



IA et Observation de la Terre

La révolution des embeddings et les modèles de fondation

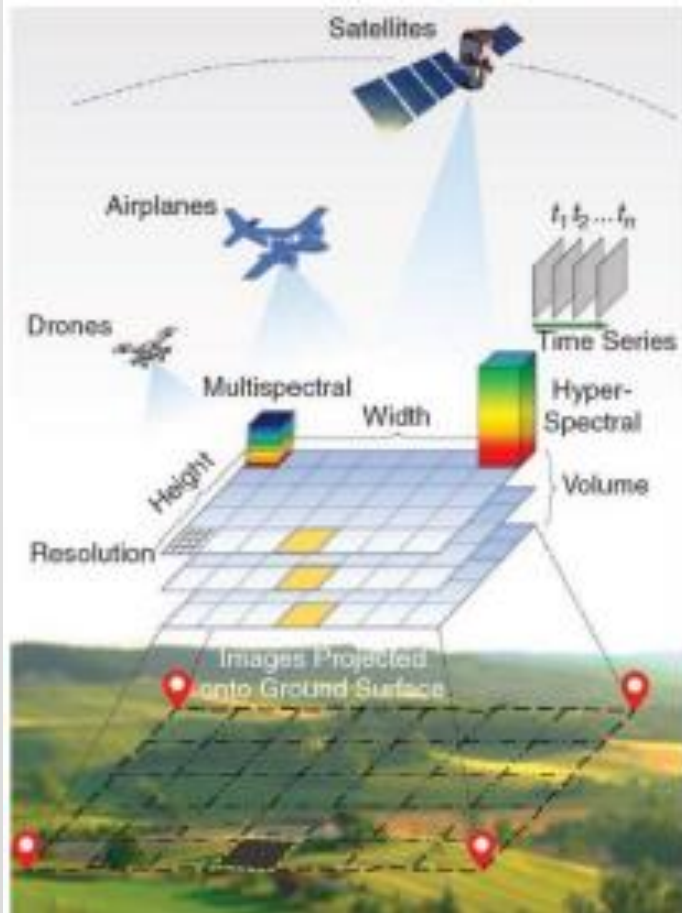
GISday
2025



Sommaire

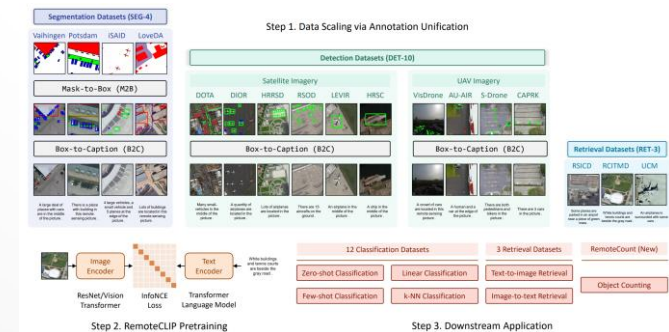
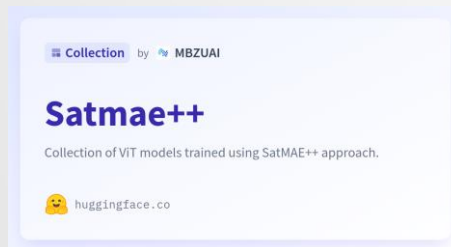
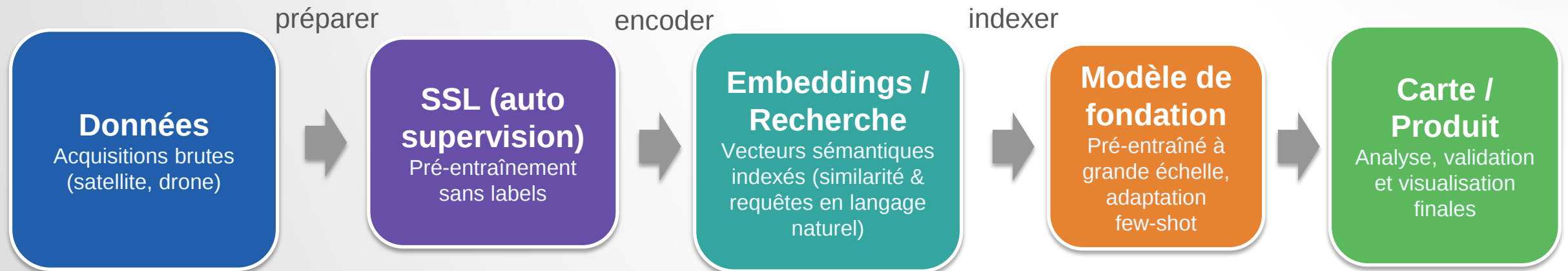
- Introduction : des labels rares à la recherche en langage naturel
- Qu'est-ce qu'une « bonne » carte ?
- L'apprentissage auto-supervisé (SSL)
- Une compréhension des embeddings
- La recherche sémantique
- Modèle de fondation pour l'observation de la Terre
- Conclusion : Quels sont mes garde-fous ?

Des labels rares à la recherche en langage naturel

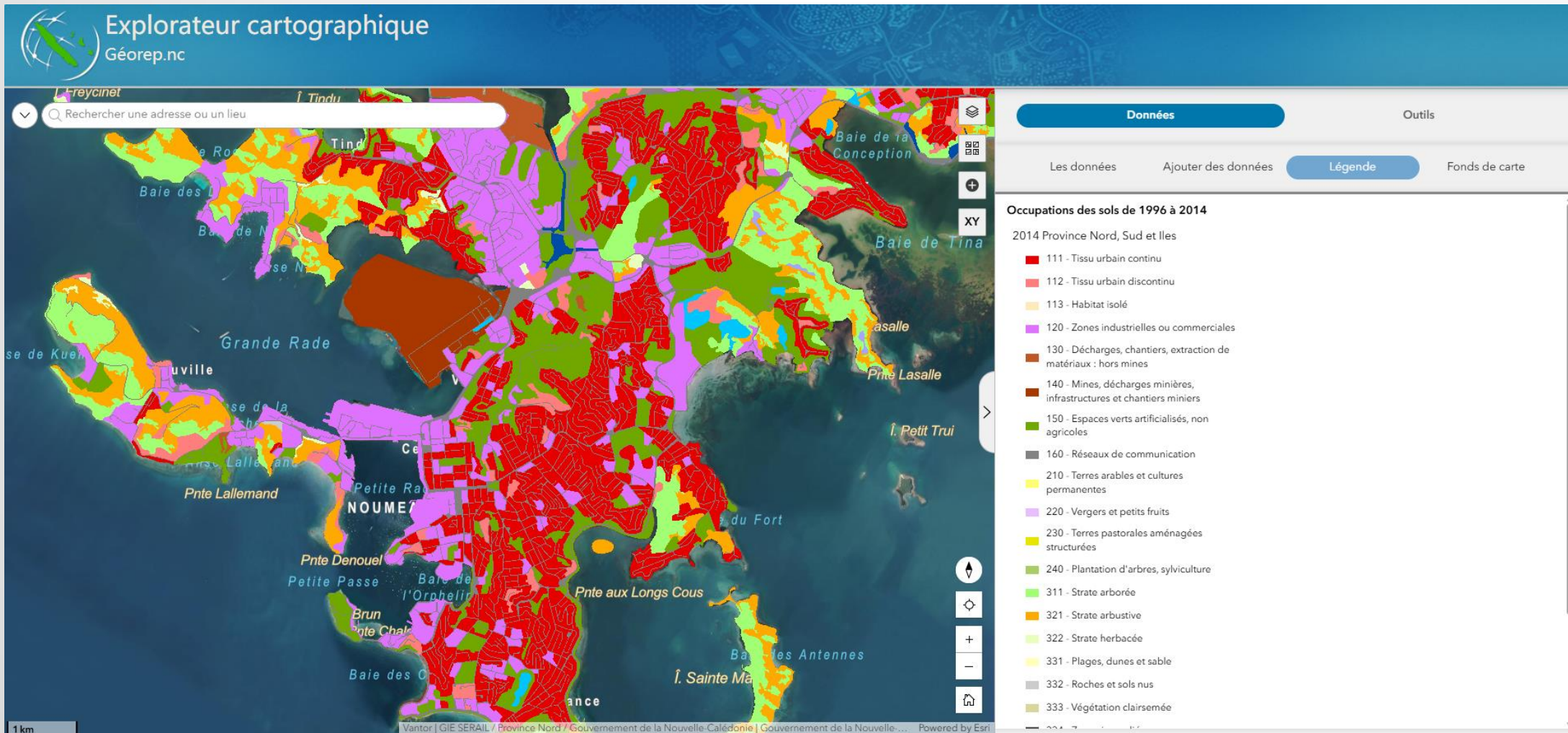


- Données EO :
 - Multimodales
 - Géolocalisée
 - Séries temporelles
 - Volume énorme
 - Beaucoup de métadonnées
 - Mesures de propriétés physiques
- Besoin de labels !
 - Donnée brute -> informations

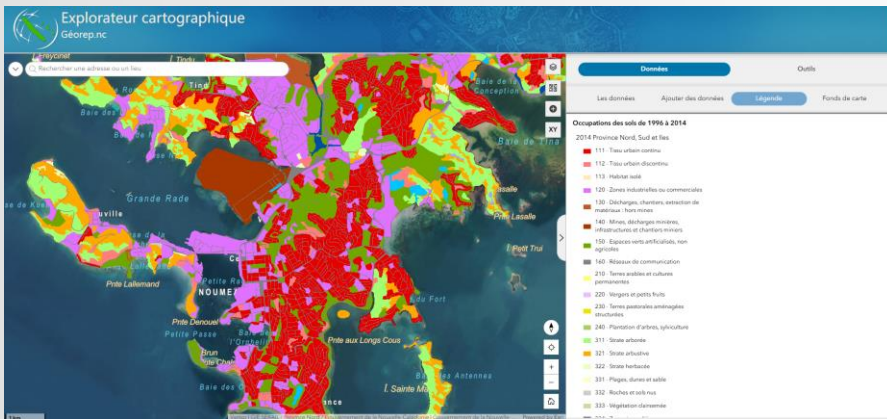
Des labels rares à la recherche en langage naturel



Qu'est-ce qu'une bonne carte ?



Qu'est-ce qu'une bonne carte ?



Toute carte thématique fiable (occupation du sol, aléas, dégâts...) part de labels de qualité. C'est souvent le goulot d'étranglement (coûteux, lents, hétérogènes).

**Une bonne carte part aussi d'une question que l'on se pose :
A quelle tâche doit répondre cette carte !**

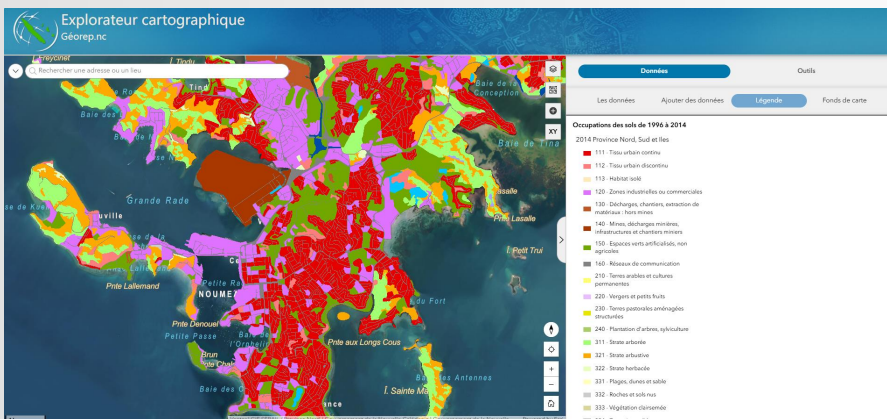
1) Tâches au niveau pixel (cartes continues)

- Segmentation sémantique (carte par classe) : occupation du sol, eau, neige/glace, zones brûlées, inondations, nuages/ombres.
- Détection de frontières / lignes : berges, routes, limites de champs.
- Estimation de fractions (sub-pixel) : abondances de matériaux/végétation par pixel.
- Régression biophysique : LAI, fAPAR, biomasse/AGB, hauteur de canopée (avec synergies), humidité des sols de surface, LST/SST, albédo, qualité de l'eau (chlorophylle-a, turbidité), AOD/NO₂/CH₄ (selon capteur).
- Cartes d'incertitude / calibration (évaluation et confiance par pixel).

2) Tâches au niveau objet (instances)

- Détection d'objets : navires (optique/SAR), avions, véhicules, éoliennes, panneaux solaires, bâtiments, réservoirs, puits, serres, tours de refroidissement.
- Instance segmentation : empreintes de bâtiments, houppiers d'arbres, polygones de champs.
- Comptage / densité : arbres, véhicules, bétail (selon résolution).
- Suivi d'objets (multi-dates) : flotte de navires, trafic portuaire/aéroportuaire.

Qu'est-ce qu'une bonne carte ?



3) Tâches au niveau scène/patch

- Classification de scène (mono-label) : type d'occupation du sol.
- Classification multi-label : co-présence (urbain + eau + végétation).
- Recherche par similarité / embeddings : "trouve des zones similaires à..." (retrieval).
- Recherche texte ↔ image (cross-modal, type CLIP) : requêtes en langage naturel sur vos archives ("quartiers résidentiels denses près du littoral").
- Légendage automatique (image captioning) et VQA géospatiale (questions-réponses visuelles).

Toute carte thématique fiable (occupation du sol, aléas, dégâts...) part de labels de qualité. C'est souvent le goulot d'étranglement (coûteux, lents, hétérogènes).

**Une bonne carte part aussi d'une question que l'on se pose :
A quelle tâche doit répondre cette carte !**

4) Tâches temporelles (séries multi-dates)

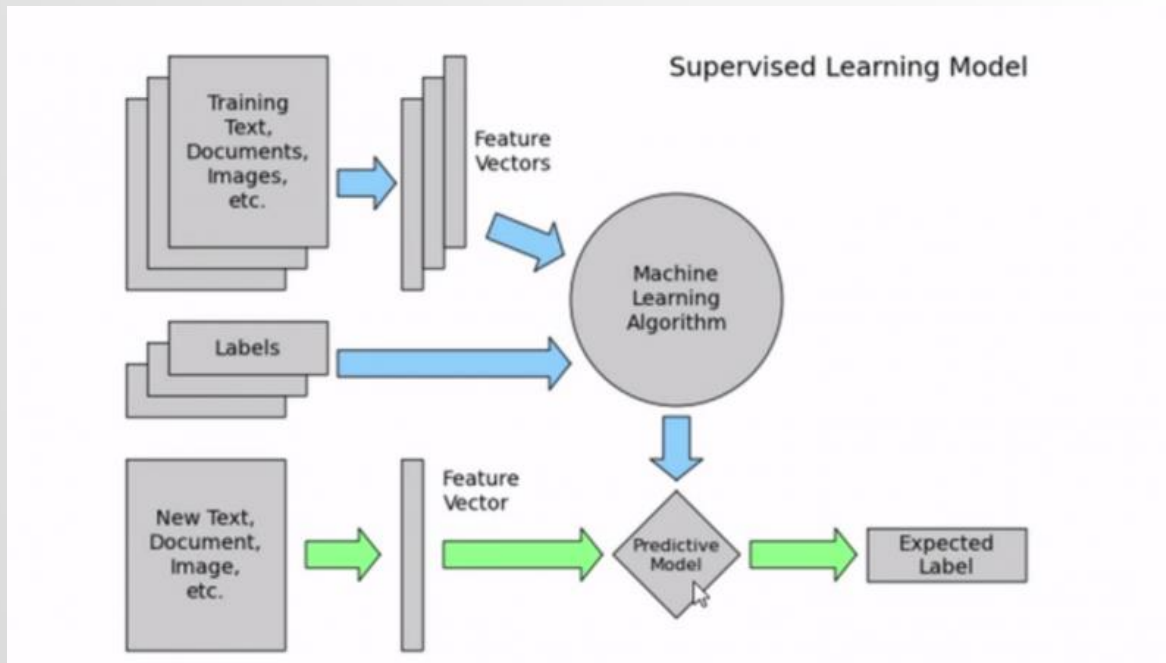
- Détection de changement
 - *Binaire* (changement / pas changement).
 - *Sémantique* (de → à : forêt → urbain, sec → inondé).
 - *Orientée objet* (apparition/disparition/déplacement d'objets).
- Détection d'événements : déforestation, brûlis, crues, glissements, constructions.
- Phénologie / stades culturels : dates clés, durée de cycle.
- Prévion : NDVI/LAI, rendements, humidité des sols, niveaux d'eau (avec données exogènes si besoin).
- Anomalies / alerte précoce : repérage de comportements atypiques sur séries.

« Réduire la dépendance aux labels humains »
SSL4EO

Le Self Supervised Learning

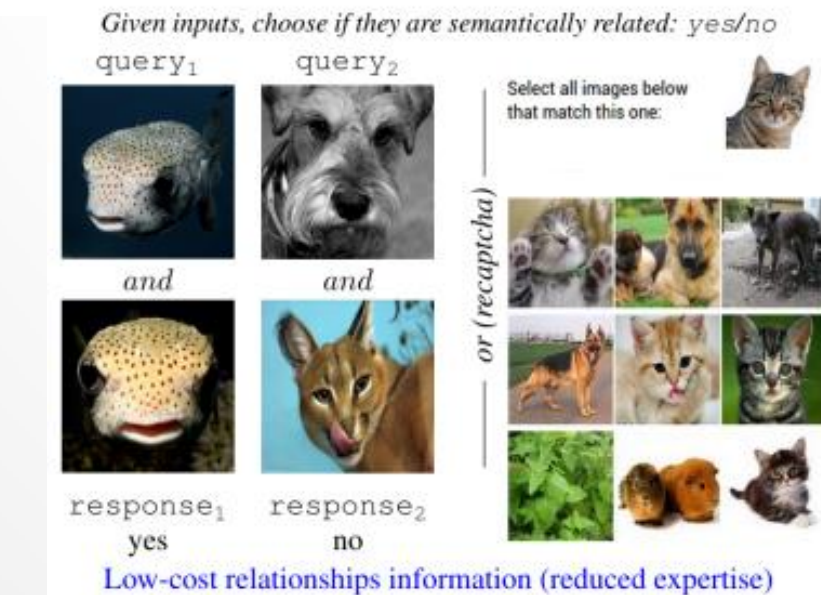
Principe du SSL

Apprendre quelque chose de ses datas sans labels disponibles...



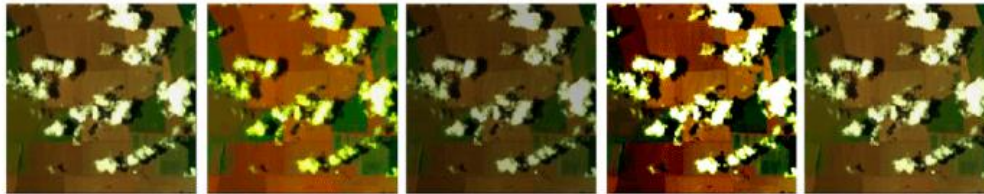
SOURCE : A Self Supervised Tour of Representation Learning

Randall Balestrieri



Data Augmentation pour créer des vues d'un input X

Color based
($\pm 20\%$)



Original

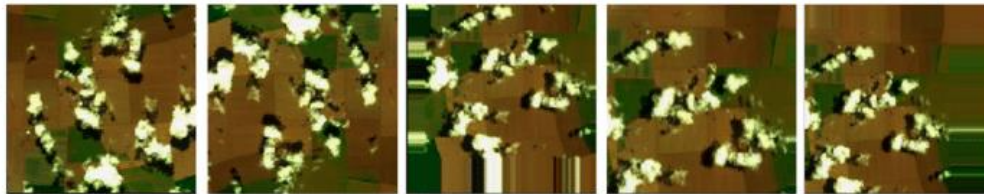
Saturation

Brightness

Contrast

Sharpness

Geometry based
($90^\circ, \pm 20\%$)



Rotate

Flip

Zoom in

Zoom out

Shift

Résultat : des features prêtes à l'emploi qui réduisent le volume de labels nécessaires et transfèrent mieux que l'ImageNet classique sur des tâches EO

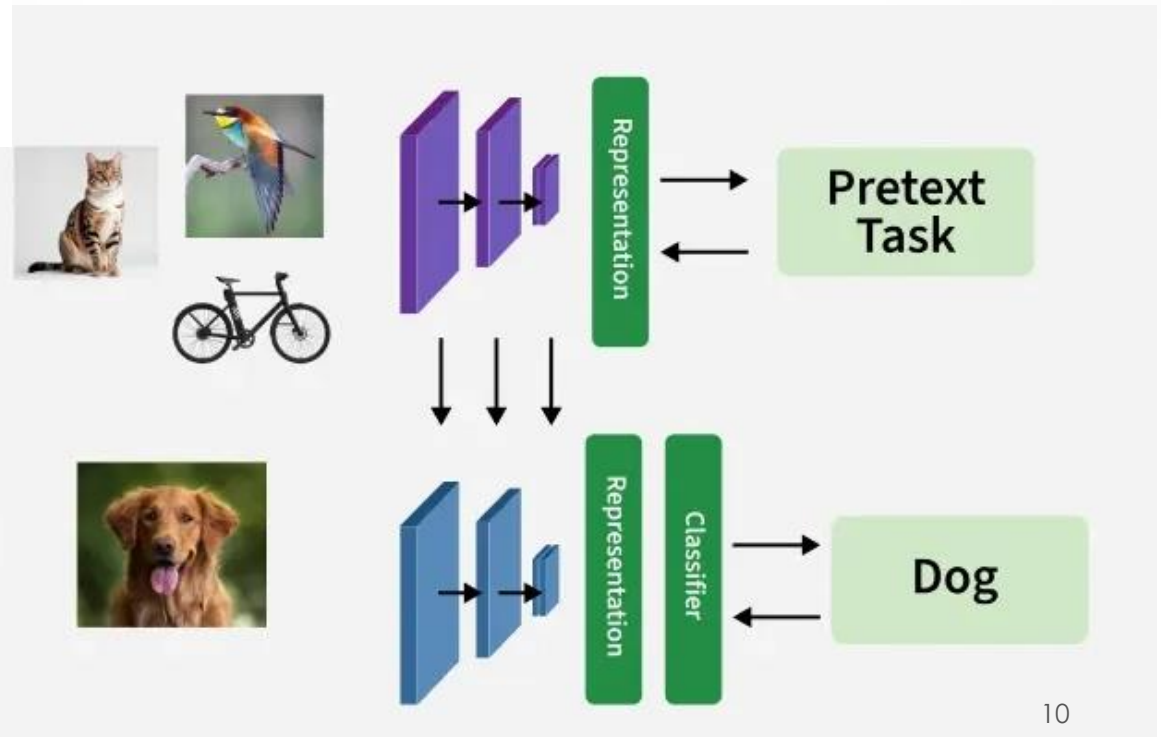
Apprentissage auto-supervisé (SSL) =

apprendre des représentations sans labels humains, en se donnant des tâches prétextes (reconstruire des pixels masqués, rapprocher des vues saisonnières du même lieu).

Deux familles utiles en EO :

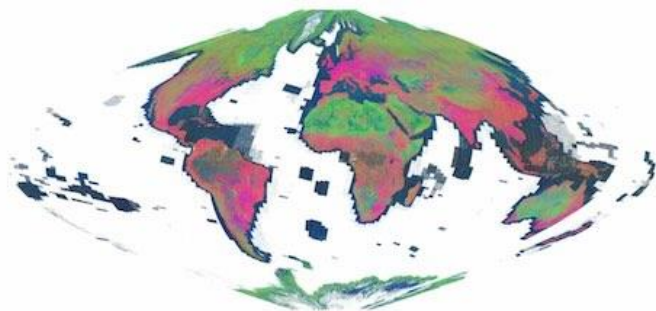
1. Contrastif spatio-temporel (SeCo : « été vs hiver au même pixel » → positives) ;
2. Masquage/ré-assemblage (SatMAE : on masque des patches multispectraux et on les reconstruit).

Pretrain



« Comment encoder l'information géographique »

Les embeddings



Core-S2RGB-SigLIP

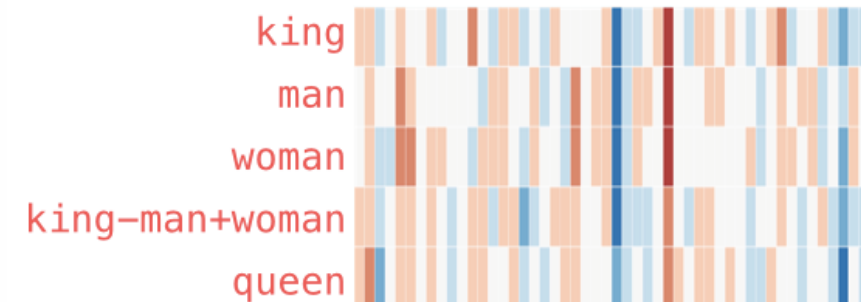


Core-S1RTC-SSL4EO



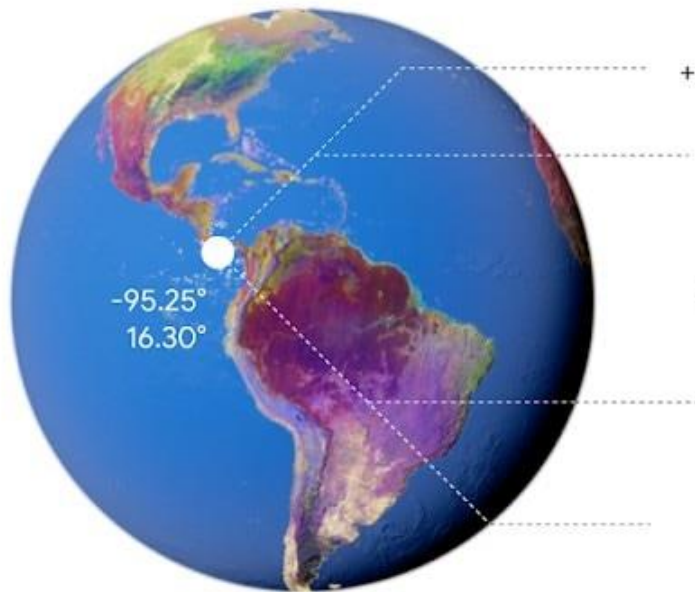
Core-S2L1C-SSL4EO

king - man + woman $\approx$ queen

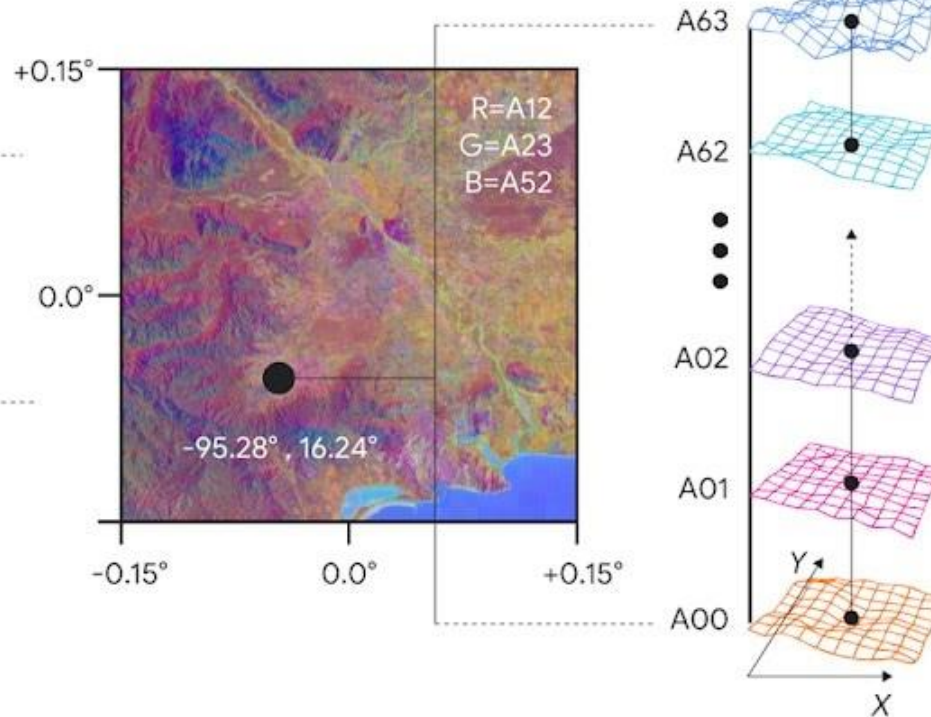


Les embeddings

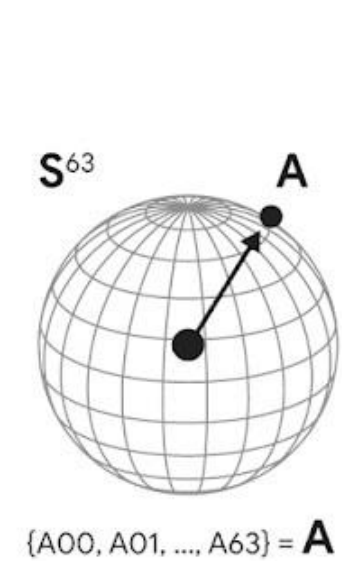
Global Embedding Field



Embedding Axes



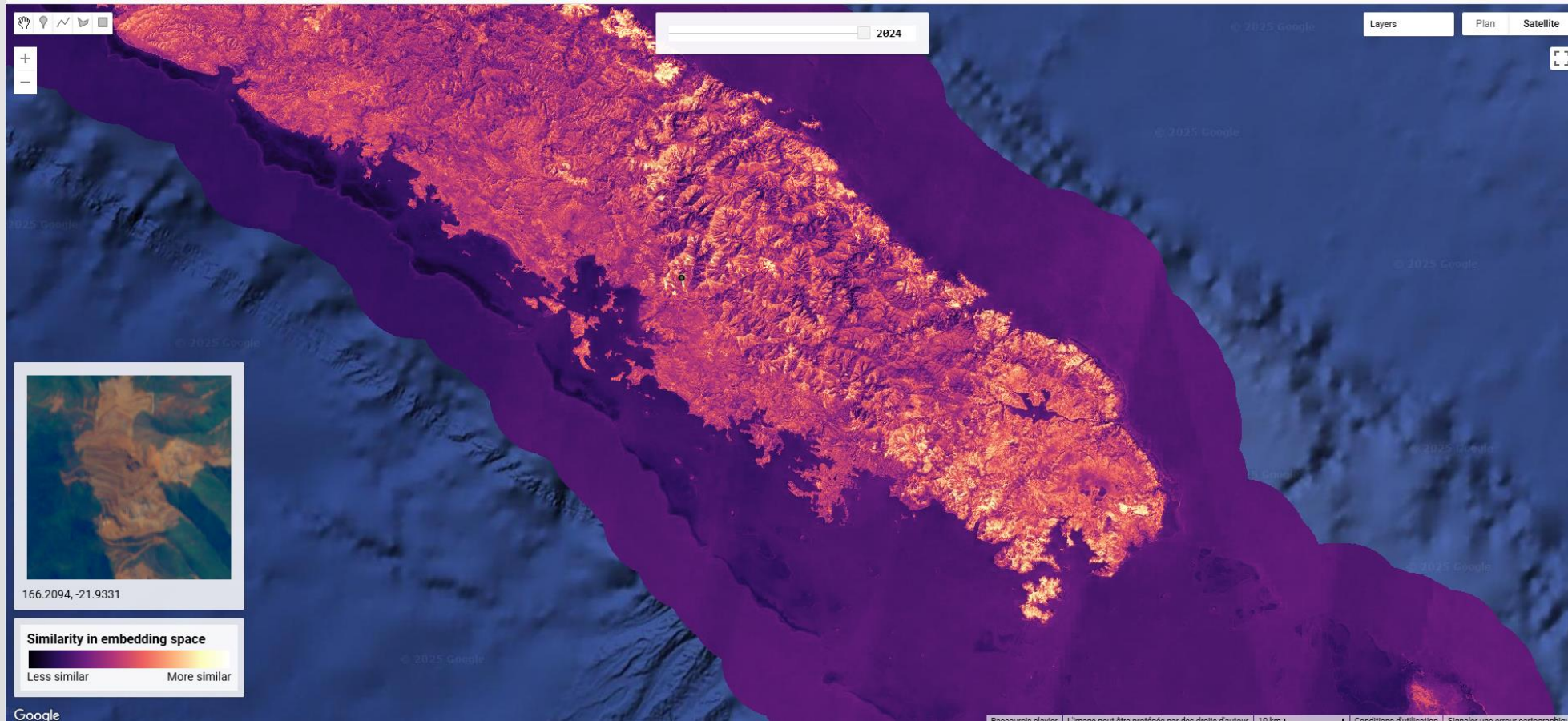
Embedding Vector



« Rechercher dans nos données »

La recherche sémantique

L'exemple d'AlphaEarth Foundations



[embedding-similarity-search](#)

[AI-powered pixels: Introducing Google's Satellite Embedding dataset | by Google Earth | Google Earth and Earth Engine | Medium](#)

« Pour l'observation de la Terre »

Modèles de fondation, why ?

Un changement de paradigme

L'Approche "Classique" (Notre défi)

Pour chaque nouvelle carte (ex: bâti, inondation), on entraîne un modèle de zéro.

Exigence énorme : des milliers de polygones/labels (vérité terrain) sont nécessaires.

Difficile à généraliser à d'autres capteurs, d'autres lieux ou d'autres saisons.

L'Approche "Fondation" (La promesse)

On pré-entraîne un modèle **gigantesque** sur des Pétaoctets de données (ex: tout HLS).

Ce modèle "comprend" la sémantique de l'imagerie EO (saisons, objets, textures).

Adaptation rapide (fine-tuning) avec très peu de labels locaux.

Définitions, exemple clé



1. Un Socle Pré-entraîné

Un modèle unique, pré-entraîné sur une quantité massive de données (HLS). Il sert de base (fondation) pour de multiples tâches.



2. L'Exemple : Prithvi-EO

Modèle d'IBM / NASA entraîné sur Harmonized Landsat-Sentinel-2. Le "GPT" de l'imagerie EO, disponible sur Hugging Face.



3. L'Adaptation (Fine-Tuning)

On le spécialise avec peu de labels pour : segmentation (bâti), classification (OCS), détection de changements...

Les 3 bénéfices majeurs

1. Apprentissage "Few-Shot"

Plus besoin de 10 000+ labels. On obtient des résultats performants avec **100 à 500 labels** (vérité terrain locale).

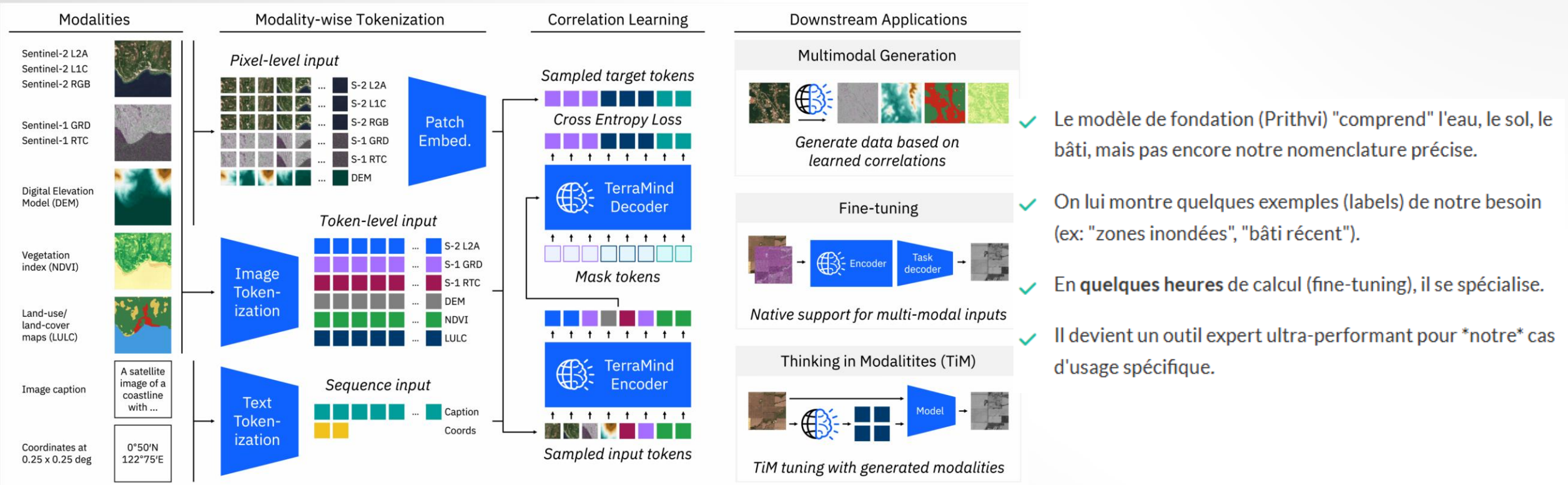
2. Multi-Temporel & Multi-Capteurs

Le modèle a déjà "vu" des milliers de scènes (saisons, nuages). Il gère mieux la **saisonnalité** et la fusion de capteurs.

3. Réduction du "Time-to-Map"

Le goulot d'étranglement (annotation) est réduit. On passe de **mois** de collecte à **jours** (ou heures) de fine-tuning.

Concept : du modèle socle au produit fini



« Les limites et garde-fous »

Conclusion : les limites

MERCI !

Thomas AVRON, APID, porteur du projet OCEAN'IA
De formation à l'IA dans le Pacifique !

