# Learning Implicit Constitutive Laws for Dynamic 3D Gaussian Splatting from Monocular Videos via Visual-Physical Alignment

Xiaoyang Liu[1] and Kai Han[1⋆]

Visual AI Lab, The University of Hong Kong, Hong Kong
xiaoyangliu@connect.hku.hk, kaihanx@hku.hk

**Abstract.** We present GCA (**G**aussian **C**onstitutive **A**lignment), a framework for learning implicit constitutive laws from monocular dynamic video of deformable objects represented by 3D Gaussians. Given a static multi-view scan for geometric initialization, our method learns intrinsic physical dynamics solely from a single fixed-viewpoint video of the moving object. Existing implicit methods often suffer from local minima under noisy supervision and lack physical interpretability, while explicit approaches rely on predefined constitutive equations, limiting generalizability and becoming unstable in monocular settings. To address these challenges, our framework unifies LoRA-based adaptation with two key alignment modules. First, we propose *Rank-based Depth-Geometric Anchors* (RDGA) to establish robust geometric constraints from monocular dynamic observations via scale-invariant rank-based depth alignment, reducing the reliance on unreliable pixel-level color supervision. Second, a *Constitutive Prior Regularizer* (CPR) integrates classical constitutive models as soft differentiable priors, regularizing the optimization while preserving the flexibility of implicit modeling—even when the actual material is absent from the hypotheses. Extensive experiments on synthetic, real-to-sim, and real-world datasets demonstrate that GCA outperforms existing methods, achieving 48% lower Chamfer Distance than the strongest baseline on synthetic benchmarks while remaining robust under monocular supervision.

**Keywords:** Constitutive laws · Differentiable physics · Monocular video · Visual-physical alignment

## 1 Introduction

Understanding the intrinsic dynamics of objects is crucial for spatial intelligence, enabling accurate digital modeling, interaction, and manipulation that follows physical laws [3, 26, 40, 53]. While humans effortlessly infer basic physical properties from videos (*e.g.*, bouncing balls or viscous fluid flow), extracting precise physical models from visual signals remains an open challenge.

⋆ Corresponding author.

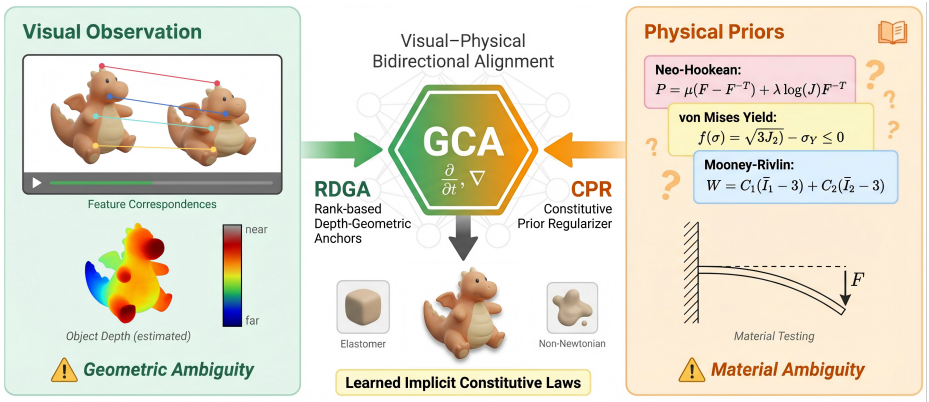



**Fig. 1: The core idea of GCA.** Learning implicit constitutive laws from dynamic video faces two key challenges: *geometric ambiguity* from sparse visual observations (left), and *material ambiguity* from unknown constitutive models and parameters (right). GCA bridges both sides through visual–physical bidirectional alignment: Rank-based Depth-Geometric Anchors (RDGA) establish robust geometric constraints via scale-invariant depth consensus, while a Constitutive Prior Regularizer (CPR) softly guides optimization using classical constitutive hypotheses. Together, they enable reliable learning of implicit constitutive laws (bottom) that generalize across diverse material types.

Prior works have employed models to understand intrinsic dynamics from visual observations [8, 23–25, 35]. A common approach combines differentiable physics simulators [13, 51] with differentiable renderers [28, 38] for optimization. Regarding constitutive law modeling, existing approaches fall into two paradigms with complementary limitations: *Explicit modeling* [24, 30, 34, 55] builds upon classical continuum mechanics with predefined constitutive models [12, 39] (*e.g.*, hyperelastic models [46]) and explicit parameters such as Young's modulus and Poisson's ratio. While enabling physical interpretability, effectiveness critically depends on correct model specification: (1) manual model selection and fine parameter tuning are required for different materials [30]; (2) complex real-world materials with deeply coupled properties remain intractable [34]; and critically, (3) these methods perform poorly under monocular supervision, as observed in our experiments (Tab. 5). *Implicit parameterization* models constitutive relations through neural networks [29, 37]. NeuMA [7] introduces the first method to align implicit constitutive models with visual observations. However, implicit models easily converge to suboptimal solutions when handling noisy and sparse supervision [49]—a problem that becomes severe in monocular settings.

These conflicting trade-offs lead to our core question: *How to reliably learn intrinsic dynamics from monocular video while preserving the generalization advantages of implicit modeling?* Following prior work in dynamic 3D Gaussian splatting [7, 56], we assume access to a static multi-view orbital scan for initial 3D geometry reconstruction. The focus of this work is to learn intrinsic constitu-

tive laws from a single fixed-viewpoint video of the object undergoing dynamic deformation. This monocular dynamic setting introduces severe geometric ambiguities (single viewpoint) and material ambiguities (underconstrained physical parameters) that make the inverse problem significantly ill-posed.

To address this challenge, we propose GCA, a framework that achieves visual-physical bidirectional alignment for learning implicit constitutive laws from monocular dynamic video. Our method unifies LoRA-based adaptation with two coordinated alignment modules: (i) Rank-based Depth-Geometric Anchors (**RDGA**) extract robust geometric constraints from sparse observations using a scale-invariant rank-based depth alignment mechanism. Unlike pixel-level color supervision, which suffers from domain shifts in monocular video, RDGA is designed to mitigate the scale-shift ambiguity of monocular depth estimators. (ii) Constitutive Prior Regularizer (**CPR**) treats classical constitutive models as *soft regularization priors* rather than hard constraints, resolving material ambiguity without sacrificing generalization, and remains effective even when the true material is absent from the hypotheses.

Extensive experiments validate that GCA achieves state-of-the-art performance: 48% lower Chamfer Distance than NeuMA on synthetic benchmarks, robust performance on a challenging real-to-sim dataset, and superior visual quality in real-world monocular experiments. We also show that naively adapting explicit methods to monocular settings leads to optimization divergence, highlighting the importance of robust alignment.

## 2 Related work

**Physics-grounded dynamic 3D generation.** NeRF-based models [15,17,27, 41] are constrained by predefined material assumptions, while recent Gaussian-based methods [16, 28, 48] show substantial progress; SpringGaus [56] reconstructs elastic dynamics but still relies on an explicit spring-mass model. Diffusion-guided approaches [24, 32, 34, 55] inherit imprecise physics priors [10, 31, 42] and typically assume rigid [35] or elastic [56] bodies. NeuMA [7] first optimizes neural constitutive laws from observational images without predefined laws, but single-modality visual optimization suffers from local minima under sparse monocular supervision; naively augmenting explicit methods (*e.g.*, PAC-NeRF [30]) with monocular depth diverges, motivating our hybrid alignment.

**Material constitutive laws.** Conventional approaches [2, 5, 9, 34, 39] enforce explicit laws via predefined nonlinear polynomial bases (*e.g.*, elastic [18]/plastic [12]/fluid [9] models) and tune parameters such as Young's modulus or Poisson's ratio, requiring manual specification [33]. Implicit neural modeling [6, 36, 43] bypasses this: NCLaw [37] pioneers hybrid NN–PDE training but requires particle-level annotations; NeuMA [7] avoids particle ground truth via LoRA-based [21] alignment with differentiable rendering, yet pure visual supervision lacks physical interpretability [1] and sparse observations introduce optimization ambiguity. Our framework softly regularizes implicit optimization with physical

priors, unifying the stability of explicit laws with the generalization of implicit ones under sparse supervision.

## 3 Method

### 3.1 Problem statement

Given static 3D Gaussians [28] of an object $\mathcal{G}(i) = \{\mathbf{p}(i), \alpha(i), \mathbf{A}(i), \mathbf{c}(i)\}$, where $\mathbf{p}(i), \alpha(i), \mathbf{A}(i), \mathbf{c}(i)$ are the center, opacity, covariance matrix, and spherical harmonic coefficients of each Gaussian primitive, and its corresponding monocular dynamic video $\{I_t\}_{t=1}^{T}$, we aim to learn implicit constitutive laws through a dynamical system $\mathcal{M}_\theta$ governed by elastoplastic dynamics [19]:

$$\rho_0 \ddot{\phi} = \nabla \cdot \mathbf{P} + \rho_0 \mathbf{b}, \quad \mathbf{P} = \mathcal{E}(\mathbf{F}_e), \quad \mathbf{F}_e = \nabla \phi, \tag{1}$$

where $\mathbf{P}$ is the first Piola–Kirchhoff stress tensor, $\rho_0$ is the object density, and $\mathbf{b}$ is the body force. $\phi$ denotes the deformation map, $\ddot{\phi}$ is its acceleration, and $\mathcal{E}$ is the elastic constitutive law.

We discretize Eq. (1) and obtain the dynamical system $\mathcal{M}_\theta$:

$$\mathbf{s}_{t+1} = \mathcal{M}_\theta(\mathbf{s}_t), \quad \forall\, t = 0, 1, \ldots, T-1, \tag{2}$$

where $\mathbf{s}_t = \{\mathbf{x}_t, \mathbf{v}_t, \mathbf{F}_e^t\}$ denotes the state at time step $t$, consisting of particle positions, velocities, and elastic deformation gradients. $\theta$ denotes the neural parameters in $\mathcal{M}$. Additional details on preprocessing Gaussians to particles are provided in the supplementary material.

To align $\mathcal{M}_\theta$ with observation $I_t$, we use a differentiable 3DGS renderer $\mathcal{R}$ producing $\hat{I}_t = \mathcal{R}(\mathbf{s}_t; \mathbf{K}, \mathbf{Q})$, where $\mathbf{K}, \mathbf{Q}$ denote camera intrinsic and extrinsic matrices. Relying solely on this rendering-based supervision, however, is insufficient to overcome the challenges posed by sparse and noisy monocular video. The inherent *geometric ambiguities* from the single viewpoint and *material ambiguities* in the dynamics make the optimization landscape intractable. To establish a robust learning pipeline, we propose GCA (see Fig. 2), which combines LoRA-based adaptation with two coordinated alignment modules—Rank-based Depth-Geometric Anchors (**RDGA**) and a Constitutive Prior Regularizer (**CPR**)—detailed below.

### 3.2 Neural material constitutive laws

Our work adopts the same dynamical system $\mathcal{M}_\theta$ as NCLaw [37] for state transitions. $\mathcal{M}_\theta$ is composed of the neural elasticity law $\mathcal{E}_{\theta_e}$, semi-implicit Euler integration [22, 47], and neural plasticity law $\mathcal{P}_{\theta_p}$. We use the basic physical prior model $\mathcal{M}_0 = \{\mathcal{E}_0, \mathcal{P}_0\}$ provided by NCLaw. To align the model with observations without compromising its fundamental capabilities, we use LoRA [21] for fine-tuning instead of training all parameters. Specifically, $\mathcal{M}_\theta = \{\mathcal{E}_{\theta_e}, \mathcal{P}_{\theta_p}\}$, where $\mathcal{E}_{\theta_e} = \mathcal{E}_0 + \Delta\mathcal{E}_{\theta_e}$ and $\mathcal{P}_{\theta_p} = \mathcal{P}_0 + \Delta\mathcal{P}_{\theta_p}$.

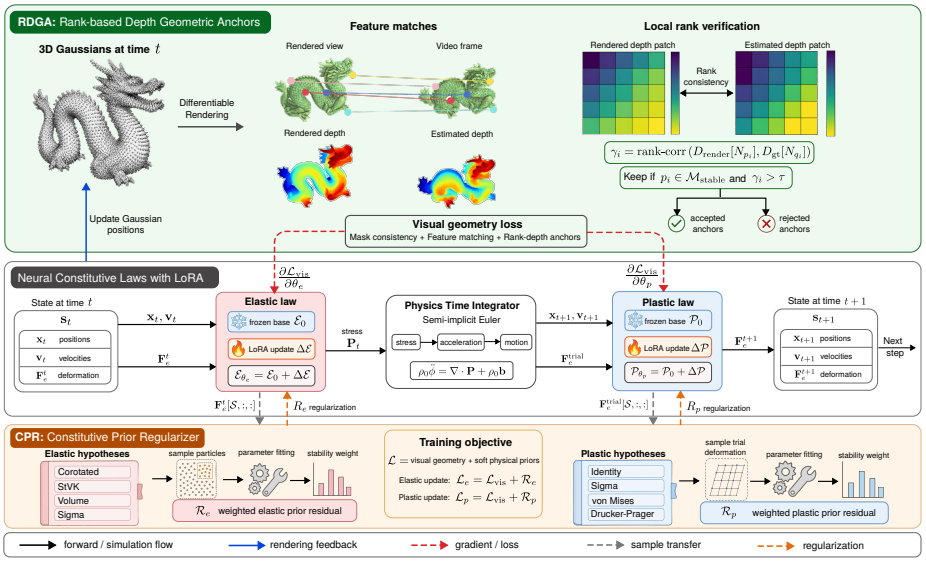



**Fig. 2: Overview of GCA.** Our framework builds on Low-Rank Adaptation (LoRA) to fine-tune neural material laws while maintaining integration with PDE-based simulation, and introduces two coordinated alignment modules: (i) Rank-based Depth-Geometric Anchors (RDGA) that resolve geometric ambiguities via scale-invariant rank-based depth alignment, and (ii) a Constitutive Prior Regularizer (CPR) that provides soft physical priors while preserving generalization.

**Remark.** While our dynamical system builds upon the NCLaw architecture [37], NCLaw itself requires *dense particle-level ground-truth supervision*, which is unavailable in visual observation settings. The core challenge addressed by GCA is not the design of neural constitutive architectures, but the *visual-physical alignment* problem: how to provide reliable supervisory signals to train such architectures from noisy, sparse, single-viewpoint video. As shown in Tab. 5, naively adapting existing methods to monocular settings leads to optimization instability, highlighting the importance of robust alignment. Moreover, as demonstrated in Tab. 4, LoRA not only improves parameter efficiency but also acts as a regularizer—full fine-tuning tends to overfit visual noise, yielding worse physical accuracy.

The dynamical system $\mathcal{M}_\theta$ advances physical states through three stages as shown in Alg. 1: (1) *Stress evaluation* via neural constitutive law $\mathcal{E}_{\theta_e}$ that computes first Piola–Kirchhoff stress $\mathbf{P}_t$ from elastic deformation gradient $\mathbf{F}_e^t$; (2) *Dynamics integration* where operator $\mathcal{I}$ implements semi-implicit Euler integration to update positions $\mathbf{x}_{t+1}$ and velocities $\mathbf{v}_{t+1}$; (3) *Plasticity update* through network $\mathcal{P}_{\theta_p}$ that modifies $\mathbf{F}_e^{\text{trial}}$ to account for plastic deformation.

**Algorithm 1** Time stepping

**Require:** State $\mathbf{s}_t = \{\mathbf{x}_t, \mathbf{v}_t, \mathbf{F}_e^t\}$
**Ensure:** Next state $\mathbf{s}_{t+1}$
  **Stress Evaluation:**
  **for** each material point $i = 1$ to $N$ **do**
    $\mathbf{P}_t^{(i)} \leftarrow \mathcal{E}_{\theta_e}(\mathbf{F}_e^{t,(i)}; \theta_e)$
  **end for**
  **Euler Integration:**
  $\mathbf{x}_{t+1}, \mathbf{v}_{t+1}, \mathbf{F}_e^{\text{trial}} \leftarrow \mathcal{I}(\mathbf{x}_t, \mathbf{v}_t, \mathbf{P}_t)$
  **Plasticity Update:**
  **for** each material point $i = 1$ to $N$ **do**
    $\mathbf{F}_e^{t+1,(i)} \leftarrow \mathcal{P}_{\theta_p}(\mathbf{F}_e^{\text{trial},(i)}; \theta_p)$
  **end for**

### 3.3 Rank-based depth-geometric anchors

Pixel-level color matching, adopted by methods like NeuMA [7], struggles under monocular video due to lighting variations, self-occlusions during rapid rotation, and geometric ambiguities. To overcome this, we propose Rank-based Depth-Geometric Anchors (RDGA), which establish robust geometric constraints by enforcing rank-based consensus between rendered depth and estimated depth, focusing on stable interior regions.

**Design rationale.** Monocular depth estimators suffer from inherent scale and shift ambiguity—absolute depth values are unreliable across frames. Standard metric losses (L1/L2) amplify this noise, as validated in Tab. 3 where replacing our rank-based formulation with L1 loss degrades performance below the baseline. Our Spearman rank correlation is scale-invariant: it only requires that relative depth orderings are preserved, which modern estimators achieve reliably even under large deformations.

Given a rendered image $I_{\text{render}} \in \mathbb{R}^{H\times W\times 3}$, a rendered depth map $D_{\text{render}} \in \mathbb{R}^{H\times W}$, and a ground-truth (GT) image $I_{\text{gt}} \in \mathbb{R}^{H\times W\times 3}$, we first compute a silhouette consistency loss:

$$\mathcal{L}_{\text{mask}} = \|M_{\text{render}} - M_{\text{gt}}\|_2^2, \tag{3}$$

where $M_{\text{render}}$ and $M_{\text{gt}}$ are object region masks on rendered and GT images.

We utilize a pre-trained depth estimation network $\mathcal{D}$ (*e.g.*, DAV [52]) to generate relative depth maps $D_{\text{gt}}$. A feature matching network $\mathcal{F}$ (*e.g.*, SuperGlue [11, 44]) extracts $N$ 2D correspondence pairs $\{(p_i, q_i)\}_{i=1}^N$, where $p_i = (x_i, y_i)$ and $q_i = (x_i', y_i')$ denote matched coordinates in $I_{\text{render}}$ and $I_{\text{gt}}$. To mitigate depth estimation errors near object boundaries, we define the edge region $\mathcal{M}_{\text{edge}}$ through morphological opening:

$$\mathcal{M}_{\text{edge}} = \mathcal{M} \circ K_{r\times r}, \tag{4}$$

where $\circ$ denotes morphological opening with an $r \times r$ rectangular kernel $K$. The stable interior region is $\mathcal{M}_{\text{stable}} = \mathcal{M} \setminus \mathcal{M}_{\text{edge}}$.

For each correspondence pair $(p_i, q_i)$, we validate depth consensus in local neighborhoods. Let $N_{p_i}$ and $N_{q_i}$ represent $k \times k$ regions centered at $p_i$ in $D_{\text{render}}$ and $q_i$ in $D_{\text{gt}}$. We compute the Spearman rank correlation coefficient $\gamma_i$ [45]:

$$\gamma_i = 1 - \frac{6 \sum_{j=1}^{k^2} (r_j - s_j)^2}{k^2(k^4 - 1)}, \tag{5}$$

where $r_j$ and $s_j$ are the ranks of the $j$-th depth value in $N_{p_i}$ and $N_{q_i}$. A consensus indicator function filters pairs:

$$\phi(p_i, q_i) = \mathbb{I}(\gamma_i > \tau). \tag{6}$$

Only pairs in $\mathcal{M}_{\text{stable}}$ satisfying $\phi(p_i, q_i) = 1$ are retained: $\mathcal{A} = \{(p_i, q_i) \mid p_i \in \mathcal{M}_{\text{stable}} \wedge \phi(p_i, q_i) = 1\}$.

The geometric alignment objective combines global and anchor-level terms:

$$\mathcal{L}_{\text{geo}} = \underbrace{\lambda_1 \left\| P_{\text{render}} - P_{\text{gt}} \right\|_2}_{\text{global alignment}} + \underbrace{\lambda_2 \sum_{(p_i, q_i) \in \mathcal{A}} (-\gamma_i)}_{\text{anchor-level supervision}}, \tag{7}$$

where $P_{\text{render}}$ and $P_{\text{gt}}$ represent matched point sets. The first term maintains global geometric consistency while the second focuses on precise local relationships through verified anchor pairs.

### 3.4 Constitutive prior regularizer

To resolve material ambiguity in sparse-view settings, we design a constitutive prior regularization process that evaluates candidate constitutive laws through parameter stability analysis.

**Constitutive hypotheses.** Following NCLaw [37], our elastic hypothesis set $\mathcal{H}_e$ includes four models: (1) corotated elasticity, (2) St. Venant–Kirchhoff (StVK) elasticity, (3) volume elasticity, and (4) sigma elasticity. The plastic hypothesis set $\mathcal{H}_p$ also includes four models: (1) identity plasticity, (2) sigma plasticity, (3) von Mises plasticity [39], and (4) Drucker–Prager plasticity [12]. These candidates cover common elastic, volumetric, and plastic response families. Full formulations are provided in the supplementary material. Importantly, CPR acts as a *soft regularization prior*—it remains effective even when the true material behavior does not match any of the hypotheses, as validated in Tab. 6.

The elastic deformation gradient $\mathbf{F}_e^t \in \mathbb{R}^{N \times 3 \times 3}$ serves as input, with physics residuals $\mathcal{R}_e, \mathcal{R}_p$ quantifying deviation from plausible laws. For each candidate law, we use $\mathbf{F}_e^t[\mathcal{S}, :, :]$ to evaluate alignment, where $\mathcal{S} \subset \{1, \ldots, N\}$ is a fixed subset of indices ($|\mathcal{S}| = 256$ by default).

The Constitutive Prior Regularizer (Alg. 2) operates through three phases. First, it performs standard elastic-plastic simulation: elastic stress prediction via

**Algorithm 2** Constitutive prior regularizer

**Require:**
$\mathbf{F}_e^t \in \mathbb{R}^{N\times 3\times 3}$, $\mathcal{H}_e = \{\mathcal{H}_e^k\}_{k=1}^K$, $\mathcal{H}_p = \{\mathcal{H}_p^m\}_{m=1}^M$, $\mathcal{E}_{\theta_e}$, $\mathcal{P}_{\theta_p}$, $\epsilon$, $\mathcal{S} \subset \{1,\ldots,N\}$
**Ensure:**
$\mathbf{F}_e^{t+1}$, $\mathcal{R}_e^t$, $\mathcal{R}_p^t$
**Elastic Stress Prediction:** $\mathbf{P}^t \leftarrow \mathcal{E}_{\theta_e}(\mathbf{F}_e^t)$
**Euler Integration:** $\mathbf{F}_e^{\text{trial}} \leftarrow \mathcal{I}(\mathbf{P}^t)$
**Plasticity Correction:** $\mathbf{F}_e^{t+1} \leftarrow \mathcal{P}_{\theta_p}(\mathbf{F}_e^{\text{trial}})$
**Parameter Solving:**
$\mathbf{F}_e^{t,\mathcal{S}} \leftarrow \mathbf{F}_e^t[\mathcal{S},:,:]$; $\mathbf{F}_e^{\text{trial},\mathcal{S}} \leftarrow \mathbf{F}_e^{\text{trial}}[\mathcal{S},:,:]$
**for** $k = 1$ **to** $K$ **do**
$\hat{\Theta}_e^k \leftarrow \arg\min_{\Theta_e^k} \|\mathcal{E}_{\theta_e}(\mathbf{F}_e^{t,\mathcal{S}}) - \mathcal{H}_e^k(\mathbf{F}_e^{t,\mathcal{S}};\Theta_e^k)\|_F^2$
$\omega_e^k \leftarrow \frac{1}{\text{Var}\{\hat{\Theta}_e^k\}+\epsilon}$
**end for**
**for** $m = 1$ **to** $M$ **do**
$\hat{\Theta}_p^m \leftarrow \arg\min_{\Theta_p^m} \|\mathcal{P}_{\theta_p}(\mathbf{F}_e^{\text{trial},\mathcal{S}}) - \mathcal{H}_p^m(\mathbf{F}_e^{\text{trial},\mathcal{S}};\Theta_p^m)\|_F^2$
$\omega_p^m \leftarrow \frac{1}{\text{Var}\{\hat{\Theta}_p^m\}+\epsilon}$
**end for**
$\mathcal{R}_e^t \leftarrow \sum_{k=1}^K \omega_e^k \|\mathcal{E}_{\theta_e}(\mathbf{F}_e^t) - \mathcal{H}_e^k(\mathbf{F}_e^t;\mathbb{E}[\hat{\Theta}_e^k])\|_F^2$
$\mathcal{R}_p^t \leftarrow \sum_{m=1}^M \omega_p^m \|\mathcal{P}_{\theta_p}(\mathbf{F}_e^{\text{trial}}) - \mathcal{H}_p^m(\mathbf{F}_e^{\text{trial}};\mathbb{E}[\hat{\Theta}_p^m])\|_F^2$

$\mathcal{E}_{\theta_e}$, Euler integration through $\mathcal{I}$, and plasticity correction using $\mathcal{P}_{\theta_p}$. Next, it solves inverse problems to estimate explicit parameters $\Theta$ for each candidate law. The credibility weights $\omega$ are computed as inverse variance measures (with smoothing factor $\epsilon$), assigning higher confidence to laws with stable parameter estimates. Finally, physical residuals $\mathcal{R}_e$ and $\mathcal{R}_p$ penalize deviations using weighted combinations.

**Overall optimization objective.** The total loss for optimizing the neural elastic model $\mathcal{E}_{\theta_e}$ and neural plasticity model $\mathcal{P}_{\theta_p}$ is

$$\mathcal{L} = \lambda_m \mathcal{L}_{\text{mask}} + \lambda_g \mathcal{L}_{\text{geo}} + \mathcal{R}_e + \mathcal{R}_p, \tag{8}$$

where $\lambda_m$ and $\lambda_g$ are balance factors. A theoretical analysis of the convergence properties is provided in the supplementary material.

## 4 Experiments

### 4.1 Experimental setup

**Datasets.** We conduct systematic validation across three dimensions. For *synthetic* experiments, we construct a challenging benchmark of six material types

(elastomers, gels, rubber, plasticine, granular materials, and non-Newtonian fluids) across diverse object geometries; the benchmark introduces (1) randomized lighting interference, (2) reduced frame rates matching real-world constraints, and (3) compound material modeling, with additional details provided in the supplementary material. For the *real-to-sim* bridge, we capture high-quality 3D Gaussian splatting models of real objects (*dragon*, *wolf*, *pudding*) and simulate dynamic sequences with complex material properties, providing ground-truth physics while retaining real-world geometric complexity. For *real-world* validation, we adopt the SpringGaus [56] dataset; while its original setup employs tri-view supervision, we strictly constrain our approach to single-view video, consistent with our problem setting. All experiments run on a single NVIDIA A800 80GB GPU.

**Baseline methods.** We evaluate against: (1) NCLaw [37] (implicit, requires particle GT); (2) NeuMA [7] (implicit, visual supervision); (3) SpringGaus [56] (explicit, real-world); and (4) PAC-NeRF [30] (explicit, multi-view), which we additionally adapt to monocular settings by augmenting with monocular depth supervision. We exclude PhysDreamer [55] and Physics3D [34] as their reliance on diffusion guidance and predefined explicit models is orthogonal to our implicit learning objective.

**Evaluation metrics.** We use: (1) Chamfer Distance (CD) [4, 14] for geometric consistency, (2) SSIM [50] for structural similarity, (3) PSNR [20] for pixel-level accuracy, and (4) LPIPS [54] for perceptual similarity.

### 4.2 Evaluation on synthetic dataset

We evaluate physical simulation accuracy by computing Chamfer Distance between predicted and ground-truth particle positions. Table 1 shows that under our challenging benchmark with color inconsistency and sparse supervision, our method achieves 48% lower average CD than NeuMA.

**Table 1: Quantitative comparison on synthetic dataset (Chamfer Distance ↓).** Our method achieves 48% lower average CD than NeuMA across diverse materials.

| Material | Elastomer | Gel | Rubber | Plasticine | Granular | Non-Newt. | |
|---|---|---|---|---|---|---|---|
| Object | Ball | Duck | Pawn | Cat | Fish | Bottle | Average |
| NCLaw [37] | 4.085 | 2.934 | 2.031 | 1.909 | 0.536 | 1.631 | 2.188 |
| NeuMA [7] | 1.123 | 1.863 | 0.517 | 0.844 | 0.322 | 1.056 | 0.954 |
| **Ours** | **0.922** | **0.702** | **0.200** | **0.318** | **0.077** | **0.757** | **0.496** |

Figure 3 illustrates temporal CD variations during simulation. Our method maintains alignment with ground truth throughout the simulation, while baselines gradually deviate. Further rendering metrics are provided in Fig. 4.

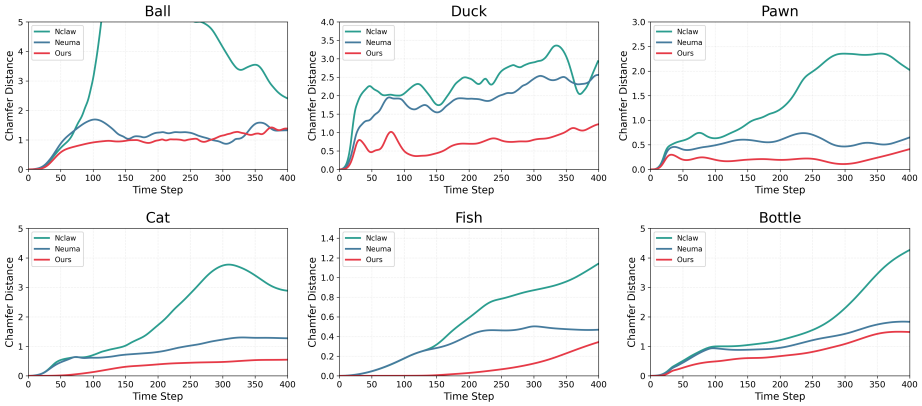



**Fig. 3: Chamfer Distance during physical simulation.** Our method consistently maintains lower CD throughout the simulation.

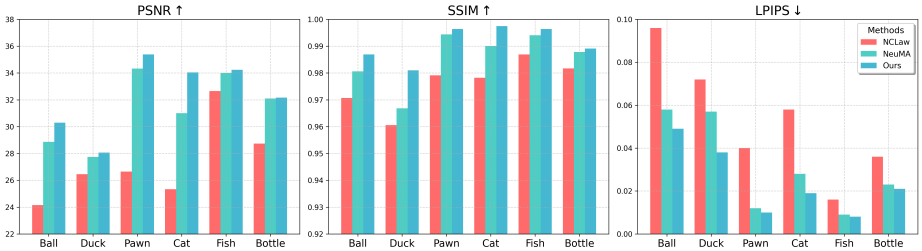



**Fig. 4: Rendering metrics on synthetic dataset.** Our method achieves superior PSNR, SSIM, and LPIPS.

### 4.3 Evaluation on real-to-sim dataset

We assess generalization to complex geometries derived from real objects using our real-to-sim dataset. Table 2 shows GCA consistently outperforms baselines. Figure 5 provides qualitative comparison, demonstrating superior ability to capture plausible dynamics for complex objects.

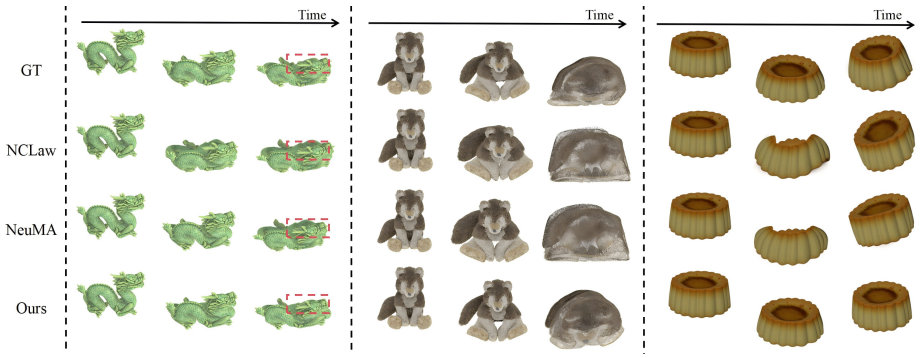



**Fig. 5: Qualitative comparison on real-to-sim dataset.** Our method accurately captures complex dynamics for real-world geometries.

**Table 2: Chamfer Distance on real-to-sim dataset (↓).** Our method achieves the lowest error across all objects.

| Object | Plast. Dragon | Sand Wolf | Gel Pudding |
|---|---|---|---|
| NCLaw [37] | 25.021 | 49.420 | 38.149 |
| NeuMA [7] | 3.527 | 9.803 | 13.804 |
| **Ours** | **2.081** | **5.842** | **0.906** |

### 4.4 Evaluation on real-world dataset

For real-world validation, we conduct monocular supervision experiments on the SpringGaus dataset [56]. While the original setup uses tri-view supervision, we strictly use single-view video. As shown in Fig. 6, GCA successfully disentangles implicit physical properties and demonstrates strong generalization under monocular constraints.

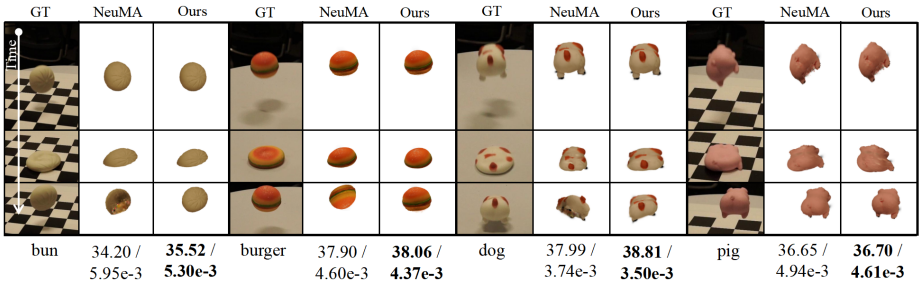



**Fig. 6: Qualitative comparison on real-world dataset.** Our method achieves better visual consistency using only monocular supervision. The PSNR and LPIPS values are computed after filtering the background.

### 4.5 Ablation and analysis

We conduct comprehensive ablation studies to validate each component and design choice.

**Component-wise ablation.** Table 3 provides a granular breakdown of each component's contribution.

**Table 3: Detailed ablation study (Chamfer Distance ↓).** Each component contributes meaningfully. Replacing rank-based depth alignment with L1 loss degrades performance *below the baseline*, validating our scale-invariant design.

| Configuration | Elast. | Gel | Rubber | Plast. | Gran. | Non-Newt. | Note |
|---|---|---|---|---|---|---|---|
| NeuMA (Baseline) | 1.123 | 1.863 | 0.517 | 0.844 | 0.322 | 1.056 | Prior best |
| w/o Global Align. | | | > 100 (Diverged) | | | | Divergence |
| w/o Anchor Superv. | 1.024 | 0.963 | 0.637 | 0.510 | 0.164 | 1.166 | Loss of detail |
| Replace Rank w/ L1 | 1.064 | 1.098 | 0.846 | 0.919 | 0.147 | 0.989 | Scale ambiguity |
| w/o RDGA | 3.842 | 3.045 | 1.975 | 1.821 | 0.559 | 1.531 | No geom. anchor |
| w/o CPR | 1.024 | **0.694** | 0.243 | 0.510 | 0.084 | 0.769 | No physics prior |
| **Ours (Full)** | **0.922** | 0.702 | **0.200** | **0.318** | **0.077** | **0.757** | Best performance |

(1) *Global alignment is foundational*: removing it causes complete optimization divergence, as the model cannot capture overall motion trends. (2) *Rank-based > Metric-based*: replacing our rank-based loss with standard L1 depth loss yields CD of 0.844 (avg.), significantly worse than our full model (0.496) and showing *degraded performance on 5 of 6 materials* compared to our rank-based formulation. This is because monocular depth estimators suffer from scale-shift ambiguity—forcing absolute scale match introduces noise rather than valid supervision. Our rank-based consensus mitigates this issue. (3) *RDGA is critical*: without it, performance degrades sharply across all materials. (4) *CPR provides*

*consistent benefit*: the regularizer improves performance on 5 of 6 materials. The single exception (Gel: 0.694 vs. 0.702) is within noise margins and reflects that simple elastic materials may not require additional physics guidance. (5) *Color supervision is unreliable under monocular shading.* In monocular videos of dynamic objects, self-occlusions and rapid rotation cause severe lighting changes and shading artifacts. Geometric consistency via depth provides a more invariant signal for learning physical dynamics.

**Computational efficiency.** Table 4 shows LoRA reduces GPU memory by 60% and training time by 87% versus full fine-tuning, while achieving superior CD (0.50 vs. 1.35). Full fine-tuning overfits visual noise.

**Table 4: Computational efficiency analysis.** LoRA fine-tuning achieves superior physical accuracy with lower cost.

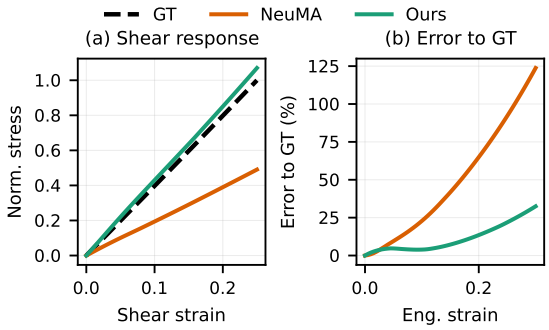
| Method | Mem. (GB) | Train (min) | Infer. (s) ↓ | CD ↓ |
|---|---|---|---|---|
| NeuMA [7] | 28.9 | 73.3 | 32.4 | 1.31 |
| Ours (Full F.T.) | 76.9 | 408.6 | 32.5 | 1.35 |
| **Ours (LoRA)** | **30.5** | **51.2** | 31.5 | **0.50** |

**Explicit methods are unstable under monocular supervision.** A natural question is whether existing explicit methods can be adapted to monocular settings by adding depth supervision. Table 5 provides empirical evidence: PAC-NeRF augmented with monocular depth becomes unstable and diverges in our monocular setting (CD $>$ 100). The rigid parameterization of explicit models (*e.g.*, Young's modulus) makes them overly sensitive to geometric noise inherent in monocular depth estimates. These results suggest that directly adding monocular depth supervision to existing explicit frameworks is insufficient—the specific design of RDGA (rank-based, edge-aware) and CPR (soft regularization) is essential.

**Table 5: Comparison with explicit methods under monocular setting (CD ↓).** Explicit methods diverge even with depth supervision; GCA remains robust.

| Method | Supervision | Ball | Cat | Status |
|---|---|---|---|---|
| PAC-NeRF [30] | Multi-view | 0.85 | 0.29 | Reference |
| NeuMA [7] | Multi-view | 0.98 | 0.35 | Reference |
| PAC-NeRF + Depth | Monocular | >100 | >100 | Diverged |
| NeuMA [7] | Monocular | 1.12 | 0.84 | Suboptimal |
| **Ours** | Monocular | **0.92** | **0.32** | Robust |

**Direct law-level validation.** Beyond trajectory-level alignment, we further evaluate whether the learned implicit constitutive law matches the underlying physical response. Using an analytic JellyDuck law as ground truth, we query the learned elasticity under held-out deformation gradients that are not used for image supervision. As shown in Fig. 7, GCA follows the ground-truth response more closely than NeuMA and reduces the average constitutive response error from 39.6% to 8.0%. This indicates that the improvement is not merely due to better trajectory fitting, but also to more accurate law-level recovery.




| Method | Uni. | Sh. | Avg. |
|---|---|---|---|
| NeuMA [7] | 52.0 | 27.1 | 39.6 |
| **Ours** | **12.0** | **3.9** | **8.0** |

Response error (%) ↓

**Fig. 7: Direct law-level validation on held-out deformation gradients.** Left: normalized constitutive response under analytic JellyDuck ground truth. Right: quantitative response error. GCA more accurately recovers the underlying constitutive response than NeuMA, reducing the average error from 39.6% to 8.0%.

**Robustness to out-of-distribution materials.** A critical concern is whether CPR becomes counterproductive when the true material is absent from the hypotheses. Table 6 addresses this with two stress tests: (1) a *blind test* where the ground-truth constitutive model is deliberately removed from the library, and (2) *composite materials* combining elasticity, plasticity, and fluid properties—behaviors not covered by any single hypothesis. In both cases, GCA still substantially outperforms the baseline, confirming that CPR provides *general physical guidance* (*e.g.*, energy consistency, stability constraints) rather than requiring exact template matches. The implicit LoRA layers successfully compensate for residual differences between the prior and actual material behavior.

**Table 6: Robustness to out-of-distribution materials (CD ↓).** GCA outperforms baselines even when the correct hypothesis is absent or for composite materials.

| Configuration | Elastomer (target excl.) | Plasticine (target excl.) | Composite (undefined) |
|---|---|---|---|
| NeuMA [7] | 1.123 | 0.844 | 3.628 |
| Ours (Target Removed) | 0.976 | 0.397 | – |
| **Ours (Full Library)** | **0.922** | **0.318** | **1.275** |

**Generalization.** The learned constitutive laws are not tied to the specific geometry used during training: once acquired, they can be applied to novel object shapes and further support multi-object interaction scenarios where different objects follow different learned material laws. Additional qualitative results are provided in the supplementary material.

## 5 Conclusion

We presented GCA, a framework for learning implicit constitutive laws from monocular dynamic video through visual-physical bidirectional alignment. By unifying LoRA-based adaptation with Rank-based Depth-Geometric Anchors (RDGA) and a Constitutive Prior Regularizer (CPR), our method achieves 48% lower Chamfer Distance than the strongest baseline on synthetic data, strong generalization on real-to-sim datasets, and superior quality in real-world monocular experiments—while remaining robust even when the true material is absent from the hypotheses.

**Limitations.** GCA still requires a static multi-view scan for geometric initialization, and extending the entire pipeline to fully monocular reconstruction is left for future work. Its geometric anchors also depend on the quality of monocular depth estimation, although the rank-based formulation mitigates scale-shift ambiguity. Finally, our current experiments focus mainly on single-object dynamics; extending the framework to dense multi-object scenes with heterogeneous materials is a promising direction.

**Acknowledgements.** This work is supported by Hong Kong Research Grants Council – General Research Fund (Grant No. 17213825), Hong Kong Innovation and Technology Commission – Innovation and Technology Fund (Grant No. ITS/488/24FP), and HKU Seed Fund for PI Research.

## References

1. Aira, L.S., Montanaro, A., Aiello, E., Valsesia, D., Magli, E.: MotionCraft: Physics-based zero-shot video generation. arXiv preprint arXiv:2405.13557 (2024)

2. Arruda, E.M., Boyce, M.C.: A three-dimensional constitutive model for the large stretch behavior of rubber elastic materials. Journal of the Mechanics and Physics of Solids (1993)
3. Billard, A., Kragic, D.: Trends and challenges in robot manipulation. Science (2019)
4. Butt, M.A., Maragos, P.: Optimum design of chamfer distance transforms. IEEE TIP (1998)
5. Cai, J., Yang, Y., Yuan, W., He, Y., Dong, Z., Bo, L., Cheng, H., Chen, Q.: Gaussian-informed continuum for physical property identification and simulation. In: NeurIPS (2024)
6. Cai, S., Mao, Z., Wang, Z., Yin, M., Karniadakis, G.E.: Physics-informed neural networks (pinns) for fluid mechanics: A review. Acta Mechanica Sinica (2021)
7. Cao, J., Guan, S., Ge, Y., Li, W., Yang, X., Ma, C.: NeuMA: Neural material adaptor for visual grounding of intrinsic dynamics. In: NeurIPS (2024)
8. Chen, H.y., Tretschk, E., Stuyck, T., Kadlecek, P., Kavan, L., Vouga, E., Lassner, C.: Virtual elastic objects. In: CVPR (2022)
9. Chhabra, R.P., Patel, S.A.: Bubbles, drops, and particles in non-Newtonian fluids. CRC press (2023)
10. Croitoru, F.A., Hondru, V., Ionescu, R.T., Shah, M.: Diffusion models in vision: A survey. IEEE TPAMI (2023)
11. DeTone, D., Malisiewicz, T., Rabinovich, A.: Superpoint: Self-supervised interest point detection and description. In: CVPR (2018)
12. Drucker, D.C., Prager, W.: Soil mechanics and plastic analysis or limit design. Quarterly of applied mathematics (1952)
13. Dubied, M., Michelis, M.Y., Spielberg, A., Katzschmann, R.K.: Sim-to-real for soft robots using differentiable fem: Recipes for meshing, damping, and actuation. IEEE Robotics and Automation Letters (2022)
14. Erler, P., Guerrero, P., Ohrhallinger, S., Mitra, N.J., Wimmer, M.: Points2surf learning implicit surfaces from point clouds. In: ECCV (2020)
15. Fang, J., Yi, T., Wang, X., Xie, L., Zhang, X., Liu, W., Nießner, M., Tian, Q.: Fast dynamic radiance fields with time-aware neural voxels. In: SIGGRAPH Asia (2022)
16. Feng, Y., Feng, X., Shang, Y., Jiang, Y., Yu, C., Zong, Z., Shao, T., Wu, H., Zhou, K., Jiang, C., Yang, Y.: Gaussian splashing: Unified particles for versatile motion synthesis and rendering. arXiv preprint arXiv:2401.15318 (2024)
17. Feng, Y., Shang, Y., Li, X., Shao, T., Jiang, C., Yang, Y.: Pie-nerf: Physics-based interactive elastodynamics with nerf. In: CVPR (2024)
18. Fung, Y.: Elasticity of soft tissues in simple elongation. American Journal of Physiology-Legacy Content (1967)
19. Fung, Y.c.: A first course in continuum mechanics. Englewood Cliffs (1977)
20. Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: Int. Conf. Pattern Recog. (2010)
21. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., et al.: LoRA: Low-rank adaptation of large language models. In: ICLR (2022)
22. Hu, Y., Fang, Y., Ge, Z., Qu, Z., Zhu, Y., Pradhana, A., Jiang, C.: A moving least squares material point method with displacement discontinuity and two-way rigid body coupling. ACM TOG (2018)
23. Huang, D.Z., Xu, K., Farhat, C., Darve, E.: Learning constitutive relations from indirect observations using deep neural networks. Journal of Computational Physics (2020)

24. Huang, T., Zeng, Y., Li, H., Zuo, W., Lau, R.W.: DreamPhysics: Learning physical properties of dynamic 3D gaussians with video diffusion priors. arXiv preprint arXiv:2406.01476 (2024)
25. Jiang, Y., Yu, C., Xie, T., Li, X., Feng, Y., Wang, H., Li, M., Lau, H., Gao, F., Yang, Y., et al.: Vr-gs: A physical dynamics-aware interactive gaussian splatting system in virtual reality. In: SIGGRAPH (2024)
26. Juarez, M.G., Botti, V.J., Giret, A.S.: Digital twins: Review and challenges. Journal of Computing and Information Science in Engineering (2021)
27. Kaneko, T.: Improving physics-augmented continuum neural radiance field-based geometry-agnostic system identification with lagrangian particle optimization. In: CVPR (2024)
28. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3D gaussian splatting for real-time radiance field rendering. ACM TOG (2023)
29. Li, X., Cao, Y., Li, M., Yang, Y., Schroeder, C., Jiang, C.: Plasticitynet: Learning to simulate metal, sand, and snow for optimization time integration. In: NeurIPS (2022)
30. Li, X., Qiao, Y.L., Chen, P.Y., Jatavallabhula, K.M., Lin, M., Jiang, C., Gan, C.: PAC-NeRF: Physics-augmented continuum neural radiance fields for geometry-agnostic system identification. In: ICLR (2023)
31. Li, Z., Tucker, R., Snavely, N., Holynski, A.: Generative image dynamics. In: CVPR (2024)
32. Lin, Y., Lin, C., Xu, J., MU, Y.: OmniphysGS: 3D constitutive gaussians for general physics-based dynamics generation. In: ICLR (2025)
33. Liu, D., Zhang, J., Dinh, A.D., Park, E., Zhang, S., Xu, C.: Generative physical ai in vision: A survey. arXiv preprint arXiv:2501.10928 (2025)
34. Liu, F., Wang, H., Yao, S., Zhang, S., Zhou, J., Duan, Y.: Physics3D: Learning physical properties of 3D Gaussians via video diffusion. arXiv preprint arXiv:2406.04338 (2024)
35. Liu, S., Ren, Z., Gupta, S., Wang, S.: PhysGen: Rigid-body physics-grounded image-to-video generation. In: ECCV (2024)
36. Lu, L., Jin, P., Pang, G., Zhang, Z., Karniadakis, G.E.: Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. Nature machine intelligence (2021)
37. Ma, P., Chen, P.Y., Deng, B., Tenenbaum, J.B., Du, T., Gan, C., Matusik, W.: Learning neural constitutive laws from motion observations for generalizable PDE dynamics. In: ICML (2023)
38. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM (2021)
39. Mises, R.v.: Mechanik der festen körper im plastisch-deformablen zustand. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse (1913)
40. Nair, S., Rajeswaran, A., Kumar, V., Finn, C., Gupta, A.: R3m: A universal visual representation for robot manipulation. arXiv preprint arXiv:2203.12601 (2022)
41. Park, K., Sinha, U., Barron, J.T., Bouaziz, S., Goldman, D.B., Seitz, S.M., Martin-Brualla, R.: Nerfies: Deformable neural radiance fields. In: ICCV (2021)
42. Poole, B., Jain, A., Barron, J.T., Mildenhall, B.: DreamFusion: Text-to-3D using 2D diffusion. In: ICLR (2023)
43. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics (2019)

44. Sarlin, P.E., DeTone, D., Malisiewicz, T., Rabinovich, A.: Superglue: Learning feature matching with graph neural networks. In: CVPR (2020)
45. Sedgwick, P.: Spearman's rank correlation coefficient. Bmj (2014)
46. Stomakhin, A., Howes, R., Schroeder, C.A., Teran, J.M.: Energetically consistent invertible elasticity. In: Symposium on Computer Animation (2012)
47. Sulsky, D., Zhou, S.J., Schreyer, H.L.: Application of a particle-in-cell method to solid mechanics. Computer physics communications (1995)
48. Tan, X., Jiang, Y., Li, X., Zong, Z., Xie, T., Yang, Y., Jiang, C.: PhysMotion: Physics-grounded dynamics from a single image. arXiv preprint arXiv:2411.17189 (2024)
49. Wang, T.H., Ma, P., Spielberg, A.E., Xian, Z., Zhang, H., Tenenbaum, J.B., Rus, D., Gan, C.: Softzoo: A soft robot co-design benchmark for locomotion in diverse environments. arXiv preprint arXiv:2303.09555 (2023)
50. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE TIP (2004)
51. Xue, T., Liao, S., Gan, Z., Park, C., Xie, X., Liu, W.K., Cao, J.: JAX-FEM: A differentiable GPU-accelerated 3D finite element solver for automatic inverse design and mechanistic data science. Computer Physics Communications (2023)
52. Yang, H., Huang, D., Yin, W., Shen, C., Liu, H., He, X., Lin, B., Ouyang, W., He, T.: Depth any video with scalable synthetic data. arXiv preprint arXiv:2410.10815 (2024)
53. Yin, H., Varava, A., Kragic, D.: Modeling, learning, perception, and control methods for deformable object manipulation. Science Robotics (2021)
54. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR (2018)
55. Zhang, T., Yu, H.X., Wu, R., Feng, B.Y., Zheng, C., Snavely, N., Wu, J., Freeman, W.T.: PhysDreamer: Physics-based interaction with 3D objects via video generation. In: ECCV (2024)
56. Zhong, L., Yu, H.X., Wu, J., Li, Y.: Reconstruction and simulation of elastic objects with spring-mass 3D gaussians. In: ECCV (2024)