引用格式：LI Ming, LIU Fan, LI Jingzhi. Combining Convolutional Attention Module and Convolutional Auto-encoder for Detail Injection Remote Sensing Image Fusion[J]. Acta Photonica Sinica, 2022, 51(6):0610005
李明,刘帆,李婧芝.结合卷积注意模块与卷积自编码器的细节注入遥感图像融合[J].光子学报,2022,51(6):0610005

结合卷积注意模块与卷积自编码器的细节注入遥感图像融合

李明,刘帆,李婧芝
(太原理工大学 大数据学院,山西 晋中 030600)

摘 要:针对现有的基于卷积自编码器的遥感图像融合方法存在光谱失真和部分细节信息丢失的情况,提出结合卷积自编码器、卷积注意模块和高斯滤波器的遥感图像融合算法。首先利用高斯滤波器获取用于模型训练的低分辨率高频图像、高分辨率高频图像和用于模型预测的低分辨率多光谱高频图像。然后用卷积自编码器学习低分辨率高频图像与高分辨率高频图像之间的非线性映射关系,将卷积注意模块引入模型训练中,使卷积自编码器更加关注图像中的关键信息。最后用训练完成的卷积自编码器获取多光谱图像缺失的细节信息,即高分辨率多光谱高频图像,并与原图像融合生成高分辨率多光谱图像。选取多组不同的卫星数据与8种融合算法进行对比实验,实验结果表明融合图像保留了更多的光谱信息和细节信息,在主观和客观上均表现出良好的性能。

关键词:遥感图像融合;特征提取;卷积自编码;多光谱图像;全色图像
中图分类号:TP391 **文献标识码**:A **doi**:10.3788/gzxb20225106.0610005

0 引言

遥感图像融合的目的是将多源图像综合成高质量图像,以提高图像信息的利用率、提升原始图像的空间分辨率和光谱分辨率,利于监测[1]。其中,全色图像具有较高的空间分辨率,多光谱图像具有丰富的光谱信息,利用其互补信息生成具有高空间分辨率的多光谱波段图像[2],提高图像的可视效果。

目前遥感图像融合的方法主要分为四种类型[2]:1)分量替换(Component Substitution, CS),如强度-色调-饱和度(Intensity Hue Saturation, IHS)变换[3],主成分分析(Principal Component Analysis, PCA)[4]和Gram-Schmidt(GS)[1]等;2)多分辨率分析(Multiresolution Analysis, MRA),如非下采样Contourlet变换(Non-Subsample Contourlet Transform, NSCT)[5]和非下采样剪切波变换(Non-Subsample Shearlet Transform, NSST)[2]等;3)变分优化(Variational Optimization, VO),如基于贝叶斯[6]和稀疏表示[7]等;4)机器学习(Machine Learning, ML)[2],如基于卷积神经网络(Convolutional Neural Network, CNN)[8]、对抗神经网络(Generative Adversarial Networks, GAN)[9]和条件随机场(Conditional Random Field, CRF)[10]等。然而大部分算法需要设计复杂的融合规则,且仍然存在不同程度的细节缺失和光谱失真情况。近年来,基于VO/ML算法与CS/MRA算法的混合模型算法被提出,并取得显著的效果。混合模型算法将光谱信息与空间细节信息分开处理,在光谱信息保留上有了很大的提升,如基于稀疏表示的细节注入(Detail injection based Sparse Representation, SR-D)[11]与基于卷积神经网络的遥感图像融合(Detail injection based Convolutional Neural Network, Di-PNN)[12]。但是仅依靠浅层网络获取的融合图像仍存在图形特征提取不完全,导致图像信息保留不全面、信息缺失和冗余信息等问题[2],使得融合图像的空间信息和光谱信息没有

基金项目:国家自然科学基金(No.61703299)
第一作者:李明(1996—),男,硕士研究生,主要研究方向为遥感图像处理、深度学习。Email: minglitut@163.com
导师(通讯作者):刘帆(1982—),女,副教授,博士,主要研究方向为遥感图像处理、深度学习、机器学习。Email: liufan@tyut.edu.cn
收稿日期:2021-12-14;**录用日期**:2022-02-07

得到有效的表达，在细节信息和光谱信息上有很大的提升空间。

为了解决上述问题，本文提出结合卷积注意模块、高斯滤波器和卷积自编码器作为新的混合模型的融合算法。一方面卷积自编码器可以从低分辨率高频图像中获取对应的高分辨率高频图像；另一方面高斯滤波器与卷积注意模块可以提高网络的学习能力，获取更准确的图像高频信息。其中，高斯滤波器可以从已知图像中，通过滤波获取图像的高低频信息[13]，并通过二次滤波获取图像的降阶图像。在模型训练过程中引入卷积注意模块提高图像关键信息的重要性和网络的学习能力，使得模型预测的图像更加接近缺失的细节信息，减少融合阶段的光谱失真，获取更加准确的高分辨率多光谱图像。

1 卷积自编码器

通过训练卷积自编码器获取低分辨率高频图像与高分辨率高频图像的非线性映射关系。卷积自编码器具有局部连接特性，并通过卷积操作实现权重共享和二维空间信息保留[14]。卷积自编码器的结构包括两个主要部分：编码和解码。

1.1 编码

编码获取输入图像的压缩版本，由卷积层和最大池化层组成编码部分网络。编码部分通过卷积操作捕捉图像中的结构信息；最大池化删除非重叠子区域中的所有非最大值，在隐藏表示上引入稀疏性，使特征检测器变得更广泛适用。卷积自编码器随机初始化 w 个卷积核，表示为 $\boldsymbol{K}=[\boldsymbol{K}_1,\boldsymbol{K}_2,\cdots,\boldsymbol{K}_w]$，通过卷积核 $\boldsymbol{K}$ 卷积操作，提取到 w 个卷积核下的图像的特征信息 T_w。计算公式为

$$T_w=\delta(F\cdot\boldsymbol{K}_w+\boldsymbol{b}_w) \tag{1}$$

式中，δ 表示编码层的激活函数，F 为被处理的图像数据，T_w 为经过卷积核操作后获取图像的特征信息，$\boldsymbol{b}_w$ 表示 w 个卷积核偏置项。编码部分由多个卷积层和最大池化层的叠加组成，上一层提取的特征信息被作为下一层的输入信息，通过多层运算获取图像的压缩版本。同时在编码部分引入卷积注意模块，处理编码部分中间步骤中的信息，实现对图像的自适应特征细化。

1.2 解码

解码将编码部分的图像压缩版本进行图像重建。由反卷积层和上采样层组成解码部分。重建图像是通过反卷积 T_w 和图像上采样实现。计算公式为

$$\tilde{F}=\delta\left(\sum_w T_w\cdot\boldsymbol{K}_w^{\mathrm{T}}+\boldsymbol{b}_w\right) \tag{2}$$

式中，$\tilde{F}$ 用来描述解码层重建的输入图像，δ 为解码层的激活函数，$\boldsymbol{K}_w^{\mathrm{T}}$ 为编码计算得到的卷积核的转置。利用 $\boldsymbol{K}_w^{\mathrm{T}}$ 对 T_w 进行反卷积操作，将反卷积后的结果相加得到新的重构特征图 T_w。解码层通过多个反卷积层和上采样层的叠加实现对输入图像的重建。

1.3 损失函数

卷积自编码器通过编码与解码的叠加实现了对图像信息的压缩和重建。与标准网络一样，训练过程中通过反向传播算法对神经网络中的权重进行微调。用低分辨率高频图像作为输入图像，高分辨率高频图像 F_{t} 作为标签图像。用均方误差作为卷积自编码器的损失函数，计算在 n 个样本数量下，卷积自编码器计算重建图像 $\tilde{F}$ 与标签图像 F_{t} 之间的均方误差。计算公式为

$$\mathrm{Loss}=\frac{1}{n}\sum_{i=0}^{n}(\widetilde{F_i}-F_{\mathrm{t}_i})^2 \tag{3}$$

2 卷积注意模块

注意力通过卷积注意模块实现，卷积注意模块（Convolutional Attention Module，CAM）是一种简单而有效的前馈卷积神经网络注意模块。给定一个中间特征信息，模块沿着通道和空间两个维度顺序推断注意映射，然后将注意映射与输入特征相加以进行自适应特征细化[15]。将卷积注意模块与卷积自编码器编码部分结合，编码部分最大池化后的信息作为卷积注意模块的输入，对信息进一步细化自适应特征和加强注意力。图1为卷积注意模块流程。

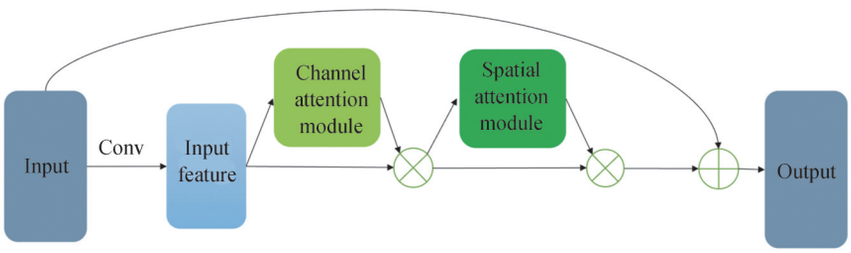


图1　卷积注意模块流程

Fig. 1　Convolutional attention module flowchart

2.1　通道注意力模块

通道注意力模块利用特征的通道关系生成通道注意图。通道注意力模块首先采用平均池化聚合空间信息，最大池化提取图像的纹理信息，生成两个不同的上下文信息，然后将获取的两个信息输入一个共享网络中生成通道注意力图$\boldsymbol{M}$[16]。共享网络由多层感知机（Multilayer Perceptron，MLP）和隐藏层组成。并通过δ函数对共享网络的输出信息进行合并生成通道注意力图$\boldsymbol{M}$，表达为

$$\boldsymbol{M}=\delta(\mathrm{MLP}(\mathrm{AvgPool}(\boldsymbol{F}))+\mathrm{MLP}(\mathrm{MaxPool}(\boldsymbol{F}))) \tag{4}$$

式中，对待处理的图像信息$\boldsymbol{F}$使用平均池化（AvgPool）和最大池化（MaxPool）操作聚合特征地图中的空间信息，生成平均池化特征和最大池化特征。然后将两个特征经过MLP和一个隐藏层，求和合并输出特征向量，如图2所示。

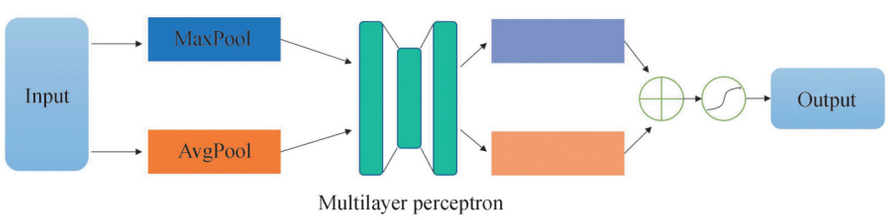


图2　通道注意力模块流程

Fig. 2　Channel attention module flowchart

获取图像的通道注意力图$\boldsymbol{M}$后，将$\boldsymbol{M}$与通道注意力模块的输入特征做乘法，获取经过通道注意力处理的特征图，并输入至空间注意力模块进一步细化。

2.2　空间注意力模块

空间注意力模块是在通道注意力模块基础上的进一步操作。沿通道轴应用池化操作可以有效地突出特征图中信息区域[15]。为了有效计算空间注意，沿通道轴进行平均池化和最大池化操作，然后通过一个标准卷积层将它们连接并卷积，生成空间注意力图，表示为

$$\boldsymbol{S}=\delta(f^{n\times n}([\mathrm{AvgPool}(\boldsymbol{M});\mathrm{MaxPool}(\boldsymbol{M})])) \tag{5}$$

式中，对通道注意力图$\boldsymbol{M}$依次进行平均池化（AvgPool）和最大池化（MaxPool）操作，并通过一个$n\times n$大小的标准卷积将它们连接，如图3所示。

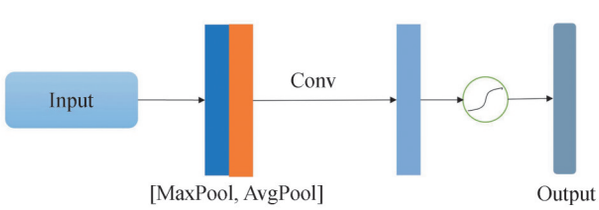


图3　空间注意力流程

Fig. 3　Spatial attention flowchart

获取图像的空间注意力图S后，先利用S与空间注意力模块输入特征图做乘积获取经过空间注意力模块处理后的特征图，再与卷积注意模块的输入信息F相加即可获取输入图像自适应特征细化后的图像。通过卷积注意模块的自适应特征细化，卷积自编码器在编码部分得以保留更多的关键信息，减少图像细节信息的丢失和光谱信息失真。

3 算法流程描述

基于理论分析结合卷积自编码器、卷积注意模块和高斯滤波器[17]作为新的遥感图像融合算法。算法的核心是通过训练卷积自编码器获取低分辨率高频图像与高分辨率高频图像的非线性映射关系；利用获取的非线性映射关系，将低分辨率多光谱图像的高频图像作为输入，获取其对应的高分辨率多光谱图像的高频信息，即多光谱图像缺失的细节信息。流程见图4。

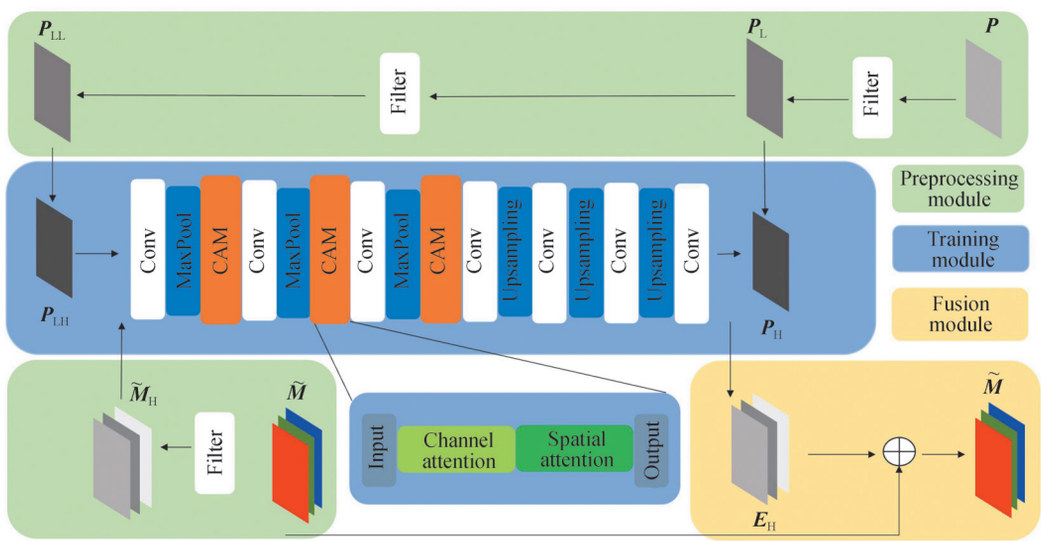


图4 算法流程

Fig. 4 Method flowchart

3.1 数据预处理

1）将全色图像P通过$P_L = P \cdot \frac{1}{2\pi\sigma^2}\exp\left(-\frac{P^{\mathrm{T}}P}{2\sigma^2}\right)$计算进行滤波处理，获取高分辨率$P$的低频信息$P_L$，并通过图像差值操作$P_H = P - P_L$获取高分辨率$P$的高频图像$P_H$。将$P_L$图像进行二次滤波操作，获取$P_L$图像的高频图像$P_{LH}$，即为高分辨率高频图像$P_H$对应的低分辨率高频图像$P_{LH}$。

2）将多光谱图像M处理至与P相同像素大小的$\widetilde{M}$图像，将$\widetilde{M}$图像中的每个波段k通过$\widetilde{M_{Lk}} = \widetilde{M_k} \cdot \frac{1}{2\pi\sigma^2}\exp\left(-\frac{\widetilde{M}_k^{\mathrm{T}}\widetilde{M}_k}{2\sigma^2}\right)$计算获取$\tilde{M}$图像的低频图像。进一步通过图像差值操作$\widetilde{M_{Hk}} = \widetilde{M}_k - \widetilde{M_{Lk}}$获取$\tilde{M}$图像不同波段的高频图像$\widetilde{M_{Hk}}$。

3.2 模型训练

模型训练是通过卷积自编码器的训练获取低分辨率高频图像和高分辨率高频图像的映射关系。在此阶段，低分辨率高频图像P_{LH}作为卷积自编码器的输入图像，高分辨率高频图像P_H作为卷积自编码器的标签图像。

1）编码部分：编码部分用于获取低分辨率高频图像P_{LH}的压缩版本。在编码部分最大池化后引入卷积注意模块，实现对编码部分中间过程图像的自适应特征细化，从而加强卷积自编码器对关键信息的关注。

2）解码部分：解码部分用于将低分辨率高频图像P_{LH}的压缩版本进行图像重建。通过上采样操作恢复图像的大小，反卷积操作实现对图像信息的恢复。并计算重建图像与标签图像P_H的均方误差，更新网络的

参数权值。

为卷积自编码器设定一定的迭代次数，模型优化器 Adadelta[18]对模型参数进行更新，获取低分辨率高频图像 $\boldsymbol{P}_{\mathrm{LH}}$ 与高分辨率高频图像 $\boldsymbol{P}_{\mathrm{H}}$ 之间的映射关系。

3.3 图像融合

1）$\widetilde{\boldsymbol{M}}$ 图像细节获取：训练好的网络模型可以获取低分辨率高频图像对应的高分辨率高频图像，因此将 $\widetilde{\boldsymbol{M}}$ 图像每个波段的高频图像 $\widetilde{\boldsymbol{M}_{\mathrm{H}k}}$ 输入至训练好的模型。通过模型预测计算，获取 $\widetilde{\boldsymbol{M}_{\mathrm{H}k}}$ 图像对应的高分辨率高频图像 $\boldsymbol{E}_{\mathrm{H}k}$。

2）图像融合：图像的融合遵循 CS 算法一般融合框架[1]，计算公式为

$$\hat{\boldsymbol{M}} = \widetilde{\boldsymbol{M}} + \boldsymbol{D} \tag{6}$$

式中，$\hat{\boldsymbol{M}}$ 即为高分辨率多光谱图像，$\boldsymbol{D}$ 为 $\widetilde{\boldsymbol{M}}$ 图像缺失的细节信息。结合式(6)，本算法中将 $\tilde{\boldsymbol{M}}$ 图像对应的高分辨率高频图像 $\boldsymbol{E}_{\mathrm{H}}$ 与 $\widetilde{\boldsymbol{M}}$ 图像对应波段相加，获取高分辨率多光谱图像 $\hat{\boldsymbol{M}}$。

4 实验结果和分析

4.1 实验平台及参数设置

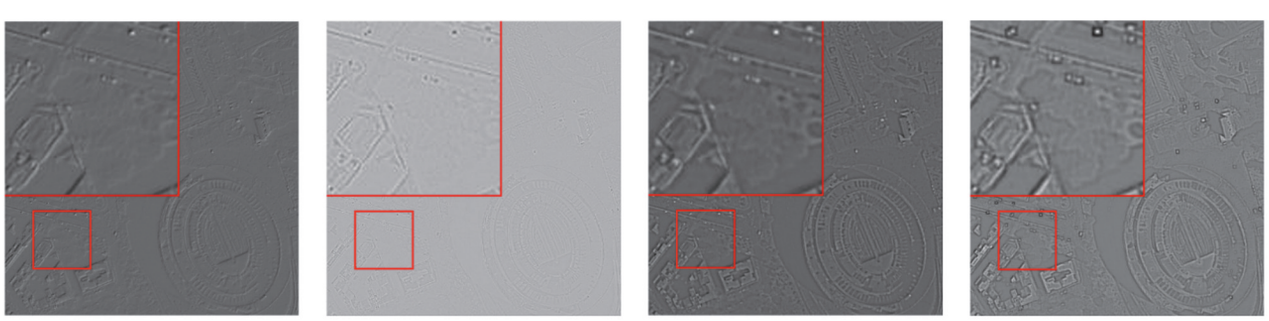
实验采用 QuickBird 和 SPOT 卫星数据集作为试验训练数据和验证数据。为了扩充样本数据集，采用滑窗方式，裁剪步长设置为 5，裁剪图像块尺寸为 8×8。在训练中模型训练批量大小为 256，训练迭代次数为 1 600，优化器 Adadelta 进行网络模型参数优化和学习率自适应。

操作系统为 Win10；CPU 为 Intel Core i5，主频 2.2 GHz；内存 500 GB。软件平台：MATLAB2016a、Pycharm。训练环境：Keras 2.1.6；Tensorflow 1.8；Python 版本为 3.6.7。

4.2 滤波器选择分析

用滤波器对模型训练数据进行预处理，克服模型训练中高低频信息冗余导致信息缺失的问题[12]。同时利用滤波器生成高分辨率高频图像对应的低分辨率高频图像。为了验证滤波器对算法的适应性，选择基于 QuickBird 卫星获取的遥感图像进行实验，在网络模型迭代次数为 150 次的情况下获取本文算法的融合结果。

图 5 为 QuickBird 数据集进行一次滤波后的高分辨率高频图像，含有大量图像细节信息，作为模型训练的标签图像。从图 5 中可以明显发现高斯滤波器、形态学滤波器和均值滤波器有较好的细节提取能力。从图 5(c)可知均值滤波器处理图像的同时对图像进行模糊化处理，图像中建筑边缘信息丢失明显；形态学滤波器相对于均值滤波器保留了更多的图像细节信息，但是从图 5(d)放大区域可以发现，最上部中汽车边缘部分有明显的像素扭曲。从图 5(a)中发现高斯滤波器处理后图像避免了细节丢失和噪声引入。从图 5(b)可以看出拉普拉斯滤波器提取细节能力比较弱。图 6 为进行二次滤波后的图像，用来模拟低分辨率高频图像。由图 4 可知训练完成的卷积自编码器可获取 $\widetilde{\boldsymbol{M}}$ 缺失的细节信息 $\boldsymbol{E}_{\mathrm{H}}$，即 $\widetilde{\boldsymbol{M}}$ 滤波后的图像为低分辨率高频图像，通过模型获取对应的高分辨率高频图像。因此 $\boldsymbol{P}$ 进行二次滤波后的图像 $\boldsymbol{P}_{\mathrm{LH}}$ 与 $\widetilde{\boldsymbol{M}_{\mathrm{H}}}$ 相关性越强，$\widetilde{\boldsymbol{M}}$



(a) Gaussian filter　(b) Laplace filter　(c) Mean filter　(d) Morphological filter

图 5　网络模型的标签图像

Fig. 5　Labeled images of the network model

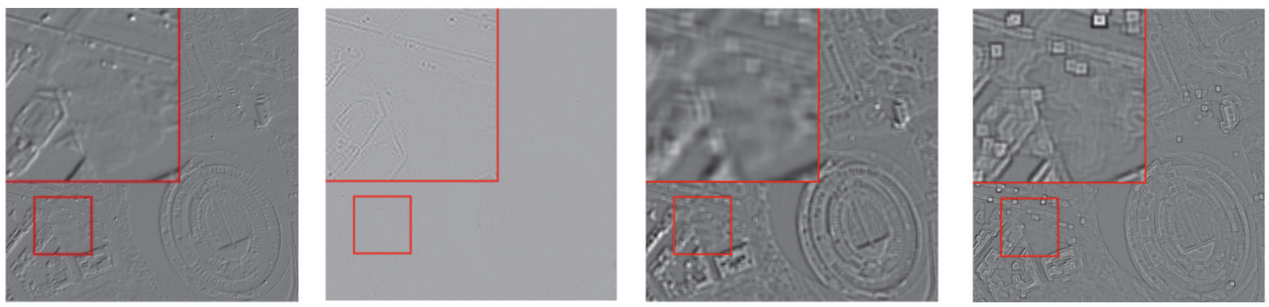


图 6 网络模型的输入图像

Fig. 6 Input images of the network model

图像能获取更完整的细节图像。将不同滤波器获取的$\boldsymbol{P}_{\mathrm{LH}}$图像与$\widetilde{\boldsymbol{M}_{\mathrm{H}}}$进行相关系数(Correlation Coefficient, CC)[19]计算。相关系数用来判断两幅图像之间的相关性,公式为

$$\mathrm{CC}(\boldsymbol{I},\boldsymbol{R})=\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_{ij}-\bar{I}\right)\left(R_{ij}-\bar{R}\right)}{\sqrt{\left(\sum_{i=1}^{M}\sum_{j=1}^{N}\left(I_{ij}-\bar{I}\right)^{2}\right)\left(\sum_{i=1}^{M}\sum_{j=1}^{N}\left(R_{ij}-\bar{R}\right)^{2}\right)}} \tag{7}$$

式中,$\boldsymbol{I}$表示原图像,$\boldsymbol{R}$为待对比图像,$\bar{I}$与$\bar{R}$表示图像均值。相关系数的值和两个图像的相关性成正比,其值越接近1,说明两幅图像相关性越强。表1记录了不同滤波器下的$\boldsymbol{P}_{\mathrm{LH}}$图像与$\widetilde{\boldsymbol{M}_{\mathrm{H}}}$的相关系数,其中高斯滤波的相似性最强。选择高斯滤波器对模型训练的图像进行预处理。为了进一步说明本文高斯滤波器有较好的适应性,进行基于高斯滤波器的图像融合实验。

表 1 $\boldsymbol{P}_{\mathrm{LH}}$与$\widetilde{\boldsymbol{M}_{\mathrm{H}}}$的相关系数

Table 1 CC for $\boldsymbol{P}_{\mathrm{LH}}$ and $\widetilde{\boldsymbol{M}_{\mathrm{H}}}$

Values	Laplace	Mean	Morphological	Gaussian
CC	0.051 2	0.104 2	0.178 1	**0.333 8**

针对不同的滤波器,在模型迭代次数一定的情况下进行了遥感图像融合。图7为高斯滤波器、拉普拉斯滤波器、均值滤波器、形态学滤波器得到的融合结果。从主观上判断,可以发现图7(a)、(d)有明显的细节注入效果,图7(b)、(c)细节信息缺失明显。同时,图7(a)与(d)相比,光谱信息保留更完全,图7(d)在放大区域可以明显发现物体边缘发生了较为明显的光谱扭曲。在表2中,采用6个参数对融合图像进行客观评价,分别为平均梯度(Average Gradient, AG)[17]、相对全局融合误差(Erreur Relative Global Adimensionnelle Synthese, ERGAS)[20]、光谱角(Spectral Angle Mapper, SAM)[21]、CC、均方根误差(Relative Average Spectral Error, RASE)[22]和通用图像质量评价指标(Universal Image Quality Index, UIQI)[23],可以发现高斯

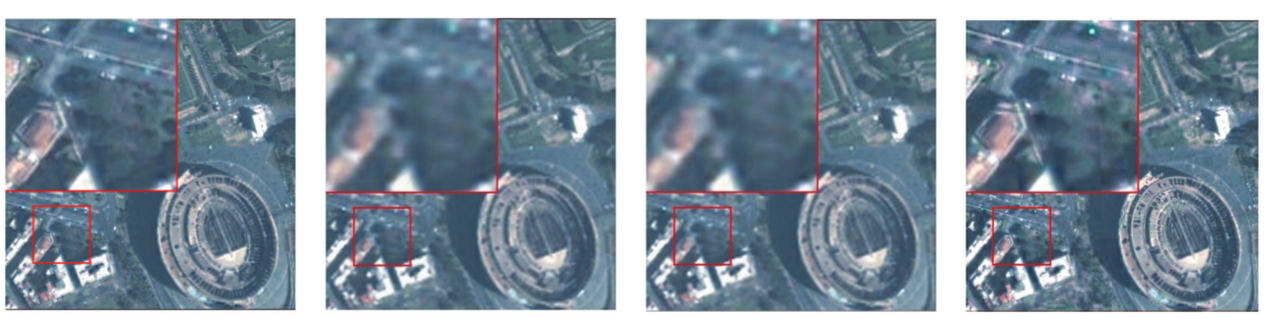


图 7 不同滤波器下的融合图像

Fig. 7 Fusion images with different filters

滤波器相比其他滤波器有更好的表现。

表2　不同滤波器下的融合图像指标

Table 2　Fusion image metrics with different filters

Values	Laplace	Mean	Morphological	Gaussian	Ideal
CC	0.723 1	0.681 8	0.899 0	**0.921 8**	1
SAM	2.119 0	2.304 3	3.765 8	**2.099 0**	0
ERGAS	8.921 6	8.150 0	5.452 6	**4.901 8**	0
UIQI	0.802 1	0.792 2	0.871 5	**0.893 0**	1
AG	0.016 5	0.029 3	**0.063 6**	0.062 9	1
RASE	8.690 2	8.732 9	10.222 0	**8.183 0**	0

4.3　迭代次数分析

在4.1节的滤波器选择中，模型迭代次数是一定的。卷积自编码器迭代次数不同，影响着模型学习图像细节结果。选取客观指标CC、ERGAS、UIQI和RASE作为本阶段实验的观测数值。如图8所示，ERGAS、RASE和CC表示图像中的光谱保持度，ERGAS和RASE值越小表明图像的光谱信息保留越完全，CC值越接近1表明图像相似度越高，图像保留信息越多；UIQI表示图像中保留的空间细节，值越接近1表明图像的空间细节信息保留更多，融合结果更好。图8中黑色曲线为融合图像的真实指标数值，绿色曲线为拟合曲线，拟合曲线更好地描述了指标的变化趋势，为迭代次数的选择提供了重要参考。在图8(a)中，CC指标总体在0.944到0.956之间浮动，由拟合曲线可以发现，指标在1 500次迭代之前迅速上升，在1 500次到1 750次时趋于稳定，在1 750次到2 000次有了下降趋势，说明迭代1 600次左右，图像的CC指标普遍最佳。在图8(b)中，RASE指标在6.4到7.2之间浮动，1 500次迭代之前，随着迭代次数的增加RASE指标逐渐下降，表

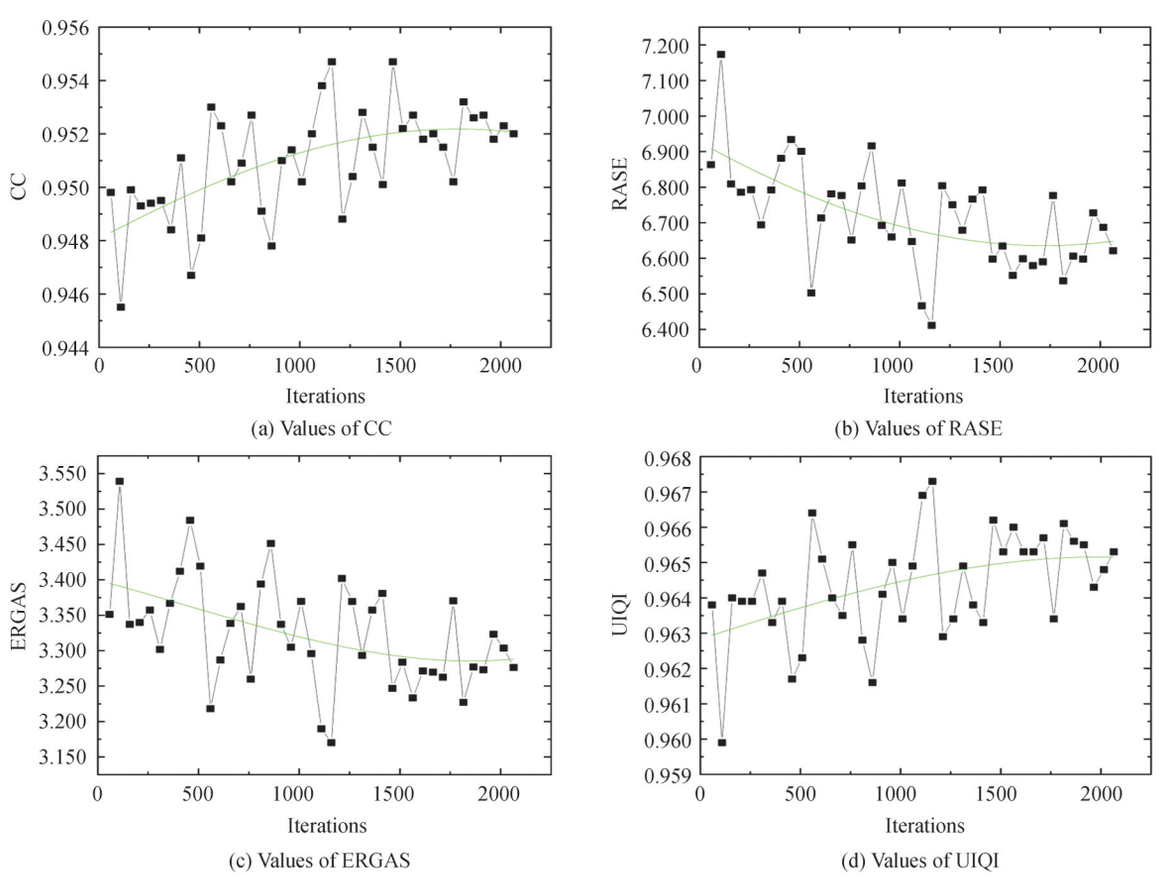


图8　不同迭代次数对应的图像指标值

Fig. 8　Image index values with different number of iterations

明图像逐渐保留更多的光谱信息，迭代1 500次到1 600次时趋于平稳，在1 600次之后RASE指标有上升趋势，可以看出，在迭代次数为1 550次左右时，RASE的值普遍更佳。在图8(c)中，ERGAS指标的变化范围在3.15与3.55之间浮动，1 500次迭代之前，ERGAS指标平稳下降，图像的融合质量逐步上升，并且在1 600次左右达到了拟合曲线的最低值，从1 750次之后指标值有了上升趋势，可以发现迭代次数在1 600左右图像的ERGAS指标值普遍较好。图8(d)中，UIQI值的浮动范围在0.95到0.97之间，可以发现，迭代次数在1 600次之前，UIQI指标值在平稳上升，并且在1 600次与1 900次之间拟合曲线趋势平稳，没有明显变化，在1 900次之后，拟合曲线出现了下降趋势，表明迭代1 700次左右时UIQI值普遍较好。结合图8，以每幅图的拟合曲线作为重要参考，可以看出合适的迭代次数在1 600次左右，此时图像的各项指标普遍较好，获取的融合图像较佳。

4.4 实验结果分析

为了证明算法的有效性，将本算法与其他8种融合算法进行比对实验。这8种融合算法分别为IHS算法[1]，BDSD算法[19]，MTF-GLP-HPM[24]算法，基于卷积神经网络的融合算法(PNN)[25]，基于卷积神经网络的细节注入算法(Di-PNN)[12]，基于稀疏表示的细节注入算法(SR-D)[11]，基于对抗生成网络的融合算法(GAN)和基于卷积自编码器网络的融合算法(CAE)[26]。

图9为QuickBird卫星获取的意大利Roma原图像融合结果和局部放大图。图9(a)为多光谱图像，分辨率为2.44 m。图9(b)为全色图像，分辨率为0.61 m。多光谱图像与全色图像大小均为512×512像素。图9(c)～(k)为不同算法下的融合结果。图9(c)为利用IHS变换获取的融合图像，在IHS变换中采用全色图像

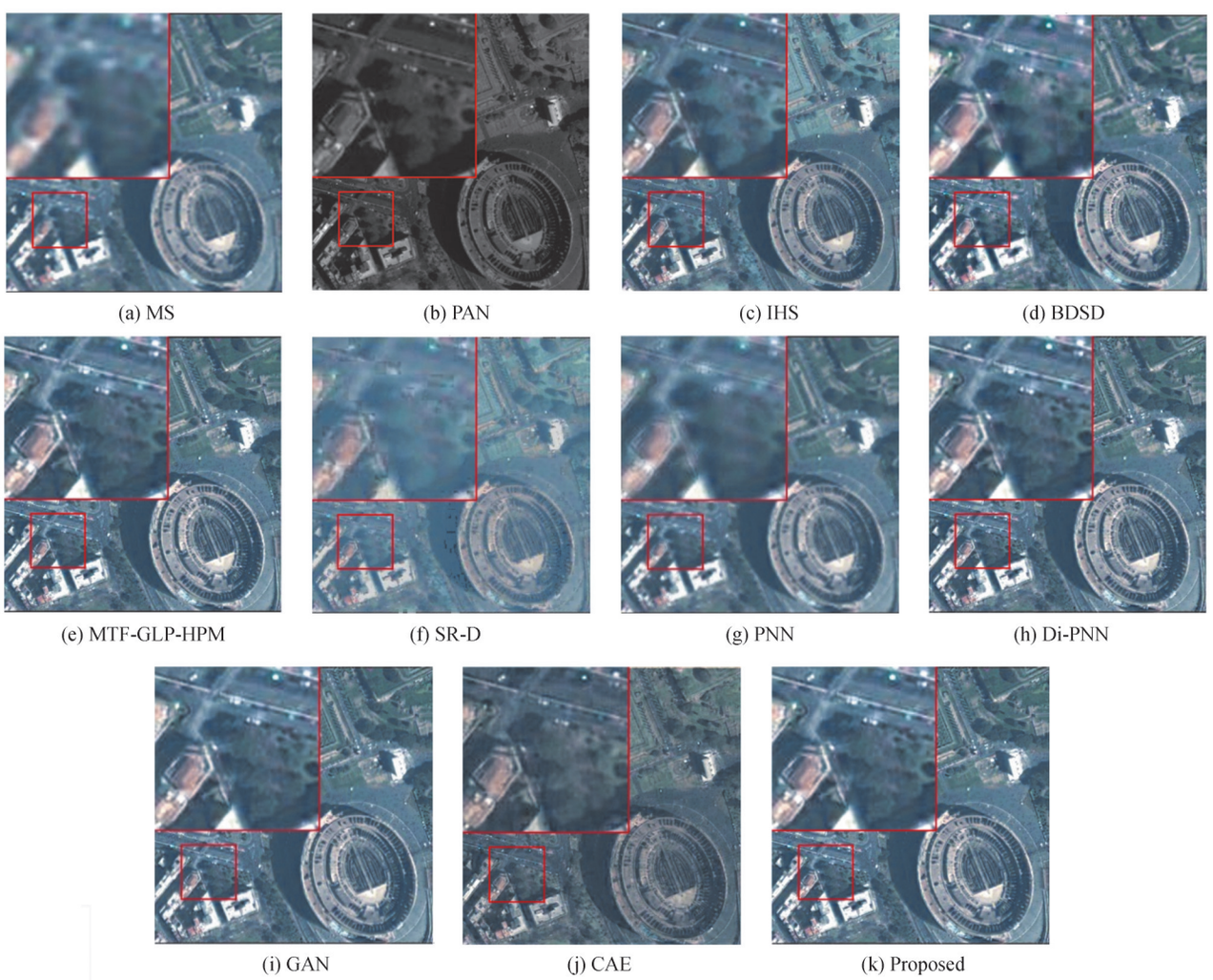


图9 Roma原图像及不同算法的融合结果

Fig. 9 Roma source images and fusion result by different methods

替换多光谱图像的强度分量以增强多光谱图像中的细节信息。可以发现，IHS变换在增强细节信息的同时，也导致了图像光谱信息的失真。从图9(c)局部放大图中可以发现树的颜色发生了失真，由绿色变成了淡绿色。图9(d)是BDSD算法的融合结果，相比IHS算法，BDSD算法对融合进行了适应性选择，保留了更多的光谱信息，但是在图9(d)放大区域中仍然存在斑点光谱失真的现象。图9(e)是MTF_GLP_HPM算法的融合结果，该算法保留了好的光谱信息与空间细节信息。可以发现，MTF_GLP_HPM算法获取的融合图像在建筑边缘存在伪影，在图像的阴影部分也出现了细节扭曲的情况。图9(f)是SR-D算法的融合结果，可以观察到图像的黑暗部分存在明显的噪声。因此，SR-D算法在提升细节信息和光谱信息的同时，也为融合结果也引入了噪声。PNN算法首次将深度学习引入遥感图像融合领域，图9(g)为PNN算法的融合结果。与多光谱图像相比，该算法在保留光谱信息的同时，加入了全色图像的细节信息。由于PNN融合算法采用三层卷积操作，在光谱信息和细节信息上仍有缺失，从图9(g)可以发现，PNN融合算法的结果图像细节信息保留不完全，并存在伪影，融合效果仍有大的进步空间。为了更好地提高PNN的融合效果，Di-PNN算法被提出，Di-PNN在细节信息保留上相比PNN算法有很大的提高。图9(h)为Di-PNN算法的融合结果，与图9(g)相比，在细节信息上有了更大的提高，保留了更多的空间细节信息与丰富的光谱信息。图9(i)为GAN算法的融合结果，主观上较图9(c)～(h)有了大的进步，但是在光谱信息保留上相比多光谱图像仍有缺失。在图9(i)局部放大图中，可以发现树木的颜色较淡，图像的阴影部分也存在一些白色噪声。图9(j)是CAE算法的融合结果，与PNN不同的是CAE采用编码解码网络，比PNN有更强的空间细节提取能力和光谱保持能力。但是图9(j)图像整体偏暗，融合结果的光谱信息缺失，因此在光谱信息保持上，CAE网络仍可以改善。受限于深度学习网络的端对端学习，PNN、GAN和CAE存在不同程度的光谱失真情况。本文提出算法的结果为图9(k)，与全色图像和多光谱图像相比，草木、建筑和海洋的边缘细节比较清晰，光谱信息和空间细节信息得到了较好的保持，主观视觉较好。

表3给出了QuickBird卫星遥感图像的客观评价指标。可以看出IHS各项指标处于劣势，BDSD算法加入MS的适应性选择后，分量替换算法的指标得到了明显提升。MTF-GLP-HPM算法与PNN算法各项指标相似，说明两者融合结果客观接近。SR-D融合算法的RASE、UIQI和AG指标较低说明图像的细节信息丢失严重，同时光谱信息也存在丢失。与PNN相比，Di-PNN在指标RASE和UIQI上有所改善，说明细节注入方式保留更多的空间细节信息，SAM、ERGAS和CC指标说明Di-PNN融合方式对光谱有较好的保存能力。GAN算法与本文算法各项指标相似，融合能力相近。但是SAM和ERGAS指标表明本文算法有较好的光谱保存能力。同时CAE算法融合的图像在RASE指标和ERGAS指标上偏高，说明图像的光谱与细节存在缺失，本文算法弥补了这些缺点并获得了较好的指标。综上，本文提出的融合算法较优于参考算法。

表3　Roma原图像融合结果性能比较

Table 3　Performance comparison of fusion results of Roma source images

Methods	ERGAS	RASE	SAM	UIQI	AG	CC
IHS	7.885 7	15.851 7	1.611 0	0.754 3	0.029 5	0.754 5
BDSD	4.756 2	9.518 1	2.188 5	0.927 1	0.028 2	0.932 2
MTF-GLP-HPM	4.643 5	9.874 5	1.433 4	0.930 3	0.036 5	0.935 4
SR-D	7.368 2	27.234 2	1.788 9	0.795 7	0.019 9	0.846 6
PNN	4.654 8	9.781 9	1.516 7	0.901 1	0.031 0	0.928 1
Di-PNN	4.086 5	8.568 9	1.361 5	**0.942 2**	0.036 0	0.921 1
GAN	3.354 7	**8.056 9**	3.056 9	0.923 5	0.041 4	**0.943 5**
CAE	10.315 5	38.779 0	2.922 5	0.650 0	0.029 8	0.673 9
Proposed	**3.242 8**	8.137 2	**1.316 7**	0.929 1	**0.043 6**	0.934 1
Ideal	0	0	0	1	1	1

图10为SPOT卫星在United Arab Emirates区域获取的原图像和融合图像的局部放大图。图10(a)为多光谱图像，其分辨率为6 m。图10(b)为对应的全色图像，分辨率为1.5 m。多光谱图像与全色图像大小均

为512×512像素。通过观察图10可以发现SPOT卫星的融合图像拥有与QuickBird卫星融合图像相似的结果。图10(c)局部放大区域中,房屋的红色房顶颜色变浅,并且草坪颜色也发生了光谱失真。图10(d)中BDSD算法有明显的细节信息丢失,建筑边缘部分模糊。图10(e)在图像的边缘部分同样也有细节丢失的情况。在图10(e)中右上角部分建筑出现了细节信息错位的情况,不容易区分房屋界限。图10(f)放大区域中可以明显发现SR-D算法存在空间细节丢失,边缘信息不完全的情况。PNN受限于浅层网络的学习能力,在图10(g)中均存在空间信息保留不完全,区域信息模糊,光谱信息丢失的情况。Di-PNN对PNN算法有了更好的优化,图10(h)在空间细节保留能力上有明显的提高,但是仍然存在细节缺失问题,融合效果还有待进一步提升。图10(i)有较好的视觉融合结果,在空间细节信息与光谱信息的保持上比上述算法有明显进步。在图10(j)中可以发现基于卷积自编码器网络的融合结果空间细节能力保持不完全,光谱信息失真,在房屋边缘部分由黄色变成了墨绿色,光谱信息扭曲。本文算法得到的图像在空间细节保留和光谱信息保存上有较好的改善。

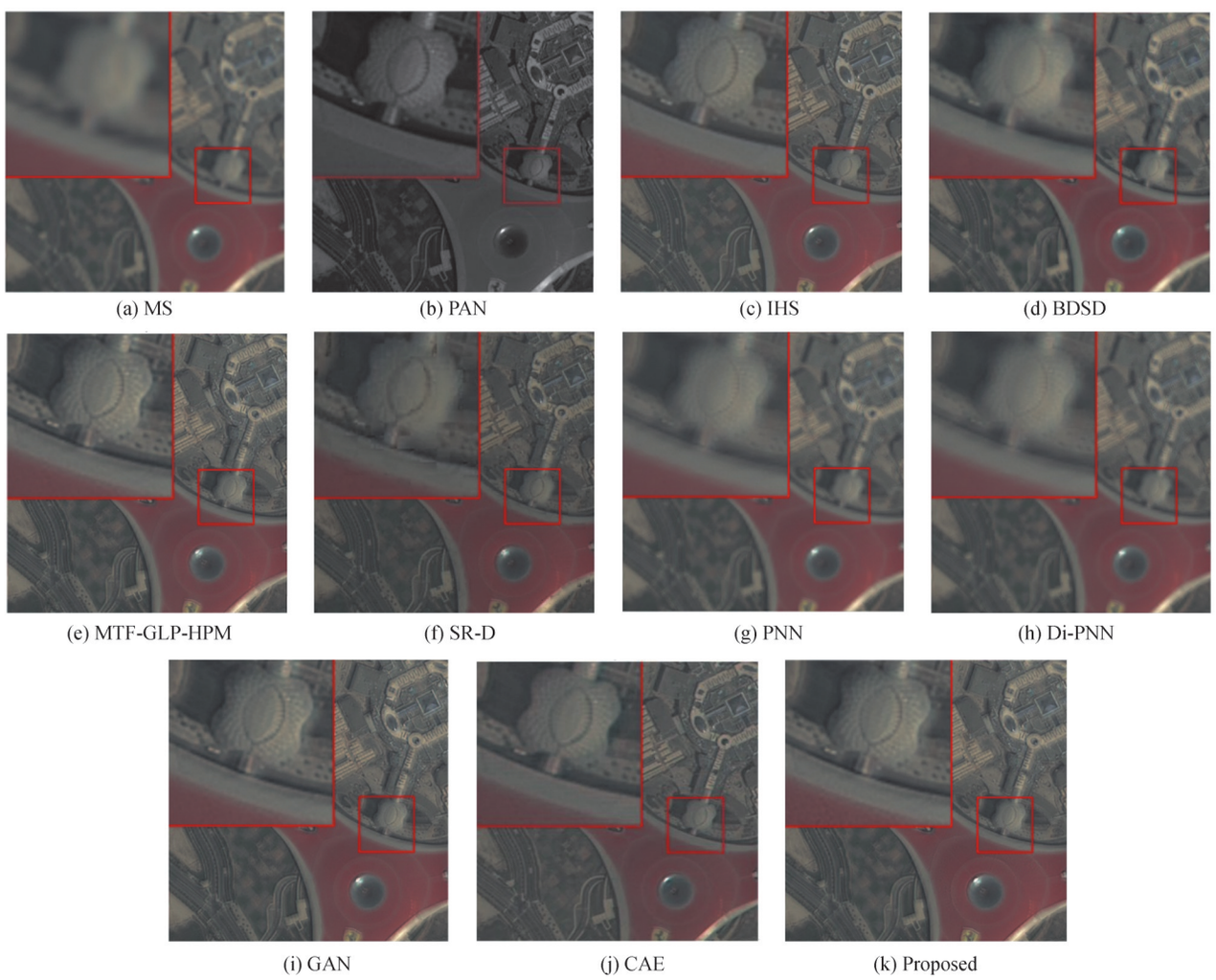


图10 United Arab Emirates原图像及不同算法的融合结果

Fig. 10 United Arab Emirates source images and fusion result by different methods

表4给出了SPOT卫星遥感图像的客观评价指标。其中IHS融合算法UIQI指标较好,表明其保留细节信息的能力,而其他指标均处于劣势,表明IHS算法在光谱信息上的缺失。BDSD算法指标的提升表明其在各项能力上优于IHS融合算法。MTF-GLP-HPM算法与PNN算法各项指标的差距很小,说明两个算法在SOPT卫星上表现能力相似,但在主观视觉上MTF-GLP-HPM算法有更优的表现能力。SR-D融合算法SAM指标偏高,光谱信息存在丢失。PNN算法在指标上没有突出的表现,Di-PNN算法指标的提升表明细

节注入对PNN算法改良的有效性。GAN算法与本文算法虽然在主观视觉上有强的相似性,但是本文算法有较好的指标,保留了更多的光谱信息与细节信息。同时本文算法相比CAE融合算法也有较好的表现能力。

表4 United Arab Emirates原图像融合结果性能比较

Table 4 Performance comparison of fusion results of United Arab Emirates source images

Methods	ERGAS	RASE	SAM	UIQI	AG	CC
IHS	5.562 4	12.301 5	0.450 1	0.842 1	0.014 4	0.841 2
BDSD	2.551 6	5.454 6	0.965 9	0.879 8	0.009 2	0.924 9
MTF-GLP-HPM	4.124 0	9.647 0	0.970 6	0.919 4	**0.016 8**	0.923 2
SR-D	8.411 7	15.447 9	1.251 5	0.859 4	0.011 9	0.886 0
PNN	4.096 8	9.601 9	1.134 1	0.920 0	0.016 7	0.924 3
Di-PNN	3.432 9	5.646 8	1.362 1	0.941 8	0.007 7	0.939 8
GAN	3.757 3	8.649 6	0.992 2	0.932 7	0.015 5	0.936 0
CAE	6.745 2	19.509 5	2.009 4	0.766 6	0.014 5	0.773 2
Proposed	**1.078 9**	**4.757 8**	**0.264 1**	**0.953 9**	0.007 7	**0.944 0**
Ideal	0	0	0	1	1	1

5 结论

本文采用结合卷积自编码器、卷积注意模块和高斯滤波器的算法,将原始图像进行滤波处理,获取低分辨率高频图像与高分辨率高频图像,接着将低分辨率高频图像重建到高分辨率高频图像作为卷积自编码器的学习目标,并通过卷积注意模块对特征信息自适应细化,提高卷积注意模块对关键信息的保留,最终训练完成的卷积自编码器能够预测更完整的图像缺失细节信息。通过采用细节注入的方式,克服了深度学习中端对端学习方式对光谱信息的损失。实验表明,本文提出的遥感图像融合算法取得较好的视觉效果和客观指标,获取的融合图像保持了好的光谱信息和空间细节信息。

参考文献

[1] VIVONE G, ALPARONE L, CHANUSSOT J, et al. A critical comparison among pansharpening algorithms[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(5): 2565-2586.

[2] VIVONE G, MURA M, GARZELLI A, et al. new benchmark based on recent advances in multi-spectral pansharpening: revisiting pansharpening with classical and emerging pansharpening methods[J]. IEEE Geoscience and Remote Sensing Magazine, 2021, 9(1): 53-81.

[3] YEE L, JUNMIN L, JIANGSHE Z. An improved adaptive intensity - hue - saturation method for the fusion of remote sensing images[J]. IEEE Geoscience and Remote Sensing Letters, 2014, 11(5): 985-989.

[4] VP S, NH Y, RL K. An efficient pan-sharpening method via a combined adaptive PCA approach and contourlets[J]. IEEE Transactions on Geoscience and Remote Sensing, 2008, 46(5): 1323-1335.

[5] JAVAN F D, SAMADZADEGAN F, MEHRAVAR S, et al. review of image fusion techniques for pan-sharpening of high-resolution satellite imagery[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 171: 101-117.

[6] VIVONE G, RESTAINO R, CHANUSSOT J. A bayesian procedure for full-resolution quality assessment of pansharpened products[J]. IEEE Transactions on Geoscience and Remote Sensing, 2018, 56(8): 4820-4834.

[7] PEI Xiaopeng. Remote sensing image fusion based on sparse representation[D]. Taiyuan: Taiyuan University of Technology, 2018.
裴晓鹏.基于稀疏表示的遥感图像融合算法[D].太原:太原理工大学, 2018.

[8] 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.

[9] OZCELIK F, ALGANCI U, SERTEL E, et al. Rethinking CNN-based pansharpening: guided colorization of panchromatic images via GANS[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 59(4): 3486-3501.

[10] YANG Y, LU H, HUANG S, et al. An efficient and high-quality pansharpening model based on conditional random fields [J]. Information Sciences, 2020, 553(1): 1-18.

[11] RONGRONG F, JIANGSHE Z, JUNMIN L, et al. Convolutional sparse representation of injected details for pansharpening[J]. IEEE Geoscience and Remote Sensing Letters, 2019, 16(10): 1595-1599.

[12] LIN H, YIZHOU R, JUN L, et al. Pansharpening via detail injection based convolutional neural networks[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2019, 12(4): 1188-1204.
[13] ITO K, XIONG K. Gaussian filters for nonlinear filtering problems[J]. IEEE Transactions on Automatic Control, 2000, 45(5): 910-927.
[14] JONATHAN M, UELI M, DAN C, et al. Stacked convolutional auto-encoders for hierarchical feature extraction[C]. Artificial Neural Networks and Machine Learning - ICANN 2011. Springer, 2011: 52-59.
[15] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
[16] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
[17] LIU Fan. Remote sensing image fusion based on wavelet kernel filter and sparse representation[D]. Xi'an: Xidian University, 2014.
刘帆. 基于小波核滤波器和稀疏表示的遥感图像融合[D]. 西安:西安电子科技大学, 2014.
[18] BERA S, SHRIVASTAVA V K. Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification[J]. International Journal of Remote Sensing, 2020, 41(7): 2664-2683.
[19] MENG X, SHEN H, LI H, et al. Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: practical discussion and challenges[J]. Information Fusion, 2019, 46: 102-113.
[20] VIVONE, G, DALLA M, GARZELLI A, et al. A benchmarking protocol for pansharpening: dataset, preprocessing, and quality assessment[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 6102-6118.
[21] ZHANG Jing. Remote sensing image fusion based on sparse representation for detail enhancement[D]. Taiyuan: Taiyuan University of Technology, 2020.
张静. 基于稀疏表示进行细节增强的遥感图像融合[D]. 太原:太原理工大学, 2020.
[22] ZHANG Jing, CHEN Hongtao, LIU Fan. Remote sensing image fusion based on multivariate empirical mode decomposition and weighted least squares filter[J]. Acta Photonica Sinica, 2019, 48(5):0510003.
张静, 陈宏涛, 刘帆. 结合多元经验模态分解和加权最小二乘滤波器的遥感图像融合[J]. 光子学报, 2019, 48(5): 0510003.
[23] ZHOU W, BOVIK A C. A universal image quality index[J]. IEEE Signal Processing Letters, 2002, 9(3): 81-84.
[24] GARZELLI A, NENCINI F, CAPOBIANCO L. Optimal MMSE pan sharpening of very high resolution multispectral images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2008, 46(1): 228-236.
[25] MASI G, COZZOLINO D, VERDOLIVA L, et al. Pansharpening by convolutional neural networks[J]. Remote Sensing, 2016, 8(7): 594.
[26] AZARANG A, MANOOCHEHRI H E, KEHTAR NAVAZ N. Convolutional autoencoder-based multi-spectral image fusion[J]. IEEE Access, 2019, 7: 35673-35683.

Combining Convolutional Attention Module and Convolutional Auto-encoder for Detail Injection Remote Sensing Image Fusion

LI Ming, LIU Fan, LI Jingzhi
(*College of Data Science, Taiyuan University of Technology, Jinzhong, Shanxi* 030600, *China*)

Abstract: Panchromatic and multispectral images can be captured by Earth observation satellites. Usually, panchromatic images have high spatial resolution and low spectral resolution, while multispectral images have low spatial resolution and high spectral resolution. For combining the spatial and spectral information of panchromatic and multispectral images, remote sensing image fusion techniques are applied and born. Although significant progress has been made in fusion algorithms, there are still problems of spectral distortion and insufficient details. To solve the above problems, this paper proposes to design a new remote sensing image fusion algorithm with convolutional auto-encoders, attention mechanism and filter as the detail processing module and additive injection fusion rule as the fusion module. Convolutional auto-encoders learns the nonlinear mapping relationship between the low-resolution image and the high-resolution image, and the high-resolution image corresponding to the low-resolution image can be obtained

after the training is completed. The introduction of attention Mechanism in the convolutional auto-encoders can improve the sensitivity of the network to information and increase the channel importance of image information. The filter plays two roles in this paper, one is to obtain the high or low frequency information of the image through the filter, and the other is to obtain the low-resolution image corresponding to the high-resolution image. The specific steps are described below. First, high-frequency images of low-resolution images and high-frequency images of high-resolution images for model training are acquired separately using Gaussian filters, while high-frequency images of low-resolution multispectral image for model prediction are acquired; then, the non-linear mapping relationship between the high-frequency image of low-resolution image and the high-frequency image of high-resolution image is learned by using convolutional auto-encoders; finally, the missing detail information of the multispectral image, i.e., the high-frequency image of high-resolution multispectral, is obtained using the convolutional auto-encoders completed by training, and fused with the original image to generate the high-resolution multispectral image. For the filter selection, experiments are conducted based on the mean filter, Laplace filter, Gaussian filter and morphological filter in this paper, and the results show that using the Gaussian filter has a better fusion effect. At the same time, experiments were conducted on the selection of the number of iterations of the network model. In this paper, the objective metrics of fused images with the different number of iterations are recorded. Since the objective indicators are floating in nature, a fitting function is used to fit the data to the objective indicators. The influence of the number of iterations on the fusion results is found by observing the trend of the fitting curve. The fitting curves show that the fusion algorithm proposed in this paper obtains the best fused image at about 1 600 iterations. This paper combines the respective advantages of Convolutional Auto-Encoders, attention mechanism and filter to perform experiments on two datasets, which are images taken by QuickBird and SPOT satellites, respectively. The resolution of the datasets is 512×512 for multispectral and 512×512 for panchromatic images. To expand the training dataset, the datasets are cropped to 8×8 size images by using a sliding window. In training the model training batch size is 256, the number of training iterations is 1 600, and the optimizer Adadelta is used for network model parameter optimization and learning rate adaptive optimization. To demonstrate the effectiveness of the algorithm proposed in this paper, it is compared with the classical fusion algorithm. Since this paper uses the additive injection of fusion rules, IHS and BDSD additive fusion algorithms are selected for comparison. PNN and GAN are typical deep learning fusion algorithms and are compared with classical deep learning fusion algorithms to demonstrate the effectiveness of the proposed fusion algorithm. The comparison with the CAE fusion algorithm can effectively prove the effectiveness of the attention mechanism and filter introduced in this paper, which can significantly improve the fused image effect. Di-PNN fusion algorithm and SR-D fusion algorithm are both detail injection fusion algorithms based on deep learning networks, and the comparison with Di-PNN and SR-D fusion can illustrate the effectiveness of the network structure in this paper. In this paper, the results of different fusion algorithms are compared in terms of subjective visual and objective metrics. The objective metrics are CC, UIQI, ERGAS, RASE, AG and SAM, where the UIQI and AG metrics describe the detail information of the image, and the ERGAS, RASE, SAM and CC metrics describe the spectral information of the image. the larger the CC, UIQI and AG metrics, the better the image quality; the smaller the ERGAS, RASE and SAM metrics, the better the image quality. By comparing with the classical fusion algorithm and using subjective visual and objective metrics, the experimental results show that the fused images in this paper retain more spectral information and detail information and show good performance both subjectively and objectively.

Key words: Remote sensing image fusion; Feature extraction; Convolutional auto-encoders; Multispectral images; Panchromatic images

OCIS Codes: 100.2000; 350.2660; 070.2615; 100.3010; 110.4234

Foundation item: National Natural Science Foundation of China (No. 61703299)