Introduction

Le levé topographique des corps de rue est une pratique courante dans le domaine de la cartographie urbaine. Il permet de recueillir des données précises sur les caractéristiques d'un terrain, telles que la pente de la rue, la largeur de la chaussée, tous les tampons des différents réseaux (eau, électricité, gaz, etc.), etc. Au fil du temps, les techniques de levé topographique ont évolué de manière significative pour devenir plus performantes en termes de quantité, de qualité et d'exhaustivité des données recueillies. Aujourd'hui, il existe une variété de technologies de levé topographique disponibles, allant des méthodes traditionnelles telles que le levé topographique manuel à l'aide d'un théodolite et d'une planche à dessin, aux technologies plus modernes comme le levé topographique assisté par ordinateur, le balayage laser, la photogrammétrie et les drones équipés de caméras. Les données recueillies lors de ces levés sont utilisées pour produire des plans topographiques, dont la façon de les faire doit s’adapter à chaque type de matériel de levé utilisé.

Après la collecte des données de levé topographique, la production d'un plan topographique implique généralement plusieurs étapes. Tout d'abord, les données doivent être triées, nettoyées et organisées pour assurer leur qualité et leur exactitude. Ensuite, les données sont généralement traitées à l'aide de logiciels de DAO (Dessin Assisté par Ordinateur) pour produire une représentation graphique précise et détaillée du terrain. Les données issues des différentes méthodes de levé topographique ont leurs propres particularités et peuvent nécessiter une adaptation de la méthode de production du plan topographique. Une estimation a été faite lors d’un PFE en prenant l'exemple de la station totale et d'un dispositif MMS, où ils prévoient 50% du temps sur terrain et 50% du temps au bureau pour le cas d'utilisation de la station totale, et un ratio temps terrain / temps bureau de 0,018 pour le MMS et de 0,2 pour une méthode photogrammétrique. [BASTARD-ROSSET 2022, p. 11] dans le cas où le traitement se fait de manière classique à l'aide de DAO.

En avançant dans le temps, la vectorisation a connu de nombreux changements et améliorations depuis ses débuts manuels. L'arrivée des logiciels de DAO a introduit l'utilisation de blocs prédéfinis pour accélérer le processus de vectorisation. Cela a permis de gagner du temps en utilisant des codifications et des blocs de symboles communs tels que les arbres, les panneaux de signalisation, etc. Les blocs étaient également utiles pour maintenir la cohérence visuelle du plan. Avec l'avancée de la technologie, la vectorisation automatique a été développée et diffère selon le type de données utilisé. Enfin, l'intelligence artificielle (IA) a été utilisée pour améliorer la précision et la vitesse de la vectorisation automatique. Les logiciels d'IA sont capables de reconnaître des objets complexes et de les vectoriser avec une grande précision. La vectorisation automatique avec l'IA permet également de gagner du temps et de réduire les erreurs liées à la vectorisation manuelle.

La photogrammétrie, dans ce contexte d'étude, est très intéressante en raison de l'exhaustivité de la captation et du transfert fidèle de l'image de la réalité terrain. D'un autre côté, il est nécessaire d'adapter le mode d'extraction de la donnée en format vectoriel pour mieux exploiter le résultat de la photogrammétrie, car le temps de production, en utilisant les outils de DAO, reste considérablement plus élevé par rapport au temps de levé sur terrain. Comment optimiser ce temps de production en utilisant des solutions automatiques de vectorisation ?

# État de l’art

Cette première partie sera consacrée à présenter, dans un premier temps, les méthodes d'acquisition de données topographiques les plus souvent utilisées. Dans un deuxième temps, les différentes techniques de vectorisation pour produire un plan topographique seront abordées. Enfin, une partie sera consacrée aux techniques modernes d'extraction de données dans le domaine de l'IA.

Cette étude de l'état de l'art se concentrera principalement sur le traitement d'image plutôt que sur le traitement de nuages de points en vue d'obtenir des résultats pratiques. Elle pourra être suivie d'une expérimentation similaire sur les nuages de points, ou d'un affinement des résultats et d'une adaptation du code de la première partie aux logiciels utilisés, à savoir Metashape, AutoCAD et QGIS.

## Acquisition de données

En effet, il existe plusieurs méthodes pour l'acquisition de données topographiques. Chacune d'entre elles présente des avantages et des inconvénients en termes de précision, de coût, de temps et de ressources nécessaires.

### Acquisition terrestre directe

Ce type d'acquisition, qui fait référence au GPS et à la station totale, est considéré comme classique pour réaliser des levés topographiques car il permet de mesurer les distances et les coordonnées directement sur le terrain. La station totale est utilisée par un opérateur pour mesurer des points caractéristiques individuels avec un haut degré de précision [Liang et al., 2016]. Comparé à d'autres méthodes, ce type d'acquisition est souvent utilisé pour délimiter les contours et géoréférencer le projet.

### Acquisition Lidar

Le Lidar (Light Detection and Ranging) est une technologie de télédétection active qui permet de mesurer les distances entre le capteur et la surface cible. Cette méthode se base sur l'émission d'un faisceau laser qui est réfléchi par la surface de la cible, permettant ainsi de calculer sa distance. On distingue trois types de Lidar. [Höfle et al. 2011]

Une image contenant diagramme

Description générée automatiquement

Figure 1 : Terminologie et classification Lidar. [Höfle et al. 2011]

* + - Le Lidar aéroporté (ALS), également connu sous le nom de Lidar aérien, est une méthode de télédétection qui utilise un faisceau laser monté sur un avion, un hélicoptère ou un drone pour capturer des données topographiques en 3D. Il est à noter qu'il y a différentes applications selon la plateforme porteuse du lidar. Parmi ces plateformes, le drone est susceptible d'être utilisé pour réaliser des levés. Selon [Pinton et al. 2020], par rapport au Lidar, la photogrammétrie montre une précision inférieure avec plus d'erreurs dues aux ombres et à la présence de discontinuités de surface, ainsi qu'une automatisation moins avancée dans le traitement des données et un temps de production plus long dans le domaine de la végétation.
    - Le Lidar terrestre (TLS) est un instrument statique qui utilise des lasers pour mesurer son environnement. Il mesure les distances et les angles avec précision grâce à un mécanisme de déviation du faisceau optique pour obtenir des observations ponctuelles 3D à partir des surfaces de l'objet. Contrairement à la station totale, le TLS est particulièrement avantageux dans des zones de travail comme les routes, car il assure la vitesse, la sécurité de l'opérateur et évite les perturbations du trafic. [Mill et al. 2011]
    - Le Lidar mobile (MMS) est une méthode d'acquisition de données topographiques 3D qui utilise des capteurs montés sur un véhicule en mouvement pour capturer des informations sur l'environnement. Selon [Kawamura et al. 2022], cette méthode est particulièrement adaptée pour le levé des routes, car elle permet de détecter le pourcentage de fissures et de colmatage à partir des données d'image.

### Acquisition Photogrammétrique

L'acquisition photogrammétrique est une méthode d'acquisition de données topographiques qui se base sur l'utilisation de photographies aériennes ou terrestres pour créer des modèles en trois dimensions de la surface. Cette technique repose sur le principe de la stéréoscopie, qui permet de mesurer la profondeur en utilisant des images stéréoscopiques prises sous différents angles.

En topographie, les différentes méthodes photogrammétriques sont similaires aux méthodes laser aériennes et terrestres en termes d'utilisation et d'étude de cas. Une étude menée par [Ganz et al. 2019] a montré que les deux techniques photogrammétrique et lasergrammétrique par drone peuvent être utilisées pour étudier la hauteur des arbres. Les deux techniques peuvent être complémentaires l'une à l'autre, comme dans l'étude menée par [Lerma et al. 2010] pour modéliser un site archéologique, ou dans le cas d'inspection des ponts [Riveiro et al. 2013] où l'objectif est de comparer les deux techniques.

Une image contenant diagramme

Description générée automatiquement

Figure 2 : Relation entre les méthodes de mesure et la taille et la précision de l'objet [Luhmann et al, 2019]

Dans le domaine de la topographie, qui s'inscrit dans le cadre de la photogrammétrie architecturale et technique, il est possible, selon [Luhmann et al. 2019] et comme indiqué dans la Figure 2, d'obtenir une précision submillimétrique en capturant des objets d'une taille d'environ dix mètres.

* + - La photogrammétrie aérienne est une technique qui consiste à utiliser des photographies aériennes pour créer des modèles en 3D de la surface. Dans le contexte de levé topographique de la rue, l'utilisation de l'avion est moins adaptée pour garantir la précision requise et capturer tous les éléments présents sur la chaussée. En revanche, l'utilisation de drone peut fournir des résultats satisfaisants en termes de précision, comme démontré dans l'étude de Tan et Li (2019), qui a permis de détecter la dégradation de la rue avec une précision du modèle 3D ne dépassant pas 1,5 cm. Néanmoins, en France, l'utilisation du drone est très réglementée et limitée, et son utilisation est interdite dans les agglomérations [1].
    - La photogrammétrie terrestre présente l'avantage de simplicité et de flexibilité, ainsi que la capacité de produire des modèles détaillés à partir de données relativement peu coûteuses et accessibles.

## La vectorisation

Après la collecte des données obtenues lors du levé de la rue, une étape de post-traitement est nécessaire pour nettoyer, corriger et parfois aligner les nuages de points dans le cas d'un levé laser. Le résultat final de cette étape est la production d'orthophotos et de nuages de points prêts à être utilisés pour la suite de la modélisation. La vectorisation consiste ensuite à créer des plans topographiques précis et détaillés à partir des données relevées sur le terrain. Au fil des années, la vectorisation a évolué et a pris plusieurs formes. Au départ, elle était effectuée manuellement par des opérateurs qualifiés à l'aide de tables à digitaliser. Bien que cette méthode soit précise, elle est très coûteuse en temps et en main-d'œuvre. Avec l'avènement de la DAO (Dessin Assisté par Ordinateur), la vectorisation a commencé à être automatisée et les logiciels de CAO (Conception Assistée par Ordinateur) ont été utilisés pour dessiner les plans topographiques. Aujourd'hui, la vectorisation se fait principalement à l'aide de logiciels de modélisation 3D qui permettent de créer des modèles précis à partir des nuages de points et des orthophotos. Cette méthode est plus rapide et plus précise que la vectorisation manuelle et permet de produire des plans topographiques détaillés en un temps record.

### La vectorisation automatique

#### La vectorisation automatique basé sur le seuillage

La vectorisation automatique basé sur le seuillage est une méthode permettant de segmenter les pixels d'une image en deux catégories, ceux qui appartiennent à l'objet à vectoriser et ceux qui n'y appartiennent pas, grâce à l'utilisation d'un algorithme. Elle est particulièrement utile pour la vectorisation de plans topographiques où l'on souhaite extraire des informations à partir d'une image de fond. Dans l’étude de [Robin & Gondeaux, 2007], une méthode est utilisée sur des images satellitaires pour l’objectif du suivi d'évolution des bâtiments à Rouen, avec des pré-traitements qui ont précédé l’opération de seuillage. Toutefois, il est important de noter que ce seuillage a été effectué sur des images de résolution de 50 cm, et uniquement pour segmenter les bâtiments. Sur des orthophotos de résolution plus importante, cette méthode peut être moins utile.

#### La vectorisation automatique par détection des contours

Ce type de vectorisation utilise des algorithmes pour détecter des caractéristiques de texture dans les images, telles que des motifs répétitifs ou des zones homogènes. Les caractéristiques sont ensuite converties en vecteurs pour créer des cartes topographiques. L'un des algorithmes les plus couramment utilisés pour la vectorisation est la transformée de Hough, qui est encore en recherche et en application depuis 59 ans [Mukhopadhyay & Chaudhuri, 2015]. Cette technique permet de convertir les pixels de l’espace coordonnées en espace paramètres, où la géométrie représente un point dans cette espace ce que nous appelons la rétroprojection. L’exemple de la ligne le plus simple a été bien détaillé dans le premier papier de cette méthode, car l’équation de la ligne est la plus simple

Y = m.x + c Équation 1 : équation de la ligne

F(m, c) = y – m.x – c = 0 Équation 2 : équation des paramètres

En utilisant les coordonnées des pixels à partir du graphique (A) et en les substituant dans l’Équation 2, nous obtenons différentes combinaisons de paramètres (m, c). Ces différentes combinaisons sont ensuite représentées par le graphique (B) dans la Figure 3. L'image (C) illustre le nombre de pixels qui appartiennent à chaque ligne correspondant à la combinaison spécifique (m, c). Dans l'image (C), les valeurs élevées indiquent les combinaisons de paramètres (m, c) qui définissent une ligne. Pour d'autres formes géométriques, le processus est similaire, mais l'équation utilisée pour générer les paramètres (m, c) diffère en fonction de la géométrie de l'objet en question.

Une image contenant diagramme, ligne

Description générée automatiquement

Figure 3 : La base de la transformée de Hough pour la détection de ligne : (A) (x, y) espace point dans l’image. (B) (m, c) espace des paramètres. (C) espace accumulateur correspondant à (B). [Illingworth & Kittler.1988]

#### La vectorisation automatique basé sur la classification

Dans la classification supervisée, l'utilisateur fournit des exemples d'images annotées pour chaque classe au modèle, qui apprend ensuite à reconnaître ces classes en fonction de ces exemples. En revanche, dans la classification non supervisée, le modèle recherche des groupes ou des motifs d'objets similaires dans l'image sans être guidé par des annotations. L'étape suivante consiste à identifier les frontières et les contours des objets, puis à effectuer une conversion vectorielle. Ces méthodes sont couramment utilisées en télédétection et dans des images à petite échelle.

## L’intelligence artificielle

L'intelligence artificielle (IA) est un domaine de l'informatique qui cherche à créer des programmes capables de simuler l'intelligence humaine, en termes de raisonnement, d'apprentissage, de compréhension, de résolution de problèmes et de prise de décision.

### L’évolution de l’IA

Dans cette section, nous explorons comment l'intelligence artificielle a évolué au fil des années pour inclure des concepts d'apprentissage machine et d'apprentissage profond. La Figure 4 représente, de manière générale, la relation entre ces concepts d’IA et quels algorithmes peuvent être utilisés

Une image contenant texte, capture d’écran, affichage, Police

Description générée automatiquement

Figure 4 : diagramme de Venn de l'apprentissage automatique et des classes. [Janiesch & al, 2021]

L'apprentissage machine représente une sous-discipline de l'IA qui permet aux machines d'acquérir des compétences à partir de données et d'exemples, plutôt que d'être explicitement programmées pour des tâches spécifiques. Cette approche favorise l'amélioration continue, car les machines peuvent ajuster leurs performances au fil du temps. Pour cela, des algorithmes sont appliqués de manière itérative en se basant sur des données d'entraînement spécifiques à un problème donné. Cela permet aux ordinateurs de découvrir des motifs subtils et des informations cachées au sein des données, sans que des instructions précises ne soient fournies. En capitalisant sur les connaissances accumulées et en identifiant des régularités au sein de vastes ensembles de données, l'apprentissage machine aboutit à la prise de décisions fiables et reproductibles.

L'apprentissage profond est une évolution importante de l'apprentissage machine, qui repose sur des réseaux de neurones artificiels et qui a démontré des performances remarquables ces dernières années. [Xin et al, 2018]. Et selon [IBM Data and AI Team, 2020], un réseau neuronal de plus de trois couches, y compris les entrées et la sortie, peut être considéré comme un algorithme d'apprentissage profond.

### L’impact de l’IA sur la détection d’objet et la segmentation

Dans le contexte de la détection et de la segmentation d'objets, l'apprentissage profond a joué un rôle révolutionnaire. Les réseaux de neurones convolutifs (CNN) ont permis de créer des modèles capables de distinguer et de localiser des objets dans des images avec une précision impressionnante lorsque le groupe de Hinton a remporté le ImageNet Large Scale Visaul Recognition Challenge (ILSVRC) avec l'apprentissage en profondeur [Wang, 2016]. La segmentation d'images en utilisant des réseaux de neurones a permis de diviser les images en régions distinctes, contribuant ainsi à des applications allant de la médecine à la vision par ordinateur

## La détection et la segmentation d’objets

La détection et la segmentation sont deux techniques importantes de l'apprentissage profond largement utilisées dans divers domaines, tels que la reconnaissance d'objets et la vision par ordinateur. Elles permettent notamment de détecter des objets ou des éléments d'intérêt dans des images ou des vidéos. Par exemple, une étude menée par [Haris & Glowacs, 2021] a utilisé ces techniques pour détecter des objets sur des images aériennes de la rue, tandis qu'une autre étude menée par [Yiğit & Uysal, 2020] a utilisé la détection pour identifier les rues sur des rasters.

### La détection

La détection en intelligence artificielle consiste à la capacité d'un système informatique à identifier et localiser des objets spécifiques dans une image, une vidéo ou un nuage de points. Cette technique utilise l'apprentissage automatique pour entraîner un modèle à reconnaître les motifs et les caractéristiques dans les données d'entrée et à identifier les objets d'intérêt en créant des boîtes englobantes appelées Bbox. Il existe différents types de détection.

#### La détection d’objets sur les données 2D (raster)

La détection d'objets dans les images optiques est une tâche essentielle mais complexe dans le domaine de l'analyse d'images aériennes et satellitaires. Elle est importante dans de nombreuses applications et a suscité un intérêt croissant ces dernières années [Cheng & Han, 2016]. Traditionnellement, les méthodes de détection reposaient sur la correspondance de modèles, les méthodes basées sur la connaissance et l'analyse d'image basée sur l'objet [Cheng & Han, 2016]. Cependant, les méthodes d'apprentissage automatique basées sur des transformateurs et des mécanismes d'auto-attention ont montré des performances élevées, mais avec un coût de calcul élevé et nécessitent un pré-entraînement. Ces méthodes ont été améliorées en utilisant différentes architectures de réseaux tels que ResNet50, ResNet152 et Res2Net50 [Kang & al, 2022]. Bien qu'il soit difficile d'appliquer ces méthodes aux images, Faster-RCNN et YOLO (You Only Look Once) sont les méthodes de référence dans ce domaine.

#### La détection des objets sur les données 3D (nuage de points)

Il existe des techniques de détection d'objets en 3D, notamment la détection sur le nuage de points. Ces dernières années, plusieurs architectures de réseaux neuronaux ont été proposées pour la détection d'objets dans un nuage de points. Selon [Guo et al., 2020], il existe deux types de méthodes. Le premier type est basé sur des propositions régionales, où le modèle propose des régions possibles contenant des objets, extrait des caractéristiques par région et détermine l'étiquette de catégorie de chaque proposition. Le deuxième type est basé sur plusieurs vues, où les caractéristiques de propositions de différents angles de vue sont fusionnées pour obtenir des boîtes tournées en 3D.

#### Les défis et les limites des modèles de détection

La plupart des détecteurs d'objets utilisant l'apprentissage automatique et les algorithmes de Deep Learning ne parviennent pas à résoudre les défis couramment rencontrés, qui ont été résumés dans l'article de [Diwan & al., 2023] :

* La multi-échelle ou les modèles de détection sont moins performants pour les entrées ayant des échelles ou des résolutions différentes.
* Déséquilibre entre les classes à détecter et l'arrière-plan
* Détection d'objets relativement plus petits.
* Le manque de la diversité d’objets et d’instances
* Localisation inexacte lors des prédictions : les boîtes englobantes sont les approximations de la vérité terrain. Généralement, les pixels d'arrière-plan sont également inclus lors des prédictions, cela affecte la précision de l'algorithme. La plupart du temps, les erreurs de localisation sont soit dues à l'occupation de l'arrière-plan dans les prédictions et à la détection d'objets similaires

### La segmentation

La segmentation consiste à diviser une image ou un nuage de points en plusieurs parties, appelées segments, selon leur contenu sémantique. Dans le domaine de l'apprentissage profond, il existe deux types principaux de segmentation : la segmentation d'instances et la segmentation sémantique.

#### La segmentation d'instances

La segmentation d'instances a pour objectif de séparer chaque instance d'un objet en identifiant chaque pixel ou point dans le nuage de points qui appartient à une instance spécifique. Ainsi, même si un objet apparaît plusieurs fois dans l'image ou le nuage de points, chaque instance doit être distinguée individuellement.

Une image contenant Caractère coloré, violet, Dessin d’enfant, art

Description générée automatiquement

(a)

(b)

Figure 5 : Exemples de segmentation d'instances. (a) sur un nuage de points [Yan & al, 2020], (b) sur une image

Dans la Figure 5, l'image (a) illustre la segmentation des arbres, où chaque nuage de points faisant partie d'un arbre est attribué à une classe distincte. De manière similaire, l'image (b) montre que chaque bande dans le passage piéton est associée à une classe spécifique.

* **La segmentation d'instances sur le nuage de points**

Les méthodes de segmentation d'instance se répartissent en deux catégories : les méthodes basées sur des propositions, qui utilisent des boîtes englobantes, et les méthodes basées sur le regroupement, qui se focalisent sur la proximité spatiale. Selon [Zhao & al, 2023], Les méthodes de regroupement dominent en utilisant des décalages de distance pour former des groupes de points. Cependant, ces méthodes tendent à provoquer la sous-segmentation et ne permettent pas toujours un raffinement précis des contours d'instances, et ils proposent une approche innovante en combinant des scènes locales pour chaque instance avec des informations globales. Cela permet d'affiner les contours d'instances et de surmonter les limitations de sous-segmentation tout en supprimant la sur-segmentation.

Contrairement au principe de cette approche, l’article de [Vu et al, 2022] propose la méthode SoftGroup, qui permet d'utiliser les prédictions sémantiques pour moduler l'affinité entre les points voisins. Les prédictions sémantiques sont utilisées pour calculer une matrice d'affinité basée à la fois sur la distance spatiale et la proximité sémantique entre les points. En d'autres termes, les points qui sont spatialement proches et qui partagent des caractéristiques sémantiques similaires sont plus susceptibles d'être regroupés ensemble en tant qu'instance.

* **La segmentation d'instances sur le raster**

La segmentation d'instances est une technique fréquemment employée dans différents domaines tels que la segmentation d'images aériennes, la photogrammétrie terrestre ou encore l'orthophotographie pour extraire des objets individuels de l'image. La segmentation d'instances peut avoir de nombreuses applications, notamment dans les domaines de la robotique, de la conduite autonome ou encore de la surveillance. Elle peut également être utile pour la détection et le suivi d'objets en temps réel dans des scènes complexes. [Hafiz & Bhat, 2020]

Une image contenant ligne, capture d’écran, diagramme, Tracé

Description générée automatiquement

Figure 6 : Chronologie des techniques notables de segmentation d'instance [Hafiz & Bhat, 2020]

Selon la Figure 6, la segmentation d'instances a commencé avec l'utilisation du modèle RCNN qui consiste à utiliser des propositions de région pour améliorer un modèle CNN pré-entraîné tel que AlexNet. Ensuite, un ensemble de classificateurs SVM spécifiques à la classe est formé sur des fonctionnalités extraites du CNN. [Hafiz & Bhat, 2020]. Dans le même article, les auteurs expliquent comment cette méthode a évolué en raison de sa lenteur, pour arriver à TensorMask qui utilise la technique de fenêtre glissante pour générer des prédictions de boîtes englobantes.

L’état de l’art actuel est le modèle SAM (Segment Anything Model), où l'idée clé derrière SAM est de créer une carte d'activation qui identifie les régions potentielles d'objets, puis applique une méthode de regroupement de région pour identifier les candidats. Enfin, une étape de fusion basée sur des modèles graphiques raffine et segmente les régions pour produire des segments d'objets cohérents et précis.

#### La segmentation sémantique

La segmentation sémantique est une technique de traitement d'image qui consiste à regrouper les pixels dans l’image ou les point dans le nuage de points en différentes catégories correspondant à des classes d'objets ou de structures présentes dans l'image.

* **La segmentation sémantique sur le nuage de points**

Il est important de noter que la segmentation sémantique sur les nuages de points est une tâche complexe en raison de la densité variable des nuages de points, des artefacts de numérisation et parfois de l'absence d'informations de couleur. Par conséquent, il est souvent nécessaire d'utiliser des approches hybrides pour combiner différentes sources d'informations et obtenir une segmentation précise. En 2017, Qi et al. ont présenté le modèle PointNet, qui prend directement les nuages de points en entrée et en sortie. Avant cela, les architectures de modèles étaient basées sur les réseaux de convolution qui nécessitaient des formats de données d'entrée très réguliers. Par exemple, des recherches ont été faites sur la transformation de données en grille voxel 3D ou en images 2D, ou encore une grille polaire a été utilisée dans [Zhang & al, 2020]. Cela a rendu les données en sortie inutilement volumineuses selon Qi et al. Une autre technique développée par Shi & Rajkumar (2020) se base sur la représentation du nuage de points sous forme de graphe, traité par un GNN (graph neural network) [Wu & al, 2020].

Il est important de considérer la couleur dans la segmentation sémantique d'un nuage de points. Certains algorithmes peuvent être utilisés dans ce cas, tels que PointNet [Qi & al, 2017], qui considère la couleur comme une caractéristique cruciale [Ye & al, 2018]. La discontinuité de couleurs peut également être utilisée [Wu & al 2021], tout comme l'encodage en HSV [Pierdicca & al, 2020], ou l'utilisation de la couleur comme attribut dans PointNet [Balado & al, 2019]. Il convient de noter que la segmentation sémantique d'un nuage de points coloré peut être plus précise, mais que cela peut également rendre la tâche plus complexe en raison de la densité variable du nuage de points et des artefacts de numérisation.

* **La segmentation sémantique sur le raster**

La segmentation sémantique sur une image raster peut être réalisée à l'aide de différents algorithmes, tels que les réseaux de neurones convolutifs (CNN) ou les méthodes de clustering. Cette technique est largement utilisée en télédétection pour l'analyse d'images satellites ou aériennes, mais également dans d'autres domaines tels que la médecine pour l'analyse d'images médicales.

La première approche de segmentation sémantique est les réseaux entièrement de La méthode de segmentation sémantique Full Convolutional Network (FCN), a été proposée pour la première fois par [Long & al, 2015]. Contrairement aux réseaux neuronaux (RN) classiques, le FCN utilise uniquement des couches convolutives pour la segmentation. Cette approche permet de conserver les informations spatiales de l'image d'entrée tout au long du réseau et de produire une carte de segmentation qui a la même résolution que l'image d'entrée. La Figure 4 présente l'architecture du réseau, avec l'étape d'inférence où le sous-échantillonnage est effectué en réduisant la taille de l'image tout en conservant les informations les plus importantes. Cette étape est appelée le Pooling.

Une image contenant graphisme, capture d’écran, diagramme, conception

Description générée automatiquement

Figure 7 : Architecture du réseau FCN

En utilisant l'approche de réseau de neurones convolutifs (FCN), [Ronneberger & al, 2015] ont proposé l'architecture U-net qui a surpassé le FCN en termes de performances. L'architecture U-net utilise un suréchantillonnage (Upsampling) pour refléter la partie de Pooling et obtenir des sorties grossières similaires à l'image d'entrée, comme illustré dans la Figure 5.

Une image contenant capture d’écran, ligne, diagramme, texte

Description générée automatiquement

Figure 8 : Exemple de l’architecture du modèle U-net

Une autre approche est celle de SegNet, développée par [Badrinarayanan & al, 2017]. Elle est conçue pour être efficace pour la segmentation sémantique pixel par pixel, notamment dans les applications de compréhension de la scène routière, qui est le sujet principal de ce mémoire. Dans l'article, plusieurs résultats d'expérimentation sont présentés, montrant une meilleure précision de segmentation par rapport à plusieurs autres modèles.

L'approche DeepLab utilise un réseau de neurones convolutif profond dilaté, avec des modifications apportées à l'architecture pour relever trois défis. Le premier défi consiste à réduire le nombre de paramètres tout en améliorant les performances. Le deuxième défi est de gérer la présence d'objets multi-échelles, qui n'est pas très pertinent dans le cas des affleurants sur un corps de rue car la taille des objets ne varie pas, sauf si le modèle est entraîné sur des images de différentes résolutions spatiales. Le troisième défi est d'augmenter la capacité du modèle à s'adapter à l'invariance aux transformations spatiales (position sur le raster, orientation). Pour cela, comme expliqué dans l'article de Chen et al. (2017) et illustré dans la Figure 6, la segmentation est dilatée pour assurer la globalité de l'objet à segmenter. Les auteurs ont ensuite utilisé les champs de Markov conditionnels (CRF) pour capturer les dépendances et améliorer la précision spatiale en prenant en compte les détails fins.

Une image contenant texte

Description générée automatiquement

Figure 9 : performances du CRF pour l'affinement de la segmentation [Chen & al, 2017]

#### Les limites de la segmentation

La segmentation, tout comme la détection, est également sujette à des limites. Les deux approches se chevauchent à certains égards, notamment en termes de nécessité de disposer de données d'entraînement. De plus on trouve :

* Les erreurs de segmentation : les modèles de segmentation peuvent parfois mal segmenter les objets, soit en les sous-segmentant (en ne couvrant pas tout l'objet), soit en les sur-segmentant (en incluant des parties non pertinentes).
* La complexité des objets : la segmentation automatique peut rencontrer des difficultés à segmenter des objets complexes, tels que des objets avec des formes irrégulières ou des objets très similaires à leur environnement.
* Le temps de traitement : les modèles de segmentation automatique peuvent nécessiter beaucoup de temps de traitement, en particulier lorsqu'ils sont appliqués à de grandes quantités de données
* Les limites des algorithmes de Deep Learning : bien que les algorithmes de Deep Learning soient très performants dans de nombreux domaines, ils ont encore des limites en termes de compréhension du contexte et de prise en compte de l'expérience passée. Il est donc possible que des erreurs de segmentation surviennent dans des situations où le modèle n'a pas été entraîné.

Pour le dernier point, il existe d'autres solutions qui peuvent être appliquées, telles que l'utilisation de plusieurs algorithmes en même temps, la combinaison de la détection et de la segmentation ou la combinaison des deux types de segmentation.

### Combinaison des approches

Dans l'état de l'art actuel, les combinaisons entre les techniques de segmentation et de détection sont couramment utilisées dans les modèles pour améliorer les performances globales. Cependant, ces techniques peuvent également être utilisées de manière indépendante pour évaluer les performances en termes de temps et de précision.

#### Combinaison de la détection et la segmentation

Cette méthode consiste à détecter les objets d'intérêt en les encadrant avec des boîtes englobantes, puis de faire la segmentation à l'intérieur de ces boîtes pour économiser du temps et de l'entraînement. Dans son mémoire d'études, [Desbiolles, 2020] propose trois solutions dans ce contexte. La première utilise deux modèles (YOLO et U-net), le premier pour la détection, suivi d'un autre pour la segmentation. La deuxième utilise un seul modèle, Mask R-CNN [He, 2017], qui combine les deux en un seul réseau pour une segmentation d'instances. La troisième solution utilise un détecteur de caractéristiques à la place d'un modèle de segmentation. Dans le cas de la reconnaissance d'objets sur un corps de rue où plusieurs formes d'objets sont présentes, la troisième solution est très difficile à mettre en œuvre, tandis que la deuxième nécessite un jeu d'entraînement détourant toutes les instances, ce qui est moins intéressant dans le cadre de notre étude. Ainsi, la première solution est la plus pertinente pour notre cas.

#### Combinaison de la segmentation d’instances et la segmentation sémantique

Cette méthode peut faciliter la résolution des problèmes de détection d'objets complexes. Dans ce cadre, la segmentation panoptique (SP) est très connue car elle englobe à la fois les deux types de segmentation [Kirillov, 2019]. Il existe également plusieurs autres versions de ce type de segmentation telles que Unified PS, Fast PS, etc. Une autre solution consiste à utiliser deux modèles : un modèle discriminatif pour la segmentation, suivi d'un regroupement d'instances en classes à l'aide d'un modèle de diffusion texte-image. Cette solution dépasse la SP en raison de ses limites en termes de vocabulaire de taille finie pour décrire le monde réel. [Xu et al., 2023]

### Modèle d’apprentissage profond (Deep Learning Model)

Tous les modèles mentionnés précédemment ont la capacité d'apprendre des modèles complexes et de généraliser à partir des données d'apprentissage, ce qui les rend particulièrement adaptés pour les tâches de classification et de prédiction. Cependant, leur utilisation nécessite une grande quantité de données d'apprentissage pour obtenir de bons résultats [Naumcheva, 2021], ainsi que des ressources informatiques importantes pour l’entraînement.

* **Préparation de données**

Cette étape est cruciale car la qualité des données utilisées pour entraîner le modèle impacte directement sa précision et sa fiabilité. Dans le cas d'un corps de rue, les données utilisées sont généralement des orthophotos ou des images brutes collectées sur le terrain, ainsi que des annotations ou masques représentant la vérité terrain des objets à détecter. Les annotations doivent être disponibles sous forme de masques raster, soit par digitalisation soit en reprenant un plan topographique. Si une orthophoto est utilisée comme image, elle doit être subdivisée en petites imagettes car il est difficile, voire impossible, de l'utiliser telle quelle pour l'entraînement. La couche d'annotation doit également être découpée en superposition avec les imagettes, comme illustré dans la Figure 7. Dans le cas d'utilisation d'images brutes, une projection des annotations issues du plan topographique sur les images est nécessaire, ou une nouvelle digitalisation doit être effectuée.

Parmi les étapes de préparation des données, le nettoyage est essentiel pour éliminer les données inutiles ou incorrectes. Une fois les données nettoyées et préparées, elles sont divisées en deux ensembles distincts : l'ensemble d'entraînement et l'ensemble de validation. L'ensemble d'entraînement est utilisé pour ajuster les poids et les biais du modèle afin de minimiser l'erreur de prédiction, tandis que l'ensemble de validation est utilisé pour évaluer la performance du modèle pendant l'entraînement et pour ajuster les paramètres de régularisation afin d'éviter le surapprentissage. Pour améliorer la robustesse du modèle et éviter le surapprentissage, il est possible d'appliquer des techniques d'augmentation de données (data augmentation) telles que la rotation, le recadrage, le zoom ou encore le changement de luminosité.

Une image contenant noir, obscurité, lune

Description générée automatiquementUne image contenant texte, croquis, art

Description générée automatiquementUne image contenant texte, croquis, capture d’écran, art

Description générée automatiquement

Figure 10 : imagette et masque, découpés et superposés

* **Création du modèle**

Cette étape consiste à sélectionner l'architecture du réseau de neurones qui sera utilisée pour traiter les données d'entraînement. Plusieurs types d'architectures de réseaux de neurones existent, chacune ayant ses propres avantages et inconvénients en fonction du type de tâche à résoudre, comme mentionné dans la partie 1.4.2.2. Les architectures les plus couramment utilisées pour la détection d'objets et la segmentation d'instances et sémantique sont les réseaux de neurones convolutifs (CNN)

Une fois l'architecture choisie, il est temps de la mettre en œuvre en utilisant une bibliothèque de Deep Learning comme TensorFlow, PyTorch, etc. TensorFlow et PyTorch sont les deux frameworks les plus populaires, mais leur choix dépend de la philosophie de création qui les influence, ce qui impacte leur utilisation et les résultats obtenus. Selon Dai et al. (2022), TensorFlow est plus prometteur dans le contexte de l'industrie, tandis que PyTorch est plus attrayant dans le milieu universitaire. Le Tableau 1 présente les principales différences entre les deux workflows.

|  |  |
| --- | --- |
| Pytorch​ | TensorFlow​ |
| Graphe dynamique (séquentiel)​ | Graphe statique (Compilé)​ |
| Moins rapide​ | Plus rapide​ |
| Pas d’API (haut-niveau)​ | API Keras​ |
| Pytorch Lightning (préparation des données d’entraînement, parallélisation des GPU, etc) | Paramétrassions manuelles |

Tableau 1 : différence entre Pytorch et TensorFlow

Dans l’article de [Dai & al, 2022], une comparaison de performances est présentée en utilisant les deux bibliothèques pour différents modèles de réseaux de neurones convolutifs (CNN), comme on peut le constater dans leTableau 2. Les résultats indiquent une performance visiblement meilleure pour TensorFlow que pour PyTorch.

Une image contenant texte, Police, nombre, ligne

Description générée automatiquement

Tableau 2 : Vitesse d'entraînement globale sur TensorFlow et PyTorch. [Dai & al, 2022].

* **L’entraînement du modèle**

Après avoir choisi le modèle, l'étape de l'entraînement suit. Elle consiste à ajuster les paramètres du modèle pour qu'il puisse apprendre à partir des données d'entraînement. Cette étape passe par l'initialisation des poids, suivie par un passage en avant, une rétropropagation, une optimisation et une validation. L'initialisation des poids consiste à attribuer des valeurs initiales aux paramètres du modèle avant le début de l'entraînement. Le passage en avant, ou l'inférence, implique l'application d'une série d'opérations mathématiques à travers les couches du modèle. Chaque couche traite les informations fournies par la couche précédente, et finalement, la sortie est générée par la dernière couche. Cette sortie est ensuite comparée à la sortie attendue pour calculer l'erreur

La rétropropagation utilise l'erreur entre la prédiction et la sortie attendue (vérité terrain qui sont les annotations) en propageant en arrière à travers le réseau, couche par couche, pour ajuster les poids de chaque neurone. L'étape de l'optimisation permet d'améliorer la performance du modèle en minimisant la fonction de perte (Loss). On peut utiliser des algorithmes d'optimisation tels que la descente de gradient qui calcule le gradient de la Loss par rapport aux poids du modèle, puis les ajuster dans la direction opposée au gradient, de sorte que la fonction de perte diminue.

L'étape de la validation permet d'évaluer les performances du modèle pendant l'entraînement en le testant sur des données qui n'ont pas été utilisées. Cette étape permet de détecter des problèmes tels que le surapprentissage, l'erreur de mise en œuvre du modèle, etc

* **Test du modèle**

Une fois que le modèle a terminé son entraînement sur les données d'apprentissage, il est temps de passer à l'étape de test pour évaluer ses performances sur des données qu'il n'a jamais rencontrées auparavant. Cette étape est cruciale pour mesurer l'efficacité du modèle en termes de précision, de rappel, de F1-score et d'autres mesures de qualité de prédiction. Il est important de noter que cette étape est similaire à l'étape de validation pendant l'entraînement, mais elle se déroule après la fin de l'entraînement.

## Solutions existantes

### Solutions logicielles

Ces logiciels ont pour fonction d'analyser des images prises depuis des avions ou des satellites pour identifier des éléments clés tels que des bâtiments, des routes, des véhicules, des arbres, et autres. Leur utilisation se répand de plus en plus dans différents domaines.

* ENVI : ce logiciel de télédétection, développé par la société américaine Harris Geospatial, souvent utilisé pour segmenter les images satellitaires, mais peut également utiliser pour la segmentation d’autres types d’images
* QGIS : sur ce logiciel, il est possible d’installer des toolbox pour différentes utilisations. Orfeo Toolbox permet de faire de la classification et de la segmentation de l’image sur QGIS [De Luca & al, 2019]
* ArcGIS : ce logiciel propose également des fonctionnalités de segmentation d'image, notamment avec le module Image Analyst qui permet de segmenter une image en utilisant des algorithmes de classification automatique basés sur des seuils, des signatures spectrales, des réseaux de neurones, etc. [Hoeser and Kuenzer,2020].

Il existe plusieurs autres logiciels de segmentation d'images, mais la plupart d'entre eux sont conçus pour des données spécifiques et il peut être difficile de les personnaliser pour répondre à des besoins particuliers. Par conséquent, leur utilisation peut être limitée dans certaines applications. Toutefois, dans les contextes où ils correspondent parfaitement aux types de données traitées, ces logiciels peuvent être très utiles.

### Solution en ligne

Il existe aujourd'hui plusieurs solutions en ligne pour la segmentation automatique d'images. Ces solutions sont généralement basées sur l'utilisation de réseaux de neurones profonds et sont de plus en plus populaires en raison de leur accessibilité et de leur facilité d'utilisation. Elles offrent une alternative pratique et rapide pour la segmentation automatique d'images, sans nécessiter l'installation d'un logiciel sur un ordinateur local.

#### Solution Open-Source

* **Segment anything Model (SAM)**

SAM est un modèle récent de segmentation d’instance d'image qui se concentre sur la segmentation de tout objet dans une image, sans limitation de forme ou de taille prédéfinie. Cette méthode utilise des réseaux de neurones profonds pour segmenter les images, offrant ainsi une solution flexible et générique pour la segmentation. Cette approche "Segment Anything" est utile pour diverses applications, telles que la segmentation d'objets complexes et non conventionnels dans les images médicales, la segmentation de bâtiments et d'infrastructures dans les images satellitaires, et la segmentation d'objets de différentes formes et tailles dans d'autres types d'images. Ce modèle a été récemment ajouté à ArcGIS en tant qu'extension Image, disponible dans certaines versions de ce logiciel. [Esri, 2023]

* **MMSegmentation**

Cet outil est nommé MMSegmentation et il s'agit d'un logiciel open-source développé par Megvii Research qui utilise les dernières technologies en matière de vision par ordinateur et d'apprentissage en profondeur pour fournir une segmentation rapide et précise des images. Il est disponible en libre accès et peut être utilisé pour une variété d'applications telles que la segmentation sémantique, la segmentation d'instances et la détection d'objets. MMSegmentation est programmé en Python et utilise la bibliothèque PyTorch pour l'apprentissage en profondeur.

* **Detectron2**

Detectron2 est une bibliothèque open source de vision par ordinateur développée par Facebook AI Research (FAIR). Cette bibliothèque permet de créer et d'entraîner des modèles avancés de détection d'objets en utilisant des réseaux de neurones de convolution (CNN).

Detectron2 est principalement orienté vers la détection d'objets, la segmentation sémantique et l'estimation de la pose 2D/3D. Cette bibliothèque est basée sur PyTorch et propose des implémentations de nombreux modèles d'apprentissage en profondeur pour ces tâches. Contrairement à MMSegmentation, qui est spécifiquement conçu pour la segmentation sémantique des images, Detectron2 couvre un large éventail de tâches de vision par ordinateur.

#### Solutions payantes

Quant aux solutions payantes, il en existe une variété qui offrent des fonctionnalités de segmentation et de classification pour les nuages de points, les orthophotos, ainsi que pour les images géographiques et non géographiques.

* Solution pour les nuages de points : Il existe plusieurs solutions pour la segmentation des nuages de points, dont The Cross Product (TCP) qui se spécialise dans la classification et la vectorisation automatique des nuages de points dans les domaines autoroutier et ferroviaire. Pointly est une autre solution qui utilise l'apprentissage profond pour générer automatiquement des classifications de nuages de points dans différents domaines, tels que les autoroutes, les bâtiments, les forêts et les routes. Elle est capable de classifier les plaques de rue, les lampadaires, les regards, les poteaux, etc.
* Solution appliquée sur les images : Après avoir contacté plusieurs entreprises proposant la détection ou la segmentation d'objets sur des images, la plupart d'entre elles ont proposé des solutions personnalisées adaptées à nos besoins. Certaines de ces solutions sont spécialisées dans la segmentation et la détection d'autres types de données, comme Celantur, qui travaille principalement sur le floutage de personnes, de plaques d'immatriculation et de véhicules, et a présenté un exemple de segmentation de fissures de bitume sur des images comme montrer dans la Figure 8. Cependant, nous avons constaté une lenteur dans la prise de décision quant à l'adoption de ces solutions, car elles nécessitent l'utilisation de nos propres données pour entraîner les modèles, afin de les adapter à notre contexte et à nos besoins spécifiques.

**Une image contenant plein air, route, asphalte, Bitume

Description générée automatiquement**

Figure 11 : segmentation des fissures sur le sol par Celantur

## De la segmentation vers la vectorisation

Pour passer de la segmentation à la vectorisation, différentes méthodes peuvent être utilisées pour convertir les résultats de segmentation en données vectorielles. Cette étape est également appelée "vectorisation à partir d'une image matricielle" ou "vectorisation à partir de nuages de points". Dans cette étude, la méthode choisie est la segmentation sur le raster, ce qui implique d'obtenir des résultats raster de la segmentation. Le processus consiste alors à convertir le raster en vecteur en utilisant des méthodes de vectorisation, telles que la vectorisation par seuillage, la vectorisation par détection de contours et la vectorisation par classification. Ces méthodes ont été développées pour résoudre les problèmes de la vectorisation automatique, notamment la qualité, la précision et la rapidité. Le résultat de la segmentation est un raster avec des formes classifiées par des valeurs, ce qui facilite la vectorisation en utilisant la classification par classe. Pour cela, plusieurs bibliothèques peuvent être utilisées, telles que GDAL, Shapely pour créer des géométries à partir de points, et OpenCV, qui peut être utilisée pour extraire des contours et des formes à partir de données rasters segmentés. Cette vectorisation conserve la rugosité des contours des formes extraites du raster. Pour les lisser, des méthodes de lissage des contours sont proposées par des bibliothèques telles que scikit-image et OpenCV, afin de retrouver des formes régulières représentant les objets détectés.