From Geospatial Data Cube to AI Cube: the Open Geospatial Engine (OGE) Approach

Peng Yue [1,2], Kaixuan Wang [1], Hanwen Xu [1], Jianya Gong [1], Longgang Xiang [3]

[1] School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan, Hubei, 430079, China - (pyue, wangkaixuan, xuhanwen, gongjy)@whu.edu.cn
[2] Hubei Province Engineering Center for Intelligent Geoprocessing (HPECIG), Wuhan University, 129 Luoyu Road, Wuhan, Hubei, 430079, China
[3] State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing (LIESMARS), Wuhan University, 129 Luoyu Road, Wuhan, Hubei, 430079, China - geoxlg@whu.edu.cn

Keywords: Geospatial Data Cube, AI Cube, Geospatial Artificial Intelligence, AI-ready SDI, Open Geospatial Engine (OGE), Digital Earth.

Abstract

The Earth Observation (EO) analytics are moving from local systems to online cloud computing platforms such as Google Earth Engine (GEE) and Open Geospatial Engine (OGE). A typical approach in existing efforts is to leverage geospatial data cubes with cloud computing to support large-scale big EO data analytics in Digital Earth systems. While online analytical processing (OLAP) can be enabled using the cube approach, it is still not clear how geospatial artificial intelligence (GeoAI) can be incorporated in data cubes to benefit the cube infrastructure. Such an investigation can consolidate the vision of an AI-ready SDI (Spatial Data Infrastructure). The paper presents a systematic approach to incorporate GeoAI models into geospatial data cubes to help create an AI Cube. It covers on-demand model retrieval, cube data and model integration, and distributed model inference. The approach is demonstrated in OGE, which is an EO cloud computing platform layered on the GeoCube implementation. The results show that such an AI Cube enriches a cube infrastructure with GeoAI capabilities, facilitates the on-demand coupling of cube data and GeoAI models, and improves the performance of GeoAI inference.

1. Introduction

The increasing volume of Earth Observation (EO) motivates the changes of the way EO data is being processed. Typical examples of big EO data processing in existing Digital Earth systems include Google Earth Engine (GEE) (Gorelick et al., 2017) and Open Data Cube (ODC) (CEOS, 2021). They harness advanced information infrastructures like cloud computing to move the traditional local processing into online processing. Since the EO data processing is characterized by multiple data types, different spatial-temporal resolutions, and various tasks, the organization and processing of big EO data is a challenging topic.

Recent process on data cubes have shown great promise in extracting information from analysis-ready data (Baumann et al., 2018; Giuliani et al., 2017; Sudmanns et al., 2020). Data cube originated in the business intelligence field, where data is organized into multi-dimensional arrays according to interested dimensions. Recently data cube has been investigated in the EO domain to create a cube infrastructure. Both the International Committee on Earth Observation Satellites (CEOS) and Open Geospatial Consortium (OGC) are working on moving the geospatial data cube forward. The EarthServer data cube proposed by Baumann et al. (Baumann et al., 2018) supports the management and processing of raster data using an array database approach. The ODC launched by the CEOS (Gomes et al., 2021) was developed for the analysis of time series images. Both solutions are mainly designed for raster data. In an Open Geospatial Engine (OGE), the GeoCube is proposed (Gao et al., 2022), which extends the capacity of data cubes to different types of geospatial data including both vector and raster data.

Recent work envisions an AI-ready SDI (Spatial Data Infrastructure) by adding AI data and models into an SDI (Yue et al., 2022). Such an SDI can be consolidated by a cube infrastructure enriched with AI capabilities, shortly named as an AI Cube. While online analytical processing (OLAP) can be enabled using the cube approach, it is still not clear how geospatial artificial intelligence (GeoAI) can be incorporated in data cubes to benefit the cube infrastructure. For example, images are still the basic unit of DL inference (Aspri et al., 2020; Fang et al., 2021; Lunga et al., 2020), making it difficult to develop fast and accurate inferences for tile-based distributed data in data cubes. Furthermore, different from traditional EO processing algorithms, GeoAI models have their own specific characteristics in data analytics. Due to the spatial heterogeneity of geographical phenomena, GeoAI models such as remote sensing (RS) intelligent interpretation models have varying inference abilities on EO imagery from different regions, seasons, and scales. One individual GeoAI model is often not general enough in various tasks and regions. Thus it is often necessary to integrate various trained models together to support various inference tasks at different places in a Digital Earth system.

The paper presents a systematic approach to incorporate GeoAI models into geospatial data cubes to help create an AI Cube. It covers on-demand model retrieval, cube data and model integration, and distributed model inference. The approach is developed in OGE, which is an EO cloud computing platform layered on the GeoCube implementation. The results demonstrate the applicability of the approach.

The rest of the paper is organized as follows. Section 2 presents the key issues towards the vision of an AI Cube. Section 3 presents how the OGE GeoCube is used to develop an AI Cube. The implementation and evaluation of the AI Cube are described in Section 4. Section 5 raises several observations that need some further discussion. And finally, Section 6 presents the conclusions and future work.

2. What is an AI Cube

A data cube generally refers to an array of multiple dimensions that facilitates online analytics, it is also known as an OLAP cube (Gray et al., 1997). The OLAP cube was introduced into the geospatial domain to accommodate various geospatial data including raster and vector data by adding the spatial dimension and spatial measure (Gao et al., 2022). In recent years, this term has been used in the EO domain to refer to a time-series multi-dimensional array. It is combined with Cyberinfrastructure to create EO data cube infrastructure, which facilitates EO data management, access, analysis, visualization, and interoperability.

The AI Cube is proposed to empower geospatial data cubes with GeoAI analytical functions. It takes benefits of analysis-ready data organization in geospatial data cubes and analytical functions from GeoAI models. An AI cube can accommodate various GeoAI models and make model inference over the cube data possible. Fig.1 gives an overall architecture for an AI Cube, which has the following major characteristics:

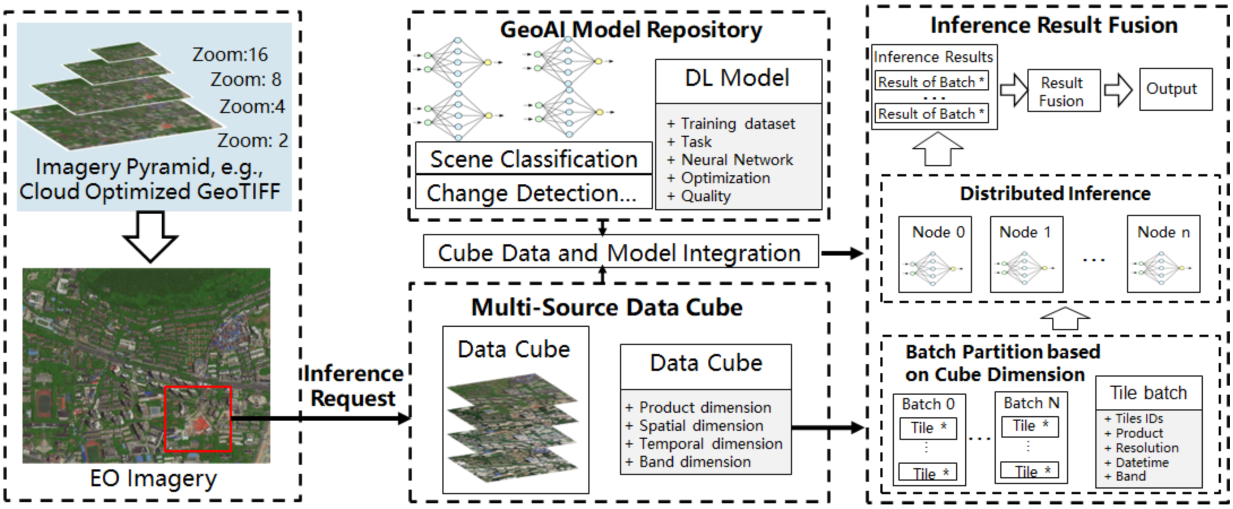


Figure 1. An AI Cube framework.

(1) On-demand model retrieval: The AI Cube is augmented with a DL model repository that stores trained GeoAI models. The repository could be a model fact table in a cube. Each model has attributes such as task, neural network, spatial and temporal applicability. These attributes help to filter appropriate models in Cube. Traditional cubes focus on using formalized dimensions to index data from fact tables. From an AI Cube perspective, the model attributes can be formalized as dimensions in cubes. Then they can be indexed using cube dimensions from a model fact table, thus promoting a data cube to a model cube. Assuming a decision support scenario where users want to know the cropland changes in the past several years in a spatial region, the change detection task, together with spatial and temporal requirements, will be used in the cube to retrieve GeoAI models from the cube on demand.

(2) Cube data and model integration: Data and model need to be coordinated to be prepared for distributed inference. On the one hand, traditional cube data retrieval and OLAP operations can be reused to efficiently get input data for models. Data are organized in the cube following well-defined ways like the star, snowflake, and fact constellation schemas (Chaudhuri and Dayal, 1997). The data organization is further combined with advanced infrastructure like cloud computing using distributed databases/files and computing technologies. To improve the performance of retrieval of EO imagery in cubes, the EO data is often organized in grids using a pyramid structure like the Cloud Optimized GeoTiff (COG) files. The OLAP operations on them can get high level product as input. On the other hand, these data still need to be pre-processed in a ready form for inference. The processing typically includes tile overlapping and band normalization. For example, in terms of pixel-level inference tasks, it goes through the following steps: performing inference separately on tiles and then stitching the results together. This may result in noticeable seam artifacts. A common solution for the pre-processing is to ensure the existence of overlapping regions between tiles and optimize the results through probability voting. The other one is band normalization, where the inference data needs to undergo a normalization process consistent with the training samples before feeding into the model.

(3) Distributed model inference: The model inference can take advantage of cube infrastructure by disseminating tiles as batches into a distributed inference engine. First, the tile IDs from multi-source EO data are retrieved from the cube dimension and packaged into several batches, where the data in each batch has the same or similar product, resolution, datetime and bands. After that, each batch is allocated to different nodes for parallel computing, and each node automatically selects the matched DL models through the batch attributes for multi-GPU inference. Finally, the inference results of cube tiles are mosaiced together and post-processed to get the final inference result. Different tasks require different post-processing methods. For example, in object detection tasks, techniques like non-maximum suppression (NMS) (Neubeck and Van Gool, 2006) or weighted box fusion (WBF) (Solovyev et al., 2021) are commonly used to eliminate multiple detection boxes for the same object. In terms of pixel-level inference tasks, steps such as image stitching, color mapping, denoising, and smoothing are typically performed to optimize the inference results (Krähenbühl and Koltun, 2011).

3. The OGE Approach

The OGE is an EO cloud computing platform for global scale data management and geoprocessing. It is developed by Wuhan University and available to the public online at http://www.openge.org.cn/. OGE integrates and manages multi-source, heterogeneous, and big spatiotemporal data on a global scale using the GeoCube model. The data are managed using the star constellation schema. It includes different dimensions, such as product, spatial, temporal, and band dimensions. The product/band dimension specifies the thematic axis using the product name (e.g., Sentinel 2 or Landsat 8), instrument name etc. Spatial dimension specifies the spatial axis using the grid code, grid name, city name etc. Temporal dimension specifies the temporal axis using the acquisition time and result time. EO images are organized as tiles in fact tables, which can be retrieved through multi-dimensional cube queries.

GeoCube supports query and processing operations as follows. The multi-dimensional cube queries are parsed and processed on the database where the tile IDs are retrieved from the dimension and fact tables. These tile IDs are partitioned and assigned to each computing unit to enable parallel data access. Cube processing includes batch processing, cloud computing, and multi-thread computing. The data processing involves multiple processes among the computing nodes and multiple threads in one node, thereby supporting hybrid parallelism.

GeoAI models are integrated into OGE GeoCube by adding new dimensions and fact tables. Fig.2 shows the cube approach to accommodate GeoAI models, where the task-class dimension is added to join the model fact table. This dimension describes the types of tasks that GeoAI models can perform and their corresponding classifications. Dimension members are defined by using combinations of tasks and/or classes in terms of classification tasks. Model facts are indexed by spatialKey, timeKey, productBandKey, taskClassKey, and modelKey. It also includes fields modelId and qualityElements. The modelId can join to the complete metadata of the model. QualityElements can provide the quality information of the model in terms of spatial measurement in each cell.

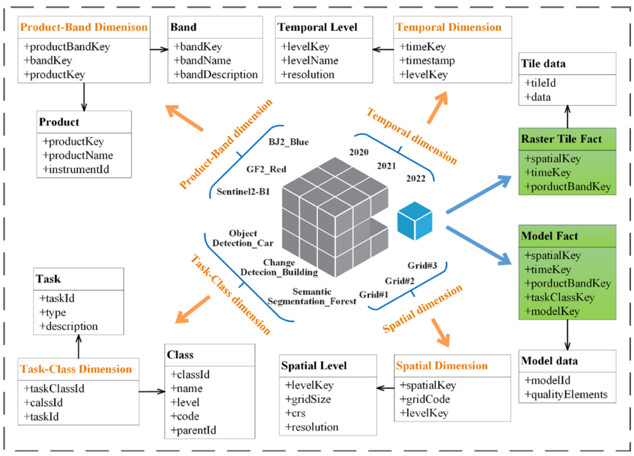


Figure 2. A cube approach to accommodate GeoAI models.

The model inference is integrated into the cube processing operations through a CPU/GPU hybrid distributed computing approach. This includes the preprocessing into a ready form and distributed inference. Preprocessing utilizes the parallel acceleration of a distributed multi-node CPU environment. The input data is initially parallelly sliced, and distributed tiles are computed in parallel by thread pools on each node. After preprocessing, the tile data is organized into several batches, with each batch consisting of multiple tensors. The main thread schedules pending inference batches, sequentially pushing batch data to multiple nodes with multiple GPUs for parallel inference. After all batches have been inferred, the main thread collects and stitches results, and performs corresponding post-processing on the resulting tiles, thus generating the final result.

4. Implementation

A prototype system based on the OGE is developed to demonstrate the applicability of the approach (Fig. 3). It follows a layered architecture: hardware layer, data layer, computing layer, and application layer. The hardware layer includes CPU/GPU high-performance computing clusters, data storage arrays, and fiber optic network links. EO data and models are imported into the data layer. The computing layer is responsible for deploying models, receiving data, and executing the inference workflow. The application layer extends the OGE Web GUI to allow users perform specific inference tasks online in a user-friendly way.

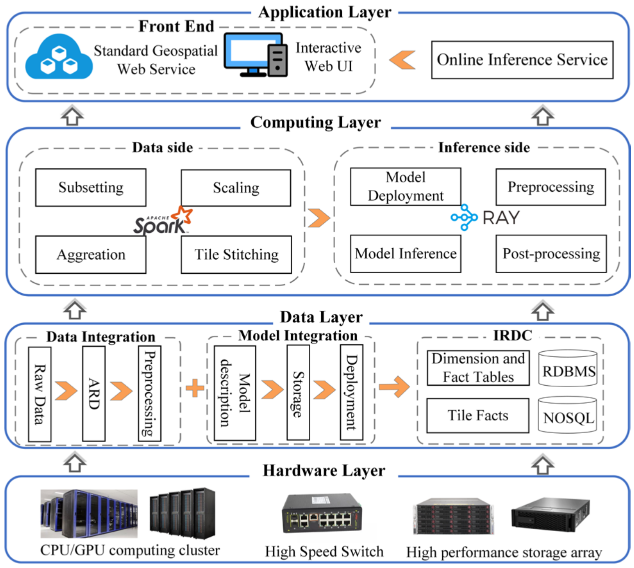
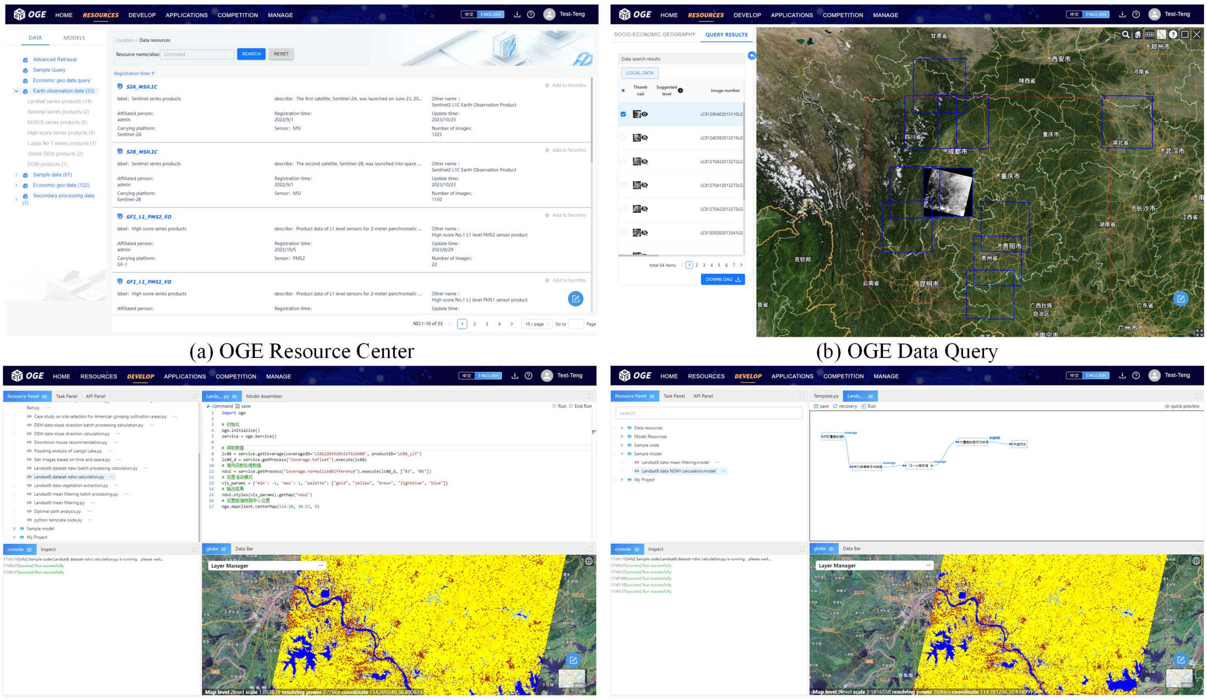


Figure 3. Implementation architecture of the system.

In the front end, OGE provides a set of Web Graphical User Interface (GUI) for data retrieval and programming (Fig. 4). Fig4a and 4b show the data resource and data query in OGE respectively. Fig4c shows the OGE support users to do programming interactively in a Web GUI, and Fig4d shows that users can build geoprocessing workflows in a drag-and-drop way with low code efforts.

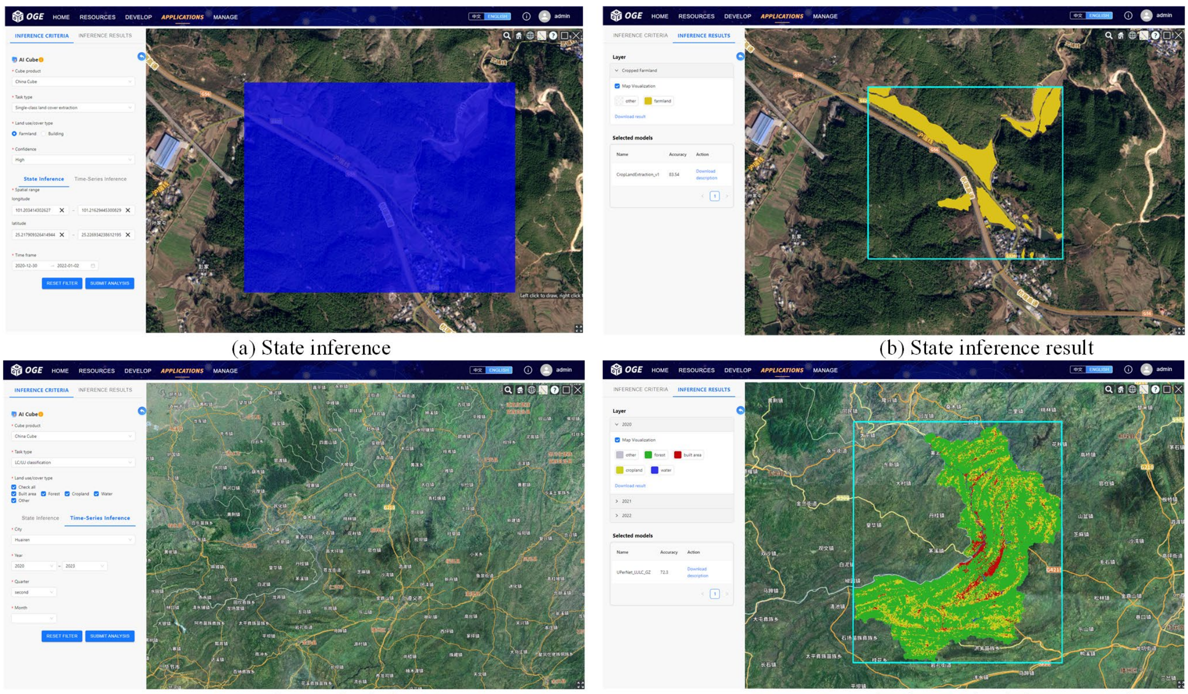
By extending the GUI to the GeoAI model inference, Fig. 5 shows a user-friendly GUI. It allows users to select tasks and spatiotemporal ranges to get inference results without delving into technical details on what data and models being used. Task types include land user and land cover (LULC) classification, single-class land cover extraction, change detection, object detection, etc. Two types of inference are supported: inference of state and time-series inference. The inference of state (Fig5a and 5b) means that the latest EO data available for the specified period are used for deriving the state of the area. The time-series inference (Fig5c and 5d) means that the cube-based OLAP operations are used to aggregate images, followed by the inference on the aggregated result. For example, the weekly images are rolled up into a quarterly image using the average method to reflect the comprehensive characteristics of that quarter. Such cube operations can be combined with inference to construct an inference pipeline, improving the performance of cube data processing.



(a) OGE Resource Center (b) OGE Data Query

(c) OGE Web Programming (d) OGE Web Workflow Builder

Figure 4. OGE Web GUI.



(a) State inference (b) State inference result

(c) Time-series inference (d) Time-series inference result

Figure 5. Snapshots of model inference.

5. Discussion

The paper presents an approach to couple geospatial data cubes with GeoAI models, thus empowering the traditional EO cubes with capabilities of AI analytical functions. Traditional EO cubes manage geospatial data to make them ready for cube OLAP analysis. Such cube analytics integrate traditional geoprocessing functions with parallel distributing paradigms in the cloud computing such as MapReduce(Dean and Ghemawat, 2008) to process EO data in distributed computing nodes. The traditional geoprocessing functions can be regarded as physics-based models. In this paper, the AI cube aims to accommodate data-driven models into the cube infrastructure. The implementation raises several observations that need some further discussion.

Metadata for GeoAI models: The AI Cube framework needs the access to and utilization of GeoAI models. The premise is that these models are described using a formal approach, enabling model management, retrieval, deployment, and invocation. Currently the interoperability of model descriptions needs more investigations. Well-defined metadata for GeoAI models can facilitate efficient exchange, sharing, and reuse of models in the cube infrastructure. The OGC TrainingDML-AI Standard Working Group (SWG) is working on extending the standards on GeoAI training data to models (Yue et al., 2022; Yue and Boyi, 2023). The OGC Testbed19 machine learning models engineering report also makes recommendations for work on model specification (Samantha and Trent, 2024). Compared to geospatial data, the descriptions of GeoAI models not only include the general metadata, but also emphasize the inputs and outputs of models. The specification of the model descriptions also needs to follow FAIR principles, which refer to findability, accessibility, interoperability, and reusability.

Matchmaker between tasks and models: The matchmaker helps to select appropriate GeoAI models from a model repository based on specific tasks. Some public repositories such as Hugging Face (Wolf et al., 2020), TensorFlow Hub, and PyTorch Hub, offer users a plethora of available DL models. There are two types of matching, explicit and implicit matching. The former one is based on keyword matching and spatiotemporal filtering. The latter one can rely on model feature repository that are created using deep feature from AI. It is worthwhile to investigate which one is better for matchmake between tasks and models.

AI ready data and model preparation: The paper unifies the organization of data and models through a multi-dimensional view. Unlike traditional EO data cubes, each cell is not limited to raster or vector data. It integrates facts and extends dimension tables to access GeoAI models. The data in the cell cannot be used directly as input for models, often requiring OLAP operations to pre-process the data. In addition, it needs to be transformed into a machine-readable format such as Tensor. The models also need to be deployed and disseminated to distributed nodes to facilitate the on-the-fly inference. Thus the model can be in an operational state, ready to accept input data at any time. This means that both data and model need to be prepared ready for inference. The results enrich the investigation of an AI-ready SDI.

Inference pipeline: Traditionally it is the labour-intensive work to apply GeoAI models to process EO data, which typically involves data pre-processing, model deployment, format tuning, and data post-processing. This ad-hoc workflow hampers the efficient application of GeoAI models in the EO domain, necessitating automated solutions to provide efficient inference decision services. The AI extension to the GeoCube builds an automated inference pipeline that fully utilizes high-performance computing resources to accelerate the application of models. This pipeline can be evaluated in the future against various cases and large scale inference to check performance and make optimization when needed.

Hybrid physical and AI modelling: It has been discussed that physical and machine learning models can be integrated to complement each other (Reichstein et al., 2019). Although they basically belong to two different paradigms: theory-driven and data-driven, each of them has its own benefits. The former one basically is interpretable, and the latter one offers the potential to find unexpected the synergy generally can happen at several places: 1) using machine learning to improve parameterizations of physical models, 2) replacing a physical sub-model with a machine learning model, 3) analysis of model-observation mismatch. The physical model in traditional geospatial data cubes can be coordinated with the GeoAI models in AI Cube to create an analytical workflow for hybrid physical and AI modelling. This will improve the modelling capabilities of geospatial data cubes.

6. Conclusions and Future Work

The paper presents an AI Cube to enrich the traditional geospatial data cube with GeoAI analytical functions. It presents the framework of an AI Cube. The development of it follows the extensions of GeoCube in OGE. The model organization and distributed inference in GeoCube are highlighted.

The prototype implementation demonstrates the applicability of on-demand model retrieval, cube data and model integration, and distributed model inference. The results show that such an AI Cube enriches a cube infrastructure with GeoAI capabilities, facilitates the on-demand coupling of cube data and GeoAI models, and improves the performance of GeoAI inference.

There are also several observations, such as specification for model descriptions, matchmaker for tasks and models, AI ready data and model preparation, and automation of inference pipeline. We will work closely with OGC TrainingDML-AI SWG to move the work foward.

Acknowledgements

The work was supported by the National Natural Science Foundation of China (No. 42071354), the Natural Science Foundation of Hubei Province (ZRMS2022000666), and the Fundamental Research Funds for the Central Universities (2042022dx0001).

References

Aspri, M., Tsagkatakis, G., Tsakalides, P., 2020. Distributed Training and Inference of Deep Learning Models for Multi-Modal Land Cover Classification. *Remote Sensing,* 12(17), 2670.

Baumann, P., Rossi, A.P., Bell, B., Clements, O., Evans, B., Hoenig, H., Hogan, P., Kakaletris, G., Koltsida, P., Mantovani, S. and Marco Figuera, R., 2018. Fostering cross-disciplinary earth science through datacube analytics. *Earth Observation Open Science and Innovation*, 91-119.

CEOS, 2021. Open Data Cube. https://www.opendatacube.org/.

Chaudhuri, S. and Dayal, U., 1997. An overview of data warehousing and OLAP technology. *ACM Sigmod record*, 26(1), 65-74.

Dean, J. and Ghemawat, S., 2008. MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107-113.

Fang, J., Liu, Q. and Li, J., 2021, April. A deployment scheme of YOLOv5 with inference optimizations based on the triton inference server. *IEEE 6th International Conference on cloud computing and big data analytics (ICCCBDA),* 441-445.

Gao, F., Yue, P., Cao, Z., Zhao, S., Shangguan, B., Jiang, L., Hu, L., Fang, Z. and Liang, Z., 2022. A multi-source spatio-temporal data cube for large-scale geospatial analysis. *International Journal of Geographical Information Science*, 36(9), 1853-1884.

Giuliani, G., Chatenoux, B., De Bono, A., Rodila, D., Richard, J.P., Allenbach, K., Dao, H. and Peduzzi, P., 2017. Building an earth observations data cube: lessons learned from the swiss data cube (sdc) on generating analysis ready data (ard). *Big Earth Data*, 1(1-2), 100-117.

Gomes, V.C., Carlos, F.M., Queiroz, G.R., Ferreira, K.R. and Santos, R., 2021. Accessing and processing Brazilian earth observation data cubes with the open data cube platform. I*SPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 4, 153-159.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202, 18-27.

Gray, J., Chaudhuri, S., Bosworth, A., Layman, A., Reichart, D., Venkatrao, M., Pellow, F., Pirahesh, H., 1997. Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals. *Data Mining and Knowledge Discovery*, 1, 29–53.

Krähenbühl, P. and Koltun, V., 2011. Efficient inference in fully connected crfs with gaussian edge potentials. *Advances in neural information processing systems*, 24.

Lunga, D., Gerrand, J., Yang, L., Layton, C. and Stewart, R., 2020. Apache spark accelerated deep learning inference for large scale satellite image analytics. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 271-283.

Neubeck, A. and Van Gool, L., 2006, August. Efficient non-maximum suppression. *18th international conference on pattern recognition (ICPR'06)*, 3, 850-855).

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N. and Prabhat, F., 2019. Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743), 195-204.

Samantha L. and Trent T., 2024. Testbed-19: Machine Learning Models Engineering Report.

Solovyev, R., Wang, W. and Gabruseva, T., 2021. Weighted boxes fusion: Ensembling boxes from different object detection models. *Image and Vision Computing*, 107, 104117.

Sudmanns, M., Tiede, D., Lang, S., Bergstedt, H., Trost, G., Augustin, H., Baraldi, A. and Blaschke, T., 2019. Big Earth data: disruptive changes in Earth observation data management and analysis?. *International Journal of Digital Earth*, 13, 832–850.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M. and Davison, J., 2019. Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771.*

Yue, P. and Shangguan, B., 2023. OGC Training Data Markup Language for Artificial Intelligence (TrainingDML-AI) Part 1: Conceptual Model Standard. https://docs.ogc.org/is/23-008r3/23-008r3.html

Yue, P., Shangguan, B., Hu, L., Jiang, L., Zhang, C., Cao, Z. and Pan, Y., 2022. Towards a training data model for artificial intelligence in earth observation. *International Journal of Geographical Information Science*, 36(11), 2113-2137.