

Optimising Onboard Earth observation AI for robustness against Sensor Degradations

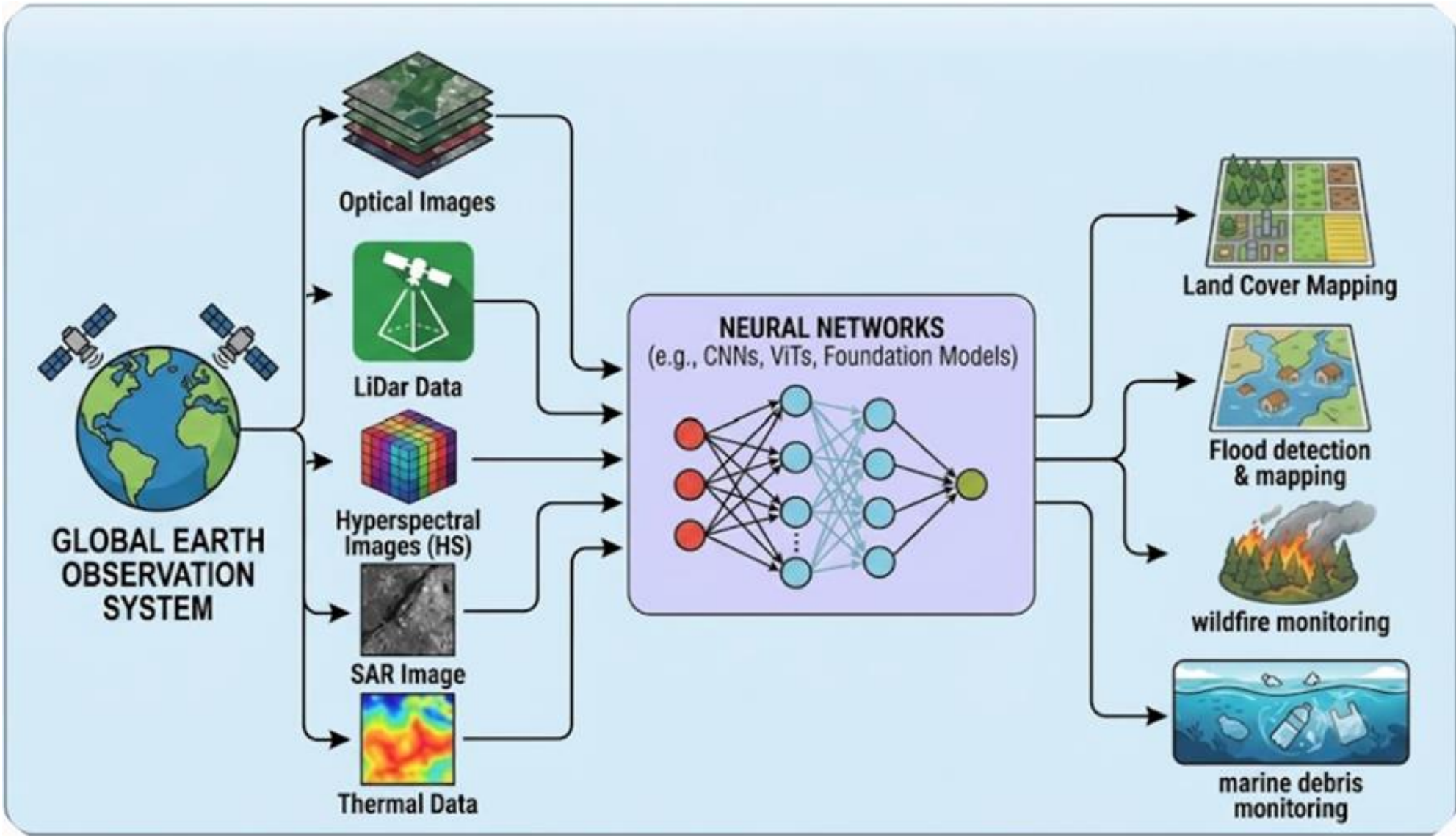
Ivar van der Spoel¹, **Lorenzo Papa**², Roberto Del Prete^{2,3}, Andre Deutz¹, Gabriele Meoni²,
Jan N. van Rijn¹

¹ LIACS, Leiden University

² Φ -lab, European Space Agency

³ Setanta Space

Introduction: AI for Earth observation (EO)



Problem: latency, bandwidth, degraded images.



Introduction: PhiSat-2 Mission[1]

- **ESA** funded mission for **€3 million**,
- **Multi-spectral Sensor**: 8 Bands (RGB, Pan, NIR1, NIR2, RE1, RE2),
- **On-board AI**: uses the **Intel Movidius Myriad 2 VPU**

(Small CNN models only!)

- **ESA's aim**: prove that AI apps can run directly on a small satellite.

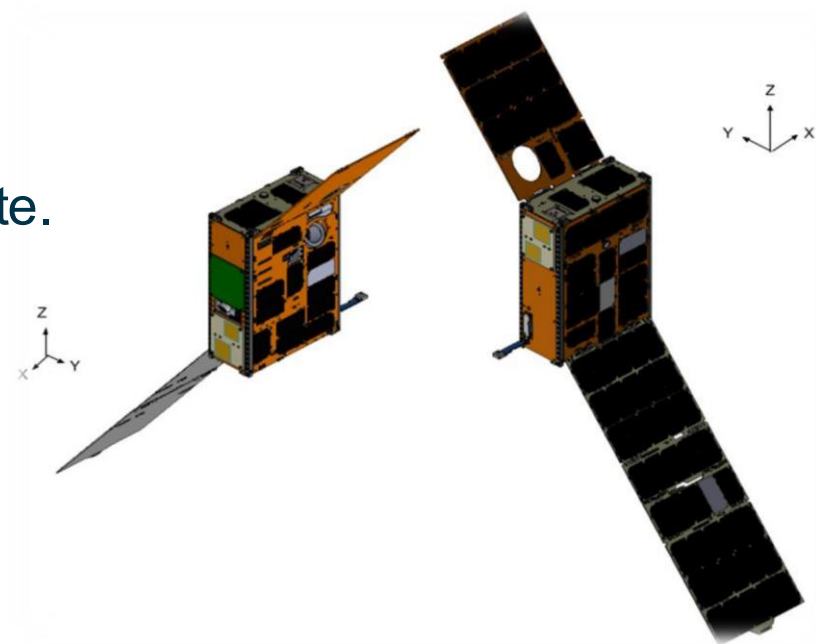


Figure 1 Φ sat-2 Deployed Configuration

Motivation: Why robustness matters in EO

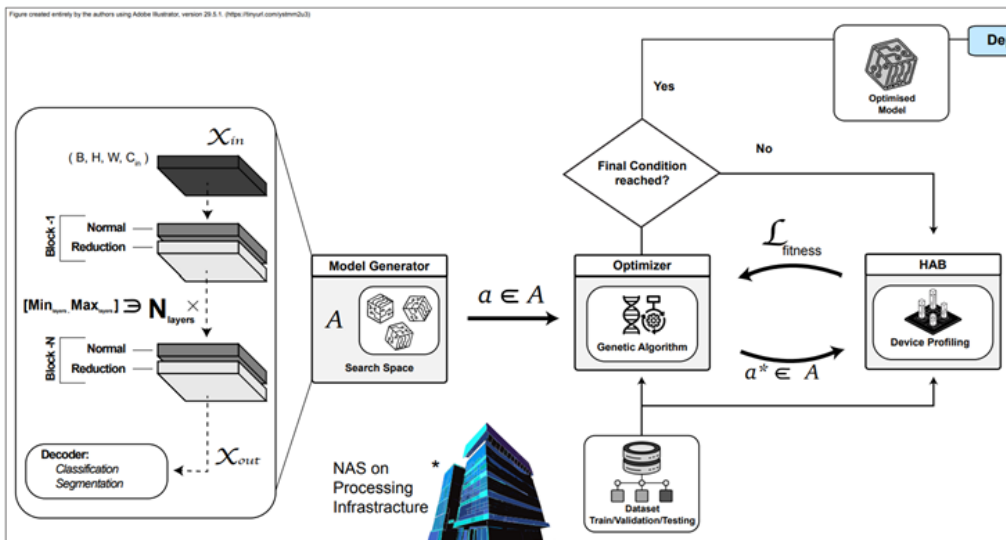
EO models are trained & evaluated on (relatively) clean datasets.

Data distribution shifts in deployment, the input distribution can change [2]

We will do this through ***Perturbation Methods $C(x)$***

Regular image: x → Apply perturbation method → Perturbed image: $C(x)$

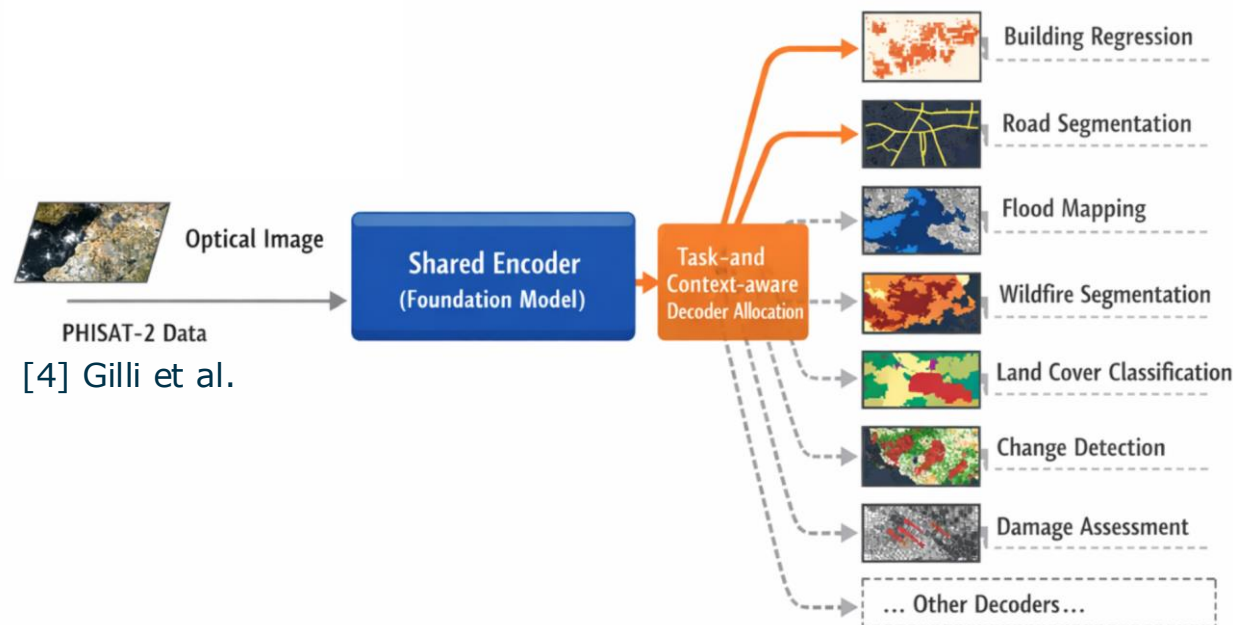
Approaches: Task-specific NAS vs Task-agnostic FM



[3] Del Prete et al.

Design models using
Neural Architecture Search (NAS)
 for a specific downstream task with
multi-objective optimization (MOO)

Use phisat-2 pre-trained
Foundation Model encoder
 Φ SatNet
 for agnostic downstream tasks



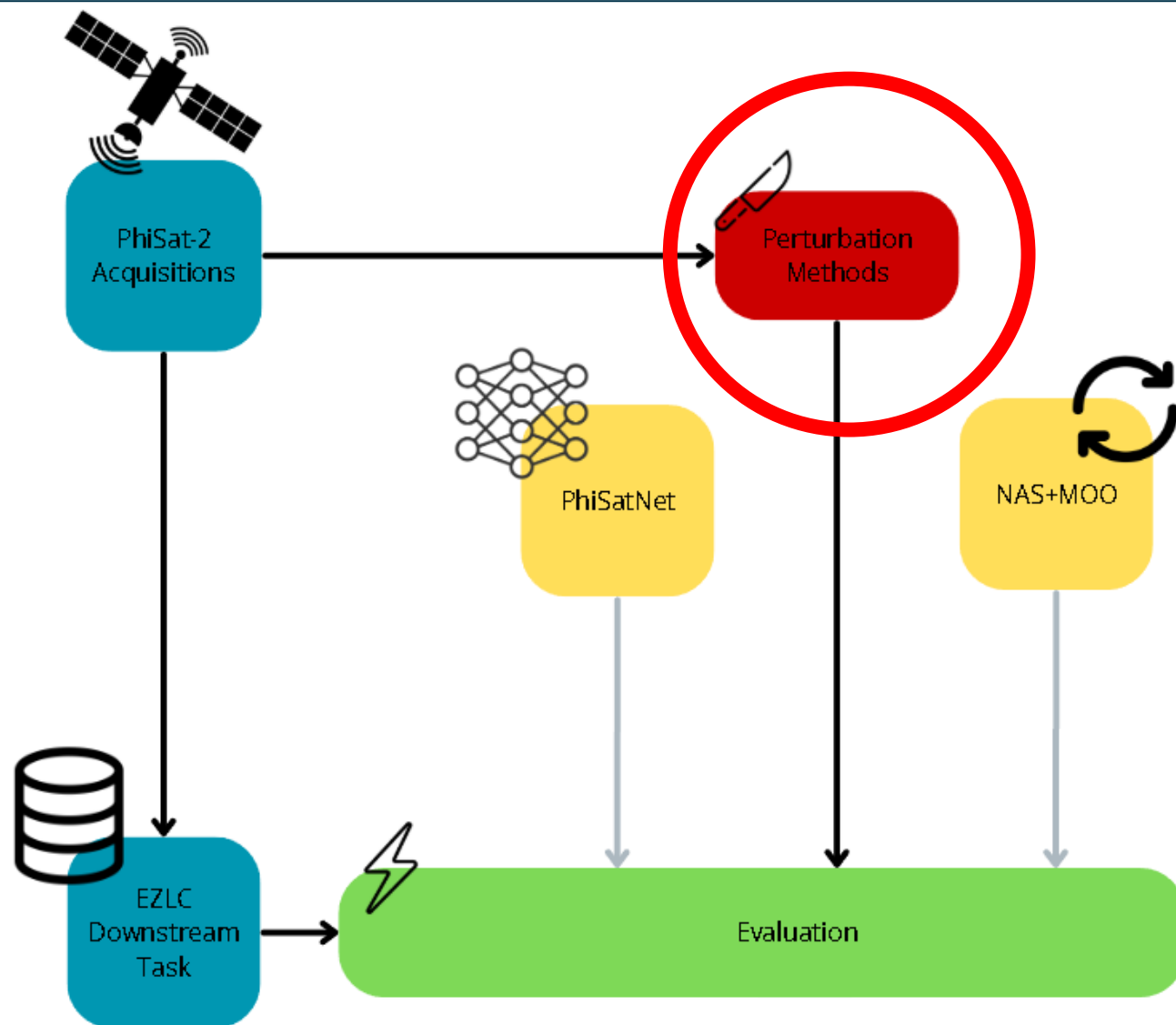
[4] Gilli et al.

"How can we evaluate and improve the robustness of compact AI models operating on real Φ Sat-2 imagery against sensor, environmental, and onboard-processing degradations?"

Main Contributions:

1. A **semantic-segmentation task** based on real Φ Sat-2 imagery.
2. A **sensor-aware robustness evaluation** framework for multispectral imagery.
3. A **robustness-aware multi-objective NAS** procedure
4. A comparison of **feature-level Φ SatNet adaptation**.

Types of degradation through perturbations



We consider 5 types of perturbation:

1. Sensor Noise
2. Optical Blur
3. Atmospheric Haze
4. Brightness
5. Band misalignment

Examples of degradation through perturbations

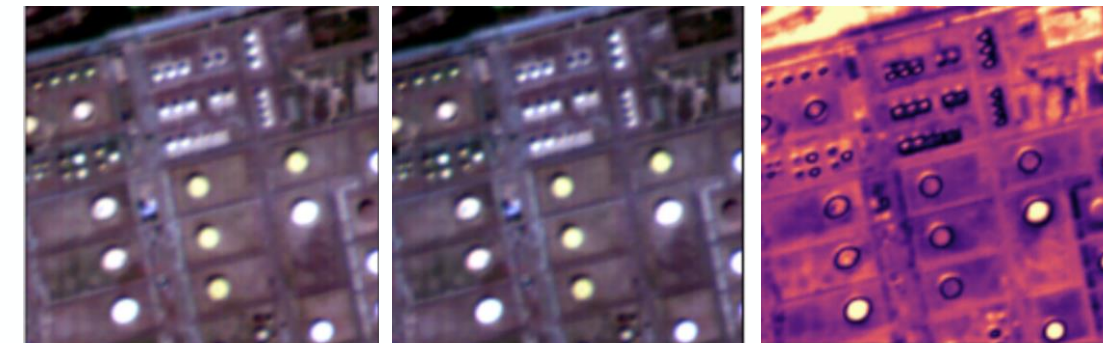
Sensor noise: measurement errors



Optical blur: loss of spatial detail



Atmospheric haze: scene mixed with air light



Data: Downstream Task - Easy LandCover (EZLC)

Masks: ESA WorldCover 2021 (Sentinel dataset)[7]

→ labels projected onto each patch for ΦSat-2

Aligned ΦSat-2 patches (8x256x256)

Resolution mismatch:

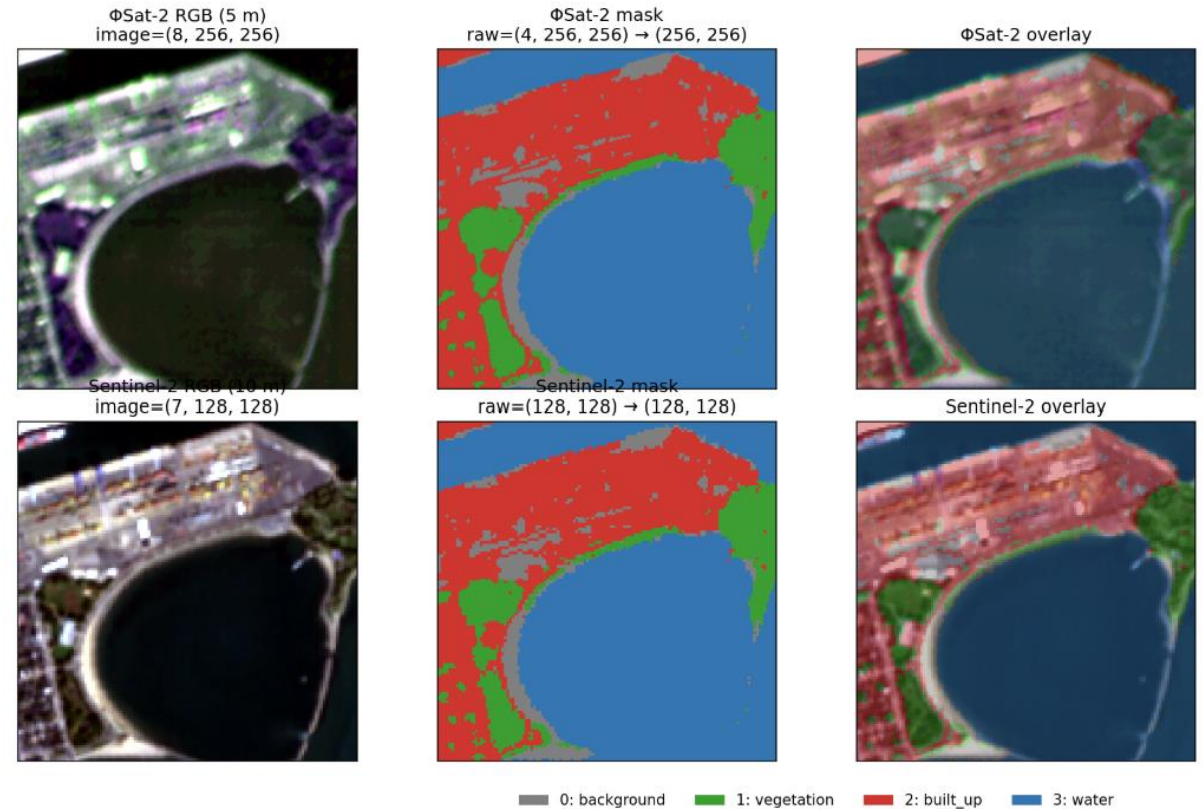
Sentinel @ 10m/pixel

ΦSat-2 @ 5m/pixel

→ Nearest Neighbor to generate masks

WorldCover classes into 4 classes:

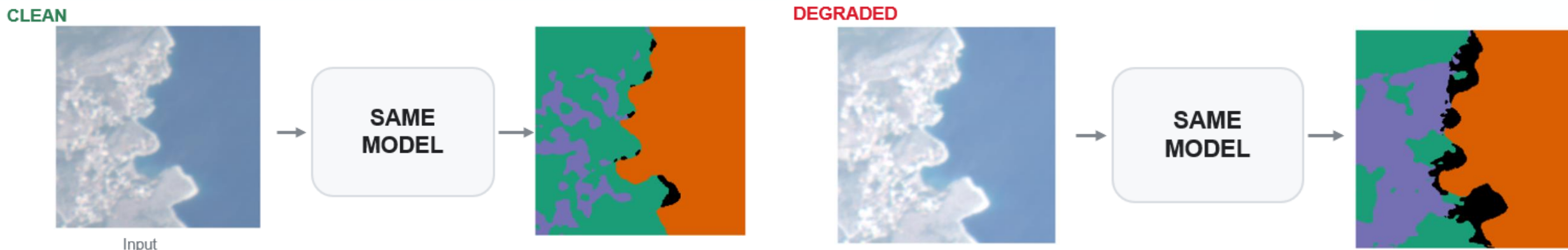
- background
- vegetation
- built-up
- Water



Stratify for specific task: EZLC (25% present each class)

How do we evaluate robustness?

Evaluate the same Φ Sat-2 observation clean and degraded, then compare both correctness and consistency.



TASK PERFORMANCE

mIoU ↑

How correct is the segmentation compared with the EZLC target?

PREDICTION CONSISTENCY

Consistency ↑

How much does the model change its prediction between the clean and degraded view?

INSIDE Φ SATNET

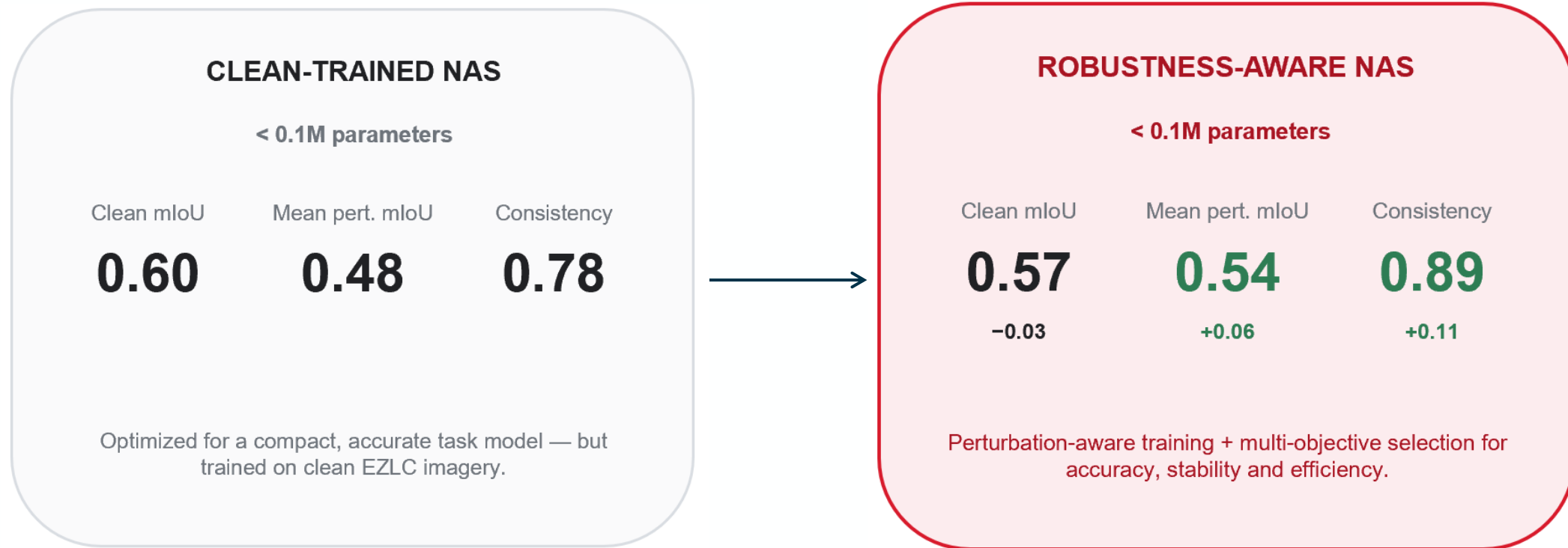
Feature similarity

Compare clean and perturbed encoder features using MSE, cosine, CKA and MMD.

Are models actually vulnerable?

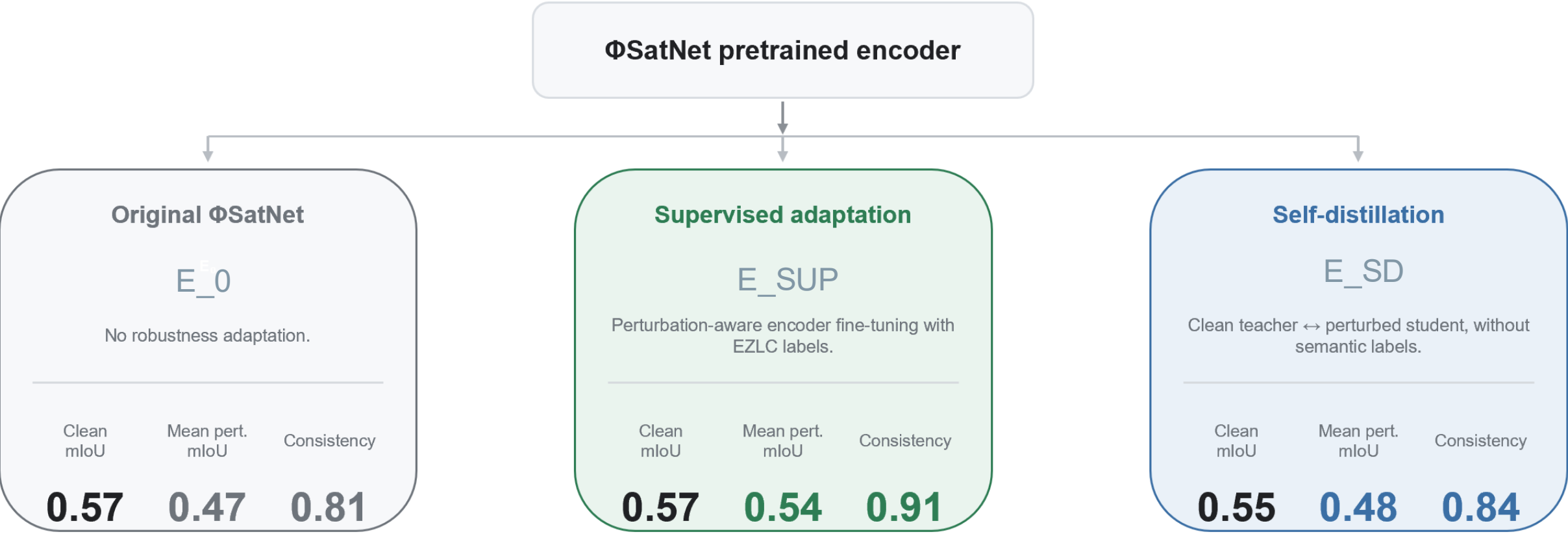
Model	Param#	Clean	Δ perturbed mIoU
PhiSatNet (Frozen)	<u>0.35M</u>	<u>0.57 mIoU</u>	0.47 (Δ -0.10)
NAS	<u>0.02M</u>	<u>0.60 mIoU</u>	0.48 (Δ -0.12)
ResNet-18[8]	12M	0.62 mIoU	0.53 (Δ -0.08)
ViT-S/16[9]	22M	0.54 mIoU	0.49 (Δ -0.05)
TerraMind[10]	<u>91M</u>	<u>0.74 mIoU</u>	0.68 (Δ -0.06)

Can NAS find a compact robust model?



Main result: +0.06 mean perturbed mIoU and +0.11 consistency, for only -0.03 clean mIoU.

Can we make the Foundation Model robust?



Supervised adaptation is strongest; self-distillation is weaker but demonstrates label-free robustness adaptation.

Which model should we actually use?

On the ground: **TerraMind!**

In space: *it depends..*

NAS with Multi-Objective Optimization:

- Best choice for a single task
- Good balance between performance, robustness and efficiency
- Search space can even be further explored!

Supervised PhiSatNet

- Best choice if you do perturbation-aware fine-tuning for every head
- Strongest robustness improvements out of NAS and FM

PhiSatNet with Self Distillation:

- When labels are scarce or to guarantee robustness to other tasks
- Weakest robustness
- Useful when reusable encoder is most important!

- 16

Thank you for listening!