

Usage de l'imagerie Satellitaire et des algorithmes de Deep Learning au service du Recensement de la population dans les DOM

Raya Berova

Insee

raya.berova@insee.fr

Gaëtan Carrere

Insee

gaetan.carrere@insee.fr

Thomas Faria

Insee

thomas.faria@insee.fr

Clément Guillo

Insee

clement.guillo@insee.fr

Tom Seimandi

Insee

tom.seimandi@insee.fr

2024-05-24

Abstract La crédibilité des chiffres produits par l’Insee est souvent remise en cause dans les départements d’Outre-Mer. Les chiffres produits par le recensement de la population sont notamment fortement critiqués en Guyane et à Mayotte par les élus locaux ce qui incite l’Insee à travailler sur des sources de données innovantes pour soutenir le discours porté par sa production statistique.

L’utilisation de l’imagerie satellitaire permet d’une part de compléter les estimations produites par l’Insee, en alignant l’évolution du bâti observée sur les images aux évolutions des estimations de population produites par le recensement de la population et d’autre part de soutenir l’opération de repérage des logements sur le terrain en anticipant en amont *via* les images les zones où les mouvements de création ou de destruction sont les plus conséquents.

Des algorithmes de *deep learning* entraînés sur ces images parviennent à détecter automatiquement les contours du bâti dans les DOM de manière très précise. L’algorithme présentant les meilleures performances a été enveloppé dans une application web à destination des agents en bureau, leur permettant de prendre des décisions à partir des zones exhibées par l’algorithme et des images satellites brutes.

La chaîne de traitement dans son ensemble allant de la récupération des images jusqu’à l’application web d’aide à la décision, en passant par l’entraînement des algorithmes de *deep learning*, requiert une pluralité de compétences et une forte technicité pour pouvoir être maintenue.

Table des matières

Introduction	3
1 Contexte	4
1.1 L'enquête cartographique et le recensement de la population	4
1.2 Soutenir les chiffres du recensement de la population à Mayotte	4
1.3 Structure globale du projet	5
2 Préparation des données	8
2.1 Les images satellites	8
2.2 Les annotations	9
3 Entraînement des algorithmes	14
3.1 Les modèles de <i>deep learning</i>	14
3.2 Réseaux de neurones convolutifs	14
3.3 Modèles de segmentation	18
3.4 Description de l'entraînement réalisé	19
4 Analyse et mise en forme des sorties	21
4.1 Traitement des polygones	23
4.2 Buffering	24
4.3 Différences de bâti d'une année à l'autre	26
4.4 Nettoyage de la soustraction	27
5 Pistes et évolutions	30
5.1 Stack technique et dette technique	30
5.2 Suite des travaux et besoins	32
Bibliographie	33

Introduction

En Guyane et à Mayotte, les chiffres du recensement de la population sont souvent remis en question par les élus et la population. La confirmation des évolutions observées par l'Insee par des sources externes est donc cruciale. Ce papier vise à présenter les travaux réalisés à l'Insee au sein du projet « *données satellites* » et à expliciter l'apport potentiel des données d'observation du sol. Le projet consiste en l'utilisation d'images satellites pour reconnaître, grâce à de l'intelligence artificielle, la position des logements, ceci afin de diriger plus finement les moyens humains déployés pour l'enquête cartographique et soutenir les estimations produites par l'opération de recensement de la population officielle de l'Insee.

L'équipe projet est constitué d'un agent basé à la Direction Inter-régionale Antilles-Guyane et de trois agents basés au Département de la Méthodologie à la Direction Générale. Ces travaux ont été initiés il y a près d'un an. Les avancées présentées dans ce qui suit sont issues de périodes de travail réalisées de manière saccadée par les membres de l'équipe qui ne peuvent y accorder qu'une partie limitée de leur temps de travail. Deux excellents stages ont été réalisés à l'été 2023, l'un à la Direction Inter-régionale Antilles Guyane, traitant de la détection de logement à partir d'images satellites de très haute résolution Berova (2023), l'autre au Département de la Méthodologie, s'appuyant sur des images de haute résolution Nabec (2023). Durant l'été 2022, un stage a été réalisé sur le sujet plus général de la couverture du sol Chabennet (2021) et a permis une première prise en main du vaste ensemble d'outils mobilisés dans ces travaux.

Ce projet mobilise des compétences variées et sa complexité demande une formation générale sur des sujets différents : manipulation d'images satellites, compréhension de la structure d'une image d'un point de vue informatique, maîtrise des outils de *deep learning* et des méthodes d'entraînement d'algorithmes, documentation à propos de l'avancée de la recherche en intelligence artificielle, compétences de traitement et de mise à disposition des résultats *via* une application... Nous tenterons donc d'aborder et de détailler tous les aspects de ce chantier. Dans un premier temps, nous allons rappeler le contexte de la mise en place de ce groupe de travail autour des données satellites, en dessinant précisément les besoins auxquels il aspire à répondre. Nous nous pencherons dans un second temps sur les données disponibles, les images satellites et leur traitement en amont de toute analyse. Nous rappellerons ensuite le principe des algorithmes de segmentation en *deep learning*, en justifiant d'une part en quoi ils sont pertinents dans le cadre de la détection du bâti, d'autre part en présentant la manière dont nous les avons appliqués à ce problème et à ces données. Cela nous permettra ensuite de détailler l'entraînement des algorithmes et de présenter les prédictions obtenues. Ces prédictions n'étant pas immédiatement exploitables, nous présenterons les traitements effectués *a posteriori* permettant de les rendre utilisables. Ce papier présentera enfin les pistes d'amélioration et de poursuite du projet, ainsi que les perspectives d'utilisation futures.

1 Contexte

1.1 L'enquête cartographique et le recensement de la population

Chaque année, le recensement de la population (RP) mobilise une centaine d'agents recenseurs dans les Départements d'Outre-Mer (DOM). Les enquêteurs viennent évaluer les logements chaque année au 1er janvier. Avant cette phase de collecte, le répertoire des immeubles localisés (RIL) dans les DOM est mis à jour grâce à une enquête cartographique réalisée en amont. Ce répertoire doit contenir une liste exhaustive des logements géolocalisés dans les DOM, parmi lesquels sont sélectionnés les logements à enquêter pour le recensement de l'année en cours. Cette enquête est spécifique aux DOM, car les bases administratives habituellement disponibles en métropole ne sont pas suffisamment fiables pour alimenter seules ce répertoire.

Un RIL de qualité permet aux enquêteurs de localiser plus facilement les logements à enquêter une année donnée. Le calcul de la population dépend du nombre de logements présents dans le RIL. En effet, cette estimation est le produit du nombre moyen de personnes par logement, obtenu via le recensement, et du nombre de logements comptabilisés. L'impact d'un bon RIL sur la qualité des estimations produites par l'Insee est donc considérable.

L'enquête cartographique se déroule d'avril à août chaque année et mobilise près de 100 enquêteurs dans les DOM, ce qui en fait une opération coûteuse. L'utilisation de l'imagerie satellite, et notamment des méthodes de détection de logements sur ces images, pourrait par exemple améliorer l'organisation de l'enquête cartographique en optimisant notamment le temps de travail des enquêteurs du RP.

Le territoire Domien est découpé en cinq groupes de rotation homogènes et, chaque année, un seul groupe est recensé. Cela signifie que les territoires concernés n'ont pas été recensés depuis cinq ans, période durant laquelle des évolutions notables peuvent survenir dans le parc immobilier, en particulier dans des zones comme Mayotte où des bidonvilles peuvent apparaître et disparaître très rapidement sous l'action des pouvoirs publics. Ainsi, le calibrage du temps de travail des enquêteurs sur ces zones dépend d'observations datées de 5 ans ce qui implique un fort risque de sous-estimation du temps de collecte nécessaire pour une zone donnée.

1.2 Soutenir les chiffres du recensement de la population à Mayotte

Comme l'ont exprimé auprès de l'équipe projet le Directeur interrégional de la Réunion-Mayotte et la Cheffe du service régional de Mayotte, ces problématiques sont d'autant plus ancrées sur le territoire de Mayotte. Répondant à l'injonction des élus de Mayotte (Loi égalité réelle Outre-Mer de 2017), l'Insee a fait évoluer la méthodologie de collecte du recensement en mettant fin aux recensements généraux sur l'île. Cette méthodologique a été remplacée par les enquêtes annuelles de recensement (EAR) avec une première édition en 2021. Cette décision s'est traduite par une période relativement longue sans actualisation des données du recensement et nous avons donc besoin de renforcer la robustesse de nos outils. Dans un contexte démographique extrêmement dynamique à Mayotte, attesté par les données d'État Civil notamment, il est indispensable de collecter l'information la plus complète possible pour confirmer et améliorer les données collectées dans le recensement.

Ce travail à partir des données satellitaires s'inscrit donc dans une réelle perspective d'amélioration de la précision du RIL mahorais et d'un meilleur calibrage de la préparation des EAR grâce à un socle d'informations actualisé pour les enquêtes cartographiques. Trois usages sont alors aujourd'hui clairement identifiés concernant Mayotte : superposer le bâti identifié par imagerie satellite à celui collecté dans les enquêtes cartographiques 2021 à 2024 et analyser les écarts potentiels observés afin d'effectuer une vérification des chiffres obtenus ; programmer des enquêtes annuelles cartographiques ultérieures enrichies afin d'évaluer le gain de connaissance capitalisé par cet élargissement d'échantillon pérennisé ; et enfin prédire avant la collecte ce que sera la charge d'enquête cartographique dans les zones enquêtées cinq années auparavant, y compris sur les zones d'habitat précaires puis mettre en place un modèle pour repérer les zones d'évolution du bâti pour identifier les évolutions sur le terrain qui déterminent la charge d'enquête prévisible des enquêteurs.

La comparaison de l'évolution de ces indicateurs de densité obtenus grâce aux images satellites avec le nombre de logements relevés d'après les enquêtes cartographiques permettra d'identifier les zones les plus divergentes pour les expertiser en bureau en amont et d'orienter la collecte en conséquence. Si l'utilisation des densités de bâti calculés grâce aux images satellites apporte des résultats probants en 2024, ils pourront être utilisés pour valider l'ajout dans le RIL de logements qui n'auraient pas été repérés en première instance.

1.3 Structure globale du projet

Schématiquement, on veut être en mesure de détecter automatiquement des changements à partir de prises de vue d'un même territoire à deux dates différentes. Pour ce faire, on entraîne un algorithme capable *in fine* de produire des masques de logements à partir d'une image donnée, c'est-à-dire une couche de polygones représentant les bâtiments habités sur une carte. En analysant les différences entre deux masques de logements produits pour un même territoire à deux moments distincts, on peut essayer d'en déduire les mouvements principaux entre ces deux dates, notamment les créations et destructions de logements. C'est ce que montre le schéma exposé Figure 1. La mesure de ces évolutions peut alors permettre de prolonger les estimations de population passées réalisées par l'Insee et permet aussi de repérer les zones sur lesquelles l'enquête cartographique devra se concentrer.



Figure 1. – Stratégie d'utilisation des algorithmes

L’entraînement de ces algorithmes de repérage des logements nécessite la constitution d’un grand nombre de couples (*images, masques*) où :

- Les images sont découpées en tuiles suffisamment petites pour pouvoir être absorbées par le modèle d’apprentissage profond. Il faut également s’assurer de l’absence de couverture nuageuse qui rendrait l’analyse impossible.
- Les masques sont des tableaux de la même dimension que l’image dessinant la présence de logements par des valeurs 0 ou 1, s’agglomérant en polygones dessinant les bâtiments. Ces masques sont constitués à partir de données provenant de l’Insee, notamment la base de données topographique fournie par l’IGN. Ces exemples sont construits à partir de données passées et serviront à réaliser le repérage sur des prises de vues plus récentes pour lesquelles ces annotations ne sont pas disponibles.

Ainsi, le projet tel qu’il existe aujourd’hui peut être décomposé en plusieurs parties distinctes (représentées en Figure 2):

- Une première chaîne de traitement comprend la constitution des couples (*images, masques*) à partir des données de l’IGN et qui vont permettre de nourrir l’algorithme.
- Une autre chaîne de traitement automatise complètement l’entraînement des algorithmes à partir des données constituées dans la chaîne en amont. L’utilisation de services tels que Argo-workflow et MLFlow permet une mise en production maîtrisée, un historique des entraînements et du contexte et une reproductibilité rigoureuse du contexte de l’entraînement.
- La dernière chaîne de traitement consiste en l’analyse des prédictions réalisées par l’algorithme et la mise à disposition des résultats visuels et statistiques pour les enquêteurs et les agents de l’Insee souhaitant comparer ces prédictions aux chiffres de population établis par l’Insee lors du recensement de la population.



Figure 2. – Articulation des parties du projet

2 Préparation des données

2.1 Les images satellites

Plusieurs sources de données d'images ont été envisagées pour ces travaux. En imagerie satellite, une distinction est faite entre les données de haute résolution et les données de très haute résolution. Dans ce qui suit, on se concentrera sur des données PLEIADES de très haute résolution. Ces données sont produites par la compagnie Airbus et sont récupérées et concaténées par l'Institut Géographique National (IGN). Ainsi, cet organisme nous fournit chaque année une couverture intégrale des territoires antillais.

Deux caractéristiques sont très importantes lorsqu'on s'intéresse aux images satellites. La première est la résolution spatiale, c'est-à-dire la surface couverte par un pixel : plus la résolution spatiale est élevée, plus la surface couverte par un pixel est faible. Pour les données PLEIADES, la résolution spatiale est de 0.5m. Les images PLEIADES récupérées sont dites panchromatiques : ce sont des images en niveau de gris de très haute résolution et des couches RGB avec une résolution supérieure sont également disponibles. Une extrapolation spatiale est réalisée par la suite pour construire une image RGB à une résolution 50 cm. A titre comparatif la résolution spatiale des images Sentinel 2 est de 10 m, donc 20 fois plus faible (voir Figure 3).



Figure 3. – Comparaison entre une image PLEIADES très haute résolution (a) et une image Sentinel 2 haute résolution (b)

La deuxième caractéristique est la résolution temporelle *i.e* la fréquence à laquelle on peut obtenir une photographie du sol d'une zone donnée. La résolution temporelle correspond au nombre de passages qu'un satellite réalise au-dessus d'un territoire donné sur une échelle de temps donnée. Plus cette dernière est élevée, plus les images à disposition seront récentes et de fait pertinentes au regard du cas d'usage souhaité. Les imageries obtenues le sont *via* des mesures optiques, ce qui implique qu'une couverture nuageuse trop élevée lors de la prise de vue satellite retardera l'acquisition pour le territoire concerné. Les territoires des DOM sont sujets à des saisons des

pluies étalées et par conséquent sont très souvent couverts. En moyenne, huit mois sont donc nécessaires pour avoir une acquisition complète sans nuage de ces territoires par prise de vue satellites PLEIADES, ce qui induit un écart entre la réalité du terrain et celle photographiée par les satellites. La Figure 4 présente la couverture totale de Mayotte en 2020 et en 2023 par les images PLEIADES sur lesquelles nous nous concentrerons par la suite.



Figure 4. – Comparaison de la couverture de Mayotte par l'imagerie PLEIADES en 2020 (a) et en 2023 (b)

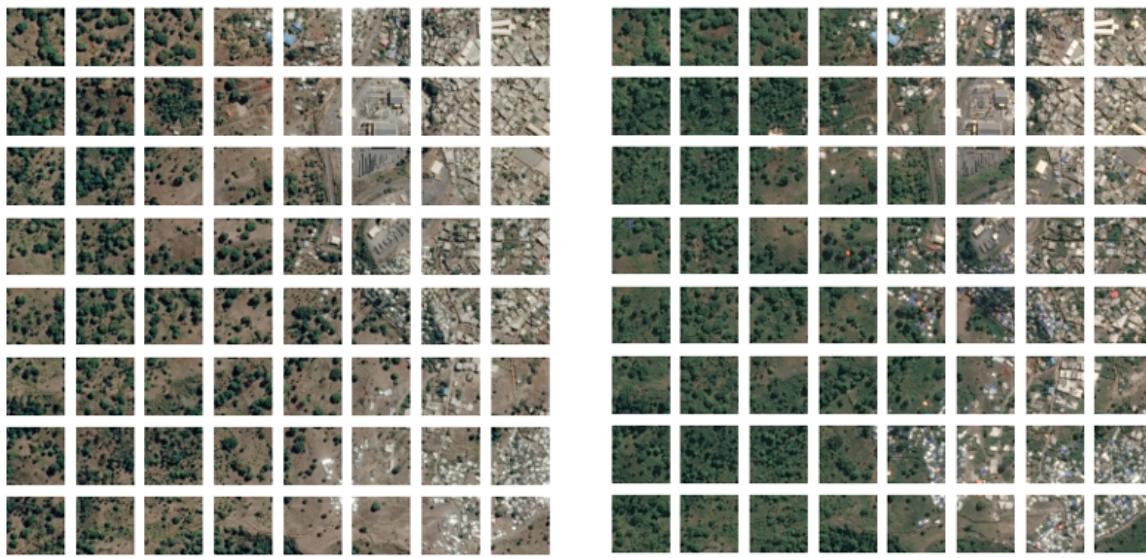
2.2 Les annotations

L'objectif ici est de constituer, à partir d'un ensemble d'images couvrant les territoires des DOM, les masques de logement associés aux images. Il est utile de noter ici que la labellisation manuelle serait la meilleure solution du point de vue de la qualité des masques générés, puisque le décalage entre la date de constitution des bases servant à annoter les images et la date de prise de vue engendrera nécessairement des masques imparfaitement synchronisés avec les images.

Le coût d'une telle labellisation réalisée manuellement est prohibitif au regard de la taille de l'équipe projet et du temps de travail disponible, c'est pourquoi une labellisation automatique a été réalisée à partir de la base de données topographique de l'IGN (BDTOPO). Cette base localise précisément chaque année le contour du bâti avec des polygones et est obtenue par la combinaison de traitements sur photographies aériennes et d'annotations réalisées à la main par des agents de l'IGN.

Cependant, la BDTOPO ne peut pas répondre aux cas d'usages mentionnés en introduction, ceci pour les raisons suivantes :

- La BTOPO millésimée produite une année donnée se veut représenter l'étalement du bâti une année donnée. Or des polygones de contour de bâti une année donnée peuvent apparaître entre deux versions différentes sans que cette évolution de la BDTOPO soit rattachée à une réelle création : cette situation peut être rencontrée si les méthodes de détection des logements de l'IGN s'affinent d'une année sur l'autre et qu'un logement construit auparavant finit par être détecté.
 - Un territoire donné est couvert par des tuiles d'images satellites obtenues au fur et à mesure des passages du satellite au-dessus des territoires et dépendent de la couverture nuageuse et de l'inclinaison des radars optiques. Elles vont donc être obtenues à des moments différents de l'année. De fait, la représentation d'un territoire par imagerie satellitaire est une mosaïque d'images obtenues à des instants différents. Il est donc difficilement envisageable que la BDTOPO coïncide parfaitement avec ces images. Elle ne peut donc pas être utilisée telle quelle. Ce découpage est mis en exergue dans la Figure 5.
 - L'actualisation de la BDTOPO par l'IGN n'est pas garantie et l'Insee doit donc internaliser ce processus de détection de bâti.



(e). – Mayotte 2017 (f). – Mayotte 2020
 Figure 5. – Représentation de la couverture par image sous forme de tuiles

Le répertoire des immeubles localisés (RIL) essentiellement constitué de la concaténation des enquêtes cartographiques des années précédentes répond à plusieurs des critiques adressées à la BDTOPO. En effet, le RIL est alimenté chaque année par l'enquête cartographique, durant laquelle les enquêteurs relèvent les créations ou suppressions de logements et précise éventuellement la localisation des logements déjà existants. Ce travail est réalisé pour 1/5ème des logements du territoire, ce cinquième correspondant aux logements qui seront recensés par la suite pour l'enquête annuelle de recensement de l'année donnée. L'enquête cartographique est réalisée chaque année

entre mai et août ce qui facilite la datation du RIL dans la mesure où ce travail de collecte des positions des logements est correctement réalisé.

Cependant les logements se voient attribuer un seul et unique point par l'enquêteur (celui de la porte d'entrée quand c'est possible) ce qui ne peut égaler la précision des contours fournis par l'IGN. Pour approximer le concept de surface de logement, il faut alors calculer des buffers autour de ces points. Une comparaison entre un masque produit *via* le RIL et un masque produit *via* la BDTOPO sur un même territoire est présentée dans la Figure 6.



Figure 6. – Masque de logements produit à partir du RIL (a) et à partir de la BDTOPO (b)

Le RIL n'étant disponible que pour un cinquième du territoire, la constitution annuelle du jeu de couples (*images, masques*) demanderait un travail considérable dans la mesure où ce découpage du territoire en cinq groupes n'est pas une grille carrée se superposant au découpage en tuiles. Cette division est basée sur le zonage administratif par îlots de l'Insee, îlots dont les contours suivent les limites urbaines et naturelles (voir Figure 8). En outre, les zones d'habitats informels, très largement rencontrées en Guyane et à Mayotte ne sont pas correctement référencées dans le RIL, ce que montre la Figure 7.



(i)

(j)

Figure 7. – Un masque du mont Baduel obtenu via le RIL (a) et celui obtenu à partir de la BDTO-PO (b)



Figure 8. – Découpage en îlots de Mayotte

Une piste de travail avortée faute de moyens avait été explorée et consistait à combiner le concept de logement délivré par le RIL et la précision des contours obtenus via la BDTOPO. Une labellisation manuelle pourrait également être envisagée pour améliorer la qualité des masques produits et se rapprocher au mieux du concept de logement mais celle-ci est très chronophage et il est

très difficile d'anticiper les gains de performance qui seraient enregistrés en entraînant un algorithme sur ces masques obtenus manuellement. De plus, dans de nombreux cas, il n'est pas aisément de s'accorder visuellement (même au sein d'une petite équipe) sur ce qui est du logement ou non, c'est pourquoi des règles de décision strictes et couvrant l'ensemble des situations visualisables devraient être produites afin que ces travaux manuels d'annotation ne soient pas trop dépendants de l'interprétation de l'opérateur qui labelliserait à la main.

In fine, une vérification manuelle des masques construits automatiquement s'impose mais cette vérification est également coûteuse et n'a pas pu être effectuée pour tout le territoire. Cependant, pour valider la pertinence des algorithmes dont le processus d'entraînement est détaillé dans ce qui suit, un ensemble d'images jugé par l'équipe comme correctement labellisé à été constitué sur Mayotte.

On retiendra ici que les masques produits automatiquement en utilisant la BDTOPO ne se limitent pas au concept de logement mais à celui de bâti. Ceci est problématique puisqu'un algorithme entraîné sur ces masques ne pourra lui aussi que détecter du bâti. Des travaux futurs devront être menés pour construire des masques de logements.

3 Entraînement des algorithmes

Nous allons dans cette partie aborder les concepts d'apprentissage profond (*deep learning*) et expliquer notre démarche et l'application de ces outils à nos problématiques précises. Néanmoins, pour la bonne compréhension du lecteur, des rappels seront faits sur l'état de l'art et les concepts sous-tendant les méthodes pertinentes dans le cadre de ce projet.

3.1 Les modèles de *deep learning*

En *machine learning*, la sous-classe des modèles de *deep learning* ou réseaux de neurones profonds désigne des modèles dont la conception est inspirée du fonctionnement du cerveau humain. Ces neurones sont seulement capables d'opérations très simples, à savoir qu'ils peuvent s'activer si le signal qu'ils reçoivent en entrée est suffisamment fort et transmettent leur activation dans ce cas. Un grand nombre de neurones correctement organisés permet alors de réaliser des opérations plus complexes, résultantes des différentes activations.

Cependant, leur utilisation n'est pas systématique et la qualité des données utilisées en entrée de ces réseaux est cruciale. En effet, le problème de segmentation est un problème compliqué et nécessite un grand nombre d'images d'entraînement. Par ailleurs, la labellisation de chaque pixel dans les images du jeu d'entraînement est très coûteuse. Les premiers modèles de *deep learning* sont en outre apparus depuis plus de 10 ans, mais malgré leurs bonnes performances, le coût d'entraînement de ces modèles était alors prohibitif. L'approche classique plus parcimonieuse qui consiste à calculer en amont un ensemble de descripteurs X à partir de l'image en entrée puis de construire un modèle à partir des descripteurs extraits était plus pertinente.

Dans la classe des modèles de *deep learning*, les réseaux de neurones convolutifs occupent une place importante car ils sont particulièrement adaptés au travail sur les images.

3.2 Réseaux de neurones convolutifs

Les réseaux de neurones convolutifs tirent leur inspiration du cortex visuel des animaux et se composent de deux parties (voir Figure 9) :

1. Une première partie composée par des couches de convolutions successives qui permet d'extraire des prédicteurs de l'image en entrée du réseau;
2. Une seconde partie permettant de classifier les pixels de l'image en entrée à partir des prédicteurs obtenus dans la première partie.



Figure 9. – Réseau de neurones convolutif

Dans la partie convulsive, l'image en entrée du réseau se voit appliquer des filtres appelés convolutions, représentés par des matrices $A = (a_{ij})$ de petite taille appelés noyaux de convolution. L'image en sortie de l'opération de convolution est obtenue à partir de chaque pixel de l'image en entrée en calculant la somme des pixels avoisinant, pondérée par les coefficients a_{ij} du noyau de convolution. Cette opération de convolution est illustrée dans la Figure 10, extraite de l'ouvrage Kim (2017).

$$\begin{array}{|c|c|c|c|} \hline 1 & 1 & 1 & 3 \\ \hline 4 & 6 & 4 & 8 \\ \hline 30 & 0 & 1 & 5 \\ \hline 0 & 2 & 2 & 4 \\ \hline \end{array} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{array}{|c|c|c|} \hline 7 & 5 & 9 \\ \hline 4 & 7 & 9 \\ \hline 32 & 2 & 5 \\ \hline \end{array}$$

Figure 10. – Exemple d'opération de convolution

Une convolution résume donc l'information contenue dans l'image. Il est d'ailleurs intéressant de remarquer que la convolution engendrera des valeurs en sortie élevées pour des pixels dont le voisinage présente la même structure que le noyau associé à cette convolution, de telle sorte que la forme du noyau donne une indication sur les parties de l'image qui seront mises en évidence par ce dernier. Une « couche » du réseau de neurone convolutif correspond en fait à l'application d'un nombre n_f de filtres convolutifs. Ainsi, en sortie d'une couche, il y a autant d'images que de filtres appliqués. Pour la couche suivante on applique alors des filtres convolutifs sur les images à n_f bandes issues de la couche précédente.

Un réseau de neurones convolutif est alors constitué de plusieurs couches. Les premières couches permettent de détecter des formes simples dans l'image (lignes horizontales, verticales, diagonales etc.) tandis que les couches suivantes, plus spécialisées, vont combiner les concepts simples appris par les couches précédentes et détecter des formes plus complexes, ce qui est schématisé dans la Figure 11.



Figure 11. – Représentation d'une convolution

Des opérations dites de « max pooling » illustrées dans la Figure 12 permettent de réduire la dimension des images obtenues entre chaque couche tout en restant proche de l'information extraite par convolution. Cette réduction de dimension permet de diminuer le nombre de calculs par couche et autorise la construction de réseaux avec un grand nombre de couches souvent plus performants. D'autres opérations telles que le padding détaillé dans la Figure 13 permettent de contrôler la dimension de l'image en sortie.



Figure 12. – Représentation du maxpool

$$\begin{array}{c}
 \begin{matrix}
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 3 & 3 & 4 & 4 & 7 & 0 & 0 \\
 0 & 9 & 7 & 6 & 5 & 8 & 2 & 0 \\
 0 & 6 & 5 & 5 & 6 & 9 & 2 & 0 \\
 0 & 7 & 1 & 3 & 2 & 7 & 8 & 0 \\
 0 & 0 & 3 & 7 & 1 & 8 & 3 & 0 \\
 0 & 4 & 0 & 4 & 3 & 2 & 2 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
 \end{matrix} \\
 6 \times 6 \rightarrow 8 \times 8
 \end{array} * \begin{matrix}
 1 & 0 & -1 \\
 1 & 0 & -1 \\
 1 & 0 & -1
 \end{matrix}_{3 \times 3} = \begin{matrix}
 10 & -13 & 1 & & & \\
 -9 & 3 & 0 & & & \\
 & & & & & \\
 & & & & & \\
 & & & & & \\
 & & & & & \\
 & & & & & \\
 & & & & &
 \end{matrix}_{6 \times 6}$$

Figure 13. – Représentation du padding

Une fonction d'activation (concept proche de la neurobiologie) non linéaire de type ReLU (Rectifier Linear Unit) est appliquée systématiquement après chaque couche (cf. Figure 14). L'utilisation de fonctions non linéaires n'est pas réservée aux seuls réseaux de neurones convolutifs mais à l'ensemble des modèles de *deep learning*. Ces fonctions non linéaires vont permettre *in fine* de former des réseaux de neurones capables de prédire des phénomènes non linéaires. Ces réseaux s'adapteront mieux au traitement des images.



Figure 14. – Fonction d’activation ReLu

Par rapport au cadre d’apprentissage classique en *machine learning*, la spécificité de ce type d’algorithmes réside dans le fait que l’extraction des caractéristiques de l’image en entrée n’est pas réalisée manuellement ou à dire d’expert mais est automatisée lors de l’entraînement du modèle. Les coefficients contenus dans les différents noyaux convolutifs font en effet partie du vecteur θ des paramètres du modèle. Ainsi, une fois le modèle entraîné, le jeu de paramètres optimal θ^* calculé décrit en fait l’ensemble des filtres convolutifs appliqués à l’image. Les descripteurs de l’image ainsi produits par les applications successives de ces filtres sont alors confrontés à la partie classifiante du réseau. Or, dans le cadre classique, les descripteurs de l’image sont obtenus manuellement : il peut par exemple s’agir de convolutions dont les paramètres sont fixés ou bien de statistiques produites à partir des 4 bandes de l’image (moyenne sur l’infrarouge, coefficient de variation, etc..).

En sortie des réseaux de neurones convolutifs classiques, la partie classifiante permet d’attribuer une catégorie à l’image en entrée. On veut, par exemple, savoir si l’image étudiée est une image de chien ou de chat. Pour ce faire, la partie classifiante part de la concaténation dans un vecteur de la sortie de la partie convulsive et retourne, par le biais d’un réseau de neurones dense (ou « *fully connected* »), un vecteur en sortie $x = (x_0, \dots, x_9)$ pour une classification sur 10 classes. Ce vecteur est ensuite transformé en une distribution de probabilité par l’opération de softmax suivante :

$$\text{Softmax } (x_i) = \frac{\exp x_i}{\sum_{j=0}^9 \exp x_j} \quad (1)$$

On classera l’image dans la catégorie ayant la plus forte probabilité calculée. La fonction softmax est infiniment dérivable et permet en fait de préserver la dérivabilité en les paramètres θ de notre modèle f_θ appliqué à une image X . La dérivabilité des fonctions d’erreurs qui seront calculées à partir des prédictions du réseau en découle.

Dans le cadre de la segmentation d’image, la sortie est plus complexe puisqu’on veut obtenir un vecteur de distribution par pixel. Nous présentons dans la suite les modèles qui apparaissent souvent dans la littérature scientifique construits à partir de briques de convolutions.

3.3 Modèles de segmentation

Les algorithmes de segmentation peuvent être vus comme des algorithmes classifiant un par un les pixels de l'image. Dans le cadre de la détection de bâti, un algorithme de segmentation prend donc une image en entrée et attribue pour chaque pixel une probabilité de présence de bâti sur ce même pixel (0,1). Le masque prédit l'est ensuite en seuillant cette probabilité et en classant en logement tous les pixels de l'image tels que la probabilité affichée par le réseau excède ce seuil.

Dans la littérature, les modèles de segmentation sont souvent basés sur une structure en forme de U i.e. composés de :

- Une partie descendante, l'encodeur, qui va permettre de transformer l'image en entrée en un vecteur numérique de taille réduite par rapport au nombre de pixels initial de l'image. La partie encodeur étant un réseau de neurones convolutif classique tel que présenté à la partie précédente.
- Une partie ascendante, le décodeur, qui va partir du vecteur obtenu précédemment (aussi appelé embedding) et remonter par opérations dites de convolutions inverses à une sortie de la même dimension que l'image.

La qualité du processus de segmentation par l'algorithme est très liée à la qualité de l'encodeur dont le but est de réécrire les images dans un espace suffisamment expressif et interprétable par le décodeur.

Les deux parties (encodeur et décodeur) sont paramétrées et sont donc améliorées lors de l'entraînement. Beaucoup de couches finissent par séparer l'image en input du masque produit en sortie. Du fait des opérations de *Max Pooling* successives dont le but premier est d'alléger le nombre de paramètres du réseau, l'information au niveau local se perd à travers les couches du réseau, et l'information vectorisée en sortie de l'encodeur ne retranscrit plus suffisamment les phénomènes locaux, lissés par cette opération d'agrégation. Dans Visin et al. (2016) et Jégou et al. (2017), les auteurs présentent des structures de modèle de segmentation permettant de pallier ce défaut. Dans ces modèles, les éléments en sortie de certaines couches servent d'entrée à plusieurs des couches suivantes. Visuellement, certaines couches sont alors court-circuitées.

Le U-net Ronneberger et al. (2015) pousse cette logique un peu plus loin en allant jusqu'à connecter les couches de la partie contractante aux couches de la partie expansive. La Figure 15 schématisse la structure du U-net. D'autres architectures telles que présentées dans Chen & al. (2017) reposent sur des formes de convolution spécifiques (*atrous convolution*) ayant pour but de minimiser l'effet résumant des opérations de *Max Pooling*. Ces structures sont très lourdes et plusieurs structures du même type mais allégées ont été produites par la suite.

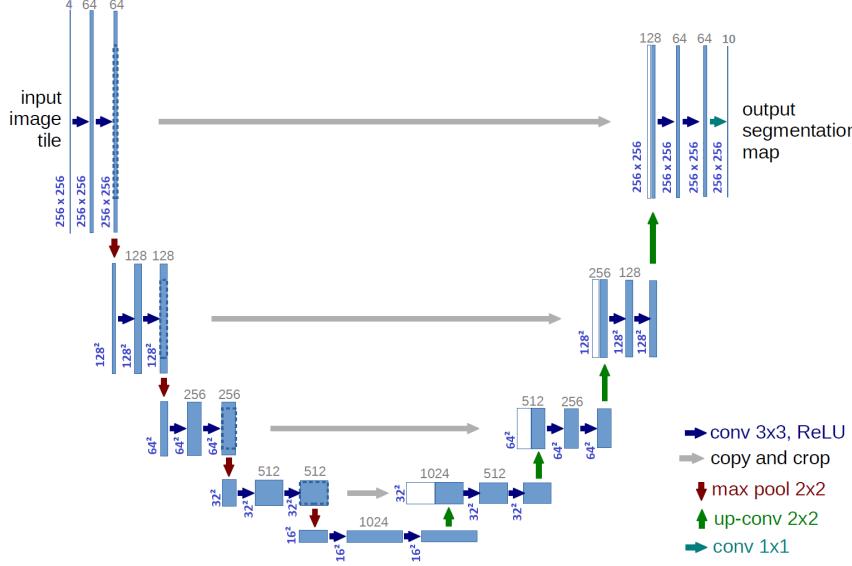


Figure 15. – Représentation schématique du Unet

Plus récemment, la construction des modèles de segmentation s'est beaucoup inspirée de celle des Large Language Model (par exemple chatGPT) dont l'efficacité n'est plus à démontrer. Par analogie avec les séquences de mots, si on considère que les images sont des séquences à deux dimensions alors on peut appliquer des structures de type « transformer » (cf. Vaswani et al. (2017)) sur ces dernières. Ainsi, dans Dosovitskiy et al. (2021) les auteurs montrent qu'on peut entraîner un modèle de classification s'appuyant uniquement sur des transformers (en opposition aux réseaux de neurones convolutifs). Dans Xie et al. (2021), les auteurs généralisent cette approche aux modèles de segmentation en utilisant un décodeur basé sur une structure de tranformer. L'avantage principal de l'utilisation de telles structures est le gain en efficience (performance à nombre de paramètres fixés), qui contraste beaucoup avec les structures basées sur des réseaux de neurones convolutifs présentées précédemment.

3.4 Description de l'entraînement réalisé

Les ramifications possibles du projet sont très nombreuses, en découle un ensemble de choix tentaculaire. L'équipe de projet après une phase d'expérimentation des modèles a donc décidé de produire au plus vite une chaîne de traitement complète en faisant des choix rapides afin de stabiliser une chaîne de production faisant office de *proof of concept*.

Un entraînement d'un réseau de type *Segformer* a été réalisé sur la base des images PLEIADES couvrant la Martinique et la Guadeloupe en 2022. Pour annoter ces images, on se sert des versions disponibles de la BDTOPO produite par l'IGN en acceptant les divergences inévitables liées aux temporalités différentes des prises de vues et de constitution des bases topographiques. Les algorithmes entraînés à partir de ce jeu d'entraînement ne peuvent alors détecter que du bâti. Un modèle de segmentation de type *Segformer* a été entrainé sur les couples ainsi obtenus.

Le choix des territoires et des années sur lesquels l'algorithme sera entraîné repose sur un trade-off entre spécialisation et généralisation : d'un côté il est souhaitable que les territoires sur lesquels

s'entraînent l'algorithme soient de même nature que les territoires sur lesquels on l'évalue, de l'autre il nous faut un jeu de situations suffisamment large pour que nos algorithmes puissent s'adapter à des situations ou structures nouvelles. Entrainer un algorithme par territoire en se servant uniquement des images couvrant ce dernier pourrait conduire à un algorithme très spécialisé mais pas suffisamment général qui ne serait pas capable de s'adapter à des zones d'habitat d'un genre nouveau.

L'entraînement des algorithmes de *deep learning* se réalise par itérations successives. A chaque itération, l'algorithme réalise des prédictions sur un petit paquet d'images sélectionnées aléatoirement. En comparant les prédictions avec les masques construits à l'étape précédente on arrive à en déduire une erreur qui dépend des paramètres θ du modèle. On peut ensuite bouger les paramètres dans le sens où l'erreur semble diminuer (en ayant calculé le gradient de l'erreur au préalable). On arrête l'entraînement quand le jeu d'images a été parcouru plusieurs fois par l'algorithme. Un entraînement dure 10 heures en moyenne.

Pour évaluer les résultats de notre algorithme pendant l'entraînement, certaines zones de Mayotte correctement labellisées ont été sélectionnées. On calcule la moyenne de l'Intersection Over Union (IOU) obtenu sur le jeu de données considéré (cf. Figure 16). L'IOU mesure la superposition des prédictions de l'algorithme aux annotations connues et est compris entre 0 (aucun recouvrement) et 1 (superposition parfaite).

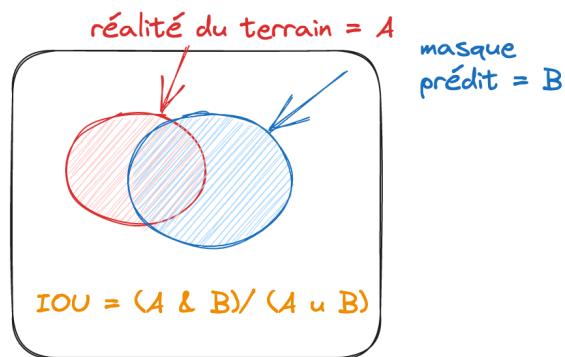


Figure 16. – Intersection over union

A l'issue de l'entraînement du modèle sélectionné, les prédictions de l'algorithme recouvrent les annotations du jeu de test à hauteur de 75%.

4 Analyse et mise en forme des sorties

Les sorties du modèle sont donc contenues dans un masque de polygones représentant les logements à un moment donné. La 45127368-afa1-446a-820f-fc64c546b2c5%fig-pred-quartier k représente sur l'île de Mayotte les prédictions obtenues en 2023. En comparant cette image avec par exemple une carte des quartiers favorisés ou non de Mayotte (45127368-afa1-446a-820f-fc64c546b2c5%fig-pred-quartier l), produite par l'Insee en 2012, on remarque que les prédictions semblent correctes, du moins globalement : les zones construites et habitées sont mises en valeur sur la carte. Cette vision générale des sorties permet de justifier de la pertinence de l'analyse puisqu'on parvient, avec ce modèle de *deep learning*, à délimiter les zones construites visibles sur des images satellites.



(k). – Prédictions originales 2023 de Mayotte



(l). – Localisation des quartiers à Mayotte (Insee 2012)

Figure 17

Le coeur du sujet réside dans l'évolution du bâti, et non simplement sa localisation, ce que la Figure 18 montre en jaune. Encore une fois, cette visualisation montre la pertinence des prédictions : il est normal d'observer des nouveaux bâtiments en 2023 en périphérie des bâtiments déjà construits en 2020, résultant de l'étalement urbain de l'île.



Figure 18. – Observation de l'étalement du bâti à Mayotte entre 2020 et 2023

Si on se concentre sur des zones plus spécifiques, c'est-à-dire en zoomant sur la représentation cartographique des polygones, on observe qu'on est effectivement capables de dessiner automatiquement les logements. L'exemple ci-dessous (Figure 19) présente les prédictions faites sur les images datant de 2023. Les polygones appliqués sur les images 2020 apparaissent à des endroits qui ne sont pas construits. Ils correspondent néanmoins à des constructions qui ont été faites entre ces deux années, comme le montre le fond d'image de 2023. Le modèle entraîné est donc capable de discerner le contour d'un logement et de le faire apparaître.



Figure 19. – Prédictions 2023 sur fond d'image 2020 (a) et sur fond d'image 2023 (b)

On remarque toutefois que le modèle n'est pas encore parfait. Premièrement parce que certains bâtiments n'apparaissent pas dans les prédictions, et parfois pas entièrement. C'est un problème qui peut être ignoré pour l'instant : ce projet n'a pas la prétention de produire une carte exhaustive, il se veut simplement être un guide pour les enquêtes cartographiques. Les résultats sont alors satisfaisants, puisqu'on met en valeur l'apparition de nouveaux bâtiments et on met en lumière la nécessité d'allouer des moyens humains pour la cartographie de cette zone particulière. De plus, augmenter la précision et l'exhaustivité des prédictions nécessite l'entraînement de nouveaux modèles, ou bien plus performants, ou bien entraînés sur d'autres images (voir la partie précédente). Deuxièmement, il semble impossible de demander aux enquêteurs de se pencher sur l'intégralité de l'île en zoomant autant. Il est donc nécessaire pour une mise à disposition des résultats correcte de produire des outils permettant une lecture d'une part plus rapide, et d'autre part plus facile.

4.1 Traitement des polygones

Un premier traitement des polygones est donc nécessaire après leur création. Il doit correspondre à une sorte de lissage et de tri des informations pertinentes, pour mettre en lumière correctement les modifications du bâti d'une année à l'autre. Ce traitement des résultats du modèle doit répondre à deux objectifs distincts, qui seront développés dans la suite de cette partie. D'une part, il faut que les bâtiments soient proprement délimités. Cela implique non seulement de réguler l'imprécision du tracé du masque, mais également de supprimer les petits polygones qui ne représentent pas de vraies habitations. D'autre part, les résultats doivent être utiles afin de répondre à la demande du projet. Il faut donc que l'outil de visualisation soit performant, clair, et qu'il soit possible d'orienter l'attention des utilisateurs sur les points critiques importants *via* par exemple la mise en place de statistiques descriptives ou de classification des îlots.

On travaille donc dans un premier temps sur le traitement des polygones de bâti prédits par le modèle. La Figure 20 représente les prédictions faites en 2023 sur une zone de Mayotte qu'on peut

qualifier de centre-ville. A gauche, les polygones des bâtiments prédis sont représentés sur un fond blanc, et à droite ces mêmes prédictions sont superposées aux images satellites de 2023.



Figure 20. – Prédiction originale 2023 d'une zone de Mayotte

On remarque trois problèmes sur ces images. Premièrement, il est clair que les polygones sont répartis en quatre images distinctes, qu'on distingue par les lignes verticales et horizontales qui traversent la ville. C'est dû à la construction du modèle, qui gère chaque image satellite indépendamment des unes des autres lors de la prédiction. Ainsi, lorsqu'un polygone est à cheval entre deux images, il apparaît et est comptabilisé comme deux polygones distincts.

Deuxièmement, on observe sur ces deux représentations que le modèle n'est pas précis dans sa délimitation des bâtiments. Tandis que les constructions sont habituellement rectangulaires ou carrées, que les façades devraient être des lignes droites et claires (du moins si elles étaient tracées manuellement), on remarque de nombreuses imperfections et sortes d'aspérités tout autour de ceux-ci. Cela ne représente pas vraiment un problème en soi, mais cela rend la lecture de la carte difficile. De plus, conserver une telle précision dans les formes et les délimitations est un problème de « précision excessive » induite par l'utilisation de frontières significatives inutilement précises.

Enfin, on remarque quelques minuscules polygones mesurant moins de dix mètres carrés. Cela peut être dû à des imprécisions ou des erreurs du modèle. Ces points ne sont pas forcément pertinents pour l'analyse de la surface construite et rendent également la lecture de la carte plus complexe. Un questionnement autour de leur conservation ou non est donc nécessaire.

4.2 Buffering

Plusieurs pistes ont été explorées pour remédier à ces problèmes. Une d'entre elles néanmoins semble répondre aux deux premières problématiques et consiste en l'application d'un buffering. Le buffer, ou zone tampon, consiste à créer une nouvelle géométrie autour d'un polygone d'origine en ajoutant ou en retirant une distance fixe, le paramètre du buffer, de son contour extérieur.

Mathématiquement, on calcule les distances entre les points du polygone d'origine et on crée de nouveaux points à la distance spécifiée.

Pour un buffer positif, la forme du polygone initial va être étendue vers l'extérieur, entraînant une forme plus arrondie et étendue. Les angles saillants peuvent être arrondis, et les contours deviennent généralement plus réguliers. La fonction `geopandas.GeoSeries.buffer()` du package GeoPandas permet de préciser d'une part la distance d'élargissement et d'autre part la résolution de l'agrandissement. En conservant la résolution par défaut, égale à 16, les angles sont fortement arrondis. Un buffer négatif, c'est-à-dire l'application de la fonction avec une distance négative, va réduire la taille du polygone initial en retirant cette distance de son contour extérieur. Cela peut entraîner des modifications comme la suppression de parties du polygone, l'arrondissement des angles concaves ou la simplification des contours.

Voici un exemple (Figure 21) sur des formes simples avec un buffer positif, issu de la documentation du package GeoPandas. En modifiant le paramètre de résolution, on obtient un lissage plus ou moins important. L'objectif ici est d'obtenir un lissage le plus important possible, pour avoir des formes géométriques plus simples et faciliter ainsi l'analyse des formes. On conserve donc la résolution par défaut.



Figure 21. – Exemple de buffer avec deux différentes résolutions

Le traitement des polygones passe ainsi par une fonction en trois étapes distinctes :

- Application d'un buffer positif
- Fusion des polygones s'interceptant
- Application d'un buffer négatif

L'agrandissement des polygones par le buffer positif permet de créer des intersections entre les polygones les plus proches, notamment les polygones séparés par les frontières entre les images satellites. La première problématique est donc corrigée par la fusion des polygones qui présentent une intersection et correspondent ainsi à un même bâtiment ou groupe de bâtiments. Cette succession de traitements permet également de lisser et simplifier les polygones, grâce notamment aux propriétés de lissage et d'arrondissement du buffering permettant de corriger les aspérités, de faire disparaître les petits « trous » dans les polygones et les imperfections les plus légères. De même, la fusion après un agrandissement permet de ne pas laisser apparaître les séparations les plus petites entre les bâtiments, et les considérer comme un seul et même bloc simplifie grandement la lecture et l'utilisation des résultats. On obtient, après ce traitement, une visualisation comme suit (Figure 22).



Figure 22. – Prédictions après application de la fonction de lissage

Le troisième problème concerne la taille des polygones tracés sur la carte. En effet, certains d’entre eux sont trop petits pour correspondre à des habitations et correspondent plutôt à du « bruit ». Il peut donc être nécessaire de mettre en place un seuil de surface en-dessous duquel on considère le polygone comme inintéressant. Il faut néanmoins être prudent. Certes, un logement est considéré comment décent par la loi française à partir de 9m² habitables mais environ 30% de la population à Mayotte n’a pas accès à l’eau courante par exemple, ce qui témoigne de l’aspect précaire d’une partie des habitations de l’île. Des bidonvilles par exemple sont susceptibles de contenir des habitations plus petites que la taille légale des bâtiments.

4.3 Différences de bâti d’une année à l’autre

On tente ensuite d’observer l’évolution du bâti par îlot en mettant en considérant trois types de bâtiments différents :

- Les bâtiments présents à l’année 2020 et à l’année 2023. C’est en quelque sorte la surface qui n’a pas évolué. On les appelle ici **la surface conservée**.
- Les bâtiments présents en 2020 mais pas en 2023. On peut les appeler **les suppressions**.
- Les bâtiments présents en 2023 qui n’existaient pas en 2020. Ce sont **les créations**.

Pour mettre en lumière les créations et les suppressions, nous avons donc pris l’ensemble de la surface construite en 2020 et/ou en 2023 à laquelle nous avons soustrait la surface conservée. On obtient donc une liste de polygones qui sont présents soit en 2020, soit en 2023, mais pas les deux. Par construction, les surfaces restantes représentent des évolutions des constructions.



Figure 23. – Soustraction des bâtiments sur les deux années d'étude sur une zone de Mayotte

La Figure 23 donne le résultat de cette opération pour un îlot. On a en rouge les créations et en bleu les suppressions déduites des prédictions 2020 et 2023 de l'algorithme. Des formes longilignes apparaissent sur la carte et semblent dessiner le contour des bâtiments. Ces formes ne traduisent pas de réels mouvements dans le bâti mais témoignent de l'inconstance de notre algorithme dans la délimitation du contour du bâti stable d'une année sur l'autre. On remarque également quelques bâtiments ou constructions trop petits pour représenter réellement des habitations. Un travail de nettoyage est donc encore nécessaire.

4.4 Nettoyage de la soustraction

On cherche donc à trier les constructions restantes. On parle ici de constructions mais ce sont évidemment des « morceaux » de constructions, le résultat de la soustraction. Ainsi, la création par exemple d'un préau ou le rajout d'une citerne de rétention d'eau se manifeste dans ce résultat par un minuscule point visible sur la carte. Pour simplifier la lecture, on parle de construction mais il faut lire « polygone visible sur la carte résultant ou bien d'une extension négligeable d'un bâtiment déjà existant, ou bien du manque de précision du modèle et impliquant un brouillage des résultats pertinents pour l'analyse ».

On se penche alors sur un indice de compacité. On remarque que les contours des bâtiments n'ont pas forcément l'aire la plus petite, mais qu'ils ont une forme caractéristique : ils sont particulièrement allongés. L'indice de compacité calcule un rapport entre le périmètre et la surface d'un polygone. Il varie donc entre 0 et 1. La valeur 0 est une ligne parfaitement allongée tandis que la valeur 1 correspond à un cercle parfait. On utilise l'indice de compacité de Gravelius Bendjoudi & Hubert (2002) dont la formule est donnée par :

$$\text{Indice de Compacité} = (4 * \pi * \text{Aire}) / (\text{Perimetre}^2)$$

Avant de filtrer les polygones, il faut déterminer un seuil. On choisit ici un seuil de 0,1. Cette décision n'est pas simple, et a été approchée à tâtons, en observant progressivement l'état des polygones affichés sur la carte, et en comparant visuellement et manuellement aux images de 2020 et 2023. On calcule de plus la somme de la surface des polygones représentés en fonction du seuil choisi (Figure 24).



Figure 24. – Evolution de la surface représentée en fonction du seuil choisi

Les résultats sont assez logiques : plus le seuil est élevé et moins les bâtiments s'affichent sur la carte. A partir d'une certaine valeur (environ 0,7), toutes les constructions sont supprimées et la carte est vide. Néanmoins, cette visualisation est assez décevante, car il est difficile d'en extraire une règle de décision pour choisir la valeur du seuil. On remarque toutefois une sorte de marche autour de la valeur 0,1, ce qui vient appuyer nos observations manuelles.

On peut donc finalement filtrer les résultats de cette soustraction et obtenir la visualisation suivante (Figure 25) des modifications réelles de l'état du bâti, qui correspond précisément aux objectifs du projet.



Figure 25. – Evolution filtrée du bâti entre 2020 et 2023 d'une zone de Mayotte

Ces polygones « nettoyés » permettent maintenant de soutenir les chiffres du recensement à Mayotte notamment, en fournissant une base de données sur les logements, sur laquelle peut s'appuyer une estimation de la population et venir corroborer les chiffres produits par les enquêtes annuelles. Ils peuvent également diriger et soutenir les enquêteurs pour l'enquête cartographique.

Néanmoins, en ce qui concerne cette deuxième tâche, les données peuvent être exploitées pour produire des statistiques et une visualisation cartographique par îlot par exemple. Ainsi, en mettant en évidence les zones géographiques qui évoluent le plus vite, en classant les îlots selon leur importance d'un appui cartographique, on peut apporter l'information la plus précise et la plus utile possible. Ce sont des travaux statistiques que nous sommes en mesure de faire.

5 Pistes et évolutions

On décrit ici l'ensemble des compétences mobilisées pour arriver au produit fini permettant de soutenir le travail de repérage des enquêteurs et d'appuyer les estimations de population de l'Insee. Un prototype de produit est disponible [ici](#) sous forme d'application web. L'ensemble des besoins nécessaires au maintien du prototype existant ainsi qu'aux améliorations potentielles de l'outil sont discutés par la suite.

5.1 Stack technique et dette technique

On met en lumière ici l'ensemble des outils mobilisés et les compétences nécessaires à l'existence de ce projet, son maintien et ses futurs développements. En premier lieu, une expertise fine sur les bases de données de l'Insee (RIL et BDTOPO) est indispensable. La bonne compréhension des bases de données et des systèmes d'informations géographiques est capitale, pour manipuler les polygones de bâti, les coordonnées (x, y) des logements ainsi que les images dont les pixels sont géolocalisés. Les images ne faisant pas partie des données classiquement manipulées, il est également nécessaire de comprendre la structure de tels objets. En outre, le package python astrovision a été développé pour faciliter la manipulation des images et des masques associés.

Des compétences de base en apprentissage statistique sont également de mise pour éviter de grossières erreurs de surapprentissage. S'ajoutent à ces dernières une compréhension un peu plus fine des algorithmes de *deep learning* ainsi qu'une capacité à comprendre et reproduire les modèles présentés dans les articles de recherche les plus récents sur le sujet.

Enfin, beaucoup d'efforts ont été réalisés pour capitaliser tous les entraînements, garantir leur reproductibilité et faciliter leur exécution. Cela nécessite une expertise poussée sur les outils de monitoring comme MLflow (cf. Figure 26) ou les outils de programmation de tâches comme Argo-workflow.



Figure 26. – Monitoring et sauvegarde des modèles et conditions d'entraînement avec MLFlow

Au niveau de la gestion des données géographiques (images, polygones, contours administratifs), un geoserver (cf. Figure 27) a été mis en place afin de pouvoir servir dynamiquement les tuiles d'images, et les prédictions générées par les algorithmes sur plusieurs années et territoires.



Figure 27. – Interface du geoserver servant les images et les prédictions de l'algorithme à une application web

Enfin, une application web développée en React permet de mettre en évidence les résultats de l'algorithme et les travaux sur la différence des masques cités plus haut pour permettre une validation des agents en bureau, planifier l'enquête cartographique et de récupérer les statistiques d'évolution produites. Cette application est également très utile pour l'équipe projet afin d'évaluer en un coup d'œil et à l'échelle de territoires entiers la pertinence de nos algorithmes.



Figure 28. – Prise de vue de l'application web en cours de développement

L'ensemble des blocs techniques présentés ici cohabitent dans les serveurs du *datalab.sspcloud* de l'Insee et leur branchement nécessite de maîtriser les déploiements d'application *via* l'outil de gestion des conteneurs Kubernetes. Une forte dette technique s'est donc accumulée sur chacun des maillons de la chaîne de traitement et le seul maintien de l'ensemble nécessiterait déjà une équipe de plusieurs personnes à plein temps. Le diagramme de la figure Figure 29 récapitule l'ensemble des outils et compétences mobilisés sur le projet. L'équipe projet actuelle n'est pas à plein temps sur le sujet et le jeu des changements de postes conduira *in fine* sous peu à sa dislocation. A ce jour, aucune division n'accueille physiquement ce projet à l'Insee.



Figure 29. – Représentation schématique de l’ensemble des éléments nécessaires à l’obtention du produit d’aide à la décision assisté par les prédictions de l’algorithme

5.2 Suite des travaux et besoins

Les usages finaux des travaux présentés (estimations de population et planification de charge) dépendent entièrement de la qualité des prédictions réalisées par l’algorithme. De nombreuses directions sont possibles pour tenter d’améliorer ces prédictions, notamment au niveau des données d’entraînement. Actuellement, l’algorithme est seulement entraîné sur l’année 2022 et sur la Martinique et la Guadeloupe. Il est nécessaire de tester la capacité de généralisation de l’algorithme en réduisant ou élargissant, temporellement ou géographiquement, le jeu d’entraînement qui nourrit son apprentissage.

Les masques construits à partir des images satellites sont des masques de bâti, et non des masques de logements, ce qui implique que l’algorithme ne peut distinguer le logement du bâti. Des travaux de labellisation manuelle pourraient venir corriger les masques produits *via* la BDTOPO en ce sens. Une telle opération est nécessairement coûteuse en moyens humains et il est difficile d’estimer le rapport entre ce coût et les gains qui seront enregistrés sur les prédictions. D’autres sources de données, hors Insee, pourraient être explorées pour constituer les masques telles qu’Open Street Map.

Au niveau des images même, les possibilités sont multiples : d’autres sources d’imagerie satellites sont disponibles, telles que les images Sentinel 2 ou les images des satellites Spot. Ces sources présentent des résolutions spatiales plus faibles que les images PLEIADES mais leur résolution temporelle est plus élevée ce qui implique qu’on pourrait les utiliser pour réaliser des estimations provisoires en l’attente de l’obtention de la couverture complète PLEIADES. Des travaux sur l’imagerie Sentinel 2 sont présentés dans le rapport Nabec (2023) et montrent qu’on peut obtenir un niveau de prédiction du bâti très satisfaisant à partir de ces images de moins bonne résolution. Des sources d’images dites stéréoscopiques ajoutent une donnée supplémentaire sur l’altitude du bâti visualisé et mèriraient également d’être expertisées.

Les réflexions sur l’algorithme sélectionné est tout aussi importante. En effet, la littérature scientifique est foisonnante sur les modèles de *Segmentation* et il est donc nécessaire de réaliser une

veille technique permanente sur le sujet. Certains modèles peuvent en théorie s'adapter à des images de résolutions différentes. Ainsi, l'entraînement de ces algorithmes pourrait être gonflé par des jeux d'images croisant de multiples sources, notamment par des jeux de données préanotés rendus disponibles par des travaux de recherche académique.

Une veille sur les travaux académiques est donc indispensable pour enrichir les travaux, ainsi qu'une veille sur les pratiques des autres instituts sur l'utilisation de l'imagerie satellitaire. Les contacts avec l'Institut Géographique National devraient être également plus touffus dans la mesure où la constitution même de la BDTOPO réside dans des travaux sur l'imagerie aérienne. Enfin, des échanges plus fréquents avec des acteurs du monde académique, voire la mise en place de projets de recherche visant exclusivement à répondre aux cas d'usages mentionnés en introduction, profiteraient grandement à l'avancée du projet.

Du point de vue opérationnel seulement, partant du principe que la méthode de constitution des données et le choix de l'algorithme sont arrêtés, l'intégration de l'outil dans le processus de production Insee n'est pas aisée. En amont déjà, l'obtention des images devrait être internalisée à l'Insee avec la création d'un service responsable de cette acquisition. De même, les temps moyens d'acquisition *i.e.* le décalage entre la date de commande de l'image satellite et son obtention effective devraient être mesurés afin de prendre la mesure du décalage potentiel entre la vérité du terrain et celle photographiée.

Des entraînements devraient aussi être réalisés assez fréquemment afin d'actualiser l'algorithme avec de nouvelles données et améliorer ses capacités de prédiction. Ensuite, les résultats des algorithmes pourraient être expertisés par les agents en bureau qui vérifieraient la pertinence des prédictions en les superposant aux images dont elles sont issues, ce qui induit nécessairement une réorganisation du travail. Cette phase d'expertise permettrait également aux agents d'émettre des propositions d'amélioration de l'outil de mise à disposition des résultats.

Toutes ces compétences et ces missions mises bout à bout, la somme des moyens nécessaires à l'évolution d'un tel projet dépasse largement les moyens qui lui sont actuellement alloués et une équipe à plein temps dédiée à ce projet permettrait de sanctuariser les outils et compétences nécessaires à son maintien puis à son développement.

Bibliographie

- Bendjoudi, H., & Hubert, P. (2002). Le coefficient de compacité de Gravelius: analyse critique d'un indice de forme des bassins versants. *Hydrological Sciences Journal*, 47(6), 921-930. <https://doi.org/10.1080/02626660209493000>
- Berova, R. (2023). *Détection automatisée de changements des bâtiments en France d'Outre-Mer à partir d'images satellites*. https://minio.lab.sspcloud.fr/cguillo/rapport_stage/Rapport_de_stage_3_A-2.pdf
- Chabennet, Q. (2021). *Détection automatique de la couverture des sols à partir d'images satellites à l'aide de méthodes de segmentation sémantique*. https://minio.lab.sspcloud.fr/cguillo/rapport_stage/rapport_stage_quentin_dmrg-1.pdf

- Chen, L.-C., & al. (2017). Rethinking Atrous Convolution for Semantic Image Segmentation. *arXiv*. <http://arxiv.org/abs/1706.05587>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., & others. (2021). An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv preprint arXiv:2010.11929*. <http://arxiv.org/abs/2010.11929>
- Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., & Bengio, Y. (2017). The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. *arXiv:1611.09326 [cs]*. <http://arxiv.org/abs/1611.09326>
- Kim, P. (2017). Convolutional Neural Network. In P. Kim (éd.), *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence* (p. 121-147). Apress. https://doi.org/10.1007/978-1-4842-2845-6_6
- Nabec, J. (2023). *Mise en place d'une méthode de détection automatique des évolutions de bâti en Outre-Mer sur des images satellites*. https://minio.lab.sspcloud.fr/cguillo/rapport_stage/Rapport_de_stage_2023-3.pdf
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv:1505.04597 [cs]*. <http://arxiv.org/abs/1505.04597>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, 30. <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>
- Visin, F., Ciccone, M., Romero, A., Kastner, K., Cho, K., Bengio, Y., Matteucci, M., & Courville, A. (2016). ReSeg: A Recurrent Neural Network-based Model for Semantic Segmentation. *arXiv:1511.07053 [cs]*. <http://arxiv.org/abs/1511.07053>
- Xie, E., Wang, W., Yuille, A. L., Anandkumar, A., & Alvarez, J. M. (2021). SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *arXiv preprint arXiv:2105.15203*. <http://arxiv.org/abs/2105.15203>