf{R}^{r \times r}$ 和 $\alpha_i \in \mathbf{R}^{n \times n}$ 分别表示卷积字典的原子和对应的稀疏表示,*表示卷积运算。针对式(4)的优化方法主要包括快速傅里叶法、直接优化方法^[33]和可学习卷积稀疏编码(Learning convolutional sparse coding, LCSC)算法^[34]。由于卷积运算是线性的,可以构造一个托普利兹矩阵 $D_i \in \mathbf{R}^{n^2 \times n^2}$,并将 α_i 向量化为 $\alpha'_i \in \mathbf{R}^{n^2}$,使 $d_i * \alpha_i = D_i \alpha'_i$ 。可将式(4)转换为

$$\min_{\{\alpha'_i\}} \frac{1}{2} \left\| \mathbf{y}' - \sum_{i=1}^m D_i \alpha'_i \right\|_2^2 + \lambda \sum_{i=1}^m \|\alpha'_i\|_1 \quad (5)$$

式中: $\mathbf{y}' \in \mathbf{R}^{n^2}$ 为向量化的 \mathbf{y} ;构造 $D \in \mathbf{R}^{n^2 \times mn^2} = \{D_1, D_2, \dots, D_m\}$; $\alpha \in \mathbf{R}^{mn^2} = \{\alpha'_1, \alpha'_2, \dots, \alpha'_m\}$,可将式(5)转化为传统的稀疏编码问题

$$\min_{\alpha} \frac{1}{2} \|\mathbf{y}' - D\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (6)$$

式(6)可以使用迭代收缩阈值算法(Iterative shrinking threshold algorithm, ISTA)^[35]求解,得到如下迭代解

$$\alpha^{(t+1)} = \mathcal{S}_{\lambda} \left(\alpha^{(t)} + \frac{1}{\mu} D^T (\mathbf{y}' - D\alpha^{(t)}) \right) \quad (7)$$

式中: μ 为梯度下降的迭代步长; $\mathcal{S}_{\lambda}(x) = \operatorname{sign}(x) \max(|x| - \lambda, 0)$ 为软阈值函数,如图1所示。由于卷积稀疏表示对整个图像操作,产生的矩阵过大,直接对式(7)求解无法实现。文献[34]使用LCSC将式中的矩阵乘法替换为卷积运算,可得

$$\alpha^{(t+1)} = \mathcal{S}_{\lambda} (\alpha^{(t)} - E * F * \alpha^{(t)} + E * \mathbf{y}) \quad (8)$$

式中 $E \in \mathbf{R}^{r \times r \times m}$ 、 $F \in \mathbf{R}^{r \times r \times m}$ 分别是实现Toeplitz矩阵 D^T 和 D 的滤波器。在LCSC中, E 和 F 作为深度网络中的卷积层是可训练的。

1.2 结合辅助信息的稀疏编码

根据最近的研究^[36],如果存在与目标图像 \mathbf{y} 相关的图像 ω ,则可

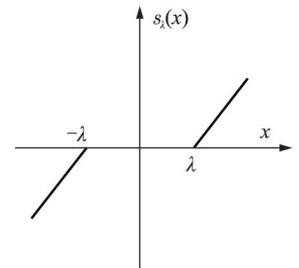


图1 软阈值函数

Fig.1 Soft thresholding function

以将相关图像作为辅助信息进一步提高稀疏约束问题的准确性。假定 $\mathbf{y} \in \mathbf{R}^n$ 和 $\boldsymbol{\omega} \in \mathbf{R}^d$ 具有相似的稀疏表示 $\boldsymbol{\alpha} \in \mathbf{R}^m$, $\mathbf{s} \in \mathbf{R}^m$, 分别对应于字典 $\mathbf{D}_y \in \mathbf{R}^{n \times m}$, $\mathbf{D}_\omega \in \mathbf{R}^{d \times m}$, $n \leq m$, $d \leq m$ 。 $\boldsymbol{\alpha}$ 不仅可以重建目标图像, 而且应该与 \mathbf{s} 相似。为了利用该辅助信息, 可以在模型中插入额外的约束项来创建结合辅助信息的 CSC 模型。稀疏表示系数 $\boldsymbol{\alpha}$ 可以求解以下 l_1 - l_1 最优化问题得到

$$\min_{\boldsymbol{\alpha}} \frac{1}{2} \|\mathbf{y} - \mathbf{D}_y \boldsymbol{\alpha}\|_2^2 + \lambda (\|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\alpha} - \mathbf{s}\|_1) \quad (9)$$

此问题已经在文献[36]中进行了理论研究, 文献[37]中介绍了求解问题的数值方法, 可使用分段迭代软阈值算法求解

$$\boldsymbol{\alpha}^{(t+1)} = \mathcal{P}_\lambda \left(\boldsymbol{\alpha}^{(t)} - \frac{1}{\mu} \mathbf{D}_y^T (\mathbf{D}_y \boldsymbol{\alpha}^{(t)} - \mathbf{y}); \mathbf{s} \right) \quad (10)$$

式中分段软阈值函数 $\mathcal{P}_\lambda(\mathbf{v}; \mathbf{s})$ 如图 2 所示, 定义如下:

(1) 如果 $s_i \geq 0$

$$\mathcal{P}_\lambda(v_i; s_i) = \begin{cases} v_i + 2\lambda & v_i < -2\lambda \\ 0 & -2\lambda \leq v_i \leq 0 \\ v_i & 0 < v_i < s_i \\ s_i & s_i \leq v_i \leq s_i + 2\lambda \\ v_i - 2\lambda & v_i \geq s_i + 2\lambda \end{cases} \quad (11)$$

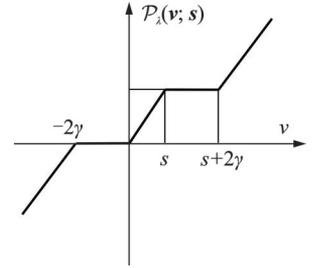


图 2 分段软阈值函数

Fig.2 Piece-wise soft thresholding function

(2) 如果 $s_i < 0$

$$\mathcal{P}_\lambda(v_i; s_i) = \begin{cases} v_i + 2\lambda & v_i < s_i - 2\lambda \\ s_i & s_i - 2\lambda \leq v_i \leq s_i \\ v_i & s_i < v_i < 0 \\ 0 & 0 \leq v_i \leq 2\lambda \\ v_i - 2\lambda & v_i \geq 2\lambda \end{cases} \quad (12)$$

同理, 使用 LCSC 将式(10)中的矩阵乘法替换为卷积运算, 可得

$$\boldsymbol{\alpha}^{(t+1)} = \mathcal{P}_\lambda \left(\boldsymbol{\alpha}^{(t)} - \mathbf{E} * \mathbf{F} * \boldsymbol{\alpha}^{(t)} + \mathbf{E} * \mathbf{y}; \mathbf{s} \right) \quad (13)$$

2 本文方法

图 3 描述了本文提出的 DUDS-FusionNet 整体框架。如图 3 所示, 该网络包括两个主要部分: 基于 CSC 的深度展开网络和双流融合网络 DSFN, 图中 UniqueIEM、AuxiliaryIEM 和 CommonIEM 分别指下

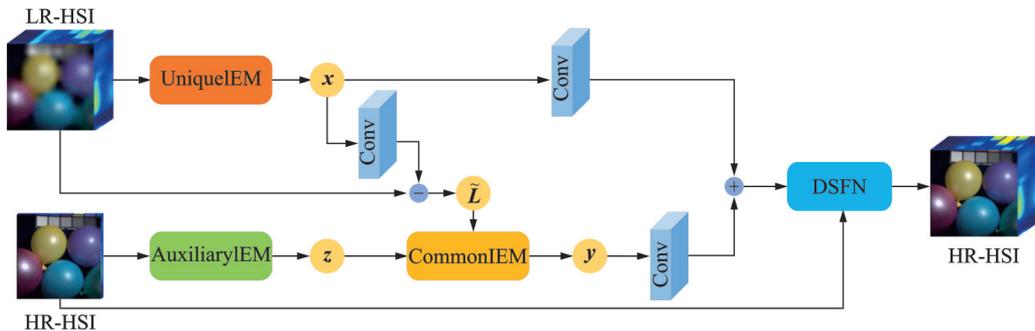


图 3 DUDS-FusionNet 网络整体架构

Fig.3 Overall architecture of DUDS-FusionNet

文的独有信息提取模块,辅助信息提取模块和共有信息提取模块。在基于CSC的深度展开网络中,首先,HR-MSI和LR-HSI通过辅助信息提取模块和独有信息提取模块得到 z 和 x 。其次,卷积单元将 x 从特征空间映射到图像空间中,计算残差图像 \tilde{L} 。然后,共有信息提取模块在辅助信息 z 的约束下从残差图像 \tilde{L} 中提取共有信息稀疏表示 y 。根据式(16),使用两个卷积单元实现字典 α 和 β ,并将 x 和 y 映射到图像空间,以获取带有先验信息的HR-HSI。在双流融合网络DSFN中融合带有先验信息的HR-HSI和HR-MSI得到最终的HR-HSI。

2.1 基于CSC的深度展开网络模块

基于带有辅助信息的卷积稀疏编码思想,本文设计了一种全新的高光谱图像融合模型,并将该模型展开为深度展开网络。在模型设计中,为了充分利用HR-MSI的信息,将其图像信息和稀疏系数信息应用于网络设计中。首先提取HR-MSI的卷积稀疏系数,具体实现公式为

$$\min_z \frac{1}{2} \|M - c * z\|_2^2 + \lambda \|z\|_1 \quad (14)$$

式中: $M \in \mathbf{R}^{b \times P \times Q}$ 表示HR-MSI; $c \in \mathbf{R}^{b \times k \times r \times r}$ 表示滤波; $z \in \mathbf{R}^{k \times P \times Q}$ 表示 M 的稀疏表示; b 表示波段数; $P \times Q$ 表示图像的高度和宽度; k 表示滤波的数量; r 表示滤波的大小。

对于LR-HSI,使用双三次插值算法^[38]将其空间尺寸上采样为与目标图像HR-HSI相同,即 $L \in \mathbf{R}^{B \times P \times Q}$, B 表示高光谱图像的波段数。 L 和 M 具有高度的空间相似性,但是 L 每个波段的光谱响应强度与 M 不同。为了描述 L 和 M 之间的关系,本文将 L 分解为两部分信息,分别是与 M 图像不相关的独有信息 $a * x$ 以及相关的共有信息 $b * y$,即 L 可以表示为 $L = a * x + b * y$,其中 a 和 b 是两种信息所对应的字典, x 和 y 分别为独有信息稀疏表示与共有信息稀疏表示。为了融合 L 和 M 的信息,使得 z 和 y 尽可能相似,得到以下模型

$$\min_{x,y} \frac{1}{2} \|L - (a * x + b * y)\|_2^2 + \lambda (\|x\|_1 + \|y\|_1 + \|y - z\|_1) \quad (15)$$

假定LR-HSI和HR-HSI图像共享稀疏表示,式中 y 由辅助信息 z 约束。因此,在已知HR-HSI的字典的情况下,即 α 和 β 已知,HR-HSI H 可以通过求解以下公式得到

$$H = \alpha * x + \beta * y \quad (16)$$

2.1.1 辅助信息提取模块

首先,优化式(14)以通过稀疏HR-MSI来获取辅助信息 z 。式(14)是 l_1 正则化最小二乘问题,根据式(7),可获得迭代公式为

$$z^{(t+1)} = \mathcal{S}_\lambda \left(z^{(t)} - \frac{1}{\mu} c^T * (c * z^{(t)} - M) \right) \quad (17)$$

式中 c^T 是 c 的转置形式。根据上述原理,式(14)可以通过辅助信息提取模块求解,式(17)的计算过程可以转换到深度网络形式,即

$$z^{(t+1)} = \mathcal{S}_\lambda \left(z^{(t)} - E_z * D_z * z^{(t)} + E_z * M \right) \quad (18)$$

式中 c^T 和 c 使用两个可自适应学习的卷积层 $D_z \in \mathbf{R}^{b \times k \times r \times r}$ 和 $E_z \in \mathbf{R}^{k \times b \times r \times r}$ 实现。式(18)可展开为图4所示。该模块包含 T 个迭代块,每个块执行式(18)中的计算流程。由于 $z^{(0)}$ 被初始化为零,第一次迭代模块的计算流程为 $z^{(1)} = \mathcal{S}_\lambda(E_z * M)$ 。

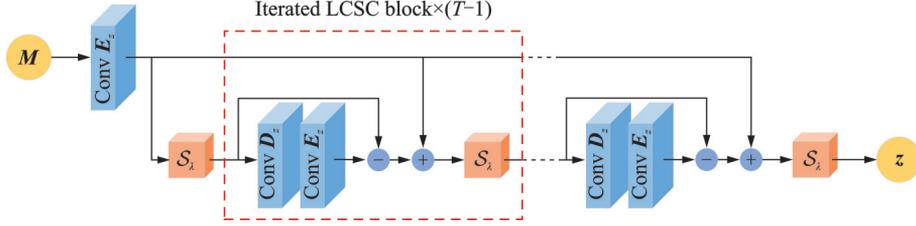


图4 辅助信息提取模块

Fig.4 Auxiliary information extraction module (AuxiliaryIEM)

2.1.2 独有信息提取模块

由于式(15)中有两个未知项,对其分别迭代优化。首先,固定 \mathbf{y} 迭代更新 \mathbf{x} ,式(15)转换为以下问题

$$\min_{\mathbf{x}} \frac{1}{2} \left\| \mathbf{L} - (\mathbf{a} * \mathbf{x} + \mathbf{b} * \mathbf{y}) \right\|_2^2 + \lambda \|\mathbf{x}\|_1 \quad (19)$$

等价代换得到

$$\min_{\mathbf{x}} \frac{1}{2} \left\| \hat{\mathbf{L}} - \mathbf{a} * \mathbf{x} \right\|_2^2 + \lambda \|\mathbf{x}\|_1 \quad (20)$$

式中 $\hat{\mathbf{L}} = \mathbf{L} - \mathbf{b} * \mathbf{y}$ 。使用迭代软阈值算法求解

$$\mathbf{x}^{(t+1)} = \mathcal{S}_\lambda \left(\mathbf{x}^{(t)} - \frac{1}{\mu} \mathbf{a}^T * (\mathbf{a} * \mathbf{x}^{(t)} - \hat{\mathbf{L}}) \right) \quad (21)$$

类似于式(17)和式(18),将此迭代过程转换为独有信息提取模块

$$\mathbf{x}^{(t+1)} = \mathcal{S}_\lambda \left(\mathbf{x}^{(t)} - \mathbf{E}_x * \mathbf{D}_x * \mathbf{x}^{(t)} + \mathbf{E}_x * \hat{\mathbf{L}} \right) \quad (22)$$

独有信息提取模块的结构如图5所示。模块的输入为 $\hat{\mathbf{L}} = \mathbf{L} - \mathbf{b} * \mathbf{y}$ 。在整体网络中,共有信息稀疏表示 \mathbf{y} 通过下节的共有信息提取模块计算得到。因此,初始化为 $\mathbf{y} = 0$,所以独有信息提取模块的输入退化为 $\hat{\mathbf{L}} = \mathbf{L}$ 。

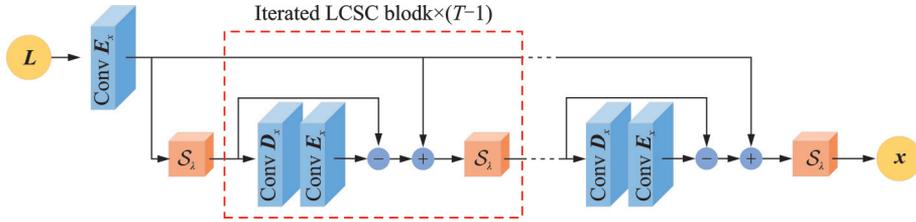


图5 独有信息提取模块

Fig.5 Unique information extraction module (UniqueIEM)

2.1.3 共有信息提取模块

最后,给定辅助信息 \mathbf{z} 和独有信息 \mathbf{x} ,就可以导出共有信息提取模块。固定 \mathbf{x} 迭代更新 \mathbf{y} ,式(15)转换为以下问题

$$\min_{\mathbf{y}} \frac{1}{2} \left\| \tilde{\mathbf{L}} - \mathbf{b} * \mathbf{y} \right\|_2^2 + \lambda (\|\mathbf{y}\|_1 + \|\mathbf{y} - \mathbf{z}\|_1) \quad (23)$$

式中 $\tilde{\mathbf{L}} = \mathbf{L} - \mathbf{a} * \mathbf{x}$ 。使用分段迭代软阈值算法求解

$$\mathbf{y}^{(t+1)} = \mathcal{P}_\lambda \left(\mathbf{y}^{(t)} - \frac{1}{\mu} \mathbf{b}^T * (\mathbf{b} * \mathbf{y}^{(t)} - \tilde{\mathbf{L}}); \mathbf{z} \right) \quad (24)$$

式中 $\mathcal{P}_\lambda(\mathbf{v}; \mathbf{s})$ 是分段软阈值函数。同理,将此迭代过程转换为共有信息提取模块

$$y^{(t+1)} = \mathcal{P}_\lambda(y^{(t)} - E_y * D_y * y^{(t)} + E_y * \tilde{L}; z) \quad (25)$$

模块结构如图6所示。使用残差结构得到稀疏特征图 $\tilde{L} = L - a * x$ 。为了使网络灵活,将 x 从特征空间投影到图像空间时,使用一个卷积层实现滤波器 a 。

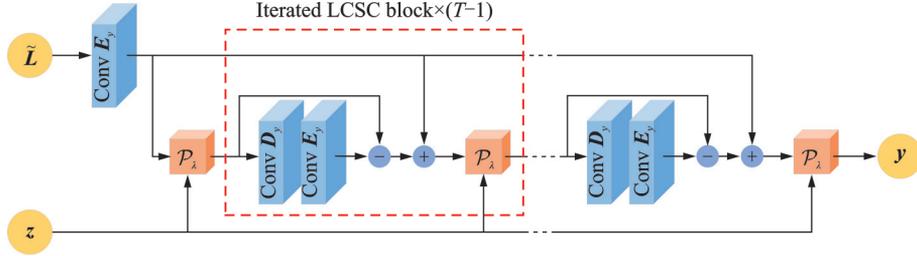


图6 共有信息提取模块

Fig.6 Common information extraction module (CommonIEM)

2.2 双流融合网络模块

为了充分利用HR-MSI的结构信息,将其再次输入网络与卷积稀疏编码获得的HR-HSI融合,以更好地提升融合效果,本文使用双流融合网络实现该操作。双流网络在特征提取和组合双线性数据的特征具有更好的提升效果,已经成功地应用于动作识别、高光谱分类和高光谱超分辨率等领域^[39-41]。

如图7所示,本文的DSFN模块将带有先验信息的HR-HSI和HR-MSI作为两个输入,包含P个残差块的分支提取HR-HSI和HR-MSI的特征信息,然后将结合的特征输入到卷积层。该网络的核心是残差块,包含两个卷积层(Conv)和批量标准化层(Batch normalization, BN),使用PReLU作为激活函数。BN层用于加快训练过程并提高图像重建性能。跳跃连接通过元素求和将输入特征图传递到残差块的输出。最后一个卷积层包含B个卷积过滤器,B为图像包含的光谱波段数。其他卷积层包含64个过滤器,每个过滤器的内核大小为3×3,步长为1。

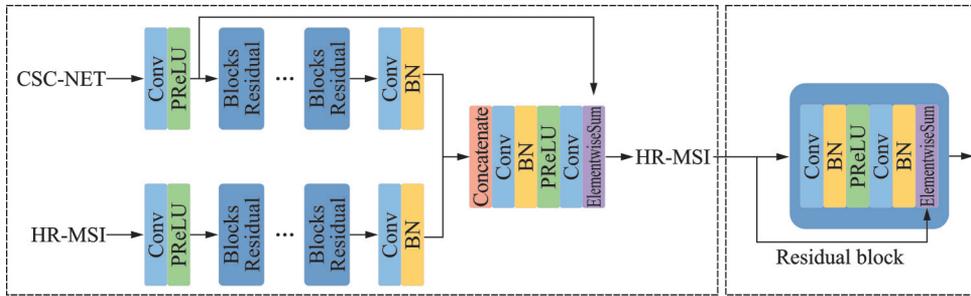


图7 双流融合网络

Fig.7 Dual stream fusion network (DSFN)

2.3 损失函数

本文的网络整体包含深度展开网络模块和双流网络模块,使用监督方式端到端训练,以反向传播的方式更新整体网络参数。深度展开模块中,信息提取子模块的优化目标由模型设定的网络结构所约束,并且在训练过程中自适应优化。为了测量网络输出与目标图像间的差距,选用 l_1 范数作为本文的损失函数为

$$\text{Loss}(\Theta) = \sum_{k=1}^K \left\| \mathcal{F}(L_{\text{up},k}, M_k; \Theta) - H_k \right\|_1 \quad (26)$$

式中:总网络为 $\mathcal{F}(\cdot)$; $L_{\text{up},k}$ 表示第 k 个上采样 LR-HSI; M_k 表示第 k 个 HR-HSI; H_k 表示第 k 个 HR-HSI; Θ 为可以以监督方式端对端训练的网络参数; K 表示训练图像对的数量。

3 实验及结果分析

3.1 实验环境及参数设置

本文使用CAVE^[42]和Harvard^[43]两个数据集作为实验数据,评估所提方法的性能。两个数据集部分图像如图8所示。其中CAVE数据集由32张在控制波段照明下拍摄的室内HSI组成,每个HSI空间分辨率为512像素×512像素,包含31个光谱波段,覆盖的光谱为400~700 nm,波长间隔为10 nm。Harvard数据集包含50张在自然光下拍摄的室内和室外HSI,图像大小为1 392像素×1 040像素,包含31个光谱波段,覆盖420~720 nm的可见光谱。本文取其左上角的1 024像素×1 024像素大小的图像用于实验。

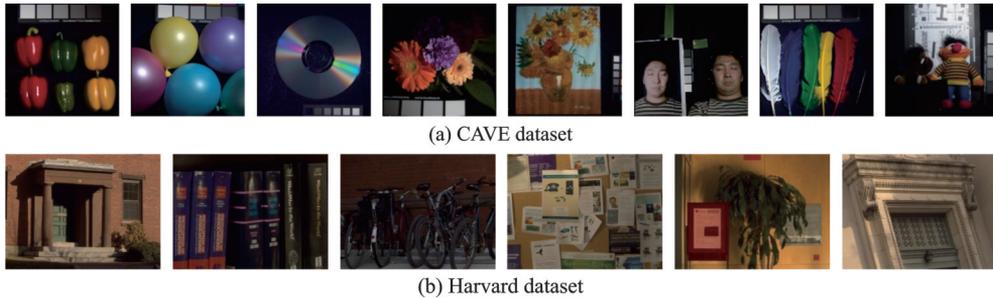


图8 部分CAVE数据集和部分Harvard数据集

Fig.8 Examples of RGB images from CAVE dataset and Harvard dataset

本文的实验数据处理如下。将来自两个数据集的HSI归一化到 $[0, 1]$ 的范围内作为参考图像。LR-HIS中 L 通过对参考图像进行下采样生成,下采样比例因子 δ 设置为8、16和32。本文分别从CAVE、Harvard数据集中选择前20、30个HSI作为训练集,剩余的HSI作为测试集,然后根据相应的训练集和测试集生成LR-HSI和HR-MSI图像。此外,为了丰富训练数据的多样性,从训练集中随机裁剪大小为 128×128 的块,并通过旋转(90° 、 180° 、 270°)和翻转来进行数据增强。在测试阶段,本文直接利用从测试集中的HR-HSI退化而来的LR-HSI和HR-MSI图像来重建目标图像。

本文模型通过Pytorch深度学习框架搭建,所有的实验都是在Ubuntu 18.04.3 LTS系统的服务器上实现,该服务器拥有内存为32 GB的Intel® Core™ i9-9900KF CPU和内存为11 GB的NVIDIA GeForce RTX 2080 GPU。此外,采用Adam算法^[44]作为优化器,batch size设置为16,初始学习率为0.000 2,迭代2 000次。为了综合评估重建的高光谱图像的质量,本文采用RMSE、PSNR、ERGAS和SAM四个指标评估算法的有效性。

RMSE^[45]常用于表示两幅图像之间的相似度。设 X 和 \tilde{X} 的大小为 $M \times N$,其定义为

$$\text{RMSE}(X, \tilde{X}) = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |X(i, j) - \tilde{X}(i, j)|^2} \quad (27)$$

式中: $X(i, j)$ 表示图像 X 在位置 (i, j) 的像素值; $\tilde{X}(i, j)$ 表示图像 \tilde{X} 在位置 (i, j) 的像素值。较小的RMSE值表示图像之间的相似度较高。

PSNR^[46]为峰值信噪比,表示两幅图像之间差异程度,定义为

$$\text{PSNR}(X, \tilde{X}) = -\frac{10}{S} \sum_{i=1}^S \lg(\text{MSE}(X_i, \tilde{X}_i)) \quad (28)$$

式中: S 表示图像的光谱波段数; $\text{MSE}(\cdot)$ 返回输入的均方误差。较高的PSNR值表明融合图像与参考图像之间的差异较小,并且保留了更多的细节。

ERGAS表示所有波段的合成误差,定义为

$$\text{ERGAS}(\mathbf{X}, \tilde{\mathbf{X}}) = \frac{100}{d} \sqrt{\frac{1}{S} \sum_{i=1}^s \frac{\text{MSE}(\mathbf{X}_i, \tilde{\mathbf{X}}_i)}{\nu_{\tilde{\mathbf{X}}_i}^2}} \quad (29)$$

式中: d 为空间下采样因子; ν 表示图像像素值的平均值。较小的ERGAS值表示两幅图像之间的光谱一致性较高。

SAM^[47]是评价频谱失真的指标,定义为

$$\text{SAM}(\mathbf{X}, \tilde{\mathbf{X}}) = \frac{1}{M} \sum_{j=1}^M \arccos \frac{\langle \mathbf{X}, \tilde{\mathbf{X}} \rangle}{\|\mathbf{X}^j\|_2 \|\tilde{\mathbf{X}}^j\|_2} \quad (30)$$

式中: M 为像素的数量; $\langle \cdot \rangle$ 表示两个向量的内积; \mathbf{X}^j 和 $\tilde{\mathbf{X}}^j$ 表示两幅图中第 j 个像素。较小的SAM值表示较少的频谱失真。

3.2 模型分析

3.2.1 网络参数选择

DUDS-FusionNet中含有2个重要参数,分别为卷积稀疏编码中迭代次数 T 以及DSFN残差模块的数量 P 。本节对2个参数进行分析。首先分析迭代次数 T ,由于3个模块具有相同的数学原理,所以3个模块中的 T 取相同值。随着参数的增大,整体性能也会逐渐提高,过大的 T 值将会增加网络的时间复杂度。本节测试了 T 从1~8不同取值的融合性能,如图9(a)所示。可以看到,本文方法在4个指标上的性能随着深度的增加而逐渐提高。当 T 处于4~6时性能趋向稳定,因此选择 $T=4$ 为卷积稀疏编码块数。另外,DSFN残差模块的数量 P 对结果的影响如图9(b)所示。较小的 P 值将会导致网络学习能力不足,相反地,较大的 P 值将会增加时间复杂度。如图所示,当 P 的值设置为6时,性能与效率最佳。

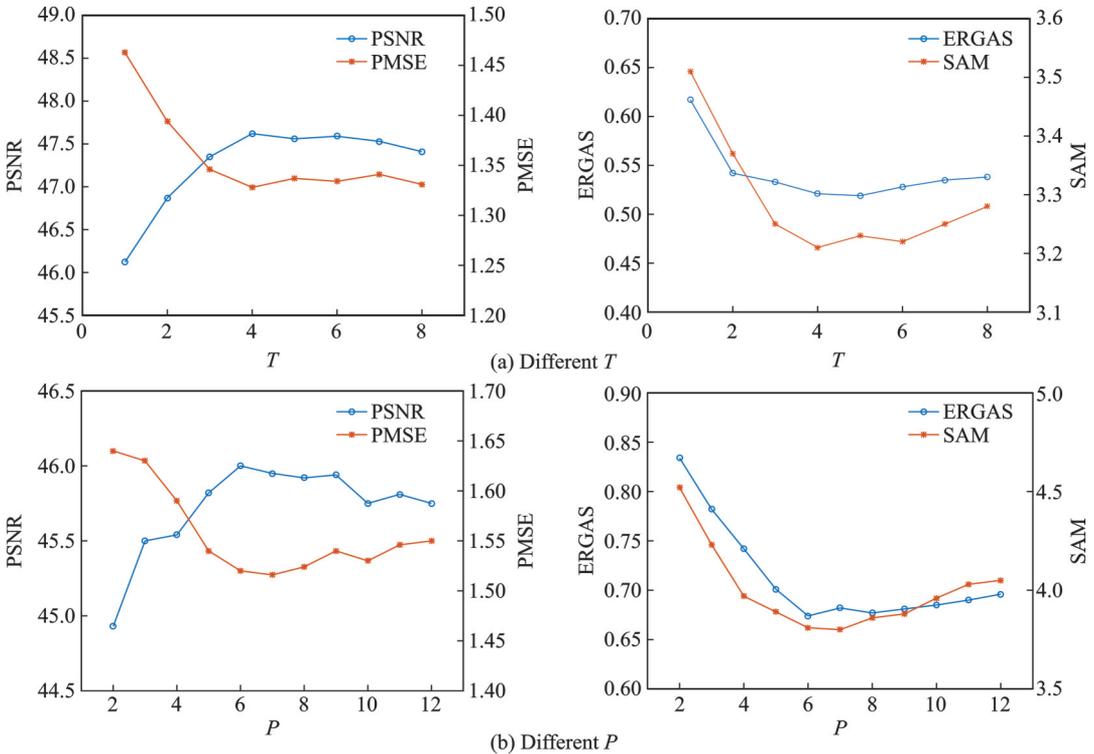


图9 不同 T 和不同 P 时在CAVE测试数据集上的4种评价指标平均值

Fig.9 Average of test results of four evaluation metrics on the CAVE test dataset with different T and P

3.2.2 消融实验

本节将对网络的两个主要部分:深度展开网络和双流融合网络做消融实验分析。实验结果如表1所示,表中 \uparrow 表示值越大越好, \downarrow 表示值越小越好。首先分析深度展开网络对结果的影响。表1中第1行描述了只使用双流融合网络、没有使用深度展开网络的重建效果,记作“w/o DU”。与第3行(提出算法)获得的结果对比可知,其指标分别上升或下降了1.608、0.205 1、0.157 9和0.602 2。图10显示了测试图像“fake_and_real_tomatoes”在550 nm波段对应不同消融实验的误差图。从视觉上看,w/oDU方法在恢复目标HR-HIS时产生的误差比DUDS-FusionNet更明显。这说明深度展开模型在整个模型中起到了重要的作用。

表1 不同方法在CAVE数据集上的平均RMSE、PSNR、ERGAS和SAM($\delta=8$)

Table 1 Average RMSE, PSNR, ERGAS, and SAM of different methods with scaling factor $\delta=8$ on CAVE dataset

Method	PSNR \uparrow	RMSE \downarrow	ERGAS \downarrow	SAM \downarrow
w/o DU	46.016	1.529 0	0.683 8	3.813 3
w/o DSFN	45.218	1.726 4	0.779 3	4.103 9
Proposed	47.624	1.323 9	0.525 9	3.211 1

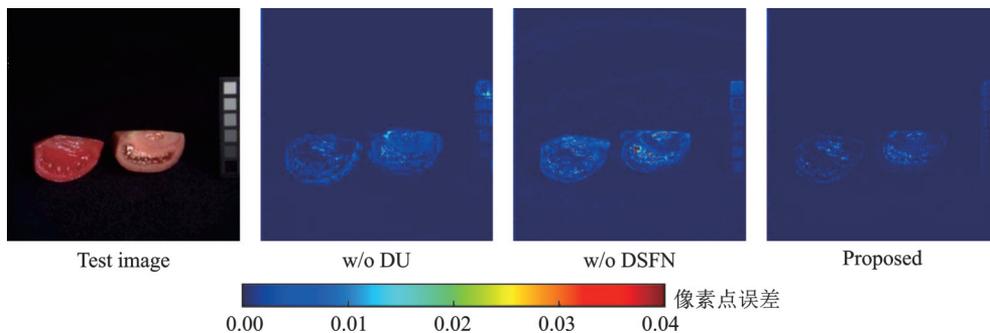


图10 CAVE数据集在550 nm波段不同消融实验的误差图($\delta=8$)

Fig.10 Error maps of different ablation experiments in 550 nm wave band on CAVE dataset($\delta=8$)

其次分析双流融合网络的影响。表1中第2行描述了只使用深度展开网络没有使用双流融合模型的重建效果,表示为“w/o DSFN”。与第3行(提出算法)获得的结果对比可知,其指标分别上升或下降了2.406、0.402 5、0.253 4和0.892 8。如图10所示,w/o DSFN方法由于在融合过程中所获取的空间信息较少,其误差图表现出在分辨率精度上有明显欠缺,这证明了DSFN在提升空间分辨率方面有显著作用。

3.3 对比试验

为了验证提出算法DUDS-FusionNet的有效性,将其与7种有代表意义的算法进行比较。包括3种基于字典的融合算法和4种基于深度学习的融合算法。3种基于字典的融合算法分别为贝叶斯稀疏表示BSR^[8]、非负结构化稀疏表示NSSR^[9]和自适应非负稀疏表示ANSR^[19]。4种基于深度学习的方法包括一个无监督学习方法UMAG-Net^[23]和3个基于模型展开的深度学习方法DHSIS^[25]、GDRRN^[28]和MHF-Net^[29]。其中前4种方法为无监督学习方法,不需要训练。对于NSSR,原子字典的个数 K 设置为75,非局部相似正则化参数 η_1 和稀疏正则化参数 η_2 设置为 $\eta_1=0.015$, $\eta_2=10^{-4}$ 。对于ANSR,原子字典的个数 K 设置为80,非局部相似正则化参数 η_1 和稀疏正则化参数 η_2 设置为 $\eta_1=10^{-2}$, $\eta_2=10^{-4}$ 。后3种方法为监督学习,为了保证测试结果的公平性,对比方法与本文所提出的方法都在同一个服务器上使用相同的数据集进行了训练。对比方法所涉及的参数以及训练迭代次数均根据参考文献中的建议进行调整,且均使用Adam优化器训练。

3.3.1 CAVE数据集评估

对CAVE数据集12张测试图像进行测试,得到每种融合方法的客观评价指标平均值。表2展示了不同算法在不同缩放因子下对CAVE数据集融合的实验结果。在表2中, \downarrow 表示值越小越好, \uparrow 表示值越大越好。通过该表可以得出,本文所提出的方法除了在比例因子为32的情况下SAM指标略低于MHF-net,都优于所有竞争方法,尤其是当比例因子较小时($\delta=8$),本文算法的优越性更加显著。

表2 不同方法在CAVE数据集上的平均RMSE、PSNR、ERGAS和SAM($\delta=8,16,32$)

Table 2 Average RMSE, PSNR, ERGAS, and SAM of different methods with three scaling factors ($\delta=8,16,32$) on CAVE dataset

Scale	Method	PSNR \uparrow	RMSE \downarrow	ERGAS \downarrow	SAM \downarrow
$\times 8$	BSR	44.152	1.757 0	0.972 0	3.422 1
	NSSR	45.513	1.443 6	0.812 2	3.873 5
	DHSIS	46.428	1.386 3	0.628 2	3.747 8
	GDRRN	46.841	1.393 9	0.589 1	3.366 2
	UMAG-Net	46.553	1.493 3	0.680 6	3.849 2
	MHF-Net	46.471	1.420 3	0.643 0	3.605 3
	ANSR	45.846	1.424 8	0.710 4	3.558 1
	Proposed	47.624	1.323 9	0.525 9	3.211 1
$\times 16$	BSR	41.573	2.317 4	0.612 3	4.876 6
	NSSR	43.676	1.775 5	0.482 3	4.434 0
	DHSIS	44.741	1.852 4	0.417 0	4.569 1
	GDRRN	44.985	1.676 7	0.375 4	4.296 1
	UMAG-Net	44.871	1.844 7	0.416 7	4.639 8
	MHF-Net	45.812	1.572 3	0.352 4	3.841 4
	ANSR	44.289	1.674 2	0.390 8	4.336 9
	Proposed	46.322	1.492 6	0.324 8	3.821 5
$\times 32$	BSR	39.162	3.125 2	0.375 4	6.755 9
	NSSR	42.227	2.210 5	0.322 2	5.687 3
	DHSIS	42.369	2.448 4	0.262 6	5.798 6
	GDRRN	43.439	2.164 2	0.247 6	5.165 8
	UMAG-Net	43.336	2.286 1	0.248 3	5.254 1
	MHF-Net	44.661	1.856 7	0.201 4	4.334 6
	ANSR	42.591	2.007 1	0.258 1	5.348 9
	Proposed	44.871	1.832 1	0.199 2	4.632 3

图11展示了不同算法对CAVE数据不同波段的平均PSNR值。通过图11可以看出,在所有测试方法中,本文方法在大部分光谱波段表现最好。主要原因在于:(1)提出的算法使用了稀疏卷积网络展开获取先验HR-HIS;(2)DSFN有效地将先验信息与空间信息融合起来,从而有助于获得更好的性能。除了运用客观指标评估以外,同时采用视觉效果展示提出方法。图12显示了测试图像“real and fake peppers”在550 nm波段的重建图像和对应的误差图,其放大因子 $\delta=8$ 。从视觉上看,本文方法在恢复目标HR-HSI的细节方面提供了最好的结果。相比与其他方法,在红框内物体表面高光部分的误差图更接近于0,这证明了本文方法在恢复目标HR-HSI的空间信息方面的有效性。

3.3.2 Harvard数据集评估

对Harvard数据集20张测试图像进行测试,得到每种融合方法的客观评价指标平均值。表3展示了不同算法使用不同缩放因子在Harvard数据集中的实验结果。与CAVE数据集不同,Harvard数据集

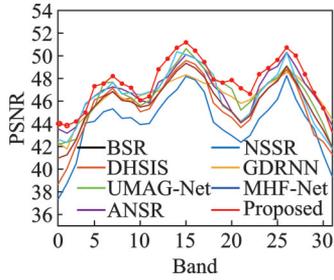


图 11 不同方法在 CAVE 测试数据集的光谱带 PSNR 值
Fig.11 Average PSNR of spectral bands of different methods on CAVE test dataset

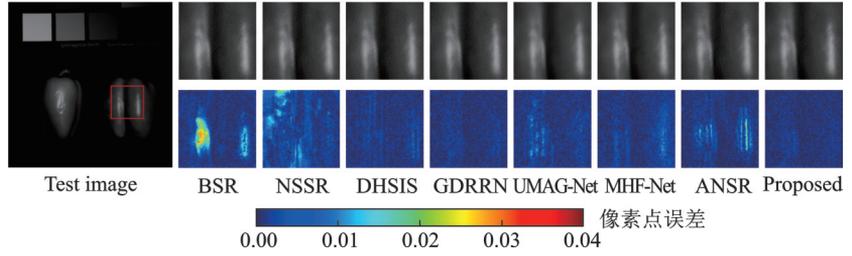


图 12 CAVE 数据集在 550 nm 波段图像的重建图及相应的误差图($\delta=8$)
Fig.12 Reconstructed images and corresponding error maps in 550 nm band images on CAVE dataset ($\delta=8$)

表 3 不同方法在 Harvard 数据集上的平均 RMSE、PSNR、ERGAS 和 SAM ($\delta=8, 16, 32$)

Table 3 Average RMSE, PSNR, ERGAS, and SAM of different methods with three scaling factors ($\delta=8, 16, 32$) on Harvard dataset

Scale	Method	PSNR \uparrow	RMSE \downarrow	ERGAS \downarrow	SAM \downarrow
$\times 8$	BSR	44.567	1.921 9	1.392 3	3.659 2
	NSSR	45.325	1.674 7	1.331 1	3.591 4
	DHSIS	46.448	1.747 4	1.315 4	3.695 6
	GDRRN	46.236	1.658 0	1.259 4	3.578 3
	UMAG-Net	46.378	1.690 2	1.219 2	3.522 6
	MHF-Net	46.499	1.736 5	1.278 6	3.464 0
	ANSR	45.951	1.662 3	1.291 4	3.396 3
	Proposed	46.720	1.628 3	1.135 3	3.365 3
$\times 16$	BSR	43.914	2.019 5	0.757 6	3.721 4
	NSSR	44.771	1.781 4	0.683 5	3.684 1
	DHSIS	45.576	1.827 0	0.660 9	3.612 0
	GDRRN	45.755	1.767 6	0.682 1	3.596 2
	UMAG-Net	46.087	1.814 3	0.635 8	3.618 3
	MHF-Net	46.053	1.779 4	0.630 6	3.566 4
	ANSR	45.659	1.891 3	0.673 5	3.651 3
	Proposed	46.306	1.743 1	0.618 8	3.565 7
$\times 32$	BSR	42.838	2.147 0	0.392 2	3.800 5
	NSSR	43.527	1.878 5	0.368 7	3.777 0
	DHSIS	45.422	1.923 2	0.348 2	3.795 8
	GDRRN	45.442	1.827 4	0.386 6	3.752 4
	UMAG-Net	45.623	1.867 0	0.383 5	3.728 1
	MHF-Net	45.811	1.885 7	0.329 7	3.669 3
	ANSR	44.852	1.901 3	0.358 5	3.700 3
	Proposed	45.905	1.836 2	0.327 6	3.677 0

的图像空间尺寸更大,轻微移动的物体在相邻波段中没有正确对齐。如表3所示,在与上述相同的实验设置下,所提出方法的数值结果在多数情况下都超过了其他对比方法。

图13显示了不同算法在Harvard数据集上光谱带的PSNR曲线。据图13可以看出,在所有测试方法中,本文方法在大部分光谱波段表现最好。图14显示了图像imgf1在放大因子 $\delta=8$ 的650 nm波段的重建图像和相应的误差图。在实际场景中,高光轮廓部分更容易产生误差。与其他方法对比可知,本文方法所重建的图像中,汽车轮廓的细节更多,高光部分更有层次,且误差显著更小,这证明了本文方法在恢复高光谱图像光谱准确性方面的有效性。

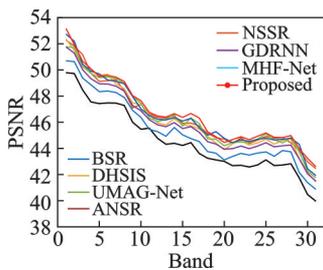


图13 不同方法在Harvard测试数据集的光谱带PSNR值

Fig.13 Average PSNR of spectral bands of different methods on Harvard test dataset

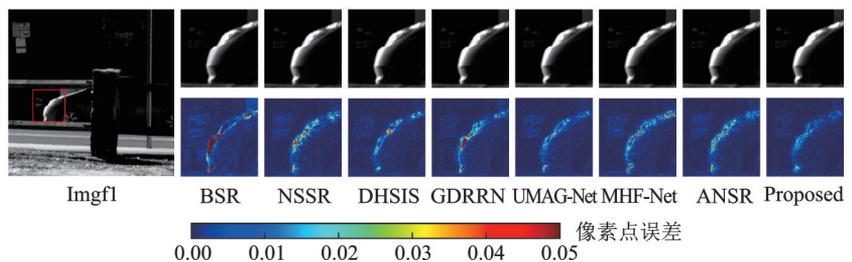


图14 Harvard数据集在650 nm波段图像的重建图及相应的误差图($\delta=8$)

Fig.14 Reconstructed images and corresponding error maps of 650 nm band images on Harvard dataset ($\delta=8$)

4 结束语

针对高光谱图像融合问题,本文提出一种基于深度展开和双流网络的高光谱图像融合网络。该网络使用卷积稀疏编码设计融合模型,并将其展开为可学习深度网络的形式,所以在深度展开的网络部分具有很好的可解释性。在设计融合模型时,将低分辨率高光谱图像分解成共有信息和独有信息,并使用共有信息与高分辨率多光谱图像建立联系,将高分辨率多光谱图像作为辅助信息加入到融合模型中。为了使用稀疏卷积编码更好的表示高光谱图像,将多个三维稀疏字典堆积成一个四维稀疏字典。为了充分挖掘高分辨率多光谱图像的结构信息,将其再次输入双流融合网络中,以获取更为精确的高分辨率高光谱图像。本文尝试性地将高光谱图像融合模型展开为可学习深度网络的形式,本研究还处于初级阶段,设置更为准确融合模型,并设计更为准确的网络展开方式是未来研究的重点。在展开过程中保留其原本数学特性的方法,并减少深度网络的参数量也是未来研究的方向。

参考文献:

- [1] FAUVEL M, TARABALKA Y, BENEDIKTSSON J A, et al. Advances in spectral-spatial classification of hyperspectral images[J]. Proceedings of the IEEE, 2013, 101(3): 652-675.
- [2] TARABALKA Y, CHANUSSOT J, BENEDIKTSSON A J, et al. Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers[J]. Transactions on Systems, Man, and Cybernetics, Part B, 2009, 40(5): 1267-1279.
- [3] UZAIR M, MAHMOOD A, MIAN A. Hyperspectral face recognition using 3D-DCT and partial least squares[C]// Proceedings of BMVC 2013. [S.l.]: BMVC, 2013.
- [4] MANOLAKIS D, SHAW G. Detection algorithms for hyperspectral imaging applications[J]. Signal Processing Magazine IEEE, 2002, 19(1): 29-43.
- [5] BIOUCAS-DIAS J M, PLAZA A, DOBIGEON N, et al. Hyperspectral unmixing overview: Geometrical, statistical, and

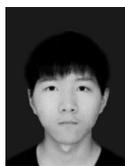
- sparse regression-based approaches[J]. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 2012, 5(2): 354-379.
- [6] AIAZZI B, BARONTI S, LOTTI F, et al. A comparison between global and context-adaptive pansharpening of multispectral images[J]. *IEEE Geoscience & Remote Sensing Letters*, 2009, 6(2): 302-306.
- [7] 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.
- [8] AKHTAR N, SHAFAIT F, MIAN A. Bayesian sparse representation for hyperspectral image super resolution[C]// *Proceedings of IEEE Conference on Computer Vision & Pattern Recognition*. Boston, USA: IEEE, 2015.
- [9] DONG W, FU F, SHI G, et al. Hyperspectral image super-resolution via non-negative structured sparse representation[J]. *IEEE Transactions on Image Processing*, 2016, 25(5): 2337-2352.
- [10] HAN X, YU J, XUE J H, et al. Hyperspectral and multispectral image fusion using optimized twin dictionaries[J]. *IEEE Transactions on Image Processing*, 2020, 29: 4709-4720.
- [11] XU Y, WU Z, CHANUSSOT J, et al. Hyperspectral images super-resolution via learning high-order coupled tensor ring representation[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(11): 4747-4760.
- [12] XU Y, WU Z, CHANUSSOT J, et al. Nonlocal coupled tensor CP decomposition for hyperspectral and multispectral image fusion[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 58(1): 348-362.
- [13] DIAN R, LI S, FANG L. 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.
- [14] LI S, DIAN R, FANG L, et al. Fusing hyperspectral and multispectral images via coupled sparse tensor factorization[J]. *IEEE Transactions on Image Processing*, 2018, 27(8): 4118-4130.
- [15] HAN X H, SHI B, ZHENG Y. Self-similarity constrained sparse representation for hyperspectral image super-resolution[J]. *IEEE Transactions on Image Processing*, 2018, 27(11): 5625-5637.
- [16] LI X, ZHANG Y, GE Z, et al. Adaptive nonnegative sparse representation for hyperspectral image super-resolution[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 4267-4283.
- [17] ZHANG K, WANG M, YANG S, et al. Spatial-spectral-graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion[J]. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 2018, 11(4): 1030-1040.
- [18] XU Y, WU Z, 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.
- [19] LI X, ZHANG Y, GE Z, et al. Adaptive non-negative sparse representation for hyperspectral image super-resolution[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 4267-4283.
- [20] PALSSON F, SVEINSSON J R, ULFARSSON M O. Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network[J]. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(5): 639-643.
- [21] ZHU Zhiyu, HOU Junhui, CHEN Jie, et al. Hyperspectral image super-resolution via deep progressive zero-centric residual learning[J]. *IEEE Transactions on Image Processing*, 2020, 30: 1423-1438.
- [22] ZHANG X, HUANG W, WANG Q, et al. SSR-NET: Spatial-spectral reconstruction network for hyperspectral and multispectral image fusion[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 59(7): 5953-5965.
- [23] LIU Shuaiqi, MIAO Siyu, SU Jian, et al. UMAG-Net: A new unsupervised multiattention-guided network for hyperspectral and multispectral image fusion[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 14: 7373-7385.
- [24] DONG W, ZHOU C, WU F, et al. Model-guided deep hyperspectral image super-resolution[J]. *IEEE Transactions on Image Processing*, 2021, 30: 5754-5768.
- [25] DIAN R, LI S, GUO A, et al. Deep hyperspectral image sharpening[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018: 1-11.
- [26] YANG J, 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*, 2021, 60: 1-17.
- [27] DIAN R, LI S, KANG X. Regularizing hyperspectral and multispectral image fusion by CNN denoiser[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 32(3): 1124-1135.
- [28] WEI W, NIE J, LI Y, et al. Deep recursive network for hyperspectral image super-resolution[J]. *IEEE Transactions on*

- Computational Imaging, 2020, 6: 1233-1244.
- [29] XIE Q, ZHOU M, 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, 2020, 44(3): 1457-1473.
- [30] DENG X, DRAGOTTI P L. Deep convolutional neural network for multi-modal image restoration and fusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(10): 3333-3348.
- [31] MARIVANI I, TSILIGIANNI E, CORNELIS B, et al. Multimodal deep unfolding for guided image super-resolution[J]. IEEE Transactions on Image Processing, 2020, 29: 8443-8456.
- [32] ZEILER M D, KRISHNAN D, TAYLOR G W, et al. Deconvolutional networks[C]//Proceedings of Computer Vision & Pattern Recognition. San Francisco, CA, USA: IEEE, 2010: 2528-2535.
- [33] WOHLBERG B. Efficient algorithms for convolutional sparse representations[J]. IEEE Transactions on Image Processing, 2015, 25(1): 301-315.
- [34] SRETER H, GIRYES R. Learned convolutional sparse coding[J]. Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Calgary, Canada: IEEE, 2018: 2191-2195.
- [35] DAUBECHIES I, DEFRISE M, MOL C D. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint[J]. Communications on Pure and Applied Mathematics, 2004, 57(11): 1413-1457.
- [36] MOTA J, DELIGIANNIS N, RODRIGUES M. Compressed sensing with prior information: Strategies, geometry, and bounds[J]. IEEE Transactions on Information Theory, 2017, 63(7): 4472-4496.
- [37] MOTA J, DELIGIANNIS N, SANKARANARAYANAN A C, et al. Dynamic sparse state estimation using l_1 - l_1 minimization: Adaptive-rate measurement bounds, algorithms and applications[C]//Proceedings of 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). South Brisbane: IEEE, 2015: 3332-3336.
- [38] KEYS R G. Cubic convolution interpolation for digital image processing[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1981, 29(6): 1153-1160.
- [39] FEICHTENHOFER C, PINZ A, ZISSERMAN A. Convolutional two-stream network fusion for video action recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 1933-1941.
- [40] XU X, WEI L, RAN Q, et al. Multisource remote sensing data classification based on convolutional neural network[J]. IEEE Transactions on Geoscience & Remote Sensing, 2017, 56(2): 937-949.
- [41] LI Yijun, HUANG Jiabin, AHUJA N, et al. Joint image filtering with deep convolutional networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2018, 41(8): 1909-1923.
- [42] YASUMA F, MITSUNAGA T, ISO D, et al. Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum[J]. IEEE Transactions on Image Processing, 2010, 19(9): 2241.
- [43] CHAKRABARTI A, ZICKLER T. Statistics of real-world hyperspectral images[C]//Proceedings of IEEE Conference on Computer Vision & Pattern Recognition. Colorado Springs, CO, USA: IEEE, 2011: 193-200.
- [44] KINGMA D, BA J. Adam: A method for stochastic optimization[J]. Computer Science, 2014. DOI: 10.48550/arXiv.1412.6980.
- [45] LEVINSON N. The wiener RMS (root mean square) error criterion in filter design and prediction[M]. [S.l.]: Wiley, 1964.
- [46] WANG Y, LI J, LU Y, et al. Image quality evaluation based on image weighted separating block peak signal to noise ratio [C]//Proceedings of International Conference on Neural Networks & Signal Processing. Nanjing, China: IEEE, 2003.
- [47] YUHAS R H, GOETZ A, BOARDMAN J W. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm[C]//Proceedings of the 4th JPL Airborne Earth Science Workshop. Pasadena: NASA, 1992.

作者简介:



刘丛(1983-), 通信作者, 男, 博士, 副教授, 研究方向: 计算智能、模式识别、图像处理, E-mail: con-gl2014@usst.edu. cn.



姚佳浩(1996-), 男, 硕士研究生, 研究方向: 高光谱图像重建, E-mail: jiahaoy-ao@163.com。