Table 1: Quality comparison of novel view synthesis for the *D-NeRF* dataset. The training times except ours are from AP-NeRF.

Method	Reposable	PSNR↑	SSIM↑	LPIPS↓	FPS↑	Resolution	Training Time
D-NeRF	No	30.48	0.9683	0.0450	< 1	400×400	20 hours
TiNeuVox-B	No	32.60	0.9783	0.0383	0.82	400×400	28 mins
Hexplane	No	29.81	0.9683	0.0400	1.37	400×400	11.5mins
K-Plane hybrid	No	31.02	0.9717	0.0495	0.52	400×400	52mins
4D-GS	No	34.39	0.9830	0.0190	141.37	800×800	-
Deformable-3D-GS	No	40.11	0.9918	0.0120	42.10	800×800	-
SC-GS	No	43.04	0.9751	-	-	400×400	-
WIM	Yes	25.21	0.9383	0.0700	0.16	400×400	11 hours
AP-NeRF	Yes	30.91	0.9700	0.0350	1.33	400×400	2.5 hours
Ours	Yes	34.49	0.9751	0.0127	119.61	800×800	6.39 hours
Ours	Yes	35.57	0.9797	0.0083	120.42	400×400	6.68 hours

Table 2: Ablation study for the number of initial superpoint M on the 'jumpingjacks' scene of *D-NeRF* dataset.

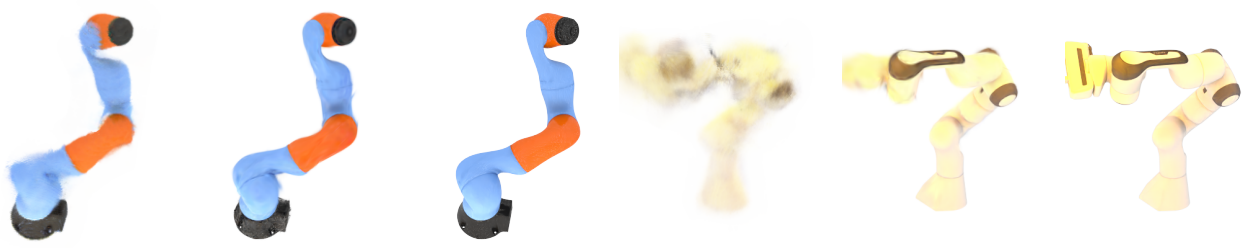
M	128	256	384	512	640	768	896	1024
PSNR↑	31.24	32.94	33.38	33.61	33.21	33.36	33.26	33.27
SSIM↑	0.9744	0.9796	0.9810	0.9801	0.9785	0.9794	0.9788	0.9788
LPIPS↓	0.0134	0.0113	0.0094	0.0090	0.0094	0.0097	0.0089	0.0096
FPS↑	120.80	118.10	106.67	130.77	107.73	104.59	120.76	119.67

Table 3: Optimization time and required resources for each scene in the *D-NeRF* dataset.

scene	hellwarrior	hook	jumpingjacks	mutant	standup	trex	average
Training Time (h)	6.35	5.37	3.58	6.27	5.05	11.7	6.39
GPU (GB)	3.48	3.84	3.67	4.23	3.63	5.63	4.08
num. of Gaussians	48.3k	164.7k	98.6k	195.5k	9.0k	230.5k	117.8k
num. of superpoints	188	184	112	51	134	42	118.5

(a) canonical space in *Dynamic* stage (b) at time 0 in *Dynamic* stage (c) canonical space in *Kinematic* stage (d) at time 0 in *Kinematic* stage (e) LBS at time 0 in *Kinematic* stage (f) ground truth at time 0

Figure 1: Compare rendered images between canonical space and the warp space of timestamp 0.



(a) AP-NeRF (b) ours (c) ground truth (d) AP-NeRF (e) ours (f) ground truth

Figure 2: Compare to AP-NeRF, ours method are more robust for complex motion.