SMARTER DATA FOR A STRONGER PLANET

# Advances in deep learning spectral models for mission-agnostic cloud detection

Arthur Vandenhoeke – 13th November 2024, Workshop on International Cooperation in Spaceborne Imaging Spectroscopy

# Challenges

Recent progress in Earth Observation

- Satellite technology (Sensors, HW, SW)
- Ground network infrastructure

→ Shift towards **small**, **affordable** and **disposable** satellites.

Increasing demands for hyperspectral data providers

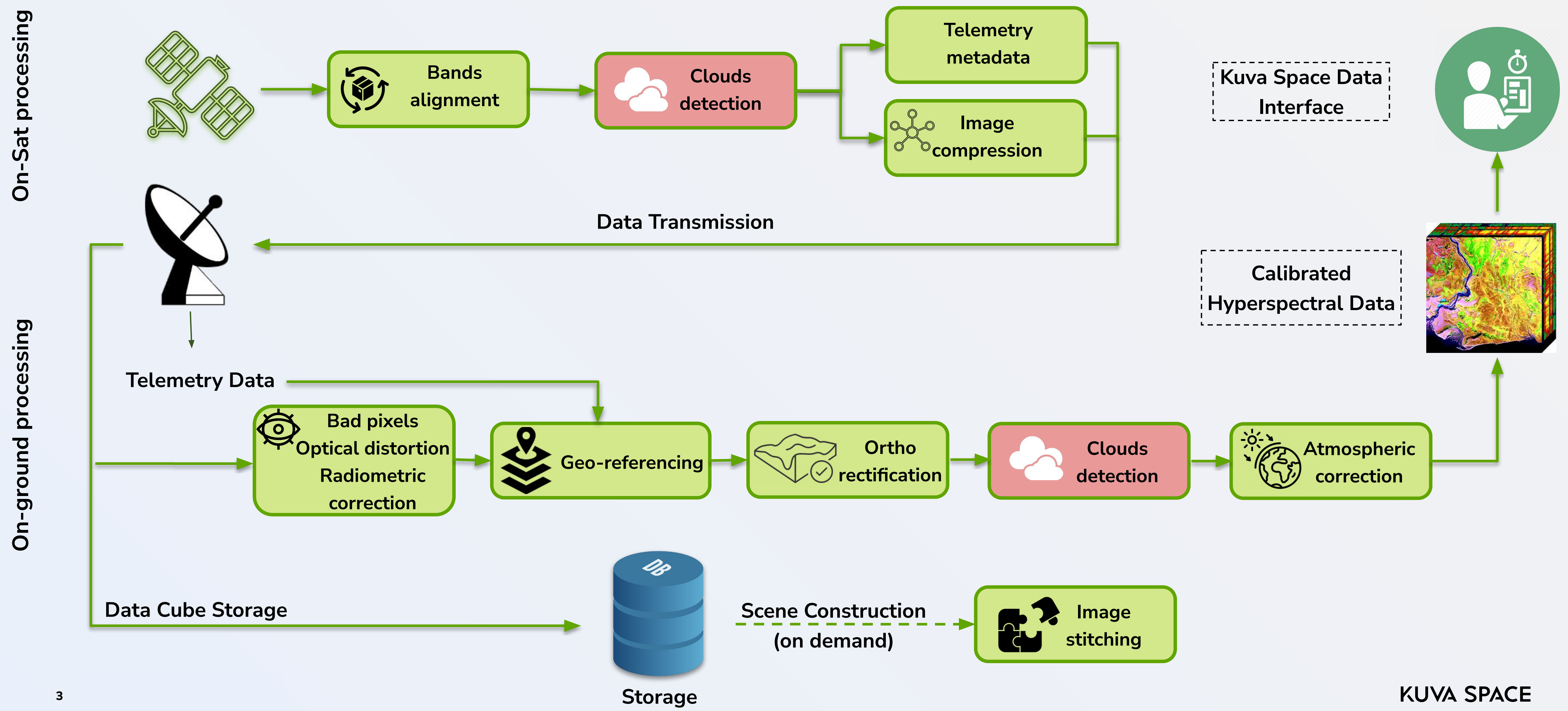
- Disaster response
- Environmental monitoring
- Security & Surveillance

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**.

# Data Processing chain

From satellite acquisition to hyperspectral product






KUVA SPACE

KUVA SPACE

SMARTER DATA FOR A STRONGER PLANET

# On-orbit cloud detection

# Approach

Limitations of Vision Transformers for Hyperspectral Imaging:

- Data Requirements: Extensive, high-quality labeled datasets
- High Computational Load: Large number of parameters
- Large Model Size: Tens-Hundreds of MBs >> uplink capacity
- Latency Constraints: Hinder near-real-time operations

Lite Vision Transformer with **Enhanced Self-Attention** †

- Reduced wavelengths (RGB-NIR only);
- Small inference latency (s to ms);
- Small number of parameters;
- Periodic fine-tuning & uplink;
- Maximum 2MB (= Max 1M parameters using float16);
- > 90% F1-Score on binary class.

† Yang, Chenglin et al. “Lite Vision Transformer with Enhanced Self-Attention.” 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2021): 11988-11998.



# On-orbit EO pipeline

## Model Training

- Datasets
    - L8-Biome (clear / thin / thick / shadow - VIS-NIR-SWIR)
    - 95-Cloud (cloud / no cloud - VNIR)
- Aggressive augmentation are key to generalization
    - Geometric augmentations (flipping, affine transformations)
    - Radiometric augmentations (brightness, gamma, blurring)

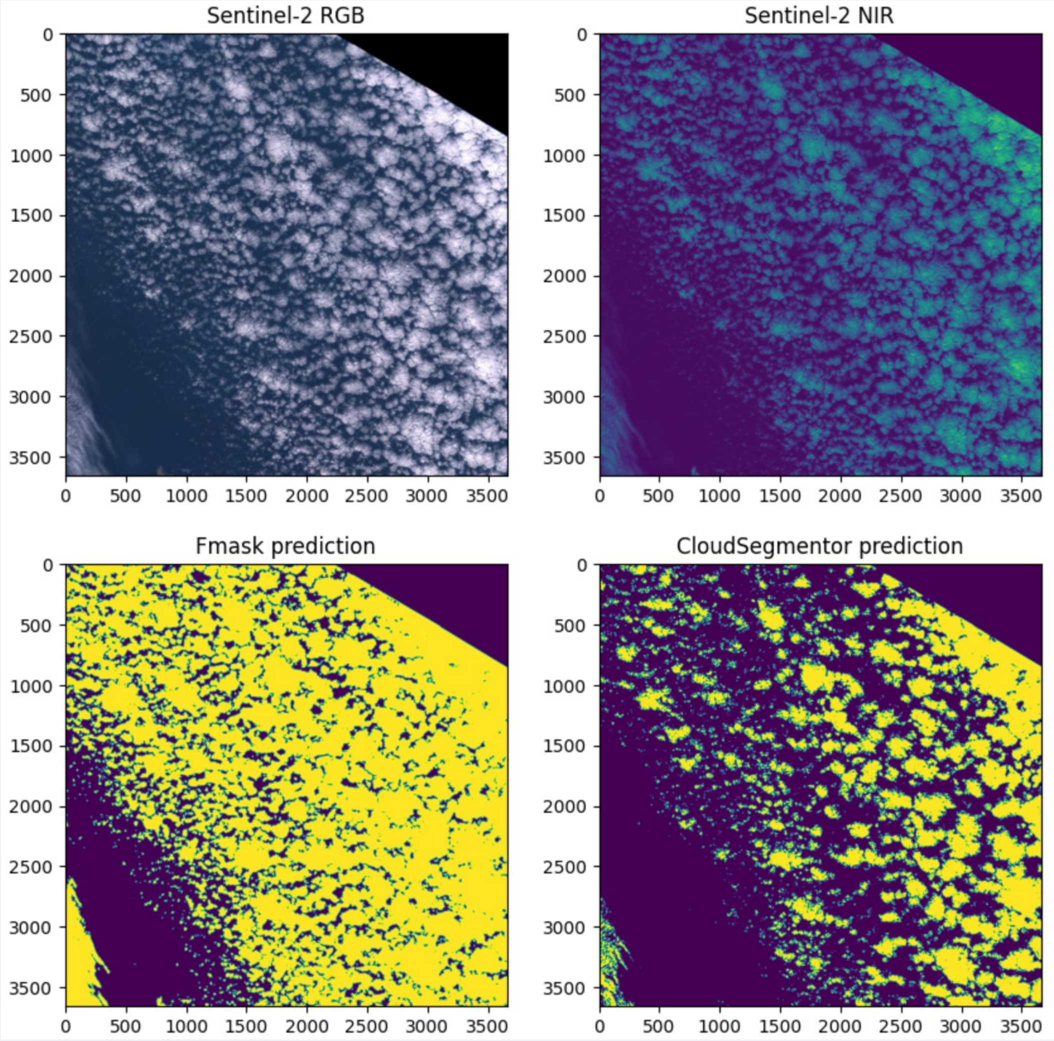
## Model Evaluation

- Benchmark LVT across various sensors:
    - PRISMA [HSI]     –> Qualitative Inspection (WIP)
    - Sentinel-2 [MSI] –> Comparison against FMASK
- LVT reaches **90.28% F1-Score** on binary classification (test set)
- Inference speed on 384 x 384 x 4 image = 2ms on NVIDIA 3090

Early results suggest our model have the capacity to **work even across different sensor technologies and satellites** despite have been developed just with Landsat imagery.

## On-board deployment

The incorporation of GPU will be investigated in a recently secured project with Copernicus Security Services.



# 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

- Ubuntu 22.04 at 30W power mode
- Jetpack 6.1 SDK with L4T PyTorch Image
- Inference latency: **32 ms / image** (4 x 384 x 384)
- = **431ms / HF-1 image** (64 x 4 x 256 x 256)

KUVA SPACE

SMARTER DATA FOR A STRONGER PLANET

# On-ground cloud detection

# On-ground Cloud Detection

## Training Workflow

**Model Selection**

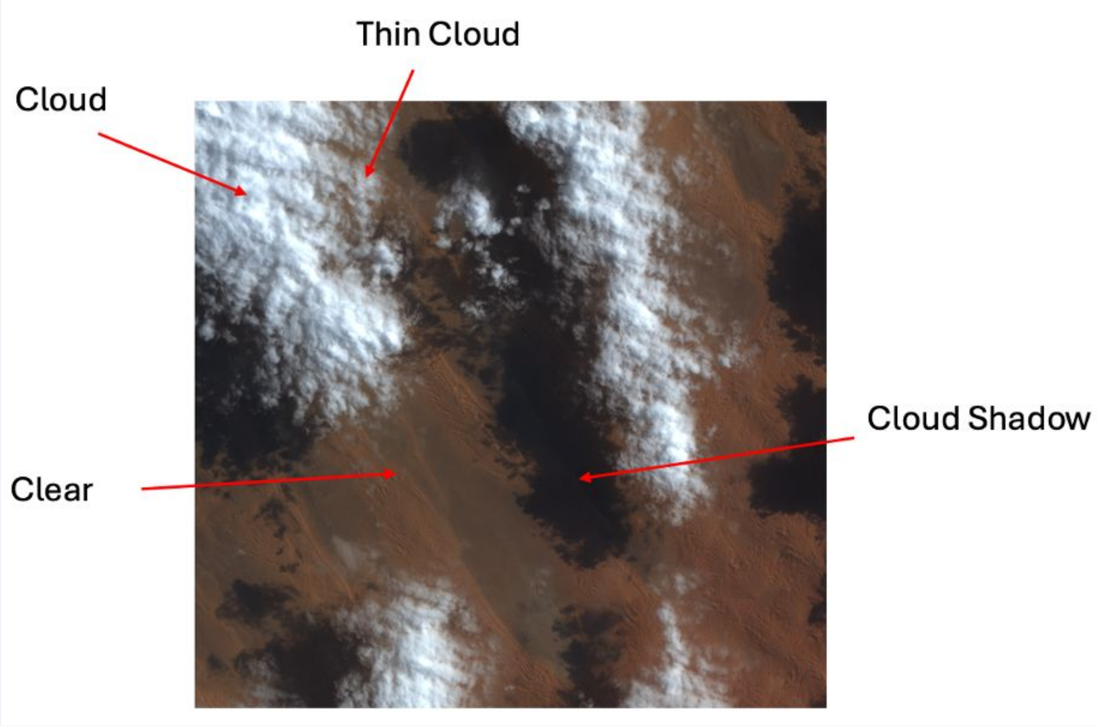
- Vision Transformer with patch_size = 4 and emb_dim = 192
- Keep model size down to 5.5M parameters (tiny)

**Model Training**

- Datasets: L8-Biome (clear / thin / thick / shadow - VIS-NIR-SWIR)
- Cosine learning rate scheduler with linear warmup
- Same augmentations as on-orbit cloud detection
  - Geometric augmentations (flipping, affine transformations)
  - Radiometric augmentations (brightness, gamma, blurring)

**Iterative Fine-Tuning**

- L8-Biome contains inconsistencies (especially cloud shadows)
- Vision Transformers are data-hungry (need >> 96 L8 images)
- Use L8-Biome to train v1 then use v1 as pre-annotator
- Encord annotation platform to manually fin-tune annotations
- Gradual increase of data set to **600 annotated PRISMA images**




Ontology definition in L8-Biome and Encord annotation platform

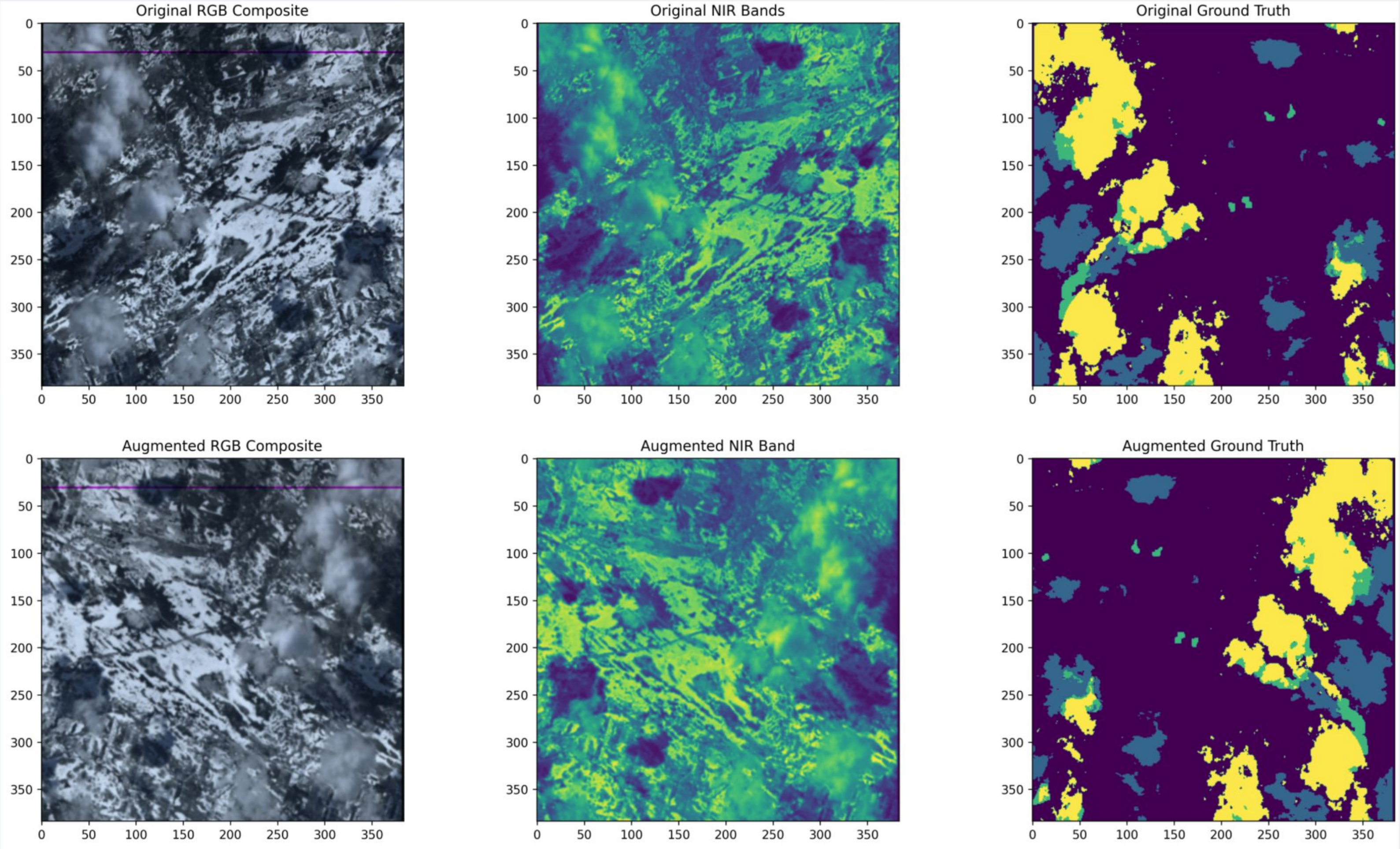


KUVA SPACE

# **On-ground** **Cloud Detection**

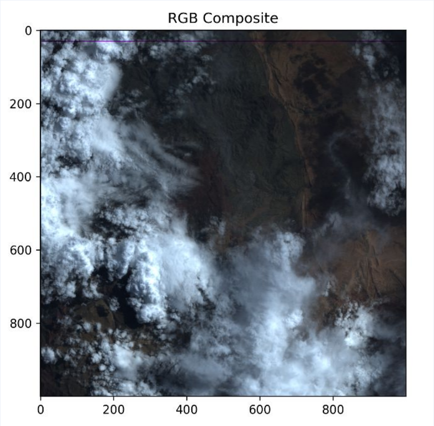
Manual PRISMA image annotations

- Sample of **PRISMA image** and its manual expert annotation (top).
  - Yellow = Thick Cloud
  - Green = Thin Cloud
  - Light Blue = Cloud Shadow
  - Dark Blue = Clear
- Bottom row = the cropped and blurred augmentation of the selected sample.
- Carefully chosen set of PRISMA images to overcome the known limitations of the L8-Biome data set.

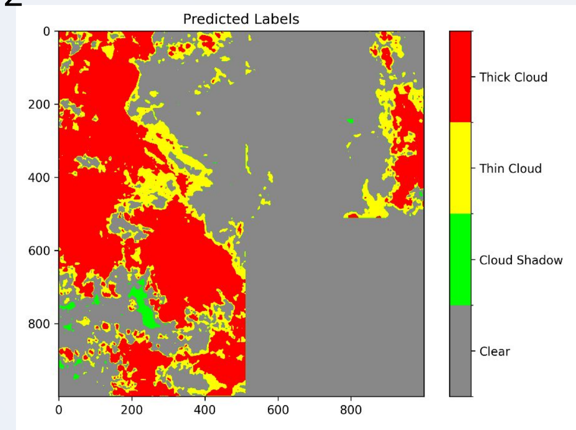


# On-ground Cloud Detection

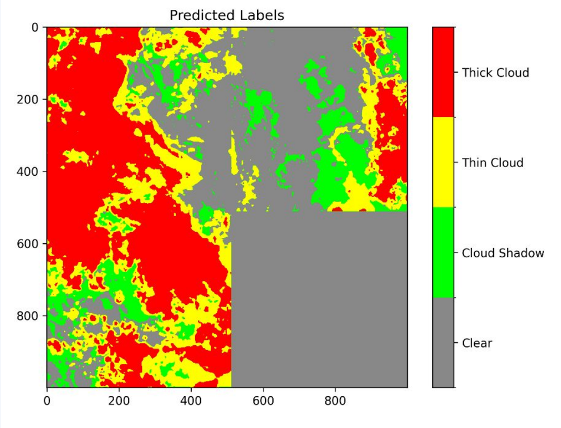
Iterative fine-tuning



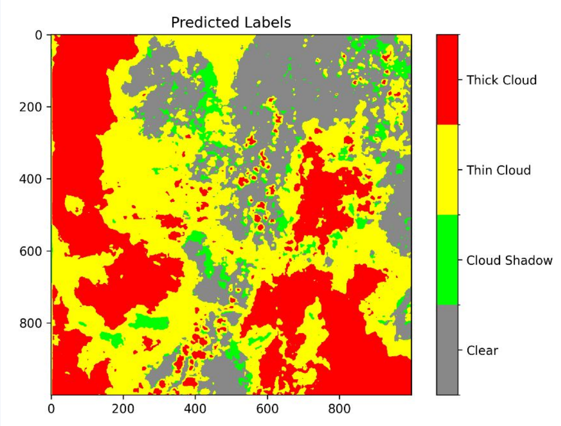

v2




v3




v4

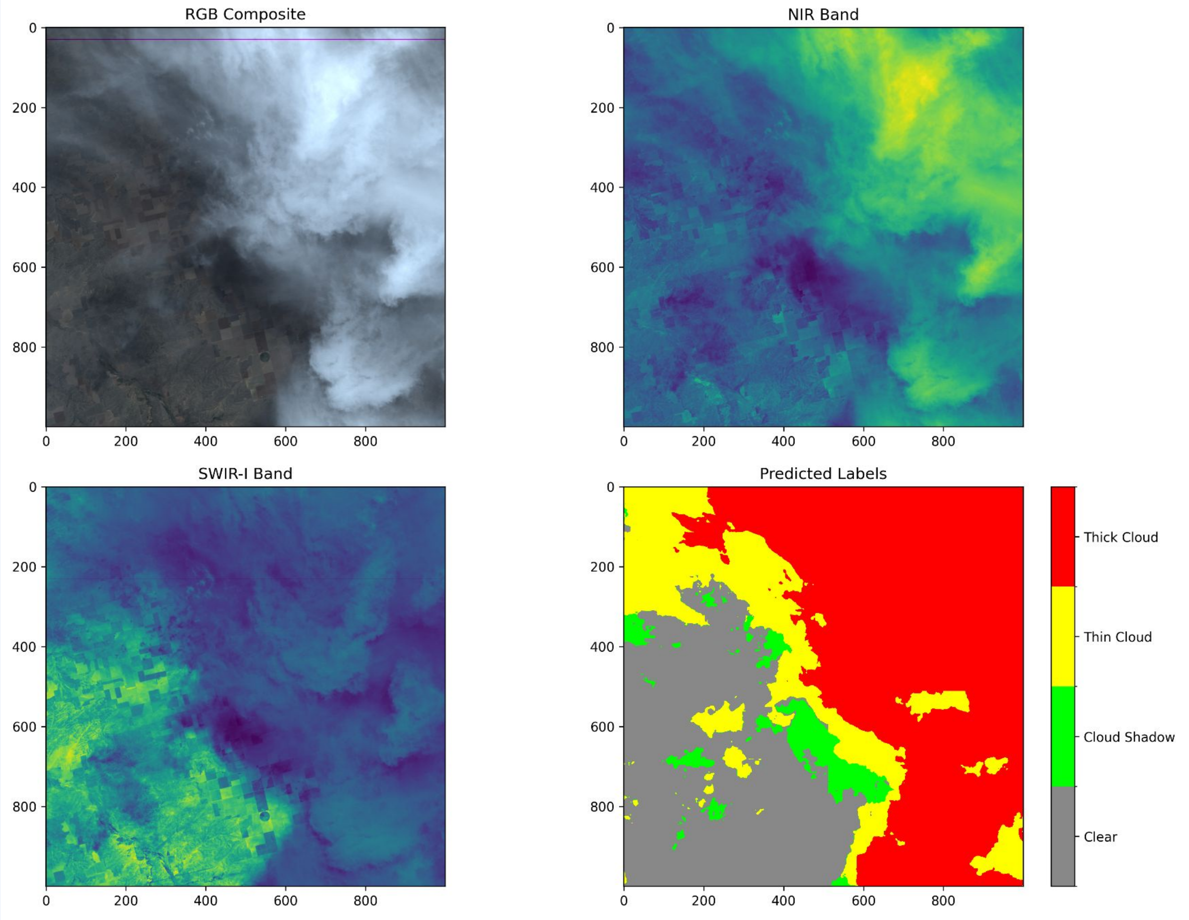


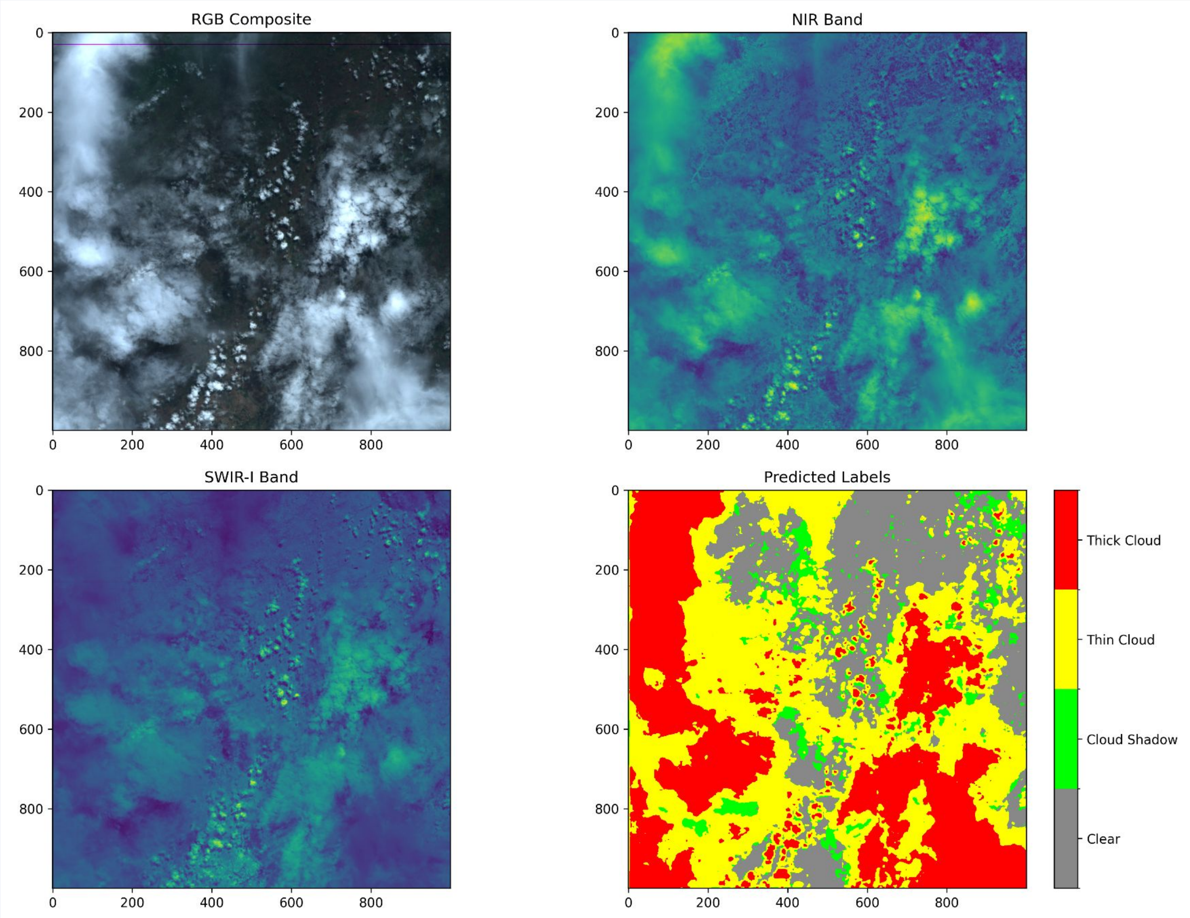

| Epoch Number | ViT version | # PRISMA images | F1-Score (%) |
|---|---|---|---|
| 271 | v1 | N/A | 92.78 |
| 374 | v2 | 250 | 94.02 |
| 593 | v3 | 500 | 94.22 |
| 600 | v4 | 600 | 94.45 |

Classification metrics across fine-tuning runs

# On-ground Cloud Detection

Inference on PRISMA imagery



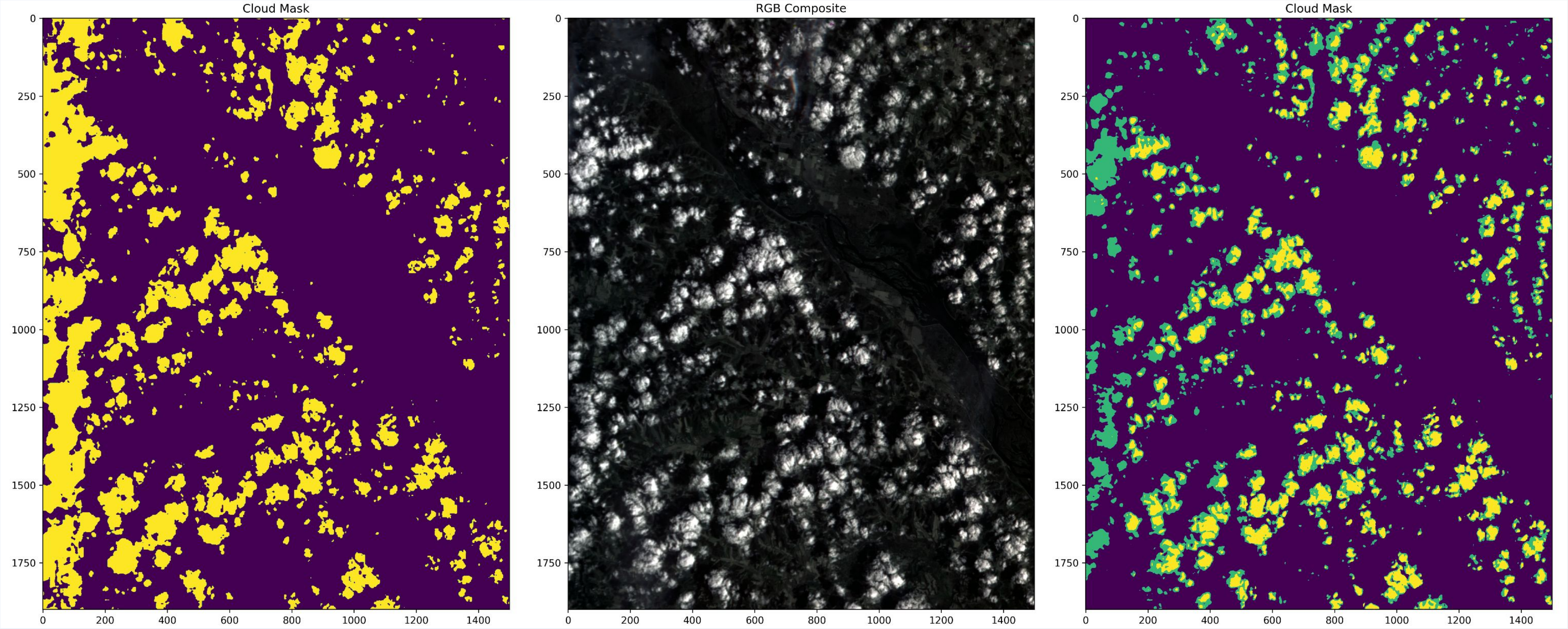






KUVA SPACE

# On-ground Cloud Detection

Application to raw HF-1A imagery (uncalibrated DN)




LVT

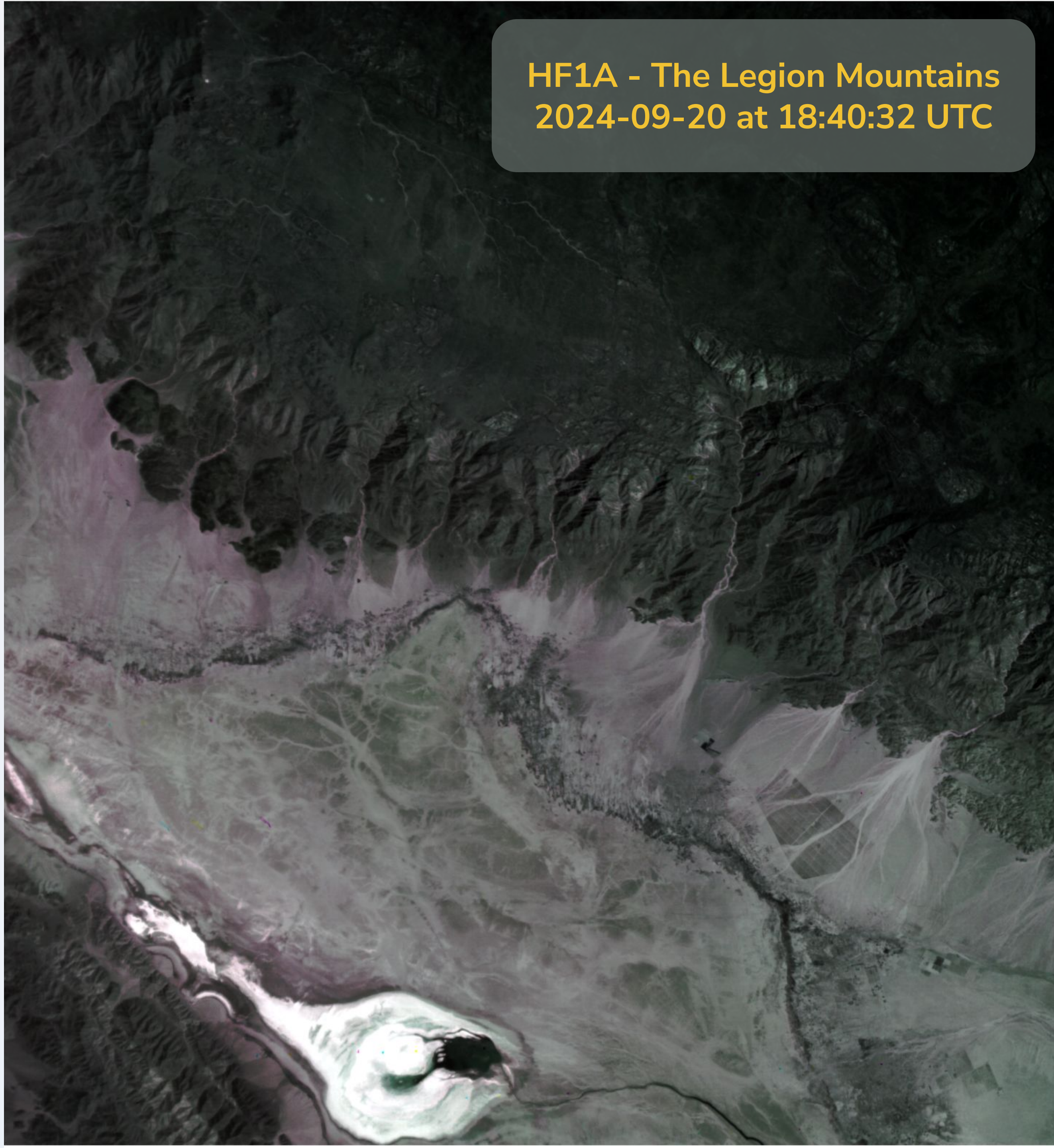
ViT

Colfax, USA - 1 November 2024 17:14:53 UTC

# Conclusion

- How much "instrument agnostic" are the current state of the art algorithms?
  - Tailored for specific EO missions
    - Fmask → Landsat + Sentinel
    - Handcrafted thresholds and band ratios
  - Requires mission-specific inputs
    - S2cloudless requires 10 S2 bands
    - CloudSEN12 adds S1 (SAR), DEM, surface water occurrence and land cover masks

- Can the AI based approaches really become instrument agnostic?
  - Yes, but data and algorithms go hand in hand
  - Key =
    - Common & minimalistic data set, with a ...
    - sufficiently complex model architecture that ...
    - lead to rich image representations (training)
  - "An uncalibrated cloud remains a cloud"
  - SSL offer promising avenues for better generalization



# KUVA SPACE

## TOGETHER TOWARDS A SUSTAINABLE PLANET AND PROSPEROUS HUMANKIND

Kuva Space Oy Otakaari 5, 02150 | Espoo, Finland

arthur.vandenhoeke@kuvaspace.com

kuvaspace.com

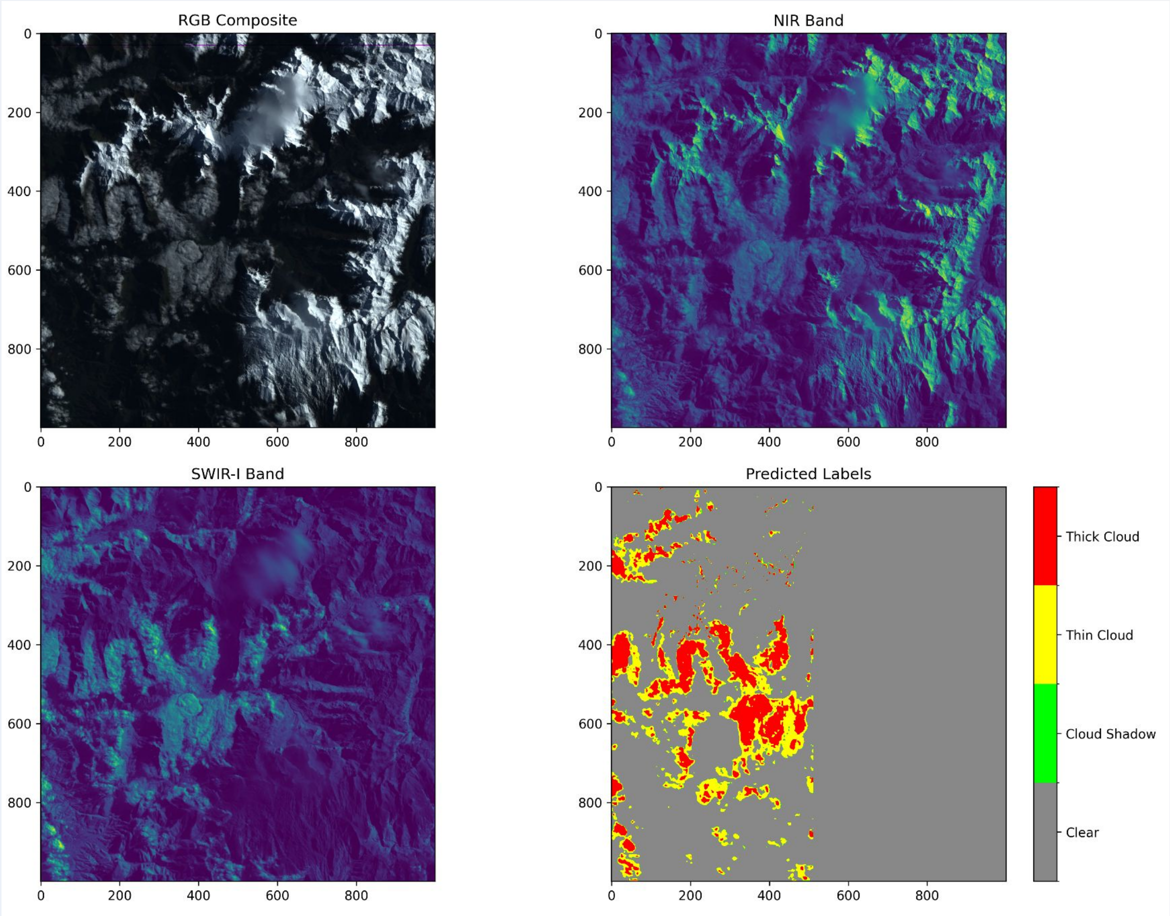
# On-ground Cloud Detection

## Known Limitations

- Missing Cirrus Band @1.37um
  - Bad distinction between ice / snow / high altitude clouds.
  - HF-2 will integrate SWIR-I

- Spectral Sensitivity Analysis
  - What if we add more bands?
  - Explainable AI to help identify the importance spectral channels

- Tradeoff size vs. Performance [WIP]
  - How does ViT/S-4 perform?
  - How about ViT/S-6?



# On-Orbit EO Pipeline

Deploying cloud detection on-orbit (CPU)

## Model Preparation

- Post-Training **Quantization** using Intel's Neural Compressor (CPU);
- Trade-off performance loss vs. memory footprint
- Sweet spot found around
  - 86.55% accuracy
  - 1.4MB memory footprint

on-orbit upload time of ~2 days

## Model Deployment

- TorchScript's Just-In-Time compiler
- Platform-agnostic TorchScript file (4.3MB)

```
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/reps/kuva-projects/topics/cloud_detection/nc_workspace/2023-11-30_12-28-39/./history.s
napshot.
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 |



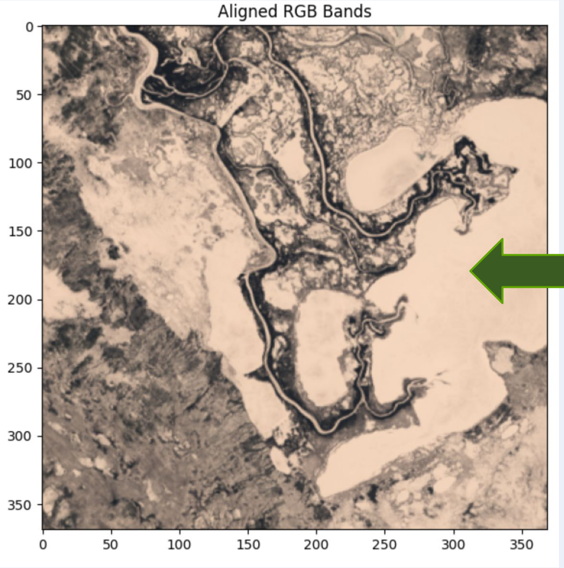
KUVA SPACE

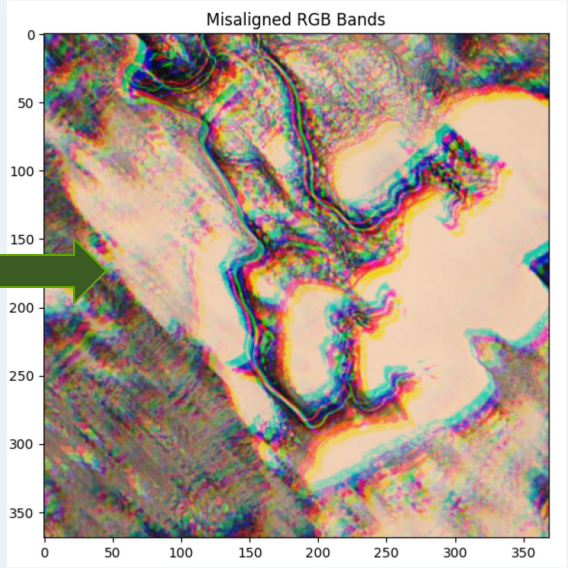
# On orbit EO pipeline

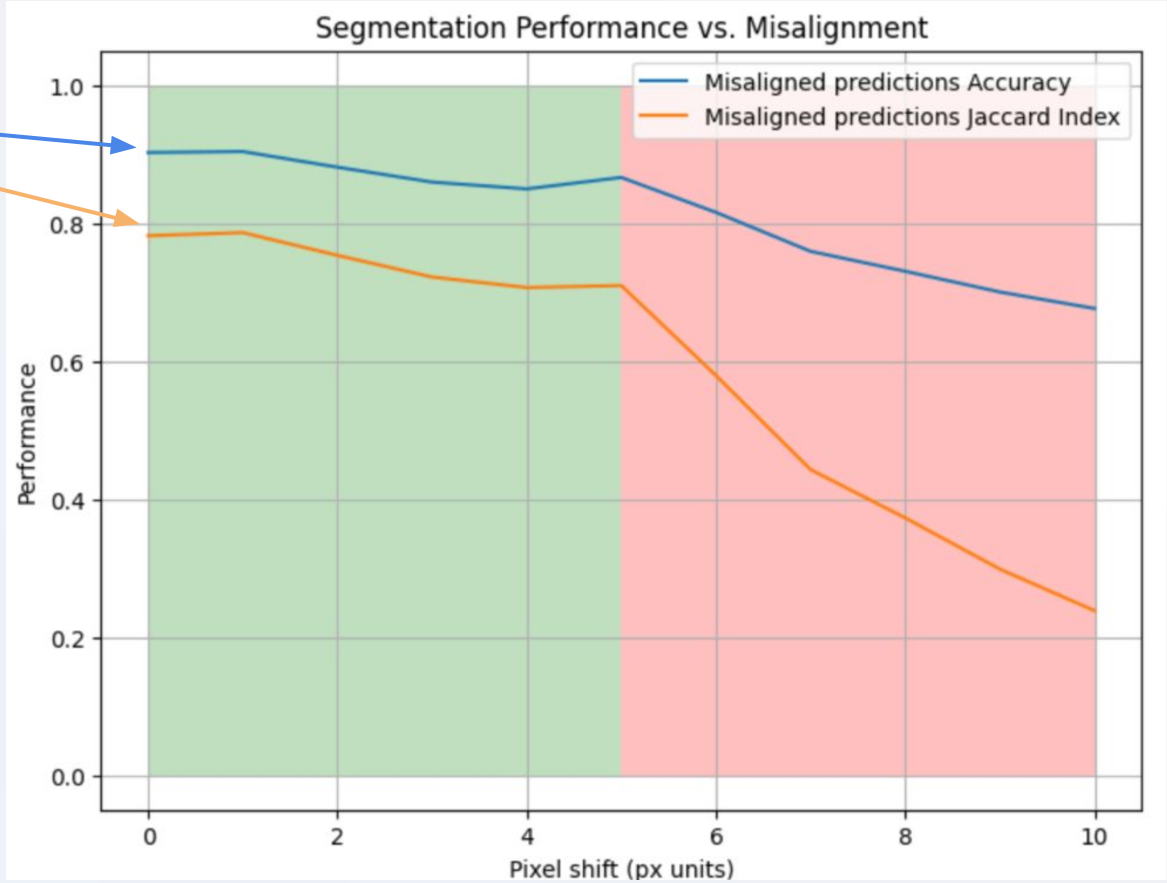
## Cloud Detection

Robustness against misalignment

- Accuracy / JI as function of channel misalignment
- Acceptable performance up to **5 pixels shift (~125m)**
- With constant v = 7.68 m/ms —> 16.28 ms timestamp error

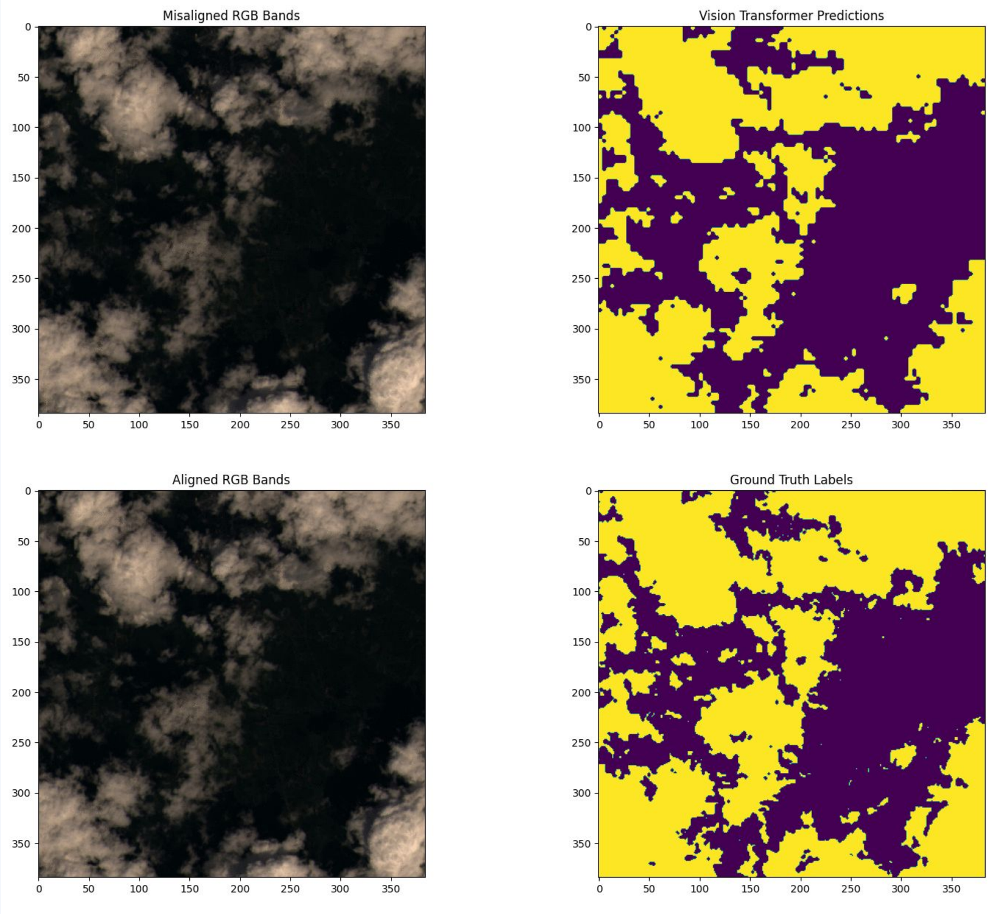


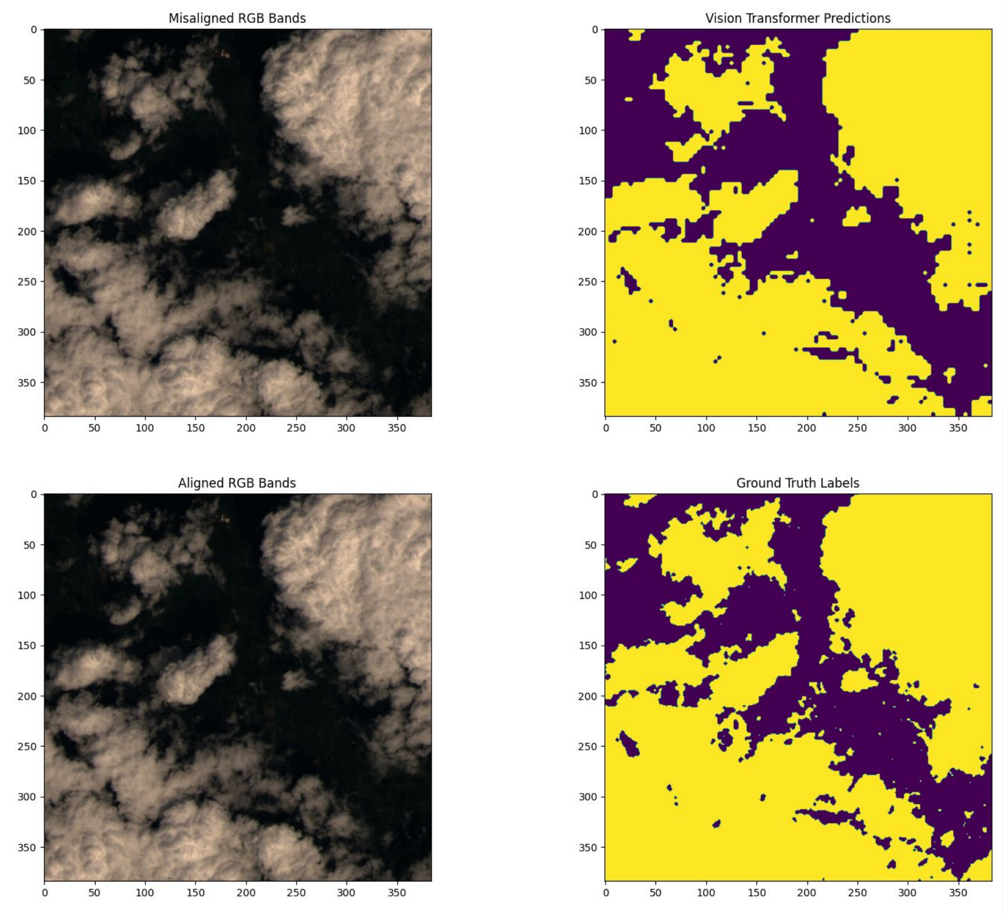



# On orbit EO Pipeline

Cloud Detection under channel misregistration (10 pixels)



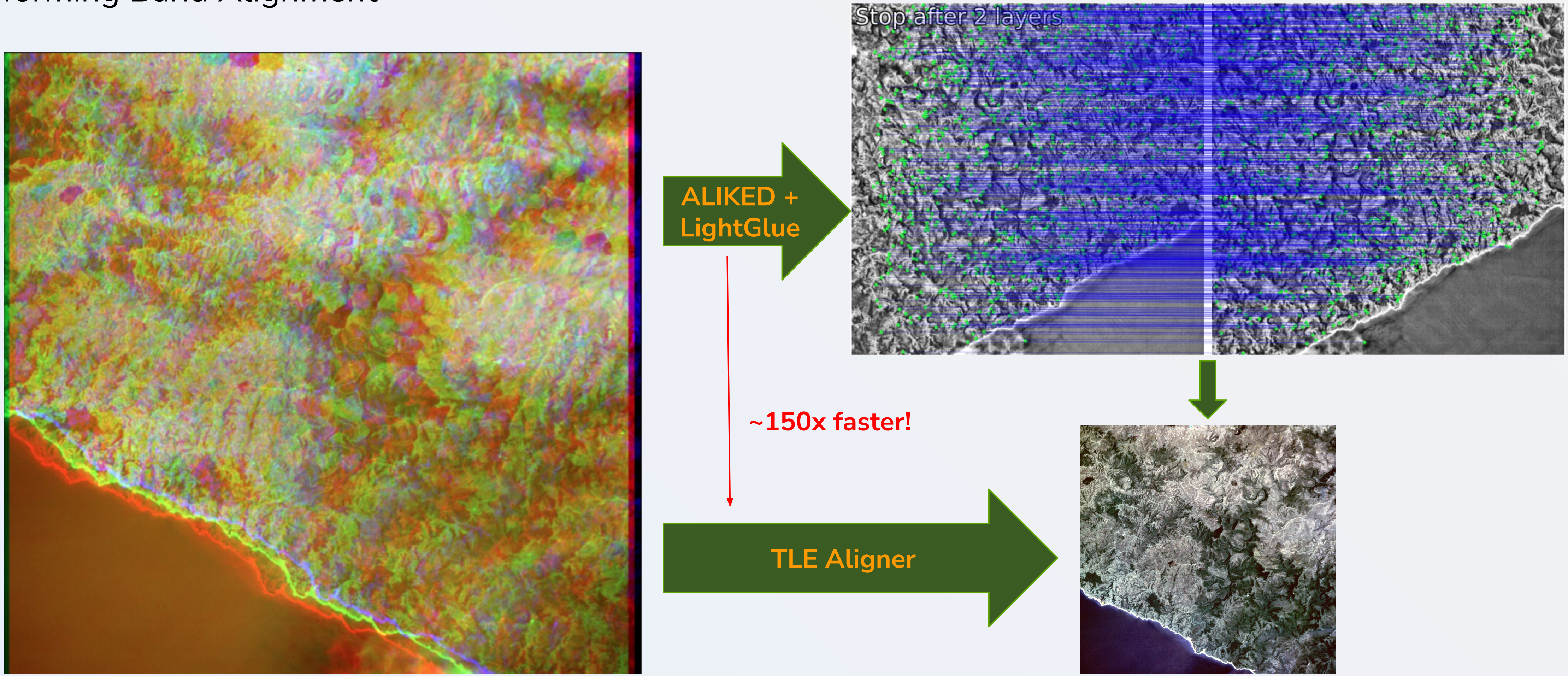






KUVA SPACE

# **On-orbit Cloud Detection**

Performing Band Alignment



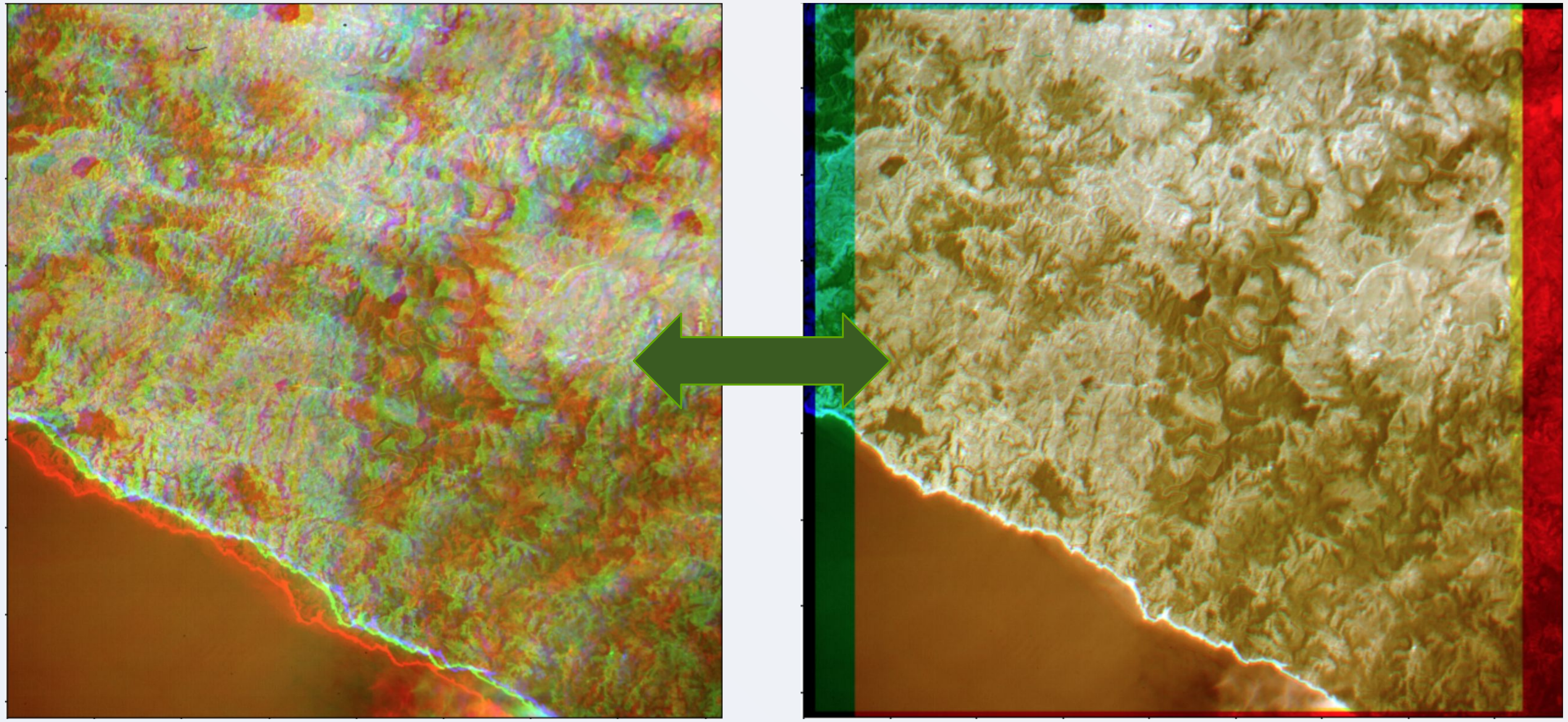
# On orbit EO pipeline

## Two-line Element Aligner

- Underlying assumptions
    - Satellite trajectory is smooth
    - Telemetry is stable
- Velocity-based alignment
    - Δt between acquisitions
    - Shift bands by integer #pixels
- Sources of uncertainties
    - Acquisition time stamps
    - Velocity of satellite

Raw unaligned image

TLE-aligned image



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

**Q**: What is the effect on downstream cloud detection?

# On-orbit EO pipeline

## Copernicus Security Services

Hyperfield for rapid response

- On-orbit processing of HF data
- On-orbit detection and monitoring
- Leverage sat-to-sat and sat-to-IOT

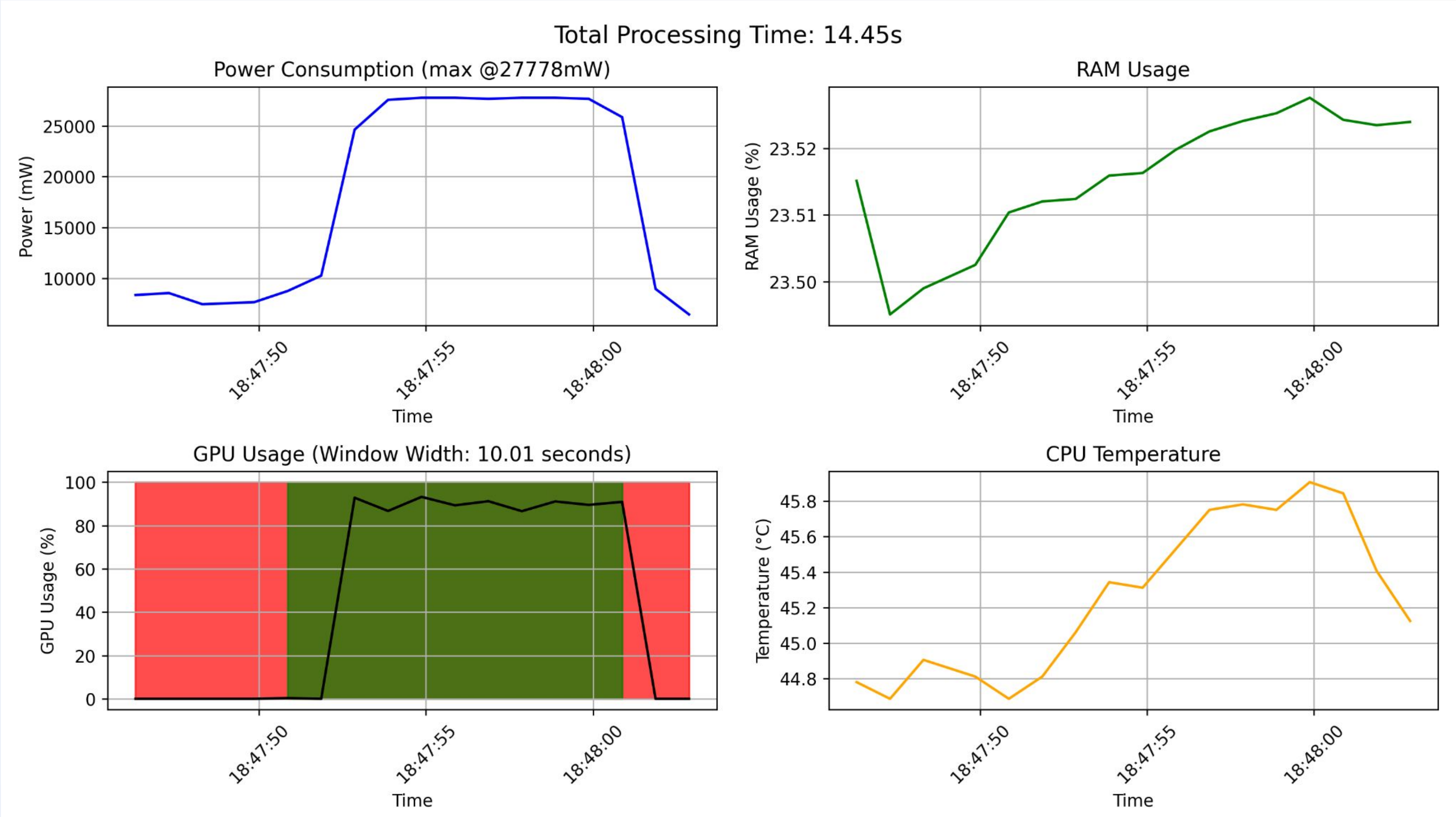
Benchmarking on-orbit GPUs

- NVIDIA AGX Orin 64GB
- NVIDIA Orin NX 16GB
- NVIDIA Jetson Nano 8GB

Various on-orbit processing scenarios

- Band alignment
- Cloud detection
- Georeferencing

Example of tiny ViT running onboard an NVIDIA AGX Orin @50W power mode

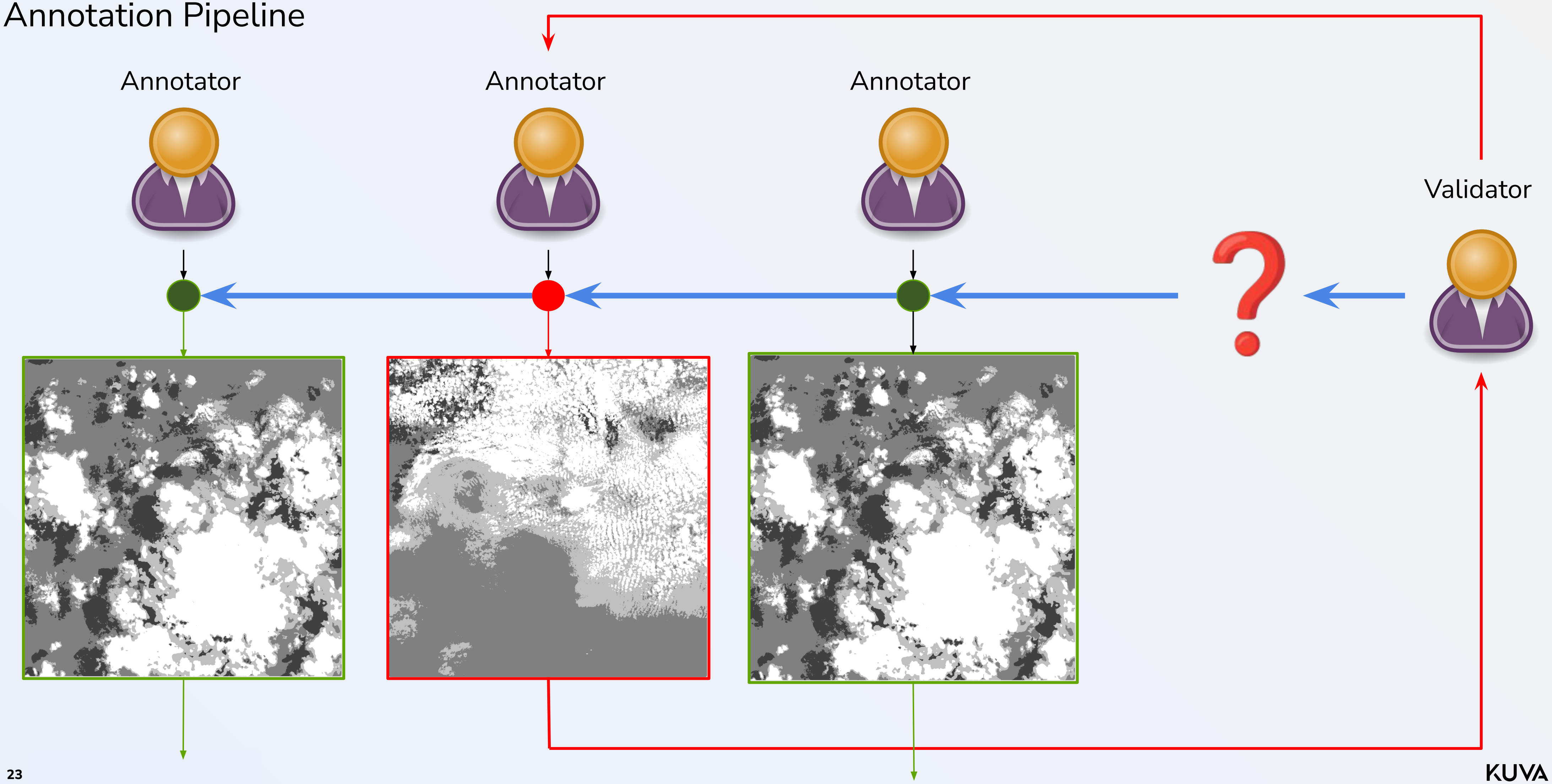





KUVA SPACE

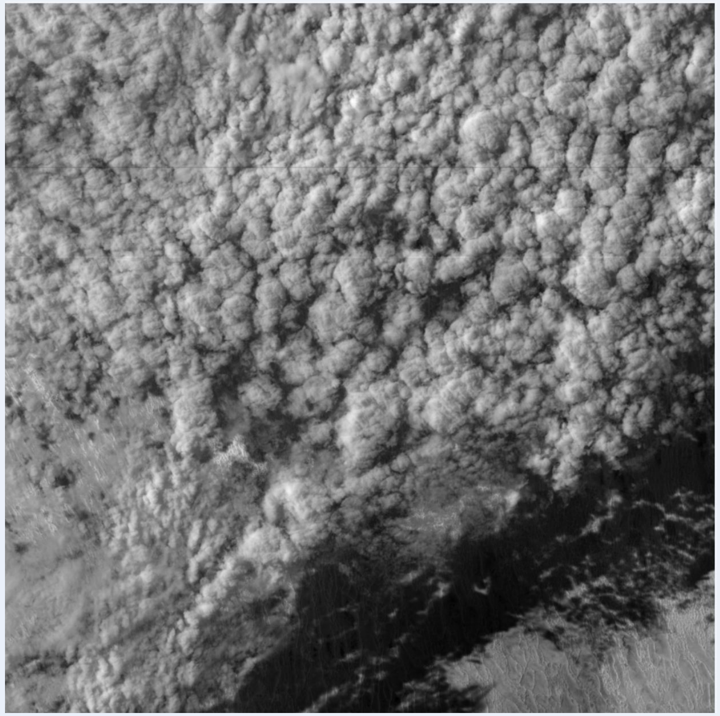
# Encord Annotations

## Annotation Pipeline

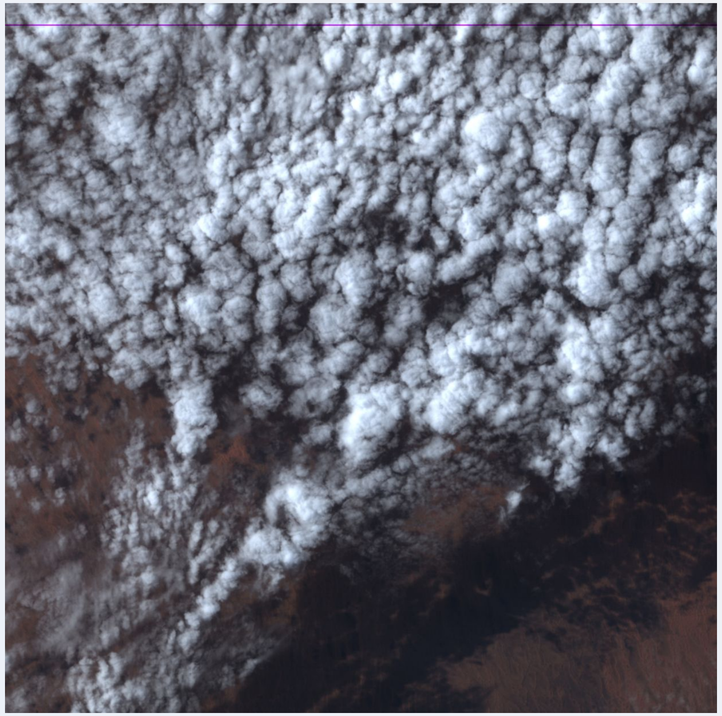


# On-ground Cloud Detection

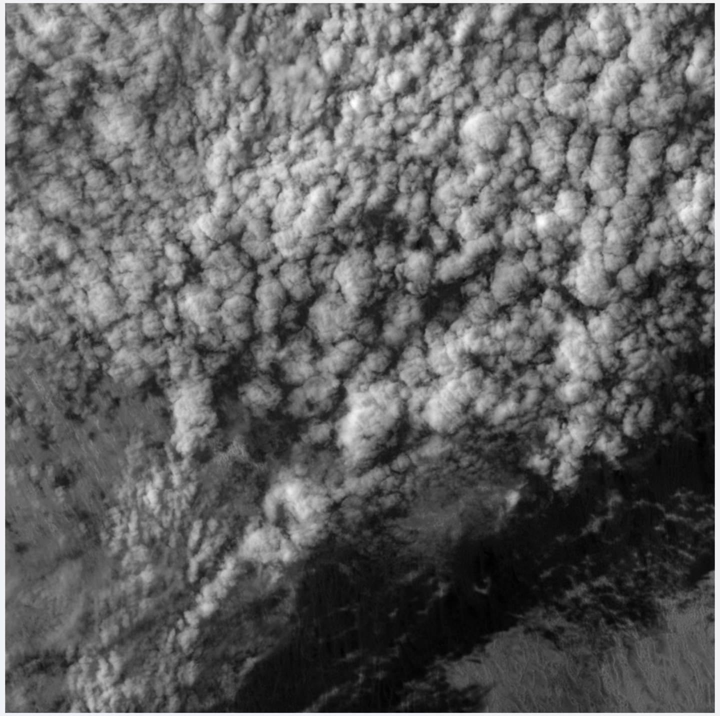
Encord Input Images



NIR



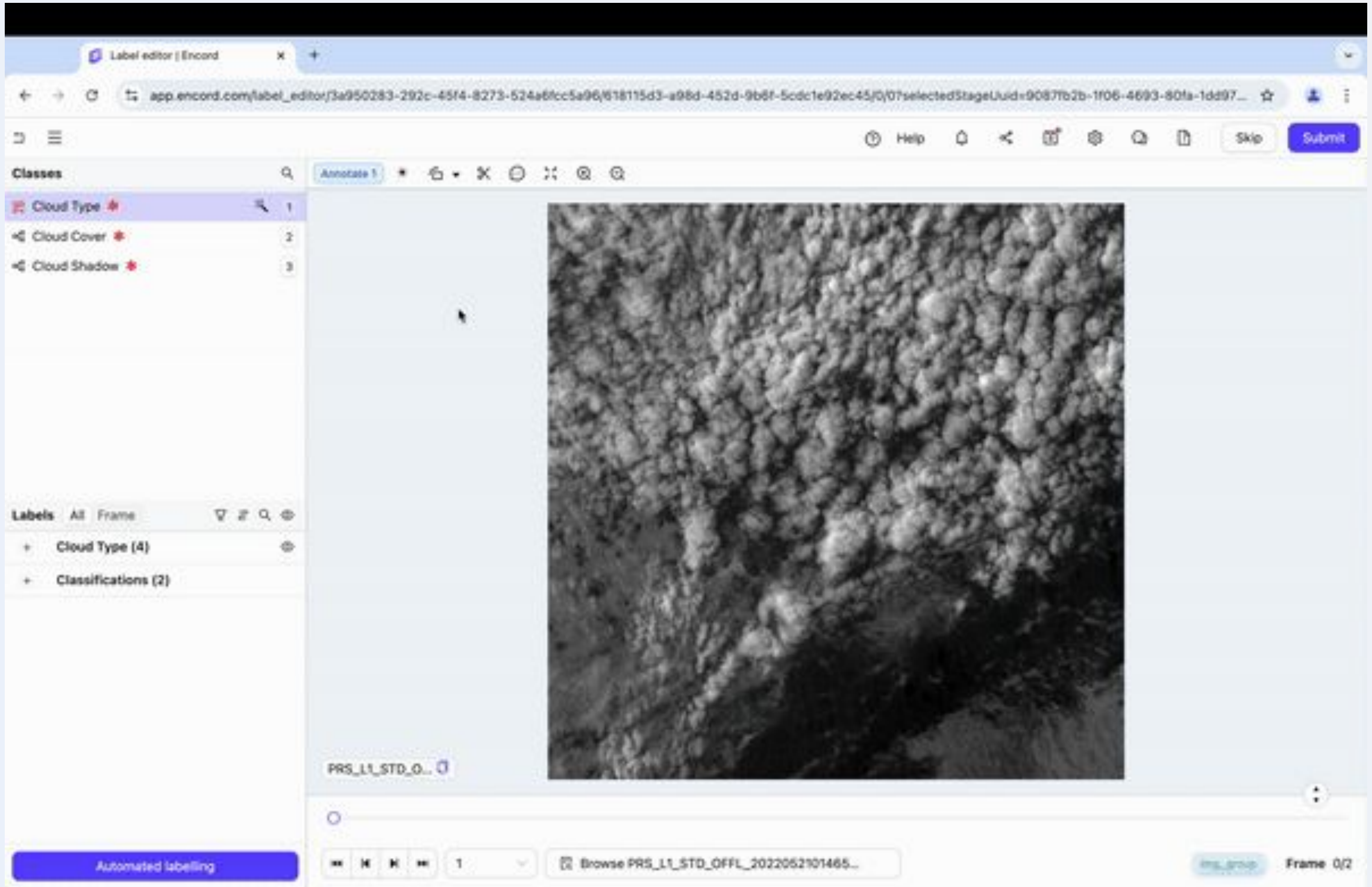
VIS (RGB)

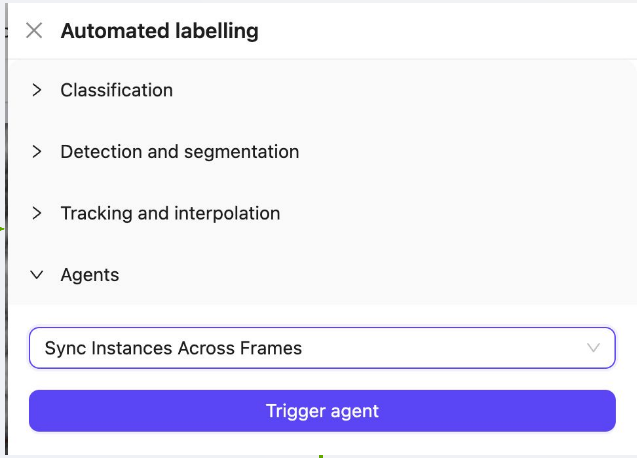


SWIR

# Encord Annotations

Thresholding the NIR bands leads to accurate cloud shadow masks



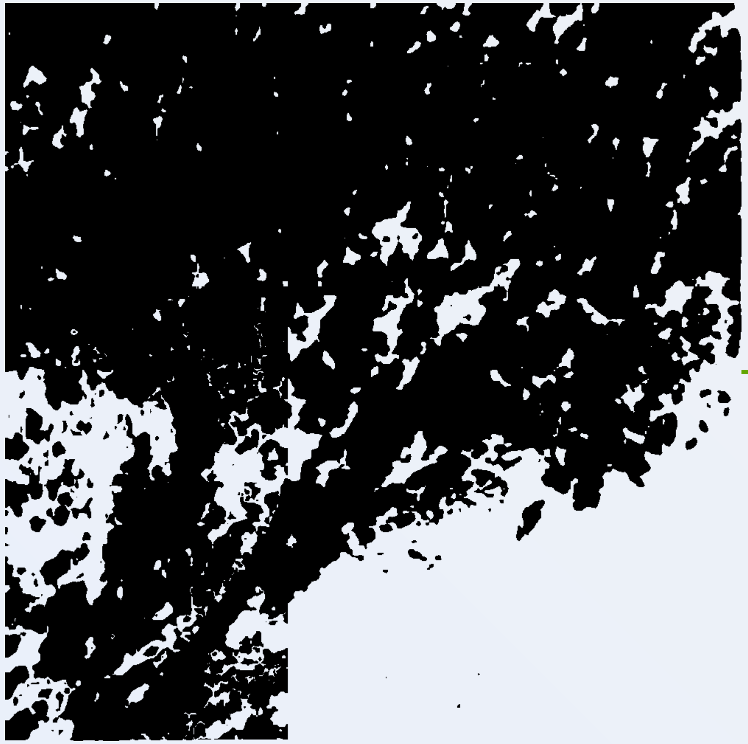

Agent tool allows to copy annotations across frames

# Encord **Annotations**

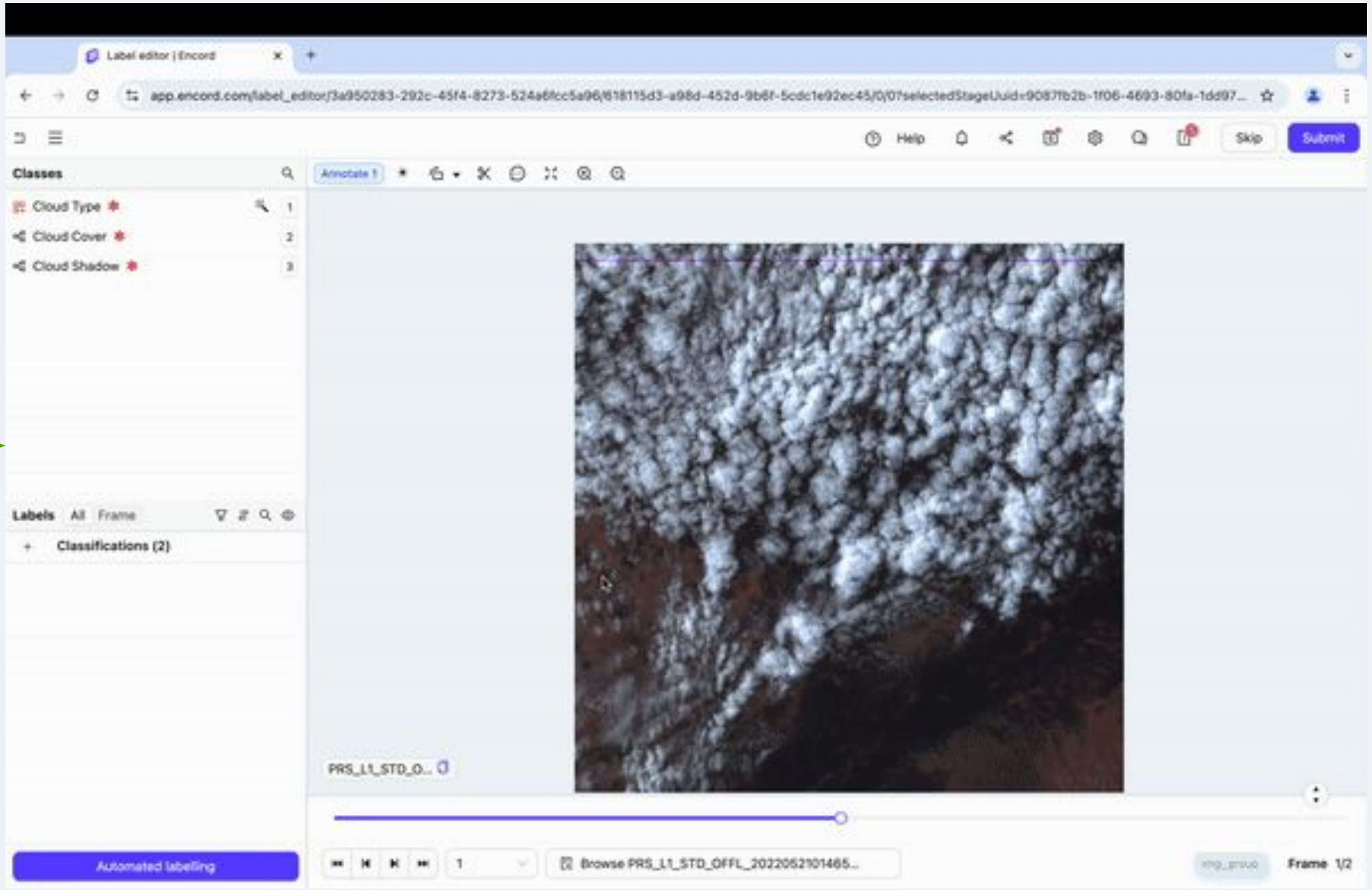
Python API for convenient pre-annotations

Annotations are orders of magnitude **faster** and increasingly **easier**

Inference using ViT on 256 x 256 patches



pre-annotated .tif



# Icons