

MSc HES-SO en Business Administration

Orientation :
Management des Systèmes d'information

Détection automatique d'objets agricoles avec du Machine Learning

Réalisé par

Vincent Schülé

Sous la direction de
Prof. Dr Dominique Genoud

Sierre, 24.08.2018

Table des matières

Table des Matières	ii
Remerciements	v
Résumé	vi
1 Introduction	1
1.1 Objectif de l'étude	2
1.2 Limitations	2
2 État de l'art	3
2.1 Réseau de neurones	4
2.1.1 CNN	4
2.1.2 R-CNN	5
2.1.3 Faster R-CNN	6
2.1.4 Segmentation d'instance	7
2.1.5 Mask R-CNN	8
2.1.5.1 ResNet	8
2.1.6 DOTA	9
2.2 Revue de littérature	9
2.2.1 Analyse de fréquence	10
2.2.2 Seuil + Cluster	10
2.2.3 Analyse basée objet	11
2.2.4 Analyse basée pixel	12
2.2.5 Prise de données	13
2.3 Discussion état de l'art	14
3 Méthodologie	15
3.1 Connaissance du métier	16
3.2 Connaissance des données	16
3.3 Préparation des données	16
3.4 Modélisation des données	17
3.5 Évaluation	17
3.6 Déploiement	17
4 Analyse	18
4.1 Connaissance du métier	18
4.1.1 But du business	18

4.1.2	Buts du Data Mining	19
4.1.3	Plan de projet	19
4.2	Connaissance des données	19
4.2.1	Collecte des données	20
4.2.2	Description des données	21
4.2.3	Exploration des données	22
4.2.3.1	Chamoson-1-4.tif	23
4.2.4	Qualité des données	23
4.3	Préparation des données	24
4.3.1	Annotations des images	24
4.3.1.1	Filtre d'image floue	25
4.3.2	Caractéristiques des données	25
4.3.3	But général des transformations	25
4.3.3.1	Transformation des images	26
4.3.3.2	Séparation en vignobles	27
4.3.4	Aligner aux bords de l'image	27
4.3.5	Séparation des images	28
4.3.6	Augmentation	29
4.4	Modélisation des données	31
4.4.1	Préparation des tests	31
4.4.2	Modèles utilisés	32
4.5	Évaluation	33
4.6	Déploiement	33
5	Résultats	35
5.1	Résultat des modèles	36
5.1.1	Référence	36
5.1.2	Faster R-CNN	37
5.1.3	Mask R-CNN	38
5.1.3.1	Évaluation métier	39
5.2	Tableau des résultats	41
6	Discussion	42
6.1	Difficultés	42
6.2	Limites et amélioration	43
7	Conclusion	44
	Références	45
	Attestation d'Authenticité	49
	Glossaire	50
	Abréviations	52
	Liste des Figures	53
	Liste des Tables	55

A Chamoson-1-4.tif

56

B Résultat sur Chamoson-1-4.tif

57

Remerciements

J'adresse mes remerciements à toutes les personnes qui m'ont aidé pendant la rédaction de ce rapport.

Je remercie en premier le professeur docteur Dominique Genoud de la HES-SO Valais-Wallis. En tant que Directeur de travail, il m'a aidé à guider cette recherche et à considérer des solutions pour avancer.

J'aimerais aussi remercier William Metz, CEO de Precision Vine, pour ses connaissances du métier intéressantes et sa collaboration dans l'acquisition des données.

Je souhaite remercier mes collègues du DUDE-LAB de l'IIG Sierre pour leurs conseils et contributions essentielles à la création de ce travail.

Finalement, je voudrais remercier ma famille et mes amis de m'avoir soutenu durant ce projet ainsi que pour la relecture de ce rapport.

Résumé

Ce projet cherche à améliorer la viticulture en utilisant des outils de Machine Learning pour détecter les objets agricoles dans un vignoble. La viticulture de précision utilise les technologies de l'information afin de perfectionner la gestion des vignes. Une meilleure gestion permet d'améliorer le rendement du vignoble ainsi que de réduire l'utilisation de produits chimiques de traitement.

L'analyse et le traitement automatique d'un vignoble nécessitent de connaître le géo-référencement de chaque ligne de vignes. Les coordonnées géographiques des vignes sont cependant longues et difficiles à obtenir correctement avec des moyens traditionnels. Le but de cette recherche est de simplifier cette tâche. Pour ce faire, la détection et la segmentation des lignes de vigne à l'intérieur d'un vignoble doivent être améliorées.

Le but de cette recherche est de faciliter l'annotation géographique des lignes de vigne. La majorité des données utilisées ont été collectées pendant le projet avec l'aide d'un drone commercial équipé d'une caméra RGB.

Cette recherche a suivi la méthodologie CRISP-DM afin de concevoir un outil de segmentation et localisation des lignes de vignes à l'intérieur d'un vignoble. Un réseau de neurones Mask R-CNN a été modélisé afin de répondre au but de la recherche. Le processus final établi reste facile à utiliser et améliore la précision par rapport au modèle de référence.

Mots clés : Machine Learning, viticulture de précision, Mask R-CNN, détection d'objet

Chapitre 1

Introduction

Depuis plusieurs années, la classification d'image s'est beaucoup améliorée grâce à de meilleures utilisations du concept de Machine Learning. La création de modèles par apprentissage nous permet de rapidement détecter et classer différents objets sans devoir écrire un algorithme précis.

Ces techniques de Machine Learning sont cependant encore peu documentées dans le domaine de la viticulture de précision. Ce domaine cherche à améliorer la production et la qualité des vignobles en améliorant les technologies utilisées pour leur gestion (AGROFLY, 2016).

TREBOUX, 2017 propose déjà une amélioration de la détection de vignoble en utilisant du Machine Learning. Nous proposons de continuer en cherchant à identifier et extraire les lignes de vigne à l'intérieur d'un vignoble. L'identification des vignes permet de mieux analyser chaque partie du vignoble et ainsi le traiter au mieux (CONESE, RAPI, ROMANI, BONORA & DAMBRA, 2005).

En viticulture de précision, des caméras multispectrales ou des LiDAR sont généralement utilisés pour la récolte de données. Pour ce projet, nous proposons d'utiliser uniquement des images Rouge, Vert et Bleu (RGB) aériennes. Ceci permettrait à notre solution d'utiliser des informations facilement disponibles. Les données de ce projet seront principalement capturées avec un drone DJI Mavic Pro mis à disposition.

1.1 Objectif de l'étude

Les buts de cette étude sont les suivants :

- Établir un modèle de reconnaissance d'objets agricoles à l'intérieur d'images de vignobles.
- Différencier entre un cep de vigne, sa couronne et la terre, herbes, routes ou autres objets alentour.
- Utiliser ce modèle doit être facile et réduire le temps de travail humain dans la détection des objets agricoles.

Pour réaliser ces buts, nous devons tout d'abord établir un état de l'art des méthodes utilisées pour reconnaître des objets agricoles dans la viticulture. À partir de cet état de l'art, nous pourrions déterminer l'ensemble des données nécessaires pour la reconnaissance d'objets viticoles et l'acquérir.

Nous devons ensuite considérer des approches de Machine Learning prometteuses qui pourraient améliorer la situation par rapport aux techniques étudiées auparavant.

Ce projet utilisera la méthodologie Cross Industry Standard Process for Data Mining (CRISP-DM) afin d'assurer une réflexion sur chaque partie importante d'un projet de Data Mining. Cette méthodologie est décrite dans le chapitre 3.

1.2 Limitations

Cette recherche s'inscrit dans le domaine de la viticulture de précision. Cela implique que nous utiliserons principalement des images de vignoble. Les limites suivantes sont aussi implémentées afin de focaliser la recherche et de permettre un résultat utile.

- Le modèle ne considère que des images de vignoble nettoyées et prétraitées. Cela afin de réduire le bruit lors de l'apprentissage du modèle.
- Les vignobles modélisés sont composés de hautes vignes linéaires. Ceci est le type de vigne utilisé principalement dans la région de la vallée du Rhône.

Chapitre 2

État de l'art

La détection des objets et leur localisation dans une image sont des éléments importants de la recherche courante en vision ordinateur. Cette discipline cherche à détecter dans une image les différents objets, leur classe ainsi que leur position dans l'image.

Une architecture souvent utilisée ces dernières années est Régions avec Réseau Neuronal Convolutif (R-CNN). Ce type d'architecture de Machine Learning s'est montré comme très performant dans plusieurs domaines pour résoudre la détection d'objet (GIRSHICK, DONAHUE, DARRELL & MALIK, 2013).

Cependant, peu de recherches ont été effectuées pour appliquer cette technique à la viticulture de précision. Dans ce domaine, les algorithmes de détection d'image documentés sont plus souvent déterministes. TREBOUX, 2017 a par contre proposé plusieurs modèles qui améliorent la détection des vignobles.

Dans cette revue de la littérature, nous ferons un tour des algorithmes et principes utilisés pour détecter les lignes de ceps à l'intérieur d'un vignoble ainsi qu'une revue de différents modèles de réseaux de neurones qui pourraient améliorer ce domaine. Nous comparerons aussi les plateformes de télédétection utilisées dans l'imagerie aérienne pour la viticulture de précision.

2.1 Réseau de neurones

Dans cette section, nous parlerons de plusieurs concepts reliés aux réseaux de neurones artificiels importants à la compréhension de ce rapport.

”Les réseaux de neurones artificiels sont des réseaux fortement connectés de processeurs élémentaires fonctionnant en parallèle. Chaque processeur élémentaire calcule une sortie unique sur la base des informations qu’il reçoit. Toute structure hiérarchique de réseaux est évidemment un réseau.” (TOUZET, 1992)

Cette définition explique que chaque neurone est une fonction basique. Un neurone peut par exemple recevoir un nombre, le multiplier par deux puis l’envoyer aux neurones suivants.

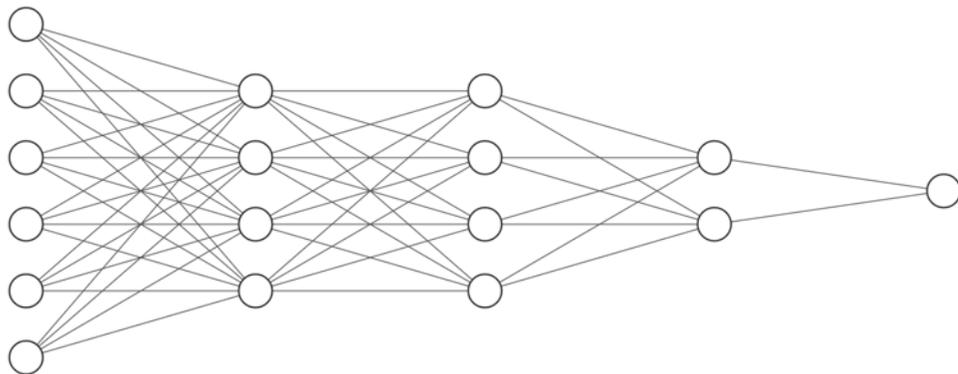


FIGURE 2.1: Réseau de neurones en cinq couches.

Les réseaux de neurones sont souvent organisés en plusieurs couches comme en figure 2.1. Ceci permet de mieux organiser le parcours de l’information et facilite l’apprentissage du modèle (TOUZET, 1992).

2.1.1 CNN

Le Réseau Neuronal Convolutif (CNN) est comparable à un réseau de neurones standard. Il est aussi composé de neurones avec différentes opérations simples (KARPATHY, 2017). Cependant, les CNN présument que les données entrantes sont en forme de grille ou matrice. Ce type de données se présente dans plusieurs domaines : longueur d’onde audio (LEE, KIM, PARK & NAM, 2017), le traitement automatique du langage naturel (LOPEZ

& KALITA, 2017) ou plus communément des images (GOODFELLOW, BENGIO & COURVILLE, 2016; KRIZHEVSKY, SUTSKEVER & HINTON, 2012; SIMONYAN & ZISSERMAN, 2014).

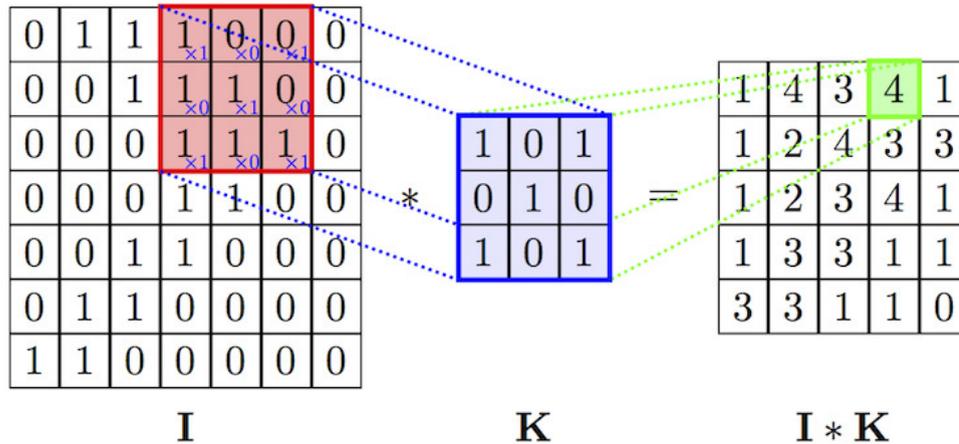


FIGURE 2.2: Couche de convolution (VELIČKOVIĆ, 2017).

Assumer une forme de grille pour les données permet au CNN de créer une couche qui les transforme à partir d'une fenêtre coulissante. Ces couches convolutives permettent de compacter l'information d'une région en une seule donnée (GOODFELLOW et al., 2016).

2.1.2 R-CNN

Le R-CNN est une architecture de détection d'objet. Le R-CNN commence par extraire des régions intéressantes de l'image, puis il utilise ces régions comme données d'entrée pour un CNN. Cette séparation en régions permet de détecter plusieurs objets de plusieurs classes différentes dans une même image. Cette solution proposée par GIRSHICK et al., 2013 a permis d'améliorer la précision des modèles de détection.

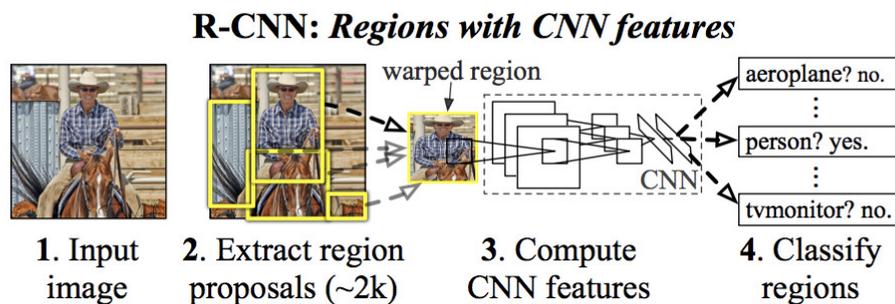


FIGURE 2.3: Architecture R-CNN (GIRSHICK, DONAHUE, DARRELL & MALIK, 2013).

Dans le R-CNN présenté en figure 2.3, les régions sont extraites grâce à une recherche sélective proposée par UIJLINGS, van de SANDE, GEVERS et SMEULDERS, 2013. Celle-ci utilise la structure de l'image et plusieurs techniques de partitionnement pour récupérer toutes les régions intéressantes possibles.

2.1.3 Faster R-CNN

Le Faster R-CNN est une amélioration du R-CNN dans sa précision et sa rapidité à l'entraînement (REN, HE, GIRSHICK & SUN, 2015).

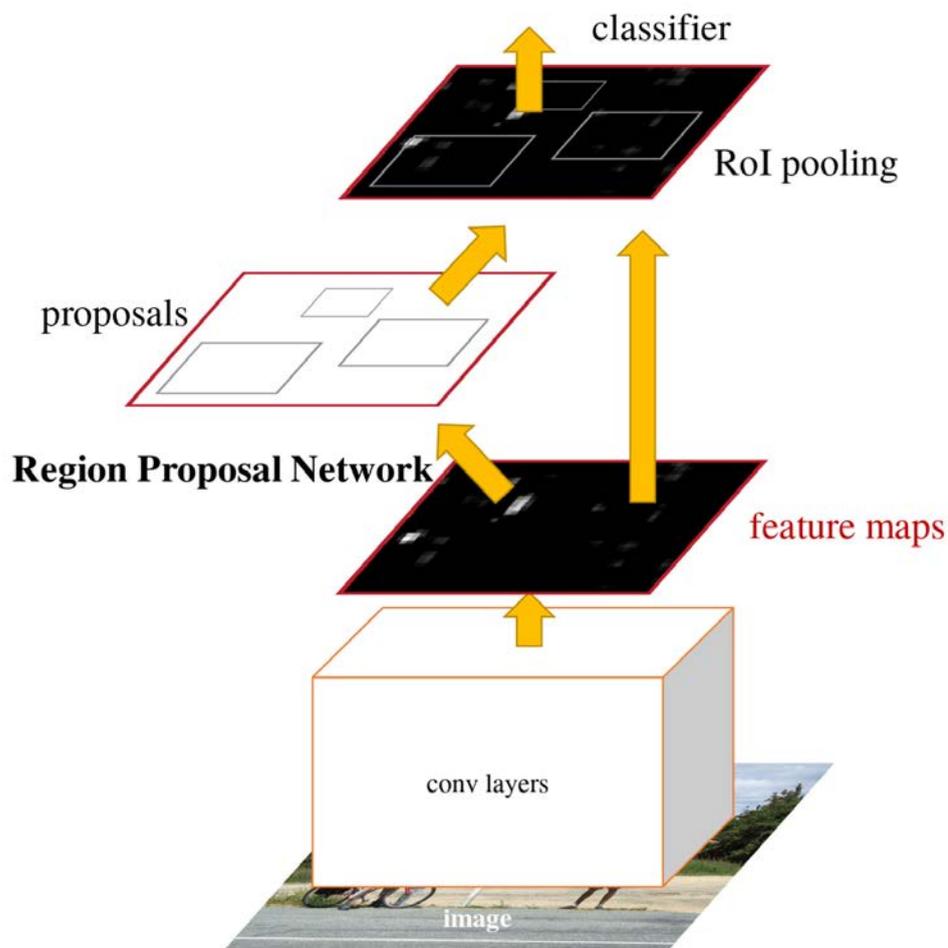


FIGURE 2.4: Architecture Faster R-CNN (REN, HE, GIRSHICK & SUN, 2015).

L'architecture Faster R-CNN (fig. 2.4) utilise la même carte de caractéristiques résultant des couches de convolution pour générer les régions intéressantes et pour ensuite les classifier. Le réseau de propositions de régions utilise des fenêtres coulissantes (fig. 2.5) de tailles et ratio différents pour analyser la carte de caractéristiques en profondeur. Ces

changements améliorent significativement la précision et la rapidité de l'architecture comparée au R-CNN (REN et al., 2015).

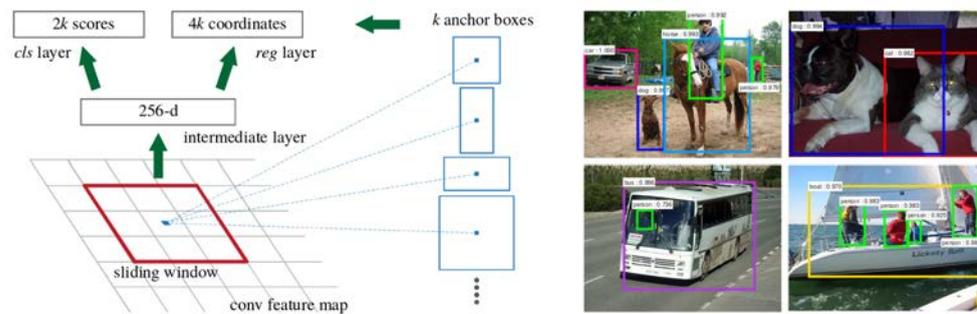


FIGURE 2.5: Faster R-CNN : Fenêtre coulissante et résultat (REN, HE, GIRSHICK & SUN, 2015).

2.1.4 Segmentation d'instance

La segmentation d'instance est un problème de vision ordinateur. La segmentation d'instance cherche à identifier chaque instance, ou exemplaire, d'une classe recherchée à l'intérieur des données présentées. En vision d'ordinateur, la segmentation d'instance cherche à classifier chaque pixel d'une image à une instance d'un objet.

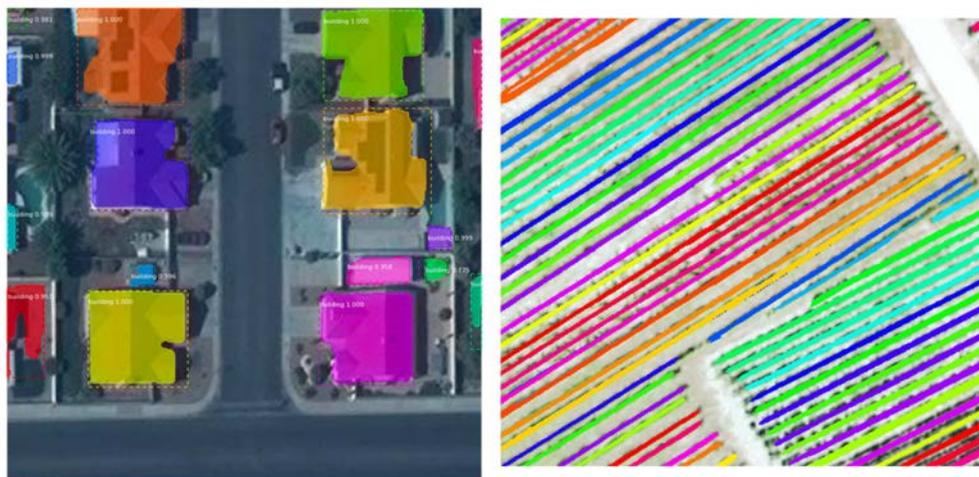


FIGURE 2.6: Exemple de segmentation d'instance. À gauche, de bâtiments (MOHANTY, 2018), à droite de vigne.

En figure 2.6, chaque bâtiment ou ligne de vigne est séparé avec un masque de prédiction. Les couleurs représentent chacune une instance différente. Cela permet de facilement identifier et visualiser chaque instance.

2.1.5 Mask R-CNN

Le Mask R-CNN est l'amélioration suivante après le Faster R-CNN. Le Mask R-CNN change la sortie du modèle final. En effet, l'architecture Faster R-CNN permet de localiser des objets distincts avec une boîte englobante. L'architecture Mask R-CNN permet, quant à elle, de segmenter chaque instance d'un objet avec un masque sémantique. Cette amélioration permet donc au modèle de faire de la segmentation d'instance (section 2.1.4) (HE, GKIOXARI, DOLLÁR & GIRSHICK, 2017).

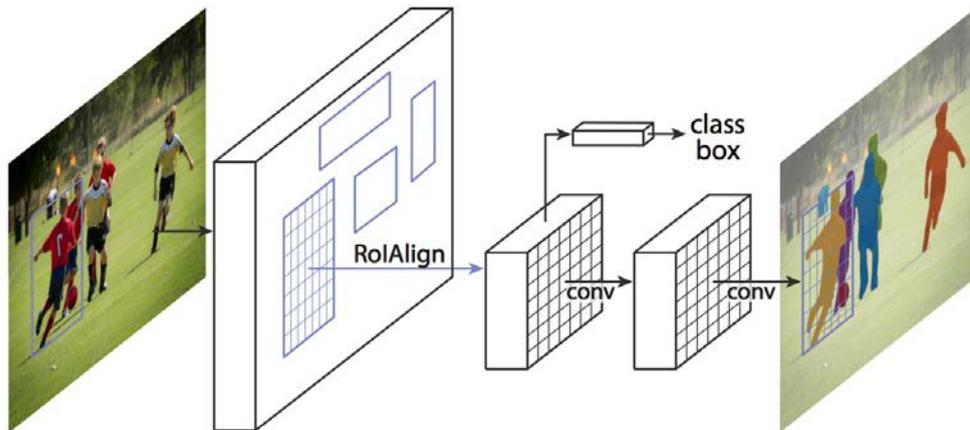


FIGURE 2.7: Architecture Mask R-CNN (HE, GKIOXARI, DOLLÁR & GIRSHICK, 2017).

Comme on peut le voir en figure 2.7, l'architecture Mask R-CNN est semblable à Faster R-CNN mais rajoute des couches en parallèle de la classification. Cette approche est différente d'autres architectures de segmentation d'instance qui classifiaient les résultats par rapport au masque généré (DAI, HE & SUN, 2015; PINHEIRO, COLLOBERT & DOLLÁR, 2015).

Cette approche en parallèle permet à chaque type d'objet de prédire leur masque sans être en concurrence avec les autres classes du modèle (HE et al., 2017).

2.1.5.1 ResNet

Les Residual Network (ResNet) sont des réseaux de neurones qui implémentent l'apprentissage résiduel. Cette innovation permet de faciliter l'apprentissage des couches neuronales profondes. L'apprentissage résiduel découle de l'observation que les couches profondes d'un réseau de neurones avaient des difficultés à converger. L'apprentissage

résiduel permet à l'information de contourner certaines couches et réduit ainsi le problème de dégradation (HE, ZHANG, REN & SUN, 2015).

2.1.6 DOTA

DOTA est un grand ensemble de données d'images aériennes prévu pour la détection d'objet. Ce set de données est intéressant pour son style d'annotation particulier. En effet, les objets sont annotés avec des boîtes englobantes orientées à l'objet en plus de celles alignées avec l'image.

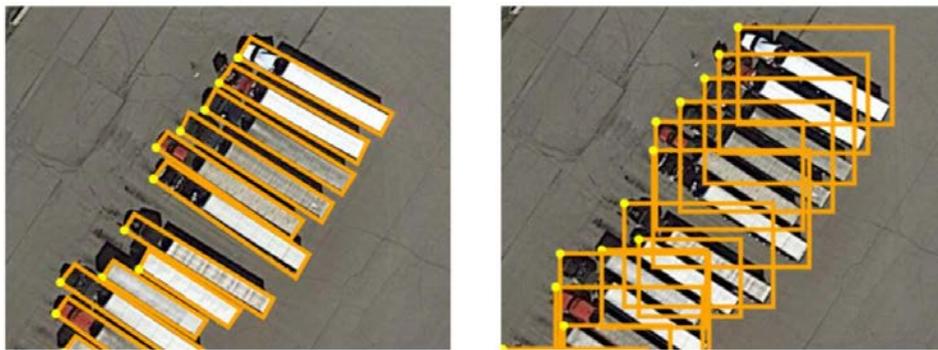


FIGURE 2.8: Styles de boîtes englobantes pour DOTA. A gauche, orientée objet. A droite, orientée image (XIA et al., 2018).

Ce type d'annotation a été choisi pour répondre à la particularité des images aériennes. XIA et al., 2018 considèrent que les objets vus du haut sont souvent dans des orientations différentes et regroupées. On peut voir en figure 2.8 qu'une boîte englobante orientée image génère beaucoup de chevauchements entre les annotations. Ceci est un problème lors de la détection d'objet, car la séparation des différentes instances est plus compliquée.

2.2 Revue de littérature

De multiples recherches en viticulture de précision tentent d'utiliser diverses techniques pour délimiter les lignes de vigne et les séparer des interlignes. Dans cette revue de la littérature, nous allons nous focaliser sur les techniques de détection utilisant des prises de vues aériennes avec des caméras RGB. Celles-ci peuvent cependant provenir de plusieurs sources : avions, drones ou satellites.

2.2.1 Analyse de fréquence

La technique d'analyse de fréquence utilise la transformation de Fourier pour détecter l'orientation des lignes de vigne et l'intervalle interligne dans une zone de vignoble prédéfinie. Cette technique fonctionne grâce aux caractéristiques régulières des vignobles. Les lignes sont en général parallèles, régulièrement espacées et de même largeur (WASSENAAR, ROBBEZ-MASSON, ANDRIEUX & BARET, 2002). Après avoir défini ces deux paramètres, l'identification se fait simplement en supprimant les labels non conformes.

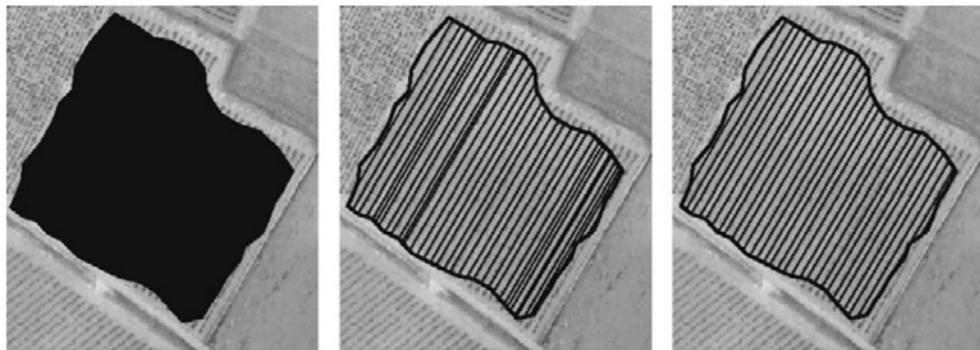


FIGURE 2.9: Illustration de la méthode d'analyse de fréquence. Utilise l'orientation et la largeur des lignes de vigne ainsi que l'espace d'interligne obtenu avec l'analyse de fréquence pour calculer les labels (DELENNE, DURRIEU, RABATEL & DESHAYES, 2010).

L'utilisation de l'algorithme Fast Fourier Transform permet d'avoir un bon résultat dans les lignes régulières. Cependant, cet algorithme ne peut pas fonctionner dans les vignes irrégulières, par exemple, en pente ou avec des lignes non parallèles. Il ne permet pas non plus de détecter un trou dans les lignes de vigne (DELENNE, DURRIEU, RABATEL & DESHAYES, 2010).

2.2.2 Seuil + Cluster

Une des premières techniques utilisées pour la détection des vignobles était un simple seuil sur l'histogramme de l'image. Chaque valeur au-dessus de ce seuil étant considérée comme vigne, chaque valeur en dessous étant considérée comme non-vigne (HALL, LOUIS & LAMB, 2003). Cette technique simple est cependant utile uniquement si l'image ne comporte pas de végétation autre que la vigne. En effet, il n'y a pas de différence utilisable de valeur entre la vigne et d'autres végétaux en imagerie RGB (COMBA, GAY, PRIMICERIO & AIMONINO, 2015). Une autre technique utilisée est de découvrir dans

l'image les structures linéaires des vignobles. Cependant, cela détecte aussi des bâtiments linéaires ou d'autres structures similaires à l'intérieur de l'image (DELENNE et al., 2010).

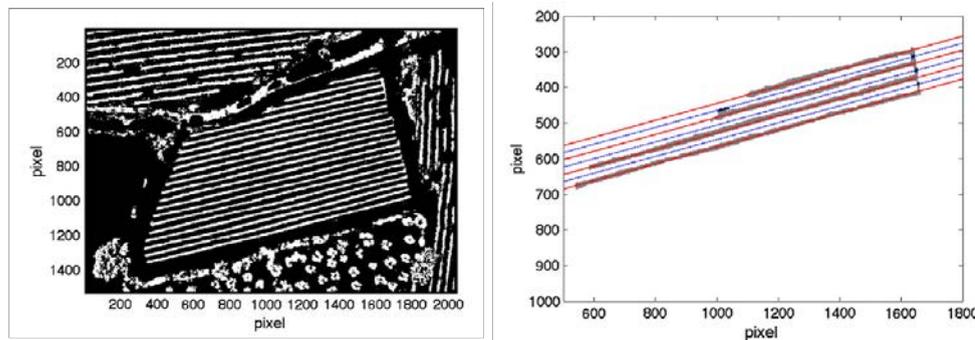


FIGURE 2.10: Illustration de la méthode de seuil et clustering. À gauche, un seuil est utilisé pour séparer la végétation, à droite cette séparation est utilisée pour détecter les différentes lignes de vigne dans l'image (COMBA, GAY, PRIMICERIO & AIMONINO, 2015).

Pour résoudre ces problèmes, COMBA et al., 2015 ont émis l'hypothèse que dans les images d'un vignoble, seules les vignes ont les caractéristiques de seuil de végétation et la structure linéaire d'un vignoble. Le résultat des deux techniques combinées permet donc de mieux filtrer la détection des vignes dans l'image.

2.2.3 Analyse basée objet

L'analyse basée objet est une technique maintenant souvent utilisée en Système d'Information Géographique (GIS). Cette technique commence par séparer une image en différents segments avec des caractéristiques similaires puis classifie ces segments selon leurs forme, taille, texture, intensité ou autres traits (BLASCHKE et al., 2014).

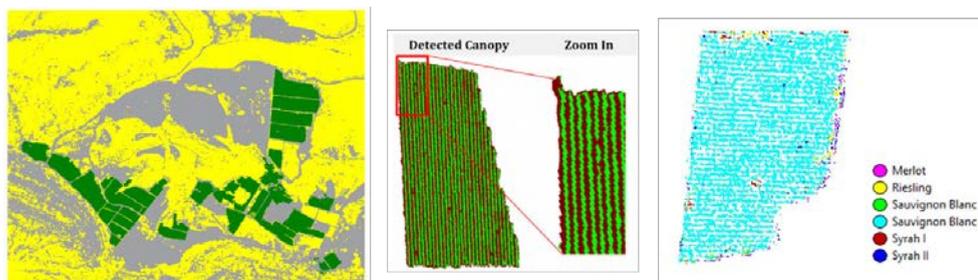


FIGURE 2.11: Illustration de la méthode d'analyse orientée objet. À gauche, les vignobles sont détectés à petite échelle, au milieu les vignes sont détectées, à droite la variété du cep est prédite selon les informations connues.(KARAKIZI, OIKONOMOU & KARANTZALOS, 2016).

Cette technique permet aussi de travailler sur plusieurs niveaux d'échelle géographique pour segmenter et classifier différents types d'objets. En effet, là où une petite échelle permet de détecter des champs, forêts ou villes, une grande échelle peut permettre de séparer chaque arbre et sa variété (SCHWARZER, ALLENBACH & JAQUET, 2009; YU et al., 2006).

KARAKIZI, OIKONOMOU et KARANTZALOS, 2016 proposent une solution de détection orientée objet pour la viticulture de précision décrite en figure 2.11. Leur solution commence par détecter les vignobles puis extrait la couronne des vignes et cette information est ensuite utilisée pour différencier la variété du vignoble.

Pour extraire les couronnes des lignes de vigne, KARAKIZI et al., 2016 ont utilisé des caractéristiques spectrales (moyenne, ratio et Indice de Végétation (NDVI)) et régionales (différences, minimums, ...) d'images multispectrales. Leur solution a permis de détecter les lignes de vigne avec une précision minimum de 96 % (KARAKIZI et al., 2016).

2.2.4 Analyse basée pixel

La classification basée pixel est une autre technique de GIS. Celle-ci cherche à catégoriser chaque pixel selon des caractéristiques qui lui sont propres. Par exemple, chaque pixel a des informations spectrales et des intensités de lumière différentes. Des indices spectraux peuvent aussi être calculés comme le NDVI (DURO, FRANKLIN & DUBÉ, 2012).

POBLETE-ECHEVERRÍA, OLMEDO, INGRAM et BARDEEN, 2017 ont effectué une comparaison de plusieurs méthodes d'analyse basée pixel pour la viticulture de précision. Leur modèle tentait de classifier chaque pixel d'un vignoble dans les catégories : plante, terre, ombre. Pour se faire, ils ont utilisé les valeurs RGB, le ratio de vert (2.1) et l'indice $2G_RBi$ (2.2).

$$G\% = \frac{G}{R + G + B} \quad (2.1)$$

$$2G_RBi = 2 * G - (R + B) \quad (2.2)$$

D'après POBLETE-ECHEVERRÍA et al., 2017, la technique de classification la plus précise avec ces données est le seuil global d'Otsu sur l'indice 2G-RBi (2.2) avec une précision moyenne de 98 %. Cet indice multiplie l'importance de la couleur verte dans l'image. Ceci le rend très utile pour détecter la végétation verte, mais est moins performant pour différencier d'autres objets.

2.2.5 Prise de données

Pour ce projet, il était aussi important d'établir l'état de l'art sur l'acquisition d'images orthogonales avec un drone ou Unmanned Aircraft Vehicle (UAV). En effet, l'acquisition d'une partie du set de données à modéliser fait partie intégrante du projet.

Différentes plateformes de télédétection ont été comparées pour la viticulture de précision par MATESE et al., 2015. Cette comparaison propose que les UAV, avions et satellites aient chacun des avantages et inconvénients différents.

Table 5. Comparative platform characteristics for different remote sensing platforms. (++ optimal, + good, o average, - poor).

		UAV	Aircraft	Satellite
Mission	Range	-	+	++
	Flexibility	++	+	-
	Endurance	-	++	++
	Cloud cover dependency	++	+	-
	Reliability	o	+	++
Processing	Payload	o	+	++
	Resolution	++	+	o
	Precision	++	+	o
	Mosaicking and geocoding effort	-	o	++
	Processing time	o	+	+

FIGURE 2.12: Comparaison de plateformes de télédétection pour la viticulture de précision (MATESE et al., 2015).

Les données prises avec un UAV sont généralement plus précises, car de très haute résolution au sol. De plus un angle de vue diagonal permet de précisément calculer l'altitude d'un point. Cependant, les UAV requièrent plus de travail pour une même zone de traitement. Les avions, ou autres aéronefs, sont bons dans tous les domaines mais n'excellent en rien. Enfin, les satellites apportent de très grandes distances de détection, mais perdent en flexibilité et demandent beaucoup de travail pour le traitement des images (MATESE et al., 2015).

2.3 Discussion état de l'art

Cette revue de la littérature nous a permis de comprendre les différentes techniques utilisées pour segmenter les lignes de vigne dans un vignoble. Ces différentes techniques proposent des avantages et inconvénients différents. Par exemple, une complexité de calcul ou de travail humain accrue pour une meilleure précision.

Pour cette recherche, nous utiliserons un seuil d'Otsu sur l'indice 2G_RB (2.2) comme décrit dans la section 2.2.4 pour créer les labels de vérité. En effet, cette technique requiert uniquement de séparer préalablement les vignobles du reste de l'image. Cette solution propose aussi un très bon niveau de précision. Une correction humaine sera tout de même effectuée sur les labels générés afin de nettoyer les résultats faux positifs générés par l'algorithme.

Nous avons pu vérifier que plusieurs techniques de prise de données font sens pour cette recherche. Nous utiliserons principalement des données provenant d'UAV. Cependant, nous utiliserons aussi des données d'aéronefs provenant de SwissTopo.

Cette recherche expérimentera avec les architectures Faster R-CNN et Mask R-CNN afin de modéliser les lignes de vigne. En effet, ces deux architectures se sont respectivement prouvées intéressantes dans la détection d'objet et la segmentation d'instance. Cependant, leur utilisation n'a pas encore été documentée dans le domaine de la viticulture de précision.

Chapitre 3

Méthodologie

Le but de ce projet est d'établir un modèle de détection d'objets agricoles dans un vignoble. Pour structurer notre rapport, nous avons décidé d'utiliser la méthodologie CRISP-DM. CRISP-DM sépare les projets de Data Mining en 6 étapes (CHAPMAN et al., 2000) :

1. Connaissance du métier
2. Connaissance des données
3. Préparation des données
4. Modélisation des données
5. Évaluation
6. Déploiement

Comme vu en figure 3.1, les différentes étapes sont liées entre elles de manière itérative. En effet, les connaissances supplémentaires que nous rapporte le travail effectué nous permettent d'améliorer les étapes précédentes. Ainsi, un meilleur modèle de prédiction permet d'améliorer les connaissances du métier et recommence le processus (CHAPMAN et al., 2000).

Ci-dessous, nous proposons une explication de chaque étape et des tâches à y effectuer.



FIGURE 3.1: Différentes phases de CRISP-DM (CHAPMAN et al., 2000).

3.1 Connaissance du métier

La connaissance du métier se focalise sur la prise de connaissance du métier et du problème à résoudre. Durant cette phase, nous définissons les buts métiers que nous voulons accomplir avec le projet. Ces buts sont après retranscrits en problème de Data Mining et en un plan capable de le résoudre (CHAPMAN et al., 2000).

3.2 Connaissance des données

L'étape connaissance des données cherche à familiariser le chercheur avec les données disponibles. Elle commence par une tâche d'acquisition des données. Ensuite, le chercheur décrit les données et identifie leurs premières caractéristiques. Cela permet de découvrir des problèmes de qualité ou de détecter des informations intéressantes présentes dans les données (CHAPMAN et al., 2000).

3.3 Préparation des données

La phase de préparation des données inclut toutes les tâches nécessaires pour transformer le set de données brut en set de données utilisable par les outils de modélisation. Cette étape est souvent très itérative avec l'étape de modélisation suivante. Cette itération

est due au fait que les problèmes deviennent plus apparents lors de la modélisation des données (CHAPMAN et al., 2000).

3.4 Modélisation des données

Dans la phase de modélisation des données, diverses techniques de modélisation sont appliquées aux données traitées pour créer un modèle. Le but de cette phase est de créer un modèle qui répond au problème de Data Mining défini auparavant efficacement. Pour ce faire, il est important de tester progressivement la capacité des modèles à atteindre leur but correctement (CHAPMAN et al., 2000).

3.5 Évaluation

Cette phase évalue le côté métier des modèles. Les modèles arrivant à cette phase sont déjà jugés de bonne qualité par la phase de modélisation. Cependant, il est nécessaire d'évaluer leur performance à résoudre les objectifs métiers ainsi que de vérifier qu'ils ne comportent pas de problèmes métier conséquents. À la fin de cette phase, nous devons savoir si le modèle peut être déployé en l'état (CHAPMAN et al., 2000).

3.6 Déploiement

Finalement, les modèles qui passent l'étape d'évaluation peuvent être déployés. La phase de déploiement permet de donner accès au modèle et à ses résultats. Cela peut être sous plusieurs formes, comme un rapport ou un service de détection (CHAPMAN et al., 2000).

Chapitre 4

Analyse

Pour cette analyse, nous allons suivre la structure présentée dans le chapitre 3. Dans l'état de l'art, nous avons défini que les lignes de vigne pouvaient être segmentées en appliquant un seuil d'Otsu sur un indice de couleur (section 2.2.4). Nous utiliserons ce résultat pour créer notre modèle de référence et nos labels de données.

Ce projet expérimente l'utilisation de deux CNN. Un Faster R-CNN (section 2.1.3) avec annotation orientée objet (section 2.1.6) et un Mask R-CNN (section 2.1.5).

4.1 Connaissance du métier

Comme vu dans le chapitre 3, la première étape d'un projet CRISP-DM est la compréhension du métier et du problème. CRISP-DM sépare cette section en 4 tâches importantes :

1. Identifier les buts du business.
2. Évaluer la situation.
3. Définir les buts du Data Mining.
4. Produire le plan du projet.

4.1.1 But du business

Ce projet a deux côtés business distincts importants. Le côté recherche et le côté client.

Du côté recherche, le but est d'améliorer les connaissances en matière de détection d'objets dans la viticulture de précision. Pour ce faire, nous devons établir le niveau de la recherche courante et l'améliorer. Ceci a été fourni dans le chapitre 2.

Du côté client, le but est de simplifier la création d'un parcours couvrant l'intégralité des plantes d'une parcelle de vigne. Afin de se rapprocher de ce but, il est nécessaire de pouvoir détecter les lignes de vigne dans une parcelle de manière fiable et efficace.

Le but de ce projet est donc la détection de lignes de vigne dans une image avec le moins d'intervention humaine possible.

4.1.2 Buts du Data Mining

Les buts de ce projet de recherche sont d'expérimenter de nouvelles solutions pour détecter des lignes de vigne et de les comparer avec les résultats de référence. Un modèle moins précis n'est pas un échec du but de la recherche, cela veut simplement dire que le modèle de référence est plus performant.

4.1.3 Plan de projet

Le plan du projet a été défini avec l'aide du superviseur afin d'orienter le projet (table 4.1). Les premiers mois ont été focalisés sur la création de l'état de l'art et la récupération des données. Ensuite, le processus d'itération a commencé et a été dirigé de façon à atteindre un résultat intéressant. Ce plan tient compte du maximum de 450 heures autorisées pour la réalisation de ce travail.

4.2 Connaissance des données

Les données sont toujours la base principale d'un projet de Data Mining. Il est donc important de bien comprendre les données disponibles et leurs caractéristiques. Pour ce projet, les données sont composées de prises de vues aériennes de vignobles linéaires. Dans cette section, nous allons commencer par la collecte des données. Nous décrirons ensuite les données à disposition et enfin nous parlerons de leur qualité.

Description	Date	Heures
Planning général	01.04	8h
État de l'art	01.04 - 31.05	40h
Compréhension du métier	22.05	8h
Connaissance des données	16.04 - 15.06	40h
Préparation des données	16.04 - 31.07	80h
Labellisation des données	30.04 - 13.07	40h
Modélisation des données	28.05 - 05.08	80h
Évaluation	30.07 - 10.08	16h
Déploiement	06.08 - 12.08	24h
Écriture du rapport	01.05 - 20.08	80h
Relecture et correction	01.08 - 23.08	12h

TABLE 4.1: Planification des ressources disponibles.

4.2.1 Collecte des données

Collecter des données est primordial pour un projet de Data Mining. C'est une étape importante qui consomme beaucoup de temps et de ressource. Pour cette recherche, la collecte des données a été effectuée d'une différente façon pour chaque source.

Des données primaires ont été collectées avec un drone DJI Mavic Pro. Pour collecter ces données, l'application Pix4D Capture a été utilisée afin de créer des missions de captures d'images. Ensuite, ces images ont été transférées dans Pix4D Mapper afin de créer des cartes orthogonales grâce à la photogrammétrie.

Après discussion avec l'entreprise Precision Vine¹, ils ont accepté de nous transmettre des images de plusieurs vignobles californiens. Compte tenu de la taille importante des données, nous avons convenu que la compagnie nous transmettrait les données sur une clé USB par Poste. Cependant, leurs données ont finalement été transmises avec le service web WeTransfer.

Enfin, des données de SwissTopo ont été récupérées directement sur le site internet swisstopo.admin.ch.

1. Startup de télédétection dans la viticulture de précision (Personne de contact William Metz).

4.2.2 Description des données

Dans la figure 4.1, nous avons représenté les différents vignobles provenant des trois sources de données. Nous les décrivons en plus de détail ci-dessous ainsi qu'en tableau 4.2.



FIGURE 4.1: Représentation des parcelles sur la carte. À gauche, les données en Californie. À droite, les données dans la vallée du Rhône. Les couleurs des labels représentent les différentes sources de données : Primaires (vert), SwissTopo (rouge), Precision Vine (noir).

Les trois sources de données du projet ont des particularités différentes. Elles ont cependant quelques points communs. Tous les sets de données sont des images orthogonales aériennes de vignobles en ligne. Toutes les images sont en format GeoTIFF et ont au moins les canaux de couleur RGB. Le format GeoTIFF permet aussi de connaître les coordonnées géographiques de chaque pixel sur Terre ainsi que le Ground Sampling Distance (GSD) où combien de centimètres carrés terrestres chaque pixel représente.

Les données récupérées de Precision Vine sont de multiples prises de vue de 7 différents vignobles. Le processus de prise de vue pour ces vignobles a été effectué plusieurs fois durant la même année. Ces différentes prises de vue permettent de voir une évolution dans le temps. Une caméra multispectrale a été utilisée pour ces prises de vue. Chaque image a donc 8 spectres différents.

Les données primaires capturées avec le drone DJI Mavic Pro puis traitées par la suite de logiciels Pix4D ont en plus le Digital Surface Model (DSM) de la zone capturée. Le DSM est une représentation de l'altitude du point le plus haut de la surface représentée par

Source	Location	Surface	Dimensions	GSD
Primaire	Chamoson	39.94 ha	30 244 x 27 000	3.07cm ²
Primaire	Salgesh	38.58 ha	24 448 x 19 981	3.27cm ²
Primaire	Saxon	37.22 ha	26 610 x 22 889	3.40cm ²
Precision Vine	Archimedes	5.08 ha	1 651 x 1 187	-
Precision Vine	Davis	15.02 ha	2 907 x 1 673	-
Precision Vine	Dierberg	33.80 ha	3 012 x 2 943	-
Precision Vine	LyonLakeCounty	16.32 ha	1 606 x 2 669	-
Precision Vine	Parish Vineyard	32.05 ha	4 348 x 3 454	-
Precision Vine	Rose Ranch	17.20 ha	2 565 x 1 584	-
Precision Vine	Safari West Vineyard	59.04 ha	3 318 x 4 702	-
SwissTopo	Aigle	144.00 ha	12 000 x 12 000	10cm ²

TABLE 4.2: Caractéristiques des sources de données.

le pixel, incluant divers obstacles ou objets. À ne pas confondre avec le Digital Ground Model (DGM) qui représente l'altitude du sol.

Les données de SwissTopo n'ont pas de caractéristiques particulières.

4.2.3 Exploration des données

Comprendre les données décrites permet au data scientist de mieux approcher le problème du projet. Pour ce faire, il est nécessaire d'explorer toutes les données en profondeur afin de pouvoir découvrir les problèmes potentiels avant de procéder à la préparation.

Les fichiers GeoTIFF ont pu être visualisés et explorés grâce aux logiciels libres GDAL et QGIS.

Dans ce projet, chaque zone géographique est sauvegardée dans son propre dossier. Pour les données provenant de Precision Vine, chaque dossier est composé de plusieurs fichiers Tiff représentant les différentes prises de vue espacées dans le temps. Les données provenant du traitement de Pix4D Mapper sont représentées de deux manières. Deux grands fichiers représentant le DSM et le RGB de toute la zone traitée et deux dossiers où ces images sont séparées en carreaux de 5 000 x 5 000 pixels. Les données récupérées sur SwissTopo sont principalement dans un grand fichier Tiff de 12 000 x 12 000 pixels avec des métadonnées supplémentaires disponibles.

4.2.3.1 Chamoson-1-4.tif

Afin de mieux visualiser les processus se rapportant aux données, nous utiliserons l'image "Chamoson-1-4.tif" 4.2. Cette image fait partie du set d'évaluation dont on parlera en section 4.4.1. Une version de taille originale de cette image est aussi disponible en annexe A.



FIGURE 4.2: Image de visualisation en entier réduite.

Comme vu en figure 4.2, nous pouvons facilement détecter les différentes lignes de vigne sur une bonne partie de l'image. Cependant, certaines zones, comme le sommet de la figure 4.2, sont trop floues pour pouvoir faire une distinction manuelle. Ceci s'explique soit par un manque de données initiales dans la région soit par une période de l'année où les plantes sont peu différenciées.

4.2.4 Qualité des données

Dans cette partie, nous allons parler de la qualité des données disponibles. Cette section nous permettra de mieux connaître les problèmes que nous devons corriger dans la section 4.3.

L'exploration des données nous a permis de remarquer qu'il est généralement possible pour un humain de distinguer les lignes de vigne du sol sur les données que nous avons à disposition. Toutefois, nous avons détecté trois problèmes qu'il nous faudra résoudre.

Le premier problème de qualité est la différence des informations disponibles entre les différents sets de données. En effet, bien que certaines informations puissent être intéressantes pour reconnaître des lignes de vigne, le fait qu'elles ne soient pas généralement disponibles les rend moins utilisables. De ce fait, les 5 canaux supplémentaires des données de Precision Vine ainsi que le DSM généré par Px4D sont problématiques.

Le deuxième problème à considérer est le format des images. Bien que les GeoTIFF puissent être lus par les bibliothèques GDAL et le programme QGIS, ces drivers sont lents et pourraient engendrer des problèmes dans le processus de modélisation. Pour ces raisons, il nous faudra transformer ces fichiers dans la section 4.3.

Le troisième problème survient lorsque les données contiennent peu de différences entre le sol et les vignes. Par conséquent, la création d'un label de vérité est difficile.

4.3 Préparation des données

La préparation des données s'occupe de traiter les problèmes et caractéristiques des données découvertes durant l'étape précédente. Cette phase est longue dans tout projet de Data Mining car il est important de limiter les défauts dans le set de données.

4.3.1 Annotations des images

La première étape de la préparation des données est d'annoter les images. Ces annotations, ou labels de vérité servent de base de connaissances pour l'apprentissage du modèle. Le but du modèle sera de se rapprocher au plus près de ces annotations. Il faut donc s'assurer de créer des labels corrects.

Nous avons commencé par utiliser le seuil d'Otsu sur l'indice 2G_RBi (2.2) décrit dans la section 2.2.4. Cette base nous a permis d'avancer rapidement dans la première partie des annotations.

Ces résultats devaient cependant être nettoyés pour enlever les faux positifs et rajouter des vignes manquées.

Finalement, chaque ligne de vigne devait être séparée des autres lignes. Cela était nécessaire parce que nous cherchons à détecter les lignes de vigne et non un vignoble.

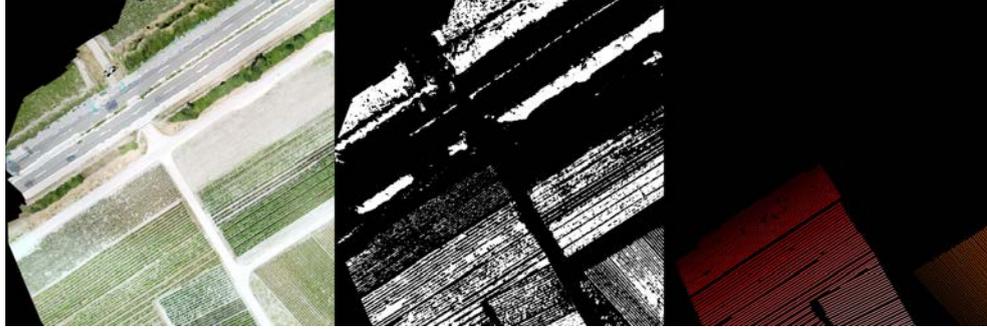


FIGURE 4.3: Annotation des données. L'image source (gauche), résultat du modèle de référence (milieu) et le label de vérité nettoyé (droite). Chaque couleur du label de vérité représente une instance différente.

Nous avons décidé de faire cette séparation en utilisant plusieurs couleurs sur une image avec couleur indexée. Cette façon d'enregistrer les annotations permet de réduire l'espace de stockage utilisé ainsi que d'être facilement visualisable. Les différentes instances sont séparées uniquement lorsque les données sont envoyées au modèle.

4.3.1.1 Filtre d'image floue

Cette étape a aussi permis d'enlever du set de données certaines images qui étaient de trop mauvaise qualité pour assurer la création des labels. Par exemple, nous n'avons ni annoté ni utilisé le sommet de la figure 4.3.

4.3.2 Caractéristiques des données

Le premier problème défini dans la section 4.2.4 était la différence d'information disponible entre les sources de données. Afin de résoudre ce problème, nous avons décidé de n'utiliser que les informations communes à toutes les sources de données. Les canaux RGB et les métadonnées de l'image (taille, GSD). Nous ferons cette réduction en même temps que la transformation des images ci-dessous.

4.3.3 But général des transformations

Pour ce projet, nous avons aussi besoin de transformer les images de plusieurs façons. Les données finales doivent être de même résolution, échelle et posséder les mêmes caractéristiques. Les lignes du vignoble doivent être alignées avec le bord de l'image.

Finalement, nous avons besoin d'augmenter considérablement notre nombre de données. Pour arriver à ces trois objectifs, nous avons établi ces différentes étapes.

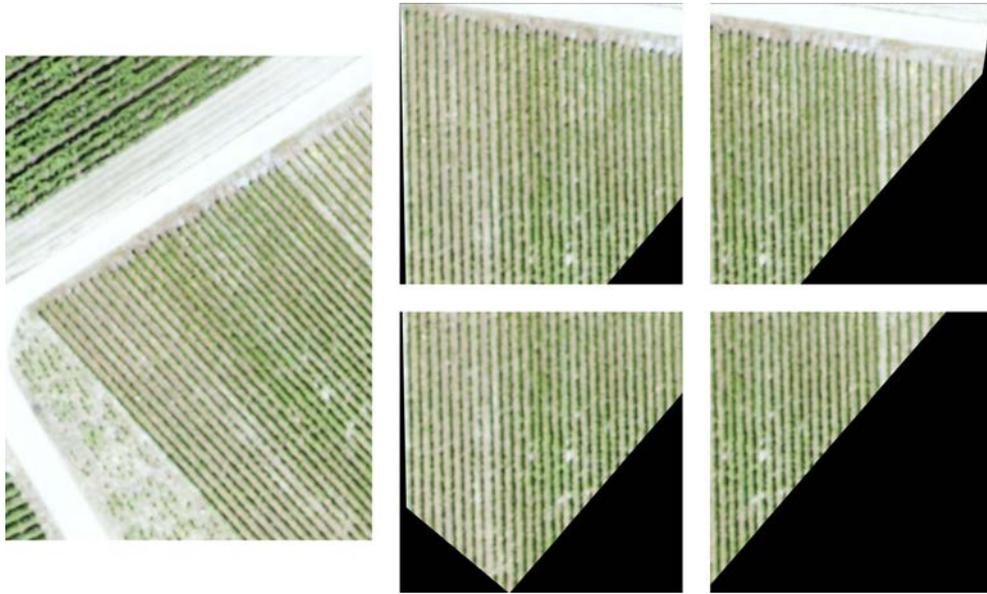


FIGURE 4.4: Transformation complète des données. À gauche une partie originale, à droite les données utilisées pour les modèles.

4.3.3.1 Transformation des images

Afin de faciliter la modélisation, il est nécessaire de transformer les images GeoTIFF au format Portable Network Graphics (PNG). En effet, les différents frameworks de Machine Learning ne supportent pas les images en GeoTIFF.

De plus, les différentes sources de données ont un GSD différent. Ceci pourrait être un problème puisque cela rajoute du bruit dans les données. En effet, une vigne avec une échelle d'un pixel par 10cm^2 ne ressemblera pas à une vigne avec une échelle d'un pixel par 3cm^2 . Il est donc nécessaire de réduire la différence d'échelle des données.

Enfin, uniquement les canaux RGB devaient être utilisés dans cette recherche afin de répondre aux limites prédéfinies.

Pour gérer ces trois transformations, l'outil `gdal_translate` a été utilisé. Cet outil fait partie de GDAL (GDAL/OGR CONTRIBUTORS, 2018).

La librairie GDAL étant libre et cross-plateforme, cette fonction peut facilement être installée et utilisée sur différentes machines ou sur un serveur de production. Cette information est importante pour le déploiement futur.

```
gdal_translate "/path/to/input_file" -b 1 -b 2 -b 3 -tr 0.04 0.04 \  
-of png "/path/to/output_file.png"
```

Ce code permet de modifier l'échelle ou GSD d'une image à un pixel par 4cm^2 , de garder uniquement les canaux RGB et de retranscrire ces informations dans une image PNG. De plus, cette commande génère aussi un fichier de métadonnées ".aux.xml". Celui-ci contient les informations nécessaires pour recréer le fichier GeoTIFF original à partir du fichier PNG.

4.3.3.2 Séparation en vignobles

Une des limites du projet est que nous considérons uniquement des vignobles comme déjà détectés auparavant. Afin de simuler cette opération, les images ont été séparées en plusieurs zones de vignobles.

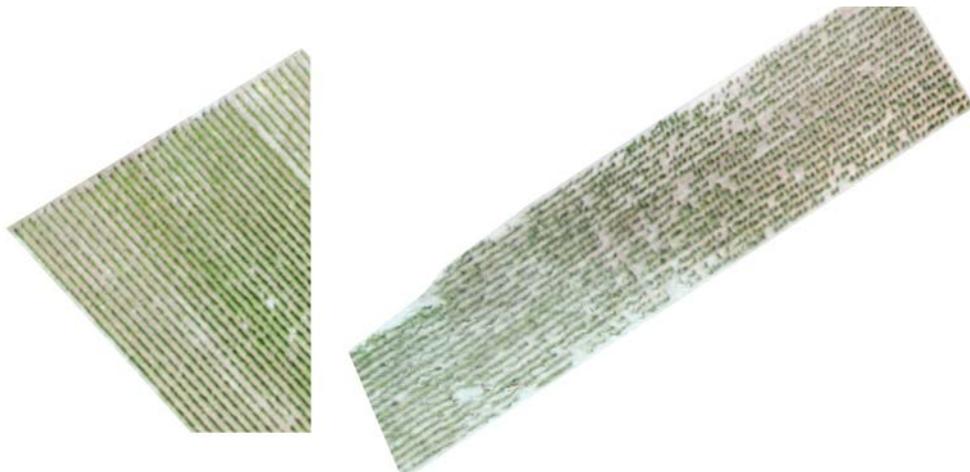


FIGURE 4.5: Représentation de la détection de deux vignobles différents.

Cette modification présentée avec la figure 4.5 permet de nettoyer les bords des vignobles. Vu que plusieurs vignobles sont présents dans chaque image, nous avons décidé de renommer les fichiers en rajoutant l'index du vignoble à la fin du nom du fichier.

4.3.4 Aligner aux bords de l'image

Après plusieurs tentatives d'itération, il s'est avéré que le modèle Mask R-CNN (Mask R-CNN) ne peut pas supporter d'image avec des vignes non alignées avec les bords de

l'image. En effet cela provoque des boîtes englobantes avec de grands chevauchements. Cela réduit considérablement la performance du modèle.

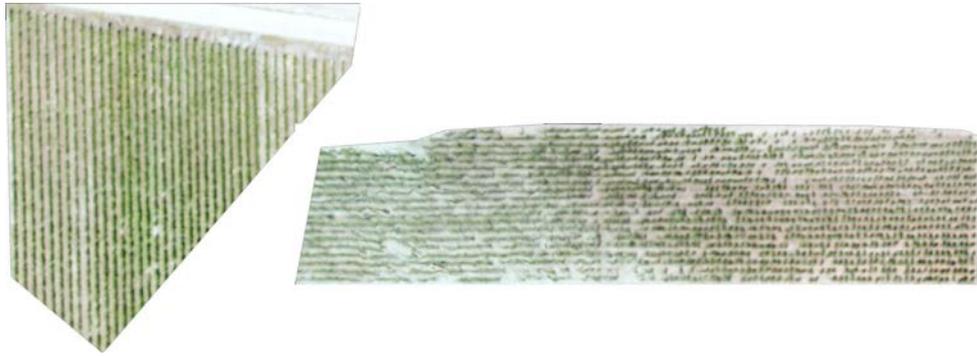


FIGURE 4.6: Les vignobles de figure 4.5 sont maintenant alignés avec l'image.

Pour résoudre ce problème, nous avons décidé d'aligner les vignes avec l'image afin d'entraîner le modèle Mask R-CNN comme présenté en figure 4.6. Ceci n'était pas un aussi grand problème lors de l'utilisation du modèle Faster R-CNN avec annotation alignée à l'objet. Nous avons cependant utilisé les données alignées pendant les tests suivants.

4.3.5 Séparation des images

Maintenant que les vignobles dans leur ensemble sont bien nettoyés, alignés et séparés. Nous pouvons résoudre le problème des tailles d'images différentes et volumineuses. Pour entraîner des R-CNN, les images doivent être de même dimension. Il est aussi plus efficient d'utiliser de petites images pour pouvoir les garder en mémoire facilement.

Pour séparer nos images, nous avons créé un outil permettant de le faire automatiquement. Il sauvegarde ainsi également les découpages effectués. Les nouvelles images ont une dimension de 1024 x 1024 pixels et une zone de chevauchement de 100 pixels entre elles. Ceci permet aux objets se retrouvant au bord d'une zone de ne pas être complètement supprimés. Au bord de l'image, le chevauchement peut être plus important pour éviter de créer des images sans information. Finalement, l'outil assure que le carreau a au moins une annotation. Ceci permet d'éviter des carreaux inutiles qui risquent d'affaiblir le modèle. Les images sont renommées "nom.x1__y1.png" où x1 et y1 sont les valeurs minimums de leur axe respectif. Le vignoble de gauche de la figure 4.6 a été séparé dans les 4 images représentées dans la figure 4.7.

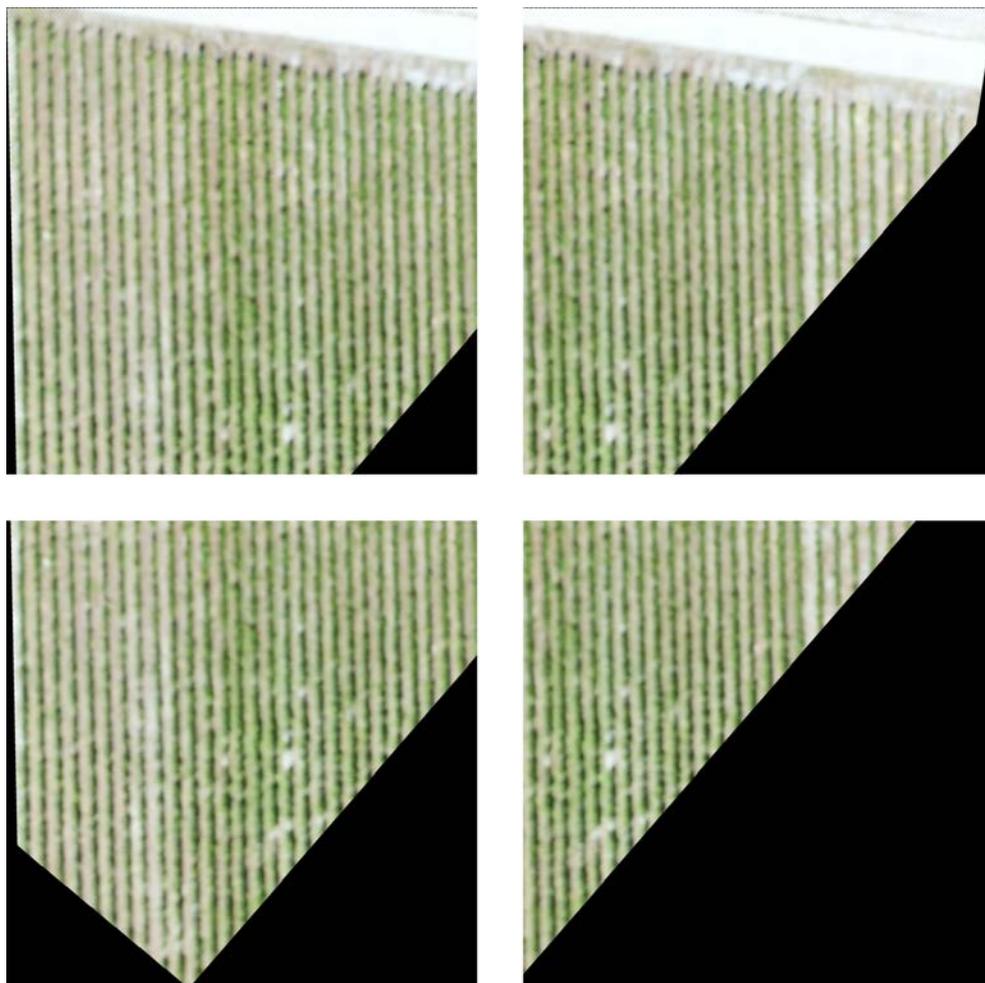


FIGURE 4.7: Séparation en carreaux du vignoble de gauche dans la figure 4.6.

4.3.6 Augmentation

Afin d'améliorer la généralisation du modèle et d'éviter le surapprentissage, il est nécessaire de procéder à de l'augmentation. Nous utilisons le processus décrit en figure 4.8 pour réaliser l'augmentation de notre ensemble de données d'apprentissage.

Pour cela, nous avons décidé d'utiliser la librairie Python `Imgaug`. `Imgaug` permet de facilement modifier les images avec de multiples transformations. Nous avons décidé d'utiliser les augmentations listées dans le tableau 4.3 sur les images. Les différentes augmentations sont appliquées en séquence indépendamment des autres. L'addition et la multiplication de lumière sont par contre exclusives, car leur combinaison était trop disruptive pour le modèle. Avec 9 augmentations distinctes, nous pouvons générer 512 images différentes à partir de la même donnée et ce, sans compter les plages variables.

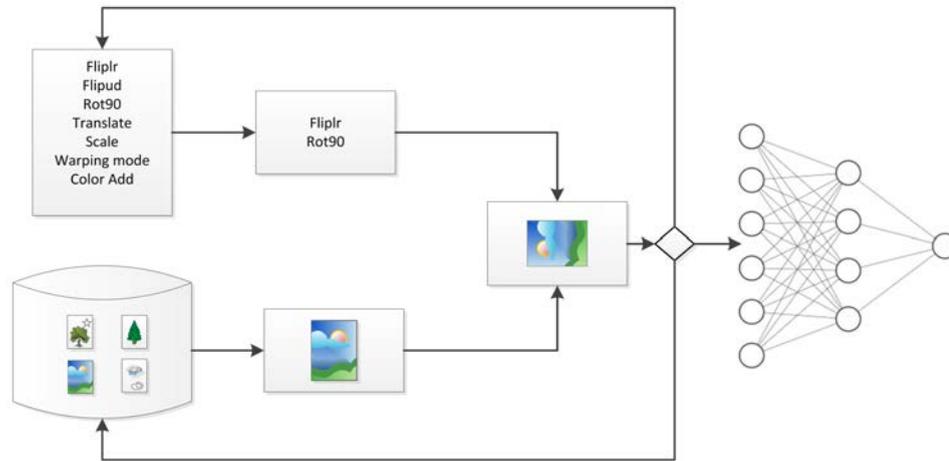


FIGURE 4.8: Processus d'augmentation. Les images sont chargées depuis le set de données, la liste des transformations et leur variable est déterminée aléatoirement puis appliquée à l'image et à ses labels. Les données transformées sont ensuite envoyées au réseau de neurones pour apprentissage. Enfin, les données sont déchargées de la mémoire et une autre image est chargée.

Augmentation	Probabilité	Variable	Description
Fliplr	50 %	-	Symétrie verticale
Flipud	50 %	-	Symétrie horizontale
Rot90	50 %	-	Rotation 90 degrés
Translate	50 %	-50.0 % à 50 %	Translation géométrique
Rotation	50 %	-10.0 à 10	Rotation en degrés
Scale	50 %	50.0 % à 100 %	Réduction de l'échelle de l'image
Warping mode	50 %	all	Remplissage des zones vides
Color add	50 %	-45.0 à 45	Ajoute à la valeur de l'image
Color multiply	50 %	50.0 % à 150 %	Multiplie la valeur de l'image

TABLE 4.3: Configuration des augmentations de données utilisées.

Afin de réduire l'espace pris par le set de données, l'augmentation se fait durant le processus d'apprentissage du modèle. Cela permet à une seule image stockée de représenter de multiples variantes en augmentant ainsi la généralisation du modèle. Bien que la même image peut représenter plusieurs variantes grâce à l'augmentation, il est important de préciser que chaque image de base n'est montrée qu'une seule fois au modèle par cycle d'apprentissage. En effet, même avec augmentation, il est également important que le modèle observe de multiples données bien distinctes.

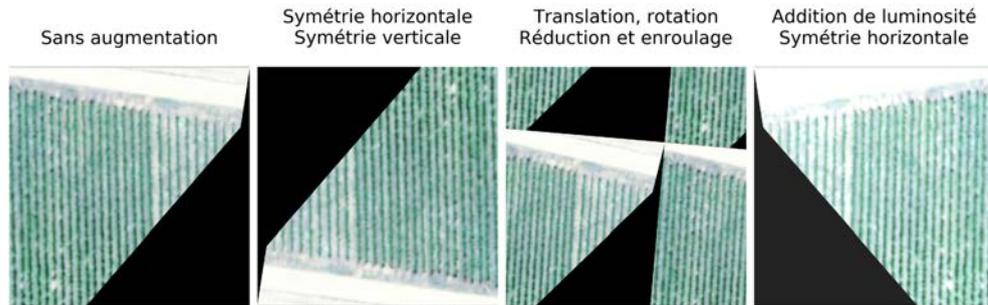


FIGURE 4.9: Augmentation des images, données de base à gauche et trois augmentations aléatoires.

4.4 Modélisation des données

La modélisation des données cherche à créer un modèle capable de prédire une information à partir des données qui lui sont envoyées. En Machine Learning, un processus d'apprentissage modifie différents poids et règles d'importance à l'intérieur du modèle afin de le former aux données qu'il reçoit. Avant de lancer l'apprentissage, nous devons toutefois assurer d'avoir planifié les tests d'évaluation du modèle.

4.4.1 Préparation des tests

Dans les projets de Data Mining, il est important de garder une partie des données en dehors de l'entraînement. En effet, il serait autrement impossible d'assurer que le modèle puisse généraliser en dehors des cas déjà présentés ou non prévus.

Pour ce projet, nous avons séparé les données en 3 sets distincts : Train, Test et Val pour entraînement, test et évaluation.

	Train	Test	Val
Images	2063	188	56

TABLE 4.4: Séparation des sets de données.

Pour réaliser ces ensembles de données, nous avons sorti des images contiguës du set d'entraînement dans chaque source de données. Les données devaient être contiguës à cause du chevauchement lors de la séparation des images (sous-section 4.3.3.2) et pour éviter que les mêmes données se retrouvent dans le set d'entraînement et les deux sets d'évaluation.

Le nombre d’images par set de données est noté dans le tableau 4.4. Le set d’entraînement est celui utilisé pour générer le modèle. Le set de test permet d’observer l’amélioration de la précision du modèle entre les différentes itérations. Le set d’évaluation permet d’avoir des données pour l’évaluation business finale. En effet, il est important d’avoir des données non utilisée pendant les diverses itérations afin d’assurer que le modèle peut fonctionner dans des cas non prévus.

L’évaluation métier sera aussi en partie réalisée sur des images en dehors du set de données et sans labels de vérité. En effet, les données de SwissTopo n’ont pas pu être labellisées par manque de temps. Nous pouvons cependant les utiliser pour assurer que le modèle fonctionne sur des données de différentes sources.

4.4.2 Modèles utilisés

Les deux architectures que nous allons utiliser pour ce projet sont le Mask R-CNN et le Fast R-CNN. Elles ont été respectivement présentées dans les sections 2.1.5 et 2.1.3. Les deux architectures sont soutenues par un ResNet101 (sous-section 2.1.5.1).

	Mask R-CNN	Faster R-CNN
Framework	TensorFlow + Keras	MXNet
Backend	ResNet101	ResNet101
Boîte	Orientée image	Orientée objet
Segmentation	oui	non
Epoch	200	100
Images/epoch	1000	2063
Seuil score	0.7	0.5
Chevauchement	0.5	0.7

TABLE 4.5: Configuration des modèles.

Afin de mener à bien ce projet, nous avons utilisé des implémentations disponibles librement sur GitHub. Cette version du Mask R-CNN a été implémentée par Waleed Abdulla de Matterport, Inc.. Nous avons utilisé cette solution, car elle est particulièrement bien documentée et utilisée dans de nombreux projets. L’implémentation du Faster R-CNN a été effectuée pour le set de données DOTA (section 2.1.6) par Zhen Zhu et Jian Ding (XIA et al., 2018).

4.5 Évaluation

La partie évaluation de CRISP-DM permet d'assurer que les modèles proposés atteignent les buts du métier. Pour cela, nous devons contrôler que les modèles améliorent l'état de l'art pour les buts métier. Une partie de l'évaluation sera présentée dans le chapitre résultat 5.

Deux buts avaient été définis pour le métier. Le premier cherchait à améliorer nos connaissances en matière de détection d'objets dans la viticulture de précision. Le second but était de simplifier la détection des vignes. Pour cela nous devons réduire la quantité de travail humain nécessaire et améliorer la précision.

Ces buts ont été atteints en partie. La précision du modèle est meilleure que celle utilisée comme référence. Nous avons aussi contribué à la viticulture de précision en développant de nouveaux modèles pour la segmentation des vignobles. Cependant, la création de masques précis demande d'aligner les vignes avec l'image, augmentant ainsi le travail humain nécessaire.

Nous avons décidé de déployer uniquement le modèle créé avec l'architecture Mask R-CNN.

4.6 Déploiement

Pour le déploiement du modèle, nous devons modifier le résultat donné à l'utilisateur. En effet, les données nécessaires pour créer le chemin de traitement sont des lignes de coordonnées. Afin de retourner ces valeurs, nous devons nous assurer de pouvoir retransformer les résultats en coordonnées géographiques.

Dans cette optique, nous avons enregistré chaque transformation effectuée sur chaque image durant la section 4.3.3. Grâce à ces informations, nous avons pu réassembler les carreaux et retourner les vignobles. Dès lors, nous avons la possibilité de rassembler les différents vignobles pour créer un seul masque d'annotation.

Enfin, nous pouvons utiliser le fichier image PNG et les fichiers de métadonnées aux.xml créés lors de l'étape de transformation (sous-section 4.3.3.1). Avec ces métadonnées, l'outil `gdal_translate` nous permet de géoréférencer ce masque d'annotation.

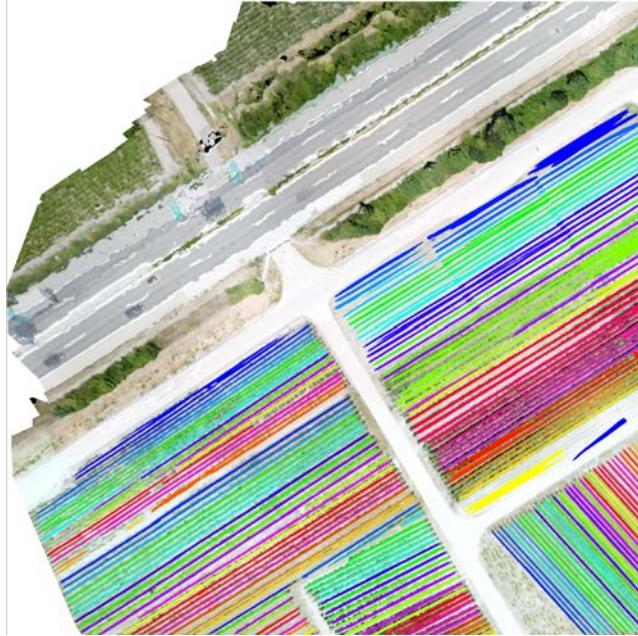


FIGURE 4.10: Masques prédits géoréférencé sur l'image de base.

```
gdal_translate "/path/to/input_file.png" -of GTiff "/path/to/output_file.tif"
```

Nous pouvons donc maintenant envoyer à l'utilisateur les labels sur une carte de la région. Comme nous pouvons le voir dans la figure 4.10, reconvertir les prédictions en GeoTIFF nous permet de bien les superposer à la carte de base (fig 4.2). Cette image est aussi accessible en format original en annexe A.

La dernière étape du déploiement est de rendre l'utilisation du modèle publique. Nous planifions un futur service web qui permettra d'accéder au modèle.

Chapitre 5

Résultats

Dans ce chapitre, nous partagerons les résultats de la recherche et du projet. Les résultats sont séparés en deux parties. Premièrement, nous testerons la précision des modèles proposés contre le modèle de référence. Ensuite, nous testerons le résultat du projet par rapport aux buts métier prédéfinis.

Pour visualiser les résultats des tests, nous utiliserons principalement la figure 5.1. Cette image est intéressante, parce qu'elle est assez précise et avec des tâches d'herbe entre les lignes de vigne. Ces tâches rendent la détection plus difficile. Comme décrit dans la section 4.3.1, les labels de vérités sont séparés en plusieurs couleurs pour représenter les différentes instances de vigne.

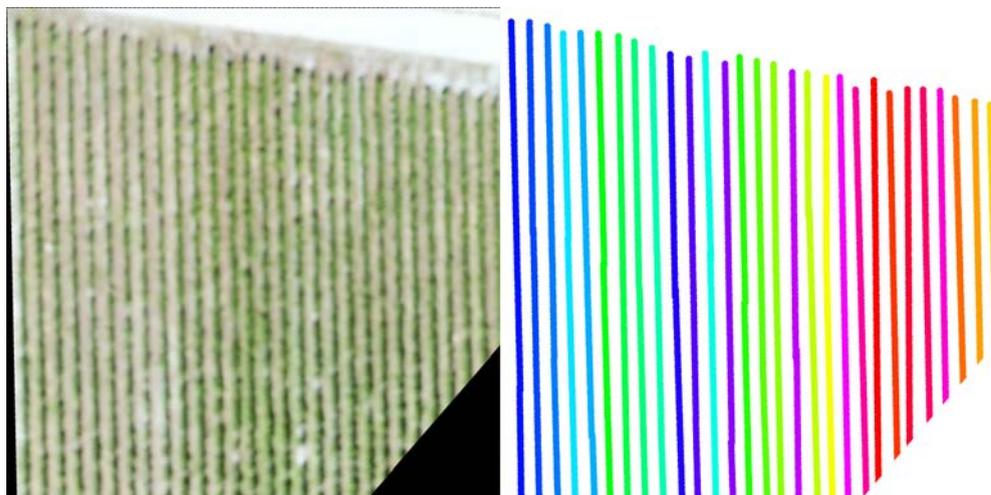


FIGURE 5.1: Image de référence pour les résultats : à gauche les données de base, à droite les labels de vérité.

Comme décrit dans la section 4.4.1, nous utiliserons les 56 images du set de données d'évaluation pour ces tests. Nous évaluerons les modèles avec la métrique Intersection sur Union (IoU) (5.1). Ce ratio permet de contrôler que la prédiction du modèle ressemble aux labels définis au départ.

$$IoU = \frac{TP}{FP + TP + FN} \quad (5.1)$$

5.1 Résultat des modèles

5.1.1 Référence

Afin de pouvoir comparer les modèles proposés, nous allons commencer par calculer et présenter les résultats du modèle de référence (section 2.2.4). Pour cela, nous avons appliqué la formule de référence sur chaque carreau du vignoble. Nous avons par la suite calculé l'IoU en comparaison avec les labels de vérité.

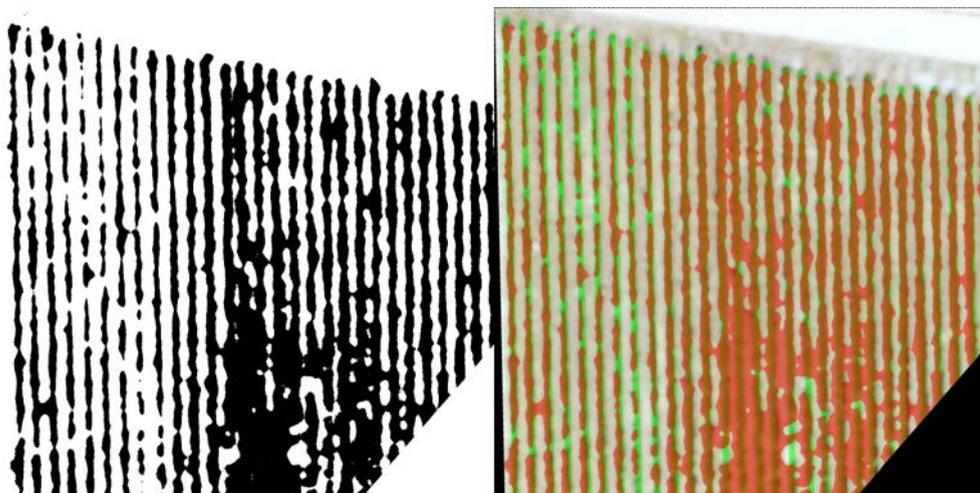


FIGURE 5.2: Résultat du modèle de référence, à gauche le masque généré, à droite la différence entre la groundtruth (vert) et le masque généré (rouge).

L'IoU moyen de cette technique est de 0.43 (table 5.1). Nous pouvons constater dans la figure 5.2 que cette technique crée beaucoup de faux positifs. En effet, cette technique utilise uniquement l'intensité des pixels et peut donc se tromper entre les vignes et d'autres végétations.

		Prédiction du modèle de référence		IoU
		Vigne (noir)	Non-vigne (blanc)	
Vérité	Vigne	10'407'926 0.7398	3'660'846 0.2602	0.4274
	Non-vigne	11'356'299 0.2702	30'665'489 0.7298	

TABLE 5.1: Matrice de confusion du modèle de référence.

5.1.2 Faster R-CNN

L'architecture Faster R-CNN propose en sortie uniquement des boîtes englobantes. Pour faciliter la comparaison avec les autres modèles, nous avons généré des masques à partir de ces boîtes. Nous considérons donc que tous les pixels à l'intérieur de la boîte englobante sont associés à une instance de ligne de vigne.

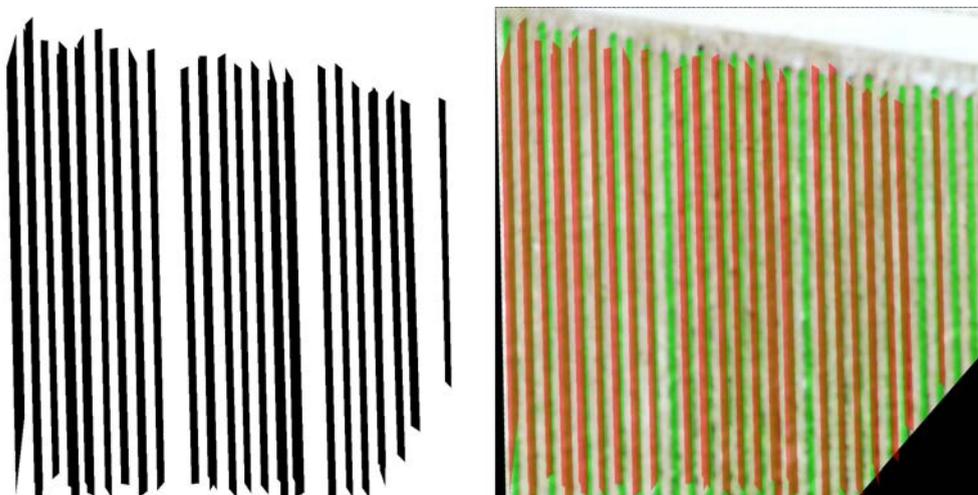


FIGURE 5.3: Résultat du modèle Faster R-CNN, à gauche le masque généré, à droite la différence entre la groundtruth (vert) et le masque généré (rouge).

Le modèle Faster R-CNN a un IoU moyen sur le set de donnée de 0.47 (table 5.2). Avec la figure 5.3, nous pouvons voir que ce modèle ne parvient pas à assurer la détection de plusieurs lignes de vigne. De plus, il a un problème pour détecter les pointes de chaque ligne, la détection semble s'arrêter juste avant chaque fin de ligne.

Ce modèle est quand même plus performant que le modèle de référence car il a un meilleur IoU et parce qu'il réduit le bruit autour des lignes de vigne.

		Prédiction de Faster R-CNN		IoU
		Vigne (noir)	Non-vigne (blanc)	
Vérité	Vigne	10'691'188 0.7599	3'377'584 0.24007667478	0.4674
	Non-vigne	9'466'660 0.2252	32'555'128 0.7747	

TABLE 5.2: Matrice de confusion du modèle Faster R-CNN.

5.1.3 Mask R-CNN

Cette architecture retournant directement des masques de prédiction, nous avons pu utiliser le résultat directement. De plus, nous pouvons facilement garder les différentes instances détectées. Nous représentons ceci avec les différentes couleurs du masque dans la figure 5.4.

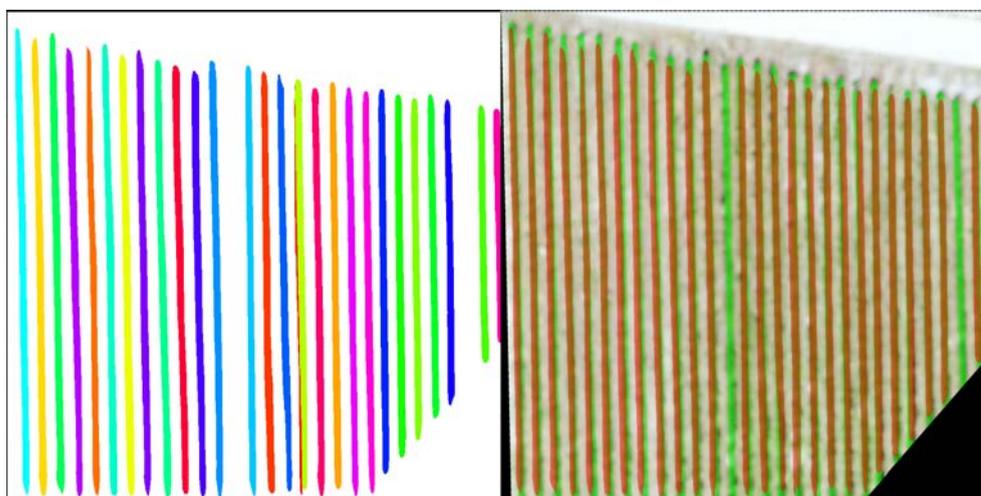


FIGURE 5.4: Résultat du modèle Mask R-CNN, à gauche le masque généré, à droite la différence entre la groundtruth (vert) et le masque généré (rouge).

Sur la figure 5.4 nous pouvons rapidement constater que ce modèle semble mieux détecter les différentes lignes que le modèle de Faster R-CNN et avec moins de bruit que le modèle de référence. Cependant, en regardant les données plus globalement (table 5.3), on remarque que ce modèle a plus de faux négatifs en contrepartie de moins de faux positif. Cette différence fait que ce modèle a un meilleur score IoU que les deux autres modèles : 0.51 .

Ce modèle ayant le meilleur IoU et permettant de facilement détecter des instances de vigne, nous le considérerons pour l'évaluation métier.

		Prédiction de Mask R-CNN		IoU
		Vigne (couleurs)	Non-vigne (blanc)	
Vérité	Vigne	9'945'536 0.7069	4'123'236 0.2931	0.5146
	Non-vigne	6'676'732 0.1579	35'611'360 0.8421	

TABLE 5.3: Matrice de confusion du modèle Mask R-CNN.

5.1.3.1 Évaluation métier

Les meilleures performances de ce modèle et la facilité d'utilisation des résultats en font un candidat propice à l'évaluation métier et peut être au déploiement.

Pour cette évaluation, nous allons principalement contrôler que le modèle soit facilement utilisable. Pour ce faire, nous utiliserons de nouvelles données non utilisées auparavant et non labellisées. Ces données proviennent de SwissTopo. Ce test montrera le résultat attendu sur des données futures par d'autres utilisateurs.

Nous avons sélectionné l'image de gauche dans la figure 5.5 parce qu'il était facile de détecter visuellement les lignes de vigne pour un humain. Cependant, vers le sommet de l'image, il n'y a que peu de différences de couleur entre l'herbe et la vigne. Cette partie est donc plus difficile à détecter automatiquement.

Pour arriver au résultat à droite de la figure 5.5, nous avons uniquement dû annoter les vignobles puis lancer le processus. Comme décrit en section 4.3, le processus suit ces différentes étapes :

1. Séparation des images en différents vignobles selon les annotations
2. Rotation des images de manière à créer le plus petit rectangle possible (et aligner les vignobles à l'image)
3. Agrandissement des images pour atteindre un GSD de $4cm^2$
4. Séparation des images en carreaux de 1024 x 1024 pixels
5. Détection des différentes lignes de vignes par le modèle Mask R-CNN
6. Inversion des différentes transformations pour revenir aux dimensions de la donnée

Le processus de détection a toujours quelques problèmes qui devraient être améliorés. Premièrement, un humain doit pour l'instant toujours délimiter les zones de vignobles

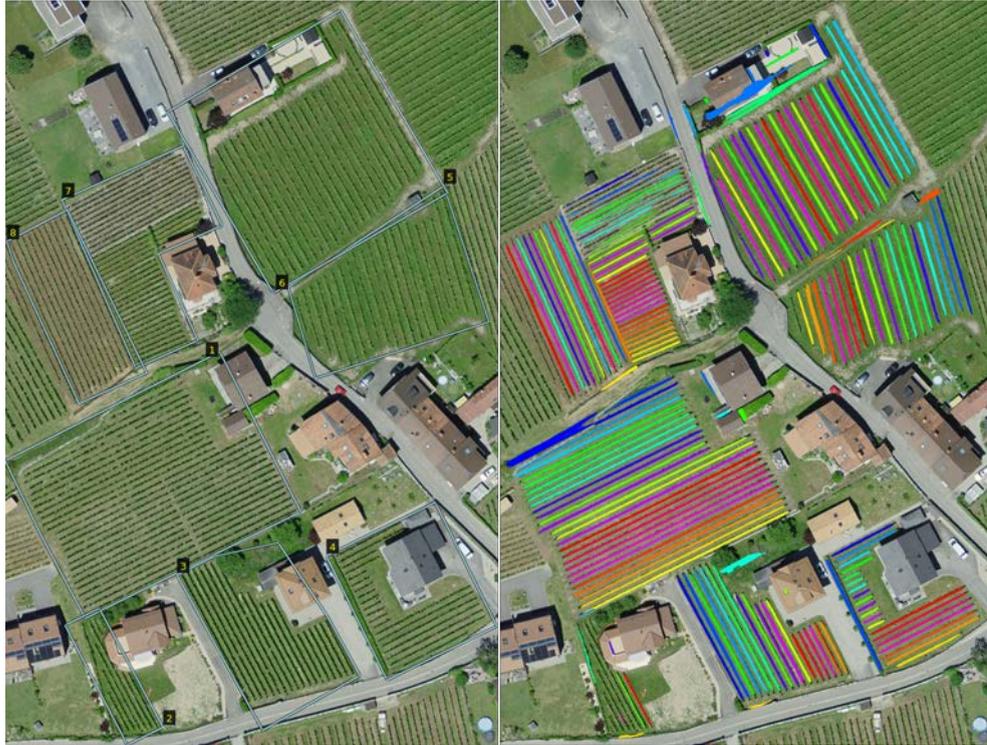


FIGURE 5.5: Évaluation métier du modèle Mask R-CNN. À gauche, l'image de entière avec annotation de séparation de vignoble. À droite, le résultat du processus de détection entier.

avec un bon alignement. Ceci pourrait être amélioré en ajoutant une couche de détection de vignoble avant la détection des lignes de vigne. Secondement, deux des 8 vignobles n'ont été annotés que très partiellement (2 et 7). Finalement, plusieurs objets non voulus ont été annotés comme vigne. Par exemple, la maison au sommet du vignoble 6. Ces deux derniers problèmes pourraient être diminués en augmentant la quantité de données vues par le modèle et en rajoutant principalement des images avec ce genre d'objets. En effet, nous émettons l'hypothèse que ce problème vienne du fait que le modèle n'ait principalement vu que des vignobles.

5.2 Tableau des résultats

Dans cette section, nous consolidons les résultats des différents modèles pour aider la comparaison.

	Référence	Faster R-CNN	Mask R-CNN
TP	10 407 926	10 691 188	9 945 536
% TP	73.98%	75.99%	70.69%
FN	3 660 846	3 377 584	4 123 236
% FN	26.02%	24.00%	29.31%
TN	30 665 489	32 555 128	35 611 360
% TN	72.98%	77.47%	84.21%
FP	11 356 299	9 466 660	6 676 732
% FP	27.02%	22.52%	15.79%
IoU	0.4274	0.4674	0.5146

TABLE 5.4: Tableau récapitulatif des résultats.

Chapitre 6

Discussion

Ce chapitre est une discussion sur l'ensemble du projet et de son déroulement. Ici, nous parlerons des difficultés rencontrées pendant la réalisation du projet. Nous discuterons aussi des limites nécessaires à la focalisation du projet et des améliorations pour de futures recherches.

6.1 Difficultés

La première difficulté rencontrée était le temps de création du set de données. Bien que ce temps ait été prévu dès le démarrage de la recherche, nous n'avions pas pensé aux problèmes de saison. En effet, les lignes de vigne sont plus facilement détectables lorsque les couronnes ont bien repoussé. Ce problème nous a empêché d'utiliser les premières tentatives de prise de données.

Le second ralentissement a été le manque de connaissances dans le Deep Learning et la classification d'images. Bien que nous ayons déjà utilisé ces techniques auparavant, il a fallu mieux les comprendre afin de mener ce projet à bout. Ces compétences acquises permettront d'améliorer les prochaines recherches.

Le dernier problème est en lien avec le précédent. Le temps nécessaire pour l'entraînement des différents modèles a été mal jugé. Nous savions qu'un réseau de neurones était long à converger mais nous n'avions pas prévu le nombre d'itérations nécessaires avant qu'un modèle soit convenable. En effet, chaque itération prenait plusieurs heures de calcul avant de pouvoir tester si le modèle était utilisable ou si un problème était survenu.

6.2 Limites et amélioration

Ce projet s'est limité dès le départ à traiter uniquement des données de vignobles afin d'éviter d'augmenter le bruit dans l'apprentissage du modèle. Il serait tout de même bénéfique d'intégrer cet outil à la suite d'une détection de vignoble. En effet, cela permettrait de réduire la quantité de travail humain nécessaire dans le prétraitement.

Ce modèle s'est aussi limité à des données RGB. D'autres informations pourraient être utilisées selon les moyens d'acquisition. Par exemple, un DSM précis pourrait apporter de très bonnes informations au modèle en utilisant la différence entre le sol et le sommet de la vigne.

Il serait aussi intéressant d'intégrer le système d'annotations orientées objet dans l'architecture Mask R-CNN. Cela pourrait permettre de mieux gérer l'orientation des vignobles. Ceci permettrait également de réduire le travail manuel nécessaire à la segmentation.

Finalement, il serait intéressant d'utiliser ce modèle pour faire de l'Active Learning. Nous pensons en effet que les prédictions sont assez bonnes pour rapidement être corrigées et les assimiler au modèle. Un processus d'Active Learning permettrait de rapidement accroître les données labellisées.

Chapitre 7

Conclusion

Nous finissons ce rapport avec une récapitulation de toutes ses différentes parties. Nous avons commencé le rapport avec un état de l'art. Celui-ci nous a été utile afin de connaître les différentes solutions déjà utilisées pour la classification d'image et la viticulture de précision. Notre recherche de projets similaires nous a permis de créer une référence ainsi que de faciliter la création des labels de vérité. Nous avons en effet utilisé l'indice spectral présenté dans la section 2.2.4 pour ces deux tâches. Du côté Machine Learning, nous avons présenté des architectures de réseaux de neurones. Cela nous a permis de découvrir les architectures Faster R-CNN (section 2.1.3) et Mask R-CNN (section 2.1.5). Nous avons utilisé ces deux architectures pour créer nos modèles de détection de lignes de vigne.

Nous avons ensuite utilisé la méthodologie CRISP-DM afin de mener à bien le projet. Cette méthodologie propose de gérer les projets de Data Mining en plusieurs itérations. Cela permet de réduire les risques provoqués par des imprévus et de rester flexible durant le projet.

L'objectif de l'étude décrit au départ a été atteint, celui-ci était décrit comme suit :

- Établir un modèle de reconnaissance d'objets agricoles à l'intérieur d'images de vignobles.
- Différencier entre un cep de vigne, sa couronne et la terre, herbes, routes ou autres objets alentour.
- Utiliser ce modèle doit être facile et réduire le temps de travail humain dans la détection des objets agricoles.

Généralement nous avons réussi à différencier les lignes de vigne du reste du vignoble avec les deux modèles appris. Cependant, nos modèles ne distinguent pas d'autres classes. Le dernier but n'a aussi que partiellement été atteint. Le modèle requiert d'installer un framework de Machine Learning et d'annoter les vignobles au départ. Néanmoins le résultat qu'il prédit permet de réduire considérablement le temps nécessaire au nettoyage des données.

L'IoU moyen final du modèle Mask R-CNN est de *0.51* et est supérieur au modèle de référence. Les erreurs reposent principalement sur les bords des vignes détectées. L'évaluation métier a été réalisée avec succès. Celle-ci nous a permis de voir comment le modèle réagissait avec une source de données qu'il n'avait jamais vues.

Nous pouvons continuer la recherche de ce projet sur plusieurs chemins différents que nous avons défini dans la section des améliorations 6.2.

Bibliographie

- AGROFLY. (2016). High Precision Forecast with Agro Drones (UAV). *SPSS*.
- BLASCHKE, T., HAY, G. J., KELLY, M., LANG, S., HOFMANN, P., ADDINK, E., ...
TIEDE, D. (2014). Geographic Object-Based Image Analysis – Towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 180–191. doi :<https://doi.org/10.1016/j.isprsjprs.2013.09.014>
- CHAPMAN, P., CLINTON, J., KERBER, R., KHABAZA, T., REINARTZ, T., SHEARER, C. & WIRTH, R. (2000). CRISP-DM 1.0 Step-by-step data mining guide.
- COMBA, L., GAY, P., PRIMICERIO, J. & AIMONINO, D. R. (2015). Vineyard detection from unmanned aerial systems images. *Computers and Electronics in Agriculture*, 114, 78–87. doi :<https://doi.org/10.1016/j.compag.2015.03.011>
- CONESE, C., RAPI, B., ROMANI, M., BONORA, L. & DAMBRA, C. (2005). BACCHUS–Methodological Approach for Vineyard Inventory and Management : VINEYARD ANALYSIS. European Commission, DG Research.
- DAI, J., HE, K. & SUN, J. (2015). Instance-aware Semantic Segmentation via Multi-task Network Cascades. *CoRR*, *abs/1512.04412*. arXiv : 1512.04412. Récupérée à partir de <http://arxiv.org/abs/1512.04412>
- DELENNE, C., DURRIEU, S., RABATEL, G. & DESHAYES, M. (2010). From pixel to vine parcel : A complete methodology for vineyard delineation and characterization using remote-sensing data. *Computers and Electronics in Agriculture*, 70(1), 78–83. doi :<https://doi.org/10.1016/j.compag.2009.09.012>
- DURO, D. C., FRANKLIN, S. E. & DUBÉ, M. G. (2012). A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sensing of Environment*, 118, 259–272. doi :<https://doi.org/10.1016/j.rse.2011.11.020>

- GDAL/OGR CONTRIBUTORS. (2018). *GDAL/OGR Geospatial Data Abstraction software Library*. Open Source Geospatial Foundation. Récupérée à partir de <http://gdal.org>
- GIRSHICK, R. B., DONAHUE, J., DARRELL, T. & MALIK, J. (2013). Rich feature hierarchies for accurate object detection and semantic segmentation. *CoRR*, *abs/1311.2524*. arXiv : 1311.2524. Récupérée à partir de <http://arxiv.org/abs/1311.2524>
- GOODFELLOW, I., BENGIO, Y. & COURVILLE, A. (2016). *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press.
- HALL, A., LOUIS, J. & LAMB, D. (2003). Characterising and mapping vineyard canopy using high-spatial-resolution aerial multispectral images. *Computers & Geosciences*, *29*(7), 813–822. doi :[https://doi.org/10.1016/S0098-3004\(03\)00082-7](https://doi.org/10.1016/S0098-3004(03)00082-7)
- HE, K., GKIOXARI, G., DOLLÁR, P. & GIRSHICK, R. B. (2017). Mask R-CNN. *CoRR*, *abs/1703.06870*. arXiv : 1703.06870. Récupérée à partir de <http://arxiv.org/abs/1703.06870>
- HE, K., ZHANG, X., REN, S. & SUN, J. (2015). Deep Residual Learning for Image Recognition. *CoRR*, *abs/1512.03385*. arXiv : 1512.03385. Récupérée à partir de <http://arxiv.org/abs/1512.03385>
- KARAKIZI, C., OIKONOMOU, M. & KARANTZALOS, K. (2016). Vineyard Detection and Vine Variety Discrimination from Very High Resolution Satellite Data. *Remote Sensing*, *8*(3). doi :[10.3390/rs8030235](https://doi.org/10.3390/rs8030235)
- KARPATHY, A. (2017). Convolutional Neural Networks (CNNs / ConvNets). Récupérée 10 juillet 2018, à partir de <http://cs231n.github.io/convolutional-networks/>
- KRIZHEVSKY, A., SUTSKEVER, I. & HINTON, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In F. PEREIRA, C. J. C. BURGESS, L. BOTTOU & K. Q. WEINBERGER (Éds.), *Advances in Neural Information Processing Systems 25* (p. 1097–1105). Curran Associates, Inc. Récupérée à partir de <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>
- LEE, J., KIM, T., PARK, J. & NAM, J. (2017). Raw Waveform-based Audio Classification Using Sample-level CNN Architectures. *CoRR*, *abs/1712.00866*. arXiv : 1712.00866. Récupérée à partir de <http://arxiv.org/abs/1712.00866>
- LOPEZ, M. M. & KALITA, J. (2017). Deep Learning applied to NLP. *CoRR*, *abs/1703.03091*. arXiv : 1703.03091. Récupérée à partir de <http://arxiv.org/abs/1703.03091>

- MATESE, A., TOSCANO, P., DI GENNARO, S. F., GENESIO, L., VACCARI, F. P., PRIMICERIO, J., ... GIOLI, B. (2015). Intercomparison of UAV, Aircraft and Satellite Remote Sensing Platforms for Precision Viticulture. *Remote Sensing*, 7(3), 2971–2990. doi :10.3390/rs70302971
- MOHANTY, S. P. (2018). CrowdAI Mapping Challenge 2018 : Baseline with Mask RCNN. <https://github.com/crowdai/crowdai-mapping-challenge-mask-rcnn>. GitHub.
- PINHEIRO, P. H. O., COLLOBERT, R. & DOLLÁR, P. (2015). Learning to Segment Object Candidates. *CoRR*, abs/1506.06204. arXiv : 1506.06204. Récupérée à partir de <http://arxiv.org/abs/1506.06204>
- POBLETE-ECHEVERRÍA, C., OLMEDO, G. F., INGRAM, B. & BARDEEN, M. (2017). Detection and Segmentation of Vine Canopy in Ultra-High Spatial Resolution RGB Imagery Obtained from Unmanned Aerial Vehicle (UAV) : A Case Study in a Commercial Vineyard. *Remote Sensing*, 9(3). doi :10.3390/rs9030268
- REN, S., HE, K., GIRSHICK, R. B. & SUN, J. (2015). Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks. *CoRR*, abs/1506.01497. arXiv : 1506.01497. Récupérée à partir de <http://arxiv.org/abs/1506.01497>
- SCHWARZER, S., ALLENBACH, K. & JAQUET, J.-M. (2009). Cartographie semi-automatisée et multi-sources de la couverture du sol. 1. Périmètre de l'agglomération franco-valdo genevoise (CCSA04, moyenne résolution). *Archives des Sciences*, 62, 107–124. ID : unige :32677. Récupérée à partir de <https://archive-ouverte.unige.ch/unige:32677>
- SIMONYAN, K. & ZISSERMAN, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, abs/1409.1556. arXiv : 1409.1556. Récupérée à partir de <http://arxiv.org/abs/1409.1556>
- TOUZET, C. (1992). *les reseaux de neurones artificiels, introduction au connexionnisme*. EC2.
- TREBOUX, J. (2017, octobre 30). *A comparison of machine learning algorithms performance for vineyards detection* (mém.de mast., Université de Fribourg).
- UIJLINGS, J. R. R., van de SANDE, K. E. A., GEVERS, T. & SMEULDERS, A. W. M. (2013). Selective Search for Object Recognition. *International Journal of Computer Vision*, 104(2), 154–171. doi :10.1007/s11263-013-0620-5
- VELIČKOVIĆ, P. (2017). Deep learning for complete beginners : convolutional neural networks with keras. Récupérée 20 juin 2018, à partir de <https://cambridgespark.com/content/tutorials/convolutional-neural-networks-with-keras/index.html>

- WASSENAAR, T., ROBBEZ-MASSON, J.-M., ANDRIEUX, P. & BARET, F. (2002). Vineyard identification and description of spatial crop structure by per-field frequency analysis. *International Journal of Remote Sensing*, 23(17), 3311–3325. doi :10.1080/01431160110076144. eprint : <https://doi.org/10.1080/01431160110076144>
- XIA, G.-S., BAI, X., DING, J., ZHU, Z., BELONGIE, S., LUO, J., ... ZHANG, L. (2018). DOTA : A Large-Scale Dataset for Object Detection in Aerial Images. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- YU, Q., GONG, P., CLINTON, N., BIGING, G., KELLY, M. & SCHIROKAUER, D. (2006). Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering & Remote Sensing*, 72(7), 799–811.

Attestation d'authenticité

Je, Vincent SCHÜLÉ, déclare sur l'honneur, que j'ai effectué ce Travail de Master seul, sans autre aide que celles dûment signalées dans les références, et que je n'ai utilisé que les sources expressément mentionnées. Je ne donnerai aucune copie de ce rapport à un tiers sans l'autorisation conjointe du Responsable de l'Orientation et du Professeur chargé du suivi du Travail de Master et de l'institution ou entreprise pour laquelle ce travail a été effectué.

Date:

Signature:

Glossaire

Active Learning Processus de Machine Learning où l'utilisateur peut corriger les prédictions du modèle durant l'apprentissage.. 43

Data Mining Processus de découvertes de nouveaux modèles dans un grand set de données. 2, 15–20, 24, 31, 44

DJI Mavic Pro Petit drone de photographie aérienne conçu par la compagnie DJI. 1, 20, 21

DOTA Large set de données pour la détection d'objet avec images aériennes. 9, 32

GDAL Librairie libre qui traduit les données géospatiales entre divers format. 22, 24, 26

GeoTIFF Format standard libre pour géo-référencer des images TIFF. 21, 22, 24, 26, 27, 34

GitHub Service internet d'hébergement de code ordinateur. 32

Machine Learning Domaine de recherche en informatique qui cherche à apprendre aux ordinateurs via des techniques de statistiques. vi, 1–3, 26, 31, 44, 45

photogrammétrie Technique de mesure utilisant des photos de différents points de vue pour recréer le relief. 20

Pix4D Capture Application mobile de la compagnie Pix4D permettant de créer des missions de captures photographiques aériens. 20

Pix4D Mapper Logiciel de la compagnie Pix4D permettant la création de cartes orthogonales à partir de photos. 20, 22

Precision Vine Startup dans la télédétection pour la viticulture de précision, William Metz, www.precisionvine.com. 20–22, 24

Python Langage de programmation interprété haut niveau. 29

QGIS Projet Libre et Open-Source de Système d'Information Géographique. 22, 24

SwissTopo Office fédéral suisse de la topographie. 14, 20–22, 32, 39

WeTransfer Web service de transfert de données. 20

Abbreviations

CNN Réseau Neuronal Convolutif. 4, 5, 18

CRISP-DM Cross Industry Standard Process for Data Mining. 2, 15, 18, 33, 44

DGM Digital Ground Model. 22

DSM Digital Surface Model. 21, 22, 24, 43

GIS Système d'Information Géographique. 11, 12

GSD Ground Sampling Distance. 21, 22, 25–27, 39

IoU Intersection sur Union. 36–39, 45

Mask R-CNN Mask R-CNN. 27, 28, 32

NDVI Indice de Végétation. 12

PNG Portable Network Graphics. 26, 27, 33

R-CNN Régions avec Réseau Neuronal Convolutif. 3, 5, 7, 8, 14, 27, 28, 32, 33, 37–40, 44

ResNet Residual Network. 9, 32

RGB Rouge, Vert et Bleu. 1, 10–12, 21, 22, 25–27, 43

UAV Unmanned Aircraft Vehicle. 13, 14

Table des figures

2.1	Réseau de neurones	4
2.2	Couche de convolution	5
2.3	Architecture R-CNN	5
2.4	Architecture R-CNN	6
2.5	Faster R-CNN : Anchor box et résultat	7
2.6	Segmentation d'instances	7
2.7	Architecture Mask R-CNN	8
2.8	DOTA : Style de boîtes englobantes	9
2.9	Analyse de fréquence	10
2.10	Seuil et cluster	11
2.11	Analyse basée objet	11
2.12	Segmentation d'instances	13
3.1	Phase de CRISP-DM	16
4.1	Carte des parcelles	21
4.2	Carreau "Chamoson-1-4.tif"	23
4.3	Annotation des données	25
4.4	Transformation général des données	26
4.5	Séparation des vignobles	27
4.6	Alignement des données	28
4.7	Séparation en carreaux	29
4.8	Processus d'augmentation	30
4.9	Augmentation des données	31
4.10	Prédiction sur carte	34
5.1	Image de référence pour les résultats	35
5.2	Résultat du modèle référence	36
5.3	Résultat du modèle Faster R-CNN	37
5.4	Résultat du modèle Mask R-CNN	38
5.5	Évaluation métier du modèle Mask R-CNN	40
A.1	Carreau "Chamoson-1-4.tif"	56
B.1	Prédiction sur carte	57

Liste des tableaux

4.1	Planification des ressources disponibles.	20
4.2	Caractéristiques des sources de données.	22
4.3	Configuration des augmentations de données utilisées.	30
4.4	Séparation des sets de données.	31
4.5	Configuration des modèles.	32
5.1	Matrice de confusion du modèle de référence.	37
5.2	Matrice de confusion du modèle Faster R-CNN.	38
5.3	Matrice de confusion du modèle Mask R-CNN.	39
5.4	Tableau récapitulatif des résultats.	41

Annexe A

Chamoson-1-4.tif



FIGURE A.1: Image de visualisation en taille originale.

Annexe B

Résultat sur Chamoson-1-4.tif

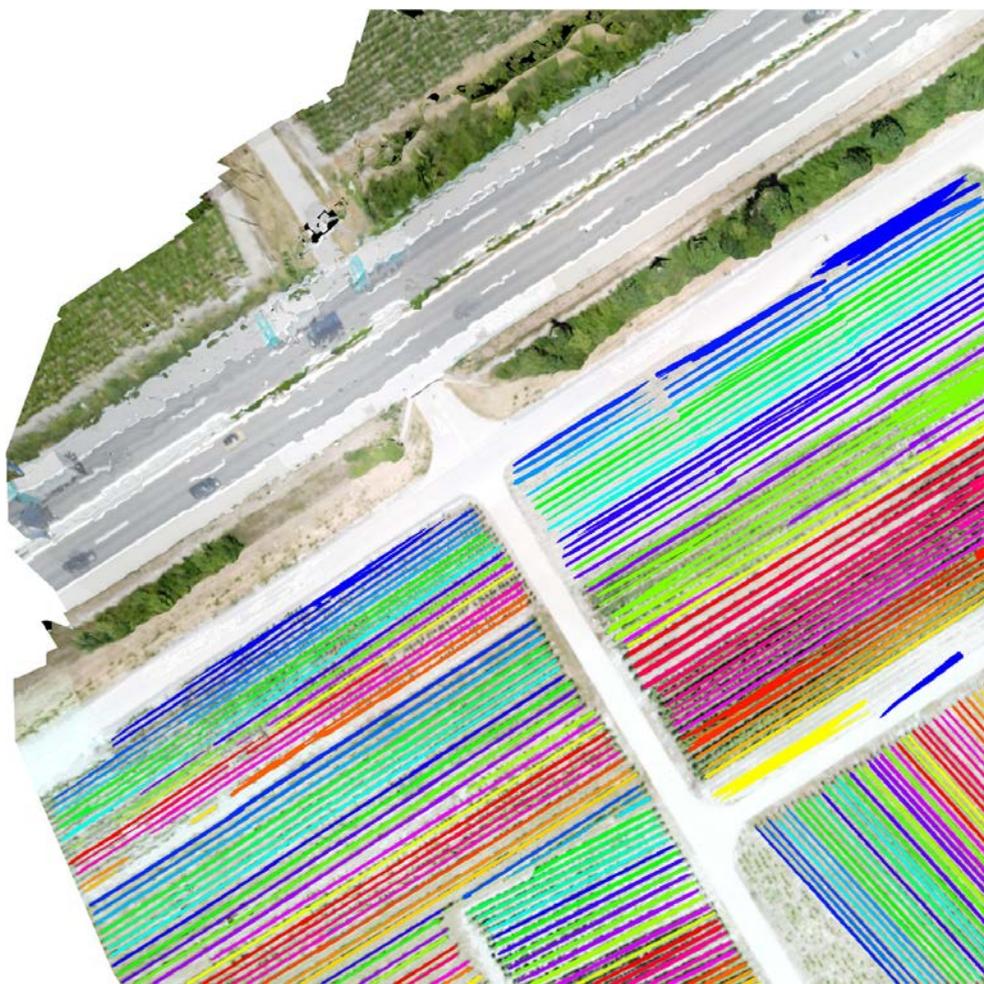


FIGURE B.1: Masques prédits géoréférencé sur l'image de base en taille originale.