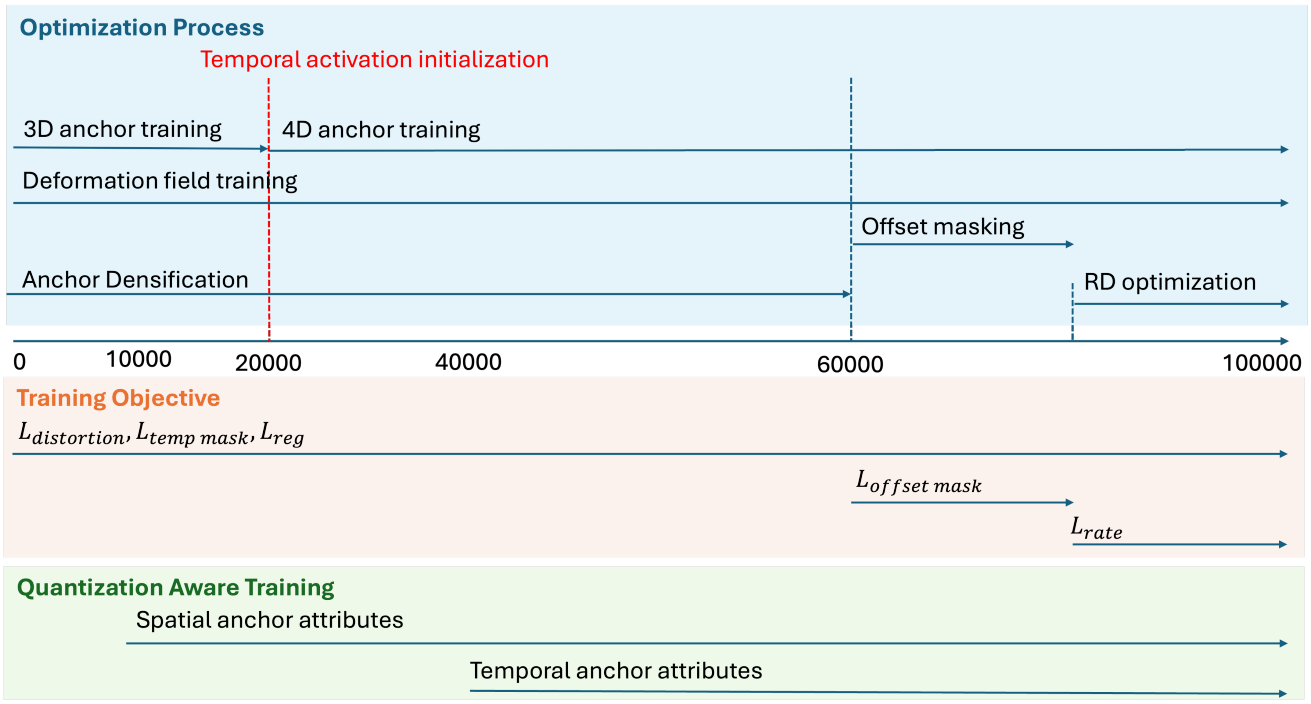


Figure A1. Detailed training process of our TED-4DGS.

1. Training Process

Fig. A1 depicts our detailed training procedure.

Warm-up Training directly from scratch with 4D anchors (i.e., anchors that already include temporal-existence parameters) makes optimization unstable. To address this issue, we adopt a staged procedure: the network is first trained with purely spatial 3D anchors, which allows a dependable spatial configuration to emerge while the deformation module begins learning motion cues from the very first iteration. We jointly optimize the canonical representation and deformation field over 60,000 iterations, with anchor densification enabled. After the 3D stage converges, temporal activation is introduced and the anchors are upgraded to their full 4D form, yielding a stable training process and improved rate–distortion performance.

Temporal activation initialization At 20,000 iterations, we initialize the temporal activation parameter τ, extending from 3D into 4D formulation.

Offset masking stage From 60,000 to 80,000 iterations, we introduce offset masking. The masking loss $L_{\text{offset-mask}}$ is enabled in this stage only.

Table A1. Hyperparameters for Neu3D and HyperNeRF datasets.

Parameter	Neu3D	HyperNeRF
Canonical feature $\boldsymbol{f}$ dimension	50	
Temporal feature ϕ dimension	8	16
Globel deformation bank $\boldsymbol{Z}$ dimension	256	128
$\lambda_{\text{temp mask}}$	2×10^{-5}	1×10^{-6}
$\lambda_{\text{offset mask}}$	1	1
λ_{tv}	0	6×10^{-4}
Channel-wise autoregressive slices	$\{5, 10, 15, 20\}$	

Rate-distortion Optimization Once masking converges, we freeze the mask and perform RD optimization. A rate penalty L_{rate} is added to the objective.

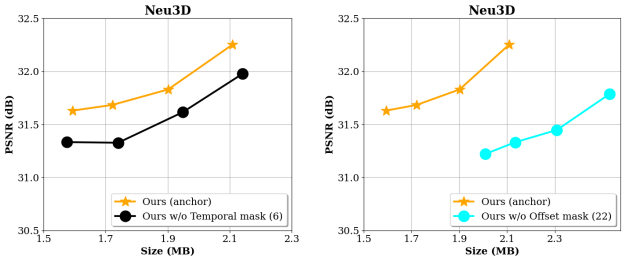
Throughout training, the deformation network remains active, ensuring coherent motion modeling.

2. Hyperparameter Settings

Table A1 summarizes the detailed hyperparameter settings used for experiments on the Neu3D and HyperNeRF datasets.

3. Ablation Study on Masking Mechanisms

Temporal Mask. Table A2 summarizes the masking results, showing that our temporal mask classifies 87 % of



(a) Temporal Mask Ablation (b) Offset Mask Ablation

Figure A2. Rate-distortion comparisons of our masking mechanisms on Neu3D dataset.

Table A2. Ablation study of temporal mask on Neu3D dataset.

Method	PSNR ↑	Size (MB)↓	Dynamic rate↓
Ours	32.25	2.11	13%
Ours w/o Temporal Mask	31.97	2.14	100%

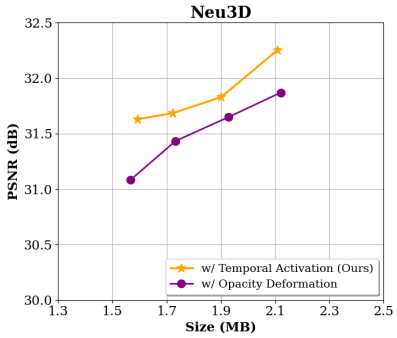


Figure A3. Rate-distortion comparison between our one-off temporal activation design and the time-varying opacity optimization.

anchors as static and thus saves storage. Figure A2a further confirms the mask's impact: disabling it from the outset raises BD-rate by 6 %, demonstrating that the temporal mask eliminates redundant temporal attributes while preserving distortion performance.

Offset Mask. The offset mask selectively disables unnecessary spatial offsets that contribute little to reconstruction quality. If all offsets associated with an anchor are masked out, the entire anchor can be pruned. This mechanism effectively reduces redundancy in both the offset representation and the anchor set. As shown in Fig. A2b, removing the offset mask leads to a 22% increase in BD-rate, demonstrating the importance of this component for achieving compact yet high-quality reconstruction.

4. Time-Varying Opacity Optimization

We compare our temporal activation design (Fig. A4 (a)) against direct opacity optimization (Fig. A4 (b)), where the deformation vectors $\boldsymbol{z}^{(t)}$, interpolated from the deformation bank $\mathbf{Z}$, are used to represent time-varying opacity. This variant offers greater flexibility, allowing the opacity of each Gaussian associated with an anchor to vary over time without the one-off activation constraint. However, it shows inferior rate–distortion results compared to our original one-off design. We attribute this to the limited expressiveness of the deformation vectors $\boldsymbol{z}^{(t)}$, which must now encode additional opacity variations for every Gaussian instead of focusing solely on geometric and motion-related deformations.

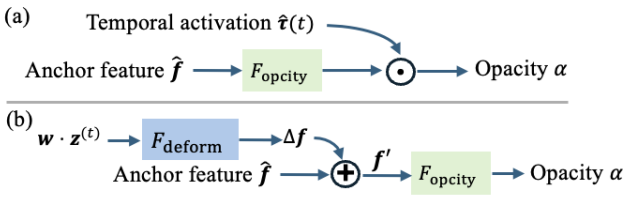


Figure A4. Illustration of (a) our temporal activation mechanism, and (b) Time-varying opacity optimization with the deformation field, where $\boldsymbol{z}^{(t)}$ denotes the deformation vector interpolated from the deformation bank $\mathbf{Z}$, and $\boldsymbol{w}$ represents the projected temporal feature $\boldsymbol{\phi}$. See Fig. 2 in the main paper for details.

5. Temporal Activation Visualizations

To study the impact of occlusion and disocclusion on temporal activation, we analyze Gaussian durations ($\Delta\tau = a_f - a_s$), defined as the time span between activation start a_s and end a_f. As shown in Fig. A5, when occlusion and disocclusion are minimal (frames 0–90), most Gaussian primitives exhibit long temporal durations and remain active throughout the sequence. In contrast, when certain primitives emerge later in the sequence (frames 310–400, e.g., occluded regions such as the interior of the banana peel), they are activated only during their visible frames by the temporal activation function, without the need to model pre-appearance states or relocating them outside the viewing frustum. Moreover, we observe that the shortest-duration Gaussian primitives are concentrated in specific regions. In addition to genuinely short-lived structures, they often serve as residual signals that compensate for inaccurately modeled fast-motion boundaries, as seen around the hands at frames 310 and 340 and the banana peel at frame 340.

6. Bitstream Composition

As shown in Fig. A6, we report how the coded bitrate is apportioned among the principal bitstream components of TED-4DGS. Bitstreams are averaged on a per-sequence basis for both the Neu3D and HyperNeRF datasets and are presented at four operating points.

The bitstream is partitioned into five main categories:

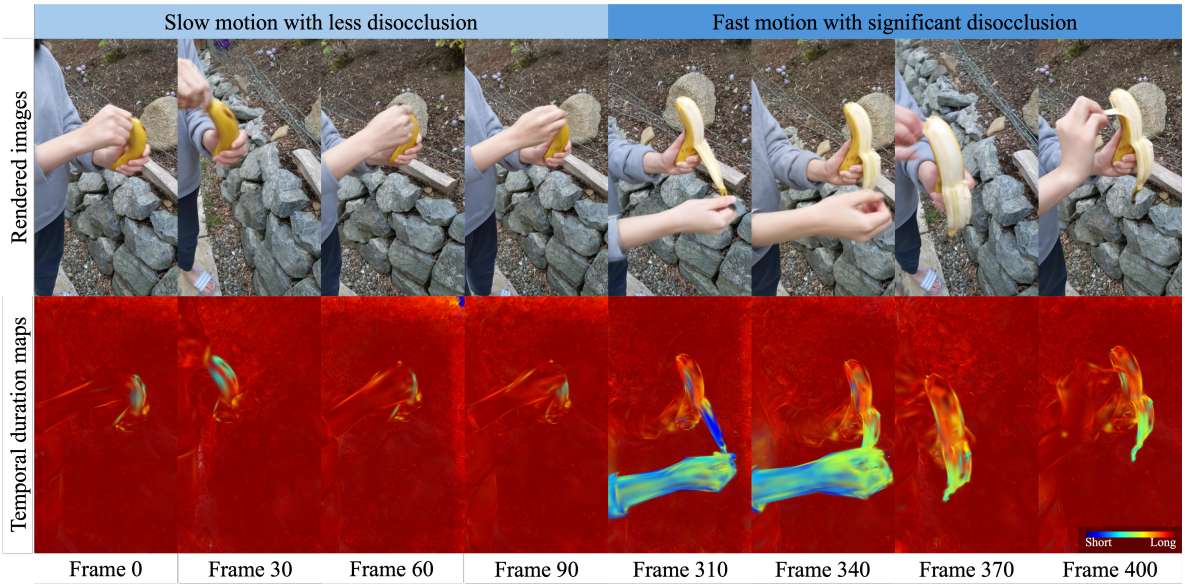
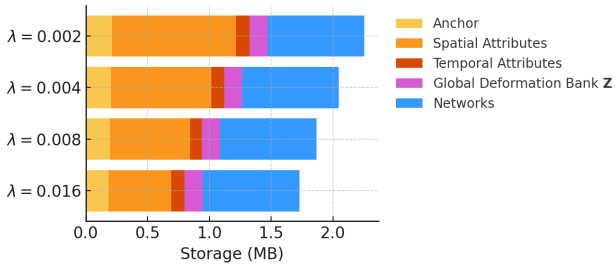
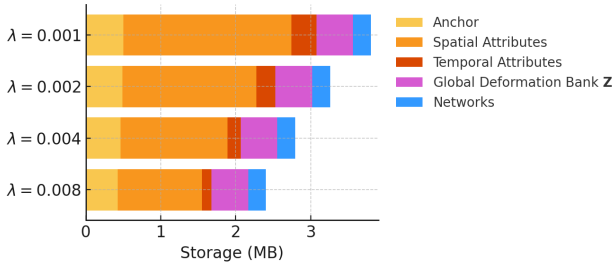


Figure A5. Comparison of rendered images and temporal duration maps at different frames within the same scene.



(a) Neu3D dataset



(b) HyperNeRF dataset

Figure A6. Bitstream composition.

(1) anchor positions $\boldsymbol{x}$, (2) spatial attributes including anchor canonical feature $\boldsymbol{f}$, scaling $\boldsymbol{l}$, offsets $\{\boldsymbol{O}_i\}$ and offset mask $\boldsymbol{M}_o$), (3) temporal attributes, such as anchor temporal feature $\boldsymbol{\phi}$, temporal activation $\boldsymbol{\tau}$, temporal mask $\boldsymbol{M}_t$), (4) global deformation bank $\boldsymbol{Z}$ and (5) the network weight of the deformation decoder F_{deform}, Scaffold MLP decoder, hyperprior decoder, and channel-wise autoregressive model.

7. Complexity Analysis

Training TED-4DGS takes approximately 130 minutes on Neu3D dataset. We then evaluate the encoding and decoding time, where the encoding time does not include training but only accounts for probability generation and entropy coding. As reported in Table A3, encoding and decoding take approximately 5.68 seconds and 3.89 seconds at the low-bitrate setting, respectively. A breakdown analysis further shows that over 90% of the runtime is consumed by CPU-based range coding, which is part of the standard arithmetic coding process and not our design, whereas coding probability generation with our INR-based hyperprior and CARM contributes less than 0.3 seconds, highlighting the efficiency of our design.

Table A3. Coding complexity analysis on Neu3D dataset. CPG: coding probability generation, RC: range coding, I/O: file I/O

Mode	Rate Point	Total(s)	CPG(s)	RC(s)	I/O(s)
Encoding	High	6.82	0.27	6.10	0.45
	Low	5.68	0.23	5.10	0.35
Decoding	High	5.15	0.18	4.85	0.12
	Low	3.89	0.16	3.63	0.10

8. Per-scene Results on Evaluation Datasets

The detailed results of our TED-3DGS on the Neu3D and HyperNeRF datasets are presented in Tables A4, A5.

Table A4. Per-scene evaluation with PSNR and Size (MB) on Neu3D.

	Cook Spinach		Cut Roasted Beef		Flame Salmon		Flame Steak		Sear Steak	
Method	PSNR	Size	PSNR	Size	PSNR	Size	PSNR	Size	PSNR	Size
4DGaussians	32.41	38.28	32.37	38.80	29.26	38.46	31.86	36.16	32.70	36.16
E-D3DGS	32.50	39.04	29.70	39.30	29.46	68.30	31.71	36.40	32.74	28.59
	32.44	3.89	31.84	3.90	29.14	3.96	31.60	3.61	32.40	3.51
Light4GS	32.58	4.76	31.88	4.93	29.17	4.45	31.74	3.69	32.54	4.08
	32.65	6.18	32.05	6.39	29.33	5.54	31.83	3.85	32.61	5.33
	32.02	4.07	31.61	3.85	28.65	4.41	32.48	3.80	32.29	4.08
ADC-GS	32.14	5.29	31.69	5.21	28.92	5.83	32.58	5.51	32.40	4.78
	32.34	6.77	31.88	6.48	29.01	6.08	32.65	7.59	32.48	5.95
	31.96	1.49	32.68	1.47	28.38	2.09	32.57	1.46	32.56	1.46
Ours	32.00	1.63	33.06	1.61	28.10	2.24	32.51	1.54	32.75	1.59
	32.27	1.79	32.76	1.76	28.66	2.53	32.49	1.68	32.98	1.74
	32.48	1.97	33.37	1.95	28.94	2.86	33.02	1.84	33.45	1.93

Table A5. Per-scene evaluation with PSNR and Size (MB) on HyperNeRF.

	3D Printer		Chicken		Broom		Banana	
Method	PSNR	Size	PSNR	Size	PSNR	Size	PSNR	Size
4DGaussians	22.12	63.50	29.03	90.57	22.14	46.70	29.06	51.73
E-D3DGS	22.77	28.50	29.83	51.00	21.70	31.60	28.98	73.20
	21.96	6.13	28.83	4.10	22.14	2.98	28.50	7.40
Light4GS	22.02	7.84	28.93	5.87	22.14	4.63	28.79	9.05
	22.08	9.50	28.97	6.98	22.16	6.10	28.99	12.98
	22.92	3.23	29.61	3.85	21.80	3.27	27.35	5.73
ADC-GS	23.01	4.05	29.69	5.12	21.89	4.73	27.53	6.90
	23.10	4.57	29.75	8.33	21.98	5.20	27.89	8.56
	22.96	2.32	28.89	1.44	21.92	2.40	27.11	3.27
Ours	23.06	2.79	29.07	1.61	21.93	2.83	27.59	3.74
	23.08	3.26	29.18	1.83	21.95	3.26	27.95	4.36
	23.13	3.90	29.26	2.09	21.94	3.77	28.34	5.10

9. Qualitative results

Figs. A7, A8 present rendering results for the Neu3D and HyperNeRF datasets with the highest rate point across different frames.

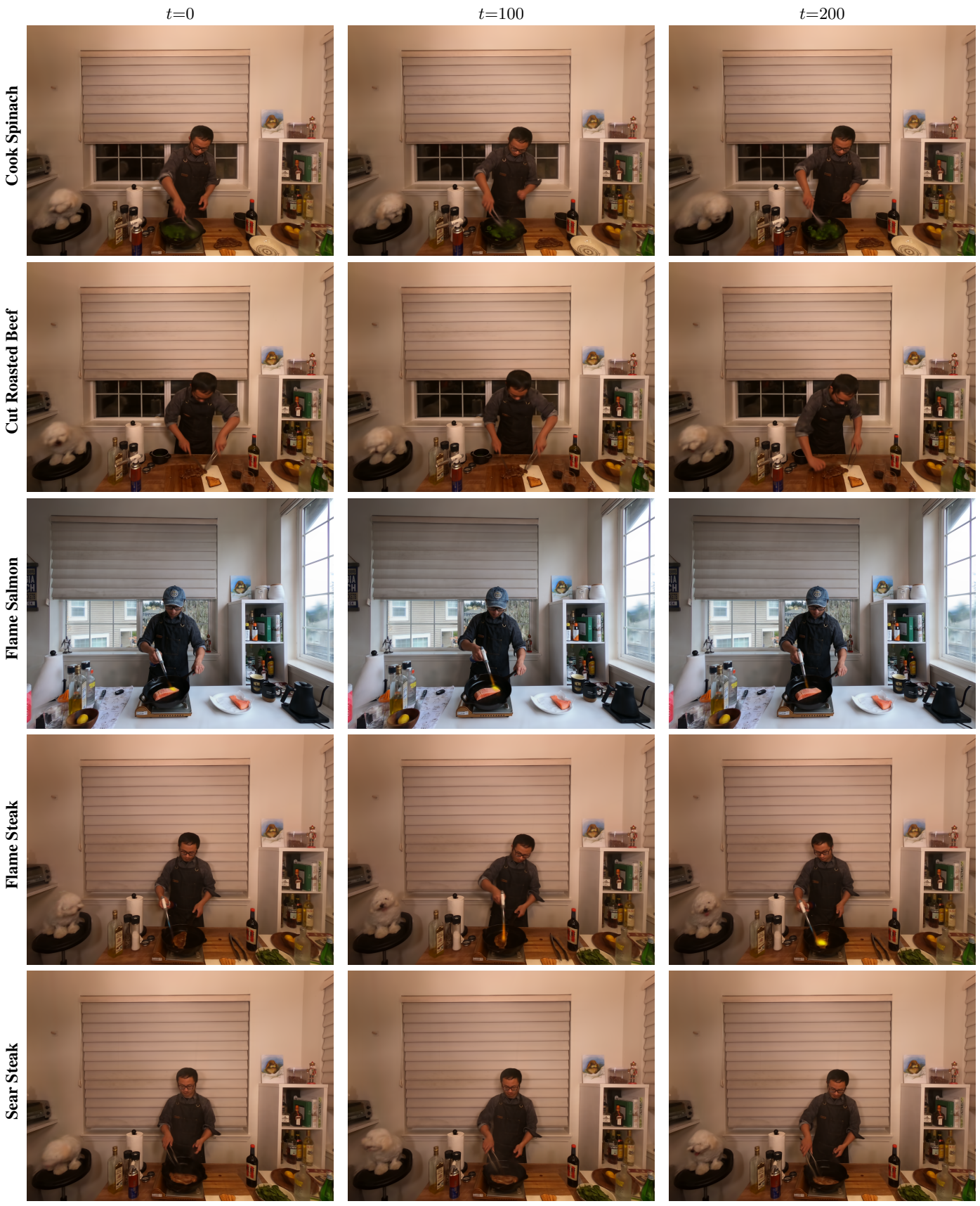


Figure A7. Qualitative results on the Neu3D dataset.

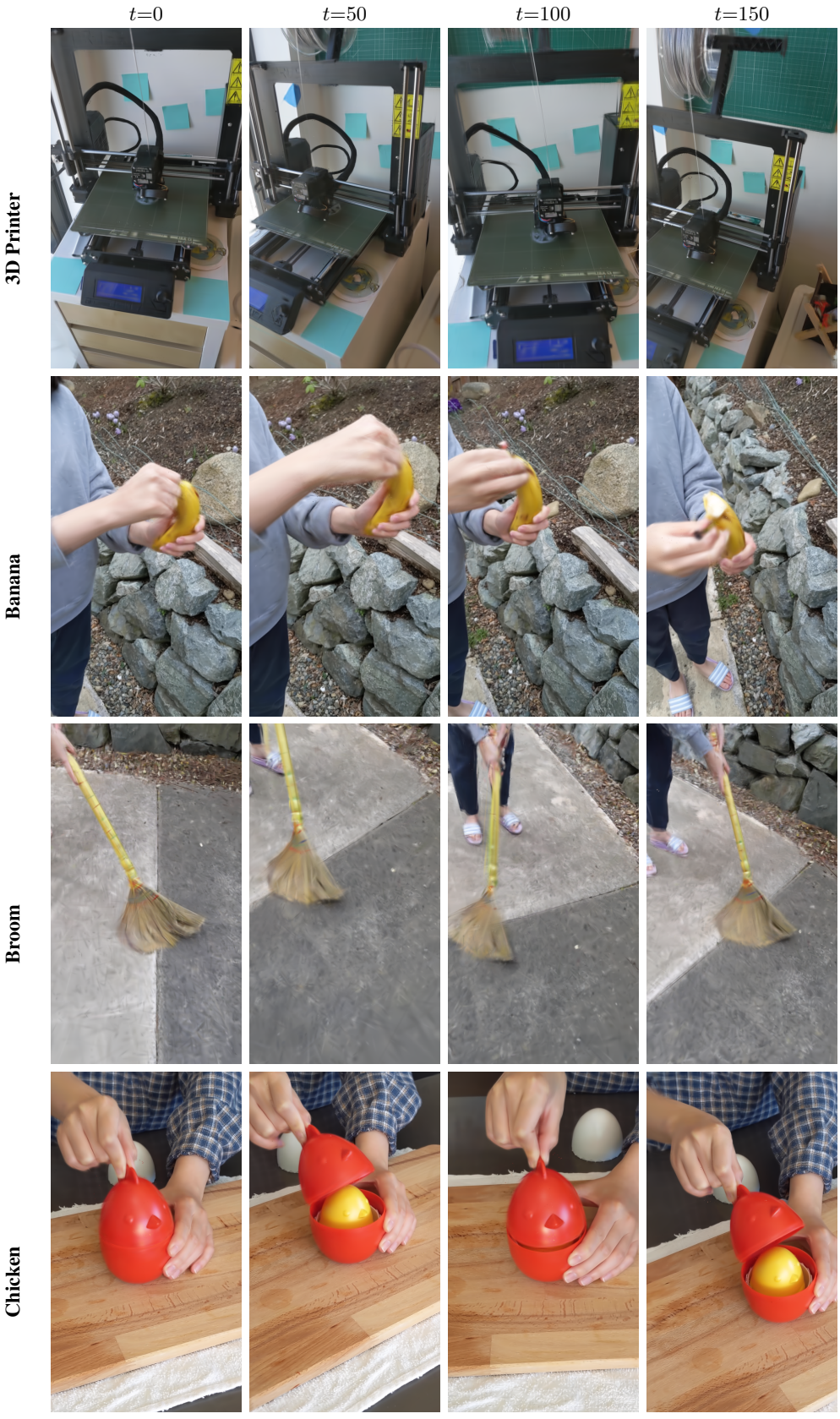


Figure A8. Qualitative results on the HyperNeRF dataset.