CyberGIS for Scalable Remote Sensing Data Fusion

Fangzheng Lyu
Department of Geography and Geographic Information Science, UIUC
Urbana, Illinois, USA
flu8@illinois.edu

Zijun Yang
Department of Geography and Geographic Information Science, UIUC
Urbana, Illinois, USA
zijuny2@illinois.edu

Zimo Xiao
Department of Geography and Geographic Information Science, UIUC
Urbana, Illinois, USA
zimox2@illinois.edu

Chunyuan Diao
Department of Geography and Geographic Information Science, UIUC
Urbana, Illinois, USA
chunyuan@illinois.edu

Jinwoo Park
Department of Geography and Geographic Information Science, UIUC
Urbana, Illinois, USA
jparkgeo@illinois.edu

Shaowen Wang
Department of Geography and Geographic Information Science, UIUC
Urbana, Illinois, USA
shaowen@illinois.edu

ABSTRACT

Satellite remote sensing data products are widely used in many applications and science domains ranging from agriculture and emergency management to Earth and environmental sciences. Researchers have developed sophisticated and computationally intensive models for processing and analyzing such data with varying spatiotemporal resolutions from multiple sources. However, the computational intensity and expertise in using advanced cyberinfrastructure have held back the scalability and reproducibility of such models. To tackle this challenge, this research employs the CyberGIS-Compute middleware to achieve scalable and reproducible remote sensing data fusion across multiple spatiotemporal resolutions by harnessing advanced cyberinfrastructure. CyberGIS-Compute is a cyberGIS middleware framework for conducting computationally intensive geospatial analytics with advanced cyberinfrastructure resources such as those provisioned by XSEDE. Our case study achieved remote sensing data fusion at high spatial and temporal resolutions based on integrating CyberGIS-Compute with a cutting-edge deep learning model. This integrated approach also demonstrates how to achieve computational reproducibility of scalable remote sensing data fusion.

CCS CONCEPTS

• **Information systems → Geographic information systems**; • **Applied computing → Environmental sciences**.

KEYWORDS

CyberGIS, Geospatial Data Science, Machine Learning

PEARC '22, July 10–14, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9161-0/22/07…$15.00
https://doi.org/10.1145/3491418.3535145

ACM Reference Format:
Fangzheng Lyu, Zijun Yang, Zimo Xiao, Chunyuan Diao, Jinwoo Park, and Shaowen Wang. 2022. CyberGIS for Scalable Remote Sensing Data Fusion. In *Practice and Experience in Advanced Research Computing (PEARC '22), July 10–14, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3491418.3535145

1 INTRODUCTION

Satellite remote sensing data is widely used in many application and science domains, and represents a major big data challenge in terms of data integration. As of January 1, 2022, there are in total 4,852 satellites orbiting and sensing the Earth for various purposes [5] . However, a specific satellite platform is often limited to having either a high spatial resolution or a high temporal resolution, but not both. The Moderate Resolution Imaging Spectroradiometer (MODIS) satellite, for example, views the entire Earth's surface every one to two days with a spatial resolution of approximately 250 meters [6]. On the other hand, the Landsat satellite provides 30-meter spatial resolution with a lower temporal resolution of every week or every two weeks [3]. To compensate for the gap, machine learning models for integration and fusion of multiple sources of remote sensing data have been extensively studied to generate remote sensing data products with both high-spatial and high-temporal resolutions (HSHT), enabling better representations and measurements of the Earth [1]. Yet, as such models are often computation- and data-intensive, significant computational challenges must be resolved to enable scalable and reproducible modeling workflows based on advanced cyberinfrastructure.

This research aims to achieve scalable and reproducible remote sensing data fusion based on advanced cyberinfrastructure via CyberGIS-Compute. We focus on addressing the limitation of either high spatial (e.g., Landsat) or temporal (e.g., MODIS) resolution. To obtain HSHT remote sensing images, remote sensing image fusion models developed by Yang et al. [9] are used to integrate high-spatial and low-temporal resolution (HSLT) images with low-spatial and high-temporal resolution (LSHT) images. CyberGIS-Compute is a cyberGIS (that is cyber-based geographic information science and systems) middleware framework for conducting high-performance and data-intensive geospatial analytics [4, 7, 8]. Supported by cyberinfrastructure resources such as Bridges-2 at the Pittsburgh

Supercomputing Center (PSC) and those provisioned by the Extreme Science and Engineering Discovery Environment (XSEDE), CyberGIS-Compute is capable of handling massive geospatial data and computationally intensive analytics. In addition, CyberGIS-Compute enables users even with little programming background to have seamless access to high-performance computing (HPC) resources through shielding the complexity of managing cyberinfrastructure [2, 7].

2 REMOTE SENSING DATA FUSION

Remote sensing data fusion has been extensively studied to generate HSHT data products by integrating HSLT images (such as Landsat) and LSHT images (such as MODIS) [1]. In this study, we utilize a robust hybrid deep learning model based on a super-resolution convolutional neural network (SRCNN) and long short-term memory (LSTM) for spatiotemporal data fusion [9]. The SRCNN is used to enhance the coarse images by restoring degraded spatial details, and the LSTM is used to learn and extract the temporal changing patterns from the time-series remote sensing images. The deep learning model requires two pairs of both fine and coarse spatial resolution images like Landsat and MODIS at the same study area. With that, given any coarse spatial resolution images taken at different times, the data fusion model is capable of generating a fine spatial resolution correspondence. This new hybrid model significantly outperforms multiple benchmark models, especially in the scenarios of rapid phenological changes. However, the lack of expertise to effectively use advanced cyberinfrastructure tailored for this model, as well as other related deep learning models have hampered their adoption and applications for large geographic areas. Thus, we investigate the solution for utilizing advanced cyberinfrastructure to enable large-scale and rapid fusion of remote sensing data with HSHT. In other words, the key focus of this work is to bridge computationally intensive deep learning models with cyberinfrastructure resources through CyberGIS-Compute, which makes the data fusion workflow scalable and reproducible, and accessible by domain experts who may have limited programming background and cyberinfrastructure knowledge.

Besides the contribution to remote sensing data fusion, this research demonstrates the power of CyberGIS-Compute as generic middleware. Not limited to our particular deep learning model, computationally intensive geospatial models can be integrated with CyberGIS-Compute to benefit from accessing advanced cyberinfrastructure resources. In the following sections, we will introduce a cyberGIS framework based on CyberGIS-Compute, and a case study of using the framework for scalable remote sensing data fusion.

3 FRAMEWORK

CyberGIS-Compute is a key component of the CyberGISX platform that provides streamlined and user-friendly access to advanced cyberinfrastructure and cyberGIS capabilities with an integrated software stack for computationally reproducible and data-intensive geospatial analytics [4, 10]. As shown in Figure 1, there are three major layers for registering and running a computational job with CyberGIS-Compute: frontend interface, middleware, and computing resources at the backend. The backend cyberinfrastructure resources include Bridges-2 at PSC and other XSEDE resources as well as Virtual ROGER - a geospatial supercomputer hosted by the CyberGIS Center for Advanced Digital and Spatial Studies at the University of Illinois. The hybrid deep learning fusion model shown in Figure 1 integrates SRCNN and LSTM. The SRCNN model contains three convolutional layers for feature extraction, non-linear mapping, and reconstruction, respectively. The LSTM model consists of two LSTM layers, each with 100 LSTM units, followed by a dropout layer and a fully-connected layer to generate the final results. A total of over 350,000 parameters are needed to be optimized during the training process of the hybrid fusion model. When utilized for generating dense time-series data, the hybrid model needs to be trained for each prediction date and thus requires substantial computational resources.

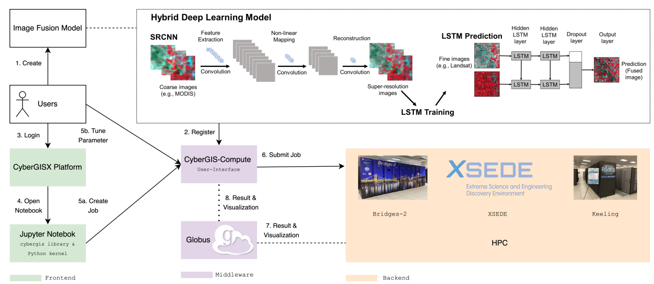

Figure 1: CyberGIS-Compute Architecture

To utilize CyberGIS-Compute, a user needs to specify the configuration of computational jobs tailored to optimize the use of cyberinfrastructure resources. A typical configuration includes: 1) job name; 2) container used to execute the job; 3) scripts, such as Python or bash script, for preprocessing, executing, and post-processing stages of a model workflow; 4) scheduler (e.g., slurm) interface for managing wall time and number of tasks; 5) input data and parameters; and 6) download path, which is the output path if there is any result files or visualization as the output of the model. CyberGIS-Compute is implemented as a library pre-installed on the CyberGISX platform. Users can take advantage of the library via an interactive user interface for submitting jobs to backend HPC resources in a Jupyter notebook environment.

4 CASE STUDY

We have applied the cyberGIS framework to fuse MODIS and Landsat data for producing HSHT images. Specifically, four major steps were taken as shown in Figure 2: a) invoke the Graphic User Interface (GUI) widget and tune a suite of parameters, b) submit the job and inspect the progress of the job with a status bar, c) monitor the real-time job status and receive the job logs once the job is done on HPC resources, and d) retrieve the output of the fusion model and download the corresponding results.

4.1 Job Configuration

Figure 2a shows the initial GUI of the CyberGIS-Compute submission tab. The cyberGIS library, which is pre-installed on the CyberGISX platform, can be used to create a job for submitting pre-registered models and associated workflow to the backend

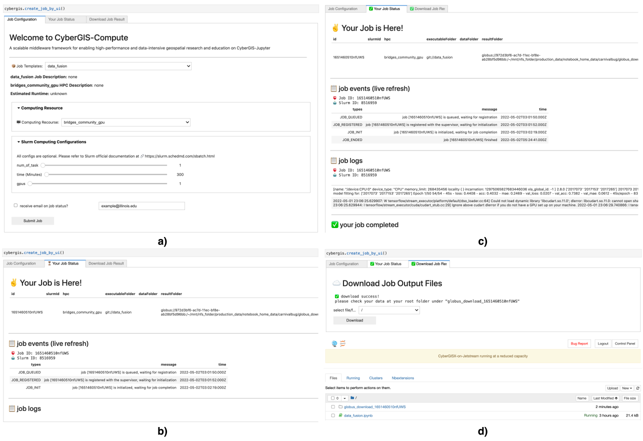

Figure 2: Workflow. a) job configuration and submission tab, b) job status after submitting the job, c) job status tab after finishing the job, d) output retrieval tab and output folder on CyberGISX

HPC resources. There are multiple configurable parameters including: 1) Job Templates that allow users to choose from existing pre-registered models such as the data-fusion model in our case study; 2) Computing Resource that enables users to choose from existing computing resources available, including XSEDE, Bridges-2, and Virtual ROGER; 3) Slurm job configuration that includes the number of repeating tasks for the job submission, wall time for the job to terminate, and the required amount of computing resources; 4) (Optional) Input parameters that vary depending on the job being run; 5) (Optional) Upload data if there is any required for the job to execute; and 6) Email address if users intend to get notification of the job status. Once clicking the Submit Job, CyberGIS-Compute will collect all the configurations and data and then submit the job to the specified HPC resources for execution.

4.2 Job Status and Output Log

A dynamic job status bar is presented to users along with a table for basic job information to inform users about the current stage of computation progress (e.g., queuing, running, finished). Figure 2b provides one example of the job status bar. The message including job id which is dynamically assigned by slurm job management system and time information can be made available as part of the status bar for users to understand the current status of high-performance computation, which could also be important for achieving computationally reproducible modeling workflows. Once a job is finished, CyberGIS-Compute will retrieve output logs from HPC resources. Figure 2c is an example of running the job successfully with CyberGIS-Compute.

4.3 Output and Visualization

Users can retrieve data fusion results within the CyberGISX platform as shown in Figure 2d. Using Globus, CyberGIS-Compute transfers the output folders and files from HPC resources into the CyberGISX cloud environment. Figure 3 shows a sample output of a predicted remote sensing image in Central Illinois on June 2, 2017 from remote sensing data fusion model. This 30-meter high spatial resolution image representing an area covering 57,323 square kilometer in central Illinois, US is fused using our model with MODIS data collected on March 14, June 2 and September 22 together with Landsat data collected on March 14 and September 22 in 2017. The data fusion workflow in the case study was enabled with Bridges-2, which is equipped with the Tesla V100-SXM2-32GB GPU. The computational process for this data fusion model takes about 2 hours and 20 minutes to finish and consumes about 4 service units (SUs) and 7.8 GB in memory on Bridges-2. With nodes and GPU cores provided by HPC resources, our computationally intensive data fusion model can be extended to significantly larger geographic area.

In our research context, all users will be able to reproduce the same fine spatial resolution image as output with the same seed for number generator. For our particular data fusion model, the reproducibility is guaranteed as the same input data is used. For other computationally intensive models, CyberGISX platform and CyberGIS-Compute can be customized to provide reproducibility and scalability support.

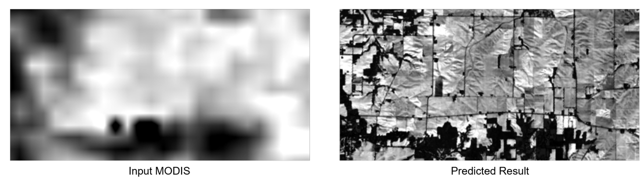

Figure 3: Input MODIS data and predicted high spatial resolution sample output on June 2, 2017 in Central Illinois, US

5 CONCLUSION AND FUTURE WORK

This research develops a novel cyberGIS framework for scalable remote sensing data fusion using advanced cyberinfrastructure. CyberGIS-Compute, a key component of the framework, serves as user-friendly middleware bridging deep learning models for data fusion with cyberinfrastructure resources. Compared with conventional methods of running a remote sensing data fusion model using local or cloud computing resources, our framework not only enables computationally scalable fusion of satellite remote sensing data with both high spatial and temporal resolutions, but also supports users with little programming background to execute sophisticated and computationally intensive models in a reproducible way.

In a variety of application and science domains, satellite remote sensing images fused from multiple sources can provide critical support for solving important problems. However, as advanced data fusion models often involve sophisticated deep learning models, the scalability of these models can be limited due to a lack of access to advanced cyberinfrastructure. Therefore, the cyberGIS framework described in this paper addresses this scalability challenge by integrating the data fusion model and advanced cyberinfrastructure. A major area of future research is to enhance the cyberGIS framework by incorporating more computation-, data- and collaboration-intensive models, such as the data fusion model in this study, to enable diverse users to have easy access to advanced cyberinfrastructure. Additional models such as more robust and larger scale data fusion model, agent based models and spatial accessibility model will be will be added to the CyberGIS-compute capability in the near future.

ACKNOWLEDGMENTS

This research is supported in part by the United States Department of Agriculture (USDA) under grant No. 2021-67021-33446. The research is also based upon work supported in part by the Institute for Geospatial Understanding through an Integrative Discovery Environment (I-GUIDE) that is supported by the National Science Foundation (NSF) under award No. 2118329. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of NSF. Our computational work used NSF XSEDE and Virtual ROGER. Virtual ROGER is a geospatial supercomputer supported by the CyberGIS Center for Advanced Digital and Spatial Studies and the School of Earth, Society and Environment at the University of Illinois Urbana-Champaign.

REFERENCES

[1] Feng Gao, Thomas Hilker, Xiaolin Zhu, Martha Anderson, Jeffrey Masek, Peijuan Wang, and Yun Yang. 2015. Fusing Landsat and MODIS Data for Vegetation Monitoring. *IEEE Geoscience and Remote Sensing Magazine* 3, 3 (2015), 47–60. https://doi.org/10.1109/MGRS.2015.2434351

[2] Fangzheng Lyu, Dandong Yin, Anand Padmanabhan, Youngdon Choi, Jonathan L. Goodall, Anthony Castronova, David Tarboton, and Shaowen Wang. 2019. Reproducible Hydrological Modeling with CyberGIS-Jupyter: A Case Study on SUMMA *(PEARC '19)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3332186.3333052

[3] NASA. n.d.. *Landsat Overview*. https://www.nasa.gov/mission_pages/landsat/overview/index.html

[4] Anand Padmanabhan, Ximo Ziao, Rebecca C. Vandewalle, Furqan Baig, Alexander Michel, Zhiyu Li, and Shaowen Wang. 2021. CyberGIS-Compute for Enabling Computationally Intensive Geospatial Research *(SpatialAPI '21)*. Association for Computing Machinery, New York, NY, USA, Article 3, 2 pages. https://doi.org/10.1145/3486189.3490017

[5] UCSUSA. 2021. *UCS Satellite Database*. http://www.ucsusa.org/nuclear-weapons/space-weapons/satellite-database

[6] USGS. n.d.. *MODIS Overview*. https://lpdaac.usgs.gov/data/get-started-data/collection-overview/missions/modis-overview/

[7] Shaowen Wang. 2010. A CyberGIS Framework for the Synthesis of Cyberinfrastructure, GIS, and Spatial Analysis. *Annals of the Association of American Geographers* 100, 3 (2010), 535–557. https://doi.org/10.1080/00045601003791243

[8] Shaowen Wang. 2016. CyberGIS and spatial data science. *GeoJournal* 81 (2016), 965–968. https://doi.org/10.1007/s10708-016-9740-0

[9] Zijun Yang, Chunyuan Diao, and Bo Li. 2021. A Robust Hybrid Deep Learning Model for Spatiotemporal Image Fusion. *Remote Sensing* 13, 24 (2021). https://doi.org/10.3390/rs13245005

[10] Dandong Yin, Yan Liu, Anand Padmanabhan, Jeff Terstriep, Johnathan Rush, and Shaowen Wang. 2017. A CyberGIS-Jupyter Framework for Geospatial Analytics at Scale *(PEARC17)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3093338.3093378