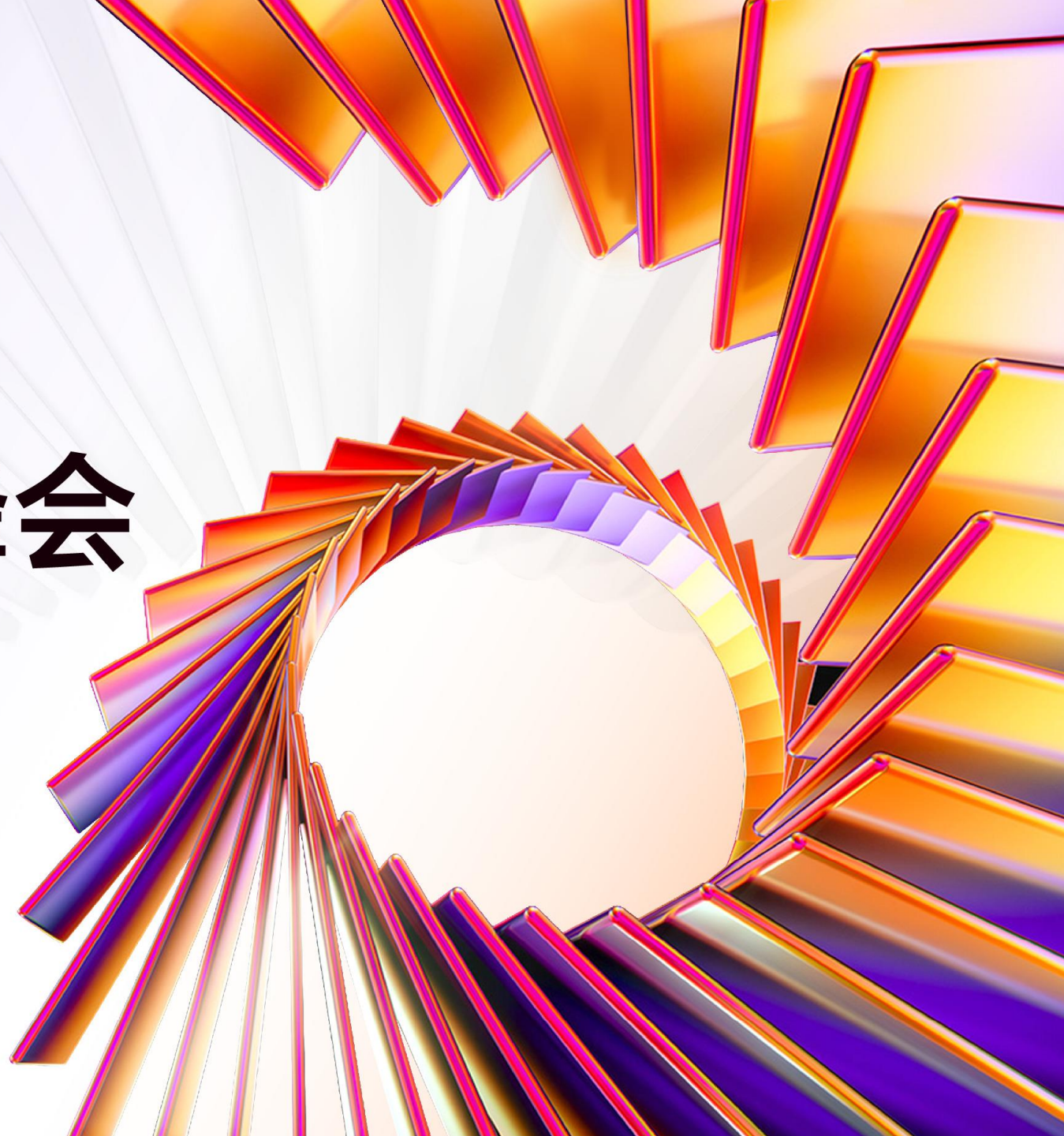
AiDD 6th 2025

2025 AI+ Development Digital Summit

AI+研发数字峰会

拥抱AI 重塑研发

05/23-24 | 上海站



AiDD 7th 2025

2025 AI+研发数字峰会

拥抱AI 重塑研发 AI+ Development Digital Summit

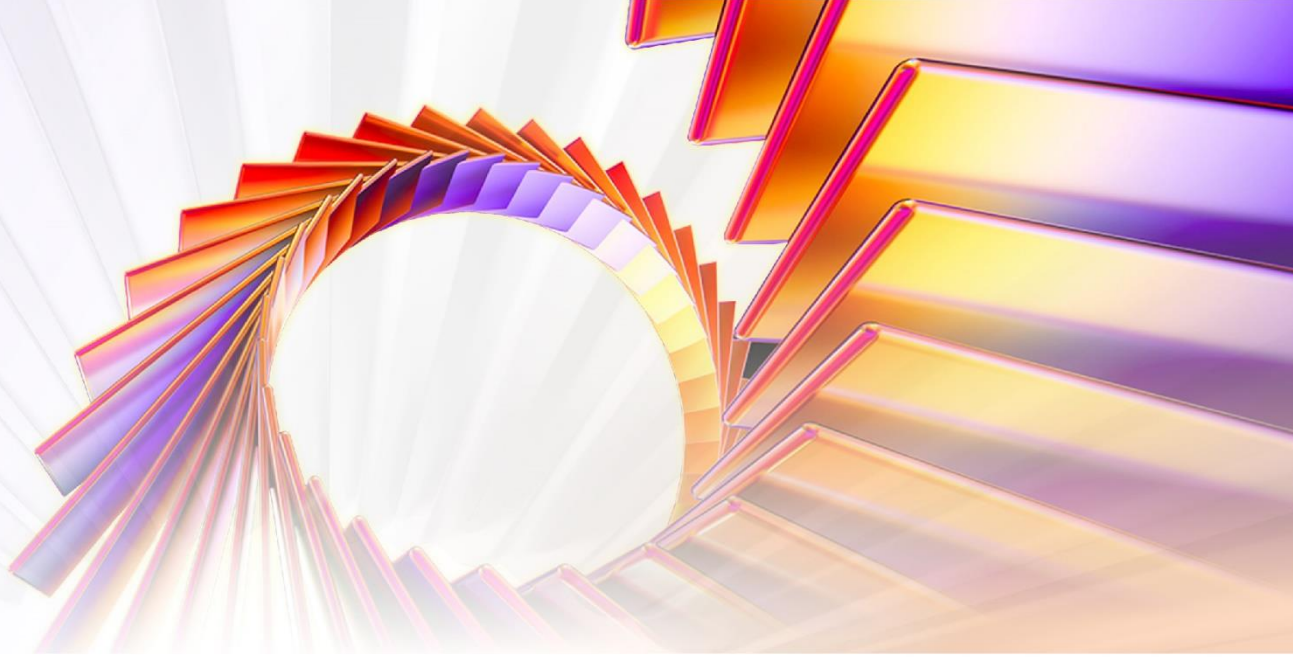
查看会议详情

下一站预告 08/08-09 | 北京站 11/14-15 | 深圳站

北京站论坛设置

- AI+ 金融业务创新
- 具身智能与机器人
- 大模型安全与对齐
- 大模型和 AI 应用评测
- 智能需求工程
- 代码生成及其改进
- 大模型应用开发框架与实践
- 智能存储与检索技术
- 智能体与研发效率工具
- AI+ 新能源汽车
- 智能体经济（Agentic Economy）
- 下一代知识工程
- AI 产品运营与出海策略
- 智能测试工具的开发与应用
- AI 前沿技术探索与实践

AiDD 6th 2025 | 05/23-24 | 上海站
2025 AI+ Development Digital Summit
AI+研发数字峰会
拥抱AI 重塑研发



基于视频生成模型Hallo和Champ的数字人构建

朱思语 | 复旦大学

朱思语

复旦大学教授

复旦大学人工智能创新与产业研究院研究员、长聘教授及博士生导师，研究方向为视频与三维生成模型。在CVPR、ICCV、ECCV、PAMI等国际会议和期刊上发表论文60余篇。博士期间，他联合创立了3D视觉公司Altizure，后被苹果公司收购。2017年至2023年，他担任阿里云人工智能实验室总监。他曾任ICCV/AAAI领域主席/程序委员会成员，荣获中国计算机学会杰出工程师奖。

目录
CONTENTS

数字人：生成式应用爆发

角色动画

hedra · DreaminaAI · LEMONSLICE · puppetry · runway · sync.

创作者克隆

Argil · captions · D-ID · Delphi · hedra · HeyGen

直播克隆或换脸

AKOOL · Deep Live Cam · HeyGen · Magicam · Pickle · SwapStream

广告和营销

arcads · captions · creatify · CreatorKit · HeyGen · icon · JOGG · makeugc · MUKUAI · ReelFarm · TikTok Symphony · TOPVIEW · VEED · vidyard · Yuzu Labs

企业

AKOOL · ANAM · Beyond Presence · CICERO · Colossyan · D-ID · elai. · hedra · HeyGen · Hour One · IMMERSIVEFOX · JOGG · synthesia · VEED

AI 虚拟主播

Hallway · HYPER

开发者工具

A2E · AKOOL · ANAM · Beyond Presence · captions · GENIES · hedra · HeyGen · Simli · sync. · tavus

数字人：生成式应用爆发

角色动画

hedra　Dreamina AI

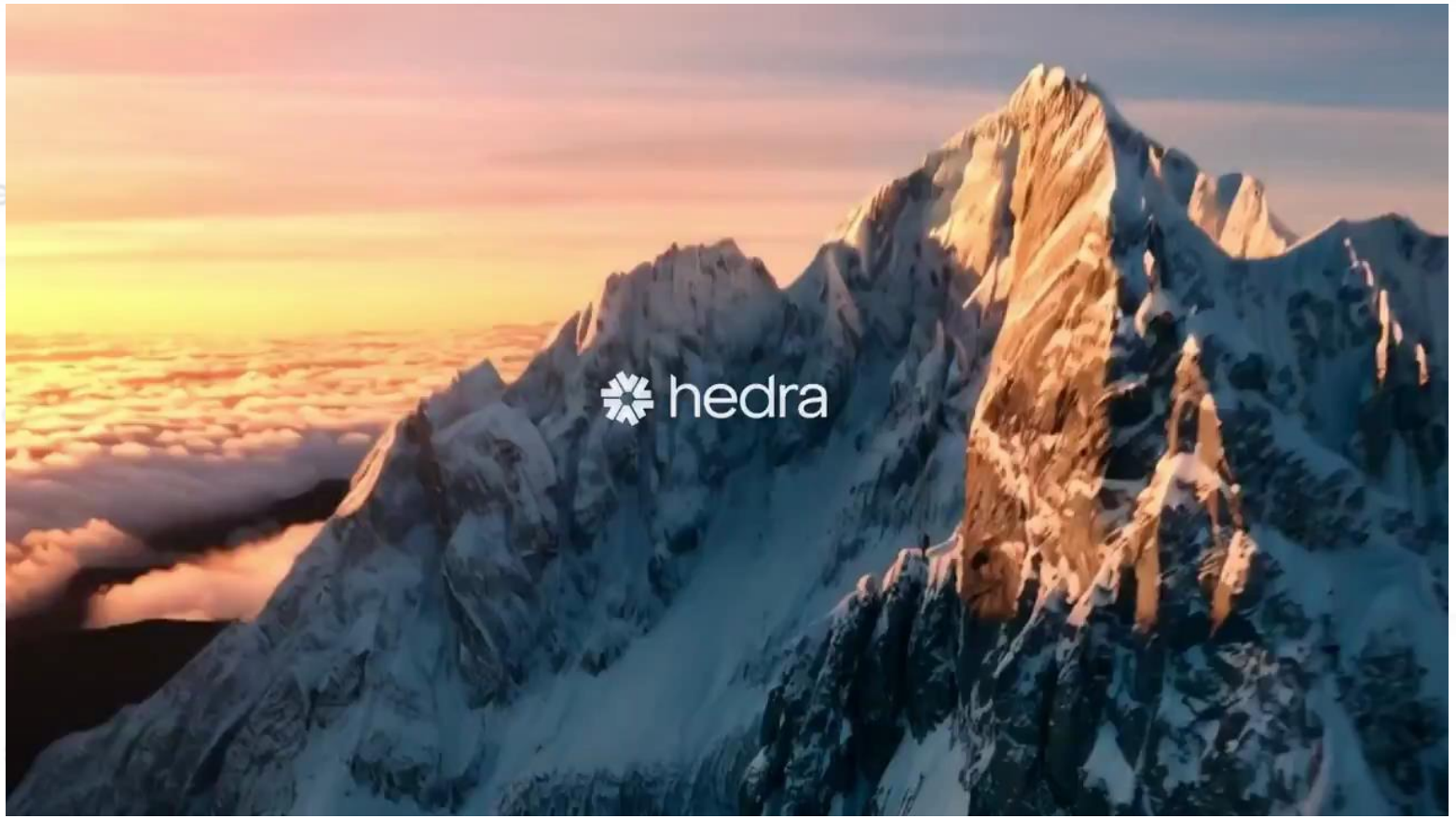
LEMONSLICE　puppetry

runway　sync.

AI 虚拟主播

Hallway　HYPER

A2E　AKOOL　ANAM



数字人：生成式应用爆发

角色动画

hedra　Dreamina AI

LEMONSLICE　puppetry

runway　sync.

AI 虚拟主播

Hallway　HYPER

A2E　AKOOL　ANAM

数字人：生成式应用爆发

创作者克隆

Argil captions D-ID

Delphi hedra HeyGen

直播克隆或换脸

AKOOL Deep Live Cam

HeyGen Magicam

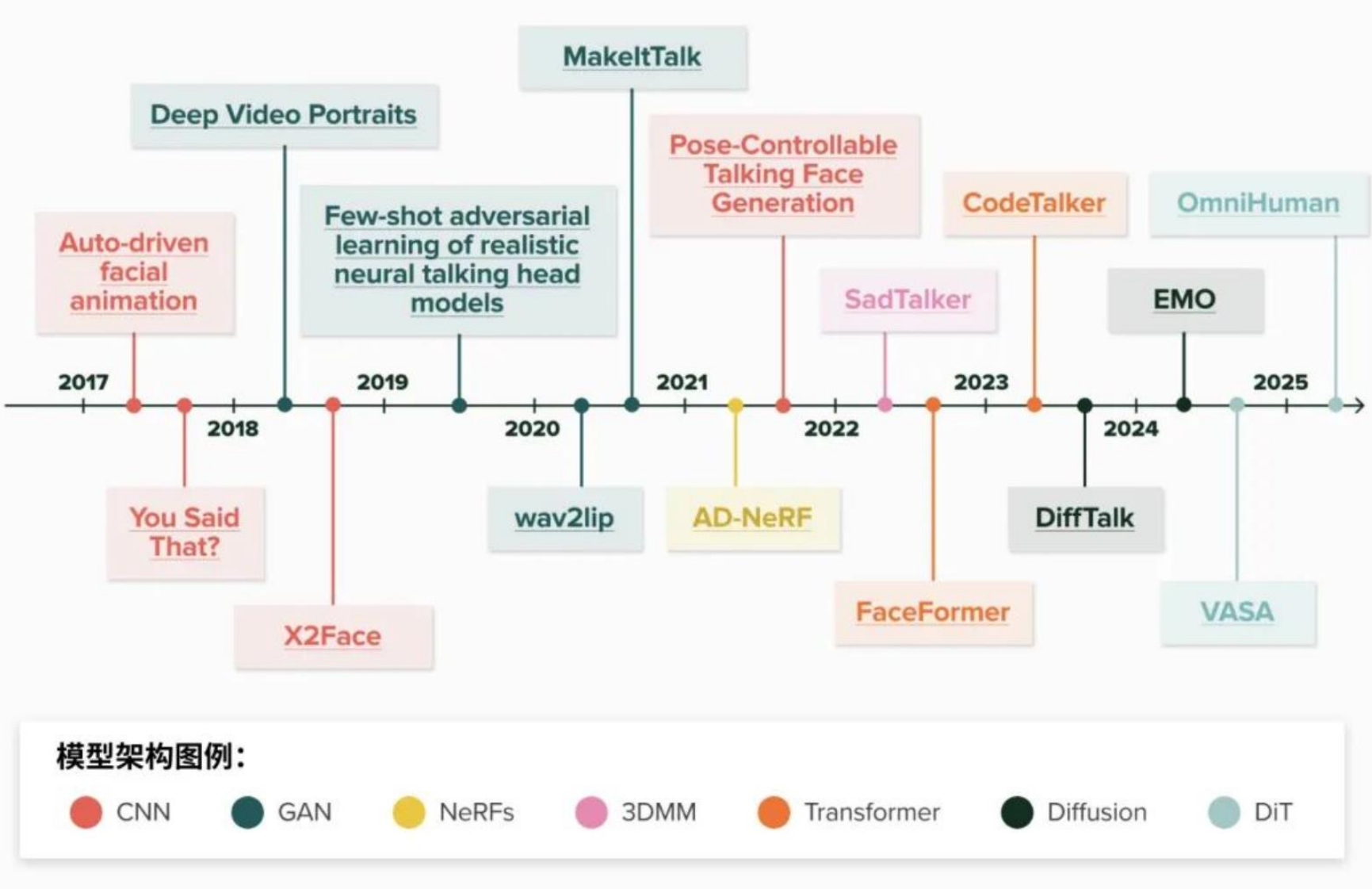
Pickle SwapStream

企业

开发者工具

A2E AKOOL ANAM Beyond Presence captions GENIES hedra HeyGen Simli sync. tav

数字人：主流技术方案

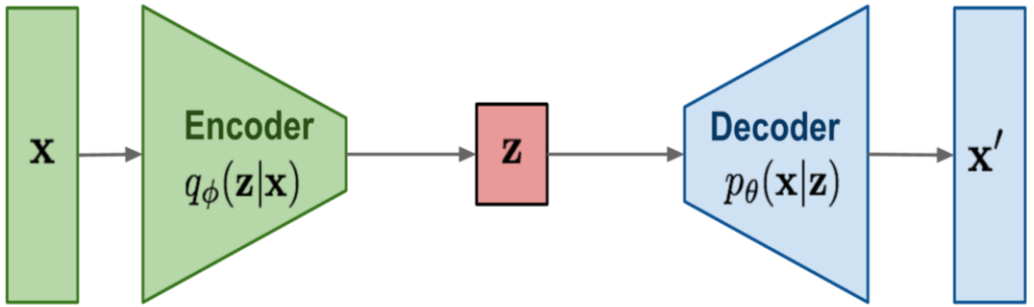

MakeItTalk
Deep Video Portraits
Pose-Controllable Talking Face Generation
CodeTalker
OmniHuman
Auto-driven facial animation
Few-shot adversarial learning of realistic neural talking head models
SadTalker
EMO
2017
2018
2019
2020
2021
2022
2023
2024
2025
You Said That?
X2Face
wav2lip
AD-NeRF
DiffTalk
FaceFormer
VASA
模型架构图例：
CNN
GAN
NeRFs
3DMM
Transformer
Diffusion
DiT

生成式模型的定义

Input

Output

VAE: maximize variational lower bound

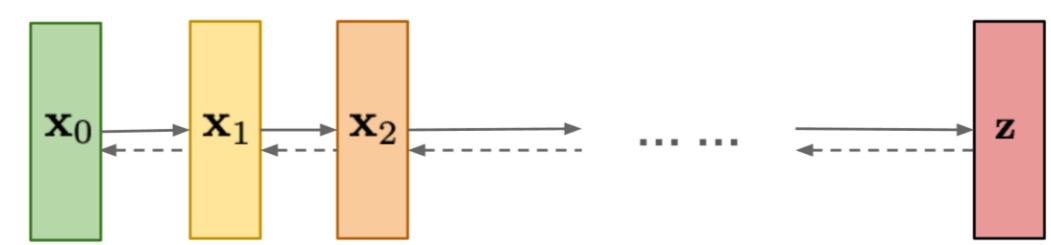


人工智能的关键概念

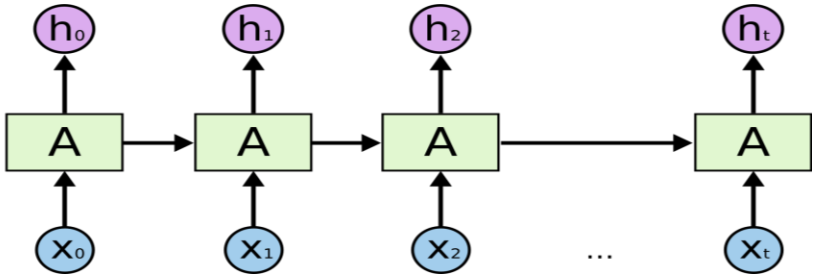
- 人工智能（1956年）
  - 以替代或超越人类智能为目标的计算机智能。
- 机器学习（1990年）
  - 人工智能的一个子领域，通过分析大量数据以识别模式，使计算机系统能自动获取知识并不断改进，从而实现准确决策或预测。
- 深度学习（2012年）
  - 机器学习的一个分支，利用多层神经网络对数据进行处理和决策。
- 生成式AI、大模型（2021年）
  - 基于特定提示、数据或条件，生成新的文本、图像、音频、视频以及分子或材料结构。

视频生成模型的快速发展

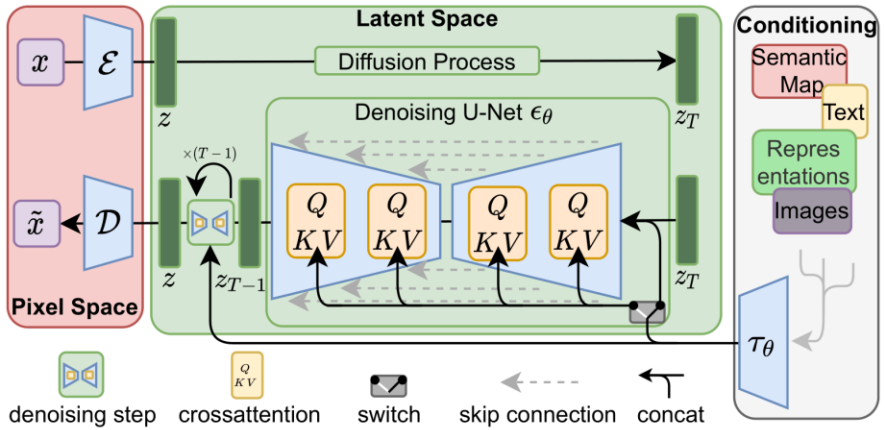
Video Diffusion models



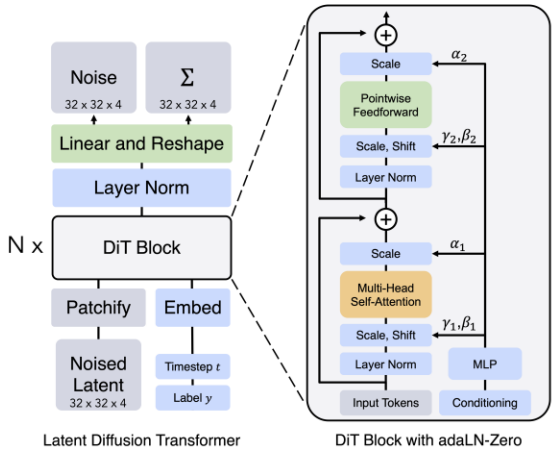
Video Auto-regressive models



Latent Space Diffusion



Diffusion through Transformer



Sora: 世界模拟器?

- Controllable video generation is still chanllenging.
- Can video generation restore the 3D physical world?

- **Appearance**
- **Geometry**
- **Motion & Dynamics**

The Bottleneck of Scaling Law

- Hard to really model the physical word. Failure case in appearance and geometry.

The Bottleneck of Scaling Law

- Hard to really model the physical word! Failure case in motion.

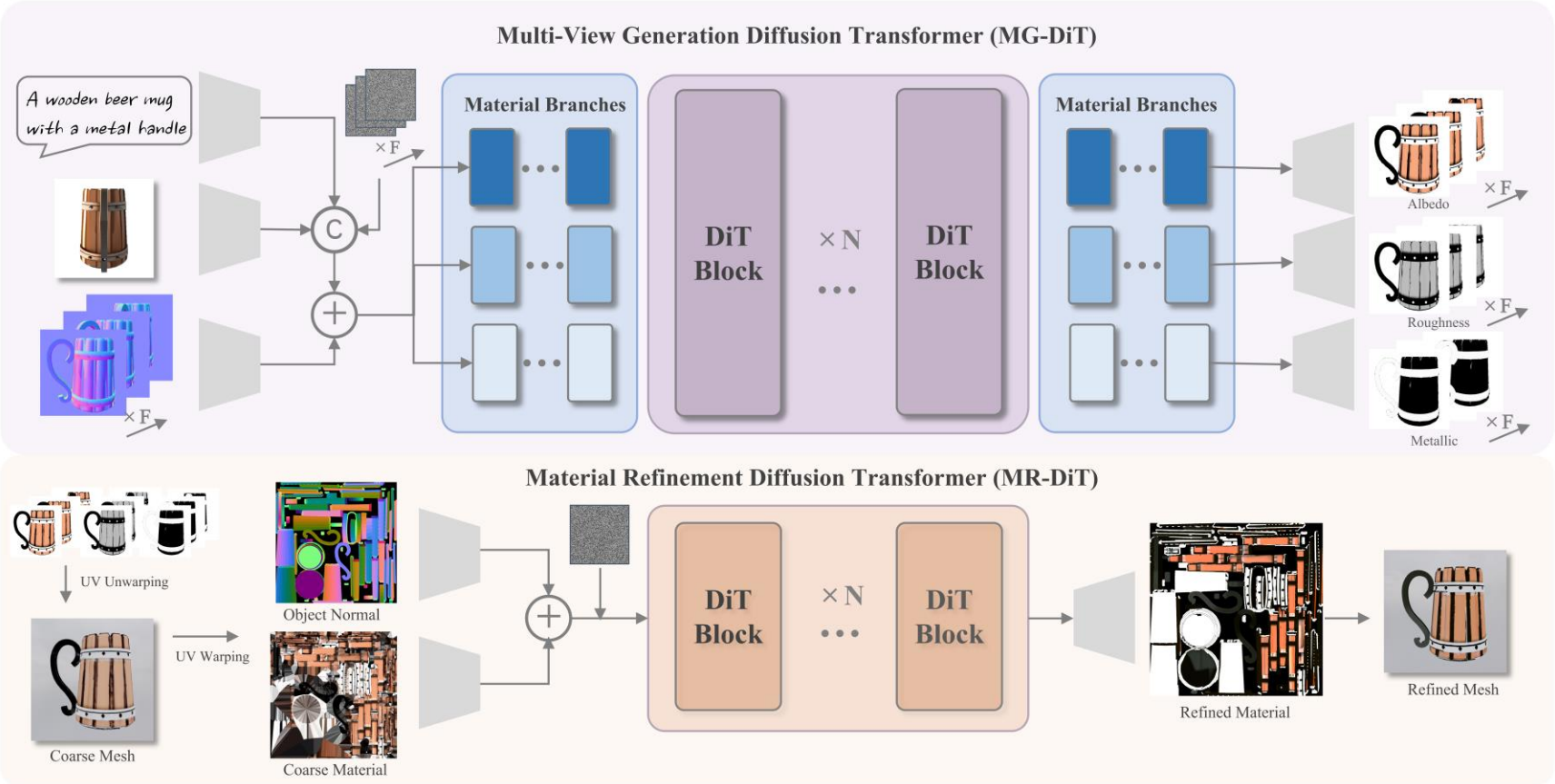
Sora: 世界模拟器?

- Controllable video generation is still chanllenging.
- Can video generation restore the 3D physical world?

- **Appearance**
- **Geometry**
- **Motion & Dynamics**

外观：纹理和材质

- **Appearance and corresponding lighting...**
- MCMat: Multiview-Consistent and Physically Accurate PBR Material Generation.

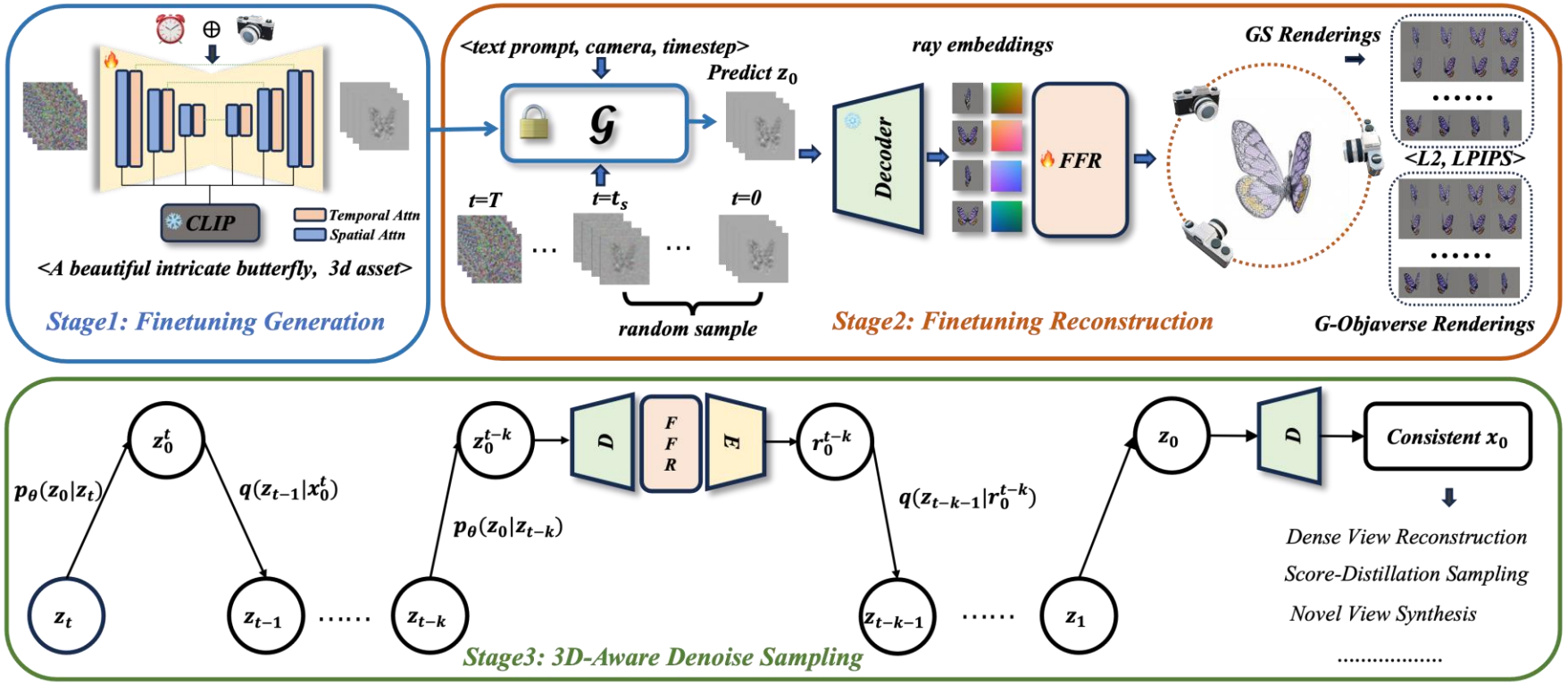


外观：纹理和材质

- **Appearance and corresponding lighting...**
- MCMat: Multiview-Consistent and Physically Accurate PBR Material Generation.

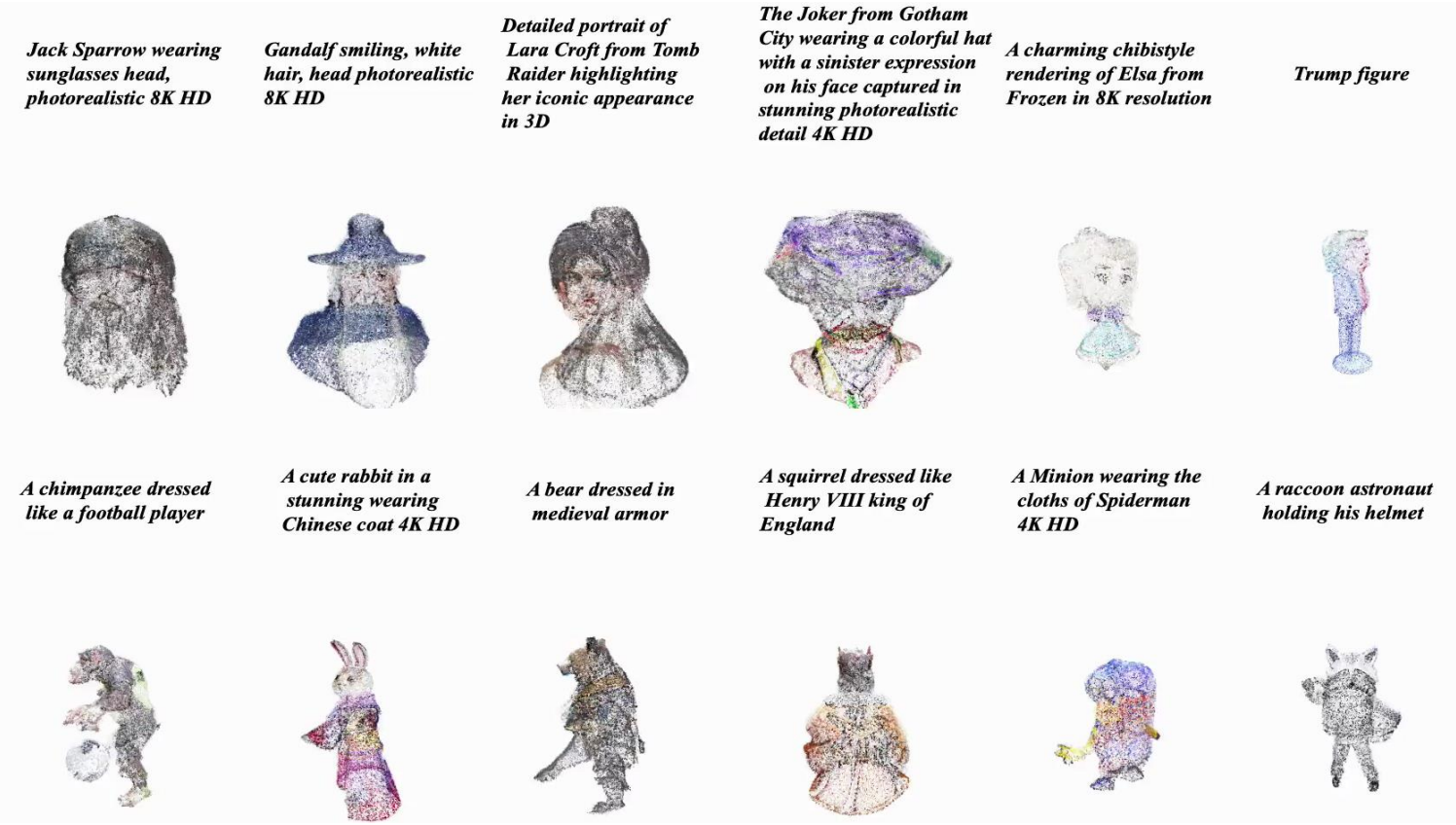
几何：三维形状

- **Directly Generate Dynamic 3D? Static 3D Generation limited to small...**
- VideoMV: Consistent Multi-View Generation Based on Large Video Generative Model.



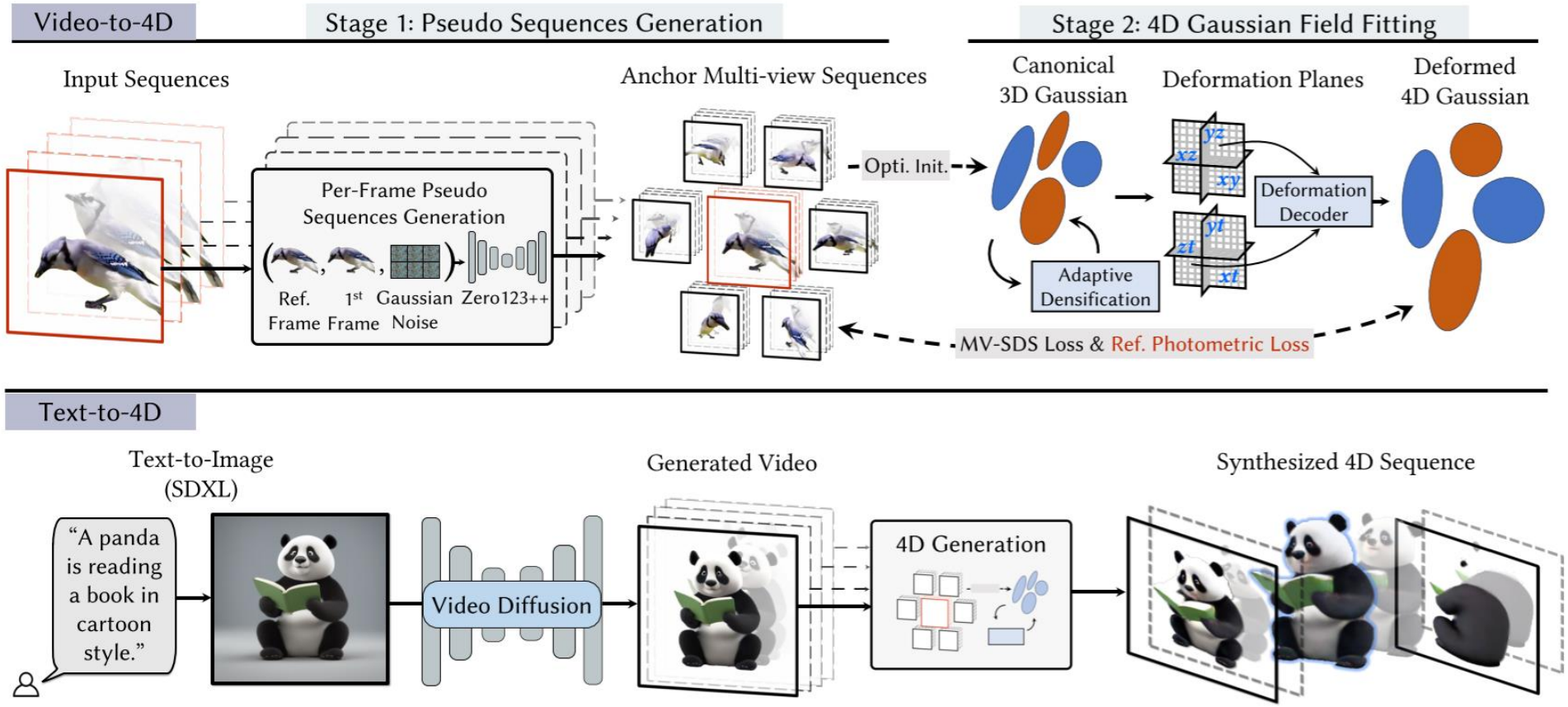
几何：三维形状

- **Directly Generate Dynamic 3D? Static 3D Generation limited to small...**
- VideoMV: Consistent Multi-View Generation Based on Large Video Generative Model.



运动和动画

- **Directly Generate Dynamic 3D? Not to mention 4D generation....**
- [ECCV 2024] STAG4D: Spatial-Temporal Anchored Generative 4D Gaussians



运动和动画

- **Directly Generate Dynamic 3D? Not to mention 4D generation....**
- [ECCV 2024] STAG4D: Spatial-Temporal Anchored Generative 4D Gaussians

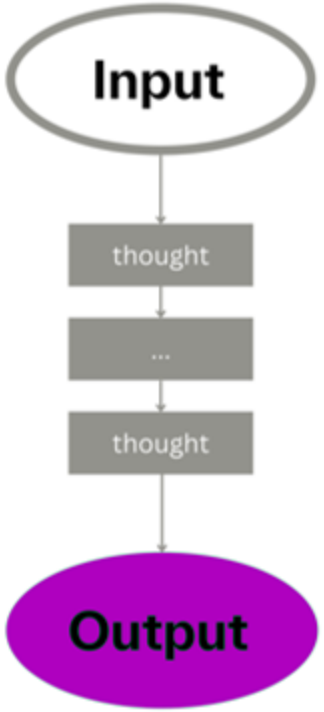
为什么难以建模三维物理世界?

- Reason #1: Latent space not fully semantic.
- Reason #2: Limited conditions; multi-modal information alignment is hard.

- **Appearance**
- **Geometry**
- **Motion & Dynamics**

控制条件多模态融合难

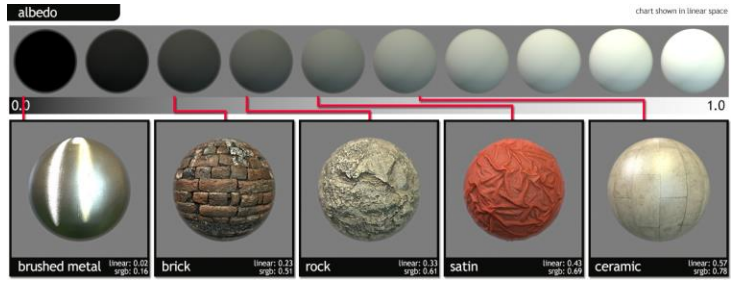
- 更好的3D物理控制条件！ 更好的3D物理条件对齐！



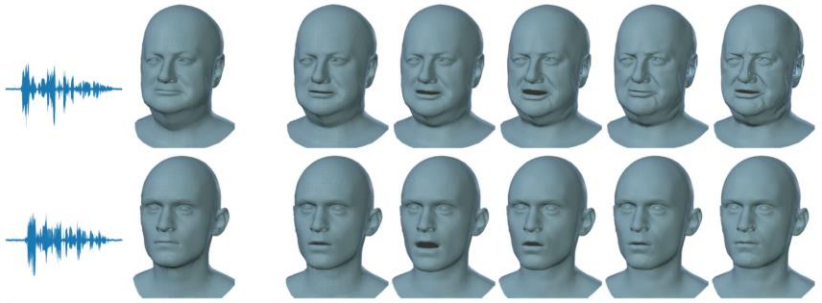
Video CoT

Geometry

Apperance & Lighting

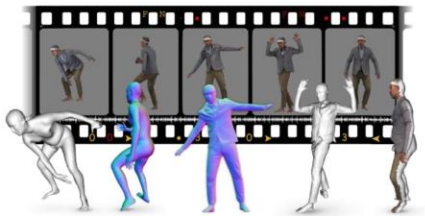


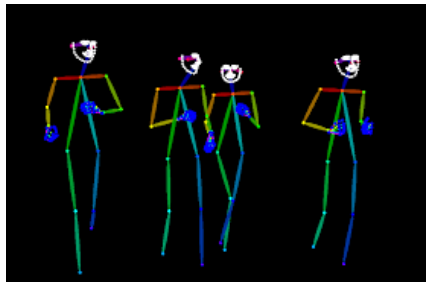
Motion & Animation

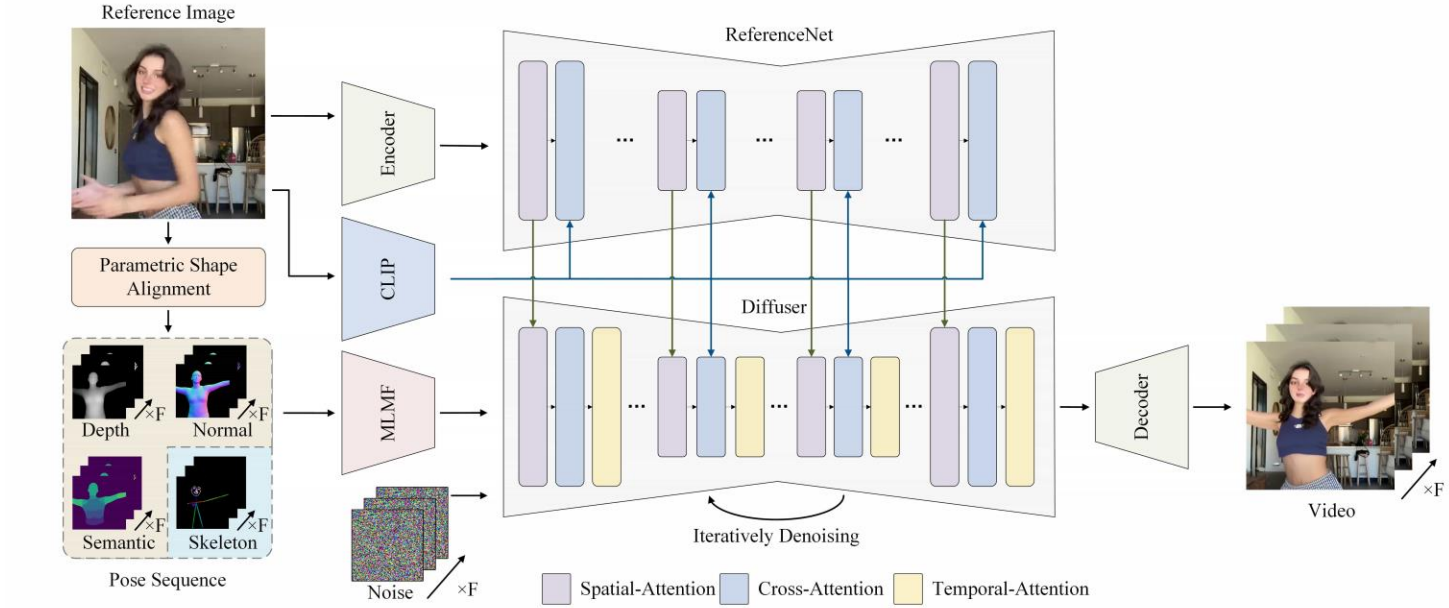


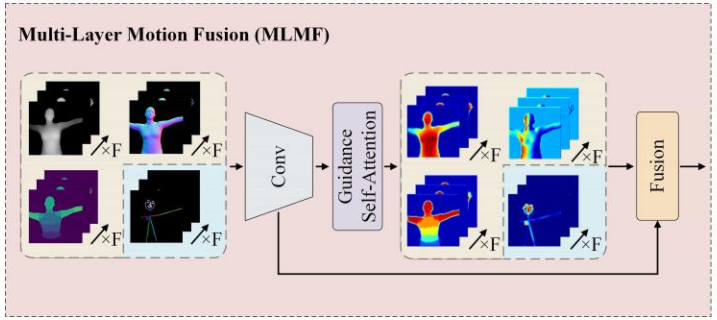
Champ：通过3D参数化控制实现数字人生成

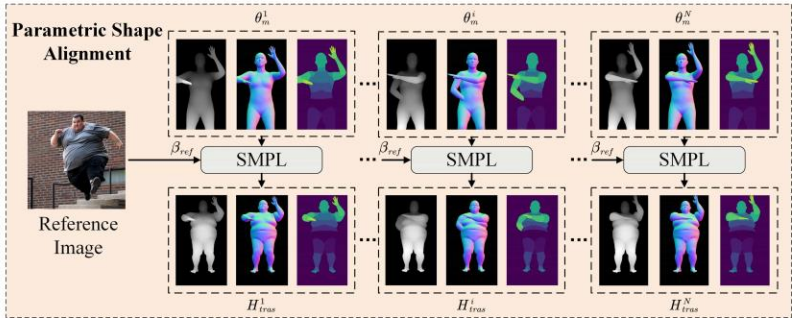
- [ECCV 2024] Champ: Controllable and Consistent Human Image Animation with 3D Parametric Guidance











Champ：通过3D参数化控制实现数字人生成

- [ECCV 2024] Champ: Controllable and Consistent Human Image Animation with 3D Parametric Guidance

Champ: Controllable and Consistent Human Image Animation with 3D Parametric Guidance

fudan-generative-vision.github.io/cha...

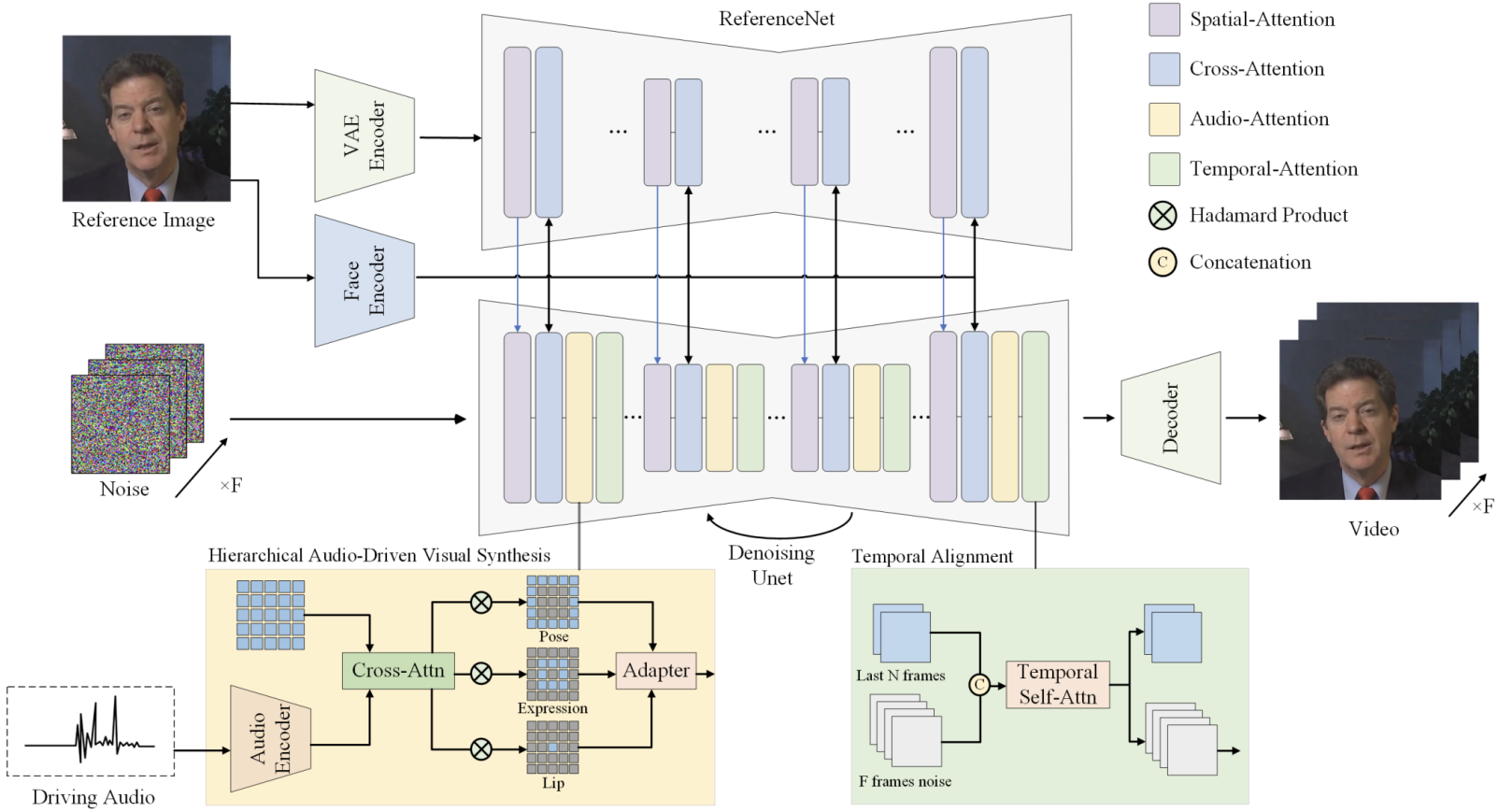
video-generation human-animation image-animatioln

Readme
MIT license
Activity
Custom properties
4.2k stars
216 watching
483 forks

Report repository

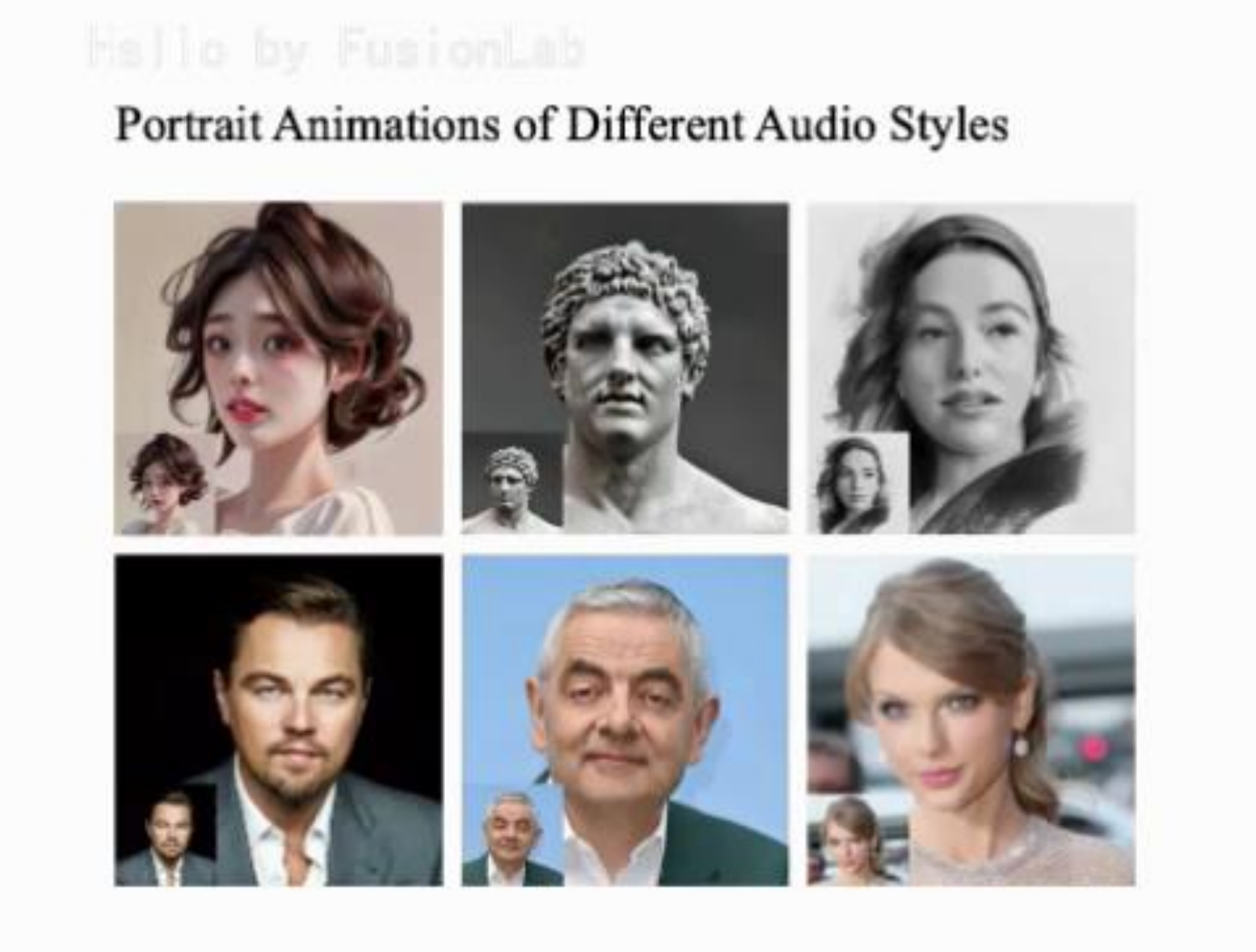
Hallo：通过参数化控制实现数字人脸生成

- Hallo: Hierarchical Audio-Driven Visual Synthesis for Portrait Image Animation



Hallo：通过参数化控制实现数字人脸生成

- Hallo: Hierarchical Audio-Driven Visual Synthesis for Portrait Image Animation



About

Hallo: Hierarchical Audio-Driven Visual Synthesis for Portrait Image Animation

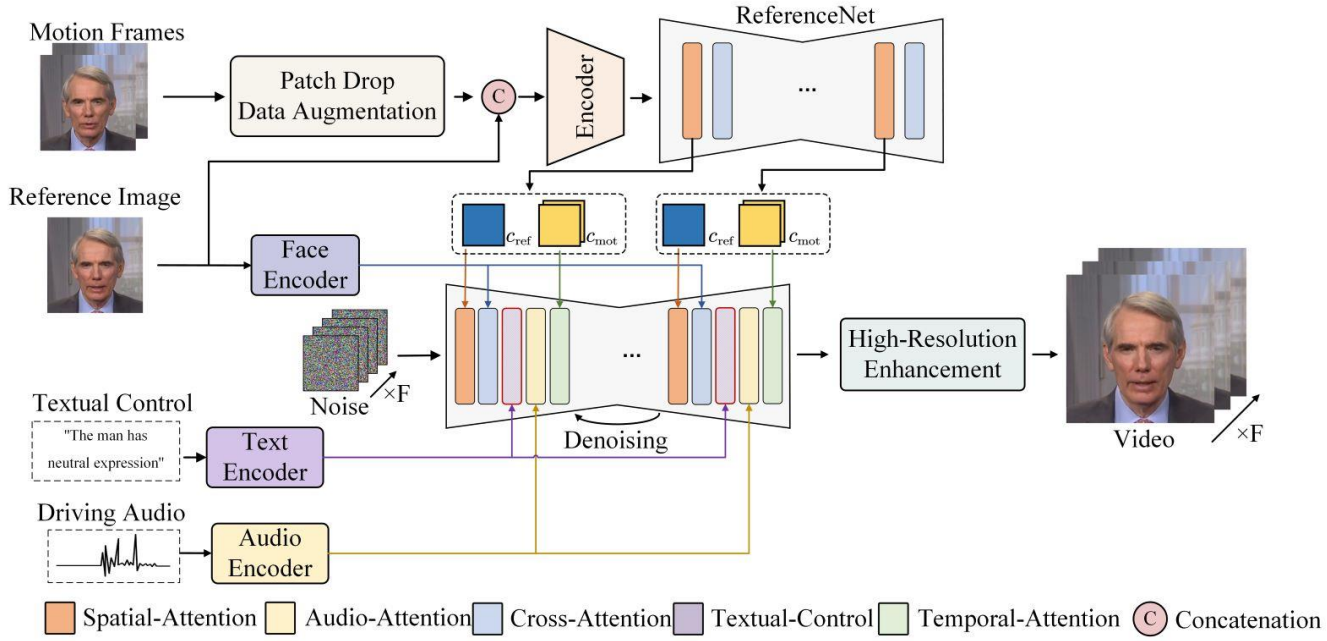
fudan-generative-vision.github.io/hallo/

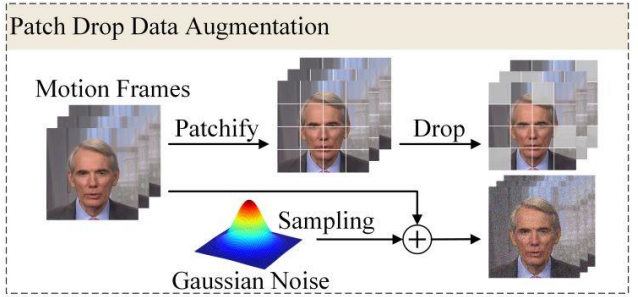
image-animation face-animation video-animation

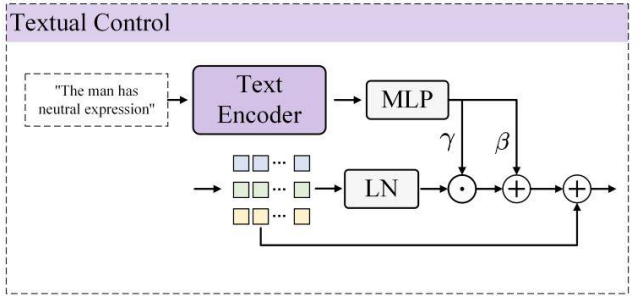
Readme
MIT license
Activity
Custom properties
8.1k stars
535 watching
1.1k forks
Report repository

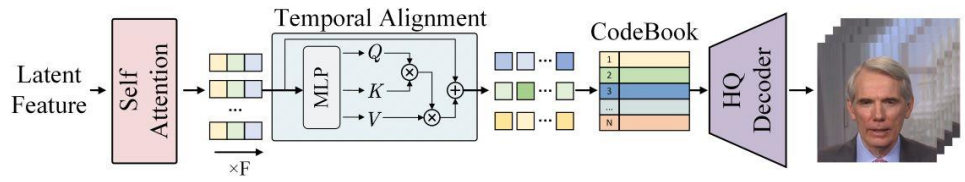
Hallo2：4K+1小时数字人

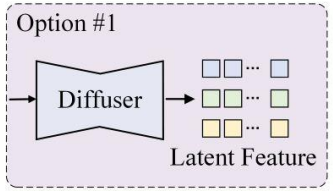
- [ICLR 2025] Hallo2: Long-Duration and High-Resolution Audio-driven Portrait Image Animation

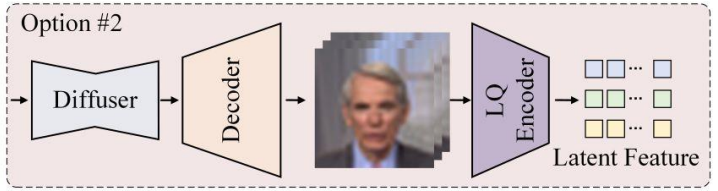






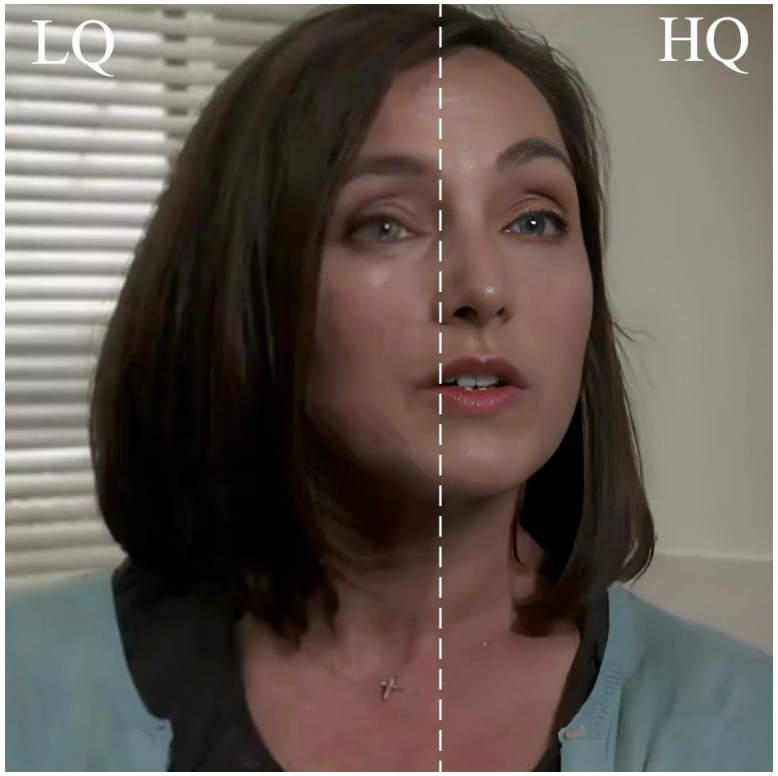


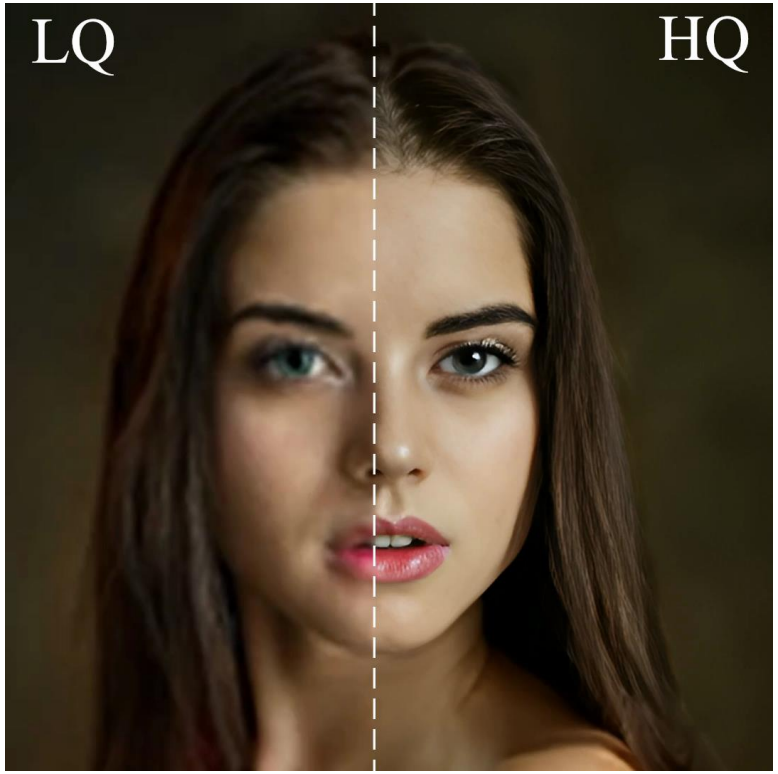




Hallo2：4K+1小时数字人

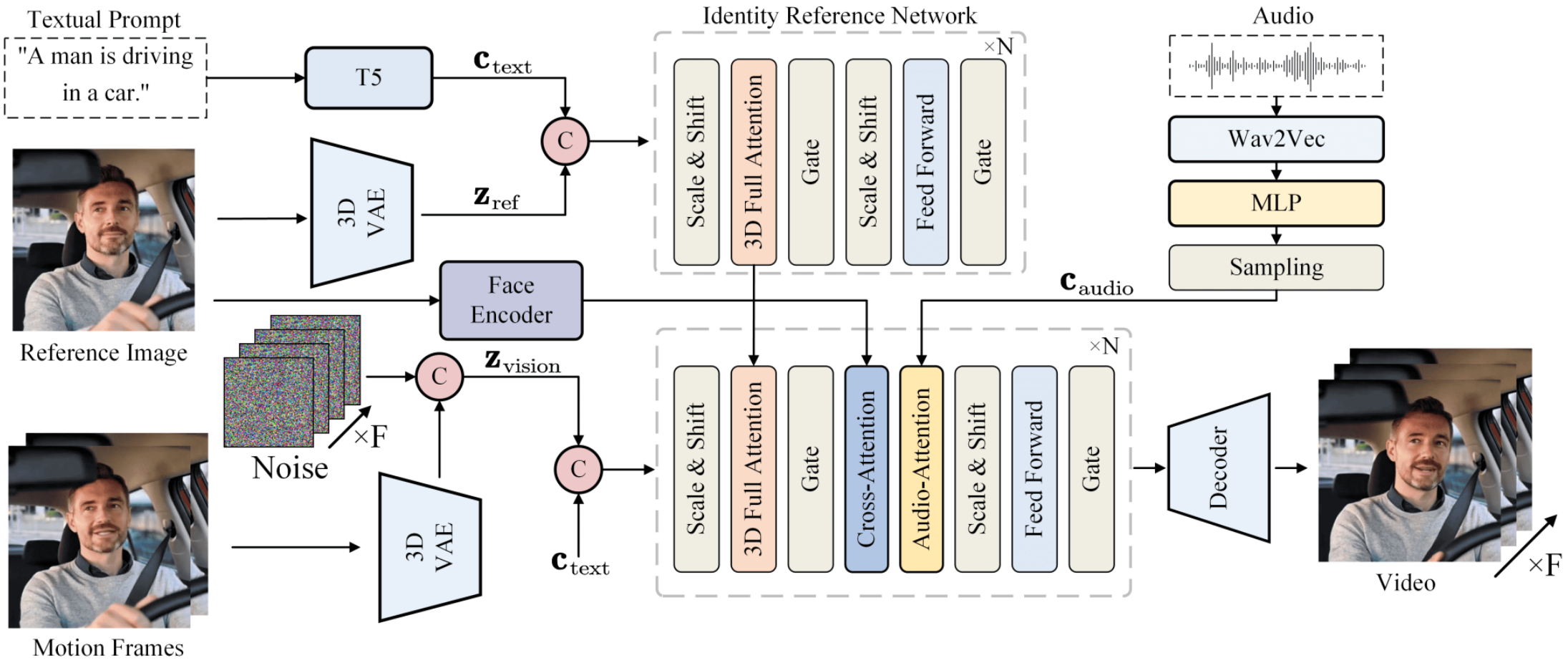
- [ICLR 2025] Hallo2: Long-Duration and High-Resolution Audio-driven Portrait Image Animation





Hallo3：DiT架构+高真实和动态数字人

- [CVPR 2025] Hallo3: Highly Dynamic and Realistic Portrait Image Animation with Vision Diffusion Transformer

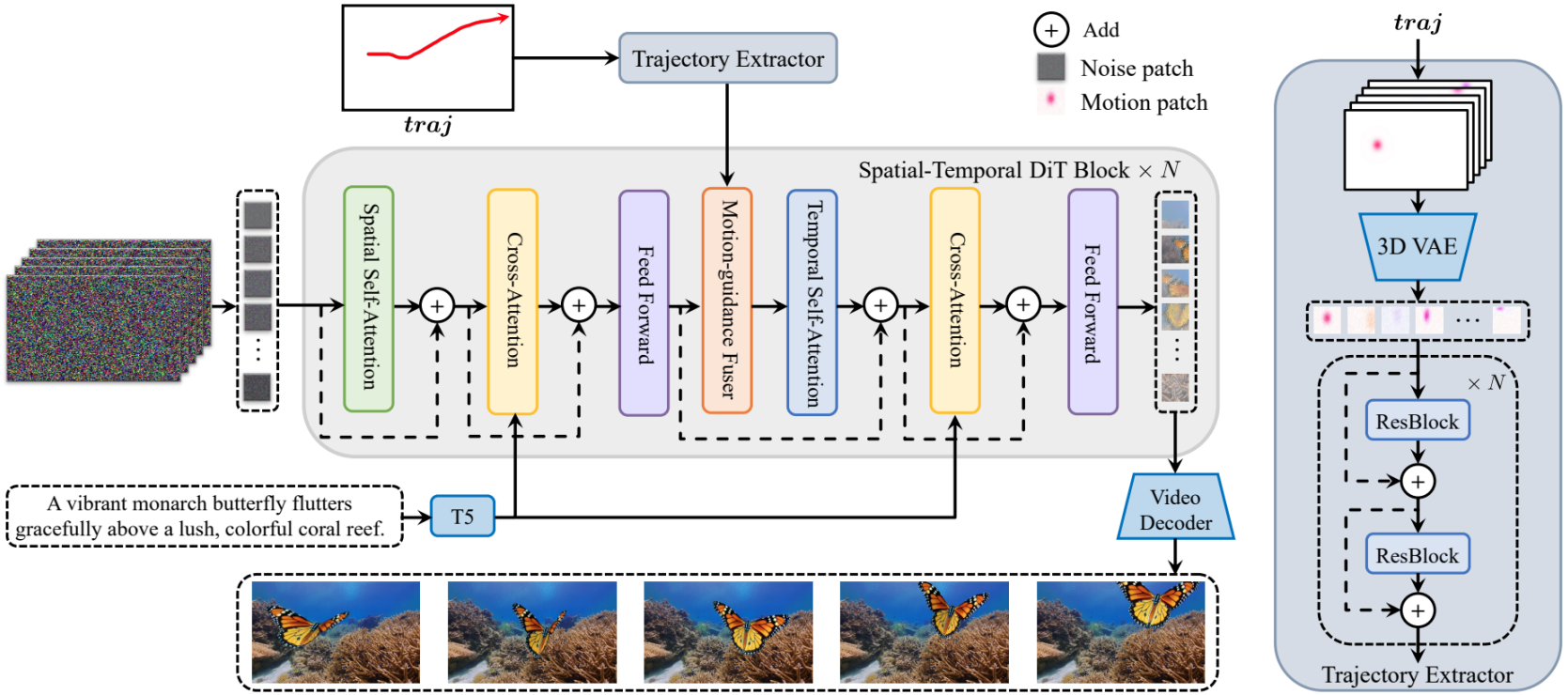


Hallo3：DiT架构+高真实和动态数字人

- [CVPR 2025] Hallo3: Highly Dynamic and Realistic Portrait Image Animation with Vision Diffusion Transformer

通用视频控制：参数化轨迹控制

- [CVPR 2025] Tora: Trajectory-oriented Diffusion Transformer for Video Generation

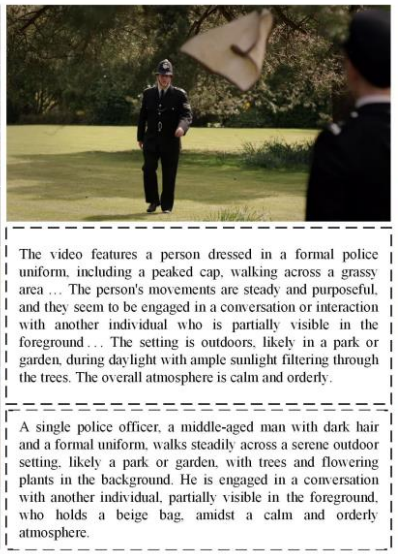


通用视频控制：参数化轨迹控制

- [CVPR 2025] Tora: Trajectory-oriented Diffusion Transformer for Video Generation

数据集: Open Human ViD

- [CVPR 2025] OpenHumanVid: A Large-Scale High-Quality Dataset for Enhancing Human-Centric Video Generation



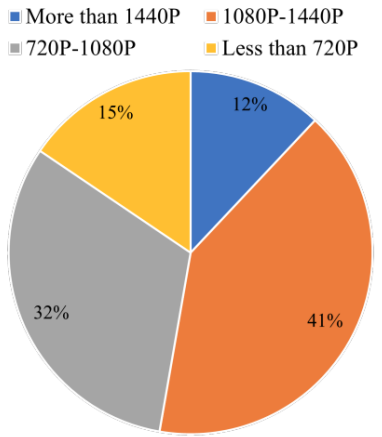
The video features a person dressed in a formal police uniform, including a peaked cap, walking across a grassy area … The person's movements are steady and purposeful, and they seem to be engaged in a conversation or interaction with another individual who is partially visible in the foreground … The setting is outdoors, likely in a park or garden, during daylight with ample sunlight filtering through the trees. The overall atmosphere is calm and orderly.

A single police officer, a middle-aged man with dark hair and a formal uniform, walks steadily across a serene outdoor setting, likely a park or garden, with trees and flowering plants in the background. He is engaged in a conversation with another individual, partially visible in the foreground, who holds a beige bag, amidst a calm and orderly atmosphere.

The video features two individuals in a hist orical setting, likely a period drama. The first person … . The second person, on the right, is a woman with blonde, styled hair, wearing … ,The background is richly decorated, with …

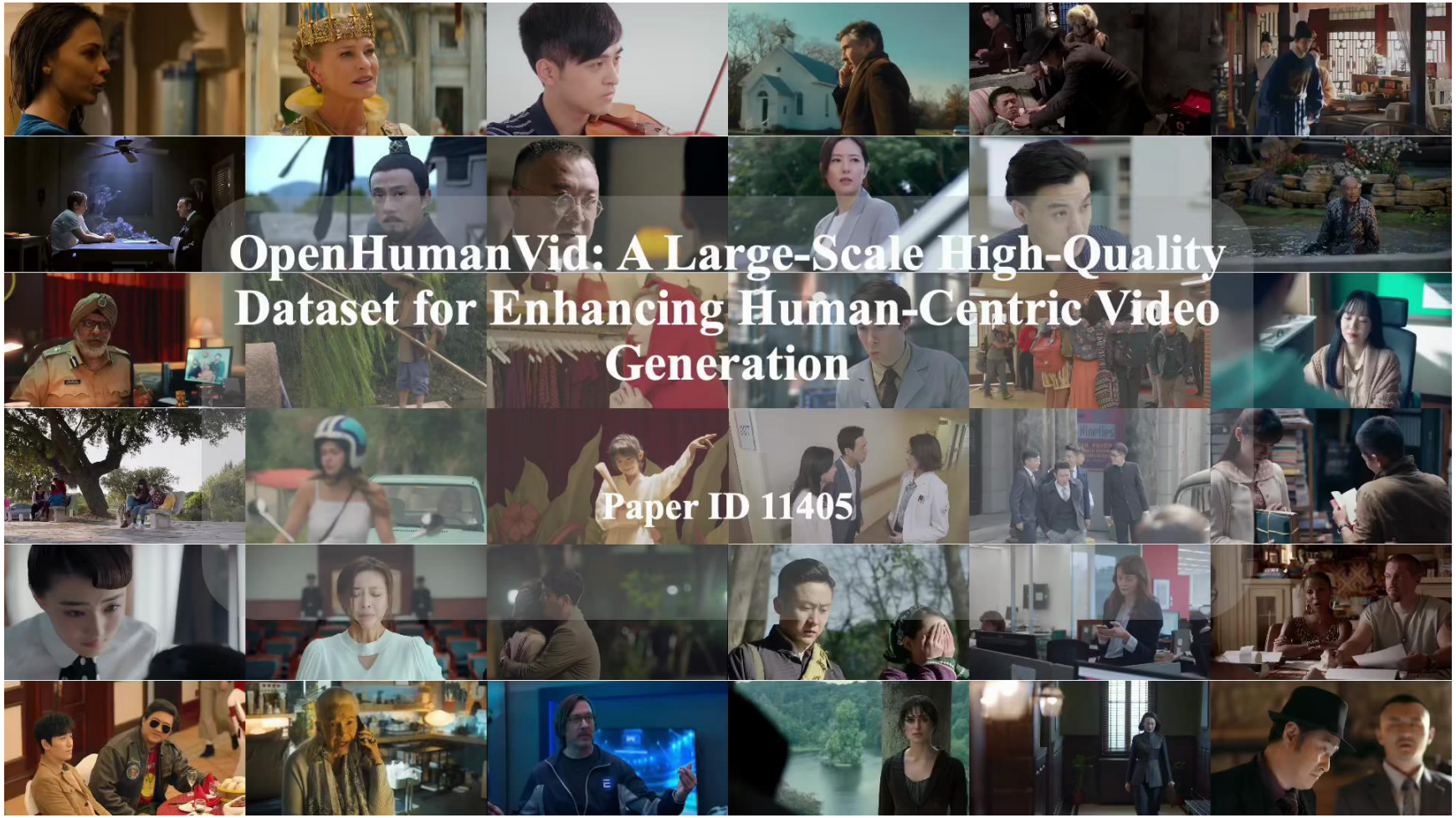
Audio

Dataset name	Year	Domain	# Videos	Total length (hours)	Caption type	Motion type	Resolution
WebVid-10M [2]	2021	Open	10M	52K	Short	Text	360P
Panda-70M [14]	2024	Open	70M	167K	Short	Text	720P
OpenVid-1M [36]	2024	Open	1M	2K	Long	Text	512P
Koala-36M [51]	2024	Open	36M	172K	Long	Text	720P
UCF-101 [46]	2012	Human action	13.3K	26.7	Short	Text	240P
NTU RGB+D [45]	2014	Human action	114K	37	-	3D pose, depth	1080P
MSP-Avatar [43]	2015	Human action	74	3	-	Speech audio, landmark, pose	1080P
ActivityNet [7]	2017	Human action	100K	849	Short	Text	-
TikTok-v4 [11]	2023	Human dance	350	1	-	Skeleton	-
Ours	2024	Human	52.3M	70.6K	Short, long, structured	Text, skeleton pose, speech audio	720P
Ours (filtered)	2024	Human	13.2M	16.7K	Short, long, structured	Text, skeleton pose, speech audio	720P



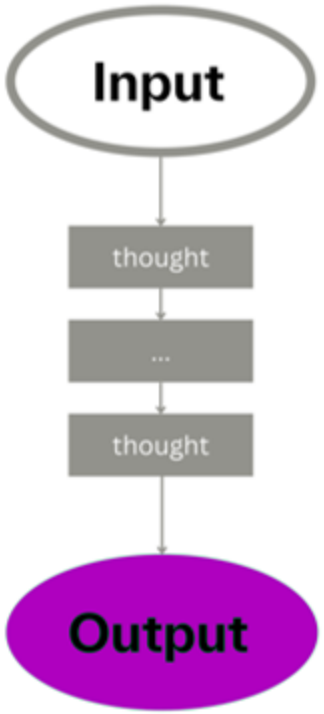
数据集: Open Human ViD

- [CVPR 2025] OpenHumanVid: A Large-Scale High-Quality Dataset for Enhancing Human-Centric Video Generation



未来工作:

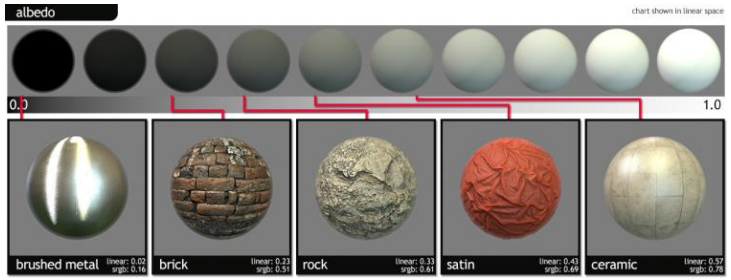
- 更多的3D物理控制条件！ 更好的3D物理条件对齐！



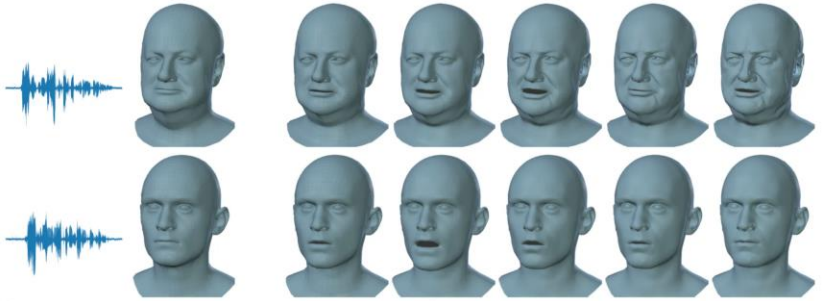
Video CoT

Geometry

Apperance & Lighting



Motion & Animation



AiDD 6th 2025

参与调研您将优先获得

AiDD定制版
《AI+软件研发精选案例》

专属学习顾问
1对1需求对接

AiDD会后小调研

AiDD峰会致力于协助企业利用AI技术深化计算机对现实世界的理解，推动研发进入智能化和数字化的新时代。作为峰会的重要共建者，您的真知灼见对我们至关重要。衷心感谢您的参与支持！

扫码参与调研

2025 AI+研发数字峰会

拥抱AI 重塑研发

科技生态圈峰会 + 深度研习

——1000 + 技术团队的共同选择

K+峰会

K+峰会 敦煌站
K+ 思考周®研习社
时间：2025.08.29-30

K+峰会 上海站
K+ 金融专场
时间：2025.09.26-27

K+峰会 香港站
K+ 思考周®研习社
时间：2025.11.17-18

K+峰会详情

AiDD峰会

AiDD峰会 上海站
AI+研发数字峰会
时间：2025.05.23-24

AiDD峰会 北京站
AI+研发数字峰会
时间：2025.08.08-09

AiDD峰会 深圳站
AI+研发数字峰会
时间：2025.11.14-15

AiDD峰会详情

2025 AI+研发数字峰会
AI+ Development Digital Summit

感谢聆听！

扫码领取会议PPT资料