Supplementary Material for RigNet

Zhiqiang Yan, Kun Wang, Xiang Li, Zhenyu Zhang,
Jun Li✉, and Jian Yang✉

PCA Lab, Nanjing University of Science and Technology, China
{Yanzq,kunwang,xiang.li.implus,junli,csjyang}@njust.edu.cn
zhangjesse@foxmail.com

A Overview

This document provides additional technical details and more visualization results. Concretely, in section B, we first elaborate our pipeline with detailed convolutions. Next, we define the metrics used in paper in section C. Then we conduct additional experiments in section D. Finally, in section E, we provide more visualizations of the final depth predictions on KITTI and NYUv2.

B Detailed Pipeline

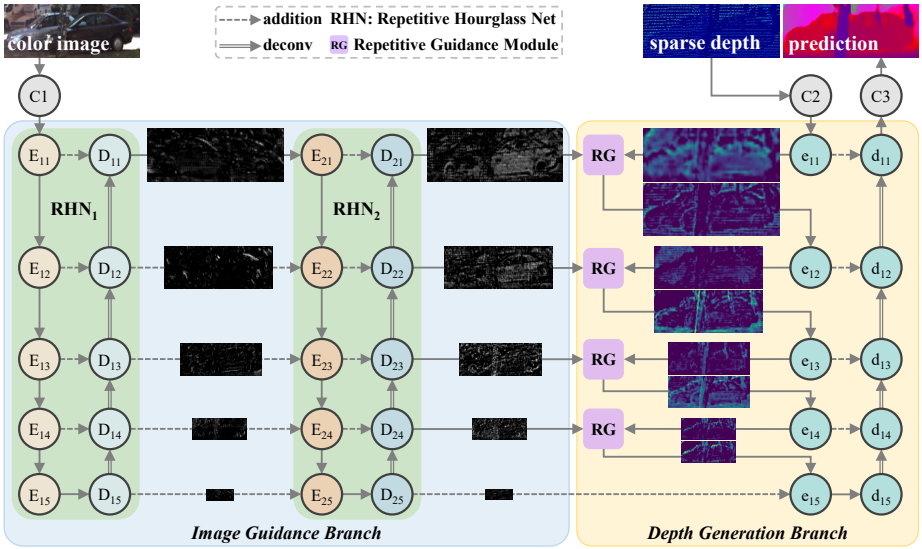
As illustrated in Fig. S1, we provide the detailed convolutions of each layer in our repetitive image guided network. Fig. S2 demonstrates the detailed components of our repetitive guidance module (RG), including the efficient guidance algorithm (EG) and adaptive fusion mechanism (AF).

C Detailed Metrics

In our paper, eight standard metrics are used for evaluation, including RMSE, MAE, iRMSE, iMAE, REL, δ_1, δ_2, and δ_3, which are defined as where GT and

- RMSE : $\sqrt{\frac{1}{m}\sum_{q\in Q_v}\|GT_q - P_q\|^2}$
- iRMSE : $\sqrt{\frac{1}{m}\sum_{q\in Q_v}\left\|\frac{1}{GT_q} - \frac{1}{P_q}\right\|^2}$
- MAE : $\frac{1}{m}\sum_{q\in Q_v}|GT_q - P_q|$
- iMAE : $\frac{1}{m}\sum_{q\in Q_v}\left|\frac{1}{GT_q} - \frac{1}{P_q}\right|$
- REL : $\frac{1}{m}\sum_{q\in Q_v}\left|\frac{GT_q - P_q}{GT_q}\right|$
- δ_i : $\frac{1}{m}\left|\max\left(\frac{GT_q}{P_q}, \frac{P_q}{GT_q}\right) < 1.25^i\right|$

P refer to ground truth depth and predicted depth, respectively. Q_v represents the set of valid pixels in GT. m is the number of the valid pixels.



Conv:
(channel_in, channel_out,
kernel_size, stride, padding)

Deconv:
(channel_in, channel_out,
kernel_size, stride, padding)

C1 : Conv(3, 32, 5, 1, 2)
C2 : Conv(1, 32, 5, 1, 2)
C3 : Conv(64, 1, 3, 1, 1)

e/E_{11} : _make_layer(BasicBlock, 64, 2, 1)
e/E_{12} : _make_layer(BasicBlock, 128, 2, 2)
e/E_{13} : _make_layer(BasicBlock, 256, 2, 2)
e/E_{14} : _make_layer(BasicBlock, 256, 2, 2)
e/E_{15} : _make_layer(BasicBlock, 256, 2, 2)

d/D_{11} : Deconv(128, 64, 3, 2, 1,1)
d/D_{12} : Deconv(256, 128, 3, 2, 1,1)
d/D_{13} : Deconv(256, 256, 3, 2, 1,1)
d/D_{14} : Deconv(256, 256, 3, 2, 1,1)
d/D_{15} : Conv(256, 256, 3, 1, 1)

E_{21} : Conv(64, 32, 3, 1, 1), Conv(32, 32, 3, 1, 1)
E_{22} : Conv(32, 32, 3, 2, 1), Conv(32, 32, 3, 1, 1)
E_{23} : Conv(32, 32, 3, 2, 1), Conv(32, 32, 3, 1, 1)
E_{24} : Conv(32, 32, 3, 2, 1), Conv(32, 32, 3, 1, 1)
E_{25} : Conv(32, 32, 3, 2, 1), Conv(32, 32, 3, 1, 1)

D_{21} : Deconv(32, 32, 3, 2, 1,1)
D_{22} : Deconv(32, 32, 3, 2, 1,1)
D_{23} : Deconv(32, 32, 3, 2, 1,1)
D_{24} : Deconv(32, 32, 3, 2, 1,1)
D_{25} : Conv(32, 32, 3, 1, 1)

Fig. S1. Overview of our repetitive image guided network with detailed convolution layers. The definition of "_make_layer" is similar with that of ResNets.

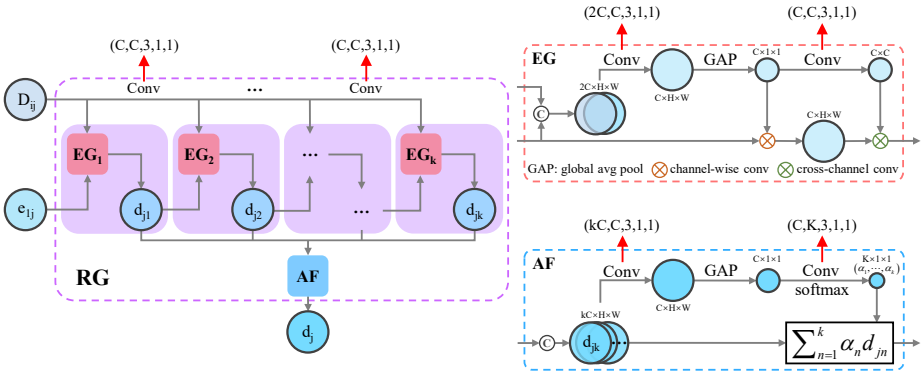


Fig. S2. Our repetitive guidance module (RG) with detailed convolutions, implemented by an efficient guidance algorithm (EG) and an adaptive fusion mechanism (AF).

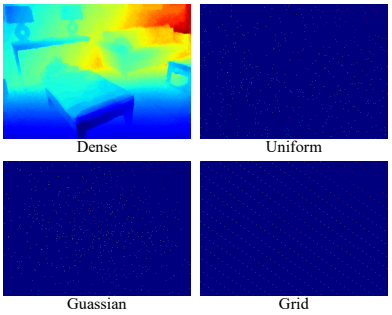


Fig. S3. Examples of different synthetic patterns on NYUv2 dataset.

Pattern	Method	RMSE	REL	δ_1	$\delta_{1.25^2}$	$\delta_{1.25^3}$
Uniform	CSPN [2]	0.117	0.016	99.2	99.9	100.0
	NLSPN [4]	0.092	0.012	99.6	99.9	100.0
	ACMNet [7]	0.105	0.015	99.4	99.9	100.0
	RigNet	0.090	0.013	99.6	99.9	100.0
Gaussian	CSPN [2]	0.121	0.017	99.1	99.8	100.0
	NLSPN [4]	0.093	0.013	99.5	99.9	100.0
	ACMNet [7]	0.110	0.017	99.3	99.9	100.0
	RigNet	0.092	0.012	99.6	99.9	100.0
Grid	CSPN [2]	0.123	0.017	99.2	99.8	100.0
	NLSPN [4]	0.095	0.013	99.5	99.9	100.0
	ACMNet [7]	0.090	0.012	99.6	99.9	100.0
	RigNet	0.087	0.010	99.7	99.9	100.0

Table S1. Performances on NYUv2 dataset with different synthetic patterns.

D More Experiments

In this section, we conduct additional experiments to further validate the generalization capability under different synthetic patterns in Fig. S3.

On NYUv2 test split, we first produce diversiform sparse depth inputs by Uniform, Gaussian, and Grid sampling manners (Fig. S3). Then we compare RigNet with three popular works with released codes and pretrained models, *i.e.*, CSPN [2], NLSPN [4], and ACMNet [7]. Note that all models are pretrained in Uniform sampling mode and fine-tuned on all three patterns. The Uniform sampling only produces 500 valid depth points. As illustrated in Table S1, **(i)** RigNet is almost superior to all other methods in the three patterns. **(ii)** In Gaussian pattern, performances of all four methods drop a little bit, since Gaussian pattern's points around edges of sparse depth map are fewer than Uniform pattern's. **(iii)** In Grid pattern, which can be seen as a simple case of Uniform with regular sampling, ACMNet and RigNet perform better while CSPN and NLSPN do not. These results show RigNet can well tackle depth inputs with different levels of sparsity.

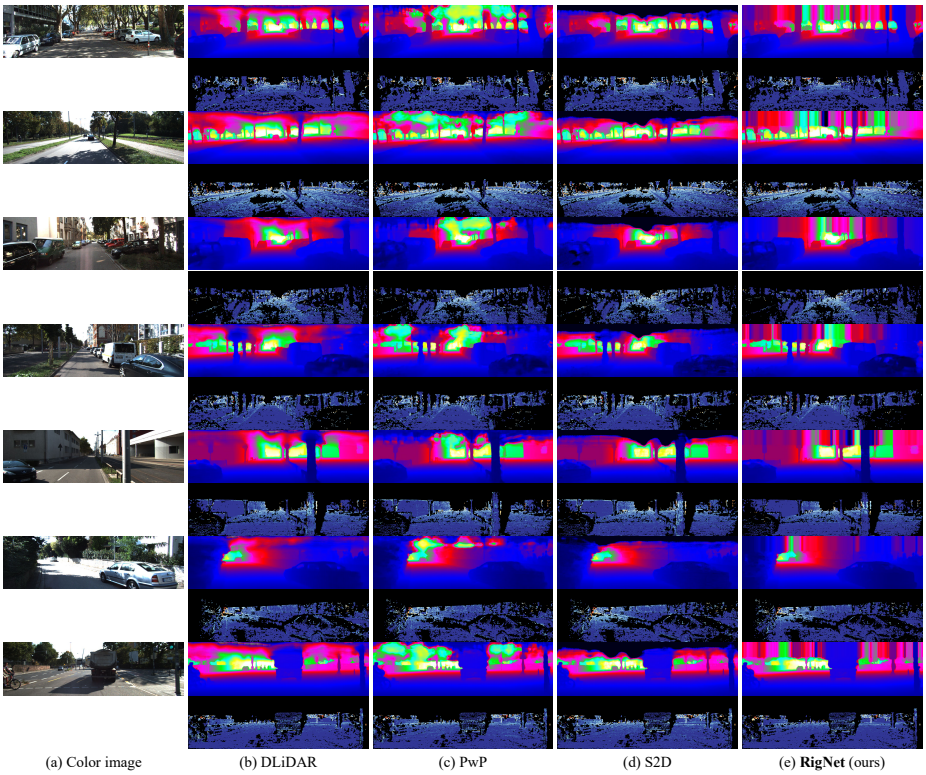


Fig. S4. More visual comparisons with others on KITTI benchmark, including DLiDAR [5], PwP [6], S2D [3], and our RigNet.

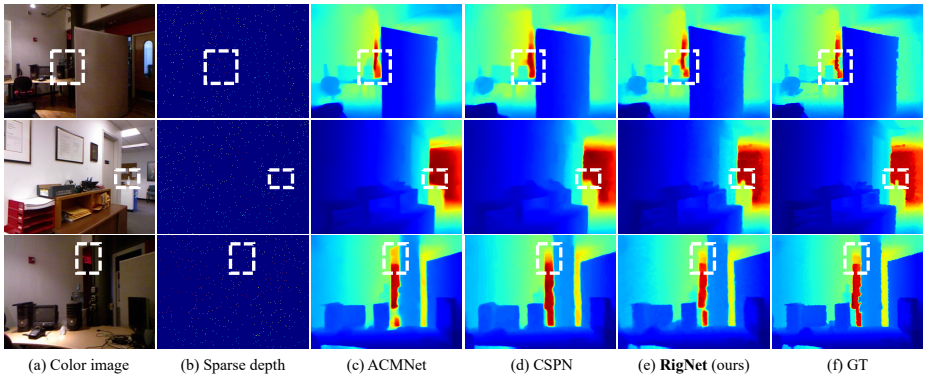


Fig. S5. More visual comparisons with existing methods on NYUv2 dataset, including ACMNet [7], CSPN [1], and our RigNet.

E More Visualizations

In this section, we provide more visualizations of the final depth predictions on KITTI benchmark (Fig. S4) and NYUv2 dataset (Fig. S5). Obviously, our RigNet can produce better results than other depth completion methods.

References

1. Chen, Y., Yang, B., Liang, M., Urtasun, R.: Learning joint 2d-3d representations for depth completion. In: ICCV. pp. 10023–10032 (2019) 4
2. Cheng, X., Wang, P., Yang, R.: Learning depth with convolutional spatial propagation network. In: ECCV. pp. 103–119 (2018) 3
3. Ma, F., Cavalheiro, G.V., Karaman, S.: Self-supervised sparse-to-dense: Self-supervised depth completion from lidar and monocular camera. In: ICRA (2019) 4
4. Park, J., Joo, K., Hu, Z., Liu, C.K., Kweon, I.S.: Non-local spatial propagation network for depth completion. In: ECCV (2020) 3
5. 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: CVPR. pp. 3313–3322 (2019) 4
6. Xu, Y., Zhu, X., Shi, J., Zhang, G., Bao, H., Li, H.: Depth completion from sparse lidar data with depth-normal constraints. In: ICCV. pp. 2811–2820 (2019) 4
7. Zhao, S., Gong, M., Fu, H., Tao, D.: Adaptive context-aware multi-modal network for depth completion. IEEE Transactions on Image Processing (2021) 3, 4