

Rapport de stage

Mastère spécialisé® Photogrammétrie, Positionnement, Mesure de Déformations

Utilisation de données Sentinel-2 et SPOT 6/7 pour la classification de l'occupation du sol



Olivier STOCKER

Septembre 2019

Non confidentiel Confidentiel IGN Confidentiel Industrie Jusqu'au ...

Jury

Président de jury :

Jean-François HANGOUET

Commanditaire :

Arnaud LE BRIS

Encadrement de stage :

Arnaud LE BRIS

Enseignant référent :

Sébastien GIORDANO

Rapporteur expert :

Clément MALLET

Responsable du MS® PPMD :

Jacques BEILIN, IGN/ENSG/DPTS

Antoine PINTE, IGN/ENSG/DIAS

Gestion du stage :

Anna CRISTOFOL, IGN/ENSG/DSHEI

© ENSG

Stage de fin d'étude du xxx au xxx

Diffusion web : Internet Intranet Polytechnicum Intranet ENSG

Situation du document :

Rapport de stage de fin d'études présenté en fin de 3^{ème} année du cycle des Ingénieurs

Nombres de pages : 60 pages dont 4 d'annexes

Système hôte : L^AT_EX

Modifications :

EDITION	REVISION	DATE	PAGES MODIFIEES
1	0	09/2016	Création

Remerciements

Je remercie avant tout Arnaud Le Bris pour sa patience et ses enseignements.

Je tiens à remercier Clément Mallet, pour m'avoir accepté au sein de son équipe, Loic Landrieu et Nathan Piasco pour leurs leçons de deep learning précieuses, ainsi qu'Emile Blettery et Stéphane Guinard pour leurs relectures.

Abstract

Cette étude porte sur le développement d'une architecture entièrement convolutive, adaptée au traitement de l'information spatiale apportée par la très haute résolution des capteurs SPOT 6 et 7. Cette architecture s'est montrée plus performante que les approches par fenêtre glissante dans la précision de la détection des objets topographiques, même en zone dense. Parallèlement ces travaux montrent que l'ajout de contraintes permet de mieux délimiter les objets et que la finesse de la vérité terrain joue un grand rôle dans cette capacité de délimitation. Cette nouvelle architecture a également permis de générer, à partir de produits existants, des cartes de couverture du sol d'une qualité prometteuse. Les différents niveaux de richesse de nomenclatures évalués ont mis en avant une capacité de constance dans la segmentation sémantique. Enfin, ces travaux ont servi d'étude préliminaire à la fusion tardive et précoce des données SPOT 6/7 et Sentinel 2, dans l'objectif d'ajouter à la richesse spatiale, déjà efficace, une dimension spectrale. L'ensemble des contraintes liées à l'implantation entièrement convolutive de la fusion et les modifications à appliquer sur notre architecture ont été listées.

Mots clés : Apprentissage profond, Cartes d'occupation du sol, SPOT 6/7, Sentinel 2, Fusion tardive et précoce, Segmentation sémantique

Abstract

This study addresses the development of a fully convolutional architecture, suitable for the processing of spatial information provided by the very high resolution of the SPOT 6/7 sensors. This approach happened to outperform the sliding window approaches regarding the precision of the detection of topographic objects, even in concentrated areas. Simultaneously, this work demonstrates that adding constraints enables a better delimitation of the objects, delimitation in which the precision of the ground truth plays a great part. This new architecture also allowed the creation of promising land-cover maps, based on existing products. The different levels of comprehensiveness of the nomenclature that have been evaluated seem to indicate a certain capacity for consistency when it comes to semantic segmentation. Finally, this work formed the preliminary study to early and late fusion of SPOT 6/7 and Sentinel 2 data. The objective being to add a spectral dimension to the efficient spatial information. All the constraints related to the fully convolutional implementation of the fusion and the modifications needed on the existing architecture have been listed for future work.

Key words: Deep Learning, Land-cover maps, SPOT 6/7, Sentinel 2, late and early fusion, semantic segmentation

Table des matières

Introduction	1
1 Classification de l'occupation du sol	3
1.1 Généralités	3
1.1.1 Occupation	3
1.1.2 Classification et segmentation sémantique	3
1.1.3 Intelligence artificielle et deep learning	3
1.1.4 Imagerie satellite	4
1.2 Cadre du projet	4
1.2.1 CES OSO et Urbain du Pôle Théia	5
1.3 État de l'art	5
1.3.1 Travaux IGN	5
1.3.2 Travaux extérieurs	6
2 Données et traitement	7
2.1 Bases de données et produits d'OCS existants	7
2.1.1 BD Topo®	7
2.1.2 RPG	7
2.1.3 OCS-GE	7
2.1.4 Urban ATLAS	7
2.1.5 OSO	7
2.2 Zones d'intérêt	8
2.2.1 Données	8
2.2.2 Pré-traitements	8
2.3 Nomenclatures	9
2.3.1 Couverture du sol	9
2.3.2 Usage du sol	11
2.3.3 Mixtes	11
3 Algorithmique	12
3.1 Réseau de neurones	12
3.2 Architecture proposée	13
3.3 Chaîne de traitement	13
3.3.1 Préparation des données d'entrée	13
3.3.2 Validation croisée	15
3.3.3 Entraînement du réseau	15
3.3.4 Inférence	17
3.4 Estimation de l'aléatoire	17
3.4.1 Aléatoire de la répartition dans la validation croisée	17
3.5 Analyse de l'architecture	19
3.5.1 Données des tests	19
3.5.2 Convolutions	19
3.5.3 Normalisation	21
3.5.4 Connexions résiduelles	22
3.5.5 Décodeur convolutif	23
3.5.6 Profondeur du réseau	23
3.5.7 Conclusion de l'étude de l'architecture	23

4	Segmentation sémantique entièrement convolutive	25
4.1	Étude méthodologique	25
4.1.1	Forêt aléatoire	25
4.1.2	Fenêtre glissante	26
4.1.3	Approche binaire	27
4.2	Transférabilité des modèles	29
4.2.1	Transfert spatial	29
4.2.2	Transfert temporel	30
4.3	Étude thématique	31
4.3.1	Détection d'instance de bâtiment	31
4.3.2	Comparaison avec l'approche seuillage	32
4.3.3	Retrait	32
4.3.4	Raffinement des nomenclatures	32
4.3.5	Vérité terrain issue de classification antérieures	36
4.4	Autres études non détaillées	39
4.5	Conclusion générale sur l'analyse des résultats	39
5	Segmentation par fusion	41
5.1	Retour sur l'état de l'art	41
5.1.1	Absence d'architecture complètement convolutive	41
5.1.2	Avantages et contraintes de la nouvelle approche	41
5.2	Adaptation à la fusion	42
5.2.1	Modification de l'architecture	42
5.2.2	Modification de la chaîne de traitement	42
5.3	Résultats de fusion	43
5.3.1	Jeu de données	43
5.3.2	Fusion des canaux SPOT	43
5.3.3	Fusion SPOT S2	45
5.3.4	Module de transformation spatiale	47
5.4	travaux futurs	48
	Conclusion	49
	Bibliographie	51
	A Détails de nomenclatures OCS-GE et UA	59

Glossaire

- ASP** Agence de Service et Paiement. 7
- BD** Base de Données. 3, 7–9, 11, 26, 28, 31, 36, 49
- BN** Normalisation par paquet ou *Batch Normalization*. 21–23, 53
- CES** Centres d’Expertise Scientifique. 1, 5, 8, 49
- CLC** CORINE Land Cover. 3
- CM** Matrice de Confusion. 17
- CNES** Centre National d’Etudes Spatiales. 4
- CNN** Réseaux de neurones convolutifs ou *Convolutionnal Neural Network*. 1, 4–6, 9, 15, 17, 19, 25, 26, 28, 31, 32, 34–37, 39, 41, 47–49, 53, 55
- CR** Connexions Résiduelles. 22, 23, 53, 55
- CS** Couverture du sol. 1, 3, 5, 7–11, 18, 26, 39, 49
- CV1** première itération de la validation croisée. 17–19, 53, 55
- DpLg** Apprentissage profond ou *Deep Learning*. 3–6, 12, 13, 16, 26, 37
- FT** Transfert d’apprentissage ou *Fine Tuning*. 29, 30, 53
- GPU** Carte graphique ou *Graphic Processing Unit*. 13, 14, 17
- GS** jeu de données global ou *Global dataSet*. 14, 15, 55
- IA** Intelligence Artificielle. 3
- IC-95** Intervalle de Confiance à 2 écarts-types. 21, 22
- IC-99** Intervalle de Confiance à 3 écarts-types. 19–22
- IGN** Institut national de l’information géographique et forestière. 1, 5, 7, 27, 37
- ILSVRC** ImageNet Large Scale Visual Recognition Competition. 4, 12
- IOTA2** Infrastructure pour l’Occupation des sols par Traitement Automatique Incorporant les Orfeo Toolbox Applications. 2, 5, 7
- IoU** intersection sur l’union ou *Intersection over Union*. 16–21, 23, 26, 29, 33–35, 38, 45, 46, 55
- LaSTIG** Laboratoire en Sciences et Technologies de l’Information Géographique. 5
- mIoU** moyenne des intersections sur l’union ou *mean Intersection over Union*. 16–23, 25, 29–34, 36, 38, 45–47, 53

- ML** Apprentissage automatique ou *Machine Learning*. 1, 3–6, 14, 25, 27, 39, 41, 49
- MLP** Multi Layer Perceptron. 6, 41
- MNS** Modèles Numériques de Surface. 1, 6, 37
- MS** Multi-Spectraux. 6, 8, 40, 41, 43, 44, 47
- OA** précision globale ou *Overall Accuracy*. 16–19
- OCS** Occupation du Sol. 1, 3–8, 12, 24, 36, 39, 49
- OCS-GE** Occupation du Sol à Grande Échelle. 1, 7, 8, 10, 11, 39
- OSO** Carte d’occupation des sols du CES OSO. 1, 2, 4, 5, 7–10, 36, 39, 40, 48, 49
- PAC** Politique Agricole Commune. 7
- PAN** Panchromatique. 4, 6, 8, 41, 43, 44
- PXS** Multi-Spectrales pansharpenés de SPOT. 43–47
- RF** Forêts d’arbre de décisions aléatoires ou *Random Forest*. 1, 5–7, 10, 25, 26, 28, 37, 38, 48, 53, 55
- RGE®** Référentiel à Grande Échelle. 7
- RNN** Réseau de neurones récurrents ou *Recurrent Neural Network*. 6, 41
- RPG** Registre Parcellaire Graphique. 7–9
- S1** Sentinel-1. 6
- S2** Sentinel-2. 1, 2, 4, 6–10, 25, 39–41, 43, 46–49, 53
- SegSem** Segmentation Sémantique. 1–3, 5–7, 9, 12–14, 17–19, 25, 27–29, 32, 34, 36–39, 41–45, 47, 49, 53
- SGD** Descente de Gradient Stochastique. 14, 16
- SVM** Séparateur à Vaste Marge. 5
- TE** jeu de données de TEst. 15, 19, 42, 45
- THR** Très Haute Résolution. 4–6, 9, 12, 19, 22, 26, 32, 39–42, 47, 49
- TOSCA** Terre, Océan, Surfaces continentales, Atmosphère. 4, 5
- TR** jeu de données d’enTRînement. 15, 19, 42
- UA** Urban Atlas. 7–9, 11, 39
- US** Usage du Sol. 3, 6–8, 11, 33, 39, 49
- VAL** jeu de données VALidation. 15, 16, 19, 42
- VT** Vérité Terrain. 7–10, 13–23, 25–40, 43, 44, 46, 47, 49, 53, 55

Introduction

L'ouverture de la donnée d'observation terrestre par le Pôle Théia aux entités publiques et scientifiques a pour objectif de promouvoir leur utilisation et leur dérivation en produit tels que les cartes d'Occupation du Sol (OCS). Ces dernières sont des outils vitaux pour le suivi des changements environnementaux et climatiques, pour la gestion du territoire, et de multiples autres applications à caractères scientifiques ou politiques.

Si ces produits étaient uniquement saisis par photo-interprétation, l'arrivée de l'Apprentissage automatique ou *Machine Learning* (ML) a permis d'automatiser, ou tout du moins d'en faciliter la production. Bien que les produits photo-interprétés comme l'Occupation du Sol à Grande Échelle (OCS-GE) font preuve d'une excellente qualité, leur mise à jour s'avère longue et fastidieuse. À l'inverse, les produits générés, uniquement par ML bénéficient d'une rapidité de production importante, au prix d'une précision dégradée et d'erreurs liées à la Segmentation Sémantique (SegSem). Le meilleur exemple actuel français de Couverture du sol (CS) est le produit Carte d'occupation des sols du CES OSO (OSO) [Inglada et al., 2017], générée automatiquement et annuellement à partir de données Sentinel-2 (S2).

La résolution spatiale de S2 ne permet pas la détection d'objets topographiques fins, en particulier dans les zones urbaines. Or, la demande de cartes d'OCS dans les villes est forte, comme le montre le développement de projets comme URCLIM [URCLIM, 2019] ou le Centres d'Expertise Scientifique (CES) Urbain du Pôle Théia. Cette étude entre dans le cadre de ce dernier, qui a pour objectif de favoriser le développement de méthodes et d'outils de télédétection appliqués au milieu urbain. Le Pôle Théia diffuse dans ce but un flux de données qui permet la visualisation des détails urbains : les images SPOT (canal panchromatique à 1,5m de résolution).

Si les algorithmes de ML classiques comme les Forêts d'arbre de décisions aléatoires ou *Random Forest* (RF) [Breiman, 2001] suffisent à l'analyse de séries temporelles S2, le fait qu'ils se basent sur des attributs prédéfinis limite leur capacité de généralisation. Les progrès permanents de ces dernières années dans les domaines du ML et de la vision par ordinateur ont popularisé les Réseaux de neurones convolutifs ou *Convolutional Neural Network* (CNN), qui sont très performants dans les tâches de classification d'images. Leur force réside dans leur capacité à détecter des textures et des motifs. Le projet interne de l'Institut national de l'information géographique et forestière (IGN) : TERMOS, applique ces méthodes pour la production d'OCS. Leur implantation d'architectures de CNN entièrement convolutifs pour le traitement d'images aériennes et de Modèles Numériques de Surface (MNS) montre d'excellents résultats.

Nos travaux se placent dans la continuité de [Postadjian et al., 2017], qui a investigué l'emploi de CNN pour la SegSem de cartes de CS à partir de données SPOT 6/7. Les résultats prometteurs pour une nomenclature à 5 classes, composée des objets topographiques les plus courants (*Eau, Cultures, Végétation, Routes et Bâtiments*) ont encouragé la poursuite de ces travaux. Néanmoins l'architecture volontairement très légère et basée sur une approche de fenêtre glissante n'exploite pas au mieux l'information spatiale contenue dans les images SPOT.

Nos travaux portent donc sur l'amélioration de cette architecture vers une approche entièrement convolutive, similaire à celle du projet TERMOS, puis au raffinement des nomenclatures afin de caractériser les performances du capteur SPOT dans la détection d'objets topographiques plus complexes et se rapprochant de la nomenclature OSO.

Cette investigation rentre dans le cadre du projet PARCELLE, qui a pour objectif de mutualiser et

capitaliser les développements sur la chaîne de traitement Infrastructure pour l'Occupation des sols par Traitement Automatique Incorporant les Orfeo Toolbox Applications (IOTA2). Les futurs développement de cette chaîne s'orientent vers l'utilisation conjointe des données SPOT et S2 pour améliorer :

1. la résolution de la SegSem,
2. la performance de la détection des éléments urbains.

Ainsi, nos travaux s'étendent également à la comparaison entre les résultats des SegSem depuis la donnée SPOT et le produit OSO, pour identifier la pertinence des deux capteurs sur les différentes classes des nomenclatures, puis à la mise en oeuvre d'une architecture prototype pour le traitement conjoint par fusion (tardive et précoce) des deux données.

Ce document s'articule en 5 parties :

- Généralités : Présentation des projets PARCELLE et CES Urbain puis de l'état de l'art,
- Données et traitement : Présentation des jeux de données, leurs pré-traitements et les nomenclatures utilisées,
- Algorithmique : Présentation de notre architecture et de la chaîne de traitement qui l'encapsule,
- Segmentation sémantique entièrement convolutive : Détails et conclusions sur l'ensemble des tests réalisés concernant l'approche complètement convolutive,
- Segmentation sémantique par fusion : Présentation des premiers travaux sur la fusion tardive/précoce et des travaux futurs.

CLASSIFICATION DE L'OCCUPATION DU SOL

CHAPITRE 1

Ce chapitre a pour objectif d'introduire les concepts généraux nécessaires à la compréhension de l'étude réalisée, de répertorier les travaux déjà effectués à l'IGN et ailleurs, puis de présenter les zones d'intérêt et leurs nomenclatures sélectionnées pour nos travaux.

1.1 Généralités

1.1.1 Occupation

L'occupation des sols est un point essentiel dans la gestion des territoires par nos sociétés. Pour aider à sa caractérisation, deux notions fondamentales ont été introduites :

- La CS, qui définit sa couverture physique (zone bâtie, Peuplements de feuillus, Peuplements de conifères, etc.) ,
- L'Usage du Sol (US), qui définit sa couverture utilitaire et anthropique (habitat résidentiel, commerces, industries, sylviculture, zones abandonnées).

Ces informations géographiques sont habituellement présentées sous forme de Base de Données (BD) cartographique. Ces bases représentent des outils essentiels pour la planification urbaine, le suivi des politiques agricoles, environnementales et climatiques. Certaines BD d'OCS mélangent souvent des objets de définitions physique et anthropique (ex. CORINE Land Cover (CLC)).

L'OCS et l'US évoluent au cours du temps. Leurs cartographies ne sont que des "images" à un instant donné (millésimes) et ont donc besoin de mises à jour régulières.

Si l'CS est compatible avec les techniques de télédétection, l'US ne présente pas cette propriété et est plus difficilement dérivable à partir d'observations de la terre.

1.1.2 Classification et segmentation sémantique

Nos travaux portent principalement sur l'identification de la couverture du sol. Dans les domaines de l'Intelligence Artificielle (IA), et plus précisément, en vision par ordinateur, les termes *classification*, *segmentation* et *segmentation sémantique* désignent des résultats différents :

- **Classification** : labellisation d'une image (1 label/image)
- **Segmentation** : partitionnement en régions de contenu homogène
- **Segmentation sémantique** : labellisation de chaque pixel d'une image (1 label/pixel)

Ces distinctions sont cruciales pour la bonne compréhension de ce document.

Dans le domaine de la télédétection, "classification" définit très généralement une opération de SegSem

1.1.3 Intelligence artificielle et deep learning

l'IA est le domaine d'étude dans lequel des systèmes sont créés pour exécuter des tâches qui nécessitent un haut niveau de compréhension. Le ML est une spécialisation de l'IA qui a pour objectif de réaliser des tâches pour lesquelles il n'a pas été explicitement codé. Il se différencie du reste de l'IA par l'absence de règles heuristiques définies par un humain (comme les IA de jeux vidéos). Les algorithmes classiques de ML apprennent à partitionner l'espace des descripteurs des données d'entrée. Ils ne peuvent pas remodeler les données d'entrée vers des représentations plus ségrégatives (notamment en ce qui concerne la description de la texture des images). Ainsi ils restent dépendant de l'humain sur l'extraction d'attributs. C'est sur ce point que le Apprentissage profond ou *Deep Learning* (DpLg) se sépare du ML. La performance des

algorithmes de deep learning réside dans cette capacité d'extraction, depuis des données d'entrée brutes, de caractéristiques discriminantes (appelées "embeddings", [Will Koehrsen, 2018]). La démonstration de cette performance a été réalisée en 2012 par le réseau "Alex-Net" [Krizhevsky et al., 2012] durant la ImageNet Large Scale Visual Recognition Competition (ILSVRC). Il a réalisé un taux de classification de 10% supérieur à ses compétiteurs de ML classiques.

Il existe plusieurs types d'algorithmes de DpLg. Les CNN sont ceux utilisés dans les travaux présentés ici. Ils sont basés sur les opérations de convolutions introduites par [LeCun et al., 1995] et particulièrement efficaces pour les tâches de reconnaissances visuelles.

1.1.4 Imagerie satellite

Plusieurs constellations de satellites d'observation orbitent autour de la terre. Elles acquièrent en permanence des images de la surface du sol, et nous offrent des données vitales pour la caractérisation des l'OCS. Dans le cadre de nos travaux, nous utilisons les données diffusées par le Pôle Théia.

1.1.4.1 Sentinel-2

Les données Sentinel-2 proviennent d'un couple de satellites, S-2A et S-2B. Ils orbitent à 786 km et fournissent des images de 290 km de largeur. Le tableau ci-dessous décrit les caractéristiques principales de leurs instruments.

	Bande	B1	B2	B3	B4	B5	B6	B7	B8	B8a	B9	B10	B11	B12
	Domaine	Vi	B	V	R	PIR	PIR	PIR	PIR	PIR	IR	IR	IR	IR
	Résolution au sol (m)	60	10	10	10	20	20	20	10	20	60	60	20	20
Longueur d'onde	centrale (nm)	442	492	559	664	704	740	780	832	864	944	1375	1612	2194
	Largeur de la bande (nm)	21	66	36	31	15	15	20	106	21	20	31	91	175

Table 1.1 – Caractéristiques des instruments sur Sentinel-2

Vi : Violet, R : Rouge, V : Vert, B : Bleu, PIR : Proche Infra-Rouge, IR : Infra-Rouge, source : [ESA, 2019]

Le temps de revisite de 2 à 3 jours permet d'obtenir des séries temporelles dense à l'échelle de la France. Des traitements ont été développés afin d'en déduire de synthèses mensuelles sans nuages. Ces produits sont principalement utilisés pour leur capacité de détection des cultures et de la végétation du fait de leur forte résolution spectrale. Les cartes OSO sont créées à partir d'images S2.

1.1.4.2 SPOT 6/7

Les satellites SPOT 6 et 7 forment une constellation de capteurs passifs qui orbitent à 694 km d'altitude. Le tableau ci-dessous décrit les caractéristiques principales de leurs instruments. Contrairement à S2

	Bande	B0	B1	B2	B3	B4
	Domaine	B	V	R	PIR	Panchromatique
	Résolution au sol (m)	6	6	6	6	1.5
Longueur d'onde	(nm)	450-520	530-590	625-695	760-890	450-745

Table 1.2 – Caractéristiques des instruments SPOT 6 et 7

R : Rouge, V : Vert, B : Bleu, PIR : Proche Infra-Rouge, source : [Astrium, 2019]

la couverture nationale n'est pas aussi régulière, avec seulement 1 à 3 images par zone et par an. En revanche la Très Haute Résolution (THR) qu'apporte le canal Panchromatique (PAN) donne accès à une information de texture que les capteurs S2 ne possèdent pas. Il y a donc une forte complémentarité spectrale et spatiale entre S2 et SPOT.

1.2 Cadre du projet

Cette étude est financée par le guichet de financement Terre, Océan, Surfaces continentales, Atmosphère (TOSCA) du Centre National d'Etudes Spatiales (CNES) par l'intermédiaire du projet PARCELLE et

s'inscrit dans le cadre des CES Urbain et OSO du Pôle Théia.

1.2.1 CES OSO et Urbain du Pôle Théia

Pôle Théia Le Pôle Théia [Theia, 2019] a pour objectif de "*faciliter l'usage des images issues de l'observation des surfaces continentales depuis l'espace*". Ce Pôle offre un système de distribution de données géomatiques pour promouvoir leur utilisation au sein de l'administration publique, et stimuler de la recherche. Il est composé de 23 CES qui ont pour mission de valoriser et de créer des produits à valeurs ajoutées, selon leurs thématiques. Ces CES sont composés d'équipes de recherches réparties dans les laboratoires français.

1.2.1.1 CES Urbain

Le CES Urbain se concentre sur l'identification des zones artificialisées du territoire. Ce sujet concerne aussi bien la délimitation de la tâche urbaine à grande échelle que l'identification des surfaces imperméables et des surfaces végétalisées à échelle plus fine en zone urbaine. L'IGN participe à la réalisation de 2 des produits du CES Urbain :

- *Tâche artificialisée* : Production annuelle d'une carte à l'échelle nationale de la tâche urbaine à une résolution de 10m à partir des données diffusées par le Pôle Théia
- *Formes urbaines* : Production annuelle d'une carte à l'échelle nationale de la tâche urbaine avec une nomenclature à 4 classes de tissu urbain, *Dense* (centres-ville), *Discontinu* (zones pavillonnaires), *Zones d'activités* (commerciales et industrielles), *Emprises de communication* (infrastructures de transport)

Les travaux de notre étude portent sur ces deux projets, avec une détection des surfaces artificialisées, puis le raffinement de la nomenclature pour atteindre le niveau d'exigence du produit *Formes urbaines*.

1.2.1.2 CES OSO

Le projet PARCELLE s'inscrit dans la continuité du projet TOSCA OSO qui a permis le développement de la chaîne de traitement IOTA2 [CESBIO, 2019]. Son objectif est de promouvoir IOTA2 sur ses aspects méthodologiques et thématiques. L'IGN contribue aux deux aspects, et en l'occurrence à l'ajout de nouveaux algorithmes et fonctionnalités. Cette étude se place en tant que travaux préalables à l'intégration d'un module de traitement de la donnée THR et d'un module de fusion.

1.3 État de l'art

1.3.1 Travaux IGN

L'IGN et son Laboratoire en Sciences et Technologies de l'Information Géographique (LaSTIG) conduisent des travaux concernant la classification de l'OCS à partir d'images de télédétection.

1.3.1.1 Deep learning et THR

Les précédents travaux du LaSTIG ont été centrés sur l'investigation de la capacité du DpLg pour la SegSem des images THR. [Postadjian et al., 2017] utilise des méthodes de classification par fenêtre glissante (approche patch-based) pour réaliser la SegSem d'une CS à 5 classes. Cette classification est effectuée par un CNN. Des comparaisons sur la rapidité d'apprentissage des modèles ont été effectuées, entre initialisations aléatoires des poids des réseaux (entraînement à partir de zéro) et *fine-tuning* (entraînement à partir d'un modèle pré-entraîné). Une variante de l'approche précédente, [Postadjian et al., 2018b], a permis de réduire les temps de calcul de 80% tout en améliorant la capacité de prédiction d'environ 2% (d'après les indices Kappa et le F-score).

Dans l'objectif d'enrichir la nomenclature utilisée dans [Postadjian et al., 2017], une technique d'adaptation de domaine a été établie par [Postadjian et al., 2018a].

D'autre part des travaux sur la détection de la tâche urbaine ont été menés par [Wendl et al., 2018]. Cet article compare les techniques de fusion tardive et les ML classiques (RF, Séparateur à Vaste Marge (SVM), etc.) pour la fusion de classification tardive. Leurs données d'entrée sont une inférence de RF sur Sentinel-2 et une inférence de [Postadjian et al., 2017] sur SPOT 6.

1.3.1.2 Projet TERMOS

Indépendamment des unités de recherches, une équipe de l'IGN s'est formée pour mettre en place et évaluer une chaîne de traitement des images THR aériennes de la BD Ortho® et des MNS associés par DpLg pour la SegSem de l'OCS. Plusieurs architectures plus opérationnelles ont été testées. EN particulier une architecture basée sur [Ronneberger et al., 2015] à été mise en oeuvre sur une de nos zones d'étude. Ils génèrent des cartes de probabilité par classe, en raison de leur approche binaire. C'est une règle heuristique (seuil) qui définit la SegSem finale. Les zones dont les probabilités pour chaque classe sont inférieures à leur seuil respectif sont labélisées comme "autres".

1.3.2 Travaux extérieurs

Depuis l'ouverture des images S2 au public, plusieurs études ont été menées pour évaluer son potentiel de SegSem de l'OCS et US. Les travaux de [Pelletier et al., 2019] introduisent des convolutions temporelles pour le traitement des séries temporelles S2. [Sabo et al., 2018] comparent les données S2 aux données Sentinel-1 (S1), LandSat et SPOT 5 par méthode de ML pour la détection de tâche urbaine. La conclusion donne aux données S2 une capacité supérieure à S1 et LandSat pour la détection de bâtis et de végétation, néanmoins la THR de SPOT 5 reste la plus performante dans ce cas d'application.

Dans l'idée de combiner la performance spectrale de S2 avec la performance spatiale de SPOT 6 et 7, l'équipe de [Benedetti et al., 2018] a élaboré une architecture DpLg à double entrée, SPOT et série temporelle S2. L'architecture traite l'image SPOT avec un CNN classique, tandis qu'elle traite la série S2 par Réseau de neurones récurrents ou *Recurrent Neural Network* (RNN). La fusion est effectuée par simple concaténation des deux embeddings avant une fonction Soft-Max. Ils ont réussi à surpasser les performances d'un RF sur l'ensemble de leur nomenclature d'OCS.

L'équipe de [Gaetano et al., 2018], du même laboratoire, a étudié la fusion par DpLg entre les canaux Multi-Spectraux (MS) et PAN. L'architecture, également à double entrée, est composée de deux CNN classiques dont les embeddings sont introduits dans un Multi Layer Perceptron (MLP) avant une fonction Soft-Max. Leurs conclusions corroborent avec le prédicat des embeddings plus discriminants du DpLg puisque les images pansharpenées obtiennent des résultats de SegSem inférieurs.

Si [Gaetano et al., 2018] utilisent des CNN classiques, [Zhang et al., 2019] ont adapté l'architecture "U-NET" de [Ronneberger et al., 2015] pour de la détection de changement d'OCS à partir d'imagerie THR aérienne. Ils ont remplacé les convolutions de l'encodeur par des convolutions dilatées (ou "à trous") et ont ainsi amélioré la performance initiale de [Ronneberger et al., 2015] sur leur jeu de données.

2.1 Bases de données et produits d'OCS existants

Sur le territoire français, un ensemble de BD décrivant l'OCS sont disponibles et seront utilisées pour produire nos Vérité Terrain (VT).

2.1.1 BD Topo®

La BD Topo® [IGN, 2019b] est une des composantes du Référentiel à Grande Échelle (RGE®) de l'IGN. Elle est conçue pour répondre aux besoins d'aménagement du territoire. Sa mise à jour est effectuée par l'IGN au gré des acquisitions aériennes et des remontées d'informations. Concernant l'OCS, elle regroupe les éléments routiers, ferroviaires et hydrographiques, les bâtiments et la végétation. L'exhaustivité de ses classes est, selon sa documentation, à minimum de 95%. Les objets y sont définis avec une précision métrique à décimétrique.

2.1.2 RPG

Le Registre Parcellaire Graphique (RPG) est une BD géographique recensant les parcelles agricoles [IGN, 2019a]. Il est produit annuellement par l'Agence de Service et Paiement (ASP) et l'IGN pour servir de référence à l'instruction des aides de la Politique Agricole Commune (PAC). Une nomenclature de 28 types de cultures décrit l'usage des parcelles.

2.1.3 OCS-GE

L'OCS-GE est une base de données de l'IGN produite à la demande. Elle sépare CS et US avec des nomenclatures respectives. Elle définit une ossature qui partitionne l'espace en éléments de même nature de CS et d'US. Le partitionnement est réalisé à partir des BD existantes puis complété par photo-interprétation et définit l'OCS plus grossièrement que la BD Topo®. Sa mise à jour est prévue tous les 4 ans.

2.1.4 Urban ATLAS

Urban Atlas (UA) est une BD géographique contenant des informations de CS générales et d'US sur les densités d'urbanisation. Cette base de données est mise à jour tous les 6 ans et n'est disponible qu'autour des grandes villes d'Europe. Elle est produite par l'Agence Européenne de l'Environnement et possède deux millésimes.

2.1.5 OSO

La OSO [CESOSO, 2019] est une SegSem par RF de l'OCS sur tout le territoire français. Sa donnée d'entrée étant les séries temporelles S2, elle possède une résolution équivalente de 10m. Elle est produite annuellement depuis 2009 grâce à la chaîne de traitement IOTA2 [Inglada et al., 2017]. La nomenclature du "millésime" 2018 comprend 23 classes. La qualité de la SegSem est synthétisée dans la matrice de confusion disponible sur leur site [CESOSO, 2019]. Les performances de la chaîne IOTA2 montrent des difficultés sur la détection des classes artificialisées. Une grande confusion entre les classes *Batis denses* et *Batis diffus* est flagrante.

2.2 Zones d'intérêt

Dans le cadre de cette étude, nous avons sélectionné trois régions d'intérêt :

- Finistère : Région d'intérêt pour les travaux de [Postadjian et al., 2017].
- Gironde : Idem, concernant le fine tuning et zone test pour le CES Urbain.
- Haute-Garonne : Région d'intérêt pour le projet LANDSENSE [International Institute of Applied Systems Analysis, 2019], le CES Urbain, le projet URCLIM [URCLIM, 2019].

2.2.1 Données

Nos images sont acquises par les constellations de satellites SPOT 6 et 7 et S2 (cf. 1.1.4). La VT est issue des BD présentées ci-dessus (cf. 2.1).

Donnée	SPOT	Sentinel	BD Topo®	RPG	OSO	OCS-GE	UA
Finistère	2014 et 2016	2016			2018		2012
Gironde	2016	2016-2017			2018		2012
Haute-Garonne	2018	2018			2018		2012

Table 2.1 – Datation des données

2.2.2 Pré-traitements

Les jeux de données ont tous subi des traitements antérieurs à notre chaîne de traitement. Ces opérations ont été réalisées par les producteurs des données (ex. : corrections radiométriques et géométriques) ou par nos soins (Pansharpening et sur-échantillonnage des images SPOT). L'ensemble des images utilisées dans le cadre du projet sont distribuées par le Pôle Théia.

2.2.2.1 Images

Concernant la donnée SPOT, 4 images sont produites.

- P : Pas de traitement particulier, utilisation de l'image directe,
- XS native : Concaténation des images MS à leur résolution native de 6m,
- XS : Concaténation des images MS après sur-échantillonnage bi-linéaire à la résolution du PAN,
- PXS : Pansharpening des images MS à partir de l'image PAN.

Les images S2 sont des synthèses mensuelles *cloud-free*, par bande, traitées par gap-filling. Nous créons 2 images à partir de ces données :

- IRGB : Concaténation des bandes B2,B3,B4 et B5 sur-échantillonnées à 1.5m,
- S2full : Concaténation de toutes les bandes sur-échantillonnées à 1.5m, excepté celles de résolution 60m (B1,B9 et B10).

En résumé nous disposons d'un jeu de données composé des images suivantes :

Nom	P	XS native	XS	PXS	IRGB	S2full
Résolution (m)	1.5	6	1.5	1.5	1.5	1.5
Canaux	PAN	R,V,B,PIR	R,V,B,PIR	R,V,B,PIR	R,V,B,PIR	voir 1.1
Nb de canaux	1	4	4	4	4	4

Table 2.2 – Récapitulatif des images pré-traitées

2.2.2.2 Vérité terrain

Les BD d'OCS contiennent régulièrement des mélanges d'objets d'CS et US. La création de cartes d'CS nécessite d'extraire uniquement les thèmes les plus intéressants au regard des nomenclatures sélectionnées (cf. 2.3 et au regard de leur compatibilité avec la télédétection).

Les pré-traitements que nous avons effectués consistaient à :

- Convertir les objets à tendance linéaire en objets surfaciques. Notre approche de rastérisation des objets (cf. 3.3.1.2) nous impose de disposer d'objets surfaciques. Or certains thèmes des BD sont stockés sous forme d'objets linéaires. Par exemple, les routes de la BD Topo.
- Éclater les thèmes trop vastes pour le raffinement des nomenclatures. En particulier le thème *végétation* de de la BD Topo).
- Créer de nouvelles classes par fusion, intersection ou dilatation de plusieurs groupes d'objets. En particulier la classe **Route** de la nomenclature A5 qui contient une fusion des routes dilatées et des parkings,
- Assembler des objets issus de différentes BD.

L'ensemble des traitements ont été réalisés à l'aide du logiciel QGIS [QGIS, 2019].

Nous avons considéré la SegSem comme étant mono classe, c'est à dire qu'à chaque pixel ne peut être attribué qu'une seule classe. Ainsi, dans les cas où des objets se superposeraient entre les différentes bases, un ordre de priorité définit quelle classe est conservée (cf. 3.3.1.2). Nos nomenclatures intègrent donc un ordre de priorité déterminé lors du pré-traitement de la VT.

2.3 Nomenclatures

Les nomenclatures présentées dans cette section ont été adaptées aux données disponibles, aux zones d'étude, aux capteurs et aux différentes études réalisées. Le tableau 2.3 lie les nomenclatures à leurs cas d'usages.

Nomenclature	Cas d'usage
A5	Comparaison d'algorithmes de SegSem
A6	Recherche de précision dans la délimitation des instances de bâtiments
U6	Recherche de précision dans la délimitation de la végétation urbaine
H13	Étude du 1 ^{er} niveau de raffinement des nomenclatures
H18	Étude du 2 ^{ème} niveau de raffinement des nomenclatures
OSO	Étude de la capacité de génération de carte d'CS à partir de VT OSO et d'image SPOT
OCS-GE LU	Étude la capacité de génération d'un produit complet
OCS-GE LC	Étude la capacité de génération d'un produit complet
D	Étude de la capacité de génération de cartes de densité urbaine UA
HM	Étude de la capacité de génération de cartes de densité urbaine UA
MSK	Étude de la capacité de génération de cartes de densité urbaine UA

Table 2.3 – Cas d'usages des nomenclatures

A : *All* toutes zones, U : Urbain, F : Finistère, H : Haute-Garonne, D : Densité, HM : *Heat Map* ou carte de chaleur, MSK : Masques. Les chiffres correspondent au nombre de classes dans la nomenclature.

2.3.1 Couverture du sol

Les nomenclatures dites "classiques" présentées dans le tableau 2.4 sont celles que nous avons constituées en regroupant et modifiant les thèmes d'objets des BD disponibles selon la méthodologie avancée précédemment.

2.3.1.1 Nomenclatures classiques

Les nomenclatures classiques présentées dans le tableau 2.4 ont été principalement élaborées pour tirer au mieux profit de la THR des images SPOT. D'autre part nous avons effectué un raffinement progressif en séparant certains thèmes d'objets larges et en y ajoutant les thèmes manquants. L'objectif étant de valider la performance des CNN sur différents niveaux de détails.

La classe *Cultures*, qui est pourtant un thème large n'a pas été dilatée en raison de nos connaissances sur les difficultés de séparation des types de cultures uniquement à partir des 4 bandes spectrales de et d'une seule date SPOT, non contemporaine du RPG disponible. Les séries temporelles S2 y sont bien plus adaptées.

Précisions sur les nomenclatures classiques :

A5	A6	U6	H13	H18
-	● Tampon bâtiment	-	● Tampon bâtiment	● Tampon bâtiment
● Eau	● Eau	● Eau	● Eau	● Eau
-	-	-	-	-
● Cultures	● Cultures	-	● Cultures	● Cultures
-	-	● Sol nu	-	-
● Routes	● Routes	● Routes	● Routes	● Chemins
			● Pistes d'aérodromes	● Routes principales
			● Parkings	● Pistes d'aérodromes
			● Cimetières	● Parkings
-	-	-	● Cimetières	● Cimetières
● Végétation	● Végétation	● Végétation haute	● Feuillus	● Feuillus
		● Végétation basse	● Conifères	● Conifères
		-	● Landes	● Landes
		-	● Haies	● Haies
		-	-	● Forêts ouvertes
		-	-	● Vignes
-	-	-	● Vergers	
-	-	-	● Peupleraies	
● Bâtiments	● Bâtiments	● Bâtiments	● Résidentiel	● Résidentiel
			● Industriel	● Industriel

- : classe non utilisée. L'ordre de priorité défini par ordre croissant de haut en bas.

Table 2.4 – Raffinement progressif des Nomenclatures classiques

- A5 : Nomenclature simple, similaire à celle utilisée par [Postadjian et al., 2017], qui décrit les objets topographiques les plus classiques et les plus courants.
- A6 : Nomenclature A5 avec addition de contrainte sur le détournement des bâtiments grâce à la classe *Tamppons de Bâtiments*. C'est une classe que nous avons créée à partir d'une dilatation arbitraire des objets bâtiments de la BD Topo®. Elle a pour objectif de mieux contraindre la détection des bâtiments et fait suite aux travaux d'identification de la tâche urbaine de [Wendl et al., 2018]
- U6 : Nomenclature pour la végétation urbaine, dont la VT est réalisée à partir d'une classification RF
- H13 : Séparation de la classe "Végétation" en ses principales sous-classes et ajout des objets *Cimetières* qui possèdent une texture particulière.
- H18 : Séparation de la classe "Végétation" en toutes ses sous-classes hormis "Bois", "Landes herbacées", "Forêts mixtes". La classe "Routes principales" contient les routes étant définies dans la BD Topo® comme "Autoroutes", "Chaussées à 1 voie", "Chaussées à 2 voies", "Échangeurs" et "Ronds points". La classe *Chemin* contient donc tous les autres éléments de la BD Topo, qui peuvent être très variables physiquement.

Le retrait des classes "Bois" et "Forêts mixtes" est dû à leur définition, trop proche des classes "Feuillus" et "Conifères". Leurs définitions étant heuristiques, il est possible de les recalculer à partir du résultat de l'inférence.

2.3.1.2 Nomenclatures existantes

Les produits OSO et OCS-GE disposent de leurs nomenclatures de CS. Elles sont disponibles en annexes.

- La nomenclature OSO a été adaptée comme un compromis entre besoins utilisateurs et performances spectrales des satellites S2. Elle est extrêmement raffinée sur les types de cultures et de végétation. Sa version 2018 contient 23 classes dont 9 sont attribuées aux cultures, et 6 aux différents types de végétation.
- La nomenclature OCS-GE de CS est elle aussi très axée sur la végétation avec 7 de ses 14 classes, mais ne se concentre pas autant sur les cultures et n'est pas spécifiquement élaborée pour un capteur.

2.3.2 Usage du sol

Le produit OCS-GE dispose également d'une nomenclature d'US composée de 16 classes, dont certaines sont très peu adaptées à la télédétection. La nomenclature est disponible en annexe A.

2.3.3 Mixtes

2.3.3.1 Nomenclature UA

Le produit UA définit une nomenclature mixte US et CS. Sa nomenclature de CS se rapproche de celle du produit OCS-GE avec des distinctions entre type de sols et de végétation. Son découpage de la tâche urbaine suit l'US avec un partitionnement par densité d'urbanisation.

Nous avons utilisé ces thèmes pour élaborer une série de nomenclatures mixtes concernant l'urbanisation. Elles utilisent les mêmes définitions que la nomenclature A5.

D	HM	MSK
● Eau	● Eau	● Eau
● Cultures	● Cultures	● Cultures
● Routes	● Routes	● Routes
● Végétation	● Végétation	● Végétation
Densité	Densité	Masques
-	● Bâtiments résidentiels	-
-	● Bâtiments industriels	-

Table 2.5 – Nomenclatures mixtes sur la densité d'urbanisation

- : classe non utilisée. L'ordre de priorité défini par ordre croissant de haut en bas.

Précision sur les nomenclatures de densités :

- Densité : Regroupe les classes de densité d'urbanisation de la base UA (codes 11100, 11210, 11220, 11230, 12240, 11300 et 12100). Ces classes ne superposant pas entre-elles, elles ne nécessitent par d'ordre de priorité.
- *Heat Map* : Equivalente à la nomenclature *D*, avec ajout des bâtiments résidentiels et industriels de la BD Topo.
- Masques : Regroupe les classes créées par intersection entre les classes de densités d'urbanisation de la base UA et les bâtiments de la BD Topo. Revient à attribuer une classe de densité à chaque bâtiment.

Ce chapitre a pour objectif de présenter la chaîne de traitement de *DpLg* mise en place pour la création de cartes d'OCS. Il permettra également de justifier les choix de conception de l'architecture du réseau de neurones.

3.1 Réseau de neurones

Le réseau [Krizhevsky et al., 2012], vainqueur du challenge ILSVRC 2012 réalise des opérations de classification. Des méthodes dites "par fenêtre glissante" ont permis d'accomplir des tâches de SegSem à partir de classification. Elles ont été utilisées dans les travaux de [Girshick et al., 2013]. [Postadjian et al., 2017] applique une même approche et attribue le résultat de la classification d'une dalle de 65 par 65 pixels à son pixel central.

Comme expliqué dans [Ronneberger et al., 2015], les deux principales limitations des SegSem par fenêtre glissante sont :

- La vitesse de calcul limitée, notamment à l'apprentissage, puisqu'une inférence de $65*65$ pixels est requise pour chaque pixel.
- Le compromis entre information contextuelle et localisation. D'un côté plus la dalle est grande, plus le contexte aidera à améliorer la précision de la classification. De l'autre, elle nécessitera plus d'opérations de synthétisation (*Max-pooling*), et donc réduira la précision de la localisation.

Si l'impact du premier inconvénient a pu être réduit à l'inférence grâce au partitionnement en superpixels dans [Postadjian et al., 2018b], l'information spatiale contenue dans la THR de SPOT6/7 reste réduite par les *Max-pooling*.

[Ronneberger et al., 2015] ont proposé une architecture axée sur la conservation de l'information spatiale par connexions résiduelles (fig 3.1), en s'inspirant de [He et al., 2015a], vainqueurs du challenge ILSVRC 2015. Cette architecture est basée sur un système encodeur / décodeur où la sortie de chaque étage de l'encodeur est reliée par une connexion résiduelle à l'étage équivalent du décodeur. L'idée étant de conserver les informations bas-niveau (avant l'encodeur) disponibles lors du décodage. Le réseau définira seul les poids à attribuer aux caractéristiques bas-niveau et haut-niveau. L'architecture dite U-NET ainsi proposée est composée de 5 étages de 2 convolutions, séparés par des *up-conv*, qui sont des convolutions qui permettent de doubler la taille des images.

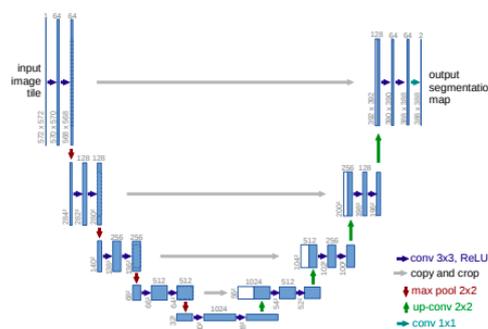


Figure 3.1 – Architecture U-NET, [Ronneberger et al., 2015]

Dans un même temps [Noh et al., 2015] ont proposé une architecture "Deconv-NET" basée sur les opérations de dé-convolution (ou convolution transposée) pour des tâches de segmentation sémantiques. La particularité supplémentaire est de conserver la position des maximaux lors des opérations de *Max-pooling* de l'encodeur pour les ré-appliquer lors des opérations d'*Unpooling*¹ Pour le détail du réseau Deconv-NET, des opérations d'*Unpooling* et de déconvolution voir le l'article de blog [Sik-Ho Tsang, 2018]. Cette technique permet de mieux conserver les informations spatiales le long des calculs.

¹.

3.2 Architecture proposée

Nous avons décidé d'implémenter une architecture inspirée du U-NET et du Deconv-NET, appelée "SO-NET" (fig 3.2. Les principales différences sont :

- L'utilisation de marges créées à la volée avant chaque convolution, pour conserver la taille de l'image d'entrée au cours de l'inférence.
- L'utilisation d'un décodeur équivalent à celui du Deconv-NET.

Au cours des premières expérimentations, les métriques de qualification de l'entraînement indiquaient que le réseau avait une forte tendance au sur-apprentissage. Cet effet apparaît lorsque le réseau dispose de trop de paramètres par rapport à la complexité de la tâche de classification (ou de SegSem). Nous avons donc décidé de supprimer les 2 derniers étages.

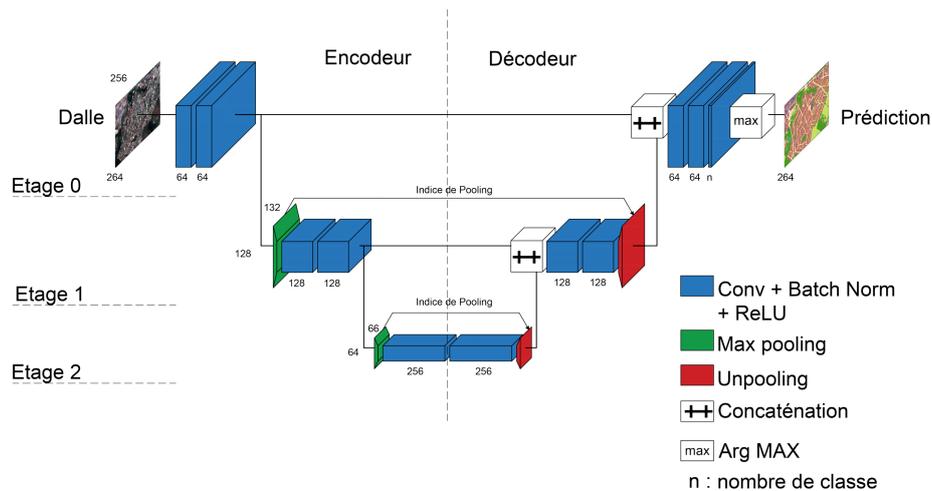


Figure 3.2 – Architecture SO-NET proposée

3.3 Chaîne de traitement

Dans le but d'entraîner et de prédire grâce à l'architecture SO-NET, nous avons développé une chaîne de traitement complète. Elle est composée de 4 processus principaux :

- **taylor** : Préparation des images satellites et de la VT
- **main** : Entraînement d'un modèle basé sur l'architecture SO-NET
- **drawer** : inférence à partir d'un modèle
- **stitcher** : Post-traitement des résultats d'inférence.

Ces processus sont des scripts *Python*. Ils sont reliés par des scripts *Shell (Bash)* qui facilitent la sélection de leurs arguments d'entrée respectifs.

L'implantation de la partie DpLg est basée sur la bibliothèque *PyTorch* et ses calculs sont effectués sur Carte graphique ou *Graphic Processing Unit (GPU)*.

3.3.1 Préparation des données d'entrée

Pour entraîner un modèle de DpLg, deux éléments sont requis :

- la donnée à traiter,
- la VT associée.

Si les pré-traitements (cf. 2.2.2) nous fournissent une donnée exploitable, il reste nécessaire de les conditionner aux formats imposés par la bibliothèque *PyTorch* [PyTorch, 2019] et aux capacités informatiques disponibles.

3.3.1.1 Images

La donnée à prédire est tout ou partie d'une image satellite. Les calculs étant effectués sur carte graphique, l'image doit pouvoir être chargée dessus. Néanmoins, avec leur taille dépassant aisément les 30Go, il est inconcevable de travailler directement avec une image complète. Elles sont ainsi segmentées en dalles.

Nous avons défini la taille de ces dalles en fonction de l'espace mémoire GPU disponible, du nombre de paramètres de notre réseau, de la taille des *batches*² (cf. 3.1) et du nombre de canaux des images.

Les dalles doivent être les plus grandes possible pour optimiser la lecture du contexte. D'un autre côté,

taille des dalles	taille des <i>batches</i>
1024*1024	1
512*512	4
256*256	16
128*128	64

Table 3.1 – Taille des *batches* en fonction de la taille des dalles

la taille des *batches* doit être la plus grande possible pour accélérer l'apprentissage. En effet, c'est en estimant l'erreur (appelée Loss) entre les prédictions des dalles du *batch* et leur VT que l'algorithme de Descente de Gradient Stochastique (SGD) est capable de réactualiser les poids des connexions [de manière à améliorer la prochaine prédiction]. La direction prise par le SGD est donc directement dépendante des données contenues dans ce *batch*. Il est impératif que les *batches* soient une représentation statistique de l'image globale pour assurer une direction optimale. Le nombre de pixels qu'il est possible de charger sur la carte graphique est constant, c'est donc un équilibre à choisir entre vitesse d'apprentissage et précision de la SegSem.

Remarque : Nous avons opté pour des *batches* de 16 dalles de 264*256 pixels. L'asymétrie permet de suivre l'orientation des images au cours des opérations de la chaîne de traitement à des fins de débogage, puisque *GDAL*, *Numpy* et *Pytorch* n'ont pas les mêmes normes de description des tableaux.

Normalisation des canaux L'apprentissage des réseaux neuronaux est permis par la rétro-propagation des gradients. Pour qu'elle soit réalisée efficacement il est important d'avoir des gradients similaires dans tout le réseau. C'est précisément l'objectif de la normalisation. Ainsi, pour s'assurer que dès la première couche de convolution les poids des connexions soient d'ordre de grandeur comparables, nous normalisons les valeurs des dalles pour chaque canal, selon l'intervalle [2%-98%] de son histogramme. Nous avons remarqué que les éléments comme les forêts de résineux, très sombres, perdent leur texture lors du bornage. Nous avons donc décidé de ne pas forcer ses valeurs. La normalisation doit être la même pour l'ensemble des dalles, nous déduisons les bornes de normalisation de l'image globale, ou depuis un texte, selon les besoins de l'utilisateur.

Nous effectuons, pour chaque canal, l'opération :

$$pix_{i,j,k} = \frac{pix_{i,j,k} - q_{k,2}}{q_{k,98} - q_{k,2}} \quad (3.1)$$

où i, j sont les coordonnées du pixel et k son canal.

3.3.1.2 Vérité terrain

La VT peut se présenter sous deux formes :

- Un ensemble de shapefiles, où chaque fichier correspond à une classe,
- Une image de valeurs entières, où chaque valeur correspond à une classe.

Dans les deux cas, l'objectif est de générer une image de vérité terrain de même emprise et de même résolution que l'image satellite d'entrée. Elle est ensuite découpée en dalles correspondant aux dalles de prédiction. À la fin des traitements, nous disposons de couples de dalles VT et satellite : ce jeu de données sera nommé jeu de données global ou *Global dataSet* (GS) dans la suite de ce document.

²*Batches* est un terme utilisé en ML pour désigner les exemples d'apprentissage utilisés dans une itération

Cas des shapefiles Lors du calcul de la Loss, la VT doit être présentée sous forme matricielle pour être comparée à la prédiction. Nous avons choisi de convertir les shapefiles en images puisqu'elles sont facilement convertibles vers une forme matricielle.

Lors de la création de l'image de VT, chaque shapefile est rasterisé selon l'emprise et la résolution de l'image satellite. Les rasters générés sont ajoutés, un à un, à l'image de VT, écrasant les valeurs déjà attribuées. La priorité des classes est donc induite par l'ordre d'ajout des shapefiles.

Cas des images Dans le cas où la VT est déjà sous forme raster, elle doit valider un ensemble de critères :

- Être accompagnée d'une nomenclature qui associe chaque valeur à un nom de classe
- Avoir 99 comme valeur de "noData"
- Commencer la numérotation des classes à 0

L'image est ensuite transformée pour correspondre à l'emprise et à la résolution de l'image de prédiction.

3.3.2 Validation croisée

Nous avons prévu une validation croisée sur le GS. L'ensemble des couples de dalles ainsi séparé en 3 jeux de données. A des fins de rapidité de temps de calcul, nous n'utilisons que la première itération de la validation croisée. **Remarque** : Nous avons décidé de répartir aléatoirement les dalles dans les 3 jeux

Nom du jeu de données	% du GS	objectif
jeu de données d'enTRaînement (TR)	64	Apprentissage du modèle
jeu de données de TEst (TE)	16	Sélection de la meilleure époque d'entraînement
jeu de données VALidation (VAL)	20	Qualification de la performance du modèle

Table 3.2 – Partitionnement du GS pour la validation croisée

de données pour conserver la répartition statistique des classes du GS.

Ce choix rend difficile la visualisation d'une large zone de test. Pour remédier à ce problème, nous avons ajouté la possibilité de créer des shapefiles d'indexation des dalles. Ils permettent de sélectionner des dalles qui seront "verrouillées" dans les différents jeux de données. Ces dalles sont alors plus considérées comme faisant partie du GS.

3.3.3 Entraînement du réseau

3.3.3.1 Initialisation des poids des connexions

[James Dellinger, 2019] recommande, pour les CNN, l'initialisation dite "Kaiming" proposée par [He et al., 2015b]. Pour l'ensemble de nos tests nous utilisons cependant l'initialisation "Xavier" proposée par [Glorot and Bengio, 2010]. Notre architecture étant peu profonde, nous n'avons pas de risque de disparition du gradient. La seule différence est un apprentissage légèrement ralenti.

3.3.3.2 Représentativité des classes

La proportion de présence entre les classes étant très variable (jusqu'à 50% de pixels *Cultures* contre 0.56% de pixels de parkings sur la région de Toulouse), il est essentiel de pondérer l'impact des classes sur le calcul de la Loss. le poids $w[class]$ associé à une classe est fonction de son pourcentage p de pixel sur l'ensemble de pixels annotés (éq. 3.2).

$$w[class] = \begin{cases} 0 & \text{si } p[class] = 0, \\ \frac{1}{p[class] \sum_{i=1}^n p[i]^2} & \text{sinon.} \end{cases} \quad (3.2)$$

où n est le nombre de classes.

3.3.3.3 Fonction Loss

La fonction de Loss permet de calculer l'erreur entre la prédiction et la VT. Nous utilisons la fonction d'entropie croisée pondérée (éq. 3.3). Cette fonction d'erreur est similaire à celle utilisée par [Ronneberger et al., 2015].

$$loss(x, class) = w[class] \left(-x[class] + \log \left(\sum_j \exp(x[j]) \right) \right) \quad (3.3)$$

où x_i est le tenseur de prédiction et w_i dérive de l'équation 3.2. Les pixels sans vérité terrain sont ignorés lors du calcul de la loss. La Loss globale est une moyenne de l'ensemble des loss du *batch*.

$$Loss = \frac{1}{n} \sum_{i=1}^n loss[i] \quad (3.4)$$

où n est la taille du *batch*

3.3.3.4 Optimiseur

L'optimiseur a pour objectif de trouver un paramétrage des poids optimisant la *Loss*, ce qui implique :

- 1 : Calculer le gradient de chaque paramètre en fonction de la Loss
- 2 : Actualiser l'ensemble des poids selon leur gradient

Une comparaison des différents algorithmes d'optimisation par gradient adapté au DpLg a été menée par [Ruder, 2016]. En conclusion, bien que l'algorithme classique de SGD converge vers les meilleurs minimums, l'algorithme "Adam" développé par [Kingma and Ba, 2014] est désigné comme le meilleur compromis entre vitesse d'apprentissage et précision de la convergence. Pour optimiser nos entraînements, nous avons sélectionné l'algorithme Adam. Les hyper-paramètres associés sont positionnés à leur valeur par défaut.

3.3.3.5 Hypers paramètres

Les pas de gradient de l'optimiseur sont définis par le pas d'apprentissage, appelé "learning rate". Avec l'utilisation d'un algorithme d'optimiseur Adam, la documentation PyTorch [PyTorch, 2019] recommande un learning rate de 10^{-3} . Pour accélérer l'apprentissage et être plus robuste aux minimums locaux, nous avons mis en place un learning rate décroissant. Il débute à 10^{-2} et se réduit d'un facteur multiplicatif de 0.7 toutes les 50 époques.

Nous n'utilisons pas de *weight decay* en raison de son incompatibilité avec la normalisation par *batch* [Ioffe and Szegedy, 2015].

3.3.3.6 Données de sortie

L'algorithme "main" enregistre les modèles entraînés et leurs métriques d'apprentissage au cours des époques.

Sauvegarde des modèles : Nous enregistrons deux modèles. Le premier à chaque époque et le deuxième lorsque la métrique moyenne des intersections sur l'union ou *mean Intersection over Union* (mIoU) du VAL enregistre un nouveau record.

métriques Les métriques enregistrées au cours de l'entraînement des modèles sont les suivantes :

Métrique	périodicité des époques d'acquisition
IoU par classe	<i>mem</i>
mIoU	<i>mem</i>
OA	<i>mem</i>
Loss	1

Table 3.3 – Métriques d'entraînement et leur périodicité d'acquisition

mem est un paramètre de l'algorithme "main" qui contrôle la fréquence du calcul des métriques. intersection sur l'union ou *Intersection over Union* (IoU); précision globale ou *Overall Accuracy* (OA).

3.3.4 Inférence

Les algorithmes d'inférence que nous avons développés nous permettent de générer, à partir d'un modèle, les données suivantes :

- Cartes de SegSem
- Cartes de probabilités par classe
- Logits³ au format H5
- Cartes de différence entre la VTet la SegSem
- Métriques d'inférence : Matrice de Confusion (CM), OA⁴, mIoU et les IoU, F-score, précision⁵ et rappel⁶ par classe.

Les cartes de différence et les métriques ne peuvent être générées que lorsqu'une VT est fournie avec les dalles d'inférence. Dans le cas où une validation croisée a été effectuée, il est possible de lier un fichier de répartition des dalles afin d'exclure celles ayant servi à l'entraînement entre les 3 jeux de données, dans le but d'obtenir des métriques indépendantes. Les cartes sont créées par fusion des dalles de prédiction.

3.4 Estimation de l'aléatoire

Notre chaîne de traitement dispose des opérations aléatoires suivantes :

- Répartition des dalles dans la validation croisée,
- Initialisation des poids CNN,
- Réparation des dalles dans les batches de chaque époque,
- Calculs GPU.

Nous avons décidé de fixer la part aléatoire de la répartition des dalles dans la validation croisée par une graine. Elle a été réutilisée pour l'ensemble des calculs et des tests réalisés dans cette étude. Ce choix a été motivé par une volonté de consistance de l'emplacement des dalles pour faciliter la visualisation. L'initialisation des poids et la répartition des dalles dans les batches n'ont pas été soumises à une graine aléatoire puisqu'ils dépendent d'hyper-paramètres variables selon les tests. Dans le cas de l'analyse du réseau, certaines architectures demandent un volume de mémoire GPU plus important, obligeant la réduction de la taille des batches. Elle varie également lorsque les images disposent de multiples canaux, comme les images S2. Concernant les calculs GPU, la bibliothèque [PyTorch, 2019] ne garantit pas la reproductibilité des calculs sur GPU. Si l'environnement de calcul GPU CuDNN propose une option de calcul déterministe, il est précisé qu'elle réduit violemment la rapidité de traitement. Nous avons préféré la vitesse de calcul à la reproductibilité.

Pour que nos analyses comparatives soient sensées et robustes, il nous est obligatoire d'estimer la part d'aléatoire dans les métriques finales. Néanmoins, dans une intention de gain de temps, nous avons décidé de ne pas calculer la part d'aléatoire sur tous les tests.

Seul un test bénéficie d'une estimation précise de l'incertitude liée à notre chaîne de traitement. Ces incertitudes seront extrapolées aux autres réalisations.

3.4.1 Aléatoire de la répartition dans la validation croisée

La totalité des tests étant réalisée avec la répartition des dalles donnée par la première itération de la validation croisée (CV1), nous en avons déterminé sa variabilité. Les 12 entraînements ont été calculés sur la zone de Toulouse, selon la nomenclature A6.

Le tableau 3.4 nous permet de conclure que les métriques globales de mIoU et d'OA sont les plus stables. Elles sont donc des indicateurs robustes pour nos tests et seront nos principaux outils de comparaison. Suivant l'analyse par classe, c'est l'IoU qui est la moins sensible aux variations aléatoires de notre chaîne de traitement.

Les classes bâtiment, route, et buffer sont des classes radiométriquement proches, avec beaucoup de frontières communes, que la résolution de l'image SPOT rendent difficile à séparer. Ces classes sont aussi de

³Logits : Prédiction non normalisées d'un modèle

⁴OA : Mesure de précision globale, qui est le ratio entre Vrais positifs sur le total d'élément

⁵Précision : Mesure de l'exactitude de la prédiction, qui est le ratio de Vrais positifs sur la somme des Vrais positifs et des Faux positifs

⁶Précision : Mesure de l'exhaustivité de la prédiction, qui est le ratio de Vrais positifs sur la somme des Vrais positifs et des Faux négatifs

Métrique	mIoU			OA		
μ	78,33			91,61		
σ	0,163			0,070		
Classe	● Buffer	● Eau	● Cultures	● Routes	● Végétation	● Bâtiments
Métrique	IoU					
μ	64,90	86,77	94,55	72,46	85,41	65,88
σ	0,404	0,365	0,076	0,575	0,176	0,444
Métrique	F-score					
μ	79,26	92,87	97,23	84,63	91,89	78,92
σ	1,338	0,214	0,094	1,138	0,627	1,544
Métrique	Précision					
μ	81,72	92,81	97,42	86,33	90,54	77,75
σ	1,098	0,494	0,327	1,120	0,577	1,577
Métrique	Rappel					
μ	75,94	93,03	96,98	81,90	93,78	81,24
σ	0,897	0,830	0,340	1,602	0,578	1,125

Table 3.4 – Moyennes (μ) et Écarts types (σ) des métriques de la CV1 selon la nomenclature A6 pour différents entraînements sur les mêmes données

petits objets, d'où l'augmentation de la sensibilité au bruit et des valeurs d'écart-type plus importantes. La classe Eau présente aussi une variabilité forte. La sur-détection de l'eau dans les zones d'ombres (non apprises ici) des milieux urbains denses est un phénomène connu. Le papier [Yao et al., 2015] en présente les principales difficultés.

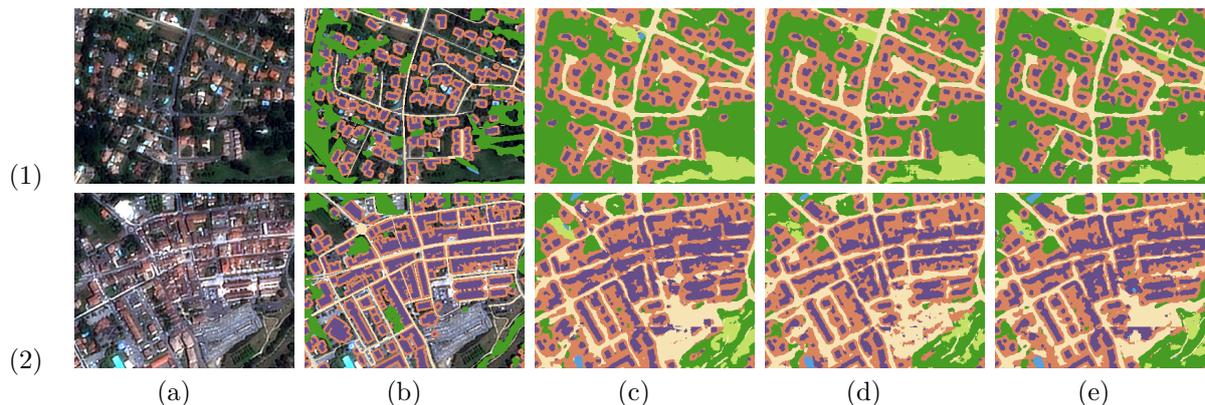


Figure 3.3 – Comparatifs visuels entre les inférences de la CV1

(1) Zone pavillonnaire, (2) Centre-ville, (a) image SPOT, (b) VT, (c) (d) (e) 3 itérations de la CV1

La variabilité de la SegSem dépend également des zones de l'image. La figure 3.3 permet de visualiser que la SegSem en zone de densité urbaine faible est plus stable entre les itérations qu'en zone dense. La délimitation des bâtiments par la classe buffer en zone dense est plus difficile et plus instable. L'îlot central de bâtiments de l'image est bien délimité dans l'image (2)(c) tandis que les deux itérations similaires n'ont pas achevé un résultat aussi efficace. Concernant les routes, le même effet est notable. Ces résultats visuels corroborent la variabilité des métriques.

Le cimetière situé en bas à droite des images du centre-ville nous permet de souligner la grande instabilité de la SegSem pour les éléments de CS faisant défaut dans la VT utilisée. L'indécision du réseau est également notable par les effets de marges exacerbés.

Étant donné que nos itérations sont indépendantes et que les tirages aléatoires suivent une loi normale, nous pouvons considérer la distribution observée de forme gaussienne ou, au moins, symétrique et unimodale. Cela nous permet de définir les intervalles de confiance qui garantissent le lien à une réelle augmentation ou diminution de performance. Une métrique supérieure de 2 écarts types à sa moyenne aura au moins de 95.5% de chance de ne pas être issue de l'aléatoire de notre chaîne de traitement. Cette

probabilité monte à 99.7% pour un Intervalle de Confiance à 3 écarts-types (IC-99).

	IC-95 (2σ)						IC-99 (3σ)					
mIoU	0,464						0,928					
OA	0,181						0,363					
	● Buf.	● Eau	● Cult.	● Rts	● Vég.	● Bât.	● Buf.	● Eau	● Cult.	● Rts	● Vég.	● Bât.
IoU	1,220	1,533	0,142	1,181	1,033	0,972	2,441	3,067	0,283	2,361	2,067	1,944
F-score	2,368	0,858	0,148	1,940	1,020	2,558	4,735	1,717	0,297	3,880	2,041	5,115
Précision	4,232	1,144	0,600	3,780	0,992	4,407	8,464	2,288	1,200	7,560	1,984	8,815
Rappel	2,674	1,964	0,612	4,635	1,645	3,454	5,349	3,928	1,225	9,269	3,289	6,909

Table 3.5 – Intervalles de confiances des métriques(%)

3.5 Analyse de l'architecture

Les outils de comparaison construits dans l'analyse de l'aléatoire de notre chaîne de traitement nous permettent d'étudier l'architecture de notre CNN. Notre objectif est de déterminer quels sont les éléments essentiels pour la SegSem à partir de données THR. Le champ réceptif, la circulation de l'information de localisation et le degré de complexité des attributs sont les trois points sur lesquels nous avons concentré l'étude. Pour chacun de ces points, nous avons étudié les composants et hyper-paramètres les plus influents, listés ci-dessous.

- Considération du contexte :
 - Nombre de convolutions
 - Tailles des noyaux de convolution
 - Gestion des bords
 - Normalisation
- Propagation de l'information spatiale :
 - Connexions résiduelles
 - Décodeur convolutif
- Niveaux de complexité des représentations :
 - Profondeur de l'architecture

3.5.1 Données des tests

Nous avons réalisé les tests de cette étude sur l'image SPOT du Finistère. Le tableau 3.6 récapitule la composition de la VT, basée sur la nomenclature A5, de cette région. Nous avons utilisé la CV1 pour tous les calculs.

Classe	Nombre de pixels	% labellisé	% total
Non Labelisé	93 880 624	0.00%	36.67%
● Eau	3 448 938	2.13%	1.35%
● Cultures	103 667 963	63.95%	40.5%
● Routes	5 498 081	3.39%	2.15%
● Végétation	42 223 656	26.04%	16.49%
● Bâtiments	7 280 738	4.49%	2.84%

Table 3.6 – Composition de la VT pour l'étude de l'architecture

La répartition entre TR, VAL et TE suit la répartition indiquée sur le tableau 3.2.

3.5.2 Convolution

Les opérations de convolutions permettent au réseau d'apprendre des attributs de texture et de contexte. Ces attributs dépendent du champ réceptif. Chaque convolution possède un champ réceptif natif correspondant à la taille de son noyau de convolution. La répétition des opérations de convolution permet d'augmenter le champ réceptif total (deux opérations de convolution à noyau $3 * 3$ auront un champ réceptif équivalent à une convolution de noyau $5 * 5$).

L'objectif est de déterminer le champ réceptif idéal pour le traitement de données satellite THR.

Nombre de convolutions	Champs réceptif	mIoU	OA	IoU				
				Routes	Cultures	Eau	Végétation	Bâtiments
1	9	82,07	92,82	68,09	91,72	90,42	78,15	81,98
2	25	83,86	93,34	71,66	92,16	91,99	79,27	84,24
3	49	84,01	93,54	72,04	92,53	91,30	79,50	84,65

Table 3.7 – Métriques selon le nombre de convolutions par étage (%)

3.5.2.1 Nombre de convolutions

Le mIoU nous permet de conclure qu'une seule convolution ne permet pas une analyse assez globale du contexte. On remarque une nette augmentation pour 2 convolutions. Si l'ajout d'une troisième convolution semble encore faire augmenter le score global, le ratio d'augmentation du champ réceptif n'est plus aussi influent. De plus, la différence de mIoU étant inférieure à un écart type, il n'est pas possible de confirmer son importance.

La classe *Cultures* est la seule classe qui montre une progression d'IoU supérieure à son IC-99.

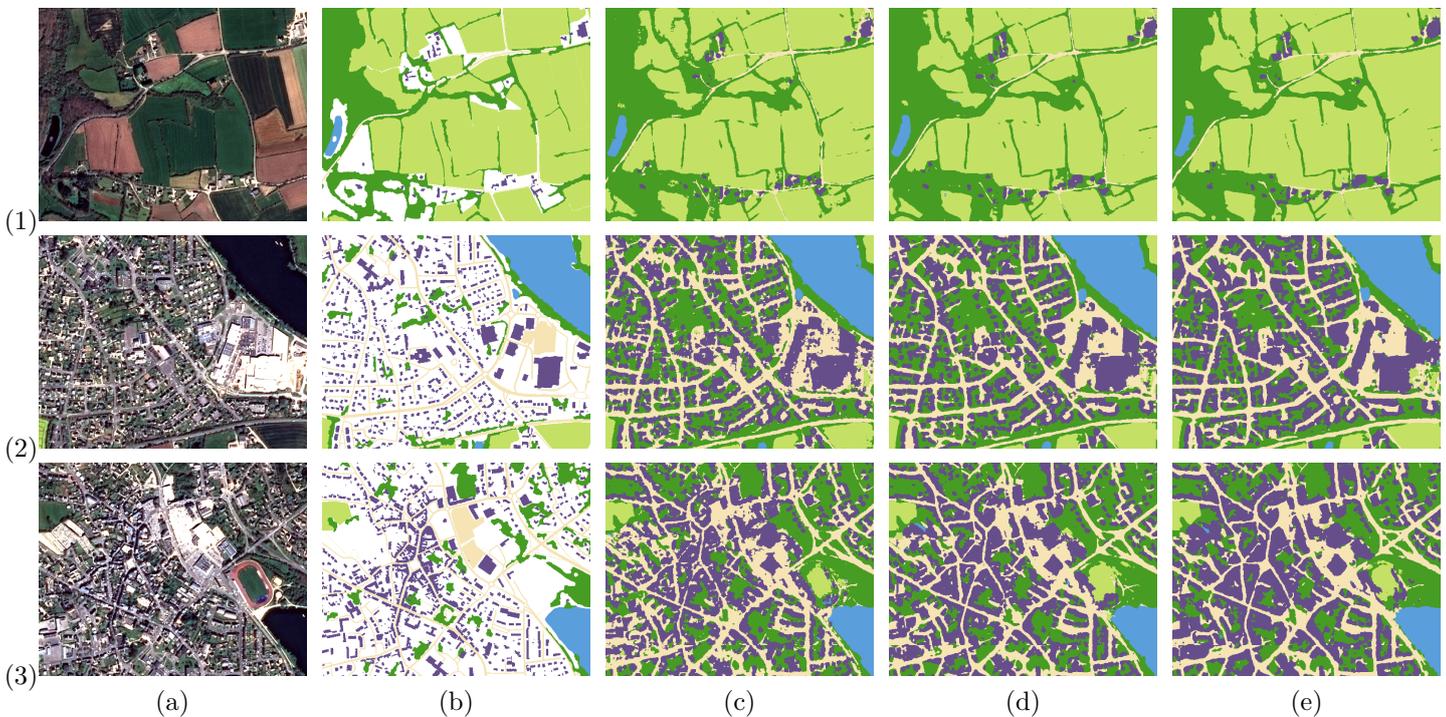


Figure 3.4 – Comparatifs visuels selon le nombre de convolutions par étage
 (1) Zone de faible densité, (2) Zone pavillonnaire, (3) Centre ville, (a) image SPOT, (b) VT, (c) 1 convolution par étage, (d) 2, (e) 3

L'analyse visuelle de la figure 3.4 permet de comprendre pourquoi les classes *Routes*, *Bâtiment* et *Eau* sont les plus impactées. Le faible champ réceptif empêche la distinction entre les parkings et les grands bâtiments. La détection des petits bâtiments de la zone (2) n'est pas affectée. Concernant l'eau, on remarque moins d'erreurs sur les ombres des bâtiments entre (c)(3) et (c)(4), ce qui explique le gain d'IoU.

Il est intéressant de noter que les points positifs et négatifs sont :

- La réduction de contexte améliore la détection des routes bordées de végétation en zone (1),
- L'ajout de convolutions a un effet régularisateur (moins de petits groupes de pixels),
- L'ajout de convolutions augmente les effets de bords entre les dalles.

Taille des noyaux	Champs réceptif	mIoU	OA	IoU				
				● Routes	● Cultures	● Eau	● Végétation	● Bâtiments
3*3	25	83,86	93,34	71,66	92,16	91,99	79,27	84,24
5*5	81	83,45	93,44	70,59	92,49	91,29	79,39	83,51
7*7	169	83,30	93,38	71,29	92,42	90,46	79,19	83,15
9*9	289	82,57	93,18	69,84	92,25	89,93	78,83	82,01

Table 3.8 – Métriques selon la taille des noyaux de convolution (%)

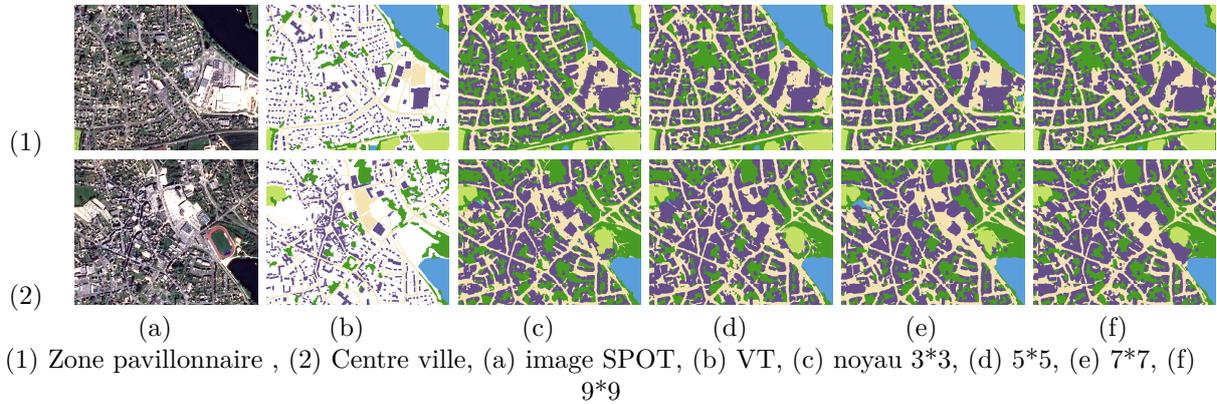


Figure 3.5 – Comparatifs visuels selon la taille des noyaux de convolution

3.5.2.2 Taille du noyau de convolution

Bien que les inférences de l'architecture à noyaux 5*5 (d) paraissent visuellement meilleures que les inférences de celle à noyaux 3*3 (c) et que leurs scores de mIoU ne soient pas séparables par l'Intervalle de Confiance à 2 écarts-types (IC-95), la décroissance du mIoU est fortement corrélée avec l'augmentation de la taille du noyau. Les classes qui demandent une forte précision, *Routes* et *Bâtiments*, sont celles qui subissent les plus grosses pertes d'IoU.

La classe *Cultures* affiche un gain d'IoU supérieur à son IC-99 pour les noyaux 3 * 3. On remarque les mêmes phénomènes que l'analyse du nombre de convolution sur la corrélation entre augmentation de régularisation et augmentation de champ réceptif.

3.5.2.3 Gestion des bords

Au vu des fluctuations sur les effets de bords avec l'augmentation du contexte, nous avons étudié l'impact des différentes méthodes de gestion des marges dans les convolutions disponibles dans la bibliothèque [PyTorch, 2019]. Nous avons testé les approches suivantes :

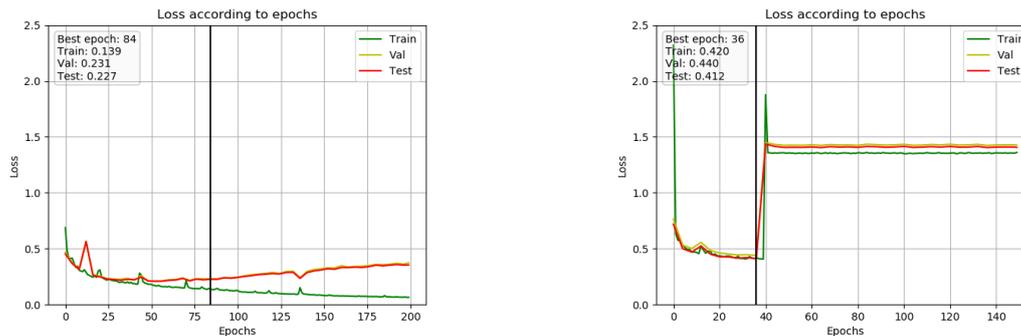
- Marges à 0,
- Marges constantes,
- Marges par réflexion (technique similaire à [Ronneberger et al., 2015]).

La seule différence notable est la perte d'un IC-95 sur la classe *eau* pour la gestion des marges "à 0" par rapport aux autres approches. Aucune différence de métrique séparable de l'aléatoire n'a été constatée entre l'approche par réflexion et par marges constantes. Visuellement les trois approches sont indifférenciables.

3.5.3 Normalisation

Bien que la Normalisation par paquet ou *Batch Normalization* (BN) soit connue pour être extrêmement importante pour assurer un apprentissage efficace, nous avons voulu le vérifier.

La figure 3.6 nous permet de comprendre que l'apprentissage n'est pas stable sans BN et qu'une divergence de gradient peut bloquer l'apprentissage. De plus l'entraînement sur les premières époques est bien plus performant pour la version avec BN.



Avec BN (à gauche) et sans BN (à droite). La barre noire indique la meilleure époque selon le mIoU.

Figure 3.6 – Comparatifs de l'évolution de la Loss au cours des époques (epochs) avec et sans BN

3.5.4 Connexions résiduelles

La compréhension du traitement de l'information spatiale qu'apporte la THR est cruciale. [Ronneberger et al., 2015] explique que les Connexions Résiduelles (CR) aident les attributs dits "de haute représentation" à mieux se localiser. Dans l'objectif de mieux comprendre l'apport de ces CR nous avons décidé de les supprimer.

	mIoU	OA	IoU				
			● Routes	● Cultures	● Eau	● Végétation	● Bâtiments
Avec RC	83,86	93,34	71,66	92,16	91,99	79,27	84,24
Sans RC	83,16	93,27	70,52	92,26	90,74	78,95	83,34

Table 3.9 – Métriques avec et sans CR

Si le score de mIoU grimpe d'un IC-99 pour la version avec CR, il est pourtant difficile de trouver quelle classe est à l'origine de ce gain. L'ajout de CR permet à la classe *Eau* de gagner un peu plus qu'un IC-95, mais le mIoU sans prendre ce gain en compte est de 83,61, soit deux écarts types. Les autres classes ne présentent pas d'augmentations significatives au regard de leurs IC-95 pour cibler une seule classe. Cela signifie que les bénéfices des CR sont répartis sur l'ensemble de classes, excepté de la classe *Cultures*, qui montre à nouveau un gain lors de la réduction de la prise en compte du contexte.

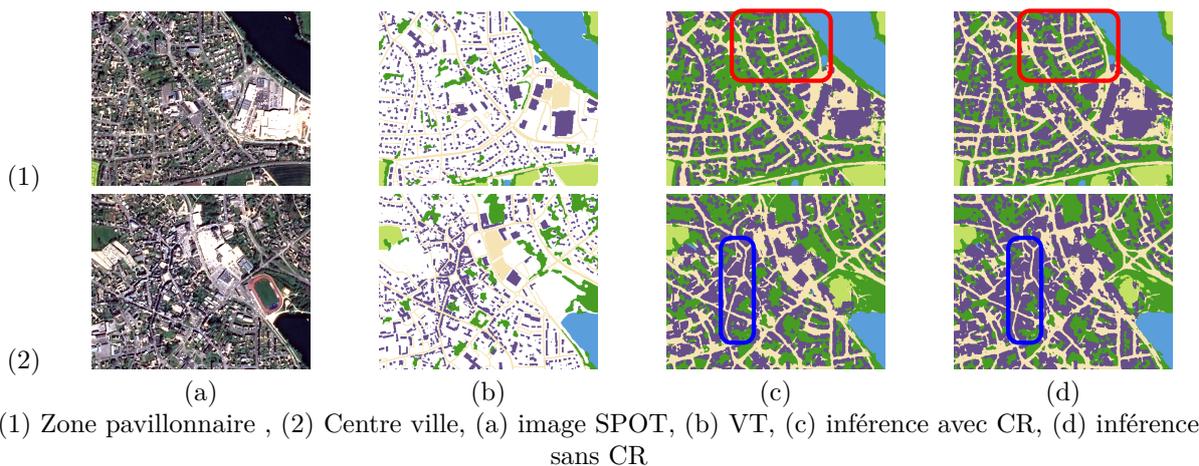


Figure 3.7 – Comparatifs visuels avec et sans CR

La figure 3.7 nous permet de constater le gain sur la délimitation des classes. Les instances de bâtiments sont bien mieux séparées dans les images (c) et en particulier dans la zone pavillonnaire (régions encadrées en rouge). Ce phénomène se manifeste sur la zone dense (2) par une meilleure délimitation des routes, qui ne présentent pas de rupture par jointure erronée entre bâtiments (régions encadrées en bleu).

3.5.5 Décodeur convolutif

L'architecture U-NET proposée par [Ronneberger et al., 2015] est composée de convolutions dans la partie décodeur. [Noh et al., 2015] utilisent des convolutions transposées dans le décodeur. Nous avons cherché à déterminer laquelle de ces deux approches permet de conserver au mieux l'information spatiale.

Néanmoins, ni les métriques ni l'analyse visuelle des résultats ne nous ont permis de départager les performances des deux approches. Nous avons conservé les convolutions transposées par choix arbitraire.

3.5.6 Profondeur du réseau

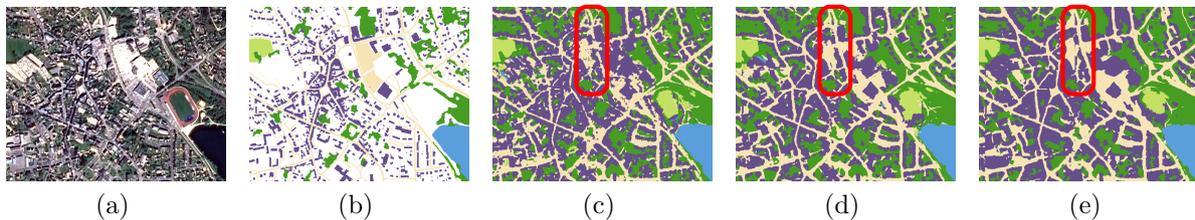
La profondeur d'une architecture est directement liée à sa capacité d'apprentissage d'attributs de haut niveau. Il est essentiel de s'assurer que l'architecture est capable de construire ses représentations. Dans notre cas, nous contrôlons la profondeur du réseau par son nombre d'étages. Néanmoins, chaque étage réduit la taille de l'image d'entrée par deux (avec les opérations de *max-pooling*), réduisant ainsi la localisation de l'information.

Pour conditionner le compromis entre gestion de localisation et représentation de haut niveau, nous avons comparé des configurations à plusieurs étages.

Nombre d'étages	mIoU	OA	IoU				
			● Routes	● Cultures	● Eau	● Végétation	● Bâtiments
1	82,39	93,02	69,00	92,00	91,02	78,47	81,49
2	83,86	93,34	71,66	92,16	91,99	79,27	84,24
3	83,38	93,35	71,74	92,35	90,91	79,04	82,87

Table 3.10 – Métriques en fonction du nombre d'étages de l'architecture

Les scores de mIoU et d'IoU sur la classe *Bâtiment* nous confirment rapidement que le compromis idéal est réalisé pour une architecture à 2 étages. Étonnamment, la classe *Route*, qui, comme constaté sur les tests précédents est une classe qui demande une forte localisation de l'information, affiche des performances égales lors de l'ajout d'un 3^{ème} étage. Nous constatons aussi que la classe *Cultures* bénéficie de cet ajout en représentation de haut de niveau.



Centre-ville, (a) image SPOT, (b) VT, (c) 1 étage, (d) 2, (e) 3

Figure 3.8 – Comparatifs visuels selon la taille des noyaux de convolution

Le manque de profondeur du réseau réduit significativement l'aspect visuel de l'image (c). En contrepartie, cette caractéristique augmente l'impact des effets de bords pour l'approche par réflexion, puisque le champ réceptif est plus grand (région encadrée en rouge). La comparaison visuelle entre les images (d) et (e) permet de visualiser la meilleure délimitation des objets de la classe *Bâtiment* pour l'architecture à 2 étages.

3.5.7 Conclusion de l'étude de l'architecture

Cette étude nous permet d'affirmer que :

- La BN est vitale pour garantir un apprentissage consistant,
- Deux convolutions de noyaux 3*3 permettent un champ réceptif idéal,
- L'utilisation de CR et de 2 étages assure une propagation de l'information spatiale jusqu'à la carte d'inférence, tout en étant capable d'apprendre des représentations de haut niveau pertinentes,
- Les connexions résiduelles aident toutes les classes, hormis les *Cultures*,
- Les classes *Bâtiments* et *Routes* sont les plus sensibles à l'information spatiale,

- La classe *Cultures* nécessite un champ réceptif fort et des attributs complexes,
- Les différentes approches de gestion des marges n'ont pas d'impact sur les métriques,
- Le type de convolutions dans le décodeur n'influe pas sur les métriques.

Ces points nous permettent de valider notre architecture de référence pour la production de carte d'OCS en particulier pour la détection des classes *Bâtiments* et *Routes*. Nous sommes cependant conscients qu'elle n'est pas optimale pour la détection de *Cultures*. La principale raison est sa destination à l'évaluation du potentiel des images SPOT, qui n'est pas le capteur le mieux adapté pour cette tâche.

Ce chapitre a pour objectif de valider le choix de notre approche au regard des méthodes de ML classiques et des approches par fenêtre glissante et du projet TERMOS. Il nous permet également de présenter la validation de la transférabilité de nos modèles dans le temps et l'espace. Enfin, il expose les travaux de raffinement de nomenclature et l'étude des configurations capteurs, en vue de la fusion tardive et précoce des données S2 et SPOT

4.1 Étude méthodologique

L'étude méthodologique nous permet de valider nos choix d'implantation en comparant nos résultats à ceux d'algorithmes de ML classiques, aux travaux de [Postadjian et al., 2017] et à l'approche CNN *binnaire*.

4.1.1 Forêt aléatoire

Cette comparaison a été réalisée afin de donner un aperçu du gain de performance entre les RF et les CNN. Les attributs sélectionnés ne sont pas optimisés pour la SegSem réalisée ici. Il est important de noter que ces tests ont été effectués après l'étude sur la détection d'instances de bâtiments, d'où la présence de la classe *Buffers*. Cela nous a permis d'étudier les performances des RF face à cette classe particulière.

La VT utilisée pour l'entraînement des RF est composée de 50 000 pixels par classe, soit 300 000 pixels au total, ce qui représente seulement 0.22% du jeu d'apprentissage du CNN. Pour chaque pixel, les attributs suivants ont été sélectionnés / calculés :

- Valeurs des canaux R, V, B, PIR
- Indice NDVI
- Variances sur 4 canaux suivant voisinages carrés de 3*3, 7*7, 10*10
- Moyennes sur 4 canaux suivant voisinages carrés de 3*3, 7*7, 10*10
- Mesure de complexité [Roger, 2006]

Résultats :

Les gains de performance sur le mIoU sont importants, et principalement dus aux gains importants sur les classes *Routes* et *Bâtiments*. Les attributs de description des textures étant générés directement par le CNN se montrent donc plus discriminants que ceux sélectionnés manuellement.

	mIoU	OA	IoU					
			● Buffer	● Eau	● Cultures	● Routes	● Végétation	● Bâtiments
classic	78,36	91,60	64,17	87,31	94,64	72,14	85,61	66,28
RF	65,17	86,77	46,43	80,07	91,6	47,19	81,76	43,98

Table 4.1 – Comparatifs des métriques entre RF et CNN

La figure 4.1 nous montre un exemple de confusion entre bâtiment et culture sur ses contours. Les gradients forts de l'image biaisent les descripteurs manuels. Ces images nous permettent aussi de souligner la difficulté de l'algorithme des RF sur la reconstruction du réseau routier.

Conclusion :

Sans surprise notre CNN affiche des performances bien supérieures aux RF. Étonnamment le résultat des RF sur la classe *Buffers*, qui est un mélange sémantique des éléments topographiques entourant les bâtiments, reste très convenable au vu de son faible champ réceptif de 10 pixels.

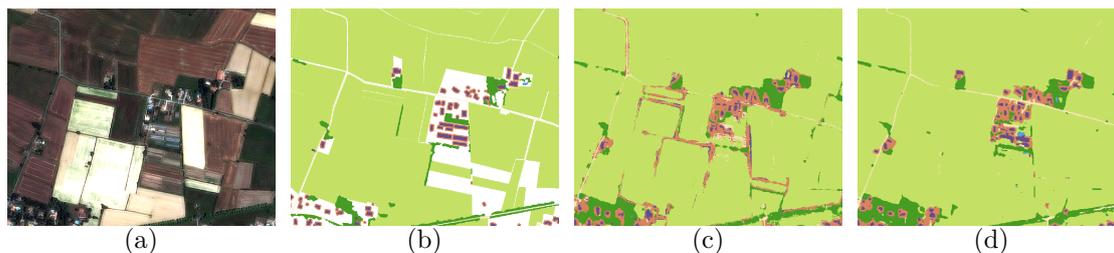


Figure 4.1 – Visuels des inférences sur zone de campagne des algorithmes RF et CNN
(a) image SPOT, (b) VT, inférences (c) CNN, (d) RF

4.1.2 Fenêtre glissante

Nos travaux se placent dans la continuité des travaux de [Postadjian et al., 2017] sur l’exploitation de la THR des capteurs SPOT. Nous avons réalisé cette étude comparative pour valider que la nouvelle architecture entièrement convolutive est réellement plus efficace dans le traitement de l’information spatiale. Il est important de préciser que les métriques proviennent de l’article [Postadjian et al., 2017] et, bien que les images soient identiques, nous ne pouvons pas garantir que la vérité terrain soit parfaitement similaire. Elle a cependant été construite à partir des mêmes éléments des mêmes BD, c’est pourquoi nous nous permettons tout de même d’effectuer des comparaisons.

L’architecture utilisée par [Postadjian et al., 2017] a été volontairement conçue pour être légère et facile à entraîner. Elle ne possède donc que 300 000 paramètres, contre 1 330 000 pour notre architecture *SO-NET*. En revanche, les progrès dans le domaine du DpLg et l’amélioration des bibliothèques comme [PyTorch, 2019] et l’approche entièrement convolutive ont permis, malgré cette différence de taille de réseau, une convergence plus rapide.

- [Postadjian et al., 2017] : 5000 époques en 61 heures soit 82 époques/h ; Convergence à partir de la 900^{ème} époque, soit 11h.
- Notre architecture : 150 époques en 2h 30 soit 60 époques /h ; Convergence avant la 100^{ème} époques soit 2h

Il est important de noter que les deux entraînements n’ont pas le même nombre de dalles d’entraînements.

Résultats : le tableau 4.1 nous montre que les métriques sont à l’avantage du CNN, en particulier sur les classes *Bâtiments* et *Routes*. L’augmentation des performances sur la classe *eau* provient de la réduction des confusions avec les ombres.

	OA	Fmoy	F-Score				
			● Eau	● Cultures	● Routes	● Végétation	● Bâtiments
CNN	93,33	91,026	95,83	95,92	83,49	88,44	91,45
FG	92	85	86	97	75	85	81

Table 4.2 – Comparatifs des métriques entre approche par fenêtre glissante et par approche entièrement convolutive

Nous pouvons constater grâce à la figure 4.2 que :

- En zone de campagne (1) notre approche permet une meilleure reconstruction des haies bordant les cultures, bien que son imprécision explique la perte d’IoU sur la classe *Cultures*,
- En zone dense (2) la difficulté de l’approche de [Postadjian et al., 2017] à distinguer la végétation urbaine et le réseau routier confirme les postulats de [Ronneberger et al., 2015] sur la difficulté de circulation de l’information spatiale dans les architectures par fenêtres glissantes,
- La zone pavillonnaire (3) en est un autre exemple,
- Les objets à signatures particulières (4) absents de la nomenclature (surfaces minérales) présentent les mêmes difficultés.

Conclusion : l’approche entièrement convolutive permet d’entraîner et d’inférer des cartes de CS plus rapidement et en exploitant la THR du capteur SPOT avec plus d’efficacité.

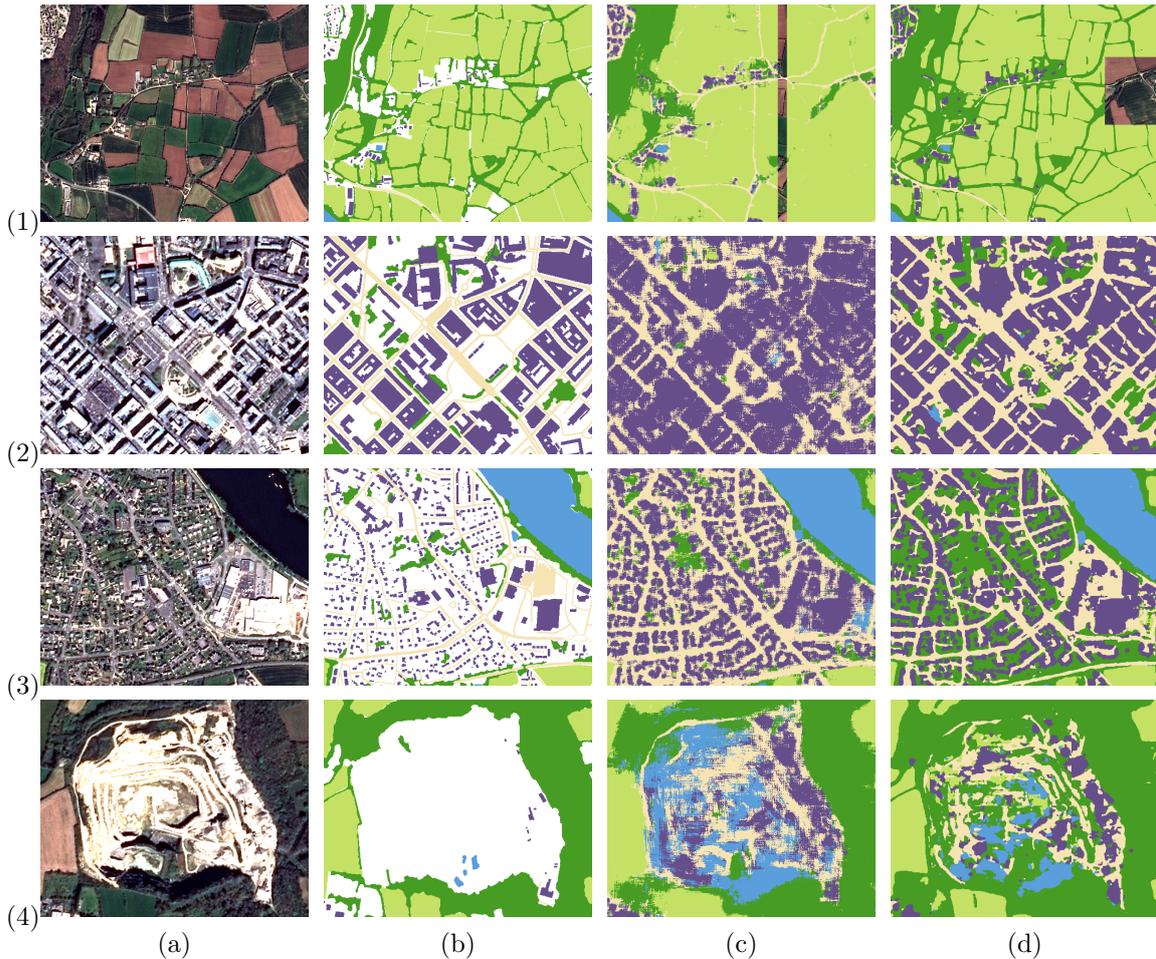


Figure 4.2 – Visuels des inférences sur entre notre approche et celle de [Postadjian et al., 2017] (1) zone de campagne, (2) zone dense, (3), zone pavillonnaire, (4) zone d'extraction minière, (a) image SPOT, (b) VT, inférences (c) [Postadjian et al., 2017], (d) notre approche

4.1.3 Approche binaire

Bien que l'approche proposée ait permis de surpasser les performances de [Postadjian et al., 2017], il existe d'autres approches entièrement convolutive pour la production de SegSem.

Notre approche considère la SegSem à nomenclature complète, c'est à dire que toutes les classes sont présentées à notre modèle lors de son entraînement. Le projet TERMOS de l'IGN s'appuie sur une approche *binaire*, qui consiste à créer un modèle pour chaque classe en utilisant une VT binaire composée de la classe considérée et d'une classe *Autres* composée de la surface inverse (tout ce qui n'est pas de la classe considérée, y compris les zones sans données). La SegSem finale est produite par fusion des cartes de probabilités générées pour chacun des modèles. Cette fusion peut être effectuée de manière heuristique (seuillage, *argmax*) ou par ML. Sa principale limitation réside dans le besoin d'un modèle par entrée de nomenclature.

Nous avons uniquement généré des inférences à partir des fusions heuristiques citées précédemment. Les seuils de probabilité utilisés sont les suivants :

- Routes : 0,65
- Cultures : 0,7
- Eau : 0,8
- Végétation : 0,6
- Bâtiments : 0,8

Résultats :

Les métriques concernant l'inférence par seuillage n'ont pas été calculées en raison de la sélection visuelle des seuils. Seule l'analyse visuelle est importante. Dans le tableau 4.3, nous constatons que les métriques générales sont en faveur de l'approche à nomenclature complète.

	mIoU	OA	IoU				
			● Eau	● Cultures	● Routes	● Végétation	● Bâtiments
Multi	83,86	93,34	71,66	92,16	91,99	79,27	84,24
Solo	82,79	93,12	71,2	92,24	88,42	77,36	84,75

Table 4.3 – Comparatifs des métriques entre approche à nomenclature complète et approche *binnaire*

La figure 4.3 nous montre que les inférences sont visuellement plus propres pour l'approche à nomenclature complète (e) sur les zones complexes comme sur les parkings que l'approche *binnaire* (d). En revanche la proximité visuelle entre le seuillage (c) et la vérité terrain (b) est évidente.



Figure 4.3 – Visuels des inférences sur zone de campagne des algorithmes RF et CNN

(a) image SPOT, (b) VT, inférences (c) par seuillage, (d) par probabilité max, (e) par nomenclature complète

La figure 4.4 présente les cartes de probabilités pour les deux approches. Pour toutes les cartes, l'approche *binnaire* fournit des résultats bien mieux délimités spatialement. On remarque que les probabilités de la végétation sont très faibles sur la zone pavillonnaire présentée. Ceci est lié au fait que la BD Topo ne décrit pas la végétation urbaine, mais uniquement les objets de végétation les plus grossiers. Le réseau n'a donc pas suffisamment d'exemples pour l'apprentissage.

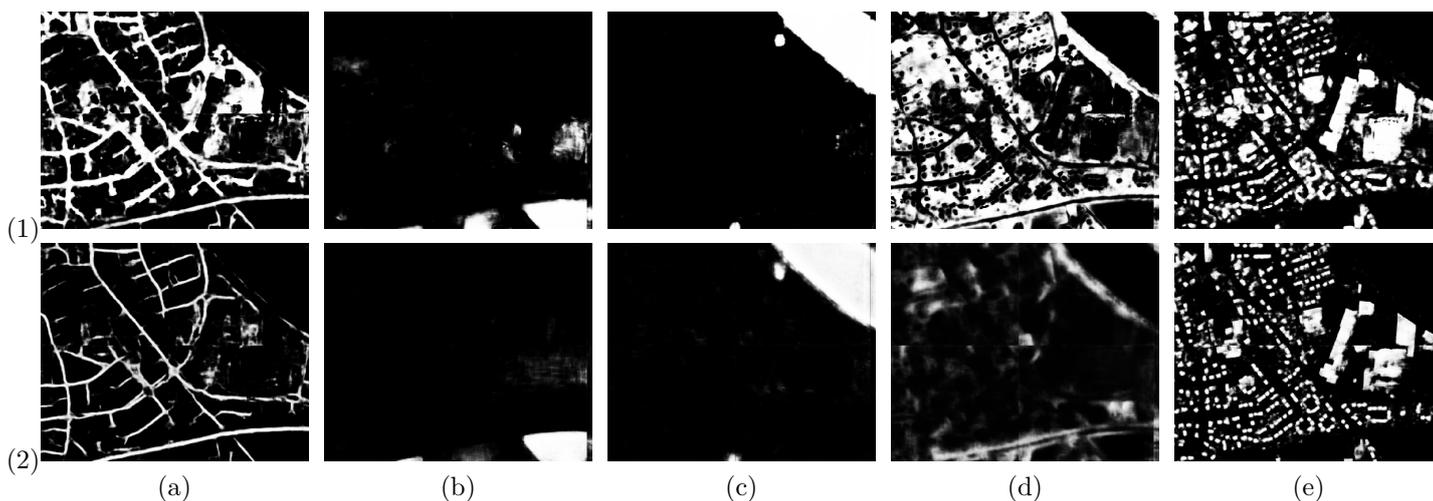


Figure 4.4 – Visuels des cartes de probabilités des approches par nomenclature complète et *binnaire*

(1) probabilités approche par nomenclature complète, (2) probabilités approche *binnaire*, pour les classes (a) *Routes*, (b) *Cultures*, (c) *Eau*, (d) *Végétation* et (e) *Bâtiments*

Conclusions : dans l'objectif d'avoir une carte de SegSem complète, notre approche s'avère plus performante et plus rapide en ne nécessitant qu'un seul modèle. Elle l'est d'autant plus au regard de notre volonté de raffinement de la nomenclature. Elle permet aussi de s'affranchir de la question de la

fusion à posteriori. Inversement, dans le cas de SegSem discontinue, on remarque que les inférences des cartes de probabilités seuillées se rapprochent de la VT. L'ajout de contraintes sur l'apprentissage a permis d'améliorer la délimitation des objets topographiques.

4.2 Transférabilité des modèles

[Postadjian et al., 2017] étudient la transférabilité de leurs modèles dans le temps et l'espace. Ils mettent en avant l'importance du Transfert d'apprentissage ou *Fine Tuning* (FT) dans le cas du transfert spatial. Dans leur cas, le modèle entraîné sur la région du Finistère a appris des représentations de l'eau avec une faible turbidité. Lors du transfert vers la Région de la Gironde, dont le fleuve possède une forte turbidité, le modèle a classifié la majeure partie du fleuve comme *Bâtiment*. Le FT a permis au modèle d'apprendre cette nouvelle représentation de l'eau avec seulement quelques échantillons. Nous avons réalisé les mêmes tests pour vérifier que la transférabilité n'était pas dégradée par l'utilisation d'une approche entièrement convolutive, et que le FT restait aussi efficace.

4.2.1 Transfert spatial

Nous avons réalisé une comparaison entre :

- un modèle entraîné sur la Gironde à partir d'une initialisation aléatoire des poids du réseau (AL),
- un modèle pré-entraîné sur le Finistère puis entraîné sur la Gironde (FT),
- une inférence à partir du modèle entraîné sur le Finistère (INF).

Résultats : le tableau 4.4 montre le gain général sur les métriques par l'utilisation de FT, et en particulier sur la classe *Bâtiments*.

	Meilleure époque	mIoU	OA	IoU				
				● Routes	● Cultures	● Eau	● Végétation	● Bâtiments
AL	152	81,59	95,01	63,62	76,32	94,09	93,97	79,92
INF	-	63,35	84,81	44,47	64,86	74,48	81,98	50,98
FT	72	83,76	93,07	69,9	83,25	94,13	89,04	82,46

Table 4.4 – Comparatifs des métriques sur les stratégies d'apprentissage

Les métriques concernant le modèle *AL* sont données à titre indicatif, puisque les zones des tests utilisées pour les modèles INF et FT sont différentes.

La figure 4.5 nous permet de montrer que :

- sur la zone de la Gironde (1) nous n'avons pas eu de confusion entre la classe *Bâtiment* et la classe *Eau* sur le fleuve lors du transfert direct (1c),
- le transfert direct présente de fortes confusions entre *Eau* et *Cultures* (2c). Ces confusions n'en sont pas réellement puisque l'image (2a) révèle des cultures inondées. Il y a donc bien une présence d'eau, bien que la VT indique une classe *Cultures*,
- le FT bénéficie à la détection des bâtiments (3), ce qui confirme les scores d'IoU

La figure 4.6 confirme que les temps d'apprentissage sont bien réduits dans le cas du FT. Dès la 30^{ème} époque le modèle FT affiche un mIoU quasiment équivalent à sa meilleure époque. On peut ainsi considérer un facteur de 5 entre le temps d'apprentissage d'un modèle par initialisation des poids aléatoires et celui d'un modèle FT.

Conclusions :

Nos modèles présentent de bonnes capacités de transférabilité spatiale sur les d'objets topographiques larges (*Cultures*, *Eau*), bien qu'une opération de FT permette d'augmenter les performances sur les objets plus fins (*Routes*, *Bâtiments*). Concernant l'apprentissage, le FT affiche une rapidité intéressante. Néanmoins il serait intéressant d'étudier :

- La quantité de dalles nécessaire à l'atteinte des performances équivalentes à un modèle *AL*
- Les performances d'un modèle mixte aux deux zones à partir d'un jeu de données composé de dalles du Finistère et de la Gironde

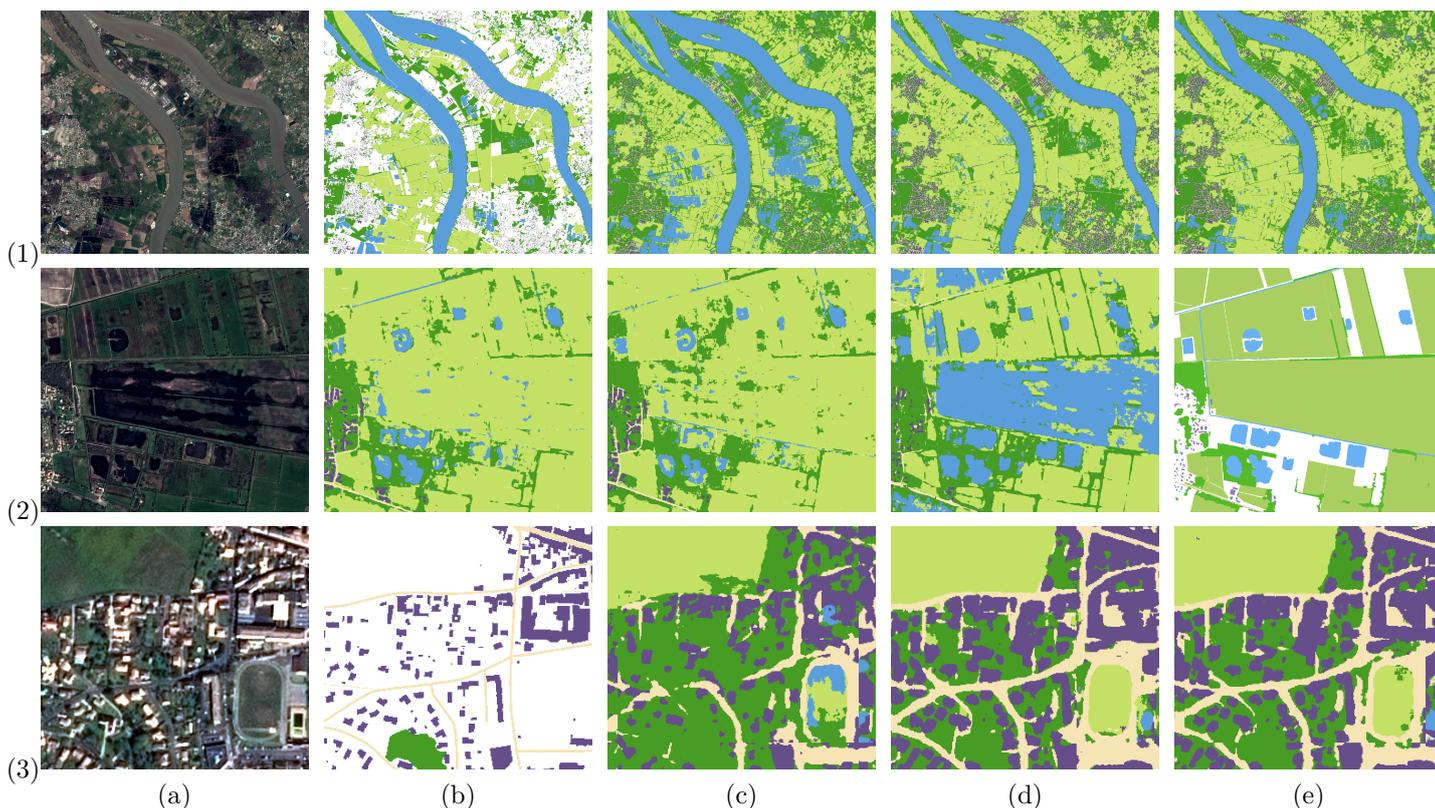


Figure 4.5 – Visuels des inférences entre transfert direct, initialisation aléatoire et FT
 (1) Gironde, (2) zone de cultures inondées, (3) zone pavillonnaire, (a) image SPOT, (b) VT, inférences
 (c) transfert direct, (d) initialisation aléatoire, (e) FT

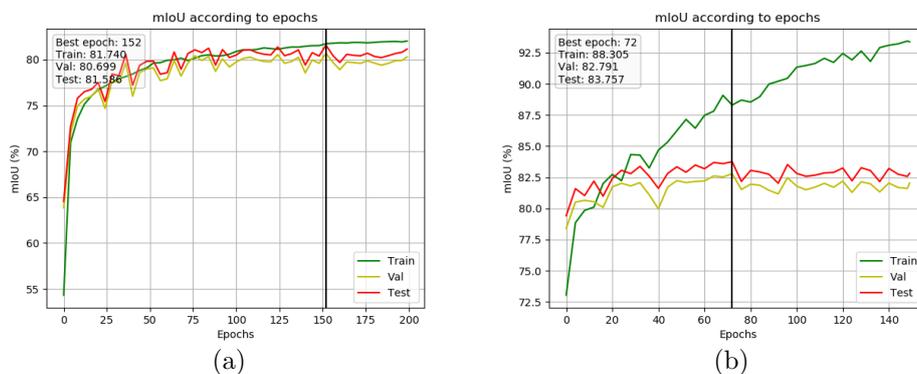


Figure 4.6 – Courbes des scores de mIoU au cours de l'entraînement des modèles pour l'étude du FT
 (a) Initialisation aléatoire, (b) FT

4.2.2 Transfert temporel

L'intérêt du transfert temporel réside dans la possibilité d'utilisation d'un modèle entraîné sur une image SPOT vers d'autres images SPOT du même environnement, acquises avec un décalage temporel. Dans notre cas nous avons étudié le transfert temporel entre une image SPOT de 2014 et une image de 2016 sur la même zone du Finistère. Le modèle de base a été entraîné sur l'image de 2014. Le tableau 4.5 expose le même comportement que pour le FT spatial, bien que les écarts entre les deux modèles soient plus réduits.

	mIoU	OA	IoU				
			● Routes	● Cultures	● Eau	● Végétation	● Bâtiments
inf	78,21	91,93	66,97	91,07	77,31	75,26	80,48
TF	82,98	93,05	70,77	91,84	91,41	78,65	82,25

Table 4.5 – Comparatifs des métriques sur les stratégies d'apprentissage

4.3 Étude thématique

4.3.1 Détection d'instance de bâtiment

Au vu des résultats de l'approche *binnaire* sur les cartes de probabilité des bâtiments et de l'étude [Wendl et al., 2018], il s'avère que l'ajout de contrainte autour des objets topographiques améliore leur détournement. C'est dans cette optique que nous avons ajouté une classe visant à mieux contraindre la classe *Bâtiment*. Cette nouvelle classe sera nommée *Tampon de bâtiments*; l'anglicisme *Buffer* sera aussi utilisé.

4.3.1.1 Mise en place des *Buffers*

La VT des *Buffers* à été créée par dilatation des objets *Bâtiments* de la BD Topo. Nous avons cherché à définir quelle valeur de dilatation permet le meilleur détournement des bâtiments. Nous avons entraîné des modèles pour les dimensions de *Buffers* suivantes : 2m, 5m, 10m et 20m.

Résultats :

Le tableau 4.6 nous permet de voir que les performances globales, en particulier le mIoU, chutent considérablement lors de l'ajout de cette nouvelle classe. A l'origine de ce phénomène se trouve la confusion entre les 3 classes *Bâtiments*, *Buffers* et *Routes*. Leur proximité sémantique évidente impacte la capacité de discrimination de notre CNN.

	mIoU	OA	IoU					
			● Buffers	● Eau	● Cultures	● Routes	● Végétation	● Bâtiments
Sans	83,86	93,34	-	91,99	92,16	71,66	79,27	84,24
2m	74,43	91,87	47,51	91,02	91,76	70,02	78,39	67,90
5m	75,91	91,58	62,99	91,53	92,32	65,14	77,55	65,94
10m	74,99	90,85	67,41	90,47	92,18	59,33	76,02	64,52
20m	72,32	89,30	65,55	90,69	91,72	50,98	73,76	61,25

Table 4.6 – Comparatifs des métriques entre les différentes tailles de *Buffers*

La figure 4.7 montre l'évolution des inférences en fonction de l'augmentation de la dilatation originale. Cette analyse visuelle permet de bien comprendre comment cette nouvelle classe interagit avec celles déjà présentes dans notre nomenclature A5, en l'occurrence sur la disparition de la végétation urbaine au profit de la classe *Buffers*. L'image (e) montre que l'utilisation de *Buffers* trop grands empêche la détection de végétation urbaine.

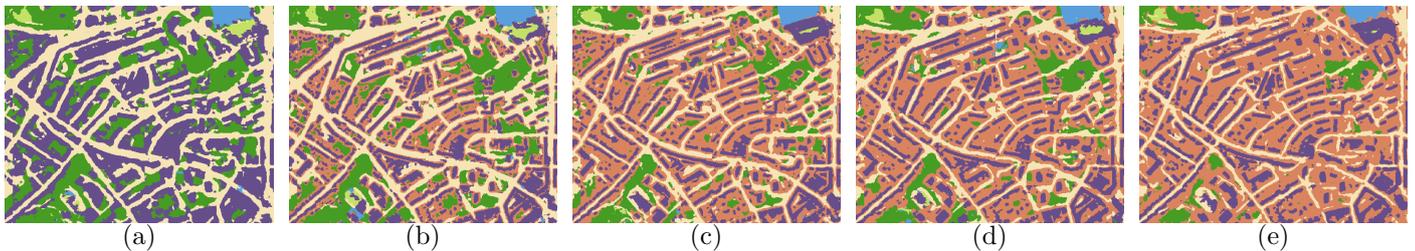


Figure 4.7 – Visuels des inférences entre les différentes tailles de *Buffers*
 (a) sans *Buffers*, *Buffers* de (b) 2m, (c) 5m, (d) 10m, (e) 20m

Conclusion :

Bien que les métriques présentées dans le tableau 4.6 indiquent qu'un *Buffer* de 2m apporte la meilleure

délimitation des routes et des bâtiments, un *Buffer* de 5m donne le meilleur mIoU. Nous avons donc opté pour ce choix et intégré cette nouvelle classe dans la nomenclature A6.

4.3.2 Comparaison avec l'approche seuillage

Grâce à l'ajout de contraintes dans la VT, nous avons pu améliorer la délimitation des bâtiments, au point de pouvoir les instancier. Cette classe supplémentaire nous a permis de retrouver, sur une approche à nomenclature complète, les performances qu'apportait l'approche de seuillage de cartes de probabilités sur la précision de la délimitation.



Figure 4.8 – Visuels des inférences entre seuillage de cartes de probabilités, nomenclature A6 et A5
(a) image SPOT, (b) VT, inférences de (c) seuillage de cartes de probabilités, (d) nomenclature A5, (e) nomenclature A6

4.3.3 Retrait

Bien que cette nouvelle classe *Buffer* offre une meilleure performance d'instanciation de bâtiments, elle ne fait pas partie des besoins utilisateurs. La question de son retrait à posteriori de l'inférence s'est posée. Nous avons donc recréé une inférence à partir des cartes de probabilités générées par la nomenclature A6, mais sans la classe *Buffers*, dans l'objectif de déterminer si sa présence a aidé à mieux séparer la végétation urbaine, des bâtiments et des routes. La figure 4.9 compare les différentes inférences ainsi générées.

L'utilisation de *Buffers* ne modifie pas les probabilités des autres classes. Nous n'avons pas décelé de



Figure 4.9 – Visuels des inférences entre seuillage de cartes de probabilités, nomenclature A6 et A5
(a) image SPOT, (b) VT, inférences de (c) seuillage de cartes de probabilités, (d) nomenclature A5, (e) nomenclature A6

changement de comportement, supérieur au bruit, entre les inférences d'un modèle selon A5 et un modèle selon A6 avec retrait des *Buffers*. Cette particularité met cependant en avant une propriété de constance entre les modèles de CNN.

4.3.4 Raffinement des nomenclatures

Jusqu'à maintenant, nous avons limité les tests de notre architecture à des SegSem selon des nomenclatures simples, composées des objets topographiques courants. L'objectif est de mesurer la performance de notre architecture dans l'utilisation de la THR pour la détection d'objets plus complexes. Les tests ont été réalisés sur la région de Toulouse.

4.3.4.1 1^{er} niveau de raffinement

Le premier niveau de raffinement concerne l'extension de la nomenclature A6 vers la nomenclature H13. Les principales modifications et leurs objectifs sont :

- l'ajout des cimetières : tentative de détection d'objets très particuliers mais peu représentés,
- la dilatation de la classe *Végétation* en *Feuillus*, *Conifères*, *Landes* et *Haies* : comprendre les limites de détection sur la finesse sémantique de la végétation à partir du capteur SPOT,
- la dilatation de la classe *Routes* en *Routes*, *Parkings* et *Piste d'aérodromes* : Comprendre les limites de détection sur la finesse sémantique de sous classes à caractéristiques radiométriques proches.
- La dilatation de la classe *Bâtiments* en *Industriel* et *Résidentiel* : extension vers des classes d'US à fort potentiel de télédétection.

Résultats :

On constate une chute importante sur les métriques de mIoU sur le tableau 4.7.

	mIoU	OA
6c	78,36	91,60
13c	58,93	87,28

Table 4.7 – Comparatifs des métriques globales entre les modèles entraînés selon les nomenclatures A6 et H13

		IoU											
6c	● Buf.	● Eau	● Cu.	● Routes			● Végétation					● Bâtiments	
	64,17	87,31	94,64	72,14			85,61					66,28	
13c	● Buf.	● Eau	● Cu.	● Rts.	● Aér.	● Par.	● Cim.	● Con.	● Feu.	● Hai.	● Lan.	● Rés.	● Ind.
	64,73	85,28	93,63	66,07	44,85	74,33	30,67	43,31	80,16	43,97	29,43	57,63	51,99

Table 4.8 – Comparatifs des IoU entre les modèles entraînés selon les nomenclatures A6 et H13

Le faible score de la classe *Cimetière* s'explique par le fait que cette classe soit définie comme une classe d'US. Elle peut être ainsi recouverte de pelouse, de routes ou de tombes (fig. 4.10). Hors la texture particulière des cimetières ne concerne que les tombes. Les autres éléments comme les pelouses sont confondus avec la classe *Landes*. De plus, il s'avère que la faible quantité de VT les concernant a entraîné un phénomène de sur-détection de ces textures sur l'ensemble de la zone.

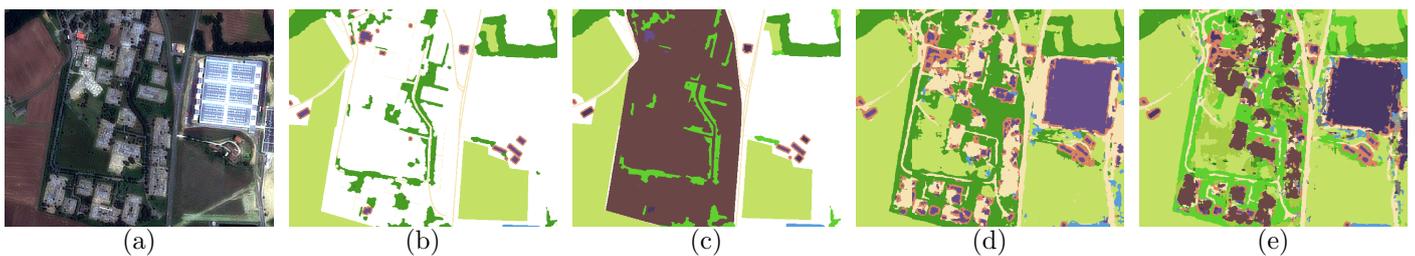


Figure 4.10 – Visuels des inférences pour le premier niveau de raffinement, sur cimetière
(a) image SPOT, (b) VT 6c, (c) VT 13c, inférences (d) 6c, (e) 13c

La dilatation de la classe *Route* montre les difficultés de notre architecture à différencier les parkings des pistes d'aérodromes. On retrouve également des confusions entre ces deux dernières classes et les toits de bâtiments industriels (fig. 4.11).

Les scores du tableau 4.8 confirment que la distinction des types de végétation est difficile à partir du capteur SPOT. Le bon score de la classe *Feuillus* est à mettre en parallèle de sa grande présence dans la VT. Enfin, la matrice de confusion 4.12 nous montre que les principales chute d'IoU sur les classes *Industriels* et *Résidentiel* provient d'une confusion mutuelle.

Conclusions :

Ce test nous permet de mettre en avant :

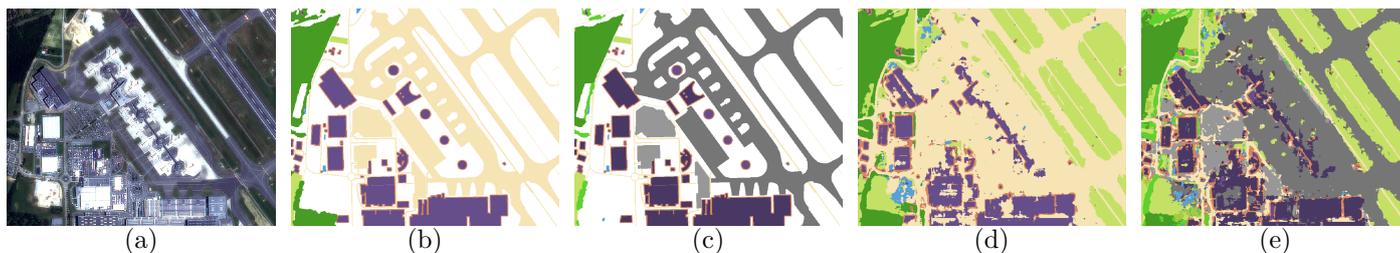


Figure 4.11 – Visuels des inférences pour le premier niveau de raffinement, sur aéroport
(a) image SPOT, (b) VT 6c, (c) VT 13c, inférences (d) 6c, (e) 13c

- Les difficultés sur les séparations de classes d’usages,
- Que pour les objets à texture particulières peu représentés une approche *binnaire* serait plus pertinente,
- Les difficultés de discriminer les objets radiométriquement proches et de texture similaires à partir du capteur SPOT (toits de bâtiments industriels et pistes d’aérodromes, feuillus et conifères).

4.3.4.2 2ème niveau de raffinement

Le 2^{ème} niveau de raffinement concerne le passage de la nomenclature H13 à la nomenclature H18. Son objectif est d’augmenter la difficulté de la SegSem. Les principaux sujets d’études portent sur la capacité de détection de classes :

- sémantiquement proches, mais de définitions différentes (chemins et routes principales),
- dont les définitions sont basées sur des proportions de représentativité d’autres classes déjà présentes dans notre nomenclatures (ex : forêts ouvertes)
- de végétations présentant des textures très particulières que SPOT pourrait permettre de détecter, mais avec de très faible échantillons d’apprentissages (vignes / peupleraies/ vergers)

Résultats :

Les modèles basés sur cette nomenclature produisent des résultats comparables aux modèles basés sur la nomenclature H13, d’une part par la chute de mIoU (tab. 4.9), et d’autre part sur le comportement dans les zones de cimetières et d’aéroports.

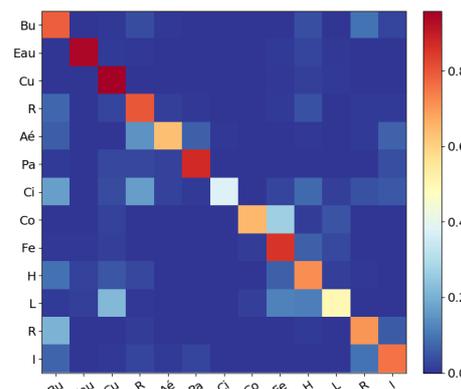


Figure 4.12 – Matrice de confusion pour la SegSem à 13 classes

	mIoU	OA
6c	78,36	91,60
18c	48,86	86,32

Table 4.9 – Comparatifs des métriques globales entre les modèles entraînés selon les nomenclatures A6 et H18

sémantiquement La classe *Forêts ouvertes*, qui est une classe dont la définition est basée sur les proportions de représentativité affichées des scores d’IoU très faibles (tab. 4.10). La matrice de confusion 4.13 confirme les principales confusions avec les autres classes de végétation. La figure 4.14 montre un exemple d’incapacité de détection de forêt ouverte. Les chemins et les routes affichent également une forte confusion, du fait de leur proximité sémantique. La figure 4.14 montre la difficulté du CNN à séparer les classes *Routes principales* et *Chemins* en éléments routiers consistants.

Concernant les objets à forte texture, la figure 4.15 nous permet de montrer que :

- les peupleraies et les vergers présentent les mêmes problèmes. Bien qu’ayant des textures particulières, ces textures restent trop proches de celles d’éléments sur-représentés comme les feuillus et les landes. Ils affichent donc des scores très faibles d’IoU (tab. 4.10),

		IoU							
6c	● Buf.	● Eau	● Routes					● Bâtiments	
	64,17	87,31	72,14				-	66,28	
18c	● Buf.	● Eau	● Chem.	● Rt. prin.	● Aéro.	● Park.	● Cim.	● Rési.	● Indus.
	64,28	86,67	12,7	64,96	42,12	69,86	30,18	56,86	45,98
6c	● Cult.	● Végétation							
	94,64	85,61							
18c	● Cult.	● Conif.	● Feu.	● F. ouv.	● Haies	● Landes	● Peupl.	● Vergers	● Vignes
	93,69	36,97	80,6	8,96	44,3	26,08	27,94	2,26	85,04

Table 4.10 – Comparatifs des métriques d'IoU entre les modèles entraînés selon les nomenclatures A6 et H18



Figure 4.14 – Visuels des inférences pour le 2^{ème} niveau de raffinement, en zone de campagne
(a) image SPOT, (b) VT 6c, (c) VT 18c, inférences (d) 6c, (e) 18c

- Les vignes affichent une performance de détection élevée et précise. Néanmoins cette portion de vignes présente un contexte très favorable de sol nu, et donc de contraste de texture fort.

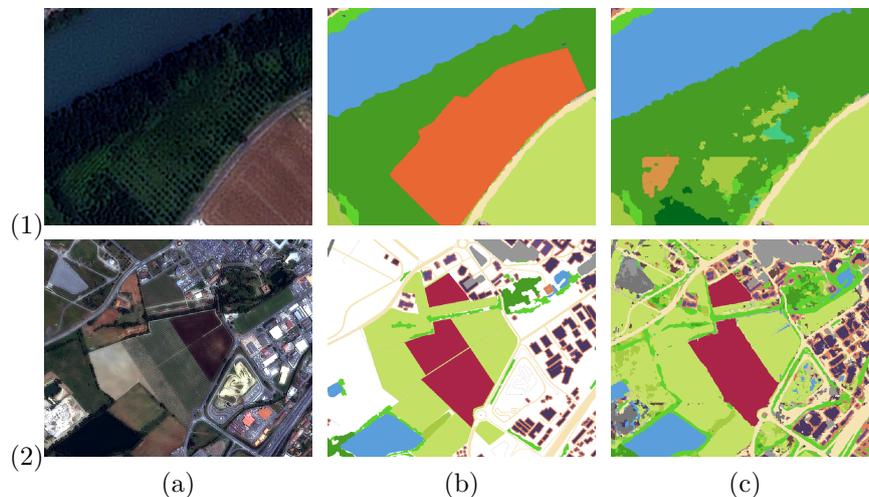


Figure 4.15 – Visuels des inférences pour le 2^{ème} niveau de raffinement, sur la végétation à forte texture
(1) Peupleraie, (2) vignes, (a) image SPOT, (b) VT, (c) inférences modèle H18

Conclusion :

Ce deuxième niveau de raffinement nous a permis de montrer que notre architecture n'a pas été capable d'apprendre des notions de proportions de représentativité des classes. Les objets à forte texture sont très bien retrouvés, mais uniquement dans les cas où ils sont les seuls à présenter ces textures. Enfin, la capacité de différenciation d'éléments proches sémantiquement est intéressante, bien qu'imparfaite. Elle peut permettre de laisser le choix au CNN de décider seul des frontières.

4.3.4.3 Stabilité des classes

Il s'avère que la propriété de constance dans la prédiction des CNN, mise en avant dans l'étude du retrait des *Buffers*, se retrouve également lors du raffinement des nomenclatures. Le tableau 4.11 présente les scores selon la nomenclature A6 des inférences des modèles entraînés avec les nomenclatures A6, H13 et H18. Les classes dilatées précédemment ont été regroupées pour permettre ce calcul.

Bien que les scores de mIoU chutent à chaque niveau de raffinement, ce tableau nous montre qu'ils remontent quasiment à la valeur du modèle entraîné sur 6 classes. Cela signifie que les erreurs commises dans les prédictions concernent majoritairement des confusions au sein des sous-classes.

Cette capacité possède un intérêt pour la modularité de production de cartes d'OCS. Néanmoins, l'utilisation de classes comme les cimetières, complètement absentes des nomenclatures, sont impossible à traduire, et les erreurs de SegSem brisent le transfert de nomenclature. La figure 4.16 illustre les conséquences, les zones rouges représentant les impossibilités de traduction.

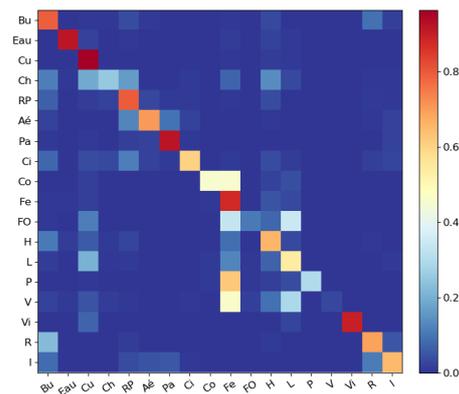


Figure 4.13 – Matrice de confusion pour la SegSem à 18 classes

	mIoU	OA	IoU					
			● Buffers	● Eau	● Cultures	● Routes	● Végétation	● Bâtiments
6c	78,36	91,60	64,17	87,31	94,64	72,14	85,61	66,28
13 as 6	76,87	91,00	63,44	84,59	93,62	70,74	83,88	64,95
18 as 6	76,37	90,63	62,96	86,58	93,44	68,69	83,93	62,65

Table 4.11 – Comparatifs des métriques selon A6 des modèles entraînés selon A6, H13 et H18



Figure 4.16 – Visuels des inférences transférées en nomenclatures A6 (a) image SPOT, (b) 6c, (c) 13c, (d) 18c. Les zones rouges représentent les impossibilités de traduction

4.3.5 Vérité terrain issue de classification antérieures

Dans la section 4.3.4.1, nous avons mis en avant, avec l'exemple des cimetières, que la capacité de généralisation des CNN permet de corriger certaines incohérences d'annotation de la VT. Nous avons utilisé cette propriété pour tenter d'améliorer :

- *A posteriori*, les cartes de SegSem OSO,
- La détection de végétation urbaine.

4.3.5.1 Amélioration de la détection de végétation urbaine

Bien que l'architecture proposée ait permis une première amélioration de la détection de la végétation urbaine (cf. 4.1.2), elle reste encore imparfaite. Les causes proviennent principalement de la VT de la classe *Végétation* issue de la BD Topo, dont la définition présente deux défauts :

- La présence d'un seuil de groupement minimal avant annotation,
- La faible présence d'annotation de végétation en milieu urbain.

Le premier point induit une absence complète de l'association des éléments individuels à la classe *Végétation* lors de l'apprentissage. Le deuxième point, en partie conséquence du premier, réduit les exemples

d'apprentissage où les classes *Bâtiments* et *Végétation* sont mitoyennes, ce qui engendre une plus faible capacité de distinction entre les deux classes.

Création d'une pseudo-VT Puisque le problème est issu d'un manque d'exemple de végétation urbaine, nous avons décidé de créer une nouvelle VT. C'est pourquoi nous avons utilisé un algorithme RF pour déduire la présence de végétation urbaine à partir d'ortho-images de l'IGN (25cm de résolution) et d'un MNS.

Les images aériennes sont réalisées au mieux tout les 3 ans, contre au moins une image SPOT par année. Il est donc important de souligner que nous nous plaçons dans un contexte où nous possédons peu d'images aériennes, employées pour améliorer la détection à partir des images satellites. De plus si la dilution d'un algorithme RF dans un CNN semble contre-intuitive, nous la justifions par :

- La très faible présence de VT qui prohibe l'utilisation de DpLg,
- La différence de résolution entre les ortho-images et l'image SPOT, qui induit un sur-échantillonnage de la SegSem du RF avant utilisation en tant que VT. Cela réduit l'importance de la bonne délimitation des objets par le RF,
- La problématique multi-classe, qui sous-entend que le CNN traite un problème plus complexe (avec plus de classes) que le RF,
- L'intégration d'informations inexistantes dans les images satellites (meilleure résolution et MNS).

La nomenclature étant dédiée au milieu urbain, nous avons volontairement confondu les classes *Végétation basse* et *Cultures*. La figure 4.17 expose la différence entre les deux VT.

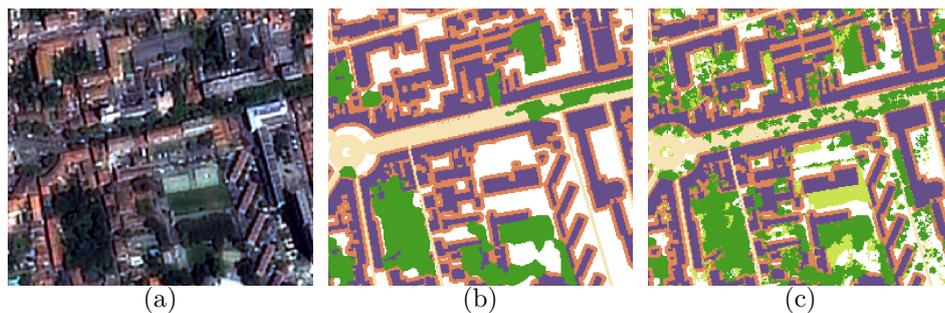


Figure 4.17 – Création d'une pseudo-VT
(a) image SPOT, (b) VT initiale, (c) pseudo-VT

Les objectifs de cette étude sont de confirmer :

- Que la précision de la VT influe sur la performance de détection,
- Qu'un processus rapide de création de VT plus fine, bien que local, bruité et imparfait améliore la détection dans notre cas d'application.

Résultats :

Les comparaisons effectuées dans le cadre de cette analyse ne peuvent qu'être visuelles du fait de l'absence de VT exacte concernant la végétation urbaine.

Le principal bénéfice de la pseudo-VT est mis en avant par la figure 4.18 : la nouvelle inférence (e) détecte aussi bien les groupements importants de végétation urbaine que l'inférence du modèle entraîné avec la VT initiale (d). Néanmoins nous notons une meilleure détection des petits groupements et des éléments individuels auparavant classés comme *Buffers*. En conséquence :

- La végétation urbaine en bordure de bâtiments est mieux détectée,
- La végétation en bordure de route est détectée.

D'autre part nous constatons que les éléments de végétation détectés suivent plus facilement les contours de la végétation de l'image : c'est la conséquence du retrait de l'effet de régularisation induit par la définition de la précédente VT. Enfin, concernant les routes, avec la contrainte qu'exerce la nouvelle VT sur la classe *Buffers*, on observe une amélioration de leur reconstruction, même en zone dense.

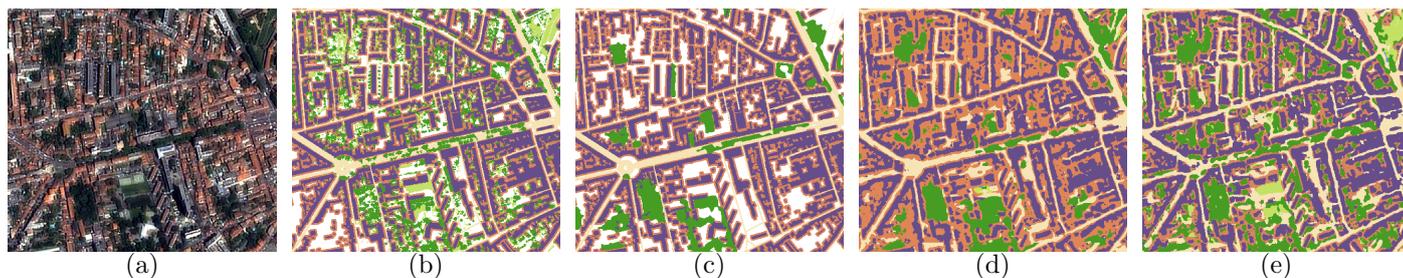


Figure 4.18 – Comparatif visuel des inférences avec et sans VT des classes de végétation urbaine issu d'un RF
(a) image SPOT, (b) VT initiale, (c) pseudo-VT, (d) inférence initiale, (e) nouvelle inférence

Étant donné que la végétation urbaine définie par l'algorithme RF empiètent sur la classe *Buffers*, nous avons réalisé un test de réduction de la taille des *Buffers*. Le tableau 4.12 permet de montrer que la réduction de la taille des *Buffers* en présence d'une VT urbaine fine permet une amélioration de la détection des routes, tout en augmentant celle des bâtiments. Néanmoins, cette meilleure performance est à mettre en perspective de la réduction de contrainte dans la VT (voir figure 4.19, et donc de pénalisation du score des classes lors de leur débordement. En effet l'analyse visuelle sur la figure 4.19 ne montre pas de changement significatif ni sur la forme des bâtiments, ni sur la régularité des routes. Dans l'exemple, l'inférence 5m (d) réalise même une meilleure reconstruction. Concernant la végétation urbaine, la comparaison visuelle montre une meilleure séparation avec la classe *Buffers* pour l'inférence à 2m (e) donc lors de la réduction de la taille des *Buffers*, en particulier pour la végétation basse.

Nous remarquons le même phénomène de réduction significative du score de IoU de la classe *Buffers* lors de la réduction de leurs tailles que dans l'étude exposée précédemment (cf. 4.3.1). Pourtant, contrairement à l'étude des *Buffers*, cette perte est compensée par le bénéfice de détection sur les classes *Bâtiments* et *Routes*, d'où une stabilité du score de mIoU. Nous n'avons pas pris en compte les score des deux classes de végétation, puisqu'en l'absence de VT exacte, il n'est pas possible de différencier un sur-apprentissage (qui induit une hausse d'IoU) d'une correction de VT (qui induit une baisse d'IoU).

Tailles <i>Buffers</i>	mIoU	OA	IoU					
			<i>Buffers</i>	<i>Eau</i>	<i>Végétation Basse</i>	<i>Routes</i>	<i>Végétation Haute</i>	<i>Bâtiments</i>
5m	64,11	73,58	46,63	83,78	71,15	60,48	41,08	65,58
2m	64,17	75,99	36,35	83,6	70,99	68,12	47,82	68,61

Table 4.12 – Comparatif des métriques selon la variation de la taille des *Buffers*

Les score d'IoU des classes *Végétation Basse* et *Végétation Haute* sont indiqués à titre indicatif et ne sont pas pris en compte pour le calcul du mIoU.

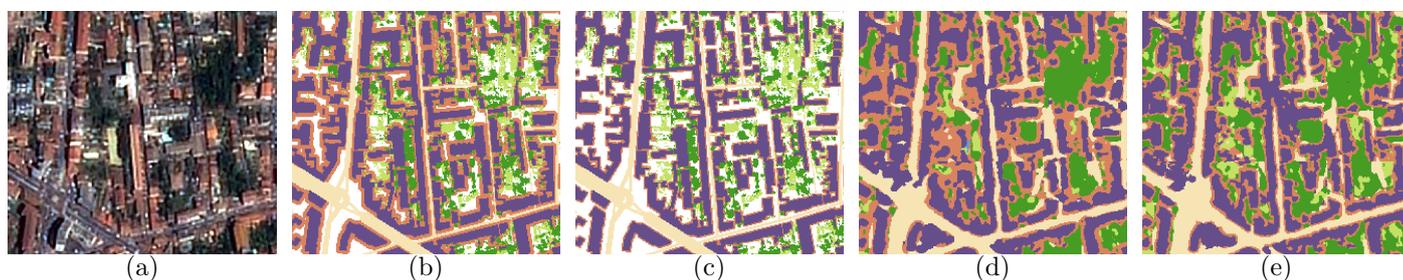


Figure 4.19 – Comparatif visuel des inférences selon la taille des *Buffers* avec VT des classes de végétation urbaine issues d'un RF

(a) image SPOT, VT avec *Buffers* (b) 5m, (c) 2m, inférence avec *Buffers* (d) 5m, (e) 2m

Conclusion :

Visuellement, la mise en place d'une pseudo-VT sur la végétation urbaine à partir d'une SegSem RF à permis d'améliorer la détection de cette dernière. Les zones prédites comme végétation urbaine étant, sur

les inférences précédentes, classées que *Buffers*, l'effet de la pseudo-VT agit comme une extraction de la texture liée à la végétation urbaine de la classe *Buffers*.

S'il est difficile de conclure sur le meilleur paramètre de la taille des *Buffers* au niveau global, il est certain que dans le cadre de la détection de la végétation urbaine, les *Buffers* de 2m sont les plus adaptés.

D'autre part l'architecture proposée semble avoir permis de corriger la pseudo-VT, mais, en l'absence de VT exacte, il ne nous est pas possible de confirmer par métriques l'amélioration observée.

4.4 Autres études non détaillées

Nous avons effectué un ensemble d'études supplémentaires pour valider notre architecture, l'utilisation de la THR de SPOT et le multi-spectral et multi-temporel de S2 dans la génération de produits d'OCS. Ces travaux, listés ci-dessous, ne sont pas détaillés dans cette version. Leurs résultats principaux figurent dans la conclusion générale de ce chapitre.

- Raffinement vers produit existant complet :
 - Production de SegSem de l'OCS-GE CS
 - Production de SegSem de l'OCS-GE CS
 - Production de SegSem de densité d'urbanisation selon UA
- Vérité terrain issue de classification :
 - Correction de la classification OSO à partir de la SPOT
- Impact de la configuration capteur :
 - Qualification des canaux SPOT avec comparaison des canaux P et XS pour la SegSem de CS
 - Qualification des canaux S2 avec comparaison des canaux XS et S2 IRVB
 - Comparaison des capteurs SPOT et S2 pour la SegSem de CS

4.5 Conclusion générale sur l'analyse des résultats

Après avoir comparé les performances des algorithmes de ML classiques, nous avons validé que l'approche entièrement convolutive est bien plus performante que l'approche par fenêtre glissante de [Postadjian et al., 2017]. Le transfert d'apprentissage est toujours applicable et confirme que les modèles peuvent être améliorés à moindre coût lors d'inférences de nouvelles régions. Nous avons également validé que l'approche de segmentation complète est plus adaptée à notre cas étude de raffinement de nomenclatures. En plus de simplifier l'entraînement, cette approche permet de s'affranchir de la question de la fusion de la donnée à posteriori.

Les résultats de [Wendl et al., 2018] et les cartes de probabilités issues de l'approche *un contre tous* nous ont permis de comprendre que l'ajout de contraintes autour des objets permet de mieux les détourner. Parallèlement, la précision de la nomenclature sur la végétation urbaine, en plus de mieux délimiter la végétation, apporte également une contrainte sur les bâtiments. Néanmoins "*le naturel revient au galop*" et sa suppression ne change pas les statistiques des autres classes, qui restent constantes.

Cette propriété de constance dans la prédiction a été retrouvée lors de l'étude du raffinement des nomenclatures. Bien que les scores généraux ont une forte tendance à chuter lors du passage des nomenclatures A6 à H18, il s'avère que les principales sources d'erreur sont inter-classes (dans les sous-groupes de nos nomenclatures). Cette propriété sous-entend la possibilité de générer, en une seule inférence, des cartes d'OCS, pour plusieurs type d'utilisateurs, En regroupant les éléments selon son besoin.

Ces produits utilisateurs peuvent aussi être directement générés à partir de notre architecture CNN. Si les inférences sur la CS de l'OCS-GE sont très encourageantes, celles de l'US montrent les limites de la télédétection sur les nomenclatures d'usages. Les attributs de densité d'urbanisation sont difficiles à déduire. Notre réseau est capable d'extraire des relations entre présence de bâtiments et densité mais comprend mal la notion de gradient de densité. Il est important de noter que notre architecture possède un champ réceptif restreint et n'est donc pas adaptée à ces problématiques de détection d'usage. L'augmentation de ce champ est une des pistes d'amélioration.

Enfin l'étude des capteurs nous a permis de montrer que l'ajout de THR est bénéfique pour l'ensemble des classes de nos nomenclatures. Il est évident au regard des métriques que l'ajout de d'information spectrale est bénéfique. L'exemple le plus flagrant étant sur les bâtiments industriels où l'apport du MS de S2 est bien plus discriminant. Concernant la classe *Cultures*, seul point de nos nomenclatures non détaillé, l'étude sur la VT issue d'OSO nous a montré que le capteur SPOT n'est pas idéal pour distinguer les types de cultures. Pour autant la comparaison entre SPOT S2 nous a montré que cette classe bénéficiera tout de même de la THR pour la délimitation de ses contours. La complémentarité entre SPOT et S2 est générale et leur fusion laisse envisager de fortes performances.

Ce chapitre présente les travaux effectués sur le sujet de la fusion. Il traite aussi bien la fusion tardive que précoce. Les modifications appliquées à l'architecture et les contraintes liées à cette tâche seront détaillées. Il est important de noter que ces travaux sont encore en cours et que les résultats présentés sont issus d'un algorithme prototype. Les conclusions restent donc provisoires.

5.1 Retour sur l'état de l'art

5.1.1 Absence d'architecture complètement convolutive

[Pelletier et al., 2019] a montré qu'il était possible de traiter les séries temporelles S2 à partir des CNN en utilisant les opérations convolutions 3D. En échange de scores surpassant ceux des algorithmes RNN, ils ont été obligés de corriger la présence des nuages sur les séries temporelles par interpolation. Partant de données déjà formatées selon les mêmes obligations, nous avons voulu étendre notre architecture entièrement convolutive vers le traitement bilatéral S2 - SPOT.

[Benedetti et al., 2018] ont investigué le traitement simultané S2 et SPOT en utilisant respectivement pour chaque entrée un RNN et un CNN, avant de combiner leurs *embeddings* par un MLP. Néanmoins cette dernière couche les force à utiliser une approche par fenêtre glissante.

Dans le même cadre [Gaetano et al., 2018] ont développé une architecture de fusion précoce entre les canaux MS et PAN de SPOT dans l'objectif de suppression des opérations de pansharpening. Chaque entrée est traitée avec un CNN distinct. Encore une fois, les *embeddings* sont fusionnées par un MLP, interdisant une approche entièrement convolutive.

Il n'existe donc pas, à notre connaissance de l'état de l'art actuel, d'architecture complètement convolutive capable de traiter la fusion. Or nous avons montré que :

1. les capteurs S2 et SPOT étaient complémentaires,
2. que les approches complètement convolutives étaient pertinentes dans l'exploitation de l'information spatiale apportée par la THR.

5.1.2 Avantages et contraintes de la nouvelle approche

La construction d'une architecture à deux entrées offre la possibilité d'étudier à la fois la fusion tardive¹ et la fusion précoce² : les cartes d'inférences ou de probabilité sont des images au même titre que les dalles SPOT ou S2.

Néanmoins une architecture entièrement convolutive apporte des contraintes dans le processus de traitement, puisque nous devons garantir :

- que les tailles des dalles soient constantes au sein de leur jeu de donnée respectif,
- que les dalles ne se superposent pas (dans le cadre d'une validation croisée),
- qu'il existe un point dans l'architecture où les tenseurs des sources ont les mêmes dimensions, en vue de leur concaténation
- que la représentation contenue dans ces tenseurs concerne la même emprise au sol au moment de la concaténation.

Or dans le cas où les sources sont les images des capteurs S2 et SPOT, leur ratio de résolution n'est pas entier (10m et 1,5m). Cela implique que pour éviter tout pré-traitements ou interpolations, il est impossible de valider, à la fois, les quarts points énoncés ci-dessus.

¹Fusion tardive : utilisation des résultats de fusion d'une SegSem précédente (carte de SegSem ou de probabilités)

²Fusion précoce : Utilisation d'images satellites en simultané, sans opération de ML au préalable

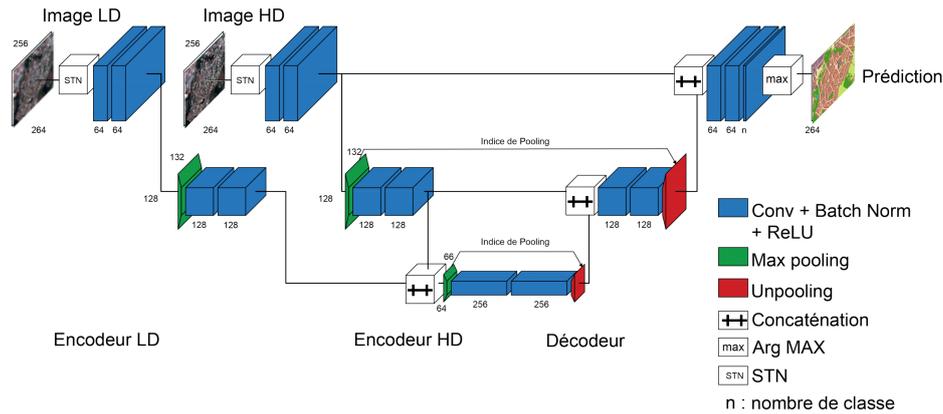


Figure 5.2 – Architecture SO-NET proposée pour la fusion avec le module *STN*

5.3 Résultats de fusion

Puisque notre architecture et nos données le permettent, nous avons étudié la fusion tardive et précoce à partir :

- des canaux PAN et MS pour éviter le pré-traitement de pansharpening (similairement à [Gaetano et al., 2018]),
- des images Multi-Spectrales pansharpenés de SPOT (PXS) et S2 *CC* pour valider les postulats sur la classification émis dans l'analyse de la configuration des capteurs.

Nous référerons les modèles de cartes de SegSem, de probabilité et de fusion précoce par, respectivement, "Tardif C", "Tardif P" et "Précoce".

5.3.1 Jeu de données

Nous avons utilisé les modèles PAN, MS, PXS, et S2 *CC*, entraînés dans le cadre de l'étude de la configuration capteur, pour générer les cartes de SegSem et de probabilités sur une nouvelle zone en commun entre les images SPOT et S2. Cela nous permet :

- d'éviter une contamination des cartes de SegSem et de probabilité par les dalles d'entraînement de la première zone,
- de s'assurer d'une quantité de VT suffisante pour l'entraînement de nos modèles de fusion.

5.3.2 Fusion des canaux SPOT

Nous avons étudié la fusion des canaux SPOT dans le but d'avoir un premier aperçu de notre architecture sur la fusion tardive et précoce, de mesurer sa capacité se passer du pansharpening et de mesurer le potentiel du traitement disjoint des canaux MS et PAN.

Résultats :

L'ensemble des traitements par fusion a permis d'obtenir de meilleurs scores que les traitements disjoints. Ces performances prouvent l'impact de la complémentarité des canaux de SPOT, même en fusion tardive.

Grâce aux images des inférences sur les zones de campagnes 5.3, que le réseau est capable, même à partir de cartes des SegSem très bruitées (1b) et (1a), de corriger la plupart des erreurs (2b). Nous remarquons que la végétation est définie plus précisément sur les inférences de la fusion précoce (2d), ce qui confirme le score sur les métriques du tableau 5.1.

Par analyse visuelle des images de la figure 5.4, nous constatons que le résultat d'inférence se rapprochant le plus de celui du modèle PXS (1b) est donné par le modèle tardif P. Si sur les bâtiments il est difficile d'effectuer une comparaison visuelle, le résultat des routes confirme les métriques du tableau 5.1.

La figure 5.5 nous montre :

Modèle	mIoU	OA	IoU					
			● Buffers	● Eau	● Cultures	● Routes	● Végétation	● Bâtiments
Tardif S	76,40	92,93	65,32	82,78	95,33	66,4	84,07	64,52
Tardif P	77,20	93,11	66,41	82,01	95,37	68,82	84,03	66,54
Précoce	77,22	93,31	65,35	85,28	95,77	66,06	85,4	65,46
PAN	68,23	90,95	61,9	61,17	92,98	50,48	82,48	60,37
MS	71,43	91,00	59,32	83,48	93,15	57,44	81,03	54,17
PXS	76,82	93,01	65,73	83,46	94,84	66,87	84,72	65,32

Table 5.1 – Comparatifs des métriques des méthodes de fusion des canaux SPOT

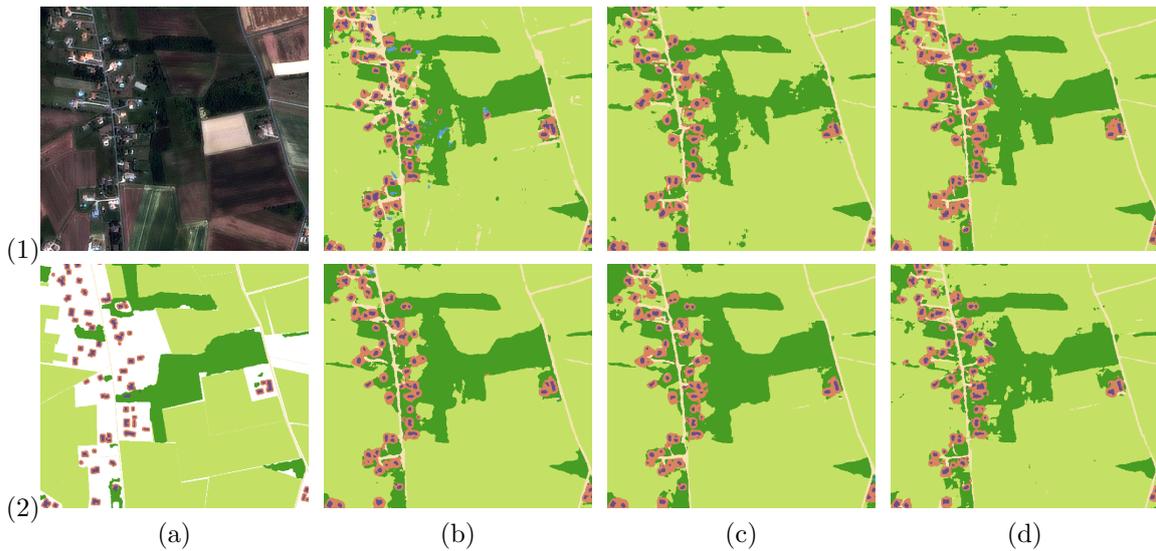


Figure 5.3 – Visuels des inférences sur zone de campagne pour la fusion des canaux SPOT
 (1a) image SPOT, (2a) VT, inférences (1b) PAN, (1c) MS, (1d) PXS, (2b) tardif C, (2c) Tardif P, (2d) Précoce

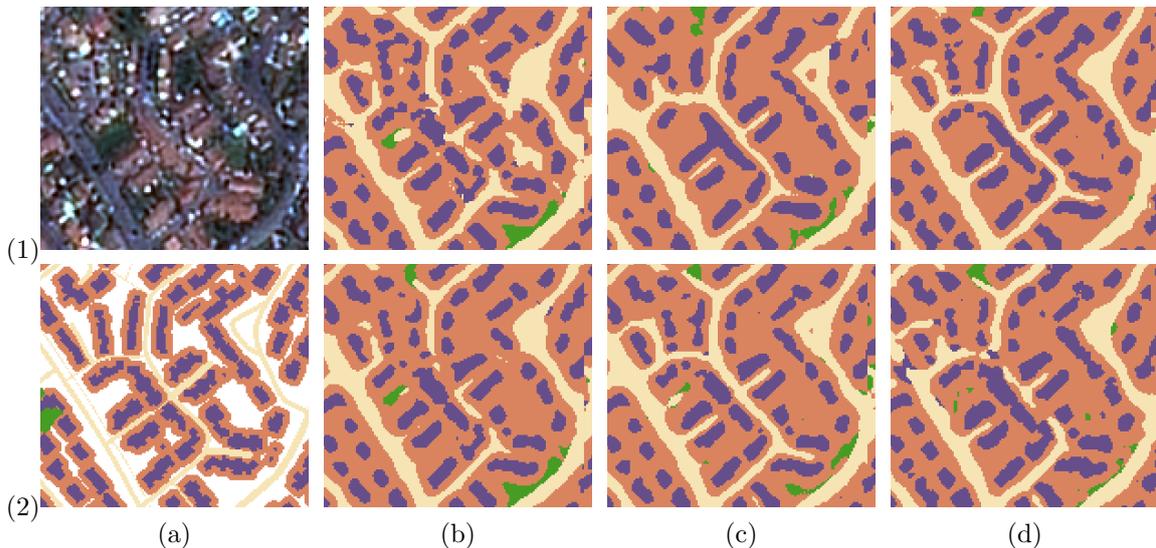


Figure 5.4 – Visuels des inférences sur zone pavillonnaire pour la fusion des canaux SPOT
 (1a) image SPOT, (2a) VT, inférences (1b) PAN, (1c) MS, (1d) PXS, (2b) tardif C, (2c) Tardif P, (2d) Précoce

- que l'espace de représentation du problème est déjà simplifié dans les cartes de SegSem et de probabilités, puisque, dès la 4^{ème} époque, les scores de leurs modèles sont déjà supérieurs à 65% de

mIoU alors qu'il est inférieur à 60% dans le cas des images,

- que les données contenues dans les cartes de SegSem sont trop pauvres pour sur-apprendre, contrairement aux deux autres,
- qu'à la vue de la croissance exponentielle du score de mIoU du TE du modèle tardif P, le réseau dispose de beaucoup trop de paramètres, et sur-apprend énormément.

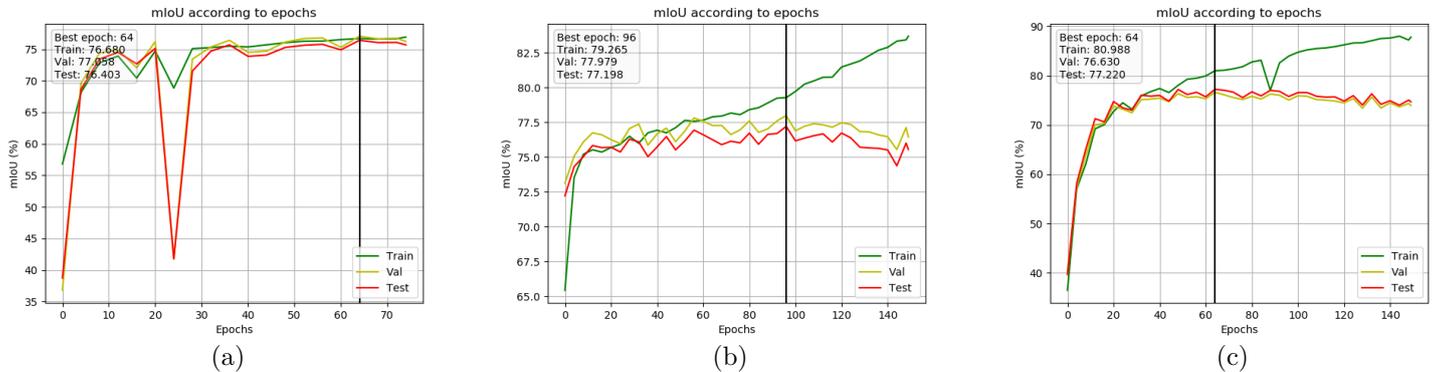


Figure 5.5 – Courbes des scores de mIoU au cours de l'entraînement des modèles de fusion (a) tardif C, (b) Tardif P, (c) Précoce

Conclusions : Il est étonnant de voir que l'apprentissage à partir de produits aussi simples que les cartes de SegSem puisse apporter des performances presque aussi efficaces que les images. Même s'il est logique que les scores des techniques tardives soient élevés, car leur traitement revient plus ou moins à "coller" deux réseaux l'un derrière l'autre.

Cela explique aussi pourquoi l'architecture est sur-dimensionnée pour la fusion tardive et qu'il serait mieux d'avoir des architectures distinctes entre fusion tardive et précoce.

Enfin, si la fusion de cartes de probabilités issues de capteurs différents permet un apprentissage aussi rapide et aussi performant que le traitement simultané (PXS), alors il est complètement envisageable de produire des cartes de probabilité pour chaque classe et chaque capteur. La génération de produits opérationnels gagnerait en modularité, puisqu'il suffirait de "piocher" dans résultats de cartes de probabilité pour les reconstruire.

5.3.3 Fusion SPOT S2

Ce premier essai de fusion entre les capteurs SPOT doit pouvoir nous donner un aperçu de la validité des postulats de notre analyse de capteurs et confirmer les comportements des fusions tardives et précoces observés dans le test précédent.

Résultats

Sur le tableau 5.2, hormis sur la classe *Végétation*, les approches de fusion affichent toutes de meilleures performances globales que les traitements individuels. Seul l'IOU de la classe *Bâtiments* du modèle Précoce n'affiche pas une métrique qualifiable de "supérieure". Même si on constate visuellement une meilleure reconstruction des bâtiments industriels grâce à la figure 5.6, ce qui confirme que le bénéfice de S2 pour ce type d'objets topographiques.

	mIoU	OA	IoU					
			● Buffers	● Eau	● Cultures	● Routes	● Végétation	● Bâtiments
classif	77,86	93,19	66,57	84,68	95,33	69,02	84,12	67,42
proba	78,59	93,46	66,1	86,1	95,67	71,25	84,43	68
image	77,73	93,46	66,93	84,86	95,79	68,08	85,36	65,34
S2pur	64,11	88,72	49,09	80,08	92,28	46,57	79,05	37,58
PXS_pur	76,82	93,01	65,73	83,46	94,84	66,87	84,72	65,32

Table 5.2 – Caption

Nous notons aussi que si les modèles tardifs ont délimité avec efficacité le bâtiment de la figure 5.6 dans la classe *Bâtiments*, le modèle Précoce le délimite en tant que *Buffers*. Ces observations corroborent avec les métriques en faveur des approches tardives.

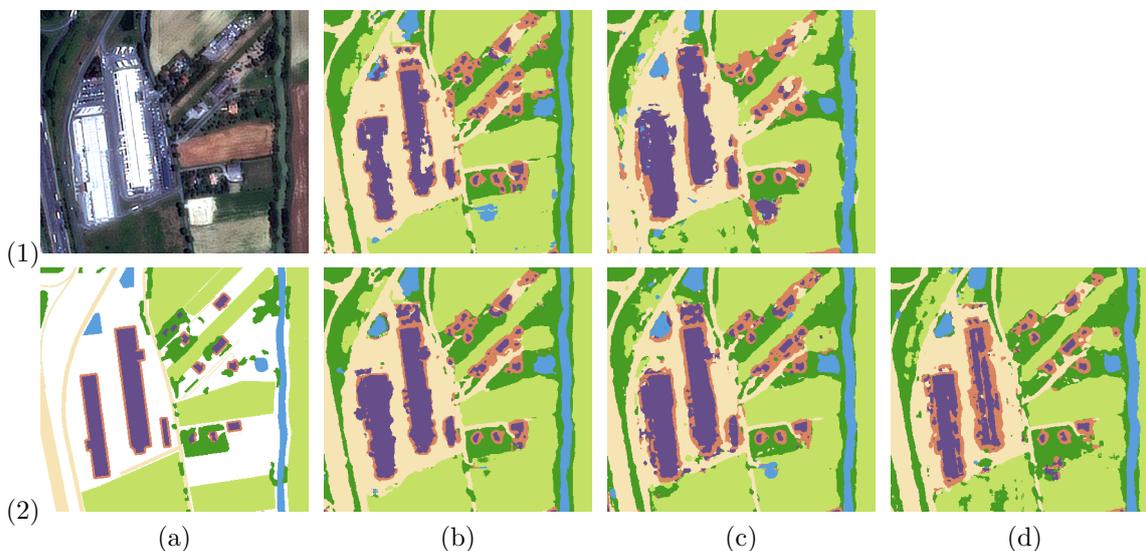


Figure 5.6 – Visuels des inférences sur zone industrielle pour la fusion SPOT et S2
(1a) image SPOT, (2a) VT, inférences (1b) PXS, (1c) S2, (2b) tardif C, (2c) Tardif P, (2d) Précoce

La fusion tardive de cartes de probabilités affiche, comme dans le test précédent, le meilleur score d'IoU sur les routes. La figure 5.8 nous confirme la meilleure reconstruction des routes de ce modèle. C'est la combinaison des gains sur les classes *Bâtiment* et *Routes* qui lui permet d'obtenir le meilleur mIoU.

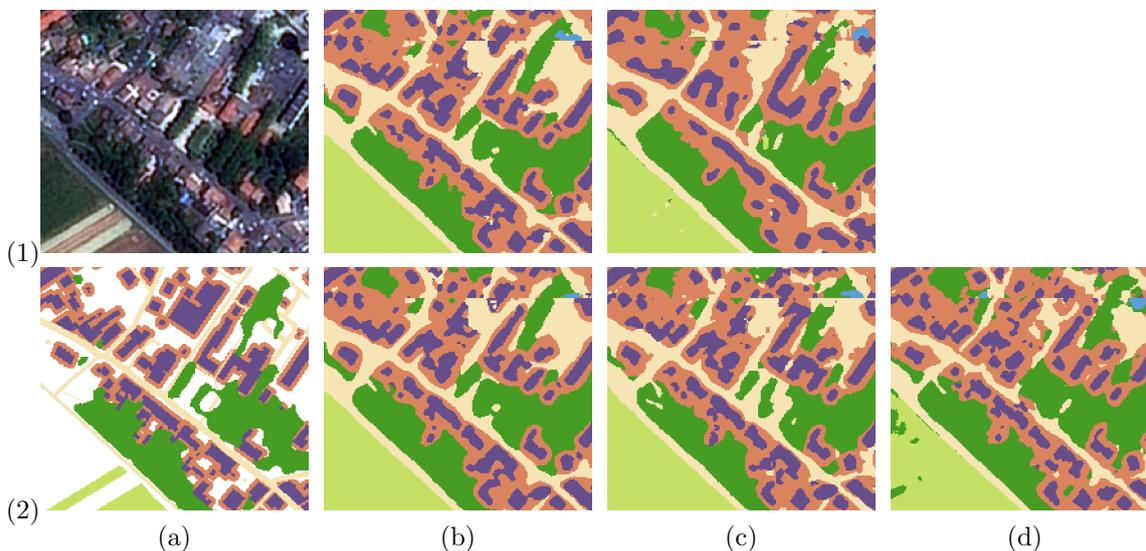


Figure 5.7 – Visuels des erreurs d'inférences pour la fusion SPOT et S2
(1a) image SPOT, (2a) VT, inférences (1b) PXS, (1c) S2, (2b) tardif C, (2c) Tardif P, (2d) Précoce

Néanmoins du fait de notre VT discontinue, les modèles tardifs P et C génèrent certaines erreurs que le modèle précoce ne présente pas. La figure 5.7 présente un exemple de matériau minéral non annoté par la VT que le modèle précoce, ayant accès aux images a pu classer comme de la route tandis que les modèles tardifs ont classé cet élément comme bâtiment.

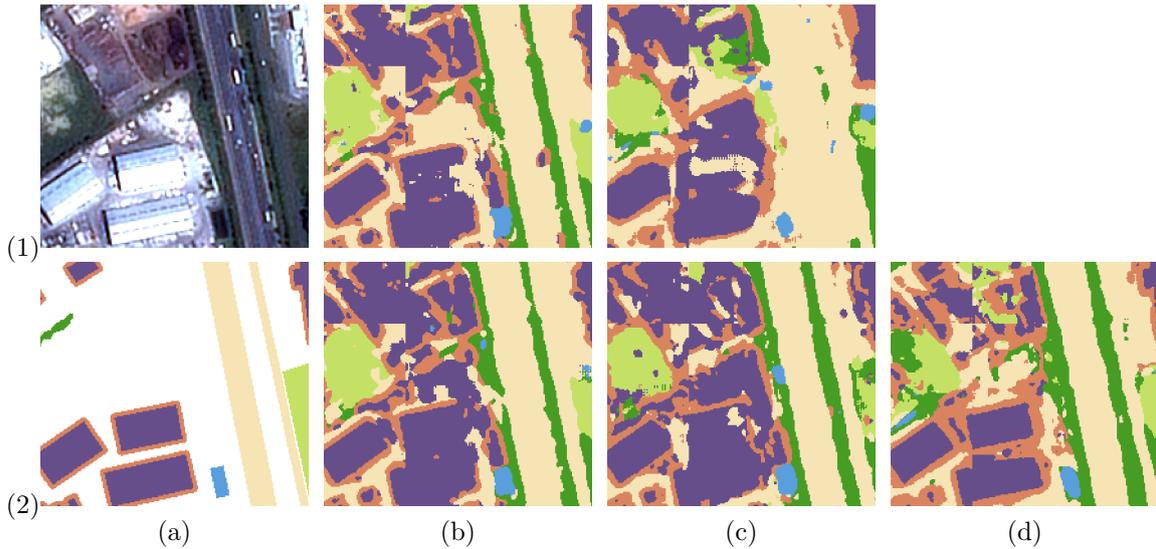


Figure 5.8 – Visuels des inférences sur zone pavillonnaire pour la fusion SPOT et S2
 (1a) image SPOT, (2a) VT, inférences (1b) PXS, (1c) S2, (2b) tardif C, (2c) Tardif P, (2d) Précoce

Conclusion : Cette première tentative de fusion entre les images des capteurs S2 et SPOT nous permet de confirmer la complémentarité des deux capteurs, au moins sur les classes de bâtiments, quelque soit l’approche de fusion utilisée.

Il est encore difficile de conclure sur l’approche la plus efficace, puisque d’une part, nos implantations ne sont pas optimales et d’autre part qu’il existe des dissonances entre résultats visuels et métriques qui rendent l’interprétation difficile.

Concernant le modèle de cartes de SegSem, nous retrouvons bien le même comportement que pour le test précédent. En effet, bien que les deux capteurs produisent des cartes bruitées, la reconstruction finale des objets topographique est satisfaisante. Nous expliquons ce phénomène par le fait que les CNN soient basés sur la reconnaissance de structures et de motifs, et donc, que le modèle tardif C arrive à détecter les zones où son "prédécesseur" a été hésitant.

5.3.4 Module de transformation spatiale

L’ajout du module *STN* n’a été efficace que pour la fusion précoce. Nous avons remarqué que les gains sur la fusion précoce sont principalement réalisés sur les classes *Routes*, *Bâtiments* et *Cultures*. Ce sont d’une part des classes qui avaient montré de forts gains sur l’ajout de THR dans nos tests précédents. Elles sont donc sensibles à la précision de localisation de l’information. D’autre part, ce sont également des classes qui bénéficiaient de l’ajout de MS (*Bâtiments* avec les bâtiments industriels, *Routes* pour les parkings, *Cultures* pour les confusions avec l’eau et les routes). Les métriques des modèles de fusion tardive ont perdu respectivement 0.72% et 0.52% de mIoU. Ce phénomène s’explique par le simple fait les cartes soient trop grossières pour réellement bénéficier du recalage spatial qu’apporte ce module. Il n’apporte de ce fait que du bruit supplémentaire et complique l’apprentissage, ce qui se répercute sur les scores finaux.

	mIoU	OA	IoU					
			● Buffers	● Eau	● Cultures	● Routes	● Végétation	● Bâtiments
Sans	77,73	93,46	66,93	84,86	95,79	68,08	85,36	65,34
Avec	78,12	93,69	66,91	84,24	95,99	69,21	86,11	66,23

Table 5.3 – Caption

Pour autant l’augmentation de score que montre le modèle précoce à l’ajout du module *STN* ne permet pas d’atteindre les scores du modèle Tardif P. Or, les deux modèles partent des mêmes sources originelles de données. Cela indique que l’architecture n’est pas optimale pour le traitement des images.

5.4 travaux futurs

La suite de nos travaux sur la fusion visera à :

- Créer deux architectures distinctes et respectivement adaptées à la fusion tardive et précoce (fig 5.9),
- Rendre l'architecture plus souple sur les résolutions des différentes sources. Jusqu'à maintenant les deux sources devaient obligatoirement posséder la même résolution, impliquant donc des sur-échantillonnages,
- Utiliser la série temporelle complète de S2. L'image utilisée étant uniquement une synthèse du mois de juillet,
- Utiliser différentes sources de classification. Nos tests se sont limités à des inférences issues de nos CNN, mais idéalement nous aimerions étudier la fusion de cartes issues de RF comme le produit OSO,
- Étendre les nomenclatures, pour mieux valider la complémentarité S2 et SPOT. En l'occurrence sur la nomenclature OSO.

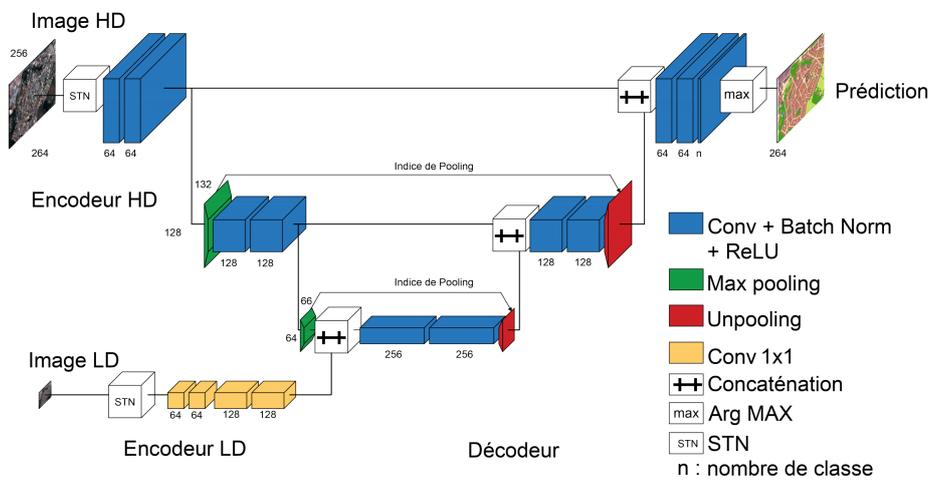


Figure 5.9 – Architecture SO-NET proposée pour la fusion précoce

Conclusion

Nos travaux concernant le projet CES Urbain du Pôle Théia nous ont permis de montrer que la THR de SPOT 6/7 utilisée conjointement avec une architecture CNN entièrement convolutive est adaptée à la détection d'objets urbains fins. Ces performances pouvant permettre l'instanciation d'éléments de diamètre inférieur à 5m, aussi bien pour la végétation que les bâtiments. Nous avons mis en avant qu'une technique de création de VT automatisée à partir ML est parfaitement envisageable puisque la généralisation des CNN corrige le bruit présent dans la VT. Il reste cependant à vérifier si la détection d'objet fins est suffisamment généralisable à :

1. d'autres environnements urbains que celui vu au cours de l'entraînement,
2. à des environnements non urbains.

Nous avons mis en évidence que les CNN étaient performants pour la détection de tâches urbaines grâce à la THR de SPOT 6/7, bien que notre architecture actuelle ne permette pas de conclure sur leur capacité à dériver des représentations discriminantes pour l'US. En revanche, concernant la CS, sa performance est incontestable.

Ayant dévoilé l'apport de la donnée S2 pour discernement de certaines classes urbaines, nous préconisons l'utilisation conjointe des deux capteurs pour améliorer la SegSem des milieux urbains.

Cet ensemble de conclusions concerne également le projet PARCELLE, dans le cadre de l'amélioration du produit OSO, puisque la détection des classes urbaines constitue son point faible actuel. En effet, l'ajout de THR permet la correction des confusions entre classes urbaines et cultures, même dans le cas extrêmement contraint de notre étude (apprentissage à partir du produit final OSO). Plus généralement, toutes les classes bénéficient de l'apport de la THR. Les cultures en profitent pour

1. leur délimitation,
2. la lecture de leur texture qui s'avère être discriminantes pour certains types (ex. vignes).

Si notre étude n'a pas amélioré la détection des routes du produit OSO, du fait de nos contraintes de tests, il ne fait pas de doute, à la vue des résultats d'apprentissage sur une VT plus précise, que l'ajout de THR permet d'en retrouver la structure, même en zone dense.

Les travaux préliminaires sur la fusion permettent de confirmer que la complémentarité de S2 et SPOT bénéficie bien à la SegSem. La fusion est néanmoins tributaire de la précision et de la richesse de la VT. Néanmoins, nos travaux ne permettent pas encore de conclure sur la meilleure approche de fusion, puisque l'architecture doit être adaptée au type de fusion utilisé (tardive ou précoce).

Enfin, nous avons mis en évidence le possible gain en modularité que la génération de carte d'OCS pourrait bénéficier :

- par la propriété de constance de la SegSem produite par les CNN, qui permet d'entraîner un modèle à partir d'une nomenclature riche, puis de modifier celle des inférences (en fonction de sa hiérarchie) pour correspondre aux besoins utilisateurs.
- grâce à la fusion de cartes de probabilités. Nos tests ont montré que les entraînements de modèles à partir de ces cartes sont extrêmement rapides et nécessitent une architecture bien plus légère que pour un entraînement à partir d'images. Il est donc possible de créer une BD de cartes de

probabilités pour chaque classe de nomenclature et chaque capteur, puis de générer un produit personnalisé à l'utilisateur en sélectionnant les probabilités liés à son besoin.

Bibliographie

- [Astrium, 2019] Astrium (2019). Spot 6 | spot 7 fiche technique.
- [Benedetti et al., 2018] Benedetti, P., Ienco, D., Gaetano, R., Ose, K., Pensa, R. G., and Dupuy, S. (2018). M3fusion : A deep learning architecture for multi- $\{Scale/Modal/Temporal\}$ satellite data fusion. *CoRR*, abs/1803.01945.
- [Breiman, 2001] Breiman, L. (2001). Random forests. *Machine learning*, 45(1) :5–32.
- [CESBIO, 2019] CESBIO (2019). Infrastructure pour l’occupation des sols par traitement automatique incorporant les orfeo toolbox applications.
- [CESOSO, 2019] CESOSO (2019). Carte d’occupation des sols du ces oso sur la france métropolitaine.
- [ESA, 2019] ESA (2019). Multispectral instrument (msi) overview.
- [Gaetano et al., 2018] Gaetano, R., Ienco, D., Ose, K., and Cresson, R. (2018). A Two-Branch CNN Architecture for Land Cover Classification of PAN and MS Imagery. *Remote Sensing*, 10(11) :1746.
- [Girshick et al., 2013] Girshick, R. B., Donahue, J., Darrell, T., and Malik, J. (2013). Rich feature hierarchies for accurate object detection and semantic segmentation. *CoRR*, abs/1311.2524.
- [Glorot and Bengio, 2010] Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256.
- [Hazırbaşı et al., 2016] Hazırbaşı, C., Ma, L., Domokos, C., and Cremers, D. (2016). Fusetnet : Incorporating depth into semantic segmentation via fusion-based cnn architecture.
- [He et al., 2015a] He, K., Zhang, X., Ren, S., and Sun, J. (2015a). Deep residual learning for image recognition. *CoRR*, abs/1512.03385.
- [He et al., 2015b] He, K., Zhang, X., Ren, S., and Sun, J. (2015b). Delving deep into rectifiers : Surpassing human-level performance on imagenet classification. *CoRR*, abs/1502.01852.
- [IGN, 2019a] IGN (2019a). Bd topo.
- [IGN, 2019b] IGN (2019b). Bd topo, descriptif de contenu.
- [Inglada et al., 2017] Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., and Rodes, I. (2017). Operational high resolution land cover map production at the country scale using satellite image time series. *Remote Sensing*, 9(1).
- [International Institute of Applied Systems Analysis, 2019] International Institute of Applied Systems Analysis (2019). Landsense.
- [Ioffe and Szegedy, 2015] Ioffe, S. and Szegedy, C. (2015). Batch normalization : Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167.
- [Jaderberg et al., 2015] Jaderberg, M., Simonyan, K., Zisserman, A., and Kavukcuoglu, K. (2015). Spatial transformer networks. *CoRR*, abs/1506.02025.
- [James Dellinger, 2019] James Dellinger (2019). Weight initialization in neural networks : A journey from the basics to kaiming.

- [Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam : A method for stochastic optimization. *arXiv preprint arXiv :1412.6980*.
- [Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and E. Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 25.
- [LeCun et al., 1995] LeCun, Y., Bengio, Y., et al. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10) :1995.
- [Noh et al., 2015] Noh, H., Hong, S., and Han, B. (2015). Learning deconvolution network for semantic segmentation. *CoRR*, abs/1505.04366.
- [Pelletier et al., 2019] Pelletier, C., Webb, G. I., and Petitjean, F. (2019). Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, 11(5).
- [Postadjian et al., 2018a] Postadjian, T., Bris, A. L., Mallet, C., and Sahbi, H. (2018a). Domain adaptation for large scale classification of very high resolution satellite images with deep convolutional neural networks. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*.
- [Postadjian et al., 2018b] Postadjian, T., Bris, A. L., Mallet, C., and Sahbi, H. (2018b). Superpixel partitioning of very high resolution satellite images for large-scale classification perspectives with deep convolutional neural networks. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1328–1331.
- [Postadjian et al., 2017] Postadjian, T., Le Bris, A., Sahbi, H., and Mallet, C. (2017). Investigating the potential of deep neural networks for large-scale classification of very high resolution satellite images. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-1/W1 :183–190.
- [PyTorch, 2019] PyTorch (2019). Pytorch site.
- [QGIS, 2019] QGIS (2019). Qgis,système d’information géographique libre et open source.
- [Roger, 2006] Roger, TRIAS-SANZ etMarc, P.-D. e. D. B. e. G. S. e. J. L. (2006). Mise à jour automatique de données d’occupation du sol à grande échelle par des méthodes d’analyse d’image.
- [Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net : Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597.
- [Ruder, 2016] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *CoRR*, abs/1609.04747.
- [Sabo et al., 2018] Sabo, F., Corbane, C., Florczyk, A. J., Ferri, S., Pesaresi, M., and Kemper, T. (2018). Comparison of built-up area maps produced within the global human settlement framework. *Transactions in GIS*, 22(6) :1406–1436.
- [Sik-Ho Tsang, 2018] Sik-Ho Tsang (2018). Review : Deconvnet — unpooling layer (semantic segmentation).
- [Theia, 2019] Theia (2019). Theia-land.
- [URCLIM, 2019] URCLIM (2019). Urban climate services.
- [Wendl et al., 2018] Wendl, C., Le Bris, A., Chehata, N., Puissant, A., and Postadjian, T. (2018). Decision fusion of spot6 and multitemporal sentinel2 images for urban area detection. pages 1734–1737.
- [Will Koehrsen, 2018] Will Koehrsen (2018). Neural network embeddings explained.
- [Yao et al., 2015] Yao, F., Wang, C., Di, D., Luo, J., Shen, Z., and Yang, K. (2015). High-resolution mapping of urban surface water using zy-3 multi-spectral imagery. *Remote Sensing*, 7 :12336–12355.
- [Zhang et al., 2019] Zhang, C., Wei, S., Ji, S., and Lu, M. (2019). Detecting large-scale urban land cover changes from very high resolution remote sensing images using cnn-based classification. *ISPRS International Journal of Geo-Information*, 8(4).

Table des figures

3.1	Architecture U-NET, [Ronneberger et al., 2015]	12
3.2	Architecture SO-NET proposée	13
3.3	Comparatifs visuels entre les inférences de la CV1	18
3.4	Comparatifs visuels selon le nombre de convolutions par étage	20
3.5	Comparatifs visuels selon la taille des noyaux de convolution	21
3.6	Comparatifs de l'évolution de la Loss au cours des époques (epochs) avec et sans BN	22
3.7	Comparatifs visuels avec et sans CR	22
3.8	Comparatifs visuels selon la taille des noyaux de convolution	23
4.1	Visuels des inférences sur zone de campagne des algorithmes RF et CNN	26
4.2	Visuels des inférences sur entre notre approche et celle de [Postadjian et al., 2017]	27
4.3	Visuels des inférences sur zone de campagne des algorithmes RF et CNN	28
4.4	Visuels des cartes de probabilités des approches par nomenclature complète et <i>binnaire</i>	28
4.5	Visuels des inférences entre transfert direct, initialisation aléatoire et FT	30
4.6	Courbes des scores de mIoU au cours de l'entraînement des modèles pour l'étude du FT	30
4.7	Visuels des inférences entre les différentes tailles de <i>Buffers</i>	31
4.8	Visuels des inférences entre seuillage de cartes de probabilités, nomenclature A6 et A5	32
4.9	Visuels des inférences entre seuillage de cartes de probabilités, nomenclature A6 et A5	32
4.10	Visuels des inférences pour le premier niveau de raffinement, sur cimetière	33
4.11	Visuels des inférences pour le premier niveau de raffinement, sur aéroport	34
4.12	Matrice de confusion pour la SegSem à 13 classes	34
4.14	Visuels des inférences pour le 2 ^{ème} niveau de raffinement, en zone de campagne	35
4.15	Visuels des inférences pour le 2 ^{ème} niveau de raffinement, sur la végétation à forte texture	35
4.13	Matrice de confusion pour la SegSem à 18 classes	36
4.16	Visuels des inférences transférées en nomenclatures A6	36
4.17	Création d'une pseudo-VT	37
4.18	Comparatif visuel des inférences avec et sans VT des classes de végétation urbaine issu d'un RF	38
4.19	Comparatif visuel des inférences selon la taille des <i>Buffers</i> avec VT des classes de végétation urbaine issues d'un RF	38
5.1	Architecture SO-NET proposée pour la fusion	42
5.2	Architecture SO-NET proposée pour la fusion avec le module <i>STN</i>	43
5.3	Visuels des inférences sur zone de campagne pour la fusion des canaux SPOT	44
5.4	Visuels des inférences sur zone pavillonnaire pour la fusion des canaux SPOT	44
5.5	Courbes des scores de mIoU au cours de l'entraînement des modèles de fusion	45
5.6	Visuels des inférences sur zone industrielle pour la fusion SPOT et S2	46
5.7	Visuels des erreurs d'inférences pour la fusion SPOT et S2	46
5.8	Visuels des inférences sur zone pavillonnaire pour la fusion SPOT et S2	47
5.9	Architecture SO-NET proposée pour la fusion précoce	48

Liste des tableaux

1.1	Caractéristiques des instruments sur Sentinel-2	4
1.2	Caractéristiques des instruments SPOT 6 et 7	4
2.1	Datation des données	8
2.2	Récapitulatif des images pré-traitées	8
2.3	Cas d’usages des nomenclatures	9
2.4	Raffinement progressif des Nomenclatures classiques	10
2.5	Nomenclatures mixtes sur la densité d’urbanisation	11
3.1	Taille des <i>batches</i> en fonction de la taille des dalles	14
3.2	Partitionnement du GS pour la validation croisée	15
3.3	Métriques d’entraînement et leur périodicité d’acquisition	16
3.4	Moyennes (μ) et Écarts types (σ) des métriques de la CV1 selon la nomenclature A6 pour différents entraînements sur les mêmes données	18
3.5	Intervalles de confiances des métriques(%)	19
3.6	Composition de la VT pour l’étude de l’architecture	19
3.7	Métriques selon le nombre de convolutions par étage (%)	20
3.8	Métriques selon la taille des noyaux de convolution (%)	21
3.9	Métriques avec et sans CR	22
3.10	Métriques en fonction du nombre d’étages de l’architecture	23
4.1	Comparatifs des métriques entre RF et CNN	25
4.2	Comparatifs des métriques entre approche par fenêtre glissante et par approche entièrement convolutive	26
4.3	Comparatifs des métriques entre approche à nomenclature complète et approche <i>binnaire</i>	28
4.4	Comparatifs des métriques sur les stratégies d’apprentissage	29
4.5	Comparatifs des métriques sur les stratégies d’apprentissage	31
4.6	Comparatifs des métriques entre les différentes tailles de <i>Buffers</i>	31
4.7	Comparatifs des métriques globales entre les modèles entraînés selon les nomenclatures A6 et H13	33
4.8	Comparatifs des IoU entre les modèles entraînés selon les nomenclatures A6 et H13	33
4.9	Comparatifs des métriques globales entre les modèles entraînés selon les nomenclatures A6 et H18	34
4.10	Comparatifs des métriques d’IoU entre les modèles entraînés selon les nomenclatures A6 et H18	35
4.11	Comparatifs des métriques selon A6 des modèles entraînés selon A6, H13 et H18	36
4.12	Comparatif des métriques selon la variation de la taille des <i>Buffers</i>	38
5.1	Comparatifs des métriques des méthodes de fusion des canaux SPOT	44
5.2	Caption	45
5.3	Caption	47

Annexes

A	Détails de nomenclatures OCS-GE et UA	59
---	---	----

DÉTAILS DE NOMENCLATURES OCS-GE ET UA

ANNEXE **A**

Couverture du sol							
CS1. Sans végétation				CS2. Avec végétation			
CS1.1 Surfaces anthropisées		CS1.2 Surfaces naturelles		CS2.1 Végétation ligneuse		CS2.2 Végétation non ligneuse	
CS1.1.1 Zones imperméables	CS1.1.2 Zones perméables	CS1.2.1 Sols nus (Sable, pierres meublées, rochers saillants, ...)	CS1.2.2 Surfaces d'eau (Eau continentale et maritime)	CS2.1.1 Formations arborées	CS2.1.2 Formations arbustives et sous-arbustives (Landes basses, formations arbustives, formations arbustives organisées, ...)	CS2.1.3 Autres formations ligneuses (Vignes et autres lianes)	CS2.2.1 Formations herbacées (Pelouses et prairies, terres arables, roseaux, ...)
		CS1.2.3 Névés et glaciers			CS2.2.2 Autres formations non ligneuses (Lichen, mousses, bananiers, bambous, ...)		
CS1.1.1.1 Zones bâties (Routes, places, parking, ...)	CS1.1.1.2 Zones à matériaux minéraux (Pierres, terre, coques, terres, pâtes, bouillottes, chemins empierrés, canalisations, salines, ...)	CS1.1.2.1 Zones à matériaux minéraux (Pierres, terre, coques, terres, pâtes, bouillottes, chemins empierrés, canalisations, salines, ...)	CS1.1.2.2 Zones à autres matériaux composites (Pôles, ...)	CS2.1.1.1 Peuplement de feuillus	CS2.1.1.2 Peuplement de conifères	CS2.1.1.3 Peuplement mixte	

- LC**
- CS1.1.1.1
 - CS1.1.1.2
 - CS1.1.2.1
 - CS1.1.2.2
 - CS1.2.1
 - CS1.2.2
 - CS2.1.1.1
 - CS2.1.1.2
 - CS2.1.1.3
 - CS2.1.2
 - CS2.1.3
 - CS2.2.1

Usage du sol		LU		
US1. Production primaire	US1.1 Agriculture			
	US1.2 Sylviculture			
	US1.3 Activités d'extraction			
	US1.4 Pêche et aquaculture			
	US1.5 Autre			
	US235 Production secondaire, tertiaire et usage résidentiel (regroupement des US2, US3 et US5 de la nomenclature nationale)			
	US4. Réseaux de transport logistiques et infrastructures	US4.1 Réseaux de transport	US4.1.1 Routier	
			US4.1.2 Ferré	
			US4.1.3 Aérien	
			US4.1.4 Eau	
			US4.1.5 Autres réseaux de transport	
		US4.2 Services de logistique et de stockage		
		US4.3 Réseaux d'utilité publique		
	US6 Autre usage	US6.1 Zones en transition		
		US6.2 Zones abandonnées		
US6.3 Sans usage				
US6.6 Usage Inconnu				

- LU**
- US1.1
 - US1.2
 - US1.3
 - US235
 - US4.1.1
 - US4.1.2
 - US4.1.3
 - US4.1.4
 - US4.2
 - US4.3
 - US6.1
 - US6.2
 - US6.3

Urban Atlas Land Use/Land Cover			
UA No.	Code	Nomenclature	Additional Information
1		Artificial surfaces	
1.1		Urban Fabric	
1.1.1	11100	Continuous urban fabric (S.L. > 80%)	HRL IMD required
1.1.2		Discontinuous Urban Fabric (S.L. 10% - 80%)	
1.1.2.1	11210	Discontinuous dense urban fabric (S.L. 50% - 80%)	HRL IMD required
1.1.2.2	11220	Discontinuous medium density urban fabric (S.L. 30% - 50%)	HRL IMD required
1.1.2.3	11230	Discontinuous low density urban fabric (S.L. 10% - 30%)	HRL IMD required
1.1.2.4	11240	Discontinuous very low density urban fabric (S.L. < 10%)	HRL IMD required
1.1.3	11300	Isolated structures	
1.2		Industrial, commercial, public, military, private and transport units	
1.2.1	12100	Industrial, commercial, public, military and private units	zoning data / field check recommended
1.2.2		Road and rail network and associated land	COTS or OSM data required
1.2.2.1	12210	Fast transit roads and associated land	COTS or OSM data required
1.2.2.2	12220	Other roads and associated land	COTS or OSM data required
1.2.2.3	12230	Railways and associated land	COTS or OSM data required
1.2.3	12300	Port areas	zoning data / field check recommended
1.2.4	12400	Airports	zoning data / field check recommended
1.3		Mine, dump and construction sites	
1.3.1	13100	Mineral extraction and dump sites	
1.3.3	13300	Construction sites	
1.3.4	13400	Land without current use	