GeoHiPO: Geolocalization via Hierarchical Preference Optimization

David Rivard[1,2], Rohit Gupta[1], Parth Parag Kulkarni[1], Mubarak Shah[1]

[1]Center for Research in Computer Vision, University of Central Florida
[2]Lake Superior State University

Worldwide Image Geolocalization

Qwen2.5-VL **Prediction**: Ani, Turkey
Ground Truth: Pompeii, Italy

Contributions

- Teaching VLMs to reason location spatially
- Hierarchically structured training data
- Preference learning using DPO
- Ground-based distance error

Dataset

Hierarchical Pair Difficulty

Easy
Large Error: NYC, USA

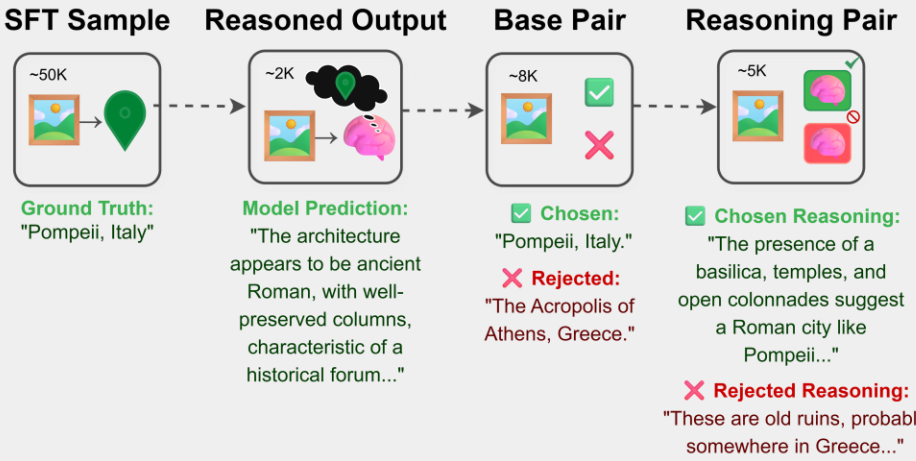
Medium
Moderate Error: Rome, Italy

Hard
Small Error: Naples, Italy

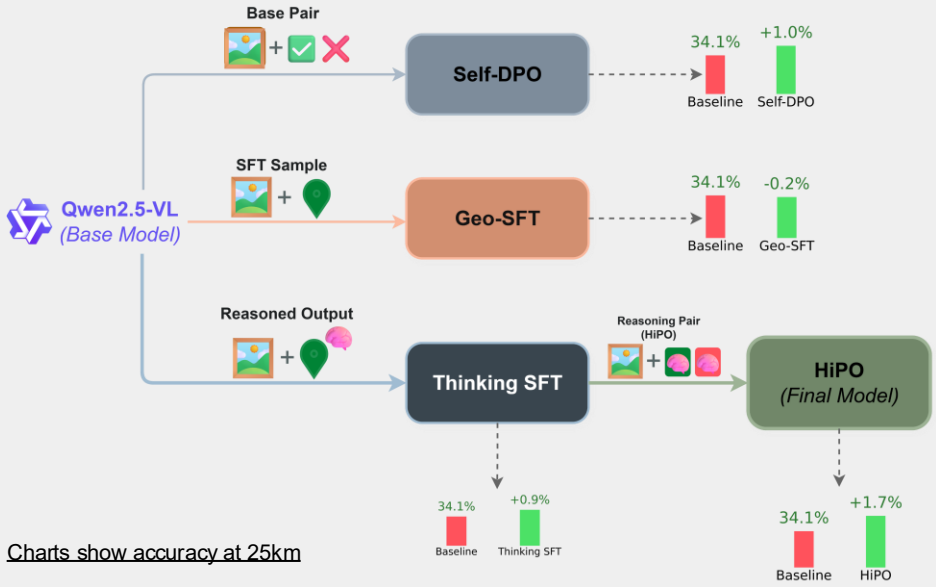
Ground Truth: Pompeii, Italy

Training: Filtered pairs from MP-16
Evaluation: IM2GPS3K benchmark dataset

Data Evolution



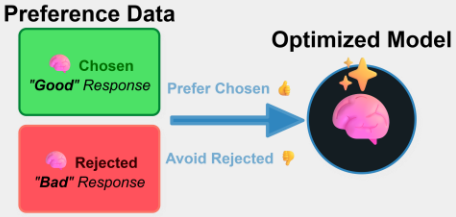
Model Evolution



Charts show accuracy at 25km

DPO?

Direct Preference Optimization



Results

Qwen2.5-VL-7B on IM2GPS3K

Model	Distance (a_r [%] @km)	
	25km	250km
Base Model	34.13	52.22
Geo-SFT	33.97	52.28
Self-DPO	35.16	53.19
Thinking SFT	35.07	53.49
HiPO	**35.87**	**54.59**

- Increased **25km** accuracy by **1.74**%
- Increased **250km** accuracy by **2.37**%
- DPO shows some improvement
- Baseline SFT decreased accuracy

Future Work

- Scale training dataset
- Test on other benchmarks
- Implement contrastive loss
- Apply more strict sample filtering
- Evaluate method on other models