Not All Frame Features Are Equal: Video-to-4D Generation via Decoupling Dynamic-Static Features

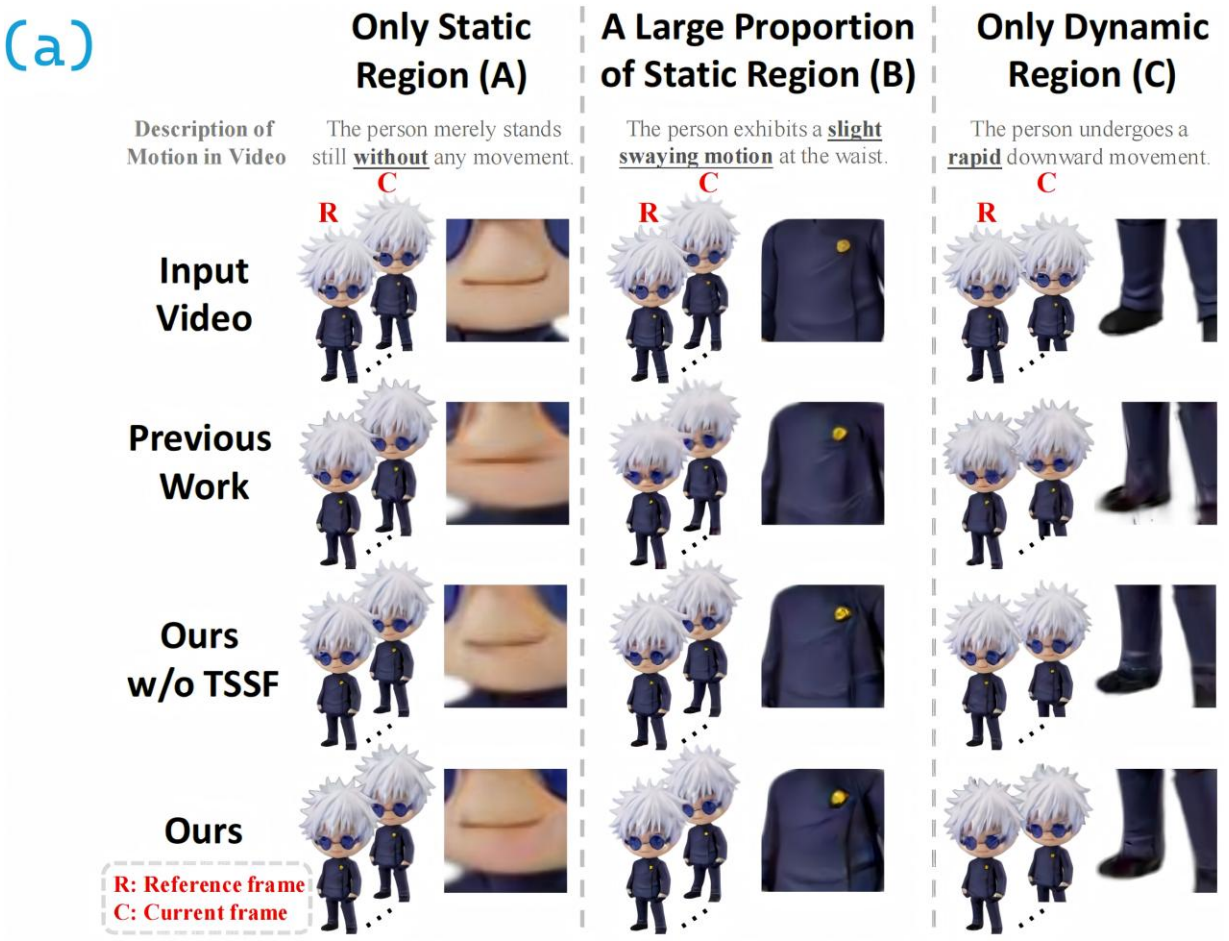
Liying Yang, Chen Liu, Zhenwei Zhu, Ajian Liu, Hui Ma, Jian Nong, Yanyan Liang

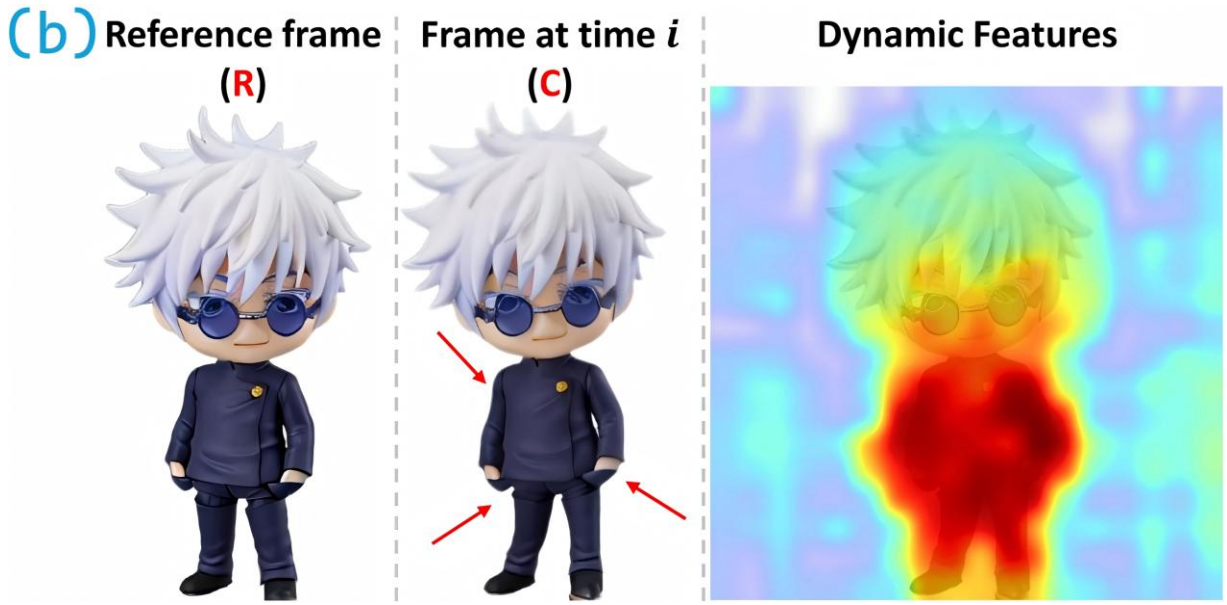
Motivation

- Existing methods directly optimize Gaussians using whole information in frames.
- If the static regions account for a large proportion in frames, existing methods often overlook information in dynamic regions and are prone to overfitting on static regions. This leads to producing results with blurry textures.

Motivation

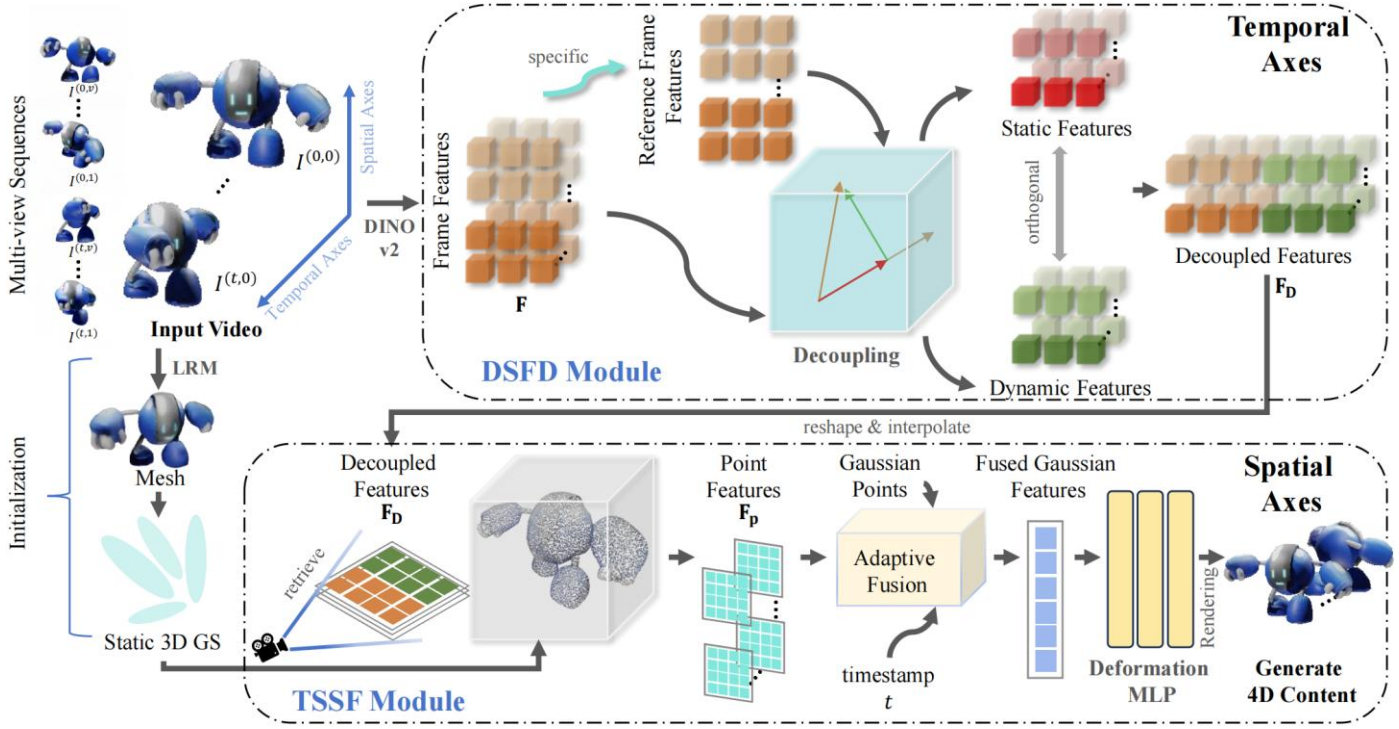
- We think **decoupling dynamic-static features** can solve this problem.





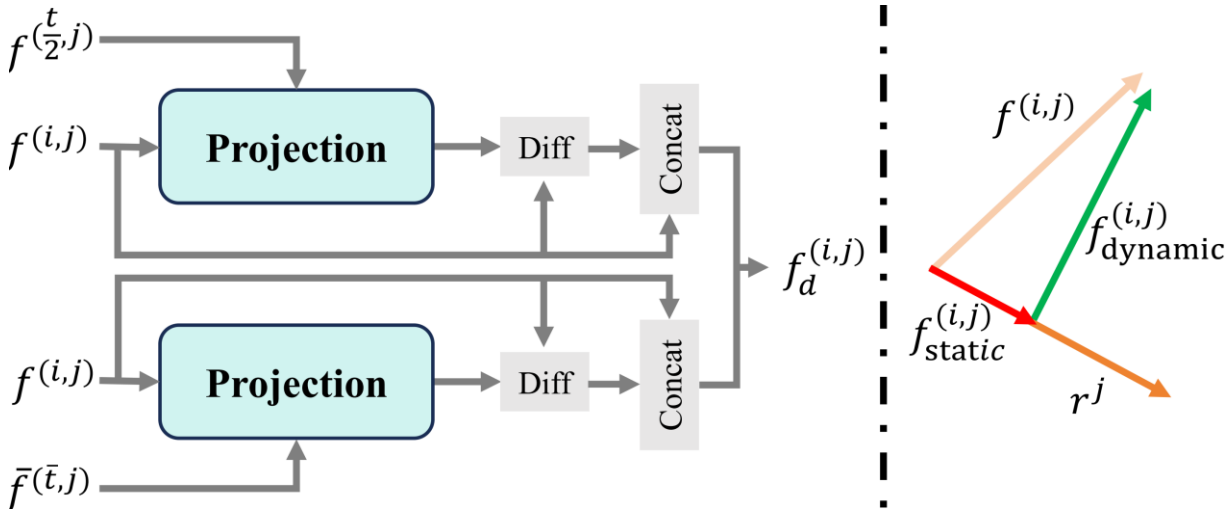
Methodology

- **Dynamic-static feature decoupling module (DSFD):** Along temporal axes, it regards the regions of current frame features that possess significant differences relative to reference frame features as dynamic features.
- **Temporal-spatial similarity fusion module (TSSF):** Along spatial axes, it adaptively selects similar information of dynamic regions.



Core of Decoupling

- We observe that regions with larger semantic magnitudes across time tend to exhibit higher semantic consistency, **while dynamic regions often yield smaller projection values** due to variations in object pose and position. Thus, **orthogonal projection** is applied between two frame features to decouple.



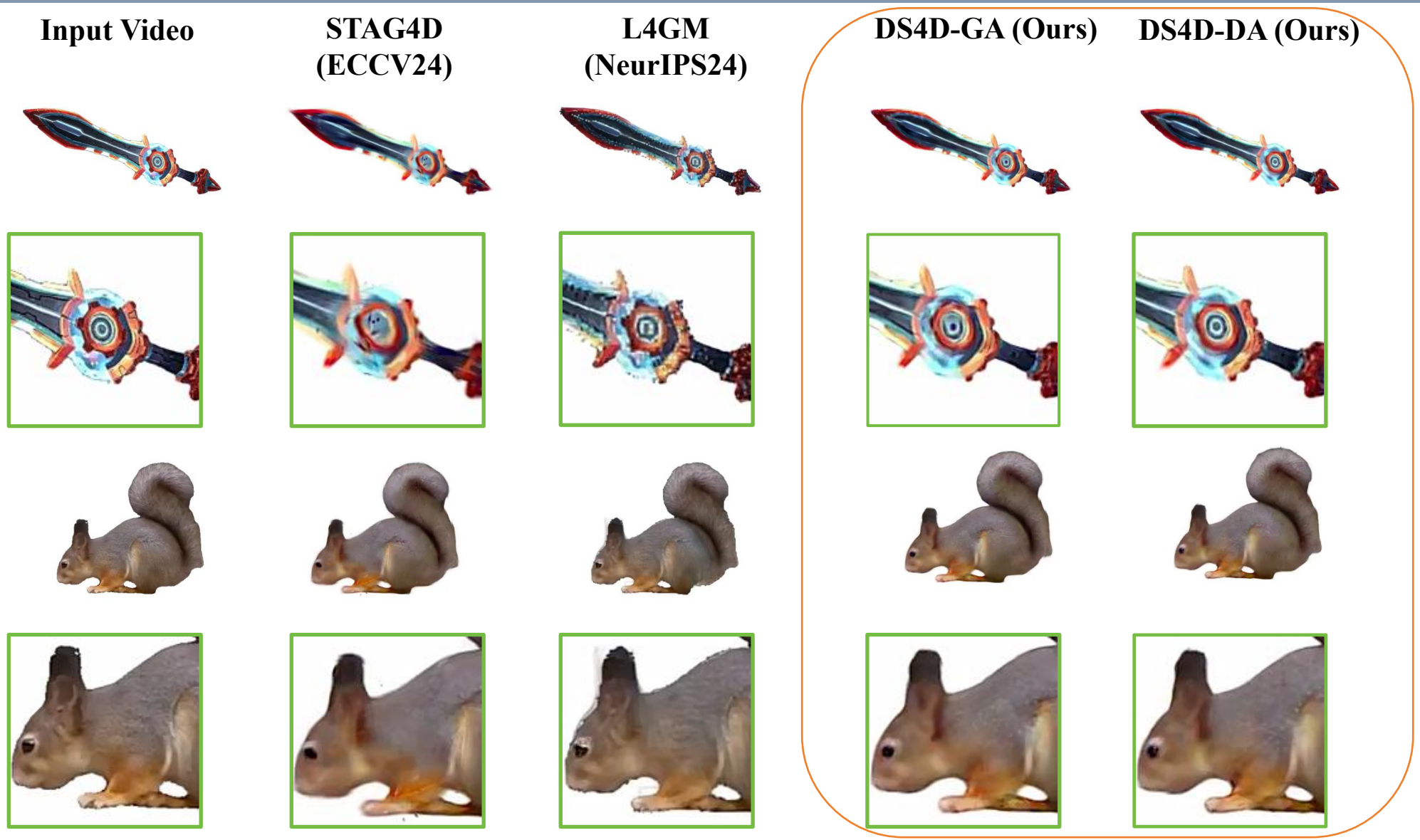
Comparisons with Existing Methods

- The best score is highlighted in bold. For a fair comparison, the experiment of 4Diffusion and L4GM on Objaverse dataset are disregarded since they are inference-based methods trained on this dataset.

Methods	Optimization	Consistent4D dataset				Objaverse dataset			
		CLIP ↑	LPIPS ↓	FVD ↓	FID-VID ↓	CLIP ↑	LPIPS ↓	FVD ↓	FID-VID ↓
Consistent4D	✓	0.9085	0.1316	1041.2242	28.6471	0.8491	0.2222	1814.5652	48.7921
Dreamgaussian4D	✓	0.9145	0.1517	844.9087	37.9977	0.8127	0.2017	1545.3009	58.3686
STAG4D	✓	0.9078	0.1354	986.8271	26.3705	0.8790	0.1811	1061.3582	30.1359
SC4D	✓	0.9117	0.1370	852.9816	26.4779	0.8490	0.1852	1067.7582	40.5130
4Diffusion	✗	0.8734	0.2284	1551.6363	149.6170	-	-	-	-
L4GM	✗	0.9158	0.1497	898.0604	31.4996	-	-	-	-
DS4D-GA (Ours)	✓	0.9206	0.1311	799.9367	26.1794	0.8868	0.1761	890.2646	26.6717
DS4D-DA (Ours)	✓	**0.9225**	**0.1309**	**784.0235**	**24.0492**	**0.8881**	**0.1759**	**870.9489**	**25.3836**

Comparisons with Existing Methods

ICCV OCT 19-23, 2025 HONOLULU HAWAII



☒

Due to previous works overfitting on static regions, they easily generate unsatisfactory results.

☑

Our methods **are the first to** decouple dynamic-static information in frames along temporal-spatial axes for mitigating such issues.

Application: Dynamic Scene Render

Ours

GT	4D-GS	Ours
		
PSNR:	27.88	29.54

Thank you for tuning in!