

ESA's 2025 Big Data Foundations for Earth Observation Training Course

Building AI models for EO analysis

[NAME]
[ORGANISATION]
[DATE]



European
Commission

researchLatvia[★]
Value Through Knowledge



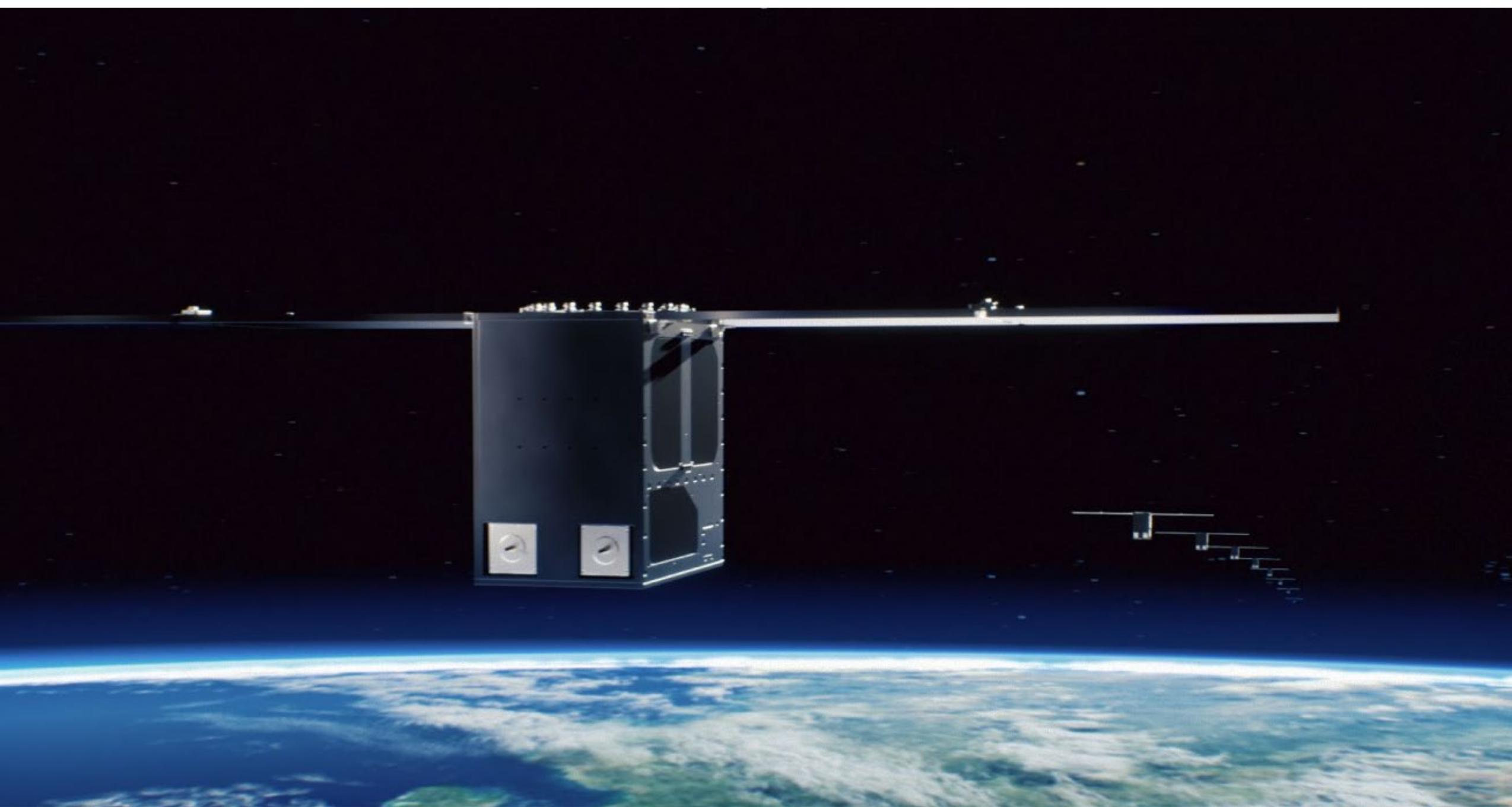
Ministry of
Education and Science
Republic of Latvia





THE CHALLENGE

Lack of continuous and affordable decision-ready planetary intelligence to manage climate change impact, food security and national threats effectively.

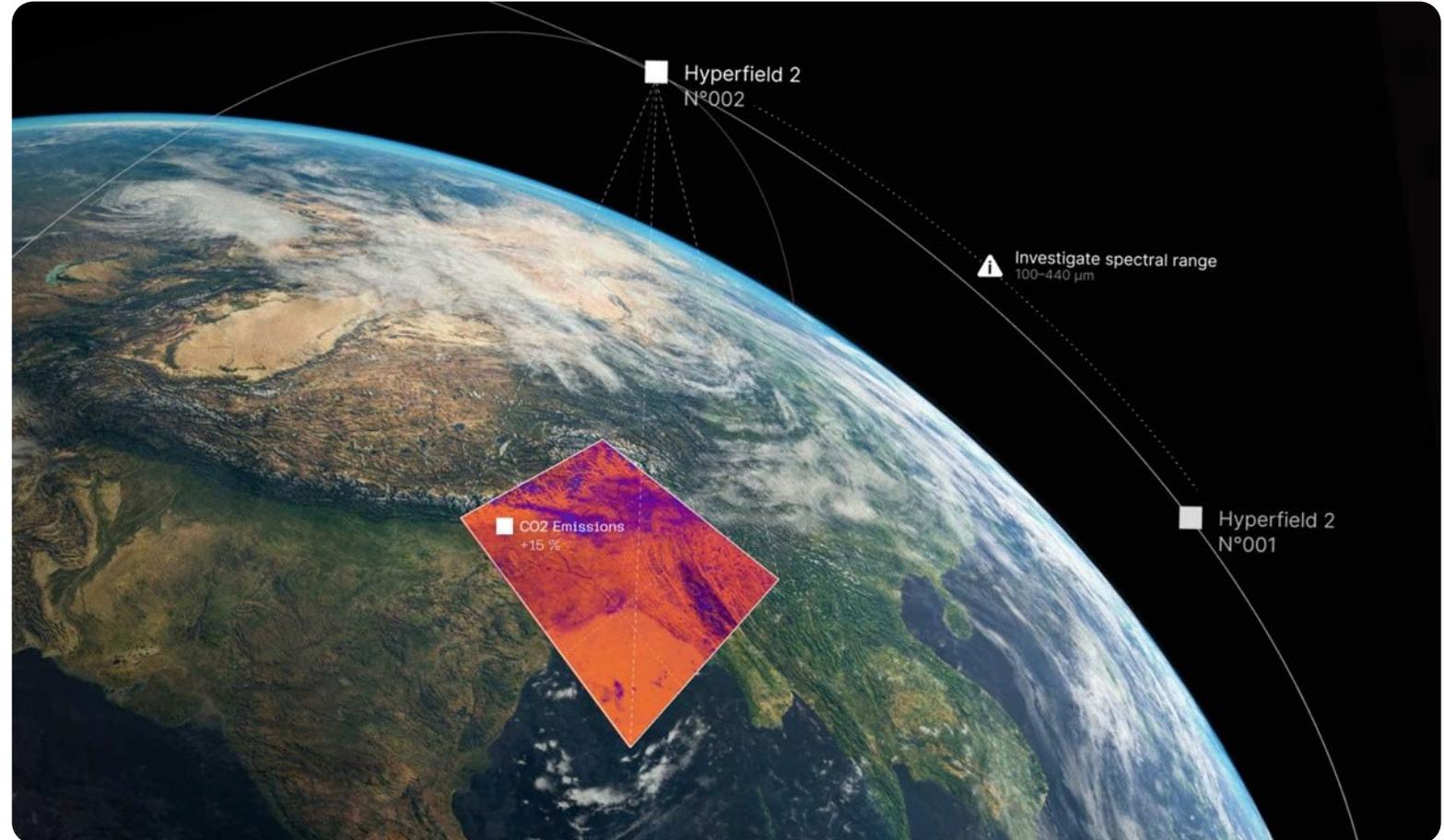


Cutting-edge space technology and AI

THE WORLD'S MOST EXTENSIVE HYPERSPECTRAL SATELLITE CONSTELLATION

Kuva Space is revolutionizing Earth observation services by delivering near real-time intelligence for various industries with robust end-to-end deep tech solution

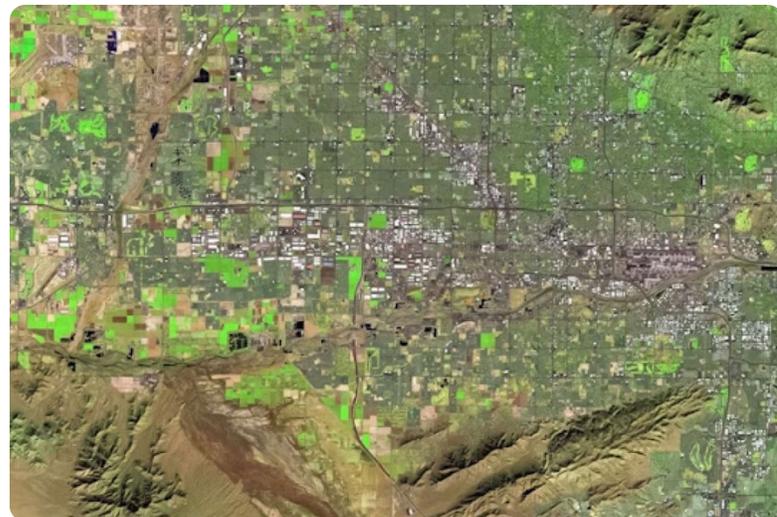
- **100 scientific-grade** microsatellites captures global hyperspectral data subdaily
- **Patented in-orbit tunable spectral camera** enables high data quality
- **Automated AI-powered analytics** streamlines processing workflow





MARKET SIZE

Targeting 235 M€ ARR by 2030



Food security

Accurate global yield forecasts and crop growth and health analysis

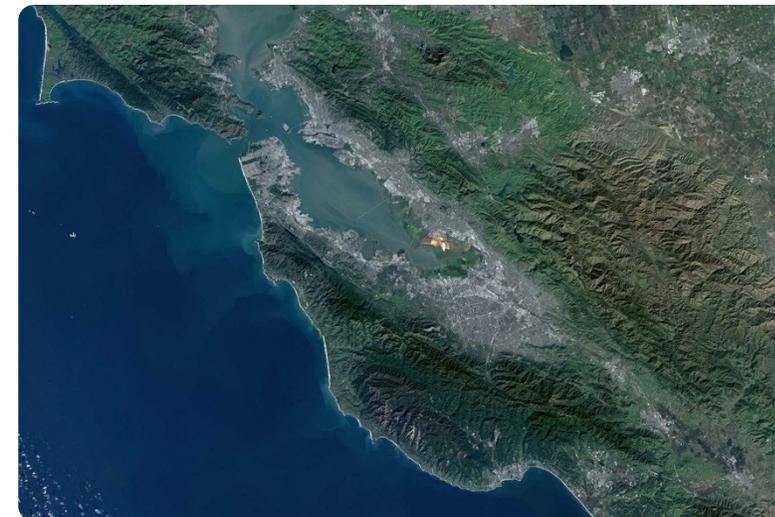
SAM: 4 B€



Carbon and environment

Reliable carbon sequestration verification and environmental monitoring

SAM: 800 M€



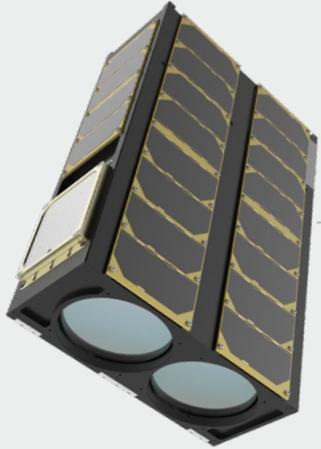
Safety and security

Automatic change detection of any material on land and at sea

SAM: 1 B€

Hyperfield constellation

SCIENTIFIC-GRADE HYPERSPECTRAL MICROSATELLITES FOR HIGH REVISIT, LARGE AREA COVERAGE, AND SUBPIXEL DETECTION ACCURACY



HYPERFIELD-1A (LAUNCH Q3/2024)
HYPERFIELD-1B (LAUNCH Q2/2025)

HYPERFIELD -1 , 1st Generation

- Scientific grade instruments in 6U CubeSat size
- Mission critical systems built by Kuva Space
- Versatile imaging modes with a patented tunable snapshot VNIR imager



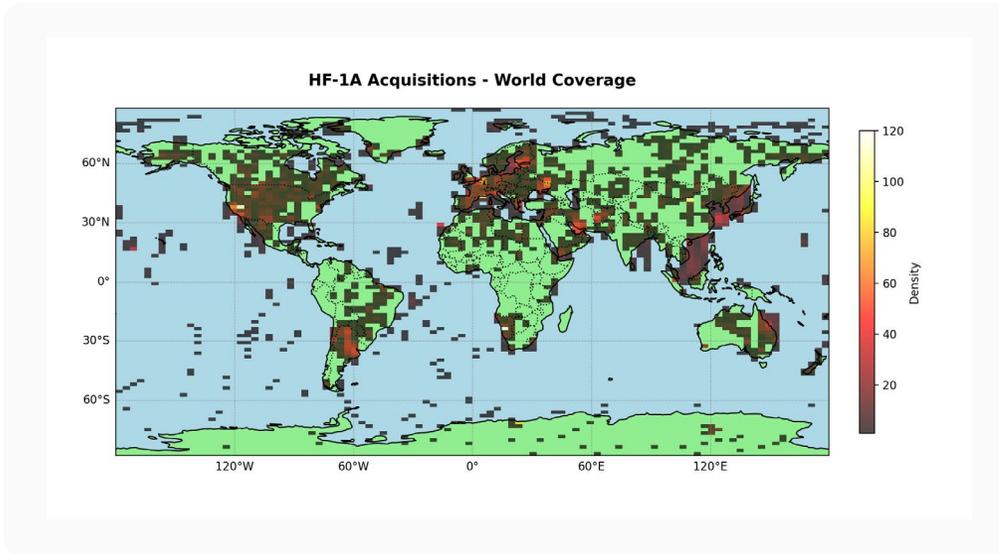
HYPERFIELD-2A (LAUNCH 1H/2026)
HYPERFIELD-2B (LAUNCH 1H/2026)

HYPERFIELD-2 , 2nd Generation

- SWIR and RGB (3 meter) imagers
- Propulsion system for precision orbital operations
- Sat-to-Sat communication
- GPU on board
- AIS enabled

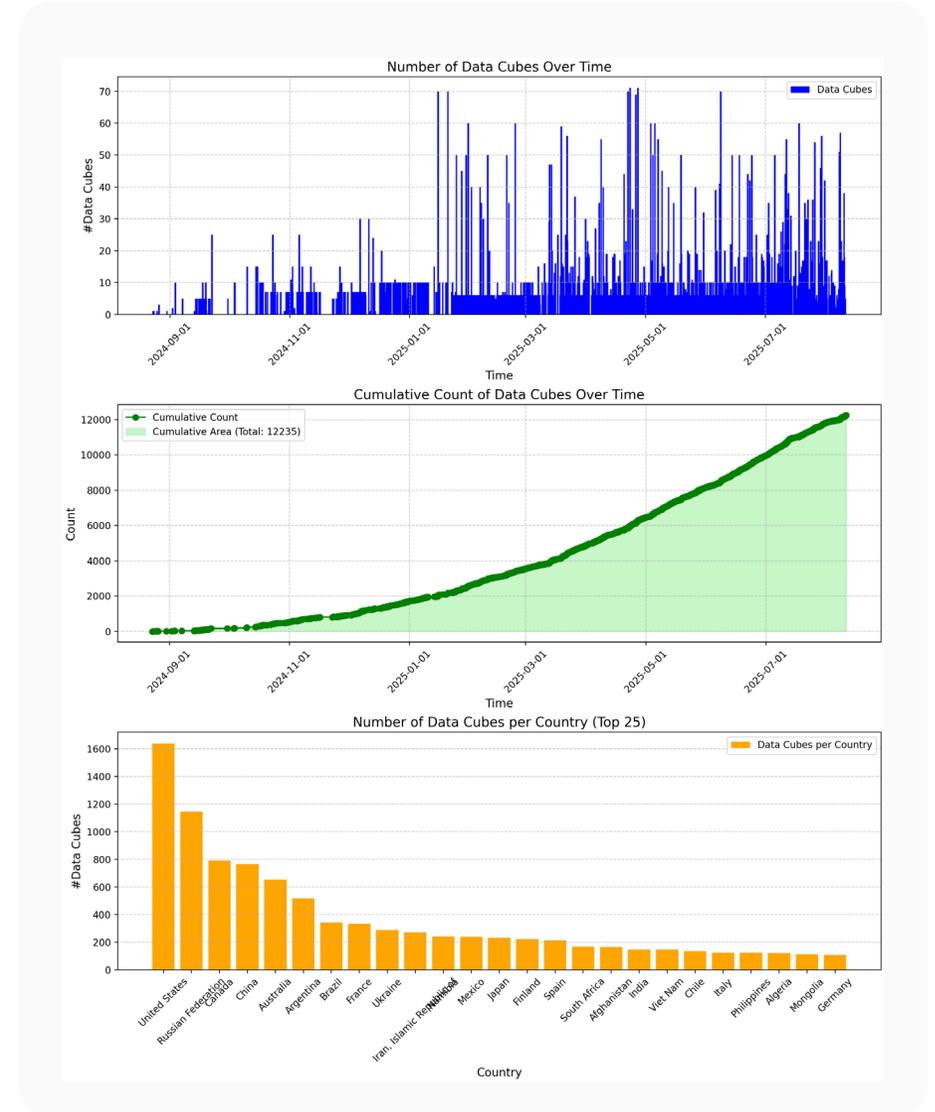
Hyperfield-1A data collection

UPDATE: 13TH AUG 2025: 12,235 ACQUISITIONS



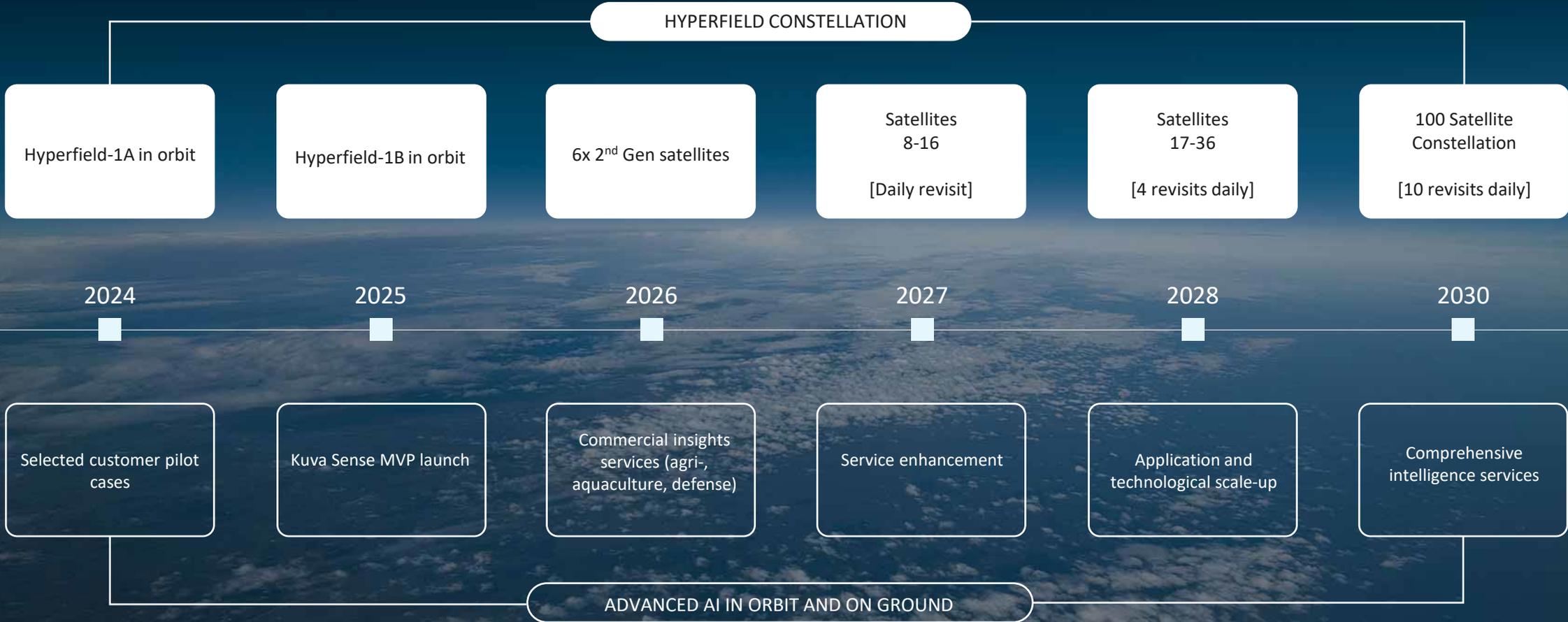
FLAG VALUE	PERFORMANCE	DESCRIPTION
Spectral accuracy	0.5%	Relative to the FWHM
Radiometric accuracy	10%	In comparison to RadCalNet
Signal-to-noise	100	Scanning mode. Tailored 100-600 SNR in pointing mode
Geometric accuracy (band-to-band co-registration)	0.2 pixels	
Spatial accuracy	1.1 pixel	CE90

* Confirmed by ESA Mission Performance Cluster (MPC)



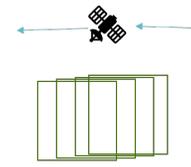
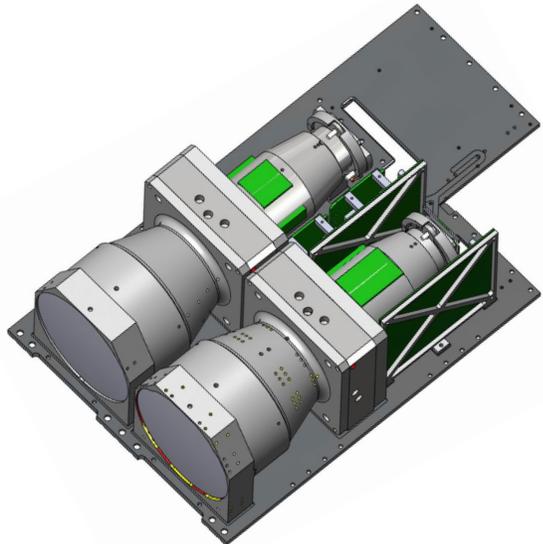
Solution roadmap

GLOBAL RAMP UP OF TECHNOLOGY AND SERVICES



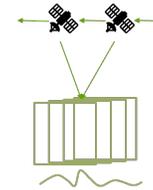
Unique hyperspectral imaging modes with proprietary camera technology

- Access to continuous data collection over large areas
- Access to tailored data acquisition in dedicated AOIs



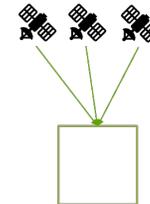
SCANNING

Continuous data acquisition
Suitable for monitoring changes over large areas and collecting archive data



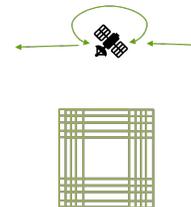
3D DSM/DTM

Tasking, Stereo data collection.
Enhance feature extraction and recognition.



OPTIMIZED STARING

Tasking. Selected targets
AI-driven data acquisition and band selection
Enhanced resolution and quality
Maximized quality and detection sensitivity for most challenging applications



OPTIMIZED SCANNING AND RESOLUTION

Tasking. Enable super-resolution
AI-driven data acquisition
Best quality data for a selected application



ADVANCED AI-POWERED ANALYTICS

Decision-ready insights in 1-2 hours



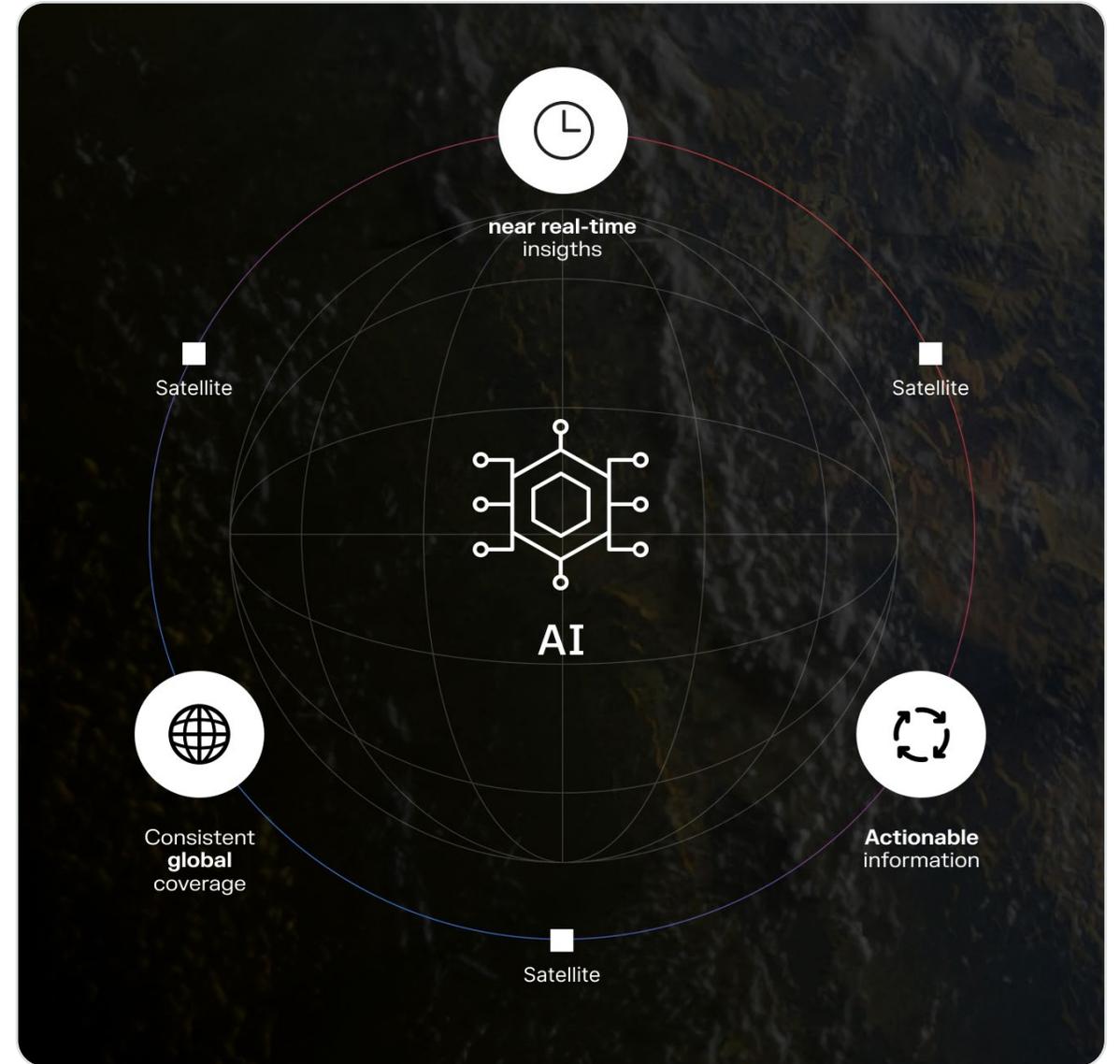
On-board processing for time-sensitive alerts and data products in less than 2 mins, cloud detection and compressions for efficient download



Automated AI-powered processing, integrating 3rd party data to generate intelligence for tens of global applications in less than 2 hours



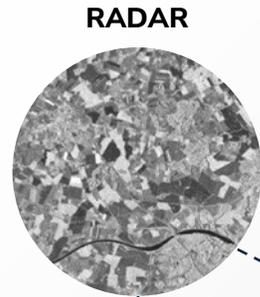
Explainable AI to enforce full transparency for automatic image capture and reliability of insights



The **missing** spaceborne capability

Radar

- Provides texture information, does not distinguish different materials or identify their states

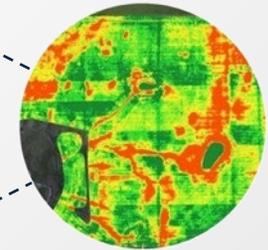


Optical / multispectral

- Low numbers of spectral bands - does not reveal changes in biochemical and physical states



HYPERSPECTRAL



Hyperspectral

- Identifies materials (physical and biochemical) accurately

Reveals the invisible

Automatic detection

Rapid Identification

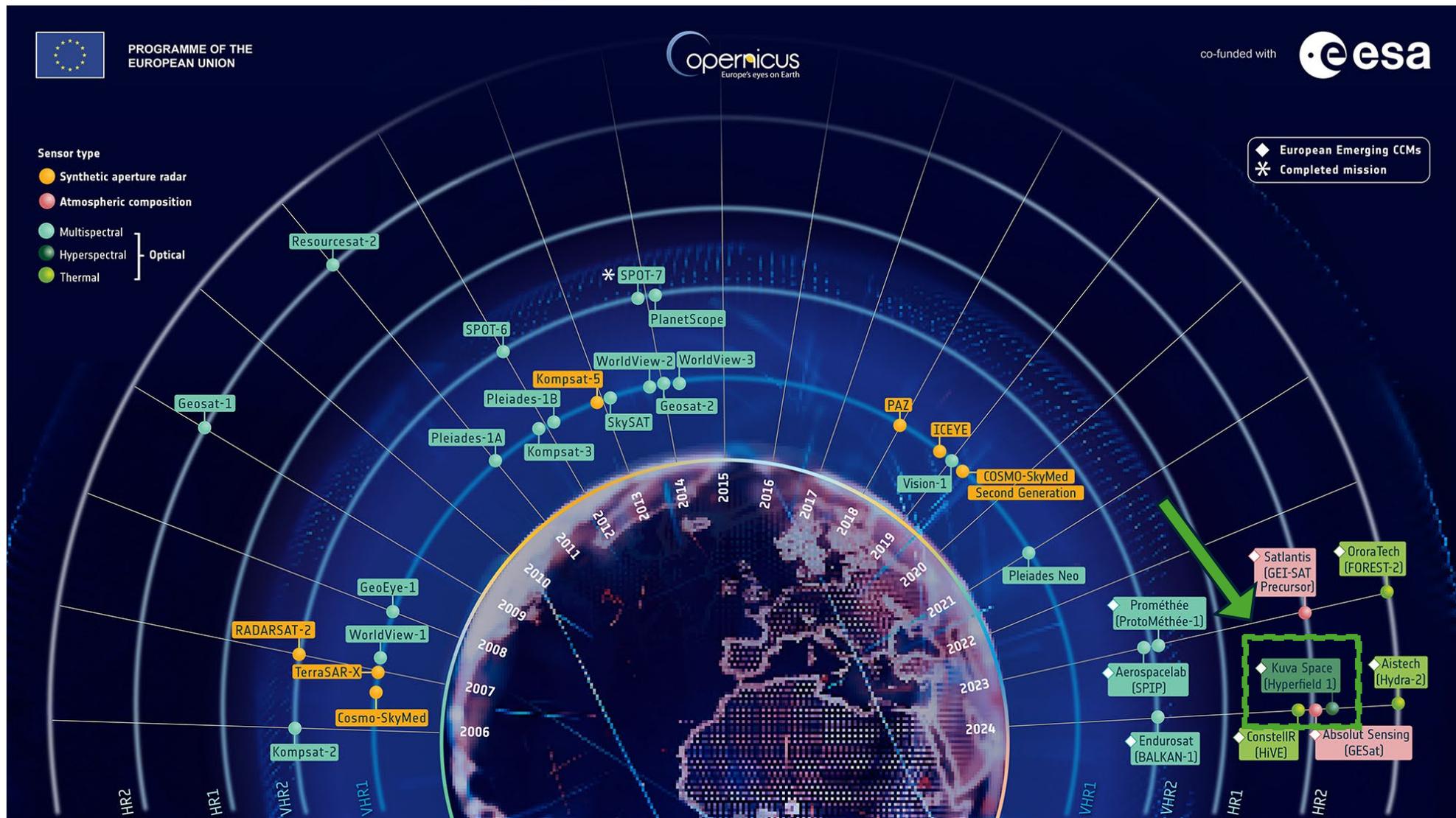
Monitoring

Change detection

Development analysis

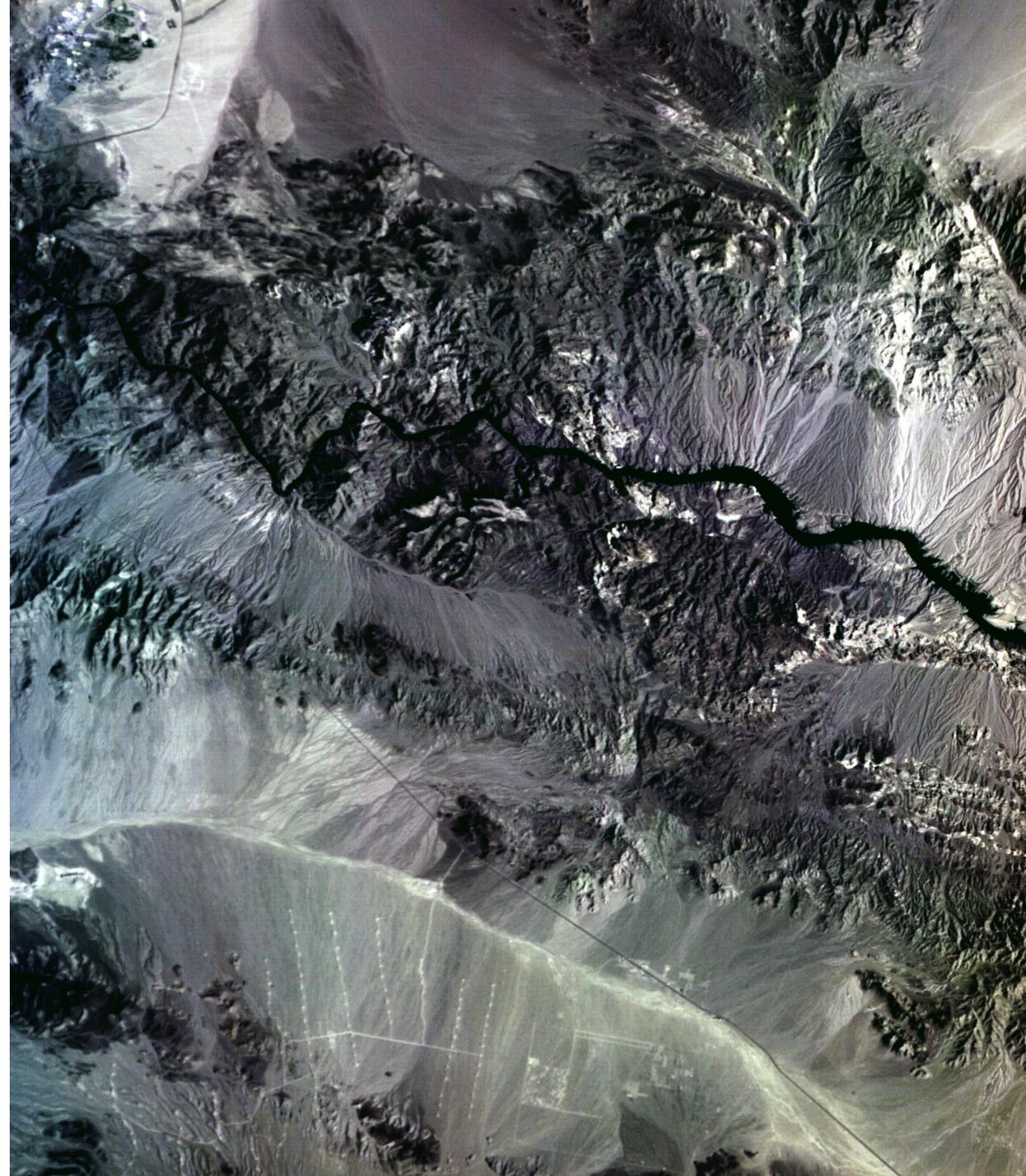
Applications: crop analysis and forecast, automatic change detection, carbon accounting, environmental monitoring

Kuva Space in the CCM Ecosystem



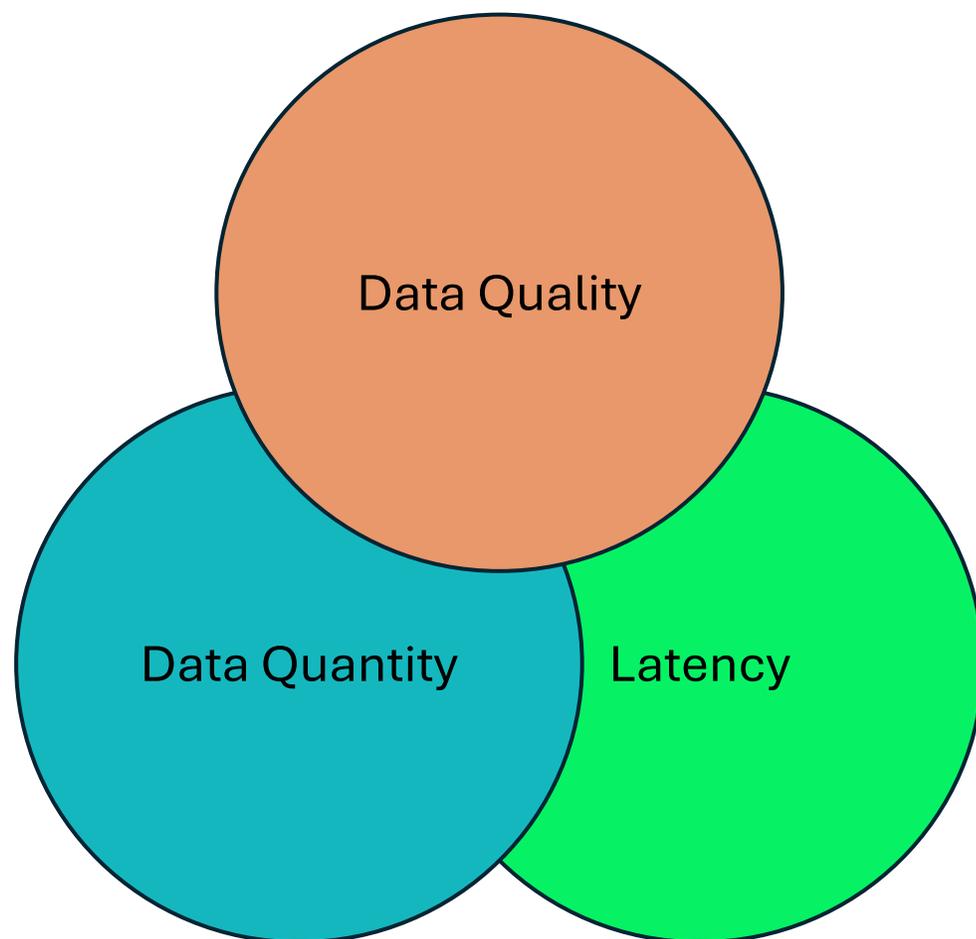
How much data are we talking about?

- Near future (2026) realistic estimation:
 - 2 x HF1 + 6x HF2 = 8 active satellites
 - Daily revisits for selected locations
 - Duty cycle = 100 cubes / day / satellite
 - 200 GB of imaging data / day (~250MB/cube)
 - **73 TB of imaging data / Y**
- Distant future (2030) realistic estimation:
 - 100 active satellites
 - **10x daily revisits** for selected locations
 - Duty cycle = 200 cubes / day / satellite
 - 8 TB of imaging data / day (~400MB/cube)
 - **~3 PB of imaging data / Y**





Today's key challenges in EO



- **Data Quality**
 - Intrinsic sensor performance
 - Quality of the spectral filter
 - Quality of the camera system
 - Higher SNR → More frames → Less bands
- **Data Quantity**
 - ~TB / day in not-so-distant future
 - Storage costs (cloud? On prem? Backups?)
 - Networking costs
 - Processing costs
- **Latency**
 - Strategic ground station coverage
 - Satellite-to-satellite communication

On-Board Processing for LEO Constellations

Turn raw downlink bottlenecks into actionable, prioritized insights at the edge

The stakes

- Downlink bottlenecks: sensors produce far more data than S-band/X-band can deliver
- Latency matters: tasking → capture → ground processing can be hours; edge triage can cut to minutes
- Duty-cycle & power: smarter selection/compression enables longer imaging and better coverage
- Regulatory & privacy: filter sensitive content in-orbit; send just what's needed

Competitive edge: deliver alerts, not files — faster decisions for customers

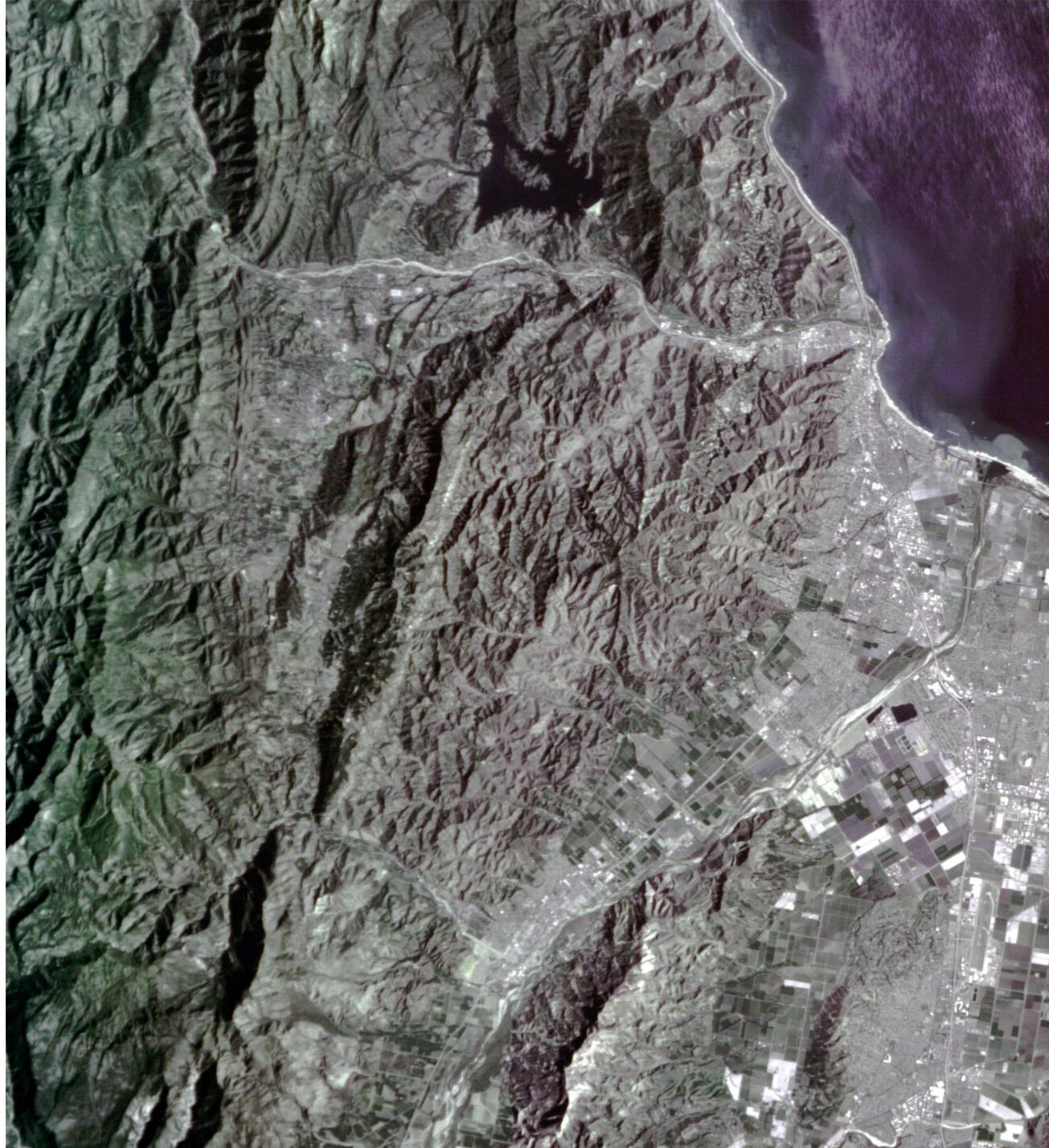
Recent advances that make it feasible

- Embedded GPUs (e.g., Jetson Orin/AGX, AMD/Xilinx Kria) with 10–100+ TOPS at <50 W envelopes
- Mature edge-AI stacks: CUDA, TensorRT, ONNX Runtime, Vulkan compute; mixed-precision quantization (FP16/INT8)
- Radiation-tolerant compute modules and thermal designs proven in recent missions
- Model compression: pruning, distillation, and tiling strategies tailored for EO workloads
- On-board codecs & fast COG/Cloud-optimized formats + ROI cropping for 10–100× data reduction

Target outcomes: 10–100× data reduction • <5 min tip-&-cue latency • Higher revisit utility • Lower \$/km² delivered

The role of AI in HS EO

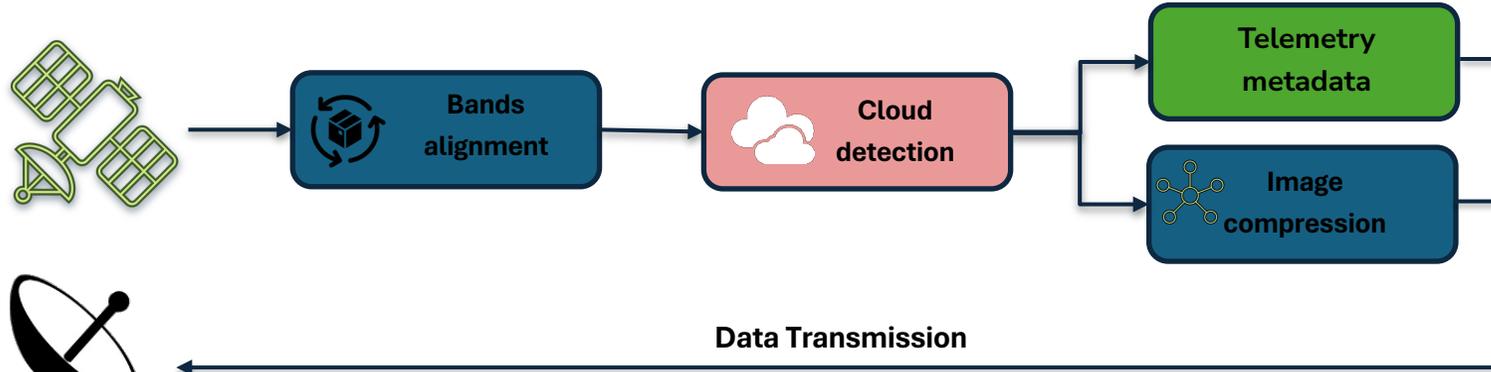
- **Enhanced data processing**
 - Deep neural networks as feature extractors
 - Excellent to cope with spectral variability
 - Ex: Band alignment → RoMa, LightGlue, etc.
- **Powerful processing frameworks**
 - Coupling between HW and SW acceleration
 - GPU-friendly image processing methods
 - Ex: Native PyTorch georef / band alignment / cloud detection / atmospheric correction
- **Actionable downstream analytics**
 - Modern DL frameworks cope well with large amounts of data (Ex: Vision Transformers)
 - Discover hidden patterns and model highly complex phenomena



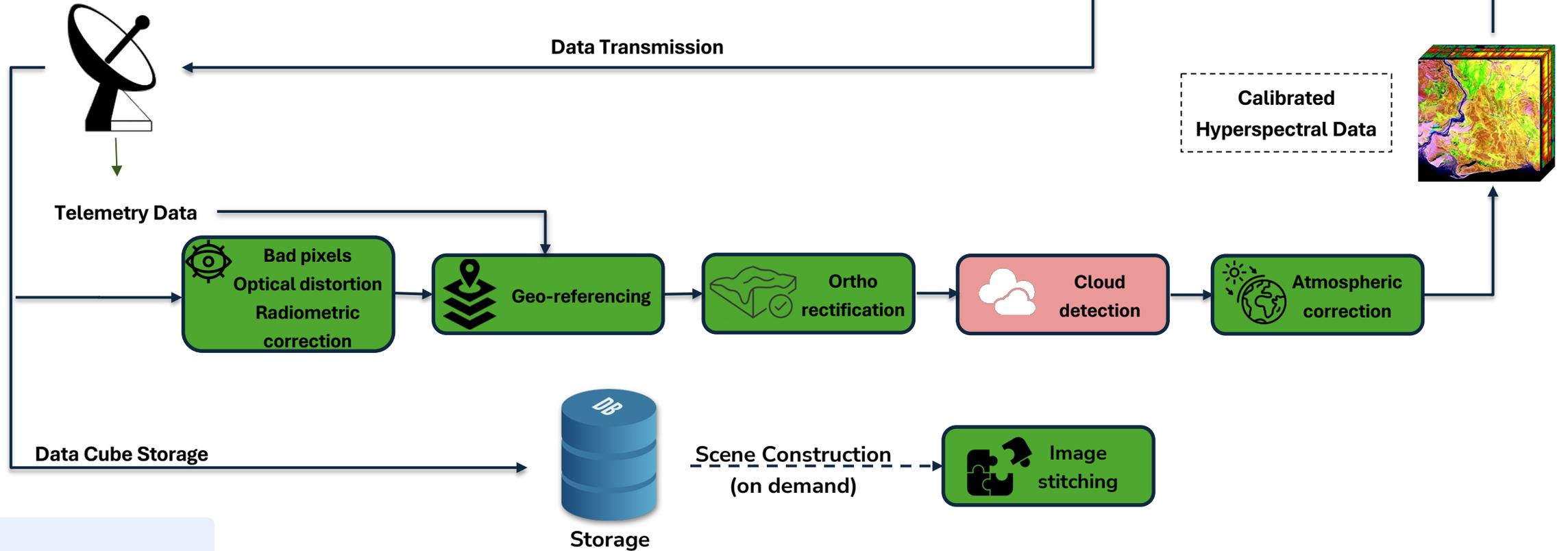
Data Processing chain

From satellite acquisition to hyperspectral product

On-Sat processing



On-ground processing



Reducing **downlink needs** using AI

High spatial / spectral res images

x

Limited bandwidth for downlink transmissions

x

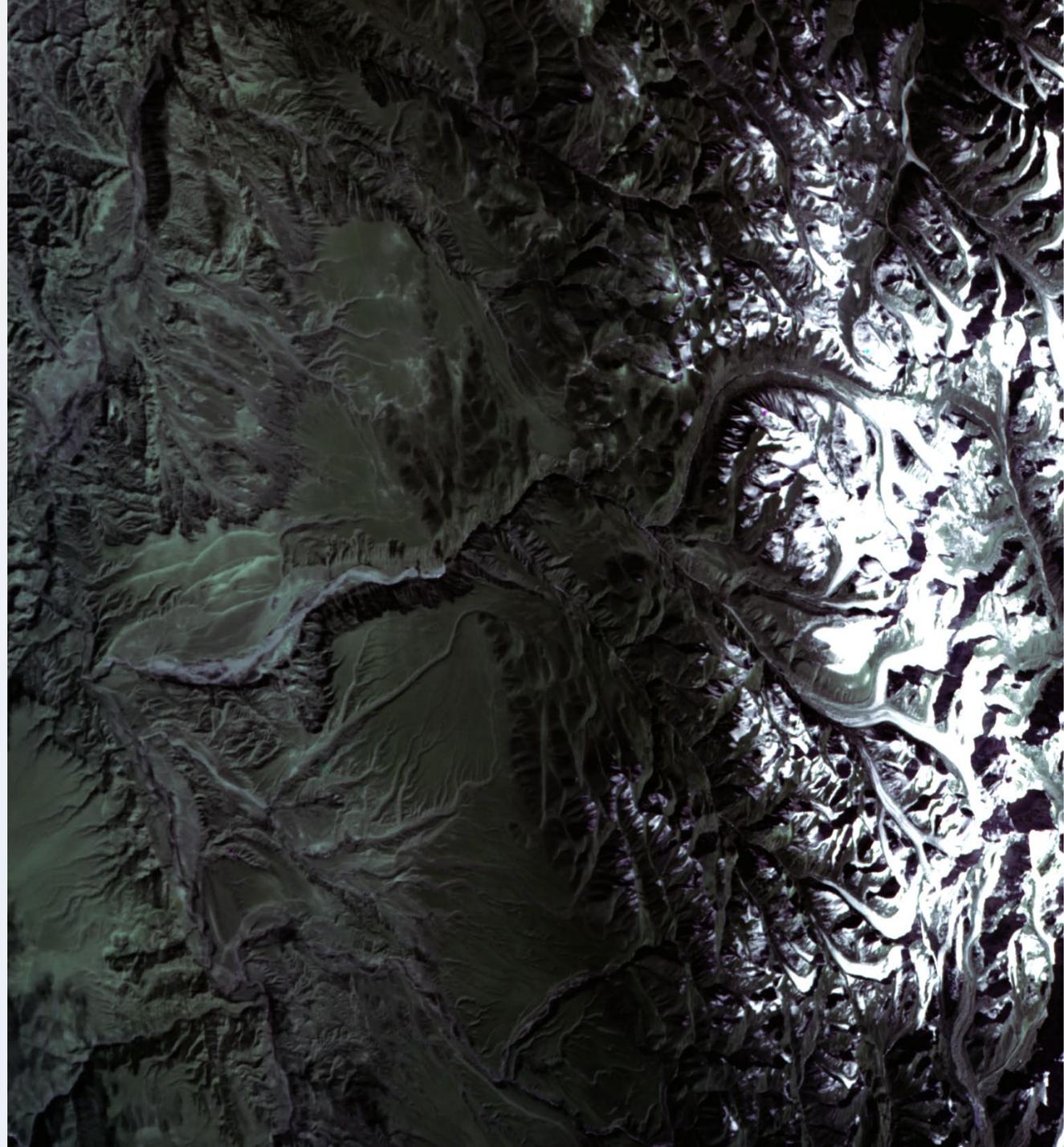
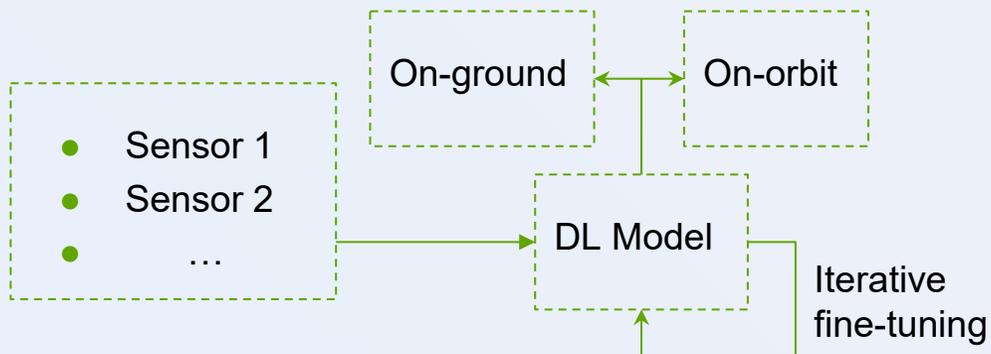
Small operational costs

Reduce the amount of data that needs to be sent to Earth by **processing hyperspectral images in orbit**.

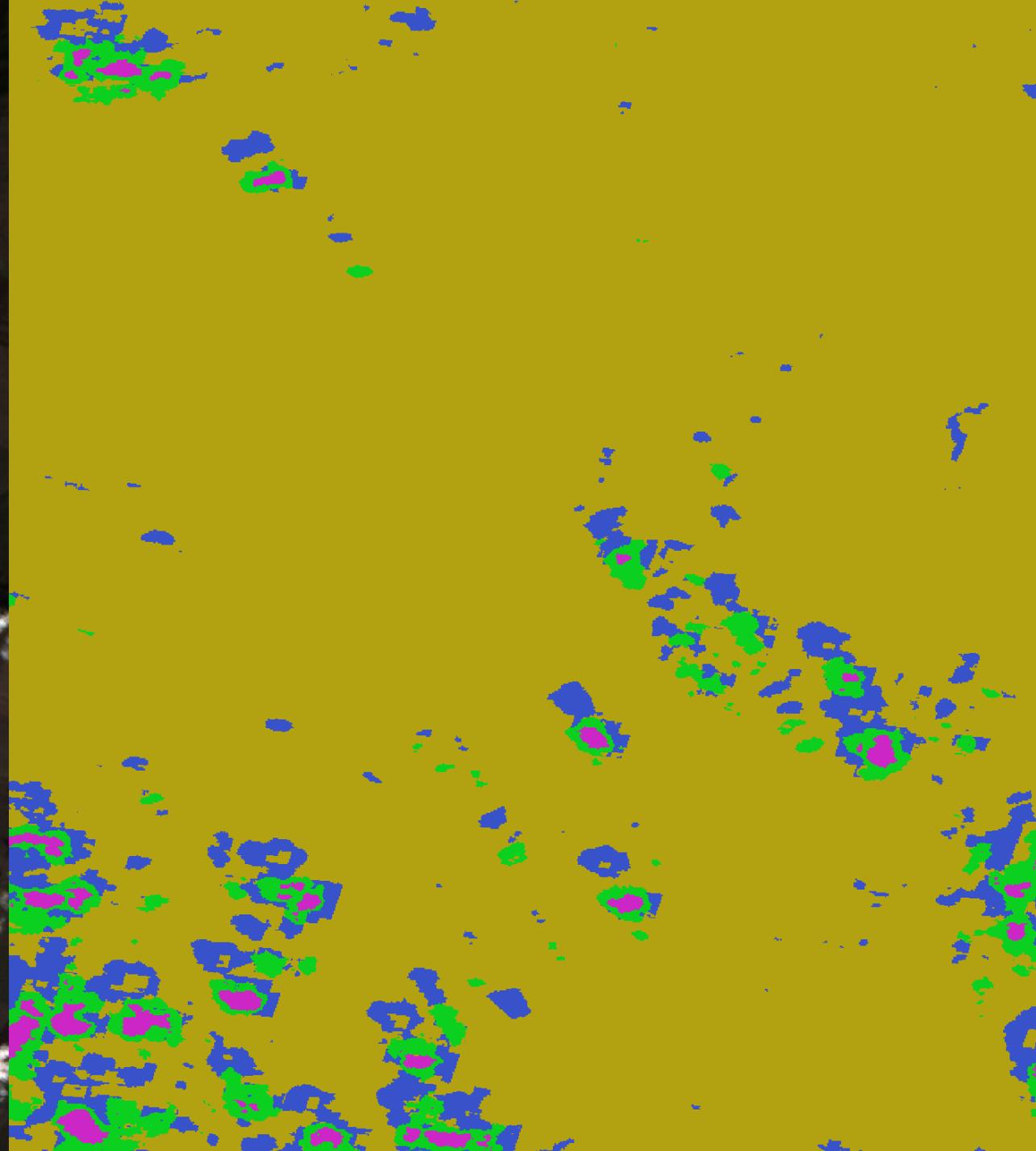
Develop an AI-based cloud detection model that can:

- **Generalize** across the constellation (robustness) ...
- use a **limited number of wavelengths** and that ...
- remain **lightweight** enough to be run onboard

Solution should come both from data and model

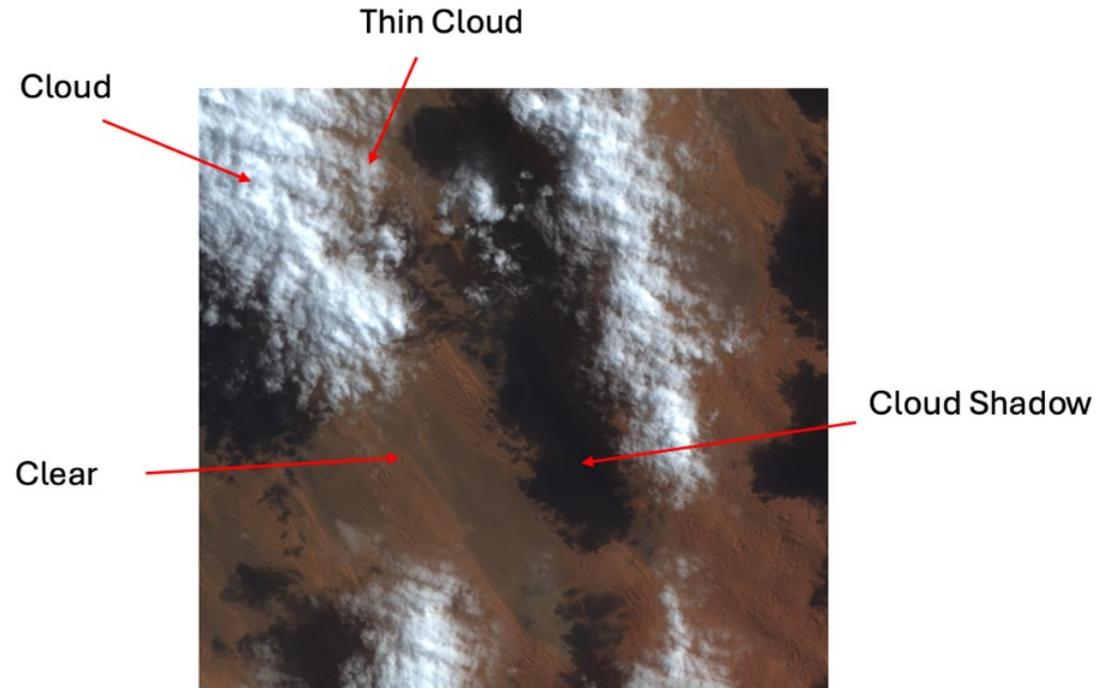
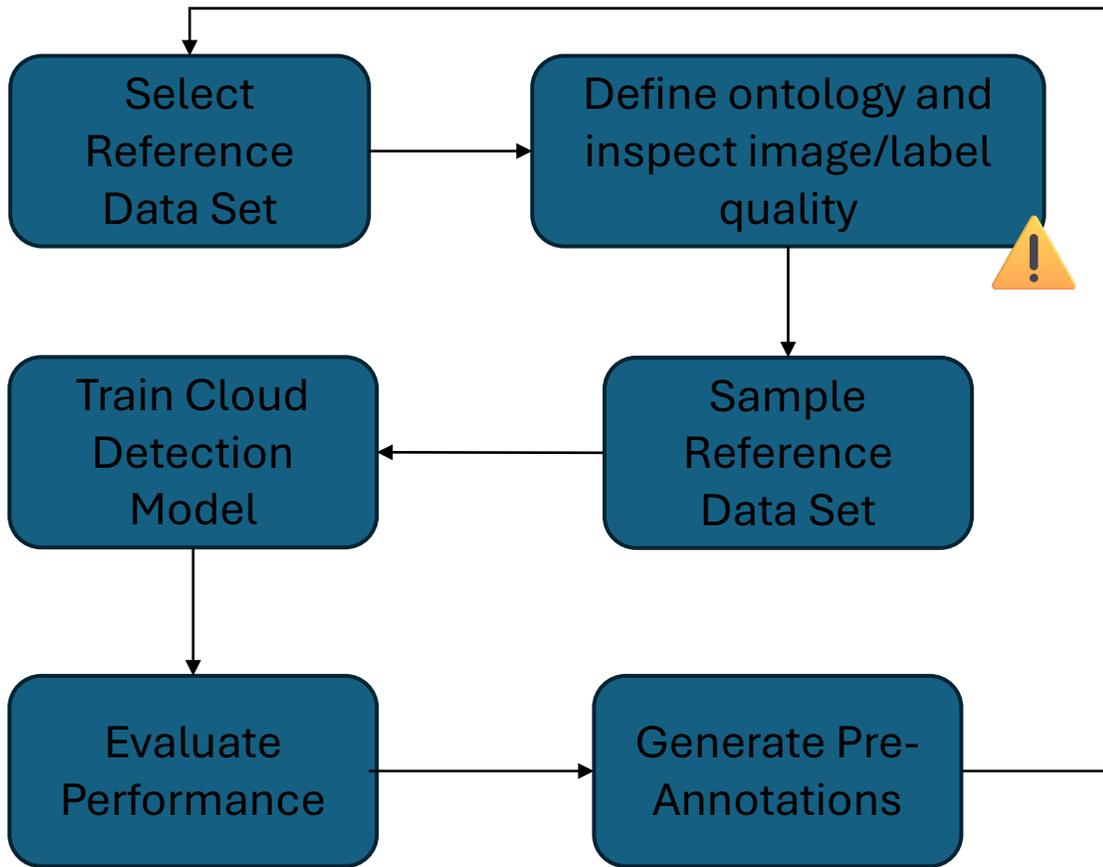


On Ground Cloud Detection



On-ground Cloud Detection

Training Workflow

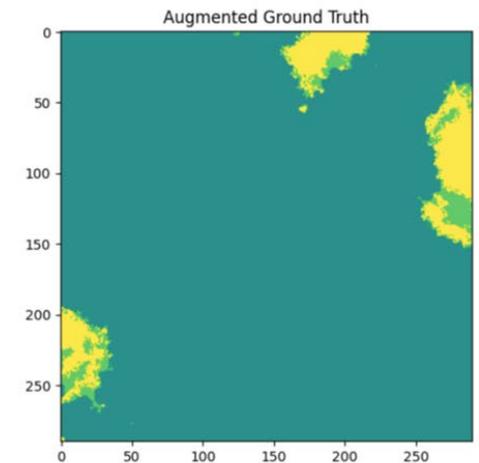
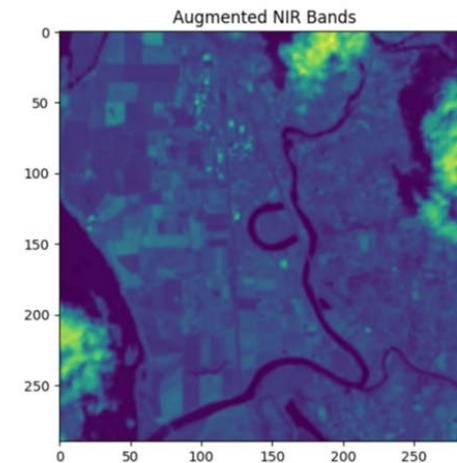
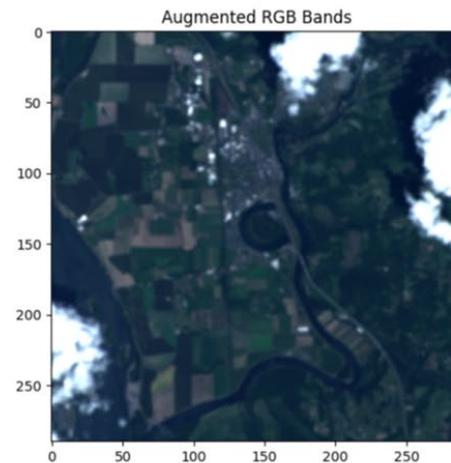
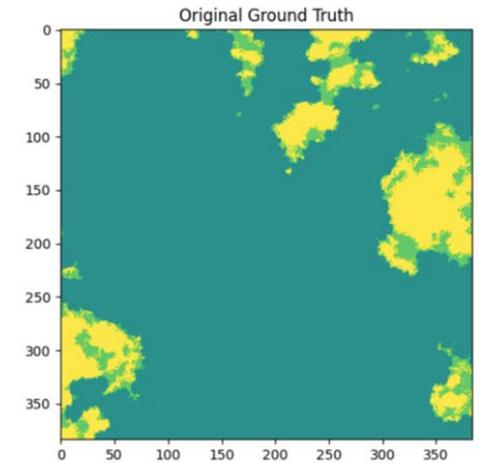
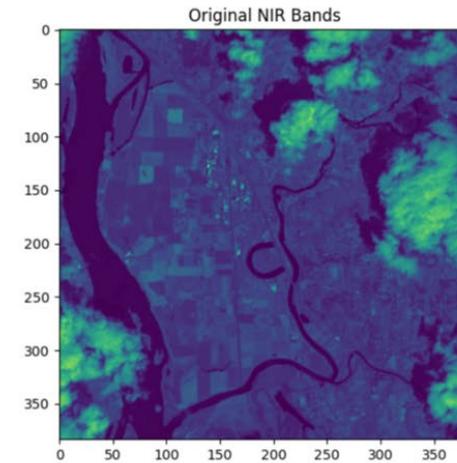
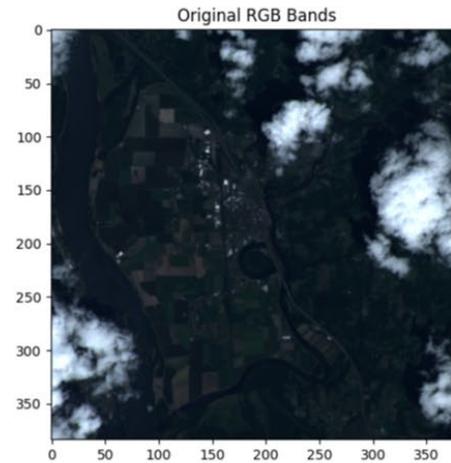


Ontology definition in [L8 Biome](#) and Encord annotation platform

On-ground Cloud Detection

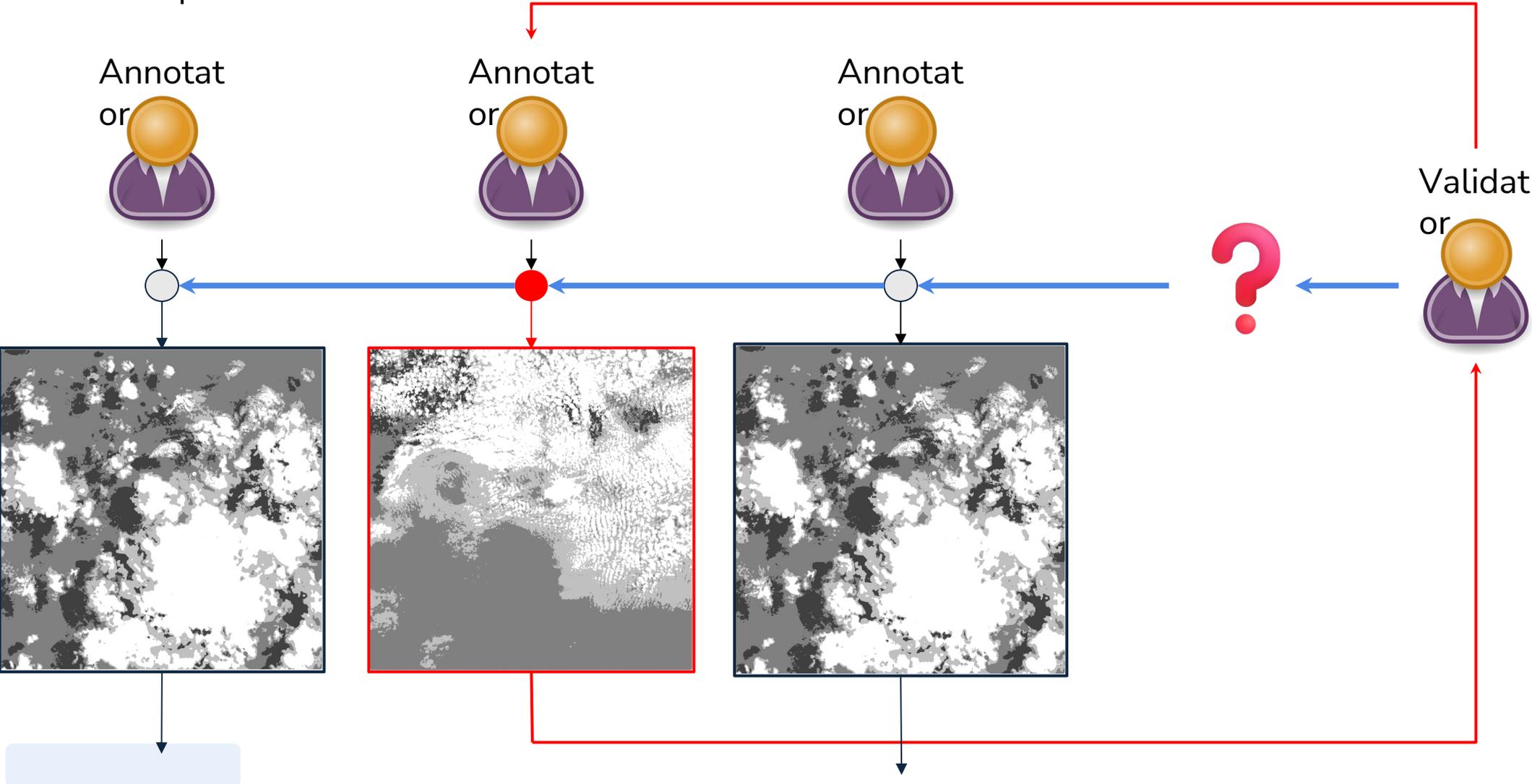
Importance of **high-quality labels**
and **rich augmentations**

- Sample of an [L8 Biome](#) image and its manual expert annotation (top).
 - Yellow = Thick Cloud
 - Light Green = Thin Cloud
 - Sapphire = Clear



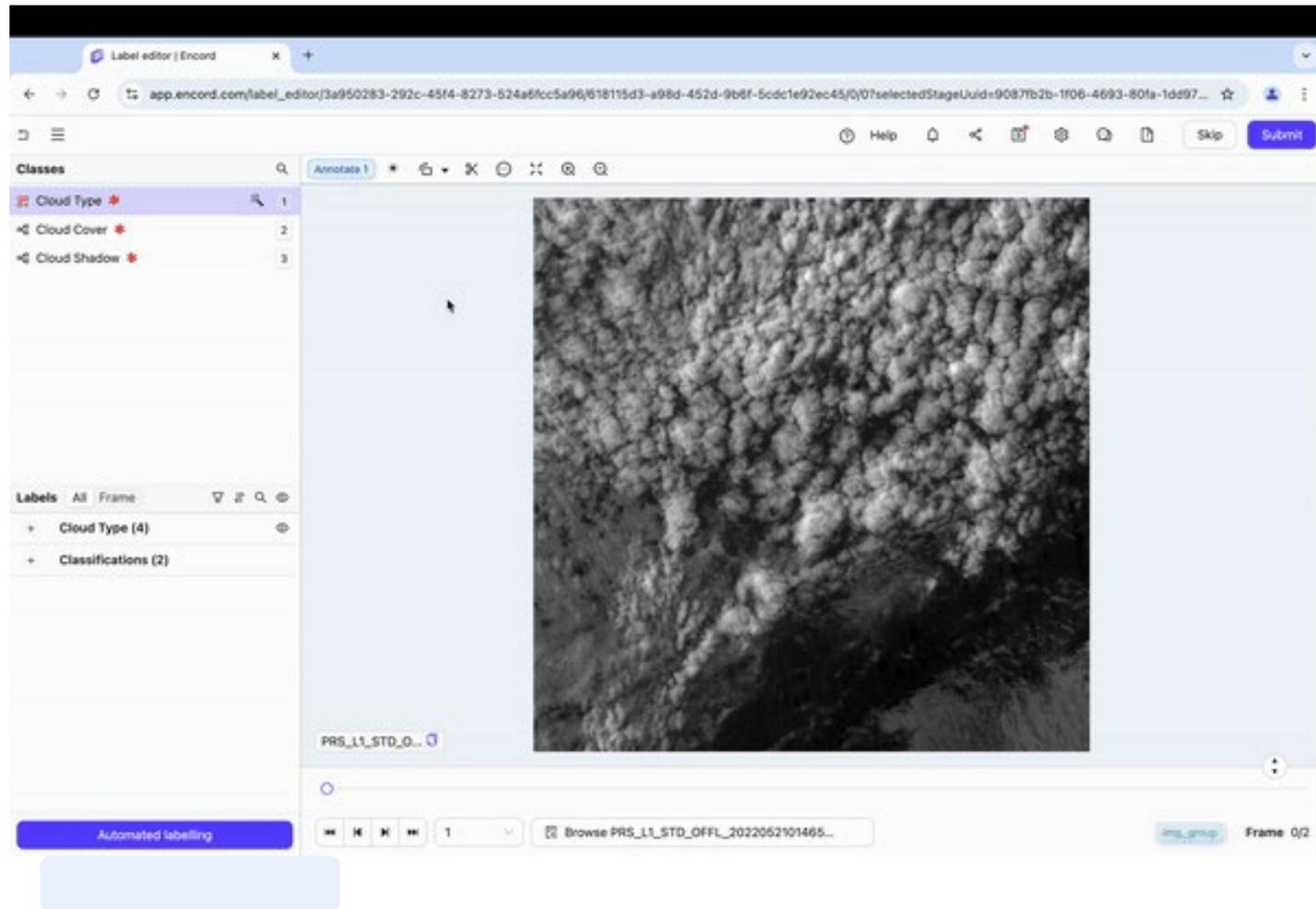
Encord Annotations

Annotation Pipeline



Encord Annotations

Thresholding the NIR bands leads to accurate cloud shadow masks



The screenshot shows the Encord Label Editor interface. On the left, a sidebar lists classes: Cloud Type (1), Cloud Cover (2), and Cloud Shadow (3). Below this, a 'Labels' section shows 'Cloud Type (4)' and 'Classifications (2)'. The main area displays a satellite image of a landscape with cloud shadows. At the bottom, there is a navigation bar with a play button, a frame counter (1), and a file path: 'Browse PRS_L1_STD_OFFL_2022052101465...'. A blue button labeled 'Automated labelling' is visible at the bottom left.

On the right, a sidebar menu is shown with the following items:

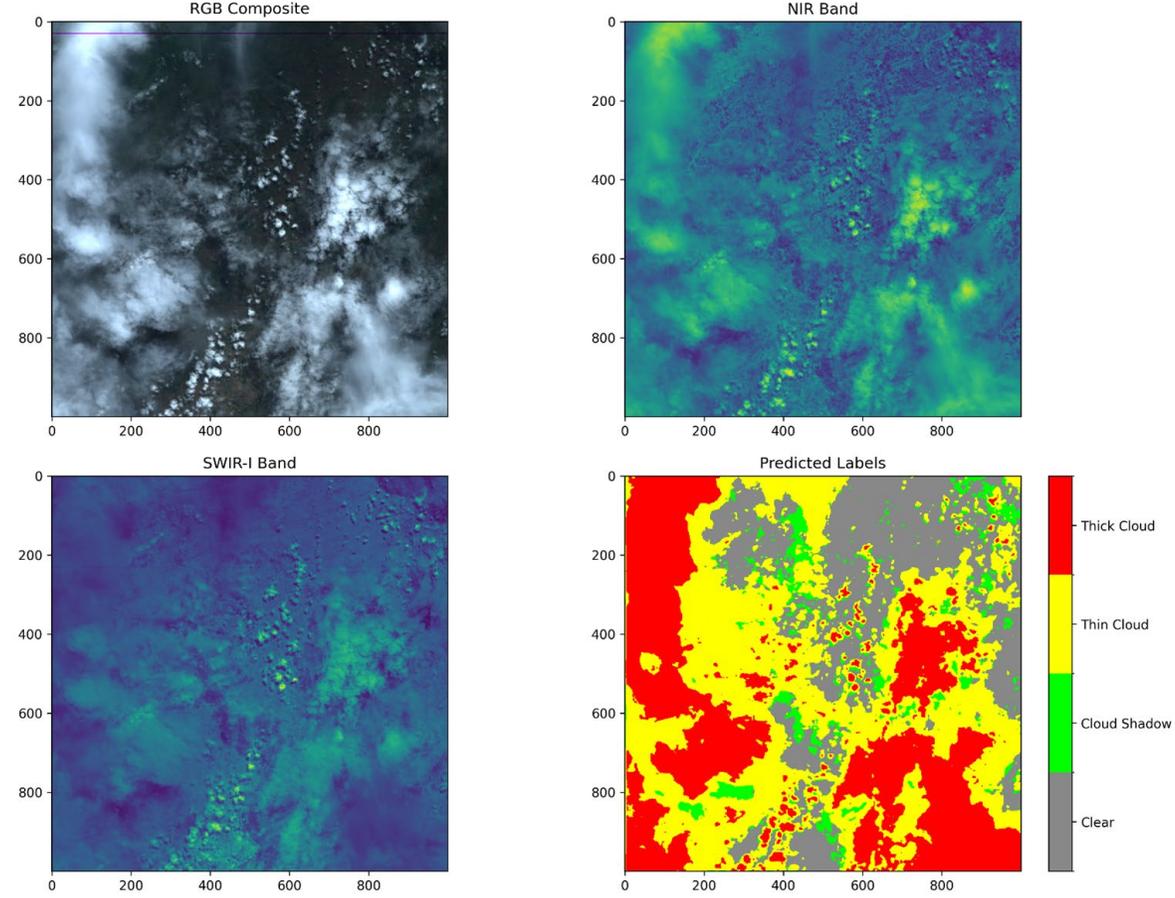
- Automated labelling
 - Classification
 - Detection and segmentation
 - Tracking and interpolation
 - Agents

Below the menu is a dropdown menu labeled 'Sync Instances Across Frames' and a blue button labeled 'Trigger agent'.

An arrow points from the 'Trigger agent' button to a dashed box containing the text: 'Agent tool allows to copy annotations across frames'.

On-ground Cloud Detection

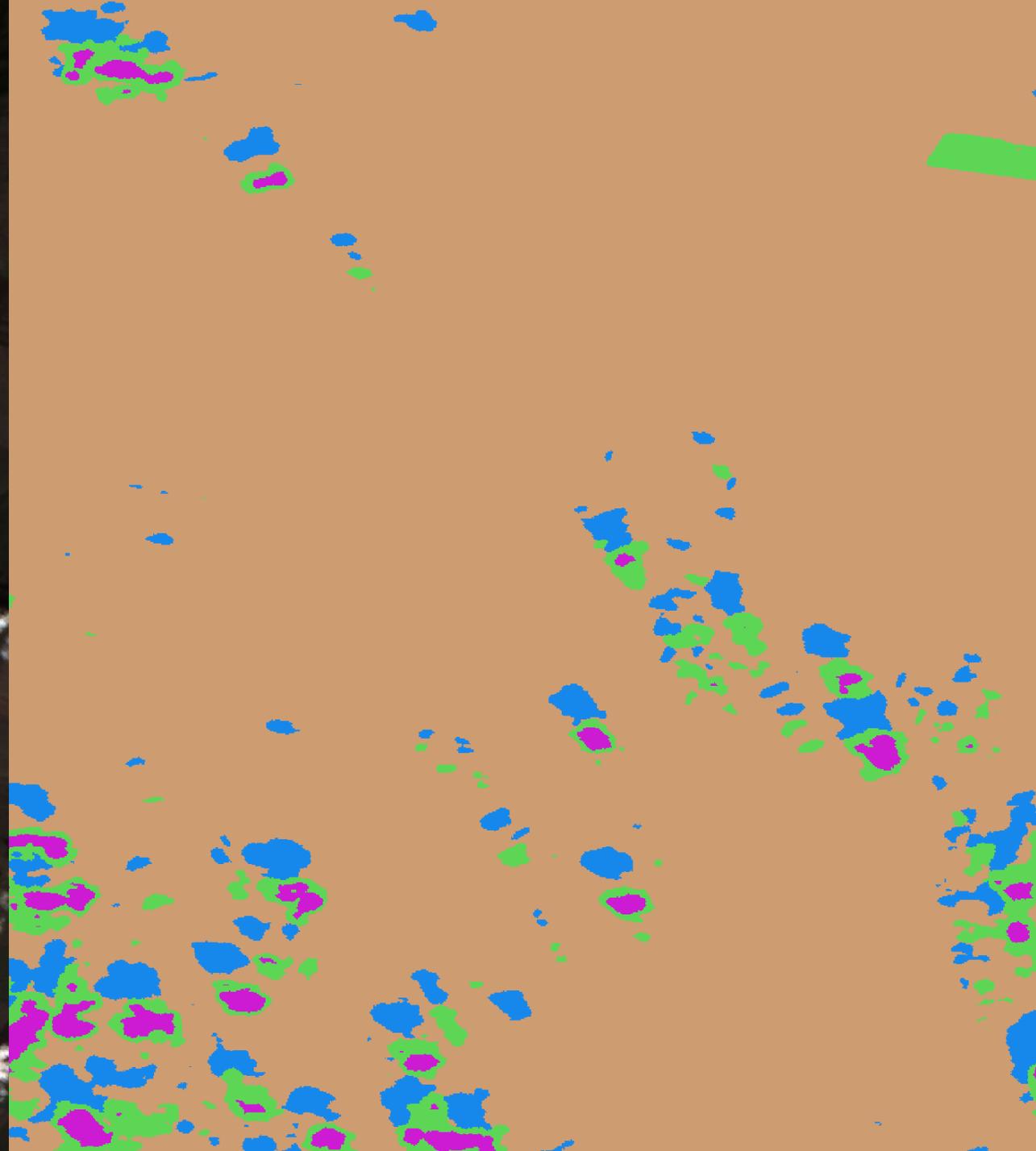
Inference on PRISMA imagery (v4)



Epoch Number	ViT version	# PRISMA images	F1-Score (%)
271	v1	N/A	92.78
374	v2	250	94.02
593	v3	500	94.22
600	v4	600	94.45

Classification performance across fine-tuning runs

On Orbit Cloud Detection



On-Orbit Cloud Detection

Making everything faster & lightweight

Repeat the training process,
with focus on:

- Model size (<1M params
→ 2MB in float16)
 - Quantization-aware training
 - Post-training quantization
(see next slide)



On-Orbit Cloud Detection

Deploying cloud detection on-orbit (CPU)

Model Preparation

- Post-Training Quantization using Intel's Neural Compressor (CPU)
 - on-orbit upload time of ~2 days
- Trade-off performance loss vs. memory footprint
- Sweet spot found around
 - 86.55% accuracy
 - 1.4MB memory footprint

```

2023-11-30 22:25:10 [INFO] Tune 86 result is: [Accuracy (int8|fp32): 0.8573|0.9042, Duration (seconds) (int8|fp32): 143.6224|152.9798], Best tune result is: n/a
2023-11-30 22:25:10 [INFO] |*****Tune Result Statistics*****|
2023-11-30 22:25:10 [INFO] |-----|
2023-11-30 22:25:10 [INFO] | Info Type | Baseline | Tune 86 result | Best tune result |
2023-11-30 22:25:10 [INFO] |-----|
2023-11-30 22:25:10 [INFO] | Accuracy | 0.9042 | 0.8573 | n/a |
2023-11-30 22:25:10 [INFO] | Duration (seconds) | 152.9798 | 143.6224 | n/a |
2023-11-30 22:25:10 [INFO] |-----|
2023-11-30 22:25:10 [INFO] Save tuning history to /home/arthur/Documents/repos/kuva-projects/topics/cloud_detection/nc_workspace/2023-11-30_12-28-39/./history.snapshot.
2023-11-30 22:25:11 [INFO] fallback [('module.lvt.module.0.1.0.mlp.fc1', 'Linear')] to fp32
2023-11-30 22:25:11 [INFO] Fx trace of the entire model failed, We will conduct auto quantization
2023-11-30 22:25:15 [INFO] |*****Mixed Precision Statistics*****|
2023-11-30 22:25:15 [INFO] |-----|
2023-11-30 22:25:15 [INFO] | Op Type | Total | INT8 | FP32 |
2023-11-30 22:25:15 [INFO] |-----|
2023-11-30 22:25:15 [INFO] | Conv2d | 33 | 31 | 2 |
2023-11-30 22:25:15 [INFO] | LayerNorm | 28 | 0 | 28 |
2023-11-30 22:25:15 [INFO] | quantize_per_tensor | 67 | 67 | 0 |
2023-11-30 22:25:15 [INFO] | Linear | 48 | 45 | 3 |
2023-11-30 22:25:15 [INFO] | dequantize | 67 | 67 | 0 |
2023-11-30 22:25:15 [INFO] | Dropout | 40 | 0 | 40 |
2023-11-30 22:25:15 [INFO] | getattr | 12 | 12 | 0 |
2023-11-30 22:25:15 [INFO] | getitem | 48 | 48 | 0 |
2023-11-30 22:25:15 [INFO] | ConvReLU2d | 1 | 1 | 0 |
2023-11-30 22:25:15 [INFO] |-----|
2023-11-30 22:25:15 [INFO] Pass quantize model elapsed time: 4957.95 ms
    
```

Validation DataLoader 0: 100% 82/82 [02:21<00:00, 1.73s/it]

Validate metric	DataLoader 0
val_loss	0.41693004965782166
val_metric	0.8655667304992676

Benchmarking inference



Raspberry Pi 3B (Cortex A-53) - Hyperfield-1

- Ubuntu 22.04.3 LTS
- PyTorch 2.1.1
- Image size (4, 384, 384)
- Inference latency: **5.43 s / image**

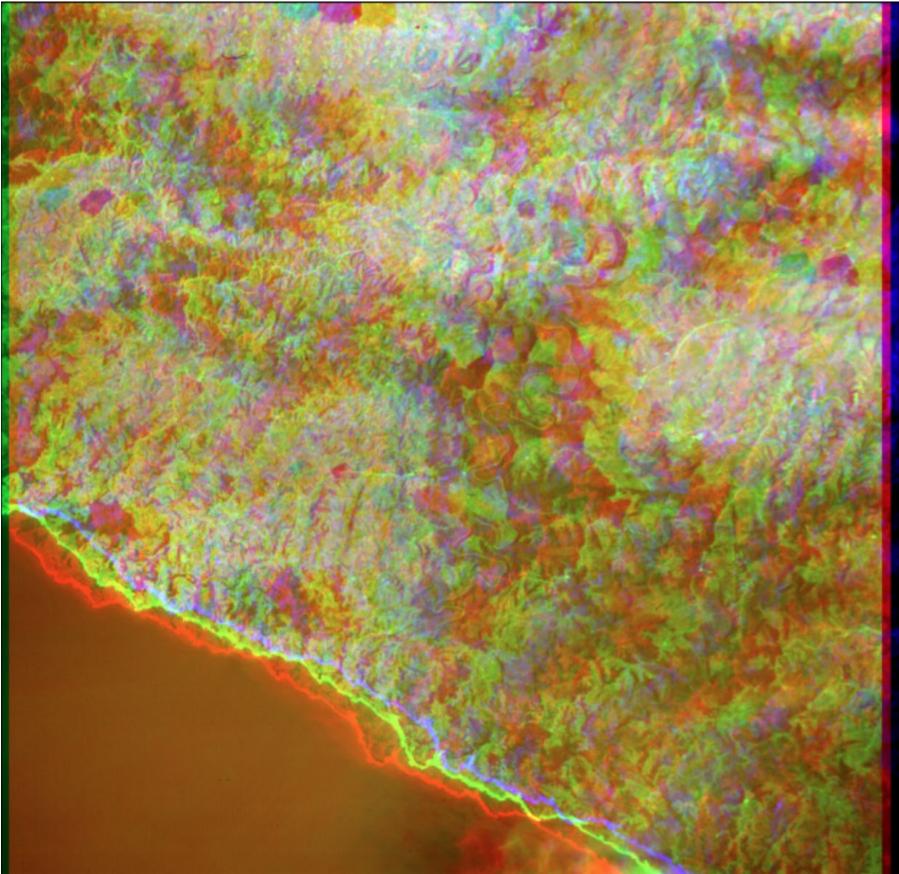


NVIDIA Jetson AGX Orin Dev Kit - Hyperfield-2

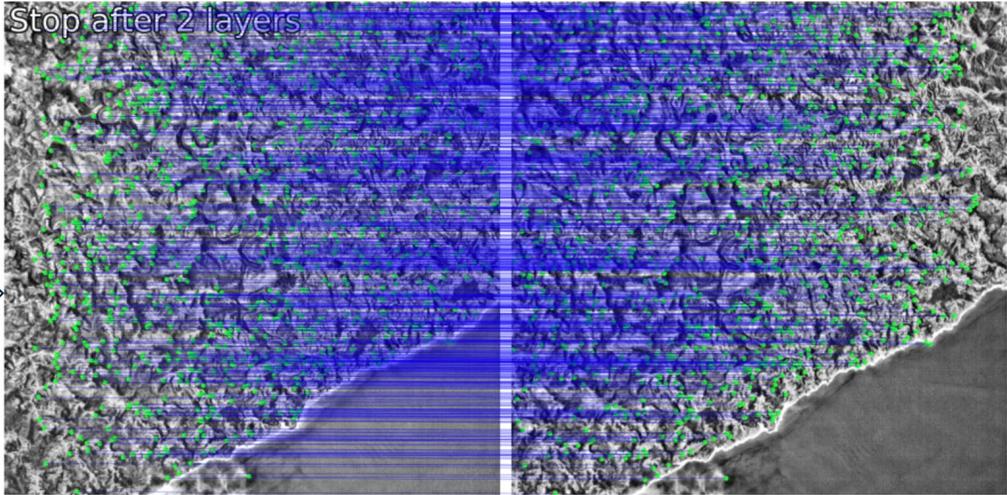
11.1 + 666.4 + 66.6

On-orbit Cloud Detection

Performing Band Alignment



ALIKED
+
LightGl
ue



~150x
faster!

TLE Aligner

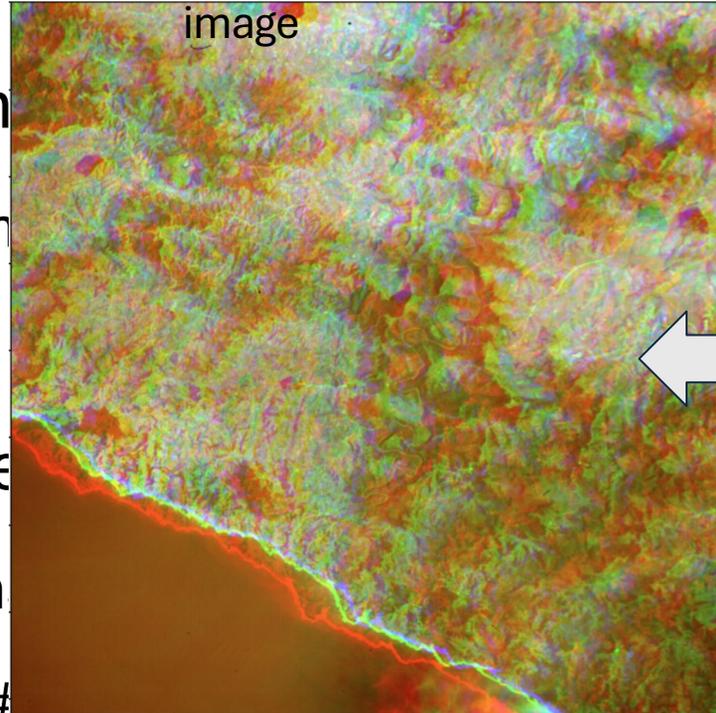


On orbit Cloud Detection

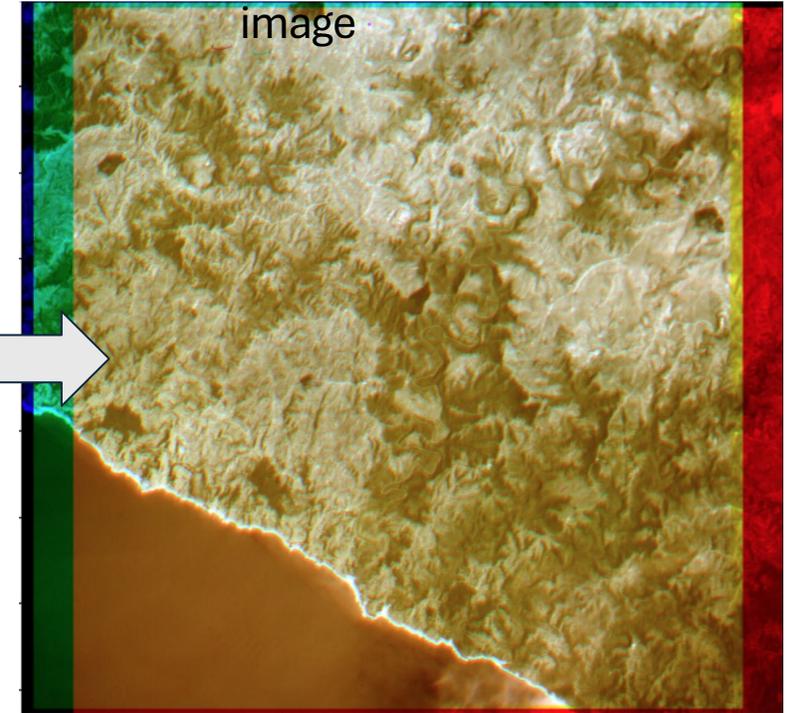
Two-line Element Aligner

- Underlying assumption
 - Satellite trajectory is sn
 - Telemetry is stable
- Velocity-based alignment
 - Δt between acquisition
 - Shift bands by integer # pixels
- Sources of uncertainties
 - Acquisition time stamps

Raw unaligned image



TLE-aligned image

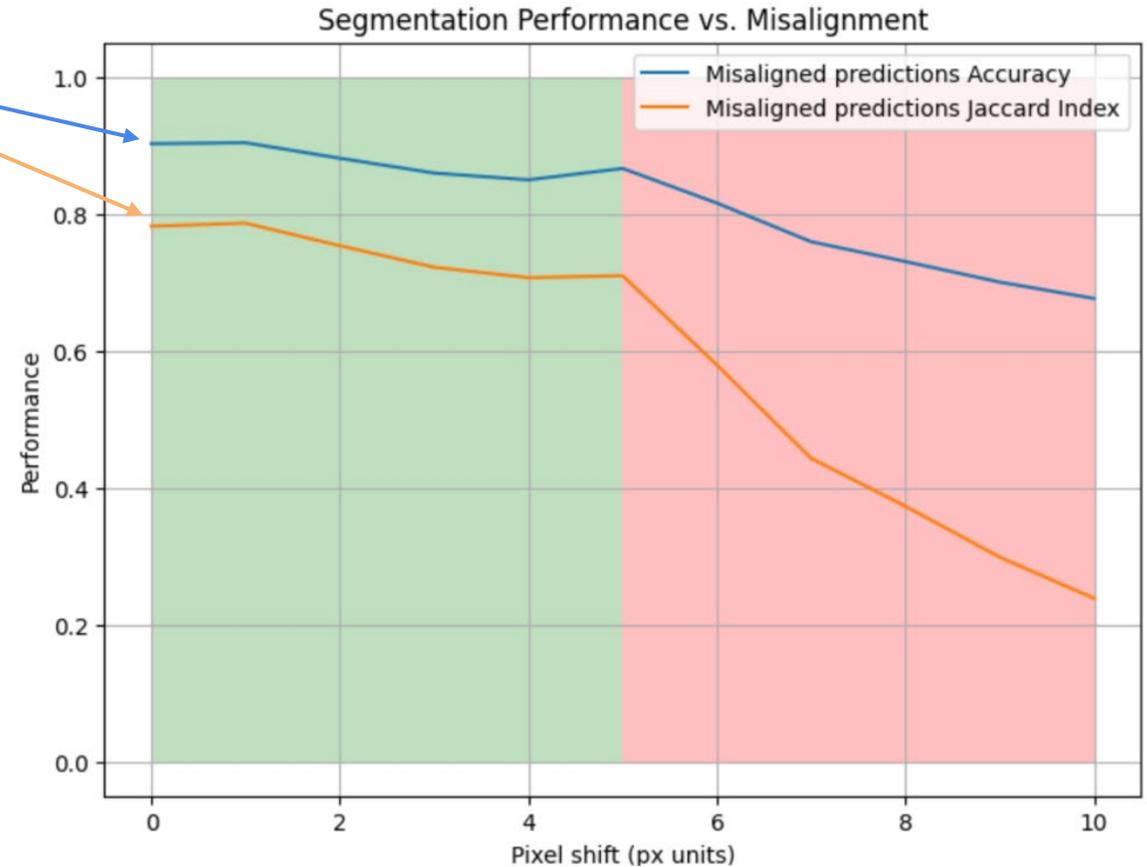
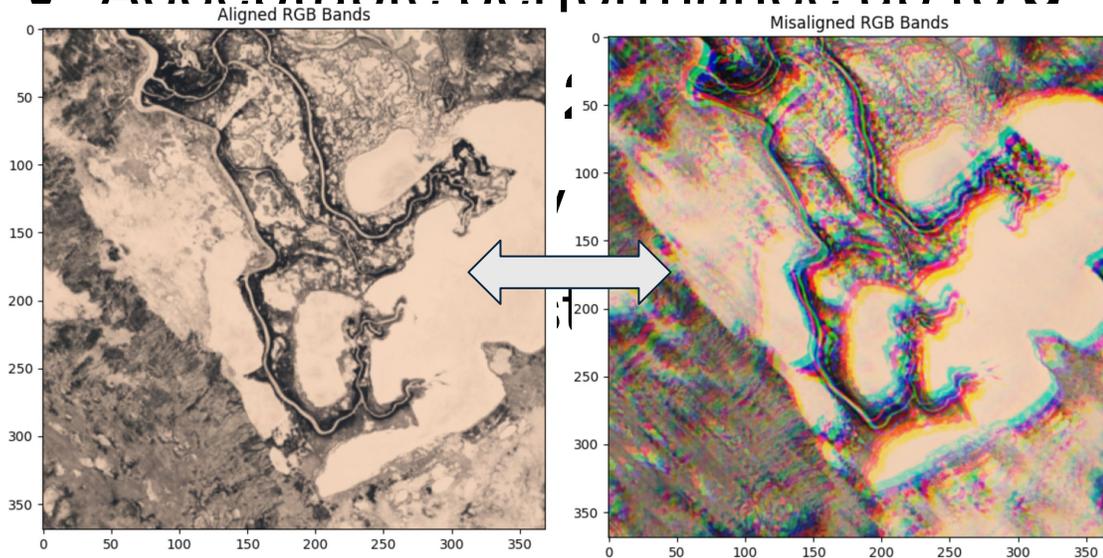


HF1A - Drakensberg Mountains (South Africa) - 10 October 2024
08:23:09 UTC

On orbit Cloud Detection

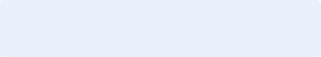
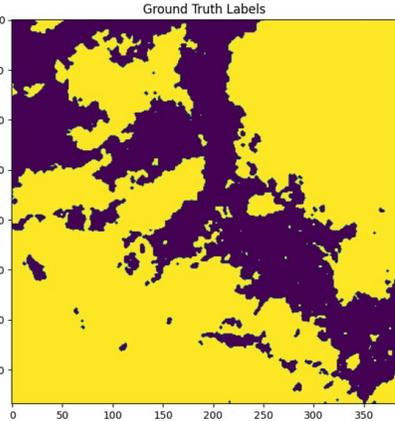
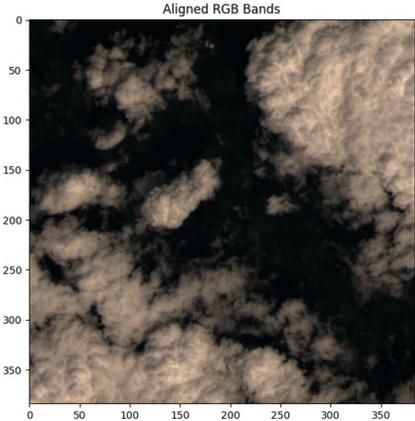
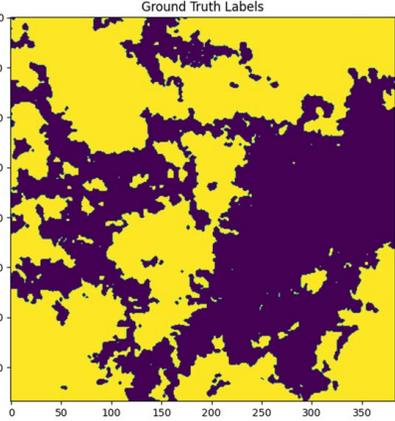
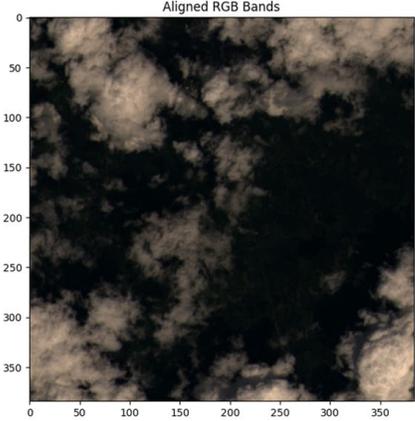
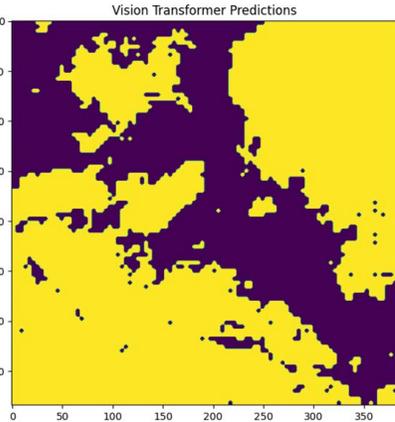
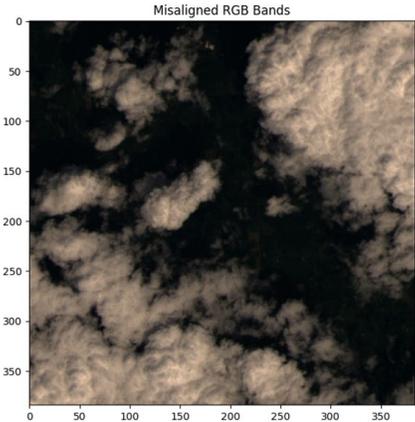
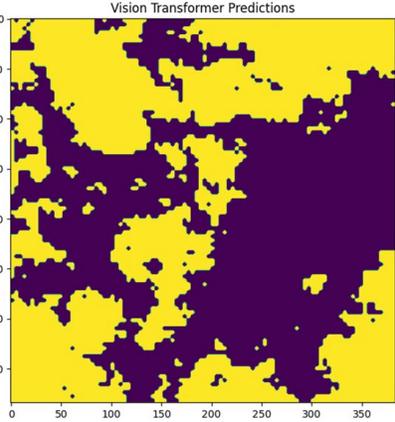
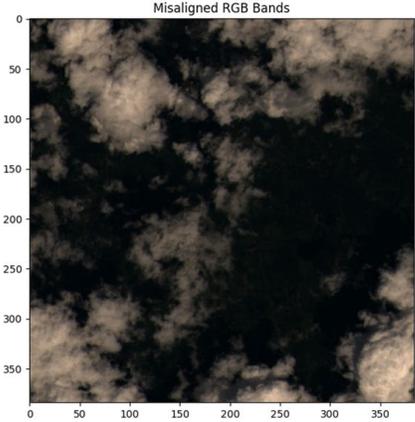
Robustness against misalignment

- Accuracy / JI as function of channel misalignment
- Acceptable performance up to 5



On orbit EO Pipeline

Cloud Detection under channel misregistration (10 pixels)



KUWO SPACE

Hyperfield 2
N°002

Hyperfield 2
N°019

Hyperfield 2
N°003

Hyperfield 2
N°020

Hyperfield 2
N°001

Hyperfield 2
N°0018

