# MoDGS: Dynamic Gaussian Splatting from Casually-captured Monocular Videos with Depth Priors

Qingming Liu*, Yuan Liu*, Jiepeng Wang, Xianqiang Lyv, Peng Wang, Wenping Wang, Junhui Hou†

*Equal Contribution † Corresponding Author

ICLR

Project Page | OpenReview

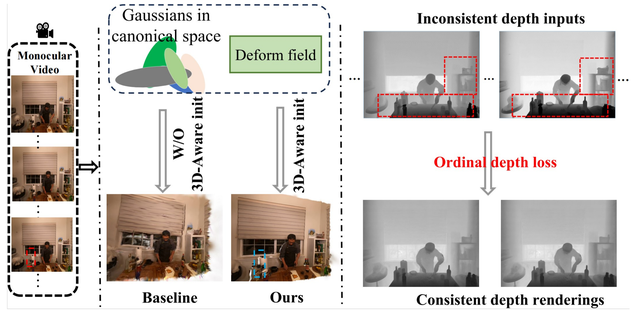
*Project Page:* ***modgs.github.io***

## Motivation:

- Novel view synthesis in dynamic scenes (DVS) is an important task in computer graphics and computer vision.
- Most of the existing DVS methods require multi-view consistency provided by pseudo monocular videos recorded by cameras with large movement or "teleporting motion"[1].
- When applied to casually captured videos with weaker multi-view consistency constraints, these methods struggle to generate high-quality novel-view renderings.

*[1] Gao et al., Monocular Dynamic View Synthesis: A Reality Check. NeurIPS 2022.*

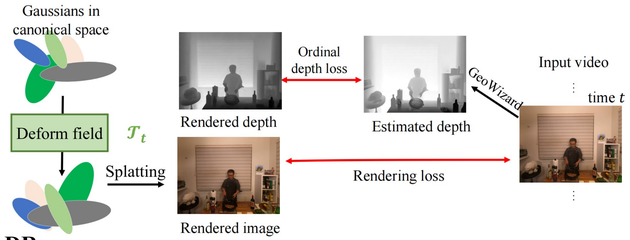
## Contribution:




- We propose a 3D-aware initialization that greatly helps the learning of the 4D representations and improves the rendering quality.
- We propose an ordinal depth loss that handles inconsistent depth inputs and effectively ensures more reliable and consistent depth outputs

## Method:

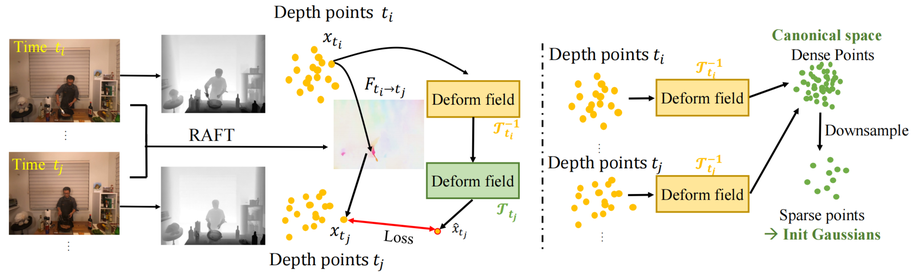
### Overview




**TL; DR:**

We use a deformation field to deform gaussians in the canonical space and propose an ordinal depth loss to deal with inconsistent depth prior inputs.

### 3D-aware Initialization

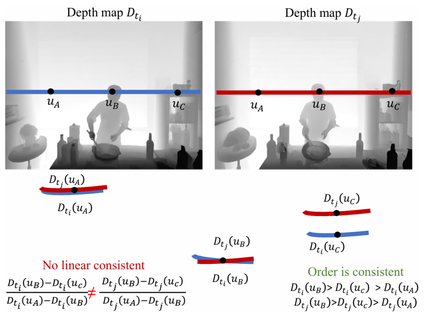



**(a) Initialization of deformation field** **(b) Initialization of Gaussians**

(a) **Initialization of deformation field.** We first lift the depth maps and a 2D flow to a 3D flow and train the deformation field for initialization.

(b) **Initialization of Gaussians in canonical space.** We use the initialized deformation field to deform all the depth points to canonical space and downsample these points to initialize Gaussians.

### Ordinal depth loss




**Intuition:**

We observe that, while depth consistency cannot be fully guaranteed, the relative order of depth values remains consistent across different frames. We first define an order indicator function:

$$\mathcal{R}(D_t(u_1), D_t(u_2)) = \begin{cases} +1, & D_t(u_1) > D_t(u_2) \\ -1, & D_t(u_1) < D_t(u_2) \end{cases}$$

Next, we approximate this indicator function using the tanh activation function, and subsequently define the ordinal depth loss as follows:

$$\ell_{\text{ordinal}} = \left\| \tanh\left(\alpha(\hat{D}_t(u_1) - \hat{D}_t(u_2))\right) - \mathcal{R}\left(D_t(u_1), D_t(u_2)\right) \right\|$$

## Experiments:

### Comparison

| Methods | DyNeRF | | | Nvidia | | |
|---|---|---|---|---|---|---|
| | PSNR↑ | SSIM↓ | LPIPS↓ | PSNR↑ | SSIM↓ | LPIPS↓ |
| HexPlane | 15.33 | 0.5593 | 0.4514 | 17.17 | 0.3675 | 0.4756 |
| SC-GS | 18.77 | 0.7359 | 0.2310 | 17.59 | 0.4679 | 0.3348 |
| D-GS | 19.55 | 0.7446 | 0.2171 | 18.07 | 0.4650 | 0.3422 |
| Ours | **22.64** | **0.8042** | **0.1545** | **19.27** | **0.5235** | **0.2581** |

Our method achieves the best performances in all metrics on both datasets.


Baseline methods fail to correctly reconstruct the 3D geometry of the dynamic scenes and produce obvious artifacts on both dynamic foreground and static background.

### Ablation Studies


Our ordinal depth loss enables a smooth reconstruction of the depth map in the interior while maintaining sharp edges at boundaries.


Compared to our 3D-aware initialization, random initialization results in more artifacts.

| 3D-aware Init | Loss | PSNR↑ | SSIM↑ | LPIPS↓ |
|---|---|---|---|---|
| × | Ordinal | 21.27 | 0.7655 | 0.1984 |
| ✓ | Pearson | 21.77 | 0.7938 | 0.1680 |
| ✓ | Ordinal | **22.96** | **0.8103** | **0.1518** |

**Our website provides additional video comparison results.**

***One more thing!*** *Qingming is actively looking for a PhD position!*