GIF Generation Project

By Rishi, Tri, and Ronel

Objective

Why GIF?

- Creating customized GIFs manually is time-consuming and requires artistic skills for many people.
- The Automated GIF generation can democratize creative content production and provide a valuable tool for marketing, social media, and entertainment industries for any person.
- It also introduces a more unique way to express yourself in a customizable way.

Background

Primary Dataset: https://github.com/ali-vilab/VGen/tree/main/data

- Consists of a bunch of videos and pictures
- Converts the images to RGB

Additional Data:

- Open Source Images

Models / Methodology

- AnimateDiffPipeline: Realistic_Vision_V5.1_noVAE/animatediff-motion-adapter-v1-5-2
- I2VGenXLPipeline
- runwayml/stable-diffusion-v1-5
- animatediff-motion-adapter-v1-5-2

Approach

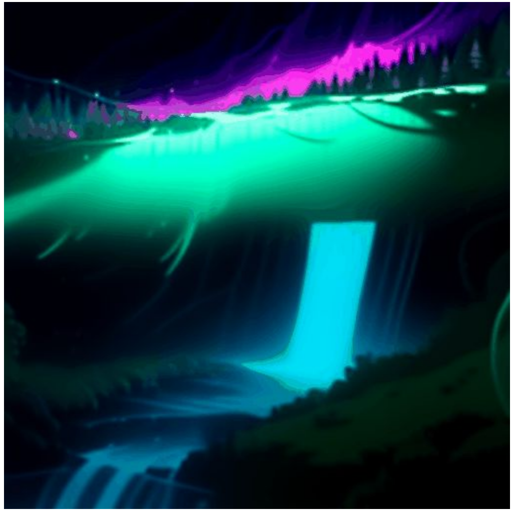
- Utilize diffusion models for generating animated content, specifically focusing on adapting and fine-tuning existing models to handle conditional inputs effectively.
- Employ conditional input strategies, such as text prompts, to guide the generation process, ensuring the resulting GIFs align with the user's specified themes or styles.
- **Post Processing:** Convert Raw Image output into GIF format

GIF generation: AnimateDiffPipeline

Prompt A: "A sunrise on waves merging with a cascading waterfall in a forest, serene atmosphere, ethereal, high-quality, detailed"

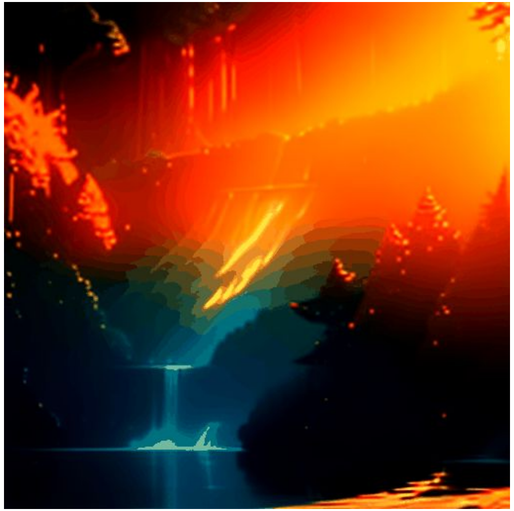
Prompt B: "Waves merging with a cascading waterfall in a forest at night, ethereal, high-quality, detailed, seamless loop"

guidance = 5



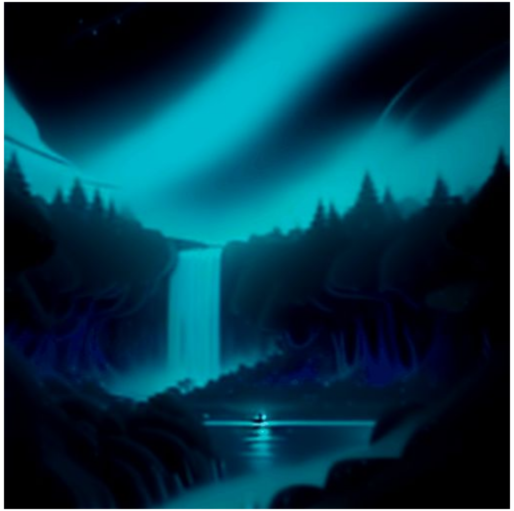
Prompt B

guidance = 8



Prompt A

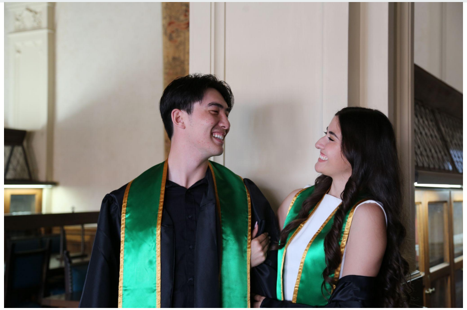
guidance = 12

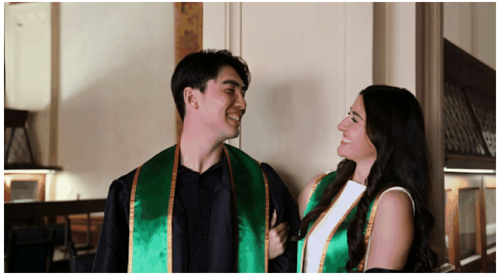


Prompt B

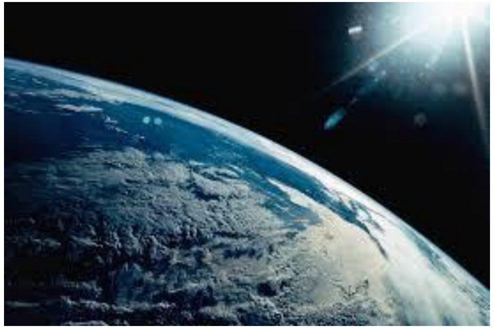
“Flawed” GIF generation: I2V-GenXL Pipeline

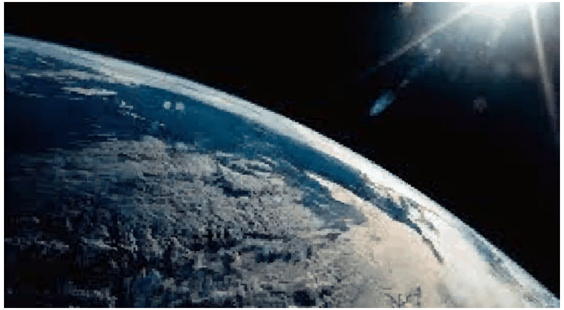
Prompt:
“A couple kissing the library”



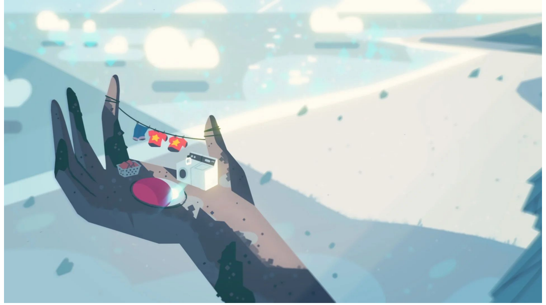


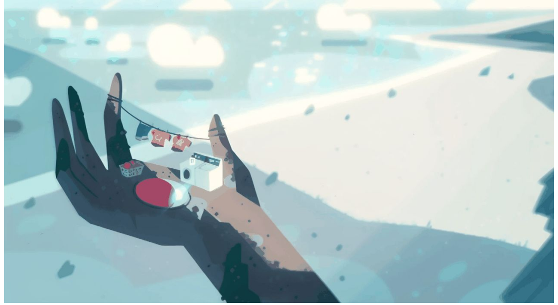
Prompt:
“The earth rotating with the glare of the sun changing. The International Space Station is orbiting in the background.”





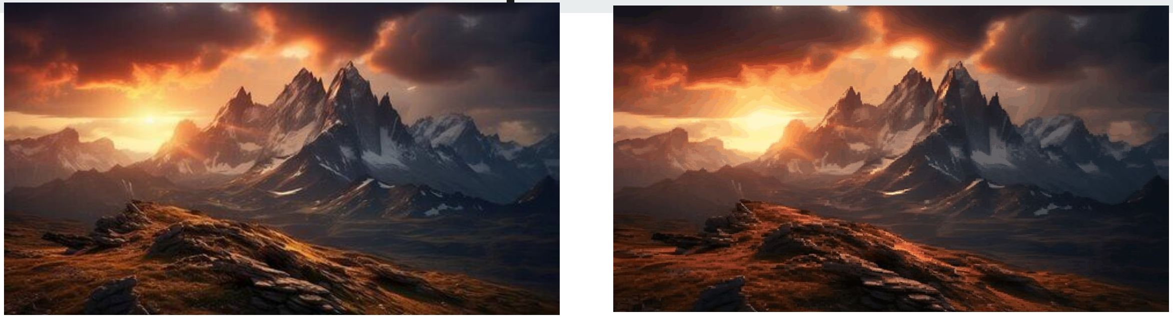
Prompt:
“Cartoon Beach with rolling waves.”



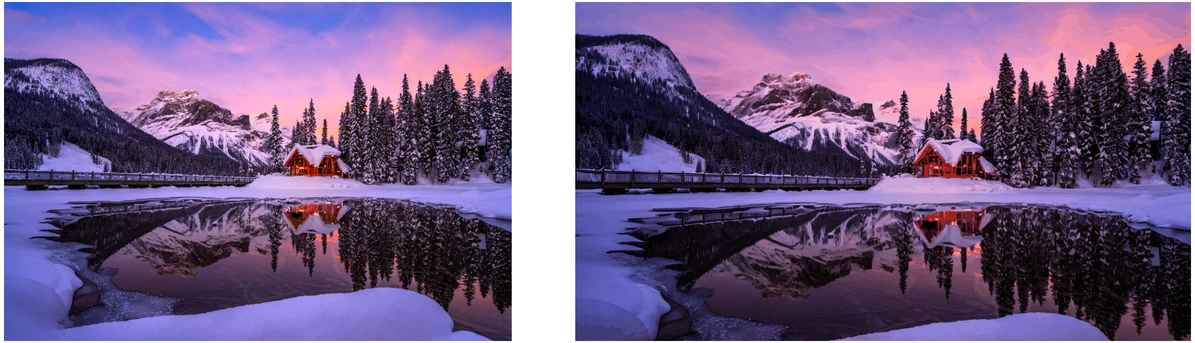


“Better” GIF generation: I2V-GenXL Pipeline

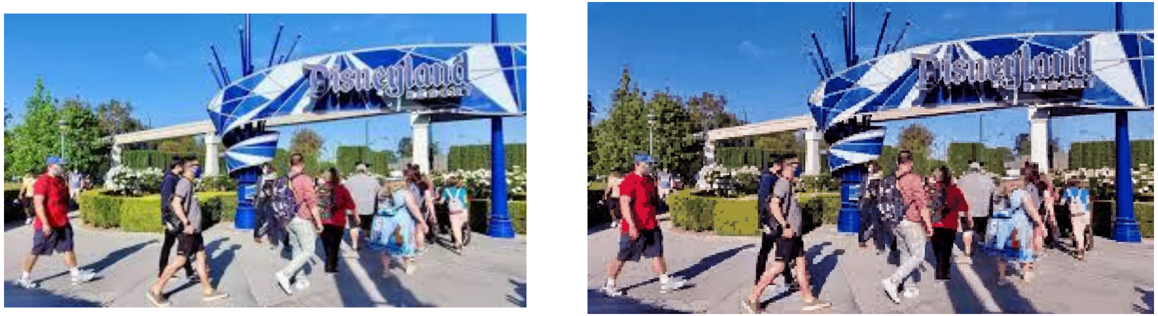
Prompt:
“ The Sunset happens behind the mountains”



Prompt:
“Night falls behind the cabin”



Prompt:
“People walking into Disneyland”



Result

- Depending on the prompt, we got different results of images
- Different levels of guidance scale produced different results.
- A lower guidance scale less fine-tuned, as we increase scale, we fine tune the model.
- The smaller the num inference step, the quality of video decreases. Saw that 50 is a good number.

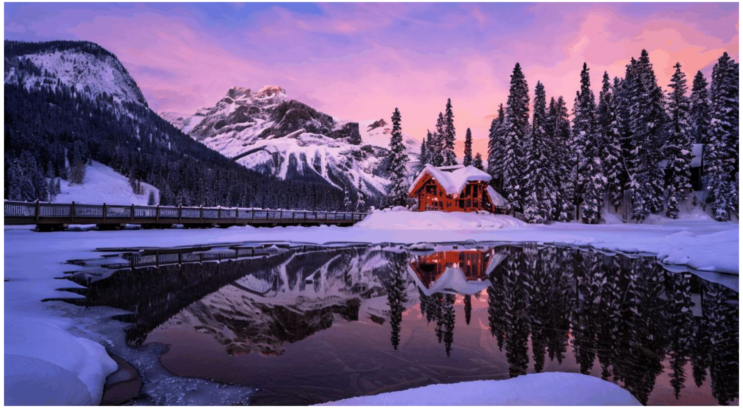
Conclusion

- Prompt Engineering
    - Small, General, Action Prompts Work Best
    - ***"Night falls behind the cabin"***
    - ***"The Sunset happens behind the mountain"***
- `num_inference_step  = 50`
- `guidance_scale = 9.0`
- Possible relationship between model size and maximum prompt length / detail

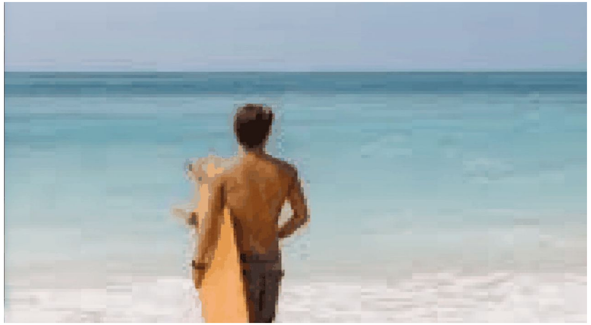
Limitations

- Generated people (facial expressions, etc.) are not perfect
- Need a prompt appropriate to the model size
- Object continuity across frames is imperfect
- Object recognition from text prompt is imperfect/unreliable

Prompt: "A Fish Jumping out of the pond and creating a rippling splash."



Prompt: "Man Walking toward the beach and getting splashed by the waves"



Further studies

- Taking an existing video, and modifying that video
- Fine-tuning models for object detection
- Objective evaluation metrics / method development
- Better Conditioning on prompts