

Semantic Aware Physics-Based Object Interaction In 3D Gaussian Splatting

Pooja Guhan
Dolby Laboratories Inc
Atlanta, USA
pooja.guhan@dolby.com

Dae Yeol Lee
Dolby Laboratories Inc
Sunnyvale, USA
daeyeol.lee@dolby.com

Guan-Ming Su
Dolby Laboratories Inc
Sunnyvale, USA
guanmingsu@ieee.org

Abstract—3D Gaussian Splatting (3DGS) has emerged as a powerful representation for photorealistic 3D scenes, but its point-wise representation poses challenges for enabling physics-based interaction among 3D objects, particularly in how to group Gaussians into coherent objects and how to control their motion during interactions. In this paper, we propose a novel framework for dynamic physics simulation in 3DGS scenes that enables efficient and physically plausible interactions among objects. We first employ semantic-aware Gaussian grouping to augment each primitive with semantic attributes distilled using a combination of SAM and CLIP features, enabling automatic identification and manipulation of object-level Gaussian clusters. We then introduce a center-of-mass dynamics module to enable object interactions and deformations that preserve the computational tractability by maintaining real-time performance constraints. We anticipate our work to serve as a foundation for bridging high-quality visual representation with physics-based simulation in interactive 3D content creation.

Index Terms—physics modeling, Gaussian splatting, interaction

I. INTRODUCTION

3D volumetric reconstruction technologies have gained immense interest due to their remarkable visual realism. By representing scenes as collections of 3D Gaussian primitives with learnable parameters, 3DGS [1] has achieved superior quality and efficiency compared to traditional neural radiance fields and mesh-based representations. Despite these advances in photorealistic rendering, current 3DGS methods remain fundamentally limited to passive visualization. They cannot simulate how objects physically interact when forces are applied or when they collide with each other. This inability to model physical interactions severely restricts the deployment of 3DGS in interactive applications where realistic object behavior is essential. Such applications include surgical training simulations, robotic manipulation and planning, physics-based gaming experiences, interactive digital twins for industrial scenarios, and immersive virtual reality environments where users expect natural object interactions.

While there has been growing interest in advancing beyond passive visualization towards physically plausible representations, existing approaches face critical limitations when handling multi-object interactions with heterogeneous materials.

The core technical challenge lies in the point-wise, unstructured nature of 3DGS representations. Unlike mesh-based

methods where object boundaries and surfaces are explicitly defined, 3DGS consists of millions of separate Gaussian primitives without inherent notions of object identity or continuous surfaces. Specifically, enabling physics-based multi-object interactions requires an accurate understanding of object surfaces and their deformation behavior under contact forces. Identifying which Gaussians belong to each object and efficiently computing their deformation during interactions are fundamental hurdles to producing plausible surface motion. To address this, we introduce a novel framework that brings physics-aware behavior to 3DGS while remaining computationally efficient. Our approach includes:

- **Semantic understanding:** We incorporate to each Gaussian a semantic attribute distilled from a 2D foundation model, enabling selection and interaction with object-level Gaussian groups.
- **Efficient dynamics modeling:** We introduce a center-of-mass (COM)-based momentum conservation formulation to simplify interaction analysis and efficiently estimate the velocity of each Gaussian particle during object interactions.

By integrating physics simulation with high-fidelity 3D scene representations, our approach opens transformative opportunities for interactive 3D content creation and computational simulation across a wide range of applications while maintaining the inherent advantages of the representation for real-time rendering and interactive scene manipulation.

II. RELATED WORK

A. Physics Applications in Gaussian Splatting

Physics-based modeling has become increasingly important in neural rendering to achieve realistic visual effects and natural object behaviors. Early works in physics-enhanced Gaussian splatting primarily focused on optical phenomena. GausSurf [2] introduced surface aware Gaussian representations that better capture geometric details, while relightable 3DGS methods [3], [4] incorporated physically-based rendering principles to model light transport and material reflectance properties under varying illumination conditions. These approaches significantly improved photorealistic rendering quality but remained limited to static scenes without dynamic material interactions. The integration of mechanical physics

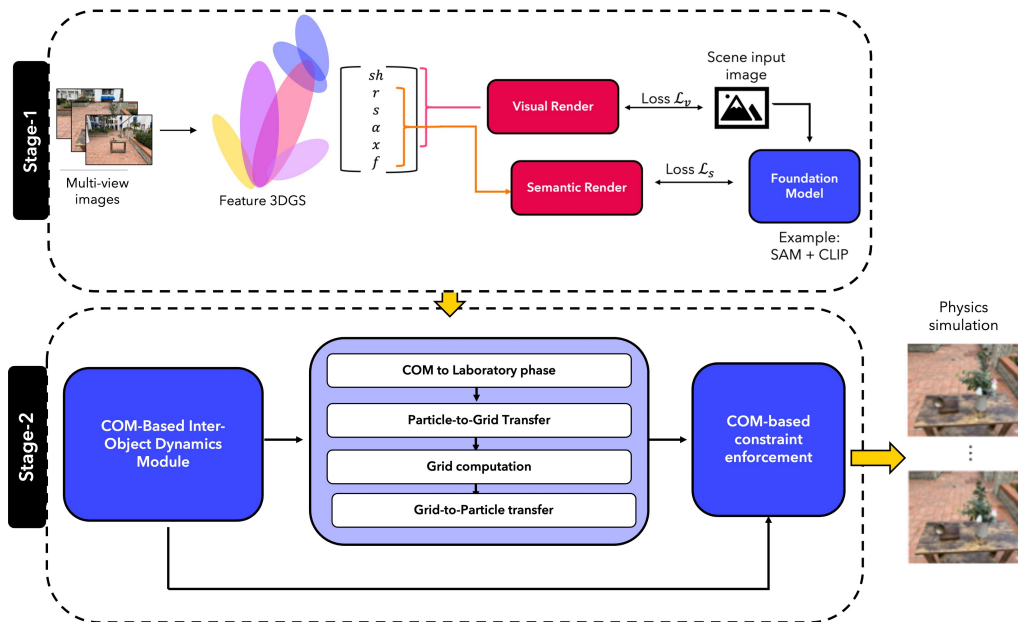


Fig. 1. Overview: In stage-1, multi-view images are processed to learn 3D Gaussian primitives augmented with semantic features using a combination of SAM, CLIP and DINOv2. Visual (\mathcal{L}_v) and semantic (\mathcal{L}_s) loss functions jointly optimize the Gaussian parameters with an additional semantic feature attribute. In Stage 2, the semantically grouped Gaussians undergo COM-based computation of inter-object dynamics followed by particle level local deformations. COM-based constraint enforcement ensures particles maintain object coherence while allowing realistic deformations. The algorithm then outputs physically plausible Gaussian trajectories for rendering the object interactions

into Gaussian representation marked a significant advancement with PhysGaussian [5], which embedded physical properties such as mass, electricity, and damping directly into Gaussian primitives. This enabled simulation of deformable objects without requiring explicit mesh reconstruction. Building upon this foundation, GASP [6] introduced more sophisticated dynamics modeling with improved temporal consistency while TRACE [7] extended the framework to support multi-view learning of physical motion patterns. Despite these advances, current methods face limitations in handling complex multi-object interactions. Most existing frameworks are designed for single-object scenarios, leaving substantial gaps in modeling realistic physical phenomenon involving multiple interacting objects.

B. Material Point Method

The material point method (MPM), originally proposed by Sulsky et al [8], has emerged as a robust computational framework for simulating a wide range of multiphysics phenomena involving solids, fluids, and granular materials. Its hybrid particle-grid formulation enables natural handling of large deformations, contact, and fracture without requiring explicit topological updates. This inherent capability allows MPM to seamlessly model topology changes, frictional and cohesive interactions, and interface evolution across diverse material systems such as elastoplastics, granular flows, foams, snow, and fluids. Moreover, extensions of MPM have been developed to represent co-dimensional materials, such as thin shells, rods, and membranes, within the same unified formulation, facilitating coupled simulations between volumetric and

lower-dimensional structures. These advancements demonstrate MPM's versatility as a general-purpose continuum solver for complex, coupled physical systems exhibiting nonlinear and discontinuous behaviors.

C. Object Interaction

Modeling realistic interactions between multiple objects with different material properties presents significant computational and algorithmic challenges. Traditional approaches often rely on explicit contact detection and resolution algorithms, which can become computationally expensive and numerically unstable for complex scenarios involving multiple contact points and varying material stiffnesses. In the context of neural rendering, several methods have attempted to address multi-body interactions. Neural rigid body dynamics learned to predict rigid body interactions from visual observation while Visual Interaction networks modeled object relationships in simplified scenarios. However, these approaches typically assume rigid body dynamics and cannot handle deformable materials or complex contact mechanics. Recent efforts such as PAC-NeRF [9] and D-NeRF [10] have shown promising results in modeling dynamic scenes, but they lack the physical accuracy required for realistic material interactions. Our work addresses these limitations by combining the representational power of 3D Gaussian splatting with the robust physics simulation capabilities of MPM to enable reliable modeling of multi-object interactions.

III. PROPOSED APPROACH

Our overarching goal is to enable dynamic movement and realistic interaction between objects within 3D Gaussian

splatting (3DGS) scenes, thereby expanding the versatility and practical applications of Gaussian splatting. To achieve this, our method consists of two stages. The first stage involves augment each Gaussian primitive with semantic features that encode its association with specific scene objects. In the second stage, we develop a novel module that incorporates the physical material properties of interacting objects and simulate their motion dynamics through a center-of-mass (COM) based computational model.

A. Stage-1: Object Selection Using Semantic Understanding

Standard Gaussian primitives include color(sh), orientation r , scale s , alpha α and position x . Since scenes typically contain multiple objects with distinct shapes, sizes and material properties, we require a mechanism to identify and group Gaussians belonging to each of the target interacting object(s). We therefore, introduce each Gaussian primitive with an additional semantic feature attribute f that encodes object-level semantic information. These features are learned during scene reconstruction from multi-view images using a two branch network. The first branch focuses on visual rendering, ensuring that the geometry and appearance of the reconstructed 3D model precisely aligns with the input views. The second branch focuses on enriching the feature primitive f of the 3D Gaussian. This is done by aggregating over CLIP [11] and DINO [12] features over segmentation masks generated by SAM [13], following a procedure similar to Feature Splatting [14]. The overall loss function for Stage-1 combines visual and semantic objectives:

$$\mathcal{L}_{stage1} = \lambda_1 \mathcal{L}_v + \lambda_2 \mathcal{L}_s \quad (1)$$

where \mathcal{L}_v , denotes the visual loss, implemented as a perceptual loss applied between the rendered image and the ground truth image and \mathcal{L}_s denotes the semantic loss, implemented as a cosine loss applied between the semantic feature estimated and SAM-enhanced clip feature. λ_1 and λ_2 are user-defined differential weighing parameters. The 3DGS representation obtained from this procedure, which we refer to as Feature3DGS, enables our algorithm to distinguish between different objects within the scene and associate material properties with semantically coherent Gaussian clusters. For object k , we define the set of its constituent Gaussians as $\mathcal{G}_k = \{g_i : semantic_id(g_i) = k\}$. Each Gaussian primitive g_i is assigned an object ID k based on the similarity of the semantic feature f with the rasterized CLIP features.

B. Stage-2: COM-Based Physics Modeling

At the end of stage 1, we have access to all Gaussians specific to the different interacting objects, each enriched with semantic information. Now, traditional MPM treats the simulation domain as a unified continuum where particles represent material points within the same physical body, making it inherently challenging to distinguish between separate objects and their boundaries. The method's grid-based transfer mechanism, while effective for internal material interactions, lack explicit

collision detection and response algorithm necessary for accurate inter-object contact handling. When multiple semantically distinct objects need to be manipulated independently, the standard MPM's unified continuum approach may not provide the object-level control desired for certain interactive applications. Therefore, to enable realistic interactions between these objects within 3DGS, we propose an approach that operates at two complementary scales: object level(course) dynamics for inter-object interactions and particle-level (fine) deformation for internal material behavior.

1) *Object-level dynamics*: In our approach, objects are treated as distinct entities with center-of-mass (COM) and material-specific characteristics computed from their constituent Gaussian primitives. The *COM* is a unique point in a system of objects that represents the average positions of the distributed mass. For the purpose of calculating the motion of a system, it behaves as if the entire mass of the system is concentrated at this point. In the COM frame of reference, the observer moves with the COM of the system, making the COM stationary relative to the observer. Additionally, the momentum of the system is always zero, simplifying the analysis of the interaction. Taking this into consideration, for each semantic object o_k , we compute center of mass (COM) properties to enable efficient object-level physics:

$$x_{COM}^{o_k} = \frac{\sum_{i \in \mathcal{G}_k} m_i x_i}{\sum_{i \in \mathcal{G}_k} m_i}, \quad v_{COM}^{o_k} = \frac{\sum_{i \in \mathcal{G}_k} m_i v_i}{\sum_{i \in \mathcal{G}_k} m_i} \quad (2)$$

where m_i , x_i , and v_i are the mass, position, and velocity of Gaussian $i \in \mathcal{G}_k$, respectively. The collision detection between the objects and its resolution is then performed in the COM frame of reference. Based on the spatial distribution of the object's constituent Gaussians, collision between two objects o_i and o_j is said to occur when

$$|x_{COM}^{o_i} - x_{COM}^{o_j}| \leq r_{o_i} + r_{o_j}, \quad (3)$$

where r_{o_i} and r_{o_j} denote the effective radii of objects o_i and o_j respectively, computed from the spatial distribution of their constituent Gaussians. Momentum conservation laws are used to detect collision between any pair of colliding objects o_i, o_j with initial velocities as $v_{initial}^{o_i}$ and $v_{initial}^{o_j}$ respectively as follows:

$$\begin{aligned} m_{o_i} v_{COM,initial}^{o_i} + m_{o_j} v_{COM,initial}^{o_j} \\ = m_{o_i} v_{COM,final}^{o_i} + m_{o_j} v_{COM,final}^{o_j} \end{aligned} \quad (4)$$

where $v_{COM,initial}^{o_k} = v_{initial}^{o_k} - v_{COM}^{o_k}$ and m_{o_k} is the mass of the object o_k where $k = i, j$. The type of collision and thereby, the energy dissipation that occurs between any two objects depends upon their materials and is determined by the coefficient of restitution e . The final velocities after collision become:

$$v_{final}^{o_i} = v_{COM}^{o_i} + \frac{m_{o_j}(1+e)}{m_{o_i} + m_{o_j}} v_{rel,initial} \quad (5)$$



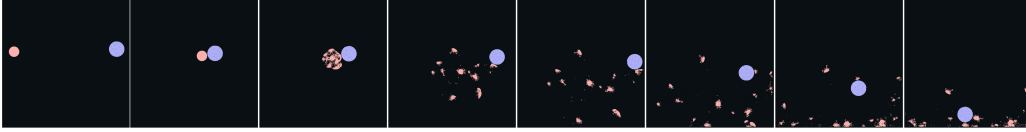
Fig. 2. Simulation of pushing a vase made of snow using our approach on a sticky table



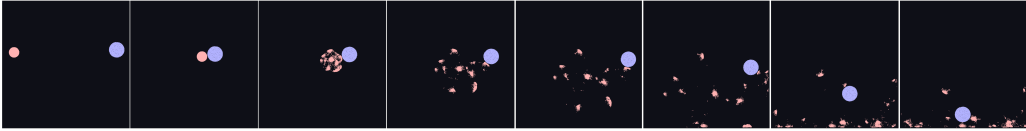
Fig. 3. Simulation of dropping a vase made of elastic material using our approach



Fig. 4. Two elastic vases made to collide against each other on a sticky table with different velocities



(a) Standard MPM



(b) Our approach

Fig. 5. Visual comparison of 2D simulation of interaction between a snow ball and a rubber ball generated using (a) Standard MPM and (b) our approach

$$v_{final}^{oj} = v_{COM}^{oj} - \frac{m_{oi}(1+e)}{m_{oi} + m_{oj}} v_{rel,initial} \quad (6)$$

where $v_{rel,initial} = (v_{initial}^{oj} - v_{initial}^{oi})$ is the relative velocity component along the collision normal.

2) *Modeling local deformations*: Object interactions inherently can lead to deformations in objects which needs to be modeled as well for accurate physics representation. In order to accomplish this, we invoke the use of MPM at the particle level since the deformations are local and not object level. This primarily consists of three key steps:

- 1) *Particle-to-grid (P2G) transfer*: This involves projecting the data from the material points or particles onto the grid for calculations where the governing equations are solved.
- 2) *Compute grid velocities and forces*: This involves computing the velocities and forces corresponding to a grid node. Grid nodes refer to fixed points on a grid used for computational purposes. The computation is based on well-defined physics equations available corresponding to different materials.
- 3) *Grid-to-particle (G2P) transfer*: This involves moving back from grid space (Eulerian) to particle space (La-

TABLE I

OPERATION COUNT BREAKDOWN FOR COMPUTATIONS DONE IN STANDARD MPM AND OUR APPROACH. α REFERS TO THE NUMBER OF PARTICLE-TO-GRID OPERATIONS PER PARTICLE, β REFERS TO THE NUMBER OF GRID OPERATIONS DONE PER GRID NODE, δ REFERS TO THE NUMBER OF GRID-TO-PARTICLE OPERATIONS PER PARTICLE, ϵ IS THE NUMBER OF CONSTRAINTS SET AND $C(N_1, N_2, r)$ IS THE CONTACT COMPLEXITY FUNCTION DEPENDING ON PARTICLE COUNTS AND GRID RESOLUTION.

Operation Type	Standard MPM	Our Approach
COM-Based Object Dynamics	-	$O(N_{objects}^2)$
Particle-to-Grid (P2G)	$O(\alpha(N_1 + N_2))$	$O(\alpha(N_1 + N_2))$
Grid-Operations	$O(\beta r^2)$	$O(\beta r^2)$
Grid-to-Particle (G2P)	$O(\delta(N_1 + N_2))$	$O(\delta(N_1 + N_2))$
Contact Resolution	$O(C_1(N_1, N_2, r))$	-
COM-Based constraint enforcement	-	$O(\epsilon(N_1 + N_2))$

grangian), i.e., we compute the new particle velocities and displacement.

To ensure there is consistency between object level and particle level dynamics, we need a mechanism that maintains object coherence while allowing appropriate internal deformation. This will prevent objects from fragmenting unrealistically while preserving the material behavior computed while modeling the local deformations.

IV. EXPERIMENTS AND RESULTS

Our algorithm facilitates the efficient physics-aware simulation of the behavior and motion of objects interacting in 3DGS. In order to test our approach, we consider the dataset from feature splatting [14]. We also generated a few additional scenarios by editing the scene in the dataset to test for different kinds of physical behaviors like dropping and pushing. These include changing materials, positions and also cloning object(s) for multi-object interactions. On these scenarios we compare and contrast the standard MPM with our proposed approach.

A. Standard MPM vs COM-Based MPM

We assume the test configuration to include objects \mathcal{S}_1 and \mathcal{S}_2 represented using N_1 and N_2 Gaussians respectively. We assume the grid resolution to be rxr nodes (i.e. r^2 total). Initial velocities for \mathcal{S}_1 and \mathcal{S}_2 to be v_1 and v_2 respectively. Time step is denoted as Δt . We define N_{object} to be equal to the number of interacting objects (For this configuration $N_{object} = 2$). We compare the computational cost of standard MPM against our approach by counting floating point operations (FLOPs) per simulation step. The analysis includes particle-to-grid transfer operations, grid velocity and force computations, contact detection and resolution, grid-to-particle interpolation and reference frame transformation (only for our method) overhead. Table I shows an analytical comparison between the two methods. Overall, the complexity of Standard MPM will be $O((\alpha + \delta)(N_1 + N_2) + \beta r^2 + C(N_1, N_2, r))$ and our approach on the other hand shows a complexity of $O(N_{objects}^2 + (\alpha + \delta + \epsilon)(N_1 + N_2) + \beta r^2)$. In terms of computational savings, the worst case complexity comparison for our approach would be $O(N_{objects}^2)$ and for standard MPM can be estimated to be $O(C(N_1, N_2, r))$ where C can be $O(N_1 + N_2)^2$ in dense contact scenarios. Therefore, our approach is more efficient than Standard MPM, especially in scenarios when $N_{objects} \ll (N_1 + N_2)$.

B. Qualitative Evaluation

We present qualitative results across different interaction scenarios. These include response to force application (Fig.2), rigid body collision (Fig.3) and multi-body collision (Fig.4). We also validated our method in a 2D simulation. Fig.5 visually compares the collision behavior of a snow ball with a rubber ball generated using standard MPM (Fig. 5(a)) and our approach (Fig.5 (b)). The snow ball is initialized with a velocity that is twice the initial velocity of the rubber ball. Apart few minor differences, our method replicates the same behavior as that generated using the standard MPM but more efficiently (as discussed in Table I). Our results demonstrate temporal coherence across simulation, with objects maintaining their structural integrity and visual quality throughout the dynamic interaction. Our approach ensures stable motion trajectories without unrealistic jittering or interpenetration artifacts.

V. CONCLUSION

We have presented a novel approach for enabling physically plausible interactions in 3D Gaussian Splatting (3DGS) scenes. Our method includes a semantic aware Gaussian grouping mechanism that leverages CLIP, DINO and SAM features to automatically identify and cluster Gaussians into coherent objects. Additionally, we also introduce a physics simulation method that operates on object and particle level for efficient and robust computations. Object level part of our approach focuses on resolution of inter-object interactions while particle level focuses on intra object deformations. Our results show the effectiveness of this approach in handling object interactions and also the computational savings involved.

REFERENCES

- [1] B. Kathariya, D. Y. Lee, T.-W. Huang, T. Shao, P. Yin, and G.-M. Su, "Gaussian splatting: State-of-the-arts and future trends," in *2025 International Conference on Computing, Networking and Communications (ICNC)*. IEEE, 2025, pp. 876–881.
- [2] J. Wang, Y. Liu, P. Wang, C. Lin, J. Hou, X. Li, T. Komura, and W. Wang, "Gaussurf: Geometry-guided 3d gaussian splatting for surface reconstruction," *arXiv preprint arXiv:2411.19454*, 2024.
- [3] Q. Zhang, C. Huang, Q. Zhang, N. Li, and W. Feng, "Su-rigs: Relightable 3d gaussian splatting from sparse views under unconstrained illuminations," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 26 859–26 868.
- [4] J. Gao, C. Gu, Y. Lin, Z. Li, H. Zhu, X. Cao, L. Zhang, and Y. Yao, "Relightable 3d gaussians: Realistic point cloud relighting with brdf decomposition and ray tracing," in *European Conference on Computer Vision*. Springer, 2024, pp. 73–89.
- [5] T. Xie, Z. Zong, Y. Qiu, X. Li, Y. Feng, Y. Yang, and C. Jiang, "Phys-gaussian: Physics-integrated 3d gaussians for generative dynamics," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 4389–4398.
- [6] P. Borycki, W. Smolak, J. Waczyńska, M. Mazur, S. Tadeja, and P. Spurek, "Gasp: Gaussian splatting for physic-based simulations," *arXiv preprint arXiv:2409.05819*, 2024.
- [7] J. Li, Z. Song, and B. Yang, "Trace: Learning 3d gaussian physical dynamics from multi-view videos," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2025, pp. 8820–8829.
- [8] D. Sulsky, Z. Chen, and H. L. Schreyer, "A particle method for history-dependent materials," *Computer methods in applied mechanics and engineering*, vol. 118, no. 1-2, pp. 179–196, 1994.
- [9] X. Li, Y.-L. Qiao, P. Y. Chen, K. M. Jatavallabhula, M. Lin, C. Jiang, and C. Gan, "Pac-nerf: Physics augmented continuum neural radiance fields for geometry-agnostic system identification," *arXiv preprint arXiv:2303.05512*, 2023.
- [10] A. Pumarola, E. Corona, G. Pons-Moll, and F. Moreno-Noguer, "D-nerf: Neural radiance fields for dynamic scenes," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 10 318–10 327.
- [11] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PmlR, 2021, pp. 8748–8763.
- [12] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging properties in self-supervised vision transformers," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 9650–9660.
- [13] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 4015–4026.
- [14] R.-Z. Qiu, G. Yang, W. Zeng, and X. Wang, "Feature splatting: Language-driven physics-based scene synthesis and editing," *arXiv preprint arXiv:2404.01223*, 2024.