



Scalable Earth Observation Analytics with Cloud Computing and AI Solutions

Living Planet Symposium 2025, Vienna

Marcin Kluczek, Jędrzej S. Bojanowski,
Mikolaj Czerkawski (Asterisk Labs)



Geo AI landscape

Geo AI models

Copernicus-FM SkySense
Scale-MAE SoftCon SpectralEarth
GeoDINO CROMA Satlas
FG-MAE AnySat Tile2Vec
GASSL SpectralGPT Terra-FM
DeCUR SeCo EarthLoc DOFA
Presto MMEarth SSL4EO-S12
Clay PhilEO Prithvi 2.0
TerraMind SatMAE SatCLIP
EarthPT GeoLangBind
Prithvi 1.0

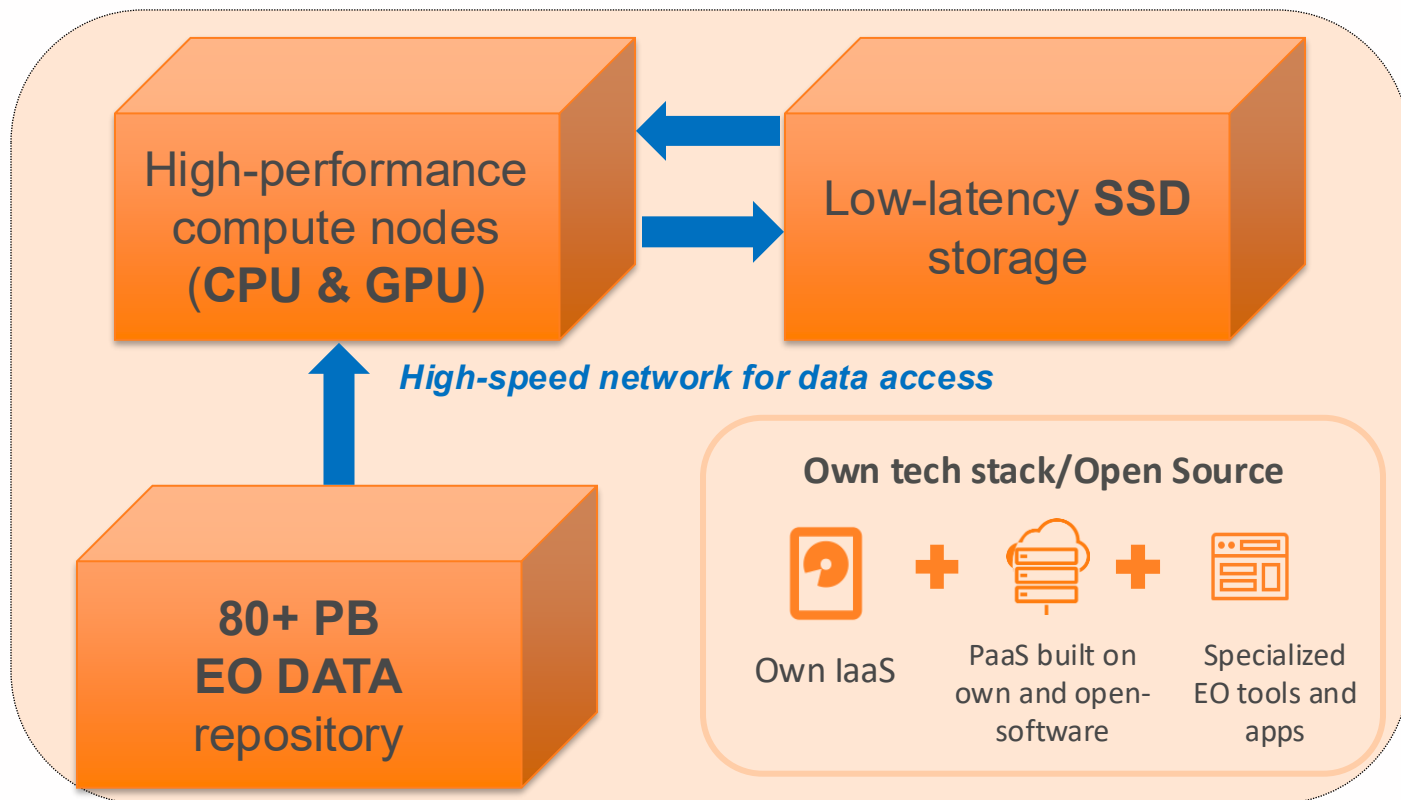
Geo AI datasets

MajorTOM SSL4EO
LAION-EO TerraMesh
Satlaspretrain BigEarthNet
MMEarth
SEN12MS-CR CloudSEN12
WorldStrat RapidAI4EO
SeCo SATLAS

Embeddings

Earth Genome
Quasara
Clay

CloudFerro computing resources for AI models & data

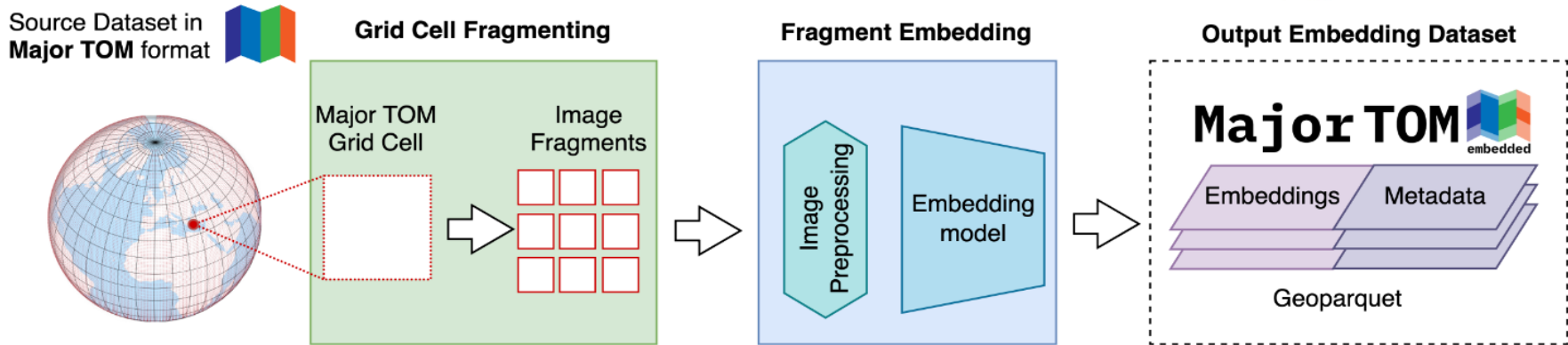


Natural place for **operational embeddings generation**: fresh data availability, highest timeliness, data already in processing pipelines



Embeddings high-level workflow

In cooperation with    asterisk labs



Model	Dataset	Bands	No. of embeddings
SigLIP	Sentinel-2 L2A	3 bands RGB	20,212,974
DINOv2	Sentinel-2 L2A	3 bands RGB	56,147,150
SSL4EO-S12	Sentinel-2 L1C	13 bands multispectral	56,147,150
SSL4EO-S1	Sentinel-1 RTC	VV + VH	36,748,875
DeCUR-S1	Sentinel-1 RTC	VV + VH	36,748,875
DeCUR-S2	Sentinel-2 L1C	13 bands multispectral	56,147,150
MMEarth	Sentinel-2 L2A	12 bands multispectral	39,727,477,454

More information



Open access satellite data embeddings for EO-AI applications

Embeddings datasets summary:

- **51 TB** of AI-embeddings generated from processed Sentinel data
- over **40 billion** embedding vectors
- processing of **147 TB** of raw satellite data
- analysis covering more than **15 million** Sentinel-1 and Sentinel-2 scenes and more than **16 trillion** pixels

Paper

Global and Dense Embeddings of Earth: Major TOM Floating in the Latent Space



arXiv [10.48550/arXiv.2412.05600](https://arxiv.org/abs/10.48550/arXiv.2412.05600)

Available open source via:



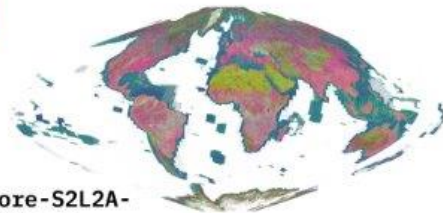
Hugging Face



Major TOM  embedded



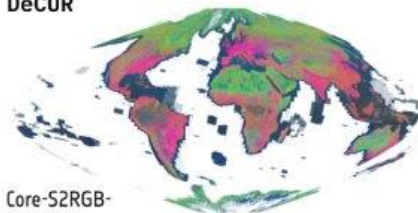
Core-S2L2A-MMEarth



Core-S1RTC-DeCUR



Core-S2L1C-DeCUR



Core-S2RGB-DINOv2



Core-S2RGB-SigLIP



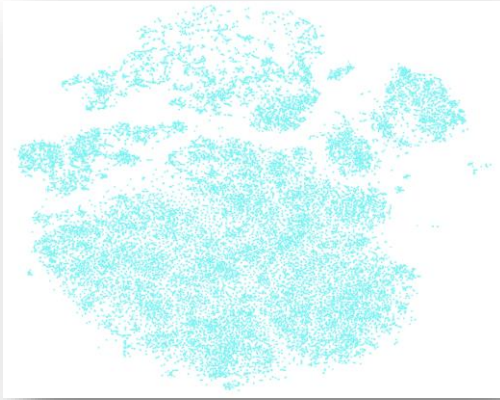
Core-S1RTC-SSL4EO



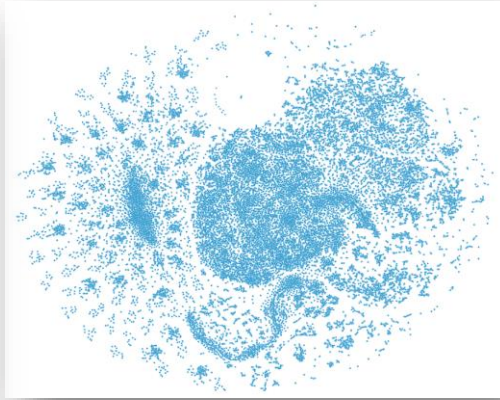
Core-S2L1C-SSL4EO

t-SNE visualization of embeddings models for Poland area

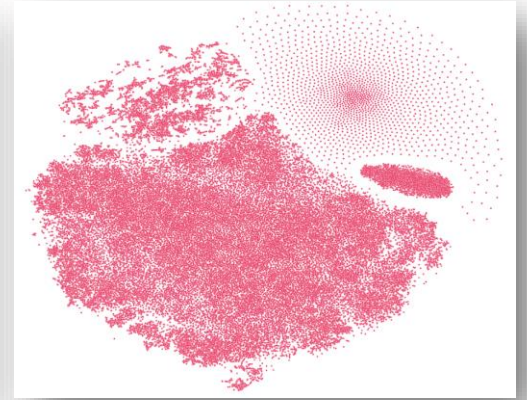
SigLIP



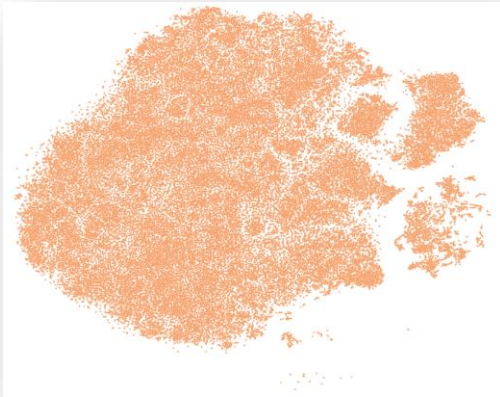
SSL4EO-S2



SSL4EO-S1



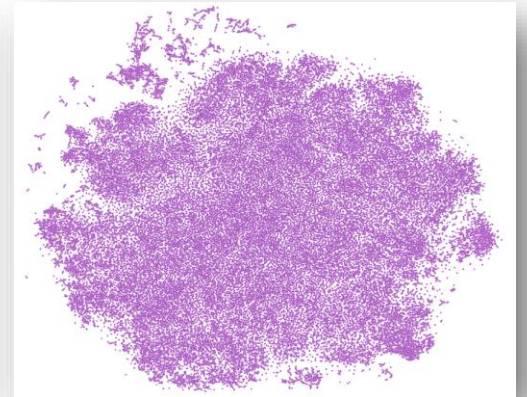
DINOv2



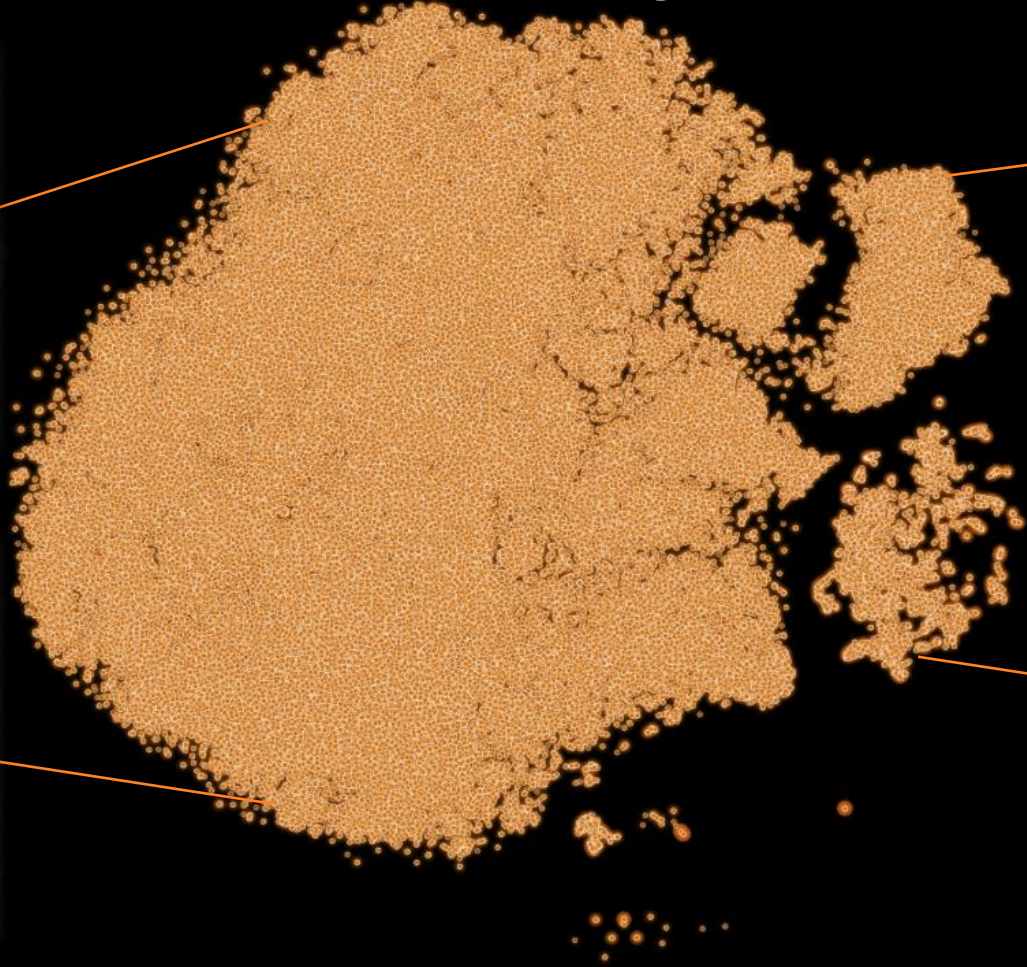
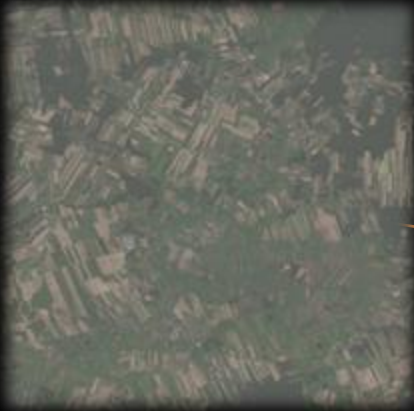
DeCUR-S2



DeCUR-S1



DINOv2 t-SNE visualization of embeddings for Poland area



Text-to-Image embedding inference using SiGLIP

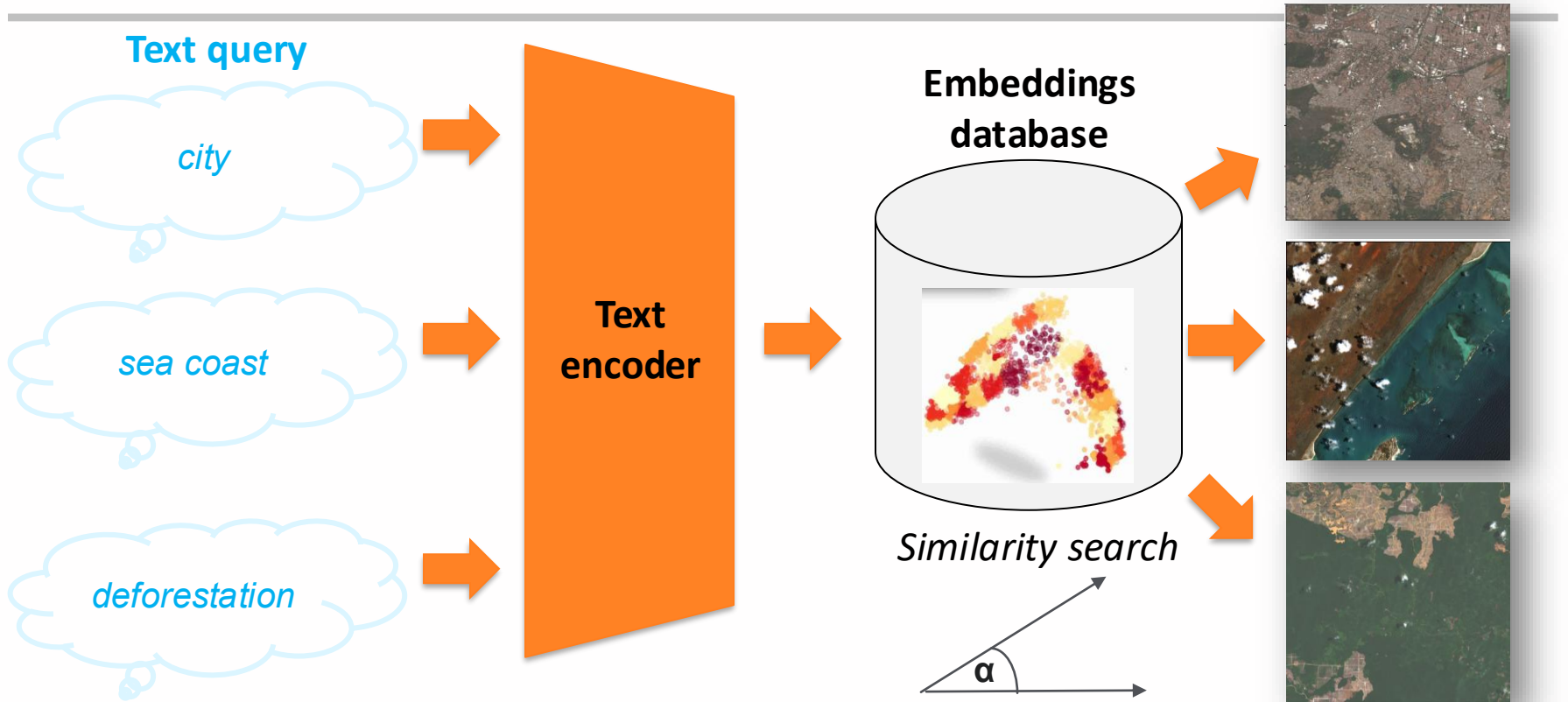
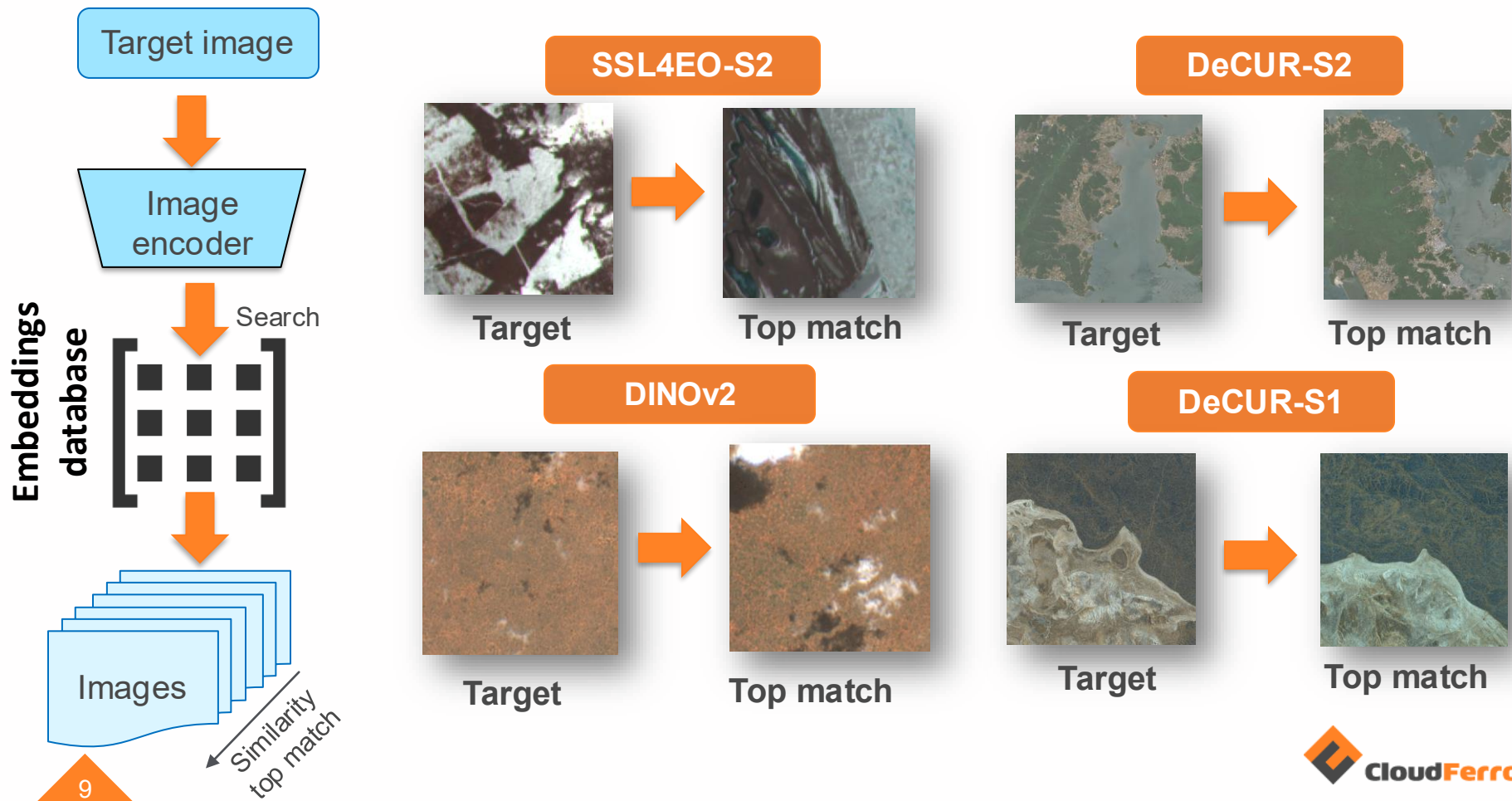
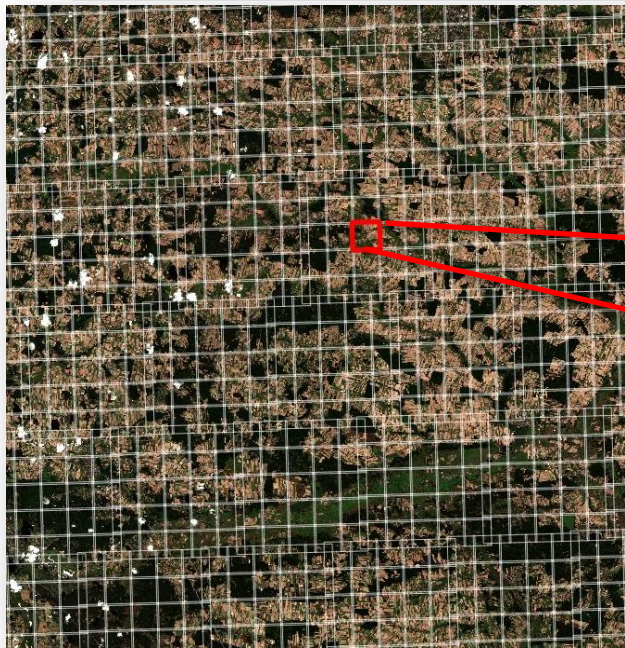


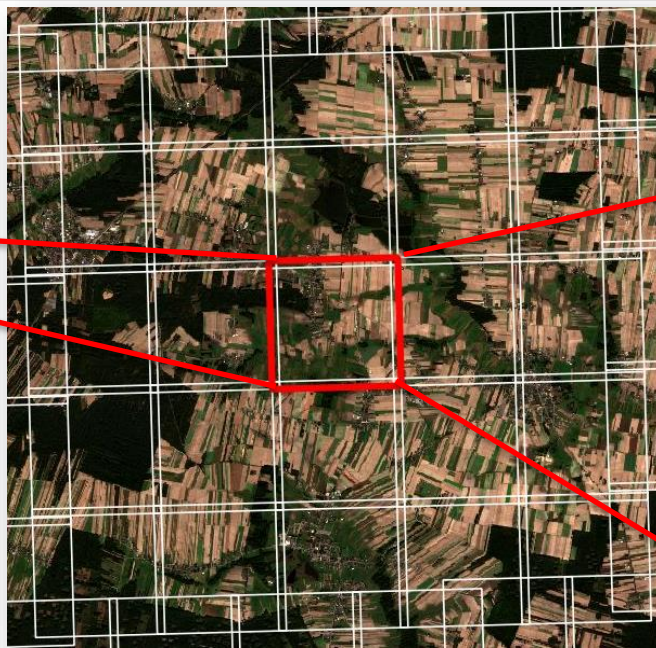
Image-to-image embeddings inference using foundation models



Embeddings with easy access and retrieval



Sentinel-2 L2A
Date: 2022-06-28



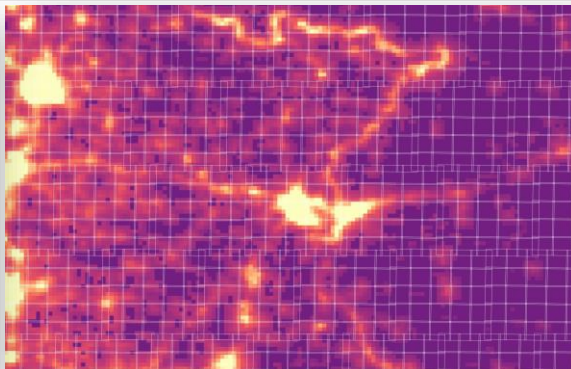
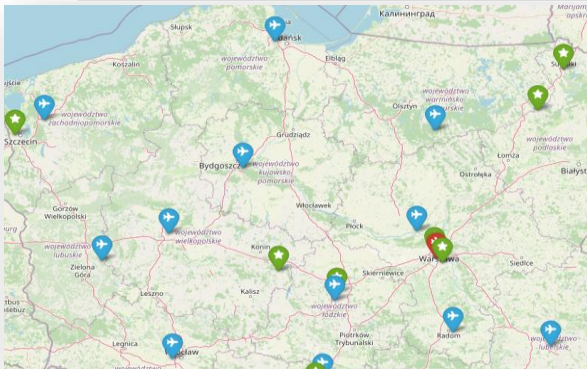
Grid cell: 576U_153R



Fragment index: (3, 2)

Embedding
array([-0.89451283, -
1.1735142...])

What's next: planned use cases and evaluations

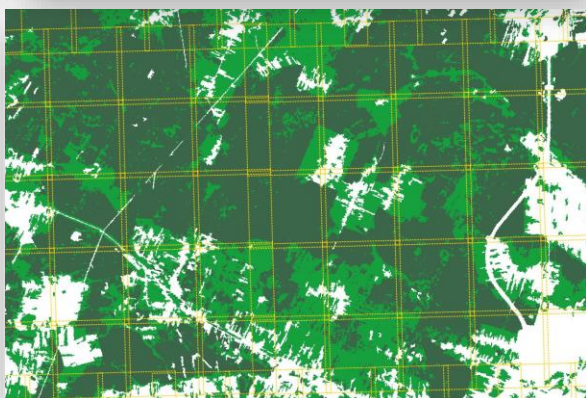
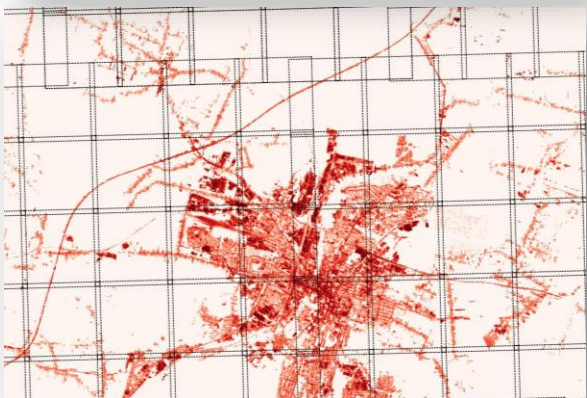


OpenStreetMap
object search

VIIRS nighttime light
modelling

HRL imperviousness
modelling

HRL forest type
classification

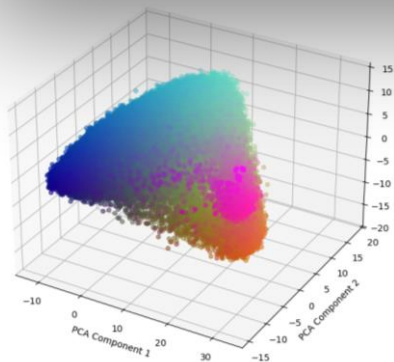


Integration of embeddings on infrastructure

```
import geopandas as gpd

s3_variables = {"endpoint_url": "https://eodata.cloudferro.com",
               "public_key": "<INSERT YOUR PUBLIC KEY>",
               "secret_key": "<INSERT YOUR SECRET KEY>"}

df = gpd.read_parquet('s3://EODATA/auxdata/MajorTOM/embeddings/
Core-S2RGB-DINov2/part_00001-00100.parquet',
                    storage_options = s3_variables)
```



Scan to learn more



CREODIAS

Data Data Explorer Cloud Pricing

CloudFerro Cloud Open Telekom Cloud

Computing:

- Computing Services
 - Virtual Machines
 - Spot Virtual Machines
 - Dedicated Servers VM
 - GPU Services
 - Bare Metals
 - Kubernetes
- Storage
- Additional Services
- CloudFerro Services status page





www.cloudferro.com

Marcin Kluczek
mkluczek@cloudferro.com



facebook.com/cloudferro



linkedin.com/company/clfr/



twitter.com/CloudFerro