CVPR JUNE 3-7, 2026 DENVER COLORADO

Enable Explicit 3D/4D Controls for Pre-trained Generative Models

Yaoyao Liu
University of Illinois Urbana-Champaign

CVPR 2026 Workshop on Long-form Video Understanding, Generation, and Action
June 4, 2026

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Humans recognize objects by understanding their 3D structure

AI models:

Input image

Prediction
e.g., vehicle

Humans:

Input image

3D structure

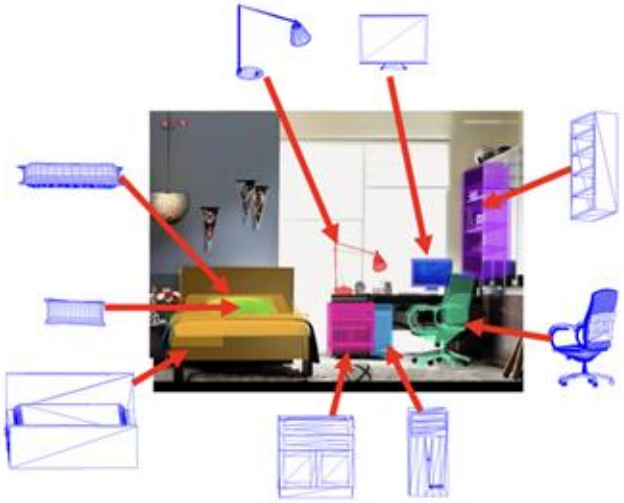
Prediction
e.g., vehicle

Advantages: more robust to occlusions, out-of-distribution cases, etc.

Can we learn AI models with the knowledge of the 3D world?

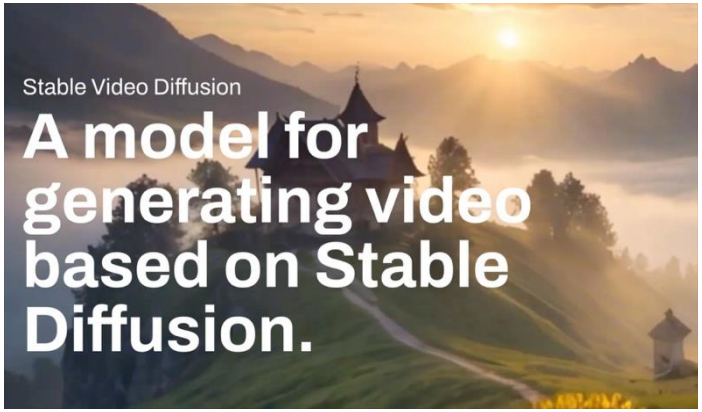
Major challenge: we don't have a large-scale dataset with 3D annotations

Dataset	# categories	# instances
PASCAL3D+ (2014) [13]	12	12,000
ObjectNet3D (2016) [24]	100	57,000
CAMERA25 (2019) [27]	6	1,000
REAL275 (2019) [27]	6	24
Objectron (2021) [28]	9	18,000
Wild6D (2022) [14]	5	2,000



Images from: https://cvgl.stanford.edu/projects/objectnet3d/

We now have powerful generation models, but they still lack explicit 3D/4D control.



Can we use pre-trained generative models to create data with 3D/4D annotations?

Image and video from: https://stability.ai/

Video generation with 3D (4D) controls

SIGGRAPH 2026 Conference Papers

go.illinois.edu/4dvideo

FreeOrbit4D: Training-Free Arbitrary Camera Redirection for Monocular Videos via Foreground-Complete 4D Reconstruction

WEI CAO, University of Illinois Urbana-Champaign, USA
HAO ZHANG, University of Illinois Urbana-Champaign, USA
FENGRUI TIAN, University of Pennsylvania, USA
YULUN WU, University of Illinois Urbana-Champaign, USA
YINGYING LI, University of Illinois Urbana-Champaign, USA
SHENLONG WANG, University of Illinois Urbana-Champaign, USA
NING YU, Eyeline Labs, USA and Netflix, USA
YAOYAO LIU*, University of Illinois Urbana-Champaign, USA

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Penn University of Pennsylvania

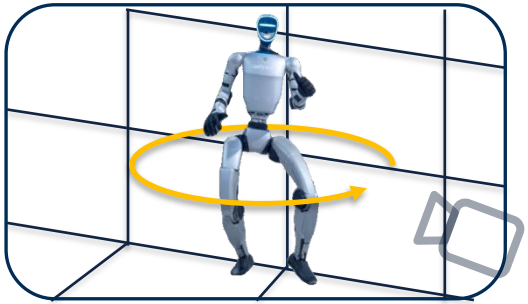
NETFLIX

Poster session today: 5:15 pm, Exhibit Hall A, No. 303

Cao, W., …, **Liu, Y**. *FreeOrbit4D: Training-free Arbitrary Camera Redirection for Monocular Videos via Foreground-Complete 4D Reconstruction.* SIGGRAPH 2026.

Can we enable 4D controls for video redirection?

Monocular Video



Bullet Time

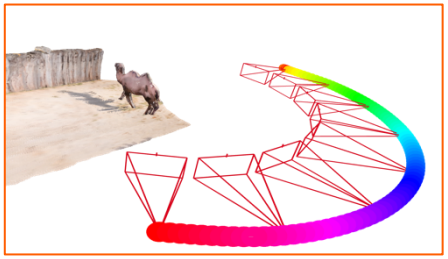
?

Existing methods: large-angle redirection remains challenging

go.illinois.edu/4dvideo

Input Video

Target Trajectory



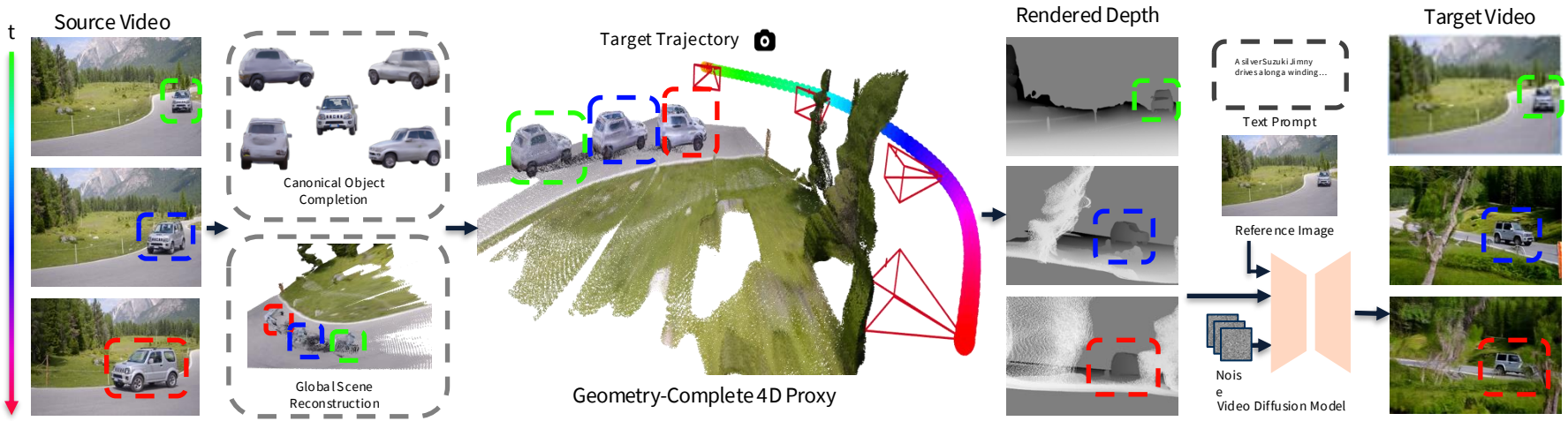
Explicit Warping
e.g., using depth maps

Implicit Control
e.g., text prompts

Our pipeline: geometry-complete 4D proxy for synthesis

go.illinois.edu/4dvideo

Training-Free



Decoupled Reconstruction

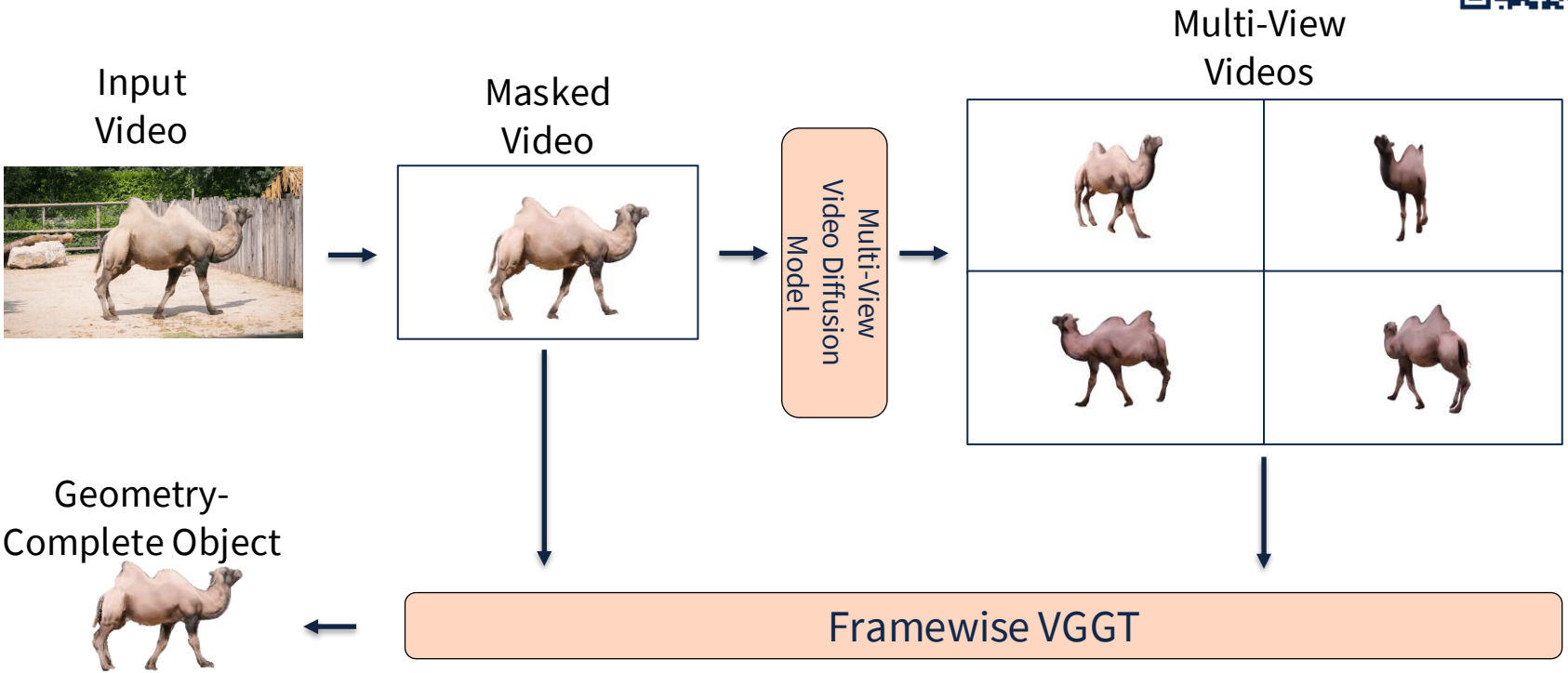
4D Proxy Alignment

Geometry-Guided Synthesis

Cao, W., …, **Liu, Y**. *FreeOrbit4D: Training-free Arbitrary Camera Redirection for Monocular Videos via Foreground-Complete 4D Reconstruction*. SIGGRAPH 2026.

go.illinois.edu/4dvideo

Canonical object completion: from masked monocular video to geometry-complete object



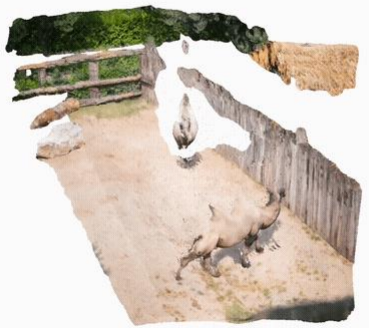
go.illinois.edu/4dvideo

Global scene reconstruction: recover background and visible foreground in world space

Input Video

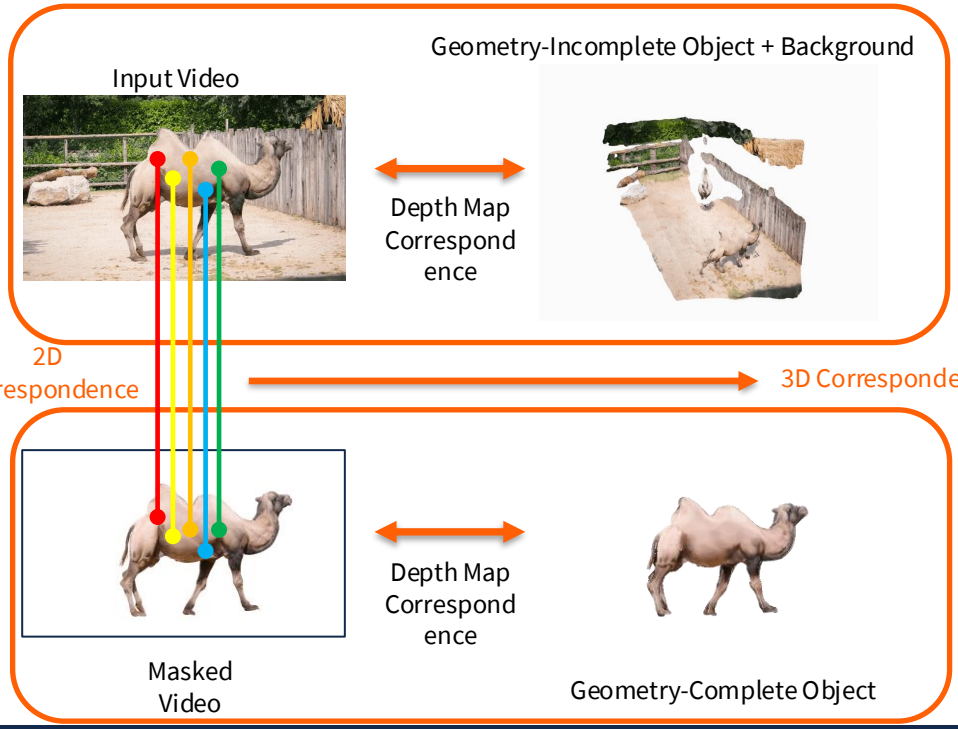
Dynamic-aware Feed-Forward Network

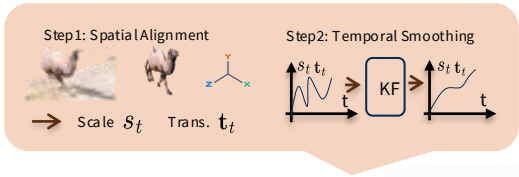
Geometry-Incomplete Object + Background

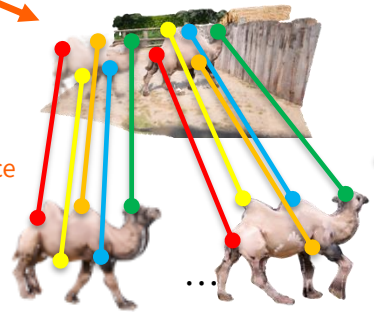


go.illinois.edu/4dvideo

3D correspondence-aware alignment:
complete geometry + correct scene placement → unified 4D proxy





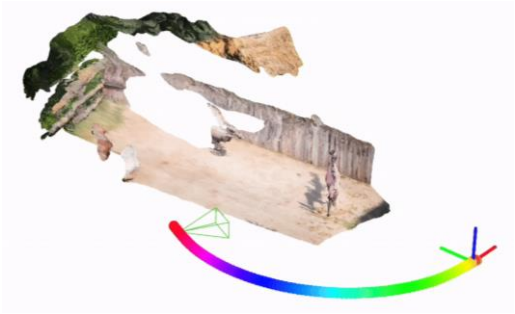


Geometry-Complete 4D Proxy

go.illinois.edu/4dvideo

Geometry-guided video synthesis:
rendered depth from the 4D proxy guides target-view generation

Geometry-conditioned Video Synthesis



Target Trajectory

Rendered Depth

Reference Image

Text Prompt

A light brown dromedary camel stands in a sunlit enclosure…

Noise

Video Diffusion Model

Target Video

Fast motion and thin structures

Input Video

ReCamMaster

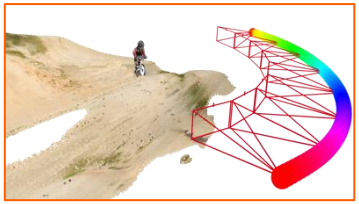
TrajectoryCrafter

Ours

EX4D

Gen3C

Large camera displacement



Input Video

ReCamMaster

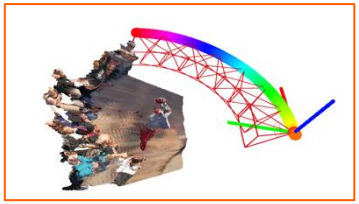
TrajectoryCrafter

Ours

EX4D

Gen3C

Complex human motion



Input Video

ReCamMaster

TrajectoryCrafter

Ours

EX4D

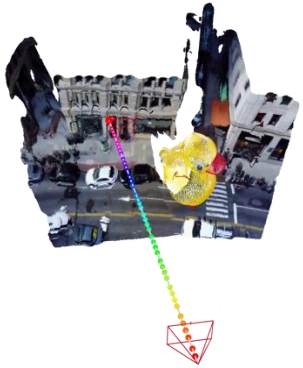
Gen3C

Flexible camera control

go.illinois.edu/4dvideo

Input Video

Target Trajectory



Target Video

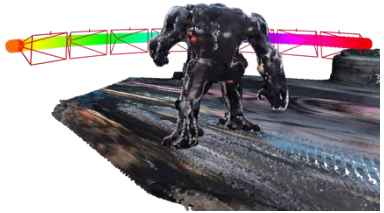
Flexible camera control

go.illinois.edu/4dvideo

Input Video

Target Trajectory

Target Video



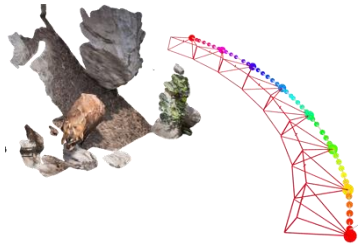
Flexible camera control

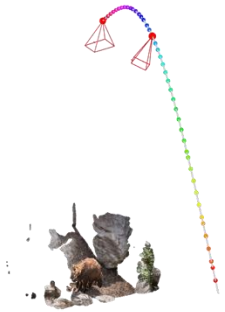
go.illinois.edu/4dvideo

Input Video

Target Trajectory

Target Video

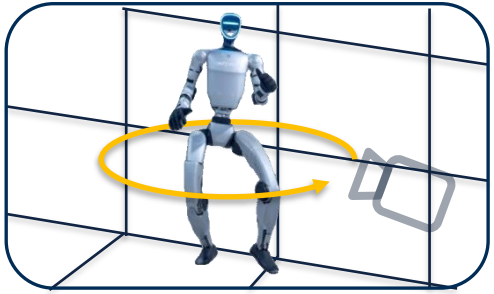




Back to our motivating example

go.illinois.edu/4dvideo

Monocular Video



Bullet Time

Target Video

Consistency across random seeds

go.illinois.edu/4dvideo

Input Video

Target Video

Seed #1

Seed #2

Appearance propagation from a single edited frame

Input Video

Reference Image

Target Video

Direct geometry editing in explicit 4D space

go.illinois.edu/4dvideo

Reference Image

Target Video

Input Video

Quantitative results

go.illinois.edu/4dvideo

Table 1. **Quantitative comparison and User Study.** We report VBench for perceptual video quality, DINO/CLIP-SIM for semantic similarity, FID-V/FVD-V for distributional fidelity, and user ratings (1–5 scale). **Bold**: best; underline: second-best.

Method	VBench ↑						Similarity & Fidelity				User Study		
	Subject Consis.	BG Consis.	Motion Smooth.	Overall Consis.	Aesth. Qual.	Imaging Qual.	DINO-SIM (↑)	CLIP-SIM (↑)	FID-V (↓ $\times 10^2$)	FVD-V (↓ $\times 10^3$)	Overall (↑)	Motion (↑)	Stab. (↑)
ReCamMaster	0.84	0.92	**0.98**	0.16	0.39	43	0.37	0.75	2.6	3.9	2.0	2.5	2.0
TrajectoryCrafter	0.80	0.91	0.94	0.19	0.47	53	0.47	0.79	2.0	3.6	2.8	3.2	2.9
EX-4D	0.76	0.89	0.94	0.16	0.42	46	0.28	0.69	3.2	3.8	2.0	2.5	2.0
GEN3C	0.79	0.88	0.95	0.18	0.42	49	0.43	0.75	2.3	**3.3**	2.4	3.5	2.3
Ours	**0.88**	**0.94**	0.96	**0.24**	**0.52**	**64**	**0.65**	**0.84**	**1.7**	3.6	**4.6**	**4.5**	**4.5**

Best in 5/6 VBench metrics ¾ Similarity & Fidelity metrics and all user-study axes

go.illinois.edu/4dbenchmark

Extension: we can create a video dataset with 4D annotations

Redirect4D-Bench: A Scalable Benchmark for Camera Redirection of Monocular Dynamic Videos with Pseudo-4D Ground Truth

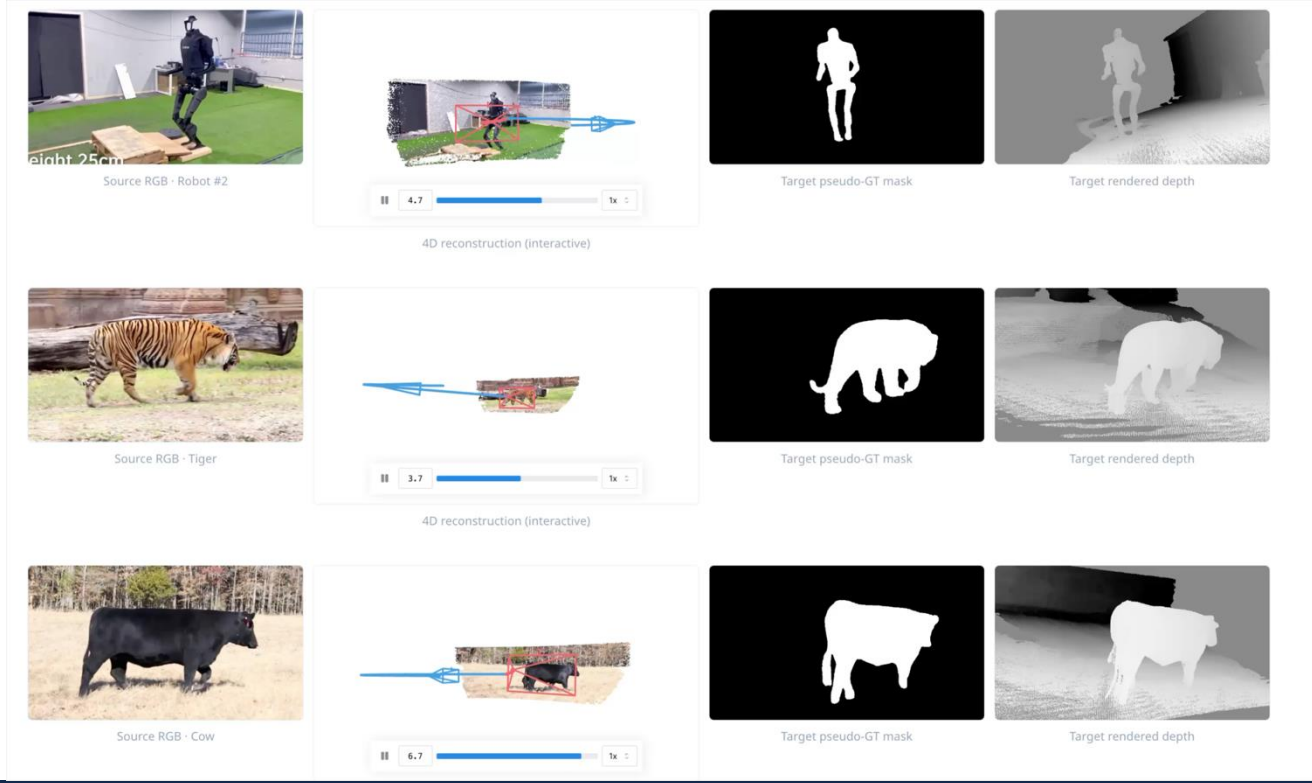
Wei Cao[1], Hao Zhang[1], Jiapeng Tang[2], Yulun Wu[1],
Yingying Li[1], Ning Yu[3], Shenlong Wang[1], Yaoyao Liu[1]

[1] University of Illinois Urbana-Champaign
[2] Technical University of Munich [3] Netflix

Cao, W., …, **Liu, Y**. *Redirect4D-Bench: A Scalable Benchmark for Camera Redirection of Monocular Dynamic Videos with Pseudo-4D Ground Truth.*

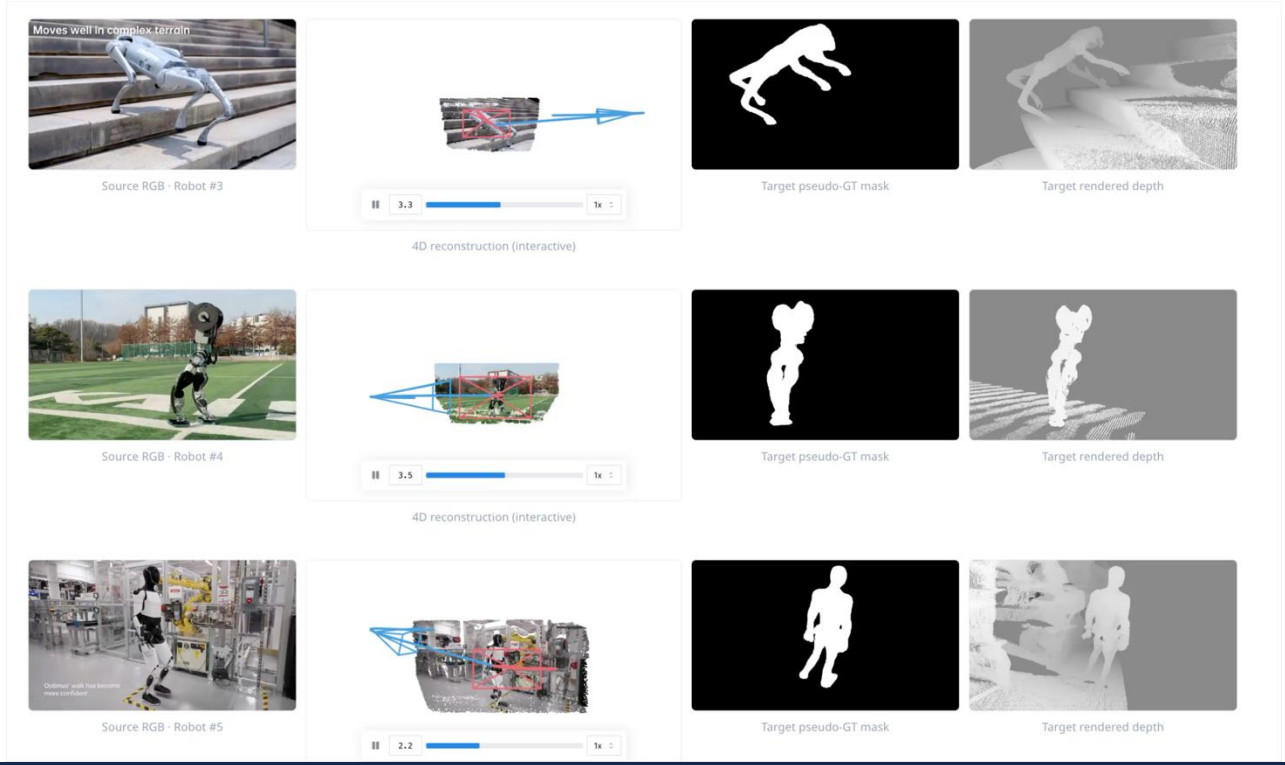
go.illinois.edu/4dbenchmark

Extension: we can create a video dataset with 4D annotations

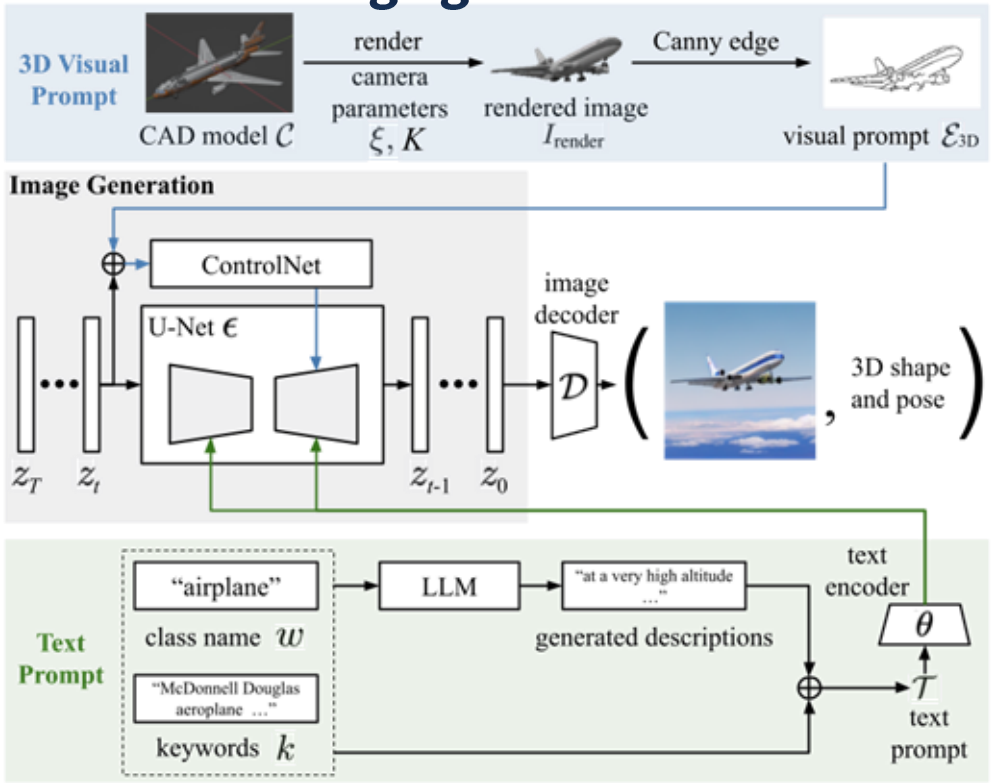


go.illinois.edu/4dbenchmark

Extension: we can create a video dataset with 4D annotations

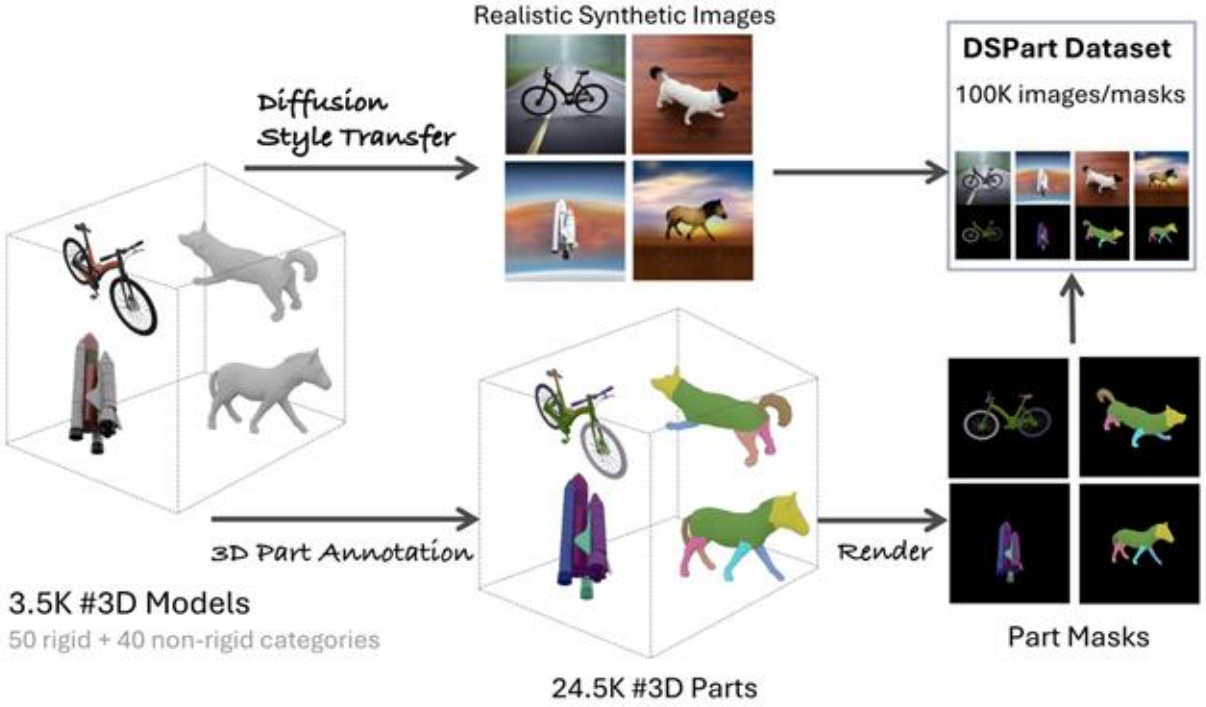


For image generation: 3D diffusion style transfer (3D-DST) enable 3D controls for image generation

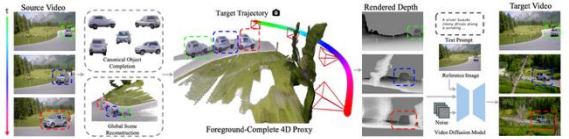


For image generation: 3D diffusion style transfer (3D-DST) enable 3D controls for image generation

Extension: generating 3D annotations based on articulated models

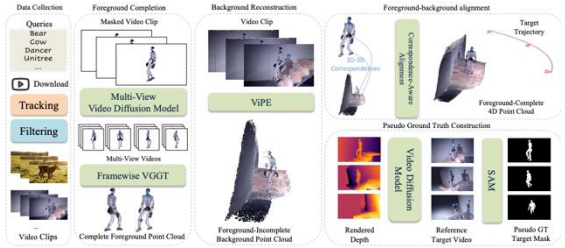


Enable 3D/4D controls for pre-trained generative models



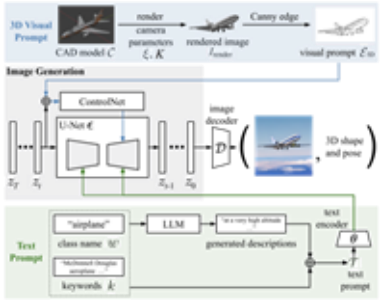
Videos with 4D controls (SIGGRAPH 2026 Conference)

go.illinois.edu/4dvideo



Videos with 4D annotations (Under review)

go.illinois.edu/4dbenchmark



Images with 3D controls (ICLR 2023 Spotlight)

go.illinois.edu/3dimage

- Cao, W., …, **Liu, Y**. *FreeOrbit4D: Training-free Arbitrary Camera Redirection for Monocular Videos via Foreground-Complete 4D Reconstruction*. SIGGRAPH 2026.
- Cao, W., …, **Liu, Y**. *Redirect4D-Bench: A Scalable Benchmark for Camera Redirection of Monocular Dynamic Videos with Pseudo-4D Ground Truth*. Under review.
- Ma, W., …, **Liu, Y.* (corresponding author)**, Yuille, A. *Generating Images with 3D Annotations Using Diffusion Models*. ICLR 2023.

Thanks!
Any questions?

lyy@illinois.edu
https://yaoyaoliu.web.illinois.edu/