S1: *RS semantic anchoring*

I.Hybrid RS image-caption pool



urban | cropland
forest | wetland
coast | ...

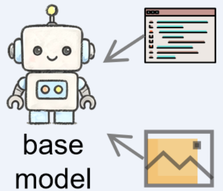
II.RS Image-Text Pair Curation



✓quality filter
deduplication
{}unified format

Elite RS-VL Pairs

III.RS-VL Alignment Training



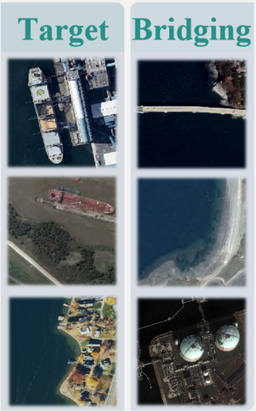
base model

Q:<image> → A:<caption>

projector alignment

RS warm-up

S2: *Domain-bridge convergence*



Shared RS priors



SMT Distillation

Teacher A:metadata grounding

metadates

Gemini-3-pro

Teacher B:SFT Synthesis

SFT.json

```
{
"id":"xxx",
"image":"xxx",
"conversations":["xxx"]
}
```

Qwen3.6-27B

Verifier:review & rewrite

MIMO-v2.5

Bridge-Conv

Target+Bridging
Multi-modal
SMT Distillation

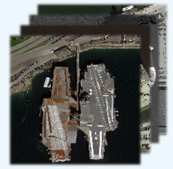
Convergence Training

LoRA Tuning

Bridging-adapted RS-MLLM

S3: *EG scenario tuning*

I.EG Data Construction



metadata-driven curation

img + text ⇄ task

Scenario-EG Set

Grained tasks

Visual evidence binding

Hard negatives

II.Harbor-EG Specialization

Perception | Spatial | Robustness | Generation

Specialization Training

LoRA Tuning

Harbor-specialized RS-MLLMs