8. LiDAR Novel View Synthesis Experimental Details

We extend the novel view synthesis comparison to more scenes in Nuscenes. In the main paper, we selected `scene-0103` as the candidate choice of scene due to sensor pose variations along the z-axis. We show LiDAR NVS results on 3 more scenes from the mini-val dataset in Table 5. We compare SMORE against NeuRAD by considering 3 scenarios- first: using NuScenes poses with no pose optimization, second: optimizing poses and third: directly using SMOR optimized poses. We observe that our method consistently outperforms NeuRAD by an order of magnitude on chamfer distance as well as median L2 depth across all scenes.

Method	Chamfer Dist. ↓	Depth ↓
scene-0061		
NeuRAD[33] w. Nuscenes Poses	1.81	0.0081
NeuRAD[33] w. Ego-pose Optimization	2.87	0.0102
NeuRAD[33] w. Our Poses	1.98	0.0022
Ours	**0.36**	**0.0005**
scene-0103		
NeuRAD[33] w. Nuscenes Poses	4.22	0.0170
NeuRAD[33] w. Ego-pose Optimization	3.26	0.0053
NeuRAD[33] w. Our Poses	2.19	0.0020
Ours	**0.26**	**0.0002**
scene-0796		
NeuRAD[33] w. Nuscenes Poses	3.80	0.0546
NeuRAD[33] w. Ego-pose Optimization	3.68	0.0245
NeuRAD[33] w. Our Poses	3.60	0.0177
Ours	**0.31**	**0.0002**
scene-1094		
NeuRAD[33] w. Nuscenes Poses	2.53	0.0566
NeuRAD[33] w. Ego-pose Optimization	3.17	0.0452
NeuRAD[33] w. Our Poses	2.41	0.0120
Ours	**0.43**	**0.0007**

Table 5. **LiDAR Novel View Synthesis on more scenes from NuScenes**. SMORE consistently outperforms NeuRAD across a variety of scenes by an order of magnitude.

9. Ego-Pose Evaluation

We conduct experiments on more scenes to evaluate the efficacy of ego-poses recovered by SMORE's optimization. Similar to the main paper, we observe that poses generated by our method significantly outperform the NuScenes ground-truth as well as those from a LiDAR odometry [38] across all scenes (See Table 6).

Pose Source	PSNR ↑	SSIM ↑	LPIPS ↓	CD ↓	Depth ↓
scene-0061					
Nuscenes-GT	24.93	0.752	0.357	1.81	0.0081
KISS-ICP[38]	23.99	0.729	0.396	2.65	0.0139
Ours	**25.98**	**0.781**	**0.306**	**1.98**	**0.0022**
scene-0103					
Nuscenes-GT	26.37	0.791	0.283	4.22	0.017
KISS-ICP[38]	24.99	0.76	0.328	2.58	0.016
Ours	**27.52**	**0.821**	**0.238**	**2.19**	**0.002**
scene-0796					
Nuscenes-GT	22.25	0.629	0.528	3.80	0.0546
KISS-ICP[38]	21.63	0.622	0.544	3.95	0.0388
Ours	**22.44**	**0.647**	**0.514**	**3.60**	**0.0177**
scene-1094					
Nuscenes-GT	23.59	0.512	0.526	2.53	0.0566
KISS-ICP[38]	22.93	0.505	0.585	2.75	0.0501
Ours	**24.28**	**0.528**	**0.492**	**2.41**	**0.0120**

Table 6. **Ego-pose evaluation by fitting NeuRAD on more scenes from NuScenes**. SMORE's refined poses provide improved geometry estimates when compared to other poses, including even the ground truth poses provided by NuScenes.

10. Bounding Box Evaluation Details

We used the standard NuScenes object detection metric, average translation error, to measure our method's improvement of the ground-truth bounding boxes. However, as our method is not an object detector, that comparison has some complications, which we explain here.

The average translation error is defined as the distance between the centers of the predicted and ground truth bounding boxes. Since our method does not predict bounding boxes, first, we need to define a "center" for them. The center has no meaning for our reconstruction, so we can choose any fixed point on each object. Specifically, we choose the point that minimizes the sum of square distances to the centers of all the input bounding boxes. Note that when we subsample the inputs for evaluation, we do not use the held-out boxes to determine the "predicted center".

Next, we must define what constitutes a "detection" for our algorithm. Our reconstructions are formed by aggregating many points over multiple sweeps, which are registered to the predicted surfaces. Due to the labeling procedure of NuScenes, some of these input bounding boxes contain very few LiDAR returns (< 50). The lack of points causes ambiguities in the registration step and can lead to instabilities, so we drop them from the optimization. In table Tab. 2, we show the results on only boxes that have been optimized by our method.

	NN Dist (m) ↓	Acc Relax ↑	Acc Strict ↑
NKSR[12] + GT tracks (10Hz) + GT ego-pose	0.086	0.89	0.76
NKSR[12] + LT3D[27] tracks + GT ego-pose	0.088	0.89	0.73
Ours + GT tracks (10 Hz) + KISS ego-pose[38]	0.074	0.93	0.82
Ours + GT tracks (10 Hz) + GT ego-pose	0.079	0.93	0.81
Ours + GT tracks (5 Hz) + GT ego-pose	0.073	0.93	0.83
Ours + GT tracks (2.5 Hz) + GT ego-pose	0.073	0.93	0.83
Ours + LT3D[27] tracks + GT ego-pose	0.083	0.92	0.79

Table 7. Surface quality evaluation on Argoverse 2.0, measured by comparing the LiDAR points to their closest points on the reconstructed surfaces.

11. Argoverse 2.0 Evaluation

We replicated the same robustness evaluation done on NuScenes on Argoverse 2.0. As with NuScenes we use a small subset of the validation dataset for our evaluation. Specifically, sequences: a7636fca-4d9e-3052-bef2-af0ce5d1df74, 0c3bad78-9f1e-395d-a376-2eb7499229fd, e50e7698-de3d-355f-aca2-eddd09c09533, 0aa4e8f5-2f9a-39a1-8f80-c2fdde4405a2 d770f926-bca8-31de-9790-73fbb7b6a890.

As with NuScenes, we tested our method with various modifications to the inputs, either downsampling the ground truth annotations or by using tracked produced by LT3d[27]. The results can be found in tables Tab. 7 and reconfirm our main findings in the NuScenes results: our method can produce high-quality reconstructions even with input annotations of significantly worse quality than the ground truth. Again, we see a significant improvement over reconstructions using the ground truth poses.

12. Failure Cases

Ground Holes in AV2 Background Reconstructions: We find (and show in Fig. 8) that the ground surface we extract from Argoverse is not as complete as those we extract from NuScenes. We believe that this is the result of the orientation of the LiDAR lasers used in each dataset collection. The LiDAR lasers in AV2 are oriented such that they focus the resolution "down-range" to make detecting vehicles and pedestrians easier. This results in less resolution on the ground. To see this, compare the distance between laser returns near the car in NuScenes and Argoverse in Fig. 5. Despite this, we still believe that good reconstructions of the ground should be possible and investigating this is an area of future research.

Registration Failures: Another source of errors for our method is when ICP produces a poor registration on a vehicle. One common cause of this is attempting to register a sweep to a vehicle that contains very few points. We use a heuristic to filter out most of these cases (dropping views of an object with fewer than 50 points) but some can still cause errors which manifest as "jittery" motion of objects. Another, harder to filter, source of error is from registering scans with low "texture". In the context of ICP, low texture means scans which do not contain corners or edges useful for exact alignment. This can occur when only the side face of a vehicle is observed, resulting in a flat plane of points which has many possible alignments to the reconstructed shape. We believe that both of these errors can be mitigated by applying stronger motion priors to the reconstructed objects in order to add constraints to the system. This is another direction for future work.