# Disentangling Object Motion for Self-supervised Depth Estimation

CV1-Final Project

Jinxi Xiao, Kecheng Ye, Yi'ang Ju, Panfeng Jiang, Haoyu Wu

*ShanghaiTech University*

## 1.Introduction

- Estimating monocular depth via unsupervised learning has emerged as a promising approach in many fields. However, one of the main difficulties would be occlusions caused by the motion of dynamic objects within the monocular inputs. In order to solve this problem, we propose some improvements based on established ideas.

## 2. Contributions

We mainly propose the following improvements:

(1) We introduce two loss functions to the Dynamic-Depth framework: L1 normal loss and normal ranking loss.

(2) Introduce pose cycle consistency constraint to enhance the ability of *Pose Net* to predict poses.

(3) Create a weighted scheme based on masks. Automatically adjust the weight of each datum according to the size of its dynamic object mask.

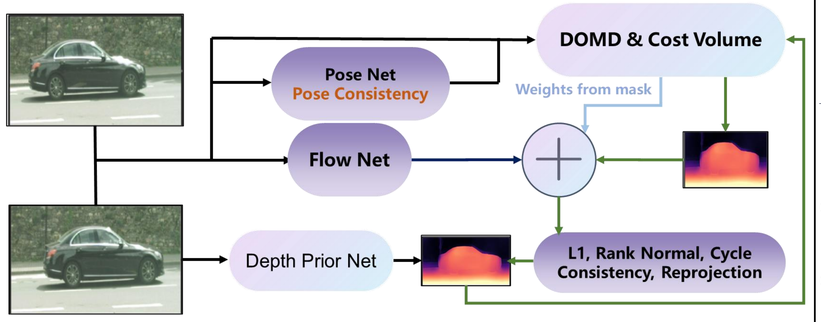
(4) Add a *Flow Net* into the framework of Dynamic-Depth, which predicts the optical flow of two images.

## 3. Methods

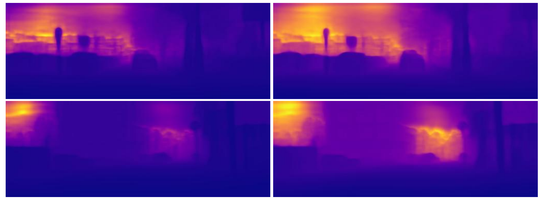
(1) Loss functions metioned in 2.1 make use of the geometry information.

(2) Pose from t to t+1 product pose t+1 to t is Indentity

(3) General pipeline




## 4. Experiments



depth map predicted by Dynamic-Depth

depth map predicted by our improved version

| model | abs_rel | sq_rel | rms | log_rms | a1 | a2 | a3 |
|---|---|---|---|---|---|---|---|
| L1 normal | 0.1134 | 1.1123 | 6.1344 | 0.1670 | 0.8730 | 0.9689 | 0.9901 |
| L1+ranking | 0.1125 | 1.1128 | 6.2354 | 0.1673 | 0.8751 | 0.9694 | 0.9904 |
| L1+ranking(weighted) | 0.1075 | 1.0613 | 6.1773 | 0.1634 | 0.8804 | 0.9707 | 0.9909 |
| Pose cycle consistency | 0.1066 | 1.0660 | 6.0124 | 0.1599 | 0.8893 | 0.9724 | 0.9910 |
| Auto doj mask weight | 0.1056 | 1.0620 | 6.0105 | 0.1595 | 0.8898 | 0.9725 | 0.9911 |
| Flow Net | 0.1091 | 1.1080 | 6.0390 | 0.1635 | 0.8842 | 0.9707 | 0.9905 |
| Dynamic-Depth | 0.1073 | 1.0856 | 6.0305 | 0.1605 | 0.8875 | 0.9720 | 0.9910 |

Numerical Results

## 5. Reference

① Z. Feng, L. Yang, L. Jing, H. Wang, Y. Tian, and B. Li, "Disentangling object motion and occlusion for unsupervised multi-frame monocular depth," arXiv preprint arXiv:2203.15174, 2022.

② L. Sun, J.-W. Bian, H. Zhan, W. Yin, I. Reid, and C. Shen, "Sc-depthv3: Robust self-supervised monocular depth estimation for dynamic scenes," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2023.

③ X. Miao, Y. Bai, H. Duan, Y. Huang, F. Wan, X. Xu, Y. Long, and Y. Zheng, "Ds-depth: Dynamic and static depth estimation via a fusion cost volume," IEEE Transactions on Circuits and Systems for Video Technology, 2023.