ZHEJIANG UNIVERSITY

MANCHESTER 1824 The University of Manchester

ACM Multimedia 2024 Melbourne, Australia

LoopGaussian: Creating 3D Cinemagraph with Multi-view Images via Eulerian Motion Field

Jiyang Li*, Zhejiang University
Lechao Cheng*, Hefei University of Technology
Zhangye Wang, Zhejiang University
Tingting Mu, The University of Manchester
Jingxuan He✉, Hefei University of Technology

*Equal contribution. ✉Corresponding author.

MM 2024 Oral

Oct. 2024

Contents

Introduction

What is a Cinemagraph?

A Cinemagraph is a combination of a still image and a video, where most of the scene is stationary, while a section moves on a continuous loop.

--Adobe

The train to Machu Picchu

Client: Chopard

Avenue Matignon

Introduction

Related Work of Cinemagraph

Traditional manual generation method

Limitations: requires a lot of manual work by skilled artist

Automatic generation method

[1]

[1]

[2]

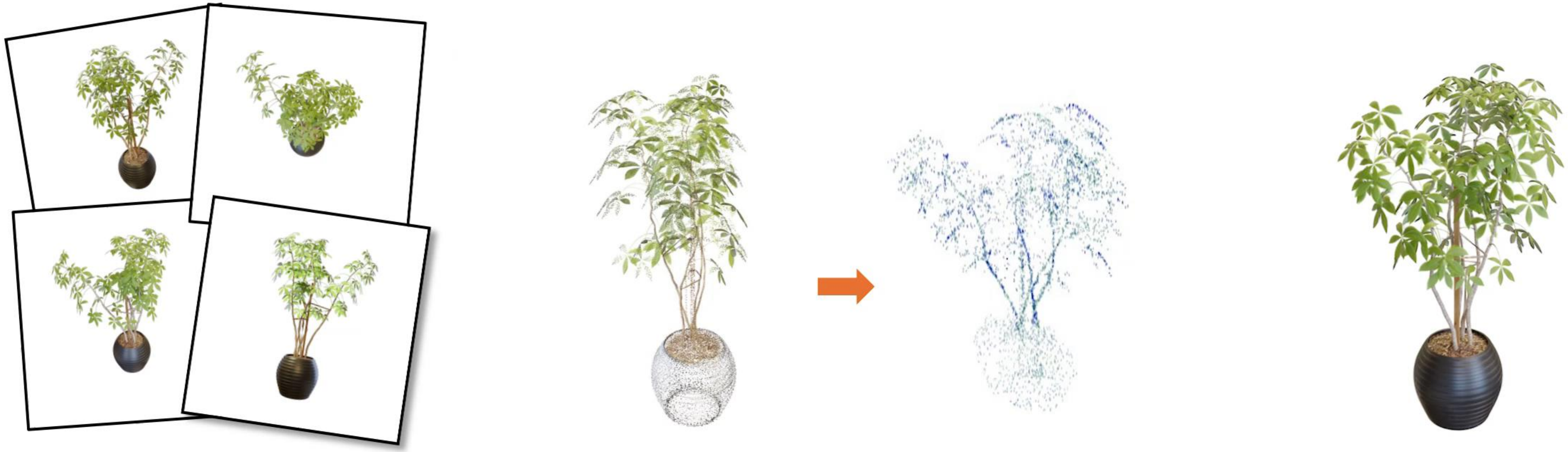
Limitations: requires pre-training on large datasets

Limitations: limited to image space, cannot change viewing point

[1] Animating Pictures with Eulerian Motion Fields.
[2] Controllable Animation of Fluid Elements in Still Images.

Purpose of Our Work

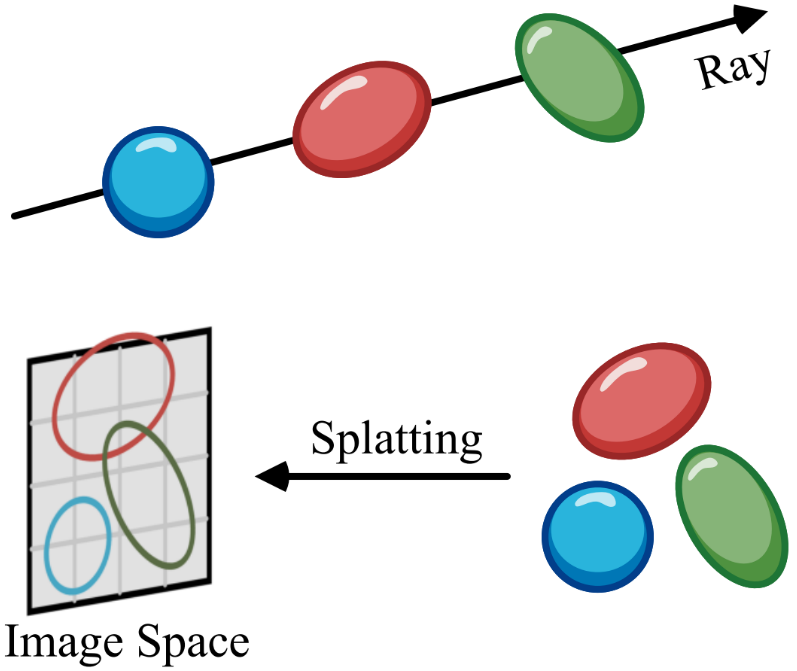
Create an authentic 3D cinemagraph from multi-view images of a stationary scene by an Eulerian motion field.



(a) Multi-view Images (b) Eulerian Motion Field (c) Loopable Video

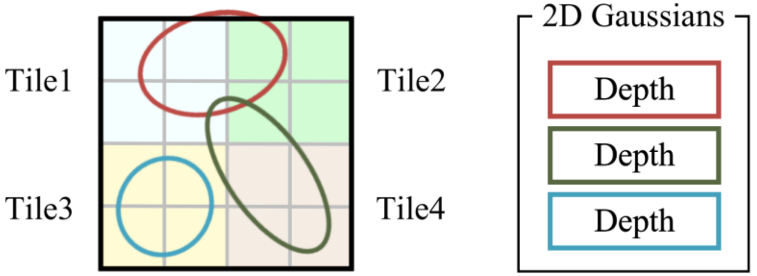
Introduction

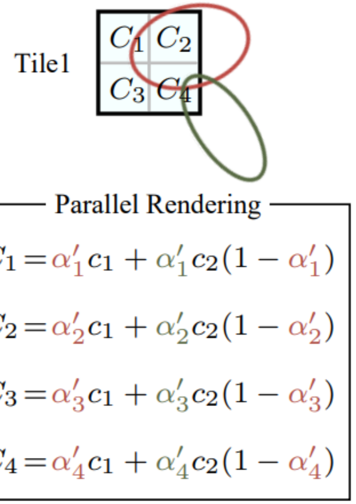
3D Gaussian Splatting



Tile-based rendering

The rendering image are divided into several tiles, and all tiles can be rendered in parallel.





Loss Function

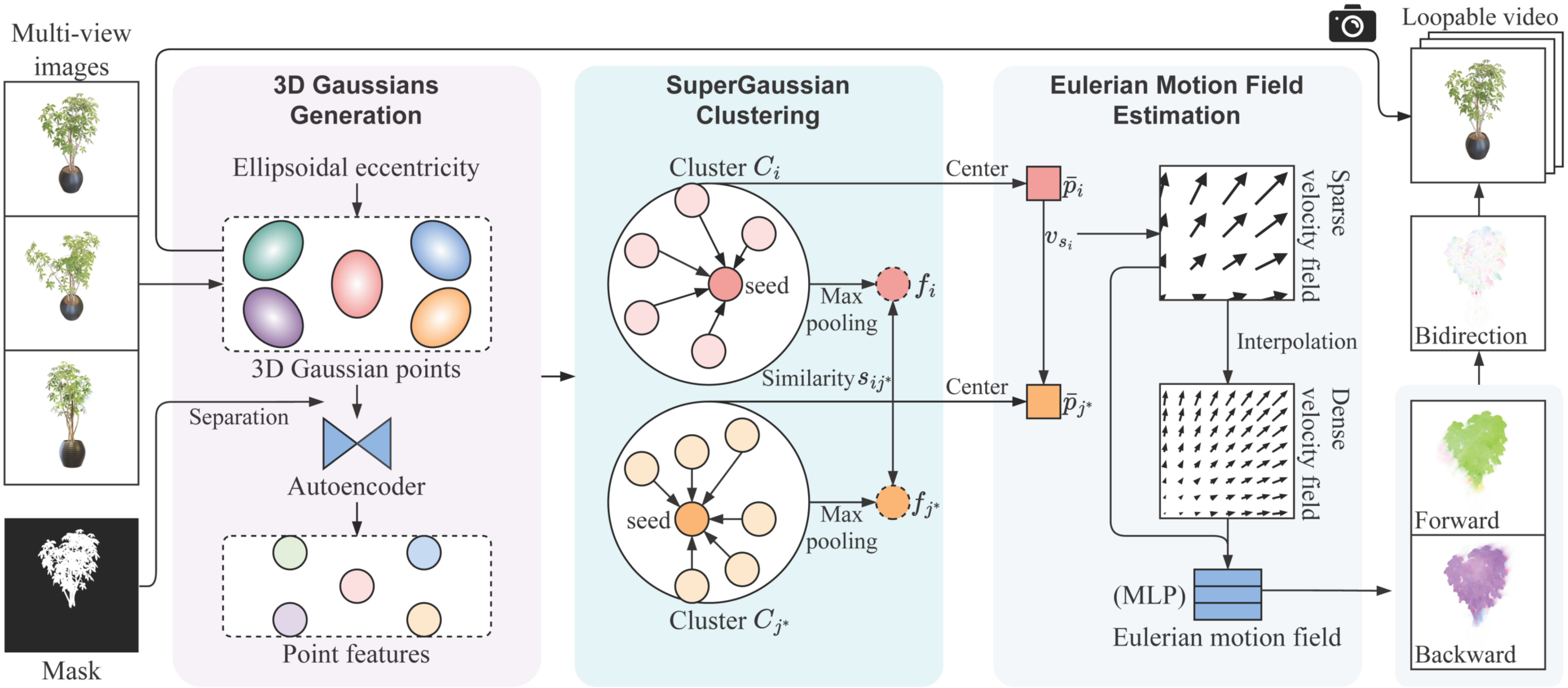
The error between the rendering image and the corresponding ground truth image is used as the loss function for training.

$$\mathcal{L} = (1 - \lambda)\mathcal{L}_1 + \lambda \mathcal{L}_{\text{D-SSIM}}$$

[1] 3D Gaussian Splatting for Real-Time Radiance Field Rendering.
[2] A Survey on 3D Gaussian Splatting.

LoopGaussian

Pipeine

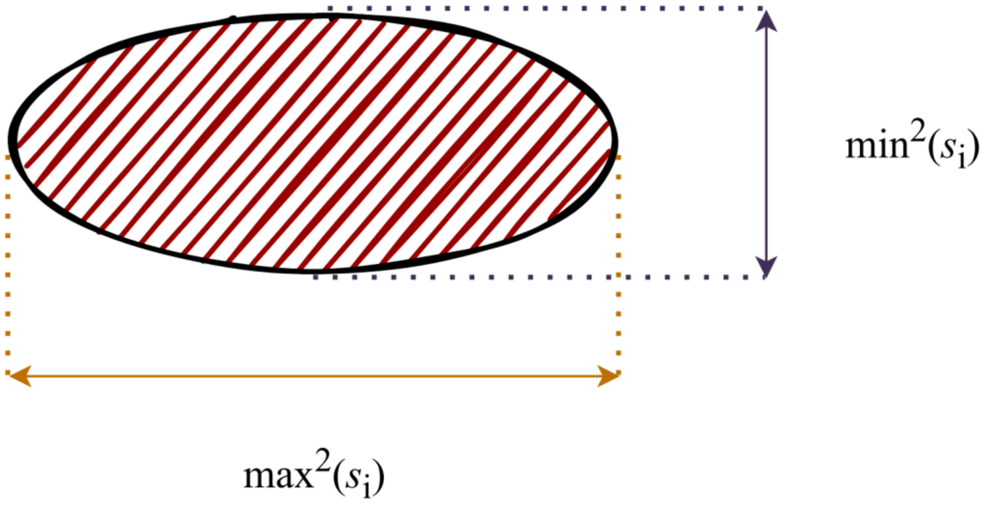


LoopGaussian

Artifact-free Scene Representation

Make Gaussian points not too sharp, as close to sphere as possible to avoid glitches when the scene deforms.

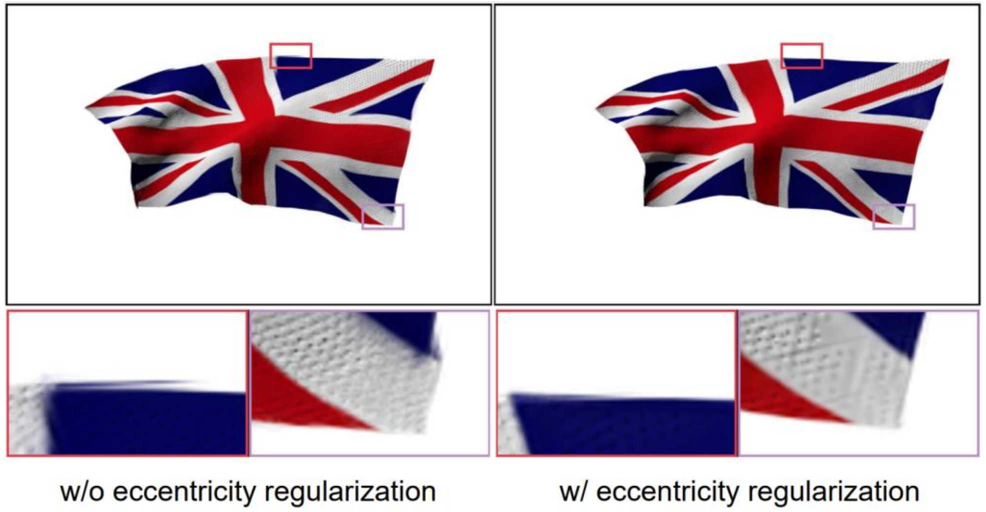
Eccentricity regularization



$$\mathcal{L}_{\text{shape}} = \frac{1}{|G|} \sum_{G_i \in G} 1 - \frac{\min^2(s_i)}{\max^2(s_i)}$$

Total Loss Function

$$\mathcal{L}_{\text{3D-GS}} = \eta\left((1-\beta)\,\mathcal{L}_1 + \beta\mathcal{L}_{\text{D-SSIM}}\right) + (1-\eta)\,\mathcal{L}_{\text{shape}}$$



Motivation

Similar objects always have similar movement trends.

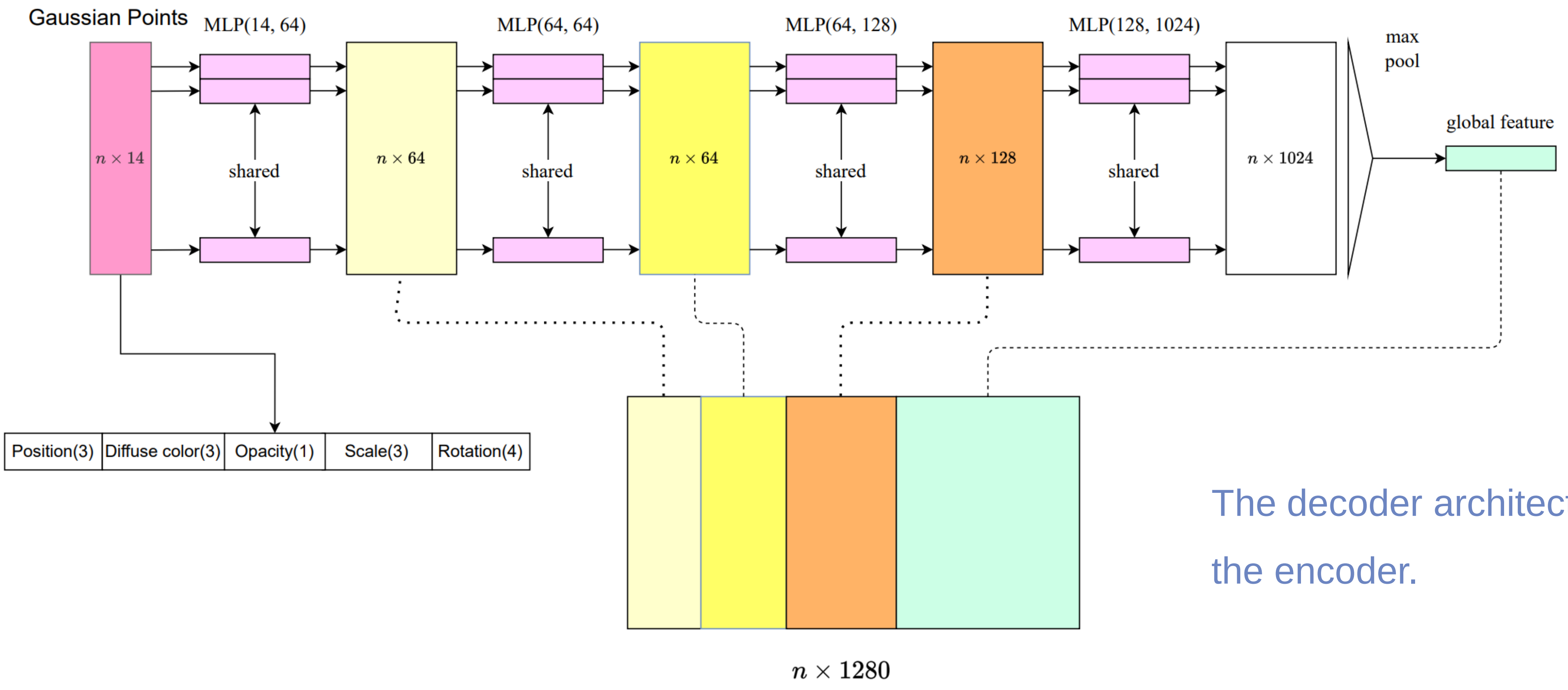
How to find similar objects in scene?

SuperGaussian Clustering

How to describe the motion of object?

Eulerian perspective v.s. Lagrangian perspective

SuperGaussian Autoencoder Architecture

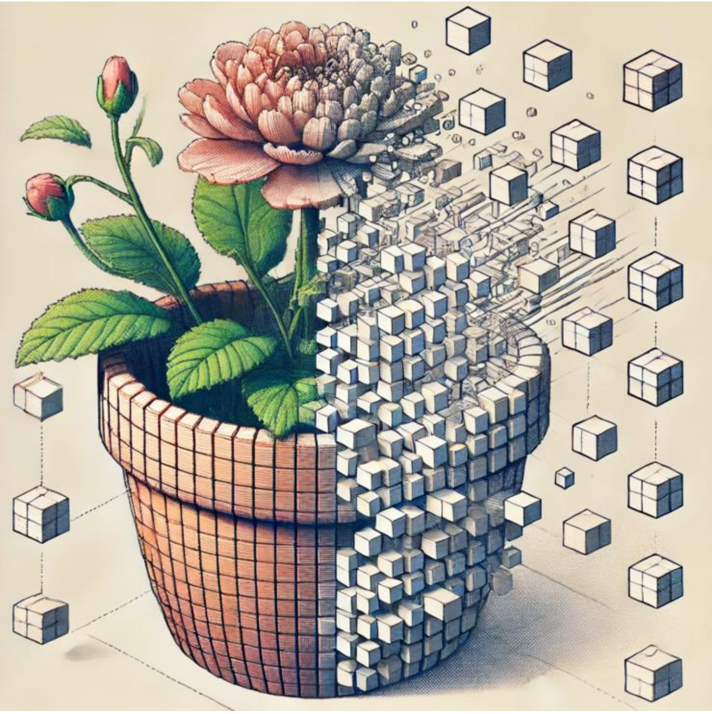


Endoer Architecture

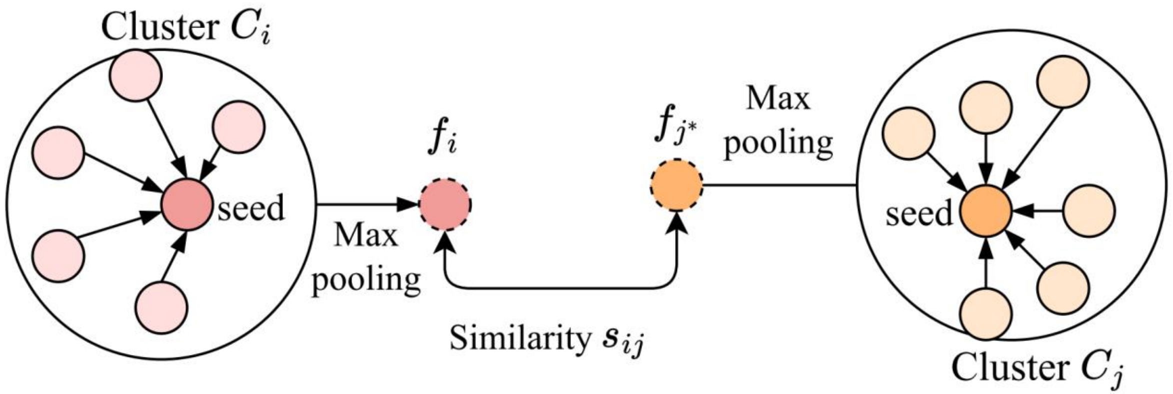
The decoder architecture is symmetric to the encoder.

SuperGaussian Clustering

Voxelization



Clustering



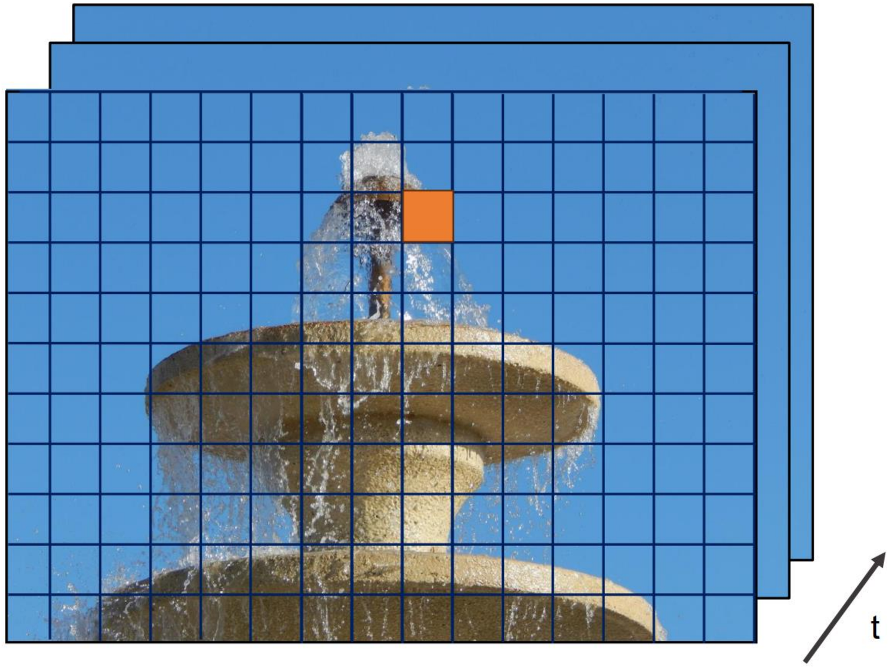
$$SG^* = \arg\min_{SG} \sum_{k=1}^{K} \sum_{SG(G_i)=k} D(G_i, G_{k'})$$

$$D(G_i, G_j) = 1 - \frac{|f_i \cdot f_j|}{\|f_i\| \cdot \|f_j\|} + \mu \frac{\|p_i - p_j\|}{R}$$

Eulerian perspective v.s. Lagrangian perspective

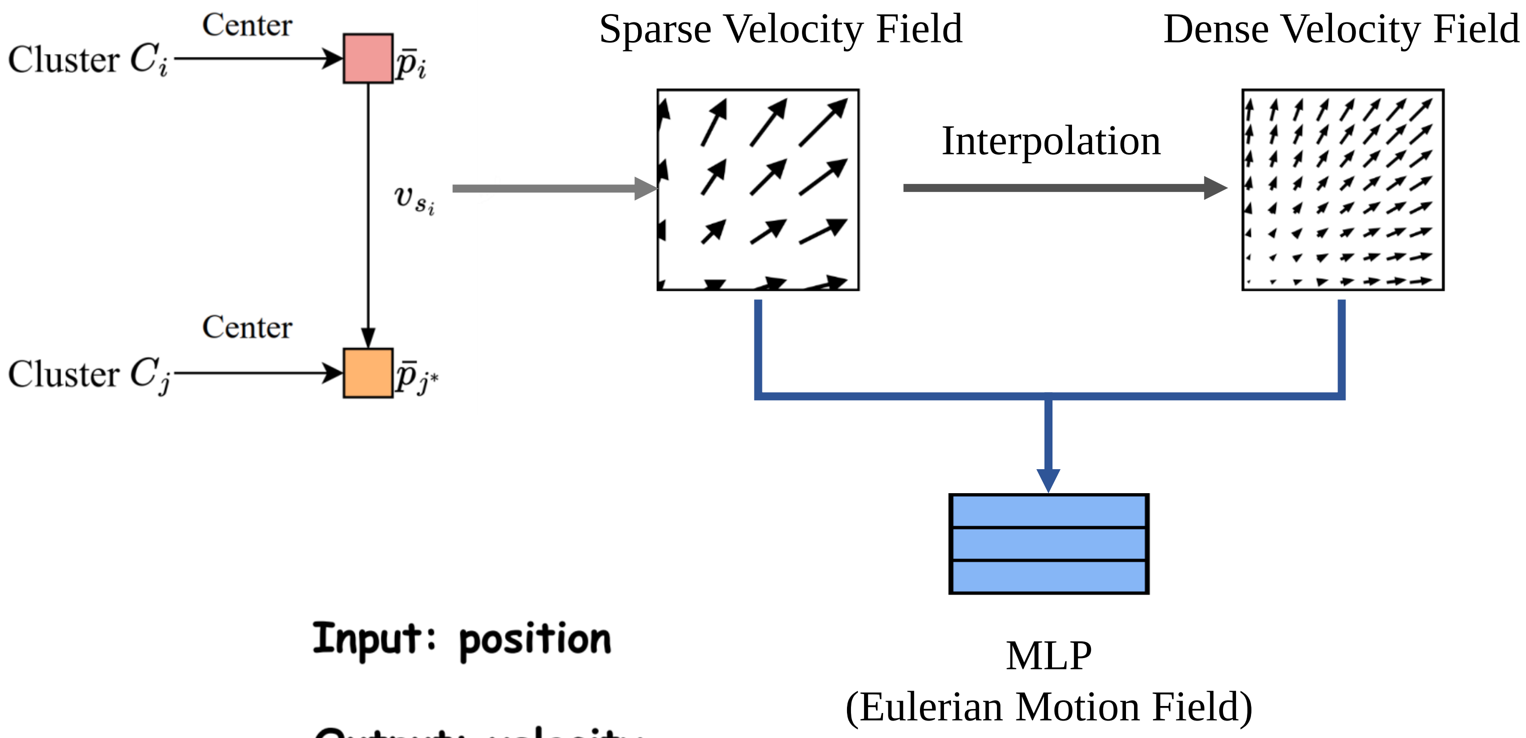
Lagrangian perspective

Eulerian perspective



- Lagrangian perspective describes the motion of the particle itself.
- Eulerian perspective describes the motion occurring at a fixed point in space.

Progressive Eulerian Motion Field Estimation



Input: position

Output: velocity

- Sparse Velocity Field Estimation

  Moving each cluster to its nearest neighbor

- Dense Velocity Field Estimation

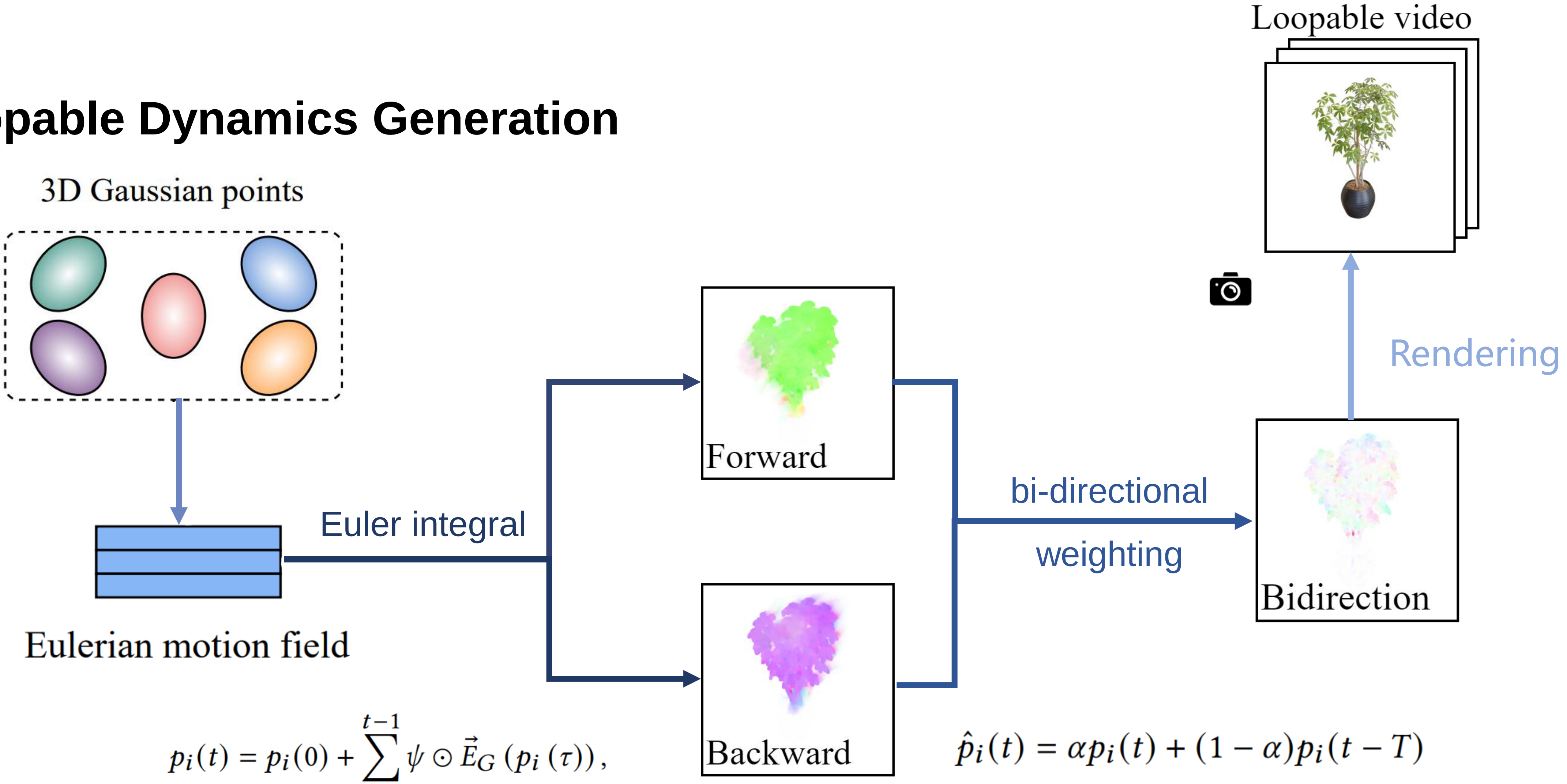
  Using Kriging interpolation

- Eulerian Motion Field Estimation

  Fitting the velocity field with an MLP

LoopGaussian

Loopable Dynamics Generation



$$p_i(t) = p_i(0) + \sum_{\tau=0}^{t-1} \psi \odot \vec{E}_G\left(p_i(\tau)\right),$$

$$\text{where} \quad p_i(\tau) = p_i(\tau - 1) + \psi \odot \vec{E}_G\left(p_i(\tau - 1)\right).$$

$$\hat{p}_i(t) = \alpha p_i(t) + (1 - \alpha) p_i(t - T)$$

$$\alpha = 1 - \frac{t}{T}$$

Experimental Results

Comparisons with 3D Cinamagraph[1]

Table 1: Comparison results of average optical flow maps.

	PSNR↑	SSIM↑	LPIPS ↓
Li [20]	22.959	0.915	0.233
Ours	**24.868**	**0.928**	**0.208**

Table 2: Comparison results of generated videos.

	FVD ↓
Li [20]	1174.948
Ours	**933.824**

[1] 3D Cinemagraphy from a Single Image. CVPR2023.

Experimental Results

The 3D cinemagraph obtained by our method can be rendered from any viewpoint.

Novel View Synthesis

Experimental Results

Ablation Study

- Comparison of different interpolation methods

  Kriging maintains the integrity of the object and motion continuity is better..

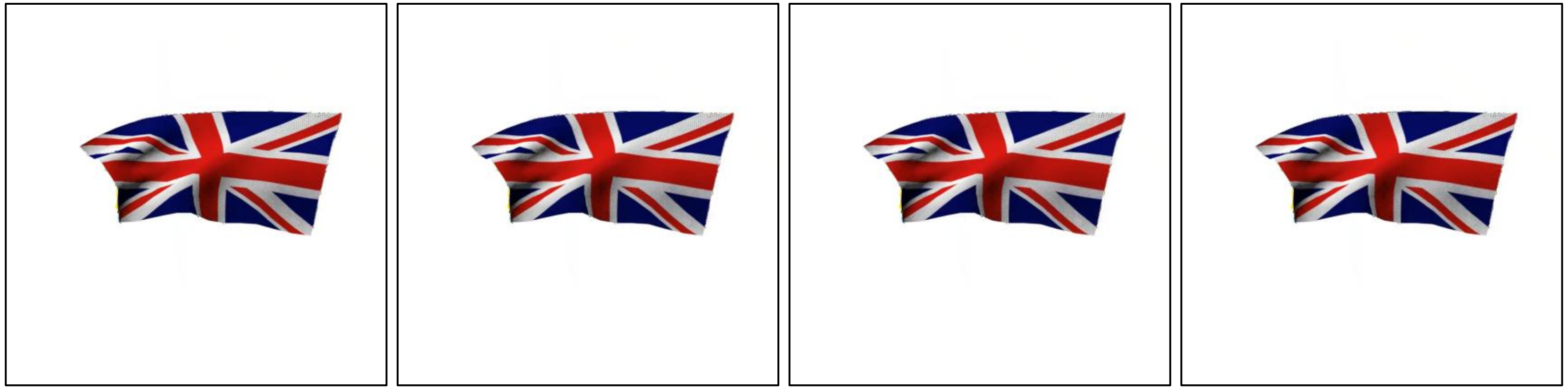
Interpolation Methods

w/o RBF Kriging

- Effect of the motion amplitude

  Higher ω results in more intense scene movement.

Motion Amplitude

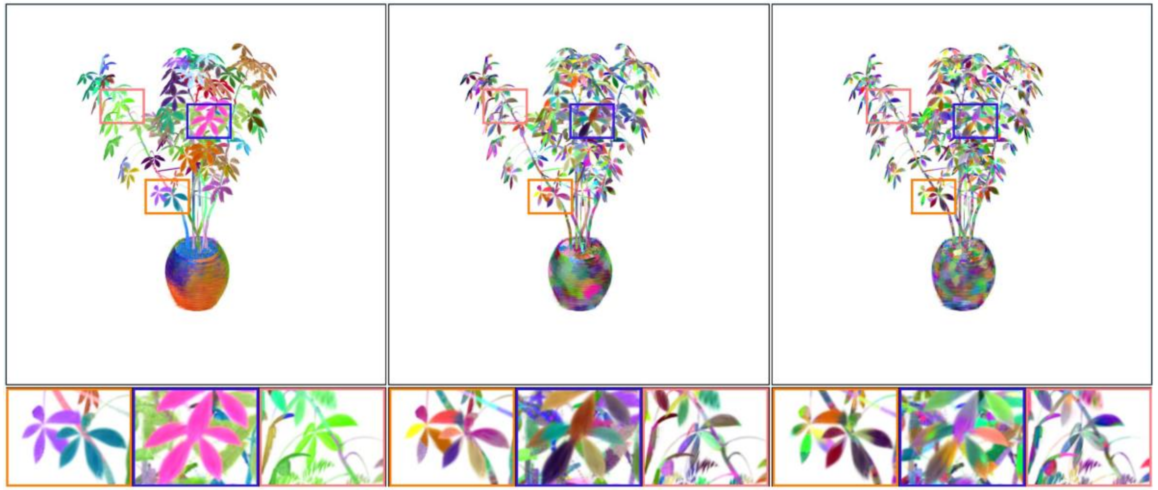


$\omega = 1.2$ $\omega = 2.0$ $\omega = 3.0$ $\omega = 4.0$

Experimental Results

Ablation Study

Voxel Resolutions



(a) $\lambda = 0.1$ (b) $\lambda = 0.04$ (c) $\lambda = 0.02$

- Voxel Resolution Selection

  Empirically chosen λ = 0.04.

  Balances scene segmentation with information preservation.

Conclusion

- We introduce LoopGaussian, a novel framework for generating authentic 3D cinemagraphs from multi-view images of static scenes.
- No extensive pre-training on large dataset required.
- Outperforms previous methods limited to 2D image space by reconstructing the 3D geometry of the scene, and experiments demonstrate the effectiveness of our method.

Limitations

- Primarily designed for single objects and faces challenges with large-scale scenarios.
- Restricted to soft non-rigid bodies like flags and tree branches.

THANK YOU FOR WATCHING

Oct. 2024