PhysSplat ⚛: Efficient Physics Simulation for 3D Scenes via MLLM-Guided Gaussian Splatting

PhysSplat: Efficient Physics Simulation for 3D Scenes via MLLM-Guided Gaussian Splatting

Supplementary Material

1. 3D Open-vocabulary Segmentation

2D Open-vocabulary Segmentation. Inspired by prior successful works, we innovatively introduce the integration of 2D open-vocabulary detector models, such as Grounding DINO, promptable 2D segmentation models, such as SAM, image tagging models like RAM. The integrated 2D open-vocabulary model can automatically segment objects within images without the need for any textual input.

Specifically, given an input image, we first employ an image tagging model, RAM to get the tags of the image. Then, given the tags, we employ Grounding DINO to generate precise boxes for objects or regions within the image by leveraging the textual information in tags as condition. Subsequently, the annotated boxes obtained through Grounding DINO serve as the box prompts for SAM to generate precise mask annotations. By leveraging the capabilities of these robust expert models, our method enables the automatic labeling of an entire image.

3D Open-vocabulary Segmentation. After 2D open-vocabulary segmentation, the segmented images contain rich semantic features for every object in the 3D scene. We effectively lift these 2D masks to segment anything in the 3D scene via radiance fields rendering. Given a pre-trained 3D scene, inspired by recent works, we preserve all attributes of the Gaussians, but add a semantic attribute to integrate semantic information for each Gaussian. Then, to assign each 2D mask a unique ID in the 3D scene, we need to associate the masks of the same identity across different views. We employ a well-trained zero-shot tracker to propagate and associate these masks.

In addition to the existing Gaussian properties, we introduce a new parameter, semantic attribute, to each Gaussian. The semantic attribute is a learnable and compact vector, which is used to distinguish semantic categories in whole 3D scene. To optimize the introduced attribute of each Gaussian, we render semantic attribute into 2D images in a differentiable manner as:

$$\mathcal{S} = \sum_{p \in \mathcal{N}} y_p \alpha_p \prod_{j=1}^{p-1} (1 - \alpha_j), \tag{1}$$

where $\mathcal{S}_k$ represents the 2D semantic labels of pixel k, derived from Gaussian point semantic attributes via α-blending. Here, y_p denotes the semantic attribute of the 3D Gaussian point p, and α_p is the influence factor of this point in rendering pixels. After associating 2D instance labels across each training view, we apply the grouping loss and 3D Gaussian reconstruction loss to supervise the optimization progress.

Extracting objects from 3DGS introduces holes, which we inpaint using LaMa. This inpainting ensures more natural results when objects undergo displacement due to external forces. The whole pipeline of 3D open-vocabulary segmentation is shown in Fig. 1.

2. Implementation Details for Baselines

In this section, we elaborate on the implementation details of baselines used for comparison to our proposed method. For PhysDreamer, we used the pre-trained models provided in the official code repository[1], as the training code is not made available. For Physics3D, we train the models using the code from official code repository[2]. For DreamPhysics, we train the models using the code from official code repository[3]. All other hyperparameters remain unchanged. The trained models are then used for qualitative evaluation.

3. User Study

We use Tencent Survey[4] to recruit participants for the human preference evaluation. The survey is fully anonymized. For each scenario, we provided video clips and asked the participants to give each video a score. A total of 41 volunteers participated in the study, including 3 professionals from the 3D art industry.

4. Video Visualization

We provide generated videos in the project page[5] for a better motion visualization. We also show the simulated interactive motion in Fig. 2.

5. Explanation of RS/AS scores

We assess each video frame's artistic value using the LAION aesthetic predictor. The final aesthetic score (AS) is the average of all frame scores, reflecting layout, color harmony, photo-realism, naturalness, and artistic quality. As

[1] https://github.com/a1600012888/PhysDreamer
[2] https://github.com/liuff19/Physics3D
[3] https://github.com/tyhuang0428/DreamPhysics
[4] https://wj.qq.com/index.html
[5] https://sim-gs.github.io/

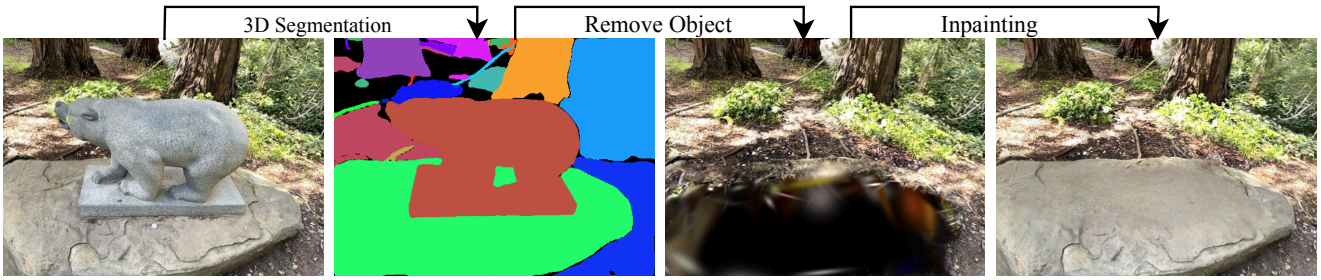

Figure 1. The whole pipeline for 3D Open-vocabulary Segmentation.

mentioned in our Supp.Mat, we provided video clips and asked the participants to give each video a score (RS) by anonymous questionnaire. A total of 41 volunteers participated in the study.

6. More Analysis about Material Property Distribution Prediction

In our paper, we train an MPDP model using part of the data from Physics3D. However, with the advancement of 3D content creation networks, such as LGM, we can generate diverse objects through these methods and utilize Physics3D for m material property distribution prediction to create additional training data. This approach has the potential to further enhance the performance of our model and represents a direction for our future work.

7. More Details about Material Point Method (MPM)

The Material Point Method (MPM) is an advanced numerical technique for simulating the behavior of continuum materials. It discretizes a material body into material points, commonly referred to as particles, which carry essential properties such as mass, velocity, deformation gradient, and stress. These particles interact with a background computational grid, which facilitates spatial derivative calculations and the application of external forces.

MPM consists of two primary phases: 1) Particle-to-Grid (P2G) Transfer: Particles transfer their properties to the grid, enabling the computation of global quantities such as forces and accelerations. 2) Grid-to-Particle (G2P) Transfer: Updated grid values, such as velocities and positions, are mapped back to the particles, ensuring their motion aligns with the computed dynamics.

This dual transfer mechanism allows MPM to efficiently handle large deformations and complex interactions in continuum materials.

Particle-to-Grid (P2G) Transfer. During this phase, the particles' properties, such as mass and momentum, are mapped to the computational grid using interpolation functions. The mass at a grid node i is computed as:

$$m_i^n = \sum_p w_{ip}^n m_p,$$

where m_p is the mass of particle p, and w_{ip}^n is the interpolation weight (often derived from a B-spline kernel) between particle p and grid node i. The momentum at the grid node is similarly updated:

$$m_i^n \mathbf{v}_i^n = \sum_p w_{ip}^n m_p \left(\mathbf{v}_p^n + \mathbf{C}_p^n(\mathbf{x}_i - \mathbf{x}_p^n)\right),$$

where $\mathbf{v}_p^n$ is the velocity of particle p, $\mathbf{C}_p^n$ represents the affine velocity field gradient, and $\mathbf{x}_i$ and $\mathbf{x}_p^n$ are the positions of the grid node and particle, respectively.

Grid Update. Once particle properties are transferred, grid velocities are updated by accounting for external forces, internal stresses, and gravity. The velocity at grid node i is computed as:

$$\mathbf{v}_i^{n+1} = \mathbf{v}_i^n - \frac{\Delta t}{m_i^n} \sum_p \boldsymbol{\tau}_p^n \nabla w_{ip}^n V_p^0 + \Delta t \mathbf{g},$$

where Δt is the time step, $\boldsymbol{\tau}_p^n$ is the stress tensor of the particle p, V_p^0 is the initial volume of the particle, and $\mathbf{g}$ is the acceleration due to gravity.

Grid-to-Particle (G2P) Transfer. After the grid is updated, the changes in velocity and momentum are transferred back to the particles. The particle velocity is updated using the grid velocities and interpolation weights:

$$\mathbf{v}_p^{n+1} = \sum_i \mathbf{v}_i^{n+1} w_{ip}^n,$$

and the new position of the particle is given by:

$$\mathbf{x}_p^{n+1} = \mathbf{x}_p^n + \Delta t \mathbf{v}_p^{n+1}.$$

Additionally, the affine velocity field gradient $\mathbf{C}_p^{n+1}$ and deformation gradient $\mathbf{F}_p^{n+1}$ are updated as:

$$\mathbf{C}_p^{n+1} = \frac{4}{(\Delta x)^2} \sum_i w_{ip}^n \mathbf{v}_i^{n+1} (\mathbf{x}_i - \mathbf{x}_p^n)^T,$$

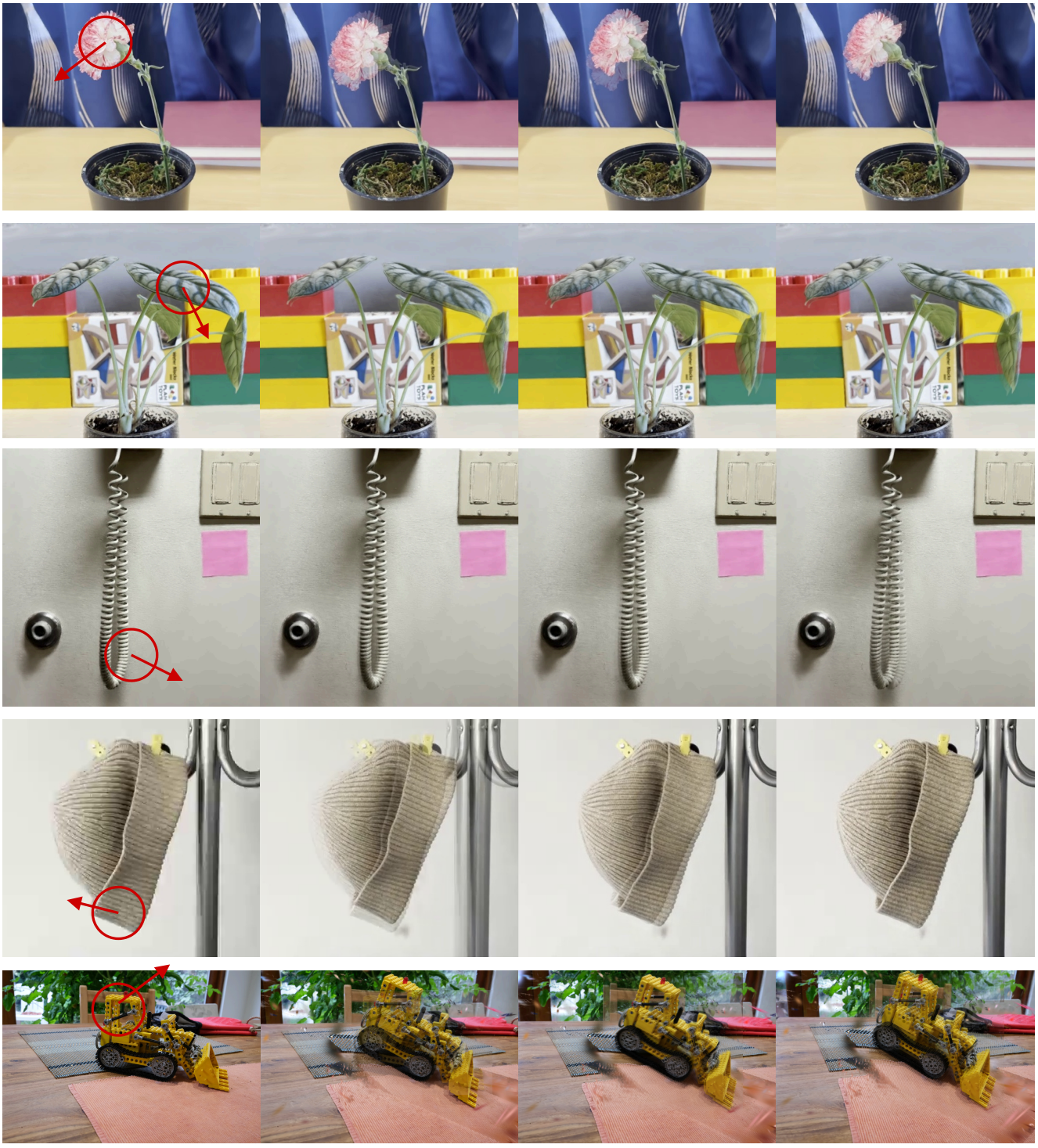

Figure 2. **More visual results** of our method.

$$\mathbf{F}_p^{n+1} = (\mathbf{I} + \Delta t \mathbf{C}_p^{n+1})\mathbf{F}_p^n.$$

The Material Point Method effectively combines Lagrangian (particle-based) and Eulerian (grid-based) approaches, making it highly suitable for simulating materials that experience large deformations, fractures, and complex interactions.

8. Analysis of Failure Cases

As noted in our limitation, segmentation failure can be a bottleneck, especially in complex environments with occluded objects. Sim Anything may struggle to segment the entire object, resulting in unnatural simulations. We will elaborate in the final paper.

9. Selection of Material Properties

Young's modulus and Poisson's ratio are essential for understanding an object's motion under forces. Young's modulus indicates stiffness, while Poisson's ratio relates lateral strain to axial strain. Predicting additional physical properties for 3D objects is a promising direction for future research.

10. Material Properties in PGAS

PGAS adjusts the sample radius based on the object's Young's modulus and curvature (Sec. 4.2 in the main paper). Young's modulus reflects material stiffness, while curvature accounts for geometric complexity, allowing for effective management of material and shape variations. Although incorporating more material properties could improve granularity, Young's modulus and curvature are sufficient for accurately modeling soft, complex-shaped objects in our experiments.

11. Ethical Statement

We confirm that all data used in this study were obtained and utilized in compliance with ethical standards. All participants provided consent, or the data were sourced from publicly available datasets with proper permissions. The use and publication of these data and models pose no societal or ethical harm. Necessary precautions were taken to respect individual rights, including privacy and ethical research principles.