

# ROB317 - Vision 3d

## TP3 : Inférence de la 3d par Deep Learning

Janvier 2026

### 1 Introduction

Le but de ce TP est d'expérimenter les capacités d'un réseau de neurones profond entraîné par auto-supervision, à prédire une carte de profondeur (3d) et un déplacement de la caméra (odométrie), à partir d'images monoculaires.

Pour cela nous allons utiliser le réseau *Monodepth2* [1], dont les poids ont été appris sur la base de données Kitti [2], en co-entraînant un réseau de prédiction de profondeur et un réseau de prédiction de pose (voir Figure 1) à partir de la fonction de loss photométrique (erreur de re-projection).

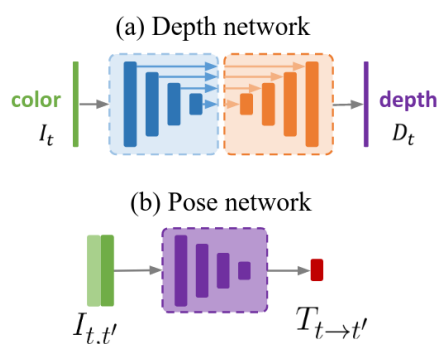


FIGURE 1 – Architecture de Monodepth2 [1].

Préliminaires informatiques :

Suivez les instructions d'installation sur le Github de Monodepth2 :

<https://github.com/nianticlabs/monodepth2>

et récupérez 2 séquences de test sur le répertoire :

[https://perso.ensta-paris.fr/~manzaner/Cours/ROB317/Images\\_TPDeep/](https://perso.ensta-paris.fr/~manzaner/Cours/ROB317/Images_TPDeep/)

Pour la dernière section, vous aurez besoin du code python suivant :

[https://perso.ensta-paris.fr/~manzaner/Cours/ROB317/TP\\_3dML.py](https://perso.ensta-paris.fr/~manzaner/Cours/ROB317/TP_3dML.py)

### 2 Prédiction de la profondeur

Expérimentez, sur plusieurs images, la prédiction de profondeur du réseau entraîné.

1. Quels types de biais peut-on attendre de la prédiction connaissant les images sur lesquelles le réseau a été entraîné (voir Figure 2) ?
2. Comment réduire certains de ces biais, sans ré-entraîner le modèle, en modifiant les images fournies en inférence ? Montrer quelques exemples. Et comment les réduire en ré-entraînant le modèle ?
3. Quels critères d'évaluation qualitative (i.e. sans disposer d'annotations de profondeur) peut-on utiliser ?

4. Quelles métriques d'évaluation quantitative peut-on utiliser si l'on dispose d'annotations de profondeur (éventuellement partielles, comme dans la base d'évaluation de Kitti) ?

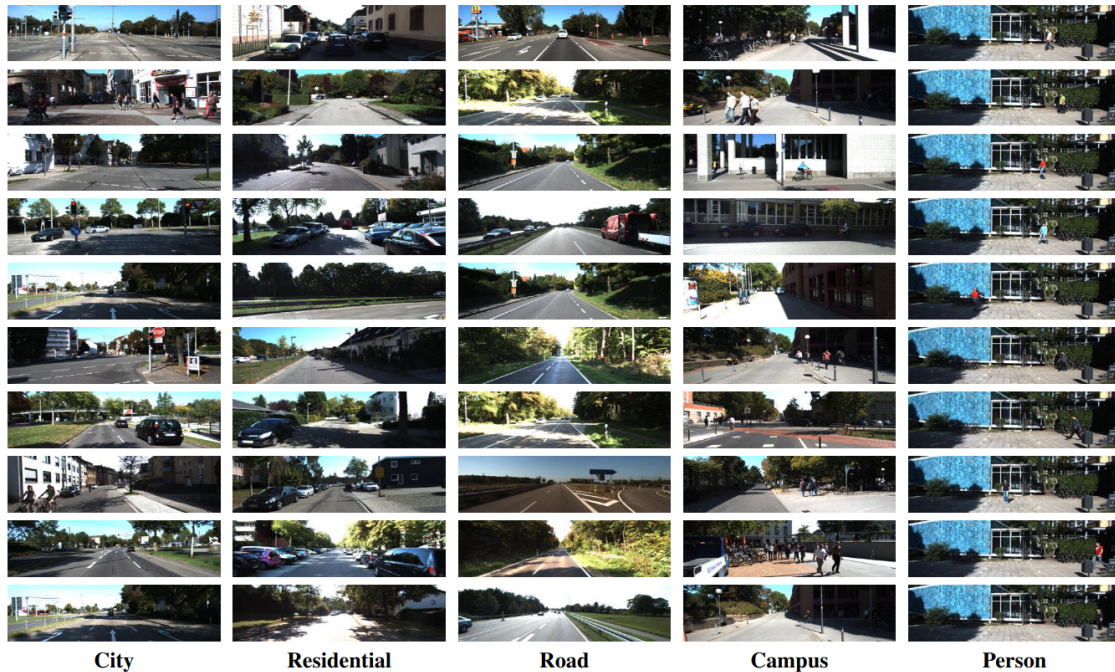


FIGURE 2 – Exemples d'images de la base Kitti [2].

### 3 Prédiction de l'odométrie

Testez la prédiction de poses sur une séquence d'images et tracez le parcours en 2d correspondant sur l'image du parcours réel obtenu par GPS+IMU dans la séquence proposée.

1. Quelle métrique proposez-vous pour évaluer l'odométrie sur une séquence ?
2. Quelles sont les limites de cette approche et comment l'améliorer ?

### 4 Pour aller plus loin

Les questions suivantes sont plus difficiles. Un code vous est mis à disposition avec des fonctions utiles pour vous aider. A vous de trouver comment les utiliser.

1. Dans Monodepth2 : retrouver comment obtenir un nuage de points 3D.
  - (a) Quel est le module/fonction qui fait la conversion depth + pose vers point 3D (back-projection) ?
  - (b) Même question pour l'opération inverse points 3D vers 2D couleur
2. Construire un nuage de points 3D couleur
  - (a) À partir d'une image RGB et de sa profondeur, comment associer une couleur à chaque point 3D ?
  - (b) Quels pixels devez-vous retirer (profondeur invalide / trop grande / etc.) pour éviter un nuage trop pollué ?
3. Nuage multi-vues + affichage Open3D

- (a) Comment ramener les points de plusieurs images dans un même repère pour fusionner les nuages afin d'obtenir un nuage total plus dense ?
  - (b) Montrez dans Open3D : (a) le nuage d'une image, puis (b) le nuage fusionné sur N images successives.
4. CLIP : carte de chaleur 3D guidée par un prompt
- (a) Pour un prompt (ex. "car" / "road"), comment obtenir une heatmap 2D de similarité CLIP sur l'image ?
  - (b) Comment "porter" cette heatmap en 3D : quelle valeur associer à chaque point 3D (score) et comment la visualiser (couleur/intensité) dans Open3D ?

## Références

- [1] C. Godard, O. Mac Aodha, M. Firman and G. J. Brostow. Digging into Self-Supervised Monocular Depth Prediction The International Conference on Computer Vision (ICCV), 2019.
- [2] A. Geiger, P. Lenz and R. Urtasun. Are we ready for autonomous driving ? The KITTI vision benchmark suite IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012.