



Technote

cerfo

FORMATION
ACCOMPAGNEMENT
RECHERCHE
EN FORESTÉRIE

NOTE TECHNIQUE NO 2023-06

15 novembre 2023



Une plateforme de test pour la cartographie automatique du réseau routier du Canada

- 2 → Contexte
- 3 → Données utilisées
- 5 → Technique développées
- 10 → Résultats et analyse
- 15 → Conclusion

CERFO.QC.CA

ÉQUIPE DE RÉALISATION
Marc-Antoine Genest, M. Sc.
Xavier Gervais, M. Sc.
Batistin Bour, M. Sc.
Charles Marseille, M. Sc.
Mathieu Varin, M. Sc.



Contexte

Une cartographie du réseau routier du Canada à jour est importante, notamment pour l'accès au territoire de façon sécuritaire et la planification et l'optimisation de trajets. Cependant, elle est présentement réalisée en photo-interprétation manuelle, par un grand nombre d'acteurs différents, elle peut donc comprendre des biais ou des erreurs. De plus, le Canada étant un vaste territoire, l'actualisation périodique du réseau routier est difficile à réaliser et nécessite des efforts conséquents.

Les récents développements dans le domaine de l'apprentissage profond et la disponibilité grandissante d'images satellitaires à haute résolution spatiale ouvrent la porte vers des solutions intéressantes pour cartographier automatiquement un territoire. En effet, différentes approches utilisant la segmentation d'images [1] ou des encodages de graphes [2] ont déjà été développées et fournissent des résultats prometteurs.

But et objectifs

Les dernières avancées de l'apprentissage profond ont motivé le Centre canadien de cartographie et d'observation de la Terre (CCCOT) et le Centre d'enseignement et de recherche en foresterie de Sainte-Foy inc. (CERFO) à développer une plateforme permettant l'entraînement et l'analyse de différentes techniques de détection de réseaux routiers, afin de les comparer différentes techniques de détection.

Plus précisément, deux techniques de détection par apprentissage profond et une technique de post-traitement ont été développées, permettant de comparer quatre méthodes :

1. →Segmentation binaire par réseau neuronal convolutif (CNN);
2. →Segmentation binaire par CNN, suivie d'une technique de post-traitement;
3. →Détection par encodage de graphes dans un tenseur (GTE);
4. →Détection par GTE suivi d'une technique de post-traitement.

De plus, deux métriques spécifiques à l'analyse réseau ont été implémentées, afin de permettre une meilleure analyse de la performance du réseau extrait. Finalement, les développements ont été publiés sur un répertoire public [3], afin de permettre la poursuite des développements, en collaboration avec d'autres équipes.

Données utilisées et prétraitements

Les images utilisées pour le développement des algorithmes proviennent du satellite WorldView-2, soit des images à quatre bandes (bleu-vert-rouge-proche infrarouge), acquises entre 2010 et 2021. Des prétraitements ont été appliqués par l'équipe du CCCOT, afin d'obtenir une réflectance avec des valeurs entre 0 et 255 (8 bits) et une résolution au sol de 50 cm. Un exemple d'une image est montré à la figure 1.



FIGURE 1. Exemple d'une image satellitaire WordView-2 acquise à Terre-Neuve

Les zones photo-interprétées totalisent 12 684 km² d'images satellitaires et 28 060 km de voies routières partout au Canada. La figure 2 présente les distributions des superficies et des longueurs des routes photo-interprétées pour les différentes provinces représentées dans la base de données.

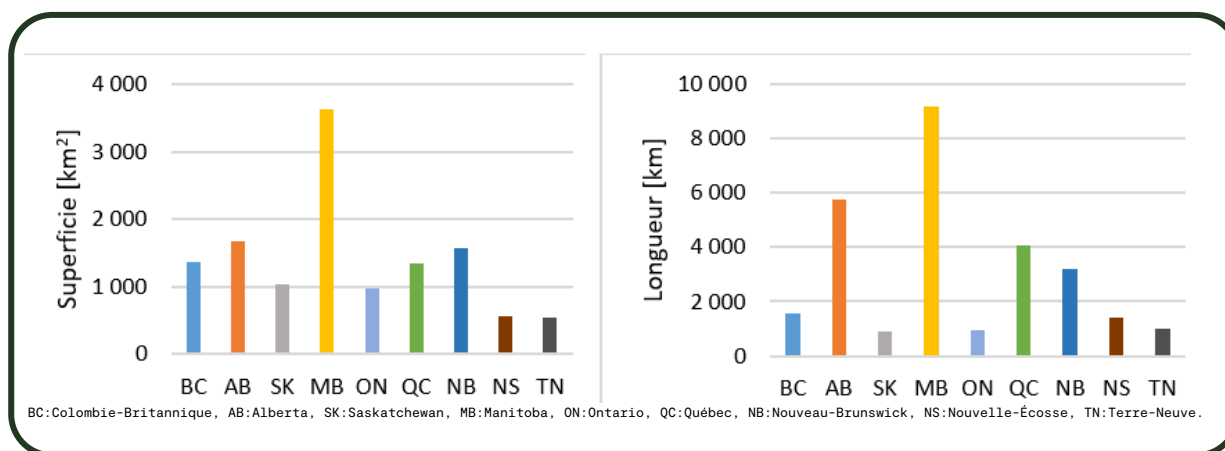


FIGURE 2. Histogrammes des superficies et des longueurs totales de route photo-interprétées par province.

Pour créer les jeux d'entraînement et de validation, les images photo-interprétées ont d'abord été découpées en imquettes de 512 par 512 pixels, telles qu'illustrées à la figure 3.



FIGURE 3. Exemple d'une imquette de 512 par 512 pixels et du masque de route photo-interprété (en rouge).

Ensuite, ces imquettes ont été standardisées, c'est-à-dire traitées de sorte que l'ensemble du jeu de données possède une moyenne de zéro et un écart-type d'un (1). Finalement, elles ont été séparées en jeux d'entraînement et de validation, avec un ratio de 80 % et 20 % respectivement, en s'assurant que les imquettes provenant d'une même image satellitaire ne se retrouvent que dans un seul des deux jeux. Cette préparation permet de créer 89 563 imquettes d'entraînement et 19 704 pour la validation. À noter que 18 zones de tests ont été gardées de côté afin de réaliser les analyses finales pour la comparaison des techniques d'extraction de réseau routier.

Techniques développées

Dans le cadre de ce projet, deux techniques par apprentissage profond ont été comparées, soit la segmentation binaire et la GTE. De plus, un post-traitement a été implémenté et appliqué sur les résultats des deux techniques d'extraction, afin de corriger les erreurs attendues et augmenter la qualité des réseaux détectés.

Dans le but d'avoir une comparaison indépendante de l'algorithme d'apprentissage profond et de pouvoir faire un meilleur parallèle entre les méthodes, l'architecture de l'algorithme et la majorité de ses hyper paramètres ont été gardés constants. En effet, seule la fonction de perte et le taux d'apprentissage ont été adaptés entre les deux techniques. L'architecture choisie, appelée R2AUnet [4], est une architecture couramment utilisée, fournissant généralement des résultats satisfaisants, d'un point de vue opérationnel, en segmentation d'images (Figure 4).

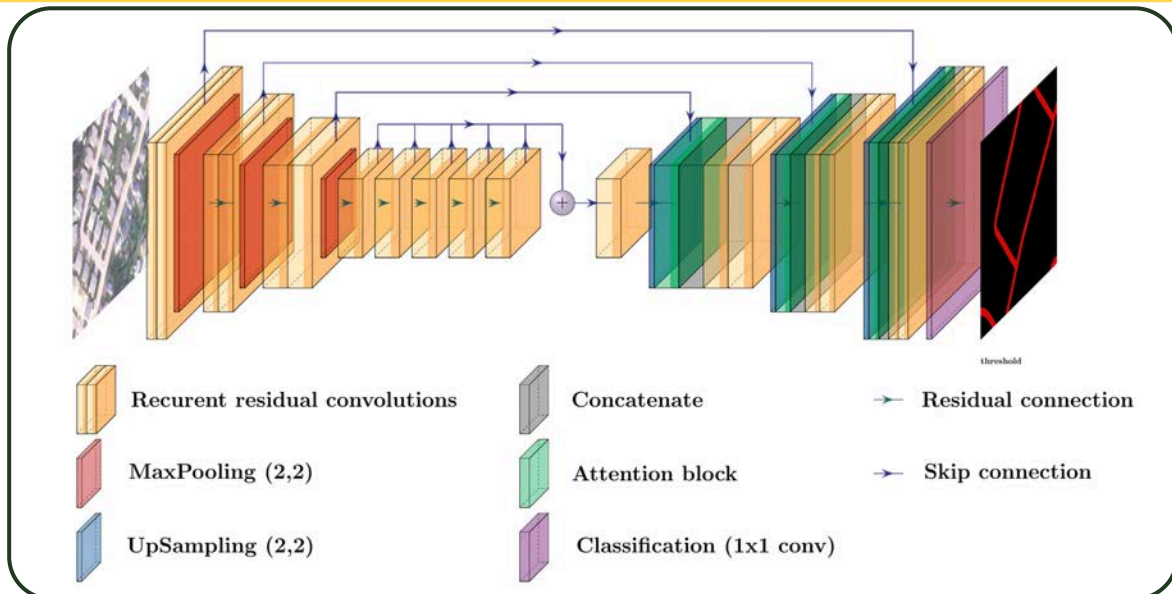


FIGURE 4. Schéma de l'algorithme R2AUnet utilisé pour le projet.

Segmentation binaire

La segmentation binaire « classique » vise à entraîner un algorithme et à prédire une valeur de 1 s'il y a présence de route et 0, dans le cas contraire. La donnée produite par l'algorithme est une matrice de probabilité de présence de route, à laquelle est appliquée une valeur seuil, pour créer un masque binaire de présence de route. La fonction de perte, utilisée pour l'entraînement de l'algorithme, est l'addition de la « dice loss » et de la « focal Tversky loss ».

$$\mathcal{L}_{Dice} = 1 - \frac{2VP}{2VP + FP + FN + \epsilon}$$

$$\mathcal{L}_{Focal\ Tversky} = \left(1 - \frac{VP}{VP + \alpha FN + (1 - \alpha)FP + \epsilon}\right)^{\frac{1}{\gamma}}$$

Où α et γ sont des hyper paramètres permettant de pénaliser davantage les faux-négatifs, fixés à 0,7 et 4/3 respectivement. VP : vrai-positif, FP : faux-positif, FN : faux-négatif et ϵ une petite constante empêchant les divisions par 0.

Cette technique traditionnelle a comme avantages d'avoir été testée sur plusieurs jeux de données et a donc fourni des pistes d'amélioration et d'optimisation. Il est également plus facile de trouver des outils disponibles en apprentissage profond, pour améliorer la convergence des algorithmes, telle que les fonctions d'augmentation de données. Finalement, un masque binaire est peu volumineux et facilite le stockage des jeux de données, des prédictions, il peut même accélérer le temps d'entraînement des algorithmes.

En revanche, un réseau routier n'est pas nécessairement représentable en deux dimensions uniquement, avec, par exemple, la présence de ponts, traverses et viaducs. Cette technique possède donc une limite théorique pour obtenir une extraction topologiquement fiable d'un réseau routier complexe.

Détection par encodage de graphes

Un graphe est une structure mathématique permettant d'encoder des objets et leurs relations grâce à deux composants : les nœuds et les liens. Les nœuds représentent des objets possédant des caractéristiques définies par le créateur du graphe, tandis que les liens représentent les relations qu'il peut y avoir entre les objets. Dans le cas d'un réseau routier, un nœud est par exemple une intersection, possédant une coordonnée géographique et un lien représente une route, qui relie deux intersections, soit une série de coordonnées géographiques créant une ligne.

Dans le but de pouvoir utiliser les réseaux neuronaux convolutifs, comme le R2AUnet, à prédire ce format de donnée, le graphe est encodé sous forme de tenseur pour créer un GTE. Cet encodage se réalise en trois étapes :

1. → **Injection de nœuds dans le graphe** : Des nœuds ont été ajoutés dans le graphe, afin d'avoir une distance minimale de 20 m entre chacun d'entre eux. Cette stratégie permet d'abord d'augmenter le nombre de nœuds dans le GTE. De plus, en prenant par hypothèse qu'une rue est linéaire sur 20 m, les liens n'ont pas besoin d'être encodés et seront interpolés par des lignes droites entre les nœuds lors du décodage;
2. → **Encodage des nœuds** : Un masque de nœuds a été créé, où chaque nœud est représenté par un carré de trois par trois pixels ayant la valeur de 1;
3. → **Encodage des angles vers les nœuds voisins** : Le but de cet encodage est d'indiquer la direction à suivre pour connecter un nœud à un autre nœud. Pour ce faire, un total de 18 canaux est nécessaire, réparti dans six groupes de trois canaux. Ces groupes représentent chacun un ensemble de 60° d'angle, soit les ensembles 0 à 60, 60 à 120, 120 à 180, 180 à 240, 240 à 300 et 300 à 360. Par la suite, l'angle entre un nœud et son voisin est calculé et encodé dans un des six groupes en fonction de la valeur de l'angle. Cet encodage est réalisé grâce à trois valeurs, soit une probabilité d'être dans le groupe (1 si dans le groupe, 0 sinon), la position relative de la composante « x » et celle de la composante « y ». Cet encodage est réalisé pour chaque voisin de chaque nœud, à la position du nœud à encoder.

Comme montré à la figure 5, le GTE requiert donc un total de 19 canaux pour représenter un graphe sous forme tensorielle. La fonction de perte utilisée pour l'optimisation de l'algorithme est l'addition de l'entropie croisée binaire sur les canaux de probabilité (nœuds et voisins) et de l'erreur quadratique moyenne sur les canaux de position relative des composantes « x » et « y ».

L'avantage que possède cette technique d'extraction de réseau routier est de pouvoir séparer deux routes qui se superposent, tel que les viaducs. Puisque les nœuds voisins sont encodés, il est possible de dissocier deux tronçons de route qui se superposent afin de ne pas les connecter ensemble, lors de la reconstitution du graphe. De plus, cette reconstitution devient une simple recherche de nœuds voisins et rend robuste la méthode de détection, aux obstructions entre les nœuds, comme les arbres, ou les ombres des bâtiments.

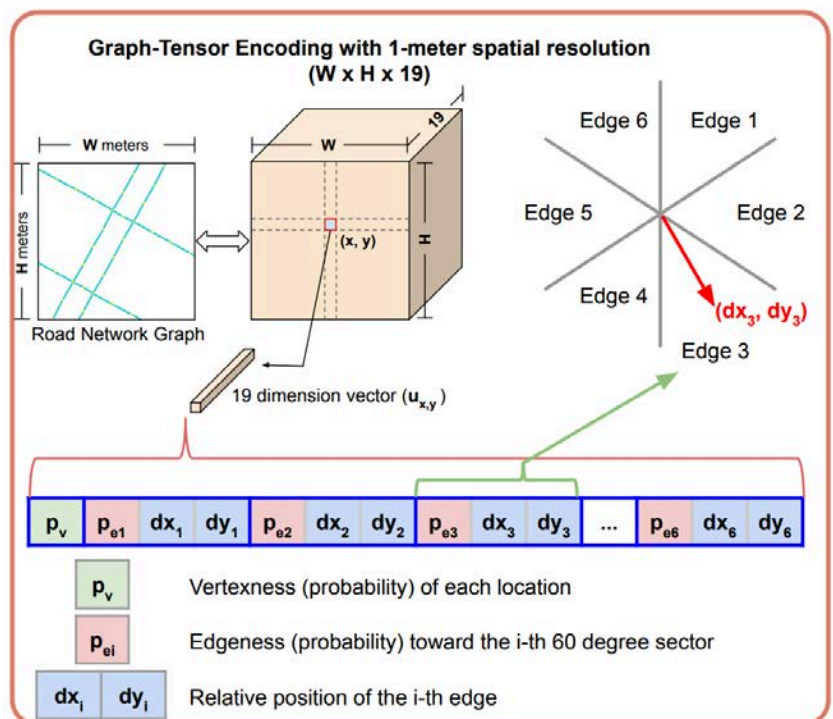


FIGURE 5. Schéma de l'encodage d'un graphe et tenseur (extrait de [2], page 5).

Cependant, puisque les directions vers les nœuds voisins sont encodées dans des canaux spécifiques, les fonctions traditionnelles d'augmentation de données de rotation et de miroir des images ne sont pas implémentées encore. En effet, elles nécessitent de modifier l'ordre des canaux et les valeurs des positions relatives, contrairement à un masque où il suffit d'effectuer la même rotation qu'à l'image. De plus, le stockage du jeu de données prend environ cinq fois plus de volume, dû aux 19 canaux d'encodage, ce qui peut rapidement devenir un enjeu, notamment au niveau de la lecture des images, lors de l'entraînement, qui devient le goulot d'étranglement du processus.

Technique de post-traitement

L'extraction de réseau routier sur imagerie visible grâce à l'apprentissage profond comporte ses limites et des erreurs étaient attendues. Les algorithmes de vision par ordinateur sont très vulnérables aux perturbations contenues dans les images [5] et peuvent facilement produire des résultats erronés. Il est donc pressenti que des erreurs d'omission seront faites en présence d'un pont de structure ou de matériaux différents de ceux présents dans les données d'entraînement ou lorsque la route est obstruée ou dans l'ombre d'un objet par exemple. De plus, des erreurs de commission sont attendues, par exemple, lorsqu'il y a présence de sentiers asphaltés ou de pistes cyclables. Des exemples de ces « perturbations » sont illustrés à la figure 6.



FIGURE 6. Exemple de structure ou d'ombres pouvant nuire à la détection des routes.

Une technique de post-traitement a été implémentée, afin de reconnecter les routes qui auraient été coupées par ce genre d'obstruction et pour supprimer les segments en trop, exécutée en cinq étapes (Figure 7) :

1. → Extraction de la centre-ligne et des extrémités du réseau détecté;
2. → Reconnexion des extrémités dont la distance entre elles va jusqu'à 100 m, trouvée par chemin de moindre coût sur la prédiction;
3. → Reconnexion des extrémités qui sont à 50 m et moins du réseau déjà détecté, trouvé par chemin de moindre coût sur la prédiction;
4. → Interpolation des largeurs des nouvelles connexions afin qu'elles aient une largeur similaire aux segments qui y sont connectés;
5. → Suppression des segments ayant une longueur inférieure à 1 km permettant d'éviter certaines erreurs de commission.

Les valeurs choisies dans les différentes étapes sont celles qui optimisent les performances d'extraction pour les modèles entraînés. Elles sont toutefois des variables dans le post-traitement et peuvent être modifiées au besoin.

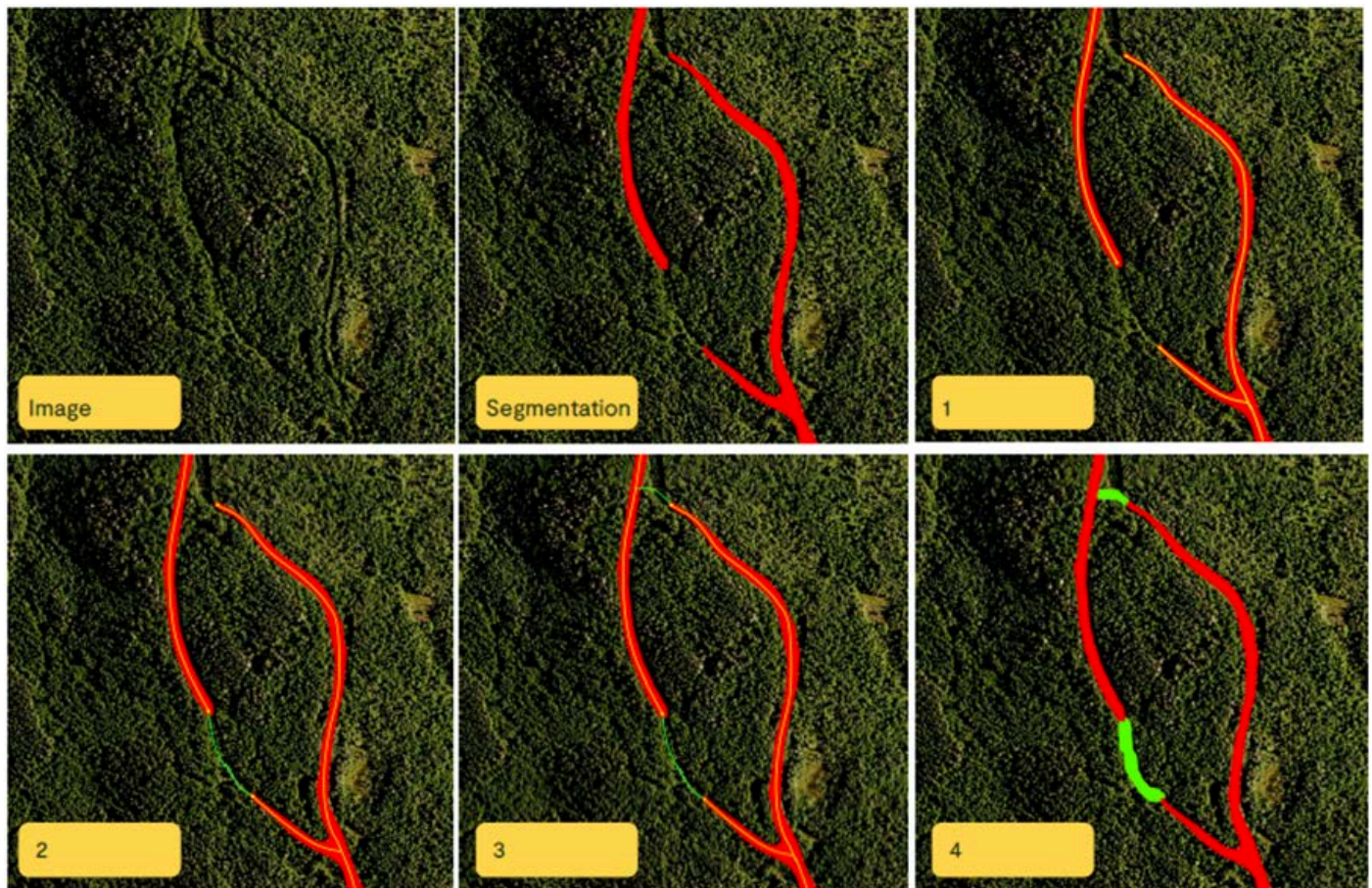


FIGURE 7. Schéma séquentiel des étapes de post-traitement d'une segmentation de réseau.

Métriques pour l'analyse des réseaux

Les métriques conventionnellement utilisées en segmentation d'images sont utiles pour évaluer les performances d'un algorithme à extraire des pixels d'une image, par exemple pour un réseau. Cependant, aucune d'entre elles ne permettent d'adresser des questions topologiques : est-ce que les point A et B sont connectés, et est-ce que l'itinéraire pour les rejoindre change entre la vérité terrain et la prédiction? C'est pourquoi les métriques APLS et TOPO [6], en plus d'autres métriques comme la précision et le f1-score, ont été calculées pour évaluer les méthodes.

APLS

APLS signifie « average path length similarity » — similitude des longueurs moyennes des trajets. Cette métrique calcule la différence de distance parcourue pour se rendre d'un point A à un point B entre la vérité terrain et le graphe prédit. Cette opération a été réalisée pour chaque trajet potentiel dans les graphes et a été moyennée afin d'obtenir l'APLS. Il s'agit donc d'un indicateur pour informer si des trajets et itinéraires peuvent être estimés directement à partir de l'inférence sur imagerie satellitaire (Figure 8).

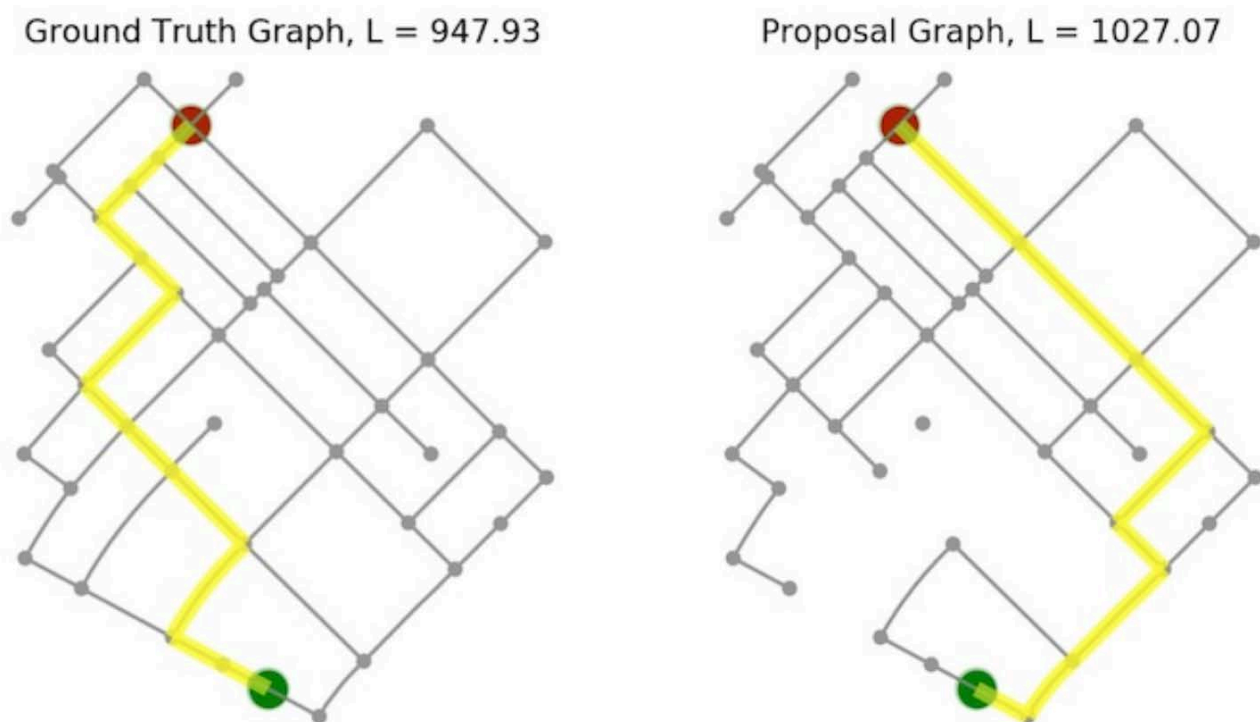


FIGURE 8. Exemple de comparaison de trajet entre un graphe provenant d'une vérité terrain et un graphe proposé (extrait de [6], page 3).

TOPO

Cette métrique permet d'évaluer la géométrie des réseaux dans une zone locale, généralement 300 mètres, et permet d'analyser si le réseau détecté tombe dans une zone tampon autour de la vérité terrain. Cela en fait donc une bonne mesure de la qualité locale du réseau et permet une flexibilité sur la position de la centre-ligne.

Résultats et analyse

Cette section décrit les méthodes d'extraction de réseau routier qui ont été comparées et analysées selon les métriques définies plus haut. Quatre types d'évaluation ont été réalisées afin de les évaluer : 1) l'impact du post-traitement sur les prédictions, 2) une analyse globale des performances sur le territoire et 3) une analyse locale selon le type de route et de milieu et 4) une analyse visuelle des réseaux extraits. Ces analyses ont été réalisées sur les 18 images tests en calculant les métriques par zone de photo-interprétation.

Impact du post-traitement

Afin d'évaluer l'impact du post-traitement, les segmentations binaires et GTE ont été comparées avant et après l'application du post-traitement, et ce, pour chacune des 8 provinces (figure 9).

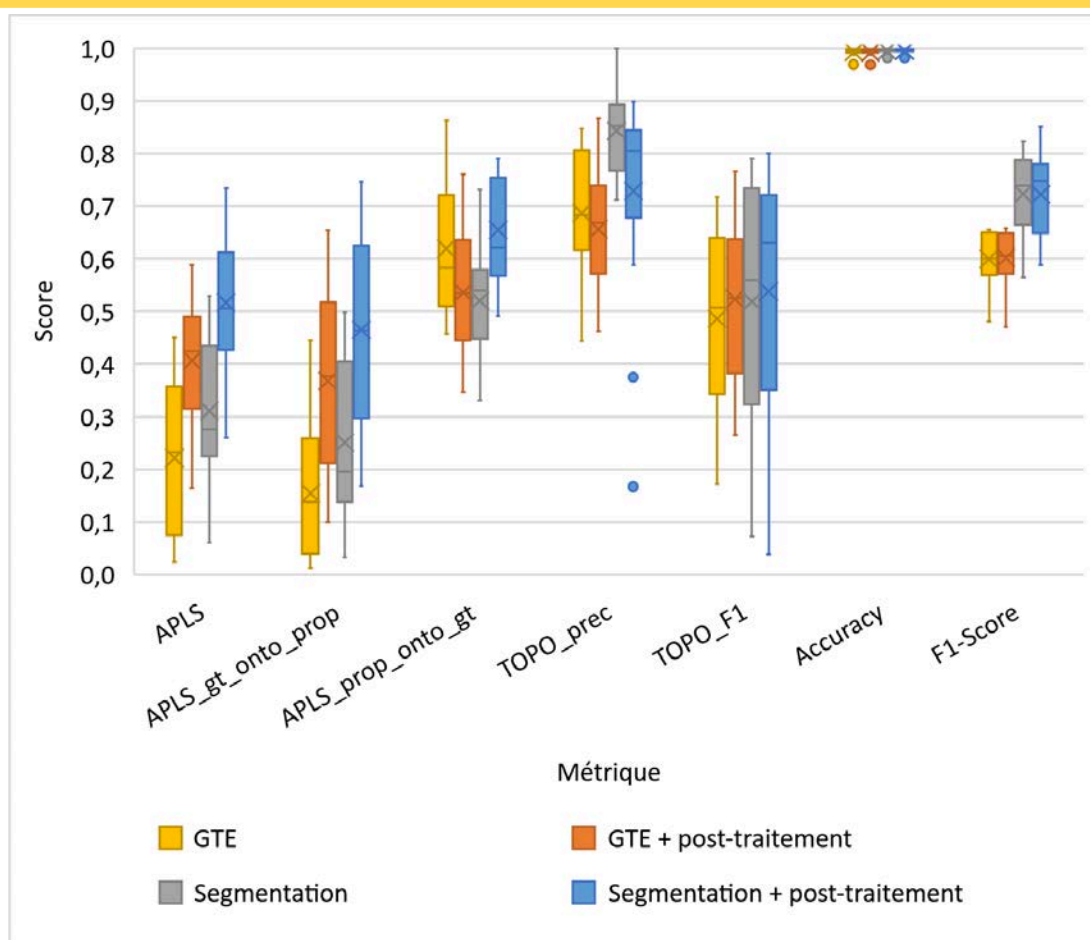


FIGURE 9. Comparaison des métriques pour les quatre méthodes d'extraction de réseaux routiers.

La première observation est l'augmentation des valeurs de l'APLS, lors de l'application du post-traitement, autant pour le GTE, que pour la segmentation. Cette augmentation de l'APLS est due en grande partie à la composante de vérité terrain sur la proposition (APLS_gt_onto_prop). Ce comportement indique une meilleure connectivité du réseau, suite à l'application du post-traitement et montre que peu de commission était présente, avant d'être corrigée.

Deuxièmement, les f1-scores, autant par pixel, que celui de la TOPO, restent stables après le post-traitement pour les deux techniques. Ceci montre que le post-traitement peut corriger des erreurs d'omission telles que des liens manquants, mais peut aussi supprimer de bons segments de routes qui sont trop loin du réseau principal pour pouvoir être reliés.

Analyse interprovincine

Les résultats ont également été compilés par province afin de comparer les performances dans l'ensemble du Canada (Figures 10 et 11).

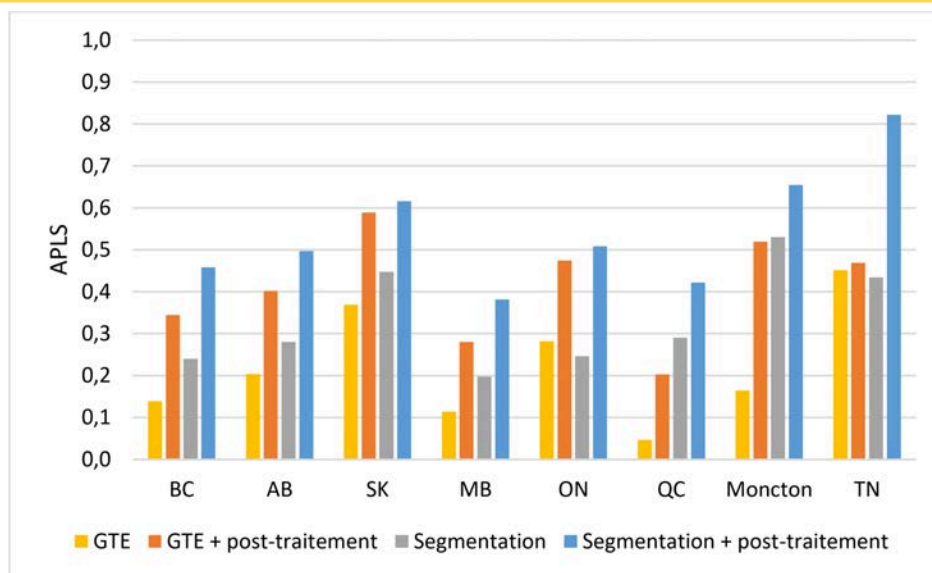


FIGURE 10. Comparaison de l'APLS des quatre méthodes sur les huit provinces-tests.

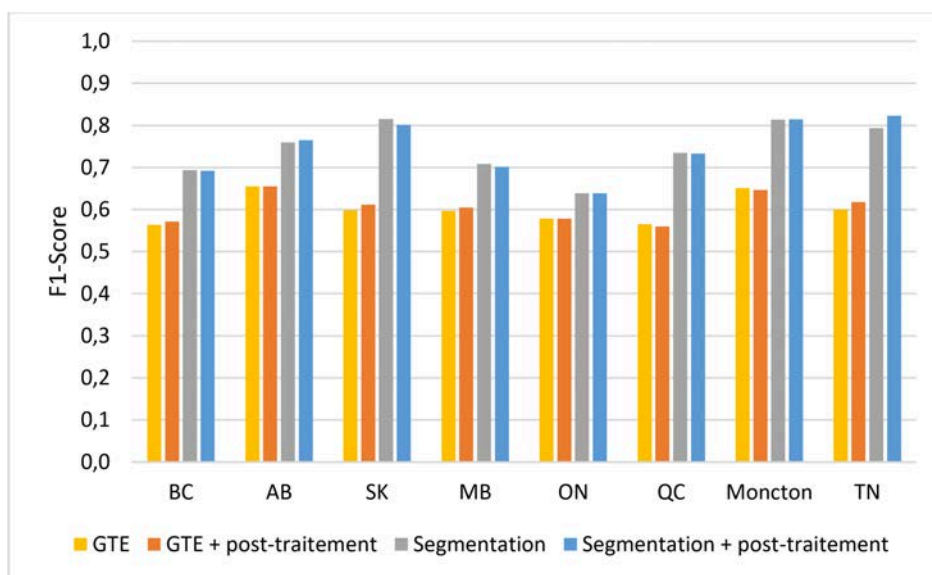


FIGURE 11. Comparaison du F1-Score des quatre méthodes sur les huit provinces-tests.

Premièrement, les résultats montrent qu'aucun biais important n'est présent entre les provinces. L'APLS, beaucoup plus élevée pour Terre-Neuve, peut être expliquée par les images tests dans cette province, couvrant deux petites villes connectées par une grande route et possédant donc moins de situations complexes à prédire. De plus, l'APLS plus faible du Manitoba vient du fait que ce sont principalement des routes agricoles ou forestières qui sont plus difficilement détectables, car généralement non pavées.

Deuxièmement, le f1-score reste stable après l'application du post-traitement, et ce, pour chacune des provinces. Ce comportement montre encore une fois que l'application du post-traitement a un gros impact pour les métriques topologiques et un impact moindre pour les métriques par pixel.

Analyse selon les type de route

Dans le but de pouvoir mesurer les performances des méthodes dans différents milieux et dans diverses conditions, quatre catégories contextuelles ont été choisies, soit autoroutes, pont et viaduc, urbain et rural. Dans la mesure du possible, pour chaque catégorie, trois zones par province ont été aléatoirement choisies pour l'analyse. La figure 12 montre des exemples visuels de cette sélection.



FIGURE 12. Exemple des catégories contextuelles choisies pour l'analyse: rural, autoroute et urbain.

Les dernières analyses ayant montré que la technique de post-traitement améliorait les résultats dans toutes les situations, seuls les méthodes avec post-traitements sont comparés avec l'APLS et le f1-score (Figures 13 et 14).

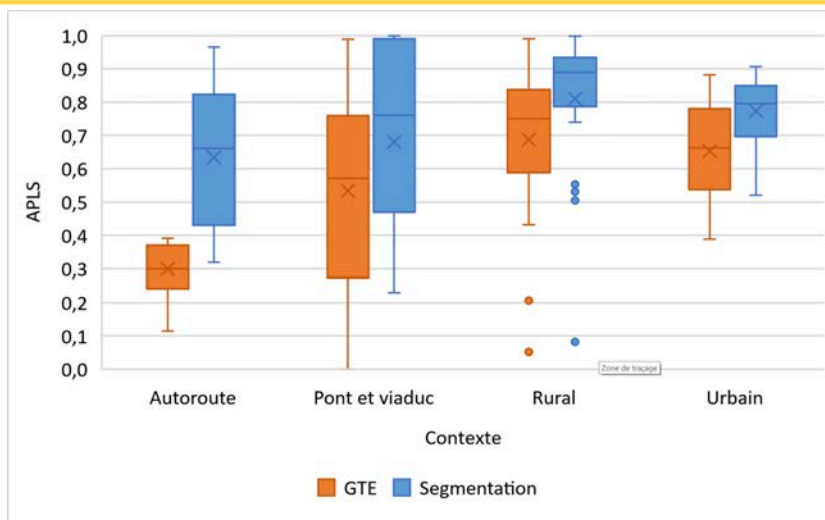


FIGURE 13. Comparaison de l'APLS des deux méthodes avec post-traitements pour les quatre catégories contextuelles.

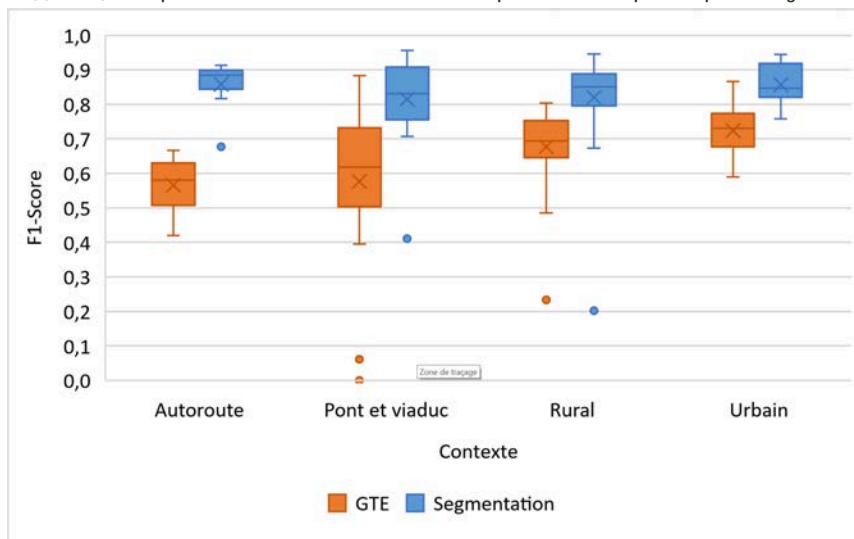


FIGURE 14. Comparaison du F1-Score des deux méthodes avec post-traitements pour les quatre catégories contextuelles.

Premièrement, les résultats montrent une baisse importante de performance pour les autoroutes, les ponts et les viaducs comparés aux catégories urbain et rural (Figure 13). Cette baisse de performance était attendue et correspond entre autres au changement de matériaux des ponts, qui crée des déconnexions. De plus, les entrées/sorties d'autoroutes posent parfois problèmes, étant soit omises ou mal connectées, ce qui nuit davantage à l'APLS. Il est cependant important de noter que seule l'APLS est variable selon les catégories contextuelles, tandis que le f1-score reste stable.

Deuxièmement, l'APLS en milieu rural, est meilleur qu'en milieu urbain. Ce comportement peut être expliqué par le fait que moins de particularités, d'objets ou d'obstructions sont présentes en milieu rural et qu'il y ait moins de routes et de trajets potentiels. L'APLS diminue donc en milieu urbain, car une seule omission peut empêcher l'accès à un quartier, ou un territoire complet. Cependant, le f1-score est légèrement plus faible en milieu rural, car les principales omissions sont des routes forestières, ou difficilement détectables, créant des erreurs sur de plus longues distances. C'est également ce qui crée un plus grand nombre de valeurs extrêmes pour l'APLS en milieu rural.

Troisièmement, une baisse importante des deux métriques est observée pour le GTE, comparativement à la segmentation binaire. Elle suit les tendances observées dans les précédentes analyses et indique encore une fois que le GTE est moins performant pour l'extraction d'un réseau routier cohérent.

Analyse visuelle

Cette section permet d'illustrer les observations faites lors des dernières analyses et de tirer des conclusions sur la justesse du réseau extrait. Cette analyse est faite sur les images test et des exemples sont donnés sur les prédictions des différentes techniques à la figure 15.

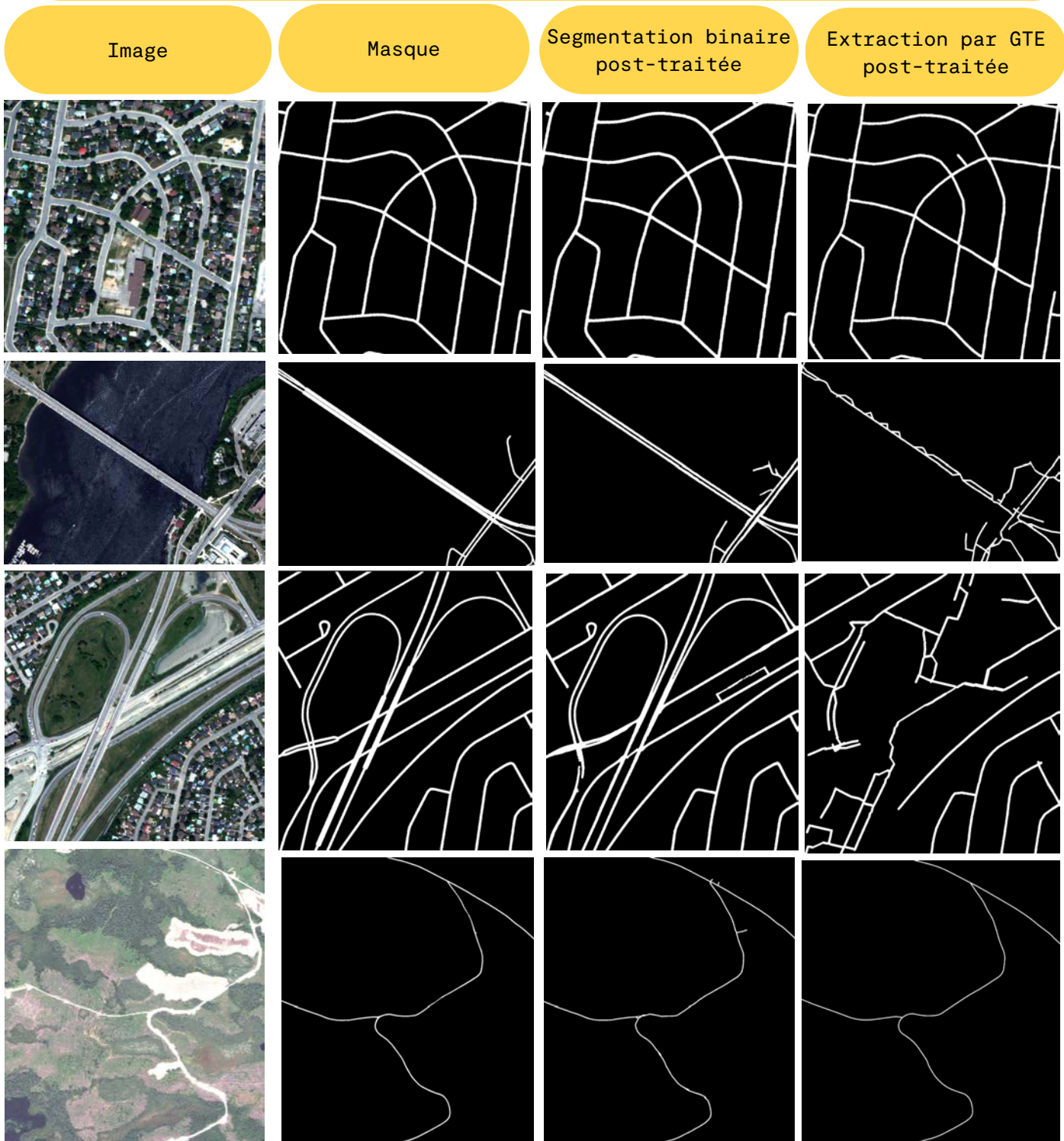


FIGURE 15. Exemple de prédictions en milieu rural et en milieu urbain.



Analyse visuelle...suite

À la suite de cette analyse, le premier point retenu est le même que pour l'analyse quantitative : le réseau extrait par segmentation binaire est généralement de meilleure qualité. En effet, particulièrement au niveau des courbes, des doubles voies, des endroits plus denses ou de formes non régulières, la segmentation binaire performe généralement mieux et fournit une meilleure prédiction.

De plus, dans le GTE, beaucoup d'erreurs a été observées lorsque l'angle de la route est proche d'un multiple de 60. En effet, à ses angles, l'encodage des nœuds voisins change d'ensemble et l'algorithme doit changer de canal de prédiction. Une ambiguïté survient alors et le signal est « dilué » entre les deux ensembles ou oscille d'un angle à l'autre, provoquant des zigzags dans le réseau extrait.

Ce problème pourrait toutefois se régler en appliquant une rotation de 90° aux images lors des prédictions et en moyennant ces prédictions avec les prédictions originales, avant le décodage du GTE. Ainsi, les angles se situant à une frontière d'un ensemble de 60° se retrouverait au centre d'un autre ensemble, et aucune ambiguïté ne subsisterait pour l'image pivotée.

Conclusion

Dans le cadre de ce projet, deux techniques d'extraction de réseau routier par apprentissage profond ont été développées, en plus de l'application d'une technique de post-traitement sur les prédictions. Les résultats provenant de ces techniques, ont été comparés expérimentalement, tant au niveau de leur facilité d'utilisation et de leurs limites, qu'au niveau de leur capacité à extraire des réseaux routiers cohérents.

D'abord, concernant la facilité d'utilisation, la segmentation classique possède plusieurs avantages. En effet, le faible volume de stockage nécessaire pour les masques binaires est avantageux pour la création et la maintenance de la base de données, le temps d'entraînement de l'algorithme et le stockage des prédictions. De plus, les fonctions d'augmentation de données sont déjà implémentées et il existe plusieurs pistes d'amélioration pour ce type de segmentation. Finalement, le principal avantage du GTE face à la segmentation est qu'il permet de considérer la superposition des routes, limite théorique de la segmentation binaire.

Ensuite, les performances de détection de réseau sont largement supérieures pour la segmentation binaire, ayant en moyenne une APLS 13,5 % plus élevée et un f1-score 14,0 % plus élevé. Ces performances supérieures sont équivalentes pour chaque province de l'ensemble de données de test et pour chacune des catégories contextuelles, tant en milieu urbain, rural, près des autoroutes ou aux passages des ponts et viaducs.

Pour conclure, l'APLS reste faible et montre que des développements sont nécessaires pour l'implémentation d'une technique automatique d'extraction du réseau routier pancanadien. Bien que la segmentation binaire fournisse présentement de meilleurs résultats, elle possède toujours une limite théorique importante. De l'autre côté, le GTE est plus difficile à utiliser et nécessite davantage de développements, mais reste une solution intéressante pour l'extraction d'un graphe. Des pistes d'amélioration comme des rotations au moment des prédictions, de l'injection de nœuds supplémentaires ou encore de l'optimisation des hyper-paramètres pourraient en effet améliorer largement les performances de cette technique.



Références

[1] Genest, Marc-Antoine et al., Développement d'une méthode cartographique de détection de réseaux par télédétection et apprentissage profond, septembre 2023, technote 2023-02.

[2] Songtao He et al., Sat2Graph: Road Graph Extraction through Graph-Tensor Encoding, 2020.

[3] https://github.com/CERFO/rncan_road_inference

[4] Qiang Zuo, Songyu Chen, Zhifang Wang, 2AU-Net: Attention Recurrent Residual Convolutional Neural Network for Multimodal Medical Image Segmentation, Secur. Commun. Networks.

[5] N. Akhtar and A. Mian, Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey, in IEEE Access, vol. 6, pp. 14410-14430, 2018.

[6] Etten, Adam, City-scale Road Extraction from Satellite Imagery, 2020.

cerfo

FORMATION
ACCOMPAGNEMENT
RECHERCHE
EN FORESTERIE

Tél. : (418) 659-4225

Courriel : info@cerfo.qc.ca

Communiquez avec notre équipe :

Marc-Antoine Genest, M. Sc. magenest@cerfo.qc.ca

Xavier Gervais, M. Sc. xgervais@cerfo.qc.ca

Batistin Bour, M. Sc. bbour@cerfo.qc.ca

Charles Marseille, M. Sc. cmarseille@cerfo.qc.ca

Mathieu Varin, M. Sc. mvarin@cerfo.qc.ca

