

RÉPUBLIQUE DU SÉNÉGAL



UNIVERSITÉ CHEIKH ANTA DIOP DE DAKAR



ÉCOLE SUPÉRIEURE POLYTECHNIQUE

DÉPARTEMENT GÉNIE INFORMATIQUE

MÉMOIRE DE FIN DE CYCLE

Pour l'obtention du :

Master Intelligence Artificielle et Big Data

SUJET :

Surveillance et Détection des Changements d'occupations du sol à partir d'images satellites grâce aux modèles de Deep Learning

Lieu de stage : IDR UMMISCO Période stage : 01/08/2023 - 01/01/2024



Année universitaire : 2022-2023

Dédicaces

Je tiens à dédier ce mémoire à ma précieuse famille, qui a toujours été un soutien inébranlable à chaque étape de ma vie. À mon oncle Mor Thiam, un homme d'une bonté incommensurable, présent dans les moments de joie comme dans les moments difficiles. À ma tendre mère, Fatou Thiam, qui m'a tout donné et qui a toujours été fière d'être ma mère. À mon père, Samba Thiam, qui a forgé en moi une mentalité de fer. À mon frère, El Hadj Thiam, qui a toujours été un soutien indéfectible et a joué son rôle de frère à la perfection. À mon cousin, Macodou Ndiaye dit Papis, toujours présent pour me prodiguer de précieux conseils et des orientations avisées. À Khady Ndao, toujours là pour illuminer mes journées et raviver mon sourire. À toute la famille Thiam. Et enfin, pour conclure, je remercie tous mes amis en commençant par Cheikh Seck, Charles Abdoulaye Ngom, Serigne Mbaye Sarr, Thérèse Daba Siga Diagne et Ndoumbé Touré, pour tous les bons moments passés ensemble.

Remerciements

Tout d'abord, je souhaite remercier chaleureusement mon maître de stage, le Professeur Alassane BAH, qui m'a offert l'opportunité de plonger dans le domaine de l'intelligence artificielle appliquée à l'environnement. Votre confiance en moi et votre soutien inébranlable ont marqué le point de départ de cette aventure académique. Ma reconnaissance envers vous et le Pr Etienne DELAY est immense, car cette opportunité a véritablement façonné ma carrière et ma passion pour ce domaine. Je tiens également à exprimer ma profonde gratitude envers mon professeur encadrant, le Dr Mandicou BA, dont l'expertise et la disponibilité ont été précieuses tout au long de ce travail. Vos conseils éclairés, votre encadrement attentif et votre vision ont été d'une valeur inestimable pour approfondir ma réflexion et mener à bien cette recherche. Toutes ces personnes ont joué un rôle clé dans la réussite de ce travail de recherche, et je leur en suis profondément reconnaissant.

Table des matières

Introduction Générale	1
1 Présentation Générale	3
1.1 Présentation de la structure d'accueil	4
1.1.1 Présentation de l'IRD et de l'UMMISCO	4
1.2 Présentation du sujet	5
1.2.1 Contexte	5
1.2.2 Problématique	7
1.2.3 Objectifs	7
1.2.4 Démarche méthodologique	8
2 Concepts Généraux	12
2.1 Généralités sur la Télédétection	14
2.1.1 Définition de la Télédétection :	14
2.1.2 Objectifs et Domaines d'Application :	18
2.1.3 Indices de Changement :	19
2.2 Généralités sur l'Intelligence Artificielle	19
2.2.1 Définition de l'Intelligence Artificielle :	20
2.2.2 Sous-Disciplines de l'IA :	20
2.2.3 Applications de l'IA :	21

TABLE DES MATIÈRES

2.2.4	Apprentissage Automatique et Apprentissage Profond :	21
2.2.5	Importance de l'IA :	21
2.2.6	Défis et Éthique :	22
2.3	Généralités sur les réseaux de neurones	22
2.3.1	Réseaux de Neurones Artificiels (ANNs)	23
2.3.2	Propagation vers l'Avant (Forward Propagation)	23
2.3.3	Rétropropagation (Backpropagation)	24
2.4	Généralités sur la segmentation en dans la vision par ordinateur :	26
2.4.1	La segmentation par instance :	26
2.4.2	La segmentation sémantique :	27
2.4.3	La segmentation panoptique :	29
2.5	Les types d'apprentissage en l'intelligence artificielle	31
2.5.1	Apprentissage supervisé	32
2.5.2	Apprentissage non supervisé	33
2.5.3	Apprentissage mixte	34
2.5.4	L'apprentissage par transfert	35
3	Etat de l'Art	36
3.1	Méthodes traditionnelles et méthodes de l'apprentissage profonds en télédétection pour la surveillance et la détection des changements d'occupations du sol.	37
3.1.1	Méthodes traditionnelles utilisées en télédétection : Une Analyse Approfondie	37
3.1.2	L'utilisation de l'apprentissage profond dans la télédétection pour la surveillance et la détection des changements d'occupations du sol.	39
3.2	Critères de sélection et méthodologie de choix	44
3.3	Avantages et limites des différentes approches	46

TABLE DES MATIÈRES

3.3.1	Méthodes Traditionnelles :	46
3.3.2	Méthodes d'apprentissage profond :	47
3.3.3	Analyse comparative :	49
4	Proposition de solution	51
4.1	Description de la collecte, de l'étiquetage, et de la validation des données	52
4.1.1	Description de la collecte	52
4.1.2	Description de l'étiquetage	52
4.1.3	Description de la validation des données	53
4.2	Description du prétraitement et de la méthode de génération des données	55
4.2.1	Description du prétraitement	55
4.2.2	Description de la méthode de génération des données	56
4.3	Analyse de l'architecture des modèles de prédiction	58
4.3.1	L'archtecture U-Net et inception-resnetv2 :	58
4.3.2	L'archtecture FCN-8 (Fully Convolutional Network 8) et VGG19 :	63
4.4	Déploiement en temps réel :	65
5	Analyse et Interprétation des résultats	67
5.1	Résultats de la prédiction des modèles sur les images satellites géospatio-temporelles pour la surveillance et la détection des changements d'occupations du sol.	68
5.1.1	Résultats expérimentaux avec U-Net et FCN-8 :	69
5.1.2	Prédiction avec U-Net et FCN-8 pour la surveillance :	72
5.1.3	Détection des changements d'occupations du sol :	75
5.1.4	Généralisation sur d'autres zones du Sénégal :	81

TABLE DES MATIÈRES

5.1.5 Déploiement en temps réel :	83
5.2 Discussion des résultats : Analyse et Interprétation des résultats	84
Conclusion et Perspectives	87

Sigles et Abréviations

IA : Intelligence Artificielle

SVM : Support Vector Machine

CAH : Classification Ascendante Hiérarchique

ANN : Artificial Neural Network

CNN : Convolutional Neural Network

FCN : Fully Convolutional Network

DL : Deep Learning

BiLSTM : Bidirectionnelle Long Short Term Memory

VGG : Visual Geometry Group

ACP : Analyse en Composantes Principales

PD : Pourcentage Difference

PCC : Pourcentage de Classification Correcte

MSE : Mean Square Error

IOU : Intersection Over Union

OA : Accuracy Overall

ReLU : Rectified Linear Unit

TP : True Positive

TN : True Negative

FP : False Positive

FN : False Negative

Sigles et Abréviations

HRMS : High-Resolution Mass Spectrometry

VHR : Very High Resolution

SIG : Système d'Information Géographique

LPIS : Land Parcel Identification System

GMV : Grande Muraille Verte

ESA : Agence Spatiale Européenne

Table des figures

1.1	Démarche méthodologique de l'étude.	8
1.2	On a les images qui sont les entrées et les masques qui sont les sorties avec les classes correspondantes.	9
2.1	Processus de la Télédétection	14
2.2	Télédétection passive	16
2.3	Télédétection active	17
2.4	Les sous-disciplines de l'intelligence artificielle	20
2.5	Réseau de neurones artificiels	23
2.6	Segmentation par instance	27
2.7	Segmentation sémantique	28
2.8	Segmentation panoptique	29
2.9	Apprentissage supervisé	32
2.10	Apprentissage non supervisé	33
4.1	L'architecture U-Net [9]	58
4.2	L'architecture inception-resnetv2 [5]	60
4.3	La répartition des classes dans les données.	62
4.4	L'architecture FCN [29]	63
4.5	L'architecture VGG19 [30]	65

TABLE DES FIGURES

5.1	La prédiction du modèle U-Net sur les images satellites avec une résolution de 8192x4320, et avec une altitude de 10 kilomètres de hauteur	73
5.2	La prédiction du modèle FCN-8 sur les mêmes images satellites avec la même résolution de 8192x4320, et avec une altitude de 10 kilomètres de hauteur	74
5.3	La détection de changement avec U-Net sur différentes périodes, avec une résolution de 8192x4320, et avec une altitude de 10 kilomètres de hauteur	76
5.4	La prédiction du modèle U-Net respectivement sur des images satellites avec différentes résolutions : 8192x4320, 3840x2160, 1920x1080, 1280x720, 1050x632 et 1024x768, et avec une altitude de 100 kilomètres de hauteur	77
5.5	La prédiction du modèle FCN-8 respectivement sur des images satellites avec différentes résolutions : 8192x4320, 3840x2160, 1920x1080, 1280x720, 1050x632 et 1024x768, et avec une altitude de 100 kilomètres de hauteur	78
5.6	La prédiction du modèle U-Net respectivement sur une image satellite avec une résolution de 8192x4320, et avec une altitude de 50 kilomètres de hauteur	79
5.7	La prédiction du modèle FCN-8 respectivement sur une image satellite avec une résolution de 8192x4320 et avec une altitude de 50 kilomètres de hauteur	79
5.8	La prédiction du modèle U-Net respectivement sur une image satellite avec une résolution de 8192x4320 et avec une altitude de 15 kilomètres de hauteur	79

TABLE DES FIGURES

5.9 La prédiction du modèle FCN-8 respectivement sur une image satellite avec une résolution de 8192x4320 et avec une altitude de 15 kilomètres de hauteur	80
5.10 La prédiction du modèle U-Net respectivement sur une image satellite avec une résolution de 8192x4320 et avec une altitude de 15 kilomètres de hauteur	80
5.11 La prédiction du modèle U-Net respectivement sur une image satellite avec une résolution de 8192x4320 et avec une altitude de 4 kilomètres de hauteur	80
5.12 La prédiction du modèle FCN-8 respectivement sur une image satellite avec une résolution de 8192x4320 et avec une altitude de 4 kilomètres de hauteur	81
5.13 La prédiction du modèle U-Net respectivement sur une image satellite avec une résolution de 3840x2160 et avec une altitude de 4 kilomètres de hauteur	81
5.14 La détection sur la zone de Dakar.	82
5.15 La détection sur la zone de Zinguinchor.	82
5.16 La détection sur la zone de Matam.	83
5.17 QR Code pour accéder aux démos vidéos.	83

Liste des tableaux

5.1	Les métriques globales	71
5.2	Rapport de classification des modèles	71
5.3	Matrices de confusion globales des différents modèles.	72

Résumé

Ce mémoire se concentre sur la surveillance et la détection des changements d'occupations du sol au Sénégal, en mettant en lumière la région de Richard Toll, qui est située dans une zone d'intérêt liée à la Grande Muraille Verte (GMV), une initiative régionale de préservation environnementale. L'approche utilisée repose sur les modèles de Deep Learning appliqués à l'analyse d'images satellites, relevant du domaine de la télédétection. La méthodologie commence par la collecte de données géospatio-temporelles, des images à 3 canaux prises à des moments différents dans le temps et dans des espaces différents, de la zone d'intérêt via Google Earth Pro.

La région étudiée est divisée en huit classes distinctes, notamment les champs cultivés, les champs non cultivés, les terres, l'eau, les bâtiments, les routes, les terrains de football et la végétation. Les architectures de l'apprentissage profond sont employées pour former des modèles qui atteignent une accuracy de 96% et un coefficient de Jaccard de 92.55% pour le premier modèle, ainsi qu'une accuracy de 91% et un coefficient de Jaccard de 84% pour le deuxième modèle, offrant ainsi des outils précis pour la surveillance des changements environnementaux dans le cadre des initiatives de conservation régionale.

Néanmoins, les limitations matérielles ont entravé l'utilisation de données plus vastes pour l'entraînement.

Mots-clés : Télédétection, Deep Learning, Surveillance, Détection de Changements, Richard Toll, Sénégal, Grande Muraille Verte.

Abstract

This dissertation focuses on the monitoring and detection of land use changes in Senegal, highlighting the Richard Toll region, which is located in an area of interest linked to the Great Green Wall (GGW), a regional environmental preservation initiative. The approach used is based on Deep Learning models applied to the analysis of satellite images, falling within the field of remote sensing. The methodology begins with the collection of geospatial-temporal data, 3-channel images taken at different points in time and in different spaces, of the area of interest via Google Earth Pro.

The study region is divided into eight distinct classes, including cultivated fields, uncultivated fields, land, water, buildings, roads, football fields and vegetation. Deep learning architectures are used to train models that achieve an accuracy of 96% and a Jaccard coefficient of 92.55% for the first model, as well as an accuracy of 91% and a Jaccard coefficient of 84% for the second model, thus providing precise tools for monitoring environmental changes as part of regional conservation initiatives.

However, hardware limitations hindered the use of larger data for training.

Keywords : Remote sensing, Deep Learning, monitoring, Change Detection, Richard Toll, Senegal, Great Green Wall.

Introduction Générale

La surveillance et la détection des changements d'occupations du sol sont des éléments cruciaux pour comprendre l'évolution de l'environnement terrestre et ses implications sur la planification et la gestion des ressources. La GMV, un ambitieux projet de restauration écologique s'étendant à travers l'Afrique pour lutter contre la désertification et la dégradation des terres [12], nécessite une surveillance précise pour atteindre ses objectifs. Au Sénégal, un pays caractérisé par une diversité de paysages et d'activités humaines, cette surveillance revêt une importance particulière pour garantir un développement durable et une utilisation judicieuse des terres. C'est dans ce contexte que notre étude se penche sur la surveillance et la détection des changements d'occupations du sol, avec un accent particulier sur la région de Richard Toll, située dans une zone d'intérêt liée à la Grande Muraille Verte (GMV).

La méthode que nous avons adoptée repose sur l'utilisation de techniques d'apprentissage profond pour l'analyse d'images satellitaires géospatio-temporelles, s'inscrivant ainsi dans le domaine de la télédétection. À travers cette approche novatrice, nous avons cherché à collecter et à analyser des données géospatio-temporelles de la région de Richard Toll en utilisant le logiciel Google Earth Pro. Cette collecte de données a été minutieusement réalisée, avec une répartition des parties de la zone en huit classes distinctes : champs cultivés, champs

non cultivés, terres, eau, bâtiments, routes, terrains de football et végétation. Cette démarche s'aligne avec les objectifs de la GMV en promouvant une utilisation durable des terres.

L'essence de notre travail repose sur l'application de méthodes d'apprentissage profond basées sur les réseaux de neurones. À travers un processus de collecte de données géospatio-temporelles, une exploitation de différentes architectures de réseaux de neurones, de l'apprentissage par transfert, et de l'ajustement fin (fine tuning) afin d'améliorer la fiabilité de la surveillance et détection des changements d'occupations du sol.

Dans les chapitres suivants de ce mémoire, nous explorerons en détail, tout d'abord, la présentation de la structure d'accueil de notre étude dans le chapitre 1. Dans le chapitre 2, nous parlerons des concepts généraux pour une meilleure compréhension de l'étude. Puis dans le chapitre 3, nous passerons en revue les différents travaux qui ont été menés à ce sujet. Ensuite, nous proposerons notre solution dans le chapitre 4, puis ferons une analyse et interprétation des résultats dans le chapitre 5. Enfin, nous allons terminer par une conclusion et une ouverture des perspectives pour étendre l'étude, en tenant compte des objectifs de la GMV.

Chapitre 1

Présentation Générale

Résumé : Dans ce premier chapitre, nous amorçons notre exploration en introduisant le domaine de notre recherche et en mettant en lumière les principales entités impliquées. Nous commençons par présenter l'IRD (Institut de recherche pour le développement) et l'UMMISCO (Unité de Modélisation Mathématique et Informatique des Systèmes Complexes), les deux institutions de recherche centrales liées à notre étude. Ensuite, nous plongeons dans le contexte spécifique de la surveillance et de la détection des changements dans la région de Richard Toll. Nous formulons la problématique de recherche qui guidera notre travail et définissons de manière précise les objectifs que nous nous efforçons d'atteindre au cours de cette étude et enfin la démarche méthodologique optée.

Sommaire

1.1	Présentation de la structure d'accueil	4
1.2	Présentation du sujet	5

CHAPITRE 1. PRÉSENTATION GÉNÉRALE

1.1 Présentation de la structure d'accueil

1.1.1 Présentation de l'IRD et de l'UMMISCO

Depuis presque 80 ans, l’Institut de Recherche pour le Développement (IRD) est un organisme de recherche public français pluridisciplinaire qui établit des partenariats équitables avec les pays du Sud et les Outre-mer français [3]. L’IRD croit fermement que les avancées scientifiques sont essentielles pour promouvoir le développement durable et humain. Cette conviction est partagée et exprimée dans toutes les régions où l’IRD est présente et travaille avec ses partenaires. Elle mène des activités de recherche, de formation et d’innovation en partenariat avec les institutions sénégalaises. Les chercheurs sénégalais et français travaillent ensemble pour définir les thématiques de recherche en fonction des priorités identifiées dans le pays et des axes scientifiques de l’institut tels que la santé, la pêche, le changement climatique, etc [4].

L’Unité de Modélisation Mathématique et Informatique des Systèmes Complexes (UMMISCO) quant à elle est une unité mixte internationale qui se trouve sous la tutelle de sept organismes universitaires dont la principale est l’IRD en France, au Maroc, au Sénégal, au Cameroun et au Vietnam. Cette unité se consacre à la conception d’approches et d’outils de modélisation mathématique et informatique pour mieux comprendre les dynamiques émergentes des ”systèmes complexes”. Ces thématiques scientifiques visent à élaborer des scénarios d’évolution les plus probables et à contribuer à la prise de décision en matière de développement durable. Dans le but d’assurer une meilleure cohérence en termes d’effectifs et de dynamisme entre les différents thèmes, pour la période 2019-2023, l’UMMISCO a regroupé leurs travaux de

CHAPITRE 1. PRÉSENTATION GÉNÉRALE

recherche en trois thèmes :

1. Applications de la modélisation de systèmes complexes
2. Science des modèles complexes
3. De la complexité des données aux modèles.

Ainsi, comme nous pouvons le voir, ces thèmes portent en particulier sur l'usage de la modélisation de systèmes complexes pour résoudre des problèmes de développement, de gestion de ressources et des milieux avec des problématiques telles que la santé où s'inscrit le sujet de ce mémoire.

1.2 Présentation du sujet

Dans cette section, nous allons examiner en détail les quatre aspects essentiels de notre travail. Ces aspects sont les suivants : le contexte dans lequel notre étude s'inscrit, la problématique que nous cherchons à résoudre, les objectifs que nous visons à atteindre, ainsi que la méthodologie que nous avons employée pour mener à bien notre recherche.

1.2.1 Contexte

Le Sénégal, situé en Afrique de l'Ouest, est un pays d'une grande diversité géographique, allant des vastes étendues de terres agricoles aux zones côtières bordées d'océan, en passant par des centres urbains en constante évolution [2]. La gestion efficace de ces ressources et la surveillance des changements environnementaux sont des défis majeurs pour le développement durable du pays, notamment dans le cadre des initiatives de la Grande Muraille Verte, qui vise à lutter contre la désertification et à promouvoir la restauration des terres dans la région.

CHAPITRE 1. PRÉSENTATION GÉNÉRALE

Parmi les régions clés du Sénégal, la région de Richard Toll, située dans le nord, se distingue par ses paysages agricoles, sa proximité avec la rivière Sénégal et son rôle économique en tant que pôle de production agricole, contribuant ainsi aux objectifs de la GMV dans la région.

La région de Richard Toll est le cœur de l'industrie sénégalaise de la canne à sucre, avec de vastes plantations qui s'étendent à perte de vue. Cependant, cette région est également sujette à des changements environnementaux et économiques constants, tels que l'expansion des terres agricoles, les fluctuations des niveaux d'eau et le développement urbain. Ces changements ont un impact direct sur les ressources naturelles, l'agriculture, l'environnement et la qualité de vie des habitants [2], et ils sont au cœur des préoccupations de la GMV pour restaurer et protéger les terres.

La surveillance et la détection des changements d'occupations du sol dans la région de Richard Toll sont essentielles pour plusieurs raisons. Premièrement, elles permettent de suivre l'évolution des terres agricoles, notamment des plantations de canne à sucre, ce qui est d'une importance capitale pour l'industrie agricole du pays et pour les efforts de restauration de la GMV. Deuxièmement, elles contribuent à la gestion des ressources en eau, en particulier la rivière Sénégal, qui est vitale pour l'irrigation des cultures, un élément clé de la GMV. Enfin, elles soutiennent la planification urbaine et le développement durable de la région, en tenant compte des besoins croissants de la population et des objectifs de la GMV en matière de gestion des terres. C'est dans ce contexte complexe que notre étude prend tout son sens en contribuant aux efforts de la GMV pour restaurer et protéger les terres dans la région de Richard Toll.

CHAPITRE 1. PRÉSENTATION GÉNÉRALE

1.2.2 Problématique

Dans le cadre des objectifs de la Grande Muraille Verte visant à restaurer et à protéger les terres dans la région du Sénégal, la question cruciale se pose : Comment développer une méthodologie efficace et précise pour surveiller et détecter les changements d'occupations du sol dans la région de Richard Toll, en tenant compte des défis liés aux variations environnementales et économiques ? Comment concevoir des outils d'analyse basés sur le Deep Learning qui répondent aux besoins de la GMV, en considérant la diversité des paysages, l'importance de l'agriculture, la gestion de l'eau et la planification urbaine dans ce contexte spécifique ?

1.2.3 Objectifs

L'objectif principal de ce mémoire est de développer une méthodologie intégrant des modèles de Deep Learning pour la surveillance et la détection des changements d'occupations du sol dans la région de Richard Toll, au Sénégal, en alignement avec les objectifs de la GMV.

Plus précisément, les objectifs sont les suivants :

1. Développement de modèles de Deep Learning robustes : Concevoir des modèles de Deep Learning capables de classifier efficacement les différentes catégories d'occupations du sol, notamment les champs agricoles, les ressources en eau et les zones urbaines, en tenant compte de la variabilité des données dans cette région, pour la restauration et la protection les terres.

2. Évaluation des impacts environnementaux et économiques : Évaluer les implications des changements d'occupations du sol détectés sur l'agriculture, la gestion de l'eau, et le développement urbain, en contribuant ainsi à une meilleure compréhension des enjeux environnementaux et économiques dans

CHAPITRE 1. PRÉSENTATION GÉNÉRALE

la région de Richard Toll.

3. Proposition de recommandations : Fournir des recommandations pratiques pour les acteurs locaux, les autorités gouvernementales et les parties prenantes, conformément aux objectifs de la GMV, afin d'améliorer la gestion des ressources, de promouvoir la restauration des terres, et de faciliter la prise de décision en matière de planification urbaine et de développement agricole dans la région.

1.2.4 Démarche méthodologique



Figure 1.1 – Démarche méthodologique de l'étude.

CHAPITRE 1. PRÉSENTATION GÉNÉRALE

Notre approche méthodologique repose sur l'apprentissage supervisé, l'apprentissage par transfert, l'ajustement fin (fine tuning) et s'articule autour de la segmentation sémantique. Cette approche suit un processus structuré composé de sept étapes, chacune étant aussi cruciale que les précédentes, comme illustré dans la Figure 1.1.

Collecte de données : La démarche commence par la collecte de données géospatio-temporelles de la région de Richard Toll à partir de Google Earth Pro qui fournit des images avec 3 canaux RVB (Rouge, Vert et Bleu). Ces données comprennent des images satellites historiques couvrant plusieurs années.

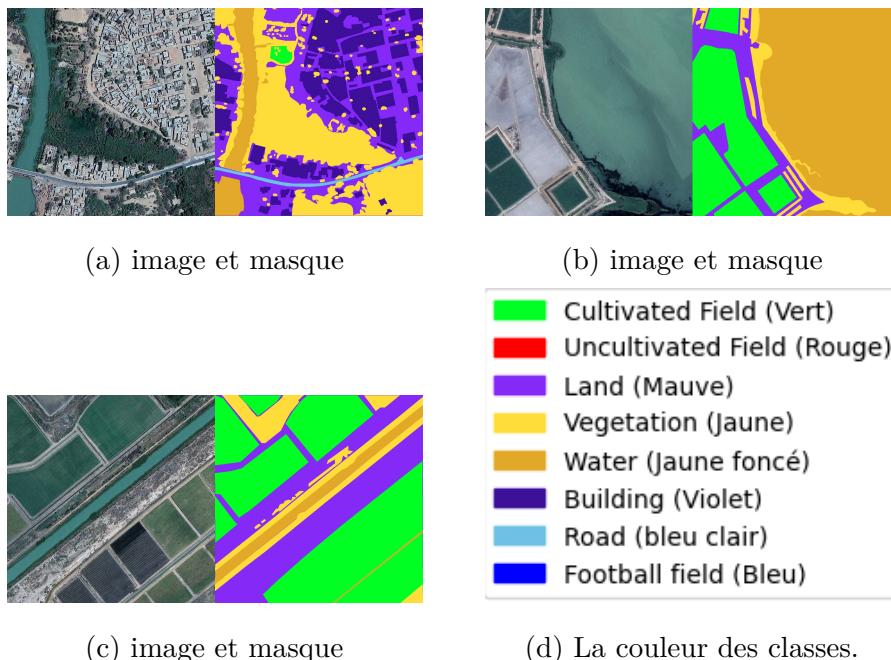


Figure 1.2 – On a les images qui sont les entrées et les masques qui sont les sorties avec les classes correspondantes.

CHAPITRE 1. PRÉSENTATION GÉNÉRALE

Annotation des données : Des annotations manuelles sont réalisées à l'aide de l'outil « Labelme » pour étiqueter les différentes zones d'intérêt, qui sont au nombre de 8. Nous avons les champs cultivés, les champs non cultivés, les bâtiments, l'eau, la végétation, le sol, les routes et les terrains de football voir Figure 1.2d.

Prétraitement des données : Les images collectées sont soumises à un ensemble de processus visant à les préparer pour une utilisation dans les modèles d'apprentissage automatiques. Ces étapes comprennent :

1. Organisation en tuile : Les images sont divisées en patches de taille 800x800 pour ensuite être organisées en tuiles dans lesquelles sont contenues les images et leurs masques correspondants.
2. Redimensionnement : Chaque image est redimensionnée à une taille spécifiée de 256x256. Les images sont également converties en format RGB (Red Green Blue) si elles ne le sont pas déjà.
3. Normalisation des pixels : Les pixels de chaque image sont normalisés pour s'assurer qu'ils sont dans la plage [0, 1]. Cela peut être important pour l'apprentissage de modèles de réseaux de neurones, car cela peut aider à accélérer la convergence de l'entraînement.
4. Découpage en patches : Chaque image est divisée en patches de la taille 256x256 avec un certain chevauchement. Cette opération permet de créer des sous-images qui peuvent être utilisées pour l'apprentissage en patch. Ces patches peuvent être plus faciles à gérer pour l'apprentissage, notamment lorsque les images sont grandes.
5. Ajout des patches au jeu de données : Les patches extraits sont ajoutés au jeu de données. Cela permet de créer un ensemble de données d'entraînement à partir des images d'origine et de leurs patches.

CHAPITRE 1. PRÉSENTATION GÉNÉRALE

6. Division en données d'entraînement et de test : Nous avons divisé nos données en données d'entraînement avec lesquelles seront entraînés les modèles et en données de test avec lesquelles, seront évalués les modèles.
7. Stockage des images : Les images prétraitées sont stockées dans un fichier au format « .npz » qui est fourni par la bibliothèque « numpy » pour une utilisation ultérieure dans les modèles d'apprentissage automatiques et la réduction de la consommation des ressources matérielles.

Choix des modèles de Deep Learning : Nous avons pris la décision de retenir deux modèles parfaitement adaptés à la tâche de segmentation sémantique, en les combinant chacun avec un autre modèle pré-entraîné pour bénéficier des avantages de l'apprentissage par transfert.

É entraînement et évaluation des modèles : Après les choix qui ont été fait, les modèles sont entraînés sur les données précédemment enregistrées dans un fichier « .npz » contenant les données d'entraînement et de test puis évalués en utilisant des métriques telles que la précision, le recall, le F1-score, la précision globale et le coefficient de Jaccard et des visualisations pour mesurer leur performance.

Déploiement en temps réel : Une fois que les métriques sont bonnes les modèles sont déployés en temps réel en utilisant des vidéos à partir des capteurs satellites pour faire la surveillance et la détection de changements en temps en réel.

Chapitre 2

Concepts Généraux

Résumé : Dans ce chapitre 2, nous abordons les généralités sur la télédétection, sur l'intelligence artificielle, sur les réseaux de neurones, sur la segmentation sémantique qui est un domaine de la vision par ordinateur et enfin terminons par une analyse comparative entre l'apprentissage supervisé, non supervisé et mixte. .

Sommaire

2.1	Généralités sur la Télédétection	14
2.1.1	Définition de la Télédétection :	14
2.1.2	Objectifs et Domaines d'Application :	18
2.1.3	Indices de Changement :	19
2.2	Généralités sur l'Intelligence Artificielle	19
2.2.1	Définition de l'Intelligence Artificielle :	20
2.2.2	Sous-Disciplines de l'IA :	20
2.2.3	Applications de l'IA :	21
2.2.4	Apprentissage Automatique et Apprentissage Pro- fond :	21
2.2.5	Importance de l'IA :	21

CHAPITRE 2. CONCEPTS GÉNÉRAUX

2.2.6	Défis et Éthique :	22
2.3	Généralités sur les réseaux de neurones	22
2.3.1	Réseaux de Neurones Artificiels (ANNs)	23
2.3.2	Propagation vers l'Avant (Forward Propagation) .	23
2.3.3	Rétropropagation (Backpropagation)	24
2.4	Généralités sur la segmentation en dans la vision par ordinateur :	26
2.4.1	La segmentation par instance :	26
2.4.2	La segmentation sémantique :	27
2.4.3	La segmentation panoptique :	29
2.5	Les types d'apprentissage en l'intelligence artificielle . . .	31
2.5.1	Apprentissage supervisé	32
2.5.2	Apprentissage non supervisé	33
2.5.3	Apprentissage mixte	34
2.5.4	L'apprentissage par transfert :	35

2.1 Généralités sur la Télédétection

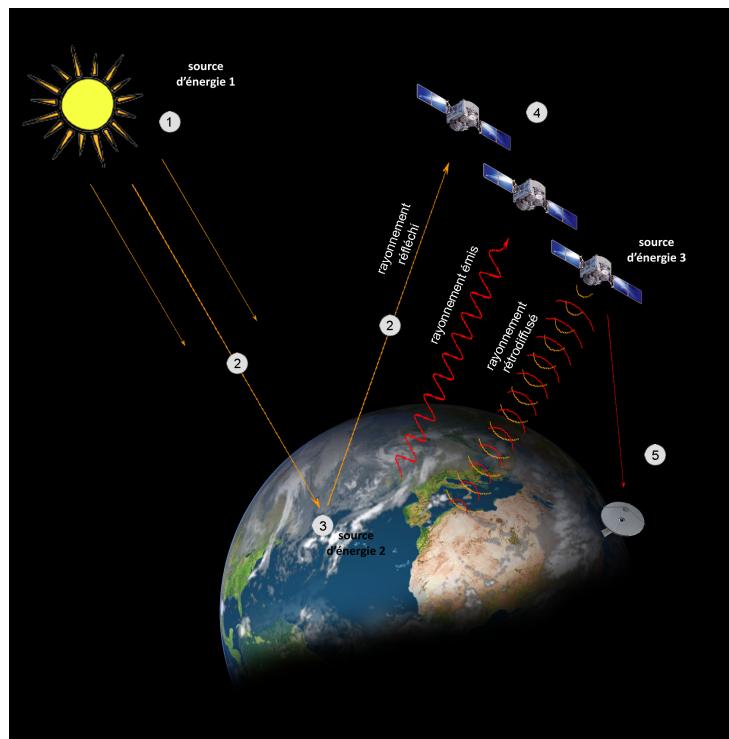


Figure 2.1 – Processus de la Télédétection

La télédétection spatiale permet d'observer le système terre / océans / atmosphère en s'appuyant sur des méthodes d'acquisition de l'information à distance basées sur les propriétés du rayonnement électromagnétique. Les capteurs satellitaires permettent de mesurer l'énergie réfléchie et/ou émise par la surface de la Terre, les océans ou par l'atmosphère, ce qui permet d'analyser et de suivre l'évolution de ces milieux au cours du temps [1].

2.1.1 Définition de la Télédétection :

La télédétection, également appelée observation de la Terre, est le processus d'acquisition d'informations sur la Terre à partir de capteurs situés à distance,

CHAPITRE 2. CONCEPTS GÉNÉRAUX

souvent dans l'espace ou à bord d'aéronefs. Ces capteurs peuvent enregistrer une variété de données, y compris des images, des spectres électromagnétiques, des températures et d'autres paramètres [38].

L'observation de la Terre par les satellites peut être décomposée en cinq étapes ou processus depuis la source d'énergie qui éclaire la surface terrestre jusqu'à la réception des données par la station satellite au sol Figure 2.1

1. La source d'énergie : L'observation de la Terre depuis les satellites implique inévitablement l'utilisation d'une source d'énergie. Trois sources d'énergie différentes sont mises à contribution en télédétection. La première, la plus répandue et couramment employée, est l'énergie solaire provenant du soleil qui éclaire la surface terrestre. La portion du rayonnement réfléchie par la surface terrestre est ensuite captée et enregistrée par les capteurs satellitaires. Ce processus illustre la télédétection optique, qui se situe principalement dans les domaines du spectre visible et du proche infrarouge. Toutefois, il est essentiel de noter que la source d'énergie n'est pas exclusivement limitée au rayonnement solaire.

La surface terrestre peut également agir comme une source d'énergie, émettant un rayonnement qui peut être détecté et enregistré par les capteurs des satellites. Ce scénario s'inscrit dans le contexte de la télédétection thermique ou de la télédétection par micro-ondes passives. En télédétection thermique, les objets et les terrains émettent naturellement un rayonnement infrarouge thermique en fonction de leur température. Les capteurs satellitaires sont en mesure de capturer ce rayonnement pour fournir des informations sur la température des objets et des surfaces terrestres. Cette méthode est précieuse pour des applications telles que la surveillance des températures des océans, la détection

CHAPITRE 2. CONCEPTS GÉNÉRAUX

des incendies de forêt et l'analyse des caractéristiques thermiques des zones urbaines.

La télédétection par micro-ondes passives, quant à elle, se base sur la réception des micro-ondes émises par la surface terrestre. Ces micro-ondes peuvent être utilisées pour étudier diverses propriétés, telles que l'humidité du sol, la végétation, la couverture nuageuse et même la topographie. Cette technique est notamment employée pour des applications liées à l'agriculture, à la météorologie et à la surveillance environnementale Figure 2.2.

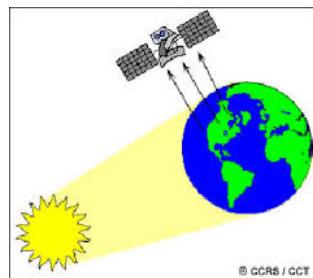


Figure 2.2 – Télédétection passive

Le capteur satellite peut également jouer le rôle de source d'énergie en émettant un rayonnement vers la surface terrestre au moyen d'une antenne spéciale, puis en enregistrant la portion du rayonnement rétrodiffusée par les objets et les surfaces. Cette procédure correspond à la télédétection active, particulièrement dans le domaine des hyperfréquences.

En télédétection active par hyperfréquences, le capteur satellite génère ses propres signaux électromagnétiques, qu'il envoie vers la Terre. Ces signaux interagissent avec les objets et les surfaces terrestres, puis sont rétrodiffusés vers le capteur. En analysant les caractéristiques des signaux rétrodiffusés, on peut extraire des informations précieuses

sur la topographie, la structure des objets, et même les propriétés des matériaux.

La télédétection active est largement utilisée pour des applications variées, comme la cartographie en relief, la surveillance des changements environnementaux, la détection des glissements de terrain, et la collecte de données topographiques précises. Elle offre des avantages essentiels pour l'observation de la Terre depuis l'espace Figure 2.3.

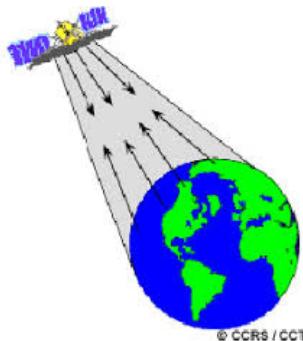


Figure 2.3 – Télédétection active

2. Les interactions du rayonnement avec l'atmosphère : Pendant son parcours depuis la source d'énergie jusqu'à la surface terrestre, puis jusqu'au capteur satellitaire, le rayonnement est soumis à des interactions avec les composants de l'atmosphère. Ces interactions sont d'une grande importance car elles peuvent altérer la qualité et la précision des données collectées.
3. Les interactions du rayonnement avec la surface terrestre : Une fois que le rayonnement atteint la surface terrestre, il entre en contact avec celle-ci. La nature de ces interactions dépend à la fois des caractéristiques du rayonnement lui-même et des propriétés spectrales des surfaces terrestres qu'il traverse. Ces interactions peuvent inclure la réflexion, l'absorption, la diffusion, et d'autres phénomènes qui fournissent des

CHAPITRE 2. CONCEPTS GÉNÉRAUX

informations cruciales sur les objets et les surfaces observés.

4. L'enregistrement du signal par le capteur satellitaire : L'énergie rayonnée par la surface terrestre, qu'elle soit réfléchie, émise ou rétrodiffusée, est captée par le capteur satellitaire. Une fois capturée, cette énergie est enregistrée et discrétisée par le capteur. On distingue deux types principaux de capteurs en fonction de leur source d'énergie :
 - Les capteurs passifs, qui exploitent les propriétés de réflexion du rayonnement solaire dans le domaine optique (visible et proche infrarouge), ainsi que les caractéristiques de l'émission dans l'infrarouge thermique et le domaine des micro-ondes. Ces capteurs sont utilisés pour caractériser les objets et les surfaces terrestres.
 - Les capteurs actifs, tels que les radars, utilisent leur propre source d'énergie pour émettre un rayonnement vers la surface terrestre, puis enregistrent la portion de rayonnement rétrodiffusée. Cette méthode est couramment employée dans le domaine des hyperfréquences.
5. La transmission et la réception des données : Lorsque le satellite se trouve dans la portée d'une station satellite terrestre, il transmet les données collectées. À ce stade, les données subissent les premiers traitements, consistant en l'application de corrections radiométriques et géométriques aux images brutes [25]. Ces traitements sont essentiels pour améliorer la qualité des données recueillies [1].

2.1.2 Objectifs et Domaines d'Application :

La télédétection trouve des applications dans de nombreux domaines, tels que la cartographie, la gestion des ressources naturelles, la surveillance environnementale, l'agriculture, la météorologie, la gestion des catastrophes, l'urbanisme, et bien plus encore. Elle permet de collecter des informations à

grande échelle et à intervalles réguliers, ce qui en fait un outil précieux pour surveiller les changements sur la Terre.

2.1.3 Indices de Changement :

Plusieurs indices sont utilisés en télédétection pour extraire des informations spécifiques des images, par exemple, l'Indice de Végétation par Différence Normalisée (NDVI) est calculé comme suit :

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (2.1)$$

où NIR représente la réflectance dans le proche infrarouge et Red représente la réflectance dans la bande rouge. Ces indices permettent d'obtenir des informations sur la végétation, la qualité de l'eau, etc [14].

La télédétection offre des moyens puissants d'acquérir des informations sur la Terre à grande échelle, ce qui en fait un outil essentiel pour la surveillance environnementale, la gestion des ressources naturelles et de nombreux autres domaines. Dans le contexte de ce mémoire, la télédétection est une composante cruciale pour l'acquisition des images satellitaires qui seront utilisées dans l'application de modèles de Deep Learning pour la surveillance et la détection des changements d'occupations du sol.

2.2 Généralités sur l'Intelligence Artificielle

L'intelligence artificielle (IA) est un domaine multidisciplinaire en plein essor qui vise à créer des systèmes informatiques capables de réaliser des tâches qui, traditionnellement, nécessitent l'intelligence humaine. L'IA englobe un large éventail de techniques et d'approches pour simuler la pensée humaine, la

CHAPITRE 2. CONCEPTS GÉNÉRAUX

perception, l'apprentissage, la résolution de problèmes et la prise de décision. Voici un aperçu des principaux concepts liés à l'intelligence artificielle :

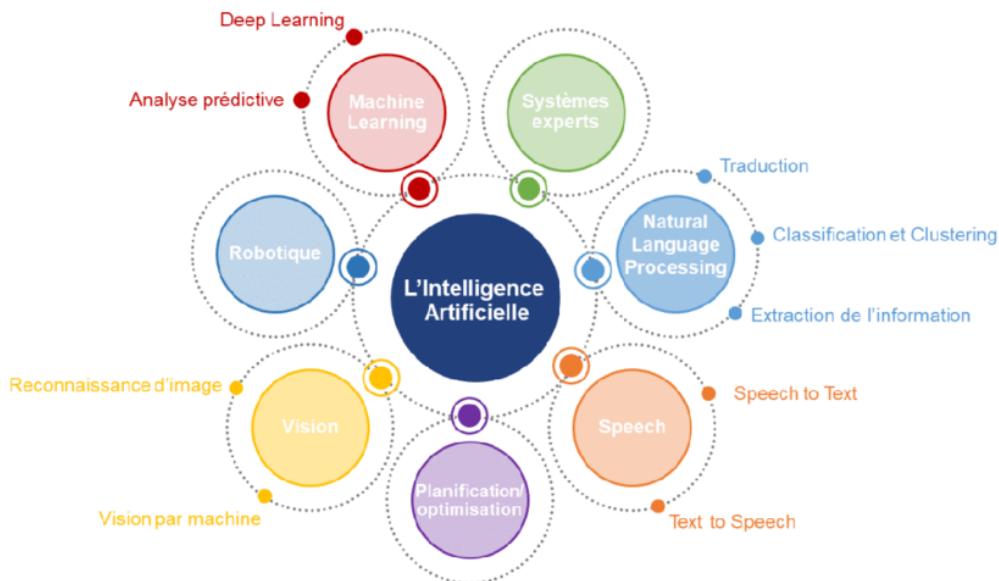


Figure 2.4 – Les sous-disciplines de l'intelligence artificielle

2.2.1 Définition de l'Intelligence Artificielle :

L'IA se réfère à la capacité d'un système informatique à effectuer des tâches qui exigent généralement une intelligence humaine. Cela inclut des activités telles que la compréhension du langage naturel, la reconnaissance de motifs, l'apprentissage, la planification, la résolution de problèmes, et bien d'autres.

2.2.2 Sous-Disciplines de l'IA :

L'IA englobe plusieurs sous-domaines, notamment l'apprentissage automatique (machine learning), la vision par ordinateur, le traitement du langage naturel (NLP), la robotique, la planification automatique, les systèmes experts,

CHAPITRE 2. CONCEPTS GÉNÉRAUX

etc. Chacune de ces sous-disciplines se concentre sur des aspects spécifiques de l'intelligence artificielle 2.4.

2.2.3 Applications de l'IA :

L'IA trouve des applications dans divers domaines, notamment la santé, les transports, la finance, la sécurité, la recherche, l'industrie, et bien d'autres. Par exemple, dans le domaine médical, l'IA est utilisée pour la détection précoce des maladies, tandis que dans l'industrie, elle peut optimiser la chaîne d'approvisionnement et les opérations de fabrication.

2.2.4 Apprentissage Automatique et Apprentissage Profond :

L'apprentissage automatique (machine learning) est une branche clé de l'IA qui consiste à développer des algorithmes capables d'apprendre à partir de données et d'améliorer leurs performances au fil du temps. L'apprentissage profond (deep learning) est une sous-discipline de l'apprentissage automatique qui utilise des réseaux de neurones artificiels profonds pour résoudre des tâches complexes, notamment la vision par ordinateur et le traitement du langage naturel.

2.2.5 Importance de l'IA :

L'IA a transformé de nombreux aspects de notre vie quotidienne et de l'industrie. Elle permet d'automatiser des tâches fastidieuses, d'analyser d'énormes quantités de données, de prendre des décisions basées sur des données et de résoudre des problèmes complexes.

CHAPITRE 2. CONCEPTS GÉNÉRAUX

2.2.6 Défis et Éthique :

L'IA soulève également des questions éthiques et de sécurité, notamment en ce qui concerne la confidentialité des données, la prise de décision automatisée, la transparence des algorithmes et les biais algorithmiques. La recherche sur l'éthique de l'IA est devenue une composante cruciale du domaine.

En résumé, l'intelligence artificielle représente une frontière fascinante de la science informatique et de la technologie, offrant un potentiel considérable pour l'innovation dans une multitude de domaines. Dans le contexte de mon mémoire, l'IA sera utilisée pour développer un modèle d'analyse des images satellitaires afin de surveiller et de détecter les changements d'occupations du sol dans la région de Richard Toll, montrant ainsi comment cette technologie peut être appliquée à des problèmes réels liés à l'utilisation des terres et à la surveillance environnementale.

2.3 Généralités sur les réseaux de neurones

Les réseaux de neurones sont des modèles d'intelligence artificielle inspirés du cerveau humain. Ils sont constitués de couches de neurones artificiels interconnectés qui traitent des données, apprennent des modèles et effectuent des tâches complexes. Grâce au deep learning, ces réseaux ont connu une révolution majeure et sont désormais utilisés dans des domaines variés tels que la vision par ordinateur, la reconnaissance vocale et la traduction automatique. Ils jouent un rôle essentiel dans la résolution de problèmes complexes en tirant des informations à partir de grandes quantités de données, ce qui ouvre des perspectives passionnantes en intelligence artificielle.

2.3.1 Réseaux de Neurones Artificiels (ANNs)

Les réseaux de neurones artificiels, ou ANNs (Artificial Neural Networks), sont une classe de modèles informatiques inspirés du fonctionnement du cerveau humain. Ils sont utilisés dans une variété d'applications, de la vision par ordinateur à la reconnaissance vocale en passant par la prédiction de séries temporelles. Les ANNs sont constitués de couches de neurones interconnectés qui sont capables d'apprendre à partir de données Figure 2.5.

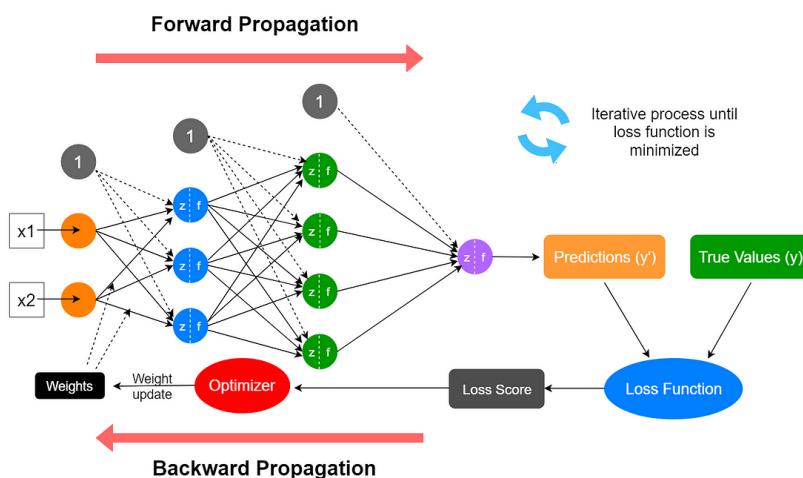


Figure 2.5 – Réseau de neurones artificiels

2.3.2 Propagation vers l'Avant (Forward Propagation)

La propagation vers l'avant est la phase dans laquelle les données sont introduites dans le réseau de neurones, et les calculs sont effectués pour obtenir une sortie. Voici comment cela fonctionne :

1. Initialisation : Chaque neurone dans le réseau est associé à un poids (w) et un biais (b) initial. Les poids représentent l'importance des connexions entre

les neurones, et les biais sont des valeurs constantes.

2. Entrée : Les données d'entrée (x) sont introduites dans la première couche du réseau, appelée couche d'entrée. Chaque neurone de cette couche est associé à une caractéristique spécifique des données.

3. Calcul : Les données et les poids sont utilisés pour calculer une somme pondérée (z) pour chaque neurone de la couche suivante. Cette somme est ensuite passée à travers une fonction d'activation (σ), généralement une fonction non linéaire comme la fonction sigmoïde ou la fonction ReLU (Rectified Linear Unit).

$$z = \sum_{i=1}^n w_i x_i + b \quad (2.2)$$

$$a = \sigma(z) \quad (2.3)$$

4. Propagation : Ce processus de calcul est répété pour chaque couche du réseau, en propageant les valeurs des neurones de la couche précédente vers la couche suivante jusqu'à ce que nous atteignons la couche de sortie.

5. Sortie : La couche de sortie génère la prédiction finale du réseau. Cela peut être une valeur unique pour les tâches de régression ou une distribution de probabilité pour les tâches de classification.

2.3.3 Rétropropagation (Backpropagation)

La rétropropagation est le processus par lequel le réseau ajuste ses poids (w) et biais (b) pour minimiser l'erreur entre les prédictions et les valeurs réelles. Voici comment cela fonctionne :

CHAPITRE 2. CONCEPTS GÉNÉRAUX

6. Calcul de l'Erreur : Tout d'abord, une fonction de perte (L) est utilisée pour mesurer l'erreur entre les prédictions du réseau et les valeurs réelles. Par exemple, pour la régression, on peut utiliser l'erreur quadratique moyenne (MSE), et pour la classification, on peut utiliser l'entropie croisée.

$$L = \text{fonction_de_perte}(\text{prédictions}, \text{valeurs_réelles}) \quad (2.4)$$

7. Calcul du Gradient : Le gradient de la fonction de perte par rapport aux poids ($\nabla_w L$) et biais ($\nabla_b L$) est calculé à l'aide de la règle de dérivation de la chaîne (règle du gradient). Cela permet de déterminer comment les poids et biais doivent être ajustés pour minimiser l'erreur.

$$\nabla_w L = \frac{\partial L}{\partial w} \quad (2.5)$$

$$\nabla_b L = \frac{\partial L}{\partial b} \quad (2.6)$$

8. Mise à Jour des Paramètres : Les poids et biais du réseau sont ajustés en fonction du gradient calculé. Le taux d'apprentissage (α) est utilisé pour contrôler la taille des ajustements. Une petite valeur de taux d'apprentissage peut rendre l'apprentissage plus stable, mais plus lent, tandis qu'une grande valeur peut accélérer l'apprentissage, mais le rendre instable.

$$w \leftarrow w - \alpha \nabla_w L \quad (2.7)$$

$$b \leftarrow b - \alpha \nabla_b L \quad (2.8)$$

9. Répétition : Les étapes 6 à 8 sont répétées pour chaque lot de données d'entraînement (mini-batch) dans l'ensemble d'entraînement. Ce processus est itératif et continue jusqu'à ce que le réseau atteigne une performance satisfaisante ou un nombre fixe d'itérations (époques).

Le processus de backpropagation permet au réseau d'ajuster ses paramètres pour s'adapter aux données d'entraînement et à apprendre des motifs complexes dans les données. Il peut y avoir de nombreuses couches et des millions de paramètres dans un réseau de neurones, mais grâce à la propagation vers l'avant et à la rétropropagation, le réseau peut apprendre à résoudre des problèmes complexes.

2.4 Généralités sur la segmentation en dans la vision par ordinateur :

La segmentation en vision par ordinateur est un domaine fondamental qui vise à extraire des informations significatives à partir d'images ou de vidéos. Elle consiste à diviser une image en régions ou en objets ayant des caractéristiques similaires. Ce processus est essentiel pour de nombreuses applications telles que la détection d'objets, la reconnaissance de formes, la surveillance vidéo, la réalité augmentée, et bien d'autres. La segmentation peut être effectuée à différents niveaux, en fonction de la complexité de la tâche. On distingue principalement trois types de segmentation : la segmentation par instance, la segmentation sémantique et la segmentation panoptique.

2.4.1 La segmentation par instance :

La segmentation par instance représente une avancée majeure en vision par ordinateur, propulsant notre capacité à comprendre et analyser les scènes visuelles vers de nouveaux sommets. Cette technique va bien au-delà de la segmentation sémantique, qui se contente de catégoriser les régions d'une image, en franchissant une étape décisive : elle distingue non seulement les



Figure 2.6 – Segmentation par instance

catégories d'objets, mais elle identifie aussi chaque occurrence individuelle de ces objets, quelle que soit leur similitude. Concrètement, la segmentation par instance assigne un identifiant unique à chaque objet détecté dans une image. Cette distinction est fondamentale. Elle signifie que si l'image contient plusieurs personnes, voitures, animaux ou autres objets, la segmentation par instance les identifiera indépendamment les uns des autres, même s'ils appartiennent à la même catégorie. Ainsi, chaque piéton, chaque voiture, chaque animal, et ainsi de suite, sera traité comme une entité individuelle, avec un identifiant spécifique. L'importance de cette distinction devient particulièrement évidente dans des domaines tels que la navigation autonome, où la sécurité et la précision de la détection d'objets sont primordiales. Dans un scénario de conduite autonome, il est essentiel de suivre chaque piéton de manière indépendante, même s'ils font tous partie de la catégorie "piéton." Cela permet d'anticiper leurs mouvements et de prendre des décisions en conséquence, garantissant un environnement de conduite sûr pour tous les usagers de la route Figure 2.6.

2.4.2 La segmentation sémantique :

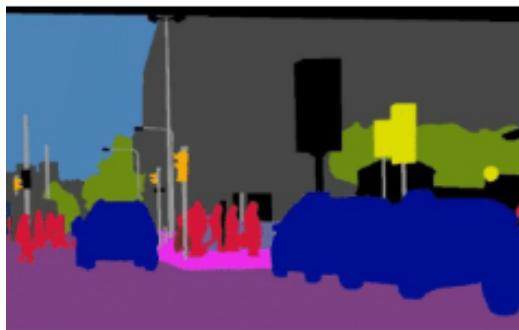


Figure 2.7 – Segmentation sémantique

La segmentation sémantique est une technique puissante en vision par ordinateur qui s'attaque au défi de comprendre en profondeur le contenu visuel d'une image. Elle se distingue par sa capacité à décomposer une image en de multiples régions, comme si elle effectuait une sorte de découpage virtuel, où chaque morceau se voit attribuer une étiquette sémantique. Concrètement, ces étiquettes correspondent aux catégories d'objets, d'entités ou de régions qui sont présentes dans l'image. Imaginons une scène urbaine pour mieux comprendre le rôle essentiel de la segmentation sémantique. Dans une telle scène, une image pourrait contenir des voitures, des arbres, des bâtiments, des piétons, des panneaux de signalisation, et bien plus encore. La segmentation sémantique aurait pour mission de distinguer chaque pixel de l'image et de l'associer à l'une de ces catégories. Ainsi, les pixels correspondant à une voiture seraient regroupés sous l'étiquette "voiture," tandis que ceux associés à un arbre seraient étiquetés comme "arbre," et ainsi de suite. Cela permet d'atteindre un niveau de compréhension automatisée de l'image qui était auparavant difficile à réaliser. Cette compréhension fine et granulaire de la scène ouvre la porte à une multitude d'applications passionnantes. Par exemple, dans le domaine de la détection d'objets, la segmentation sémantique peut servir à identifier les zones où des objets d'intérêt sont présents. Dans la

cartographie, elle peut être utilisée pour extraire des informations précises sur la distribution des éléments dans une région donnée. Dans la surveillance de la circulation, elle peut contribuer à analyser la circulation routière et à identifier des anomalies potentielles Figure 2.7.

2.4.3 La segmentation panoptique :



Figure 2.8 – Segmentation panoptique

La segmentation panoptique constitue une avancée récente et prometteuse en vision par ordinateur, combinant habilement les avantages des deux techniques précédentes, à savoir la segmentation sémantique et la segmentation par instance. Concrètement, elle va plus loin en fournissant des informations aussi bien sur la catégorie sémantique des objets que sur leurs instances individuelles dans une image donnée. En d'autres termes, la segmentation panoptique offre une double perspective : elle indique quels objets sont présents dans l'image, les classant dans des catégories sémantiques spécifiques, tout en distinguant chaque occurrence individuelle de ces objets. Cette dualité d'informations permet d'obtenir une compréhension plus profonde de l'environnement visuel, offrant un aperçu détaillé à la fois sur la nature des objets présents et sur la manière dont ils sont répartis dans l'image. L'importance de la segmentation panoptique ne peut être surestimée, en particulier dans des domaines critiques

CHAPITRE 2. CONCEPTS GÉNÉRAUX

tels que la conduite autonome. Dans ce contexte, une perception précise et complète de l'environnement est cruciale pour garantir la sécurité des passagers et des autres usagers de la route. En fournissant une vue d'ensemble détaillée de la scène, cette méthode permet aux véhicules autonomes de prendre des décisions plus éclairées et de réagir de manière plus anticipée aux divers éléments de leur environnement. La segmentation panoptique a rapidement gagné en importance et suscité un vif intérêt dans le domaine de la recherche en vision par ordinateur. Cette approche novatrice ouvre des perspectives passionnantes pour des applications allant de la conduite autonome à la surveillance automatisée et bien au-delà, renforçant ainsi notre capacité à interagir de manière intelligente avec le monde visuel qui nous entoure Figure 2.8.

En plus de ces types de segmentation, il existe d'autres approches comme la segmentation basée sur la région, qui regroupe les pixels en régions homogènes en termes de couleurs ou de textures. La segmentation par contour vise à détecter les limites des objets en se basant sur les variations de luminosité ou de couleur. La segmentation hiérarchique crée une hiérarchie de régions, permettant de passer d'une vue globale à une vue plus détaillée de l'image.

la segmentation en vision par ordinateur est un domaine vaste et diversifié, offrant de nombreuses techniques et approches pour extraire des informations significatives à partir d'images ou de vidéos. Chacun de ces types de segmentation joue un rôle essentiel dans différentes applications, contribuant ainsi à l'avancée de la vision par ordinateur [32].

2.5 Les types d'apprentissage en l'intelligence artificielle

L'apprentissage machine est un pilier fondamental de l'intelligence artificielle, et il existe diverses approches dont nous allons aborder que quatre d'entre eux qui sont l'apprentissage supervisé, l'apprentissage non supervisé, l'apprentissage mixte et l'apprentissage par transfert. Ces méthodes visent à permettre aux systèmes informatiques d'acquérir des connaissances à partir des données, mais elles diffèrent considérablement dans leurs objectifs et leurs applications. Cette analyse comparative examine ces trois approches pour mieux comprendre leurs caractéristiques, leurs avantages et leurs limites dans le domaine de l'intelligence artificielle. Elle explore comment l'apprentissage supervisé, en utilisant des données étiquetées, permet aux systèmes de prédir des résultats précis, comment l'apprentissage non supervisé identifie des structures et des modèles cachés dans les données non étiquetées, et comment l'apprentissage mixte combine ces deux approches pour des solutions plus complètes. En comprenant ces différences, on peut choisir la méthode qui convient le mieux à un problème donné en intelligence artificielle.

2.5.1 Apprentissage supervisé

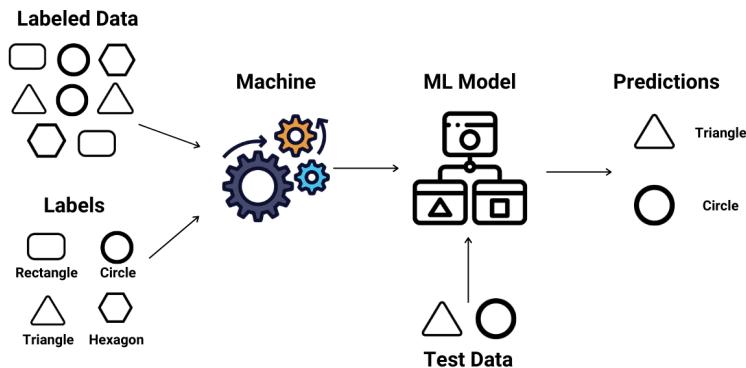


Figure 2.9 – Apprentissage supervisé

L'apprentissage supervisé est l'un des paradigmes les plus couramment utilisés en apprentissage machine. Dans ce type d'apprentissage, le modèle est formé sur un ensemble de données étiquetées, ce qui signifie que chaque exemple de données est associé à une étiquette ou à une classe spécifique. Le modèle apprend à faire correspondre les caractéristiques des données d'entrée à leurs étiquettes correspondantes. Les principales caractéristiques de l'apprentissage supervisé comprennent :

Types d'algorithmes : Les algorithmes d'apprentissage supervisé incluent la régression linéaire, les machines à vecteurs de support (SVM), les arbres de décision, les réseaux de neurones, et bien d'autres.

Applications courantes : L'apprentissage supervisé est utilisé dans de nombreuses applications, telles que la classification d'images, la prédiction de prix, la détection de fraudes, la reconnaissance vocale et la traduction automatique.

Avantages : Il offre une grande précision dans la prédiction des étiquettes sur de nouvelles données, une grande adaptabilité aux différents types de problèmes et il est largement utilisé dans les applications du monde réel.

Inconvénients : Il nécessite un ensemble de données étiquetées, ce qui peut être coûteux à obtenir. De plus, il peut ne pas fonctionner efficacement sur des données non linéaires ou hautement dimensionnelles.

2.5.2 Apprentissage non supervisé

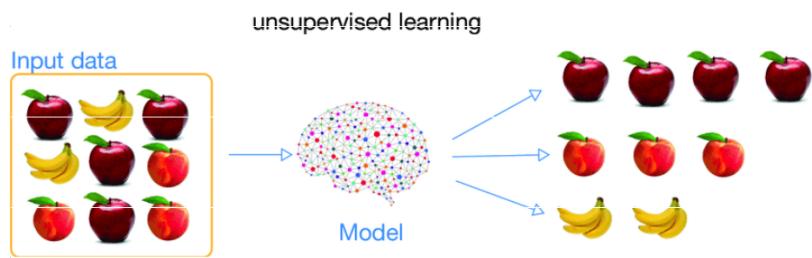


Figure 2.10 – Apprentissage non supervisé

L'apprentissage non supervisé est utilisé lorsque les données ne sont pas étiquetées, ce qui signifie qu'aucune classe n'est attribuée aux exemples. Dans ce cas, l'objectif est généralement de découvrir des structures ou des modèles intrinsèques aux données. Les principales caractéristiques de l'apprentissage non supervisé comprennent :

Techniques de clustering : Les techniques de clustering, telles que la classification ascendante hiérarchique (CAH) et la k-moyenne (k-mean), sont utilisées pour regrouper les données en fonction de leurs similitudes intrinsèques.

CHAPITRE 2. CONCEPTS GÉNÉRAUX

Applications courantes : L'apprentissage non supervisé est utilisé pour la segmentation d'images, la recommandation de contenu, l'analyse de données biologiques, la réduction de dimensionnalité et la détection d'anomalies.

Avantages : Il peut révéler des informations cachées dans les données, identifier des groupes ou des tendances naturels, et il ne nécessite pas d'étiquetage coûteux. Inconvénients : L'interprétation des résultats peut être subjective, et il peut être difficile de déterminer le nombre optimal de clusters.

2.5.3 Apprentissage mixte

L'apprentissage mixte combine des éléments de l'apprentissage supervisé et non supervisé. Il est souvent utilisé lorsque l'ensemble de données est partiellement étiqueté, ce qui signifie qu'une partie des exemples est étiquetée, tandis que d'autres ne le sont pas. L'apprentissage mixte peut être divisé en deux catégories principales :

Semi-Supervisé : Dans cette approche, le modèle est formé à la fois sur des données étiquetées et non étiquetées. Les données étiquetées sont utilisées pour guider l'apprentissage, tandis que les données non étiquetées peuvent aider à améliorer la généralisation du modèle.

Renforcement : L'apprentissage par renforcement est une forme d'apprentissage mixte où un agent apprend à interagir avec un environnement pour maximiser une récompense. Il est couramment utilisé dans les domaines de la robotique et des jeux.

L'apprentissage mixte offre l'avantage de tirer parti des données non étiquetées pour améliorer les performances du modèle tout en utilisant les données

CHAPITRE 2. CONCEPTS GÉNÉRAUX

étiquetées lorsque cela est possible.

2.5.4 L'apprentissage par transfert :

L'apprentissage par transfert en intelligence artificielle consiste à utiliser des modèles pré-entraînés sur une tâche source pour améliorer les performances sur une tâche cible similaire. Il permet de gagner du temps et des ressources en transférant des connaissances d'une tâche à une autre, ce qui est particulièrement utile avec des ensembles de données limités. L'adaptation du modèle peut se faire en ajustant les poids du modèle, en ajoutant des couches supplémentaires, ou en gelant certaines parties du modèle. Cette technique est couramment utilisée en vision par ordinateur, en traitement du langage naturel et dans d'autres domaines de l'apprentissage automatique. Elle offre des avantages significatifs tout en présentant des défis potentiels, notamment la nécessité de trouver des tâches sources appropriées et de gérer correctement l'adaptation [28].

Chapitre 3

Etat de l'Art

Résumé : Dans ce chapitre, nous commencerons par l'étude des travaux existants avant de plonger dans le choix de notre approche et pour enfin terminer avec une étude comparative des différentes approches.

Sommaire

3.1	Méthodes traditionnelles et méthodes de l'apprentissage profonds en télédétection pour la surveillance et la détection des changements d'occupations du sol.	37
3.2	Critères de sélection et méthodologie de choix	44
3.3	Avantages et limites des différentes approches	46

La revue de la littérature revêt une importance cruciale dans le processus de recherche. Elle permet de contextualiser la nouvelle étude en examinant ce qui a déjà été fait dans le domaine. Cette étape fournit un aperçu des connaissances existantes, identifie les lacunes de recherche et aide les chercheurs à comprendre les méthodes et les résultats des études antérieures. En outre, elle renforce la crédibilité de la recherche en montrant qu'elle s'appuie sur une base solide et établie. En somme, la revue de la littérature éclaire, guide et justifie la

nouvelle recherche.

3.1 Méthodes traditionnelles et méthodes de l'apprentissage profonds en télédétection pour la surveillance et la détection des changements d'occupations du sol.

Dans cette partie nous faisons la revue de la littérature sur les méthodes traditionnelles utilisée en télédétection et sur l'utilisation de l'apprentissage profond dans la télédétection pour la surveillance et la détection des changements d'occupations du sol.

3.1.1 Méthodes traditionnelles utilisées en télédétection : Une Analyse Approfondie

Dans un premier article, on décrit une méthode de cartographie de l'occupation du sol et de ses changements dans la région méridionale de la Sénégambie en utilisant des données de télédétection [17]. Cette approche rétrospective se fonde sur la comparaison des signatures spectrales considérées comme des "références", qui proviennent d'observations sur le terrain, avec celles des classes obtenues à partir de la classification multispectrale non supervisée d'images prises à des dates antérieures. Le résultat de cette méthodologie est la création de cartes représentant les principales catégories d'occupation du sol à chaque date d'étude, ainsi que l'analyse de leur évolution temporelle. Par contre ces techniques présentent des limites telles que la perte de la signification sémantique des classes, la création de classes principalement basées

sur des caractéristiques spectrales, et l'absence d'informations contextuelles pour une interprétation précise des changements observés.

La surveillance et la détection des changements d'occupations du sol se fait souvent avec des méthodes telles la comparaison pixel à pixel ou "image-image". Les méthodes de comparaison "image-image" sont basées sur la comparaison directe des valeurs radiométriques des pixels entre deux images prises à différentes dates. Dans d'autres cas, on utilise des techniques de statistique multivariées telles que l'ACP combinée à la comparaison pixel à pixel [13], comme dans une étude où des chercheurs utilisaient une technique qui combine deux méthodes statistiques simples, la technique de la différence en pourcentage (PD) et l'analyse en composantes principales (ACP), pour fusionner les données et détecter les changements dans les images résultantes. Elle offre des performances quantitatives raisonnables, avec une précision globale allant jusqu'à 91% et un coefficient Kappa allant jusqu'à 66%. Malgré leurs performances raisonnables, elles sont souvent confrontées à des défis, notamment la nécessité de normaliser les données radiométriques, ce qui peut être complexe lorsque les images sont prises à des saisons différentes et présentent une variabilité de qualité radiométrique. Elles peuvent être sensibles aux variations des conditions atmosphériques, ne tiennent pas compte de la sémantique des classes, ce qui peut entraîner des faux positifs ou des résultats moins précis. De plus, elles ont tendance à être limitées dans leur capacité à identifier la nature des changements, se concentrant principalement sur la détection de changements dans les valeurs des pixels [13].

La recherche antérieure a identifié ces limites, notamment dans des études telles que celles de Singh (1986), Alphan (2003), Coppin et al. (2004) et Yuan et al. (2005). Ces auteurs ont souligné les défis liés à la normalisation radio-

métrique et au manque d'informations sur la nature des changements dans les méthodes de comparaison "image-image" [22]. En réponse aux limites des méthodes "image-image", les méthodes de comparaison post-classification ont été développées. Ces approches comparent les classifications de l'occupation du sol qui sont produites indépendamment pour chaque image à différentes dates. Elles sont moins sensibles aux variations saisonnières, car elles se basent sur les classes d'occupation du sol plutôt que sur les valeurs radiométriques brutes. De plus, elles permettent d'identifier la nature des changements, c'est-à-dire quelles classes d'occupation du sol ont changé entre les deux dates [22]. Des travaux antérieurs, tels que ceux de Gupta et al. (1985), ont exploré ces méthodes de comparaison post-classification. Ils ont montré que ces approches sont moins sensibles aux variations saisonnières et peuvent fournir des informations importantes sur la nature des changements dans les images satellitaires [22].

3.1.2 L'utilisation de l'apprentissage profond dans la télédétection pour la surveillance et la détection des changements d'occupations du sol.

La surveillance et la détection des changements d'occupations du sol suscitent un intérêt croissant, avec de nombreuses recherches examinant l'application de l'apprentissage profond à cette fin.

Dans une première étude, les chercheurs ont développé une architecture siamoise multimodale pour la détection de changement. Le réseau se compose de deux branches d'encodeur pour chaque capteur et d'un partie décodeur intégrant les fonctionnalités des couches précédentes. Chaque branche d'encodeur traite le signal bitemporel de sa modalité correspondante. échantillons,

SAR et optique multispectrale, en deux passes. Les fonctionnalités extraites sont transmises via des connexions sautées dans un U-Net comme mode puis concaténé [27]. Cette approche leur a valu des résultats qu'on pourrait qualifiés d'acceptable où ils ont une précision de 60.2%, un recall de 56.1% et un F1-score, qui est la moyenne harmonique des deux métriques, de 58.1% [27].

Dans une deuxième étude, les chercheurs ont testé plusieurs approches et parmi elles, il y'en a qui est basée sur le réseau siamois-CNN. Ce modèle prend deux images hétérogènes en entrée et utilise deux flux de réseau CNN parallèles mais non couplés pour extraire des descripteurs indépendants à partir de chaque patch d'entrée. Ces descripteurs sont ensuite combinés en une représentation multimodale unique, utilisée pour apprendre la fonction de coût de classification binaire. Cette approche est testé sur 5 datasets où ils ont eu des résultats commes suit : une accuracy de 87%, 86.5%, 98.7%, 87.7% et 83.7% respectivement sur les datasets SAR/Optical, Optical/Optical, SAR/Optical, Quickbird/IKONOS et Quickbird/IKONOS. Malgré ces résultats on peut remarquer le fait que les limitations potentielles peuvent inclure des performances variables en fonction des sources d'imagerie et des résolutions spatiales, ainsi que des contraintes de données spécifiques [35].

Dans une troisième étude, les chercheurs ont développé un algorithme de détection de changement sur des images multi-spectrales à haute résolution a été développé en utilisant une architecture de type U-Net au niveau des caractéristiques. Pour améliorer les performances de l'algorithme, une méthode de pan-sharpening à faible complexité a été introduite, permettant l'utilisation conjointe d'images panchromatiques et multi-spectrales. Les images HRMS ont été intégrées dans le modèle U-Net au niveau des caractéristiques, composé de deux étapes : un réseau de soustraction au niveau des caractéristiques et

le modèle U-Net lui-même. Le réseau de soustraction au niveau des caractéristiques a été utilisé pour extraire des images de différence dynamique (DI) à la fois pour les caractéristiques de bas niveau et de haut niveau. Les résultats obtenus ont démontré que l'algorithme de détection de changement sur les images HRMS surpassait les performances des méthodes existantes. Les mesures de performance ont révélé un score F-1 moyen de 0,62, un pourcentage de classification correcte (PCC) de 98,78% et un coefficient kappa de 61,60% pour les jeux de données de test. Ces résultats indiquent une précision élevée et une capacité de détection de changement efficace de l'algorithme proposé. Cependant, des limites potentielles telles que la dépendance aux données HRMS et la qualité des images panchromatiques et multi-spectrales peuvent influencer les performances dans différents scénarios [37].

Dans une quatrième étude, l'article explore l'utilisation de réseaux neuronaux convolutifs CNN pour classifier l'utilisation des terres à partir d'images orthorectifiées multispectrales (spectres visibles) de très haute résolution. Elle se base sur l'abondance de données VHR disponibles aujourd'hui. Les auteurs notent que l'apparence des caractéristiques terrestres (par exemple, forêts, parkings, aéroports, zones résidentielles, autoroutes) peut varier en fonction de nombreux facteurs, tels que l'heure de la prise de vue, les réglages du capteur, le traitement de l'image, et le contexte géographique et culturel. Le texte souligne également l'importance de tester la généralisabilité des classifieurs sur des ensembles de données non vus auparavant. Les résultats de la thèse montrent que l'utilisation de réseaux neuronaux convolutifs pour la classification de l'utilisation des terres à partir d'images VHR est prometteuse. En particulier, le texte mentionne un taux de précision moyen de 96% sur le jeu de données UC Merced en utilisant un réseau de neurones convolutionnel avec

CHAPITRE 3. ETAT DE L'ART

une validation croisée à 5 volets. De plus, ces résultats sont testés sur d'autres images VHR non liées, montrant ainsi que le modèle généralise bien au-delà des données d'entraînement initiales. Une des principales limites abordées est le besoin de larges ensembles de données étiquetées pour l'apprentissage profond, ce qui peut être difficile à obtenir pour la classification de l'utilisation des terres. Cependant, la thèse explore des techniques telles que le "refinement learning" pour réutiliser des réseaux préalablement formés sur d'autres tâches. Une autre limite est que les apparences des caractéristiques terrestres peuvent varier, ce qui peut rendre la classification plus complexe [36].

Dans une cinquième étude, [21]L'étude a utilisé les données chronologiques Sentinel-2 obtenues auprès de l'Agence spatiale européenne (ESA) pour analyser la classification de l'utilisation des sols. Les données comprenaient des images multispectrales dans 13 bandes spectrales couvrant le spectre visible, le proche infrarouge et l'infrarouge à ondes courtes. La résolution spatiale des données variait en fonction de la bande, allant de 10 m à 60 m. Les données des séries chronologiques couvraient l'année agronomique 2017-2018 et comprenaient 14 caractéristiques extraites de 30 pas temporels. Les données de terrain fournies par le ministère de l'Agriculture, du Développement rural, de l'Urgence climatique et de la Transition écologique ont été utilisées pour former et tester les modèles de classification. Une architecture bidirectionnelle de réseau de mémoire à long terme (BilsTM) a été utilisée pour la classification de l'utilisation des sols, avec deux couches BilsTM et une couche de sortie softmax. La carte de classification obtenue à l'aide du modèle 2-BilsTM a montré une distribution spatiale des classes d'utilisation des sols conformément au système d'identification des parcelles (LPIS) de l'Espagne. La carte identifiait différentes catégories telles que les arbres fruitiers et les agrumes dans les

CHAPITRE 3. ETAT DE L'ART

zones côtières, les terres arables de l'ouest avec diverses cultures céréalier, les cultures de riz à l'est près de l'Albufera Natural, et la végétation naturelle, les fruits secs et les vignobles principalement dans les zones intérieures. La majeure partie de la zone a été classée comme zone de végétation naturelle, suivie par les cultures permanentes et les cultures annuelles. L'algorithme de classification a atteint une précision globale de 98,7% par rapport à l'ensemble de tests, surpassant les autres modèles et algorithmes de classification. Les scores de précision, de rappel et F-1 ont été calculés pour chaque classe, la précision la plus élevée étant atteinte pour les rizières et la plus faible pour les arbres fruitiers. Le réseau 2-BilsTM s'est bien comporté pour les classes de végétation naturelle mais avait un pouvoir discriminatoire légèrement inférieur pour les cultures permanentes par rapport aux cultures annuelles.

Dans une sixième étude, Les chercheurs ont proposé une approche basée sur une grille non chevauchante pour entraîner le FCN-8 avec des poids VGG-16. Le réseau a été utilisé pour segmenter des images satellites en quatre classes distinctes : forêt, zones construites, terres agricoles et eau. Le FCN-8 prend des caractéristiques discriminantes apprises par l'encodeur à plus basse résolution et les projette sémantiquement dans l'espace des pixels à plus haute résolution, permettant ainsi une classification dense. L'ensemble de données utilisé pour l'expérimentation était le jeu de données Gaofen-2, qui contenait 150 images provenant de plus de 60 villes différentes en Chine. Une comparaison a été effectuée en utilisant des données de référence disponibles ainsi que des images segmentées à l'aide d'un logiciel SIG commercial appelé eCognition. Avec l'approche basée sur une grille non chevauchante proposée, le FCN-8 a obtenu des performances significativement meilleures par rapport au logiciel eCognition. La précision moyenne du modèle était de 91,0%, et

l’intersection moyenne sur l’union (IoU) était de 0,84. En revanche, le logiciel eCognition a obtenu une précision moyenne de 74,0% et un IoU moyen de 0,60. Bien que le FCN-8 ait montré des performances supérieures, cette méthode n’est pas sans limites. Les erreurs se produisent principalement aux frontières des classes d’utilisation des terres et de couverture des terres. Ces erreurs aux limites sont analysées en détail. Les performances du modèle peuvent dépendre fortement de la qualité et de la résolution des images satellites utilisées. Des images de moindre qualité pourraient entraîner une baisse de la précision du modèle. Les performances peuvent varier en fonction de la diversité géographique des régions étudiées [7]

3.2 Critères de sélection et méthodologie de choix

Lorsqu’il s’agit de choisir des approches, il est essentiel de prendre en considération plusieurs critères pour assurer la pertinence et l’efficacité des méthodes employées dans la surveillance et la détection des changements d’occupations du sol.

Dans le cadre de mon mémoire, nous avons opté pour une approche basée sur l’apprentissage profond qui a montré ses preuves et son efficacité d’après plusieurs études précédentes, comme celles mentionnées ci-dessus. Nous avons voulu classifier de manière efficace nos différentes classes dans nos images géospatio-temporelles, sans perdre le contexte de ces dernières d’où la nécessité de faire de l’apprentissage supervisé dans notre cas, ce qui nous permettra de faire une comparaison des images géospatio-temporelles déjà classifiées dans le but de mettre en évidence les zones ayant subi un changement sans perdre d’informations des classes concernées et de garder la sémantique des classes qui est très importante dans ce processus. D’où l’adoption de la segmentation

sémantique dans notre cas d'étude qui nous permet de bien mettre en évidence la sémantique de nos classes.

Puis nous avons choisi d'utiliser deux architectures qui sont adaptées pour la segmentation sémantique, les architectures de U-Net qui a montré ses preuves dans l'étude de Wiratama et al. [37], et FCN (Fully Convolutional Network) qui lui aussi montre de bonnes performances que peut en témoigner l'article de Abu Bakar et al. [7]. Après Nous avons procédé à leurs combinaisons à des modèles pré-entraînés.

U-Net est un réseau de neurones à convolution qui a été initialement développé pour la segmentation d'images biomédicales à l'université de Fribourg en Allemagne [26, 39]. Cette architecture est réputée pour sa grande efficacité dans la détection précise des contours des objets dans les images [11, 34]. Cette précision est essentielle pour notre étude, car elle nous permet de délimiter de manière nette et précise les différentes classes et zones dans nos images géospatio-temporelles, ce qui est fondamental pour la surveillance et la détection des changements.

D'autre part, l'architecture entièrement convolutif (FCN) est principalement utilisée pour la segmentation sémantique [18, 24]. Elle offre la possibilité de comprendre la sémantique des classes présentes dans nos images, ce qui est particulièrement utile pour l'analyse des changements survenus au fil du temps. Cette capacité à conserver la sémantique des classes dans nos résultats est cruciale pour mettre en évidence et comprendre les changements qui se sont produits dans notre zone d'étude.

Une fois que le modèle génère les masques correspondants aux images temporelles, nous avons mis en place une post-classification pour la mise en évidence des zones qui ont subi un changement car les masques générés gardent la sémantique des classes ce qui rend cette approche intéressante.

3.3 Avantages et limites des différentes approches

Dans la revue de littérature, nous avons plusieurs approches de détection de changements d'occupations du sol à partir d'images satellites qui ont été présentées. Chacune de ces approches comporte ses avantages et ses limites. Voici une étude comparative des différentes approches, les méthodes traditionnelles et les méthodes de Deep Learning, en mettant en évidence leurs points forts et leurs faiblesses :

3.3.1 Méthodes Traditionnelles :

les méthodes traditionnelles en télédétection, y compris la comparaison "image-image" et l'analyse en composantes principales (ACP) offrent une simplicité d'application mais présentent des limitations importantes. Elles sont sensibles aux variations saisonnières et aux conditions atmosphériques, ne tiennent pas compte de la sémantique des classes, et se concentrent principalement sur les valeurs radiométriques des pixels, ce qui limite leur capacité à comprendre la nature des changements.

Les méthodes de comparaison post-classification se distinguent en permettant une meilleure compréhension des changements en se basant sur les classes d'occupation du sol plutôt que sur les valeurs radiométriques brutes. Cependant, elles nécessitent des classifications préalables de haute qualité pour des performances optimales.

3.3.2 Méthodes d'apprentissage profond :

3.3.2.1 Architecture Siamoise Multimodale :

Cette approche est capable de traiter des données hétérogènes provenant de capteurs différents. Elle est adaptée à la fusion de données de sources variées pour la détection de changement. Les performances en termes de précision et de rappel peuvent être modérées par rapport à d'autres méthodes. La détection de changement peut être moins précise.

3.3.2.2 Réseau Siamois-CNN :

Les performances sont élevées, avec des précisions impressionnantes. Cette méthode est efficace pour la détection de changement, même avec des données hétérogènes. La performance peut varier en fonction des sources d'imagerie et des résolutions spatiales. Des contraintes de données spécifiques peuvent se poser.

3.3.2.3 Modèle U-Net au Niveau des Caractéristiques :

Cette méthode offre une grande précision et est efficace pour identifier la nature des changements dans les images. Elle est capable de surpasser les méthodes existantes. Elle peut dépendre des données HRMS et de la qualité des images panchromatiques et multi-spectrales. Les performances peuvent varier dans différents scénarios.

3.3.2.4 Réseaux Neuronaux Convolutifs (CNN) :

Les CNN montrent des performances prometteuses avec une précision élevée. Ils peuvent généraliser efficacement au-delà des données d'entraînement initiales, ce qui est essentiel pour la télédétection. Le besoin de larges

ensembles de données étiquetées pour l'apprentissage profond peut être difficile à satisfaire. De plus, la classification peut être complexe en raison de la variation des apparences des caractéristiques terrestres.

3.3.2.5 Architecture bidirectionnelle de réseau de mémoire à long terme (BiLSTM) :

L'approche décrite dans la cinquième étude offre une cartographie précise de l'utilisation des sols en utilisant des données Sentinel-2, avec une grande variété de bandes spectrales et une extraction de caractéristiques poussée. Cela garantit des résultats fiables avec une précision globale élevée de 98,7%. Cependant, la méthode peut être sensible à la résolution spatiale variable des données et dépend fortement des données de terrain pour l'entraînement, ce qui peut limiter son applicabilité. De plus, la qualité des images panchromatiques et multispectrales peut influencer les performances de la méthode.

3.3.2.6 Le modèle FCN-8 avec VGG-16 :

Le modèle FCN-8 avec VGG-16 en tant qu'encodeur présente plusieurs avantages. Il offre des performances exceptionnelles, avec une précision moyenne de 91,0% et un IoU moyen de 0,84, surpassant ainsi un logiciel SIG commercial. De plus, le modèle effectue une classification dense, ce qui signifie qu'il attribue une classe à chaque pixel de l'image, offrant ainsi une granularité fine pour la segmentation des terres. En utilisant des poids pré-entraînés VGG-16, il exploite des caractéristiques profondes, ce qui peut améliorer sa capacité de discrimination. L'étude inclut également une analyse détaillée des erreurs aux frontières des classes d'utilisation des terres, ce qui permet de mieux comprendre ces erreurs. Cependant, le modèle peut être sensible à la qualité et à la résolution des images satellites utilisées, ce qui

peut affecter ses performances. Il présente également des erreurs aux limites des classes d'utilisation des terres et de couverture des terres, bien que ces erreurs aient été analysées.

3.3.3 Analyse comparative :

Les méthodes traditionnelles en télédétection, telles que la comparaison "image-image" et l'ACP, offrent une simplicité d'application, mais elles présentent des limitations importantes. Elles sont sensibles aux variations saisonnières et aux conditions atmosphériques, ne tiennent pas compte de la sémantique des classes et se concentrent principalement sur les valeurs radiométriques des pixels, limitant ainsi leur capacité à comprendre la nature des changements. Les méthodes de comparaison post-classification, bien qu'offrant une meilleure compréhension des changements en se basant sur les classes d'occupation du sol, nécessitent des classifications préalables de haute qualité pour des performances optimales.

En revanche, les méthodes d'apprentissage profond proposent des approches plus sophistiquées. L'architecture Siamoise Multimodale permet de traiter des données hétérogènes, mais ses performances en termes de précision et de rappel peuvent être modérées. Le réseau Siamois-CNN offre des performances élevées, avec des précisions impressionnantes, mais elles peuvent varier en fonction des sources d'imagerie et des résolutions spatiales. Les modèles U-Net au niveau des caractéristiques offrent une grande précision et sont efficaces pour identifier la nature des changements, mais dépendent des données HRMS et de la qualité des images panchromatiques et multi-spectrales, ce qui peut influencer les performances. Les réseaux neuronaux convolutifs (CNN) montrent des performances prometteuses avec une précision élevée, mais ils

CHAPITRE 3. ETAT DE L'ART

nécessitent de larges ensembles de données étiquetées pour l'apprentissage profond, et la classification peut être complexe en raison de la variation des apparences des caractéristiques terrestres. Enfin, l'architecture bidirectionnelle de réseau de mémoire à long terme (BiLSTM) offre une cartographie précise de l'utilisation des sols avec des données Sentinel-2, mais elle peut être sensible à la résolution spatiale variable des données et dépend fortement des données de terrain pour l'entraînement, tout en étant influencée par la qualité des images panchromatiques et multispectrales. le modèle FCN-8 avec VGG16 offre des performances exceptionnelles, avec une précision moyenne et un IoU moyen élevé, surpassant ainsi un logiciel SIG commercial. Cependant, il peut être sensible à la qualité des images satellites, présente des erreurs aux limites des classes d'utilisation des terres.

Chapitre 4

Proposition de solution

Résumé : Dans ce chapitre, dans un premier temps nous abordons la manière dont la collecte a été effectuée, le processus d'étiquetage et les choses à prendre en compte pour la validation. Dans un deuxième temps, nous faisons la description du prétraitement et de la méthode de génération des données. Enfin pour terminer, nous analysons l'architecture des modèles utilisés.

Sommaire

4.1	Description de la collecte, de l'étiquetage, et de la validation des données	52
4.2	Description du prétraitement et de la méthode de génération des données	55
4.3	Analyse de l'architecture des modèles de prédiction	58
4.4	Déploiement en temps réel :	65

La phase allant de la collecte des données à leur validation est une étape fondamentale dans le développement de modèles de Deep Learning. Elle englobe la réunion de données de qualité, leur prétraitement pour qu'elles

soient adaptées à l'apprentissage automatique, l'augmentation des données pour accroître la diversité de l'ensemble d'entraînement, la séparation des données en ensembles d'entraînement et de test, et enfin, la validation pour s'assurer de la qualité et de l'équilibre des données. Cette étape est cruciale pour garantir que le modèle sera performant et capable de généraliser à de nouvelles données.

4.1 Description de la collecte, de l'étiquetage, et de la validation des données

4.1.1 Description de la collecte

L'étude ayant été mené sur la zone de Richard Toll, la collecte des images satellitaires géospatio-temporelles a été effectuée sur cette zone à l'aide du logiciel « Google Earth Pro ». Le terme "images satellitaires géospatio-temporelles" fait référence à des images satellitaires qui contiennent des informations non seulement sur la position géographique (l'aspect géospatial) mais aussi sur l'horodatage (l'aspect temporel) car nous avons collecté les données dans différentes positions géographiques et à des moments différents dans le temps. Les images satellitaires collectées sont de résolutions (1024x768) et (3840x2160) et sont constituées de 3 canaux RVB (Rouge Vert Bleu). Pour les images de taille (3840x2160), elles sont partitionnées en taille de (800x800) pour être organisées en tuiles.

4.1.2 Description de l'étiquetage

Ensuite, nous avons effectué des étiquetages manuels grâce au logiciel « LabelMe » en délimitant avec des polygones les zones d'intérêt qui ont été

réparties en huit classes distinctes : Le sol, la végétation, l'eau, les bâtiments, les routes, les champs cultivés, les champs non cultivés et les terrains de football. Une fois que la délimitation des zones terminée, un fichier « json » contenant les coordonnées de chaque classe est généré. Puis à l'aide d'un script en python, on utilise ce fichier pour générer les masques correspondants aux images, qui seront considérés comme les labels plus tard.

4.1.3 Description de la validation des données

La validation des données est une étape essentielle dans notre projet de collecte de données, en particulier lorsqu'il s'agit de données utilisées pour des tâches d'apprentissage automatique, telles que la segmentation d'images. Elle garantit la qualité, la fiabilité et la pertinence de nos données collectées. Voici comment la validation des données s'applique à notre étude :

Qualité des données : La validation des données consiste à vérifier la qualité de nos images satellitaires collectées. Cela implique de s'assurer que nos images ne présentent pas de problèmes techniques tels que des distorsions, des artefacts, des problèmes d'étalonnage ou des anomalies. Des images de mauvaise qualité peuvent affecter négativement les performances de nos modèles d'apprentissage automatique.

Concordance avec l'étiquetage : Nous avons mentionné que nous avons effectué des étiquetages manuels en délimitant des zones d'intérêt à l'aide de polygones pour huit classes différentes. La validation des données consiste à vérifier si les zones étiquetées correspondent correctement aux caractéristiques de la zone de Richard Toll. Cela implique de s'assurer que les polygones ont été correctement tracés et que les étiquetages correspondent à la réalité.

Consistance temporelle : Comme nous travaillons avec des images satellitaires géospatio-temporelles, il est important de valider la cohérence

CHAPITRE 4. PROPOSITION DE SOLUTION

temporelle de nos données. Cela signifie de vérifier que les changements réels dans la zone sont correctement reflétés dans les images collectées à différents moments.

Évaluation des étiquettes : Nous avons généré des masques correspondants à nos images à l'aide d'un fichier JSON contenant les coordonnées de chaque classe. La validation des données inclut la vérification de la précision de ces masques. Les masques doivent correspondre correctement aux zones étiquetées et ne pas contenir de lacunes ni d'erreurs significatives.

Partition des données : Si nous avons partitionné nos images à haute résolution en tuiles de (800x800), la validation des données consiste à vérifier que cette partition est correcte et qu'aucune information importante n'a été perdue ou mal découpée.

Évaluation de la variabilité des données : Enfin, la validation des données doit également prendre en compte la variabilité de nos données. Les données peuvent varier en fonction de la saison, des conditions météorologiques, de la qualité des images, etc. Il est important de s'assurer que ces variations sont bien comprises et gérées dans notre étude.

La validation des données est cruciale pour garantir la qualité et la fiabilité de nos données, ce qui à son tour contribuera à des résultats plus précis et fiables dans notre étude de surveillance et de détection des changements d'occupations du sol à partir d'images satellites.

4.2 Description du prétraitement et de la méthode de génération des données

4.2.1 Description du prétraitement

Le prétraitement des données est une étape cruciale dans l'apprentissage profond car il permet de faciliter l'apprentissage aux modèles. Voici les différentes étapes qui ont été effectuées à la fin de prétraitement dans le cadre de notre étude :

Redimensionnement : Les images sont redimensionnées pour qu'elles aient la même taille. Cela garantit que toutes les images sont cohérentes en termes de dimensions.

Normalisation : Les images sont normalisées à l'aide de la fonction `scaler.fit_transform(__)`. Cela consiste à mettre à l'échelle les valeurs des pixels pour qu'elles se situent dans la plage [0, 1]. La normalisation est effectuée pour chaque "patch" de l'image, ce qui est important pour la formation de modèles d'apprentissage automatique.

Découpage en patches : Les images sont divisées en "patches" plus petits, généralement de taille 256x256 pixels. Cela permet de découper les images en de petites régions, ce qui est courant dans les tâches de segmentation d'images.

Construction des ensembles de données : Les "patches" d'images sont ajoutés à `image_dataset`, et les "patches" de masques correspondants sont ajoutés à `mask_dataset` qui représentent des arrays. Cela prépare les données pour être utilisées dans des modèles d'apprentissage automatique, tels que les réseaux de neurones convolutifs pour la segmentation d'images.

Conversion des étiquettes en 2D : La fonction `rgb_to_2D_label(__)`,

défini dans le code, prend une étiquette au format RVB (Rouge-Vert-Bleu) et la convertit en une étiquette 2D où chaque pixel est représenté par un entier indiquant la classe à laquelle il appartient (par exemple, sol, végétation, eau, etc.).

Conversion des masques en étiquettes : Cette conversion est effectuée pour chaque masque dans le mask_dataset. Les masques RVB sont convertis en étiquettes 2D en utilisant la fonction `rgb_to_2D_label()`, et les étiquettes sont ajoutées à la liste labels.

Conversion en catégories : Les étiquettes 2D sont converties en catégories à l'aide de `to_categorical`. Cela permet de transformer les étiquettes en un format adapté à l'entraînement de modèles de classification, où chaque catégorie est représentée par un vecteur binaire.

Division des données en ensembles d'entraînement et de test : Les données d'images (X) et les étiquettes (y) sont divisées en ensembles d'entraînement (X_{train} , y_{train}) et de test (X_{test} , y_{test}) en utilisant `train_test_split`. Cela permet d'obtenir un ensemble d'entraînement pour entraîner le modèle et un ensemble de test pour évaluer ses performances.

Sauvegarde des données prétraitées : Les données prétraitées sont sauvegardées dans un fichier .npz en utilisant `np.savez` de numpy. Cela permet de stocker les données de manière organisée et de les charger ultérieurement pour l'entraînement de modèles sans avoir à refaire le prétraitement.

4.2.2 Description de la méthode de génération des données

Pour l'augmentation de nos données, nous avons choisi d'utiliser `ImageDataGenerator` qui est essentielle dans le contexte de l'augmentation des données pour améliorer les performances des modèles d'apprentissage automatique, en particulier lorsqu'il y a une quantité limitée de données d'entraînement.

CHAPITRE 4. PROPOSITION DE SOLUTION

Cette classe offre un moyen efficace de diversifier le jeu de données existant en créant de nouvelles variantes des images originales [19]. Voici un aperçu plus détaillé des transformations effectuées et basées sur un procédé empirique :

Rotation : La transformation de rotation permet de varier l'orientation des images. En autorisant une rotation aléatoire jusqu'à 15 degrés dans le sens horaire ou antihoraire, les modèles peuvent mieux généraliser et reconnaître des objets sous différents angles.

Translation : Les translations horizontales et verticales ajoutent de la diversité spatiale aux données. En déplaçant aléatoirement les images vers la gauche, la droite, le haut ou le bas (jusqu'à 10% de leur largeur ou hauteur), les modèles apprennent à traiter des objets dans différentes positions.

Cisaillement : La distorsion en cisaillement introduit des modifications mineures dans la forme des objets. Cela peut être utile pour améliorer la robustesse de la détection d'objets face à des déformations.

Zoom : La transformation de zoom permet de rapprocher ou d'éloigner aléatoirement les objets dans les images. Cela aide le modèle à s'adapter à des échelles différentes.

Retournements : Les retournements horizontaux et verticaux créent des versions miroir des images. Cela aide à l'entraînement des modèles pour reconnaître des objets indépendamment de leur orientation.

Remplissage : Lorsque des pixels sont créés à la suite de certaines transformations, le mode "Reflect" remplit ces pixels en utilisant la réflexion des pixels voisins. Cette méthode de remplissage est généralement appropriée pour les images naturelles, car elle préserve la cohérence des textures et des motifs.

4.3 Analyse de l'architecture des modèles de prédiction

Dans le cadre de ce mémoire, deux modèles ont été utilisés, et nous avons procédé par le transfert d'apprentissage (transfer learning) en utilisant des modèles de classification déjà pré-entraîné sur ImageNet. Il s'agit de U-Net avec comme encodeur Inception-ResNetV2, et FCN-8 (Fully Convolutional Network 8) avec comme encodeur VGG19.

4.3.1 L'archctecture U-Net et inception-resnetv2 :

L'archctecture U-Net se caractérise par les éléments suivants 4.1 :

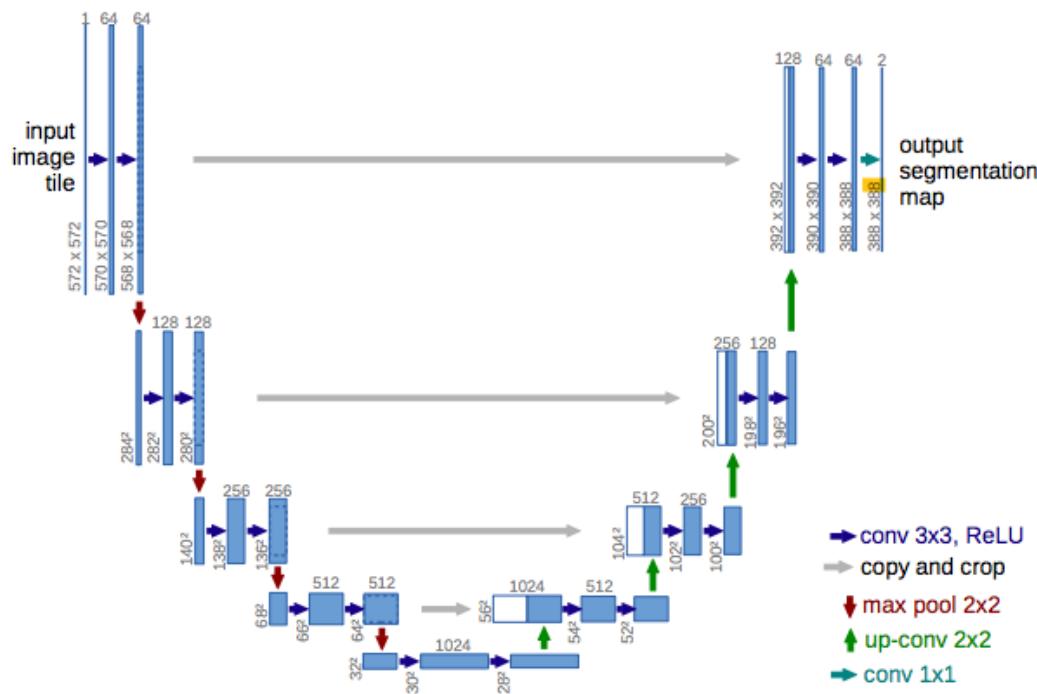


Figure 4.1 – L'archctecture U-Net [9]

D'un point de vue visuel, son architecture est en forme de "U" et se divise en trois sections : la contraction, le goulot d'étranglement et l'expansion. La

CHAPITRE 4. PROPOSITION DE SOLUTION

première section, appelée encodeur, a pour objectif de saisir le contexte d'une image. Elle est composée d'un empilement de couches de convolution et de max-pooling, permettant de capturer les caractéristiques de l'image tout en réduisant sa dimension pour économiser des paramètres. Cette phase comprend l'application répétée de deux couches de convolution 3×3 , chacune étant suivie d'une fonction d'activation ReLU et d'une normalisation par lots (batch normalization). Ensuite, une opération de max-pooling 2×2 est utilisée pour réduire la dimension spatiale. Le goulot d'étranglement, également connu sous le nom de "pont", relie l'encodeur au décodeur et assure la continuité du flux d'informations. Il est constitué de deux couches de convolution 3×3 , chacune suivie d'une activation ReLU. Le deuxième bloc est le décodeur, qui permet une localisation précise grâce à des convolutions transposées et rétablit la taille d'origine de l'image. Le décodeur commence par un sur-échantillonnage de la carte des caractéristiques, suivi d'une couche de convolution 2×2 transposée. Ensuite, deux couches de convolution 3×3 sont appliquées, chaque convolution étant suivie d'une activation ReLU. La sortie du dernier décodeur passe par une couche de convolution 1×1 avec une fonction d'activation sigmoïde. U-Net utilise une fonction de perte pour chaque pixel de l'image. La fonction Softmax est appliquée à chaque pixel, suivi d'une fonction de perte. Cela transforme la tâche de segmentation en un problème de classification, où chaque pixel doit être classé dans l'une des classes [9].

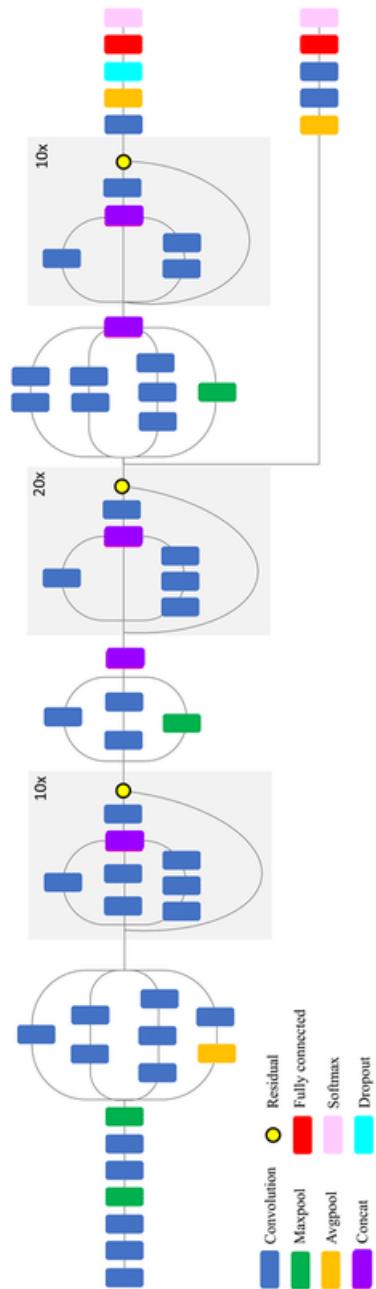


Figure 4.2 – L'architecture inception-resnetv2 [5]

Inception-ResNetv2 est une architecture de réseau de neurones qui combine

les avantages des blocs Inception et des connexions résiduelles de ResNet. Elle fonctionne en extrayant des caractéristiques riches à partir d'une image grâce à des modules Inception, tout en utilisant des connexions résiduelles pour améliorer la stabilité de l'apprentissage profond. Cette combinaison permet au modèle de capturer des informations diverses et hiérarchiques pour des tâches de vision par ordinateur, telles que la classification d'images ou la détection d'objets [5].

Lorsqu'Inception-ResNetv2 est utilisé comme encodeur dans le modèle U-Net, il joue un rôle essentiel dans l'extraction de caractéristiques à partir de l'image d'entrée pour la tâche de segmentation. Cette architecture permet de capturer des caractéristiques à différentes échelles spatiales grâce à ses modules Inception-ResNet, tout en maintenant l'efficacité de la formation grâce à des connexions résiduelles. Les caractéristiques extraites sont ensuite transmises au décodeur U-Net, où elles guident la reconstruction de l'image segmentée, contribuant ainsi à une segmentation précise [31].

Étant donné un ensemble de données déséquilibré Figure 4.3, nous avons appliqué la fonction de perte Dice en lui attribuant des poids équivalents, que nous avons ensuite additionné avec la fonction de perte Focal [15, 23]. Cette approche vise à contraindre le modèle à considérer toutes les classes comme importantes. Ensuite, nous avons choisi l'optimiseur Adam, qui est une méthode de taux d'apprentissage adaptatif, ce qui signifie qu'elle calcule les taux d'apprentissage individuels pour différents paramètres. Son nom est dérivé de l'estimation adaptative du moment (Adaptive moment estimation), et la raison pour laquelle il est appelé ainsi est parce qu'Adam utilise des estimations des premier et deuxième moments de gradient pour adapter le taux d'apprentissage pour chaque poids du réseau neuronal [6, 10].

La formule du Dice Loss est donnée par [15] :

CHAPITRE 4. PROPOSITION DE SOLUTION

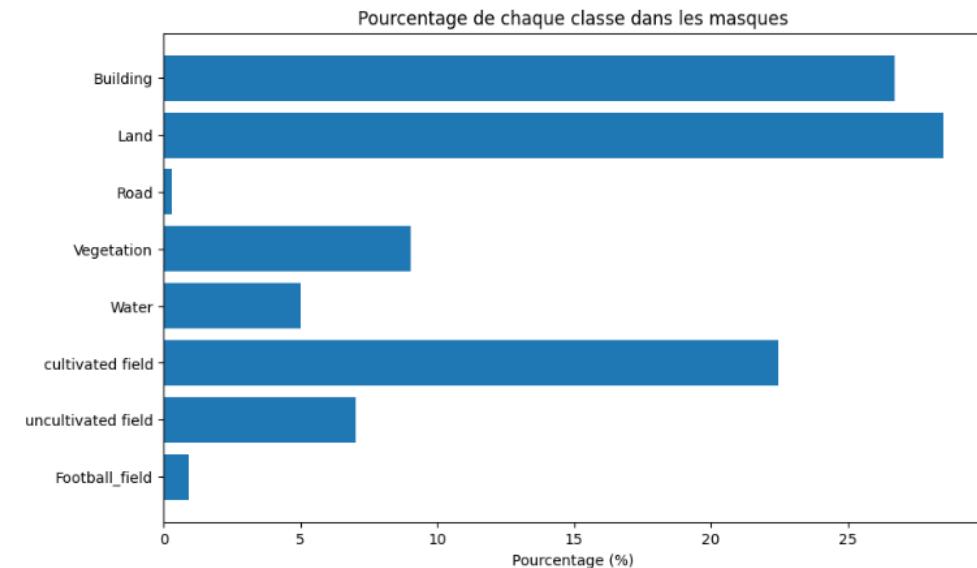


Figure 4.3 – La répartition des classes dans les données.

$$DSC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \quad (4.1)$$

Où :

TP représente les vrais positifs.

FP représente les faux positifs.

FN représente les faux négatifs.

La formule du Focal Loss est donnée par [15] :

$$\text{Focal Loss} = -(1 - \hat{y})^\gamma \cdot \log(\hat{y}) \quad (4.2)$$

Où :

\hat{y} est la probabilité prédictée par le modèle pour la classe cible.

γ est un paramètre de modulation qui peut être ajusté.

4.3.2 L'archctecture FCN-8 (Fully Convolutional Network 8) et VGG19 :

FCN, ou Fully Convolutional Network (Réseau Entièrement Convolutif en français), est une architecture de réseau de neurones convolutifs (CNN) utilisée pour la segmentation sémantique d'images. FCN a été développé pour traiter des images de taille variable et produire une carte de segmentation de la même taille que l'image d'entrée [29, 33], comme c'est décrit dans la figure 4.4 :

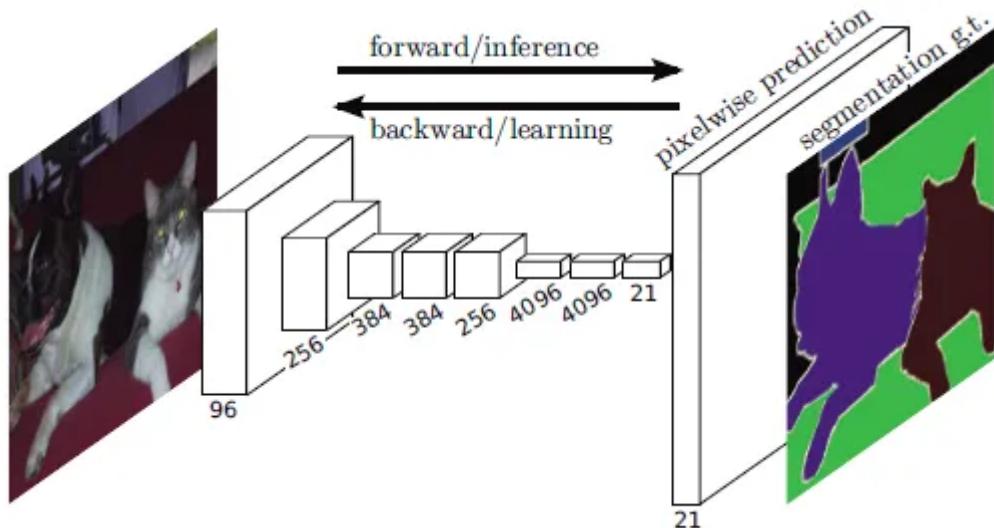


Figure 4.4 – L'archctecture FCN [29]

FCN-8 est basée sur l'idée de prendre un réseau pré-entraîné (comme VGG16 ou resNet) et de le transformer en un réseau entièrement convolutif. Pour ce faire, FCN-8 utilise des couches de déconvolution (transposées) pour augmenter la résolution de la sortie par rapport au modèle original. Cela permet au réseau de générer des cartes de segmentation de la même taille que l'image d'entrée, ce qui est essentiel pour la segmentation pixel par pixel.

CHAPITRE 4. PROPOSITION DE SOLUTION

FCN-8 effectue également une fusion des informations à plusieurs échelles en intégrant des cartes de caractéristiques de couches intermédiaires. Cela permet au réseau de prendre en compte des informations de différentes échelles spatiales pour une meilleure segmentation sémantique. En fin de compte, FCN-8 génère des cartes de segmentation qui attribuent une classe à chaque pixel de l'image, ce qui en fait un outil puissant pour la détection d'objets et d'autres tâches de vision par ordinateur[29].

[30]Dans notre cas le réseau pré-entraîné choisi est VGG19, c'est un réseau de neurones profond qui se distingue par sa profondeur, comme décrit dans la figure 4.5. Elle repose sur des couches de convolution 3x3 et des couches de pooling pour extraire des caractéristiques complexes des images. Les couches convolutionnelles détectent des détails simples et progressent vers des caractéristiques plus complexes. VGG19 suit une structure uniforme avec des couches convolutives et de pooling en cascade, aboutissant à des couches entièrement connectées pour la classification. C'est un modèle puissant pour la classification d'images et la caractérisation d'objets grâce à sa capacité à capturer des informations hiérarchiques à différentes échelles.

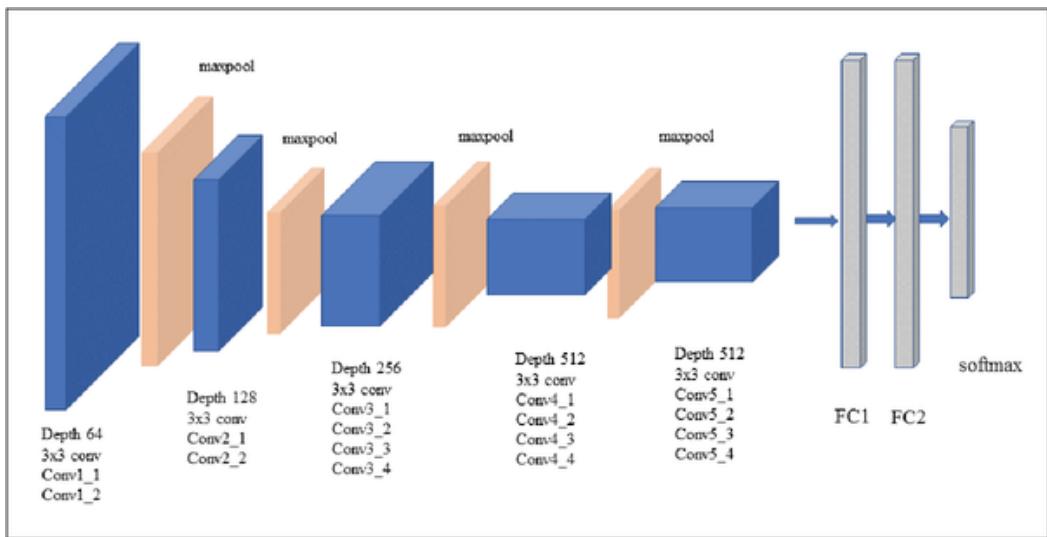


Fig. 3. VGG19 network architecture

Figure 4.5 – L'architecture VGG19 [30]

Étant donné que c'est le même ensemble de données déséquilibré, nous avons appliqué les mêmes hyperparamètres comme dans le réseau U-Net.

Une fois que ceci fait, nous passons au fine-tuning qui consiste à l'ajustement ou à la modification de paramètres de nos modèles déjà entraînés pour les adapter à notre tâche [40].

4.4 Déploiement en temps réel :

Dans cette partie, nous aborderons la détection de ces changements en temps réel à l'aide de techniques de traitement d'image. Cette capacité a des implications importantes dans la gestion des ressources naturelles, la télédétection et la surveillance environnementale, offrant ainsi des opportunités pour une planification plus durable et une meilleure compréhension de notre environnement. Nous allons explorer les grandes lignes de la détection de changements d'occupations du sol en temps réel, en mettant l'accent sur son

CHAPITRE 4. PROPOSITION DE SOLUTION

importance et ses applications en suivant les étapes suivantes :

Import des bibliothèques et définition des paramètres :

Les bibliothèques nécessaires sont importées, notamment OpenCV (cv2).

Les noms de fichiers d'entrée et de sortie sont définis, ainsi que le codec vidéo à utiliser.

Vérification des paramètres vidéo :

Les paramètres vidéo des deux flux d'entrée sont vérifiés pour s'assurer qu'ils sont identiques. Initialisation de la sortie vidéo :

Un objet VideoWriter est initialisé pour enregistrer la vidéo de sortie avec les paramètres spécifiés.

Boucle principale de traitement vidéo :

Une boucle while est utilisée pour lire en continu les frames des deux vidéos simultanément. Traitement des frames :

Pour chaque paire de frames lues : Les frames sont divisées en patches (régions plus petites) pour l'analyse. Les patches sont soumis à un modèle de Deep Learning pour prédire les changements d'occupations du sol. Les prédictions sont combinées pour former une image représentant les changements entre les deux frames. Cette image est enregistrée dans la vidéo de sortie.

Affichage et interruption de la boucle :

Les résultats sont affichés en temps réel. La boucle se termine si la touche 'q' est pressée comme définie dans le code.

Libération des ressources :

Une fois la boucle terminée, les ressources vidéo sont libérées, et les fenêtres OpenCV sont fermées.

Chapitre 5

Analyse et Interprétation des résultats

Résumé : Dans ce chapitre, nous faisons une présentation de nos résultats en premier lieu. Puis dans un deuxième temps nous passons à la discussion de ces résultats en analysant et interprétant ces derniers.

Sommaire

5.1 Résultats de la prédiction des modèles sur les images satellitaires géospatio-temporelles pour la surveillance et la détection des changements d'occupations du sol	68
5.1.1 Résultats expérimentaux avec U-Net et FCN-8 : . .	69
5.1.2 Prédiction avec U-Net et FCN-8 pour la surveillance :	72
5.1.3 Détection des changements d'occupations du sol : .	75
5.1.4 Généralisation sur d'autres zones du Sénégal : . . .	81
5.1.5 Déploiement en temps réel :	83
5.2 Discussion des résultats : Analyse et Interprétation des résultats	84

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

5.1 Résultats de la prédiction des modèles sur les images satellitaires géospatio-temporelles pour la surveillance et la détection des changements d'occupations du sol.

La phase de présentation des résultats constitue une étape cruciale de notre étude dédiée à la surveillance et à la détection des changements d'occupations du sol à partir d'images satellitaires géospatio-temporelles. Dans cette section, nous dévoilerons les fruits de notre travail acharné et des analyses méticuleuses que nous avons menées pour évaluer les performances de nos modèles prédictifs.

L'utilisation d'images satellitaires géospatio-temporelles représente un atout essentiel dans la surveillance de l'environnement et la compréhension des transformations qui s'y opèrent au fil du temps. Ces données permettent de surveiller les changements dans des domaines variés, de l'urbanisme à l'agriculture en passant par la gestion des ressources naturelles. Les modèles que nous avons développés visent à accroître l'efficacité de cette surveillance en automatisant la détection des changements d'occupations du sol, une tâche complexe et exigeante.

Dans cette section, nous examinerons en détail les résultats de nos modèles, en mettant en lumière leur performance, leur précision et leur capacité à détecter les changements d'occupations du sol de manière fiable et efficiente. Nous présenterons également des analyses approfondies pour interpréter ces résultats, en mettant en évidence les implications pratiques de notre travail.

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

5.1.1 Résultats expérimentaux avec U-Net et FCN-8 :

Après avoir effectué les étapes cruciales du prétraitement détaillées dans le chapitre 4, nous sommes passés à l'entraînement, qui consiste à utiliser l'ensemble d'entraînement pour ajuster les paramètres des modèles. Les modèles apprennent à partir des données d'entraînement en minimisant la fonction de perte (la somme du Dice loss et du Focal loss) qui mesure l'écart entre les prédictions des modèles et les images réelles. Cela implique généralement la rétropropagation du gradient (backpropagation) pour mettre à jour les poids des modèles. Puis à l'évaluation, qui consiste à évaluer l'ensemble de test pour mesurer la performance des modèles. Les mesures de performance dépendent du type de problème, mais pour notre cas nous avons adopté les métriques telles que :

1. Précision (Precision) : - La précision mesure le nombre de prédictions positives correctes (VP - vrais positifs) parmi toutes les prédictions positives faites par le modèle.

- Formule : $\text{Precision} = \frac{TP}{TP+FP}$
- La précision indique la capacité du modèle à ne pas attribuer à tort la classe positive à des exemples négatifs [20].

2. Rappel (Recall) : - Le rappel mesure le nombre de prédictions positives correctes (VP - vrais positifs) parmi toutes les instances réellement positives.

- Formule : $\text{Recall} = \frac{TP}{TP+FN}$
- Le rappel indique la capacité du modèle à identifier correctement tous les exemples positifs [20].

3. F-mesure (F1-Score) ou Dice loss : - La F-mesure est une métrique qui

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

combine à la fois la précision et le rappel en une seule mesure en utilisant la moyenne harmonique.

- Formule : $F1\text{-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$
- La F-mesure est utile lorsque vous voulez trouver un équilibre entre la précision et le rappel [15, 20].

4. Exactitude (Accuracy) : - L'exactitude mesure la proportion totale de prédictions correctes parmi toutes les prédictions.

- Formule : $\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$
- L'exactitude est une métrique globale qui évalue la capacité du modèle à classer correctement les exemples positifs et négatifs [8].

5. Précision Globale (Accuracy Overall [AO]) : - Il n'y a pas de définition standard de la "précision globale". Cependant, il peut s'agir de l'exactitude globale, qui mesure la proportion totale de prédictions correctes, comme décrit ci-dessus.

6. Coefficient de Kappa (Cohen's Kappa) : - Le coefficient de Kappa mesure la concordance entre les prédictions du modèle et les classes réelles en tenant compte du hasard. - Formule : $\kappa = \frac{P_o - P_e}{1 - P_e}$
- P_o est la proportion d'accord observée, et P_e est la proportion d'accord attendue par hasard [16].

7. Coefficient de Jaccard (Indice de Jaccard) : - Le coefficient de Jaccard mesure la similarité entre deux ensembles en utilisant l'intersection et l'union des ensembles.

- Formule : $\text{Jaccard} = \frac{TP}{TP + FP + FN}$

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

- Il est couramment utilisé pour évaluer la similarité entre ensembles, comme dans l'index de Jaccard pour la segmentation d'images [15].

Dans notre cas, ayant effectué l'entraînement des deux modèles qu'on a choisi, les métriques qu'on a obtenu, sont présentées dans les tableaux 5.1 et 5.2.

Métriques	Accuracy	Jaccard	Kappa	AO
UNET	96%	92.55%	94.98%	96.10%
FCN8	91%	84%	88.55%	91.15%

Tableau 5.1 – Les métriques globales

Métriques	Precision	Recall	F1-score
Modèles	U-Net FCN-8	U-Net FCN-8	U-Net FCN-8
Land	0.95 0.85	0.93 0.88	0.94 0.87
Vegetation	0.92 0.89	0.96 0.87	0.94 0.88
Building	0.95 0.91	0.97 0.91	0.96 0.91
Road	0.93 0.73	0.97 0.78	0.95 0.75
Water	0.97 0.92	0.95 0.86	0.96 0.89
Cultivated field	0.98 0.97	0.97 0.94	0.98 0.95
Uncultivated field	0.97 0.94	0.99 0.94	0.98 0.94
Football field	0.96 0.96	0.98 0.97	0.97 0.96

Tableau 5.2 – Rapport de classification des modèles

Dans le tableau 5.3, nous avons la matrice de confusion des deux modèles :

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

modèles	TP	TN	FP	FN
UNET	96.10%	99.44%	3.89%	0.55%
FCN8	91.15%	98.73%	8.84%	1.26%

Tableau 5.3 – Matrices de confusion globales des différents modèles.

TP(Vrai Positif) : Le nombre d'échantillons positifs correctement prédis comme positifs.

FP(Faux Positif) : Le nombre d'échantillons négatifs incorrectement prédis comme positifs.

TN(Vrai Ngatif) : Le nombre d'échantillons négatifs correctement prédis comme négatifs.

FN(Faux Ngatif) : Le nombre d'échantillons positifs incorrectement prédis comme négatifs.

5.1.2 Prédiction avec U-Net et FCN-8 pour la surveillance :

Notre objectif est de surveiller l'évolution des champs agricoles dans la région de Richard Toll sur une période allant de mai à octobre 2019. Cette surveillance vise à déterminer quels sont les champs qui sont actuellement cultivés (représentés en vert) et ceux qui ne le sont pas (représentés en rouge). Pour ce faire, nous avons utilisé deux modèles de segmentation d'image, à savoir U-Net (5.1) et FCN-8 (5.2).

L'utilisation de ces modèles nous permet d'analyser des images de la région capturées à différents moments au cours de cette période. Grâce à U-Net et FCN-8, nous sommes en mesure de distinguer avec précision les zones qui ont été consacrées à la culture et celles qui sont restées non cultivées. Ces modèles sont spécialement conçus pour segmenter les images, ce qui signifie qu'ils peuvent identifier et isoler les champs agricoles des autres éléments de

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

l'environnement.

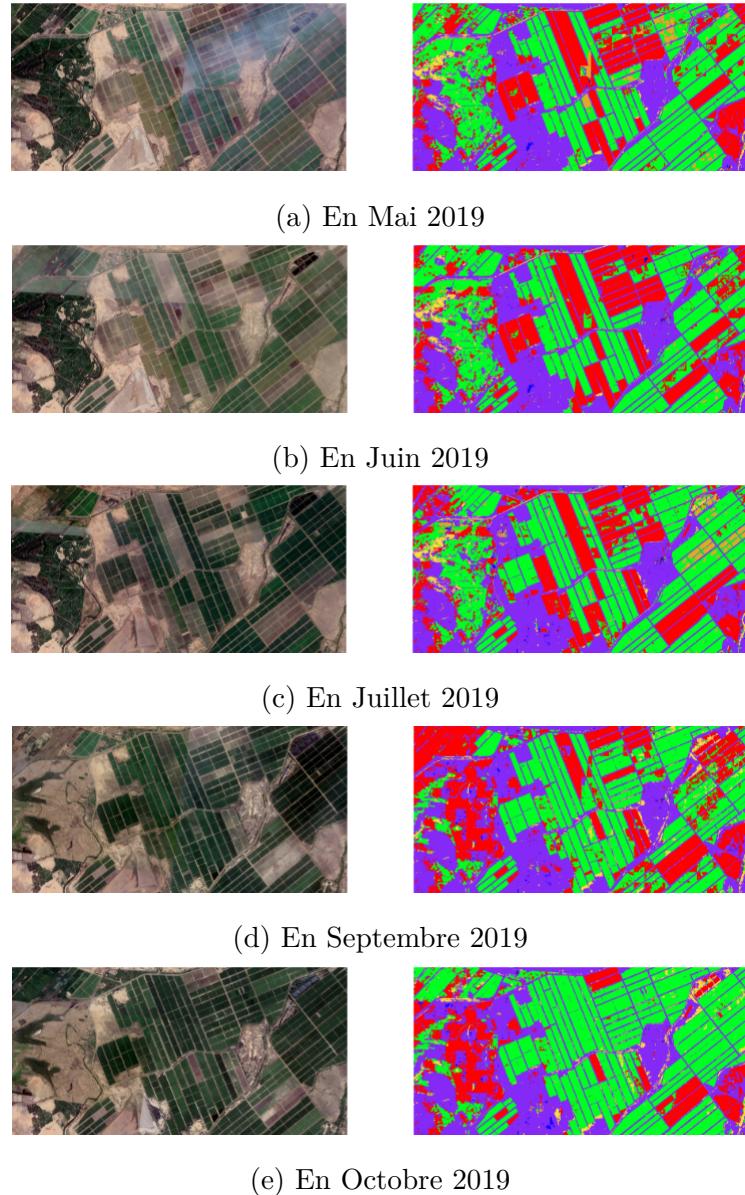


Figure 5.1 – La prédiction du modèle U-Net sur les images satellites avec une résolution de 8192x4320, et avec une altitude de 10 kilomètres de hauteur

La figure 5.1 illustre les résultats de l'application du modèle U-Net, tandis que la figure 5.2 présente les résultats obtenus avec le modèle FCN-8. Ces

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

images fournissent une visualisation des zones de champs agricoles cultivées en vert et des zones non cultivées en rouge.

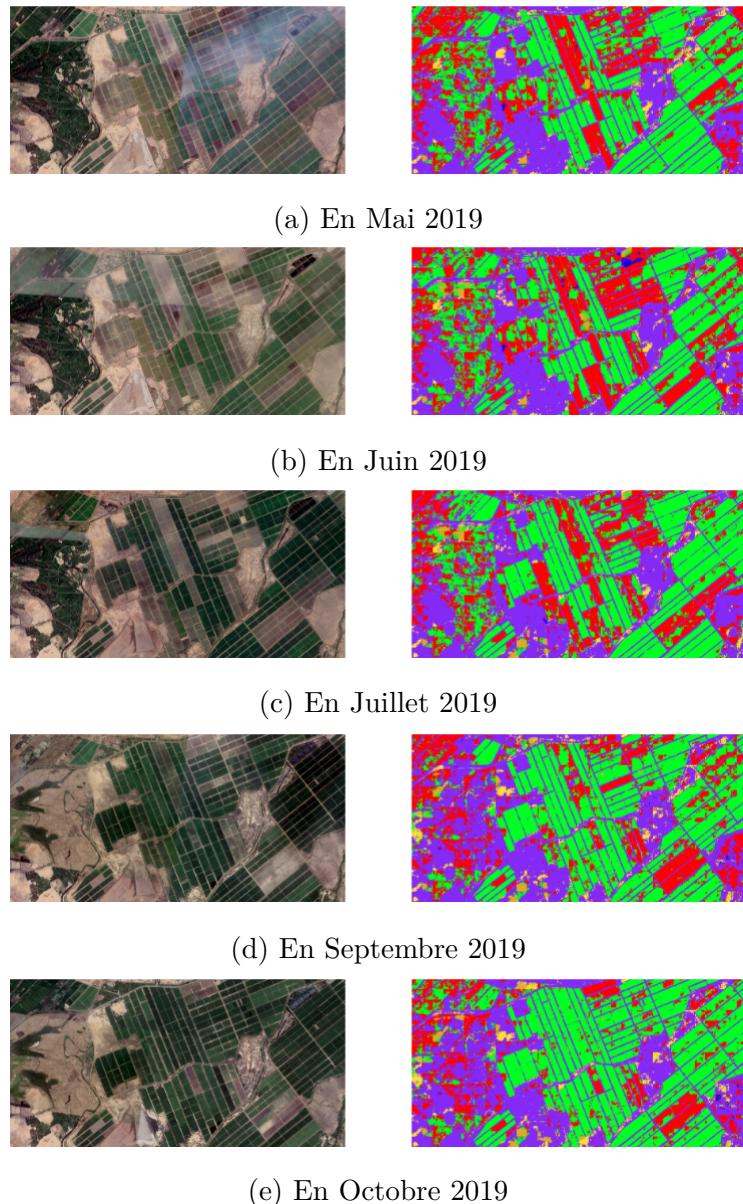


Figure 5.2 – La prédiction du modèle FCN-8 sur les mêmes images satellites avec la même résolution de 8192x4320, et avec une altitude de 10 kilomètres de hauteur

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

Cette analyse nous permet de suivre l'évolution des activités agricoles dans la région, offrant des informations importantes pour la planification et la gestion des ressources agricoles. Les modèles U-Net et FCN-8 sont des outils efficaces pour cette tâche, facilitant ainsi une évaluation détaillée de la situation des champs au fil du temps.

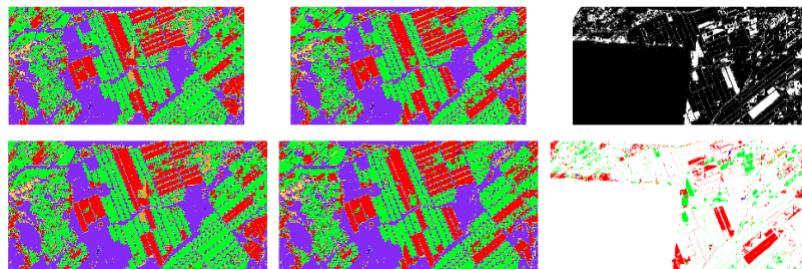
5.1.3 Détection des changements d'occupations du sol :

Pour mettre en évidence les zones de changement, nous effectuons une comparaison pixel par pixel en utilisant les masques générés par notre modèle U-Net, qui a montré sa précision à travers des métriques fiables. Ces masques sont essentiels pour conserver à la fois le contexte et la sémantique des objets dans la scène, garantissant ainsi une détection précise et claire des modifications.

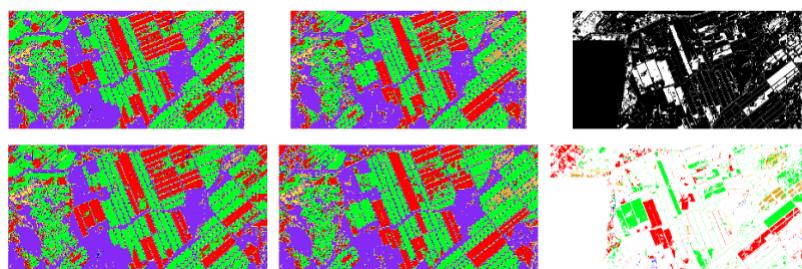
Nos modèles de segmentation avancés nous permettent de générer des masques agissant comme des cartes d'étiquetage pour chaque pixel de l'image. Ces masques facilitent la distinction entre les objets et les zones d'arrière-plan, ce qui simplifie la comparaison détaillée entre les images.

Lorsque nous comparons deux images, à savoir l'image de référence et celle après la survenue du changement, nous utilisons les masques générés par notre modèle U-Net. Ces masques nous aident à distinguer comment chaque pixel a évolué ou est resté inchangé, sans nécessiter la superposition des images. Nous avons les cartes binaire où la couleur noire représente pas de changements et la blanche représente les changements et aussi une carte multiple qui est plus précise où chaque changement est identifié puis mis à sa classe, les zones présentant des changements significatifs sont clairement identifiées, car les pixels correspondants dans les masques sont étiquetés différemment voir 5.3d.

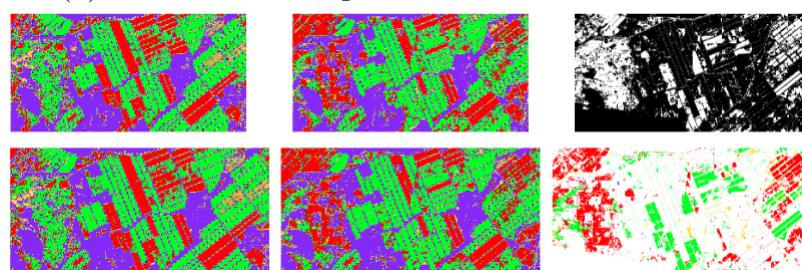
CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS



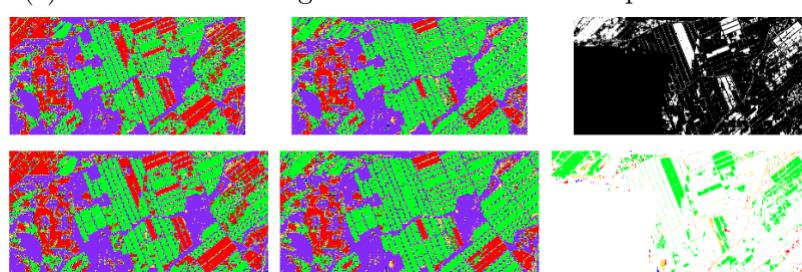
(a) Détection de changement entre Mai et Juin 2019



(b) Détection de changement entre Juin et Juillet 2019



(c) Détection de changement entre Juillet et Septembre 2019



(d) Détection de changement entre Septembre et Octobre 2019

Figure 5.3 – La détection de changement avec U-Net sur différentes périodes, avec une résolution de 8192x4320, et avec une altitude de 10 kilomètres de hauteur

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

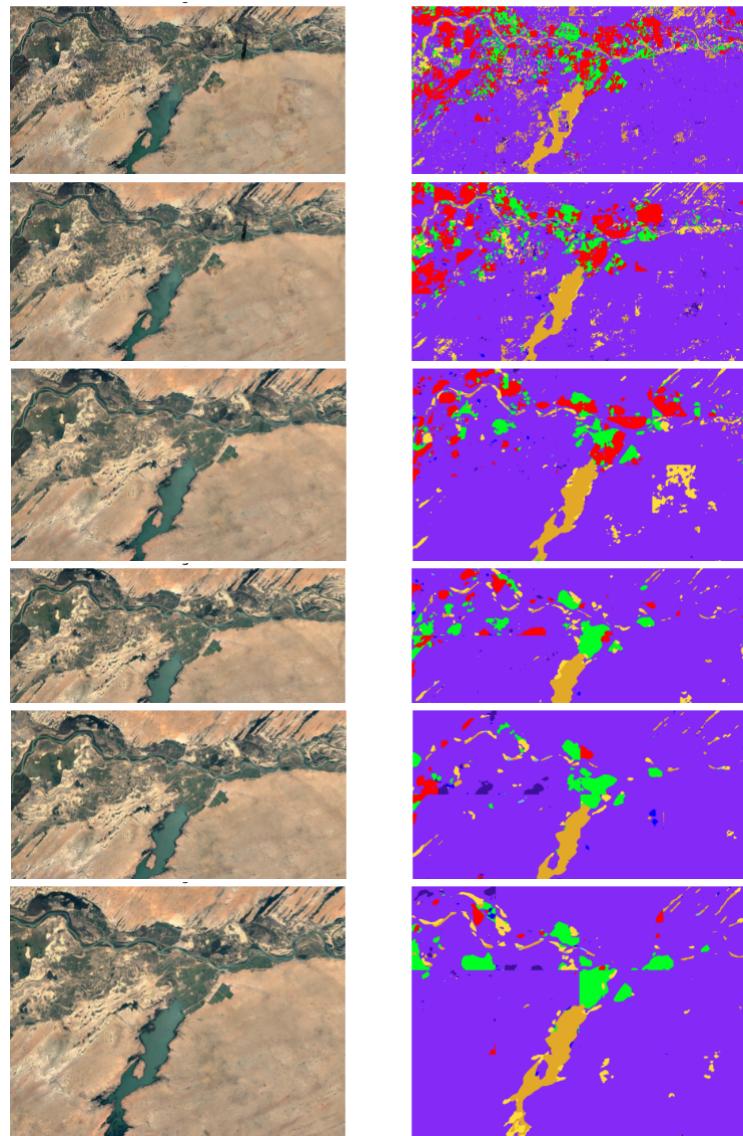


Figure 5.4 – La prédiction du modèle U-Net respectivement sur des images satellites avec différentes résolutions : 8192x4320, 3840x2160, 1920x1080, 1280x720, 1050x632 et 1024x768, et avec une altitude de 100 kilomètres de hauteur

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

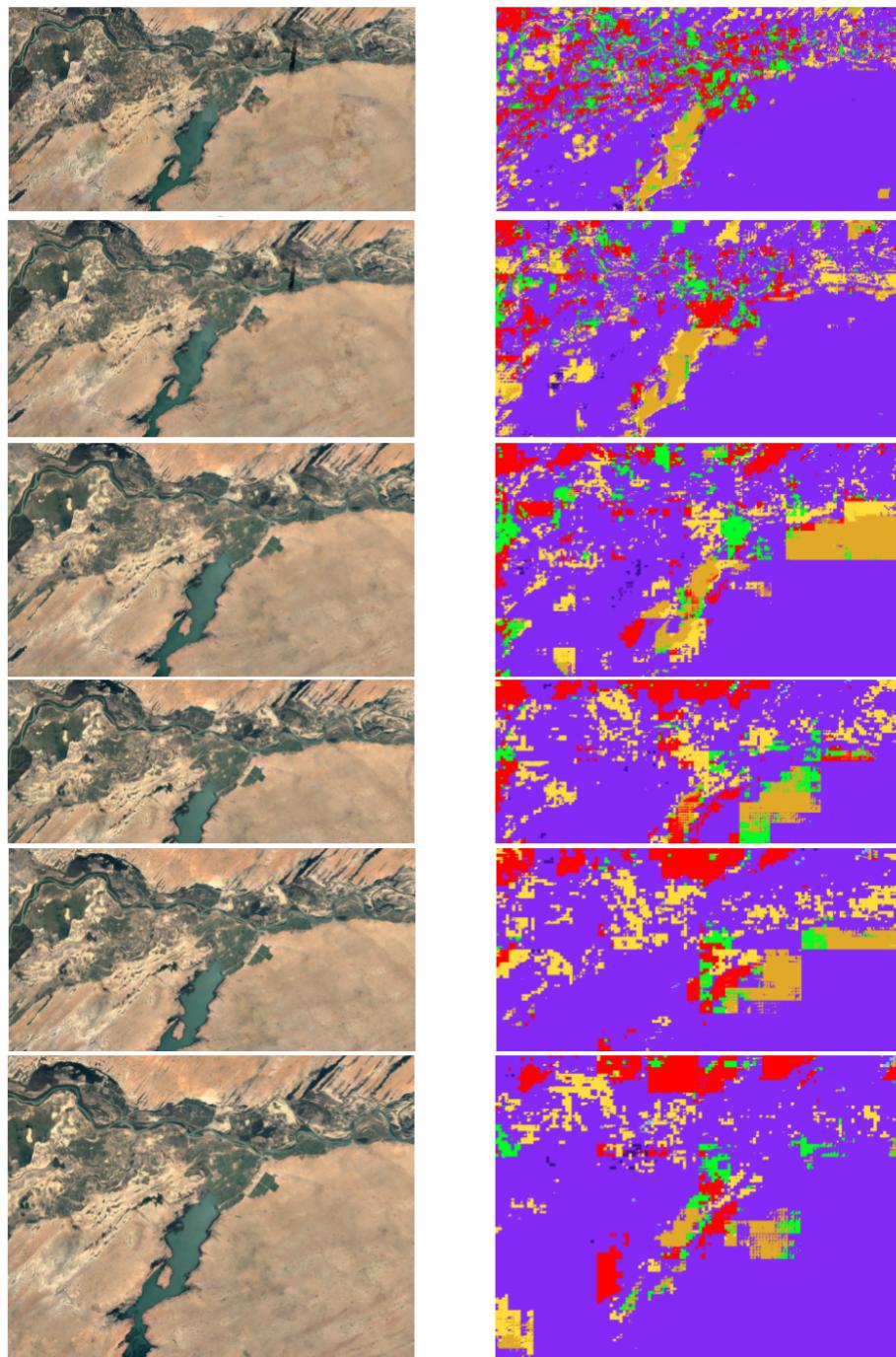


Figure 5.5 – La prédiction du modèle FCN-8 respectivement sur des images satellites avec différentes résolutions : 8192x4320, 3840x2160, 1920x1080, 1280x720, 1050x632 et 1024x768, et avec une altitude de 100 kilomètres de hauteur

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

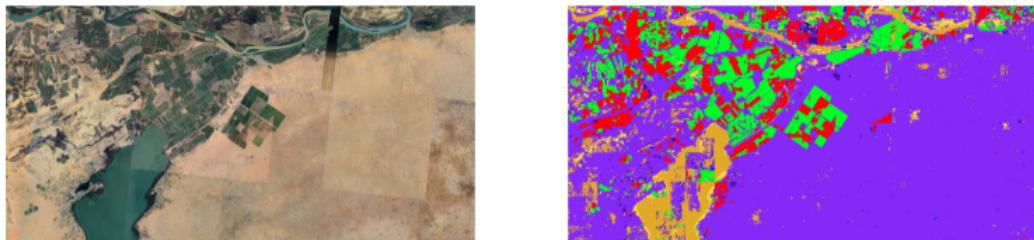


Figure 5.6 – La prédiction du modèle U-Net respectivement sur une image satellite avec une résolution de 8192x4320, et avec une altitude de 50 kilomètres de hauteur

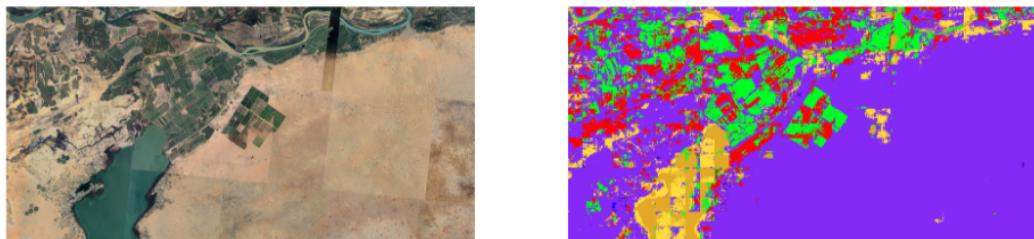


Figure 5.7 – La prédiction du modèle FCN-8 respectivement sur une image satellite avec une résolution de 8192x4320 et avec une altitude de 50 kilomètres de hauteur

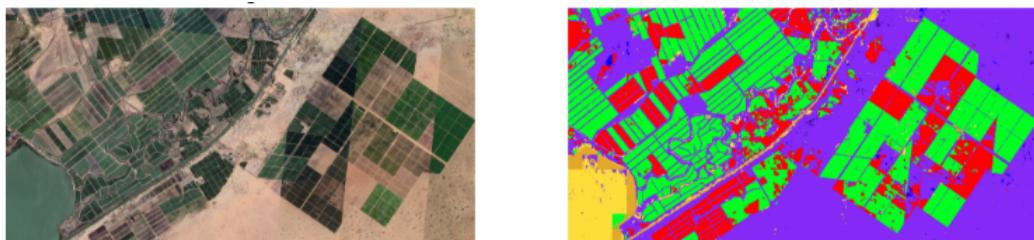


Figure 5.8 – La prédiction du modèle U-Net respectivement sur une image satellite avec une résolution de 8192x4320 et avec une altitude de 15 kilomètres de hauteur

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

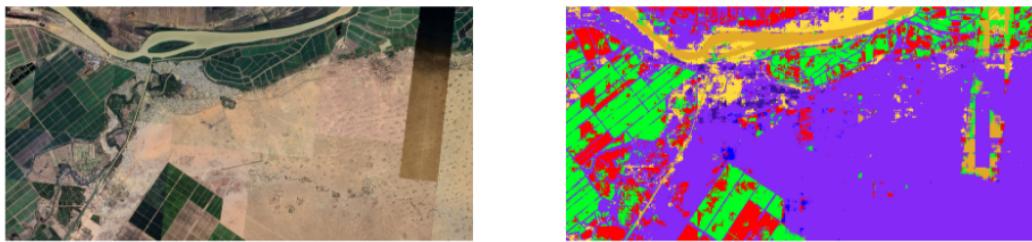


Figure 5.9 – La prédiction du modèle FCN-8 respectivement sur une image satellite avec une résolution de 8192x4320 et avec une altitude de 15 kilomètres de hauteur

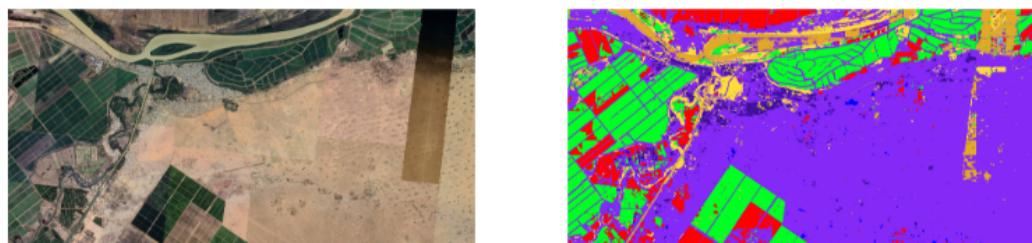


Figure 5.10 – La prédiction du modèle U-Net respectivement sur une image satellite avec une résolution de 8192x4320 et avec une altitude de 15 kilomètres de hauteur

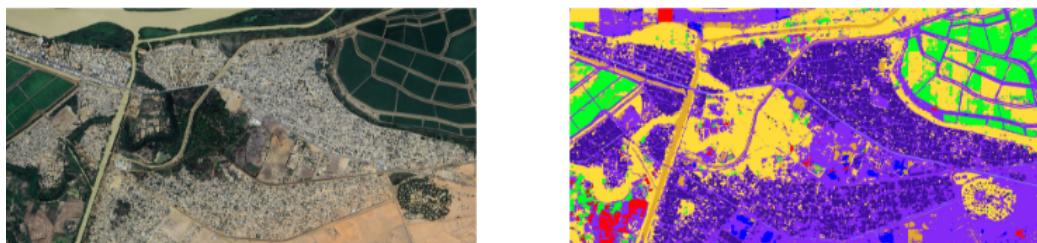


Figure 5.11 – La prédiction du modèle U-Net respectivement sur une image satellite avec une résolution de 8192x4320 et avec une altitude de 4 kilomètres de hauteur

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

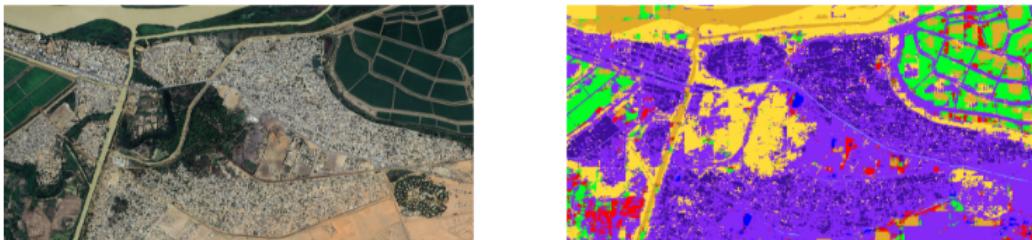


Figure 5.12 – La prédiction du modèle FCN-8 respectivement sur une image satellite avec une résolution de 8192x4320 et avec une altitude de 4 kilomètres de hauteur

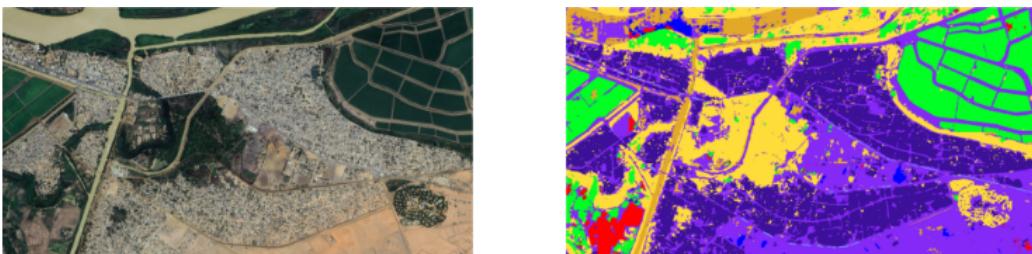


Figure 5.13 – La prédiction du modèle U-Net respectivement sur une image satellite avec une résolution de 3840x2160 et avec une altitude de 4 kilomètres de hauteur

Force est de constater que, nos modèles sont plus précis dans leurs détection lorsqu'ils reçoivent des images de hautes résolutions et plus l'altitude est diminuée et plus on peut voir les détails. Cette affirmation est faite sur une base de procédé empirique voir 5.4, 5.5, 5.6, 5.7, 5.8, 5.10, 5.12 et 5.13.

5.1.4 Généralisation sur d'autres zones du Sénégal :

Le modèle U-Net entraîné sur les images satellites de Richard Toll a été testé sur quelques zones du Sénégal. Nous remarquons que le modèle parvient à bien généralisé sur des zones hors de Richard Toll, même si les résultats peuvent être améliorés.

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

5.1.4.1 Dakar

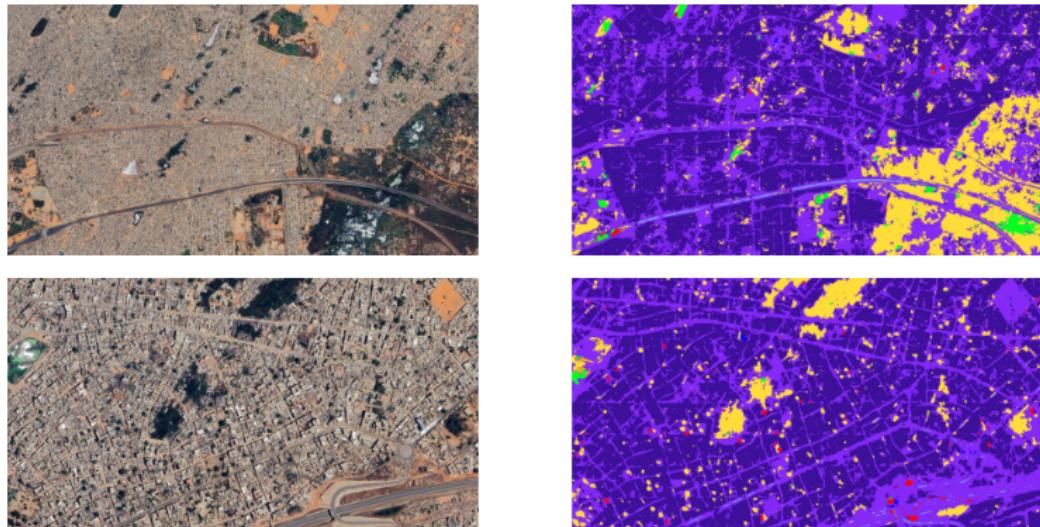


Figure 5.14 – La détection sur la zone de Dakar.

5.1.4.2 Zinguinchor

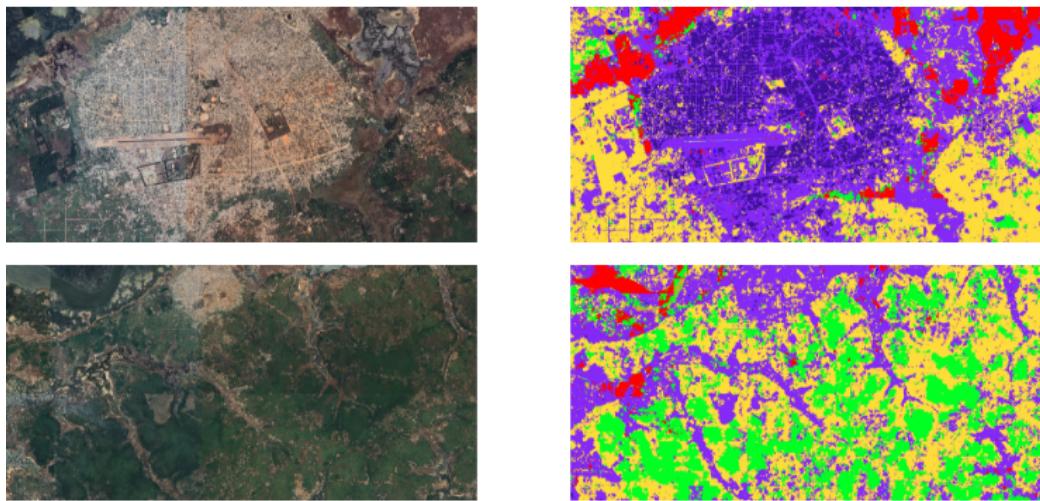


Figure 5.15 – La détection sur la zone de Zinguinchor.

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

5.1.4.3 Matam

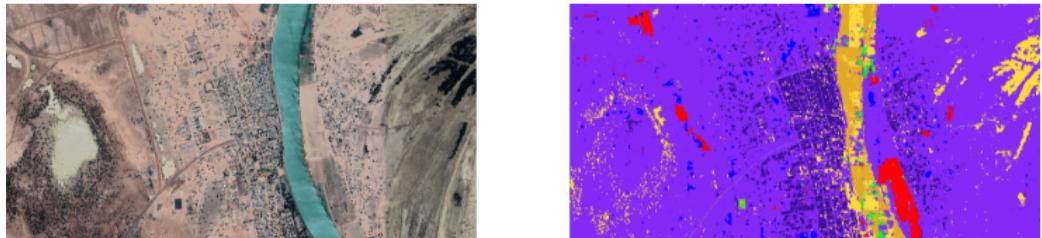


Figure 5.16 – La détection sur la zone de Matam.

5.1.5 Déploiement en temps réel :

À des fins de démonstration, des vidéos ont été capturées dans la région de Richard Toll, puis soumises à un processus de prétraitement comme décrit dans la partie « proposition de solution » avant d'être acheminées vers le modèle U-Net, spécifiquement choisi à cette fin.



Figure 5.17 – QR Code pour accéder aux démos vidéos.

5.2 Discussion des résultats : Analyse et Interprétation des résultats

Suite à la présentation des résultats de notre étude, en examinant notre travail à la lumière de l'état de l'art précédemment présenté, nous pouvons identifier des similitudes et des différences significatives dans les approches de détection de changement d'occupation du sol. Les études antérieures ont abordé la détection de changement à l'aide de diverses architectures de réseau et de modalités d'imagerie. Dans le cadre de notre analyse, nous comparons nos deux modèles, U-Net et FCN-8, à ces études, tout en mettant en évidence les caractéristiques qui distinguent notre approche.

Lorsque nous comparons nos résultats avec les études de l'état de l'art, nous pouvons identifier des similitudes et des différences significatives dans la performance de nos modèles U-Net et FCN-8.

Notre modèle U-Net a atteint une précision (accuracy) de 96%, ce qui le place parmi les modèles les plus précis dans la détection de changement.

Le coefficient de Jaccard de 92.55% pour U-Net est un indicateur fort de la capacité du modèle à mesurer la similarité entre les ensembles de pixels prédicts et réels. Cependant, l'Étude 2 a également montré des résultats impressionnantes dans ce domaine.

Le coefficient de Kappa de 94.98% pour U-Net est un indicateur de la précision du modèle en tenant compte du hasard. Ce score élevé suggère une très faible probabilité que les résultats du modèle soient dus au hasard. Cependant, l'Étude 4 a obtenu des performances comparables en termes de précision.

L'accuracy overall (OA) de 96.10% pour U-Net est une métrique globale qui résume la performance du modèle sur l'ensemble des classes. Les résultats

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

de l'Étude 2 ont également montré une excellente performance globale.

D'un autre côté, notre modèle FCN-8 a également montré des performances solides, bien que légèrement inférieures à U-Net.

FCN-8 a atteint une précision (accuracy) de 91%, ce qui est notable bien que légèrement inférieur à U-Net. Cependant, l'Étude 4 a également obtenu une précision élevée.

Le coefficient de Jaccard de 84% pour FCN-8 est un score respectable, bien que légèrement inférieur à U-Net. Cela signifie que FCN-8 a montré une bonne correspondance avec les zones de changement réelles, bien que l'Étude 2 ait obtenu des résultats similaires.

Le coefficient de Kappa de 88.55% pour FCN-8 témoigne de sa capacité à obtenir des résultats significatifs tout en minimisant l'influence du hasard, bien que l'Étude 1 ait également obtenu des résultats comparables en termes de précision.

L'accuracy overall (OA) de 91.15% pour FCN-8 indique une solide performance globale dans la détection de changement, bien que l'Étude 4 ait obtenu des performances similaires.

Lorsque nous comparons les performances de nos deux modèles, U-Net et FCN-8, dans la détection de changement, quelques distinctions clés émergent :

Précision (Accuracy) : U-Net a démontré une précision plus élevée, atteignant 96%, par rapport à FCN-8 qui a obtenu 91%. Cela suggère que U-Net a montré une tendance à être plus précis dans la classification des champs agricoles.

Coefficient de Jaccard : Le coefficient de Jaccard, mesurant la similarité entre les zones de changement prédites et réelles, a montré un avantage significatif pour U-Net avec un score de 92.55%, par rapport à FCN-8 qui a atteint 84

CHAPITRE 5. ANALYSE ET INTERPRÉTATION DES RÉSULTATS

Coefficient de Kappa : U-Net a également surpassé FCN-8 en termes de coefficient de Kappa, obtenant 94.98% par rapport à 88.55% pour FCN-8. Le coefficient de Kappa mesure la précision du modèle en tenant compte du hasard.

Accuracy Overall (OA) : L'Accuracy Overall, une métrique globale de performance sur toutes les classes, a montré une nette supériorité pour U-Net, avec un score de 96.10%, tandis que FCN-8 a obtenu 91.15

U-Net a affiché des performances supérieures par rapport à FCN-8 dans la détection de changement sur les images. Il est important de noter que nos performances ont été obtenues avec des images collectées à partir de Google Earth Pro, ce qui renforce la pertinence de nos modèles dans un contexte d'utilisation de données provenant de différentes sources. Les comparaisons effectuées nous aident à contextualiser nos résultats par rapport à l'état de l'art et à identifier les forces spécifiques de nos modèles dans la détection de changement.

Conclusion et Perspectives

La surveillance et la détection des changements d'occupations du sol sont des aspects cruciaux pour mieux comprendre l'évolution de l'environnement terrestre et pour planifier judicieusement l'utilisation des ressources. Dans le cadre du projet Grande Muraille Verte (GMV), qui vise à lutter contre la désertification et la dégradation des terres en Afrique, cette surveillance revêt une importance particulière. Notre étude s'est concentrée sur la région de Richard Toll au Sénégal, située dans la zone d'influence de la GMV, et a utilisé des techniques d'apprentissage profond pour analyser des images satellitaires géospatio-temporelles.

L'objectif de notre travail était de collecter et d'analyser des données géospatio-temporelles de la région en utilisant Google Earth Pro, en classant les zones en huit catégories distinctes, alignées avec les objectifs de la GMV. Pour ce faire, nous avons employé des méthodes basées sur les réseaux de neurones, notamment l'apprentissage par transfert et l'ajustement fin, pour améliorer la fiabilité de la surveillance et de la détection des changements d'occupation du sol.

Cependant, notre étude a révélé plusieurs défis auxquels nous avons dû faire face. Tout d'abord, les limitations de ressources matérielles ont restreint notre capacité à utiliser une quantité plus importante de données pour l'entraînement de nos modèles. Cela a inévitablement influencé la performance

de nos algorithmes. De plus, l'annotation fastidieuse des données pour la segmentation sémantique a demandé un effort considérable en termes de temps et de main-d'œuvre.

Un autre défi majeur auquel nous avons été confrontés est le déséquilibre des classes. Dans le contexte de la surveillance des changements d'occupation du sol, certaines catégories peuvent être sous-représentées par rapport à d'autres, ce qui peut entraîner des biais dans les modèles. Cela nécessite des stratégies d'équilibrage des données pour garantir une représentation équitable de toutes les classes.

Malgré ces défis, notre étude a abouti à des résultats encourageants et ouvre la voie à des perspectives futures prometteuses. Il est impératif de continuer à investir dans des ressources matérielles adéquates pour améliorer la performance des modèles. De plus, l'augmentation des données géospatio-temporelles à disposition et l'utilisation de l'apprentissage par transfert offrent des opportunités d'amélioration significative.

Dans le contexte du projet GMV [12], il est crucial d'étendre notre étude pour couvrir un spectre plus large de régions et d'environnements, en prenant en compte les spécificités de chaque zone d'influence de la GMV. Cela contribuera à une surveillance plus complète de l'évolution des terres et à une meilleure prise de décision pour la gestion durable des ressources. En outre, la collaboration avec des organismes locaux et internationaux pour partager des données et des méthodologies renforcera l'efficacité de la surveillance dans le cadre du projet GMV.

Références

- [1] ENVCAL - SUIVI DE L'ENVIRONNEMENT PAR TÉLÉDÉTECTION. <https://e-cours.univ-paris1.fr/modules/uved/envcal/html/rayonnement/index.html>.
- [2] Richard-toll. <https://fr.wikipedia.org/wiki/Richard-Toll>. 14 septembre 2022.
- [3] IRD. <https://www.ird.fr>, 2021, September.
- [4] IRDSN. www.ird.fr/senegal/presentation, 2021, September.
- [5] Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery - Scientific Figure on ResearchGate. https://www.researchgate.net/figure/Schematic-diagram-of-InceptionResNetV2-model-compressed-view_fig9_326421398, 24 Oct, 2023.
- [6] Optimisation pour l'apprentissage profond. http://www2.ift.ulaval.ca/~pgiguere/cours/DeepLearning/03-Deep_Learning_Optimization.pdf, consulté le 24 octobre 2023.
- [7] Ovi Paul et al. Abu Bakar Siddik Nayem, Anis Sarker. Lulc segmentation of rgb satellite image using fcn-8. <https://arxiv.org/ftp/arxiv/papers/2008/2008.10736.pdf>, 24 Aug 2020.

RÉFÉRENCES

- [8] Ahmed Negida Alireza Baratloo, Mostafa Hosseini and Gehad El Ashal. Simple Definition and Calculation of Accuracy, Sensitivity and Specificity. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4614595/>, 2015.
- [9] Nada Belaidi. U-Net : le réseau de neurones populaire en Computer Vision. <https://blent.ai/blog/a/unet-computer-vision>, 8 mars 2022.
- [10] Vitaly Bushaev. Adam — latest trends in deep learning optimization. <https://towardsdatascience.com/adam-latest-trends-in-deep-learning-optimization-6be9a291375c>, Oct 22, 2018.
- [11] Andrew Joseph Davies. Semantic Segmentation of Aerial Imagery Using U-Net in Python. <https://towardsdatascience.com/semantic-segmentation-of-aerial-imagery-using-u-net-in-python-552705238514>, Mar 31, 2022.
- [12] Agence Panafricaine de la Grande Muraille Verte. Forum de l’investissement pour la gmv. <https://www.grandemurailleverte.org/>, Consulté le 28 octobre 2023.
- [13] IRD CIRAD ENSG INA-PG ENSAM. Télédétection des changements et sig application à l’étude de l’évolution des modes d’occupation des sols à la reunion entre 1989 et 2002. https://agritrop.cirad.fr/551257/1/document_551257.pdf, Novembre 2003.
- [14] Gaël Kerfendal et Yann Kerfendal. Télédetection : Les différents satellites et indices. <https://veillecarto2-0.fr/2021/04/12/teledetection-les-differents-satellites-et-indices/>. 10 octobre 2023.
- [15] Julian LE GOUIC. Computer vision, machine learning. <https://www.quantmetry.com/blog/>

RÉFÉRENCES

- choix-fonction-de-perte-en-computer-vision-partie-2/,
12/07/2022.
- [16] J-P. BOUTIN I. BERGERI, R. MICHEL. POUR TOUT SA-
VOIR OU PRESQUE SUR LE COEFFICIENT KAPPA... [https://www.jle.com/fr/MedSanteTrop/2002/62.6/634-636%20Pour%20tout%20savoir%20ou%20presque%20sur%20le%20coefficient%20Kappa%20\(Bergeri\).pdf](https://www.jle.com/fr/MedSanteTrop/2002/62.6/634-636%20Pour%20tout%20savoir%20ou%20presque%20sur%20le%20coefficient%20Kappa%20(Bergeri).pdf), 2002.
- [17] Oumar Sy et Tidiane Sané Ibrahima Diédhiou, Catherine Mering. Carto-
graphier par télédétection l'occupation du sol et ses changements. <https://journals.openedition.org/echogeo/20510#toctoin3>, Consulté le
24 octobre 2023.
- [18] Evan Shelhamer Jonathan Long and Trevor Darrell. Fully Convo-
lutional Networks for Semantic Segmentation. https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Long_Fully_Convolutional_Networks_2015_CVPR_paper.pdf, 2014.
- [19] TOM KELDENICH. Data Augmentation, améliorer rapidement son
modèle de Deep Learning. <https://inside-machinelearning.com/data-augmentation-ameliorer-rapidelement-son-modele-de-deep-learning/>,
22 FÉVRIER 2021.
- [20] Rohit Kundu. F1 Score in Machine Learning : Intro Calculation. <https://www.v7labs.com/blog/f1-score-guide#:~:text=F1%20score%20is%20a%20machine%20learning%20evaluation%20metric%20that%20measures,prediction%20across%20the%20entire%20dataset.,>
December 16, 2022.
- [21] Francisco Javier García-Haro et al. Manuel Campos-Taberner. Under-
standing deep learning in land use classification based on Sentinel-2 time

RÉFÉRENCES

- series. <https://www.nature.com/articles/s41598-020-74215-5>, 14 octobre 2020.
- [22] Mathieu De Maeyer et Eléonore Wolff Matthieu Kayembe Wa Kayembe. Cartographie de la croissance urbaine de Kinshasa (R.D. Congo) entre 1995 et 2005 par télédétection satellitaire à haute résolution. <https://journals.openedition.org/belgeo/7349#tocto1n4>. 10 octobre 2023.
- [23] Evis Sala et al. Michael Yeung. Unified Focal loss : Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. <https://arxiv.org/pdf/2102.04525.pdf>, 24 novembre 2021.
- [24] Mohammad Yousuf Saifi ; Jimmy Singla ; Nikita. Deep Learning based Framework for Semantic Segmentation of Satellite Images. <https://ieeexplore.ieee.org/document/9076551>, 20 avril 2020.
- [25] BELGIAN EARTH OBSERVATION. corrections géométriques et radiométriques. <https://eo.belspo.be/fr/corrections-geometriques-et-radiometriques>. 09 octobre 2023.
- [26] Philipp Fischer Olaf Ronneberger and Thomas Brox. U-Net : Convolutional Networks for Biomedical Image Segmentation . <https://arxiv.org/pdf/1505.04597.pdf>, 2015.
- [27] Xiao Xiang Zhu Patrick Ebel, Sudipan Saha. FUSING MULTIMODAL DATA FOR SUPERVISED CHANGE DETECTION. <https://isprs-archives.copernicus.org/articles/XLIII-B3-2021/243/2021/isprs-archives-XLIII-B3-2021-243-2021.pdf>, 2021.
- [28] Data Science. Transfer Learning : Qu'est-ce que c'est ? <https://datascientest.com/transfer-learning>, consulté le 27 octobre 2023.

RÉFÉRENCES

- [29] Towards Data Science. Review : FCN — Fully Convolutional Network (Semantic Segmentation). <https://towardsdatascience.com/review-fcn-semantic-segmentation-eb8c9b50d2d1>, 5 octobre 2018.
- [30] Dr. Info Sec. VGG-19 Convolutional Neural Network. <https://blog.techcraft.org/vgg-19-convolutional-neural-network/>, March 6, 2021.
- [31] Paweł Siciarz and Boyd McCurdy. U-net architecture with embedded Inception-ResNet-v2 image encoding modules for automatic segmentation of organs-at-risk in head and neck cancer radiation therapy based on computed tomography scans. <https://iopscience.iop.org/article/10.1088/1361-6560/ac530e>, 22 June 2022.
- [32] Ilyes Talbi. La revue ia. <https://larevueia.fr/question-ce-que-la-segmentation-dimages/>. 24 décembre 2023.
- [33] Great Learning Team. Fully Convolutional Network (Semantic Segmentation). <https://www.mygreatlearning.com/blog/fcn-fully-convolutional-network-semantic-segmentation/>, 31 octobre 2022.
- [34] Tensorflow. LSegmentation des images. <https://www.tensorflow.org/tutorials/images/segmentation?hl=fr>, Consulté le 20 octobre 2023.
- [35] Redha Touati. D'étection de changement en imagerie satellitaire multimodale, PAIRWISE DESCRIPTORS LEARNING FOR MULTIMODAL CHANGE DETECTION USING PSEUDO-SIAMESE CNN NETWORK MODEL . <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/22662?locale-attribute=fr>. Avril, 2019.
- [36] Nagesh Kumar Uba. Land Use and Land Cover Classification Using Deep Learning Techniques. <https://arxiv.org/ftp/arxiv/papers/1905/1905.00510.pdf>, April 2016.

RÉFÉRENCES

- [37] JONGSEOK LEE WAHYU WIRATAMA and DONGGYU SIM. Change Detection on Multi-Spectral Images Based on Feature-level U-Net. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8952681>, 21 Janvier 2020.
- [38] Wikipedia. Télédétection. <https://fr.wikipedia.org/wiki/T%C3%A9l%C3%A9d%C3%A9tection#:~:text=La%20t%C3%A9l%C3%A9d%C3%A9tection%20est%20l'ensemble,qu'ils%20%C3%A9mettent%20ou%20r%C3%A9fl%C3%A9chissent>. 15 octobre 2023.
- [39] Wikipédia. U-Net. <https://fr.wikipedia.org/wiki/U-Net>, 22 avril 2023.
- [40] Wikipédia. Toggle the table of contents Fine-tuning (deep learning). [https://en.wikipedia.org/wiki/Fine-tuning_\(deep_learning\)](https://en.wikipedia.org/wiki/Fine-tuning_(deep_learning)), last updated 21 October 2023.