

Supplementary Material

Shima Salehi

Joshua Peebles

1. Protocol and Scope

All experiments use chronological observed/future splits. The dynamic Gaussian representation, control keypoints, keypoint-to-Gaussian bindings, appearance, opacity, and renderer are fitted only on the observed prefix and frozen before forecasting. Router calibration further divides the observed prefix 75/25: temporary experts fit the first part and predict the disjoint observed tail; the calibrated router is then frozen and paired with deployment experts trained on the complete observed prefix. No future frame or metric is used for reconstruction, grouping, expert fitting, routing, checkpoint selection, or early stopping. The calibration tail is excluded only from temporary motion-expert fitting; it remains part of the frozen reconstruction. Every reported checkpoint is cold reloaded before evaluation.

The primary D-NeRF [2] protocol uses an 80/20 split. Forecasting stability is evaluated with expert/router seeds 17, 42, and 123; routing, horizon, and split analyses otherwise use seed 17. The standalone temporal-basis ablation uses seed 42 and no router. Split-sensitivity experiments use independently reconstructed 70/30 and 90/10 representations. All routed runs use $G = 16$ soft motion groups, four horizon bins, and a 256-parameter route table. Peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and learned perceptual image patch similarity (LPIPS) are averaged within a scene and then equally across scenes. Exact implementation settings are given below.

2. Additional Implementation Details

This section reports only settings not specified in the main paper. Soft-group assignment uses temperature 0.18 and 30 deterministic k -means updates. The coefficient Transformer uses AdamW with weight decay 10^{-4} ; the structured Transformer uses Adam. Both use cosine learning-rate decay. The router uses zero weight decay and gradient-norm clipping at 1.0. The coefficient expert weights pose, velocity, acceleration, and gate terms by 1, 0.5, 0.1, and 10^{-5} . The structured expert uses unit pose and acceleration weights;

its radius-normalized rigid and non-rigid physical terms are weighted 0.10 and 0.02. Per-channel standard deviations are lower-bounded by 10^{-4} .

For the important-pixel mask defined in the main paper, the foreground and dynamic-change thresholds are $\tau_{\text{fg}} = 0.04$ and $\tau_{\text{dyn}} = 0.025$, respectively. The masked mean-squared-error term has weight 1.0. The adjacent-bin smoothness, moving-mass, and moving-expert entropy weights are $\lambda_s = 2 \times 10^{-3}$, $\lambda_m = 10^{-4}$, and $\lambda_e = 10^{-4}$, respectively.

3. Router Visualization

Figure 1 visualizes the saved Jumping Jacks router using observed-prefix information only. For visualization, the left panel colors each keypoint by its maximum-probability group and scales its marker by assignment confidence in a two-dimensional principal-component projection. This hard color label is used only for visualization; inference retains the complete soft assignment vector. The right panel shows the learned expert weights for every group and horizon bin.

4. Forecasting Stability

Table 1 evaluates three expert/router initialization seeds while holding the frozen Gaussian reconstruction fixed. Thus, the variation measures forecasting-head optimization stability rather than reconstruction variability. The scene-unweighted mean is 27.110 ± 0.279 dB. Relative to persistence, the three-seed model gains 2.409 dB; a scene-resampled bootstrap gives a 95% confidence interval of $[0.883, 4.173]$ dB, and an exact one-sided paired sign-flip test gives $p = 0.0039$.

5. Horizon and Quartile Analysis

Each sealed 80/20 suffix is divided chronologically into four approximately equal quartiles. Table 2 compares MoRE-GS with persistence. PSNR improves by 3.12, 3.17, 3.02, and 1.73 dB from Q1 to Q4, respectively, with better SSIM and LPIPS throughout. The smaller Q4 margin indicates increasing long-horizon uncertainty.

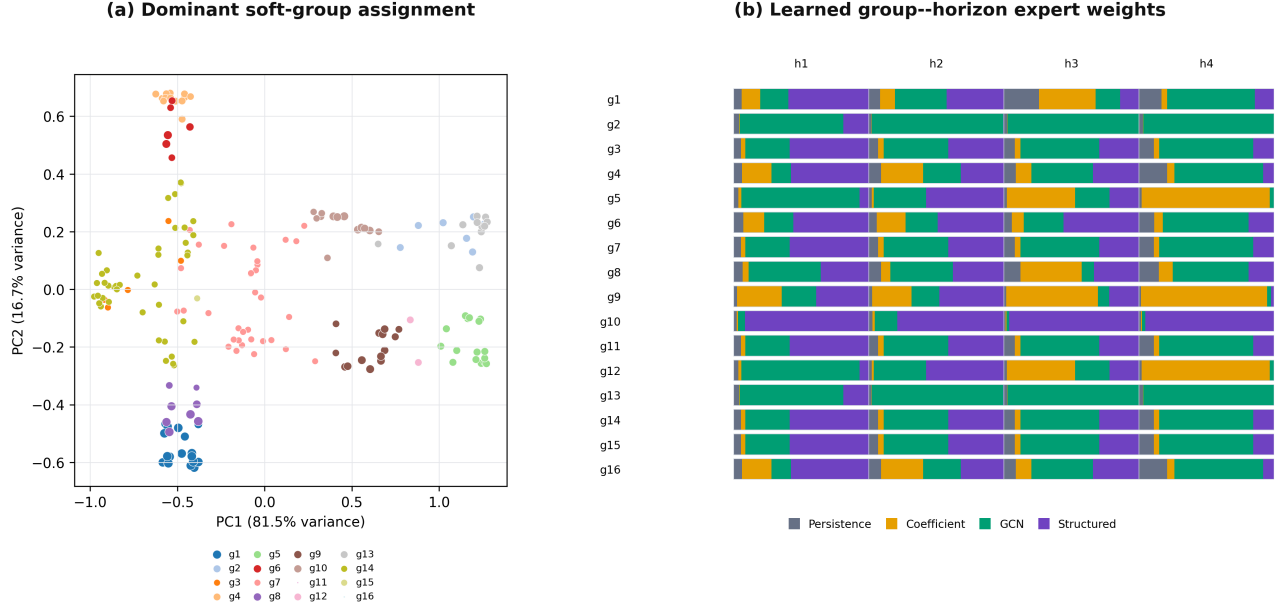


Figure 1. Router visualization for $G = 16$: (a) dominant soft-group assignment, with color indicating group and marker size indicating confidence; inference retains full soft assignments. (b) Group-horizon expert weights.

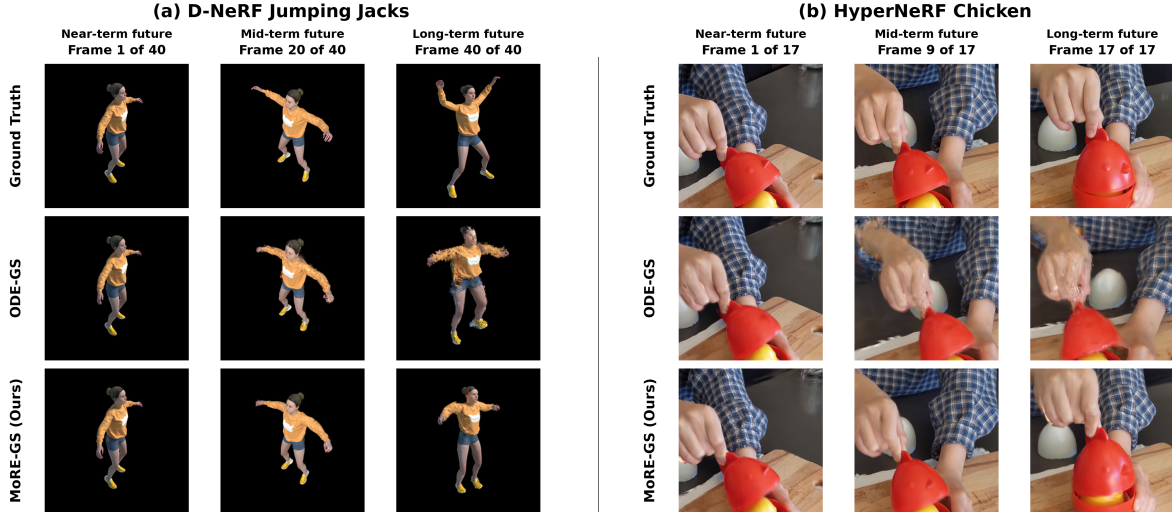


Figure 2. Near-, middle-, and long-horizon forecasts on D-NeRF Jumping Jacks and HyperNeRF Chicken.

Table 3 compares matched $G = 16$ runs. Group-horizon routing is 0.46 dB lower in Q1, changes PSNR by +0.05, +0.24, and +0.00 dB in Q2-Q4, and improves SSIM in every quartile and LPIPS in Q2-Q4. Its full-suffix PSNR is 0.06 dB lower, while SSIM and LPIPS improve by 0.0012 and 0.0008. This one-seed result supports horizon conditioning as a secondary structural/perceptual refinement, not a dominant gain.

6. Temporal-Basis Ablation

Table 4 compares the random orthonormal basis with DCT-II [1] while retaining all coefficients and matching all other settings. The random basis improves mean PSNR by 0.185 dB. We retain it because it preserves trajectory information without imposing DCT-II’s fixed frequency ordering and achieves better aggregate performance across all three metrics. This ablation compares basis ordering; it does not establish

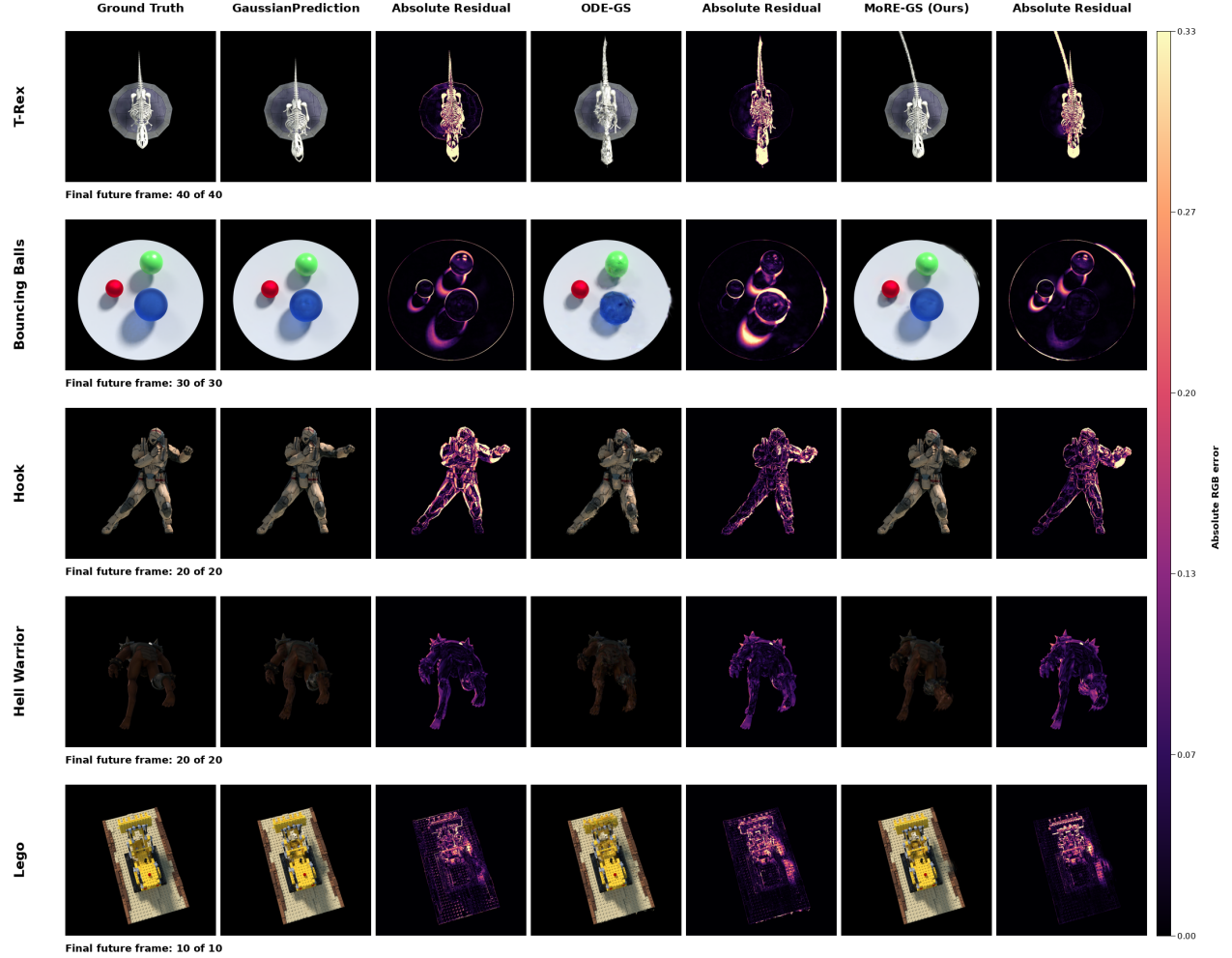


Figure 3. Final-horizon forecasts and absolute RGB residuals on the five D-NeRF scenes not visualized in the main paper.

Table 1. $G = 16$ stability over seeds 17, 42, and 123; values are mean $\pm$ standard deviation.

Scene	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
T-Rex	20.61 ± 0.64	0.934 ± 0.012	0.059 ± 0.012
Mutant	33.19 ± 0.27	0.975 ± 0.001	0.015 ± 0.001
Jumping Jacks	21.20 ± 0.10	0.922 ± 0.003	0.072 ± 0.002
Stand Up	30.36 ± 0.23	0.956 ± 0.001	0.035 ± 0.001
Bouncing Balls	27.60 ± 0.57	0.972 ± 0.003	0.037 ± 0.005
Hook	26.50 ± 0.07	0.938 ± 0.001	0.041 ± 0.001
Hell Warrior	30.54 ± 0.27	0.922 ± 0.003	0.078 ± 0.003
Lego	26.89 ± 0.95	0.945 ± 0.003	0.043 ± 0.003
Mean	27.11 ± 0.28	0.946 ± 0.003	0.047 ± 0.003

necessity relative to direct, unprojected input.

7. Observed/Future Split Robustness

We compare MoRE-GS with GaussianPrediction [3] under matched 70/30 and 90/10 conditions. Each split

Table 2. Quartile forecasts for persistence and MoRE-GS. Best values per row and metric are bold.

Quartile	Persistence			MoRE-GS		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
Q1	28.31	0.955	0.032	31.43	0.970	0.024
Q2	24.51	0.937	0.050	27.68	0.951	0.040
Q3	23.23	0.929	0.061	26.25	0.941	0.050
Q4	22.39	0.921	0.070	24.12	0.928	0.065

uses its own observed-only Gaussian reconstruction, identical future indices for both forecasters, forecasting seed 17, and all eight scenes. Tables 5 and 6 report all scene-level metrics.

MoRE-GS improves every aggregate metric, with PSNR gains of 2.066 dB at 70/30 and 1.635 dB at 90/10. GaussianPrediction achieves higher PSNR on Jumping Jacks under both splits and on Hell Warrior

Table 3. Matched $G = 16$ quartile forecasts for group-only and group-horizon routing. Best values use unrounded scores.

Quartile	Group-only			Group-horizon		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
Q1	31.89	0.969	0.024	31.43	0.970	0.024
Q2	27.63	0.950	0.041	27.68	0.951	0.040
Q3	26.00	0.939	0.052	26.25	0.941	0.050
Q4	24.12	0.927	0.066	24.12	0.928	0.065
Mean	27.53	0.947	0.045	27.47	0.948	0.045

Table 4. Matched temporal-basis ablation. Best values per scene and metric are bold.

Scene	DCT-II			Random orthonormal		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
T-Rex	19.04	0.924	0.067	19.05	0.930	0.063
Mutant	32.70	0.969	0.016	32.31	0.968	0.017
Jumping Jacks	20.61	0.907	0.086	20.82	0.910	0.084
Stand Up	28.92	0.951	0.040	29.37	0.953	0.038
Bouncing Balls	26.47	0.970	0.043	26.99	0.973	0.039
Hook	26.45	0.940	0.040	26.97	0.941	0.039
Hell Warrior	29.60	0.902	0.095	30.17	0.918	0.082
Lego	25.30	0.938	0.050	24.90	0.936	0.052
Mean	26.14	0.938	0.055	26.32	0.941	0.052

Table 5. D-NeRF forecasting under the matched 70/30 observed/future split. Best values per scene and metric are bold.

Scene	GaussianPrediction			MoRE-GS		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
T-Rex	19.52	0.933	0.063	19.64	0.929	0.067
Mutant	25.31	0.941	0.037	28.97	0.957	0.024
Jumping Jacks	19.32	0.907	0.095	18.53	0.837	0.153
Stand Up	22.92	0.926	0.068	27.60	0.945	0.044
Bouncing Balls	23.17	0.955	0.052	27.73	0.977	0.032
Hook	22.74	0.902	0.068	24.52	0.919	0.059
Hell Warrior	28.52	0.904	0.098	28.82	0.900	0.101
Lego	23.12	0.915	0.069	25.33	0.932	0.057
Mean	23.08	0.923	0.069	25.14	0.924	0.067

at 90/10. MoRE-GS achieves higher PSNR on seven of eight scenes at 70/30 and five of eight scenes at 90/10, with an additional tie on T-Rex at 90/10. Changing the split affects both the available reconstruction evidence and the forecast length.

8. Additional Qualitative Results

Figure 3 complements the main-paper qualitative comparison with the five D-NeRF scenes not visualized there. Mutant, Stand Up, and Jumping Jacks are omitted because they already appear in the main paper. All residuals use one shared display scale. Future ground truth is used only for this post-evaluation visualization and residual computation.

Table 6. D-NeRF forecasting under the matched 90/10 observed/future split. Best values per scene and metric are bold.

Scene	GaussianPrediction			MoRE-GS		
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
T-Rex	22.62	0.952	0.039	22.62	0.954	0.039
Mutant	27.90	0.956	0.023	30.39	0.966	0.017
Jumping Jacks	23.00	0.938	0.051	21.87	0.928	0.064
Stand Up	30.75	0.964	0.024	32.22	0.971	0.020
Bouncing Balls	29.90	0.980	0.020	34.04	0.988	0.015
Hook	25.86	0.933	0.037	28.66	0.957	0.026
Hell Warrior	33.59	0.952	0.049	33.13	0.946	0.056
Lego	26.39	0.929	0.044	30.15	0.961	0.031
Mean	27.50	0.951	0.036	29.13	0.959	0.033

References

- [1] Nasir Ahmed, T. Natarajan, and K. R. Rao. Discrete cosine transform. *IEEE Transactions on Computers*, C-23(1):90–93, 1974. [2](#)
- [2] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-NeRF: Neural radiance fields for dynamic scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10318–10327, 2021. [1](#)
- [3] Boming Zhao, Yuan Li, Ziyu Sun, Lin Zeng, Yujun Shen, Rui Ma, Yinda Zhang, Hujun Bao, and Zhaopeng Cui. GaussianPrediction: Dynamic 3d gaussian prediction for motion extrapolation and free view synthesis. In *ACM SIGGRAPH Conference Papers*, pages 84:1–84:12, 2024. [3](#)