

Rapport de stage
Master 2e année 2020-21 :
Système embarqué pour le traitement du signal, de
l'image et du son

Réalisé au laboratoire du LIMOS
dans le cadre du projet ACCRIL

Par Philippe Bauwens

Sujet :
Segmentation sémantique sur
orthophotographie pour générer des cartes
en relief à destination des instructeurs en
locomotion

Remerciement :

Je souhaiterais dans un premier temps remercier Jean-Marie Favreau, mon tuteur de stage, qui a été très enthousiaste et investi dans le suivi de mon travail. En gardant contact avec des échanges réguliers et agréables, il a su me guider et m'aider tout au long de mes recherches et expérimentations.

De même j'aimerais remercier Jérémy Kalsron, doctorant au LIMOS, qui a suivi mon travail, avec autant d'intérêt et de curiosité, en participant aux échanges, posant des questions et en me formant sur la prise en main de QGIS.

Aussi, je remercie le LIMOS qui a mis tout en œuvre pour maintenir un travail en présentiel durant cette crise du Covid-19. Ils ont su nous offrir un cadre de travail propice dans leur loco avec une salle et du matériel de qualité.

Par ailleurs, je souhaite remercier les personnes extérieures avec qui j'ai pu échanger comme Sylvain Huin de la startup FeelObject qui développe une tablette avec un dispositif en relief à destination des déficients visuelles. Mais aussi Didier Leotoing, instructeur en locomotion à la retraite, qui nous a parlé de son travail et de ses techniques afin de mieux appréhender leurs besoins.

Bien évidemment, je remercie tous les autres stagiaires et doctorants avec qui je partageais les bureaux. Nous avons un très bon cadre de travail et une très bonne ambiance.

Table des matières

Chapitre 1 : Context	
1 - Limos	4
2 - ACCRIL	4
2 - Thèse de Gauthier Fillières-Riveau	4
4 - Sujet	4
3 - Plan	5
Chapitre 2 : Création du jeu de données	
1 - Description du jeu de données	5
2 - Sémantique	6
3 - Réalisation de la vérité terrain segmentée	6
3.1 Choix des couches	6
3.2 Ajustements et corrections	7
4 - Orthophotographie	7
5 - Enrichissement avec OpenStreetMap	7
5.1 - Filaire de la chaussée	8
5.2 - Point des passages piétons	9
5.3 - Contours du bâti	10
5.4 - Affinage	10
5.5 - Inconvénients	11
5.6 - Intégration	11
6 - Génération du jeu de données	11
Chapitre 3 : L'architecture	
1 - Description d'un GAN	12
2 - Description d'un cGAN	13
3 - Description de pix2pix	14
3.1 Générateur	15
3.2 Discriminateur	17
Chapitre 4 : Evaluation	
1 - Stratégie	18
2 - Evaluation visuelle	18
3 - Métriques	18
1 - Premiers tests	19
2 - Amélioration du jeu de données	21

3 - Batch size	22
4 - Label Smoothing	23
5 - Noisy Label	25
6 - Combinaison de Label Smoothing et Noisy Label	26
7 - Conclusions des expérimentations	27
Chapitre 6 : Conclusion	29
Conclusion personnelle	30
Glossaire	30
Annexe	30-34
Référence	35

Table des figures

figure 1-2 : échantillon d'un couple d'image	5
figure 3 : Exemple d'une segmentation sémantique	6
Figure 4 : Représentation de l'enrichissement du filaire de la chaussée	8
Figure 5 : Représentation de l'enrichissement des points pour passages piétons	9
Figure 6 : Schéma de la segmentation avec les enrichissements du bâti	10
Figure 7 : Résultat de la génération aléatoire des points (roses) le long des axes routiers.	12
Figure 8 : Schéma représentant le fonctionnement d'un GAN	13
Figure 9 : Schéma représentant le fonctionnement d'un GAN	13
Figure 9 : Schéma représentant l'adaptation CGAN pour pix2pix.	15
Figure 10: Schéma représentant la structure du cgan de pix2pix,	15
Figure 11 : Schéma des couches du générateur	16
Figure 12 : Schéma représentant les couches du discriminateur	17
Figure 13 : Analyse de chaque classe entre la vérité terrain à gauche et la prédiction à droite	18
Figure 14 : Échantillon du premier modèle réalisé basé sur pix2pix sans amélioration	19
Figure 15 : Échantillons du premier modèle basé sur pix2pix sans amélioration, dans une situation d'obstruction de la vision par la végétation	19-20
Figure 16 : Échantillon du premier modèle basé sur pix2pix sans amélioration, dans une situation d'absence de donnée d'enrichissement	20
Figure 17 : Tableau de l'évaluation du premier modèle	21
Figure 18 : Échantillon du jeu de données avec lissage à gauche et sans lissage à droite.	22
Figure 19 : Graphique de l'évaluation du modèle avec Label Smoothing et le nouveau jeu donnée non lissé	23
Figure 20 : Graphique de l'évaluation du modèle avec Label Smoothing et l'ancien jeu donnée lissé	24
Figure 21 : Graphique de l'évaluation du modèle avec Noisy Label et le nouveau jeu donnée non lissé	25
Figure 22 : Graphique de l'évaluation du modèle avec Label Smoothing et Noisy Label sur le nouveau jeu donnée non lissé	26
Figure 23 : Graphique de l'évaluation du modèle avec Label Smoothing et Noisy Label sur le jeu donnée lissé	26

Chapitre 1 : Context

1 - LIMOS

Le Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes (LIMOS) est une Unité Mixte de Recherche (UMR 6158) en informatique, et plus généralement en Sciences et Technologies de l'Information et de la Communication (STIC).

Le LIMOS est principalement rattaché à l'Institut des Sciences de l'Information et de leurs Interactions (INS2I) du CNRS et de façon secondaire à l'Institut des Sciences de l'Ingénierie et des Systèmes (INSIS). Il a pour tutelles académiques l'Université Clermont Auvergne et l'Ecole Nationale Supérieure des Mines de Saint-Etienne (EMSE), et comme établissement partenaire l'école d'ingénieur SIGMA.

2 - ACCRIL

L'objectif du projet de recherche d'Aide à la Conception de Carte en Relief pour les Instructeurs en Locomotion (ACCRIL) est de concevoir une chaîne de traitement qui produise une carte en relief adaptée aux personnes en situation de déficience visuelle, en combinant à la fois des photos vues du ciel et des données géographiques telles que celles d'OpenStreetMap.

Ce projet est né d'une collaboration entre le LIMOS de l'Université Clermont Auvergne (UCA) et l'équipe GeoVIS de l'UMR LASTIG (IGN), et a dans un premier temps été principalement animé par la thèse de Gauthier Fillières-Riveau [1] en 2017.

Les échanges réguliers de l'équipe avec l'Association des Instructeurs de Locomotion pour personnes Déficiantes Visuelles (AILDV), ainsi qu'avec les instructeurs et instructrices en locomotion du Centre de Rééducation pour Déficiants Visuels (CRDV) permettent de constamment affiner et préciser le projet.

3 - Thèse de Gauthier Fillières-Riveau

Gauthier Fillières-Riveau a dans sa thèse réalisé les toutes premières explorations et expérimentations pour générer un premier jeu de données d'entraînement ainsi qu'un premier modèle d'apprentissage profond afin de faire de la segmentation sémantique sur des orthophotographies [1]. Il a étudié le principe d'enrichissement dans les images afin d'améliorer les performances du modèles ainsi que d'autres solutions pour évaluer, assembler et corriger les prédictions.

4 - Sujet

La segmentation sémantique, dans ce sujet, a pour but de trouver, identifier et délimiter les objets de voirie piétonne et automobiliste. Pour cela on se base sur des orthophotographies, qui sont des photos aériennes, afin d'avoir une vue d'ensemble sur les environnements à étudier.

Les recherches menées dans ce stage ont pris pour base ce premier travail réalisé par Gauthier Fillières-Riveau en amont [1] et vise à poursuivre ce qui a été commencé. La première piste de poursuite est la réalisation d'une plus grande base de données

d'apprentissage afin de repousser les problèmes de sur-apprentissage. Puis, dans un second temps, étudier et améliorer l'enrichissement qui devra améliorer les performances de segmentation. Et pour finir comprendre et modifier un modèle d'apprentissage profond afin de voir les capacités et les résultats qu'il est possible d'obtenir.

5 - Plan

Ce rapport commence par présenter le jeu de données, ses caractéristiques et sa conception. Puis il présente l'architecture en commençant par sa compréhension et puis sa structuration. Ensuite, il explique les stratégies d'évaluations mises en place pour mesurer les performances des modèles. Et pour finir, il présente les différentes expérimentations qui ont été mises en place avec chacune leurs différences et leurs résultats associés.

Chapitre 2 : Création du jeu de données

Dans les recherches menées précédemment [1], le jeu de données était basé sur 15 cartes de carrefour de Clermont-Ferrand en haute résolution réalisées par des instructeurs en locomotion. Ce sont des professionnels qui aident et accompagnent les déficients visuelles à appréhender et se déplacer dans les environnements quotidiens et dangereux comme notamment les carrefours. Ces cartes ont été subdivisés en vignette de 256 pixels de côté et leur nombre a ensuite été augmenté en apportant des rotations. Ainsi, le jeu de données obtenues comporte 375 images. La problématique ici, est d'obtenir un jeu de données avec un nombre plus conséquent d'images afin d'atteindre les meilleures performances que le modèle peut offrir.

1 - Description du jeu de données

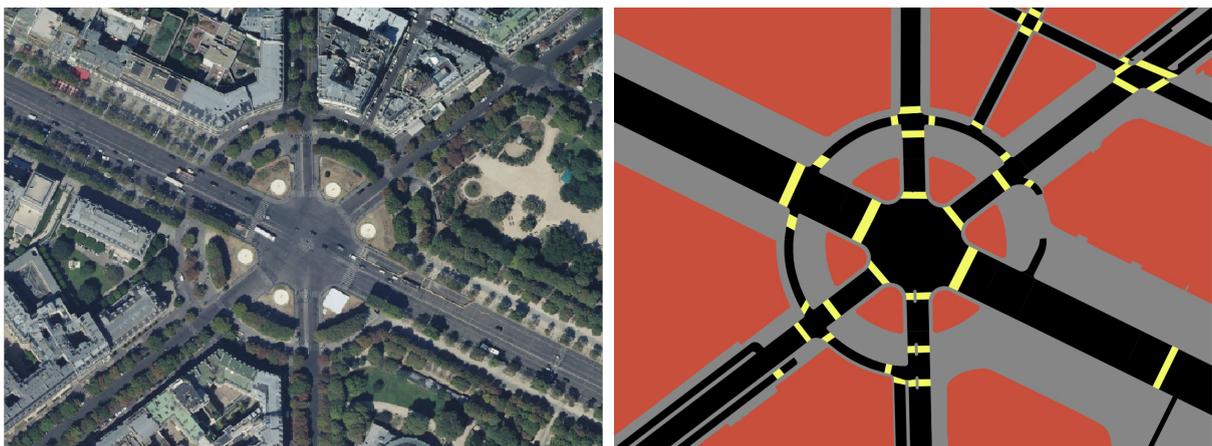


Figure (a)

Figure (b)

- (a) une orthophotographie du rond-point des Champs Elysées
(b) La vérité terrain de la segmentation sémantique du rond-point des Champs Elysées.

Le jeu de données nécessaire à l'entraînement est constitué de deux ensembles. Le premier est une représentation ortho-photographique (Figure a), communément appelée orthophoto, du terrain à étudier. Une orthophoto est une photographie prise

perpendiculairement par rapport au sol (cliché aérien) et où la géométrie a été redressé de façon à pouvoir la superposer sur une carte. Cet ensemble servira de base, pour le modèle, sur lequel il devra faire des prédictions.

Le second ensemble devra être une vérité terrain de la segmentation sémantique (Figure b) que devra générer le modèle. Il servira de base à comparer avec les prédictions afin de mesurer si ces dernières se rapprochent des résultats attendus.

Ainsi le jeu de données est constitué de couple d'image, une orthophoto avec sa segmentation sémantique associée.

2 - Sémantique

Dans la démarche du projet ACCRIL, les éléments recherchés sont ceux qui ont un intérêt dans la description des voiries piétonnes et véhiculaires mais aussi les environnements avec des caractéristiques particulières. Les principaux sont la chaussée, les trottoirs, les passages piétons et les îlots directionnels. Ces premiers objets sont les surfaces les plus importantes pour les instructeurs en locomotion car ce sont sur celles-ci que les déficients visuels vont apprendre à évoluer. Avec ceux-ci, il est nécessaire d'ajouter les informations des bâtiments, réseaux ferroviaires, fleuves qui peuvent être des obstacles, mais aussi des repères tant au contact qu'au son.

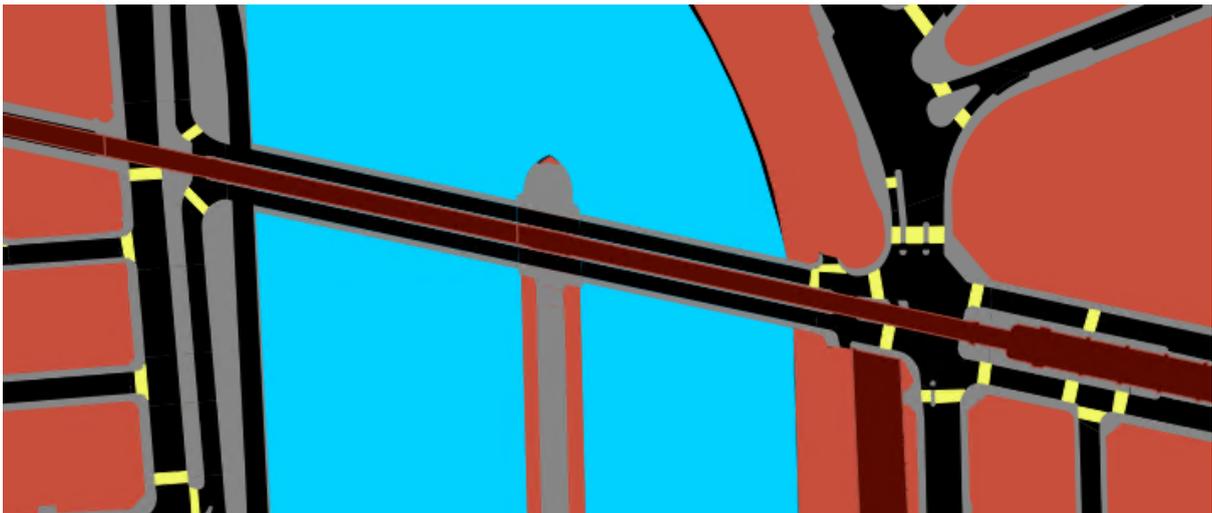


Figure 3 : Segmentation sémantique du pont de Bir-Hakeim avec la présence en bleu de la Seine, le réseau ferroviaire en bordeaux, chaussée en noir, trottoirs en gris et passage piéton en jaune.

Afin d'avoir une sémantique compréhensible facilement par des personnes extérieures, il a été choisi de décrire les éléments suivant avec des couleurs associées :

- La chaussée en **noir**
- Les trottoirs en **gris**
- Les passages piétons en **jaune**
- Les terre-plein et îlots directionnels en **vert foncé**
- Les bâtiments en **rouille**
- Le réseau ferroviaire en **rouge bordeaux**
- Les espaces vert en **vert claire**
- Les voies d'eau en **bleu**

Il a été remarqué par la suite que les modèles retournent de mauvais résultats au moment de la différenciation entre trottoir et îlots (chap 5 section 2), ainsi il a été décidé de les fusionner sous la même sémantique afin d'éviter ce conflit de segmentation et ainsi obtenir de meilleure performance dans les évaluations. Néanmoins, il est possible en un simple post-traitement de re-séparer cette classe en deux si cette information s'avérait utile.

3 - Réalisation de la vérité terrain segmentée

La plateforme opendata de Paris [13], met à disposition l'ensemble des jeux de données d'informations géographiques publiés par les services de la ville et ses partenaires sous licence ODbL. Ces données couvrent une très grande partie du territoire parisien et sont le sujet d'étude pour créer un jeu de données d'entraînement. Ce qui a motivé le choix pour ces jeux de données est qu'il était très complet, recouvrait une grande superficie et avait un niveau de précision qualitatif.

3.1 Choix des couches

Parmi tous les groupes de données proposés sur la plate-forme, seules les suivantes sont retenues : *Emprise îlots privés, Emprise trottoirs, Voies en escalier, Passages piétons, Aires mixtes véhicules et piétons, Terre-pleins, îlots directionnels, Chaussées, Rampes de souterrain, Emprises espaces verts, Espaces verts 'frais', Emprises ferroviaires, Emprises bâties et non bâties, Pistes cyclable et couloirs de bus, Voies d'eau, Rampes de viaducs.* Cet ensemble de groupe est suffisant pour couvrir la région parisienne avec les informations désirées. Ces jeux de données ont été assemblés sur un logiciel de système d'information géographique (SIG) nommé QGIS. Représenté sous la forme de couche, il est alors possible de les agencer, les superposer et les découper afin de couvrir la plus grande superficie de l'espace parisien.

3.2 Ajustements et corrections

Ces données ne sont néanmoins pas parfaites et certaines régions ne sont alors pas couvertes. Certains espaces manquants ont été corrigés à la main comme des informations de chaussées ou encore des emprises de passages piétons et d'îlots directionnels. Par exemple, afin de détecter les passages piétons incohérents, c'est à dire ceux qui ne joignent pas deux voiries piétonnes (trottoirs ou îlots), une opération de tampon (augmentation de l'air de l'objet) a été appliquée afin qu'ils chevauchent les objets adjacents, de cette manière il est possible de trouver les éléments qui ne sont pas en contact sur leur deux extrémités et il est alors possible de retrouver les objets et de corriger l'imperfection.

La principale contrainte dans ces jeux de données est les informations apportées avec les polygones d'emprises géographiques. Le manque d'information a réduit l'espace utilisable car il n'était pas possible de corriger automatiquement ces erreurs de représentations. Parmi ces erreurs on retrouve notamment les superpositions de niveau sur la chaussée et les trottoirs mais aussi les voies ferroviaires (ex : pont, passage souterrain). De par l'absence d'information précise sur le niveau des polygones d'emprise par rapport aux autres, il n'était pas possible de corriger ces incohérences. Leur nombre étant conséquent, il a été choisi de réduire la superficie utilisable. Ces incohérences apparaissent fréquemment

au niveau de l'autoroute périphérique de la ville, où par ailleurs l'information se dégrade et devient de plus en plus manquante car on arrive à l'extrémité de la superficie représentée par les jeux de données.

De cette manière, il a été possible de construire une représentation segmentée, sur une vaste superficie, plutôt fidèle, de la vérité terrain en ce qui concerne les éléments de voirie comme les espaces piétons, les espaces de chaussée, les espaces mixtes, les voies d'eau et les infrastructures de type bâtiment, espaces verts et ferroviaires.

4 - Orthophotographie

La base de données d'orthophotographies est issue de la base de données topologique de l'IGN qui est en libre accès sur leur site internet [15]. La région de Paris est couverte dans son entièreté avec une date de capture datant de Mars 2020 et d'une précision de 20cm par pixels. Toutes les illustrations montrées dans ce rapport et les jeux de données sont issues de cette base de données.

5 - Enrichissement avec OpenStreetMap

D'après les études menées précédemment [1], l'utilisation d'enrichissement sur les orthophotographies est bénéfique pour les performances du modèle. Pour cela, il a été choisi d'utiliser des informations issues d'OpenStreetMap (OSM) [14]. Tous les enrichissements sont issus d'OpenStreetMap ainsi que ceux montrés dans les illustrations.

Ce choix a été motivé par plusieurs facteurs.

Pour commencer, c'est une base de données qui recouvre une très grande partie du globe avec des informations très précises sur les différentes voies routières et piétonnes existantes ainsi que bien d'autres renseignements. Ainsi il sera possible d'apporter cet enrichissement ailleurs que sur la région parisienne.

Puis aussi car c'est une base de données en libre accès et qui est maintenue par une communauté active et dynamique.

Cependant elle a aussi des inconvénients provenant du principe de libre contribution des utilisateurs qui n'utilisent pas tout le temps la même nomenclature pour décrire les éléments qu'ils ajoutent. Certains tags (nom de champ dans la base de données) sont tout de même largement utilisés par la communauté ce qui facilite la lecture de cette dernière. En effet, OSM est construit comme une base de données sous forme de tableau avec des tags comme identifiant. Néanmoins un travail de mise au point de filtre a été réalisé afin de n'extraire que les informations utiles mais aussi afin de pallier ces variantes de nomenclature et ainsi couvrir un maximum de cas.

Comme couleur d'affichage il a été choisi d'utiliser un rouge saturé. Cette couleur n'étant que très rarement présente dans les milieux urbain ou naturel, cela permet au modèle de ne pas confondre l'enrichissement avec l'information de l'orthophoto.

Dans cette optique d'enrichissement, plusieurs choix ont été mis à l'étude.

5.1 - Filaire de la chaussée

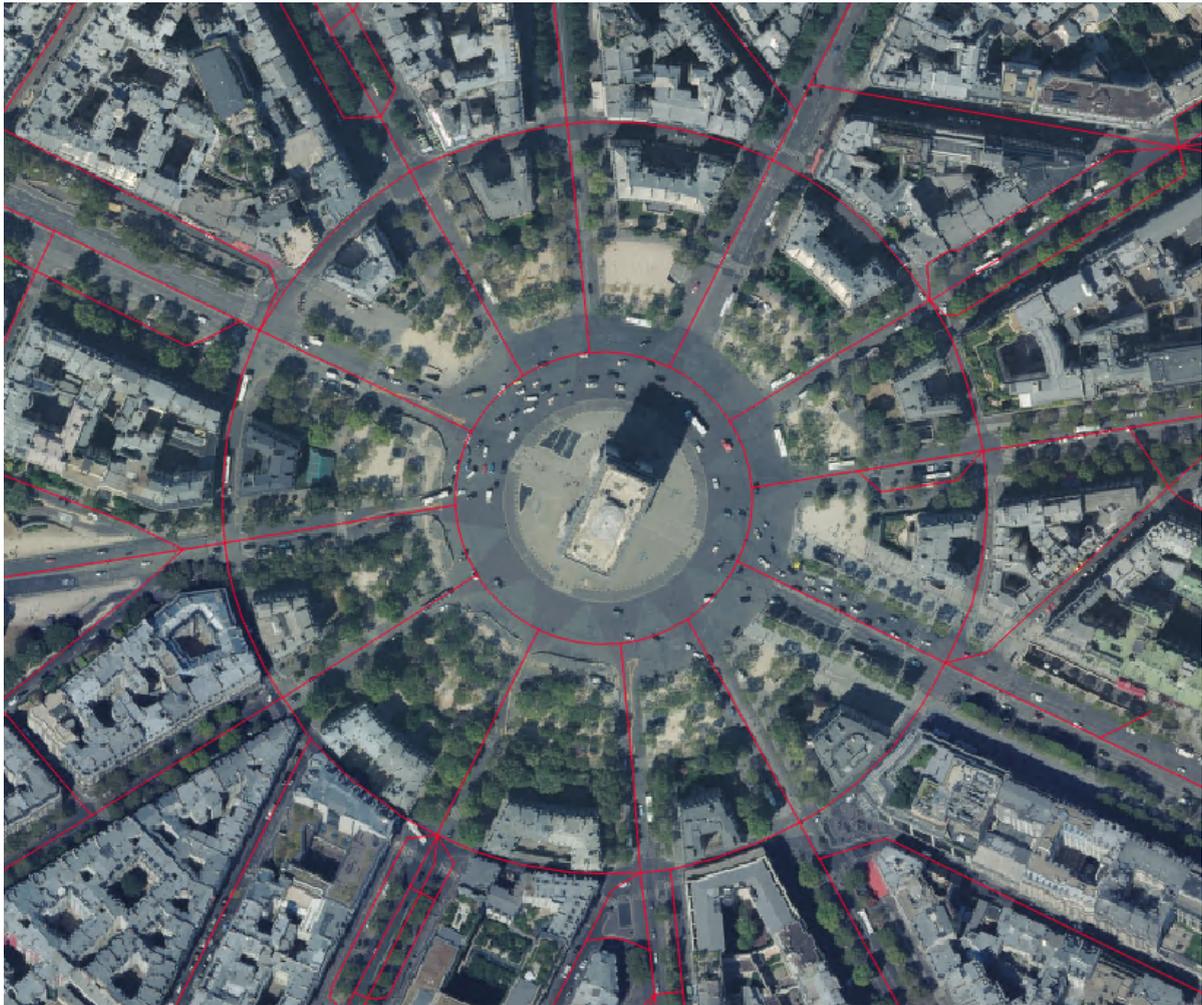


Figure 4 : Représentation de l'enrichissement du filaire de la chaussée sur le rond-point des champs de l'Elysée

Le premier, qui a commencé à être évalué dans l'étude précédente [1], est l'ajout d'un tracé sous une forme filaire (fig 4) pour représenter la position des différentes chaussées. Dans l'étude précédente il a été montré que cet enrichissement permettait au modèle d'être plus robuste à la présence d'obstruction visuelle sur le sol par la végétation (fig 15) ou la présence de bâtiment incliné dû à l'orientation des prises des images. Dans les résultats obtenus on remarque nettement la présence des chaussées dans des rues fortes étroites et presque inaccessibles visuellement par la végétation. Cela permet aussi au modèle de décider si les végétations en bord de chaussées peuvent être considérées comme des espaces verts ou non.

5.2 - Point des passages piétons

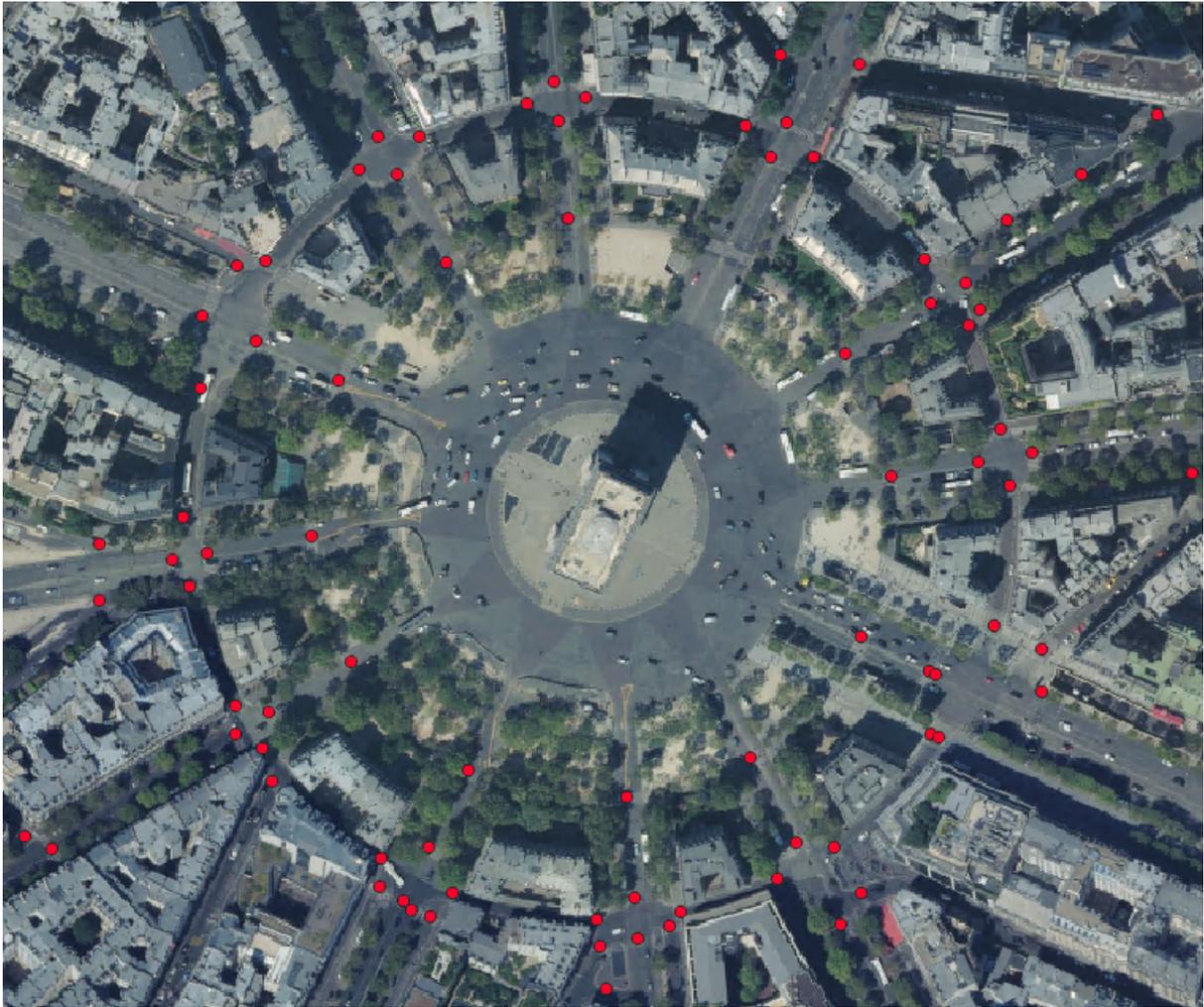


Figure 5 : Représentation de l'enrichissement des points pour passages piétons sur le rond-point des champs de l'Elysée.

Afin que le modèle reconnaisse distinctement les différents enrichissements, il a été choisi de leur appliquer des formes différentes. Ainsi un second enrichissement a été étudié et ajouté sous forme de point (fig 5), les passages piétons. Il a été remarqué que lorsque ce dernier n'est pas ou peu visible, le modèle les détecte difficilement. Ainsi avec cet ajout, le modèle parvient à interpréter les passages piétons même dans des situations où ils sont peu ou pas visibles.

5.3 - Contours du bâti

Le dernier enrichissement qu'il était question d'ajouter est celui de la limitation au sol du bâti (fig 6). Il a été remarqué sur les prédictions du modèle que la présence de végétation en bordure de route et de bâti amène une certaine confusion sur la délimitation entre le bâti et le trottoir. Ce qui donne un aspect ondulé et granuleux à cette démarcation. Cette information peut être récupérée dans OSM mais aussi dans les données du Cadastre. Cet enrichissement est encore à explorer.

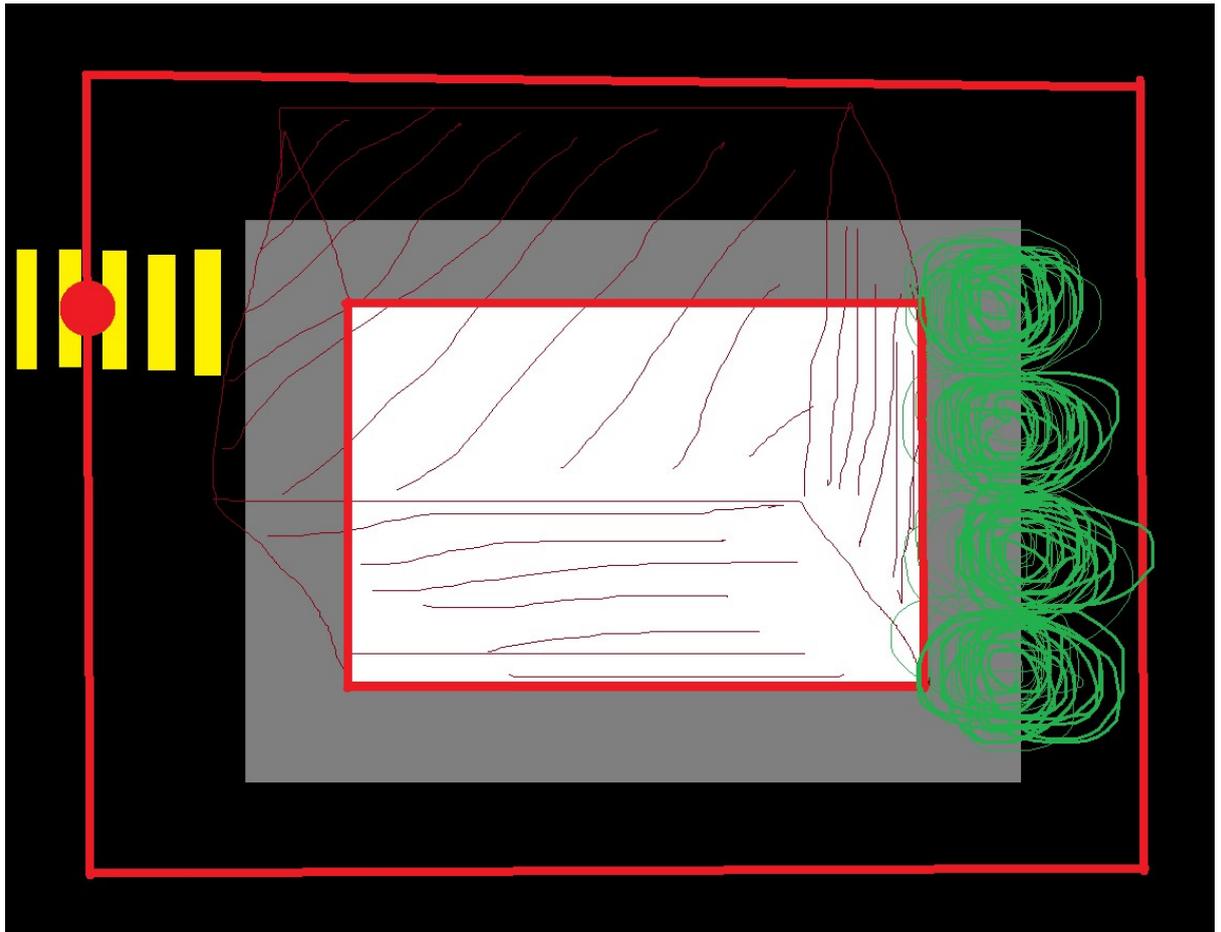


Figure 6 : Schéma de la segmentation avec les enrichissements de chaussée (filaire extérieur), passage piéton (point) et bâti (filaire intérieur)

5.4 - Affinage

L'intégration de l'enrichissement demande cependant quelques ajustements malgré les filtres appliqués en amont. Afin d'éviter des incohérences et confusion lors de l'entraînement il faut que l'information qui indique l'enrichissement soit présente dans la segmentation de la vérité terrain. En effet, certaines informations sont manquantes ou de trop. Pour le filaire de la chaussée, certains passages privés ou entrée de garage sont présents mais l'information d'emprise géographique ne l'est pas. Il était donc nécessaire de détecter ces objets afin de les supprimer. Pour les points de passage piéton il a aussi été remarqué que certains sont très éloignés de la vérité terrain. Pour cela, à partir d'une certaine distance de l'emprise, il a été décidé de les supprimer. Pour les points manquants, rien n'a été fait, de même que pour les filaires manquants.

5.5 - Inconvénients

Néanmoins, l'enrichissement n'apporte pas que des avantages. Dans des situations où il serait manquant, un modèle entraîné avec enrichissement deviendra dépendant de ce dernier. Cette situation a été constatée avec le filaire des chaussées (fig 16), où en l'absence de ce dernier, le modèle prédit de mauvais résultats. C'est à dire que même avec

une chaussée bien visible, il va tout de même tenter de classer la chaussée comme autre chose. En revanche, pour les points des passages piétons, le modèle parvient à utiliser l'enrichissement sans pour autant en être dépendant. C'est-à-dire qu'il est en mesure de prédire un passage piéton visible sans son enrichissement (fig 16). Mais aussi, avec un enrichissement, quand il n'est pas visible. La raison de cette dépendance et de cette différence entre les deux enrichissements est encore à déterminer.

5.6 - Intégration

Pour ajouter cet enrichissement aux images, deux choix ont été envisagés.

Le premier et le plus facile à implémenter est d'ajouter ces éléments directement dans l'image avec des épaisseurs faibles afin de ne pas cacher l'information et avec des saturations de couleurs non naturelles dans les images afin de bien dissocier l'image de son enrichissement.

Le second est d'ajouter un, ou plusieurs, canaux à l'image. Il est envisageable d'ajouter un canal aux images et d'y incruster les enrichissements sous forme de nuances de gris. Dans des cas où les enrichissements seraient nombreux, il serait aussi possible d'ajouter plusieurs canaux, éventuellement un canal par type d'enrichissement afin qu'ils n'y ait pas de perte d'information en cas de superposition. Cette approche est à étudier pour de futurs travaux.

6 - Génération du jeu de données

Dans l'étude précédente [1], il a été choisi d'utiliser des images de 256 pixels de côté afin de répondre à la spécificité d'entrée du modèle et cette décision a été conservée pour le jeu de données. Cela correspond à une représentation au sol de 51 mètres par 51 mètres environ (précision de 20 cm par pixel).

La segmentation réalisée sur QGIS couvre un vaste espace de la région parisienne à l'exception des quelques zones volontairement exclues pour manque de données ou à cause de droit de vision sur certains lieux. Pour découper cet espace, des coordonnées ont été générées aléatoirement (Figure 7), avec des restrictions de proximité entre eux, le long des axes routiers représentés par nos lignes issues d'OSM. Pour chaque point, un ensemble de translations et de rotations sont appliqués afin d'obtenir un jeu d'images différentes pour une région restreinte. Cette opération a été réalisée avec l'outil de rendu, présent dans QGIS. Ainsi il a été possible de générer environ 180 mille couples d'images. Un couple étant la version segmentée et l'orthophoto enrichie.

Des opérations de post-traitement ont néanmoins été nécessaires pour corriger et supprimer certains défauts.

Dans un premier temps un tri a été effectué afin de supprimer toute image possédant des régions n'étant pas couvertes par une sémantique (le blanc représente l'absence de sémantique dans la segmentation).

Ensuite il a été remarqué que lors du rendu un effet de lissage est appliqué par QGIS, rendant l'image propre pour des yeux humains mais impliquant l'ajout de nombreuses couleurs intermédiaires au niveau des bordures entre segmentation. Pour cela, une nouvelle correction est appliquée afin d'affilier chaque pixel lissé à sa sémantique originelle (fig 18).



Figure 7 : Résultat de la génération aléatoire des points (roses) le long des axes routiers. Chaque point permet de réaliser 50 vignettes en ajoutant un offset et une rotation depuis les coordonnées initiales.

Chapitre 3 : L'architecture

1 - Description d'un GAN

Les structures d'apprentissage profond de la famille des Generative Antagonist Network (GAN) [2] sont constituées de deux réseaux de neurones. Un générateur et un discriminateur. Comme illustré en figure 8, le générateur doit, à partir d'un vecteur de bruit aléatoire, créer un contenu en sortie. Le discriminateur doit, à partir d'une entrée, déterminer si cette dernière peut être considérée comme une véritable information, ou une fausse, générée par le générateur. De cette manière, ces deux réseaux sont alors entraînés simultanément et en compétition. D'un côté, le générateur doit produire un contenu qui puisse dupé le discriminateur. De l'autre côté, le discriminateur doit parvenir à démasquer les entrées produites par le générateur. De cette manière, le générateur ne fait pas qu'apprendre les spécificités du jeu de données mais aussi de produire un contenu cohérent et capable de ressembler à ce qu'y est attendu.

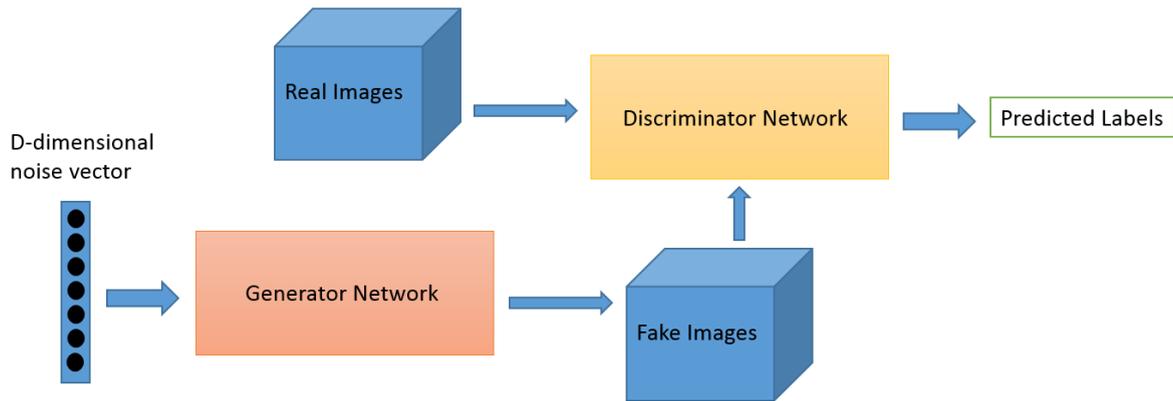


Figure 8 : Schéma représentant le fonctionnement d'un GAN Crédit : O'Reilly

2 - Description d'un cGAN

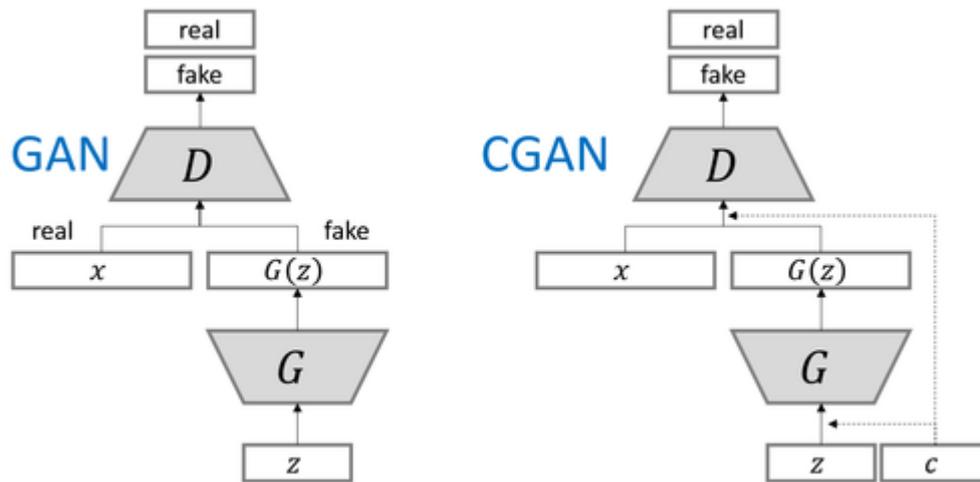


Figure 9 : Schéma représentant le fonctionnement d'un GAN (à gauche) et d'un CGAN (à droite)
Crédit : LoGAN - Generating Logos with Generative Adversarial Neural Network Conditioned on color

Un Conditional Generative Antagonist Network (CGAN) (fig 9) est une variante [3] de la famille des GAN. L'architecture est la même à l'exception des données en entrée du générateur et du discriminateur. Dans un CGAN, on concatène une information supplémentaire en entrée des modèles. Cette information est une base sur laquelle s'appuyer pour générer et comparer les résultats. De cette manière, le générateur va s'appuyer sur cette nouvelle donnée pour produire un résultat et le discriminateur peut comparer les résultats tout en connaissant la base sur laquelle ils ont été produits. Ainsi cela rend la tâche plus compliquée pour le générateur, car ce dernier devra produire des résultats encore plus fidèles à la vérité terrain pour duper le discriminateur.

3 - Description de pix2pix

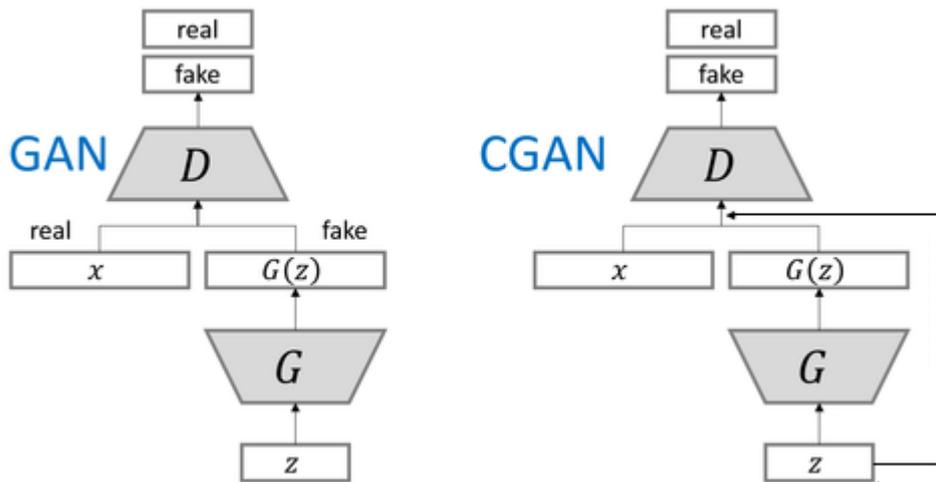


Figure 9 : Schéma représentant l'adaptation CGAN pour pix2pix. Pix2Pix, utilise l'information de départ du générateur comme condition dans le discriminateur.

Crédit : LoGAN - Generating Logos with Generative Adversarial Neural Network Conditioned on color

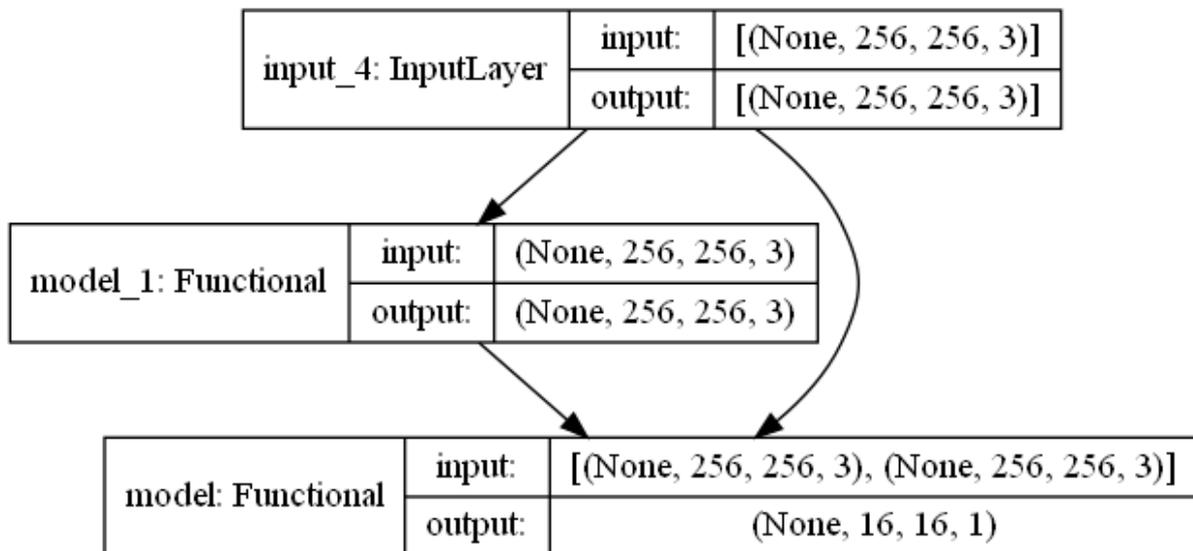


Figure 10: Schéma représentant la structure du cgan de pix2pix, *InputLayer* étant l'image en entrée, *model_1* est le générateur et *model* le discriminateur.

Les cGAN ont déjà été utilisés pour des tâches de segmentation sémantique comme en [4]. Pour cette étude, il a été choisi de travailler sur l'architecture du cGAN de pix2pix [5] qui est un modèle de transfert d'image à image. Ce qui en fait un bon candidat pour un travail de segmentation sémantique. Pix2pix a déjà fait partie de l'étude en amont [1] et a montré de bon résultat. L'objectif ici est de voir ce qu'il est possible d'en faire avec le nouveau jeu de données et s'il est possible de l'améliorer.

La famille des cGAN utilise habituellement un vecteur de bruit (fig 8) en entrée afin d'introduire de l'aléatoire dans les résultats. D'après le papier de pix2pix [5], les modèles apprennent à ignorer le bruit. Alors dans leurs architectures ils ont ajouté un bruit sous la forme de Dropout dans plusieurs couches du générateur à l'entraînement et aux tests. Cependant il a été constaté une faible stochasticité des résultats, ce qui dans cette étude est un point positif, car l'aléatoire aidera le modèle durant l'entraînement pour explorer les

caractéristiques des images sans pour autant devenir trop fort pour produire des résultats trop incohérents.

Pour les fonctions de perte, l'architecture de Pix2pix utilise deux fonctions : binary cross entropy et le mean absolute error.

Les images du jeu de données doivent être comprises entre -1 et 1 à cause de la fonction d'activation en sortie du générateur qui est une tangente hyperbolique retournant des résultats compris entre -1 et 1. Le discriminateur devant comparer deux images, elles doivent être sur le même intervalle.

Le choix des outils d'implémentation ont été portés sur Python avec le framework Keras utilisant Tensorflow [16].

3.1 Générateur

Dans cette architecture, le générateur est un U-Net (fig 11). C'est un modèle de la forme encodeur-décodeur, mais avec des passerelles (skips) entre les couches d'encodage et de décodage. De cette manière, lors du décodage, le modèle ne perd pas l'information générale de l'entrée qui a été réduite à un vecteur de caractéristiques lors de l'encodage. Il a été montré que les U-Net obtiennent de bien meilleures performances que ses CNN prédécesseurs

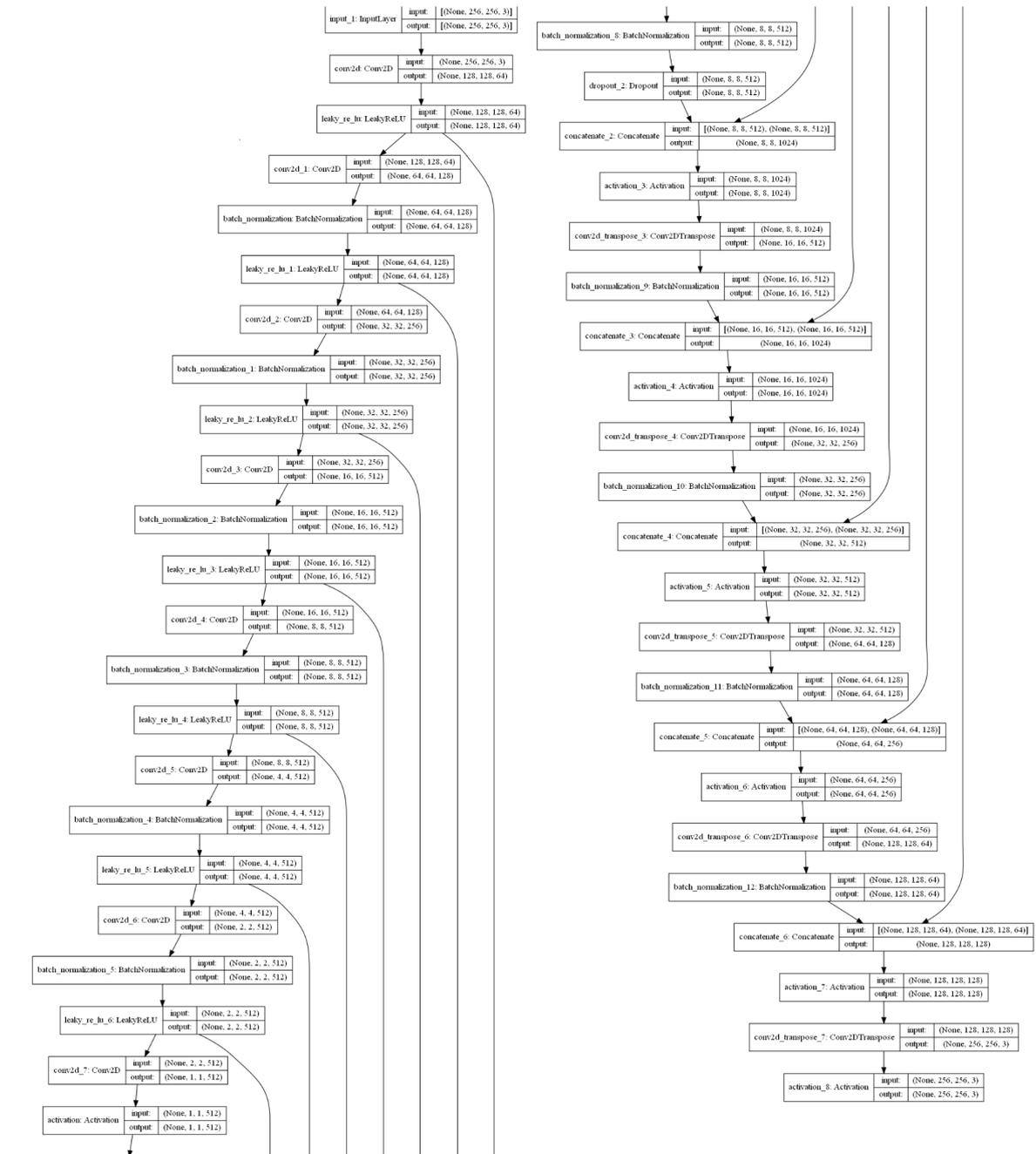


Figure 11 : Schéma des couches du générateur
 Encodeur : C64-C128-C256-C512-C512-C512-C512
 Decoder : CD512-CD1024-CD1024-C1024-C1024-C512-C256-C128
 C = Convolution-BatchNorm-LeakyReLU
 CD = Convolution-BatchNormDropout-LeakyReLU

3.2 Discriminateur

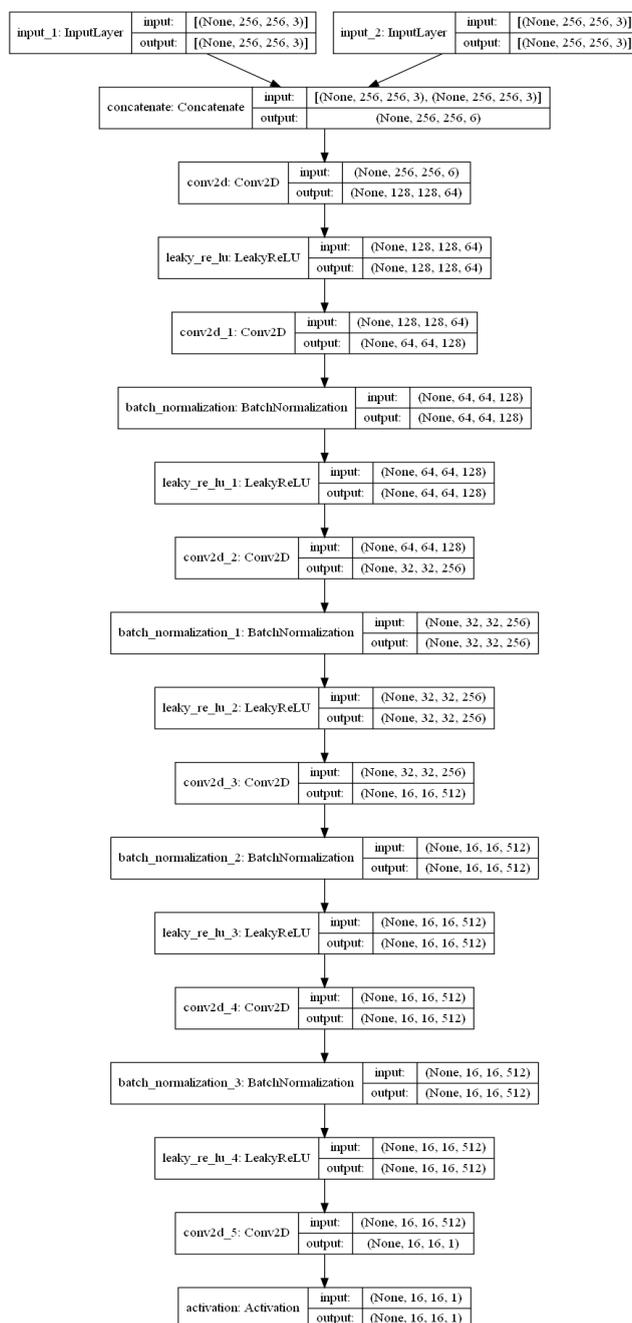


Figure 12 : Schéma représentant les couches du discriminateur
C64-C128-C256-C512
C = Convolution-BatchNorm-LeakyReLU

Le Discriminateur (fig 12) quand à lui est un classifieur convolutif “PatchGan” [5], qui au lieu de déterminer si une image, dans son entièreté, est considéré comme réelle ou pas, il va travailler sur des patch de l’image, minimisant l’impacte de ces décision à l’échelle des patch.

Chapitre 4: Evaluation

1 - Stratégie

Pour chaque entraînement, des sauvegardes du générateur sont réalisées tous les 10 epochs, soit 20 sauvegardes pour 200 epochs.

Durant l'entraînement des échantillons d'images avec leur prédictions et vérité terrain sont enregistrés en même temps que le modèle afin d'avoir un aperçu de l'évolution sur le jeu d'entraînement dans le temps [11].

2 - Evaluation visuelle

Les premières expérimentations ont été évaluées sur un petit jeu de données (101 images) sélectionné à la main. Ce jeu de test comprenait des images avec des situations variées et challengeantes (ombres portées, travaux, enrichissement manquant, carrefour complexe, couleur nombreuse, parking, piscine...).

Ces premières évaluations ont été réalisées en visuelle afin d'avoir un aperçu humain, dans un premier temps, pour mesurer la précision des modèles.

Cependant une problématique s'est rapidement fait ressentir. Il devenait difficile de déterminer si une version d'un modèle était mieux qu'une autre quand les différences s'amointrissent. Ainsi un algorithme de mesure par métrique a été mis en place.

3 - Métriques

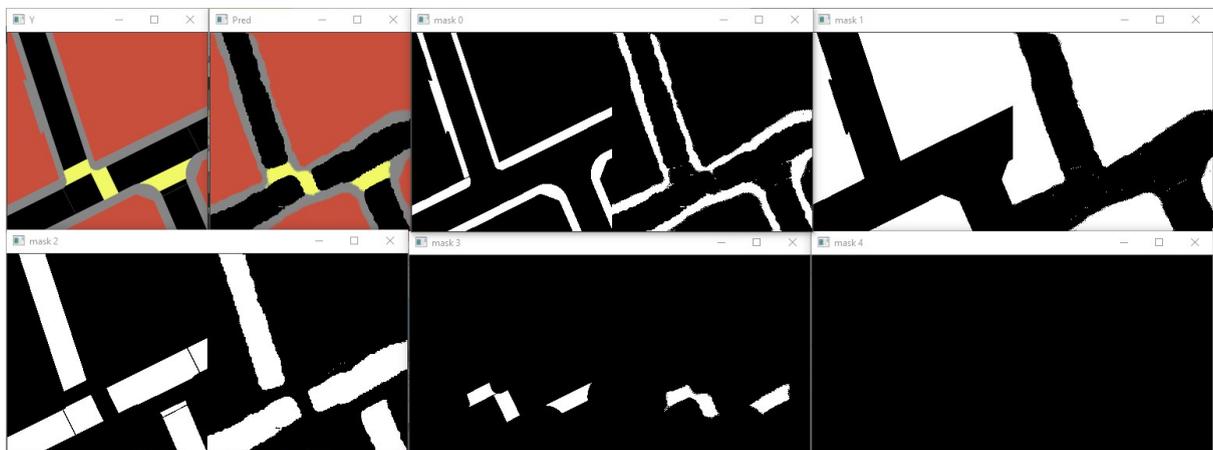


Figure 13 : Analyse de chaque classe indépendamment entre la vérité terrain à gauche et la prédiction à droite

Pour une évaluation sur un travail de segmentation sémantique plusieurs méthodes sont largement utilisées. Dans cette étude, trois métriques ont été testées [12] : la moyenne des intersections sur les unions (IoU) [12], la précision par pixel et la précision par classe. Il a été remarqué que les résultats entre elles étaient similaires voire presque identiques. Ainsi seule IoU a été retenue.

Cependant rien ne pourra remplacer la perception et le jugement humain. L'évaluation des GANs a toujours été une question récurrente et il n'y pas vraiment de solution miracle.

L'architecture utilisée produit des résultats sous forme d'image RGB avec des valeurs comprises entre -1 et 1. Il a été constaté que les outils de métriques dans Tensorflow ne parvenaient pas à se lancer avec cet intervalle des valeurs. Deux solutions étaient envisageables. La première et celle qui a été choisie est de réaliser la fonction en question. Elle isole chaque classe présente dans l'image et applique les opérations pour réaliser l'évaluation. La seconde option était de changer le format de sortie en "one hot encoding" pour permettre aux fonctions de Tensorflow de fonctionner, mais elle n'a pas été retenue car la structure du réseau avec le discriminateur ne le permet pas.

Les résultats sont représentés sous forme de graphe permettant ainsi d'avoir un aperçu sur les différentes sauvegardes. Pour la métrique de la moyenne de l'intersection sur l'union, on récupère la moyenne sur toute les images, la médiane, le minimum et le maximum.

Le jeu de données comprenant environ 180 milles images et seulement 100 milles étant utilisé pour l'entraînement, il a été décidé d'en utiliser 25 milles, non parcourues durant l'entraînement, pour construire le jeu de tests. Ce choix a été motivé principalement sur une contrainte de temps afin d'éviter de trop long traitement.

Chapitre 5 : Expérimentation

1 - Premiers tests

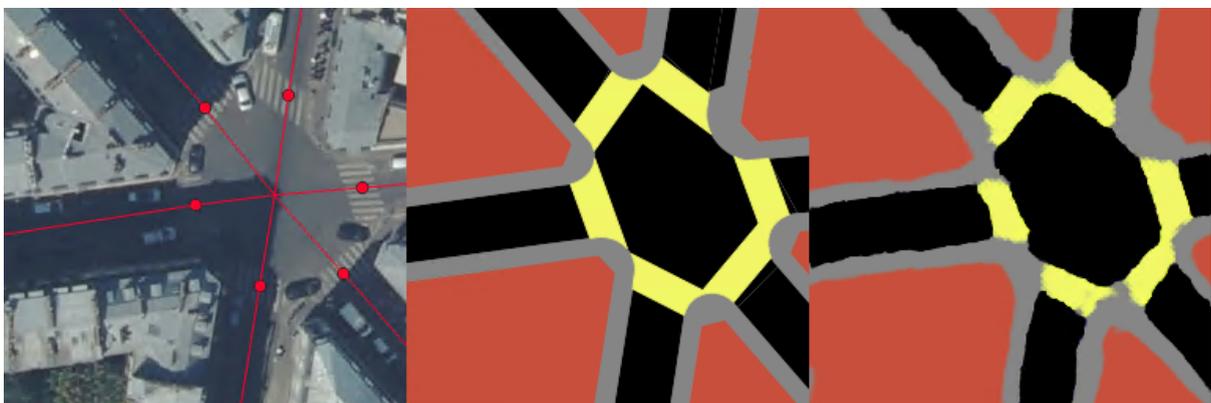
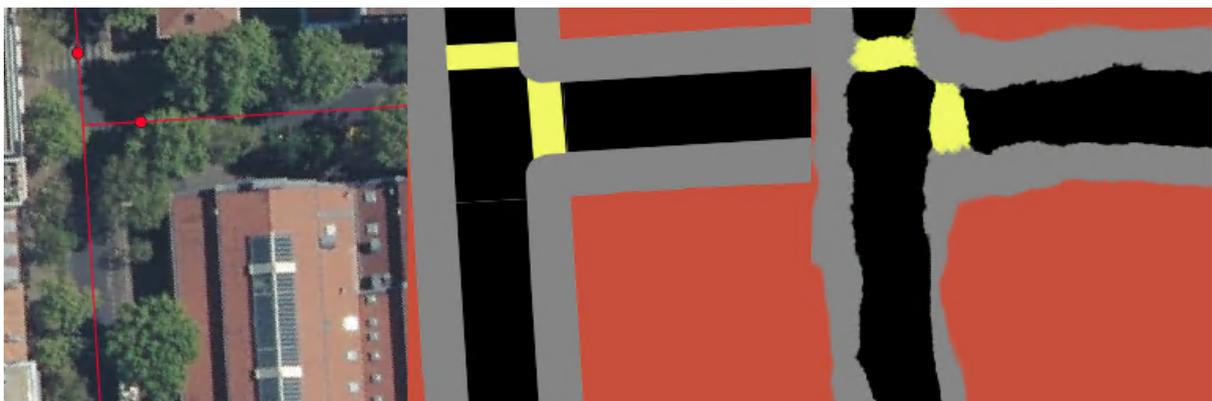


Figure 14 : Échantillon du premier modèle réalisé basé sur pix2pix sans amélioration



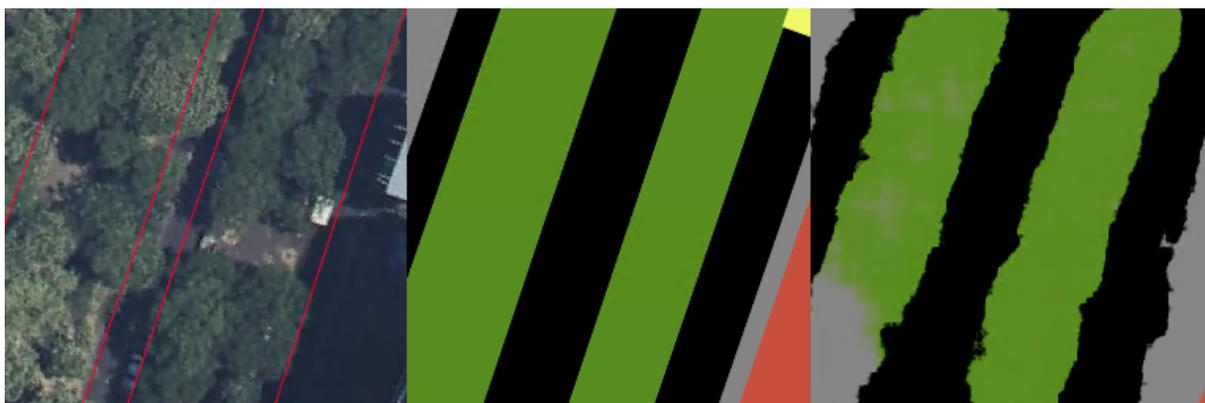


Figure 15 : Échantillons du premier modèle basé sur pix2pix sans amélioration, dans une situation d'obstruction de la vision par la végétation

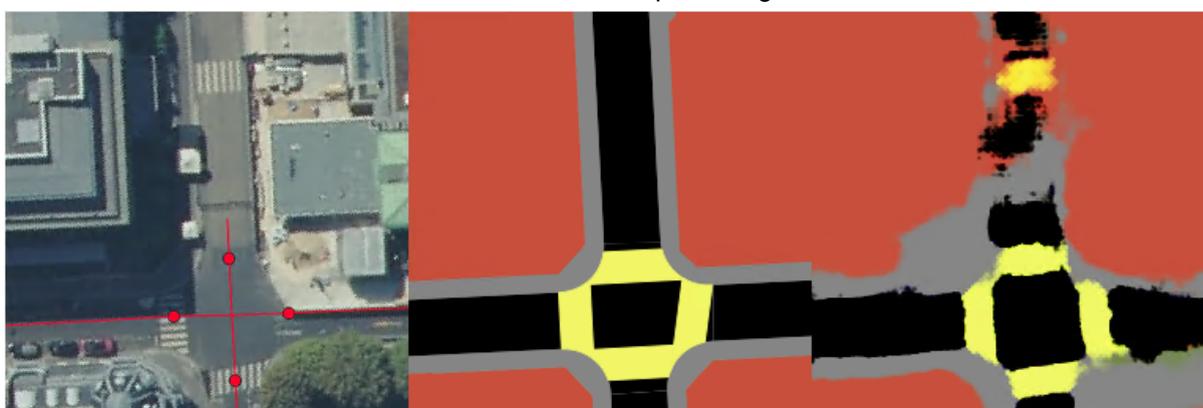


Figure 16 : Échantillon du premier modèle basé sur pix2pix sans amélioration, dans une situation d'absence de donnée d'enrichissement

Les premières expérimentations furent exclusivement évaluées à l'œil humain. Les premières constatations ont montré des résultats surprenants. En effet, avec seulement 10 epochs, le modèle parvenait à segmenter plus ou moins correctement les différents objets (figure X). Les limites entre les objets restaient très imprécises avec des ondulations notamment dû à la présence de végétation le long des voies routières (figure X). Mais la précision de la segmentation restait très satisfaisante. Il a cependant été remarqué la présence d'artéfact plus ou moins important qui peuvent éventuellement être retirés avec du traitement d'image en post-traitement.

De plus, la détection des terre-pleins, quand ces derniers sont étroits, semble être compliquée pour le modèle. Il parvient parfois à les détecter mais rarement entièrement et ne semble pas s'améliorer avec le temps voir même préfère les supprimer de sa segmentation.

Lors de ces premiers tests, il a été constaté que l'enrichissement apporte un vrai avantage pour le modèle car ce dernier parvient tout à fait à l'interpréter. Les petites rues de Paris où la visibilité est complètement obstruée par la végétation sont bien reconnues par le modèle (figure X).

Cependant avec ce petit jeu de test challengeant il a aussi été remarqué que sans enrichissement le modèle ne parvenait pas entièrement à interpréter les différents objets associés. Les chaussées n'étant pas enrichi provoque des conflits dans la segmentation (fig 16) même quand ce dernier est très bien visible (figure X). Ce qui n'est pas le cas en revanche pour l'enrichissement des passages piétons où ces derniers semblent tout y gagner. Ils sont reconnus avec et sans enrichissement (figure X). La raison de cette différence reste une question ouverte.

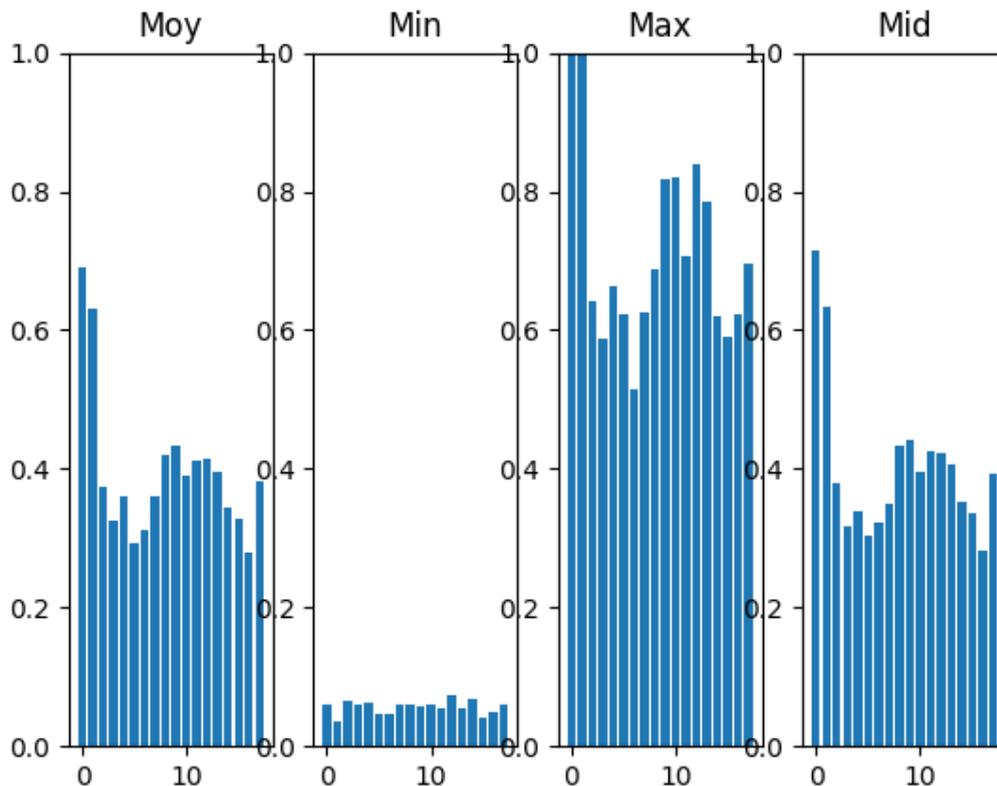


Figure 17 : Tableau de l'évaluation du premier modèle

Les premiers résultats d'évaluations par métrique (fig 17) ont révélé de l'instabilité et du sur-apprentissage lors de l'entraînement. L'instabilité se matérialise par la chute drastique des performances. Pour le sur-apprentissage, on peut considérer que la baisse globale au fil de l'apprentissage est un signe de ce dernier.

2 - Amélioration du jeu de données

Il a été constaté que le modèle ne retournait pas des résultats avec un nombre de couleurs égal au nombre de classes présentes dans l'image. Cette information a mis en évidence que le jeu de données n'était pas du format voulu. En effet, un lissage sur les contours des segmentations venait ajouter des couleurs intermédiaires.

Ainsi, une nouvelle version du jeu de données sans ce lissage a été réalisée (fig 18).

Deux évaluations ont été réalisées avec des modèles qui ont appris sur le jeu de données avec lissage et les autres sur la version sans lissage.

Cependant il a été constaté qu'avec un nouveau jeu de données, la présence d'artéfacts dans les prédictions était plus importante. Ces derniers devaient être moins visibles avec le lissage des couleurs.

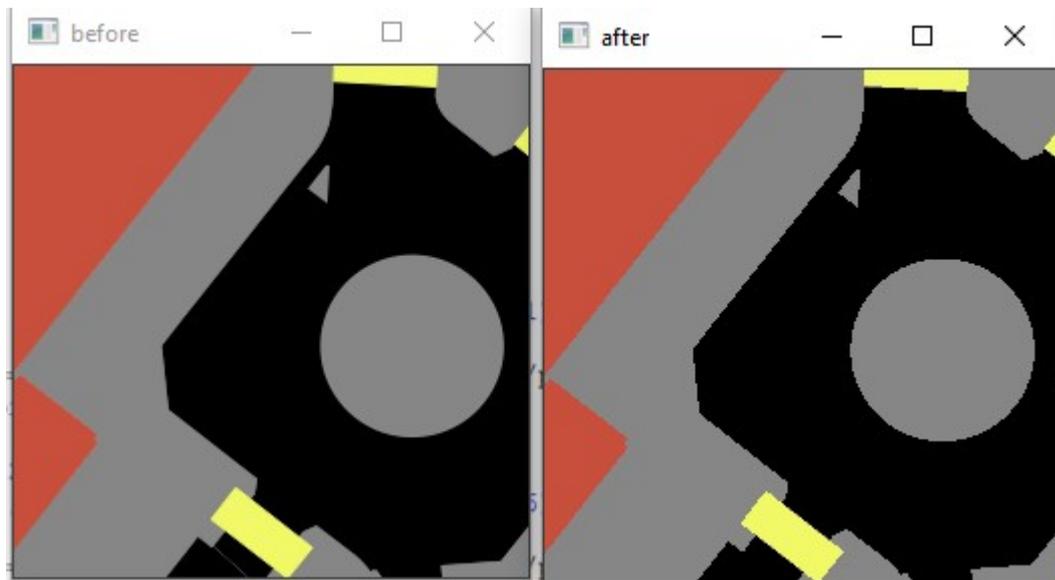


Figure 18 : Échantillon du jeu de données avec lissage à gauche et sans lissage à droite.

De plus, il a été décidé de fusionner la classe des trottoirs avec celle des terre-pleins et îlots directionnels. Ces derniers causent de nombreux conflits lors de la segmentation et réduisent les statistiques lors de l'évaluation.

Ce changement dans le jeu de données a apporté un gain dans les évaluations.

Par ailleurs, une nomenclature a été établie pour chaque image. On y associe dans son nom ses coordonnées géographiques ainsi que la rotation apportée. De cette manière il sera possible de reproduire le même jeu de données mais avec des propriétés différentes comme de l'enrichissement ou un changement de sémantique etc...

3 - Batch size

L'entraînement d'un GAN devient compliqué de par sa composition [8]. Comme il est formé de deux réseaux qui doivent apprendre simultanément en concurrence, il est important de maintenir un batch size faible afin d'éviter des instabilités. C'est à dire qu'avec une batch size élevée, il peut se produire un problème qui bloquera l'apprentissage, si un des deux modèles apprend trop d'un coup par rapport à l'autre il se peut qu'il se bloque dans cette situation et plus rien n'évoluera ensuite. C'est pour cela qu'il est important de maintenir un batch size faible.

Lors du premier entraînement, le batchsize avait été fixé à 1. Mais étant donné que le jeu

d'entraînement est composé de 100 milles images les temps d'entraînement s'envolent. Ainsi le choix a été fait de tester un batch size supérieur : 4. Les résultats sont sans appel. Ainsi, les temps passent de 32h d'entraînement par sauvegarde (10 epoch) à 8h. Ce qui fait un total de 7 jours d'apprentissage pour arriver au 200 epochs. L'entraînement a été réalisé sur une Geforce RTX 3090.

La qualité des prédictions n'est pas affectée mais le temps d'apprentissage est divisé par 4. D'un autre point de vue, le Label Smoothing et le Noisy Label sont des procédés qui aident à la stabilisation de l'apprentissage des GANs. Il serait peut être envisageable d'augmenter encore le batch size, mais c'est une question qui a été laissé ouverte.

4 - Label Smoothing

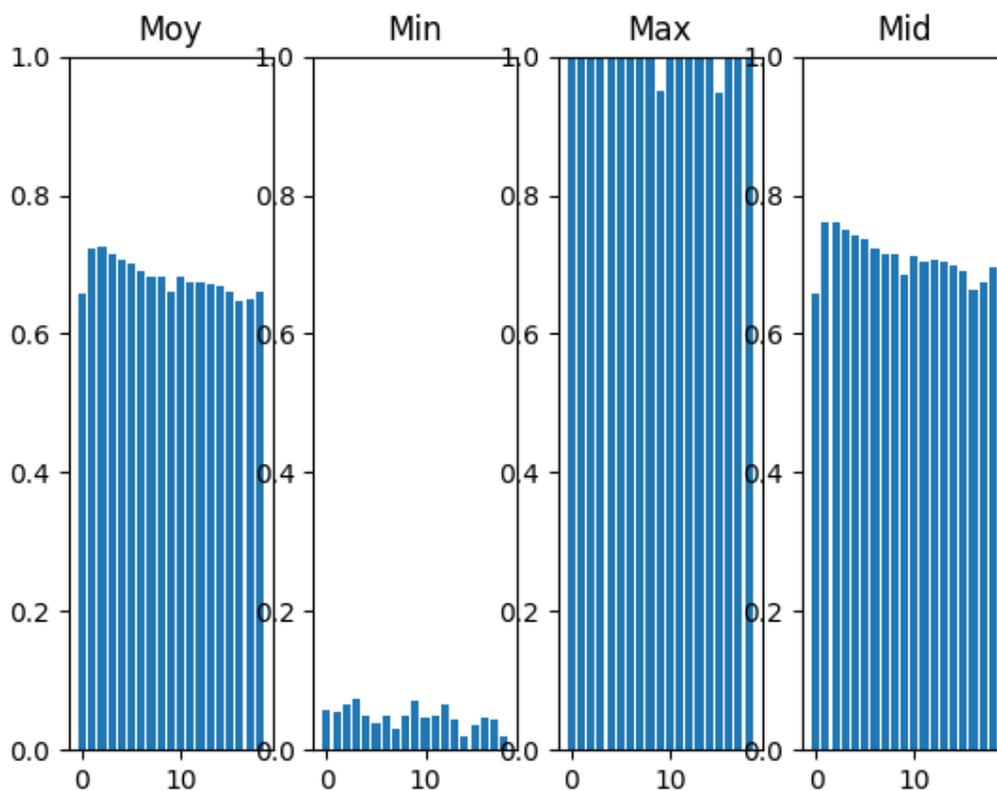


Figure 19 : Graphique de l'évaluation du modèle avec Label Smoothing et le nouveau jeu donnée non lissé

La seconde expérimentation a été réalisée avec l'ajout du principe de "Label Smoothing" [9]. Ce procédé permet de retarder le sur-apprentissage (overfitting) et aussi résoudre les problèmes de sur-confiance (overconfidence) dans des tâches de classification. Dans ce modèle, seul le discriminateur est un classifieur. Il prédit si l'information en entrée peut être considérée comme réelle ou non. C'est une classification booléenne. Les réponses sont soit 0 soit 1. Le "Label Smoothing" vient lisser les valeurs. Au lieu d'avoir des valeurs fixes, ces dernières vont être étalées. De cette manière nous aurons une valeur aléatoire comprise dans un intervalle autour de la valeur initiale.

Le Label Smoothing remplace le vecteur y_{hot} encodé en "one-hot" avec une mixture de y_{hot} et une distribution uniforme:

$$y_{\text{ls}} = (1 - \alpha) * y_{\text{hot}} + \alpha / K$$

où K est le nombre de classe, et α un hyper paramètre qui détermine la valeur du lissage. Si $\alpha = 0$, on obtient le vecteur original. Si $\alpha = 1$, on obtient une distribution uniforme.

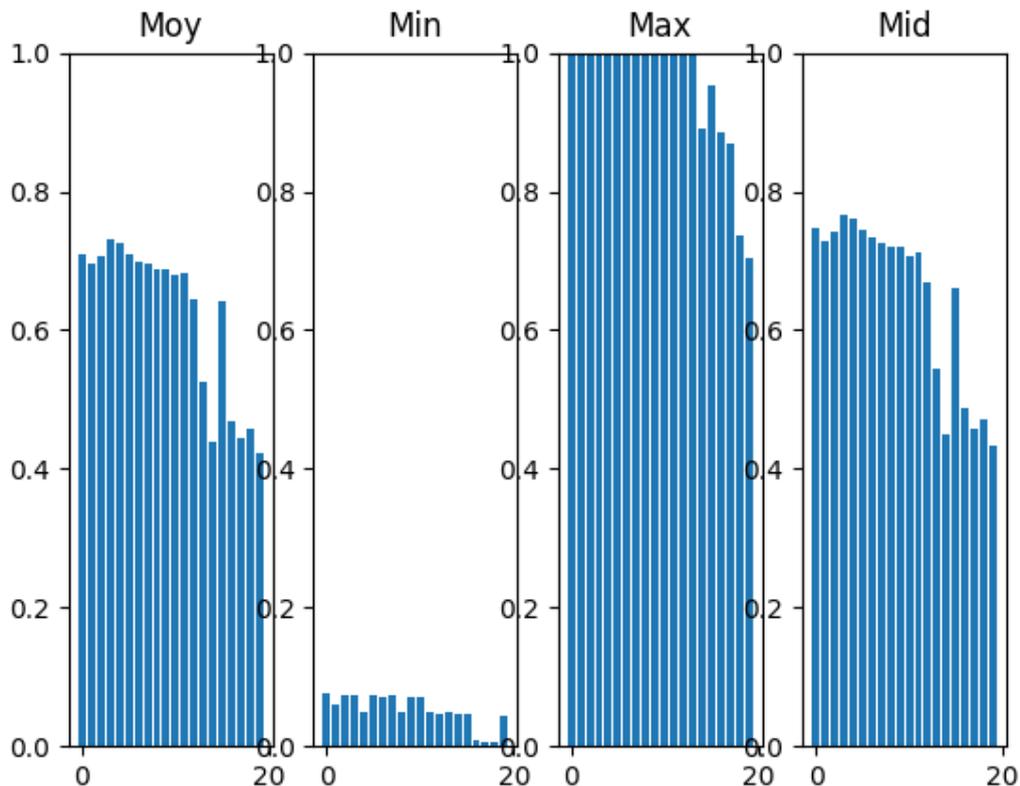


Figure 20 : Graphique de l'évaluation du modèle avec Label Smoothing et l'ancien jeu de données lissé

Avec cet ajout, il a été décidé de le tester sur les deux jeux de données. Dans un premier temps, il a été remarqué, sur l'ancien jeu de données (figure 20), une stabilisation dans les performances suivit d'une chute sur la fin d'entraînement. Puis l'évaluation sur le nouveau jeu de données (figure 21) montre une absence de rupture de stabilité tout du long de l'apprentissage.

Cependant il a été constaté de meilleur résultat sur l'ancien jeu de données que sur le nouveau avec une différence de 3% pour les meilleures sauvegardes.

Néanmoins, la présence d'une baisse progressive sur l'évaluation montre un sur-apprentissage croissant qui commence dès la 3-4e sauvegarde soit 30-40 epochs.

5 - Noisy Label

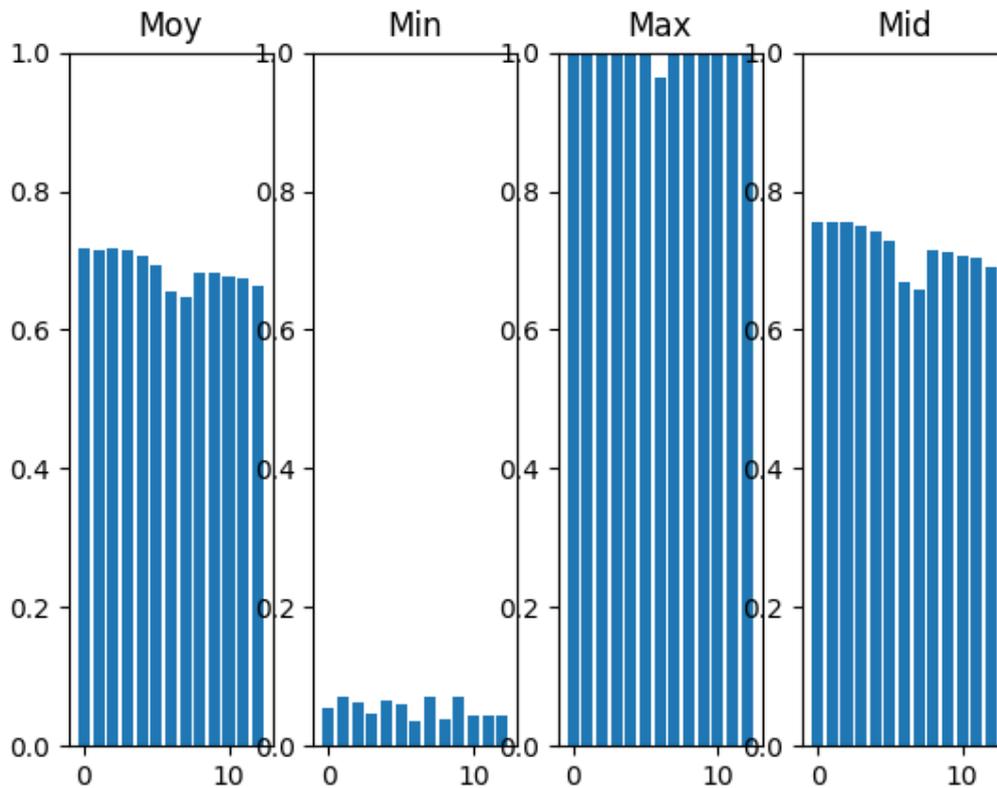


Figure 21 : Graphique de l'évaluation du modèle avec Noisy Label et le nouveau jeu donnée non lissé

Dans une troisième expérimentation, le principe de "Noisy Label" [9] a été utilisé. Les valeurs de vérité terrain que l'on donne au discriminateur sont toujours vraies. Avec ce procédé on introduit une erreur dans les valeurs. Ainsi des valeurs à 1 seront à 0 et inversement. Dans cette expérimentation il a été introduit un biais de 5% appliqué de manière aléatoire.

Ce procédé a pour but de réduire la sur-confiance du modèle mais aussi de repousser le sur-apprentissage.

Sur la figure 21, le modèle s'est stabilisé à l'exception de deux sauvegardes qui font un petit bond vers le bas avant de revenir. En revanche, la progression est décroissante dès les premières sauvegardes et les résultats sont identiques à ceux du Label Smoothing.

6 - Combinaison de Label Smoothing et Noisy Label

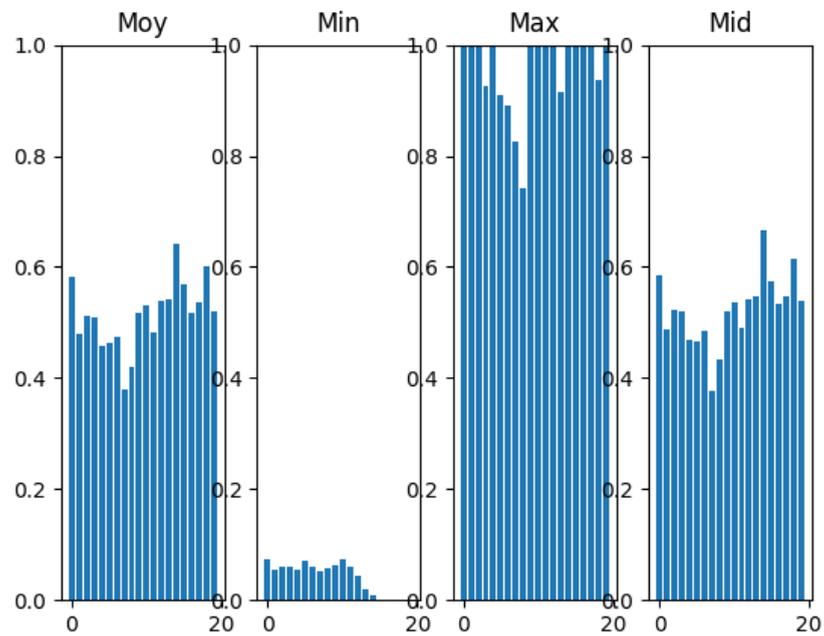


Figure 22 : Graphique de l'évaluation du modèle avec Label Smoothing et Noisy Label sur le nouveau jeu donnée non lissé

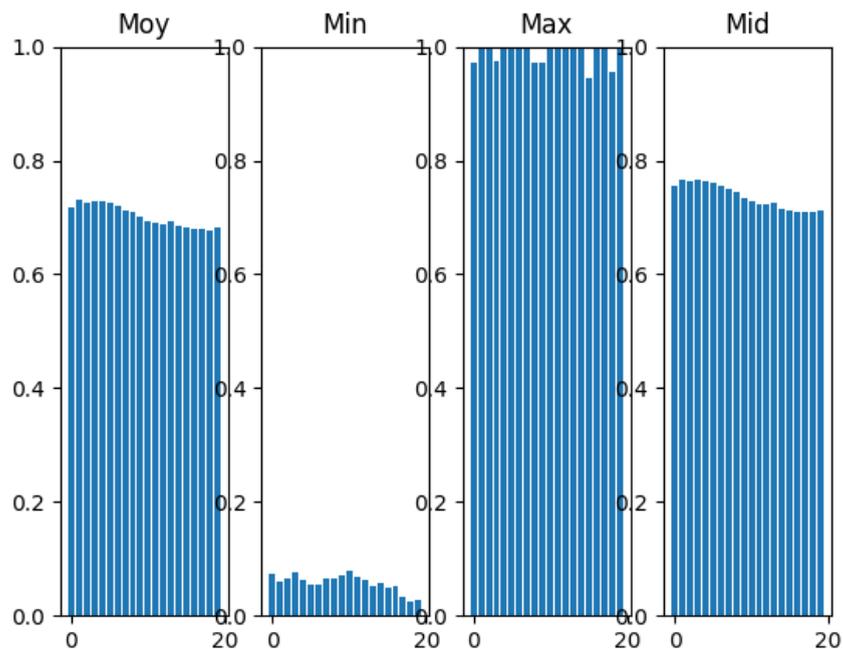


Figure 23 : Graphique de l'évaluation du modèle avec Label Smoothing et Noisy Label sur le jeu donnée lissé

Dans une dernière expérimentation, il a été testé l'implémentation des deux principes. Ce qui est recherché est d'accroître l'apprentissage en repoussant le sur-apprentissage afin d'obtenir de meilleures performances. Il a été constaté que chacun apportait une stabilité lors de l'apprentissage, mais le sur-apprentissage était présent assez rapidement.

Sur ces deux graphiques, une première chose apparaît. Sur le jeu de données non lissée (fig 22), l'apprentissage est devenu instable très rapidement sans réussir à se stabiliser. En revanche sur l'ancien jeu de données lissé (fig 23), il a su se maintenir tout du long des 200 epochs.

Cependant la décroissance des performances survient toujours aussi tôt (10 epochs plus tard). Les performances quant à elles ne sont pas meilleures.

7 - Conclusions des expérimentations

Avec ces expérimentations, des réponses à certains problèmes ont été apportés mais d'autres questionnements sont apparus.

Dans un premier temps, la stabilité d'entraînement des GANs est un enjeu qui ne doit pas être mis de côté. Le Label Smoothing et le Noisy Label ont montré leur efficacité, dans une certaine mesure, à maintenir l'entraînement stable sur 200 epochs.

Il a été remarqué par ailleurs que 200 epochs n'était pas nécessaire car le sur-apprentissage semble survenir dès 40-50 epochs et les expérimentations n'ont pas été en mesure de le repousser plus loin.

Dans un second temps, les évaluations ont révélé un maximum entre 0.70 et 0.74 en fonction des expérimentations avec les meilleurs résultats.

Cependant certains résultats suscitent l'interrogation. Dans la dernière expérimentation, le nouveau jeu de données non lissée a perdu sa stabilité alors que l'ancien lissé a su le garder.

De plus, les résultats sont légèrement meilleurs sur l'ancien jeu de données que sur le nouveau. La possibilité de lisser les contours entre les segmentations semble moins contraignante pour le modèle.

Par ailleurs, il est bon de signaler que même avec le nouveau jeu de données non lissé, les modèles qui s'entraînent dessus donnent des prédictions avec un nombre supérieur de couleurs qu'il n'y a de classe, mais ils en génèrent moins que sur l'ancien jeu de données. Cela peut s'expliquer que le modèle génère une image en sortie et ses possibilités ne sont pas conditionnées à une série de possibilités que sont les possibles classes. Une solution serait d'adapter l'architecture à ce type de sortie en "one hot encoding".

Pour finir, une partie importante des erreurs de segmentations commises par les modèles surviennent dans la distinction entre bâti, trottoirs et chaussée quand la vision est réduite. L'enrichissement des chaussées a montré son efficacité et un autre enrichissement pourrait augmenter très nettement les performances. Il a été évoqué en 5.3. Ajouter les contours des bâti pourrait encore réduire l'incertitude dans ces démarcations.

Chapitre 6 : Conclusion

Dans cette étude la réalisation d'un jeu de données avec un très grand nombre d'éléments est un succès avec 180 milles couples d'image. La principale contrainte dans l'apprentissage profond sur de l'imagerie est la taille du jeu d'apprentissage qui est souvent très limité. La démarche de Paris de mettre à libre disposition leur ressources géographique a été fort utile et cette pratique commence à se démocratiser dans d'autres villes. Ce qui laisse imaginer les nombreuses possibilités à venir. Par ailleurs diversifier la base d'apprentissage avec de nouvelles villes permettra au modèle d'appréhender d'autres architectures et réseaux routiers.

De plus, l'enrichissement des images avec des informations issues d'OpenStreetMap a montré une efficacité à double tranchant. Dans un premier temps, un modèle entraîné avec un enrichissement en filaire pour décrire la chaussée est très dépendant de ce dernier, s'il n'est pas présent, cela produit des conflits de segmentation. En revanche, quand ce dernier est présent, les résultats montrent une réelle efficacité. Une interrogation persiste néanmoins concernant l'enrichissement des passages piétons sous forme de point. Pour ce dernier, le modèle semble très bien s'adapter avec et sans enrichissement. Cette différence reste une question grande ouverte.

De plus, et afin d'améliorer encore les performances du modèle, l'intégration de l'enrichissement du contours du bâti devrait grandement aider et sera pour de futur recherche. De même que la superposition de l'enrichissement sur l'image, l'idée d'ajouter une dimension à l'image pour y ajouter les données pourra faire l'objet d'une future étude.

Les expérimentations ont elles aussi apporté leurs lots d'informations. Dans un premier temps, remédier à l'instabilité de l'apprentissage des GAN avec les procédés de Label Smoothing et de Noisy Label.

Des éléments dans les résultats sont encore incompris, comme notamment dans le chapitre expérimentation - 6. Néanmoins il a été constaté que repousser le sur-apprentissage n'était pas une solution apportant une amélioration significative des performances. Le modèle atteint un maximum très rapidement lors de l'apprentissage. Faut-il affiner le modèle pour longtemps dans le temps d'une autre manière, cela reste une question grande ouverte.

Par ailleurs, les différences observées entre le jeu de donnée non lissé et lissé semble montrer que le jeu de donnée lissé est moins contraignant pour le modèle avec une légère hausse des performances. Le format de sortie du modèle est un sujet d'étude qui permettra peut être d'améliorer la qualité des prédictions.

Pour finir, les modèles retournent des prédictions avec une évaluation comprise entre 70 et 74 % mais à l'œil humain les résultats sont surprenant. La végétation est probablement un des plus gros défis avec l'ombre des bâtiments. Cela cause des troubles sur la démarcation du bâti, trottoir et chaussé mais malgré tout les modèles semblent y être relativement robuste. Il a été montré dans ce rapport que de nombreuses autres améliorations peuvent encore être apportées afin d'accroître la qualité des prédictions. Quelques questionnements restent encore ouverts pour d'autres études.

Conclusion personnelle

J'ai été ravi d'avoir travaillé sur ce sujet. Les débuts étaient plutôt flou mais plus on avance plus on en découvre. Jean-Marie et Jérémie ont su m'épauler, m'aider et me guider tout du long de ces 6 mois. Le LIMOS est un très bon lieu de travail et ils font tout leur possible pour qu'il le reste même durant cette crise sanitaire.

Ce stage m'a donné l'envie de continuer dans la recherche sur de l'intelligence artificielle. Les opportunités avec cette technologie sont très prometteuses. J'ai envie de suivre et participer à son évolution.

J'ai appris de nombreuses connaissances en lisant les nombreux papiers autour de mon sujet et j'ai aussi amélioré ma stratégie de travail

Glossaire

Batchsize : désigne le nombre d'éléments que l'on utilise par étape d'apprentissage.

Label Smoothing : Lissage d'étiquette, faire en sorte que les étiquettes soient moins précises et élargissent leur contenu

Noisy Label : bruitage d'étiquette, introduction d'un bruit ou d'erreur dans les étiquettes représentant les données.

Couple d'image : un ensemble de deux images, constitué d'une orthophotographie enrichie avec des données issues d'OSM avec sa segmentation sémantique associée.

QGIS : Logiciel de système d'information géographique permettant de manipuler des objets géographiques sous forme de couche afin de réaliser des cartes avec des légendes.

Orthophotographie : plus communément appelé orthophoto, ce sont des clichés aériens idéalement pris orthogonalement par rapport au sol où la géométrie a été redressée de façon à pouvoir superposer l'image sur une carte.

Segmentation sémantique : procédé au travers duquel une image est décomposée en différentes régions étiquetées. Cette étiquette est définie par la sémantique.

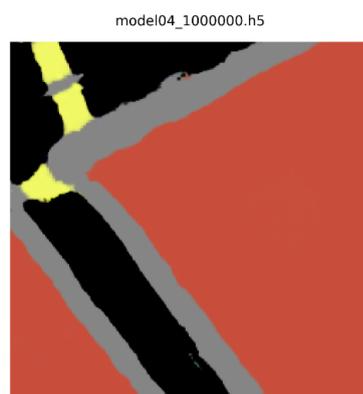
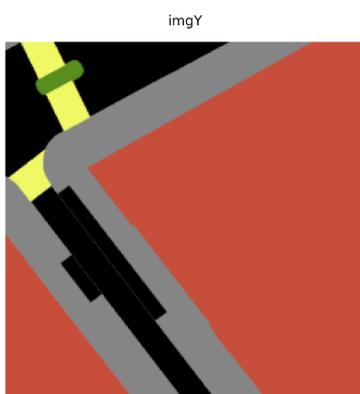
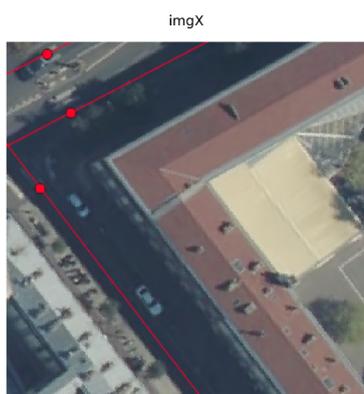
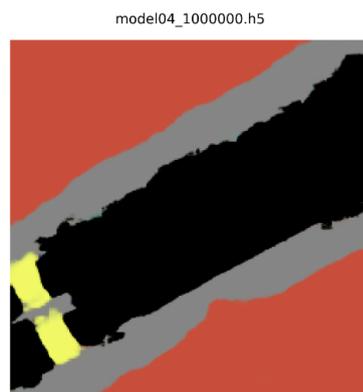
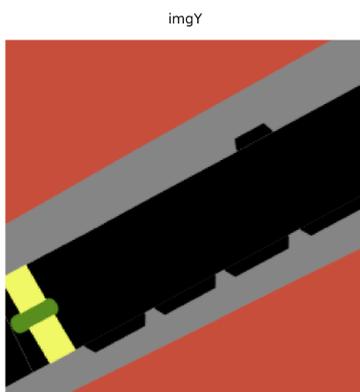
Métrique : Outil de mesure pour évaluer la qualité d'un modèle selon les mesures utilisées

Dropout : Le Dropout est un outil permettant de bloquer un nœud aléatoirement durant l'entraînement. Il permet de réduire le sur-apprentissage et améliore la généralisation des modèles.

Annexe

Échantillons de prédictions de la 4eme sauvegarde du modèle avec Label Smoothing sur un jeu de données lissé.

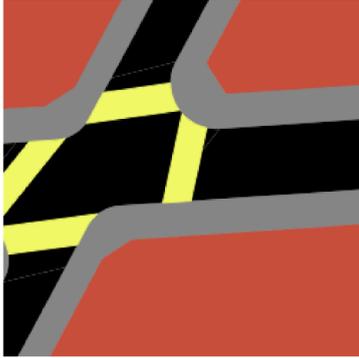
Sur ces véritables terrains, les classes d'îlots, de terre-plein et de trottoirs sont encore différenciés.



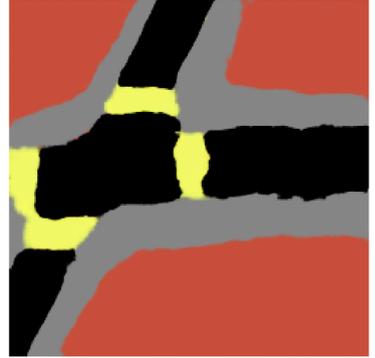
imgX



imgY



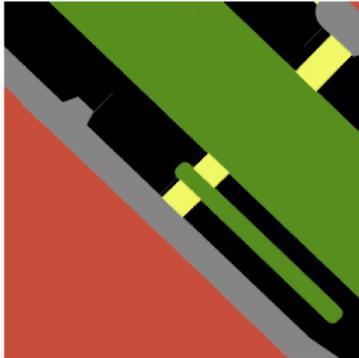
model04_1000000.h5



imgX



imgY



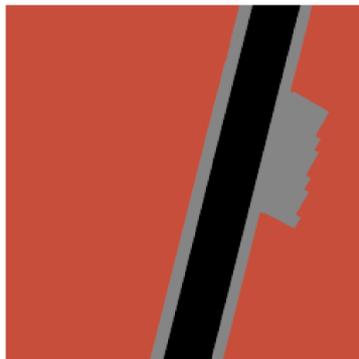
model04_1000000.h5



imgX



imgY



model04_1000000.h5



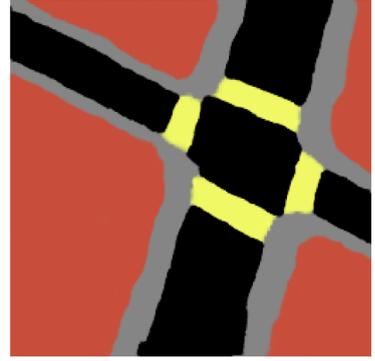
imgX



imgY



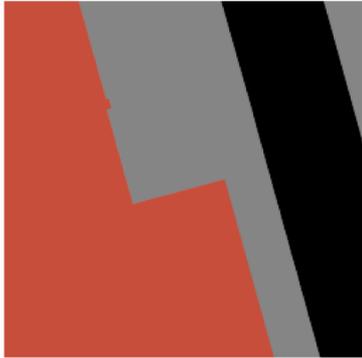
model04_1000000.h5



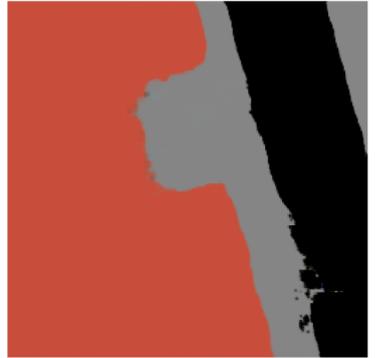
imgX



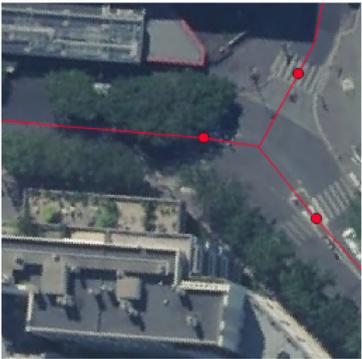
imgY



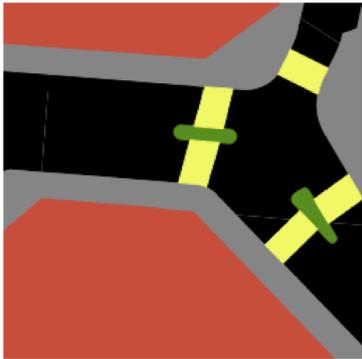
model04_1000000.h5



imgX

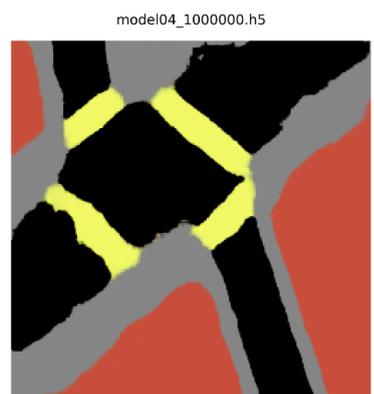
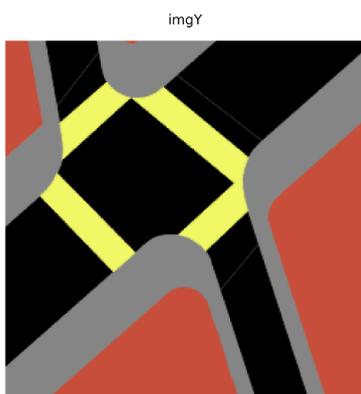
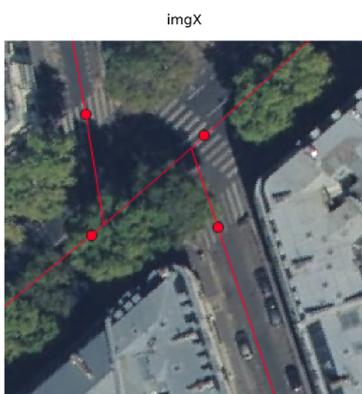
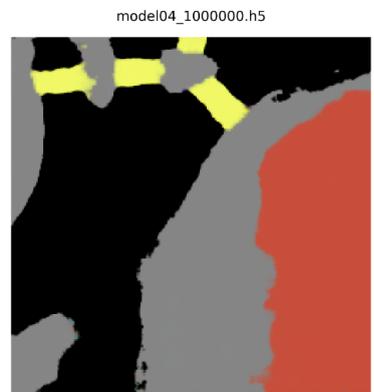
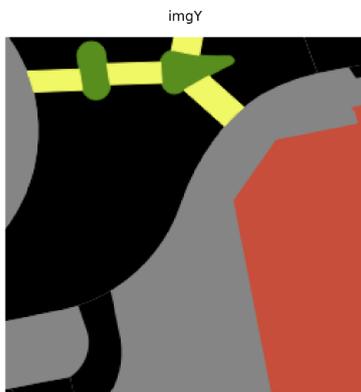
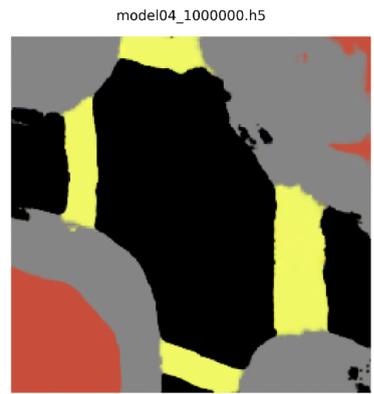
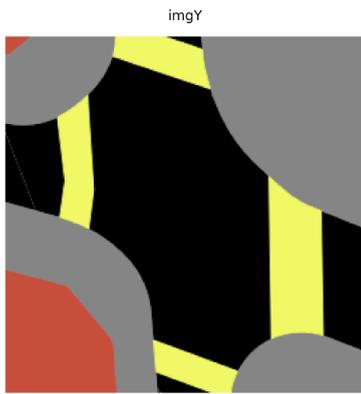


imgY



model04_1000000.h5

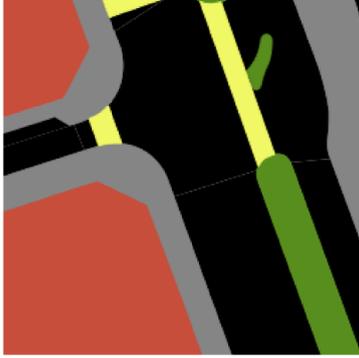




imgX



imgY



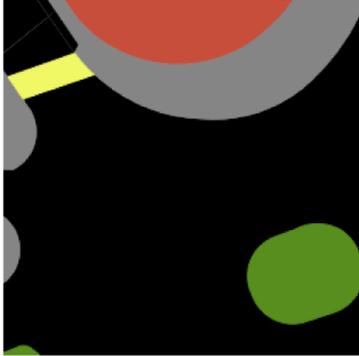
model04_1000000.h5



imgX



imgY



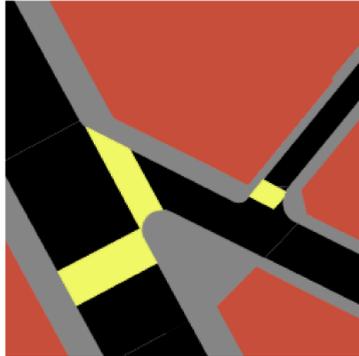
model04_1000000.h5



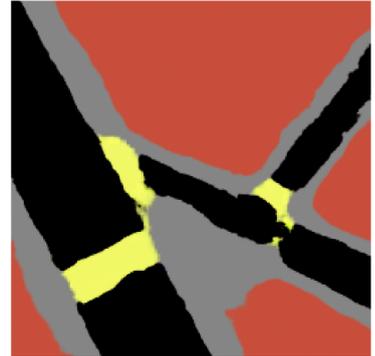
imgX



imgY



model04_1000000.h5



Références et poursuite de lecture :

- [1] Fillières-Riveau, G., Favreau, J. M., Barra, V., & Touya, G. (2020). Génération de cartes tactiles photoréalistes pour personnes déficientes visuelles par apprentissage profond. *Revue Internationale de Géomatique*, 30(1-2), 105-126.
<https://hal.uca.fr/hal-03208233v1>
- [2] Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems* 27 (2014). : <https://arxiv.org/abs/1406.2661>
- [3] Others GAN type : <https://github.com/eriklindernoren/Keras-GAN>
- [4] Luc, Pauline, et al. "Semantic segmentation using adversarial networks." *arXiv preprint arXiv:1611.08408* (2016). : <https://arxiv.org/pdf/1611.08408.pdf>
- [5] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.:
<https://arxiv.org/pdf/1611.07004>
- [6] pix2pix github : <https://github.com/phillipi/pix2pix>
- [7] How to implement pix2pix gan models from scratch with keras :
<https://machinelearningmastery.com/how-to-implement-pix2pix-gan-models-from-scratch-with-keras/>
- [8] Salimans, Tim, et al. "Improved techniques for training gans." *Advances in neural information processing systems* 29 (2016): 2234-2242. : <https://arxiv.org/abs/1606.03498>
- [9] Soumith GAN Hack github : <https://github.com/soumith/ganhacks>
- [10] How to code generative adversarial network hacks :
<https://machinelearningmastery.com/how-to-code-generative-adversarial-network-hacks/>
- [11] How to evaluate GAN :
<https://machinelearningmastery.com/how-to-evaluate-generative-adversarial-networks/>
- [12] Evaluating image segmentation models :
<https://www.jeremyjordan.me/evaluating-image-segmentation-models/>
- [13] OpenDataParis : <https://opendata.paris.fr/pages/home/>
- [14] OpenStreetMap : <https://www.openstreetmap.fr/>
- [15] IGN : <https://www.ign.fr/>
- [16] Keras-Tensorflow : <https://www.tensorflow.org/guide/keras?hl=fr>