GP-4DGS: Probabilistic Gaussian Splatting from Monocular Video via Variational Gaussian Processes
–Supplementary Documents–

Mijeong Kim[1] Jungtaek Kim[3] Bohyung Han[1,2]
[1]ECE and [2]IPAI, Seoul National University, Korea [3]University of Wisconsin–Madison, USA
{mijeong.kim, bhhan}@snu.ac.kr, jungtaek.kim@wisc.edu

A. Effect of Time-series Feature Extractor

Motivation In Variational Gaussian Processes (VGPs), the initialization of inducing points $\mathcal{Z}$ is a key factor in summarizing the 4D deformation field. To represent complex motion from a large candidate set of trajectories, it is essential to select points that capture the most informative dynamics.

Standard approaches often use velocity differences as a heuristic for k-nearest neighbor (KNN) selection. While efficient, velocity-based features only reflect instantaneous change, failing to account for long-term temporal patterns or higher-order dependencies within a trajectory. To bridge this gap, we employ Chronos [1], a pre-trained time-series foundation model, to generate embeddings that encode the global temporal deformation of each Gaussian primitive.

Time-series feature extraction For a trajectory $\boldsymbol{\xi}_k \in \mathbb{R}^{T\times 3}$ of the k-th primitive, we extract features for each spatial dimension $i \in \{x, y, z\}$ independently. This preserves axis-specific semantics while capturing temporal dynamics within each dimension. Chronos produces a 256-dimensional embedding $\mathbf{f}_{k,i}$ per axis, which we concatenate into a single trajectory descriptor as follows:

$$\mathbf{F_k} = [\mathbf{f}_{k,x} \parallel \mathbf{f}_{k,y} \parallel \mathbf{f}_{k,z}] \in \mathbb{R}^{768}, \tag{A}$$

By using learned representations, $\mathbf{F}_k$ captures complex motion cues such as periodic patterns and non-linear shifts that simple velocity vectors overlook.

Inducing point selection We construct the final set of inducing points $\mathcal{Z}$ by combining these spatial features with temporal sampling. First, we apply KNN in the feature space $\mathbf{F}$ to identify M_{spatial} representative trajectory landmarks. These are then combined with M_{time} points sampled from the normalized time interval $[0, 1]$ via a Cartesian product:

$$\mathcal{M} = \mathcal{M}_{\text{spatial}} \times \mathcal{M}_{\text{time}}, \quad |\mathcal{M}| = M_{\text{spatial}} \times M_{\text{time}}. \tag{B}$$

This strategy ensures that the selected inducing points are placed on semantically significant trajectories, leading to more accurate reconstruction and fast convergence in VGPs.

B. Effect of Periodic Component in GPs

The choice of mean function and kernel design is critical for enabling long-term extrapolation and short-term motion prediction for future frames. We systematically compare constant versus periodic components to understand their complementary roles in this selection.

B.1. Periodic Mean

Priors on the output means also can effect on extrapolation, where they have minimal effect on the observed data region, but their strength is manifested in unobserved regions as inductive bias. The prior over inducing point values is defined as follows [1]:

$$f_i(\mathbf{x}) \sim \mathcal{GP}(m_i(\mathbf{x}), k_i(\mathbf{x}, \mathbf{x}')). \tag{C}$$

where m_i is the mean function and the most common choice in standard a constant mean function as follows:

$$m_i(\mathbf{x}) = c_i, \tag{D}$$

where c is a learnable mean offset. To enable temporal extrapolation on periodic motions, a **periodic mean** serves as an effective solution:

$$m_i(\mathbf{x}) = c_i + A_i \sin\left(\frac{2\pi t}{T_i} + \phi_i\right), \tag{E}$$

where $\mathbf{x} = (\boldsymbol{p}, t)$, c is the mean offset, A_i is the amplitude, T_i is the period, and ϕ_i is the phase shift. All parameters are learned jointly during training along with parameters in kernel.

[1]This generalizes the standard form $m_i(\mathbf{x}) = 0$ commonly used in Variational Gaussian Processes by introducing a learned mean function

Table A. Comparison prior and kernel design for extrapolation.

Prior	Kernel	Long-term	Short-term	Recommendation
Constant	Constant			No extrapolation
Constant	Periodic		✓	Non-periodic motion
Periodic	Constant	Limited		Artifacts on non-periodic
Periodic	Periodic	✓	✓	Periodic motion

B.2. Periodic Kernel

The temporal component of the kernel is critical for capturing motion patterns over time. We also compared two designs for the temporal kernel structure: constant kernel and periodic kernel. The **constant kernel** employs a stationary Matérn kernel over spatial and temporal dimensions jointly as follows:

$$k_i^{\text{temporal}}(\mathbf{x}, \mathbf{x}') = \sum_{j \in \{x,y,z\}} k_{i,j}^{\text{matérn}}((p_j, t), (p'_j, t')), \quad \text{(F)}$$

The **periodic kernel** factorizes spatial and temporal correlations as follows;

$$k_i^{\text{temporal}}(\mathbf{x}, \mathbf{x}') = \sum_{j \in \{x,y,z\}} k_{i,j}^{\text{matérn}}(p_j, p'_j) \cdot k_{i,j}^{\text{periodic}}(t, t'), \quad \text{(G)}$$

where $k^{\text{periodic}}(t, t')$ captures recurring temporal patterns.

B.3. Long-term & Short-term Extrapolation

For time-wise extrapolation, we compare the effects of periodic mean and periodic kernel. As summarized in Table A, the periodic mean is effective for long-term extrapolation on periodic motions, while the periodic kernel provides more general performance for both periodic and non-periodic motions, enabling short-term extrapolation. Constant kernels without periodic structure fail entirely, halting predictions at the trajectory boundary

Figure A demonstrates the effectiveness of periodic components on periodic motions. Without any periodic structure, GPs fail to extrapolate beyond the observed time range. The periodic mean significantly improves long-term extrapolation, capturing the cyclical patterns well into the future. For non-periodic motions, Figure B shows the periodic mean introduces spurious local oscillations and underfits the observed data on non-periodic sequences.

C. Additional Analysis and Comparison

Visualization of GP-GS trajectory To supplement Figure 5 in the main paper, we provide detailed visualizations of Gaussian primitive's trajectories across all dimensions (3D translation and 6D rotation representation) to show the impact of our GP algorithm, in Figure C. This demonstrates the quality of our variational Gaussian process predictions for both translation (X, Y, Z coordinates) and rotation components. Figure D shows trajectories with non-periodic motions, demonstrating that our method also handles these challenging scenarios effectively.

Table B. Ablation study on GP-GS optimization for the *paper-windmill* scene. Our GP-GS optimization enables proper convergence and produces accurate priors that improve reconstruction.

Method	mPSNR ↑	mSSIM ↑	mLPIPS ↓
Baseline	19.56	0.558	0.21
w/o GP-GS optimization	19.22	0.541	0.17
w/ GP-GS optimization	19.88	0.560	0.19

Table C. Ablation study on kernel designs and initialization.

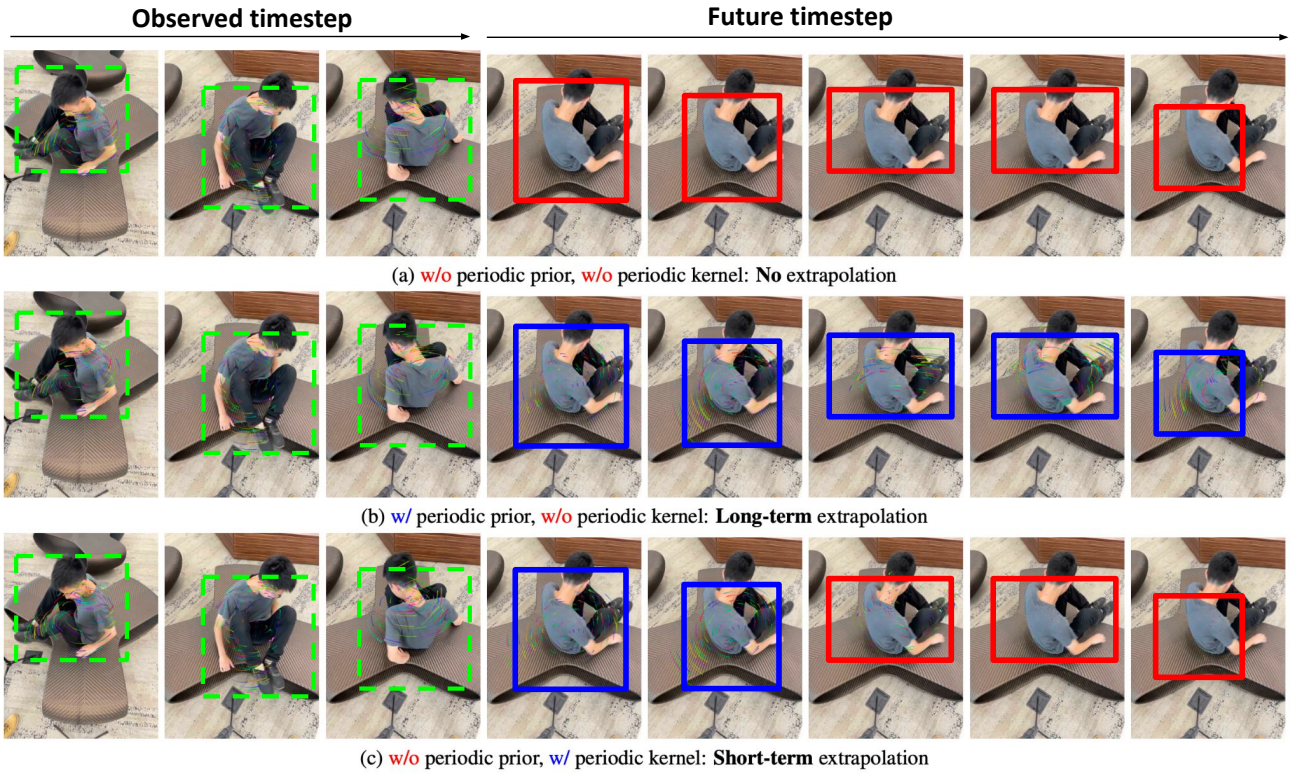
Kernel design	Extrapolation		m-PSNR ↑
	5 frames	15 frames	
w/o decomposition	12.61	10.6	15.0
Temporal kernel → RBF	13.8	11.1	16.4
Temporal kernel → Matérn	13.1	10.4	17.1
Temporal kernel → spectral-mixture	15.7	14.1	**17.5**
Ours	**15.9**	**14.2**	17.3

Effect of GP-GS optimization Table B shows the effect of the GP-GS optimization on the *paper-windmill* scene. Without joint optimization, the GP fails to learn accurate trajectory priors, resulting in worse performance than baseline methods. Our GP-GS optimization enables proper convergence and produces accurate priors that substantially improve reconstruction.

Additional analysis of kernel choice We choose Matérn kernel over RBF to better handle discontinuities (Fig. E), which is important for modeling deformation of multi-objects. As shown in Table C, different temporal kernels achieve similar NVS performance, while periodic kernels perform better for long-horizon extrapolation. Rather than advocating a single optimal kernel, we emphasize that kernel choice should reflect the desired modeling properties. Spatial-temporal decomposition is essential; without it, if any single axis value among (p_x, p_y, p_z, t) falls outside the observed range, prediction collapses to prior.

D. Additional Details

We use an initial learning rate of 10^{-2} for both spatial and temporal inducing points in the GPs. The initial length scales are set to 0.001 and 0.002 for the spatial and temporal kernels, respectively. We employ an exponential decay schedule with a decay rate of 0.95 per epoch. The stochastic variational inference uses a batch size of 5000. For GP-GS optimization, we run 1000 inner iterations for GP optimization. All GP inputs and outputs are normalized for more

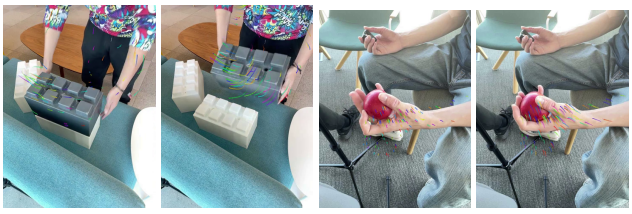


(a) w/o periodic prior, w/o periodic kernel: **No** extrapolation

(b) w/ periodic prior, w/o periodic kernel: **Long-term** extrapolation

(c) w/o periodic prior, w/ periodic kernel: **Short-term** extrapolation

Figure A. **Ablation of periodic component for extrapolation** with **predicted motion trajectories via GPs**. Periodic priors enable long-term extrapolation, while periodic kernels enable short-term extrapolation. The first three columns show the seen frame period, columns four through five show short-term motion estimation, and columns six through eight show long-term motion estimation. **The background for future frames is set to the last frame of the training sequence.**



(a) local-fluctuation (b) short-term extrapolation

Figure B. Examples of local fluctuation in non-periodic motion with periodic prior. Thus, we do not use a periodic prior for non-periodic motion, and a periodic kernel is enough to predict for short-term extrapolation.

stable optimization. The Gaussian process illustration in Figure 1 of the main paper is adapted from this cite[2].

E. Breakdown Results

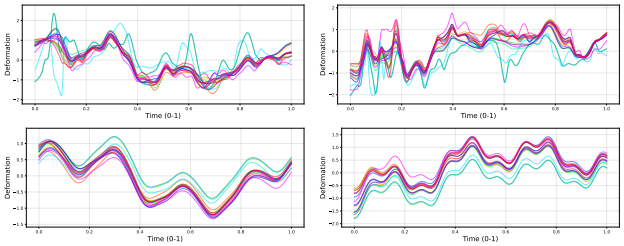
To supplement Table 1 in the main paper, we provide detailed results; Table D and Tabletab:pck show the breakdown results on DyCheck. Note that PCK can be only measured on 5 scenes because SoM [4] only provides the 3D tracking information on these scene. Our method demonstrates substantial improvements on the *wheel* and *space-out* scenes, which feature high occlusion but simple geometry that can be effectively inferred by our GP algorithm. Tthe challenging subset uses subsampled frames with reduced overlap to evaluate robustness under sparse observations. The sampling for the challenging subset follows the diffusion-based method [2], which is an orthogonal algorithm, whereas ours focuses on motion modeling, not visual priors. Our method consistently outperforms baselines, with larger improvements under challenging conditions.

[2] https://www.lancaster.ac.uk/stor-i-student-sites/thomas-newman/2022/05/05/gaussian-processes-in-regression/

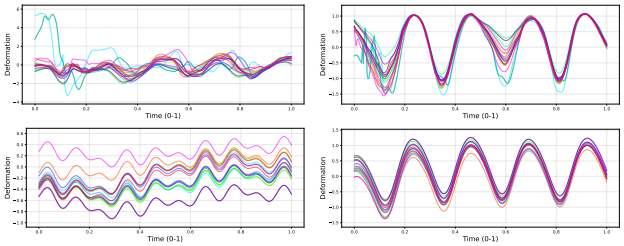
References

[1] Abdul Fatir Ansari, Lorenzo Stella, Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Syndar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Yuyang Wang. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024. 1

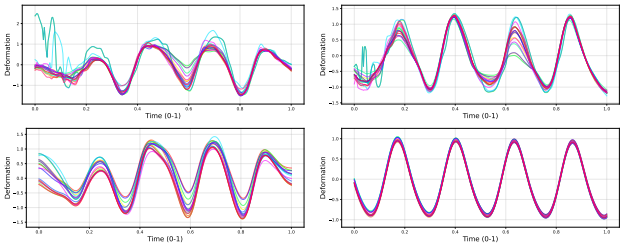
[2] Jiaxin Huang, Sheng Miao, Bangbang Yang, Yuewen Ma, and Yiyi Liao. Vivid4d: Improving 4d reconstruction from monocular video by video inpainting. In *ICCV*, 2025. 3



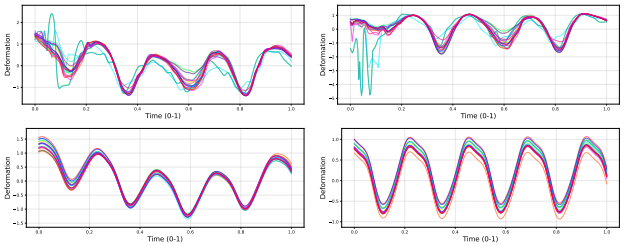
(a) X Translation (b) Y Translation



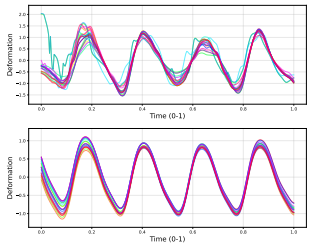
(c) Z Translation (d) 6D Rotation Dim. 1



(e) 6D Rotation Dim. 2 (f) 6D Rotation Dim. 3

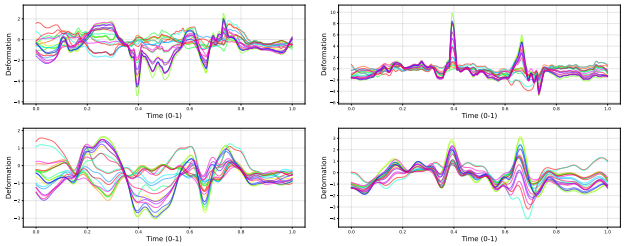


(g) 6D Rotation Dim. 4 (h) 6D Rotation Dim. 5

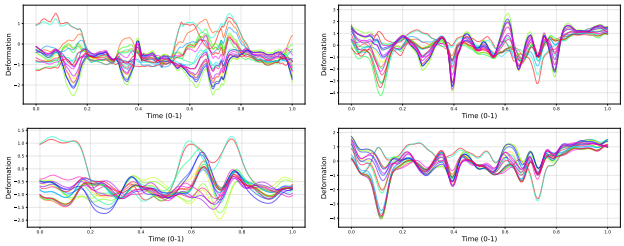


(i) 6D Rotation Dim. 6

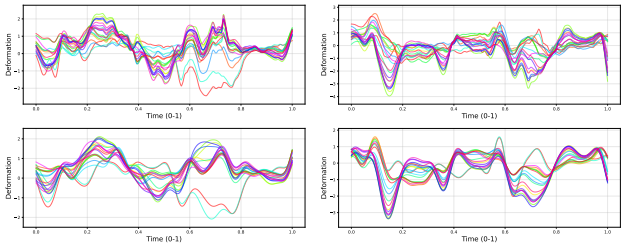
Figure C. Comprehensive trajectory comparison across all transformation dimensions in the *spin* scene: **Original (top) vs. GP (bottom)**. Our variational Gaussian process approach successfully captures and denoises both translational and rotational dynamics, properly giving the priors on reconstructing trajectories.



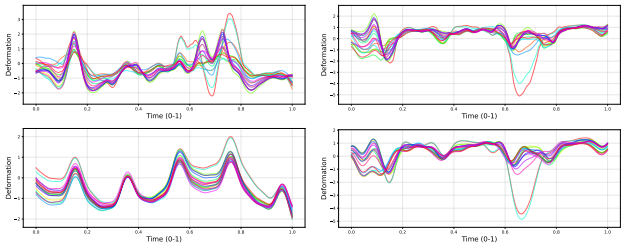
(a) X Translation (b) Y Translation



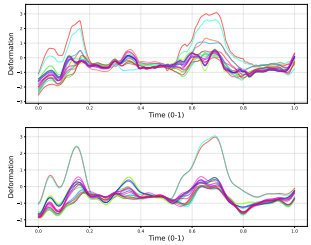
(c) Z Translation (d) 6D Rotation Dim. 1



(e) 6D Rotation Dim. 2 (f) 6D Rotation Dim. 3



(g) 6D Rotation Dim. 4 (h) 6D Rotation Dim. 5



(i) 6D Rotation Dim. 6

Figure D. Comprehensive trajectory comparison across all transformation dimensions in the *Wheel* scene: **Original (top) vs. GP (bottom)**. Our variational Gaussian process approach successfully captures and denoises both translational and rotational dynamics, properly giving the priors on reconstructing trajectories, where even with non-periodic motion.

[3] Colton Stearns, Adam Harley, Mikaela Uy, Florian Dubost, Federico Tombari, Gordon Wetzstein, and Leonidas Guibas. Dynamic gaussian marbles for novel view synthesis of casual monocular videos. In *SIGGRAPH Asia*, 2024. 5

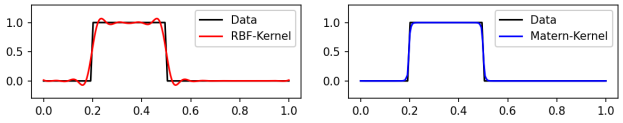


Figure E. Kernel comparison on discontinuous data. RBF kernel (left) fails, but Matérn kernel (right) accurately fits step functions.

Table D. Per-scene breakdown results on DyCheck. Our method shows large improvements on highly occluded scenes (*wheel*, *space-out*) with simple geometry.

Scene	Method	mPSNR ↑	mSSIM ↑	mLPIPS ↓
Apple	Gaussian Marbles [3]	16.84	0.702	0.68
	SoM [4]	16.65	0.752	0.56
	Ours	16.79	0.754	0.55
Block	Gaussian Marbles [3]	16.50	0.650	0.51
	SoM [4]	16.69	0.660	0.45
	Ours	17.01	0.675	0.44
Paper	Gaussian Marbles [3]	15.96	0.296	0.58
	SoM [4]	19.56	0.558	0.21
	Ours	19.88	0.560	0.19
Spin	Gaussian Marbles [3]	17.84	0.515	0.48
	SoM [4]	17.31	0.707	0.31
	Ours	17.56	0.728	0.27
Teddy	Gaussian Marbles [3]	13.01	0.550	0.66
	SoM [4]	13.44	0.546	0.61
	Ours	13.55	0.576	0.59
Space-out	Gaussian Marbles [3]	15.19	0.560	0.54
	SoM [4]	19.41	0.617	0.37
	Ours	19.83	0.628	0.38
Wheel	Gaussian Marbles [3]	15.55	0.531	0.53
	SoM [4]	16.41	0.620	0.37
	Ours	17.01	0.620	0.35

Table E. Tracking results on the DyCheck dataset with Percentage of Correct Keypoints (PCK @ 10 cm). We evaluate 3D keypoint tracking following the evaluation protocol on SoM [4] Our GP-GS optimization enables GP-4DGS to reconstruct more accurate 3D scene flow compared to baseline methods.

Scene	**SoM [4] (%)**	**GP-4DGS (ours, %)**
paper-windmill	93.6	**97.1**
apple	65.3	**69.7**
spin	92.3	**93.6**
teddy	84.4	**85.9**
block	75.9	**82.2**
Average	82.3	**85.7**

[4] Qianqian Wang, Vickie Ye, Hang Gao, Weijia Zeng, Jake Austin, Zhengqi Li, and Angjoo Kanazawa. Shape of Motion: 4D reconstruction from a single video. In *ICCV*, 2025. 3, 5

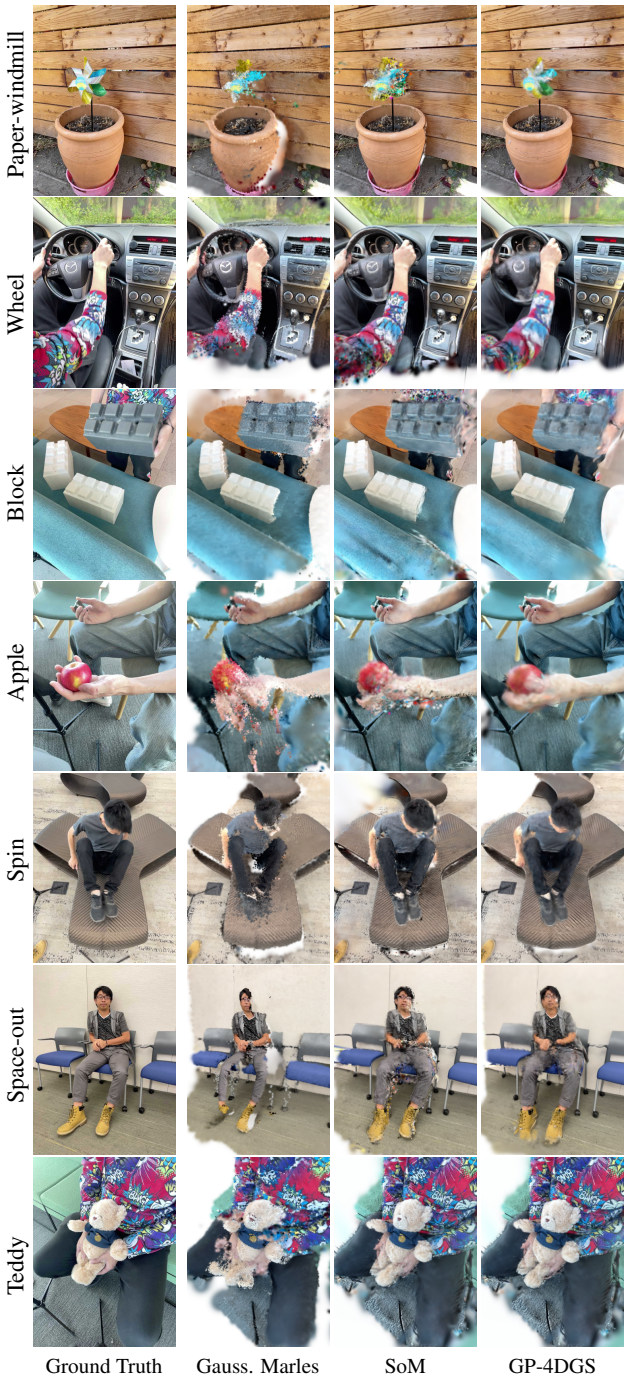


Figure F. Qualitative comparison of novel view synthesis on the DyCheck dataset. Our method produces sharper details and more accurate geometry compared to baselines, particularly in regions with less observation.