⚽ REALDROP: PREDICTING REALISTIC IMPACT RESPONSES OF DROPPING

Anonymous authors
Paper under double-blind review



Figure 1: RealDrop reconstructs physically interactive 3D objects from a single image and predicts how they realistically respond when dropped.

ABSTRACT

Single-image physical reconstruction aims to recover a 3D asset that not only matches appearance but also responds realistically to physical interactions. We focus on object dropping, a particularly revealing setting where impact can trigger diverse and coupled responses, including rebound, deformation, fracture, component motion, liquid release, and splash. However, existing physics-aware reconstruction methods often rely on deformation-dominated physical representations, causing materially complex objects to collapse to overly simple or physically incorrect impact behaviors. We present **RealDrop**, a framework for predicting realistic dropping responses from a single image. RealDrop constructs part-aware physical assets and introduces a Stress-Responsive Particle System (SRPS) that supports appropriate solid responses from elastic rebound to spatially coherent fracture, together with an SPH-based formulation for liquid dynamics. These representations allow distinct physical behaviors to coexist and interact within heterogeneous objects. We further introduce **PhysLab**, which turns physical-property inference from appearance-based prediction into consequence-driven verification by executing predicted properties through simulation and iteratively correcting those that produce physically inconsistent outcomes. Extensive experiments demonstrate that RealDrop produces substantially more realistic and expressive impact responses across diverse single-image inputs.

1 INTRODUCTION

Single-image physical reconstruction aims to recover a physically interactive 3D object from a single photograph. Recent physics-aware Gaussian methods have made substantial progress by

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

coupling reconstructed appearance with physical simulation (Xie et al., 2024; Zeng et al., 2026; Lv et al., 2026). Among commonly studied interactions, dropping provides a particularly revealing test of physical reconstruction, since a brief impact can expose fundamentally different material and structural responses. Depending on the object, impact may lead to rebound, deformation, fracture, component motion, liquid release, or splash, and these responses are often coupled rather than mutually exclusive. Existing methods, however, frequently reduce such diverse outcomes to relatively simple deformation-driven behaviors. For example, an apple-filled bowl may bounce as a single entity after impact, instead of the bowl fracturing while the apples remain intact and roll away.

This limitation stems in part from the restricted response space of the underlying physical representations. Most existing approaches ultimately rely on standard MPM (Lv et al., 2026) or spring-mass formulations (Wang et al., 2026; Jiang et al., 2025), which are primarily designed for continuous elastic or plastic deformation. While such representations can reproduce plausible bouncing and deformation, they do not inherently capture the broader range of responses required by realistic dropping. This limitation is particularly critical in the single-image setting: because a static image provides no observation of how the object actually reacts to external forces, inferred physical properties can only be meaningfully assessed through the behavior produced by simulation. If the simulator itself cannot express the expected outcome, improving parameter estimation alone is insufficient, as different inferred properties may still collapse to similarly incorrect responses. We therefore argue that an expressive physical representation is a prerequisite for meaningful physical inference.

Fragile fracture provides a representative example of this response-space limitation. Conventional MPM and spring-mass formulations do not inherently model brittle separation. In spring-mass systems, fracture is commonly introduced by breaking connections once their deformation exceeds a prescribed threshold, while damage-based MPM progressively weakens material until separation occurs (Huang et al., 2026). Such local failure mechanisms can produce two opposite errors under impact: if the failure condition is not reached, an object that should break may remain connected and rebound; if failure is activated densely over a stressed region, the object may disintegrate into excessively small pieces rather than coherent fragments. More broadly, realistic dropping requires a representation that can preserve ordinary deformation and rebound when appropriate, produce coherent fracture when failure occurs, and support distinct responses across different components, including rigid motion and liquid dynamics.

To address these challenges, we introduce RealDrop, a single-image physical reconstruction framework for predicting realistic impact responses under dropping. RealDrop decomposes the reconstructed object into semantic parts and equips each component with an appropriate physical representation, enabling heterogeneous objects to exhibit distinct yet coupled behaviors under the same impact. For solid components, we develop a Stress-Responsive Particle System (SRPS) that supports stress-dependent responses: it preserves elastic deformation and rebound when appropriate, while producing spatially coherent fracture and meaningful fragments when material failure occurs. For liquid components, we introduce an SPH-based formulation that models liquid release, flow, and splash. Together, these representations substantially expand the physical response space beyond deformation-dominated simulation, allowing RealDrop to reproduce complex dropping events that jointly involve fracture, deformation, rigid-body motion, and liquid dynamics.

To reliably infer the physical properties required by this expressive representation, we further introduce PhysLab, a simulation-in-the-loop refinement framework. Physical properties inferred from appearance alone can be semantically plausible yet produce incorrect physical consequences—for example, an inaccurate material assignment or fracture parameter may cause an object that should break to remain intact. Rather than relying on a single feed-forward prediction, PhysLab treats the inferred part-wise properties as executable physical hypotheses, evaluates the resulting simulated responses, and iteratively corrects inconsistent material assignments and parameters using simulation feedback. By closing the loop between physical inference and observable consequences, PhysLab enables physical properties to be determined by how well they reproduce the expected behavior, rather than by appearance alone. Our contributions are summarized as follows:

- We introduce an expressive physics simulation framework for realistic dropping interactions, combining SRPS for stress-dependent solid responses with SPH-based liquid simulation to support coupled effects including rebound, deformation, fracture, flow, and splash.
- We propose PhysLab, a simulation-in-the-loop physical-property inference framework that iteratively verifies and refines part-wise properties according to their simulated consequences, correcting physically inconsistent predictions from single-pass inference.

- We introduce part-aware physical reconstruction for heterogeneous objects, assigning component-specific material models and parameters so that different parts can exhibit distinct yet coupled impact responses within the same asset.

2 Related Works

2.1 Physics-Aware 3D Reconstruction and Simulation

Physics-aware 3D reconstruction aims to recover simulation-ready assets that encode not only geometry and appearance, but also physical properties governing their response to external interaction(Rho et al., 2026; Kim et al., 2026; Zhang et al., 2024; Li et al., 2023). Early approaches infer these properties from observed motion, where temporal changes provide direct evidence for inverse physics (Lin et al., 2025b; Yang et al., 2025b; Xie et al., 2025). Spring-Gaus (Zhong et al., 2024) integrates a spring-mass model with 3D Gaussians to reconstruct elastic objects from multi-view videos and simulate their future motion. PhysTwin (Jiang et al., 2025) further combines spring-mass simulation, generative shape priors, and Gaussian rendering to recover complete geometry, appearance, and dense physical properties from sparse interaction videos. More recently, ReconPhys (Wang et al., 2026) formulates this problem in a feed-forward setting and jointly predicts Gaussian appearance and physical attributes from a single monocular video. These methods benefit from observed deformation or interaction trajectories, which provide target-specific evidence for physical system identification.

A more challenging setting is to infer physical properties from single image (Xu et al., 2025; Lv et al., 2026; Lin et al., 2025a). DreamPhysics (Huang et al., 2025) optimizes physical properties of dynamic Gaussians using motion priors distilled from video diffusion models and then performs MPM-based simulation (Liu et al., 2025; Lin et al., 2025b). Related approaches such as Physics3D (Liu et al., 2024) similarly exploit generative video priors to infer physical parameters for Gaussian assets. OmniPhysGS (Lin et al., 2025b) broadens the range of supported materials by representing each Gaussian with a mixture of constitutive domain experts, covering elastic, plastic, viscoelastic, and fluid behaviors. PhysGM (Lv et al., 2026) removes per-scene optimization by jointly predicting 3D Gaussians and physical properties from a single image in a feed-forward manner.

2.2 Vision-Language and Agentic Physical Reasoning

The ambiguity of single-image physical reconstruction has also motivated the use of vision-language and multimodal models as semantic priors (Qiu et al., 2024; Cao et al., 2026a;b;c). Rather than estimating physical parameters only from low-level visual cues, these methods exploit high-level knowledge of object category, material, and expected interaction behavior. Phys4DGen (Lin et al., 2025a), for example, uses multimodal reasoning to identify material composition and initialize properties for downstream simulation. PhysSplat (Zhao et al., 2025) similarly leverages multimodal reasoning to estimate physical attributes and their spatial distributions. Such semantic priors are particularly useful when appearance alone is insufficient to distinguish materials with similar geometry or texture. More recent methods further incorporate simulation into the reasoning process. PhyMAGIC (Meng et al., 2025) combines image-to-video generation, LLM-based confidence reasoning, and differentiable simulation, iteratively refining motion hypotheses using simulation feedback. PhysMAS (Qin et al., 2026) adopts a multi-agent formulation for compositional 4D Gaussians, performing part-level material reasoning, binding inferred properties to per-particle MPM fields, and screening candidate forward simulations. These methods demonstrate that physical reasoning can benefit from closing the loop between semantic inference and simulation, instead of relying on a single one-shot prediction.

3 Method

Given a single input image, RealDrop reconstructs a part-aware 3D Gaussian object and assigns each semantic part an independent physical description. Based on the inferred material type, solid components are simulated using our Stress-Responsive Particle System (SRPS), while liquid components are handled by an SPH-based liquid simulation. To infer physical properties without ground-truth interactions, we introduce PhysLab, a VLM-guided iterative inference pipeline. The agent first predicts the expected dropping interaction and an initial set of part-aware physical parameters. RealDrop then performs physics simulation and renders the resulting interaction, which is returned to the agent together with the $\mathrm{CLIP}_{\mathrm{sim}}$ score. The agent iteratively updates the physical parameters based on this visual feedback until the simulated interaction is consistent with the hypothesized behavior. In this way, RealDrop combines part-aware reconstruction, expressive physics simulation, and simulation-guided physical-property inference within a unified single-image framework.

3.1 Part-Aware Physical Asset Construction

We first construct a part-decomposed 3D asset following OmniPart (Yang et al., 2025c). Given the input image, OmniPart obtains 2D part masks and uses them to condition an autoregressive structure

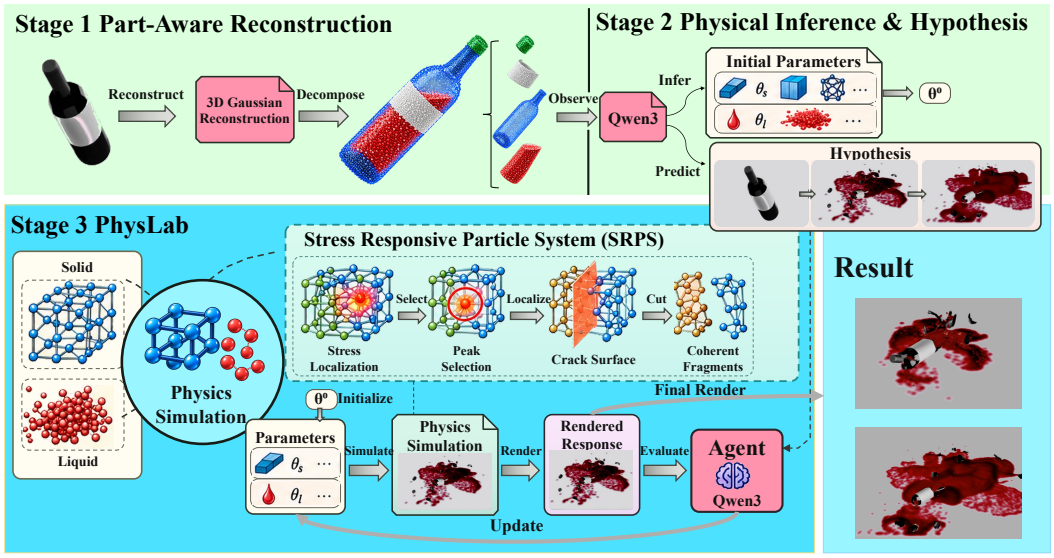

Figure 2: **Overview of RealDrop.** Given a single image, RealDrop reconstructs a part-aware Gaussian asset and initializes component-wise physical properties and expected impact behaviors. PhysLab then treats these properties as executable hypotheses and iteratively refines them through simulation and rendered feedback. The resulting physical asset combines our Stress-Responsive Particle System (SRPS) for solid responses with an SPH-based liquid model (Bender & Koschier, 2015), enabling realistic and coupled dropping responses across heterogeneous components.

planner that predicts a set of part-level 3D bounding boxes. A spatially conditioned generation model then jointly reconstructs the individual parts within the predicted layout, producing a geometrically coherent part-aware 3D object. Built upon TRELLIS (Xiang et al., 2025), the resulting asset can be decoded into 3D Gaussian primitives, which we use as the appearance representation throughout our pipeline.

The visible part decomposition alone, however, does not provide the internal structure required for physics simulation. We therefore use a VLM agent to further inspect each semantic part and determine whether it should be modeled as a solid body or a hollow shell. In practice we used Qwen 3 (Yang et al., 2025a) as the VLM agent. For hollow parts, the agent additionally determines whether an internal material is expected. For example, a wine bottle is represented by a hollow glass shell together with an internal liquid component. When such hidden contents are inferred, we use inpainting to complete the occluded content and introduce it as an additional semantic part for subsequent simulation.

We then convert the completed part-aware asset into simulation particles. Each part is sampled on a shared Cartesian lattice, where each particle stores its reference position $\mathbf{X}_i$, reference volume V_i^0, part identity $p(i)$, and solid or liquid phase label. The reconstructed Gaussians are associated with simulation particles of the corresponding parts through the reconstruction-to-physical coordinate transformation. During simulation, particle motion is transferred back to the bound Gaussians, allowing the reconstructed appearance to follow the physical evolution of each part.

3.2 Physics Simulation

As discussed in Sec. D (in Appendix-Preliminary), conventional MPM primarily models deformation through continuous constitutive updates, while a standard spring-mass system preserves a fixed material graph. Fracture therefore requires additional failure mechanisms. A common extension is to weaken material or remove individual bonds once a local failure criterion is satisfied. Under strong impact, however, many nearby particles or bonds can satisfy such criteria simultaneously, leading to diffuse separation rather than coherent fragments. To address this limitation, we introduce a Stress-Responsive Particle System (SRPS), which combines energy-driven bond failure with stress-guided, spatially coherent crack propagation. As illustrated in Fig. 3, SRPS maintains persistent material connectivity for solids and explicitly evolves this connectivity as cracks form and propagate.

Solid and liquid phases share particle positions $\mathbf{x}_i$, velocities $\mathbf{v}_i$, masses m_i, reference volumes V_i^0, and force buffers. Solid particles additionally maintain reference positions $\mathbf{X}_i$ and a persistent material graph $\mathcal{E}_0$, with $\mathcal{E}(t) \subseteq \mathcal{E}_0$ denoting the surviving bonds. Liquid particles instead interact

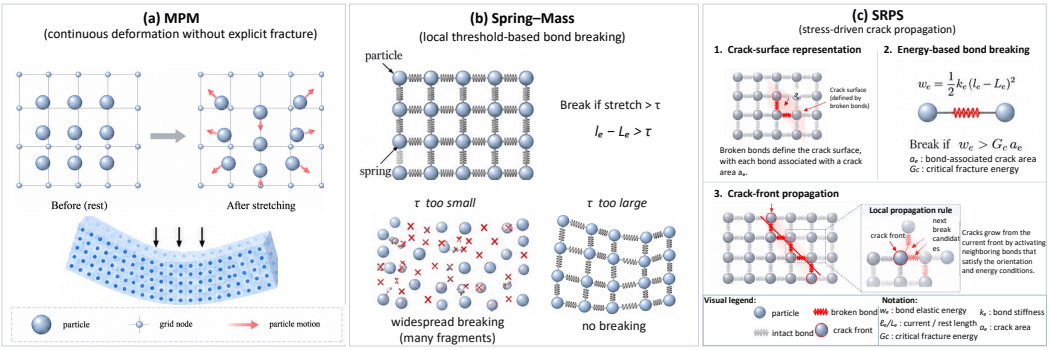

Figure 3: Mechanistic comparison with Spring–Mass and MPM. SpringMass breaks bonds using local stretch thresholds, often causing either excessive fragmentation or no fracture, while standard MPM preserves a continuous material representation and does not natively form crack surfaces. Our method uses stress-guided bond failure and energy-consistent crack-front propagation to produce spatially coherent fractures.

through dynamically updated spatial neighborhoods. Forces from both phases are accumulated before joint integration.

Energy-Driven Bond Failure For a solid bond $e = (i, j)$, let L_e^0, R_e, and l_e denote its initial, natural, and current lengths. We define its elastic energy and effective fracture area as

$$U_e = \frac{1}{2}k_e(l_e - R_e)^2 \qquad A_e = \kappa\frac{V_i^0 + V_j^0}{2L_e^0} \tag{1}$$

where U_e is the elastic energy stored in bond e, A_e is its effective fracture area, k_e denotes the bond stiffness, and κ calibrates the accumulated bond areas against the corresponding solid cross section. A bond becomes eligible for tensile failure only when it is stretched and stores sufficient energy to overcome the material fracture energy:

$$l_e > R_e \qquad \mathcal{Q}_e = U_e - G_{c,e}A_e > 0 \tag{2}$$

where $G_{c,e}$ is the fracture energy per unit area, and $\mathcal{Q}_e$ measures the energetic driving force for bond failure. This criterion determines whether a bond has sufficient energy to fail, but does not by itself determine where a coherent crack should form.

Stress-Guided Crack Initiation We therefore estimate the local tensile stress from the surviving material connections. Let

$$\mathcal{N}_t(i) = \{j \mid (i, j) \in \mathcal{E}(t)\} \tag{3}$$

denote the intact bonded neighbors of particle i at simulation step t, where $\mathcal{E}(t)$ is the set of surviving bonds. The unit direction of bond (i, j) is defined as

$$\hat{\mathbf{d}}_{ij} = \frac{\mathbf{x}_j - \mathbf{x}_i}{l_{ij}} \tag{4}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the current particle positions and $l_{ij} = \|\mathbf{x}_j - \mathbf{x}_i\|$ is the current bond length. We estimate the local stress tensor of particle i as

$$\boldsymbol{\sigma}_i = \frac{1}{2V_i^0}\sum_{j\in\mathcal{N}_t(i)} k_{ij}(l_{ij} - R_{ij})l_{ij}\hat{\mathbf{d}}_{ij} \otimes \hat{\mathbf{d}}_{ij} \tag{5}$$

where V_i^0 denotes the reference volume associated with particle i, k_{ij} is the stiffness of bond (i, j), R_{ij} is its natural length, and $\otimes$ denotes the outer product. After K averaging passes over the reference material graph, we extract the maximum principal tensile stress and its corresponding eigenvector. Regions with high tensile stress provide crack nucleation candidates, while the principal tensile direction determines the local crack orientation.

Finite Crack Propagation Directly breaking every energetically eligible bond can still produce excessive fragmentation. We therefore apply local peak selection and retain only locally dominant failure candidates. For each eligible bond e, we require

$$\mathcal{Q}_e \geq \eta \max_{e'\in\mathcal{N}_{\text{tip}}(e)} \mathcal{Q}_{e'}, \qquad \eta \in (0, 1] \tag{6}$$

Here, $\mathcal{Q}_e$ is the fracture driving energy of bond e, $\mathcal{N}_{\text{tip}}(e)$ denotes the neighboring bonds considered when testing whether e is a local peak, e' indexes a bond in this neighborhood, and η controls the strictness of the local-maximum selection. Only bonds satisfying this condition are retained as failure candidates.

Each crack is represented by a union of disks with center $\mathbf{c}_s$, normal $\mathbf{n}_s$, and radius r_s. A bond can extend an existing crack only when its reference-space segment intersects one of these finite disks. For a candidate bond (i, j), let $\mathbf{X}_i$ and $\mathbf{X}_j$ denote the reference positions of its two endpoints, and define their signed distances to the plane of crack disk s as

$$d_i = (\mathbf{X}_i - \mathbf{c}_s) \cdot \mathbf{n}_s \qquad d_j = (\mathbf{X}_j - \mathbf{c}_s) \cdot \mathbf{n}_s \tag{7}$$

We require

$$d_i d_j < 0 \qquad \left\| \mathbf{X}_i + \frac{d_i}{d_i - d_j}(\mathbf{X}_j - \mathbf{X}_i) - \mathbf{c}_s \right\|_2 \leq r_s \tag{8}$$

where d_i and d_j indicate on which side of the disk plane the two bond endpoints lie. The first condition ensures that the bond crosses the disk plane, while the second requires the resulting segment–plane intersection point to fall within the finite disk radius r_s. A confirmed failure adds a new disk at the intersection point and inherits the identity of its parent crack. This finite-surface constraint localizes bond removal around an evolving crack rather than allowing neighboring bonds to fail independently throughout the stressed region. The final fragments are given by the connected components of the surviving material graph.

For compressible materials, intact bonds can additionally undergo irreversible length shortening:

$$R_e^+ = \min \left[R_e, \max \left((1 - p_{\max}) L_e^0, \frac{l_e}{1 - \epsilon_y} \right) \right] \tag{9}$$

where R_e is the current natural length of bond e, R_e^+ is the updated natural length, L_e^0 is its initial length, l_e is its current length, ϵ_y denotes the compressive yield strain, and $p_{\max}$ bounds the maximum permanent shortening relative to the initial length. This update is applied before force evaluation so that elastic force, stress estimation, and fracture evaluation use the same natural length.

SPH-Based Liquid Simulation While solid particles in SRPS maintain persistent material connections, liquid particles continuously rearrange during simulation. We therefore adopt an SPH-based (Bender & Koschier, 2015) liquid simulation, where each liquid particle carries position $\mathbf{x}_i$, velocity $\mathbf{v}_i$, and mass m_i. Let $\mathcal{L}$ denote the set of liquid particles. Instead of fixed material bonds, liquid interactions are defined by spatial neighborhoods that are updated at every simulation step. For particle i, we define

$$\mathcal{N}_i(t) = \left\{ j \in \mathcal{L} \;\middle|\; \|\mathbf{x}_j(t) - \mathbf{x}_i(t)\|_2 < h \right\} \tag{10}$$

where h denotes the kernel support radius. This dynamic neighborhood allows liquid particles to continuously rearrange, enabling liquid flow and free-surface deformation.

The density of each liquid particle is estimated from its neighboring liquid particles and nearby solid boundaries:

$$\rho_i = \sum_{j \in \mathcal{N}_i(t)} m_j W(\mathbf{x}_i - \mathbf{x}_j, h) + \rho_i^{\mathrm{b}} \tag{11}$$

where W denotes the smoothing kernel and ρ_i^{b} accounts for the contribution of nearby solid boundaries. Pressure forces are computed from the estimated density to resist compression. Gravity, viscosity, and capillary forces further determine the liquid motion. We use iterative pressure correction in the incompressible solver to maintain approximately constant density during simulation. Solid and liquid components interact within the same physics simulation. Moving solid particles provide dynamic boundaries for neighboring liquid particles and exchange forces with them. When SRPS removes material connections during fracture, the solid geometry separates accordingly. Liquid leakage naturally occurs once the resulting opening becomes sufficiently large, without explicitly prescribing liquid outlets. This enables RealDrop to simulate interactions such as a glass container fracturing upon impact while its liquid contents escape and flow.

Continuous Liquid Appearance We decouple the liquid simulation discretization from its visual representation. Rather than rendering individual simulation particles as isolated Gaussian primitives, we reconstruct a continuous surface from the simulated particle distribution and represent the resulting interface using surface-aligned anisotropic Gaussians. This decoupling allows the visual resolution of the liquid to be controlled independently of the simulation particle count and produces a continuous appearance under Gaussian rendering.

3.3 PHYSLAB:SIMULATION-GUIDED PHYSICAL PROPERTY INFERENCE

A single image provides no ground-truth observation of how an object responds to physical interaction, making its physical properties inherently under-constrained. To address this ambiguity, we introduce PhysLab, a VLM-guided iterative physical property inference pipeline that estimates part-wise physical properties through repeated physics simulation and visual feedback. Rather than relying on a single prediction of numerical parameters, PhysLab evaluates whether the resulting simulated

interaction agrees with the physical behavior expected from the input image. As illustrated in Fig. 2, the VLM agent (Yang et al., 2025a) first analyzes the input image and formulates a hypothesis of the expected dropping interaction for each semantic part. The hypothesis describes observable physical outcomes, including whether a part should remain rigid, deform, fracture, or behave as a liquid. Based on this hypothesis, the agent predicts an initial physical configuration for each part, including whether it is solid or liquid and its corresponding simulation parameters.

Given the physical properties, RealDrop performs physics simulation and renders the resulting interaction. The rendered frames and the $\mathrm{CLIP}_{\mathrm{sim}}$ score are provided to the agent as feedback. The agent compares the simulated interaction with the initial hypothesis, identifies inconsistencies in the observed physical behavior, and updates the corresponding part-wise parameters. The updated configuration is then simulated and rendered again. Repeating this process allows the physical parameters to be progressively refined according to their observable simulation outcomes. In practice, PhysLab performs at most five refinement rounds.

4 Experiments

4.1 Experimental Setup

Datasets and Evaluation Protocol We evaluate RealDrop on 50 randomly selected objects sourced from Objaverse (Deitke et al., 2023) and PhysAssets (Lv et al., 2026). Each test case starts from a single-view image and a motion prompt. Our evaluation covers heterogeneous multi-part objects, fragile fracture, and liquid dynamics, allowing us to examine both material-dependent responses and coupled physical events. For all methods, we match the input image, motion prompt, external action, ground condition, camera trajectory, sequence duration, timestamps, and rendering resolution whenever supported by the corresponding method. All experiments are implemented on Nvidia A100.

Table 1: Quantitative comparison. For baselines without reproducible public implementations, we report the $\mathrm{CLIP}_{\mathrm{sim}}$ values provided in their original papers for reference. UPR is unavailable for PhyMAGIC and PhysMAS as their implementations are not publicly released and this metric is not reported in their papers.

Method	$\mathbf{CLIP}_{\mathbf{sim}}\uparrow$	UPR↑
OmniPhysGS (Lin et al., 2025b)	0.2091	4%
DreamPhysics (Huang et al., 2025)	0.2291	5%
PhyMAGIC (Meng et al., 2025)	0.2520	-
PhysGM (Lv et al., 2026)	0.2748	9%
PhysMAS (Qin et al., 2026)	0.2787	-
Phys4DGen (Lin et al., 2025a)	0.2731	6%
RealDrop (Ours)	**0.3054**	**76%**

Evaluation Metrics We report CLIP similarity ($\mathrm{CLIP}_{\mathrm{sim}}$) (Radford et al., 2021) and User Preference Rate (UPR), following prior physical 4D generation works (Lv et al., 2026; Lin et al., 2025a; Qin et al., 2026). For $\mathrm{CLIP}_{\mathrm{sim}}$, a frozen CLIP ViT-B/32 encoder computes the cosine similarity between each rendered frame and the full motion prompt; frame-level scores are averaged within each sequence and then equally across test cases. UPR is measured through pairwise human evaluation with 100 participants, who select the video exhibiting more realistic physical behavior. Importantly, the two metrics capture complementary aspects of quality: $\mathrm{CLIP}_{\mathrm{sim}}$ primarily measures high-level semantic consistency with the motion description, but may assign similar scores to physically distinct outcomes, such as a bottle bouncing intact versus breaking and spilling. UPR therefore better captures perceived physical realism, particularly when semantically similar dropping results exhibit substantially different impact responses.

4.2 Comparison with State-of-the-Art Methods

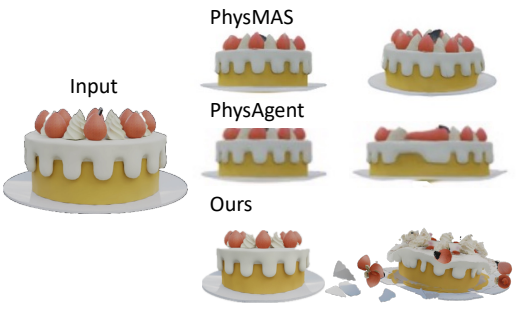

Figure 5: Comparison with PhysMAS(Qin et al., 2026) and PhysAgent(Yang et al., 2026).

Quantitative Comparison Table 1 shows that RealDrop achieves the best performance on both $\mathrm{CLIP}_{\mathrm{sim}}$ and UPR. While the gain in $\mathrm{CLIP}_{\mathrm{sim}}$ is moderate, recent SOTA methods are tightly clustered within a narrow range (approximately 0.21–0.28), suggesting limited sensitivity of this metric to physical correctness. In particular, physically distinct outcomes, such as an intact bottle bouncing versus breaking and spilling, can remain semantically similar in CLIP space. By contrast, the substantially larger improvement in UPR indicates that RealDrop produces physical responses that are more consistent with human judgments of realistic interaction. This gap further suggests that the benefit of our method lies primarily in impact-induced physical realism, which is not fully captured by semantic similarity alone. Our PhysLab further improves the results by refining the initial physical configuration through simulation feedback and correcting implausible behaviors that are difficult to infer from appearance.

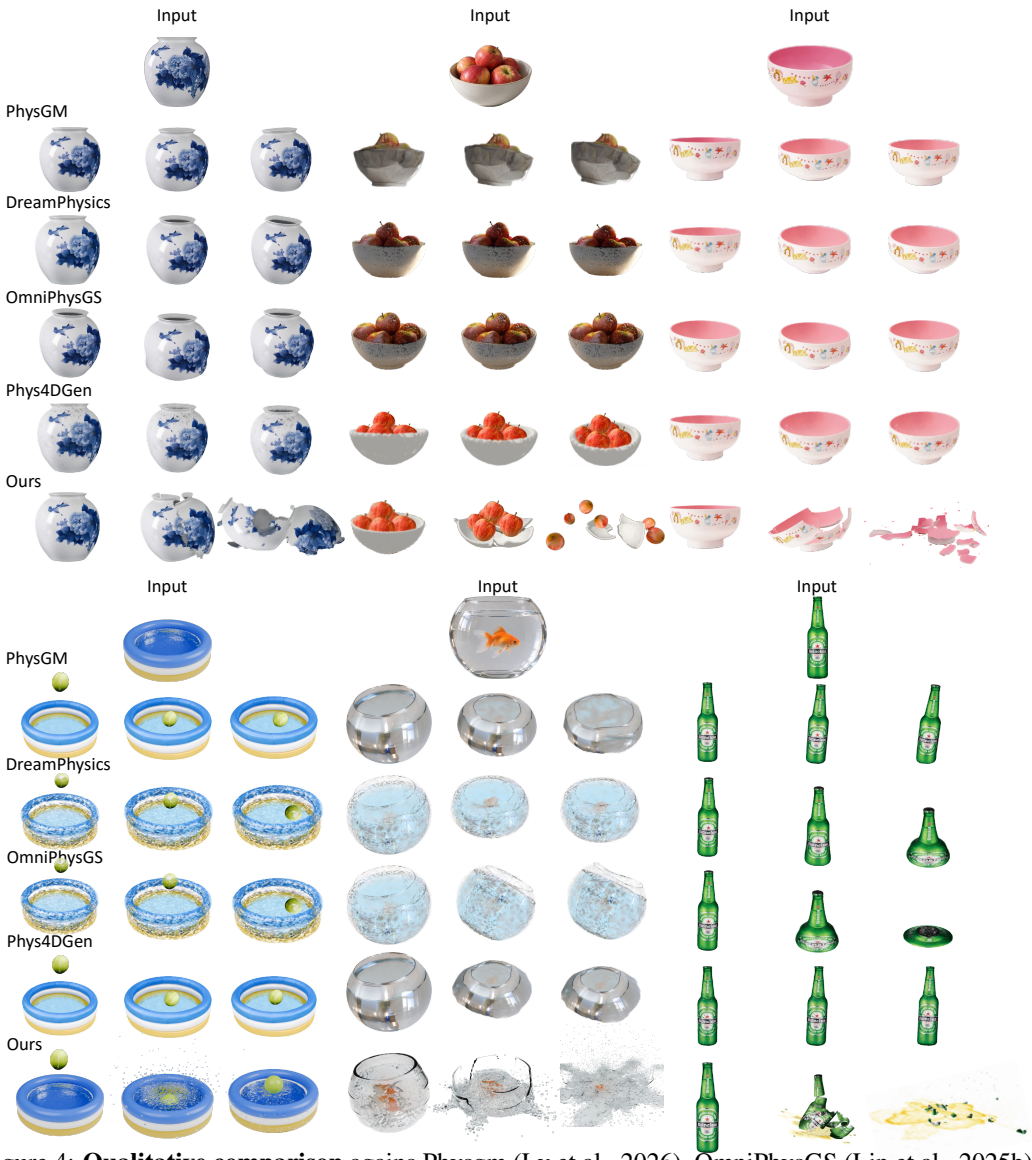

Figure 4: **Qualitative comparison** agains Physgm (Lv et al., 2026), OmniPhysGS (Lin et al., 2025b), Phys4DGen (Lin et al., 2025a), and DreamPhysics (Huang et al., 2025)

Qualitative Comparison We evaluate all methods on in-the-wild single images to examine whether they can predict the correct physical consequences of impact. As shown in Fig. 4, existing methods largely remain within a homogeneous deformation regime: fragile objects bend or distort after impact, but rarely undergo coherent fracture. RealDrop instead produces distinct impact-induced state transitions, including brittle fracture and component-dependent material responses. For an apple-filled ceramic bowl, the bowl fractures into pieces while the apples remain intact and roll out naturally; for a beer bottle, the glass shatters while the contained liquid is released and flows outward. These examples highlight a key difference: RealDrop does not merely deform an object globally, but resolves how different physical components respond and interact after impact. Additional qualitative results are provided in Fig. 9(in Appendix).

4.3 Ablation Studies

We ablate the key components of RealDrop, including part-aware modeling, fracture-aware simulation, peak selection, and liquid modeling. Quantitative results are reported in Tab. 2, while qualitative comparisons are shown in Figs. 6, 7, and 11.

Part-aware modeling. Part-aware modeling is essential for composite objects whose components exhibit different physical responses. Without explicit part-level modeling, all components are forced to share the same physical properties, creating conflicts between incompatible materials. As shown in Fig. 11(in Appendix), the ceramic bowl and the apples are assigned identical material properties. A

Table 2: Ablation studies of our proposed RealDrop.

Variant	$\mathbf{CLIP_{sim}}$ ↑	UPR↑
w/o Part-Aware Modeling	0.2705	14%
w/o Fracture-Aware Simulation	0.2732	9%
w/o Peak Selection	0.2848	19%
w/o Liquid Modeling	0.2978	21%
Full Model	**0.3054**	37%

Table 3: Comparison of different physical-property inference strategies. UPR denotes how often the full agent is preferred to the corresponding alternative.

Inference Strategy	$\mathbf{CLIP_{sim}}$ ↑	UPR↑
Single-Pass VLM	0.2712	15%
24-Layer Transformer	0.2634	7%
Full Model	**0.3054**	78%

fragile assignment would unrealistically fracture the apples, while a non-fragile assignment prevents the bowl from breaking. The full model resolves this ambiguity by assigning component-specific physical properties, allowing the bowl to fracture while the apples remain intact after impact.

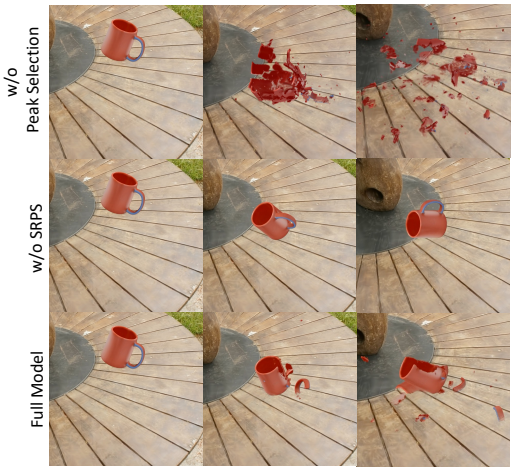

Figure 6: Ablation of fracture-aware simulation.

Fracture-aware simulation. Fracture-aware simulation is necessary for modeling brittle failure under impact. As shown in Fig. 6, removing this component causes the mug to remain largely connected after hitting the ground, reducing the response to deformation rather than fracture. In contrast, the full model enables impact-induced breakup and produces fragments consistent with the expected behavior of fragile materials.

Liquid modeling. Liquid modeling is critical for reproducing coherent liquid motion after impact. Without it, Gaussian particles disperse independently and fail to maintain a consistent liquid structure, as shown in Fig. 7. The full model preserves collective liquid behavior, enabling plausible liquid release, flow, and spreading after the container is broken.

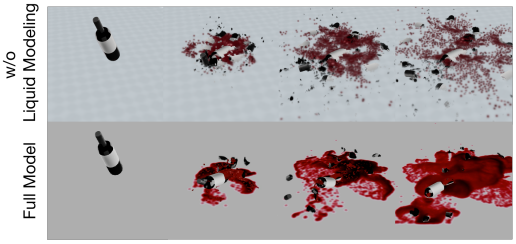

Figure 7: Ablation of the liquid modeling.

4.4 Validation of PhysLab

We further validate PhysLab against two alternative physical-property inference strategies: a Single-Pass VLM and a 24-Layer Transformer. Quantitative results are reported in Tab. 3, with qualitative examples shown in Fig. 11(in Appendix). Both alternatives struggle in challenging multi-material cases. The 24-Layer Transformer correctly identifies the bowl as rigid but inaccurately estimates the fracture parameter G_c, preventing fracture under impact. The Single-Pass VLM instead assigns soft material properties to the bowl. These failures show that one-shot inference can produce visually plausible properties that lead to incorrect outcomes. PhysLab addresses this issue through simulation-guided refinement. It evaluates the behavior produced by an initial physical hypothesis and iteratively updates the predicted properties based on simulation feedback, correcting both material assignments and fracture parameters that are difficult to infer from appearance alone. This demonstrates that its advantage comes from closed-loop physical verification rather than increased feed-forward model capacity.

5 Conclusion



Figure 8: RealDrop generalizes to in-the-wild single-image inputs and produces object-dependent impact responses.

We presented RealDrop, a single-image physical reconstruction framework for predicting realistic impact responses under object dropping. RealDrop expands the response space of existing deformation-dominated methods through part-aware physical modeling, a Stress-Responsive Particle System for adaptive solid responses from rebound to coherent fracture, and SPH-based liquid simulation for release, flow, and splash.

REFERENCES

Jan Bender and Dan Koschier. Divergence-free smoothed particle hydrodynamics. In *Proceedings of the 14th ACM SIGGRAPH/Eurographics symposium on computer animation*, pp. 147–155, 2015.

Ziang Cao, Zhaoxi Chen, Liang Pan, and Ziwei Liu. Physx-3d: Physical-grounded 3d asset generation. *Advances in Neural Information Processing Systems*, 38:93771–93784, 2026a.

Ziang Cao, Fangzhou Hong, Zhaoxi Chen, Liang Pan, and Ziwei Liu. Physx-anything: Simulation-ready physical 3d assets from single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5839–5848, 2026b.

Ziang Cao, Yinghao Liu, Haitian Li, Runmao Yao, Fangzhou Hong, Zhaoxi Chen, Liang Pan, and Ziwei Liu. Physx-omni: Unified simulation-ready physical 3d generation for rigid, deformable, and articulated objects. *arXiv preprint arXiv:2605.21572*, 2026c.

Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3d objects. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13142–13153, 2023.

Daniel Holz, Stefan Rhys Jeske, Fabian Löschner, Jan Bender, Yin Yang, and Sheldon Andrews. Multiphysics simulation methods in computer graphics. In *Computer Graphics Forum*, volume 44, pp. e70082. Wiley Online Library, 2025.

Bei Huang, Yixin Chen, Ruijie Lu, Gang Zeng, Hongbin Zha, Yuru Pei, and Siyuan Huang. Gaussianfluent: Gaussian simulation for dynamic scenes with mixed materials. *arXiv preprint arXiv:2601.09265*, 2026.

Tianyu Huang, Haoze Zhang, Yihan Zeng, Zhilu Zhang, Hui Li, Wangmeng Zuo, and Rynson WH Lau. Dreamphysics: Learning physics-based 3d dynamics with video diffusion priors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 3733–3741, 2025.

Hanxiao Jiang, Hao-Yu Hsu, Kaifeng Zhang, Hsin-Ni Yu, Shenlong Wang, and Yunzhu Li. Phystwin: Physics-informed reconstruction and simulation of deformable objects from videos. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7219–7230. IEEE, 2025.

Hyunjin Kim, Ri-Zhao Qiu, Guangqi Jiang, and Xiaolong Wang. Egophys: Learning generalizable physics models of deformable objects from egocentric video. *arXiv preprint arXiv:2606.16202*, 2026.

Xuan Li, Yi-Ling Qiao, Peter Yichen Chen, Krishna Murthy Jatavallabhula, Ming Lin, Chenfanfu Jiang, and Chuang Gan. Pac-nerf: Physics augmented continuum neural radiance fields for geometry-agnostic system identification. *arXiv preprint arXiv:2303.05512*, 2023.

Jiajing Lin, Zhenzhong Wang, Dejun Xu, Shu Jiang, Yunpeng Gong, and Min Jiang. Phys4dgen: Physics-compliant 4d generation with multi-material composition perception. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 10398–10407, 2025a.

Yuchen Lin, Chenguo Lin, Jianjin Xu, and Yadong Mu. Omniphysgs: 3d constitutive gaussians for general physics-based dynamics generation. In *International Conference on Learning Representations*, volume 2025, pp. 70160–70187, 2025b.

Fangfu Liu, Hanyang Wang, Shunyu Yao, Shengjun Zhang, Jie Zhou, and Yueqi Duan. Physics3d: Learning physical properties of 3d gaussians via video diffusion. *arXiv preprint arXiv:2406.04338*, 2024.

Zhuoman Liu, Weicai Ye, Yan Luximon, Pengfei Wan, and Di Zhang. Unleashing the potential of multi-modal foundation models and video diffusion for 4d dynamic physical scene simulation. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11016–11025. IEEE, 2025.

Chunji Lv, Zequn Chen, Donglin Di, Weinan Zhang, Hao Li, Chen Wei, Yinjie Lei, and Changsheng Li. Physgm: Large physical gaussian model for feed-forward 4d synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 29855–29865, 2026.

Siwei Meng, Yawei Luo, and Ping Liu. Phymagic: Physical motion-aware generative inference with confidence-guided llm. *arXiv preprint arXiv:2505.16456*, 2025.

Jiang Qin, Chunji Lv, Yangguang Wei, Yang Gao, Ming Liu, Lizhong Ding, Ye Yuan, Yinjie Lei, and Changsheng Li. Physmas: Physics-grounded multi-agent synthesis of compositional 4d gaussians. *arXiv preprint arXiv:2609.07174*, 2026.

Ri-Zhao Qiu, Ge Yang, Weijia Zeng, and Xiaolong Wang. Feature splatting: Language-driven physics-based scene synthesis and editing. *arXiv preprint arXiv:2404.01223*, 2024.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Daniel Rho, Jun Myeong Choi, Matthew Thornton, Biswadip Dey, and Roni Sengupta. Monophysics: Estimating geometry, appearance, and physical parameters from monocular videos. *arXiv preprint arXiv:2605.30320*, 2026.

Boyuan Wang, Xiaofeng Wang, Yongkang Li, Zheng Zhu, Yifan Chang, Angen Ye, Guosheng Zhao, Chaojun Ni, Guan Huang, Yijie Ren, et al. Reconphys: Reconstruct appearance and physical attributes from single video. *arXiv preprint arXiv:2604.07882*, 2026.

Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. Structured 3d latents for scalable and versatile 3d generation. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21469–21480. IEEE, 2025.

Tianyi Xie, Zeshun Zong, Yuxing Qiu, Xuan Li, Yutao Feng, Yin Yang, and Chenfanfu Jiang. Physgaussian: Physics-integrated 3d gaussians for generative dynamics. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4389–4398. IEEE, 2024.

Ziyang Xie, Zhizheng Liu, Zhenghao Peng, Wayne Wu, and Bolei Zhou. Vid2sim: Realistic and interactive simulation from video for urban navigation. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1581–1591. IEEE, 2025.

Xinli Xu, Wenhang Ge, Dicong Qiu, ZhiFei Chen, Dongyu Yan, Zhuoyun Liu, Haoyu Zhao, Hanfeng Zhao, Shunsi Zhang, Junwei Liang, et al. Gaussianproperty: Integrating physical properties to 3d gaussians with lmms. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7231–7240. IEEE, 2025.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025a.

Yehui Yang, Bo Zhao, Junzhe Cao, Hui Ma, Yue Sun, Wenjin Wang, and Zitong Yu. Physagent: A multi-agent framework for reliable remote heart rate estimation. *arXiv preprint arXiv:2608.00066*, 2026.

Yu Yang, Zhilu Zhang, Xiang Zhang, Yihan Zeng, Hui Li, and Wangmeng Zuo. Physworld: From real videos to world models of deformable objects via physics-aware demonstration synthesis. *arXiv preprint arXiv:2510.21447*, 2025b.

Yunhan Yang, Yufan Zhou, Yuan-Chen Guo, Zi-Xin Zou, Yukun Huang, Ying-Tian Liu, Hao Xu, Ding Liang, Yan-Pei Cao, and Xihui Liu. Omnipart: Part-aware 3d generation with semantic decoupling and structural cohesion. In *Proceedings of the SIGGRAPH Asia 2025 Conference Papers*, pp. 1–12, 2025c.

Bin Zeng, Johannes Künzel, Anna Hilsmann, and Peter Eisert. Phys-3d: Physics-constrained real-time crowd tracking and counting on railway platforms. *arXiv preprint arXiv:2602.23177*, 2026.

Tianyuan Zhang, Hong-Xing Yu, Rundi Wu, Brandon Y Feng, Changxi Zheng, Noah Snavely, Jiajun Wu, and William T Freeman. Physdreamer: Physics-based interaction with 3d objects via video generation. In *European Conference on Computer Vision*, pp. 388–406. Springer, 2024.

Haoyu Zhao, Hao Wang, Xingyue Zhao, Hao Fei, Hongqiu Wang, Chengjiang Long, and Hua Zou. Physsplat: Efficient physics simulation for 3d scenes via mllm-guided gaussian splatting. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5242–5252. IEEE, 2025.

Licheng Zhong, Hong-Xing Yu, Jiajun Wu, and Yunzhu Li. Reconstruction and simulation of elastic objects with spring-mass 3d gaussians. In *European Conference on Computer Vision*, pp. 407–423. Springer, 2024.

A Ethics Statement

This work adheres to the ICLR Code of Ethics. In this study, no human subjects or animal experimentation was involved. All datasets used, including Objaverse (Deitke et al., 2023) and PhysAssets (Lv et al., 2026), were sourced in compliance with relevant usage guidelines, ensuring no violation of privacy. We have taken care to avoid any biases or discriminatory outcomes in our research process. No personally identifiable information was used, and no experiments were conducted that could raise privacy or security concerns. We are committed to maintaining transparency and integrity throughout the research process.

B Reproducibility Statement

We have made every effort to ensure that the results presented in this paper are reproducible. The experimental setup, including training steps, model configurations, and hardware details, is described in detail in the paper. We have also provided a full description of RealDrop, to assist others in reproducing our experiments.

Additionally, Dataset for our experiment, such as Objaverse (Deitke et al., 2023) and PhysAssets (Lv et al., 2026), are publicly available, ensuring consistent and reproducible evaluation results.

We believe these measures will enable other researchers to reproduce our work and further advance the field.

C LLM Usage

Large Language Models (LLMs) were used to aid in the writing and polishing of the manuscript. Specifically, we used an LLM to assist in refining the language, improving readability, and ensuring clarity in various sections of the paper. The model helped with tasks such as sentence rephrasing, grammar checking, and enhancing the overall flow of the text.

It is important to note that the LLM was not involved in the ideation, research methodology, or experimental design. All research concepts, ideas, and analyses were developed and conducted by the authors. The contributions of the LLM were solely focused on improving the linguistic quality of the paper, with no involvement in the scientific content or data analysis.

The authors take full responsibility for the content of the manuscript, including any text generated or polished by the LLM. We have ensured that the LLM-generated text adheres to ethical guidelines and does not contribute to plagiarism or scientific misconduct.

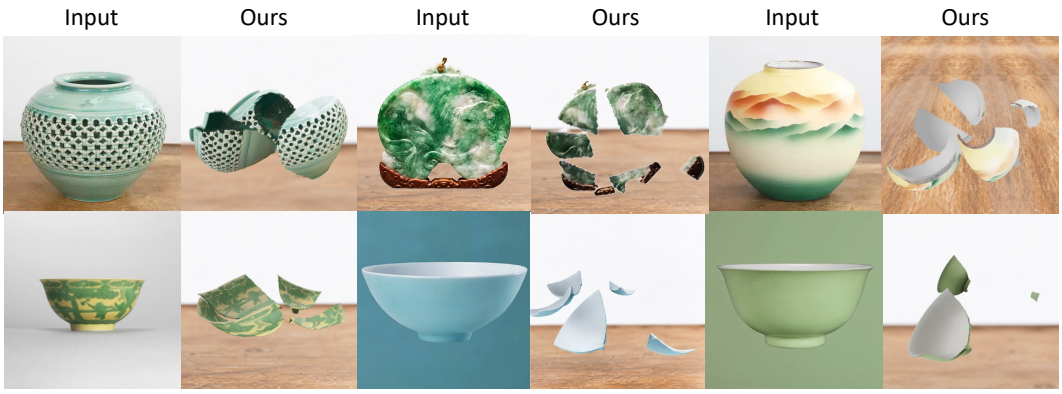

Figure 9: Qualitative results of our method. Our method produce realistic fractures fragile material.

Appendix

D PRELIMINARIES

Material Point Method The Material Point Method (MPM) represents an object using a set of Lagrangian particles carrying physical quantities such as mass, velocity, deformation, and material properties (Holz et al., 2025; Lv et al., 2026; Huang et al., 2026). At each simulation step, particle states are transferred to a background Eulerian grid through particle-to-grid (P2G) transfer, where forces and velocities are updated according to the constitutive model. The updated grid states are then transferred back to the particles through grid-to-particle (G2P) transfer, after which particle positions are advected. In its standard formulation, MPM updates material deformation through a continuous constitutive model without explicitly changing material connectivity, and therefore primarily describes deformation and elastic responses.

Spring-Mass System A spring-mass system discretizes an object into particles connected by springs (Zhong et al., 2024; Jiang et al., 2025). Particles carry mass, position, and velocity, while springs encode local connectivity and elasticity. For a spring connecting particles i and j, the elastic force is

$$\mathbf{f}_{ij} = -k_{ij}\left(\|\mathbf{x}_i - \mathbf{x}_j\| - l_{ij}^0\right)\frac{\mathbf{x}_i - \mathbf{x}_j}{\|\mathbf{x}_i - \mathbf{x}_j\|} \tag{12}$$

Here, k_{ij} denotes the spring stiffness and l_{ij}^0 denotes the rest length. A damping force is commonly added to suppress oscillations. The forces from connected springs and external forces are accumulated on each particle and integrated over time to update its velocity and position. With fixed spring connectivity, the system can stretch, compress, and recover elastically, but cannot represent structural separation unless an additional connection-breaking mechanism is introduced.

D.1 FURTHER COMPARISON

We further compare RealDrop with PhysMAS (Qin et al., 2026) and PhysAgent (Yang et al., 2026) on the cake example in Fig. 5. Since their implementations are not publicly available, we use the common instance shown in their papers for comparison. Both baselines produce largely coupled motion across the entire scene, failing to separate the distinct physical responses of its components. In particular, they do not reproduce the joint event in which the supporting plate fractures while the soft cream collapses. RealDrop instead models the components with different physical representations, enabling these heterogeneous responses to occur simultaneously and yielding a substantially more plausible impact outcome.

Table 4: Comparison with recent physics-aware 3D reconstruction and simulation methods, including PhysGM (Lv et al., 2026), Phys4DGen (Lin et al., 2025a), PhyMAGIC (Meng et al., 2025), GaussianFluent (Huang et al., 2026), and ReconPhys (Wang et al., 2026).

Capability	PhysGM	Phys4DGen	PhyMAGIC	GaussianFluent	ReconPhys	Ours
Single-image reconstruction	✓	✓	✓	×	×	✓
Automatic physical inference	✓	✓	✓	×	✓	✓
Heterogeneous material reconstruction	×	✓	×	NA	×	✓
Test-time physics validation	×	×	✓	×	×	✓
Fragile fracture	×	×	×	✓	×	✓

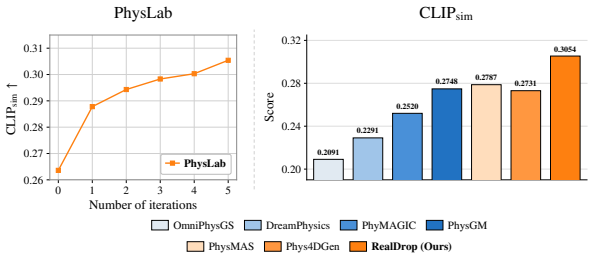

Figure 10: Left: CLIPsim across PhysLab iterations, with iteration 0 denoting the initial physical-property inference. Right: quantitative comparison with existing methods using CLIPsim. Higher is better.

D.2 In-the-Wild Generalization

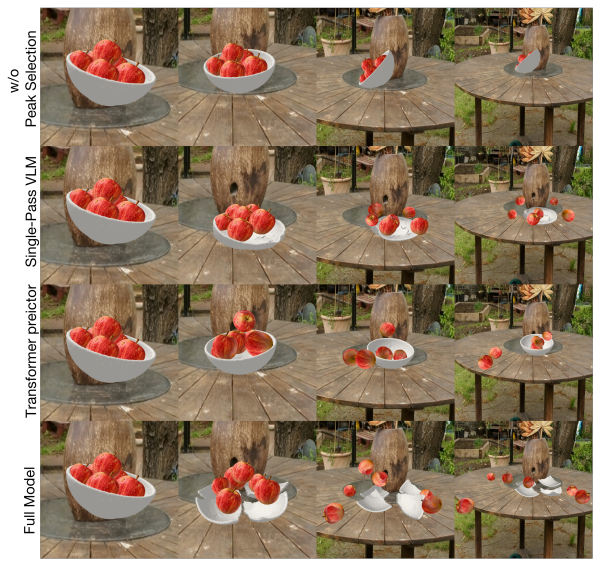

Figure 11: Ablation of part-aware modeling and physical-property inference. Without part-aware modeling, the bowl and apple share the same material properties, preventing plausible multi-material behavior. The 24-layer and the Single-Pass VLM assigns incorrect soft material. PhysLab refines these predictions through simulation feedback and recovers the desired physical behavior.

We further evaluate RealDrop on unconstrained in-the-wild single images spanning unseen object categories and material compositions. As shown in Fig. 8, RealDrop preserves distinct impact responses without object-specific tuning: brittle objects such as ceramic plates undergo coherent fracture, while heterogeneous biological objects such as watermelon break into multiple pieces with a distinct fragmentation pattern. These results demonstrate that RealDrop generalizes beyond the controlled evaluation setting while preserving material-dependent impact behavior.

D.3 Beyond Dropping: Stretching, and Bending

Although RealDrop primarily targets dropping scenarios, the reconstructed physical assets are not restricted to impact-driven motion and can respond to other external forces. As shown in Fig. 12, paper progressively deforms and tears under stretching, while a baguette bends under sustained loading. These results demonstrate that the same physical representation can express distinct material responses under different interaction modes without modifying the reconstruction pipeline, extending RealDrop beyond collision-induced dynamics.

D.4 LIMITATIONS

RealDrop remains dependent on the quality of the initial Gaussian reconstruction. Errors in geometry or part decomposition, especially in regions unobserved from the single input image, can propagate into physical simulation and are difficult to recover in the current framework. Moreover, evaluating physical realism remains an open challenge. Existing semantic metrics such as $\mathrm{CLIP}_{\mathrm{sim}}$ are not sufficiently sensitive to distinct physical outcomes, while human preference studies are inherently subjective. Developing standardized metrics that directly measure the correctness of physical events and interactions therefore remains an important direction for future work.

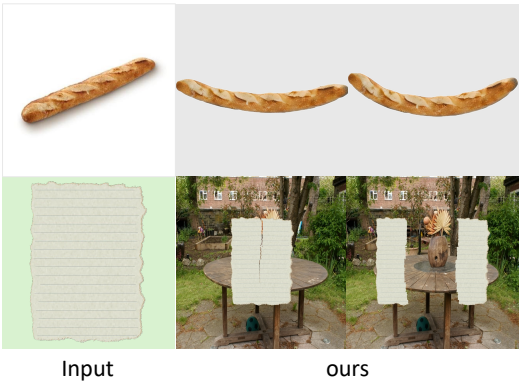

Figure 12: Physical interactions beyond dropping. RealDrop supports diverse external manipulations, including paper stretching and bending.