Supplementary material

This supplementary material is for the paper, *Spacetime Surface Regularization for Neural Dynamic Scene Reconstruction*, namely *4DRegSDF*. We will further provide details of architecture design, visual examples of spacetime surface sampling, and qualitative reconstruction results in Dycheck dataset. Note that the reference numbers including tables and figures are equivalent to those in the manuscript.

A. Reconstruction results

We provide the reconstruction results from Dycheck dataset [22] as in Fig. 7. It demonstrate that while the previous study (TiNeuVox [19]) properly renders the color information. However, it turns out that the surface geometry has severe artifacts. As this dataset states, this dataset is traeted as a challenging dataset due to the lack of effective multi-view factor [22]. However, our method outperforms the color rendering and surface geometry estimation as in Fig. 7.

B. Visualization of surface samples

We illustrate our spacetime surface sampling process as in Fig. 10. In Fig. 10-(a), we visualize the 3D surface sample results from [12, 39, 73], which is formulated as below:

$$\mathbf{p} = \mathbf{p}_i - s \cdot \frac{\partial s}{\partial \mathbf{p}_i}, \tag{11}$$

where $\mathbf{p}$ is the location of the point and s is the SDF value at $\mathbf{p}$. In contrast to this 3D surface sampling, in Fig. 10-(b), we apply our spacetime surface sample method that takes gradient steps along the temporal axis as well as spatial axis along the Signed Distance Function 'hills' as below,

$$[\mathbf{p}; t] = [\mathbf{p}_i; t_i] - s \cdot \frac{\partial s}{\partial [\mathbf{p}_i; t_i]}, \tag{12}$$

This equation clearly indicates that our method can update the time information such that we can sample the surface in various spatio-temporal domain as illustrated in Fig. 10-(b).

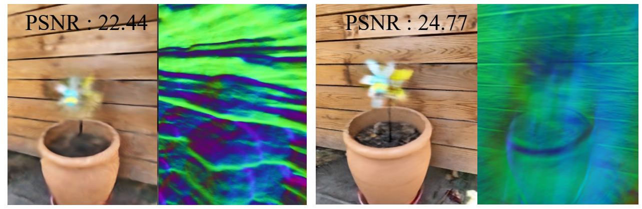
Based on the 4D gradient step, we iterate the 4D surface sampling for 5 times as visualized in Fig. 10-(c). While the sampled points at the first step are remarkably close to the surface, we are able to obtain the refined surface samples after a few iterations.

C. Implementation details

For fair comparison, architecture and hyper-parameter setup is crucial to verify the effectiveness of our spacetime surface regularization for neural scene reconstruction.

Architecture. we utilize the same network architecture designed by the NeuS [70] for Eq. 3. Also, we exploit the deformation network by D-NeRF [54] for the inverse deformation network (Eq. 6). The minor difference is that our network $f(\cdot)$ infers the deformation displacement $\Delta\mathbf{p}$. As stated in Sec. 4.1 of the manuscript, we do not explicitly define the deformation network, instead, the network $f(\cdot)$ jointly infers deformation displacement $\Delta\mathbf{p}$ as well as the SDF value s. Except such minor differences, we do not change any additional hyper-parameters for our architectural setup.

Training. For training, the size of the batch is 1024. For each ray, we sample 64 points with hierarchically samples 64 points following NeuS [70], *a.k.a.* PDF sampling. Also, we follow the training scheme by D-NeRF [54] which takes 800k iterations which takes 48 hours in total. However, to initiate our spacetime surface regularization, we need to have warm start for 10k iterations without applying our regularization loss (Eq. 9). This is to sample surface which is estimated from our warm-started networks. After 10k iteration, our networks, $f(\cdot)$ and $g(\cdot)$, are optimized with the all combinations of the loss as stated in Eq. 10 of the manuscript.



TiNeuVox 4DRegSDF (ours)

Figure 7. Qualitative results in the Dycheck dataset [22].

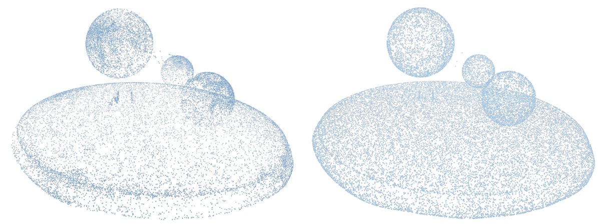

w/o surface rejection w/ surface rejection

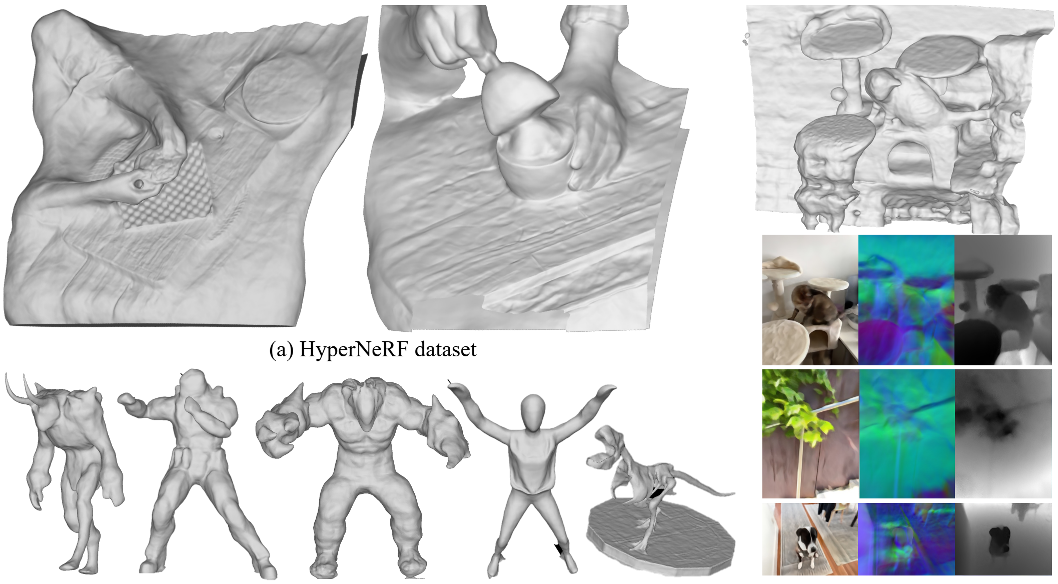
Figure 8. Qualitative results on uniform surface sampling.

	Bouncing balls	Chickens	Haru-sit
PSNR (↑)	41.12 (↑ 0.23)	30.10 (↑ 0.30)	30.89 (↑ 0.23)
SSIM (↑)	0.994 (-)	0.855 (↑ 0.120)	0.870 (↑ 0.070)
MS-SSIM (↑)	0.998 (↑ 0.002)	0.967 (↑ 0.020)	0.939 (↑ 0.012)
LPIPS (↓)	0.021 (↓ 0.010)	0.077 (↓ 0.011)	0.051 (↓ 0.015)
RMSE (↓)	-	-	0.283 (↓ 0.007)

Table 7. Ablation study for uniform surface sampling.

	D-NeRF dataset			HyperNeRF dataset		Dycheck dataset	
	Bouncing balls	Hook	Mean	Chicken	Mean	Sriracha-tree	Mean
D-NeRF [50]	0.10	0.11	0.06	0.15	0.205	0.18	0.251
TiNeuVox [17]	**0.02**	**0.05**	**0.04**	0.18	0.144	**0.06**	0.222
NeuS+D [64]	0.03	0.13	0.05	0.13	0.221	0.15	0.262
Ours	**0.02**	**0.04**	**0.04**	**0.07**	**0.135**	0.08	**0.213**

Table 8. LPIPS evaluation. Mean is average metric of all scenes.

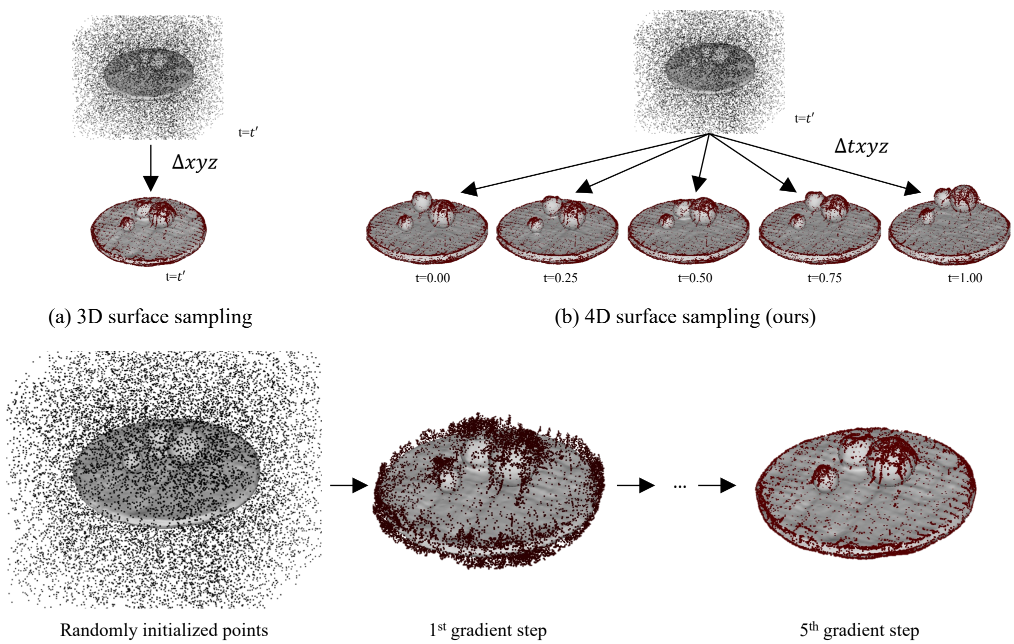


(a) HyperNeRF dataset

(b) D-NeRF dataset

(c) Dycheck dataset

Figure 9. Qualitative results of our method in (a) HyperNeRF dataset, (b) D-NeRF dataset, and (c) dycheck dataset.



(a) 3D surface sampling

(b) 4D surface sampling (ours)

(c) Our surface sample results for each iteration at $t = 0.5$

Figure 10. Visualization of the surface sample process. Our spacetime surface sample (Eq. 5) takes 5 iteration steps for refinement.

D. Uniform surface sampling

To make samples uniform, we follow the rejection sampling presented in [73] as visualized on Fig. 9-(c). We also re-train our method and present the results in Table 7. Overall, we observed that performance improved in all metrics. We will include these results as an ablation study in the final manuscript.

E. LPIPS metric

For SSIM and MS-SSIM, we follow the official metrics provided by the original papers, SSIM by D-NeRF [50] and MS-SSIM by HyperNeRF [48]. Meanwhile, for LPIPS, we provide our results in Table 8.