

## PROPOSITION DE SUJET DE THÈSE

### Intitulé : Deep neural networks for 3D prediction in the wild

Référence : TIS-DTIS-2019-xxx  
(à rappeler dans toute correspondance)

**Laboratoire d'accueil à l'ONERA :**

Domaine : iTraitement de l'Information et Systèmes      Lieu (centre ONERA) : Palaiseau

Département : Traitement de l'Information et Systèmes

Unité : Image Vision Apprentissage      Tél. +33 1 80 38 65 73

Responsable ONERA :      Email : pauline.trouve@onera.fr

Pauline Trouvé-Peloux, Bertrand Le Saux      bertrand.le\_saux@onera.fr

**Directeur de thèse envisagé :**

Nom : Frédéric Champagnat et Bertrand Le Saux (Codirecteurs)

Adresse : ONERA

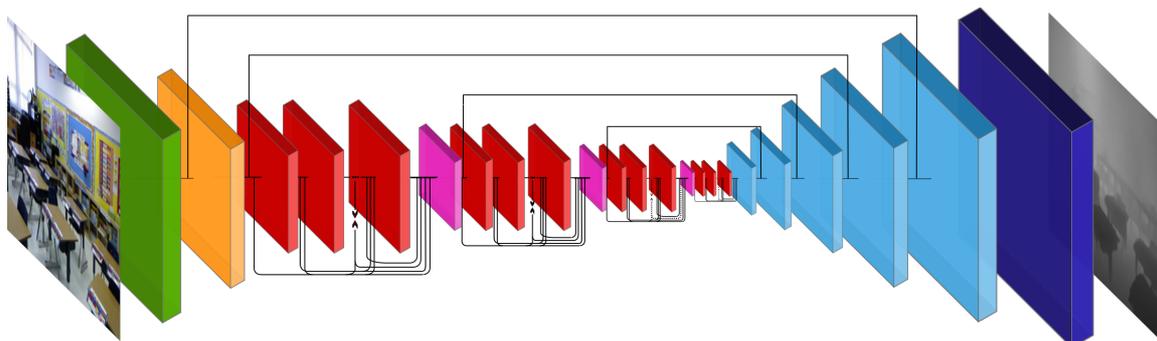
Tél. :      Email : frederic.champagnat@onera.fr

Sujet :

**Objective**

3D estimation is crucial for scene understanding (autonomous driving...) and accurate 3D reconstruction (3D mapping, robotics...). However, open environments, outdoors or indoors ("in the wild") abound with challenging situations which require robust and efficient systems. Recently, 3D from a single image has reached great performances thanks to deep neural networks. This might be a data acquisition paradigm shift, away from stereo vision and active laser scanners. We at ONERA/DTIS developed D3-Net [Carvalho et al., 2018a-b], one of the top state-of-the-art approaches to depth estimation by deep learning, awarded by a Best Paper Award at RFIAP2018, the French machine learning conference.

The objective of the thesis project is two-fold. First, to push further this research work by developing convolutional neural networks (CNNs) for *directly estimating 3D point clouds*, instead of depth rasters. Indeed, 3D point clouds are the standard in 3D data acquisition with laser and photogrammetry, and hence in 3D perception. Second, to co-design a smart system able to predict depth in the wild by combining specific camera optics with the adequate deep learning model.



*D3Net, take from [Carvalho et al., 2018a]*

In details, 3D point-cloud prediction from a single image in the wild raises several problems

### **Problem #1: How to build convolutional network models for prediction of 3D real points ?**

While 2D processing with convolutional neural networks (CNN) has now reached practical solutions, handling 3D data is still an open issue, due to the irregular and sparse nature of the 3D points. Hence the need for developing new CNN models able to predict 3D real points. A particular care will be given to the design of the loss function, as the same geometry may admit different point cloud representations. Solutions based on Optimal Transport will be investigated [Fan et al., 2017].

### **Problem #2: How to predict accurate depth in the wild, i.e. challenging environments?**

In many real-life scenarios, input data might be corrupted while stakes of the prediction are high. Thus, it is crucial to model and correct the image capture defaults. These defaults are well-identified. For example, images captured in the wild are often blurry due to motion and distorted due to the rolling shutter acquisition mode of the camera. How to correct those images and be insensitive to these defaults while predicting 3D? Specific architecture networks able to perform multi-task learning will be investigated.

At the other end of the process, it is also decisive to predict only accurate and reliable 3D points. Various approaches will be compared, including simultaneous Bayesian uncertainty prediction as in [Carvalho et al., 2018a], [Kendall & Gal, 2017], and understanding of image hints which allow to estimate 3D: defocus, edges, scene statistics. Moreover, when image sequences are available, as often in robotics, using several images taken at short intervals will allow to select the most reliable estimates.

### **Problem #3: Unconventional optics for an absolute depth prediction**

Deep learning models provide high quality predictions but there are based on the dataset which was used for training. Thus, to improve the network ability to generalize, it is important to have a reliable prediction based on measure, that can be provided by a sensor which adds physically-based depth cue information on the image. For instance, the impact of defocus blur on networks for depth prediction has been evaluated in [Carvalho et al., 2018a], and improvement of the network generalization ability using defocused images has been demonstrated. In future works, several optical tricks from computational imaging, which reinforce the variation of the defocus blur with depth, will be explored: chromatic or astigmatic aberration, or coded aperture [Trouvé-Peloux et al. 2018].

As defocus is tightly related to the sensor parameters (focus, focal length, aperture etc), they influence the depth estimation performance. Then what are the best sensor settings for depth estimation using neural networks ? More generally how can the sensor parameters be included in a neural network optimization framework ? These questions, investigated in this thesis, will be the first steps towards a co-design of optical and processing using neural networks.

### **Workplan and practical information**

The research work plan will consists in: design and development deep learning models for 3D point cloud prediction from a single image; coding (python) and experiments with CNNs using open libraries such as Pytorch or Tensorflow; application to robotics and design of a smart sensor prototype by co-design of the camera optics and the associated processing. This thesis will take place at the ONERA center in Palaiseau (near Paris).

### **References:**

[Carvalho et al., 2018b] On Regression Losses for Deep Depth Estimation *M. Pinheiro de Carvalho, B. Le Saux, P. Trouvé-Peloux, F. Champagnat, A. Almansa* IEEE Int. Conf. on Image Processing (ICIP 2018), Athens, Greece, October 2018.

[Carvalho et al., 2018a] Deep Depth from Defocus: how can defocus blur improve 3D estimation using

dense neural networks? *M. Pinheiro de Carvalho, B. Le Saux, P. Trouvé-Peloux, F. Champagnat, A. Almansa* ECCV / Workshop on 3D Reconstruction in the Wild, Munich, Germany, September 2018.

[**Trouvé-Peloux et al., 2018**]. Turning a conventional camera into a 3D camera with an add-on, *P. Trouvé-Peloux, J. Sabater, A. Bernard-Brunel, F. Champagnat, G. Le Besnerais, T. Avignon*, Applied Optics, **57**, 10, 2018.

[**Fan et al., 2017**] A Point Set Generation Network for 3D Object Reconstruction from a Single Image. *Haoqiang Fan, Hao Su, Leonidas J. Guibas* CVPR 2017, Hawaii, USA, July 2017.

[**Kendall & Gal, 2017**] What uncertainties do we need in Bayesian deep learning for computer vision ? NIPS 2017, Long Beach, Cal., USA, December 2017.

**Collaborations extérieures :**

### **PROFIL DU CANDIDAT**

**Formation :** Ms. Eng. (CS, EE, ...), M.Sc.

French Grandes Écoles, Master 2 recherche learning / computer vision

**Spécificités souhaitées :** Machine Learning, Deep Learning, Image Processing., Optics  
Programming experience (python, etc.)

**ANNEXE À LA PROPOSITION DE SUJET DE THÈSE**

Les rubriques suivantes doivent être dûment renseignées :

**1. Titre de la thèse**

**Deep neural networks for 3D prediction in the wild**

**2. Domaine et thématique scientifique principale**

Perception et Traitement de l'Information / Intelligence Artificielle et Décision / Robotique

**3. Contexte de l'étude (en 1 à 2 pages)**

- a. à l'ONERA (préciser notamment les personnes participant à l'encadrement en plus du responsable ONERA)

F. Champagnat, B. Le Saux, P. Trouvé-Peloux

Les travaux s'intègrent dans le cadre du projet Delta, notamment le lot 1 pour la partie théorique de conception de réseaux de neurones et le lot 2 pour le cadre des applications robotiques.

Par ailleurs, ces travaux continuent et utilisent les résultats de la thèse de Marcela Carvalho (soutenance attendue automne 2019)

- b. à l'extérieur

A. Almansa (Univ. Paris Descartes)

- c. bibliographie succincte

cf. sujet.

**4. Description des travaux (en 1 à 2 pages)**

- a. plan de thèse prévisionnel

La première année de thèse sera consacrée dans un premier temps à une étude bibliographique concernant les réseaux pour la prédiction 3D de manière générale, puis plus précisément sur ceux qui traitent de données non denses, sous forme de nuages de points. Pour cela le doctorant pourra s'appuyer sur le manuscrit de thèse de Marcela Carvalho, ainsi que sur les travaux réalisés dans le cadre du stage 2019, sur cette même thématique, si celui-ci a été pourvu. Le doctorant sera également sensibilisé par ses encadrants et ses propres recherches, aux dégradations possibles des images (flou de bougé, rolling shutter) ainsi qu'au domaine du *computational imaging* dans lequel des optiques non conventionnelles sont utilisées pour accroître les capacités des imageurs. Les premières approches de l'état de l'art concernant la modélisation de capteur dans un réseau de neurones seront également investiguées.

Puis, le doctorant s'intéressera en particulier à la thématique de l'estimation non dense de la 3D à partir d'images et sera amené à développer ses propres entraînements et tests sur les jeux de données sans flou de l'état de l'art (NYUV2, DiW, etc). L'étude des différentes fonctions de coûts possibles et l'évaluation du gain de cette approche par rapport aux méthodes denses pourront conduire à une publication dans un congrès international.

A la fin de la première année et au début de sa deuxième année, le doctorant pourra s'intéresser à des données "in the wild" en développant des jeux de données spécifiques, notamment en robotique, présentant du bruit, du flou de bougé des dégradations liées au rolling shutter. Le traitement temporel de ces données est une piste qui sera exploitée par le doctorant, ainsi que les approches de réseaux de neurones multi-tâches qui viseront à estimer conjointement la 3D et à restaurer des images de bonne qualité par exemple par correction géométrique grâce à l'estimation du mouvement du porteur, déconvolution et/ou

supér-résolution temporelle. Des comparaisons entre les résultats de approches par apprentissage et des méthodes plus classiques et non supervisées de correction d'images pourront être mise en oeuvre.

Les nouveaux algorithmes robustes à des données "in the wild", la base de données ainsi créés et la comparaison avec l'état de l'art des approches non supervisées pourront être diffusés lors de conférence de rang A ou de journaux.

A mi-thèse, jusqu'à la moitié de la troisième année, le doctorant sera amené à développer des approches mettant en oeuvre une interaction entre le capteur et le réseau de neurones, pour proposer des imageurs non conventionnels associés à des traitements par apprentissage. Pour étudier l'influence du flou de défocalisation sur les performances, le doctorant pourra commencer son étude sur la base de données de Depth-from-Defocus (ONERA DFDIW) réalisée dans la thèse de M. Carvalho à l'aide de capteurs chromatiques, mais également modifier le banc de mesure avec des capteurs présentant d'autres solutions optiques (lame cylindrique, pupille codée etc).

Un état de l'art des approches de modélisation du capteur dans les réseaux de neurone sera effectué de manière plus poussée qu'en première année. Le doctorant sera alors amené à simuler les dispositifs expérimentaux de l'ONERA pour déterminer typiquement le réglage conduisant aux meilleures performances du système complet. Il pourra alors comparer ses résultats de simulation avec des données expérimentales. Cette première preuve de concept d'optimisation conduira à une publication dans une revue d'optique appliquée ou de traitement d'images/computational imaging.

La fin de la dernière année de thèse sera consacrée à la rédaction du manuscrit et à la mise en place d'une démonstration sur un capteur le plus compact possible pour envisager suivant le temps restant son intégration dans un système robotique.

b. techniques à mettre en oeuvre

Apprentissage automatique dont deep learning sur des données denses

Apprentissage multi-tâche pour la correction des images et l'estimation de la 3D sur des données souffrant des dégradations classiques (flou de bougé, bruit, wobble lié au rolling shutter)

Modélisation du capteur dans un réseau de neurones

Optimisation d'un système d'imagerie combiné à un réseaux de neurones

c. résultats attendus

Les différentes techniques mises en oeuvre conduiront à des publications dans des journaux et des conférences de références en vision par ordinateur, optique et robotique

Outils logiciels et prototype de caméra pour la vision 3D (optique + traitement)

## 5. Financement envisagé

Cocher dans la colonne de droite

Type de bourse	
ONERA	<input checked="" type="checkbox"/>
DGA	<input checked="" type="checkbox"/>
Région	<input type="checkbox"/>
Contrat doctoral	<input type="checkbox"/>
CIFRE	<input type="checkbox"/>

CNES	<input type="checkbox"/>
Autre	<input type="checkbox"/>

Commentaire éventuel :

## 6. Avis de l'Adjoint Scientifique du Département