MosicFormer: A Novel approach for Remote Sensing Spatiotemporal Data Fusion

Instruction

湖泊是水圈的重要组成部分，是生物圈、大气圈和岩石圈在物质循环和能量转移方面相互作用的枢纽，湖泊面积的变化反映了全球气候和环境的变化 (Yamazaki et al., 2015)。然而，由于气候变化和人类活动，湖泊发生了巨大变化，有研究显示湖泊面积和湖水储量正在下降(Yao et al., 2023)。水体动态的监测变得尤为重要。随着遥感(RS)和地理信息技术(GIS、GPS)的发展和应用，可以采用一些有效的技术手段对大空间区域湖泊水体的动态变化进行监测 (Wang et al., 2023)。

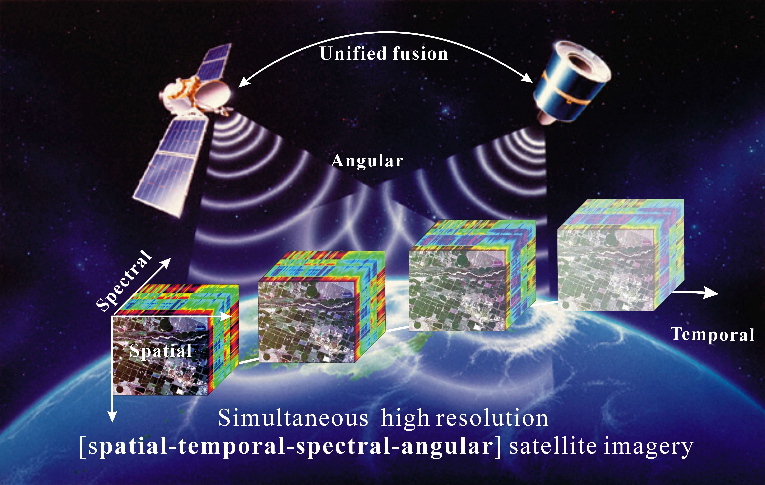
快速变化的内陆水域对传感器的空间和时间分辨率都有严格的要求。Some High-resolution optical satellite time series data, with both fine spatial and temporal detail, are essential for accurate modeling and monitoring of dynamics of surface water bodies. However, a key limitation arises from the inherent tradeoff between spatial resolution and temporal revisit rates, which restricts the satellites' capacity to capture the Earth's surface in high spatiotemporal resolution (Feng et al., 2006). For example, the Moderate Resolution Imaging Spectroradiometer (MODIS), a widely used sensor, provides images with resolutions ranging from 250m to 1000m. While this resolution may not be sufficient for accurately mapping water bodies, it offers daily image coverage. On the other hand, Landsat, which has the longest continuous space-based observation data since the 1970s, provides 30m spatial resolution images(Roy et al., 2014). However, its 16-day revisit cycle and frequent cloud cover limit its utility for capturing rapid changes in surface water change(Sun et al., 2024). To address this issue, previous research has proposed spatiotemporal fusion (STF) methods that improve spatiotemporal resolution by merging high-spatial-resolution images, which are temporally sparse, with low-spatial-resolution images that are captured more frequently, to obtain more accurate information on water body dynamics.

Generally, STF methods are pursued through four main approaches: the spatiotemporal weighting function methods, as represented by the spatial and temporal adaptive reflectance fusion model (STARFM) (Feng et al., 2006), wavelet transformation (Acerbi-Junior et al., 2006; Malenovský et al., 2007), unmixing-based data fusion (Gevaert and García-Haro, 2015), sparse representation (Huang and Song, 2012; Wei et al., 2015) and machine learning methods (Moosavi et al., 2015). 各种STF算法已在spatiotemporal data augmentation of water bodies有了广泛的应用。具体来说，STF用于生成类似 Landsat 的多光谱图像，然后使用无监督、监督或基于对象的分类算法对其进行处理以得出地表水图 (Chen et al., 2018; Dao et al., 2019; Heimhuber et al., 2018; Tan et al., 2019; Zhang et al., 2014)。Among the various data fusion techniques, STARFM and its derivatives (ESTARFM) are likely the most widely used algorithms for generating synthetic surface reflectance with both high spatial and temporal resolutions, owing to their reliable prediction performance (Emelyanova et al., 2013). STARFM identifies similar neighborhoods for the target pixel and makes predictions by weighting these neighborhoods based on spatial, spectral, and temporal proximity, while ignoring the mixture and changes in land cover types in the coarse resolution pixels. However, STARFM faces several issues: first, it struggles to address abnormal cases involving land-cover type changes or disturbance events not captured in a single Landsat image; second, it is less effective in making predictions in heterogeneous landscapes. Subsequently, many studies have suggested improvements to the STARFM algorithm, and several modified models of STARFM have been proposed. Hilker et al. (2009) introduced the Spatial and Temporal Adaptive Algorithm for Reflectance Change Mapping (STAARCH), which identifies temporal variations using a dense set of MODIS data. ESTARFM, which was proposed by Zhu et al. (2010), introduced several improvements over the original STARFM algorithm, with the most notable being the use of a conversion coefficient to enhance prediction accuracy in heterogeneous landscapes. Wang et al. (2023)使用ESTARFM模型融合Landsat与MODIS影像后，提取湖泊水体，探究近20年来湖泊水位变化关系。Dao et al. (2019)同样使用ESTARFM生成无云Landsat/MODIS 合成数据，用于划定洪水期间的淹没区域。Li et al. (2021) 提出了STSWM方法，除了MODIS-Landsat影像对之外，还使用surface water occurrence (SWO) 和 DEM提供水面以上和水面以下的地形信息，生成了30m空间分辨率的地表水图，可检测到由于云层覆盖和数据缺乏地区的地表水变化。

随着深度学习在近年的流行，基于学习的时空融合方法也随之出现。Song et al. (2018)提出了一种基于卷积神经网络（CNN）的时空融合模型，通过学习MODIS和下采样Landsat影像之间的端到端映射来自动提取有效的影像特征。然而这种方法仅使用了3个隐藏层，难以准确模拟MODIS和Landsat影像之间复杂的非线性对应关系。为了改进这种浅层模型，Zheng et al. (2019)提出了新的基于超深卷积神经网络（VDCN）的方法，在低空间分辨率的MODIS和Landsat数据之间训练了一个非线性映射VDCN。最近，Filali Boubrahimi et al. (2024)提出了一种新的基于机器学习的方法（Hydro-GAN）将低分辨率 MODIS 数据映射到高分辨率 Landsat-8 图像时提高边界精度，提高了生成水体多边形的准确性。

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Approach | Type | Target resolution | Spatial interpolation | Temporal gap filling | Radiometric correction | Input data (revisit frequency) | Strengths | Limitations |
| STARFM | Weight function–based method | 30m; daily | N/A | √ | √ | MODIS (MOD09GHK; daily), Landsat (Landsat-7ETM+; 16 days) | Predicts robust performance; capable of generating synthetic surface reflectance with high spatial and temporal resolution. | The assumptions of STARFM are violated in complex heterogeneous regions, leading to low accuracy in land-cover change areas and complex heterogeneous regions. |
| Semi-physical model | Semi-physical fusion approach | 30m; daily | √ | N/A | √ | MODIS (MODIS BRDF/Albedo; 16 days), Landsat (Landsat ETM+ L1G; 16 days) | Easy to implement without scene-based adjustments, applicable to any high-resolution satellite data, and usable on a global scale | Requires precise data registration, solar geometry, cloud detection, and ideally atmospheric correction. |
| STAARCH | Hybrid method (weight function-based and unmixing method) | 30m; 8 days | N/A | √ | √ | MODIS (MOD09/MYD09; 8 days), Landsat (Landsat ETM; 16 days) | Improved STARFM, enhancing the level of detail in the generated synthetic Landsat images | Does not specifically investigate the impact of cloud shadows on surface reflectance; does not allow for the detection or monitoring of individual disturbance events. |
| ESTARFM | Weight function–based method | 30m; daily | √ | N/A | N/A | MODIS (MOD09GQ; daily), Landsat (Landsat 8 OLI C1 Level 2; 16 days) | Improves the accuracy of predicting fine-resolution reflectance, especially for heterogeneous landscapes, while preserving spatial details | Needs manual selection of cloud-free image pairs, limiting its use for dense time-series generation. |
| STRUM | Hybrid method (weight function-based and unmixing method) | 30m; daily | √ | N/A | √ | MODIS (MODIS MCD43A4 BRDF; 8 days), Landsat (Landsat 8 OLI; 8 days) | Suitable for data fusion applications requiring Landsat-like surface reflectance, especially when high-resolution images are limited | Lacks comprehensive comparison with other models. |
| FSDAF | Hybrid method (weight function-based and unmixing method) | 30m | √ | N/A | √ | MODIS (MOD09GA Collection 5; daily), Landsat (Landsat 7 ETM +; 16 days) | Requires minimal input data (only one high-resolution image); suitable for heterogeneous landscapes; capable of predicting both gradual changes and land cover type changes. | Fails to capture subtle land cover changes. |
| RASTFM | Weight function–based method | 30m | √ | √ | √ | MODIS (MOD09; 8 days), Landsat (Landsat-7; 16 days) | Capable of generating accurate high-resolution imagery; efficiently operates automatically; applicable to other satellite data, such as Sentinel-2A/B MSI and Landsat-8 OLI | Cannot predict subtle changes; inaccurate for new features absent in prior images. |
| VDCNSTF | Deep learning–based method | 30m | √ | √ | N/A | MODIS (MOD09GA Collection 5; daily), Landsat (Landsat-5 TM; 16 days) | Achieves superior performance compared to other learning-based methods | Significant room for improvement in spectral accuracy and spatial detail restoration. |
| VIPSTF | Hybrid method (weight function-based and unmixing method) | 30m | √ | √ | N/A | MODIS (MOD09GA Collection 5; daily), Landsat (Landsat 7 ETM+; 16 days) | Enhances spatiotemporal fusion performance; applicable to both heterogeneous sites and sites with temporal land cover changes; low computational cost | Combining with other spatiotemporal fusion methods is complex and requires further validation and development. |
| STAIR | Weight function–based method | 30m | √ | √ | N/A | MODIS (MCD43A4; daily), Landsat (Landsat 7 and 8 Level 2; 16 days) | Effectively fills missing pixel values in input images caused by clouds or sensor damage | Computational efficiency needs improvement. |
| STSWM | Hybrid method (weight function-based and unmixing method) | 30m; 8 days | N/A | N/A | √ | MODIS (MOD09A1; 8 days), Landsat (Landsat 7 and 8 Level 2; 16 days) | Suitable for spatiotemporal surface water mapping in spatially heterogeneous landscapes and for predicting abrupt surface water changes, capable of predicting water bodies in cloud- and shadow-contaminated areas. | The higher the percentage of cloud and shadow areas in MODIS scenes, the lower the accuracy of STSWM maps. |

因为用的这篇文章的数据集，所以大体写法参考这个文章：10.1007/s11432-019-2785-y



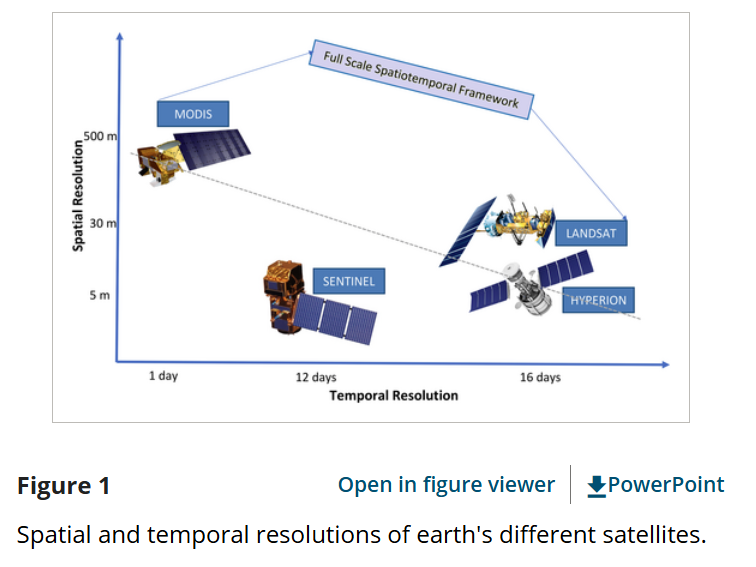


图1. Data Fusion概念图

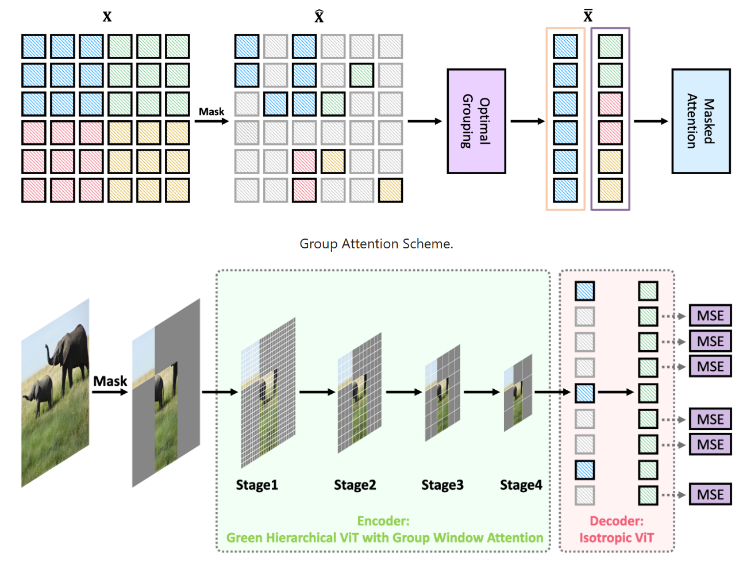


图2. Mosic概念示意

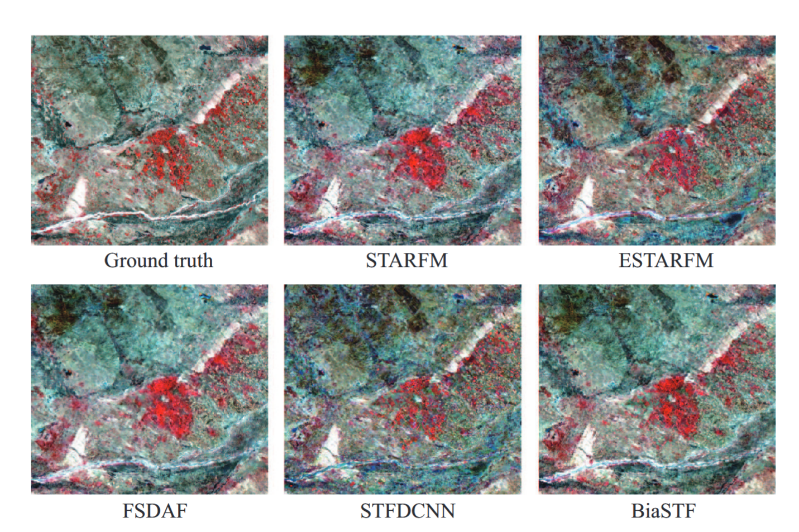


图3. 测试集的结果

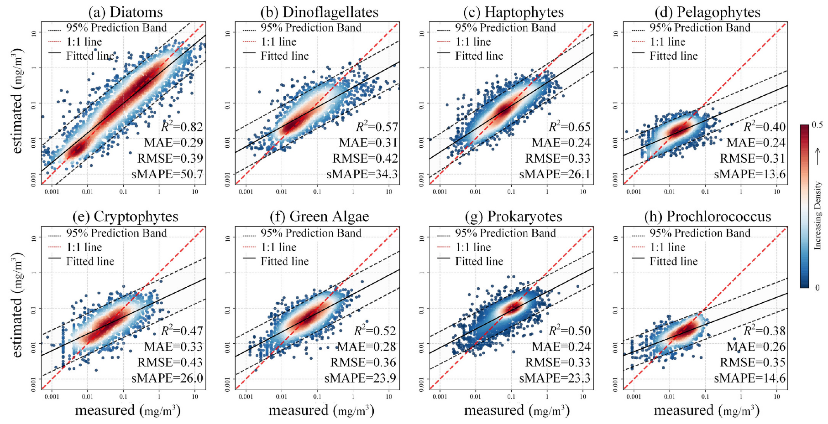


图4. 定量验证结果

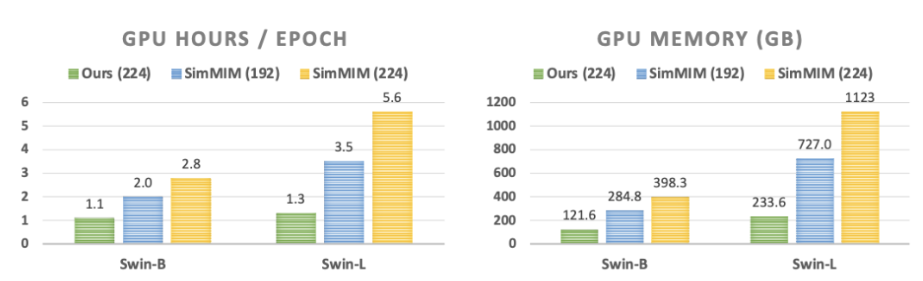


图5. 与其它模型的指标比较

Acerbi-Junior, F.W., Clevers, J.G.P.W., Schaepman, M.E., 2006. The assessment of multi-sensor image fusion using wavelet transforms for mapping the Brazilian Savanna. International Journal of Applied Earth Observation and Geoinformation, 8(4): 278-288.

Chen, B., Chen, L., Huang, B., Michishita, R., Xu, B., 2018. Dynamic monitoring of the Poyang Lake wetland by integrating Landsat and MODIS observations. ISPRS Journal of Photogrammetry and Remote Sensing, 139: 75-87.

Dao, P.D., Mong, N.T., Chan, H.-P., 2019. Landsat-MODIS image fusion and object-based image analysis for observing flood inundation in a heterogeneous vegetated scene. GIScience & Remote Sensing, 56(8): 1148-1169.

Emelyanova, I.V., McVicar, T.R., Van Niel, T.G., Li, L.T., van Dijk, A.I.J.M., 2013. Assessing the accuracy of blending Landsat–MODIS surface reflectances in two landscapes with contrasting spatial and temporal dynamics: A framework for algorithm selection. Remote Sensing of Environment, 133: 193-209.

Feng, G., Masek, J., Schwaller, M., 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.

Filali Boubrahimi, S., Neema, A., Nassar, A., Hosseinzadeh, P., Hamdi, S.M., 2024. Spatiotemporal Data Augmentation of MODIS-Landsat Water Bodies Using Adversarial Networks. Water Resources Research, 60(3): e2023WR036342.

Gevaert, C.M., 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.

Heimhuber, V., Tulbure, M.G., 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.

Hilker, T. et al., 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.

Huang, B., Song, H., 2012. Spatiotemporal Reflectance Fusion via Sparse Representation. IEEE Transactions on Geoscience and Remote Sensing, 50(10): 3707-3716.

Li, X. et al., 2021. Monitoring high spatiotemporal water dynamics by fusing MODIS, Landsat, water occurrence data and DEM. Remote Sensing of Environment, 265: 112680.

Malenovský, Z. et al., 2007. Scaling dimensions in spectroscopy of soil and vegetation. International Journal of Applied Earth Observation and Geoinformation, 9(2): 137-164.

Moosavi, V., Talebi, A., Mokhtari, M.H., Shamsi, S.R.F., Niazi, Y., 2015. A wavelet-artificial intelligence fusion approach (WAIFA) for blending Landsat and MODIS surface temperature. Remote Sensing of Environment, 169: 243-254.

Roy, D.P. et al., 2014. Landsat-8: Science and product vision for terrestrial global change research. Remote Sensing of Environment, 145: 154-172.

Song, H., Liu, Q., Wang, G., Hang, R., 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.

Sun, C. et al., 2024. Synthesizing Landsat images using time series model-fitting methods for China’s coastal areas against sparse and irregular observations. GIScience & Remote Sensing, 61(1): 2421574.

Tan, Z., Wang, X., Chen, B., Liu, X., Zhang, Q., 2019. Surface water connectivity of seasonal isolated lakes in a dynamic lake-floodplain system. Journal of Hydrology, 579: 124154.

Wang, Z., Liu, X., Li, W., He, S., Zheng, T., 2023. Temporal and Spatial Variation Analysis of Lake Area Based on the ESTARFM Model: A Case Study of Qilu Lake in Yunnan Province, China. Water, 15(10): 1800.

Wei, Q., Bioucas-Dias, J., Dobigeon, N., Tourneret, J.Y., 2015. Hyperspectral and Multispectral Image Fusion Based on a Sparse Representation. IEEE Transactions on Geoscience and Remote Sensing, 53(7): 3658-3668.

Yamazaki, D., Trigg, M.A., Ikeshima, D., 2015. Development of a global ~90m water body map using multi-temporal Landsat images. Remote Sensing of Environment, 171: 337-351.

Yao, F. et al., 2023. Satellites reveal widespread decline in global lake water storage. Science, 380(6646): 743-749.

Zhang, F., Zhu, X., Liu, D., 2014. Blending MODIS and Landsat images for urban flood mapping. International Journal of Remote Sensing, 35(9): 3237-3253.

Zheng, Y., Song, H., Sun, L., Wu, Z., Jeon, B., 2019. Spatiotemporal Fusion of Satellite Images via Very Deep Convolutional Networks, Remote Sensing.

Zhu, X., Chen, J., Gao, F., Chen, X., 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.