# Reangle-A-Video: 4D Video Generation as Video-to-Video Translation

Hyeonho Jeong*, Suhyeon Lee*, Jong Chul Ye (* denotes equal contribution)

Project page: https://hyeonho99.github.io/reangle-a-video

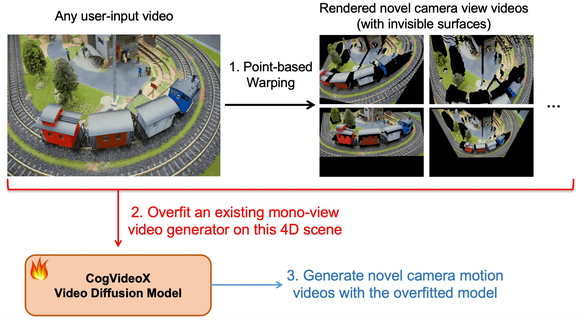
ICCV HONOLULU HAWAII OCT 19-23, 2025

KAIST Adobe

Project

## TL;DR

Reangle-A-Video achieves **video novel view synthesis** by applying NeRF logic to video diffusion models!




## Key Contributions

- A unified framework for generating synchronized multi-view videos from a single input video.
- Unlike concurrent works (RecamMaster, TrajectoryCrafter), we do not require training on large-scale 4D datasets.
- We frame multi-view video generation task as video-to-videos translation and decompose a dynamic 4D scene into:
  (1) View-specific appearance (starting image)
  (2) View-invariant motion (image-to-video generation)

## Quantitative Results

Quantitative comparisons using VBench, MEt3R, FID and FVD

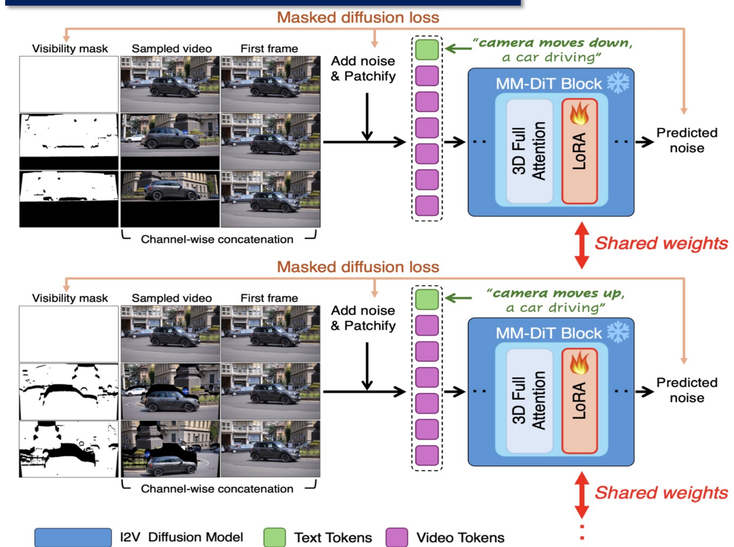
| (a) Static View Transport | Subject Consistency | Background Consistency | Temporal Flickering | Motion Smoothness | Dynamic Degree | Aesthetic Quality | Imaging Quality | MEt3R↓ | FID↓ | FVD↓ |
|---|---|---|---|---|---|---|---|---|---|---|
| Generative Camera Dolly [70] | 0.8849 | 0.9107 | 0.8731 | 0.9119 | 0.7608 | 0.4052 | 0.5524 | 0.1237 | 155.15 | 5264.7 |
| Vanilla CogVideoX [85] | 0.9448 | **0.9398** | 0.9738 | 0.9859 | 0.7286 | 0.4997 | **0.6475** | 0.0539 | 79.621 | 3664.2 |
| Reangle-A-Video (Ours) | **0.9516** | 0.9327 | **0.9764** | **0.9907** | **0.7657** | **0.5160** | 0.6414 | **0.0412** | **53.448** | **2690.9** |
| **(b) Dynamic Camera Control** | | | | | | | | | | |
| NVS-Solver [88] | 0.9037 | 0.9325 | 0.9049 | 0.9521 | 0.8809 | 0.5023 | **0.6411** | 0.1090 | 95.815 | 3516.5 |
| Trajectory Attention [79] | 0.8984 | 0.9288 | 0.9342 | 0.9658 | **0.8889** | 0.4854 | 0.5990 | 0.0965 | 109.20 | 3624.9 |
| Reangle-A-Video (Ours) | **0.9140** | **0.9364** | **0.9386** | **0.9794** | 0.8884 | **0.5238** | 0.6271 | **0.0648** | **74.194** | **3019.7** |

Quantitative evaluation on stochastic control guidance

| Warped image inpainting | MEt3R ↓ | SED ↓ | TSED ↑ |
|---|---|---|---|
| w/o stochastic control guidance | 0.1431 | 1.1966 | 0.5241 |
| w/ stochastic control guidance | **0.1184** | **1.1844** | **0.5588** |

References
- Trajectory Attention (ICLR'25)
- NVS-Solver (ICLR'25)
- MEt3R (CVPR'25)
- PhotoConsistent-NVS (ICCV'23)

## Method: Multi-View Motion Learning

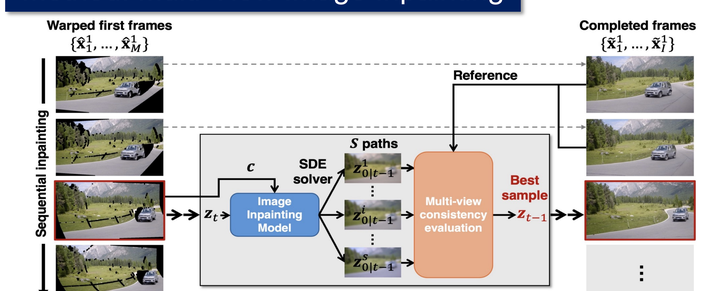



Diffusion loss: $\mathbb{E}_{\epsilon\sim\mathcal{N}(0,I),z_t\sim p,t,c}\left[\|\epsilon-\epsilon_\theta(z_t,t,c)\|_2^2\right]$

Masked diffusion loss: $\mathbb{E}_{\epsilon,z_t^{1:N},m_{\text{down}}^{1:N},t,c}\left[\|\epsilon\odot m_{\text{down}}^{1:N}-\epsilon_\theta(z_t^{1:N},t,c)\odot m_{\text{down}}^{1:N}\|_2^2\right]$

→ To address finetuning on warped videos, masked diffusion loss excludes invisible regions (pixels) in its loss computation

## Method: Multi-View Image Inpainting




## Video Novel View Synthesis Results