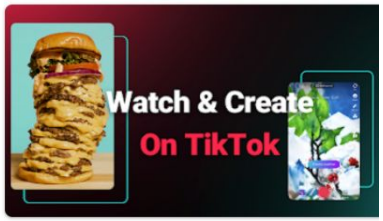
CS577 - Final Project

Explore the Application of Diffusion Model in Video Content Creation

Haiyun Xiao / Yu Li / Haichen Pang

Why we choose this area

In recent years, short videos on social media have become ubiquitous and extremely popular



TikTok
TikTok Pte. Ltd.
4.3 ★

YouTube
Google LLC
4.1 ★

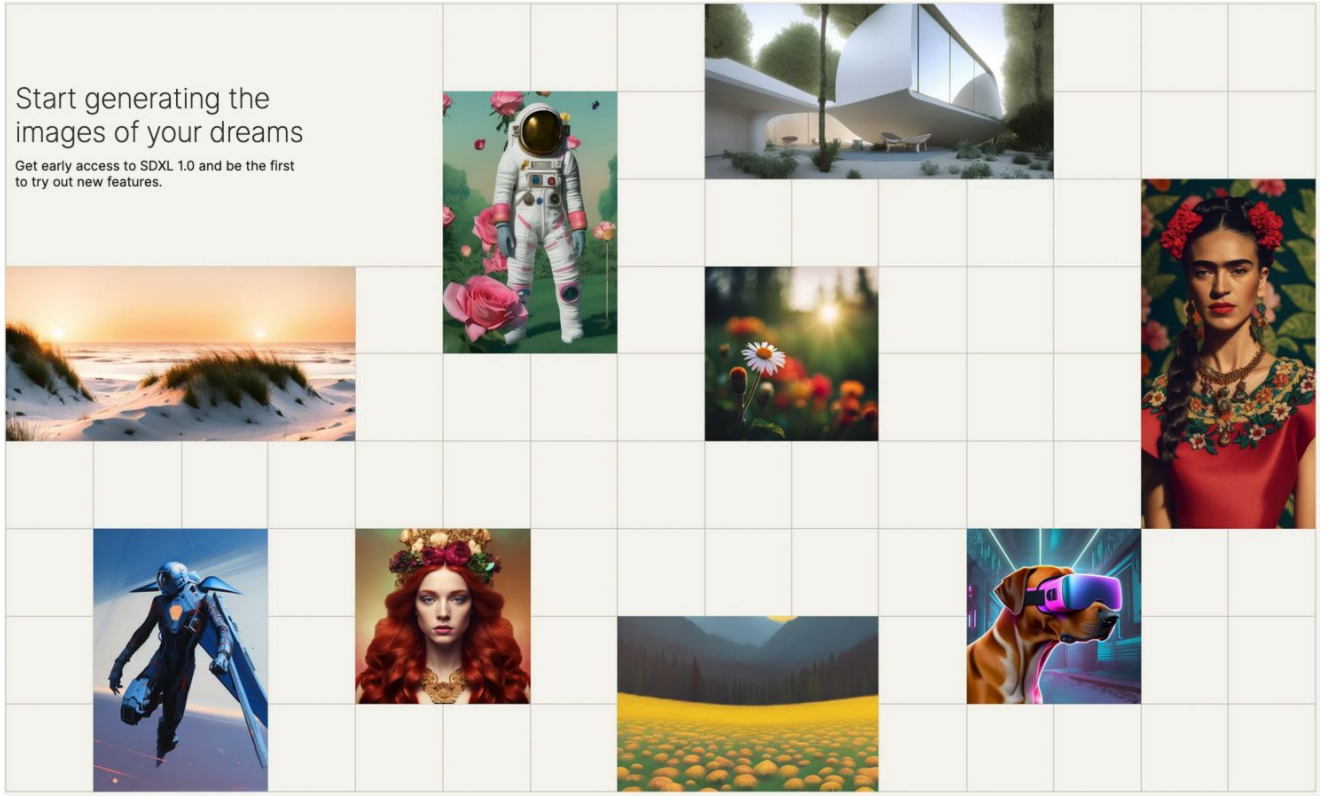
X
X Corp.
3.9 ★

Instagram
Instagram
4.2 ★

Is it possible to automatically generate videos?

https://play.google.com/store/apps?hl=en&gl=US

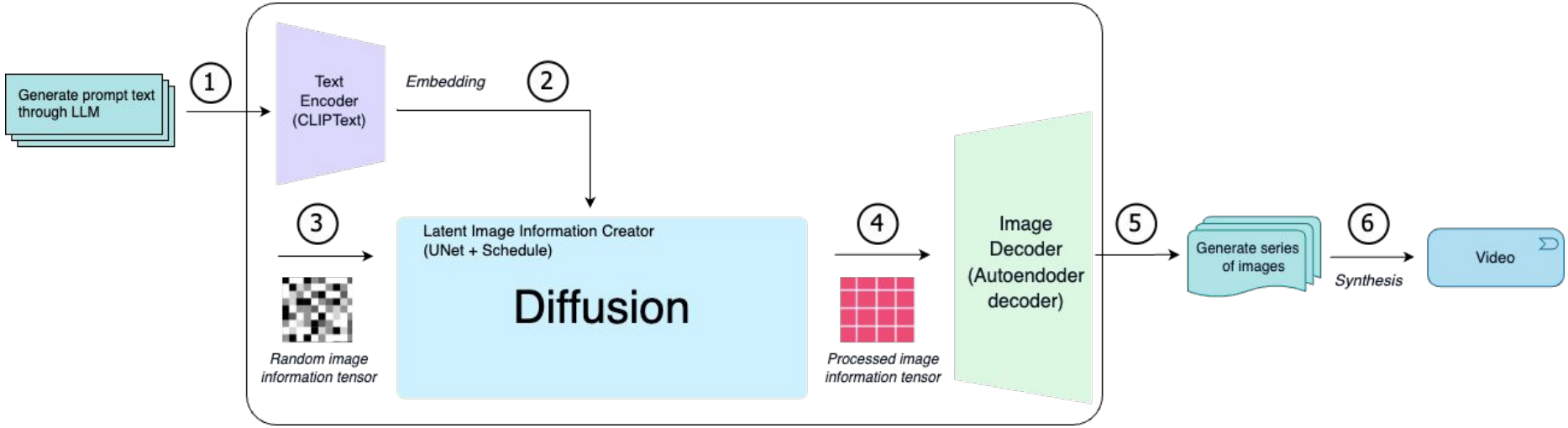
Inspiration from Stable Diffusion



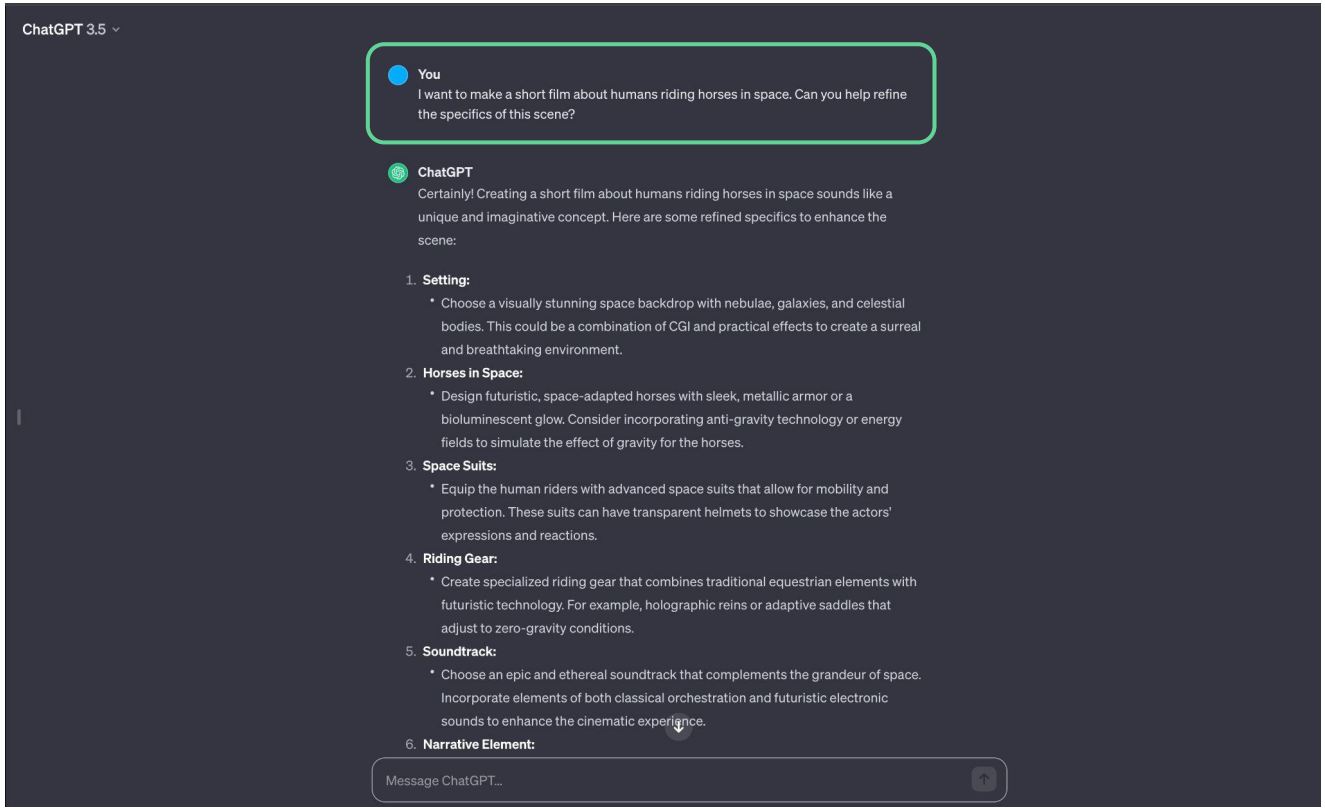
Is it possible to generate videos in the same way?

https://dreamstudio.ai/

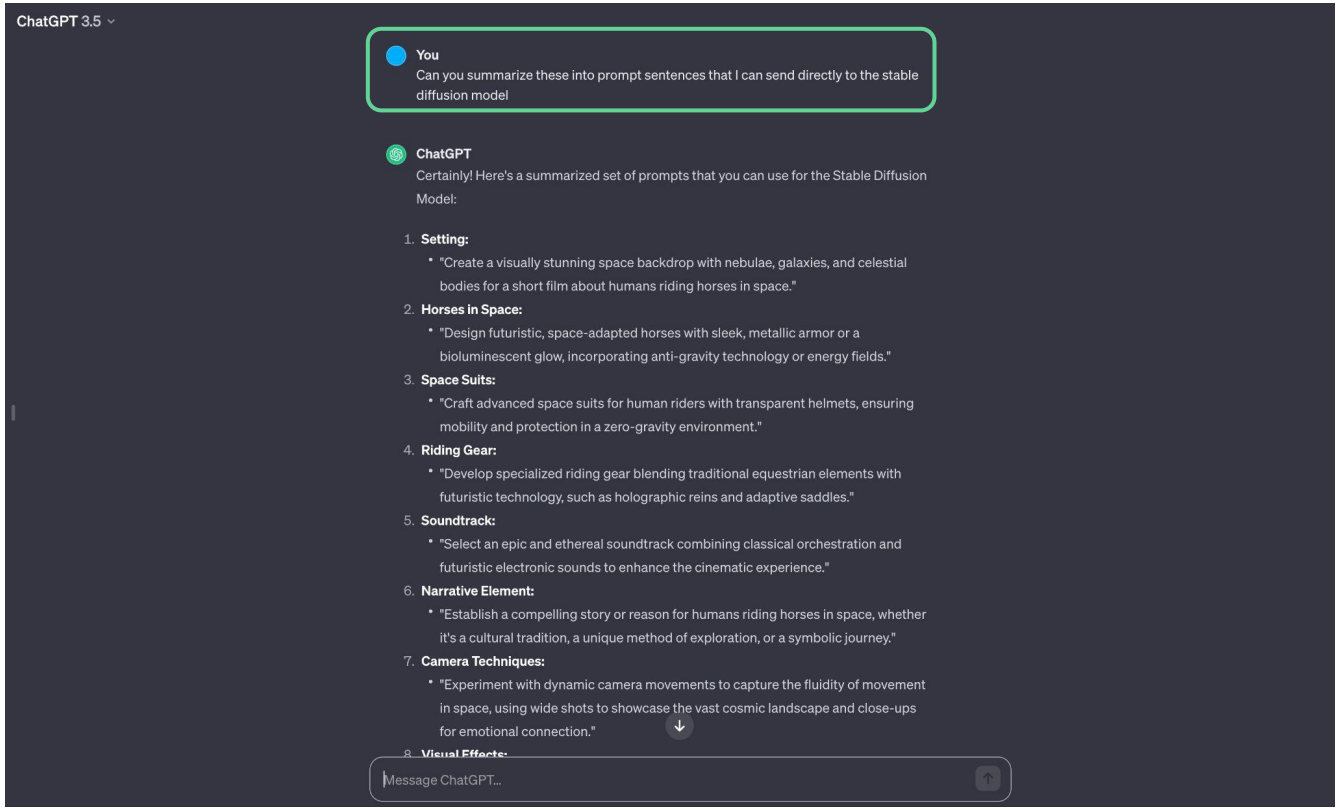
Pipeline


Generate prompt text through LLM
1
Text Encoder (CLIPText)
Embedding
2
3
Random image information tensor
Latent Image Information Creator (UNet + Schedule)
Diffusion
4
Processed image information tensor
Image Decoder (Autoendoder decoder)
5
Generate series of images
6
Synthesis
Video

Step 1 - Use ChatGPT to refine the video scene



Step 1 - Summarize and output the prompt words



Step 2
Generate Pictures

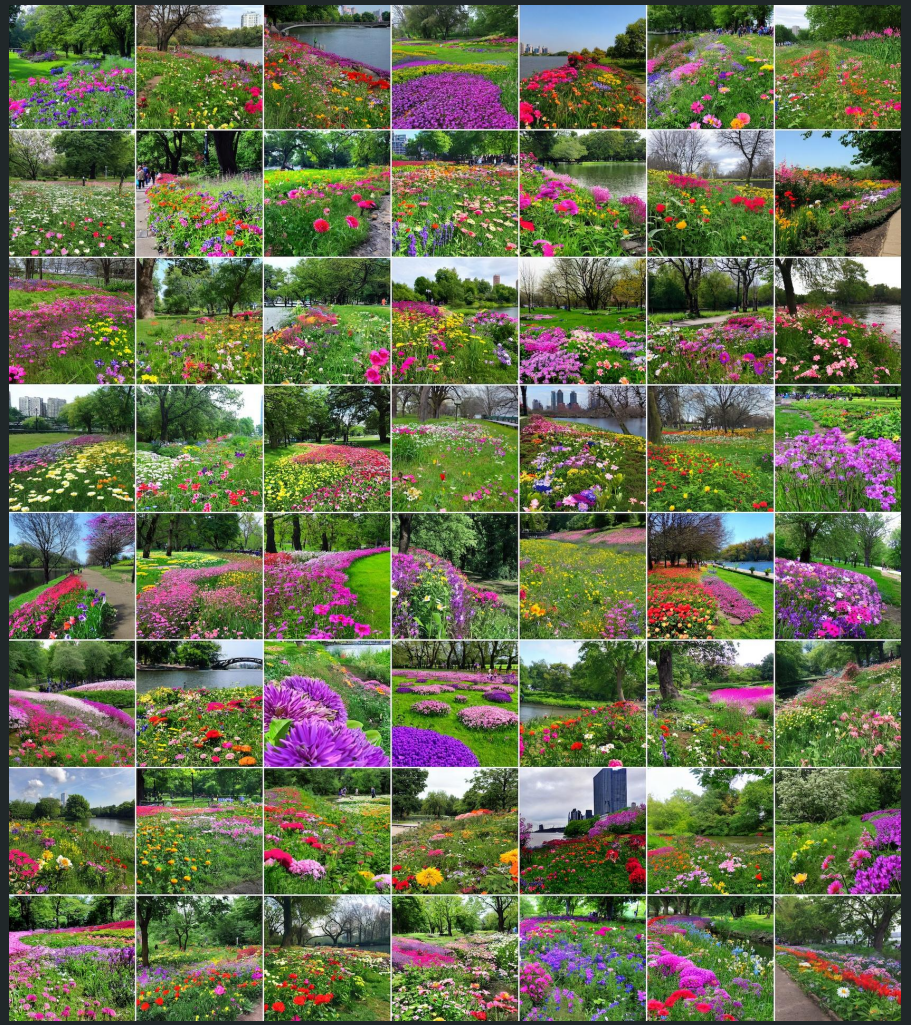
Experiment 1

Prompt words: "a photograph of an astronaut riding a horse"

Step 2
Generate Pictures

Experiment 2

Prompt words: "Many flowers growing in the riverside park"



Step 2
Generate Pictures

Experiment 3

Prompt words: "A very exciting and fun speed racing game"

Step 3 - Synthesize pictures from experiment 1 into video

Step 3 - Synthesize pictures from experiment 2 into video

Step 3 - Synthesize pictures from experiment 3 into video

Conclusion

- The experimental process confirmed the **feasibility** of the pipeline
- It is **not** easy to **control** the generation process and **fine-tune** it
- The **coherence** and **naturalness** of the generated video are **not** enough
- **Suitable** for generating **comics-like** videos that do not require high frame rates
- Next we want to investigate whether we can generate videos directly from the diffusion model

https://github.com/byxhy/Explore-Stable-Diffusion