

DÉPARTEMENT DE GÉOMATIQUE APPLIQUÉE

Faculté des lettres et sciences humaines

Université de Sherbrooke

Caractérisation de trottoirs :
Extraction automatique d'entités géographiques sur des images
panoramiques urbaines par réseaux de neurones convolutifs

Mémoire de recherche - GPA804

Par RÉMI TAVON

Candidat à la maîtrise en sciences géographiques

Cheminement de type recherche en environnements géographiques

Sous la direction de

Yacine Bouroubi (Université de Sherbrooke)

Lynda Bellalite (Université de Sherbrooke)

Samuel Foucher (Centre de recherche informatique de Montréal)

Sherbrooke

24 janvier 2021

Résumé

L'état des trottoirs a un impact déterminant sur la mobilité des personnes à mobilité réduite, leur participation à la vie sociale et leur état de santé physique. Un suivi automatisé de ces trottoirs, à l'instar de celui qui est actuellement développé pour le réseau routier dans plusieurs grandes villes, permettrait d'alléger les coûts liés aux relevés manuels à des fins d'entretien régulier. Le présent projet vise à bonifier les facteurs pris en compte par un outil de planification de trajets afin qu'il réponde mieux aux besoins de mobilité de ce type d'utilisateurs.

Depuis l'avènement des réseaux de neurones convolutifs (CNN) en reconnaissance visuelle, les applications liées à ces nouveaux outils d'apprentissage machine ne cessent de se multiplier. Les CNN présentent un intérêt majeur dans l'application des techniques de vision par ordinateur en télédétection. L'objectif principal des travaux présentés ici consiste à développer une méthode permettant de caractériser automatiquement les trottoirs. Pour ce faire, des images de scènes urbaines *Google Street View* couvrant deux villes du Québec ont été acquises, puis annotées afin d'y associer la vérité terrain. Une fois entraînés sur ces données, les CNN ont servi à détecter et caractériser les trottoirs sur un ensemble de test, couvrant 25 km du centre-ville de Sherbrooke. Une approche d'apprentissage par transfert est privilégiée afin de tirer profit de grands jeux de données annotées, comme *Cityscapes*. La segmentation des trottoirs sur le site d'évaluation finale atteint un indice de Jaccard de 91 %. La caractérisation, quant à elle, n'atteint que 31 %, notamment à cause d'un déséquilibre de classes dans les données d'entraînement. Une récurrence de surapprentissage révèle enfin le besoin d'accroître la quantité de ces données servant à l'apprentissage des modèles.

Mots clés : universal accessibility, universal design, pavement distress, crack detection, deep convolutional neural networks.

Table des matières

Résumé.....	2
Liste des acronymes.....	iii
Liste des figures.....	iv
Liste des tableaux.....	ix
Remerciements.....	x
Chapitre 1. Problématique.....	1
1.1. Mise en contexte.....	1
1.2. Cadre théorique.....	3
1.2.1. Conception pour tous : le cas des trottoirs.....	3
1.2.2. Traitement d'image par apprentissage profond.....	12
1.3. Objectifs et hypothèse de recherche.....	22
1.3.1. Objectifs secondaires.....	23
1.3.2. Hypothèse de recherche.....	23
1.4. Pertinence de l'étude.....	23
1.5. Limites du projet.....	24
Chapitre 2. Méthodologie.....	25
2.1. Organigramme méthodologique.....	25
2.2. Sources d'informations.....	33
2.3. Données utilisées.....	25
2.4. Choix des sites d'étude.....	25
2.4.1. Sites d'entraînement.....	26
2.4.2. Site d'évaluation finale.....	30
2.5. Tests de prétraitement.....	33
2.6. Comparaison des méthodes de transfert d'apprentissage.....	34
2.6.1. Extraction de caractéristiques.....	35
2.6.2. Affinement de l'apprentissage.....	35
2.6.3. Affinement partiel.....	36
2.6.4. Déroulement et résultats des tests.....	37
2.7. Choix d'architectures.....	38
2.7.1. Deeplabv3+.....	38
2.7.2. MobileNetV2.....	40
2.8. Entraînement de modèles par segmentation sémantique.....	42

2.8.1.	Tâche 1 : segmentation des trottoirs	42
2.8.2.	Tâche 2 : caractérisation de l'état du revêtement.....	44
2.9.	Post-traitement	46
2.10.	Évaluation finale.....	47
2.11.	Présentation et analyse des résultats	47
2.12.	Interprétation des résultats.....	48
Chapitre 3.	Résultats : présentation et analyse	48
3.1.	Tests de prétraitement	48
3.2.	Comparaison des méthodes de transfert d'apprentissage.....	50
3.3.	Tâche 1 : segmentation des trottoirs.....	51
3.4.	Tâche 2 : caractérisation de l'état du revêtement.....	55
3.5.	Évaluation finale	57
3.3.1.	Post-traitement	63
Chapitre 4.	Interprétation des résultats et discussion.....	63
4.1.	Objectif n°1 : recours à l'apprentissage profond.....	63
4.1.1.	Niveau de complexité de la tâche	64
4.1.2.	Débalancement des classes	65
4.1.3.	Manque de données d'entraînement	66
4.1.4.	Faux positifs : une lacune possible des réseaux travaillant en silo.....	69
4.2.	Objectif n°2 : robustesse du modèle sur le site d'évaluation finale	70
4.3.	Objectif n°3 : avantages, inconvénients et limitations	72
4.3.1.	Recours à l'apprentissage profond.....	72
4.3.2.	Utilisation de panoramas <i>GSV</i>	72
4.4.	Retour sur l'hypothèse de recherche	73
4.5.	Discussion	74
4.5.1.	Alternatives aux algorithmes utilisés	74
4.5.2.	Alternatives aux panoramas <i>GSV</i>	75
Conclusion	78
Références	81

Liste des acronymes

CDPH : Convention relative aux droits des personnes handicapées

CNN : réseaux de neurones convolutifs

GSV: Google Street View

IoU: *Intersection over Union* ou indice de Jaccard

MDH-PPH 2 : modèle de développement humain et Processus de production du handicap

PRM : personne à mobilité réduite

RIPPH : réseau international sur le processus de production du handicap

Liste des figures

Figure 1: Modèle de développement humain - Processus de production du handicap 2 (Fougeyrollas, 2010, p.152).....	5
Figure 2: Facteurs influençant le choix d'un cheminement piéton. Adapté de Gendron (2014, p.119).....	8
Figure 3: Les cinq dimensions relatives à l'accès. Adapté de Fougeyrollas et al. (2019, p.169).....	9
Figure 4: Exemples d'aménagements accessibles en lien avec les trottoirs (Beneficial Designs Inc., 2001, p.4-16 à 4-27).....	10
Figure 5: Deux véhicules utilisés lors de l'auscultation traditionnelle de trottoir. Tiré de Blanchette (2011, p.30).....	11
Figure 6: Changement de paradigme entre la programmation classique et l'apprentissage machine. Adapté de Chollet (2017a, p.5)	12
Figure 7: Composition d'un neurone, l'unité computationnelle de base des couches d'un réseau. Chaque valeur d'entrée x_1, x_2, \dots, x_r est multipliée par un poids $w_{i,1}, w_{i,2}, \dots, w_{i,r}$. Un coefficient b_i , qui s'apparente à l'ordonnée à l'origine dans une équation linéaire, s'ajoute à ces poids. Une fonction d'activation f est ensuite appliquée à cette somme pondérée a_i , produisant la valeur de sortie y_i . (tiré de Tanikic, 2012, p.165)	13
Figure 8: Quatre tâches courantes en vision par ordinateur (Tiré de Wu et al. 2020, p.39)	15
Figure 9: Taux d'erreur obtenus par les réseaux de neurones convolutifs de 2010 à 2015 lors du concours ImageNet Large Scale Visual Recognition Challenge (ILSVCR). Le concours a pris fin en 2017 étant donné que la problématique était considérée résolue (tiré de Goodfellow et al., 2016, p.24).....	17
Figure 10: Exemple d'une convolution 2D. En haut, les deux matrices qui servent à effectuer le produit scalaire. En bas, six sorties possibles résultant du déplacement du filtre de convolution sur les valeurs d'entrée (input) (Tiré de Goodfellow et al. 2016, p.327) .	18
Figure 11: Comparaison de plusieurs réseaux de neurones profonds issus de l'état de l'art selon leur performance sur ImageNet, le nombre d'opérations que nécessite l'inférence et le nombre de paramètres qui les constitue. (tiré de Canziani, 2018)	19

Figure 12: Structure de SegNet, une architecture de type encodeur-décodeur utilisée en segmentation sémantique (tiré de Badrinarayanan et al., 2017, p.4)	20
Figure 13: Exemple de différentes techniques d'augmentation de données permettant de réduire le surapprentissage (tiré de Ahmad et al. 2017, p.6)	21
Figure 14: 2953 points d'extraction de panoramas Google Street View dans le quartier Ahuntsic-Cartierville au nord de l'île de Montréal (Qc)	27
Figure 15: 2455 points d'extraction de panoramas Google Street View à Saint-Hyacinthe (Qc)	27
Figure 16: Zones (en bleu) annotées en vue d'être ignorées durant l'entraînement. Celles-ci sont volontairement exclues du calcul de perte, car le chevauchement entre deux panoramas voisins diminue l'importance des prédictions de modèle pour ces zones.	29
Figure 17: 2609 points d'extraction de panoramas Google Street View sur le site d'étude au centre-ville de Sherbrooke (Qc).....	30
Figure 18: Organigramme méthodologique du projet de recherche	32
Figure 19: Exemple d'un panorama rogné, montrant seulement un côté de la route, avec et sans la distorsion originale causée par la projection équirectangulaire. Sur la photo de gauche, il est inhabituel de percevoir un trottoir avec une telle déformation.	34
Figure 20: Transfert d'apprentissage de type « Extraction de caractéristiques ». Seule la couche de sortie est réentraînée pour la tâche cible.	35
Figure 21: Transfert d'apprentissage de type « Affinement de l'apprentissage ». La couche de sortie est d'abord réentraînée, puis l'apprentissage se poursuit pour le corps du réseau, après copie des paramètres pour cette portion.	36
Figure 22: Transfert d'apprentissage de type « Affinement partiel ». La couche de sortie est d'abord réentraînée, puis l'apprentissage se poursuit pour une partie seulement du réseau.	37
Figure 23: Structure de Deeplabv3+, une architecture de segmentation sémantique de type encodeur-décodeur (tiré de Chen et al., 2018, p.4).....	39
Figure 24: Structure du module « Regroupement de pyramides spatiales à trous » (Atrous Spatial Pyramid Pooling) contenu dans l'encodeur de Deeplabv3+ (tiré de Chen et al., 2017, p.1)	40

Figure 25: Trois approches de composition des architectures de segmentation sémantique. (a) regroupement de pyramides spatiales, (b) Encodeur-décodeur, (c) Encodeur-décodeur avec convolution à trous, un hybride entre (a) et (b) utilisé dans Deeplabv3+ (tiré de Chen et al., 2018, p.2)	40
Figure 26: Performances, taille (nombre de paramètres) et nombre d'opérations de 3 réseaux mobiles de la famille MobileNet et d'un réseau de classification classique, ResNet-101. Chaque réseau a été testé selon deux configurations différentes.(tiré de Sandler et al., 2018, p.4517)	41
Figure 27: Exemple d'un panorama Google Street View. Les cadres blancs indiquent les imagettes extraites dans chaque panorama après rognage pour les fins de l'entraînement. N.B. La taille réelle de l'image originale est 6656 x 3328 pixels.	43
Figure 28: Échantillons représentatifs de chaque classe pour l'état du revêtement (gauche à droite) : bon état, lézardes superficielles, crevasses profondes/larges, très endommagé/impraticable	45
Figure 29: Exemple d'un échantillon utilisé durant l'entraînement de modèles pour la tâche 2 (caractérisation de l'état du revêtement)	45
Figure 30: Chaîne de traitement finale intégrant les modèles les plus performants pour chacune des deux tâches (segmentation et caractérisation)	46
Figure 31: Deux images avec copie distordue (originale) et corrigée avec les prédictions obtenues par Deeplabv3+ préentraîné sur Cityscapes.	49
Figure 32: Évolution de l'indice de Jaccard pour les tests d'affinement complet et partiel. Ce dernier test comportait trois sous-tests dans lesquels la dernière couche gelée était définie à trois emplacements différents (pour plus de détails sur ces couches, voir Chollet, 2017a). Les courbes sont aplanies pour faciliter la visualisation. Selon les résultats en cours d'entraînement, ceux-ci étaient prolongés plus ou moins longtemps, d'où le nombre d'époques d'entraînement variable. Les résultats pour la méthode d'extraction des caractéristiques ne sont pas présentés.	50
Figure 33: Superposition d'un panorama Google Street View et d'un masque de segmentation (en rose) pour les pixels de la classe «Trottoir». Ce masque a été produit par le modèle Deeplabv3+ entraîné sur les données d'entraînement du présent projet.	52

Figure 34 : Exemples de masques imparfaits (en rose) produit par le modèle de segmentation des trottoirs.	53
Figure 35: Évolution de l'indice de Jaccard en fonction de l'époque d'entraînement, pour les ensembles d'entraînement et de validation. Cet entraînement a permis de produire le modèle utilisé pour la segmentation des trottoirs. L'apparence bruitée de la courbe de validation est causée par le niveau de difficulté élevé de la tâche d'apprentissage jumelée aux propriétés de la métrique utilisée, qui pénalise fortement les mauvais recouvrements entre prédictions et cibles.....	54
Figure 36: Évolution de l'indice de Jaccard en fonction de l'époque d'entraînement d'un modèle de caractérisation de l'état du revêtement avant l'exclusion de panoramas ne présentant aucun défaut.....	55
Figure 37: Évolution de l'indice de Jaccard en fonction de l'époque d'entraînement, pour les ensembles d'entraînement et de validation. Cet entraînement a permis de produire le modèle utilisé pour la segmentation des défauts de revêtement sur les trottoirs.	56
Figure 38: Exemples de vrais positifs (défauts réels) détectés par la chaîne de traitement finale (après post-traitement). Chaque défaut est entouré d'une boîte, d'un chiffre unique identifiant ce défaut sur l'image de référence, ainsi que de l'aire, en pixel, du défaut. ...	58
Figure 39: Exemples de faux positifs liés à la détection d'un défaut mineur plutôt que majeur. Chaque défaut est entouré d'une boîte, d'un chiffre unique identifiant ce défaut sur l'image de référence, ainsi que de l'aire, en pixel, du défaut. Ces exemples sont issus de la chaîne de traitement finale (après post-traitement).....	59
Figure 40 : Exemples de faux positifs liés à une commission du modèle de segmentation du trottoir. Selon la méthodologie adoptée, le modèle de caractérisation n'avait pas été entraîné à rejeter des zones qui ne correspondait pas au trottoir. Chaque défaut est entouré d'une boîte, d'un chiffre unique identifiant ce défaut sur l'image de référence, ainsi que de l'aire, en pixel, du défaut. Ces exemples sont issus de la chaîne de traitement finale (après post-traitement).	60
Figure 41 : Exemples de faux positifs liés à la sous-représentation d'images captées au printemps et à l'automne. Chaque défaut est entouré d'une boîte, d'un chiffre unique identifiant ce défaut sur l'image de référence, ainsi que de l'aire, en pixel, du défaut. Ces exemples sont issus de la chaîne de traitement finale (après post-traitement).	61

Figure 42 : Exemples de faux positifs liés à la présence d'une ombre ou d'une bouche d'égout, à des travaux de réfection, à un revêtement atypique (pavé uni, asphalte) ou à d'autres événements rares. Chaque défaut est entouré d'une boîte, d'un chiffre unique identifiant ce défaut sur l'image de référence, ainsi que de l'aire, en pixel, du défaut. Ces exemples sont issus de la chaîne de traitement finale (après post-traitement).	62
Figure 43: Exemple d'un panorama Google Street View où une ambiguïté existe entre la limite du trottoir et celle de la rue, à cause de la présence de gravier.	64
Figure 44: 5 des 25 échantillons photo soumis pour évaluation auprès d'une experte. Le classement donné par celle-ci a été comparé à celui de l'annotateur et a permis de diminuer le niveau de subjectivité pour l'annotation subséquente sur les données d'entraînement (tâche de caractérisation).	65
Figure 45: Comparaison du nombre de pixels annotés pour la classe «trottoir» par rapport à deux jeux de données de référence en segmentation sémantique appliquée à l'imagerie urbaine.....	67
Figure 46: Chaîne de traitement pour l'acquisition de données d'entraînement tel qu'utilisé par Ma et al. (2017). Les panoramas GSV sont acquis automatiquement à partir de données publiques existantes pour la ville de New York. (tiré de Ma et al., 2017, p.3)	68
Figure 47: Structure d'un réseau de neurones multitâches en segmentation sémantique tel que proposé par Bischke et al. (2019), p.1482.....	70
Figure 48: Impact de l'augmentation de données au moment de l'inférence (courbe « train + test »). La mesure en ordonnée correspond à l'aire sous la courbe ROC, une mesure de performance commune en classification. (tiré de Shorten et al., 2019, p.36).....	71
Figure 49: Système PaveVision3D avec résolution de 1 mm . a) extérieur et (b) fonctionnement (tire de Luo et al., 2014, p.122)	76

Liste des tableaux

Tableau 1: Définition des classes pour la caractérisation de l'état du revêtement des trottoirs	45
Tableau 2: Comparaison des hyperparamètres et performances pour les tests des méthodes de transfert d'apprentissage pour la tâche de segmentation des trottoirs	50
Tableau 3: Hyperparamètres ayant produit le modèle Deeplabv3+ retenu pour la segmentation des trottoirs	54
Tableau 4: Répartition relative des pixels annotés pour entraînement en fonction de la classe d'appartenance	55
Tableau 5: Hyperparamètres ayant produit le modèle MobileNetv2 retenu pour la caractérisation des trottoirs	57
Tableau 6: Répartition relative des causes d'erreur ayant causé des faux positifs sur le site d'évaluation finale	59

Remerciements

Merci à mon directeur, Yacine Bouroubi, de m'avoir donné l'opportunité de compléter ce projet de maîtrise dans un domaine totalement nouveau et inconnu pour moi. Merci également pour le jumelage avec l'équipe VISI du CRIM, dont l'apport a été inestimable.

Merci à ma codirectrice, Lynda Bellalite, pour son expertise et son encadrement tout au long du projet.

Merci à mon codirecteur, Samuel Foucher, d'avoir accepté un néophyte au sein de l'institution respectable qu'est le CRIM. Son soutien, celui de Pierre-Luc Saint-Charles, de Justine Boulent et de tous les autres membres de l'équipe VISI ont été déterminants dans la réussite du projet.

Merci à Mathieu Saint-Laurent pour le prêt de ses cartes graphiques personnelles et tous ses efforts pour faciliter les traitements sur ses équipements.

Merci à mes collègues du A6-1028 qui ont mis de la joie dans mes journées quand la motivation et les résultats n'étaient pas au rendez-vous.

Merci à ma conjointe Rosalie et à mes enfants Léonard et Henri pour leur soutien moral.

Merci enfin à Miles Davis, Horace Silver, Dave Brubeck et Scott Joplin de m'avoir accompagné tout au long de mon projet.

Chapitre 1. Problématique

1.1. Problématique

La caractérisation automatique des trottoirs grâce aux images panoramiques urbaines présente des débouchés prometteurs, en particulier à l'égard des avancées récentes en apprentissage profond. Quoiqu'une telle méthode ne puisse atteindre un niveau de précision équivalent à une auscultation traditionnelle, ce type de caractérisation permettrait néanmoins à des villes de petite et moyenne tailles d'obtenir une évaluation sommaire de l'état du revêtement de ses trottoirs, de manière récurrente, à peu de frais. De plus, une caractérisation par images panoramiques urbaines pourrait être complémentaire à une auscultation spécialisée telle que pratiquée actuellement.

Au confluent des efforts investis en conception universelle et des progrès récents en apprentissage profond, un horizon encourageant se dessine donc pour l'accès des municipalités à des données fréquemment mises à jour concernant leur réseau de trottoirs. Enfin, de telles données permettraient l'amélioration des plateformes Web de planification d'itinéraires adaptés. Dans ce contexte, le présent projet mise sur le potentiel des techniques en traitement d'image par intelligence artificielle, issues de l'état de l'art, afin d'aborder, voire résoudre, la problématique soulevée.

1.2. Mise en contexte

À l'occasion de la dernière enquête canadienne sur l'incapacité, 7,2 % des Canadiens de plus de 15 ans ont déclaré avoir un « trouble de mobilité qui limitait leurs activités quotidiennes » (Statistiques Canada, 2013). En Belgique, une étude estime que cette proportion atteint près de 33 % (Ministère de la Région de Bruxelles-Capitale, 2008). Cette différence significative relève d'une définition élargie du concept de mobilité réduite qui, dans le cas de la Belgique, tient aussi compte des enfants, des personnes âgées, voire des

femmes enceintes ou des livreurs. Cette approche, mue par une volonté d'inclusion et d'intégration, déteint sur les politiques d'accessibilité universelle telles qu'adoptées par plusieurs villes à l'international.

Ainsi, l'approche traditionnelle selon laquelle l'individu avait la pleine responsabilité de s'adapter à son environnement physique a évolué vers une approche où l'environnement devrait, au contraire, être adapté à l'individu (Beneficial Designs Inc., 2001). Au Québec, la Loi assurant l'exercice des droits des personnes handicapées en vue de leur intégration scolaire, professionnelle et sociale (Publications Québec, 2018) est le principal outil législatif favorisant l'atteinte d'une telle vision. Toutefois, les défis sont encore nombreux afin de réduire les obstacles auxquels font face les personnes à mobilité réduite (PMR).

Au chapitre des solutions visant à faciliter les déplacements des PMR, en particulier les personnes en fauteuil roulant, différentes solutions technologiques existent afin de planifier des trajets exempts d'obstacles. Toutefois, les plateformes actuelles de trajets adaptés présentent une lacune majeure : ils ne prennent pas en compte l'état du revêtement des trottoirs et autres corridors piétonniers (traverses piétonnes, sentiers, bordure de route sans trottoir) dans leurs algorithmes, autrement que par l'approvisionnement par le public (approvisionnement par la foule, *crowdsourcing*). Comme le démontrent plusieurs études cliniques (voir section 1.3.1 « Conception pour tous : le cas des trottoirs »), cette variable (état du revêtement) a un impact non négligeable sur la santé à long terme des usagers en fauteuil roulant. De plus, un trottoir dont le revêtement est fortement détérioré peut constituer une entrave majeure au passage d'une personne en fauteuil roulant.

À l'heure actuelle, la caractérisation des trottoirs d'une ville comme l'état du revêtement nécessite des relevés de terrain fort coûteux. Dans une ville de région comme Sherbrooke, le recours à l'auscultation systématique des infrastructures de transport commence à peine et cible uniquement le réseau routier (Dubé, 2018). Quant aux trottoirs, l'évaluation de l'état de leur revêtement dépend, entre autres, des plaintes formulées à la Ville (Ville de Sherbrooke, 2018). Une approche par plainte, moins coûteuse qu'un relevé systématique par du personnel qualifié, fait cependant place à une connaissance fragmentaire de l'état des infrastructures et introduit un biais, car elle se base sur la perception des usagers plutôt que sur l'état réel des infrastructures.

En ce sens, la caractérisation automatique des trottoirs par traitement d'images présente des débouchés prometteurs, en particulier à l'égard des avancées récentes en apprentissage profond. Ces avancées, combinées aux nouvelles sources de données de télédétection, pourraient procurer des bénéfices inestimables dans le déploiement d'une politique d'accessibilité universelle, voire en gestion des infrastructures urbaines de façon globale.

1.3. Cadre théorique

1.3.1. Conception pour tous : le cas des trottoirs

Depuis la fin des années 1960, des chercheurs et groupes de défense des droits des personnes « handicapées » ont imposé une vision du handicap qui renverse l'approche selon laquelle l'individu avec une incapacité est l'unique responsable de son handicap. Ce changement de vision a mené, au Québec, à l'adoption de la Loi assurant l'exercice des droits des personnes handicapées (Publications Québec, 2018), qui devient également l'acte de naissance de l'Office des personnes handicapées du Québec (OPHQ).

Quelques années plus tard, en 1984, la politique *À part... égale* (OPHQ, 1984) réitère à nouveau l'importance de la dimension contextuelle dans la création du handicap. Ainsi, l'État québécois reconnaît l'influence d'acteurs externes comme les municipalités et les organismes publics et privés dans l'atténuation des conséquences du handicap et dans la pleine participation sociale des personnes avec incapacité (Nations Unies, 2006).

La conception universelle (ou conception pour tous) découle directement de cette reconnaissance unilatérale de l'environnement comme facilitateur ou obstacle à l'émancipation des personnes avec incapacité. Cette discipline vise à guider plusieurs corps de métiers (urbaniste, entrepreneur, architecte, décideur), afin qu'ils considèrent les besoins du plus grand nombre d'usagers potentiels. Dès lors, cette approche tente de minimiser les barrières physiques et sociales lors de leur utilisation de produits, de services, d'installations et d'environnements (Beneficial Designs Inc., 2001).

Avant d'aborder plus spécifiquement les notions d'accès inclusif¹ et de conception universelle en lien avec les trottoirs, les deux sections suivantes visent à introduire les pierres d'assise sur lesquelles repose ce domaine central dans la planification urbaine. Ainsi, le processus de production du handicap et le concept de mobilité réduite sont autant de notions qui sont abordées préalablement aux concepts centraux des présents travaux.

Processus de production du handicap et droits humains

Le handicap, de la perspective promue par la Convention relative aux droits des personnes handicapées (CDPH) (Nations Unies, 2006) et le Réseau international sur le processus de production du handicap (RIPPH), est d'abord et avant tout un enjeu de droits humains. Son ancrage dans le respect des différences et l'idéologie des droits humains le distingue de l'approche biomédicale du handicap développée par l'Organisation mondiale de la santé (OMS) (OMS, 2001). Celle-ci, sans rejeter l'impact de facteurs environnementaux, aborde la question du handicap du point de vue de l'individu. Sans s'y opposer complètement, les travaux de la CDPH et du RIPPH permettent de mieux cerner le rôle crucial et incontournable de l'État dans la participation sociale des personnes en situation de handicap. Ainsi, le regard interdisciplinaire qu'offre le cadre d'analyse issu des droits humains sera privilégié dans le cadre des présents travaux, d'autant plus que l'étendue des travaux publiés depuis les années 1980 qui en découlent offre un socle théorique solide pour aborder le concept de conception universelle qui le sous-tend.

Le modèle de développement humain et Processus de production du handicap (MDH-PPH 2) (Fougeyrollas, 2010) définit le handicap comme l'interaction entre trois domaines conceptuels : les facteurs personnels, les facteurs environnementaux et les habitudes de vie (Figure 1). Les facteurs personnels réfèrent aux aptitudes fonctionnelles de l'individu (voir, courir, entendre) et à son identité (appartenance à un groupe, une religion, une idéologie). Les facteurs environnementaux (voies de circulation, édifices, attitudes, législation), quant à eux, s'appliquent à trois niveaux : le micro-environnement d'une personne (lieu de vie, poste de travail, famille nucléaire), son méso-environnement (quartier, communauté) et son macro-environnement (société). Enfin, les habitudes de vie incorporent les activités

¹ Le choix de ce terme, plutôt qu'« accessibilité universelle » découle des travaux du Centre interdisciplinaire de recherche et réadaptation et intégration sociale (CIRIS) (Fougeyrollas *et al.*, 2019).

quotidiennes (se laver, s'habiller, prendre ses repas, se déplacer) ainsi que les rôles sociaux de chacun (travailler, étudier, militer dans un parti politique, être parent, conjoint, etc.).

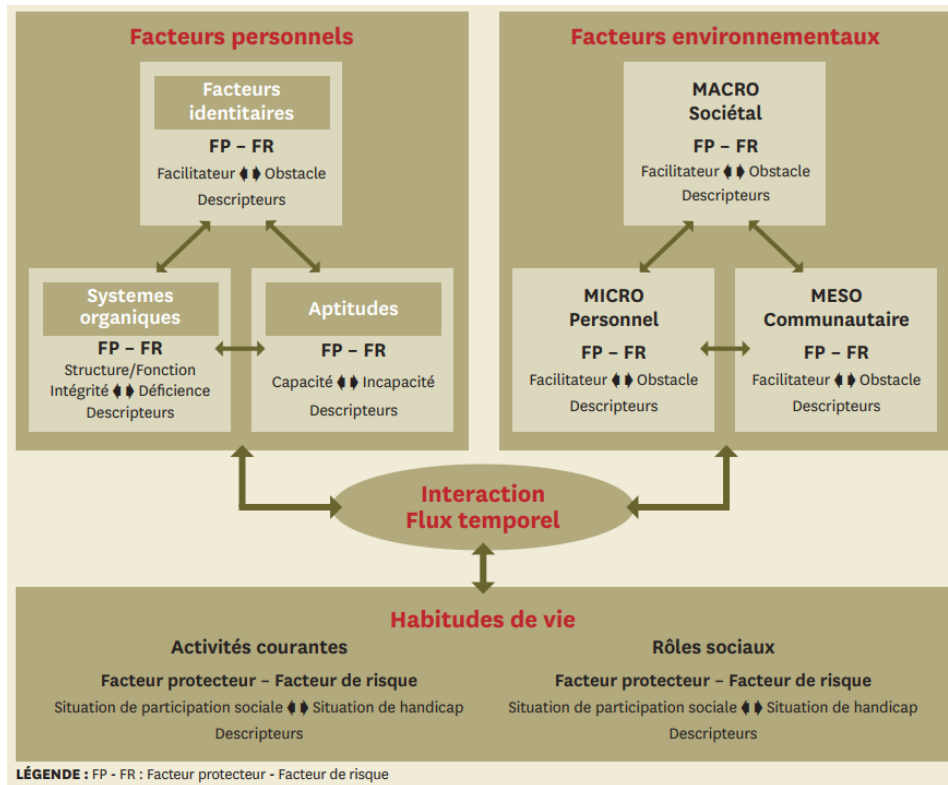


Figure 1: Modèle de développement humain - Processus de production du handicap 2 (Fougeyrollas, 2010, p.152)

Les facteurs environnementaux, centraux dans le présent projet, peuvent être des facilitateurs ou des obstacles lorsqu'ils sont conjugués avec les facteurs personnels et les habitudes de vie d'une personne. Par conséquent, l'impact du cadre bâti, par exemple, sur la situation de handicap d'une personne n'est pas uniquement attribuable aux caractéristiques qui le composent (présence d'une rampe d'accès vers l'entrée d'un édifice, trottoir exempt d'obstacle, etc.), mais également aux facteurs personnels et aux habitudes de vie de la clientèle qui est susceptible de l'utiliser.

Plutôt que de servir d'outil diagnostique pour évaluer le degré d'atteinte d'un organe ou d'une fonction corporelle, le MDH-PPH2 propose une manière de quantifier le niveau de participation sociale des individus. Cette dimension de participation sociale permet, enfin, d'évaluer l'exercice effectif des droits des personnes ayant des incapacités.

Personnes à mobilité réduite (PMR) : un large spectre de la population

Selon un guide de référence belge, une personne à mobilité réduite (PMR) se définit comme :

« Toute personne confrontée à la difficulté de se déplacer et soumise à des entraves qu'elle ne peut surmonter seule, ou très difficilement. Les difficultés de mobilité peuvent être liées à l'âge, à la maladie, à un état (femme enceinte, personne chargée de commissions, livreurs), à un handicap temporaire ou définitif. Elles portent atteinte à l'autonomie de la personne et limitent ses déplacements et sa capacité d'intégration à la société en général. [...] » (Ministère de la Région de Bruxelles-Capitale, 2008, p.1.1)

Selon cette définition, une personne en fauteuil roulant est une PMR au même titre qu'une personne de petite taille, qu'un livreur, qu'une femme enceinte ou qu'un parent accompagné de ces enfants. Cette définition rejoint, d'une part, celle du « passer à mobilité réduite » décrite par le Parlement européen (Parlement européen, 2001). D'autre part, elle s'accorde avec le MDH-PPH2, quoiqu'elle s'étende également à des situations de handicap qui ne sont ni significatives ni persistantes. Ainsi, lorsqu'un enfant en bas âge, par exemple, peine à parcourir un segment de trottoir sans bateau pavé, celui-ci peut vivre une situation de handicap, au sens du MDH-PPH2, même si cet enfant n'est atteint d'aucune déficience. À l'inverse, le MDH-PPH2 permet également de constater que, si le milieu fréquenté par une PMR lui permet de se déplacer aisément (ex. : présence de bateaux pavés), le handicap est diminué ou absent, même si l'incapacité liée à l'individu demeure la même. Ce raisonnement démontre donc qu'un environnement sans obstacle favorise une plus grande autonomie (ex. : accès à un travail, à un établissement d'enseignement, à l'épicerie, à un arrêt d'autobus, etc.) et une meilleure participation sociale (Beneficial Designs Inc., 2001).

Au Québec, selon l'Enquête canadienne sur l'incapacité de 2012 (Statistiques Canada, 2013), 9,6 % de la population québécoise de plus 15 ans est aux prises avec au moins une incapacité, c'est-à-dire une difficulté à réaliser certaines activités du quotidien à cause d'un état ou d'un problème de santé qui persiste pendant six mois ou plus. Parmi ces personnes, environ 50 % d'entre elles comptent une incapacité liée à la mobilité. Ainsi, environ 340 000 Québécoises et Québécois ayant une incapacité persistante éprouvent de la difficulté à marcher sur un terrain plat pendant 15 minutes sans se reposer ou à monter ou

descendre une douzaine de marches sans se reposer. Ces proportions se reflètent également à l'échelle de la région estrienne et de la Ville de Sherbrooke (OPHQ, 2018).

En ce qui a trait à la proportion de la population répondant à la définition de PMR retenue en début de section, la Belgique estime celle-ci à 33 % de sa population (Ministère de la Région de Bruxelles-Capitale, 2008). Bien que les données publiées par l'OPHQ ne permettent pas d'établir un tel décompte, la proportion de PMR au Québec concerne donc certainement plus que les 5 % recensés officiellement. En outre, plusieurs villes du Québec prennent en compte les besoins de tous les types de PMR, qu'il s'agisse de personnes en fauteuil roulant ou de parents avec une poussette (Ville de Montréal, 2017) dans leurs projets d'infrastructures. Les décideurs municipaux sont donc bien au fait que la construction d'infrastructures conçues afin de répondre aux besoins de personnes en situation de handicap profite également à l'ensemble de la population (Fougeyrollas *et al.*, 2019).

Au Québec, les personnes de 65 ans et plus sont deux à trois fois plus susceptibles d'avoir une incapacité (Statistiques Canada, 2013). Le vieillissement de la population est donc un facteur qui pousse d'autant plus les municipalités à considérer les besoins des PMR de manière élargie.

Conception universelle : le cas des trottoirs

Les Québécoises et Québécois avec une incapacité se déplacent plus souvent à pied et en transport en commun que les personnes sans incapacité (OPHQ 2017). Pour ce type de déplacements, les trottoirs et autres corridors piétonniers (traverses piétonnes, sentiers aménagés et non aménagés, autres accotements de route) jouent un rôle névralgique afin de permettre l'accès à un arrêt de transport en commun, à un commerce, un établissement ou encore à un véhicule. Une récente publication de la Ville de Montréal souligne d'ailleurs l'importance capitale de l'accès à ce type d'infrastructure pour son utilisation par la clientèle piétonne (Figure 2).

Par exemple, un trottoir dont le revêtement est en mauvais état, avec des joints trop larges ou encore recouvert de glace ou de neige représente un obstacle fréquent pour les usagers en fauteuil roulant du Québec (OPHQ, 2011), qui sont au nombre de 45 000 environ (OPHQ, 2017). Au moment de planifier la construction de nouvelles infrastructures comme

les trottoirs, les principes de la conception universelle permettent ainsi de réduire au minimum les obstacles auxquels ces personnes peuvent être confrontés.

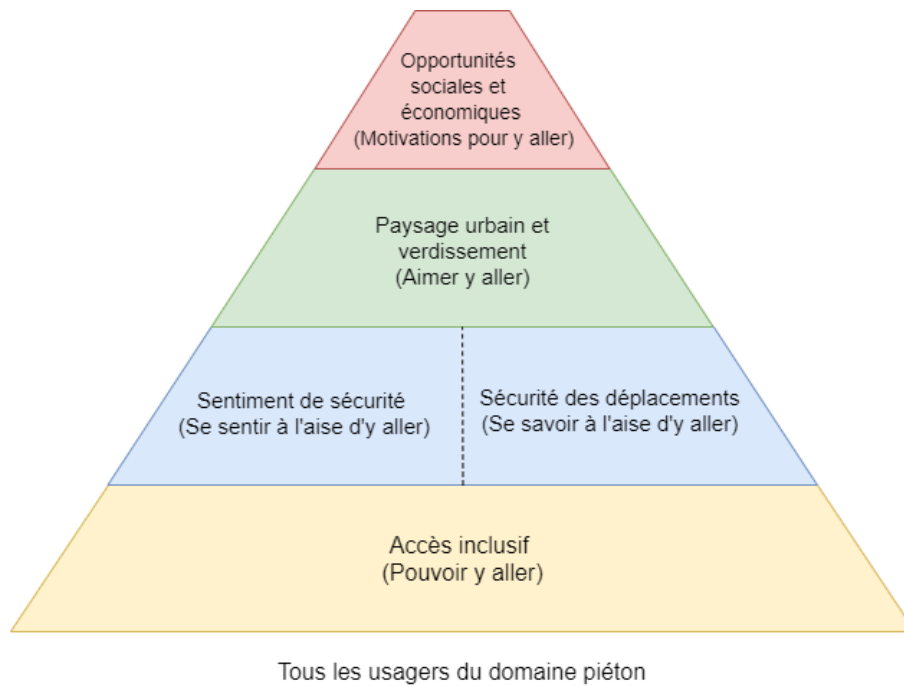


Figure 2: Facteurs influençant le choix d'un cheminement piéton. Adapté de Gendron (2014, p.119).

La conception universelle est une discipline qui s'intéresse à la création d'objets et d'environnements esthétiques et fonctionnels permettant un accès inclusif (Fougeyrollas *et al.*, 2019). L'accès inclusif, tel que proposé par Rocques *et al.* (2011) et adapté par Fougeyrollas *et al.* (2019) est :

[...] la caractéristique d'un produit, procédé, service, information ou composante de l'environnement physique et social qui, dans un but d'équité et de respect des droits humains, permet à toute personne de réaliser des activités de façon autonome et d'obtenir des résultats équivalents. (Fougeyrollas *et al.* 2019, p. 171)

Cette définition se rapporte par ailleurs au concept général de l'accès, qui s'exprime quant à lui par cinq dimensions : la disponibilité, l'accessibilité, l'acceptabilité, l'abordabilité et l'utilisabilité (Figure 3). Le présent projet s'intéresse particulièrement à la dimension de l'accessibilité, c'est-à-dire l'« état des composantes physiques spatiales, architecturales et technologiques de l'environnement permettant, selon leurs capacités ou préférences, à une personne ou à un groupe de population de réaliser leurs activités » (Fougeyrollas *et al.* 2019, p.167). Cette dimension intègre, entre autres, les normes de conception sans obstacle.

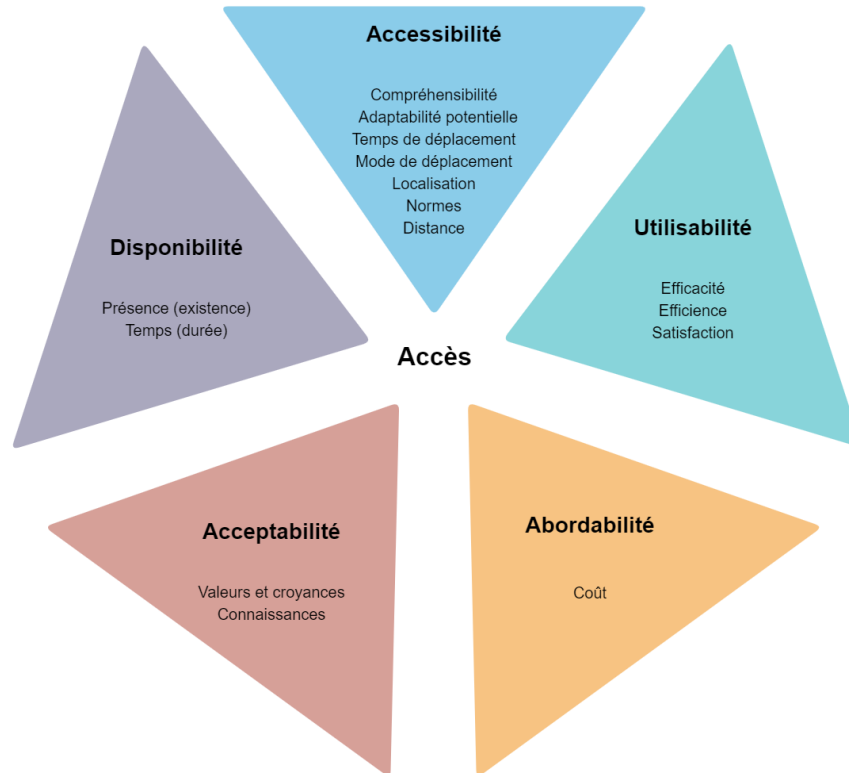


Figure 3: Les cinq dimensions relatives à l'accès. Adapté de Fougeyrollas et al. (2019, p.169)

Guides normatifs de conception sans obstacle

De nombreux pays comme le Canada (Ville de Toronto, 2004; Ville de Québec 2010), les États-Unis (Beneficial Designs Inc., 2001), la Suisse (Centre suisse pour la construction adaptée aux handicaps, 2003; Moro, 2007) et la Belgique (Association Nationale pour le Logement des Handicapés, 1999; Services publics de Wallonie, 2006, Ministère de la Région de Bruxelles-Capitale, 2008; Centre de recherches routières, 2009) proposent des guides normatifs sur la conception sans obstacle qui ciblent les besoins spécifiques des PMR et établissent des balises pour réduire les barrières physiques dans l'environnement urbain (Figure 4). Dans ces guides, les personnes en fauteuil roulant ressortent comme la catégorie d'utilisateurs confrontée au niveau de contraintes le plus élevé dans leurs déplacements. Pour cette raison, ces utilisateurs sont d'un intérêt particulier dans la mise en œuvre d'une politique d'accessibilité universelle et font l'objet de la présente étude.

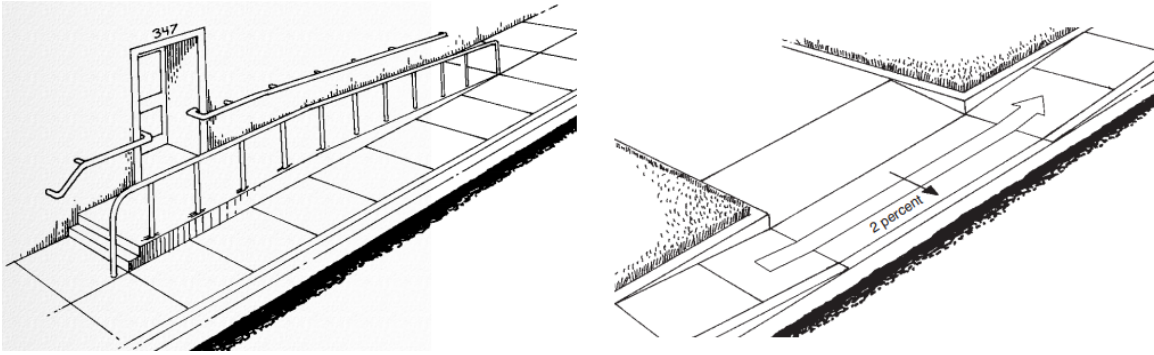


Figure 4: Exemples d'aménagements accessibles en lien avec les trottoirs (Beneficial Designs Inc., 2001, p.4-16 à 4-27)

En dépit de balises claires qui figurent dans ces guides, des études cliniques (Duvall *et al.*, 2013; Wolf *et al.* 2007) démontrent que la recommandation par les guides normatifs d'une surface lisse, stable et antidérapante n'est pas suffisamment précise et fait place à l'interprétation en ce qui a trait, par exemple, au seuil minimal toléré pour la qualité du revêtement de surface (Duvall *et al.*, 2013).

De surcroît, malgré l'impact reconnu de l'état du revêtement sur l'accessibilité des trottoirs, l'importance d'un suivi régulier de la détérioration de ces infrastructures n'est énoncée dans aucun des guides consultés. Cette lacune souligne la nécessité d'élaborer des normes non seulement en matière de conception des infrastructures, mais également au niveau de leur auscultation après qu'elles aient été construites.

Enfin, le cadre de référence du projet actuel, en dépassant les exigences généralement prescrites par les guides normatifs, se démarque d'autres initiatives similaires.

Technologies : pas de solutions adaptées sans données

Diverses plateformes Web permettent la planification de trajets adaptés. Dans les dernières années, de tels projets ont vu le jour en France (Île-de-France, année inconnue), en Allemagne (Mobasheri *et al.*, 2017), en Suisse (Mack, 2017) et à Québec par une équipe de l'Université Laval (Mostafavi, 2014). En mars 2019, Google a également annoncé l'ajout d'une fonctionnalité permettant la création de trajets dits accessibles dans sa populaire plateforme *Google Maps* pour six grandes villes à l'international (Akasaka, 2018). À terme, ces différentes solutions pourraient grandement faciliter les déplacements des personnes en fauteuil roulant. Cependant, les données, permettant la planification de trajets adaptés, sont limitées. Si celles-ci permettent aux usagers d'éviter des zones à forte

pente ou d'assurer l'accessibilité d'un arrêt d'autobus, par exemple, la prise en compte de variables comme l'état de revêtement des trottoirs s'avère plus ardue.

En effet, l'intégration de cette variable nécessite un suivi récurrent de la dégradation des infrastructures. Au Québec, les méthodes permettant ce type d'auscultation impliquent le recours à des firmes d'entrepreneurs privés, comme SNC-Lavalin inc., qui possèdent l'expertise et l'équipement spécialisé pour cette tâche (Figure 5). Ces équipements permettent d'évaluer au millimètre près la dégradation du revêtement, mais nécessite de parcourir à basse vitesse l'ensemble des infrastructures visées par une auscultation. À titre d'exemple, la Ville de Montréal a ausculté l'ensemble de son réseau de trottoirs en 2019, au coût estimé de 500 000 \$ (Labbé, 2018). Montréal figure toutefois parmi les rares municipalités qui peuvent dédier une enveloppe budgétaire afin d'effectuer ce type d'étude. De plus, lorsqu'elles sont effectuées, leur faible récurrence rend rapidement désuètes les dernières données acquises. À Montréal, l'auscultation des trottoirs n'avait pas été effectuée depuis 2010.



Figure 5: Deux véhicules utilisés lors de l'auscultation traditionnelle de trottoir. Tiré de Blanchette (2011, p.30).

À Sherbrooke, quoique la Ville ait recours à l'auscultation systématique par camion pour le réseau routier (Dubé, 2018), l'évaluation de l'état des trottoirs dépend, entre autres, des plaintes formulées à la Ville (Ville de Sherbrooke, 2018). Une approche par plainte, moins coûteuse qu'un relevé systématique par des entrepreneurs spécialisés, limite cependant l'acquisition de données aux zones fréquentées par les PMR. Par cette approche, les données obtenues sur l'état des infrastructures demeurent donc fragmentaires et ne permettent pas d'établir un bilan pour l'ensemble du réseau piétonnier. Enfin, cette

approche est empreinte d'une grande subjectivité, car elle s'en remet à la perception des usagers plutôt qu'à l'état réel des infrastructures.

1.3.2. Traitement d'image par apprentissage profond

L'apprentissage profond (*deep learning*) est une forme d'intelligence artificielle, qui découle des efforts déployés depuis les années 1950 pour automatiser les tâches intellectuelles habituellement effectuées par des humains (Chollet, 2017a). L'apprentissage machine (*machine learning*), une forme d'intelligence artificielle qui englobe l'apprentissage profond, exploite des algorithmes qui requièrent généralement trois éléments fondamentaux : 1) des données en entrée (ex. : nombres, texte, bandes sonores, images, vidéos), 2) des exemples du résultat attendu et 3) une mesure d'efficacité permettant de quantifier la performance de l'algorithme et d'en ajuster les paramètres afin de maximiser cette mesure. L'apprentissage machine s'écarte donc du paradigme de la programmation classique qui cherche à produire des résultats à partir de données et de règles explicitement programmées (voir Figure 6).

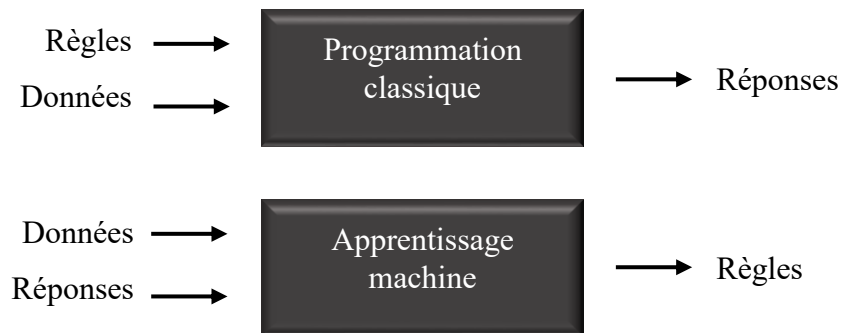


Figure 6: Changement de paradigme entre la programmation classique et l'apprentissage machine. Adapté de Chollet (2017a, p.5)

L'apprentissage profond fait appel à une série d'algorithmes appelés « réseaux de neurones profonds ». Ceux-ci sont dotés d'un grand nombre de couches, conférant une « profondeur » au réseau, impliquant de nombreux neurones et donc de paramètres matérialisant le modèle (Figure 7).

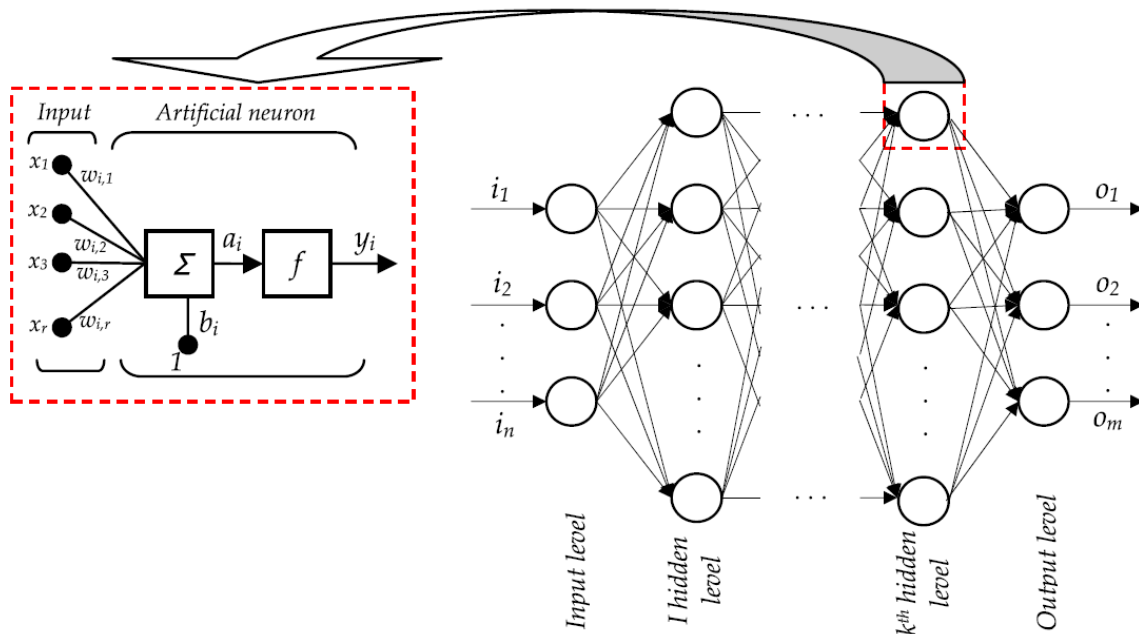


Figure 7: Composition d'un neurone, l'unité computationnelle de base des couches d'un réseau. Chaque valeur d'entrée x_1, x_2, \dots, x_r est multipliée par un poids $w_{i,1}, w_{i,2}, \dots, w_{i,r}$. Un coefficient b_i , qui s'apparente à l'ordonnée à l'origine dans une équation linéaire, s'ajoute à ces poids. Une fonction d'activation f est ensuite appliquée à cette somme pondérée a_i , produisant la valeur de sortie y_i . (tiré de Tanikic, 2012, p.165)

Trois principaux facteurs expliquent la popularité actuelle de l'apprentissage profond :

- 1) la taille croissante des jeux de données à l'ère du *Big Data* ;
- 2) le décuplement des capacités de calcul des unités de traitement graphique ;
- 3) l'amélioration de l'architecture et des méthodes d'entraînement des réseaux de neurones (Goodfellow et al., 2016).

Réseaux de neurones artificiels : architectures et apprentissage automatique

L'architecture des réseaux de neurones comprend deux ramifications principales : les réseaux séquentiels (*feedforward*) à couches multiples et les réseaux récurrents. Le premier type d'architecture est davantage appliqué en vision par ordinateur (*computer vision*), tandis que les réseaux récurrents sont particulièrement utiles pour des applications qui traitent des séquences de texte ou de caractères. Étant donné que ce projet fait appel à des images, la présente section abordera seulement le premier type d'architecture.

Les réseaux séquentiels constituent un type de réseau composé de couches hiérarchiques, où les résultats à la sortie de chaque couche servent d'entrée pour la couche suivante. Au

terme de ce processus, l'algorithme dégage une hiérarchie de concepts ou de représentations qui ont un niveau de complexité et d'abstraction croissant (Goodfellow *et al.*, 2016). Par exemple, un réseau de neurones séquentiel peut extraire, dans les premières couches, des caractéristiques pertinentes (*features*) de l'image comme la présence de contours, motifs et formes. À partir de cette transformation, le réseau peut réaliser différentes tâches pour lesquelles il est programmé : associer une classe à une image, à une région d'intérêt, à chaque pixel d'une image, etc. (Figure 8). Toutefois, avant d'atteindre les performances souhaitées, le réseau doit se soumettre au processus d'apprentissage automatique.

L'apprentissage automatique est l'« ajustement » automatique de l'algorithme permettant d'extraire la meilleure correspondance possible entre les données initiales et la réponse produite (*mapping*). Le succès des réseaux de neurones profonds dépend largement de l'automatisation de l'apprentissage, car il évite le travail laborieux et souvent peu robuste qui consiste à définir manuellement des caractéristiques pertinentes pour des tâches de reconnaissance visuelle (*feature engineering*) (Chollet 2017a; Zhu *et al.*, 2017).

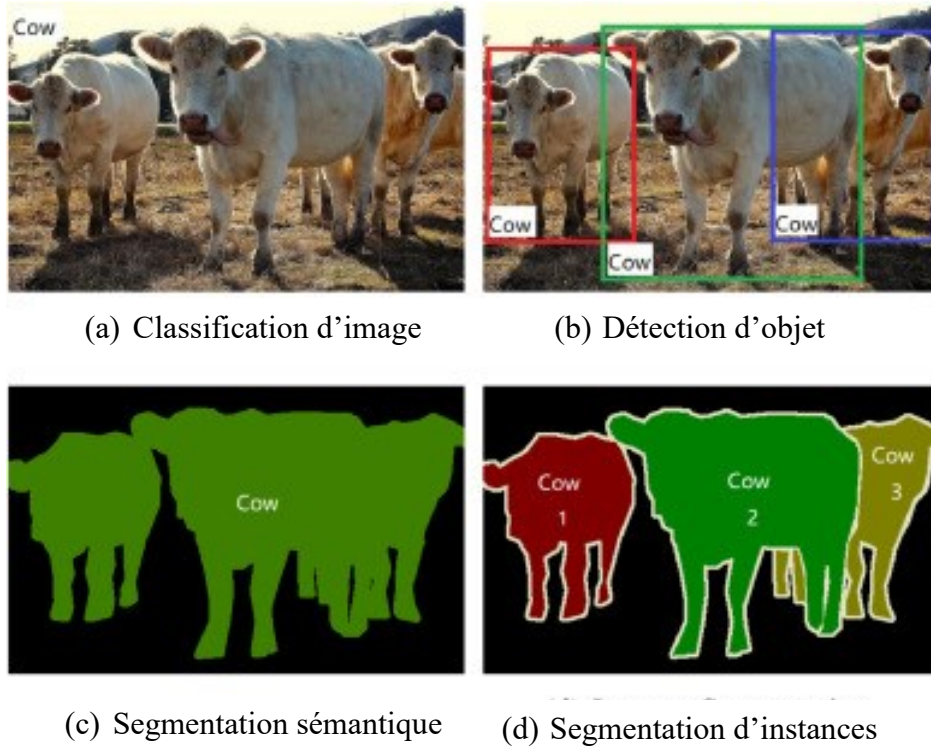


Figure 8: Quatre tâches courantes en vision par ordinateur (Tiré de Wu et al. 2020, p.39)

L'apprentissage d'un réseau de neurones procède par boucles itératives nommées « époques », où chaque époque représente une itération à travers toutes les données d'entraînement. D'abord, une valeur initiale, aléatoire ou transférée d'un réseau entraîné (voir section suivante), est fixée pour chacun des poids. Le réseau produit un premier ensemble de prédictions à partir des données d'apprentissage (*forward pass*). Puis, l'algorithme mesure une fonction de perte (*loss function*) telle que l'erreur quadratique moyenne (*Mean Squared Error*) en comparant la valeur obtenue avec la cible réelle. À la troisième étape, un processus fondamental en apprentissage profond intervient, soit : la rétropropagation (*backpropagation*). Cette étape, effectuée automatiquement par les logiciels en apprentissage profond, permet de calculer un gradient d'erreur qui indique la proportion de l'erreur attribuable à chaque poids (Bengio *et al.* 2015). La dernière étape du processus d'apprentissage permet d'ajuster la valeur des poids de chaque neurone en les déplaçant proportionnellement à la magnitude du gradient calculé, mais en direction opposée à celui-ci. Cette étape s'appelle l'optimisation.

Au fur et à mesure que les poids sont réajustés, l'algorithme tente donc de minimiser l'écart entre les valeurs prédites et les valeurs réelles. Au moment de l'optimisation, une

fraction seulement du gradient d'erreur mesuré est considérée dans la mise à jour de la valeur des poids. Cette proportion est appelée le taux d'apprentissage. Le processus d'entraînement s'arrête lorsque la fonction de perte est suffisamment faible ou lorsque le nombre d'époques atteint le seuil prédéfini. Selon François Chollet (2017a), un membre de l'équipe *Google Brain*, l'algorithme d'optimisation est la clef de voûte des méthodes d'apprentissage profond.

Lorsque l'apprentissage s'effectue sur un jeu de données restreint, le processus d'optimisation a tendance à mener au surapprentissage (*overfitting*). En effet, à mesure que la fonction de perte diminue par rapport aux données d'apprentissage, le modèle peut commencer à apprendre (par cœur) des motifs spécifiques aux données d'entraînement, ce qui nuit à sa capacité de généralisation (Chollet, 2017a). Les efforts de recherche pour tenter de pallier ce problème ont mené à plusieurs débouchés comme des techniques de régularisation lors de l'apprentissage (Guo *et al.* 2016; Hinton *et al.* 2012).

Applications en traitement d'image

Depuis les résultats exemplaires obtenus par l'équipe de M. Geoffrey Hinton au concours *ImageNet Large-Scale Visual Recognition Challenge* (ILSVRC) (Russakovsky *et al.*, 2015) en 2010 (Krizhevsky *et al.* 2012), le recours aux réseaux de neurones convolutifs (CNN) en vision par ordinateur (Chollet 2017a; LeCun *et al.* 2015; Guo *et al.* 2016) et en télédétection (Zhang *et al.*, 2016a; Dahmane *et al.*, 2016; Penatti *et al.*, 2015) ne cesse de croître.

En effet, le CNN de l'équipe de l'Université de Toronto a obtenu un taux d'erreur du top 5 de 15,3 %, c'est-à-dire que la véritable classe de l'image figurait dans une des 5 premières classes proposées par le réseau à l'exception de 15,3 % des exemples de test. Cette victoire, la première pour un CNN à ce concours, creusait l'écart avec le taux d'erreur du meilleur résultat précédent à plus de 10 %, démontrant la robustesse et l'efficacité de ce type de réseau.

Depuis 2010, les architectures de réseaux soumises à ce même concours ont produit des résultats de plus en plus probants (Zeiler et Fergus, 2013; Simonyan et Zisserman, 2015, Szegedy *et al.* 2015; He *et al.* 2016; Huang *et al.* 2017), faisant chuter le taux d'erreur sur

le jeu de données *ImageNet* (Russakovsky *et al.*, 2015) à moins de 5 % (voir Figure 9) (Gershgorn 2017).

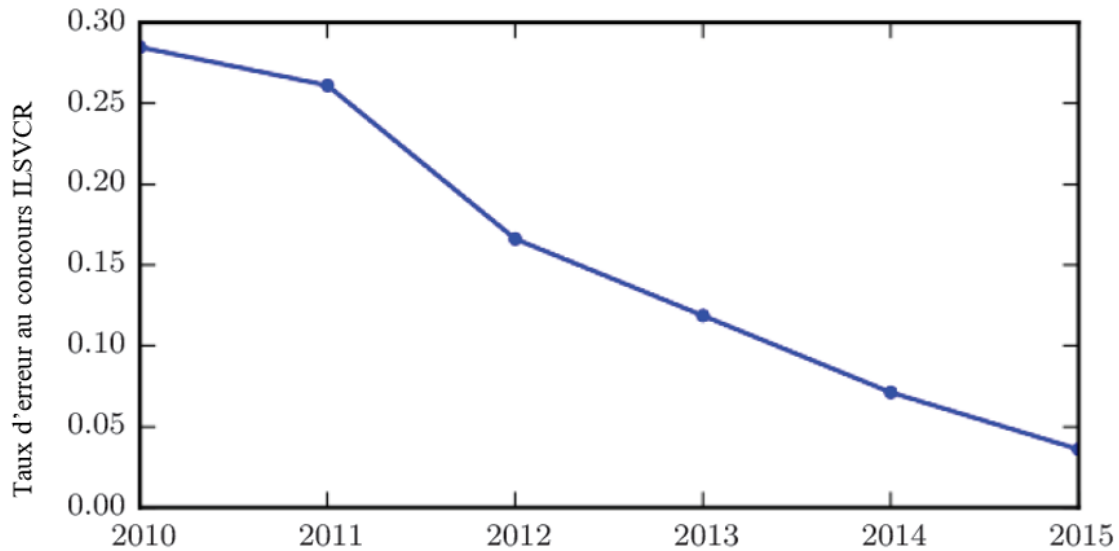


Figure 9: Taux d'erreur obtenus par les réseaux de neurones convolutifs de 2010 à 2015 lors du concours *ImageNet Large Scale Visual Recognition Challenge (ILSVCR)*. Le concours a pris fin en 2017 étant donné que la problématique était considérée résolue (tiré de Goodfellow *et al.*, 2016, p.24)

Les CNN, un type de réseaux de neurones généralement séquentiel, ont maintes fois démontré leur supériorité à s'attaquer à des problèmes perceptuels comme la reconnaissance visuelle par rapport à d'autres méthodes d'apprentissage machine, comme les arbres de décision, la régression logistique et les machines vectorielles de soutien (SVM). L'opération de convolution auquel font appel les CNN consiste en un produit scalaire de deux matrices, celle formée par les valeurs numériques des pixels de l'image, s'il s'agit de la première couche du réseau, et celle formée par les valeurs d'un filtre de convolution (Figure 10). L'utilisation de la convolution dans les CNN permet de détecter des motifs souvent répétitifs sur les images tout en réduisant le nombre de paramètres constituant le réseau et le temps de calcul.

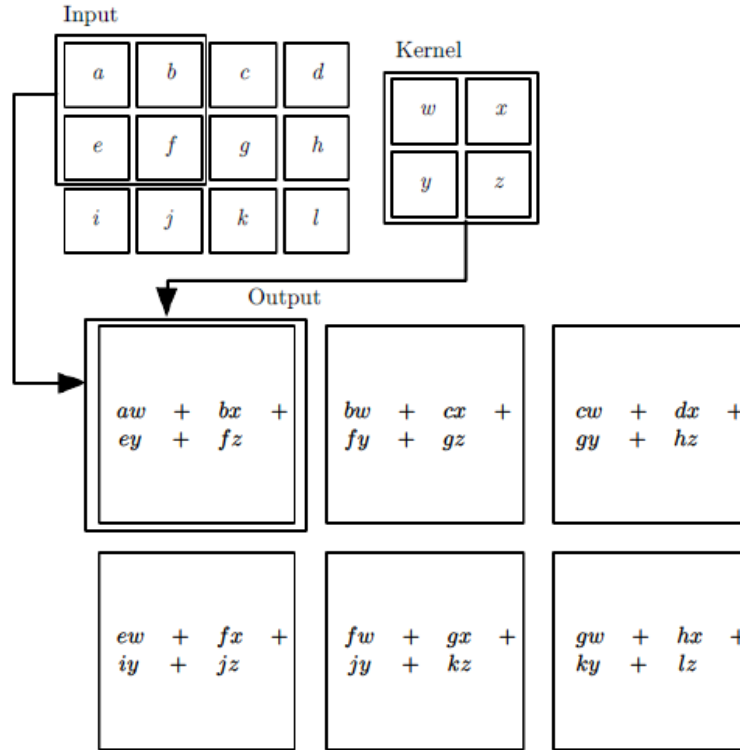


Figure 10: Exemple d'une convolution 2D. En haut, les deux matrices qui servent à effectuer le produit scalaire. En bas, six sorties possibles résultant du déplacement du filtre de convolution sur les valeurs d'entrée (input) (Tiré de Goodfellow et al. 2016, p.327)

Architectures de CNN récentes

Depuis le succès de l'architecture *AlexNet*, premier réseau de neurones profond proposé à l'épreuve ILSVRC (Krizhevsky *et al.* 2012), de nombreuses architectures ont été élaborées et proposées pour lui succéder. Parmi celles-ci, de nombreux CNN avec un nombre de paramètres et une complexité souvent inégalés ont retenu l'attention et ont émergé comme des modèles repoussant l'état de l'art. Canziani *et al.* (2018) a effectué une analyse approfondie des performances de plusieurs CNN et autres réseaux de neurones et a comparé la taille de chaque modèle selon la quantité de paramètres qui le compose (Figure 11). Cette analyse permet de constater que les modèles de classification récents comme *Xception* (Chollet, 2017b) dépassent largement les performances d'*AlexNet*, tout en réduisant considérablement le nombre de paramètres qui les composent.

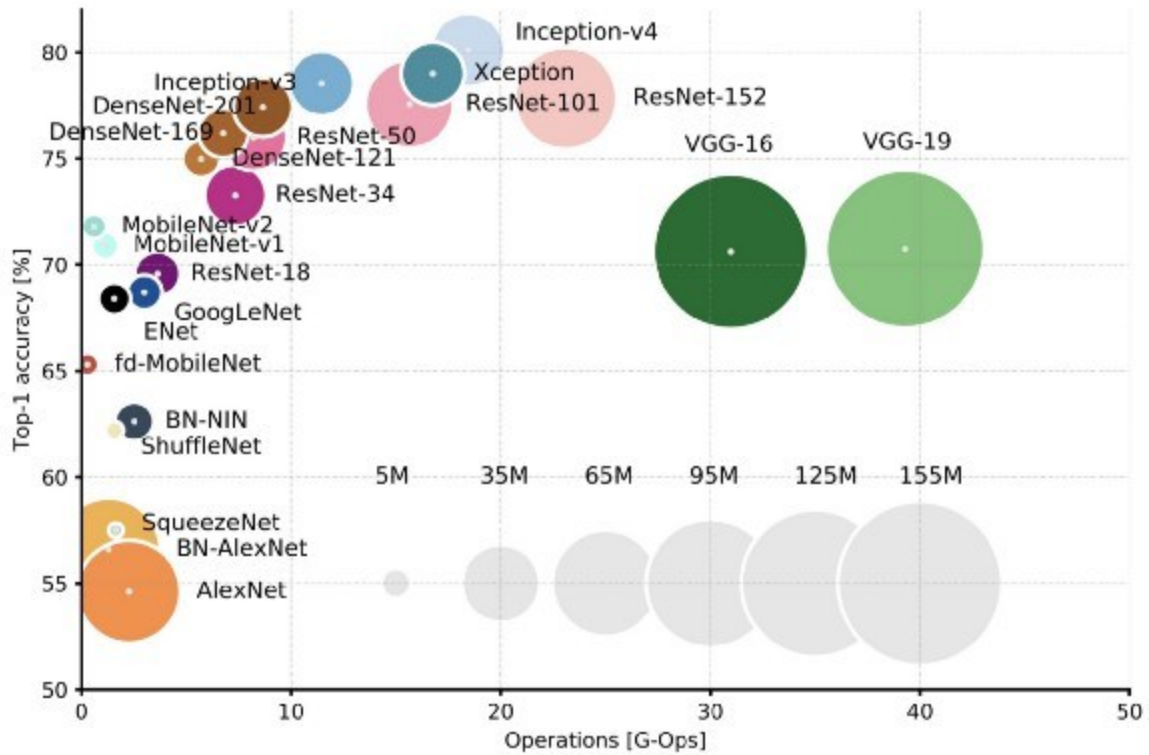


Figure 11: Comparaison de plusieurs réseaux de neurones profonds issus de l'état de l'art selon leur performance sur ImageNet, le nombre d'opérations que nécessite l'inférence et le nombre de paramètres qui les constitue. (tiré de Canziani, 2018)

Dans le cas de la segmentation sémantique, des architectures spécifiques à cette tâche sont développées afin de faire porter la prédiction sur chaque pixel plutôt que pour l'image entière, dans le cas de la classification. Ces architectures conservent tout de même plusieurs points communs avec leurs cousines élaborées pour la classification. Ainsi, des modèles comme *FCN* (Long *et al.*, 2015) et *Deeplabv3+* (Chen *et al.*, 2018) recyclent notamment des architectures de classification afin d'assurer le traitement « sémantique » d'une image. Par exemple, tel que présenté dans la section 2.5, *Deeplabv3+* intègre le réseau de classification *Xception* (Chollet, 2017b) comme « encodeur ».

Les encodeurs constituent les premières couches de traitement des images dans un contexte de segmentation sémantique. Ils en extraient des représentations pertinentes et simplifiées qui permettent à un décodeur, en aval, de projeter ces représentations sur l'ensemble des pixels de l'image originale, passant d'une information de basse résolution (sortie de l'encodeur) à une prédiction dense, de haute résolution (sortie du décodeur). En d'autres mots, l'encodeur permet au réseau de développer une « compréhension » minimale des

pixels de l'image en entrée pour la tâche en cours tandis que le décodeur utilise cette analyse afin de prédire la classe à attribuer au niveau de chaque pixel de l'image entière, en sortie du réseau (Figure 12). L'approche par encodeur-décodeur, bien qu'elle ne soit pas unique, est souvent adoptée dans les architectures de segmentation sémantique récentes et en a plus d'une fois assuré le succès.

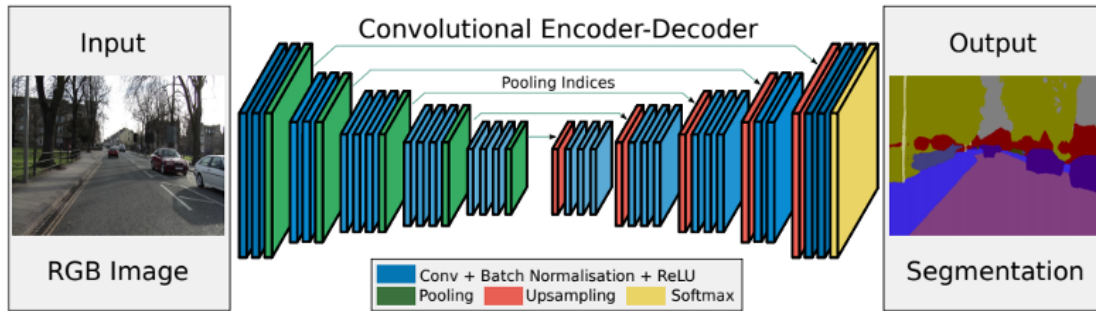


Figure 12: Structure de SegNet, une architecture de type encodeur-décodeur utilisée en segmentation sémantique (tiré de Badrinarayanan et al., 2017, p.4)

En 2018, *Deeplabv3+* se situe premier parmi les modèles au code ouvert et documenté dans le classement de *Cityscapes* (Cordts et al., 2016) (voir sous-section suivante), le jeu de données le plus rapproché de la problématique visée par le présent projet. Par conséquent, cette architecture est retenue pour la segmentation des trottoirs sur les panoramas *GSV* du projet.

Lorsque la revue de littérature a été effectuée, l'architecture *MobileNetV2* (Sandler et al., 2018) se démarquait comme le modèle mobile avec une performance inégalée sur les jeux de données publics *ImageNet* (Russakovsky et al., 2015) et *COCO* (Lin et al., 2014) (voir sous-section suivante).

Jeux de données et performances des CNNs

À l'heure actuelle, la disponibilité de jeux de données étiquetées suffisamment volumineux pour entraîner un CNN à une tâche spécifique demeure la principale pierre d'achoppement dans la recherche et le développement de nouvelles applications liées à l'apprentissage profond (Lawrence, 2015; Zhang et al., 2016a; Lam et al., 2018; Dahmane et al., 2016). En effet, un jeu de données de petite taille mène souvent au surapprentissage. En télédétection, certains jeux de données existent pour des applications spécifiques, dont

celles liées à l'imagerie satellitaire (Penatti *et al.*, 2015; Van Etten *et al.*, 2018; Demir *et al.*, 2018; Lam *et al.*, 2018). Toutefois, le développement de nouveaux jeux de données adaptés à une vaste gamme de problématiques dans ce domaine demeure un besoin important (Lam *et al.*, 2017).

Dans ce contexte, deux stratégies sont fréquemment utilisées : 1) augmentation des données, 2) apprentissage par transfert (Yosinski *et al.*, 2014; Zhang *et al.*, 2016a; Dahmane *et al.*, 2016; Penatti *et al.*, 2015).

D'une part, l'augmentation des données, une méthode largement utilisée (Simard *et al.* 2003; Krizhevski *et al.* 2012; Zhu *et al.* 2017), consiste à multiplier la taille de l'ensemble d'entraînement en effectuant des modifications à ses échantillons (ex. : rotation, flou, changement de contraste, d'échelle ou de luminosité et transformation projective), sans en altérer la classe d'appartenance (Figure 13).

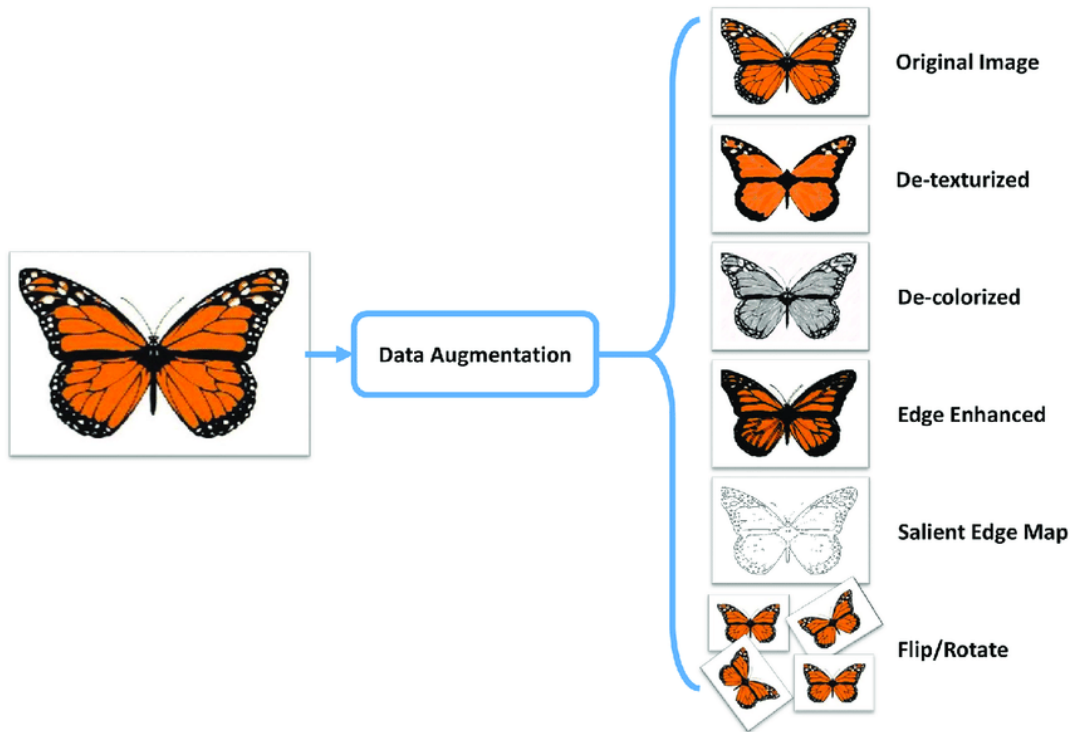


Figure 13: Exemple de différentes techniques d'augmentation de données permettant de réduire le surapprentissage (tiré de Ahmad *et al.* 2017, p.6)

D'autre part, l'apprentissage par transfert permet de réutiliser les poids d'un réseau préentraîné sur un jeu de données comme *ImageNet* (Russakovsky *et al.*, 2015). Ces

paramètres servent ensuite de point de départ pour l'entraînement de réseaux visant à accomplir des tâches ayant une similitude plus ou moins importante avec la tâche initiale (Yosinski *et al.* 2014; Razavian *et al.* 2014; Guo *et al.* 2016).

Selon la revue de littérature effectuée, il n'existe aucune publication répertoriée sur la caractérisation automatique des trottoirs par apprentissage profond. Ainsi, le cadre théorique et la méthodologie du présent projet se sont basés sur des articles publiés dans un domaine apparenté, celui de la détection de la dégradation des chaussées (*pavement distress detection*). Les plus récentes publications dans ce domaine (Some 2016; Zhang *et al.*, 2016b; Gopalakrishnan *et al.*, 2017; Gopalakrishnan *et al.*, 2018; Du *et al.*, 2020) évoquent clairement les avantages liés à l'apprentissage profond, en particulier ceux des CNN.

Plus encore, plusieurs publications (Zhang *et al.*, 2016b; Gopalakrishnan *et al.*, 2017) démontrent la supériorité des CNN en comparaison avec d'autres algorithmes d'apprentissages machines (*Random Forest, SVM, Boosting*), en particulier lorsque ces algorithmes sont utilisés pour évaluer des jeux de données avec une grande variabilité au niveau du type de revêtement, des conditions d'éclairage et des types de défauts à détecter (orientation des fissures, variabilité morphologique). Pour ces raisons, les réseaux de neurones convolutifs sont également privilégiés pour le projet.

Dans ce domaine, il n'existe actuellement pas de consensus permettant de privilégier l'entraînement d'un réseau en entier (Zhang *et al.*, 2016b; Pandey et Dholay, 2019) ou l'approche de transfert d'apprentissage (Some 2016, Gopalakrishnan *et al.* 2017). Toutefois, étant donné les bénéfices généralement attribués au transfert d'apprentissage (*transfer learning*) (Yosinski *et al.*, 2014), cette technique a été retenue dans le cadre du projet.

1.4. Objectifs et hypothèse de recherche

L'objectif principal est de développer une méthode permettant de qualifier l'état du revêtement des trottoirs à partir d'images panoramiques urbaines à haute résolution.

1.4.1. Objectifs secondaires

1. Recourir aux techniques d'apprentissage profond dirigé afin de (1) segmenter les trottoirs parmi d'autres éléments d'une scène urbaine et (2) caractériser l'état du revêtement, sur des images panoramiques du type *Google Street View* ;
2. Évaluer la robustesse du modèle d'apprentissage profond développé en l'appliquant au site d'étude ;
3. Dégager les avantages, les inconvénients et les limitations associées à l'apprentissage profond dans la caractérisation des trottoirs à partir des images panoramiques du type *Google Street View*.

1.4.2. Hypothèse de recherche

La classification d'images panoramiques urbaines par apprentissage profond est en mesure de fournir une bonne approximation de l'état du revêtement des trottoirs, en vue de faciliter les déplacements des personnes à mobilité réduite et la planification des travaux d'entretien des trottoirs.

1.5. Pertinence de l'étude

Le présent projet survient alors que plusieurs villes du Québec s'intéressent aux besoins des personnes à mobilité réduite sur leur territoire. Quoique plusieurs plans d'action en accessibilité universelle proposent des solutions afin de faciliter leurs déplacements (Ville de Montréal, 2015; Ville de Sherbrooke, 2016), des lacunes subsistent en ce qui concerne, notamment, la caractérisation de l'état du revêtement des trottoirs (Mostafavi, 2018).

Tel que mentionné à la section 1.3.1 « Conception pour tous [...] », les quelque 45 000 usagers en fauteuil roulant constituent une population particulièrement vulnérable à des aménagements urbains inadaptés. Le projet actuel offre donc une opportunité inédite de mettre la géomatique et la télédétection, deux domaines en forte expansion, au service de cette frange de la population.

À terme, la méthode qui est développée par ce projet pourrait bonifier les algorithmes de planification de trajet en permettant d'ajouter l'état du revêtement comme variable prise en compte lors de l'élaboration de trajets adaptés. Cette méthode a également le potentiel d'enrichir les données d'auscultation de trottoirs que possèdent les villes afin d'aider à l'entretien et à la réfection de ces infrastructures.

1.6. Limites du projet

Le recours au traitement d'images par apprentissage profond ne permet pas de caractériser l'ensemble des propriétés des trottoirs. Ainsi, toutes les variables ne seront pas prises en compte. Les variables liées notamment à l'état de la surface (uniformité de la surface), la hauteur des seuils de l'accès universel (abaissement du trottoir ou bateau-pavé) de même que les pentes longitudinales et transversales exigent des données altimétriques fines qu'il n'est pas possible d'extraire d'une image numérique (du moins, du type de celles utilisées dans ce projet).

De plus, le présent projet porte uniquement sur un type de corridor piétonnier : les trottoirs. Quoique les traverses piétonnes, les sentiers aménagés et non aménagés et les accotements de route sont également visés par la problématique du projet, ces corridors piétonniers sont exclus pour des raisons techniques. En effet, la délimitation de plusieurs de ces corridors piétonniers est plus difficile à extraire par traitement d'images automatique.

Chapitre 2. Méthodologie

2.1. Données utilisées

Les photos panoramiques 360° *Google Street View* (GSV) constituent une source de données privilégiées afin d'aborder la problématique de la caractérisation des corridors piétonniers pour les raisons suivantes :

- couverture géospatiale référencée de villes de 20 pays sur quatre continents ;
- résolution spatiale supérieure à toute image de télédétection conventionnelle;
- tuiles de panoramas prémosaïquées et prétraitées pour atténuer les délimitations entre chaque image (Anguelov *et al.*, 2010) ;
- plusieurs CNN préentraînés sur ce type d'images sont ouverts au public.

Les images de rue *GSV* sont prises à l'aide de véhicules, surmontés de caméras, qui quadrillent les villes du monde entier. La météo et la densité de population sont les deux facteurs principaux qui déterminent l'endroit et le moment de la prise des clichés (Google, 2018).

2.2. Choix des sites d'étude

Une recherche exhaustive afin d'identifier une banque de plusieurs centaines d'images utiles pour l'entraînement de réseaux de neurones en lien avec la tâche de caractérisation des trottoirs s'est montrée infructueuse. Aucune donnée ouverte de grandes villes à proximité du site d'étude (Sherbrooke, Montréal, New York, Boston, etc.) ne présentait les attributs nécessaires pour la constitution d'une telle base de données. Les tentatives pour localiser des données pertinentes issues de plateformes d'approvisionnement par la foule (*crowdsourcing*) ont également été vaines. Ces démarches ont donc démontré l'importance

de monter une banque d'images annotées propre au présent projet, à partir des panoramas *GSV*.

Après avoir effectué une évaluation visuelle préliminaire sur les images *Google Street View* de plusieurs quartiers des villes du Québec, un enjeu important a été identifié : les secteurs avec trottoirs fortement endommagés sont sous-représentés par rapport aux autres secteurs. Ainsi, les recherches de secteurs à viser par l'extraction, en particulier pour les sites d'entraînement, se sont focalisées sur des secteurs propices à une forte détérioration des trottoirs (ex. : quartiers anciens ou défavorisés). Malgré cela, moins de 1 % des images *GSV* (consulté via la plateforme *maps.google.ca*) présentaient des défauts majeurs.

2.2.1. Sites d'entraînement

Des panoramas *GSV* du quartier Ahuntsic-Cartierville à Montréal et de Saint-Hyacinthe (Figure 14 et Figure 15) ont été utilisés afin de constituer les ensembles de données pour l'entraînement des modèles. Ces zones ont été sélectionnées sur la base d'articles de presse (Anonyme, 2018; Martineau, 2018) dénotant une densité élevée de défauts majeurs sur les trottoirs de ces secteurs.



Figure 14: 2953 points d'extraction de panoramas Google Street View dans le quartier Ahuntsic-Cartierville au nord de l'île de Montréal (Qc)



Figure 15: 2455 points d'extraction de panoramas Google Street View à Saint-Hyacinthe (Qc)

Acquisition des panoramas Google Street View

L'acquisition des panoramas *GSV* a été effectuée grâce à deux scripts adaptés de Doersch (2012) et Gronat *et al.*, (2011). Un premier script *JavaScript* permet de trouver un panorama à partir d'une coordonnée tandis que le second, en *Python*, permet d'extraire le fichier *jpg* du panorama, identifié grâce au script précédent, à partir des serveurs de Google. Le processus d'identification des panoramas pour un périmètre donné suit quatre étapes principales :

1. Dans un logiciel SIG comme QGIS, téléchargement des données vectorielles de *Open Street Maps* pour la zone visée par l'extraction ;
2. Conserver uniquement les entités qui correspondent aux routes (lignes) ;
3. Convertir les données sous forme de points avec l'extension *QChainage* (1 point / 5 mètres)
4. Exporter les coordonnées de latitude et longitude pour l'ensemble des points dans le format utilisé pour le script de recherche des panoramas (Tavon, 2020).

Un script *Matlab*, traduit en *Python*, a permis de télécharger les panoramas sous forme de tuiles, assemblées, puis enregistrées sous forme de *jpg*, à partir de la liste générée par l'étape précédente d'identification. Ces panoramas peuvent être obtenus à cinq niveaux de zoom (n°1-résolution minimale, n°5-résolution maximale). Les panoramas avec zoom n°4 répondaient le mieux aux besoins du présent projet, avec une résolution de 6656 x 3328 pixels. À cette résolution, le niveau de détail est appréciable et la taille du fichier (~3.5 Mb par image) permet une lecture rapide par les processus informatiques au moment des entraînements par réseaux de neurones. Des tests préliminaires ont démontré que le gain de détails obtenu par les images au zoom n°5 (13312 x 6656 pixels) était trop marginal pour être utile. De plus, la lecture répétée de ces fichiers volumineux aurait occasionné un ralentissement lors de l'entraînements des réseaux. En outre, en adaptant les scripts originaux, plusieurs métadonnées ont pu être récupérées : latitude, longitude, date, orientation du véhicule, adresse civique la plus proche, etc. Dans l'ensemble, un total de 8408 panoramas *GSV* ont été extraits des serveurs de Google pour couvrir les sites d'entraînement.

Annotation des données

Quoique la Ville de Sherbrooke, entre autres, partage des données ouvertes sur les limites des trottoirs de sa municipalité, ses données n'ont pu être utilisées comme vérité terrain étant donné la nature des panoramas urbains, captées au niveau de la rue. L'échelle de ces panoramas n'est pas orthogonale (un pixel ne représente pas toujours les mêmes dimensions sur l'image) comme une orthophoto, par exemple. Dans ce contexte, l'annotation a été réalisée seulement à partir des images extraites. Deux outils d'annotation ont été utilisés : Cordts (2016), Saint-Charles (2020).

Étant donné que les véhicules *GSV* effectuent une capture à tous les dix mètres de réseau routier approximativement, il existe un chevauchement important entre deux photos contiguës. Afin d'accélérer l'annotation et de guider l'apprentissage des réseaux et à la suite d'une recommandation du Centre de recherche en informatique de Montréal (CRIM), certaines régions dépeintes à la Figure 16, correspondant à des zones éloignées, étaient donc annotées de manière à être ignorées lors de l'entraînement. En effet, ces zones étaient souvent mieux représentées sur d'autres panoramas. Au moment de produire un masque de segmentation pour les trottoirs, les réseaux n'étaient donc pas pénalisés s'ils attribuaient l'une ou l'autre des classes d'entraînement, « non-trottoir » ou « trottoir », à ces pixels.

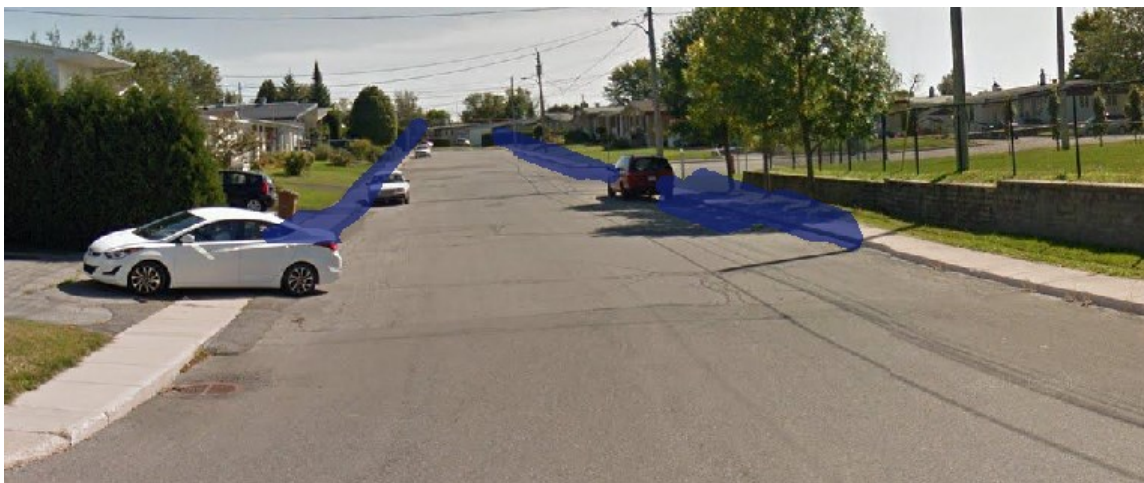


Figure 16: Zones (en bleu) annotées en vue d'être ignorées durant l'entraînement. Celles-ci sont volontairement exclues du calcul de perte, car le chevauchement entre deux panoramas voisins diminue l'importance des prédictions de modèle pour ces zones.

2.2.2. Site d'évaluation finale

La centre-ville de Sherbrooke a été sélectionnée comme site d'évaluation finale, notamment pour sa proximité du lieu de recherche. De plus, la municipalité rend accessibles diverses données portant sur ses infrastructures urbaines (ex. : réseau de trottoirs et routes, mobilier urbain) qui peuvent être utiles dans le contexte de ce projet. Étant donné que la méthodologie du projet fait appel à l'apprentissage profond, les données acquises à partir de ce site d'étude constitueront l'ensemble de test, c'est-à-dire qu'elles serviront uniquement à évaluer les réseaux de neurones entraînés sur des données indépendantes du site d'étude. Ces données ne participeront donc pas à la phase d'entraînement afin d'éviter de biaiser les résultats. Au total, 2609 panoramas ont été acquis pour le site d'évaluation finale (Figure 17). Ces panoramas ont servi à l'évaluation finale des modèles, après entraînement.



Figure 17: 2609 points d'extraction de panoramas Google Street View sur le site d'étude au centre-ville de Sherbrooke (Qc)

2.3. Organigramme méthodologique

Le projet de recherche suit une progression conventionnelle, à partir de la recherche bibliographique, jusqu'à la discussion et la conclusion (Figure 18). Le corps du projet consiste en l'entraînement de réseaux de neurones afin de développer deux modèles de segmentation sémantique. En amont de cette étape centrale figurent l'acquisition des données, ainsi que l'annotation et le prétraitement de la partie servant à l'entraînement des réseaux. En aval, une chaîne de traitement est montée afin d'évaluer la performance des modèles sur le site d'évaluation finale.

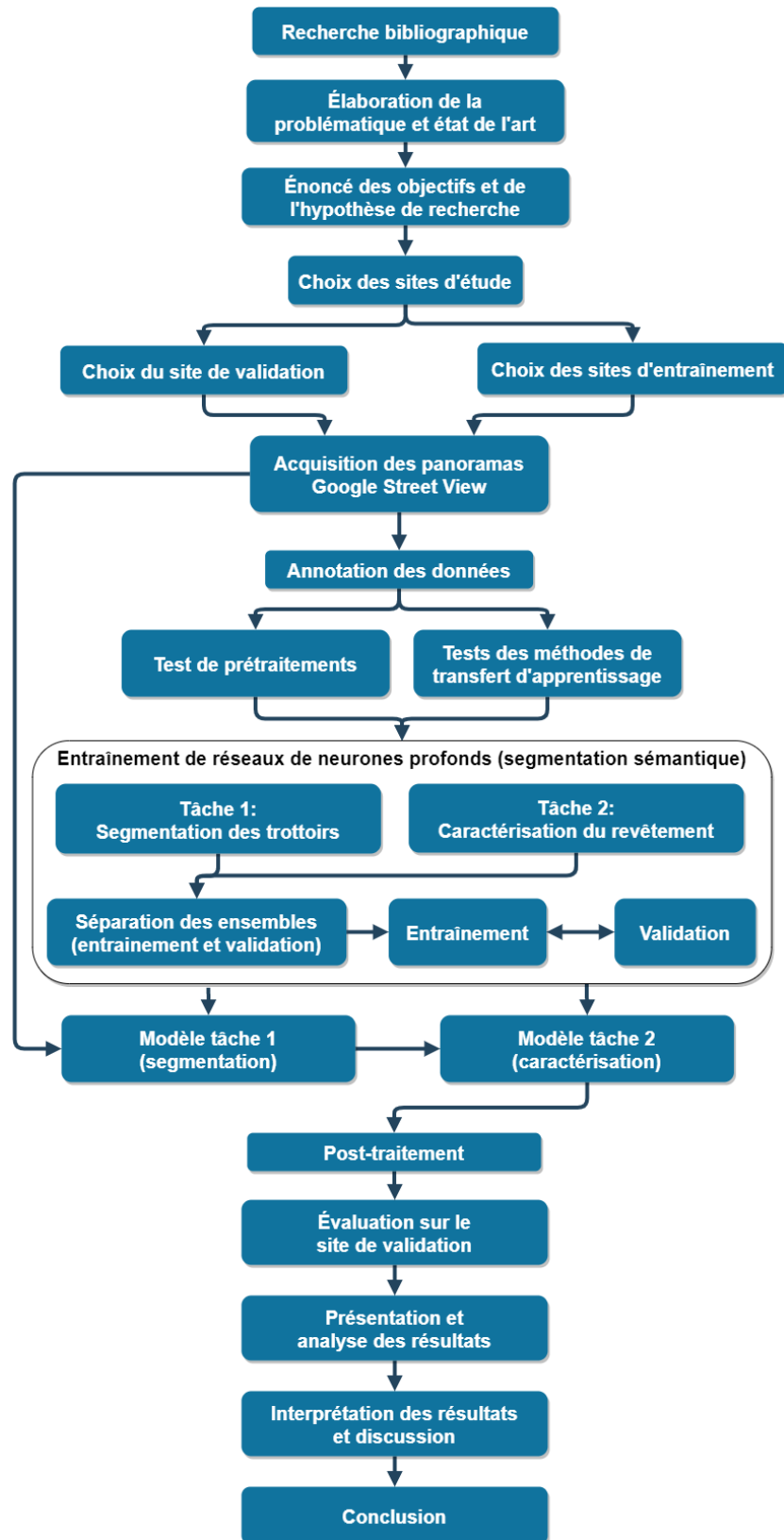


Figure 18: Organigramme méthodologique du projet de recherche

2.4. Sources d'informations

Les informations ayant permis l'élaboration de la problématique, de la méthodologie et de la discussion du présent projet proviennent de différentes catégories de documents : articles de revues scientifiques, articles de presse, politiques publiques, normes de construction et de conception, monographies et sites Web.

Afin de trouver ces documents, les banques de référence consultées sont *IEEE Xplore*, *Computers & Applied Sciences Complete* et *Collection of computer science bibliographies* en plus d'outils de recherche génériques comme *Google* et *Google Scholar*.

Les principales rubriques utilisées durant ces recherches sont les suivantes : traitement d'images, intelligence artificielle, accessibilité universelle, planification urbaine, approvisionnement par la foule (*crowdsourcing*), et villes intelligentes.

2.5. Tests de prétraitement

Considérant que les panoramas *GSV* présentent une perspective 360°, certains éléments, comme les trottoirs, revêtent un aspect distordu (Figure 19) à cause de la projection équirectangulaire utilisée. Afin d'assurer un traitement efficace des images, un prétraitement pour corriger cette distorsion a été envisagé. Au besoin, une correction géométrique afin de récupérer l'aspect normal de la portion centrale de chaque panorama *GSV* aurait été appliquée grâce à un outil comme *Pannellum* (Petroff, 2020).

Une comparaison de la performance du modèle *Deeplabv3+*, retenu pour la segmentation des trottoirs (voir section 2.7 « Choix d'architecture »), préentraîné sur *Cityscapes* (Cordts, 2016) a été effectuée afin d'évaluer l'impact de cette distorsion sur l'affinement de l'apprentissage. Une première série de test a fait appel à 20 images différentes. Si les résultats s'étaient avérés ambigus, d'autres images auraient été ciblés.

Enfin, les résultats de cette comparaison ont permis de décider si un prétraitement s'avérait nécessaire ou non. À titre d'information, les résultats de ce test ont été évalués uniquement visuellement étant donné que la vérité terrain n'avait pas été acquise au moment de sa

réalisation. De plus, ce test a seulement été effectué pour la tâche de segmentation des trottoirs, car aucun réseau préentraîné connu n'effectue la caractérisation du revêtement des trottoirs (voir 2.7.2 « MobileNetV2 »). À la suite des tests, cette correction n'a pas été jugée nécessaire.



Figure 19: Exemple d'un panorama rogné, montrant seulement un côté de la route, avec et sans la distorsion originale causée par la projection équirectangulaire. Sur la photo de gauche, il est inhabituel de percevoir un trottoir avec une telle déformation.

2.6. Comparaison des méthodes de transfert d'apprentissage

Le transfert d'apprentissage permet de tirer profit d'un réseau préentraîné. Ces réseaux bénéficient d'un apprentissage intensif sur une grande quantité de données, permettant d'atteindre un niveau de généralisation élevé. En d'autres mots, les réseaux préentraînés développent des capacités d'extraction qui sont utiles même lorsqu'utilisés pour des tâches différentes de celle visée par l'entraînement initial. En segmentation sémantique, ces entraînements sont souvent effectués sur de grands jeux de données publics tels que *COCO* (Lin, 2014) et *Cityscapes* (Cordts, 2016). Ainsi, la capacité de généralisation des réseaux préentraînés persiste même après que ceux-ci aient été entraînés une seconde fois pour une nouvelle tâche (ex. : segmentation des trottoirs au Québec plutôt qu'en Allemagne pour *Cityscapes*). Différentes approches existent en transfert d'apprentissage en fonction de la proximité de la nouvelle tâche avec la tâche d'origine.

2.6.1. Extraction de caractéristiques

Dans sa plus simple expression, le transfert d'apprentissage permet d'extraire des cartes de représentations (*feature maps*) en amont de la couche finale de classification appartenant à la tâche d'origine. Cette approche est à privilégier lorsque les tâches d'origine et de destination sont semblables. Le cas échéant, une nouvelle couche de classification terminale peut être entraînée selon les classes fixées par la tâche cible (Figure 20). Par exemple, un réseau entraîné sur les 19 classes de *Cityscapes* peut être réaffecté à reconnaître uniquement les trottoirs, dans un contexte géographique différent, sans que l'ensemble du réseau soit réentraîné. Cette approche est possible lorsque les caractéristiques apprises sur le jeu de données original peuvent être recyclées telles quelles pour la tâche cible.

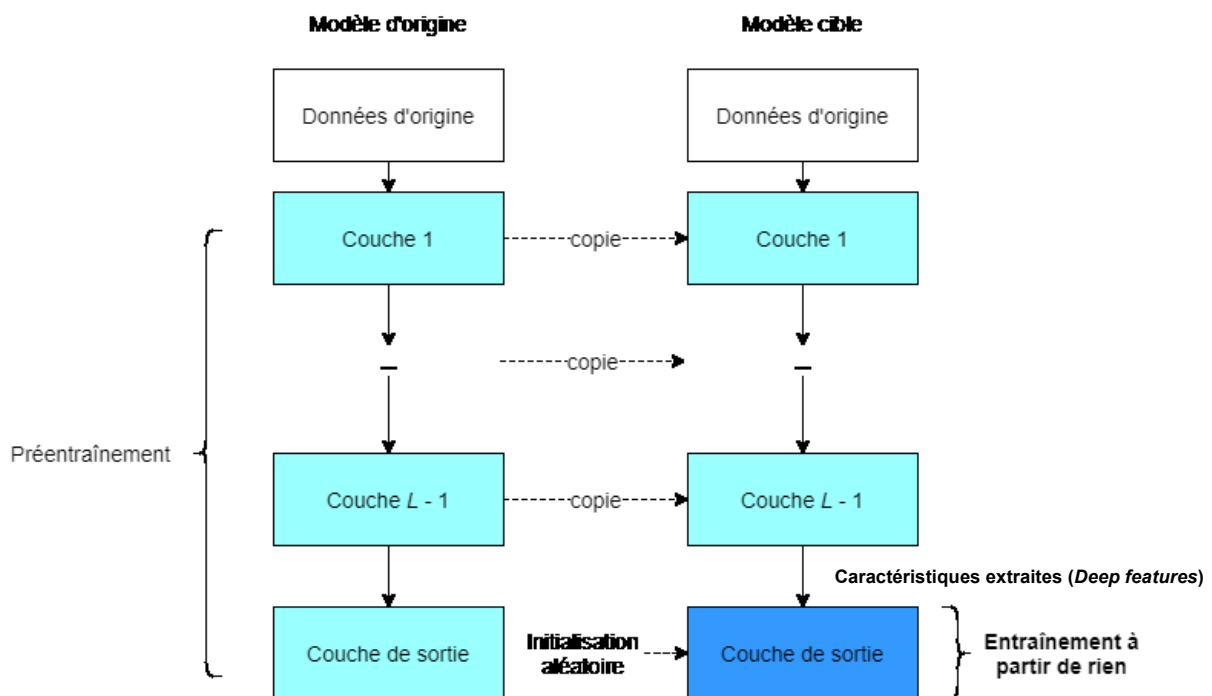


Figure 20: Transfert d'apprentissage de type « Extraction de caractéristiques ». Seule la couche de sortie est réentraînée pour la tâche cible.

2.6.2. Affinement de l'apprentissage

Une autre approche de transfert d'apprentissage, l'affinement de l'apprentissage (*fine tuning*) (Figure 21), se fait en deux temps. Une nouvelle couche de classification finale est

d'abord réentraînée alors que l'ensemble des couches en amont est maintenu gelé. Puis, l'ensemble des couches, incluant celles qui étaient d'abord gelées, est réentraîné avec un taux d'apprentissage très faible (1-10 % du taux d'origine). Cette approche, gourmande en ressources et nécessitant un jeu de données de destination de plus grande taille, permet généralement d'atteindre de meilleurs résultats.

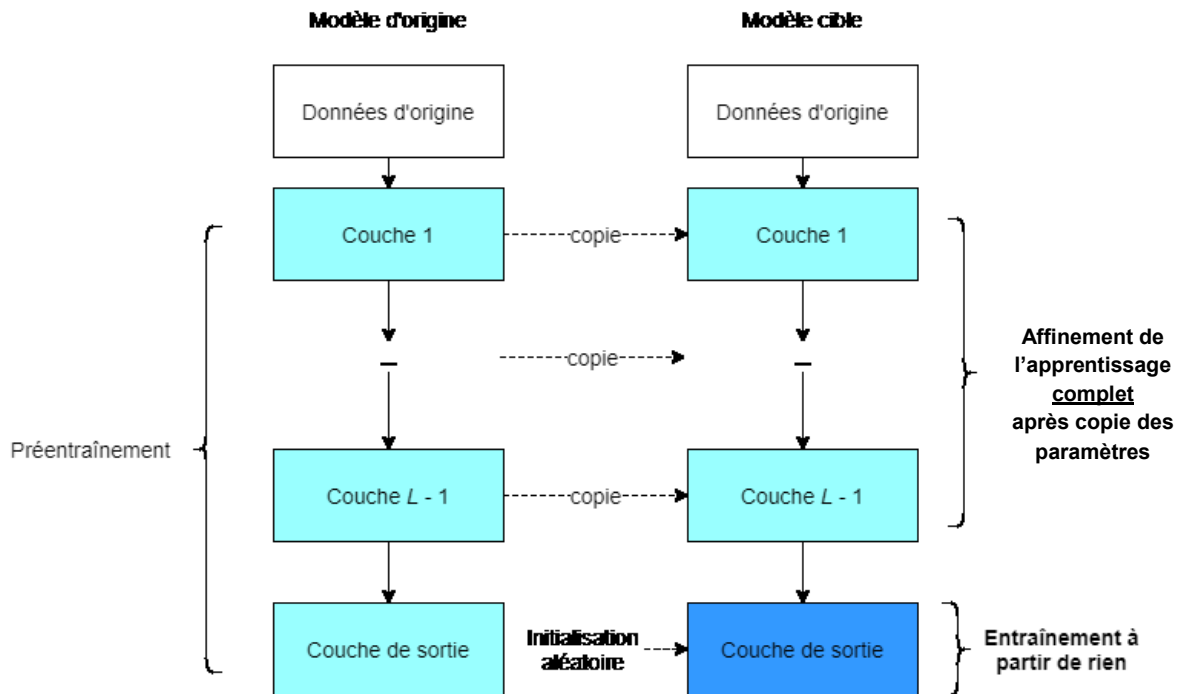


Figure 21: Transfert d'apprentissage de type « Affinement de l'apprentissage ». La couche de sortie est d'abord réentraînée, puis l'apprentissage se poursuit pour le corps du réseau, après copie des paramètres pour cette portion..

2.6.3. Affinement partiel

Une troisième approche, l'affinement partiel, constitue un hybride entre l'extraction de caractéristiques et l'affinement de l'apprentissage (Figure 22). Selon Chollet (2017a), un gel partiel des couches fondamentales peut être bénéfique lorsqu'un affinement de l'apprentissage est effectué avec un jeu de données de petite taille. Ce gel éviterait au modèle de détériorer ses représentations de bas niveau apprises sur des jeux de données de grande taille. Pour plus d'information, se référer à Chollet (2017a, p.152).

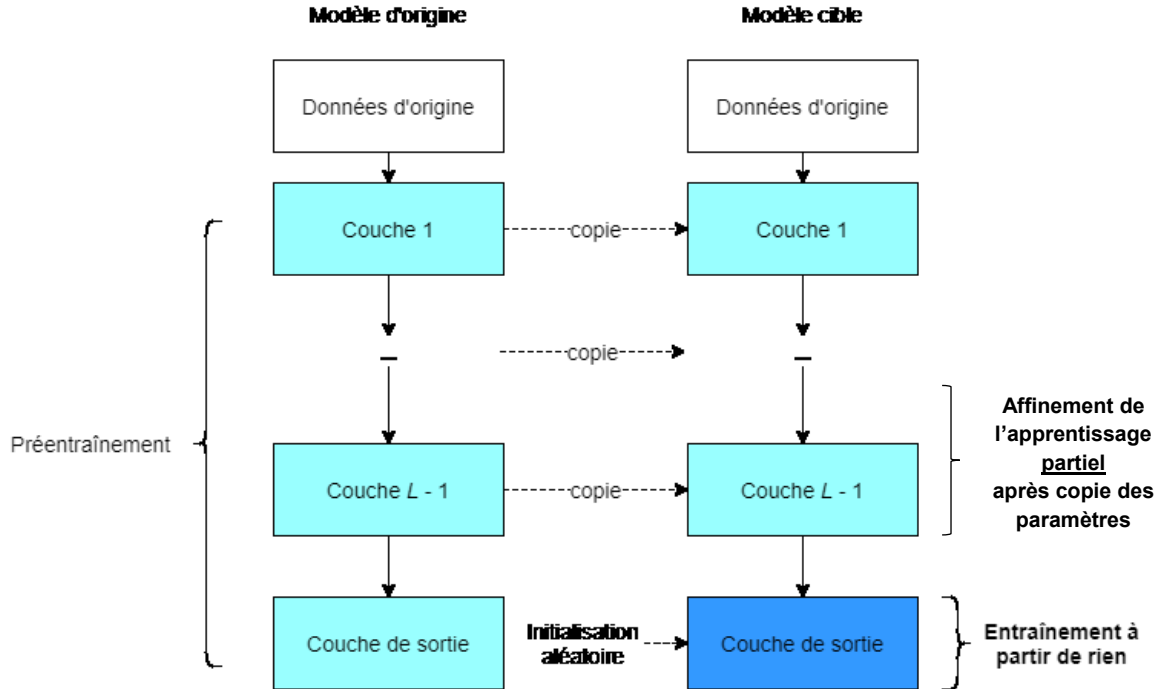


Figure 22: Transfert d'apprentissage de type « Affinement partiel ». La couche de sortie est d'abord réentraînée, puis l'apprentissage se poursuit pour une partie seulement du réseau.

2.6.4. Déroulement et résultats des tests

En raison de la variété des méthodes de transfert d'apprentissage et de l'absence d'un consensus à l'égard de la démarche la plus appropriée, il est apparu opportun de procéder à un test de comparaison des trois méthodes afin de sélectionner celle qui s'avérerait la plus performante. Pour le test d'affinement complet, l'entièreté du réseau a été réentraînée, c'est-à-dire l'encodeur et le décodeur. En ce qui a trait au test d'affinement partiel, trois emplacements qui correspondent environ aux trois premiers quarts de l'encodeur sont sélectionnés. Les couches en amont de ces sections sont ensuite gelées afin d'effectuer un réentraînement uniquement sur les dernières couches du réseau. Dans le cadre de cette seconde ronde de test, le réentraînement vise donc à nouveau l'encodeur (la portion qui n'est pas gelée) autant que le décodeur.

Selon les résultats obtenus lors des tests conduits pour ces trois différentes approches de transfert d'apprentissage, une méthode a été privilégiée pour les entraînements successifs. Étant donné que ces tests ne constituaient pas le cœur du présent projet, un seul modèle est entraîné selon les différentes approches, et ce, uniquement pour la tâche de segmentation

(*Deeplabv3*). Si les résultats étaient demeurés ambigus, plusieurs modèles auraient pu être entraînés afin de se baser sur une moyenne des performances plutôt que sur un résultat unique. De plus, ces tests auraient été répétés, au besoin, pour *MobileNetV2* (tâche de caractérisation).

2.7. Choix d'architectures

Plusieurs architectures récentes issues de la recherche d'architecture automatisée (*Neural Architecture Search*) ont été publiées et dépassent l'état de l'art sur des jeux de données de référence. Toutefois, la composition de ces architectures est souvent difficilement compréhensible pour un humain étant donné que les différents éléments qui les composent sont organisés d'une manière automatique et aléatoire (Sandler *et al.*, 2018). De surcroît, d'éventuelles modifications légères à ces architectures afin de les optimiser à la tâche désirée ou pour des tests spécifiques revêtent une difficulté supplémentaire. Pour cette raison, le projet actuel s'est penché uniquement sur des architectures issues de la recherche manuelle.

Les deux architectures de segmentation sémantique retenues afin de s'atteler aux tâches de segmentation et de caractérisation des trottoirs sont respectivement *Deeplabv3+* (Chen 2018) et *MobileNetV2* (Sandler *et al.*, 2018).

2.7.1. Deeplabv3+

Composé de 65 couches (encodeur seulement) et 23 millions de paramètres, *Deeplabv3+* (Figure 23) est le fruit d'efforts de recherche déployés par une équipe de *Google* depuis 2015. La première architecture de segmentation sémantique, *Deeplabv1*, s'est vue graduellement bonifiée afin de constituer un modèle à la fois performant et léger en nombre de paramètres, lorsque comparé à d'autres modèles semblables. *Deeplabv3+* récupère les avancées de plusieurs modèles concurrents, notamment une approche par encodeur-décodeur et le recours à la convolution à trous dans la portion finale de l'encodeur (Figure

24 et Figure 25). Comme encodeur, *Deeplabv3+* récupère et adapte une architecture de classification parmi les plus performantes : *Xception* (Chollet, 2017b).

Ces deux éléments servent principalement l'objectif de permettre à cette architecture de traiter des images de haute résolution, et ce, à différentes échelles. *Deeplabv3+* parvient donc à segmenter une image urbaine de *Cityscapes* (Cordts et al., 2016) avec précision, qu'il s'agisse d'extraire une bicyclette ou une route, au pixel près. De plus, *Deeplabv3+* profite d'une architecture pleinement convolutive telle que proposée par Long *et al.* (2015), lui permettant de traiter des images de tailles variables en entrée.

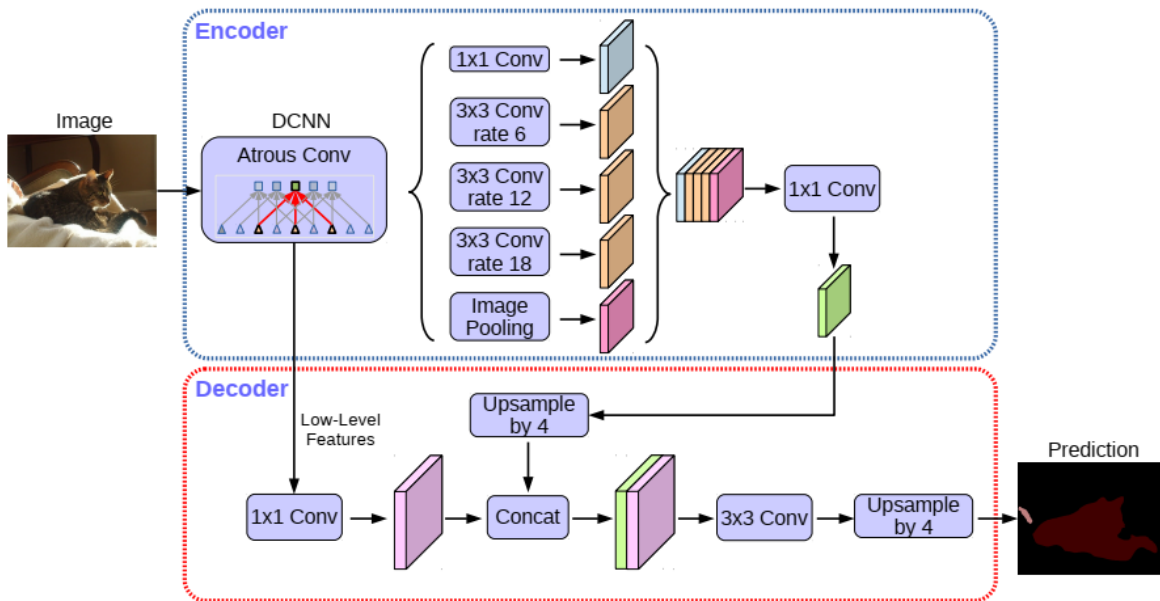


Figure 23: Structure de *Deeplabv3+*, une architecture de segmentation sémantique de type encodeur-décodeur (tiré de Chen *et al.*, 2018, p.4)

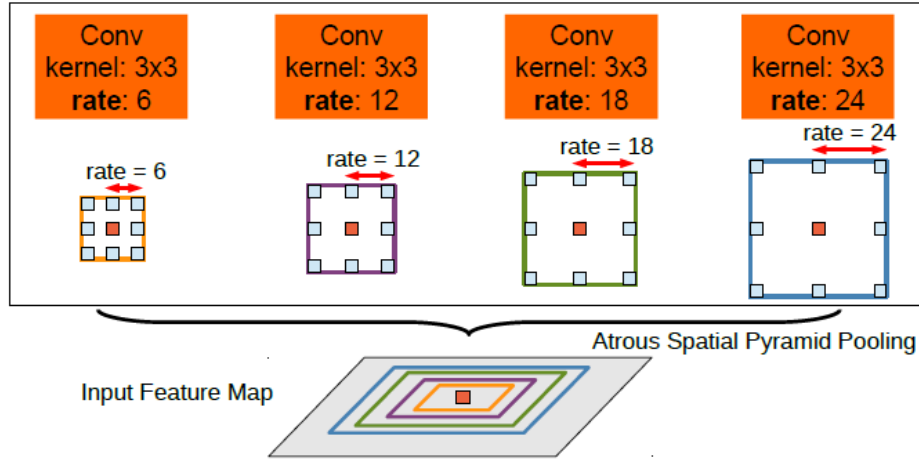


Figure 24: Structure du module « Regroupement de pyramides spatiales à trous » (Atrous Spatial Pyramid Pooling) contenu dans l'encodeur de Deeplabv3+ (tiré de Chen et al., 2017, p.1)

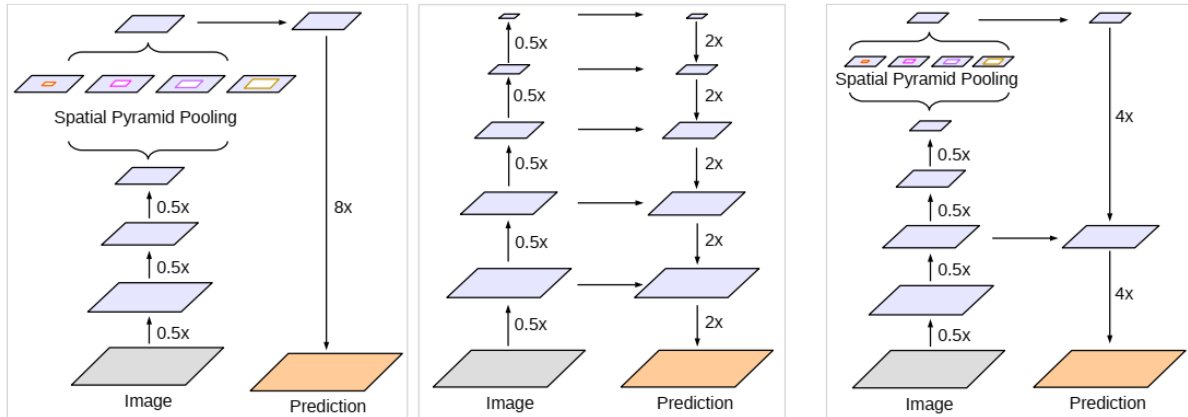


Figure 25: Trois approches de composition des architectures de segmentation sémantique. (a) regroupement de pyramides spatiales, (b) Encodeur-décodeur, (c) Encodeur-décodeur avec convolution à trous, un hybride entre (a) et (b) utilisé dans Deeplabv3+ (tiré de Chen et al., 2018, p.2)

2.7.2. MobileNetV2

Pour assurer la caractérisation des trottoirs, la recherche d'une architecture adéquate a porté sur une famille spécifique d'architectures : les architectures mobiles. Celles-ci sont optimisées afin de minimiser le nombre de paramètres qui les constituent tout en conservant une performance acceptable. Elles sont conçues pour être intégrées, par exemple, dans diverses applications de reconnaissance visuelle directement sur des appareils mobiles.

Cette famille d'architectures a été ciblée pour la tâche de caractérisation des trottoirs étant donné qu'aucun jeu de données public de grande taille n'existe pour préentraîner un réseau

à la caractérisation des trottoirs. Comme les données d’entraînement sont exclusivement produites dans le cadre de ce projet, la taille finale du jeu de données est donc limitée et risque de provoquer le surentraînement des réseaux. Le recours à une architecture mobile, de taille minimale en nombre de paramètres, permet donc de limiter ce risque (Goodfellow *et al.* 2016).

L’architecture *MobileNetV2* (Sandler *et al.*, 2018) est composée de 11 couches et 2.1 millions de paramètres. Elle intègre plusieurs éléments innovants comme un type de convolution décomposée qui permet de réduire de 8 à 9 fois le nombre de paramètres d’un filtre de convolution standard. *MobileNetV2* introduit également diverses avancées architecturales inspirées de structures existantes, comme les connexions résiduelles et le goulot d’étranglement linéaire. Tous ces éléments, amplement expliqués et détaillés par Sandler *et al.* (2018), font de *MobileNetV2* une architecture particulièrement légère, mais dont la performance est comparable à des réseaux classiques (Figure 26). Pour cette raison, cette architecture a été retenue afin de s’atteler à la tâche de caractérisation de l’état de revêtement des trottoirs.

Network	OS	ASPP	MF	mIOU	Params	MAdds
MNet V1	16	✓		75.29	11.15M	14.25B
	8	✓	✓	78.56	11.15M	941.9B
MNet V2*	16	✓		75.70	4.52M	5.8B
	8	✓	✓	78.42	4.52M	387B
MNet V2*	16			75.32	2.11M	2.75B
	8		✓	77.33	2.11M	152.6B
ResNet-101	16	✓		80.49	58.16M	81.0B
	8	✓	✓	82.70	58.16M	4870.6B

Figure 26: Performances, taille (nombre de paramètres) et nombre d’opérations de 3 réseaux mobiles de la famille *MobileNet* et d’un réseau de classification classique, *ResNet-101*. Chaque réseau a été testé selon deux configurations différentes. (tiré de Sandler *et al.*, 2018, p.4517)

2.8. Entraînement de modèles par segmentation sémantique

2.8.1. Tâche 1 : segmentation des trottoirs

Séparation des ensembles

La banque d'images annotées, acquises à Montréal et Saint-Hyacinthe, comprenait un total de 394 panoramas de 6656 x 3328 pixels. Parmi ces données d'entraînement, 80 % (309 panoramas) ont constitué l'ensemble d'entraînement tandis que 20 % (85 panoramas) ont été réservées à des fins de validation après chaque époque d'entraînement. Cette proportion a été fixée selon les pratiques courantes en apprentissage machine (Chollet, 2017a).

Les panoramas ont été sélectionnés sur la base d'une évaluation visuelle visant la surreprésentation d'images comprenant des trottoirs avec un revêtement endommagé. Cette approche a été retenue afin de réduire en amont le déséquilibre des classes, un obstacle courant pour l'apprentissage des modèles (Johnson *et* Khoshgoftaar, 2019; Masko *et* Hensman, 2015).

Entraînement

Deeplabv3+ a été entraîné afin d'effectuer une segmentation sémantique à deux classes : (1) non-trottoir et (2) trottoir. Ce modèle a été utilisé comme étape de prétraitement pour la tâche de caractérisation finale (voir section 2.8.2 « Tâche 2 : Caractérisation de l'état du revêtement »). Ainsi, ce modèle devait permettre de limiter l'analyse du second modèle à une région d'intérêt, c'est-à-dire uniquement les pixels composant les trottoirs.

Pour les besoins du projet, la couche de classification originale (*Cityscapes*, 19 classes) a été tronquée, puis remplacée par une couche à 2 classes (« trottoirs » et « non-trottoirs »). Puis, un affinement de l'apprentissage a été effectué à partir de poids préentraînés sur le jeu de données *Cityscapes* (Cordts *et al.*, 2016) afin de favoriser une convergence rapide et une meilleure robustesse des caractéristiques apprises (Yosinski *et al.*, 2014). À titre d'information, l'approche d'affinement d'apprentissage sélectionnée à cette étape dépend des résultats des tests préalables décrits à la section 2.6 « Comparaison des méthodes de transfert d'apprentissage ».

L'entraînement de ce modèle a été effectué sur les images de Montréal et de Saint-Hyacinthe, avec des mini-lots (*mini-batch*) de 8 images de 544 x 544 pixels chacun. Ces images ont été obtenues en rognant les images de taille originale (Figure 27) afin d'éliminer des zones où les trottoirs étaient systématiquement absents ou très éloignés. L'augmentation de la taille des mini-lots a été priorisée à l'augmentation des dimensions des images afin d'assurer une stabilité dans le calcul de la normalisation par mini-lots, tel que démontré par Chen *et al.* (2018).



Figure 27: Exemple d'un panorama Google Street View. Les cadres blancs indiquent les imagerie extraites dans chaque panorama après rognage pour les fins de l'entraînement. N.B. La taille réelle de l'image originale est 6656 x 3328 pixels.

La fonction de perte *Lovasz-Softmax* (Berman *et al.*, 2018) a été utilisée, car elle est particulièrement adaptée pour la segmentation sémantique. Comme le mentionne Berman *et al.* (2018), cette fonction permet, entre autres, de gérer intrinsèquement le déséquilibre des classes, contrairement à la traditionnelle fonction *Cross-Entropy*, qui nécessite un balancement via une pondération manuelle. En ce qui a trait à l'optimisation du modèle, l'optimiseur *Adabound* (Luo *et al.*, 2019) a été préconisé étant donné, d'une part, sa rapidité à converger par rapport à la fonction de descente du gradient stochastique (*SGD*, en anglais) et, d'autre part, les performances démontrées par Luo *et al.* (2019) par rapport à la fonction *Adam*.

Le meilleur modèle a été sélectionné sur la base des performances obtenues sur l'ensemble de validation, toujours issu des images de Montréal et Saint-Hyacinthe. L'évaluation de cette performance a été effectuée grâce à l'indice de Jaccard (Jaccard, 1908), une métrique couramment utilisée en segmentation sémantique également appelée *Intersection over Union* (IoU) (Long *et al.*, 2015). Pour un ensemble d'images donné, cet indice évalue la similarité entre la prédiction et la vérité terrain, et ce, pour chaque classe. Il est défini par l'équation suivante :

$$\text{indice de Jaccard} = \frac{VP}{FP+VP+FN} \quad (1)$$

où VP, FP et FN signifient respectivement vrai positif, faux positif (erreur de type I) et faux négatif (erreur de type II).

Validation

L'étape de validation, intégrée à l'entraînement des modèles, permet d'évaluer un modèle en cours d'apprentissage sur des données indépendantes, grâce à une fonction de perte et une métrique (l'indice de Jaccard, le cas échéant). Ainsi, après chaque époque d'entraînement (itérations d'entraînement couvrant l'intégralité de l'ensemble d'entraînement), l'algorithme d'apprentissage évalue le modèle, dont les paramètres viennent d'être mis à jour, sur l'ensemble de validation afin de retenir, ou non, ce modèle comme meilleur modèle. De plus, par économie de ressources, les entraînements étaient interrompus après 600 époques sans amélioration du modèle, et ce, basé sur la valeur de perte obtenue sur l'ensemble de validation (Gopalakrishnan, 2017).

2.8.2. Tâche 2 : caractérisation de l'état du revêtement

Un deuxième modèle a été produit afin de traiter les pixels rendus par le masque de segmentation. La tâche programmée pour ce modèle était la segmentation des pixels de trottoirs en quatre classes, présentées au Tableau 1 et à la Figure 28. Vu la quantité limitée de panoramas utilisés durant l'entraînement et la surreprésentation des pixels classés « Bon état », la tâche a été réduite à deux classes : (1) bon état ou défaut mineur et (2) défaut majeur.

Séparation des ensembles

Étant donné que le modèle de caractérisation reçoit, en entrée, un masque de segmentation des trottoirs (i.e. tous les pixels « non-trottoirs » sont masqués), produit en amont, celui-ci a été entraîné uniquement sur des tronçons de panoramas incluant des pixels de trottoirs (Figure 29). De plus, des tests préliminaires ont démontré l'importance de conserver uniquement les panoramas comprenant un défaut (classe 2, 3 ou 4). Cela a permis de pallier une forte sous-représentation de portions de trottoirs endommagés. Au total, l'ensemble d'entraînement a été constitué de 337 images et l'ensemble de validation, 85.

Tableau 1: Définition des classes pour la caractérisation de l'état du revêtement des trottoirs

Variable	Impact de défaut sur une personne en fauteuil roulant	Classe	Id	Id, après regroupement
État du revêtement	Aucun	Bon état	1	1b
	Mineur	Lézardes superficielles	2	
	Majeur	Crevasses profondes/larges	3	2b
		Très endommagé/impraticable	4	



Figure 28: Échantillons représentatifs de chaque classe pour l'état du revêtement (gauche à droite) : bon état, lézardes superficielles, crevasses profondes/larges, très endommagé/impraticable



Figure 29: Exemple d'un échantillon utilisé durant l'entraînement de modèles pour la tâche 2 (caractérisation de l'état du revêtement)

Entraînement

Le réseau de neurones retenu pour cette tâche ultime est *MobileNetV2* (Sandler *et al.*, 2018). Comme pour la tâche de segmentation des trottoirs, un affinement de l'apprentissage a été effectué avec des poids préentraînés sur *Cityscapes* (Cordts *et al.*, 2016). Comme pour *Deeplabv3+*, la couche de classification de *MobileNet2* a été remplacée afin de convenir aux besoins de la tâche de caractérisation (« défaut majeur », « bon état ou défaut mineur »). L'étape de validation s'est déroulée de manière identique à celle de la tâche de segmentation des trottoirs.

2.9. Post-traitement

Après entraînement, les deux modèles ayant obtenu les meilleures performances sur les ensembles de validation pour la segmentation des trottoirs et la caractérisation de leur revêtement ont été intégrés dans une chaîne de traitement. Celle-ci a permis l'évaluation sur le site d'évaluation finale (Figure 30).

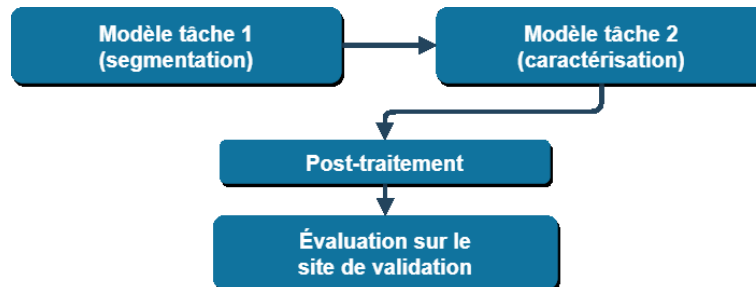


Figure 30: Chaîne de traitement finale intégrant les modèles les plus performants pour chacune des deux tâches (segmentation et caractérisation)

D'abord, le post-traitement a seulement été appliqué pour les données du site d'évaluation finale. Ainsi, les étapes préliminaires (production des modèles à partir des données d'entraînement) ont uniquement servi à la production de deux modèles de segmentation sémantique. Les données des sites d'entraînement (Montréal et Saint-Hyacinthe) n'ont donc pas été visées par le post-traitement.

Au moment du post-traitement, un seuillage était effectué sur les valeurs de la carte des représentations produite par l'avant-dernière couche des réseaux de segmentation

sémantique. Ce filtrage servait à éliminer de faibles valeurs d'activation associées à certains pixels dans le masque de segmentation final.

Par la suite, les prédictions sous forme de masque étaient vectorisées afin de traiter les défauts comme des objets, des polygones le cas échéant, plutôt que comme un ensemble de pixels. Cette conversion comporte plusieurs avantages, comme le fait de pouvoir filtrer certains défauts en fonction de leurs dimensions. À ce titre, les défauts de revêtement avec une superficie prédite de moins de 400 pixels ont été supprimés. Enfin, la vectorisation des prédictions a permis de classer chaque défaut comme un vrai positif, un faux positif, ou un faux négatif. Le résultat de ce décompte était enfin exprimé avec des métriques (voir section suivante) plus faciles à interpréter que l'indice de Jaccard, basé sur le résultat par pixel plutôt que par défaut détecté.

2.10. Évaluation finale

Les ressources disponibles dans le cadre du projet n'ont pas permis d'annoter les 2609 panoramas du site d'évaluation finale sur terrain. Seule une évaluation visuelle des défauts sur cet ensemble d'images (centre-ville de Sherbrooke, voir Section 4) a permis d'évaluer la précision et le rappel du modèle, avec une incertitude de l'ordre de 10 %. Ces deux métriques sont mesurées de la façon suivante :

$$\text{Précision} = \frac{VP}{FP+VP} \quad (2) \qquad \text{Rappel} = \frac{VP}{FN+VP} \quad (3)$$

où VP, FP et FN signifient respectivement vrai positif, faux positif et faux négatif.

2.11. Présentation et analyse des résultats

Au terme des tests, entraînements et évaluations prévus, les résultats sont présentés et mis en perspective selon l'approche méthodologique choisie. D'abord, le bilan des tests de prétraitement et des méthodes de transfert d'apprentissage sont exposés. Puis, les indices de Jaccard obtenus pour chacune des deux tâches ciblées sont partagés et accompagnés

d'exemples visuels de prédictions obtenues à partir des meilleurs modèles retenus. Enfin, les résultats issus de l'évaluation de ces modèles sur le site d'évaluation finale sont présentés.

2.12. Interprétation des résultats

L'interprétation des résultats permet de revenir sur l'hypothèse et d'en discuter à la lumière de la littérature scientifique. Comme cette section est rédigée plusieurs mois plus tard, certaines avancées récentes méritent une attention particulière. Si elles s'avèrent pertinentes, leur intégration dans de futurs efforts de recherche pourrait être proposée.

Chapitre 3. Résultats : présentation et analyse

Les entraînements des modèles ont eu lieu entre le 1^{er} avril et le 8 août 2019. Ceux-ci ont été réalisés sur des cartes graphiques Nvidia Tesla V100 des superordinateurs du centre de calcul Béluga géré par Calcul Canada. L'entraînement de modèles pour la segmentation des trottoirs a duré entre 24h et 48h. Les entraînements pour la tâche de caractérisation ont duré environ 1 heure. Les scripts de programmation ont été développés dans le langage Python (v3.7), avec la bibliothèque d'apprentissage profond Pytorch. Le code développé de ce projet de maîtrise est disponible sur la plateforme *GitHub* (github.com/remtav).

3.1. Tests de prétraitement

Ce test préliminaire a démontré qu'un réseau *Deeplabv3+* préentraîné sur *Cityscapes* fournissait des prédictions équivalentes, voire de qualité supérieure, lorsque le réseau recevait des images distordues en entrée (Figure 31). Ainsi, l'hypothèse selon laquelle les poids d'un réseau préentraîné sur des images sans distorsion seraient inadaptés pour extraire des représentations pertinentes sur des images similaires, mais présentant une distorsion géométrique n'a pas été confirmée. En fait, la comparaison effectuée sur 20 images durant le test de prétraitement a clairement démontré que le modèle préentraîné

fournissait des prédictions d'aussi bonne, voire de meilleure qualité lorsque les images présentaient cette distorsion. À la lumière de ces résultats, le recours à une étape de prétraitement afin de corriger ladite distorsion a été écarté.

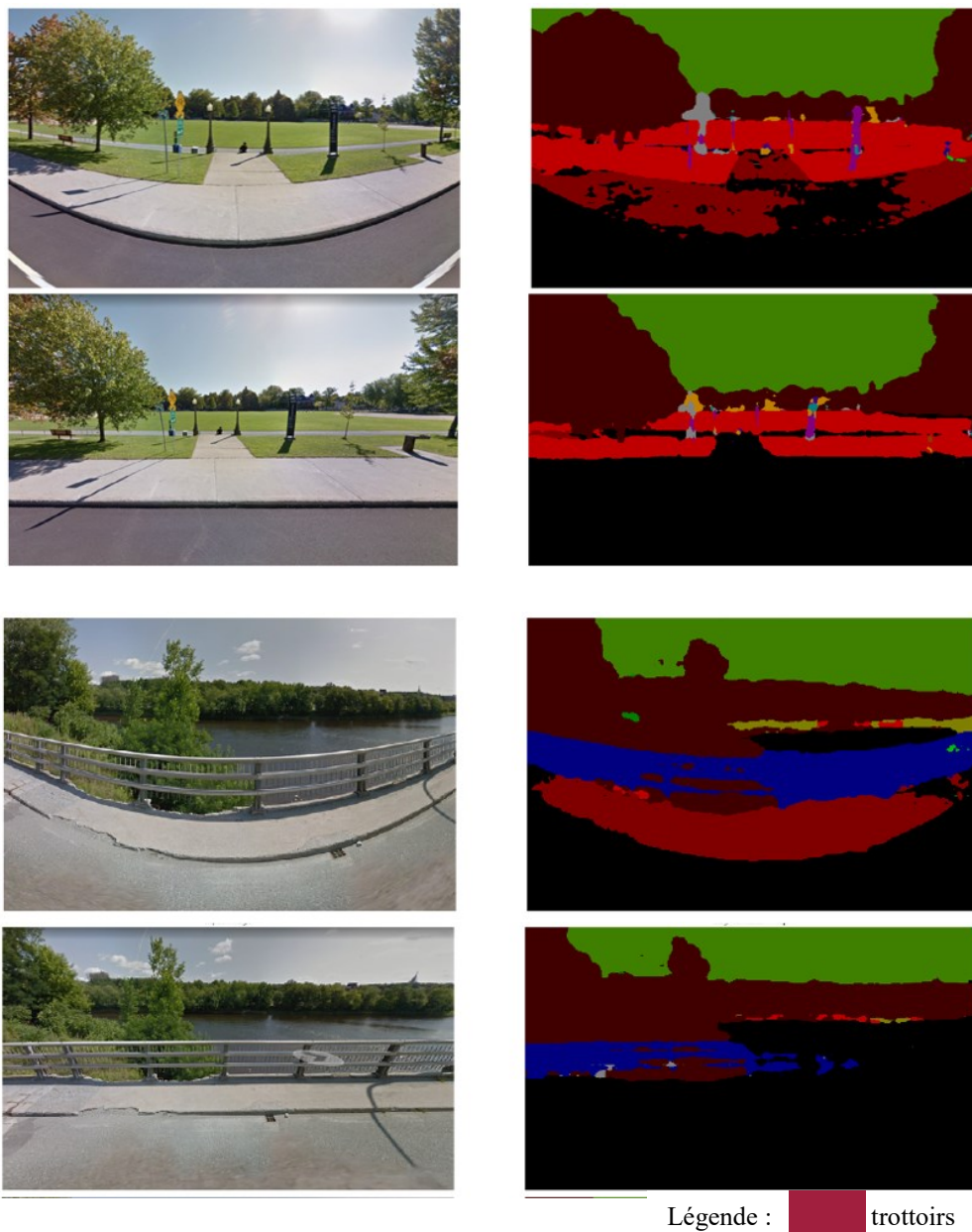


Figure 31: Deux images avec copie distordue (originale) et corrigée avec les prédictions obtenues par Deeplabv3+ préentraîné sur Cityscapes.

3.2. Comparaison des méthodes de transfert d'apprentissage

Les indices de Jaccard obtenus lors des tests des méthodes de transfert d'apprentissage sur la tâche de segmentation indiquent que l'affinement complet donne de meilleurs résultats que les méthodes de transfert concurrentes (Tableau 2 et Figure 32).

Tableau 2: Comparaison des hyperparamètres et performances pour les tests des méthodes de transfert d'apprentissage pour la tâche de segmentation des trottoirs

Technique de transfert d'apprentissage	Dimension des mini-lots	Dernière couche gelée	Performance (indice de Jaccard)
Extraction de caractéristiques	4	Toutes couches sauf la dernière couche	0.50
Affinement complet	8	Aucun gel	0.93
Affinement partiel	8	Bloc 4	0.82
	8	Bloc 12	0.65
	8	Bloc 18	0.69

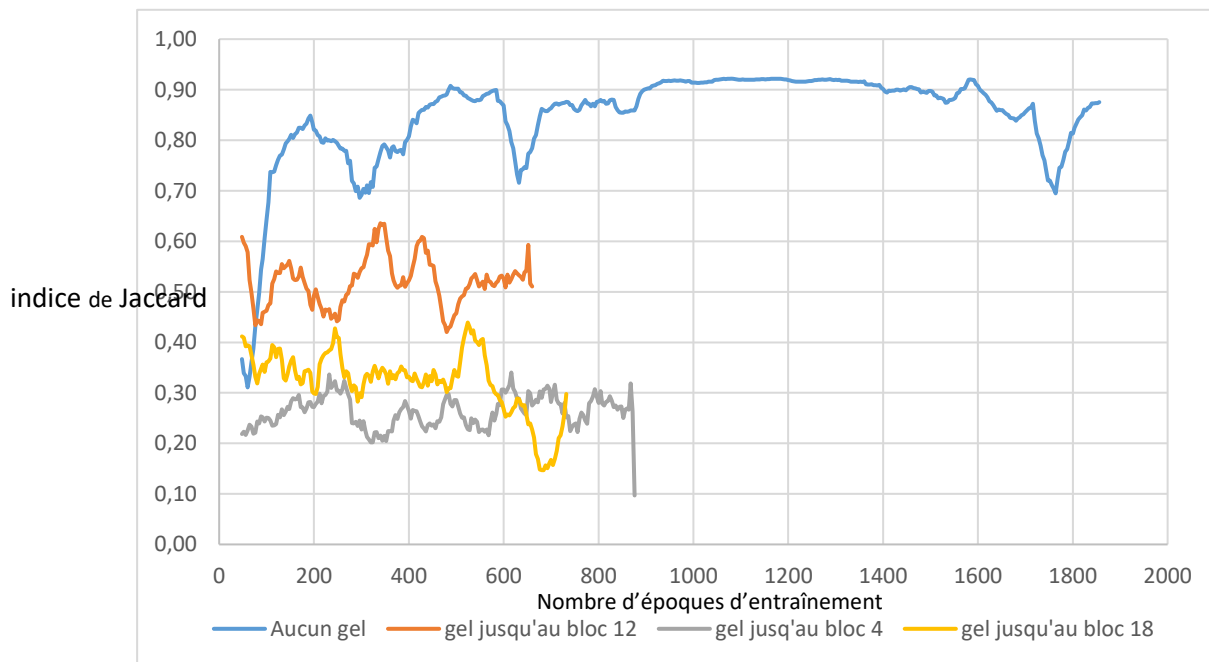


Figure 32: Évolution de l'indice de Jaccard pour les tests d'affinement complet et partiel. Ce dernier test comportait trois sous-tests dans lesquels la dernière couche gelée était définie à trois emplacements différents (pour plus de détails sur ces couches, voir Chollet, 2017a). Les courbes sont aplanies pour faciliter la visualisation. Selon les résultats en cours d'entraînement, ceux-ci étaient prolongés plus ou moins longtemps, d'où le nombre d'époques d'entraînement variable. Les résultats pour la méthode d'extraction des caractéristiques ne sont pas présentés.

Étant donné l'écart significatif entre les résultats du modèle avec affinement complet par rapport aux autres modèles, la comparaison des méthodes d'apprentissage ne s'est pas étendue au-delà des 5 modèles dont les performances sont présentées dans cette section. Ainsi, l'approche de l'affinement complet a été privilégiée dans l'ensemble des entraînements subséquents du projet, autant pour la segmentation que pour la caractérisation.

3.3. Tâche 1 : segmentation des trottoirs

De nombreux tests ont permis de choisir la configuration optimale d'hyperparamètres pour l'affinement de l'apprentissage du modèle de segmentation. Les hyperparamètres ayant démontrés l'impact le plus important sur les performances du modèle sont, en ordre décroissant d'importance, le nombre d'époques d'entraînement, la taille du mini-lot (*mini-batch*) et la taille de l'image à l'entrée du réseau.

Le meilleur modèle après réentraînement a enregistré un indice de Jaccard de 93 % (voir exemple de masque de trottoirs prédit aux Figure 33 et Figure 34) sur l'ensemble de validation (voir section 2.2.2 « Site d'évaluation finale »). Celui-ci a été obtenu à la 1255^e époque d'entraînement, après 28h d'exécution. La Figure 35 présente l'évolution de l'indice de Jaccard durant l'entraînement de ce modèle. Le Tableau 3 présente les hyperparamètres fixés pour l'entraînement du modèle retenu en segmentation des trottoirs.

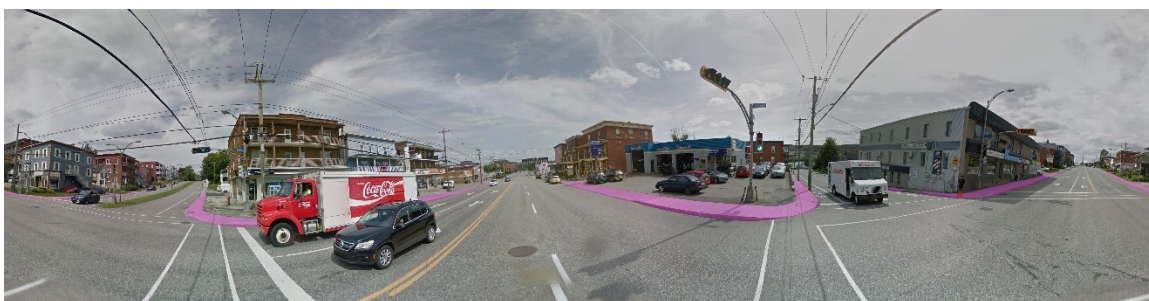


Figure 33: Superposition d'un panorama Google Street View et d'un masque de segmentation (en rose) pour les pixels de la classe «Trottoir». Ce masque a été produit par le modèle Deeplabv3+ entraîné sur les données d'entraînement du présent projet.



Figure 34 : Exemples de masques imparfaits (en rose) produit par le modèle de segmentation des trottoirs.

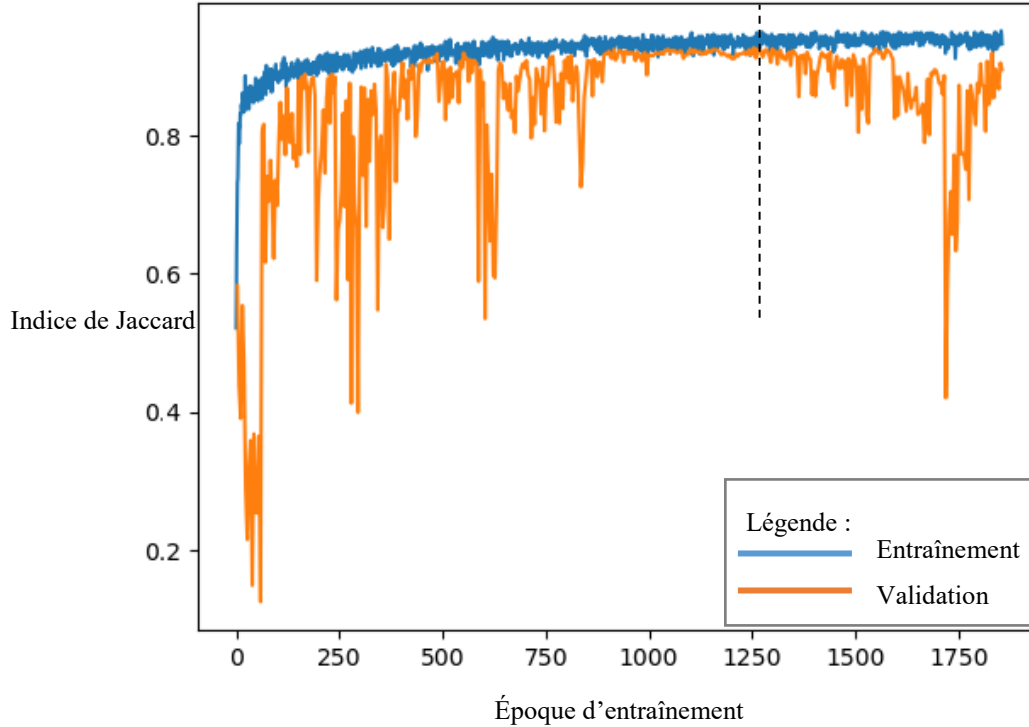


Figure 35: Évolution de l'indice de Jaccard en fonction de l'époque d'entraînement, pour les ensembles d'entraînement et de validation. Cet entraînement a permis de produire le modèle utilisé pour la segmentation des trottoirs. L'apparence bruitée de la courbe de validation est causée par le niveau de difficulté élevé de la tâche d'apprentissage jumelée aux propriétés de la métrique utilisée, qui pénalise fortement les mauvais recouvrements entre prédictions et cibles.

Tableau 3: Hyperparamètres ayant produit le modèle Deeplabv3+ retenu pour la segmentation des trottoirs

Type de paramètre	Nom	Valeur
Modèle	Encodeur	<i>Xception</i> – 65 couches avec regroupement de pyramides spatiales à trous (<i>ASPP</i>)
	Décodeur	N.A.
	<i>Output stride</i>	8
Données	Taille de l'image	544 x 544 pixels
	Augmentations affines	Rognage, rotation, translation, redimensionnement
	Augmentations radiométriques	Luminosité, contraste
Entraînement	Nombre d'époques maximal	2400
	Taille de mini-lot	8
	Préentraîné	<i>Cityscapes</i>
	Fonction de perte	Multi-class Lovasz-Softmax (Berman <i>et al.</i> , 2018)
	Taux d'apprentissage	0.0007
	Type de <i>scheduler</i>	Cosine Cyclic Learning Rate (SGDR)
	Cycle	30 époques
Fonction d'optimisation	<i>Adabound</i> (Luo <i>et al.</i> , 2019)	

3.4. Tâche 2 : caractérisation de l'état du revêtement

Dans un deuxième temps, le modèle *MobileNetV2* a été entraîné pour la tâche de caractérisation de l'état du revêtement. Pour cette tâche, le déséquilibre des classes a posé un défi important. La proportion des pixels annotés selon chaque classe (Tableau 4) montre une sous-représentation élevée des pixels de la vérité terrain associés aux classes « défaut mineur » et « défaut majeur » (classes 2, 3 et 4) pour la tâche de caractérisation. Lors des tests d'entraînement initiaux, ce déséquilibre a eu un impact délétère sur l'apprentissage des modèles. En effet, en entraînant ceux-ci à partir du jeu de données d'entraînement entier, un surapprentissage important était observé (Figure 36).

Tableau 4: Répartition relative des pixels annotés pour entraînement en fonction de la classe d'appartenance

Classe – état du revêtement	Classes à l'entraînement	Pourcentage (%)
Bon état	Bon état	2
Lézardes superficielles	Défaut mineur	< 1
Crevasses profondes/larges	Défaut majeur	< 1
Très endommagé/impraticable		< 1
Autre	Autre	97

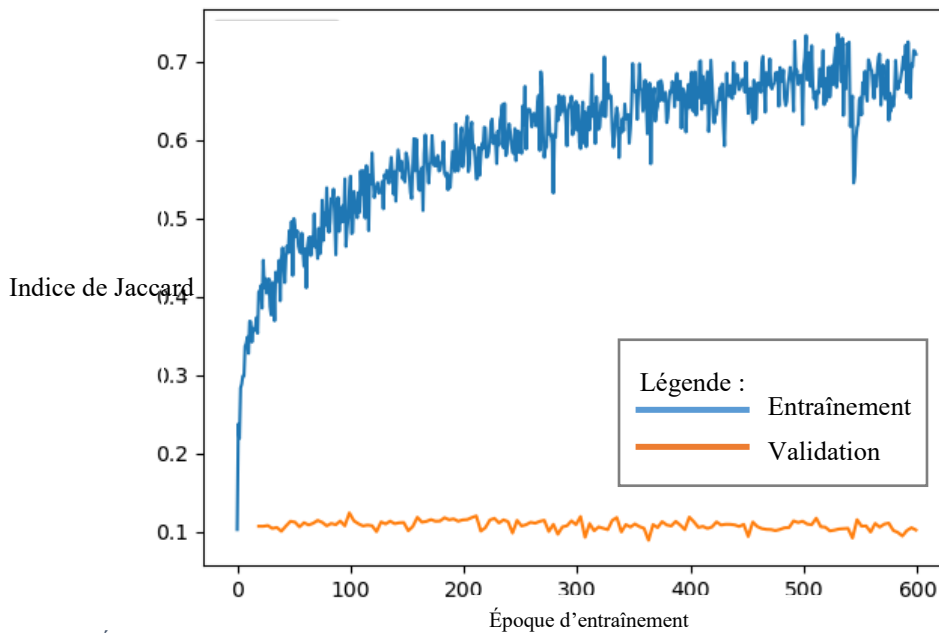


Figure 36: Évolution de l'indice de Jaccard en fonction de l'époque d'entraînement d'un modèle de caractérisation de l'état du revêtement avant l'exclusion de panoramas ne présentant aucun défaut

Cette observation a mené à l'exclusion de certains panoramas ne présentant aucun défaut afin de mieux équilibrer la représentation des classes dans les données d'entraînement. Malgré ces mesures de sous-échantillonnage, le modèle retenu pour la caractérisation n'a obtenu qu'un indice de Jaccard de 31 % sur l'ensemble de validation, avant post-traitement.

La Figure 37 présente l'évolution de l'entraînement pour le modèle retenu. L'éloignement progressif des courbes d'entraînement et de validation démontre que, malgré l'utilisation d'un réseau léger comme *MobileNetV2* (2.1M de paramètres), l'entraînement a mené à du surapprentissage. Le Tableau 5 présente les hyperparamètres fixés pour l'entraînement du modèle retenu.

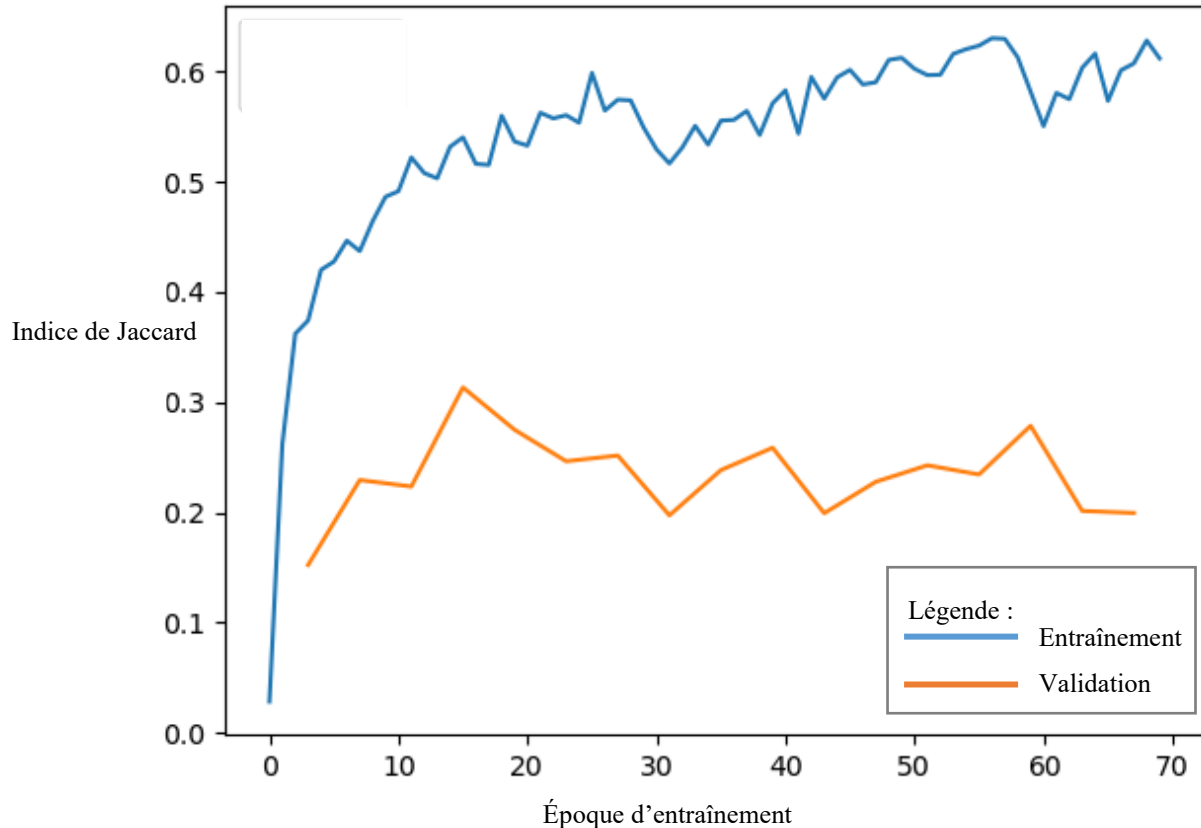


Figure 37: Évolution de l'indice de Jaccard en fonction de l'époque d'entraînement, pour les ensembles d'entraînement et de validation. Cet entraînement a permis de produire le modèle utilisé pour la segmentation des défauts de revêtement sur les trottoirs.

Tableau 5: Hyperparamètres ayant produit le modèle *MobileNetv2* retenu pour la caractérisation des trottoirs

Type de paramètre	Nom	Valeur
Modèle	Encodeur	<i>MobileNetv2</i> avec regroupement de pyramides spatiales (<i>SPP</i> , en anglais)
	Décodeur	N.A.
Données	Taille de l'échantillon	576 x 1024 pixels
	Augmentations affines	Rognage, rotation, translation, redimensionnement
	Augmentations radiométriques	Luminosité, contraste
Entraînement	Nombre d'époques maximal	200
	Taille de mini-lot	8
	Préentraîné	<i>Cityscapes</i>
	Fonction de perte	Multi-class Lovasz-Softmax (Berman <i>et al.</i> , 2018)
	Taux d'apprentissage	0.0007
	Type de <i>scheduler</i>	Cosine Cyclic Learning Rate (SGDR)
	Cycle	30 époques
	Fonction d'optimisation	<i>Adabound</i> (Luo <i>et al.</i> , 2019)

3.5. Évaluation finale

Après post-traitement, la précision sur les images du site d'étude a atteint 8 % et le rappel, 90 % (voir équations à la section 2.10 « Évaluation finale »). Ce faible score de précision, une métrique sensible à des erreurs de faux positifs, indique que le modèle de caractérisation a relevé plusieurs défauts qui n'avaient pas lieu d'être détectés. Quant au rappel, une métrique qui fluctue en fonction des cas de faux négatifs, le score élevé obtenu traduit une forte capacité du modèle de caractérisation de trouver des défauts lorsqu'ils sont bel et bien présents sur l'image (Figure 38).

Les causes associées aux faux positifs étaient multiples (Tableau 6). La majorité de ces erreurs de commission durant la caractérisation des trottoirs étaient attribuables à la détection d'un défaut mineur plutôt que majeur (Figure 39) ou à des faux positifs générés par le modèle de segmentation des trottoirs (Figure 40). Plusieurs faux positifs étaient également causés par une sous-représentation d'images captées au printemps et à

l'automne (Figure 41). Enfin, divers contextes avec une faible occurrence dans les données d'entraînement ont causé une part considérable de faux positifs (Figure 42).

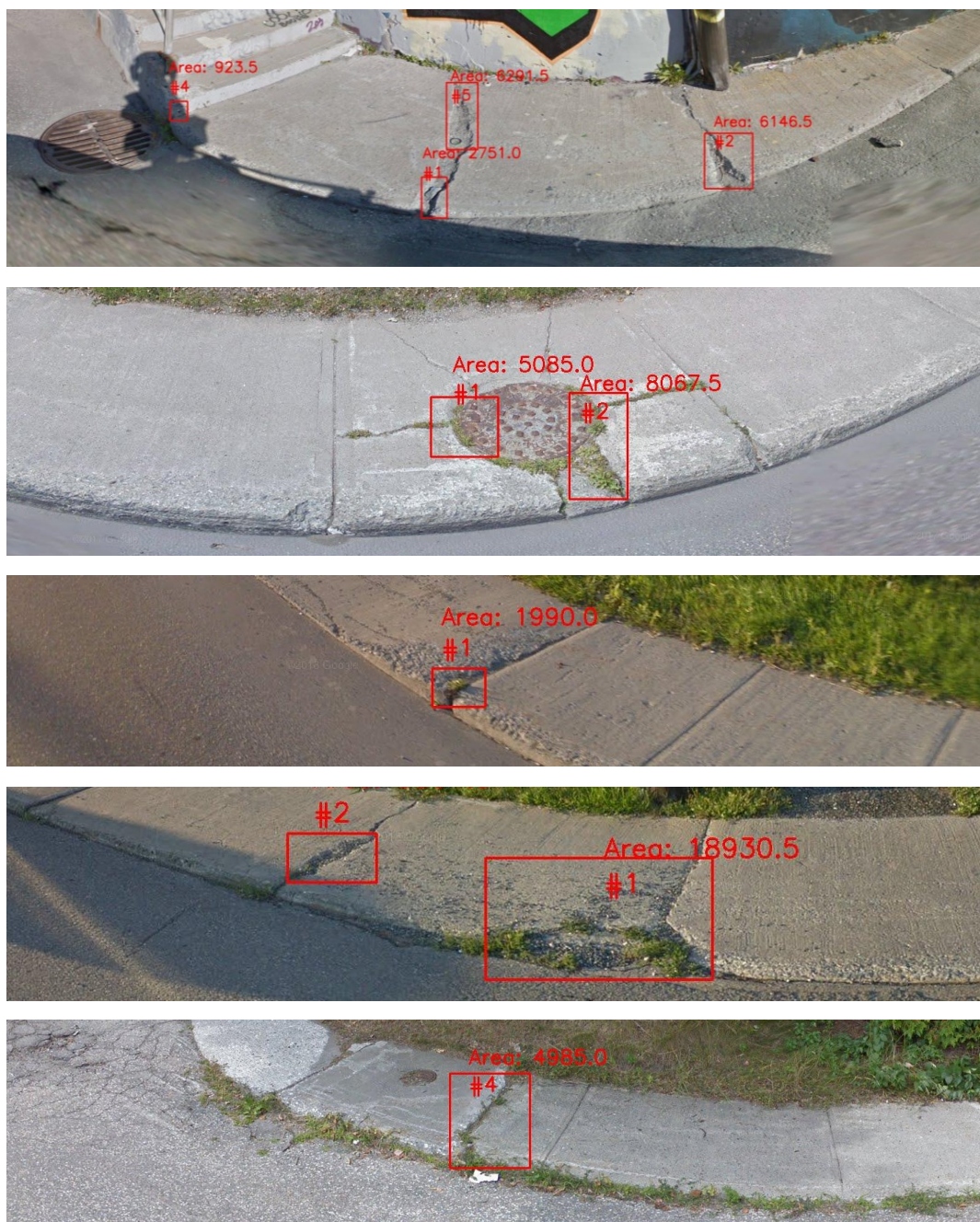


Figure 38: Exemples de vrais positifs (défauts réels) détectés par la chaîne de traitement finale (après post-traitement). Chaque défaut est entouré d'une boîte, d'un chiffre unique identifiant ce défaut sur l'image de référence, ainsi que de l'aire, en pixel, du défaut.

Tableau 6: Répartition relative des causes d'erreur ayant causé des faux positifs sur le site d'évaluation finale

Défaut mineur	Faux positifs générés par le modèle de segmentation du trottoir	Sous-représentation d'images captées au printemps et à l'automne	Autre
42 %	36 %	14 %	8 %

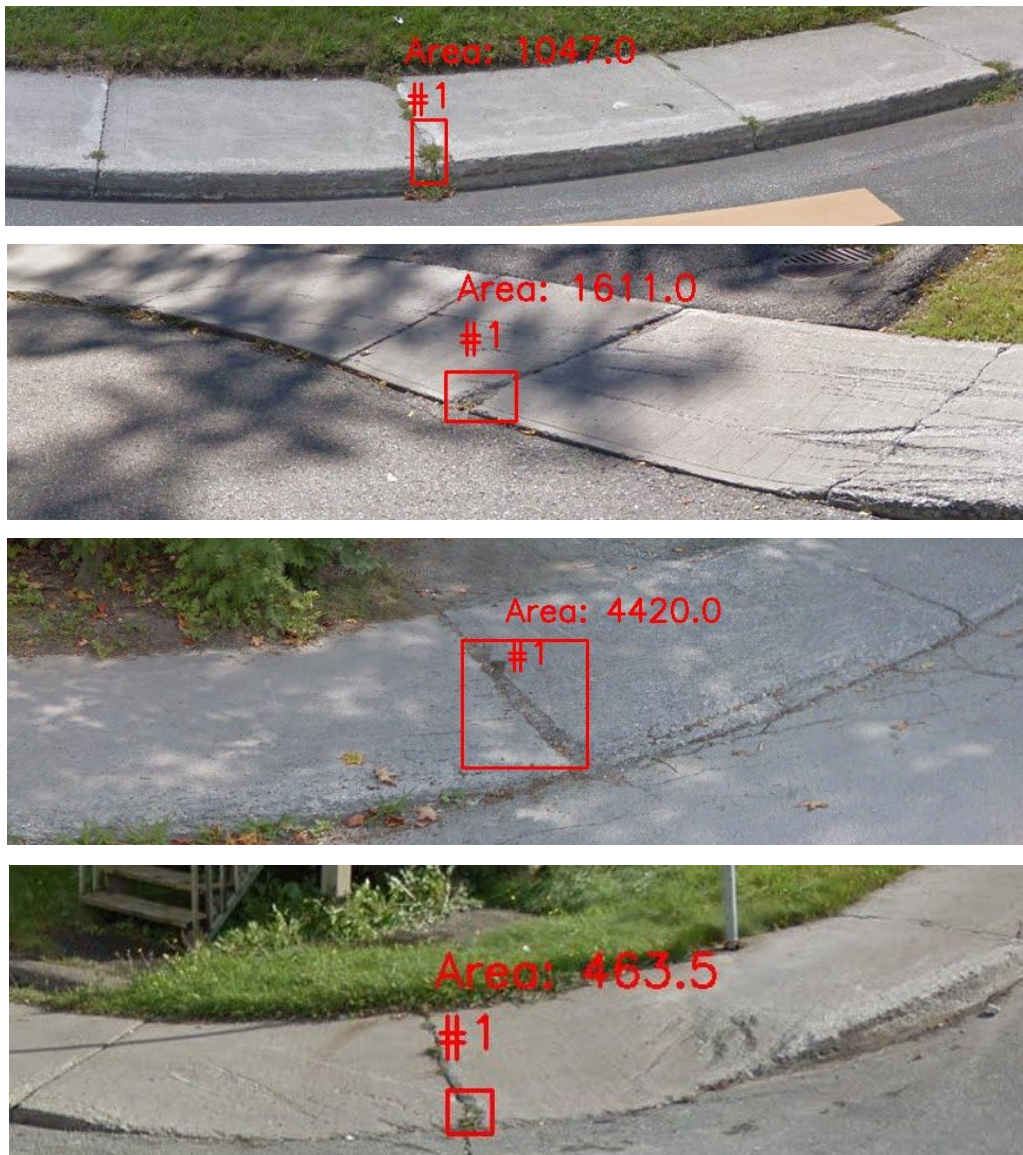


Figure 39: Exemples de faux positifs liés à la détection d'un défaut mineur plutôt que majeur. Chaque défaut est entouré d'une boîte, d'un chiffre unique identifiant ce défaut sur l'image de référence, ainsi que de l'aire, en pixel, du défaut. Ces exemples sont issus de la chaîne de traitement finale (après post-traitement).

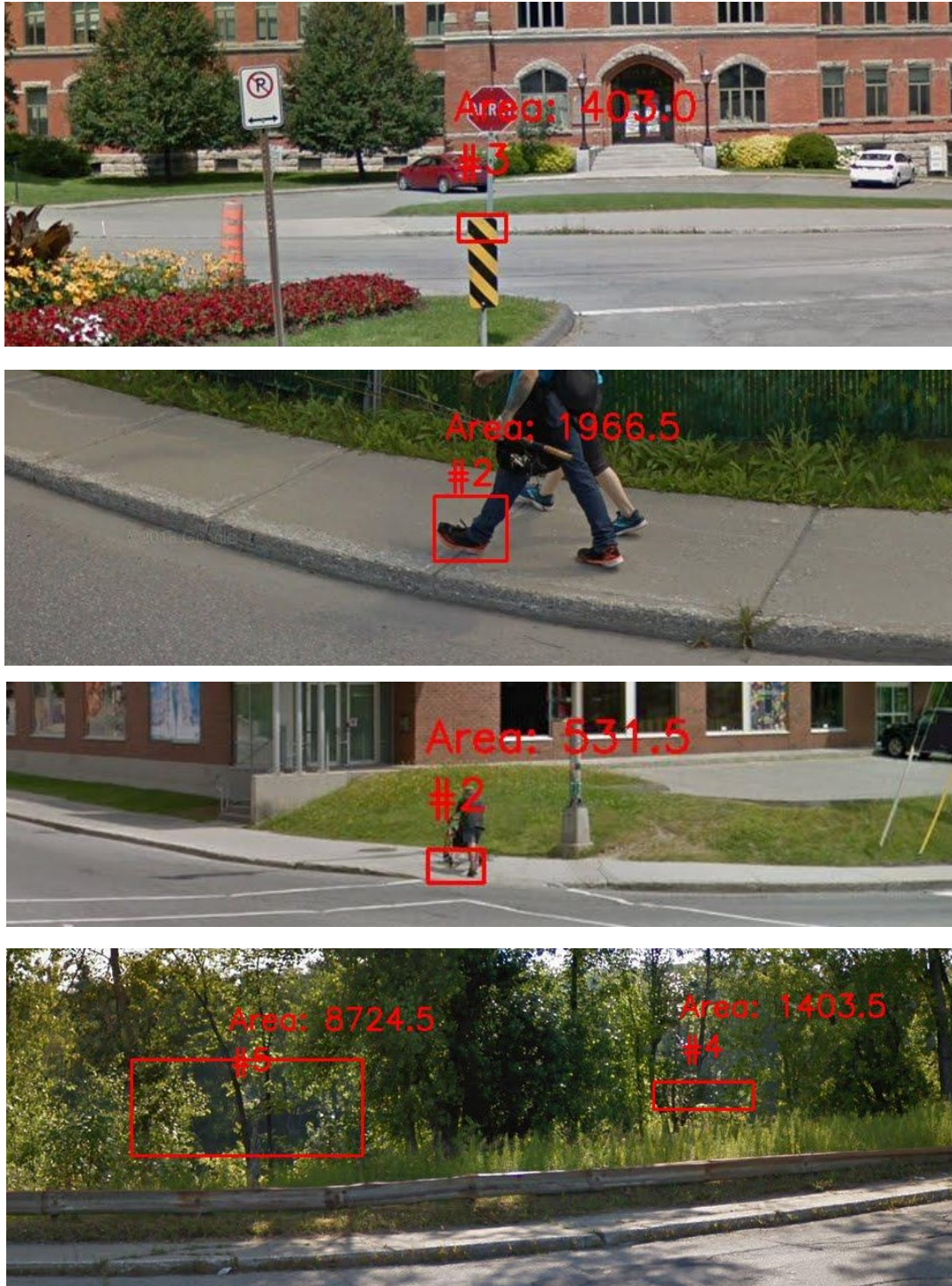


Figure 40 : Exemples de faux positifs liés à une commission du modèle de segmentation du trottoir. Selon la méthodologie adoptée, le modèle de caractérisation n'avait pas été entraîné à rejeter des zones qui ne correspondait pas au trottoir. Chaque défaut est entouré d'une boîte, d'un chiffre unique identifiant ce défaut sur l'image de référence, ainsi que de l'aire, en pixel, du défaut. Ces exemples sont issus de la chaîne de traitement finale (après post-traitement).



Figure 41 : Exemples de faux positifs liés à la sous-représentation d'images captées au printemps et à l'automne. Chaque défaut est entouré d'une boîte, d'un chiffre unique identifiant ce défaut sur l'image de référence, ainsi que de l'aire, en pixel, du défaut. Ces exemples sont issus de la chaîne de traitement finale (après post-traitement).

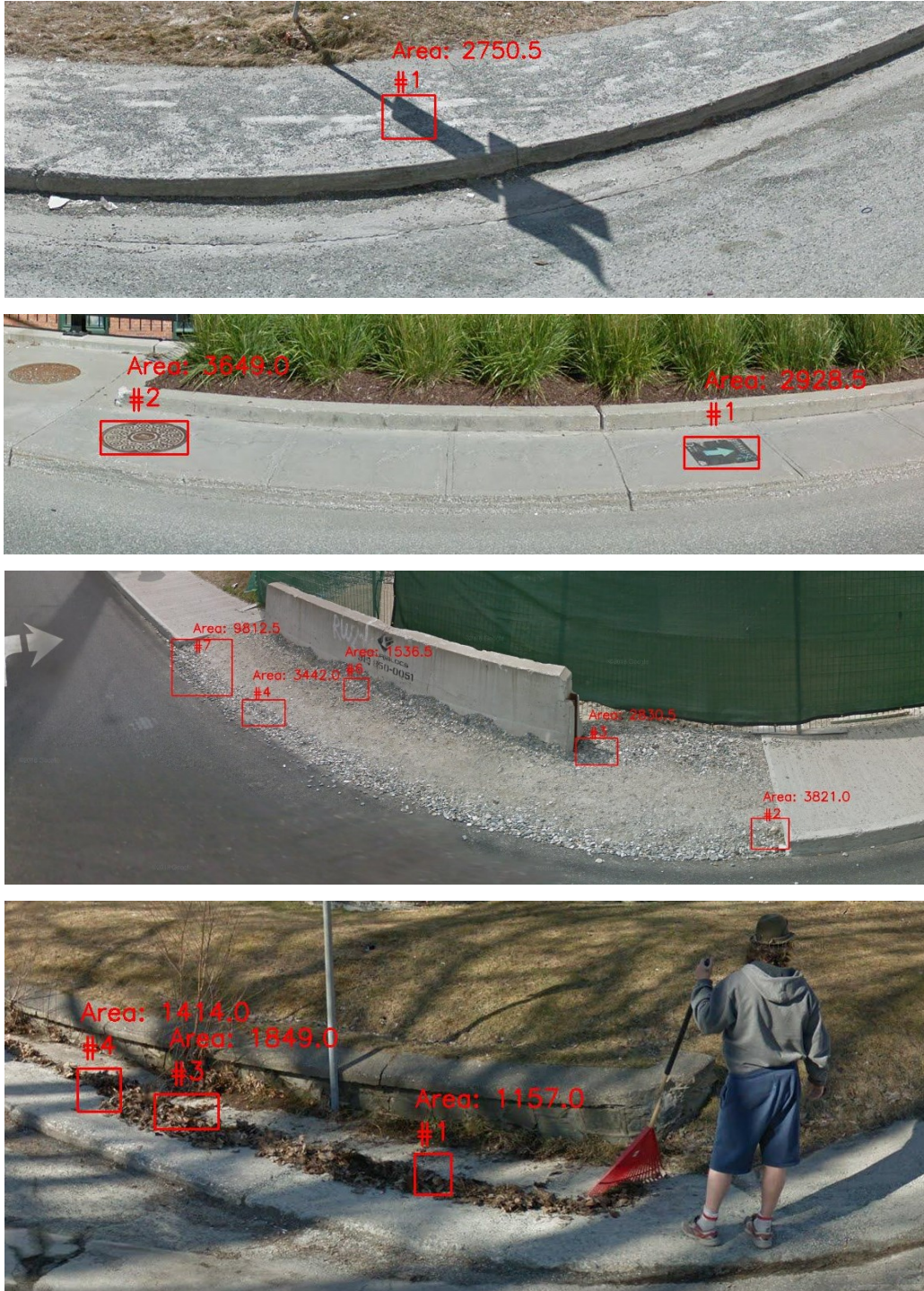


Figure 42 : Exemples de faux positifs liés à la présence d'une ombre ou d'une bouche d'égout, à des travaux de réfection, à un revêtement atypique (pavé uni, asphalté) ou à d'autres événements rares. Chaque défaut est entouré d'une boîte, d'un chiffre unique identifiant ce défaut sur l'image de référence, ainsi que de l'aire, en pixel, du défaut. Ces exemples sont issus de la chaîne de traitement finale (après post-traitement).

3.5.1. Post-traitement

Une analyse visuelle des prédictions avant et après post-traitement a montré l'efficacité des techniques utilisées. En effet, le seuillage ainsi que le filtre appliqué par rapport à la taille des zones extraites comme « Revêtement en mauvais état » ont permis de réduire considérablement la quantité de faux positifs, avec un faible impact sur les faux négatifs.

Chapitre 4. Interprétation des résultats et discussion

4.1. Objectif n°1 : recours à l'apprentissage profond

L'objectif n°1 visant le recours à l'apprentissage profond afin de détecter les trottoirs et caractériser les trottoirs n'a été que partiellement atteint. Un indice de Jaccard de 93 % pour la segmentation des trottoirs est tout à fait comparable aux performances obtenues par Chen *et al.* (2018) pour la classe « trottoir » (indice de Jaccard de 87 %) sur le jeu de données *Cityscapes*. Toutefois, les résultats en ce qui a trait à la caractérisation suggèrent qu'il est plus aisé d'entraîner un réseau de neurones convolutif à segmenter des trottoirs parmi d'autres éléments d'une scène urbaine (voitures, bâtiments, végétation, etc.) que de segmenter des pixels de trottoirs selon l'état de leur revêtement. Trois causes sous-jacentes à ce constat sont proposées :

1. Les caractéristiques devant être apprises par un modèle peuvent revêtir différents niveaux de complexité. Cette complexité peut expliquer qu'un modèle excelle davantage pour la segmentation des trottoirs que pour leur caractérisation ;
2. Un déséquilibre des classes constitue un obstacle à la performance d'un réseau de neurones. Des données d'entraînement comprenant une sous-représentation de pixels pour une classe donnée (ex. : « trottoirs avec défaut majeur ») nuisent à la convergence de l'entraînement ;
3. Les modèles démontrent des résultats mitigés ou peinent à converger (sous-apprentissage) si la quantité de données d'entraînement est insuffisante.

Des explications détaillées pour ces trois causes suivent dans les sections ci-dessous. Par la suite, une approche alternative est proposée afin de diminuer les faux positifs causés par l'utilisation de réseaux distincts pour effectuer des tâches corrélées, celles de la segmentation des trottoirs et de la caractérisation de l'état de leur revêtement.

4.1.1. Niveau de complexité de la tâche

La première cause explique la difficulté qu'a un modèle à distinguer des textures de matériaux dont les caractéristiques dans le domaine RGB sont similaires (voir Figure 43). Dans le cas du présent projet, l'annotation de la vérité terrain pour la tâche de caractérisation des trottoirs a représenté un plus grand défi que l'annotation de trottoirs dans des images de scènes urbaines. En effet, il est plus aisé de distinguer un trottoir d'une route, qu'un revêtement de trottoir faiblement ou fortement endommagé. Malgré la mise en place d'un protocole d'annotation pour la tâche de caractérisation (Figure 44), plusieurs cas limites, dont l'aspect ambigu rendait difficile l'attribution d'une classe plutôt qu'une autre, composaient la banque d'images annotées finale. Cette complexité pour la tâche de classification s'est donc probablement reflétée dans les performances médiocres des modèles entraînés (voir Tableau 4, « défaut mineur »).



Figure 43: Exemple d'un panorama Google Street View où une ambiguïté existe entre la limite du trottoir et celle de la rue, à cause de la présence de gravier.






<i>Échantillon photo</i>					
<i>Classement expert</i>	Bon état	Impraticable	Crevassé profond/large	Lézardé	Impraticable
<i>Classement annotateur</i>	Crevassé profond/large	Crevassé profond/large	Crevassé profond/large	Crevassé profond/large	Impraticable

Figure 44: 5 des 25 échantillons photo soumis pour évaluation auprès d'une experte. Le classement donné par celle-ci a été comparé à celui de l'annotateur et a permis de diminuer le niveau de subjectivité pour l'annotation subséquente sur les données d'entraînement (tâche de caractérisation).

4.1.2. Débalancement des classes

Le Tableau 4 illustre la proportion des pixels annotés selon chaque classe associée à la tâche de caractérisation des trottoirs. Celui-ci permet de constater une sous-représentation élevée des pixels de la vérité terrain associés aux classes « défaut mineur » et « défaut majeur » (classes 2, 3 et 4) pour la tâche de caractérisation.

Lors des tests d'entraînement initiaux, ce débalancement a eu un impact délétère sur l'apprentissage des modèles. En effet, en entraînant ceux-ci à partir du jeu de données d'entraînement entier, un surapprentissage important était observé (Figure 36). Cette observation a mené à l'exclusion de certains panoramas ne présentant aucun défaut afin de mieux équilibrer la représentation des classes dans les données d'entraînement. Malgré ces mesures de sous-échantillonnage, le modèle retenu pour la caractérisation n'a obtenu qu'un indice de Jaccard de 31 % (avant post-traitement).

Le débalancement des classes revêt une importance particulière pour expliquer les résultats dans la tâche principale de ce projet, c'est-à-dire la caractérisation. Les résultats de Ma et al. (2017) démontrent que cet enjeu est également un défi important pour l'évaluation de l'état des chaussées.

Dans de futurs essais, de nouvelles stratégies pour contrer le débalancement des classes pourraient être testées. D'ailleurs, une évaluation systématique des méthodes courantes pour contrer les impacts négatifs de ce débalancement sur les performances des réseaux de neurones convolutifs (Buda *et al.*, 2018) fournit plusieurs pistes utiles pour la suite. Celle-ci observe, par exemple, que le sous-échantillonnage tel qu'appliqué dans le présent projet s'avère peu efficace en comparaison avec le suréchantillonnage. Ainsi, il serait avantageux de répéter des entraînements pour la tâche de caractérisation en assurant une représentativité égale des échantillons « défaut majeur » par suréchantillonnage des données contenant cette classe plutôt que par sous-échantillonnage des données des classes majoritaires (« aucun défaut » et « défaut mineur »). Les résultats obtenus par Japkowicz *et al.* (2002) montrent par ailleurs que ces stratégies pourraient réduire les erreurs de commission, sans pour autant hausser les erreurs d'omission.

4.1.3. Manque de données d'entraînement

La Figure 45 présente la quantité de pixels annotés dans la classe « trottoirs » pour la tâche de segmentation initiale. La banque d'images utilisée pour cette tâche, comparable à des jeux de données de référence, semble donc suffisante pour permettre à des modèles d'atteindre des performances comparables à ceux obtenus sur les jeux de données évoqués.

Ce comparatif n'est pas possible pour la tâche de caractérisation des trottoirs, étant donné la nature inédite du présent projet. Par contre, plusieurs articles soulignent l'importance de la quantité de données d'entraînement pour l'apprentissage de réseaux de neurones et l'impact direct de la disponibilité de ces données afin de prévenir le surapprentissage et augmenter la robustesse des modèles (Hara *et al.*, 2014; Some, 2016; Ma *et al.*, 2017).

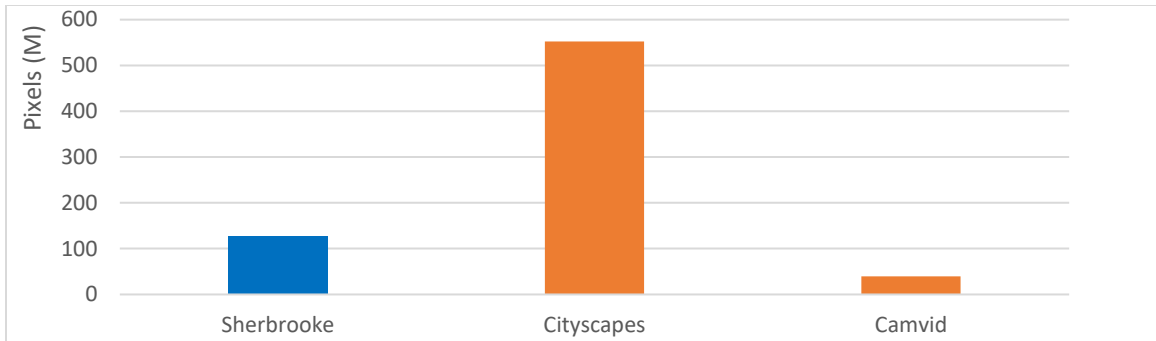


Figure 45: Comparaison du nombre de pixels annotés pour la classe «trottoir» par rapport à deux jeux de données de référence en segmentation sémantique appliquée à l'imagerie urbaine.

Depuis la création d'*AlexNet* en 2010 (Krizhevsky *et al.* 2012), de nombreuses solutions ont été développées afin de pallier un jeu de données de faible taille. Plusieurs d'entre elles, comme le transfert d'apprentissage, la normalisation par lot (*batch normalization*), l'augmentation de données et le recours à l'architecture légère *MobileNetV2* pour la tâche de caractérisation, ont déjà été mises en place dans le présent projet. Toutefois, de nombreuses avenues restent à être explorées afin de pallier davantage ce manque de données ou encore de faciliter l'acquisition de nouvelles données.

Pallier le manque de données d'entraînement

Si certaines techniques d'augmentation des données permettent d'adresser un déséquilibre des classes, tel qu'abordé dans la section précédente, plusieurs d'entre elles ont comme simple objectif de réduire le surapprentissage en gonflant artificiellement une banque d'images d'entraînement. Dans son relevé exhaustif des techniques d'augmentation de données actuelles, Shorten *et al.* (2019) énumère et évalue plusieurs de ces techniques.

Selon le niveau de complexité lié à leur implémentation et leur impact relatif sur le surapprentissage, certaines pourraient être de meilleures candidates pour de futurs essais. Par exemple, une altération aléatoire des couleurs sur les images d'origine serait une méthode d'augmentation simple à intégrer, mais ses bénéfices demeureraient limités. À l'inverse, le recours à des réseaux génératifs (*generative adversarial networks*) afin de créer de nouveaux exemples complètement artificiels serait une voie avec un fort potentiel, mais sa mise en place dans l'algorithme d'entraînement requiert des efforts substantiels (Shorten *et al.*, 2019).

Augmenter le jeu de données stratégiquement

En dépit des bienfaits potentiels liés aux stratégies d'augmentation mentionnées ci-dessus, l'agrandissement réel de la banque d'images d'entraînement annotées demeure l'avenue la plus prometteuse. Celle-ci pourrait diminuer les erreurs de commission et améliorer les performances globales pour la tâche de caractérisation, où le surapprentissage a été observé.

Toutefois, les méthodes de recensement ayant permis d'identifier des zones pour l'acquisition de données du présent projet ont démontré des limites claires. Les emplacements où l'état de revêtement du trottoir est fortement dégradé sont souvent épars : le ratio « nombre de défauts majeurs trouvés » par « kilomètres de trottoirs recensés » s'est avéré très faible, de l'ordre de 1/100. A posteriori, l'investissement de temps afin d'assurer l'acquisition d'un nombre suffisant d'exemples de « défauts majeurs » semble déraisonnable, en particulier pour une acquisition d'envergure comme celle qui est nécessaire pour le recours à l'apprentissage profond.

Ainsi, de futures acquisitions pourraient être ciblées, par exemple, en fonction de données d'auscultation des trottoirs existantes, comme celles acquises en 2019 par la Ville de Montréal (Labbé, 2018) ainsi que dans d'autres villes du Québec. Des démarches auprès de ces municipalités ainsi qu'avec *SNC Lavallin*, la principale entreprise qui réalise ces relevés au Québec, devraient donc être conduites. Pour Ma *et al.* (2017), ce jumelage entre données existantes et acquisition de panoramas *GSV* s'est avéré une méthode efficace pour créer une banque de plus de 700 000 images d'entraînement en vue de la caractérisation de l'état de la chaussée (Figure 46).



Figure 46: Chaîne de traitement pour l'acquisition de données d'entraînement tel qu'utilisé par Ma *et al.* (2017). Les panoramas GSV sont acquis automatiquement à partir de données publiques existantes pour la ville de New York. (tiré de Ma *et al.*, 2017, p.3)

Des futurs efforts pourraient également faire appel à une chaîne de traitement itérative telle que réalisée par Lutnik *et al.* (2018) afin d'utiliser des prédictions comme points de départ pour l'annotation de nouvelles données. Quoique de telles itérations ont été réalisées au cours du projet, aucune chaîne de traitement n'a été montée afin d'en accélérer le processus. De plus, ces efforts doivent préalablement prévoir suffisamment de panoramas à partir desquels effectuer une prédiction pour assurer une meilleure efficacité durant ces itérations entre prédictions et annotation. Enfin, étant donné la présence marquée de faux positifs dans les résultats présentés, l'annotation de prédictions présentant ce type d'erreurs pourrait être priorisée afin d'agrandir la banque d'images annotées et améliorer les performances des modèles.

4.1.4. Faux positifs : une lacune possible des réseaux travaillant en silo

En outre, comme l'indique la proportion de faux positifs issus du modèle de segmentation, la caractérisation du revêtement des trottoirs est d'autant plus ardue lorsque les tâches de segmentation et de caractérisation sont assumées par des réseaux indépendants. Dans ce contexte, une erreur de commission par le réseau responsable de segmenter les trottoirs a un impact direct, en aval, sur les performances du réseau effectuant la caractérisation du revêtement. Quoiqu'une telle approche visant à appliquer un masque afin de limiter les traitements à une région d'intérêt (ex. : pixels de trottoirs seulement) ait démontré de bons résultats (Some, 2016), d'autres approches sont susceptibles de donner de meilleurs résultats dans des projets futurs. Bischke *et al.* (2019), par exemple, démontre l'efficacité de soumettre un seul réseau de neurones convolutif à l'exécution de deux tâches semblables par le recours à un apprentissage commun et simultané (Figure 47). Dans le contexte de la caractérisation de l'état du revêtement des trottoirs, cette approche « multitâches » permettrait à un seul réseau d'apprendre à la fois à distinguer un pixel « trottoir » d'un pixel « non-trottoir » et à classer le pixel « trottoir » entre différentes classes d'état du revêtement. En développant des caractéristiques communes à ces deux tâches corrélées, il est probable que les cas de faux positifs causés par l'entraînement indépendant des deux réseaux du présent projet diminuent considérablement.

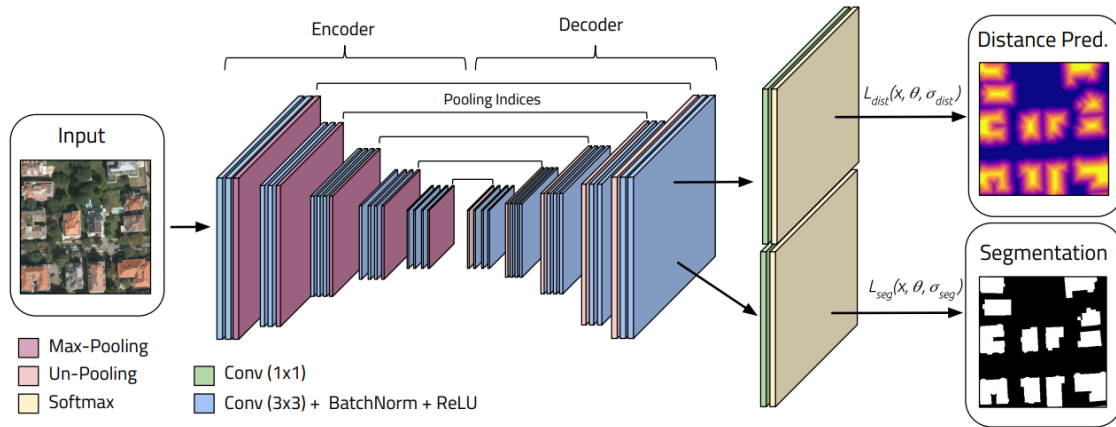


Figure 47: Structure d'un réseau de neurones multitâches en segmentation sémantique tel que proposé par Bischke et al. (2019), p.1482

4.2. Objectif n°2 : robustesse du modèle sur le site d'évaluation finale

Les performances médiocres du modèle de caractérisation sur les panoramas du site d'étude principal rendent difficile l'évaluation de la robustesse de ce modèle. En effet, les occurrences multiples de faux positifs sont davantage attribuables à la faiblesse initiale du modèle qu'à un écart de performance entre les données des sites d'entraînement (incluant l'ensemble de validation) et celles du site d'évaluation finale. Néanmoins, le débalancement des classes, tel que soulevé précédemment, a un impact sur la robustesse du modèle lorsque celui-ci est évalué sur l'ensemble de validation (Buda *et al.*, 2018). De futurs modèles qui contrent plus efficacement l'impact néfaste de ce débalancement pourraient donc gagner en robustesse. En outre, l'agrandissement du jeu de données d'entraînement pour la tâche de caractérisation recèle le plus grand potentiel afin d'améliorer globalement les performances, donc la robustesse sur le site d'évaluation finale.

En ce qui a trait au modèle de segmentation des trottoirs, une évaluation à la fois visuelle et quantitative révèle des lacunes au niveau de la robustesse de ce modèle. En fait, le taux d'erreurs de commissions sur les données du site d'évaluation finale était plus élevé que sur l'ensemble de validation des sites d'entraînement (pour plus de détails sur la séparation des ensembles, voir section 2.8.1 « *Tâche 1 : segmentation des trottoirs* »). Cela signifie

que les modèles entraînés peinaient à apprendre des caractéristiques permettant une performance stable sur des données n'ayant pas été utilisées durant l'entraînement. Cette baisse de performance est courante et est attribuable à une utilisation répétée de l'ensemble de validation pour ajuster les hyperparamètres (Goodfellow *et al.*, 2016). Il en résulte une faible surestimation de la robustesse du modèle par rapport à l'évaluation finale, sur les données du site d'évaluation finale. Comme pour la tâche de caractérisation, le modèle de segmentation pourrait donc être plus robuste si la banque d'images d'entraînement était agrandie, en couvrant plus de villes québécoises, par exemple.

Enfin, une utilisation accrue des stratégies d'augmentation de données au moment de l'inférence (*test-time augmentation*), c'est-à-dire au moment de produire des prédictions sur de nouvelles données, semble efficace pour améliorer la robustesse (Figure 48) (Shorten *et al.*, 2019). Celle-ci permettrait de combiner plusieurs prédictions issues d'une même image ayant subi différentes transformations géométriques et radiométriques afin de hausser la qualité de la prédiction finale, après regroupement.

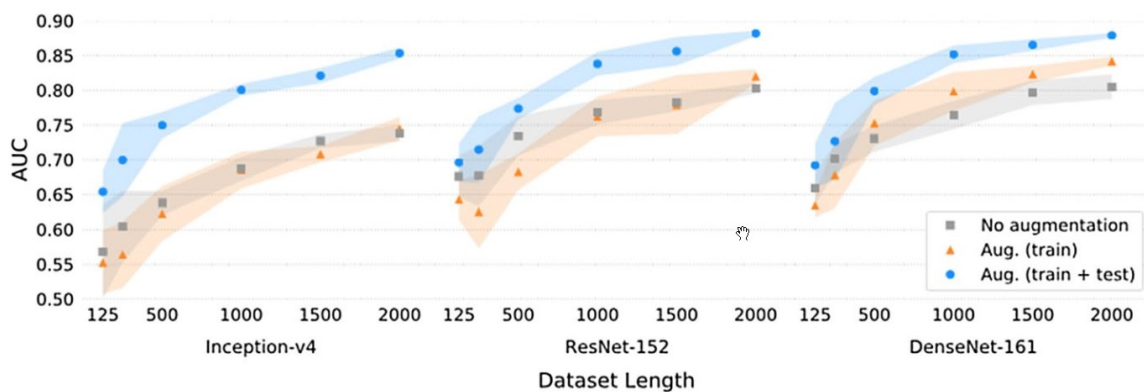


Figure 48: Impact de l'augmentation de données au moment de l'inférence (courbe « train + test »). La mesure en ordonnée correspond à l'aire sous la courbe ROC, une mesure de performance commune en classification. (tiré de Shorten *et al.*, 2019, p.36)

4.3. Objectif n°3 : avantages, inconvénients et limitations

4.3.1. Recours à l'apprentissage profond

Dans le présent projet, le recours à l'apprentissage profond s'est distingué comme une solution adaptée pour effectuer la tâche de segmenter les trottoirs au sein d'une scène urbaine sur des panoramas *Google Street View*. Toutefois, le plein potentiel de ces algorithmes d'intelligence artificielle n'a pas pu être évalué pour la caractérisation de l'état du revêtement des trottoirs. En effet, le surapprentissage observé révèle donc un inconvénient majeur aux algorithmes d'apprentissage profond : leurs performances dépendent de la quantité de données d'entraînement qu'ils ingèrent.

En outre, la grande disponibilité de modèles d'apprentissage profond entraînés sur des jeux de données liés au développement des véhicules autonomes constitue un atout de taille pour la segmentation d'éléments urbains comme les trottoirs. D'ailleurs, la revue de littérature effectuée indique que plusieurs chercheurs valorisent ces réseaux de neurones préentraînés lorsque les données sont insuffisantes pour une tâche apparentée (Brostow et al., 2009; Sengupta et al., 2012; Kim et al., 2017).

4.3.2. Utilisation de panoramas *GSV*

D'emblée, l'utilisation de photos *GSV* comporte une limite importante : la licence rattachée aux données *GSV* ne permet pas l'utilisation de ces données à l'extérieur du contexte académique. Ainsi, si la méthode devait être déployée à l'échelle commerciale, de nouveaux clichés devraient être acquis. Dans ce contexte, mis à part le coût potentiellement prohibitif de l'acquisition de ces données, le présent modèle pourrait devoir être adapté selon l'équipement utilisé et le protocole d'acquisition afin de générer des résultats similaires.

De plus, afin d'obtenir un portrait juste de l'état des infrastructures à l'étude, la présente méthode dépend d'un renouvellement fréquent des données. En ce qui a trait aux panoramas *GSV*, plus une ville est petite, plus le délai de renouvellement risque d'être important (Ma et al., 2017). Les panoramas pour la Ville de Sherbrooke ciblés par le

présent projet s'étendent d'ailleurs d'octobre 2011 pour les rues moins fréquentées à août 2018 pour les artères principales. Spécifiquement, 22 % et 56 % des panoramas ont été extraits en 2018 et 2017, respectivement. Les 22 % restants dataient d'avant 2017, ce qui indique une limitation importante à l'utilisation de tels panoramas. Ainsi, si la méthode développée dans le cadre du présent projet s'avérait parfaitement efficace pour caractériser les trottoirs, il n'en demeure pas moins que le portrait réalisé ne représenterait pas forcément l'état des trottoirs au moment de la caractérisation. Le projet *MobiliSIG*, apparenté au présent projet, doit également composer avec cette limite importante (Mostafavi, 2018). Ces limites peuvent être levées si la Ville de Sherbrooke décide d'acquérir annuellement des images proximales spécifiquement pour cette application.

De surcroît, il importe de mentionner que la solution proposée n'est pas adaptée au traitement en temps réel de nouvelles données tel que retenu par Tsai *et al.* (2010). Si cet aspect comporte un intérêt certain pour la caractérisation rapide de l'état du revêtement des trottoirs sur de grandes quantités de données, la méthodologie du présent projet devra être remaniée en profondeur pour pouvoir s'adapter à une telle contrainte.

Enfin, les métadonnées extraites conjointement avec les 2609 panoramas ont permis de constater que la majorité des panoramas *GSV* ont été acquis durant les mois d'été (76 % des panoramas en juillet et août). Pour le présent projet, cela constitue sans doute un avantage en lien avec l'utilisation des panoramas *GSV*, étant donné que le revêtement des trottoirs est davantage exposé qu'au printemps, par exemple, alors qu'il reste encore du gravier fin sur les trottoirs.

4.4. Retour sur l'hypothèse de recherche

À la lumière de ce qui a été présenté ci-dessus, les résultats du présent projet ne permettent pas de démontrer que la classification d'images panoramiques urbaines par apprentissage profond est en mesure de fournir une bonne approximation de l'état du revêtement des trottoirs dans les limites de la méthode retenue.

4.5. Discussion

4.5.1. Alternatives aux algorithmes utilisés

Le surapprentissage observé durant l'entraînement des modèles de caractérisation de l'état du revêtement des trottoirs indique que des travaux subséquents devront considérer cet écueil important au moment de l'élaboration de leur méthodologie. Une solution simple, si les ressources le permettent, serait de décupler la quantité de données annotées afin de minimiser ce problème. Dans ce cas, de futurs projets de recherche devront donc accorder une importance particulière à la planification de l'acquisition et l'annotation de ces données, coûteuses à produire. Autrement, différentes techniques, comme le suréchantillonnage, pourraient également contrer l'effet délétère du déséquilibre des classes afin d'évaluer adéquatement le potentiel des réseaux de neurones.

Enfin, rappelons que la littérature consultée indique sans équivoque la supériorité d'algorithmes d'apprentissage machine par rapport à des méthodes conventionnelles de classification de textures et de matériaux pour distinguer différents éléments d'un panorama urbain (ex. : piéton, voiture, bâtiment, trottoir) (Ma *et al.*, 2017), ainsi que différents stades de dégradation d'un même revêtement (Some 2016; Zhang *et al.*, 2016; Gopalakrishnan *et al.*, 2017; Gopalakrishnan *et al.*, 2018) comme le béton des trottoirs. Toutefois, une approche de segmentation sémantique par apprentissage profond ne constitue pas la seule voie à considérer pour des travaux futurs. Plusieurs alternatives en apprentissage profond ou, plus largement, en apprentissage machine revêtent un intérêt particulier afin d'évaluer adéquatement ces méthodes d'intelligence artificielle pour aborder la problématique visée par le présent projet. Par exemple, le recours aux arbres de décision pourrait s'avérer plus approprié (Ma *et al.*, 2017).

En apprentissage profond, la classification et la détection d'objets, c'est-à-dire l'attribution d'une classe pour une image entière ou une région d'intérêt, respectivement, pourraient également être explorées. En effet, les tâches de segmentation et de caractérisation des trottoirs, avec la méthodologie retenue pour ce projet, ne nécessitent pas une précision de l'ordre du pixel. Une approche de classification, par exemple, pourrait traiter des zones de

petite taille (ex. 500 x 500 pixels) rognées à partir d'un panorama entier et aboutir à une caractérisation dont la précision spatiale demeurerait raisonnable dans le contexte de la problématique abordée. En outre, la constitution d'une banque d'images d'entraînement se trouverait facilitée par le recours à la classification, car l'effort d'annotation serait réduit à l'attribution d'une classe pour chaque zone ciblée.

Une telle approche permet également de tirer profit d'innovations constantes issues des nombreuses publications scientifiques qui ciblent le domaine de la classification par apprentissage profond. À titre d'exemple, l'état de l'art pour des jeux de données de référence comme *ImageNet* est constamment surpassé par de nouvelles architectures comme *EfficientNet* (Tan *et al.*, 2019). Celles-ci sont publiées à un rythme soutenu en comparaison avec les architectures de segmentation sémantique, un domaine dont les avancées dépendent souvent des travaux préalablement publiés dans le domaine de la classification (Long *et al.*, 2015; Chen *et al.*, 2018). De plus, la classification demande une capacité de calcul plus faible que la segmentation sémantique, ce qui confère un avantage supplémentaire à cette approche. Enfin, advenant le maintien d'une approche par apprentissage profond, le recours au transfert d'apprentissage s'impose afin de maximiser les performances de futurs modèles.

Pour conclure, plutôt que de démontrer l'inefficacité de l'apprentissage profond pour aborder la problématique de caractérisation des trottoirs, les présents résultats suggèrent plutôt que de futurs efforts de recherche sont nécessaires pour évaluer le plein potentiel de l'apprentissage machine, voire de l'apprentissage profond pour la problématique visée.

4.5.2. Alternatives aux panoramas GSV

De la même manière qu'une approche par segmentation sémantique ne constitue pas l'approche unique à considérer, le recours à d'autres données que les panoramas GSV devrait faire l'objet d'une réflexion approfondie. À titre d'exemple, certains chercheurs ont utilisé des images acquises à partir d'un téléphone intelligent pour caractériser l'état de revêtement des chaussées (Zhang *et al.*, 2016b; Maeda *et al.*, 2018). Un modèle entraîné sur ce type de données comporte l'avantage qu'il pourrait être utilisé conjointement avec des images acquises via une application de contribution par la foule (*crowdsourcing*),

débouchant sur une banque d'images d'entraînement de grande ampleur. Les données laser issues de la technologie *PaveVision3D* (Figure 49) émergent également comme une solution d'intérêt pour la caractérisation de l'état de la chaussée (Luo *et al.*, 2014; Zhang *et al.*, 2017). Ce type de données présente plusieurs avantages en vue de la caractérisation de revêtements en béton. Par exemple, la qualité des données n'est pas autant affectée par les conditions lumineuses et météorologiques au moment de l'acquisition. La résolution de 1 mm de ces données confère une précision d'une grande valeur pour la caractérisation de la surface d'un trottoir, mais dresse également certaines embûches en lien avec les ressources informatiques nécessaires pour en effectuer le traitement.

De plus, certaines entreprises comme *Jakarta* (jakarto.com) produisent des données d'imagerie urbaine RGB semblables à celles acquises par *Google*. Toutefois, celles-ci sont jumelées spatialement et temporellement à une acquisition *Lidar* (*Light Detection and Ranging*). Ces données *Lidar* comportent un riche potentiel pour la caractérisation de l'état du revêtement des trottoirs, notamment par la plus grande versatilité qu'elles confèrent au niveau de la méthodologie. Par exemple, les panoramas RGB pourraient être utilisés d'une manière semblable à celle retenue dans le présent projet, mais un post-traitement incluant les données *Lidar* permettrait d'obtenir davantage d'informations sur le défaut identifié. Par ailleurs, l'utilisation des nuages de points, en conjonction ou non avec les panoramas RGB, offre un potentiel considérable pour la caractérisation des trottoirs, étant donné la description au niveau de la texture et le géoréférencement précis qu'ils permettent.



Figure 49: Système *PaveVision3D* avec résolution de 1 mm . a) extérieur et (b) fonctionnement (tire de Luo *et al.*, 2014, p.122)

Conclusion

Des perspectives prometteuses se dégagent du présent projet, notamment par le débroussaillage qu'il a permis. De nombreuses stratégies technologiques restent à évaluer et ont le potentiel de générer des résultats de plus en plus probants. À ce titre, l'entraînement de réseaux de neurones en segmentation sémantique avec des panoramas urbains RGB suggère qu'une approche alternative en apprentissage profond, avec des réseaux de classification par exemple, permettrait de faciliter l'évaluation approfondie du potentiel de ces nouveaux outils d'intelligence artificielle pour la caractérisation de l'état de revêtement des trottoirs. De plus, il semble clair qu'une banque d'images annotées comptant quelques milliers d'exemplaires plutôt que quelques centaines s'impose afin d'éviter les écueils liés au surapprentissage. Ce phénomène, qui a empêché l'obtention de modèles de caractérisation performants et robustes sur de nouvelles données, doit donc être planifié d'entrée de jeu pour de futurs projets. Parmi les solutions proposées dans la section précédente, l'apport de bases de données existantes en lien avec l'auscultation des trottoirs demeure une piste plus que pertinente afin d'automatiser l'annotation d'une grande quantité de données d'entraînement.

Par la suite, des efforts complémentaires à l'agrandissement de la banque d'image d'entraînement pourraient être investis afin d'améliorer le résultat final. Par exemple, des méthodes classiques de traitement d'image comme l'application de filtres *HSV* pourraient prendre place à l'étape de post-traitement. Cela permettrait, entre autres, d'éliminer certains faux positifs liés à l'attribution de la classe « trottoir » à des pixels de végétation par le modèle de segmentation, par exemple.

Enfin, le présent projet s'est intéressé à une caractérisation de l'état du revêtement des trottoirs visant à la fois l'accessibilité des infrastructures pour les PMR et pour des fins de gestion des infrastructures urbaines à l'instar d'une auscultation traditionnelle. A posteriori, même si ces deux types de caractérisation présentent des points communs, celles-ci doivent faire l'objet de méthodologies distinctes. En effet, l'évaluation de l'état du revêtement pour assurer l'accessibilité et l'utilisabilité (voir section 1.3.1 « Conception pour tous : le cas des trottoirs ») d'un trottoir pour une personne en fauteuil roulant est une

tâche qui nécessite une connaissance approfondie des besoins de ces usagers. L'auscultation telle qu'effectuée par certaines municipalités nécessite, quant à elle, une caractérisation descriptive et détaillée en vue de mesurer précisément les défauts et leur orientation, ainsi que d'identifier la cause et le niveau de la dégradation (Wang *et al.*, 2017; Blanchette, 2011). À la différence d'une caractérisation pour des fins d'accessibilité universelle, ce type d'évaluation intègre également une dimension préventive afin de favoriser une intervention rapide avant que l'infrastructure visée ne se détériore davantage (Tsai *et al.*, 2010).

Ainsi, malgré la ressemblance entre ces deux tâches, des travaux subséquents devront choisir l'une ou l'autre de ces problématiques, car, à la lumière des résultats du présent projet, ces deux types de caractérisation demeurent fondamentalement différents et nécessitent leur méthodologie propre.

En ce qui a trait à l'apport du projet dans le domaine de l'accessibilité universelle, une méthode permettant la caractérisation de l'état du revêtement des trottoirs pour assurer le passage sûr des personnes en fauteuil roulant ne constitue pas une solution exhaustive afin de répondre aux besoins de ces usagers. La solution proposée, même si les résultats avaient été exemplaires, comporte plusieurs limites majeures :

1. Elle se limite à un type d'infrastructures piétonnes, les trottoirs, alors que les usagers concernés circulent également sur d'autres types de corridors piétonniers (traverses piétonnes, sentiers aménagés et non aménagés, autres accotements de route) ;
2. L'état de revêtement constitue une variable parmi plusieurs autres (type de revêtement, pente longitudinale, pente transversale, présence d'obstacles temporaires ou permanents, présence de glace durant l'hiver) qui doivent être considérées pour assurer l'accès des trottoirs ;
3. La caractérisation issue d'une méthode telle que développée par le présent projet vise ultimement à permettre la création de trajets adaptés dans plusieurs villes via une application pour téléphone intelligent comme *MobiliSIG*. En soi, ce volet requiert ses propres efforts de recherche et ressources afin de développer un produit utile à grande échelle.

Des travaux ultérieurs devront donc prendre en considération ces limites et envisager une méthode qui puisse les surmonter, du moins en partie.

De plus, comme le projet *MobiliSIG*, les efforts futurs devront éventuellement intégrer la participation de la clientèle à mobilité réduite concernée. Celle-ci pourra, par exemple, évaluer elle-même le degré d'accessibilité d'un trottoir selon des facteurs personnels (ex.: force physique) et des habitudes de vie propre à chacun (Fougeyrollas, 2019). Une fissure considérée « défaut mineur » dans le cadre de ce projet pourrait représenter un irritant majeur dans certains contextes. Par exemple, une telle situation pourrait survenir si le défaut se trouve dans une côte à forte pente ou si l'utilisateur n'a pas la capacité de le franchir. À l'inverse, un défaut majeur sur le revêtement du trottoir n'est pas nécessairement un obstacle si l'individu peut le contourner aisément et en toute sécurité.

Somme toute, il convient de souligner qu'aucune recherche répertoriée n'applique les méthodes de traitement d'images par apprentissage profond à la problématique de la mobilité des personnes en fauteuil roulant. En soi, les travaux réalisés revêtent donc un caractère innovant, car la méthodologie développée pourra servir de pierre d'assise pour des travaux de recherche futurs.

Références

1. Ahmad, J., Muhammad, K., & Baik, S. W. (2017). Data augmentation-assisted deep learning of hand-drawn partially colored sketches for visual search. In *PloS one*, vol. 12, n°8.
2. Akasaba, R. (2018). Introducing “wheelchair accessible” routes in transit navigation. Google Blog, <https://www.blog.google/products/maps/introducing-wheelchair-accessible-routes-transit-navigation/>
3. Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R. & Weaver, J. (2010). Google Street View: Capturing the world at street level. In *Computer*, 43(6), 32-38.
4. Association Nationale pour le Logement des Handicapés. (1999). Accessibilité des voiries – un espace public pour tous: les cheminements horizontaux. <http://www.anlh.be/accessvoirie/acc06.htm>
5. Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. In *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, n°12, pp. 2481-2495.
6. Beneficial Designs Inc., Kirschbaum, J. B., Axelson P. W., Longmuir, P. E., Mispagel, K. M., Stein, J. A. and Yamada, D. A. (2001). Designing Sidewalks and Trails for Access. Part II of II: Best Practices Design Guide. 484 p.
7. Bischke, B., Helber, P., Folz, J., Borth, D., and Dengel, A. (2019, September). Multi-task learning for segmentation of building footprints with deep neural networks. In *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 1480-1484.
8. Blanchette, A. (2011). Formation technique : Gestion des bordures et trottoirs. Conférences et présentations (INFRA), Québec, 69 p. https://ceriu.qc.ca/system/files/Auscultation_et_inspection_des_trottoirs.pdf
9. Brostow, G. J., Fauqueur, J., and Cipolla, R. (2009). Semantic object classes in video: A high-definition ground truth database. In *Pattern Recognition Letters*, vol. 30, n°2, pp. 88-97.
10. Buda, M., Maki, A., and Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. In *Neural Networks*, vol.106, pp. 249-259.
11. Canziani, A., Molnar, T., Burzawa, L., Sheik, D., Chaurasia, A. and Culurciello, E. (2018). An analysis of deep neural network. Medium.com. <https://medium.com/@culurciello/analysis-of-deep-neural-networks-dcf398e71aae>

12. Centre de recherches routières. (2009). Manuel relatif à la réalisation pratique des passages pour piétons: Passages pour piétons adaptés aux personnes à mobilité réduite, F47, feuillet 10 version 1, 13 p.
13. Centre suisse pour la construction adaptée aux handicaps. (2003). Directives «Voies piétonnes adaptées aux handicapés», 32 p.
14. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. In *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, n°4, pp. 834-848.
15. Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 801-818.
16. Cheng, H., Wang, J., Hu, Y., Glazier, C., Shi, X., & Chen, X. (2001). Novel approach to pavement cracking detection based on neural network. In *Transportation Research Record: Journal of the Transportation Research Board*, (1764), pp. 119-127.
17. City of Toronto. (2004). City of Toronto Accessibility Design Guidelines. 137 p.
18. Chollet, F. (2017). Deep learning with python. *Manning Publications Co.*.
19. Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. arXiv preprint, 1610-02357.
20. Conseil national de recherche Canada. (2005). Code de construction du Québec, Chapitre I – Bâtiment, et Code national du bâtiment – Canada, 1281 p.
21. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R. & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213-3223.
22. CRIM (Centre de recherche en informatique de Montréal). (2019). Communication personnelle.
23. Dahmane, M., Foucher, S., Beaulieu, M., Riendeau, F., Bouroubi, Y. & Benoit, M. (2016). Object detection in pleiades images using deep features. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 1552-1555.
24. Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Basu, S., ... & Raska, R. (2018). Deepglobe 2018: A challenge to parse the earth through satellite images. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. pp. 172-179.

25. Doersch, C., Singh, S., Gupta, A., Sivic, J., & Efros, A. (2012). What makes paris look like paris?. In *ACM Transactions on Graphics*, vol. 31, n°4.
26. Du, Y., Pan, N., Xu, Z., Deng, F., Shen, Y., & Kang, H. (2020). Pavement distress detection and classification based on YOLO network. In *International Journal of Pavement Engineering*, pp. 1-14.
27. Dubé, C. (2018). 46 % du réseau routier de Sherbrooke qualifié de très mauvais ou mauvais. Estrieplus.com, http://estrieplus.com/contenu-reseau_routier_sherbrooke_etat_bilan-1355-44100.html
28. Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning. 31 p.
29. Duvall, J., Cooper, R., Sinagra, E., Stuckey, D., Brown, J. and Pearlman, J. (2013). Development of surface roughness standards for pathways used by wheelchairs. *Transportation Research Record: Journal of the Transportation Research Board*, volumen 2387, p.149-156.
30. Fougeyrollas, P. (2010). La funambule, le fil et la toile: transformations réciproques du sens du handicap. Québec, Presses Université Laval, 315 p.
31. Fougeyrollas, P., Fiset, D., Dumont, I., Grenier, Y., Boucher, N. & Gamache, S. (2019). Réflexion critique sur la notion d'accessibilité universelle et articulation conceptuelle pour le développement d'environnements inclusifs. 9-9. 17 p.
32. Geiger, A., Lenz, P., Stiller, C., & Urtasun, R. (2013). Vision meets robotics: The KITTI dataset. *The International Journal of Robotics Research*, 32(11), pp. 1231-1237.
33. Gendron, P.-É. (2014). L'accessibilité universelle : un point de départ pour des aménagements piétons de qualité. La ville sous nos pieds : connaissances et pratiques favorables aux mobilités piétonnes, actes du 4e colloque francophone international du GERI COPIE, 20-22 novembre 2013, Centre – Urbanisation Culture Société, Institut national de la recherche scientifique, Montréal, pp. 109-123.
34. Goodfellow, I., Bengio, Y., Courville, A. and Bengio, Y. (2016). *Deep learning* (Vol. 1). Cambridge: MIT press. 800 p.
35. Gopalakrishnan, K., Khaitan, S. K., Choudhary, A., & Agrawal, A. (2017). Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. *Construction and Building Materials*, 157, 322-330.
36. Gopalakrishnan, K. (2018). Deep Learning in Data-Driven Pavement Image Analysis and Automated Distress Detection: A Review. *Data*, 3(3), 28.
37. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S. and Lew, M. S. (2016). Deep learning for visual understanding: A review. In *Neurocomputing*, vol. 187, p. 27-48.

38. Hara, K. (2014). Scalable methods to collect and visualize sidewalk accessibility data for people with mobility impairments. In *Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology*. pp. 1-4.
39. He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, p. 770-778.
40. Hensman, P., & Masko, D. (2015). The impact of imbalanced training data for convolutional neural networks. Degree Project in Computer Science, KTH Royal Institute of Technology. 28 p.
41. Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In *CVPR*, vol. 1, n°2, p. 3.
42. Île de France. (année inconnue). Vianavigo. <https://www.vianavigo.com/accueil>
43. Japkowicz, N., and Stephen, S. (2002). The class imbalance problem: A systematic study. In *Intelligent data analysis*, vol. 6, n°5, pp. 429-449.
44. Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, vol. 6, n°1, 27 p.
45. Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097-1105.
46. Labbé, J. (2018). Montréal prendra le pouls de ses trottoirs en 2019. Radio-Canada. <https://ici.radio-canada.ca/nouvelle/1138564/operation-auscultation-reseau-arteriel-raav-pietons>
47. Lam, D., Kuzma, R., McGee, K., Dooley, S., Laielli, M., Klaric, M., Bulatov, Y. and McCord, B. (2018). xView: Objects in Context in Overhead Imagery. *arXiv preprint arXiv:1802.07856*.
48. Lawrence, N. (2015) OpenAI won't benefit humanity without data-sharing. The Guardian, <https://www.theguardian.com/media-network/2015/dec/14/openai-benefit-humanity-data-sharing-elon-musk-peter-thiel>.
49. LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. In *Nature*, vol. 521, n°7553, p. 436.
50. Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... and Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In *European conference on computer vision*, Springer, Cham, pp. 740-755. *cocococo*

51. Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431-3440.
52. Mack, D. (2017). *WE-MAP* (Doctoral dissertation, Haute Ecole de Gestion and Tourisme).. <https://www.hevs.ch/en/rad-institutes/institute-of-information-systems/projects/wemap-crowdsourced-map-based-mobile-application-for-the-wheelchair-accessibility-in-switzerland-15169>
53. Maeda, H., Sekimoto, Y., and Seto, T. (2016). Lightweight road manager: smartphone-based automatic determination of road damage status by deep neural network. In *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Mobile Geographic Information Systems*. pp. 37-45.
54. Magnusson, M. L., Pope, M. H., Wilder, D. G. and Areskou, B. (1996). Are occupational drivers at an increased risk for developing musculoskeletal disorders?. *Spine*, vol. 21, n°6, pp. 710-717.
55. Ministère de la Région de Bruxelles-Capitale (2008). *Vademecum : personnes à mobilité réduite dans l'espace public*, 74 p.
56. Mobasheri, A., Deister, J. and Dieterich, H. (2017). Wheelmap: The wheelchair accessibility crowdsourcing platform. In *Open Geospatial Data, Software and Standards*, vol. 2, n°1, p. 27.
57. Moro, I. (2007). *Un espace public pour tous - Guide pour une planification cohérente*. République et Canton de Genève, [Département du territoire] : Equiterre, 117 p.
58. Mostafavi, M. A. and Fiset, D. (2014) MobiliSIG: Une technologie géospatiale innovante au service des personnes à mobilité réduite. In *Géomatique*, vol. 41, n°1, pp. 8-11.
59. Mostafavi, M. A. (2018). Communication personnelle.
60. Nations Unies (2006), *Convention relative aux droits des personnes handicapées*, Résolution 61/106, 13 décembre 2006, Nations Unies. 38 p.
61. Office des personnes handicapées du Québec (1984). *À part... égale : L'intégration sociale des personnes handicapées: un défi pour tous*. Drummondville, Secrétariat général, L'Office, 447 p.
62. Office des personnes handicapées du Québec (2011). *Avis de l'Office des personnes handicapées du Québec : Au regard des recommandations de l'Avis de santé publique sur la circulation des aides à la mobilité motorisées sur le réseau routier*. Drummondville, Secrétariat général, L'Office, 447 p.
63. Office des personnes handicapées du Québec (2017). *Évaluation de l'efficacité de la politique gouvernementale À part entière: pour un véritable exercice du droit à*

l'égalité : les déplacements des personnes handicapées : l'accès aux transports et l'accessibilité des bâtiments et des lieux publics, Drummondville, Secrétariat général, L'Office, 169 p.

64. Office des personnes handicapées du Québec (2018). Estimations de population avec incapacité en 2012 : région administrative de l'Estrie, Drummondville, Directon de l'évaluation et du soutien à la mise en oeuvre de la loi, L'Office, 76 p.
65. Organisation mondiale de la Santé (2001). Classification internationale du fonctionnement, du handicap et de la santé: CIF. Genève, 304 p.
<http://apps.who.int/iris/handle/10665/42418>
66. Parlement européen (2001). Directive 2001/85/CE du Parlement européen et du Conseil du 20 novembre 2001 concernant des dispositions particulières applicables aux véhicules destinés au transport des passagers et comportant, outre le siège du conducteur, plus de huit places assises, et modifiant les directives 70/156/CEE et 97/27/CE. Article 2.21. <https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=CELEX%3A02001L0085-20130701>
67. Pauly, L., Hogg, D., Fuentes, R., and Peel, H. (2017). Deeper networks for pavement crack detection. In *Proceedings of the 34th ISARC, IAARC*, pp. 479-485.
68. Penatti, O. A., Nogueira, K. and dos Santos, J. A. (2015). Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 44-51.
69. Petroff (2020), A Lightweight Panorama Viewer for the Web, Pannellum.
<https://mpetroff.net/software/pannellum/>
70. Publications Québec. (2018). Loi assurant l'exercice des droits des personnes handicapées en vue de leur intégration scolaire, professionnelle et sociale. Chapitre E-20.1. <http://legisquebec.gouv.qc.ca/fr/ShowDoc/cs/E-20.1>
71. Razavian, A. S., Azizpour, H., Sullivan, J., and Carlsson, S. (2014). CNN features off-the-shelf: an astounding baseline for recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, p. 806-813.
72. Régie du bâtiment du Québec. (2010). Normes de conception sans obstacles. In Code de construction du Québec, chapitre 1 - Bâtiment, et Code national du bâtiment - Canada 2005. 66 p.
73. Rocque, S., Langevin, J., Chalghoumi, H., and Ghorayeb, A. (2011). Accessibilité universelle et designs contributifs dans un processus évolutif. *Revue développement humain, handicap et changement social*, vol. 19, n° 3, pp. 7-24.

74. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... and Berg, A. C. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, vol. 115 n°3, pp. 211-252.
75. Saint-Charles, P.-L. (2020). Thelper in Github, <https://github.com/plstcharles/thelper>
76. Sengupta, S., Sturges, P., and Torr, P. H. (2012). Automatic dense visual semantic mapping from street-level imagery. In *Intelligent Robots and Systems (IROS)*, 2012 IEEE/RSJ International Conference, pp. 857-862.
77. Services publics de Wallonie. (2006) Les manuels du MET, Guide de bonnes pratiques pour l'aménagement de cheminements piétons accessibles à tous, n°10, 90 p.
78. Shorten, C., and Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. In *Journal of Big Data*, vol. 6, n°1, 60 p.
79. Simard, P. Y., Steinkraus, D. and Platt, J. C. (2003). Best practices for convolutional neural networks applied to visual document analysis. In *Proceedings of the Seventh International Conference on Document Analysis and Recognition*, vol. 2, pp. 958–962.
80. Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. In *Proc. International Conference on Learning Representations*. arXiv preprint arXiv:1409.1556.
81. Some, L. (2016). Automatic image-based road crack detection methods.
82. Statistiques Canada. (2013). Enquête canadienne sur l'incapacité 2012. Catalogue no. 89-654-X, 86 p.
83. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. and Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9.
84. Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning?. In *IEEE transactions on medical imaging*, vol. 35, n° 5, pp. 1299-1312.
85. Tan, M., and Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*.
86. Tanikić, D., & Despotovic, V. (2012). Artificial intelligence techniques for modelling of temperature in the metal cutting process. In *Metallurgy-Advances in Materials and Processes*. 25 p.

87. Tavon, R. (2020). Google Street View panorama search and download scripts. Github, https://github.com/remtav/download_gsv_pano.
88. The Centre for Universal Design. (1999). Guidelines for Use of the Principles of Universal Design, North Carolina State University, 1 p.
89. Tsai, Y. C., Kaul, V., and Mersereau, R. M. (2010). Critical assessment of pavement distress segmentation methods. In *Journal of transportation engineering*, vol. 136, n°1, pp.11-19.
90. Van Etten, A., Lindenbaum, D., & Bacastow, T. M. (2018). Spacenet: A remote sensing dataset and challenge series. arXiv preprint arXiv:1807.01232.
91. Ville de Montréal. (2015). Le Plan d'action 2015-2018 en accessibilité universelle de la Ville de Montréal, 28 p.
92. Ville de Montréal (2017). Aménagements piétons universellement accessibles: Guide d'aménagement durable des rues de Montréal, Fascicule 5. Direction des transports. Montréal, 230 p.
93. Ville de Québec. (2010). Guide pratique d'accessibilité universelle: Trottoirs et liens piétonniers, p. 81-84.
94. Ville de Sherbrooke. (2016). Plan d'action 2016 pour l'intégration des personnes handicapées. 17 p.
95. Ville de Sherbrooke. (2018). Communication personnelle.
96. Ville de Toronto (2004). Accessibility Design Guidelines. Toronto, 137 p.
97. Wang, X., and Hu, Z. (2017). Grid-based pavement crack analysis using deep learning. In 2017 4th In *International Conference on Transportation Information and Safety (ICTIS)*, pp. 917-924.
98. Wolf, E., Cooper, R. A., Pearlman, J., Fitzgerald, S. G. and Kelleher, A. (2007). Longitudinal assessment of vibrations during manual and power wheelchair driving over select sidewalk surfaces. In *Journal of Rehabilitation Research & Development*, vol. 44, n°4, p. 573-580.
99. Wu, X., Sahoo, D., & Hoi, S. C. (2020). Recent advances in deep learning for object detection. In *Neurocomputing*. vol. 396, pp. 39-64.
100. Yosinski, J., Clune, J., Bengio, Y. and Lipson, H. (2014). How transferable are features in deep neural networks?. In *Advances in neural information processing systems*, p. 3320-3328.
101. Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision*, Springer, Cham, p. 818-833.

102. Zhang, L., Zhang, L. and Du, B. (2016). Deep learning for remote sensing data: A technical tutorial on the state of the art. In *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, n°2, p.22-40.
103. Zhang, L., Yang, F., Zhang, Y. D., and Zhu, Y. J. (2016). Road crack detection using deep convolutional neural network. In *2016 IEEE international conference on image processing (ICIP)*, pp. 3708-3712.
104. Zhang, A., Wang, K. C., and Ai, C. (2017). 3D shadow modeling for detection of descended patterns on 3D pavement surface. In *Journal of Computing in Civil Engineering*, vol. 31, n°4.
105. Zhu, X. X., Tuia, D., Mou, L., Xia, G. S., Zhang, L., Xu, F. and Fraundorfer, F. (2017). Deep learning in remote sensing: a comprehensive review and list of resources. In *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, n°4, 8-36.