Depth Completion from Sparse LiDAR Data with Depth-Normal Constraints – Supplementary Materials

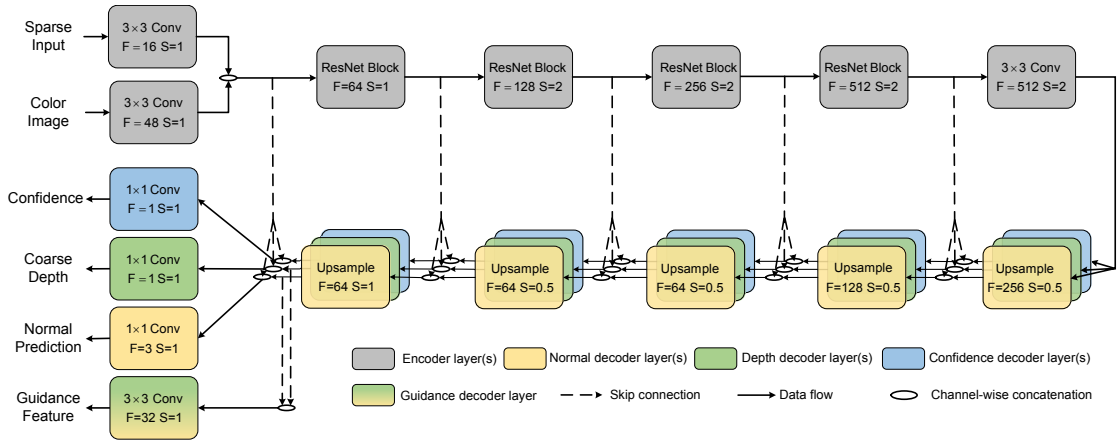

Figure S-1: The specific architecture of our prediction network.

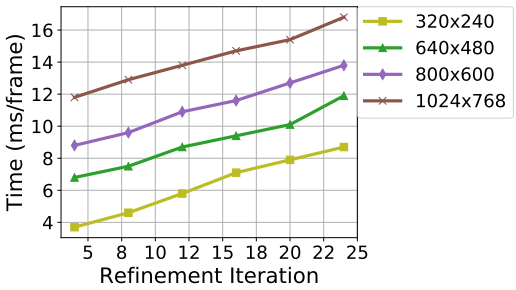

Figure S-2: The running time with different refinement iterations and input resolutions.

	RMSE	MAE	iRMSE	iMAE
Ours+Prev. guidance	859.57	325.03	3.02	1.78
Ours+RGB guidance	844.16	279.74	2.94	1.50
Ours+L1	812.62	221.18	2.25	1.03
Ours+Huber	818.59	227.51	2.27	1.07
Ours	811.07	236.67	2.45	1.11

Table S-1: More ablation experimental results.

1. Prediction Network Architecture

The design of the backbone (prediction network) follows the basic principles adopted by [4]. We employ a variant of ResNet[2] as the encoder and consecutive upsampling layers as the decoder, as illustrated in Fig. S-1. The decoder branches for different modalities (*i.e.*, coarse depth maps, surface normals and confidences of sparse inputs) diverge from the end of the encoder, and skip connections are applied between corresponding layers in encoder and decoders to preserve more details.

2. More Experimental Results

2.1. Running Time Analysis

Compared with the previous CNN-based methods, our framework spends extra running time in the space transformation and the refinement phase. Additional time analysis on these two parts is performed altogether. In the experiment, we fix the diffusion kernel size to 3×3 and test on different input resolutions. Fig. S-2 plots the running time w.r.t the refinement iterations. It can be found that our proposed modules (required for enforcing geometric constraints) can run in real-time, which is suitable to couple with different prediction backbones.

2.2. More Ablation Study

We further demonstrate the effectiveness of proposed modules and cast more light on our method by providing more ablation results in Table S-1. We take the 3 channel color image as the guidance rather than our proposed guidance feature map (Ours+RGB guidance) and see an inferior performance due to the absence of geometric information in color images. It is also been validated that the model with guidance feature map generated from last decoder layers (Ours) achieves better performance than that with feature maps from previous layers (Ours+Prev. guidance). Moreover, we test different loss functions, *i.e.*, L1 loss (Ours+L1) and Huber loss (Ours+Huber). Compared with using L2

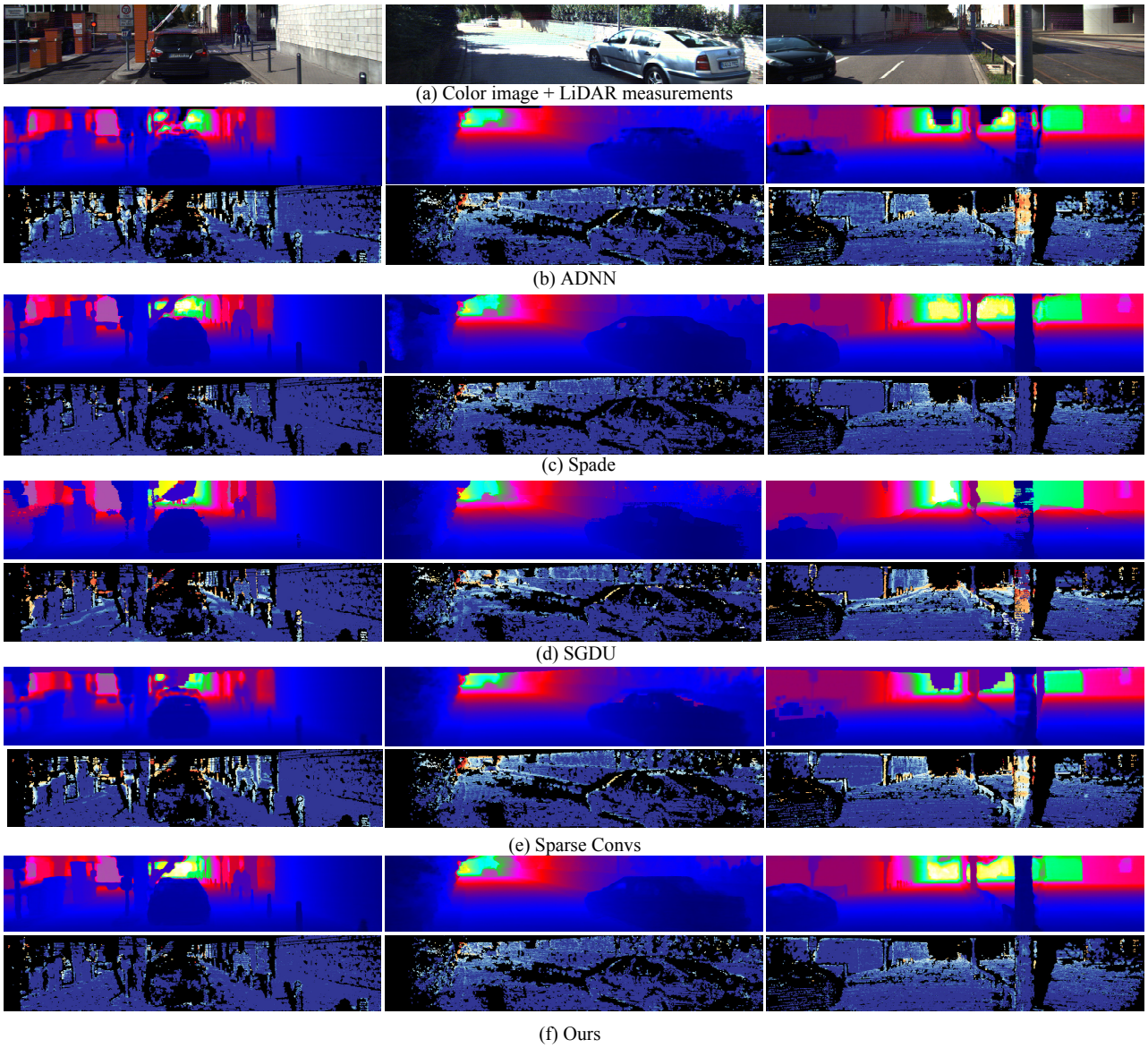

Figure S-3: More quantitative comparison results with other methods *i.e.*, ADNN [1], Spade [3], SGDU [5], and Sparse Convs [6]. For each method, we provide the whole completion results as well as error maps for better comparison.

loss (Ours), L1 and Huber loss achieve a lower MAE while a little bit higher RMSE as illustrated in Table S-1.

2.3. More Qualitative Comparison Results on KITTI

Fig. S-3 demonstrates more qualitative comparison results with the other latest competitive methods.

2.4. More Qualitative Results on NYU

Our method achieves good generalization capability to indoor scenes as well. Fig. S-4 provides more qualitative evaluation results on NYU-Depth-v2 dataset. As we can see, the prediction network (in first phase) occasionally generates illogical outputs (in coarse depth map) especially near the sparse depth inputs. Our refinement network regularizes the depth completion with the constraints between depths and surface normals. The accurate sparse inputs and the skeletons (jointly built by the surface normals and coarse depths) complement each other in the diffusion process, thus the refined depth maps are much more accurate and smoother.

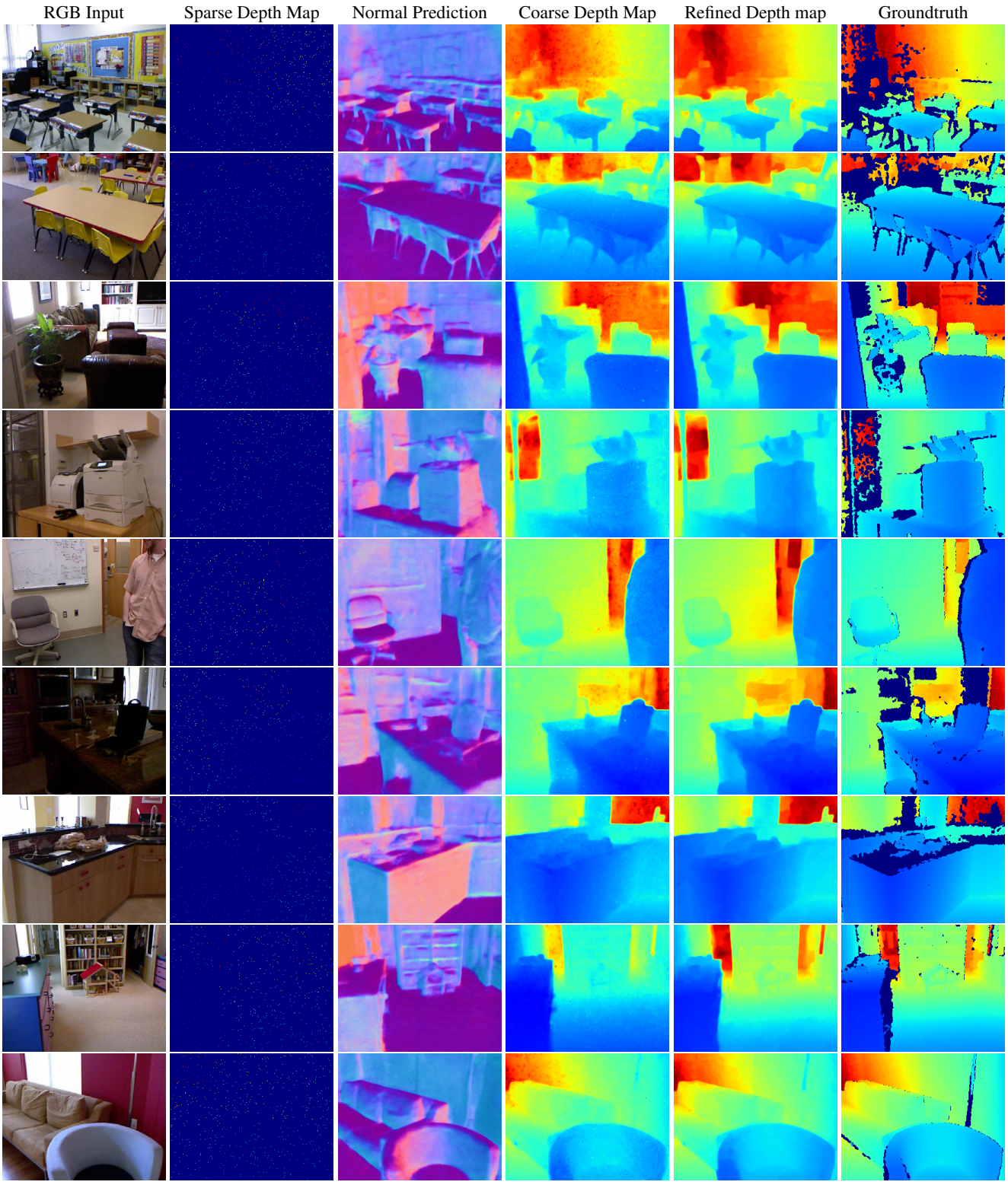

Figure S-4: More Qualitative evaluation results on NYU-Depth-v2 dataset. Each sparse depth map contains about 500 points randomly sampled from the denser groundtruth. Zoom in for better vision.

References

[1] N. Chodosh, C. Wang, and S. Lucey. Deep convolutional compressed sensing for lidar depth completion. In *Asian Conference on Computer Vision*, pages 499–513. Springer, 2018.

[2] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[3] M. Jaritz, R. de Charette, E. Wirbel, X. Perrotton, and F. Nashashibi. Sparse and dense data with cnns: Depth completion and semantic segmentation. *arXiv preprint arXiv:1808.00769*, 2018.

[4] F. Mal and S. Karaman. Sparse-to-dense: Depth prediction from sparse depth samples and a single image. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–8. IEEE, 2018.

[5] N. Schneider, L. Schneider, P. Pinggera, U. Franke, M. Pollefeys, and C. Stiller. Semantically guided depth upsampling. In *German Conference on Pattern Recognition*, pages 37–48. Springer, 2016.

[6] J. Uhrig, N. Schneider, L. Schneider, U. Franke, T. Brox, and A. Geiger. Sparsity invariant cnns. In *2017 International Conference on 3D Vision (3DV)*, pages 11–20. IEEE, 2017.