Apprentissage multitâche auto-supervisé pour la segmentation d'images

Lichun Gao
Département d'informatique
Institut polytechnique de Worcester
Worcester, MA
lgao2@wpi.edu

Chinmaya Khamesra
Département de robotique
Institut polytechnique de Worcester
Worcester, MA
ckhamesra@wpi.edu

Uday Kumbhar
Département de la science des données
Institut polytechnique de Worcester
Worcester, MA
ukumbhar@wpi.edu

Ashay Aglawe
Département de la science des données
Institut polytechnique de Worcester
Worcester, MA
alaglawe@wpi.edu

***Résumé* - *Grâce* aux percées réalisées dans le domaine de l'IA et de l'apprentissage profond, les techniques de vision par ordinateur s'améliorent rapidement. La plupart des applications de vision par ordinateur nécessitent une segmentation sophistiquée de l'image afin de comprendre ce qu'elle contient et de faciliter l'analyse de chaque section. L'entraînement des réseaux d'apprentissage profond pour la segmentation sémantique nécessite une grande quantité de données annotées, ce qui représente un défi majeur dans la pratique, car la production de ces données est coûteuse et nécessite beaucoup de travail. L'article présente**
1. Techniques auto-supervisées pour améliorer les performances de la segmentation sémantique en utilisant l'apprentissage multitâche avec la prédiction de la profondeur et la normalisation de la surface. 2. Évaluation des performances des différents types de techniques de pondération (UW, Nash-MTL) utilisées pour l'apprentissage multitâche. L'ensemble de données NY2D a été utilisé pour l'évaluation des performances. Selon notre évaluation, la méthode Nash-MTL est plus performante que l'apprentissage d'une seule tâche (segmentation sémantique).

I. Introduction

L'apprentissage profond est désormais reconnu comme une stratégie standard pour des problèmes tels que la classification, la segmentation et la détection, car la vision par ordinateur et l'apprentissage automatique se sont rapidement améliorés. Un grand nombre de techniques de pointe reposent sur l'apprentissage supervisé, qui nécessite l'étiquetage manuel des données, ce qui est à la fois long et coûteux.

Il est possible de trouver des images et des vidéos non étiquetées en grandes quantités pour un prix modique. Malheureusement, leur potentiel est rarement exploité au maximum. L'apprentissage non supervisé est utilisé pour découvrir des modèles cachés dans des données non étiquetées, mais il n'est pas conçu pour résoudre un problème spécifique. Par conséquent, il ne tient pas compte d'informations importantes nécessaires à la réalisation de tâches visuelles (par exemple, la segmentation).

L'apprentissage auto-supervisé a le potentiel de surmonter les limitations et de capitaliser sur les avantages de l'apprentissage supervisé et non supervisé. Il s'agit d'une sorte d'apprentissage supervisé dans lequel les étiquettes sont générées automatiquement à partir de données non étiquetées. Par conséquent, contrairement à l'apprentissage non supervisé, l'apprentissage auto-supervisé se concentre sur l'optimisation d'une tâche spécifique, forçant réseau à acquérir des connaissances sémantiques sans avoir à gérer des problèmes supplémentaires liés aux étiquettes.

La majorité des recherches antérieures sur l'apprentissage supervisé et auto-supervisé se sont concentrées sur une seule tâche à la fois. Cela donne de bons résultats, mais ne tient pas compte d'un grand nombre de données pertinentes. Lorsque de nombreuses tâches sont entraînées simultanément, les connaissances spécifiques au domaine sont davantage exploitées, ce qui permet une meilleure généralisation.

A. Contribution à la recherche

Les travaux antérieurs ont porté sur des techniques de pondération telles que la grille, la recherche, la pondération de l'incertitude (UW) et la moyenne de pondération dynamique. Ces travaux ont permis de conclure que la méthode de pondération de l'incertitude (UW) était plus performante que la méthode de pondération dynamique (DWA). Dans le prolongement de ces travaux, nous avons essayé la méthode de pondération Nash-MTL, qui a donné les meilleurs résultats par rapport à la méthode de pondération UW.

II. Travaux connexes

Les deux tâches auto-supervisées sur lesquelles nous nous concentrons sont la prédiction de la normalité de la surface et la prédiction de la profondeur. Nous remplissons tâches de segmentation sémantique en nous basant sur le domaine de l'apprentissage profond. En raison des multiples tâches incluses, nous appliquons l'apprentissage multi-tâches pour améliorer l'efficacité de notre expérience. Nous aborderons les détails dans les sections suivantes.

A. Apprentissage auto-supervisé

L'apprentissage auto-supervisé est un outil utile qui nous permet d'apprendre l'ampleur de plus de données, ce qui aide le modèle à grande échelle à être formé même en l'absence d'étiquettes. L'apprentissage auto-supervisé apprend le signal à partir des données originales. L'objectif de l'apprentissage auto-supervisé est de prédire la partie cachée ou la partie qui n'a pas été observée pour la valeur d'entrée dans le contexte des parties non cachées. L'important est que l'apprentissage auto-supervisé génère automatiquement des étiquettes de vérité de terrain. Notre solution pour les tâches 1) prédiction de la normale de la surface et
2) la prédiction de la profondeur utilise l'apprentissage auto-supervisé. La limite de l'apprentissage auto-supervisé est principalement liée à la sous-performance par rapport à l'apprentissage supervisé. Cependant, ce problème n'influence pas notre expérience basée sur nos tâches.

1) Prédiction de la normale de surface : La prédiction de la normale de surface consiste à prédire l'orientation de la surface des objets présents dans une scène. De nombreux chercheurs ont travaillé dans ce domaine et ont trouvé de nombreuses méthodes pour accomplir cette tâche. Eigen et al. ont utilisé les normales de surface dans leur étude sur l'architecture convolutionnelle. Qi et al. ont également appliqué à leur étude la prédiction conjointe de la profondeur et l'estimation des normales de surface. D'après l'étude précédente, nous pouvons également dire que la prédiction des normales de surface s'accompagne généralement d'une tâche d'estimation des normales de surface.

la prédiction de la profondeur. Compte tenu de cette partie, nous incluons les deux tâches dans notre étude.

Pour la prédiction de la normale de surface, nous aurons la formule suivante :

$$L(I, Y) = -\sum_{i=1}^{(\Sigma)(M\times M)} \sum_{k=1}^{K} (K(y_i = k) \log F_{i,k}(I))$$

Où $F_{i,k}(I)$ est la probabilité pour le ième pixel, qui devrait avoir la normale définie par le kième code. $K(y_i = k)$ représente la fonction indicatrice et $Y = y_i$ est l'ensemble des étiquettes de la vérité de terrain pour la prédiction de la normale de surface. En outre, $M = M(t)$ et $K = K(t)$.

2) Prédiction de la profondeur : La prédiction de la profondeur est une tâche indispensable à la compréhension de la scène 3D. La tâche est ambiguë dans une certaine mesure en raison de l'échelle globale. La prédiction de la profondeur joue un rôle important dans la conduite autonome, c'est pourquoi de nombreux chercheurs se concentrent sur ce domaine.

dès à présent. La prédiction de la profondeur dans le cadre de l'apprentissage supervisé a montré que la segmentation sémantique est un processus complexe. Étant donné que la segmentation sémantique nécessite une étiquette de vérité terrain,

Les résultats sont prometteurs, mais le coût est énorme car nous devons utiliser les capteurs et les ressources humaines pour effectuer cette tâche. Pour rendre cette tâche rentable, l'objectif de la recherche s'est déplacé vers

l'apprentissage auto-supervisé et l'apprentissage non supervisé. Sur la base des travaux de Godard et al., nous commençons à apprendre la profondeur en

en capturant les images de différents points de vue à l'aide d'équations géométriques simples. Grâce à cette amélioration, Sudeep et al. ont découvert que le goulot d'étranglement des performances de prédiction de la profondeur était la faible résolution de l'image. Par conséquent, Vitor et al. nous ont présenté une architecture de réseau neuronal spécialisée dans l'estimation de la profondeur monoculaire autosupervisée. Avec le développement de la technologie, nous disposons désormais de nombreux ensembles de données utiles. Par exemple, le benchmark Dense Depth for Automated Driving (DDAD) (Profondeur dense pour la conduite automatisée). Les résultats ont été améliorés pour les travaux suivants.

Dans notre formation, nous utiliserons les images gauche (I^l) et droite (I^r) pour prédire, en transformant limage droite en image gauche. Nous reconstruirons l'image de gauche sur la base de l'image de droite.

formule suivante, où $\hat{I}^l$ désigne notre gauche reconstruite

et bilinéaire désigne la fonction d'interpolation bilinéaire :

$$\hat{I}^l = bilinear(I^r, d)$$

Nous obtiendrons ensuite une perte de reconstruction similaire à celle de la

L2 :

$$L_{profondeur}(I^r, I^l, \hat{I}^l) = \frac{1}{N} \sum_{i=0}^{N-1} (I^l_i - \hat{I}^l_i)^2$$

Nous utiliserons la fonction sigmoïde pour convertir la sortie de la couche convolutionnelle finale en disparité. Enfin, nous pouvons obtenir la profondeur à partir de la disparité.

B. Segmentation sémantique

La segmentation sémantique est l'une des tâches sur lesquelles les chercheurs se concentrent le plus dans le domaine de la vision par ordinateur : 1) la classification, 2) la localisation et 3) la segmentation. Pour être plus précis, la segmentation sémantique est le processus de classification des pixels et de génération d'une étiquette pour les éléments suivants

chaque pixel. La segmentation sémantique est importante car elle permet de dériver la corrélation de l'image d'entrée et de supprimer le bruit. Dans les méthodes d'apprentissage profond, le réseau neuronal convolutif est fréquemment utilisé pour effectuer cette tâche. Long et al. nous ont le premier réseau entièrement convolutif pour la segmentation sémantique. Il s'agissait d'une étape importante parce qu'elle

améliore considérablement l'efficacité et la précision. Vijay et al. avaient a proposé le SegNet, qui est basé sur le CNN. D'autres solutions incluent UNet, PSPNet, PANet et DANet. Toutefois, compte tenu limites des exigences de calcul, bon nombre de ces réseaux ne sont pas adaptés à l'industrie. Pour répondre aux problèmes de l'industrie, des réseaux plus petits tels que ENet, MobileNet, etc. sont proposés.

La formule de perte d'entropie croisée pour la segmentation sémantique est la suivante :

$$L_{seg}(S, \hat{S}) = -\frac{1)}{N} \sum_{i0=}^{N-1} S_i \log(\hat{S}_i)$$

S_i est la vérité terrain pour le i-ième pixel. $\hat{S}_i$ désigne la classe prédiction que :

$$\hat{S}_i = e^{zsi} / \sum_s e^{zsi,s}$$

Où zs est la sortie de la dernière couche convolutive du décodeur. s est le nombre de classes sémantiques.

C. Apprentissage multitâche

L'apprentissage multitâche est un paradigme d'apprentissage important dont l'objectif est d'exploiter les informations essentielles de plusieurs tâches relatives. L'objectif principal de l'apprentissage multitâche est d'améliorer les performances générales de toutes les tâches. Dans le domaine de l'image, Marvin et al. ont proposé Multinet, la première architecture pour la classification, la détection et la segmentation. Elle comprend le codeur (VGG) et le décodeur (décodeur de classification et décodeur de segmentation).

Parmi tous les modèles d'apprentissage multitâches développés, celui qui correspond le mieux à notre travail est l'apprentissage visuel multitâche auto-supervisé proposé par Carl et al. Il a entraîné simultanément différents types de tâches auto-supervisées complémentaires afin d'obtenir les meilleurs résultats d'apprentissage. Ces tâches sont les suivantes 1) la prédiction de la position relative, 2) la prédiction des couleurs, 3) l'apprentissage d'un seul échantillon et 4) la segmentation du mouvement. Deux structures sont utilisées dans ce travail, ce qui rend le réseau plus flexible. Ce travail est orienté vers l'ingénierie mais nous fournit des idées de recherche significatives.

Comme nous utilisons des tâches auto-supervisées, la perte pour notre travail est la somme pondérée des pertes spécifiques à chaque tâche :

$$L_{total} = \lambda 1 L_{surface} + \lambda 2 L_{profondeur} + \lambda 3 L_{seg}$$

Où λ_1, λ_2, λ_3 représente le poids pour 1) la prédiction de la normalité de la surface, 2) la prédiction de la profondeur, 3) la segmentation sémantique.

III. MÉTHODE PROPOSÉE

A. Nash-MTL

Dans l'apprentissage multitâche (AMT), un modèle conjoint est formé pour faire des prédictions pour plusieurs tâches en même temps. L'apprentissage conjoint permet de gagner du temps et de l'argent en réduisant les coûts de calcul et en augmentant l'efficacité des données. Cependant, comme les gradients de ces différentes tâches peuvent entrer en conflit, l'apprentissage d'un modèle conjoint pour l'apprentissage multitâche se traduit souvent par des performances inférieures à celles de ses homologues monotâches.

Une solution courante à ce problème consiste à combiner les gradients par tâche dans une direction de mise à jour commune en utilisant une heuristique spécifique. Nous proposons ici de considérer l'étape de combinaison des gradients comme un jeu de négociation dans lequel les tâches négocient pour parvenir à un accord sur une direction commune de mise à jour des paramètres. Sous certaines hypothèses, le problème de négociation a une solution unique connue sous le nom de solution de négociation de Nash, que nous proposons d'utiliser comme une approche fondée sur des principes pour l'apprentissage multitâche.

Sur la base des conclusions de Nash, nous proposons Nash-MTL, un nouvel algorithme d'optimisation MTL dans lequel les gradients sont combinés à chaque étape à l'aide de la solution de négociation de Nash. Nous commençons par caractériser la solution de négociation de Nash pour MTL et développons un algorithme efficace pour approcher sa valeur. Ensuite, dans les cas convexe et non convexe, nous analysons théoriquement notre approche et établissons des garanties de convergence. Enfin, nous démontrons empiriquement que l'approche NashMTL permet d'obtenir des résultats de pointe pour toute une série de défis.

B. Pondération de l'incertitude

L'apprentissage multitâche s'intéresse au problème de l'optimisation d'un modèle en fonction d'objectifs multiples. Pour combiner les pertes multi-objectifs, l'approche consisterait simplement à effectuer une somme linéaire pondérée des pertes pour chaque tâche individuelle.

Cette méthode présente toutefois un certain nombre d'inconvénients. La performance du modèle, en particulier, est extrêmement sensible à la sélection des poids. Ces hyperparamètres de poids sont coûteux à régler, ce qui prend souvent plusieurs jours par essai. Par conséquent, il est préférable de trouver une approche plus pratique capable d'apprendre les poids optimaux.

Considérons un réseau qui apprend à partir d'une image d'entrée à prédire la profondeur par pixel et la classe sémantique. Nous constatons qu'à une certaine pondération optimale, le réseau commun est plus performant que les réseaux séparés formés pour chaque tâche séparément. Le réseau est moins performant pour l'une des tâches lorsque les pondérations sont proches de la valeur optimale. Cependant, la recherche de ces pondérations optimales est coûteuse et devient de plus en plus difficile avec des modèles plus importants.

L'incertitude de la tâche saisit la confiance relative entre les tâches, reflétant l'incertitude inhérente à la tâche de régression ou de classification. Elle dépend également de la manière dont la tâche est représentée ou mesurée. Nous proposons que dans un problème d'apprentissage multitâche, nous puissions utiliser l'incertitude homoscédastique comme base pour la pondération des pertes.

IV. EXPÉRIMENTATIONS

A. Ensemble de données

Nous avons utilisé l'ensemble de données NYU-Depth V2, qui se compose de séquences vidéo provenant d'une variété de scènes d'intérieur. Il se compose de 1449 paires densément étiquetées d'images RGB et dept alignées et de 407 024 images étiquetées.

B. L'architecture

Le modèle a été construit à l'aide de PyTorch. L'architecture se compose de deux étapes : La pyramide de mise en commun spatiale et le réseau encodeur-décodeur. La pyramide de mise en commun spatiale capture les images à plusieurs échelles, ce qui est nécessaire pour les tâches de vision par ordinateur. L'architecture du réseau utilise un réseau en U, dans lequel les sorties de la couche de codage sont combinées avec les entrées des couches de décodage par concaténation. Resnet est défini comme l'épine dorsale du modèle. La normalisation par lots avec ReLU est utilisée. La dernière couche est suivie de fonctions d'activation spécifiques à la tâche.

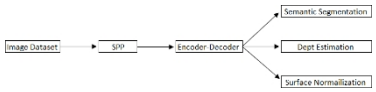

Fig. 1. Architecture du modèle

Le modèle a été entraîné pour 200 époques/itérations. Tout d'abord, le modèle a été entraîné en utilisant une seule tâche, c'est-à-dire Segnet. Ensuite, le modèle a été entraîné à l'aide de l'apprentissage multitâche en utilisant la prédiction de la profondeur et la prédiction de la normale de surface en utilisant des méthodes de pondération comme Nash-MTL et UW (Uncertainty Weighing). Le delta M a été calculé pour comparer le modèle de base (Segnet) à d'autres méthodes multitâches avec différentes approches de pondération.

V. RÉSULTATS

Semantic Segmentation (Baseline)	
MIOU	
Train	Test
0.752	0.751

Fig. 2. Segmentation sémantique (base)

Dans la figure 2, le MIOU pour la segmentation sémantique (Segnet) pour le test et la formation est calculé.

METHODS	Semantic Loss		Mean IOU		Pixel Accuracy	
	Train	Test	Train	Test	Train	Test
Nash-MTL	0.291	0.237	0.8061	0.8044	0.9199	0.9189
UW	0.386	0.3327	0.746	0.748	0.8888	0.8891

Fig. 3. Segmentation sémantique

Dans la Fig(3), la perte sémantique a été évaluée pour les techniques Nash-MTL et UW et Nash-MTL a moins de perte sémantique ou plus de MIOU lorsqu'il est exécuté sur l'ensemble de données de test. La figure 4 montre la perte sémantique par époque.

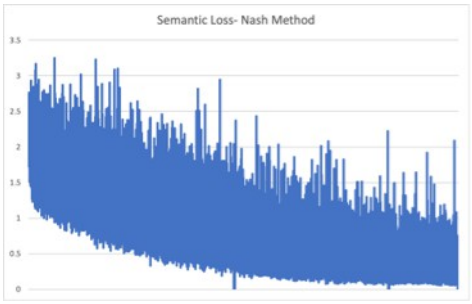

Fig. 4. Perte sémantique : NASH-MTL 200 époques

METHODS	Dept Loss		Absolute Error		Relative Error	
	Train	Test	Train	Test	Train	Test
Nash-MTL	0.2914	0.2724	0.2147	0.2724	0.1092	0.1092
UW	0.2157	0.2854	0.2157	0.2854	0.1094	0.1097

Fig. 5. Estimation du département

Dans la figure (5), la perte de profondeur a été évaluée pour les techniques Nash-MTL et UW et Nash-MTL a une perte de profondeur inférieure lorsqu'il est exécuté sur l'ensemble de données de test. La figure (4) montre la perte de profondeur par époque.

Dans la figure 7, la perte normale a été évaluée pour les techniques Nash-MTL et UW. Nash-MTL a une perte normale inférieure lorsqu'elle est utilisée sur l'ensemble de données de test. La figure (8) montre la perte normale par époque.

Dans la figure 9, le Delta M est utilisé comme mesure pour évaluer les techniques de pondération par rapport à la ligne de base. Nash MTL a un Delta-M inférieur à celui d'UW lorsque Segnet est utilisé comme référence pour la comparaison. La figure 10 illustre la segmentation sémantique et l'estimation de la profondeur d'une image RVB.

VI. Discussions

L'apprentissage multitâche permet de traiter plusieurs tâches en même temps, généralement à l'aide d'un seul réseau neuronal. La méthode MTL présente des avantages tels que l'amélioration de la précision des pixels, l'augmentation de l'IOU moyen et la diminution de la perte d'apprentissage, mais au prix d'une augmentation du temps d'inférence du modèle. En MTL, la méthode de pondération de Nash est plus performante que la plupart des nouvelles méthodes de pondération des tâches publiées. La combinaison de l'estimation de la profondeur et de la segmentation sémantique en tant que tâches peut donner de meilleurs résultats. Dans notre travail, nous nous concentrons principalement sur la segmentation sémantique, la prédiction des normales de surface et la prédiction de la profondeur. Il est possible que nous ayons d'autres combinaisons de tâches qui permettront de remplir plus efficacement les tâches orientées objet.

En outre, les fonctions et les modèles que nous avons utilisés sont les plus adaptés à notre travail, mais pas les meilleurs si l'on considère les problèmes de temps et de fonctionnement. De plus, nous ne disposons pas de notre propre GPU et nous avons donc utilisé Turing pour réaliser notre expérience. Cela a limité l'échelle des données que nous pouvions sélectionner, ce qui peut une influence sur notre précision et nos pertes.

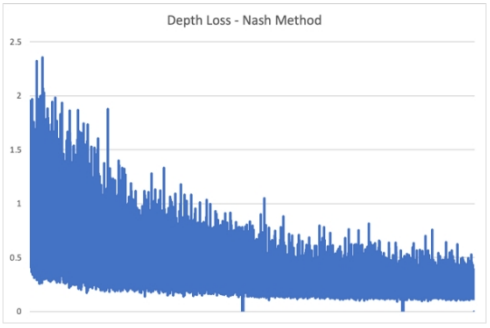

Fig. 6. Perte en profondeur : Nash-MTL 200 époques

METHODS	Normal Loss		Angle Distance				Within t°					
			Median		Mean							
	Train	Test	Train	Test	Train	Test	Train			Test		
Nash-MTL	0.291	0.237	0.8061	0.8044	0.9199	0.9189	0.3574	0.656	0.769	0.3053	0.6141	0.7522
UW	0.386	0.3327	0.746	0.748	0.8888	0.8891	0.405	0.708	0.8117	0.4237	0.7463	0.8527

Fig. 7. Surface normale

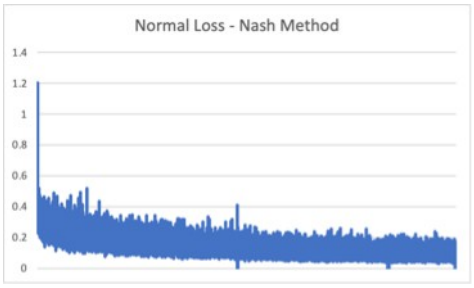

Fig. 8. Perte normale de surface : Nash-MTL 200 époques

METHODS	Delta M
Nash-MTL	-47.571
UW	-32.677

Fig. 9. Delta M

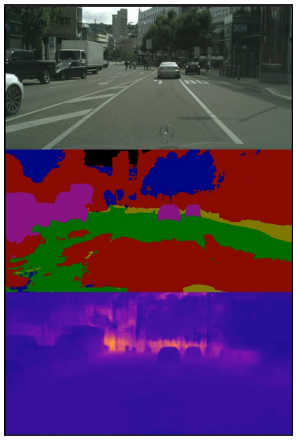

Fig. 10. Visualisation

VII. CONCLUSIONS ET TRAVAUX FUTURS

Nous avons utilisé le paradigme de l'apprentissage multitâche pour stimuler de multiples tâches autosupervisées dans le domaine de la vision par ordinateur. Nous avons appris plusieurs choses : 1) l'apprentissage auto-supervisé est important pour la vision par ordinateur. Sans lui, le coût de l'étiquetage sera énorme. L'apprentissage multitâche, lorsqu'il est utilisé avec la méthode de pondération Nash-MTL, est plus performant que la segmentation sémantique à tâche unique (Segnet).
De meilleurs résultats pourraient être obtenus en utilisant l'augmentation des données. Expérimenter différentes méthodes de pondération et trouver le temps d'inférence optimal. Le travail pourrait être étendu à de nouveaux ensembles de données de segmentation sémantique comme waymo (opendatasetv130 : perception dataset).

RÉFÉRENCES

[1] Navon, et al. Multi-task learning as a Bargaining Game. *arXiv:2202.01017*, 2022.

[2] Doersch, A.Zisserman. Apprentissage visuel multitâche auto-supervisé. *Actes de la conférence internationale de l'IEEE sur la vision par ordinateur*, 2017.

[3] Eigen, R.Fergus. Prédire la profondeur, les normales de surface et les étiquettes sémantiques avec une architecture convolutive multiéchelle commune. *Actes de la conférence internationale de l'IEEE sur la vision par ordinateur*, 2015.

[4] Eigen, C.Puhrsch, R.Fergus. Depth map prediction from a single image using a multi-scale deep network. *Advances in neural information processing systems, 27*. 2014.

[5] Long, E. Shelhamer, T. Darrell. Fully convolutional networks for semantic segmentation (Réseaux entièrement convolutifs pour la segmentation sémantique). *Actes de la conférence IEEE sur la vision informatique et la reconnaissance des formes*. 2015.

[6] Novosel, P.Viswanath, B.Arsenali. Renforcer la segmentation sémantique avec l'apprentissage multitâche auto-supervisé pour les applications de conduite autonome.

[7] Hoyer, et al. Three ways to improve semantic segmentation with self-supervised depth estimation (Trois façons d'améliorer la segmentation sémantique avec l'estimation de profondeur auto-supervisée). *Actes de la conférence IEEE/CVF sur la vision informatique et la reconnaissance des formes*. 2021.

[8] Teichmann, et al. Multinet : Real-time joint semantic reasoning for autonomous driving. *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018.

[9] Thoma. A Survey of Semantic Segmentation. *arXiv : 1602.06541*, 2016.

[10] Pillai, R.Ambrus, A.Gaidon. Superdepth : Estimation de profondeur monoculaire auto-supervisée et super-résolue. *2019 International Conference on Robotics and Automation*. IEEE, 2019.

[11] Badrinarayanan, A.Kendall, R.Cipolla. Segnet : A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495*. 2017

[12] Guizilini, et al. 3d packing for self-supervised monocular depth estimation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.

[13] Nekrasov, T.Dharmasiri, A.Spek, T.Drummond, C.Shen, et I.Reid. Real-Time Joint Semantic Segmentation and Depth Estimation Using Asymmetric Annotations. *arXiv : 1809.04766*, 2019.

[14] Qi, et al. GeoNet : Geometric Neural Network for Joint Depth and Surface Normal Estimation (Réseau de neurones géométriques pour l'estimation conjointe de la profondeur et de la normalité de la surface). *Actes de la conférence internationale de l'IEEE sur la vision informatique et la reconnaissance des formes*, 2018.

[15] Wang, D.Fouhey, A.Gupta. Conception de réseaux profonds pour l'estimation de la normale de surface. *Actes de la conférence de l'IEEE sur la vision informatique et la reconnaissance des formes*, 2015.

[16] Zhang, Q.Yang. A Survey on Multi-Task Learning. *arXiv : 1707.08114*, 2021.