An AI Foundation Model for Earth Observation, Techniques and Applications Including Flood Rapid Mapping

Nikolaos Dionelis[1], Jente Bosmans[1], Ziyang Zhang[3,1], Riccardo Musto[2], Giancarlo Paoletti[2], Casper Fibaek[1], Nicolas Longépé[1]

1 - European Space Agency (ESA), Φ-lab, ESRIN
2 - Leonardo Labs, Italy
3 - Lancaster University, UK

Workshop on “Resilience to Natural Hazards through AI Solutions”
International Telecommunication Union (ITU)

Contents

1. **ESA Φ-lab** and presenter's background
2. **EO Foundation Model** PhilEO
3. **Evaluation framework** for EO Foundation Models
4. **Techniques**
   - Confidence quantification & assessment
     a) Classification, b) Regression
5. **Applications**
   - Flood Rapid Mapping
   - Building density regression, road segmentation, land cover mapping, estimation of age of buildings, etc

The ESA Φ-lab Offices

Φ-lab Explore Office

Explores the innovation universe and connects together EO and digital revolution

A team of Researchers
and innovation seed funding (FutureEO)

Φ-lab Invest Office

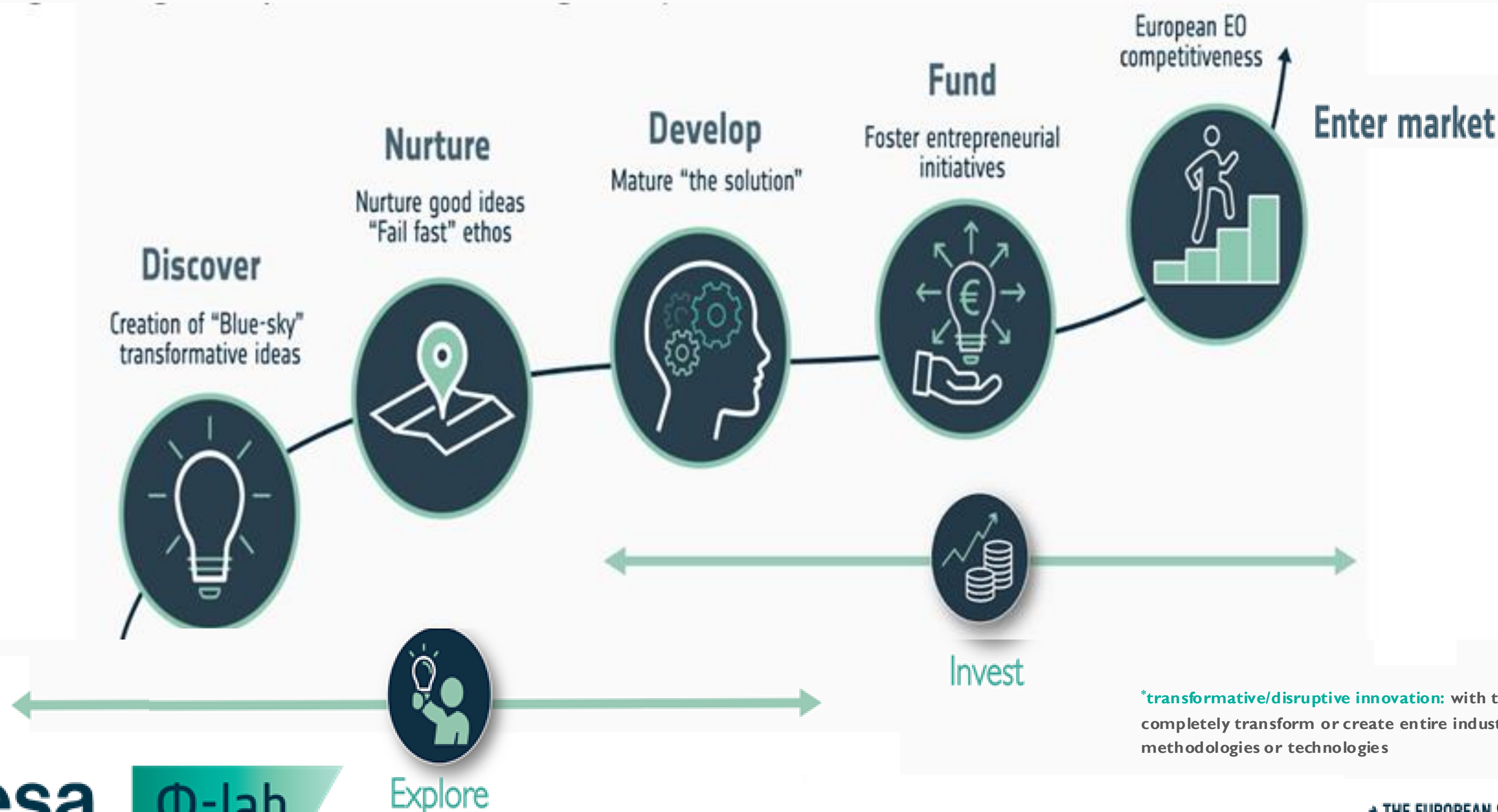
Stimulates competitiveness by fostering the growth of entrepreneurial initiatives through investment actions from ESA Member States and private investors

A team of business innovators and
a commercial co-funding programme (InCubed)

The ESA Φ-lab – What ?

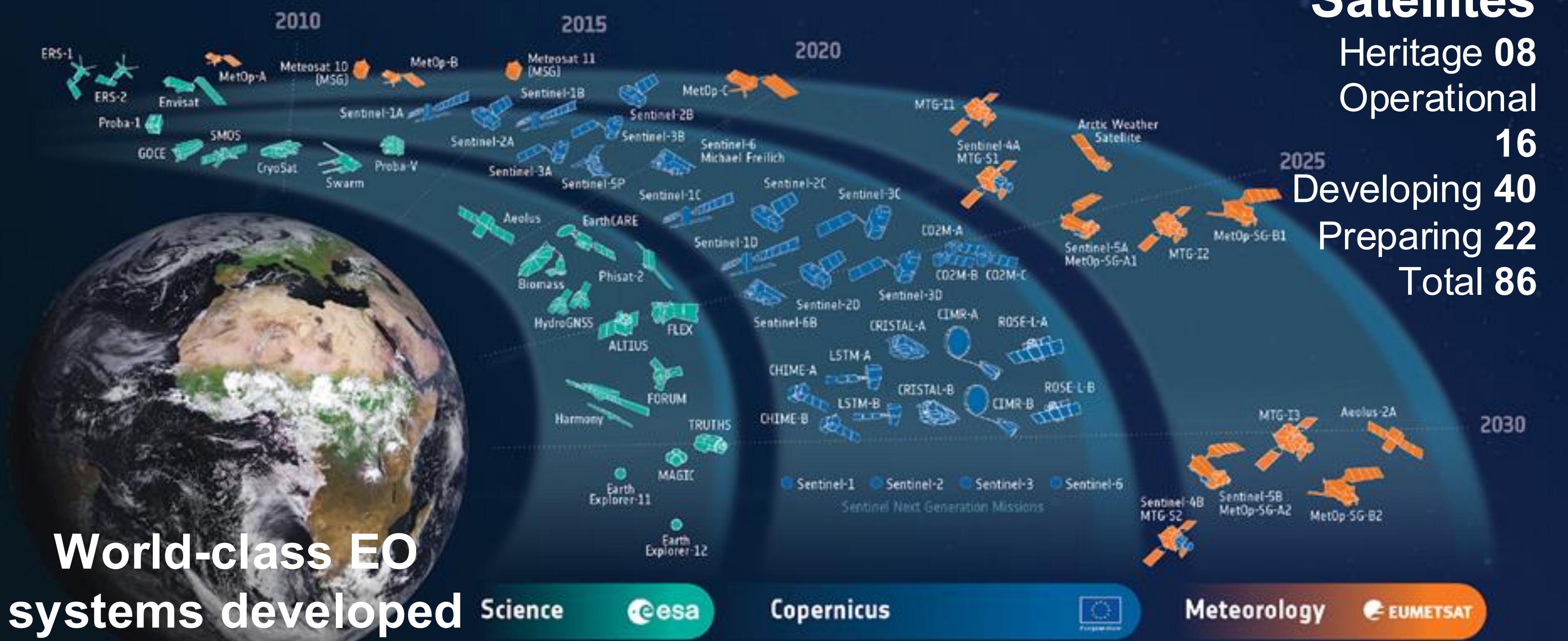
Accelerate the future of Earth Observation via transformative innovation* and commercialisation actions strengthening Europe's world-leading competitiveness

Uniquely in ESA Φ-lab **innovate and apply** under-one-roof



*transformative/disruptive innovation: with the ability to completely transform or create entire industries via new methodologies or technologies

ESA's Earth Observation Missions

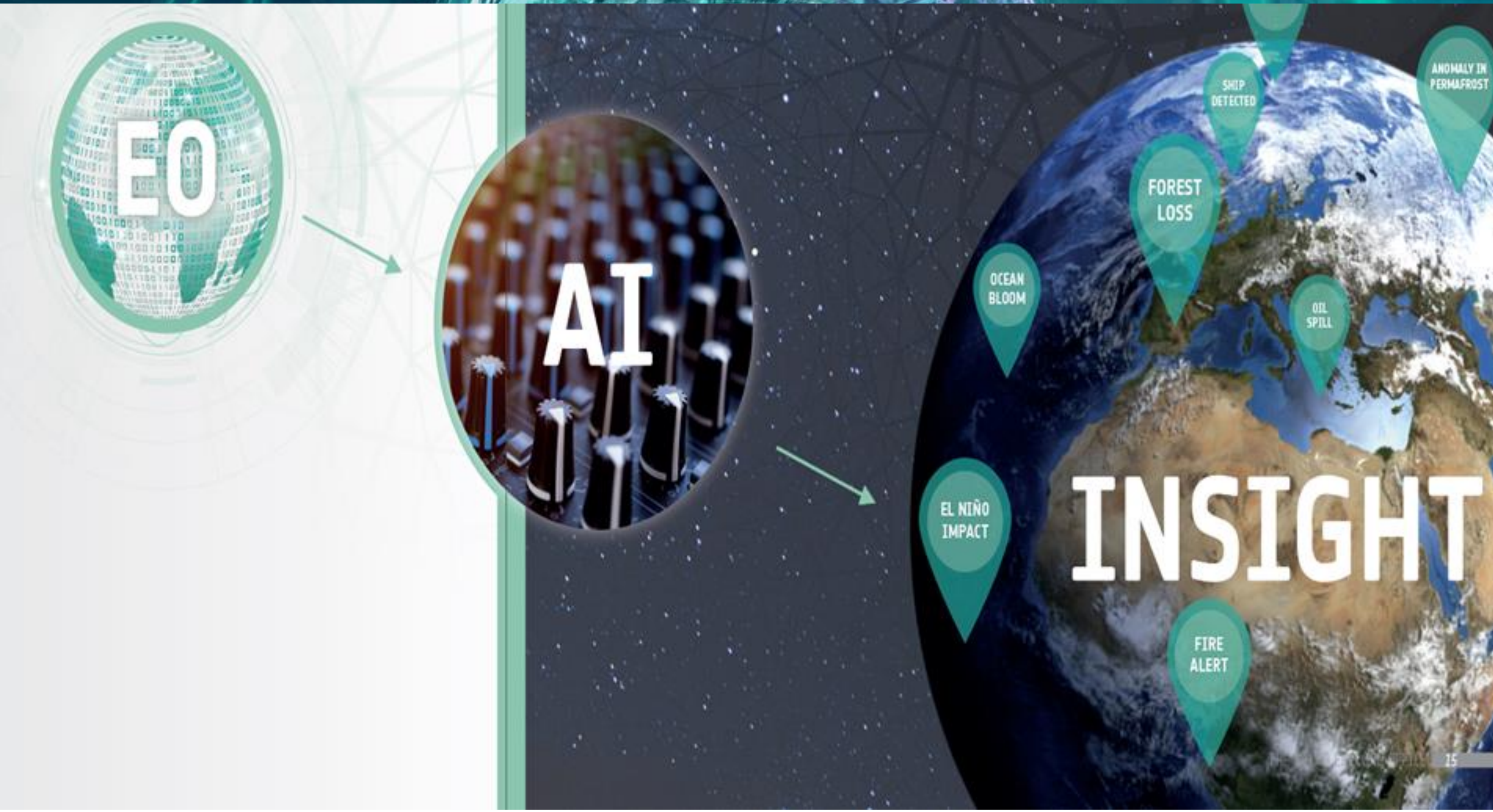


Satellites

Heritage **08**
Operational **16**
Developing **40**
Preparing **22**
Total **86**

World-class EO systems developed with European and global partners to address scientific & societal challenges

The ESA Φ-lab – Why ?



From Earth Observation to Earth Action

From data to actionable information

Innovation Technologies axis and Applications

AXIS I
Augmented Intelligence

Foundation Models
Generative AI
Decision Intelligence, AI Agent
Explainability (xAI)
Physics-Informed ML
Digital Assistant and Twins
AI4EO for Climate, Health, and Human

AXIS II
Innovative Computing Paradigms

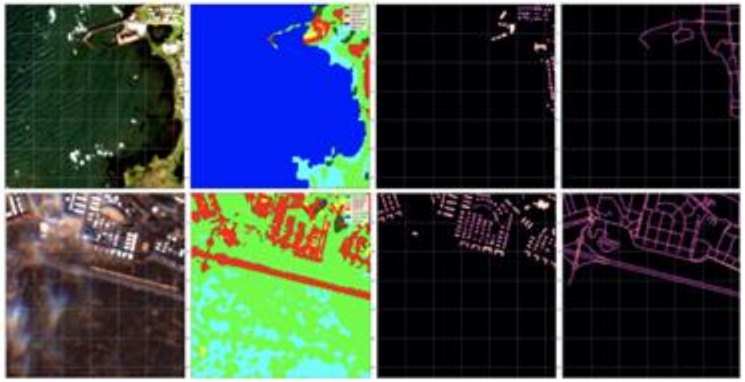
On-Board AI
Quantum Computing
Hybrid HPC Computing
Neuromorphic
Biocomputing and others

AXIS III
Innovative Computing Paradigms

Cognitive Space
VR/AR Immersive Visualisation
Web 3.0
IOT
Distributed Ledgers/Blockchain

Presenter’s background - Nikolaos Dionelis

- **European Space Agency:**
    - Internal Research Fellow at EOP-S Φ-lab Explore (since 04/2023)
  - **Research:** AI for Earth Observation (EO) (AI4EO)
    - Foundation Models & downstream tasks for EO
  - **Devise new** and adapt existing SoTA models
  - **Programming & extensive experimentation:**
    - Training, evaluation on large-scale data
    - Comparison against benchmarks
  - **Write invitations to proposals** & evaluate

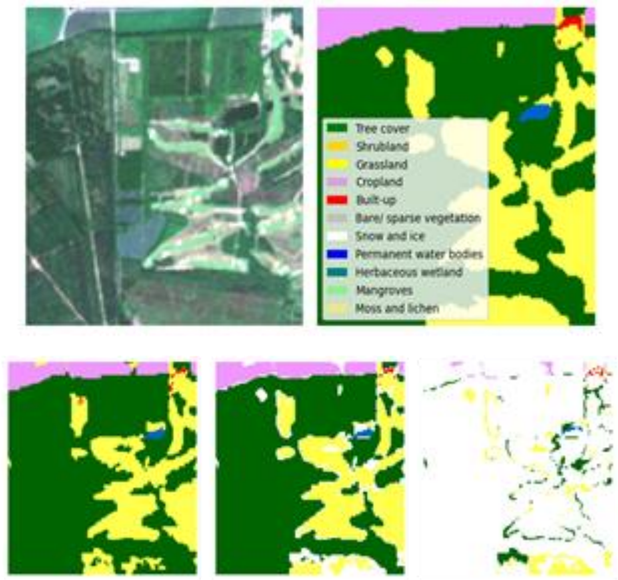


Sentinel-2, land cover labels, **building** density, road density

Foundation Model PhilEO

- **Problem:** Extract information from unlabelled satellite data in order to solve a number of downstream tasks, e.g. land cover

- **Method:** Develop a Foundation Model for EO
  - Use large amounts of geo-aware satellite data to pretrain and then fine-tune for downstream tasks
  - Geo-aware self-supervised learning
    - Geo-location: Longitude & latitude prediction
    - Reconstruction loss: Masking, MAE
  - Architectures: ViT, U-Net



From top left: Sentinel-2, Prediction by our model, Ground truth, *Correct* classifications of the model, Incorrect classifications

Foundation Model PhilEO (Ctd)

- Datasets:
  - Global dataset PhilEO Globe, 500 GB (new)
  - Subset of MajorTOM, FastTOM, 2 TB
  - Scale-up to MajorTOM, 23 TB
    - In collaboration with Leonardo Labs

- Downstream tasks:
  A. Flood Rapid Mapping
  B. Building density pixel-wise regression, road regression segmentation, land cover semantic segmentation

Challenge: Limited labelled data

Dataset	Modality	Number of Patches	Sensing Type	Comments
Core-S2L2A	Sentinel-2 Level 2A	2,245,886	Multi-Spectral	General-Purpose Global (about 23 TB)
Core-S2L1C	Sentinel-2 Level 1C	2,245,886	Multi-Spectral	General-Purpose Global (about 23 TB)

Foundation Model PhilEO (Ctd)

- Fine-tune on: Dataset PhilEO Bench, 400 GB (new)
  - Global, labelled

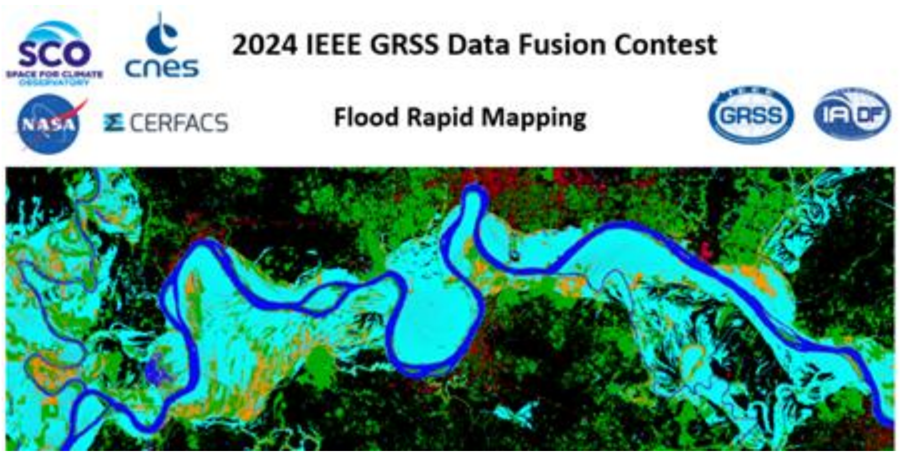
- Evaluate PhilEO FM:
  - PhilEO Bench
  - Achieve improved performance on downstream tasks

Left: S-2 data: Visualization in **RGB** (3 bands) Middle: Prediction by model. Right: Ground truth labels: Classes like **Cropland**

- RGB colours defined by **ESA WorldCover**

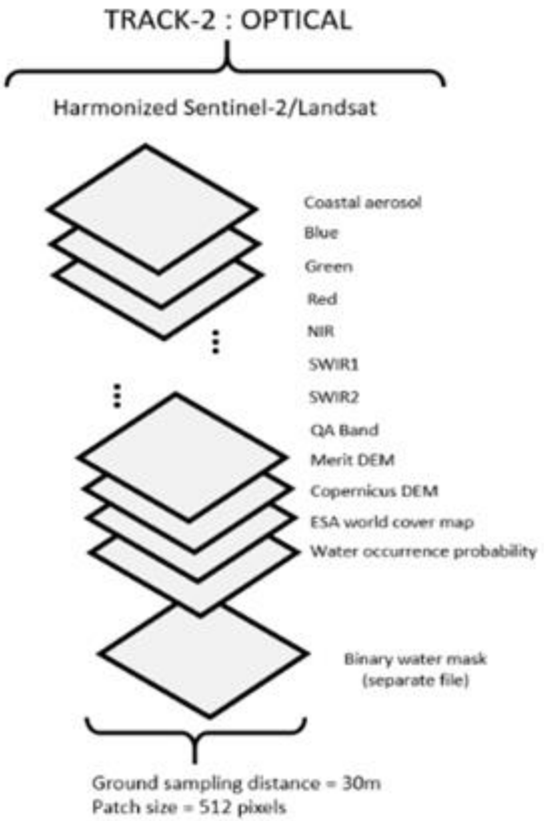
Foundation Model PhilEO (Ctd)

- **Results:**
    - Flood Rapid Mapping
        - Joint work with: Ziyang Zhang, ESA Φ-lab Visiting researcher, Lancaster University, UK
        - IEEE GRSS Flood Rapid Mapping dataset
            - IGARSS 2024 Challenge
            - Track-2: Flood rapid mapping with optical data
                - Sentinel-2 satellite multi-spectral images
            - http://www.grss-ieee.org/community/technical-committees/2024-ieee-grss-data-fusion-contest



Foundation Model PhilEO (Ctd)

- Flood Rapid Mapping (Ctd):
  - IEEE GRSS Flood Rapid Mapping labelled dataset
  - Harmonized Landsat Sentinel-2 (HLS) data
    - Multi-spectral 30m res images
  - **Labels:** Flood extent labeled by Copernicus Emergency Management Service (CEMS)
  - http://ieee-dataport.org/competitions/2024-ieee-grss-data-fusion-
  - **Evaluation:** F1-score
    - 0.84348, ZiyangZhang, 28
    - 0.89843, Henryljp, 1
    - Difference: 0.05495
      - Approximately 5 points

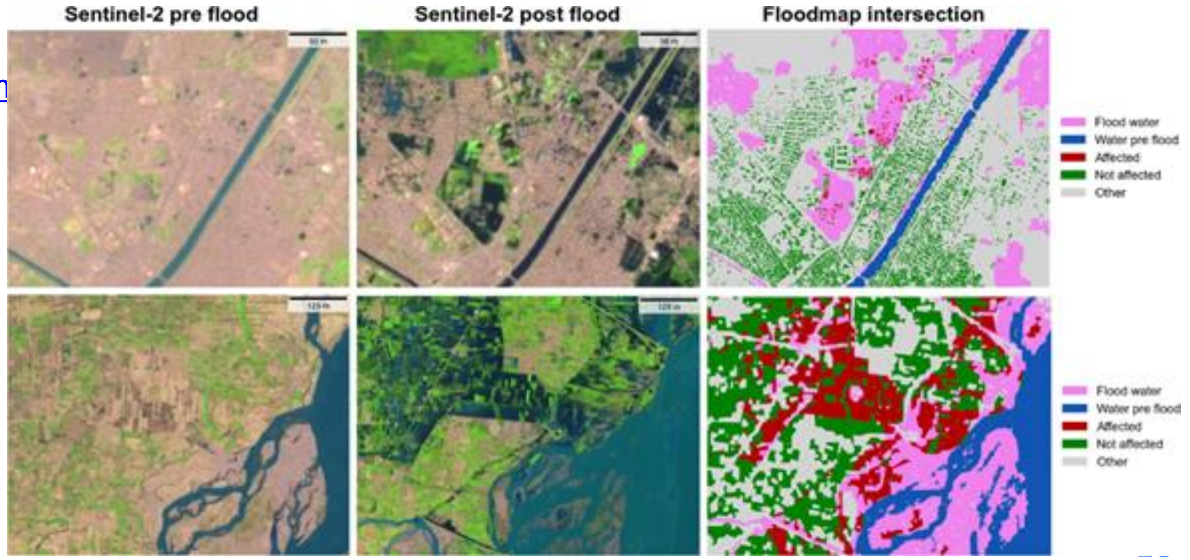


Foundation Model PhilEO (Ctd)

- Flood Rapid Mapping (Ctd):
  - EO Foundation Model approach
    - Fine-tuning
  - PhilEO Foundation Model
  - Ensemble of EO Foundation Models
    - Like Mixture of Experts (MoE)
    - PhilEO, Satlas, http://github.com/allenai/satlaspretrain_models
  - F1-score, 0.84348, ZiyangZhang, 28
  - http://codalab.lisn.upsaclay.fr/competitions/16699#results
  - Paper: Ongoing work: Ziyang Zhang, Nikolaos Dionelis, Plamen Angelov, and Nicolas Longépé, "*An Interpretable Ensemble Geoscience Foundation Model for Flood Mapping*," To be submitted, 2025.
  - http://github.com/Ziyang-cyber/challenge_philab

Foundation Model PhilEO (Ctd)

- **Results (Ctd):**
  - Floods
  - WorldFloods database
    - E. Portalés-Julià, et al., “*Global flood extent segmentation in optical satellite images*,” Nature Scientific Reports, 13, 20316, 2023.
      - http://www.nature.com/articles/s41598-023-47595-7
    - Sentinel-2 multi-spectral data
    - Flood segmentation masks
    - 509 pairs
  - http://spaceml-org.github.io/n



Φ-lab

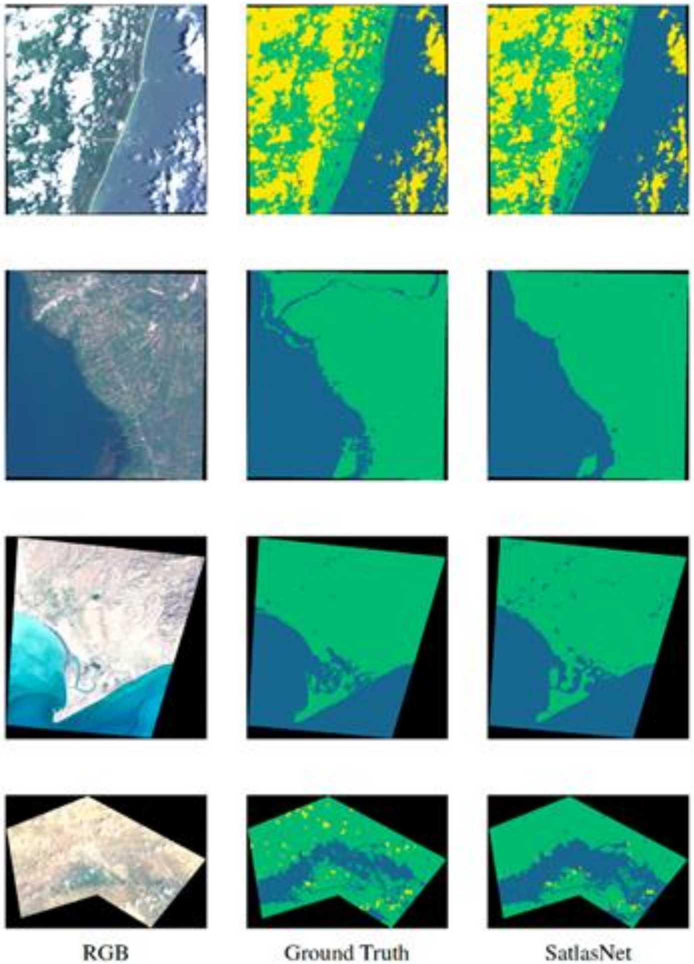
Foundation Model PhilEO (Ctd)

- WorldFloods database (Ctd):
  - Ensemble: 4 models: PhilEO, Satlas, DOFA, Prithvi
  - Semantic segmentation: 3 classes: Water, land, cloud
  - Evaluation metric: Intersection over Union (IoU)
  - Evaluation of: i) **PhilEO** model, and ii) **Ensemble**

Model	IoU land (%)	IoU water (%)	IoU cloud (%)	Mean IoU (%)
Prithvi	86.92	73.76	66.71	75.80
DOFA	83.98	69.90	69.10	74.33
Satlas	87.28	75.28	76.78	79.78
U-Net	87.39	77.21	80.02	81.54
DeepLabv3+	88.10	77.45	81.21	82.25
PhilEO	85.58	74.09	77.61	79.09
Ensemble	**88.61**	**78.05**	**81.37**	**82.67**

Foundation Model PhilEO (Ctd)

- WorldFloods database (Ctd):
  - Potential future work: Change detection approach in addition to fine-tuning method



Results on the WorldFloods dataset.

Model	IoU land	IoU water	IoU cloud	Mean
Satlas	87.28	75.28	76.78	79.78
PhilEO	85.58	74.09	77.61	79.09
Ensemble	**88.61**	**78.05**	**81.37**	**82.67**

Φ-lab

Foundation Model PhilEO (Ctd)

- WorldFloods database (Ctd):
  - Evaluation of EO Foundation Models: Generalize across different spectral bands

Model	# Bands	IoU land (%)	IoU water	IoU cloud	Mean IoU
Prithvi	6	86.55	72.00	67.17	75.05
Prithvi	13	86.92	73.76	66.71	**75.80 (↑0.75)**
DOFA	9	82.01	68.58	53.93	68.17
DOFA	13	83.98	69.90	69.10	**74.33 (↑6.16)**
PhilEO	10	86.39	73.51	67.31	75.74
PhilEO	13	85.58	74.09	77.61	**79.09 (↑3.35)**
Satlas	9	87.75	74.39	68.38	76.84
Satlas	13	87.28	75.28	76.78	**79.78 (↑2.94)**
U-Net	13	87.39	77.21	80.02	81.54
DeepLabv3+	13	**88.10**	**77.45**	**81.21**	82.25

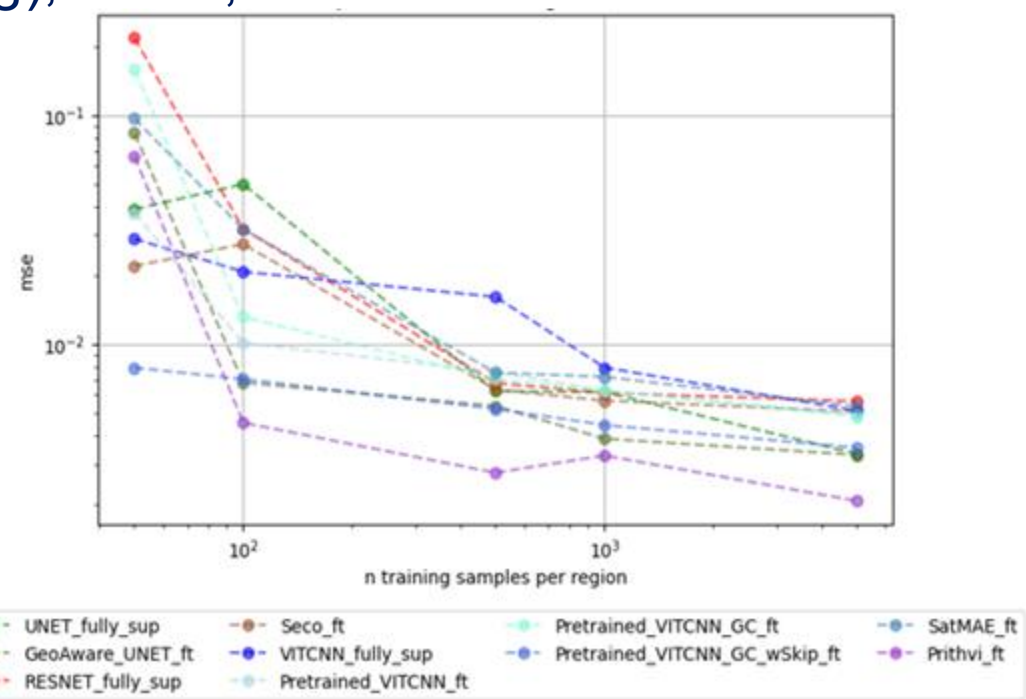
Foundation Model PhilEO (Ctd)

- Summary & number of parameters of models
  - Foundation Models for EO and Geospatial AI

Model	Pretraining dataset	Pretraing image type	# Parameters (M)
U-Net	ImageNet	RGB	7.8
DeepLab v3+	ImageNet	RGB	45.7
PhilEO	MajorTOM subset	S-2	45.9
Satlas	SatlasPretrain	S-2	89.8
Prithvi	Harmonized Landsat S-2	Landsat S-2	127
DOFA	DOFA	S-1, S-2, GaofenNAIP, EnMAP	151

Foundation Model PhilEO (Ctd)

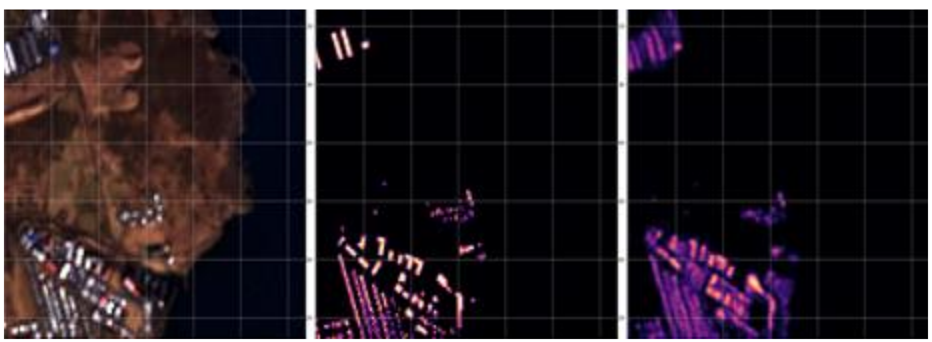
- **Results (Ctd):**
  - Building density regression
    - Dataset: Custom, labelled
    - Evaluation metric: Mean squared error (MSE)
    - PhilEO (GeoAware_UNET_ft) outperforms ResNet (supervised learning), SeCo, SatMAE

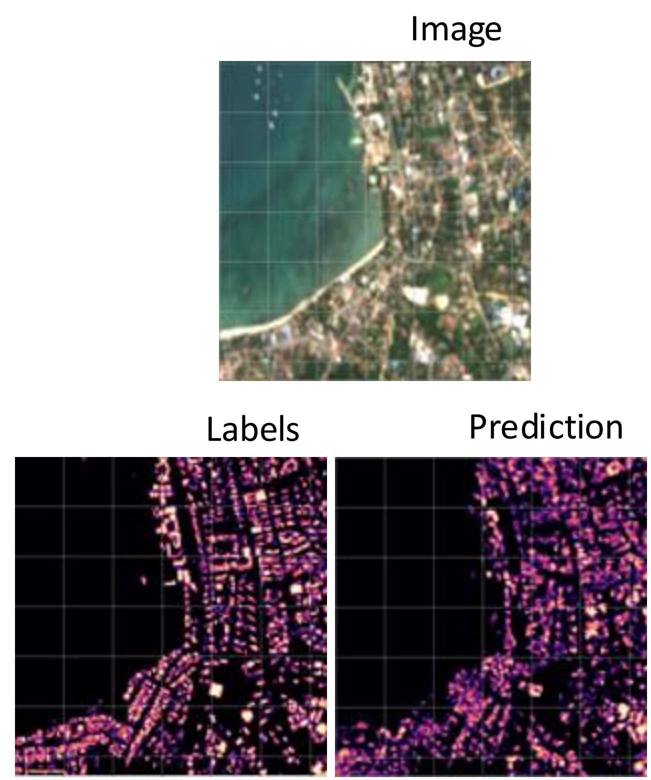


Foundation Model PhilEO (Ctd)

- Building density regression (Ctd)
  - Qualitative evaluation
    - Images:
      - Left: Sentinel-2 input
      - Middle: Ground truth
      - Right: Prediction

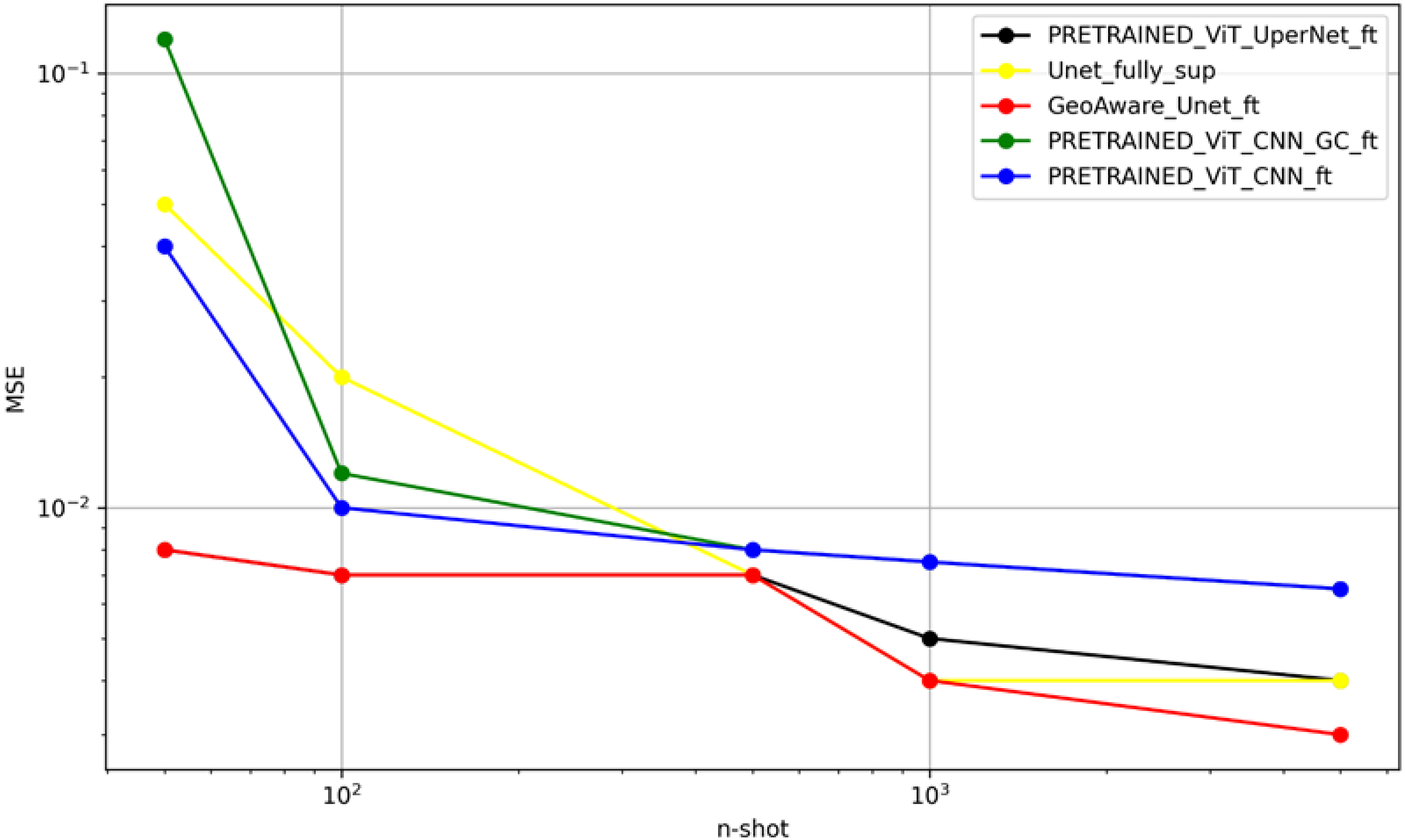
Geo-Aware U-Net:





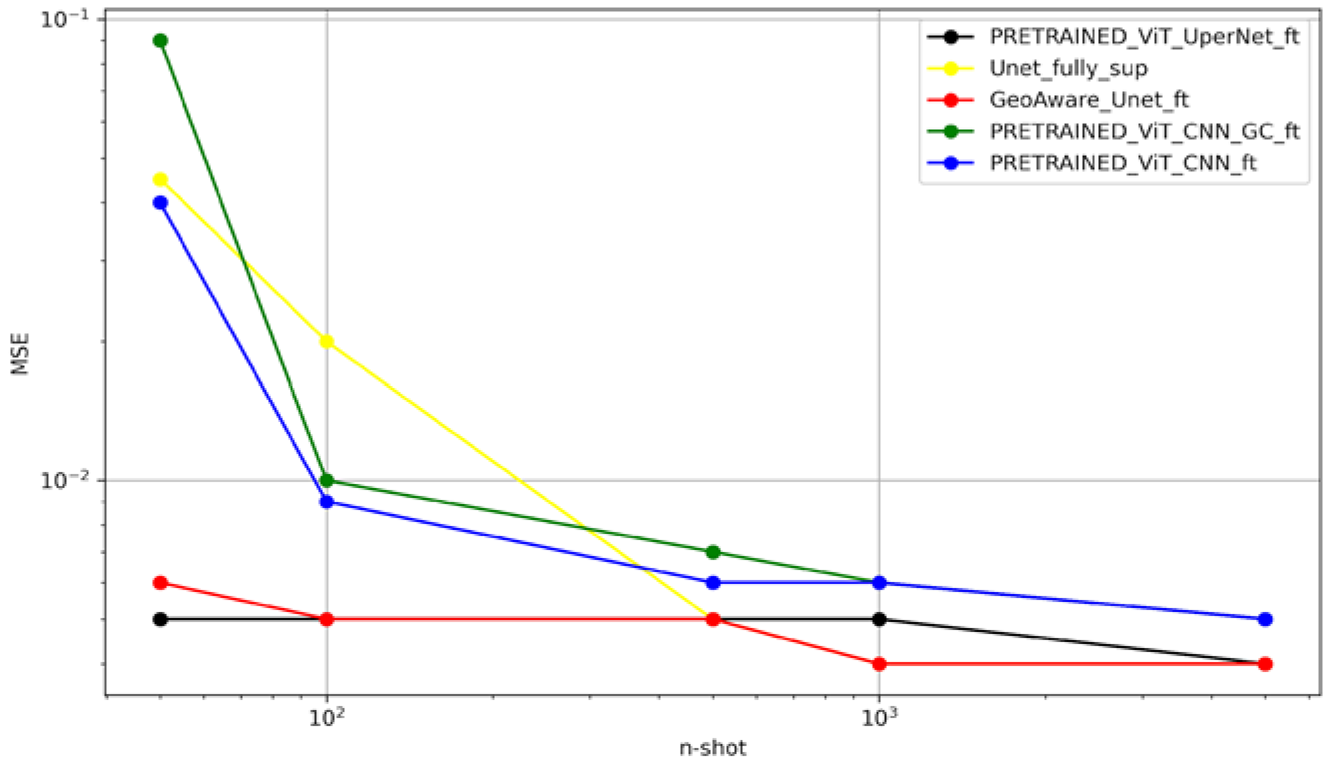
Foundation Model PhilEO (Ctd)

- Building density regression (Ctd):
  - Decoder: UPerNet



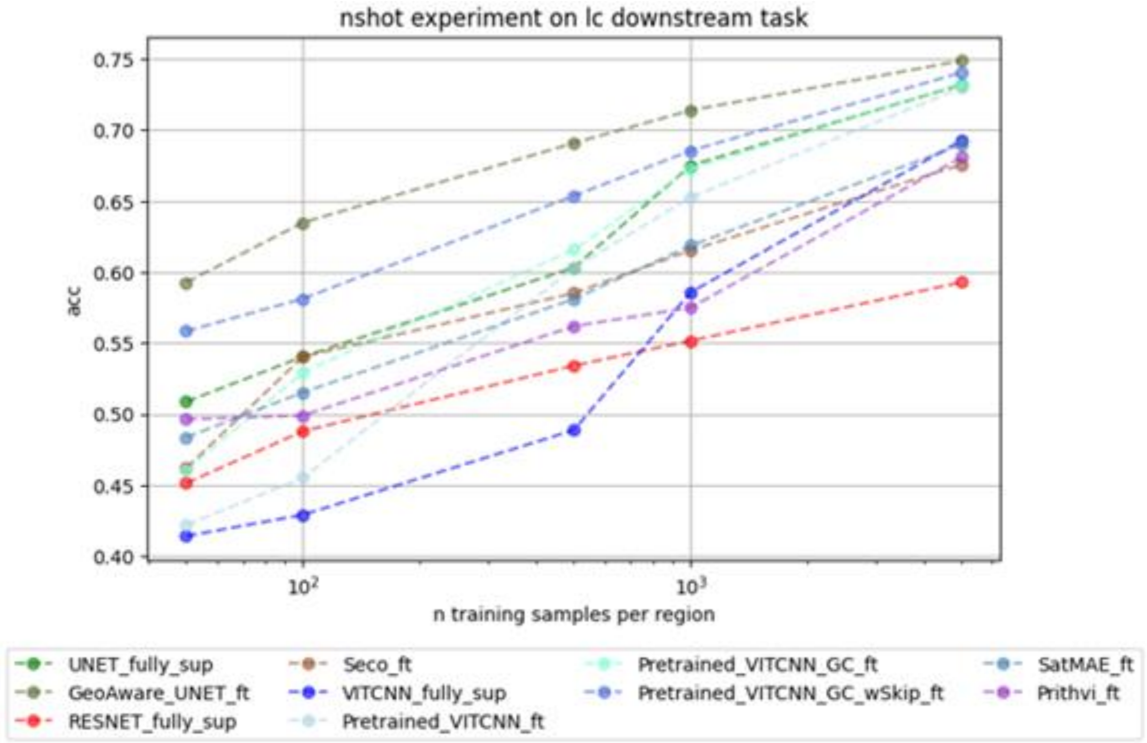
Foundation Model PhilEO (Ctd)

- **Results (Ctd):**
  - Road density regression
    - Evaluation metric: MSE
    - Decoder: UPerNet
    - Roads: Evacuation planning, emergency disaster management



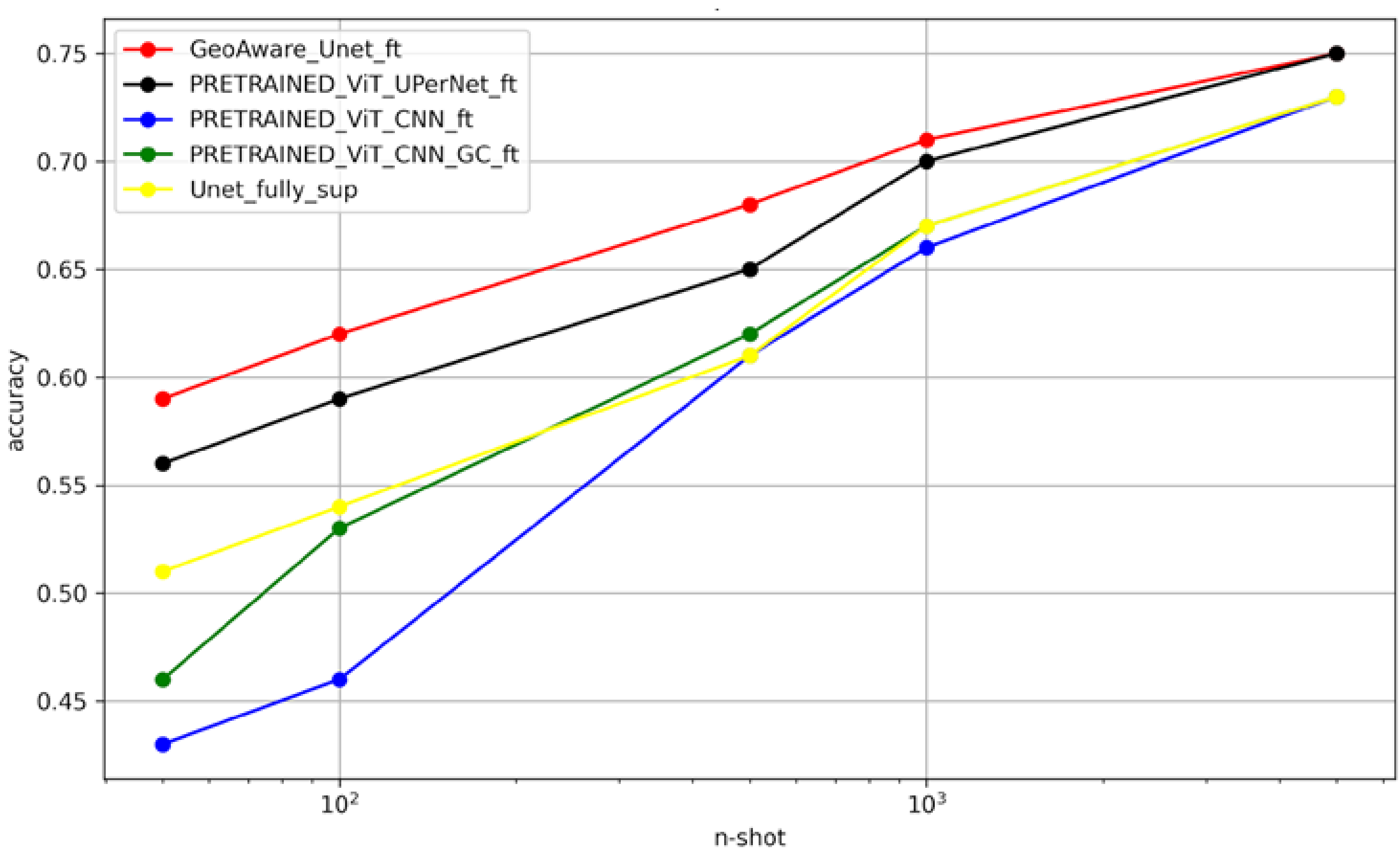
Foundation Model PhilEO (Ctd)

- **Results (Ctd):**
  - Semantic segmentation land cover classification (lc)
    - Dataset: ESA WorldCover (11 classes)
    - Evaluation metric: Accuracy (acc)
    - PhilEO (GeoAware_UNET_ft) outperforms other models



Foundation Model PhilEO (Ctd)

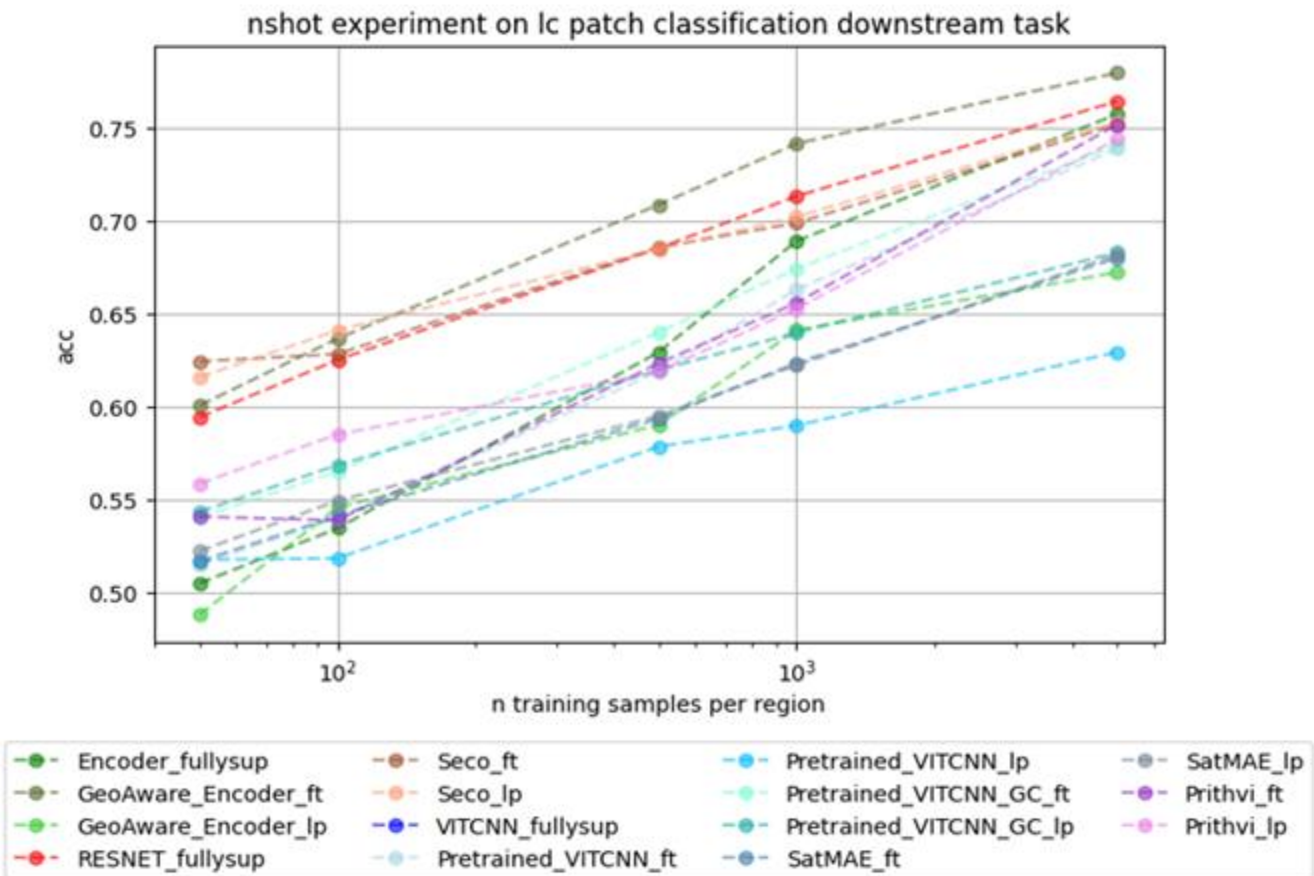
- Semantic segmentation land cover classification (Ctd):
  - Decoder: UPerNet
  - http://collections.sentinel-hub.com/worldcover/readme.html, http://collections.sentinel-hub.com/worldcover



Φ-lab

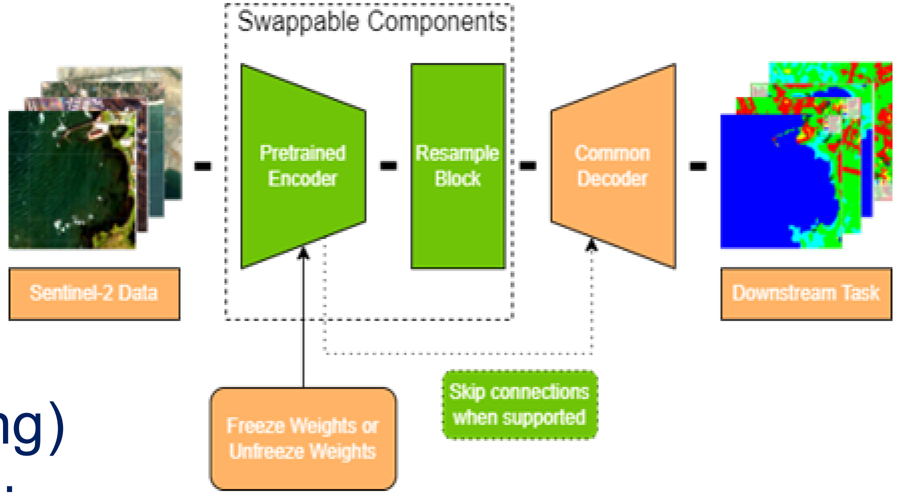
Foundation Model PhilEO (Ctd)

- Land cover classification (Ctd):
  - Patch classification: Labels at the image level
    - Majority class in image



Evaluation framework for EO Foundation Models

- **Problem:** Compare the performance of different models across downstream tasks
  - Tasks:
    A. Building density pixel-wise regression
    B. Road regression segmentation
    C. Land cover semantic segmentation
  - Models:
    A. PhilEO
    B. Prithvi
    C. SatMAE
    D. SeCo
    E. ResNet (supervised learning)
- **Method:** Fine-tune each model using same-architecture decoder per task



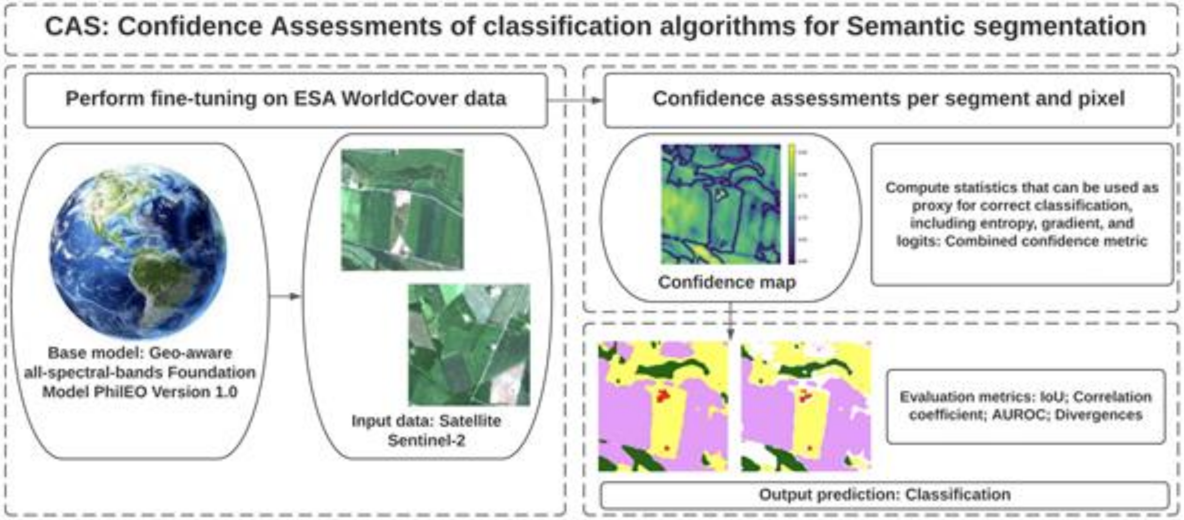
Evaluation framework for EO Foundation Models (Ctd)

- **Method (Ctd):**
  - Decoder: Convolutional, UPerNet
  - Dataset PhilEO Bench: Labelled, global, downstream tasks
    - Regression, Classification
  - Measure sample efficiency of FMs: n-shot fine-tuning
  - Confidence quantification & assessment
    - ESA WorldCover: Noisy labels, accuracy 77%
  - Ongoing work:
    - Add downstream tasks to PhilEO Bench: Sen1Floods11 dataset, HLS Burn Scars
      - e.g., TerraMind model, http://arxiv.org/pdf/2504.11171
      - mIoU 90.78, Sen1Floods11

Techniques: a) Confidence quantification & assessment

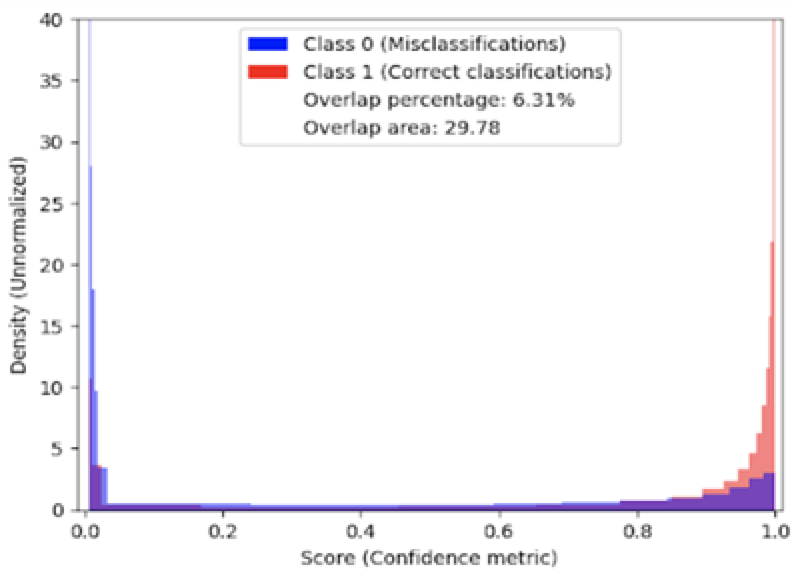
- **Problem:** Confidence quantification for semantic segmentation use cases: Achieve improved performance
  - Confidence assessment: Indicator of performance, *a priori*
  - Low confidence:
    - Further model training
    - Collect more data
  - Anomaly detection/ Out-of-Distribution (OoD), change detection
  - Noisy labels

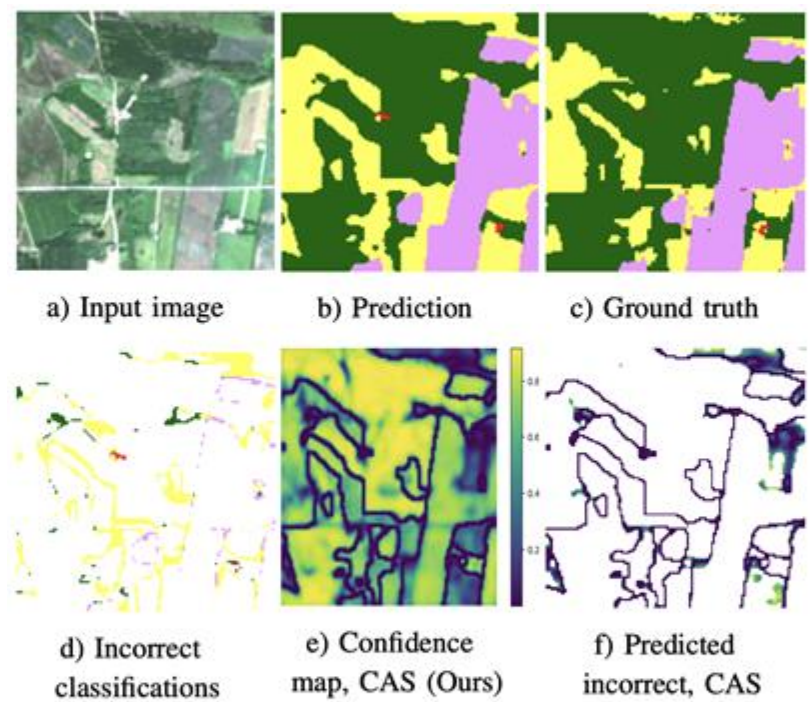
- **Method:** Model Confidence Assessment for Semantic segmentation (CAS)



Techniques: a) Confidence quantification & assessment (Ctd)

- **Method (Ctd):** Features of correct classification:
  - Softmax output probability, Entropy, Diff btwn 1st & 2nd
  - Spread of high probability pixels within segment
  - Gradient of DNN model
- **Results:** Evaluation: Correlation btwn IoU & confidence for segment
  - Confidence score histograms
    - Separability: Correct, incorrect



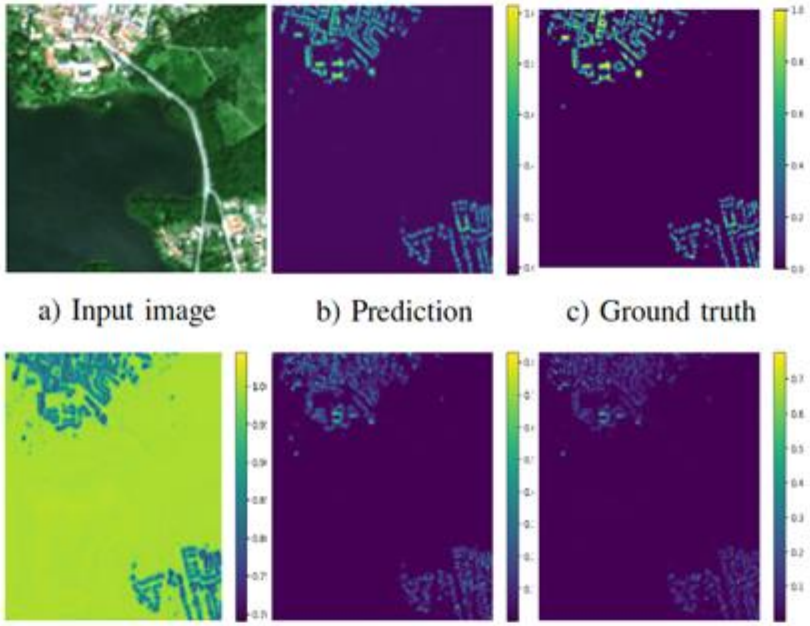


Techniques: b) Confidence quantification & assessment

- **Problem:** Pixel-wise regression downstream tasks: Conf quantification
- **Method:** Model Confidence-Aware Regression Estimation (CARE)
  - Confidence-aware fine-tuning and inference
  - 2 heads: Continuous output (regression), confidence
    - Variation in regression: Modelled by confidence metric
- **Results:** Building density
  - Error btwn confidence metric & error in regression

EVALUATION OF THE PROPOSED MODEL CARE ON THE DATASET PHILEO BENCH BUILDING DENSITY ESTIMATION (SENTINEL-2 DATA, L2A).

$n=$	10000	7500	5000	1000	500	100	50
Error, Mean	0.00683	0.00761	0.00759	0.0138	0.0167	0.0246	0.0261
Error, Med.	0.00150	0.00190	0.00184	0.0065	0.0088	0.0147	0.0159
MSE	0.00301	0.00316	0.00326	0.0039	0.0043	0.0051	0.0034

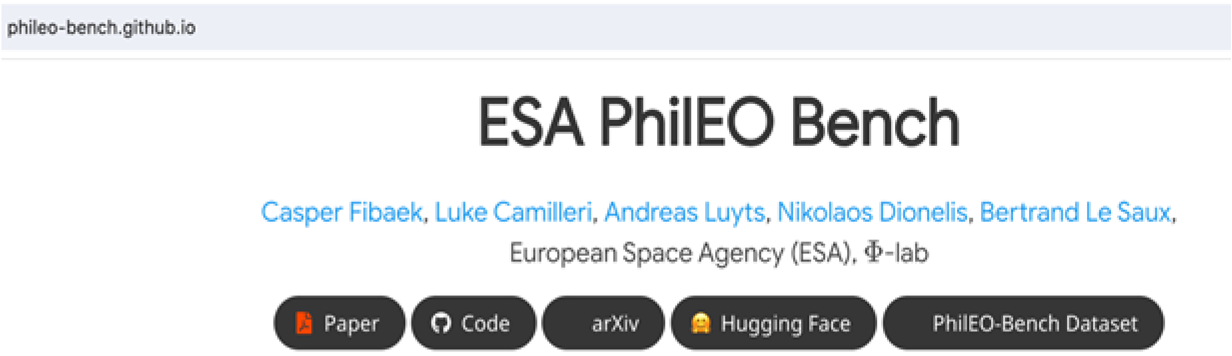


a) Input image b) Prediction c) Ground truth

d) Confidence map, CARE (Ours) e) Abs. error btw prediction & gt f) Abs. error btw pred. uncertainty & (e)

Applications

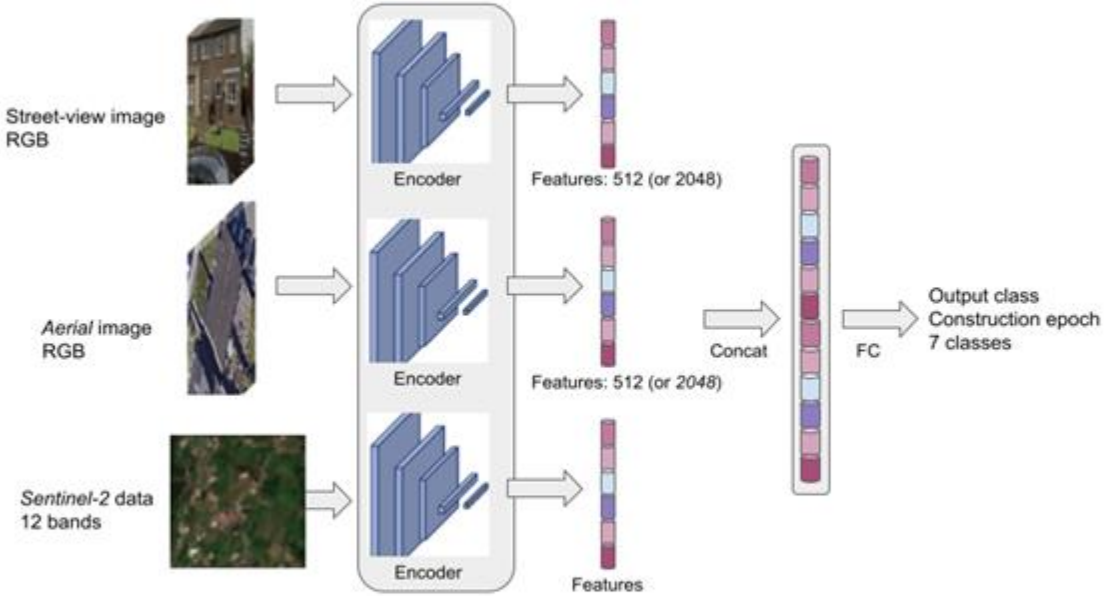
- Downstream tasks:
  - Flood Rapid Mapping
  - Building density regression, road segmentation, land cover mapping
    - Evaluation framework for Sentinel-2 FMs: PhilEO Bench
  - Estimation of construction year of buildings
    - Data fusion, multiple modalities
  - EO datasets:
    - Noisy labels
  - Fine-tuning:
    - Confidence quantification



GitHub: http://github.com/ESA-PhiLab/PhilEO-Bench
Landing page: http://phileo-bench.github.io/

Estimation of age of buildings

- **Problem:** Estimation of construction year of buildings from multi-modal dataset
  - Data fusion: Cross-view dataset, street & satellite VHR
    - Labelled dataset MapYourCity
  - Energy efficiency, age of buildings
- **Method:** Late data fusion: Concatenate features
  - 3 modalities: Inference without street-view

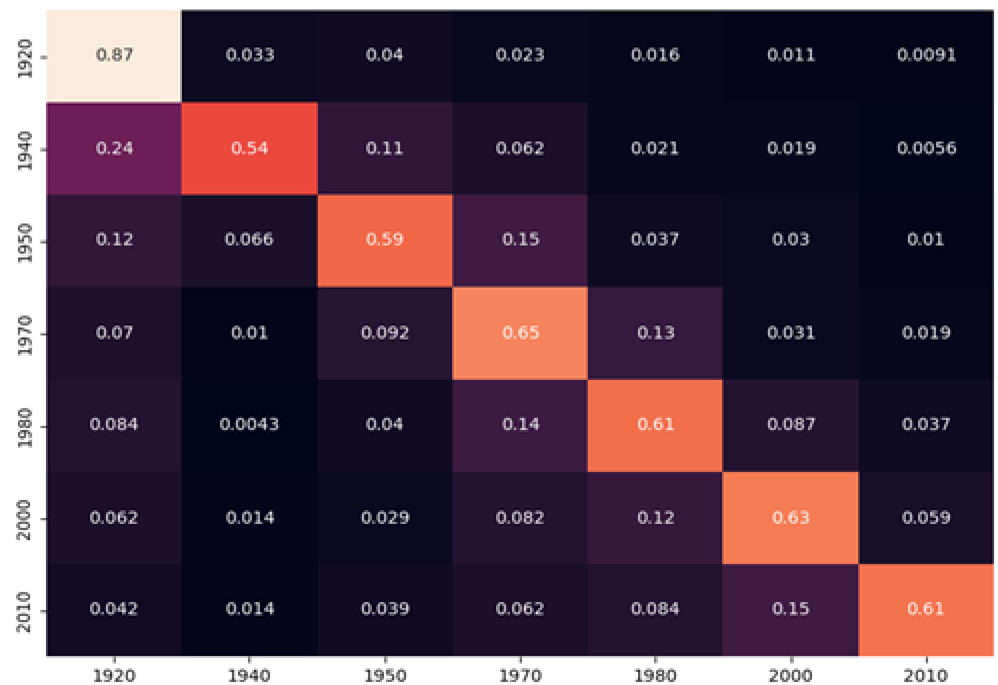


Flowchart of the deep learning based model we use - End-to-end method
*FC = Fully Connected layers

Estimation of age of buildings (Ctd)

- **Results:** Evaluation metric: Classification confusion matrix
    - Diagonals: Mean Producer's Accuracy (MPA)
  - Large scale: Generalization to previously unseen cities
  - 7 classes for construction epoch of buildings: e.g., “Before 1920”
  - Comparison with single modality: Street only, improvement ~7%
- http://github.com/AI4EO/MapYourCity

Model	Performance of model
Accuracy	67.83%
Precision	68.00%
Recall	67.84%
F1-score	67.56%
Mean of diagonals of confusion matrix	64.03%



Main references

- C. Fibaek, L. Camilleri, A. Luyts, N. Dionelis, and B. Le Saux, “PhilEO Bench: Evaluating Geo-Spatial Foundation Models,” in Proceedings International Geoscience and Remote Sensing Symposium (IGARSS), 2024. http://arxiv.org/pdf/2401.04464
- B. Le Saux, C. Fibaek, L. Camilleri, A. Luyts, N. Dionelis, G. Cascarano, L. Bagaglini, and G. Pasquali, “The PhilEO Geospatial Foundation Model Suite,” European Geosciences Union (EGU), 2024. http://meetingorganizer.copernicus.org/EGU24/EGU24-17934.html
- N. Dionelis and N. Longepe, “Fine-Tuning Foundation Models with Confidence Assessment for enhanced Semantic segmentation,” IEEE Geoscience and Remote Sensing Letters, 10.1109/LGRS.2024.3504293, 2024. http://ieeexplore.ieee.org/document/10759697
- N. Dionelis, C. Fibaek, L. Camilleri, A. Luyts, J. Bosmans, and B. Le Saux, “Evaluating and Benchmarking Foundation Models for Earth Observation and Geospatial AI,” W ECCV FOCUS, 2024. http://arxiv.org/pdf/2406.18295
- N. Longepe, et al., “PhilEO Foundation Model using MajorTOM and an Evaluation Framework including Confidence Assessment,” AGU, 2024. http://agu.confex.com/agu/agu24/meetingapp.cgi/Paper/1683553
- N. Dionelis, et al., “PhilEO Foundation Model,” ESA W Digital Twins, Accepted, Oral presentation, 2025
- Nikolaos Dionelis, Francesco Pro, Luca Maiano, Irene Amerini, and Bertrand Le Saux, “Learning from Unlabelled data with Transformers: Domain Adaptation for Semantic Segmentation of High Resolution Aerial Images,” in Proceedings IGARSS, 2024. http://arxiv.org/pdf/2404.11299
- Francesco Pro, Nikolaos Dionelis, Luca Maiano, Bertrand Le Saux, and Irene Amerini, “A semantic segmentation-guided approach for ground-to-aerial image matching,” IGARSS, 2024. http://arxiv.org/pdf/2404.11302