StoryForge Project Report

30-06-2024

Team - JARVIS

Parikshit Gehlaut R Eshwar

Abstract

This report presents the development of a Text-to-Video (T2V) generation model created during the StoryForge. Our model leverages Stable Diffusion for Text-to-Image (T2I) generation and Stable Video Diffusion (SVD) for Image-to-Video (I2V) generation.

The ability to generate videos from textual descriptions has significant applications in fields such as entertainment, education, and virtual reality. This project aims to develop a Text-to-Video (T2V) model that can generate high-quality videos based on given text prompts. By integrating Stable Diffusion for generating images and SVD for creating video sequences, our model aims to push the boundaries of current generative models..

Methodology

Our model consists of two main components: the Text-to-Image (T2I) generation module and the Image-to-Video (I2V) generation module.

Text-to-Image

We used Stable Diffusion to generate images from textual descriptions. Stable Diffusion is a powerful model known for producing high-quality images from text.

Image-to-Video

For the I2V generation, we applied Stable Video Diffusion (SVD) to transition from static images to dynamic video sequences. This technique ensures smooth transitions and temporal coherence.

Result

Below are some examples of the images and videos generated by our model from given text prompts.

Text-to-Image

- Text Prompt: "`Epic anime artwork of a wizard atop a mountain at night casting a cosmic spell into the dark sky made out of colorful energy, highly detailed, ultra sharp, cinematic, 100mm lens, 8k resolution.`"
- Image :

Image-to-Video

- Text Prompt: “ winnie the pooh wearing a christmas hat, walking through the street while it's snowing, close up, 4k, artstation, realistic”
- Video Generated: https://youtu.be/4qFsjO75Lz0?feature=shared

Conclusion and Future Work

Our Text-to-Video generation model shows promising results in creating videos from textual descriptions. Following points will be our Future work:

- Currently for our video diffusion model, we are using an inbuilt pipeline from Hugging face, our next step will be coding our model for generated video from image.
- Audio Integration is also part of a project that we are not able to for many reasons.
- Of Course following the latest research to further improve our model.

References

Special thanks to **Stability.ai**

- https://github.com/divamgupta/stable-diffusion-tensorflow
- https://github.com/kjsman/stable-diffusion-pytorch
- https://github.com/Stability-AI/generative-models
- https://huggingface.co/stabilityai/stable-video-diffusion-img2vid-xt

Code

Link : https://github.com/Challenger-84/StoryForge-JARVIS