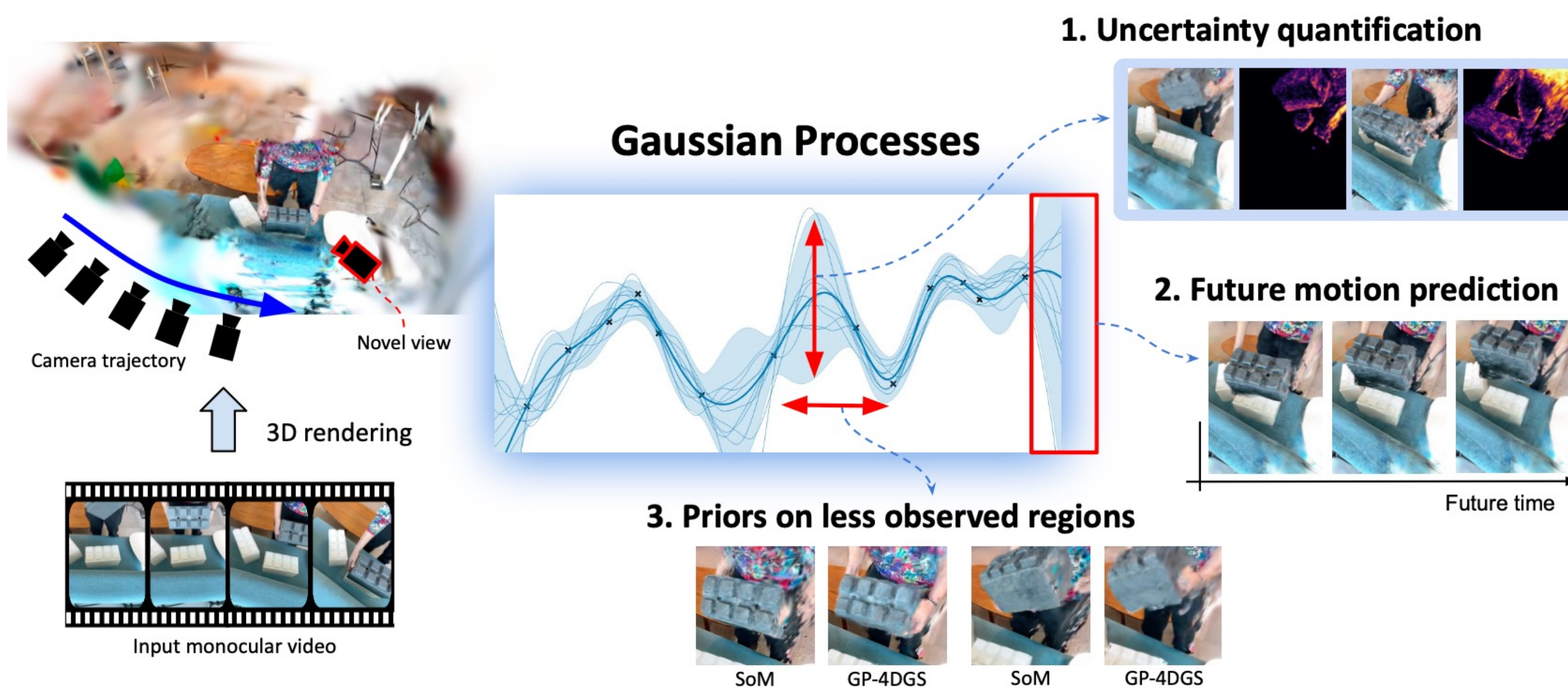


Mijeong Kim¹ Jungtaek Kim² Bohyung Han¹
¹ Seoul National University, Korea ² University of Wisconsin–Madison, USA



Task: Dynamic Novel View Synthesis

- Goal:** Reconstructing dynamic 4D scenes from cell-phone monocular video
- Challenge:**
 - Too sparse information for reconstruction**
 - Existing 4DGS algorithms model unables uncertainty quantification, future motion prediction, and adaptive motion priors.



TL;DR: We introduce **probabilistic deformation modeling** of 4D Gaussian Splatting from monocular video via **Variational Gaussian Processes**.

4D Gaussian Splatting

- 3D Gaussian Splatting** represents static scenes as a set of primitives.
- 4D Gaussian Splatting** adds per-primitive motion over time to model dynamic scenes.

Modeling

[Gaussian Processes]

- A probabilistic model over functions** that provides mean prediction and uncertainty estimate via kernel-based correlation.
- Kernel:** Encodes smoothness and periodicity through correlations between inputs.

1. Probabilistic Deformation Modeling

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

Model Gaussian primitive deformation as a probabilistic function over space and time, where we adopt multi-input / multi-output Gaussian Processes.

- Input:** canonical position & target time $\mathbf{x} = (x_0, y_0, z_0, t) \in \mathbb{R}^4$
- Output:** deformation offset $\mathbf{y} = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_d(\mathbf{x})) \in \mathbb{R}^d$

2. Spatiotemporal Correlation

$$k_i(\mathbf{x}, \mathbf{x}') = k_{i,\text{spatial}}(\mathbf{x}_s, \mathbf{x}'_s) + k_{i,\text{temporal}}(\mathbf{x}, \mathbf{x}')$$

Decomposed kernel:

- Spatial kernel:** matern only
- Temporal kernel:** matern&periodic

$$k_{i,\text{spatial}}(\mathbf{x}_s, \mathbf{x}'_s) = \sigma_{s,i}^2 \frac{2^{1-\nu_i}}{\Gamma(\nu_i)} r_{s,i}^{\nu_i} K_{\nu_i} r_{s,i}$$

$$k_{i,\text{temporal}}(\mathbf{x}, \mathbf{x}') = \sum_{j \in \{x, y, z\}} k_{i,j}(j, j') k_{i,\text{periodic}}(t, t')$$

Training & Inference

1. Scalable Variational Inference

Variational GPs: Inducing point approximation reduces complexity from $O(N^3)$ to $O(NM^2+M^3)$, enabling scalable GP inference over numerous primitives.

- Training:** Variational posterior $q(\mathbf{u}_i) = \mathcal{N}(\mathbf{m}_i, \mathbf{S}_i)$ over inducing points is optimized jointly with kernel hyperparameters θ via ELBO.

$$\mathcal{L}_{\text{ELBO}} = \sum_{i=1}^d [\mathbb{E}_q[\log p(\mathbf{y}_i | \mathbf{u}_i)] - \text{KL}[q(\mathbf{u}_i) \| p(\mathbf{u}_i)]]$$

- Inference:** $\bar{\mu}_i^* = \mathbf{k}_*^{(i)\top} (\mathbf{K}_{ZZ}^{(i)})^{-1} \mathbf{m}_i$ and $\bar{\sigma}_i^{*2} = \mathbf{k}_i^* - \mathbf{k}_*^{(i)\top} \Sigma_i \mathbf{k}_*^{(i)}$

2. Self-Reinforcing Optimization

- GP-GS optimization:** GP is trained on high-confidence primitives (Stage 1) and GS is guided from GP posterior (Stage 2).
- For training GPs,** selecting the confident data from 4D Gaussian Splatting.



Algorithm 1 GP-GS Optimization Strategy

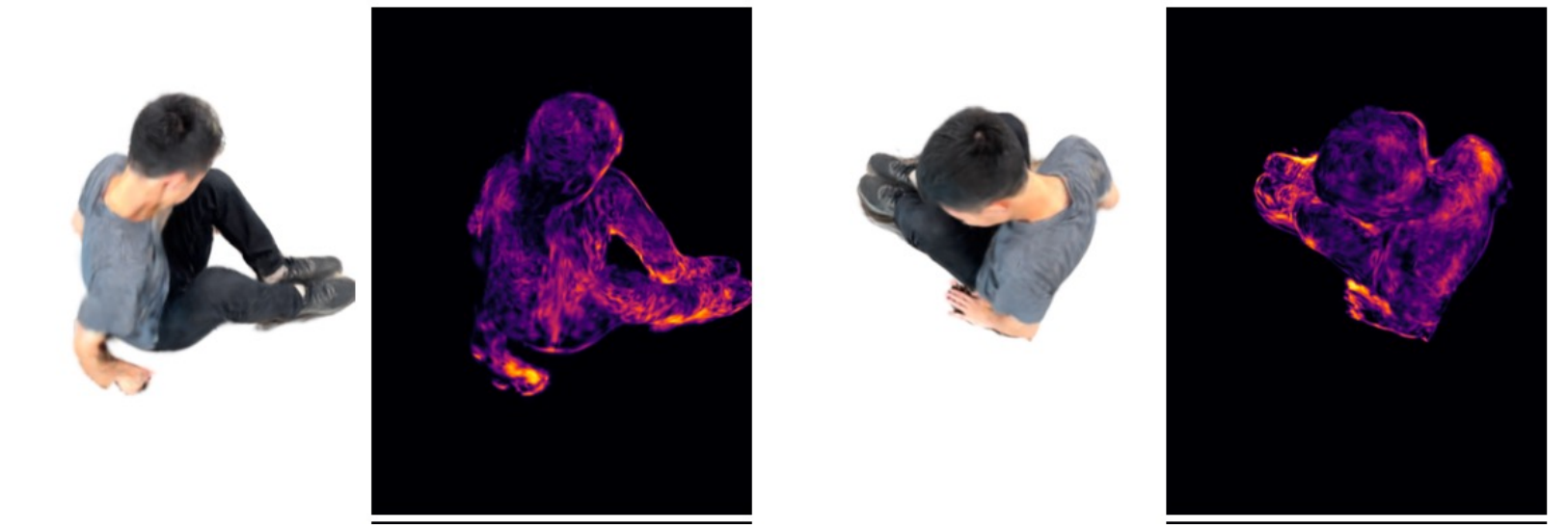
Input: Training images $\mathcal{T}$, Gaussian primitives $\{\gamma_k\}_{k=1}^N$
Output: Optimized GS model with GP regularization

- 1: **while** not converged **do**
- 2: **Stage 1: GP Training** // For every N_{GP} iterations
- 3: Compute confidence: $C_k = \sum_{\mathbf{r} \in \mathcal{T}} \omega_k^c(\mathbf{r})$
- 4: Select confident data: $\mathcal{D}_C = \{(\mathbf{x}_i, \mathbf{y}_i) : C_k > \tau_C\}$
- 5: Train variational GPs on $\mathcal{D}_C$ with Eq. (14)
- 6: **Stage 2: GS Optimization**
- 7: Infer variational GP posteriors and cache $\bar{\mu}_{(k,t)}(\mathbf{x})$
- 8: Compute loss: $\mathcal{L}_{\text{GP}} = \mathbb{E}[\delta_k \cdot \|\mathbf{y}_{(k,t)} - \bar{\mu}_{(k,t)}\|^2]$
- 9: Update GS: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{recon}} + \lambda_{\text{GP}} \mathcal{L}_{\text{GP}}$
- 10: **end while**
- 11: **return** optimized GS and GP models

Application & Results

1. Uncertainty Quantification

- GP posterior variance naturally provides uncertainty estimates for motion prediction, where less-observed regions exhibit higher uncertainty

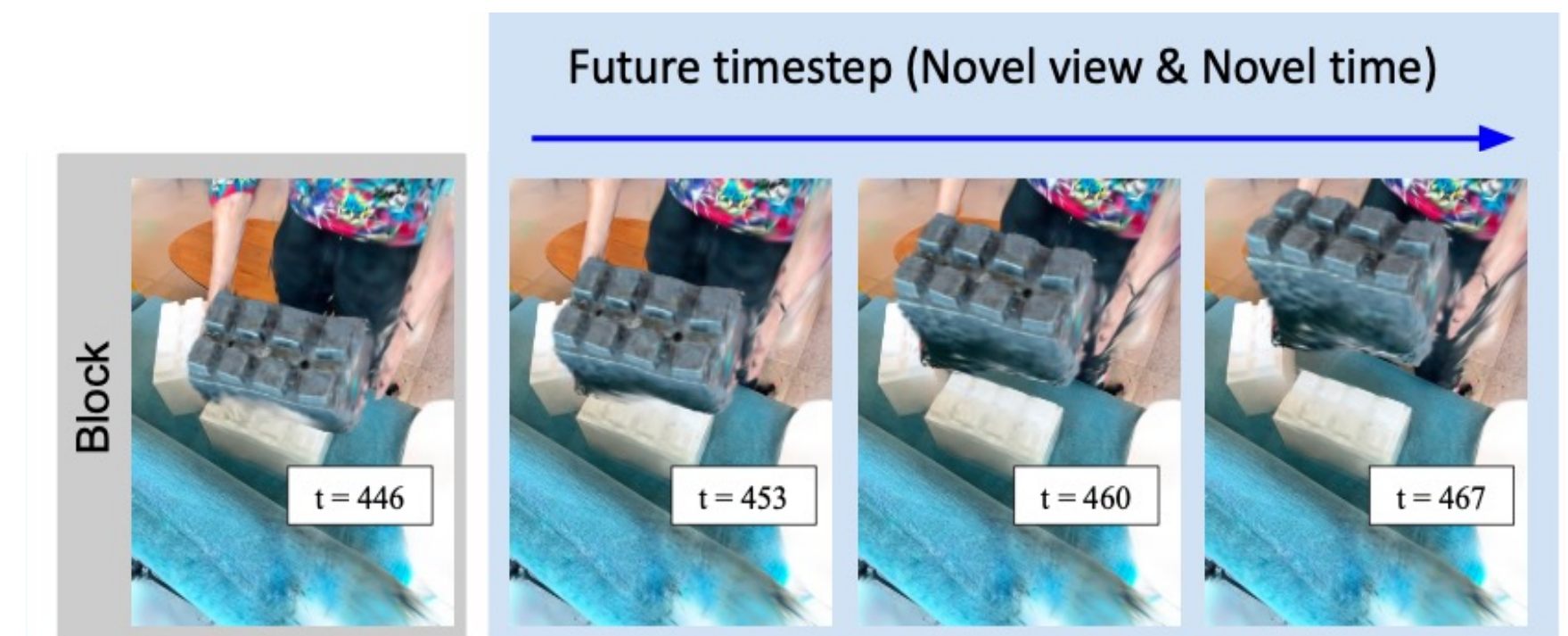


$$U_k = \text{Var}(\{\mathbf{p}_k^{(s)}\}_{s=1}^S) \quad \hat{\mathbf{U}}(\mathbf{r}) = \sum_{k=1}^K U_k \omega_k^\pi(\mathbf{r})$$

Method	AUSE-MSE [40] ($\times 10^{-2}$) ↓		
	Top 20 frames	Top 40 frames	All frames
Random	9.76	9.30	10.98
UA-4DGS [19]	7.60	8.11	8.62
GP-4DGS (ours)	7.22 (-0.38)	8.00 (-0.11)	8.49 (-0.13)

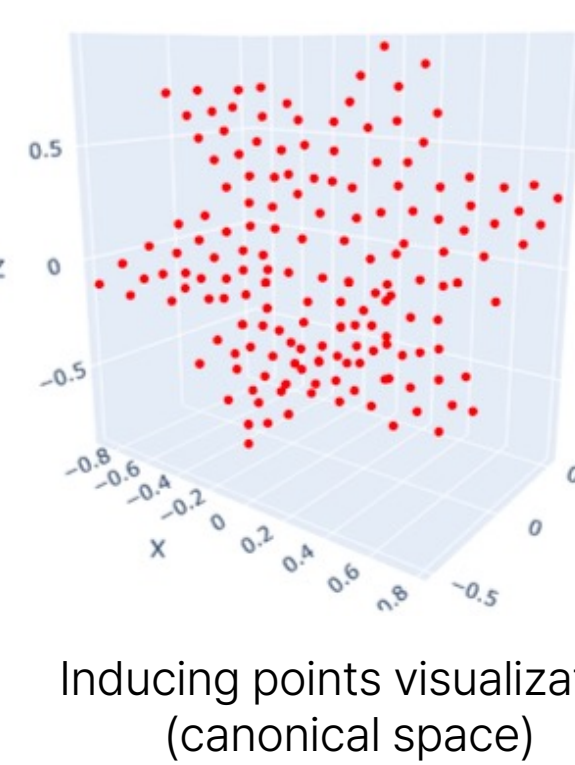
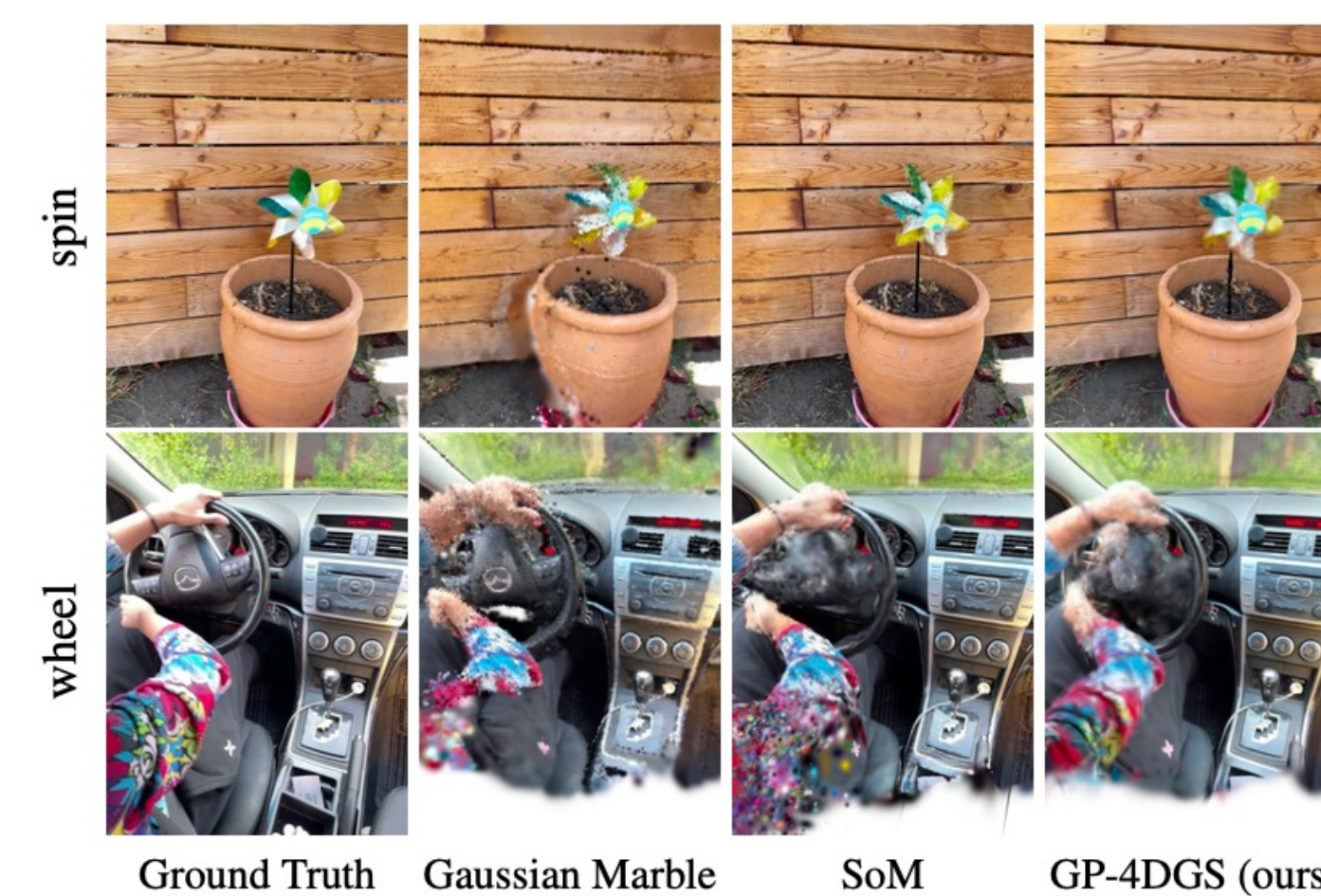
2. Future Motion Prediction

- Querying the trained GP at **future timestamp t^f** (x_0, y_0, z_0, t^f) enables motion prediction beyond the observed training sequence without additional training.



3. Adaptive Motion Priors for Less Observed Area

- GP-4DGS outperforms baselines and improves 3D tracking from 82.3% to 85.7%.



Data	Method	mPSNR ↑	mSSIM ↑	mLPIPS ↓
All	Gaussian Marbles [42]	15.84	0.54	0.57
	SoM [46]	17.09	0.65	0.39
	GP-4DGS (ours)	17.38	0.65	0.37
	SC-GS [23]	14.13	0.48	0.49
SoM 5	D-3DGS [51]	11.92	0.49	0.66
	4DGS [48]	13.42	0.49	0.56
	TNeRF [46]	15.60	0.55	0.55
	DynIBar [22]	15.99	0.59	0.51
	Gaussian Marbles [42]	13.41	0.48	0.55
	SoM [46]	16.03	0.54	0.58
	SoM [46]	16.73	0.64	0.43
	GP-4DGS (ours)	16.92	0.66	0.41
Challenging subset [14]	Gaussian Marbles [42]	14.05	0.40	0.61
	SoM [46]	14.56	0.46	0.53
	GP-4DGS (ours)	15.02	0.46	0.51

Conclusion

- Conclusion:** Our novel probabilistic framework capture spatiotemporal correlations by leveraging GP-based motion priors.
- Impact:** Uncertainty and future-motion outputs can inform downstream decision-making modules, such as planning in autonomous driving.

Organizers

