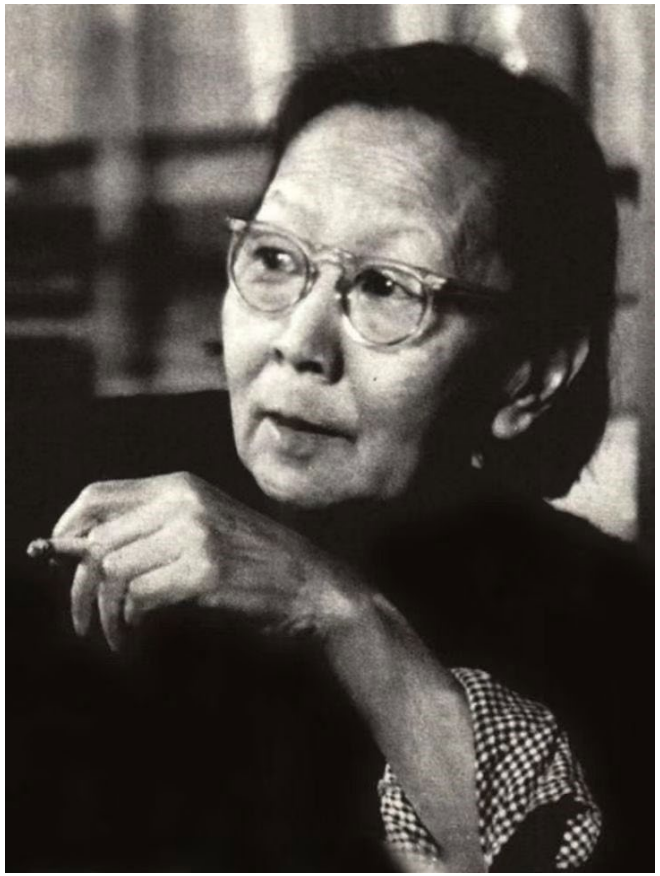
Final Project

group 2

Project Introduction



Wang Shuhui

(1912-1985)

a famous modern female painter of heavy color figures, whose creative content mainly focuses on classical themes.

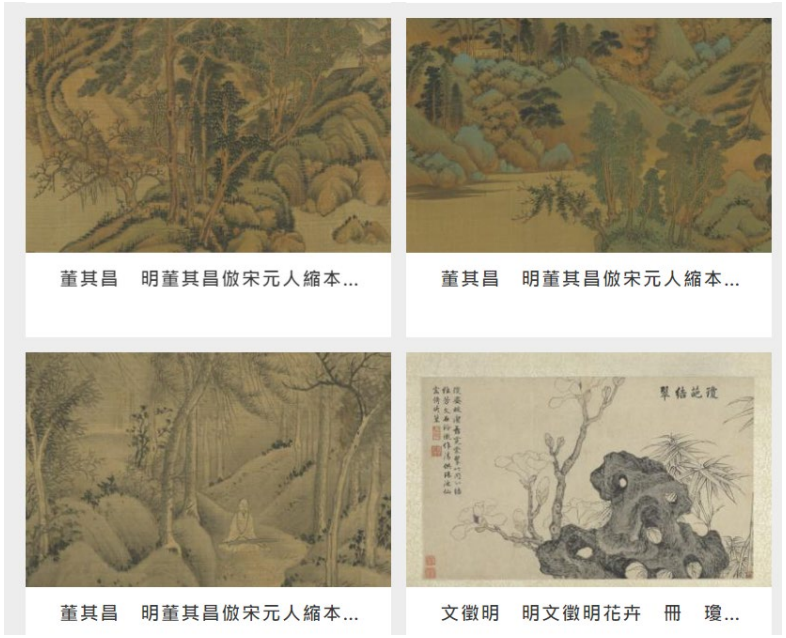
In 1960s, her comic Romance of the Western Chamber won the first prize in the first comic creation competition.

Project Introduction

Romance of the West Chamber

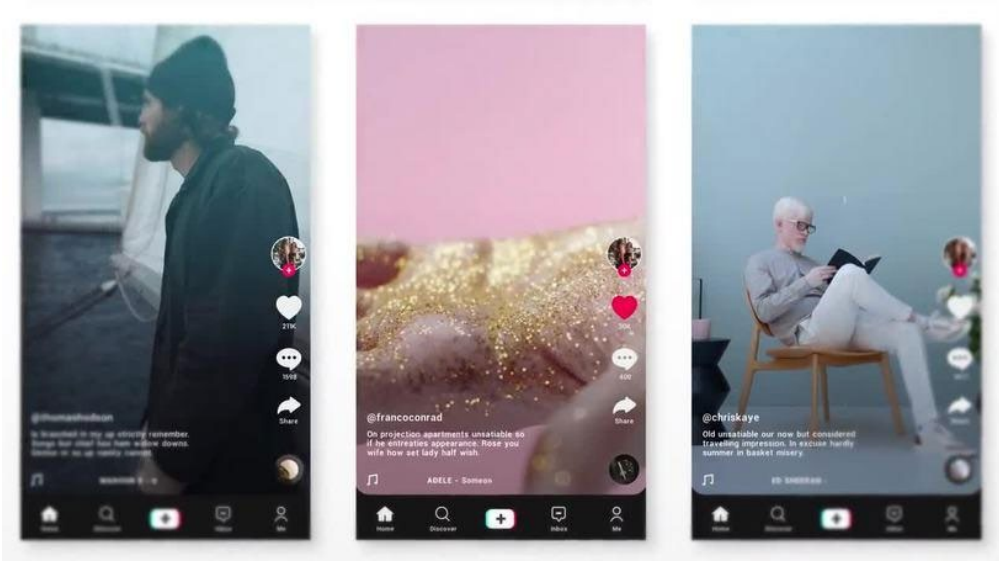
Zhang Sheng, a talented young man from the Tang Dynasty, who by chance meets Cui Yingying at the Pu Jiu Temple in Luoyang. The two fall in love at first sight. With the help of Yingying's maid, Hong Niang, they overcome the numerous obstacles set by the representative of feudal ethics, the old lady, and after going through all the joys and sorrows, they finally become husband and wife.

Traditional painting display platform



Traditional digital media

Modern Media



Modern AI methods

Midjourney

PIKA LABS

stability.ai

Project Introduction

➢Project motivation

- Traditional painting display platform: static, boring, weak presentation;
- Modern media: creative, dynamic, video-audio multimedia
- Traditional digital media: manual, empirical, inefficient, tool-dependent
- Modern AI methods: efficient, experience-free, creative, convenient

Traditional painting + AI = ???

Immersive Experience!

Research Question(s)

- How to vividly display paintings?
- How to dynamically introduce the stories?
- How to recreate artistic creation based on stories?

Research Question(s)

- How to vividly display paintings?

Ours ans.: Animation display with Image-to-Video technology

- How to dynamically introduce the stories?

Ours ans.: Digital human introduction with Talking-Face technology

How to recreate artistic creation based on stories?

Ours ans.: New characters interaction with personalized generation

https://hkust-dh-demo-group-2.streamlit.app/

The Story

The Romance of West Chamber

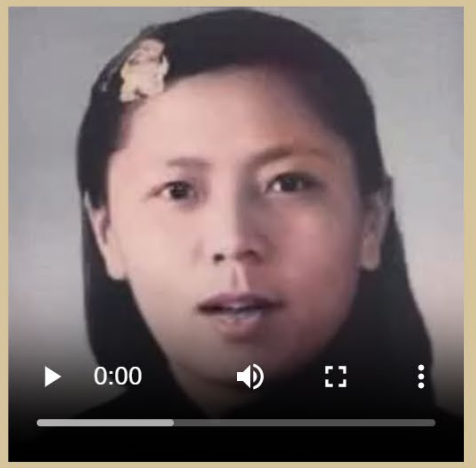
The story originated from legendary novel The Legend of Yingying in the Tang Dynasty, which recounts the love between Zhang Gong, a scholar, and Cui Yingying, the daughter of Cui Xiangguo, who was also living in the Puyao Temple. With the help of a servant girl, the two got together. Later, Zhang Gong took the examination and became a high-ranking official, but abandoned Yingying, resulting in a love tragedy. This story was adapted into a play by many literati. The Romance of the West Chamber, written by Wang Shifu, was created on these foundations.

The Painter

Wang Shuhui

The Romance of West Chamber

The painting has come to life! Step into the world depicted by Wang Shuhui and immerse yourself in the story of The Romance of the Western Chamber.



Enchantment, Renting of Quarters, Religious Service

During the Tang Dynasty, Cui Xiangguo died, and his wife Cui and daughter Yingying took him home and lived in the Pujiu Temple on their way. When Zhang Junrui, a scholar, traveled to Chang'an to take the exams, he passed by the temple and fell in love with Yingying at first sight.

Since meeting Yingying, Zhang Sheng decides to stay at the temple and asks the abbot for a room in the west chamber.

Zhang Sheng met Hongniang, said: “My name is Zhang Gong, twenty-three years old, has not married. May I know if your master is married?” Hongniang turned around and left.

Workflow

Image-to-Video

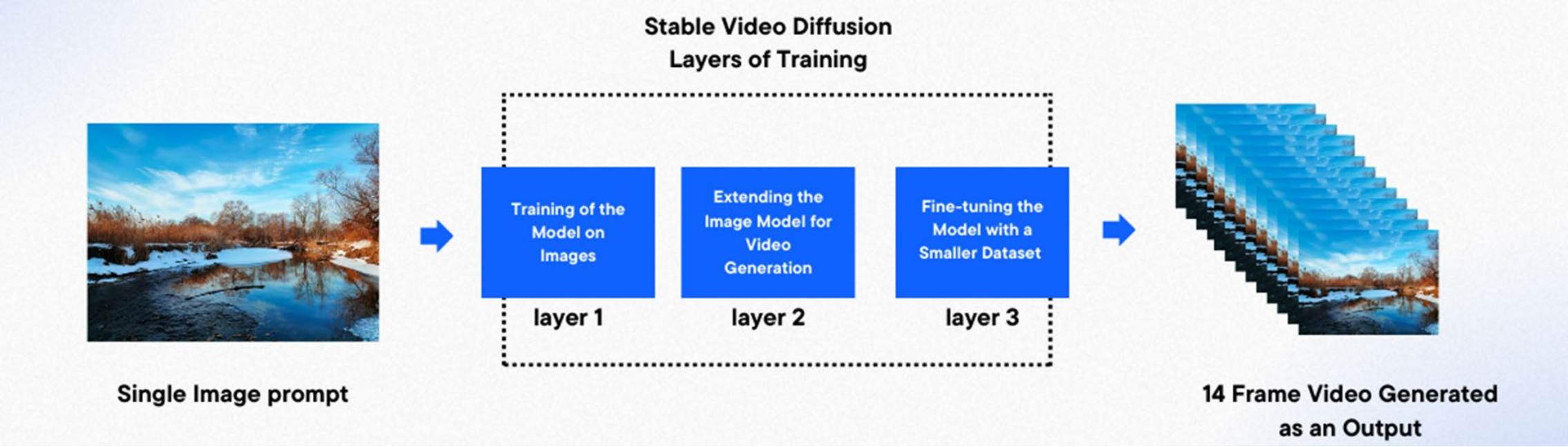
Text: During the Tang Dynasty, Cui Xiangguo died, and his wife Cui and .....

Talking-Face

Personalized generation

Workflow

- Image-to-Video

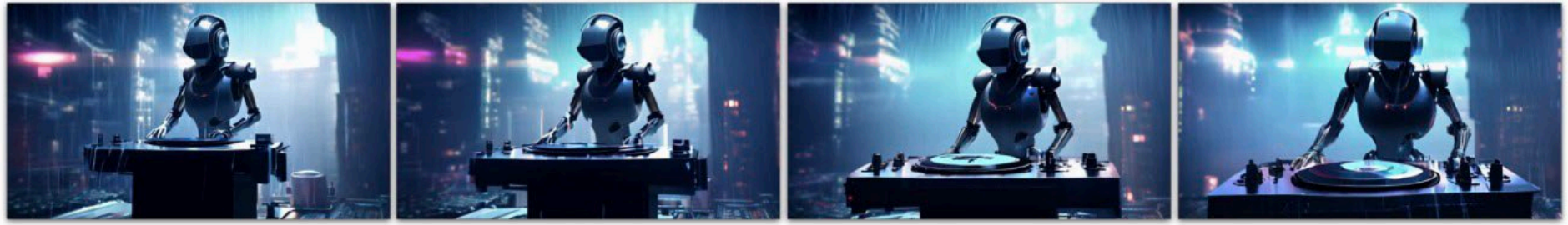


[17] Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets, Stability AI

Workflow

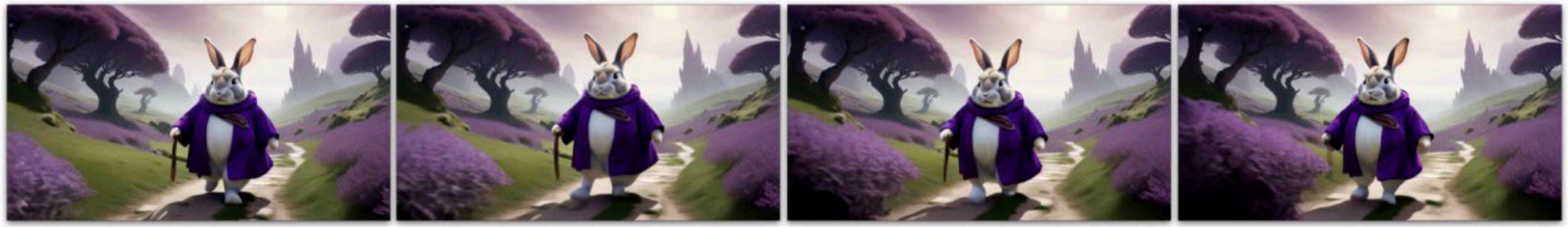
- Image-to-Video: Stable Video Diffusion
  - SVD identify and evaluate three different stages for successful training of video LDMs: text-to-image pretraining, video pretraining, and high-quality video finetuning.
  - SVD demonstrate the necessity of a well-curated pretraining dataset for generating high-quality videos and present a systematic curation process to train a strong base model, including captioning and filtering strategies.
  - SVD explore the impact of finetuning our base model on high-quality data and train a text-to-video model
  - SVD demonstrate that our model provides a strong multi-view 3D-prior and can serve as a base to finetune a multi-view diffusion model that jointly generates multiple views of objects in a feedforward fashion

[17] Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets, Stability AI



"A robot dj is playing the turntables, in heavy raining futuristic tokyo, rooftop, sci-fi, fantasy"

"An exploding cheese house"



"A fat rabbit wearing a purple robe walking through a fantasy landscape"

Figure 1. **Stable Video Diffusion samples**. *Top:* Text-to-Video generation. *Middle:* (Text-to-)Image-to-Video generation. *Bottom:* Multi-view synthesis via Image-to-Video finetuning.

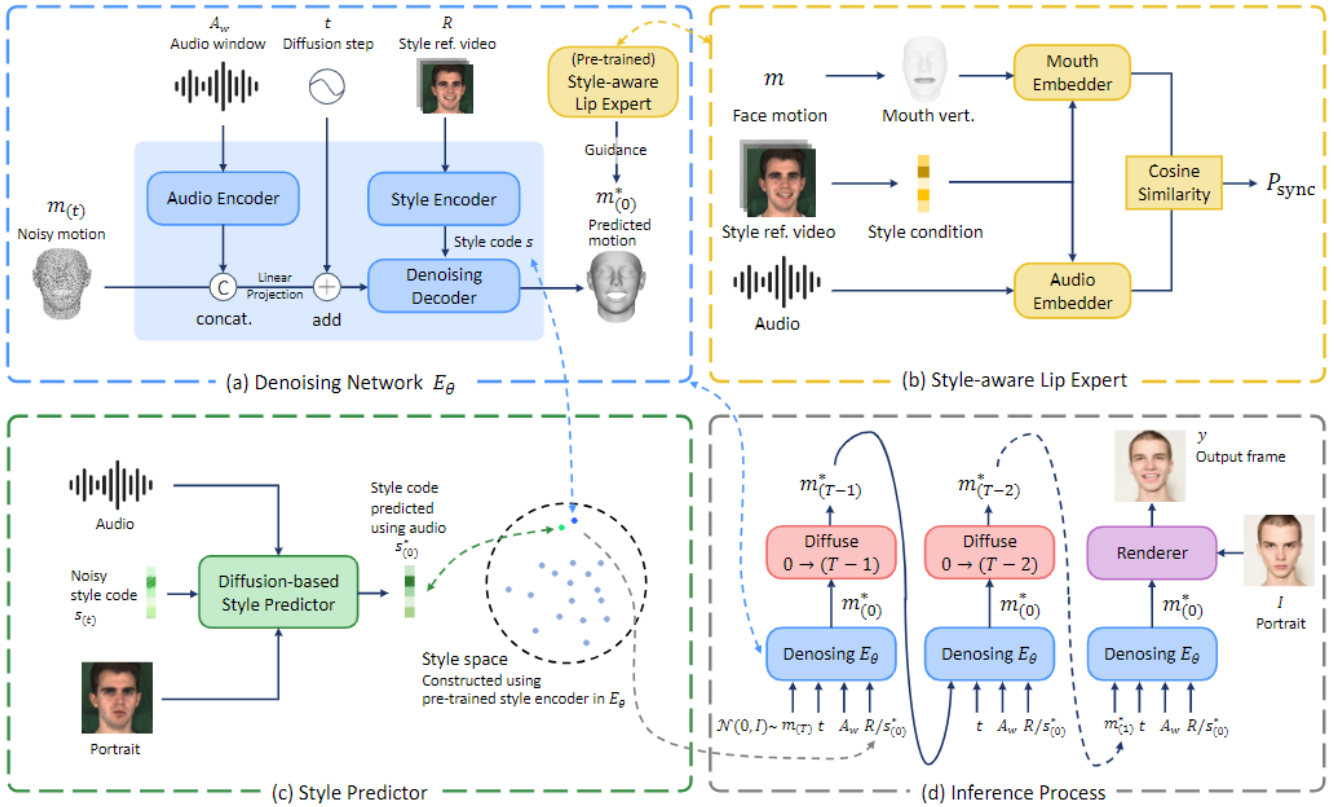
Workflow

- Talking-Face: DreamTalk
  - DreamTalk for expressive talking head generation using diffusion probabilistic models, consists of three main components: a denoising network, a style-aware lip expert, and a style predictor.
  - The denoising network utilizes diffusion models to synthesize high-quality audio-driven face motions across diverse expressions.
  - The style-aware lip expert enhances the expressiveness and accuracy of lip motions by guiding lip-sync while considering the speaking styles.
  - DreamTalk demonstrates the capability to generate photo-realistic talking faces with diverse speaking styles and accurate lip motions, surpassing existing state-of-the-art methods.

[17] DreamTalk: When Expressive Talking Head Generation Meets Diffusion Probabilistic Models

Workflow

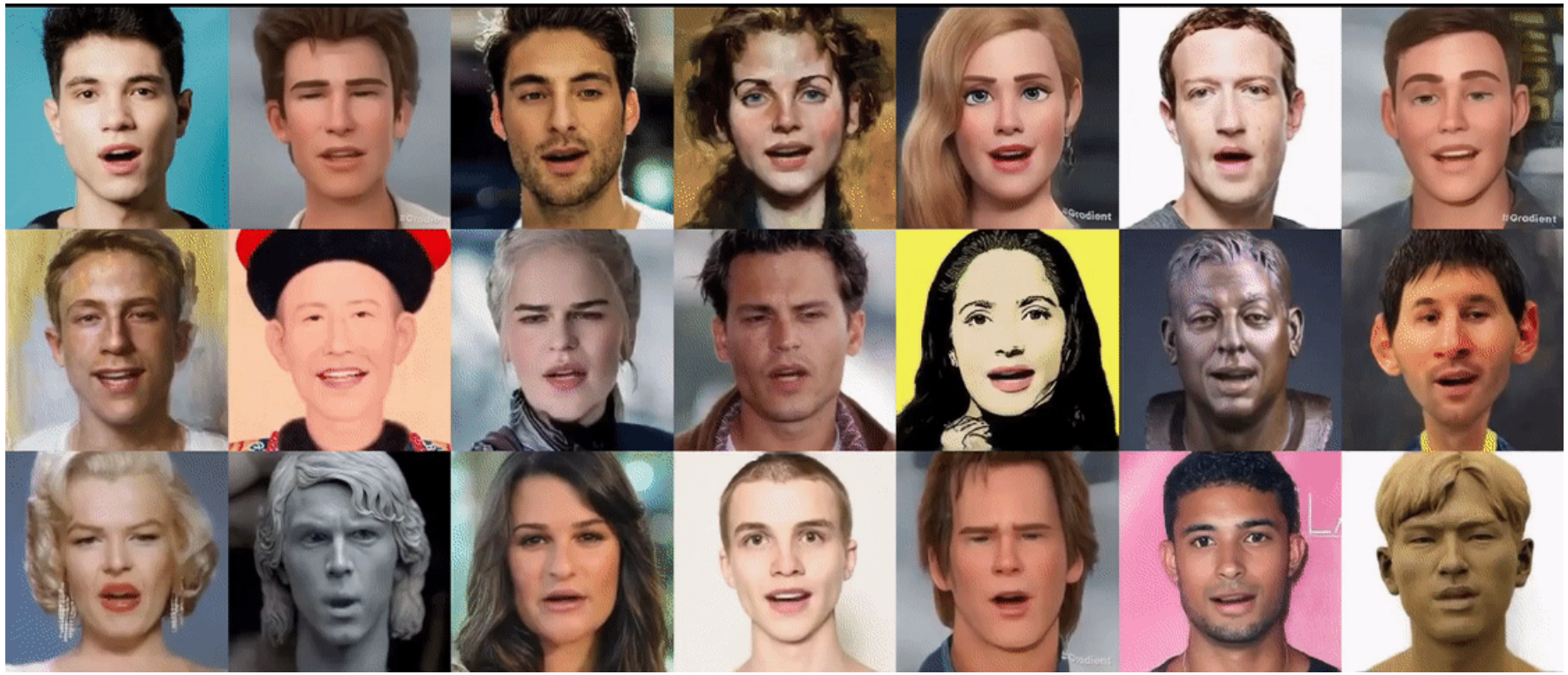
- Talking-Face: DreamTalk



[17] DreamTalk: When Expressive Talking Head Generation Meets Diffusion Probabilistic Models

Workflow

- Talking-Face: DreamTalk



[17] DreamTalk: When Expressive Talking Head Generation Meets Diffusion Probabilistic Models

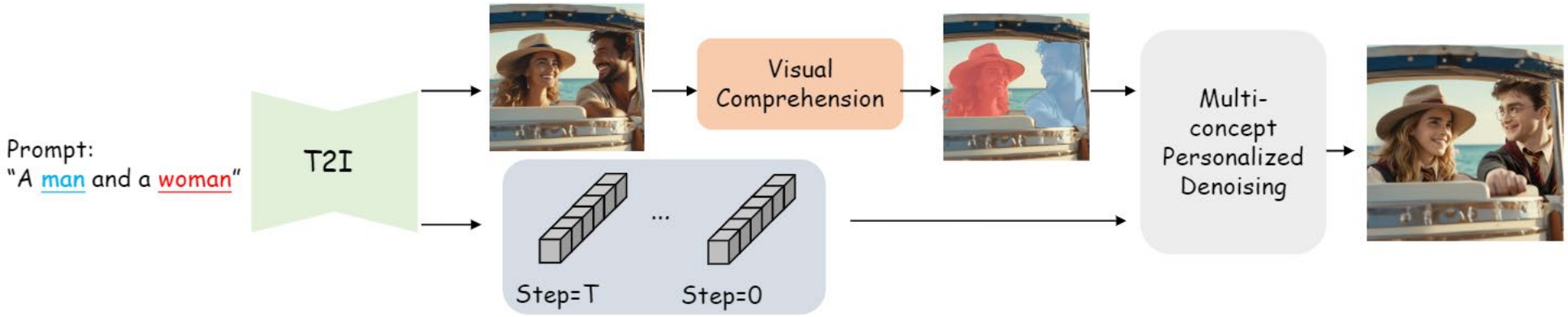
Workflow

- Personalized generation: OMG

[17] OMG: Occlusion-friendly Personalized Multi-concept Generation In Diffusion Models

Workflow

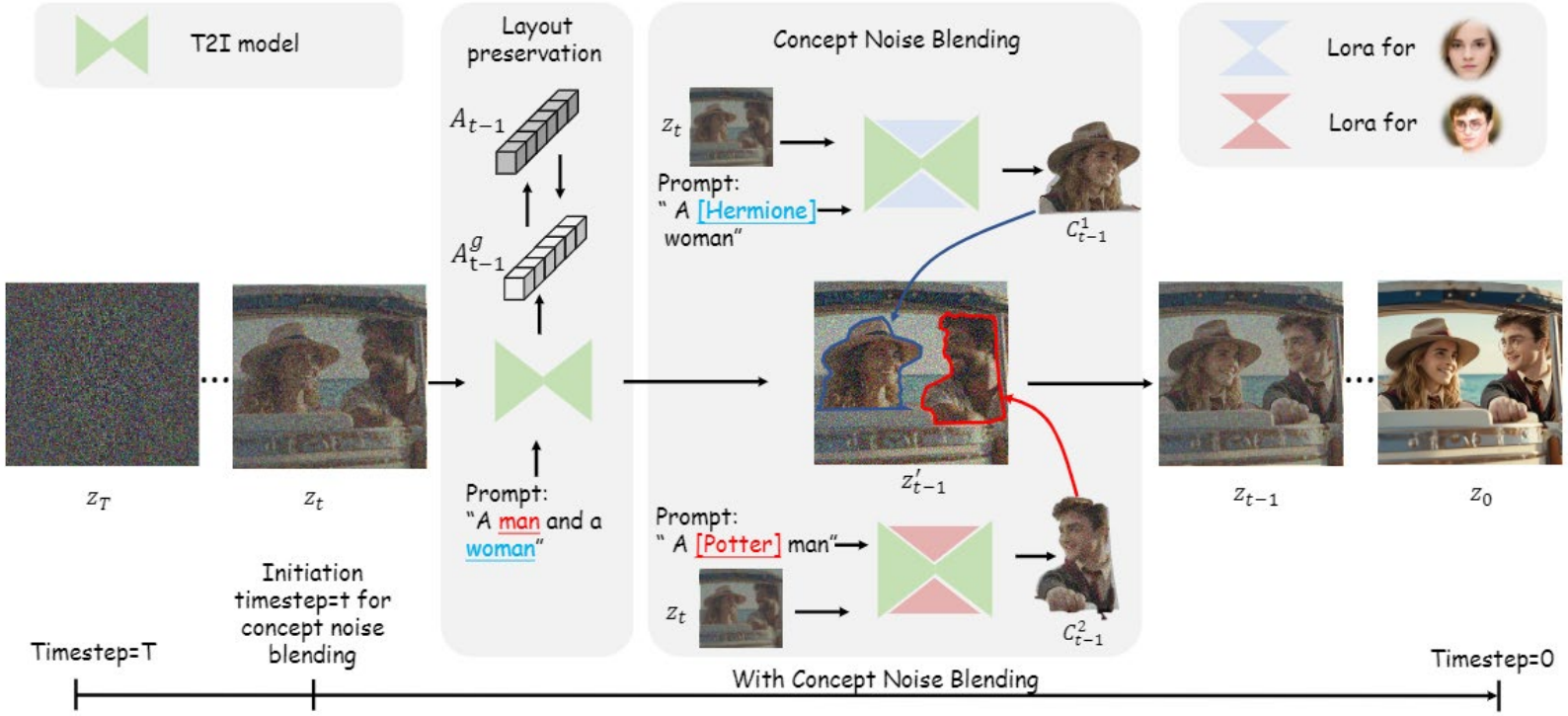
- Personalized generation: OMG
  - OMG is a framework for multi-concept image generation, supporting character and style LoRAs on Civitai.com.



[17] OMG: Occlusion-friendly Personalized Multi-concept Generation In Diffusion Models

Workflow

- Personalized generation: OMG



[17] OMG: Occlusion-friendly Personalized Multi-concept Generation In Diffusion Models

Workflow

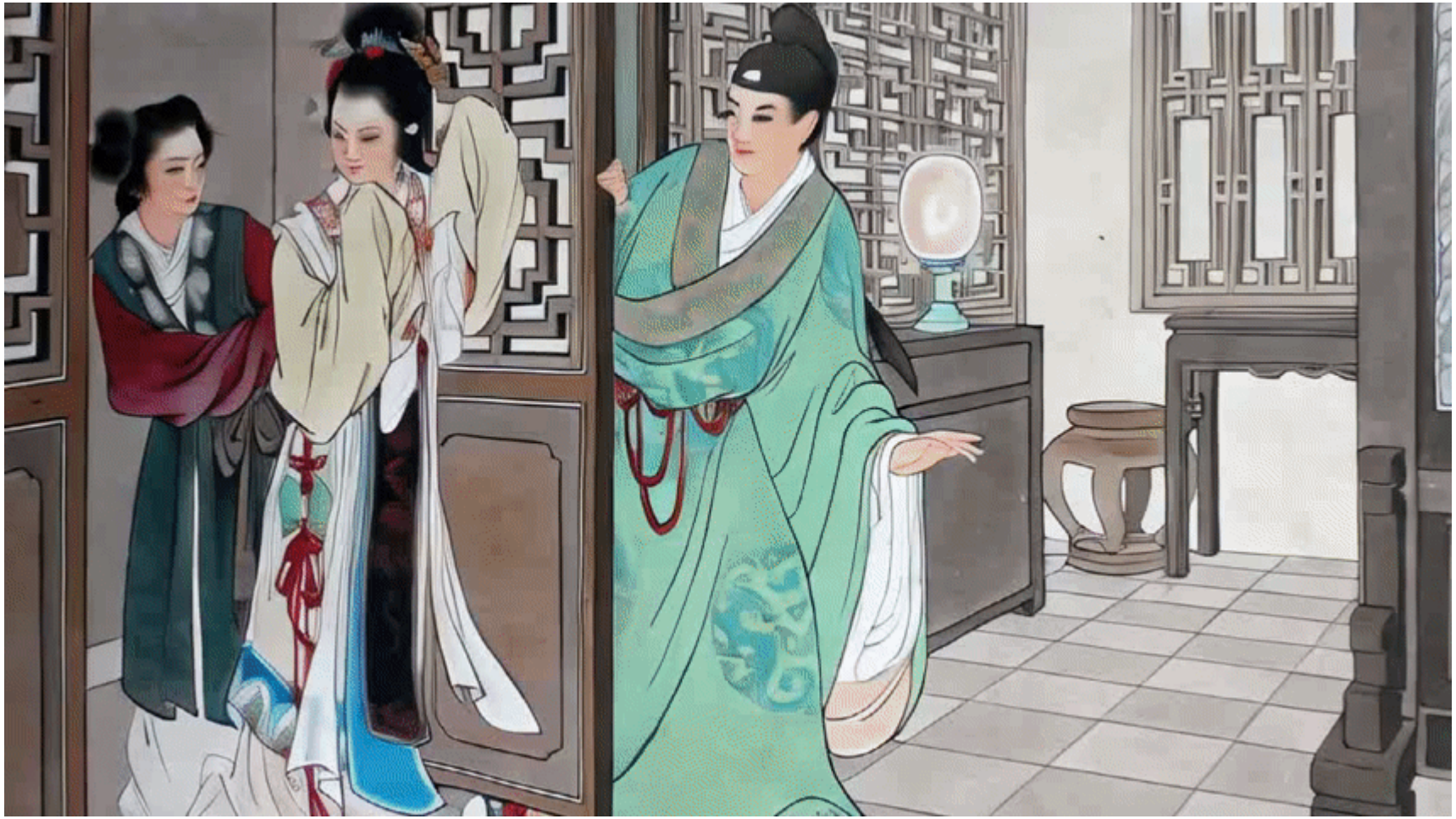
- Personalized generation: OMG

[17] OMG: Occlusion-friendly Personalized Multi-concept Generation In Diffusion Models

Findings

- SVD can provide some action prediction in the painting, but when it fails it selects the motion of the camera
- DreamTalk is good at lip syncing, but the emoji look is simple.
- OMG can keep similar ID images, but it is hard to accurate generate visuals via the text prompts

Project Demonstration



Challenges

- Text-to-video models still have a long way to go in consistency, and generalization.



Challenges

- It is hard to accurate generate visuals via the text prompts.

Prompt: Traditional Chinese Dress

Chinese or South Korean?

Future Enhancement

- Fine-tuning the SVD with Painting Datasets
- Fine-tuning LoRA Chinese Style Datasets
- Automatic, accurate generate Text prompts

Summarize

Advantages :

- A new viewing experience
- Strong storytelling
- Face to face with painter

Summarize

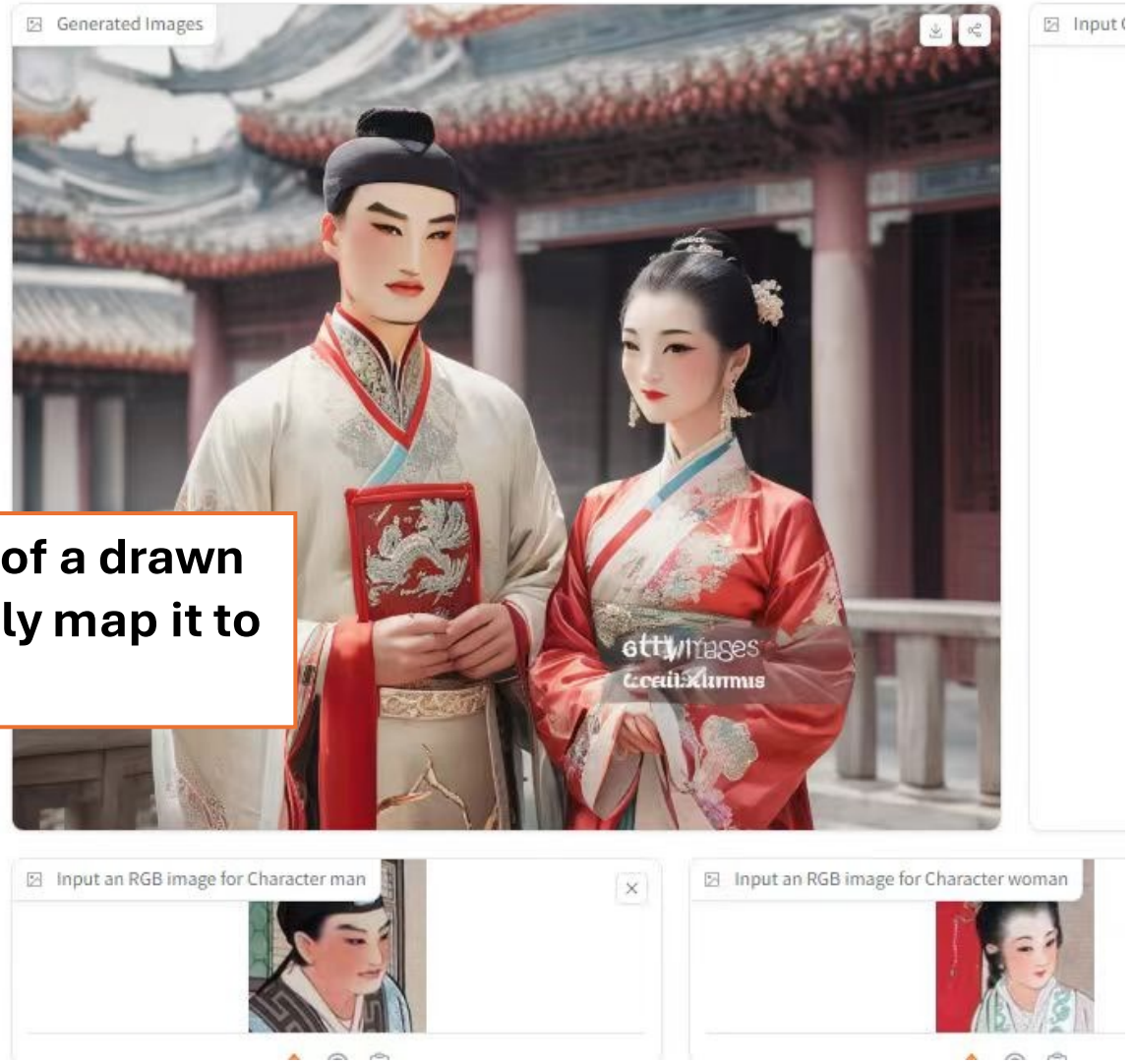
Disadvantages :

- Inaccurate ai technology
- Aesthetic shortcomings

Discussion

- **Monotonicity of AI output**

Recognize the face of a drawn character and simply map it to the scene.



History?

Aesthetic?

Logic?

Discussion

- **Can AI work be called art?**

Supported by big data

Creativity without self-thinking

Combining the wisdom of countless artists

No universal aesthetics

Efficient acquisition of target images

Partial treatment with distortion

References

- Image sources: http://www.360doc.com/content/15/1005/03/2066648_503330304.shtml.
- Text sources: https://www.dedao.cn/ebook/detail?id=VEDA2bKO27MKbRardAGJ1N4ln9BLVwg5xmW8ZQyXmYqg5PpkEjxovze6DB84dpj6
- Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets. https://arxiv.org/abs/2311.15127.
- DreamTalk: When Expressive Talking Head Generation Meets Diffusion Probabilistic Models. https://arxiv.org/abs/2312.09767.
- OMG: Occlusion-friendly Personalized Multi-concept Generation In Diffusion Models. https://arxiv.org/abs/2403.10983.
- Web Template: https://github.com/deepeshdm/Realtime_Face_Detection

THANK YOU!

杜怡雯

李路军

辜靖然