Unleashing the potential of remote sensing foundation models via bridging data and computility islands

Yansheng Li,[1] Jieyi Tan,[1,*] Bo Dang,[1] Mang Ye,[2] Sergey A. Bartalev,[3] Stanislav Shinkarenko,[3] Linlin Wang,[1] Yingying Zhang,[4] Lixiang Ru,[4] Xin Guo,[5] Liangqi Yuan,[6] Lei Yu,[4] Jingdong Chen,[4] Ming Yang,[4,*] José Marcato Junior,[7] and Yongjun Zhang[1,*]

[1]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[2]School of Computer Science, Wuhan University, Wuhan 430072, China
[3]Space Research Institute, Russian Academy of Sciences, Moscow 119421, Russia
[4]Ant Group, Hangzhou 310013, China
[5]Shanghai Academy of Artificial Intelligence for Science, Shanghai 200232, China
[6]College of Engineering, Purdue University, West Lafayette, IN 47907, USA
[7]Faculty of Engineering, Architecture, and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande 79070-900, Brazil
*Correspondence: tanjieyi@whu.edu.cn (J.T.); m.yang@antgroup.com (M.Y.); zhangyj@whu.edu.cn (Y.Z.)

Received: November 5, 2024; Accepted: February 19, 2025; Published Online: June 2, 2025; https://doi.org/10.1016/j.xinn.2025.100841

Citation: Li Y., Tan J., Dang B., et al., (2025). Unleashing the potential of remote sensing foundation models via bridging data and computility islands. The Innovation **6(6)**, 100841.

DATA AND COMPUTILITY ISLANDS IN REMOTE SENSING FOR EO

The rapid advancement of Earth observation (EO) capabilities is driving an explosive increase in remote sensing data. There is an urgent need for advanced processing techniques to unleash their application value.[1] Generalist EO intelligence refers to the ability to provide unified support for qualitative interpretation, quantitative inversion, and interactive dialogue across diverse EO data and tasks. It has attracted significant attention recently, prompting academia, industry, and government to invest substantial resources.[2] Through developing remote sensing foundation models (RSFMs), generalist EO intelligence can ultimately offer humanity a shared spatial-temporal intelligence service in various fields (e.g., agriculture, forestry, and oceanography).[3] However, a critical question remains: have we truly unleashed the potential of RSFMs for generalist EO intelligence? Despite the vast volume of remote sensing data, their distribution is often fragmented and decentralized due to privacy concerns, storage bottlenecks, industrial competition, and geo-information security. This fragmentation leads to data islands, which limit the full utilization of multi-source remote sensing data. Moreover, computility (i.e., computational resources) typically develops in isolation, inadequately supporting the large-scale training and application of RSFMs.

Limitation of the cloud-based architecture for RSFMs

Cloud-based architecture for remote sensing, exemplified by Google Earth Engine, is a widely adopted paradigm. It refers to the centralized storage and processing of remote sensing data on a cloud platform. However, this centralized paradigm demands substantial costs, typically affordable only by tech giants. Furthermore, this architecture raises several concerns, such as privacy concerns, monopolization, and ambiguous data ownership. Additionally, the data, storage, and computility on a single cloud platform are limited and difficult to scale, resulting in scalability challenges that hinder the growth and applicability of RSFMs.

Opportunities and challenges of the collaborative architecture for RSFMs

As we approach the upcoming era of sixth-generation mobile networks (6G), advancements in communication technologies are expected to provide significant network support, facilitating the collaborative architecture for RSFMs to address these critical challenges.[4] Techniques such as federated learning and split learning can facilitate cross-cloud collaborative pre-training and fine-tuning of RSFMs, eliminating the need to transfer raw data. Recently, the success of the large language model INTELLECT-1, trained across 30 clouds distributed globally, provides valuable insights into this novel paradigm. Additionally, techniques such as multi-agent reinforcement learning and vision-and-language navigation can enhance intelligent task planning and inference. However, current research on collaborative architectures in remote sensing remains insufficient to fully support generalist EO intelligence. Challenges such as heterogeneity and trustworthiness continue to hinder effective collaboration. In practice, collaborative architectures can be integrated throughout the lifespan of RSFMs. In this commentary, we outline potential research directions in two key phases: cross-cloud collaborative training and collaborative inference of RSFMs for generalist EO intelligence (Figure 1).

CROSS-CLOUD COLLABORATIVE TRAINING OF RSFMs

Cross-cloud collaborative training of RSFMs enables cloud platforms to conduct local training and exchange model parameters or intermediate results, thereby eliminating the need to transfer raw data. Federated learning and split learning are privacy-preserving distributed collaborative technologies. Specifically, they can empower RSFMs in two stages: self-supervised collaborative pre-training and cloud-personalized collaborative fine-tuning. Local knowledge derived from private remote sensing data on cloud platforms can be aggregated and integrated with global expertise from public data. Furthermore, cloud platforms can fine-tune task-specific models based on pre-trained RSFMs to deliver more personalized services.

Challenges and future perspectives

Performance degradation induced by mixed heterogeneity. Highly mixed heterogeneity manifests in data volumes, modalities, temporal variations, geographic distributions, model architectures, and computility across cloud platforms. The unification across three axes (i.e., architectures for multi-modal data, pre-training objectives, and tasks) may mitigate it. Furthermore, adaptive collaboration methods (e.g., model fusion) can be developed at the data, model, and cloud levels.

Time-consuming and resource-intensive communication. Frequent transmission of massive parameters in RSFMs across cloud platforms incurs high communication overhead, especially with limited bandwidth. Employing advanced gradient compression (e.g., joint sparsification and quantization) and communication scheduling (e.g., asynchronous communication and bandwidth-aware scheduling) methods can help accelerate this process.

Concerns about trustworthiness. The rise of generative models has exacerbated this issue, necessitating the establishment of robustness against attacks. Byzantine-robust federated learning can help identify and defend against untrustworthy collaborators based on model updates, while differential privacy can prevent others from inferring raw data from gradient information by adding controlled noise. Evaluating participant contributions for fairness using methods like Shapley values and enhancing interpretability are also crucial avenues for improving trustworthiness.

COLLABORATIVE INFERENCE OF RSFMs FOR GENERALIST EO INTELLIGENCE

Collaborative inference of RSFMs involves the cooperation of multiple institutions and agents. The future of RSFMs toward generalist EO intelligence is centered around three key areas of collaborative inference: cross-cloud, cloud-edge, and cross-edge collaboration.

Cross-cloud collaborative inference facilitates global high-resolution dynamic mapping by pooling resources and integrating data across institutions. Institutions utilize their own data and computility to map their assigned zones, collaboratively contributing to creating global high-resolution maps. This method embodies zonal intelligence, overcoming previous limitations in storage, computility, and data availability.

Cloud-edge collaborative inference addresses the limitations of edge devices in complex environments by leveraging powerful RSFMs on cloud platforms. For example, during an emergency response in a disaster, drones equipped with vision-and-language navigation communicate with the cloud using images or text. The cloud then provides route and task instructions. When unfamiliar

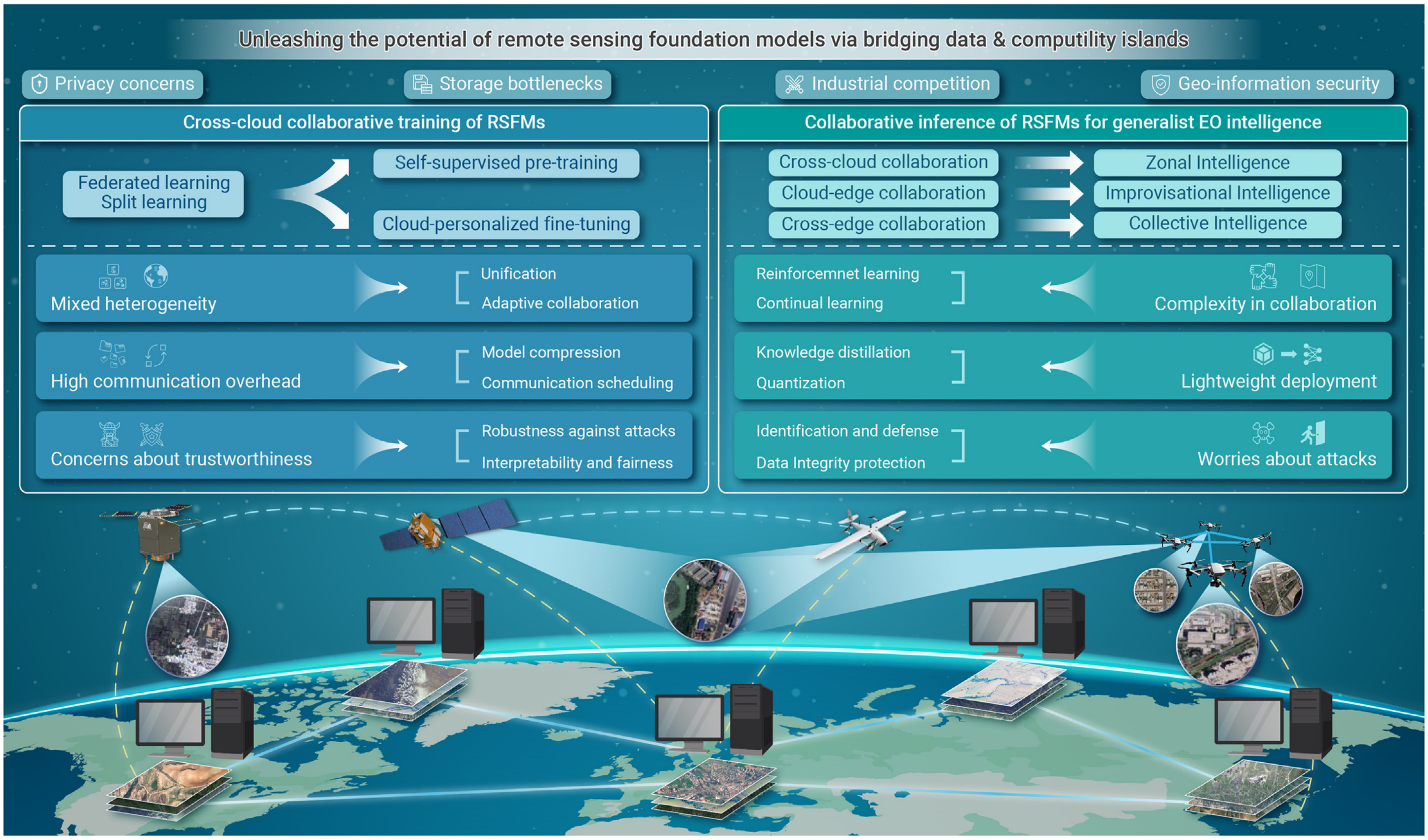

Figure 1. The framework for cross-cloud collaborative training and collaborative inference of RSFMs for generalist EO intelligence

scenarios arise, such as debris blocking the path, drones transmit images to the cloud's RSFMs for human-machine collaborative interpretation. The cloud then assists drones by providing few-shot samples, which are used for prompt-based inference on drones. This collaborative process enables improvisational intelligence.

Cross-edge collaborative inference enables space-air-ground integrated collaboration by leveraging the strengths of a multi-agent system.[5] It involves multiple edge devices (e.g., drones and satellites), each contributing through autonomous planning and collaborative perception. By fusing data from different time points, spatial resolutions, spectral ranges, and sensor observation angles, collaboration among these agents improves the accuracy and adaptability of EO, demonstrating the power of collective intelligence.

Challenges and future perspectives

Complexity in collaboration. Optimizing the scheduling process is crucial for ensuring efficient task allocation, resource utilization, and seamless coordination. Communication limitations pose challenges to maintaining real-time synchronization. Variations in altitudes and trajectories of devices increase the complexity of the georeferencing process. Reinforcement learning can autonomously manage task scheduling. Continual learning on edge devices enhances the evolution of collaborative systems and enables real-time, adaptive improvements in resource-limited environments.[5]

Lightweight deployment. Current edge devices face significant constraints in processing power, memory capacity, and energy efficiency, making it challenging to deploy RSFMs. For example, the basic version of SkySense incorporates 2.06 billion parameters, which requires professional graphics processing units (GPUs) in servers for inference.[2] Techniques such as knowledge distillation and quantization can be employed. In particular, they should adapt to the capability for processing multi-modal data and handling the large-size nature of remote sensing images.

Worries about attacks. Security concerns, including adversarial samples, data poisoning, and backdoor attacks, are pervasive in collaborative inference. The communication process within the collaborative system involves the transmission of images and model parameters. Therefore, identifying and defending against these threats is critical, particularly in black-box scenarios. Ensuring data integrity protection is essential for maintaining system security.

CONCLUSION

Recently, RSFMs have demonstrated significant potential in advancing generalist EO intelligence. However, the persistent challenges of data and computility islands severely restrict the full potential of RSFMs. Therefore, we strongly advocate for global collaboration among institutions to construct and develop RSFMs through cross-cloud, cloud-edge, and cross-edge collaboration, thereby establishing a collaborative architecture to bridge these islands. Future research should focus on developing collaborative architectures at both hardware and software levels to ensure seamless integration. Ultimately, collaborative approaches will enhance the scalability and efficiency of RSFMs, fostering deeper insights and understanding in EO research and unleashing the potential of RSFMs for generalist EO intelligence.

REFERENCES

1. Xu, Y., Wang, F., An, Z. et al. (2023). Artificial intelligence for science-bridging data to wisdom. *Innovation* **4**:100525.
2. Guo, X., Lao, J., Dang, B. et al. (2024). Skysense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (IEEE), pp. 27672–27683.
3. Zhao, T., Wang, S., Ouyang, C. et al. (2024). Artificial intelligence for geoscience: Progress, challenges, and perspectives. *Innovation* **5**:100691.
4. Wang, F., Yao, D., Li, Y. et al. (2023). Ai-enhanced spatial-temporal data-mining technology: New chance for next-generation urban computing. *Innovation* **4**:100405.
5. Liu, R., Diao, B., Huang, L. et al. (2024). Continual learning in the frequency domain. In Advances in Neural Information Processing Systems, pp. 85389–85411.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under grants 42030102 and 42371321 and by the Ant Group.

DECLARATION OF INTERESTS

The authors declare no competing interests.