arXiv:2601.00504v1 [cs.CV] 1 Jan 2026

MotionPhysics：用于文本引导仿真的可学习运动蒸馏

Miaowei Wang、Jakub Zadro˙zny、Oisin Mac Aodha、Amir Vaxman

信息学院，爱丁堡大学，爱丁堡 EH8 9AB，英国 m.wang-123@sms.ed.ac.uk，{jakub.zadrozny、oisin.macaodha、a.vaxman}@ed.ac.uk

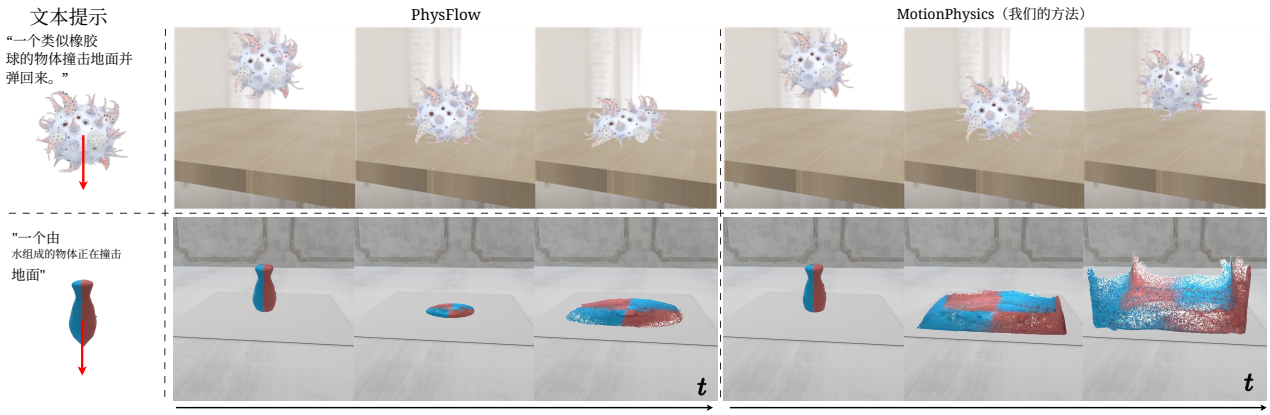

图 1：**MotionPhysics** 可自动估计合理的材料参数，以支持多种材料和物体类型的动态三维仿真。与先前工作（如 PhysFlow（Liu 等，2025a））相比，它能更准确地遵循用户输入的提示（左），尤其适用于人工智能生成物体（上：弹性仿真）、人工设计物体（下：水仿真）以及真实世界扫描。图像背景来自 Li 等人（2023年）。

摘要

要准确模拟现有的 3D物体和种类繁多的材料，通常需要专业知识，并耗费大量时间调节物理参数，才能实现所需的动态行为。我们提出了*MotionPhysics*，这是一种端到端可微框架，可根据用户提供的自然语言提示，为选定的目标三维场景推断合理的物理参数，无需真实轨迹或标注视频提供指导。我们的方法首先利用多模态大语言模型估计材料参数值，并将其约束在合理范围内。我们进一步提出一种可学习运动蒸馏损失，从预训练视频扩散模型中提取稳健的运动先验，同时最大限度地减少外观与几何归纳偏置对仿真的影响。我们在三十多个场景中评估了*MotionPhysics*，涵盖真实世界、人工设计和人工智能生成的 3D物体，涉及多种材料，包括弹性固体、金属、泡沫、砂，以及牛顿流体和非牛顿流体。结果表明，该方法能够在自然语言提示的引导下生成视觉逼真的动态仿真，超越最先进水平，并自动确定物理上合理的参数。

代码和项目页面见：https://wangmiaowei.github.io/MotionPhysics.github.io/。

引言

合理物理参数的设定对于逼真的 3D 仿真至关重要（Gottstein 2004年）。例如，杨氏模量决定材料的刚度，而屈服应力则标志着不可逆塑性变形的开始。传统上，识别此类参数往往依赖专家直觉或耗时的反复试错（Stomakhin 等，2013），导致仿真流程耗时且非专业人士难以使用。

这推动了大量致力于自动物理参数估计的研究。早期方法依赖直接观测（Asenov 等，2019；Wu 等人，2016；Jaques、Burke 和 Hospedales，2020；Ma 等，2023）。近期的新视角合成研究，例如神经辐射场（NeRF）（Mildenhall 等，2021）和高斯泼溅（GS）（Kerbl 等，2023），提供了其他策略。GS 通过高斯核显式表示场景几何，从而能够轻松集成现有模拟器，例如用于流体的 PBD（Feng 等，2025）、用于弹性体的扩展位置基动力学（Jiang 等，2024），以及用于一般材料的 MPM（Xie 等，2024）。这种集成重新激发了人们对系统辨识的兴趣；系统辨识旨在从合成物体的多视角视频中恢复物理参数，

（Jatavallabhula 等，2021；Cai 等，2024；Li 等，2023）。当无法获得真实动力学时，近期方法会通过文本或图像提示引导视频扩散模型，以推断合理的参数值（Liu 等，2025a；Zhang 等，2024；Lin 等，2025；Huang 等，2025）。

尽管视频扩散模型有望实现零样本参数估计，但其性能仍然有限。近期研究（Kang 等，2025；Bansal 等，2024）表明，无论是开源（Yang 等，2025；Chen 等，2024b；Wang 等，2025b；Zheng 等，2024）还是闭源（Pika 2024；Bar-Tal 等，2024；Luma AI 2024）视频生成系统，都无法生成哪怕遵循基本物理常识的视频。因此，对于许多物体和场景，例如 AI生成形状（图 1，上）、人工设计模型（图 1，下），或从新视角观察以及处于不同受力条件下的真实世界物体（图 7），当前基于视频扩散的方法通常会推断出错误的仿真参数。

由于计算需求巨大，且需要涵盖广泛新颖物体的多样化真实运动（GT）数据，构建能够生成物理上合理输出的完整视频扩散模型并不现实。相反，我们利用现有的预训练视频扩散模型来辅助分布外场景与物体仿真。我们的目标是在用户提供的高层语言指令引导下，从预训练模型中蒸馏出合理的运动线索，同时减轻模型在形状和外观方面的归纳偏差。为此，我们提出了一种新颖的可学习运动蒸馏（LMD）损失，从预训练视频扩散模型中提取纯粹的运动信号，以引导我们的可微仿真。具体而言，LMD 通过在训练期间结合几何形状和外观方面的增强扰动，利用轻量级、可训练的运动提取器，最小化仿真与扩散模型预测结果之间的外观和几何形状差异。

准确初始化仿真参数至关重要，因为较差的初始值会浪费计算资源并妨碍收敛。我们在 PhysFlow 的多模态初始化基础上，加入约束感知提示，将特定领域的参数边界范围（例如金属、泡沫和橡皮泥的典型杨氏模量或密度范围）嵌入其中。通过强制大语言模型在这些限制范围内选择数值，我们得以利用其关于真实世界材料的内部知识。这种方法既能将模拟限定在物理上合理的范围内，也能抑制大语言模型幻觉和捏造（Farquhar 等，2024；Hao 等，2024；Walters 和 Wilder，2023）。

我们的最终目标是确保“你描述的内容正是你仿真的内容”。我们在 30 多种仿真场景中验证了本框架，涵盖弹性材料、橡皮泥、金属、泡沫、沙土、牛顿流体和非牛顿流体。我们的主要贡献包括：(i) 引入可学习运动蒸馏损失，以分离并利用真实的运动信号，并在合理的材料范围值触发下采用大语言模型初始化；(ii) 构建了一个全自动的文本引导系统，在仿真性能上达到最先进水平（SOTA），并在人工设计、人工智能生成和真实世界物体上无缝超越现有方法。

相关工作

三维动力学生成。扩散（Liu 等人，2022年）和流匹配（Jin 等人，2024年）模型已能根据文本或图像提示生成中等时长、高质量的视频合成，SORA（Liu 等人，2024b）和 GOKU（Chen 等人，2025）便是其中的代表。为将这些动态先验扩展至四维（3D +时间），多项研究将扩散与双重形状–纹理表示相结合，采用 NeRF 风格的编码器或高斯泼溅（GS），在没有显式四维监督的情况下生成视角一致的动态场景（Singer 等人，2023年；Yuan 等人，2024年；Ling 等人，2024年；Zhao 等人，2023年）。D-NeRF（Pumarola 等人，2021年）和 HyperNeRF（Park 等人，2021年）等隐式方法往往存在渲染速度慢、用户控制有限的问题，而显式泼溅流程（例如 4DGS（Wu 等人，2024年）、HYBRID3D–4DGS（Oh 等人，2025）及其他实时系统（Yang 等人，2024年；Duan 等人，2024年））则提供了快速且易于编辑的替代方案。这些泼溅框架进一步推广至多种输入模态：多视图静态重建（Chen 等人，2024a；Huang 等人，2024a）、动态捕获（Xu 等人，2024年）、单张未标定图像（Yi 等人，2024年；Smart 等人，2024年），以及网格到高斯场的转换（Waczy´nska 等人，2024年）。尽管取得了这些进展，大多数方法仍缺乏物理基础，因为它们并未对由力和材料属性驱动的形变或运动进行建模，因此真实感有限（Bansal 等人，2025；Zhang 等人，2025；Cao 等人，2024年；Wang 等人，2025c）。

基于物理的动态生成。 将物理定律嵌入生成模型（Zhong 等，2024）能够产生更逼真的交互；使用可微分模拟器，则可以进行基于梯度的运动合成，同时调整材料参数。简单的弹簧–质量系统（如 SPRINGGAUS（Zhong 等，2024）和 PHYSTWIN（Jiang 等，2025））能够有效捕捉弹性变形，而可微分物质点法（MPM）（Jiang 等，2016）则擅长模拟多种材料行为。例如，PHYSMOTION（Tan 等，2024）利用 MPM 实现单图像动力学；PACNERF（Li 等，2023）则通过粒子融合 NeRF 和 MPM，在复杂场景中权衡效率与保真度。PHYS-GAUSSIAN（Xie 等，2024）进一步结合 3DGS 与 MPM 提升视觉质量，但仍需要手动调优材料设置。系统辨识可以恢复这些设置，但需要真实值视频（Cai 等，2024）或标记点（Ma 等，2023），这限制了其可扩展性。因此，无需任何真实值动力学即可自动估计材料属性，仍是现实应用中的一项开放挑战。

物理参数估计。为引入物理真实感，近期方法（Zhang 等，2024；Liu 等，2024a；Huang 等，2025；Lin 等，2025；Liu 等，2025a）利用视频扩散先验（Blattmann 等，2023；Meng 等，2024）来推断弹性和塑性等材料属性。PHYSDREAMER（Zhang 等，2024）对真实场景中的弹性行为建模，PHYS3D（Liu 等，2024a）则将这一方法扩展到塑性变形。DREAMPHYSICS 和 PHYSFLOW 能处理更广泛的材料，而 OMNIPHYSGS 将本构模型整合到每个粒子中，以支持异质相互作用。这些方法通过对可微模拟器进行反向传播来优化材料参数，所采用的方式包括——

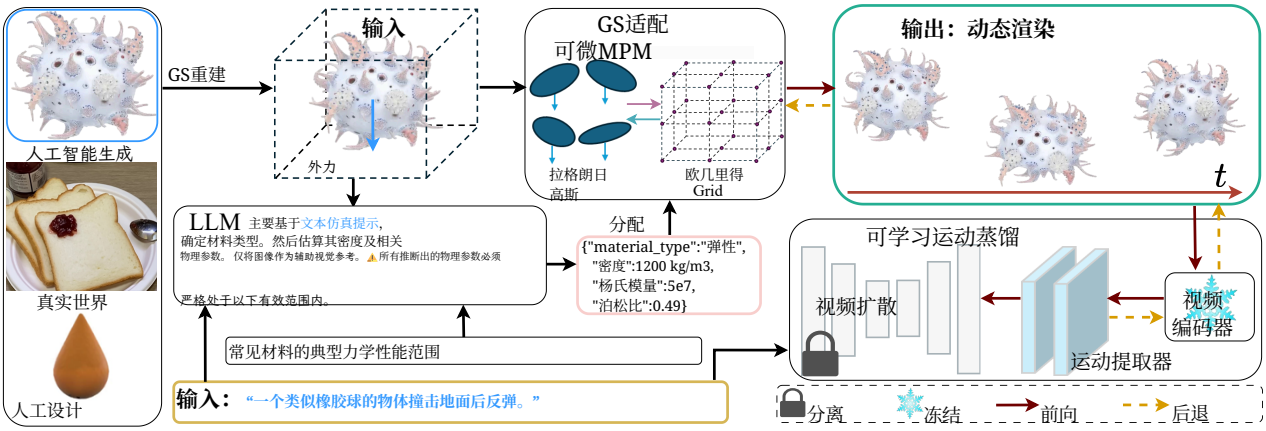

图 2：**概述**。 MotionPhysics 通过自动估计不同输入场景（包括人工智能生成、真实世界及人工设计的资产）的物理参数，根据文本引导的输入提示模拟符合物理规律的动力学。

——（i）直接感知目标，例如模拟帧与生成帧之间的图像相似度（Zhang 等，2024）或光流散度（Liu 等，2025a）；或（ii）从扩散模型导出的分数蒸馏损失（Huang 等，2025；Liu 等，2024a）。然而，这些方法依赖真实世界视频素材，其中的物体往往固定不动、受到遮挡或受噪声影响（Wang 等人，2025a），因此与日益普及的人工设计和人工智能生成资产相比，其适用性有限。人工设计物体往往缺乏一致纹理（图1，下），而人工智能生成的网格（Zhao 等，2025；Tochilkin 等，2024）可能具有非典型几何形状或外观（图1，上），从而干扰外观驱动的监督。为解决这些局限，我们提出一种可学习运动蒸馏 损失，从受文本提示引导的预训练扩散模型中提取运动线索，同时抑制外观和几何偏差。尽管有些工作（Jeong、Park 和 Ye，2024；Zhai 等，2024）提取帧间运动以进行基于视频的迁移，我们的方法侧重于通过文本引导估计物理参数，以用于基于物理的模拟。文本提示通常会通过指明材料类型和物体类别，隐含地传达物理属性。我们利用预训练大语言模型和合理的材料取值范围，更有效地自动估计这些材料属性；这是先前工作中忽视的关键方面（Huang 等，2025；Liu 等，2025a；Lin 等，2024）。

方法

问题陈述

我们考虑真实、合成、人工设计（通常缺乏高质量纹理）或由 AI 生成（通常几何形状或外观较为罕见）的物体和场景。这些输入可以是多视角静态图像、动态视频、单张图像或网格。与传统网格表示相比，3DGS 更适合重现真实世界场景，并支持高质量实时渲染。PGSR、GIC（Cai 等，2024）、Splat3R 和 GaMes 等方法可以将输入转换为由三维高斯组成的集合，$\mathcal{G} := \{\mathbf{x}_g,\ \sigma_g,\ \mathbf{\Sigma}_g,\ \mathcal{S}_g\}$，其中每个泼溅元素 g 由其中心 $\mathbf{x}_g \in \mathbb{R}^3$、不透明度 $\sigma_g \in [0,1]$、协方差 $\mathbf{\Sigma}_g \in \mathbb{R}^{3\times3}$和颜色系数 $\mathcal{S}_g$定义。

为模拟离散时间 t 内的动态行为，我们将随时间变化的 splat 表示为 $\mathcal{G}^t := \{\mathbf{x}_g^t,\ \sigma_g^t,\ \mathbf{\Sigma}_g^t,\ \mathcal{S}_g^t\}$。我们采用可微的 GS适配 MLS-MPM 模拟器（Xie 等，2024），该模拟器使用马尔可夫（Richey，2010）更新 $\mathcal{T}$ 演化 splat 状态，其中包括粒子到网格和网格到粒子的映射（为清晰起见省略速度）：

$$(\mathbf{x}_g^{t+1},\ \mathbf{\Sigma}_g^{t+1},\ \mathcal{S}_g^{t+1}) = \mathcal{T}(\mathbf{x}_g^t,\ \mathbf{\Sigma}_g^t,\ \mathcal{S}_g^t \mid \boldsymbol{\theta},\ \mathbf{f}^{\text{ext}}), \quad (1)$$

其中，$\mathbf{\Sigma}_g^t$ 和 $\mathcal{S}_g^t$ 通过变形梯度F^{t+1}演化，该梯度能够描述拉伸、旋转和剪切（Xie 等，2024），而$\mathbf{f}^{\text{ext}}$ 表示外力。遵循 PHYSFLOW（Liu 等，2025a），所有高斯点共享相同的材料参数，定义为 $\boldsymbol{\theta} := \{\rho, c, \boldsymbol{\theta}_c\}$，其中 ρ 为密度，c为材料类别（例如弹性材料、橡皮泥、金属、泡沫、砂、牛顿流体或非牛顿流体），对应不同的材料本构模型；$\boldsymbol{\theta}_c$ 包含相关的类别特定系数（详情见补充材料）。目标：我们的目标是自动推断 完整的材料参数集 $\boldsymbol{\theta}$，依据自然语言提示 $\mathcal{P}_{\text{text}}$，无需真实动力学、动作捕捉标记或视频的监督。获得这些参数后，即可在各种力场和物理条件下进行高保真且多样化的 3D 动态仿真（见补充材料）。

预备知识

分数蒸馏采样（SDS）（Poole 等，2022）最初旨在从大规模二维扩散模型中提炼三维先验，以用于文本到三维生成（Lin 等，2023）。近期的扩展工作（Lin 等，2025；Huang 等，2025；Liu 等，2024a）将 SDS 应用于使用扩散模型 ϕ优化物理参数 $\boldsymbol{\theta}_c$。设 $z_0 = \mathcal{E}_\phi(\{\mathcal{I}_l\})$ 为视频编码器 $\mathcal{E}_\phi$对视频帧 $\{\mathcal{I}_l\}$ 进行潜在编码的结果。在视频扩散步骤 k，z_0 会受到噪声 $\epsilon \sim \mathcal{N}(0,\ I)$ 的扰动：

$$z_k = \sqrt{\bar{\alpha}_k}\, z_0 \ +\ \sqrt{1-\bar{\alpha}_k}\, \epsilon, \quad \bar{\alpha}_k = \prod_{i=1}^{k} \alpha_i, \quad \alpha_i \in [0,1].$$

SDS 更新的计算方式如下：

$$s_{\text{SDS}}(k,\epsilon) \;=\; w_k\left(\epsilon_\phi(z_k,k\mid\mathcal{P}_{\text{text}}) \;-\; \epsilon\right), \qquad (2)$$

其中，ϵ_ϕ 是来自 ϕ 的预测噪声，w_k 是依赖于步数的权重。相对于 $\boldsymbol{\theta}_c$ 的梯度为：

$$\nabla_{\boldsymbol{\theta}_c} L_{\text{SDS}} \;=\; \mathbb{E}_{k,\epsilon}\Big[\, s_{\text{SDS}}(k,\epsilon)\,\frac{\partial\{\mathcal{I}_l\}}{\partial\{\mathcal{G}^t\}}\,\frac{\partial\{\mathcal{G}^t\}}{\partial\boldsymbol{\theta}_c}\Big]. \qquad (3)$$

其中，$\{\mathcal{I}_l\}$ 由 3DGS 从 $\{\mathcal{G}^t\}$中进行可微渲染得到，而 $\{\mathcal{G}^t\}$ 通过 GS适配的 MLS-MPM 求解器模拟依赖于 $\boldsymbol{\theta}_c$。该目标使模拟视频与扩散模型的数据分布保持一致，从而从后者中蒸馏物理先验，以优化参数 $\boldsymbol{\theta}_c$。

方法概述

如图 2所示，我们首先将静态物体或场景重建为初始 GS 表示 $\mathcal{G}^0$。与此同时，我们向多模态大语言模型（例如 GPT-4（Achiam 等，2023））提供为每种可能的材料类型所规定的参数范围，主要以用户文本提示 $\mathcal{P}_{\text{text}}$ 为条件，并辅以一张参考渲染图像，从而获得初始材料参数估计 $\boldsymbol{\theta}^{\text{ini}}$。利用 $\boldsymbol{\theta}^{\text{ini}}$ 和外力 $\mathbf{f}^{\text{ext}}$，我们的可微 GS适配 MLS-MPM 求解器对散点随时间的动态进行模拟，生成序列 $\{\mathcal{G}^t\}$。我们将其中的稀疏子集进行阿尔法混合，得到视频帧 $\{\mathcal{I}_l\}$。为优化 $\boldsymbol{\theta}^{\text{ini}}$，我们引入可学习运动蒸馏损失（公式 5），该损失从以相同 $\mathcal{P}_{\text{text}}$为条件的预训练视频扩散模型 ϕ 中提取动态先验。通过迭代执行模拟、渲染和基于梯度的优化，我们最终收敛到一组物理上合理的参数集 $\boldsymbol{\theta}_c$，忠实地复现用户指定的运动。

通过多模态大语言模型进行初始化

准确识别材料类型 c对于在 MLS-MPM 中选择合适的本构模型至关重要。给定用户提示，例如“一个橡胶球状物体撞击地面后弹起。”，我们的目标是推断材料类型 c、密度 ρ以及一组相应的特定材料参数 $\boldsymbol{\theta}_c$。如果提示发生变化，推断出的材料及参数也应相应调整，从而实现灵活的用户驱动仿真（见补充材料中的图 A4）。

为此，我们利用 GPT-4（Achiam 等，2023）主要根据文本、其次根据参考图像来估计初始材料参数 $\boldsymbol{\theta}^{\text{ini}}$。然而，像 PhysFlow 那样直接向 GPT-4 提问，可能会导致其产生幻觉（Farquhar 等，2024；Hao 等，2024）或编造数值（Walters 和 Wilder，2023）（见图 6）。大语言模型本质上编码了大量真实世界
材料知识，因此，为了减少幻觉，我们向 GPT-4 提供了包含取值范围约束的提示模板，这些约束基于标准材料属性手册（Callister 和 Rethwisch 2020；Ashby 和 Jones 2012；Mitchell 和 Soga 2005；Gibson 和 Ashby 1999；Rumble 2024）。这种依据可靠资料的做法能引导大语言模型给出切合实际的数值，避免在某些情况下出现不合理的预测和仿真失败（图 6）。补充材料

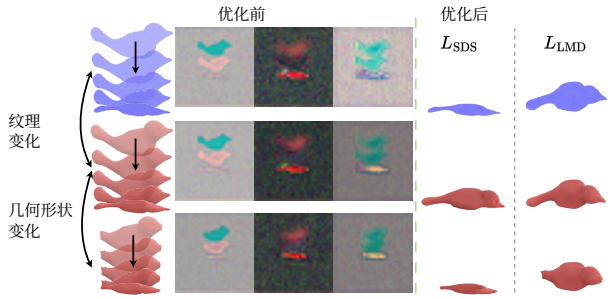

图3：**结构相似性**。左：在优化前，具有相同运动模式的物体（中间为通过主成分分析可视化的潜在编码）呈现出不同的纹理和几何形状。右：应用我们的 L_{LMD}后，动力学表现趋于一致，且基本不受纹理和几何形状的影响。

包含完整的提示模板和典型参数范围。例如，弹性材料的杨氏模量范围为 10^7–4×10^{11}帕，橡皮泥的杨氏模量为$10^6-5\times10^6$ 帕。

可学习运动蒸馏

通过上述基于大语言模型的方法，可以可靠地推断密度 ρ 和材料类别 c。然而，类别特定系数 $\boldsymbol{\theta}_c$ 只是粗略的近似，无法满足精确仿真的要求，因此需要额外的监督，因为大语言模型缺乏动态建模和仿真能力。沿用先前的工作（Liu 等，2024a；Huang 等，2025），我们采用视频扩散模型 ϕ来监督运动优化。然而，基于扩散的预测会将运动与训练数据中的外观和几何偏差固有地纠缠在一起，因此难以提取纯粹的运动信号，尤其是对于人工设计或人工智能生成物体而言；它们的外观或结构可能超出训练分布（见图5）。

一个关键观察是，尽管外观或形状发生变化，在相同初始物理参数下模拟出的相同运动，仍会产生由视频编码器 $\mathcal{E}_\phi$ 得到的全局一致潜在编码。如图 3（左）所示，无论鸟的颜色从红色变为蓝色（上），还是其形状失去尾巴和喙（下），潜变量（通过主成分分析投影可视化）都具有相同的整体结构，仅在局部细节上有所不同。这一规律在不同扩散模型中均成立（见补充材料），并启发我们借助可学习运动提取器 M，通过动态平滑局部潜在空间中的差异（即模型预先学习的分布与目标场景之间在外观和几何形状上的差距），直接从“干净”潜变量而非噪声中提取运动信号。

具体而言，为了强化运动学习，我们在初始表示 $\mathcal{G}^0$（包括核中心和颜色系数）中加入高斯噪声 ϵ'（见补充材料），得到 $\tilde{\mathcal{G}}^0$。由此生成增强后的渲染结果 $\{\tilde{\mathcal{I}}_l\}$和潜-˜变量 $\tilde{z}_0=\mathcal{E}_\phi(\{\mathcal{I}_l\})$。随后，我们计算单步去噪潜变量 $\tilde{z}_k^0=\frac{1}{\sqrt{\bar{\alpha}_k}}\left(\tilde{z}_k-\sqrt{1-\bar{\alpha}_k}\,\epsilon_\phi(\tilde{z}_k,k\mid\mathcal{P}_{\text{text}})\right)$，，再通过几何投影将 $\tilde{z}_k$ 投回干净潜变量流形。

我们将蒸馏后的运动目标和预测分别记为：

$$y_{\text{target}}=M(\tilde{z}_0),\quad y_{\text{pred}}=M(\tilde{z}_k^0), \qquad (4)$$

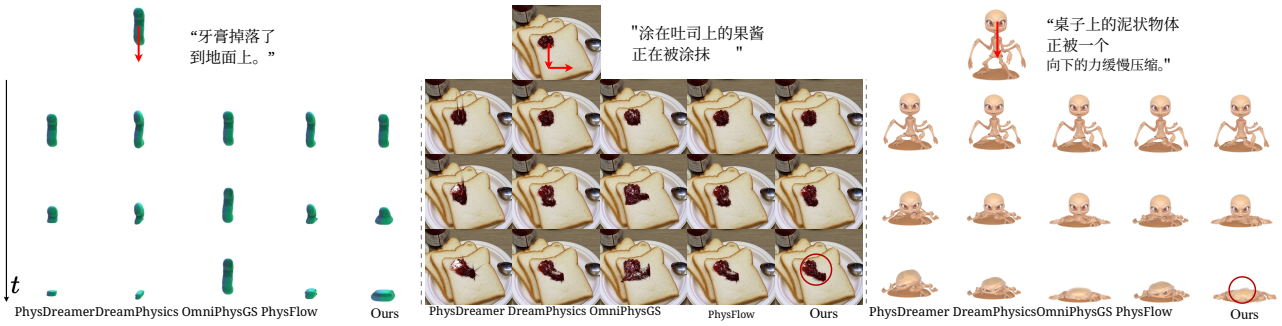

图4：**定性评估**。我们在多种仿真案例中将我们的方法与几种基线方法进行比较，包括人工设计物体（例如牙膏，左）、真实世界场景（例如果酱，中）以及人工智能生成物体（例如外星人，右）。红色箭头表示输入力，红色圆圈突出显示差异的关键区域。更多结果请参见补充材料。

其中，可学习运动提取器 M 是一个轻量级的两层卷积网络，初始化为恒等映射，并以较小的学习率 2×10^{-5}进行训练。由此得到的可学习运动蒸馏（LMD）损失采用 Charbonnier 变体以提高数值稳定性（Barron 2019）：

$$L_{\text{LMD}} = w_k \sqrt{\|y_{\text{pred}} - y_{\text{target}}\|^2 + \beta^2}, \tag{5}$$

加上较小的常数 $\beta = 10^{-3}$，并通过指数移动平均（Tarvainen 和Valpola 2017）保持 M 同步。最后，估计系数 $\boldsymbol{\theta}_c$ 的梯度：

$$\nabla_{\boldsymbol{\theta}_c} L_{\text{LMD}} = \mathbb{E}_{k,\epsilon,\epsilon'}\left[\frac{\partial L_{\text{LMD}}}{\partial\{\tilde{\mathcal{I}}_l\}} \frac{\partial\{\tilde{\mathcal{I}}_l\}}{\partial\{\tilde{\mathcal{G}}^t\}} \frac{\partial\{\tilde{\mathcal{G}}^t\}}{\partial \boldsymbol{\theta}_c}\right]. \tag{6}$$

如图 3（右）所示，我们的 L_{LMD} 损失能够捕捉外观和几何形状中一致的运动模式，而L_{SDS} （公式 3）在优化后会产生不一致的结果。

实现细节

我们基于英伟达的 WARP 实现构建可微分模拟器（ Macklin 2022）。遵循 Xie 等，2024 的方法，我们通过各向异性正则化缓解细长伪影，并填充固体物体的内部体积，以增强仿真真实感。为稳定梯度传播并提升长 MPM 展开序列的训练速度，我们采用帧增强方案（Huang 等，2025；Zhang 等，2024）。给定 $M \times L$ 总帧数（其中 M =为 8），我们将其拆分为 M 个交错子序列 $V_i = \{I_i, I_{i+M}, \ldots, I_{i+M(L-1)}\}$，分别对应 i = 1、.....、M，并在这些 M 组之间交替进行监督。每次仿真持续 5 秒，共生成 150 个渲染帧，每帧包含 256 个内部子步。采用帧增强时，对于每个子序列，我们会在相邻帧之间执行 $256 \times M$ 次中间更新，并仅在最后一步计算梯度。此外，我们使用 CogVideoX 模型（Yang 等，2025）并结合无分类器引导（CFG=100），遵循 Phys-Flow 对可学习运动蒸馏进行蒸馏。训练约经过 40 次迭代后收敛，在 NVIDIA A100 80 GB GPU 上，每次前向—反向传播约需 28 秒。我们的框架支持对边界条件（例如，见图 1）和力的施加（见图 7）进行多种手动指定，从而能够精确控制材料响应的时空变化。

结果与讨论

评估设置

数据集。 我们在三种类型的数据集上进行实验。1）**人工设计：**我们评估了八个 PAC-NeRF 模型（环面、鸟、橡皮泥、猫、奖杯、液滴、字母奶油和牙膏），这些模型呈现均匀色彩，而非细致纹理。我们使用来自 GIC（Cai 等，2024）的静态 3DGS 重建结果。由于我们的重点是使用各种材料提示进行文本引导物理仿真（见补充材料），而非系统辨识，因此不使用其渲染的动态帧作为真实标签。这些帧依赖于手动指定的参数，无法涵盖真实分布的全部多样性。

2) **真实世界：**我们纳入了 PhysDreamer 的四个场景（Alo- casia、康乃馨、帽子和电话）（Zhang 等，2024），以及另外四个场景：来自 InstantNGP 的狐狸（M¨uller 等，2022）、来自 NeRFStudio 的飞机（Tancik 等，2023）、来自 Mip-NeRF 360 的厨房（Barron 等，2022），以及来自 PhysFlow 的果酱和沙堡。对于所有真实世界场景，我们都使用 PhysFlow 提供的文本提示。

3) **人工智能生成：** 我们使用来自混元3D（Zhao 等，2025）的网格 Urchin、Alien、Gentleman 和斧头，并通过 GaMes（Waczy´nska 等，2024）对其进行进一步处理，以获得相应的 3DGS 表示。

基线方法。 我们与以下方法进行比较：1）*PhysDreamer*（Zhang 等，2024），通过模拟渲染结果与生成视频之间的图像外观损失来估计材料属性和初始速度，仅支持弹性材料。2）*DreamPhysics*（Huang 等，2025），采用分数蒸馏采样（SDS）。3）*PhysFlow*（Liu 等，2025a），通过最小化模拟视频与生成视频之间的光流损失进行优化。4）*OmniPhysGS*（Lin 等，2025），在视频扩散的引导下，使用本构 3D 高斯对异质物体进行建模。我们使用开源实现，其中前三种方法基于 NVIDIA WARP，而 OmniPhysGS 依赖速度较慢且稳定性较差的 PyTorch 模拟器（见表 2），因此我们缩短其视频长度，以保持帧率一致。所有基线方法均采用相同的模拟设置（边界条件、外力和文本提示）。在材料初始化方面，我们按照 PhysFlow 的做法使用 GPT-4 预测结果，但 OmniPhysGS 除外，它

直接优化本构模型，而不指定具体的材料参数。为保证公平性，所有初始化和优化（包括我们的方法）均只运行一次。

指标。 我们开展了一项双选强迫选择（2AFC）用户研究（Macmillan 和 Creelman，2005年；Zhang 等，2024年），其中 79 名参与者比较 15 组随机选取的并排视频对：在相同场景、提示和施加的力下，将我们的方法与基线方法进行比较，并回答以下问题：Q1）**物理真实感**：哪个视频展现出对施加的力更真实的物理响应？以及 Q2）**提示词遵循度**：哪个视频对用户提示的遵循度更高？（详见补充材料）。我们还通过 VBench（Huang 等人，2024b）中的总体一致性（OC），对提示词遵循度进行客观量化；该指标使用 ViCLIP（Wang 等人，2023b）进行测量，以捕捉提示与视频之间的全局语义对齐程度；同时采用 CLIPSIM（Wu 等人，2021）（与 OmniPhysGS 中相同），计算提示与每个视频帧的 CLIP（Radford 等人，2021）嵌入之间余弦相似度的平均值。此外，我们使用 PhysFlow 中基于能量最小化原理的能量约束运动评分（ECMS）来评估运动真实性。

结果

为评估泛化能力，我们在不同场景和设置下改变输入力，从而产生显著的变形和动态响应（见补充材料）。定性示例见图 A1，各基线方法上的平均用户偏好见表 1（详细投票结果见补充材料）。我们的方法在物理真实感和提示词遵循度方面均优于所有基线方法，在所有数据集上的偏好率均超过 50%，在人工设计和人工智能生成的场景上更超过 80%，展现出对新颖几何形状和纹理的强大泛化能力。例如，在果酱实验中（图 A1，中），我们的方法能够捕捉牛顿流体的黏性，呈现平滑且无空洞铺展的效果，不同于基线方法（红圈）。类似地，在牙膏实验中（图 A1，左），我们的方法再现了非牛顿行为：起初流动并铺展，随后沉降为稳定的堆积体，而其他方法则未能做到。

在表 2 中，得益于轻量级运动提取器，我们的方法在所有客观定量指标上的平均得分均为最高，同时保持了有竞争力的优化速度（在图 5 下方所示的鸟场景上进行测试）。然而，这些指标并未完全符合人类感知。例如，水滴场景（见补充材料）在 CLIPSIM 和 ECMS 上的得分略低于部分基线方法，

	对比	人工设计 ↑	真实世界 ↑	人工智能生成 ↑
Realism	我们的方法与 PhysDreamer 对比	**96.77%**	77.63%	**93.80%**
	我们的方法与 DreamPhysics 对比	**80.27%**	69.77%	**94.65%**
	我们的方法与 OmniPhysGS	**97.62%**	**92.73%**	**84.60%**
	我们的方法 vs. PhysFlow	**90.30%**	66.13%	**84.20%**
Adherence	我们的方法 vs. PhysDreamer	**95.88%**	**86.08%**	**96.45%**
	我们的方法 vs. DreamPhysics	**82.35%**	75.05%	**93.80%**
	我们的方法 vs. OmniPhysGS	**96.48%**	**91.43%**	**80.75%**
	我们的方法 vs. PhysFlow	**91.30%**	71.53%	**85.60%**

表 1：**用户研究结果**。平均百分比值超过 80%（以粗体突出显示）表示对物理真实感（上）和提示词遵循度（下）的强烈偏好。

指标	PhysDreamer	DreamPhysics	OmniPhysGS	PhysFlow	Ours
总体一致性 $\times 10^{-2}\uparrow$	17.00	18.02	17.10	17.96	**18.18**
CLIPSIM $\times 10^{-2}\uparrow$	21.62	21.64	21.27	21.32	**21.69**
ECMS ↓	27.48	15.76	13.70	13.07	**11.37**
优化时间 ↓	16.48分钟	18.20分钟	~ 9 h	18.14分钟	18.39分钟

表2：**定量评估**。 最佳结果以**粗体**显示；“优化时间”是初始化后的优化时间（分钟/小时）。

尽管生成了预期的水花飞溅行为。我们将这一差距归因于两个因素：1）当前的视频-文本一致性和运动指标无法区分相同场景、力和提示下的动态差异；2）预训练指标模型侧重于静态外观，缺乏捕捉文本所指定的多样材料和运动模式的能力（详细分析见补充材料）。一种补救方法是像系统辨识那样比较物理参数。然而，获取物理上合理结果的真实值分布颇具挑战。类似局限性在三维生成任务中也曾被指出（Yu 等，2025；Tang 等，2024）。因此，我们强调视觉对比，并在后续实验中展示定性评估。

消融研究

LMD **的影响。** 为展示我们可学习运动蒸馏（LMD）目标函数的优势，我们比较了三种损失：1）PhysFlow 的光流损失 L_{Flow}，从生成视频中提取光流；2） SDS 损失 L_{SDS}（式 3）；以及 3）我们的 LMD 损失 L_{LMD} （式5）。在图5中，由于提示指定的材料与原始仿真数据中的材料相同（橡皮泥：上方，橡皮泥；鸟：下方，弹性），我们使用 PAC-NeRF 的手动调优输出（Li 等，2023）作为材料行为的粗略参考，尽管该方法最初是为系统辨识而开发的。使用 L_{LMD} 优化的模型与

这一参考最为接近，证实了其在材料参数优化方面的优越性（定量结果见表3， L_{LMD} 的更多消融实验见补充材料）。请注意，PhysFlow 项目页面（Liu等人，2025b）上的结果与手动参考十分相似，因为他们在配对的真实视频上训练人工设计物体，并对其进行系统辨识。

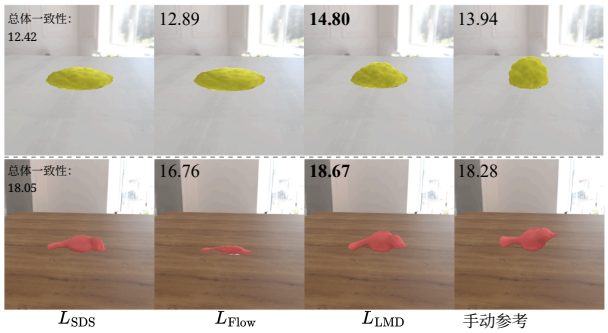

图5： L_{LMD} 的消融实验。相应的总体一致性（OC $\times 10^{-2}\uparrow$）分数也一并给出。

初始化的影响。 我们的方法采用特定于材料的范围提示，与 PhysFlow 的朴素

我们的方法初始化 →PhysFlow 初始化	我们的方法初始化（无优化）	$L_{\text{LMD}} \rightarrow L_{\text{SDS}}$	**Ours**
18.04	18.05	18.13	**18.18**

表 3：**消融结果**。 使用总体一致性（$\times 10^{-2} \uparrow$）报告。

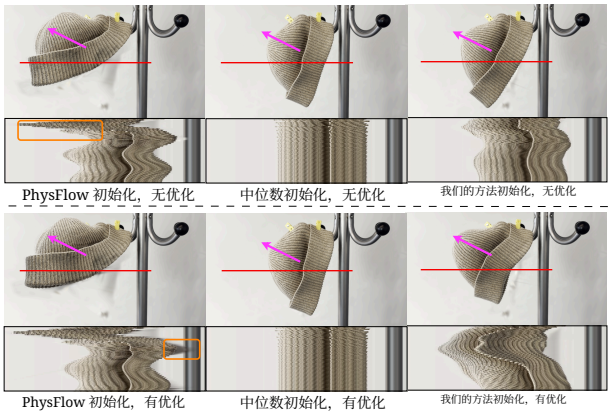

图 6：**初始化消融**。我们通过时空切片展示振荡：时间位于 y 轴上，物体的横截面（顶部“物体”中的红线）位于 x 轴上，由此揭示振幅和频率上的振荡。

大语言模型推理不同。为验证这一点，我们在图 6 中对 Hat 场景进行了六种条件下的模拟：PhysFlow 初始化（使用 LMD）并优化、PhysFlow 初始化（使用 LMD）但不优化、我们的方法初始化（使用 LMD）并优化、我们的方法初始化（使用 LMD）但不优化，以及中位值初始化（使用 LMD）并优化、中位值初始化（使用 LMD）但不优化（取约束值范围的中位数）。PhysFlow 的初始化会导致早期形变过度夸张，即使经过优化，仍会产生不真实的伪影，如橙色矩形框所示。中位值初始化对上限非常敏感。在此情况下，较大的杨氏模量 2×10^{11} 会导致物体表现得过于刚硬，形变极小，无法遵循给定的力和文本描述，即使经过优化也如此。相比之下，我们的方法提供了稳定且合理的起点（见表 2）。与 LMD 结合后，生成的材料动力学既准确又具有视觉说服力。这些结果表明，我们的范围提示能够引导大语言模型选择合适的参数值，而不是凭空臆造错误的猜测。

对变化仿真条件的鲁棒性。我们在不同外力和轻微摄像机扰动下评估每个场景。如图 7（上）所示，电话受到方向不同的力，并且视角发生了偏移。我们的方法仍能在提示的引导下产生正确的弹性变形：“电话线正在轻轻振动”；而 PhysFlow 几乎保持静止，因为生成的视频本身未能提供正确的运动，无法作为其损失函数的真实标签（L_{Flow}）（见补充材料中的图 A6）。如图 7（下）所示，我们将飞机螺旋桨的转速提高至五倍。得益于我们设定范围内金属较高的杨氏模量和屈服强度，我们的仿真保持了结构完整性，且未达到屈服应力。相比之下，即使使用相同的提示，PhysFlow 仍会使螺旋桨脱落（红圈标出）

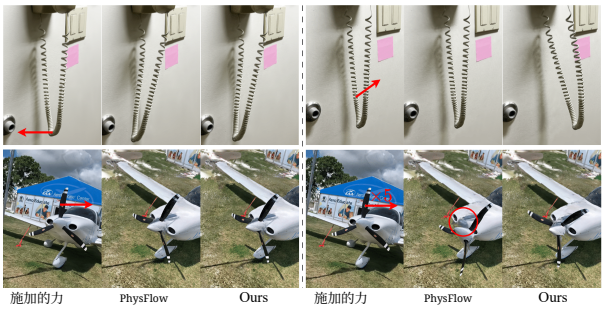

图 7：**对变化的仿真条件的鲁棒性**。（顶部）改变力的方向和视点方向。
（底部）螺旋桨的旋转速度提高了五倍。

提示相同：“飞机螺旋桨正在旋转”。

异质材料扩展。 我们的方法可无缝扩展到包含不同材料的多物体场景。通过结合 GS 分割（Wang 等人，2025a）与 SAM2（Ravi 等人，2024），多模态大语言模型能够根据单条文本提示推断每个物体的材料属性。如图 8 所示，斧头被视为金属，玩具则被视为弹性橡胶。在我们的 MPM框架中，每个粒子都携带自身的材料参数，从而实现高保真的异质动力学。例如，橡胶玩具会在受力时变形，而斧头则保持刚性。PhysDreamer 中的弹性伪影以红色标出（见随附视频）。请注意，在撰写本文时，OmniPhysGS 尚无可用的多物体仿真代码。将我们的流程应用于电影制作或电子游戏等复杂场景，还需要进行超出本文范围的进一步工程开发。



图 8：**异质材料扩展**。 与其他方法相比，我们的方法产生的视觉伪影更少。

结论

MotionPhysics 是一种新颖的框架，利用视频扩散模型和多模态大语言模型，在简单文本提示的引导下驱动三维动态场景仿真。可学习运动蒸馏模块能够提取干净的运动线索，而基于大语言模型的嵌入则用于初始化特定材料的参数先验，从而实现高保真、具有物理基础的动画。未来，我们计划支持仅凭文本全自动配置精细仿真，并将运动蒸馏损失拓展至其他动画任务，例如角色绑定和形变。

局限性。 我们的方法未对阴影效果进行建模，而这本可以提升视觉真实感。此外，尽管我们的

估计参数能够实现合理的模拟，但其目的并非进行准确的真实世界材料测量。

致谢

作者谨此感谢华为技术有限公司对本研究的慷慨支持。

参考文献

Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F. L.; Almeida, D.; Altenschmidt, J.; Alt- man, S.; Anadkat, S.; 等。2023年。GPT-4 技术报告。*arXiv:2303.08774*。

Asenov, M.; Burke, M.; Angelov, D.; Davchev, T.; Subr, K.; 和 Ramamoorthy, S。2019年。Vid2param：从视频中建模动力学参数。《机器人与自动化快报》。

Ashby, M. F.; and Jones, D. R. H. 2012年。工程材料 *1*：性能、应用与设计导论。第4版。Bansal, H.; Lin, Z.; Xie, T.; Zong, Z.; Yarom, M.; Bitton, Y.; Jiang, C.; Sun, Y.; Chang, K.-W.; and Grover, A. 2024年。Videophy：评估视频生成中的物理常识。*arXiv:2406.03520*。Bansal, H.; Peng, C.; Bitton, Y.; Goldenberg, R.; Grover, A.; and Chang, K. 2025。VideoPhy-2：一项以动作为中心、具有挑战性的视频生成物理常识评估。在视频生成国际学习表征会议研讨会论文集中。Bar-Tal, O.; Chefer, H.; Tov, O.; Herrmann, C.; Paiss, R.; Zada, S.; Ephrat, A.; Hur, J.; Liu, G.; Raj, A.; 等。2024年。Lumiere：用于视频生成的时空扩散模型。在 *SIGGRAPH Asia 2024* 会议论文集中。Barron, J. T. 2019。通用且自适应的鲁棒损失函数。在计算机视觉与模式识别会议中，4331–4339。Barron, J. T.; Mildenhall, B.; Verbin, D.; Srinivasan, P. P.; and Hedman, P. 2022。Mip-NeRF 360：无界抗锯齿神经辐射场。在计算机视觉与模式识别会议（计算机视觉与模式识别会议）中，5470–5479。Blattmann, A.; Dockhorn, T.; Kulal, S.; Mendelevitch, D.; Kilian, M.; Lorenz, D.; Levi, Y.; English, Z.; Voleti, V.; Letts, A.; 等。2023年。稳定视频扩散：将潜在视频扩散模型扩展至大型数据集。*arXiv:2311.15127*。Bonet, J.; and Wood, R. D. 1997。用于有限元分析的非线性连续介质力学。剑桥大学出版社。Cai, J.; Yang, Y.; Yuan, W.; He, Y.; Dong, Z.; Bo, L.; Cheng, H.; and Chen, Q. 2024年。用于物理属性识别与仿真的高斯信息连续体。神经信息处理系统进展。Callister, W. D.; and Rethwisch, D. G. 2020。材料科学与工程：导论。Wiley，第10版。Cao, M.; Tang, H.; Zhao, H.; Guo, H.; Liu, J.; Zhang, G.; Liu, R.; Sun, Q.; Reid, I.; and Liang, X. 2024年。PhysGame：揭示游戏视频中的物理常识违背现象。*arXiv:2412.01800*。

Chen, D.; Li, H.; Ye, W.; Wang, Y.; Xie, W.; Zhai, S.; Wang, N.; Liu, H.; Bao, H.; and Zhang, G. 2024a。PGSR：用于高效、高保真表面重建的平面高斯泼溅。*arXiv:2406.06521*。Chen, H.; Zhang, Y.; Cun, X.; Xia, M.; Wang, X.; Weng, C.; and Shan, Y. 2024b。Videocrafter2：突破高质量视频扩散模型的数据限制。在计算机视觉与模式识别会议中，7310–7320。Chen, S.; Ge, C.; Zhang, Y.; Zhang, Y.; Zhu, F.; Yang, H.; Hao, H.; Wu, H.; Lai, Z.; Hu, Y.; 等。2025。Goku：基于流的视频生成基础模型。*arXiv:2502.04896*。

Duan, Y.; Wei, F.; Dai, Q.; He, Y.; Chen, W.; 和 Chen, B。2024年。4D-Rotor 高斯泼溅：面向动态场景的高效新视角合成。载于*ACM*计算机图形学特别兴趣小组 *2024* 会议论文，1–11。

Farquhar, S.; Kossen, J.; Kuhn, L.; 和 Gal, Y.2024年。使用语义熵检测大语言模型中的幻觉。《自然》，630(8017): 625–630。Feng, Y.; Feng, X.; Shang, Y.; Jiang, Y.; Yu, C.; Zong, Z.; Shao, T.; Wu, H.; Zhou, K.; Jiang, C.; 和 Yang, Y. 2025。高斯泼溅：用于多样化运动合成与渲染的统一粒子。发表于计算机视觉与模式识别会议 *(CVPR)*。Gibson, L. J.; 和 Ashby, M. F. 1999年。胞状固体：结构与性质。第二版。

Gottstein, G. 2004年。 材料科学的物理基础，第3卷。

Hao, G.; Wu, J.; Pan, Q.; 和 Morello, R. 2024年。量化复杂适应性社交网络中大语言模型幻觉传播的不确定性。《科学报告》，14(1): 16375。Huang, B.; Yu, Z.; Chen, A.; Geiger, A.; 和 Gao, S. 2024a. 面向几何精确辐射场的二维高斯泼溅。载于*SIGGRAPH 2024* 会议论文集。美国计算机协会。

Huang, T.; Zhang, H.; Zeng, Y.; Zhang, Z.; Li, H.; Zuo, W.; 和 Lau, R. W. H. 2025. DreamPhysics：利用视频扩散先验学习基于物理的三维动力学。载于人工智能促进协会。Huang, Z.;He, Y.;Yu, J.; Zhang, F.;Si, C.; Jiang, Y.; Zhang, Y.; Wu, T.; Jin, Q.; Chanpaisit, N.; 等。2024b. Vbench：面向视频生成模型的综合基准测试套件。载于计算机视觉与模式识别会议论文集，21807–21818。

Jaques, M.； Burke, M.； Hospedales, T. 2020。Physics-as-Inverse-Graphics：从视频中无监督估计物理参数。载于国际学习表征会议。Jatavallabhula, K. M.； Macklin, M.； Golemo, F.； Voleti, V.； Petrini, L.； Weiss, M.； Considine, B.； Parent-Levesque, J.； Xie, K.； Erleben, K.； Paull, L.； Shkurti, F.； Nowrouzezahrai, D.； Fidler, S. 2021。gradSim：用于系统辨识和视觉运动控制的可微分仿真。国际学习表征会议。Jeong, H.； Park, G. Y.； Ye, J. C. 2024。VMC：使用时间注意力适配进行视频运动定制，以实现文本生成视频扩散模型。载于计算机视觉与模式识别会议。

Jiang, C.；Schroeder, C.；Teran, J.；Stomakhin, A.；Selle, A. 2016。用于仿真连续体材料的物质点法。载于*ACM SIGGRAPH 2016* 课程，1–52。计算机协会。Jiang, H.；Hsu, H.-Y.；Zhang, K.；Yu, H.-N.；Wang, S.；Li, Y. 2025。PhysTwin：从视频中进行可变形物体的物理信息重建与仿真。*ICCV*。Jiang, Y.；Yu, C.；Xie, T.；Li, X.；Feng, Y.；Wang, H.；Li, M.；Lau, H.；Gao, F.；Yang, Y.；等。2024。Vr-gs：虚拟现实中的物理动力学感知交互式高斯溅射系统。载于*ACM SIGGRAPH 2024* 会议论文集。Jin, Y.；Sun, Z.；Li, N.；Xu, K.；Jiang, H.；Zhuang, N.；Huang, Q.；Song, Y.；Mu, Y.；Lin, Z. 2024。用于高效视频生成建模的金字塔流匹配。*arXiv:2410.05954*。Kang, B.；Yue, Y.；Lu, R.；Lin, Z.；Zhao, Y.；Wang, K.；Gao, H.；Feng, J. 2025。视频生成距离世界模型还有多远？物理定律视角。载于第*42*届国际机器学习会议（*ICML*）论文集。PMLR。Kerbl, B.；Kopanas, G.；Leimkuehler, T.；Drettakis, G. 2023。用于实时辐射场渲染的三维高斯溅射。*ACM* 图形学汇刊。Li, X.；Qiao, Y.-L.；Chen, P. Y.；Jatavallabhula, K. M.；Lin, M.；Jiang, C.；Gan, C. 2023。PAC-NeRF：用于与几何无关的系统辨识的物理增强连续神经辐射场。载于国际学习表征会议。Lin, C.-H.；Gao, J.；Tang, L.；Takikawa, T.；Zeng, X.；Huang, X.；Kreis, K.；Fidler, S.；Liu, M.-Y.；Lin, T.-Y. 2023。Magic3d：高分辨率文本到三维内容生成。载于计算机视觉与模式识别会议。Lin, J.；Wang, Z.；Xu, D.；Jiang, S.；Gong, Y.；Jiang, M. 2024。Phys4DGen：结合多材料组成感知的符合物理规律的四维生成。*arXiv*预印本 *arXiv:2411.16800*。Lin, Y.；Lin, C.；Xu, J.；MU, Y. 2025。OmniPhysGS：用于通用基于物理的动力学生成的三维本构高斯。载于国际学习表征会议。Ling, H.；Kim, S. W.；Torralba, A.；Fidler, S.；Kreis, K. 2024。对齐你的高斯：使用动态三维高斯和组合扩散模型进行文本到四维生成。载于计算机视觉与模式识别会议，8576–8588。Liu, F.；Wang, H.；Yao, S.；Zhang, S.；Zhou, J.；Duan, Y. 2024a。Physics3d：通过视频扩散学习三维高斯的物理属性。*arXiv:2406.04338*。Liu, Y.；Zhang, K.；Li, Y.；Yan, Z.；Gao, C.；Chen, R.；Yuan, Z.；Huang, Y.；Sun, H.；Gao, J.；等。2024b。SORA：关于大型视觉模型的背景、技术、局限性与机遇的综述。*arXiv:2402.17177*。Liu, Z.；Ning, J.；Cao, Y.；Wei, Y.；Zhang, Z.；Lin, S.；Hu, H. 2022。视频 Swin Transformer。载于计算机视觉与模式识别会议，3202–3211。Liu, Z.；Ye, W.；Luximon, Y.；Wan, P.；Zhang, D. 2025a。释放多模态基础模型和视频扩散在四维动态物理场景仿真中的潜力。*CVPR*。

Liu, Z.; Ye, W.; Luximon, Y.; Wan, P.; and Zhang, D. 2025b. 释放多模态基础模型与视频扩散在四维动态物理场景仿真中的潜力。计算机视觉与模式识别会议。参见 https://zhuomanliu.github.io/ PhysFlow/。Luma AI. 2024. Luma Dream Machine — AI 视频生成器。https://lumalabs.ai/dream-machine。访问日期：2025年5月4日。Ma, P.; Chen, P. Y.; Deng, B.; Tenenbaum, J. B.; Du, T.; Gan, C.; and Matusik, W. 2023. 从运动观测中学习神经本构定律，以实现可泛化的偏微分方程动力学。收录于*ICML*。PMLR。Macklin, M. 2022. Warp：用于 GPU 模拟和图形学的高性能 Python 框架。https://github.com/nvidia/warp。英伟达 GPU 技术大会（GTC）。Macmillan, N. A.; and Creelman, C. D. 2005. 《检测理论：用户指南》。新泽西州马瓦：Lawrence Erlbaum Associates，第二版。Meng, F.; Liao, J.; Tan, X.; Shao, W.; Lu, Q.; Zhang, K.; Cheng, Y.; Li, D.; Qiao, Y.; and Luo, P. 2024. 迈向世界模拟器：为视频生成打造基于物理常识的基准。*arXiv:2410.05363*。Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. NeRF：将场景表示为神经辐射场以实现视图合成。*ACM* 图形学通讯，65(1)：99–106。Mitchell, J. K.; and Soga, K. 2005. 《土壤行为基础》。第三版。M¨uller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. 采用多分辨率哈希编码的即时神经图形基元。*ACM* 图形学汇刊，41(4)：102:1–102:15。Oh, S.; Lee, Y.; Jeon, H.; and Park, E. 2025. 用于快速动态场景表示的混合式3D-4D高斯泼溅。*arXiv:2505.13215*。Park, K.; Sinha, U.; Hedman, P.; Barron, J. T.; Bouaziz, S.; Goldman, D. B.; Martin-Brualla, R.; and Seitz, S. M. 2021. HyperNeRF：用于拓扑变化神经辐射场的高维表示。*ACM* 图形学汇刊（*TOG*），40(6)：1–12。Pika. 2024. Pika。https://pika.art/。访问日期：2025-05-04。Poole, B.; J ain, A.; Barron, J. T.; and Mildenhall, B. 2022. DreamFusion：使用二维扩散的文本到3D。收录于国际学习表征会议。Pumarola, A.; Corona, E.; Pons-Moll, G.; and Moreno- Noguer, F. 2021. D-NeRF：用于动态场景的神经辐射场。收录于计算机视觉与模式识别会议。Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. 从自然语言监督中学习可迁移视觉模型。收录于*ICML*。Ravi, N.; Gabeur, V.; Hu, Y.-T.; Hu, R.; Ryali, C.; Ma, T.; Khedr, H.; R¨adle, R.; Rolland, C.; Gustafson, L.; et al. 2024. SAM 2：图像与视频中的万物分割。收录于国际学习表征会议。Richey, M. 2010. 马尔可夫链蒙特卡罗方法的演变。《美国数学月刊》。

Rumble, J. R., ed. 2024. *CRC*《化学与物理手册》。第105版。Singer, U.; Sheynin, S.; Polyak, A.; Ashual, O.; Makarov, I.; Kokkinos, F.; Goyal, N.; Vedaldi, A.; Parikh, D.; John- son, J.; and Taigman, Y. 2023. 文本到4D动态场景生成。收录于*ICML*。Smart, B.; Zheng, C.; Laina, I.; and Prisacariu, V. A. 2024. Splatt3r：从未校准的图像对进行零样本高斯泼溅。*arXiv:2408.13912*。Stomakhin, A.; Schroeder, C.; Chai, L.; Teran, J.; and Selle, A. 2013. 一种用于雪模拟的物质点法。*ACM* 图形学汇刊Tan, X.; Jiang, Y.; Li, X.; Zong, Z.; Xie, T.; Yang, Y.; and Jiang, C. 2024. PhysMotion：从单张图像出发、以物理学为基础的动力学。*arXiv* 预印本 *arXiv:2411.17189*。Tancik, M.; Weber, E.; Ng, E.; Li, R.; Yi, B.; Kerr, J.; Wang, T.; Kristoffersen, A.; Austin, J.; Salahi, K.; Ahuja, A.; McAllister, D.; and Kanazawa, A. 2023. Nerfstudio：用于神经辐射场开发的模块化框架。收录于*ACM*计算机图形学特别兴趣小组 *2023* 会议论文集。Tang, J.; Chen, Z.; Chen, X.; Wang, T.; Zeng, G.; and Liu, Z. 2024. Lgm：用于高分辨率3D内容创作的大型多视图高斯模型。收录于欧洲计算机视觉会议。Tarvainen, A.; and Valpola, H. 2017. 平均教师是更好的榜样：权重平均一致性目标可提升半监督深度学习的效果。神经信息处理系统进展，30。Teed, Z.; and Deng, J. 2020. RAFT：用于光流的循环全对场变换。收录于欧洲计算机视觉会议。Tochilkin, D.; Pankratz, D.; Liu, Z.; Huang, Z.; Letts, A.; Li, Y.; Liang, D.; Laforte, C.; Jampani, V.; and Cao, Y.-P. 2024. TripoSR：从单张图像快速进行3D物体重建。*arXiv:2403.02151*。Waczy´nska, J.; Borycki, P.; Tadeja, S.; Tabor, J.; and Spurek, P. 2024. GaMes：基于网格的高斯泼溅自适应与修改。*arXiv:2402.01459*。Walters, W. H.; and Wilder, E. I. 2023. ChatGPT生成的书目引文及其错误。《科学报告》，13(1)：14045。Wang, J.; Yuan, H.; Chen, D.; Zhang, Y.; Wang, X.; and Zhang, S. 2023a. ModelScope 文本生成视频技术报告。*arXiv* 预印本 *arXiv:2308.06571*。Wang, M.; Zhang, Y.; Ma, R.; Xu, W.; Zou, C.; and Morris, D. 2025a. DecoupledGaussian：用于基于物理的交互的物体—场景解耦。收录于计算机视觉与模式识别会议。Wang, Y.; Chen, X.; Ma, X.; Zhou, S.; Huang, Z.; Wang, Y.; Yang, C.; He, Y.; Yu, J.; Yang, P.; et al. 2025b. Lavie：采用级联潜扩散模型的高质量视频生成。《国际计算机视觉杂志》。Wang, Y.; He, Y.; Li, Y.; Li, K.; Yu, J.; Ma, X.; Li, X.; Chen, G.; Chen, X.; Wang, Y.; et al. 2023b. InternVid：面向多模态理解与生成的大规模视频文本数据集。收录于国际学习表征会议。Wang, Z.; Li, X.; Kumar, A.; and Smith, J. 2025c. NS-4Dynamics：基于物理先验的神经符号4D重建。*arXiv:2502.07007*。维基百科贡献者。2024。材料属性 — 维基百科，自由百科全书。访问日期：2025-05-02。Wu, C.; Huang, L.; Zhang, Q.; Li, B.; Ji, L.; Yang, F.; Sapiro, G.; and Duan, N. 2021。Godiva：根据自然语言描述生成开放领域视频。*arXiv:2104.14806*。Wu, G.; Yi, T.; Fang, J.; Xie, L.; Zhang, X.; Wei, W.; Liu, W.; Tian, Q.; and Wang, X. 2024。用于实时动态场景渲染的4D高斯泼溅。收录于计算机视觉与模式识别会议。Wu, J.; Lim, J. J.; Zhang, H.; Tenenbaum, J. B.; and Freeman, W. T. 2016。Physics 101：从无标签视频中学习物体物理属性。收录于英国机器视觉会议。Xie, T.; Zong, Z.; Qiu, Y.; Li, X.; Feng, Y.; Yang, Y.; and Jiang, C. 2024。Physgaussian：用于生成式动力学的物理集成3D高斯。收录于计算机视觉与模式识别会议，4389–4398。Xu, J.; Fan, Z.; Yang, J.; and Xie, J. 2024。Grid4D：用于高保真动态高斯泼溅的4D分解哈希编码。*arXiv:2410.20815*。Yang, Z.; Pan, Z.; Zhu, X.; Zhang, L.; Jiang, Y.-G.; and Torr, P. H. 2024。4D高斯泼溅：使用原生4D基元建模动态场景。*arXiv:2412.20720*。Yang, Z.; Teng, J.; Zheng, W.; Ding, M.; Huang, S.; Xu, J.; Yang, Y.; Hong, W.; Zhang, X.; Feng, G.; et al. 2025。CogVideoX：采用专家变换器的文生视频扩散模型。收录于国际学习表征会议。Yi, T.; Fang, J.; Wang, J.; Wu, G.; Xie, L.; Zhang, X.; Liu, W.; Tian, Q.; and Wang, X. 2024。GaussianDreamer：通过连接2D和3D扩散模型，实现从文本快速生成3D高斯。收录于计算机视觉与模式识别会议。Yu, Q.; Li, X.; Tang, Y.; Han, X.; Hu, L.; Hao, Y.; and Chen, M. 2025。Fancy123：通过即插即用变形从单张图像生成高质量3D网格。收录于计算机视觉与模式识别会议论文集（*CVPR*），595–604。Yuan, Y.-J.; Kobbelt, L.; Liu, J.; Zhang, Y.; Wan, P.; Lai, Y.- K.; and Gao, L. 2024。4dynamic：基于混合先验的文生4D。*arXiv:2407.12684*。Zhai, Y.; Lin, K.; Yang, Z.; Li, L.; Wang, J.; Lin, C.-C.; Doermann, D.; Yuan, J.; and Wang, L. 2024。运动一致性模型：通过解耦运动-外观蒸馏加速视频扩散。神经信息处理系统进展，37：111000–111021。Zhang, C.; Cherniavskii, D.; Zadaianchuk, A.; Tragoudaras, A.; Vozikis, A.; Nijdam, T.; Prinzhorn, D. W. E.; Bodracska, M.; Sebe, N.; and Gavves, E. 2025。Morpheus：通过真实物理实验评测视频生成模型的物理推理能力。*arXiv:2504.02918*。Zhang, T.; Yu, H.-X.; Wu, R.; Feng, B. Y.; Zheng, C.; Snavely, N.; Wu, J.; and Freeman, W. T. 2024。Physdreamer：通过视频生成实现与3D物体的基于物理的交互。收录于欧洲计算机视觉会议。

Zhao, Y.; Yan, Z.; Xie, E.; Hong, L.; Li, Z.; and Lee, G. H. 2023。Animate124：将单张图像动画化为4D动态场景。*arXiv:2311.14603*。Zhao, Z.; Lai, Z.; Lin, Q.; Zhao, Y.; Liu, H.; Yang, S.; Feng, Y.; Yang, M.; Zhang, S.; Yang, X.; et al. 2025。混元3D 2.0：扩展扩散模型，实现高分辨率纹理3D资产生成。*arXiv:2501.12202*。Zheng, Z.; Peng, X.; Yang, T.; Shen, C.; Li, S.; Liu, H.; Zhou, Y.; Li, T.; and You, Y. 2024。Open-Sora：让所有人都能高效制作视频。*arXiv:2412.20404*。Zhong, L.; Yu, H.-X.; Wu, J.; and Li, Y. 2024。利用弹簧-质点3D高斯重建和仿真弹性物体。欧洲计算机视觉会议。

MotionPhysics：用于文本引导仿真的可学习运动蒸馏

补充材料

组织结构

本补充材料包括本构模型概述；基于大语言模型的完整材料初始化提示，以及不同材料类型的参数范围；运动不变 GS 扰动的详细信息；边界条件；与其他方法的定性比较；更多消融实验；以及额外的用户研究结果，包括投票统计和问卷屏幕截图。我们建议观看补充视频，以了解完整的动态模拟结果。

本构模型

这里，我们简要介绍仿真设置中使用的物理模型。建议读者参考相关资料以获取更多信息。

连续介质力学中的控制方程

连续介质力学使用依赖时间的变形映射 $\mathbf{x} = \psi(\mathbf{X}, t)$来描述材料运动，它将未变形参考构形 Ω^0 中的每个材料点 $\mathbf{X}$ 映射到时间 t时变形构形 Ω^t中的位置 $\mathbf{x}$。局部运动和形状变化由变形梯度刻画：

$$\mathbf{F}(\mathbf{X}, t) = \nabla_{\mathbf{X}} \psi(\mathbf{X}, t), \tag{7}$$

它描述了局部拉伸、旋转和剪切效应（Bonet 和 Wood，1997）。

变形场 ψ 的演化由质量守恒和动量守恒控制。后者给出弹性动力学方程：

$$\rho_0 \ddot{\psi} = \nabla \cdot \mathbf{P} + \rho_0 \mathbf{b}, \tag{8}$$

其中：

- $\ddot{\psi}$ 表示材料点的加速度。
- ρ_0 是参考（初始）质量密度。
- $\mathbf{P}$ 是第一皮奥拉-基尔霍夫应力张量。
- $\mathbf{b}$ 是单位质量的体力。

该偏微分方程组无法完整描述该系统。本构定律用于将应力$\mathbf{P}$与材料状态联系起来，即与材料的形变方式以及在非弹性情况下与其形变历史联系起来。

本构定律：弹性与塑性

本构定律描述材料如何响应形变。一般而言，对于固体，遵循先前的研究（Jiang 等，2016；Ma 等，2023），我们区分以下类型：

- **弹性定律：** 将（可能为弹性的）变形梯度 $\mathbf{F}^e$ 与应力关联起来，决定形变中可逆（可恢复）部分的表现。

材料类型	弹性定律	返回映射（塑性积分）
弹性	固定共旋弹性	恒等映射
橡皮泥	StVK弹性	含损伤的 von Mises 塑性
金属	StVK弹性	von Mises塑性
Foam	StVK弹性	黏塑性返回映射
Sand	StVK弹性	Drucker-Prager 塑性
牛顿流体	可压缩 Neo-Hookean 弹性 恒等映射	
非牛顿流体 StVK 弹性		Herschel-Bulkley 塑性

表 A1：代表性本构模型及相应的塑性积分方案（返回映射），遵循 PhysFlow（Liu 等，2025a）中的实现。弹性定律决定可恢复应力响应，而返回映射算法则将试探状态投影回塑性屈服面。

- **塑性流动与积分（返回映射）：** 当应力状态超过由材料决定的屈服准则时，就会发生永久（不可逆）形变。屈服条件、流动法则和硬化定律决定塑性状态的演化。在实践中，每个时间步都会使用返回映射 算法，以数值方式施加这些约束。

一般形式如下：

$$\text{Elastic Law:} \quad \mathbf{P} = \widehat{P}(\mathbf{F}^e), \tag{9}$$
$$\text{Yield Criterion:} \quad f_Y(\mathbf{P}) < 0, \tag{10}$$
$$\text{Plastic Update:} \quad \mathbf{F}^e \leftarrow \mathrm{ReturnMap}(\mathbf{F}^e). \tag{11}$$

经典材料模型与返回映射

表 A1 总结了计算力学和本文中常用的经典材料模型及返回映射（塑性积分）方案。下面简要而准确地概述所用的各个本构模型及其塑性积分方案：

固定共旋弹性

$$\mathbf{P} = 2\mu(\mathbf{F} - \mathbf{R}) + \lambda J(J-1)\mathbf{F}^{-T}, \tag{12}$$
$$\text{No plastic yields,} \quad f_Y(\mathbf{P}) \equiv -1 < 0, \tag{13}$$
$$\mathbf{F}^e \leftarrow \mathbf{F}, \quad \text{Identity mapping (no evolution).} \tag{14}$$

St.维南–基尔霍夫弹性 (StVK)

$$\boldsymbol{\tau} = \mathbf{U}\left[2\mu\,\boldsymbol{\epsilon} + \lambda\,\mathrm{tr}(\boldsymbol{\epsilon})\mathbf{I}\right]\mathbf{U}^T, \quad \boldsymbol{\epsilon} = \log(\boldsymbol{\Sigma}), \tag{15}$$
$$f_Y(\tau) \text{ depends on specific plastic model below,} \tag{16}$$
$$\mathbf{F}^e \leftarrow \mathrm{ReturnMap}(\mathbf{F}^e). \tag{17}$$

von Mises塑性 （冯·米塞斯屈服准则和径向返回映射）

$$f_Y(\tau) = \|\mathbf{s}\| - \sqrt{\frac{2}{3}}\sigma_Y < 0, \quad \mathbf{s} = \mathrm{dev}(\boldsymbol{\tau}), \tag{18}$$
$$\mathbf{F}^e \leftarrow \mathrm{RadialReturnMap}_{\mathrm{vonMises}}(\mathbf{F}^e). \tag{19}$$

含损伤的 von Mises 塑性（屈服应力随累积塑性应变降低）

$$f_Y(\tau) = \|\mathbf{s}\| - \sqrt{\frac{2}{3}}\sigma_Y(\epsilon^p) < 0, \quad \frac{d\sigma_Y}{d\epsilon^p} < 0, \tag{20}$$

$$\mathbf{F}^e, \sigma_Y \leftarrow \text{RadialReturnMap}_{\text{DamagedVM}}(\mathbf{F}^e, \epsilon^p). \tag{21}$$

黏塑性返回映射（泡沫）（黏性正则化 von Mises 屈服定律）

$$f_Y(\tau) = \|\mathbf{s}\| - \sqrt{\frac{2}{3}}\sigma_Y < 0, \tag{22}$$

$$s = s^{\text{trial}} - \frac{y}{1 + \frac{\eta}{2\mu\Delta t}} \frac{s^{\text{trial}}}{\|s^{\text{trial}}\|}, \tag{23}$$

$$\mathbf{F}^e \leftarrow \text{ViscoplasticReturnMap}(\mathbf{F}^e, \eta). \tag{24}$$

Drucker-Prager 塑性（砂）（摩擦锥屈服准则）

$$f_Y(\tau) = \|\mathbf{s}\| + \alpha\,\text{tr}(\tau) - k < 0, \tag{25}$$

$$\mathbf{F}^e \leftarrow \text{ReturnMap}_{\text{DruckerPrager}}(\mathbf{F}^e, \alpha, k). \tag{26}$$

可压缩 Neo-Hookean 弹性（牛顿流体）（弱可压缩弹性；无独立塑性步骤）

$$\mathbf{P} = \frac{\mu}{J^{2/3}}(\mathbf{F} - J^{-1/3}\mathbf{R}) + \frac{\kappa}{2}(J^2 - 1)\mathbf{F}^{-T}, \tag{27}$$

$$\text{No plastic yields}, \quad f_Y(\mathbf{P}) \equiv -1 < 0, \tag{28}$$

$$\mathbf{F}^e \leftarrow \mathbf{F}, \quad \text{Identity mapping}. \tag{29}$$

Herschel-Bulkley 塑性（非牛顿流体）（率相关黏塑性流体）

$$f_Y(\tau) = \|\mathbf{s}\| - \sqrt{\frac{2}{3}}\sigma_Y < 0, \tag{30}$$

$$s = s^{\text{trial}} - \frac{s^{\text{trial}} - \sqrt{\frac{2}{3}}\sigma_Y}{1 + \frac{\eta}{2\mu\Delta t}} \frac{s^{\text{trial}}}{\|s^{\text{trial}}\|}, \tag{31}$$

$$\mathbf{F}^e \leftarrow \text{ReturnMap}_{\text{HerschelBulkley}}(\mathbf{F}^e, \eta, \sigma_Y). \tag{32}$$

这套全面的本构模型使我们的模型能够捕捉各种现实材料类型丰富的弹塑性行为，从而支持高保真模拟。

基于大语言模型的材料初始化

遵循 PhysFlow（Liu 等，2025a），我们考虑了七种具有代表性的材料类型，每种材料类型都由一组典型参数表征，以涵盖范围广泛的常见材料：

0.1 **弹性：** 杨氏模量 E 和泊松比 ν

0.2 **橡皮泥：** E、ν和屈服应力 τ_Y

0.3 **金属：** E、ν和屈服应力 τ_Y

0.4 **泡沫：** E、ν和塑性黏度 η

0.5 **砂：** 摩擦角 θ_{fric}

0.6 **牛顿流体：** 流体黏度 μ和体积模量 κ

0.7 **非牛顿流体：** 剪切模量 μ、体积模量κ、屈服应力 τ_Y和塑性黏度 η

我们参考了标准材料属性手册中的物理参数范围限制（Callister 和 Rethwisch 2020；Ashby 和 Jones 2012；Mitchell 和 Soga 2005；Gibson 和 Ashby 1999；Rumble 2024），所选范围如表 A2 所示。

我们使用 OpenAI 的 GPT-4（Achiam 等，2023年）这一多模态大语言模型（LLM），根据文本仿真提示和可选的图像参考推断材料类型和物理参数。以下是我们在流程中使用的完整查询提示：

输入：

- **文本仿真提示：**
  “一把斧头正在撞击地面。”
- **参考图像：**



问：这是什么物体？主要依据文本仿真提示，确定最合适的材料类型。然后估算其密度（kg/m3）和相关物理参数。请仅将图像作为次要视觉参考。

警告：材料类型及所需参数（含国际单位制单位）：

- Elastic: E (Pa), ν (unitless)
- Plasticine: E, ν, τ_Y (Pa)
- Metal: E, ν, τ_Y (Pa)
- Foam: E, ν, τ_Y, η (unitless)
- Sand: θ_{fric} (°)
- Newtonian fluid: μ (Pa·s), κ (kPa)
- Non-Newtonian fluid: μ, κ, τ_Y (Pa), η (unitless)

警告：所有推断出的物理参数都必须严格落在以下有效范围：

- *Elastic:*
  - E: $1 \times 10^7 - 4 \times 10^{11}$ Pa

材料类型	参数	国际单位制单位	取值范围
弹性	E (杨氏模量)	Pa	$10^7 - 4 \times 10^{11}$ (Callister 和 Rethwisch 2020)
	ν (泊松比)	(无量纲)	0.1 – 0.5 (Callister 和 Rethwisch 2020)
橡皮泥	E	Pa	$1 \times 10^6 - 5 \times 10^6$ (Ashby 和 Jones 2012)
	ν	(无量纲)	0.3 – 0.4（Ashby 和 Jones 2012年）
	τ_Y （屈服应力）	Pa	$5 \times 10^3 - 2 \times 10^4$ （Ashby 和 Jones 2012年）
金属	E	Pa	$4.5 \times 10^{10} - 4.0 \times 10^{11}$ （Callister 和 Rethwisch 2020）
	ν	（无量纲）	0.25 – 0.35（Callister 和 Rethwisch 2020）
	τ_Y	Pa	$1 \times 10^7 - 2 \times 10^9$ （Callister 和 Rethwisch 2020）
Foam	E	Pa	$10^3 - 10^7$ （Gibson 和 Ashby 1999）
	ν	（无量纲）	0 – 0.3（Gibson 和 Ashby 1999）
	τ_Y	Pa	$1 \times 10^4 - 1 \times 10^6$ (Gibson 和 Ashby 1999)
	η	(无量纲)	0.1 – 1 (Gibson 和 Ashby 1999)
Sand	θ_{fric}	°	27 – 45° (Mitchell 和 Soga 2005)
牛顿流体	μ (黏度)	Pa·s	$10^{-3} - 10$（Rumble 2024）
	κ（体积模量）	Pa	$1 \times 10^9 - 5 \times 10^9$ （Rumble 2024）
非牛顿流体	μ	Pa·s	$10^{-3} - 10^3$ （Rumble 2024）
	κ	Pa	$1 \times 10^9 - 5 \times 10^9$ （Rumble 2024）
	τ_Y	Pa	$1 - 2 \times 10^3$ (Ashby 和 Jones 2012)
	η	(无量纲)	0.1 – 1 (Ashby 和 Jones 2012)
常见	ρ	kg·m^{-3}	$10 - 2.3 \times 10^4$ (Rumble 2024; 维基百科贡献者 2024年)

表 A2：七种模拟材料类型的材料参数、国际单位制单位和取值范围（含来源）。

  - ν: 0.1 – 0.5
- *Plasticine:*
  - E: $1 \times 10^6 - 5 \times 10^6$ Pa
  - ν: 0.3 – 0.4
  - τ_Y: $5 \times 10^3 - 2 \times 10^4$ Pa
- *Metal:*
  - E: $4.5 \times 10^{10} - 4.0 \times 10^{11}$ Pa
  - ν: 0.25 – 0.35
  - τ_Y: $1 \times 10^7 - 2 \times 10^9$ Pa
- *Foam:*
  - E: $1 \times 10^3 - 1 \times 10^7$ Pa
  - ν: 0 – 0.3
  - τ_Y: $1 \times 10^4 - 1 \times 10^6$ Pa
  - η: 0.1 – 1
- *Sand:* θ_{fric}: 27 – 45°
- *Newtonian fluid:*
  - μ: 1×10^{-3} – 10 Pa·s
  - κ: $1 \times 10^6 - 5 \times 10^6$ kPa
- *Non-Newtonian fluid:*
  - μ: $1 - 1 \times 10^6$ mPa·s
  - κ: $1 \times 10^6 - 5 \times 10^6$ kPa
  - τ_Y: $1 - 2 \times 10^3$ Pa
  - η: 0.1 – 1
- 密度 ρ: $10 - 2.3 \times 10^4$ kg/m3

请严格按照以下格式回复：

```
{
  "材料_类型":          "...",
  "密度": ...,
  "E": ...,
  “nu”: ...,
  “tau_Y”: ......,
  “mu”: ...,
  “kappa”: ...,
  “eta”: ...,
  “theta fric”: ......_
}
```

仅包含与推断出的材料类型相关的字段。

大语言模型回答：

```
{
  "材料_类型": "金属",
  "密度": 7850,
  "E": 2.1e11 ,
  "nu": 0.30,
  "tau_Y": 2.5e8
}
```

此外，遵循 PhysFlow 的做法，我们对 GPT-4 的预测应用相同的线性缩放或对数缩放，随后进行相同的截断，以确保我们的可微分 MLS-MPM 仿真具备数值稳定性。

我们提出的方法使大语言模型（LLM）能够直接根据自然语言提示和视觉线索推断仿真参数，为后续的参数优化提供合理的初始化，以复现用户描述的场景。

运动不变 GS 扰动

为促进训练过程中的运动蒸馏，我们对高斯中心$\mathbf{x}_g$和球谐系数$\mathcal{S}_g$都添加了可控噪声：

$$\tilde{\mathbf{x}}_g = \mathbf{x}_g + \epsilon\mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{33}$$

$$\tilde{\mathcal{S}}_g = \mathcal{S}_g + 2\epsilon\mathcal{N}(\mathbf{0}, \mathbf{I}), \tag{34}$$

其中，$\epsilon = 0.1$ 设置扰动尺度。这些带噪高斯仅用于在优化过程中计算运动蒸馏损失，在仿真和评估期间则会被舍弃。通过轻微改变几何形状和外观，这种有针对性的增强促使模型学习在小幅 GS 扰动下仍保持稳定的运动特征，从而获得更鲁棒且语义更丰富的运动表征。

边界条件与力施加

我们采用多种边界条件模块和力施加方案，以支持复杂的材料行为和场景交互。如表 A3 所示，我们的框架包含六种边界条件和五种不同的外力模块，可对粒子和网格行为进行精确的时空控制。最后，我们在表 A4 中总结了真实世界数据集案例的仿真设置，这些设置与 PhysFlow 补充材料中提供的设置有所不同。

更多定性结果

我们在图 A1 中提供了更多定性比较，并在下文展示了进一步的消融研究。

运动模式一致性。 在相同的初始物理参数下经历相同的模拟运动时，纹理或形状的变化会产生由视频编码器生成的全局一致潜在编码。如图 A2 所示，这种运动模式一致性出现在文本生成视频模型使用的不同视频编码器中，包括 ModelScope（Wang 等，2023a）（左）和 CogVideoX（Yang 等，2025）（右）。

额外消融实验 L_{LMD}。 我们展示了主论文图 3（右）对应的模拟序列。如图 A3 所示，我们的可学习运动蒸馏损失 L_{LMD} 能够在几何形状和纹理变化下持续蒸馏运动模式，而使用 L_{SDS} 进行优化则会产生相互偏离的模拟结果，无法遵循用户提供的输入。

对变化文本提示的鲁棒性。 我们评估了每种方法对同一物体不同仿真文本描述的遵循程度。如图 A4 所示，我们的方法能够在多种提示下可靠地产生预期效果，证明其能够减少几何形状或纹理带来的偏差，而是优先实现用户提供的文本提示中指定的运动。

交互式评估。 遵循先前的工作（Zhang 等，2024年），我们仅在单个物体上训练模型一次。在评估期间，我们引入新的仿真条件，以生成多样化的交互场景（图 A5）。初始训练完成后（左侧青色箭头），物体能够在新的施加的力作用下发生形变（右侧白色箭头）。

PhysFlow 的失败分析。 我们进一步分析 PhysFlow 的失败案例。尽管 PhysFlow 同样基于提取运动信号这一思路，但其方法依赖于从生成视频中估计出的光流。然而，这些生成视频往往无法反映仿真条件（例如施加的力）所决定的实际运动，即使在真实世界场景中也是如此，如图 A6 所示。

用户研究详情

在我们的用户研究中，每对视频的左右顺序均经过随机排列，且所有视频都进行了匿名处理，以防参与者根据排列顺序得知哪个视频是我们的方法或某种基线方法。参与者也不知道研究中涉及多少种基线方法。

我们评估了 15 个不同的场景，每个场景分别与 4 种基线方法进行比较，共得到 60 对视频。这些视频对被随机分为 4 份不同的问卷，每份包含 15 对视频，每个场景对应一对。参与者通过主链接访问调查问卷后，会被随机分配到其中一份问卷。共有 79 名用户参与研究，平均每种方法、每个场景约有 20 次用户评估。所有测试案例的原始投票统计数据见表 A5。

我们在下文的图中展示了问卷界面的屏幕截图。图 A7 展示了初始说明页面，其中包含经批准的伦理声明、研究指南以及预计完成时间。图 A8 展示了文本提示、力可视化和视频比较界面。图 A9 展示了用于收集主观评分的两个评估问题。我们在提交文件中附上了原始回答，以及各表单中分配到左侧/右侧位置的视频（详情请参阅用户研究文件夹）。

定量指标详情

本文对主论文中提出的三项指标——总体一致性（OC）、CLIPSIM 和 ECMS——进行详细分析。

考虑一个针对用户提示 $\mathcal{P}_{\text{text}}$ 生成的视频 $I = \{I_1, I_2, \ldots, I_L\}$，其中 I_l 表示时间步 l 对应的帧。为评估视频与提示的契合程度，我们采用两项指标：VBench（Huang 等人，2024b）中的总体一致性（OC）和 CLIPSIM（Wu 等人，2021），定义如下：

$$\text{OC}(I, \mathcal{P}_{\text{text}}) = \frac{\langle \text{ViCLIP}(I), \text{ViCLIP}(\mathcal{P}_{\text{text}}) \rangle}{|\text{ViCLIP}(I)|\,|\text{ViCLIP}(\mathcal{P}_{\text{text}})|}, \tag{35}$$

其中，ViCLIP（Wang 等人，2023b）表示文本和视频联合嵌入模型。

类别	模式 / 名称	描述和参数
基于网格	网格钳制（盒）	长方体速度狄利克雷钳制。**参数：** v、边界范围、轴。
	包围盒 -	全局包围盒约束。将速度投影至内侧；可选摩擦项。
表面碰撞器	表面碰撞器——粘性	接触时粘住：设置 $\mathbf{v} = 0$。摩擦必须为零。
	表面碰撞器——滑移	法向速度为零；保留切向速度。
	表面碰撞器 – 摩擦（默认）	使用库仑摩擦进行单向速度投影（仅向内）。
	表面碰撞器 – 切割	在空间区域内阻尼或钳制速度；移除材料。
力	添加恒力 -	在区域内施加恒定体力。时变激活。
	添加冲量	对选定粒子施加瞬时冲量。
	力粒子平移 -	对粒子施加平移。可按轴和组选择。
	力粒子旋转 -	对粒子施加旋转。可选择组。
	释放粒子 -	根据时间，按索引顺序依次释放冻结粒子。

表 A3：边界条件与外力模块概览

场景	力模块（方向）	持续时间
海芋	添加冲量 (0.44, 0,0)	2 个子步
康乃馨	添加冲量（-0.1，0，0）	1 个子步
Hat	添加冲量（2，-2，2）	1 秒
电话	添加恒力（15，15，-15）	0.75 秒
Fox	添加冲量 $(0,-0.5,0.25) \rightarrow (0,0,-0.5) \rightarrow (0,0.5,0.25)$	$1 \rightarrow 1 \rightarrow 1$子步
飞机	力粒子旋转 $-50 \rightarrow -5$	$0.8 \rightarrow 1$秒
厨房	力粒子平移 (0，0，0。- 1)	5 秒
Jam	力粒子平移 (0，0。3，- 0) $\rightarrow$ (0。3，0，0)	$2.7 \rightarrow 2.3$秒
沙堡	释放粒子 $n_{\text{layer}} = 200$	5 秒

表 A4：真实世界数据集模拟中施加的力及相应持续时间。持续时间以子步或秒为单位报告。

$$\text{CLIPSIM}(I, \mathcal{P}_{\text{text}}) = \sum_l \frac{\langle \text{CLIP}(I_l), \text{CLIP}(\mathcal{P}_{\text{text}}) \rangle}{|\text{CLIP}(I_l)|\,|\text{CLIP}(\mathcal{P}_{\text{text}})|}. \quad (36)$$

遵循 OmniPhysGS（Lin 等，2025）的做法，我们使用 CLIP（Radford 等人，2021）的 ViT-L/14 变体来计算 CLIPSIM。

为量化运动真实性，我们参照 Phys- Flow（Liu 等，2025a），使用能量约束运动评分(ECMS)，其定义为

$$\text{ECMS} = \frac{1}{\sum_l \|F_{l,l+1}\|} + \sum_l \|F_{l,l+1} - F_{l-1,l}\|^2 + \|\nabla^2 F_{l,l+1}\|^2.$$

其中 $F_{l,\ l+1}$ 是 l 和 $l+1$帧之间的光流。我们使用 raft-sintel 预训练权重集，通过 RAFT（Teed 和 Deng，2020）直接从生成视频 I 中估计$F_{l,\ l+1}$。

我们承认，尽管总体一致性、CLIPSIM 和 ECMS 提供了有用的定量见解，但它们仍不完美，且可能与人类判断存在偏差。

总体一致性依赖于 ViCLIP，这是一种基于字幕-视频配对数据集训练的学习型视频-文本对齐模型，而这类数据集通常缺乏不同材料所呈现的多样化动态行为。因此，当外观、场景布局、几何形状、光照、环境和受力条件相同时，它对不同材料的视频给出的分数相近（见主论文中的表 2）。

CLIPSIM 会针对提示独立评估每一帧，因此无法判断时间动态是否体现了指定的材料属性。例如，在水滴场景中（图 A10，上），PhysDreamer（Zhang 等，2024）取得了最高的 CLIPSIM 分数，尽管它将水滴建模为弹性物体，而非液体——因为每一帧在视觉上都像水滴，CLIPSIM 忽略了错误的时间行为。

相比之下，ECMS 完全忽略提示词遵循度。它只关注轨迹匹配，因此可能会奖励不切实际的平滑运动，与预期的材料响应相悖。例如，OmniPhysGS（Lin 等，2025）在图 A10（底部）中取得了最高的 ECMS 分数，尽管牙膏既没有“塌陷”，液滴也没有完全“飞溅”，导致运动与提示相矛盾。

开发更完善的评估指标是未来工作的重要方向。一项关键进展是设计一种提示遵循度指标，能够在固定场景和受力条件下捕捉时间动态。另一项进展则是设计一种全面的物理真实性指标，依据文本提示共同评估材料属性和动态行为。不过，这些改进超出了本研究的范围。

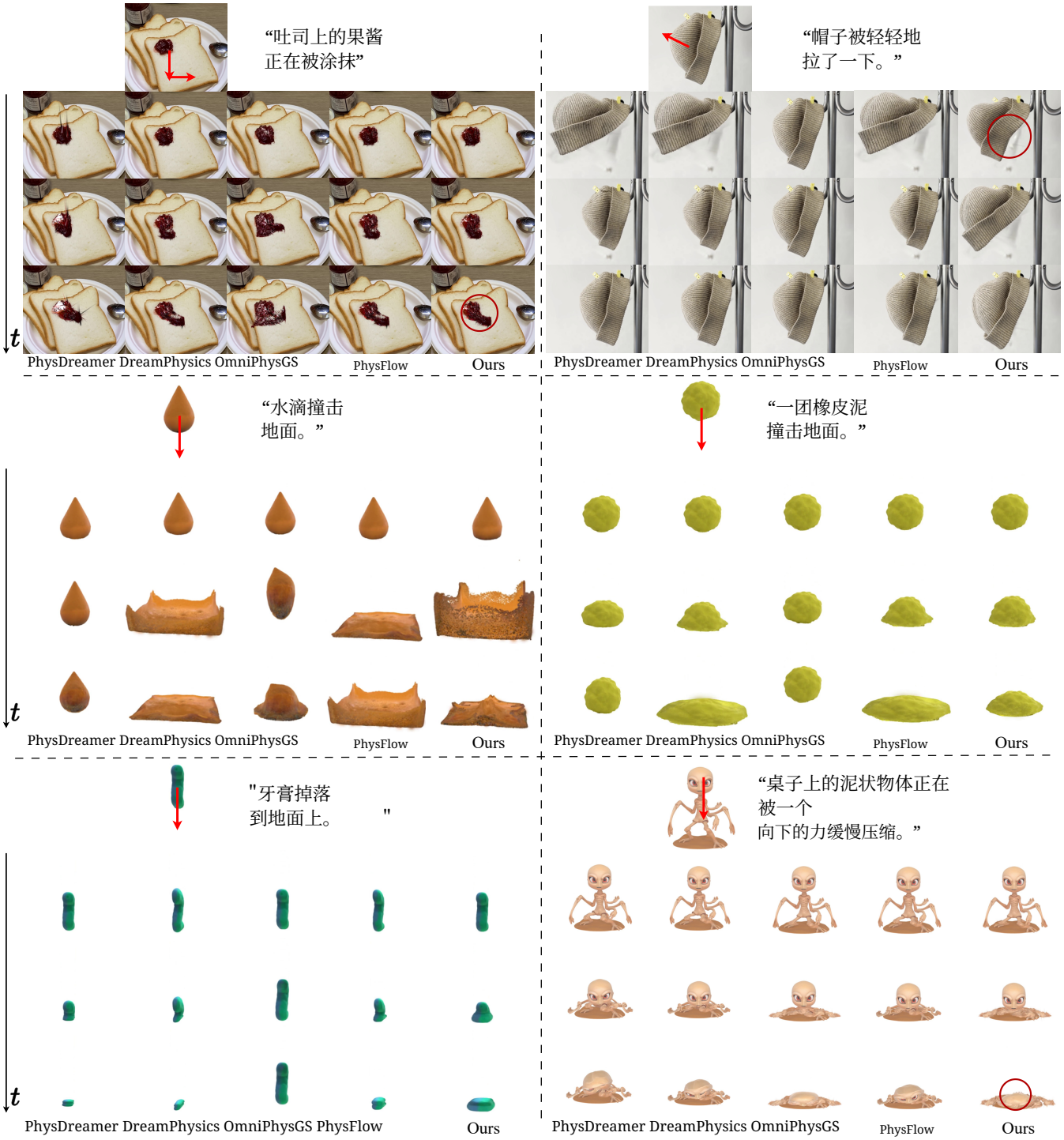

图 A1：**定性评估**。我们在多种仿真案例上比较了所有基线方法，包括 PhysDreamer（Zhang 等，2024）、DreamPhysics（Huang 等，2025）、OmniPhysGS（Lin 等，2025）和 PhysFlow（Liu 等，2025a）。这些案例涵盖真实世界场景（果酱、帽子）、合成类人物体（液滴、橡皮泥、牙膏）以及生成物体（外星人）。红色箭头表示输入力，红色圆圈标出显著区域。我们的方法实现了最逼真的动态仿真，能与纹理提示和施加的力高度吻合。

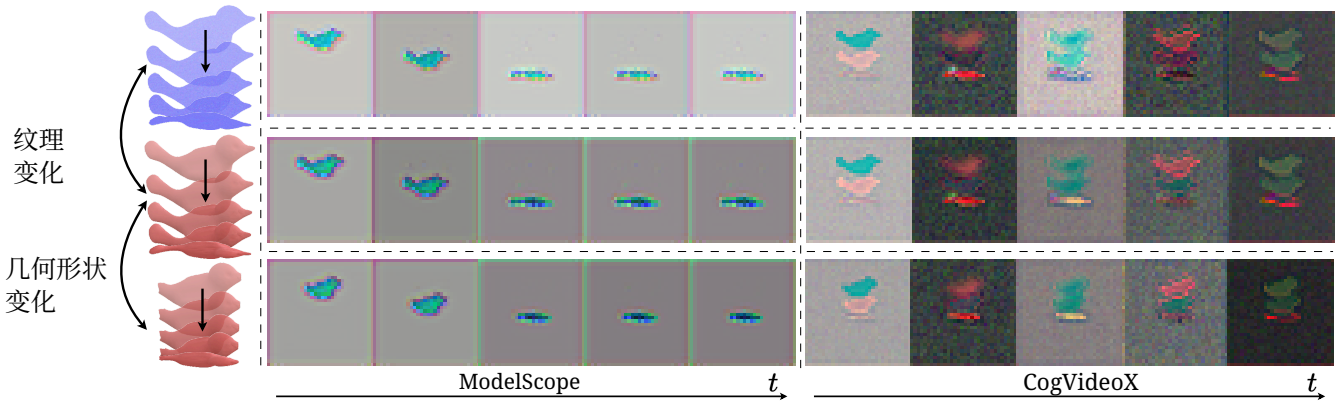

图 A2：**结构相似性补充材料**。通过主成分分析可视化的视频潜变量，在视频扩散模型的不同视频编码器中保留了全局结构，具体包括 ModelScope（左）和 CogVideoX（右）；纹理（上）和几何形状（下）两种变体仅在局部、部件级细节上有所不同。这些变体遵循相同的运动序列，由我们的大语言模型估算出相同的初始化参数后进行模拟，并在此基础上进行后续优化。

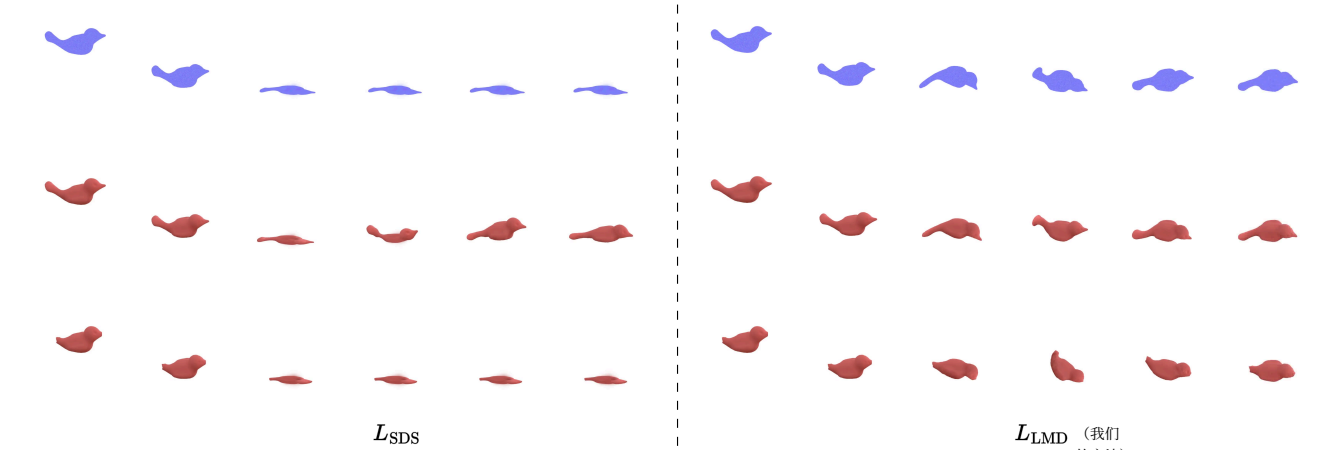

图 A3：**额外消融实验：** L_{LMD}。我们将可学习运动蒸馏损失 L_{LMD} 与传统的分数蒸馏损失 L_{SDS}进行比较，使用相同的初始化参数和提示“一只弹性鸟正在撞击地面。”优化后， L_{LMD} 在纹理和几何形状两种变体中均实现了连贯的弹性运动模式，忠实地体现了文本提示。相比之下， L_{SDS} 未能产生预期的动力学效果，导致两种变体表现出不同的行为（完整动力学过程请参见补充视频）。

物理真实感	外星人	海芋海胆	Bird	Cat	奶油	液滴	Hat	Jam	Lego	飞机	橡皮泥	沙堡	电话	牙膏	Sum
我们的方法优于 PhysDream	**26 / 28**	**13 / 13**	**18 / 19**	**13 / 13**	**13 / 13**	**18/19**	**19 / 19**	**25/ 28**	**28 /28**	**25/ 28**	**16/ 19**		**17 / 19**	**11 / 13**	3 / 13	**27 / 28**	**272 / 300**
我们的方法优于 DreamPhysics	**19/ 19**	**11 / 19**	**25 / 28**	**18 / 28**	**11 / 19**	**17 /19**		**24 / 28**	**16/ 19**	**19 /19**	**19/ 19**	**24/ 28**	**13 / 13**	3 / 19	**21 / 28**	**16 / 19**	**256 / 324**
我们的方法优于 OmniPhysGS	**9 / 13**	**18 / 19**	**19 / 19**	**19 / 19**	**19 / 13**	**13 /13**		**19 /19**	**16/ 19**	**12 /13**	**13/ 19**	**19/ 19**	**24 / 28**	**26 / 28**	**19 / 19**	**13 / 13**	**258 / 279**
我们的方法优于 PhysFlow	**13 / 19**	**21 / 28**	**13 / 13**	**19 / 19**	**20 / 28**	**27 /28**		**11 / 13**	**10/ 13**	**18 /19**	**12/ 13**	**12/ 13**	**18 / 19**	8 / 19	**13 / 19**	**18 / 19**	**233 / 282**
提示词遵循度	外星人	海芋海胆	Bird	Cat	奶油	液滴	Hat	Jam	Lego	飞机	橡皮泥	沙堡	电话	牙膏	Sum
我们的方法 优于 PhysDreamer	**26 / 28**	**13 / 13**	**19 / 19**	**13 / 13**	**13 / 13**	**18/19**	**17 / 19**	**25/ 28**	**27 /28**	**26/ 28**	**15/ 19**		**18 / 19**	**12 / 13**	6 / 13	**27 / 28**	**275 / 300**
我们的方法优于 DreamPhysics	**18/ 19**	**14 / 19**	**26 / 28**	**23 / 28**	8 / 19	**18 / 19**	**24 / 28**		**18/ 19**	**16/ 19**	**17 / 19**	**23 /28**	**13 / 13**	5 / 19	**24 / 28**	**17 / 19**	**264 / 324**
我们的方法优于 OmniPhysGS	**8 / 13**	**18 / 19**	**19 / 19**	**19 / 19**	**19 / 13**	**13 /13**		**19 / 19**	**15/ 19**	**11 /13**	**14/ 19**	**19/ 19**	**26 / 28**	**25 / 28**	**19 / 19**	**13 / 13**	**257 / 279**
我们的方法优于 PhysFlow	**15 / 19**	**22 / 28**	**12 / 13**	**18 / 19**	**21 / 28**	**27 / 28**		**12 / 13**	**10/ 13**	**17 /19**	**12/ 13**	**11/ 13**	**18 / 19**	**11 / 19**	**13 / 19**	**18 / 19**	**237 / 282**

表 A5：用户研究中的明确投票数（我们的票数 / 总票数）。**粗体** 表示我们的方法获得了多数票的情况。

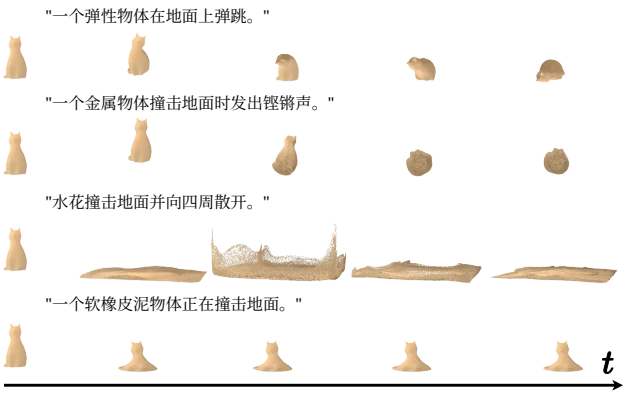

图 A4：**对变化文本提示的鲁棒性**。 我们的方法使用同一个合成猫场景，根据不同的文本提示成功模拟了多种动态效果。从上到下，结果成功展示了弹性、金属、牛顿流体和橡皮泥材料对应的行为。

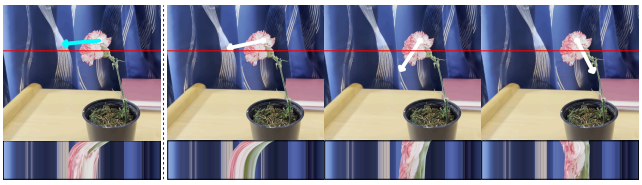

图 A5：**交互式评估**。 我们通过底部显示的时空切片将产生的振荡可视化。

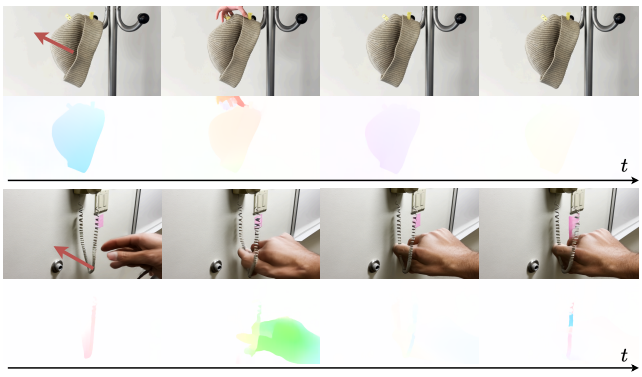

图 A6：**PhysFlow 失败**。Hat（顶部）和电话（底部）场景中生成的视频帧及其对应光流的可视化。光流未能反映仿真所需的真实物理运动。

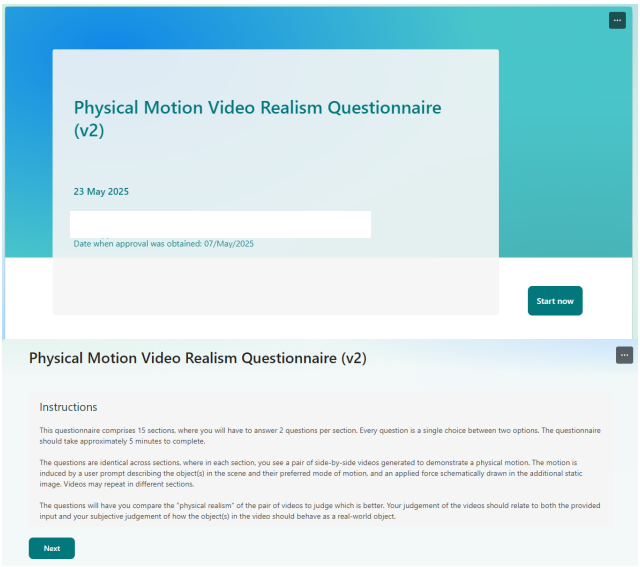

图 A7：**说明页面**。用户研究问卷介绍页面的截图。经相关研究伦理委员会批准公开的个人或机构信息已用白色矩形遮挡。

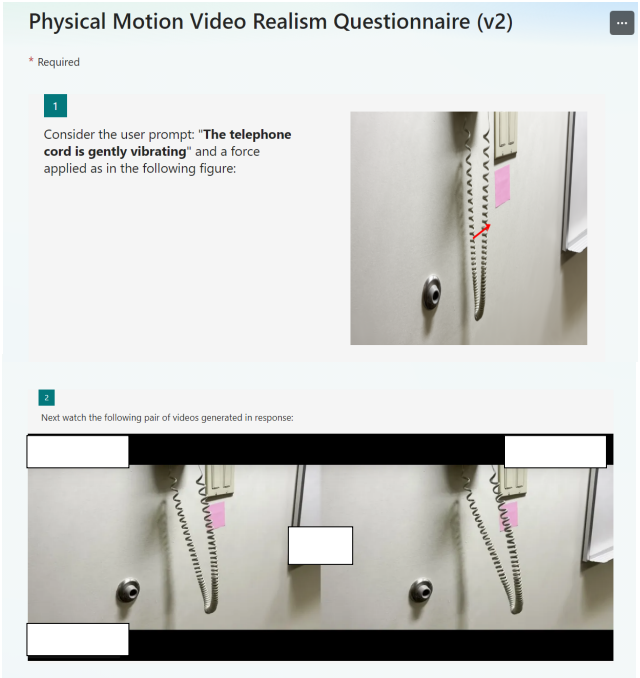

图 A8：**视频比较界面**。界面截图，显示文本提示、施加的力的方向以及用于比较的成对模拟视频。敏感或可识别身份的信息已遮挡。

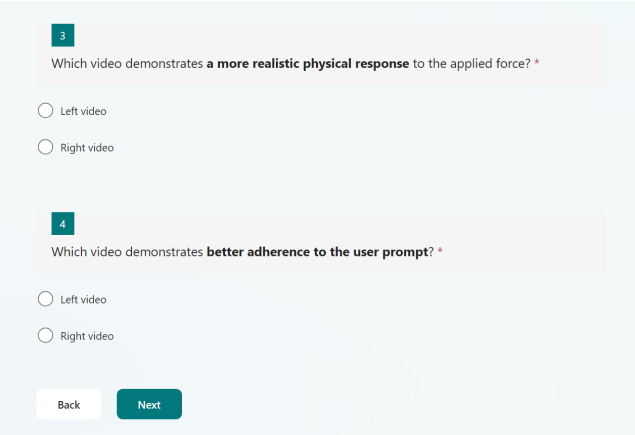

图 A9：**评估问题**。用于收集用户偏好的两个问题截图，是所报告主观评估指标的基础。

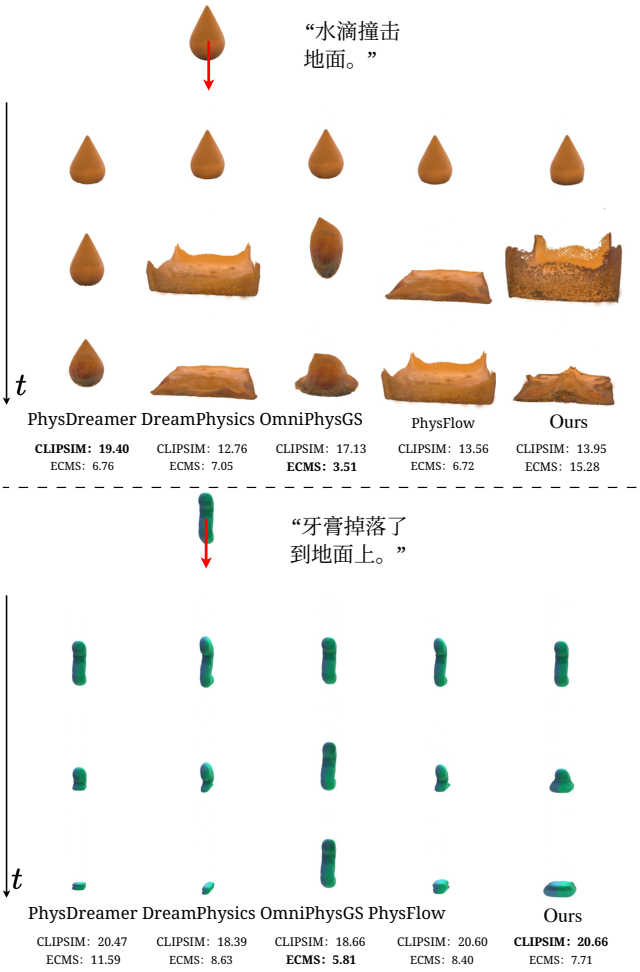

图 A10：液滴（上）和牙膏（下）场景的定性结果，并标注了 CLIPSIM(%)(↑) 和 ECMS (↓) 指标。每个场景的最佳结果以**粗体**显示。