



HAL
open science

Detection, Localisation et Tracking d'Objets basé Deep Learning pour un Fauteuil Roulant Intelligent.

Redouane Khemmar, Louis Lecrosnier, Romain Rossi, Jean Yves Ertaud, France Rouvray, Benoit Decoux, Dupuis Yohan, Nicolas Ragot

► **To cite this version:**

Redouane Khemmar, Louis Lecrosnier, Romain Rossi, Jean Yves Ertaud, France Rouvray, et al.. Detection, Localisation et Tracking d'Objets basé Deep Learning pour un Fauteuil Roulant Intelligent.. Handicap 2020 Technologies pour l'autonomie et l'inclusion, Nov 2020, Paris, France. hal-02989535

HAL Id: hal-02989535

<https://hal.science/hal-02989535v1>

Submitted on 5 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Détection, Localisation et Suivi d'Objets basé Deep Learning pour un Fauteuil Roulant Intelligent

1st Louis Lecrosnier
UNIRouen, Normandy University.
ESIGELEC/IRSEEM
Saint Etienne du Rouvray, France
louis.lecrosnier@esigelec.fr

2nd Redouane Khemmar
UNIRouen, Normandy University.
ESIGELEC/IRSEEM
Saint Etienne du Rouvray, France
redouane.khemmar@esigelec.fr

3rd Nicolas Ragot
UNIRouen, Normandy University.
ESIGELEC/IRSEEM
Saint Etienne du Rouvray, France
nicolas.ragot@esigelec.fr

4th Romain Rossi
UNIRouen, Normandy University.
ESIGELEC/IRSEEM
Saint Etienne du Rouvray, France
romain.rossi@esigelec.fr

5th Jean-Yves Ertaud
UNIRouen, Normandy University.
ESIGELEC/IRSEEM
Saint Etienne du Rouvray, France
Jean-Yves.Ertaud@esigelec.fr

6th Benoit Decoux
UNIRouen, Normandy University.
ESIGELEC/IRSEEM
Saint Etienne du Rouvray, France
benoit.decoux@esigelec.fr

7th Dupuis Yohan
UNIRouen, Normandy University.
ESIGELEC/IRSEEM
Saint Etienne du Rouvray, France
yohan.dupuis@esigelec.fr

Résumé—Dans cet article, nous présentons nos travaux sur le développement d'un système d'aide à la conduite pour un fauteuil roulant électrique intelligent afin d'améliorer l'autonomie de personnes à mobilité réduite. Notre cas d'utilisation, construit à partir d'une étude clinique, repose sur la détection, l'estimation de la profondeur, la localisation et le suivi d'objets de l'environnement intérieur du fauteuil à savoir : portes, poignées de porte, interrupteurs d'éclairage, etc. Cela a pour objectif non seulement d'améliorer la perception autour du fauteuil roulant mais aussi d'asservir ce dernier pour se déplacer, d'une manière semi-autonome, vers ces cibles. Tout d'abord, nous présentons une adaptation à notre cas d'usage de l'algorithme de détection d'objet YOLO V3. Par la suite, nous présenterons notre approche d'estimation de profondeur en utilisant une caméra Intel RealSense. Enfin, comme troisième et dernière étape de notre approche, nous présentons notre approche de suivi d'objets 3D basée sur l'algorithme SORT. Afin de valider l'ensemble des développements de notre approche, nous avons réalisé des expérimentations dans un environnement intérieur contrôlé. La détection, l'estimation de la distance et le suivi des objets sont expérimentés à l'aide de notre propre jeu de données. Ce dernier comprend des couloirs, portes, poignées et interrupteurs. Un des scénarios étudié pour valider la plateforme développée comprend non seulement la détection et le suivi d'objets, mais aussi le déplacement du fauteuil roulant vers l'une de ces cibles désignée comme points d'intérêts.

Mots clés—deep learning, détection et suivi d'objets, estimation de distance, fauteuil roulant électrique, mobilité intelligente.

Ce travail est financé dans le cadre du projet INTERREG VA FMA ADAPT "Assistive Devices for empowering disAbled People through robotic Technologies" <http://adapt-project.com>. Le Programme FMA est un programme de Coopération Territoriale Européenne qui vise à financer des projets de coopération ambitieux dans la région frontalière entre la France et l'Angleterre. Le Programme est financé par le Fonds Européen de Développement Régional (FEDER)

I. INTRODUCTION

La détection, la reconnaissance, la localisation et le suivi d'objets sont des tâches très importantes en robotique et en vision par ordinateur. Ces procédés sont réalisés grâce à l'utilisation de différents moyens de mesure (caméra, LIDAR, RADAR) et d'algorithmes (filtres, reconnaissance de formes, extraction de caractéristiques, classification, etc.). Les travaux présentés dans cet article portent sur la détection, la localisation et le suivi d'objets pour un fauteuil roulant intelligent basé sur la mobilité intelligente. Notre contribution vise à développer la détection d'objets, l'estimation de la profondeur et le suivi de cible du fauteuil roulant en utilisant des approches d'apprentissage profond, ou deep learning. Nous avons réalisé les phases d'apprentissage à l'aide d'un jeu de données créé dans cette optique, puis nous avons validé l'approche complète au sein de la plateforme de navigation autonome de l'ESIGELEC. Cet article est organisé comme suit. L'introduction est présentée dans la section I. Dans la section II, nous présentons un état de l'art de la détection et du suivi d'objets traitant les approches par apprentissage profond. L'architecture du fauteuil roulant intelligent est présentée, quant à elle, dans la section III. Notre approche liée à la détection, de l'estimation de profondeur et du suivi (ou tracking) d'objets est ensuite détaillée en section IV. La section V conclura l'article.

II. TRAVAUX ASSOCIÉS

A. Détection d'Objets

Les méthodes basées sur l'apprentissage profond qui offrent les meilleures performances pour la détection d'objets (comme

les algorithmes Fast R-CNN [1], Faster R-CNN [2] et Mask R-CNN [3]) se décomposent en deux modules présents dans un réseau de neurones convolutif (CNN) [5] (ensemble de cellules appelées neurones qui, travaillant ensemble, permettent indépendamment des prédictions à partir des entrées du réseau). Le premier module renvoie les coordonnées des régions d'intérêt (Region of Interest (ROI)) où un objet (quelle que soit sa nature) serait présent dans l'image. C'est la proposition de ROI. Le deuxième module est dédié à la détection. Il fournit une prédiction de classe de la région proposée.

L'algorithme YOLOv3 (You Only Look Once) [4] est l'un des algorithmes de détection d'objets en apprentissage profond les plus performants. YOLO est basé sur le principe de régression au lieu de classification. Cela signifie que tout le processus de localisation d'une ROI englobant un objet et de classification de cet objet se fait en une seule fois. Cela permet une détection d'objet en temps réel sur un GPU standard. Comme YOLO, SSD [6] combine des méthodes de classification et de localisation des ROI, évitant ainsi le rééchantillonnage fastidieux des pixels et des caractéristiques extraites de l'image et ce pour chaque boîte englobante. Pour effectuer l'extraction des caractéristiques de l'image, SSD s'appuie sur l'architecture VGG-16 [7] pour les premières couches du réseau. Les couches suivantes sont progressivement réduites en taille pour permettre une détection multi-échelle (contrairement à YOLO). Chacune de ces couches produit un ensemble de prédictions de détection.

B. Base de données

Pour la détection d'objets, de nombreux jeux de données sont disponibles librement. Nous pouvons en citer quelques-uns : ImageNet [15] avec 14.000.000 d'instances par image. Il comprend 1000 classes [16] mais l'accessibilité de ce jeu de données reste limitée. SIFT10M [17] qui comprend 11 Millions d'images annotées. Il est structuré sous la forme de points de référence d'images connues (méthode SIFT). Open Images [18], qui est un jeu de données communautaire avec plus de 9000000 d'images. Les objets sont étiquetés avec des ROI. Il comprend 7881 classes différentes [19]. Les classes sont très variées et en constante expansion. COCO [20] avec 1.500.000 d'image. Les objets sont étiquetés et catégorisés. Il comprend 80 classes d'objets [21]. COCO-stuff [22] est une version étendue de COCO avec plus de 1800000 d'image. Il comprend 181 classes [23]. PASCAL VOC [26] est dédié à la détection des piétons avec 500000 d'images étiquetées. Elle comprend des classes [27] de 20 objets différents. CIFAR-100 [28] avec des images à 60000. Elle comprend des classes étiquetées à 100 [29] objets. INRIA [30] est une base de données dédiées à la détection de piétons. Caltech [31] est également dédié à la détection de personnes et de piétons. Dans l'ensemble, ces jeu de données sont relativement riches concernant les objets les plus couramment détectés (piétons, voitures, bus, vélos, etc.). Le jeu de données COCO est également la référence en termes de détection d'objets et est idéal pour comparer différents modèles de détection d'objets. Ce dernier propose des bases de données différentes adaptées

aux phases d'entraînement et à l'inférence des réseaux de neurones. Yolov3, par exemple, utilise ImageNet pour former les 53 premières couches de son réseau et établir les bases de données de référence. Cet jeu de données étant très dense, c'était un choix judicieux pour préparer l'entraînement de Yolov3. Ce modèle utilise ensuite d'autres bases de données pour effectuer la détection et la classification, comme COCO par exemple.

III. ARCHITECTURE DE LA PLATEFORME BASÉE FAUTEUIL ROULANT ELECTRIQUE

Le fauteuil roulant électrique robotique du laboratoire IR-SEEM est un Invacare, modèle Bora, dont toute l'électronique propriétaire a été retirée et remplacée par plusieurs éléments : 1. Un PC embarqué sous Linux Ubuntu 16.04 LTS, 2. Un pilote moteur de Roboteq, 3. Un joystick Xbox à la place de l'original qui peut être connecté au fauteuil roulant par une connexion USB ou Bluetooth, 4. Un routeur WIFI pour obtenir un réseau local sans fil sur le fauteuil roulant, 5. Une IHM embarquée avec un écran tactile. Le logiciel du fauteuil roulant est entièrement développé sous le middleware robotique ROS. La description du fauteuil roulant est illustrée dans la figure 1. Un ordinateur embarqué est utilisé, avec Ubuntu 16.04 comme système d'exploitation et 8 Go de RAM et 250 Go de SSD, pour l'ensemble du traitement embarqué. L'interaction entre le fauteuil roulant et l'utilisateur s'effectue par le biais d'un écran tactile qui est connecté à l'ordinateur embarqué par une connexion HDMI. Un pilote de moteur de Roboteq est utilisé pour piloter les moteurs. Le fauteuil roulant est connecté et dispose de son propre routeur Wi-Fi. Un joystick Xbox One est utilisé pour contrôler manuellement le fauteuil. Une caméra Intel RealSense D435 est utilisée pour la vision embarquée, et fournit des images couleur et de profondeur .

IV. DÉTECTION, LOCALISATION ET SUIVI D'OBJETS PAR DEEP LEARNING

A. Détection d'Objets

Pour effectuer la détection d'objets, nous nous sommes appuyés sur le puissant réseau de neurones YOLOv3. Dans le cadre du projet ADAPT, la détection et la classification des éléments spécifiques à l'environnement intérieur (tels que les portes, les poignées de porte et les interrupteurs) est un point essentiel. Comme ces classes sont sous-représentées dans le jeu de données d'entraînement de YOLOv3 (c'est-à-dire ImageNet [24]), nous avons composé un jeu de données personnalisé que nous avons utilisé pour effectuer l'entraînement du réseau. Nous avons extrait 755 images de portes du jeu de données MCIndoor20000 [25], composé d'images étiquetées contenant divers objets d'intérieurs. Nous n'avons pas pu trouver un jeu de données ouvert avec une représentation suffisante des interrupteurs, des poignées de porte segmentés et étiquetés. Pour cette raison, nous avons développé notre propre base de données au sein de l'ESIGELEC constituée de 1885 images, que nous avons combiné avec les images de portes de la base de données MCIndoor20000. Nous avons supervisé l'étiquetage de 2640 images provenant de deux ensembles de données

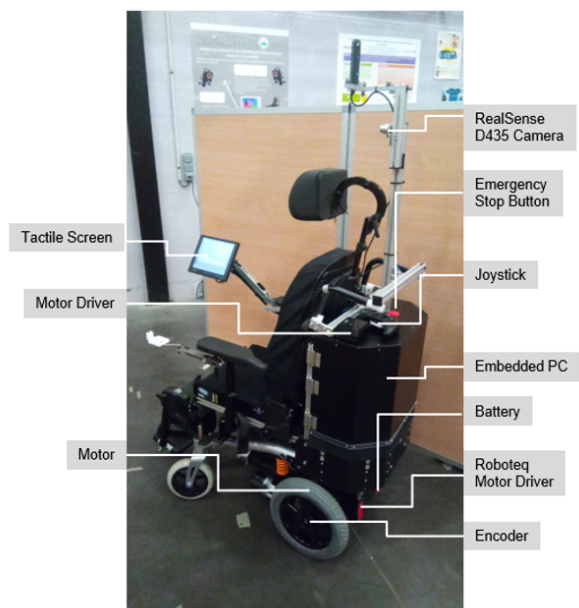


FIGURE 1: Fauteuil roulant électrique et robotique de l'IR-SEEM.

combinés en utilisant un outil d'étiquetage semi-automatique que nous avons développé. Nous avons finalement procédé au re-entraînement de YOLOv3 sur les classes requises. Pour ce processus d'apprentissage par transfert, nous n'avons entraîné que les couches de classification du réseau neuronal. La figure 2 montre les résultats qualitatifs du processus de détection après avoir re-entraîné YOLOv3 sur la reconnaissance de portes, des poignées de porte et des interrupteurs d'éclairage.

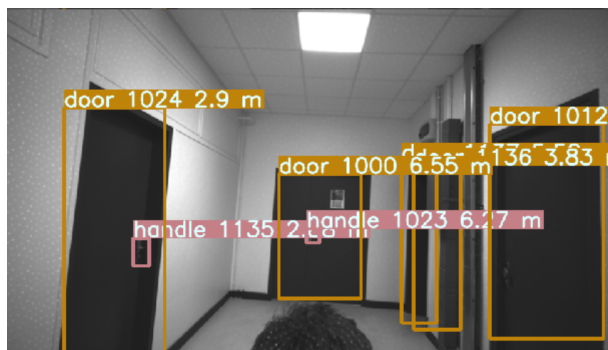


FIGURE 2: Détection d'objets en environnement Indoor du fauteuil roulant comprenant portes et poignées de portes.

B. Estimation de profondeur

Une fois l'étape de détection réalisée, une distance est associée à chaque élément extrait. La caméra Realsense *D435* peut fournir une image de profondeur avec une précision de l'ordre de 10% pour de courtes distances (0, 10 – 20m). Néanmoins, on observe une différence de champ de vue entre la paire de caméras infrarouges stéréoscopiques fournissant les images de profondeur, et le capteur couleur. Il est donc

nécessaire de mettre en correspondance les images couleurs et infrarouge. Celle-ci est réalisée directement à l'aide des données constructeur. A l'issue de cette étape, on extrait les valeurs de distance pour chaque pixel des ROI retournées par YOLOv3. On applique ensuite un filtre médian sur ces valeurs de manière à écarter les zones où l'information de profondeur peut être incomplète ou erronée. La figure 2 illustre le résultat du processus d'extraction de ROI et l'attribution d'une information de profondeur associée.

C. Tracking d'Objets

En utilisant la combinaison de position et profondeur associées à chaque élément détecté par l'algorithme de détection, nous nous appuyons sur l'algorithme SORT (Simple Online and Realtime Tracking) [32] [33] pour suivre les différents objets de la scène lorsque le fauteuil roulant est en mouvement. Sur un flux vidéo, plusieurs objets (portes, interrupteurs et poignées de porte) sont détectés et se voient attribuer une distance donnée. SORT inspecte les objets détectés, et détermine si un objet donné est nouvellement vu, ou si le déplacement de l'objet est une conséquence des mouvements du fauteuil roulant. Cet algorithme basé sur un filtre de Kalman fournit finalement un numéro d'identification unique à chaque nouvel objet détecté. SORT garde une trace de plusieurs objets simultanément et filtre les positions d'objets bruyants associés aux boîtes de délimitation en mouvement.

Enfin, nous utilisons les données odométriques provenant de la caméra Realsense *T265* pour estimer le déplacement du fauteuil roulant. Nous combinons ces informations avec la position de l'objet pour visualiser une carte sémantique 3D de l'environnement contenant les objets détectés et suivis, comme le montre la figure 3.

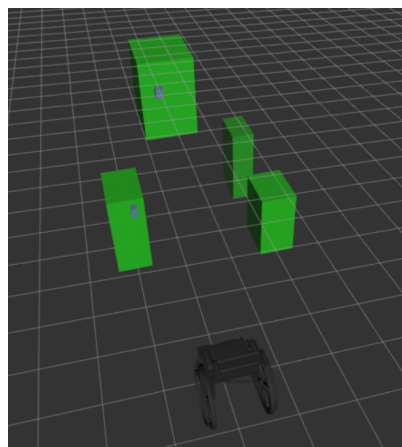


FIGURE 3: 3D local semantic map

V. VALIDATION EXPÉRIMENTALE

Dans l'optique de valider ces algorithmes, nous avons réalisé un jeu de données expérimental au sein du laboratoire IRSEEM. Un ensemble quatre portes comportant couleurs et formes de poignées différentes ont été placées dans la zone de couverture d'un système de capture de mouvement. Une

fois équipées de marqueurs réfléchissants, portes et fauteuil roulant sont localisés par un système Vicon fournissant leur position et orientation avec une précision millimétrique à une fréquence de 100Hz [34]. De cette manière, on obtient une vérité terrain permettant de calculer à haute vitesse la distance entre le fauteuil roulant et les portes.

Notre ensemble de porte a été placé suivant un arc de cercle (cf. Fig. 4), avec pour optique d'évaluer la capacité de détection lorsque de multiples éléments sont présents simultanément sur une image. Lors de cette expérience, le fauteuil roulant d'ADAPT se déplace dans la scène et, en changeant d'orientation, se retrouve à plusieurs reprises en vis à vis des différentes portes. On évalue alors l'erreur entre la distance estimée à l'issue d'une détection par YOIOv3, et la vérité terrain fournie par le système vicon.

La figure 5 présente ces résultats. Pour un total de 650 détections de portes, on observe la distance réelle et la distance estimée par la caméra Realsense D435, ainsi que la différence de ces deux valeurs. La table I fournit les résultats numériques de cette expérimentation. On y note une erreur moyenne de 18.1 cm, soit 3.8% d'erreur sur l'estimation de la distance des objets. Ces valeurs sont en deçà des données fournies par le constructeur.

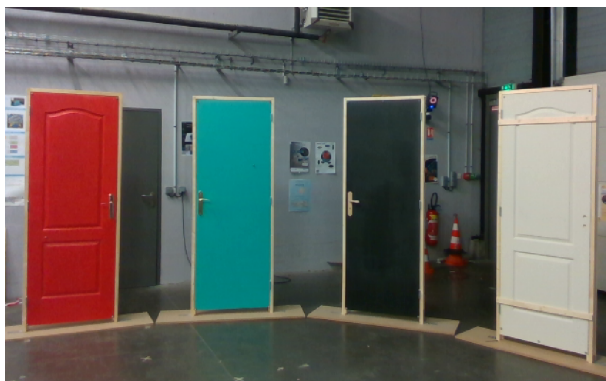


FIGURE 4: Disposition des portes dans le jeu de données de validation

CONCLUSION

Nous avons présenté dans ce document la détection, l'estimation de la profondeur, la localisation et le suivi des objets

Erreurs d'estimation de profondeur	
Médiane (cm)	15.6
Moyenne (cm)	18.1
Ecart type (cm)	13.5
Médiane (%)	3.2
Moyenne (%)	3.8
Ecart type (%)	2.6

TABLE I: Evaluation de l'estimation de profondeur sur 650 détections de portes. Ecart entre distance estimée par la caméra et vérité terrain. Erreur absolue en cm, reportée en pourcentage. Valeurs moyennes, médianes et écart type.

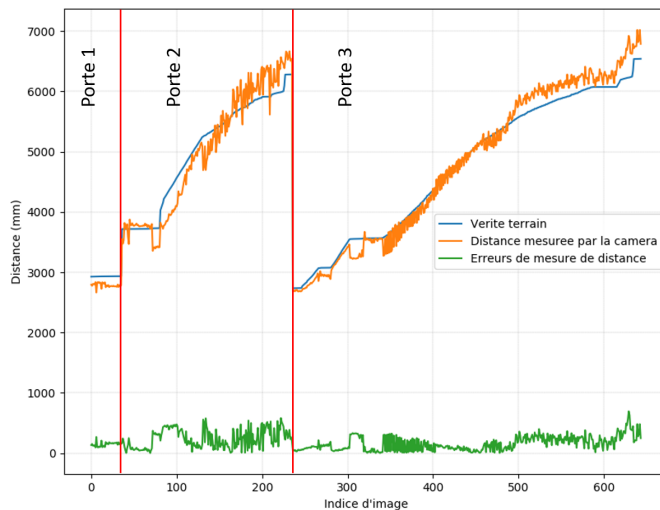


FIGURE 5: Résultats quantitatifs de l'estimation de profondeur par la caméra Realsense D435. Distance mesurée entre porte et caméra d'après les images de profondeurs (orange), Vérité terrain (bleu), et écart entre vérité terrain et mesure caméra (vert)

sur la base d'un apprentissage approfondi pour une mobilité intelligente des fauteuils roulants. La détection d'objets est basée sur l'approche YOLO V3. Nous avons mesuré l'erreur d'estimation de distance avec les objets détectés. Enfin, nous avons amélioré une version de l'algorithme SORT afin d'effectuer le suivi des objets. L'estimation de la position de l'objet est améliorée grâce à l'utilisation du filtre de Kalman étendu. Pour valider l'ensemble des développements, nos modèles ont été réentraînés en utilisant une composition de données ouvertes et internes : *MCIndoor20000* et *ESIGELEC*. La détection et le suivi d'objets ont été évalués sous le jeu de données *ESIGELEC* pour valider la détection et le suivi d'objets pour un fauteuil roulant en environnement intérieur. En utilisant notre modèle reentraîné sur notre jeu de données personnalisé, notre implémentation offre de bonnes performances en matière d'inférence sur l'environnement intérieur du fauteuil roulant. L'ensemble des développements est intégré sur la plateforme via une carte Nvidia Jetson TX2 comme calculateur principal pour tous les algorithmes d'apprentissage profond (détection, estimation et suivi d'objets).

REMERCIEMENTS

Le travail présenté dans ce document est soutenu par le projet ADAPT qui est réalisé dans le cadre du projet INTERREG VA FMA ADAPT "Assistive Devices for empowering disAbled People through robotic Technologies" <http://adapt-project.com/index.php>. Le programme Interreg FCE est un programme de coopération territoriale européenne qui vise à financer des projets de coopération de haute qualité dans la région frontalière de la Manche entre la France et l'Angleterre. Le programme est financé par le Fonds européen de développement régional (FEDER)). Nous tenons à remercier les ingénieurs du Laboratoire de Navigation Autonome

(LNA) de l'IRSEEM pour leur soutien et le Service Ingénierie Recherche et Développement (SIRD) de l'ESIGELEC pour leur aide dans la phase de test. Ce travail a bénéficié des moyens de calcul du mésocentre CRIANN (Centre Régional Informatique et d'Applications Numériques de Normandie).

RÉFÉRENCES

- [1] GIRSHICK, Ross. Fast r-cnn. In : Proceedings of the IEEE international conference on computer vision. 2015. p. 1440-1448.
- [2] REN, Shaoqing, HE, Kaiming, GIRSHICK, Ross, et al. Faster r-cnn : Towards real-time object detection with region proposal networks. In : Advances in neural information processing systems. 2015. p. 91-99.
- [3] HE, Kaiming, GKIOXARI, Georgia, DOLLÁR, Piotr, et al. Mask r-cnn. In : Proceedings of the IEEE ICCV 2017. p. 2961-2969.
- [4] Redmon, Joseph and Farhadi, Aliar, YOLOv3 : An Incremental Improvement 2018, arXiv
- [5] PELLEGRINI, Thomas, FONTAN, Lionel, et SAHRAOUI, Halima. Réseau de neurones convolutif pour l'évaluation automatique de la prononciation.
- [6] LIU, Wei, ANGUELOV, Dragomir, ERHAN, Dumitru, et al. Ssd : Single shot multibox detector. In ECCV 2016. p. 21-37.
- [7] HASSAN, M. VGG16 : Convolutional Network for Classification and Detection. 2019.
- [8] GODARD, Clément, MAC AODHA, Oisín, et BROSTOW, Gabriel J. Unsupervised monocular depth estimation with left-right consistency. In IEEE CVPR 2017. p. 270-279.
- [9] ZHOU, Tinghui, BROWN, Matthew, SNAVELY, Noah, et al. Unsupervised learning of depth and ego-motion from video. In IEEE CVPR 2017. p. 1851-1858.
- [10] EIGEN, David, PUHRSCH, Christian, et FERGUS, Rob. Depth map prediction from a single image using a multi-scale deep network. In : Advances in neural information processing systems. 2014. p. 2366-2374.
- [11] LIU, Fayao, SHEN, Chunhua, LIN, Guosheng, et al. Learning depth from single monocular images using deep convolutional neural fields. IEEE transactions on pattern analysis and machine intelligence, 2015, vol. 38, no 10, p. 2024-2039.
- [12] MAYER, Nikolaus, ILG, Eddy, HAUSSER, Philip, et al. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In : Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. p. 4040-4048.
- [13] TONIONI, Alessio, TOSI, Fabio, POGGI, Matteo, et al. Real-time self-adaptive deep stereo. In IEEE CVPR 2019. p. 195-204.
- [14] MAYER, Nikolaus, ILG, Eddy, HAUSSER, Philip, et al. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In IEEE CVPR 2016. p. 4040-4048.
- [15] DENG, Jia, DONG, Wei, SOCHER, Richard, et al. Imagenet : A large-scale hierarchical image database. In IEEE CVPR 2009. p. 248-255.
- [16] JEONG, Hyeok-June, PARK, Kyeong-Sik, et HA, Young-Guk. Image preprocessing for efficient training of YOLO deep learning networks. In : 2018 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2018. p. 635-637.
- [17] YANG, Jie, ZHAO, Wan-Lei, DENG, Cheng-Hao, et al. Fast Nearest Neighbor Search Based on Approximate k-NN Graph. In : International Conference on Internet Multimedia Computing and Service. Springer, Singapore, 2017. p. 327-338.
- [18] KUZNETSOVA, Alina, ROM, Hassan, ALLDRIN, Neil, et al. The open images dataset v4 : Unified image classification, object detection, and visual relationship detection at scale. arXiv preprint arXiv :1811.00982, 2018.
- [19] DOYLE, Scott, FELDMAN, Michael D., SHIH, Natalie, et al. Cascaded discrimination of normal, abnormal, and confounder classes in histopathology : Gleason grading of prostate cancer. BMC bioinformatics, 2012, vol. 13, no 1, p. 282.
- [20] LIN, Tsung-Yi, MAIRE, Michael, BELONGIE, Serge, et al. Microsoft coco : Common objects in context. In : European conference on computer vision. Springer, Cham, 2014. p. 740-755.
- [21] CHEN, Xinlei, FANG, Hao, LIN, Tsung-Yi, et al. Microsoft coco captions : Data collection and evaluation server. arXiv preprint arXiv :1504.00325, 2015.
- [22] CAESAR, Holger, UIJLINGS, Jasper, et FERRARI, Vittorio. Cocomstuff : Thing and stuff classes in context. In : Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. p. 1209-1218.
- [23] LI, Jie, RAVENTOS, Allan, BHARGAVA, Arjun, et al. Learning to fuse things and stuff. arXiv preprint arXiv :1812.01192, 2018.
- [24] DENG, Jia, DONG, Wei, SOCHER, Richard, et al. Imagenet : A large-scale hierarchical image database. In : 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009. p. 248-255.
- [25] Bashiri, Fereshteh S., et al. "MCIndoor20000 : A fully-labeled image dataset to advance indoor objects detection." Data in brief 17 (2018) : 71-75.
- [26] VICENTE, Sara, CARREIRA, Joao, AGAPITO, Lourdes, et al. Reconstructing pascal voc. In : Proceedings of the IEEE conference on computer vision and pattern recognition. 2014. p. 41-48.
- [27] ZHU, Chao, BICHOT, Charles-Edmond, et CHEN, Liming. Multi-scale color local binary patterns for visual object classes recognition. In : 2010 IEEE CVPR, p. 3065-3068.
- [28] KRIZHEVSKY, Alex, NAIR, Vinod, et HINTON, Geoffrey. Cifar-10 and cifar-100 datasets. URL : <https://www.cs.toronto.edu/kriz/cifar.html>, 2009, vol. 6.
- [29] CUI, Yin, JIA, Menglin, LIN, Tsung-Yi, et al. Class-balanced loss based on effective number of samples. In : Proceedings of the IEEE CVPR. 2019. p. 9268-9277.
- [30] EVERINGHAM, Mark, VAN GOOL, Luc, WILLIAMS, Christopher KI, et al. The PASCAL visual object classes challenge 2007 (VOC2007) results. 2007.
- [31] GRIFFIN, Gregory, HOLUB, Alex, et PERONA, Pietro. Caltech-256 object category dataset. 2007.
- [32] BEWLEY, Alex, GE, Zongyuan, OTT, Lionel, et al. Simple online and realtime tracking. In : 2016 IEEE ICIP p. 3464-3468.
- [33] WOJKE, Nicolai, BEWLEY, Alex, et PAULUS, Dietrich. Simple online and realtime tracking with a deep association metric. In : 2017 IEEE ICIP, p. 3645-3649.
- [34] Merriaux, P., Dupuis, Y., Bouteau, R., Vasseur, P., Savatier, X. (2017). A study of vicon system positioning performance. Sensors, 17(7), 1591.