Evaluating Foundation Multimodal Models for High-Resolution Earth Observation: Object Recognition and Contextual Understanding

Enrique Fernández-Laguilhoat, Dr. Ivan Tankoyeu, Dr. Sergey Sukhanov

FlyPix AI GmbH |Robert-Bosch-Str. 7, 64293 Darmstadt, Germany
www.flypix.ai | info@flypix.ai | +49 6151 2776497

ESA-NASA International Workshop on AI Foundation Model for EO | 5-7 May 2025 | ESA-ESRIN | Frascati, Italy

Summary

- **Hypothesis:** General vision-language models underperform at EO tasks without domain adaptation.
- **Goal**: Evaluate zero-shot performance of open-source vision-language models on high-res satellite imagery.
- **Key Findings:**
  - **Captioning**: Models precise captions, often richer than humans
  - **Classification**: Class recognition is successful for objects, but less effective for areas
  - **Task Variance**: Performance for some models fluctuates drastically depending on task (e.g. Phi4 - good captions - bad class identifying)
- **Takeaway:** Foundation models show promise but require EO-specific tuning for reliable geospatial analysis.

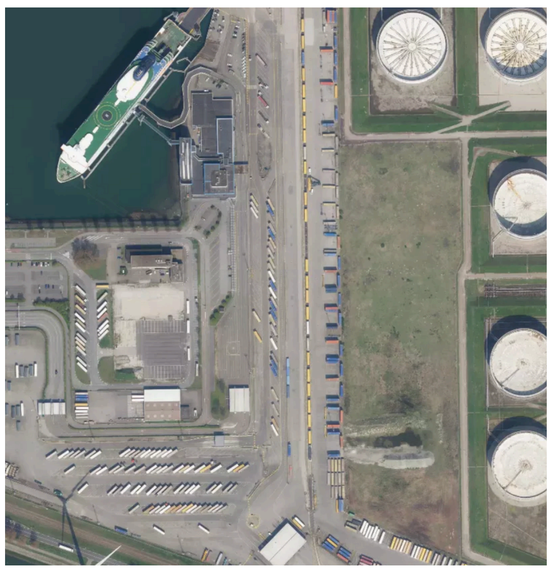
Methodology

Three tasks are given to a pair of humans and 5 open-source models across 15 EO images. Performance of the models is compared to each human and averaged. Humans are compared against each other to establish a baseline.

Models: Gemma-4, Phi-4, Janus-Pro, Qwen 2.5 & SMolVLM were chosen as popular state-of-the-art open-source VLM options.

Tasks:

- **Task 1:** Describe the image
- **Task 2:** List object classes (e.g. “car”, “house”, “tree”)
- **Task 3:** List area classes (e.g. “water”, “forest”, “sand”)



Describe the image

Human: This image shows land mostly covered with crop fields ...

Phi4: The image is an aerial view of a coastal area with a large body of water ...

Gemma3: This is a satellite or aerial view of an arid or semi-arid landscape featuring agricultural activity ...

List object classes

Human: ship, container, reservoir, parking lot, truck, helipad, car, building, tree, pipe

JanusPro7B: airports, boat, buildings, cars, parks, road, storage tanks

SmolVLM: tank, ship, car, truck, building, road, grass, water, train, train track

Evaluate similarity with **BERTScore** + **METEOR**

Lemmatize classes and evaluate with **Precision** + **Recall** + **F1**

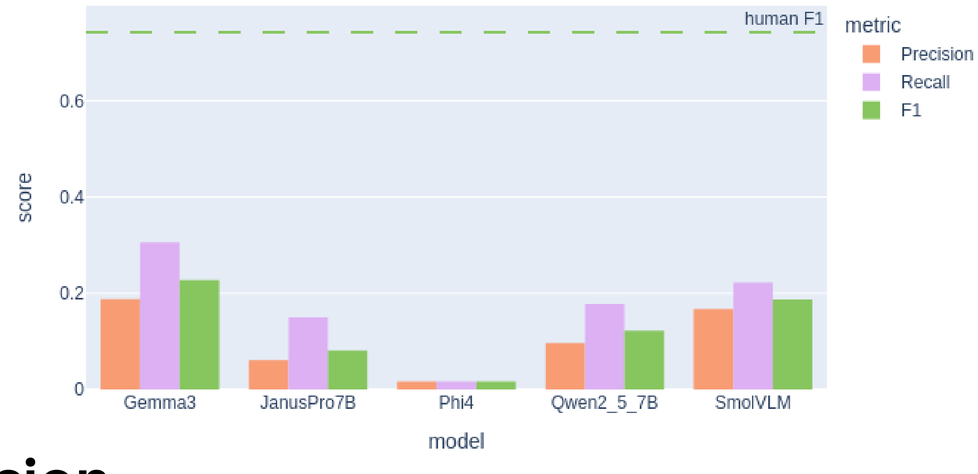
METEOR: Scores text by matching words with stems, synonyms, and order.
BERTScore: Uses BERT embeddings to measure semantic similarity.

Results

Describe image

List objects

List areas



Discussion

- **Human-aligned evaluation:** Performance reflects similarity to professionals, not absolute correctness
- **Task inconsistency**: Models can excel in one task (e.g., Phi-4 captions) but fail in others (e.g., areas). There is no best-for-all model.
- **Caption richness**: Model descriptions are often more detailed and verbose than human ones
- **Object vs. area identification**:
  - **Objects**: Decent accuracy shown across all models
  - **Areas**: Largest deviation from human behavior
- **Training data bias**: Each model’s outputs reflect its training data, creating distinct but self-consistent behaviors

image	Gemma3	Phi4
	**overall impression: the image is** a high-angle, aerial view of a port or industrial area. it shows a complex of infrastructure related to shipping ...	**the image is an aerial view of** an industrial area with a large body of water on the left side. there is a large green and white ship docked at a ...
	**overall impression: the image is** an aerial view, likely taken by a drone or satellite, showcasing a rural or semi-rural residential area. it appears ...	**the image is an aerial view of** a residential area with houses, roads, and green spaces. there are several buildings, some with flat roofs and ...
	**overall impression: the image is** a high-angle aerial view of a coastal recreational area, combining a marina, beach, and associated ...	**the image is an aerial view of** a coastal area with a marina. there are numerous boats docked in the marina, which is located on the ...

Learn More

- **Code+Data**: github.com/flypixai/research-vlm4eo
- **Contact:** sergey@flypix.ai
- **FlyPix AI on DESP:** GeoAI Service https://platform.destine.eu/services/service/geoai/