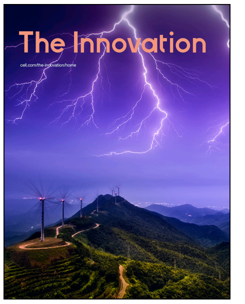
Journal Pre-proof



Artificial intelligence-assisted remote sensing observation, understanding, and decision

Han Xu, Chen Lu, Hao Zhang, Zhenfeng Shao, Guangcan Liu, Jiayi Ma

PII: S2666-6758(25)00219-X

DOI: https://doi.org/10.1016/j.xinn.2025.101016

Reference: XINN 101016

To appear in: *The Innovation*

Received Date: 19 April 2025

Revised Date: 29 June 2025

Accepted Date: 1 July 2025

Please cite this article as: Xu, H., Lu, C., Zhang, H., Shao, Z., Liu, G., Ma, J., Artificial intelligence-assisted remote sensing observation, understanding, and decision, *The Innovation* (2025), doi: https://doi.org/10.1016/j.xinn.2025.101016.

Artificial intelligence-assisted remote sensing observation, understanding, and decision

Han Xu[1], Chen Lu[1], Hao Zhang[2], Zhenfeng Shao[3], Guangcan Liu[1], and Jiayi Ma[2,*]

[1]School of Automation, Southeast University, Nanjing 210096, China

[2]Electronic Information School, Wuhan University, Wuhan 430072, China

[3]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

[*]Correspondence: jyma2010@gmail.com (J.M.)

Remote sensing underpins environmental monitoring and Earth science. Recent expansion of satellites and observation platforms drives a substantial increase in multi-source remote sensing data. The integrated land-air-space multi-sensor stereoscopic observation heralds a new era of intelligent photogrammetry and digital infrastructure. However, the inherent complexity of multi-source data (spanning spatial, spectral, and temporal domains) poses significant challenges for observation, interpretation, and understanding. [1] The rapid advancement of artificial intelligence (AI) injects new vitality into the intelligent interpretation of remote sensing by reshaping systems: from overcoming imaging limitations through enhanced visual observation, to elevating knowledge dimensions via semantic understanding, and ultimately enabling intelligent decision-making (Figure 1). This commentary examines how AI enhances visual observation, facilitates the semantic transition from pixels to knowledge, and empowers intelligent decision-making. These advancements provide theoretical and methodological support for the paradigm shift from data acquisition to cognitive services.

AI-ENHANCED VISUAL OBSERVATION IN REMOTE SENSING IMAGERY

Remote sensing imagery fundamentally differs from natural imagery due to disparities in sensor-specific geometric configurations, physical radiation mechanisms, and imaging platforms. These discrepancies manifest as variations in spatial/spectral resolution, radiometry, and viewing geometry, necessitating integrating multi-source data for comprehensive observation. For example, registering and fusing synthetic aperture radar (SAR) and multi-spectral data enhances disaster response. SAR delivers weather-robust monitoring of structural deformation, while multi-spectral data detects vegetation stress, floods, and fires. The reliability of multi-source data can be ensured by physics-guided quality assessment and AI-embedded screening. Then, physics-informed AI can harmonize multi-sensor data by leveraging underlying physical priors to guide registration and fusion. However, these multifaceted variances complicate model training, necessitating AI models that structurally adapt to highly heterogeneous data. Deep models enable cross-modal invariant representation and complex geometric transformations for robust registration and derive comprehensive feature extraction and complementary fusion. They jointly underpin visual perception for downstream semantic understanding and decision.

AI facilitates multi-source data registration by extracting cross-modal descriptors and correcting geometric transformations. First, deep models, particularly large foundation models pre-trained on massive diverse data with extensive adaptive capabilities, can learn modal-invariant descriptors. These descriptors transcend sensor-specific characteristics, simplifying nonlinear cross-modal mapping and facilitating the establishment of registration metrics and deformation estimation. Second, with gradient-descent framework, AI overcomes challenges of non-convex and high-dimensional optimization. [2] It enables complex model construction and solution strategies to improve transformation correction accuracy and robustness. Networks can bypass constrained transformation models by parameterizing deformation fields into pixel-level fields with physical regularization. This approach empowers comprehensive solution spaces for high-fidelity modeling and solution of complex transformations.

Figure 1: The framework for remote sensing data observation, understanding, and decision

Multi-source fusion faces two fundamental challenges: managing heterogeneous data discrepancies and optimizing integration process. AI assists in deep, comprehensive representation learning across cross-modal data, [3] prerequisites critical for effective fusion. It can disentangle modal-specific attributes while preserving spatial fidelity and spectral discriminability. It avoids feature entanglement inherent in oversimplified, manual prior-dependent representations. Regarding integration, learning-based frameworks can supplant empirical parameterization with end-to-end optimization. The incorporation of comprehensive multi-dimensional constraints and interactive mechanisms enables self-calibrating cross-modal feature embeddings, realizing collaborative multi-dimensional optimization. Critically, data-driven generalization ensures robust integration of diverse modalities under varying conditions. These advancements position AI as a pivotal enabler for generating high-quality fused data.

FROM PIXELS TO KNOWLEDGE: AI-DRIVEN SEMANTIC UNDERSTANDING OF REMOTE SENSING DATA

Remote sensing imagery provides rich visual data, yet actionable insights require higher-level interpretation. Semantic understanding converts raw pixels into geospatial knowledge, advancing from observation to understanding. A framework for comprehensive geospatial understanding requires integrated “object-scene-system” cognition. Object-level semantic parsing encounters challenges in attribute inference to differentiate diverse ground objects. Scene-level semantic association necessitates establishing semantic relationships among diverse categories, enabling holistic interpretation of spatial interactions. System-level semantic evolution extends analysis to multi-temporal dynamics, interpreting evolving processes.

Object-level semantic parsing leverages AI to bridge the gap between low-level visual features and high-level semantic knowledge through intrinsic feature construction and multi-granularity attribute inference. Annotating remote sensing data is labor intensive and requires domain expertise, creating significant bottlenecks. Semi-supervised learning offers partial mitigation of these challenges. By integrating attributes from high-spatial and high-spectral resolution data, deep models construct discriminative and expressive intrinsic semantic features. Deep models—particularly foundation models powered by in-context learning, emergent reasoning, and cross-domain generalization-enable multi-granularity attribute inference. It addresses ambiguity arising from spectral heterogeneity within identical classes or spectral homogeneity across distinct classes.

At the scene level, AI facilitates semantic association through relational inference and knowledge embedding. Graph neural networks with adaptive edge weights infer heterogeneous relationships via semantic graphs, revealing spatial and functional dependencies among objects. Concurrently, knowledge-guided semantic constraints integrate domain-specific knowledge graphs into deep models. This integration encodes geographic priors as structured graph representations, guiding models to learn coherent semantic logic and resolve contextual ambiguities.

At the system level, semantic evolution leverages AI to achieve advanced cognition through architectures and causal interpretability, i.e., identifying and quantifying input-decision relationships. Multidimensional spatio-temporal mechanisms bridge short-term dynamics and long-term trends, enabling comprehensive characterization of intricate geospatial processes.[4] Interpretable causal mechanisms integrate two complementary approaches: causal saliency analysis quantifies causal attribution through counterfactual gradient propagation and dynamic causal graph evolution employs temporal graph convolutional networks to model time-varying causal topologies. This framework deciphers latent driving forces and supports decision-making through evolutive causal reasoning.

Visual observation and semantic understanding are intrinsically interdependent, enabling collaborative geospatial analysis. High-fidelity visual features provide accurate inputs for robust semantic understanding. Concurrently, semantic understanding guides visual models to focus on regional relevance and maintain semantic consistency across constituent regions for refinement.

FROM KNOWLEDGE TO INTELLIGENCE: AI-ARCHITECTED GEOSPATIAL DECISION

Geospatial decision-making establishes data-knowledge-intelligence closed loops operating in dynamic environments and cross-platform collaboration. AI enables a shift from rigid paradigms to flexible decision frameworks that integrate autonomous and collaborative intelligence, transforming semantic insights into actionable knowledge, e.g., turning geospatial observations into proactive geological decisions. For instance, in flood monitoring, AI fuses multi-source data to generate near-real-time risk maps and trigger automated alerts. For seismic response, AI-enhanced sensing enables rapid ground motion assessment with dynamic rerouting of critical logistics. In landslide or mining areas, predictive models forecast slope instability to trigger preemptive shutdowns or diversions. These implementations reveal how decision layers grounded in multi-source semantic understanding and adaptive intelligence drive autonomous, context-sensitive action planning.

Adaptive decision-making for emergencies and non-stationary environmental evolution—driven by sensor-derived distribution shifts—propels autonomous system development. AI drives the transition from conventional static processing to dynamic closed-loop geospatial intelligence systems. The core involves joint perception-decision learning rather than treating them as isolated tasks. It integrates semantic knowledge with executable policies through adaptive adversarial reward mechanisms. These reward strategies incorporate multi-constraint optimization, encompassing geospatial knowledge priors, semantic perception, and decision-making objectives. This triad facilitates knowledge-guided perception-decision integration and enables self-improving learning. Concurrently, online decision evolution mechanisms coupled with incremental meta-learning architectures rapidly assimilate multi-sensor remote sensing data. Historical decision feedback loops dynamically refine policies by incorporating spatiotemporal pattern recognition, enabling continuous adaption to non-stationary geospatial state evolution. Collectively, these advances establish robust, self-optimizing autonomous decision frameworks for complex and evolving remote sensing environments.

The land-air-space collaborative sensing and AI's paradigm-shifting impetus drive the transformation of decision architectures-from isolated system to cross-platform collaborative frameworks. This evolution enables efficient remote sensing data sharing, secure multi-platform sharing, and distributed decision interoperability. Collaborative intelligence is operationalized through distributed federated learning frameworks, facilitating optimized resource allocation across platforms. By integrating physics-informed multi-agent systems, geospatial knowledge is encoded into reward functions that steer decentralized policy optimization. Further integration of physics-informed federated learning with game-theoretic equilibrium frameworks enables dynamic equilibrium strategies. This synthesis resolves resource competition while ensuring decision consistency across platforms, achieving optimal multi-platform coordination. This approach shifts from isolated decision-making toward robust federated equilibrium, supporting efficient collaboration, scalable model deployment, and secure knowledge transfer across ecosystems.

FUTURE TRENDS

Despite AI's acceleration of remote sensing analysis, sensor-data dependency remains a bottleneck. Future systems must integrate cross-domain knowledge spanning environmental science, socio-economics, and related fields. Incorporating textual and leveraging cross-modal alignment in foundation models enables absorption, transfer, and reasoning over cross-domain knowledge pertinent to remote sensing tasks. It enhances generalization and adaptability to unseen scenarios and tasks. Domain gaps can be mitigated by combining remote sensing characteristics with large-model capabilities via adapter-based architectures and unifying multi-modal features through projection, contrastive learning, or graph alignment. Moreover, current AI-driven remote sensing remains constrained by data-dependent paradigms rooted in historical correlations. This limitation underscores the need for cognitively adaptive systems featuring biologically inspiration, interactive learning, cognitive evolution, and dynamic human-machine interaction.

The black-box nature of AI raises interpretability and security concerns,[5] conflicting with the verifiability of remote sensing physical mechanisms. Developing interpretable architectures is essential to clarify processing and decision pathways, and quantify model behavior risks. Potential directions include embedding domain knowledge or physics-informed constraints, implementing inherently transparent structures, or designing post-hoc explanation methods. Addressing dynamic data drift from environmental shifts and adversarial attacks necessitates real-time defense strategies and security mechanisms. Integrating detection systems with adaptive updating or online learning enhances reliability. Co-designed architectures incorporating adversarial training, uncertainty estimation, and redundancy strategies are critical for reliability. These advancements enable intelligent remote sensing with high accuracy, transparency, and resilience.

References

1 Bian, J., Li, A., Nan, X. et al. (2025). Stereoscopic remote sensing observation for mountains: Advancing monitoring, modeling, and management. Innov. *6*, 100838. doi: 10.1016/j.xinn.2025.100838.

2 Ekvall, M., Bergenstråhle, L., Andersson, A. et al. (2024). Spatial landmark detection and tissue registration with deep learning. Nat. Methods *21*, 673–679. doi: 10.1038/s41592-024-02199-5.

3 Hong, D., Zhang, B., Li, X. et al. (2024). Spectralgpt: Spectral remote sensing foundation model. IEEE Trans. Pattern Anal. Mach. Intell. *46*, 5227–5244. doi: 10.1109/tpami.2024.3362475.

4 Zheng, Z., Ermon, S., Kim, D. et al. (2025). Changen2: Multi-temporal remote sensing generative change foundation model. IEEE Trans. Pattern Anal. Mach. Intell. *47*, 725–741. doi: 10.1109/tpami.2024.3475824.

5 Mehdi, S., and Tiwary, P. (2024). Thermodynamics-inspired explanations of artificial intelligence. Nat. Commun. *15*, 7859. doi: 10.1038/s41467-024-51970-x.

FUNDING AND ACKNOWLEDGMENTS

This work was funded by the National Natural Science Foundation of China (U21B2027), Natural Science Foundation of Jiangsu Province (BK20241280), and Start-up Research Fund of Southeast University (RF1028623006 and RF1028624061).

DECLARATION OF INTERESTS

The authors declare no competing interests.


AI-Enhanced Visual Observation
AI-Driven Semantic Understanding
AI-Architected Geospatial Decision
Multi-Source Heterogeneous Data
Registration
Transformation?
Fusion
AI
Object-Level Semantic Parsing
Hilly Terrain
Farmland
Lake
Village
Scene-Level Semantic Association
System-Level Semantic Evolution
Time
Autonomous Intelligence
Collaborative Intelligence
From Pixels to Knowledge
From Knowledge to Intelligence