Machine Learning for Earth Observation 2026

Domain Knowledge- and Large Model-Driven Remote Sensing-Based Intelligent Lithology Interpretation

Wei Han

June 22, 2026

China University of Geosciences (Wuhan)

Outline

1. Research background and challenges

2. Knowledge and LLM-driven lithology interpretation

3. LithoBench: Benchmarking LLMs for Lithology Interpretation

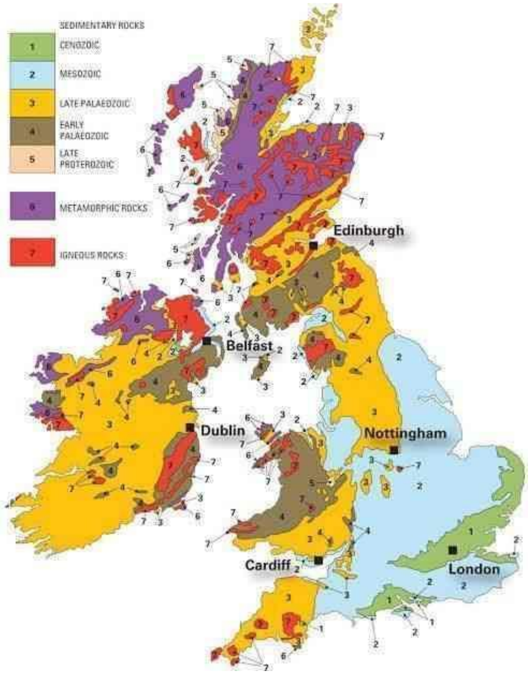
4. Future directions and conclusion

Research Background

Lithology is a fundamental geological component, it is very important for understanding geological conditions and supporting resource exploration, engineering construction and etc.

Remote Sensing enables fast, efficient and cost-effective identification of key geological elements and processes.

Geological survey



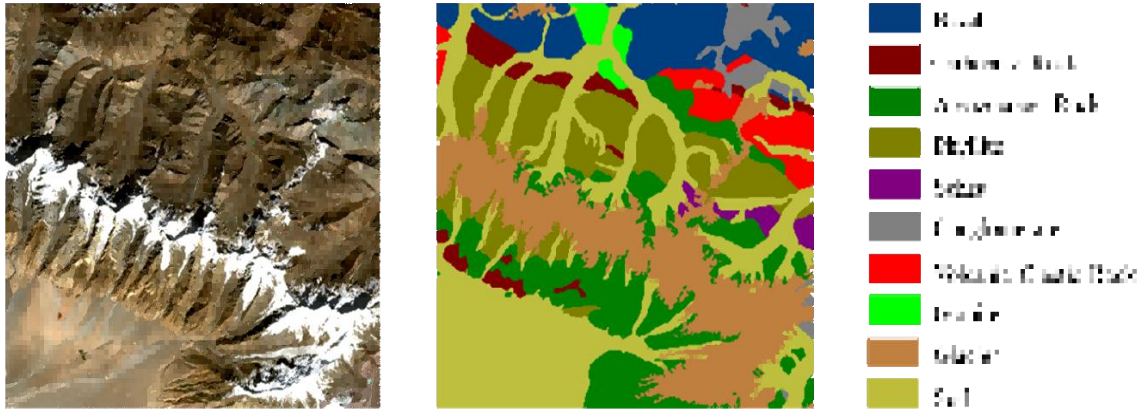
Geological mapping

Disaster early warning

Key challenges: (1) Intrinsic lithological ambiguity

(a) Different lithologies in the Gaofen 2 images

Many types
high inter-class similar

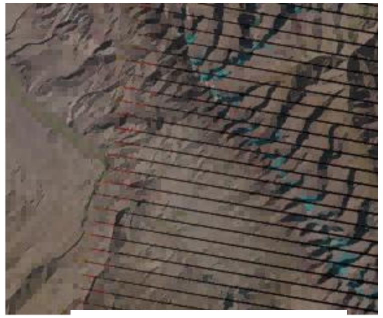


Complex distributions
sample imbalance

(b) Distributions of different types in a large region (Landsat 8 image)

Key challenges: (2) Complexity of Image Acquisition

Blur

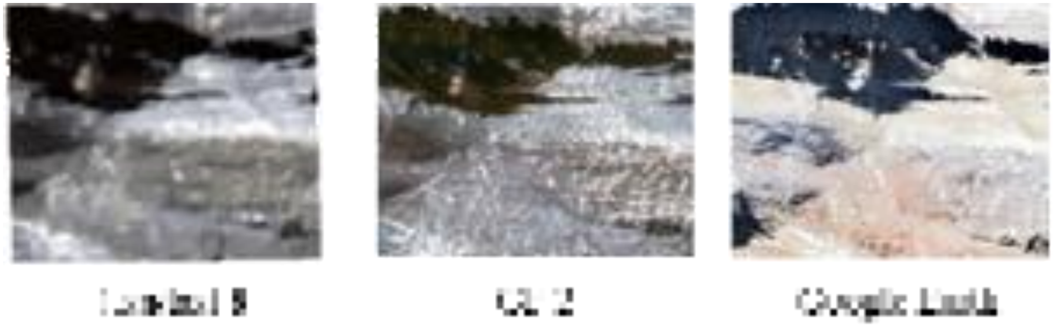


Striping noise

clouds

haze

Multiple noise

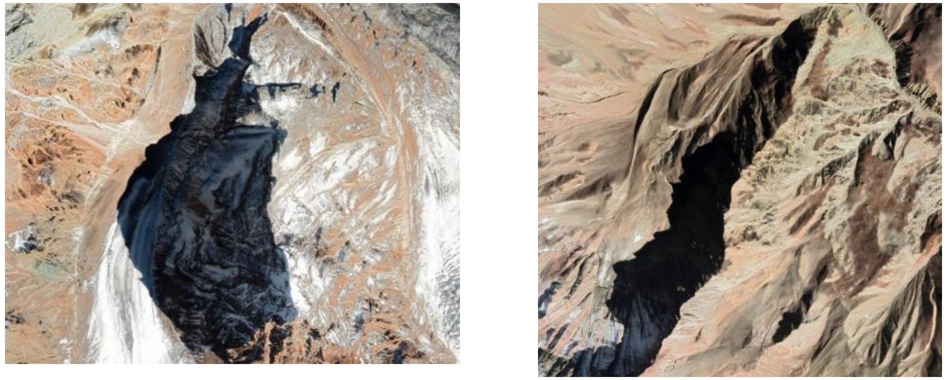


Carbonate rock

Slate

Differences across imaging sensors

Key Challenges: (3) Complex Background Conditions



(a) Topographic and geomorphic effects

Slate

Granite

(b) Vegetation cover

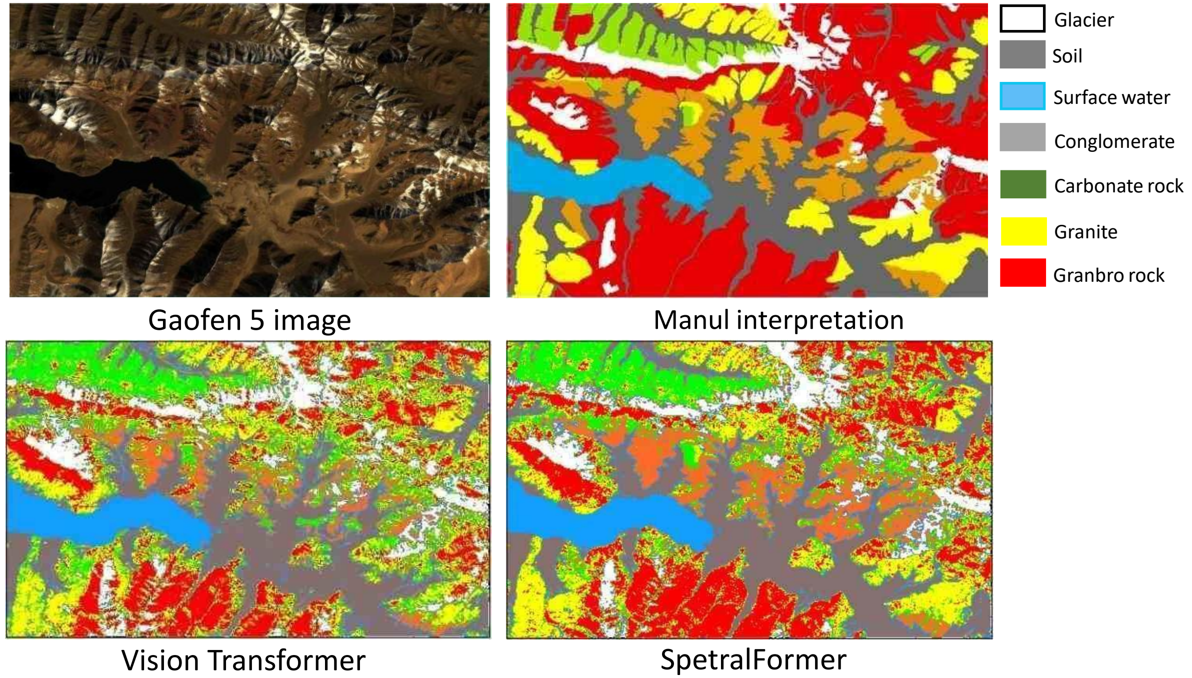
Results: Limited Identification Accuracy

(1) Intrinsic lithological ambiguity

(2) Complexity of image acquisition

(3) Complicated background

Terrible performance



Deep Learning models: fragmented, lack continuity
The overall mIoUs is lower than 50% and the accuracies is lower than 70%.
Manul interpretation: more consistent and high-quality.

Wei Han, et al., "A survey of machine learning and deep learning in remote sensing of geological environment: challenges, advances, and opportunities," ISPRS P&RS， vol. 202, pp. 87-113, 2023. (ESI hot paper)

Outline

1. Research background and challenges

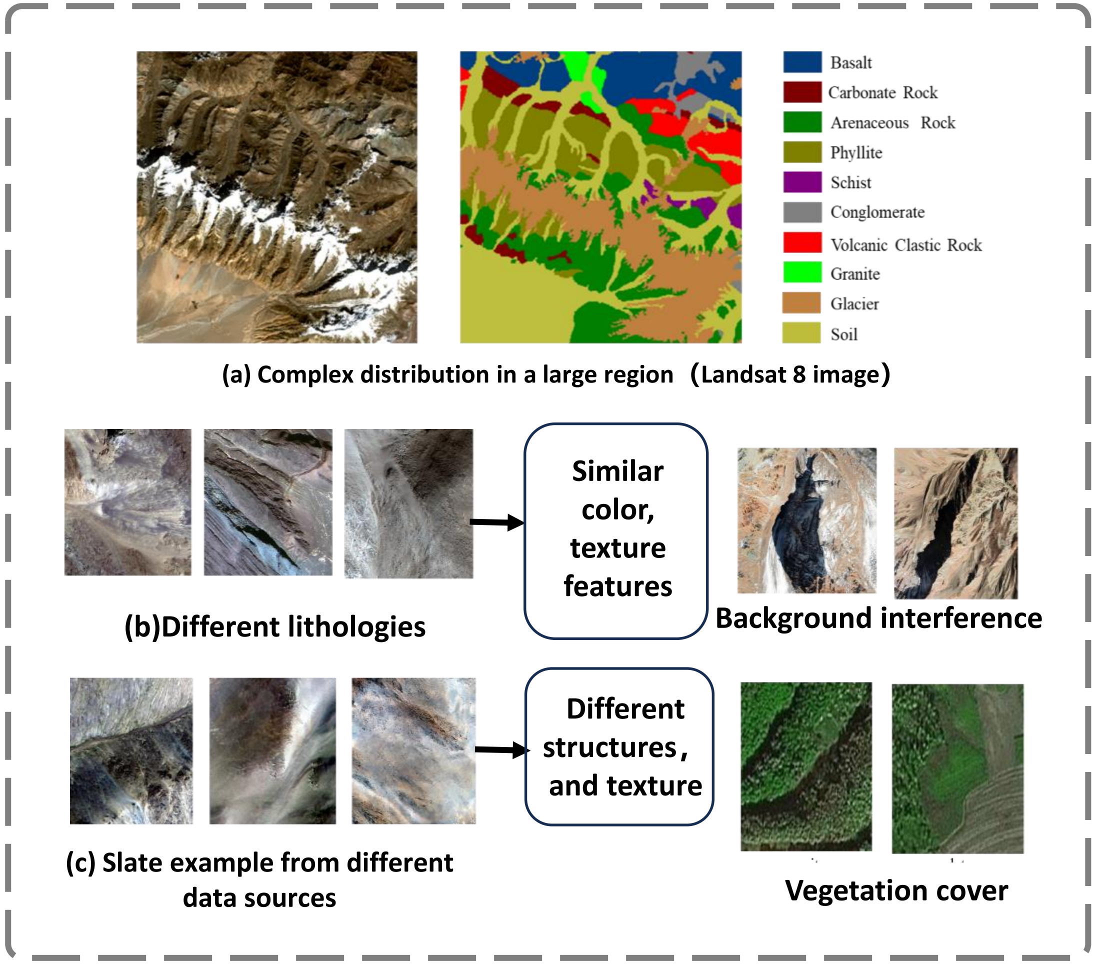
2. Knowledge and LLM-driven lithology interpretation

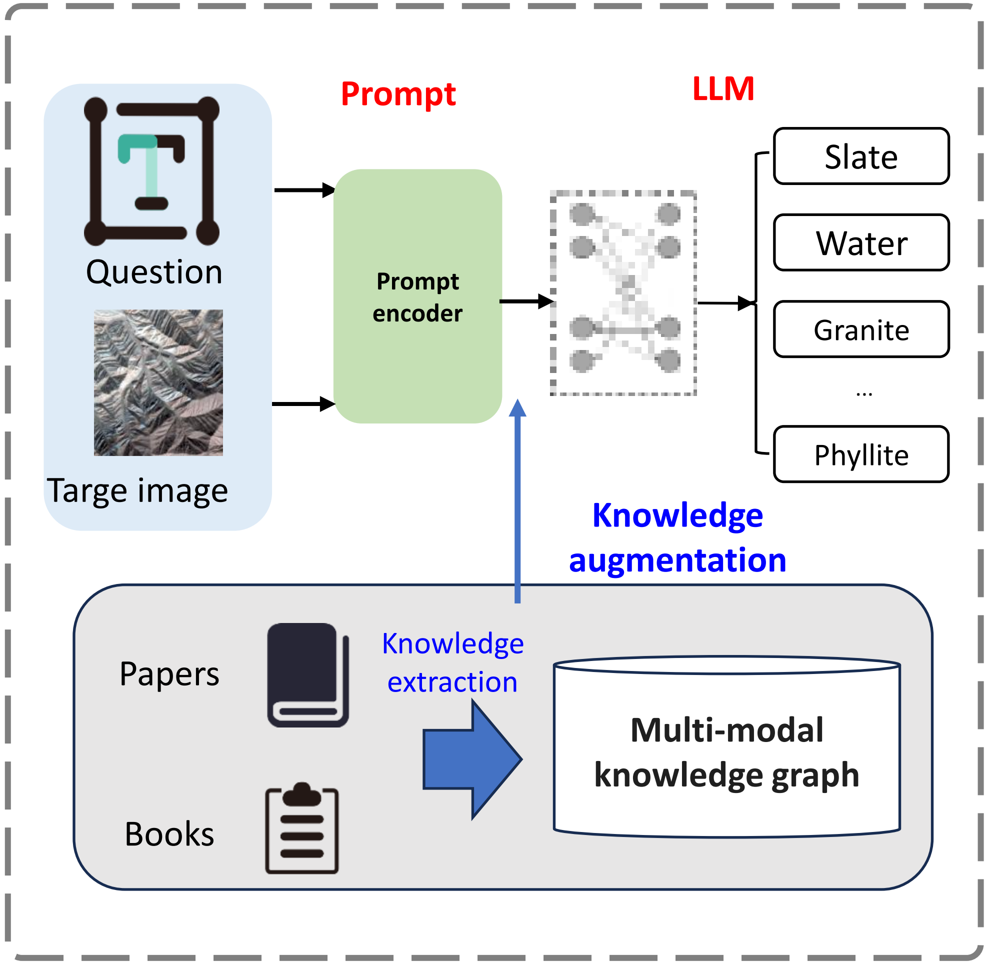
3. LithoBench: Benchmarking LLMs for Lithology Interpretation

4. Future directions and conclusion

Knowledge and LLM-driven Lithology Interpretation

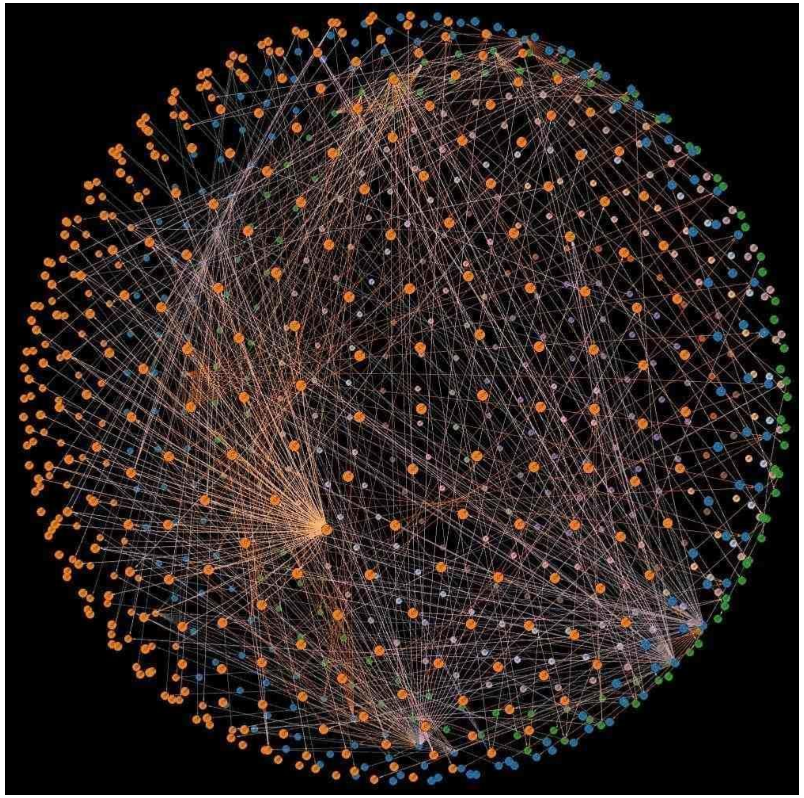
We construct a large-scale "lithology-mineral-spectrum" knowledge graph, and then establish a large model to identify lithology from remote sensing images.



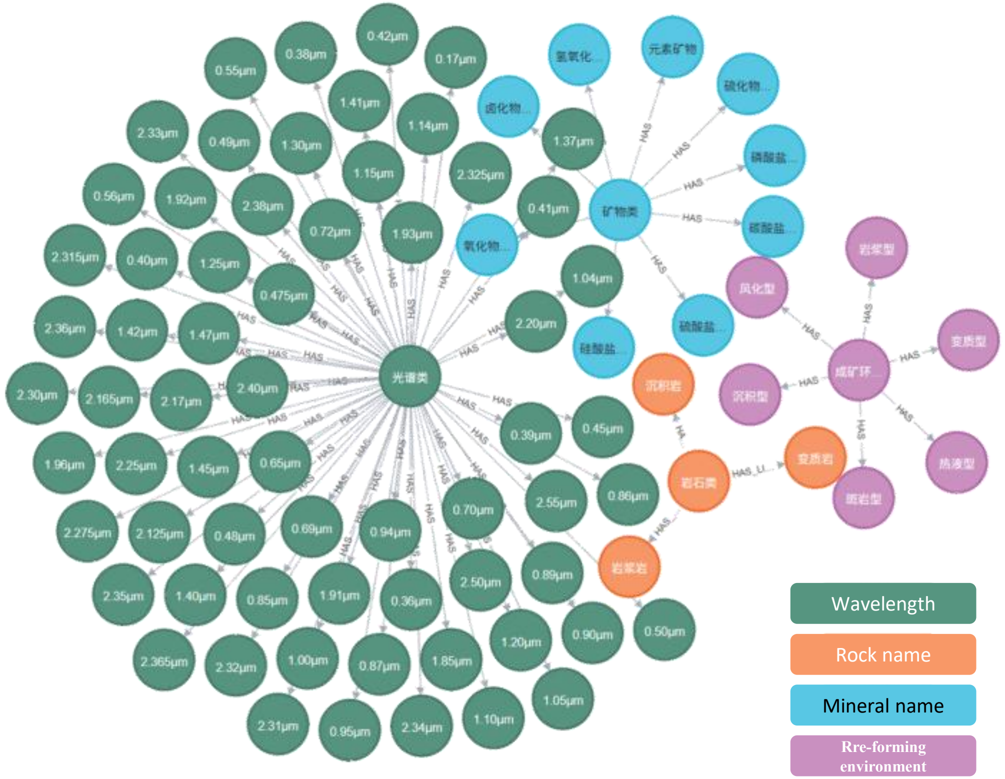


Knowledge and LLM-driven Lithology Interpretation

We built a novel, multimodal domain knowledge graph from more than 40k papers and 10 spectral databases, covering over 1000 types of minerals and rocks with their composition, visual features and spectral characteristics.



Knowledge graph visualization



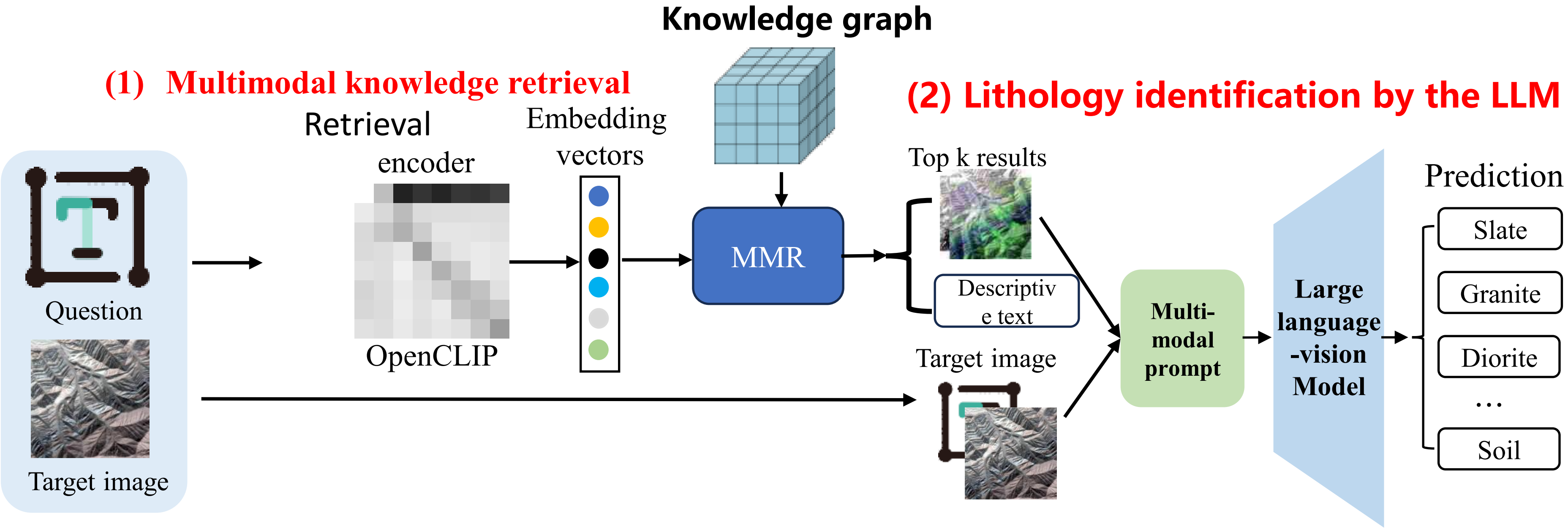
Four key entity types

Knowledge and LLM-driven Lithology Interpretation

Then, we developed a knowledge-driven large model by incorporating retrieval-augmented generation(RAG) method to interpret lithology.

① Input a question and an image
② Retrieve relevant knowledge

① Construct a multimodal prompt
② Use LLAVA-7B as the reasoning model



Chen X, Liu Y, **Han W***, et al. A vision-language foundation model-based multi-modal retrieval-augmented generation framework for remote sensing lithological recognition. ISPRS P&RS, 2025, 225: 328-340.

Knowledge and LLM-driven Lithology Interpretation

Experimental results

Data: We selected over 1000 samples from the large dataset to construct two small sets.

Comparison methods: classic DL models and LLaVA-Mistral-7B.

Comparative study

Gaofen 2 data

Table 2: The OA results of comparison experiments in our research area(%)*

Method Name	AR	CR	DR	GR	ST	SL	SM	WB	OA
Squeezenet	25.5	37.5	42.4	64.8	20.0	43.8	56.7	63.4	44.9
Resnet	24.8	36.1	51.3	48.1	70.0	50.0	32.4	65.8	45.1
Densenet(baseline)	22.6	70.8	37.0	69.1	76.7	67.2	56.8	63.4	54.4
LLaVA-V1.6-Mistral-7B	36.8	**75.0**	37.0	64.8	33.3	61.7	64.9	**1.0**	55.8
Our Method	**63.9**	72.2	**82.2**	**75.3**	**93.3**	**90.7**	**78.3**	**1.0**	**79.7**

* Arenaceous Rocks (AR),Carbonate Rock (CR),Diorite (DR),Granite (GR), Silt (ST), Slate (SL),Soil Mass (SM),Water Bodies (WB).

Landsat 8 data

Table 3: The OA results of comparison experiments in Landsat8 research area(%)*

Method Name	AR	CR	DR	GR	ST	SL	SM	WB	OA
Squeezenet	5.0	54.2	27.2	38.4	41.0	59.2	40.3	83.0	42.0
Resnet	26.0	**66.7**	28.4	13.1	**67.0**	0.1	29.9	**95.4**	38.0
Densenet(baseline)	20.5	36.4	29.6	**59.6**	46.0	67.0	52.7	55.4	46.3
LLaVA-V1.6-Mistral-7B	25.0	45.8	38.2	34.3	39.0	49.5	49.3	80.0	44.0
Our Method	**35.0**	44.3	**38.3**	48.5	63.0	**69.0**	**56.9**	86.5	**53.9**

* Arenaceous Rocks (AR),Carbonate Rock (CR),Diorite (DR),Granite (GR), Silt (ST), Slate (SL),Soil Mass (SM),Water Bodies (WB).

Ablation study

Table 5: Ablation experiment of the LLaVA-V1.6-Mistral-7B-RAG in our research area(%)

LLaVA-V1.6-Mistral-7B	RAG by text description	RAG by image	OA
√			55.8
√	√		75.7
√		√	76.8
√	√	√	**79.7**

Compared to the LLaVA-Mistral-7B and the DL models, our method improved the performance by more than 20% in Gaofen 2 data, and 7.6% in Landsat 8.

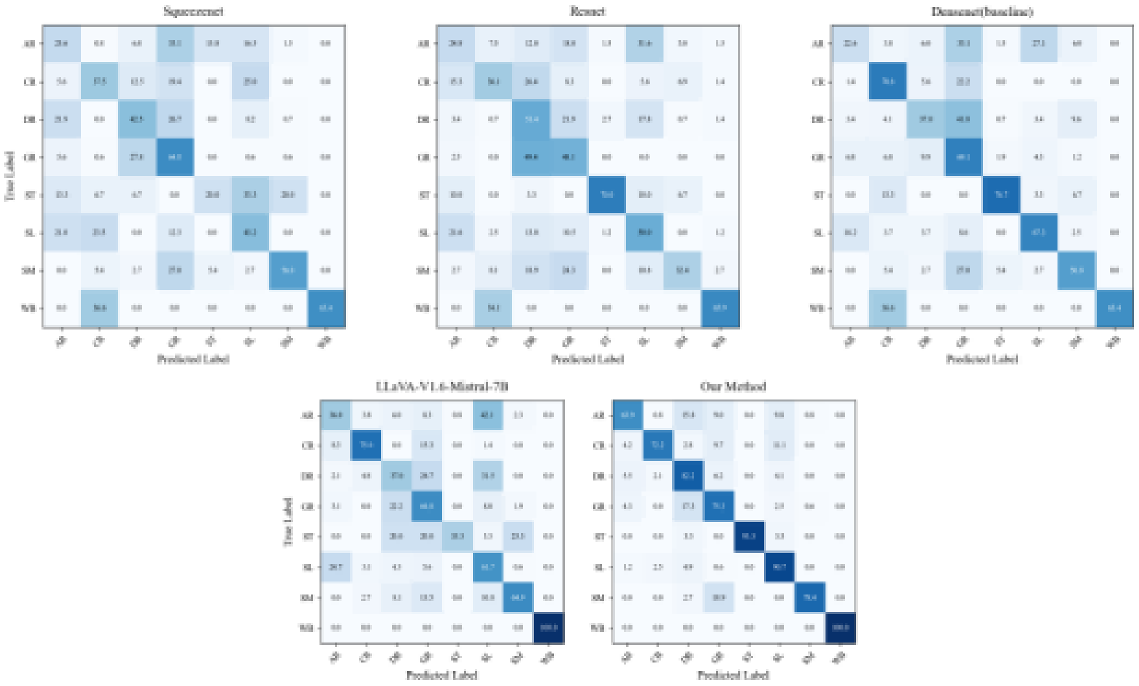
Knowledge and LLM-driven Lithology Interpretation

Experimental results

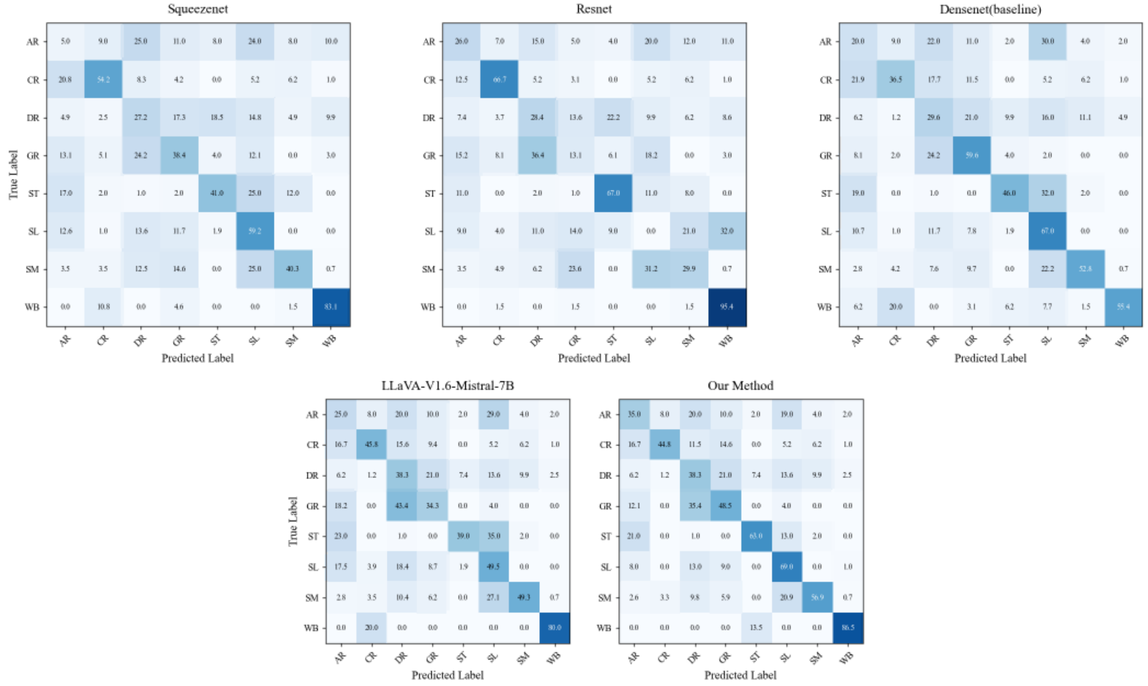
Data: We selected over 1000 samples from the large dataset to construct two small sets.

Comparison method: classic DL models and LLaVA-Mistral-7B.

Gaofen 2 data



Landsat 8 data



Compared to the LLaVA-Mistral-7B and the DL models, our method improved the performance by more than 20% in Gaofen 2 data, and 7.6% in Landsat 8.

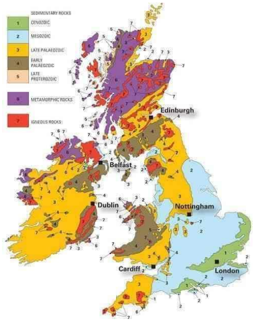
LithoBench: Benchmarking LLMs for Lithology Interpretation

1、Limitations of existing benchmarks

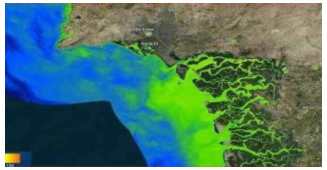
Dataset	Domain	Type	Samples	Labels	Task
HyRANK	RS	HSI	5 scenes	Pixel labels	Land-cover classification
WorldView-3	Litho-RS	Lithology	RS scenes	Lithology labels	Lithological mapping
TASI Lithology	Litho-RS	Lithology	3 sites	Lithology labels	Lithology classification
PRISMA	Litho-RS	Lithology	RS scenes	Lithology labels	Lithological mapping
OpenEarthMap	RS	Segmentation	5,000 images	2.2M labels	Land-cover segmentation
FAIR1M	RS	Detection	>15,000 images	>1M labels	Fine-grained object detection
SAMRS	RS	Segmentation	105,090 images	1.67M labels	Segmentation and detection
Geo-Bench	Geo-RS	GFM benchmark	Multiple datasets	Task labels	Classification and segmentation
PANGAEA	Geo-RS	GFM benchmark	Multiple datasets	Task labels	Geospatial analysis
RSGPT	RS-VL	Image captioning	RS images	2,585 captions	RS image captioning
EarthVQA	RS-VQA	Visual QA	6,000 images	208,593 QA pairs	Reasoning, counting, analysis
RSVBench	RS-VL	VL benchmark	29,614 images	123,221 QA pairs	Captioning, grounding, QA
RSVLM-QA	RS-VQA	Visual QA	13,820 images	162,373 QA pairs	Captioning and counting
OmniEarth	Geo-VL	VLM benchmark	9,275 images	44,210 instructions	Perception, reasoning, robustness
MSEarth	Earth Sci.	Scientific VLM	>289K images	QA / captions	Captioning, selection, reasoning

Existing benchmarks focus on general classification and recognition.

2、Needs for real world applications

Geological survey **Geological mapping**

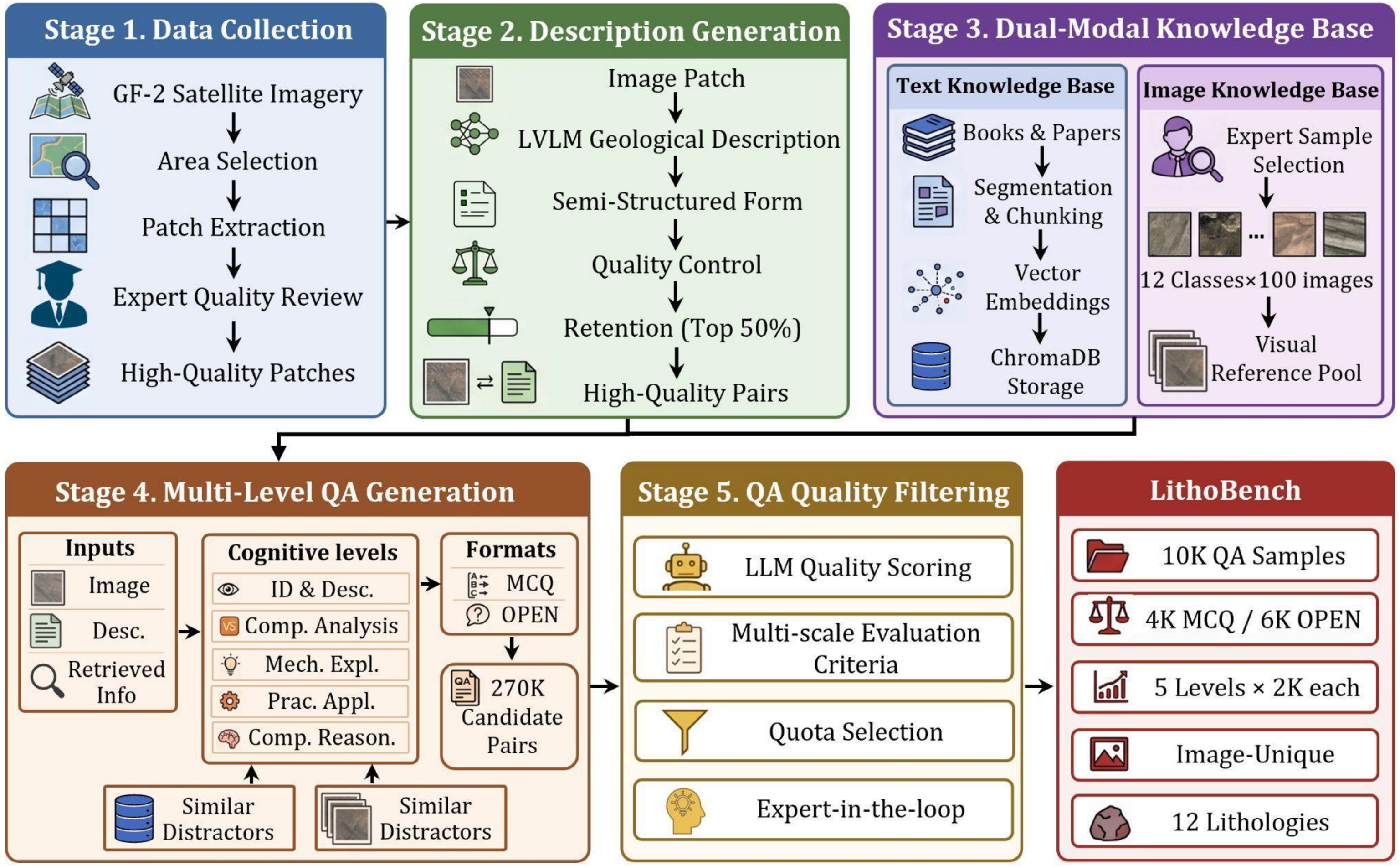
Disaster early warning **Resource monitoring**

Models need expert-level understanding of lithological evidence.

Therefore, a multi-level benchmark is urgently needed for expert-level RS lithology understanding and advanced geoscientific reasoning.

LithoBench: Benchmarking LLMs for Lithology Interpretation

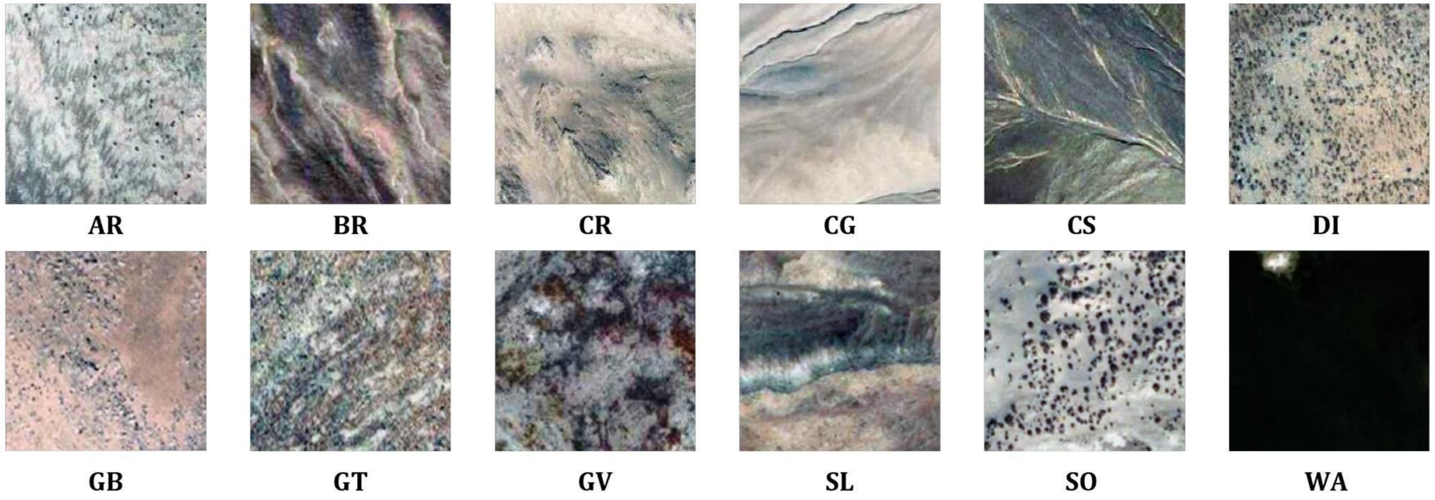
We built a multimodal benchmark dataset to **evaluate the LLMs‘ geological understanding capability** across five tasks: **identification and description, comparative analysis, mechanism explanation, practical application, and comprehensive reasoning**.



Statistic	Quantity
Lithology classes	12
Image patches	112,163
Image–caption pairs	56,000
Candidate instances	269,790
Final Q&A instances	10,000
Multiple-choice questions	4,000
Open-ended questions	6,000
Cognitive levels	5
Instances per level	2,000
MCQs per level	800

LithoBench: Benchmarking LLMs for Lithology Interpretation

Word cloud



Representative images in LithoBench

LithoBench: Benchmarking LLMs for Lithology Interpretation

1 Identification & Description

Question: Based on the **visible tone**, **texture**, and **structures**, **identify and describe** the lithology here? Describe the color, grain features, patches or spots, surface continuity, and any obvious layering or linear structures.

2 Comparative Analysis

Question: Which **visual features** can help experts **distinguish** this surface from similar gneiss-like or crystalline surfaces, especially in terms of **tone**, **texture**, and **structural organization**?

3 Mechanism Explanation

Question: **Why** is this image more consistent with granite than sedimentary or schistose rocks? Explain using **tone**, **coarse texture**, **structure**, **mineral clues, and crystallization features**.

4 Practical Application

Question: Based on the **exposed** surface texture, **blocky** weathering features, and weak linear structures, is this area **suitable for high-quality decorative stone mining**? What **field checks** should be conducted before mining?

5 Practical Application

Question: Given the coarse speckled texture, clear light–dark mineral contrast, and lack of obvious bedding or foliation, what is the most likely **geological origin** of this rock? How do these features **support the interpretation**?

6 Multiple-choice Question

Question: **Which lithology** best matches the **gray tone**, fine texture, weak speckles, and **scattered** bands?

Options:

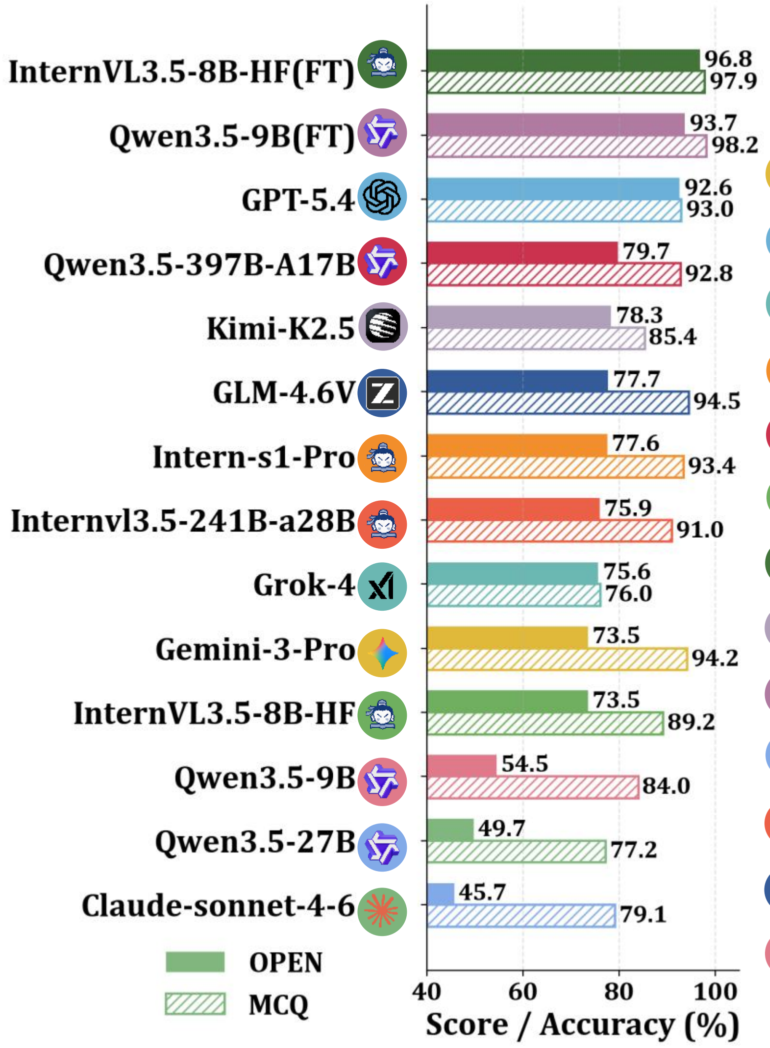
A. Crushed stone B. Conglomerate

C. Breccia D. Arenaceous rocks

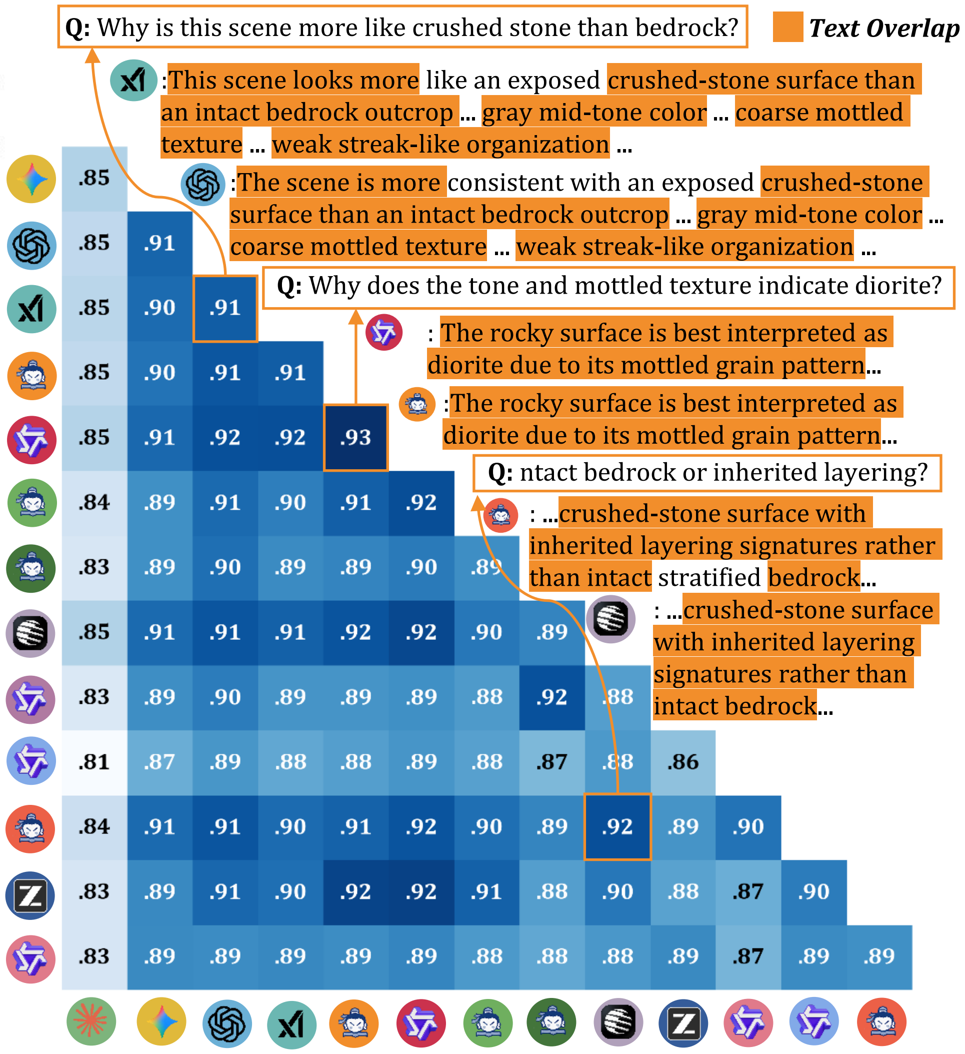
Core Design: Five capability levels × two question types

Open-ended / Multiple-choice questions

LithoBench: Benchmarking LLMs for Lithology Interpretation



（1） Mainstream LLM performance on LithoBench



（2） Similarity Across LLM Responses

1. Fine-tuning leads to the best overall performance: Fine-tuned models achieve the highest scores on both open-ended and multiple-choice questions.

2. LLM responses show high similarity: Mainstream LLMs mainly rely on visual features to predict the lithology categories, leading to high response similarity.

3. Higher-order reasoning remains limited: Many LLMs still struggle with application and comprehensive reasoning tasks.

Future directions and conclusion

Summary

Data — Built multimodal lithology datasets and knowledge resources.

Model — Integrated geological knowledge with LLM-based interpretation.

App — Supported lithology mapping and geological applications.

Future Directions

Build multi-scale image–knowledge–text datasets for lithology, mines, hazards, and resources.

Develop geological LLMs aligned with expert reasoning and domain knowledge.

Expand to geological mapping, hazard assessment, resource monitoring.

The Innovation: A Journal to See the Unseen and Change the Unchanged CellPress

"*The Innovation*" was launched in **May 2020** and rigorously follows the same peer review, editorial, and publishing standards of ***Cell***. There are **197** Editorial members from **21** countries and their research interests cover all disciplines of science. So far, the authors are from 58 countries and the papers had been cited by authors from 139 countries.

"*The Innovation*" has been indexed by **DOAJ, ADS, Scopus, WHO, Google Scholar, PubMed, ESCI, INSPEC, EI, CAS Journal Ranking (Q1)**. So far, **> 45%** citations are outside of China. Not surprisingly, the rejection rate has risen to **90%.** From submission to publishing, the mean time duration was **62 days** in 2020-2023. It got CiteScore 2025 (**=32.6**) and Impact 2025 Factor (**=39.5**), respectively.

Website: www.the-innovation.org www.cell.com/the-innovation

The Innovation Geoscience: An international journal of Earth science

INNOVATION
PRESS

Editors-in-Chief

Prof. Hai Cheng
Academician
Xi'an Jiaotong University

Prof. Christoph Spötl
Academician
University of Innsbruck

The Innovation Geoscience is an Open Access journal and a sister journal of *The Innovation*, aiming to boost geoscience research and promote collaboration with other disciplines and the sustainable development for human society. It welcomes submissions from all fields of geoscience and from any source, including but not limited to: Geography, Ecology, Geology, Geochemistry, Geophysics and Space Physics, Atmosphere and Oceanography, Environmental Science, and GeoscienceX for multidiscipline collaboration, et al.

The Innovation Geoscience has published its first issue on **26th June 2023** and rigorously follows the same peer review, editorial, and publishing standards of *The Innovation* which got CiteScore 2022 (=23.6) on June 8 2023 and Impact Factor (=32.1) on June 28, 2023 , respectively. There are 65 Editorial members from 12 countries and 1/3 of them are outside of China. And their research interests cover all disciplines of geoscience.

Website:
www.the-innovation.org/geoscience
Submission:
https://mc03.manuscriptcentral.com/innovation-geo

Thank You for Your Attention！