

# Utilisation d'images satellites par l'INSEE

Séminaire PROGYSAT 27 JUIN 2023

Philippe DORELON

Chef du Service Territorial de Guyane



## INSTITUT DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

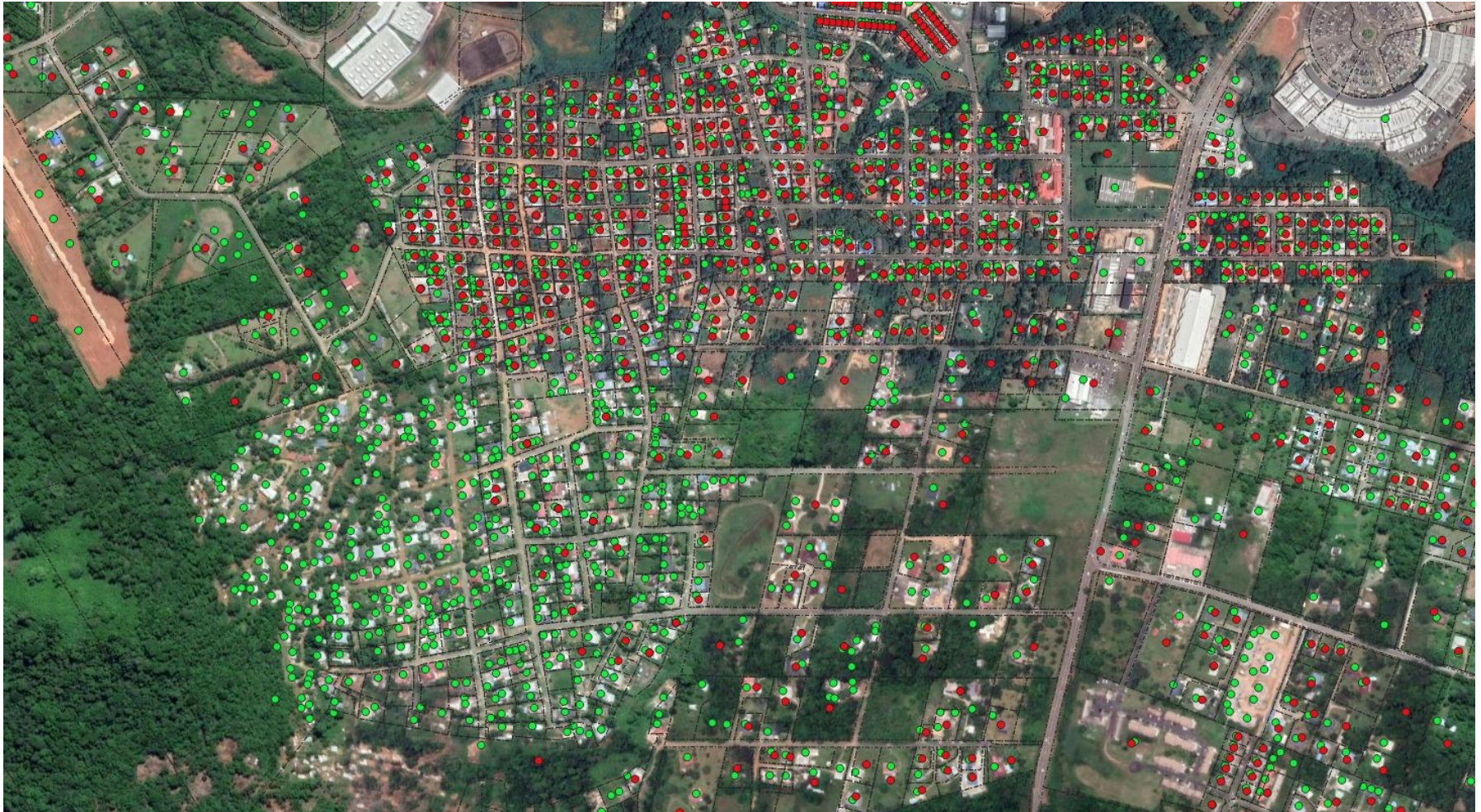
- 6000 agents répartis sur le territoire
  - En Guyane, 25 agents en bureau et 30 enquêteurs
- Gestion de répertoires : individus, fichier électoral, entreprises
- Production de statistiques issues de différentes enquêtes
  - Recensement de la population
  - Enquêtes auprès des ménages
  - Enquêtes auprès des entreprises
- Publication d'études à partir des statistiques produites

- **INSEE chargé de réaliser le recensement de la population avec les communes**
  - L'INSEE pilote, organise, contrôle, exploite
  - Les communes réalisent la collecte
- **Recensement par sondage dans les communes de plus de 10000 habitants (8/22)**
  - 8 % des logements recensés chaque année
- **Recensement exhaustif tous les 5 ans dans les petites communes**

- **Un sondage implique une base de sondage dans laquelle on va tirer les logements à recenser**
- **En France métropolitaine :**
  - Mise à jour avec utilisation des permis de construire : positionnement avec l'adresse et la parcelle cadastrale
- **En Outre-Mer**
  - Problèmes d'adressage et d'exhaustivité, habitat illégal important en particulier en Guyane et à Mayotte
  - Nécessité de réaliser une enquête cartographique pour mettre à jour le RIL

- **Opération lourde à réaliser chaque année**
  - Des enquêteurs sur le terrain actuellement
  - Pour mettre à jour un cinquième du territoire
- **Besoin d'avoir des données externes pour cibler le travail et produire des plans pour la collecte**
  - Images aériennes, satellites
  - Données de promoteurs immobiliers, de la Poste, des communes

- **Des tentatives d'utilisations d'images depuis 15 ans**
  - Couverture IGN
  - Données en ligne : Google, Bing,...
  - Données locales aériennes réalisées par SIAGE
  - Obtention d'images par les communes, des services de l'État
  - Via DINAMIS et désormais SEAS



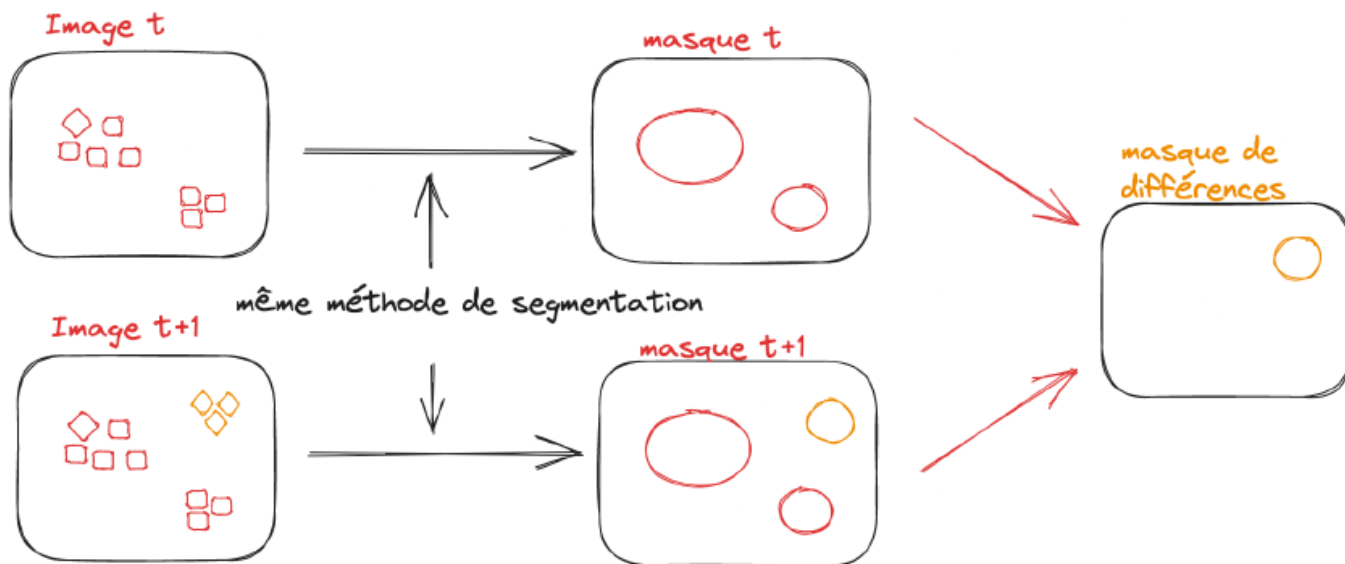




- **Ciblage par contrôle visuel et échanges avec les communes**
  - On ne va pas dans les quartiers stables
- **Souhait de détecter automatiquement les zones en évolution**
  - Travaux initiaux par l'INSEE au niveau national
  - Mise en œuvre de techniques de deep learning appliquées aux images
    - **L'INSEE a participé à un hackathon sur la détection du bâti**
    - **Des résultats encourageants qu'on souhaite appliquer dans les DOM en particulier en Guyane et à Mayotte**

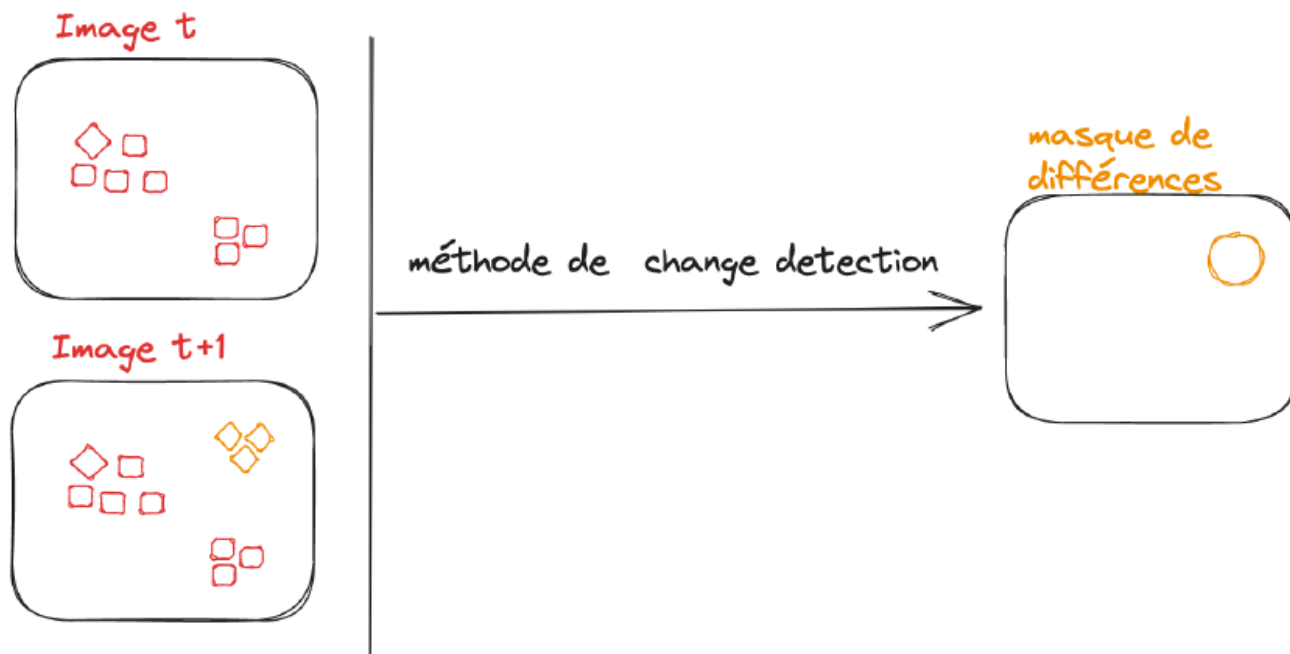
Première stratégie envisagée :

- Segmenter les images couvrant un même territoire à deux dates différentes avec le même procédé et calculer un masque de différence !



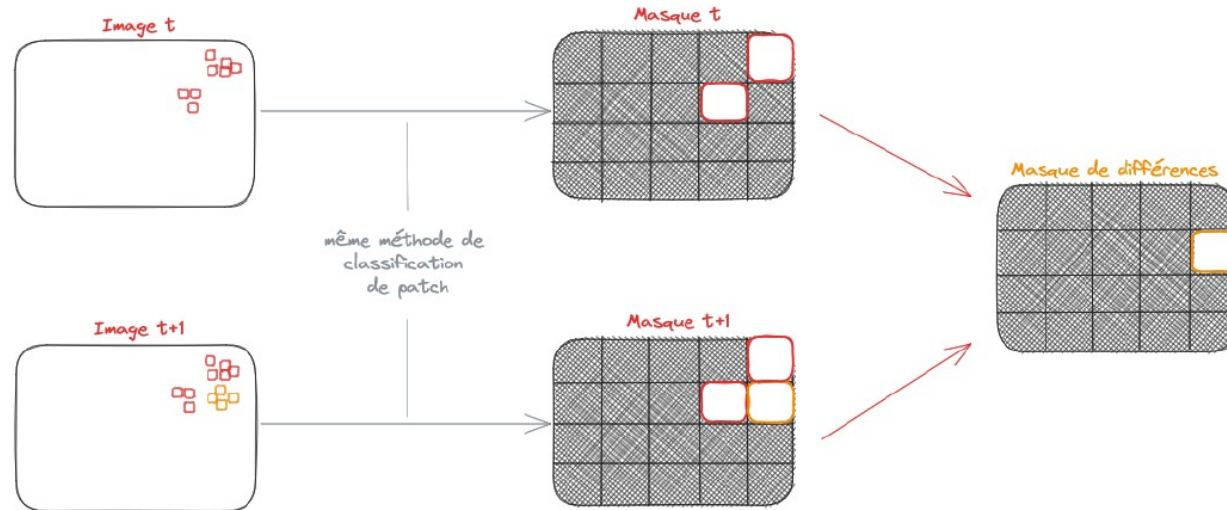
Deuxième stratégie envisagée :

- Calculer directement un masque de changement à partir des deux images (sans nécessairement segmenter au préalable)



Troisième stratégie envisagée :

- Découper l'image en une grille de patches et faire une classification binaire : présence de bâtiments ou non. Puis construction d'un masque de l'ensemble des patches blancs (avec bâtiments) et noirs (sans bâtiment) de l'image.



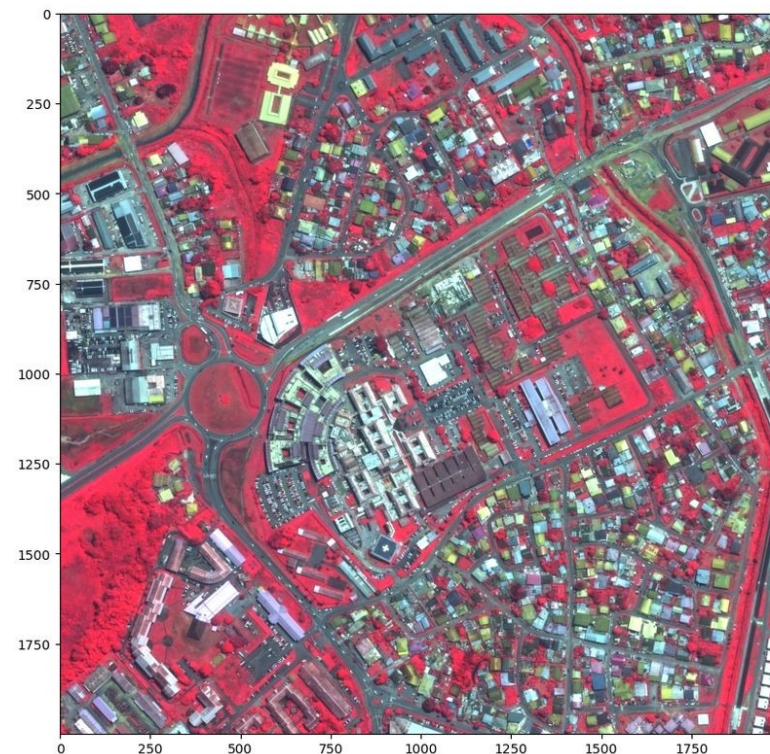
- On optera pour la première stratégie dans un premier temps, basée sur la segmentation préalable



## – Deux grands types de méthodes :

- Les méthodes dites "spontanées" qui vont s'appuyer directement sur les valeurs des pixels d'une image donnée
- Des méthodes plus algorithmiques qui vont s'appuyer sur un ensemble d'images sélectionnées et des informations extérieures sur la position des logements
  - **On les évoquera dans la partie suivante**

- La bande infra rouge est connue pour bien discriminer les végétaux des constructions
- La bande infrarouge n'est pour le moment pas disponible pour tous nos territoires d'où les essais qui suivent



- On obtient une image en niveau de gris en moyennant sur les 3 couches R, G, B
- On obtient donc un seul et unique tableau de valeurs comprises entre 0 et 255 que l'on peut également représenter de cette façon
- Il n'y a plus qu'à « seuiller » les intensités lumineuses pour obtenir un masque convenable

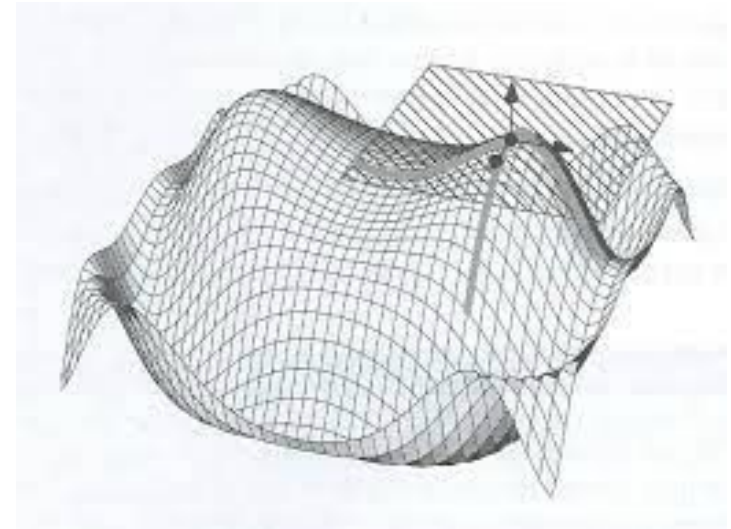




image 2020



masque seuillage RGB 2020

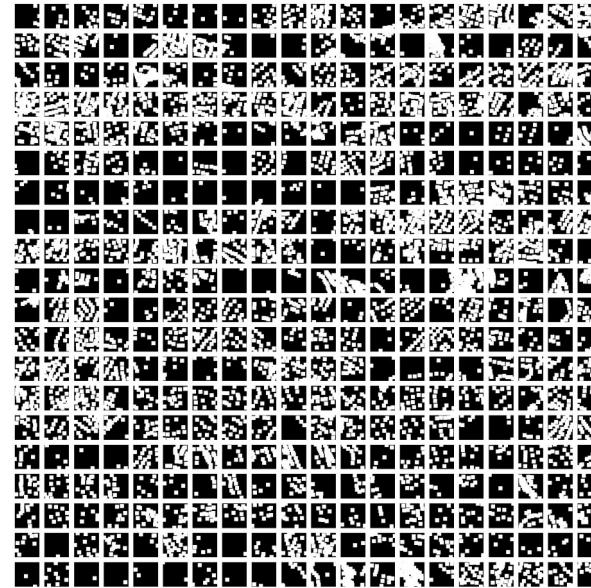


## – Problème posé par le seuillage

- le seuil dépend de l'image considérée, difficile de trouver un procédé automatique . . .

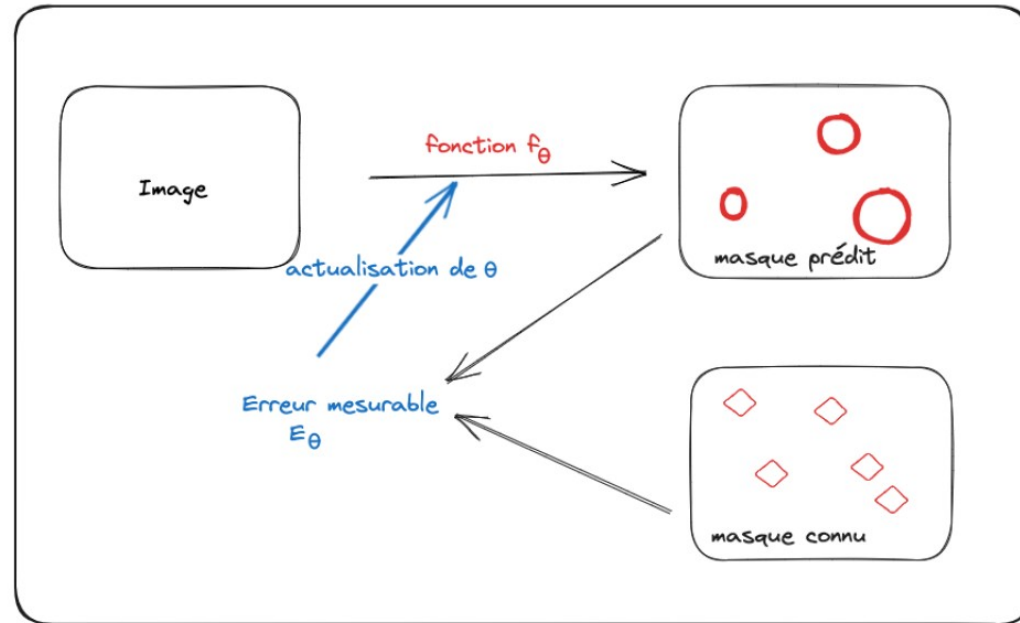
- **On souhaite entraîner un modèle (de deep learning) à détecter automatiquement des logements sur des images**
- **Un tel modèle doit être entraîné sur des exemples annotés et de bonne qualité**
- **Il faut donc produire ces annotations et filtrer les images de mauvaise qualité**

- **Obtenir ces annotations peut être coûteux, mais on dispose de données permettant de s'approcher :**
  - **RIL (INSEE)**
    - **permet d'avoir des points millésimés précisément mais les approximations de contours de logements sont grossières**
  - **La BDTOPPO (IGN)**
    - **moins "millésimable" mais les contours de bâtis obtenus sont plus précis**
  - **D'autres sources de données sont mobilisables pour les annotations (e.g données Open Street Map ou jeux de données déjà annotés issus de la recherche..)**

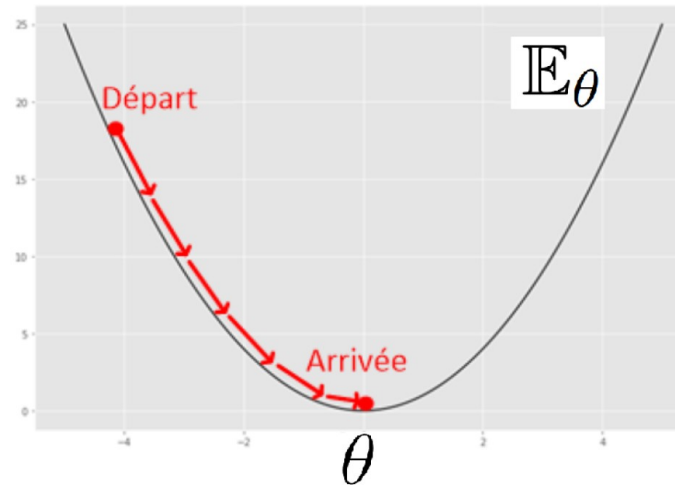


- On peut désormais entraîner des modèles sur nos données

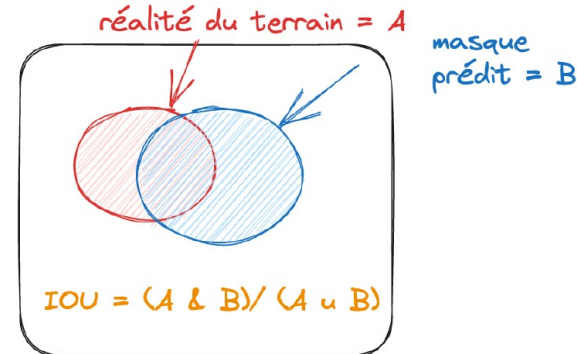
- On entraîne un modèle à prédire un masque correct à partir des couples image/masque obtenus



- Fonctionnement par itération successives, on applique le modèle à des petits paquets de couples images/masques, et on bouge les paramètres de sorte à diminuer l'erreur commise sur le paquet
- Au bout de plusieurs itérations, les paramètres du modèle convergent et ce dernier devrait être en mesure de prédire l'emplacement des logements à partir de seules images



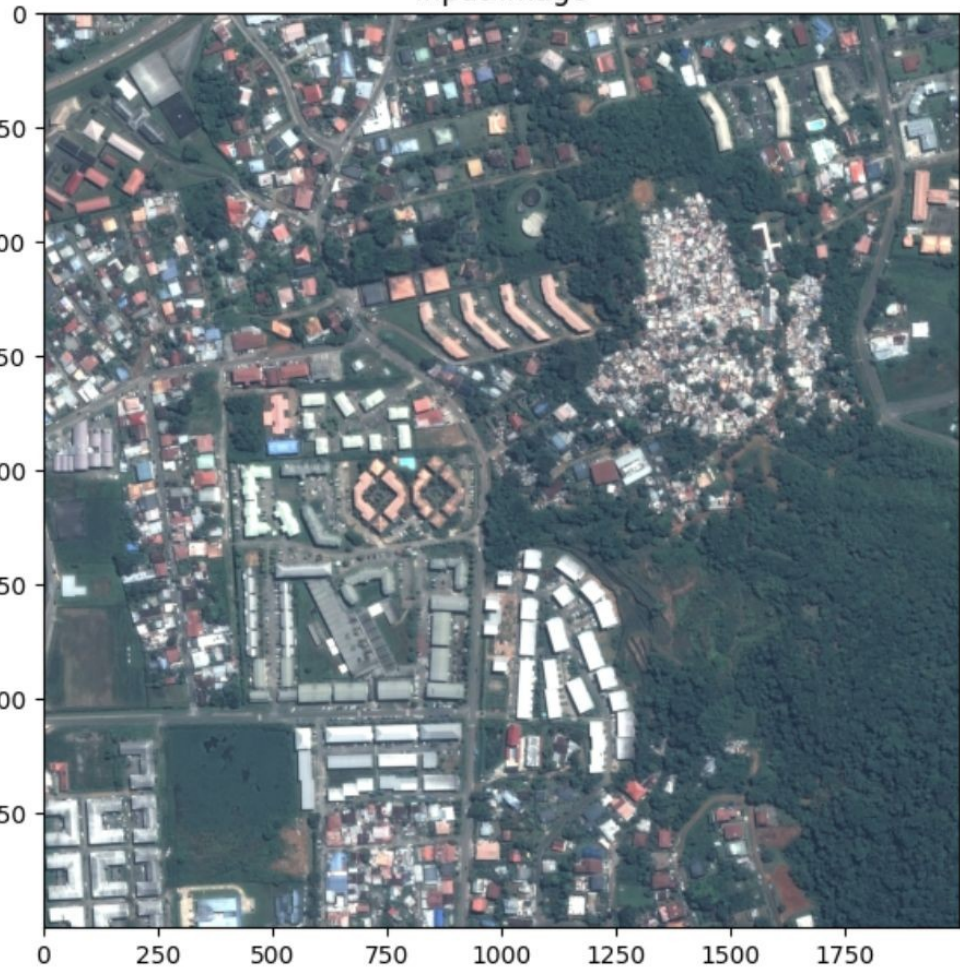
- On s'entraîne ici sur l'ensemble des images Pleiades disponibles pour la Martinique en 2022 :
  - Chaque image est découpée en patches de taille 250 x 250 pixels
  - On se sert du RIL avec zone tampon de 10 m pour labelliser les images
  - 10 000 images au total dont :
    - 80 % pour l'entraînement de l'algorithme (jeu d'apprentissage)
    - 20 % pour l'évaluation des performances (jeu de validation)
- 10 h d'entraînement sont nécessaires
- Comment évaluer les résultats ?
  - IOU : Intersection Over Union



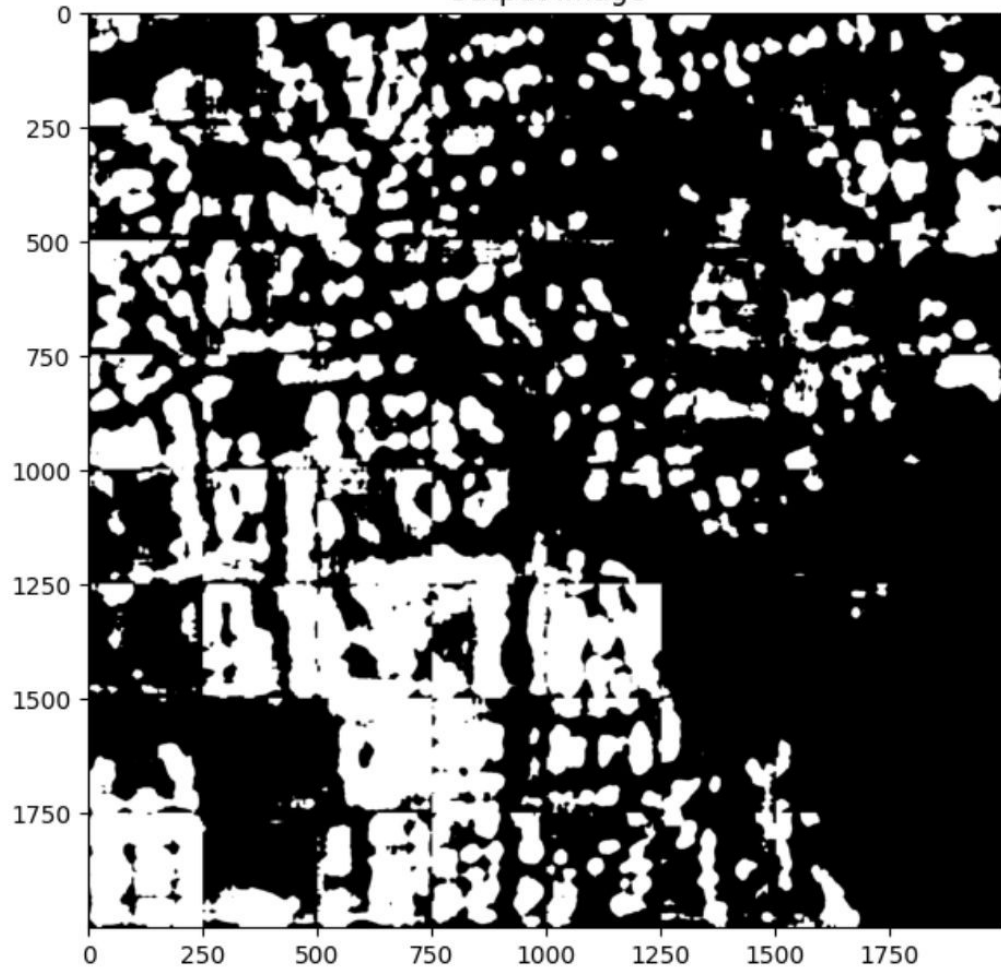
- On a besoin d'un jeu d'images test correctement labellisées pour évaluer plus finement nos résultats
  - peut prendre du temps
- C'est la seule façon de valider nos résultats qualitativement
- Recherche manuelle de cas intéressants sur lesquels tester nos méthodes
- Travail de labellisation à la main ?

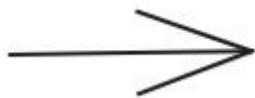


Input Image

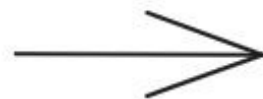


Output Image





modèle



- Résultats cohérents dans l'ensemble, plus difficile de détecter des changements que de segmenter
- Nécessaire de produire beaucoup plus d'exemples de tests
- Une grosse limite potentielle
  - la poursuite des travaux est conditionnée à l'obtention des images Pleiades sans nuages avec une couverture et dans un temps raisonnables

- **Découpage des images en petit patches et classification des patches (moins exigeant que la segmentation)**
- **Approche "object detection" à implémenter**
- **Entraînement sur jeux de données déjà annotés**
- **Entraînement des algorithmes à détecter directement les changements (difficulté pour trouver suffisamment d'exemples)**
- **Même type de travaux sur les données Sentinel 2 en cours**
- **Échanges avec l'IGN à venir**

- **Ces travaux ne s'arrêtent pas à la seule détection de changements, d'autres utilisations sont envisageables pour l'INSEE :**
  - Estimation de population à partir des images Pleiades, Sentinel 2, Sentinel 1 (radar)
  - Mesure d'étalement urbain, consommation d'espace
  - Coopération à l'international e.g avec Afristat
    - Ces méthodes poussées permettent de pallier l'absence de répertoire statistique fiable

## Retrouvez-nous sur

[insee.fr](https://www.insee.fr)



UTILISATION D'IMAGES SATELLITES PAR L'INSEE 27 JUN 2023



Mesurer pour comprendre