

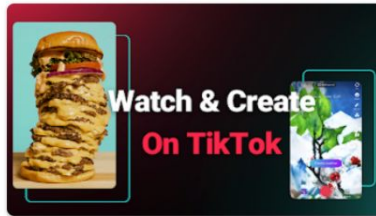
CS577 - Final Project


Explore the Application of Diffusion Model in
Video Content Creation

Haiyun Xiao / Yu Li / Haichen Pang

Why we choose this area

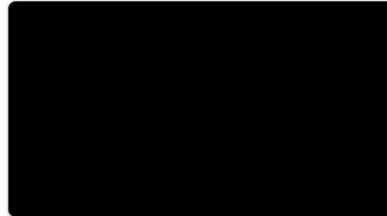
In recent years, short videos on social media have become ubiquitous and extremely popular



 TikTok
TikTok Pte. Ltd.
4.3 ★



 YouTube
Google LLC
4.1 ★



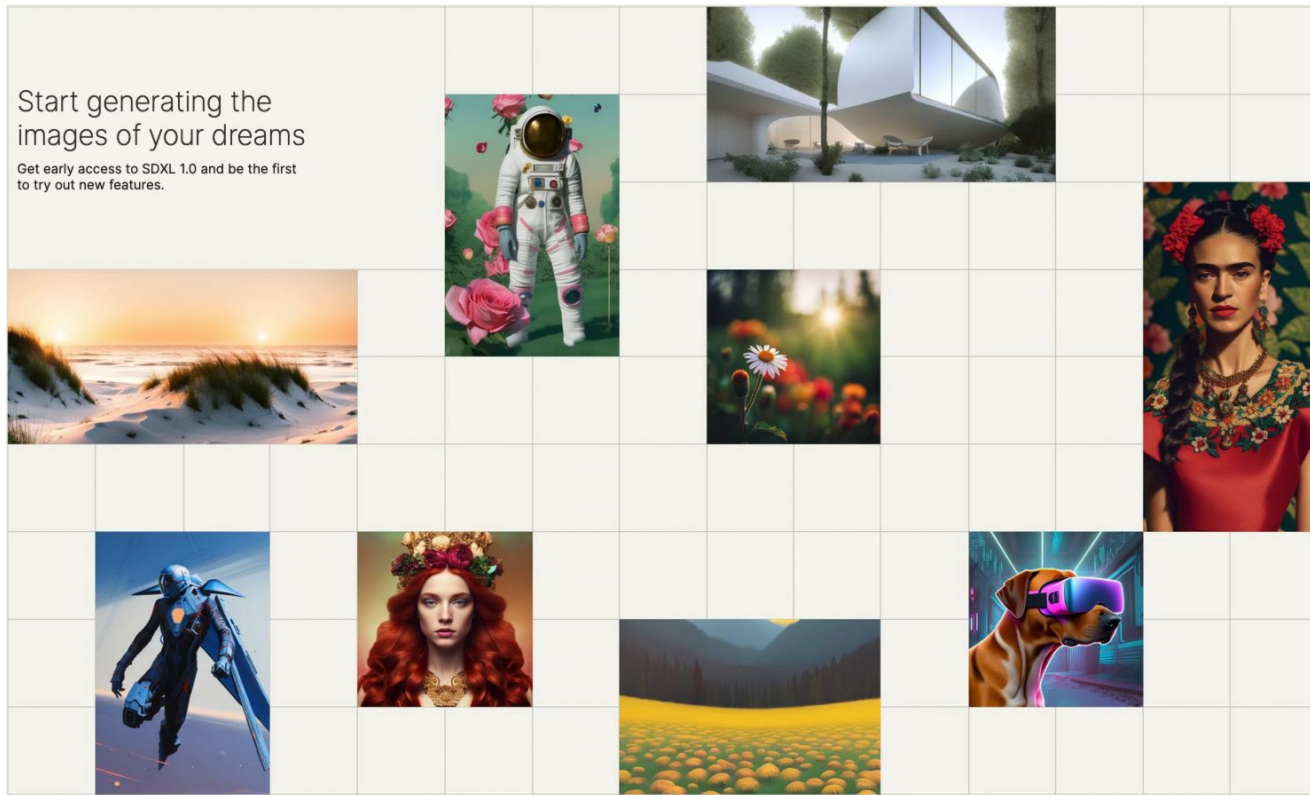
 X
X Corp.
3.9 ★



 Instagram
Instagram
4.2 ★

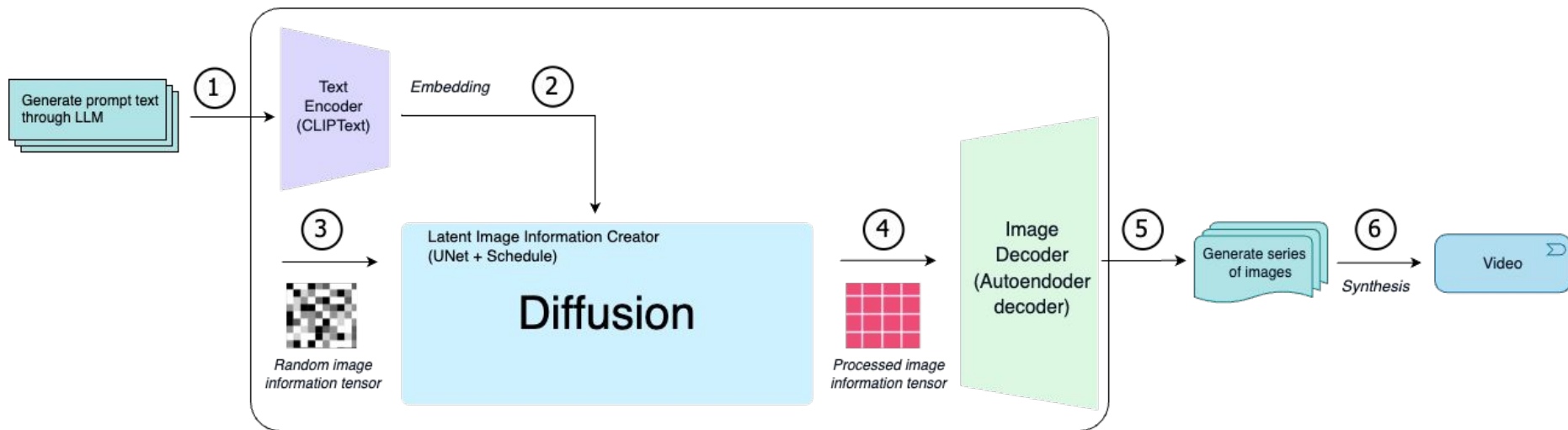
Is it possible to automatically generate videos?

Inspiration from Stable Diffusion

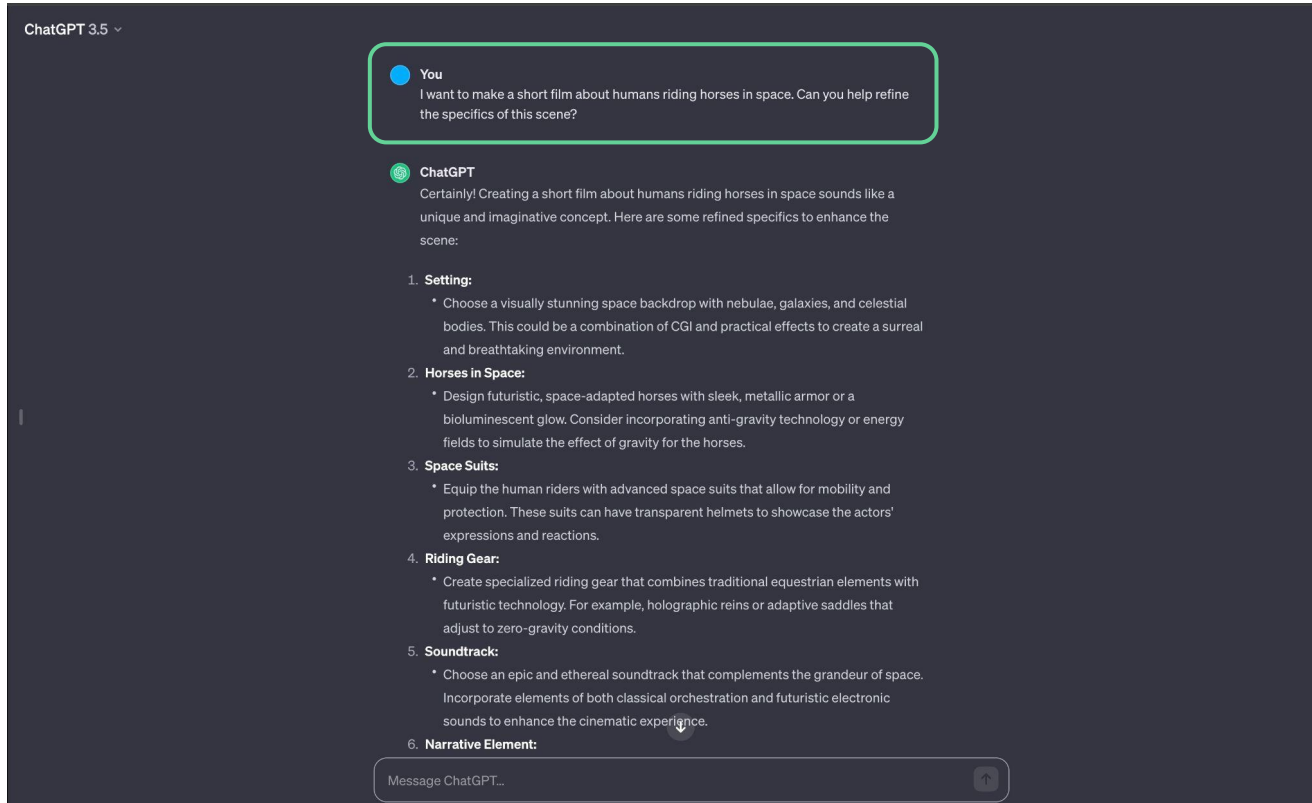


Is it possible to generate videos in the same way?

Pipeline



Step 1 - Use ChatGPT to refine the video scene



Step 1 - Summarize and output the prompt words

ChatGPT 3.5 ▾

You
Can you summarize these into prompt sentences that I can send directly to the stable diffusion model

ChatGPT
Certainly! Here's a summarized set of prompts that you can use for the Stable Diffusion Model:

- Setting:**
 - "Create a visually stunning space backdrop with nebulae, galaxies, and celestial bodies for a short film about humans riding horses in space."
- Horses in Space:**
 - "Design futuristic, space-adapted horses with sleek, metallic armor or a bioluminescent glow, incorporating anti-gravity technology or energy fields."
- Space Suits:**
 - "Craft advanced space suits for human riders with transparent helmets, ensuring mobility and protection in a zero-gravity environment."
- Riding Gear:**
 - "Develop specialized riding gear blending traditional equestrian elements with futuristic technology, such as holographic reins and adaptive saddles."
- Soundtrack:**
 - "Select an epic and ethereal soundtrack combining classical orchestration and futuristic electronic sounds to enhance the cinematic experience."
- Narrative Element:**
 - "Establish a compelling story or reason for humans riding horses in space, whether it's a cultural tradition, a unique method of exploration, or a symbolic journey."
- Camera Techniques:**
 - "Experiment with dynamic camera movements to capture the fluidity of movement in space, using wide shots to showcase the vast cosmic landscape and close-ups for emotional connection."

Visual Effects:

Message ChatGPT...

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



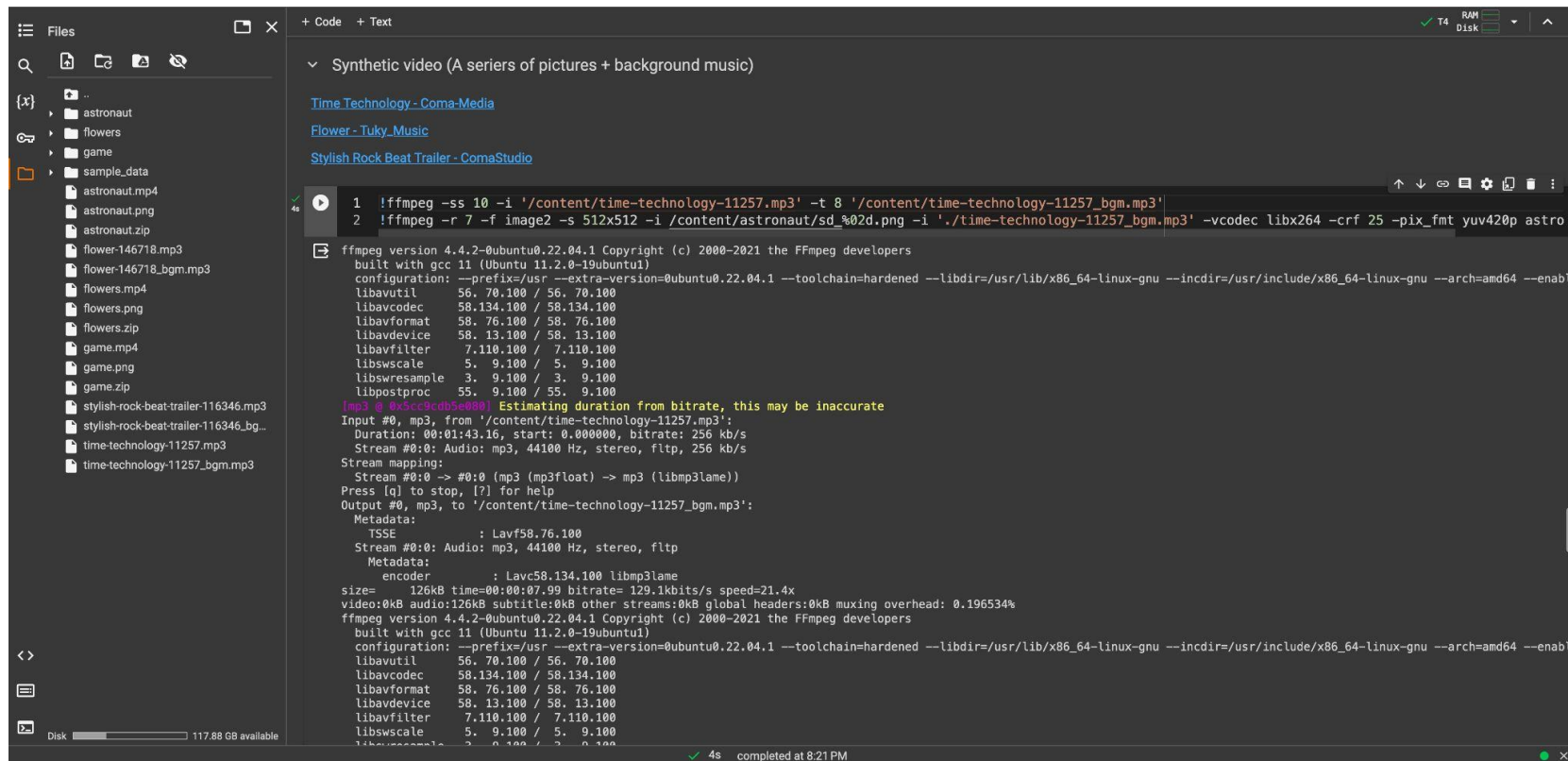
Step 3 - Synthesize pictures from **experiment 3** into video



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>



The screenshot shows a terminal window with a file explorer on the left and a code editor on the right. The file explorer displays a directory structure with files like 'astronaut.png', 'flowers.png', 'game.png', and 'time-technology-11257_bgm.mp3'. The code editor shows a terminal session where a synthetic video is being created using FFmpeg. The command executed is: `1 !ffmpeg -ss 10 -i '/content/time-technology-11257.mp3' -t 8 '/content/time-technology-11257_bgm.mp3'` and `2 !ffmpeg -r 7 -f image2 -s 512x512 -i /content/astroaut/sd_%02d.png -i '/time-technology-11257_bgm.mp3' -vcodec libx264 -crf 25 -pix_fmt yuv420p astro`. The output shows the FFmpeg version (4.4.2-0ubuntu0.22.04.1) and configuration details. The video is being processed with a duration of 00:01:43.16 and a bitrate of 256 kb/s. The output file is 'astro'.

```
Files
{X}
  astronaut
  flowers
  game
  sample_data
  astronaut.mp4
  astronaut.png
  astronaut.zip
  flower-146718.mp3
  flower-146718_bgm.mp3
  flowers.mp4
  flowers.png
  flowers.zip
  game.mp4
  game.png
  game.zip
  stylish-rock-beat-trailer-116346.mp3
  stylish-rock-beat-trailer-116346_bgm.mp3
  time-technology-11257.mp3
  time-technology-11257_bgm.mp3

+ Code + Text
Synthetic video (A series of pictures + background music)

Time Technology - Coma-Media
Flower - Tuky_Music
Stylish Rock Beat Trailer - ComaStudio

1 !ffmpeg -ss 10 -i '/content/time-technology-11257.mp3' -t 8 '/content/time-technology-11257_bgm.mp3'
2 !ffmpeg -r 7 -f image2 -s 512x512 -i /content/astroaut/sd_%02d.png -i '/time-technology-11257_bgm.mp3' -vcodec libx264 -crf 25 -pix_fmt yuv420p astro

ffmpeg version 4.4.2-0ubuntu0.22.04.1 Copyright (c) 2000-2021 the FFmpeg developers
built with gcc 11 (Ubuntu 11.2.0-19ubuntu1)
configuration: --prefix=/usr --extra-version=0ubuntu0.22.04.1 --toolchain=hardened --libdir=/usr/lib/x86_64-linux-gnu --incdir=/usr/include/x86_64-linux-gnu --arch=amd64 --enabl
libavutil 56. 70.100 / 56. 70.100
libavcodec 58.134.100 / 58.134.100
libavformat 58. 76.100 / 58. 76.100
libavdevice 58. 13.100 / 58. 13.100
libavfilter 7.110.100 / 7.110.100
libswscale 5. 9.100 / 5. 9.100
libswresample 3. 9.100 / 3. 9.100
libpostproc 55. 9.100 / 55. 9.100
[mp3 @ 0x5cc9cbb5e000] Estimating duration from bitrate, this may be inaccurate
Input #0, mp3, from '/content/time-technology-11257.mp3':
Duration: 00:01:43.16, start: 0.000000, bitrate: 256 kb/s
Stream #0:0: Audio: mp3, 44100 Hz, stereo, fltp, 256 kb/s
Stream mapping:
Stream #0:0 -> #0:0 (mp3 (mp3float) -> mp3 (libmp3lame))
Press [q] to stop, [?] for help
Output #0, mp3, to '/content/time-technology-11257_bgm.mp3':
Metadata:
TSSE : Lavf58.76.100
Stream #0:0: Audio: mp3, 44100 Hz, stereo, fltp
Metadata:
encoder : Lavc58.134.100 libmp3lame
size= 126kB time=00:00:07.09 bitrate= 129.1kbits/s speed=21.4x
video:0kB audio:126kB subtitle:0kB other streams:0kB global headers:0kB muxing overhead: 0.196534%
ffmpeg version 4.4.2-0ubuntu0.22.04.1 Copyright (c) 2000-2021 the FFmpeg developers
built with gcc 11 (Ubuntu 11.2.0-19ubuntu1)
configuration: --prefix=/usr --extra-version=0ubuntu0.22.04.1 --toolchain=hardened --libdir=/usr/lib/x86_64-linux-gnu --incdir=/usr/include/x86_64-linux-gnu --arch=amd64 --enabl
libavutil 56. 70.100 / 56. 70.100
libavcodec 58.134.100 / 58.134.100
libavformat 58. 76.100 / 58. 76.100
libavdevice 58. 13.100 / 58. 13.100
libavfilter 7.110.100 / 7.110.100
libswscale 5. 9.100 / 5. 9.100
libswresample 3. 9.100 / 3. 9.100
libpostproc 55. 9.100 / 55. 9.100

4s completed at 8:21 PM
```