



# LLM-Driven Multimodal Knowledge Graph Construction for Bridges

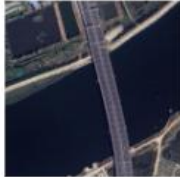
Chao Wu<sup>1</sup>, Yewen Fan<sup>1</sup> †, Hope Irvine<sup>2</sup>, Alex Robertson<sup>3</sup>

<sup>1</sup> The State Key Lab. LIESMARS, Wuhan University  
{<sup>2</sup> School of Engineering; <sup>3</sup> School of Computing}, Newcastle University

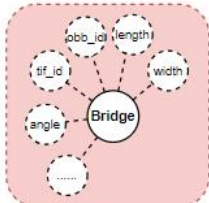


# BackGround

- Bridges are vital components of modern transportation networks and serve as essential enablers of economic and social activity. To ensure their safety, reliability, and resilience — especially in the face of natural disasters and major traffic accidents — it is crucial to leverage GeoAI for continuous monitoring and analysis.
- LLMs have opened up new possibilities for automating knowledge graph construction.
- Existing GeoKGs are lacking the ability to fully represent real-world spatial features.
- Combining vector data (spatial structure) and raster data (real-time imagery) is essential and unexplored.



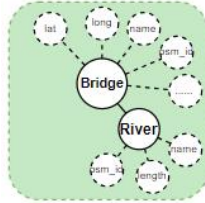
*real-time*



*Raster Bridge Graph*



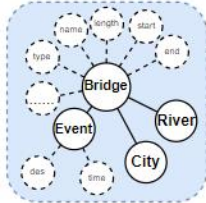
*spatial*



*Vector Bridge Graph*



*textual*



*Text Bridge Graph*



# Related Work

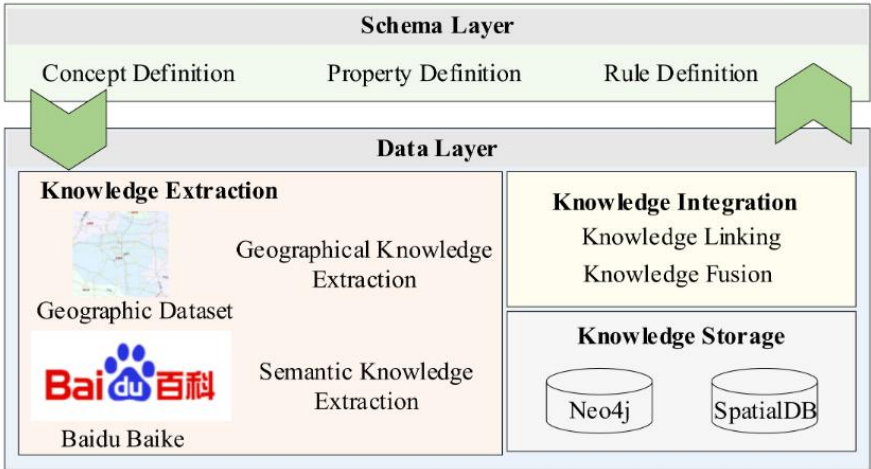


Fig.1. Geographic knowledge graph construction framework[1]

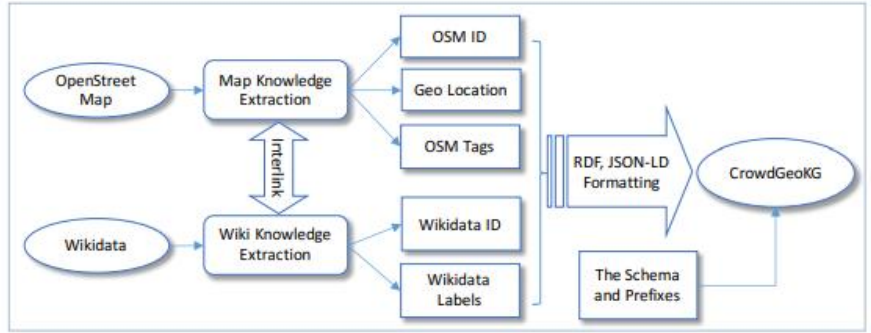


Fig.2. Wiki and OSM knowledge graph construction process[2]

## 1、 Achievements of Traditional GeoKG

Projects like OSMSemanticNetwork, LinkedGeoData, and GeoWordNet link geographic entities from OSM with topology and Wikipedia attributes.

## 2、 Limitations in Real-Time Use

OSM and Wikipedia updates are often slow, making it hard for traditional GeoKG to provide up-to-date, real-time information.

## 3、 Lack of Automation in Multimodal GeoKG Construction

Current approaches require many manual, tedious steps and lack an automated system to handle the full process efficiently.



# Related Work

## 1、 Context-Aware Language Interaction

LLM agents can understand and interact with human language in a nuanced, context-sensitive manner.[3]

## 2、 Integration with Geospatial Systems

Studies such as Remote Sensing ChatGPT[4] demonstrate that combining LLMs with geospatial systems enables rapid and accurate Big Data analysis.[5]

## 3、 Foundation for Next-Generation GeoAI

The ability to manage and analyze complex geospatial data positions LLM agents as a core technology for future multimodal and real-time applications.

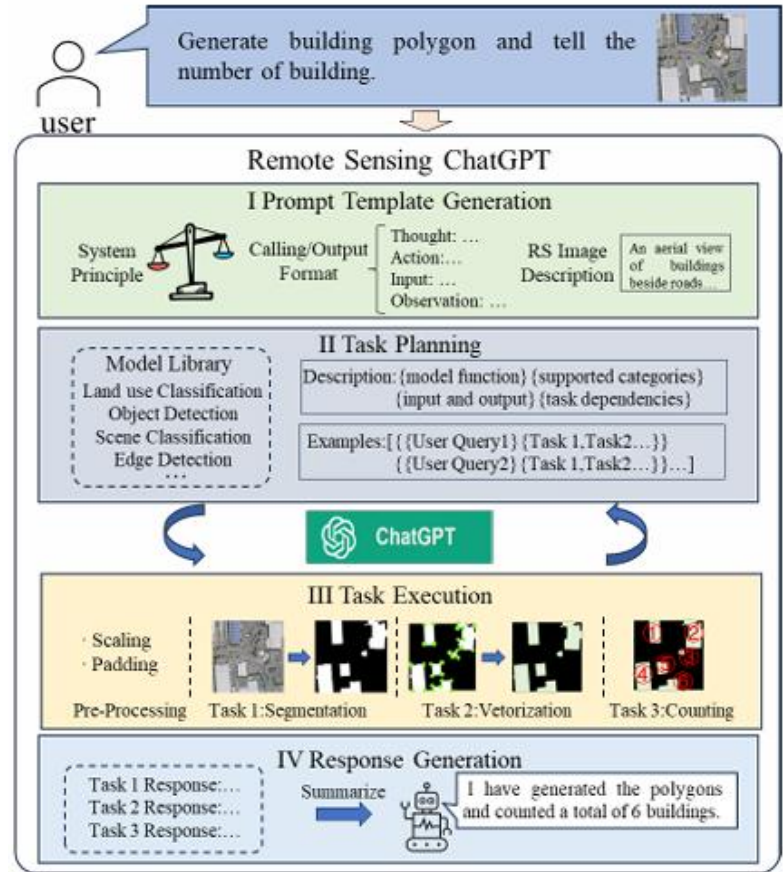
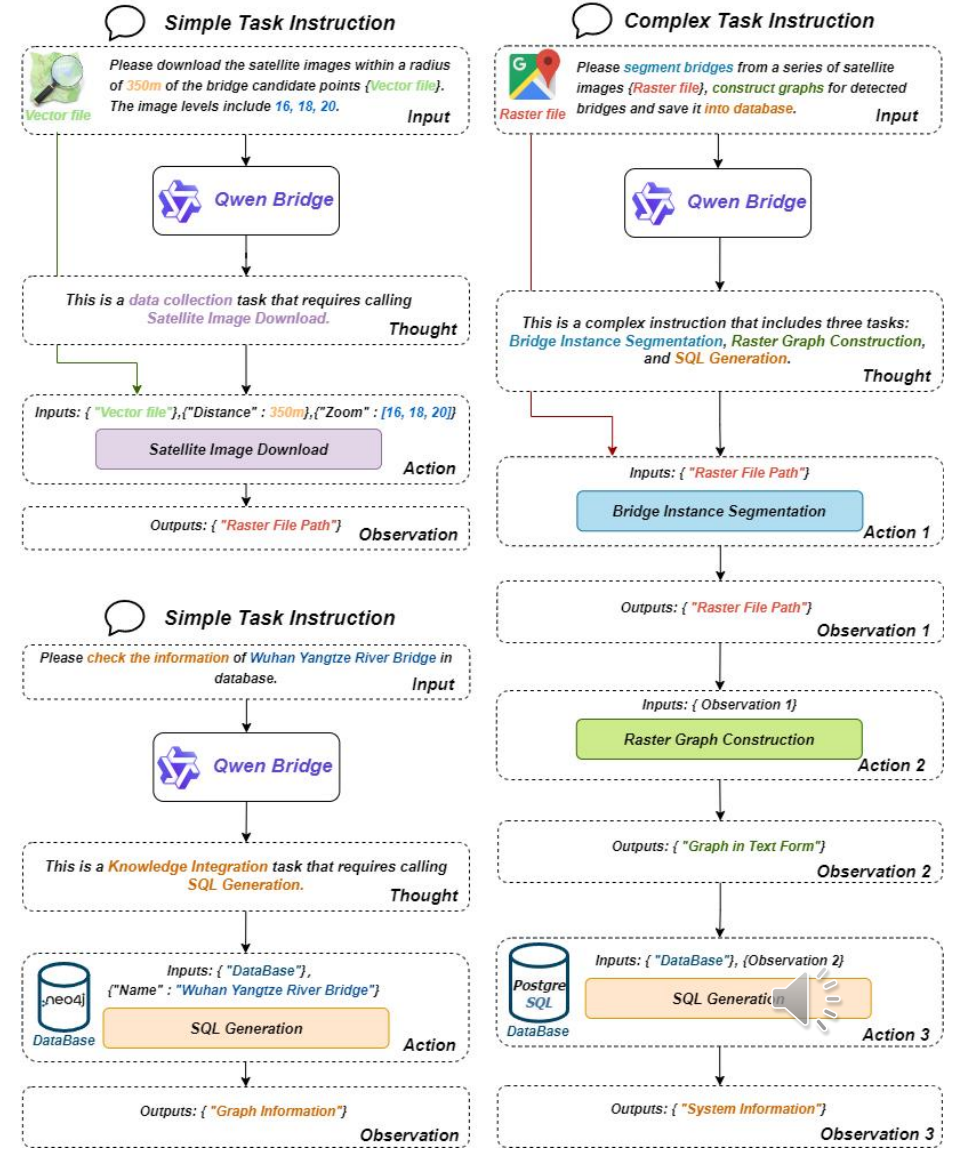
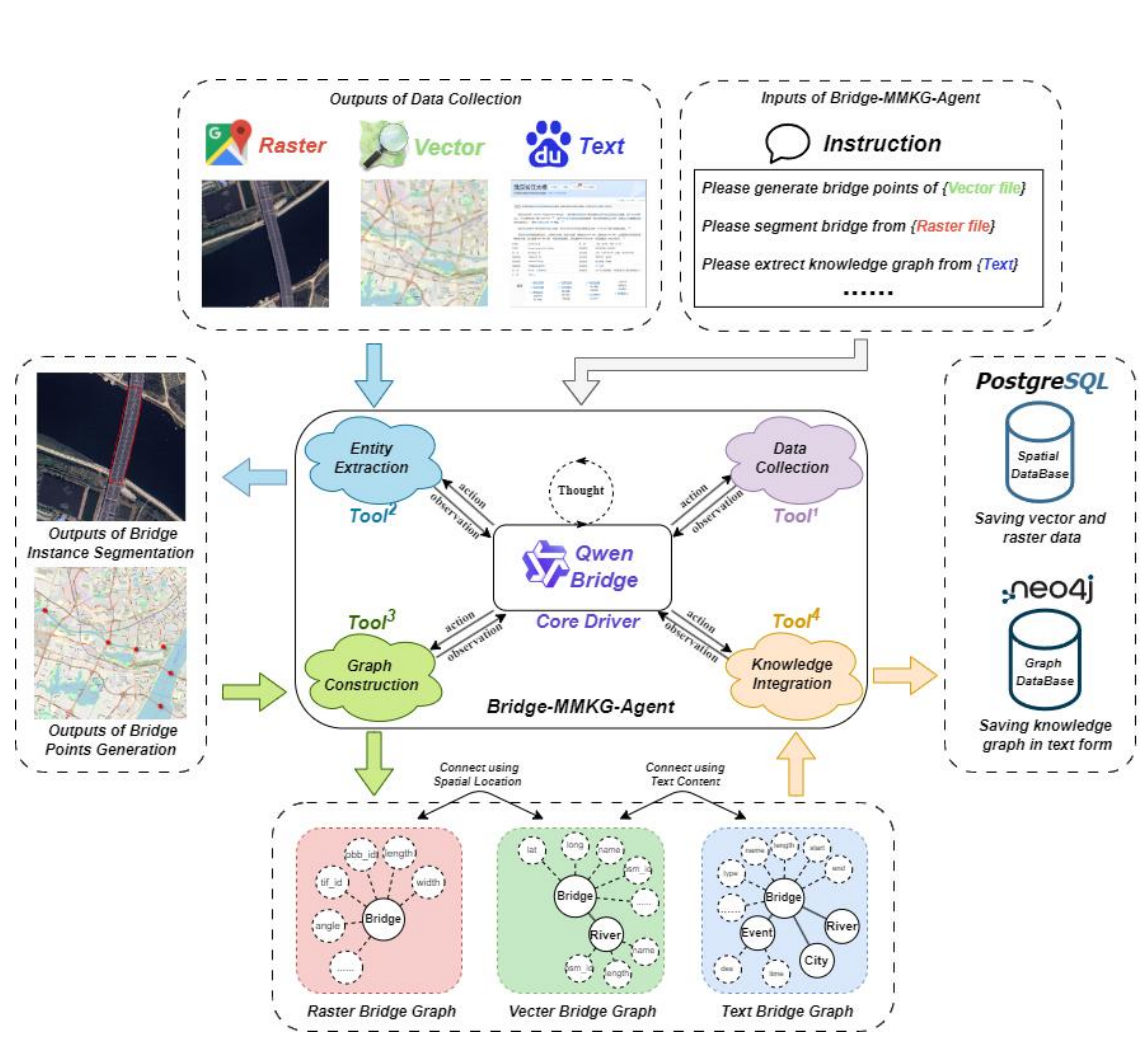


Fig.3. Workflow of Remote Sensing ChatGPT. [4]



# Methods

## The Flowchart and Task Workflow of the proposed framework



# Methods

## Toolkits for Multimodal Knowledge Graph Construction

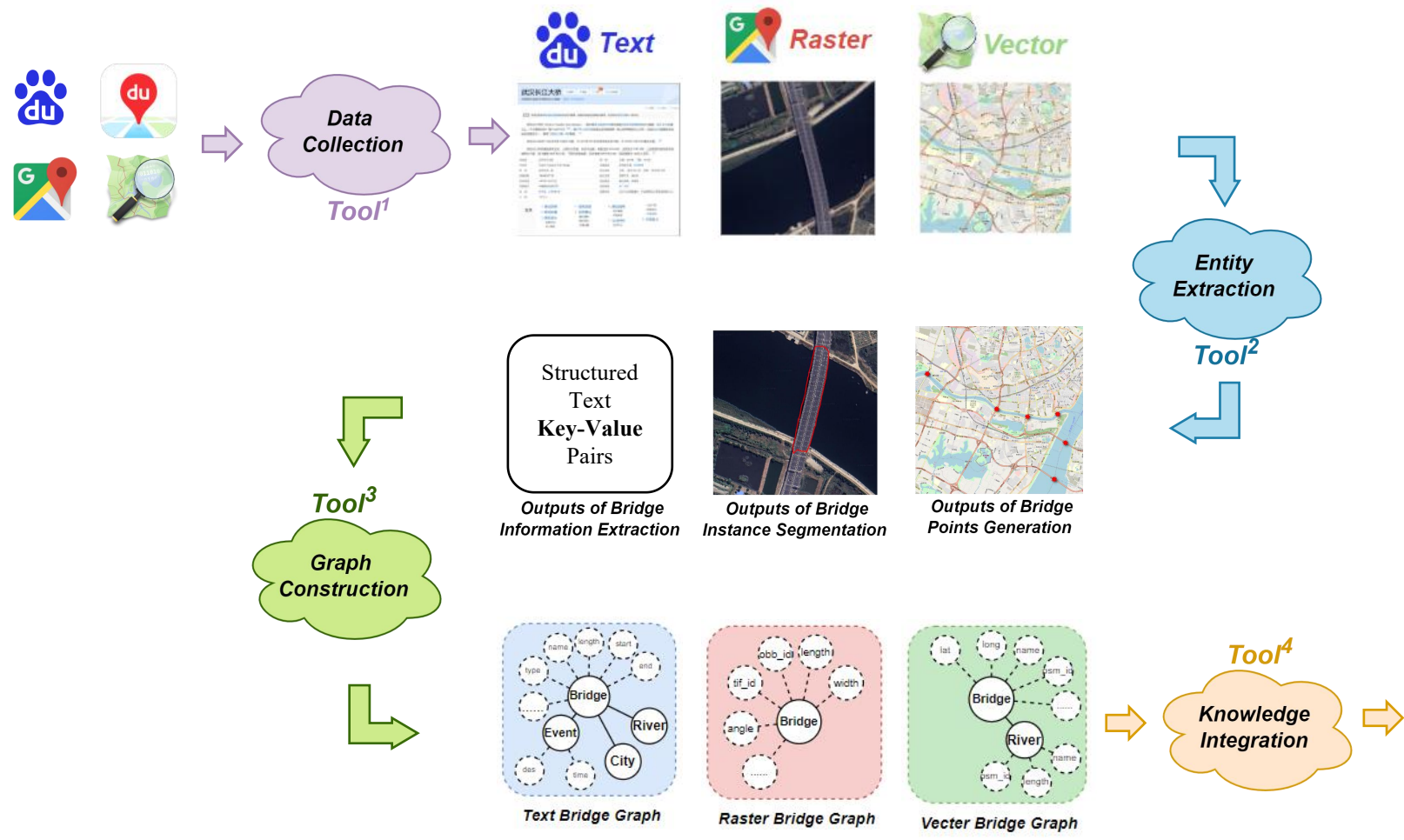
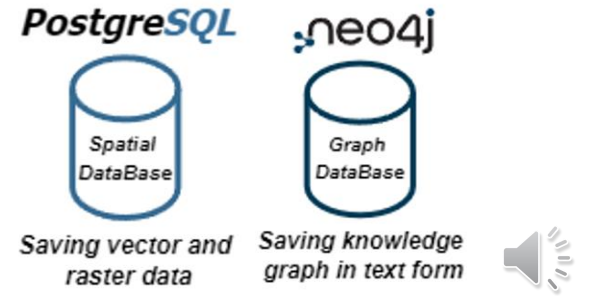


TABLE I  
SUPPORTED TOOLS IN BRIDGE-MMKG-AGENT

Tools	Modality	Method
Satellite Image Download <sup>1</sup>	Raster	Google API
Textual Content Download <sup>1</sup>	Text	Baidu Baike API
Geographic Information Retrieval <sup>1</sup>	Text	Baidu Map API
Bridge Points Extraction <sup>2</sup>	Vector	Arcpy Scripts
Bridge Instance Segmentation <sup>2</sup>	Raster	<i>SAM-Bridge</i> <sup>5</sup>
Bridge Information Extraction <sup>2</sup>	Text	<i>Qwen-Bridge</i> <sup>6</sup>
Vector Graph Integration <sup>3</sup>	Vector	Arcpy Scripts
Raster Graph Integration <sup>3</sup>	Raster	GDAL Scripts
Graph Connection <sup>3</sup>	All	Python Scripts
SQL Generation <sup>4</sup>	Text	<i>Qwen-Bridge</i> <sup>6</sup>

<sup>1</sup> Tools of data collection <sup>2</sup> Tools of entity extraction  
<sup>3</sup> Tools of graph construction <sup>4</sup> Tools of knowledge integration  
<sup>5</sup> A fine-tuned version of SAM for Prompt-free Bridge Segmentation.  
<sup>6</sup> A fine-tuned version of Qwen2.5 as core driver.



# Methods

---

## *Qwen-Bridge* training details

### 1、 Data Collection

- Used tool functions, inputs/outputs, and dependencies from our tool library as prompts.
- Provided example samples to GPT-4o to simulate user queries and generate format compliant responses.

### 2、 Dataset Composition

- ~**1K** training samples, evenly split between Chinese and Engli.
- **180** test samples: 120 simple task instructions + **60** complex task instructions.

### 3、 Model Training & Validation Details

- Training framework: LLaMA-Factory<sup>[6]</sup>.
- Models: LLaMA3, GLM4, Qwen2.5.
- Evaluation method: Calculated accuracy by checking whether **the structured JSON output** from the model exactly **matched the expected results** in the test set.



# Methods

## Example of the training data

```
{
  "instruction": "Please extract the bridge entity from waterways {Vector file1} and roads {Vector file2}.",
  "input": {
    "Vector file1": "./data/osm_file/hunan-latest-free/gis_osm_waterways_free_1.shp",
    "Vector file2": "./data/osm_file/hunan-latest-free/gis_osm_roads_free_1.shp"
  },
  "output": {
    "thought": "This is a simple instruction that includes one task: Bridge Point Extraction.",
    "action": [
      {
        "tool_name": "Bridge Point Extraction",
        "tool_path": "./tools/bridge_point_extraction.py",
        "tool_parameters": {
          "input": [
            "./data/osm_file/hunan-latest-free/gis_osm_waterways_free_1.shp",
            "./data/osm_file/hunan-latest-free/gis_osm_roads_free_1.shp"
          ],
          "output": "./output/bridge_point_extraction/hunan-latest-free/intersection_river_roads.shp"
        }
      }
    ]
  },
  "system": "You are a helpful AI assistant",
  "history": []
},
```

**the structured JSON output**

→ **path and name of the tool**

→ **input of the tool**

→ **output of the tool**



## *SAM-Bridge* training details

### 1、 Data Collection

- 6,000 bridge instance segmentation images from *SAMRS*<sup>[7]</sup> dataset (600 for testing).
- 3,200 non-bridge segmentation images (600 for testing).
- **Purpose:** Improve bridge vs. non-bridge discrimination.

### 2、 Model Training

- Model: ViT-H SAM.
- Training framework: *CellSeg1*<sup>[8]</sup>.
- Fine-tuned using curated bridge and non-bridge imagery.

Using this data combination to train the SAM model with *CellSeg1* produces a model capable of efficiently *segmenting bridges without any prompts*.



# Results and Analysis

To evaluate the effectiveness of the *Bridge-MMKG-Agent*, we designed experiments targeting its two core capabilities:

- (1) accurate task planning and tool invocation by the *Qwen-Bridge* model
- (2) precise segmentation performance of the *SAM-Bridge* model.

TABLE II  
CORRECTNESS OF BRIDGE-MMKG-AGENT ON TASK PLANNING

Tools	LLaMA3 <sup>1</sup>	GLM4 <sup>2</sup>	Qwen2.5 <sup>3</sup>
Overall Correctness	61.1%	79.4%	<u>83.3%</u>
Data Collection	77.8%	<u>88.9%</u>	86.1%
Entity Extraction	66.7%	83.3%	<u>88.9%</u>
Graph Construction	69.4%	<u>86.1%</u>	80.6%
Knowledge Integration	41.7%	58.3%	<u>83.3%</u>
Complex Tasks	46.7%	71.7%	<u>80.0%</u>

<sup>1</sup> Llama-3-8B-Instruct <sup>2</sup> GLM-4-9B-Chat <sup>3</sup> Qwen2.5-7B-Instruct

## Segmentation performance of the *SAM-Bridge*

- mAP@50IoU: 56.81.
- IoU@50IoU: 38.35.
- Precision is 7.3 points below SOTA (per SAMRS paper).

**Analysis** – **Qwen2.5** shows **the highest overall accuracy** in task planning, excelling in Knowledge Integration and complex tasks, while GLM4 demonstrates strong data handling, and LLaMA3 performs moderately in specific areas like Data Collection.

**Analysis** – Performance is acceptable given limited fine-tuning and cost, but low precision suggests dataset noise, highlighting the need for **cleaner data and enhanced augmentation**.

# Conclusion

---

## ❑ **Advancement in Bridge Monitoring:**

The *Bridge-MMKG-Agent* integrates the *Qwen-Bridge* LLM with AI tools to process vector, raster, and text data for real-time, multimodal bridge knowledge graphs.

## ❑ **Efficient Workflow with Practical Impact:**

Modular design handles both simple and complex tasks, and its proven effectiveness in Chinese case studies demonstrates clear improvements in operational efficiency and decision-making for bridge management.

## ❑ **Future Potential:**

The *Bridge-MMKG-Agent* offers a practical demonstration that can inform and inspire future GeoAI applications in infrastructure monitoring.



# Reference

---

- [1] Guo X, Qian H, Wu F, et al. A method for constructing geographical knowledge graph from multisource data[J]. Sustainability, 2021, 13(19): 10602.
- [2] Chen J, Deng S, Chen H. Crowdgeokg: Crowdsourced geo-knowledge graph[C]//Knowledge Graph and Semantic Computing. Language, Knowledge, and Intelligence: Second China Conference, CCKS 2017, Chengdu, China, August 26–29, 2017, Revised Selected Papers 2. Springer Singapore, 2017: 165-172.
- [3] L. Wang, C. Ma, X. Feng, Z. Zhang, H. Yang, J. Zhang, Z. Chen, J. Tang, X. Chen, Y. Lin et al., “A survey on large language model based autonomous agents,” Frontiers of Computer Science, vol. 18, no. 6, p. 186345, 2024.
- [4] H. Guo, X. Su, C. Wu, B. Du, L. Zhang, and D. Li, “Remote sensing chatgpt: Solving remote sensing tasks with chatgpt and visual models,” in IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, 2024, pp. 11 474 – 11 478.
- [5] L. Zhu, J. Wu, B. Wang, G. Zhang, J. Wang, S. Chen, and M. Tan, “Rs-agent: Large language models guided agent system for remote sensing image generation,” in IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium, 2024, pp. 7020–7024
- [6] Y. Zheng, R. Zhang, J. Zhang, Y. Ye, Z. Luo, Z. Feng, and Y. Ma, “Llamafactory: Unified efficient fine-tuning of 100+ language models,” in Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations). Bangkok, Thailand: Association for Computational Linguistics, 2024.
- [7] D. Wang, J. Zhang, B. Du, M. Xu, L. Liu, D. Tao, and L. Zhang, “Samrs: Scaling-up remote sensing segmentation dataset with segment anything model,” Advances in Neural Information Processing Systems, vol. 36, 2024.
- [8] P. Zhou, B. Du, and Y. Xu, “Cellseg1: Robust cell segmentation with one training image,” arXiv preprint arXiv:2412.01410, 2024.





Thank You !

