

# 基于深度学习与超分辨率重建的遥感高时空融合方法<sup>\*</sup>

张永梅<sup>1</sup>, 滑瑞敏<sup>1</sup>, 马健喆<sup>2</sup>, 胡 蕾<sup>3</sup>

(1. 北方工业大学信息学院, 北京 100144; 2. 香港理工大学电子与信息工程系, 香港 00852;

3. 江西师范大学计算机信息工程学院, 江西 南昌 330022)

**摘 要:**针对遥感影像的“时空矛盾”,提出一种改进 STARFM 的遥感高时空融合方法。利用 SRCNN 对低分辨率影像进行超分辨率重建,由于所融合的 2 组影像分辨率差距过大,网络训练困难,先将 2 组影像均采样至某一中间分辨率,使用高分辨率影像作为低分辨率影像的先验知识进行 SRCNN 重建,再将得到的中间分辨率影像重采样后以原始高分辨率影像作为先验知识进行第 2 次 SRCNN 重建,得到的最终重建图像相比原先使用插值法重采样所得图像,在 PSNR 和 SSIM 上均有提升,缓解了传感器差异所造成的系统误差。STARFM 融合方法在筛选相似像元与计算权重时均使用专家知识提取人工特征,基于 STARFM 时空融合的基本思想,以 SRCNN 作为基本框架自动提取特征,实验结果表明,其 MSE 值相比原方法更低,进一步提高了遥感时空融合的质量,有利于充分利用遥感影像。

**关键词:**时空融合;改进 STARFM;SRCNN;自动特征提取

**中图分类号:**TP751

**文献标志码:**A

**doi:**10.3969/j.issn.1007-130X.2020.09.008

## A high spatial temporal fusion method based on deep learning and super resolution reconstruction

ZHANG Yong-mei<sup>1</sup>, HUA Rui-min<sup>1</sup>, MA Jian-zhe<sup>2</sup>, HU Lei<sup>3</sup>

(1. School of Information Science and Technology, North China University of Technology, Beijing 100144;

2. Department of Electronic & Information Engineering, The Hong Kong Polytechnic University, Hong Kong 00852;

3. School of Computer Information Engineering, Jiangxi Normal University, Nanchang 330022, China)

**Abstract:** Aiming at the "space-time conflict" of remote sensing images, a high spatial-temporal fusion algorithm based on improved STARFM is proposed. SRCNN is used for the super-resolution reconstruction of low-resolution images. Due to the large difference in resolution between the two groups of fusion images, the network training is difficult. Firstly, both of the two groups are sampled to an intermediate resolution, and low-resolution images are reconstructed by SRCNN with high-resolution images as their prior knowledge. Secondly, the obtained intermediate resolution images are resampled, and then they are reconstructed by SRCNN with original high-resolution images as their prior knowledge. The resulting reconstructed images have higher PSNR and SSIM than the images resampled by interpolation, alleviating the systematic error caused by the sensor difference. The STARFM fusion method uses expert knowledge to extract artificial features in selecting "Spectrally Similar Neighbor Pixels" and computer their weights. Based on the basic concept of STARFM, an automatic feature extraction method using SRCNN as the basic framework is realized. The experimental results show that this method has lower MSE value than the original STARFM, which further improves the quality of spatial-temporal fu-

<sup>\*</sup> 收稿日期:2019-11-07;修回日期:2020-02-22

基金项目:国家自然科学基金(61371143,61662033);教育部高等教育司产学研合作协同育人项目(201801121002);全国高等学校计算机教育研究会 2019 年度课题(CERACU2019R05);教育部科技发展中心“天诚汇智”创新促教基金(2018A03029);2019 年度北京市教委基本科研业务费(110052971921/002)

通信地址:100144 北京市北方工业大学信息学院

Address: School of Information Science and Technology, North China University of Technology, Beijing 100144, P. R. China

sion.

**Key words:** spatial-temporal fusion; improved STARFM; SRCNN; automatic feature extraction

## 1 引言

受限于卫星传感器的硬件条件与发射成本,在可获取的单一遥感数据中,其空间分辨率和时间分辨率成为一对矛盾体<sup>[1-3]</sup>,若想要尽可能地提高遥感影像的空间分辨率,结果则可能引发“时间数据缺失”<sup>[4]</sup>,反之难免又会牺牲影像的空间分辨率,制约了遥感影像的应用。在早期发射升空的卫星传感器所得到的数据中这种表现尤为明显<sup>[5]</sup>,造成了一直以来积累下来的庞大数据难以直接利用。

多源遥感图像融合是指将不同传感器获得的同一场景的图像或者同一传感器在不同时刻获得的同一场景图像,经过相应处理后,再运用某种融合技术得到一幅合成图像的过程<sup>[6]</sup>,其为解决上述遥感影像时空矛盾问题提供了一条出路,即遥感影像的时空融合方法。

传统的时空融合模型包括基于变换的模型和基于像元重构的模型<sup>[4]</sup>。基于变换的模型是指对遥感影像进行数据变换(如小波分解等)后,对变换所得数据进行融合处理,最后再进行反变换从而得到想要的未知时刻高分辨率影像。如 Malenovsky 等<sup>[7]</sup>利用小波变换对 MODIS (MODerate-resolution Imaging Spectroradiometer) 影像的低频信息和 Landsat 影像的高频信息进行融合,验证了 MODIS 与 Landsat 影像融合的可能性,但其所得结果的分辨率只达到了 240 m,且在有变化发生的区域效果较差;基于像元重构的模型主要是将低空间分辨率图像进行像元分解,计算出不同时相上遥感影像的变化关系,同时还需求出不同分辨率影像间的变化关系,再通过插值得到未知时刻的高空间分辨率图像。如在时空融合领域影响最为深远的时空自适应反射率融合模型 STARFM (Spatial Temporal Adaptive Reflectance Fusion Model)<sup>[8]</sup>属于基于像元重构的模型,该模型在融合过程中不仅考虑了空间的差异性,而且还考虑了时间的差异性,是目前应用最广泛的时空融合模型之一,但在地物覆盖复杂、像元混合严重时往往精度较差。

随着机器学习的蓬勃发展,基于学习的遥感影像时空融合模型应运而生,如赵永光等<sup>[9]</sup>提出了一种利用稀疏表示对低空间分辨率图像进行超分辨率重建,再对数据进行高通滤波后融合的方法,取

得了良好的融合效果。但是,通常基于学习的模型使用浅层学习较多,同时方法复杂度较高,运行效率低,较难实现大区域的融合,且在空间分辨率上不宜相差过大,一般只有 4 倍左右<sup>[2]</sup>。

为了实现融合质量的进一步提升,本文选择将学习能力更强、特征提取更有力的深度学习方式引入遥感时空融合中。在现有基于学习的遥感影像时空融合的基础上,通过对 SRCNN (Super Resolution Convolutional Neural Network) 网络进行二次学习的方法来重建低分辨率图像。SRCNN 网络结构简单,可以相对缓解本文方法复杂度的提升,而二次学习的方式可以缓解融合时分辨率差距过大的影响。本文基于 STARFM 的基本思想,在融合过程中采用神经网络自动提取特征,学习映射关系,代替原本在滑动窗口内筛选相似像元与计算其权重的过程。实验表明,本文方法改善了遥感影像时空融合的质量。

## 2 基于改进 STARFM 的遥感高时空融合方法

### 2.1 基于深度学习的超分辨率重建方法

由于衍射现象的存在,通常成像系统所能达到的最高分辨率受限于成像光学器件本身,传统改善图像质量方法难以复原出系统在截止频率以外的信息<sup>[10]</sup>,而超分辨率重建 SR (Super resolution Restoration) 技术采用信号处理方法,能够重建出成像系统截止频率之外的信息,从而得到高于成像系统分辨率的图像<sup>[11]</sup>。

传统的超分辨率重建技术分为基于重建的方法和基于浅层学习的方法<sup>[12]</sup>。基于重建的方法将多幅低分辨率图像进行亚像素精度对齐来求出不同分辨率图像之间的运动偏移量,从而构建出空间运动参数,再通过各种先验约束条件与最优解来重建出高分辨率图像<sup>[13]</sup>。基于浅层学习的方法旨在通过学习获得高低分辨率图像之间的映射关系来指导图像的重建,通常分为特征提取、学习和重建 3 个阶段,且各阶段相互独立<sup>[14]</sup>,导致浅层学习特征提取与表达能力有限,有待进一步加强。

近年来,深度学习技术发展迅猛。在超分辨率重建问题中使用深度学习网络,在结构上仍需参考传统超分辨率重建方法的思想进行设计,使其作为

预测器能够输出较为准确的预测值<sup>[15]</sup>。

香港中文大学 Dong 等人<sup>[16]</sup>于 2016 年首次将卷积神经网络 CNN (Convolution Neural Network) 应用于超分辨率重建中, 提出了一种新的网络 SRCNN。首先从深度学习与传统稀疏编码 SC (Sparse Coding) 之间的关系入手, 将网络分为图像块提取、非线性映射和图像重建 3 个阶段, 使其分别对应于深度卷积神经网络框架中的 3 个卷积层, 统一于该神经网络之中, 从而实现了由低分辨率图像到高分辨率图像的超分辨率重建。该网络直接学习低分辨率图像和高分辨率图像之间的端到端映射, 除了优化之外, 几乎不需要预处理和后处理。

在 SRCNN 中使用的 3 个卷积层卷积核大小分别为  $9 \times 9$ ,  $1 \times 1$  和  $5 \times 5$ , 前 2 个输出的特征个数分别为 64 和 32。SRCNN 将稀疏编码的过程视为一种卷积运算, 网络设计简单, 并且重建效果相比作为浅层学习代表的 SCSR (Sparse Coding based Super Resolution)<sup>[17]</sup>方法也有较大改善, 是一种具有参考性的超分辨率重建方法。本文将通过二次学习方式, 将其引入遥感影像时空融合中, 从而提高低分辨率遥感图像重建效果。

本文采用 SRCNN 方法对低空间分辨率数据进行超分辨率重建, 以代替 STARFM 方法中低空间分辨率图像的重采样过程, 然而由于本文使用的 MOD09Q1 遥感影像只有 250 m 空间分辨率, 像元混合现象严重, 而 Landsat8 遥感影像具有 30 m 空间分辨率, 二者之间的分辨率相差 8~9 倍, 描述空间细节信息的能力差距较大; 并且由于得到遥感影像的不同传感器之间存在着各种差异(几何误差), 使得图像之间即使同一时刻同一位置的像素反射率也有较大差异, 故将 SRCNN 直接引入至遥感影像的时空融合比较困难。

基于学习模型的融合方法中影像之间的空间分辨率差距不宜过大, 一般在 4 倍左右, 差距过大容易造成深度学习神经网络的训练困难, 对低分辨率图像进行分辨率增强的效果欠佳, 像元混合严重的现象较难缓解。

为解决上述问题, 本文选择了一种进行二次学习的方法, 先将低空间分辨率遥感影像从 250 m 分辨率经过学习重建至 90 m 分辨率, 再重建至 30 m 分辨率, 从而保证了 2 次学习过程中分辨率差距均在 4 倍之内。超分辨率重建方法流程如图 1 所示。

首先将具有 30 m 空间分辨率的 Landsat8 遥

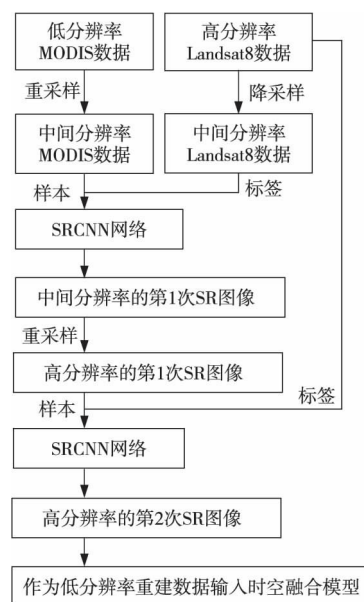


Figure 1 Flow chart of the super-resolution reconstruction method

图 1 本文超分辨率重建方法技术路线

感影像使用最近邻法降采样到 90 m 分辨率, 同时将具有 250 m 空间分辨率的 MOD09Q1 遥感影像使用双三次插值法重采样到 90 m 分辨率。如图 2 所示, 经过采样后的遥感影像虽然形式上具有了相同的分辨率, 但在空间细节上差距仍较大。本文使用 SRCNN 对采样后的 2 组数据进行第 1 次学习, 得到 90 m 分辨率的 MOD09Q1 重建图像。将得到的 90 m 分辨率重建图像重采样至 30 m 分辨率, 与具有 30 m 分辨率的 Landsat8 原始数据一起利用 SRCNN 进行第 2 次学习, 得到具有 30 m 分辨率的 MOD09Q1 重建图像, 本文将该重建图像作为时空融合模型中的低分辨率输入数据。

传统 STARFM 方法的输入只将 MODIS 数据通过插值法重采样到 Landsat8 数据的分辨率即可, 本文在使用 SRCNN 网络进行超分辨率重建时, 将 Landsat8 数据作为 MOD09Q1 数据的先验知识进行学习, 一方面超分辨率重建的精度明显优于插值法重采样结果, 提升了 STARFM 的输入质量; 另一方面, 在学习过程中, 能够缓解由不同传感器所造成的各种误差, 使重建得到的图像在风格与相似性上也与作为输入的高分辨率图像更加接近。故该方法能够尽可能还原出高低分辨率数据之间原本的差异信息, 使得融合结果更加准确。

本文进行二次学习的 2 个 SRCNN 网络具有相同的结构, 区别只在于输入和输出不同, 而且输入和输出的不同之处只体现在分辨率上, SRCNN 网络的训练过程通过图像切块进行, 故二次学习的

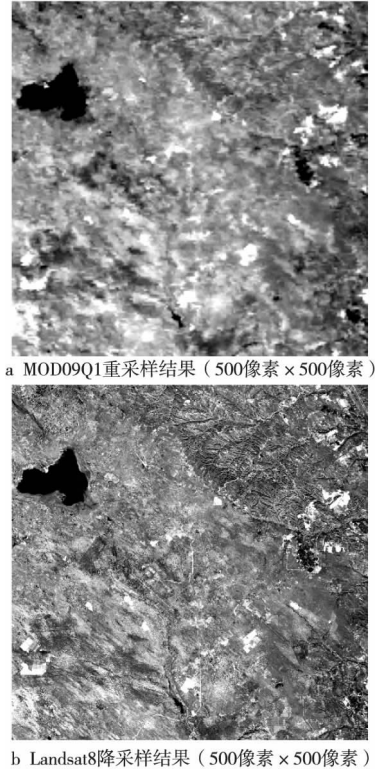


Figure 2 Comparison of MOD09Q1 and Landsat8 both converted to 90-meter resolution in the study area (2018. 10. 29)

图2 研究区域(2018. 10. 29)不同传感器采样至 90 m 分辨率的遥感影像对比图

2 个网络参数格式具有完全一致性,可以将训练好的用于第 1 次学习的网络参数迁移到用于第 2 次学习的网络中,提高运行效率,缓解深度学习所带来的时间复杂度的升高。

本文重建方法提高了高时间分辨率、低空间分辨率数据的空间分辨率,已得到了一定精度的高时空分辨率数据,再将其输入本文的改进 STARFM 方法中,相当于进行了双层时空融合,从而保证了较好的融合效果。

## 2.2 时空自适应反射率融合模型

Gao 等<sup>[8]</sup>提出的 STARFM 方法是遥感影像时空融合领域应用最广泛的一种方法。该方法在忽略几何误差和大气校正误差的情况下,假设低空间分辨率(如 MODIS)像元的反射率可以表示为高空间分辨率(如 Landsat)对应像元反射率的线性组合。此时如果将低空间分辨率的遥感影像重采样到分辨率相同的高分辨率遥感影像,则低空间分辨率像元与高空间分辨率对应位置像元存在如下关系:

$$M(x, y, t) = L(x, y, t) + \epsilon \quad (1)$$

其中,  $M(x, y, t)$ 、 $L(x, y, t)$  分别表示在  $t$  时刻

MODIS 影像和 Landsat 影像在坐标为  $(x, y)$  处像元的反射率,  $\epsilon$  表示不同分辨率图像由于传感器不同等所造成的系统误差。假设该  $\epsilon$  不随时间变化,若已有  $t_0$  和  $t_k$  时刻的 MODIS 数据,以及  $t_0$  时刻的 Landsat 数据,则未知的  $t_k$  时刻的 MODIS 数据可由式(2)求得:

$$L(x, y, t_k) = L(x, y, t_0) + M(x, y, t_k) - M(x, y, t_0) \quad (2)$$

STARFM 方法为避免假设所带来的像元混合现象,通过滑动窗口技术引入了邻近像元信息,即:

$$L(x_{w/2}, y_{w/2}, t_0) = \sum_{i=1}^w \sum_{j=1}^w \sum_{k=1}^n w_{i,j,k} \times (L(x_i, y_j, t_k) + M(x_i, y_j, t_0) - M(x_i, y_j, t_k)) \quad (3)$$

其中,  $w$  是滑动窗口大小,  $n$  表示不同时间点的遥感影像进行融合的数量,  $w_{i,j,k}$  为相应位置处像元对应的权值大小。在 STARFM 方法中,还需对上述滑动窗口中的相似像元进行搜索,并排除劣质像元,只有筛选得到的像元才配置非零权重,而在对筛选后相似像元配置权重时,需要综合考虑像元间的光谱距离  $S_{i,j,k}$ 、时间距离  $T_{i,j,k}$  和空间距离  $D_{i,j,k}$  等,其计算公式分别如下所示:

$$S_{i,j,k} = |L(x_i, y_j, t_k) - M(x_i, y_j, t_k)| \quad (4)$$

$$T_{i,j,k} = |L(x_i, y_j, t_0) - M(x_i, y_j, t_k)| \quad (5)$$

$$d_{i,j,k} = \sqrt{(x_i - x_{w/2})^2 + (y_j - y_{w/2})^2} \quad (6)$$

$$D_{i,j,k} = 1.0 + d_{i,j,k}/A \quad (7)$$

其中,  $A$  是一个常量,定义了  $S_{i,j,k}$ 、 $T_{i,j,k}$  与  $D_{i,j,k}$  之间重要程度的比例。该相似像元的权重  $w_{i,j,k}$  计算公式为:

$$w_{i,j,k} = S_{i,j,k} \times T_{i,j,k} \times D_{i,j,k} \quad (8)$$

或

$$w_{i,j,k} = \ln(S_{i,j,k} \times B + 1) \times \ln(T_{i,j,k} \times B + 1) \times D_{i,j,k} \quad (9)$$

其中,  $B$  是取决于传感器分辨率间差距的比例因子。求出滑动窗口内所有筛选后相似像元权重  $w_{i,j,k}$  后,再对其进行归一化处理,即可求得联合权重矩阵  $W$ ,从而预测出未知时刻高分辨率遥感影像像素的分辨率。

由于 MODIS 影像和 Landsat 影像间分辨率差距过大,在重采样的过程中像元混合严重,制约了 STARFM 方法预测的准确率<sup>[9]</sup>,本文将引入改进的 SRCNN 来克服这一现象,提高时空融合的质量。

另外,STARFM 在引入滑动窗口技术时权重的计算取决于光谱距离、时间距离与空间距离的组合,可视为一种根据专家经验知识提取的手工特



征。本文基于 STARFM 的基本思想,采用 SRCNN 作为基本框架,利用神经网络来自动计算该特征,充分利用深度学习强大的特征提取与表达能力,进一步提高时空融合的质量。

### 3 实验结果与分析

#### 3.1 研究区域及数据源

本文选择在遥感时空融合领域中经典的 MODIS 和 Landsat 数据进行融合,充分利用二者分别在时间分辨率和空间分辨率上的优势。MODIS 数据选择了空间分辨率为 250 m、重访周期为 1 天的 MOD09Q1(8 天合成数据)的第 1 波段红色波段和第 2 波段近红外波段数据;Landsat 数据选取了空间分辨率为 30 m、重访周期为 16 天的 Landsat8 相应的第 4、5 波段<sup>[18]</sup>,所举图例均以红色波段实验数据作为代表。本文的研究区域位于陕蒙交界一带,地理坐标  $38^{\circ}47'59''N \sim 39^{\circ}12'5''N$ ,  $110^{\circ}20'35''E \sim 110^{\circ}49'6''E$ ,实验截取的地理范围为  $45\text{ km} \times 45\text{ km}$ 。

本文实验所用数据均来源于 NASA 官网 USGS (<https://earthexplorer.usgs.gov/>),选取了 2015 年~2019 年该研究区域内所有云量少于 1% 的数据。使用 MCTK (MODIS Conversion Toolkit) 工具将 MODIS 数据从 HDF 数据格式转换为 TIF 数据格式,并重投影到 UTM-WGS84 坐标系,所有数据均使用 ENVI 软件根据对应坐标裁剪到指定区域并进行配准。从图 3 可以看到,低分辨率的 MOD09Q1 影像在部分地区像元混合现象较严重。从图 4 可以看到该研究区域地物类型复杂,包括水体、山脉、沙地、林地、耕地、建筑物、道路等,且部分地物相间分布。该研究区域属于使用传统时空融合方法较难处理的类型。

#### 3.2 实验方法

##### 3.2.1 超分辨率重建实验方法

传统 SRCNN 在处理自然图像时是一种单幅图像超分辨率重建技术,即样本和标签是基于同一幅图像变换所得,在训练之前先将样本图像进行降采样得到分辨率仅为原始图像  $1/3$  的低分辨率图像;再将该降采样图像使用双三次插值法放大至原样本图像尺寸,此时仍视其为低分辨率图像,并作为 SRCNN 的输入样本。在进行训练时, SRCNN 将没有经过降采样处理的原样本图像作为高分辨率先验知识进行学习,再把训练得到的网络作用于

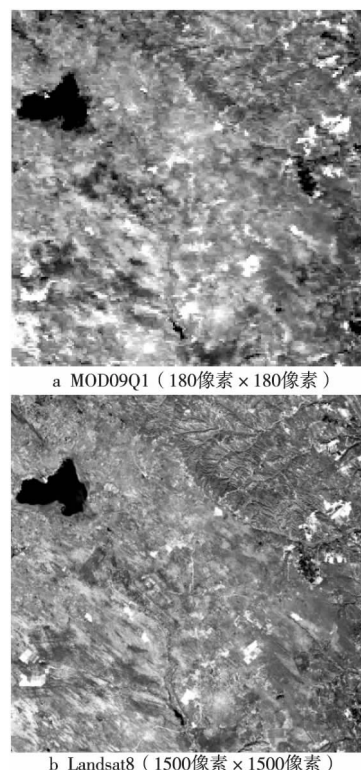


Figure 3 Comparison of MOD09Q1 and Landsat8 in the study area (2018, 10, 29)

图 3 研究区域(2018, 10, 29)不同传感器遥感影像对比图

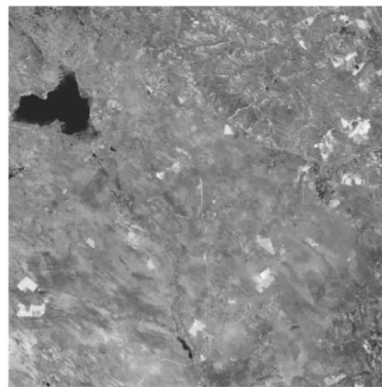


Figure 4 Landsat8 true color image in the study area (2019, 04, 23)

图 4 研究区域(2019, 4, 23)Landsat8 真彩色遥感影像图  
相同类型的测试集图像上进行检测<sup>[16]</sup>。

本文在引入 SRCNN 时,进行了 2 种尝试。以上述第 1 次学习为例,方法 1 是直接使用高分辨率遥感影像进行单幅图像超分辨率重建,即使用上述传统的 SRCNN 只训练降采样后的 90 m 分辨率 Landsat8 图像这一组数据,再将所得到的网络作用于同一时刻同一地区的低分辨遥感影像,即经过重采样的 90 m 分辨率 MOD09Q1 数据,对比 90 m 分辨率的 Landsat8 数据检测实验结果;方法 2 是使用已有的高分辨率图像作为同一时刻同一地区的低分辨率图像的先验知识进行学习,本文采用经

过采样后所得同为 90 m 分辨率的 Landsat8 图像和 MOD09Q1 图像 2 组数据来训练 SRCNN,通过该网络可以得到重建后 90 m 分辨率的 MOD09Q1 图像,同样对比降采样的 90 m 分辨率 Landsat8 数据检测实验结果。

在测试集上得到的实验结果对比图如图 5 所示(像素数没对应是由于外围元素卷积运算时不填充),从中可以看到,同样以 MOD09Q1 数据信息作为基础,相比方法 1,方法 2 所得图像在风格上与 Landsat8 数据更为接近。本文采用均方误差 MSE (Mean Squared Error) 作为损失函数来训练 SRCNN。实验结果中方法 1 的 MSE 值为 0.087 575 69,方法 2 的 MSE 值为 0.042 040 63,方法 2 相比方法 1 的实验结果具有更丰富的空间细节信息。

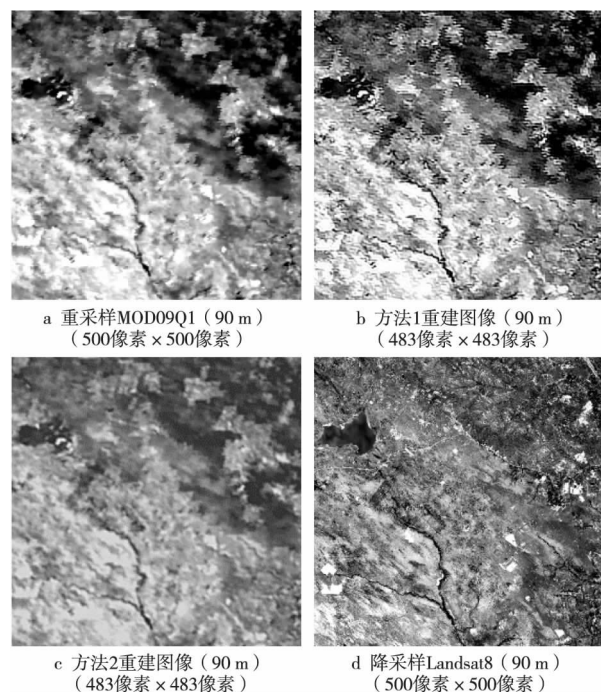


Figure 5 Comparison of restructured images with 90-meter resolution by different methods in the study area (2015. 7. 1)

图 5 研究区域(2015. 7. 1)不同方法重建得到 90 m 分辨率的实验结果对比图

本文使用方法 2 在第 1 次学习后得到了具有较高质量 90 m 分辨率的 MOD09Q1 重建图像,将该图像重采样至 30 m 分辨率后,再次利用 SRCNN 网络使用方法 2 进行第 2 次超分辨率重建。本次重建使用 30 m 分辨率的原始 Landsat8 遥感影像作为其先验知识,得到如图 6c 所示 30 m 分辨率的第 2 次超分辨率重建结果。对比相应的 Landsat8 数据,其 MSE 值为 0.029 617 50。本文将该结果作为 STARFM 模型中低分辨率遥感影像的输入。

图 6 所示为本文通过二次学习进行超分辨率

重建得到的实验结果与原始数据对比图,从中可以看到,图 6a~图 6d 空间细节逐渐丰富。从整体角度来看,可以认为实验是以图 6d 作为图 6a 的先验知识训练神经网络的,中间先经过图 6b 最终得到图 6c。图 6d 作为先验知识空间细节信息最为丰富,但时间分辨率为 16 天,而实验得到的图 6c 在空间细节上仅次于图 6d,且时间分辨率只有 1 天,本文将其与图 6d 进行时空融合,可以得到时间分辨率为 1 天、空间细节信息更为丰富的合成影像,进一步提升了遥感影像时空融合的精度。

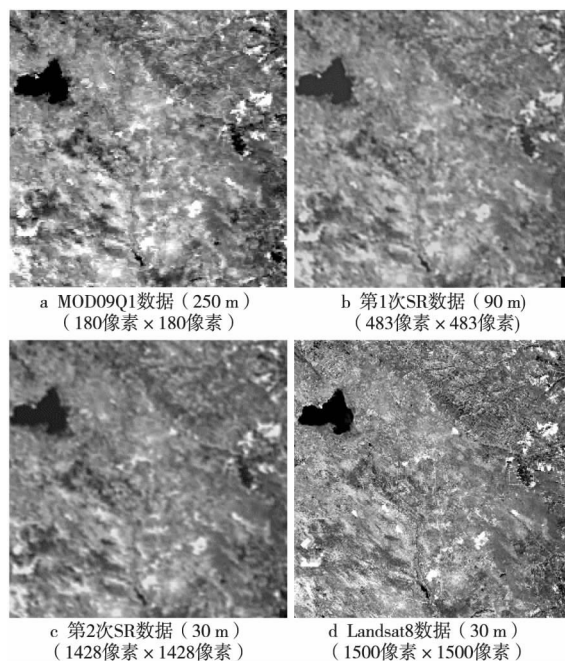


Figure 6 Comparison of original data and reconstructed results in the study area (2018. 10. 29)

图 6 研究区域(2018. 10. 29,红色波段) 实验数据与实验结果对比图

### 3. 2. 2 时空融合实验方法

STARFM 作为应用最为广泛的时空融合模型,其基本原理经过众多实际应用的考验,证实了其正确性与优越性,并且在该过程中,其所存在的一些不足也被学者们加以改善。本文在 3. 2. 1 节通过二次学习使用 SRCNN 进行超分辨率重建的方法,改善了其在低空间分辨率数据重采样过程中像元混合严重所导致的精度不足现象。本文采用 SRCNN 作为基本框架,通过深度学习神经网络方法来提取图像特征,发现并学习高低分辨率数据之间的映射关系。

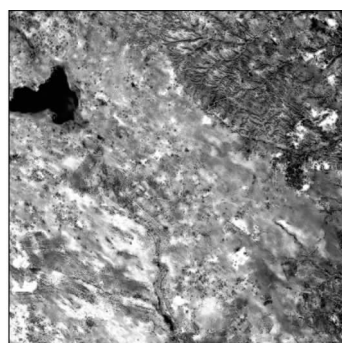
STARFM 的基本思想就是利用已知时刻的高低空间分辨率数据和未知时刻的低空间分辨率数据来对未知时刻的高空间分辨率数据进行预测,如式(2)所示。STARFM 在具体实现时为了减轻

像元混合现象的影响,通过滑动窗口技术引入邻近像元信息来计算中心像元的反射率,根据特定规则筛选出相似像元,并排除劣质像元,根据图像对间的光谱距离、时间距离以及窗口内部的空间距离由特定的公式来计算出像元对应的权重,得到滑动窗口的权值矩阵<sup>[8]</sup>。

STARFM 在实现过程中无论是在滑动窗口内筛选或排除像元时所用的规则,还是计算权重时所用的特定公式,都属于专家知识范畴,即手工提取的特征。本文采用神经网络进行深度学习来替代该过程,实现特征的自动提取。

基于 STARFM 的基本思想,在已知  $t_0$  时刻高分辨率数据与  $t_k$  时刻低分辨率数据时,通过式(2)计算所得到的数据可视为预测理论值,本文将其作为学习样本,并将对应时刻的高分辨率实际数据作为标签,输入 SRCNN 中进行训练,即可学习得到预测理论值与实际标签间所存在的映射关系。STARFM 所使用的滑动窗口计算方法也是对该映射的一种直接计算方式。

图 7a 为 STARFM 方法融合所得预测结果,本文使用上述网络进行实验,得到的实验结果如图 7b 所示。



a STARFM融合预测结果(30 m)  
(1428像素×1428像素)



b 本文方法融合预测结果(30 m)  
(1428像素×1428像素)

Figure 7 Comparison of fusion images by STARFM and the proposed method in the study area (2018, 10, 29)

图 7 研究区域(预测 2018. 10. 29, 红色波段)  
STARFM 融合方法与本文改进融合方法对比图

从图 7 中可以明显看出,原先在 MOD09Q1 及其重建图像中只显示为图像斑块的地方(像元混合严重),融合结果具有了与 Landsat8 原始影像相同的空间细节信息,预测结果对比真实结果 MSE 值为 0.008 333 66。实验图中 7a 与图 7b 都是基于同样的数据进行融合,即低分辨率图像均经过本文 2.1 节超分辨率重建处理。可以看到,使用手工特征的 STARFM 方法融合预测的结果虽然也明显呈现了更多的空间细节,图像风格也与 Landsat8 数据更加接近,但在部分区域图像斑块仍然较多,在细节上与本文的实验结果(图 7b)仍有一定差距,其作为融合预测结果与真实数据相比的 MSE 值为 0.017 497 47。故采用神经网络自动提取特征的方法相比利用滑动窗口手工提取特征的方法提升了融合质量。

### 3.2.3 实验结果与分析

本文已经从主观评价的角度对实验结果进行了分析。此外,本文采用峰值信噪比 PSNR(Peak Signal-to-Noise Ratio)和结构相似性 SSIM(Structural Similarity Index)作为客观评价指标来对实验结果进行进一步分析,其计算公式分别如下所示:

$$PSNR(f_1, f_2) = 10 \times \lg \frac{255^2}{MSE(f_1, f_2)} \quad (10)$$

$$SSIM(f_1, f_2) = \frac{(2\mu_1\mu_2 + C)(2\sigma_{12} + C_2)}{(\mu_1^2 + \mu_2^2 + C_1)(\sigma_1^2 + \sigma_2^2 + C_2)} \quad (11)$$

其中,  $MSE(f_1, f_2)$  表示 2 幅图像的均方误差,  $\mu_1$ 、 $\mu_2$ 、 $\sigma_1$ 、 $\sigma_2$  分别表示 2 幅图像的均值和方差,  $\sigma_{12}$  表示 2 幅图像间的协方差,  $C_1$ 、 $C_2$  是用来维持稳定的常数。

PSNR 是一种应用非常广泛的图像客观评价指标, PSNR 越大表示图像的失真越小,然而其作为一种基于误差敏感的图像质量评价指标,并未考虑到人眼的视觉特性(人眼对空间频率较低的对比差异敏感度较高,对亮度对比差异的敏感度较色度高,对一个区域的感知结果会受到其周围邻近区域的影响等),因而经常出现评价结果与人的主观感觉不一致的情况。而 SSIM 分别从亮度、对比度和结构 3 方面来度量图像相似性, SSIM 值越大表示图像相似性越高、失真越小,故本文采用 PSNR 和 SSIM 来衡量实验结果的优劣。

表 1 给出了本文的客观指标评价结果。对比第 2~4 行数据可知,在重建至中间分辨率时,使用的 2 种重建方法的 PSNR 差距不大,但在 SSIM

上方法 2 优势明显。通过对比第 5 行数据和第 6 行数据可知,本文采用的通过二次学习方式进行 SRCNN 重建方法相比 STARFM 所使用的简单重采样方法,PSNR 和 SSIM 都高,尤其是在 SSIM 上提升效果显著。通过对比最后 2 行数据可知,本文采用 SRCNN 自动提取特征进行融合的方法相比 STARFM 使用滑动窗口技术人工提取特征进行融合的方法,PSNR 和 SSIM 都高,融合效果更好。通过对比客观评价指标,表明本文改进方法提高了遥感数据时空融合的质量。

Table 1 Objective evaluation results  
表 1 客观指标评价结果

实验图像	PSNR	SSIM
简单重采样至 90 m 分辨率	29.960 263	0.458 496 53
图 5b 方法 1 重建至 90 m 分辨率	29.965 455	0.354 222 75
图 5c 方法 2 重建至 90 m 分辨率	29.966 763	0.740 235 38
简单重采样至 30 m 分辨率	30.150 611	0.801 174 62
二次学习重建至 30 m 分辨率	30.199 254	0.926 231 97
STARFM 融合结果	30.251 345	0.790 622 25
本文方法融合结果	30.318 176	0.970 218 47

4 结束语

本文对传统 STARFM 方法进行了 2 点改进,一方面基于学习的时空融合模型方法,对低空间分辨率数据进行超分辨率重建来代替原来的直接重采样方法。本文为了缓解分辨率差距过大的影响,通过二次学习方式,利用 SRCNN 网络实现了低空间分辨率数据的超分辨率重建,并且在重建过程中使用了不同传感器的高分辨率数据作为先验知识,丰富了低分辨率图像的细节信息。另一方面,本文基于 STARFM 的基本思想,以 SRCNN 作为框架,在融合过程中采用深度学习自动提取特征,相比原来利用滑动窗口手工提取特征的方法,明显改善了融合质量。

本文提出的基于深度学习和超分辨率重建的时空融合方法,虽然取得了良好的融合结果,但算法时间复杂度较高。在进行超分辨率重建时,本文选取了网络设计相对简单的 SRCNN,避免了其它更为复杂的网络,在保证重建质量的前提下不再提升复杂度。本文下一步将深入研究 SRCNN 的并行化方法,尝试将 2 次超分辨率重建过程融入到深度学习中,并完善端到端的训练网络,以降低本文方法的时间复杂度。

本文在进行具体的时空融合时使用了神经网

络自动提取特征,学习从预测理论值到对应标签间的映射关系,前期虽然需要付出较多的时间成本进行学习,但在学习到映射关系后即可快速得到需要进行融合预测的图像。相对而言,传统 STARFM 方法虽然不需要前期学习的时间成本,但每次进行融合预测时都需要复杂度较高的大量计算,理论上,如若需要融合预测大量的数据,本文方法在时间复杂度上相对而言具有优势。

参考文献:

[1] Li D R, Tong Q X, Li R X, et al. Current issues in high-resolution earth observation technology [J]. Science China Earth Science, 2012, 55: 1043-1051.

[2] Liu Jian-bo, Ma Yong, Wu Yi-tian, et al. Review of methods and applications of high spatiotemporal fusion of remote sensing data [J]. Journal of Remote Sensing, 2016, 20(5): 1038-1049. (in Chinese)

[3] Song H H, Huang B. Spatiotemporal satellite image fusion through one-pair image learning [J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 51(4): 1883-1896.

[4] Dong Wen-quan, Meng Ji-hua. Review of spatiotemporal fusion model of remote sensing data [J]. Remote Sensing for Land & Resources, 2018, 30(2): 1-11. (in Chinese)

[5] Huang Bo, Zhao Yong-quan. Research status and prospect of spatiotemporal fusion of multi-source satellite remote sensing imagery [J]. Acta Geodaetica et Cartographica Sinica, 2017, 46(10): 1492-1499. (in Chinese)

[6] Huang Wei. Research on pixel-level image fusion [D]. Shanghai: Shanghai Jiao Tong University, 2008. (in Chinese)

[7] Malenovsky Z, Bartholomeus H M, Acirbi-Junior F W, et al. Scaling dimensions in spectroscopy of soil and vegetation [J]. International Journal of Applied Earth Observation and Geoinformation, 2007, 9(2): 137-164.

[8] 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.

[9] Zhao Yong-guang, Huang Bo, Wang Chao-liang. Multi-temporal MODIS and Landsat reflectance fusion method based on super-resolution reconstruction [J]. Journal of Remote Sensing, 2013, 17(3): 590-608. (in Chinese)

[10] Zhang Xiao-ling. Research on compression and super-resolution restoration of remote sensing image [D]. Beijing: Beijing University of Technology, 2006. (in Chinese)

[11] Lu Wan-yun, Wang Ji-zhou. Survey of super resolution processing method of remote sensing image [J]. Science of Surveying and Mapping, 2016, 41(12): 53-58. (in Chinese)

[12] Su Heng, Zhou Jie, Zhang Zhi-hao. Survey of super-resolution image reconstruction methods [J]. Acta Automatica Sinica, 2013, 39(8): 1202-1213. (in Chinese)

[13] Chantas G K, Galatsanos N P, Woods N A. Super-resolution



- based on fast registration and maximum a posteriori reconstruction[J]. IEEE Transactions on Image Processing, 2007,16(7):1821-1830.
- [14] Freeman W T, Pasztor E C, Carmichael O T. Learning low-level vision [J]. International Journal of Computer Vision, 2000,40(1):25-47.
- [15] Sun Xu, Li Xiao-guang, Li Jia-feng, et al. Review on deep learning based image super-resolution restoration algorithms [J]. Acta Automatica Sinica, 2017,43(5):697-709. (in Chinese)
- [16] Dong C, Loy C C, He K M, et al. Image super-resolution using deep convolutional networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016,38(2):295-307.
- [17] Yang J C, Wright J, Huang T S, et al. Image super-resolution via sparse representation [J]. IEEE Transactions on Image Processing, 2010,19(11):2861-2873.
- [18] Xie Deng-feng, Zhang Jin-shui, Pan Yao-zhong, et al. Fusion of MODIS and Landsat 8 images to generate high spatial-temporal resolution data for mapping autumn crop distribution [J]. Journal of Remote Sensing, 2015,19(5):791-805. (in Chinese)
- 与 Landsat 反射率融合方法[J]. 遥感学报, 2013,17(3):590-608.
- [10] 张晓玲. 遥感图像的压缩和超分辨率复原技术研究[D]. 北京:北京工业大学, 2006.
- [11] 陆婉芸, 王继周. 遥感影像超分辨率处理方法与研究进展[J]. 测绘科学, 2016,41(12):53-58.
- [12] 苏衡, 周杰, 张志浩. 超分辨率图像重建方法综述[J]. 自动化学报, 2013,39(8):1202-1213.
- [15] 孙旭, 李晓光, 李嘉峰, 等. 基于深度学习的图像超分辨率复原研究进展[J]. 自动化学报, 2017,43(5):697-709.
- [18] 谢登峰, 张锦水, 潘耀忠, 等. Landsat 8 和 MODIS 融合构建高时空分辨率数据识别秋粮作物[J]. 遥感学报, 2015,19(5):791-805.

## 附中文参考文献:

- [2] 刘建波, 马勇, 武易天, 等. 遥感高时空融合方法的研究进展及应用现状[J]. 遥感学报, 2016,20(5):1038-1049.
- [4] 董文全, 蒙继华. 遥感数据时空融合研究进展及展望[J]. 国土资源遥感, 2018,30(2):1-11.
- [5] 黄波, 赵涌泉. 多源卫星遥感影像时空融合研究的现状及展望[J]. 测绘学报, 2017,46(10):1492-1499.
- [6] 黄伟. 像素级图像融合研究[D]. 上海:上海交通大学, 2008.
- [9] 赵永光, 黄波, 汪超亮. 基于超分辨率重建的多时相 MODIS

## 作者简介:



张永梅(1967-), 女, 山西太原人, 博士, 教授, CCF 会员(10935S), 研究方向为图像处理和人工智能。E-mail: zhangym@ncut.edu.cn

ZHANG Yong-mei, born in 1967, PhD, professor, CCF member(10935S), her research interests include image processing, and artificial intelligence.



滑瑞敏(1994-), 男, 山西临汾人, 硕士生, 研究方向为图像处理和人工智能。E-mail: hua8078@foxmail.com

HUA Rui-min, born in 1994, MS candidate, his research interests include image processing, and artificial intelligence.