Aggregating Feature Point Cloud for· Depth Completion
Supplementary Materials

Zhu Yu[1], Zehua Sheng[1], Zili Zhou[1], Lun Luo[1],
Si-Yuan Cao[2,1*] Hong Gu[4], Huaqi Zhang[4], Hui-Liang Shen[1,3*]
[1]College of Information Science and Electronic Engineering, Zhejiang University
[2]Ningbo Innovation Center, Zhejiang University
[3]Key Laboratory of Collaborative Sensing and Autonomous Unmanned Systems of Zhejiang Province
[4]vivo Mobile Communication Company Ltd.
{yu_zhu, shengzehua, zhou_zili, luolun, cao_siyuan, shenhl}@zju.edu.cn
{guhong, zhanghuaqi}@vivo.com

The supplementary materials provide more experimental details and more visualization results.

1. Detailed Metrics

In our paper, we adopt eight standard metrics for evaluation, including RMSE, MAE, iRMSE, iMAE, REL, $\delta_i(i = 1.25, 1.25^2, 1.25^3)$. Formally, the metrics are defined as follows:

- Root mean square error (RMSE):

$$\sqrt{\frac{1}{|\mathcal{V}|}\sum_{\mathbf{x}\in\mathcal{V}}\left(\hat{\mathbf{D}}(\mathbf{x})-\mathbf{D}(\mathbf{x})\right)^2} \tag{1}$$

- Mean absolute error (MAE):

$$\frac{1}{|\mathcal{V}|}\sum_{\mathbf{x}\in\mathcal{V}}\left|\hat{\mathbf{D}}(\mathbf{x})-\mathbf{D}(\mathbf{x})\right| \tag{2}$$

- Root mean square error of the inverse depth (iRMSE):

$$\sqrt{\frac{1}{|\mathcal{V}|}\sum_{\mathbf{x}\in\mathcal{V}}\left(\frac{1}{\hat{\mathbf{D}}(\mathbf{x})}-\frac{1}{\mathbf{D}(\mathbf{x})}\right)^2} \tag{3}$$

- Mean absolute error of the inverse depth (iMAE):

$$\frac{1}{|\mathcal{V}|}\sum_{\mathbf{x}\in\mathcal{V}}\left|\frac{1}{\hat{\mathbf{D}}(\mathbf{x})}-\frac{1}{\mathbf{D}(\mathbf{x})}\right| \tag{4}$$

- Mean relative error (REL):

$$\frac{1}{|\mathcal{V}|}\sum_{\mathbf{x}\in\mathcal{V}}\frac{\left|\hat{\mathbf{D}}(\mathbf{x})-\mathbf{D}(\mathbf{x})\right|}{\mathbf{D}(\mathbf{x})} \tag{5}$$

- Thresholded accuracy ($\delta_i(i = 1.25, 1.25^2, 1.25^3)$)

$$\max\left(\frac{\hat{\mathbf{D}}(\mathbf{x})}{\mathbf{D}(\mathbf{x})},\frac{\mathbf{D}(\mathbf{x})}{\hat{\mathbf{D}}(\mathbf{x})}\right) < 1.25^i \tag{6}$$

where $\mathbf{D}$ and $\hat{\mathbf{D}}$ refer to ground truth depth map and predicted depth map, respectively; $\mathbf{x}$ denotes the indexes of depth map; $\mathcal{V}$ represents the set of valid pixels in $\mathbf{D}$; $|\mathcal{V}|$ is the number of valid pixels.

2. Camera Models

Let denote the image coordinate of an pixel as $\mathcal{C} = (u_i, v_i)$, whose corresponding depth value is d_i, and the coordinate in the 3D world system as $\mathcal{P} = (x_i, y_i, z_i)$. Denoting the camera intrinsic matrix as $K \in \mathbb{R}^{4\times4}$, then $\mathcal{C}$ can be transformed to $\mathcal{P}$ as follows

$$[x_i, y_i, z_i, 1]^T = K^{-1}\cdot[u_i \times d_i, v_i \times d_i, d_i, 1]^T. \tag{7}$$

3. More Qualitative Comparison Results

In this section, we provide more visualization results for qualitative comparison analysis.

References

[1] Xinjing Cheng, Peng Wang, and Ruigang Yang. Depth estimation via affinity learned with convolutional spatial propagation network. In *Proceedings of the European Conference on Computer Vision*, pages 103–119, 2018. 2, 3, 4

[2] Mu Hu, Shuling Wang, Bin Li, Shiyu Ning, Li Fan, and Xiaojin Gong. Towards precise and efficient image guided depth completion. 2021. 3, 4

[3] Jaewon Kam, Jungeon Kim, Soongjin Kim, Jaesik Park, and Seungyong Lee. Costdcnet: Cost volume based depth completion for a single rgb-d image. In *Proceedings of the European*

*Corresponding author.

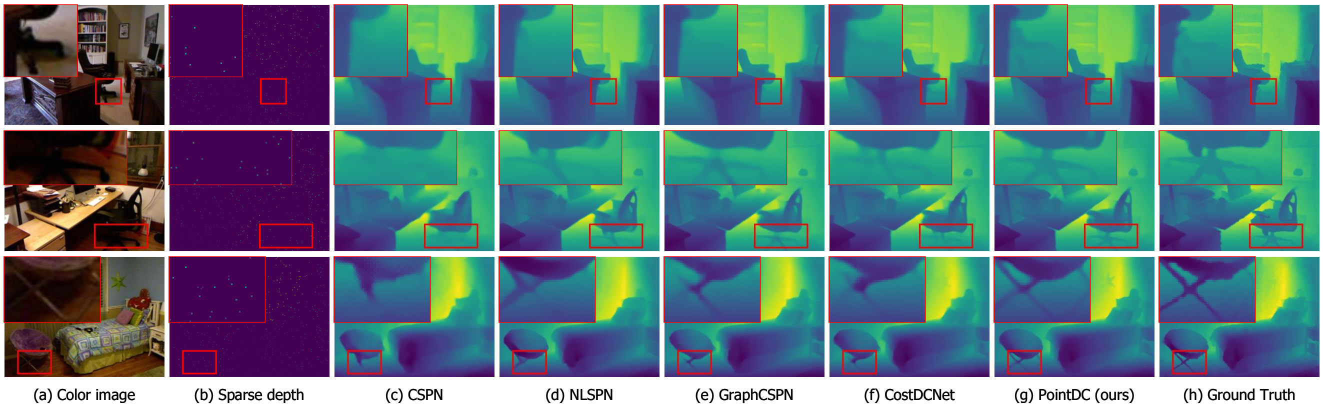

Figure 1. **Qualitative depth completion results on the NYU Depth V2 dataset [6] with 500 sampling points.** (a) Color image, (b) Sparse depth, (c) CSPN [1], (d) NLSPN [5], (e) GraphCSPN [4], (f) CostDCNet [3], (e) PointDC (ours), (h) Ground truth.

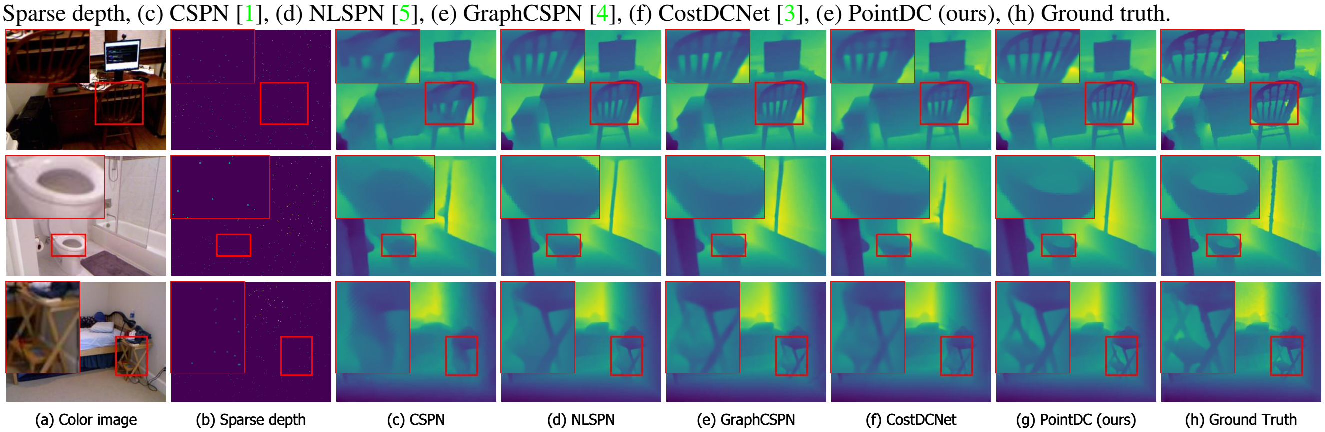

Figure 2. **Qualitative depth completion results on the NYU Depth V2 dataset [6] with 300 sampling points.** (a) Color image, (b) Sparse depth, (c) CSPN [1], (d) NLSPN [5], (e) GraphCSPN [4], (f) CostDCNet [3], (e) PointDC (ours), (h) Ground truth.

Conference on Computer Vision, pages 257–274, 2022. 2, 3, 4

[4] Xin Liu, Xiaofei Shao, Bo Wang, Yali Li, and Shengjin Wang. Graphcspn: Geometry-aware depth completion via dynamic gcns. In *Proceedings of the European Conference on Computer Vision*, pages 90–107, 2022. 2, 3, 4

[5] Jinsun Park, Kyungdon Joo, Zhe Hu, Chi-Kuei Liu, and In So Kweon. Non-local spatial propagation network for depth completion. In *Proceedings of the European Conference on Computer Vision*, pages 120–136, 2020. 2, 3, 4

[6] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgbd images. *Proceedings of the European Conference on Computer Vision*, 7576:746–760, 2012. 2, 3, 4

[7] Jie Tang, Fei-Peng Tian, Wei Feng, Jian Li, and Ping Tan. Learning guided convolutional network for depth completion. *IEEE Transactions on Image Processing*, 30:1116–1129, 2020. 3, 4

[8] Jonas Uhrig, Nick Schneider, Lukas Schneider, Uwe Franke, Thomas Brox, and Andreas Geiger. Sparsity invariant cnns. In *International Conference on 3D Vision*, 2017. 3, 4

[9] Shanshan Zhao, Mingming Gong, Huan Fu, and Dacheng Tao. Adaptive context-aware multi-modal network for depth completion. *IEEE Transactions on Image Processing*, 30:5264–5276, 2021. 3, 4

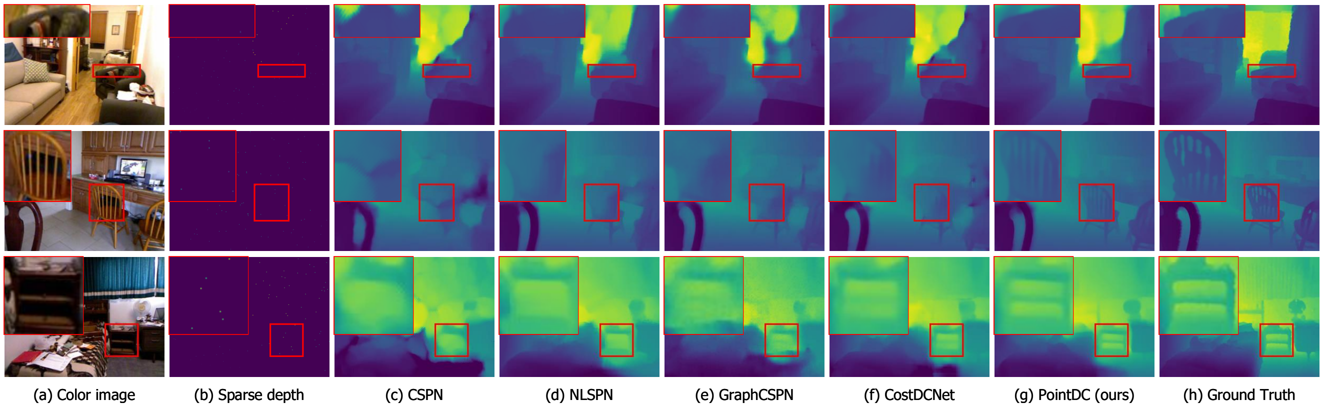

Figure 3. **Qualitative depth completion results on the NYU Depth V2 dataset [6] with 100 sampling points.** (a) Color image, (b) Sparse depth, (c) CSPN [1], (d) NLSPN [5], (e) GraphCSPN [4], (f) CostDCNet [3], (e) PointDC (ours), (h) Ground truth.

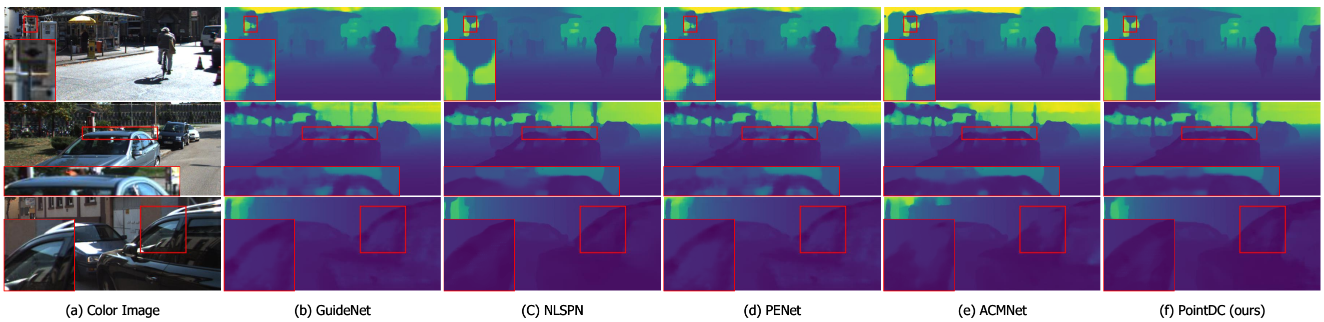

Figure 4. **Qualitative depth completion results on the KITTI DC Dataset [8] with sampling ratio as** 1**.** (a) Color image, (b) GuideNet [7], (d) NLSPN [5], (d) PENet [2], (e) ACMNet [9], (f) PointDC (ours).

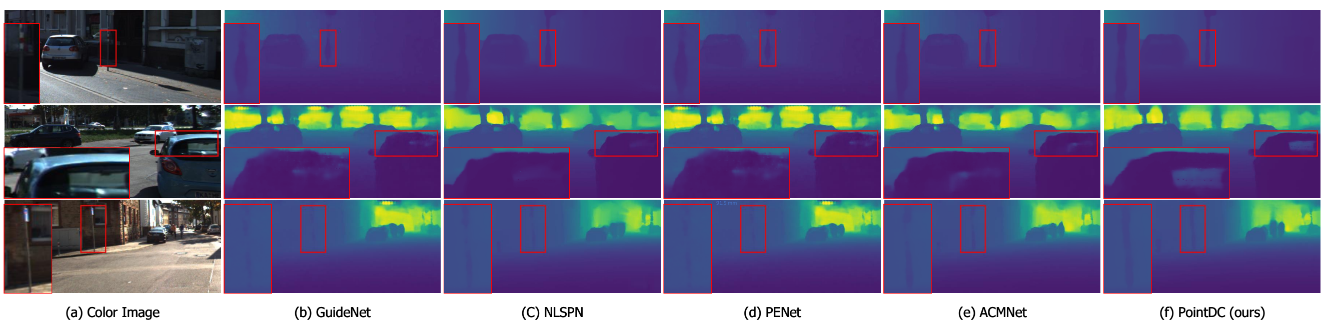

Figure 5. **Qualitative depth completion results on the KITTI DC Dataset [8] with sampling ratio as** 0.4**.** (a) Color image, (b) GuideNet [7], (d) NLSPN [5], (d) PENet [2], (e) ACMNet [9], (f) PointDC (ours).

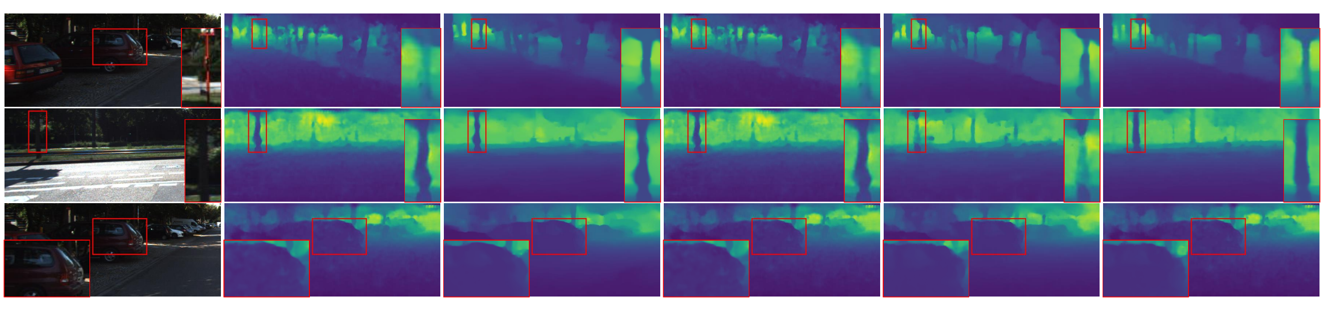

Figure 6. **Qualitative depth completion results on the KITTI DC Dataset [8] with sampling ratio as** 0.1**.** (a) Color image, (b) GuideNet [7], (d) NLSPN [5], (d) PENet [2], (e) ACMNet [9], (f) PointDC (ours).

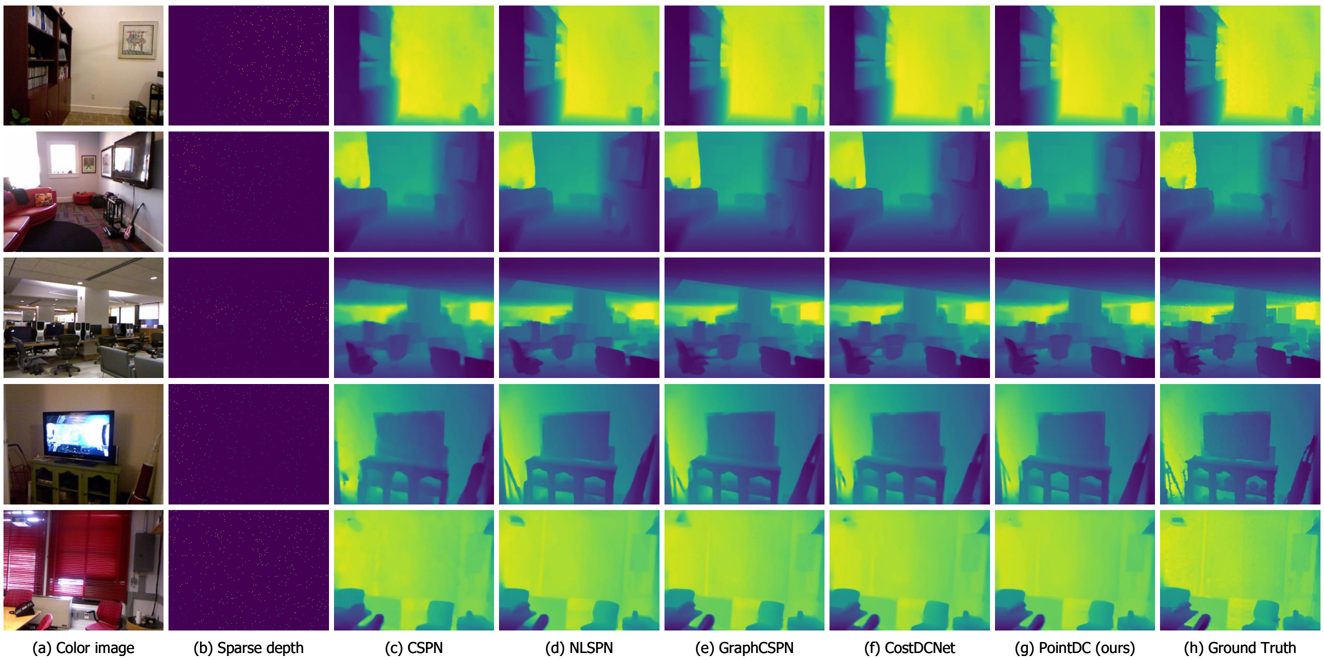

Figure 7. **More qualitative depth completion results on the NYU Depth V2 dataset [6].** (a) Color image, (b) Sparse depth, (c) CSPN [1], (d) NLSPN [5], (e) GraphCSPN [4], (f) CostDCNet [3], (e) PointDC (ours), (h) Ground truth.

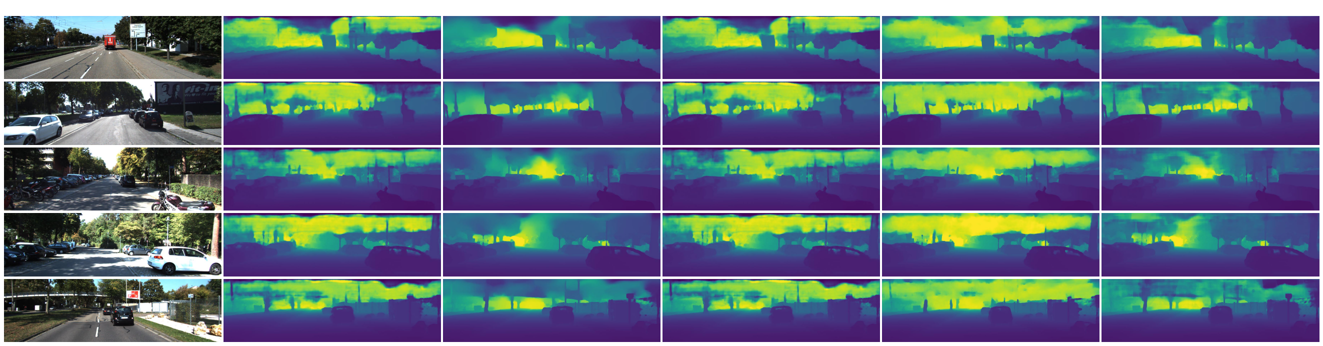

Figure 8. **More qualitative depth completion results on the KITTI DC Dataset [8].** (a) Color image, (b) GuideNet [7], (d) NLSPN [5], (d) PENet [2], (e) ACMNet [9], (f) PointDC (ours).