PacGDC: Label-Efficient Generalizable Depth Completion with Projection Ambiguity and Consistency

Supplementary Material

6. More Implementation Details

Training Details. *Zero-shot Depth Completion:* The training data simply concentrate all available training datasets following [50, 51], without any explicit balancing strategies as in [82]. Due to resource constraints, the training resolution is set to 320×320, though higher resolutions could further enhance performance, as observed in image analysis tasks [24, 59].

Due to the challenges posed by our highly diverse data setting, we modify the 2×2 convolutions in the up/downsampling layers of SPNet to 3×3 convolutions, as odd-sized kernels generally provide better stability [60]. This modification results in a slight increase in computational cost, as shown in Tab. 8. Notably, "Ours-T" even outperforms "SPNet-L" while requiring half the inference time and only 17% of the parameters.

Additionally, we impose a constraint that the minimum number of sparse depth pixels during training is two. This allows us to simplify the absolute term in the G2-MonoDepth loss [50] to L1 loss. The updated loss function $\mathbb{L}$, which measures the discrepancy between predictions $\widetilde{d}$ and our pseudo depth labels $\hat{d}$, is expressed as follows:

$$\begin{aligned}\mathbb{L}(\widetilde{d},\hat{d}) = &\frac{1}{\eta}\sum_{i=1}^{\eta}|T(\widetilde{d}_i)-T(\hat{d}_i)| + \frac{1}{\eta}\sum_{i=1}^{\eta}|\widetilde{d}_i-\hat{d}_i| \\ &+\frac{0.5}{\eta}\sum_{r=0}^{3}\sum_{i=1}^{\eta}|\nabla(\rho_r(T(\widetilde{d}_i)-T(\hat{d}_i)))|,\end{aligned} \tag{4}$$

where T is the standardize operation with mean deviation in [50]. The function ρ_r is the nearest neighbor interpolation at the $1/2^r$ resolution. ∇ is the Sobel gradient in height and width directions. η denotes the number of valid pixels in dense labels.

Few-shot Depth Completion: In our few-shot experiments, we do not employ additional refinement strategies, such as SPN-like modules [7, 29, 45, 54] or depth enhancement methods [48, 49]. This ensures that our model retains SPNet's efficiency. The training resolution is set to a randomly cropped 256×1216. The loss function is updated to the commonly used L1+L2 loss, following the standard practice in most intra-domain learning methods [54, 55, 78].

Testing Details. The details of the test datasets are provided in Tab. 7. For the uniform sampling experiment, test images are resized to a height of 320 pixels. In the sensor-captured experiment, the VOID and KITTI datasets follow standard protocols, with VOID maintaining its original resolution of 480×640 and KITTI using a bottom center-cropped resolution of 256×1216. The final results on KITTI are obtained by averaging predictions from both original and horizontally flipped inputs following implementations in [54, 55].

Datasets	Indoor	Outdoor	Label	Size
ETH3D [40]	✓	✓	Laser	454
Ibims [20]	✓		Laser	100
NYUv2 [42]	✓		RGB-D	654
DIODE [46]	✓	✓	Laser	771
Sintel [3]	✓	✓	Synthetic	1064
KITTI [11]		✓	Stereo	1000
VOID [58]	✓	✓	RGB-D	800

Table 7. The details of test datasets.

Method	Speed↑ (Image/s)	Param.↓ (M)	Memo.↓ (MB)	RMSE↓ (mm)	MAE↓ (mm)
SPNet-T	**126.6**	**35.0**	330	2342	857
Ours-T	121.8	39.7	**242**	**2143**	**792**
SPNet-L	**60.2**	**235.5**	**1176**	2271	791
Ours-L	58.7	254.4	1246	**1966**	**731**

Table 8. The inference costs under "Tiny" (T) and "Large" (L) configurations, including speed, parameters, and memory usage. Notably, the results of "Ours-T" are copied from the ablation study only using 25% training data (in gray).

7. More Quantitative Results

Zero-shot Depth Completion on DDAD Dataset. We further evaluate PacGDC on the DDAD [12] dataset, comparing to more generalizable and supervised baselines, following the standard protocol of VPP4DC [1]. The baseline results are directly taken from relevant papers. As shown in Tab. 9, the results further validate the effectiveness of PacGDC for zero-shot generalization.

Method	*RMSE* ↓	*MAE* ↓	Method	*RMSE* ↓	*MAE* ↓
BP-Net [45]	8903	2712	Marigold-DC [47]	6449	2364
VPP4DC [1]	10247	2290	DMD^{3}C [23]	6609	1842
OGNI-DC [81]	6876	1867	**Ours**	**5918**	**1140**

Table 9. **Zero-shot depth completion** on DDAD dataset under VPP4DC protocol.

Few-shot Comparison with Other Baselines. We supplement Tab. 4 with additional few-shot baselines. The base-

line results on KITTI validation set are directly taken from their original papers. As shown in Tab. 10, the results further demonstrate the superiority of our model in few-shot depth completion.

Method	Shot	*RMSE* ↓	*MAE* ↓	*iRMSE* ↓	*iMAE* ↓
DepthPrompt [32]	100	1798	602	-	-
Ours	100	**911**	**229**	**2.54**	**0.96**
DDPMDC [35]	11000	966	291	3.63	1.48
Ours	1000	**830**	**220**	**2.28**	**0.91**

Table 10. **Few-shot depth completion** on KITTI with 64 line LiDAR, supplementing Tab. 4.

In-Domain Evaluation on the KITTI Dataset. We further conduct standard in-domain evaluation by fine-tuning the pre-trained zero-shot PacGDC model on the entire KITTI training set (*i.e.,* 86K samples). As presented in Tab. 11, despite adopting a plain backbone without specialized components such as spatial propagation networks (SPNs), PacGDC delivers competitive performance on the KITTI validation set, comparable to recent state-of-the-art methods. Moreover, we submit the results of the fully fine-tuned model to the official KITTI test set leaderboard.

Method	Plain	*RMSE* ↓	*MAE* ↓	*iRMSE* ↓	*iMAE* ↓
BEV@DC [79]		720	**187**	**1.88**	**0.80**
TPVD [68]		**719**	187	-	-
BEV@DC [79]	✓	762	198	2.06	0.86
TPVD [68]	✓	764	**198**	-	-
UniDC [30]	✓	824	209	-	-
Ours	✓	**759**	203	**2.06**	**0.85**

Table 11. **In-domain evaluation** on KITTI validation set.

8. More Ablation Study

Different Depth Foundation Models. We evaluate our approach with four different depth foundation models: DepthAnything (DA) [69], DepthPro [2], DepthAnythingV2 (DAV2) [70], and DistillAnyDepth (DistillAD) [14]. As shown in Tab. 12, PacGDC consistently yields performance improvements over the baseline (without PacGDC), further validating the generality and effectiveness of our method.

It is worth noting that this experiment was newly introduced in response to reviewer feedback. Accordingly, our "Large" model continues to use DA and DepthPro, as reported in Tab. 6, rather than the combination of DA, DepthPro, and DAV2 used in Tab. 12.

9. More Visual Results

Zero-Shot Depth Completion. We further provide visual examples of zero-shot scenarios in Fig. 8, covering a range of datasets and sparsity levels: DIODE with 1% sparsity, ETH3D with 10% sparsity, KITTI with 4-line LiDAR, and VOID with 1500 feature points derived from a VIO system. Across these scenarios, characterized by diverse scene semantics, varying scales, and different forms of depth sparsity, PacGDC consistently achieves higher accuracy in predicting metric depth maps compared to existing baselines.

DA [69]	DepthPro [2]	DAV2 [70]	DistillAD [14]	*RMSE* ↓	*MAE* ↓
				2484	990
✓				2277	857
✓	✓			2241	854
	✓	✓		2243	852
		✓	✓	2276	859
✓	✓	✓		**2232**	**848**
✓	✓	✓	✓	2279	874

Table 12. **Ablation study** on different depth foundation models. Results with our data synthesis pipeline are shaded in gray.

Few-Shot Depth Completion. Visual results for few-shot scenarios are presented in Figs. 9 and 10, using models trained with 1, 10, 100, and 1000 samples. To provide a comprehensive analysis, we also separately showcase results for 8-, 16-, 32-, and 64-line LiDAR inputs under the same few-shot training settings. Leveraging the strong pre-trained weights from our synthesis pipeline, our model demonstrates significant qualitative improvements over in-domain learning baselines across all levels of supervision.

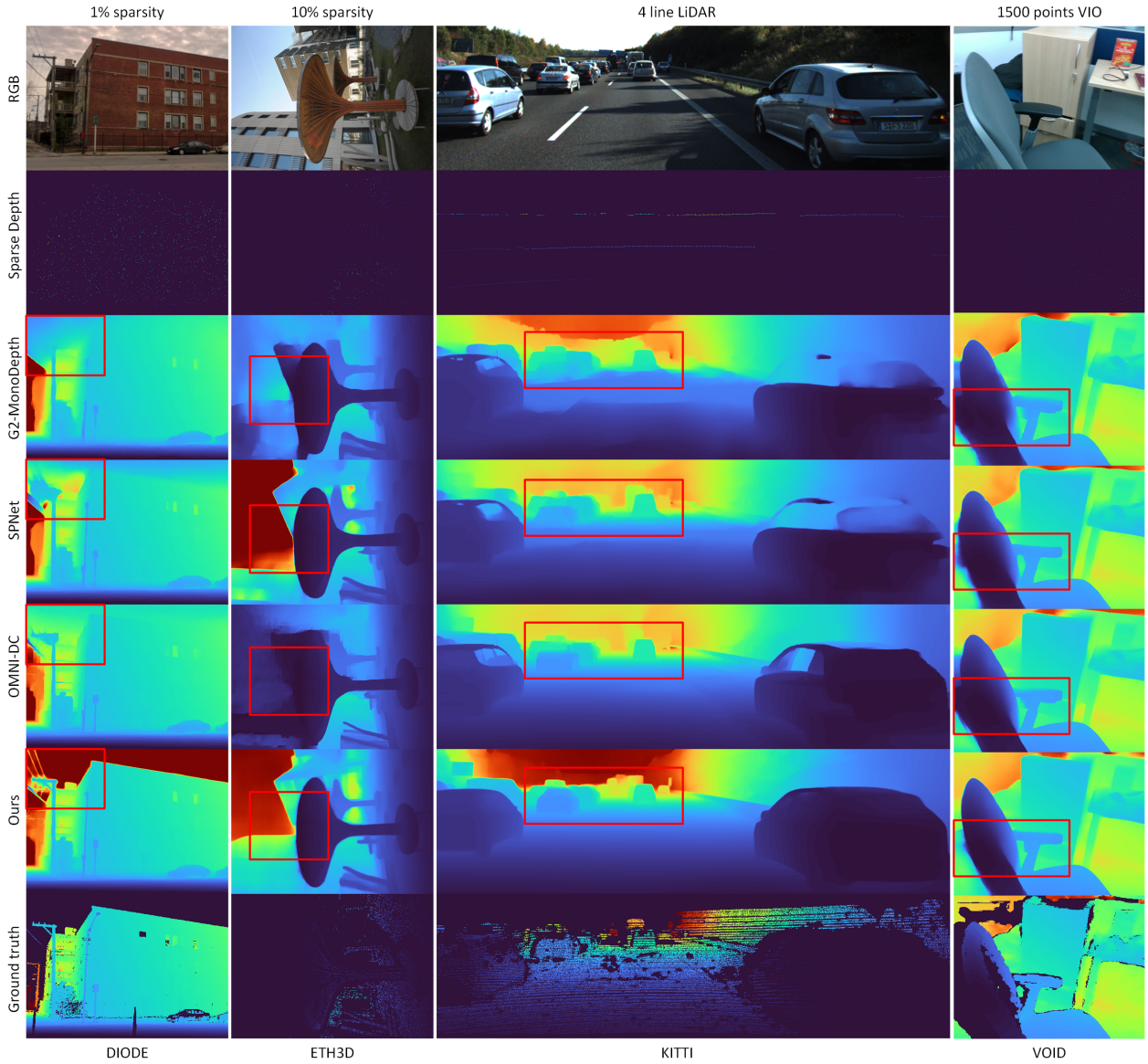

Figure 8. **Zero-shot depth completion** on unseen scenarios with different scene semantics/scales and depth sparsity/patterns.

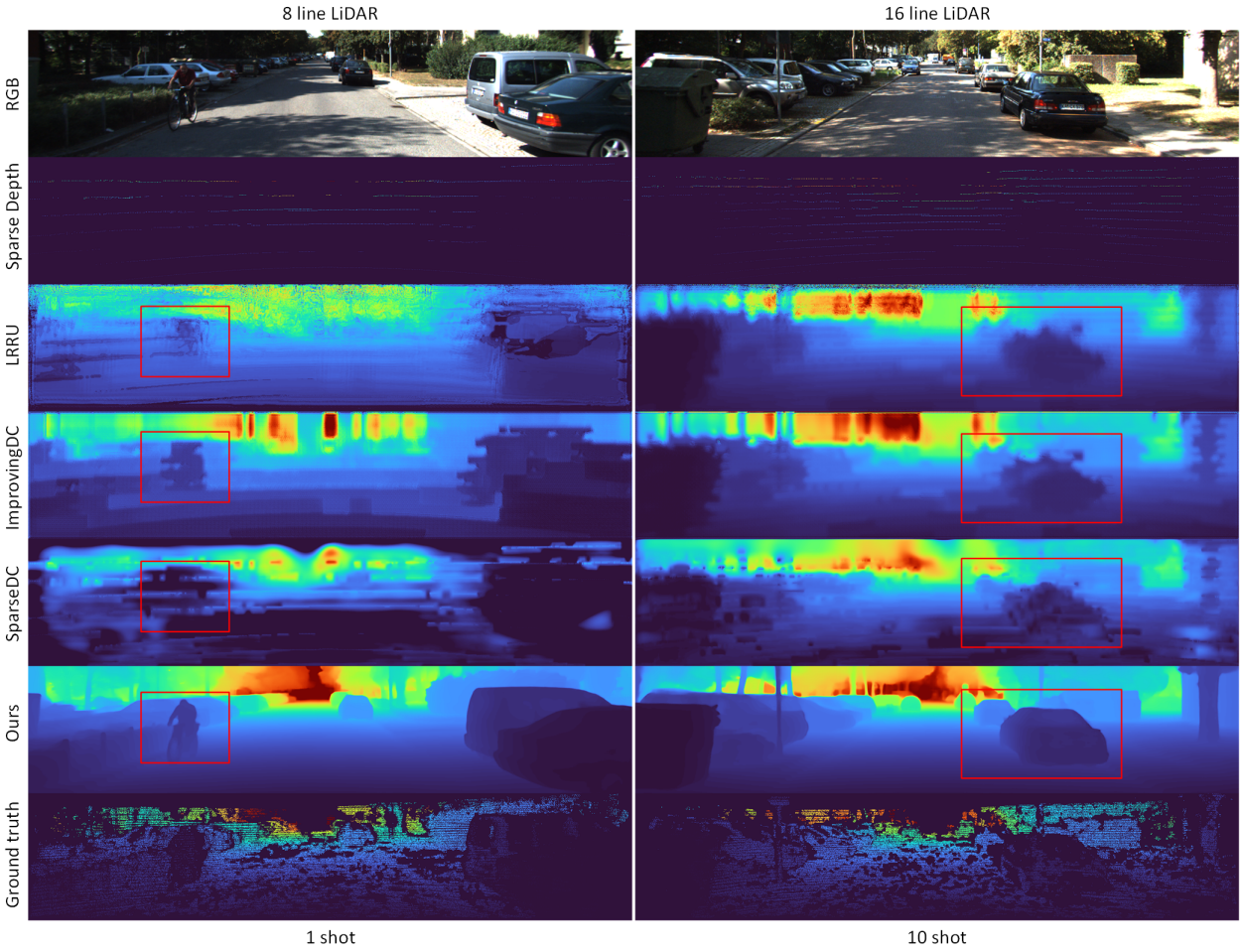

Figure 9. **Few-shot depth completion** on KITTI with 8- and 16-lines LiDAR, using models trained with 1 and 10 samples, respectively.

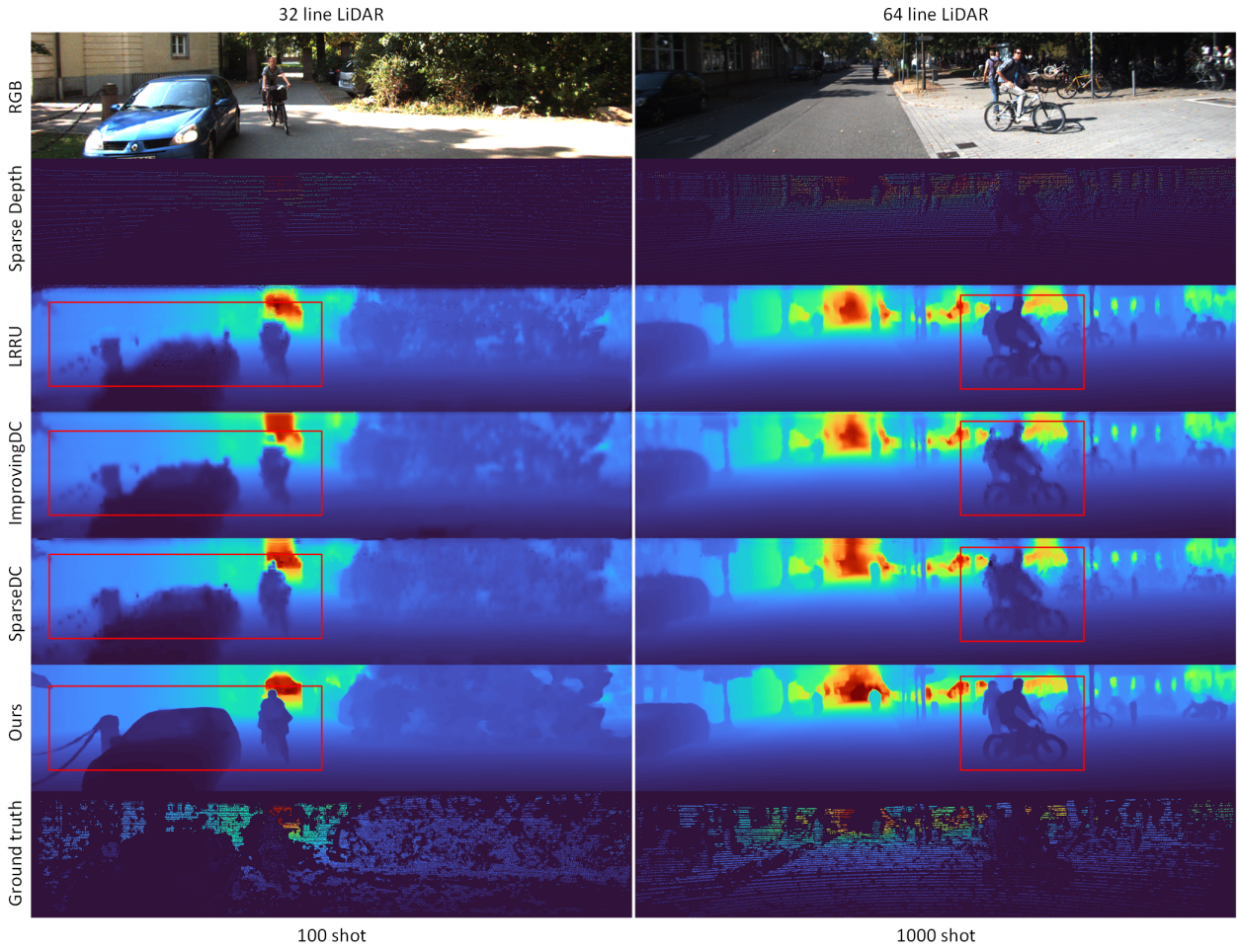

Figure 10. **Few-shot depth completion** on KITTI with 32- and 64-lines LiDAR, using models trained with 100 and 1000 samples, respectively.