

Toward Material-Agnostic System Identification from Videos

Yizhou Zhao¹ Haoyu Chen¹ Chunjiang Liu¹ Zhenyang Li² Charles Herrmann³
Junhwa Hur³ Yinxiao Li³ Ming-Hsuan Yang^{3,4} Bhiksha Raj¹ Min Xu^{1*}
¹Carnegie Mellon University ²University of Alabama at Birmingham ³Google ⁴UC Merced
<https://github.com/Skaldak/MASIV>

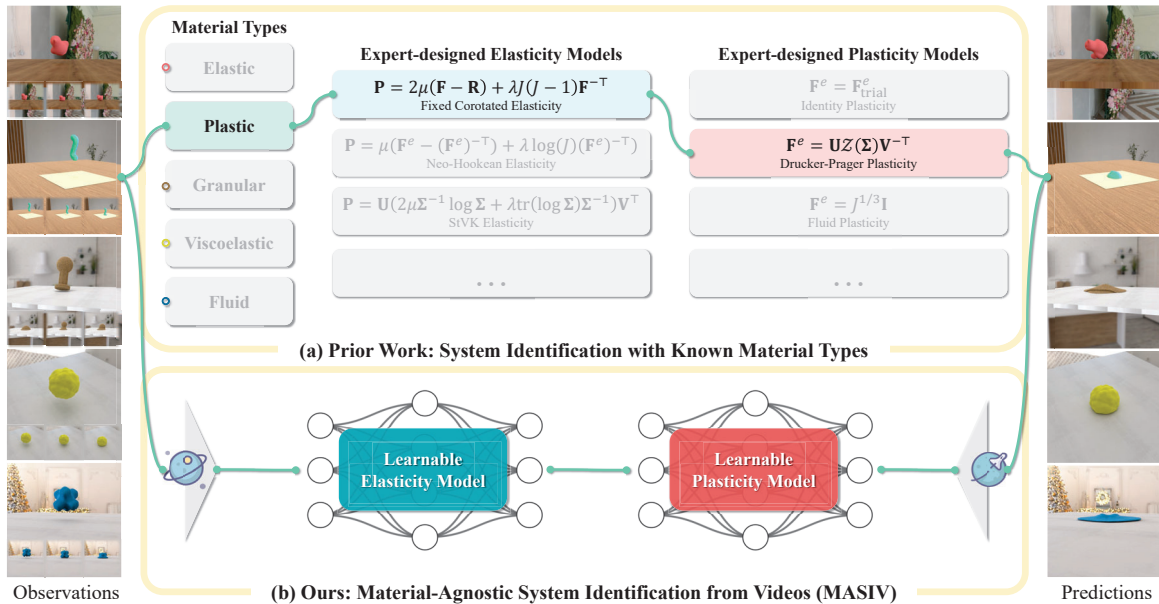


Figure 1. System identification from video observations seeks to understand the intrinsic dynamics of a given scene. (a) Existing methods rely on scene-specific material priors. (b) Our MASIV removes this requirement, grounding object motion in a material-agnostic manner.

Abstract

System identification from videos aims to recover object geometry and governing physical laws. Existing methods integrate differentiable rendering with simulation but rely on predefined material priors, limiting their ability to handle unknown ones. We introduce MASIV, the first vision-based framework for material-agnostic system identification. Unlike existing approaches that depend on hand-crafted constitutive laws, MASIV employs learnable neural constitutive models, inferring object dynamics without assuming a scene-specific material prior. However, the absence of full particle state information imposes unique challenges, leading to unstable optimization and physically implausible behaviors. To address this, we introduce dense geometric guidance by reconstructing continuum particle trajectories,

providing temporally rich motion constraints beyond sparse visual cues. Comprehensive experiments show that MASIV achieves state-of-the-art performance in geometric accuracy, rendering quality, and generalization ability.

1. Introduction

From the flow of water down a stream to the deformation of a rubber ball upon impact, the physical world operates with a set of underlying principles that govern motion and interaction. Humans intuitively grasp these principles through observation, allowing us to see a scene and instinctively predict its outcome or re-imagine it with different initial dynamics. Material identification forms the basis of this ability, requiring us to determine an object’s composition and understand how that influences its response to forces. This capability is fundamental to world modeling, the process of constructing internal representations that help us understand

*Corresponding author

and anticipate the evolution of the environment. However, replicating the ability to generalize physical laws to diverse scenarios remains a challenge for artificial intelligence.

To bridge this gap, researchers have explored vision-based system identification [6, 8, 9, 20, 23, 30, 38], which aims to infer governing physical laws from visual observations. Typically, they optimize a parameterized physical model, by end-to-end integrating differentiable renderers, such as Neural Radiance Fields (NeRF) [60] and 3D Gaussian Splatting (3DGS) [34], with differentiable simulators [26, 27, 31, 78, 91]. To enforce physically meaningful constraints on the simulation, existing approaches [6, 38, 100] often rely on material-specific constitutive laws, such as elasticity, plasticity, or viscosity models. These laws define how an object’s material properties influence its response to external forces, allowing system identification methods to estimate parameters such as stiffness, damping, or friction. While effective in controlled settings, this dependence on hand-crafted material models limits adaptability. Moreover, these methods assume knowing the material type for a given scene, making it difficult for them to generalize across in-the-wild scenarios where material properties are unknown.

Directly targeting these limitations, we introduce Material-Agnostic System Identification from Videos (MASIV), a vision-based approach that infers object dynamics free of predefined material priors. As illustrated in Fig. 1, rather than relying on expert-designed elasticity and plasticity models, MASIV employs fully learnable neural constitutive models inspired by NCLaw [59]. However, directly applying neural constitutive models to vision-based system identification poses significant challenges. Similar to the learning process of neural PDE solvers [5, 22, 44, 50], the original formulation of NCLaw assumes access to complete state observations of simulated particles, including position, velocity, deformation gradient, and affine momentum. Since visual observation fails to provide such complete information, simply relying on frame-wise pixel supervision provides insufficient constraints. This leads to unstable optimization and poor convergence [48].

We combine two main insights to enable this optimization. First, despite the lack of ground truth full particle state, frame-wise observations still allow reasonable estimates for some of them, such as position and velocity. Existing dynamic reconstruction methods [42, 55, 56] factorize moving objects with their canonical representation and corresponding deformations, which enables per-frame position tracking. Further, the commonly adopted motion basis [61, 94, 101] provides a low-rank encoding of object motion, facilitating state interpolation between frames. This formulation allows us to reconstruct particle trajectories over dense time steps. Second, although trajectory estimations are inherently limited to visible object regions,

internal deformations can be inferred by propagating surface motions inward. Our intuition is that nearby particles inside an object should be deformed similarly, akin to the As-Rigid-As-Possible (ARAP) prior.

We use these insights to guide the design of our framework. Concretely, we reconstruct dynamic Gaussians from multi-view video inputs, simultaneously establishing geometry and deformation models of the scene. The geometry model is then transformed into a solid continuum, where the deformation model is fine-tuned to encode both the exterior and the interior motion. To identify intrinsic dynamics, we guide the learning of the neural constitutive model using two complementary cues: appearance cues from video frames and geometric constraints from the fine-tuned deformation model. In this way, MASIV distills dense particle motion from sparse visual observations and guides the learning of neural constitutive models for stable, material-agnostic system identification.

Our contributions can be summed up as follows:

- We present a novel framework, MASIV, that determines the intrinsic dynamics of an object from videos in a material-agnostic manner. In contrast, all other current vision-based system identification methods [6, 8, 38] necessitate predefined material priors.
- We introduce temporally dense geometry guidance to constrain object behaviors with neural constitutive models, facilitating smoother convergence and improved performance for material-agnostic system identification.
- Through extensive qualitative and quantitative experiments, we demonstrate that MASIV achieves superior performance in both observable state recovery and future state prediction, even without explicit material priors.

2. Related Work

Dynamic reconstruction is a fundamental task in computer vision, aimed at reconstructing high-fidelity representations of complex dynamic scenes from diverse inputs such as monocular and multi-view videos [1, 7, 19, 41, 53, 76, 83, 88, 92, 95, 96]. Earlier research primarily focused on augmenting canonical Neural Radiance Fields (NeRF) [60] with a deformation field [64], which were later enhanced by incorporating volume-preserving regularization techniques [62, 63]. While primarily focused on the novel view synthesis task, their neural implicit representations often introduce noisy deformations, limiting the accuracy of recovered geometries required for reliable physics-based property estimation [38]. With the explicit representation of the scene using Gaussian ellipsoids, 3D Gaussian Splatting (3DGS) [34] achieves efficient dynamic scene reconstruction. Building on the principles of 3DGS, several studies have extended its application to 4D dynamic reconstruction by handling each frame separately [56] or mod-

eling the entire scene with a canonical 3D Gaussian point cloud coupled with a robust deformation framework that accurately maps it to the target scene [36, 86, 93].

Dynamic simulation integrates physics principles into 3D representations to generate realistic motion and interactions. A typical approach to synthesizing dynamic 3D scenes involves integrating a 3D generation pipeline with a video generation model [3, 49, 69, 75]. Recent research incorporates 3D Gaussian kernels as both visual and physical representations, embedding Newtonian dynamics into the Gaussian framework to enable seamless rendering and simulation while introducing physical constraints to ensure plausible dynamics [4, 16, 38, 48, 51, 67, 89, 100]. Building on this, Gaussian Splashing [14] introduces a position-based dynamics framework within 3D Gaussian Splatting, enabling interactions among solids, fluids, and deformable objects. Additionally, motion-conditioned simulation methods [18, 43, 74, 85, 87] use trajectory-based guidance to predict object motion, showcasing how generative models can enhance dynamic scene synthesis. However, these methods primarily rely on pre-trained generative models, which often lack a fundamental understanding of physics. More recent works leverage neural physics modeling, incorporating video priors and generative models to infer object dynamics, allowing physics-driven motion synthesis without explicit solvers [15, 28, 54, 81, 98]. However, these methods are constrained to specific material types, limiting their generalization across diverse objects. To address this, we introduce a material-agnostic approach that unifies system identification across different object types, ensuring broader applicability beyond predefined materials.

System identification aims to understand the physical laws governing the 3D world and is an essential task for simulation [37, 40, 45, 68, 79, 100] and robotic manipulation [46, 66, 72, 73, 99]. However, accurately recovering both object geometries and physical properties remains a significant challenge. Existing methods often assume known geometries and rely on finite element methods (FEM) or mass-spring systems for dynamic simulation [80, 82], limiting their ability to handle complex, real-world materials. Neural network-based methods [39, 70, 90] offer greater flexibility by leveraging data-driven models for system identification, but they struggle to generalize across diverse material properties and environmental conditions. More recently, differentiable physics-based simulations have emerged as a promising direction [11, 12, 17, 21, 27, 29, 30, 32, 58, 65]. However, these methods often require intricate modeling to bridge the gap between simulated and real-world behavior. To improve adaptability, some approaches focus on neural constitutive modeling, embedding learned material properties within physical simulations to refine expert-designed models [8, 59]. Other methods leverage the Material Point Method (MPM)

Algorithm 1 MPM state transfer $\mathcal{M}$

Input: $\mathbf{s}_n = \{\mathbf{x}_n, \mathbf{v}_n, \mathbf{F}_n^e\}$

Output: $\mathbf{s}_{n+1} = \{\mathbf{x}_{n+1}, \mathbf{v}_{n+1}, \mathbf{F}_{n+1}^e\}$

- 1: for $i = 1 \dots Q$, $\mathbf{P}_n(i) = \mathcal{E}(\mathbf{F}_n^e(i))$
 - 2: $\mathbf{x}_{n+1}, \mathbf{v}_{n+1}, \mathbf{F}_{n+1}^{e, \text{trial}} = \mathcal{I}(\mathbf{x}_n, \mathbf{v}_n, \mathbf{F}_n^e, \mathbf{P}_n)$
 - 3: for $i = 1 \dots Q$, $\mathbf{F}_{n+1}^e(i) = \mathcal{P}(\mathbf{F}_{n+1}^{e, \text{trial}}(i))$
-

and 3D Gaussian-based representations to enhance geometry reconstruction and material property estimation [6, 38, 71, 100]. Building upon GIC [6], our method presents a material-agnostic approach that eliminates reliance on predefined material priors. Rather than assuming material properties, our model infers object dynamics directly from visual observations, making it adapt better to a wide range of unknown or heterogeneous materials.

3. Method

Given visual observations $\mathbf{I} = \{\mathbf{I}_0, \mathbf{I}_1, \dots, \mathbf{I}_T\}$ over T frames capturing a dynamic object, our goal is to recover its explicit geometric representation and the governing evolution rules of the underlying dynamical system. Formally, we represent the object’s geometry using a set of Gaussian kernels [34] and integrate neural constitutive laws [59] with the material point method (MPM) [26, 31, 78] to roll out their motion over time. Unlike [6, 8, 38, 100], we do not assume a known material type of the object, while exploring Material-Agnostic System Identification from Videos (MASIV). Fig. 2 shows an overview of our pipeline.

3.1. Preliminaries

Material Point Method (MPM) discretizes the object of interest into Q material points and models their evolution over time via a state transfer equation

$$\mathbf{s}_{n+1} = \mathcal{M}(\mathbf{s}_n), \forall n = 0, 1, \dots, N. \quad (1)$$

$\mathbf{s}_n$ denotes the particle state at time step n . N is the total number of simulation steps with the simulation interval τ and observed T frames, where $N = T/\tau \gg T$.

The transition process $\mathcal{M}$ follows a time-stepping scheme, as outlined in Algorithm 1. The state of particle i at time step n is described by its position $\mathbf{x}_n(i)$, velocity $\mathbf{v}_n(i)$, and the elastic component $\mathbf{F}_n^e(i)$ of its deformation gradient $\mathbf{F}_n(i)$. At each time step, the elastic constitutive law $\mathcal{E}$ computes the first Piola-Kirchhoff stress $\mathbf{P}_n(i)$ from the elastic deformation gradient $\mathbf{F}_n^e(i)$. Subsequently, the time integration scheme $\mathcal{I}$ updates all particle states. The updated trial elastic deformation gradient, denoted as $\mathbf{F}_{n+1}^{e, \text{trial}}(i)$, needs to be further corrected by the plastic constitutive law $\mathcal{P}$, which applies plasticity constraints and yields the final elastic deformation gradient $\mathbf{F}_{n+1}^e(i)$. More details of MPM can be found in Sec. A.2 of Supp. Mat.

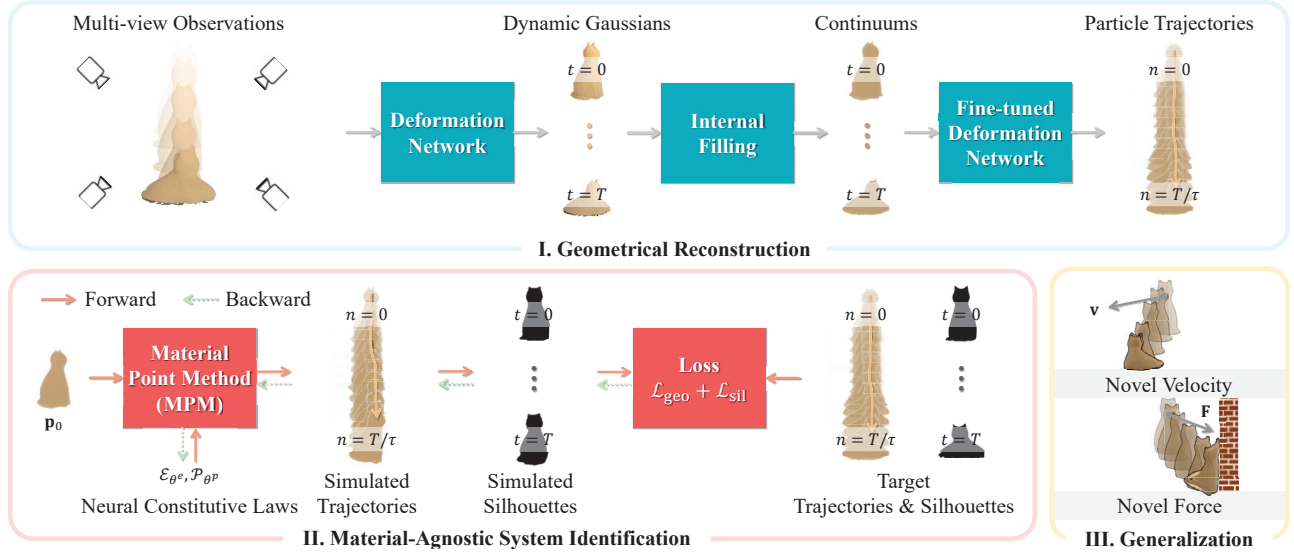


Figure 2. **The overview of MASIV.** Our pipeline comprises three phases. Phase I, Geometrical Reconstruction, infers geometrical representations (Sec. 3.2) and dense trajectories (Sec. 3.3) for the captured object. Phase II, Material-Agnostic System Identification (Sec. 3.4), grounds dynamic characteristics by exploiting visual observations with reconstructed motion clues. From this, we obtain a generalizable digital twin in Phase III, which can simulate novel interactions, such as new velocities or forces.

Existing efforts [6, 8, 38, 100] for video-based system identification commonly assume known constitutive laws $\mathcal{E}$ and $\mathcal{P}$, with a particular backbone form, e.g., neo-Hookean elasticity with identity plasticity return function. This simplifies the problem by estimating a small number of physical properties, such as Young’s modulus, fluid viscosity, friction angles, etc., at the cost of sacrificing flexibility. In contrast, we explore the same problem without this material-specific assumption.

3.2. Dynamic Gaussian Reconstruction

We start by reconstructing dynamic Gaussians over T frames. We consider this reconstruction as temporally sparse because the number of simulation steps is hundreds of times more than the number of observed video frames T , i.e., $N = T/\tau \gg T$. Following [6, 36, 86, 93], we maintain a set of canonical Gaussians $\mathcal{G}^* = \{\mu^*, \Sigma^*, \mathbf{c}^*, \sigma^*\}$ and use a neural network $\mathcal{D}$ to warp them over time to represent observed frames. Specifically, a basis network maps a given time step t to B bases, producing deformation basis for position $\mathbf{B}^\mu(t) \in \mathbb{R}^{B \times 3}$ and scale $\mathbf{B}^s(t) \in \mathbb{R}^{B \times 1}$. A coefficient network then estimates the weight of each basis, $\mathbf{w}(\mu^*, t) \in \mathbb{R}^B$, from the center coordinates of canonical Gaussian kernels μ^* and the time step t . This results in deformed Gaussian kernels at each time step t :

$$\mu_t = \mu^* + \mathbf{w}(\mu^*, t)\mathbf{B}^\mu(t), \quad (2)$$

$$s_t = s^* + \mathbf{w}(\mu^*, t)\mathbf{B}^s(t). \quad (3)$$

We employ the same optimization strategy in [6, 93] to minimize the L1 and Structural Similarity Index Mea-

sure (SSIM) losses between rendered frames $\hat{\mathbf{I}}_t$ and corresponding ground-truth observations $\mathbf{I}_t$. We also regard all Gaussian kernels as isotropic and apply L1 regularization on scales for simplicity and reconstruction quality [10, 97, 100]. The overall objective function is formulated as

$$\min_{\mathcal{G}^*, \mathcal{D}} \left[\mathcal{L}_1(\hat{\mathbf{I}}_t, \mathbf{I}_t) + \lambda_{\text{SSIM}} \mathcal{L}_{\text{SSIM}}(\hat{\mathbf{I}}_t, \mathbf{I}_t) + \lambda_s \|s_t\|_1 \right], \quad (4)$$

with hyperparameters λ_{SSIM} and λ_s .

3.3. Continuum Trajectory Estimation

To bridge the gap between unevenly distributed Gaussian particles and the corresponding solid continuum, we follow [6] to fill internal volumes, forming continuum particles $\{\mathbf{p}_t\}$ with uniform scale s and density σ . Existing methods for material-specific system identification often rely on sparse visual [6, 8, 38] or geometrical [6] cues along T frames to optimize a small set of parameters. However, since we do not assume a known constitutive backbone or material types, these sparse signals may lead to unconstrained behaviors at unobserved time steps or even cause training instability. Inspired by system identification [22, 33, 59] using state data per simulation step, we propose leveraging particle trajectories over N simulation steps as dense geometrical supervision to better constrain the simulated behaviors of continuum particles.

Given that the deformation network trained in Sec. 3.2 can produce rough trajectories for dynamic Gaussians, we fine-tune it to establish trajectories for continuum particles. Specifically, we initialize the canonical space of particles

$\mathbf{x}_0$ with the filled continuum $\mathbf{p}_1$ at $t = 1$ and apply deformations similar to Eq. (2),

$$\mathbf{x}_t = \mathbf{x}^* + \mathbf{w}(\mathbf{x}^*, t)\mathbf{B}^{\mathbf{x}}(t), \quad \forall t = 0, 1, \dots, T, \quad (5)$$

resulting in per-frame deformed particles $\{\mathbf{x}_t\}$. Then we optimize this deformation with all filled continuums $\{\mathbf{p}_t\}$ using the Chamfer Distance [13, 57]

$$\min_{\mathbf{x}^*, \mathcal{D}'} [\mathcal{L}_{\text{CD}}(\{\mathbf{x}_t\}_{t=0}^T, \{\mathbf{p}_t\}_{t=0}^T)], \quad (6)$$

where $\mathcal{D}'$ denotes the fine-tuned deformation network. Unlike the deformation network $\mathcal{D}$ trained in Sec. 3.2, this fine-tuned version $\mathcal{D}'$ is aware of the internal particles absence in Gaussians. Further, the learned motion basis with temporal positional embeddings enables it for temporal interpolation, allowing us to probe the motion between observed discrete time steps $t = 0, 1, \dots, T$. Formally, we have particle positions at each simulation time step

$$\mathbf{x}_t = \mathbf{x}^* + \mathbf{w}(\mathbf{x}^*, t)\mathbf{B}^{\mathbf{x}}(t), \quad \forall t = 0, \tau, \dots, N\tau, \quad (7)$$

where τ is the time interval between two simulation steps. We then consider these temporally dense particle trajectories as pseudo ground truth for later optimization.

3.4. Material-Agnostic System Identification

Inspired by NCLaw [59], we parameterize the elastic and plastic constitutive laws using neural networks, denoted as $\mathcal{E}_{\theta^e}$ and $\mathcal{P}_{\theta^p}$, where θ^e and θ^p are their respective network parameters. Both networks share the same multi-layer perceptron (MLP) architecture with a single hidden layer. Further, they incorporate two physical priors, frame indifference and undeformed state equilibrium, enforced by using rotation-invariant input representations and eliminating bias terms, respectively. While we adopt the same parameterization for material-agnostic modeling, we introduce a more challenging setting by relaxing the input requirements from dense, complete state sequences $\{\mathbf{s}_t\}_{n=0}^N$ to sparse visual observations $\{\mathbf{I}_t\}_{t=0}^T$. To this end, we follow an analysis-by-synthesis scheme. For the initial particle state $\mathbf{s}_0$, we use positions $\mathbf{x}_0$ derived from Eq. (7), an optimizable velocity $\mathbf{v}_0$, and identity deformation gradients $\mathbf{F}_0$. After we evolve particle positions $\{\hat{\mathbf{x}}_t\}_{t=\tau}^{N\tau}$ through time integration as described in Eq. (1), we optimize θ^e and θ^p with

$$\min_{\theta^e, \theta^p} (\mathcal{L}_{\text{geo}} + \mathcal{L}_{\text{sil}}), \quad (8)$$

where $\mathcal{L}_{\text{geo}}$ supervises particle positions $\hat{\mathbf{x}}_t$ with estimated trajectories $\mathbf{x}_t$ and $\mathcal{L}_{\text{sil}}$ compares rendered masks $\hat{\mathbf{M}}_t$ with object silhouettes $\mathbf{M}_t$ extracted from frame $\mathbf{I}_t$

$$\mathcal{L}_{\text{geo}} = \mathcal{L}_{\text{traj}} = \mathcal{L}_1(\{\hat{\mathbf{x}}_t\}_{t=0}^{N\tau}, \{\mathbf{x}_t\}_{t=0}^{N\tau}), \quad (9)$$

$$\mathcal{L}_{\text{sil}} = \mathcal{L}_1(\{\hat{\mathbf{M}}_t\}_{t=0}^{N\tau}, \{\mathbf{M}_t\}_{t=0}^{N\tau}). \quad (10)$$

While not assuming known material types in our problem setting, we note that MASIV still requires a pre-trained constitutive model from [59] as a stable initialization.

4. Experiments

4.1. Experimental Settings

Datasets. We conduct our experiments on the PAC-NeRF dataset [38] and the Spring-Gaus dataset [100]. The PAC-NeRF dataset consists of synthetic objects simulated using the Material Point Method (MPM) and includes a variety of material types such as elastic bodies, plasticine, sand, and Newtonian and non-Newtonian fluids. This dataset provides 45 multi-view RGB video sequences that allow for the estimation of both object geometry and physical properties and approximately 14 frames per viewpoint. The Spring-Gaus dataset contains two subsets, namely, Synthetic and Real. The Synthetic subset focuses on the reconstruction and simulation of elastic objects, containing 30 frames in each of 10 viewpoints. The Real subset provides both dense-view static and sparse-view dynamic captures of elastic objects. We evaluate the performance of observable state simulation using both datasets, while the Spring-Gaus dataset is additionally utilized for future state prediction. To obtain object masks, we apply off-the-shelf matting with [35, 47].

Baselines. To assess the performance of our approach, we compare it against four state-of-the-art methods for system identification. PAC-NeRF [38] combines neural radiance fields with differentiable physics to jointly infer geometry and material properties. Spring-Gaus [100] integrates a spring-mass system into Gaussian representations for efficient elastic object simulation. NeuMA [8] fine-tunes a neural adapter on expert-designed physical models to better align reconstructed dynamics with visual observations. GIC [6] leverages a two-stage pipeline to obtain the geometry representation and physical properties sequentially. These baselines represent the most relevant prior work and provide a strong comparison for evaluating our method. We note that they all necessitate material priors. To be concrete, PAC-NeRF and GIC roughly categorize materials into five groups and adopt a certain set of physical models for each group. Spring-Gaus generally focuses on elastic objects, thereby being material-specific. And NeuMA leverages a scene-specific constitutive model as the material prior.

Metrics. We evaluate our method using several standard metrics to assess both geometric reconstruction quality and physical property estimation accuracy. Metrics including 1) Chamfer Distance (CD) [13, 57] with units expressed in 10^3 mm^2 ; 2) Peak Signal-to-Noise Ratio (PSNR) [24]; 3) Structural Similarity Index Measure (SSIM) [84]. The CD quantifies the similarity between two point distributions, while PSNR and SSIM assess the visual quality.

		PAC-NeRF						Spring-Gaus							
Method		Newtonian	Non-Newtonian	Elasticity	Plasticine	Sand	Mean	Torus	Cross	Cream	Apple	Paste	Chess	Banana	Mean
CD ↓	PAC-NeRF [38]	0.277	0.236	0.238	0.429	0.212	0.278	4.92	1.10	0.77	1.11	3.14	0.96	2.77	2.11
	Spring-Gaus [100]	-	-	-	-	-	-	0.17	0.48	0.36	0.38	0.19	1.80	2.60	0.85
	NeuMA [8]	-	-	-	-	-	-	4.59	0.06	0.08	0.04	0.71	0.05	0.03	0.79
	GIC [6]	0.243	0.195	0.178	0.196	0.250	0.212	0.13	0.13	0.14	0.15	0.17	0.41	0.03	0.17
	MASIV (Ours)	0.233	0.198	0.192	0.201	0.229	0.210	0.08	0.10	0.19	0.16	0.13	0.19	0.08	0.13

Table 1. **Observable state simulation on PAC-NeRF and Spring-Gaus Synthetic datasets.** Despite being the only technique that does not use expert-designed physical models (noted by the gray row), MASIV performs better than or on par with existing baselines.

	Method	Bun	Burger	Dog	Pig	Potato	Mean
PSNR $\uparrow$	Spring-Gaus [100]	30.69	34.01	32.10	34.97	32.72	32.90
	NeuMA [8]	31.27	23.78	25.61	25.40	-	26.51
	GIC [6]	34.68	40.45	37.17	38.32	39.95	38.11
	Ours	40.26	42.12	38.97	40.84	43.40	41.12
SSIM $\uparrow$	Spring-Gaus	0.992	0.994	0.994	0.996	0.992	0.994
	NeuMA	0.994	0.993	0.995	0.995	-	0.994
	GIC	0.995	0.997	0.997	0.997	0.996	0.996
	Ours	0.998	0.997	0.997	0.998	0.998	0.997

Table 2. **Observable state simulation on Spring-Gaus Real dataset.** MASIV achieves state-of-the-art performance.

Implementation details. In dynamic reconstruction, we follow [6] to optimize canonical Gaussians $\mathcal{G}^*$ and the deformation network $\mathcal{D}$ for 40K iterations. Then the deformation network is fine-tuned to estimate particle trajectories from filled continuum sequences for 10K iterations. In system identification, we set the simulation time interval τ to 1/200 the frame duration. We adopt the RAdam [52] optimizer to optimize the initial velocity $\mathbf{v}_0$ for 100 steps and constitutive model parameters θ^e, θ^p for 1000 steps. All experiments are conducted on a single NVIDIA A100 GPU.

4.2. Comparison Results

Observable state simulation. In Tab. 1, we compare MASIV with existing vision-based system identification methods for observable state simulation. The results indicate that MASIV outperforms all other methods in terms of CD, highlighting its effectiveness in reconstructing object dynamics across diverse materials. Further, Tab. 2 evaluates on real data. Thanks to the more flexible constitutive representation, MASIV achieves state-of-the-art results. This demonstrates the potential of MASIV as an adaptive and generalizable approach for system identification.

Future state simulation. In Tab. 3, we provide a detailed evaluation of future state simulation performance. The results indicate that MASIV consistently outperforms all other methods while slightly underperforming GIC. Since NeuMA [8] does not originally experiment on the synthetic subset of Spring-Gaus, we report our reproduction results. To be concrete, we use the closest pre-trained model, Jelly, from NCLaw [59] as the material prior and conduct LoRA [25] fine-tuning with hyperparameters from its orig-

	Method	Torus	Cross	Cream	Apple	Paste	Chess	Banana	Mean
CD ↓	PAC-NeRF [38]	2.47	3.87	2.21	4.69	37.70	8.20	66.43	17.94
	Spring-Gaus [100]	2.38	1.57	2.22	1.87	7.03	2.59	18.48	5.16
	NeuMA [8]	1.17	0.40	1147.4	0.05	44.26	0.33	0.30	170.56
	GIC [6]	0.75	1.09	0.94	0.22	2.79	0.77	0.12	0.95
	MASIV (Ours)	0.16	0.38	1.51	2.64	1.75	1.70	0.24	1.20
PSNR ↑	PAC-NeRF [38]	17.46	14.15	15.37	19.94	12.32	15.08	16.04	15.77
	Spring-Gaus [100]	16.83	16.93	15.42	21.55	14.71	16.08	17.89	17.06
	NeuMA [8]	16.61	19.67	2.32	26.25	13.42	20.85	23.16	17.47
	GIC [6]	20.24	30.51	19.15	26.89	16.31	18.44	29.29	22.98
	MASIV (Ours)	22.42	33.73	16.45	21.19	15.73	17.28	26.56	21.91
SSIM ↑	PAC-NeRF [38]	0.919	0.906	0.858	0.878	0.819	0.848	0.886	0.870
	Spring-Gaus [100]	0.919	0.940	0.862	0.902	0.872	0.881	0.904	0.897
	NeuMA [8]	0.942	0.948	0.889	0.964	0.889	0.933	0.964	0.933
	GIC [6]	0.942	0.939	0.909	0.948	0.894	0.912	0.964	0.930
	MASIV (Ours)	0.955	0.967	0.852	0.898	0.891	0.878	0.965	0.915

Table 3. **Future state simulation on Spring-Gaus Synthetic dataset.** MASIV outperforms all baselines except for GIC.

inal implementation. However, we observe unstable optimization processes for most categories. This is probably rooted in its sparse pixel supervision, which only operates once per frame. With dense geometrical constraints supervising hundreds of times per frame, MASIV achieves significantly better results on both geometrical accuracy and rendering quality. This suggests that MASIV effectively captures the underlying dynamics governing future states, allowing it to generate predictions that remain structurally and visually consistent over time. Unlike baseline approaches that rely on predefined material priors, MASIV demonstrates the ability to infer relevant properties directly from visual observations, making it more adaptable to a wide range of scenarios. On the other hand, while dense trajectory guidance enables overfitting to observations, the lack of category-level regularization, unlike hand-crafted constitutive models, limits its generalization under scarce data. This leads to inferior performance than GIC, which can benefit from its pre-known constitutive model as an oracle. More analysis on the enhanced generalization ability of MASIV with more data is in Sec. C of Supp. Mat.

4.3. Ablation Study

We investigate the impact of geometrical objectives here, while referring the readers to more ablation studies in Sec. B of Supp. Mat. Fig. 3 validates the effectiveness of dense trajectory supervision. As shown in Fig. 4, we evaluate three geometrical objectives by replacing the $\mathcal{L}_{\text{geo}}$ term in opti-

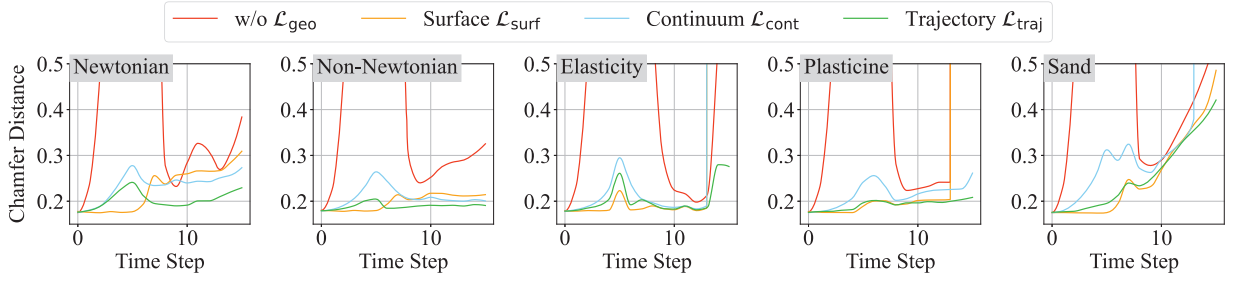


Figure 3. **Ablation study of geometrical objectives $\mathcal{L}_{\text{geo}}$ on PAC-NeRF dataset.** We compare three types of geometrical objectives against a baseline without geometrical guidance and plot their simulation errors using Chamfer Distance (CD). Overall, dense trajectory guidance achieves the lowest and most stable errors. The figure is best viewed in color.

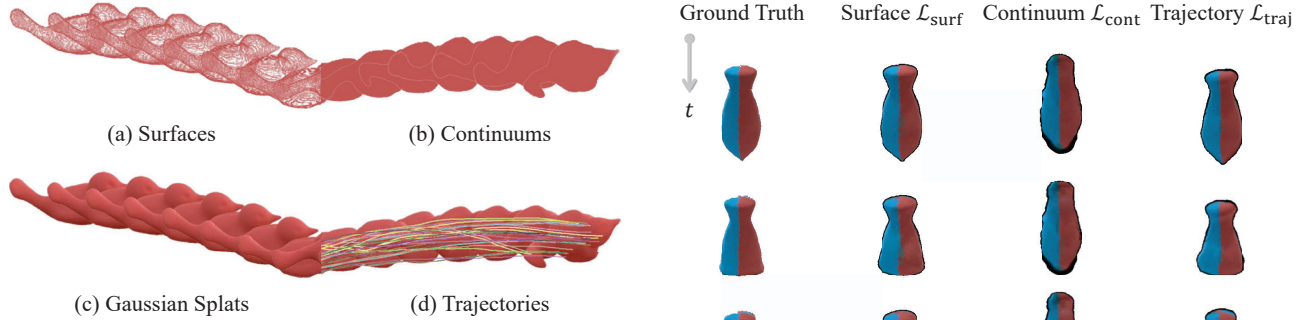


Figure 4. **Visualization of geometrical objectives $\mathcal{L}_{\text{geo}}$ on PAC-NeRF Bird.** The fine-tuned deformation network $\mathcal{D}'$ captures smooth and accurate motion for continuums, including low-frequency translation of the body and high-frequency swing of the tail. We space adjacent frames in the horizontal axis for clarity.

mizing the neural constitutive model objective, Eq. (8). We define them formally as follows. (1) **Surface**: per-frame surface alignment between the simulated continuum and the filled continuum,

$$\mathcal{L}_{\text{surf}} = \mathcal{L}_{\text{CD}} \left(\{S(\hat{\mathbf{x}}_t)\}_{t=0}^T, \{S(\mathbf{p}_t)\}_{t=0}^T \right), \quad (11)$$

where S denotes the surface extraction by thresholding Gaussian opacities [6]. (2) **Continuum**: similar to Surface while between solid continuums instead of surfaces,

$$\mathcal{L}_{\text{cont}} = \mathcal{L}_{\text{CD}} \left(\{\hat{\mathbf{x}}_t\}_{t=0}^T, \{\mathbf{p}_t\}_{t=0}^T \right). \quad (12)$$

(3) **Trajectory**: per-simulation-step trajectory alignment between the simulated continuum $\{\hat{\mathbf{x}}_t\}_{t=0}^{N\tau}$ and the continuum predicted by the deformation network $\{\mathbf{x}_t\}_{t=0}^{N\tau}$, as defined in Eq. (9). By analyzing per-frame errors, we first confirm the necessity of geometrical guidance, as the baseline (red curve) exhibits significantly higher CD errors across all material types. Surface-based supervision (yellow curve) shows moderate error reduction but remains less stable over time, especially in Plasticine and Sand. Continuum-based supervision (blue curve) further reduces errors but struggles with certain materials, e.g., Sand. These error overshoots

Figure 5. **Qualitative comparison of geometrical objectives $\mathcal{L}_{\text{geo}}$ on PAC-NeRF Cream.** Using sparse surfaces or solid continuum as geometrical cues can lead to physically implausible deformations, while trajectory guidance mitigates these issues by introducing temporally denser constraints. Better viewed when zoomed in.

are probably due to supervising relatively complex neural constitutive models with sparse signals, which might lead to overfitting since the simulated behaviors between observed discrete time steps are not well constrained. Finally, by enforcing dense alignment across simulation steps, trajectory-based supervision (green curve) achieves the lowest and most stable errors across all materials.

4.4. Qualitative Analysis

Dynamics reconstruction. We visualize the reconstructed Cream object by using the alternative objectives in Fig. 5. Here, we observe physically implausible deformations in results for surface and continuum supervision. Under uniform force, the object exhibits inconsistent distortion, espe-

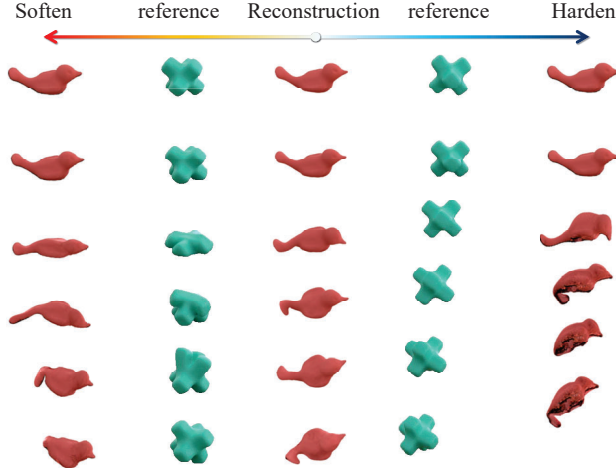


Figure 6. **Intra-class generalization.** We first optimize the geometry for PAC-NeRF Bird and then use the constitutive models optimized for reference elastic objects to predict novel dynamic behaviors, thereby softening or hardening the Bird respectively.

cially in the boundary between blue stripes and red stripes. We attribute this to two main reasons. 1) Chamfer Distance based geometrical constraint focuses mainly on the overall structure. While it encourages alignment between simulated and reference shapes, they do not enforce point-wise correspondence. A single particle can thus be matched with different ground truth particles in different time steps. As a result, local mismatches can accumulate, leading to inaccurate deformations. 2) Sparse supervision fails to constrain intermediate motion dynamics. Surface and continuum objectives provide alignment only at discrete time steps, leaving large gaps in supervision. Without explicit constraints on how the object deforms over time, the neural constitutive model may interpolate arbitrarily, leading to unstable and unrealistic motion patterns.

Dynamics generalization. We further study the generalization ability for MASIV. Fig. 6 examines how MASIV can be used to modify the dynamic characteristics of a single object (PAC-NeRF Bird). Specifically, we first reconstruct the geometrical representation of the Bird and optimize neural constitutive models for two elastic cross-shaped objects. These optimized constitutive models are then used as references to drive the simulation of the Bird, which deforms accordingly, exhibiting fluid-like softness on the left and increased rigidity and bounce on the right. Additionally, we see continuous transitions between soft and rigid behaviors when interpolating the neural constitutive model weights. In Fig. 7, we extend this study to inter-class generalization by applying neural constitutive models trained on different materials to a novel object, the PAC-NeRF Trophy. The results show that despite sharing the same initial state, each material model induces distinct deformation patterns. These

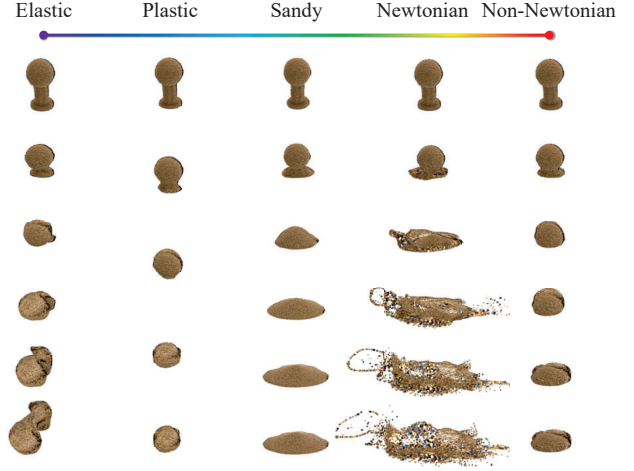


Figure 7. **Inter-class generalization.** We train multiple neural constitutive models with different materials and apply them to the PAC-NeRF Trophy. While sharing the same initial state, the Trophies with different materials exhibit distinct behaviors over time.

observations confirm that MASIV effectively captures generalizable intrinsic dynamics. More qualitative results can be found in Sec. D of Supp. Mat.

4.5. Discussion

While MASIV excels in material-agnostic system identification, several limitations remain. First, it relies on multi-view video input, restricting applicability to single-view or sparse observations. Second, our approach assumes that internal geometries and deformations can be inferred from surface motion, following a spatially coherent model. This assumption may break down in materials with complex internal flows, such as granular media or multiphase fluids. Finally, MASIV does not explicitly disentangle material properties from external forces in system identification, assuming a uniform gravity field for all objects. Future works aim to relax these assumptions for broader applicability.

5. Conclusion

We present MASIV, a material-agnostic system identification framework. By integrating neural constitutive models with dense geometry guidance, MASIV effectively captures object dynamics across diverse materials. Our approach reconstructs dynamic Gaussians from multi-view videos and fine-tunes a deformation model within a solid continuum. This enables us to extract dense motion cues, which, combined with visual observations, guide the learning of neural constitutive models. Through this formulation, MASIV generalizes beyond traditional system identification methods that rely on pre-defined material priors. Extensive experiments highlight its ability to reconstruct observable dynamics and to generalize for novel interactions.

Acknowledgment

This work was supported in part by U.S. NSF grant DBI-2238093, and in part by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean Government (MSIT) (No. RS-2024-00457882, National AI Research Lab Project).

References

- [1] Benjamin Attal, Jia-Bin Huang, Christian Richardt, Michael Zollhoefer, Johannes Kopf, Matthew O’Toole, and Changil Kim. Hyperreel: High-fidelity 6-dof video with ray-conditioned sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16610–16620, 2023. 2
- [2] Genesis Authors. Genesis: A universal and generative physics engine for robotics and beyond, 2024. 3
- [3] Sherwin Bahmani, Ivan Skorokhodov, Victor Rong, Gordon Wetzstein, Leonidas Guibas, Peter Wonka, Sergey Tulyakov, Jeong Joon Park, Andrea Tagliasacchi, and David B Lindell. 4d-fy: Text-to-4d generation using hybrid score distillation sampling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7996–8006, 2024. 3
- [4] Piotr Borycki, Weronika Smolak, Joanna Waczyńska, Marcin Mazur, Sławomir Tadeja, and Przemysław Spurek. Gasp: Gaussian splatting for physic-based simulations. *arXiv preprint arXiv:2409.05819*, 2024. 3
- [5] Johannes Brandstetter, Daniel Worrall, and Max Welling. Message passing neural pde solvers. *arXiv preprint arXiv:2202.03376*, 2022. 2
- [6] 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. *arXiv preprint arXiv:2406.14927*, 2024. 2, 3, 4, 5, 6, 7
- [7] Yuanhao Cai, Jiahao Wang, Alan Yuille, Zongwei Zhou, and Angtian Wang. Structure-aware sparse-view x-ray 3d reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11174–11183, 2024. 2
- [8] Junyi Cao, Shanyan Guan, Yanhao Ge, Wei Li, Xiaokang Yang, and Chao Ma. Neuma: Neural material adaptor for visual grounding of intrinsic dynamics. *Advances in Neural Information Processing Systems*, 37:65643–65669, 2025. 2, 3, 4, 5, 6
- [9] Pradyumna Chari, Chinmay Talegaonkar, Yunhao Ba, and Achuta Kadambi. Visual physics: Discovering physical laws from videos. *arXiv preprint arXiv:1911.11893*, 2019. 2
- [10] Hanlin Chen, Chen Li, and Gim Hee Lee. Neusg: Neural implicit surface reconstruction with 3d gaussian splatting guidance. *arXiv preprint arXiv:2312.00846*, 2023. 4
- [11] Hsiao-yu Chen, Edith Tretschk, Tuur Stuyck, Petr Kadlec, Ladislav Kavan, Etienne Vouga, and Christoph Lassner. Virtual elastic objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15827–15837, 2022. 3
- [12] Tao Du, Kui Wu, Pingchuan Ma, Sebastien Wah, Andrew Spielberg, Daniela Rus, and Wojciech Matusik. Diffpd: Differentiable projective dynamics. *ACM Transactions on Graphics (TOG)*, 41(2):1–21, 2021. 3
- [13] Philipp Erler, Paul Guerrero, Stefan Ohrhallinger, Niloy J Mitra, and Michael Wimmer. Points2surf learning implicit surfaces from point clouds. In *European Conference on Computer Vision*, pages 108–124. Springer, 2020. 5
- [14] Yutao Feng, Xiang Feng, Yintong Shang, Ying Jiang, Chang Yu, Zeshun Zong, Tianjia Shao, Hongzhi Wu, Kun Zhou, Chenfanfu Jiang, et al. Gaussian splashing: Unified particles for versatile motion synthesis and rendering. *arXiv preprint arXiv:2401.15318*, 3, 2024. 3
- [15] Yutao Feng, Yintong Shang, Xuan Li, Tianjia Shao, Chenfanfu Jiang, and Yin Yang. Pie-nerf: Physics-based interactive elastodynamics with nerf. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4450–4461, 2024. 3
- [16] Zhoujie Fu, Jiacheng Wei, Wenhao Shen, Chaoyue Song, Xiaofeng Yang, Fayao Liu, Xulei Yang, and Guosheng Lin. Sync4d: Video guided controllable dynamics for physics-based 4d generation. *arXiv preprint arXiv:2405.16849*, 2024. 3
- [17] 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 (TOG)*, 39(6):1–15, 2020. 3
- [18] Daniel Geng, Charles Herrmann, Junhwa Hur, Forrester Cole, Serena Zhang, Tobias Pfaff, Tatiana Lopez-Guevara, Carl Doersch, Yusuf Aytar, Michael Rubinstein, et al. Motion prompting: Controlling video generation with motion trajectories. *arXiv preprint arXiv:2412.02700*, 2024. 3
- [19] Sharath Girish, Tianye Li, Amrita Mazumdar, Abhinav Shrivastava, David Luebke, and Shalini De Mello. Queen: Quantized efficient encoding of dynamic gaussians for streaming free-viewpoint videos. *arXiv preprint arXiv:2412.04469*, 2024. 2
- [20] Shanyan Guan, Huayu Deng, Yunbo Wang, and Xiaokang Yang. Neurofluid: Fluid dynamics grounding with particle-driven neural radiance fields. In *International conference on machine learning*, pages 7919–7929. PMLR, 2022. 2
- [21] Eric Heiden, Miles Macklin, Yashraj Narang, Dieter Fox, Animesh Garg, and Fabio Ramos. Disect: A differentiable simulation engine for autonomous robotic cutting. *arXiv preprint arXiv:2105.12244*, 2021. 3
- [22] Maximilian Herde, Bogdan Raonic, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efficient foundation models for pdes. *Advances in Neural Information Processing Systems*, 37:72525–72624, 2025. 2, 4
- [23] Florian Hofherr, Lukas Koestler, Florian Bernard, and Daniel Cremers. Neural implicit representations for physical parameter inference from a single video. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2093–2103, 2023. 2

- [24] Alain Hore and Djemel Ziou. Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE, 2010. 5
- [25] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3, 2022. 6
- [26] 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 (TOG)*, 37(4):1–14, 2018. 2, 3
- [27] Yuanming Hu, Luke Anderson, Tzu-Mao Li, Qi Sun, Nathan Carr, Jonathan Ragan-Kelley, and Frédo Durand. DiffTaichi: Differentiable programming for physical simulation. *arXiv preprint arXiv:1910.00935*, 2019. 2, 3
- [28] Tianyu Huang, Haoze Zhang, Yihan Zeng, Zhilu Zhang, Hui Li, Wangmeng Zuo, and Rynson WH Lau. Dreamphysics: Learning physical properties of dynamic 3d gaussians with video diffusion priors. *arXiv preprint arXiv:2406.01476*, 2024. 3
- [29] Zhiao Huang, Yuanming Hu, Tao Du, Siyuan Zhou, Hao Su, Joshua B Tenenbaum, and Chuang Gan. Plasticinelab: A soft-body manipulation benchmark with differentiable physics. *arXiv preprint arXiv:2104.03311*, 2021. 3
- [30] Krishna Murthy Jatavallabhula, Miles Macklin, Florian Golemo, Vikram Voleti, Linda Petrini, Martin Weiss, Brendan Considine, Jérôme Parent-Lévesque, Kevin Xie, Kenny Erleben, et al. gradsim: Differentiable simulation for system identification and visuomotor control. *arXiv preprint arXiv:2104.02646*, 2021. 2, 3
- [31] Chenfanfu Jiang, Craig Schroeder, Joseph Teran, Alexey Stomakhin, and Andrew Selle. The material point method for simulating continuum materials. In *ACM SIGGRAPH 2016 Courses*, pages 1–52, 2016. 2, 3, 1
- [32] Takuhiro Kaneko. Improving physics-augmented continuum neural radiance field-based geometry-agnostic system identification with lagrangian particle optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5480, 2024. 3
- [33] Namgyu Kang, Jaemin Oh, Youngjoon Hong, and Eunbyung Park. Pig: Physics-informed gaussians as adaptive parametric mesh representations. *arXiv preprint arXiv:2412.05994*, 2024. 4
- [34] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023. 2, 3, 1
- [35] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023. 5
- [36] Agelos Kratimenos, Jiahui Lei, and Kostas Daniilidis. Dynmf: Neural motion factorization for real-time dynamic view synthesis with 3d gaussian splatting. In *European Conference on Computer Vision*, pages 252–269. Springer, 2025. 3, 4
- [37] Jinxi Li, Ziyang Song, and Bo Yang. Nvfi: neural velocity fields for 3d physics learning from dynamic videos. *Advances in Neural Information Processing Systems*, 36, 2024. 3
- [38] 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. 2, 3, 4, 5, 6
- [39] Yunzhu Li, Jiajun Wu, Russ Tedrake, Joshua B Tenenbaum, and Antonio Torralba. Learning particle dynamics for manipulating rigid bodies, deformable objects, and fluids. *arXiv preprint arXiv:1810.01566*, 2018. 3
- [40] Yifei Li, Tao Du, Kui Wu, Jie Xu, and Wojciech Matusik. Diffcloth: Differentiable cloth simulation with dry frictional contact. *ACM Transactions on Graphics (TOG)*, 42(1):1–20, 2022. 3
- [41] Zhengqi Li, Qianqian Wang, Forrester Cole, Richard Tucker, and Noah Snavely. Dynibar: Neural dynamic image-based rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4273–4284, 2023. 2
- [42] Zhan Li, Zhang Chen, Zhong Li, and Yi Xu. Spacetime gaussian feature splatting for real-time dynamic view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8508–8520, 2024. 2
- [43] Zhengqi Li, Richard Tucker, Noah Snavely, and Aleksander Holynski. Generative image dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24142–24153, 2024. 3
- [44] Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/JMS Journal of Data Science*, 1(3):1–27, 2024. 2
- [45] Junbang Liang, Ming Lin, and Vladlen Koltun. Differentiable cloth simulation for inverse problems. *Advances in Neural Information Processing Systems*, 32, 2019. 3
- [46] Xiao Liang, Fei Liu, Yutong Zhang, Yuelei Li, Shan Lin, and Michael Yip. Real-to-sim deformable object manipulation: Optimizing physics models with residual mappings for robotic surgery. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 15471–15477. IEEE, 2024. 3
- [47] Shanchuan Lin, Andrey Ryabtsev, Soumyadip Sengupta, Brian L Curless, Steven M Seitz, and Ira Kemelmacher-Shlizerman. Real-time high-resolution background matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8762–8771, 2021. 5
- [48] 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. 2, 3
- [49] Huan Ling, Seung Wook Kim, Antonio Torralba, Sanja Fidler, and Karsten Kreis. Align your gaussians: Text-to-4d

- with dynamic 3d gaussians and composed diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8576–8588, 2024. [3](#)
- [50] Phillip Lippe, Bas Veeling, Paris Perdikaris, Richard Turner, and Johannes Brandstetter. Pde-refiner: Achieving accurate long rollouts with neural pde solvers. *Advances in Neural Information Processing Systems*, 36:67398–67433, 2023. [2](#)
- [51] 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. [3](#)
- [52] Liyuan Liu, Haoming Jiang, Pengcheng He, Weizhu Chen, Xiaodong Liu, Jianfeng Gao, and Jiawei Han. On the variance of the adaptive learning rate and beyond. *arXiv preprint arXiv:1908.03265*, 2019. [6](#)
- [53] Qingming Liu, Yuan Liu, Jiepeng Wang, Xianqiang Lv, Peng Wang, Wenping Wang, and Junhui Hou. Modgs: Dynamic gaussian splatting from causally-captured monocular videos. *arXiv preprint arXiv:2406.00434*, 2024. [2](#)
- [54] Shaowei Liu, Zhongzheng Ren, Saurabh Gupta, and Shenglong Wang. Physgen: Rigid-body physics-grounded image-to-video generation. In *European Conference on Computer Vision*, pages 360–378. Springer, 2025. [3](#)
- [55] Zhicheng Lu, Xiang Guo, Le Hui, Tianrui Chen, Min Yang, Xiao Tang, Feng Zhu, and Yuchao Dai. 3d geometry-aware deformable gaussian splatting for dynamic view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8900–8910, 2024. [2](#)
- [56] Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. *arXiv preprint arXiv:2308.09713*, 2023. [2](#)
- [57] Baorui Ma, Zhizhong Han, Yu-Shen Liu, and Matthias Zwicker. Neural-pull: Learning signed distance functions from point clouds by learning to pull space onto surfaces. *arXiv preprint arXiv:2011.13495*, 2020. [5](#)
- [58] Pingchuan Ma, Tao Du, Joshua B Tenenbaum, Wojciech Matusik, and Chuang Gan. Risp: Rendering-invariant state predictor with differentiable simulation and rendering for cross-domain parameter estimation. *arXiv preprint arXiv:2205.05678*, 2022. [3](#)
- [59] Pingchuan Ma, Peter Yichen Chen, Bolei Deng, Joshua B Tenenbaum, Tao Du, Chuang Gan, and Wojciech Matusik. Learning neural constitutive laws from motion observations for generalizable pde dynamics. In *International Conference on Machine Learning*, pages 23279–23300. PMLR, 2023. [2](#), [3](#), [4](#), [5](#), [6](#)
- [60] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. [2](#)
- [61] Clinton A Mo, Kun Hu, Chengjiang Long, and Zhiyong Wang. Continuous intermediate token learning with implicit motion manifold for keyframe based motion interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13894–13903, 2023. [2](#)
- [62] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5865–5874, 2021. [2](#)
- [63] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M Seitz. Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *arXiv preprint arXiv:2106.13228*, 2021. [2](#)
- [64] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10318–10327, 2021. [2](#)
- [65] Yiling Qiao, Junbang Liang, Vladlen Koltun, and Ming Lin. Differentiable simulation of soft multi-body systems. *Advances in Neural Information Processing Systems*, 34: 17123–17135, 2021. [3](#)
- [66] Yi-Ling Qiao, Alexander Gao, and Ming Lin. Neuphysics: Editable neural geometry and physics from monocular videos. *Advances in Neural Information Processing Systems*, 35:12841–12854, 2022. [3](#)
- [67] Ri-Zhao Qiu, Ge Yang, Weijia Zeng, and Xiaolong Wang. Language-driven physics-based scene synthesis and editing via feature splatting. In *European Conference on Computer Vision*, pages 368–383. Springer, 2024. [3](#)
- [68] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019. [3](#)
- [69] Jiawei Ren, Liang Pan, Jiaxiang Tang, Chi Zhang, Ang Cao, Gang Zeng, and Ziwei Liu. Dreamgaussian4d: Generative 4d gaussian splatting. *arXiv preprint arXiv:2312.17142*, 2023. [3](#)
- [70] Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *International conference on machine learning*, pages 8459–8468. PMLR, 2020. [3](#)
- [71] Yidi Shao, Mu Huang, Chen Change Loy, and Bo Dai. Gausim: Registering elastic objects into digital world by gaussian simulator. *arXiv preprint arXiv:2412.17804*, 2024. [3](#)
- [72] Haochen Shi, Huazhe Xu, Samuel Clarke, Yunzhu Li, and Jiajun Wu. Robocook: Long-horizon elasto-plastic object manipulation with diverse tools. *arXiv preprint arXiv:2306.14447*, 2023. [3](#)
- [73] Haochen Shi, Huazhe Xu, Zhiao Huang, Yunzhu Li, and Jiajun Wu. Robocraft: Learning to see, simulate, and shape elasto-plastic objects in 3d with graph networks. *The International Journal of Robotics Research*, 43(4):533–549, 2024. [3](#)
- [74] Xiaoyu Shi, Zhaoyang Huang, Fu-Yun Wang, Weikang Bian, Dasong Li, Yi Zhang, Manyuan Zhang, Ka Chun

- Cheung, Simon See, Hongwei Qin, et al. Motion-i2v: Consistent and controllable image-to-video generation with explicit motion modeling. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024. 3
- [75] Uriel Singer, Shelly Sheynin, Adam Polyak, Oron Ashual, Iurii Makarov, Filippos Kokkinos, Naman Goyal, Andrea Vedaldi, Devi Parikh, Justin Johnson, et al. Text-to-4d dynamic scene generation. *arXiv preprint arXiv:2301.11280*, 2023. 3
- [76] Colton Stearns, Adam Harley, Mikaela Uy, Florian Dubost, Federico Tombari, Gordon Wetzstein, and Leonidas Guibas. Dynamic gaussian marbles for novel view synthesis of casual monocular videos. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–11, 2024. 2
- [77] Alexey Stomakhin, Craig Schroeder, Lawrence Chai, Joseph Teran, and Andrew Selle. A material point method for snow simulation. *ACM Transactions on Graphics (TOG)*, 32(4):1–10, 2013. 1
- [78] Deborah Sulsky, Shi-Jian Zhou, and Howard L Schreyer. Application of a particle-in-cell method to solid mechanics. *Computer physics communications*, 87(1-2):236–252, 1995. 2, 3
- [79] Priya Sundareshan, Rika Antonova, and Jeannette Bohgl. Diffcloud: Real-to-sim from point clouds with differentiable simulation and rendering of deformable objects. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10828–10835. IEEE, 2022. 3
- [80] Tetsuya Takahashi and Ming C Lin. Video-guided real-to-virtual parameter transfer for viscous fluids. *ACM Transactions on Graphics (TOG)*, 38(6):1–12, 2019. 3
- [81] Xiyang Tan, Ying Jiang, Xuan Li, Zeshun Zong, Tianyi Xie, Yin Yang, and Chenfanfu Jiang. Physmotion: Physics-grounded dynamics from a single image. *arXiv preprint arXiv:2411.17189*, 2024. 3
- [82] Bin Wang, Longhua Wu, KangKang Yin, Uri M Ascher, Libin Liu, and Hui Huang. Deformation capture and modeling of soft objects. *ACM Trans. Graph.*, 34(4):94–1, 2015. 3
- [83] Chaoyang Wang, Lachlan Ewen MacDonald, Laszlo A Jeni, and Simon Lucey. Flow supervision for deformable nerf. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21128–21137, 2023. 2
- [84] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 5
- [85] Zhouxia Wang, Ziyang Yuan, Xintao Wang, Yaowei Li, Tianshui Chen, Menghan Xia, Ping Luo, and Ying Shan. Motionctrl: A unified and flexible motion controller for video generation. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024. 3
- [86] GuanJun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20310–20320, 2024. 3, 4
- [87] Weijia Wu, Zhuang Li, Yuchao Gu, Rui Zhao, Yefei He, David Junhao Zhang, Mike Zheng Shou, Yan Li, Tingting Gao, and Di Zhang. Draganything: Motion control for anything using entity representation. In *European Conference on Computer Vision*, pages 331–348. Springer, 2024. 3
- [88] Wenqi Xian, Jia-Bin Huang, Johannes Kopf, and Changil Kim. Space-time neural irradiance fields for free-viewpoint video. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9421–9431, 2021. 2
- [89] 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*, pages 4389–4398, 2024. 3
- [90] Zhenjia Xu, Jiajun Wu, Andy Zeng, Joshua B Tenenbaum, and Shuran Song. Densephysnet: Learning dense physical object representations via multi-step dynamic interactions. *arXiv preprint arXiv:1906.03853*, 2019. 3
- [91] Tianju Xue, Shuheng Liao, Zhengtao Gan, Chanwook Park, Xiaoyu Xie, Wing Kam Liu, and Jian Cao. Jax-fem: A differentiable gpu-accelerated 3d finite element solver for automatic inverse design and mechanistic data science. *Computer Physics Communications*, 291:108802, 2023. 2
- [92] Zeyu Yang, Hongye Yang, Zijie Pan, and Li Zhang. Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting. *arXiv preprint arXiv:2310.10642*, 2023. 2
- [93] 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*, pages 20331–20341, 2024. 3, 4
- [94] Nianjin Ye, Chuan Wang, Haoqiang Fan, and Shuaicheng Liu. Motion basis learning for unsupervised deep homography estimation with subspace projection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 13117–13125, 2021. 2
- [95] Heng Yu, Joel Julin, Zoltan A Milacski, Koichiro Niinuma, and László A Jeni. Dylin: Making light field networks dynamic. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12397–12406, 2023. 2
- [96] Heng Yu, Joel Julin, Zoltán Á Milacski, Koichiro Niinuma, and László A Jeni. Cogs: Controllable gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21624–21633, 2024. 2
- [97] Vladimir Yugay, Yue Li, Theo Gevers, and Martin R Oswald. Gaussian-slam: Photo-realistic dense slam with gaussian splatting. *arXiv preprint arXiv:2312.10070*, 2023. 4
- [98] Tianyuan Zhang, Hong-Xing Yu, Rundui Wu, Brandon Y Feng, Changxi Zheng, Noah Snively, Jiajun Wu, and William T Freeman. Physdreamer: Physics-based interaction with 3d objects via video generation. In *European Conference on Computer Vision*, pages 388–406. Springer, 2025. 3

- [99] Dongzhe Zheng, Siqiong Yao, Wenqiang Xu, and Cewu Lu. Differentiable cloth parameter identification and state estimation in manipulation. *IEEE Robotics and Automation Letters*, 2024. [3](#)
- [100] 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*, pages 407–423. Springer, 2025. [2](#), [3](#), [4](#), [5](#), [6](#)
- [101] Xiaowei Zhou, Can Yang, and Weichuan Yu. Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE transactions on pattern analysis and machine intelligence*, 35(3):597–610, 2012. [2](#)