Single Noisy Image Depth Estimation with Multi-Scale Feature Fusion Module

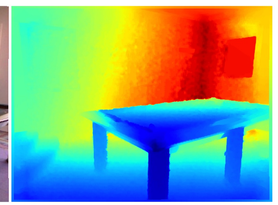
Yuxiao Chen
yuxiaoc@stanford.edu

Qingxi Meng
qingxi@stanford.edu

Introduction

- Monocular depth estimation is ill-posed due to the lack of information [1]
- Stereo information more expensive and unsuitable for deployment [2]
- Gaussian noise stereo image reconstruction
  - Explore median and bilateral denoise filter
  - Explore multi-scale feature fusion module (MFF) [3]

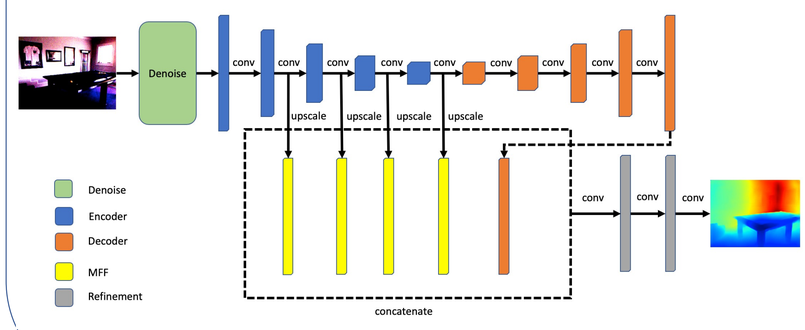
Ground Truth RGB



Ground Truth Depth

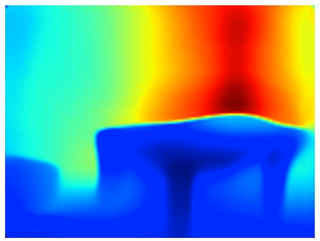
Model Architecture

- Denoise: Median / Bilateral
- Depth Estimation: encoder (E), decoder (D), multi-scale feature fusion module (MFF), and refinement module (R)

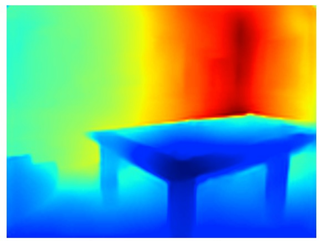


Experiments

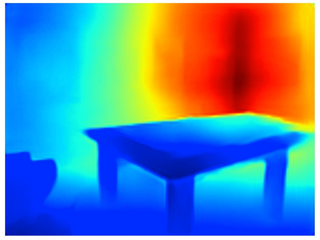
- Trained on 8000 uniformly sampled rgb-depth pairs
  - Add Gaussian noise σ=0.1 to the RGB image input
  - Use depth loss, normal loss, and gradient loss
- Evaluate with 150 unseen rgb-depth image pairs



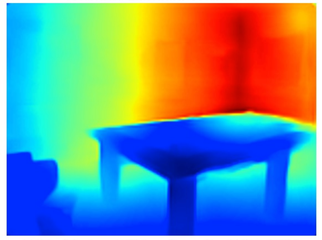
Without Denoise Without MFF



Without Denoise With MFF

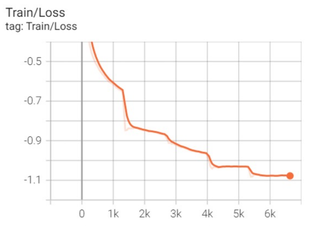


With Median Filter Denoise with MFF

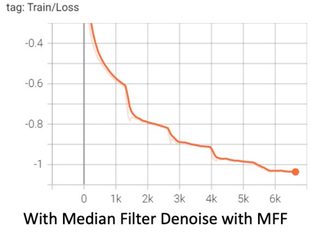


With Bilateral Filter Denoise With MFF

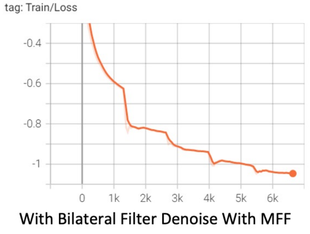
Without Denoise Without MFF



Without Denoise With MFF



With Median Filter Denoise with MFF



With Bilateral Filter Denoise With MFF

Discussion

	Model Trained without Denoise without MFF	Model Trained without Denoise with MFF	Model Trained with Median Filter Denoise with MFF	Model Trained with Bilateral Filter Denoise with MFF
Root Mean Squared Error (RMSE)	0.660	0.594	0.697	0.574
Mean Absolute Error (MAE)	0.409	0.371	0.451	0.366
Mean Relative Error (REL)	0.151	0.140	0.159	0.136

Bilateral denoise > no denoise > median denoise > no MFF

Challenges / Limitations
- Denoise filter can smooth out edges, impede the depth-estimation performance
- Real noise have unknown distribution

Summary

Summary
- Training with MFF yields lower loss than without MFF
- Denoise slightly improves depth estimation

Future Work
- Explore other denoise models (Non-Local Mean, DnCNN)
- Explore other depth estimation methods (self-supervised)

Reference
[1] S. Zia, B. Yu¨ksel, D. Yu¨ret, and Y. Yemez. Rgb-d object recognition using deep convolutional neural networks, 2017
[2] M. M. Johari, C. Carta, and F. Fleuret. Depthinspace: Ex- ploitation and fusion of multiple video frames for structured-light depth estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 6039– 6048, 2021.
[3] J. Hu, M. Ozay, Y. Zhang, and T. Okatani. Revisiting single image depth estimation: Toward higher resolution maps with accurate object boundaries. In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pages 1043– 1051. IEEE, 2019.