

单样本对卷积神经网络遥感图像时空融合

李云飞¹, 李军¹, 贺霖²

1. 中山大学 地理科学与规划学院 广东省城市化与地理环境空间模拟重点实验室, 广州 510275;

2. 华南理工大学 自动化科学与工程学院, 广州 510640

摘要: 遥感图像时空融合是一种生成兼具高时空分辨率的合成遥感数据的技术。近年来, 产生了一些基于卷积神经网络的时空融合方法。这些方法效果良好, 但需要较多的图像样本对训练模型, 限制了它们的应用。针对此问题, 本文提出了一种单样本对卷积神经网络时空融合方法 (SS-CNN)。该方法以高空间分辨率图像的波段平均图像提供的空间信息激励卷积神经网络建立高、低空间分辨率图像间的超分关系, 进而利用该超分关系映射求解目标高空间分辨率图像。在实验中使用两个模拟数据集和一个真实数据集对该方法进行了测试, 并与两种常用的时空融合方法做了比较。实验结果表明, SS-CNN 在单样本对训练的情况下, 可以较好地预测地物的物候变化和类型的变化, 且在异质性高、地块破碎的区域表现良好。其不足之处在于会在地物边界上会造成轻微的模糊, 将来需针对此问题做进一步改进。

关键词: 遥感, 时空融合, 遥感图像, 单样本对, 卷积神经网络

引用格式: 李云飞, 李军, 贺霖. 2022. 单样本对卷积神经网络遥感图像时空融合. 遥感学报, 26(8): 1614-1623

Li Y F, Li J and He L. 2022. Convolutional neural network based single image pair method for spatiotemporal fusion. National Remote Sensing Bulletin, 26(8): 1614-1623 [DOI: 10.11834/jrs.20219348]

1 引言

随着越来越多对地观测卫星的发射, 我们能获取到的不同类型遥感图像数据日益增多。卫星遥感传感器在朝着高空间、高时间、高光谱分辨率的方向发展。然而目前来说, 空间、时间和光谱3种分辨率之间的相互制约在硬件技术上还没有被解决 (黄波和赵涌泉, 2017), 即没有单一卫星遥感传感器可以提供时序密集且空间分辨率较高的数据。同时, 云干扰也导致很多高空间分辨率遥感图像无法使用, 使得可用的高空间分辨率数据变得更少。而对于植物物候监测, 土地利用和土地覆盖变化监测, 农业评估等地表动态研究来说, 在时间序列上密集, 同时空间分辨率又满足要求的遥感数据集常常是必须的 (Zhu 等, 2016)。为了解决这个矛盾, 遥感数据时空融合技术应运而生。其目的是将高空间低时间分辨率的遥感数据 (如 Landsat, SPOT 数据) 与低空间高时间分辨率的遥感数据 (如

MODIS, AVHRR 数据) 进行融合, 得到高空间同时高时间分辨率的合成数据。目前已有很多研究应用到时空融合方法, 例如作物估产 (Liao 等, 2019), 作物监测 (Nduati 等, 2019), 植被生产力评估 (Zhang 等, 2019), 蒸腾量估算 (Knipper 等, 2019), 环境动态监测 (Heimhuber 等, 2018), 悬浮颗粒物监测 (Pan 等, 2018) 等。为了叙述的清晰方便, 下文中将高空间低时间分辨率的图像称为细图像且用 F 表示, 将低空间高时间分辨率的图像称为粗图像且用 C 表示, $k-1, k, k+1$ 表示数据获取时相, \hat{F}_k 来表示最终的预测图像。现有时空融合方法最少需要用 F_{k-1}, C_{k-1} 和 C_k 3 幅图像来预测 \hat{F}_k , 这类方法可被称为“单样本对方法”, 需要多于 3 幅图像的方法可称为“多样本对方法”。

目前为止, 国内外学者已经提出了多种时空融合方法。其中大多数方法可归为如下 3 类: 基于加权的方法, 基于解混的方法, 基于学习的方法

收稿日期: 2019-10-18; 预印本: 2020-06-07

基金项目: 国家重点研究发展计划 (编号: 2017YFB0502900); 国家自然科学基金 (编号: 61771496, 61901208, 61571195); 广东省自然科学基金 (编号: 2016A030313254, 2016A030313516, 2017A030313382)

第一作者简介: 李云飞, 研究方向为多源遥感图像融合。E-mail: liyf18213483@163.com

通信作者简介: 李军, 研究方向为遥感图像处理等。E-mail: lijun48@mail.sysu.edu.cn

(Zhu等, 2016)。其中, 基于加权的方法中最有代表性的为STARFM (Gao等, 2006)。该方法假设所有粗图像 (C_{k-1} 和 C_k) 的像元都是纯像元, 进而认为地表反射率在不同时相粗图像上的变化就等于其在对应时相细图像上的变化, 因此将粗图像上的反射率变化量直接加至 F_{k-1} ; 然后使用邻域校正策略寻找每个像元的相似像元, 并给这些相似像元赋权加和, 得到最终预测值。此类方法还包括 STAARCH (Hilker等, 2009), ESTARFM (Zhu等, 2010), SADFAT (Weng等, 2014) 等。这些方法的主要问题为在进行邻域校正时, 其相似像元来自于 F_{k-1} , 此处隐含一个假设, 即地表在 $k-1$ 时相和 k 时相地物类型一致。如果在 $k-1$ 时相和 k 时相有地物类型变化, 例如发生了洪水, 或者农田变城市, 像元在 F_{k-1} 上的相似关系和在 F_k 上将完全不同, 此假设就会不合理并带来误差。

基于解混的方法将混合像元分解理论引入了时空融合, 如 ESTDFM (Zhang等, 2013), MSTDFM (Wu M等, 2015), STRUM (Gevaert 和 García-Haro, 2015) 和 FSDAF (Zhu等, 2016) 等。这类方法一般先对 F_{k-1} 进行分类, 然后假设 C_{k-1} 和 C_k 的像元都是由这些类别线性混合得到。然后基于此假设计算每个类别的反射率变化, 进而得到预测图像 \hat{F}_k 。可以看出, 这类方法的隐含假设与基于加权的方法一致。因此, 这类方法和基于加权的方法有着同样的问题。

基于学习的方法是将机器学习方法引入到时空融合中, 如 SPSTFM (Huang 和 Song, 2012), SPFMOL (Song 和 Huang, 2013), EBSCDL (Wu B等, 2015), 增强 SPFMOL (Li等, 2018), 时空谱联合稀疏方法 (Zhao等, 2018) 等。以 SPSTFM 为例, 该方法先使用稀疏表示建立 $\Delta F_{k-1,k+1}$ 与 $\Delta C_{k-1,k+1}$ (即细尺度的反射率变化和粗尺度的反射率变化) 之间的关系, 基于此关系利用 $\Delta C_{k-1,k}$ 和 $\Delta C_{k+1,k}$ 预测 $\Delta F_{k-1,k}$ 和 $\Delta F_{k+1,k}$, 进而得到预测图像 \hat{F}_k 。可以看出, 这类方法并不具有类似于前两种方法的假设, 所以其适用的范围更广。除了稀疏表示, 其他一些机器学习方法也被应用到这个领域中, 例如回归树 (Boyte等, 2018), 随机森林 (Ke等, 2016) 等。

近几年, 卷积神经网络也被应用到时空融合领域当中。目前基于卷积神经网络的时空融合方

法主要有两种, 分别是 STFDCNN (Song等, 2018) 和 StfNet (Liu等, 2019)。其中, STFDCNN 采用超分卷积神经网络来提高粗图像的空间分辨率, 然后用高通模型将其与相邻时相的细图像融合得到预测图像。需要注意的是, 此方法需要较多的训练数据, 基于单样本对训练 STFDCNN 效果相对有限。StfNet 提出在超分粗尺度的反射率变化图像时加入细图像来弥补空间信息的缺失, 并通过双目标联合训练来提高超分结果的准确性。需要指出的是, 此方法是一个多样本对方法, 无法在单样本对的情况下使用。上述两种方法在实际数据上都表现出了较好的性能, 但其共同的问题在于对先验数据的需求较多, 难以应用于单样本对的情况。如前所述, 高空间分辨率遥感数据缺失严重, 在某些区域, 比如华南的广东, 广西等省份, 常年被云遮蔽, 很可能全年都很难得到可用的高空间分辨率图像。因此, 样本数极为有限, 严重限制了基于卷积神经网络的时空融合方法的应用潜力。

为了解决上述的问题, 本文提出了一种单样本对卷积神经网络 (SS-CNN) 的时空融合方法, 本方法可以在单样本对的情况下使用, 并取得良好的融合效果。

2 方法

2.1 方法基本原理

图1是本文的单样本对卷积神经网络时空融合方法 (SS-CNN) 所使用的卷积神经网络基本结构。先将多波段细图像进行波段平均得到波段平均图像, 然后将其与插值后的多波段粗图像垒叠起来送入卷积层, 通过网络学习获取重建的多波段图像。

在时空融合问题中, 通常粗图像和细图像的空间分辨率差异巨大 (例如 MODIS 图像 500 m 分辨率, 而 Landsat 图像 30 m 分辨率), 这导致粗图像中的空间细节严重缺失, 再加上遥感图像本身纹理复杂, 直接使用自然图像的卷积神经网络超分方法难以从粗图像上恢复出细图像的细节 (Liu等, 2019)。因此, 在网络超分粗图像时加入细图像的波段平均图像来弥补空间细节的缺失。而用细图像的波段平均图像的目的在于防止细图像光谱对粗图像光谱有所干扰。具体来说, 超分过程

中,粗图像提供光谱信息,细图像的波段平均图像来提供空间细节信息。在网络的训练过程中,我们先用待预测细图像相邻时相的图像对(如

F_{k-1} 和 C_{k-1})训练卷积神经网络,建立粗图像与细图像之间的超分重建关系,然后再利用该关系预测细图像。

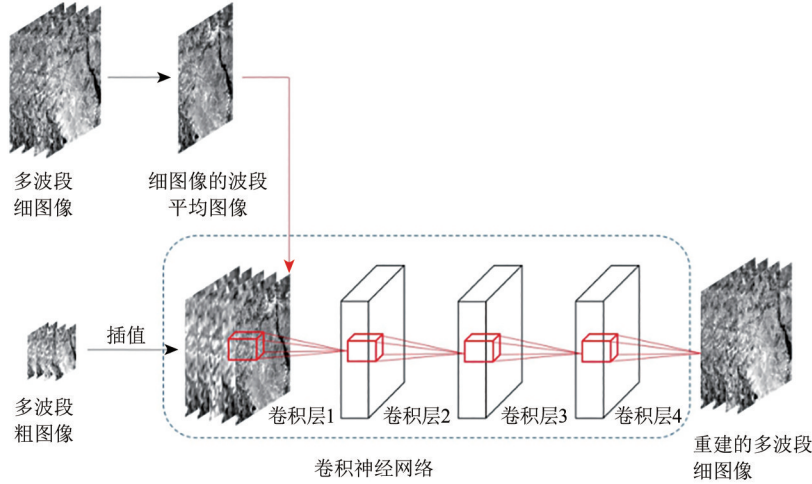


图1 本文方法SS-CNN所使用的网络基本结构

Fig. 1 Network structure used in the proposed method

图2是本文方法的流程示意图。 A^F 和 A^C 分别表示细图像和粗图像的波段平均图像。

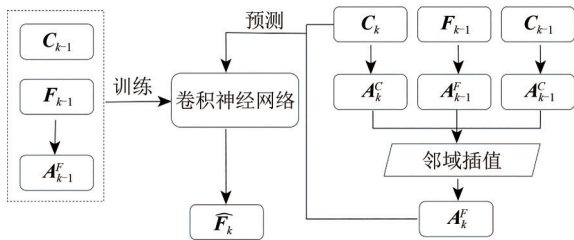


图2 本文方法流程示意图

Fig. 2 Flowchart of the proposed method

2.2 网络训练与结果预测

网络训练集可以表示为如下形式:

$$T = \{T_x, T_y\} \quad (1)$$

式中, T_x 表示训练集中的输入图像, T_y 表示训练集中的目标图像。其中:

$$T_x = \{C_{k-1}, A_{k-1}^F\} \quad (2)$$

$$T_y = \{F_{k-1}\} \quad (3)$$

式中, A_{k-1}^F 是 F_{k-1} 的波段平均图像, 其与粗图像 C_{k-1} 全叠作为网络的输入。采用均方差函数MSE (Mean Square Error) 作为损失函数, 其可以表示为:

$$\ell(\theta) = \frac{1}{N} \sum \left\| f(\theta | T_x) - T_y \right\|_F^2 \quad (4)$$

式中, $f(\cdot)$ 表示卷积神经网络的映射函数, θ 表示网络要优化求解的参数。卷积神经网络的结构如

图1所示,共4个卷积层,前3层的输出被Relu函数激活后传入下一层。同时选择Adam方法作为优化器来最小化损失函数 $\ell(\theta)$,求解 θ ,Adam方法可以在网络训练初期快速减小损失函数,从而大大提高网络训练效率(Kingma和Ba,2015)。网络训练完成后,就可以预测目标图像 \hat{F}_k 。如前所述,网络的输入是粗图像和对应的细图像的波段平均图像,因此预测 \hat{F}_k 需要先获取其波段平均图像 A_k^F 。本文提出一种简单的邻域加权方法获取 A_k^F 。

假设 A^F 的像元可由插值后的 A^C 的像元加权表示,且其权重不随时间变化,即

$$A_k^F(x, y) = \sum_{i=-s}^s \sum_{j=-s}^s w_{ij} \cdot A_k^C(x+i, y+j) \quad (5)$$

式中, w_{ij} 表示权重, s 为邻域大小, x 和 y 为像元坐标。为了获取 w_{ij} ,基于前述假设,可以利用 $k-1$ 时相图像,即

$$A_{k-1}^F(x, y) = \sum_{i=-s}^s \sum_{j=-s}^s w_{ij} \cdot A_{k-1}^C(x+i, y+j) \quad (6)$$

求解得到。最后,便可利用 A_k^F 和 C_k 进行预测,

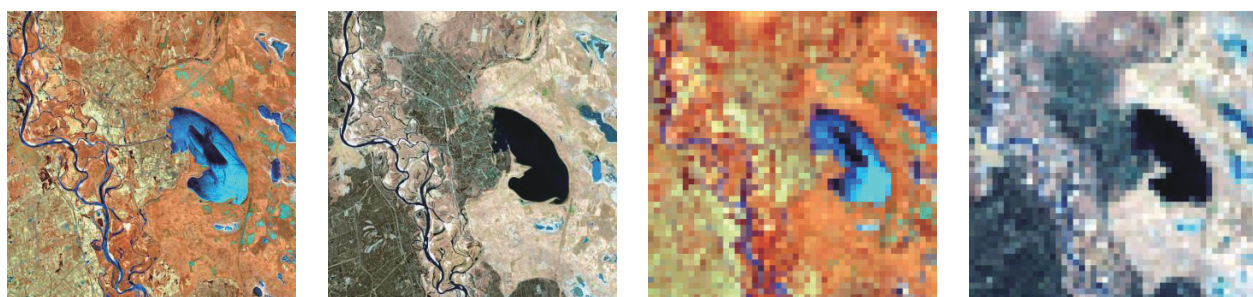
$$\hat{F}_k = f((C_k, A_k^F)) \quad (7)$$

3 实验与分析

为验证本文提出的SS-CNN方法,本章采用模拟数据与真实数据进行测试。模拟数据实验使用Landsat 8图像作为细图像(30 m空间分辨率),将

Landsat 8 图像降采样 20 倍 (600 m 空间分辨率) 后充当粗图像 (Zhu 等, 2016), 实验目的是排除传感器误差, 天气情况, 大气状况等一系列因素影响, 测试 SS-CNN 在理想数据上的效果。真实数据实验则使用真实的 Landsat 7 图像和 MODIS 图像。实验选择了两个对比方法: 第 1 个是 FSDAF (Zhu

等, 2016), FSDAF 是一个经典的基于解混的时空融合方法, 具有良好的融合效果; 另一个对比方法是 STFDCNN (Song 等, 2018), STFDCNN 理论上可以在单样本对的情况下使用 (StfNet 则不能)。这两种对比方法的参数设置与其出处文献中所述相同, 以保证其效果。



(a) 2017-03-08 高分辨率图像 (b) 2017-04-09 高分辨率图像 (c) 2017-03-08 低分辨率图像 (d) 2017-04-09 低分辨率图像
(a) Fine image of 2017-03-08 (b) Fine image of 2017-04-09 (c) Coarse image of 2017-03-08 (d) Coarse image of 2017-04-09

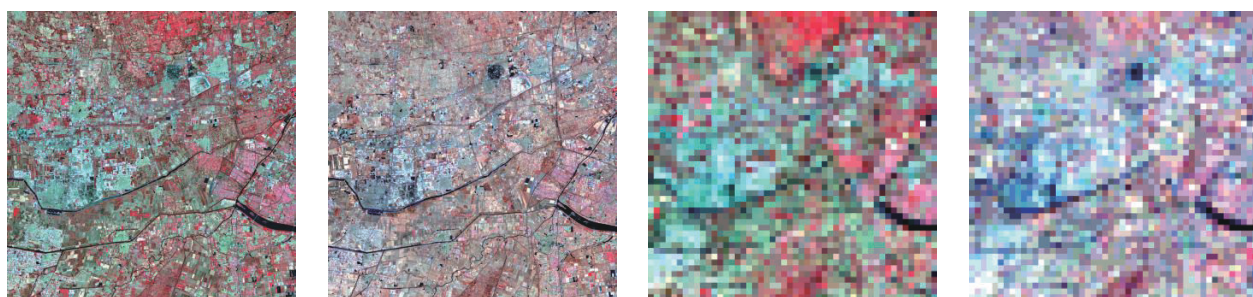
图3 巴彦查干数据集

Fig. 3 The Bayanchagan dataset

3.1 实验数据集

实验采用的模拟数据集分别来自黑龙江省杜尔伯特蒙古族自治县巴彦查干乡 ($46^{\circ}7'N$, $124^{\circ}26'E$) 和天津市西青区 ($39^{\circ}31'N$, $116^{\circ}28'E$), 如图 3 和图 4 所示。其中图 3 (a) 和图 3 (b) 分别是巴彦查干乡 2017 年 3 月 8 日和 2017 年 4 月 9 日的 Landsat 8 图像, 图 3 (c) 和图 3 (d) 则是图 3 (a) 和图 3 (b) 降采样之后得到的低分辨率图像。图 4 (a) 和图 4 (b) 分别是天津西青区 2017 年 10 月 30 日与 2017 年 12 月 17 日的 Landsat 8 图像, 图 4 (c) 和

图 4 (d) 为对应的低空间分辨率图像。两个数据集中的 Landsat 图像尺寸都为 1000×1000 , 包含 6 个波段, 分别是蓝光波段 ($0.45-0.51 \mu m$), 绿光波段 ($0.53-0.59 \mu m$), 红光波段 ($0.64-0.67 \mu m$), 近红外波段 ($0.85-0.88 \mu m$), 短波红外波段 1 ($1.57-1.65 \mu m$), 短波红外波段 2 ($2.11-2.29 \mu m$)。所有 Landsat 图像的原始 DN 值都已经转化为反射率, 并做了暗像元法大气校正, 且将反射率归一化到 0—255 的范围内。



(a) 2017-10-30 高分辨率图像 (b) 2017-12-17 高分辨率图像 (c) 2017-10-30 低分辨率图像 (d) 2017-12-17 低分辨率图像
(a) Fine image of 2017-10-30 (b) Fine image of 2017-12-17 (c) Coarse image of 2017-10-30 (d) Coarse image of 2017-12-17

图4 天津数据集

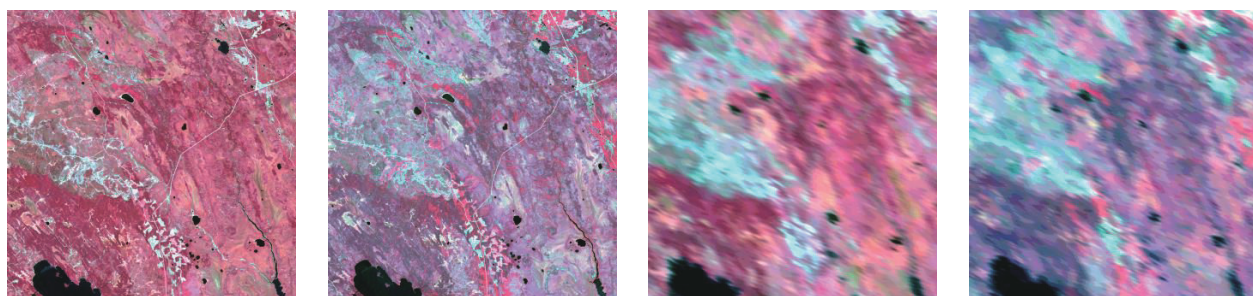
Fig. 4 The Tianjin dataset

巴彦查干地区在 2017 年 3 月 8 日之前下了大雪, 导致其旁边的向阳湖以及周围的其他水体都结了冰, 并且地面上也有残余冰雪存在。而到了

2017 年 4 月 9 日, 冰雪完全消融。选择这个数据集主要来测试本文 SS-CNN 对于地物类型变化的预测能力, 同时从图 3 (a) 和图 3 (b) 中也可以看出,

这段时间内巴彦查干地区的植被已经开始生长，这两个时相之间存在很明显的物候变化。而天津西青区从2017-10-30—2017-12-17日存在明显的物候变化，这点可以从图4(a)和图4(b)中看出。同时，由于它是城镇区域，地块破碎，地表异质性很高，这也给时空融合带来了一定的挑战。我们选择这个数据集用来测试本文SS-CNN在异质性高的区域物候变化的预测能力。

真实实验使用Gao等(2006)在文献中使用的



(a) 2001-05-24 高分辨率图像 (b) 2001-08-12 高分辨率图像 (c) 2001-05-24 低分辨率图像 (d) 2001-08-12 低分辨率图像
(a) Fine image of 2001-05-24 (b) Fine image of 2001-08-12 (c) Coarse image of 2001-05-24 (d) Coarse image of 2001-08-12

图5 BOREAS数据集

Fig. 5 The BOREAS dataset

后文中将各数据集的第一时相和第二时相的高分辨图像分别记为已知图像和待预测图像。

3.2 实验结果评价指标

为了对实验结果进行定量评估，选择4个评价指标：(1) 均方根误差 (RMSE)，用来评价预测图像和真实图像各个波段图像之间的总体反射率差异；(2) 结构相似度 (SSIM)，用来评价预测图像和真实图像各个波段图像之间的空间结构相似性 (Wang 等, 2004)；(3) ERGAS (Erreur Relative Global Adimensionnelle de Synthèse)，用来评价预测图像多个波段整体的光谱质量 (Renza 等, 2013)；(4) 光谱角 (SAM) (Park 等, 2007)，来评价预测图像和真实图像之间的多个波段整体的光谱误差。越小的RMSE, ERGAS, SAM和越大的SSIM代表预测结果越好。

3.3 参数设置

本文SS-CNN在进行邻域插值时将 s 经验性地设置为1，即一个边长为3的正方形窗口。另外，由于计算机GPU内存的限制，无法将整个图像直接输入训练，所以在实验中我们将其裁成带重叠的图像块输入训练，图像块大小为 32×32 ，重叠宽

“BOREAS (Boreal Ecosystem-Atmosphere Study southern study area)”数据集，如图5所示。该数据集来自 54°N , 124°W ，此区域植被生长期短，物候变化迅速 (Gao 等, 2006)。其中Landsat图像是ETM+图像 (30 m空间分辨率)，MODIS图像为MOD09GHK图像 (500 m空间分辨率)，两种图像都包括近红，红和绿3个波段，图像大小为1200像素 \times 1200像素。

度为16。网络的基本参数设置如表1所示，学习率设置为0.0001。本次实验代码在Tensorflow和Keras框架下编写，并在GTX 1080Ti GPU上运行。

表1 网络基本参数

Table 1 Essential parameters of the CNN used in the proposed SS-CNN

网络层	卷积核尺寸	卷积核步长	激活函数
卷积层1	7 \times 7 \times 96	(1,1)	Relu
卷积层2	5 \times 5 \times 96	(1,1)	Relu
卷积层3	3 \times 3 \times 96	(1,1)	Relu
卷积层4	3 \times 3 \times 6(或3)	(1,1)	—

3.4 结果与分析

图6展示了SS-CNN邻域插值方法预测的波段平均图像与真实的波段平均图像。从目视效果来看，3个数据集邻域插值预测的图像总体上都接近真实图像。但也可以看出，由于邻域插值带来了模糊效应，与真实图像相比，预测图像在地物边界较为模糊。考虑到插值是为图像融合过程提供空间细节，因此我们采用SSIM指标对预测的波段平均图像与真实波段平均图像的空间相似性做了定量评价，同时与已知图像的波段平均图像做了对比，结果如表2所示。可以看出，3个数据集预测图像的SSIM值都要更高，这说明邻域插值是有效的。

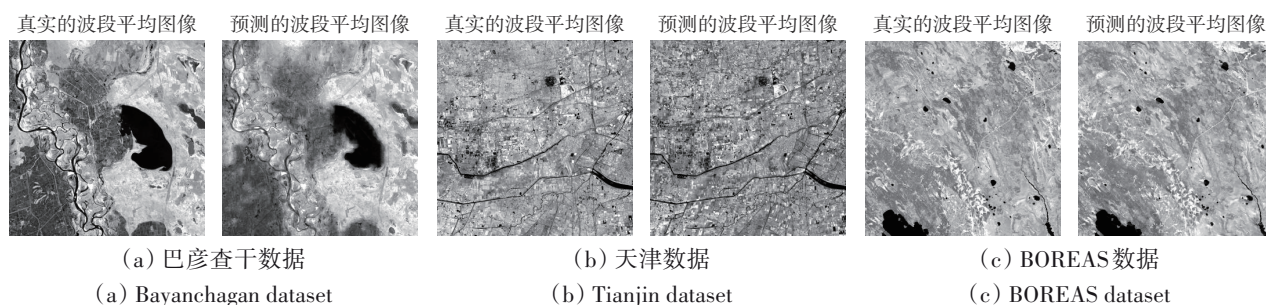


图6 基于邻域插值预测的波段平均图像结果

Fig. 6 The obtained band-average images

表2 邻域插值预测的波段平均图像定量评价

Table 2 Quantitative evaluation of the results of neighboring interpolation

	巴彦查干数据集		天津数据集		BOREAS数据集	
SSIM	已知图像的 波段平均图像	预测的 波段平均图像	已知图像的 波段平均图像	预测的 波段平均图像	已知图像的 波段平均图像	预测的 波段平均图像
	0.5125	0.8858	0.8741	0.9221	0.9536	0.9540

注：黑体表示最优值。

图7和图8分别显示了两个模拟数据实验的完整结果图像及子区域图像。图7(a)和(b)分别是巴彦查干乡2017年3月8日和2017年4月9日真实的Landsat 8图像，图7(c)–(e)分别是SS-CNN，STFDCNN和FSDAF的结果图像。可以看到，首先，SS-CNN取得得视觉效果最好，尤其是从子区域图像来看，SS-CNN成功预测到了冰雪已经消融，表明SS-CNN有预测地物类型变化的能力。其次，FSDAF的预测结果相当突出，然而FSDAF并没

有预测到冰雪消融带来的地物变化。此外，STFDCNN预测的结果图像出现了严重的畸变，已经不具有对比价值。这说明虽然STFDCNN从理论上来说可以在一对先验数据的情况下使用，但过少的训练数据会严重影响其效果。需要指出的是，SS-CNN的预测效果看起来较为模糊，这是因为邻域插值方法预测的波段平均图像模糊造成的。最后，表3是巴彦查干数据集实验结果的定量评价，可以看到SS-CNN几乎在所有指标上都取得了最优。

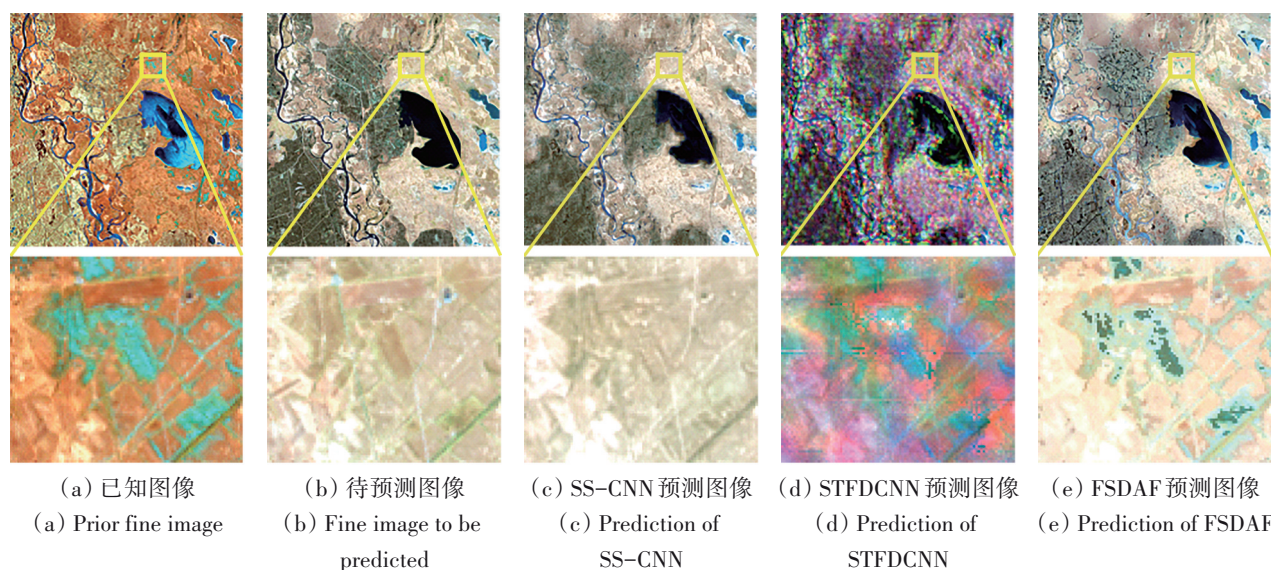


图7 巴彦查干数据集实验结果

Fig. 7 Experimental result from the Bayanchagan dataset

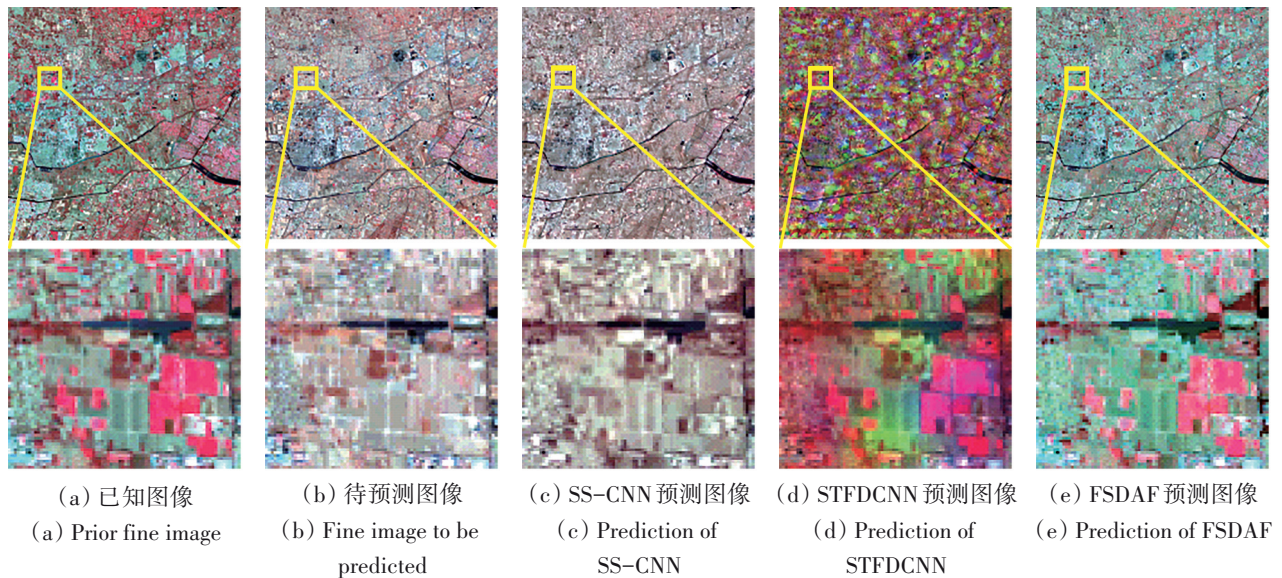


图8 天津数据集实验结果

Fig. 8 Experimental result from the TianJin dataset

表3 巴彦查干数据集实验结果评价指标

Table 3 Quantitative results of Bayanchagan dataset

评价指标	波段	FSDAF	STFDCNN	SS-CNN
RMSE	波段1	0.0150	0.0156	0.0127
	波段2	0.0199	0.0269	0.0149
	波段3	0.0277	0.0403	0.0194
	波段4	0.0423	0.0711	0.0305
	波段5	0.0483	0.0763	0.0433
	波段6	0.0386	0.0804	0.0383
SSIM	波段1	0.8475	0.8363	0.8860
	波段2	0.7960	0.6977	0.8754
	波段3	0.7481	0.6144	0.8666
	波段4	0.7486	0.5666	0.8608
	波段5	0.8360	0.7019	0.8645
	波段6	0.8737	0.6497	0.8747
ERGAS		1.2787	2.0656	1.0377
SAM		0.1143	0.2742	0.1030

注: 黑体表示最优值。

图8 (a) 和 (b) 分别是天津西青区2017年10月30日和2017年12月17日真实的Landsat 8图像, 图8 (c)—(e) 分别表示SS-CNN, STFDCNN和FSDAF的预测图像。同样SS-CNN的结果在视觉效果上更接近于真实图像, 说明SS-CNN在预测物候变化方面效果良好。FSDAF方法结果在本数据集上的预测结果视觉效果有限。STFDCNN的结果出现了严重的畸变。从子区域图像上来看, SS-CNN抓住了这两个时相之间大部分的物候变化,

尽管由于地物异质性高及部分地物面积过小导致有些地方预测并不正确。表4是天津数据集实验结果的定量评价, 从整体评价指标ERGAS和SAM上来看, SS-CNN的预测结果更好。

表4 天津数据集实验结果评价指标

Table 4 Quantitative results of Tianjin dataset

评价指标	波段	FSDAF	STFDCNN	SS-CNN
RMSE	波段1	0.0057	0.0079	0.0057
	波段2	0.0070	0.0433	0.0065
	波段3	0.0099	0.0385	0.0097
	波段4	0.0159	0.0204	0.0137
	波段5	0.0156	0.0235	0.0180
	波段6	0.0148	0.0189	0.0153
SSIM	波段1	0.9720	0.9464	0.9718
	波段2	0.9586	0.3741	0.9617
	波段3	0.9242	0.4486	0.9293
	波段4	0.8539	0.7736	0.8807
	波段5	0.8781	0.7607	0.8395
	波段6	0.8827	0.8257	0.8738
ERGAS		0.8250	2.3149	0.8221
SAM		0.0919	0.1619	0.0756

注: 黑体表示最优值。

图9展示了真实数据集上的实验结果, 图9 (a) 和 (b) 分别是BOREAS数据集2001年5月24日和2001年8月12日真实的Landsat 7图像, 图9 (c)—(e) 分别表示SS-CNN, STFDCNN和FSDAF的预测图像。从图9可以看出, SS-CNN的预测结果更

接近真实情况, 尤其是从子区域图像上来看; FSDAF也取得了较好的视觉效果, STFDCNN的结果仍然存在明显的畸变。表5是对于图9结果的定量比较, 其中FSDAF对于第一波段的预测是最好

的, SS-CNN对于第二波段的预测最好, 从ERAGS和SAM这两个整体指标上来看, SS-CNN取得了更好的效果, 并在数值上取得了较大的领先。

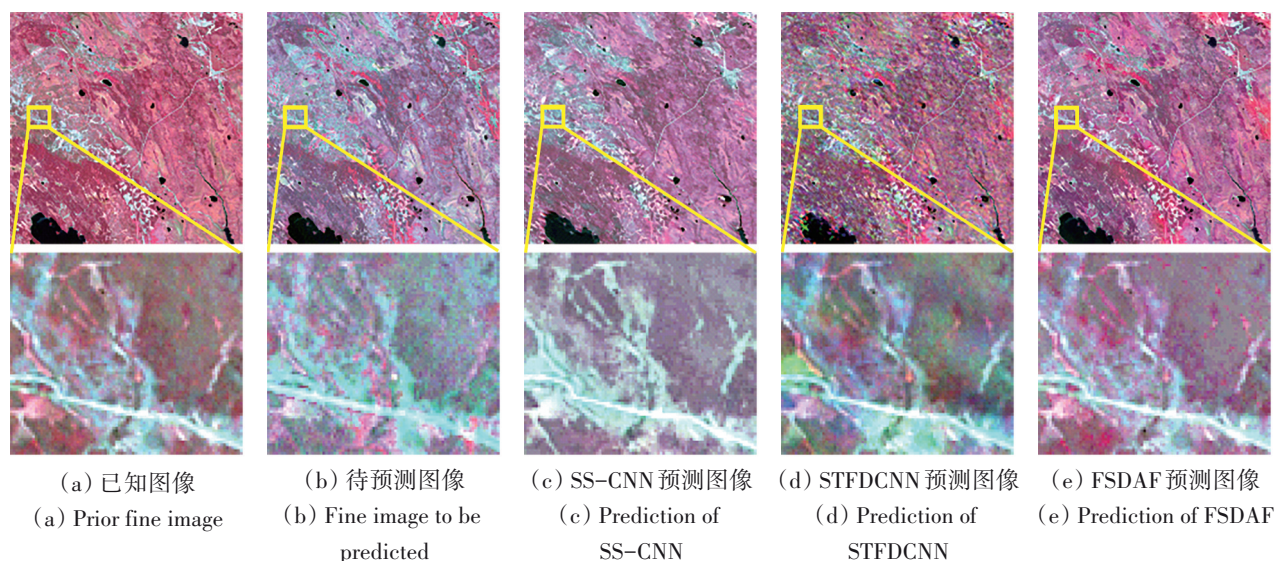


图9 BOREAS数据集实验结果

Fig. 9 Experimental result from the BOREAS dataset

表5 BOREAS数据集实验结果评价指标

Table 5 Quantitative results of BOREAS dataset

评价指标	波段	FSDAF	STFDCNN	SS-CNN
RMSE	波段1	0.0058	0.0141	0.0061
	波段2	0.0086	0.0233	0.0081
	波段3	0.0272	0.0351	0.0271
SSIM	波段1	0.9711	0.9168	0.9684
	波段2	0.9421	0.7419	0.9471
	波段3	0.8882	0.8205	0.8848
ERGAS		1.0372	2.5941	1.0010
SAM		0.0543	0.1107	0.0518

注: 黑体表示最优值。

从实验结果可以总结出以下3点:

(1) SS-CNN在单样本对的情况下, 可以较好地预测地物的物候变化和类型的变化, 且在异质性强、地块破碎的城镇区域可以做出良好的预测。

(2) 现有的基于卷积神经网络的时空融合方法在单样本对的情况下效果难以保障, 如STFDCNN在3个数据集中的预测结果都有较为严重的畸变。

(3) 需要指出的是, 由于SS-CNN采用邻域插值方法, 会导致地物边界的预测出现模糊现象。

这是SS-CNN以后需要重点提高的部分。

4 结论

针对目前基于卷积神经网络时空融合方法需要较多训练样本的问题, 本文提出了一种基于单样本对训练的卷积神经网络时空融合方法(SS-CNN), 本方法使用细图像的波段均值图像提供空间细节信息, 利用粗图像来提供光谱信息, 进而结合卷积神经网络建立粗图像与细图像之间的超分关系。在预测阶段, 提出采用邻域插值方法预测待求时相细图像的波段平均图像, 并基于训练好的网络对目标细图像进行预测。在实验部分, 使用了两个模拟的Landsat 8数据集和一个真实的Landsat 7-MODIS数据集测试了SS-CNN的效果, 结果证明SS-CNN在预测地物类型变化, 物候变化时候都表现良好, 并且在异质性高的区域具有较好的表现。但是SS-CNN用来预测空间细节信息邻域插值方法在地物边界上会出现模糊现象, 并影响最终的预测结果。将来会针对此问题做进一步研究。

参考文献(References)

Boyte S P, Wylie B K, Rigge M B and Dahal D. 2018. Fusing MODIS

- with Landsat 8 data to downscale weekly normalized difference vegetation index estimates for central great basin rangelands, USA. *GIScience and Remote Sensing*, 55(3): 376-399 [DOI: 10.1080/15481603.2017.1382065]
- Gao F, Masek J, Schwaller M and Hall F. 2006. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8): 2207-2218 [DOI: 10.1109/TGRS.2006.872081]
- Gevaert C M and García-Haro F J. 2015. A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion. *Remote Sensing of Environment*, 156: 34-44 [DOI: 10.1016/j.rse.2014.09.012]
- Heimhuber V, Tulbure M G and Broich M. 2018. Addressing spatio-temporal resolution constraints in Landsat and MODIS-based mapping of large-scale floodplain inundation dynamics. *Remote Sensing of Environment*, 211: 307-320 [DOI: 10.1016/j.rse.2018.04.016]
- Hilker T, Wulder M A, Coops N C, Linke J, McDermid G, Masek J G, Gao F and White J C. 2009. A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS. *Remote Sensing of Environment*, 113(8): 1613-1627 [DOI: 10.1016/j.rse.2009.03.007]
- Huang B, and Song H H. 2012. Spatiotemporal reflectance fusion via sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 50(10): 3707-3716 [DOI: 10.1109/TGRS.2012.2186638]
- Huang B, Zhao Y Q. 2017. Research status and prospect of spatiotemporal fusion of multi-source satellite remote sensing imagery. *Acta Geodaetica et Cartographica Sinica*, 46(10): 1492-1499 (黄波, 赵涌泉. 2017. 多源卫星遥感影像时空融合研究的现状及展. *测绘学报*, 46(10): 1492-1499) [DOI: 10.11947/j.AGCS.2017.20170376]
- Ke Y H, Im J, Park S and Gong H L. 2016. Downscaling of MODIS one kilometer evapotranspiration using Landsat-8 data and machine learning approaches. *Remote Sensing*, 8(3): 215 [DOI: 10.3390/rs8030215]
- Kingma D P and Ba J. 2015. Adam: a method for stochastic optimization//3rd International Conference on Learning Representations. San Diego: ICLR.
- Knipper K R, Kustas W P, Anderson M C, Alfieri J G, Prueger J H, Hain C R, Gao F, Yang Y, McKee L G, Nieto H, Hipps L E, Alsina M M and Sanchez L. 2019. Evapotranspiration estimates derived using thermal-based satellite remote sensing and data fusion for irrigation management in California vineyards. *Irrigation Science*, 37(3): 431-449 [DOI: 10.1007/s00271-018-0591-y]
- Li D C, Li Y R, Yang W F, Ge Y Q, Han Q J, Ma L L, Chen Y H and Li X. 2018. An enhanced single-pair learning-based reflectance fusion algorithm with spatiotemporally extended training samples. *Remote Sensing*, 10(8): 1207 [DOI: 10.3390/rs10081207]
- Liao C H, Wang J F, Dong T F, Shang J L, Liu J G and Song Y. 2019. Using spatio-temporal fusion of Landsat-8 and MODIS data to derive phenology, biomass and yield estimates for corn and soybean. *Science of the Total Environment*, 650: 1707-1721 [DOI: 10.1016/j.scitotenv.2018.09.308]
- Liu X, Deng X W, Chanussot J, Hong D F and Zhao B J. 2019. *StfNet*: a two-stream convolutional neural network for spatiotemporal image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9): 6552-6564 [DOI: 10.1109/TGRS.2019.2907310]
- Nduati E, Sofue Y, Matniyaz A, Park J G, Yang W and Kondoh A. 2019. Cropland mapping using fusion of multi-sensor data in a complex Urban/Peri-Urban area. *Remote Sensing*, 2(11): 207 [DOI: 10.3390/rs11020207]
- Pan Y Q, Shen F and Wei X D. 2018. Fusion of landsat-8/OLI and GO-CI data for hourly mapping of suspended particulate matter at high spatial resolution: a case study in the yangtze (Changjiang) estuary. *Remote Sensing*, 2(10): 158 [DOI: 10.3390/rs10020158]
- Park B, Windham W R, Lawrence K C and Smith D P. 2007. Contaminant classification of poultry hyperspectral imagery using a spectral angle mapper algorithm. *Biosystems Engineering*, 96(3): 323-333 [DOI: 10.1016/j.biosystemseng.2006.11.012]
- Renza D, Martinez E and Arquero A. 2013. A new approach to change detection in multispectral images by means of ERGAS index. *IEEE Geoscience and Remote Sensing Letters*, 10(1): 76-80 [DOI: 10.1109/LGRS.2012.2193372]
- Song H H and Huang B. 2013. Spatiotemporal satellite image fusion through one-pair image learning. *IEEE Transactions on Geoscience and Remote Sensing*, 51(4): 1883-1896 [DOI: 10.1109/TGRS.2012.2213095]
- Song H H, Liu Q S, Wang G J, Hang R L and Huang B. 2018. Spatiotemporal satellite image fusion using deep convolutional neural networks. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3): 821-829 [DOI: 10.1109/JSTARS.2018.2797894]
- Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600-612 [DOI: 10.1109/TIP.2003.819861]
- Weng Q H, Peng F and Feng G. 2014. Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. *Remote Sensing of Environment*, 145(8): 55-67 [DOI: 10.1016/j.rse.2014.02.003]
- Wu B, Huang B and Zhang L P. 2015. An error-bound-regularized sparse coding for spatiotemporal reflectance fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 53(12): 6791-6803 [DOI: 10.1109/TGRS.2015.2448100]
- Wu M Q, Huang W J, Niu Z and Wang C Y. 2015. Generating daily synthetic Landsat imagery by combining Landsat and MODIS data. *Sensors*, 15(9): 24002-24025 [DOI: 10.3390/s150924002]
- Zhang M, Lin H, Wang G, Sun H and Cai Y T. 2019. Estimation of vegetation productivity using a Landsat 8 time series in a heavily Urbanized area, central China. *Remote Sensing*, 11(2): 133 [DOI: 10.3390/rs11020133]
- Zhang W, Li A N, Jin H A, Bian J H, Zhang Z J, Lei G B, Qin Z H and Huang C Q. 2013. An enhanced spatial and temporal data fusion

- model for fusing Landsat and MODIS surface reflectance to generate high temporal Landsat-like data. *Remote Sensing*, 5(10): 5346-5368 [DOI: 10.3390/rs5105346]
- Zhao C Y, Gao X B, Emery W J, Ying W and Li J. 2018. An integrated spatio-spectral-temporal sparse representation method for fusing remote-sensing images with different resolutions. *IEEE Transactions on Geoscience and Remote Sensing*, 56(6): 3358-3370 [DOI: 10.1109/TGRS.2018.2798663]
- Zhu X L, Chen J, Gao F, Chen X H and Masek J G. 2010. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment*, 114(11): 2610-2623 [DOI: 10.1016/j.rse.2010.05.032]
- Zhu X L, Helmer E H., Gao F, Liu D S, Chen J and Lefsky M A. 2016. A flexible spatiotemporal method for fusing satellite images with different resolutions. *Remote Sensing of Environment*, 172: 165-177 [DOI: 10.1016/j.rse.2015.11.016]

Convolutional neural network based single image pair method for spatiotemporal fusion

LI Yunfei¹, LI Jun¹, HE Lin²

1. *Guangdong Provincial Key Laboratory of Urbanization and Geo-simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China;*
2. *School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China*

Abstract: Spatiotemporal fusion is a feasible way to provide synthetic satellite images with high spatial and high temporal resolution simultaneously. In recent years, some efficient STF methods based on Convolutional Neural Networks (CNNs) have been developed. However, these methods require a significant number of training image pairs, where each pair generally consists of a high spatial resolution image and a low spatial resolution image. Such a requirement limits the applicability of STF methods to actual scenarios because image pairs for training are not widely available in many cases. To overcome this important limitation, we introduce a CNN-based single image pair method for STF of remotely sensed images. Our method, called SS-CNN, uses the spatial information provided by the average image (obtained across available spectral bands) of the high spatial resolution image to perform CNN-based Super-Resolution Mapping (SRM) between the low and high spatial resolution images. The proposed SS-CNN has been tested in experiments using two simulated and one real dataset and compared with two commonly used spatiotemporal fusion methods. The experimental results show that SS-CNN can predict the phenological changes and land cover changes well. Plus, its performance in heterogeneous areas is remarkable. The disadvantage is that it will slightly blur the boundary, which needs to be further improved in the future.

Key words: remote sensing, spatio-temporal fusion, remote sensing images, single image pair, Convolutional Neural Networks (CNN)

Supported by National Key Research and Development Program of China (No. 2017YFB0502900); National Natural Science Foundation of China (No. 61771496, 61901208, 61571195); Guangdong Provincial Natural Science Foundation (No. 2016A030313254, 2016A030313516, 2017A03031338)