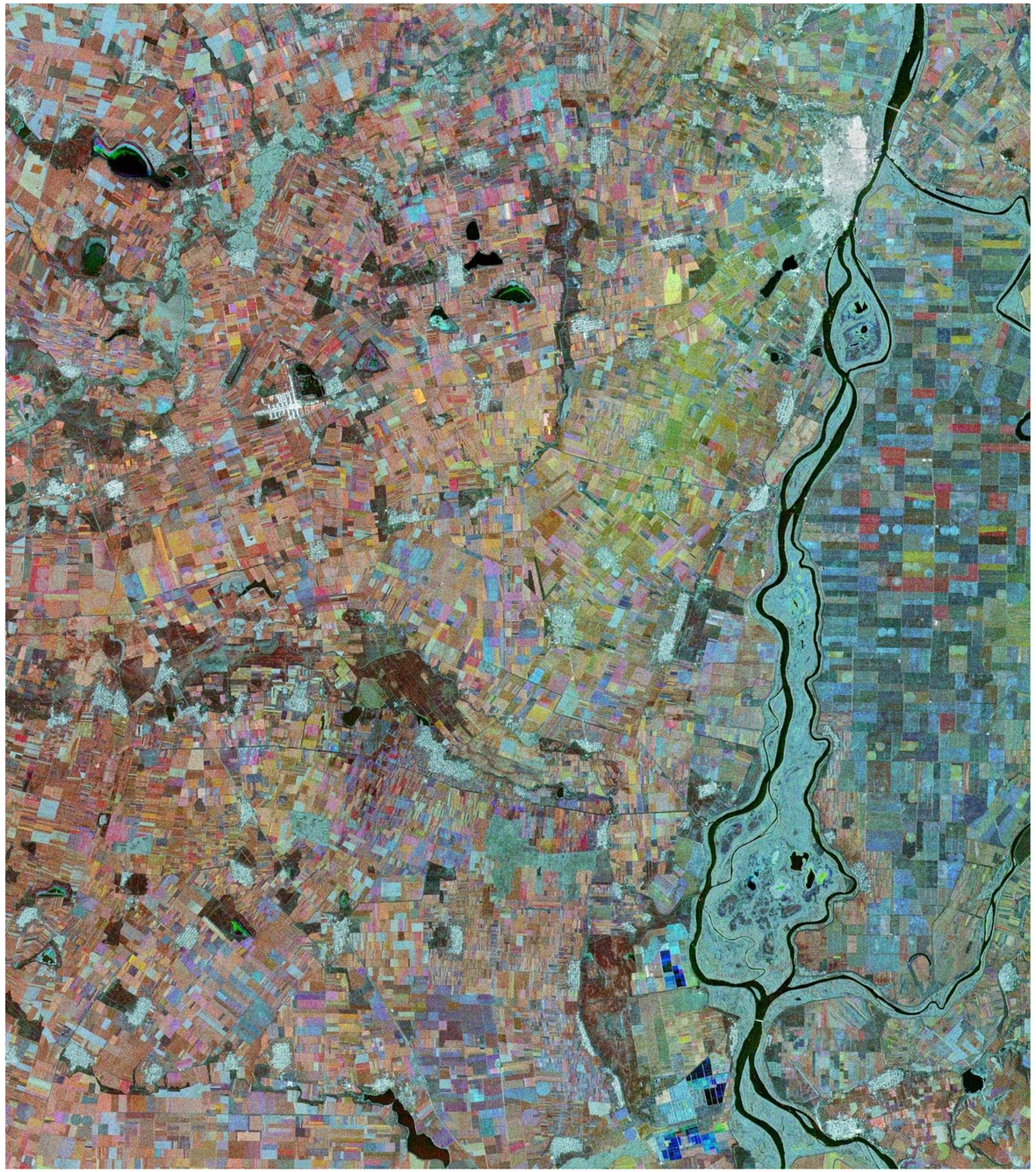
TerraMesh-Masks: Open-Vocabulary Semantic Segmentation for Earth Observation

Benedikt Blumenstiel, Hugues Devimeux,
Johannes Jakubik, Konrad Schindler

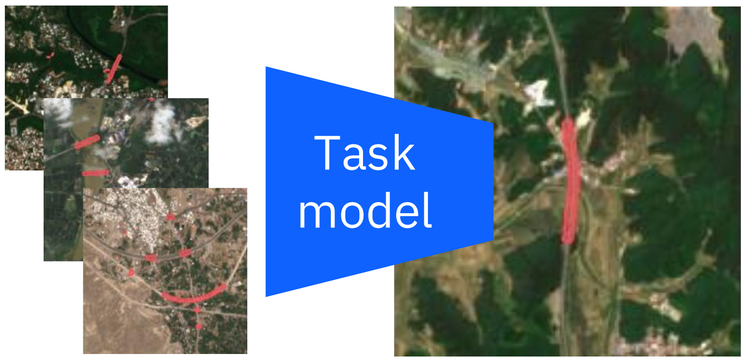


IBM **Research**

ETH zürich

Open-Vocabulary Segmentation

Earth Observation today: Each new task needs labels and a fine-tuned model.



Finetuning for *bridges*

In general computer vision, SAM 3 and other models segment any concept from a text prompt.

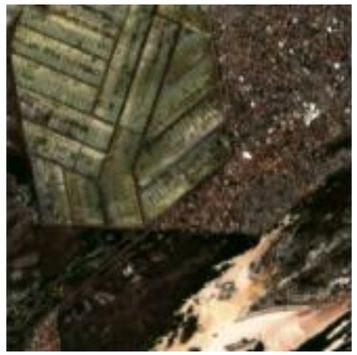


Kangaroo



Pinguin

But generalist models fail on Earth Observation data: They often miss the target or over-segment.



Salt pond



Running track

TerraMesh-Masks & TerraMesh-Masks-Eval

16M Binary masks

44k Unique labels

537k Synonyms

\+ Human-reviewed evaluation set



Railway corridor
◎ Russia
⇄ Railway track, train transportation line, paved rail route, linear rail



Coral reef
◎ Canada
⇄ Reef, shallow coral formation, marine reef structure, coastal reef system



Water retention dam
◎ Cambodia
⇄ Hydraulic dam, reservoir embankment, water control structure



International airport
◎ United States
⇄ Airport runway complex, large aviation facility, aircraft tarmac and



Urban park
◎ France
⇄ City park, public green space, urban recreational area, small urban parkland



Glacier
◎ Peru
⇄ Ice field, permanent ice cover, mountain glacier, high-altitude ice mass



River
◎ Indonesia
⇄ Stream, narrow water channel, inland waterway, small river segment



Quarry
◎ Philippines
⇄ Open-pit mine, stone extraction site, large excavation area, mineral extraction pit



Hospital complex
◎ Austria
⇄ Medical facility, large healthcare infrastructure, hospital campus



Landfill
◎ Chile
⇄ Waste disposal site, municipal solid waste dump, large landfill area

Dataset creation

Data sources

Class selection

Inputs

LLM processing

NDVI-based filtering

TerraMesh-Masks

TerraMesh images

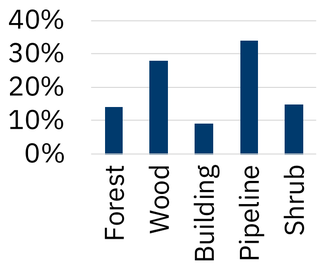
Manual review

TerraMesh-Masks-Eval

Heuristics

Subtypes: ["natural","wood"]
Location: (-91.2, 37.3)
Area: 5.1km2
Quality heuristic: 0.55

Probabilities



Prompt

"You are a remote-sensing annotator…"
Subtypes: ["natural","wood"]
country: "United States"
continent: "Northern America"

+

RGB with purple polygon overlay

Scoring

Quality reason: "The image shows a clear and coherent polygon representing a wooded area…"
Mask quality: 0.95

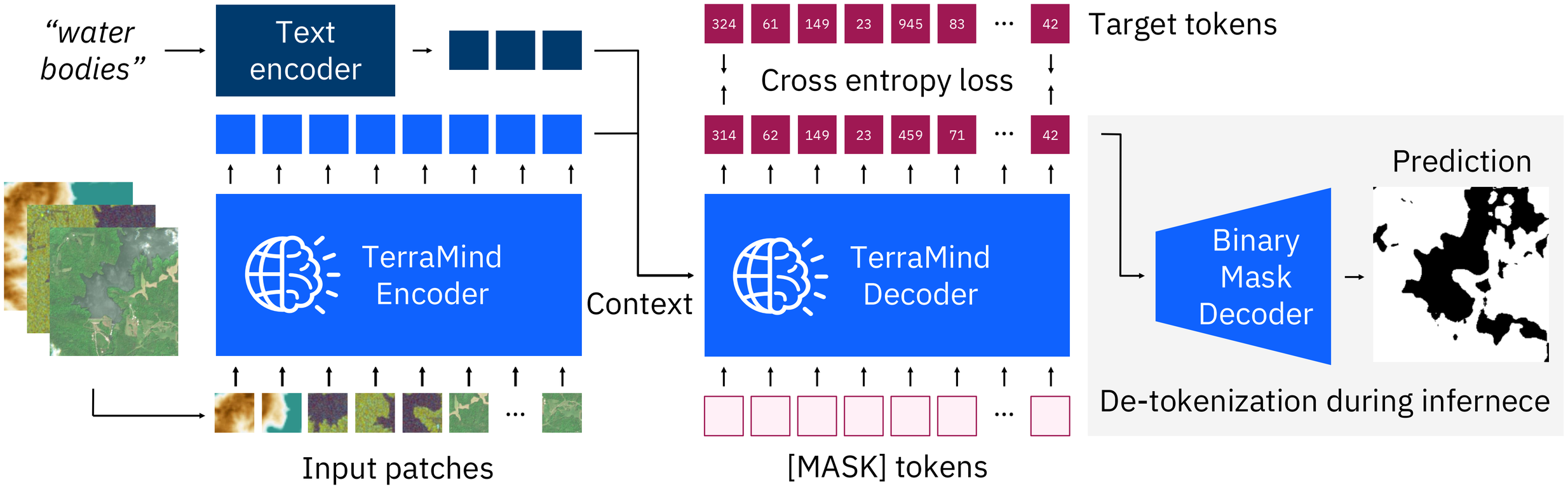
Labelling

Class label: "temperate deciduous woodland"
Class synonyms: "['dense forest canopy cover', 'mixed hardwood forest stand']"

NDVI agreement: 0.9568

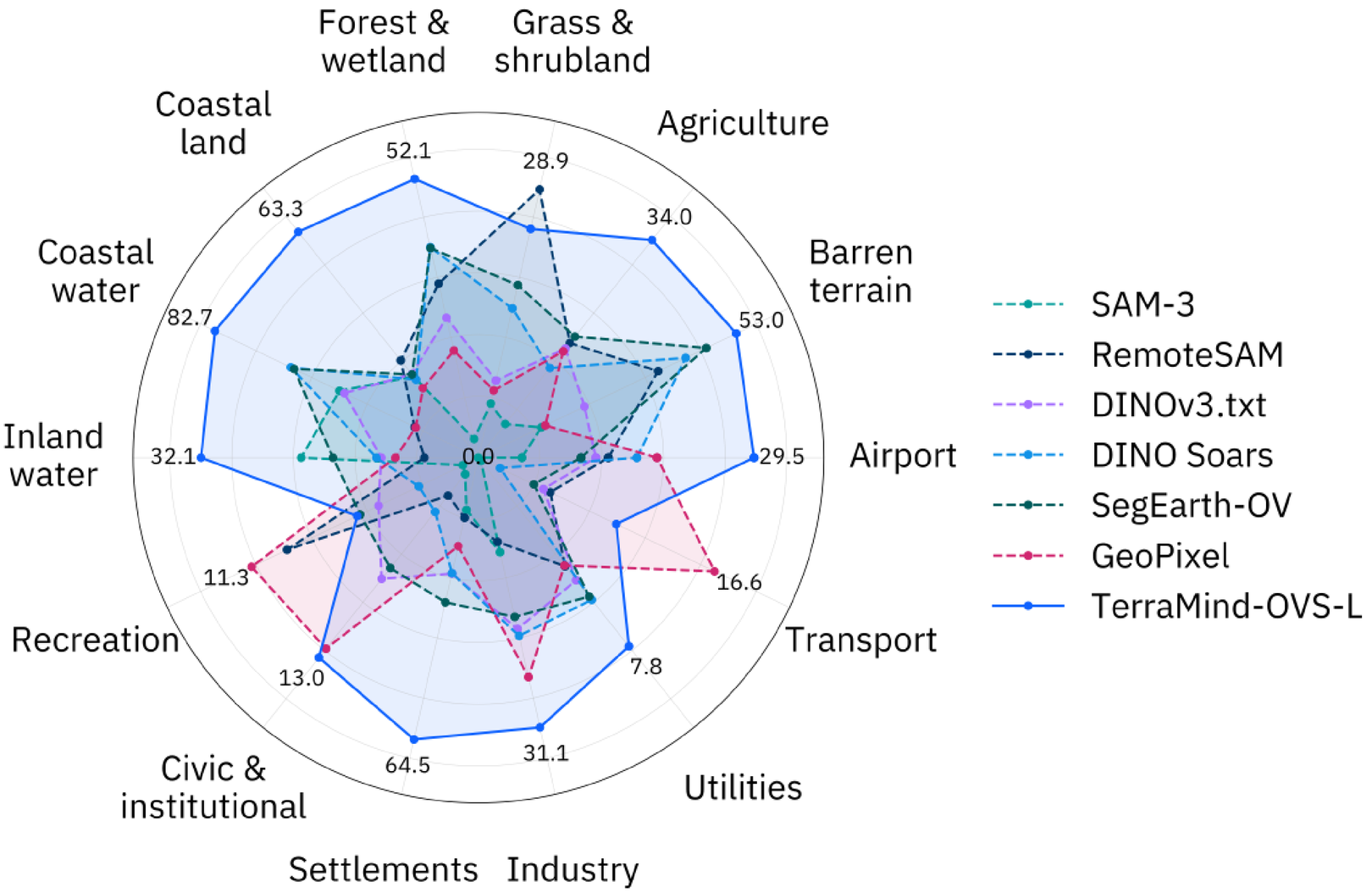
Blumenstiel, B., Fraccaro, P., Marsocci, V., Jakubik, J., ... & Longépé, N. (2025). Terramesh: A planetary mosaic of multimodal earth observation data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 2419-2427).

TerraMind adaptation for Open-Vocabulary Segmentation

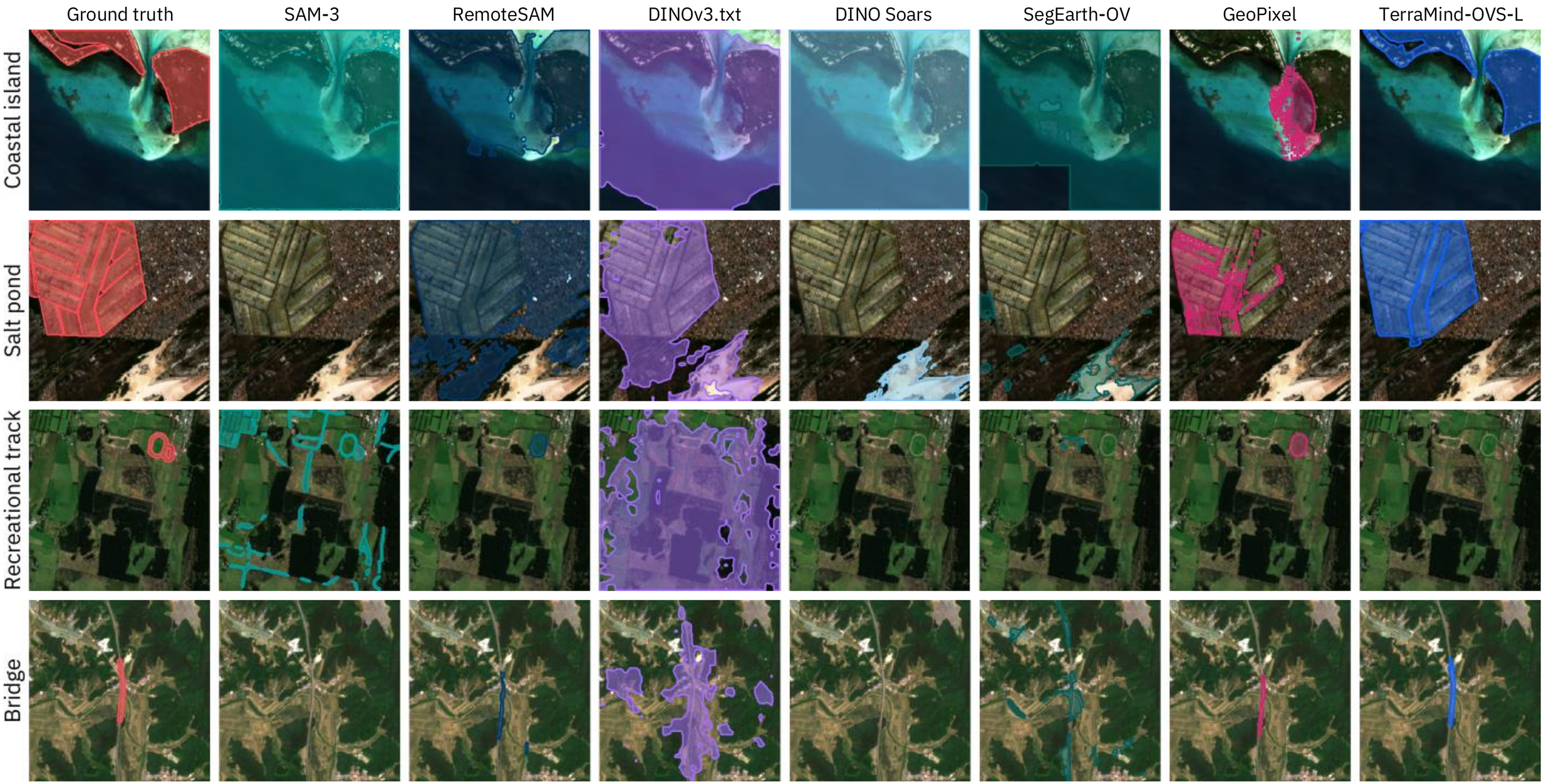


Jakubik, J., Yang, F., Blumenstiel, B., Scheurer, E., Sedona, R., ... & Longépé, N. (2025, October). Terramind: Large-scale generative multimodality for earth observation. In *2025 IEEE/CVF International Conference on Computer Vision* (pp. 7383-7394). IEEE.

TerraMesh-Masks-Eval



IoU per model and label group. Models are tested in an open-vocabulary setting on 200+ class labels that are grouped into 14 areas.


Ground truth
SAM-3
RemoteSAM
DINOv3.txt
DINO Soars
SegEarth-OV
GeoPixel
TerraMind-OVS-L
Coastal island
Salt pond
Recreational track
Bridge



Thank you!

Hugging Face

arXiv:2611.08081