

基于卷积神经网络的时空融合算法

陈翔宇

成都信息工程大学应用数学学院, 四川 成都

收稿日期: 2022年3月14日; 录用日期: 2022年4月8日; 发布日期: 2022年4月19日

摘要

对于遥感问题, 时空融合算法旨在解决卫星传感器无法同时获取高空间分辨率、高时间分辨率的遥感图像的缺陷, 大多数情况下, 单个卫星传感器的遥感图像无法满足需求, 因此衍生出许多融合高空间分辨率、低时间分辨率遥感图像和低空间分辨率、高时间分辨率的遥感图像的时空融合方法, 其中效果最为显著的便是基于深度学习的时空融合方法, 本文在此基础上, 利用卷积神经网络构建一种新的时空融合算法, 并得到了更小的误差, 应用效果比较好, 值得应用及推广。

关键词

遥感, 时空融合, 空间分辨率, 时间分辨率, 神经网络

Spatio-Temporal Fusion Algorithm Based on Convolutional Neural Network

Xiangyu Chen

Department of Applied Mathematics, Chengdu University of Information Technology, Chengdu Sichuan

Received: Mar. 14th, 2022; accepted: Apr. 8th, 2022; published: Apr. 19th, 2022

Abstract

For remote sensing, the spatio-temporal fusion algorithm aims to solve the defect that satellite sensors cannot obtain remote sensing images with high spatial resolution and high temporal resolution at the same time. In most cases, the remote sensing images of a single satellite sensor cannot meet the demand, therefore, many methods that merge high spatial resolution images with low temporal resolution and high temporal resolution images with low spatial resolution were created, the most significant effect is the spatiotemporal fusion method based on deep learning. On this basis, this paper constructs a new spatiotemporal fusion algorithm by using convolutional neural

network, and obtains less error. The application effect is relatively good, which is worthy of application and promotion.

Keywords

Remote Sensing, Spatio-Temporal Fusion, Spatial Resolution, Temporal Resolution, Neural Network

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

高空间、时间分辨率的卫星遥感图像对研究地球地表问题具有重大意义，比如识别地表土地覆盖类型[1]、监测植被季节性变化[2]、估算作物产量[3]等。而人们可以获得的免费卫星遥感图像的途径也越来越多，因此对于时空融合模型的研究越来越流行，以 Modis 为代表高时间分辨率、低空间分辨率的卫星图像，以及以 LandSat 为代表的低时间分辨率、高空间分辨率的卫星图像成为人们的主要研究对象。

在过去的十几年，随着时空融合算法的越来越多元化，时空融合算法可大致分为五类[4]：基于分解的方法、基于权函数的方法、基于贝叶斯的方法、基于学习的方法以及混合的方法，且发展方向更倾向于学习的方法，而学习的方法又分为传统的学习方法和深度学习的方法，深度学习方法取得了更高的精度，本文选用的数据集为天津数据集[5]，地理位置为北纬 39.8625 度，东经 117.8591 度，用已有数据预测缺失数据，并与真实数据对比，并将误差与已有的时空融合算法进行对比，最终结果显示本文提出的方法拥有更小的误差。

2. 数据集及算法结构

2.1. 数据集

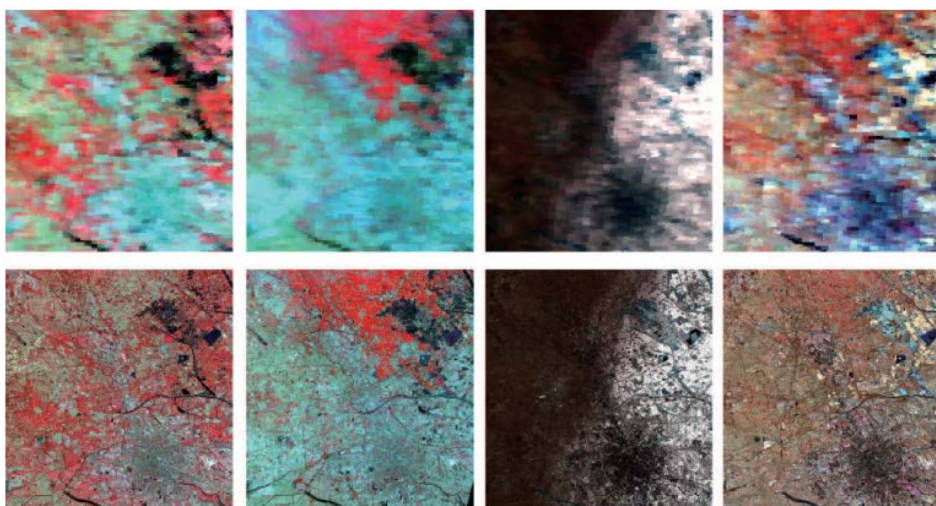


Figure 1. Sample of Tianjin dataset

图 1. 天津数据集样本

天津是中国的一个直辖市，天津数据集[5]包括从2013年9月至2019年9月的27个LandSat-Modis图像对[6]，同一时间点分别有一张LandSat图像和一张Modis图像，其中LandSat图像为高空间分辨率、低时间分辨率图像，即图像分辨率高，但相邻两个时间点间隔长，而Modis图像为高时间分辨率、低空间分辨率，即虽然分辨率低，但是图像的时间连续性很强，根据实验目的，仅选择与LandSat图像对应时间节点的Modis图像，数据集的样本图像如图1所示。由于数据量较大，每张图片包含6个波段，本次实验仅用每张图片的第一个波段。

2.2. 算法结构

2.2.1. 模型结构

模型由两部分构成，第一部分为超分辨率模型，第二部分为融合模型，均基于卷积神经网络，超分辨率模型输入为两个不同时间点的低分辨率图像的残差图，输出为两个不同时间点的高分辨率图像的残差图，例如：设第一、第二个时间点的低分辨率图像为 M_1 、 M_2 ，第一、第二个时间点的高分辨率图像为 L_1 、 L_2 ，则输入为 M_1 、 M_2 的残差 ΔM_{21} ，输出为 L_1 、 L_2 的残差 ΔL ，公式如下：

$$\Delta M = M_2 - M_1, \quad \Delta L = L_2 - L_1$$

将最终的输出结果，即 $\Delta \hat{L}$ 叠加在 L_1 上得到由 L_1 预测的 \hat{L}_{21} ，显而易见， \hat{L}_2 的结果与 M_1 、 M_2 及 L_1 有很大关系，为使结果更加精确，同理，我们也可以得到由 L_3 预测的第二个时间点的图像 \hat{L}_{23} ，这便是超分辨率模型的目标。

融合模型的目的便是将由超分辨率模型得到结果结合起来，得到一张更加精确的第二个时间节点的图，同样适用卷积神经网络，输入为第一、第二个时间点的低分辨率图像为 M_1 、 M_2 的残差 ΔM_{21} 、第二、第三个时间点的低分辨率图像为 M_2 、 M_3 的残差 ΔM_{32} 以及真实的 L_1 、 L_3 ，标签为 L_2 ，模型的输出为一个权重向量矩阵 α ，将权重向量矩阵 α 与 \hat{L}_{21} 和 \hat{L}_{23} 的组合为模型的最终输出 \hat{L}_2 ， \hat{L}_2 的公式如下：

$$\hat{L}_2 = \alpha * \hat{L}_{21} + (1 - \alpha) * \hat{L}_{23}$$

整个算法的结构如图2所示。

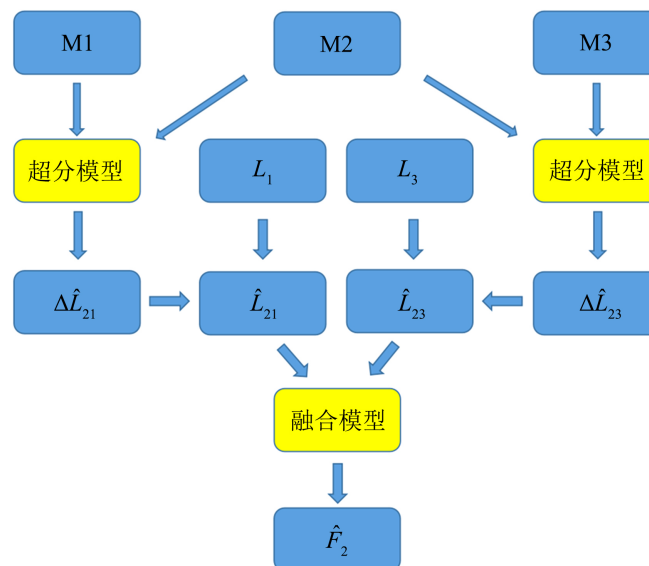


Figure 2. Algorithm structure
图2. 算法结构

最终的预测结果即是第二个时间点的高分辨率图像，模型的目标如图 3 所示，超分辨率模型和融合模型的具体配置将在第三节中提到。

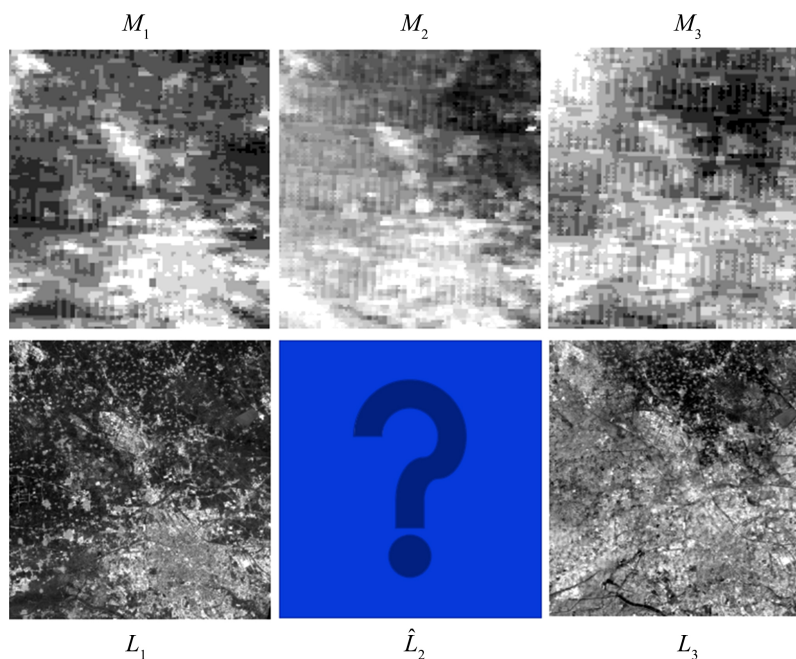


Figure 3. Target of the algorithm

图 3. 算法目标

2.2.2. 卷积神经网络

在图像处理的问题中，可以把输入的图片看作由像素点组成矩阵，由卷积核对矩阵进行卷积处理以达到特征提取的目的，与全连接不同，卷积神经网络对输入矩阵进行局部连接处理，一般认为，图像的空间相关性与相邻的像素点关系较为密切，而全连接则是对全局进行卷积处理，就本文而言，用全连接处理存在两大问题，首先是参数量，若使用全连接，则参数量太过巨大，比如输入图片大小为 1000×1000 ，假设隐含层数目与输入层一致，则会产生 10^6 个参数，这对于训练时间及机器配置要求都是极高的，而另一个问题则是根据生物学的视觉特征，人们对于外界事物信息的提取往往是由局部到全局的，就同一张图片而言，距离过远的两个像素点往往没有直接的关联，在很多情况下使用全局视野的模式下往往不能有效的提取信息，甚至会对最终的判别带来干扰，使用局部视野可以很好的解决这些问题，同样是 1000×1000 的图片，假如使用 10×10 的卷积核，隐含层与输入层数一致的情况下，用于卷积核的权值是共享的，因此仅产生了 100 个参数，大大节省了训练的时间及成本，而且局部像素具有较强的关联性，可以使得最终结果更加精确，本文也是采用这种方法。

一般情况下，卷积层的深度越深越好，一方面，图像经过一个 5×5 的卷积核与经过两个 3×3 的卷积核能达到一样的效果，甚至后者的表现力会更强，而前者产生了 25 个参数，后者仅产生了 18 个参数，以较小的算力达到同等甚至更好的效果，另一方面，加深卷积层数可以将原图所包含的信息分层次地向下一层传递，使得学习效果效率更高，但卷积的层数也不是可以无限加深的，随着层数的加深，可能出现梯度消失的情况，在本问题中，经过试验，选用四层卷积得到的结果更好。

此外，卷积神经网络的另一大特点是可以向训练过程引入非线性变换，通过激活函数的引入，简化了模型结构，不但加快了训练速度，还降低了模型的计算复杂度。卷积层公式如下：

$$Z_i = \phi(W_i * X_i + \delta_i)$$

其中，* 表示卷积运算， W_i 和 δ_i 分别表示权重和模型偏差， $\phi(\cdot)$ 表示激活函数。

3. 算法配置

本算法包含两个模型，即超分辨率模型和融合模型，均使用四层卷积，两个模型结构基本相同，不同的是前者输入为单层，后者输入为四层的张量，虽然结构相似，但由于两者的参数是不共享的，因此分为两个模型。

3.1. 相关参数的设置

在超分辨率模型中，初始卷积核选择 Xavier 初始化方法，四层卷积前三层为 leakyrelu 激活函数，由于输入为残差图，因此输入矩阵存在负值，用 leakyrelu 激活函数效果更好，第四层无激活函数，如表 1 所示，优化算法使用 Adam。损失函数为：

$$f_1(\theta) = \left\| \hat{G}(X; \theta) - \Delta L \right\|_F^2$$

其中 $\hat{G}(X; \theta)$ 为超分辨率模型的输出结果， ΔL 为相邻两个时间点的真实高分辨率图像的残差。

在融合模型中，初始卷积核选择 Xavier 初始化方法，四层卷积前三层为 leakyrelu 激活函数，由于输入包含残差图，因此输入矩阵存在负值，用 leakyrelu 激活函数效果更好，第四层为 sigmoid 函数使输出结果的像素值在 0~1 的范围，如表 2 所示，优化算法使用 Adam。损失函数为：

$$f_2(\theta) = \left\| \hat{L}(X; \theta) - L \right\|_F^2$$

其中 $\hat{L}(X; \theta)$ 为融合模型的输出结果与 \hat{L}_{21} 和 \hat{L}_{23} 的组合， L 为真实的高分辨率图像。

这种配置在与其他配置对比中取得了更好的结果。

Table 1. Configuration of super-resolution model

表 1. 超分辨率模型配置

Layer	Filter size	Stride	Activation function
Conv1	7*7*128	(1, 1)	leakyrelu
Conv2	5*5*128	(1, 1)	leakyrelu
Conv3	3*3*128	(1, 1)	leakyrelu
Conv4	3*3*1	(1, 1)	

Table 2. Configuration of fusion model

表 2. 融合模型配置

Layer	Filter size	Stride	Activation function
Conv1	7*7*128	(1, 1)	leakyrelu
Conv2	5*5*128	(1, 1)	leakyrelu
Conv3	3*3*128	(1, 1)	leakyrelu
Conv4	3*3*1	(1, 1)	sigmoid

3.2. 模型评估

对模型的评估选择均方根误差 RMSE 和相关系数 CC 两种指标，均方根误差 RMSE 对模型的预测效果做定量的统计值，有关计算公式如下：

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

其中 n 表示预测的样本量， y_i 表示标签值， \hat{y}_i 表示预测值。MSE 或 RMSE 越小，说明模型精度越高。相关系数 CC 表示预测图像与真实图像的相关性，计算公式如下：

$$\text{CC} = \frac{E(X,Y) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}}$$

CC 最大值为 1，CC 值越接近 1，说明输出与标签的相关性越高。超分辨率模型在不同配置下的效果如表 3 所示：

Table 3. Super-resolution models with different configurations

表 3. 超分辨率模型的不同配置

权值初始化方式	前三层激活函数	第四层激活函数	MSE	CC
ones	leakyrelu	无	0.00245	0.1155
0~1 均匀分布	leakyrelu	无	0.00197	0.5809
标准正态分布	relu	无	0.00115	0.6355
标准正态分布	leakyrelu	无	0.00149	0.1378
Xavier	relu	无	0.00118	0.6012
Xavier	leakyrelu	无	0.00114	0.7971

融合模型在不同配置下的效果如表 4 所示：

Table 4. Fusion models with different configurations

表 4. 融合模型的不同配置

权值初始化方式	前三层激活函数	第四层激活函数	MSE	CC
ones	leakyrelu	sigmoid	0.00053	0.5356
0~1 均匀分布	leakyrelu	sigmoid	0.00068	0.7259
标准正态分布	relu	sigmoid	0.00042	0.8267
标准正态分布	leakyrelu	sigmoid	0.00051	0.7469
Xavier	relu	sigmoid	0.00041	0.8324
Xavier	leakyrelu	sigmoid	0.00038	0.8948

4. 算法结果及不同算法比较

BiasSTF [6]是一种基于卫星传感器偏置的时空融合算法，并在与以往经典的时空融合算法如ESTARFM、FSDAF、STFDCNN等的对比中获得了较好的结果，以往的时空融合算法在融合步骤多是以低空间分辨率的遥感图形的残差 ΔM_{21} 、 ΔM_{32} 作为融合决策，对预测图像 \hat{L}_{21} 和 \hat{L}_{23} 加以权重，此时的权重相当于对 \hat{L}_{21} 和 \hat{L}_{23} 进行线性变换，比如若 ΔM_{21} 的值较大，则为 \hat{L}_{21} 赋予一个较小的权重，区别在于融合决策的不同，但总的来说这类方法缺少弹性，而本文的融合阶段使用卷积神经网络学习出一个权重矩阵 α ，为权重矩阵引入了非线性(激活函数的使用)，模型潜力更大，这是本文的一大创新。以本文的数据集为基准，天津数据集的第22、23、24天有较大的地表变化，以这三天作为测试集，本文提出的方法最终结果MSE为0.00038，相关系数CC为0.8948，以BiasSTF算法得出的结果MSE为0.00055，相关系数CC为0.8582，本文的两个评价指标均优于BiasSTF算法，最终的输出结果如图4所示：

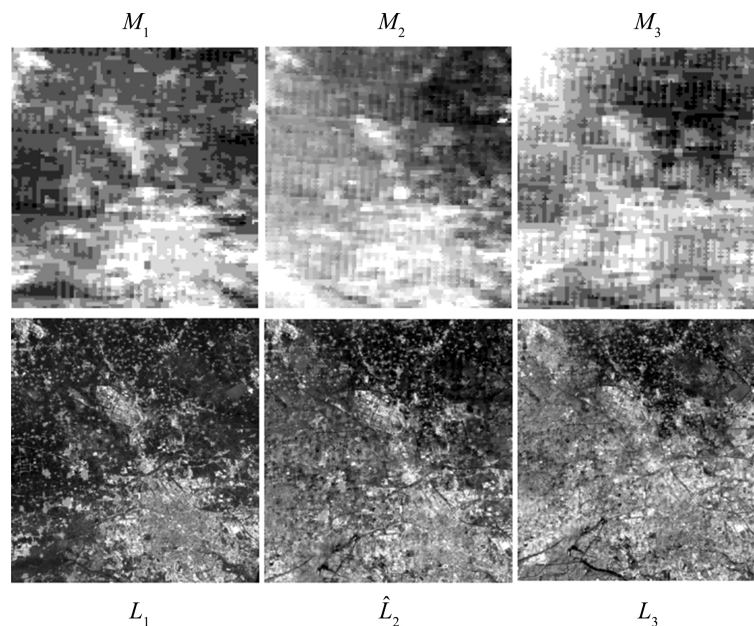


Figure 4. Final result
图4. 最终结果

5. 结束语

本文以卷积神经网络为基础，区别于其他的基于卷积神经网络的时空融合方法，本文在融合阶段也使用了卷积神经网络来预测权重矩阵，为权重的生成引入了非线性，对模型配置进行了优化，并与近年的时空融合算法在天津数据集上进行对比，经试验证明，使用Xavier的权值初始化方法，leakyrelu激活函数能取得较优结果，并在与BiasSTF时空融合算法的对比中取得了更好的结果，相较于以往的时空融合算法具有更好的RMSE以及CC，该算法可应用于各种对地表监测的场景。

参考文献

- [1] Johnson, M.D., Hsieh, W.W., Cannon, A.J., Davidson, A. and Bédard, F. (2016) Crop Yield Forecasting on the Canadian Prairies by Remotely Sensed Vegetation Indices and Machine Learning Methods. *Agricultural & Forest Meteorology*, **218-219**, 74-84. <https://doi.org/10.1016/j.agrformet.2015.11.003>
- [2] Eck, M.A., Murray, A.R., Ward, A.R., et al. (2020) Influence of Growing Season Temperature and Precipitation Ano-

-
- malies on Crop Yield in the Southeastern United States. *Agricultural and Forest Meteorology*, **291**, 108053. <https://doi.org/10.1016/j.agrformet.2020.108053>
- [3] Li, X., Zhou, Y., Asrar, G.R., *et al.* (2017) Response of Vegetation Phenology to Urbanization in the Conterminous United States. *Global Change Biology*, **23**, 2818-2830. <https://doi.org/10.1111/gcb.13562>
- [4] Zhu, X., Cai, F., Tian, J., *et al.* (2018) Spatiotemporal Fusion of Multisource Remote Sensing Data: Literature Survey, Taxonomy, Principles, Applications, and Future Directions. *Remote Sensing*, **10**, 527. <https://doi.org/10.3390/rs10040527>
- [5] Li, J., Li, Y., He, L., *et al.* (2020) Spatio-Temporal Fusion for Remote Sensing Data: An Overview and New Benchmark. *Science China. Information Sciences*, **63**, Article ID: 140301. <https://doi.org/10.1007/s11432-019-2785-y>
- [6] Li, J., Li, Y., He, L., *et al.* (2020) A New Sensor Bias-Driven Spatio-Temporal Fusion Model Based on Convolutional Neural Networks. *Science China. Information Sciences*, **63**, 16. <https://doi.org/10.1007/s11432-019-2805-y>