Geospatial remote sensing interpretation: From perception to cognition

Sheng Wang,[1] Wei Han,[1] Xiaohan Zhang,[1] Jun Li,[1] and Lizhe Wang[1,*]
[1]School of Computer Science, China University of Geosciences, Wuhan 430078, China
*Correspondence: lizhe.wang@foxmail.com

Received: October 31, 2023; Accepted: March 8, 2024; Published Online: March 14, 2024; https://doi.org/10.59717/j.xinn-geo.2024.100056

Citation: Wang S., Han W., Zhang X., et al., (2024). Geospatial remote sensing interpretation: From perception to cognition. The Innovation Geoscience **2(1)**: 100056.

Geographical space is a vast and complex system involving multiple elements and non-linear interactions of these elements, and rich in geographical phenomena, processes and patterns. Deep learning (DL) approaches are increasingly applied to extract useful information, patterns, and insights from the ever-growing stream of geospatial data, but DL-based Remote Sensing (RS) interpretation often struggles to precisely perceive and recognize complex geographical spaces when the system behavior is dominated by the spatiotemporal context. Rather than duplicated development of interpretation models, insufficient mining of geographic patterns, or isolated applications of geographic information extraction, we argue that data-driven RS intelligent interpretation requires the guidance of geographic ideas and a unified architecture.

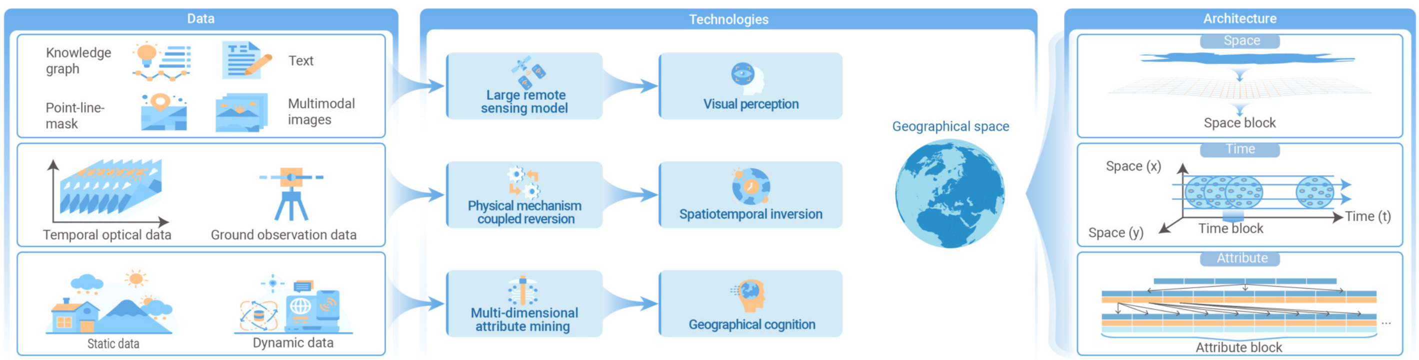

Figure 1. The architecture and technologies of the cognitive remote sensing.

To overcome these limitations, we introduce a new geospatial RS interpretation from perception to cognition (i.e., cognitive RS) building on a unified multi-scale "space-time-attribute" architecture. Its outcome contains a multi-scale RS information base and a knowledge base that facilitates an understanding of the morphology, types, and indicators of complex geospatial (visual perception) and a perspective of the structure, evolution, and trends (geographical cognition), respectively. Through an integrated RS interpretation architecture with multiple scales and stages (including visual perception of geographical phenomena, spatiotemporal inversion of geographical processes, and geographical patterns cognition), the cognitive RS can obtain high-precision and trustworthy geospatial information, and mine geospatial distribution patterns, evolutionary mechanisms, and dynamical trends, thereby serving geospatial thematic applications such as land use and land cover (LULC), urban spatial planning, and quantitative agricultural RS.

ARCHITECTURE OF THE COGNITIVE RS

Existing research has demonstrated that the process understanding and hybrid modeling of complex geographic space urgently requires the coupling of physical mechanisms and geographical knowledge.[1,2] Although the new RS technologies with three-dimensional observation, real-time perception, and spatiotemporal synergy have been able to realize all-around fine observation of the earth's surface environment,[3] precise understanding and functional perspectives of geographical space still lack a unified architecture. In addition, traditional RS interpretation units of pixels and objects are difficult to correspond to the real geographic entities with geographic attributes, which exacerbates the spatial and temporal structural differences between the geospatial RS interpretation and the geographic cognition of human beings. To bridge this gap, we introduce geo-parcels with geographic attributes to correlate the RS image space to the geographic information space. Distinguished from the object-oriented approach, we define a geo-parcel as the smallest geographic entity in a geographic space that is visually perceivable and has a clear land use/land cover type under certain spatial scale constraints.

Within the spatial constraints of structured geo-parcels, through shallow visual perception and deep pattern mining, the objective of the cognitive RS is to achieve a precise understanding of the morphology, types, and indicators of geographic entities, as well as the functional perspective of the structure, evolution, and trends of geographic space. In Figure 1, considering that geographic space in the context of multi-source heterogeneous big data observations is a complex system,[4] we construct a unified "space-time-attribute" architecture to model complex geographic spaces at different levels, providing the foundation for visual perception, spatiotemporal inversion, and geographic cognition.

Spatial structure

The spatial structure aims to take high-resolution RS images as the basis, extract irregular geo-parcels as spatial constraints through visual perception algorithms, and construct a multi-level spatial architecture of pixel-parcel-scene to realize bottom-up scale hierarchical mapping of the complex geographic space. The morphological boundaries of the geo-parcels are the external manifestations of the spatial scale mapping results.

Time structure

Under the impression of human activities, geological effects, climate change, and other factors, the observation data in the geographic entities are in a dynamic changing process. When the spatial structure is relatively stable, the time-series changes in the feature values of the geographic entities' attributes are due to the changes in the distribution of data on each time block. Therefore, based on time-series satellite data and sensor monitoring data, the spatiotemporal inversion and attribute feature fusion of geographic entities are crucial for exploring the geospatial temporal changes in the time dimension by constructing a multi-level time structure such as "year-month-day". The characterization data of the time block can be calculated as single-moment values, or as the mean, standard deviation, maximum, and minimum values of the observed values over a period of time.

Attribute structure

After the visual perception of geographical phenomena and spatiotemporal inversion of geographical processes, geo-parcels are enriched with attribute data, including morphology, types, and indicators. Multi-source observation data can form a series of attribute values after aggregating on the parcel units with different spatial block sizes, which helps to mine the correlation relationships and potential patterns within the geographical space. The multi-dimensional attribute table of geographical space is obtained by calculating the mean, multitude, maximum, minimum, and histogram of the attribute data under the spatial constraints of the geo-parcels.

TECHNOLOGIES OF THE COGNITIVE RS

The emergence of an advanced large RS model[5] (also called the pre-training foundation model) greatly improves the visual perception ability of complex geographical space, with the characteristics of intelligence, generalization, and transportability. With the coupling of physical mechanism, the spatiotemporal inversion can be able to support the integrated inversion of multi-source and multi-temporal RS data, improve the accuracy of the index inversion, and reduce the uncertainty. In addition, considering various types of static (such as natural resource endowment and socioeconomics data) and dynamic data (such as human activities and social production data), constructing multi-dimensional dynamic attribute tables for structured geographic entities and carrying out pattern mining are the keys to penetrate the pattern, evolution, and trend of complex geographical space.

Visual perception

Towards the needs of geospatial perception for accuracy, generalization, and portability, a multimodal large RS model is proposed to obtain category-independent segmentation maps of RS scenes.To take full advantage of the generalization and zero-shot capabilities of the visual foundation model, a general Transformer architecture is employed for training. Point-line-mask prompt data, multimodal RS images, structured and unstructured text, and a generic RS knowledge graph are introduced through multimodal prompt learning. Considering the spatial autocorrelation in the first law of geography, these multimodal data are organized and aligned for model training. Based on this, for different geographic scenarios, the constructed large RS model is introduced for scene segmentation to obtain structured geo-parcels.

Spatiotemporal inversion

In response to the problem that single-temporal RS data cannot reflect the changing patterns of surface objects, a spatiotemporal computational inversion method for the geo-parcels is proposed. Under the boundary constraint of fine geo-parcel units, the RS data with synergistic high temporal resolution can discriminate complex parcel types and quantitatively invert relevant parameter indicators.Multimodal time-series spectral RS data, ground observation data, and phenological evolution mechanism are fused to perform spatiotemporal inversion to obtain an information base of geo-parcel, including various attribute information such as land cover types and quantitative indicators. These indicators can be further applied to agriculture and other fields, such as crop generation detection, planting structure assessment, and crop planting planning.

Geographical cognition

To achieve comprehensive cognition of complex natural scenes, it is highly demanding to mine the evolution rules and spatial patterns of various geospatial scenes. After obtaining geo-parcels with semantic, attribute, and functional information (e.g., morphology, topography, climate, soil, etc.), geo-parcels are further aggregated through correlation mining of the spatial topology, semantic and attribute to construct structured geographic scenarios. In addition, the static attribute table of the geo-parcels is reorganized to infer implied parcel attributes. The dynamic time-series attribute table is constructed to simulate the rapid evolution of the geographic state under the coupled internal and external dynamics, to predict future development trends and guide current scientific decisions. Then the spatiotemporal evolution laws can be parsed through Markov chains, Bayesian, and decision trees. Finally, the results of historical scene cognition are stored in the knowledge base.

APPLICATIONS OF THE COGNITIVE RS

Land use and land cover

Limited by the traditional RS interpretation units, the smallest image units of pixels or objects are difficult to correspond to real geographic entities in terms of morphology and attributes. Thus, for a long time, LULC products have always coexisted in a carrier (pixel or object) in an ambiguous and confusing way. By introducing multimodal RS data and advanced cognitive RS architectures, we aim to achieve precise LULC classification and comprehensive understanding of complex geospatial patterns.

Urban spatial planning

The variety of surface elements, complex spatial distribution, and rapid changes have exacerbated the difficulty of urban spatial planning. The traditional single-stage, single-scale, and pixel-or object-level RS interpretation is hard to realize the quantitative expression of multi-scale urban spatial information such as space, time and attributes. Therefore, the introduction of the multi-scale cognitive RS architecture is the most promising solution for perception, cognition and decision-making in urban space.

Quantitative agricultural RS

Farmland is a crucial geographic element, but traditional RS interpretation units often struggle to characterize complex and irregular farmland boundaries. Considering the visual perception, spatiotemporal inversion, and geographical cognition capabilities of cognitive RS, parcel-based quantitative agricultural RS is the most promising solution in refined cropland boundary extraction, crop growth parameters inversion, and crop planting adaptability evaluation.

CONCLUSION

In this commentary, we describe the architecture and technologies of the cognitive RS, a new generation of geospatial RS interpretation with multi-scale in space, time, and attribute, which integrates geographic space perception, geographic process inversion, and geographic pattern mining. It aims to perceive the morphology, types, and indicators of geospatial entities and penetrate the distribution patterns, evolution mechanisms, and dynamics trends of geographical space. With the support of the large RS model, physical mechanism coupled inversion, pattern mining, the cognitive RS will promote the integrated perception and cognition of geographical space, serving comprehensive geospatial thematic applications, such as precise LULC, urban spatial planning, and quantitative agricultural RS.

REFERENCES

1. Reichstein, M., Camps-Valls, G., Stevens, B., et al. (2019). Deep learning and process understanding for data-driven Earth system science. Nature 566(7743): 195–204. DOI: 10.1038/s41586-019-0912-1.
2. Chen, M., Qian, Z., Boers, N., et al. (2023). Iterative integration of deep learning in hybrid Earth surface system modelling. Nat Rev Earth Environ 4(8): 568–581. DOI: 10.1038/s43017-023-00452-7.
3. Wang, L., Zuo, B., Le, Y., et al. (2023). Penetrating remote sensing: Next generation remote sensing for transparent Earth. The Innovation 4(6): 100519. DOI: 10.1016/j.xinn.2023.100519.
4. Wang, S., Han, W., Huang, X., et al. (2024). Trustworthy remote sensing interpretation: Concepts, technologies, and applications. ISPRS J. Photogramm. Remote Sens 209: 150–172. DOI: 10.1016/j.isprsjprs.2024.02.003.
5. Zhang, L., Zhang, L., Yuan, Q., et al. (2023). Large Remote Sensing Model: Progress and Prospects. Geomatics and Information Science of Wuhan University 48(10): 1574–1581. DOI: 10.13203/j.whugis20230341.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (41925007, U21A2013, 42201415), and the Open Research Program of the International Research Center of Big Data for Sustainable Development Goals (CBAS2023ORP03).

DECLARATION OF INTERESTS

The authors declare no competing interests.