Génération de cartes tactiles photoréalistes pour personnes déficientes visuelles par apprentissage profond

Gauthier Fillières-Riveau[1], **Jean-Marie Favreau**[1], **Vincent Barra**[1], **Guillaume Touya**[2]

1. *LIMOS UMR 6158, Université Clermont Auvergne, 1 rue Chebarde, TSA 60125 CS 60026 63178 Aubière cedex, France*

2. *LASTIG, Univ Gustave Eiffel, ENSG, IGN, 73 av. de Paris, 94165 Saint-Mandé cedex, France*

RÉSUMÉ. *Les cartes tactiles photoréalistes sont un des outils mobilisés par les personnes en situation de déficience visuelle pour appréhender leur environnement urbain proche, notamment dans le cadre de la mobilité, pour la traversée de carrefours par exemple. Ces cartes sont aujourd'hui principalement fabriquées artisanalement. Dans cet article, nous proposons une approche permettant de produire une segmentation sémantique d'une imagerie aérienne de précision, étape centrale de cette fabrication. Les différents éléments d'intérêt tels que trottoirs, passages piétons, ou îlots centraux sont ainsi localisés et tracés dans l'espace urbain. Nous présentons en particulier comment l'augmentation de cette imagerie par des données vectorielles issues d'OpenStreetMap permet d'obtenir par une technique d'apprentissage profond (réseau adverse génératif conditionnel) des résultats significatifs. Après avoir présenté les enjeux de ce travail et un état de l'art des techniques existantes, nous détaillons l'approche proposée, et nous étudions les résultats obtenus, en comparant en particulier les segmentations obtenues sans et avec enrichissement par données vectorielles. Les résultats sont très prometteurs.*

ABSTRACT. *Photo-realistic tactile maps are one of the tools used by visually impaired people to understand their immediate urban environment, particularly in the context of mobility, for crossing crossroads for example. These maps are nowadays mainly hand-made. In this article, we propose an approach to produce a semantic segmentation of precision aerial imagery, a central step in this manufacturing process. The different elements of interest such as sidewalks, pedestrian crossings, or central islands are thus located and traced in the urban space. We present in particular how the augmentation of this imagery by vector data from OpenStreetMap leads to significant results using a deep learning technique (conditional generative adversarial network). After presenting the stakes of this work and a state of the art of existing techniques, we detail the proposed approach, and we study the results obtained, in particular by comparing the segmentations obtained without and with enrichment by vector data. The results are very promising.*

MOTS-CLÉS : *apprentissage machine, enrichissement de données, information géographique volontaire, segmentation sémantique.*

KEYWORDS: *machine learning, data enrichment, VGI, semantic segmentation.*

DOI: 10.3166/rig.2020.00104

1. Introduction

La production de cartes à l'usage des déplacements urbains s'appuie sur une longue tradition, et bénéficie aujourd'hui des possibilités offertes par les bases de données géographiques, notamment lorsque les modes de consultation sont nomades. Elle offre aux usagers piétons de la ville un moyen efficace de se localiser et de réaliser des déplacements.

Parmi ces usagers, les *personnes en situation de déficience visuelle* (PSDV) sont régulièrement confrontées à un manque d'accessibilité des infrastructures urbaines. Si l'on observe une amélioration des aménagements urbains pour prendre en compte les besoins spécifiques de ces usagers (généralisation des bandes d'éveil de vigilance, des feux sonores, ou encore des abaissements de trottoir), la signalisation de ces aménagements et leur documentation ne sont que peu développées.

Ce sont les *instructeurs de locomotion* (IL) qui en France accompagnent les PSDV dans leur appropriation de l'espace urbain, pour les aider à développer leur mobilité et leur locomotion. Proposées à une fréquence régulière pendant toute leur instruction, les séances se déroulent généralement sur le terrain, pour apprendre à maîtriser les codes et savoir-faire d'un déplacement en sécurité dans la ville.

Cette transmission s'appuie en grande partie sur la capacité des PSDV à se représenter l'espace urbain. Ce savoir-faire est développé et consolidé par les IL qui s'appuient sur un échange permanent avec les apprenants, mais également sur la manipulation de représentations plus ou moins schématiques des structures urbaines, généralement sous forme de maquettes puis de plans et cartes en relief.

L'un des défis principaux d'une PSDV lorsqu'elle s'approprie les déplacements piétons urbains concerne la traversée des carrefours, l'une des zones de plus grand danger. Il est donc fréquent de travailler sur des supports construits à cette échelle, et ces supports ne peuvent pas être les mêmes que pour les personnes avec une vision normale (Hennig *et al.*, 2017).

1.1. Problématiques

La fabrication de ces cartes à l'usage d'un public en situation de déficience visuelle est essentiellement un processus manuel, et réalisé par des professionnels formés spécialement à la fabrication de documents pour ces publics, les adaptateurs-transcripteurs (AT). Leur travail consiste généralement à adapter un contenu initialement prévu pour des voyants, d'une part par la transcription des textes en braille, et d'autre part par adaptation des contenus graphiques pour la mise en relief. Ces processus de transcription sont longs, et nécessitent une compétence experte (Kern, 2016) souvent peu disponible.

À chaque type de document correspondent des conventions d'adaptation spécifiques. Ainsi, l'adaptation des cartes et plans est fortement liée à la pratique des IL, qui pourront avoir besoin de solliciter des informations géographiques à l'échelle du déplacement piéton : position précise des feux, tracé des passages piétons, localisation des portes d'entrées des bâtiments. Les pratiques des IL et des AT étant comparables à une pratique

artisanale, on observe également à travers le territoire français une grande variété de pratiques dans les modes de représentation adoptés. C'est le constat que nous avons notamment fait à l'occasion d'une collecte de cartes réalisée en juin 2019[1]. Cette grande variation est d'autant plus problématique que la découverte d'un document en relief par une PSDV n'est pas aussi intuitive que la découverte d'un document graphique. En particulier, chaque changement de représentation peut nécessiter une étape de découverte accompagnée.

Nous avons par exemple relevé 8 manières différentes de représenter le nord sur ces cartes (voir figure 1). La représentation des voies peut être soit décrite par un simple tracé (figure 2b, produite par les professionnels du CRDV[2]), soit par un tracé de chaque côté (figures 2a, produite par les professionnels du CTRDV[3]). Les trottoirs peuvent être très précisément décrits (figure 3a), ou au contraire juste évoqués (figure 2a). Certaines cartes utilisent plusieurs textures avec une sémantique associée (figure 3a), d'autres cherchent au contraire une représentation la plus schématique possible (figure 2a).

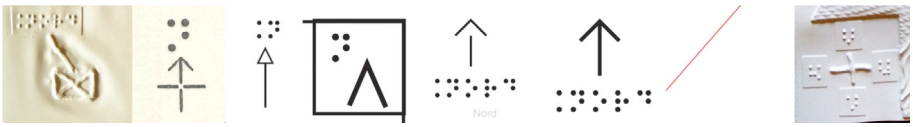

Figure 1. Exemples de représentations de l'orientation des cartes adaptées pour les déficients visuels

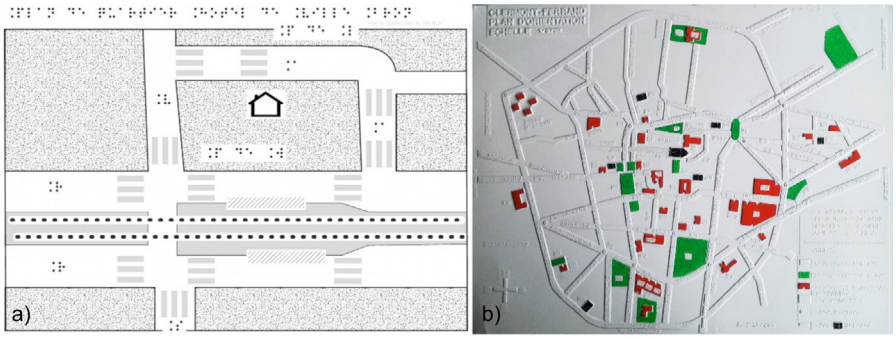

Figure 2. Deux cartes issues de la collecte. a) plan du quartier de l'hôtel de ville de Bron, dessin pour thermogonflage (©CRTDV). b) plan de Clermont-Ferrand thermoformé (©CRDV)

1. https://compas.limos.fr/collectes/ : 113 cartes collectées auprès d'adaptateurs transcripteurs et d'instructeurs en locomotion
2. http://www.crdv.org : Centre de rééducation pour déficients visuels (Clermont-Ferrand)
3. https://ctrdv.fr/ : Centre technique régional pour la déficience visuelle (Lyon)

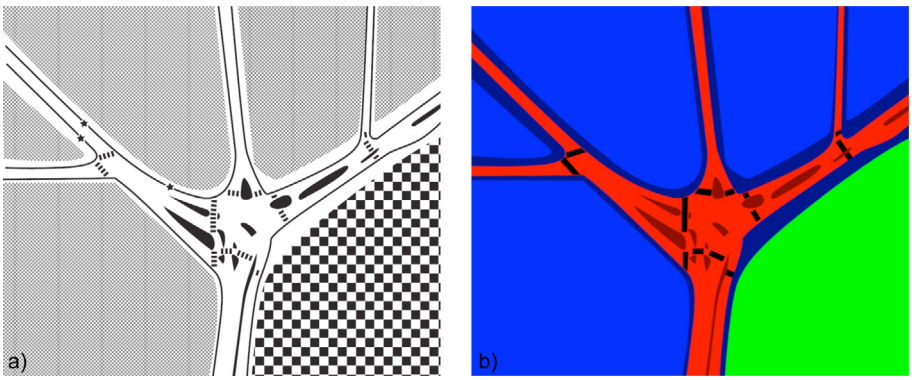

Figure 3. Carte adaptée, et sa segmentation correspondante. a) carte de carrefour adaptée par une professionnelle (©CRDV) pour un PSDV ; b) segmentation correspondante dans nos travaux

On retrouve cependant dans la grande majorité d'entre elles des éléments incontournables des aménagements urbains : routes, terre-pleins et îlots routiers, trottoirs, parcelles closes et/ou bâtiments, espaces verts, passages piétons, tracés des tramways et des voies ferrées, arrêts de bus.

Nous focalisons notre attention sur le sous-ensemble de cartes de notre collecte les plus significatives des usages des IL. Il s'agit de cartes que l'on pourrait qualifier de « photoréalistes », dans le sens où les éléments qui y figurent sont généralement représentés par leur empreinte réelle au sol. Construire une telle carte revient donc à réaliser une segmentation sémantique du territoire, où chaque élément est étiqueté sémantiquement. D'après les entretiens que nous avons eu avec plusieurs AT ayant réalisé ces cartes, on sait que leur fabrication s'appuie généralement sur l'utilisation d'orthophotographies, et sur une exploration minutieuse du terrain. Il s'agit d'un processus fastidieux, peu réalisé par manque de temps et de moyens, malgré les besoins exprimés par les IL.

La production automatique de cette segmentation est de notre expérience l'une des briques indispensables d'une suite logicielle qui serait destinée à faciliter la conception de cartes pour la mobilité et la locomotion des PSDV, augmentant ainsi les outils à disposition des IL.

Dans la littérature, on constate que la production automatique de cartes s'appuie essentiellement sur des bases de données géographiques telles qu'OpenStreetMap (OSM), une ressource pertinente à la fois pour la production de cartes classiques (Girres et Touya, 2010 ; Touya *et al.*, 2017), mais également dans le cadre de classification par apprentissage machine (Chen et Zipf, 2017 ; Kurath *et al.*, 2017). OpenStreetMap est pertinent car la base couvre le monde entier et elle est localement plus détaillée que les bases de données des agences nationales de cartographie comme l'IGN en France.

Cependant, le contenu de ces bases de données est limité pour la génération de cartes pour PSDV, à l'échelle de ce que les contributeurs d'OSM appellent *micromapping*[4]. En effet, les trottoirs et îlots centraux ne sont quasiment pas présents dans les bases de données, et un grand nombre d'informations sont décrites de manière ponctuelle (passages piétons) ou linéaire (routes), leur emprise réelle n'étant pas renseignée. Si d'autres bases de données contiennent certaines de ces informations, elles ne sont pas satisfaisantes pour notre usage, car elles ne sont généralement pas disponibles de manière ouverte[5], ou ne couvrent qu'un petit territoire[6].

En s'inspirant du processus de fabrication de ces cartes par les professionnels de la déficience visuelle, nous proposons dans cet article une chaîne de traitement s'appuyant à la fois sur les orthophotographies des carrefours représentés, et sur leur enrichissement par une connaissance experte du terrain, via les données d'OSM.

1.2. Approches existantes

La perception de la carte par les PSDV est dépendante de nombreux éléments, dont on trouve la trace dans la littérature dédiée à la génération automatique de cartes pour ce public (Wabinski et Moscicka, 2019).

Certains travaux se concentrent sur la sélection et la mise en relief de la carte, avec une représentation des données telles que présentes dans la base de référence (Cervenka *et al.*, 2016) ou après généralisation (Stampach et Mulíčková, 2016 ; Touya *et al.*, 2019). D'autres se focalisent sur l'augmentation de la représentation par l'ajout d'informations sonores (Josselin *et al.*, 2016 ; Boularouk *et al.*, 2017). Certains travaux combinent ces deux approches, mais la carte est représentée avec un style symbolique des données (Watanabe *et al.*, 2014). Dans le cas d'utilisation que nous envisageons, la représentation se doit d'être fidèle aux réalités du terrain, malgré les informations d'emprise incomplètes ou non présentes dans la base de données de référence. D'autre part, puisque le premier usage identifié de ces cartes est celui des IL qui accompagnent les PSDV sur le terrain, l'enrichissement par information sonore n'est donc pas nécessaire dans un premier temps.

Un des moyens pour enrichir les données d'OpenStreetMap pour obtenir l'information source nécessaire à la constitution de ces cartes serait d'utiliser des techniques d'enrichissement de données spatiales vectorielles. Des travaux ont montré que l'on pouvait reconstruire des structures complexes du réseau routier simplement en utilisant les polylignes représentant le centre des voies routières (Touya, 2010 ; Savino

4. Cartographie d'éléments à petite échelle, comme les portes de bâtiments, les cheminements piéton, etc. Voir https://wiki.openstreetmap.org/wiki/Micromapping
5. Bases de données privées, notamment utilisées par les acteurs travaillant à la voiture intelligente, avec les cartes HD où chaque voie de circulation est décrite par une trace géolocalisée, augmentée d'informations complémentaires relevant de la signalisation et de l'usage de la voie. Voir par exemple https://www.tomtom.com/products/hd-map/
6. Dans la dynamique de publication en OpenData, voir par exemple https://opendata.paris.fr/

et al., 2010) : on peut ainsi retrouver les ronds-points, les pattes d'oies, les voies séparées par un terre-plein central ou classer les carrefours en fonction de leur forme. Pour reconstruire l'emprise de la chaussée et des trottoirs, les travaux sur la génération procédurale de modèles très détaillés de routes (Cura *et al.*, 2015) pourraient être réutilisés pour enrichir les routes linéaires présentes dans OpenStreetMap. Sur le même principe, des travaux ont proposé la génération procédurale des feux et panneaux de signalisation à partir des routes d'OpenStreetMap (Taal et Bidarra, 2016).

L'approche que nous présentons dans cet article est plutôt basée sur une technique d'apprentissage profond[7], aujourd'hui considérée comme incontournable. En effet, l'arrivée progressive des réseaux de neurones à convolution (CNN) dans de nombreux domaines applicatifs – notamment en imagerie (Guo *et al.*, 2016) – a rendu caduque un grand nombre de méthodes classiques de segmentation issues des mathématiques et heuristiques (Shotton *et al.*, 2008). Le fait que les approches par réseaux de neurones permettent d'intégrer notamment dans le processus d'entraînement le savoir-faire expert a conforté notre exploration de ces techniques.

Le mécanisme des couches de mise en commun (ou *pooling*) des CNN n'est cependant pas complètement adapté à la segmentation sémantique, car il entraîne une perte d'information de localisation et de contexte initial. Le paradigme des *Fully Connected Networks* (FCN) (Long *et al.*, 2014) est une réponse à ce constat, car leur structure permet de prendre en compte à la fois des informations globales et locales.

On trouve dans la littérature deux approches basées sur ce paradigme pour répondre aux problèmes de segmentation : les réseaux de type *encodeur-décodeur* (Ronneberger *et al.*, 2015) adaptés de réseaux utilisés en classification et ceux utilisant des *convolutions dilatées* (Yu et Koltun, 2016 ; Chen *et al.*, 2018), basées sur une conception dédiée aux problèmes de segmentation. Ces approches font référence dans le domaine de la segmentation sémantique. Cependant, nous avons constaté que ces approches n'étaient pas pertinentes sur les données que nous utilisons. En effet, l'approche U-Net (Ronneberger *et al.*, 2015) que nous avons mise en place lors de nos expérimentations préalables a montré un très mauvais score de classification, notamment des passages piétons et des terres-pleins, soit très mal détectés (< 15 %) soit par faux positifs abondants (> 70 %). Ce constat n'est pas surprenant, car l'approche U-Net considère que les données d'entraînement sont parfaitement segmentées. Or, les cartes que nous avons collectées comportent des biais et incertitudes liés à leur fabrication. Certaines zones des images sont ainsi étiquetées avec la mauvaise sémantique, ce qui biaise l'apprentissage d'un réseau comme U-Net, dégradant d'autant ses performances sur les segmentations attendues. De plus, les modèles de type U-Net nécessitent généralement un très grand nombre d'exemples d'apprentissage, ce qui n'est pas possible pour nous car notre collecte de carte a été limitée.

---

7. Ensemble de méthodes d'apprentissage automatique utilisant des architectures complexes de transformations non linéaires, notamment les réseaux de neurones, pour modéliser des données avec un haut niveau d'abstraction.

À l'inverse, on obtient des résultats plus intéressants pour les mêmes jeux de données en utilisant un réseau de la famille des *Generative Adversarial Networks* (GAN) (Goodfellow *et al*., 2014), où deux réseaux (un générateur et un discriminateur) sont entraînés à tour de rôle. En particulier, les GAN conditionnels (ou cGAN) (Mirza et Osindero, 2014) permettent d'introduire dans le mécanisme d'apprentissage des conditions issues des jeux de données, et donc de mieux contrôler que l'image générée par le générateur ressemble à ce qu'on attend. Ces approches ont notamment été utilisées pour générer des images complexes ou transférer le style d'une image vers une autre, par exemple avec des applications dans lesquelles une carte est générée par le réseau (Isola *et al*., 2016 ; Zhu *et al*., 2017 ; Kang *et al*., 2019).

Nous proposons dans la partie 2 suivante une approche s'appuyant sur un cGAN, où les données orthophotographiques initiales sont consolidées par association de données OSM. Cet enrichissement sémantique permet d'entraîner un réseau de neurones en s'appuyant sur le jeu de données collectées. Le résultat de cet apprentissage est alors un réseau de neurones capable de produire une segmentation sémantique automatique de carrefours pour la production de cartes en relief.

Nous présentons dans la partie 3 une description des expérimentations réalisées, tout d'abord avec les détails de la préparation des données et de l'implémentation de la chaîne de traitement, puis avec une discussion des résultats obtenus sur les carrefours de la ville de Clermont-Ferrand. Nous détaillons dans la partie 4 les étapes nécessaires à l'intégration de ces résultats de segmentation pour la production d'une carte en relief. Enfin, la partie 5 propose une synthèse et ouvre différentes perspectives de recherche.

2. Méthode

Après une introduction du réseau utilisé, nous exposerons le type de données servant à l'entraînement du réseau, pour enfin présenter les moyens d'évaluation permettant d'apprécier les résultats de notre approche.

2.1. GAN conditionnel

L'approche que nous proposons s'appuie sur un cGAN appelé Pix2Pix (Isola *et al*., 2016) dont le principe est d'apprendre à modéliser la mise en correspondance entre une image d'entrée et une image de sortie. Dans cette approche, le réseau comporte deux composantes principales : le *générateur* et le *discriminateur*.

Le générateur utilisé est un encodeur-décodeur U-Net générant une image ayant les mêmes dimensions qu'une image d'entrée, après lui avoir appliqué une succession de modifications. Le rôle du discriminateur est de déterminer si l'image de sortie a été produite par le générateur ou si c'est une « vraie » image, en utilisant un discriminateur markovien spécifique (patchGAN).

L'entraînement du réseau s'effectue en deux étapes, la première pour entraîner le discriminateur, la seconde pour entraîner le générateur (figure 4).

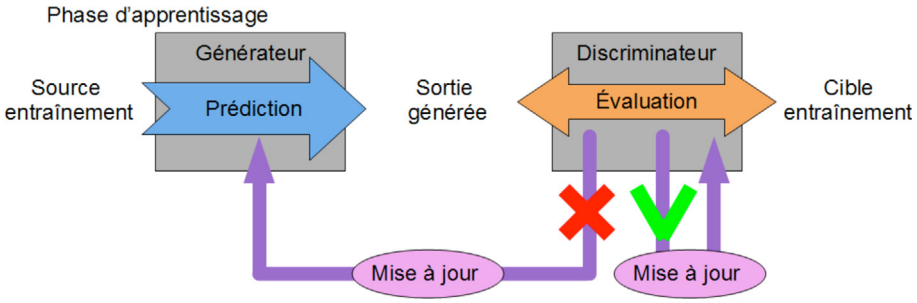

Figure 4. Schéma simplifié du processus d'entraînement du GAN conditionnel

Le principe général consiste à améliorer d'abord le discriminateur. On poursuit ensuite par l'amélioration du générateur, qui cherche à tromper le discriminateur en faisant passer l'image qu'il génère pour une image cible. S'il réussit alors le discriminateur est ajusté, sinon ce sont les poids du générateur qui sont modifiés. Le réseau s'entraîne en itérant sur le jeu de données.

L'utilisation la plus répandue de GAN est la génération d'image, notamment à partir de descriptions textuelles (Zhang *et al.*, 2016) ou de segmentation (Isola *et al.*, 2016 ; Wang *et al.*, 2017).

2.1.1. Données sources

Les données d'entrée du réseau sont les images à segmenter. L'originalité de notre approche consiste à préparer ces données d'entrée en combinant deux jeux de données : des images issues de prise de vue aériennes (*orthophotographies* ou *orthophotos*), et des données géographiques vectorielles extraites d'OSM.

Les orthophotos comportent plusieurs spécificités intéressantes qui facilitent leur utilisation en segmentation par apprentissage machine (Kaiser *et al.*, 2017) : elles sont géométriquement rectifiées pour prendre en compte les déformations du système d'acquisition et recalées dans un système de coordonnées exploitable par un système d'information géographique (SIG). Pour homogénéiser leur traitement, elles sont également radiométriquement corrigées (luminosité, saturation et contraste des couleurs unifiées).

La base de données géographique volontaire OSM complète avantageusement ces premières données, car elle propose une grande quantité d'informations vectorielles et sémantiques, d'une pertinence et précision significatives (Haklay, 2010 ; Zielstra et Zipf, 2010). Distribué sous licence libre, OSM est ouverte à la contribution de tout volontaire. Sa représentation par géométrie vectorielle augmentées d'un ensemble d'étiquettes (association clé/valeur) et de relations entre ces objets géométriques permet un enrichissement constant, notamment par le raffinement régulier des conventions de description adoptées par la communauté des contributeurs de manière collégiale.

Nous proposons dans ce travail d'ajouter des données issues d'OSM par superposition aux orthophotos, en les traçant avec une couleur définie suivant leur classe et suffisamment saturées pour être distinctes des couleurs naturelles des orthophotos. Nous avons ainsi choisi pour les routes un rouge saturé (RGB 255 ; 0 ; 0) complètement absent des données orthophotos que nous avons utilisées. Les tracés vectoriels sont réalisés avec une largeur minimale (un pixel) dans le but de minimiser l'impact sur les données orthophotos : nous ne souhaitons pas, par exemple, que les routes cachent les passages piétons.

Ces images sont ensuite découpées en *vignettes* carrées, en choisissant une définition à la fois assez grande pour y représenter de manière identifiable les éléments urbains d'intérêt (passages piétons, îlots centraux, etc), mais également assez petites pour être prises en charge par les réseaux de neurones (typiquement 256 pixels de côté).

2.1.2. Données cibles

Les données cibles du réseau de neurones sont les segmentations de référence qui serviront aux phases d'entraînement et de tests du réseau. Parmi les 113 cartes de la collecte que nous avons réalisée, nous avons sélectionné 15 cartes satisfaisant aux besoins de notre contexte applicatif. Ces cartes sont celles dessinées à l'échelle du carrefour, et transmises dans un format numérique exploitable. Si de prime abord, ce nombre peut sembler faible pour entraîner un réseau de neurones, la précision de couverture orthophotographique (voir section 3.1) et le format vectoriel permettent l'export avec une grande résolution. Ainsi, la subdivision en carrés de 256 pixels de côté permet d'obtenir de nombreux échantillons d'entraînement. Nos données d'entrée et cibles ayant les mêmes définitions, nous avons également consolidé les données par les techniques d'augmentation, notamment la rotation des vignettes (Taylor et Nitschke, 2017).

Afin de préparer les cartes collectées au format attendu par le cGAN, nous avons transposé les motifs et régions présentes dans les cartes en noir et blanc (figure 3a) par un code couleur définissant les catégories des éléments représentés (figure 3b). De plus, les cartes étaient non géoréférencées et préparées pour une impression au format A4. Nous avons dû les recaler sur les données géographiques. Ce recalage a été fait de manière semi-automatique dans un SIG. Le découpage en images carrées se fait alors sur le modèle de celui des données sources.

2.2. Définition de l'erreur et de sa mesure

L'évaluation d'un réseau de neurones consiste à comparer le résultat produit par le modèle avec l'image de référence attendue, et ce pour toutes les paires d'images d'un jeu de données qui aura été réservé à cet effet, et non utilisé pour l'entraînement.

Afin de mesurer la qualité de la segmentation obtenue, on évalue l'erreur de classification de chaque pixel en le classant suivant son appartenance à un ensemble de classes prédéfinies. Cette classification est observée à la fois dans l'image résultat et

dans l'image de référence, l'erreur correspondant alors à une différence entre la classe détectée et la classe réelle.

Les ontologies géographiques et les liens sémantiques sont des approches classiques en cartographie (Couclelis, 2010 ; Dobesova et Brus, 2011), sur lesquelles nous nous appuyons pour classer les erreurs de segmentation.

On définit l'intensité de l'erreur sémantique suivant le nombre de liaisons sémantiques séparant la catégorie déterminée par la segmentation de celle correspondant à la classe issue de l'image de référence. Dans l'exemple de la figure 5, l'erreur de classification de la classe passage piéton avec celle de la route est moindre qu'avec l'élément terre-plein ou trottoir, et elle est maximale pour l'association avec la classe parcelle.

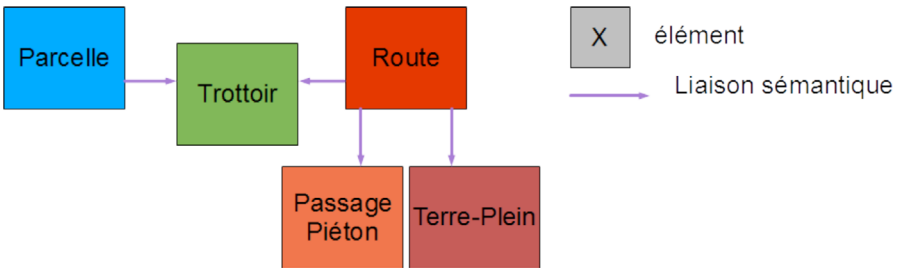

Figure 5. Représentation non exhaustive de liens sémantiques entre des éléments de notre segmentation

L'utilisation des cartes produites s'articulant autour de la navigation piétonne, nous avons choisi de minorer les erreurs d'éléments ne perturbant pas cette navigation, et au contraire de majorer celles entravant ces déplacements. Nous proposons donc une classification d'erreur en trois niveaux :

– **erreur faible :** il existe une proximité sémantique forte entre les catégories et la navigation n'est pas ou peu influencée,

– **erreur modérée :** il existe une proximité sémantique forte entre les catégories et la navigation est influencée, ou il n'existe pas de proximité sémantique forte entre les catégories mais il n'y aura aucune influence sur la navigation,

– **erreur forte :** il n'existe pas de proximité sémantique entre les catégories et la navigation est perturbée.

3. Implémentation et résultats

L'implémentation de la méthode a été réalisée suivant le processus décrit par le schéma (figure 6) : préparation des données, puis apprentissage suivi du test, et enfin mesures d'erreur du modèle afin de pouvoir analyser nos résultats.

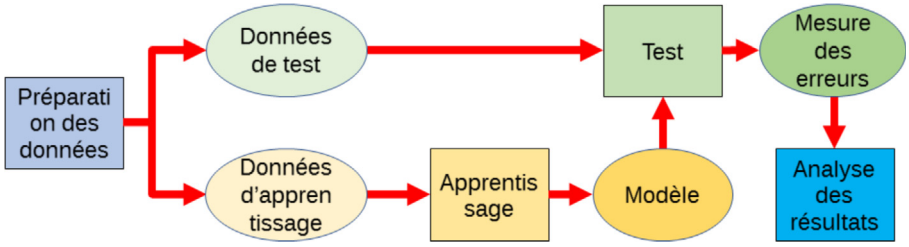

Figure 6. Schéma global du processus implémenté

Le cGAN utilisé est basé sur l'implémentation TensorFlow (bibliothèque Python pour l'apprentissage profond) du modèle Pix2Pix proposé par (Isola *et al.*, 2016)[8], et la phase d'entraînement a été réalisée sur un serveur de calcul équipé d'un processeur Intel® Xeon® CPU E5-2640 v4 @ 2.40 GHz et d'une carte NVIDIA® GP104GL.

La préparation des données a permis d'obtenir deux jeux de 375 paires d'images sources/cibles. Le premier jeu de données comporte uniquement les orthophotos simples, alors que le second comprend les orthophotos enrichies des données OSM, afin d'étudier l'apport de l'enrichissement sémantique. Dans les deux cas, nous avons utilisé 300 paires pour l4 entraînement et les 75 restantes pour l'évaluation des modèles, par une répartition aléatoire.

L'évaluation de l'erreur a été réalisée en suivant la méthode présentée section 2.2, pour produire la pondération illustrée par le tableau 1. Dans ce tableau, nous avons proposé une classification en trois niveaux (faible, modéré et fort), que nous avons raffiné en 5 niveaux (de 1 à 5), afin de proposer une visualisation plus fine de sa spatialisation (voir figure 8).

3.1. Données utilisées

Nous avons utilisé une orthophoto haute résolution (ORTHO HR®) coproduite par l'Institut national de l'information géographique et forestière (IGN) et le Centre régional Auvergne-Rhône-Alpes de l'information géographique (CRAIG) pour le Puy-de-Dôme. Elle est disponible sous licence ouverte[9] et propose une résolution de 20 cm à 25 cm par pixel, ce qui permet de décrire au mieux l'échelle du carrefour qui nous intéresse, même à l'intérieur des vignettes de 256 pixels de côté (figure 7).

8. https://github.com/affinelayer/pix2pix-tensorflow
9. http://professionnels.ign.fr/orthohr-par-departements\#tab-3

Tableau 1. Classification de l'importance des erreurs en fonction de l'association des classes

Classe	**route**	**passage piéton**	**terre-plein**	**trottoir**	**parcelle/bâti**	**espace vert**
route	-	faible (1)	modérée (3)	faible (2)	haute (5)	haute (5)
passage piéton	faible (1)	-	faible (2)	faible (2)	haute (5)	haute (5)
terre-plein	modérée (3)	faible (2)	-	haute (4)	haute (5)	haute (5)
trottoir	faible (2)	faible (2)	haute (4)	-	modérée (3)	modérée (3)
parcelle/bâti	haute (5)	haute (5)	haute (5)	modérée (3)	-	modérée (3)
espace vert	haute (5)	haute (5)	haute (5)	modérée (3)	modérée (3)	-

Figure 7. Vignettes de 256 pixels de côtés extraites de l'ortho HR utilisée pour l'apprentissage

Les données OSM de la région Auvergne ont été téléchargées en 2018 depuis le serveur Geofabrik[10]. La préparation des données a été réalisée avec le logiciel libre de conception et d'analyse cartographique QGIS[11].

Un plugin QGIS écrit en Python nous permet d'isoler chaque classe de données géographiques en une couche vectorielle correspondante, par filtrage des différentes associations de clé/valeur dans les étiquettes OSM. Un style prédéfini est ensuite appliqué pour chaque classe. Ainsi, en agrégeant la couche regroupant le tracé des routes à la couche composée de l'orthophoto, nous obtenons l'enrichissement décrit dans la section 2.2.1. Les zones correspondant à nos cartes sont alors exportées avec et sans enrichissement pour générer le jeu de données d'entrée (voir figure 8).

Figure 8. La même vignette sans et avec augmentation par données OSM

Les cartes collectées ont été préparées en deux étapes. Tout d'abord, elles ont été coloriées grâce à un logiciel de dessin vectoriel[12] afin de préparer les cartes texturées au format attendu des cartes de segmentation (figure 3). Nous avons ensuite effectué une

10. http://download.geofabrik.de/europe/france/auvergne.html
11. https://www.qgis.org/fr/site/, version 2.18
12. Inkscape : https://inkscape.org

mise en correspondance manuelle des données réelles et de la segmentation, les déformations et simplifications effectuées par les professionnels lors de l'adaptation rendant difficile l'utilisation des outils de géoréférencement de QGIS.

On note cependant que l'emprise grossière et les simplifications de géométries de certains éléments, contenues dans certaines des cartes collectées, apportent des imperfections et débordements pour certaines classes (voir figure 9).

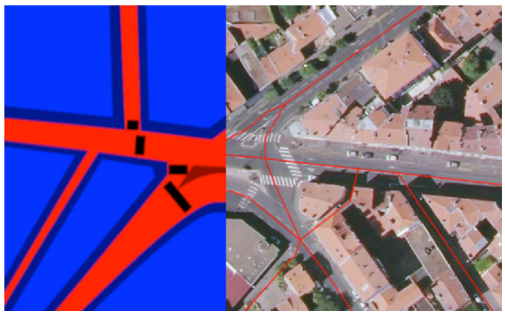

Figure 9. Juxtaposition d'une segmentation issue d'une carte collectée et d'une ortophoto augmentée de données OSM, illustrant la schématisation introduite par le tracé artisanal sur la carte collectée

Enfin, nous avons utilisé les paramètres du modèle Pix2Pix proposés par défaut, à l'exception du nombre de cycle complet (*epoch*) pour lequel nous avons testé différentes valeurs et qui donne de bons résultats à partir de 200 epochs avec le jeu de données enrichies utilisé. Les résultats sont issus d'un entraînement stoppé à 300 epochs. Les temps d'entraînement sont très longs, car il faut environ deux semaines pour atteindre 300 epochs.

La figure 10 présente le résultat de la segmentation obtenue sans et avec utilisation des données OSM sur un jeu de test non utilisé pour l'apprentissage.

3.2. Analyse des résultats

La comparaison des résultats obtenus sans et avec utilisation des données OSM sur notre jeu de données de test permet de visualiser les améliorations apportées par l'enrichissement. Les erreurs sont relevées et classées en comparant l'écart de sémantique entre l'image de référence et celle générée par le modèle, tel que présenté section 2.2.

Les résultats présentés sur le tableau 2 confirment la nette amélioration apportée par l'enrichissement. Le relevé des erreurs constatées met en évidence une amélioration de la cohérence de la segmentation.

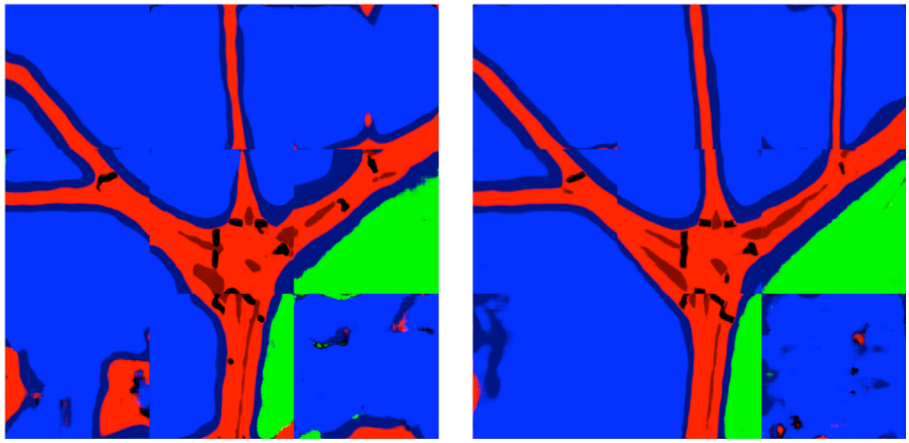

Figure 10. Résultat obtenus par les deux réseaux entraînés sur les données de test (non utilisées pour l'apprentissage) composées de 9 tuiles adjacentes. À gauche, résultat obtenu par le réseau entraîné sans augmentation, à droite par le réseau avec augmentation par données OSM

Tableau 2. Résultats sur le jeu de test : pour les données sans et avec enrichissement, pourcentage de pixels dans chacune des classes.

Mesure	**Modèle sans OSM**	**Modèle avec OSM**
Bonne classification	70,76 %	81,9 %
Erreur faible	13,73 %	10,54 %
Erreur modérée	7,06 %	5,36 %
Erreur forte	8,45 %	2,21 %

Nous remarquons en particulier que le gain de précision et de cohésion provient principalement de la diminution du nombre d'erreurs de forte importance. C'est ici majoritairement l'association route et parcelles/bâtis qui est en cause : sans enrichissement, certains toits de bâtiments et cours privatives sont identifiés comme route, ce qui ne se produit plus en utilisant les données OSM. Nous constatons également que la segmentation des routes avec enrichissement ne souffre plus des problèmes d'occlusion (par exemple par des arbres) et de détection dans les zones recouvertes d'ombre. En observant la répartition des erreurs (figure 11) on constate qu'une grande partie des incohérences dans la segmentation sont minimisées avec l'enrichissement et que les erreurs restantes correspondent à une imperfection de contours, notamment dans les zones proches des bords des vignettes.

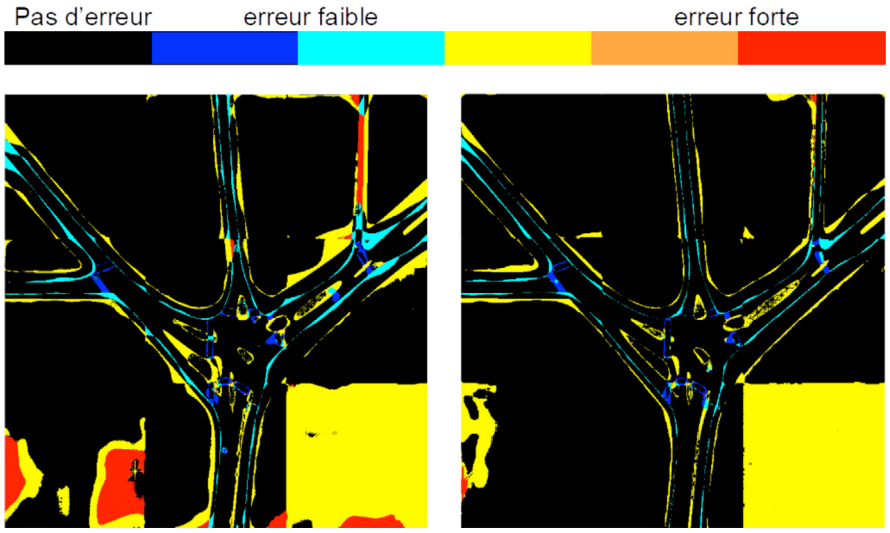

Figure 11. Carte de la localisation des erreurs suivant leur importance, sans enrichissement à gauche et avec enrichissement à droite, sur un ensemble de 3 × 3 vignettes correspondant au découpage d'entraînement. Les cinq couleurs d'erreur correspondent aux niveaux encodés de 1 à 5 dans le tableau 1

Une grande partie des erreurs de forte importance toujours présentes malgré l'utilisation des données OSM résulte également des biais présents dans nos cartes de référence. En effet, les professionnels concentrent en priorité l'information portée par la carte aux abords directs des intersections, en délaissant et en simplifiant les informations mineures en périphérie de carte, par exemple au niveau de rues peu fréquentées (figure 12). Notre segmentation incluant les données OSM travaillant sans faire la distinction entre le centre et la périphérie de la carte, on obtient donc des segmentations très différentes dans ces zones.

Figure 12. Biais sur la classification des routes entre la carte adaptée manuellement qui ne contient pas toutes les routes (gauche), l'orthophoto enrichie (centre) et le résultat généré par le réseau entraîné avec les données OSM (droite)

3.3. Étude d'un contexte de données manquantes

Afin de mesurer l'importance de l'accroche aux données OSM dans l'apprentissage, nous avons également testé le modèle appris sur des images test incluant des erreurs dans les données enrichies, avec l'absence d'un tronçon de route (figure 13), voire même l'absence complète de tracé avec l'utilisation de données non enrichies, ou encore le mauvais placement du tracé. À l'exception du cas de l'absence partielle de tracé de route, les performances chutent en deçà de celles sans enrichissement, ce qui montre l'influence des données OSM enrichies sur l'apprentissage par le réseau.

Cela met également en évidence le fait que la méthode tire parti des informations liées à la sémantique, telles que la liaison passage piéton/route (quart inférieur gauche figure 13). En effet, un passage piéton bien détecté en présence du tracé de la route ne l'est plus en retirant ce tracé.

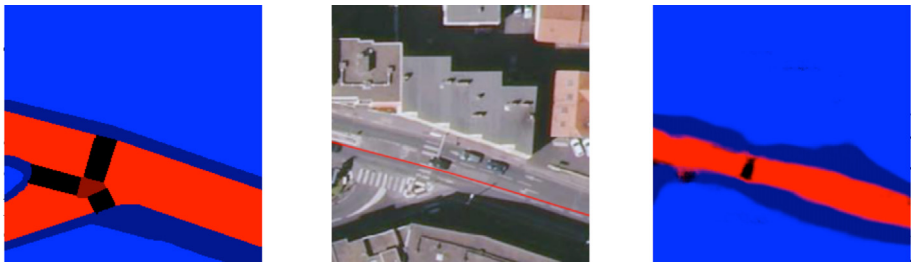

Figure 13. Erreur issue d'erreur de données OSM sur la classification. Référence (gauche), orthophoto avec données OSM incomplètes (centre), résultat du réseau entraîné avec enrichissement (droite)

4. De la segmentation à la carte tactile

Nous avons vu que les résultats de segmentation pouvaient comporter des erreurs de classification au bord et que pour générer une carte tactile autour d'un carrefour qu'un IL souhaite faire découvrir à une PSDV, nous avons souvent besoin d'utiliser plusieurs vignettes, du fait de la taille limitée de nos vignettes. Quand nous cherchons à assembler ces vignettes, des incohérences peuvent donc apparaître, avec des pixels qui deviennent voisins mais non classés de manière similaire (voir par exemple la figure 11). Pour corriger ces problèmes, nous proposons d'harmoniser les pixels au bord, en utilisant les valeurs de segmentation dans les deux images car il est très rare dans nos observations que les erreurs au bord arrivent dans les deux images voisines. Ainsi, pour ces pixels au bord, nous utilisons une fenêtre glissante centrée sur le bord qui regarde les valeurs autour du bord pour déterminer la classe corrigée des pixels de bord (figure 14).

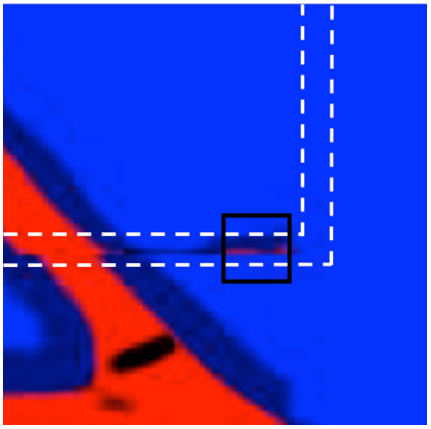

Figure 14. Méthode de correction des erreurs au bord lors de la fusion des vignettes : pour l'ensemble des pixels au bord (dans la zone entre les pointillés blancs), on utilise la valeur la plus fréquente dans une fenêtre glissante (carré noir)

Un autre des défauts constatés dans les résultats de segmentation est la nature granuleuse, non lisse, des contours des régions segmentées. Or, ces contours doivent être bien nets, lisses, dans les cartes tactiles pour être bien perçus par le toucher. Il reste donc une étape de simplification, lissage, schématisation de ces contours, que nous n'avons pas traité dans ce travail. Toutefois, comme expliqué dans nos travaux précédents (Touya *et al.*, 2019), nous pensons que les techniques de simplification utilisées en généralisation cartographique peuvent nous permettre d'obtenir de manière automatique des contours de régions qui répondent aux exigences d'une carte tactile pour PSDV, proches de ceux que produisent les AT. Pour s'en persuader, nous nous appuyons sur les résultats obtenus pour simplifier le contour de bâtiments segmentés sur des orthophotos qui présentaient le même type d'irrégularité que nos régions (Lokhat et Touya, 2016).

Pour finir, une fois que nous disposons de régions aux contours bien définis, nous pouvons mettre ces contours à disposition d'un AT qui va pouvoir finaliser la carte et l'habiller (par exemple en ajoutant du texte en braille). Une fois ce travail de finalisation terminé, il faut transformer cette carte vectorielle en deux dimensions, souvent au format SVG du fait des logiciels de dessin utilisés par les AT, en un modèle 3D prêt à être imprimé en trois dimensions. Pour faciliter cette dernière étape, nous avons développé un outil en licence libre, appelé SVG-to-STL[13], qui permet de transformer une carte au format SVG en modèle au format STL, compatible avec de nombreuses imprimantes 3D.

13. https://github.com/ACTIVmap/svg-to-st

5. Conclusion et perspectives

Dans cet article, nous avons présenté la problématique de la conception de cartes en relief photoréalistes pour déficients visuels à l'échelle du carrefour. Nous avons identifié les besoins et pratiques de ces usages, puis présenté une approche utilisant des données sémantiques et géométriques issues d'OSM pour enrichir une représentation de carrefours par orthophotos permettant de produire une segmentation sémantique. Les résultats de cette étape indispensable à la génération de cartes en relief à cette échelle ont été discutés, à la fois d'un point de vue de la qualité des résultats obtenus, mais également en envisageant leur intégration dans une chaîne permettant de produire les cartes définitives. Ces résultats prometteurs constituent une avancée dans la construction d'un processus semi-automatique de production de cartes en relief, notamment pour un usage en orientation et mobilité.

L'enrichissement présenté dans cet article n'exploite que le tracé des voies automobiles. L'utilisation d'autres éléments OSM tels que le contour parcellaire ou la localisation des passages piétons sera l'étape naturelle de l'extension de ce travail, pour améliorer le guidage sémantique de la segmentation. On peut aussi imaginer l'ajout des emplacements des ronds-points ou séparateurs de chaussée qui peuvent être détectés automatiquement à partir des routes (Touya, 2010 ; Savino *et al.*, 2010).

L'une des problématiques qui devra être explorée concerne les typologies de villes. Dans ce premier travail, nous avons utilisé des données d'entraînement et de test issues d'une même ville, au tissu urbain propre, relativement éloigné de celui que l'on pourrait rencontrer dans une ville de grande taille, comme Lyon ou Paris. Est-il possible d'entraîner un unique réseau pour toutes ces configurations ou devra-t-on imaginer une approche intégrant une prise en compte explicite de cet aspect ? C'est une question qui reste aujourd'hui ouverte, et que l'on pourra explorer à condition de disposer de jeux de données complémentaires à celui présenté dans la section 3.1.

La difficulté de cette exploration réside principalement dans le manque de données d'entraînement disponibles. L'introduction de cartes réalisées par d'autres professionnels que ceux ayant dessinés le jeu de données présenté dans la section 3.1 pourrait entraîner une difficulté du réseau à converger, si l'interprétation du territoire était trop différente. En effet, notre collecte de cartes montre une très grande diversité dans les choix de *design* cartographiques faits par les professionnels de la transcription pour PSDV. Il faudrait alors envisager des approches alternatives, plus robustes à ces variations.

Une autre solution consisterait à apprendre avec d'autres sources de données comme les modèles numériques de surfaces (MNS), qui ne seraient pas utilisées par la suite lors de la phase de segmentation avec le modèle appris. En utilisant un réseau secondaire pour apprendre à générer ces sources annexes, ce second modèle pourrait alors être utilisé comme une forme d'extracteur d'informations complémentaires sur les données d'entrée (Piasco *et al.*, 2018).

Bibliographie

Boularouk S., Josselin D., Altman E. (2017). Ontology for a Voice Transcription of Open Street Map Data: The Case of Space Apprehension by Visually Impaired Persons. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 11, n° 5, p. 585-589.

Cervenka P., Brinda K., Hanousková M., Hofman P., Seifert R. (2016). Blind Friendly Maps. In K. Miesenberger, C. Bühler, P. Penaz (Eds.), *Computers Helping People with Special Needs*, vol. 9759, p. 131-138. Springer International Publishing.

Chen J., Zipf A. (2017). DeepVGI: Deep learning with volunteered geographic information.In *Proceedings of the 26th International Conference on World Wide Web Companion*.

Chen L., Papandreou G., Kokkinos I., Murphy K., Yuille A.L. (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, n° 4, p. 834-848.

Couclelis H. (2010). Ontologies of geographic information. *International Journal of Geographical Information Science*, vol. 24, p. 1785-1809.

Cura, R., Perret J., Paparoditis N. (2015). StreetGen: In-Base Procedural-Based Road Generation. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, II-3- W5: 409-16. Copernicus GmbH, 2015. https://doi.org/10.5194/isprsannals-II-3-W5-409-2015.

Dobesova Z., Brus J. (2011). Coping with cartographical ontology. *Proceedings of SGEM*, p. 377-384.

Girres J.-F., Touya G. (2010). Quality assessment of the french openstreetmap dataset. *Transactions in GIS*, vol. 14, n° 4, p. 435-460.

Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S.*et al.* (2014). Generative Adversarial Networks. arXiv e-prints, p. arXiv:1406.2661.

Guo Y., Liu Y., Oerlemans A., Lao S., Wu S., Lew M.S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, vol. 187, p. 27-48.

Haklay M. (2010). How good is volunteered geographical information? a comparative study of openstreetmap and ordnance survey datasets. *Environment and Planning B: Planning and Design*, vol. 37, n° 4, p. 682-703.

Hennig S., Zobl F., Wasserburger W.W. (2017). Accessible Web Maps for Visually Impaired Users: Recommendations and Example Solutions. *Cartographic Perspectives*, vol. 88. https://doi.org/10.14714/CP0.1391.

Isola P., Zhu J., Zhou T., Efros A.A. (2016). Image-to-image translation with conditional adversarial networks. *CVPR 2017*.

Josselin D., Roussel D., Boularouk S., Saidi A., Matrouf D., Bonin O., *et al.* (2016, juin). Sonorous cartography for sighted and blind people. *AGILE'2016-19th AGILE International Conference on Geographic Information Science*. Helsinki, Finland.

Kaiser P., Wegner J. D., Lucchi A., Jaggi M., Hofmann T., Schindler K. (2017). Learning aerial image segmentation from online maps. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, n° 11, p. 6054-6068.

Kang Y., Gao S., Roth R.E. (2019). Transferring Multiscale Map Styles Using Generative Adversarial Networks. *International Journal of Cartography*, vol. 5, n° 2-3, p. 115-141.

Kern R. (2016). Cartographie et malvoyance - du papier au numérique. *Cartes & Géomatique*, vol. 229-230, p. 167-174.

Kurath S., Das Gupta R., Keller S. (2017, 01). Osmdeepod - object detection on orthophotos with and for vgi, vol. 1, p. 173-188.

Lokhat I., Touya G. (2016). Enhancing Building Footprints with Squaring Operations. *Journal of Spatial Information Science*, vol. 13, p. 33-60.

Long J., Shelhamer E., Darrell T. (2014). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, n° 4, p. 640-651.

Mirza M., Osindero S. (2014). Conditional generative adversarial nets. *CoRR*, vol. abs/ 1411.1784.

Piasco N., Sidibé D., Gouet-Brunet V., Demonceaux C. (2018). Apprentissage de modalités auxiliaires pour la localisation basée vision. *Reconnaissance des Formes, image, Apprentissage et Perception (RFIAP)*. Juin, Marne-la-Vallée, France.

Ronneberger O., Fischer P., Brox T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Proceedings of International Conference on Medical image computing and computer-assisted intervention*, p. 234-241.

Savino, S., Rumor M., Zanon M., Lissandron I. (2010). Data Enrichment for Road Generalization through Analysis of Morphology in the CARGEN Project.*Proceedings of 13th ICA Workshop on Generalisation and Multiple Representation*. Zurich, Switzerland, 2010.

Shotton J., Johnson M., Cipolla R. (2008, June). Semantic texton forests for image categorization and segmentation. *2008 ieee conference on computer vision and pattern recognition*, p. 1-8.

Stampach R., Mulíčková E. (2016). Automated Generation of Tactile Maps. *Journal of Maps*, vol. 12, n° sup 1, p. 532-540. https://doi.org/10.1080/17445647.2016.1196622.

Taal, F., Bidarra R. (2016). Procedural generation of traffic signs.*Proceedings of the Eurographics Workshop on Urban Data Modelling and Visualisation* (pp. 17-23). Eurographics Association.

Taylor L., Nitschke G. (2017). Improving deep learning using generic data augmentation. *CoRR*, vol. abs/1708.06020.

Touya, G. (2010). Enrichissement Automatique de Données Par Analyse Spatiale Pour La Généralisation de Réseaux. *Revue Internationale de Géomatique*, vol. 20, n° 2, p. 175-200.

Touya G., Antoniou V., Christophe S., Skopeliti A. (2017). Production of Topographic Maps with VGI: Quality Management and Automation.*Mapping and the Citizen Sensor*, edited by G. Foody, L. See, S. Fritz, P. Mooney, A. M. Olteanu-Raimond, C. C. Fonte, and V. Antoniou, 61-91. London: Ubiquity Press, 2017. https://doi.org/10.5334/bbf.d

Touya G., Christophe S., Favreau J.-M., Ben Rhaiem A. (2019). Automatic derivation of on-demand tactile maps for visually impaired people: first experiments and research agenda. *International Journal of Cartography*, vol. 5, n° 1, p. 65-91.

Wabinski, J., Moscicka A. (2019). 'Automatic (Tactile) Map Generation—A Systematic Literature Review'. *ISPRS International Journal of Geo-Information*, vol. 8, n° 7, p. 293. https://doi.org/10.3390/ijgi8070293.

Wang T., Liu M., Zhu J., Tao A., Kautz J., Catanzaro B. (2017). High-resolution image synthesis and semantic manipulation with conditional gans. *CVPR 2018*.

Watanabe T., Yamaguchi T., Koda S., Minatani K. (2014, juillet). Tactile Map Automated Creation System Using OpenStreetMap. *Proceedings of 14th International Conference on Computers Helping People with Special Needs*. Paris, France.

Yu F., Koltun V. (2016). Multi-scale context aggregation by dilated convolutions. *International Conference on Learning Representations 2016*.

Zhang H., Xu T., Li H., Zhang S., Huang X., Wang X.*et al.* (2016). Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks.*ICCV 2017*.

Zhu, J.-Y., Park T., Isola P., Efros A.A. (2017). 'Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks'. *IEEE International Conference On Computer Vision (ICCV)*.

Zielstra D., Zipf A. (2010). A comparative study of proprietary geodata and volunteered geographic information for Germany. *13th agile international conference on geographic information science, 2010*.