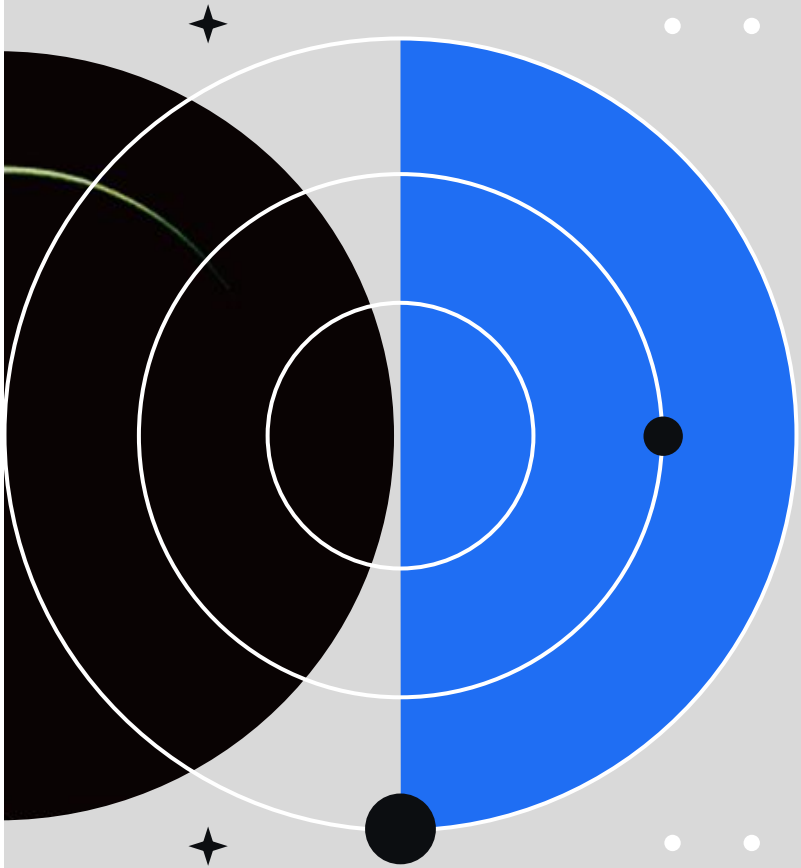
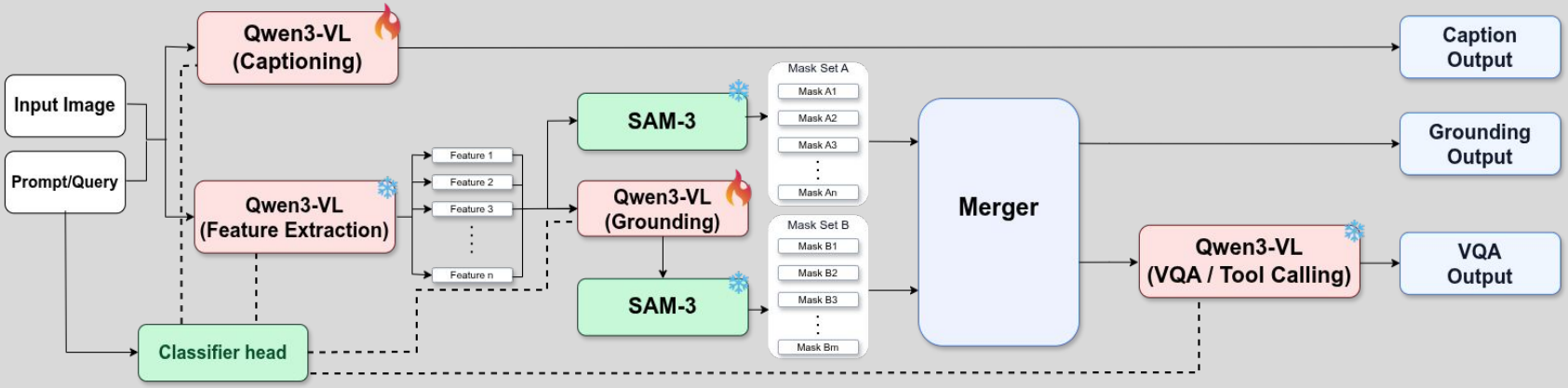


GeoNLI: NATURAL LANGUAGE INTERPRETATION OF SATELLITE IMAGERY

TEAM 27

SOLUTION DESIGN



1 ResNet50 Modality Classifier

Automatic RGB/SAR/false-color detection and prompt injection

2 IoM-Based Mask Merger

Novel graph algorithm for dual-stream segmentation fusion

3 ReAct Tool Orchestration

LangChain agent for geometric reasoning and mathematical calculations

4 Domain-Adaptive LoRA Modules

3 specialized adapters for Qwen3-VL-8B instruct (SAR grounding, optical/SAR caption)

SOFTWARE STACK & MODELS

FOUNDATIONAL MODELS

Qwen3-VL-8B-Instruct
Vision-language backbone

SAM-3 (ViT-H/16)
Dual-stream segmentation

ResNet50
Modality classification

FRAMEWORK & TOOLS

PyTorch
Model inference and training

Unsloth
LoRA fine-tuning and adapter hot-swapping

LangChain
ReAct agent framework

OpenCV, NumPy
Geometric computations

DATASETS

1. MMRS-SARV2
2. SpaceNet6
3. SL4EO-S12
4. VRSBench
5. EarthMind-Bench

APIs Used

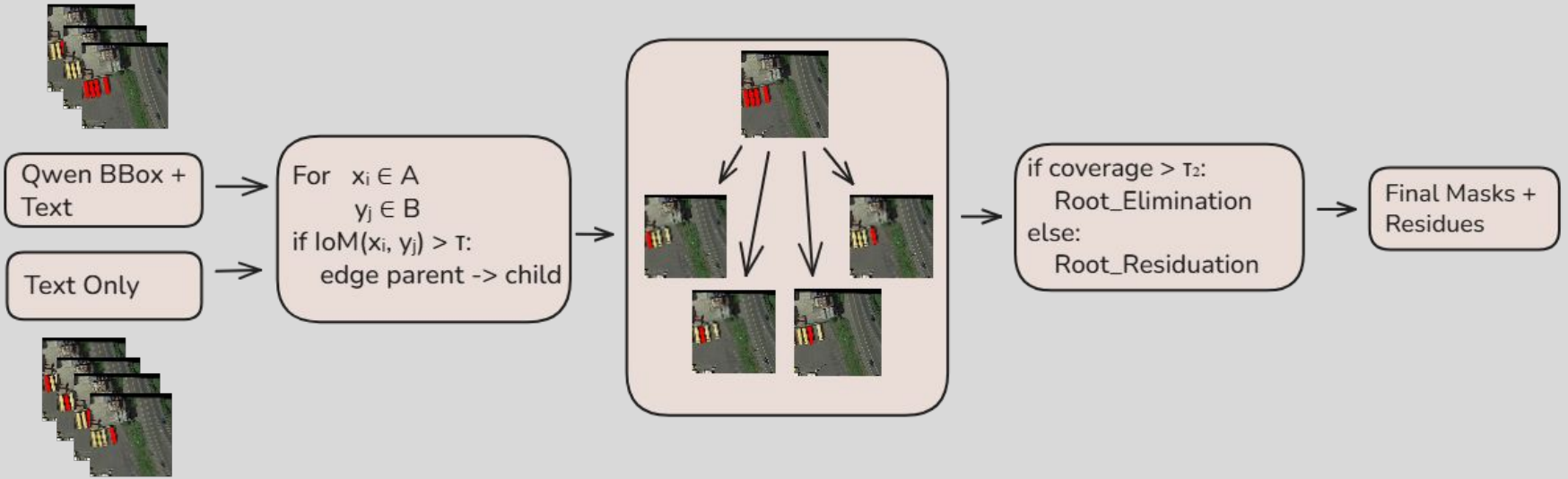
No Commercial API used.
Fully self-hosted,
offline-capable

INNOVATIONS

1. DUAL-STREAM SEGMENTATION + MERGER

Challenge: Single stream segmentation approach misses objects or is too coarse

$$\mathrm{IoM}(parent, child) = \frac{\mathrm{Area}(parent \cap child)}{\mathrm{Area}(child)}$$



INNOVATIONS

2. DOMAIN-ADAPTIVE LoRA (3 specialized adapters)

ADAPTER	DATASET	RANK	PARAMS
SAR Grounding	MMRS SARSv2	r=16	50M
Optical Caption	EMBench RGB	r=16	50M
SAR Caption	Multi Source	r=16	50M

- Swappable based on modality classifier (<500ms overhead)
- Consistent BERT-BLEU improvements: +3-5% over base model

3. ReAct TOOL ORCHESTRATION FOR GEOMETRIC REASONING:

- **Problem:** VLMs hallucinate numerical values
- **Solution:** LangChain ReAct agent with symbolic tools
- **Tools:** distance computation, relative position, safe arithmetic

TEST PHILOSOPHY AND KEY OUTCOMES

TESTING PHILOSOPHY

- **Prototyping Phase**: We used a benchmark of **100 examples** from existing datasets covering:
  - Different Sized Objects
  - Single vs. multiple instances
  - Terrains such as mountains
  - Measurement tasks (area, distance), orientation, counting tasks

- **Systematic Validation** For exhaustive testing we validated using
  - VRSBench Validation Set
  - 20% Hold out on the EarthMind SAR and MMRS datasets

TASK	METRIC	SCORE
Captioning	Bert Bleu-4	0.701
Grounding	CP × Mean IoU	0.334
Grounding	Mean IoU	0.748
Binary Attr	Exact Match	0.782
Numeric Attr	Exp rel Error	0.705
Semantic Attr	Bert Bleu-1	0.637
Captioning **(SAR)**	Bert Bleu-4	0.776
Grounding **(SAR)**	CP × Mean IoU	0.519
Grounding **(SAR)**	Mean IoU	0.785

Eval(GeoNLI) Metrics on VRSBench and MMRS SARV2

LIMITATIONS AND FUTURE WORK

CONCURRENCY HANDLING

- Paged attention limits reliable shared KV-cache for concurrent requests
- Limited GPU memory and bandwidth restrict reliable concurrent requests

AMBIGUITY AND OCCLUSION

- False positives on heavily occluded, partially visible, semantically similar objects
- Feature overgeneralization in dense scenes

FINE-SCALE INSTANCE RECALL

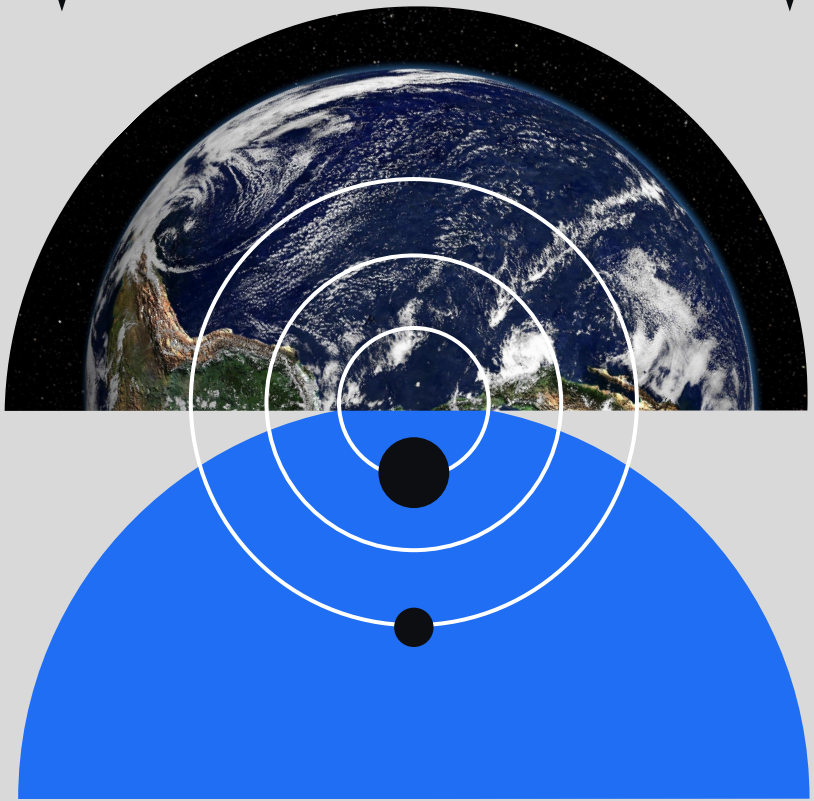
- Misses some minuscule objects in very dense scenes

SYSTEM PROMPT SENSITIVITY

- Compact models (8B) are naturally sensitive to system prompts and instructions
- Task performance varies with prompt phrasing and task-specific instructions

FUTURE WORK

- Finetune Qwen3 VL for better Tool Calling or Employ another suited model



THANKS!

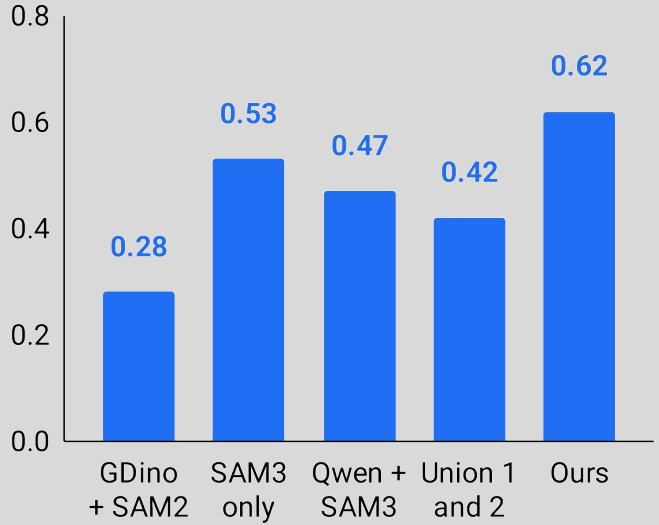
DO YOU HAVE ANY QUESTIONS?

APPENDIX

TRAIN	TEST	OURS	QWEN3	EARTHMIND	GEOPIXEL
EMBench RGB	VRSBench	0.701	0.678	0.714	0.655
EMBench Sar	EarthMind Sar	0.803	0.751	0.787	0.761

Captioning Bert Bleu-4 Scores

CP × Mean IoU (α = 5)



Grounding Metric for Different Approaches On VRSBench

PARAM	TRAIN	TEST	BB-4
50M	EMBench RGB	VRSBench	0.701
50M	EMBench Sar Pair	EMBench Sar Pair	0.803
50M	EMBench Sar+RGB	VRSBench	0.696
50M	EMBench Sar+RGB	EMBench Sar Pair	0.801
1.5M	EMBench Sar+RGB	VRSBench	0.692
1.5M	EMBench Sar+RGB	EMBench Sar Pair	0.797

LoRA Ablation Studies for RGB and SAR