

How Deep Learning help Change Detection and upgrade Data Quality.



Kruwialis Gaël
Spatial Data Information | IGN-NGI
24/10/2019

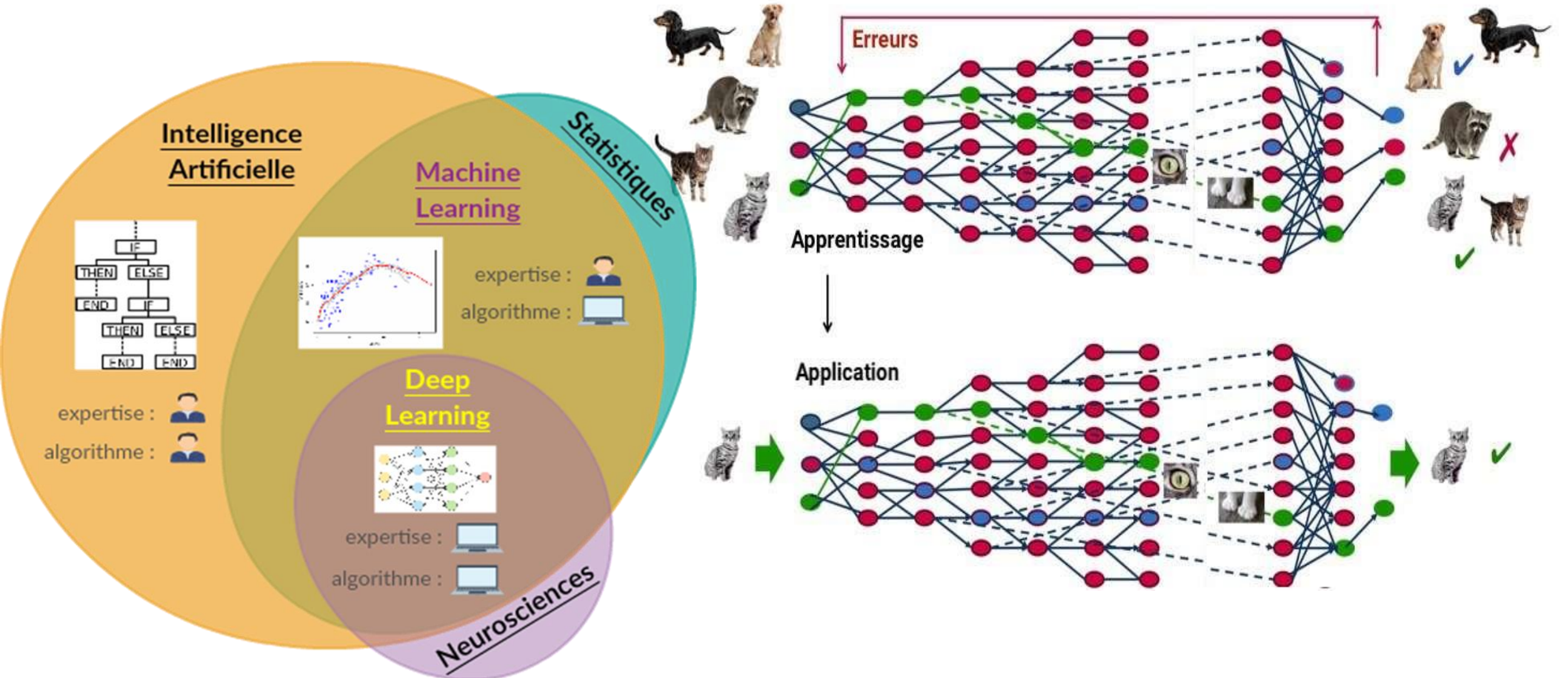
Etape



- Principe
- Objectif (IGN)
- Données Sources
- Outils
- Résultats
- Conclusions
- Perspective

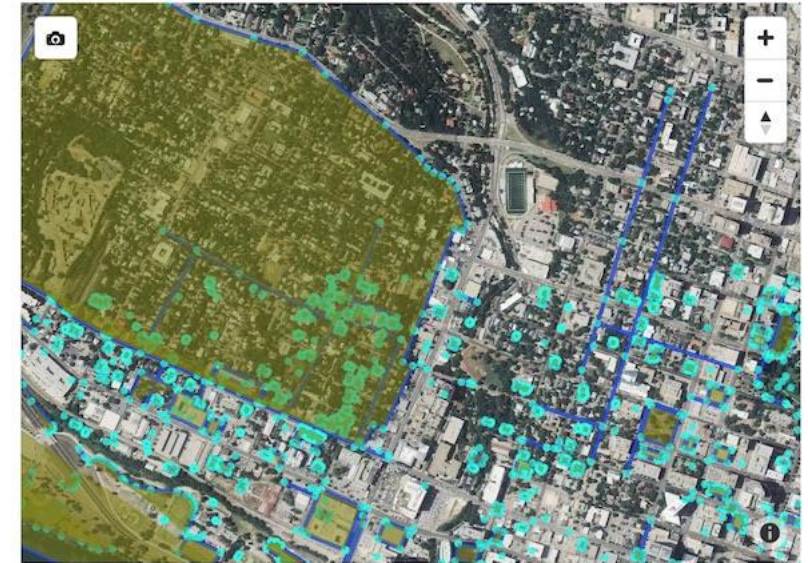
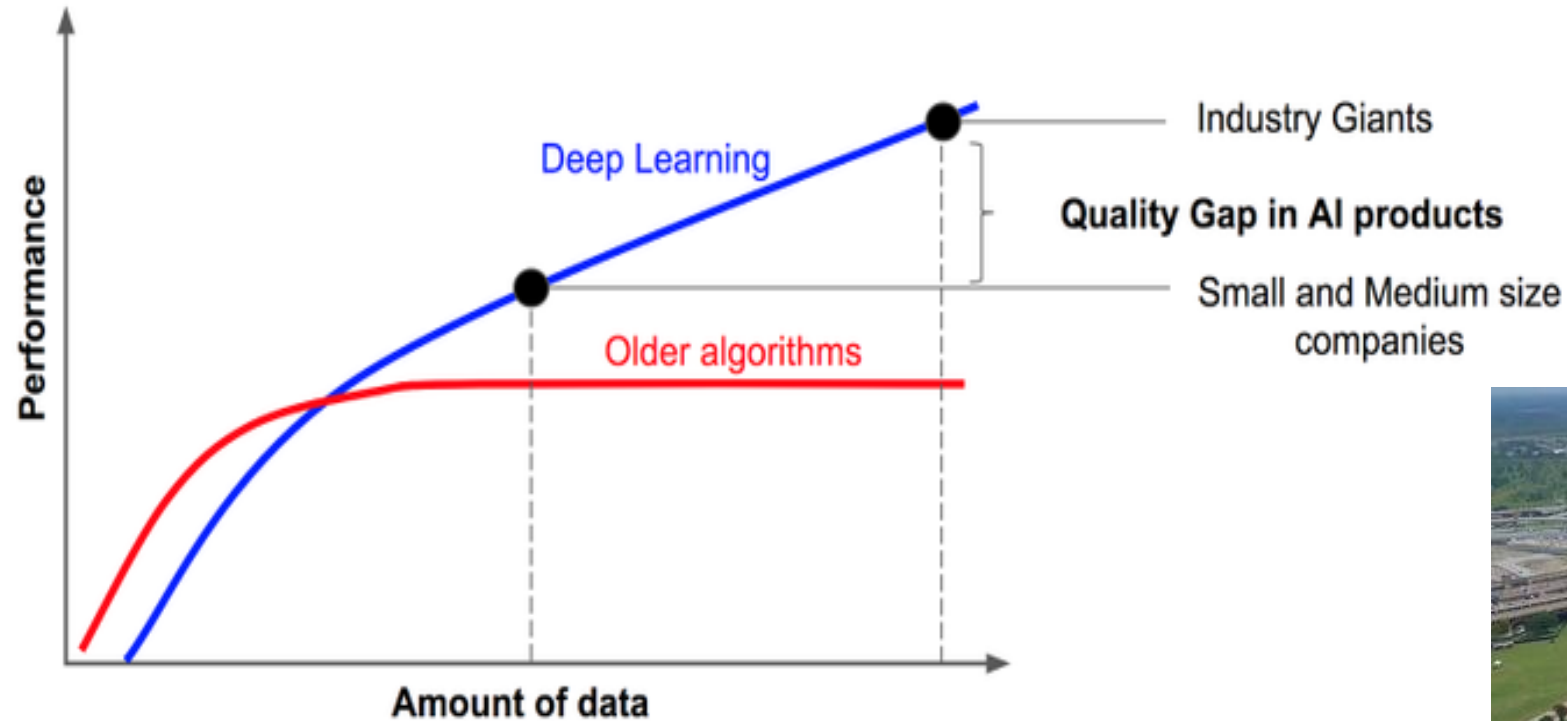
Principe

Comment ça fonctionne



Principe

Le Deep Learning nécessite de très grands jeux d'entraînement de haute qualité.



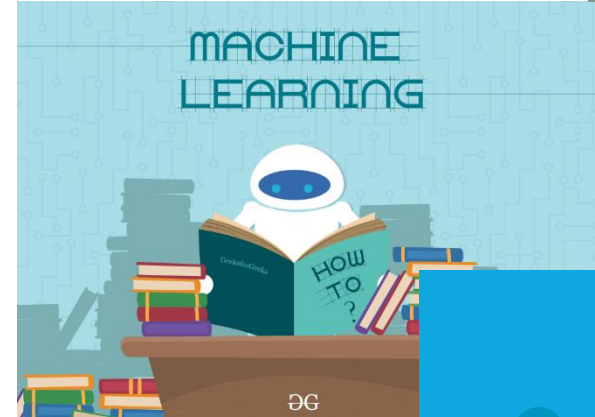
1.065 Km² Building



146.817 Km Road

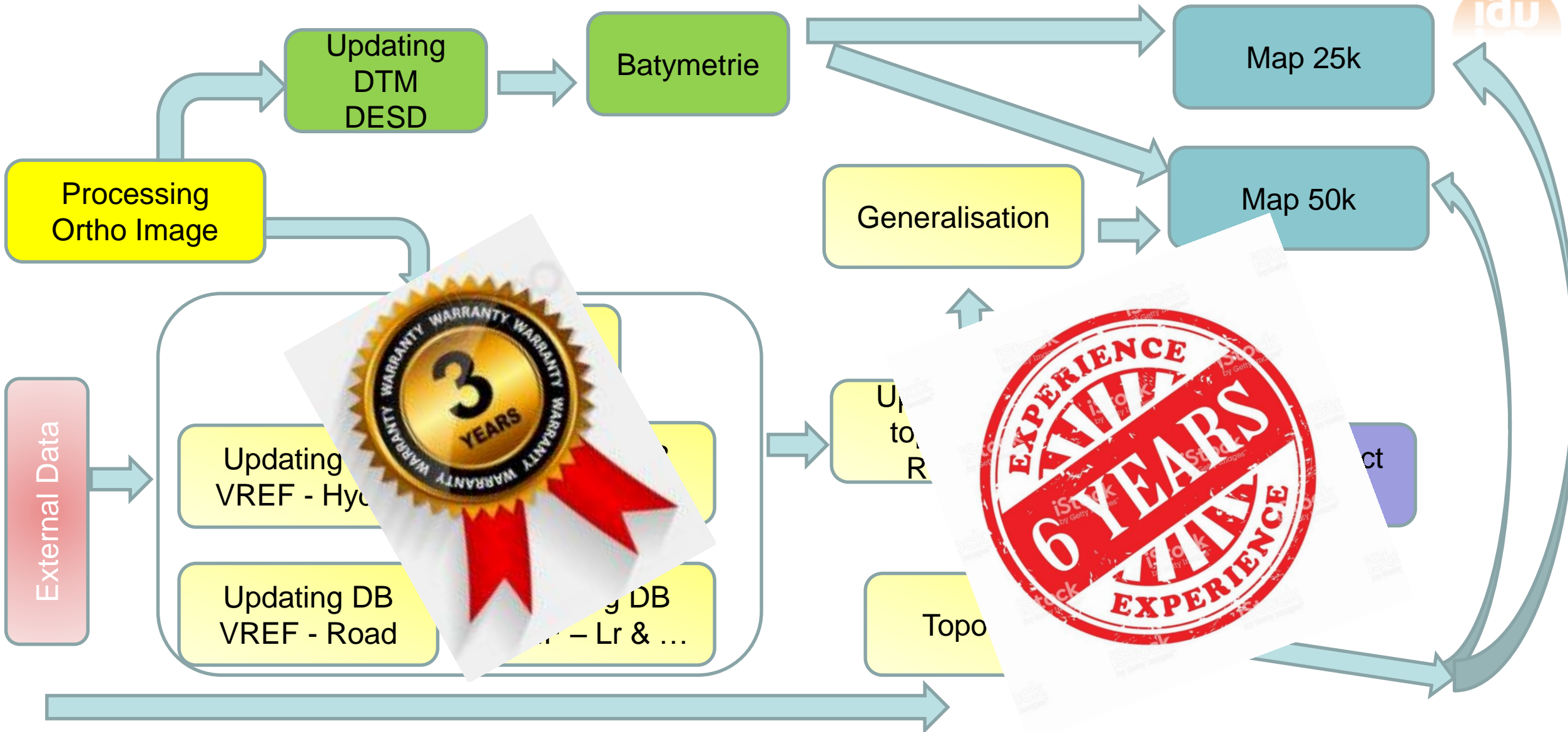
Objectif (IGN)

- Se familiariser avec les techniques de Deep Learning dans le cadre de la détection de changements.
- Étoffer nos procédures de contrôle qualité
- Adapter outils d'acquisition de données sur base de cycles de mise à jour plus courts



Objectif (IGN)

Notre projet => Construire des modèles sur des layer



Objectif (IGN)

Notre projet => Construire des modèles sur des layer



Brunnel	Brunnel
Bâtiment	Gebouw
Construction particulière polygonale	Polygonale bijzondere constructie
Géométrie polygonale supplémentaire	Bijkomende polygonale geometrie
Plan d'eau	Waterloopoppervlak
Segment de cours d'eau	Waterloopsegment
Surface de cours d'eau	Watervlak
Zone d'occupation du sol sur la surface terrestre	Bodembedekkingszone op het aardoppervlak
Segment de chemin de terre	Aardewegsegment
Segment de route	Padsegment
Segment de voie non carrossable (sentiers)	Wegsegment

Objectif (IGN)

Notre projet => Construire des modèles sur des layer



■ Chaque couche présente une richesse d'attribut classifiable visuellement

• Forme du bâtiment

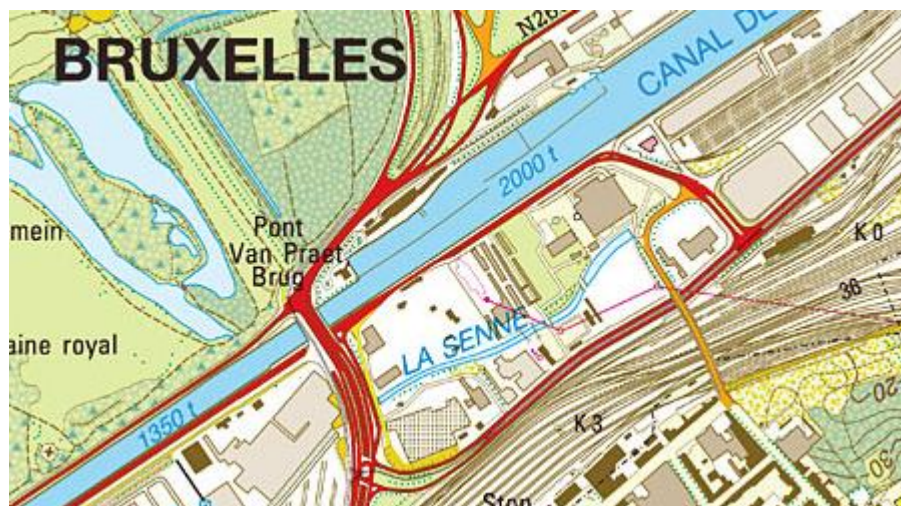
Classification qui indique le type de **bâtiment** sur base de ses caractéristiques extérieures.

- (1) Bâtiment fortifié **TOC**
- (2) Château **TOC**
- (3) Abbaye **TOC**
- (4) Bâtiment d'église **TOC**
- (5) Petite chapelle **TOC**
- (6) Moulin à vent **TOC**
- (7) Bâtiment d'architecture industrielle **TOC**
- (8) Structure gonflable **TOC**
- (9) Serre **TOC**
- (10) Château d'eau **TOC**
- (11) Autre tour

• Largeur d'un cours d'eau

Largeur d'un **segment de cours d'eau**.

- (1) largeur $\leq 1m$
Le **segment de cours d'eau** a une largeur maximale d'1 mètre.
- (2) $1m < \text{largeur} \leq 2m$
Le **segment de cours d'eau** a une largeur supérieure à 1 mètre et inférieure ou égale à 2 mètres.
- (3) $2m < \text{largeur} \leq 3m$
Le **segment de cours d'eau** a une largeur supérieure à 2 mètres et inférieure ou égale à 3 mètres.

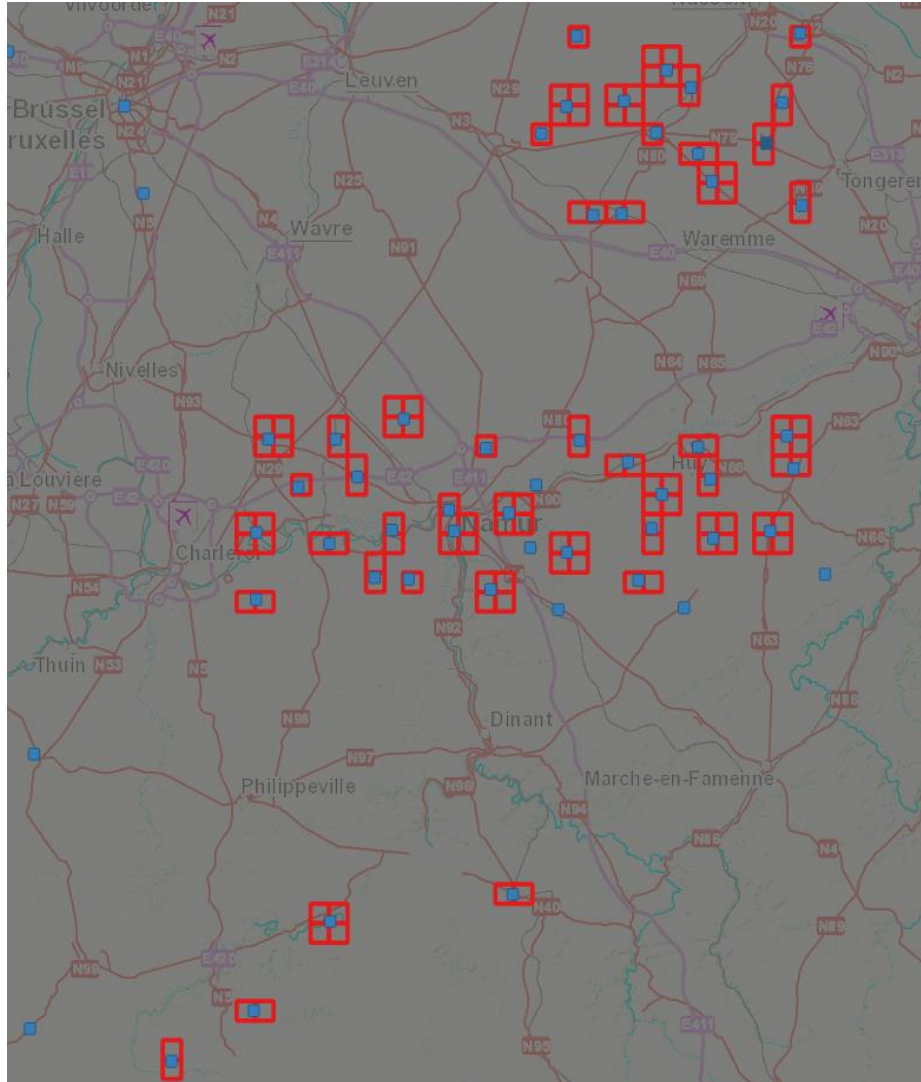


• Occupation du sol

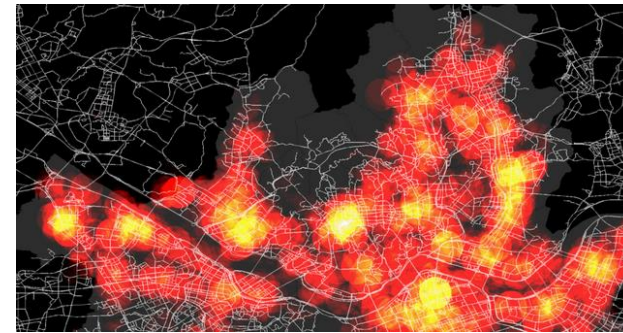
Classification qui indique quelle **occupation du sol** on trouve dans une **zone d'occupation du sol**.

- (3) Couvert mixte de feuillus et conifères sans dominant (FS)
- (7) Pépinière ou oseraie (Os)
- (8) Verger (V)
- (9) Broussailles (B)
- (17) Prairie permanente ou pré de fauche (P)
- (18) Pelouse (G)
- (21) Terrain sablonneux (Z)
- (22) Rochers (Rx)
- (51) Conifères (S) ou Couvert mixte à conifères dominants (Sf)
- (52) Feuillus (F), Couvert mixte à feuillus dominants (Fs) ou Peupleraie (Pp)
- (53) Lande (La) ou Lande associée à d'autres végétations.

Données Sources



Sampling



47 zones de 1km²
+ 118 ortho

Données Sources

QC terrain



- ☒ Layers
 - ☒ RO_RoadSegment
 - ☒ RO_DirtRoadSegment
 - ☒ RO_PathSegment
 - ☒ RO_ConnectingRoadSurface
 - ☐ HY_WatercourseSurface
 - ☐ HY_WaterSurface
 - ☒ HY_Wetland
 - ☐ HY_WatercourseSegment
 - ☐ CO_Brunnel
 - ☐ CO_ParticularPolyConstruction
 - ☒ CO_Building
 - ☒ LC_GroundLevelLandcoverZone
 - landcover
 - ☐ <Null>
 - ☐ 3
 - ☐ 5
 - ☐ 7
 - ☐ 8
 - ☐ 9
 - ☐ 15
 - ☐ 17
 - ☐ 18
 - ☐ 21
 - ☐ 22
 - ☐ 51
 - ☐ 52
 - ☐ 53
 - ☒ ExtractZone_Buffer



- Préparation des données vectorielles
 - 1 homogénéisation de nom des layer
 - 2 extractions de chaque DB et fusion des layer identiques dans une seule DB
 - 3 nettoyages des attributs : Réseau routier => éléments souterrain, gués, ...
 - 4 spatialisations des vecteurs linéaires
 - 5 exports en geojson.

Données Sources



■ Préparation des rasters.

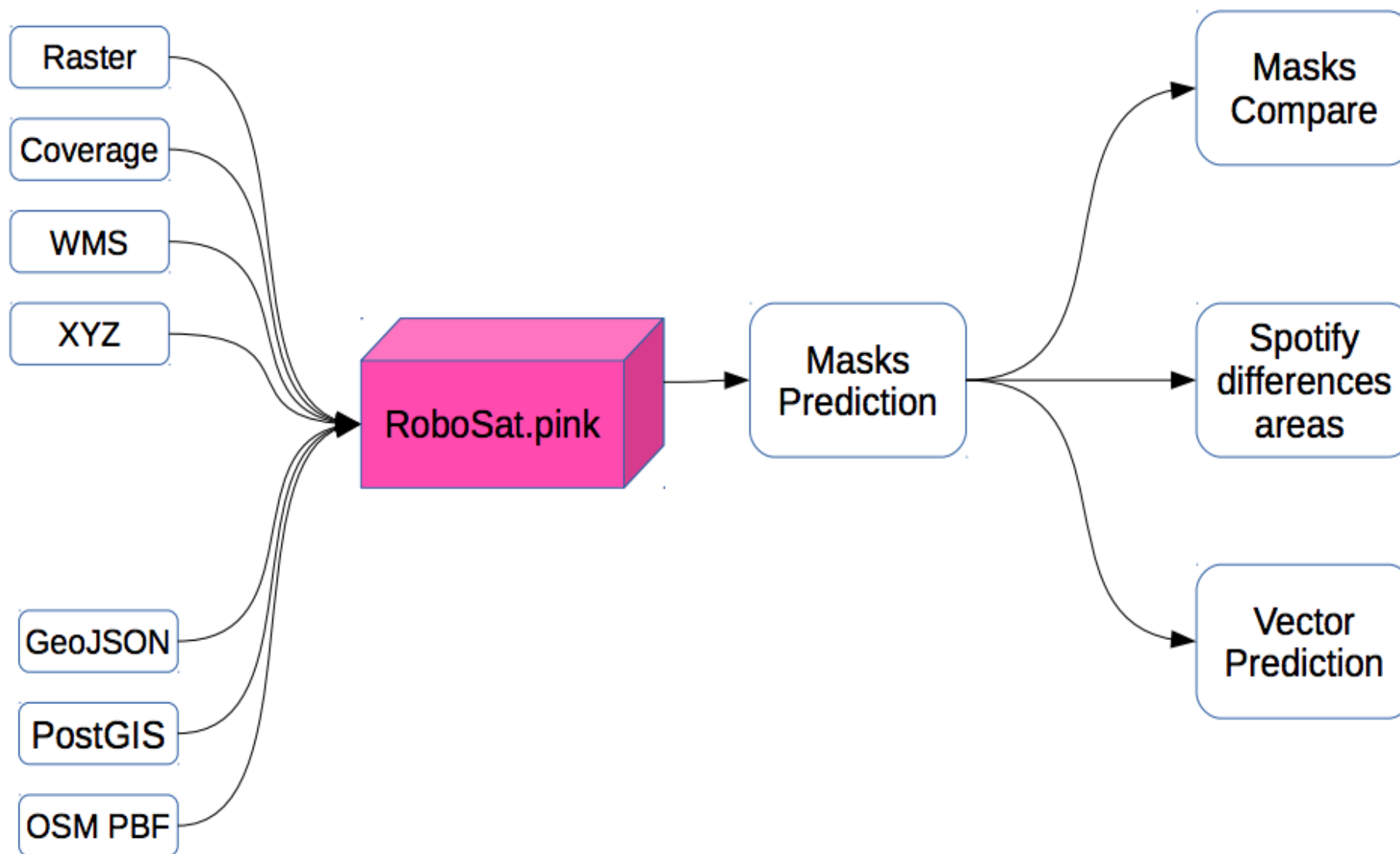
- 1 Générer liste tuile utile (gdal)
- 2 Filtrer les RGB & NIR
- 3 Tester la « conformité » : CRS disponible (gdalsrsinfo)
Correction (gdal_edit) (sur serveur Labo).

■ Outils Deep Learning

■ [RoboSat.pink](#)

Initialement créé dans le giron de MapBox, puis forké pour permettre la segmentation sémantique sur images aériennes et satellitaires et extrait des éléments tels que: bâtiments, parkings, routes, eau, ...

- Outils de ligne de commande pour créer votre propre flux de travail
- Basé sur normes géospatiales afin de faciliter l'interopérabilité et la préparation des données
- Prend en charge les images RGB et multi bandes et permet la fusion de données
- Outils Web-UI pour afficher, hiérarchiser ou sélectionner facilement des résultats
- Déploiement simple et robuste



- Label => objet vectoriel rasteriser servant à l'entraînement
- Images => données raster (ortho -satellite) servant à la détection
- Mask => raster résultant de la détection du modèle
- Modèle => résultat neuronal issue de l'entraînement sur base de différent paramètre

8 étapes :

- Fichier de configuration : xyz.toml (< 1 minute)
 - Permet de définir quel algorithme utilisé et les paramètres utiles
 - Les images sources et les bandes
 - Définir ce que le modèle va apprendre (bâtiment – route – land cover)
- Tuilage : rsp tile (~ 1-2 heure 25 GB - zoom 19 – CPU 2 cores - 12 GB)
 - Permet de générer des tuiles des ortho (512x512 px)
- Définition de la zone d'étude => rsp cover (< 1 minute)
 - Permet (très facilement) d'ajuster l'extent des labels avec ceux des images
- Rastérisation => rsp rasterize (< 5 minutes)
 - Transforme les objets vectoriels en image pour comparer des objets de nature semblable

- Préparation de subset de données => rsp subset (< 1 minute)
 - Prépare des lots d'image aléatoires pour l'entraînement et la validation
- Entraînement => rsp train (~30 minutes – 1 GPU 8GB/ epoch)
 - En deux temps :
 - Sur base d'un lot de données ou le label identifie (correctement) la présence de ce qu'on doit trouver dans les images => **phase d'apprentissage**
 - De ce qu'il créer comme lien – corrélation – affinité, il le teste avec les labels => si c'est bon il garde c'est « connexion neuronale » sinon il rejette
=> **phase de validation**

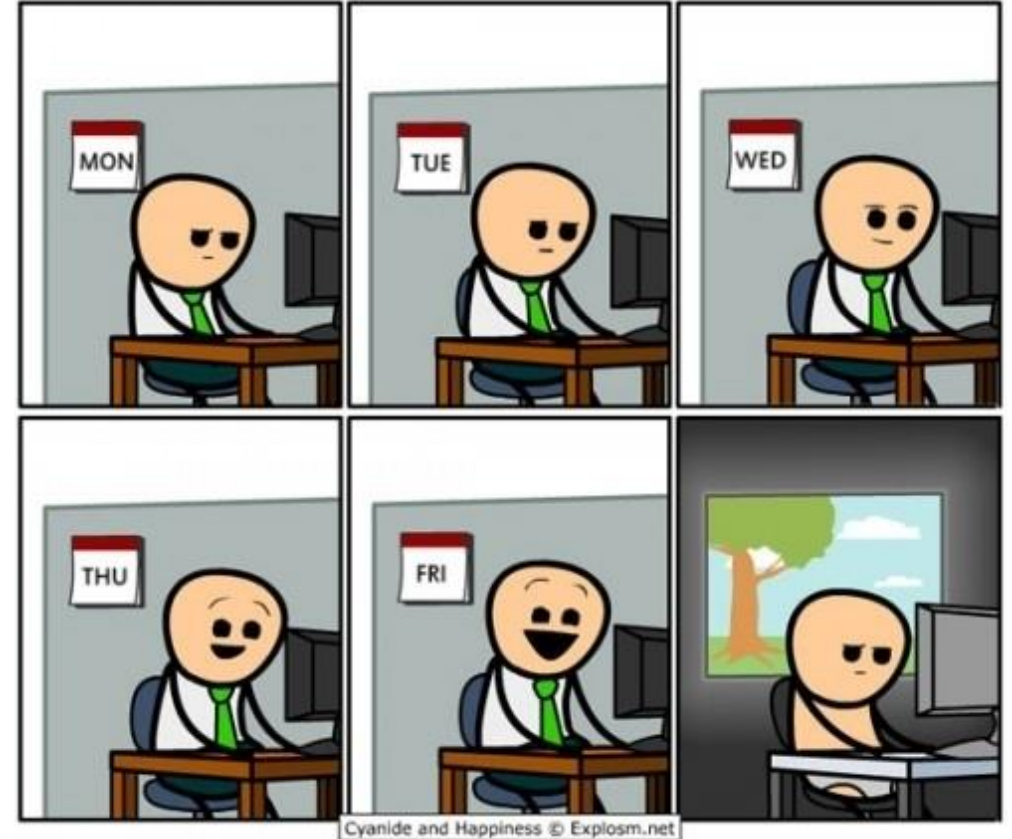
- Détermine le nombre de cycles (epoch)
- Résultat disponible dans les logs

```
---  
Epoch: 43/55 -- UUID: 44282ea2-f050-11e9-b316-fa163e68d5b7  
Loss: 0.013  
build_east iou 0.888  
[Saving checkpoint]  
Loss: 0.016  
build_east iou 0.881  
---  
Epoch: 44/55 -- UUID: c84fd2e4-f054-11e9-b316-fa163e68d5b7  
Loss: 0.013  
build_east iou 0.888  
[Saving checkpoint]  
Loss: 0.016  
build_east iou 0.880
```

Outils



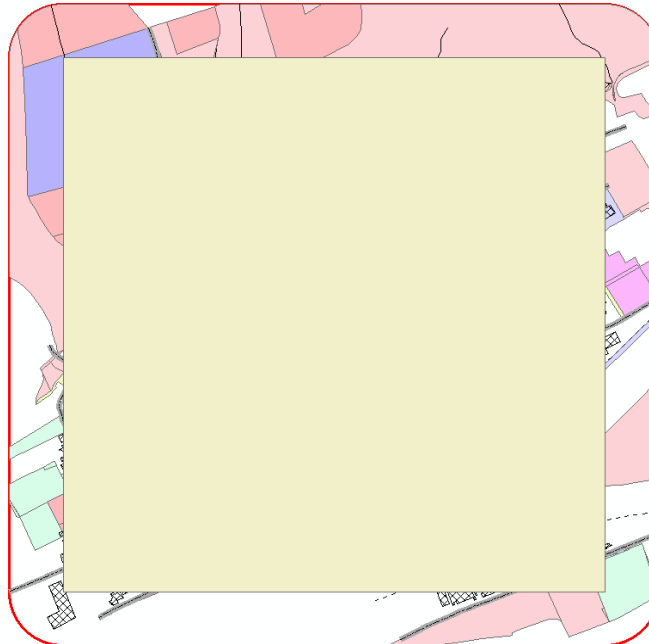
- On teste le modèle en live => rsp predict (~15 minutes /50 km²)
 - Fournis les Mask
 - Il existe plusieurs modes :
 - **Compare** : permet de visualiser les images & label & Mask
 - **List** : visualise uniquement les données litigieuses
- Récupération des résultats => rsp vectorize (< 10minutes)
 - Fournis un geojson



Outils

Premières impressions :

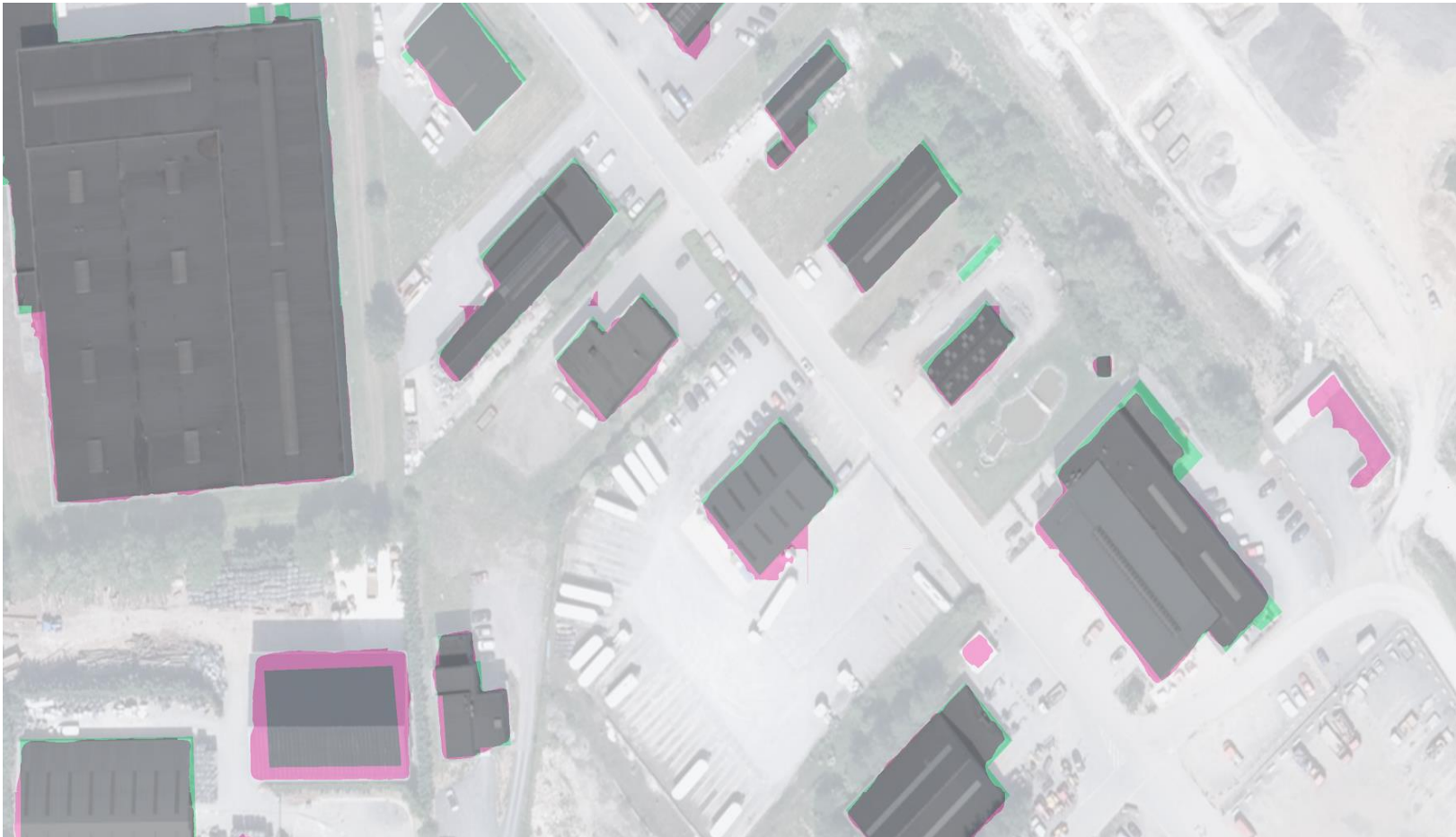
- Outils identifie rapidement si les sources de données sont bonnes ou pas
 - Manque quelques précisions sur les messages d'erreur
 - Piste d'optimisation : Permettre d'exclure les données litigieuses
- Garbage IN
 - Garbage OUT



Résultats



- <http://54.37.151.54/rsp/east/build/compare/> Bâtiment (+/- 0.85)



Résultats

- <http://54.37.151.54/rsp/east/build/compare/> Bâtiment (+/- 0.85)



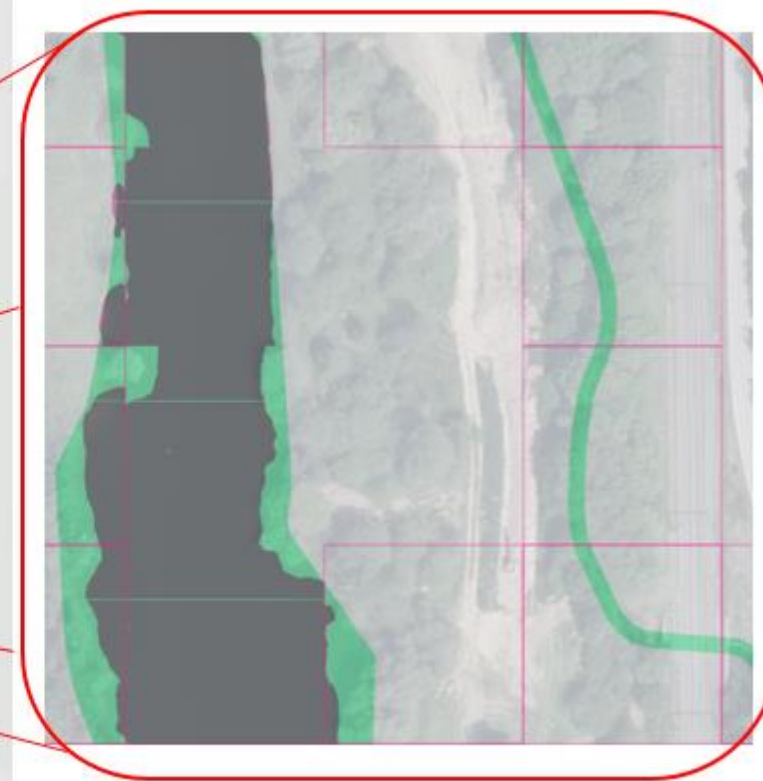
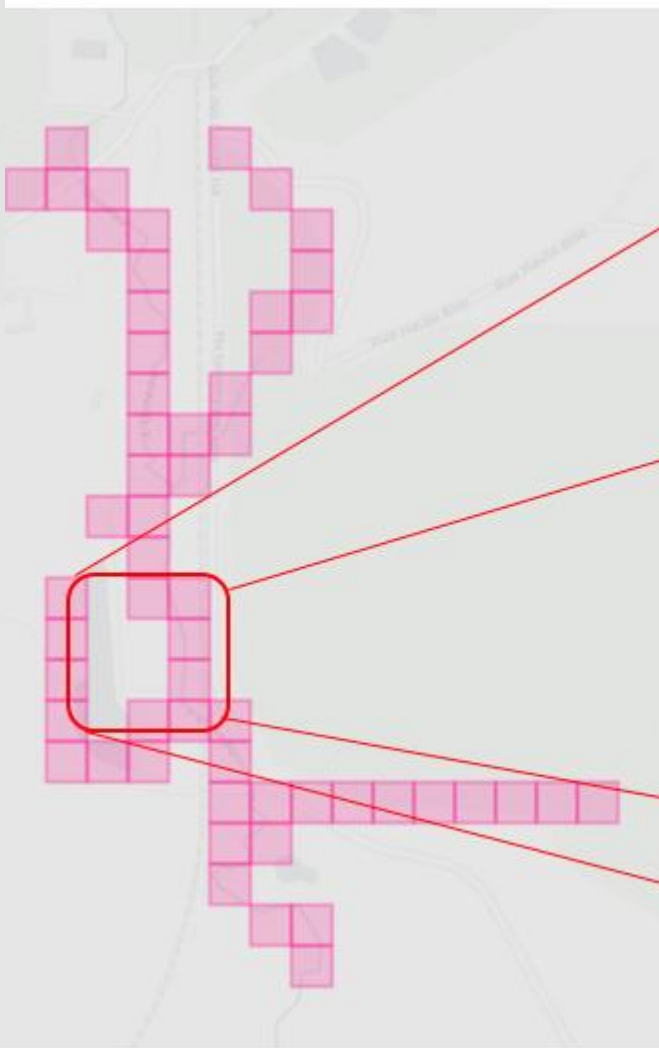
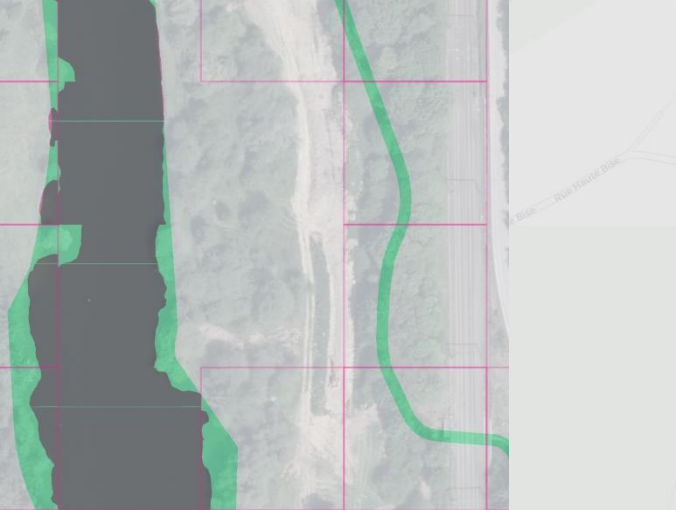
Résultats



- Mask de route : http://54.37.151.54/rsp/east/road/masks_val_road/



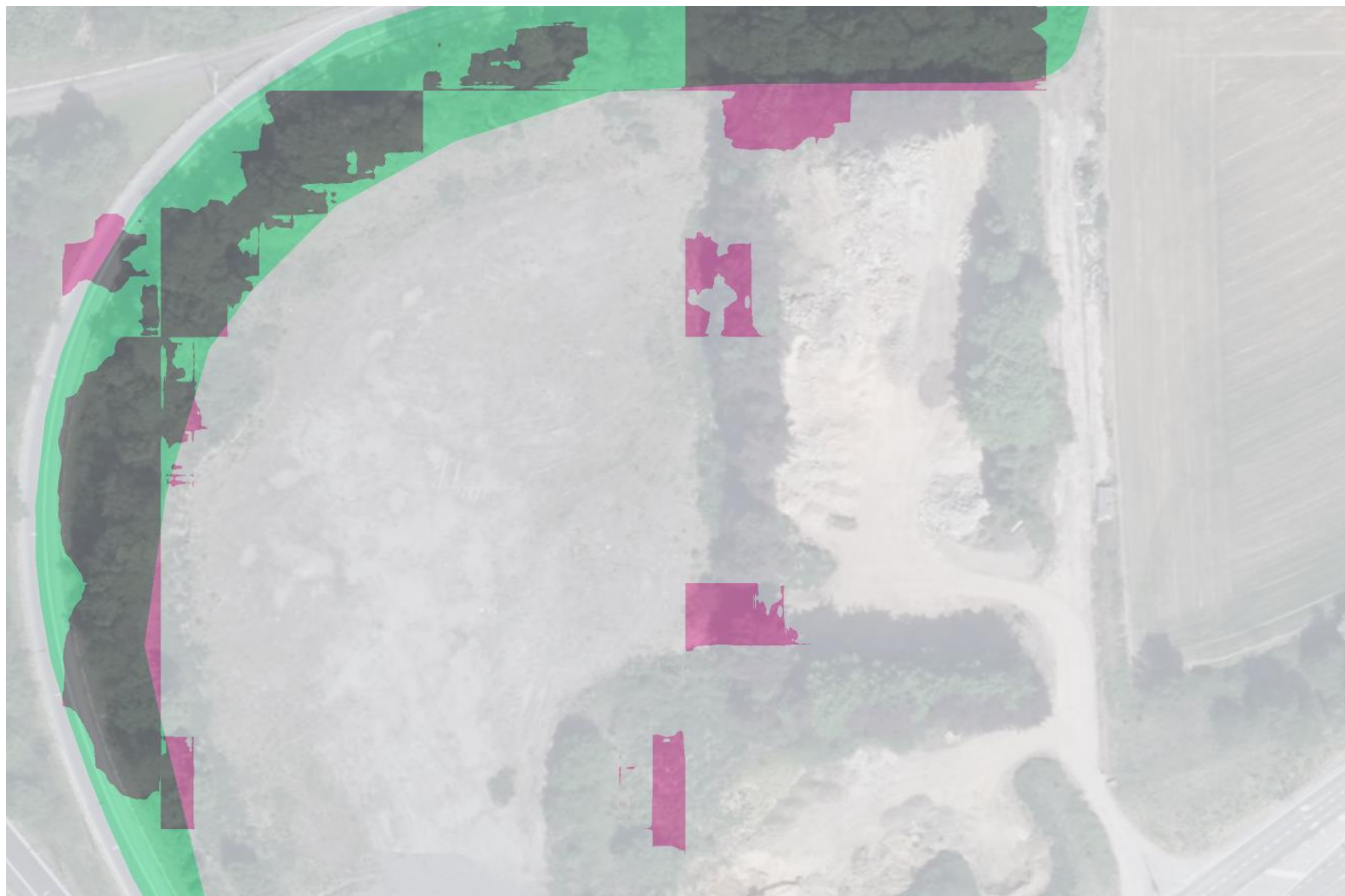
<http://54.37.151.54/rsp/east/hydro/compare2/>



Résultats



- Land Cover : feuillus <http://54.37.151.54/rsp/east/feuillus/compare/>



Conclusion



- Ressources internes à l'IGN sont pertinentes pour acquérir des modèles
 - ⇒ Actuellement préparation importante est nécessaire + une mobilisation de process contraignant et peu efficace.
 - ⇒ Quelques calculs statistiques (Label/Mask) utiles pour améliorer la performance
- Génération modèle et application rapide à mettre en place.
 - ⇒ Sur base d'une zone d'entraînement de 50 km² il faut 8 heures par thème pour construire un modèle
 - ⇒ En extrapolant, pour le **Q**uality **C**ontrol d'un planche du 1/50.000, => +/- 3h
- Prochainement intégration dans les Workflow de production
 - Définir les zones d'échantillonnage pour QC
 - Valider prioriser les demandes de correction de Cartoweb
 - ⇒ Change Détection entre « Ortho 20XX » et ITGI

Perspective



■ Améliorer la sélection des DataSet

- Affiner les attributs
- Exclure les recouvrements

■ Intégration de ressources complémentaires

- Ajout des images NIR
- Ajout du nDSM

■ Optimisation :

- Des flux entre source des données – prétraitement – entraînement,
- Du hardware pour diminuer génération des caches de tuiles – temps d'entraînement

THANK YOU
ANY QUESTIONS?

