

RECENT ADVANCES IN OPEN-SOURCE VIDEO GENERATION AND EDITING

Qifeng Chen

The Hong Kong University of Science and Technology

Sora by OpenAI



Sora + Our Seeing-and-Hearing



Open-source Video Generation and Editing

Video Foundation Models

(High Quality Foundation Models)

Single Domain

Open Domain

World Simulator

Advanced Functions

Controllability

Multimodality

Editing

Character

Local Region

Pose

Visual + Audio

Temporal Consistency

- Video Generation with High Dynamics
- Multimodality Joint Generation

VideoCrafter

An **Open-source** Toolkit for
Text-to-Video Generation and Editing

 VideoCrafter Public

 Edit Pins ▾

 Watch 66 ▾

 Fork 276 ▾

 Starred 3.9k ▾

 main ▾

 4 Branches  1 Tags

 Go to file

t

Add file ▾

 Code ▾

About





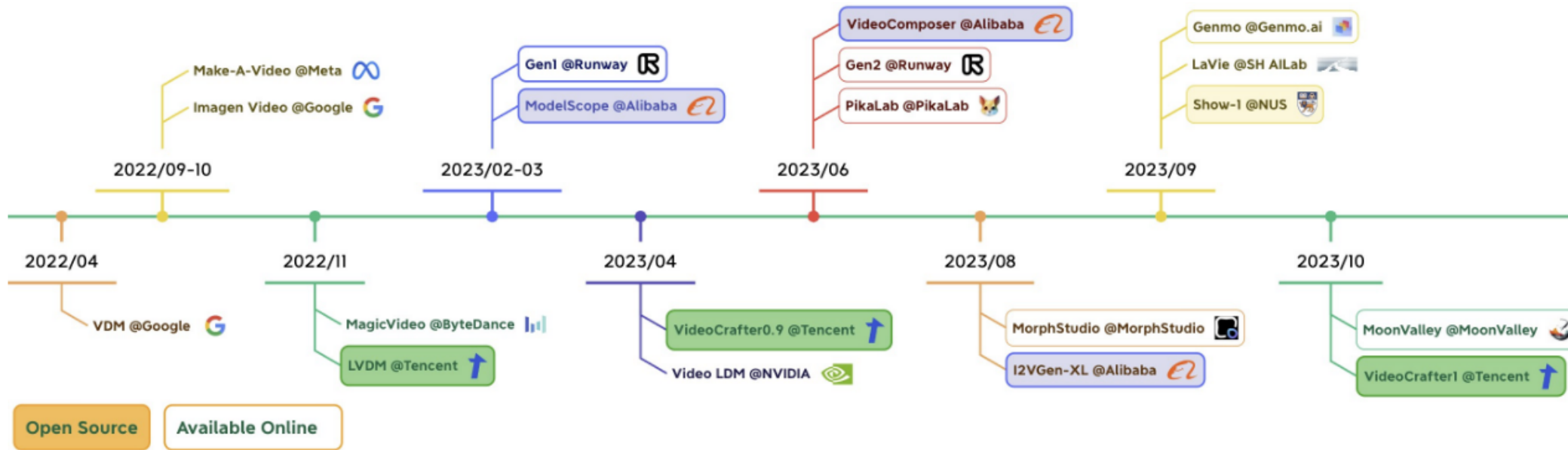
VideoCrafter1: Open Diffusion Models for High-Quality Video Generation

Haoxin Chen ^{1*}, Menghan Xia ^{1*}, Yingqing He ^{1,2*}, Yong Zhang ^{1*}, Xiaodong Cun ^{1*}
Shaoshu Yang ^{1,3}, Jinbo Xing ^{1,4}, Yaofang Liu ^{1,5}, Qifeng Chen ²,
Xintao Wang ^{1†}, Chao Weng ¹, Ying Shan ¹

¹ Tencent AI Lab ² Hong Kong University of Science and Technology

³ Chinese Academy of Sciences ⁴ The Chinese University of Hong Kong ⁵ City University of Hong Kong

Diffusion Models for Video Generation



INSTRUCTION

Evaluation Quality Dimension

subject consistency

background consistency

motion smoothness

aesthetic quality

imaging quality

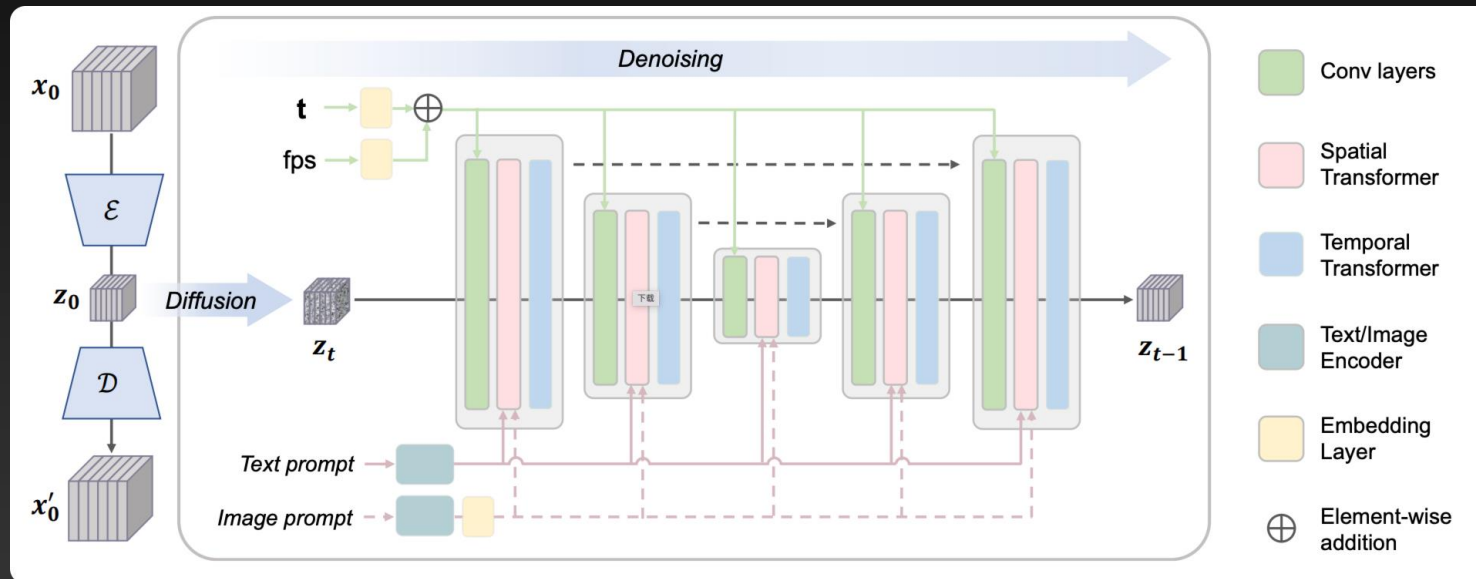
dynamic degree

Model Name (clickable) ▲	Quality Score ▲	Selected Score ▲	subject consistency ▲	background consistenc
Sora	79.69%	79.69%	94.96%	95.84%
Gen-2	78.79%	78.79%	97.61%	97.61%
Pika	78.26%	78.26%	96.76%	98.95%
VideoCrafter-1.0	78.14%	78.14%	95.1%	98.04%
Show-1	77.5%	77.5%	95.53%	98.02%
LaVie-Interpolation	75.86%	75.86%	92.0%	97.33%
LaVie	75.75%	75.75%	91.41%	97.47%
ModelScope	74.91%	74.91%	89.87%	95.29%
VideoCrafter-0.9	71.53%	71.53%	86.24%	92.88%

Foundation Models in Video Generation: *VideoCrafter1*



Foundation Models in Video Generation: VideoCrafter1



Latent Video Diffusion Models (LVDM 2022.11)

- 1024 × 576 resolution
- Image Video Joint Training: 20 million videos + 600 million images.
- Text-to-Video + Image-to-Video
 - The first open-source generic I2V model that can strictly preserve the content and structure of the input reference image while animating it into a video.

Latent Video Diffusion Models for High-Fidelity Long Video Generation

Yingqing He¹

Tianyu Yang²

Yong Zhang²

Ying Shan²

Qifeng Chen¹

¹The Hong Kong University of Science and Technology ²Tencent AI Lab

ScaleCrafter: Tuning-free Higher-Resolution Visual Generation with Diffusion Models

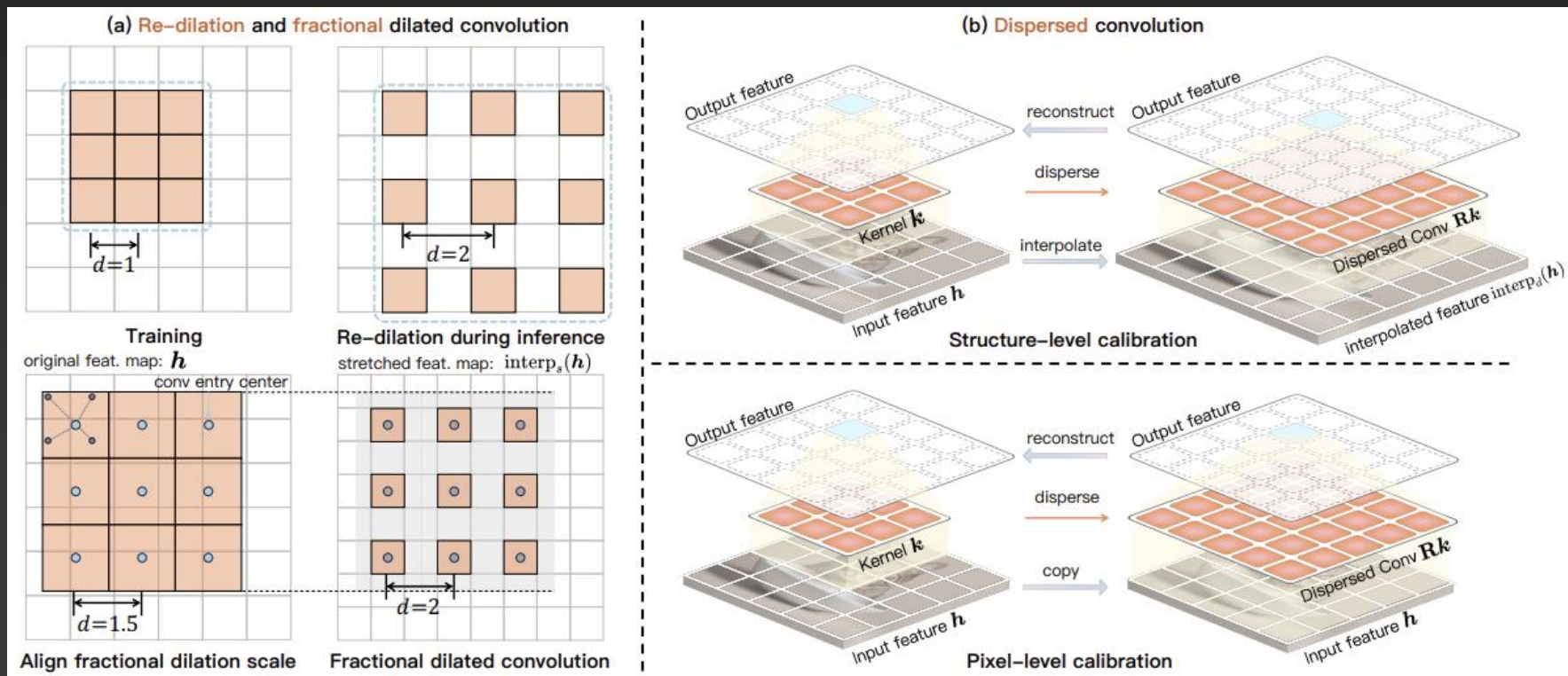
Yingqing He^{*1} Shaoshu Yang^{*2} Haoxin Chen³ Xiaodong Cun³ Menghan Xia³ Yong Zhang^{#3} Xintao Wang³ Ran He² Qifeng Chen^{#1} Ying Shan³

¹ Hong Kong University of Science and Technology ² Chinese Academy of Sciences ³ Tencent AI Lab

(* Equal Contribution # Corresponding Author)

ICLR 2024 Spotlight

4K/2K Image/Video
Generation



Text-to-2K videos (2048 × 1152)



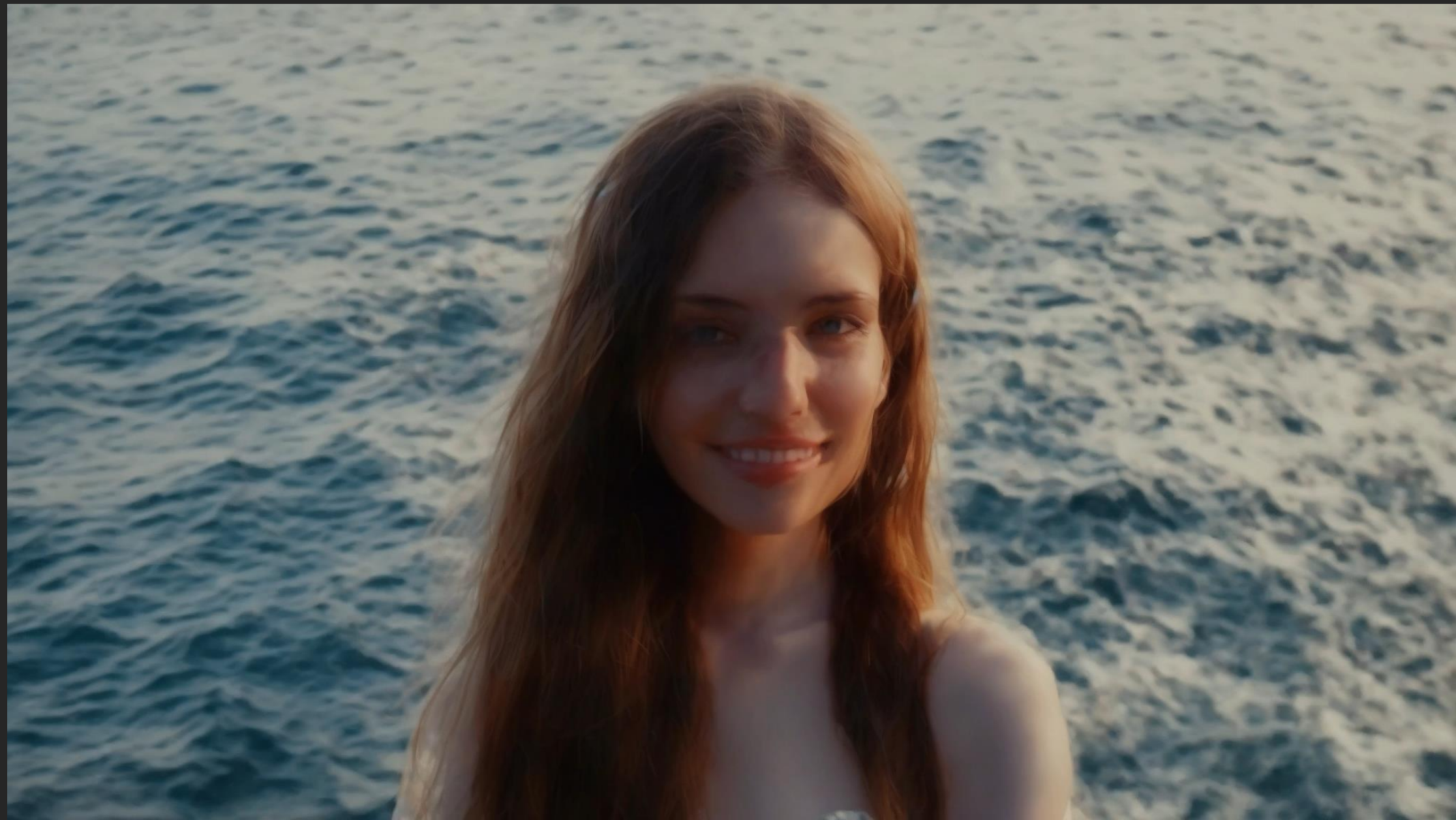
Text-to-2K videos (2048 × 1152)



Text-to-2K videos (2048 × 1152)



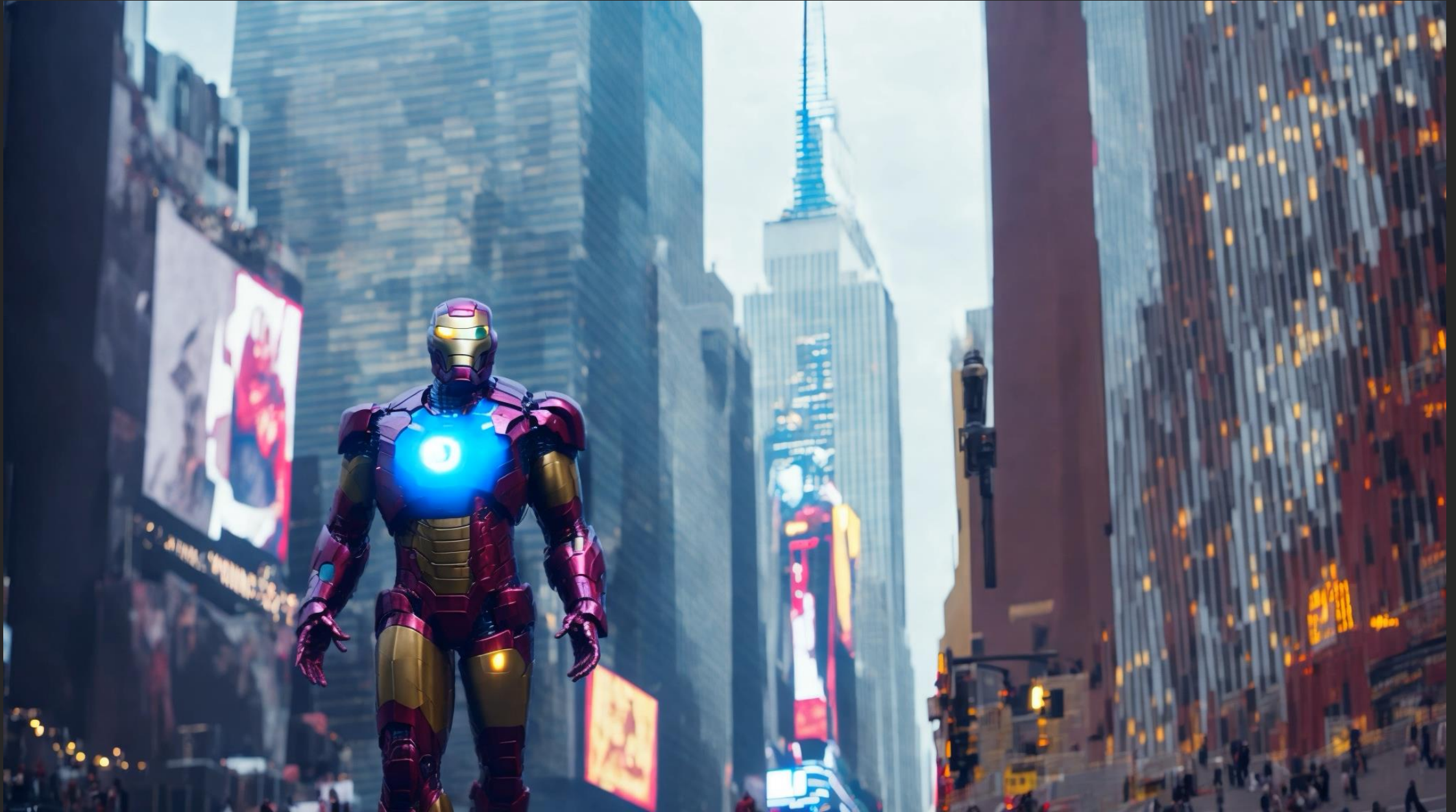
Text-to-2K videos (2048 × 1152)



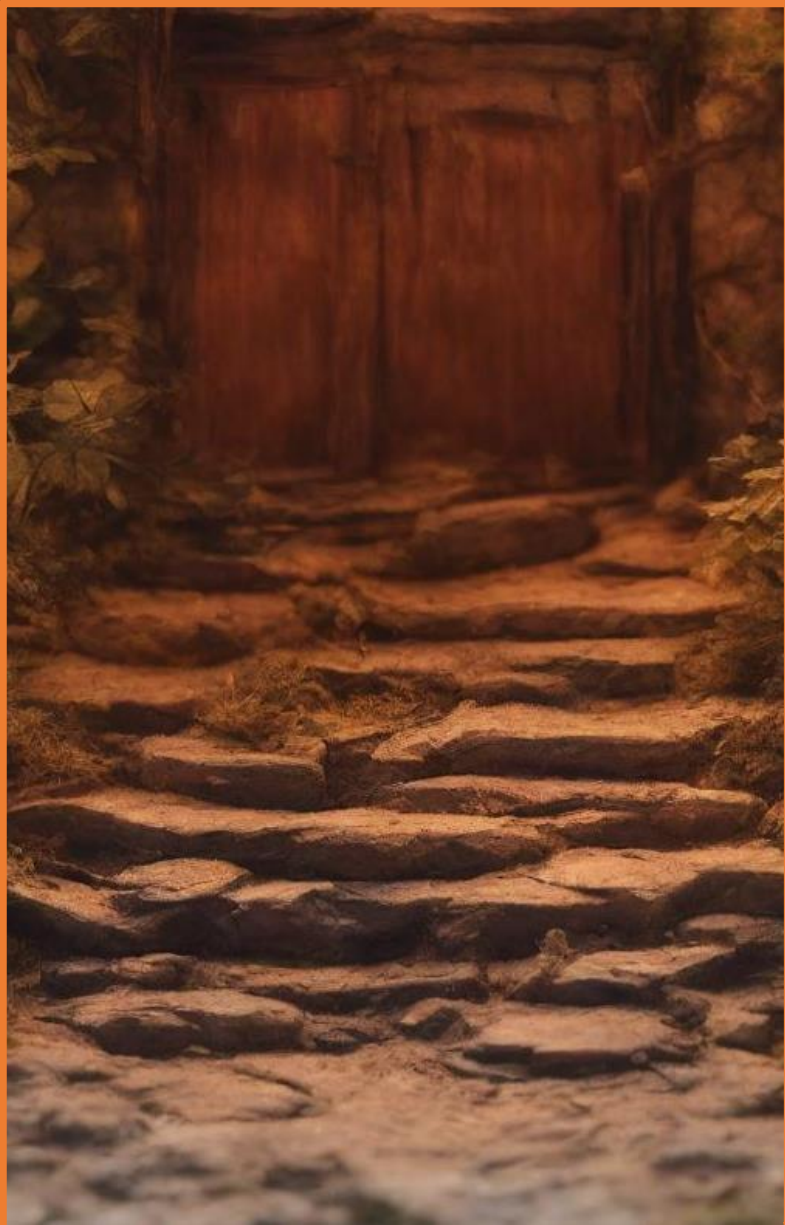
Text-to-2K videos (2048 × 1152)



Text-to-2K videos (2048 × 1152)



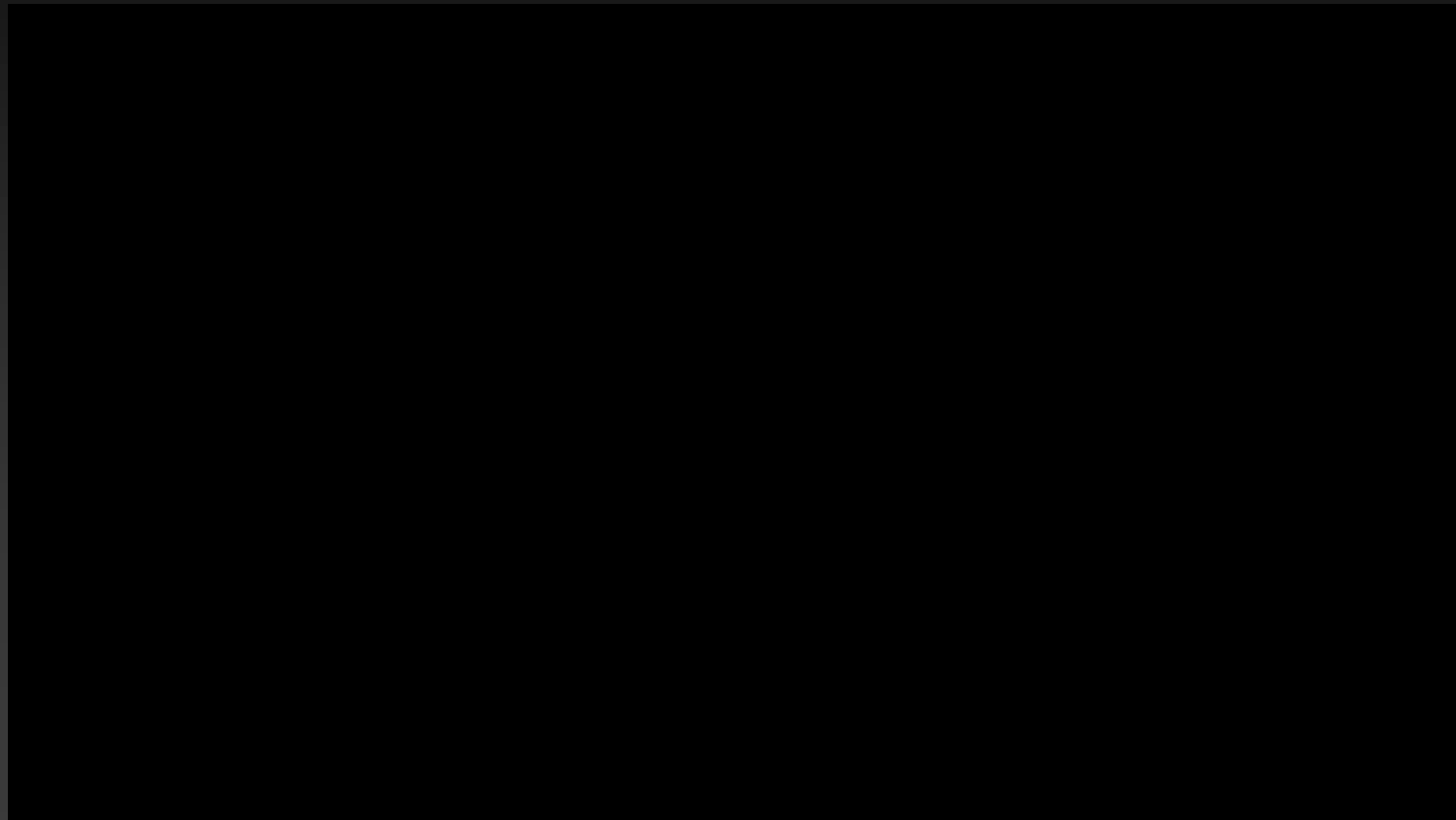
Text-to-4K Images



Controllability

High-Controllability: *Animate-A-Story* for Character Control

Yann LeCun's
Journey in China



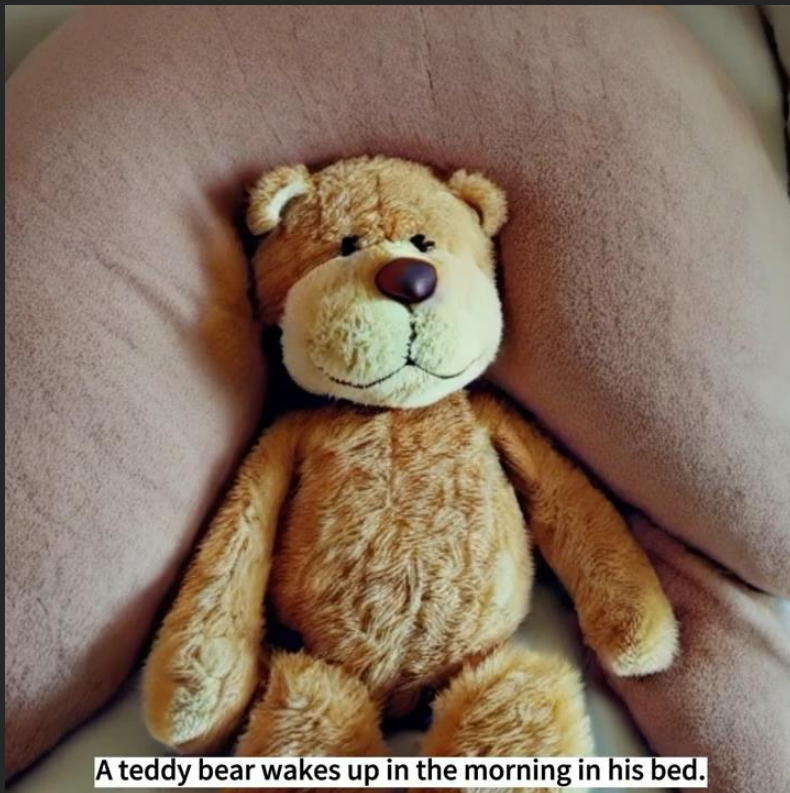
Demo

Controllable Character



A day of a Teddy bear

Duck
Kingdom



A teddy bear wakes up in the morning in his bed.

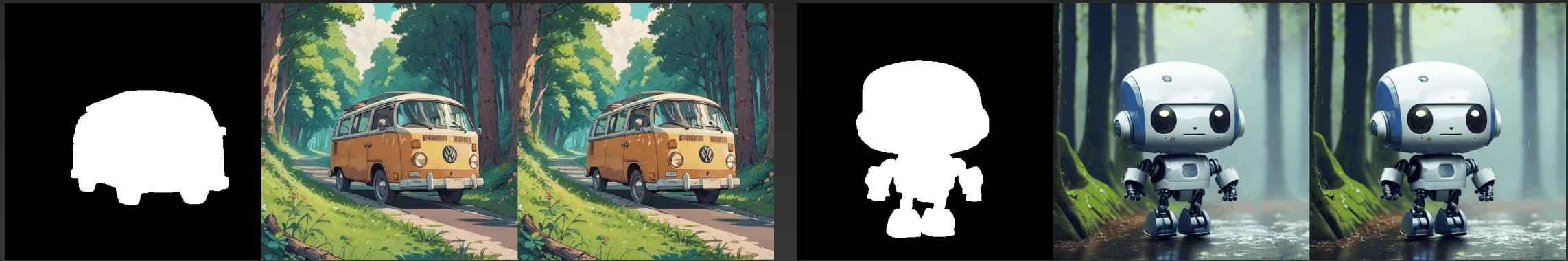
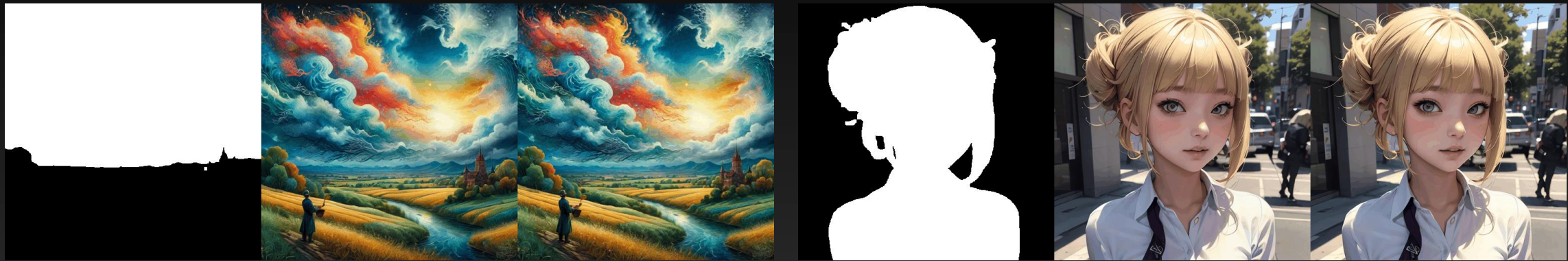


There is a Duck Kingdom

High-Controllability: Follow-Your-Click (in submission)



High-Controllability: *Follow-Your-Click*



FOLLOW-YOUR-POSE : POSE-GUIDED TEXT-TO-VIDEO GENERATION USING POSE-FREE VIDEOS

AAAI 2024

 **FollowYourPose** Public

 Watch **23**

 Fork **63**

 Starred **889**

 main

 1 Branch  1 Tags

 Go to file

 t

Add file

 Code

About

 **mayuelala** Update README.md

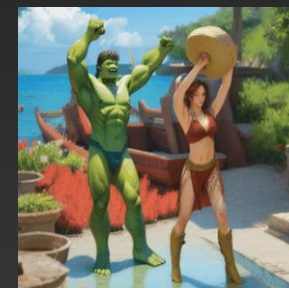
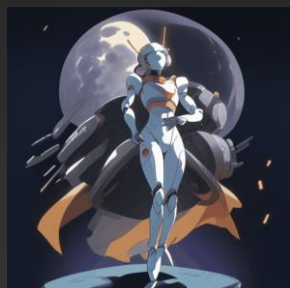
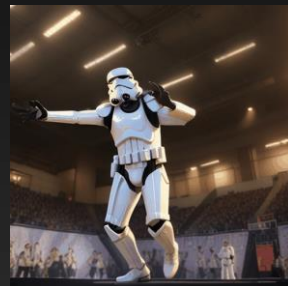
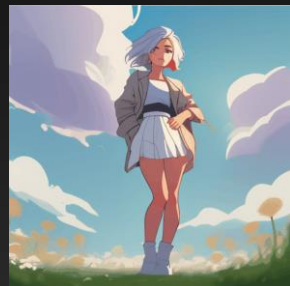
ca0bdb8 · 3 months ago  63 Commits

 configs release code last year

 followyourpose Update pipeline_followyourpose.py last year

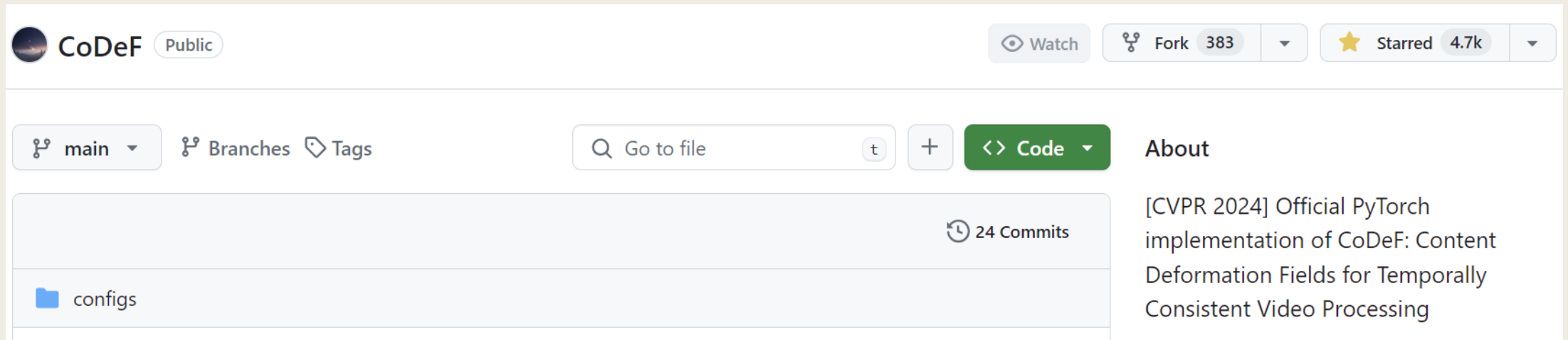
[AAAI 2024] Follow-Your-Pose: This repo is the official implementation of "Follow-Your-Pose : Pose-Guided Text-to-Video Generation using Pose-Free Videos"

High-Controllability: Follow-Your-Pose



CODEF: CONTENT DEFORMATION FIELDS FOR TEMPORALLY CONSISTENT VIDEO PROCESSING CVPR 2024 (HIGHLIGHT)

Hao Ouyang*, Qiuyu Wang*, Yuxi Xiao*, Qingyan Bai, Juntao Zhang, Kecheng Zheng, Xiaowei Zhou, Qifeng Chen†, Yujun Shen† (*equal contribution, †corresponding author)



The screenshot shows the GitHub repository page for 'CoDeF'. The repository is public and has 383 forks and 4.7k stars. The main branch is selected. The repository description is '[CVPR 2024] Official PyTorch implementation of CoDeF: Content Deformation Fields for Temporally Consistent Video Processing'. The repository has 24 commits and a 'configs' folder is visible in the file list.

CoDeF Public

Watch Fork 383 Starred 4.7k

main Branches Tags

Go to file t + Code

About

[CVPR 2024] Official PyTorch implementation of CoDeF: Content Deformation Fields for Temporally Consistent Video Processing

24 Commits

configs



DEMO









FATEZERO: FUSING ATTENTIONS FOR ZERO-SHOT TEXT-BASED VIDEO EDITING ICCV 2023 (ORAL)

Chenyang Qi, Xiaodong Cun, Yong Zhang, Chenyang Lei, Xintao Wang, Ying Shan, and Qifeng Chen

The screenshot shows the GitHub repository page for 'FateZero' by user 'ChenyangQiQi'. The repository is public and has 13 watchers, 97 forks, and 1k stars. The main branch is selected. A commit by 'mayuelala' is visible, titled 'Update README.md', with a commit hash of '9b6cb7b' and 118 commits. The repository description is '[ICCV 2023 Oral] "FateZero: Fusing Attentions for Zero-shot Text-based Video Editing"'. The navigation bar includes options for Code, Issues (7), Pull requests, Actions, Projects, Wiki, Security, and Insights.

ChenyangQiQi / FateZero

Type to search

Code Issues 7 Pull requests Actions Projects Wiki Security Insights

FateZero Public

Watch 13 Fork 97 Starred 1k

main Branches Tags

Code

About

[ICCV 2023 Oral] "FateZero: Fusing Attentions for Zero-shot Text-based Video Editing"

mayuelala Update README.md 9b6cb7b · 7 months ago 118 Commits

Results



"silver jeep → posche car"
Shape Editing



"+ Van Gogh style"
Style Editing



"squirrel, Carrot → rabbit, eggplant"
Attribute Editing

Thank you!

More information at <https://cqf.io>