Predicting Sharp and Accurate Occlusion Boundaries in Monocular Depth Estimation Using Displacement Fields

Michaël Ramamonjisoa[1,*] Yuming Du[1,*] Vincent Lepetit[1]

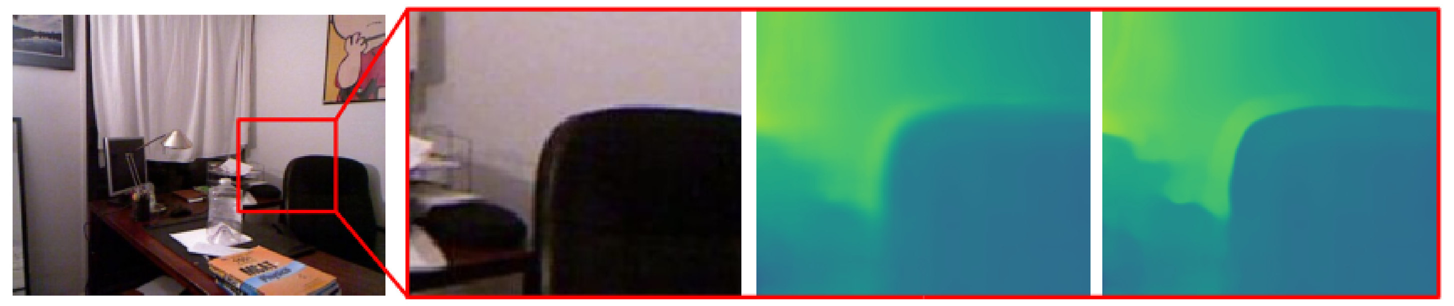
[1] LIGM (UMR 8049), IMAGINE, Ecole des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France

https://michaelramamonjisoa.github.io/projects/DisplacementFields

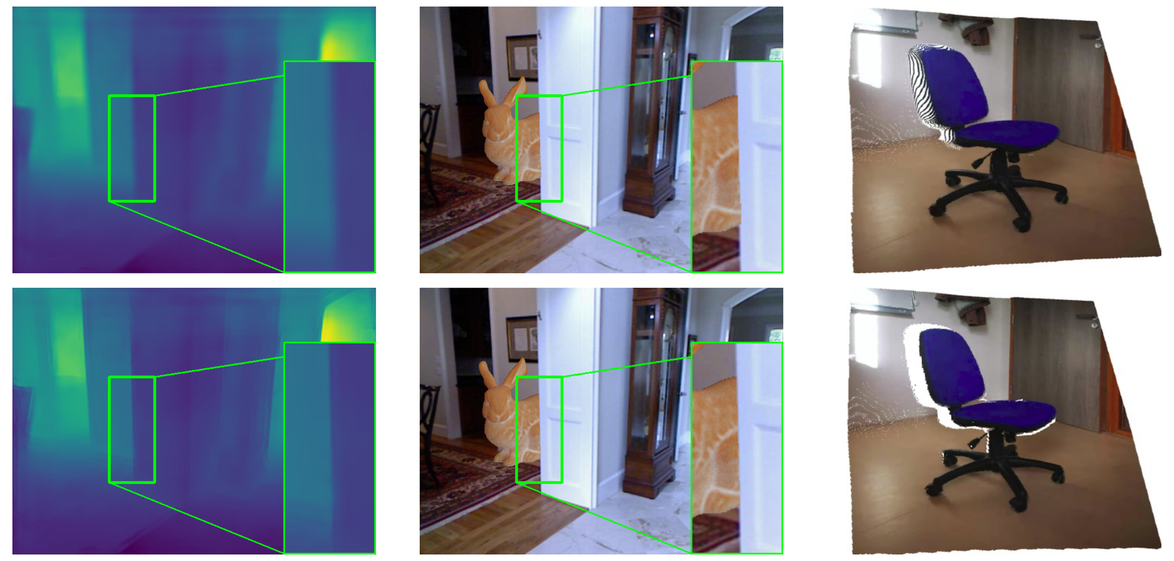
CVPR SEATTLE WASHINGTON JUNE 16-18 2020

Contributions

- We propose a simple method to predict depth maps from a single color image with high accuracy along the Occlusion Boundaries (OB). Our method can be plugged on top of any Monocular Depth Estimation (MDE) method:

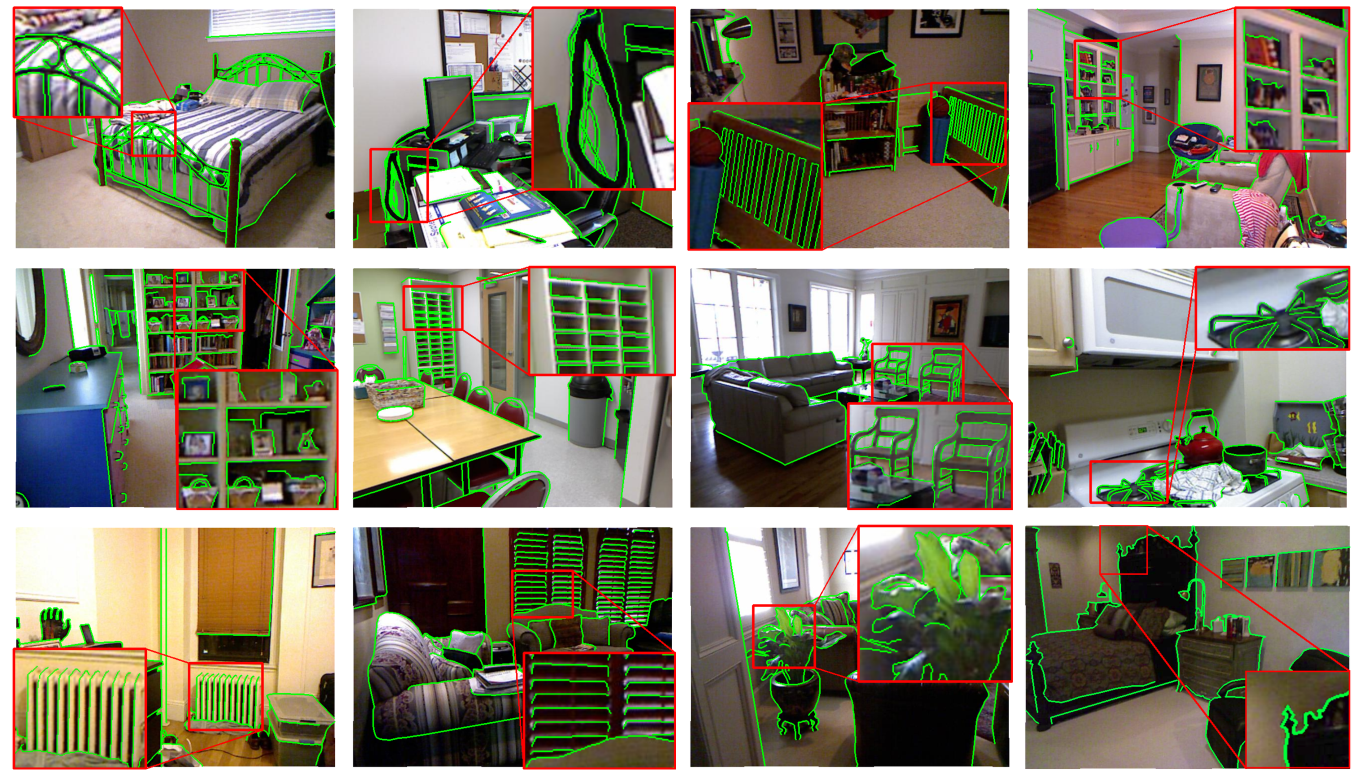


Possible applications to Augmented Reality and object extraction



- We introduce the NYUv2-OC++ dataset. This dataset contains manual, fine-grained annotations for occlusion boundaries across all 654 images from the evaluation set of the NYUv2 dataset [2]. Note that we use this dataset for evaluation only, not for training.

NYUv2-OC++



Method

Problem formulation

We learn to recover **sharp** depth edges by refining depth maps predicted by MDE methods.

- This has been done by predicting an additive residual ΔD [3]:

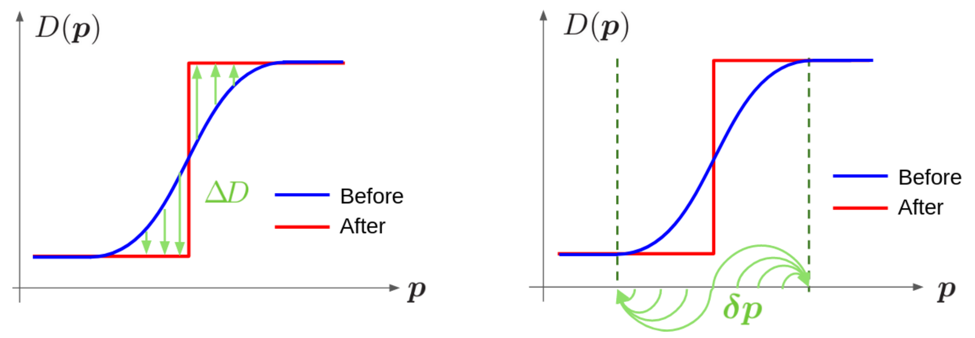
$$\text{for each image location } \boldsymbol{p}, \quad D(\boldsymbol{p}) \leftarrow D(\boldsymbol{p}) + \Delta D(\boldsymbol{p}) \, , \tag{1}$$

However this approach can introduce artifacts.

- Instead, we propose to predict a field $\boldsymbol{\delta p}$ of local 2D transformations to refine depth maps, which we call **displacement fields**.

$$\forall \boldsymbol{p}, \quad D(\boldsymbol{p}) \leftarrow D(\boldsymbol{p} + \boldsymbol{\delta p}(\boldsymbol{p})) \quad \text{where } \boldsymbol{\delta p}(\boldsymbol{p}) \in \mathbb{R}^2 \tag{2}$$

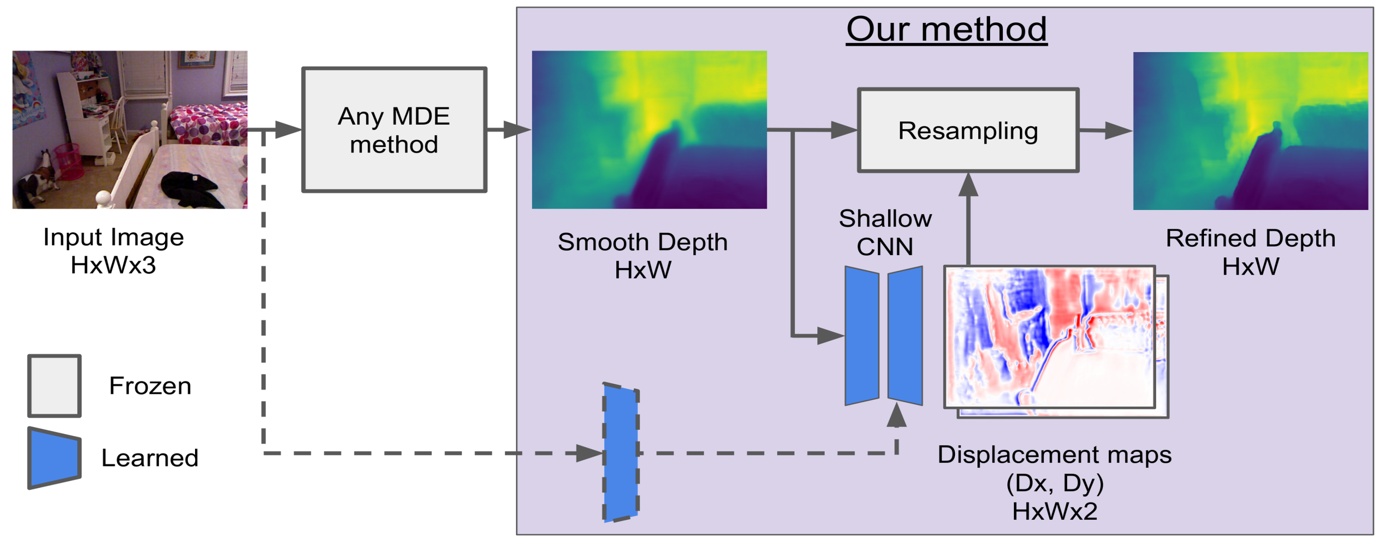
This results in a **bounded error** in depth and hence avoids artifacts that additive residuals introduce.

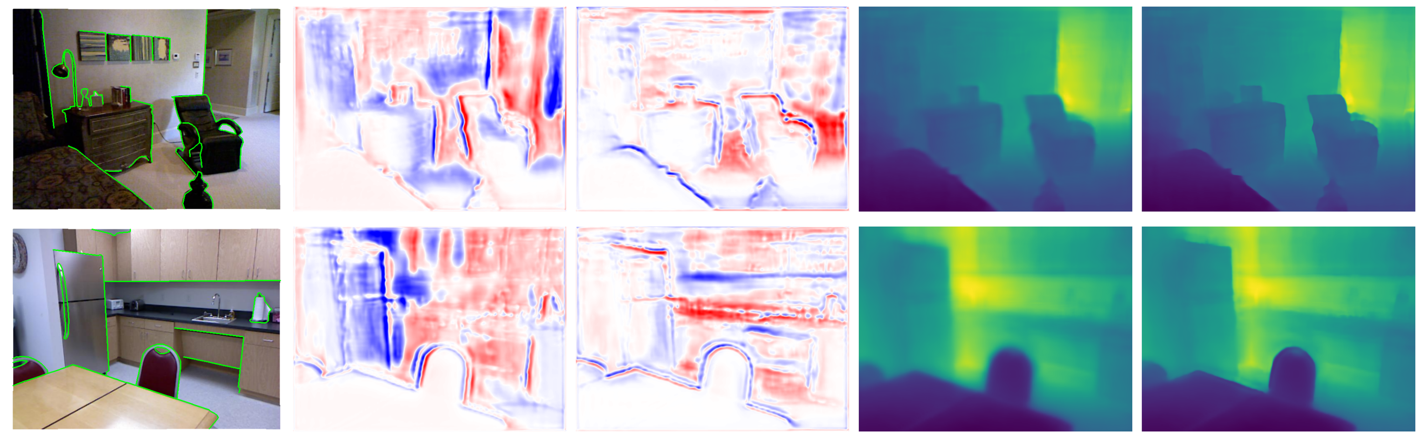


- Results on a toy problem:

Implementation

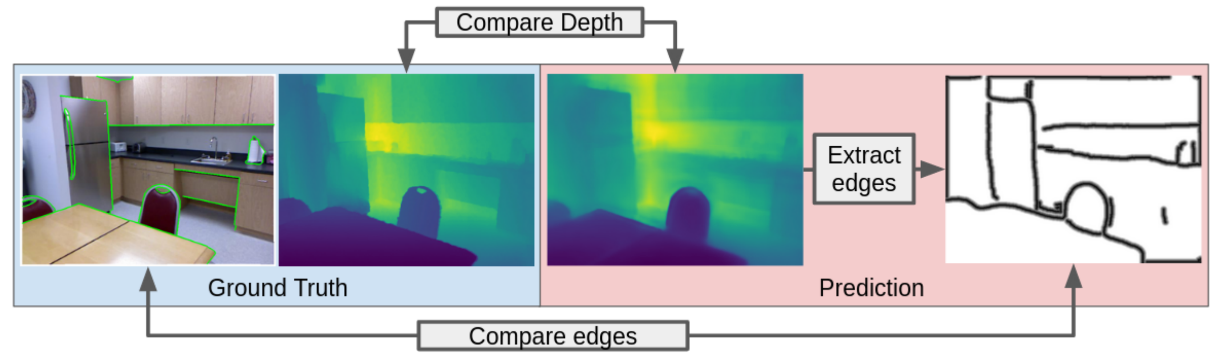
- We train a network to predict a 2D displacement field from an initial depth map.
- The refined depth map is computed by displacing pixels according to Eq. (2), using the predicted displacement fields $\boldsymbol{\delta p}$ and a differentiable resampling block.
- Using the input image for guidance is optional. This slightly improves the contour localisation.

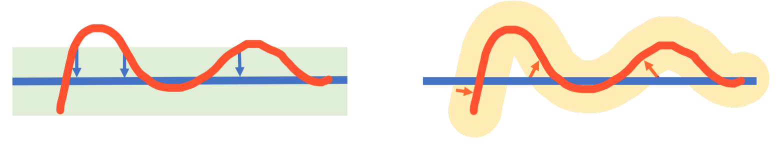




Evaluation

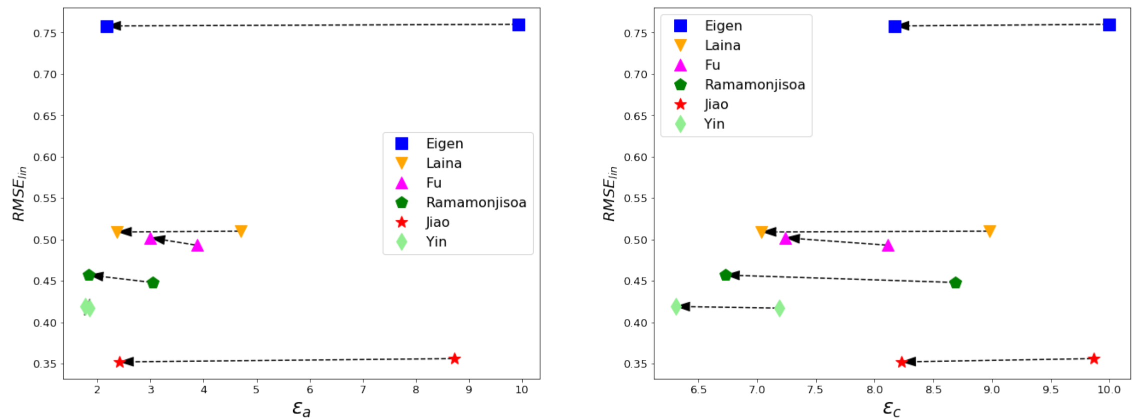
- Global depth reconstruction is evaluated on NYUv2-Depth [2].
Occlusion Boundaries accuracy is evaluated on NYUv2-OC++ using edge extraction and chamfer distance metrics ϵ_a and ϵ_c [4].



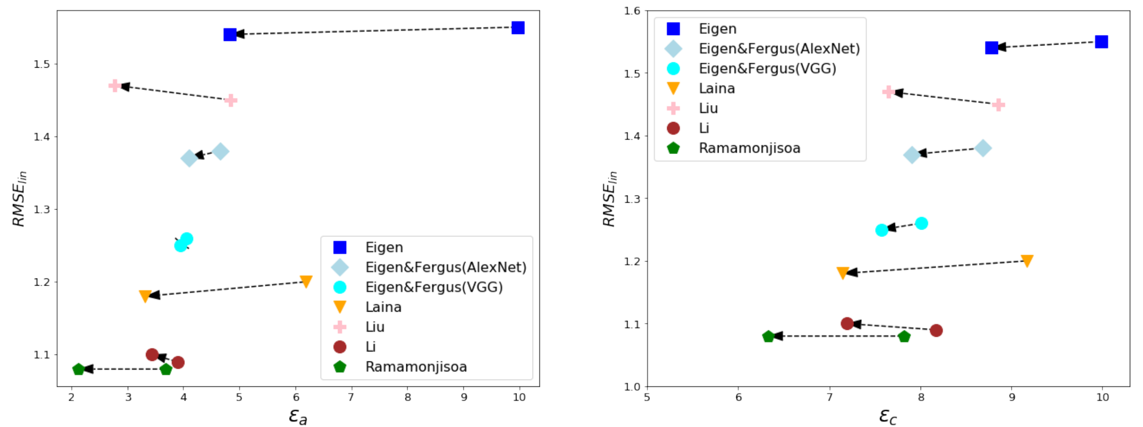


- Occlusion Boundaries are improved with minimal impact on global depth reconstruction

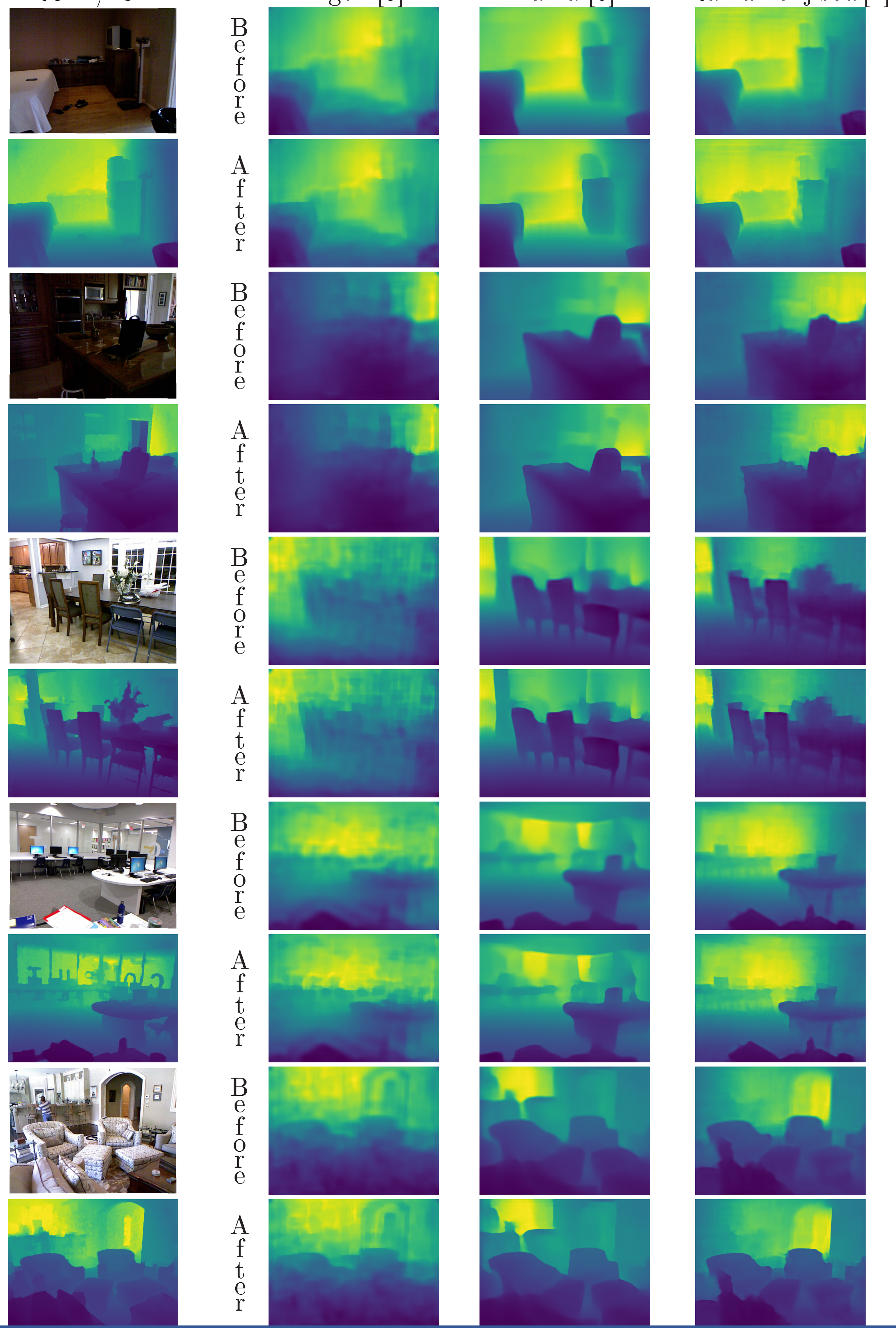
NYUv2 [2]



iBims [4]



Qualitative results



References

[1] M. Ramamonjisoa and V. Lepetit. Sharpnet: Fast and Accurate Recovery of Occluding Contours in Monocular Depth Estimation, *ICCV Workshop*, 2019

[2] N. Silberman, D. Hoiem, P. Kohli and R. Fergus, Indoor Segmentation and Support Inference from RGBD Images, *ECCV*, 2012

[3] X. Chen, X. Chen, Z-J. Zha, Structure-Aware Residual Pyramid Network for Monocular Depth Estimation *IJCAI*, 2019

[4] T. Koch, L. Liebel, F. Fraundorfer and M. Körner, Evaluation of CNN-Based Single-Image Depth Estimation Methods, *ECCV Workshops*, 2018

[5] D. Eigen, C. Puhrsch, and R. Fergus, Depth Map Predictionfrom a Single Image Using a Multi-Scale Deep Network. *NIPS*, 2014

[6] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari and N. Navab, Deeper Depth Prediction with Fully Convolutional Residual Networks, *3DV*, 2016