Sparse-to-Dense Depth Completion Revisited: Sampling Strategy and Graph Construction–Supplemental Material*

Xin Xiong[1], Haipeng Xiong[1], Ke Xian[1], Chen Zhao[1], Zhiguo Cao[1], and Xin Li[2]

[1] School of AIA, Huazhong University of Science and Technology, China
{xiong_xin,hpxiong,kexian,hust_zhao,zgcao}@hust.edu.cn
[2] Lane Department of CSEE, West Virginia University, USA.
xin.li@mail.wvu.edu

1 Implementation Details

In all of our experiments, we use Mean Square Error(MSE) as the loss function:

$$Loss = \frac{1}{N} \sum_{k \in K_v} ||D_k^{gt} - D_k^{pre}||^2 \tag{1}$$

where K_v is the set of valid points, N is the number of valid points. D_k^{gt} and D_k^{pre} are ground truth depth value and predicted depth value, respectively.

We complete all of our experiments based on two RTX TITAN GPUs. All of our models except the model fine-tuned on NYUDv2 dataset [8] are trained with a batch size of 8 for about 50 epochs and use Adam [4], where $\beta_1 = 0.9, \beta_2 = 0.999$, $\epsilon = 10^{-8}$, and weight decay is set to 10^{-6}. Batch size is set to 4 when the model is fine-tuned on NYUDv2 dataset. As for learning rate, we set the initial learning rate to 10^{-3} in all of our experiments. The learning rate decays exponentially with epochs and the learning rate for the t_{th} epoch can be written as:

$$LR_t = 10^{-3} * (1 - \frac{t-1}{50})^{0.9} \tag{2}$$

where LR_t is the learning rate for the t_{th} epoch.

* Xin Xiong and Haipeng Xiong contributed equally. Zhiguo Cao is the corresponding author. This work was supported in part by the National Key R&D Program of China (No.2018YFB1305504) and the National Natural Science Foundation of China (Grant No. U1913602). Xin Li's work is partially supported by the DoJ/NIJ under grant NIJ 2018-75-CX-0032, NSF under grant OAC-1839909, IIS-1951504 and the WV Higher Education Policy Commission Grant (HEPC.dsr.18.5).

2 Results on NYUDv2

2.1 NYUDv2 Dataset

The NYUDv2 dataset [8] provides RGB images and dense depth maps captured by Microsoft Kinect from 464 indoor scenes. Its raw data contains more than $100k$ samples and following [7], [2], [6], we use about 46k samples as training data. Besides, NYUDv2 provides an officially labeled subset containing 1449 samples (654 for testing). To fill missing values, the depth values are in-painted using the official toolbox, which adopts the colorization scheme [5]. Following [10] and [6], we have down-sampled the origin images to half-resolution and center-cropped to the dimension of 320×256 with additional paddings.

Fig. 1 and Fig. 2 show the results of our baseline model with different sampling strategies. Fig. 3 and Fig. 4 show the results of our full model with different sampling strategies.

2.2 Cross Dataset Evaluation: From KITTI to NYUDv2

We fine-tune the models trained on KITTI[3] with the 795 training samples in NYUDv2 subset under different sampling strategies. Specifically, we fix the feature extractor and only fine-tune the last layer. Fig. 5 and Fig. 6 show the outdoor-indoor transfer results of our baseline and full model with different sampling strategies.

Table 1. Result of models transferred from KITTI dataset to NYUDv2 dataset using different sampling strategies. All models are pre-trained on KITTI dataset. For method, Ours GNN is our full model with graph neural network. Mode "D" means directly evaluating the model on NYUDv2 test set with 654 samples and mode "F" means fine-tuning the model on NYUDv2 subset with 795 samples. For methods with *, we test the pre-trained model provided by their author.

Method	Mode	Sample	Rel↓	RMSE↓	δ_1↑	δ_2↑	δ_3↑
*Van et al. [9]	D	Random	0.090	0.487	85.7	93.6	97.2
Ours Baseline	D	Random	0.087	0.365	90.6	97.7	99.3
Ours GNN	D	Random	0.061	0.310	94.0	98.4	99.5
Ours Baseline	D	R2	0.060	0.261	95.8	99.0	**99.8**
Ours Baseline	D	Plastic	0.060	0.262	95.7	99.0	99.7
Ours Baseline	D	Golden	0.058	0.256	95.7	98.7	99.7
Ours Baseline	D	Halton2,3	0.070	0.294	94.6	98.7	99.7
Ours GNN	D	R2	**0.044**	**0.238**	**96.6**	**99.2**	**99.8**
Ours GNN	D	Plastic	**0.044**	**0.238**	**96.6**	**99.2**	**99.8**
Ours GNN	D	Golden	0.045	0.240	96.4	99.1	99.7
Ours GNN	D	Halton2,3	0.048	0.254	96.1	99.1	99.7
Ours Baseline	F	Random	0.075	0.275	93.9	98.9	99.8
Ours Baseline	F	R2	0.047	0.189	97.3	99.6	**99.9**
Ours Baseline	F	Plastic	0.047	0.189	97.3	99.6	**99.9**
Ours Baseline	F	Golden	0.044	0.177	97.4	99.6	**99.9**
Ours Baseline	F	Halton2,3	0.055	0.211	96.6	99.5	**99.9**
Ours GNN	F	Random	0.050	0.212	97.1	99.5	**99.9**
Ours GNN	F	R2	0.036	0.165	98.2	**99.7**	**99.9**
Ours GNN	F	Plastic	0.036	0.166	98.2	**99.7**	**99.9**
Ours GNN	F	Golden	**0.034**	**0.158**	**98.3**	**99.7**	**99.9**
Ours GNN	F	Halton2,3	0.040	0.177	98.1	**99.7**	**99.9**

3 Results on Matterport3D

3.1 Supplementary note for Matterport3D

Matterport3D [1] is an indoor large-scale RGB-D dataset with $10.8k$ real panoramic views and 90 real indoor scenes. We use the same training and testing split as Zhang [11]. There are about 104k samples for training and 474 samples for testing. The ground truth depth map of Matterport3D is generated from Zhang [11] using multi-view reconstruction method. It should be noticed that Matterport3D provides raw depth map produced by the sensor. Those depth maps always have holes in which depth valves are missed. Different from recoverving depth from sparse points, the depth completion task on Matterport3D aims to recover dense depth map from those raw depth with holes.

Fig. 7 shows the results of our baseline model and full model on Matterport3D.

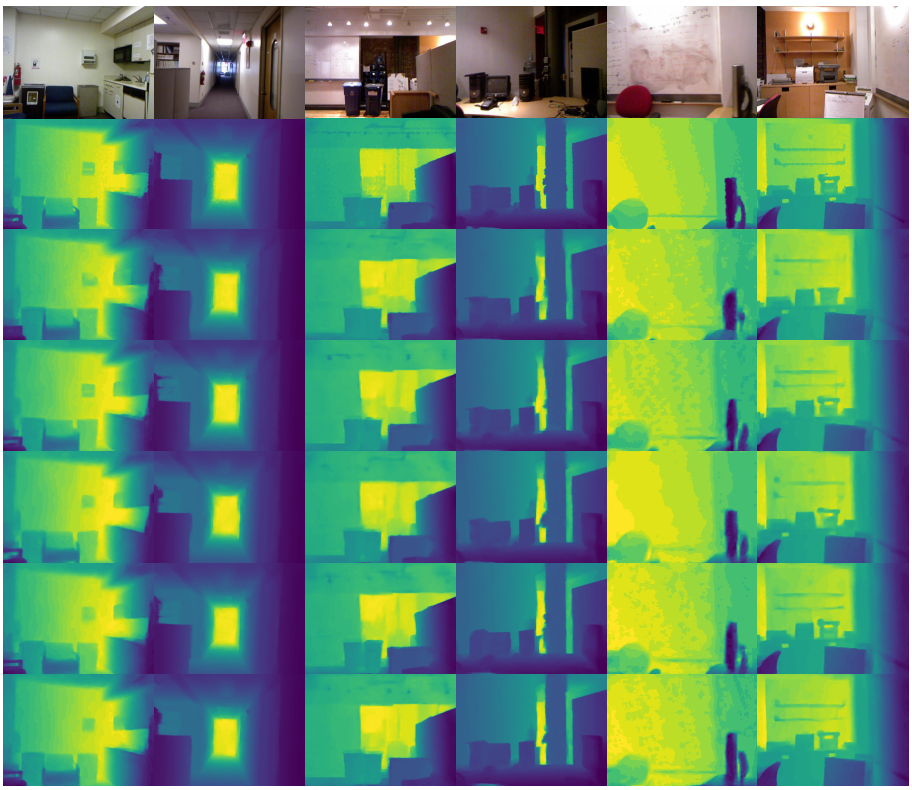

Fig. 1. Results of our baseline models with different sampling strategies on NYUDv2 test set. From top to bottom are the RGB image, the ground truth depth map and the results of our baseline model with Random, R2, Plastic, Halton and Golden sampling strategy, respectively.

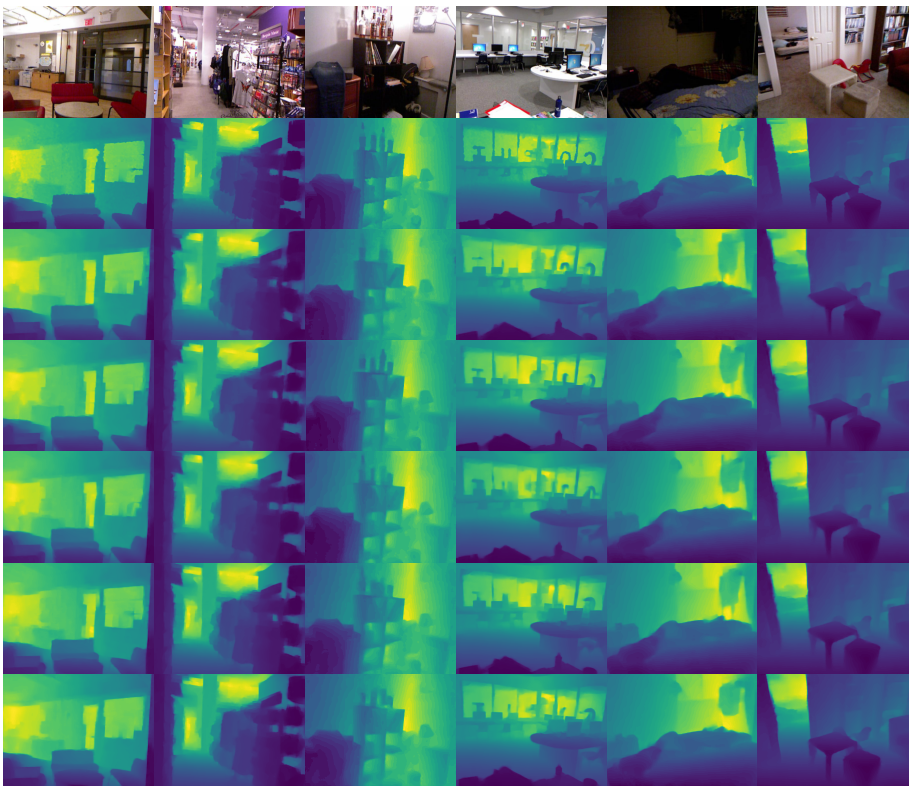

Fig. 2. Results of our baseline models with different sampling strategies on NYUDv2 test set. From top to bottom are the RGB image, the ground truth depth map and the results of our baseline model with Random, R2, Plastic, Halton and Golden sampling strategy, respectively.

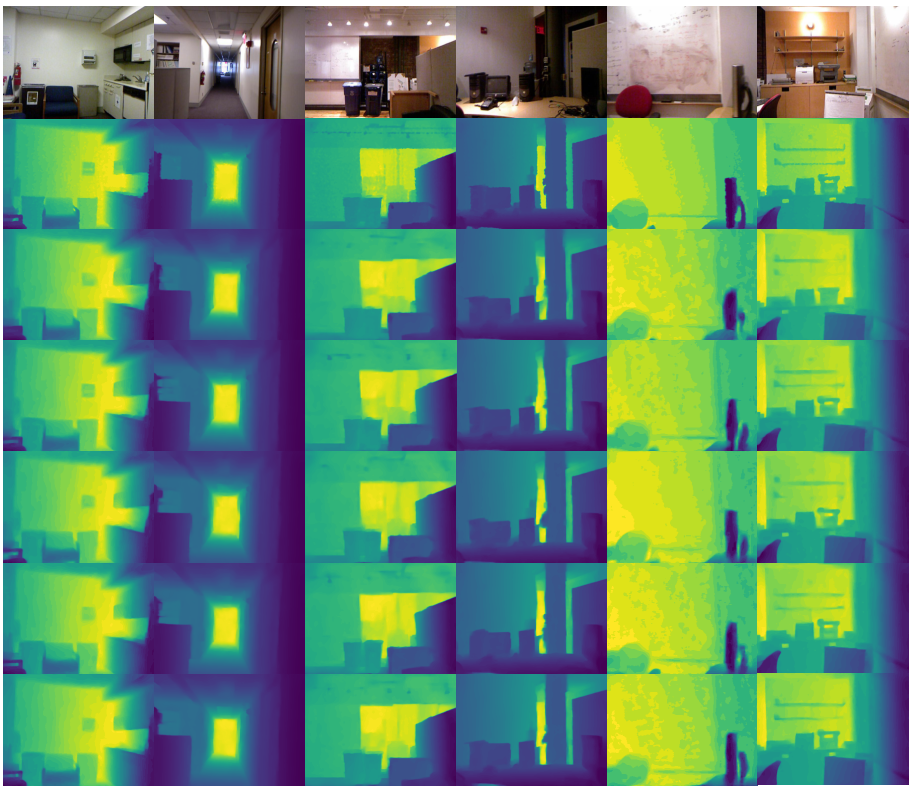

Fig. 3. Results of our full models with different sampling strategies on NYUDv2 test set. From top to bottom are the RGB image, the ground truth depth map and the results of our full model with Random, R2, Plastic, Halton and Golden sampling strategy, respectively.

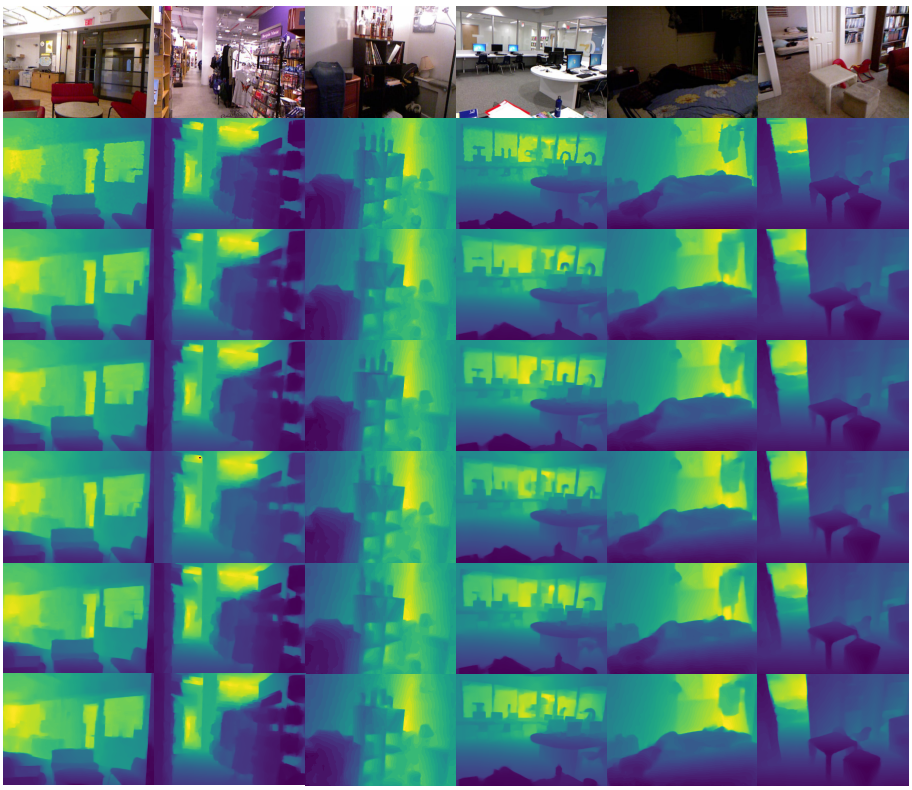

Fig. 4. Results of our full models with different sampling strategies on NYUDv2 test set. From top to bottom are the RGB image, the ground truth depth map and the results of our full model with Random, R2, Plastic, Halton and Golden sampling strategy, respectively.

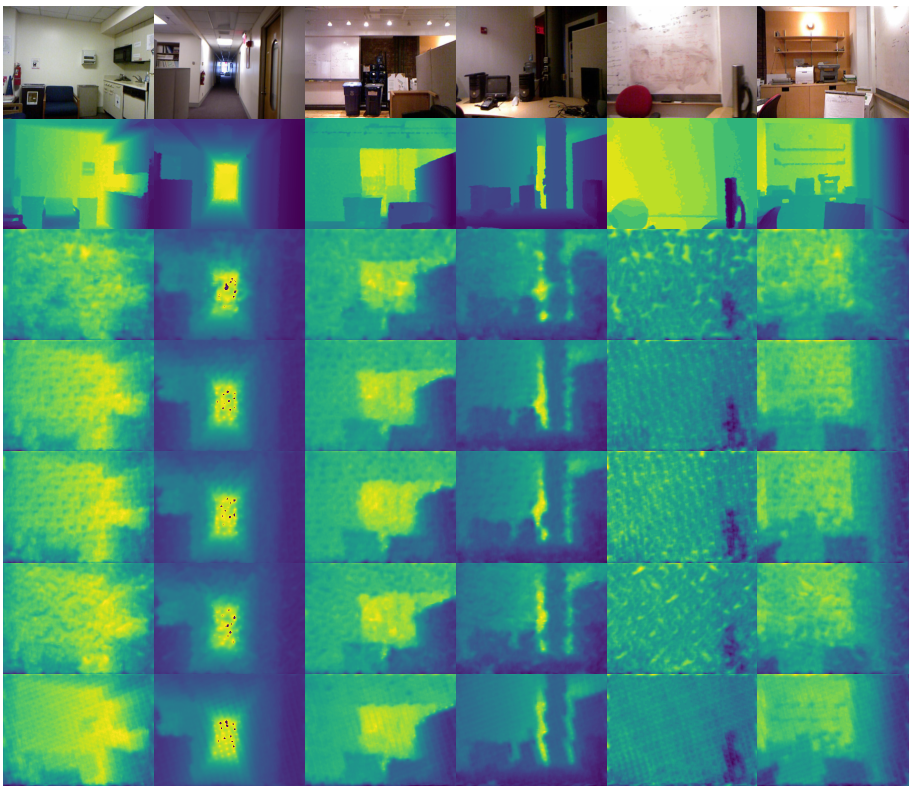

Fig. 5. Results of our baseline model trained on KITTI[3] and fine-tuned on NYUDv2 sub set while tested on NYUDv2 test set. From top to bottom are the RGB image, the ground truth depth map and the results of our full model with Random, R2, Plastic, Halton and Golden sampling strategy, respectively.

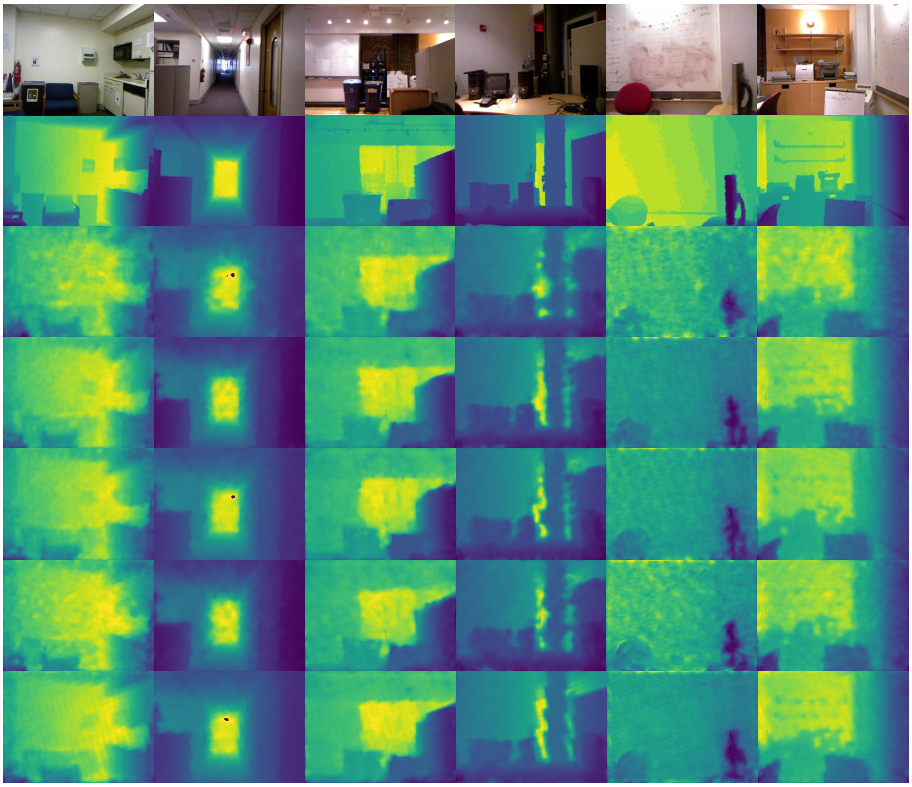

Fig. 6. Results of our full model trained on KITTI[3] and fine-tuned on NYUDv2 sub set while tested on NYUDv2 test set. From top to bottom are the RGB image, the ground truth depth map and the results of our full model with Random, R2, Plastic, Halton and Golden sampling strategy, respectively.

3.2 Cross Dataset Evaluation: From NYUDv2 to Matterport3D

Matterport3D dataset provides both the raw depth map and the ground truth depth map, and we can do the cross dataset evaluation from NYUDv2 to Matterport3D. We utilize the raw depth map as sparse depth input.

Fig. 8 shows the cross dataset evaluation results of our full model trained on NYUDv2 with different sampling strategies.

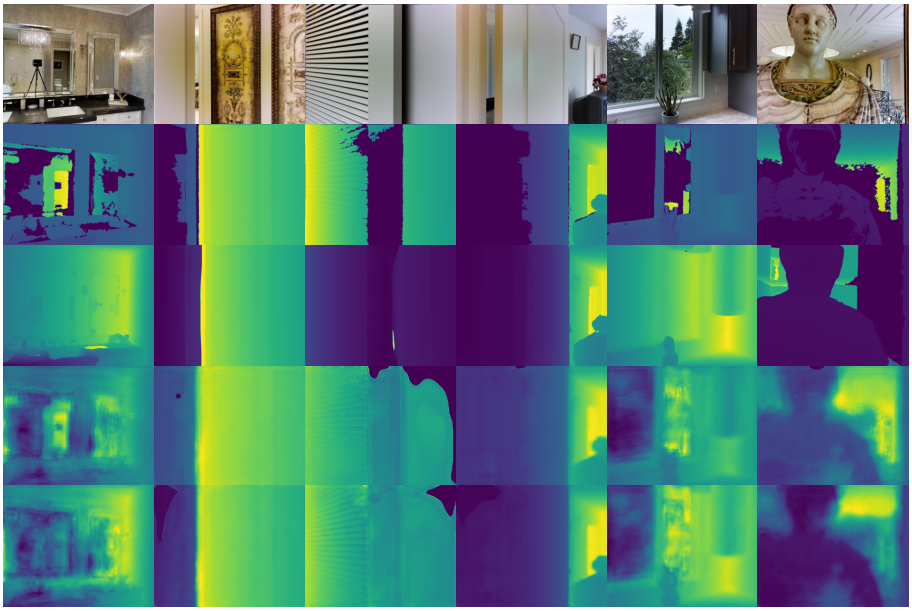

Fig. 7. Results of our models on Matterport3D test set. From top to bottom are the RGB image, the raw depth map, the ground truth depth map, the results of our baseline model and the results of our full model.

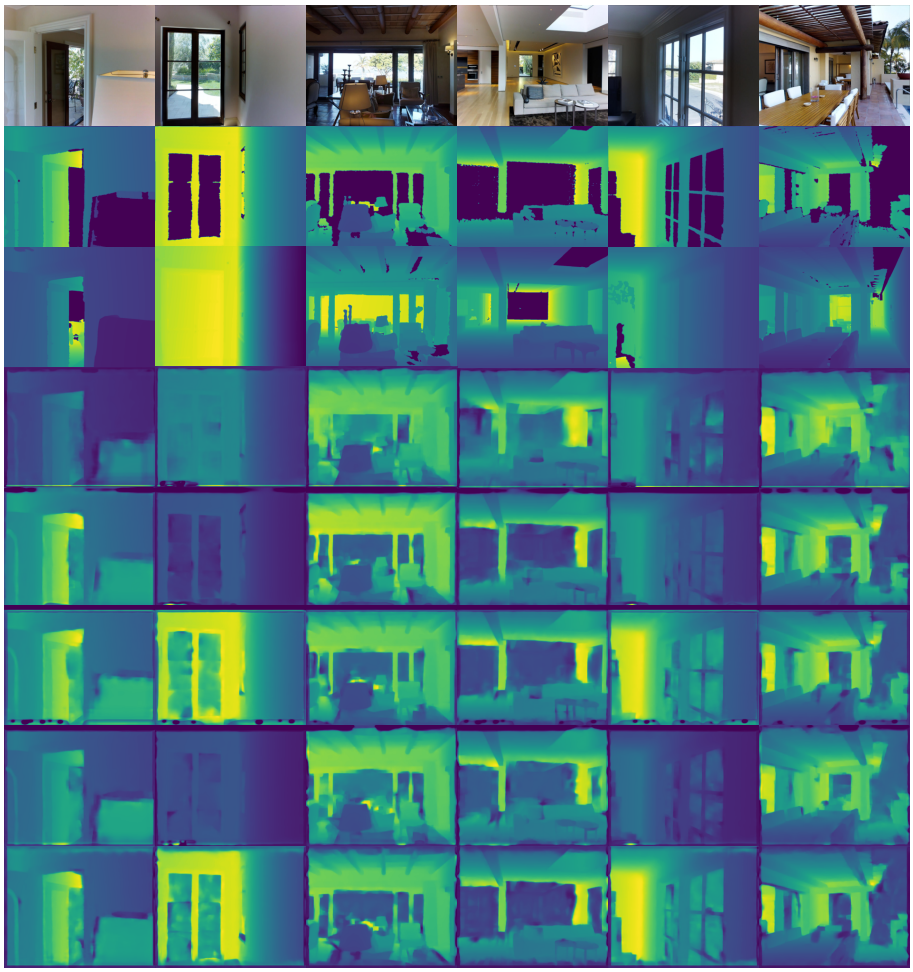

Fig. 8. Results of our full model trained on NYUDv2 while tested on Matterport3D. From top to bottom are the RGB image, the raw depth map, the ground truth depth map, and the results of our full model trained on NYUDv2 with Random, R2, Plastic, Halton and Golden sampling strategy, respectively.

References

1. Chang, A., Dai, A., Funkhouser, T., Halber, M., Niessner, M., Savva, M., Song, S., Zeng, A., Zhang, Y.: Matterport3d: Learning from rgb-d data in indoor environments. International Conference on 3D Vision (3DV) (2017)
2. Cheng, X., Wang, P., Yang, R.: Depth estimation via affinity learned with convolutional spatial propagation network. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 103–119 (2018)
3. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 3354–3361. IEEE (2012)
4. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. Proceedings of the International Conference for Learning Representations (ICLR) (2014)
5. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. In: ACM SIGGRAPH 2004 Papers, pp. 689–694 (2004)
6. Mal, F., Karaman, S.: Sparse-to-dense: Depth prediction from sparse depth samples and a single image. In: 2018 IEEE International Conference on Robotics and Automation (ICRA). pp. 1–8. IEEE (2018)
7. Qiu, J., Cui, Z., Zhang, Y., Zhang, X., Liu, S., Zeng, B., Pollefeys, M.: Deeplidar: Deep surface normal guided depth prediction for outdoor scene from sparse lidar data and single color image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3313–3322 (2019)
8. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from rgbd images. In: European conference on computer vision. pp. 746–760. Springer (2012)
9. Van Gansbeke, W., Neven, D., De Brabandere, B., Van Gool, L.: Sparse and noisy lidar completion with rgb guidance and uncertainty. In: 2019 16th International Conference on Machine Vision Applications (MVA). pp. 1–6. IEEE (2019)
10. Xu, Y., Zhu, X., Shi, J., Zhang, G., Bao, H., Li, H.: Depth completion from sparse lidar data with depth-normal constraints. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2811–2820 (2019)
11. Zhang, Y., Funkhouser, T.: Deep depth completion of a single rgb-d image. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 175–185 (2018)