DIFFERENTIABLE CROSS-OBJECT CONTACT RECONSTRUCTION FROM REAL-WORLD COLLISION VIDEOS

Anonymous authors
Paper under double-blind review

ABSTRACT

Reconstructing real-world collisions from video requires identifying both how objects deform and how they exchange momentum through contact. Real-world recordings expose these coupled dynamics under occlusion, reconstruction errors, and mismatch between physical behavior and simulation models. We present an object-aware differentiable framework that jointly fits object-level material parameters and inter-object contact parameters from calibrated multi-view video. Separated Material Point Method states preserve each object's constitutive response, while contact geometry derived from reconstructed surfaces drives two-way impulse exchange. Differentiating through a finite projected contact solver connects surface and image observations to material response, friction, and optional adhesion. We also introduce Contact4D, a calibrated multiview dataset of real-world collisions between objects in free fall across diverse material pairings, with persistent instance labels and reconstructed surface observations. Comparisons with adapted reconstruction baselines show improved agreement with observed motion, geometry, and appearance. Matched ablations demonstrate the contribution of inter-object contact and the treatment of ambiguous surface observations, while numerical audits assess the consistency of contact derivatives. Together, the method and dataset support joint material and better contact reconstruction from real-world collision videos.

1 INTRODUCTION

Inferring physical properties from video enables simulation models that explain observed motion. Differentiable simulation and rendering have advanced this goal for individual objects across diverse material models (Li et al., 2023; Cai et al., 2024; Cao et al., 2024; Rho et al., 2026a). Collisions provide additional evidence: deformation reflects material response, and momentum exchange constrains how the objects influence one another. Recovering a collision therefore requires identifying the coupled dynamics of the objects and their contact.

However, real-world collisions make this coupled reconstruction challenging. Mutual occlusion during rapid impacts can leave reconstructed contact surfaces incomplete: errors in contact gaps and normals alter simulated momentum exchange, while ambiguous surface correspondences can bias parameter fitting. Synthetic benchmarks enable controlled evaluation with known physical parameters under specified simulation models, but their dynamics remain constrained by the assumptions of the underlying simulator. Real-world collision recordings provide a complementary empirical test of whether inferred material and contact models reproduce motion and deformation observed in physical experiments (Wu et al., 2016; Jiang et al., 2025). In this work, we study real-world object–object impacts initiated by free fall, which contain large occlusion and deformation, using calibrated multi-view video to jointly fit both objects' motion and contact. This setting evaluates physical reconstruction in the presence of both model mismatch and observation errors, without being biased by the simulation generator.

Our approach connects reconstructed contact geometry, two-way momentum exchange, and visual supervision in a differentiable simulation. Separated Material Point Method (MPM) (Jiang et al., 2016) states retain each object's constitutive response. A local surface field provides consistent contact gaps and normals, and a finite projected solve exchanges normal and frictional impulses, with support for rigid objects and optional interfacial adhesion. To handle reconstruction errors

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.

and occlusions in real-world collisions, we continuously update the surface field (Sec. 4.3) and use predictive penetration checks (Sec.4.4) to detect contact before interpenetration occurs. Both operations are differentiable and integrated into the simulation. Differentiating through the executed updates then enables joint estimation of object-level material parameters and inter-object contact parameters, including friction and adhesion strength. The fitted parameters describe a coupled model under the prescribed material families and contact laws.

In this work, we introduce a framework for reconstructing real-world multi-object collisions from multi-view videos. We also present Contact4D, to the best of our knowledge, the first real-world video dataset of collisions between multiple objects in free fall. Our extensive evaluation reveals how key design choices, including the contact operator, the treatment of uncertain surface observations, and the gradients used for inverse fitting, shape reconstruction quality. Our contributions are:

- An object-aware formulation for reconstructing real-world collisions, combining reconstructed-surface contact, two-way projected impulses, and finite-unroll differentiation with geometric and image supervision.
- Contact4D, a calibrated multi-view dataset of real-world two-object collisions across material-pair types, with persistent instance labels and reconstructed surface observations.
- An evaluation of coupled reconstruction through comparisons on real-world scenes, matched contact and observation ablations, and controlled inverse-fitting and numerical tests of contact derivatives.

2 Related Work

2.1 Visual Inverse Physics and Multi-Object Reconstruction

Differentiable simulation and rendering enable recovery of geometry and physical properties from video across radiance-field, Gaussian, spring–mass, and continuum representations (Li et al., 2023; Cai et al., 2024; Rho et al., 2026a; Zhong et al., 2024; Jiang et al., 2025). Recent methods further broaden material and observation models through learned constitutive components, monocular reconstruction, and spatially varying responses (Cao et al., 2024; Zhao et al., 2025; Rho et al., 2026b; Deng et al., 2026). For interacting scenes, CoupNeRF models coupled multi-material dynamics (Li et al., 2024), while MOSIV estimates per-object physical parameters from multi-object videos and introduces a synthetic contact-rich benchmark (Liu et al., 2026). We instead focus on collisions captured by calibrated real-world multiview video, where uncertain contact geometry and surface observations must be handled explicitly.

2.2 Differentiable Contact and Coupled Simulation

MPM supports large deformation through a hybrid particle–grid discretization (Jiang et al., 2016). Extensions enable two-way rigid coupling (Hu et al., 2018) and end-to-end differentiation for inverse design and control (Hu et al., 2019; 2020). Differentiable contact has been formulated using smooth analytical forces (Geilinger et al., 2020), coupled MPM–rigid–cloth systems (Liu et al., 2023), convex frictional contact (Han et al., 2023; Zong et al., 2024), asynchronous GPU coupling (Yu et al., 2025), and global complementarity over contact impulses (Ménager & Carpentier, 2026). AS-DiffMPM further incorporates differentiable rigid contact into visual reconstruction (Vasile et al., 2025).

Related finite-element methods couple rigid and deformable bodies with frictional contact (Fernández-Fernández et al., 2024), differentiate contact dynamics through adjoint formulations (Huang et al., 2024), and study gradient stability across contact regimes (Zeng et al., 2026). We instead differentiate through a finite projected contact solver for parameter fitting from reconstructed surfaces and images.

2.3 Dynamic Gaussians and Physical Scene Representations

3D Gaussian Splatting provides an explicit, differentiable scene representation (Kerbl et al., 2023), with dynamic extensions that track primitives over time or model their deformation (Luiten et al.,

2024; Yang et al., 2024). These methods reconstruct motion from visual evidence without requiring a physical simulator.

Physical Gaussian methods instead couple rendered primitives to simulated states. Prior work combines Gaussians with continuum mechanics, spring–mass models, particle-based simulation, and material-conditioned dynamics (Xie et al., 2024; Zhong et al., 2024; Cao et al., 2024; Cai et al., 2024; Jiang et al., 2025; Lin et al., 2025). Our pipeline uses fixed particle–Gaussian bindings and appearance attributes to render simulated motion, enabling coupled multi-object fitting from geometric and image supervision.

3 PRELIMINARIES

3.1 MATERIAL POINT METHOD

The Material Point Method (MPM) evolves particles using a background grid (Jiang et al., 2016; Hu et al., 2018). Each substep transfers particle mass and momentum to the grid (P2G), updates grid velocities using material stresses, gravity, boundaries, and contact, and transfers the result back to particles (G2P). For particle p and grid node j at substep n, with interpolation weight w_{jp}, representative transfers are

$$m_j^g = \sum_p w_{jp} m_p, \qquad v_p^{n+1} = \sum_j w_{jp} v_j^{g,n+1}.$$

Here m_p and m_j^g are particle and grid-node masses, and v_p^{n+1} and $v_j^{g,n+1}$ are their updated velocities. G2P also advances particle positions and deformation states. We retain separate particle states and grid velocities for each object, with contact exchanging momentum before G2P (Section 4.4).

3.2 DIFFERENTIABLE GAUSSIAN RENDERING

3D Gaussian Splatting alpha-composites projected Gaussians in depth order (Kerbl et al., 2023). Suppressing pixel and camera indices, the rendered color is

$$\hat{I} = \sum_{i \in \mathcal{V}} T_i \alpha_i c_i, \qquad T_i = \prod_{k<i} (1 - \alpha_k).$$

Here T_i is the transmittance reaching Gaussian i after the preceding Gaussians, while α_i and c_i are its projected opacity and color. We bind Gaussians to simulated particles and fix their appearance during physical fitting; differentiable rendering then connects particle motion to image residuals.

4 METHOD

We reconstruct the motion and deformation of two colliding objects from calibrated real-world multiview video by jointly fitting object-level material parameters and inter-object contact parameters. Labeled visual reconstruction provides surface observations and initial particle states; a differentiable simulator evolves these states through two-way contact, and particle-bound Gaussians render the resulting motion.

Let

$$\mathcal{I} = \{I_{t,c} \mid t = 0, \ldots, T-1,\ c = 1, \ldots, C\}$$

denote the input images over T frames and C views, with calibrated camera projections $\{\Pi_c\}_{c=1}^C$ and per-object masks $\mathcal{M}$. For object labels $a \in \{1, 2\}$, visual reconstruction supplies $\mathcal{Q} = \{Q_{t,a}\}_{t,a}$, where $Q_{t,a}$ is the reconstructed surface of object a at frame t. The observation set is $\mathcal{O} = (\mathcal{I}, \mathcal{M}, \mathcal{Q})$. The initial state X_0 contains the first-frame particle geometry, object labels, bound Gaussian appearance, and initial kinematics. We fit the physical parameters θ through

$$\mathcal{X}_\theta = \mathcal{S}_\theta(X_0), \qquad \theta^* = \arg\min_\theta \mathcal{L}_{\text{obs}}\left((\mathcal{X}_\theta, \mathcal{R}(\mathcal{X}_\theta)), \mathcal{O}\right).$$

Here $\mathcal{X}_\theta$ is the coupled simulated trajectory. Its surfaces enter the geometric loss, while $\mathcal{R}(\mathcal{X}_\theta)$ supplies images and masks for rendering losses (Section 4.5). Figure 1 summarizes observation preparation, simulation, and parameter fitting.

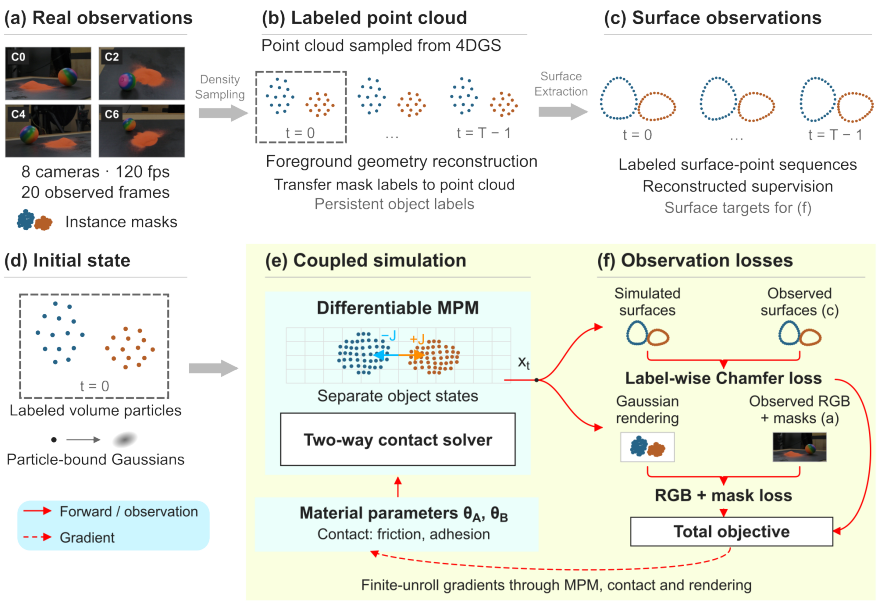

Figure 1: **Pipeline for fitting real-world collisions.** (a–c) Multi-view RGB observations are reconstructed into labeled point clouds and per-frame surface observations. (d–e) Labeled particles initialize a differentiable MPM simulation with separate object states and two-way contact. (f) Surface, RGB, and mask losses supervise material and contact parameters through finite-unroll gradients. Solid and dashed red arrows denote forward flow and gradients, respectively.

4.1 Reconstruction Observations and Initialization

From calibrated multiview images, we reconstruct 4D Gaussians and transfer persistent object labels from image masks (Figure 1(a,b)). Following GIC's density-grid particle filling (Cai et al., 2024), we extract labeled first-frame volumetric particles for simulation and labeled surfaces $Q_{t,a}$ at each frame for supervision (Figure 1(c,d)).

For continuum object a, each labeled MPM particle p carries position x_p, deformation gradient F_p, and reference volume V_p, alongside velocity and mass. The objects retain separate grid velocities $v_j^{g,a}$ until contact exchanges momentum between them. Each object has a prescribed Neo-Hookean, Kelvin–Voigt, von Mises, or Drucker–Prager material family and one spatially homogeneous parameter set (Jiang et al., 2016; Stomakhin et al., 2013; Ram et al., 2015; Klár et al., 2016).

An exact rigid object instead uses one state $q_r = (c_r, R_r, v_r, \omega_r)$ for center of mass, rotation, and linear and angular velocities. Its samples follow the shared pose; total mass m_r and world-space inertia I_w are computed from the initial volumetric filling. Appendix B gives the full particle state and rigid kinematics.

4.2 Particle-Bound Gaussian Observation Model

Gaussians bound to first-frame particles render the simulated trajectory. Their centers follow continuum particles or the shared rigid pose, while fitted appearance remains fixed during physical optimization. The resulting images and masks define the rendering losses in Section 4.5; Appendix B gives the binding and appearance details.

4.3 Contact Geometry Representation on Reconstructed Surfaces

Surface field under deformation. The first-frame density-grid reconstruction supplies volumetric particles and a set of boundary-particle indices $\mathcal{B}_b^0$ for each object b. At simulation substep n, we

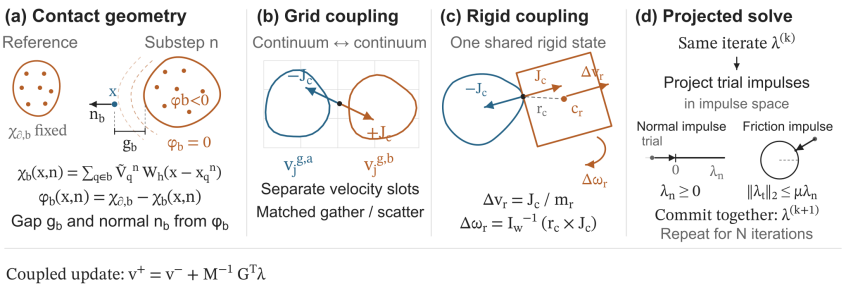

Figure 2: **Two-way contact within one simulation substep** (Figure 1(e)). (a) Surface geometry defines contact gaps and normals. (b) Separate continuum grids exchange opposing impulses. (c) Rigid coupling converts off-center impulses into translation and rotation. (d) A projected solve enforces normal and friction constraints. Bottom: coupled velocity update and differentiation through the executed iterations.

evaluate a local density-like field from the current particle positions and volume weights:

$$\chi_b(x,n) = \sum_{q\in b} \widetilde{V}_q^n W_h\left(x - x_q^n\right),$$
$$\widetilde{V}_q^n = V_q \max\left(\det F_q^n, J_{\min}\right).$$

Here q indexes particles of object b, $\widetilde{V}_q^n$ is the current volume weight, V_q is the reference volume, $J_{\min} > 0$ is a numerical volume floor, and W_h is a compactly supported Wendland C^2 kernel with fixed scale h. We set the boundary isovalue $\chi_{\partial,b}$ to the mean initial field value at the prepared boundary particles and retain it throughout the simulation. The field is reevaluated at each substep, so its level sets follow changes in particle positions and deformation. For a rigid object, particle positions follow the shared pose in Section 4.1, and their volume weights remain equal to their reference volumes.

Consistent gap and normal. The current implicit boundary is the zero level set of

$$\phi_b(x,n) = \chi_{\partial,b} - \chi_b(x,n).$$

Wherever the spatial gradient is valid, we evaluate a first-order signed-gap estimate and outward normal from this same field:

$$g_b(x,n) = \frac{\phi_b(x,n)}{\|\nabla_x \phi_b(x,n)\|_2}, \qquad n_b(x,n) = \frac{\nabla_x \phi_b(x,n)}{\|\nabla_x \phi_b(x,n)\|_2}.$$

Positive $g_b(x,n)$ denotes separation and negative values denote overlap. The gap controls contact activation and the normal sets the impulse direction. Both come from the current field as the objects deform. We suppress the substep argument below for readability.

Surface queries. The prepared boundary indices define a persistent material surface whose particle positions evolve with the simulation. In the prepared-particle initialization, these particle centers serve as contact query points. When dynamic surface refresh is enabled, per-object occupancy grids and local sparsity or stretch tests propose additional candidates; local field and gradient checks retain candidates near the boundary, using reference values interpolated from the persistent surface. The resulting query set is evaluated against the other object's current field within a fixed active band. These contact candidates are separate from the prepared surface-particle indices used for geometric supervision, which are evaluated at their simulated positions. For rigid objects, all surface queries follow one shared translation and rotation, and their impulses act on the same rigid state (Section 4.4). The refresh mode, candidate budget, query band, collision clearance, and retry policy are specified in the experimental configuration. Validity checks record degenerate gradients, empty neighborhoods, and out-of-domain queries; discrete candidate selection is held fixed during each backward replay.

4.4 TWO-WAY PROJECTED CONTACT

Each candidate from Section 4.3 supplies a normal for the contact Jacobian G and a signed gap for the velocity target below. Let v collect the labeled grid velocities and any rigid generalized velocity. The Jacobian maps them to relative contact velocities $u = Gv$, while G^T scatters impulses with opposite signs to the objects. With generalized mass M and accumulated contact impulse λ,

$$v^+ = v^- + M^{-1}G^\mathsf{T}\lambda, \qquad K = GM^{-1}G^\mathsf{T}.$$

Superscripts $-$ and $+$ denote velocities before and after contact. The contact-space response K maps impulses to changes in relative velocity: $u^+ - u^- = K\lambda$. Figure 2 shows the geometry, impulse transfer, and projected solve.

Friction and adhesion. For non-adhesive contact, the normal impulse obeys a velocity-level constraint and the tangential impulse lies in a Coulomb disk:

$$0 \le \lambda_n \;\perp\; G_n v^+ - b \ge 0, \qquad \|\lambda_t\|_2 \le \mu\lambda_n.$$

Here λ_n and λ_t are the normal and tangential components of λ, G_n is the normal row of G, b is the normal-velocity target, and μ is the pairwise friction coefficient; $\perp$ means the two nonnegative terms have zero product. For selected pairs, adhesion permits a bounded tensile impulse over a fixed range, with strength that may be fitted. Appendix B gives the adhesive projection.

Predictive nonpenetration. For signed gap g, post-contact normal velocity $v_n^+ = G_n v^+$, and substep duration Δt, we predict $g^+ \approx g + \Delta t v_n^+$. With zero restitution, we set

$$b(g) = \begin{cases} -\max(g - d_{\min}, 0)/\Delta t, & g > 0, \\ 0, & g \le 0, \end{cases}$$

where $d_{\min}$ is the collision clearance. For $g > d_{\min}$, the target allows approach toward the clearance; for $g \le d_{\min}$, including overlap, the zero target suppresses further first-order closing. The gap comes from the current surface field, so the target follows the simulated deformation.

We use a fixed number of synchronous projected-Jacobi iterations. Writing $\lambda^{(k)}$ for the accumulated impulse at iteration k, each iteration reads the same iterate, projects trial normal and tangential impulses, and commits the updates as $\lambda^{(k+1)}$. Contacts sharing grid nodes or a rigid state are coupled through K; the final iterate defines the contact update differentiated during fitting.

For continuum pairs, G gathers velocities from the separate labeled grids and G^T scatters equal-and-opposite impulses with matched weights. For a rigid contact at x_c, let $r_c = x_c - c_r$ and let J_c act on the rigid object. Its contact-point velocity and shared-state update are

$$v_c^r = v_r + \omega_r \times r_c, \qquad \Delta v_r \;\; = J_c/m_r, \qquad \Delta\omega_r = I_w^{-1}(r_c \times J_c).$$

Here Δv_r and $\Delta\omega_r$ are changes in rigid linear and angular velocity. The continuum grid receives $-J_c$, and all rigid contacts update the same translation and rotation. Appendix B gives the gather, scatter, and iteration details.

Time integration and differentiation. Each video-frame interval uses MPM substeps with labeled P2G, grid and contact updates, G2P, and rigid-pose integration when applicable. We backpropagate through the executed contact iterations and remaining simulation updates, holding contact-candidate and topology choices fixed within each backward replay. Appendix B details the substeps and backward replay; Appendix D reports derivative checks.

4.5 OBSERVATION OBJECTIVE

At each observed frame, we compare simulated surface samples $P_{t,a}$ with reconstructed targets $Q_{t,a}$ using a label-wise symmetric Chamfer loss. We exclude ambiguous samples near the cross-label contact seam from surface correspondences while retaining their particles in the simulation and contact solve. On the remaining supervision sets, the geometric loss is

$$\mathcal{L}_{\text{geo}} = \sum_{t,a} w_a \left[\frac{1}{|P_{t,a}|} \sum_{p \in P_{t,a}} \min_{q \in Q_{t,a}} \|x_p - q\|_2 \;\; + \frac{1}{|Q_{t,a}|} \sum_{q \in Q_{t,a}} \min_{p \in P_{t,a}} \|q - x_p\|_2 \right].$$

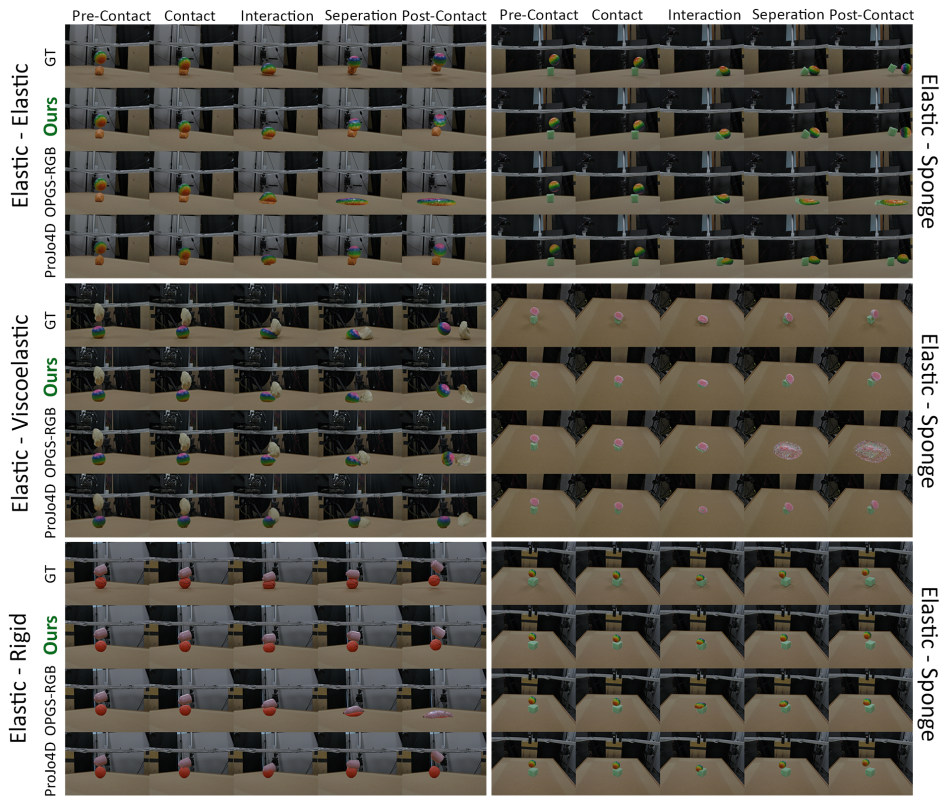

Figure 3: Selected Contact4D collisions from approach to late response. Rows compare observed RGB, Ours, OPGS-RGB, and ProJo4D-adapted, whose single-object optimizer is run independently for each object.

Here q indexes target surface points, and w_a balances objects with different surface areas. The foreground image loss $\mathcal{L}_{\text{rgb}}$ combines RGB L_1 and structural similarity (SSIM); $\mathcal{L}_{\text{mask}}$ compares rendered and observed masks. The observation objective is

$$\mathcal{L}_{\text{obs}} = w_g \mathcal{L}_{\text{geo}} + w_i \mathcal{L}_{\text{rgb}} + w_m \mathcal{L}_{\text{mask}}.$$

We estimate initial object velocities before fitting material and contact parameters. Appendix B specifies the losses and optimization protocol; Section 5.5 compares observation choices.

5 Experiments

We compare full reconstructions on Contact4D, then isolate pair contact and surface supervision through matched ablations.

5.1 Contact4D and Evaluation Protocol

Contact4D. Contact4D contains calibrated multiview recordings of real-world two-object free-fall collisions across material pairs, including rigid-object interactions. RGB views and instance masks provide persistent labels; we reconstruct per-frame surfaces and initial volumetric particles (Section 4.1). Appendix A gives the capture setup (Figure 5) and evaluation protocol.

Metrics. Worst-object IoU is the lower of the two objects' mean visible-mask IoUs. Centroid error is normalized by the image diagonal; surface CD is label-wise symmetric Chamfer-L_1 distance normalized by the initial union-surface diagonal. Foreground PSNR and LPIPS measure rendering quality under the same cameras and masks. Post-contact measurements are summarized as mean $\pm$ standard error (SEM) across scenes.

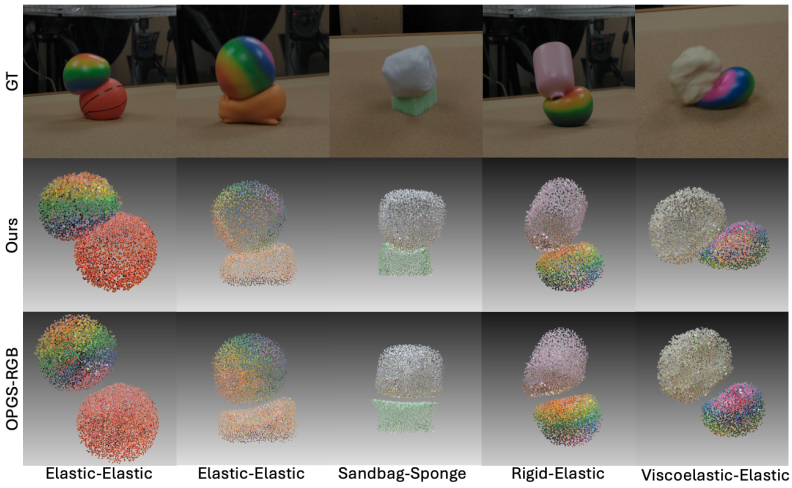

Figure 4: Contact at impact. Rows show observed RGB, our predicted point clouds, and OPGS-RGB point clouds. The point clouds use a different viewpoint to expose the contact region.

Quantitative baseline. OPGS-RGB is a Contact4D reimplementation of OmniPhysGS's shared MPM core with RGB L_1+DSSIM supervision and a predeclared scene-specific expanded physical domain; it has no projected pair-contact solver (Lin et al., 2025). The comparison shares initial-motion estimates and evaluation settings. We found no public official code for CoupNeRF or MO-SIV at the time of writing (Li et al., 2024; Liu et al., 2026); their results are therefore outside the Contact4D comparison.

5.2 Reconstructing Real-world Collisions

Table 1 tests post-impact agreement with observed motion, surfaces, and images. It reports Ours on full Contact4D and OPGS-RGB on its accepted subset; failure rates cover all attempted runs. The comparison below uses Ours scores on the same accepted subset.

Table 1: Post-contact reconstruction. Ours is reported on full Contact4D and OPGS-RGB on its accepted subset. Values are scene mean $\pm$ SEM on the stated sets. Failure rates cover all attempted runs; OPGS-RGB failures are physical-domain audit rejections. Matched Ours comparisons are given in the text.

Scene set	Method	IoU ↑	Centroid ↓	Surface CD ↓	PSNR ↑	LPIPS ↓	Failure rate ↓
Full Contact4D	Ours	0.6206 ± 0.0306	0.00912 ± 0.00110	0.02496 ± 0.00346	14.076 ± 0.268	0.3427 ± 0.0163	**0%**
Accepted subset	OPGS-RGB	0.2998 ± 0.0339	0.03128 ± 0.00380	0.08563 ± 0.01309	12.363 ± 0.325	0.4884 ± 0.0122	15%

Ours yields accepted reconstructions throughout Contact4D, with mean IoU 0.6206, surface CD 0.02496, and a 0% failure rate. OPGS-RGB completes its runs, but 15% of predicted trajectories fail the numerical divergence audit of its predeclared simulation domain. On its accepted subset, Ours has higher mean IoU and lower surface CD.

5.3 Qualitative Results

Figure 3 compares observed and rendered sequences at matched times. ProJo4D-adapted serves as a single-object physics-guided baseline by optimizing each object independently and combining the renders without cross-object contact (Rho et al., 2026a).

Figure 4 focuses on impact. OPGS-RGB shows visible gaps, while ours maintains contact without visible interpenetration in these examples. Its shared MPM grid may mix object momentum before reconstructed surfaces meet.

5.4 Contact Response and Post-Impact Motion

To test the role of inter-object momentum exchange, Table 2 compares projected contact, smooth-penalty contact, and no pair contact within our pipeline. Initial states, fitting, and evaluation are matched; ground contact remains active in every variant.

Projected contact enforces normal and Coulomb bounds on accumulated impulses. Smooth penalty instead uses smoothed penetration and closing-velocity damping; both use the same contact geometry and two-way impulse transfer. Appendix C gives the penalty update.

Table 2: Contact-operator ablation on a matched scene set. Values are scene mean $\pm$ SEM; penetration is $P_{95}/\Delta x$ and is undefined without pair contact.

Variant	IoU ↑	Centroid ↓	Surface CD ↓	Penetration ↓	PSNR ↑	LPIPS ↓
Projected contact	$\mathbf{0.6094 \pm 0.0249}$	$\mathbf{0.00809 \pm 0.00054}$	$\mathbf{0.02310 \pm 0.00171}$	$\mathbf{0.00570 \pm 0.00326}$	$\mathbf{15.0820 \pm 0.2893}$	$\mathbf{0.3462 \pm 0.0114}$
Smooth penalty	0.5884 ± 0.0260	0.00883 ± 0.00061	0.02449 ± 0.00158	0.02232 ± 0.00236	15.0407 ± 0.2993	0.3501 ± 0.0121
No pair contact	0.3832 ± 0.0314	0.02047 ± 0.00295	0.05477 ± 0.00767	–	13.8530 ± 0.3909	0.4378 ± 0.0166

Penetration is the post-contact 95th percentile of $\max(0, -g_{\min})/\Delta x$, where $g_{\min}$ is the minimum candidate gap and Δx is grid spacing; it is undefined without pair contact.

Removing pair contact lowers mean IoU from 0.6094 to 0.3832 and increases centroid and surface errors, showing the contribution of coupled impulses to post-impact fitting. Smooth penalty recovers much of this gain. Projection gives modestly better mean reconstruction and lowers normalized penetration from 0.02232 to 0.00570 while specifying normal and friction constraints directly. The shared-rigid subset in Table 6 shows similar reconstruction for the two contact solvers.

5.5 Supervision Near the Contact Interface

Table 3 tests the effect of uncertain surface correspondences near contact by comparing filtered and unfiltered Chamfer supervision with rendering, along with geometry-only and rendering-only variants. All use projected contact and matched fitting; filtering removes only supervision correspondences, not particles from the dynamics.

Table 3: Observation-loss ablation with projected contact on a matched scene set. Variants share initialization and fitting settings; values are scene mean $\pm$ SEM. Filtering removes surface correspondences while retaining their particles in the dynamics.

Supervision	IoU ↑	Centroid ↓	Surface CD ↓	PSNR ↑	LPIPS ↓
Filtered CD + render	$\mathbf{0.6127 \pm 0.0244}$	$\mathbf{0.00858 \pm 0.00062}$	$\mathbf{0.02321 \pm 0.00171}$	$\mathbf{15.0593 \pm 0.2957}$	$\mathbf{0.3510 \pm 0.0121}$
All-surface CD + render	0.5934 ± 0.0254	0.00932 ± 0.00062	0.02468 ± 0.00160	14.9600 ± 0.3074	0.3554 ± 0.0126
All-surface CD only	0.5953 ± 0.0255	0.00957 ± 0.00069	0.02420 ± 0.00167	14.8951 ± 0.3059	0.3594 ± 0.0132
Render only	0.5931 ± 0.0252	0.00937 ± 0.00062	0.02502 ± 0.00162	14.9473 ± 0.3086	0.3558 ± 0.0127

Filtered Chamfer plus rendering performs best on all mean metrics, improving IoU from 0.5934 to 0.6127 and reducing surface CD from 0.02468 to 0.02321 over the unfiltered variant. Other combinations remain close, indicating that the clearest gain comes from removing ambiguous contact-seam correspondences.

Numerical validation. Appendix D reports contact residuals, invariant checks, and derivative comparisons for the executed contact map.

6 Conclusion and Limitations

We presented a differentiable framework and Contact4D for reconstructing real-world multi-object free-fall collisions by jointly estimating object-level material and inter-object contact parameters. Experiments and ablations show that modeling inter-object contact and uncertain surface observations improves agreement with captured motion, geometry, and appearance, while numerical audits support the contact gradients. Our method relies on reconstructed contact geometry and prescribed spatially homogeneous material families, so reconstruction errors or model mismatch can affect the fitted parameters. Future work could jointly refine geometry and physical parameters, infer material families from video, and model spatially varying properties.

REFERENCES

Junhao Cai, Yuji Yang, Weihao Yuan, Yisheng He, Zilong Dong, Liefeng Bo, Hui Cheng, and Qifeng Chen. GIC: Gaussian-informed continuum for physical property identification and simulation. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Junyi Cao, Shanyan Guan, Yanhao Ge, Wei Li, Xiaokang Yang, and Chao Ma. NeuMA: Neural material adaptor for visual grounding of intrinsic dynamics. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Yinan Deng, Jianqiao Song, Yisi Zhang, Yuhan Wang, Jiahui Wang, and Yufeng Yue. PhysReal: Learning real-world deformable object physics via hybrid constitutive modeling. *arXiv preprint arXiv:2609.07532*, 2026.

José Antonio Fernández-Fernández, Ralph Lange, Stefan Laible, Kai O. Arras, and Jan Bender. STARK: A unified framework for strongly coupled simulation of rigid and deformable bodies with frictional contact. In *IEEE International Conference on Robotics and Automation*, pp. 16888–16894, 2024. doi: 10.1109/ICRA57147.2024.10610574.

Moritz Geilinger, David Hahn, Jonas Zehnder, Moritz Bächer, Bernhard Thomaszewski, and Stelian Coros. ADD: Analytically differentiable dynamics for multi-body systems with frictional contact. *ACM Transactions on Graphics*, 39(6), 2020. doi: 10.1145/3414685.3417766.

Xuchen Han, Joseph Masterjohn, and Alejandro Castro. A convex formulation of frictional contact between rigid and deformable bodies. *IEEE Robotics and Automation Letters*, 2023.

Yuanming Hu, Yu Fang, Ziheng Ge, Ziyin Qu, Yixin Zhu, Andre Pradhana, and Chenfanfu Jiang. A moving least squares material point method with displacement discontinuity and two-way rigid body coupling. *ACM Transactions on Graphics*, 37(4):150:1–150:14, 2018. doi: 10.1145/3197517.3201293.

Yuanming Hu, Jiancheng Liu, Andrew Spielberg, Joshua B. Tenenbaum, William T. Freeman, Jiajun Wu, Daniela Rus, and Wojciech Matusik. ChainQueen: A real-time differentiable physical simulator for soft robotics. In *IEEE International Conference on Robotics and Automation*, 2019.

Yuanming Hu, Luke Anderson, Tzu-Mao Li, Qi Sun, Nathan Carr, Jonathan Ragan-Kelley, and Frédo Durand. DiffTaichi: Differentiable programming for physical simulation. In *International Conference on Learning Representations*, 2020.

Zizhou Huang, Davi Colli Tozoni, Arvi Gjoka, Zachary Ferguson, Teseo Schneider, Daniele Panozzo, and Denis Zorin. Differentiable solver for time-dependent deformation problems with contact. *ACM Transactions on Graphics*, 43(3):31:1–31:30, 2024. doi: 10.1145/3657648.

Chenfanfu Jiang, Craig Schroeder, Joseph Teran, Alexey Stomakhin, and Andrew Selle. The material point method for simulating continuum materials. In *ACM SIGGRAPH 2016 Courses*, pp. 24:1–24:52, 2016. doi: 10.1145/2897826.2927348.

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. *arXiv preprint arXiv:2503.17973*, 2025.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):139:1–139:14, 2023. doi: 10.1145/3592433.

Gergely Klár, Theodore Gast, Andre Pradhana, Chuyuan Fu, Craig Schroeder, Chenfanfu Jiang, and Joseph Teran. Drucker–prager elastoplasticity for sand animation. *ACM Transactions on Graphics*, 35(4):103:1–103:12, 2016. doi: 10.1145/2897824.2925906.

Jin Li, Yang Gao, Wenfeng Song, Yacong Li, Shuai Li, Aimin Hao, and Hong Qin. CoupNeRF: Property-aware neural radiance fields for multi-material coupled scenario reconstruction. *Computer Graphics Forum*, 43(7):e15208, 2024. doi: 10.1111/cgf.15208.

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. In *International Conference on Learning Representations*, 2023.

Yuchen Lin, Chenguo Lin, Jianjin Xu, and Yadong Mu. OmniPhysGS: 3d constitutive gaussians for general physics-based dynamics generation. *arXiv preprint arXiv:2501.18982*, 2025.

Chunjiang Liu, Xiaoyuan Wang, Qingran Lin, Albert Xiao, Haoyu Chen, Shizheng Wen, Hao Zhang, Lu Qi, Ming-Hsuan Yang, Laszlo A. Jeni, Min Xu, and Yizhou Zhao. MOSIV: Multi-object system identification from videos. *arXiv preprint arXiv:2603.06022*, 2026.

Min Liu, Gang Yang, Siyuan Luo, and Lin Shao. SoftMAC: Differentiable soft body simulation with forecast-based contact model and two-way coupling with articulated rigid bodies and clothes. *arXiv preprint arXiv:2312.03297*, 2023.

Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. In *International Conference on 3D Vision*, pp. 800–809, 2024.

Etienne Ménager and Justin Carpentier. Frictional contact solving for material point method, 2026. Preprint, HAL-05488708.

Daniel Ram, Theodore Gast, Chenfanfu Jiang, Craig Schroeder, Alexey Stomakhin, Joseph Teran, and Pirouz Kavehpour. A material point method for viscoelastic fluids, foams and sponges. In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 2015.

Daniel Rho, Jun Myeong Choi, Biswadip Dey, and Roni Sengupta. ProJo4D: Progressive joint optimization for sparse-view inverse physics estimation. *Transactions on Machine Learning Research*, 2026a.

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*, 2026b.

Alexey Stomakhin, Craig Schroeder, Lawrence Chai, Joseph Teran, and Andrew Selle. A material point method for snow simulation. *ACM Transactions on Graphics*, 32(4):102:1–102:10, 2013. doi: 10.1145/2461912.2461948.

Federico Vasile, Ri-Zhao Qiu, Lorenzo Natale, and Xiaolong Wang. Gaussian-augmented physics simulation and identification with complex colliders. In *Advances in Neural Information Processing Systems*, 2025.

Jiajun Wu, Joseph J. Lim, Hongyi Zhang, Joshua B. Tenenbaum, and William T. Freeman. Physics 101: Learning physical object properties from unlabeled videos. In *British Machine Vision Conference*, 2016.

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

Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3d gaussians for high-fidelity monocular dynamic scene reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

Chang Yu, Wenxin Du, Zeshun Zong, Alejandro Castro, Chenfanfu Jiang, and Xuchen Han. A convex formulation of material points and rigid bodies with gpu-accelerated async-coupling for interactive simulation. *arXiv preprint arXiv:2503.05046*, 2025.

Ziqiu Zeng, Gang Yang, Zhenhao Huang, Yulin Li, Jason Pho, Siyuan Luo, and Fan Shi. Fast and reliable gradients for deformables across frictional contact regimes. *arXiv preprint arXiv:2603.16478*, 2026.

Yizhou Zhao, Haoyu Chen, Chunjiang Liu, Zhenyang Li, Charles Herrmann, Junhwa Hur, Yinxiao Li, Ming-Hsuan Yang, Bhiksha Raj, and Min Xu. Toward material-agnostic system identification from videos. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5944–5956, 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*, 2024.

Zeshun Zong, Chenfanfu Jiang, and Xuchen Han. A convex formulation of frictional contact for the material point method and rigid bodies. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1831–1838, 2024. doi: 10.1109/IROS58592.2024.10801598.