

DÉPARTEMENT DE GÉOMATIQUE APPLIQUÉE
FACULTÉ DES LETTRES ET SCIENCES HUMAINES
UNIVERSITÉ DE SHERBROOKE

ESSAI

SEGMENTATION SÉMANTIQUE EN TEMPS RÉEL À
PARTIR D'UN NANO-ORDINATEUR : ÉTUDE DES
PERFORMANCES ET DES LIMITES

*Essai présenté pour l'obtention du grade de Maître en sciences (M.Sc.),
cheminement géodéveloppement durable*

VINCENT LE FALHER

LONGUEUIL
SEPTEMBRE 2020

Remerciements

Je tiens à remercier ...

DRAFT

Table des matières

Liste des figures	2
Liste des tableaux	2
1 Introduction	1
1.1 Mise en contexte	1
1.2 Problématique	2
1.3 Objectifs	3
2 Cadre théorique (état des connaissances, revue de la littérature) – 5-9 pages	4
2.1 Cadre théorique au sujet du nano-ordinateur	4
2.2 Cadre théorique au sujet de l'apprentissage profond et de la segmentation sémantique – 3-5 pages	5
3 Matériel et méthodes	5
3.1 Site d'étude - 1 page	5
3.2 Données — 3-4 pages	6
3.3 Matériel et logiciels — 2-3 pages	11
3.4 Méthodologie — 10-15 pages	11
4 Résultats – 10-15 pages	12
5 Interprétation et discussion des résultats – 5-7 pages	13
6 Conclusion et recommandations – 3-4 pages	13

Liste des figures

1 Organigramme de la méthodologie	13
---	----

Liste des tableaux

1 Tableau des données	8
---------------------------------	---

1 Introduction

1.1 Mise en contexte

Voici le plan qui est utilisé pour rédiger la mise en contexte de l'essai.

- Présentation du projet avec PJCCI; contexte, objectif et brève synthèse de la solution proposée.
- Mise en contexte de l'essai par rapport au projet avec PJCCI;
- Introduction à la détection d'objets en temps réel;
- Introduction aux objets connectés, aux nano-ordinateurs;
- Introduction aux réseaux de neurones;
- Introduction à la segmentation sémantique avec les réseaux de neurones convolutifs entiers (FCN);
- Conclusion avec les avantages et les défis du projet PJCCI.

La compagnie LES PONTS JACQUES CARTIER ET CHAMPLAIN INCORPORÉE (PJCCI) désire évaluer la mise en service de la piste multifonctionnelle (vélos, piétons, etc.) du pont Jacques-Cartier, à Montréal, durant l'hiver. Pour ce faire, la piste doit rester sécuritaire et dégagée, malgré les événements météorologiques.

L'université de Sherbrooke, qui participe à cette initiative, propose de mettre en place sur le pont une plateforme de détection innovatrice qui consiste à installer plusieurs paires d'objets connectés ultralégers et performants (des nano-ordinateurs) à différents endroits du pont. Chacun de ces nano-ordinateurs possède trois différents types de capteurs : vision, son, et météorologiques (température, humidité, etc.). Chaque nano-ordinateur d'une paire perçoit le même environnement, mais d'une perspective différente que son homologue : la caméra pointe vers la même surface, mais d'un autre point de vue; les sons et les données météorologiques sont captés dans le même voisinage. Les données collectées par les capteurs sont traitées en temps réel par des algorithmes de détections performants qui sont adaptés à ce type de problématiques : les réseaux de neurones, du domaine de l'intelligence artificielle. La déduction de l'état de la surface de la piste (sèche, mouillée, glacée, etc.) se fait en fusionnant les différentes perceptions (multi-cibles) de chaque capteur (multi-capteurs).

L'objet principal de cet essai consiste à étudier la capacité du nano-ordinateur du fabricant NVIDIA, le Jetson nano [15], à exécuter, en temps réel, un modèle de réseau de neurones entraîné à faire de la segmentation sémantique (classification) d'images de haute résolution qui sont perçues avec la caméra. Les résultats de cette étude permettront de déterminer le modèle de réseau de neurones le plus adapté pour répondre aux besoins du volet vision du projet pour PJCCI.

La détection d'objets et de surface en temps réel est de plus en plus précise et efficace depuis que les performances des systèmes informatisés permettent l'exécution d'algorithmes exigeants, en majeure partie depuis l'utilisation des processeurs graphiques "GPU" [5] [7] [2] [9] [19] [11].

Les systèmes informatiques performants sont de plus en plus miniatures, on parle de nano-ordinateurs et des objets connectés ("Internet of Things" ou "IoT") [4] [18]. Ils permettent la détection en temps réel à des endroits, dans des situations et dans des conditions qui n'étaient pas envisageables il y a encore 10 ans ([19] [3] [1] [4]).

Les réseaux de neurones ont aussi très rapidement progressé depuis 2012 [2], permettant d'offrir des alternatives aux solutions de détection et de classifications, entre autres [17]. Les réseaux

de neurones convolutifs entiers ("FCN" en anglais, pour "Fully Convolutional Network") sont les derniers à avoir émergé ("state-of-art") [19] et à profiter au domaine de la vision et de la détection d'objets ([14] [19]).

La segmentation sémantique est une forme de classification d'image, pixel par pixel, qui tire profit des dernières évolutions de la classification supervisée grâce aux réseaux de neurones convolutifs entiers, et se permet d'être déduite en temps réel avec des nano-ordinateurs ([12] [4]). Les images doivent être de très haute résolution, ce qui nécessite d'avoir à disposition un système informatique capable de fournir une puissance de calcul appropriée, particulièrement pour la manipulation de la mémoire et des nombres flottants pendant l'inférence [13]. Leur application par des nano-ordinateurs est un défi en raison de la faible consommation d'énergie (Watts) et de la puissance de calcul limitée de ces derniers [6].

Pour PJCCI, les avantages d'une telle plateforme seraient multiples, et on peut en énumérer plusieurs, sans se limiter à : contrôler l'épandage de sel ; surveiller les conditions de la piste multifonctionnelle ; suivre les effets du gel et du dégel ; optimiser les coûts des opérations d'entretien (déplacements, quantité) ; offrir aux usagers des conditions d'accès sécurisées et optimales même en hiver ; effets environnementaux atténués ; détecter ce qu'un spécialiste humain ne pourrait pas ou aurait des difficultés à détecter ; prise de décision et gestion proactive ; planification.

D'un autre côté, les défis ne sont pas à sous-évaluer : la détection doit être précise, fiable et consistante, tout cela afin d'assurer aux usagers un service de qualité dans un contexte sécuritaire.

1.2 Problématique

Voici les problèmes que l'essai va contribuer à résoudre.

- Défi pour le domaine du transport actif et durable d'être soutenu par des solutions technologiques fiables (opérationnelles) pour pouvoir offrir des services de qualité et sécuritaire sur l'ensemble des quatre saisons ;
- Capacités d'un nano-ordinateur, tel que le NVIDIA Jetson nano, de faire de la segmentation sémantique en temps réel avec des vidéos de hautes résolutions et d'images par seconde élevée ;
- Problématiques des solutions existantes pour la segmentation sémantique en temps réel sur nano-ordinateur ;

Dans le cadre du projet pour PJCCI, une plateforme technologique sera mise à la disposition des gestionnaires du pont afin de les aider à prendre les décisions les plus responsables et raisonnables possibles. Mais la mise en opération d'une solution innovante et fiable, qui concilie des algorithmes d'apprentissage profond, du temps réel, des nano-ordinateurs, et des conditions climatiques variables, est complexe. Dans une certaine mesure l'essai va contribuer à la recherche de solutions afin de répondre au défi pour le domaine du transport actif et durable d'être soutenu par des solutions technologiques fiables (opérationnelles), l'objectif étant de pouvoir offrir des services de qualité et sécuritaire sur l'ensemble des quatre saisons.

La seconde problématique que l'essai va contribuer à résoudre concerne les limites d'un nano-ordinateur. Un nano-ordinateur est un ordinateur miniaturisé en taille, mais aussi limité en capacité. Il existe différents fabricants et modèles, de spécifications variées, pour répondre à différents besoins. Le dernier né est le modèle "Jetson nano" du fabricant "NVIDIA ", disponible depuis juin 2019 au prix très abordable de 99\$US, et qui sera le matériel utilisé dans le cadre de cet essai. La

compagnie NVIDIA a conçu ce matériel spécialement pour l'inférence de modèles d'apprentissage profond sur une plateforme mobile (drone) ou proche des données ("edge" en anglais). L'inférence nécessite une architecture et une puissance machine différente de celle nécessaire pour l'entraînement. Les modèles de réseaux de neurones sont adaptés et optimisés pour l'inférence. L'essai va permettre de préciser les capacités du Jetson nano pour l'inférence de diverses architectures de réseaux de neurones convolutifs entiers (FCN en anglais) et la segmentation sémantique en temps réel avec des vidéos de différentes propriétés (résolutions et nombre d'images par seconde). Il existe des tests encourageants ([16] [14] [5]), qui seront utilisés comme modèle, même si ceux-ci sont limités à des types d'application qui ne sont pas les mêmes que pour l'essai.

Il est difficile de trouver des jeux de données pour entraîner les réseaux de neurones convolutifs entiers adaptés à la problématique. La technique de "Data augmentation" permet de démarrer d'un modèle qui a déjà appris avec un jeu d'images important (milliers d'images), et de lui faire apprendre davantage, en lui fournissant un plus petit jeu d'images (centaines d'images) de la nouvelle zone d'étude. Par exemple un modèle peut avoir appris à classifier des images de la Californie, États-Unis. Pour lui permettre de classifier des images de la Ville de Sherbrooke, il est souhaitable de lui fournir un nouveau jeu de données spécifique à cette ville afin qu'il s'adapte (ses paramètres) à cette région. Dans le contexte de cet essai, les données acquises sur le terrain seront fournies aux différents modèles qui seront évalués, et qui seront ré-entraînés avec ce nouveau jeu d'images adapté à la zone d'étude.

La paramétrisation (des "hyper-paramètres") des réseaux de neurones est très "subtile" et "intuitive" et requière de l'expérience. C'est un processus d'essais-erreurs qui est très coûteux en temps, et risqué puisqu'il n'y a aucune garantie de succès. La technique de "Transfer Learning" permet d'hériter d'un modèle qui est déjà entraîné et configuré, et de l'adapter pour répondre à ses besoins. Cette technique permet un gain en temps et en énergie (et en argent) important puisque le temps de conception (architecture et configuration) et le temps d'entraînement, de validation et de tests sont diminués de façon non négligeable. La problématique pour l'essai est de trouver le modèle qui est le plus adapté pour répondre au besoin, et il en existe des milliers [10]. La recherche dans la littérature permet heureusement de limiter les choix et donner des pistes ([19] [14] [16]). La problématique de la conception existe toujours, car le modèle a besoin d'être étudié, adapté et ré-entraîné, jusqu'à l'obtention de résultats probants. Mais la paramétrisation des hyper-paramètres n'est plus nécessaire (supposément), ce qui est très avantageux.

1.3 Objectifs

L'objectif principal de l'essai est d'évaluer la capacité du nano-ordinateur "NVIDIA Jetson nano" à exécuter, en temps réel, un modèle de réseau de neurones à convolution entier (FCN) permettant la segmentation sémantique d'une piste multifonctionnelle. La segmentation sera limitée à deux classes : une classe pour la piste et une autre pour les autres éléments présents dans l'image. Pour y arriver, différents sous-objectifs ont été établis.

Voici en bref les sous-objectifs, avec un peu plus de détails par la suite pour chacun d'eux :

- Évaluer les limites de la plateforme, matérielle et applicative ;
- Trouver des moyens d'optimiser la plateforme, au besoin, d'un point de vue matériel, mais aussi applicatif ;
- Permettre un accès à distance sécurisé au nano-ordinateur ;
- Documenter l'approche, les tests, et les résultats ;

— Tester le mode multi-cible (optionnel).

Le premier sous-objectif est de déterminer quelles sont les limites de la plateforme, d'un point de vue matériel (GPU, CPU, mémoire, transfert mémoire, consommation, etc.), mais aussi applicatif (entraînement, inférence). Cette phase du projet va permettre d'exécuter différents modèles déjà existants, sans modification, en tenant compte des éléments documentés dans la littérature [14] [19] [16]. Selon le déroulement de cette étape, un ou plusieurs modèles seront sélectionnés.

Un autre sous-objectif est d'optimiser ou d'adapter la plateforme, d'un point de vue matériel, mais aussi applicatif, afin d'avoir les meilleures performances et résultats possibles pendant l'entraînement et l'inférence.

Comme les résultats devront être disponibles en tout temps, une connexion à distance sécurisée devra être mise en place. Cette connexion permettra aussi de pouvoir prendre le contrôle du nano-ordinateur à distance et de l'administrer.

L'approche, les tests, et les résultats seront documentés. Il y aura beaucoup d'activités relatives à la conception et aux tests, le cheminement complet ne sera pas fourni. Une synthèse sera préférée et les informations les plus pertinentes seront incluses. Les détails de l'installation de l'environnement de développement et des applications, librairies et autres dépendances nécessaires seront inclus, ainsi que ceux de la configuration. Dans le cas où l'objectif principal n'est pas atteint, ou partiellement, la/les raison/s de l'échec seront spécifiées et des pistes de solutions potentielles proposées.

Optionnellement, si le temps le permet, le mode multi-cible pourrait être testé. L'idée est d'avoir deux mêmes nano-ordinateurs qui pointent vers la même cible, et de combiner les deux résultats afin d'améliorer les chances de déduction. Il serait envisageable d'utiliser une deuxième cible "virtuelle" sur le même nano-ordinateur dans un premier temps, afin de ne pas dépendre d'un deuxième nano-ordinateur.

2 Cadre théorique (état des connaissances, revue de la littérature) – 5-9 pages

Il y a deux sections, la première qui concernent le nano-ordinateur et ensuite la seconde, l'apprentissage profond et la segmentation sémantique.

2.1 Cadre théorique au sujet du nano-ordinateur

Voici le plan qui est utilisé pour rédiger le cadre théorique au sujet du nano-ordinateur.

- historique et évolution ; une brève présentation de l'historique des nano-ordinateurs, de leur apparition à leur place aujourd'hui.
- usages ; quelques exemples d'usages des nano-ordinateurs, dans un contexte professionnel.
- architecture ; brève présentation, fonctionnement et comparaison des architectures matérielles des nano-ordinateurs, leurs coûts, leurs avantages et limitations.

2.2 Cadre théorique au sujet de l'apprentissage profond et de la segmentation sémantique – 3-5 pages

Voici le plan qui est utilisé pour rédiger le cadre théorique au sujet des réseaux de neurones et de la segmentation sémantique.

- historique ; une brève présentation de l'historique et du contexte de l'intelligence artificielle, l'apprentissage machine et les réseaux de neurones.

Les concepts de l'Intelligence artificielle (AI) existent depuis les années 1950 et ont continué à se développer jusqu'à leur popularité des 10 dernières années ;

- popularité depuis 10 ans ; argumentation autour des raisons de la renaissance de l'apprentissage machine.

Trois raisons principales ont permis à ce domaine de sortir du champ de la recherche pour celui de l'industrie et la mise en production : amélioration de la capacité et la puissance des machines ; jeux de données plus larges ; algorithmes plus avancés ;

- domaine ; où se situent les réseaux de neurones et la segmentation sémantique dans la hiérarchie de l'IA.

L'apprentissage profond est un sous-domaine de celui de l'apprentissage machine qui est un sous-domaine de celui de l'intelligence artificielle.

La segmentation sémantique d'images ou de vidéos avec des algorithmes d'apprentissage profond fait partie du domaine de la télédétection par la vision.

- applications ; quelques exemples d'applications des réseaux de neurones selon leur type, dont les réseaux de neurones à convolution entiers.

Les applications sont plus sophistiquées, la segmentation sémantique en fait partie.

- principes ; présentation des principes théoriques de la segmentation sémantique.

La segmentation sémantique d'images est une technique élaborée de classification supervisée d'images.

À noter qu'il n'est pas prévu expliquer le fonctionnement des réseaux de neurones, tels que les différentes fonctions (activation, perte), les hyper-paramètres, les différentes architectures et les types de couches.

3 Matériel et méthodes

3.1 Site d'étude - 1 page

Voici le plan qui est utilisé pour rédiger au sujet du site d'étude.

- Brève présentation du pont Jacques-Cartier, et de la piste multifonctionnelle ;
- Présentation des difficultés de l'usage de la piste l'hiver et des défis et raisons (technique, politique, sécurité) de pouvoir la conserver ouverte toute l'année, en lien avec les objectifs de l'essai

3.2 Données -- 3-4 pages

Voici le plan qui est utilisé pour rédiger au sujet des données.

- Présentation des réseaux de neurones qui seront utilisés dans le cadre de l'essai, leurs caractéristiques.
- Présentation des sources de données (images) qui seront utilisées pour l'apprentissage, la validation et les tests des réseaux de neurones pour le ré-entraînement.
- Présentation des sources de données (images et vidéos) qui seront utilisées pour l'inférence.

Données

- Précision des différentes sources consultées :
 - Les ressources mises à disposition par le constructeur du Jetson nano, NVIDIA, font référence à des jeux de données qui sont disponibles publiquement.
 - En complément des ressources de NVIDIA, deux références scientifiques seront principalement étudiées, car leurs recherches ont été faites avec le Jetson nano ([14] et [5]). Beaucoup de références ont été publiées ces deux dernières années sur le sujet de la segmentation sémantique, ils existent donc de multiples alternatives inspirantes.
 - Internet est une mine de données. Il existe des forums et des blogues dans lesquels des utilisateurs publient leurs expérimentations de la segmentation sémantique en temps réel avec le Jetson nano ([8]), ou plus génériquement la segmentation sémantique. Des sites comme "modelzoo.co" ou "kaggle.com" sont des entrepôts de données. Une autre option est d'effectuer une recherche d'images ou de vidéos de la piste multifonctionnelle du pont Jacques-Cartier via les sites de recherche tels que Google.
 - L'Association des piétons et cyclistes du pont Jacques-Cartier existe depuis de nombreuses années pour promouvoir le transport actif et conserver la piste multifonctionnelle du pont Jacques Cartier ouverte durant l'hiver. Ils fournissent, via leurs sites Internet, des collections de vidéos et d'images qui pourraient être utilisées. Il serait aussi possible d'entrer en contact avec l'association et leur demander de prendre de nouvelles vidéos. Voir "<http://pontjacquescartier365.com>", et "<https://www.flickr.com/photos/pontjacquescartier>".
 - Une autre possibilité serait d'hériter des acquisitions faites par un autre étudiant de l'université de Sherbrooke, soit déjà archivée, soit collectée prochainement. Mon directeur de projet Mickaël G. m'a informé qu'un étudiant de Sherbrooke va avoir besoin de collecter le trafic automobile sur le campus de l'Université de Sherbrooke, à Sherbrooke.
 - Enfin il y a l'acquisition des vidéos spécifiquement pour le projet PJCCI. Comme il n'y a aucune date de planifiée pour la capture des vidéos, l'essai devra s'arranger pour dépendre le moins possible d'elles durant la préparation et le développement, et s'attendre à les recevoir pour le ré-apprentissage et les tests, en fin d'essai.
 - Tout au long de l'essai, mon directeur Mickaël sera une ressource importante afin de vérifier que les sources de données, les prétraitements et les traitements sont adéquats aux attentes du projet pour PJCCI.
- Récapitulatif des jeux de données utilisés pour l'essai, grâce à un tableau : les réseaux de neurones et le nom des jeux de données d'imageries et de vidéos respectifs ; leur source ; le nombre d'images ; leur résolution, leur nombre d'images par secondes dans le cas des

vidéos.

- Présentation des images utilisées pour le ré-entraînement des modèles, la validation et les tests ; les traitements nécessaires des vidéos et des images.
- Présentation des vidéos et des images qui seront utilisées pour l'inférence ; les traitements nécessaires des vidéos et des images.
- Mention de la méthodologie d'acquisition des nouvelles données sur le site d'étude (même si potentiellement elles ne seront pas utilisées pendant l'essai ?).

Approche prévue pour le traitement des données

Il y a deux phases à cet essai : l'inférence avec des modèles déjà prêts et l'inférence avec des modèles ré-entraînés. Les données utilisées par l'inférence sont des vidéos (d'une certaine résolution et d'un certain nombre d'images par seconde), et celles pour l'entraînement sont des images. Dans les deux cas, les images pour l'entraînement ou l'inférence doivent être d'une taille bien précise, celle avec laquelle le modèle a été, ou sera, entraîné. La résolution et la qualité de l'image-vidéo seront nivelées vers le bas afin de déterminer la limite inférieure acceptable pour la détection la plus efficace et fiable possible. La résolution et le nombre d'images par seconde de la vidéo sont contrôlés par le logiciel ("driver" en anglais) de la caméra, et sont configurables.

Tout cela signifie que les vidéos ou nouvelles images devront être traitées pour répondre à une certaine taille et résolution requise par le modèle, tout en conservant une qualité élevée (nombre de pixels, niveaux de couleurs). De nouvelles images pour l'entraînement seront extraites des vidéos, et annotées.

Certains framework d'apprentissage profond (par exemple "Keras") offrent l'option d'augmenter automatiquement le jeu de données avec des techniques d'augmentation de données (par exemple la rotation, le redimensionnement, l'effet miroir), ce qui est très utile et non négligeable.

Voici le tableau de synthèse des données, incluant la référence avec leur réseaux de neurones.

TABLE 1: Tableau des données

	Spécification	Description
1	<p>réseau : U-Net</p> <p>jeu de données : Membrane (origine isbi challenge)</p> <p>nombre d'images : 30</p> <p>résolution/s : 512x512</p>	<p>C'est le jeu de données pour le réseau U-Net. Il est utilisé dans le benchmark de NVIDIA pour l'inférence avec le Jetson nano. Les images sont de type médicale.</p> <p>À noter que le framework "Keras" s'occupe de l'augmentation de données.</p> <p>https://github.com/zhixuhao/unet/tree/master/data/membrane</p>
2	<p>réseau : SegNet</p> <p>jeu de données : CamVid</p> <p>vidéos : 10 minutes</p> <p>résolution/s : HD</p>	<p>SegNet est un réseau qui a été créé pour la segmentation sémantique de vidéos. Il a été entraîné avec le jeu de données de CamVid, qui procurent des vidéos de la route avec la même perspective que le conducteur du véhicule. Un modèle entraîné est disponible pour le Jetson nano.</p> <p>https://github.com/PengKiKi/camvid</p>
3	<p>réseau : MFANet</p> <p>jeu de données : Cityscapes</p> <p>nombre d'images : 5000</p> <p>résolution/s : 1280x1024</p>	<p>MFANet est un réseau qui a été créé en 2019 pour la segmentation sémantique sur des appareils tel que le Jetson nano. Il a été entraîné avec le jeu de données de Cityscapes, qui procurent des images de scènes urbaines. Différentes stratégies d'augmentation de données sont utilisées. Des tests ont été fait avec le Jetson nano.</p> <p>leejy@ustb.edu.cn</p>
4	<p>réseau : MAVNet</p> <p>jeu de données : Penstock</p> <p>nombre d'images : 135</p> <p>résolution/s : 1280x1024</p>	<p>C'est l'un des deux jeux de données pour le réseau MAVNet. Les images sont celles de "conduites forcées", des voies d'eau de régulation, et sont préparées pour la segmentation sémantique. Des tests ont été fait avec le Jetson nano.</p> <p>https://github.com/tynguyen/MAVNet/tree/master/data/TN_penstock</p>
5	<p>réseau : MAVNet</p> <p>jeu de données : Penstock</p> <p>nombre d'images : 135</p> <p>résolution/s : 1280x1024</p>	<p>C'est l'un des deux jeux de données pour le réseau MAVNet. Les images sont celles de drones volant à l'intérieur d'un bâtiment, et préparées pour la segmentation sémantique. Des tests ont été fait avec le Jetson nano.</p> <p>https://github.com/tynguyen/MAVNet/tree/master/data/perch_drone</p>
6	<p>réseau : RESNet18</p> <p>jeu de données : Cityscapes</p> <p>nombre d'images : 25 000</p> <p>résolution/s : 360x720, 512x256, 1024x512, 2048x1024</p>	<p>Cityscapes est un jeu de données qui fournit des images de rues spécifiquement destinées pour la segmentation sémantique. Il peut être utilisé par de nombreux réseaux.</p> <p>RESNet18 a été entraîné avec ce jeu et est disponible en diverses résolutions pour le Jetson Nano.</p> <p>https://github.com/tynguyen/MAVNet/tree/master/data/perch_drone</p>

	Spécification	Description
7	réseau : RESNet18 jeu de données : DeepScenes nombre d'images : 15 000 résolution/s : 576x320, 864x480	DeepScene propose un modèle et un jeu de données. Le modèle est entraîné avec différents jeux de données, comme Cityscapes, SUN-RGBD, Synthia. Le jeu de données fournit des images de forêt, qui est destinée pour la segmentation sémantique. RESNet18 a été entraîné avec ce jeu et est disponible en deux résolutions pour le Jetson Nano. http://deepscene.cs.uni-freiburg.de
8	réseau : RESNet18 jeu de données : Multi-Human nombre d'images : 25 043 résolution/s : 512x320, 640x360	Le jeu de données Multi-Human fournit des images contenant des humains, et qui est destinée pour la segmentation sémantique. RESNet18 a été entraîné avec ce jeu et est disponible en deux résolutions pour le Jetson Nano. https://lv-mhp.github.io/dataset
9	réseau : RESNet18 jeu de données : Pascal VOC nombre d'images : 11 530 résolution/s : 320x320, 512x320	Le jeu de données Pascal VOC fournit des images de classes variées tel que des personnes, des animaux, des véhicules, et des objets classiques, et qui peut être utilisé pour la segmentation sémantique. RESNet18 a été entraîné avec ce jeu et est disponible en deux résolutions pour le Jetson Nano. http://host.robots.ox.ac.uk/pascal/VOC/voc2012/index.html
10	réseau : RESNet18 jeu de données : SUN RGB-D nombre d'images : 10 335 résolution/s : 512x400, 640x512	Le jeu de données SUN RGB-D fournit des images de scènes d'intérieur de maison, et qui est destiné pour la segmentation sémantique. RESNet18 a été entraîné avec ce jeu. https://synthia-dataset.net
11	réseau : DeepScene jeu de données : Synthia nombre d'images : 220 000 résolution/s : 1280x760	Le jeu de données Synthia fournit des images (et vidéos) de scènes de rue comme celui de Cityscapes, et qui est destiné pour la segmentation sémantique. DeepScene a été entraîné avec ce jeu. Il n'a pas été testé avec le Jetson Nano. http://3dvision.princeton.edu/datasets.html
12	jeu de données : Association des piétons et cyclistes pont Jacques-Cartier nombre d'images : 313 résolution/s : variées	L'Association des piétons et cyclistes du pont Jacques-Cartier a une collection d'images et de vidéos de la piste multifonctionnelle du pont Jacques-Cartier. Ce n'est pas un jeu de données qui est prêt à être utilisé pour l'apprentissage tel-qu'il, il doit être préparé. Mais c'est une source de données qui est très importante pour l'essai. Il est envisagé de contacter l'association au besoin afin de leur demander leur collaboration pour la collecte d'autres d'images ou vidéos. https://www.flickr.com/photos/pontjacquescartier http://pontjacquescartier365.com/videos-pont-jacques-cartier

	Spécification	Description
13	<p>jeu de données : images et vidéo sur Internet</p> <p>nombre d'images : entre 30-50</p> <p>résolution/s : variées</p>	<p>Internet est une source de données non négligeable en terme de données. Quelques images et vidéos de la piste multifonctionnelles du pont Jacques-Cartier, autres que celles fournies par L'Association des piétons et cyclistes du pont Jacques-Cartier, sont disponibles. Ce n'est pas un jeu de données qui est prêt à être utilisé pour l'apprentissage tel-quel, il doit être préparé. Mais c'est une source de données qui est très importante pour l'essai.</p> <p>https://google.ca</p>
14	<p>jeux de données : Kaggle</p>	<p>Le site Kaggle offre une vingtaine de jeux de données offert par la communauté pour faire de la segmentation sémantique, et qui sont prêt à être utilisé. Les jeux de données n'ont pas été évalués.</p> <p>https://www.kaggle.com/search?q=</p>

3.3 Matériel et logiciels — 2-3 pages

Voici le plan qui est utilisé pour rédiger au sujet du matériel.

- Présentation du nano-ordinateur qui sera utilisé, le NVIDIA Jetson nano ;
- Usage du nano-ordinateur durant l'essai.

L'objet d'étude de cet essai est un nano-ordinateur. Un nano-ordinateur est un ordinateur miniaturisé en taille, mais aussi limité en capacité. Il existe différents fabricants et modèles, de caractéristiques techniques variées, pour répondre à différents besoins. Le dernier né est le modèle "Jetson nano" du fabricant "NVIDIA", disponible depuis juin 2019 au prix très abordable de 99\$US. La compagnie NVIDIA a conçu ce matériel spécialement pour différentes applications d'inférence de modèles d'apprentissage profond sur une plateforme mobile (drone) ou proche des données ("edge" en anglais). Ce modèle a été choisi afin de répondre à l'intérêt que suscitent ses capacités et ses limites. Une image du Jetson nano et un tableau de ses caractéristiques techniques seront ajoutés.

L'architecture matérielle sera étudiée et présentée avec l'aide d'images, de diagrammes et de textes explicatifs. Les éléments clés seront identifiés.

Afin d'optimiser les performances du Jetson nano, une recherche des périphériques les plus adaptés pour répondre aux besoins de performance (et de budget) de l'essai est essentielle, telle que l'alimentation, le stockage, la caméra. Des images des périphériques seront incluses, et les caractéristiques principales seront présentées dans des tableaux.

Il est à noter que le NVIDIA Jetson nano est déjà en ma possession. La liste des équipements est en cours et sera commandée par le collaborateur "Vision météo".

Logiciels — 1-2 pages De même que pour les périphériques, les logiciels qui seront utilisés seront résumés dans un tableau, où il sera indiqué leur nom, le type de licence, leur version, leurs avantages et limitations, comme pour le système d'exploitation, l'environnement de développement pour l'apprentissage profond, l'inférence, les logiciels de traitements vidéos et d'images.

- Présentation du système d'exploitation qui sera installé ;
- Présentation des frameworks d'apprentissage profond qui vont être utilisés (PyTorch, torchvision) ;
- Présentation des bibliothèques pour l'inférence (TensorRT, onnx) ;

3.4 Méthodologie — 10-15 pages

Voir la section 3.4 et l'organigramme de la figure 1 pour la représentation graphique de la méthodologie, et dont les phases peuvent être résumées de la façon suivante :

- Recherche des références, des modèles et des données, ainsi que l'équipement pour le nano-ordinateur et des logiciels nécessaires.
- Installation sur le Jetson nano du système d'exploitation, de l'environnement de développement et de tests pour l'inférence.
- Itération entre les étapes suivantes :
 - Inférence avec le Jetson nano en utilisant les modèles et les sources de données sélectionnées.
 - Adaptation des modèles à différentes résolutions d'images et à la zone d'étude.

— Traitement des données afin de les adapter au requis des modèles.

L'organigramme de la figure 1 présente la démarche de la réalisation du projet.

L'objectif principal de l'essai est de déterminer la capacité et les limites du nano-ordinateur d'inférer en temps réel des modèles de réseau de neurones à convolution entier pour la segmentation sémantique de vidéos. La stratégie qui sera appliquée sera de tester avec divers modèles et divers niveaux de qualité vidéos, en espérant trouver le compromis qui répond le mieux à cet objectif.

1. Afin de s'assurer du bon fonctionnement du nano-ordinateur et d'avoir des résultats de référence propre à notre environnement, l'inférence sera testée avec des modèles existants et pré-entraînés pour la segmentation sémantique, avec les images et les vidéos provenant des références, et dont les caractéristiques et les résultats sont disponibles.
2. En espérant que les tests de l'étape #1 précédente donnent les résultats documentés dans les articles de références, ils seront repris avec les mêmes modèles, mais avec les images et les vidéos du site d'étude possédant la meilleure qualité acquise (1080p/i, 30FPS). Les données sources (images et vidéos) devront subir certains prétraitements à ce effet, afin de répondre aux requis des modèles.
3. Selon les résultats de l'étape #2, les tests se concentreront sur l'inférence avec des vidéos, en réduisant progressivement la résolution (760p/i, 576p/i, 480p/i, 360p/i) et le nombre d'images par seconde (20FPS, 10FSP, 1FPS).
4. Les étapes intermédiaires de l'étape #3 précédente seront de 1) valider les résultats de l'inférence avec des images avant de tester avec les vidéos, et 2) évaluer si les modèles de réseaux de neurones à convolution entiers doivent et/ou peuvent être adaptés facilement, en tenant compte de l'échéancier de l'essai, et ce afin de répondre à l'objectif principal.

L'accès aux connaissances et à l'expérience de mon directeur de projet dans le domaine de l'apprentissage profond, ainsi que l'adhésion au "Deep Learning Institute" offerte par NVIDIA, sont les soutiens à ma disposition pour arriver à compléter les objectifs de cet essai avec succès, et à trouver des solutions de contournement aux problèmes qui seront rencontrés.

À noter que l'acquisition des données du site d'étude se fera par l'intermédiaire d'une autre partie. La présentation du matériel et de la méthode d'acquisition des vidéos et des images terrain est donc exclue de cet essai.

4 Résultats – 10-15 pages

Voici le plan qui est utilisé pour rédiger les résultats.

- Pour chaque modèle et résolution utilisés, la segmentation sémantique de certaines images et vidéos sera présentée. La segmentation qui a réussi, celle qui est moins précise, et celle qui a échoué seront soulignées. Un résumé du % de succès vs des échecs sera fait, selon les modèles et les résolutions.
- En complément de la section précédente, les performances du Jetson nano pour les divers scénarios de test seront résumés avec différents indicateurs. Ceux qui ont échoué ou n'ont pas été possibles en raison des limitations du nano-ordinateur seront indiqués.

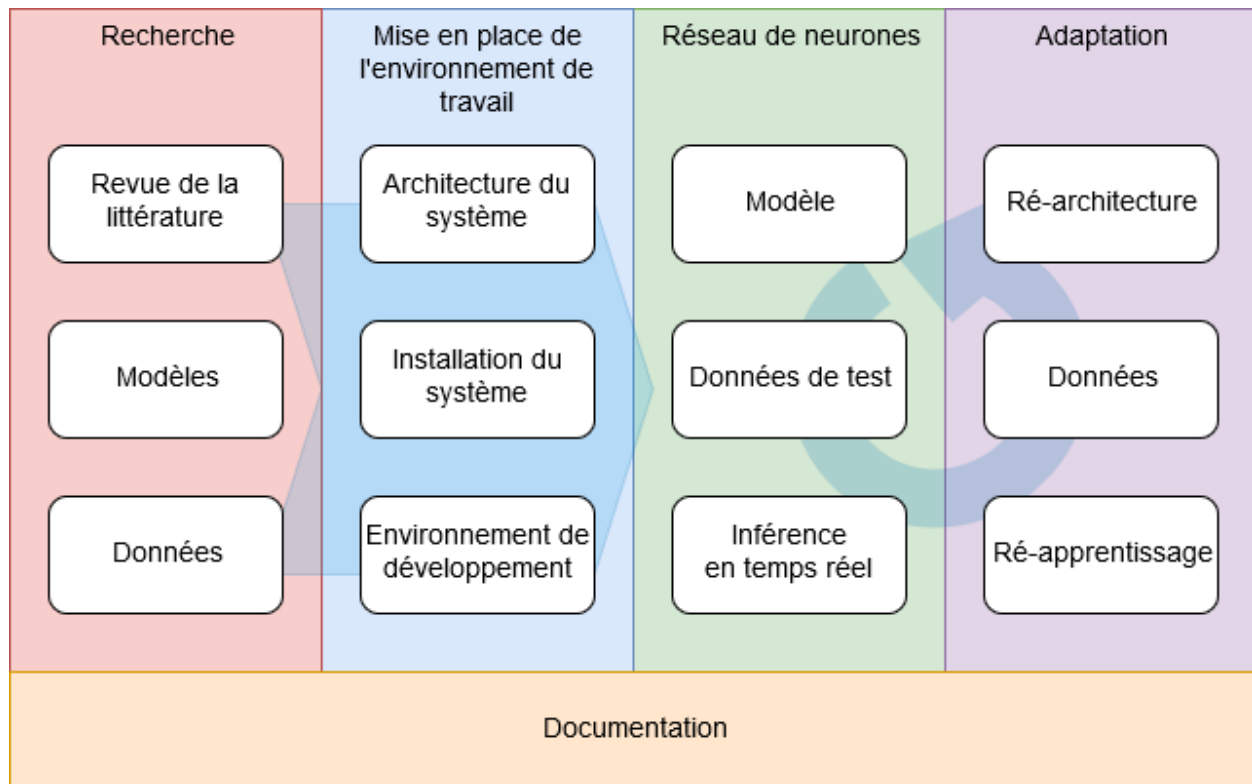


FIGURE 1 – Organigramme de la méthodologie

- Enfin les performances de l'inférence et des modèles de réseaux de neurones pour la segmentation sémantique seront listées. Des indicateurs de performance classiques et tirés de la littérature seront utilisés.

5 Interprétation et discussion des résultats – 5-7 pages

Voici le plan qui est utilisé pour rédiger le cadre théorique au sujet du nano-ordinateur.

- Interprétation et discussion des résultats de la segmentation sémantique en temps réel de la zone d'études ;
- Interprétation et discussion des performances du nano-ordinateur "Jetson nano" pour l'inférence en temps réel dans le contexte de l'essai.
- Interprétation et discussion des performances de l'inférence et des modèles de réseaux de neurones pour la segmentation sémantique en temps réel de la zone d'étude.

6 Conclusion et recommandations – 3-4 pages

Voici le plan qui est utilisé pour rédiger la conclusion.

- Synthèse des réussites et des échecs de l'essai par rapport aux objectifs ;

- Synthèse des capacités et des limites du nano-ordinateur "Jetson nano" pour l'inférence en temps réel à des fins de segmentation sémantique de vidéos ;
- Synthèse des capacités et des limites des modèles de réseaux de neurones pour la segmentation sémantique en temps réel de la zone d'études ;
- Recommandations ;

DRAFT

Bibliographie

- [1] S. ABOUZHAR, M. SADIK et E. SABIR. "IoT-Empowered Smart Agriculture : A Real-Time Light-Weight Embedded Segmentation System". In : *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2017), p. 319-332. DOI : 10.1007/978-3-319-68179-5_28.
- [2] A. BEAM. *Deep Learning 101 - Part 1 : History and Background*. 2017. URL : https://beamandrew.github.io/deeplearning/2017/02/23/deep_learning_101_part1.html.
- [3] M. BERNAS, B. PLACZEK et A. SAPEK. "Edge Real-Time Medical Data Segmentation for IoT Devices with Computational and Memory Constrains". In : *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (2017), p. 119-128. DOI : 10.1007/978-3-319-67077-5_12.
- [4] B. BLANCO-FILGUEIRA et al. "Deep Learning-Based Multiple Object Visual Tracking on Embedded System for IoT and Mobile Edge Computing Applications". In : *IEEE Internet of Things Journal* (juin 2019), p. 5423-5431. DOI : 10.1109/JIOT.2019.2902141.
- [5] C. P. CHONG, C. A. T. SALAMA et K. C. SMITH. "Real-Time Edge Detection and Image Segmentation". In : *Analog Integrated Circuits and Signal Processing* (1992), p. 117-130. DOI : 10.1007/BF00142412.
- [6] M. COPEL. *What's the Difference Between Deep Learning Training and Inference ?* Août 2016. URL : <https://blogs.nvidia.com/blog/2016/08/22/difference-deep-learning-training-inference-ai/>.
- [7] T. DETTMERS. *Deep Learning in a Nutshell : History and Training*. Déc. 2015. URL : <https://devblogs.nvidia.com/deep-learning-nutshell-history-training/>.
- [8] F. DUSTIN. *Realtime Semantic Segmentation on Jetson Nano in Python and C++*. Oct. 2019. URL : <https://www.linkedin.com/pulse/realtime-semantic-segmentation-jetson-nano-python-c-dustin-franklin>.
- [9] JIACONDA. *A Concise History of Neural Networks*. Avr. 2019. URL : <https://towardsdatascience.com/a-concise-history-of-neural-networks-2070655d3fec>.
- [10] J. Y. KOH. *Model Zoo - Deep Learning Code and Pretrained Models for Transfer Learning, Educational Purposes, and More*. 2018. URL : <https://modelzoo.co/>.
- [11] A. KURENKOV. *A 'Brief' History of Neural Nets and Deep Learning*. 2015. URL : <https://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/>.
- [12] J. LONG, E. SHELHAMER et T. DARRELL. "Fully Convolutional Networks for Semantic Segmentation". In : *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Juin 2015, p. 3431-3440. DOI : 10.1109/CVPR.2015.7298965.
- [13] M. MODY et al. "Low Cost and Power CNN/Deep Learning Solution for Automated Driving". In : *Proceedings - International Symposium on Quality Electronic Design, ISQED*. 2018, p. 432-436. DOI : 10.1109/ISQED.2018.8357325.

- [14] T. NGUYEN et al. “MAVNet : An Effective Semantic Segmentation Micro-Network for MAV-Based Tasks”. In : *arXiv :1904.01795 [cs]* (juin 2019). URL : <http://arxiv.org/abs/1904.01795>.
- [15] NVIDIA. *Jetson Nano*. Mar. 2019. URL : <https://developer.nvidia.com/embedded/jetson-nano>.
- [16] NVIDIA. *Jetson Nano : Deep Learning Inference Benchmarks*. Avr. 2019. URL : <https://developer.nvidia.com/embedded/jetson-nano-dl-inference-benchmarks>.
- [17] D. PATHAK et M. EL-SHARKAWY. “Architecturally Compressed CNN : An Embedded Realtime Classifier (NXP Bluebox2.0 with RTMaps)”. In : *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*. Jan. 2019, p. 0331-0336. DOI : 10.1109/CCWC.2019.8666495.
- [18] N. SHARMA, M. SHAMKUWAR et I. SINGH. *The History, Present and Future with Iot. Intelligent Systems Reference Library*. Springer Science and Business Media Deutschland GmbH, 2019. ISBN : 18684394 (ISSN). DOI : 10.1007/978-3-030-04203-5_3. URL : https://www.scopus.com/inward/record.uri?eid=2-s2.0-85060126104%5C&doi=10.1007%5C%2F978-3-030-04203-5_3%5C&partnerID=40%5C&md5=1878f15afc39fbca5142a5f680e0f3c7.
- [19] J. ZHENG et al. “Real-Time Semantic Segmentation Network for Edge Deployment”. In : *Proceedings of 2019 Chinese Intelligent Systems Conference*. Sous la dir. d’Y. JIA, J. DU et W. ZHANG. Springer Singapore, 2020, p. 243-249. ISBN : 978-981-329-697-8 978-981-329-698-5. DOI : 10.1007/978-981-32-9698-5_28. URL : http://link.springer.com/10.1007/978-981-32-9698-5_28.