



HAL
open science

Identification et classification d'espèces d'arbres non-natifs par réseau de neurones

Nathan Herrig

► **To cite this version:**

Nathan Herrig. Identification et classification d'espèces d'arbres non-natifs par réseau de neurones. Géographie. 2021. dumas-03359537

HAL Id: dumas-03359537

<https://dumas.ccsd.cnrs.fr/dumas-03359537v1>

Submitted on 30 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Identification et classification d'espèces d'arbres non-natifs par réseau de neurones

HERRIG Nathan



Source: BORGNIET Laurent, INRAE.

Jury:

- **Sylvain Bigot** : Enseignant-chercheur à l'Université Grenoble-Alpes et à l'IGE.
- **Pauline Dusseux** : Enseignante-chercheure à l'Université Grenoble-Alpes et à l'UMR Pacte.

Soutenance: **Le 9 septembre 2021** à l'IUGA

Stage: **INRAE** du 01/04/2021 au 31/08/2021

Maitre de stage: **Laurent Borgniet**

Lieu du stage : **INRAE - LESSEM**

Année universitaire 2020-2021

INRAE

Interreg
Alpine Space
ALPTREES
European Regional Development Fund



UGA
Université
Grenoble Alpes



GEODES
GEOGRAPHIE INFORMATION INTERFACE
DURABILITÉ ENVIRONNEMENTS

Notice bibliographique

Projet de Fin d'Études Master GAED - GEOÏDES

Auteur : HERRIG Nathan

Titre du Projet de Fin d'Études : Identification et classification d'espèces d'arbres non-natifs par réseau de neurones

Date de soutenance : 09/09/2021

Organisme d'affiliation : Institut d'Urbanisme et de Géographie Alpine de l'Université Grenoble Alpes

Organisme dans lequel le stage a été effectué : INRAE Grenoble

Directeur du Projet de Fin d'Études : Mr. Laurent BORGNIET

Collation : Nombre de pages : 61 / Nombre d'annexes : 4 / Nombre de références bibliographiques : 51 / Nombre de figures : 24

Mots-clés analytiques : CNN, SIG, Télédétection, NNT

Mots-clés géographiques : Grenoble, Berges de l'Isère, Ile Falcon

Remerciements

Je voudrais commencer par remercier mon maître de stage Laurent Borgniet qui m'a proposé un sujet des plus intéressants et qui m'a accueilli avec la plus grande des bienveillances. Je le remercie également pour l'accompagnement fourni tout au long du stage, pour les connaissances qu'il m'a apportées et pour son aide immédiate à chaque fois demandée. Enfin je voudrais le remercier, lui et toute l'équipe pour la bonne humeur permanente qui règne dans l'unité LESSEM, ce fut une expérience forte agréable que j'espère pouvoir renouveler.

Ma gratitude va également à Vincent Breton qui s'est tenu disponible pour la récolte de données floristiques auprès de ses partenaires, mais également pour le temps qu'il m'a consacré lors de la récolte des données terrain, il m'a permis d'apprendre l'usage des GPS de haute précision ainsi que la reconnaissance de nombreux végétaux.

J'adresse aussi mes remerciements à Eric Maldonado qui s'est investi à fond dans le projet pour toute la partie informatique complexe. Il a été d'une réactivité sans pareil et a lancé la quasi-totalité des tests de Machine Learning qui seront évoqués ci-dessous. Je le remercie également pour avoir fait marcher ses contacts au sein de l'UGA pour me permettre d'accéder au Master CCI (compétences complémentaires en informatique) qui va me permettre dans le courant de l'année à venir d'approfondir et d'améliorer le projet.

Un grand merci à Pauline Dusseux, mon enseignante référente qui m'a accompagné tout au long de mon stage, me demandant des nouvelles et me fournissant de précieux conseils, merci également pour avoir contacté Mr Genoud pour le Master CCI. Et enfin merci également pour la relecture de mon mémoire.

Mes remerciements vont également aux enseignants chargés du Master GEOIDES Sylvain Bigot et Sandra Rome qui, pour l'ensemble de leurs élèves, fournissent un suivi régulier qui a été très apprécié par l'ensemble de la promotion à laquelle j'ai eu la chance de faire partie. Cela permet une véritable qualité d'études, et particulièrement en ces temps difficiles pour l'ensemble des étudiants.

Enfin je porte mes derniers remerciements à mes amis et à ma famille qui m'ont accompagné et m'ont aidé à surmonter les épreuves que la vie place sur notre chemin.

Plan

-Liste des abréviations	6
<u>I. Introduction</u>	
1.1 Contexte du projet ALPTREES	7
1.2 Objectifs du projet ALPTREES	8
1.3 Problématique	9
<u>II. Etat de l'art</u>	
2.1 Sciences de l'information Géographique et végétation	11
2.2 Intelligence Artificielle, Machine Learning et Réseaux de Neurones	12
2.3 La place des Machine Learning et Réseaux de Neurones en Sciences de l'information Géographique	13
<u>III. Données et méthode</u>	
3.1 Les sites d'étude	15
3.2 Les espèces retenues	16
3.3 Les données mobilisées et produites	19
3.4 Classifications	23
3.4.1 Classifications pixel	24
3.4.2 Classifications objet	26
3.5 Fonctionnement des Réseaux de Neurones	29
3.6 Récolte et préparation des données	31
3.7 Choix du Réseau de Neurones	34
3.8 Traitements Réseaux de Neurones	34
<u>IV. Résultats et discussions</u>	
4.1 Analyse des résultats	36
4.2 Quelle suite pour l'outil?	38
4.3 Quelles pistes pour l'améliorer?	39
4.4 Les limites de l'outil	40
<u>V. Conclusion</u>	42
-Bibliographie	43
-Table des figures	48
-Annexes	50

Liste des abréviations:

- **CNN:** Convolutional Neural Network (réseaux de neurones convolutifs)
- **IA:** Intelligence Artificielle
- **INRAE:** Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement.
- **ML:** Machine Learning
- **NB :** Nota Bene
- **NNT :** Non-native trees (arbres non-indigènes)
- **PIR:** Proche Infrarouge
- **ROI:** Aires d'entraînement
- **RVB:** Rouge / Vert / Bleu
- **SCR:** Système de Coordonnées de Références Projeté
- **SIG:** Sciences de l'Information Géographique

I. Introduction

1.1 Contexte du projet ALPTREES

Les derniers rapports du GIEC indiquent une augmentation de l'intensité du changement climatique (Masson-Delmotte et al., 2021). Les écosystèmes alpins sont particulièrement sensibles au changement climatique. Une étude menée en 2017 par des chercheurs de l'université Savoie-Mont-Blanc et coordonnée par l'université d'Umea en Suède a démontré que les écosystèmes alpins sont fortement impactés avec notamment une remontée de la végétation forestière vers les étages alpins supérieurs créant un déséquilibre entre besoins des écosystèmes et ressources au sol (Mayor et al., 2017).

Les Alpes sont une région dynamique, dans laquelle le développement d'infrastructures anthropiques y est conséquent (tourisme, transport, urbain, agriculture, etc...) et provoque une pression sur les espaces naturels. Cette pression se traduit par une fragmentation des milieux naturels ainsi que par une pollution atmosphérique ce qui engendre une baisse de la biodiversité au sein des écosystèmes. ("Les Alpes, une région menacée | WWF France,").

A cela s'ajoute un phénomène qui n'est pas nouveau mais qui n'a été pris en compte que depuis une dizaine d'années : la colonisation de milieux par des espèces exotiques envahissantes (NNT) occasionnant la disparition d'espèces locales ("Espèces exotiques envahissantes en Auvergne,"). Leur caractère envahissant présente donc un risque de plus pour les écosystèmes alpins.

On peut prendre pour exemple le *Buddleia de David* (*Buddleja Davidii*) qui s'installe sur les zones rudérales, remaniées et en friche et se développe à grande vitesse, en monopolisant l'entièreté de l'espace disponible de par sa dissémination (Tallent and Watt, 2009).

En revanche, les espèces non natives (NNT) peuvent présenter des caractéristiques avantageuses face au changement climatique. On peut reprendre l'exemple évoqué ci-dessus; le *Buddleia de David* (*Buddleja Davidii*), par son développement racinaire peut être un atout dans la stabilisation d'un milieu érodé.

Cela illustre très bien toute la complexité du problème; la végétation évolue au sein des Alpes apportant à la fois son lot de bénéfices et d'impacts négatifs sur les différents milieux qui les composent. Il est donc nécessaire de qualifier le rôle des espèces exotiques envahissantes et ce grâce à une bonne connaissance de celles-ci ("Risk and opportunity of nnt – Alpine Space,"). Pour évaluer les risques qu'elles représentent il est nécessaire de les identifier, de les localiser, et de comprendre leur fonctionnement.

Le projet ALPTREES Alpine Space s'intéresse aux espèces végétales non-natives de l'espace alpin dans son ensemble.



Figure 1 : Carte de présentation des régions alpines concernées par le projet ALPTREES.
Source: ("about – Alpine Space,")

Les Alpes (figure 1) sont une chaîne de montagne s'étendant du Sud-Est de la France jusqu'en Slovénie. C'est un territoire particulier avec un fort relief, et une variabilité climatique importante. Ce territoire étendu comprend une grande variété d'écosystèmes ainsi qu'une faune et une flore diverse et disparate d'une région à l'autre. (Larousse, Alpes)

On recense de nos jours un certain nombre de projets (Lévêque et al., 2012) liés de près ou de loin aux espèces non indigènes dans les Alpes, qu'elles soient végétales ou animales. Cependant, ils sont la plupart du temps spécifiques à des zones géographiques restreintes. Les résultats et recommandations ne sont donc applicables que dans l'espace concerné.

Le projet ALPTREES a pour but de centraliser les connaissances sur espèces non-natives au sein du territoire alpin ; pour pouvoir par la suite les gérer au mieux ("about – Alpine Space,").

1.2 Objectifs du projet ALPTREES

Les objectifs du projet ALPTREES sont les suivants:

- Définir une approche transnationale pour étudier les caractéristiques des espèces non-indigènes et leur rôle dans les écosystèmes alpins,
- Fournir une stratégie transnationale et un système d'aide à la décision pour une meilleure gestion des espèces non-indigènes,
- Faire de ce système d'aide à la décision un outil adapté de gestion face au changement climatique afin de trouver un équilibre entre avantages qu'elles représentent et les risques encourus.

Ils comprennent les implémentations suivantes:

- « Développer une base de données complète sur les espèces d'arbres non indigènes,
- Prédire la répartition actuelle et potentielle des espèces d'arbres non indigènes dans l'Espace alpin selon des scénarios de changement climatique,
- Déterminer leur potentiel invasif,
- Analyser les différents services écosystémiques fournis par les espèces d'arbres non indigènes pour évaluer les compromis entre les risques et les avantages,
- ALPTREES formulera des recommandations de gestion pour les espèces d'arbres non indigènes dans différents scénarios climatiques et économiques. »

Source: ("about – Alpine Space,")

Le projet que nous avons mis en place s'inscrit dans le Work Package n°1 « Collecte de données et évaluation des risques spécifiques au site » ("Work packages – Alpine Space,") du projet ALPTREES.

L'idée est de fournir aux gestionnaires environnementaux un outil leur permettant d'établir efficacement des cartes d'espèces exotiques envahissantes (cartes de répartition, de dissémination, de suivi, d'évolution, de risques, etc...). Pour cela, il est nécessaire de créer une base de données géolocalisée de ces espèces, transposable sur les outils classiques de SIG.

1.3 Problématique

Notre étude est un travail exploratoire faisant appel aux compétences d'un géographe et nécessitant des développements informatiques, et sollicitant des techniques novatrices. Nous avons développé de nouvelles méthodes de télédétection mobilisant des outils numériques visant à faciliter la récolte de données spatiales sur les espèces exotiques envahissantes.

L'objectif au travers de l'outil est de détecter et classier de manière précise et automatisée la répartition des espèces non-natives au sol à partir d'images aériennes de différentes résolutions. Si l'outil est suffisamment développé il pourrait à terme permettre des études diachroniques de tous types de végétaux, ainsi que des modèles de prévisions.

Il serait donc potentiellement un outil d'inventaire, de suivi, d'étude et de gestion de la végétation transposable dans des situations et territoires différents.

Pour le réaliser nous avons découpé le projet en deux temps :

A. Récolter et construire une base de données géolocalisée d'arbres non-natifs sur différents sites sur l'espace de la Métropole grenobloise.

B. Mettre au point une méthodologie de "détection-classification" par "Réseaux de Neurones" de ces espèces non-indigènes sur des images haute résolution.

En l'état actuel des connaissances est-il possible, à l'aide des technologies actuelles d'identifier puis de classer de manière automatisée différentes espèces d'arbres à partir d'images de télédétection?

II. Etat de l'art

2.1 Sciences de l'information géographique, télédétection et végétation

Aujourd'hui la géographie est une discipline multithématique comprenant la compréhension de l'ensemble des systèmes naturels qui régissent le fonctionnement de notre planète ainsi que l'analyse du développement anthropique et l'impact de celui-ci sur les territoires. Elle comprend également la représentation spatiale de tous types d'études grâce aux SIG.

Les SIG constituent des logiciels grâce auxquels on peut traiter différents ensembles de données de manière spatiale et statistique. Ils sont aujourd'hui utilisés par tous types d'organismes nécessitant des études spatiales, on peut citer par exemple les opérateurs développant les réseaux de fibre optique pour lesquels anticiper l'organisation de ces réseaux s'avère indispensable. Ils jouent de ce fait un rôle important dans la caractérisation des territoires et ainsi dans l'aménagement de ceux-ci, ce sont des outils d'acquisition, de traitement et de gestion de données qui permettent une vulgarisation de la réalité tout en fournissant des informations géoréférencées ("SIG: Tout savoir sur les Systèmes d'Information Géographique | Esri France,"), ("SIG Système d'Information Géographique ou Géomatique,").

La cartographie des espèces invasives est une nécessité car celles-ci sont aujourd'hui considérées comme l'une des plus grandes menaces pour les êtres humains à cause des dégâts irréversibles qu'elles peuvent provoquer sur les écosystèmes (Alexander et al., 2016). Aux Etats-Unis, les dégâts provoqués par les espèces non-natives ont été chiffrés à près de 137 milliards de dollars par an, chiffre sous-estimé selon l'auteur car de nombreuses externalités négatives ne sont pas prises en compte (Pimentel et al., 2000). Il est donc important de surveiller de près l'évolution de ces espèces. Pour cela il existe des moyens de les cartographier à l'aide de la télédétection et des SIG.

Le terme télédétection désigne les techniques d'acquisition et d'analyse d'images aéroportées de par les informations de rayonnement renvoyées par celles-ci à l'aide de divers capteurs. Elle permet tous types d'études au sol à distance (Larousse). Alors que la cartographie spatiale et temporelle des végétaux demande des moyens sur le terrain, la télédétection permet une analyse spatiale à distance. On peut prendre pour exemple l'étude des chercheurs Mirik et Ansleys qui en tentant de détecter le chardon musqué (qui est une espèce envahissante) aux Etats-Unis (Texas), ont pu déterminer à quelle période de l'année la réflectance du charbon était la plus forte (Mirik et al., 2013). Cela pourrait permettre (par exemple) d'établir des cartes de répartition annuelles ainsi que des cartes d'évolution sur plusieurs années.

Les techniques de SIG et de télédétection sont donc un atout dans la connaissance de l'occupation au sol notamment grâce à des outils de plus en plus performants tels que les capteurs LiDAR ou hyperspectraux. Un comparatif entre télédétection à résolution modérée, haute, hyperspectrale, et LiDAR a été établi par Huang et Asner sur les plantes exotiques envahissantes et le constat est le suivant : alors qu'une résolution modérée permet difficilement de discriminer des espèces au sol, les résolutions supérieures permettent de réaliser des classifications au sol assez fiables. Néanmoins la télédétection présente une limite courante quand il s'agit d'étudier la végétation : les bandes spectrales représentant les différentes espèces sont très étroites et les confusions peuvent être nombreuses si l'espèce en question ne présente pas de caractéristiques spécifiques ou s'il elle ne se trouve pas en terrain dégagé (Huang and Asner, 2009).

Bien qu'inventaires et études des espèces non-natives se multiplient, la connaissance sur leur répartition et leur propagation se trouve être encore de nos jours assez limitée.

2.2 Intelligence Artificielle, Machine Learning et Réseaux de Neurons

C'est l'apparition du « BIG DATA » à la fin des années 90 qui a permis l'apparition des Intelligences artificielles puis des Machine Learning; ils représentent des ensembles de données importants qui ne peuvent être traités par des machines classiques (Cox and Ellsworth, 1997).

En premier lieu, il est important de distinguer les IA des Machine Learning ainsi que les Machine Learning des Réseaux de Neurons profonds. En effet les trois termes sont souvent utilisés à l'oral de manière aléatoire, et pourtant ils représentent des méthodes différentes plus ou moins poussées. Les Machine Learning sont un sous-ensemble d'Intelligences Artificielles tandis que les CNN sont des sous-ensembles de ML.

“L'Intelligence Artificielle est partout dans la presse. Le Machine Learning est partout dans les entreprises. Le Deep Learning est partout dans la recherche.”
(“Intelligence Artificielle, Machine Learning et Deep Learning,” Ahmad 2019)

L'Intelligence Artificielle renvoie donc à des algorithmes capables d'imiter des actions humaines à condition de lui fournir des références sur le sujet en question. Elle regroupe un spectre très large d'algorithmes plus ou moins performants dont les résultats doivent être au minimum corrects ou similaires à ceux donnés par un expert humain (Lambert, 2018).

Le Machine Learning désigne lui, un IA à laquelle on ajoute un certain nombre de données statistiques qui permettent à l'ordinateur de composer des schémas transposables sur de nouvelles bases non référencées. C'est là qu'intervient le terme "Learning"; en effet plus on fournit une base de données conséquente à l'ordinateur, plus celui-ci sera capable d'apprendre en autonomie et de fournir des résultats de plus en plus précis. Les résultats prendront donc la forme de statistiques s'approchant de la réalité, néanmoins on parle ici de valeurs chiffrées tandis que la vérité terrain est souvent bien plus complexe. A noter qu'il est important de fournir au Machine Learning une base de données fiable et précise car si on y introduit des anomalies, celui-ci le prendra dans sa base de données comme argent comptant, ce qui peut créer des aberrations dans les résultats. Il est donc doté d'une capacité d'apprentissage, mais celle-ci reste limitée (Lambert, 2018).

Enfin le Deep Learning (ou réseaux de neurones profonds) est un algorithme inspiré des réseaux cognitifs du cerveau humain. Celui-ci reçoit de l'information, la compare aux informations existantes et en extrait celles qui sont pertinentes pour l'étude en question. Informatiquement parlant, c'est un réseau qui va analyser les données de pixels bruts d'une image et qui va apprendre à en extraire les caractéristiques grâce à des bases d'entraînement. Alors qu'il faut fournir au Machine Learning des caractéristiques à étudier, les réseaux de neurones ne nécessitent qu'une seule base de données dans laquelle ils vont mettre en place des modèles de calculs sur plusieurs couches de profondeur (en créant des niveaux d'abstraction) en fonction des similitudes détectées entre les différentes informations. Bien entendu plus un CNN sera étudié et élaboré à partir d'hyperparamètres plus il sera efficace, cela permet d'accélérer les processus et de les rendre plus fiables, l'intervention humaine demeure par conséquent bien nécessaire autant dans l'élaboration d'un CNN que dans la vérification des résultats (Lambert, 2018), ("Les réseaux de neurones,").

Les réseaux de neurones résultent donc l'évolution des Machines Learning qui eux même sont l'évolution de l'intelligence artificielle.

2.3 La place des Machine Learning et Réseaux de Neurones en Sciences de l'Information Géographique

Les Machine Learning et CNN sont de plus en plus utilisés en SIG car ils permettent un apprentissage rapide et une détection d'informations efficaces. Ils permettent d'étudier, de capter des similitudes et de catégoriser des caractéristiques spatiales fournissant un certain nombre de données pouvant par la suite favoriser un système d'aide à la décision ou encore d'établir des modèles de prévision. On peut prendre pour exemple ESRI qui a présenté à une

conférence ESRI à Paris en 2019 un code Python de Deep Learning qui a permis de détecter des fissures sur une voie routière à partir d'une vidéo de cette même route puis de les classer en fonction de leur forme et leur taille (Messaoudi, 2019). On peut facilement imaginer l'atout que pourrait représenter un tel outil pour les services publics routiers. Les utilisations des ML et des CNN en SIG peuvent donc être multiples.

Tandis que les techniques traditionnelles de classification ne permettent pas de différencier les espèces entre elles au sein de peuplements, les CNN sont en capacité d'étudier et d'apprendre les caractéristiques renvoyées par chaque pixel et ainsi d'établir des modèles permettant la reconnaissance de tous types d'objets. Des études récentes (Nijhawan et al., 2017), (Zhang et al., 2019), (Langford et al., 2019), (Ayhan et al., 2020) ont démontré qu'il était possible de détecter et de classer différentes espèces végétales. Bien entendu, ces études sont menées sur des espèces ciblées dans des zones ciblées mais on peut imaginer que d'ici quelques années nous serons en capacité de combiner SIG, télédétection et CNN pour réaliser des suivis d'espèces fiables au sein de peuplements.

On voit émerger ces dernières années de plus en plus d'études tentant de discriminer différentes espèces végétales par CNN ; on peut citer l'exemple du travail des chercheurs allemands Heinzl et Koch qui ont en 2012 réussi à classer des espèces comme le Pin (*Pinus*), le Chêne (*Quercus*) ou encore le Hêtre (*Fagus Sylvatica*) à des précisions allant de 80 à 90% grâce à des données lidar et hyperspectrale (Heinzl and Koch, 2012). Cette étude date de 2012, il a été démontré récemment que les CNN sont efficaces pour établir des modèles étant capables d'extraire de nombreuses propriétés de la végétation à partir d'images aériennes, certains parlent même de nouvelle ère en matière de télédétection de la végétation (Kattenborn et al., 2021). Avec un matériel adapté, ces classifications par CNN peuvent atteindre des performances hors normes. Une étude menée par l'Institut de technologie spatiale au Pakistan sur de l'Infrarouge thermique hyperspectral (plus de 3000 bandes) a réussi à classer un panel d'espèces à près de 99% de précision sur la base d'entraînement à l'aide d'un CNN (Hasan et al., 2019). Cela démontre qu'avec des moyens conséquents il est possible de créer des bases d'entraînement fonctionnelles.

Les CNN sont donc un outil automatisé d'avenir qui peuvent traiter des bases de données d'envergure et qui gagnent en efficacité au fur et à mesure que la technologie informatique évolue. C'est également un outil adapté aux classifications de végétation.

III. Données & Méthode

3.1 Les sites d'étude

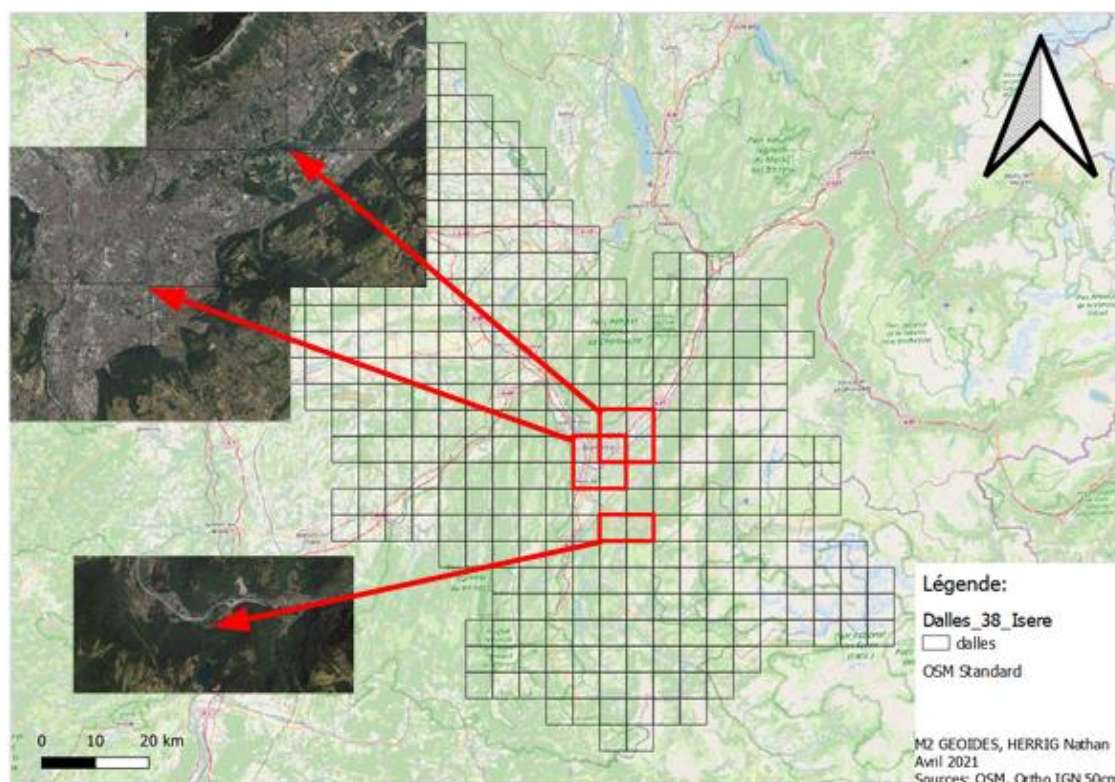


Figure 2 : Carte des sites retenus pour l'étude.
Source : HERRIG Nathan, M2 GAED-GEOIDES.

Les sites (Figure 2) ont été ciblés en fonction de leur occupation au sol et de la donnée disponible. Nous avons établi qu'il était pertinent de sélectionner différentes zones en milieu urbain, péri-urbain et naturel. La métropole grenobloise a donc été retenue en tant que zone urbaine pour ses données disponibles (orthophotos IGN à 50 et 20cm, proximité, relevés floristiques géolocalisés de la ville), à noter que la végétation ornementale est plus facile à étudier du fait de l'isolation des individus.

Les berges de l'Isère ont été retenues en tant que zone péri-urbaine (orthophotos IGN à 50 et 20cm, proximité et relevés floristiques géolocalisés),

Enfin la zone de compensation écologique de l'Île Falcon a été retenue en tant que zone naturelle du fait des études INRAE en cours qui ont déjà permis de produire plusieurs orthophotos par vols de drones en 2019 et 2020 à 3, 5 et 10cm de résolution.

3.2 Les espèces retenues

Nous avons établi les critères suivants pour la sélection d'espèces à compiler dans notre base de données :

- visibilité sur images aériennes,
- relevés en quantité suffisante (au moins 300 individus),
- signature radiométrique particulière,
- taille du houppier supérieure à 4m²,
- caractère invasif pour la plupart¹.

La plupart des espèces retenues ont d'abord été sélectionnées au sein du tableau de NNT ALPTREES puis d'autres ont été ajoutées pour les besoins de l'étude. Voici une rapide présentation dans les figures 3 et 4 des espèces principales étudiées.

¹ Le cadre de cette étude s'était à l'origine fixé comme objectif de recenser uniquement des espèces invasives. Toutefois nous avons été amenés à étendre notre choix et à intégrer d'autres espèces en raison de leur forte présence sur les zones étudiées.

Robinier Faux-Acacia

Robinia Pseudoacacia
Fabaceae



Erable Négundo

Acer Negundo
Aceraceae



Pin Noir d'Autriche

Pinus Nigra
Pinaceae



Robinier Faux Acacia:

Le Robinier Faux-Acacia est une espèce issue d'Amérique du Nord et plus précisément de l'Est Etats-Unis.

C'est une espèce pionnière à croissance rapide qui s'installe sur des terrains sinistrés et qui initie leur régénération en fixant l'azote par ses racines. (Chambras, 2019)

Il a été introduit en Europe au XVIIème siècle.

Nous l'avons sélectionné pour sa forme particulière ainsi que pour sa forte dissémination.

Erable Négundo:

L'Érable Négundo est une espèce issue d'Amérique du Nord; son aire de répartition d'origine s'étend du Sud du Mexique jusqu'au Sud du Canada.

C'est une espèce qui s'installe principalement en zones humides et notamment au bord des cours d'eau. Elle est considérée comme envahissante car elle colonise les plaines alluviales empêchant la croissance d'autres plantes du fait de son couvert végétal dense.

Il a été introduit en Europe au XVIIIème siècle. (Dumas, 2019)

Nous l'avons sélectionné pour sa présence sur les berges de l'Isère ainsi que pour sa signature radiométrique particulière.

Pin Noir d'Autriche:

L'aire de répartition des Pins Noirs est assez complexe, elle s'étend de l'extrême ouest de l'Europe jusqu'en Turquie et de 35 à 48° Nord. Les Pins Noirs sont subdivisés en 5 espèces dont le Pin Noir d'Autriche.

Le Pin Noir et ses sous-espèces présentent une forte adaptabilité à différents milieux, on peut les trouver aussi bien à 250 m d'altitude qu'à plus de 1500 m d'altitude, il est capable de s'installer sur des sols pauvres et secs mais également de résister à des gels pouvant descendre jusqu'à -30°C ("Pinus nigra Arnold,")

Nous l'avons sélectionné pour sa forte présence sur notre territoire ainsi que pour sa signature radiométrique particulière.

Figure 3 : Espèces retenues pour l'étude.

Sources images : ("Le robinier faux-acacia (Robinia pseudoacacia),"), ("8 idées de Acer negundo | feuille d'érable, érable negundo, feuille,"), ("Pinus Nigra", limited)

Buddleia de David

Buddleja Davidii
Buddlejaceae



Tilleul à Feuilles Courtes

Tilia Cordata
Malvaceae



Frêne Elevé

Fraxinus Excelsior
Oleaceae



Buddleia de David:

Le Buddleia de David est une espèce végétale endémique de Chine.

C'est une espèce considérée comme toxique et invasive du fait de son caractère pionnier. Elle s'installe principalement sur les terrains sinistrés ou en friche mais s'adapte à de nombreux territoires, elle est capable de s'installer jusqu'à 3000 m d'altitude. Elle supprime les autres espèces pionnières du fait d'une résistance particulière à la pollution de l'ozone.

Elle s'est naturalisée sur le reste des zones tempérées au cours du XIXème siècle et a été fortement implantée en Europe en tant que plante ornementale au cours du XXème siècle. (Tallent and Watt, 2009)

Nous l'avons sélectionné du fait de sa forte présence sur la zone Falcon ainsi que pour sa forme arbustive.

Tilleul à Feuilles Courtes:

Le Tilleul à feuilles courtes ou Tilleul à petites feuilles est une espèce endémique d'Europe, son aire de répartition couvre quasiment le continent dans son entièreté allant du Nord de l'Espagne jusqu'en Russie.

C'est donc un arbre ultra-implanté dans nos territoires, il peut atteindre jusqu'à 500 ans de longévité et peut s'installer jusqu'à 1500 m d'altitude. C'est une espèce de mésophile qui apprécie les sols relativement riches. Néanmoins, il s'adapte également à des sols un peu plus secs et aérés. ("Le tilleul à petites feuilles (*Tilia cordata*),")

Il ne fait pas partie d'espèces considérées comme invasives mais nous l'avons tout de même sélectionné pour sa forte présence afin d'augmenter notre base de données.

Frêne Elevé:

Le Frêne élevé est également une espèce endémique d'Europe dont l'aire de répartition couvre quasiment tout le territoire Européen, allant, comme le Tilleul à petites feuilles, du Nord de l'Espagne jusqu'en Russie.

C'est une espèce post-pionnière très résistante au froid et dont l'adaptabilité est forte. En effet sa tolérance écologique dépasse l'entendement puisqu'on le trouve aussi bien en montagne, qu'en plaine ou encore en bord de mer. Il est néanmoins sensible aux sécheresses. ("Frêne élevé," 2021)

Lui non plus ne fait pas partie des espèces considérées comme invasives, il a été sélectionné pour les mêmes raisons que le Tilleul à petites Feuilles ainsi que pour sa forme particulière.

Figure 4 : Espèces retenues pour l'étude.

Sources images : ("Buddleja davidii illustration - circa 1898 | Buddleja davidii, Botanical illustration, Botanical drawings,"), ("Le tilleul à petites feuilles (*Tilia cordata*),"), ("Frêne | Remèdes de Grand-mère | Ash leaf, Botanical flowers, Botanical illustration,")

NB: D'autres taxons tels que l'Ailante Glanduleux, le Noyer Noir, le Myrobolan, le Sumac Vinaigrier et autres espèces toxiques et invasives ont été recensés, étudiés et utilisés mais les données se sont avérées insuffisantes pour la suite du projet. Celles-ci ont tout de même été conservées dans les données de réserve.

3.3 Les données mobilisées et produites

La création d'une base de données conséquente d'espèces non-natives est l'un des objectifs principaux du projet ALPTREES. Cette base de données est principalement fournie par les données provenant de l'application iNaturalist qui est une application participative dont vous pouvez trouver tous les résultats actuels en consultant le site: (<https://www.inaturalist.org/projects/alptrees>).

Malheureusement les relevés floristiques réalisés par smartphones que ce soit sur iNaturalist, iPlant ou autres ne répondent pas aux critères de précision demandés par notre projet ceux-ci étant précis de quelques mètres à des dizaines de mètres tandis que nous recherchons une précision inférieure à 50 centimètres.

Nous avons donc dû constituer notre propre base de données. Les premières données nous ont été transmises par le PIFH (Pôle d'Informations Flore-Habitats-Fonge) sur le secteur de la Romanche ainsi que sur le secteur des Berges de l'Isère. Nous avons également contacté la Métro Grenobloise pour obtenir les relevés floristiques de la ville, la plupart des arbres étant ornementaux ils sont en général isolés les uns des autres et donc plus facilement distinguables par télédétection, d'où l'intérêt d'étendre l'étude à une zone urbaine. Enfin des données floristiques étaient déjà à disposition au sein d'INRAE, elles ont été récoltées à l'occasion d'études sur l'île Falcon et notamment sur une étude portant sur la télédétection de *Buddleia Davidii* sur un site de restauration écologique géré par EDF.

Ces données prennent donc la forme de points géolocalisés de manière précise, reportés sur des logiciels tels que QGis. Ces points nous permettent d'identifier les espèces retenues sur des photos aériennes ou sur des images de drones. Elles ont été vérifiées sur le terrain. Ainsi nous avons sélectionné une gamme de points que nous avons validés, invalidés, ou corrigés. Sur les relevés floristiques de la Métro grenobloise figuraient par exemple quelques arbres qui avaient été coupés.

Nous avons ensuite complété cette base de données sur le terrain sur les différents sites étudiés à l'aide de GPS de haute précision. La méthode est simple; il s'agit en premier lieu de définir un panel d'espèces à identifier (espèces retenues partie 3.2). Puis, sur le terrain, de les reconnaître, de les reporter sur des logiciels

tels que Qfield, et de les transposer sur QGIS en les superposant aux images aériennes. La démarche d'utilisation des GPS est expliquée dans les tutoriels en annexes n°3 et n°4.

Enfin nous sommes également allés sur le terrain (zone Falcon) pour réaliser des vols de drones. En effet le projet implique non seulement d'avoir une base de données de végétaux conséquente, mais également une base de données d'images aériennes et de drones multiples, l'intérêt étant de tester un code de Machine Learning sur des images à différentes résolutions. En plus des images aériennes IGN à 50 cm et 20 cm de résolution à disposition, nous avons réalisé des vols drone au-dessus de la zone de compensation écologique de l'île Falcon permettant la création d'orthophotos à 3cm de résolution (et également à 5 et 10cm de résolution en les dégradant). A noter que ces vols n'ont pu être réalisés que sur cette zone, les vols de drones en zone urbaine étant règlementés et nécessitant des autorisations.

Les vols de drone ont été réalisés avec le drone DJI Phantom 4 multispectral composé de 6 caméras (RVB, Red Edge, PIR) qui permettent une prise de vue sur plusieurs longueurs d'ondes permettant diverses analyses en fonction des besoins de l'étude. Il est doté d'un positionnement RTK en temps réel qui géotague chaque image et améliore la précision des traitements photogrammétriques.

Les images sont ensuite traitées avec le logiciel Agisoft Metashape v. 1.7.3 permettant la création d'orthophotos à haute résolution (*figure 5*). La première étape (pour le vol réalisé le 15/06/2021) est d'aligner les 3651 photos à l'aide de 175 markers préétablis (*figure 9*) sur le terrain puis disposés sur toute la zone photographiée. Ils permettent de réduire les erreurs de position grâce à des points de contrôle (en rouge) et des points de validation (en jaune). On crée ensuite un nuage de points par corrélation grâce à l'algorithme SIFT. Puis on y applique un géoréférencement grâce à des points pris sur le terrain par GPS différentiel (DGPS) et basé sur un SCR en Lambert 93 EPSG 21:54. On réalise ensuite une densification au nuage de points qui nous permet de créer un DSM (DigitalSurfaceModel). Enfin on génère une orthophotographie à l'aide du DSM.

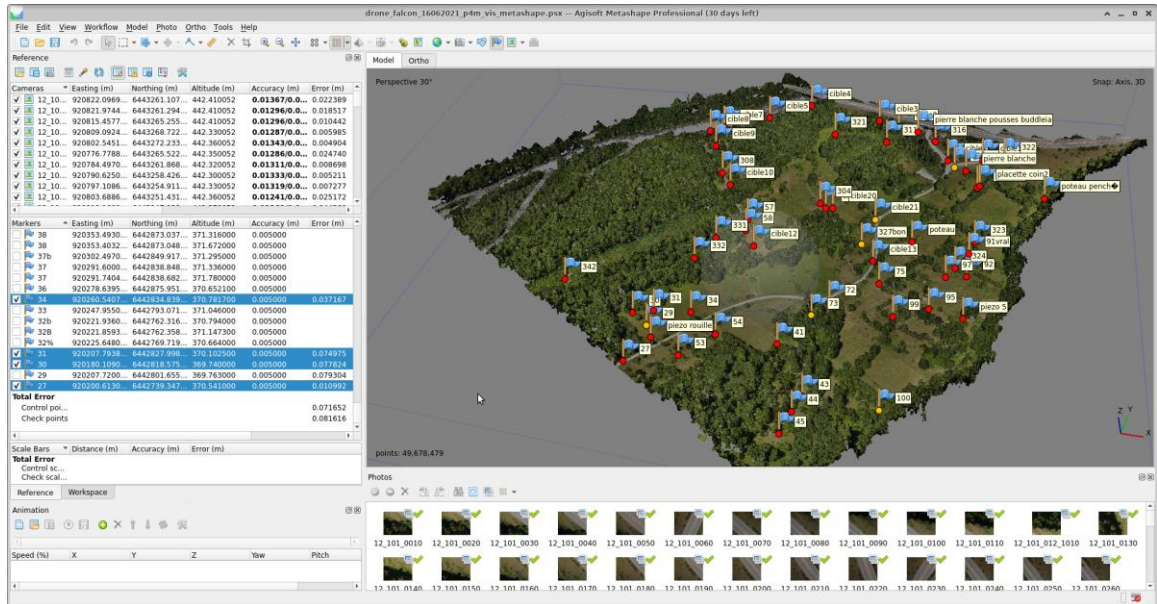


Figure 5 : Capture d'écran de la réalisation d'une orthophotographie sur Agisoft Metashape.
Source : HERRIG Nathan, M2 GAED – GEOIDES.

Ainsi on obtient des orthophotos (figure 6) de haute précision comme celles ci-dessous avec des résolutions à 3, 5, (et 10cm en dégradant l'image) on distingue facilement les différentes entités composant l'occupation au sol.

Présentation des orthophotos créées à partir d'images de drones et générées avec Agisoft Metashape:

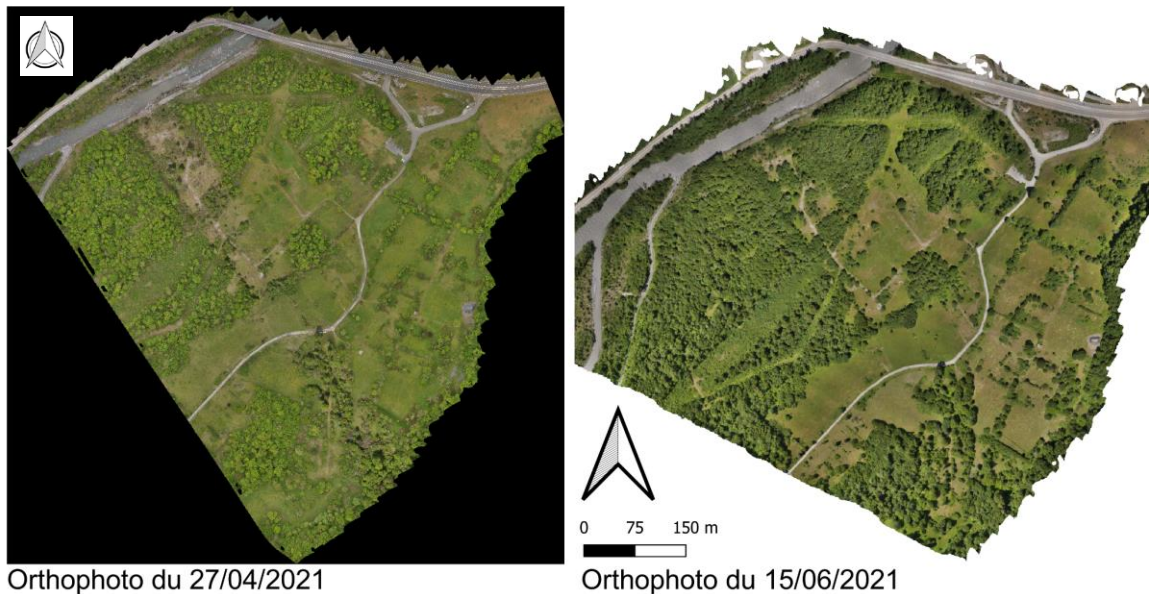


Figure 6 : Orthophotos de la zone de l'île Falcon réalisées à partir de vols de drones.
Source : BORNIET Laurent, INRAE & HERRIG Nathan, M2 GAED – GEOIDES.

Nous avons donc sélectionné un panel de données (*figure 7*) nous permettant en premier lieu de créer une base de données d'arbres géolocalisés sur des images à différentes résolutions, celle-ci étant nécessaire à l'apprentissage d'un Machine Learning ou d'un réseau de neurones convolutifs.

	Ile falcon (Saint-Barthélemy de Séchilienne)	Métropole grenobloise	Berges de l'Isère (Saint-Martin d'Hères)
PIFH	<ul style="list-style-type: none"> PIFH_SudGrenoble 	X	<ul style="list-style-type: none"> PIFH_NordGrenoble
Ville de Grenoble	X	<ul style="list-style-type: none"> Arbres_Grenoble (Relevés floristiques de la Ville) 	X
INRAE	<ul style="list-style-type: none"> Ortho IGN 38_2018_20cm Dalles_38 Ortho DRONE Falcon_2016_5cm Ortho DRONE Falcon_2020_10cm Relevés GPS Falcon_2020 	<ul style="list-style-type: none"> Ortho IGN 38_2018_20cm Dalles_38 	<ul style="list-style-type: none"> Ortho IGN 38_2018_20cm Dalles_38
Données Terrain	<ul style="list-style-type: none"> Relevés GPS Falcon 22/04/2021 + 27/04/2021 + 15/06/2021 Ortho DRONE Falcon_27/04/2021 Ortho DRONE Falcon_15/06/2021 	<ul style="list-style-type: none"> Relevés GPS Grenoble 19/04/2021 	<ul style="list-style-type: none"> Relevés GPS Berges de l'Isère 19/04/2021

Figure 7 : Tableau des données principales utilisées durant la réalisation de l'étude.
Source : HERRIG Nathan, M2 GAED – GEOIDES.

NB: D'autres données ont été utilisées et mises en forme mais ne figurent pas dans les 3 principales zones étudiées, elles sont néanmoins prêtes à l'emploi en cas de réhabilitation au sein du projet et constituent des données de réserve.

3.4 Classifications

Il existe deux types d'outils de classification au sein des SIG qui permettent d'identifier l'occupation au sol sur des images : la classification supervisée pixel et la classification supervisée objet.

La première (pixel) est issue d'une analyse spectrale de chaque pixel (les surfaces qui composent l'image renvoient une signature radiométrique) contenu dans l'entité spatiale.

La seconde fonctionne sur le même principe mais est composée de prétraitements (segmentation) qui vont créer des groupes de pixels de taille variable dont la signature radiométrique est homogène.

Pour ce faire on suit la méthode suivante. On crée en premier lieu des ROI qui sont des aires d'entraînement, c'est-à-dire que l'on indique à l'algorithme utilisé des vérités terrain pour lui fournir une base d'entraînement. On entre ensuite les images à classifier, on sélectionne l'algorithme à faire tourner et les ROI d'entraînement que l'on peut séparer pour créer une gamme de ROI de validation, qui nous permettront ensuite d'établir les matrices de confusion.

Parmi eux nous avons testé les 3 suivants: **OTB, Dzetsaka et Ecognition.**

	Classification supervisée Objet	Classification supervisée Pixel
OTB		- LibSVM
DZETSAKA		- Gaussian Mixture Model - Random Forest - K-Nearest Neighbour
Ecognition	- K-Nearest Neighbors	

Figure 8 : Tableau récapitulatif des algorithmes de classification testés sur différents logiciels.
Source : HERRIG Nathan, M2 GAED – GEOIDES.

La figure 8 est un tableau récapitulatif des différents algorithmes testés avec les différents outils. Ils ne fonctionnent pas tous de la même manière: par exemple le LibSVM (Support Vector Machine) est un classificateur qui marche avec un mappage de données sur un espace d'attributs qui va ensuite être

séparé (“IBM Docs,” 2021), tandis que le KNN (K-Nearest Neighbour) va étudier les caractéristiques d’un échantillon selon les entrées et les distances aux autres échantillons (“Méthode des k plus proches voisins,” 2021). Dans le cas de notre étude c’est l’algorithme KNN en classification objet sur Ecognition qui s’est avéré être le plus efficace.

3.4.1 Classifications Pixels

La figure 9 est le résultat de premier test avec l’algorithme LibSVM réalisée sur OTB (en libre-service sur QGis) sur une orthophoto IGN à 20cm de résolution, afin de discriminer trois classes de l’occupation du sol : l’urbain, la végétation et l’eau. On peut d’ores et déjà remarquer que le nombre de classes est trop faible par rapport à la réalité terrain, en effet l’ensemble des pixels ont classifié alors que l’occupation au sol est bien plus complexe. Cela induit la nécessité d’apporter un nombre de classes suffisant pour classer une image aérienne au complet et de manière fiable. L’autre solution est d’ajouter à l’image une couche de masque qui permet de passer en NO DATA les parties de l’image qui ne sont pas pertinentes pour l’étude en question. Malgré une discrimination de la végétation plutôt correcte (supérieure à 80%), l’outil a immédiatement été abandonné du fait d’un temps de traitement trop long (plus d’une semaine pour réaliser la carte ci-dessus).

Classification réalisée avec OTB sur la métropole Grenobloise (Dalle 910-460):

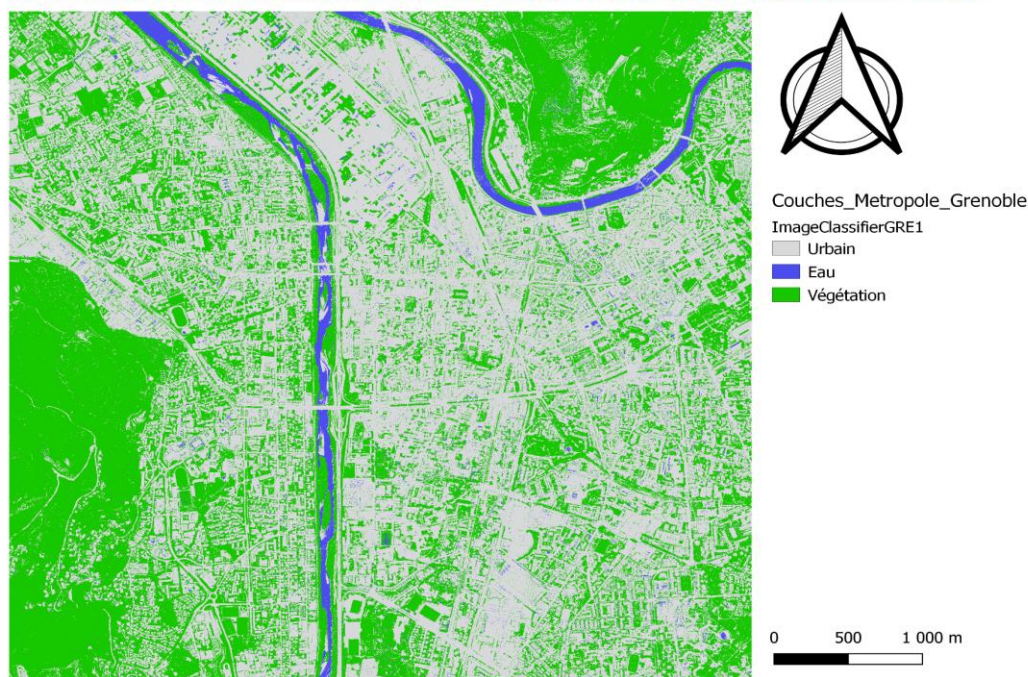


Figure 9 : Classification de la Métropole Grenobloise réalisée sur OTB.
Source : HERRIG Nathan, M2 GAED – GEOIDES.

Voici la classification réalisée avec l'outil DZETSAKA (figure 11), un module complémentaire installable en libre-service sur QGis. Nous avons classé la dalle selon la nomenclature suivante (figure 10):

Urbain	Toits
	Bâti
	Route
Aquatique	Fleuve
	Bassins
Végétation	Pelouses
	Champs
	Végétation Haute
Ombres	Ombres
No Data	No Data

Figure 10 : Nomenclature de la classification de la Métropole Grenobloise réalisée sur DZETSAKA.

Source : HERRIG Nathan, M2 GAED – GEOIDES.

Les résultats sont peu satisfaisants sur l'image en sortie, nous avons donc essayé à nouveau en changeant de paramètres et de ROI mais en vain. Le constat est le suivant: l'outil fonctionne sur des images à basse résolution mais devient confus sur des images à haute résolution. Bien que le projet soit axé sur le test d'images à différentes résolutions, les images en ayant une supérieure à 20 cm ont été éliminées d'entrée de jeu, celles-ci ne permettant pas de différencier les végétaux par individus à l'œil nu et le nombre de pixels par individus étant trop faible pour créer des imagerie d'entraînement pour un modèle. Il a donc également rapidement été abandonné.

Classification réalisée avec Dzetsaka sur la métropole Grenobloise (Dalle 910-460):

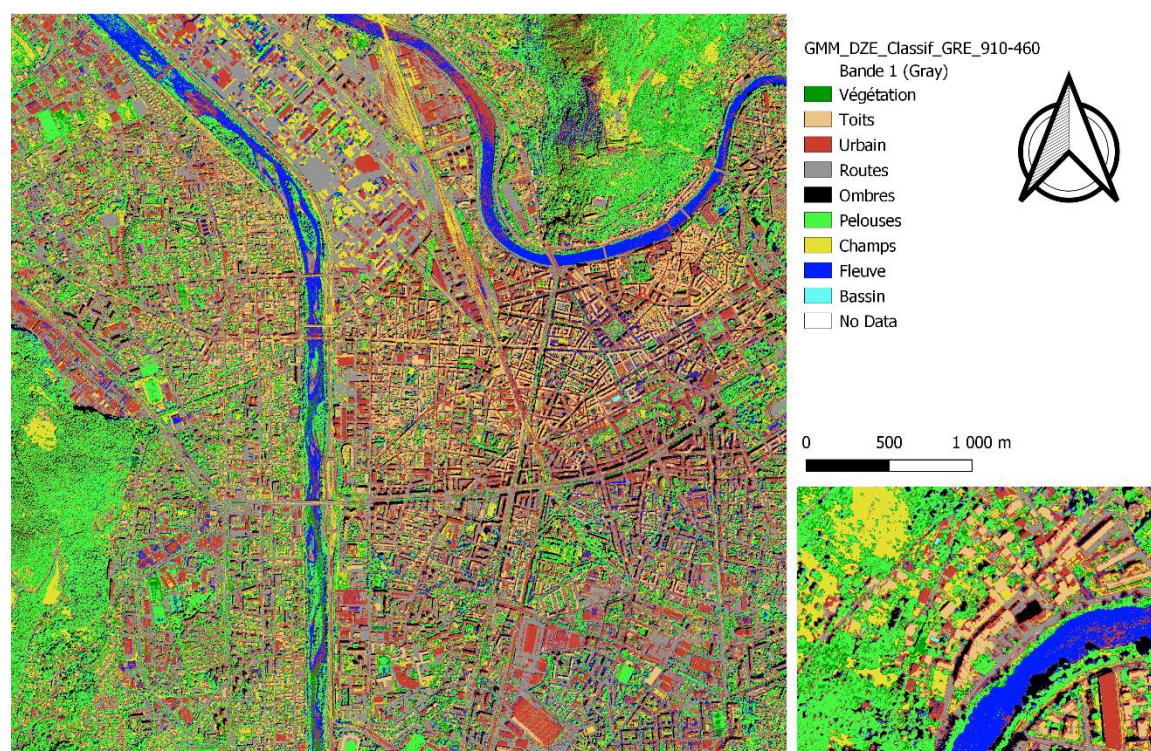


Figure 11 : Classification de la Métropole Grenobloise réalisée sur DZETSAKA.
Source : HERRIG Nathan, M2 GAED – GEOIDES.

NB : La matrice de confusion générée par Dzetsaka n'a pas été ajoutée car elle présentait des résultats aberrants.

3.4.2 Classifications Pixels

Le dernier outil de classification testé est Ecognition (*figure 12*). Il est basé sur l'algorithme K-Nearest-Neighbors qui vérifie et répertorie le nombre de voisins ayant le même label. Nous avons classé la dalle selon la même nomenclature que sur Dzetsaka (*figure 10*).

Nous avons constaté des anomalies mais les résultats sont de loin les plus satisfaisants comparés aux classifications précédentes. Nous avons donc tenté une seconde classification sur ce logiciel en discriminant végétation haute et végétation arbustive puis directement par espèces mais nous n'avons obtenu aucun résultat, les cartes établies étant incohérentes. Cela démontre ipso facto la nécessité d'apporter des techniques de télédétection nouvelles dans la reconnaissance et la classification d'espèces végétales.

Cette classification a par la suite été utilisée de nouveau pour réaliser des couches de masque.

Classification réalisée avec ECognition sur la métropole Grenobloise (Dalle 910-460):

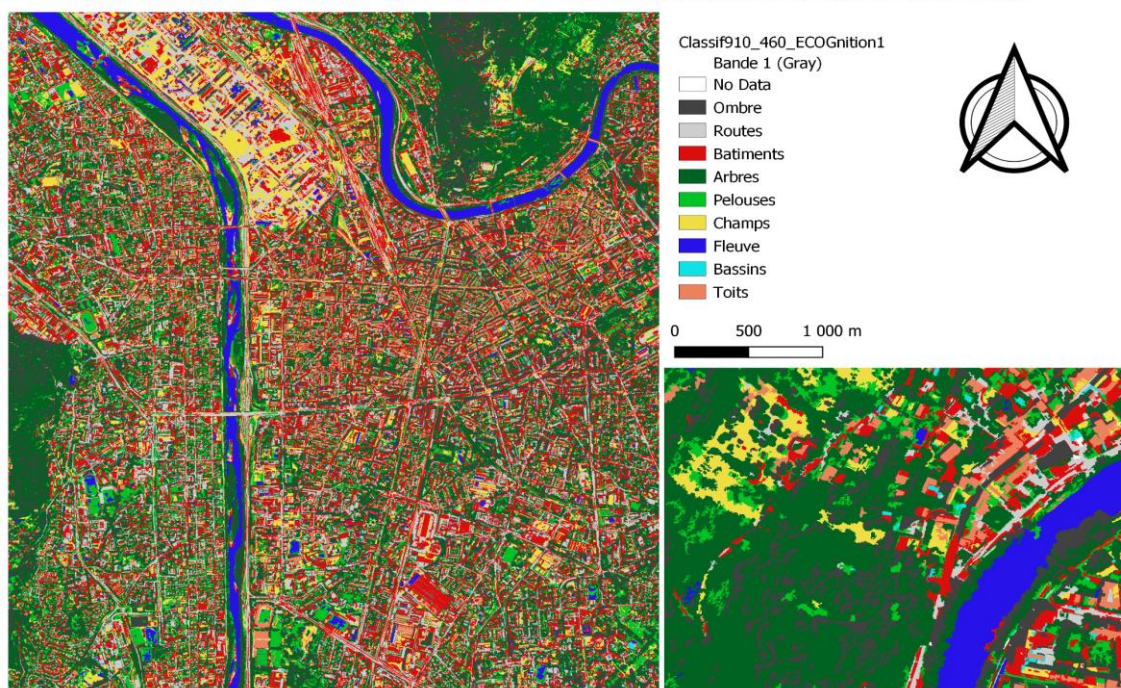


Figure 12 : Classification de la Métropole Grenobloise réalisée sur ECognition.
Source : HERRIG Nathan, M2 GAED – GEOIDES.

User Class \ Sample	Ombre	Routes	Batiments	Arbres	Pelouses	Champs	Fleuve	Bassins	Toits	Sum
Ombre	110	0	0	33	0	0	0	0	0	143
Routes	0	69	17	0	0	0	1	0	1	88
Batiments	0	18	88	0	0	0	1	0	0	109
Arbres	0	0	4	84	3	0	0	0	0	91
Pelouses	0	0	3	0	44	3	1	0	1	52
Champs	0	0	4	0	9	68	0	1	0	82
Fleuve	0	0	3	0	9	0	86	0	0	98
Bassins	0	0	0	0	0	0	0	39	0	39
Toits	0	1	0	0	0	0	0	0	96	97
Unclassified	0	1	0	0	0	0	0	0	0	1
Sum	110	89	119	117	65	73	87	40	99	
Producer	1	0.77528089888	0.73949579832	0.71794871795	0.67692307692	0.93150684932	0.98850574713	0.975	0.9696969697	
User	0.76923076923	0.78409090909	0.80733944954	0.92307692308	0.84615384615	0.82926829268	0.87755102041	1	0.98969072165	
Heiliden	0.86956521739	0.77966101695	0.77192982456	0.80769230769	0.75213675214	0.87741935484	0.92972972973	0.98734177215	0.97959183673	
Short	0.76923076923	0.63888888889	0.62857142857	0.67741935484	0.60273972603	0.7816091954	0.86868686869	0.975	0.96	
KIA Per Class	1	0.74750662795	0.69840323973	0.68174749557	0.65446318387	0.92368451177	0.98690113633	0.97371879106	0.96551575499	
Overall Accuracy	0.855									
KIA	0.83539240641									

Figure 13 : Matrice de confusion de la classification de la zone de Grenoble réalisée sur ECognition
Source : HERRIG Nathan, M2 GAED – GEOIDES.

La classification suivante a été réalisée sur la zone de l'île Falcon (figure 14). Après avoir testé les différents outils de classification et sélectionné ECognition pour réaliser la suivante nous avons déterminé un échantillonnage à 300 ROIs d'entraînement et 300 ROI de vérification pour déterminer quelle précision nous pouvions obtenir dans l'objectif de faire de ces classifications des couches de masque permettant ainsi au futur CNN de se concentrer sur les zones de végétation (végétation en valeur 1, le reste en valeur 0). On peut voir sur la matrice de confusion (figure 15) que quelques anomalies entre les classes d'eau et les classes de routes et de sol nu persistent. Néanmoins on a peu de

confusion entre végétation haute et végétation herbacée avec une précision pour les deux classes avoisinant les 99% sur les ROI de validation.

Classification réalisée avec ECognition sur la zone de l'Ile Falcon



Figure 14 : Classification de la zone de l'Ile Falcon réalisée sur ECognition.
Source : HERRIG Nathan, M2 GAED – GEOIDES.

User Class \ Sample	Végétation Haute	Végétation Herbacée	Ombres	Routes/Bâti/Sol nu	Eau	Sum
Végétation Haute	595	5	1	0	0	601
Végétation Herbacée	3	595	0	0	0	598
Ombres	1	0	599	0	0	600
Routes/Bâti/Sol nu	1	0	0	580	48	629
Eau	0	0	0	20	552	572
unclassified	0	0	0	0	0	0
Sum	600	600	600	600	600	600
Producer	0.99166666667	0.99166666667	0.99833333333	0.96666666667	0.92	
User	0.99001663894	0.99498327759	0.99833333333	0.92209856916	0.96503496503	
Hellden	0.99084096586	0.99332220367	0.99833333333	0.94385679414	0.94197952218	
Short	0.98184818482	0.98673300166	0.99667221298	0.8936825886	0.89032258065	
KIA Per Class	0.98957899125	0.98959200666	0.99791666667	0.95782370308	0.90115321252	
Overall Accuracy	0.97366666667					
KIA	0.96708333333					

Figure 15 : Matrice de confusion de la classification de la zone de l'Ile Falcon réalisée sur ECognition
Source : HERRIG Nathan, M2 GAED – GEOIDES.

Ecognition est un outil performant pour réaliser des classifications d'images mais il ne nous permet pas de discriminer les végétaux qu'à l'échelle des peuplements et non à l'échelle des espèces. Nous nous sommes donc tournés vers les réseaux de neurones.

3.5 Fonctionnement des Réseaux de Neurones Convolutifs

Les classifications objet et pixel ne permettant pas de discriminer les peuplements à l'échelle des espèces, nous avons décidé d'engager un projet basé sur un code de réseaux de neurones (CNN) pour traiter nos images. L'idée est de fournir à l'ordinateur une banque d'images issues d'orthophotos des différentes espèces au sol, puis une fois le modèle entraîné, de tenter de le faire fonctionner sur une orthophoto pour en faire une carte de classification. Pour cela nous nous sommes rendus sur le site "DATA Scientist" sur lequel nous avons pu apprendre le fonctionnement d'une chaîne de traitements CNN basique comme celle représentée sur la figure 16 ci-dessous (qui représente un exemple simplifié). Celle-ci se décompose selon les étapes suivantes:

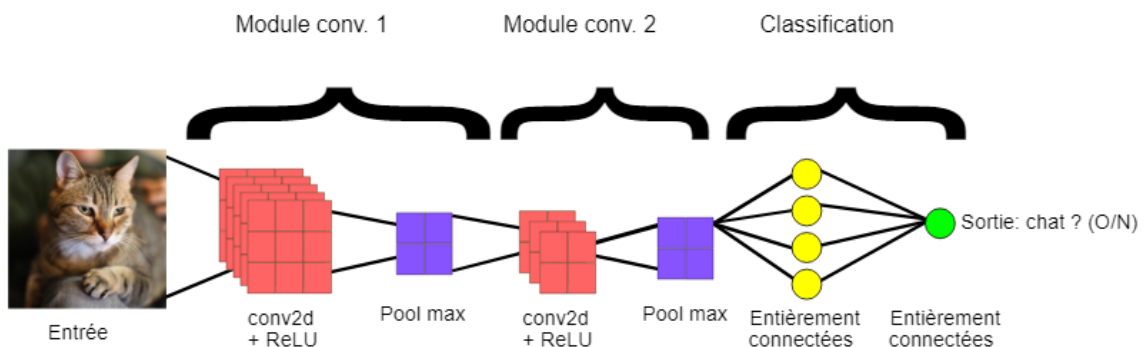


Figure 16 : Schéma d'un réseau de neurones convolutifs classique.
Source: ("ML Practicum,")

1ère étape: La convolution:

Le réseau extrait des tuiles en fonction du nombre de pixels par image, puis leur applique des filtres qui vont lui permettre de créer une première "carte de caractéristiques convoluées".

Cette action va être répétée plusieurs fois car en appliquant ces filtres, le réseau de neurones apprend les valeurs optimales des matrices de ces mêmes filtres. Plus il y a de filtres, plus le CNN est en capacité d'en extraire les caractéristiques, mais plus son temps d'apprentissage augmente (système de rétropropagation).

2ème étape: La fonction ReLu (unité de rectification linéaire)

Celle-ci s'intercale après chaque opération de convolution, le CNN applique une transformation qui introduit la non-linéarité dans le projet (elle consiste à prendre une fonction linéaire et de l'interrompre à la valeur 0, au-dessus de cette valeur, c'est également une fonction linéaire mais si la fonction renvoie à une valeur négative alors celle-ci est limitée à 0 et permet ainsi une fonction non-linéaire comme schématisé sur la figure 17.

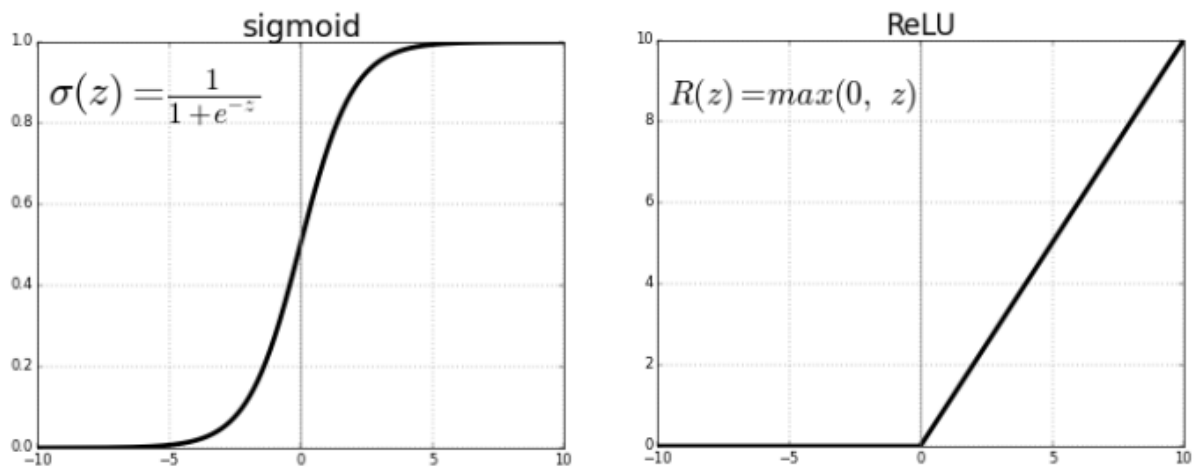


Figure 17 : Schémas des fonctions sigmoïdes et ReLU.
Source: ("Les réseaux de neurones,")

3ème étape: Le Pooling:

La figure 18 représente l'étape du « Pooling », c'est-à-dire que le CNN sous-échantillonne les caractéristiques convoluées diminuant ainsi le nombre de dimensions de la carte de caractéristiques ce qui permet de réduire les temps de traitement.

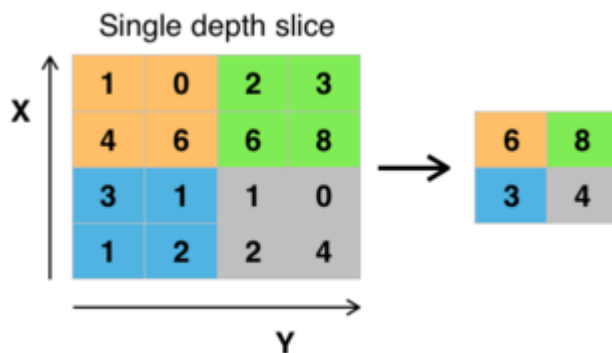


Figure 18 : Schéma de l'étape de « Pooling » dans un réseau de neurones.
Source: ("Convolutional neural network | Deep Learning | DataScientest,")

4ème étape: Les couches connectées:

Quel que soit le nombre de couches, tous les nœuds sont interconnectés, leur rôle est d'effectuer une classification basée sur l'étude des caractéristiques extraites par convolution. On voit sur le schéma ci-dessous (figure 19) les deux entrées qui représentent l'image à analyser et la base d'entraînement dont les caractéristiques vont être comparées dans les couches cachées pour donner un résultat en sortie. Ce résultat va être donné sous forme de probabilités.

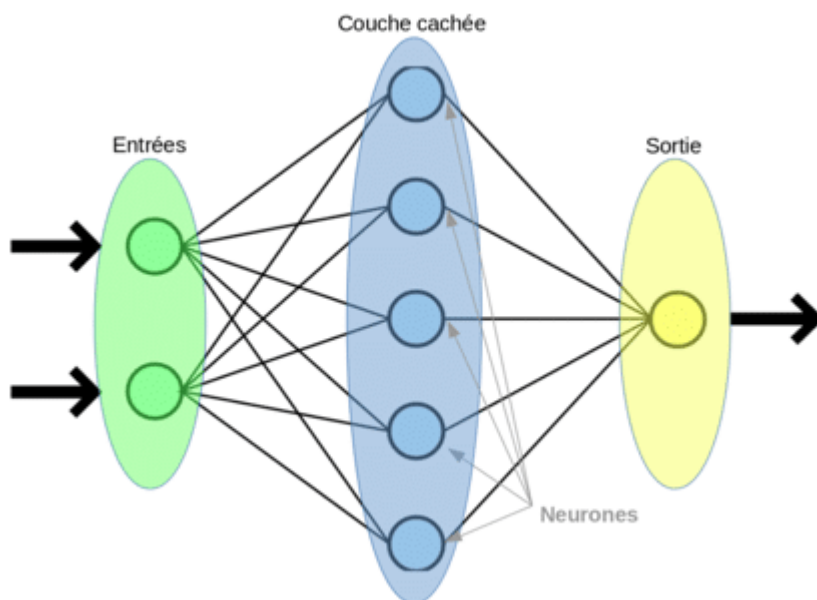


Figure 19 : Schéma des couches connectées dans un réseau de neurones convolutif.
 Source : ("Convolutional neural network | Deep Learning | DataScientest,")

L'objectif de cette chaîne de traitement est donc d'obtenir une carte de pixel où chaque pixel est remplacé par une probabilité, ces probabilités seront ensuite compilées sous forme de statistiques et validées à partir d'un certain seuil.

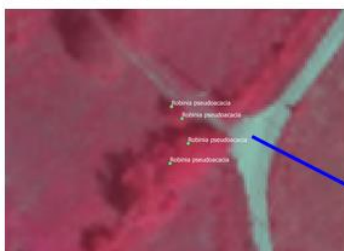
Cela nécessite donc de définir en premier lieu des sites d'études, des espèces à étudier et de par ce biais des types de données à exploiter.

3.6 Préparation des données

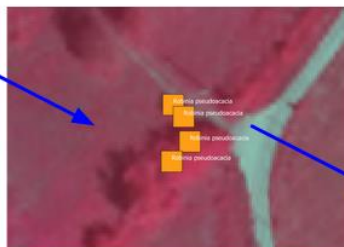
Les données ont été préparées sur le logiciel QGis v.3.18. L'idée première fut de reporter les points GPS sur QGis et d'en faire une couche vecteur de points pour chaque espèce étudiée, puis de créer des zones tampons de 2m sur 2m à partir des points géolocalisés. Ces zones tampons servent d'emprise grâce auxquels on va extraire la partie de l'image en question (*figure 20*).

Traitements (QGIS)

Points RFA Falcon



Tampons RFA Falcon



Imagettes RFA Falcon



Figure 20 : Traitement de données réalisées sur QGis.

Source : HERRIG Nathan, M2 GAED – GEOIDES.

Cependant la forme et la taille des houppiers est extrêmement variable que ce soit d'une espèce à l'autre ou au sein d'une même espèce, les emprises n'étaient donc pas toujours calquées sur les zones escomptées. Ainsi nous avons pris la décision de reporter les points puis de créer manuellement des polygones (voir *figure 21* ci-dessous) de taille variable en fonction de la résolution (la résolution permet d'inclure un certain nombre de pixels par imagettes, en effet, sur une image à 20 cm, un houppier ne prendra que quelques pixels en fonction de la taille de l'arbre, nous avons donc démultiplié notre base de donnée en créant des polygones adaptés en taille aux différentes résolutions). Cela permet également d'extraire plusieurs imagettes sur le même houppier si celui-ci est suffisamment imposant et donc d'augmenter la base de données avec un nombre de points identique.

Nous avons donc créé une première base de 8 classes toutes composées de 300 imagettes d'entraînement sur les orthophotos IGN de la Métropole Grenobloise à 20 cm de résolution. La taille des imagettes a d'abord été définie à 21x21 pixels soit des polygones de 4,20 x 4,20m. Cela a posé une première limite technique: des polygones en 11x11 pixels étaient invisibles car ils n'auraient renvoyé que peu d'informations d'où la décision de commencer directement en 21x21. Cependant cela pose un problème réel; un grand nombre d'espèces étudiées n'ont pas un houppier d'une telle ampleur.

La démarche a donc été la suivante:

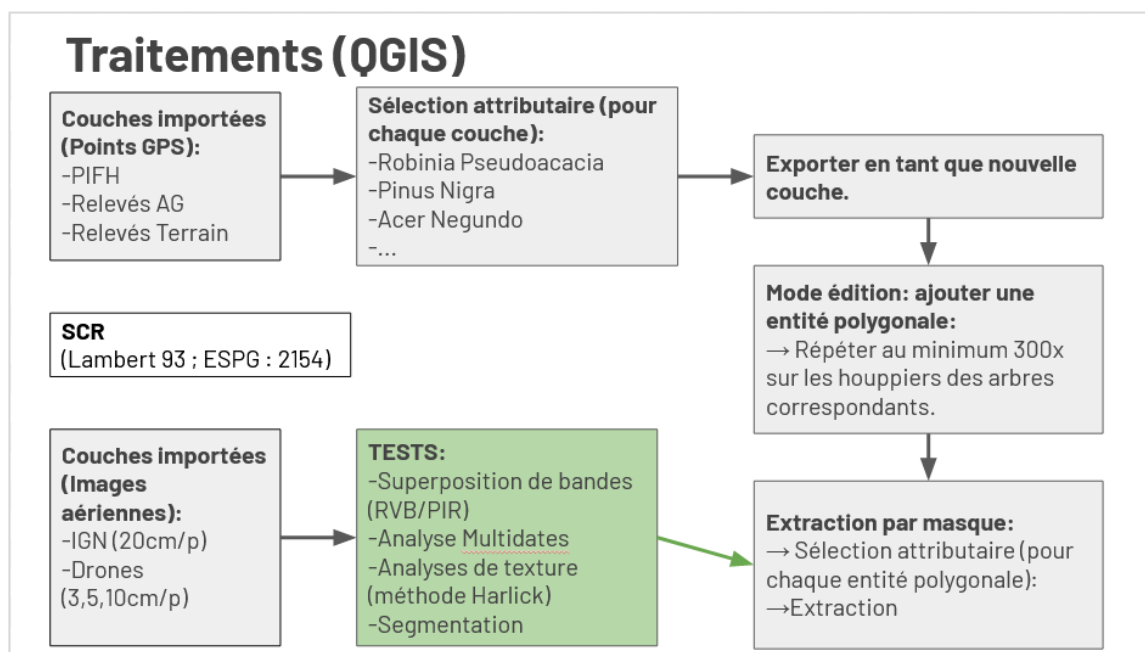


Figure 21 : Traitement de données réalisées sur QGIS.

Source : HERRIG Nathan, M2 GAED – GEOIDES.

En opérant de la sorte on obtient des dossiers de classes d'espèces qui font office de base d'entraînement pour le Machine Learning. Bien entendu, elles doivent être rangées selon une arborescence cohérente.

Comme montré sur le schéma ci-dessus (figure 21), d'autres tests ont été effectués; nous avons tenté d'augmenter le nombre de bandes des imagerie en superposant du proche-infrarouge à du RVB et en effectuant une analyse de texture Haralick, mais malheureusement celles-ci ne fonctionnaient pas dans le code du Machine Learning qui avait déjà commencé à être rédigé à ce moment-là, cependant elles ont été gardées car le code est encore susceptible d'évoluer et elles pourront peut-être à terme servir à augmenter les pourcentages de précision que ce soit sur les bases d'entraînement ou sur les résultats en sortie.

Enfin nous avons également effectué des segmentations afin de créer des objets pixels dont la signature radiométrique est homogène. L'idée était de créer manuellement les polygones d'extraction au sein des polygones de segmentés. Cela a posé deux limites: la première est simplement logique; le réseau de neurones doit être en capacité de reconnaître non seulement une certaine variété d'espèces, mais également de reconnaître tous les individus au sein d'une espèce, nous avons donc émis l'hypothèse qu'il valait mieux introduire une variabilité et non une homogénéité dans la base d'entraînement. La seconde est d'ordre technique; les paramètres d'échelle de la segmentation doivent être réduits au maximum afin que celle-ci cerne correctement la limite des houppiers et la signature radiométrique au sein d'un peuplement peut être très variable, les

polygones étaient donc trop resserrés pour pouvoir y introduire des polygones d'extraction.

Une fois la base de données d'images d'entraînement constituée, nous avons pu l'implémenter au sein du réseau de neurones.

3.7 Choix du CNN

Après quelques recherches effectuées sur différents sites et forums, il nous a semblé pertinent de débiter en Deep Learning avec Keras, un code CNN optimisé pour être accessible aux débutants et visiblement suffisamment performant ("Réseaux de neurones en partant de zéro en Python | by Omar Aflak | France School of AI | Medium," n.d.). A noter qu'un code CNN peut être écrit de bout en bout mais il est plus aisé de prendre un modèle de base et d'y ajouter des implémentations et des modifications en fonction du besoin du projet.

Plusieurs arguments sont mis en avant sur les différents forums visités, en premier lieu le code Keras propose une interface épurée et suffisamment explicite pour être comprise par toute personne s'intéressant aux CNN. Il est également facilement modifiable et permet ainsi d'effectuer des essais rapidement et d'en comprendre les résultats. Keras dispose également d'un certain nombre de modèles pré-entraînés à tester directement pour vérifier son environnement de travail ainsi que pour pouvoir decrypter le code et ses différentes fonctions ("Prototyper un réseau de neurones avec Keras," n.d.).

Nous l'avons importé depuis une bibliothèque Tensorflow qui permet d'ajouter facilement différentes extensions et d'assembler des morceaux de codes. En effet Tensorflow est une librairie qui centralise tous les codes relatifs au CNN, elle permet d'avoir un environnement de travail stable et fonctionnel assez rapidement ainsi que de tester plusieurs langages informatiques tels que Python ou encore C++ ("Et si la création d'un (petit) réseau de neurones artificiels n'était pas si complexe ?," 2017).

3.8 Traitements CNN

Nous nous sommes donc inspirés d'un CNN existant que nous avons adapté à notre projet (code à retrouver en annexe n°2).

La première étape a été de compiler notre base de données au format "numpy", c'est un format structuré en tableau unidimensionnel dans lequel on retrouve les fichiers images associés à un identifiant ce qui permet au CNN de définir les classes intégrées.

La seconde étape est de copier le code, (annexe n°3). Le code Keras est simple mais est écrit pour un jeu de données type qui fonctionne immédiatement et qui permet aux nouveaux utilisateurs de s'entraîner sur une base de données pré-entraînée. Il nécessite donc d'être adapté à notre propre jeu de données,

celui-ci étant différent du jeu de données type et nécessitant donc des traitements différents.

Une fois le code écrit, nous avons effectué des premiers tests sur la base d'entraînement associée aux images aériennes IGN de la métropole Grenobloise à 20 cm de résolution (imassettes en 21x21/pixel). La base d'entraînement est divisée en deux parties; une partie de données d'entraînement et une partie de données de vérification.

Nous sommes arrivés à une précision ne dépassant pas les 70% de précision sur la base d'entraînement ce qui est insuffisant, nous avons donc déduit que les imassettes d'entraînement en 21x21/pixel ne contenaient pas suffisamment d'informations pour en tirer des spécificités d'une classe à l'autre.

Ce test a marqué un tournant dans nos essais; en effet les résultats étant peu probants nous nous sommes concentrés sur notre seconde base de données moins fournie constituée sur la base des orthophotos de réalisées grâce aux vols de drones sur la zone de compensation écologique de l'île Falcon mais contenant des imassettes en 51x51/pixel pour une résolution de 3,5 cm. Nous avons lancé le même test et avons immédiatement obtenu des résultats allant de 70 à 80% de précision sur la base d'entraînement.

Les résultats étant un peu plus encourageants, nous avons pris la décision d'augmenter ce jeu de données car plus la base donnée au CNN est fournie, plus celui-ci est en mesure d'apprendre et d'en dégager des caractéristiques spectrales spécifiques pour chaque classe. Il a donc fallu effectuer à nouveau les traitements évoqués dans la partie 3.6.

Une fois la base augmentée et les premiers tests effectués nous avons testé de nouveaux paramètres au sein du code pour tenter d'augmenter le pourcentage de précision sur la base d'entraînement.

L'idée première fut d'effectuer des "rotations" et des "flips" pour démultiplier notre base de données. Le concept est simple: les imassettes sont extraites de polygones carrés et présentent donc un certain nombre de pixels qui renvoient des informations spectrales. Si on effectue une rotation sur ces imassettes la série d'informations bascule et permet donc d'augmenter la base. Les "flips" fonctionnent sur le même principe que les "rotations" sauf qu'au lieu de faire tourner l'imassette on la retourne inversant ainsi les valeurs comprises dans les pixels eux-même compris dans les imassettes. Ces deux méthodes permettent donc de rendre les bases de données plus conséquentes et ainsi de fournir le modèle à entraîner.

Nous avons ensuite testé différentes valeurs pour le "batch processing" et pour les "epochs".

Le "batch processing" consiste à envoyer les imassettes d'entraînement par lots, c'est-à-dire qu'on l'on propage un certain nombre d'échantillons (ici les imassettes) par succession de groupes dont le nombre est défini au préalable. Cela permet de réduire le besoin de mémoire de calcul pour l'ordinateur car il ingère la base d'entraînement par système de vagues. Le modèle s'entraîne

donc plus rapidement, et ainsi le poids des paramètres se met à jour après le passage de chaque lot ce qui permet d'éviter le surentraînement. Néanmoins, si les lots ou les "packages" sont de taille trop réduite, le modèle risque de créer des erreurs dans la rétropropagation du gradient. Cette technique consiste à corriger les erreurs selon l'importance de la contribution de chaque élément aux erreurs. Le batch size désigne donc la taille des lots envoyés au CNN.

En langage informatique, une "epoch" (ou époque en français) désigne une passe avant et une passe arrière. Pour donner un exemple simple reprenons le batch size évoqué ci-dessus: si la taille de l'échantillon est de 100 et que le batch size est de 10 alors il faudra 10 itérations pour terminer une "epoch". On peut ensuite démultiplier ces "epochs" qui vont déterminer le nombre de passes (avant/arrière) et donc le nombre de fois où l'échantillon total va être analysé.

Enfin, le « dropout » est une commande qui désactive des neurones dans les itérations de manière aléatoire et selon une probabilité définie au préalable. Cela permet d'entraîner plusieurs modèles en parallèle ce qui permet à chaque neurone un apprentissage complet en évitant ce qu'on appelle la coadaptation c'est-à-dire la corrélation d'apprentissage entre les neurones proches, cela leur fournit une certaine indépendance.

IV. Résultats & discussions

4.1 Analyse des résultats

Voici les résultats obtenus après avoir testé notre base d'entraînement d'images en 51x51/pixels testée avec les différents paramètres évoqués ci-dessus.

Bach Size	Epochs	Résultats	Paramètres
10	20	0.79	Avec 600 images tests/4 classes
10	20	0.76	Avec 600 images tests/4 classes
10	20	0.74	Avec 600 images tests/4 classes + Sans le dropout
10	20	0.76	Avec 600 images tests/4 classes + Ajout dropout entre les convolutions
10	50	0.82	Avec 600 images tests/4 classes
10	70	0.81	Avec 600 images tests/4 classes
20	20	0.83	Avec 600 images tests/4 classes
20	20	0.79	Avec 400 images tests/4 classes
20	20	0.81	Avec 600 images tests/4 classes + 3 rotations
20	20	0.821	Avec 600 images tests/4 classes + 3 rotations + 3 flips
20	30	0.81	Avec 600 images tests/4 classes
20	30	0.807	Avec 600 images tests/4 classes
20	30	0.826	Avec 600 images tests/4 classes + 3 rotations
20	40	0.78	Avec 600 images tests/4 classes + 3 rotations
20	50	0.83	Avec 600 images tests/4 classes
20	120	0.837	Avec 600 images tests/4 classes
20	120	0.843	Avec 600 images tests/4 classes
20	200	0.843	Avec 600 images tests/4 classes
20	1000	0.832	Avec 600 images tests/4 classes
30	30	0.76	Avec 600 images tests/4 classes
40	100	0.854	Avec 600 images tests/4 classes
50	40	0.8	Avec 600 images tests/4 classes
100	40	0.76	Avec 600 images tests/4 classes

Figure 22 : Tableau récapitulatif des résultats CNN obtenus en fonction de différents paramètres. Source : MALDONADO Éric, INRAE & HERRIG Nathan, M2 GAED – GEOIDES.

On constate sur le tableau ci-dessus (*figure 22*) la totalité des résultats fluctue entre 75% de précision et 85% de précision sur la base d'entraînement. Ces résultats sont encore insuffisants pour créer un outil fonctionnel mais sont néanmoins très encourageants pour la suite, ils montrent que le modèle est capable de reconnaître à plus de 80% (avec certains paramètres) les caractéristiques composantes des différentes classes d'arbres qui sont ici le Buddleia de David, le Robinier-Faux-Acacia, le Frêne Élevé et le Tilleul à Petites Feuilles. Cela signifie que les caractéristiques extraites permettent d'identifier à hauteur de plus de 4/5 les espèces étudiées entre elles et au sein d'une même classe. Cela démontre néanmoins le besoin d'améliorer encore la base d'entraînement afin de pouvoir se diriger vers une classification suffisamment fiable pour être utilisée sur des cas concrets.

Le premier constat est le suivant: plus la base d'entraînement est fournie, plus celle-ci est fonctionnelle. En effet, cela permet au CNN d'en extraire un plus grand nombre de caractéristiques associées et ainsi d'introduire une variabilité qui permet de reconnaître plus d'individus que ce soit au sein d'une classe ou d'une classe à l'autre. Cela démontre la nécessité d'une phase de terrain élaborée permettant d'avoir une base de données géolocalisée complète. Cela démontre également la nécessité d'utiliser des méthodes de "Data Augmentation" qui permettent de gonfler les bases de données et qui peuvent pallier des problèmes de main d'œuvre ou de matériel. A noter tout de même que la base de données doit absolument être fiable afin de ne pas courir le risque d'entraîner un modèle basé sur des informations erronées.

En revanche, difficile d'établir un lien entre précision de l'outil et paramétrage des "epochs" et du "batch size". Un certain nombre de tests ont été effectués et nous n'avons pu trouver un paramétrage qui fasse significativement augmenter la précision de la reconnaissance sur la base d'entraînement.

Enfin, le dropout ne montre pas non plus de résultats significatifs en termes de précision mais celui-ci semble nécessaire pour éviter le surentraînement du modèle pouvant générer des erreurs dans les résultats en surévaluant les caractéristiques de classe.

4.2 Quelle suite pour l'outil?

Jusqu'ici les résultats sont prometteurs, bien qu'ils ne dépassent pas les 90% de précision nous avons tout de même pu tirer des premières conclusions qui nous permettent de faire le point et ainsi réfléchir à de nouvelles pistes pour l'améliorer.

L'idée pour la suite serait de produire une carte de classification en sortie. Pour cela il sera nécessaire de continuer à développer le code et la base de données pour arriver à ce résultat. Deux types de cartes seraient envisageables:

- Une carte de probabilités d'occurrences de classes, c'est à dire que l'on insère en entrée (dans le CNN) une image à analyser (pour commencer les orthophotos drone de l'île Falcon) à laquelle il applique les différents traitements (convolution) et qui nous permettrait d'obtenir en sortie cette même image mais où les pixels sont remplacés par des probabilités. Il s'agirait ensuite de définir un seuil où les probabilités sont suffisantes pour affirmer qu'il y a bien telle ou telle espèce à tel ou tel endroit.
- L'autre possibilité serait d'avoir une carte binaire en sortie où le CNN nous donne simplement pour chaque pixel la mention TRUE ou FALSE pour une classe prédéfinie et à partir d'un seuil donné. A noter que ces deux solutions nécessitent également d'établir un protocole pour les recoloriser et en faire des cartes de classification vecteur ou raster.

Quel que soit le type de carte en sortie, celle-ci nécessiterait une seconde carte de vérification superposable. Pour cela nous avons commencé à créer une carte de numérisation manuelle en vecteur, c'est-à-dire que nous avons repris tous les points géolocalisés et identifié les espèces recensées, puis les avons ensuite polygones. L'idée est ensuite de remplir ces polygones de points de vérification à comparer avec les cartes en sortie.

4.3 Les limites de l'outil

Malgré l'avancée majeure que représentent les CNN en termes d'analyse d'images, ceux-ci peuvent présenter quelques limites à étudier.

Comme exprimé dans l'analyse des résultats, il est de rigueur de créer une base de données fiable afin de ne pas biaiser le modèle. La première limite n'est donc pas informatique mais humaine, il est nécessaire de définir au préalable une technique d'exploitation de données afin d'introduire une homogénéité dans la méthode.

La seconde limite est d'ordre quantitative; en effet, un tel travail nécessite en premier lieu de la main d'œuvre terrain pour pouvoir collecter efficacement un certain nombre de points géolocalisés et pour pouvoir quadriller le terrain étudié permettant la reconnaissance d'un maximum d'espèces et évitant ainsi les No Data et la surévaluation en sortie de modèle.

Un réseau de neurones nécessite également une puissance de calcul importante, plus le modèle est fourni plus le besoin augmente. Aujourd'hui peu d'organismes disposent réellement de puissance de calcul suffisante pour établir des modèles à grande échelle (espace métropole), l'exemple le plus courant reste aujourd'hui celui de Google avec la création de la voiture autonome celle-ci étant basée sur un modèle de Deep Learning et dont la base de données est fournie tous les jours par nos soins lorsque nous utilisons leur moteur de recherche (lorsque nous cliquons sur des images de passage piéton pour accéder à un site nous entraînons leur modèle). Cela donne une idée simple des moyens considérables à développer pour créer un modèle complexe sur un cas concret.

Une limite est également survenue dans le prétraitement des données ; les imagerie d'entraînement nécessitent d'être en nombre suffisant. Comme expliqué précédemment les imagerie en 21x21/pixels se sont avérées insuffisantes en termes d'informations (bien que la base n'ai pas été totalement inefficace les résultats étaient trop peu probants pour en tirer des conclusions). A noter que cette limite n'est pas valable pour toute étude réalisée par CNN mais les études portées sur le végétal nécessitent une quantité d'information importante du fait de la variabilité importante des caractéristiques physiques qui les composent (phénologie, taille, etc.).

Enfin, nous avons rencontré un problème de luminosité sur les images de drones créées par nos soins. En effet, celle-ci varie en fonction de la météo journalière, les imagerie d'une espèce extraites sur des orthophotos prises à des dates différentes peuvent présenter des informations spectrales différentes. Néanmoins cela présente tout de même un point positif qui est l'augmentation de l'amplitude d'information au sein des classes d'arbres concernées. C'est-à-dire que si le modèle est suffisamment fourni, il pourrait à terme, être en mesure d'identifier les espèces en question sur des images présentant des caractéristiques lumineuses différentes.

4.4 Quelles pistes pour l'améliorer?

Pour dépasser les 90% de précision sur la base d'entraînement nous avons réfléchi à quelques pistes pour l'améliorer.

En premier lieu nous estimons qu'il sera important pour la suite d'ajouter de l'information en ajoutant des bandes dans les couches (bien que cela implique de modifier le code). Pour cela nous pouvons envisager plusieurs solutions comme une analyse multispectrale en ajoutant par exemple une couche de proche infrarouge, ou une analyse de texture Haralick qui permet de dégager des informations texturales (*Figure 23*).

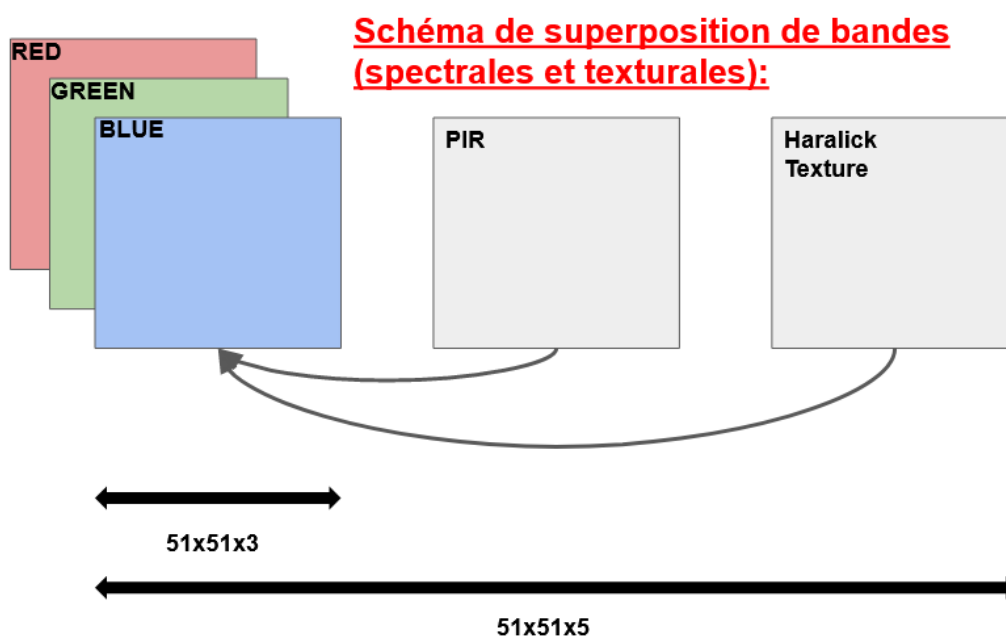


Figure 23 :

Source :

Ensuite, il y'a une nécessité d'avoir des bases de données conséquentes. Les méthodes de Data Augmentation sont nombreuses et n'ont pas toutes été appliquées, on peut donc envisager des techniques nouvelles telles que la multi-visualisation d'un objet. C'est-à-dire que l'on reprend les images brutes prises (ici par le drone) et qu'on en extrait l'emprise voulue pour créer des imagettes d'entraînement. Par ce biais on peut facilement multiplier une base de données par 10 (Liu and Abd-Elrahman, 2018).

Une telle analyse de données nécessite des machines performantes et bien que les machines locales soient relativement efficaces les temps de traitement peuvent s'avérer extrêmement longs. Pour pallier ce problème nous pouvons réutiliser les classifications tests dont les résultats étaient satisfaisants pour pondérer les différentes classes en leur donnant toutes une valeur de 0 sauf pour la végétation haute (valeur 1) ce qui permet à la machine de se

concentrer sur les zones d'intérêt pour notre projet. Nous avons obtenu la carte de masque suivante (figure 24):

Couche de Masque pour extraction de la végétation (Métropole Grenobloise):

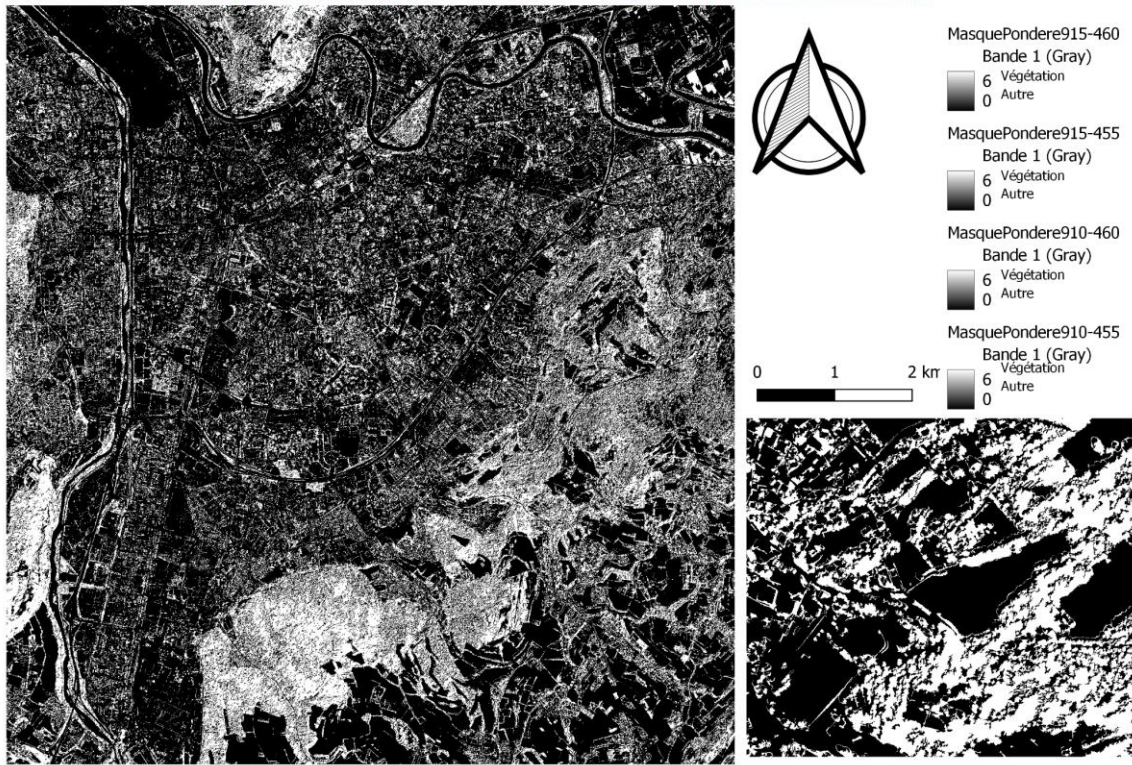


Figure 24 : Couche de masque sur la zone de la Métropole Grenobloise réalisée sur QGIS
Source : HERRIG Nathan, M2 GAED – GEOIDES.

Une autre idée est d'ajouter une implémentation 3D ou des modèles de canopée (CHM) ce qui impliquerait une nouvelle prise d'images consécutive pour établir un modèle propre. Cela nous permettrait d'ajouter des caractéristiques de taille et ainsi d'analyser la végétation par strates.

L'idée suivante est de ramener l'étude à une échelle plus réduite sur un panel d'espèces complet. C'est à dire définir une zone réduite (par exemple de 100m sur 100m) et d'y faire un inventaire floristique complet nous permettant à la fois de fournir le CNN d'une base pouvant couvrir la zone sans "no data" et nous permettant de construire facilement une carte de vérification des résultats.

Enfin, l'apparition de techniques nouvelles ouvre certaines perspectives. On peut parler par exemple du transfert learning qui est la suite logique des réseaux de neurones convolutifs. Cela consiste à réutiliser des résultats obtenus par d'autres modèles et à en faire une base pour le modèle suivant. Cette méthode est une fois de plus fortement inspirée des réseaux cognitifs de l'être humain. En effet, de nombreuses études ont démontré qu'un enfant qui apprend plusieurs langues pendant sa période d'apprentissage aura moins de difficultés à en apprendre d'autres à l'âge adulte: le transfert learning suit cette logique d'apprentissage ("Transfer Learning," 2020).

VI. Conclusion

Cette étude a donc pour objectif de créer un outil de télédétection-classification par Réseau de neurones convolutifs pour les espèces invasives du

De nombreuses pistes ont été explorées puis abandonnées en cours de route tandis que d'autres sont plutôt prometteuses. Il faut donc retenir qu'un outil de cette ampleur nécessite des bases de données conséquentes et fiables ainsi que des images d'une qualité suffisante et présentant un certain nombre d'informations.

En conclusion on peut affirmer que les CNN sont un outil d'avenir dans la reconnaissance d'espèces végétales par télédétection, bien que de nombreuses limites techniques soient encore présentes. On voit toutefois les études sur le sujet progresser d'années en années. Ils seront à l'avenir un véritable outil d'analyse et d'aide à la décision dans la gestion de la biodiversité et notamment dans la gestion des espèces invasives. L'identification et le suivi de celles-ci vont représenter dans le futur un véritable enjeu quant à la gestion des espaces naturels. Leur propagation soulève de nombreuses questions quant au rôle qu'elles auront à jouer dans l'évolution des écosystèmes face au changement climatique.

Cette étude montre également la nécessité d'associer informatique et écologie et donc d'engager une passerelle multidisciplinaire pour pallier aux difficultés environnementales actuelles. Bien que l'informatique propose de nos jours de nombreuses solutions théoriques pour tous types de thématiques, on comprend que dans des cas concrets tels que l'étude des végétaux il est nécessaire d'y ajouter une expertise de la part d'agents compétents pour appliquer les solutions envisagées sur le terrain.

NB: Ce mémoire ne marque pas la fin de l'étude, j'ai le plaisir de pouvoir partir dans un nouveau Master à l'IM2AG sur le campus de Grenoble dans lequel je vais pouvoir élargir mon panel de compétences en informatique qui vont m'être nécessaires pour continuer à développer cet outil.

Bibliographie:

NB : La majeure partie de la bibliographie est constituée de divers sites internet et forums les informations recherchées étant des informations techniques concernant principalement les réseaux de neurones et les fonctions qui les régissent.

Articles :

- Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, Ö., Yu, R., Zhou, B. (Eds.), 2021. « Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. » *Cambridge University Press*.
- Mayor, J.R., Sanders, N.J., Classen, A.T., Bardgett, R.D., Clément, J.-C., Fajardo, A., Lavorel, S., Sundqvist, M.K., Bahn, M., Chisholm, C., Cieraad, E., Gedalof, Z., Grigulis, K., Kudo, G., Oberski, D.L., Wardle, D.A., 2017. « Elevation alters ecosystem properties across temperate treelines globally. » *Nature* 542, 91–95.
- Tallent, N., Watt, M., 2009. « The Invasive *Buddleja davidii* (Butterfly Bush). » *Bot. Rev.* 75, 292–325.
- Lévêque, C., Tabacchi, É., Menozzi, M.-J., 2012. « Les espèces exotiques envahissantes, pour une remise en cause des paradigmes écologiques. » *Sci. Eaux Territ.* Numéro 6, 2–9.
- Alexander, J.M., Lembrechts, J.J., Cavieres, L.A., Daehler, C., Haider, S., Kueffer, C., Liu, G., McDougall, K., Milbau, A., Pauchard, A., Rew, L.J., Seipel, T., 2016. Plant invasions into mountains and alpine ecosystems: current status and future challenges. « *Alp. Bot.* » 126, 89–103.
- Pimentel, D., Lach, L., Zuniga, R., Morrison, D., 2000. « Environmental and Economic Costs of Nonindigenous Species in the United States. » *BioScience* 50, 53–65. 2
- Mirik, M., Ansley, R.J., Steddom, K., Jones, D.C., Rush, C.M., Michels, G.J., Elliott, N.C., 2013. « Remote Distinction of A Noxious Weed (Musk Thistle: *Carduus Nutans*) Using Airborne Hyperspectral Imagery and the Support Vector Machine Classifier. » *Remote Sens.* 5, 612–630
- Huang, C., Asner, G.P., 2009. « Applications of Remote Sensing to Alien Invasive Plant Studies. » *Sensors* 9, 4869–4889.

- Cox, M., Ellsworth, D., 1997. « Application-controlled demand paging for out-of-core visualization, in: Proceedings. » *Visualization '97* pp. 235-244,.
- Nijhawan, R., Sharma, H., Sahni, H., Batra, A., 2017. « A Deep Learning Hybrid CNN Framework Approach for Vegetation Cover Mapping Using Deep Features, » in: *International Conference on Signal-Image Technology Internet-Based Systems (SITIS)*, pp. 192–196.
- Zhang, J., Wang, C., Li, C., Qin, H., 2019. « Example-based rapid generation of vegetation on terrain via CNN-based distribution learning. » *Vis. Comput.* 35, 1181–1191.
- Langford, Z.L., Kumar, J., Hoffman, F.M., Breen, A.L., Iversen, C.M., 2019. « Arctic Vegetation Mapping Using Unsupervised Training Datasets and Convolutional Neural Networks. » *Remote Sens.* 11, 69.
- Ayhan, B., Kwan, C., Budavari, B., Kwan, L., Lu, Y., Perez, D., Li, J., Skarlatos, D., Vlachos, M., 2020. « Vegetation Detection Using Deep Learning and Conventional Methods. *Remote Sens.* » 12, 2502.
- Heinzl, J., Koch, B., 2012. « Investigating multiple data sources for tree species classification in temperate forest and use for single tree delineation. » *Int. J. Appl. Earth Obs. Geoinformation* 18, 101–110.
- Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. « Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. » *ISPRS J. Photogramm. Remote Sens.* 173, 24–49.
- Hasan, M., Ullah, S., Khan, M.J., Khurshid, K., 2019. « COMPARATIVE ANALYSIS OF SVM, ANN AND CNN FOR CLASSIFYING VEGETATION SPECIES USING HYPERSPPECTRAL THERMAL INFRARED DATA, » in: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. pp. 1861–1868.
- Chambras, M., 2019, « Dynamique des populations de robinier faux-acacia (*Robinia pseudoacacia*) en régions Nord-Pas-de-Calais et Picardie 48. »
- Dumas, Y., 2019. « Que savons-nous de l'Érable négondo *Acer negundo* L. ? » *Naturae*.
- Liu, T., Abd-Elrahman, A., 2018. « Deep convolutional neural network training enrichment using multi-view object-based analysis of Unmanned Aerial systems imagery for wetlands classification. » *ISPRS J. Photogramm. Remote Sens.* 139, 154–170.

Sitographie :

- Les Alpes, une région menacée | WWF France [WWW Document],
URL <https://www.wwf.fr/espaces-prioritaires/alpes>
- Espèces exotiques envahissantes en Auvergne - Pourquoi ce site ?,
Espèces Exot. Envah. En Auvergne-Rhône-Alpes.
URL <http://eee-auvergnerhonealpes.fr/wordpress/index.php/presentation/>
- Risk and opportunity of nnt – Alpine Space [WWW Document],
URL <https://www.alpine-space.eu/projects/alptrees/en/home/risk-and-opportunity-of-nnt>
- About – Alpine Space [WWW Document],
URL <https://www.alpine-space.eu/projects/alptrees/en/home/about>
- Larousse, É., Alpes - LAROUSSE [WWW Document].
URL <https://www.larousse.fr/encyclopedie/mont/Alpes/104958>
- Work packages – Alpine Space [WWW Document],
URL https://www.alpine-space.eu/projects/alptrees/en/the-project/work-packages?fbclid=IwAR1M86vxwShwGoH4SIBeS4i_OrofSZipCHBQDozGyMShUUC-NtIL28NJ_Z8
- SIG : Tout savoir sur les Systèmes d'Information Géographique | Esri France [WWW Document],
URL <https://www.esrifrance.fr/sig11.aspx>
- SIG Système d'Information Géographique ou Géomatique : outils de cartographie numérique, représentation spatiale des informations [WWW Document],
URL <http://www.sig-geomatique.fr/sig-sig.html>
- Larousse, É., Définitions : télédétection - Dictionnaire de français Larousse [WWW Document].
URL <https://www.larousse.fr/dictionnaires/francais/t%C3%A9l%C3%A9d%C3%A9tection/77051>
- Intelligence Artificielle, Machine Learning et Deep Learning : Quels liens et quelles différences ?, 2019. Geek Mais Pas Que.
URL <https://www.geekmaispasque.com/2019/10/intelligence-artificielle-machine-learning-deep-learning-liens-differences/>
- R, L., 2018. Différence entre Intelligence Artificielle, Machine Learning et Deep Learning. Pensée Artif.
URL <https://penseeartificielle.fr/difference-intelligence-artificielle-machine-learning-deep-learning/>
- Les réseaux de neurones [WWW Document],
URL https://ml4a.github.io/ml4a/fr/neural_networks/

- MESSAOUDI, 2019. Le Deep Learning et les SIG s'associent pour des résultats meilleurs [WWW Document].
URL <https://veillecarto2-0.fr/2019/10/24/le-deep-learning-et-les-sig-sassocient-pour-des-resultats-meilleurs/>
- Pinus nigra Arnold [WWW Document],
URL http://archive.wikiwix.com/cache/index2.php?url=http%3A%2F%2Fwww.na.fs.fed.us%2Fpubs%2Fsilvics_manual%2FVolume_1%2Fpinus%2Fnigra.htm
- Le tilleul à petites feuilles (Tilia cordata) [WWW Document],
URL https://viagallica.com/v/tilleul_a_petites_feuilles.htm
- Frêne élevé, 2021. Wikipédia.
URL https://fr.wikipedia.org/wiki/Fr%C3%AAne_%C3%A9lev%C3%A9
- IBM Docs [WWW Document], 2021.
URL <https://prod.ibmdocs-production-dal-6099123ce774e592a519d7c33db8265e-0000.us-south.containers.appdomain.cloud/docs/fr/spss-modeler/SaaS?topic=models-how-svm-works>
- Méthode des k plus proches voisins, 2021. Wikipédia.
URL https://fr.wikipedia.org/wiki/M%C3%A9thode_des_k_plus_proches_voisins
- Réseaux de neurones en partant de zéro en Python | by Omar Aflak | France School of AI | Medium [WWW Document],
URL <https://medium.com/france-school-of-ai/math%C3%A9matiques-des-r%C3%A9seaux-de-neurones-code-python-613d8e83541>
- Prototyper un réseau de neurones avec Keras [WWW Document], Dieux Code.
URL <https://lesdieuxducode.com/blog/2019/1/prototyper-un-reseau-de-neurones-avec-keras>
- Et si la création d'un (petit) réseau de neurones artificiels n'était pas si complexe ? [WWW Document], 2017.
URL <https://www.imagile.fr/creation-dun-petit-reseau-de-neurones-artificiels-netait-complexe/>
- Transfer Learning: Qu'est-ce que c'est ?, 2020. Form. Data Sci. DataScientestcom.
URL <https://datascientest.com/transfer-learning>

Sources images :

- About – Alpine Space [WWW Document],
URL <https://www.alpine-space.eu/projects/alptrees/en/home/about>
- Le robinier faux-acacia (*Robinia pseudoacacia*) [WWW Document],
URL https://viagallica.com/v/robinier_faux-acacia.htm
- 8 idées de Acer negundo | feuille d'érable, érable negundo, feuille [WWW Document], Pinterest.
URL <https://www.pinterest.fr/antoinepinel1/acer-negundo/>
- Pin noir d'Autriche ou noir pine, *Pinus nigra* subsp. *laricio* (*Pinus laricio*). Chromolithographie après une illustration botanique de Hermann Adolph Koehler's *Plantes médicinales*, sous la direction de Gustav Pabst, Koehler, Allemagne, 1887 Photo Stock [WWW Document]. Alamy.
URL <https://www.alamyimages.fr/photo-image-pin-noir-d-autriche-ou-noir-pine-pinus-nigra-subsp-laricio-pinus-laricio-chromolithographie-apres-une-illustration-botanique-de-hermann-adolph-koehlers-plantes-medicinales-sous-la-direction-de-gustav-pabst-koehler-allemande-1887-102721748.html>
- *Buddleja davidii* illustration - circa 1898 | *Buddleja davidii*, Botanical illustration, Botanical drawings [WWW Document],
URL
[https://www.pinterest.fr/pin/766386061573067300/?amp_client_id=CLIENT_ID\(&mweb_unauth_id=&url=https%3A%2F%2Fwww.pinterest.fr%2F&pin%2F766386061573067300%2F&expand=true](https://www.pinterest.fr/pin/766386061573067300/?amp_client_id=CLIENT_ID(&mweb_unauth_id=&url=https%3A%2F%2Fwww.pinterest.fr%2F&pin%2F766386061573067300%2F&expand=true)
- Le tilleul à petites feuilles (*Tilia cordata*) [WWW Document],
URL https://viagallica.com/v/tilleul_a_petites_feuilles.htm
- Frêne | Remèdes de Grand-mère | Ash leaf, Botanical flowers, Botanical illustration [WWW Document],
URL <https://www.pinterest.fr/pin/310185493089269043/>
- ML Practicum: Image Classification [WWW Document], Google Dev.
URL <https://developers.google.com/machine-learning/practica/image-classification/convolutional-neural-networks?hl=fr>
- Les réseaux de neurones [WWW Document],
URL https://ml4a.github.io/ml4a/fr/neural_networks/
- Convolutional neural network | Deep Learning | DataScientest [WWW Document],
URL <https://datascientest.com/convolutional-neural-network>

Table des figures:

- **Figure 1** : Carte de présentation des Alpes faite par ALPTREES.
Source: (“about – Alpine Space,”)_____ **8**
- **Figure 2** : Carte des sites retenus pour l'étude.
Source : HERRIG Nathan, M2 GAED-GEOIDES._____ **15**
- **Figure 3** : Espèces retenues pour l'étude.
Sources images : (“Le robinier faux-acacia (Robinia pseudoacacia),”), (“8 idées de Acer negundo | feuille d'érable, érable negundo, feuille,”), (Pin noir,)_____ **17**
- **Figure 4** : Espèces retenues pour l'étude.
Sources images : (“Buddleja davidii illustration - circa 1898 | Buddleja davidii, Botanical illustration, Botanical drawings,”), (“Le tilleul à petites feuilles (Tilia cordata),”), (“Frêne | Ash leaf, Botanical flowers, Botanical illustration,”)_____ **18**
- **Figure 5** : Capture d'écran de la réalisation d'une orthophotographie sur Agisoft Metashape.
Source : HERRIG Nathan, M2 GAED – GEOIDES._____ **21**
- **Figure 6** : Orthophotos de la zone de l'île Falcon réalisées à partir de vols de drones.
Source : BORGNIET Laurent, INRAE & HERRIG Nathan, M2 GAED – GEOIDES. _____ **21**
- **Figure 7** : Tableau des données principales utilisées durant la réalisation de l'étude.
Source : HERRIG Nathan, M2 GAED – GEOIDES. _____ **22**
- **Figure 8** : Tableau récapitulatif des algorithmes de classification testés sur différents logiciels.
Source : HERRIG Nathan, M2 GAED – GEOIDES._____ **23**
- **Figure 9** : Classification de la Métropole Grenobloise réalisée sur OTB.
Source : HERRIG Nathan, M2 GAED – GEOIDES. _____ **24**
- **Figure 10** : Nomenclature de la classification de la Métropole Grenobloise réalisée sur DZETSAKA.
Source : HERRIG Nathan, M2 GAED – GEOIDES._____ **25**
- **Figure 11** : Classification de la Métropole Grenobloise réalisée sur DZETSAKA.
Source : HERRIG Nathan, M2 GAED – GEOIDES. _____ **26**
- **Figure 12** : Classification de la Métropole Grenobloise réalisée sur ECognition.
Source : HERRIG Nathan, M2 GAED – GEOIDES._____ **27**
- **Figure 13** : Matrice de confusion de la classification de la zone de Grenoble réalisée sur Ecognition
Source : HERRIG Nathan, M2 GAED – GEOIDES._____ **27**

- **Figure 14** : Classification de la zone de l'île Falcon réalisée sur ECognition.
Source : HERRIG Nathan, M2 GAED – GEOIDES. _____ **28**
- **Figure 15** : Matrice de confusion de la classification de la zone de l'île Falcon réalisée sur Ecognition
Source : HERRIG Nathan, M2 GAED – GEOIDES. _____ **28**
- **Figure 16** : Schéma d'un réseau de neurones convolutifs classique.
Source: ("ML Practicum,") _____ **29**
- **Figure 17** : Schémas des fonctions sigmoïdes et ReLu.
Source: ("Les réseaux de neurones,") _____ **30**
- **Figure 18** : Schéma de l'étape de « Pooling » dans un réseau de neurones.
Source: ("Convolutional neural network | Deep Learning | DataScientest,") _____ **30**
- **Figure 19** : Schéma des couches connectées dans un réseau de neurones convolutif.
Source : ("Convolutional neural network | Deep Learning | DataScientest,") _____ **31**
- **Figure 20** : Traitements des données effectuées sur Qgis.
Source : HERRIG Nathan, M2 GAED - GEOIDES. _____ **32**
- **Figure 21** : Traitements des données effectuées sur Qgis.
Source : HERRIG Nathan, M2 GAED - GEOIDES. _____ **33**
- **Figure 22** : Tableau récapitulatif des résultats CNN obtenus en fonction de différents paramètres.
Source : MALDONADO Eric, INRAE & HERRIG Nathan, M2 GAED – GEOIDES _____ **37**
- **Figure 23** : Schéma de superposition de bandes (spectrales & texturales)
Source : HERRIG Nathan, M2 GAED – GEOIDES. _____ **40**
- **Figure 24** : Couche de masque sur la zone de la Métropole Grenobloise réalisée sur QGis
Source : HERRIG Nathan, M2 GAED – GEOIDES. _____ **41**

ANNEXE n°1 : Présentation de l'organisme d'accueil

Mon stage s'est déroulé à INRAE, l'Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement sur la période d'Avril à Août 2021.

INRAE est donc un institut de recherche issu de la récente fusion (2020) entre INRA (l'Institut National de la Recherche pour l'Agronomie) et IRSTEA (l'Institut national de Recherche en Sciences et Technologies pour l'Agriculture et l'Environnement). Il porte le statut d'EPST (Établissement à caractère Scientifique et Technologique) et est sous la double tutelle du Ministère de la recherche et du Ministère de l'agriculture.

L'organisme a été créé en 1946 à la suite de la Seconde Guerre Mondiale avec la problématique suivante: améliorer les méthodes d'agriculture et d'élevage par la technologie et la science afin de subvenir aux besoins alimentaires du pays. Les objectifs vont évoluer au fur et à mesure des décennies en accord avec les problématiques de chaque époque. En effet avec les années, les études menées par INRAE se sont dirigées vers des projets intégrant un certain code d'éthique, portant sur des questions qualitatives et ajoutant une dimension environnementale (autonomie, durabilité, développement local...).

INRAE compte aujourd'hui près de 8400 agents titulaires et plus de 3000 agents contractuels, répartis sur 18 centres de recherches situés sur 14 départements sur le territoire français. L'organisme au complet est composé de 268 unités de recherche qui sont des unités de recherche, d'expérimentations et de services. Son budget avoisine les 1 milliard d'euros pour des recettes d'environ 30 millions d'euros en 2018. Enfin, l'Institut se place comme l'un des plus prolifiques au monde dans sa catégorie avec plus de 5700 publications de rang A jusqu'en 2018.

Les problématiques étudiées par INRAE sont aujourd'hui diverses et variées, elles peuvent porter sur de nombreuses thématiques comme le changement climatique, l'eau, l'alimentation, la santé, la biodiversité ou encore les biotechnologies en intégrant des dimensions environnementales, sociales et économiques: "selon son décret fondateur (n° 2019-1046 du 10 octobre 2019), INRAE est en charge de diverses missions: réaliser, organiser et coordonner, à son initiative ou à la demande de l'Etat, tous travaux de recherche scientifique et technologique dans les domaines de l'agriculture, de l'alimentation, de la forêt, de l'environnement, de l'eau, de la biodiversité, de la bioéconomie, de l'économie circulaire, de la gestion durable des territoires et des risques dans les champs de compétence précités." (INRAE, rubrique "nous connaître").

Le pôle Grenoblois est composé de trois unités:

-SDAR (Services Déconcentré d'Appui à la Recherche)

-ETNA (Érosion Torrentielle Neige et Avalanches)
-LESSEM (Laboratoire EcoSystèmes et Sociétés En Montagne)
Mon stage s'est déroulé dans l'unité LESSEM, celle-ci étant consacrée à l'étude des sociétés et écosystèmes de montagne.

A noter que l'INRAE est également impliqué dans de nombreux partenariats, on peut prendre pour exemple le laboratoire OSUG (Observatoire des Sciences de l'Univers de Grenoble) situé sur le campus à Grenoble en partenariat avec l'UGA ou encore le laboratoire PAVAL (Pathogenèse et vaccination lentivirales) qui compte plusieurs partenariats avec différentes universités dont l'université de Louisiane Lafayette.

Enfin l'INRAE s'est engagé comme partenaire dans le projet ALPTREES pour apporter sa pierre à l'édifice dans la caractérisation des risques et avantages potentiels des arbres dits non-natifs et dispose d'un coordinateur de projet.

ANNEXE n°2 : CODE du CNN

```
from tensorflow import keras
import tensorflow as tf
from tensorflow.keras import layers
from skimage.util import random_noise
from keras.layers import GaussianNoise

from PIL import Image
import os
import numpy as np

from random import seed
from random import randint

#path_to_files = "G:" + os.path.sep + "ALPTREES_Nathan" + os.path.sep +
"ImagettesFalcon2021"
path_to_files = "Imagettes_3bandes_Falcon"

# liste des images qui vont servir de l'entraînement
vectorized_images = []
classif_images = []

classe = 0
x_train = np.full((2000,51,51,4), 0)

# on tire 400 images aléatoirement
seed(1)
liste_value=[]
```

```

for _ in range(400):
    value = randint(0, x_train.shape[0])
    liste_value.append(value)

# liste des images qui vont servir de test
vectorized_images_test = []
classif_images_test = []

# pour chaque répertoire espèces
# on parcourt chaque image, la met sous forme de numpy array
# on enregistre sa classe dans une liste dédiée
# on met l'image dans la liste

for repertoire in enumerate(os.listdir(path_to_files)):
    path_to_sfiles = path_to_files + os.path.sep + str(repertoire[1])
    nb = 0
    for file in enumerate(os.listdir(path_to_sfiles)):
        if file[1].endswith(".tif"):
            f = path_to_sfiles + os.path.sep + str(file[1])
            image = Image.open(f)
            image_array = np.array(image)
            image_array = image_array.astype("float32") / 255
            #pas de gain réel avec ce flip
            #flip_3 = np.fliplr(flip_2)

            # si l'image a le numéro d'une image qui sert à l'entraînement
            if nb not in liste_value:
                x_train[nb] = image_array
                rot_90 = tf.image.rot90(image_array, k=1)
                rot_180 = tf.image.rot90(image_array, k=2)
                rot_270 = tf.image.rot90(image_array, k=3)
                flip_1 = np.fliplr(image_array)
                flip_2 = np.flipud(image_array)
                # image originelle
                vectorized_images.append(image_array)

                # image originelle avec flip vertical
                vectorized_images.append(flip_1)

                # image originelle avec flip horizontal
                vectorized_images.append(flip_2)

                # image avec rotation de 90 degrés
                vectorized_images.append(rot_90)

                # image avec rotation de 90 degrés avec flip vertical
                vectorized_images.append(np.fliplr(rot_90))

```

```

# image avec rotation de 90 degrés avec flip horizontal
vectorized_images.append(np.flipud(rot_90))

# image avec rotation de 180 degrés
vectorized_images.append(rot_180)

# image avec rotation de 180 degrés avec flip vertical
vectorized_images.append(np.fliplr(rot_180))

# image avec rotation de 180 degrés avec flip horizontal
vectorized_images.append(np.flipud(rot_180))

# image avec rotation de 270 degrés
vectorized_images.append(rot_270)

# image avec rotation de 180 degrés avec flip vertical
vectorized_images.append(np.fliplr(rot_270))

# image avec rotation de 180 degrés avec flip horizontal
vectorized_images.append(np.flipud(rot_270))

# pas de gain
# gaussian = np.random.normal(0, 0.5,
(image_array.shape[0],image_array.shape[1],image_array.shape[2]))
#noisy_image = image_array + gaussian
#vectorized_images.append(noisy_image)

shape = [51, 51, 4]

classif_images.append(classe)
classif_images.append(classe)
classif_images.append(classe)
classif_images.append(classe)
classif_images.append(classe)
classif_images.append(classe)
classif_images.append(classe)
classif_images.append(classe)
classif_images.append(classe)
classif_images.append(classe)
classif_images.append(classe)
classif_images.append(classe)

# si l image est dans le jeu de test on l'ajoute à la liste des images de tests
else:
    vectorized_images_test.append(image_array)

```

```

        classif_images_test.append(classe)
        #vectorized_images_test.append(x_train[int(i-1)])
        #classif_images_test.append(y_train[int(i-1)])
    nb +=1

    classe += 1

x_train = np.array(vectorized_images)

# save as DataX or any other name. But the same element name is to be used while
loading it back.
np.savez("./y_train.npz",DataY=classif_images)
np.savez("./x_test.npz",DataX=vectorized_images_test)
np.savez("./y_test.npz",DataY=classif_images_test)

path = "./y_train.npz"
with np.load(path,allow_pickle=True) as data:
    y_train = data['DataY']

path = "./x_train.npz"
with np.load(path,allow_pickle=True) as data:
    x_train = data['DataX']

path = "./x_test.npz"
with np.load(path,allow_pickle=True) as data:
    #load DataX as train_data
    x_test = data['DataX']

```

Antenne GPS Trimble

Mobile CrossCall

Application Qfield



Contacts

- Vincent Breton (LESSEM)
- Frédéric Bray
- Hugues François

Installation du matériel

- Smartphone Crosscall Trekker X3

Code sim : 0000

Déverrouillage : signe Z comme zorglub

Ce mobile est livré avec un système de charge fiolaire + un système de charge par induction :

Attention au sens du mobile sur le chargeur

- Compte Google

Un compte google est indispensable pour accéder à l'app Store, nous en avons créé un :

- Irstea38@gmail.com
- PW : irstea!38

- Configuration de l'antenne R2 sur le Smartphone :

A. Paramétrer les communications

Dans Paramètres/réseau mobile, nom des points d'accès → + Ajouter

- Nom : TERIA
- APN : orange.m2m.spec
- Nom d'utilisateur: orange
- Mot de passe : orange

B. Associer l'antenne en Bluetooth

L'antenne Trimble s'associe très facilement dans les paramètres bluetooth : Il suffit d'aller dans Bluetooth, et de taper sur le R2 pour le connecter

C. Application Trimble GNSS Statuts

Cette application permet la prise en charge des coordonnées GPS ainsi que des corrections RTK pour lesquelles nous avons pris un abonnement. L'installer depuis le [Playstore](#).

- Commencer par connecter l'antenne R2 dans le menu sous accueil.
- Paramétrage des corrections en temps réel
 - Menu Correction
 - Sélectionner Internet
 - Internet Source : NTRIP
 - URL : 78.24.131.136
 - Port : 2101
 - NTRIP Source : PRS 3.0
 - Username : IRSTEA1
 - Password : 02031803

Une fois ces paramètres testés, se rendre dans Accueil et aller dehors pour vérifier que le GPS envoie des coordonnées d'une précision centimétrique.

Le temps d'attente est d'environ 2 minutes.

!! La doc originale comporte une ERREUR de port 21901 au lieu de 2101 !!

- Récupération des données GPS depuis l'antenne vers d'autres applis

Il faut activer la réception des données GPS par l'antenne au lieu de celles du téléphone.

Pour cela, aller dans *paramètres / à propos du système*

Taper 4 fois sur *version du noyau* (version de build)

Cela active le *Mode Développeur* !

Aller dans *Paramètres / Options pour les développeurs / Partie débogage*

Puis *Sélectionner l'application de position fictive*

Et Sélectionner l'application qui passe les données GPS : *Trimble GNSS Status*

Voilà, normalement maintenant, les données peuvent être reçues par Qfield !

Application Qfield

(version 0.10.9 Kesch installée)

Doc : <http://www.qfield.org/docs/fr/index.html>

1. Intro

Qfield a besoin d'un projet Qgs issu de Qgis.

Mes premiers tests ont été très simple :

Avoir une couche de vecteur (bd-topo, bâtiments)

Pour que ça marche, il faut créer le projet Qgis, et avoir un répertoire data au même niveau dans lequel on va mettre toutes les données dont on a besoin.

NB: Qfield supporte les données Shapefile/Tif, mais pas jp2000

2. Préparation du projet Qgis avec l'extension QfieldSync

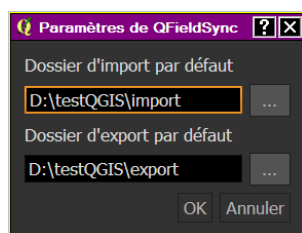
QField Sync permet de synchroniser un projet facilement

Pour préparer une mission terrain, Créer

- Répertoire Export : Contendra les données à basculer vers Qfield
- Répertoire Import : Contendra les données récupérées au retour de terrain par qfield

Paramétrer l'extension Qfield pour prendre en compte ces dossiers :

Extension/ Qfield Sync/Préférences



!/ Préparer le projet qgis normalement, en évitant la surcharge de couches !/

Préparer les couches à alimenter sur le terrain. Utiliser les paramètres de champ pour avoir par exemple un x automatique, ou une date automatique.

Quand tout est prêt, aller dans Extension/QField Sync/Paquet pour Qfield

Cliquez sur créer : Des fichiers vont être créés dans le répertoire Export

Copier ces fichiers, et les mettre dans l'appareil android en le connectant (USB).

Mettre les fichiers directement dans

Ordinateur\TREKKER-X3\Mémoire de stockage interne\Android\data\ch.opengis.qfield\files

Attention, nous avons essayé de créer un répertoire pour éviter d'en mettre partout ☐ celui-ci n'est pas accessible en écriture...

Débrancher le câble, ouvrir le projet Qfield transféré. Normalement, toutes les données qui étaient affichées apparaissent.

3. Sur le terrain : Saisie de données

Pour créer de nouveaux points, aller dans le menu/Mode/Dessin

La liste des couches du projet apparaît. Sélectionner la couche à alimenter.

Sur la carte, une cible apparaît au centre de la carte.

Pour centrer, cliquer sur le bouton bleu GPS

→ un crayon apparaît en bas à droite.

Utilisation du GPS EMLID RS2

(Différentiel Base Rover)

Application Reach View 3



Contacts

- Laurent Borgniet (LESSEM)
- Nathan Herrig (stagiaire donc provisoire)

Installation du matériel

- Téléchargez l'application Reach View 3 sur Google Play ou Apple Store
- Accédez aux *paramètres Wi-Fi* sur votre appareil
- Connectez-vous au point d'accès Reach. Il apparaît comme la portée : XX : XX
- Entrer le mot de passe "emlidreach"
- Lancer Reach View 3
- Choisissez Portée dans la liste (Reach Base dans l'onglet [receivers](#))

Caractéristiques

-Un matériel facile d'utilisation et optimisé pour Android - GPS différentiel RTK

→ Récolte des points GPS précise et efficace

-Des données optimisées en sortie et déjà mises en forme pour utilisation sur un logiciel tiers (SIG)

-Système de coordonnées modifiable WGS84 ou projeté RGF93

Grandes étapes

I. Installation du matériel (terrain)

1. Installer la base ([Reach Base](#))
2. Allumer la base et attendre la connexion au réseau satellite et Wi-fi
3. Allumer le téléphone associé et le connecter en Wi-fi à la base
4. Entrer sur l'application Reach View et le connecter à nouveau à la base (Onglet [Receivers](#))
5. Entrer manuellement les coordonnées XYZ de la base (Onglet [Receivers](#))

II. Utilisation du matériel (terrain)

1. Allumer le second GPS [Reach Rover](#)
2. Quitter l'application Reach View 3
3. Paramètres > Wi-fi > se connecter à Reach Rover
4. Entrer sur l'application Reach View 3 et le connecter à nouveau au GPS Reach Rover (Onglet [Receivers](#))

5. Créer nouveau projet et entrer paramètres dans l'onglet **Survey** (Nom, description, etc...)
6. Ouvrir le projet (**Survey**)
7. Déplacez-vous avec le Reach Rover jusqu'au point souhaité
8. Ajouter points à l'aide du bouton "+" → l'identifiant point s'inscrit automatiquement, vous n'avez qu'à remplir l'espace "commentaire"
9. Cliquez sur "**Measures**" et attendez une dizaine de secondes
10. Pour quitter le projet cliquez sur l'onglet "<- Projets"

TOUTES LES MODIFICATIONS S'ENREGISTRENT AUTOMATIQUEMENT ET PEUVENT ETRE EDITEES.

III. Extraction des données

1. Allumer la base RTK (**Reach Base**)
2. Allumer le téléphone associé
3. Le connecter au Wi-fi du Reach Base
4. Aller sur l'onglet **Survey**
5. Entrer dans les paramètres du projet (3 petits points en haut à droite de l'onglet projet)
6. Cliquer sur "**Export**"
7. Plusieurs choix possibles pour exporter les données (Bluetooth, Drive, Mail, ...) → Mails à favoriser (plus simple)
8. Envoyer mail avec le document exporté en .csv
9. Connecter le téléphone à un réseau ouvert pour envoi du mail (**les mails restent dans la boîte d'envoi quand le téléphone est connecté à Reach Base**)
10. Réceptionner le mail (tableau .csv avec l'ensemble des coordonnées + id et descriptif, tout est automatisé).