[SplineGS]

Robust Motion-Adaptive Spline for Real-Time Dynamic 3D Gaussians from Monocular Video


Jongmin Park[1*]


Minh-Quan Viet Bui[1*]


Juan Luis Gonzalez Bello[1]


Jaeho Moon[1]


Jihyong Oh[2†]


Munchurl Kim[1†]

1: **KAIST**, 2 : **Chung-Ang Univ.**

* : Co-first authors (equal contribution), † : Co-corresponding authors

VICLab
Video and Image Computing Lab

CMLab
Creative Vision and Multimedia

Dynamic Novel View Synthesis (Dynamic NVS)

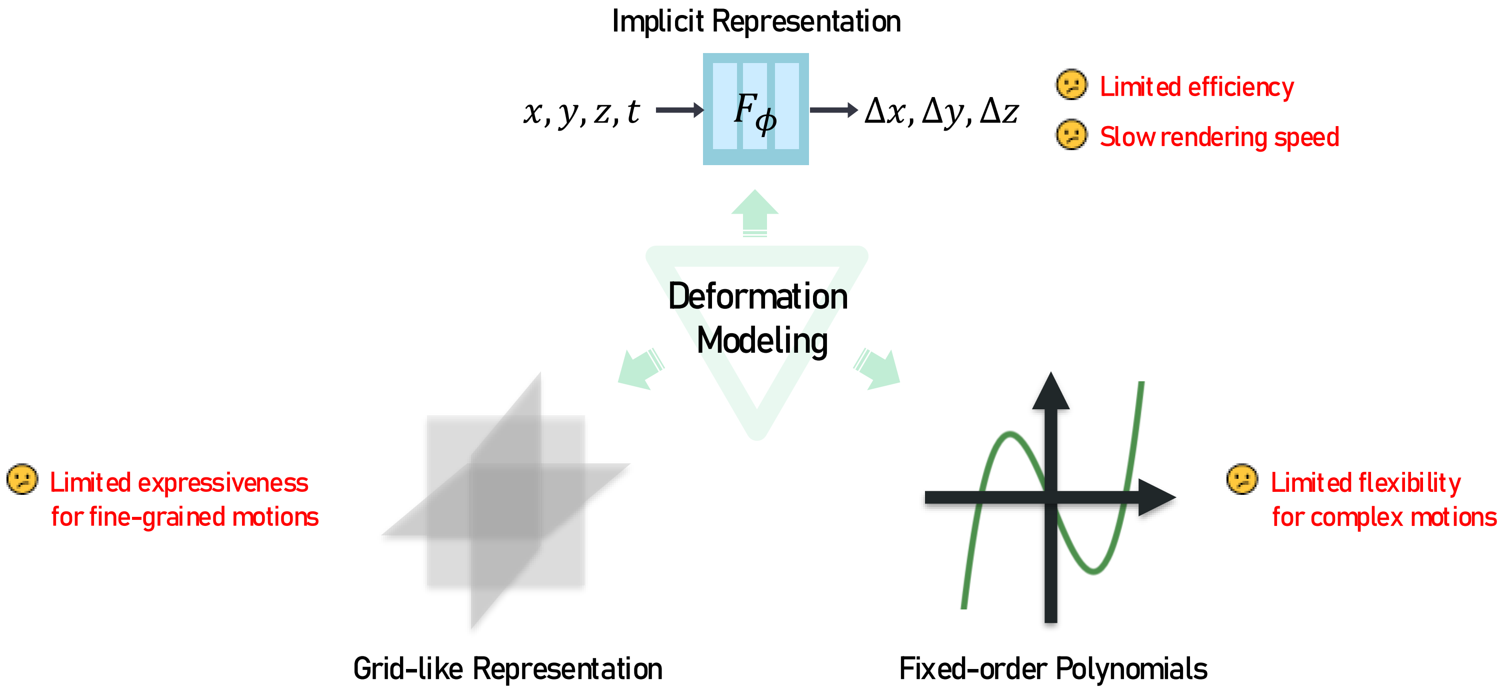
- **Goal:** To generate video frames from any viewpoint in a dynamic scene.

- Requires accurate motion modeling of moving objects.
- Demands precise estimation of camera parameters.

Motivation

❑ Limitation of prior models

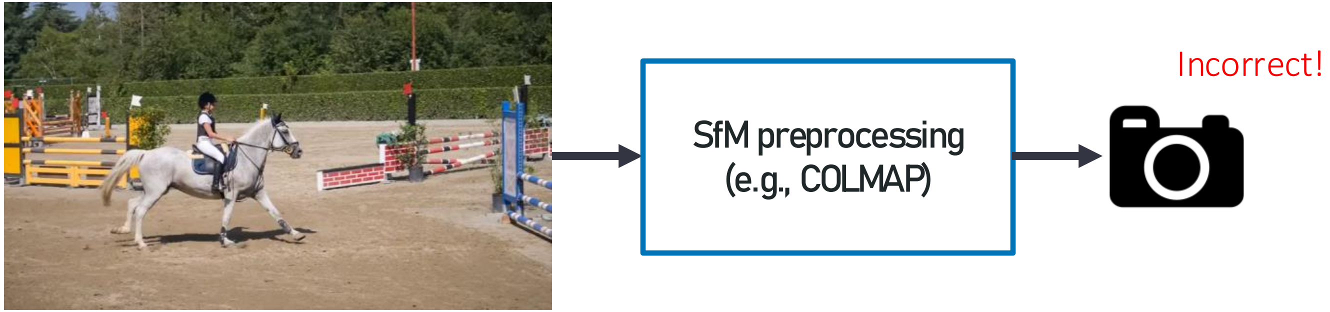
- Existing deformation methods face several challenges.



Motivation

Limitation of prior models

- Most dynamic NVS methods rely on Structure-from-Motion (SfM) algorithms (e.g., COLMAP).
- However, it often produce imprecise prediction for in-the-wild monocular videos.



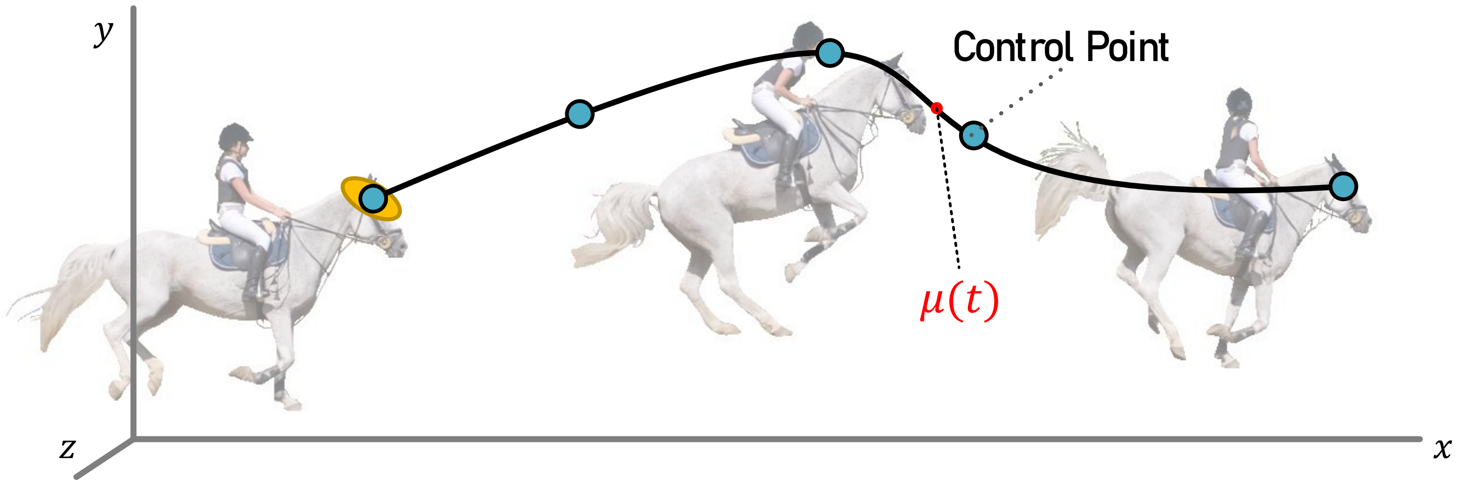
in-the-wild videos with large motion

Proposed Method: SplineGS

Motion-Adaptive Spline for dynamic 3D Gaussians

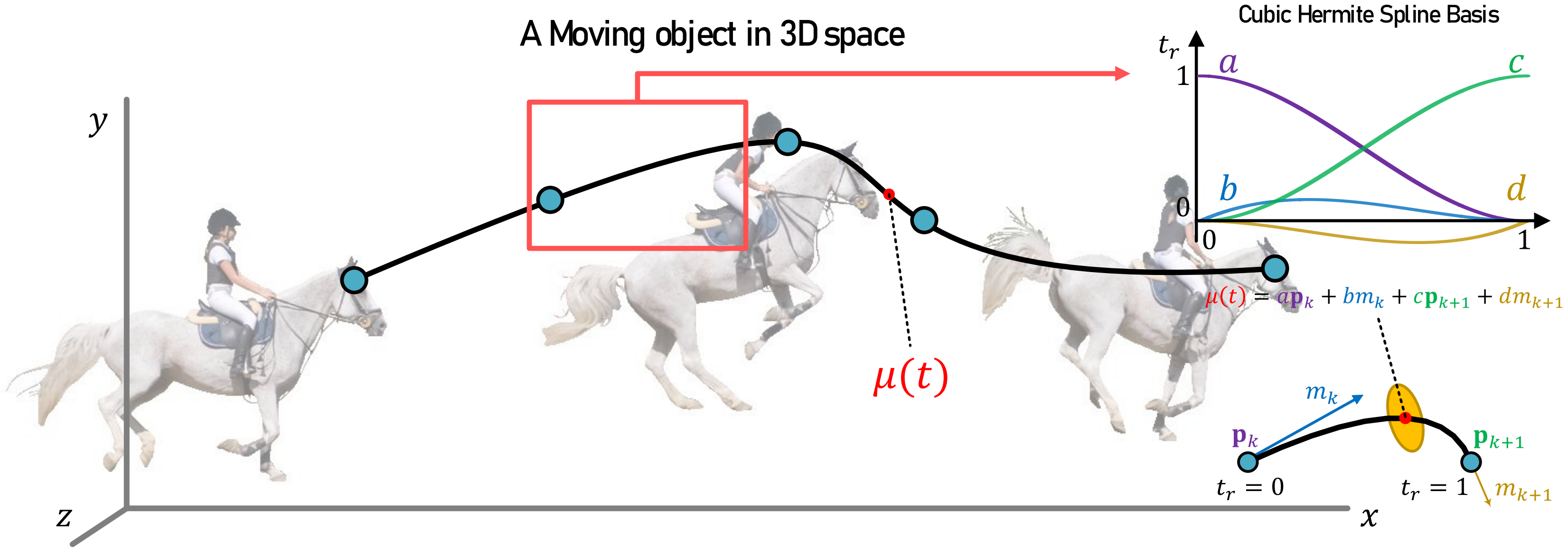
- Cubic Hermite spline functions for time-varying smooth and continuous deformation.

A Moving object in 3D space



Proposed Method: SplineGS

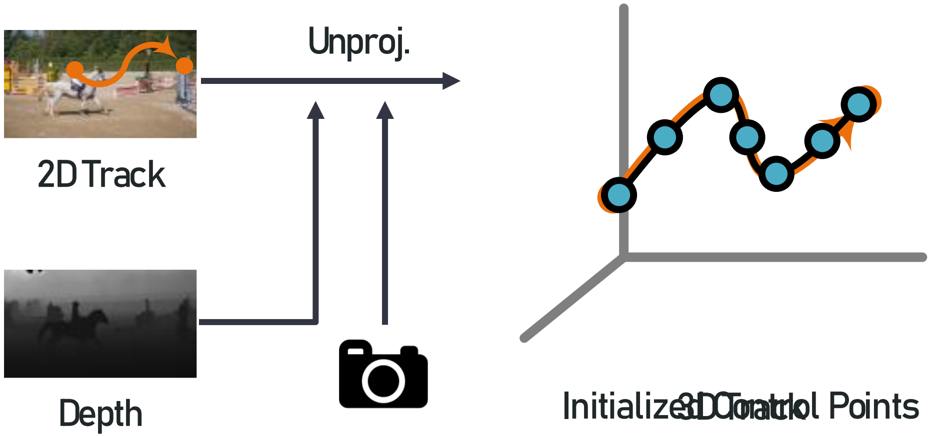
- **Motion-Adaptive Spline for dynamic 3D Gaussians**
  - Cubic Hermite spline functions for time-varying smooth and continuous deformation.



Proposed Method: SplineGS

❑ Motion-Adaptive Spline for dynamic 3D Gaussians

- To initialize the control points, we leverage 2D tracking* and metric depth**.
- We lift 2D tracks into 3D space using metric depth and camera parameters.

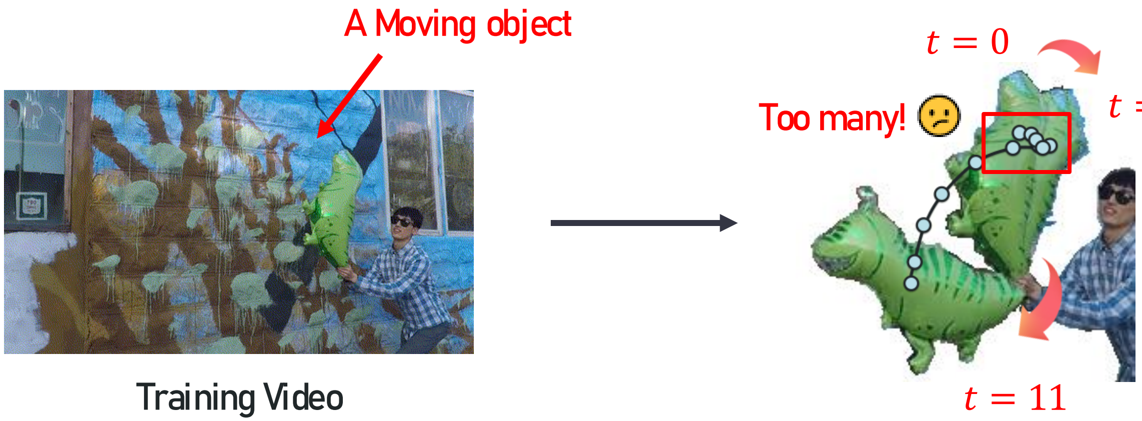


*CoTracker: It Is Better to Track Together, Karaev *et al.*, ECCV 2024
**UniDepth: Universal Monocular Metric Depth Estimation, Piccinelli *et al.*, CVPR 2024

Proposed Method: SplineGS

Motion-Adaptive Spline for dynamic 3D Gaussians

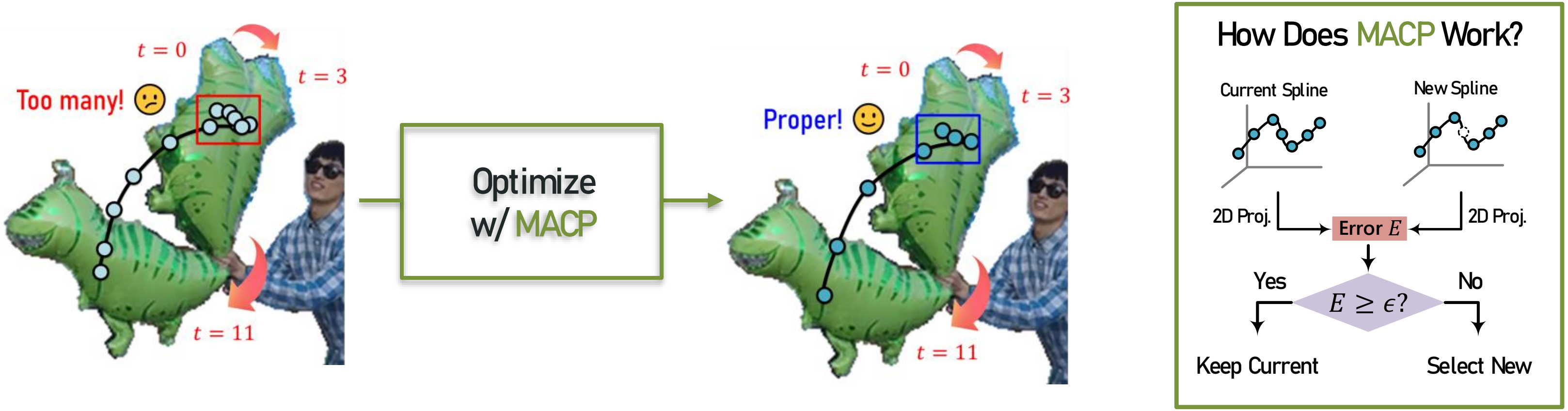
- An excessive number of control points may cause over-fitting and slow rendering speed.
- The number of control points should be adaptively adjusted based on object dynamics.



Proposed Method: SplineGS

❑ Motion-Adaptive Control points Pruning (MACP)

- ➥ MACP can generate sparser control points while ensuring no performance degradation.
- ➥ MACP reduces the number of control points and replaces the original set if the error is acceptable.



Proposed Method: SplineGS

Camera Parameter Estimation

- SplineGS estimates camera parameters for joint optimization with the Gaussian attributes.

$$\gamma(t) \longrightarrow F_\theta \longrightarrow [\hat{R}_t | \hat{T}_t]$$

Camera Extrinsic

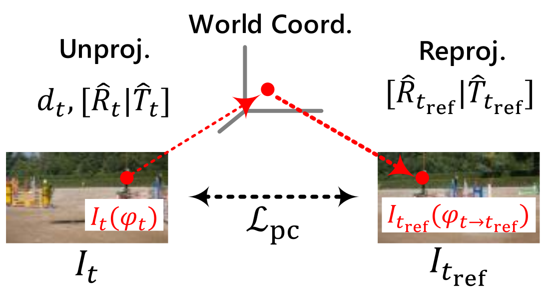
$$\text{Learn. Param.} \longrightarrow \hat{f}$$

Camera Intrinsic

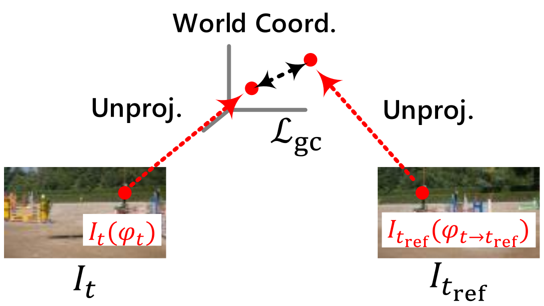
Proposed Method: SplineGS

Camera Parameter Estimation

- SplineGS estimates camera parameters for joint optimization with the Gaussian attributes.
- To enforce two types of consistency: Photometric and Geometric.



Photometric consistency



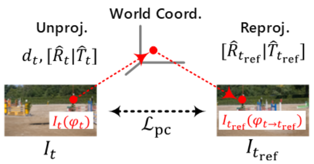
Geometric consistency

Proposed Method: SplineGS

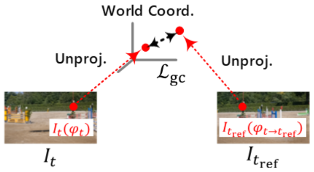
Optimization

- Two-stage optimization
  - Warm-up stage: To optimize camera parameters only
  - Main training stage: To jointly optimize camera parameters and Gaussian attributes

Warm-up Stage



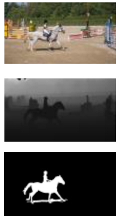
Photometric consistency



Geometric consistency

Main training Stage

Gaussian Optim.



- Rendered RGB
- Rendered depth
- Rendered motion mask

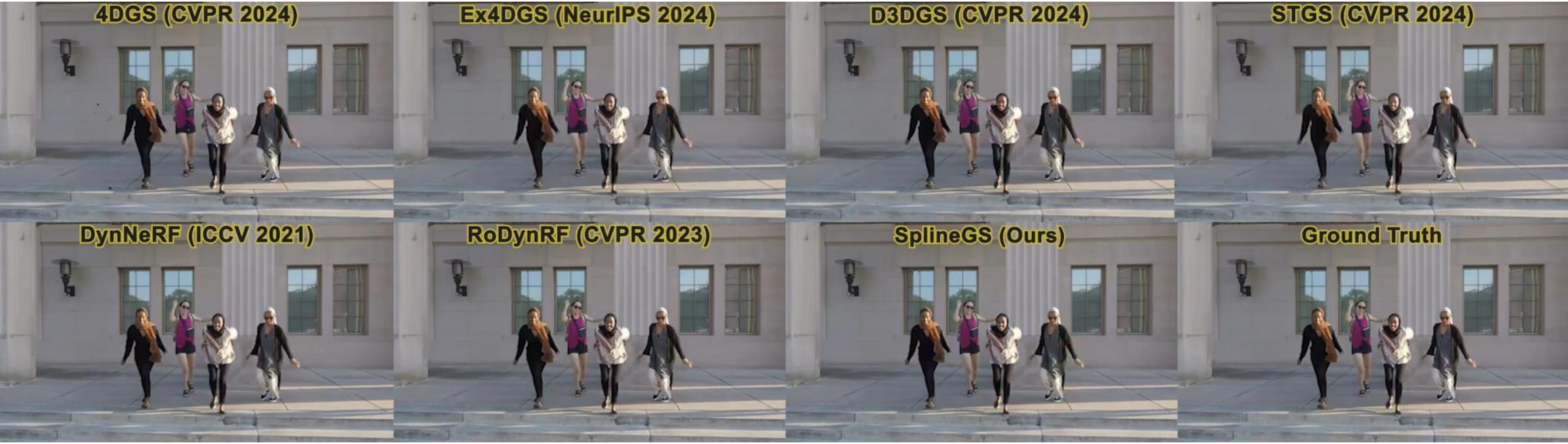
Cam. Param. Optim.

- Photometric consistency
- Geometric consistency
- Rendered-depth-based Photometric consistency

Experiment Results

❑ Novel View Synthesis

➥ On Nvidia Dataset



Experiment Results

❑ Novel View Synthesis

➥ On Nvidia Dataset

1st / 2nd

Methods	PSNR↑/LPIPS↓	FPS↑	Training Time (hr) ↓
DynNeRF (ICCV'21)	26.10 / 0.082	0.05	>24
D3DGS (CVPR'24)	23.02 / 0.195	25	1.0
4DGS (CVPR'24)	23.64 / 0.123	95	**0.25**
SP-GS (ICML'24)	25.70 / 0.247	180	6.0
Casual-FVS (ECCV'24)	24.57 / 0.081	48	**0.25**
E-D3DGS (ECCV'24)	23.36 / 0.129	45	2.6
Shape-of-Motion (arXiv'24)	24.21 / 0.102	255	1.1

Methods	PSNR↑/LPIPS↓	FPS↑	Training Time (hr) ↓
DynPoint (NeurIPS'23)	26.53 / 0.068	0.8	0.5
RoDynRF (CVPR'23) w/ COLMAP	25.89 / 0.067	0.45	>24
RoDynRF (CVPR'23) w/o COLMAP	25.38 / 0.079	0.45	>24
MoSca (arXiv'24) w/ COLMAP	26.55 / 0.070	40	0.8
MoSca (arXiv'24) w/o COLMAP	26.61 / 0.069	40	1.0
Ours (SplineGS) w/ COLMAP	26.90 / 0.060	**400**	0.5
Ours (SplineGS) w/o COLMAP	**27.21** / **0.053**	**400**	1.5

Experiment Results

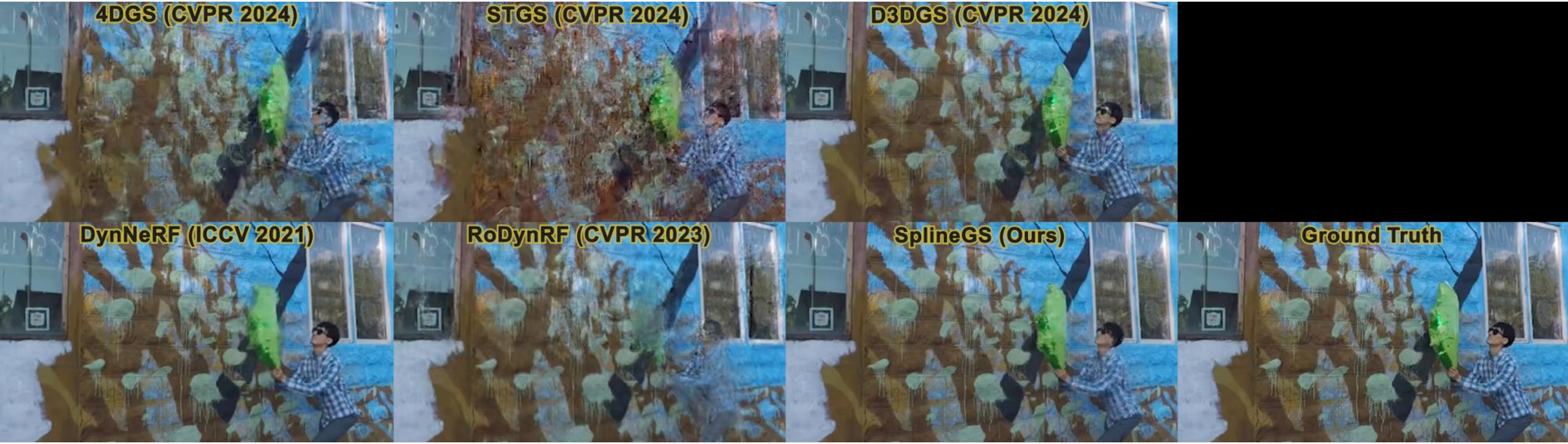
- **Novel View Synthesis**
  - On DAVIS Dataset (Spiral & Zoom)

RoDynRF (CVPR'23)

Ours (SplineGS)

Experiment Results

- **Novel View & Time Synthesis**
  - On Nvidia Dataset



Experiment Results

Novel View & Time Synthesis

- On Nvidia Dataset

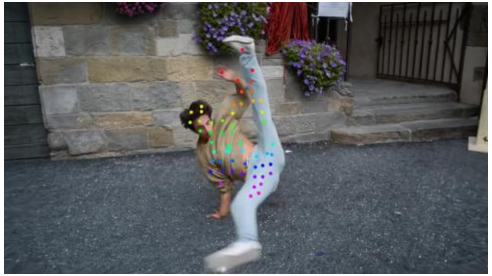
1st / 2nd

Methods	PSNR↑	LPIPS↓	tOF↓
DynNeRF (ICCV'21)	23.36	0.219	0.921
4DGS (CVPR'24)	17.07	0.459	6.314
D3DGS (CVPR'24)	19.63	0.343	3.225
STGS (CVPR'24)	15.72	0.474	2.105
RoDynRF (CVPR'23)	21.58	0.221	2.138
Ours (SplineGS)	**25.92**	**0.098**	**0.703**

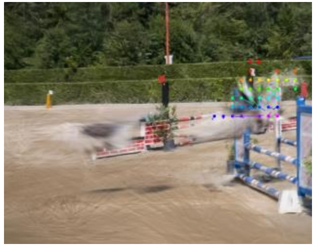
Experiment Results

❑ 3D Motion Tracking

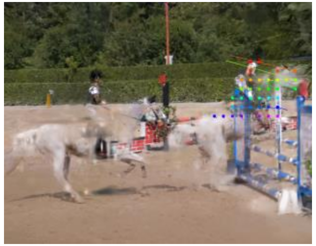
➥ On DAVIS Dataset





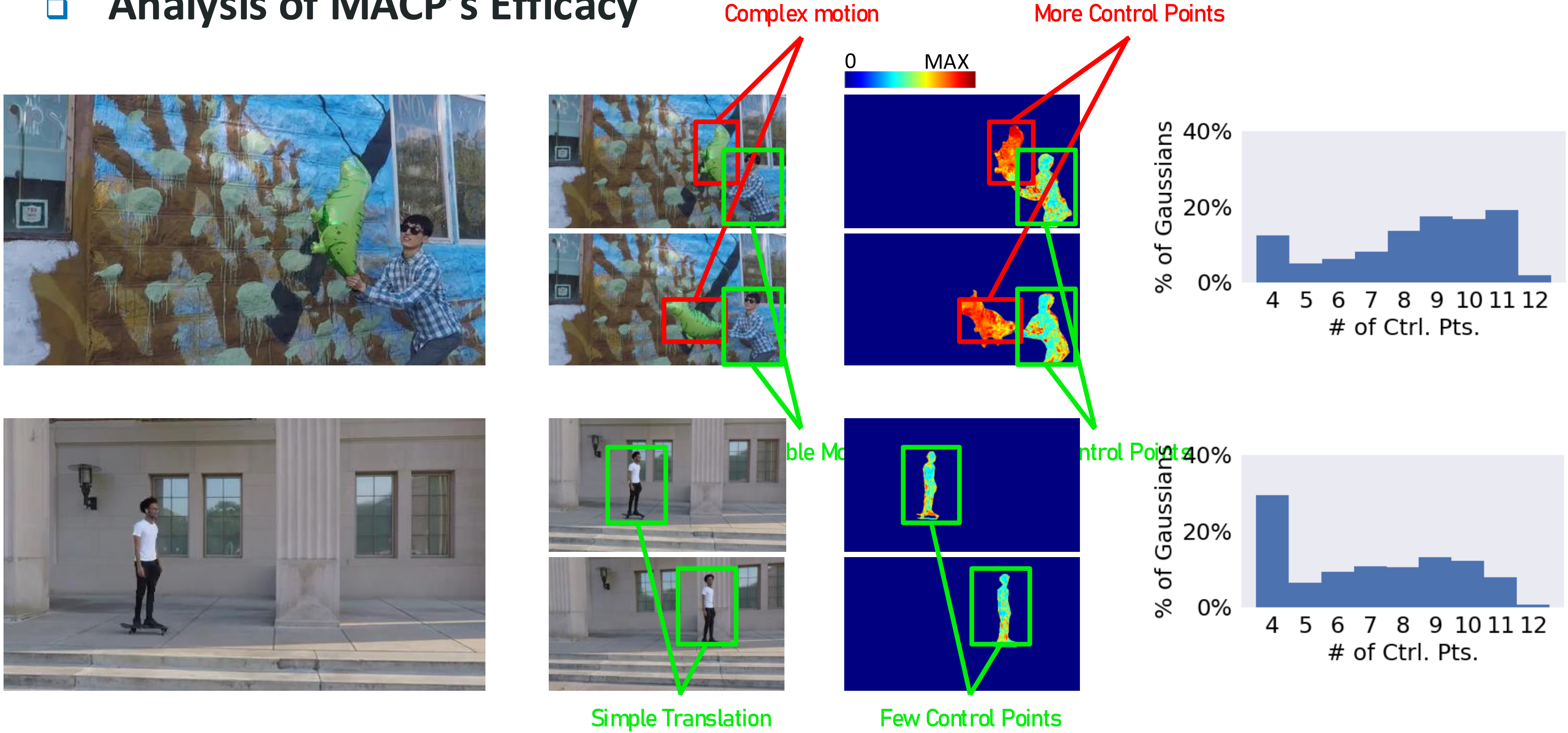
D3DGS (CVPR'24)



STGS (CVPR'24)

Experiment Results

❑ Analysis of MACP's Efficacy



[SplineGS]

Robust Motion-Adaptive Spline for Real-Time Dynamic 3D Gaussians from Monocular Video

Thank You!

☜ For more details, please visit here!

VICLab
Video and Image Computing Lab

KAIST
KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY
1971
한국과학기술원

CHUNG-ANG UNIVERSITY
1916
CAU

Lab
Creative Vision and Multimedia