Learning Physics-Grounded 4D Dynamics with Neural Gaussian Force Fields

Shiqian Li[1*], Ruihong Shen[1*], Junfeng Ni[2], Chang Pan[1], Chi Zhang[1]✉, Yixin Zhu[1]✉
[1]Peking University [2]Tsinghua University

ICLR

Background

Human intuitive physics vs. AI models

1. From infancy, humans develop robust intuitive physics understandings.
2. AI models frequently violates fundamental physical laws like gravity, solidity, and object permanence.


Permanence and consistency

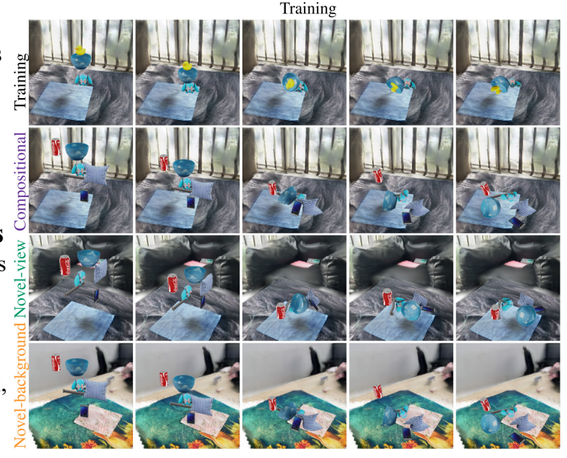

Gravity and support


Sora's failure case

Previous Methods

1. **Video Generation Models (Sora,Cosmos) :** Produce good visuals but overfit to appearance, hallucinating physics during complex multi-object interactions.
2. **Physical Simulators (PhysGaussian, PhysGen3D) :** Achieve strong physical fidelity but suffer from inflexibility and prohibitive computational costs.

GSCollision Dataset



· **Massive scale**
640k rendered 4D videos (~4TB).

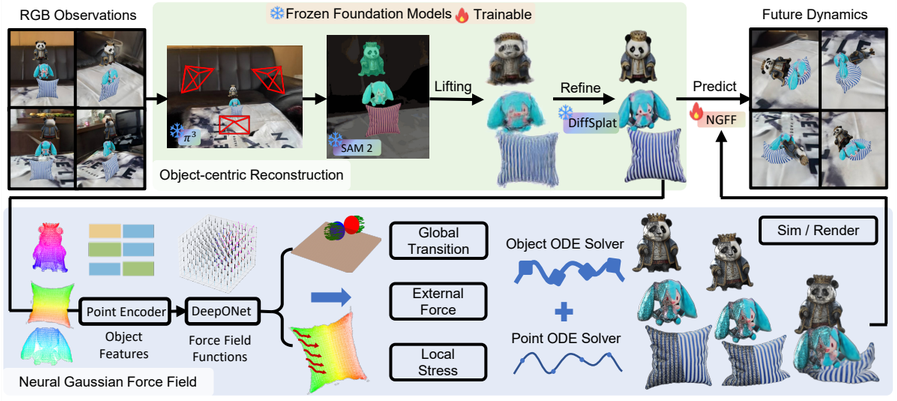
· **Diverse physics**
10 everyday objects spanning soft to rigid materials.

· **Complex dynamics**
3,200 multi-object scenes (collisions, sliding, containment).

· **Rigorous testing:**
Built-in spatial, temporal, and compositional generalization splits.

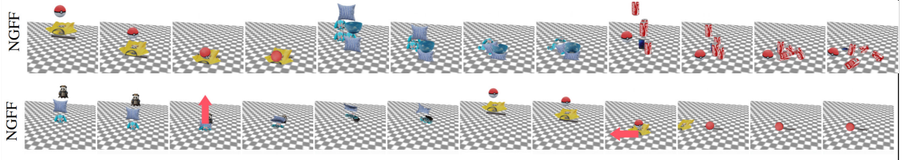
Our Method

Unify **object-centric** feed-forward 3D Gaussian reconstruction via **force fields** and Neural ODE solvers for efficient, **4D** dynamics prediction and rendering.



Dynamics prediction

Robust generalization and interactive force responses.



Video generation

Physically-consistent 4D video synthesis across novel views.

Quantitative results

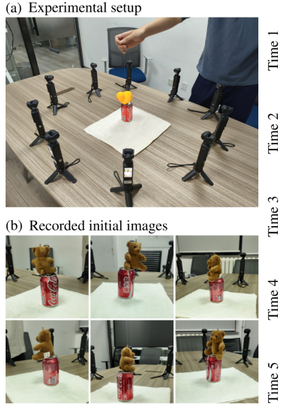
Superior physical accuracy with fast inference.

Model	Spatial			Temporal			Compositional			Time
	RMSE ↓	FPE ↓	R ↑	RMSE ↓	FPE ↓	R ↑	RMSE ↓	FPE ↓	R ↑	(s) ↓
VLM-MPM	0.31	0.77	0.30	0.33	0.90	0.30	0.36	0.90	0.31	39.29
GCN	0.13	0.48	0.41	0.17	0.59	0.40	0.15	0.51	0.35	0.35
Pointformer	0.10	0.39	0.62	0.13	0.54	0.60	0.16	0.59	0.43	**0.18**
NGFF w/o deform.	0.11	0.46	0.60	0.14	0.60	0.58	0.13	0.55	0.52	0.30
NGFF	**0.08**	**0.33**	**0.66**	**0.11**	**0.42**	**0.65**	**0.10**	**0.41**	**0.57**	0.36

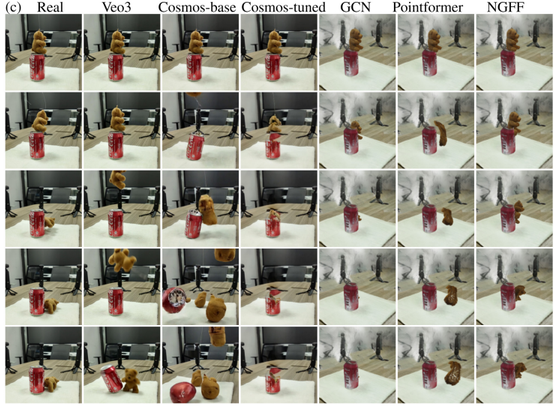
Model	Comp.				NV				All			
	VLM Eval.		Human Eval.		VLM Eval.		Human Eval.		VLM Eval.		Human Eval.	
	PhysR	PhotoR	PhysR	PhotoR	PhysR	PhotoR	PhysR	PhotoR	PhysR	PhotoR	PhysR	PhotoR
Cosmos	0.34	**0.42**	0.29	0.43	0.39	0.42	0.26	0.39	0.20	0.32	0.28	0.41
Cosmos (tuned)	0.26	0.35	**0.57**	**0.58**	**0.49**	**0.40**	**0.63**	0.62	0.24	0.36	0.59	0.58
NGFF-V	**0.47**	**0.42**	0.56	0.55	0.44	0.38	0.55	0.54	**0.30**	0.35	0.55	0.55
Veo3	–	–	–	–	–	–	–	–	0.29	**0.41**	0.53	**0.64**
PhysGen3D	–	–	–	–	–	–	–	–	0.19	0.35	**0.57**	0.58

Real-world validation

Sim2real transfer for real-world interactions.


(a) Experimental setup

(b) Recorded initial images



Contact

Paper | Website | Codebase | Dataset | vimeo