Monocular Depth Estimation

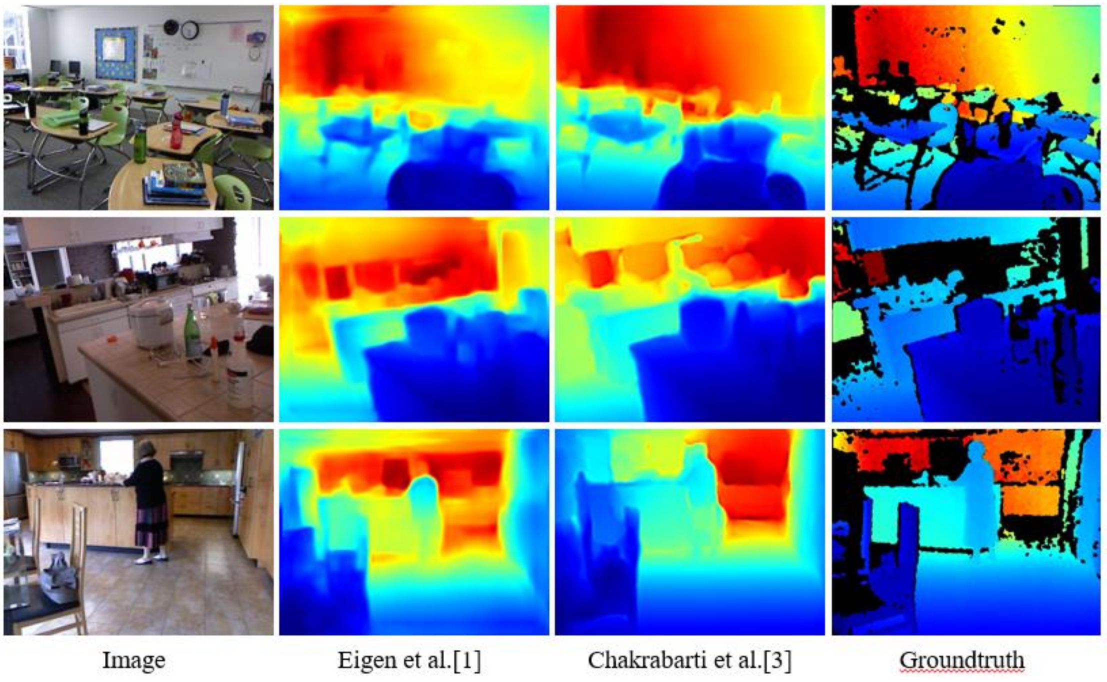

Figure 1: Example of different monocular depth estimation methods on the NYU Depth V2[4] dataset.

Depth estimation has been an essential component for scene understanding over a long period. One challenging setting category resorts to with monocular image. Monocular depth estimation, though inherently ill-posed, can be solved with specific scene condition. Recent years, thanks to the development of deep neural network, many attempts have been made to address the monocular depth estimation problem in a supervised learning manner. Most of them formulate it as a pixel level regression problem by Convolutional Neural Network (CNN), with promising results achieved. Most of these works focus on designing better network structures, loss functions, and additionally post processing steps.

In this project, we plan to estimating depth from monocular image using only the labeled dataset of NYU Depth V2[4], which contains 1449 image-depth pairs in total and can be found at https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html. With most common settings, 795 samples are used for training and 654 for testing, the train/test split can be found at http://horatio.cs.nyu.edu/mit/silberman/indoor_seg_sup/splits.mat. For performance metric, we follow Eigen et al.[1], which is also used by most other methods. We do not pay much attention to the final performance, so do not just simple reimplement an existing paper, show us your original idea that trying to solve something (e.g. speed up the model deploying, improve blur depth results, simultaneously estimate semantic labels, etc.) and show us your exploratory process. To begin with, we recommend to read[1,2] and the first work for pixel level task learning[5], then you may find some fancy ideas by reading more papers(e.g. from cvf(http://openaccess.thecvf.com/menu.py), arXiv(https://arxiv.org/list/cs.CV/recent), etc).

References

[1] D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network. In Advances in neural information processing systems, pages 2366–2374, 2014.

[2] D. Eigen and R. Fergus. Predicting depth, surface normal and semantic labels with a common multi-scale convolutional architecture. In Proceedings of the IEEE International Conference on Computer Vision, pages 2650–2658, 2015.

[3] A. Chakrabarti, J. Shao, and G. Shakhnarovich. Depth from a single image by harmonizing overcomplete local network predictions. In Advances in Neural Information Processing Systems, pages 2658–2666, 2016.

[4] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. Computer Vision–ECCV 2012, pages 746–760, 2012.

[5] J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.