Grounding Creativity in Physics: A Brief Survey of Physical Priors in AIGC

Siwei Meng[1] , **Yawei Luo**[2✉] , **Ping Liu**[1✉]

[1]Department of Computer Science, University of Nevada, Reno
[2]School of Software Technology, Zhejiang University
siweim@unr.edu, yaweiluo@zju.edu.cn, pino.pingliu@gmail.com

Abstract

Recent advancements in AI-generated content have significantly improved the realism of 3D and 4D generation. However, most existing methods prioritize appearance consistency while neglecting underlying physical principles, leading to artifacts such as unrealistic deformations, unstable dynamics, and implausible objects interactions. Incorporating physics priors into generative models has become a crucial research direction to enhance structural integrity and motion realism. This survey provides a review of physics-aware generative methods, systematically analyzing how physical constraints are integrated into 3D and 4D generation. First, we examine recent works in incorporating physical priors into static and dynamic 3D generation, categorizing methods based on representation types, including vision-based, NeRF-based, and Gaussian Splatting-based approaches. Second, we explore emerging techniques in 4D generation, focusing on methods that model temporal dynamics with physical simulations. Finally, we conduct a comparative analysis of major methods, highlighting their strengths, limitations, and suitability for different materials and motion dynamics. By presenting an in-depth analysis of physics-grounded AIGC, this survey aims to bridge the gap between generative models and physical realism, providing insights that inspire future research in physically consistent content generation.

1 Introduction

Recent advancements in AI-generated content (AIGC) have significantly enhanced 3D and 4D content generation, with applications spanning gaming, simulation, animation, and robotics. Traditional AI-driven 3D generation methods primarily focus on improving geometric fidelity and rendering efficiency, leveraging representations such as Neural Radiance Fields (NeRF) [Mildenhall *et al.*, 2020] and 3D Gaussian Splatting (GS) [Kerbl *et al.*, 2023]. More recent models, such as DreamFusion [Poole *et al.*, 2023], integrate Diffusion Models (DMs) to improve synthesis realism. However, these models primarily optimize for visual quality, often neglecting physical plausibility, leading to artifacts such as implausible deformations, unstable motion, and inconsistent object interactions.

The integration of physics priors into generative models is an emerging yet underdeveloped direction in 3D/4D generation. Most generative models are trained on datasets lacking explicit physical constraints, failing to capture material properties, object dynamics, and force interactions. Consequently, generated content often deviates from real-world physical laws, limiting its applicability in simulation-driven applications. To address this gap, recent research has explored differentiable physics-based models such as Material Point Method (MPM) [Jiang *et al.*, 2016; Hu *et al.*, 2018], Finite Element Method (FEM) [Zienkiewicz and Taylor, 2000], and differentiable physics engines into the generative pipeline. These approaches enable physics-informed content generation, ensuring structural integrity, dynamic realism, and physically consistent interactions. However, a systematic review of these physics-based advances in generative models remains absent.

Existing surveys on 3D and 4D generation primarily focus on three aspects (Table 1): (1) scene representations and rendering [Li *et al.*, 2025b], which discuss different 3D representations and rendering optimizations; (2) generative models [Cao *et al.*, 2023; Liu *et al.*, 2024b], which analyze synthesis techniques from text and images; and (3) applications and scalability [Li *et al.*, 2025b], covering gaming, animation, and robotics. However, none of these works systematically explore the role of physics priors in generative models. A recent survey [Liu *et al.*, 2025a] provides the closest discussion to our work, categorizing physics-aware generation into explicit physics simulation-based (PAG-E) and implicit physics-informed (PAG-I) methods. However, its focus remains on physics-aware video and 3D content generation, with limited discussion on 4D generation and dynamic scene modeling.

To bridge this gap, this survey provides an overview of physics-grounded generative models, categorizing recent advances into static 3D generation, dynamic 3D generation, and 4D generation. Specifically, we further organize existing methods based on representation types and generation paradigms, covering vision-based, NeRF-based, and GS-based dynamic 3D generation approaches (Figure 1). By synthesizing insights from physics-based simulation and generative models, we aim to establish a systematic perspective on integrating physics priors with AI-driven content generation, offering new di-

[✉] denotes the co-corresponding author.

Works	Focused Tasks	Physics Priors	4D Generation	Released Date
[Cao *et al.*, 2023]	Generative Models in AIGC	✗	✗	Mar 2023
[Li *et al.*, 2025b]	3D Generation	✗	✗	Jan 2024
[Liu *et al.*, 2024b]	3D Generation	✗	✗	Feb 2024
[Banerjee *et al.*, 2024]	Physics-informed Computer Vision Models	✓	✗	Oct 2024
[Liu *et al.*, 2025a]	Physical Simulation with Generative Models	✓	✗	Jan 2025
Ours	**Physics-aware AIGC**	✓	✓	**Feb 2025**

Table 1: Comparison between this work and previous surveys.

rections for research in this evolving field. A curated list of all related papers mentioned in this work can be found at: https://github.com/mengsiwei/Awesome-Physical-AIGC-lists.

Scope of This Survey

In this survey, we first explain the foundation of generative models for 3D content (Section 2.1), and introduce common 3D representations (Section 2.2) and 4D representations methods (Section 2.3). We also provide the introduction of main physical simulation techniques, including MPM, FEM, and DiffTaichi (Section 2.4). Then, we present a taxonomy of recent research that grounding physics in AIGC, categorizing it into three key areas: static 3D generation (Section 3.1), dynamic 3D generation (Section 3.2), and 4D generation (Section 3.3). Next, we present datasets (Section 4.1) and evaluation metrics (Section 4.2) in this field, and provide a detailed comparison results of several approaches on synthetic PAC-NeRF dataset in Table 3. Finally, our survey discuss non-negligible challenges and possible future directions for this exciting new area of research (Section 5).

2 Background

2.1 Generative Models for 3D Content Generation

Generative models have significantly advanced 3D content creation by using deep learning to synthesize realistic structures. Generative Adversarial Networks (GANs) [Goodfellow *et al.*, 2014] and Diffusion Models [Ho *et al.*, 2020] are two key approaches, each with distinct strengths. Originally designed for 2D generation, GANs have been adapted to 3D by incorporating voxels, point clouds, and meshes. The generator synthesizes 3D data from latent codes, while the discriminator distinguishes real from generated samples. Though capable of producing high-quality shapes, GANs suffer from mode collapse and training instability, issues mitigated by subsequent advances. Meanwhile, DMs provide a more stable alternative with superior sample diversity. They follow a two-step denoising process: first adding Gaussian noise in forward diffusion, then iteratively removing it in reverse. In 3D generation, these models employ diverse representations to produce high-resolution, geometrically consistent outputs [Poole *et al.*, 2023; Xu *et al.*, 2024; Liu *et al.*, 2024a].

2.2 3D Representations

3D representations include explicit and implicit methods for modeling and rendering complex objects and scenes.

Explicit Methods. Explicit methods including point clouds, meshes, and 3D Gaussian Splatting, offer diverse approaches to modeling and rendering complex objects and scenes. Point clouds represent objects as collections of discrete points with attributes like color and surface normals, with advanced methods like SynSin [Wiles *et al.*, 2020] and Neural Point-based Rendering [Dai *et al.*, 2020] leveraging differentiable pipelines for optimization. Meshes define objects through vertices, edges, and faces in polygon networks, enabling accurate shape description and property refinement through differentiable rendering. 3D Gaussian Splatting [Kerbl *et al.*, 2023] employs learnable 3D Gaussian kernels optimized via multi-view supervision, providing efficient real-time rendering capabilities. These methods, each with their unique attributes and techniques, serve different applications while contributing to the advancement of 3D modeling and rendering technology.

Implicit Methods. Implicit methods such as Signed Distance Fields (SDF) and Neural Radiance Fields (NeRF), define the shapes and boundaries of objects not through explicit geometric components, but via functions that describe spatial occupancy, enabling continuous and detailed descriptions of geometries for realistic visualizations and complex operations. Both techniques offer unique advantages: SDFs [Shim *et al.*, 2023] enable efficient rendering and precise geometric manipulations like blending and smoothing of surfaces, making them valuable for computer-aided design systems and dynamic simulations, whereas NeRFs [Mildenhall *et al.*, 2020] enable highly photorealistic rendering of 3D scenes from novel viewpoints by modeling scenes as continuous volumetric fields within a neural network. These complementary approaches provide developers and researchers with powerful tools for advancing the state of 3D rendering and geometric computation.

2.3 4D Representations

4D representations incorporate spatiotemporal information for dynamic scene synthesis and reconstruction. Three notable approaches have emerged in this field: K-Planes, D-NeRF, and 4D Gaussian Splatting (4DGS). K-Planes method represents 4D scenes by factorizing the 4D space into six planes, three for spatial dimensions and three for spatiotemporal variations [Fridovich-Keil *et al.*, 2023]. K-Plances encodes 4D information by separating static and dynamic components and achieves fast optimization without relying on MLP-based decoders. D-NeRF [Pumarola *et al.*, 2021] extends the NeRF framework by conditioning the radiance field on time as an additional dimension, modeling scene dynamics through time-dependent radiance fields and deformation fields to handle temporal changes. Splatting methods [Wu *et al.*, 2024;

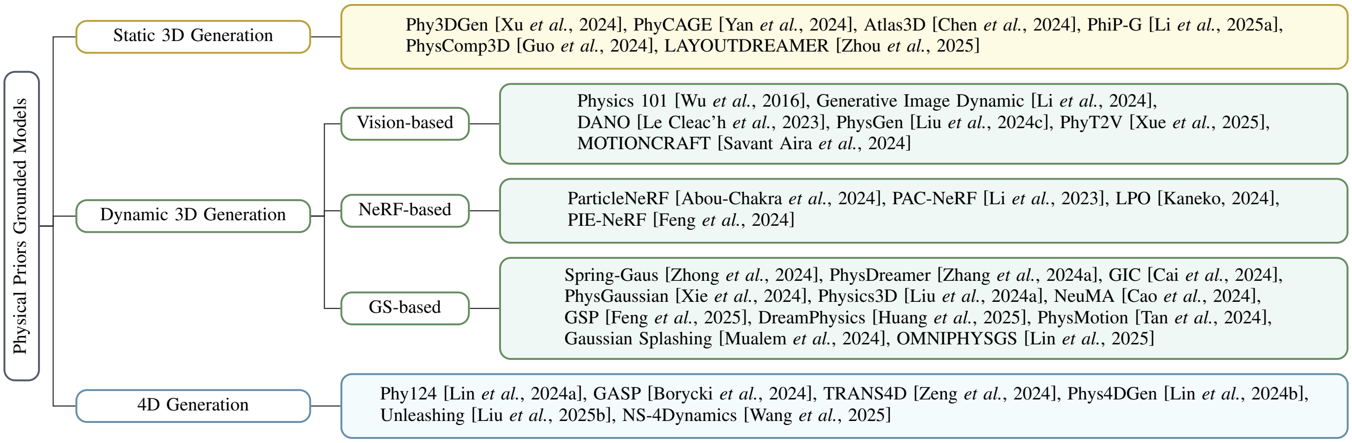
Physical Priors Grounded Models
Static 3D Generation
Phy3DGen [Xu et al., 2024], PhyCAGE [Yan et al., 2024], Atlas3D [Chen et al., 2024], PhiP-G [Li et al., 2025a], PhysComp3D [Guo et al., 2024], LAYOUTDREAMER [Zhou et al., 2025]
Dynamic 3D Generation
Vision-based
Physics 101 [Wu et al., 2016], Generative Image Dynamic [Li et al., 2024], DANO [Le Cleac'h et al., 2023], PhysGen [Liu et al., 2024c], PhyT2V [Xue et al., 2025], MOTIONCRAFT [Savant Aira et al., 2024]
NeRF-based
ParticleNeRF [Abou-Chakra et al., 2024], PAC-NeRF [Li et al., 2023], LPO [Kaneko, 2024], PIE-NeRF [Feng et al., 2024]
GS-based
Spring-Gaus [Zhong et al., 2024], PhysDreamer [Zhang et al., 2024a], GIC [Cai et al., 2024], PhysGaussian [Xie et al., 2024], Physics3D [Liu et al., 2024a], NeuMA [Cao et al., 2024], GSP [Feng et al., 2025], DreamPhysics [Huang et al., 2025], PhysMotion [Tan et al., 2024], Gaussian Splashing [Mualem et al., 2024], OMNIPHYSGS [Lin et al., 2025]
4D Generation
Phy124 [Lin et al., 2024a], GASP [Borycki et al., 2024], TRANS4D [Zeng et al., 2024], Phys4DGen [Lin et al., 2024b], Unleashing [Liu et al., 2025b], NS-4Dynamics [Wang et al., 2025]

Figure 1: A taxonomy of generative models grounded with physical priors.

Yang *et al.*, 2024; Zhang *et al.*, 2024b] propose two primary approaches: deformation-based transformation of 3D Gaussian kernels in image space, and direct integration of the time dimension into 3D Gaussian kernels for temporal-spatial coherence. These 4DGS methods offer lightweight representations of spatiotemporal information while maintaining high-quality dynamic scene reconstructions through the compactness of Gaussian kernels.

2.4 Physical Simulation

While 3D and 4D representations focus on modeling static and dynamic scenes respectively, physical simulation methods are essential for understanding and predicting how these representations evolve under physical laws and material properties. The Material Point Method (MPM), Finite Element Method (FEM), and DiffTaichi framework employ different approaches to model and compute such dynamic material behaviors across various applications.

Material Point Method. MPM combines Lagrangian material points with Eulerian Cartesian grids, where material properties like density and velocity are initially stored in particles and then interpolated onto a stationary grid [Jiang *et al.*, 2016]. In MPM, material is viewed as collections of particles with properties such as density, velocity, and force as continuous functions of position. These particles encode local material information and move through a stationary Eulerian grid, which facilitates the computation of spatial derivatives and enforces physical laws. MPM simulation computes the dynamics of continuum materials through a three-stage process: particle-to-grid transfer, grid computation, and grid-to-particle transfer. Initially, material properties are stored in Lagrangian particles and interpolated to a fixed Eulerian grid using shape functions. Then, the governing physical equations are solved on the grid including conservation of momentum and mass. After solving, the updated physical quantities are interpolated back from the grid to the particles, making the particles move and update their states. This hybrid approach enables the computation of spatial derivatives and enforcement of physical laws on the grid, after which the updated quantities are transferred back to the particles for state updates, creating an effective cycle for simulating complex material behaviors with high precision and stability.

Finite Element Method. FEM denotes a classical numerical approach which is widely used in engineering and computer graphics, tackles the simulation of deformable objects by subdividing larger systems into finite elements through space discretization [Zienkiewicz and Taylor, 2000]. FEM operates by formulating boundary value problems and approximating solutions through the minimization of associated error functions using variational calculus.

DiffTaichi. DiffTaichi provides comprehensive differentiable programming capabilities [Hu *et al.*, 2020]. It provides a two-scale automatic differentiation system that supports megakernels, imperative programming, and flexible indexing, and simplifies the implementation and optimization of simulayion methods, particularly MPM. This simulator demonstrates high productivity and efficient convergence in gradient-based learning and optimization tasks.

3 Categories of Physics Prior

To systematically study physics-grounded AIGC methods, we introduce a taxonomy based on the spatiotemporal modeling granularity and task semantics of the generated content. This categorization reflects how different approaches encode and enforce physical principles across varying spatial scales and temporal resolutions, which are fundamental to the generation of physically plausible outcomes. Building on this taxonomy, as illustrated in Figure 1, we divide existing methods into three categories: static 3D generation, dynamic 3D generation, and 4D generation. Static 3D generation focuses on spatial physical plausibility under fixed configurations, such as maintaining geometric integrity or modeling the mechanical properties of objects in equilibrium. Dynamic 3D generation extends this by introducing short-term temporal dynamics, modeling how objects evolve, move, or deform over time, often using explicit or learnable motion representations. 4D generation aims at continuous spatiotemporal synthesis, capturing long-range temporal dependencies and enforcing global physical consistency throughout dynamic evolution. This taxonomy not only aligns with the increasing complexity of physical modeling across tasks but also offers a principled lens through which to compare methods that operate at different levels of physical abstraction. In the following subsections, we discuss each of these three categories in details.

3.1 Static 3D Generation with Physics

Advancements in 3D generation have enhanced both geometric accuracy and physical plausibility. Early works [Tang *et al.*, 2024; Yi *et al.*, 2024] focused on high-fidelity shapes, while recent approaches integrate physical realism. Phy3DGen [Xu *et al.*, 2024] incorporates solid mechanics into a 3D diffusion model, using a differentiable physics network for refinement. During the optimization phase, Phy3DGen employs sample points for pretraining and implement a co-training strategy to simultaneously optimize the geometry and physics networks. By leveraging FEM [Zienkiewicz and Taylor, 2000] and a co-optimization strategy, it ensures both visual accuracy and physical consistency.

Generating physically coherent multi-component structures from a single image remains challenging. PhysComp3D [Guo *et al.*, 2024] optimizes physical compatibility by decomposing mechanical properties, external forces, and rest-shape geometry. Using plastic deformation parametrization and implicit differentiation, it refines object structures but relies on predefined material properties and mesh representations. Extending this, PhyCAGE [Yan *et al.*, 2024] employs Gaussian Splatting and physics-based SDS optimization to enhance efficiency. By generating multi-view images and refining Gaussian positions through simulation-enhanced SDS gradients, it improves the physical coherence of 3D assets.

Text-based 3D generation faces challenges in maintaining stability due to missing physical constraints. Atlas3D [Chen *et al.*, 2024] addresses this by integrating differentiable physics simulation into an SDS-based framework, predicting rigid body dynamics and enforcing standability and stable equilibrium losses for structural consistency. While effective for individual objects, it lacks scene-level physical interactions. To address this, PhiP-G [Li *et al.*, 2025a] introduces a physics-grounded text-to-3D scene generation framework, integrating a multi-agent text processor and a relational database-based physics pool for object interactions. Additionally, its physical magnet module applies vector approximations to align assets in semantically and physically consistent 3D scenes. Continuing this progress, LAYOUTDREAMER [Zhou *et al.*, 2025] leverages 3DGS to create compositional scenes by converting text into a direct scene graph, which guides objects layout and physical interactions. LAYOUTDREAMER excels in generating complex multi-object scene, allowing users to conveniently edit and expand disentangled scenes.

3.2 Dynamic 3D Generation

Dynamic 3D generation focuses on capturing 3D content motion over time. According to 3D representation categories, this subsection discuss three dominant paradigms: vision-based dynamic 3D generation, NeRF-based dynamic 3D generation, and GS-based dynamic 3D generation. Vision-based dynamic methods typically rely on information from video sequence or single image and infer the dynamic changes of 3D structure by analyzing the movements between consecutive frames. NeRF-based dynamic 3D generation expends NeRF representation and introduces time variables to represent motion deformation. GS-based generation methods represents the dynamic changes of 3D scenes through Gaussians which is suitable for handing the movements of both rigid and non-rigid objects.

Vision-based Dynamic 3D Generation

Several works explore motion modeling through frequency analysis and optical flow. Li et al. [Li *et al.*, 2024] introduce a frequency-domain motion model, integrating spatial and frequency attention within LDM UNet for motion prediction. Their feature pyramid softmax-splatting strategy generates, allowing the model to generate future frames based on predicted spectral volumes. However, this method only keeps 16 low-frequency components to reduce computational complexity, which causes distortion when showing thin objects or large amounts of content. MOTIONCRAFT [Savant Aira *et al.*, 2024] extends this by mapping optical flow fields between RGB and latent space in Stable Diffusion, using physics simulators to animate input images. This approach improves zero-shot video generation with realistic fluid dynamics, rigid motion, and multi-agent interactions, avoiding extensive data and computation costs.

To ensure physically consistent motion, differentiable physics-based approaches explicitly model object dynamics. DANO [Le Cleac'h *et al.*, 2023] estimates mass, center of mass, and inertia matrix of the object from a density field, and then introduces the differentiable contact model to calculate the friction force generated by the motion collision of the object. Their proposed Monte Carlo method computes the contact force by sampling the neural density field and computing the gradient to approximate the outward normal. This approach can interact with existing simulation engines to optimize the control trajectory of neural objects. However, DANO only considers rigid body objects, while the simulation of articulated and soft body contacts remains a research challenge. Physics 101 [Wu *et al.*, 2016] instead infers physical properties from video, categorizing them into intrinsic (unobservable) and descriptive (visually detectable) attributes. By combining visual recognition, physics interpretation, and world simulation, it extracts properties from unlabeled video scenarios, such as sliding, falling, and floating.

Beyond explicit physics modeling, PhysGen [Liu *et al.*, 2024c] integrates model-based physics simulation into image-to-video generation. It first employs GPT-4V [OpenAI, 2023] for scene perception, extracting materials, object composition, and physical attributes from single input image. Rigid-body physics then simulates object dynamics, followed by motion-guided rendering for realistic and controllable training-free video synthesis. However, models still struggle to understand the interaction and movement of multiple objects. To enforce physics-grounded video generation, PhyT2V [Xue *et al.*, 2025] embeds physical rules into the text prompts by local chain-of-thought (CoT) and global step-back reasoning. PhyT2V expends the exsiting video generation models to out-of-distribution domains with sufficient and appropriate contexts via LLM, enabling significant quality improvement.

NeRF-based Dynamic 3D Generation

Dynamic NeRF methods [Fang *et al.*, 2022; Gao *et al.*, 2021; Park *et al.*, 2021] require full image sequences for training, limiting adaptability to dynamic scenes. ParticleNeRF [Abou-Chakra *et al.*, 2024] addresses this by introducing a particle-based encoding, updating particle positions via backpropagated photometric loss. They use a lightweight physics system to manage particle collisions and motion, enabling continuous adaptation to objects deformation. Unlike traditional static grid encodings, ParticleNeRF allows faster adaptation to dynamic scenes involving rigid bodies, articulated objects, and deformable entities with higher fidelity and efficiency. Further, PAC-NeRF [Li *et al.*, 2023] extends NeRF's capabilities by estimating geometry and physical parameters from multi-view videos using a hybrid Eulerian-Lagrangian representation. It combines NeRF density fields with MPM physical simulation, enabling differentiable rendering and simulation without predefined object structures.

Despite its advantages, PAC-NeRF relies on first-frame grid representations, restricting optimization across sequences. LPO [Kaneko, 2024] overcomes this by introducing Lagrangian particle optimization, which refines particle positions and features across entire videos. Additionally, it incorporates physical constraints from the MPM to iteratively correct geometric structures from video sequences, which addresses the bottleneck of geometry learning in sparse-view settings.

While these methods enhance motion modeling, NeRF remains computationally demanding for elastic dynamics simulation, especially for complex deformable objects. PIE-NeRF [Feng *et al.*, 2024] addresses this by integrating physics-based simulation into NeRF, employing quadratic generalized least squares (Q-GMLS) [Martin *et al.*, 2010] for nonlinear dynamics and large deformations. Using spatial reduction and real-time neural graphics primitives (NGP) [Müller *et al.*, 2022], it enables interactive manipulation and real-time rendering, achieving both physical accuracy and visual fidelity.

GS-based Dynamic 3D Generation

Traditional NeRF-based dynamic models assume known material properties, limiting abilities to simulate heterogeneous objects. Spring-Gaus [Zhong *et al.*, 2024] introduces a 3D Spring-Mass model with learnable mass points and springs, enabling elastic reconstruction from multi-view videos. By integrating Gaussian kernels with the Spring-Mass model, it decouples appearance and geometry, efficiently capturing geometry, motion, and physical properties.

To model action-conditioned dynamics, PhysDreamer [Zhang *et al.*, 2024a] learns dynamics priors from video generation models. The framework distills dynamics priors through differentiable Material Point Method (MPM) simulation and rendering, optimizing physical parameters such as the Young's modulus field, which controls object stiffness. Additionally, to enhance computational efficiency, PhysDreamer introduces an accelerated simulation strategy by employing K-means clustering, creating "driving particles" that reduce computational overhead while maintaining physical fidelity. Extending this, Physics3D [Liu *et al.*, 2024a] integrates viscoelastic MPM with Score Distillation Sampling (SDS) to simulate a wide range of materials, from elastic to plastic behaviors. PhysMotion [Tan *et al.*, 2024] introduces physics-based simulation MPM to guide intermediate 3D representations from a single image. Unlike PhysDreamer and Physics3D, which focus on learning physical properties from video diffusion models, PhysMotion combines 3DGS and refines the coarse simulation using a 2D image DM with cross-frame attention.

For fluid-solid coupling, Gaussian Splashing (GSP) [Feng *et al.*, 2025] utilizes 3D Gaussian kernels as particles, tracking fluid surfaces and interpolating deformations onto Gaussian kernels. By incorporating surface tension and specular effects, it achieves physically consistent rendering of solid-fluid interactions. In underwater dynamic modeling, Gaussian Splashing for underwater imagery [Mualem *et al.*, 2024] extends 3D Gaussian Splatting with learnable backscatter and attenuation effects, introducing a depth-aware rasterization pipeline for robust underwater reconstruction. Additionally, Gaussian Splashing also provides a novel underwater dataset TableDB, consisting 172 images with a resolution of 1384 × 918, which are unbounded in camera-to-scene distances.

To address discontinuities in video diffusion-based dynamics, DreamPhysics [Huang *et al.*, 2025] introduces Motion Distillation Sampling (MDS), improving motion realism over standard SDS. It employs KAN-based material fields and frame boosting, enabling text- and image-conditioned physical simulations without requiring ground truth videos. PhysGaussian [Xie *et al.*, 2024] further bridges Newtonian dynamics and 3D Gaussian kernels, leveraging continuum mechanics-driven deformation models to align physical simulation and visual rendering, handling a variety of materials including metals and non-Newtonian fluids. However, PhysGaussian still require manual tuning physical properties, which is time-consuming and highly relies on expert knowledge. To enhance the flexibility and generalizability of the constitutive model, OMNIPHYSGS [Lin *et al.*, 2025] introduces a general physics-based 3D dynamic synthesis framework by treating each 3D asset as a collection of learnable constitutive 3D Gaussians. Then, OMNIPHYSGS leverages a pre-trained video diffusion model [Poole *et al.*, 2023] to supervise the estimation of material weighting factors, enabling the synthesis of physically plausible dynamics across a broad spectrum of materials.

For generalizable modeling, NeuMA [Cao *et al.*, 2024] integrates physical laws with learned corrections, introducing Neural Constitutive Laws (NCLaw) for adaptive physics simulation. It incorporates Particle-GS, which binds simulation particles to Gaussian kernels, improving visual-grounded dynamic modeling. GIC (Gaussian-Informed Continuum) [Cai *et al.*, 2024] further refines physical property estimation, using motion decomposition networks and a coarse-to-fine density field generation strategy, enhancing dynamic scene reconstruction and continuum mechanics-based physical optimization.

3.3 4D Generation

4D generation aims to reconstruct 3D presentations from input conditions such as text, image, and video sequences. Currently, most 4D generation works heavily rely on powerful 3D generation works, which have high computational costs and face challenges in understanding real-world dynamics.

To improve 4D generation models modeling and understanding ability, Phy124 [Lin *et al.*, 2024a] eliminates diffusion models, enabling fast, physics-driven 4D content generation by converting an image into a 3D Gaussian representation and applying MPM to simulate Gaussian field dynamics. First, it transforms the input image into a static 3D Gaussian representation. In the second stage, the Material Point Method (MPM) [Jiang *et al.*, 2016; Hu *et al.*, 2018] is used to simulate the physical dynamics of the 3D Gaussian field, where each Gaussian kernel is treated as a discrete material point with physical properties such as mass, density, and volume. By using MPM for physics-grounded dynamics, Phy124 eliminates the need for diffusion models in 4D generation, significantly speeding up the process.

Similarly to Phy124, which focuses on 4D content generation by modeling dynamic 3D objects evolving over time, GASP [Borycki *et al.*, 2024] extends Gaussian-based modeling for real-time 3D simulation. They leveraged the GaMeS framework [Waczyńska *et al.*, 2024], utilizing flat Gaussian representations to map Gaussian components into triangle face representations, treating each 3D point as a discrete entity. Their method supports real-time simulations, efficiently handling both static and dynamic 3D scenes. Phys4DGen [Lin *et al.*, 2024b] further extends Phy124 by integrating a PPM for material-aware physics simulation, segmenting material groups from images and inferring properties via GPT-4o [OpenAI, 2023], enhancing recognition and simulation fidelity.

For text-to-4D synthesis, TRANS4D [Zeng *et al.*, 2024] introduces multi-modal large language model priors to generate detailed and physically plausible 4D scene data from original textual input. Based on the planning 4D scene data, they calculate the 3DGS transformation function at each timestep. After obtaining the physics-aware planning, they use a geometry-aware Transition Network to process them to produce the final output which serves as a reference for 4D transition. TRANS4D enables MLLMs to initialize realistic 4D scenes with multiple interacting objects and generate geometric-awareness transitions, helping to generate more realistic 4D game scenes. Similarly, Liu et al. [Liu *et al.*, 2025b] propose a comprehensive 4D simulation framework, integrating GPT-4 material inference with optical flow-based loss for optimizing physical properties. By combining multi-modal foundation models and video diffusion models, they achieve high-fidelity dynamic simulation across diverse material types.

Beyond generation, understanding and reasoning about 4D dynamic scenes remain critical for enabling physics-aware AIGC. NS-4Dynamics [Wang *et al.*, 2025] serves as the first neural-symbolic models for explicit 4D scene reconstruction. NS-4Dynamics incorporates physical priors into the scene parsing process, further enhancing the realism and accuracy of 4D dynamic generation. Additionally, this work proposes the SuperCLEVR-Physics dataset, designed for video question answering tasks focused on object dynamics and interaction properties, further bridging the gap between 4D generative models and physics-aware reasoning.

4 Benchmarks

In this section, we discuss the benchmarks, including datasets and evaluation metrics used to evaluate physics-aware generation models. Additionally, we present a quantitative comparison of the state-of-the-art physics-aware dynamic 3D generation models on synthetic dataset.

4.1 Datasets

Physics-aware generative models are typically evaluated using two types of datasets: synthetic and real-world. Each type serves a distinct role in assessing model robustness and generalization. Synthetic datasets provide controlled dynamic environments with access to ground-truth physical properties such as object mass, density, and material parameters. These datasets are particularly well-suited for benchmarking models in complex scenarios involving multi-object collisions, elastic or plastic deformations, and non-Newtonian fluid behavior. In contrast, real-world datasets capture the diversity and unpredictability of natural scenes, offering a means to evaluate a model's ability to transfer from simulation to reality and generalize to unconstrained, noisy conditions.

Synthetic Dataset

Synthetic PAC-NeRF consists of 9 instances with deformable objects, plastics, granular, metal, and Newtonian/Non-Newtonian fluids [Li *et al.*, 2023]. Each scene depicts the process of objects falling freely, colliding, and bouncing back, captured from 11 viewpoints with ground truth data generated by the MLS-MPM framework [Hu *et al.*, 2018]. Synthetic Spring-Gaus includes fourteen 3D models [Zhong *et al.*, 2024], all of which are generated from PAC-NeRF [Li *et al.*, 2023] and OmniObject3D [Wu *et al.*, 2023] approaches. This dataset features elastic object sequences captured from 10 viewpoints across 30 frames at a resolution of 512×512.

Real-world Dataset

Real-world PAC-NeRF captures a deformation ball falling onto a table using a capture system comprising four synchronized Intel RealSense D455 cameras. The real-world data in PAC-NeRF are RGB images at a resolution of 640×480 and at a rate of 60 frames per second. Real-world Spring-Gaus contains both static scenes and dynamic multi-view videos [Zhong *et al.*, 2024]. Static scenes include 50-70 images from various viewpoints, while dynamic scenes are recorded from three viewpoints at a resolution of 1980×1080. Physics 101 contains $17,408$ videos of 101 objects made of 15 different materials [Wu *et al.*, 2016]. Each material category has 4 to 12 objects of different sizes and colors, with recorded physical properties such as mass, volume, and density. VIDEOPHY evaluates whether generated videos adhere to physical commonsense [Bansal *et al.*, 2025]. It comprises 688 human-verified high-quality captions, with 344 prompts for the test set and 344 prompts for train set. This dataset maintains a balanced distribution of the state of matter and complexity across both sets.

Physical Properties

From the definition in MPM simulator [Jiang *et al.*, 2016], seven material types include Newtonian and non-Newtonian fluids, elasticity, plasticine, metal, foam, and sand are characterized by specific physical parameters (Table 2). For instance, Newtonian fluids are defined by viscosity and bulk modulus, while Non-Newtonian fluids also consider shear modulus and yield stress. For the synthetic dataset, each object's material

Material Types	Physical Properties
Newtonian fluid	Fluid viscosity μ, Bulk modulus κ
Non-Newtonian fluid	Shear modulus μ, Bulk modulus κ, Yield stress τ_Y, Plastic viscosity η
Elasticity	Young's modulus E, Poisson's ratio ν
Plasticine	Young's modulus E, Poisson's ratio ν, Yield stress τ_Y
Metal	Young's modulus E, Poisson's ratio ν, Yield stress τ_Y
Foam	Young's modulus E, Poisson's ratio ν, Plastic viscosity η
Sand	Friction angle θ_{fric}

Table 2: A taxonomy of seven common material types simulated with various physical properties.

type is predefined, while the real-world dataset usually lack the information of object material.

Evaluation Metrics

Evaluating physics-prior guided generation methods requires assessing three key aspects: semantic coherency, physical consistency, and model performance, to ensure alignment with physical principles. For semantic coherency, we adopt the CLIP score [Radford *et al.*, 2021] to quantify the alignment between generated content and textual descriptions by computing cross-modal embeddings. Additionally, we use the Semantic Adherence (SA) metric, which assigns a binary score ($\text{SA} \in 0, 1$), where $\text{SA} = 1$ indicates that the generated video is semantically grounded in its corresponding text caption. For physical consistency, we report the Mean Absolute Error (MAE) between the generated physical properties and ground-truth values. We also introduce the Physical Commonsense (PC) metric, a binary indicator ($\text{PC} \in 0, 1$) denoting whether the generated content adheres to intuitive physical laws commonly understood by humans. Lastly, for overall model performance, we use Peak Signal-to-Noise Ratio (PSNR) to evaluate the quality of generated visual content, while Structural Similarity Index Measure (SSIM) quantifies the perceptual similarity between generated and reference images, especially in dynamic scenes.

Comparison Results

We select the state-of-the-art approaches in physics-aware dynamic 3D generation, and each utilizing different representations and architectures to integrate physical priors. Specifically, PAC-NeRF [Li *et al.*, 2023] and LPO [Kaneko, 2024] are NeRF-based approaches, while GIC [Cai *et al.*, 2024] and unleashing [Liu *et al.*, 2025b] are GS-based approaches. Table 3 provides a comparative analysis of four physics-aware 3D generation methods on the Synthetic PAC-NeRF dataset, which includes nine objects composed of Newtonian fluids, Non-Newtonian fluids, elasticity-based materials, plasticine, and sand. Besides, the dataset provides ground-truth physical simulation data, allowing for a quantitative evaluation of these methods based on key physical parameters such as viscosity, yield stress, sheer modulus, bulk modulus, elasticity modulus, Poisson's ratio, and friction angle.

The comparison reveals that Unleashing consistently achieves the lowest AE for most fluid and plasticine materials, including Newtonian fluids (Droplet, Letter), Non-Newtonian fluids (Toothpaste), and plasticine (Playdoh, Cat). LPO demonstrates superior performance in yield stress modeling for Non-Newtonian fluids, achieving the best bulk modulus ($\Delta\kappa$) results for Cream. GIC is the strongest performer in elasticity-based materials, obtaining the lowest AE in Torus and Bird (ΔE, $\Delta\nu$). Unleashing is competitive in estimating sand friction angle ($\Delta\theta_{\text{fric}}$) for the Trophy object, indicating its effectiveness.

Across different methods, PAC-NeRF exhibits higher errors, particularly in yield stress ($\Delta\tau_Y$) and elasticity (ΔE), probably due to its fixed first-frame optimization. LPO refines particle positions, improving performance in Non-Newtonian fluids but struggling with elasticity-based materials. GIC excels in elasticity modeling, achieving the lowest ΔE values for Torus and Bird. Unleashing demonstrates the most robust overall performance, excelling in Newtonian and Non-Newtonian fluids as well as plasticine, showcasing its effectiveness in capturing complex material behaviors.

Overall, the results highlight the importance of integrating physics-aware priors in generative models, as different methods exhibit distinct strengths depending on material properties. Unleashing demonstrates the most robust performance across various material types, particularly in modeling fluid dynamics and plasticine deformations. GIC excels in elasticity-based materials due to its Gaussian Splatting-based scene reconstruction, effectively capturing non-rigid deformations. LPO refines PAC-NeRF by optimizing particle positions in time-varying sequences, improving its accuracy in Non-Newtonian fluids, though it struggles with elasticity modeling. PAC-NeRF, despite pioneering a hybrid Eulerian-Lagrangian representation, suffers from larger errors due to its reliance on first-frame initialization, limiting its adaptability to dynamic materials.

5 Challenges and Future Directions

Despite significant progress in physics-grounded generative models, several challenges remain that hinder their broader applicability and effectiveness. Current approaches often struggle with accurately modeling physical interactions, maintaining long-term dynamic consistency, and generalizing to diverse materials and real-world scenarios. Addressing these limitations requires advancements in dataset construction, model design, and integration with physics-aware reasoning.

One major challenge is the lack of accurate physical parameter annotations in existing datasets, which limits the scalability of data-driven generative models. Most datasets lack expert-annotated physical properties and fail to cover a diverse range of material behaviors and dynamic interactions. While recent works have explored leveraging large language models (LLMs)

Materials	Objects	GT	Parameters	NeRF-based Methods		GS-based Methods	
				PAC-NeRF	LPO	GIC	Unleashing
Newtonian fluid	Droplet	200	$\Delta\mu$	9	41	1	**0.89**
		1.0×10^5	$\Delta\kappa$	8.0×10^3	2.8×10^4	8.2×10^4	**3.0×10^3**
	Letter	100	$\Delta\mu$	16.15	**2**	4.95	2.27
		1.0×10^5	$\Delta\kappa$	3.5×10^4	1.3×10^4	**0**	7.0×10^3
Non-Newtonian fluid	Cream	1.0×10^4	$\Delta\mu$	1.11×10^5	2.6×10^3	**3.0×10^2**	1.21×10^4
		1.0×10^6	$\Delta\kappa$	5.7×10^5	**3.2×10^5**	4.8×10^5	5.3×10^5
		3.0×10^3	$\Delta\tau_Y$	160	40	**20**	1730
		10	$\Delta\eta$	8000	**0.8**	3.4	7.5
	Toothpaste	5.0×10^3	$\Delta\mu$	1510	340	810	**70**
		1.0×10^5	$\Delta\kappa$	5.122×10^4	7.88×10^4	7.6×10^4	**1.3×10^4**
		200	$\Delta\tau_Y$	28	38	**26**	46
		10	$\Delta\eta$	0.23	**0.20**	0.90	28.99
Elasticity	Torus	1.0×10^6	ΔE	4.0×10^4	9.0×10^4	**1.0×10^4**	3.9×10^4
		0.3	$\Delta\nu$	0.022	0.007	**0.005**	0.297
	Bird	3.0×10^5	ΔE	2.2×10^5	1.9×10^4	**8.0×10^3**	1.2×10^4
		0.3	$\Delta\nu$	0.027	0.047	**0.016**	0.171
Plasticine	Playdoh	2.0×10^6	ΔE	1.84×10^6	7.2×10^5	4.2×10^5	**2.11×10^5**
		0.3	$\Delta\nu$	0.028	0.063	**0.022**	0.098
		1.54×10^4	$\Delta\tau_Y$	150	7600	**20**	41
	Cat	1.0×10^6	ΔE	8.39×10^5	8.03×10^5	**2.0×10^3**	3.87×10^5
		0.3	$\Delta\nu$	0.007	**0.003**	0.004	0.124
		3.85×10^3	$\Delta\tau_Y$	280	650	**90**	850
Sand	Trophy	40	$\Delta\theta_{fric}$	3.9°	2.25°	2.0°	**0.5°**

Table 3: Performance comparison of dynamic 3D generation methods on the synthetic PAC-NeRF dataset.

for physical reasoning [Liu *et al.*, 2024c; Lin *et al.*, 2024b; Liu *et al.*, 2025b], these approaches struggle with capturing intricate physical dependencies. The effectiveness of LLMs in physical reasoning heavily depends on well-constructed prompts, and current models [OpenAI, 2023] exhibit inconsistency in physics-related tasks. A promising research direction is to enhance learning and reasoning capabilities within LLMs.

Another fundamental challenge lies in embedding physical constraints into generative models, particularly in understanding the relationship between dynamic movements and physical laws. While properties such as mass, velocity, and friction play a crucial role in realistic motion synthesis, effectively integrating these parameters into neural representations remains an open problem. 3D Gaussian Splatting (3DGS)-based approaches [Xie *et al.*, 2024; Cai *et al.*, 2024; Borycki *et al.*, 2024; Liu *et al.*, 2025b] attempt to encode physical parameters by extending Gaussian kernels with additional dimensions and incorporating physics-aware optimization in differentiable networks. However, for 4D dynamic generation, existing modeling techniques are still underdeveloped, and further exploration is needed to improve their ability to capture complex object deformations and temporal consistency.

Future research should focus on enhancing the integration of physics-aware priors in generative models, improving their adaptability to diverse materials and dynamic environments. For example, advancing Sim2Real transfer and Embodied AI will further bridge the gap between simulated and real-world interactions, enabling more physically consistent and generalizable generative models.

6 Conclusion

This survey reviews physics-grounded generative models for 3D and 4D content generation, categorizing methods based on their representation and generation paradigms. We analyze their strengths and limitations, highlighting how physical priors improve visual realism and structural consistency. Quantitative comparisons reveal gaps in physical accuracy and generalization. Despite progress, challenges remain in modeling multi-object complex physical interactions and enhancing dataset diversity in this area. Future research should focus on improving physics-aware learning, integrating differentiable physics, and advancing Sim2Real transfer to bridge the gap between simulation and real-world applications.

References

[Abou-Chakra *et al.*, 2024] Jad Abou-Chakra, Feras Dayoub, and Niko Sünderhauf. ParticleNeRF: A particle-based encoding for online neural radiance fields. In *WACV*, 2024.

[Banerjee *et al.*, 2024] Chayan Banerjee, Kien Nguyen, Clinton Fookes, and Karniadakis George. Physics-informed computer vision: A review and perspectives. *ACM COMPUT SURV*, 2024.

[Bansal *et al.*, 2025] Hritik Bansal, Zongyu Lin, Tianyi Xie, Zeshun Zong, Michal Yarom, Yonatan Bitton, Chenfanfu Jiang, Yizhou Sun, Kai-Wei Chang, and Aditya Grover. VideoPhy: Evaluating physical commonsense for video generation. *ICLR*, 2025.

[Borycki *et al.*, 2024] 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.

[Cai *et al.*, 2024] 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. *NeurIPS*, 2024.

[Cao *et al.*, 2023] Yihan Cao, Siyu Li, Yixin Liu, Zhiling Yan, Yutong Dai, Philip S Yu, and Lichao Sun. A comprehensive survey of ai-generated content (AIGC): A history of generative ai from gan to chatgpt. *arXiv preprint arXiv:2303.04226*, 2023.

[Cao *et al.*, 2024] Junyi Cao, Shanyan Guan, Yanhao Ge, Wei Li, Xiaokang Yang, and Chao Ma. NeuMA: Neural material adaptor for visual grounding of intrinsic dynamics. *NeurIPS*, 2024.

[Chen *et al.*, 2024] Yunuo Chen, Tianyi Xie, Zeshun Zong, Xuan Li, Feng Gao, Yin Yang, Ying Nian Wu, and Chenfanfu Jiang. Atlas3D: Physically constrained self-supporting text-to-3d for simulation and fabrication. *NeurIPS*, 2024.

[Dai *et al.*, 2020] Peng Dai, Yinda Zhang, Zhuwen Li, Shuaicheng Liu, and Bing Zeng. Neural point cloud rendering via multi-plane projection. In *CVPR*, 2020.

[Fang *et al.*, 2022] Jiemin Fang, Taoran Yi, Xinggang Wang, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Matthias Nießner, and Qi Tian. Fast dynamic radiance fields with time-aware neural voxels. In *SIGGRAPH*, 2022.

[Feng *et al.*, 2024] Yutao Feng, Yintong Shang, Xuan Li, Tianjia Shao, Chenfanfu Jiang, and Yin Yang. PIE-NeRF: Physics-based interactive elastodynamics with nerf. In *CVPR*, 2024.

[Feng *et al.*, 2025] Yutao Feng, Xiang Feng, Yintong Shang, Ying Jiang, Chang Yu, Zeshun Zong, Tianjia Shao, Hongzhi Wu, Kun Zhou, Chenfanfu Jiang, and Yin Yang. Gaussian Splashing: Unified particles for versatile motion synthesis and rendering. *CVPR*, 2025.

[Fridovich-Keil *et al.*, 2023] Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-Planes: Explicit radiance fields in space, time, and appearance. In *CVPR*, 2023.

[Gao *et al.*, 2021] Chen Gao, Ayush Saraf, Johannes Kopf, and Jia-Bin Huang. Dynamic view synthesis from dynamic monocular video. In *ICCV*, 2021.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *NeurIPS*, 2014.

[Guo *et al.*, 2024] Minghao Guo, Bohan Wang, Pingchuan Ma, Tianyuan Zhang, Crystal Elaine Owens, Chuang Gan, Joshua B Tenenbaum, Kaiming He, and Wojciech Matusik. Physically compatible 3d object modeling from a single image. *NeurIPS*, 2024.

[Ho *et al.*, 2020] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 2020.

[Hu *et al.*, 2018] 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. *TOG*, 2018.

[Hu *et al.*, 2020] Yuanming Hu, Luke Anderson, Tzu-Mao Li, Qi Sun, Nathan Carr, Jonathan Ragan-Kelley, and Fredo Durand. DiffTaichi: Differentiable programming for physical simulation. In *ICLR*, 2020.

[Huang *et al.*, 2025] Tianyu Huang, Yihan Zeng, Hui Li, Wangmeng Zuo, and Rynson WH Lau. DreamPhysics: Learning physical properties of dynamic 3d gaussians with video diffusion priors. *AAAI*, 2025.

[Jiang *et al.*, 2016] Chenfanfu Jiang, Craig Schroeder, Joseph Teran, Alexey Stomakhin, and Andrew Selle. The material point method for simulating continuum materials. In *SIGGRAPH*. 2016.

[Kaneko, 2024] Takuhiro Kaneko. Improving physics-augmented continuum neural radiance field-based geometry-agnostic system identification with lagrangian particle optimization. In *CVPR*, 2024.

[Kerbl *et al.*, 2023] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *TOG*, 2023.

[Le Cleac'h *et al.*, 2023] Simon Le Cleac'h, Hong-Xing Yu, Michelle Guo, Taylor Howell, Ruohan Gao, Jiajun Wu, Zachary Manchester, and Mac Schwager. Differentiable physics simulation of dynamics-augmented neural objects. *RAL*, 2023.

[Li *et al.*, 2023] Xuan Li, Yi-Ling Qiao, Peter Yichen Chen, Krishna Murthy Jatavallabhula, Ming Lin, Chenfanfu Jiang, and Chuang Gan. PAC-NeRF: Physics augmented continuum neural radiance fields for geometry-agnostic system identification. In *ICLR*, 2023.

[Li *et al.*, 2024] Zhengqi Li, Richard Tucker, Noah Snavely, and Aleksander Holynski. Generative Image Dynamics. In *CVPR*, 2024.

[Li *et al.*, 2025a] Qixuan Li, Chao Wang, Zongjin He, and Yan Peng. PhiP-G: Physics-guided text-to-3d compositional scene generation. *arXiv preprint arXiv:2502.00708*, 2025.

[Li *et al.*, 2025b] Xiaoyu Li, Qi Zhang, Di Kang, Weihao Cheng, Yiming Gao, Jingbo Zhang, Zhihao Liang, Jing Liao, Yan-Pei Cao, and Ying Shan. Advances in 3d generation: A survey. *IJCV*, 2025.

[Lin *et al.*, 2024a] Jiajing Lin, Zhenzhong Wang, Yongjie Hou, Yuzhou Tang, and Min Jiang. Phy124: Fast physics-driven 4d content generation from a single image. *arXiv preprint arXiv:2409.07179*, 2024.

[Lin *et al.*, 2024b] Jiajing Lin, Zhenzhong Wang, Shu Jiang, Yongjie Hou, and Min Jiang. Phys4DGen: A physics-driven framework for controllable and efficient 4d content generation from a single image. *arXiv preprint arXiv:2411.16800*, 2024.

[Lin *et al.*, 2025] Yuchen Lin, Chenguo Lin, Jianjin Xu, and Yadong Mu. OMNIPHYSGS: 3d constitutive gaussians for general physics-based dynamics generation. *ICLR*, 2025.

[Liu *et al.*, 2024a] 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.

[Liu *et al.*, 2024b] Jian Liu, Xiaoshui Huang, Tianyu Huang, Lu Chen, Yuenan Hou, Shixiang Tang, Ziwei Liu, Wanli Ouyang, Wangmeng Zuo, Junjun Jiang, et al. A comprehensive survey on 3d content generation. *arXiv preprint arXiv:2402.01166*, 2024.

[Liu *et al.*, 2024c] Shaowei Liu, Zhongzheng Ren, Saurabh Gupta, and Shenlong Wang. PhysGen: Rigid-body physics-grounded image-to-video generation. *ECCV*, 2024.

[Liu *et al.*, 2025a] Daochang Liu, Junyu Zhang, Anh-Dung Dinh, Eunbyung Park, Shichao Zhang, and Chang Xu. Generative physical ai in vision: A survey. *arXiv preprint arXiv:2501.10928*, 2025.

[Liu *et al.*, 2025b] Zhuoman Liu, Weicai Ye, Yan Luximon, Pengfei Wan, and Di Zhang. Unleashing the potential of multi-modal foundation models and video diffusion for 4d dynamic physical scene simulation. *CVPR*, 2025.

[Martin *et al.*, 2010] Sebastian Martin, Peter Kaufmann, Mario Botsch, Eitan Grinspun, and Markus Gross. Unified simulation of elastic rods, shells, and solids. *TOG*, 2010.

[Mildenhall *et al.*, 2020] 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. *ECCV*, 2020.

[Mualem *et al.*, 2024] Nir Mualem, Roy Amoyal, Oren Freifeld, and Derya Akkaynak. Gaussian splashing: Direct volumetric rendering underwater. *arXiv preprint arXiv:2411.19588*, 2024.

[Müller *et al.*, 2022] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *TOG*, 2022.

[OpenAI, 2023] OpenAI. GPT-4V(ision) System Card, 2023. https://openai.com/index/gpt-4o-system-card/.

[Park *et al.*, 2021] 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 *ICCV*, 2021.

[Poole *et al.*, 2023] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. DreamFusion: Text-to-3d using 2d diffusion. *ICLR*, 2023.

[Pumarola *et al.*, 2021] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-NeRF: Neural radiance fields for dynamic scenes. In *CVPR*, 2021.

[Radford *et al.*, 2021] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, 2021.

[Savant Aira *et al.*, 2024] Luca Savant Aira, Antonio Montanaro, Emanuele Aiello, Diego Valsesia, and Enrico Magli. MotionCraft: Physics-based zero-shot video generation. *NeurIPS*, 2024.

[Shim *et al.*, 2023] Jaehyeok Shim, Changwoo Kang, and Kyungdon Joo. Diffusion-based signed distance fields for 3d shape generation. In *CVPR*, 2023.

[Tan *et al.*, 2024] 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.

[Tang *et al.*, 2024] Jiaxiang Tang, Jiawei Ren, Hang Zhou, Ziwei Liu, and Gang Zeng. DreamGaussian: Generative gaussian splatting for efficient 3d content creation. *ICLR*, 2024.

[Waczyńska *et al.*, 2024] Joanna Waczyńska, Piotr Borycki, Sławomir Tadeja, Jacek Tabor, and Przemysław Spurek. GaMeS: Mesh-based adapting and modification of gaussian splatting. *arXiv preprint arXiv:2402.01459*, 2024.

[Wang *et al.*, 2025] Xingrui Wang, Wufei Ma, Angtian Wang, Shuo Chen, Adam Kortylewski, and Alan Yuille. Compositional 4d dynamic scenes understanding with physics priors for video question answering. *ICLR*, 2025.

[Wiles *et al.*, 2020] Olivia Wiles, Georgia Gkioxari, Richard Szeliski, and Justin Johnson. SynSin: End-to-end view synthesis from a single image. In *CVPR*, 2020.

[Wu *et al.*, 2016] Jiajun Wu, Joseph J Lim, Hongyi Zhang, Joshua B Tenenbaum, and William T Freeman. Physics 101: Learning physical object properties from unlabeled videos. In *BMVC*, 2016.

[Wu *et al.*, 2023] Tong Wu, Jiarui Zhang, Xiao Fu, Yuxin Wang, Jiawei Ren, Liang Pan, Wayne Wu, Lei Yang, Jiaqi Wang, Chen Qian, et al. Omniobject3D: Large-vocabulary 3d object dataset for realistic perception, reconstruction and generation. In *CVPR*, 2023.

[Wu *et al.*, 2024] 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 *CVPR*, 2024.

[Xie *et al.*, 2024] Tianyi Xie, Zeshun Zong, Yuxing Qiu, Xuan Li, Yutao Feng, Yin Yang, and Chenfanfu Jiang. PhysGaussian: Physics-integrated 3d gaussians for generative dynamics. In *CVPR*, 2024.

[Xu *et al.*, 2024] Qingshan Xu, Jiao Liu, Melvin Wong, Caishun Chen, and Yew-Soon Ong. Precise-physics driven text-to-3d generation. *arXiv preprint arXiv:2403.12438*, 2024.

[Xue *et al.*, 2025] Qiyao Xue, Xiangyu Yin, Boyuan Yang, and Wei Gao. PhyT2V: LLM-guided iterative self-refinement for physics-grounded text-to-video generation. *CVPR*, 2025.

[Yan *et al.*, 2024] Han Yan, Mingrui Zhang, Yang Li, Chao Ma, and Pan Ji. PhyCAGE: Physically plausible compositional 3D asset generation from a single image. *NeurIPS*, 2024.

[Yang *et al.*, 2024] Zeyu Yang, Hongye Yang, Zijie Pan, and Li Zhang. Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting. *ICLR*, 2024.

[Yi *et al.*, 2024] Taoran Yi, Jiemin Fang, Junjie Wang, Guanjun Wu, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Qi Tian, and Xinggang Wang. GaussianDreamer: Fast generation from text to 3d gaussians by bridging 2d and 3d diffusion models. In *CVPR*, 2024.

[Zeng *et al.*, 2024] Bohan Zeng, Ling Yang, Siyu Li, Jiaming Liu, Zixiang Zhang, Juanxi Tian, Kaixin Zhu, Yongzhen Guo, Fu-Yun Wang, Minkai Xu, et al. Trans4D: Realistic geometry-aware transition for compositional text-to-4d synthesis. *arXiv preprint arXiv:2410.07155*, 2024.

[Zhang *et al.*, 2024a] Tianyuan Zhang, Hong-Xing Yu, Rundi Wu, Brandon Y Feng, Changxi Zheng, Noah Snavely, Jiajun Wu, and William T Freeman. PhysDreamer: Physics-based interaction with 3d objects via video generation. In *ECCV*, 2024.

[Zhang *et al.*, 2024b] Xinjie Zhang, Zhening Liu, Yifan Zhang, Xingtong Ge, Dailan He, Tongda Xu, Yan Wang, Zehong Lin, Shuicheng Yan, and Jun Zhang. MEGA: Memory-efficient 4d gaussian splatting for dynamic scenes. *arXiv preprint arXiv:2410.13613*, 2024.

[Zhong *et al.*, 2024] Licheng Zhong, Hong-Xing Yu, Jiajun Wu, and Yunzhu Li. Reconstruction and Simulation of Elastic Objects with Spring-Mass 3D Gaussians. *ECCV*, 2024.

[Zhou *et al.*, 2025] Yang Zhou, Zongjin He, Qixuan Li, and Chao Wang. LAYOUTDREAMER: Physics-guided layout for text-to-3d compositional scene generation. *arXiv preprint arXiv:2502.01949*, 2025.

[Zienkiewicz and Taylor, 2000] Olgierd Cecil Zienkiewicz and Robert Leroy Taylor. *The finite element method: solid mechanics*. Butterworth-heinemann, 2000.