

Apprentissage de profondeur par flou de défocalisation: réseaux de neurones pour l'estimation de la profondeur mono-image

Soutenance de thèse – Marcela CARVALHO

25 novembre 2019, 14h

Université Paris Descartes, Paris

45 rue des Saints-Pères, 7ème étage, Salle 749,

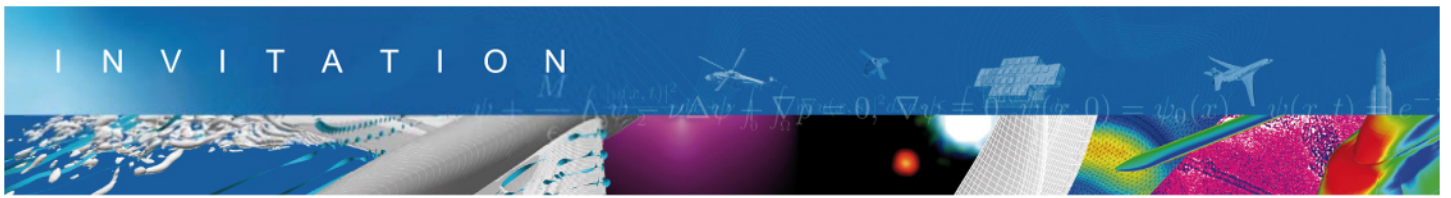


Devant le jury composé de :

- Christine GUILLEMOT (rapporteuse),
- Thierry CHATEAU (rapporteur),
- Patrick PEREZ (examineur),
- Gabriele FACCIOLO (examineur),
- David PICARD (examineur),
- Frédéric CHAMPAGNAT (directeur).
- Andrés ALMANSA (co-directeur),
- Bertrand LE SAUX (co-encadrant),
- Pauline TROUVÉ-PELOUX (co-encadrante).

Résumé :

L'estimation de profondeur à partir d'une seule image est maintenant cruciale pour plusieurs applications, de la robotique à la réalité virtuelle. Les approches par apprentissage profond dans les tâches de vision par ordinateur telles que la reconnaissance et la classification d'objets ont également apporté des améliorations au domaine de l'estimation de profondeur. Dans cette thèse, nous développons des méthodes pour l'estimation de profondeur avec un réseau de neurones profond en explorant différents indices, tels que le flou de défocalisation et la sémantique. Nous menons également plusieurs expériences pour comprendre la contribution de chaque indice à la performance du modèle et sa capacité de généralisation. Dans un premier temps, nous proposons un réseau de neurones convolutif efficace pour l'estimation de la profondeur ainsi qu'une stratégie d'entraînement basée sur les réseaux génératifs adversaires conditionnels. Notre méthode permet d'obtenir des performances



parmi les meilleures sur les jeux de données de référence. Ensuite, nous proposons d’explorer le flou de défocalisation, une information optique fondamentalement liée à la profondeur. Nous montrons que ces modèles sont capables d’apprendre et d’utiliser implicitement cette information pour améliorer les performances et dépasser les limitations connues des approches classiques d’estimation de la profondeur par flou de défocalisation. Nous construisons également une nouvelle base de données avec des images réelles focalisées et défocalisées que nous utilisons pour valider notre approche. Enfin, nous explorons l’utilisation de l’information sémantique, qui apporte une information contextuelle riche, en apprenant à la prédire conjointement avec la profondeur par une approche multi-tâche. Nous démontrons la validité des concepts proposés avec plusieurs jeux de données contenant des images en intérieur, en extérieur et aériennes.

Abstract:

Depth estimation from a single image is a key instrument for several applications from robotics to virtual reality. Successful Deep Learning approaches in computer vision tasks as object recognition and classification also benefited the domain of depth estimation. In this thesis, we develop methods for monocular depth estimation with deep neural network by exploring different cues: defocus blur and semantics. We conduct several experiments to understand the contribution of each cue in terms of generalization and model performance. At first, we propose an efficient convolutional neural network for depth estimation along with a conditional Generative Adversarial framework. Our method achieves performances among the best on standard datasets for depth estimation. Then, we propose to explore defocus blur cues, which is an optical information deeply related to depth. We show that deep models are able to implicitly learn and use this information to improve performance and overcome known limitations of classical Depth-from-Defocus. We also build a new dataset with real focused and defocused images that we use to validate our approach. Finally, we explore the use of semantic information, which brings rich contextual information while learned jointly to depth on a multi-task approach. We validate our approaches with several datasets containing indoor, outdoor and aerial images.

Mots clés :

Vision par ordinateur, Apprentissage profond, Robotique, Télédétection;
Estimation de profondeur, Réseaux de neurones, Flou de défocalisation, Classification sémantique, Apprentissage multi-tâche.