第 58 卷 第 22 期/2021 年 11 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于增强超分辨率网络的单对图像时空融合

李奇泽，何超琦，魏静波*

南昌大学信息工程学院，江西 南昌 330031

摘要 由于高质量的对地观测需要时空连续的高分辨率遥感图像，故对时空融合的研究广泛开展，并且集中在 Landsat 和 MODIS 卫星之间。目前已经提出了使用卷积神经网络进行时空融合的方法，但是网络较浅，故融合性能有限。针对应用最广泛的单对图像时空融合问题，建立了一种基于深度神经网络的新时空融合方法。首先，基本网络框架由两个级联的 4 倍上采样器构成以近似 Landsat 和 MODIS 卫星之间的空间差异和传感器差异。然后，利用卷积神经网络学习重建图像与真实图像之间的残差，使重建图像与真实图像更接近。接着，使用高通调制策略进行时间上的预测。最后，将所提方法在不同的 Landsat 和 MODIS 卫星图像上进行了测试，并与多种时空融合算法进行了比较。实验结果表明，与现有融合算法相比，所提方法的重建效果更好，且处理速度更快。

关键词 遥感；卷积神经网络；深度残差网络；时空融合；Landsat

中图分类号 TP751 **文献标志码** A **doi**：10.3788/LOP202158.2228006

Spatiotemporal Fusion of One-Pair Image Based on Enhanced Super-Resolution Network

Li Qize, He Chaoqi, Wei Jingbo*

Information Engineering School, Nanchang University, Nanchang, Jiangxi 330031, *China*

Abstract Due to high-quality earth observation requires spatiotemporal continuous high-resolution remote sensing images, the research on spatiotemporal fusion is widely carried out and focused on Landsat and MODIS satellites. At present, the method of spatiotemporal fusion using convolutional neural networks has been proposed, but the network is shallow, so the fusion performance is limited. Aiming at the most widely used one-pair image spatiotemporal fusion, a new spatiotemporal fusion method based on deep neural network is established. Firstly, the basic network framework consists of two cascaded upsamplers with quadruple magnification to approximate the spatial difference and sensor difference between Landsat and MODIS satellites. Then, the residual error between the reconstructed image and the real image is learned by the convolutional neural network to make the reconstructed image closer to the real image. Moreover, the time prediction is carried out by highpass moduation strategy. Finally, the proposed method is tested on different Landsat and MODIS satellite images and compared with many spatiotemporal fusion algorithms. The experimental results show that, compared with the existing fusion algorithms, the reconstruction effect of the proposed method is better and the processing speed is faster.

Key words remote sensing; convolutional neural networks; deep residual networks; spatiotemporal fusion; Landsat

OCIS codes 280.4788; 280.4750; 280.4991

1 引 言

卫星图像已成为大区域对地观测的主要数据来源。其中，Landsat 系列搭载的多光谱传感器的地面分辨率为 30 m×30 m，重访周期为 16 天，适用于观测森林、农作物、湿地、土壤、水等自然资源。但

收稿日期：2020-12-08；**修回日期**：2021-01-08；**录用日期**：2021-01-27

基金项目：国家自然科学基金(61860130)

通信作者：*wei-jing-bo@163.com

是，大多数监测数据受到云或雨的影响而无法使用，常常要等待数月才能得到一张 Landsat 图像，故难以及时估计植物等生长状况。一种相对经济的解决方案是使用时空融合算法，将不同时间和空间分辨率的图像进行组合，充分利用不同遥感源中时空信息之间的互补性，以获得具有高空间分辨率和高时间分辨率的合成图像。

进行时空融合至少需要知道一对完整数据。由于低分辨率图像常常具有较短的重访周期（对应较高的时间分辨率），因此卫星可能在 t_1 和 t_2 时刻都拍摄了图像。然而，高分辨率图像具有较长的重访周期，因此可能只有 t_1 时刻的图像而没有 t_2 时刻的图像。时空融合就是利用 t_1 时刻的高低分辨率图像对和 t_2 时刻的低分辨率图像来预测 t_2 时刻缺失的高分辨率图像的方法。在一些算法中，至少需要两组已知时刻的图像对。首先，根据 t_1 和 t_3 时刻的两组高低分辨率图像对建立映射关系。然后，输入 t_2 时刻的低分辨率图像以合成 t_2 时刻缺失的高分辨率图像。

目前，在一些时空融合算法中要求输入多对高低分辨率图像[1]，如文献[2-3]中的方法，并且在选择参考图像时要求时间相近。然而，由于在实际情况中存在研究区域的不利天气因素和配准困难等原因，故获取多个成对遥感数据是非常困难的。因此，考虑到方法的一般性，本文研究了单对图像的时空融合方法，其是多对图像融合的基础。根据优化策略，现有的面向单对或最多两对已知图像的时空融合方法可以分为：基于加权的方法、基于解混的方法、基于字典对学习的方法和基于神经网络的方法。

在基于加权的方法中，假定类别或相邻位置的像素共享相似的像素值。首先，定义已知图像中相似像素的位置。然后，基于相似性规则对预测时间的低分辨率图像中的像素进行加权，以预测新像素。时空自适应反射率数据融合模型（STARFM）[4]、增强型的 STARFM（ESTARFM）[5] 和映射反射率变化的时空自适应（STAARCH）算法[6] 是典型的基于加权的方法。其中，STARFM 的混合权重由光谱差异、时间差异和位置距离确定。

基于解混的方法可以融合时空图像的原因是其可以使用已知时刻的高分辨率图像来计算端元的丰度矩阵。Wu 等[7] 提出了一种时空数据融合算法（STDFA），该算法可以提取分类覆盖图，并利用最小二乘法预测了类别的平均反射率。Xu 等[8] 提出了一种混合方法，该方法利用先验光谱以平滑每个分类中利用 STARFM 得到的预测图像。Zhu 等[9] 提出了一种柔性时空数据融合（FSDAF）算法，该算法中利用薄板样条插值的方法来预测图像。

在基于字典对学习的方法中，利用字典学习和非解析优化的方式在稀疏域中预测缺失图像[10]。对于使用相似传感器同时捕获的图像，如果使用精心设计的字典，则它们的稀疏编码系数可能非常相似。当用设计后的字典对其他时间的图像对进行编码时，这些图像的编码系数仍具有相似性。基于稀疏表示的时空反射融合模型（SPSTFM）[11]、误差有界半耦合字典学习（EBSCDL）[12] 和快速迭代收缩阈值算法（FISTA）[13] 是典型的基于字典对学习的方法。本课题组也研究了该主题，并提出了基于稀疏贝叶斯学习[14] 和压缩感知[15] 的时空融合方法。

卷积神经网络（CNN）和深度学习已被用于时空融合。Dai 等[16] 提出了一种两层融合策略，在每一层中都使用 CNN 来学习图像之间的非线性映射关系。Song 等[3] 提出了两个 5 层 CNN，以解决 MODIS 与 Landsat 图像之间的对应关系复杂和空间分辨率差距大的问题。在预测阶段，文献[3]、[16]设计了一个融合模型，该模型中采用了高通调制和加权的方法以充分利用先前图像中的信息[10]。

虽然利用 CNN 可以改善时空融合算法，但改善后的算法并没有表现出能够超越现有算法的能力。在相关的研究文章中，并没有将新提出的算法与较新的字典学习类算法（如 EBSCDL）进行比较。但是，结合所有算法与传统基准算法 STARFM 的比较结果，可以间接判断出这些利用 CNN 改善后的算法的效果尚达不到 EBSCDL 的时空融合效果。

经过分析，认为现有使用 CNN 进行时空融合的算法性能不佳的原因有两点。一方面，现有 CNN 算法的网络复杂度低，无法有效实现空间差异建模。现有算法的网络架构基础通常来自超分辨卷积神经网络（SRCNN），随后在此基础上进行扩展，但最终的网络层数不会超过 5 层。要想改善重建质量，需要进一步增加网络层数以实现较为复杂的映射过程。另一方面，现有的基于 CNN 的算法仅仅使用了一个网络同时解决空间和时间上的融合问题，没有单独考虑不同卫星传感器之间的差异。如果能将传感器之间的差异进行单独建模，则会提升重建结果质量。

在本文中，提出了一种使用深度神经网络进行时空融合的新方法，它是通过对单对时空图像中的

空间差异和传感器差异分别建模并耦合在一起，进而实现时空融合的。首先，对增强型深度超分辨率(EDSR)网络[17]进行训练以建模空间差异和传感器差异。然后，使用 EDSR 结果与真实遥感图像对 CNN 进行有限地训练以建模传感器差异进行残差修正。此外，整个过程中只使用了一对参考图像进行训练。

2　所提方法

本节介绍了一种新的时空融合算法，该方法首先利用深度神经网络对不同卫星图像的空间差异和传感器差异进行建模。然后，对重建结果进行残差修正。最后，利用高通调制方法进行时间上的预测。

本文的目标是利用已知 t_1 时刻的 Landsat 图像 L_1、t_1 时刻的 MODIS 图像 M_1 和 t_2 时刻的 MODIS 图像 M_2，来预测 t_2 时刻的 Landsat 图像 L_2。图 1 展示了所提时空融合算法的流程。首先，使用 EDSR 网络模型来建模 MODIS 和 Landsat 图像间的空间分辨率差异和传感器差异，此时的输入为 M_1 和 M_2，输出为上采样 16 倍后的结果 L_1' 和 L_2'。然后，将 L_1' 和 L_2' 输入到设计好的 CNN 中进行残差修正，以得到较接近真实 Landsat 图像的结果 T_1^{L} 和 T_2^{L}。最后，利用高通调制方法对 t_1 到 t_2 时刻的时间变化信息进行预测，得到 t_2 时刻的预测结果 L_2。

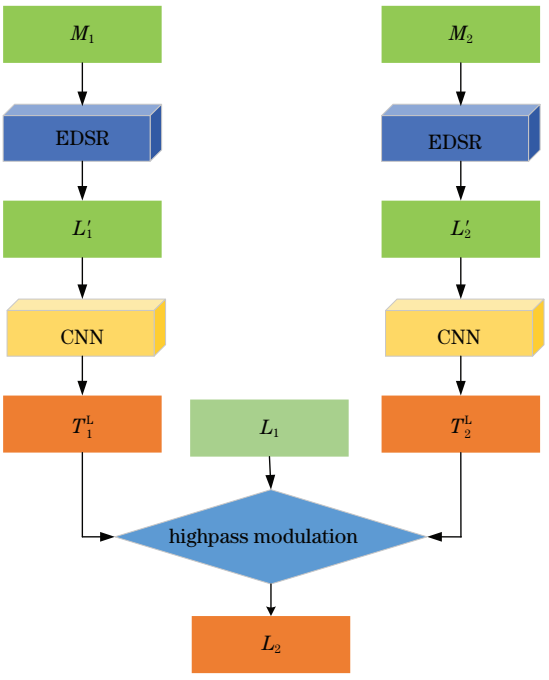

图 1　所提方法的流程图

Fig. 1　Flow chart of proposed method

2.1 空间差异和传感器差异网络

单对图像时空融合的核心任务是超分辨率。以 Landsat-7 和 MODIS 为例，卫星的飞行高度一致，拍摄时刻集中在当地时间的 10:00～10:30，且平台搭载的红、绿、蓝、近红外传感器的光谱范围很接近，故两种卫星的最主要差别是空间分辨率。

MODIS 的空间分辨率为 500 m×500 m，而 Landsat-7 的空间分辨率为 30 m×30 m，这不利于卷积核的设计。由于典型的卷积核大小都是整数，故此处使用 16 倍上采样器来近似地重建目标图像。

针对 Landsat 和 MODIS 这样大分辨率差异的时空融合问题，设计一个具有 16 倍上采样功能的融合模型是非常具有挑战性的。通常，典型的上采样器只具有 4 倍的最大分辨率，直接训练一个 16 倍上采样器会面临训练难以收敛的问题。为了实现具有高重建性能的 16 倍上采样器，将两个 4 倍上采样器进行了级联。

通过使用残差块成功在残差网络(ResNet)中加深了传统的神经网络。EDSR 网络移除了批处理标准化操作后，较适合进行图像重建，在解决底层视觉任务方面表现出强大的性能。因此，设计的 4 倍上采样网络借鉴了 EDSR 超分辨率网络的网络架构。

除了空间分辨率上的差异，光谱响应不同、光学弥散造成的解析力不一致、拍摄时间对应的太阳高度角和地物反射率不同等因素均会造成不同卫星传感器之间产生很大的差异。由于在实际中使用的两颗卫星过境时间差别在半小时以内，并且每次的准确过境时间是由飞控系统控制的，即过境时间具有不确定性，因此拍摄时间对应的太阳高度角和地物反射率不同导致的差异是无法建模的。此外，光谱响应不同和光学弥散造成的解析力下降与地物内容有关，且光谱响应不同和光学弥散问题无法套用统一公式进行修正。CNN 自发展以来，展现出了很强的自学习能力和表征能力，可以学习已有图像中的相似地物场景来尽可能地表征传感器之间的差异。

图 2 为用于建模空间差异和传感器差异的级联 EDSR 网络。该网络主要包含卷积层 Conv、64 个残差单元 ResBlock 和两个上采样模块 Upsample。每个残差单元由两个 Conv、一个线性整流单元 ReLU 和一个恒等比例缩放层 Mult 组成，且残差单元中没有使用批归一化层 BN。Upsample 模块由两个 Conv 和两个 Shuffle 层组成，其中 Shuffle 层的作用是将低分辨率特征图进行有效放大。每个

Upsample 模块可进行 4 倍上采样操作，网络中通过两个 Upsample 模块完成了 16 倍上采样操作。

在 EDSR 网络训练过程中，训练集的输入和输出分别为真实的 MODIS 和 Landsat 图像，共使用 1296 个图像对进行训练。同时，在训练过程中使用早停策略，该策略的作用是当验证集的精度在一定周期内不再增长或者下降时，及时停止网络训练。因此，使用早停策略能让网络在保持当前数据量和网络结构的情况下，在最短的时间内较好地完成训练数据拟合，避免网络复杂而训练数据有限所造成的过拟合。

网络中每个卷积层的卷积核大小为 3×3，特征通道数量为 256。残差单元中将 ReLU 作为激活函数。网络的输入图像大小为 4 pixel×4 pixel，输出图像大小为 64 pixel×64 pixel，整个过程中使用单波段图像进行训练和测试，即每个波段都有自己的模型参数。

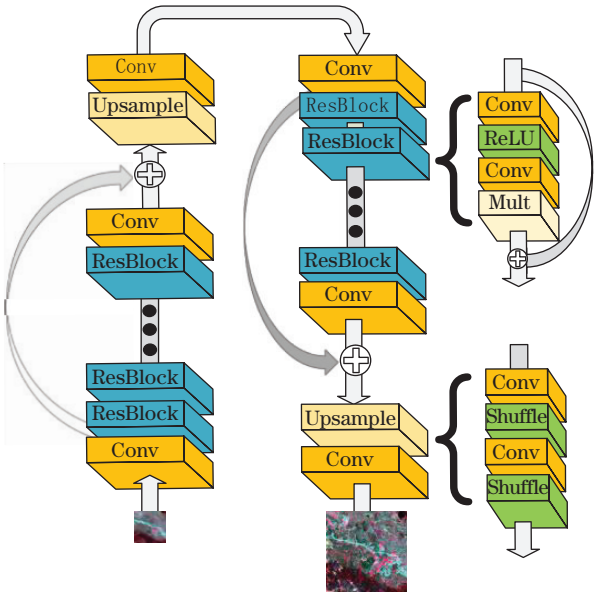

图 2　用于建模空间差异和传感器差异的级联 EDSR 网络

Fig. 2　Cascaded EDSR network for modeling spatial differences and sensor differences

2.2　残差修正

在不考虑时相变化的理论情况下，重建的真实 Landsat 图像可以由 MODIS 图像、空间差异和传感器差异组成，具体表达式为

$$L_k = M_k + \Delta S + \Delta B, k \in \mathbf{N}^+, \quad (1)$$

式中：M_k 表示任意时刻的 MODIS 图像；ΔS 表示空间差异；ΔB 表示传感器差异；L_k 为需要重建的 k 时刻的 Landsat 图像。

然而，在使用 EDSR 网络对 MODIS 数据进行 16 倍上采样以后，生成的结果 L_1' 和 L_2' 中每个像素可能与真实的 Landsat 图像中的几种地物相对应，故这并不是一个非常准确的预测。因此，在生成结果与其对应时刻的真实 Landsat 图像之间引入一个残差项 R，进而(1)式可以改写为

$$L_k = M_k + \Delta S + \Delta B + R, k \in \mathbf{N}^+。 \quad (2)$$

从而，进一步提出了一种基于 CNN 的方法来有效地学习 L_1' 和 L_2' 与真实 Landsat 图像之间的残差，对 L_1' 和 L_2' 进行进一步修正，使得生成结果与真实的 Landsat 图像更接近，结构如图 3 所示。该网络包含三个 Conv 和一个 ResBlock，其中所有的卷积核大小设置为 3×3，每个卷积层包含 128 个卷积核，在残差单元中将 ReLU 作为激活函数，网络的输入和输出均为 64 pixel×64 pixel 大小的图像块。

所用训练集来自 EDSR 网络的输出结果和其对应的真实时刻的 Landsat 图像，网络采用 Adam 优化器进行优化，初始学习率为 0.0001，训练周期为 500 个。从单对图像中提取出 1296 个图像块用于训练，整个过程中使用单波段数据进行训练和验证，并且在拟合实验数据过程中使用早停策略。

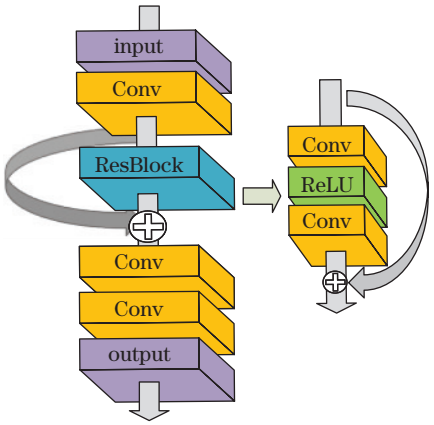

图 3　用于残差修正的 CNN 网络

Fig. 3　CNN network for residual correction

2.3　时间预测

尽管通过残差修正的结果 T_1^L 和 T_2^L 在空间分辨率上与真实 Landsat 相同，但是仅使用 t_1 时刻的图像对训练会导致 T_2^L 中不具备 t_1 时刻到 t_2 时刻的时间变化信息。因此，借鉴文献[3]中的做法，利用高通调制方法将 t_1 时刻到 t_2 时刻的时间变化信息传递给 T_2^L，具体表达式为

$$L_2 = [T_2^L / T_1^L] \times L_1, \quad (3)$$

式中：T_2^L / T_1^L 为高通调制比例系数。计算过程是逐像素进行的，最后可得到预测图像 L_2。

3 实　　验

3.1 实验方案

为了验证所提方法的有效性，在两个数据集上将所提方法与目前已有的时空融合方法进行了比较。选择的比对算法包括 STARFM[4]、SPSTFM[11]、FISTA[13]、EBSCDL[12]、FSDAF[9] 和 STFDCNN[3]。在比较过程中使用的所有比对算法均保留其原始参数设置。对于 STARFM 预测结果中出现的−9999 数值，统一修正为 0，以改善该算法性能。单对图像训练所能提供的数据量有限，因此 STFDCNN 算法的性能受到了限制。

实验使用的第一组数据为文献[4]中的实验数据，该图像大小为 1200 pixel×1200 pixel，且仅有绿色、红色和近红外三个波段。两对 Landsat-7 和 MODIS 数据分别于 2001 年 5 月 24 日和 8 月 12 日获取。将 8 月 12 日的 Landsat-7 图像设置为未知的待重建图像，利用 5 月 24 日的一对图像和 8 月 24 日的 MODIS 图像来合成，并将合成结果和真实数据比对以确定算法的性能。从 Landsat-7 ETM+ 传感器和 MODIS 多光谱(MS)传感器中选择地表反射率产品中的绿色、红色和近红外波段的部分数据构成融合的源图像，并生成了大小为 1200 pixel×1200 pixel 的预测结果来进行详细比较。为了适应单对图像的情况，比对算法 STARFM、SPSTFM、FISTA、EBSCDL、FSDAF 和 STFDCNN 在训练时只使用了已知时刻的一对图像而没有使用差值图像对。

实验使用的另外一组数据来自 Landsat-5 和 MODIS 匹配的 LGC 数据集[18]。选取 2004 年 4 月 16 日和 2004 年 5 月 2 日获取的图像对进行实验，将 2004 年 5 月 2 日的 Landsat 图像设为目标图像，利用 4 月 16 日的一对图像和 5 月 2 日的 MODIS 图像来合成，并将合成结果和真实数据比对以确定算法的性能。从 Landsat-5 TM 传感器和 MODIS 多光谱(MS)传感器中选择地表反射率产品中的蓝色、绿色、红色、近红外、中红外和热红外波段的部分数据构成融合的源图像，并将图像裁剪成和 Landsat-7 数据同样的 1200 pixel×1200 pixel 大小来进行实验，以便于验证实验结果。

3.2 客观指标评价

对预测图像进行辐射误差、结构相似性(SSIM)和光谱损失评估。为了客观地比较算法差异，使用均方根误差(RMSE)来评估重建结果的辐射误差，使用 SSIM 来评估结构间的相似性，还采用了一些光谱指标来测量光谱损失，包括光谱角(SAM)、相对平均光谱误差(RASE)、相对全局合成误差(ERGAS)和基于四元数理论的质量指数(Q4)。RMSE、SAM、RASE 和 ERGAS 的值越小越好，而 SSIM 和 Q4 的值越大越好。

为了能够进行细节比较，将最终 1200 pixel×1200 pixel 大小的预测结果裁剪为 4 个 600 pixel×600 pixel 大小的块进行比较。表 1～4 为 Landsat-7 和 MODIS 重建图像块的客观评价结果比较，其中最好的评价结果均已被加粗显示，band1、band2、band3 分别表示绿色、红色和近红外波段。可以发现，这些方法均能有效地重建未知时刻的高分辨率图像。SPSTFM 和 FISTA 产生了非常相似的结果，而基于字典学习的方法 EBSCDL 的各项指标均优于 SPSTFM 和 FISTA。STFDCNN 由于受到单对图像训练的限制，故表现较差。虽然对 STARFM 的结果进行了修正，但是 FSDAF 和所提方法仍然具有较大的优势。所提方法在不同的评价指标上均

表 1　Landsat-7 重建结果中第一个图像块的指标评价

Table 1　Index evaluation of the first image block in Landsat-7 reconstruction result

Index	Band	STARFM	SPSTFM	FISTA	EBSCDL	FSDAF	STFDCNN	Proposed
RMSE	band1	0.0067	0.0068	0.0068	0.0069	0.0064	0.0099	**0.0062**
	band2	0.0098	0.0112	0.0112	0.0104	0.0092	0.0118	**0.0082**
	band3	0.0242	0.0270	0.0270	0.0231	0.0232	0.0304	**0.0222**
SSIM	band1	0.8669	0.7877	0.8653	0.8622	**0.8811**	0.8031	0.8028
	band2	0.8090	0.7861	0.7864	0.7857	0.8277	0.7675	**0.8511**
	band3	0.7488	0.7877	0.7880	0.7866	0.7718	0.7499	**0.8028**
SAM		0.0477	0.0567	0.0567	0.0510	0.0465	0.0574	**0.0418**
RASE		0.1670	0.1878	0.1877	0.1622	0.1593	0.2131	**0.1509**
ERGAS		0.1990	0.2182	0.2181	0.2076	0.1875	0.2523	**0.1682**
Q4		0.7949	0.7661	0.7663	0.8128	0.8068	0.7761	**0.8490**

表 2 Landsat-7 重建结果中第二个图像块的指标评价

Table 2 Index evaluation of the second image block in Landsat-7 reconstruction result

Index	Band	STARFM	SPSTFM	FISTA	EBSCDL	FSDAF	STFDCNN	Proposed
RMSE	band1	0.0063	0.0065	0.0065	0.0062	0.0059	0.0096	**0.0056**
	band2	0.0092	0.0102	0.0102	0.0098	0.0086	0.0123	**0.0080**
	band3	0.0304	0.0343	0.0343	0.0293	0.0289	0.0324	**0.0275**
SSIM	band1	0.7118	0.9123	0.9125	0.9098	0.9158	0.8586	**0.9194**
	band2	0.8733	0.8558	0.8561	0.8558	0.8816	0.8172	**0.8968**
	band3	0.7118	0.7482	0.7484	0.7458	0.7529	0.7394	**0.7627**
SAM		0.0492	0.0569	0.0569	0.0521	0.0485	0.0606	**0.0430**
RASE		0.2057	0.2309	0.1877	0.1991	0.1938	0.2293	**0.1853**
ERGAS		0.2115	0.2392	0.2390	0.2167	0.1974	0.2708	**0.1930**
Q4		0.7659	0.6783	0.6785	0.7821	0.7842	0.7923	**0.8247**

表 3 Landsat-7 重建结果中第三个图像块的指标评价

Table 3 Index evaluation of the third image block in Landsat-7 reconstruction result

Index	Band	STARFM	SPSTFM	FISTA	EBSCDL	FSDAF	STFDCNN	Proposed
RMSE	band1	0.0059	0.0066	0.0066	0.0063	0.0056	0.0100	**0.0053**
	band2	0.0097	0.0115	0.0114	0.0105	0.0087	0.0125	**0.0072**
	band3	0.0270	0.0301	0.0301	0.0260	0.0262	0.0421	**0.0252**
SSIM	band1	0.8636	0.8652	0.8655	0.8635	0.8835	0.7961	**0.8883**
	band2	0.7918	0.7229	0.7231	0.7625	0.8153	0.7300	**0.8501**
	band3	0.7736	0.7658	0.7658	0.7985	0.7854	0.7843	**0.8108**
SAM		0.0676	0.1204	0.1203	0.0754	0.0644	0.0842	**0.0616**
RASE		0.2133	0.2313	0.2312	0.2077	0.2034	0.3233	**0.1934**
ERGAS		0.2390	0.2819	0.2817	0.2529	0.2171	0.3220	**0.1875**
Q4		0.9066	0.8738	0.8739	0.9106	0.9111	0.8162	**0.9235**

表 4 Landsat-7 重建结果中第四个图像块的指标评价

Table 4 Index evaluation of the fourth image block in Landsat-7 reconstruction result

Index	Band	STARFM	SPSTFM	FISTA	EBSCDL	FSDAF	STFDCNN	Proposed
RMSE	band1	0.0058	0.0060	0.0060	0.0057	0.0053	0.0088	**0.0049**
	band2	0.0088	0.0100	0.0100	0.0092	0.0081	0.0104	**0.0064**
	band3	0.0287	0.0311	0.0311	0.0279	0.0281	0.0410	**0.0270**
SSIM	band1	0.8626	0.8849	0.8852	0.8827	0.8918	0.8315	**0.8968**
	band2	0.8158	0.8004	0.8007	0.7988	0.8287	0.7953	**0.8723**
	band3	0.7244	0.7656	0.7656	0.7605	0.7452	0.7187	**0.7772**
SAM		0.0498	0.0572	0.0572	0.0537	0.0499	0.0604	**0.0395**
RASE		0.1871	0.2028	0.2028	0.1821	0.1803	0.2631	**0.1718**
ERGAS		0.2020	0.2130	0.2129	0.2044	0.1847	0.2496	**0.1547**
Q4		0.8007	0.7746	0.7747	0.8127	0.8049	0.7315	**0.8476**

表现出较好的性能，RMSE 和 SSIM 的结果表明所提方法具有较低的辐射误差和较高的 SSIM，SAM 等光谱指标表明利用所提方法得到的重建图像结果产生的光谱损失最少。相对于 STARFM、SPSTFM、FISTA、EBSCDL 和 STFDCNN，所提方法有明显的优势，且各项指标稳定优于 FSDAF，重建结果较好。

表 5～8 为 Landsat-5 和 MODIS 重建图像块的客观评价结果比较，其中最好的评价结果均已被加粗显示，band1、band2、band3、band4、band5、band6 分别表示绿色、红色、近红外、蓝色、中红外和热红外波段。可以发现，EBSCDL 的评价结果仍然稳定优于 SPSTFM 和 FISTA，且相比于 STFDCNN 也有明显的优势。表 7 的评价结果表明所提方法在第三

个图像块上取得较好的性能，具有较小的辐射误差和光谱角，同时具有较高的结构相似性。表 5、表 6 和表 8 的结果都表明所提方法在 RMSE 评价和部分光谱损失评价都弱于 FSDAF 和 STARFM。但是，所提方法在所有图像的各个波段上的 SSIM 评价均较好，表明所提方法的重建结果保持了较好的结构重建信息，与真实图像具有较小的误差和较大的相关性。

表 5 Landsat-5 重建结果中第一个图像块的指标评价

Table 5 Index evaluation of the first image block in Landsat-5 reconstruction result

Index	Band	STARFM	SPSTFM	FISTA	EBSCDL	FSDAF	STFDCNN	Proposed
RMSE	band1	0.0144	0.0203	0.0202	0.0160	**0.0140**	0.0238	0.0148
	band2	0.0176	0.0243	0.0242	0.0196	**0.0170**	0.0321	0.0181
	band3	0.0258	0.0348	0.0347	0.0288	**0.0245**	0.0402	0.0253
	band4	0.0129	0.0156	0.0156	0.0140	0.0125	0.0236	**0.0116**
	band5	0.0300	0.0368	0.0366	0.0289	**0.0265**	0.0460	0.0279
	band6	0.0298	0.0389	0.0388	0.0309	**0.0281**	0.0499	0.0296
SSIM	band1	0.8636	0.8456	0.8463	0.8526	0.8634	0.8132	**0.8726**
	band2	0.8566	0.8459	0.8488	0.8486	0.8546	0.7962	**0.8696**
	band3	0.8131	0.8139	0.8179	0.8034	0.8177	0.7582	**0.8317**
	band4	0.8666	0.8573	0.8593	0.8490	0.8632	0.8082	**0.8758**
	band5	0.8006	0.8842	0.8866	0.8788	0.8648	0.8254	**0.8863**
	band6	0.7901	0.8617	0.8640	0.8535	0.8375	0.7972	**0.8651**
SAM		0.0516	0.0516	0.0610	0.0613	**0.0505**	0.0862	0.0518
RASE		0.1386	0.1386	0.1975	0.1553	**0.1337**	0.2322	0.1410
ERGAS		0.1403	0.1403	0.2035	0.1569	**0.1362**	0.2423	0.1449
Q4		0.8811	0.8811	0.8062	0.8639	**0.8869**	0.7761	0.8826

表 6 Landsat-5 重建结果中第二个图像块的指标评价

Table 6 Index evaluation of the second image block of Landsat-5 reconstruction result

Index	Band	STARFM	SPSTFM	FISTA	EBSCDL	FSDAF	STFDCNN	Proposed
RMSE	band1	**0.0146**	0.0218	0.0217	0.0167	**0.0146**	0.0219	0.0150
	band2	0.0181	0.0256	0.0255	0.0203	**0.0180**	0.0295	0.0187
	band3	0.0254	0.0338	0.0336	0.0283	**0.0252**	0.0389	0.0258
	band4	0.0130	0.0177	0.0176	0.0143	0.0128	0.0215	**0.0116**
	band5	0.0296	0.0350	0.0349	0.0298	**0.0284**	0.0464	0.0294
	band6	**0.0353**	0.0372	0.0370	0.0367	0.0360	0.0516	0.0364
SSIM	band1	0.8559	0.8231	0.8247	0.8330	0.8325	0.7909	**0.8635**
	band2	0.8526	0.8336	0.8367	0.8375	0.8281	0.7836	**0.8656**
	band3	0.8176	0.8064	0.8096	0.8031	0.8077	0.7633	**0.8287**
	band4	0.8618	0.8506	0.8528	0.8430	0.8462	0.7996	**0.8793**
	band5	0.7839	0.8626	0.8658	0.8533	0.8200	0.7884	**0.8589**
	band6	0.7747	0.8434	**0.8461**	0.8239	0.7813	0.7635	0.8335
SAM		0.05380	0.0538	0.0619	0.0615	**0.0519**	0.0851	0.0526
RASE		**0.1452**	0.4152	0.2026	0.1642	0.1454	0.2307	0.1517
ERGAS		**0.1490**	**0.1490**	0.2145	0.1697	0.1497	0.2383	0.1573
Q4		0.8819	0.8819	0.8156	0.8652	0.8805	0.7919	**0.8828**

表 7 Landsat-5 重建结果中第三个图像块的指标评价

Table 7 Index evaluation of the third image block in Landsat-5 reconstruction result

Index	Band	STARFM	SPSTFM	FISTA	EBSCDL	FSDAF	STFDCNN	Proposed
RMSE	band1	0.0141	0.0234	0.0234	0.0154	0.0142	0.0186	**0.0136**
	band2	0.0170	0.0285	0.0285	0.0189	0.0171	0.0271	**0.0168**
	band3	0.0261	0.0355	0.0354	0.0281	0.0255	0.0387	**0.0254**
	band4	0.0141	0.0206	0.0206	0.0147	0.0135	0.0207	**0.0106**
	band5	0.0278	0.0385	0.0384	0.0276	0.0260	0.0396	**0.0259**
	band6	0.0295	0.0377	0.0376	0.0302	**0.0287**	0.0458	0.0288
SSIM	band1	0.7282	0.7201	0.7228	0.7217	0.6934	0.7021	**0.7559**
	band2	0.7549	0.7292	0.7325	0.7429	0.7150	0.7024	**0.7777**
	band3	0.7960	0.7961	0.7998	0.7935	0.7817	0.7655	**0.8152**
	band4	0.6960	0.6986	0.7020	0.6886	0.6920	0.6549	**0.7363**
	band5	0.7200	0.8163	0.8201	0.8117	0.7764	0.7653	0.8223
	band6	0.7219	0.8062	0.8089	0.7970	0.7600	0.7574	**0.8121**
SAM		0.0475	0.0475	0.0616	0.0542	0.0477	0.0735	**0.0467**
RASE		0.1374	0.1374	0.1992	0.1503	**0.1369**	0.2098	0.1374
ERGAS		**0.1351**	**0.1351**	0.2063	0.1488	0.1359	0.2040	0.1357
Q4		0.8993	0.8993	0.8081	0.8890	0.8987	0.8250	**0.9071**

表 8 Landsat-5 重建结果中第四个图像块的指标评价

Table 8 Index evaluation of the fourth image block in Landsat-5 reconstruction result

Index	Band	STARFM	SPSTFM	FISTA	EBSCDL	FSDAF	STFDCNN	Proposed
RMSE	band1	0.0129	0.0170	0.0169	0.0146	**0.0127**	0.0218	0.0135
	band2	**0.0160**	0.0207	0.0206	0.0184	**0.0160**	0.0299	0.0172
	band3	0.0232	0.0302	0.0301	0.0267	**0.0224**	0.0435	0.0248
	band4	0.0104	0.0131	0.0130	0.0114	0.0100	0.0164	**0.0098**
	band5	0.0288	0.0316	0.0315	0.0294	**0.0276**	0.0469	0.0294
	band6	**0.0361**	0.0372	0.0371	0.0378	0.0375	0.0527	0.0373
SSIM	band1	0.7879	0.7807	0.7829	0.7745	0.7741	0.7370	**0.8060**
	band2	0.8024	0.7879	0.7916	0.7844	0.7809	0.7288	**0.8139**
	band3	0.8228	0.8127	0.8162	0.8015	**0.8264**	0.7553	0.8218
	band4	0.7626	0.7597	0.7626	0.7455	0.7475	0.7185	**0.7878**
	band5	0.7322	0.8197	**0.8226**	0.8073	0.7747	0.7444	**0.8118**
	band6	0.7229	0.7878	**0.7904**	0.7697	0.6992	0.7187	0.7799
SAM		**0.0461**	**0.0461**	0.0626	0.0535	0.0464	0.0755	0.0472
RASE		0.1258	0.1258	0.1637	0.1451	**0.1236**	0.2383	0.1364
ERGAS		0.1301	0.1301	0.1700	0.1497	**0.1291**	0.2431	0.1413
Q4		0.8943	0.8943	0.8561	0.8731	**0.8976**	0.7506	0.8839

3.3 视觉比较

从视觉上对实验结果进行了分析，图 4 中展示了 Landsat-7 和 MODIS 的第三个图像块中的重建结果。图 5 展示了 Landsat-5 和 MODIS 的第三个图像块中的重建结果。所有图像的原始波段顺序是近红外、红色、绿色，被显示的红色、绿色、蓝色波段是从重建的近红外、红色、绿色波段映射而来。由于这些图像的编码方式为 16 位编码方式，为了能够显示，将真实的 Landsat 图像作为参考图像，并对其进行 2%（去掉的像素值最大和最小的像素数目占总像素数目的比例）拉伸。然后，并将拉伸所使用的阈值应用到其他图像进行非线性拉伸，即所有图像都使用相同的阈值拉伸至 0～255 之间。经过上述处理后，在将重建图像与真实图像进行比较时，可以直接从图像中的色彩偏差中读取光谱损失。

图 4 为 Landsat-7 第三个图像块中近红外、红色、绿色波段的合成结果。对图 4 逐个色彩进行比较能够定性判断出每个波段的重建质量。在图 4 中，SPSTFM 和 FISTA 的整体呈现为红色，和原图的红色、绿色相间有明显差异，这说明两种方法对近

红外波段的重建误差较大。STFDCNN 仅使用一对图像训练，重建结果中出现大量黑蓝色部分，与 Landsat-7 原图的绿色区域不一致，说明其对红色波段的重建误差较大。其他图像中的绿色区域与 Landsat-7 原图中绿色区域较一致，即其他方法对红色波段的重建误差较小。图像中没有单纯的蓝色，难以直接评价出各方法对于绿波段的重建质量的差异，但可以发现 EBSCDL 的绿色中融入了较多的蓝色，说明其对绿色波段的重建不如所提算法。虽然 STARFM 和 FSDAF 颜色整体一致，但是所提方法的重建结果更接近真实图像，说明所提方法与真实图像结构信息上更加接近，且具有较小的误差。

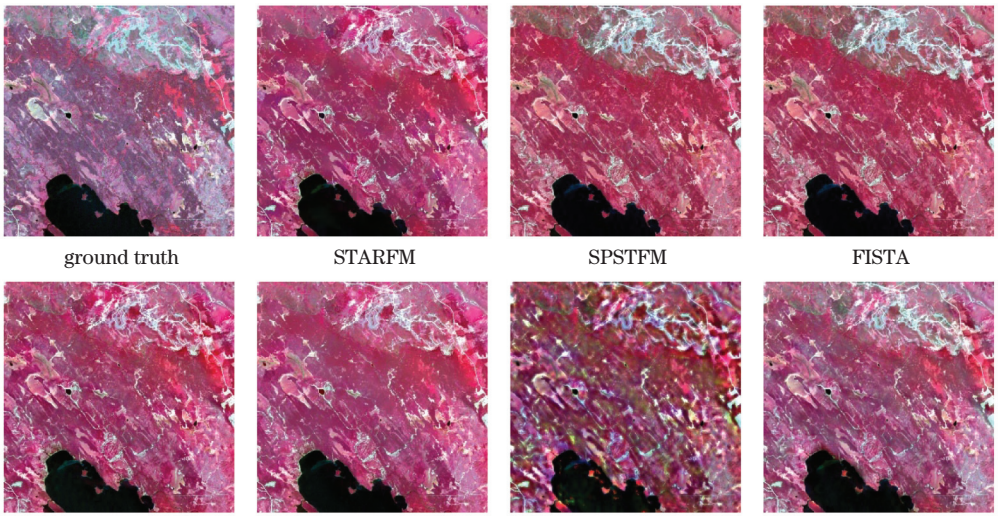

图 4　Landsat-7 第三个图像块中近红外、红色、绿色波段的合成结果

Fig. 4　Synthesis result of near infrared, red and green bands in the third image block of Landsat-7

图 5 为 Landsat-5 第三个图像块中近红外、红色、绿色波段的合成结果。对图 5 进行目视分析可以发现，所有重建结果的整体颜色均相近，说明各方法重建出的光谱差异都比较低。SPSTFM 和 FISTA 的重建结果仍然保持一致，且重建结果中缺失了 Landsat-5 原始图像中很多绿色部分。EBSCDL 的重建结果略优于 SPSTFM 和 FISTA。观察局部放大的区域可以发现，所提算法与 STARFM、EBSCDL、FSDAF 和 STFDCNN 相比，可以重建出更多的细节部分，与真实图像更加接近。仔细观察 STARFM 重建结果的左上角部分，可以观察到重建结果中产生了大量蓝色部分，这不利于后续的地物分类应用。

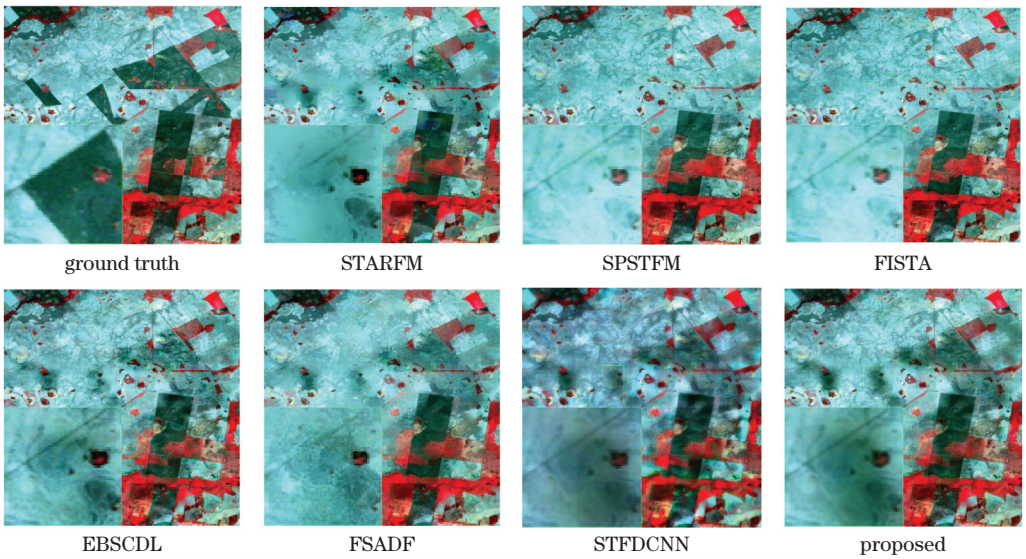

图 5　Landsat-5 第三个图像块中近红外、红色、绿色波段的合成结果

Fig. 5　Synthesis result of near infrared, red and green bands in the third image block of Landsat-5

4 讨 论

实验的数字结果和视觉结果均表明，基于所提方法得到的结果图像具有较高的质量，相比于基于字典学习的融合算法具有明显提升。在提供同样的训练集和验证集的情况下，相比于基于深度学习的融合算法 STFDCNN，所提方法优势更加明显。

如果实验中能够积累更多的训练数据，那么所提方法就有可能获得更高的性能。当训练数据集变丰富时，基于深度学习的方法的重建结果会变优，而基于字典学习的方法和基于解混的方法则无法有效学习。

此外，所提方法在使用上有较为明显的时间优势。在训练的过程中，通过设置早停策略以便在最短的时间内获得最优的重建效果。在预测一个近红外、红色、绿色的三波段图像时，所提方法需要使用 2080Ti 型号的图形处理器(GPU)计算 2～3 h 来生成 1200 pixel × 1200 pixel 大小的图像。使用 MATLAB 进行编程但核心的字典学习和优化是使用 C 代码实现的 EBSCDL 算法大约需要 19 h 才能获得同样尺寸的图像，且这一速度无法提升，因为每一次的耦合字典学习都依赖于已知时刻的图像对，即和图像内容密切相关。由于字典学习对数据规模的限制，它无法重复利用已有知识，进而无法离线进行。基于深度学习提出的 STFDCNN 算法，由于采用的是类似于 SRCNN 的浅层网络结构，故其学习能力与泛化能力较差。因此，在提供相同数据集的情况下，与所提方法相比，STFDCNN 需要迭代更多次数来训练网络。在实际操作中，STFDCNN 算法在预测一个三波段图像时，需要在 2080Ti 型号的 GPU 计算大约 12 h 才能获得同样尺寸的图像。

5 结 论

针对目前使用卷积神经网络进行时空融合时存在的网络浅且融合性能有限的问题，提出了基于增强超分辨网络的时空融合方法。首先，使用深度 CNN 融合不同时空分辨率的遥感图像，通过引入 EDSR 网络来生成目标图像，解决了高分辨率和低分辨率图像之间的空间差异和传感器差异。然后，在 EDSR 网络之后补充了一个小型神经网络来进行残差修正，以缩小重建图像与真实 Landsat 图像之间的差距。接着，采用高通调制的方法来进行不同时间上的预测，并得到最终融合结果。最后，将所提方法与一些基于权重的时空融合方法、基于解混的方法、基于字典学习的融合方法和基于深度学习的方法进行了比较，并测试了 Landsat-7 和 MODIS 的近红外、红色和绿色波段以及 LandSat-5 和 MODIS 的蓝色、绿色、红色、近红外、中红外和热红外波段。实验结果中的客观评价结果表明，所提方法在辐射、结构和光谱保真度方面优于比对算法。视觉比较结果显示，与比对算法相比，所提方法可以重建出更多的细节，展示出了该方法的有效性。

参 考 文 献

[1] Jiang Z T, He Y T. Infrared and visible image fusion method based on convolutional auto-encoder and residual block[J]. Acta Optica Sinica, 2019, 39(10): 1015001.
江泽涛, 何玉婷. 基于卷积自编码器和残差块的红外与可见光图像融合方法[J]. 光学学报, 2019, 39(10): 1015001.

[2] Chen Y, Cao R Y, Chen J, et al. A new cross-fusion method to automatically determine the optimal input image pairs for NDVI spatiotemporal data fusion[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(7): 5179-5194.

[3] Song H H, Liu Q S, Wang G J, et al. Spatiotemporal satellite image fusion using deep convolutional neural networks[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(3): 821-829.

[4] Gao F, Masek J, Schwaller M, et al. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance[J]. IEEE Transactions on Geoscience and Remote Sensing, 2006, 44(8): 2207-2218.

[5] Zhu X L, Chen J, Gao F, et al. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions[J]. Remote Sensing of Environment, 2010, 114(11): 2610-2623.

[6] Hilker T, Wulder M A, Coops N C, et al. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS[J]. Remote Sensing of Environment, 2009, 113(8): 1613-1627.

[7] Wu M Q, Niu Z, Wang C Y, et al. Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model[J]. Journal of Applied Remote Sensing, 2012, 6(1): 063507.

[8] Xu Y, Huang B, Xu Y Y, et al. Spatial and temporal

image fusion via regularized spatial unmixing[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(6): 1362-1366.

[9] Zhu X L, Helmer E H, Gao F, et al. A flexible spatiotemporal method for fusing satellite images with different resolutions[J]. Remote Sensing of Environment, 2016, 172: 165-177.

[10] He C Q, Li Q Z, Liu H L, et al. Remote sensing images mosaicking method based on spatiotemporal fusion[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1415002.
何超琦, 李奇泽, 刘华霖, 等. 基于时空融合的遥感图像镶嵌方法[J]. 激光与光电子学进展, 2021, 58(14): 1415002.

[11] Huang B, Song H H. Spatiotemporal reflectance fusion via sparse representation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2012, 50(10): 3707-3716.

[12] Wu B, Huang B, Zhang L P. An error-bound-regularized sparse coding for spatiotemporal reflectance fusion[J]. IEEE Transactions on Geoscience and Remote Sensing, 2015, 53(12): 6791-6803.

[13] Liu X, Deng C W, Zhao B J. Spatiotemporal reflectance fusion based on location regularized sparse representation[C]//2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), July 10-15, 2016, Beijing, China. New York: IEEE Press, 2016: 2562-2565.

[14] Wei J B, Wang L Z, Liu P, et al. Spatiotemporal fusion of remote sensing images with structural sparsity and semi-coupled dictionary learning[J]. Remote Sensing, 2016, 9(1): 21.

[15] Wei J B, Wang L Z, Liu P, et al. Spatiotemporal fusion of MODIS and Landsat-7 reflectance images via compressed sensing[J]. IEEE Transactions on Geoscience and Remote Sensing, 2017, 55(12): 7126-7139.

[16] Dai P Y, Zhang H Y, Zhang L P, et al. A remote sensing spatiotemporal fusion model of Landsat and MODIS data via deep learning[C]//IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, July 22-27, 2018, Valencia, Spain. New York: IEEE Press, 2018: 7030-7033.

[17] 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), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1132-1140.

[18] Emelyanova I V, McVicar T R, van Niel T G, et al. Assessing the accuracy of blending Landsat-MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: a framework for algorithm selection[J]. Remote Sensing of Environment, 2013, 133: 193-209.