Analyzing Diffusion Model Applications in Video Content Creation

Haiyun Xiao, Yu Li, Haichen Pang

September 2023

Proposal Study Planner - Collaboration

1 Executive Summary

With the popularity of generative artificial neural networks, one can observe numerous applications, such as text-to-image, text-to-speech, and text-to-video. The objective is to further investigate potential future applications in video content creation. The initial plan involves researching and evaluating various mainstream diffusion models, comparing their unique features and differences. Subsequently, the aim is to integrate these models with content creation, emphasizing their unique attributes. This endeavor is aimed at expediting content production and facilitating video creators in delivering higher quality output.

2 Survey of Background Literature

2.1 Overview

At present, diffusion models have found diverse applications in data generation. One of the most prominent applications is the Stable Diffusion model[1], which is particularly noteworthy in the field of image generation, such as in art design area. In the era of short videos, the potential use of stable diffusion for video generation holds promise. This could potentially alleviate the burden on content creators, allowing them to focus primarily on script writing, while the large model takes care of all other aspects.

2.2 Background

Diffusion models can be applied to a variety of tasks, including image denoising, inpainting, super-resolution, and image generation[2]. It can learn the probability distribution of a given dataset, then use its knowledge to reproduce the sample via the inputs. Latent diffusion model is a main variants of diffusion models. It operate by repeatedly reducing noise in a latent representation space and then converting that representation into a complete image. It produces many interesting text conditional generators, which can be used for images, specs, music, videos, and many more.

3 Proposed Methodology

What we envision is that as a content video creator, you only need to write down the text script of the story you are going to shoot, and then send it to the big model. First, the large language model will automatically generate the title and introduction of the video based on the script, and will also add more text details to the steps. Next, latent diffusion model generates the required video based on the input text details. Of course, in this project, we will focus on the video generation part in the latter part.

4 Research Plan

- Research the current mainstream diffusion models
- Try to reproduce and restore some diffusion model structures and applications of the paper
- Compare the differences between these models to identify their unique features and advantages
- Apply them to the field of video creation

References

[1] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2021.

[2] Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.