

L'usage de l'IA par les communautés THEIA

27&28 Janvier 2026, Marseille



Biodiversité



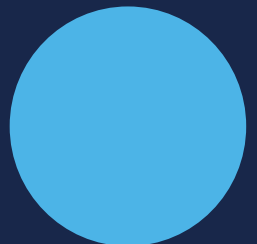
Ocean



Surfaces
continentales



Terre solide



Atmosphere

CONSORTIUMS D'EXPERTISE SCIENTIFIQUE

(1) Des communautés scientifiques structurées autour de 5 domaines d'expertise
= constitué de groupes d'experts thématiques



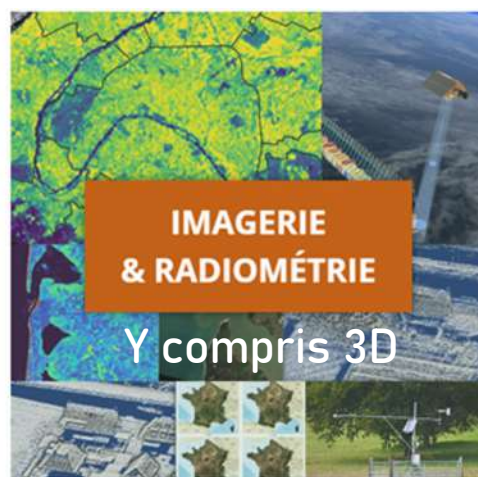
Variables
& Produits



Données in situ



Logiciels & services



1. Optique
2. Radar
3. Thermique
4. 3D
5. Hyperspectral
6.



1. Glacier
2. Neige
3. Permafrost ?
4. Avalanche ?
5.



1. Bilan hydrique
2. Caractéristiques hydrodynamiques
3. Propriétés des eaux continentales
4.



1. Propriétés des sols
2. Propriétés des forêts
3. Propriétés de la vég.
4. Propriétés des agroecosystems
5.



1. Occupation des sols
2. Espaces forestiers
3. Espaces semi-naturels
4. Espaces Urbains
5. Espaces littoraux
6.

+ **Réseaux « métiers »** - GDR spécialisés où les usages de l'IA sont discutés :
- GDR IASIS – Information Apprentissage Signa Image Vision
- GDR MAGIS – Méthodes et applications pour la Géomatique et l'Information Spatiale

CENTRES DE DONNÉES ET DE SERVICES (CDS)

Développer une offre de services aux données afin de répondre aux besoins des communautés (remontés via les AAP THEIA) :



Accès aux données

a) Services de découverte / accès aux données et productions THEIA + support à la fairisation (+ formation)



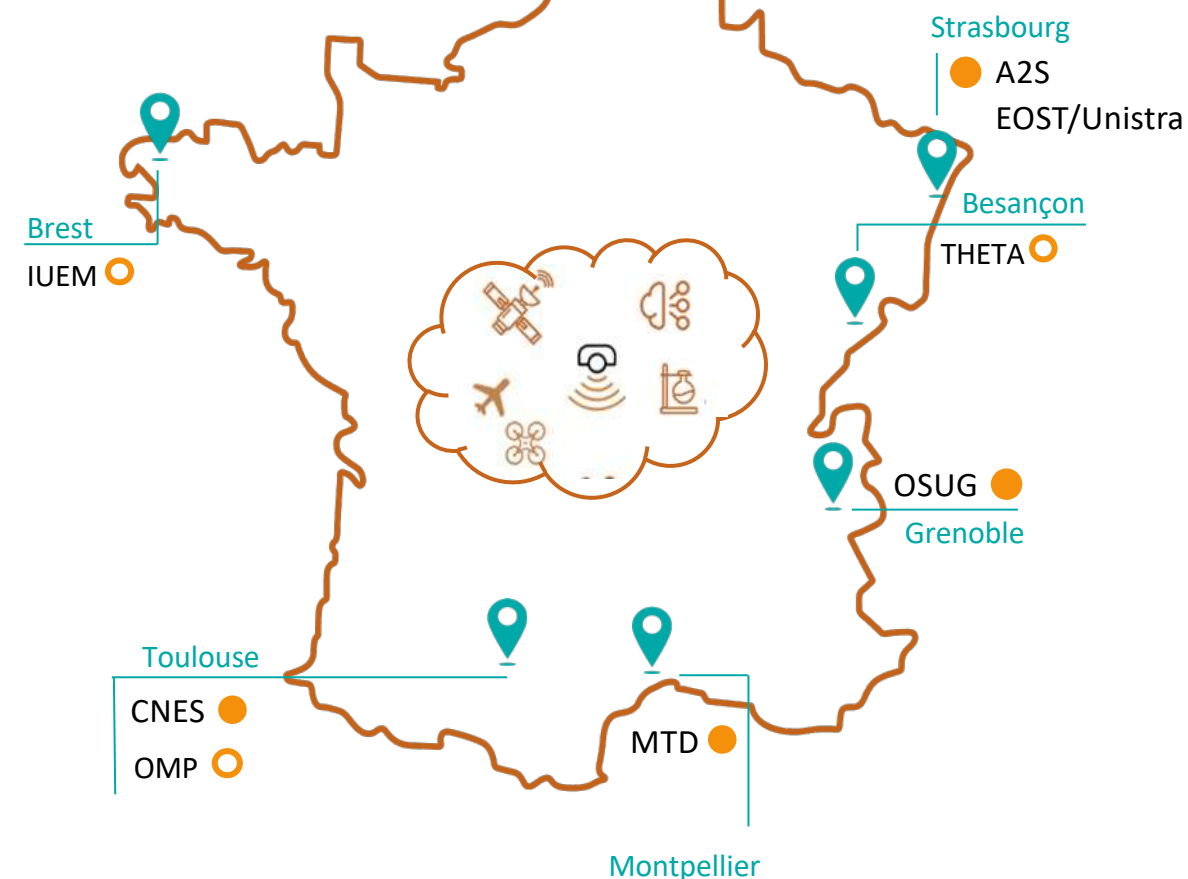
Production régulière de données

b) Services de transformation (production systématique, production à la demande)



Analyse et traitement à la demande

c) Services interactifs d'analyse croisée de données multi-sources et hétérogènes -- *DataLabs*





user non-expert

Production à la demande

via interface Web



user expert

Production systématique produits ARD

via webGIS ou API



Produits corrigés Sentinel2 :
- Réflectance L2A
- Synthèse mensuelle L3A

user expert

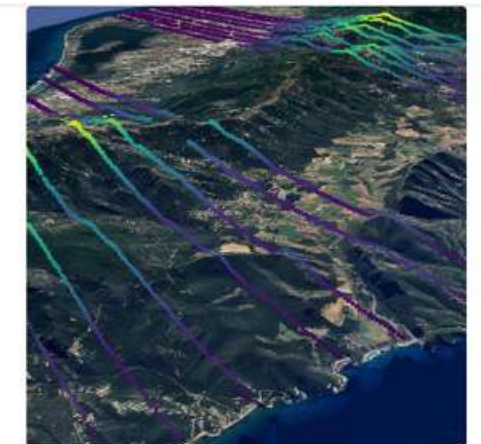
Accès à des produits thématiques

via QGIS (mode stream / via API STAC)



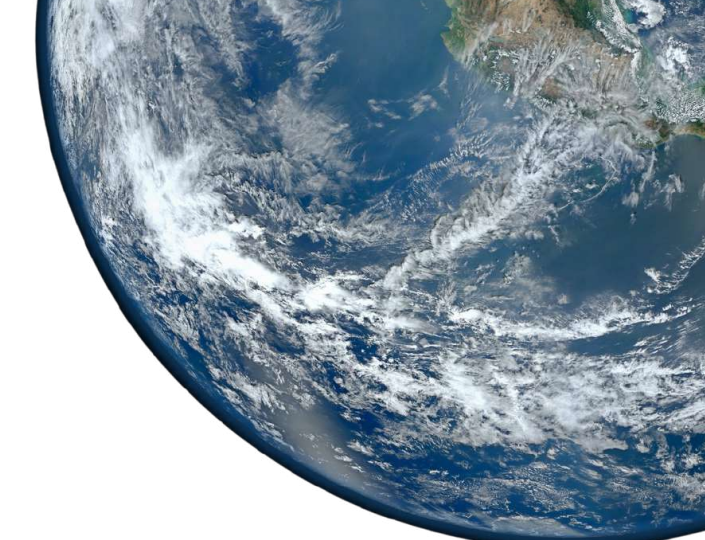
FORMS-T

Cartes de Hauteur, Biomasse, Volume des arbres



GeoGEDI L2A

Données 3D

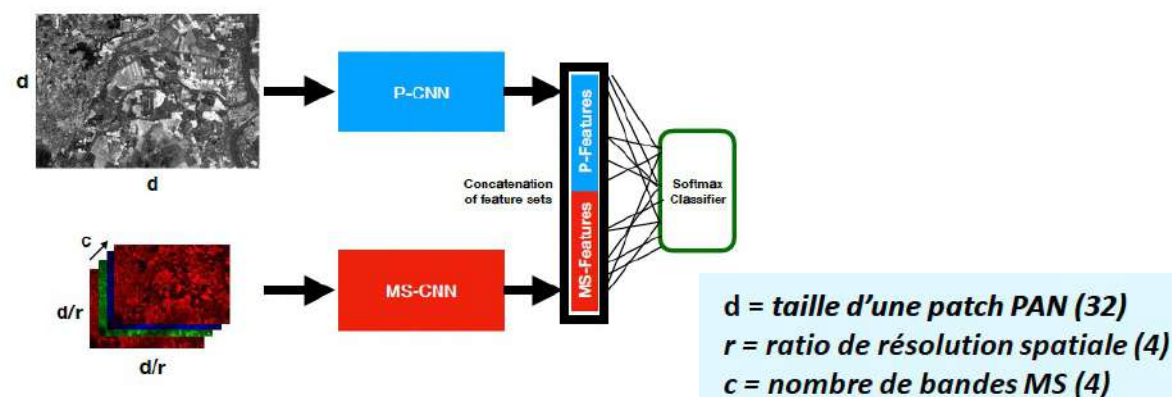


Depuis 2018 :

A two-branch Deep Learning architecture for land cover classification of PAN and MS imagery

R. Gaetano³, D. Ienco², K. Ose⁴, R. Cresson⁴

Figure 1. General Overview of MultiResoLCC. The model is based on a two-branch CNN architecture to deal with PAN and MS information sources at their native resolution.



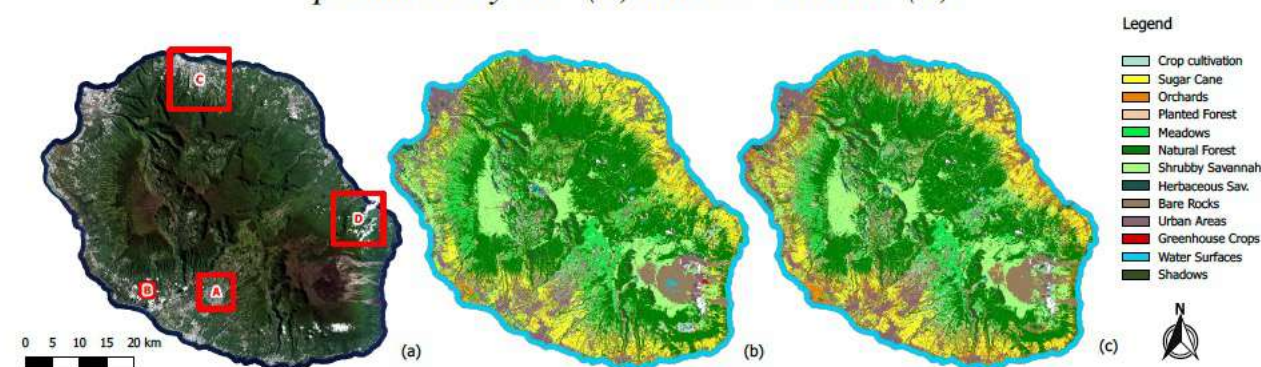
- ❖ Architecture « **end-to-end** », apprentissage « **from scratch** »
- ❖ Un **module CNN dédié** pour chaque source, avec **concaténation des features** extraites (**late fusion**)
- ❖ La **classification** est produite à la **résolution de la PAN (1.5m)**

[4] R. Gaetano, D. Ienco, K. Ose, R. Cresson: "A Two-Branch CNN Architecture for Land Cover Classification of PAN and MS Imagery". Remote Sensing 10(11): 1746 (2018)

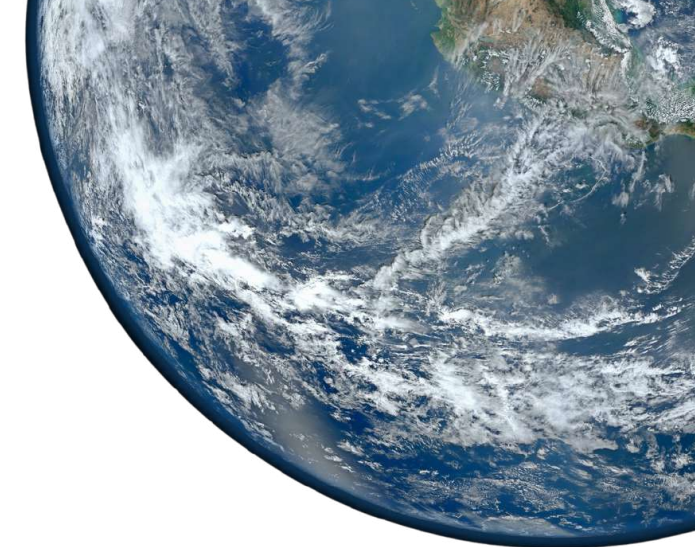
$M^3Fusion$: A Deep Learning Architecture for Multi-{Scale/Modal/Temporal} satellite data fusion

P. Benedetti¹, D. Ienco², R. Gaetano³, K. Ose⁴, R. Pensa⁵, S. Dupuy⁶

FIGURE 4. Source VHRS scene (a) (see Fig. 3 for details in red boxes), maps produced by RF (a) and $M^3Fusion$ (b).

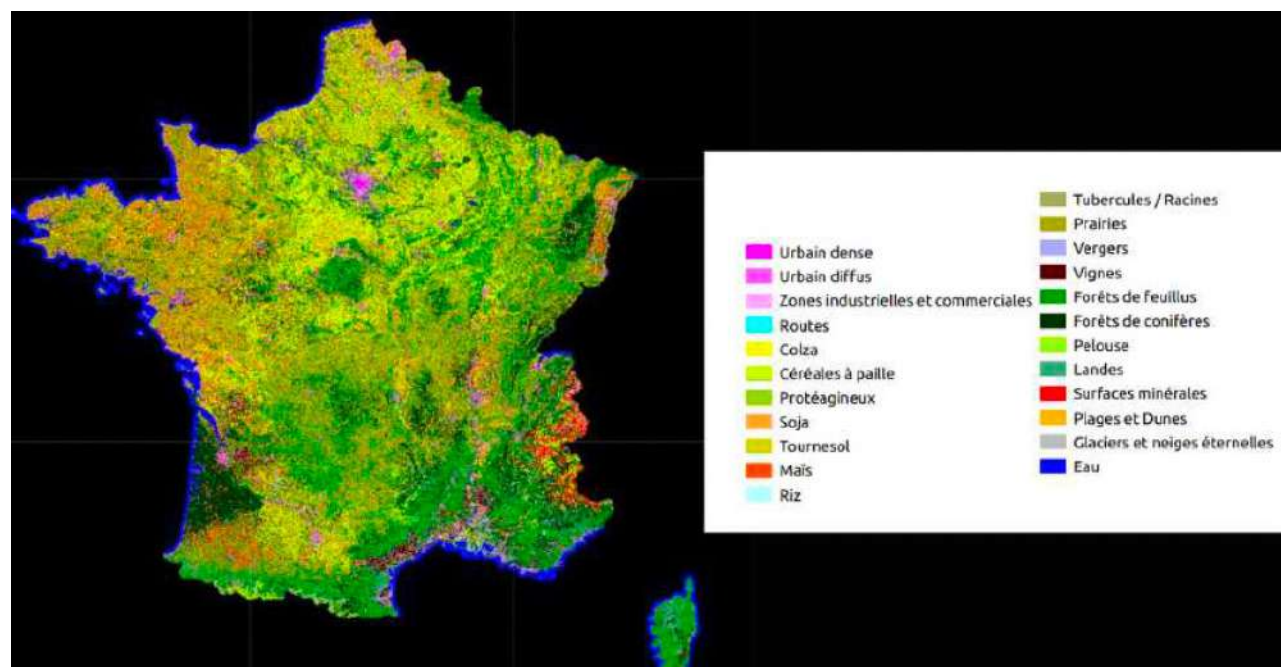


Classification d'occupation des sols



Plus récemment – depuis 2022 :

CDS THEIA-CNES – produit Occupation des sols
Chaine de traitement IOTA2



→ Iota2 latest release – Deep Learning

Plus récemment :

IGN – produit de prédiction du sol
CoSIA (Couverture du Sol par IA)



selon 15 classes :

- Bâtiment
- Zone imperméable, Zone perméable
- Piscine, Surface eau, Neige
- Serre, Sol nu
- Conifère, Feuillu, Broussaille, Pelouse, Culture, Terre labourée, Vigne

Multimodal and Multitemporal Land Use/Land Cover Semantic Segmentation on Sentinel-1 and Sentinel-2 Imagery: An Application on a MultiSenGE Dataset

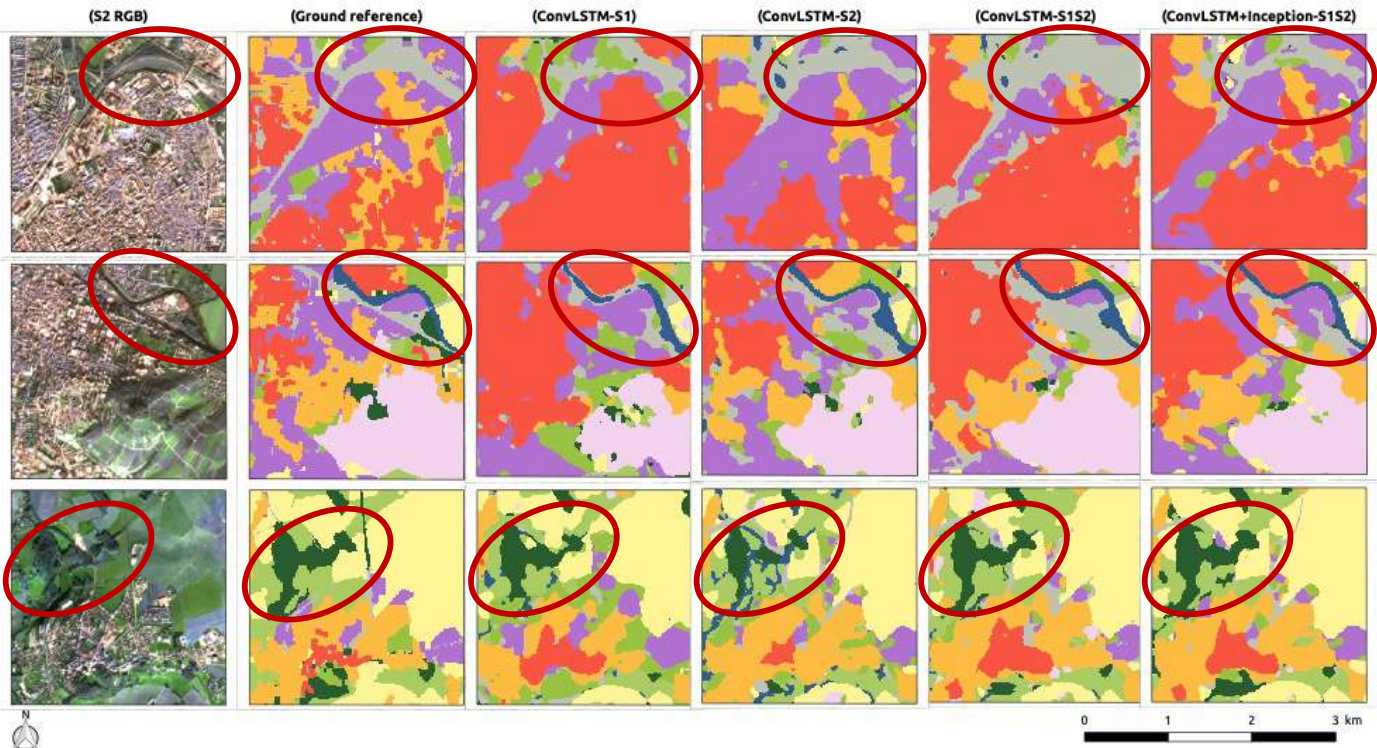
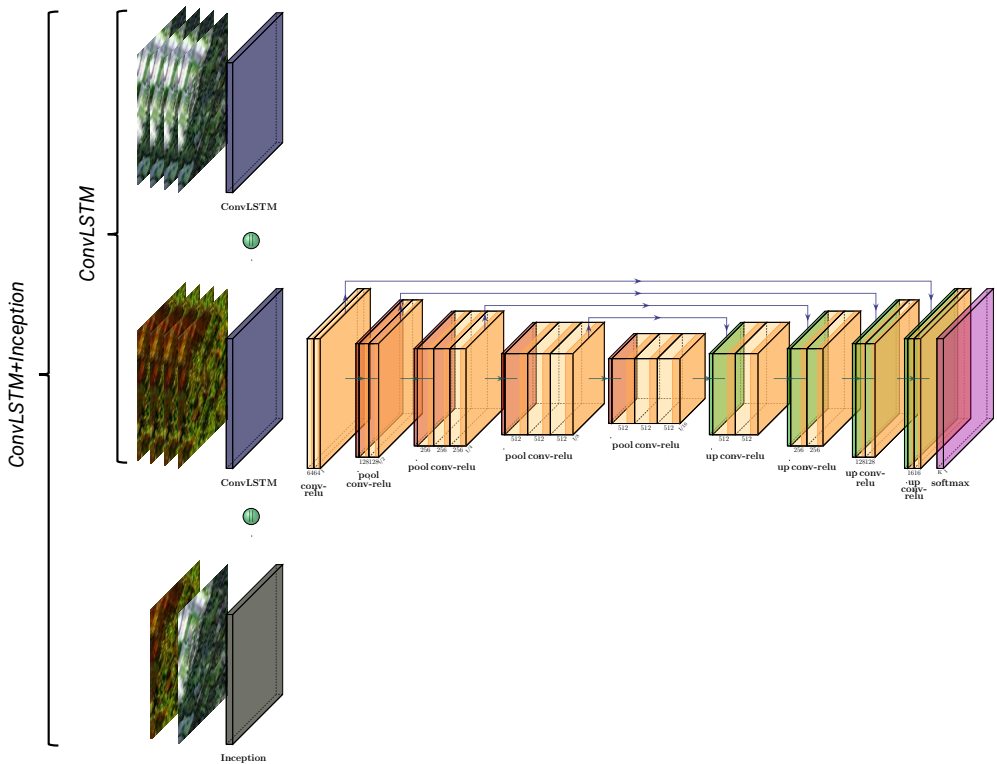
Romain Wenger^{1,*}, Anne Puissant¹, Jonathan Weber², Lhassane Idoumghar² and Germain Forestier²

¹ LIVE UMR 7362 CNRS, University of Strasbourg, F-67000 Strasbourg, France

² IRIMAS UR 7499, University of Haute-Alsace, F-68100 Mulhouse, France

* Author to whom correspondence should be addressed.

□ Résultats qualitatifs – 10 classes



Légende

Dense Built-Up (1)
Sparse Built-Up (2)
Specialized Built-Up Areas (3)
Specialized Vegetative Areas (4)
Large Scale Networks (5)
Arable Lands (6)
Vineyards and Orchards (7)
Grasslands (8)
Forests and semi-natural areas (9)
Water Surfaces (10)

Segmentation sémantique

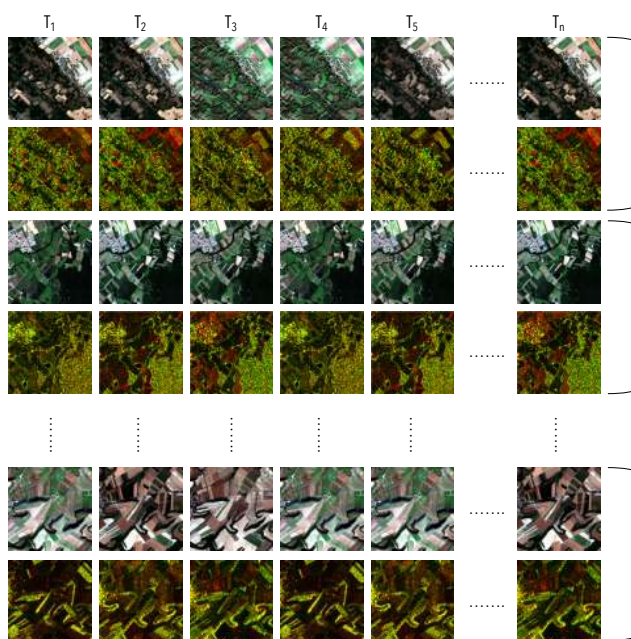
Besoin de jeux de données IA

→ Comme jeux de données THEIA

Produit final

- 8157 patches multitemporels et multimodaux de 256x256 pixels
 - ✓ Toutes les acquisitions disponibles pour Sentinel-1
 - ✓ Toutes les images disponibles sur la plateforme Theia en fonction de la couverture nuageuse
- Pour chaque patch → fichier GeoJSON :
 - ✓ Classes dans le patch
 - ✓ Projection UTM/WGS84
 - ✓ Patches Sentinel-2 associés
 - ✓ Patches Sentinel-1 associés

Segmentation sémantique et classification
(*scene classification*)



Légende :
 T_n : profondeur temporelle du patch
 P_n : nombre de patches (série temporelle Sentinel-1 et Sentinel-2, donnée de référence, GeoJSON)

TYPES DE DONNÉES

Données satellitaires

Jeux de données pour l'IA

Données d'expérimentation

Données in situ

Données de longue traîne

Données aéroportées

Logiciels & services



Laboratoire image, ville, environnement | LIVE

IRIMAS HAUTE-ALSACE Institut de Recherche en Informatique, Mathématiques, Automatique et Signal

DATA TERRA

Theia

Université de Strasbourg

CNRS INSU

anr

Identifiant DOI 10.25577/563Q-QD29

DataCite Metadata HTML XML

License CC BY NC

How to cite :

LIVE & IRIMAS & THEIA | Data Terra. (2026): Artificial Intelligence benchmark datasets for Land Cover Classification from Satellite Imagery. EOST. (Collection)

doi:10.25577/563Q-QD29

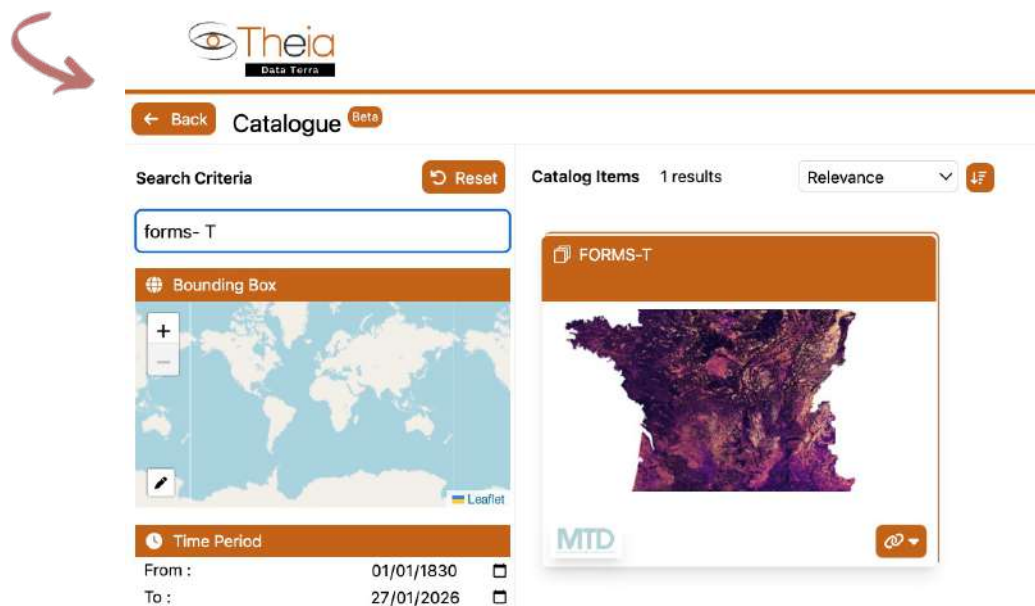
RIS Citation BibTeX Citation Copy

Tâche 2.1 : Forest mapping using deep learning

Task lead: Philippe Ciaï

Presentation: Martin Schwartz

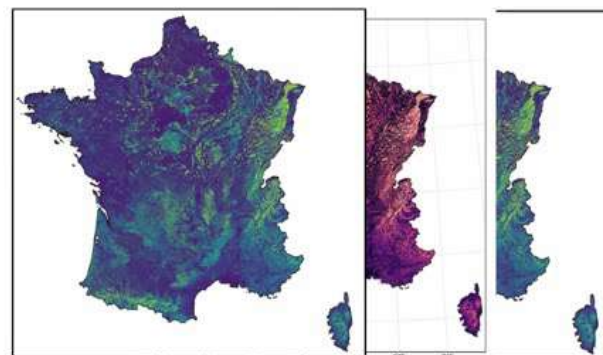
Team : Ibrahim Fayad, Loïc Landrieux, Fajwel Fogel, Cedric Vega, Nikola Besic, Thomas Boudras, Alvin Opler



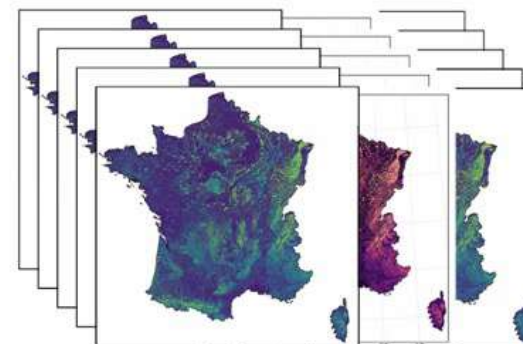
→ Collection STAC – CDS THEIA-MTD
+ ressources sur gitlab

FORMS

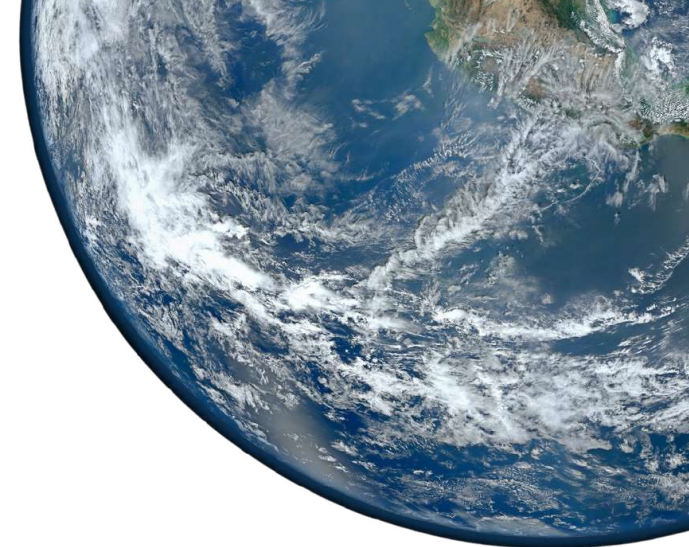
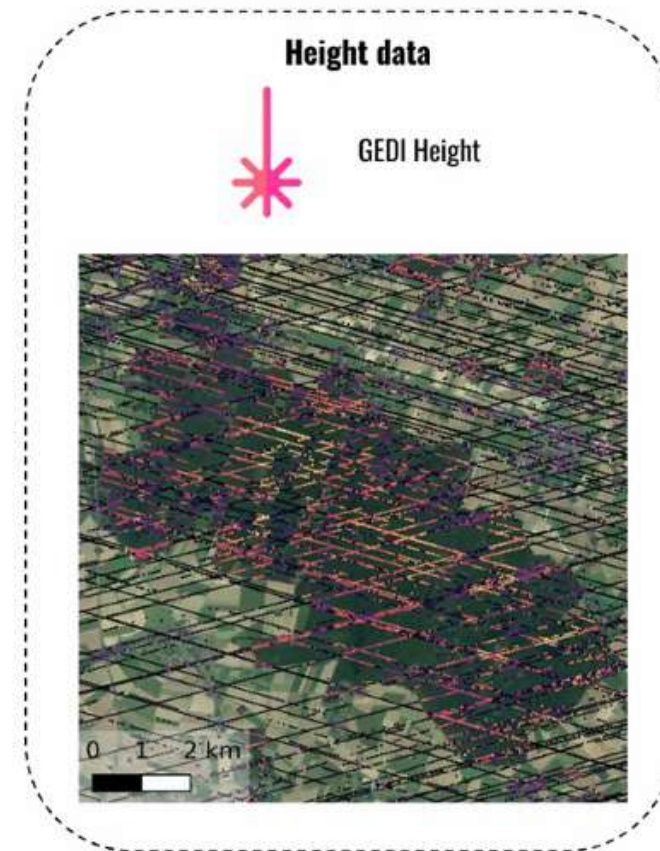
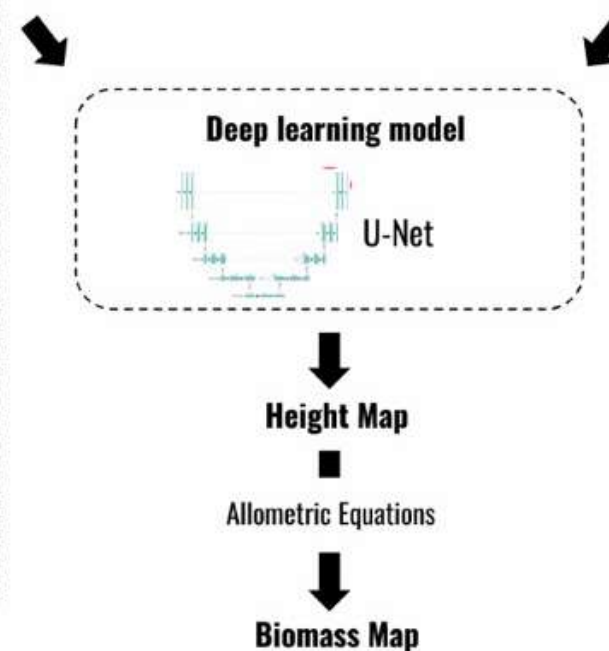
FORest Multiple Satellite
Martin Schwartz



FORMS-H / FORMS-B / FORMS-V
Height, Biomass, Wood Volume for 2020
(Schwartz et al., 2023 ESSD)



FORMS-T (timeseries)
Height, Biomass, Wood Volume for 2018-2024
(Schwartz et al., 2025 RSE)



Open-Canopy project

Martin Schwartz, Fajwel Fogel

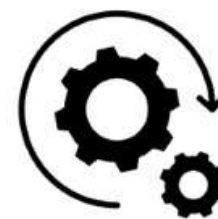
PEPR FORESTT –
Programme OFVi



SPOT

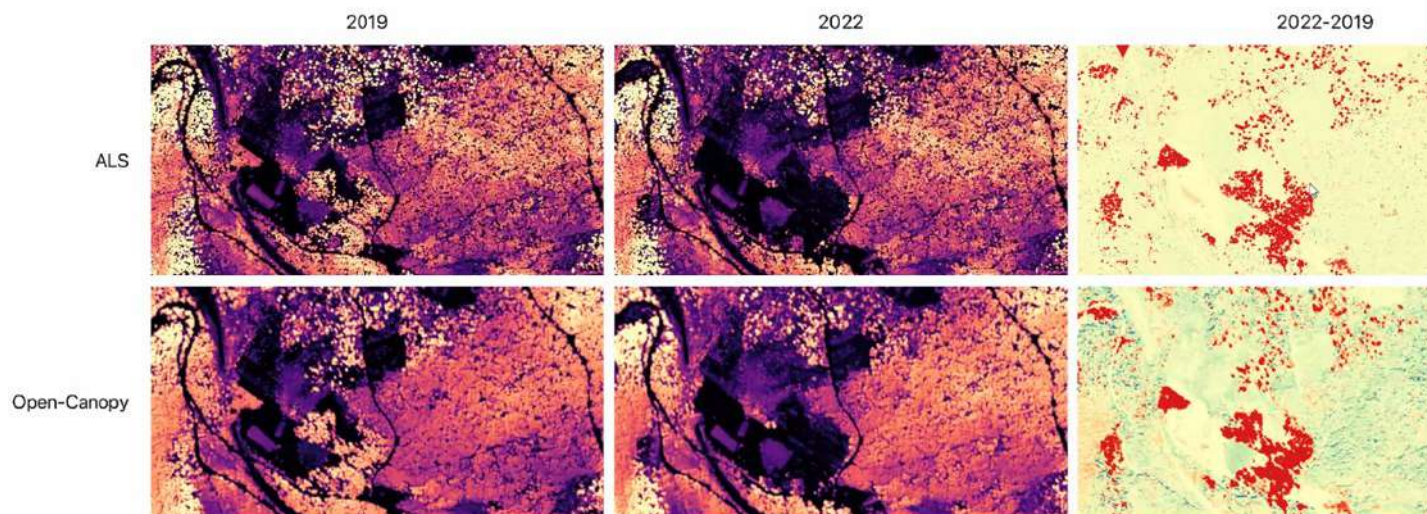


Lidar HD



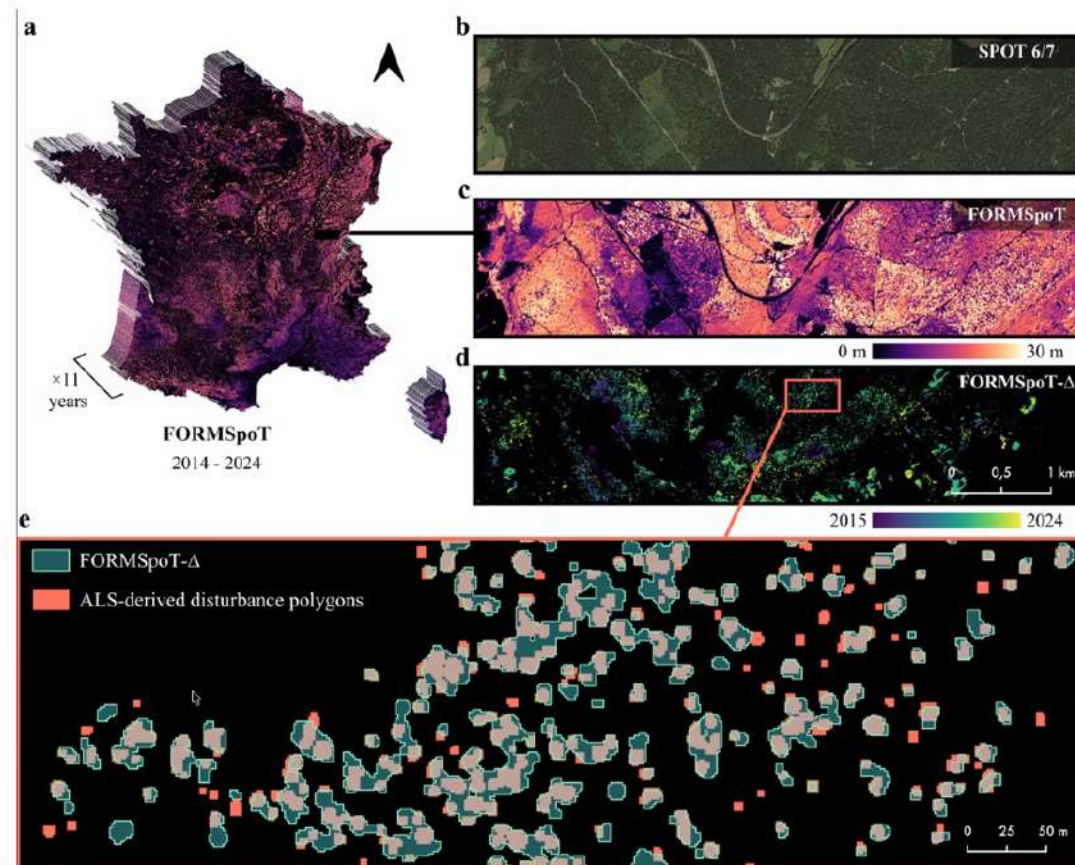
Vision transformer model

Open-Canopy disturbances detection

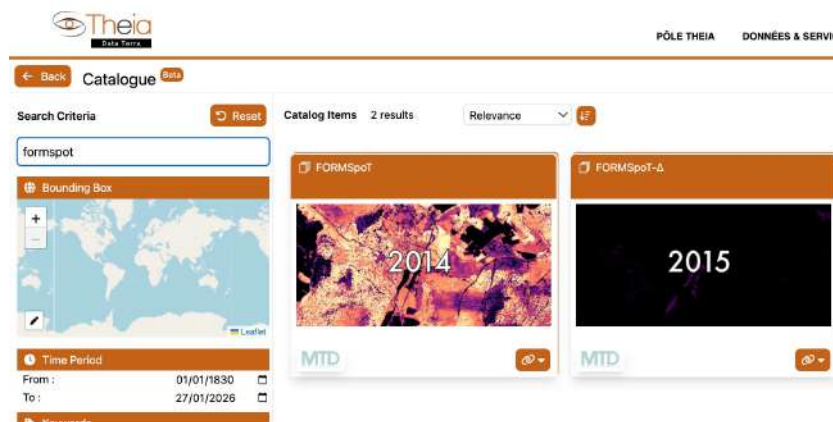


Cartes FORMSpOT

- Forest Mapping with Spot Time Series (Anciennement "opencanopy")
- Article soumis hier
- Cartes disponibles sur le catalogue Théia :
- <https://doi.theia.data-terra.org/FORMSpOT/>



→ Deux collection STAC – CDS THEIA-MTD
FORMSpOT et FORMSpOT-delta



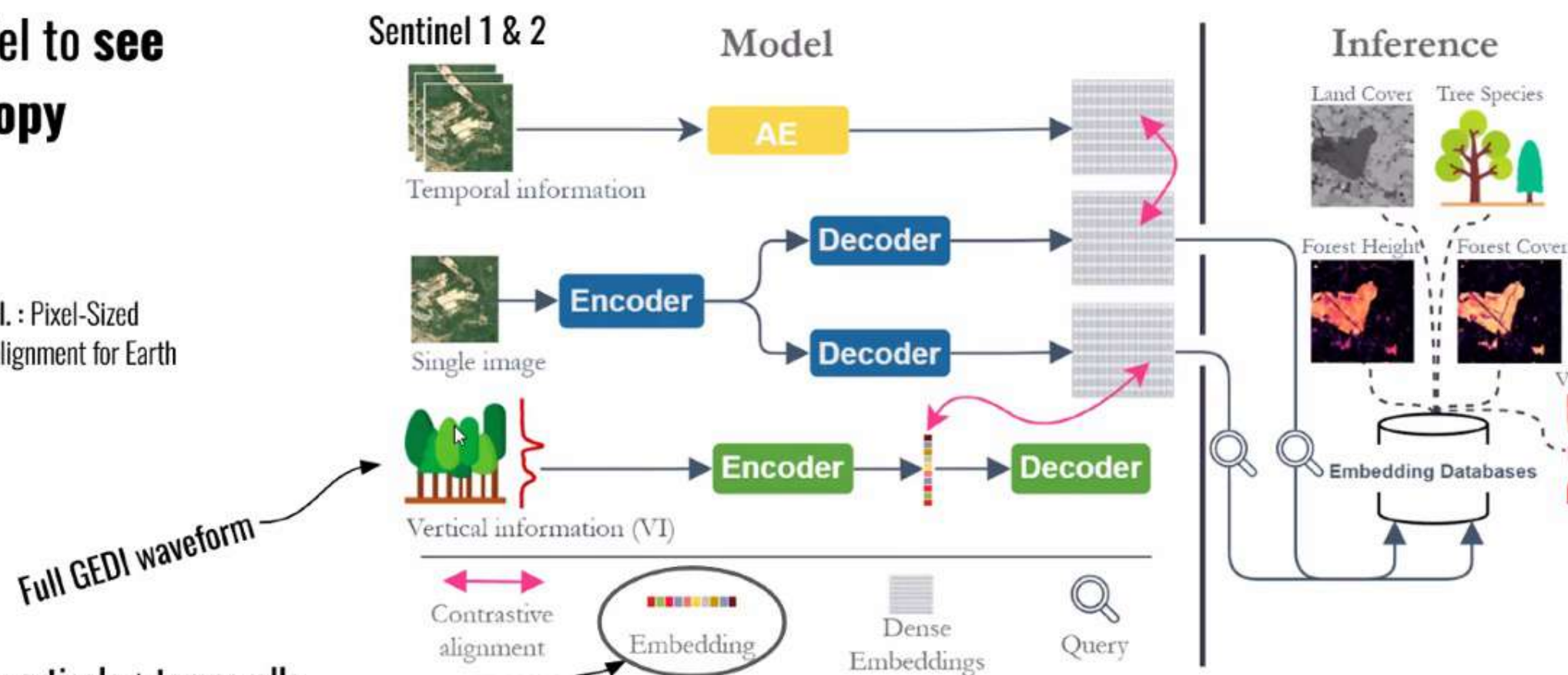
DUNIA

Dense Unsupervised Nature Interpretation Algorithm

Ibrahim Fayad

- Foundation model to **see below the canopy**

- Papier ICML 2025 Fayad et al. : Pixel-Sized Embeddings via Cross-Modal Alignment for Earth Observation Applications



Information sur la structure verticale + temporelle

Cartographie des essences en milieu urbain



Test 1 - Random Forest approach

IMPROVING URBAN TREE SPECIES CLASSIFICATION WITH HIGH RESOLUTION SATELLITE IMAGERY AND MACHINE LEARNING

Romain Wenger¹, Clément Bressant¹, Lucie Roettelé¹, Germain Forestier², Anne Puissant¹

¹ LIVE UMR 7362 CNRS, University of Strasbourg, F-67000 Strasbourg

² IRIMAS UR 7499, University of Haute-Alsace, F-68100 Mulhouse



22 dates - 10 bands



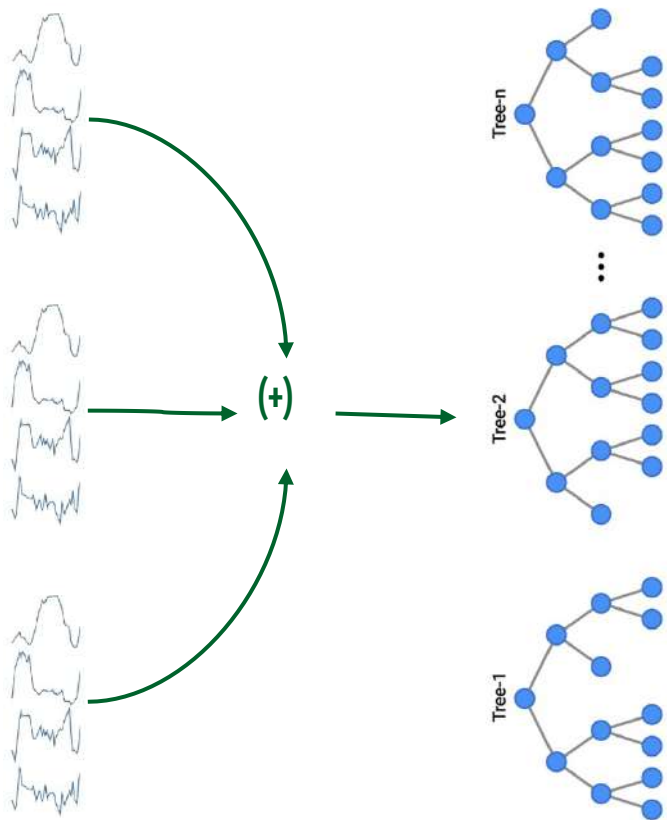
53 dates - 4 bands



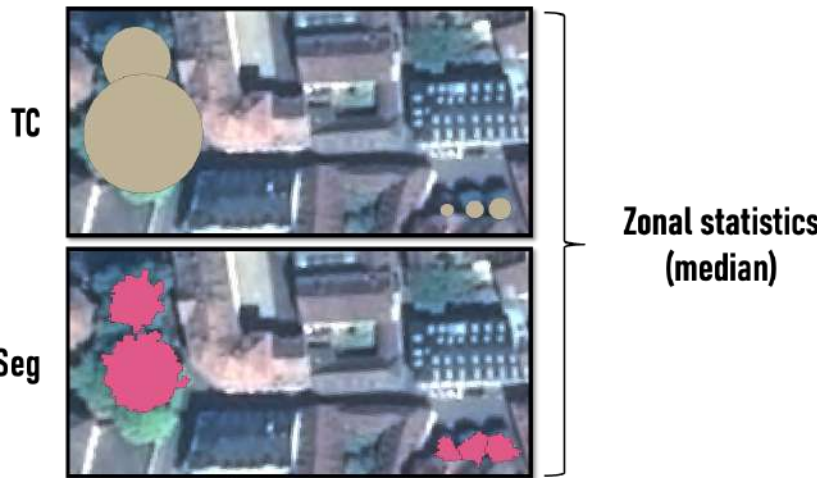
4 dates - 4 bands



Reference Data



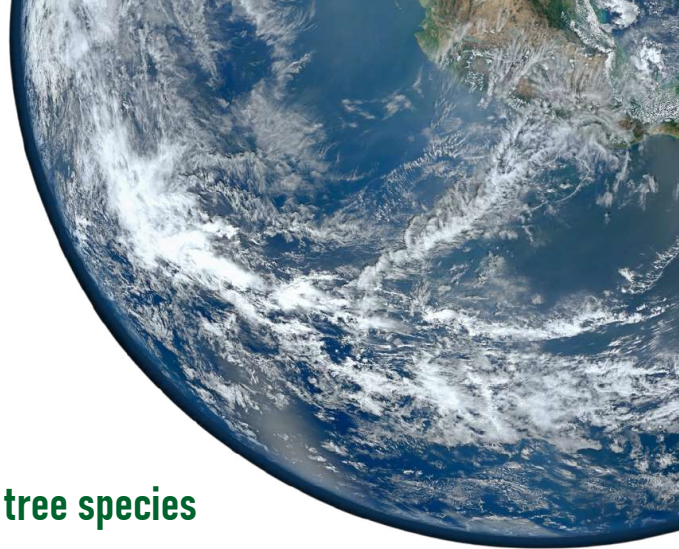
Two approaches tested: **tree crown (TC) buffer** and **watershed segmentation (Seg)**



Scenario	Sensors Used			NDVI/DSM	
	S2	Pléiades	Planet	NDVI	DSM
1 (Seg1, TC1)	✓	✓	✓	✓	✓
2 (Seg2, TC2)	✓	✓		✓	✓
3 (Seg3, TC3)		✓	✓	✓	✓
4 (Seg4, TC4)		✓		✓	✓
5 (Seg5, TC5)	✓			✓	
6 (Seg6, TC6)			✓	✓	
7 (Seg7, TC7)	✓	✓	✓		
8 (Seg8, TC8)	✓		✓		

- ✓ Machine learning and freely available SITS give interesting results to map urban tree species (i.e. 70% F1-Score)
- ✓ Pleiades imagery (VHSR) does **not improve** that much the classification
- ✓ NDVI and DSM does **not improve** the results
- ✓ Better to use temporal satellite imagery (i.e. trees phenology)
- ⇒ Can we use **deep learning** with **time series** to improve these classifications ?

Cartographie des essences en milieu urbain

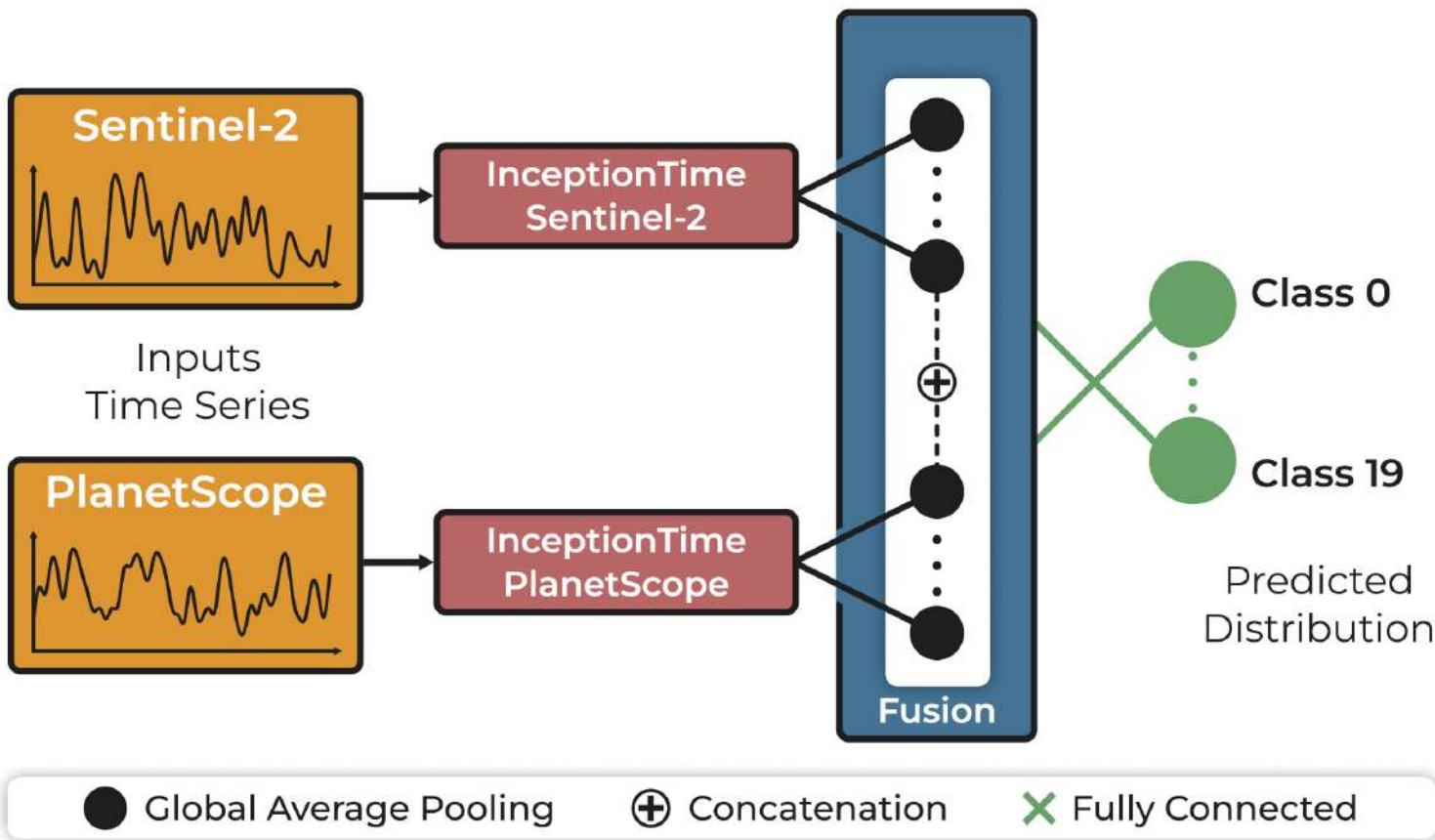


Test 2 - SITS analysis with deep learning

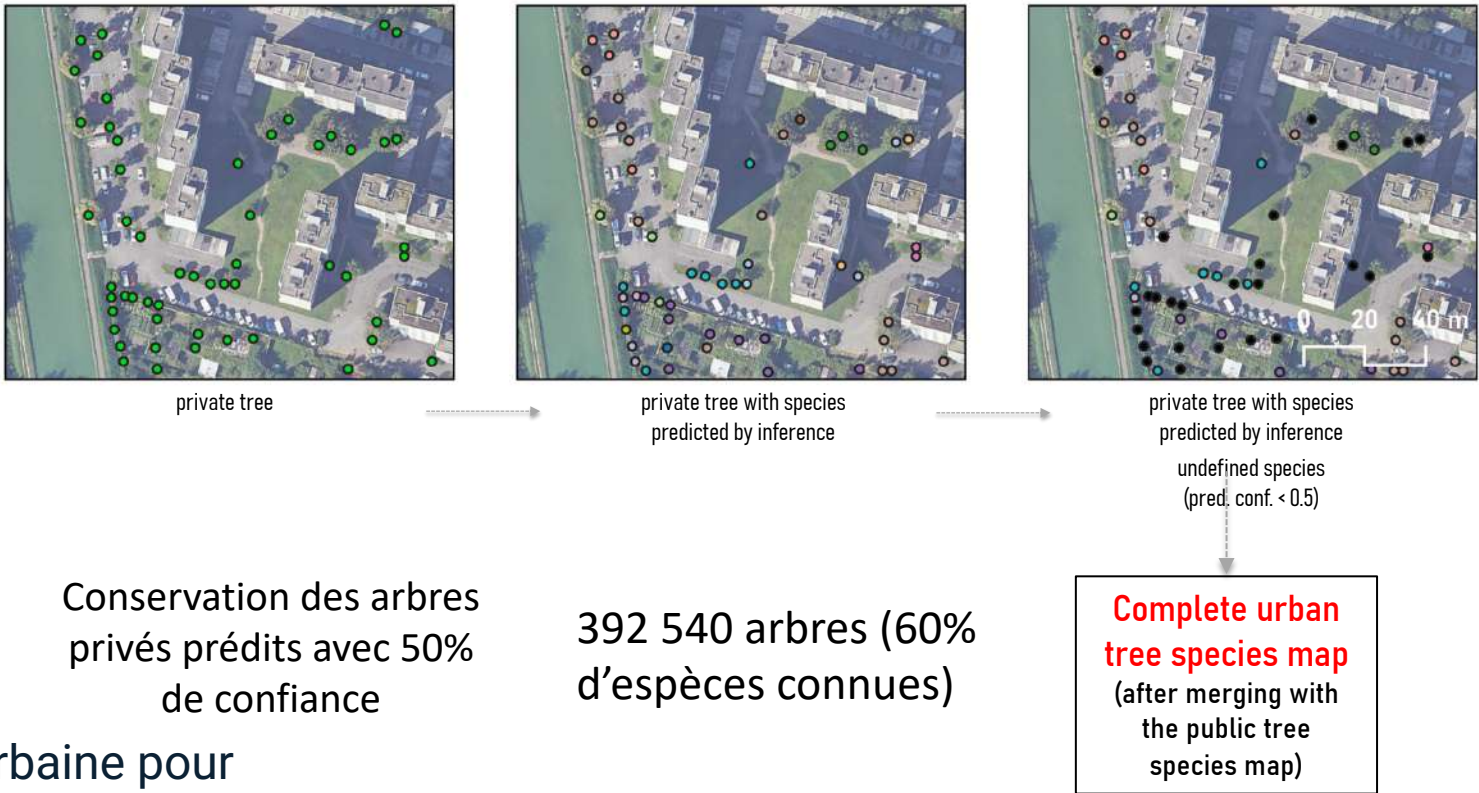
Run encoding	S2 SITS	PS SITS	Number of species	Accuracy*
Dual-InceptionTime	✓	✓	20	0.656±0.005
InceptionTime-S2	✓		20	0.603±0.003
InceptionTime-PS		✓	20	0.615±0.004

* average value of 5-fold processes.

- ⇒ This time note for the 10 but the **20 most representative tree species**
- ✓ Best results for **H-InceptionTime**
- ✓ Low results for LITE, not adapted for this task
- ✓ Almost every species F1-Score are higher for H-InceptionTime



Test 3 – Inférence sur le domaine privé



Cartographie des essences en milieu urbain

Valorisation des jeux de données



Data in Brief
Volume 61, August 2025, 111777

Data Article

Urban tree species benchmark dataset for time series classification


Clément Bressant ^a, Romain Wenger ^a  , David Michéa ^b, Anne Puissant ^{a c}

Urban tree species benchmark dataset for time series classification

Citation / Citation html text ris bibtex 

“BRESSANT Clément (Laboratoire Image, Ville, Environnement (Strasbourg - France)), WENGER Romain (Laboratoire Image, Ville, Environnement (Strasbourg - France)), PUISSANT Anne (Laboratoire Image, Ville, Environnement (Strasbourg - France)), MICHEA David (École & Observatoire des Sciences de la Terre (Strasbourg - France)) (2025).Urban tree species benchmark dataset for time series classification.
<https://doi.org/10.57932/c17e2153-c687-4c28-824c-2be148fb9e4d>

DOI: <https://doi.org/10.57932/c17e2153-c687-4c28-824c-2be148fb9e4d> Télécharger le dataset(zip)



PÔLE THEIA DONNÉES & SERVICES APPLICATIONS RESSOURCES ACTU

← Back Catalogue Beta

Collection

Search Criteria Reset

Enter keyword

Bounding Box

+

−

Map

Time Period

From : 01/01/2022 📅

To : 27/01/2026 📅

Keywords

Sources

Platforms

Instruments

Urban Tree Species Benchmark Dataset For Time Series Classification


EasyData

We propose a benchmark dataset for urban tree species classification based on multi-source optical Satellite Image Time Series. The dataset provides, on the city of Strasbourg (France), surface reflectance values extracted from co-registered Sentinel-2 and PlanetScope imagery on public trees.

Show more

Catalog Items 1 results Relevance 📄

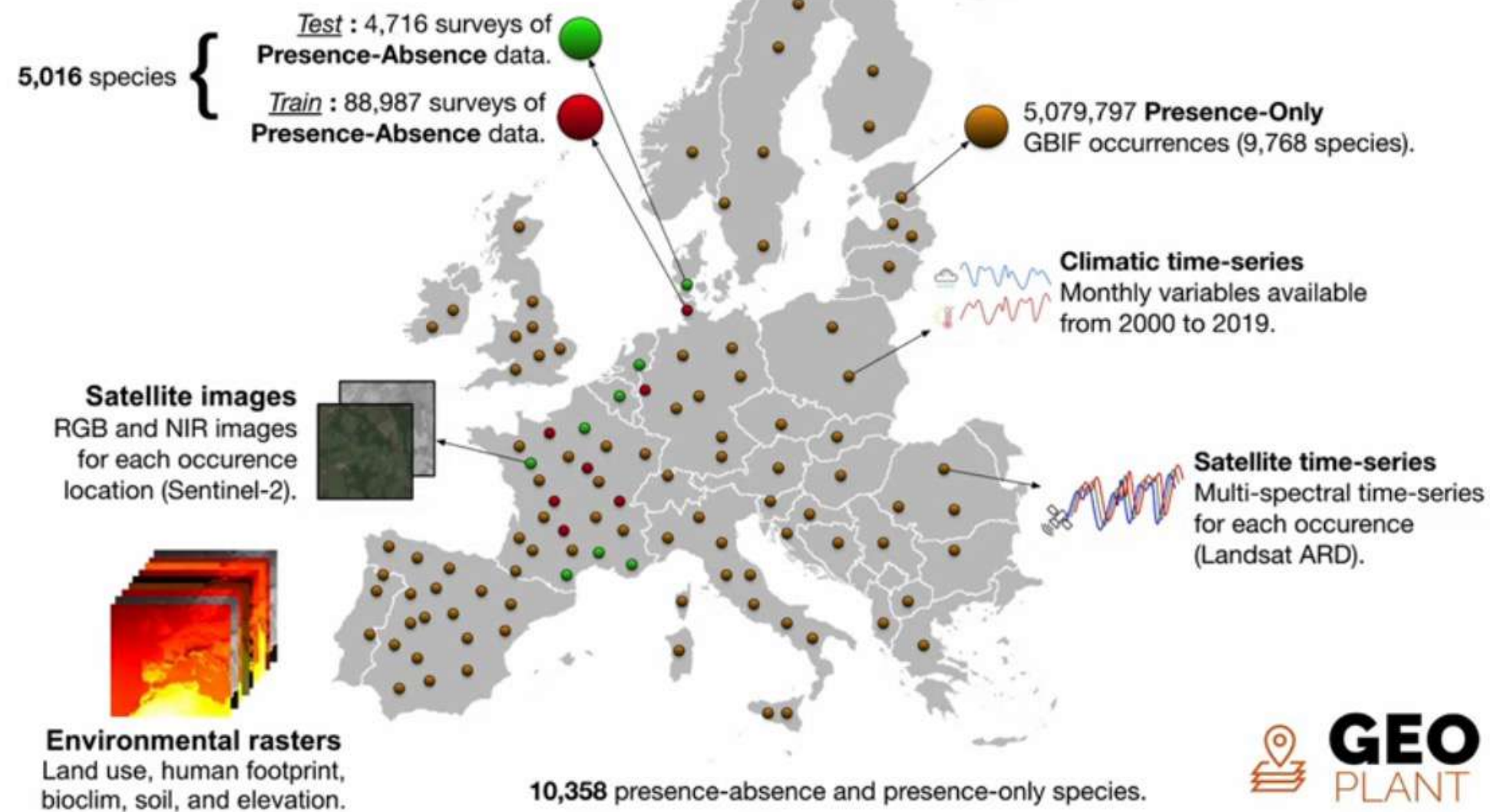
Urban tree species benchmark dataset for time series classification



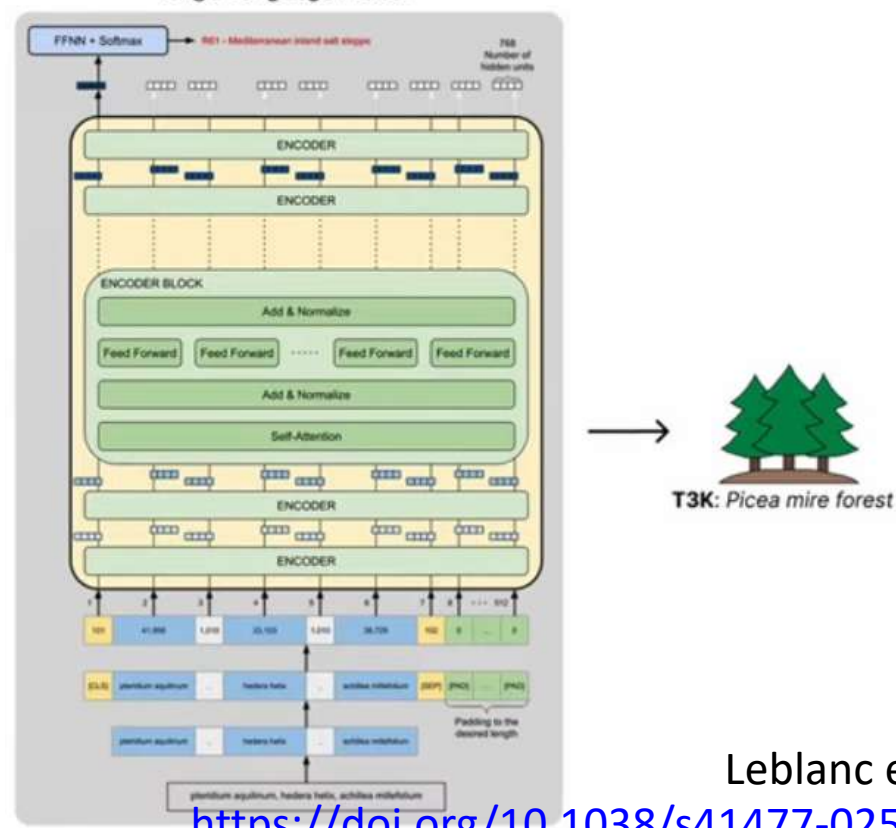
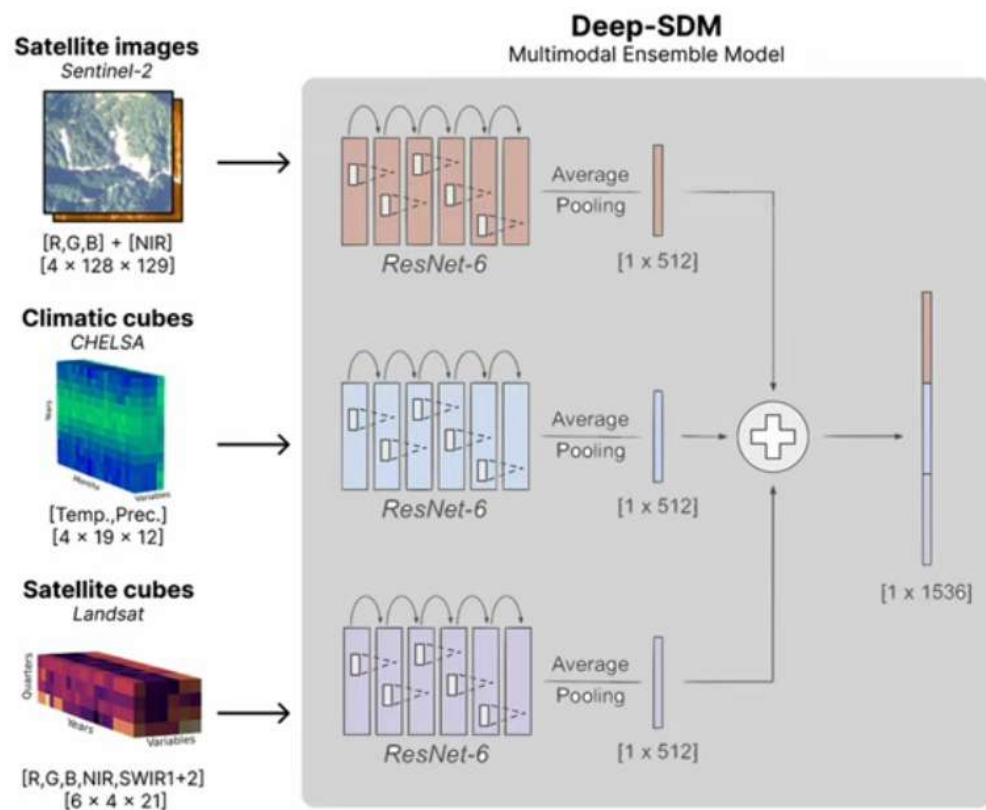
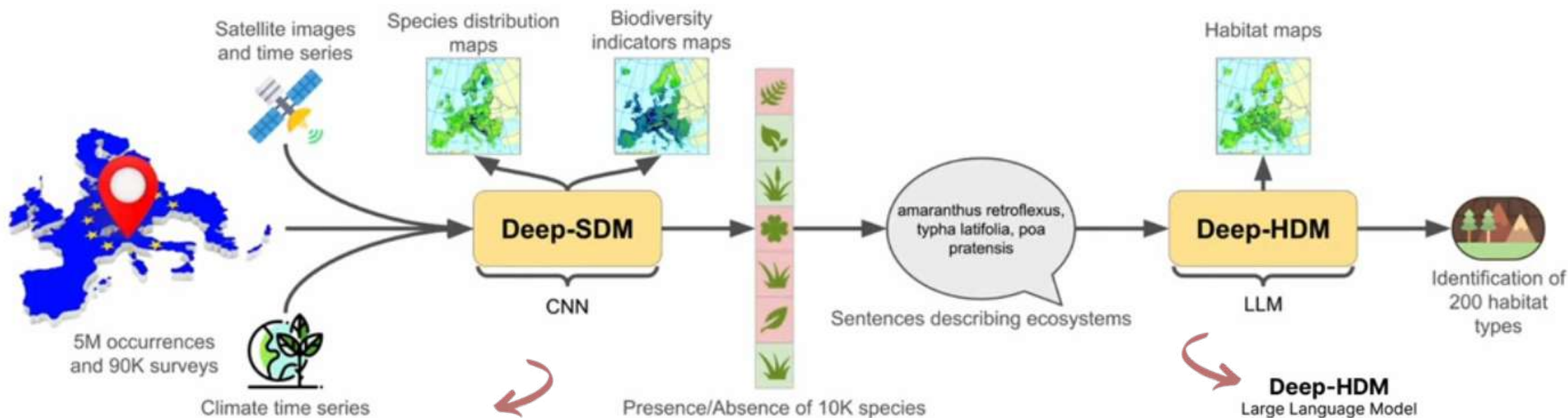
📄 📄

A deep-learning framework for enhancing habitat identification based on species composition

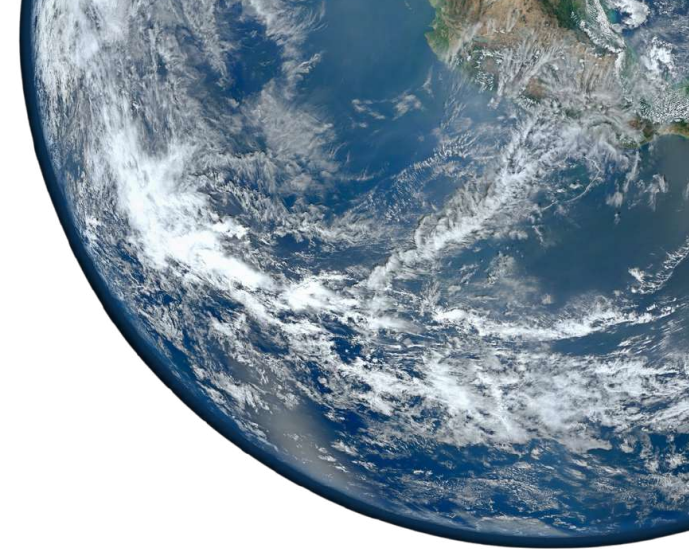
Methodology – Dataset



A deep-learning framework for enhancing habitat identification based on species composition



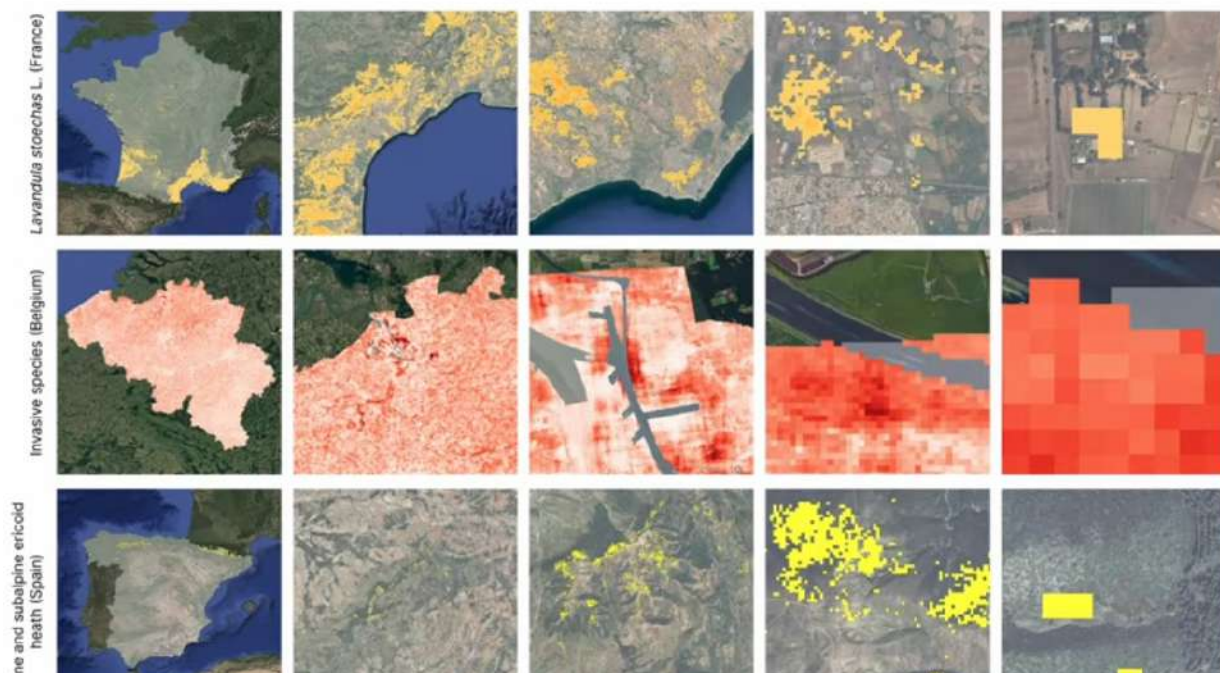
A deep-learning framework for enhancing habitat identification based on species composition



GeoPl@ntNet est développé et maintenu par l'équipe Pl@ntNet dans le cadre d'un accord de consortium entre le CIRAD, l'INRAE, l'INRIA et l'IRD.



Example Maps



10,000+ plant species

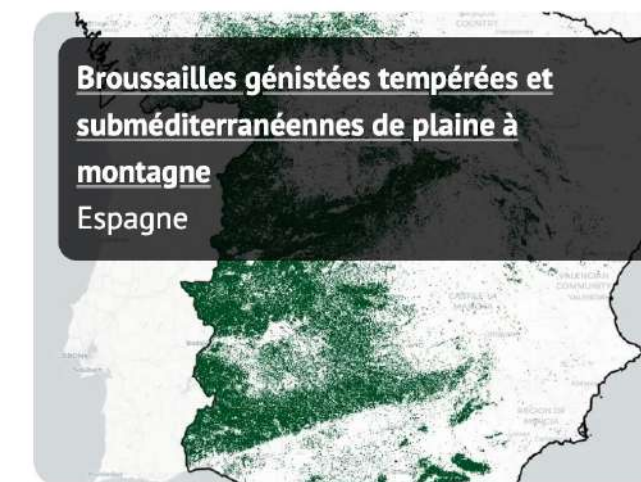
7 biodiversity indicators

200+ habitat types

Prédiction des espèces



Prédiction de la distribution des habitats



Indicateurs de biodiversité

Web application



geo.plantnet.org

Dataset and Species Distribution Model



github.com/plantnet/GeoPlant

Habitat Identification Model



github.com/cesar-leblanc/PlantBERT

GeoPl@ntNet: A Platform for Exploring Essential Biodiversity Variables

[Lukas Picek](#), [César Leblanc](#), [Alexis Joly](#), [Pierre Bonnet](#), [Rémi Palard](#), [Maximilien Servajean](#)

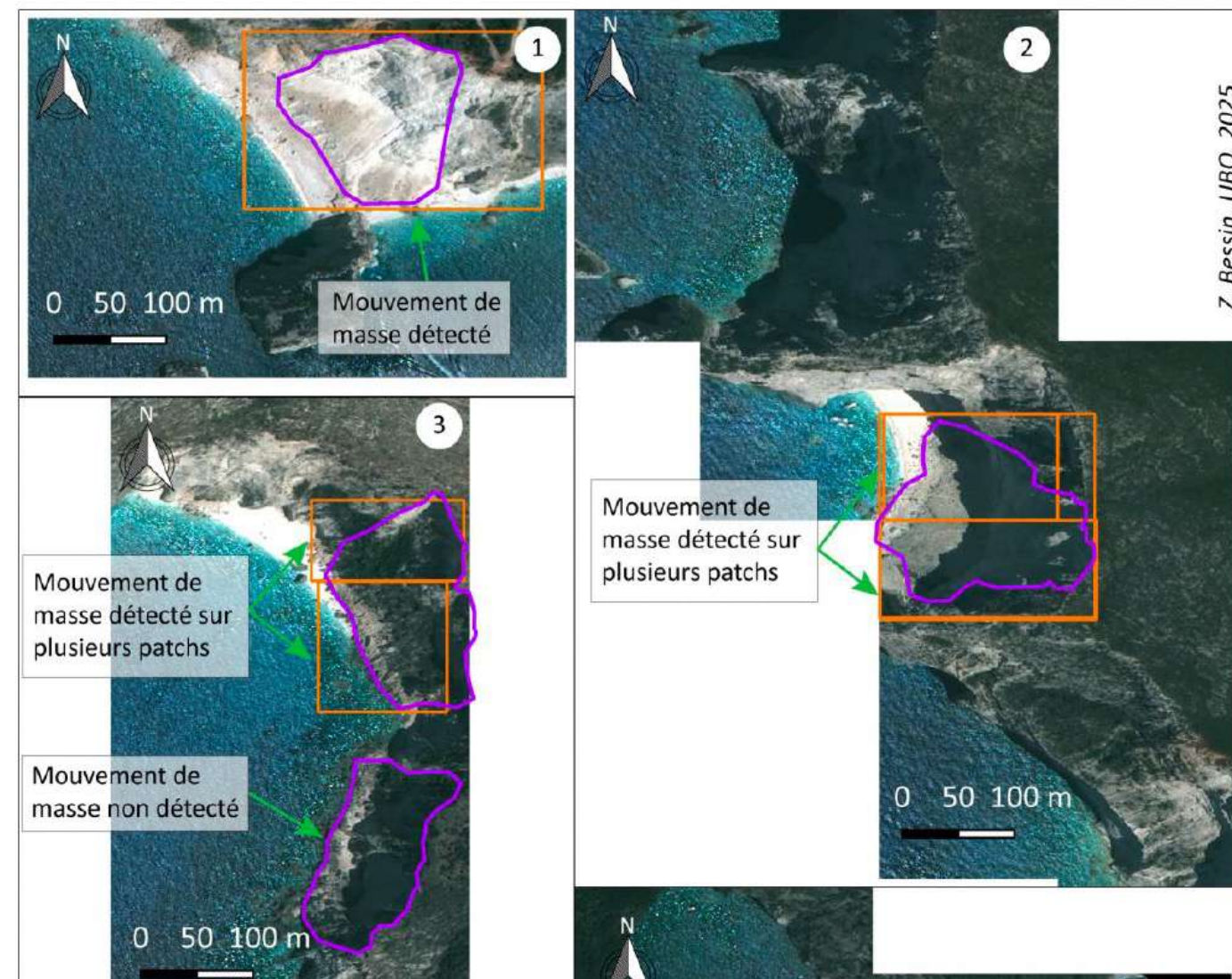
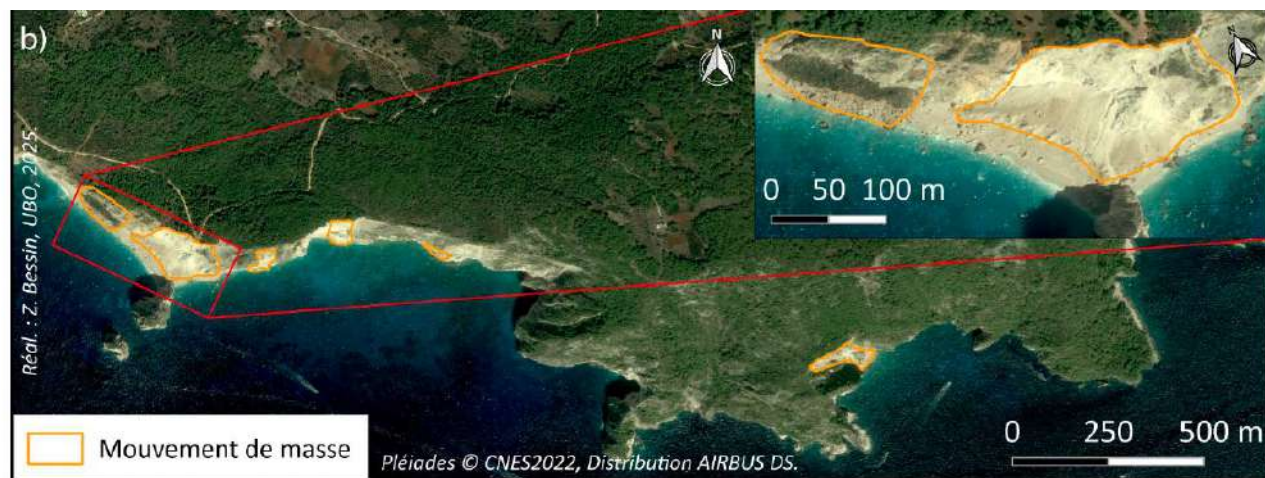
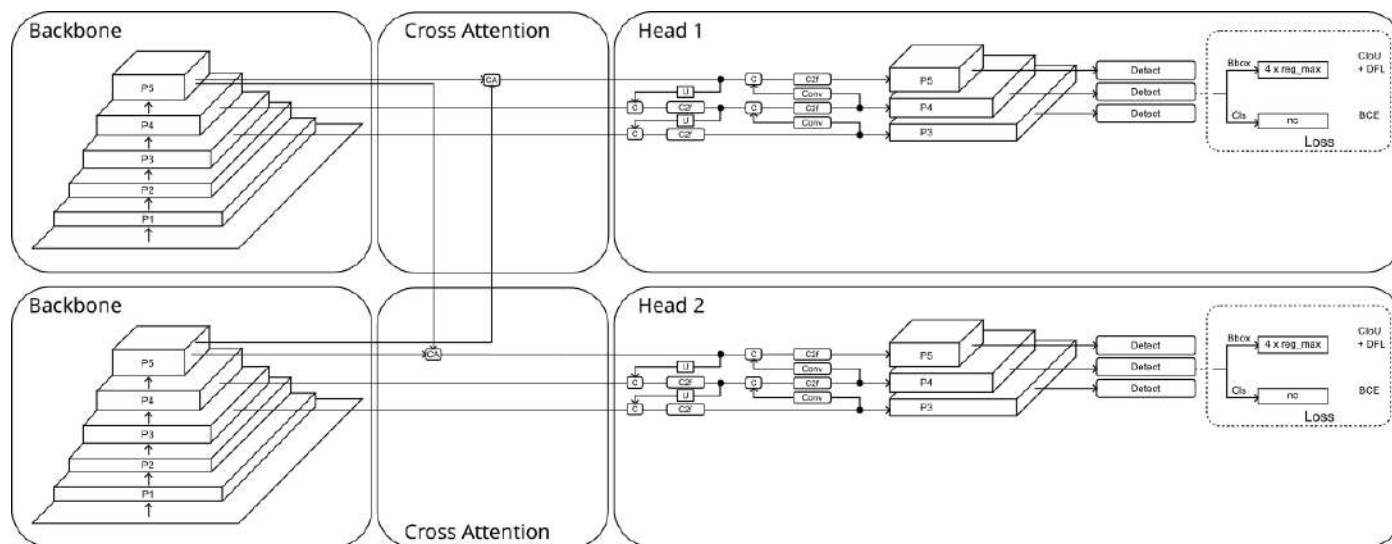
Détection de mouvements de masse sur le littoral

Thèse de Zoé Bessin (UBO, 2025) : exploitation d'images Pléiades obliques

En collaboration avec IRISA

Détection des mouvements de masse en tant qu'objets (falaise) : utilisation de l'apprentissage profond sur des données 2D

- Première architecture, YOLOv8
- YOLOv8+stereo. --> pour exploiter des images stéréoscopiques



Exemples de prédictions de mouvements de masse (MM) réalisées avec YOLOv8. 1)

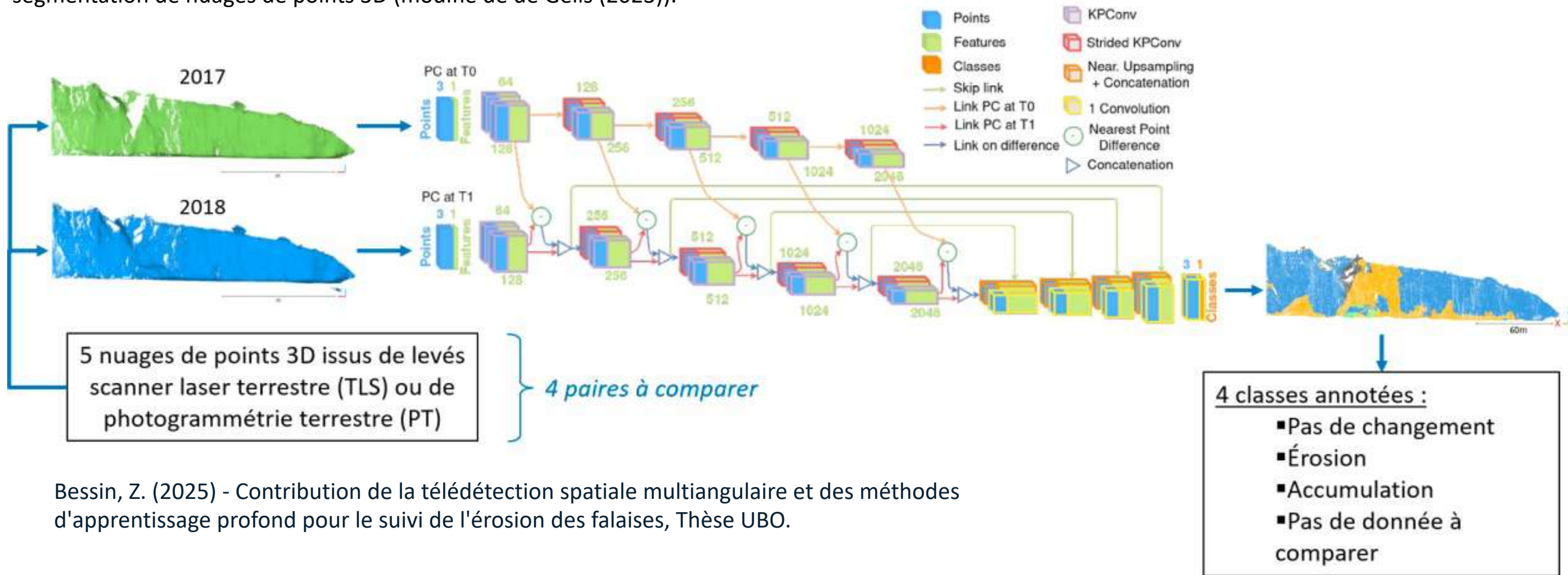
Détection de mouvements de masse sur le littoral

Thèse de Zoé Bessin (UBO, 2025)

En collaboration avec IRISA

Détection des changements à partir de nuages de points

Architecture Encoder Fusion Siamese Kernel Point Convolution pour la segmentation de nuages de points 3D (modifié de de Gélis (2023)).



Bessin, Z. (2025) - Contribution de la télédétection spatiale multiangulaire et des méthodes d'apprentissage profond pour le suivi de l'érosion des falaises, Thèse UBO.

CESBIO – CDS THEIA-CNES :

Super résolution et fusion de données

- Des méthodes d'apprentissage permettent :
 - D'améliorer la résolution apparente des images :
 - Super-résolution
 - On montre à un réseau profond de neurones des images à 10 et 20m (S2) et à 5m de résolution (VENμS) simultanées du même endroit, et il apprend à généraliser
 - Certains vont jusqu'à amener des données Sentinel-2 de 10 à 1m de résolution
 - Outil S2_Superrésolution (Julien Michel)
 - https://github.com/Evoland-Land-Monitoring-Evolution/sentinel2_superresolution
 - Bientôt installé comme **traitement à la demande** dans  GEODES
- De fusionner des séries temporelles de plusieurs capteurs

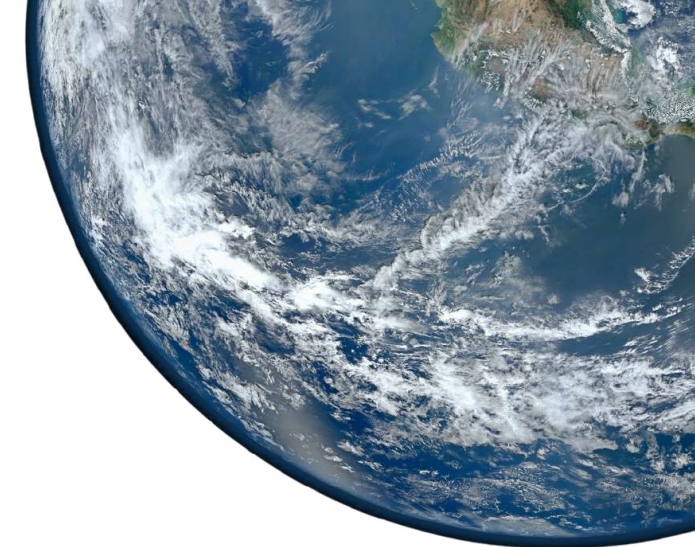


Image Classification



Object Detection



Semantic Segmentation



Instance Segmentation



Conclusions :

- Un usage de l'IA très diversifié
- Des réseaux spécialisés comme le GDR IASIS avec des événements autour de l'IA (communauté computer vision)
- Des approches qui mobilisent de plus en plus des données hétérogènes (images remote, in-situ mais aussi texte, données participatives, données climatiques, biodiversité, etc)
- Une valorisation croissante des jeux de données pour l'IA – via les catalogues de données
- Des besoins exprimés par les communautés pour accéder à des modèles d'IA (pré-entraînés), besoins en discussion – Via prochaine AG THEIA en mars 2026
- Usage de l'IA pour l'amélioration de la recherche dans les catalogues, l'enrichissement des metadonnées, etc – en lien avec les autres pôles de données / Data Terra



MERCI POUR VOTRE ATTENTION

Anne Puissant – Directrice du pôle THEIA, directrice adjointe Data-Terra
anne.puissant@data-terra.org

