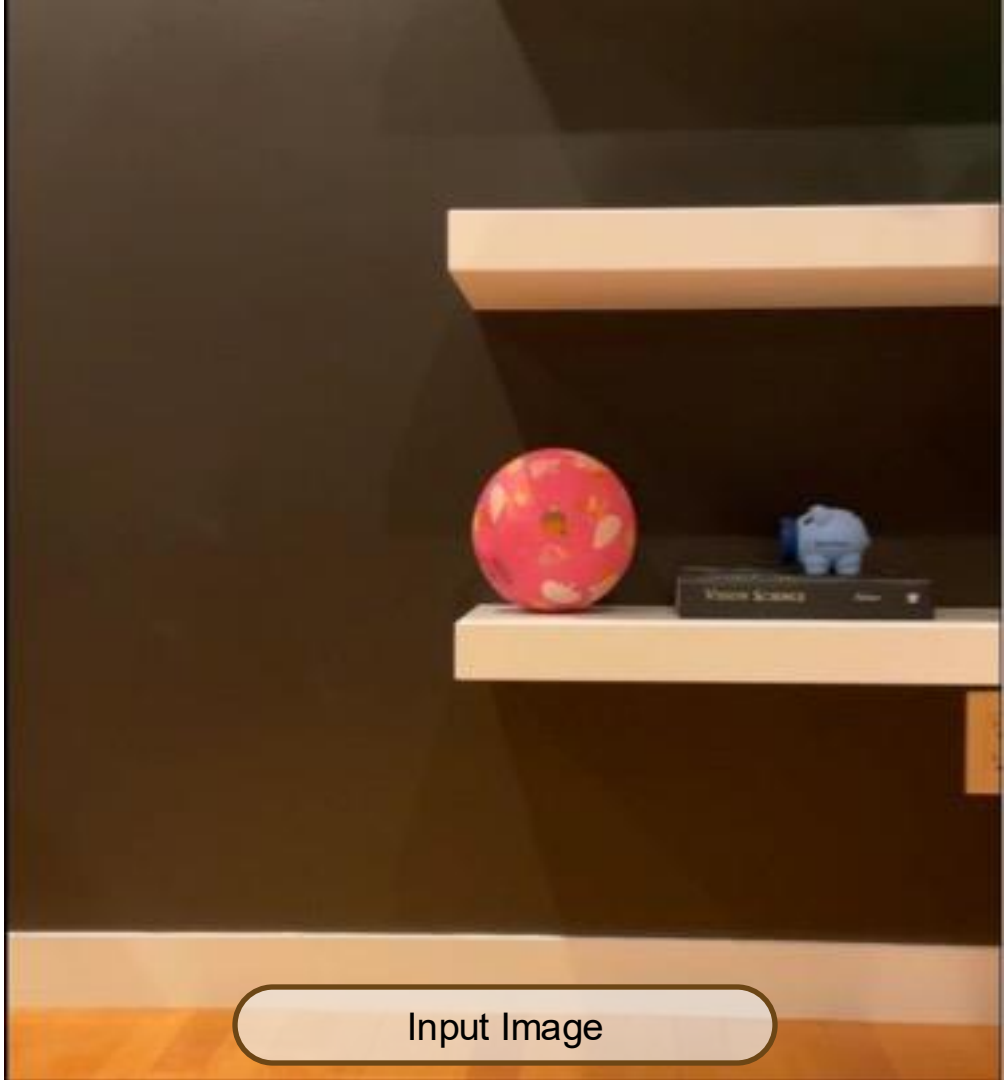
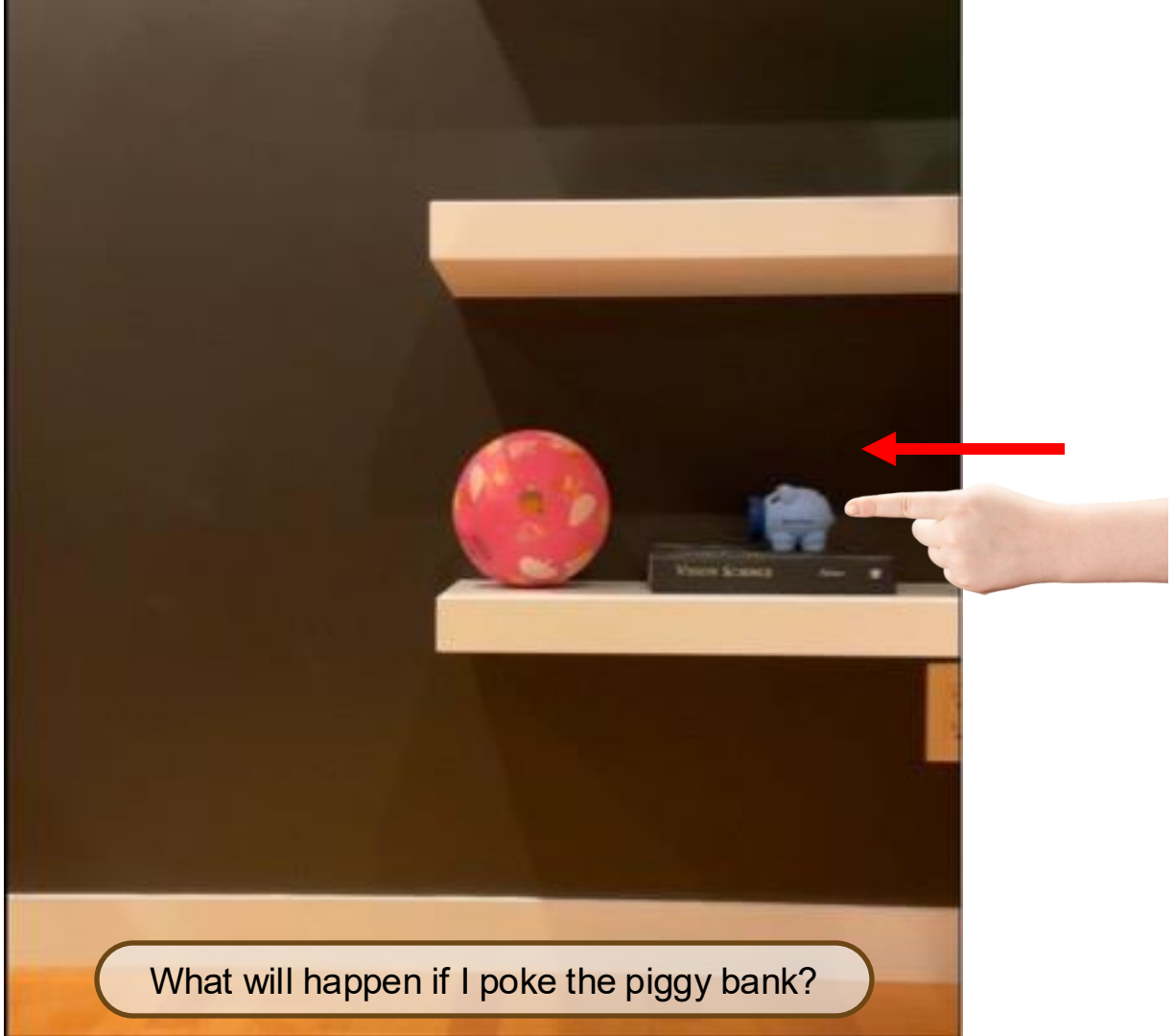
Physical Inductive Biases for Interactive Image and Video

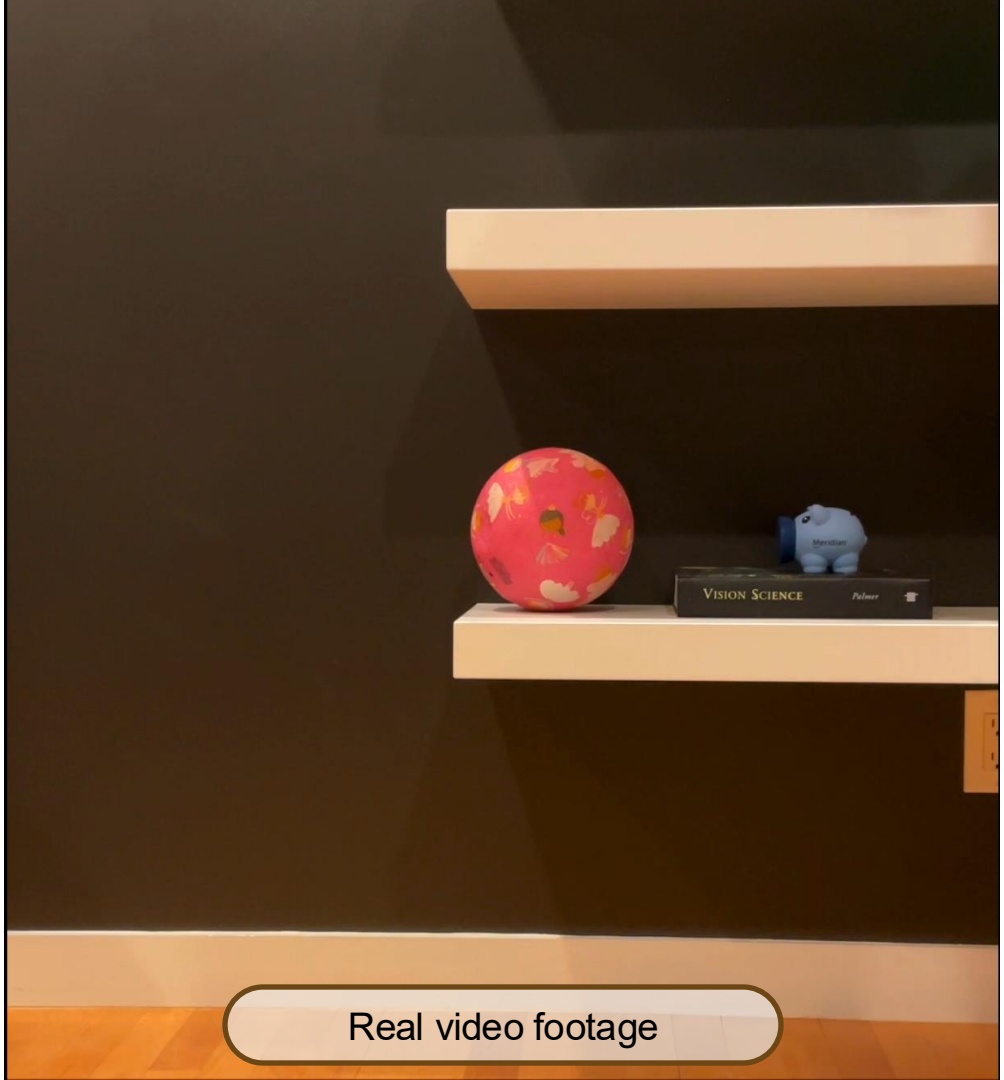
Shenlong Wang

CVPR 2025 Ind3D Workshop

Jun 12, 2025

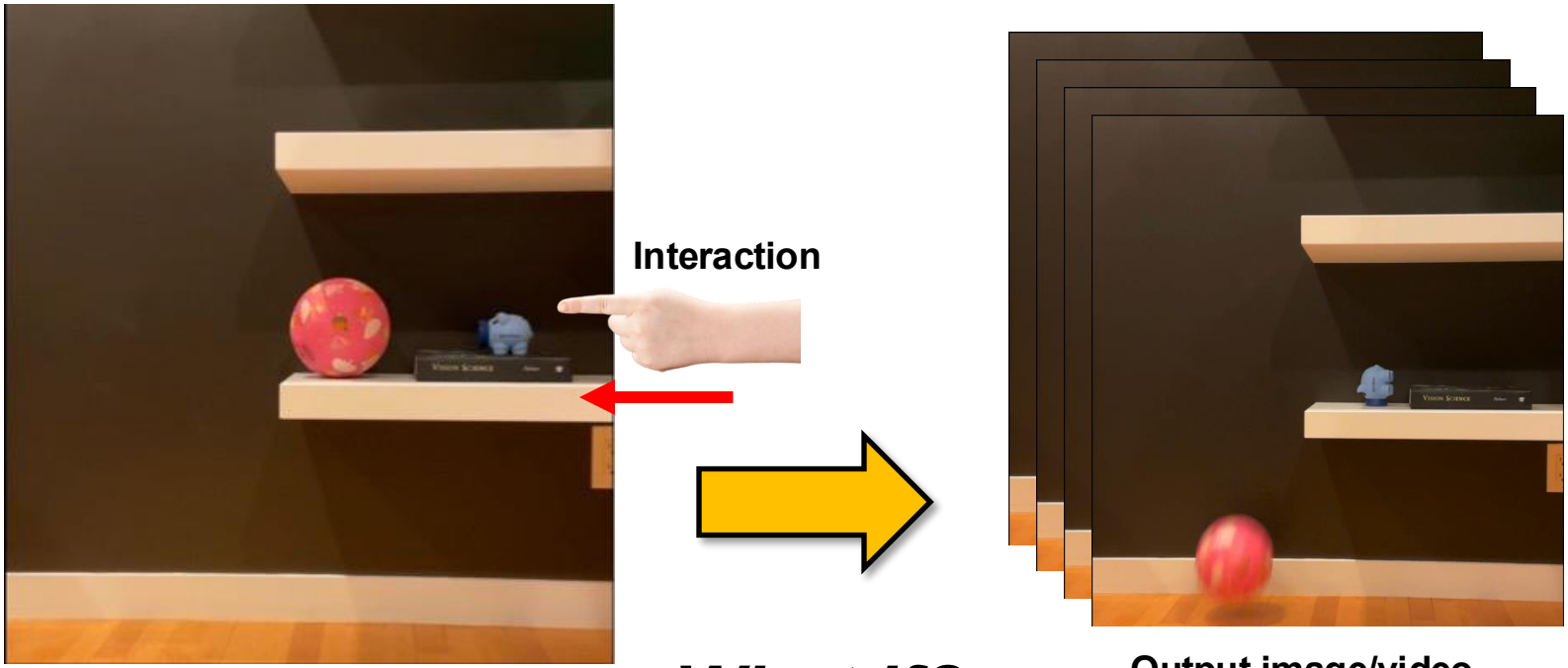

Input Image


What will happen if I poke the piggy bank?

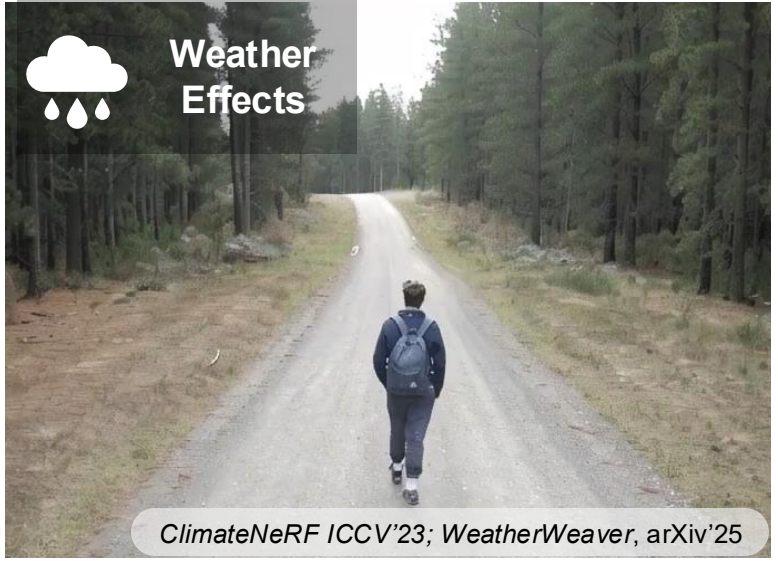


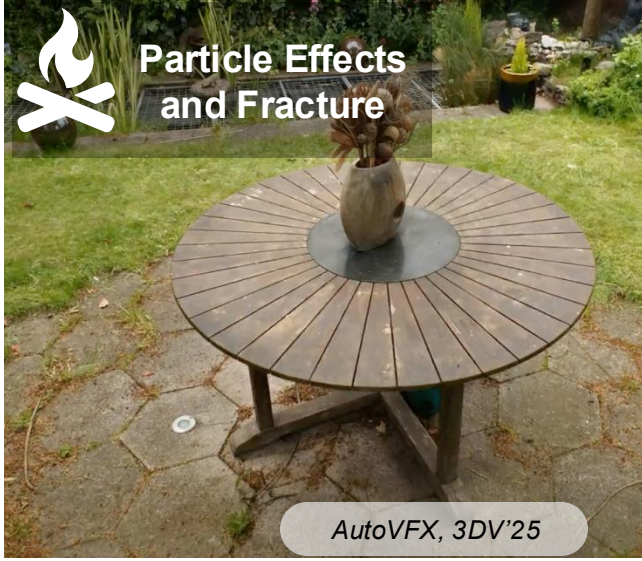
Real video footage

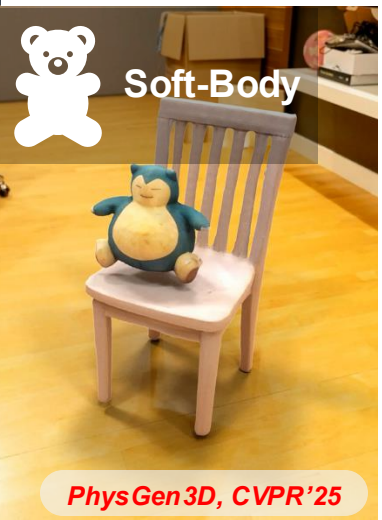
Goal: turn image(s) into an interactive world




Rigid Dynamics
PhysGen, ECCV'24


Weather Effects
ClimateNeRF ICCV'23; WeatherWeaver, arXiv'25


Particle Effects and Fracture
AutoVFX, 3DV'25


Soft-Body
PhysGen3D, CVPR'25


Lighting Effects
UrbanIR, 3DV'25; IRIS, CVPR'25

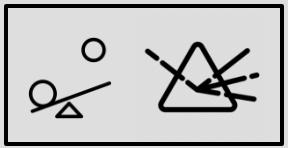

Articulation
DRAWER, CVPR'25

Key Idea: Perceive, Simulate and Render

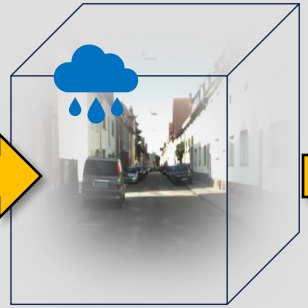
Input Image(s)

User interaction

Scene Models
(*Geometry, Semantics, Appearance, Physics…*)



Physical Simulation
(*Lighting, Dynamics…*)



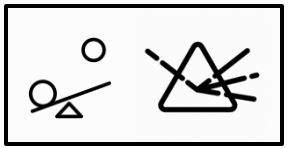
Updated Scene

Output Images

Perceive ***Simulate*** ***Render***

Key Idea: Perceive, Simulate and Render

Scene Models
(***Geometry, Semantics, Appearance, Physics…***)



Physical Simulation
(***Lighting, Dynamics…***)

A suitable representation that facilitates physics simulation

A good inductive bias in physics for our task

Perceive ***Simulate*** ***Render***

Poking an image

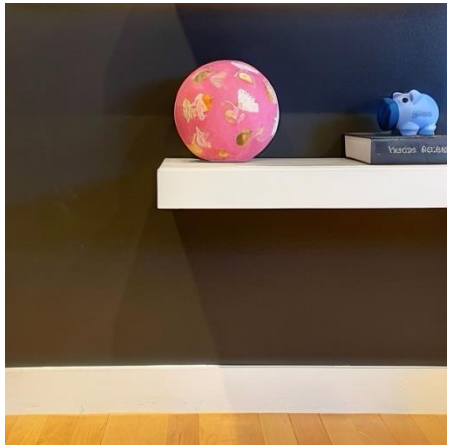
Shaowei Liu

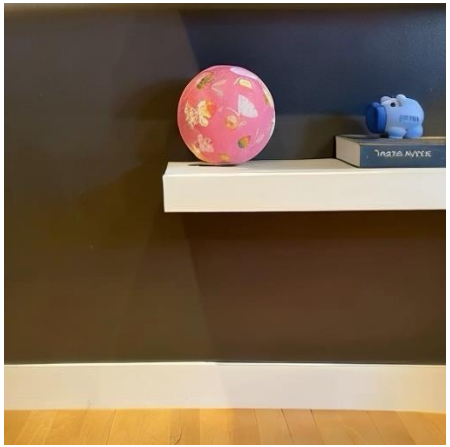
PhysGen: Rigid-Body Physics-Grounded Image-to-Video Generation Shaowei Liu, Zhongzheng Ren, Saurabh Gupta*, Shenlong Wang*, ECCV 2024

Generative Image2Video models…

Shaowei Liu

photorealistic animation of a piggy bank, placed on a smooth surface, being poked firmly by an unseen force from the right side, causing it to slide and rotate from right to left. As the piggy bank moves, it hits the pink ball, displacing the ball slightly before the ball rolls and falls off the surface onto the ground. The piggy bank continues sliding a little further and comes to a stop exactly where the ball was initially located. The motion should be smooth, with realistic physics, lighting, and shadows.





SEINE **I2VGEN-XL** **DynamiCrafter** **Kling AI**

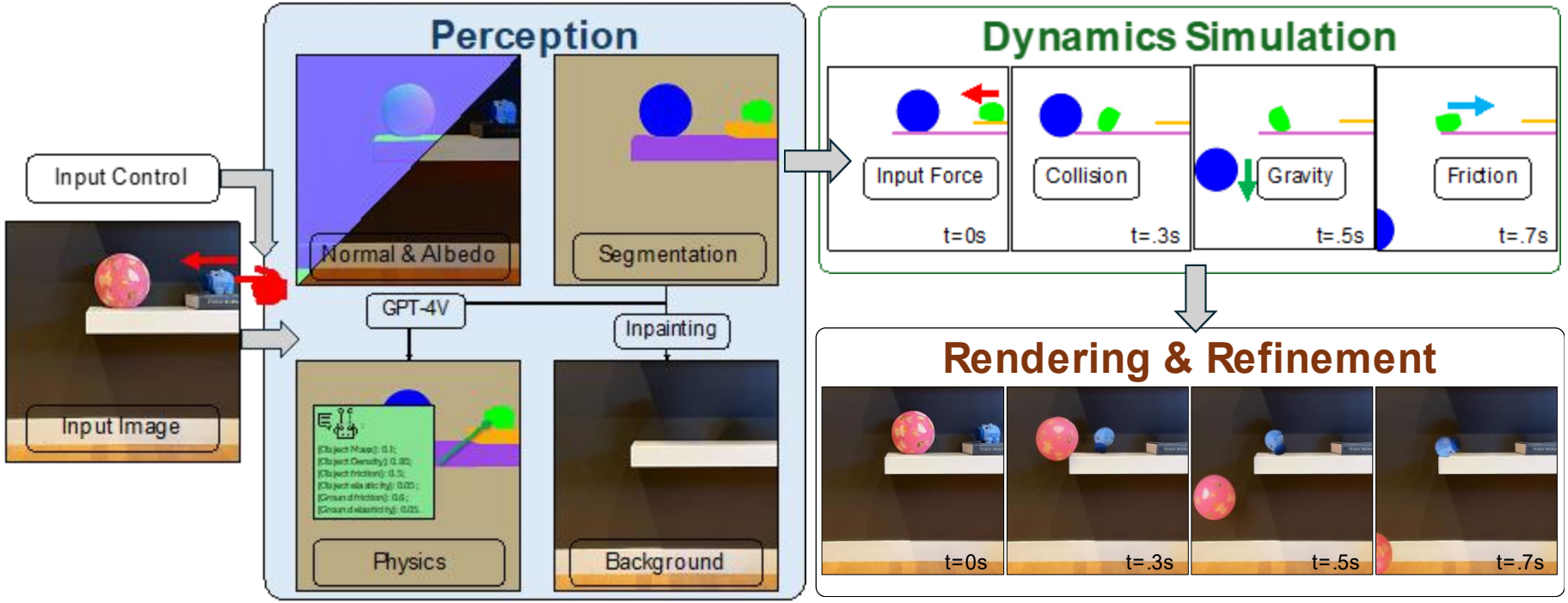
Latest open-sourced model

Latest closed-source model

Sota i2v diffusion models by Aug 2024

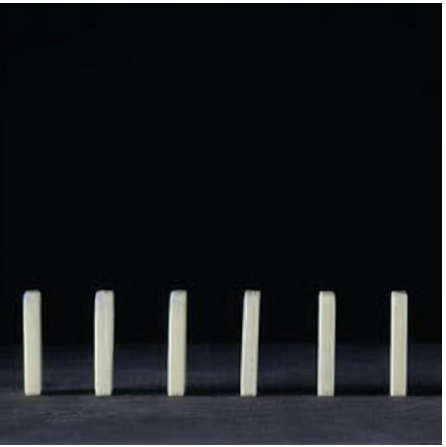
Perceive → Simulate → Render → Video

Shaowei Liu

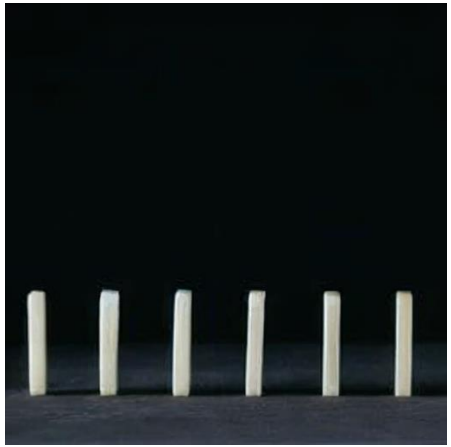


PhysGen: Rigid-Body Physics-Grounded Image-to-Video Generation Shaowei Liu, Zhongzheng Ren, Saurabh Gupta*, Shenlong Wang*, ECCV 2024

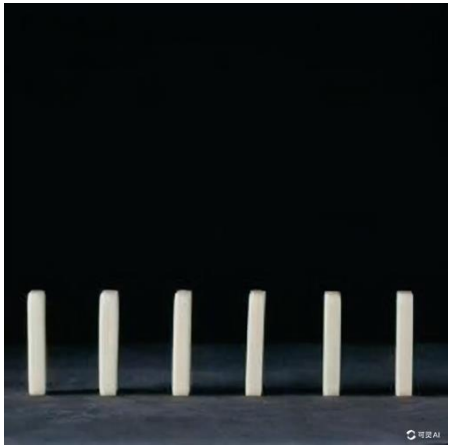
Qualitative results



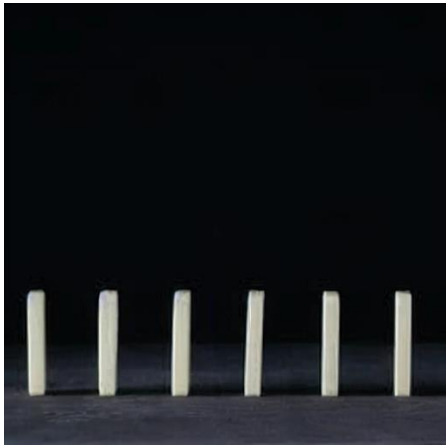
SEINE



I2VGEN-XL



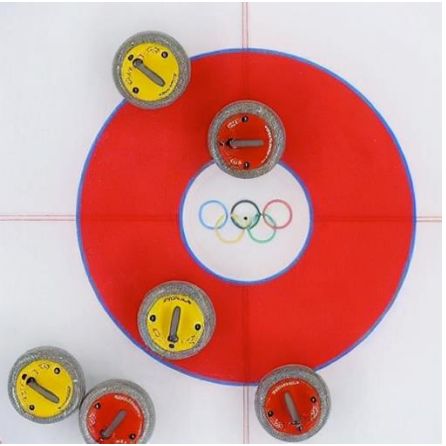
KlingAI



Ours

PhysGen: Rigid-Body Physics-Grounded Image-to-Video Generation Shaowei Liu, Zhongzheng Ren, Saurabh Gupta*, Shenlong Wang*, ECCV 2024

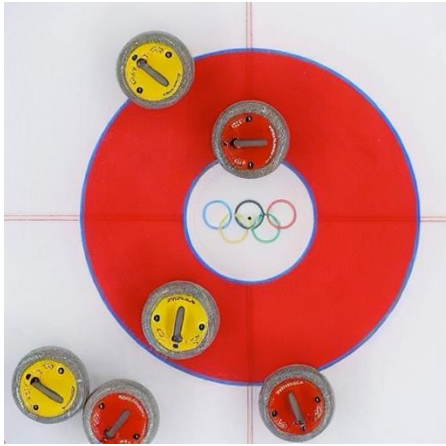
Qualitative results



SEINE

I2VGEN-XL

DynamiCrafter



Ours

PhysGen: Rigid-Body Physics-Grounded Image-to-Video Generation Shaowei Liu, Zhongzheng Ren, Saurabh Gupta*, Shenlong Wang*, ECCV 2024

Qualitative results

SEINE

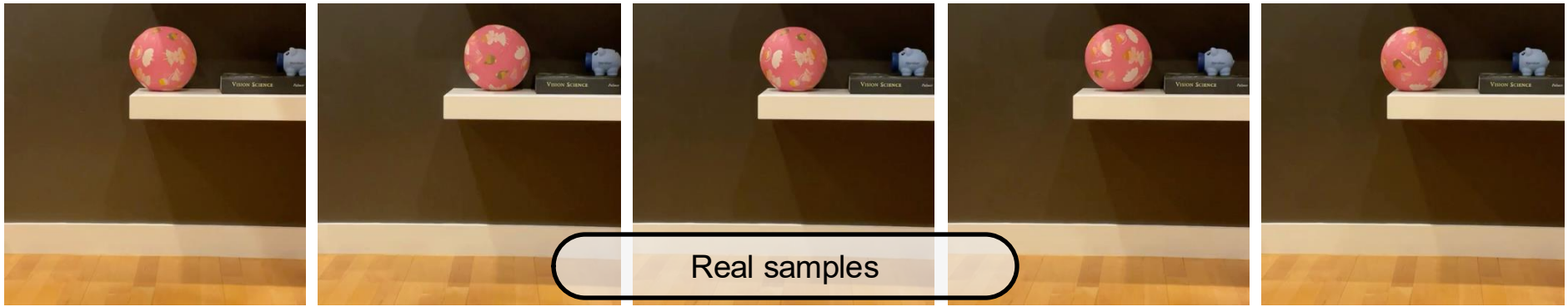
I2VGEN-XL

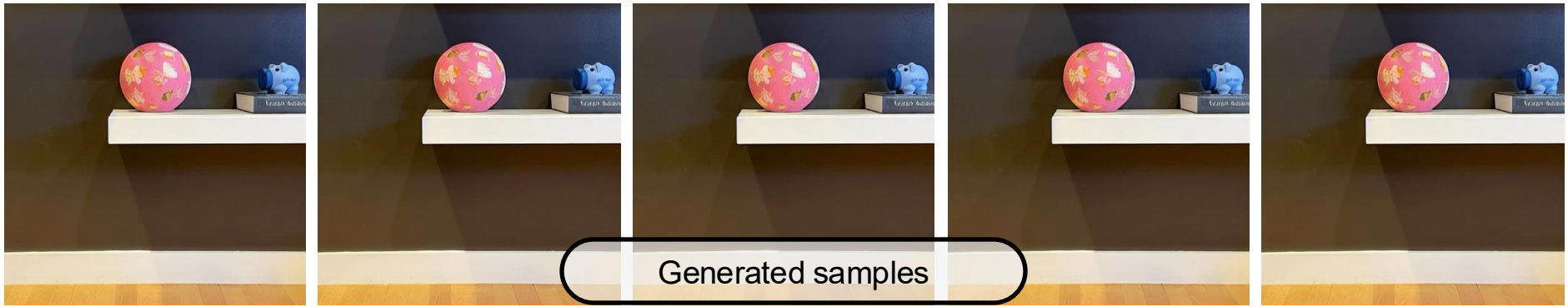
DynamiCrafter

Ours

PhysGen: Rigid-Body Physics-Grounded Image-to-Video Generation Shaowei Liu, Zhongzheng Ren, Saurabh Gupta*, Shenlong Wang*, ECCV 2024

Controllability & Diversity

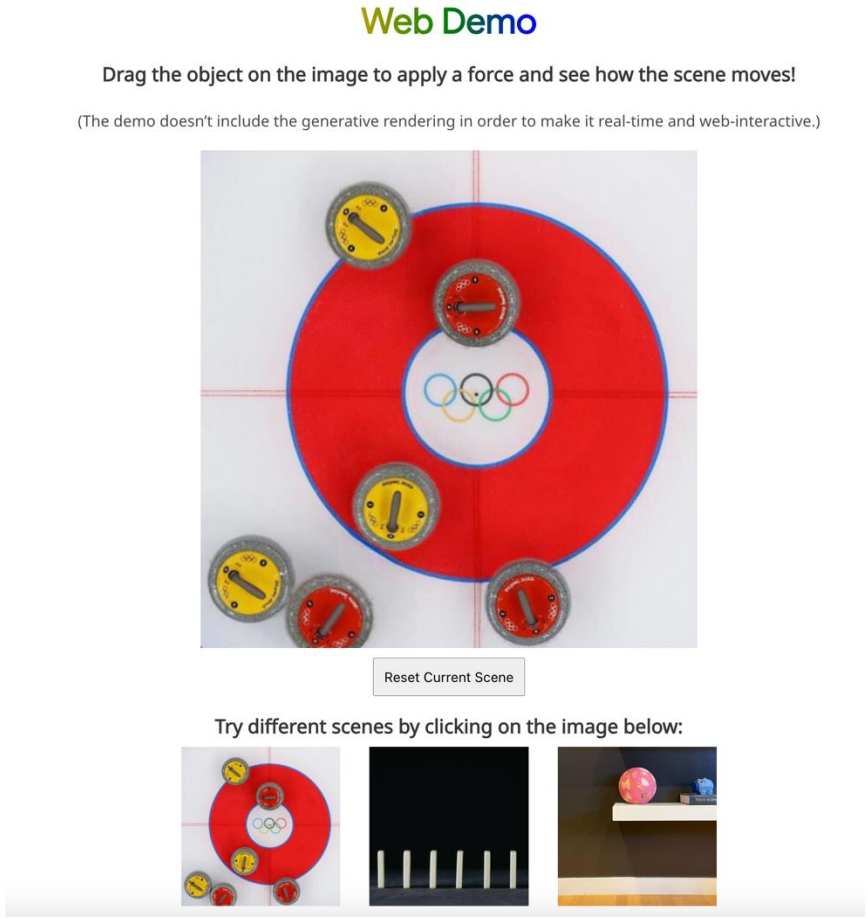




PhysGen: Rigid-Body Physics-Grounded Image-to-Video Generation Shaowei Liu, Zhongzheng Ren, Saurabh Gupta*, Shenlong Wang*, ECCV 2024

Demo

https://stevenlsw.github.io/physgen/



PhysGen: Rigid-Body Physics-Grounded Image-to-Video Generation Shaowei Liu, Zhongzheng Ren, Saurabh Gupta*, Shenlong Wang*, ECCV 2024

Extend to 3D dynamics and multiphysics

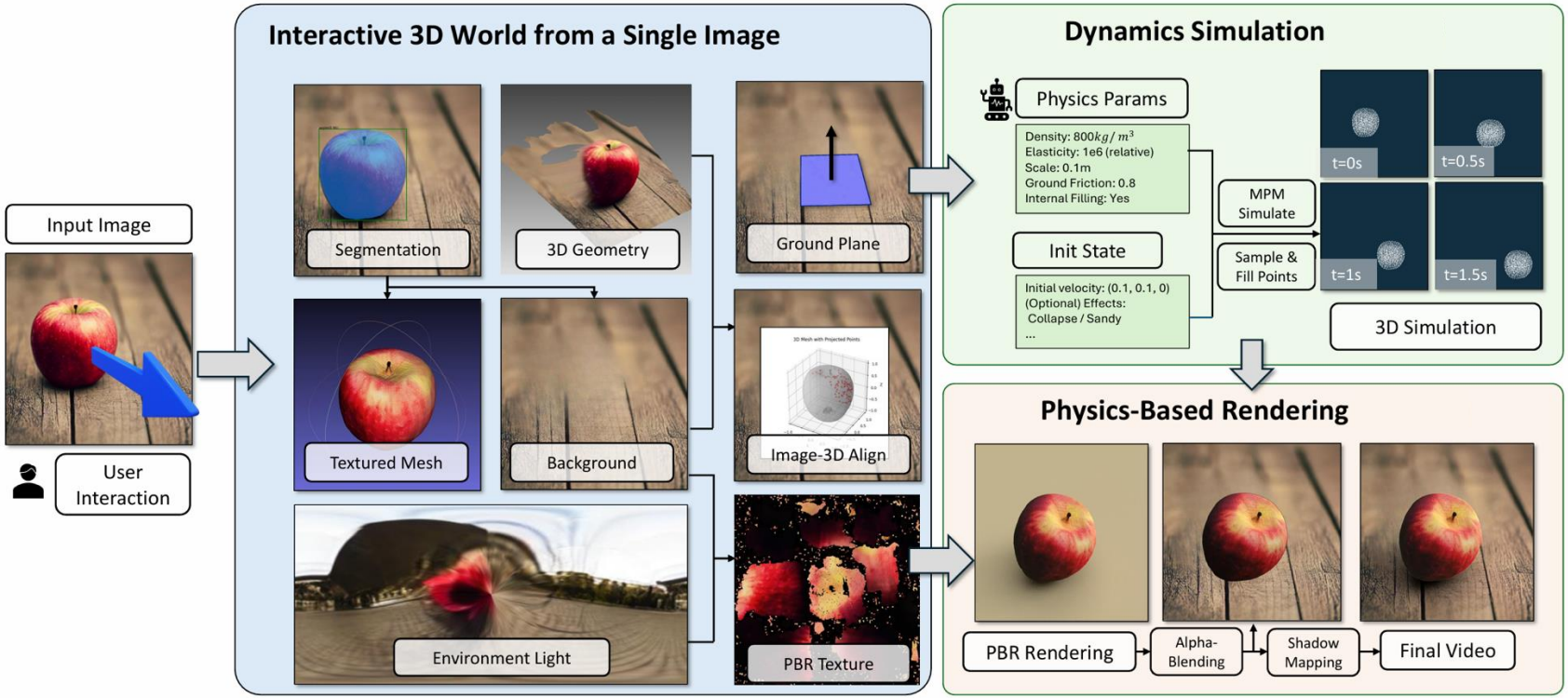
Boyuan Chen, Hanxiao Jiang, Shaowei Liu, Saurabh Gupta, Yunzhu Li, Hao Zhao, Shenlong Wang,
PhysGen3D: Crafting a Miniature Interactive World from a Single Image, CVPR 2025

Perceive → Simulate → Render


Boyuan Chen


Hanxiao Jiang

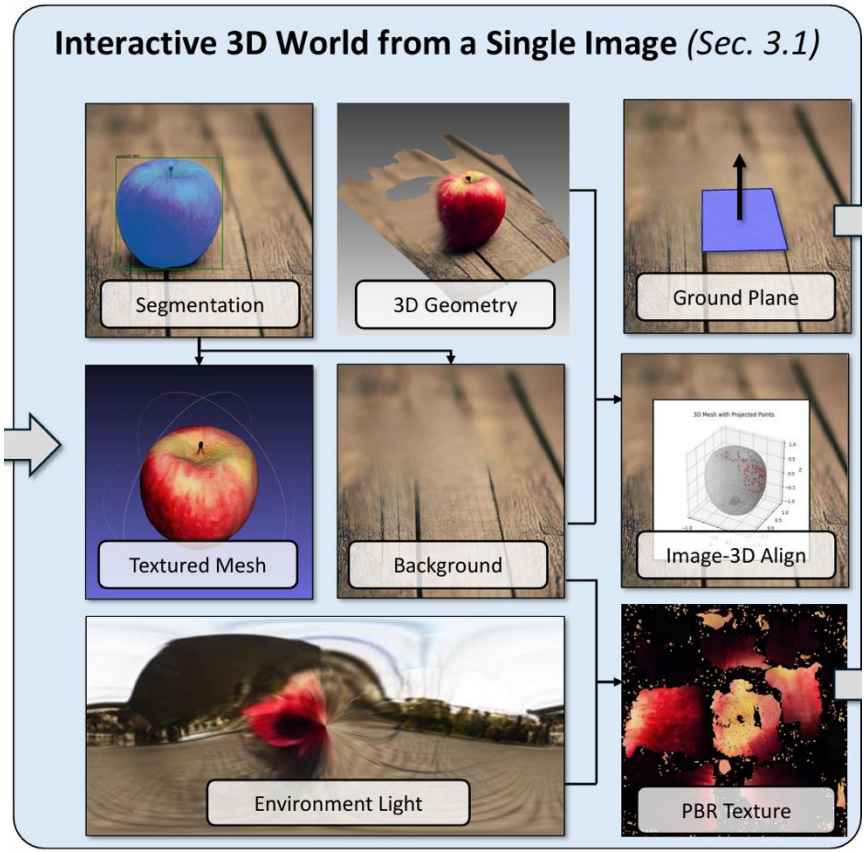

Shaowei Liu



Boyuan Chen, Hanxiao Jiang, Shaowei Liu, Saurabh Gupta, Yunzhu Li, Hao Zhao, Shenlong Wang,
PhysGen3D: Crafting a Miniature Interactive World from a Single Image, CVPR 2025

Image → Perceive → Simulate → Render → Video

- Segmentation: ***SegAnything v2***
- Scene Geometry: ***Dust3R***
- Ground Plane: ***Neural vanishing points***
- Textured Mesh: ***InstantMesh***
- Background: ***SDXL***
- Image-Mesh Align: ***SuperGlue + PnP***
- Light: ***Diffusion Light***
- PBR Material: ***Mitsuba3 InvRendering***
- Physical understanding***: GPT-4o***



Comparisons: 3D rigid body rolling

"Red apple rolls on the table."

Kling AI*

Runway Gen-3

Pika 1.5

Ours

*SOTA video models by Nov 2024

* *Kling AI uses privileged information from user-provided motion guidance.*

Comparisons: multiple objects

"The book falls and the orange rolls to the front."

Kling AI

Runway Gen-3

Pika 1.5

Ours

*SOTA video models by Nov 2024

** Kling AI uses privileged information from user-provided motion guidance.*

Comparisons: soft-body & scene interaction

"The toy falls off the chair."

Kling AI

Runway Gen-3

Pika 1.5

Ours

**SOTA video models by Nov 2024*

** Kling AI uses privileged information from user-provided motion guidance.*

Comparisons: material editing

The dog deflates and collapses.

Kling AI

Runway Gen-3

Pika 1.5

Ours

***SOTA video models by Nov 2024**

** We use a specific-purpose model from Pika for such physical effects comparison.*

Comparisons: Multi-Objects

"Three stuffed animal jump forward."

Kling AI

Runway Gen-3

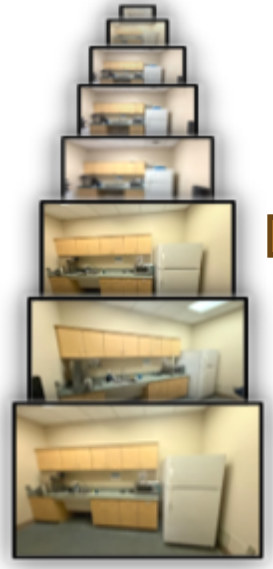
Pika 1.5

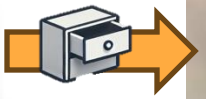
Ours

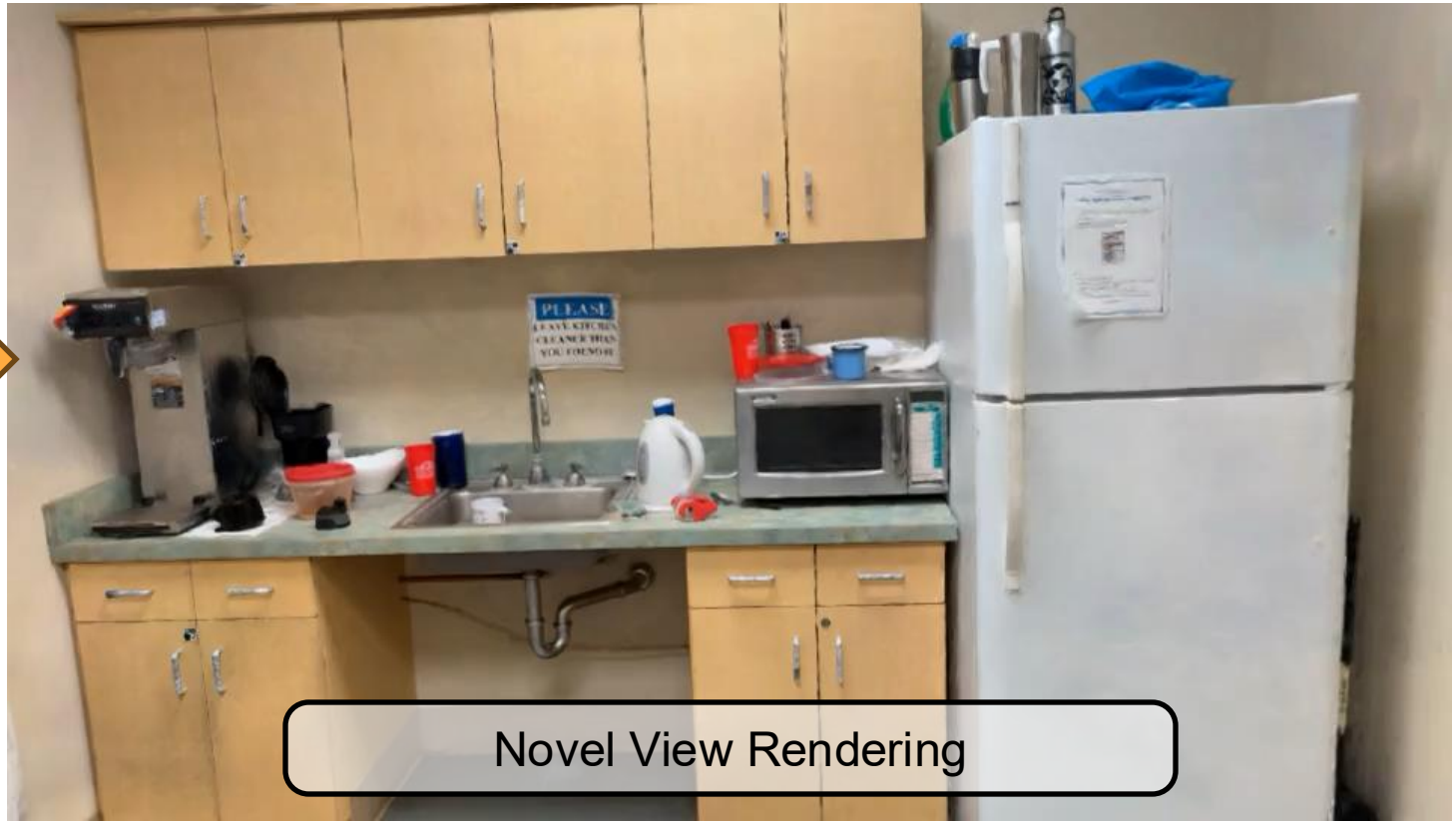
**SOTA video models by Nov 2024*

** Kling AI uses privileged information from user-provided motion guidance.*

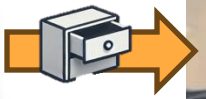
Video → Interactive Environment





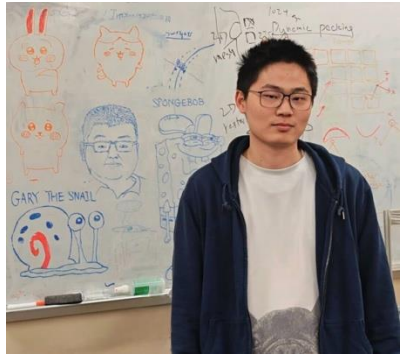


Video → Interactive Environment

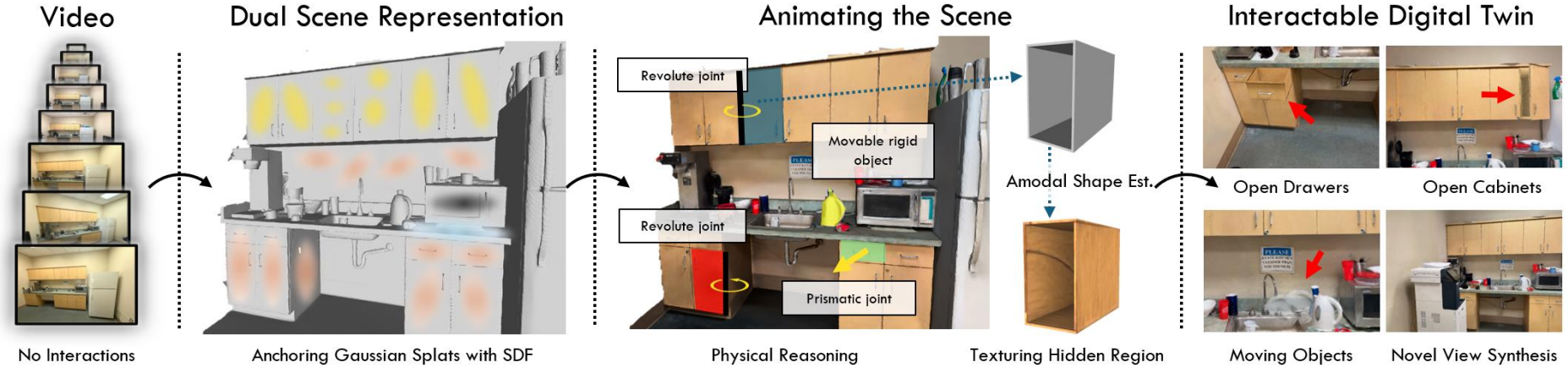


Interactive Scene with Full Articulation

Overview of DRAWER

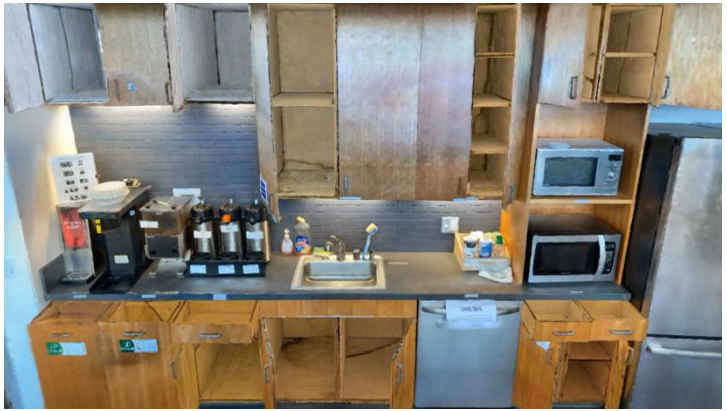


Hongchi Xia



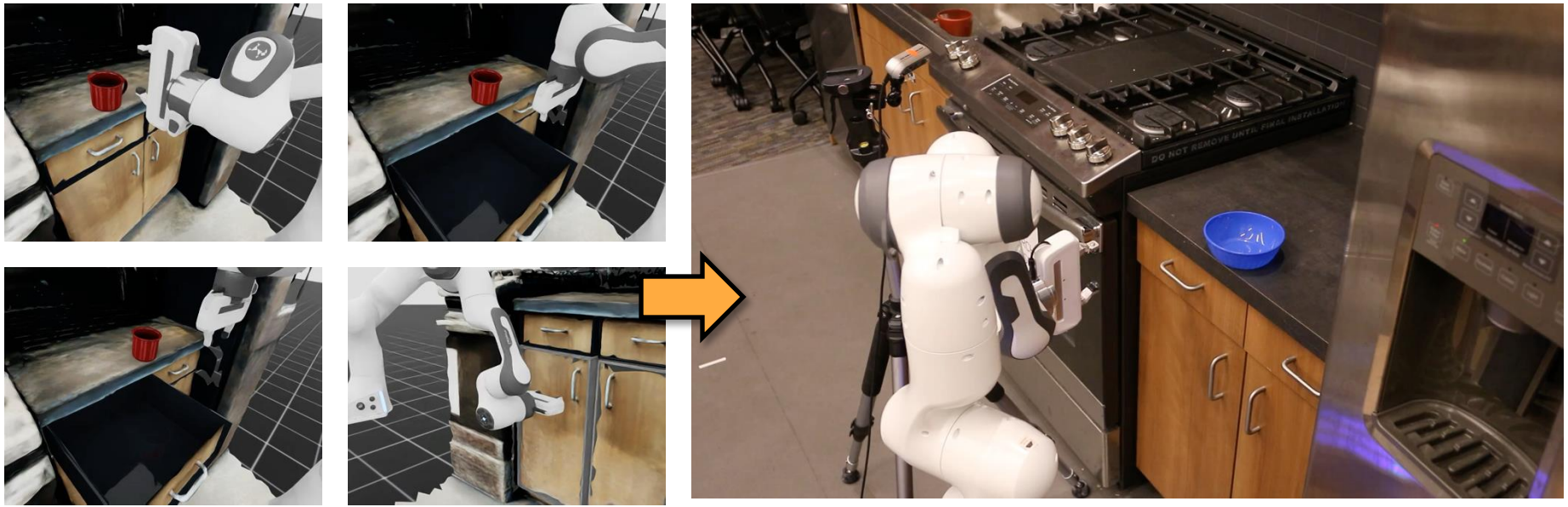
DRAWER: Digital Reconstruction and Articulation With Environment Realism Hongchi Xia, Entong Su, Marius Memmel, Arhan Jain, Raymond Yu, Numfor Mbiziwo-Tiapo, Ali Farhadi, Abhishek Gupta, Shenlong Wang, Wei-Chiu Ma International Conference on Computer Vision and Pattern Recognition (CVPR), 2025

Interactive Environment



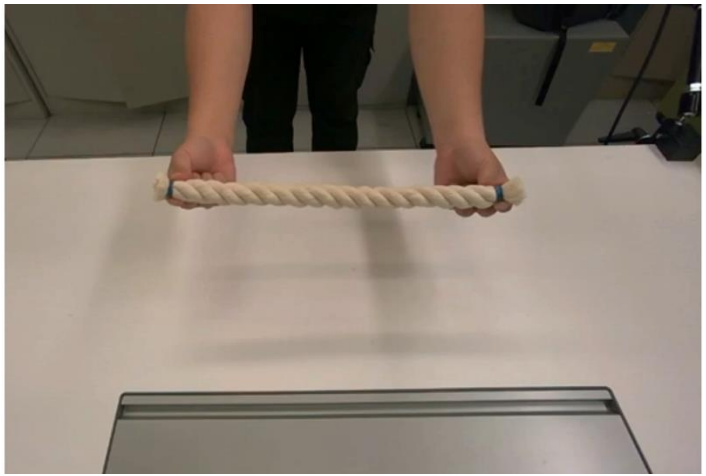
Application: Gaming

Application: Robot Learning via Real2Sim2Real

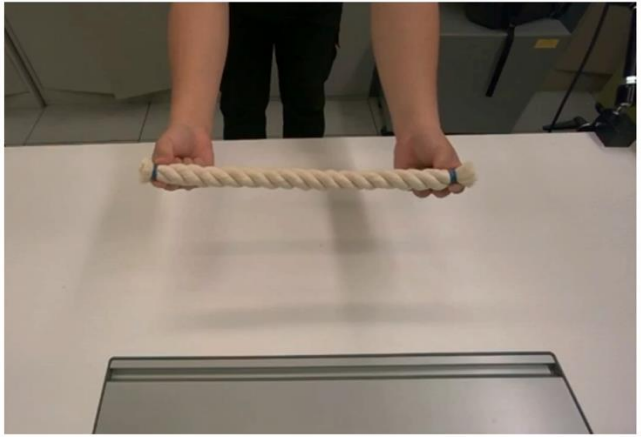


What if we observe dynamic interaction?

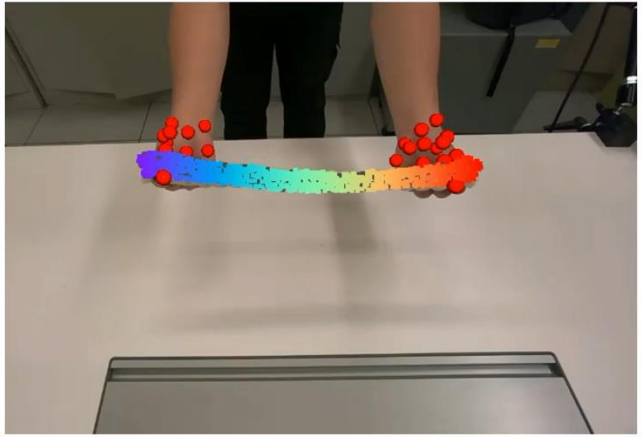
Hanxiao Jiang



PhysTwin: Physics-Informed Reconstruction and Simulation of Deformable Objects from Videos,
Hanxiao Jiang, Hao-Yu Hsu, Kaifeng Zhang, Hsin-Ni Yu, Shenlong Wang, Yunzhu Li, arXiv 2025



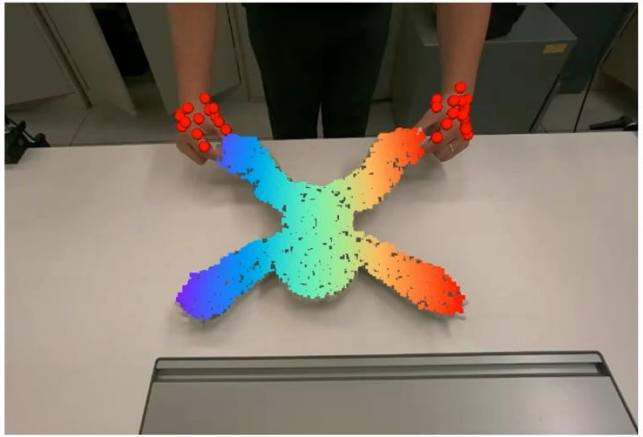
Observation



Reconstructed state/action

Interactable digital twin

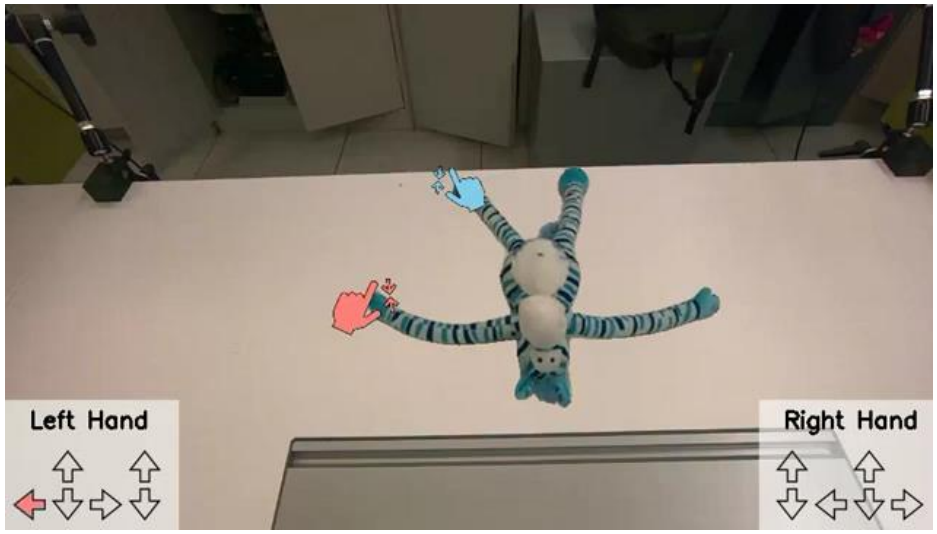
Observation



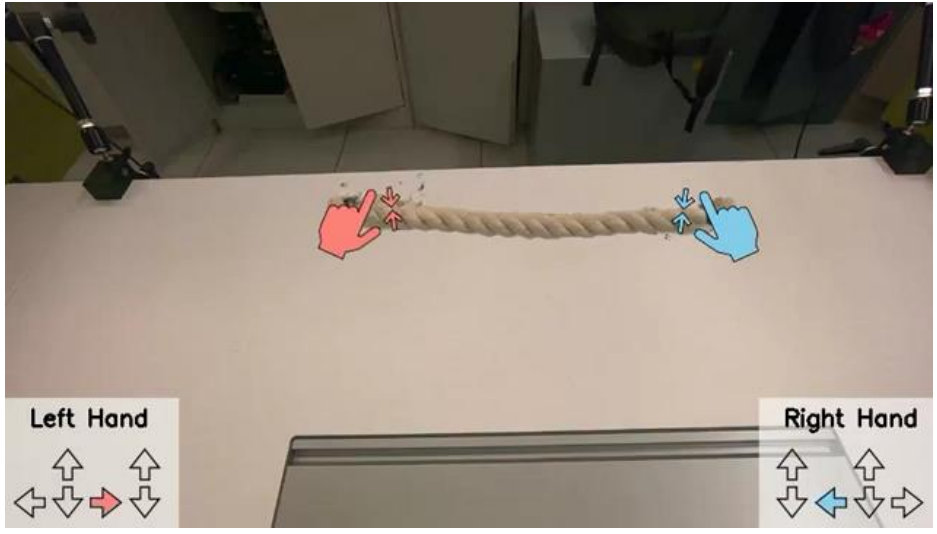
Reconstructed state/action

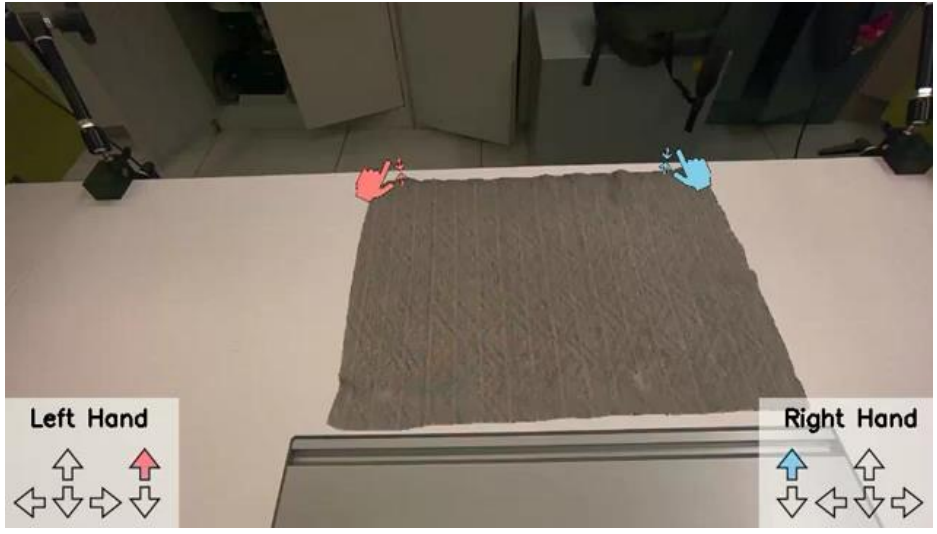
Interactable digital twin

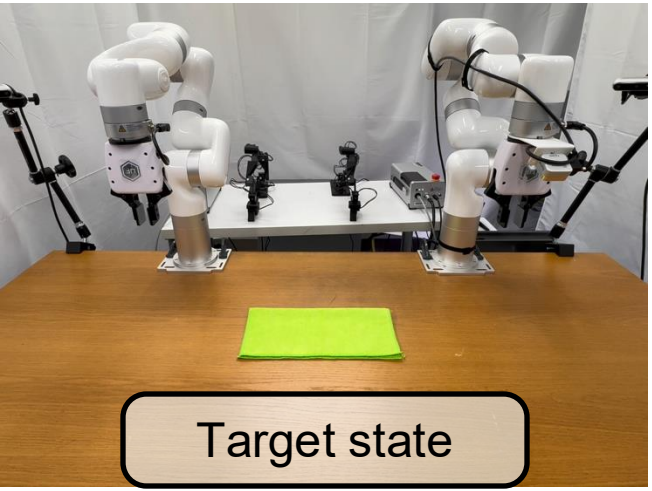
Hanxiao Jiang, Hao-Yu Hsu, Kaifeng Zhang, Hsin-Ni Yu, Shenlong Wang, Yunzhu Li
PhysTwin: Physics-Plausible Reconstruction and Resimulation from Videos

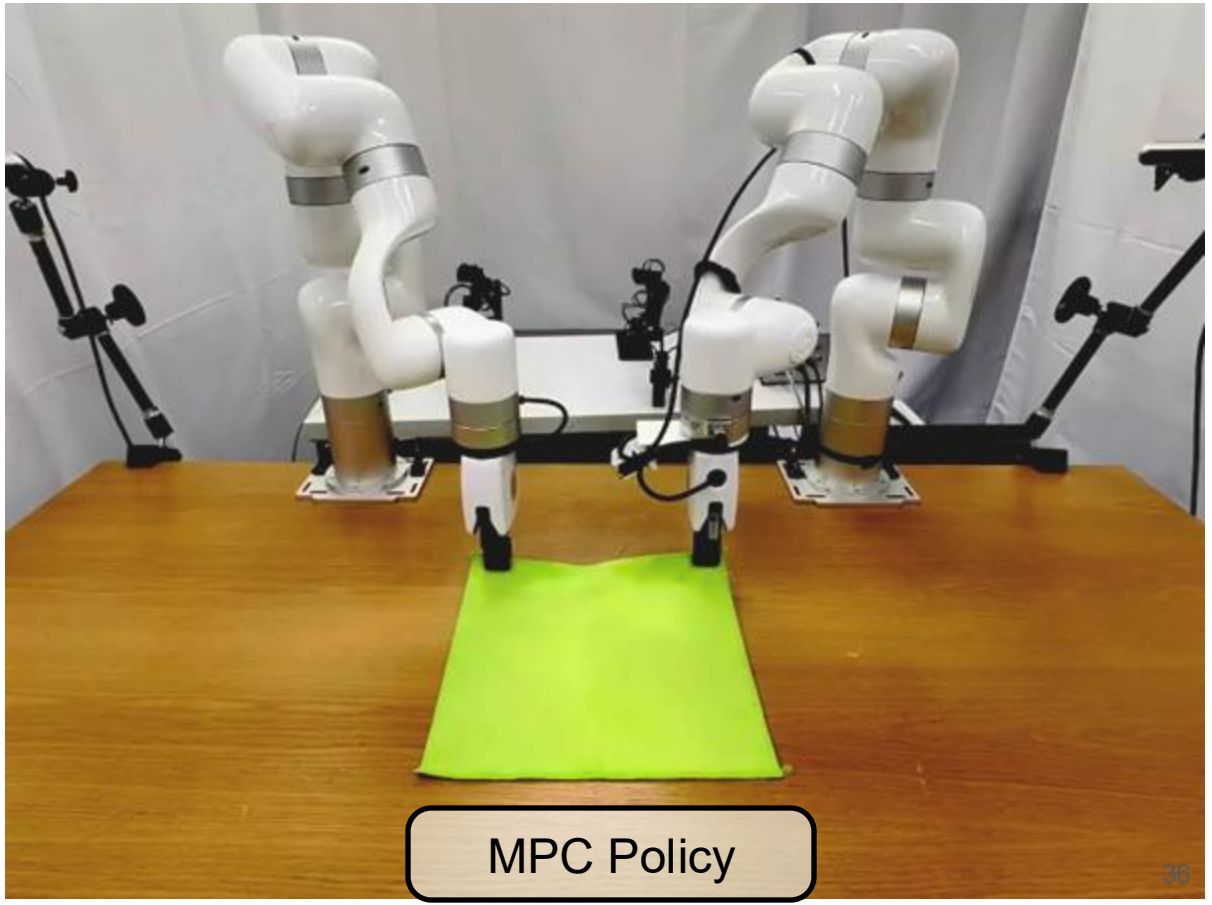

Left Hand
Right Hand

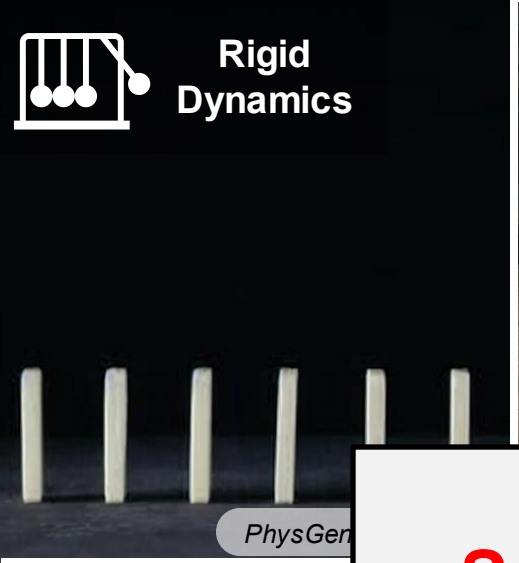

Left Hand


Left Hand
Right Hand

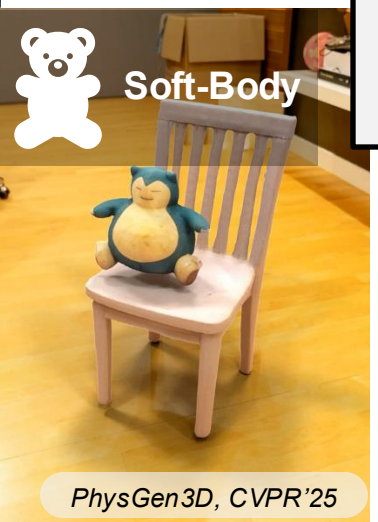

Left Hand
Right Hand

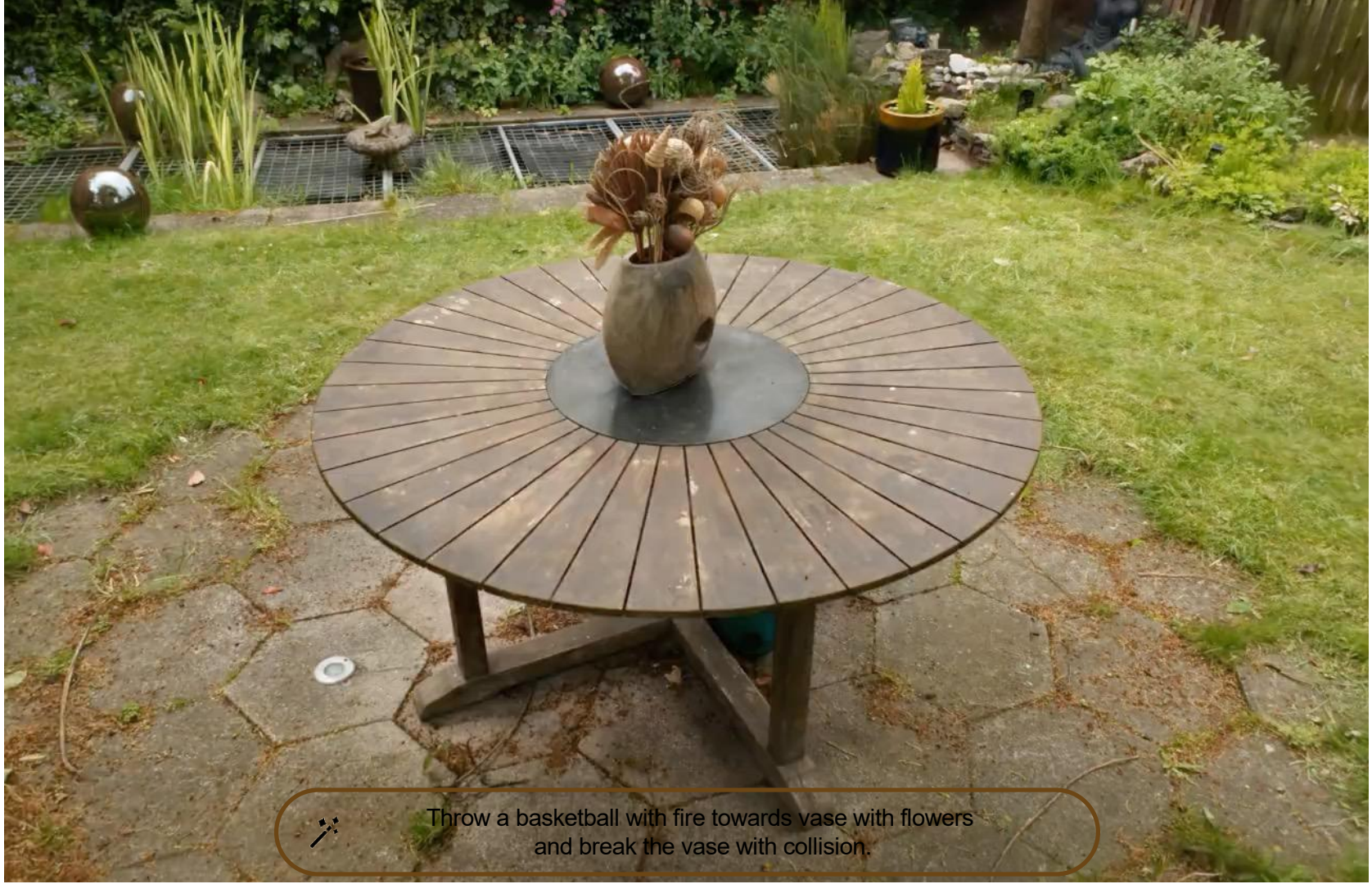

Target state


MPC Policy



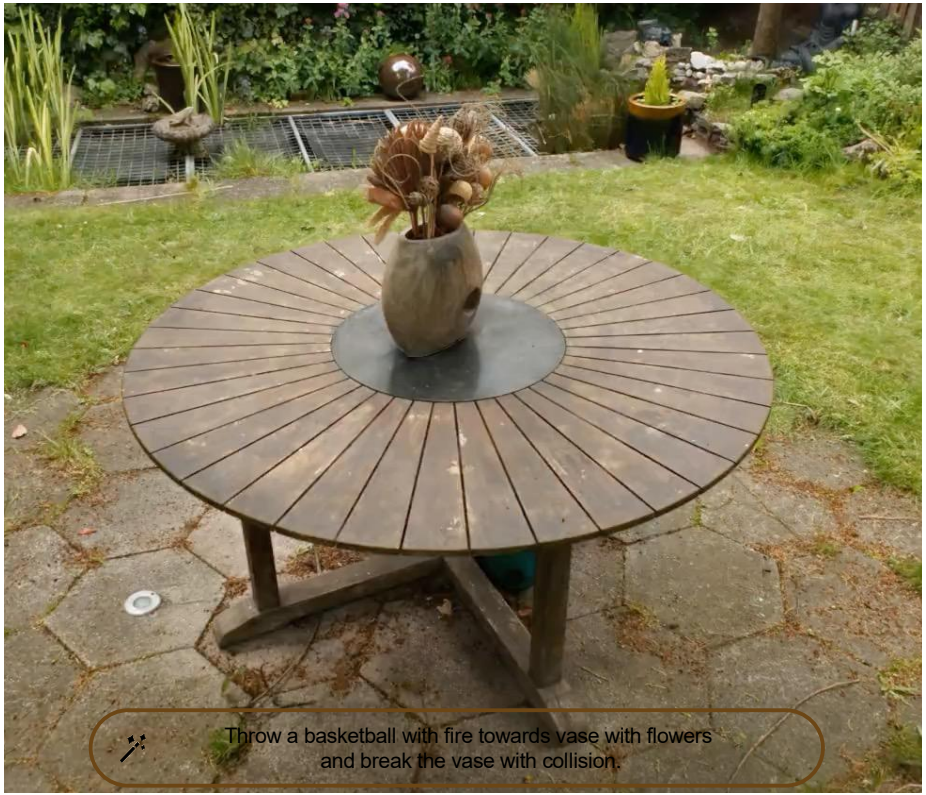
Specialist interactive videos

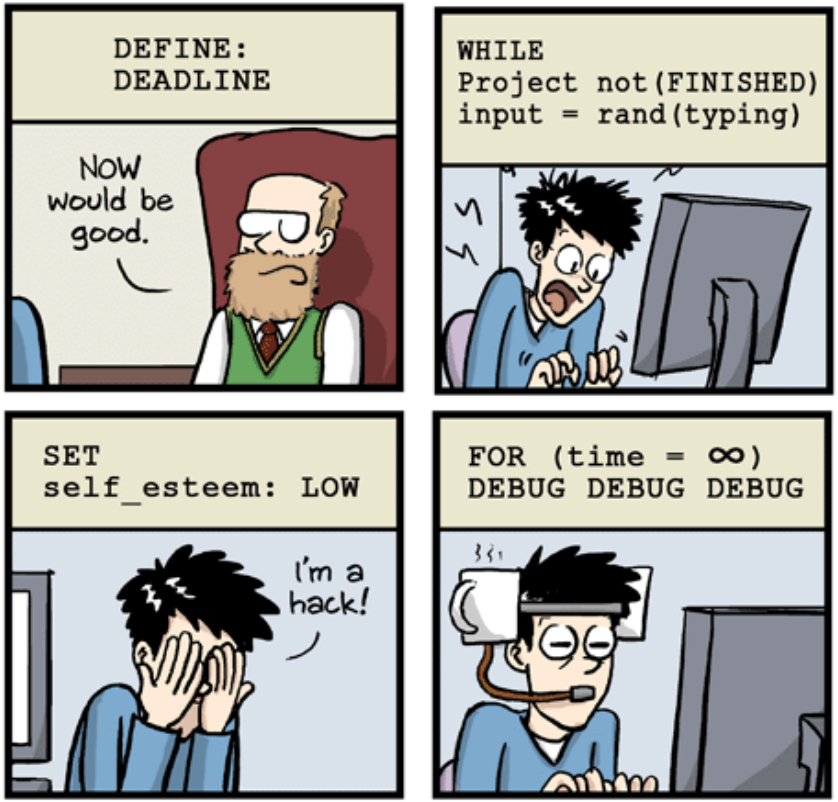




Throw a basketball with fire towards vase with flowers
and break the vase with collision.

Someone needs to do heavy lifting!

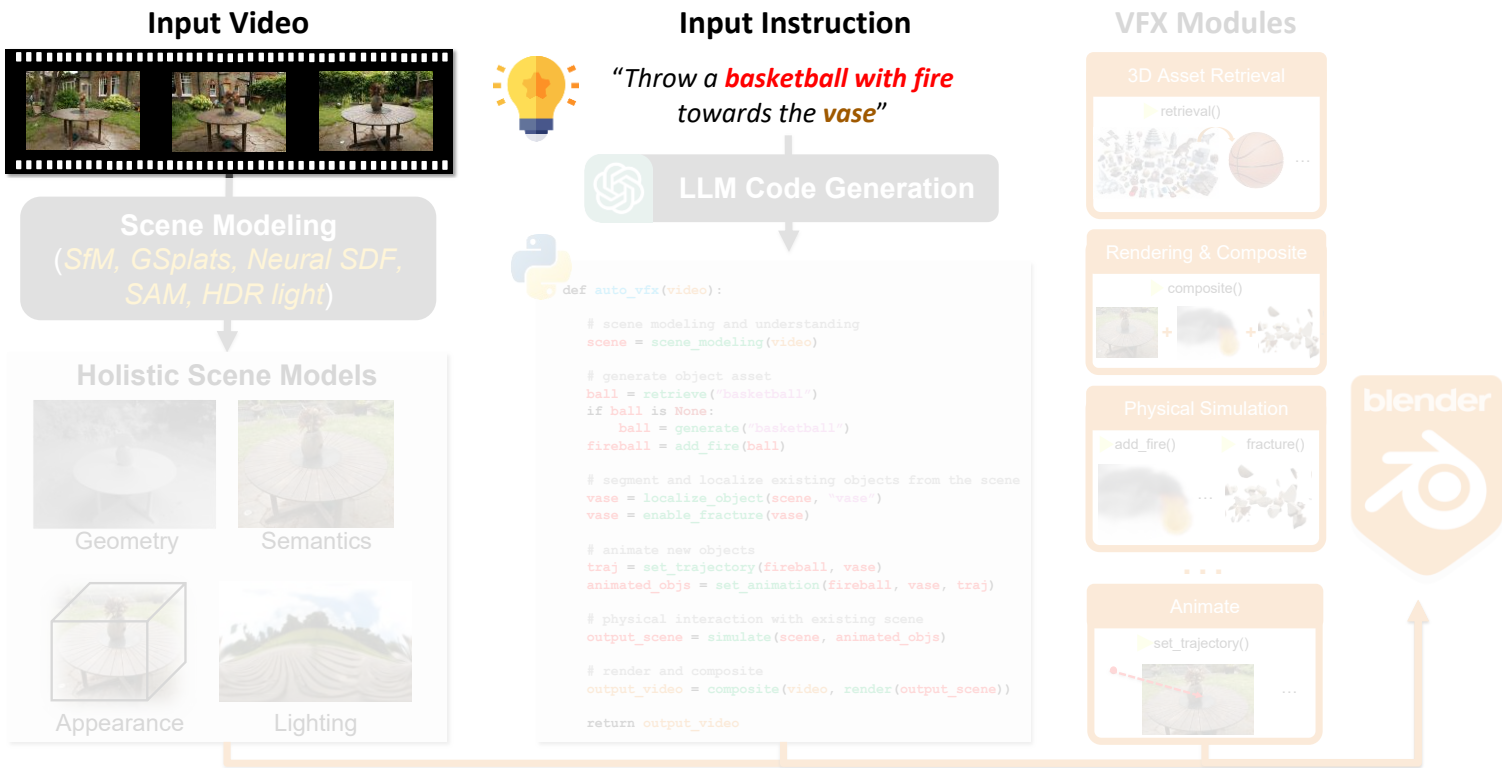




PhD Comics

AutoVFX: Generalist Interactive Video

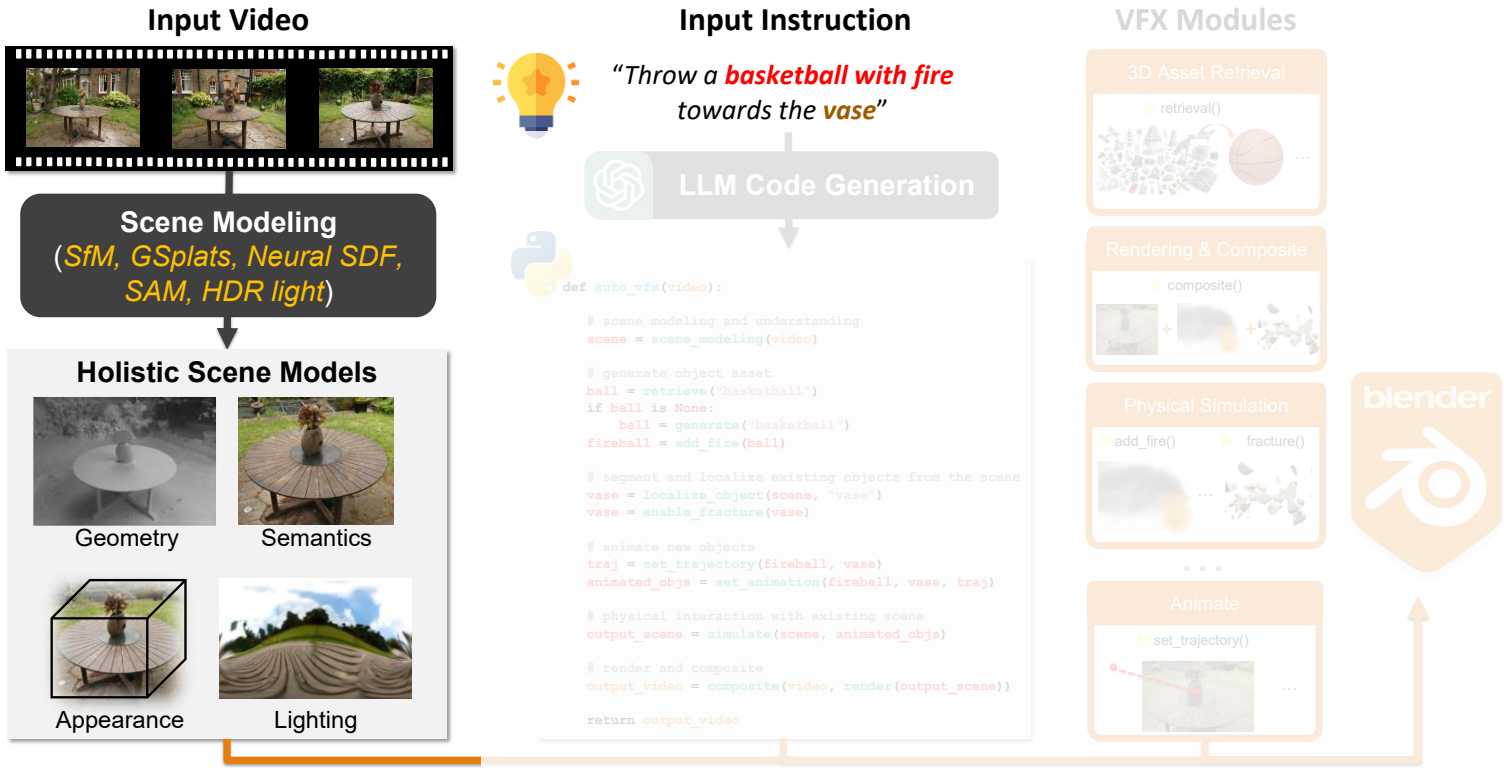
Hao-Yu Hsu



AutoVFX: Physically Realistic Video Editing from Natural Language Instructions, ***3DV 2025***

AutoVFX: Generalist Interactive Video

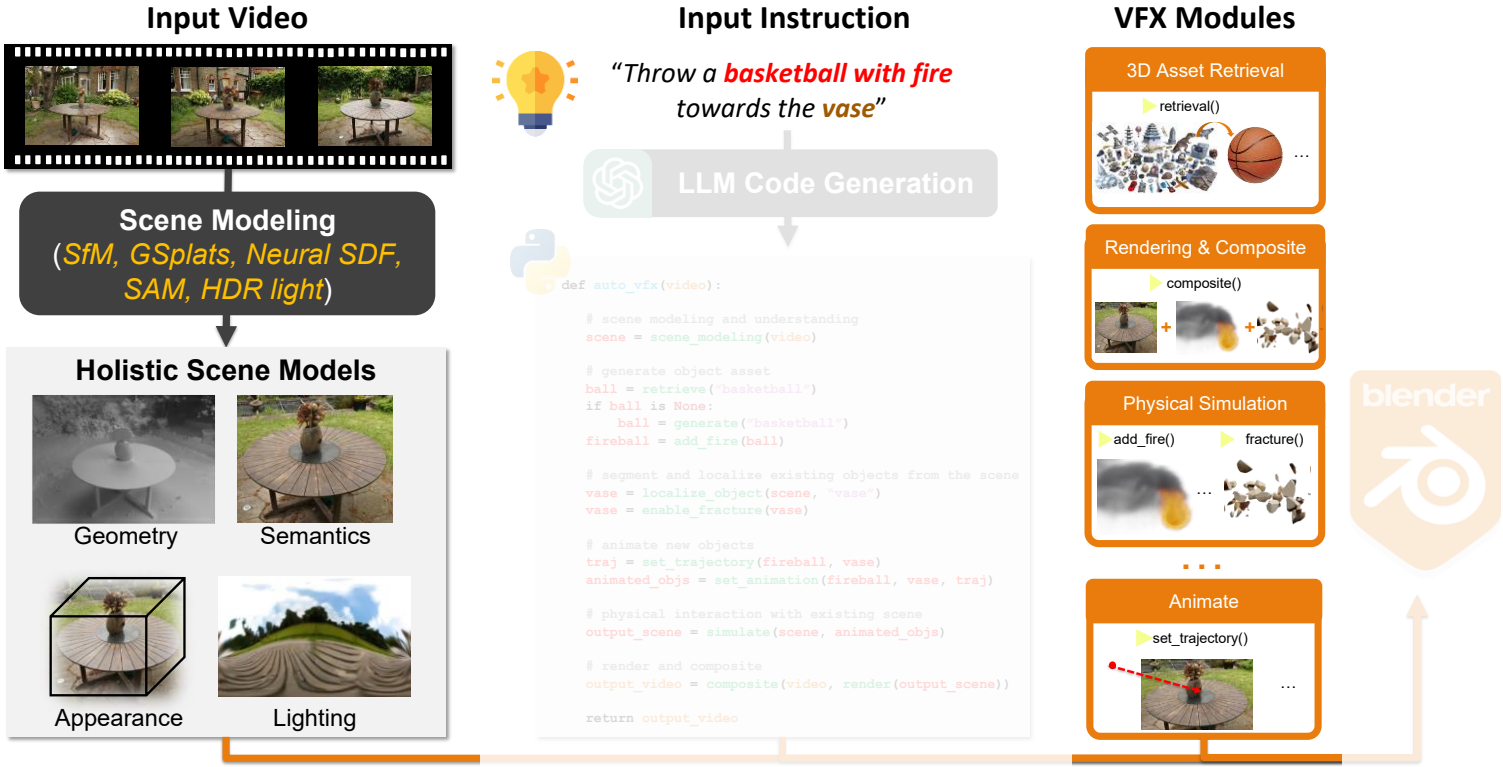
Hao-Yu Hsu



AutoVFX: Physically Realistic Video Editing from Natural Language Instructions, 3DV 2025

AutoVFX: Generalist Interactive Video

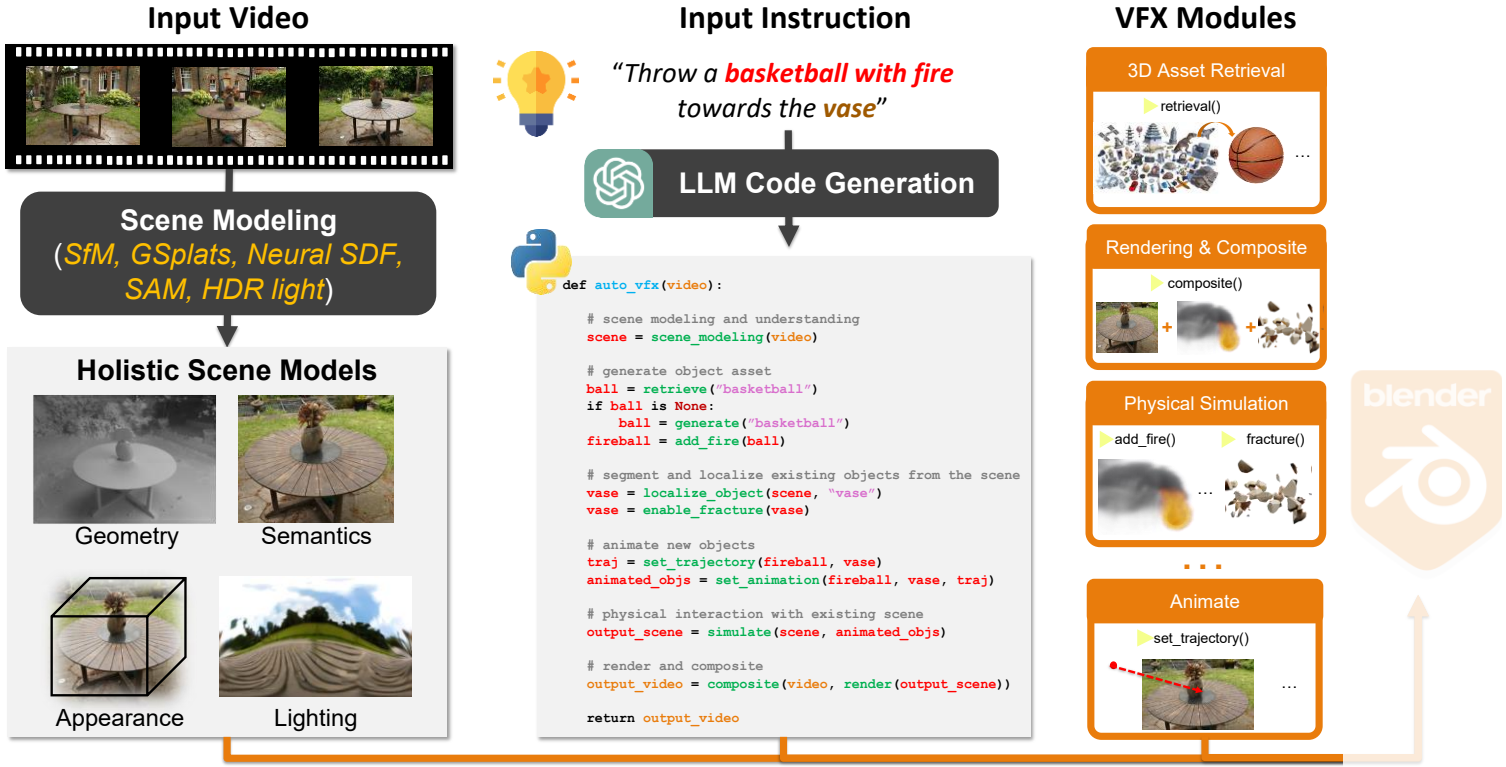
Hao-Yu Hsu



AutoVFX: Physically Realistic Video Editing from Natural Language Instructions, 3DV 2025

AutoVFX: Let the LLM agent code for us

Hao-Yu Hsu



AutoVFX: Physically Realistic Video Editing from Natural Language Instructions, 3DV 2025

AutoVFX: Let the LLM agent code for us

Hao-Yu Hsu

Input Video

Scene Modeling
(*SfM, GSplats, Neural SDF, SAM, HDR light*)

Holistic Scene Models

Geometry

Semantics

Appearance

Lighting

Input Instruction

*"Throw a **basketball with fire** towards the **vase**"*

LLM Code Generation

```python
def auto_vfx(video):

    # scene modeling and understanding
    scene = scene_modeling(video)

    # generate object asset
    ball = retrieve("basketball")
    if ball is None:
        ball = generate("basketball")
    fireball = add_fire(ball)

    # segment and localize existing objects from the scene
    vase = localize_object(scene, "vase")
    vase = enable_fracture(vase)

    # animate new objects
    traj = set_trajectory(fireball, vase)
    animated_objs = set_animation(fireball, vase, traj)

    # physical interaction with existing scene
    output_scene = simulate(scene, animated_objs)

    # render and composite
    output_video = composite(video, render(output_scene))

    return output_video
```

VFX Modules

3D Asset Retrieval

retrieval()

Rendering & Composite

composite()

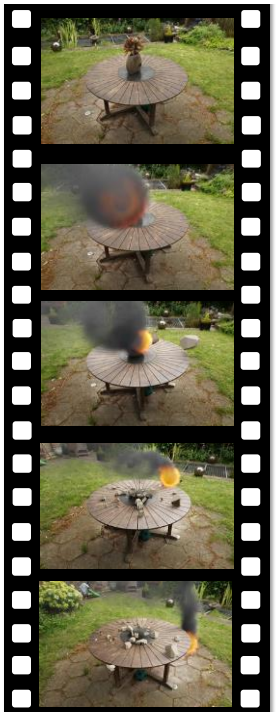
Physical Simulation

add_fire() fracture()

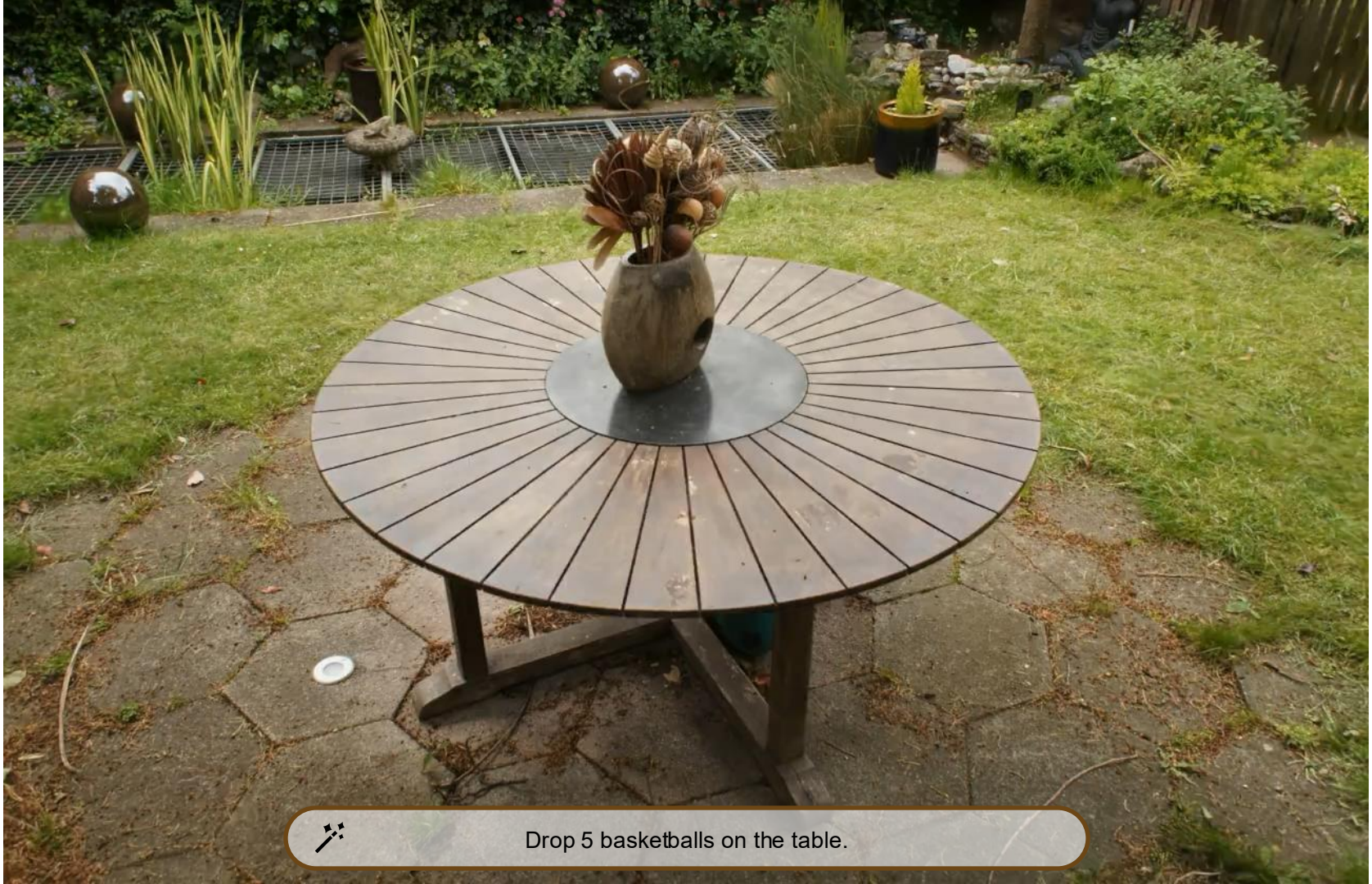
Animate

set_trajectory()

blender



AutoVFX: Physically Realistic Video Editing from Natural Language Instructions, 3DV 2025


Drop 5 basketballs on the table.


Make the vase with flowers to be like a mirror.


Insert an animated Pikachu on the table.


Put a Tony Stark on the floor covered with smoke.


Drop four barrels onto the floor: one mirror-like, one with fabric textures,
one resembling pavement, and one unchanged.


Insert a physics-enabled Benz G 20 meters in front of us with random 2D rotation. Add a Ferrari moving forward.

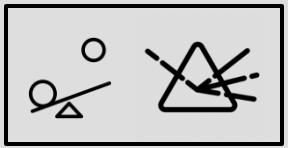

Drop a Tesla cybertruck with fire randomly in front of our vehicles from 3 meters high.

Perceive, Simulate and Render

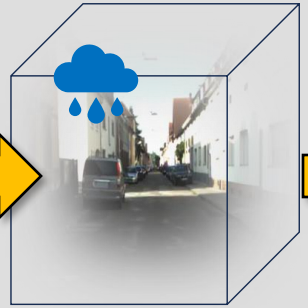
Input Image

User interaction

Scene Models
(*Geometry, Semantics, Appearance, Physics…*)



Physical Simulation
(*Lighting, Dynamics…*)



Updated Scene

Output Images

Perceive *Simulate* *Render*

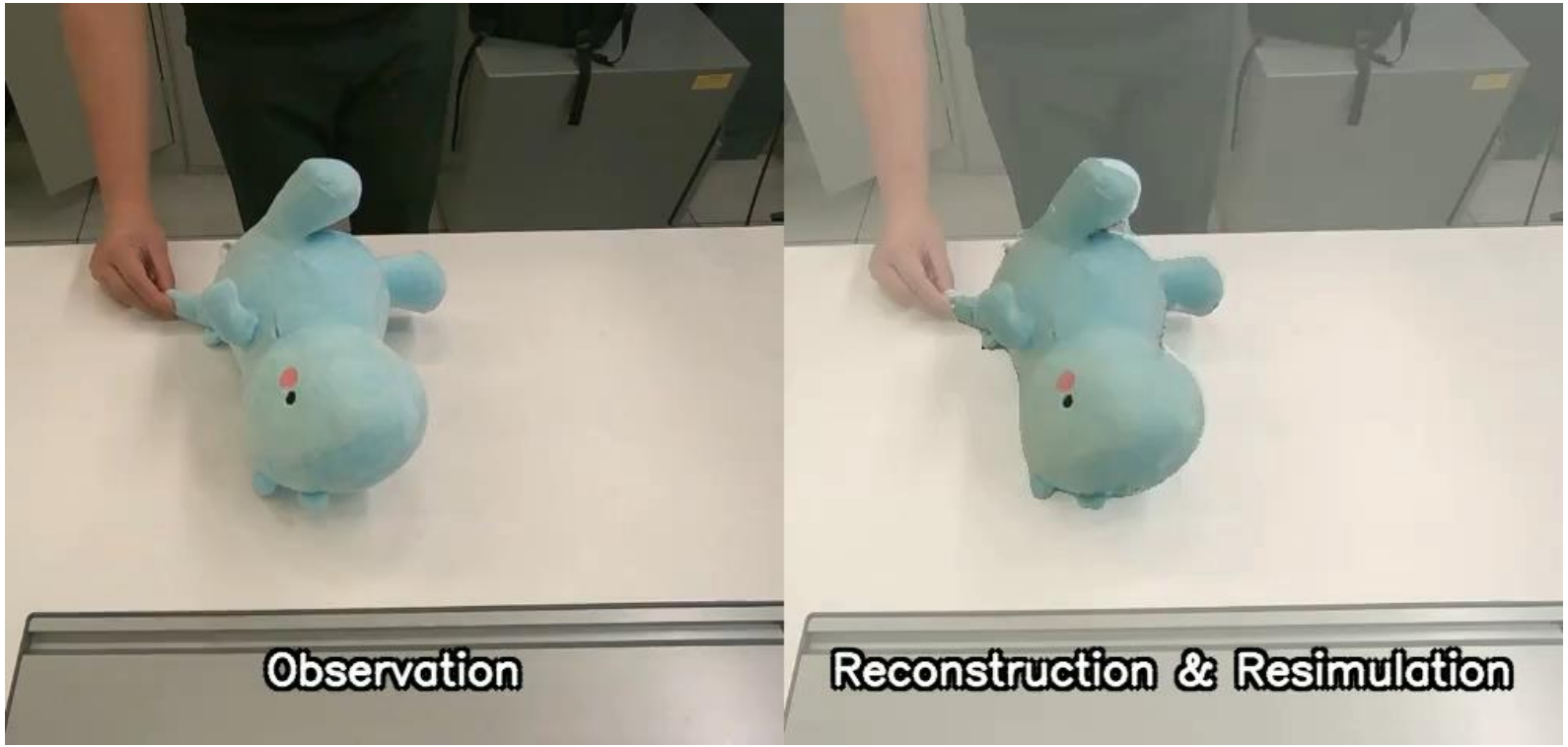
Projects summary

- PhysGen (ECCV 2024): https://stevenlsw.github.io/physgen/
- Video2Game (CVPR 2024): https://video2game.github.io/
- ClimateNeRF (ICCV 2023): https://climatenerf.github.io/
- AutoVFX (3DV 2025): https://haoyuhsu.github.io/autovfx-website/
- UrbanIR (3DV 2025): https://urbaninverserendering.github.io/
- PhysTwin (arXiv): https://jianghanxiao.github.io/phystwin-web/
- PhysGen3D (CVPR 2025, ***ExHall D Poster #71 4pm-6pm Fri Jun 13***): https://by-luckk.github.io/PhysGen3D/
- DRAWER (CVPR 2025, ***ExHall D Poster #68, 10:30-12:30 Sun 15 Jun***): https://xiahongchi.github.io/DRAWER/
- IRIS (CVPR 2025, ***ExHall D Poster #28 Fri 13 Jun 10:30am-12:30pm***): https://irisldr.github.io

All our projects have open-sourced code.

Lemon-picking

Lemon-picking

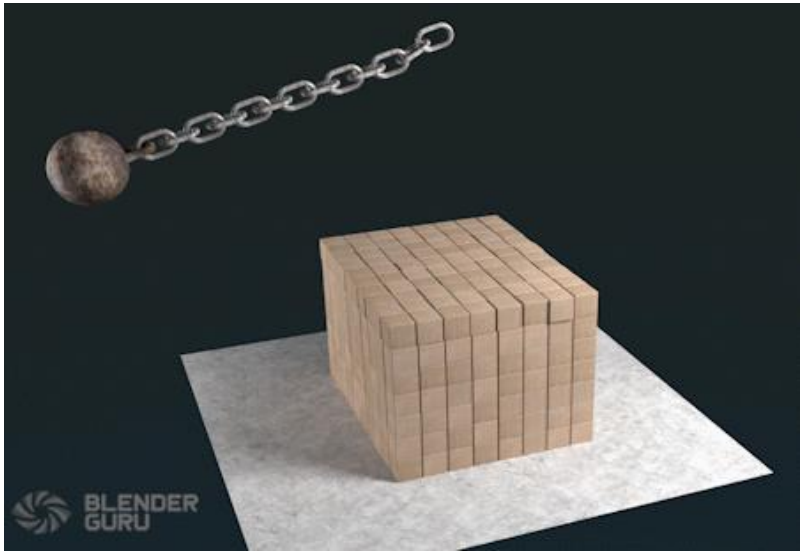


The Two Cultures

"*World Models*"
Data-driven, Generation, Implicit

Sora

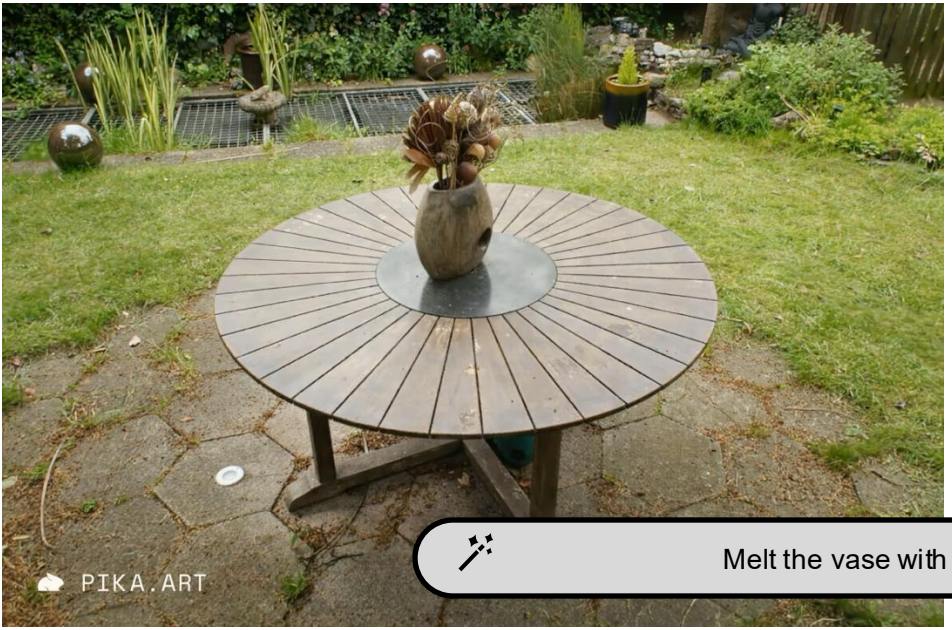
"*Digital Twins*"
Model-based, Simulation, Explicit



Blender

The Two Cultures

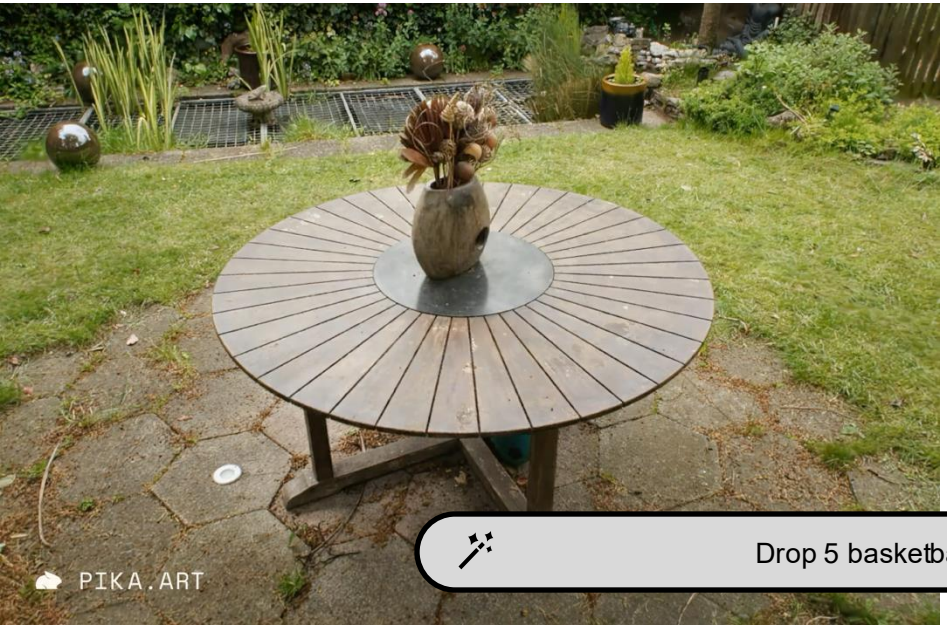
Pika 1.5 Effects



AutoVFX

The Two Cultures

Pika 1.5



AutoVFX

Rain Synthesis

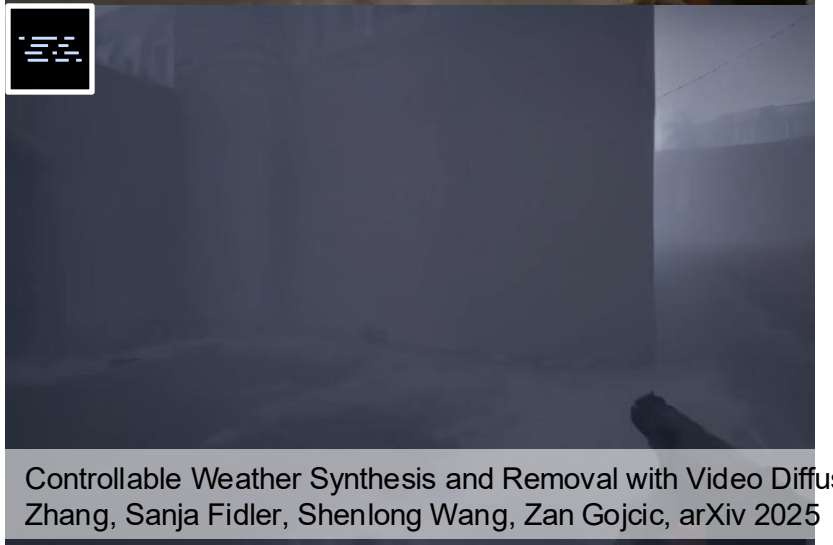
Fog Synthesis

Snow Synthesis

Controllable Weather Synthesis and Removal with Video Diffusion Models, Chih-Hao Lin, Zian Wang, Ruofan Liang, Yuxuan Zhang, Sanja Fidler, Shenlong Wang, Zan Gojcic, arXiv 2025

Rain Synthesis

Fog Synthesis

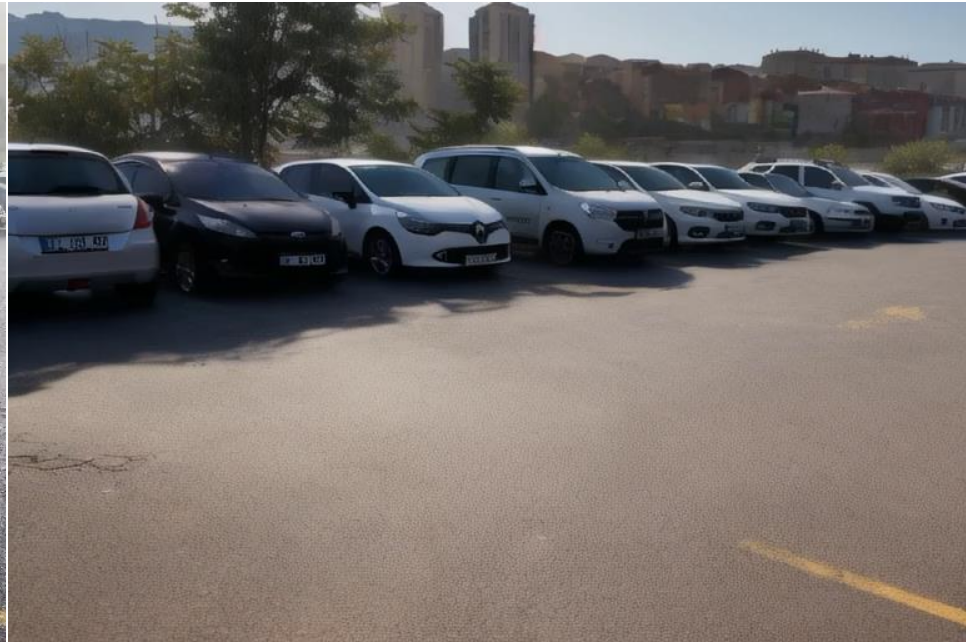


Snow Synthesis

Controllable Weather Synthesis and Removal with Video Diffusion Models, Chih-Hao Lin, Zian Wang, Ruofan Liang, Yuxuan Zhang, Sanja Fidler, Shenlong Wang, Zan Gojcic, arXiv 2025

Input Video

Weather Removal



WEATHER REMOVAL MODEL

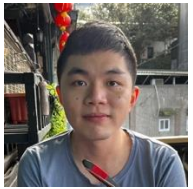
Input Video

Weather Removal

WEATHER REMOVAL MODEL

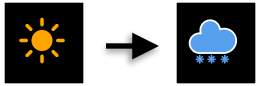
Controllable Weather Synthesis and Removal with Video Diffusion Models, Chih-Hao Lin, Zian Wang, Ruofan Liang, Yuxuan Zhang, Sanja Fidler, Shenlong Wang, Zan Gojcic, arXiv 2025

The Bitter Lessons?



Chih-Hao Lin

Input Video



WeatherWeaver (generative model)

ClimateNeRF (physical sim)

WeatherWeaver (arXiv): https://research.nvidia.com/labs/toronto-ai/WeatherWeaver
ClimateNeRF (ICCV23): https://climatenerf.github.io

Projects summary

- AutoVFX (3DV 2025): https://haoyuhsu.github.io/autovfx-website/
- UrbanIR (3DV 2025): https://urbaninverserendering.github.io/
- PhysGen (ECCV 2024): https://stevenlsw.github.io/physgen/
- PhysGen3D (CVPR 2025): https://by-luckk.github.io/PhysGen3D/
- Video2Game (CVPR 2024): https://video2game.github.io/
- DRAWER (CVPR 2025): https://xiahongchi.github.io/DRAWER/
- ClimateNeRF (ICCV 2023): https://climatenerf.github.io/
- PhysTwin (arXiv): https://jianghanxiao.github.io/phystwin-web/

All our projects have open-sourced code.

Open Questions

- ***Synergizing modeling and interaction***: Model the physical world to enable physical interaction ←→ interact with the physical world to better model the world.
- ***Need for perfection***: Physical simulation expects nearly error-free scene understanding – what can we do when we haven't yet solved computer vision?
- ***The two cultures***: How to better harness the best of model-based simulation and data-driven generation.

Open Questions

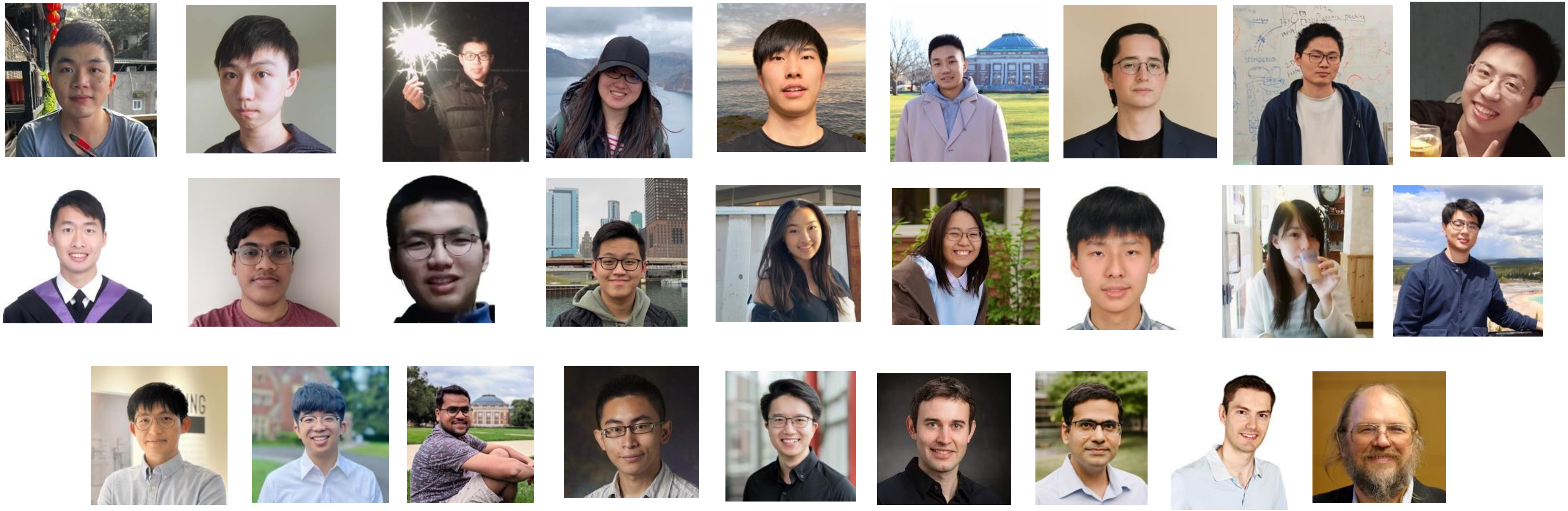
- ***Need for perfection***: Physical simulation expects nearly error-free scene understanding – what can we do when we haven't yet solved computer vision?
- ***The two cultures***: How to better harness the best of model-based simulation and data-driven generation.
- ***Physical understanding for physical AI***: Model the physical world to enable interaction ←→ interact with the physical world to better model the world.
- ***Broaden impact*** beyond content creation, gaming, and robotics.

What is a good inductive bias for content creation?

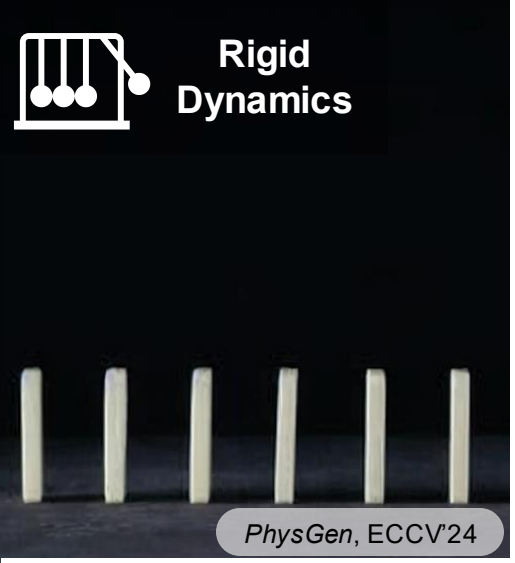
*in the scaling-law era

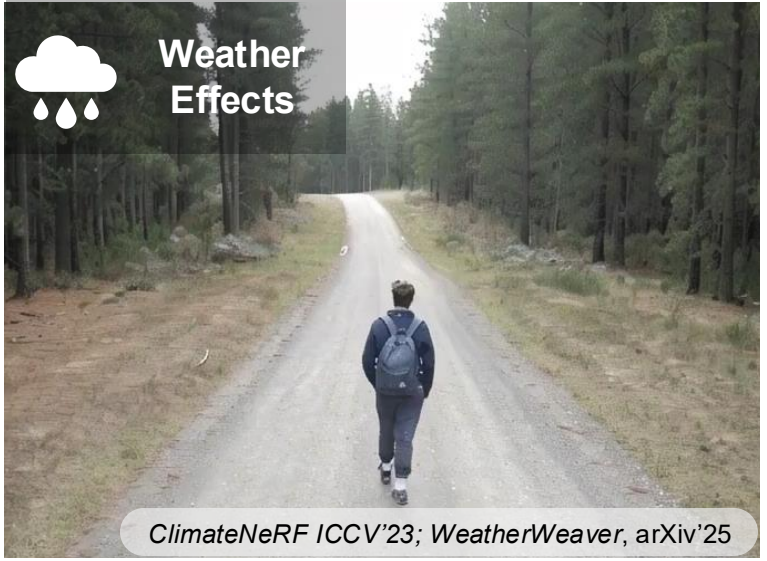
Universal Laws Newtonian physics, Conservation laws	**Statistical Priors** Laplacian smooth, As-rigid-as-possible Reg, Bilateral smooth
Hard Constraints Collision-free constraint, Kinematic Chain, Watertightness	**Soft Regularizers** Penetration loss
Easy to impose, costly to learn Rigid-body dynamics	**Hard to impose** Hair dynamics, Physical-based lighting
Essential for Downstream Grounded physics for robotic-manipulation simulators	**Optional for Downstream** Grounded physics for creating TikTok videos

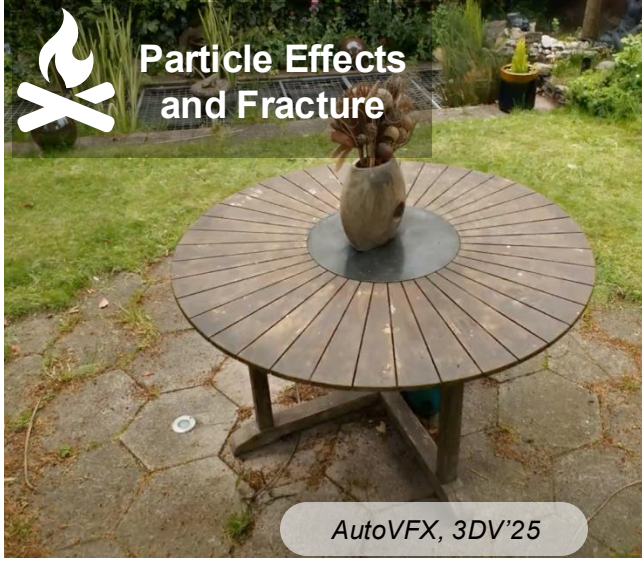
Acknowledgement

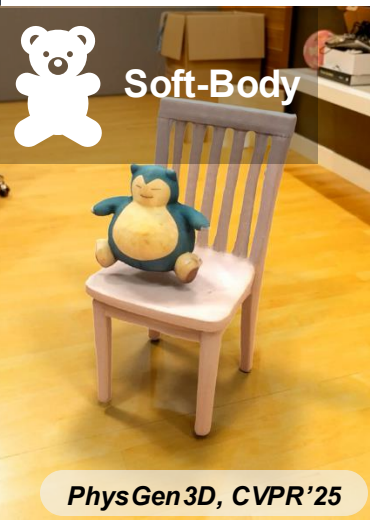


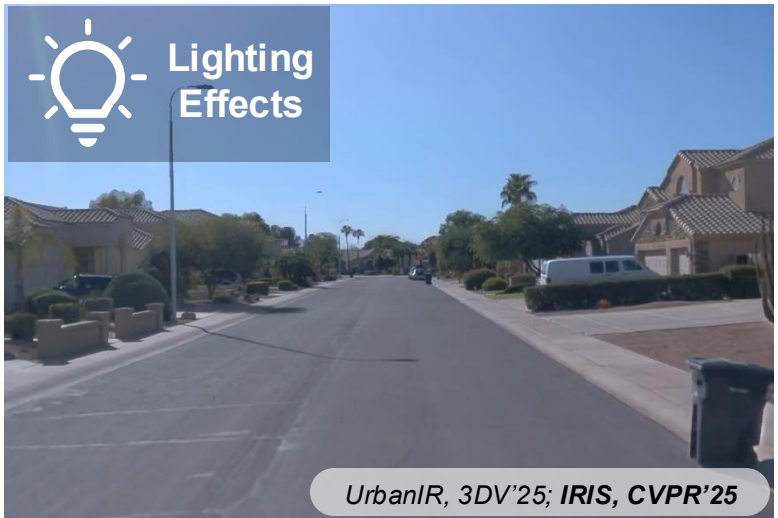
Acknowledgement


Rigid Dynamics
PhysGen, ECCV'24


Weather Effects
ClimateNeRF ICCV'23; WeatherWeaver, arXiv'25


Particle Effects and Fracture
AutoVFX, 3DV'25


Soft-Body
PhysGen3D, CVPR'25


Lighting Effects
UrbanIR, 3DV'25; IRIS, CVPR'25

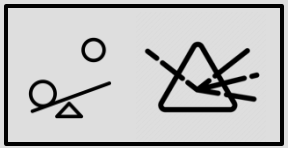

Articulation
DRAWER, CVPR'25

Perceive, Simulate and Render

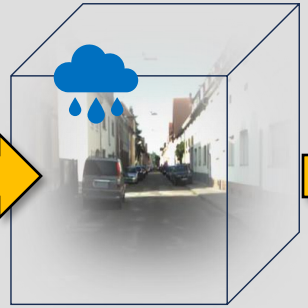
Input Image

User interaction

Scene Models
(*Geometry, Semantics, Appearance, Physics…*)



Physical Simulation
(*Lighting, Dynamics…*)



Updated Scene

Output Images

Perceive ***Simulate*** ***Render***