

基于双重注意力机制的遥感图像时空融合方法

范学晶, 薛笑荣, 杜意超

引用本文

范学晶, 薛笑荣, 杜意超. 基于双重注意力机制的遥感图像时空融合方法[J]. 计算机科学, 2024, 51(11A): 240200097-6.

FAN Xuejing, XUE Xiaorong, DU Yichao. Spatiotemporal Fusion Method for Remote Sensing Images Based on Dual Attention Mechanisms [J]. Computer Science, 2024, 51(11A): 240200097-6.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于双重标签分配的遥感有向目标检测方法](#)

Remote Sensing Oriented Object Detection Method Based on Dual-label Assignment
计算机科学, 2024, 51(11A): 240100058-9. <https://doi.org/10.11896/jsjcx.240100058>

[基于多任务学习的复杂城市遥感图像道路提取](#)

Road Extraction from Complex Urban Remote Sensing Images Based on Multi-task Learning
计算机科学, 2024, 51(11A): 240300095-8. <https://doi.org/10.11896/jsjcx.240300095>

[面向嵌入式平台的光学遥感图像舰船检测识别](#)

Ship Detection and Recognition of Optical Remote Sensing Images for Embedded Platform
计算机科学, 2024, 51(6A): 230700117-7. <https://doi.org/10.11896/jsjcx.230700117>

[基于改进FCOS的遥感图像舰船目标检测](#)

Ships Detection in Remote Sensing Images Based on Improved FCOS
计算机科学, 2024, 51(6A): 230700166-7. <https://doi.org/10.11896/jsjcx.230700166>

[结合多尺度卷积块与密集卷积块的遥感图像融合](#)

Remote Sensing Image Fusion Combining Multi-scale Convolution Blocks and Dense Convolution Blocks
计算机科学, 2024, 51(6A): 230400110-6. <https://doi.org/10.11896/jsjcx.230400110>

基于双重注意力机制的遥感图像时空融合方法

范学晶 薛笑荣 杜意超

辽宁工业大学电子与信息工程学院 辽宁 锦州 121001

(734057293@qq.com)

摘要 由于遥感成像技术条件的限制,难以同时获得既有高时间分辨率又具有高空间分辨率序列的遥感影像,通过时空融合技术可以生成同时具有高时间和高空间分辨率的遥感图像。近年来,时空融合方法层出不穷,这些方法效果良好,但在特征提取有效信息方面仍有不足。针对此问题,提出了一种基于双重注意力机制改进的深度神经网络(ADCSTFN),使得模型在全局的保留和细节场景的重建能力都得到了提高。在实验中,采用 Landsat 和 MODIS 数据为研究对象,使用两个开源数据集和一个本地数据集对该方法进行测试,并与 4 种常用的时空融合方法进行比较。实验结果表明,文中提出的残差网络和双重注意力机制方法能更好地提取图像的有效信息,使用深监督损失函数缓解反向传播时梯度消失的问题,优化了学习过程,使得融合后的效果有了明显的提高。

关键词: 时空融合;遥感图像;双重注意力机制;Landsat;MODIS

中图分类号 TP391

Spatiotemporal Fusion Method for Remote Sensing Images Based on Dual Attention Mechanisms

FAN Xuejing, XUE Xiaorong and DU Yichao

School of Electronic and Information Engineering, Liaoning University of Technology, Jinzhou, Liaoning 121001, China

Abstract Due to the limitations of remote sensing imaging technology conditions, it is difficult to obtain sequences of remote sensing images with both high temporal and high spatial resolution simultaneously. However, spatiotemporal fusion techniques can generate remote sensing images with high temporal and high spatial resolution. In recent years, various spatiotemporal fusion methods have emerged, showing good performance but still falling short in effectively extracting meaningful information in feature extraction. A deep learning model based on dual attention mechanism improvement (ADCSTFN) is proposed to address this issue, which enhances the model's ability to preserve global information and reconstruct detailed scenes. In the experiments, Landsat and MODIS data were used as the research subjects, and the proposed method was tested using two open-source datasets and a local dataset, and compared with four commonly used spatiotemporal fusion methods. Experimental results show that the residual network and dual attention mechanism proposed in this paper better extract effective information from images. The use of a deep supervision loss function mitigates the issue of vanishing gradients during backpropagation, optimizes the learning process, and significantly improves the fusion results.

Keywords Spatiotemporal fusion, Remote sensing images, Dual attention mechanism, Landsat, MODIS

1 引言

随着遥感技术的发展,目前卫星遥感已进入高空间、高时间、高光谱分辨率的时代^[1-2]。然而,不同传感器获取的遥感图像具有不同的特征,如空间分辨率、光谱波段和观测周期,这使得图像之间存在不一致性,限制了它们在时空分析中的应用。遥感图像时空融合旨在将多源遥感数据整合到一个一致的时空框架中,以获得更全面和准确的地表信息。利用高时空的遥感数据对研究目标进行连续的监测,记录其变化的过程,探究其潜在模式并预测其发展趋势的研究越来越多^[3]。

时空融合的目的是提供了一种高效、低成本的方式来融合具有时空互补性的两类遥感图像,生成同时具有高时空分

辨率的融合数据,以解决该问题^[4]。目前国内外基于时空融合的结果用于长时间序列数据的补充、定量反演分析等方面取得了较好的结果^[5]。

近年来基于多源遥感数据的时空融合技术得到快速的发展与应用^[6-7],目前国内外基于已有的数据资源,针对多源遥感时空融合技术做了大量的研究。基于时空融合技术中应用模型的不同,可以将目前所用的方法大体分为 3 类,即基于变换模型的融合、基于重建模型的融合、基于学习模型的融合。

基于变换模型的融合主要是以数据变换模型为基础,对变换后的数据进行信息的筛选和加工,最后再经过反变换,获得未知时刻的高分辨率数据。其中用得较多的是小波变换模型,小波变换是对影像进行小波分解,然后对小波分解后的各

基金项目:辽宁省科技计划项目(2021JH2/10200023);辽宁省教育厅科研重点项目(LJKZ0618)

This work was supported by the Science and Technology Plan Project of Liaoning Province(2021JH2/10200023) and Key Project of Scientific Research of Education Department of Liaoning Province(LJKZ0618).

通信作者:薛笑荣(xr_986@163.com)

层进行融合处理,再通过小波变换来实现影像的融合^[8]。Gu等^[9]在MODIS时序影像和TM影像支撑下,采用小波变换融合方法获取30m空间分辨率的NDVI时间序列数据,并利用该数据有效地区分出各种秋季作物的空间分布。Zhu等提出了一种灵活的时空融合模型^[10](Flexible Spatiotemporal Data Fusion, FSDAF)。

基于重建模型的融合,大多是基于像元分解重构的时空融合模型,应用最广的是由GAO等^[11]提出的时间自适应的反射率融合模型STARFM(Spatial Temporal Analysis Reflection Fusion Model),用于将不同分辨率和时间间隔的遥感影像进行融合。Zhu等基于已有的STARFM模型提出了增强型时空自适应反射融合模型ESTARFM^[12](Enhanced Spatial and Temporal Adaptive Reflectance Fusion Model),该算法考虑像元反射率的时间变化特征,结合光谱解混理论,更好地预测异质景观下的反射率,Busetto等^[13](于2008)在前两者研究的基础上,考虑了周边像元与目标像元间的光谱差异,减少了空间距离近但光谱相差较大的像元。

基于学习模型的融合,主要是建立高、低分辨率影像之间的对应关系,以对预测日期下的高分辨率影像进行预测。Zeyde等^[14]利用了Yang算法的思想。Song等^[15]在Zeyde的方法的基础上进一步提出了通过仅学习一对图像的时空融合模型。为了减少输入数据的使用,Song和Huang之后提出使用一对高分辨率和低分辨率的影像进行字典学习训练一个字典对,参考Dong等提出的基于深度卷积网络的超分辨率重建的方法,Song等建立了一种基于深度卷积网络的时空融合框架^[16](Spatiotemporal Fusion using Deep Convolutional Neural Network, STFDCNN)。与基于稀疏表示的方法相比,此算法能够更有效地从大规模遥感数据中提取和表达信息,并且产生更加精确的预测结果。Tan等提出了一种新的深度卷积时空融合网络^[17](Deep Convolutional Spatiotemporal Fusion Network, DCSTFN),充分使用卷积神经网络从输入影像中提取信息并进行整合,以获取高时空分辨率图像。基于单样本对训练的卷积神经网络时空融合方法^[18](Single-sample-based Convolutional Neural Network for Spatiotemporal Fusion of Remote Sensing Images, SS-CNN),利用光谱信息结合卷积神经网络建立粗图像与细图像之间的超分关系。基于深度学习,提出了一种残差卷积神经网络模型,利用全新的网络架构来融合来自Landsat-8和MODIS图像的NDVI检测结果,从而显著提高融合结果^[19]。

本文提出了引入残差网络模块和双重注意力机制,对数据进行特征提取,再利用制作好的训练集样本对网络结构进行训练,最后使用测试集数据,得到融合后的影像。

本文旨在探讨遥感图像时空融合的方法,以解决遥感数据时空不一致性带来的挑战。首先,介绍了几种经典的时空融合方法,包括基于传统时空融合模型(STARFM, ESTARFM, FSDAF)和基于深度学习的方法(DCSTFN),随后对本文方法(ADCSTFN)进行了详细介绍。最后,将上述提到的已有方法与文中提出的方法进行了对比实验。

实验结果表明,本文方法在不同数据集环境下的性能都有较好的效果,能够更好地提取图像的有效信息,深监督损失函数缓解反向传播时梯度消失的问题,作为约束条件优化了整个学习过程,使得融合后的效果有了明显的提高。

2 研究区域及数据集概况

2.1 研究区域概况

研究区选择的是辽宁省锦州市,位于辽宁西部中心城市、辽宁重要的工业城市和东北地区交通枢纽城市,地理坐标处在东经120°44′—122°36′,北纬40°47′—42°08′之间,是辽宁沿海经济带重要的港口城市。该研究区域是温带季风气候,西北高东南低,地表覆盖类型丰富,包括人工建筑、森林、海洋、湿地、耕地。

2.2 实验数据

Landsat-8是美国地质调查局(USGS)和国家航空航天局(NASA)合作开发和维护的一颗卫星,它携带了一台名为“Operational Land Imager”(OLI)的遥感仪器以及“Thermal Infrared Sensor”(TIRS)热红外传感器。这颗卫星于2013年2月11日发射,是Landsat卫星系列中的第八颗卫星。Landsat-8以前被称为Landsat Data Continuity Mission(LDCM),它的任务是继承和延续先前Landsat卫星的遥感数据记录,提供高质量的地球观测数据。

MODIS(Moderate Resolution Imaging Spectroradiometer)是一种遥感传感器,由美国国家航空航天局(NASA)的地球观测系统科学项目(EOS)开发并在多颗卫星上部署,旨在监测地球的大气、陆地和海洋等方面的变化。MODIS数据集包括了来自这些传感器的遥感数据,为科学家、政府机构和研究人员提供了丰富的地球观测信息。

本文主要采用的是2015年无云的Landsat-8数据和MODIS09A1数据作为最终的实验数据,并对Landsat-8数据影像进行了辐射定标、大气校正、几何校正以及重采样等预处理方法,MODIS数据影像进行重投影与重采样。表1列出了本文实验所需要数据的具体信息。

表1 实验所用数据

Table 1 Data used in experiment

数据类型	行列号	时间		
Landsat-8	120/31	2015-02-06	2015-03-10	2015-03-26
	120/32	2015-05-13	2015-09-26	2015-12-07
	h27v04	2015-01-01	2015-01-25	2015-02-02
	h26v04	2015-02-26	2015-03-10	2015-03-22
MODIS09A1		2015-05-09	2015-06-18	2015-07-20
		2015-08-21	2015-09-22	2015-10-16
		2015-11-01	2015-11-25	2015-12-11

3 遥感时空融合的方法

3.1 STARFM模型

时间自适应的反射率融合模型是一种空间和时间自适应反射率融合模型,用于将不同分辨率和时间间隔的遥感影像进行融合。它主要用于预测和插值时间序列数据,特别是在处理非线性、非平稳的时间序列数据时表现出较高的预测精度。该算法可以将低分辨率高频率(例如MODIS)和高分辨率低频率(例如Landsat)的影像进行融合,结合输入数据,确定移动窗口的大小,查找附近相似的像元,计算附近相似像元的距离加权值和转换系数,以生成高分辨率高频率的影像数据,即:

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

其中, w 为移动窗口, $(x_{w/2}, y_{w/2})$ 为中心像元, W_{ijk} 为权重函数, $M(x_i, y_i, t_0)$ 是低分辨率图像在 t_0 时刻的分辨率, $L(x_i, y_i, t_k)$ 是高分辨率图像在 t_k 时刻的分辨率, $M(x_i, y_i, t_k)$ 是低分辨率图像在 t_k 时刻的分辨率。

3.2 ESTARFM 模型

增强自适应的反射率融合模型是一种时间序列预测模型,主要用于解决具有季节性和趋势性的数据预测问题。它是在 STARFM 模型的基础上改进的。输入两对 Landsat 和 MODIS 数据和预测时期的 MODIS 数据,首先确定移动窗口的大小,查找附近相似的像元,然后根据输入数据计算附近相似像元的时间权重,最后根据以上步骤得到的相似像元和相似像元的时间权重计算中心像元,以生成高分辨率高频率的影像数据。中心像元的表达式为:

$$C(x_{\frac{w}{2}}, y_{\frac{w}{2}}, t_0) = t_1 C_1(x_{\frac{w}{2}}, y_{\frac{w}{2}}, t_1) + t_2 C_2(x_{\frac{w}{2}}, y_{\frac{w}{2}}, t_2) \quad (2)$$

其中, w 为移动窗口, $(x_{w/2}, y_{w/2})$ 为中心像元, t_1 和 t_2 为两个基础时相的时间权重, C_1, C_2 为高分辨率数据。

3.3 FSDAF 模型

灵活的时空数据融合方法,是一种通过捕获包括缓慢和剧烈的土地覆盖类型变化,更准确地预测异质地区的细分辨率图像的方法,只需要很少的输入数据。输入数据包括 t_1 处获取的一对粗分辨率和细分辨率图像以及 t_2 处获取的一幅粗分辨率图像。输出数据是 t_2 时刻预测的细分辨率图像。

首先使用无监督分类器 ISODATA 对输入的细分辨率图像 t_1 进行分类,然后估算粗分辨率图像中各类从 t_1 到 t_2 的时间变化,根据光谱线性混合理论,粗像元的时间变化是粗像元内每个类别时间变化的加权和。

$$\Delta C(x_i, y_i) = \sum_{c=1}^L f_c(x_i, y_i) * \Delta F(c) \quad (3)$$

然后从时序变化中预测细分辨率每个波段的时间变化 F_2^{TP} 和残差 R 。

$$F_2^{TP}(x_{ij}, y_{ij}) = F_1(x_{ij}, y_{ij}) + \Delta F(c) \quad (4)$$

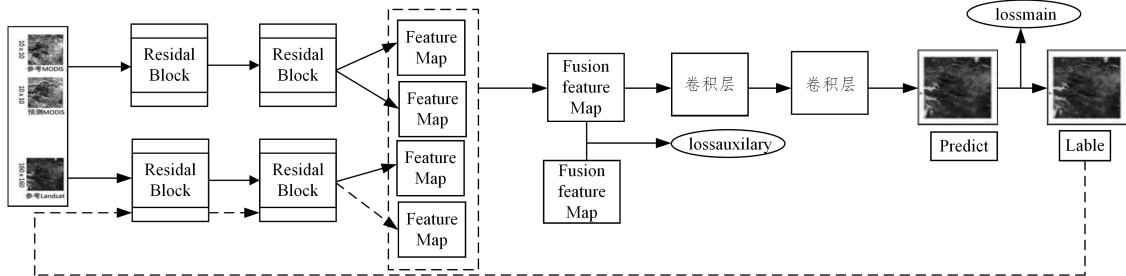


图1 ADCSTFN 模型的总体架构设计

Fig. 1 Overall architectural design of ADCSTFN model

一开始 backbone 加入双重注意力机制,效果并不明显,后来发现是因为网络深度过深,导致梯度回传过小,因此进一步引入了残差连接,缓解了梯度消失的问题。

ResNet 网络结构的基本模块,即残差块(residual block)的目的是通过添加额外的连接来解决深度神经网络训练中的梯度消失和梯度爆炸等问题。通过引入快捷连接,允许某一层的输出直接跳过一个或多个层,连接到后续层的输入。这样做的好处是,即使某些层不做任何有意义的变换,它们仍然可以传递之前层的信息,而不会对梯度产生过多的损失。这可以用一个公式来表示:

$$H(X) = F(X) + X \quad (8)$$

$$R(x_i, y_i) = \Delta C(x_i, y_i) - \frac{1}{m} \left[\sum_{j=1}^m F_2^{TP}(x_{ij}, y_{ij}) - \sum_{j=1}^m F_1(x_{ij}, y_{ij}) \right] \quad (5)$$

利用薄板样条(TPS)插值器从 t_2 粗图像中预测出高分辨率图像;通过空间距离权重对所有相似像素的变化信息求和,以生成高分辨率高频率的影像数据。

$$F_2(x_{ij}, y_{ij}) = F_1(x_{ij}, y_{ij}) + \sum_{c=1}^n \omega_c \Delta F(x_c, y_c) \quad (6)$$

其中, w 为窗口大小, c 为移动窗口中心像元位置。

3.4 DCSTFN 模型

基于深度卷积网络融合高时空遥感图像,提出利用卷积神经网络(CNN)从具有 MODIS 和 Landsat 的遥感图像中提取高时空分辨率图像。DCSTFN 模型由 3 个主要部分组成: MODIS 图像的扩展、Landsat 图像中高频分量的提取以及提取特征的融合。DCSTFN 模型在输入和输出之间建立了一个复杂但直接的非线性映射。

首先共享网络是对 MODIS 进行扩展。这两幅 MODIS 图像都通过子网络提取低频分量,同时利用反卷积 Deconv(d2,3,2)操作将 MODIS 图像扩展到 Landsat 图像的大小。第二部分是从 Landsat 图像中提取高频成分。将卷积层应用于 Landsat 图像斑块,提取细节边缘和纹理等高空间频率信息。池化层用于过滤高频信息。最后对提取的抽象特征进行融合。该方程定义了从参考时间 t_k 到预测时间 t_0 的 MODIS 和 Landsat 图像之间的时间地面覆盖变化,计算式如下:

$$Landsat(t_0) = MODIS(t_0) + Landsat(t_k) - MODIS(t_k) \quad (7)$$

3.5 ADCSTFN 模型

本文算法是基于 ResNet^[20]、DANet^[21] (Dual Attention Network) 和 DCSTFN 算法提出的 ADCSTFN (Attention Deep Convolutional Spatiotemporal Fusion Network) 算法。算法模型如图 1 所示。

本文采用的是 ResNet-18。在 ResNet-18 中,每个卷积层之后都跟有一个批量归一化(Batch Normalization)层,以加速训练并提高模型的稳定性。此外,ResNet-18 还摒弃了 dropout 层,这是一种正则化技术,用于防止过拟合,但在 ResNet 中不是必须的。

ResNet-18 引入了残差连接,允许网络学习残差函数,从而更容易训练深度网络。这有助于解决传统深度网络中的梯度消失问题,使得网络能够更好地收敛。由于残差连接的设计,ResNet-18 等 ResNet 模型能够轻松地训练非常深的网络,而不会导致性能下降。这使得它们在图像处理任务中能够更好地捕捉图像中的特征。相对于更深的 ResNet 模型(如

ResNet-50, ResNet-101 等), ResNet-18 具有相对较小的模型体积, 但仍然能够提供相当不错的性能。这使得 ResNet-18

成为一种在计算资源受限的环境中使用的合适选择。图 2 是 ResNet-18 网络结构示意图。

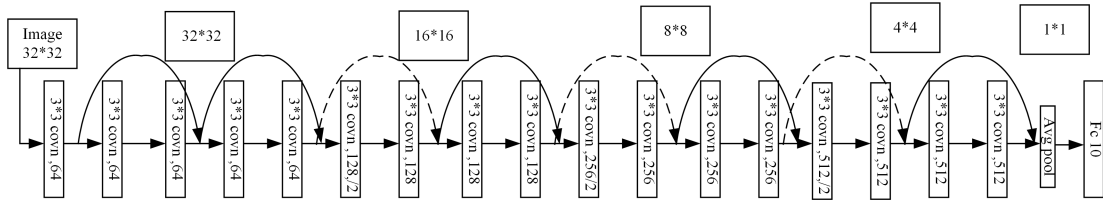


图 2 ResNet-18 网络结构示意图

Fig. 2 ResNet-18 network architecture diagram

双重注意力机制 (Dual Attention Network, DANet) 是一种用于特征空间和通道空间双重注意力权重学习的网络结构。分别对空间维度和通道维度上的语义依赖关系进行建模, 使用位置注意力模块来学习特征的空间相互依赖性, 并设计通道注意力模块来模拟通道相互依赖性, 通过在局部特征上建模丰富的上下文依赖关系, 显著改善了分割结果。对这两个注意力模块的输出进行融合, 进一步增强图像的特征表示。这种机制可以更加有效地捕捉输入数据中的重要特征, 提高模型的性能。传统的注意力机制主要关注输入序列中的元素之间的关系, 例如在自然语言处理 (NLP) 中的 Transformer 模型中, 注意力机制用于计算输入序列中每个位置的权重, 以便更好地捕捉长距离依赖关系。然而, 这些注意力机制往往忽略了输入数据在不同空间维度上的特征。

深度监督损失函数是在深度学习领域中用于指导神经网络训练的一种重要概念。在神经网络的训练过程中, 损失函数扮演着至关重要的角色, 它衡量了模型预测值与真实值之间的差距。通过优化损失函数, 我们可以调整网络的参数, 使得模型能够更准确地学习数据的内在规律。

在传统的监督学习中, 我们通常会使用如均方误差 (MSE) 或交叉熵损失函数来评估模型的性能。然而, 在深度学习中, 由于模型结构的复杂性, 单一的损失函数往往无法满足所有训练需求。因此, 研究者们提出了多种多样的深度监督损失函数, 以适应不同的学习任务和数据特性, 使用了辅助损失, 缓解了反向传播时梯度消失的问题, 使得辅助损失优化了学习过程。

3.6 时空融合定量评价指标

为了评估融合实验的效果, 采用目视对比、定量评价指标分析融合影像与真实影像, 采用了 3 种时空融合最常用的精度评价指标。

均方根误差 (RMSE): 可以比较融合图像和真值图像之间的差异。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2} \quad (9)$$

其中, RMSE 的值越小, 融合图像的质量越好。X 和 Y 分别

为预测期的融合影像和真实影像; i 为影像中第 i 个像素; n 为影像的像素总数。

峰值信噪比 (PSNR): 评价各波段的重建质量, 是信号的最大功率与残差功率的比值。

$$PSNR = 10 \log_{10} \left(\frac{MAX^2}{MSE} \right) \quad (10)$$

其中, PSNR 的值越大, 融合图像的质量越好。max 为信号峰值, 在图像中即为最大像素值。

结构相似性 (SSIM): 是一种衡量两幅图像相似度的指标。

$$SSIM(X, Y) = \frac{(2\mu_X \mu_Y + C_1)(2\sigma_{XY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_X^2 + \sigma_Y^2 + C_2)} \quad (11)$$

其中, SSIM 越高, 融合图像的质量越好。 $\bar{X}\bar{Y}$ 分别为预测图像和真实图像的均值, σ_X 和 σ_Y 分别为预测图像和真实影像的标准差; σ_{XY} 为融合影像和真实影像的协方差, C_1 和 C_2 通常使用一个接近于 0 的常数。

4 结果与分析

本文所使用的数据集是上文提到的 Jinzhou 数据集和开源的 AHB 和 Daxing 数据集进行对比实验。首先对 Jinzhou 数据集进行大气校正、辐射定标、几何校正的处理, 然后对输入图像进行裁剪和切片处理, 再进行分组实验, 使用随机裁剪的方法来训练模型, 每次迭代中从训练区域中随机裁剪出具有相同空间分辨率的训练图像。在整个实验过程中, 训练数据和测试数据是不重复的。

将本文方法与传统方法和深度学习的时空融合方法进行充分比较, 包括 STARFM, ESTARFM, FSDAF, DCSTFN 以及本文提出的 ADCSTFN 模型。图 3—图 5 分别给出了使用不同数据集在不同算法中生成的高空间高时间分辨率的影像。图 3—图 5 中, (a) 显示的是原图, (b) 显示的是 STARFM 方法的结果影像, (c) 显示的是 ESTARFM 方法的结果影像, (d) 显示的是 FSDAF 方法的结果影像, (e) 显示的是深度学习方法的结果影像, (f) 是本文提出的改进方法的结果影像。

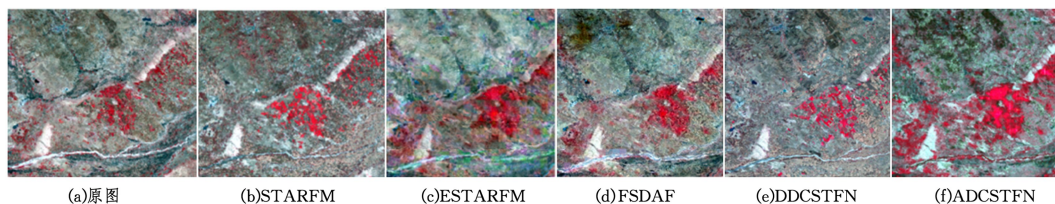


图 3 在 AHB 数据集下与先进方法的对比

Fig. 3 Comparison with advanced methods on AHB dataset

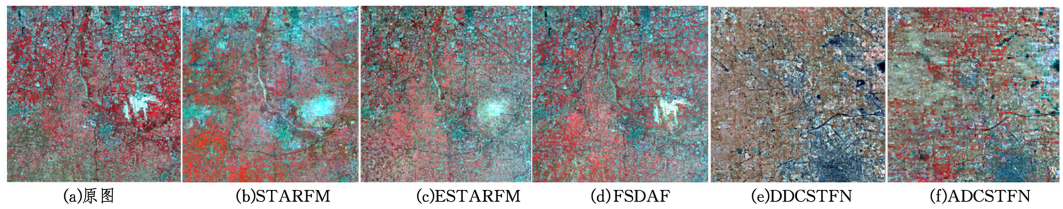


图4 在 Daxing 数据集下与先进方法的对比

Fig. 4 Comparison with advanced methods on Daxing dataset

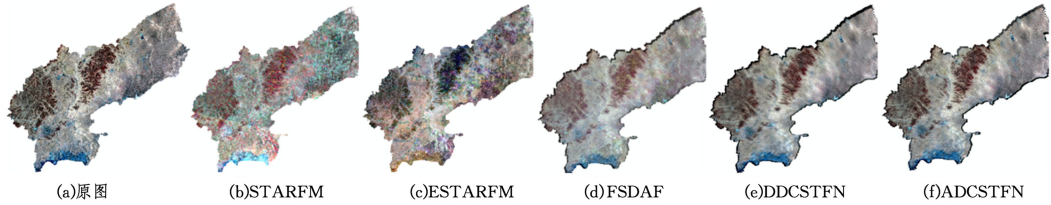


图5 在 Jinzhou 数据集下与先进方法的对比

Fig. 5 Comparison with advanced methods on Jinzhou dataset

从可视化的结果来看,通过对 5 种时空融合模型预测结果进行目视比对。整体上看,5 种融合方法的结果色调与真实影像接近,纹理相对清晰,没有明显的质量问题。本文提出的 ADCSTFN 模型,在这几个模型中效果是最好的,地物信息清晰可见,纹理清晰。从细节上来看,传统的方法中 FSDAF 模型重建影像与真实影像最为接近,影像颜色真实,且各类地物对比明显,STARFM 和 ESTARFM 方法相对有所偏差。FSDAF 模型的效果最好,ESTARFM 模型次之,STARFM 模型的效果最差,在地物的纹理特征等方面相对较差,存在一定的斑块化现象。与传统的模型方法相比,对于基于深度学习的融合方法,在经过多次训练的结果下,融合生成的图像质量也普遍较高。

本文也对不同的数据集结果分别进行了精度评价,以此来进一步分析实验。AHB, Daxing 和 Jinzhou 数据集的详细评价参数分别如表 2—表 4 所列。

表 2 AHB 数据集下的评价指标

Table 2 Evaluation indicators on Daxing dataset

	SSIM	PSNR	RMSE
STARFM	0.687	22.200	22.946
ESTARFM	0.702	24.130	19.422
FSDAF	0.681	22.412	20.444
DCSTFN	0.820	13.368	13.368
ADCSTFN	0.861	49.530	8.737

表 3 Daxing 数据集下的评价指标

Table 3 Evaluation indicators on Daxing dataset

	SSIM	PSNR	RMSE
STARFM	0.805	27.942	10.221
ESTARFM	0.826	29.543	9.032
FSDAF	0.811	28.941	9.144
DCSTFN	0.816	30.166	8.566
ADCSTFN	0.957	53.912	5.522

表 4 Jinzhou 数据集下的评价指标

Table 4 Evaluation indicators on Jinzhou dataset

	SSIM	PSNR	RMSE
STARFM	0.7670	32.7900	958.200
ESTARFM	0.9472	40.1310	423.420
FSDAF	0.9717	43.2151	290.460
DCSTFN	0.9704	42.8912	296.951
ADCSTFN	0.9846	46.4110	198.000

由表 2—表 4 中可以看出,在这两个不同数据集上进行不同方法模型对比,本文方法都达到了很好的效果。SSIM 与其他方法对比是最好的,在 AHB 数据集上是 0.861,在 Daxing 数据集上提高到了 0.957,在 Jinzhou 数据集上提高到了 0.984。PSNR 和 RMSE 这两个评价指标也是本文提出的 ADCSTFN 在不同的数据集上都有较好的效果。总的来看,本文算法无论是从可视化影像来看还是从数据指标来看,都是可行的并且有较好的效果。

结束语 本文通过 5 种时空融合算法,在 3 个数据集上进行实验,生成高空间、高时间分辨率的遥感影像,介绍了图像融合算法和深度学习算法的相关理论知识,其中主要介绍了 3 种 (STARFM, ESTARFM, FSDAF) 经典的算法和深度学习 (DCSTFN) 的算法,并对这几种算法进行了复现,同时与本文算法进行了对比。将所使用的数据处理方法做详细介绍并进行大量的实验,而后在广泛使用的数据集下进行对比实验,本文算法无论是在实验结果的视觉层面和数据上均取得了好的效果,说明本文算法是可行的。由于本文的研究是在 DCSTFN 模型的基础上提出的改进,因此本文的实验结果也主要是针对该方法进行对比。

遥感图像时空融合技术自 2006 年被提出以来,已经发展十多年。早期的研究主要集中在基于变换模型、基于重建模型和基于学习模型的融合方法。随着深度学习技术的发展,基于深度学习的融合方法得到了广泛应用。然而,遥感图像时空融合领域仍然面临一些挑战,例如缺乏大规模的数据集,以及如何在保持光谱信息的同时提高图像的空间分辨率。

未来的研究可以从以下几个方面进行:1) 开发新的数据集,以提供更多的信息和测试算法的性能;2) 研究新的融合方法,以提高融合效果和算法的稳定性;3) 探索深度学习技术在遥感图像时空融合中的应用,以利用其强大的特征提取和融合能力。总的来说,遥感图像时空融合技术在过去的十多年里取得了显著的进展,但仍然有一些问题需要解决。随着科技的进步和数据资源的丰富,相信在不久的将来,遥感图像时空融合技术将会有更大的发展和应用。

参考文献

[1] GONG P. Some Frontier Issues in Remote Sensing Science and

- Technology[J]. *Journal of Remote Sensing*, 2009, 13(1):13-23.
- [2] TONG Q, ZHANG B, ZHANG L. Frontier Advances in Hyperspectral Remote Sensing in China[J]. *Journal of Remote Sensing*, 2016, 20(5):689-707.
- [3] WANG S. Research on Spatiotemporal Fusion Algorithms Based on Multi-source High-Resolution Satellite Remote Sensing Images [D]. Beihua Aerospace Industry College, 2023.
- [4] LI C, SONG H, ZHANG K, et al. Conditional Generative Adversarial Networks for Spatiotemporal Fusion of Remote Sensing Images[J]. *Journal of Image and Graphics*, 2021, 26(3):714-726
- [5] SUN Y. Research on Spatiotemporal Fusion Algorithms for Remote Sensing Data-A Case Study with Landsat and MODIS Data [C]//China University of Mining and Technology, 2019.
- [6] HUANG B, ZHAO Y. Current Status and Prospects of Spatiotemporal Fusion Research on Multi-source Satellite Remote Sensing Images[J]. *Journal of Geomatics*, 2017, 46(10):1492-1499.
- [7] ZHANG H K, HUANG B, ZHANG M, et al. A generalization of spatial and temporal fusion methods for remotely sensed surface parameters[J]. *International Journal of Remote Sensing*, 2015, 36(17):4411-4445.
- [8] SHEVYRNOGOV A, TREFOIS P, VYSOTSKAYA G. Multi-satellite Data Merge to Combine NOAA AVHRR Efficiency with Landsat-6 MSS Spatial Resolution to Study Vegetation Dynamics[J]. *Advances in Space Research*, 2000, 26(7):1131-1133.
- [9] GU X H, HAN L J, WANG J A. Remote Sensing Estimation of Corn Area by Wavelet Fusion of Medium and Low Resolution Images[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2012, 28(3):203-209.
- [10] 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.
- [11] GAO F, MASEK J, SCHWALLER M, et al. On the blending of the Landsat and MODIS surface reflectance; predicting daily Landsat surface reflectance[C]//IEEE Transactions on Geoscience and Remote Sensing. 2006:2207-2218.
- [12] ZHU X, CHEN J, GAO F. An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions[J]. *Remote Sensing of Environment*, 2010, 114(11):2610-2623.
- [13] BUSETTO L, MERONI M, COLOMBO R. Combining medium and coarsespatial resolution satellite data to improve the estimation of sub-pixel NDVI time series[J]. *Remote Sensing of Environment*, 2008, 112(1):118-131.
- [14] ZEYDE R, ELAD M, PROTTER M. On single image scale-up using sparse-representations[C]//Proceedings of the 7th International Conference on Curves and Surfaces. Avignon, France: Springer, 2012:711-730.
- [15] SONG 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.
- [16] SONG H, LIU Q, WANG G. 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.
- [17] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition(CVPR). 2015:770-778.
- [18] LI Y, LI J, HE L. Single-sample-based Convolutional Neural Network for Spatiotemporal Fusion of Remote Sensing Images [J]. *Journal of Remote Sensing*, 2022, 26(8):1614-1623.
- [19] SUN Z, OUYANG X, LI H, et al. Spatiotemporal Data Fusion Model of NDVI Based on Deep Learning[J]. *Journal of Resources and Ecology*, 2024, 15(1):214-226.
- [20] TAN Z Y, YUE P, DI L P, et al. Deriving high spatiotemporal remote sensing images using deep convolutional network[J]. *Remote Sensing*, 2018, 10(7):1066.
- [21] FU J, LIU J, TIAN H, et al. Dual Attention Network for Scene Segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). 2018:3141-3149.



FAN Xuejing, born in 1996, postgraduate. Her main research interests include image processing and pattern recognition.



XUE Xiaorong, born in 1974, Ph.D., professor, is a member of CCF(No. 14043M). His main research interests include image processing and pattern recognition, etc.