Optimizing Generative Video Quality with Diffusion Models

Jeevan Vasanthan (jv1564), Sai Akhil Tekuri (st5050), Vyom Yadav (vy366)

Problem Statement

Stable diffusion models are powerful generative neural networks used for image synthesis. This is very useful in non-photorealistic rendering, commonly utilized for game development, animation, and illustration purposes. For this project, we are interested in extending these concepts to video input. We plan on taking real-time video input and generating non-photorealistic videos with the Pixar art style. To accomplish this, we will have to explore the many challenges presented such as minimizing video flickering and maintaining video quality.

Literature Survey

The paper, "Diffutoon: High-Resolution Editable Toon Shading via Diffusion Models", proposes an effective video stylization approach that involves a personalized stable diffusion model for anime stylization, motion modules for temporal consistency, and different ControlNet models such as colorization and edge detection to enhance the controllability of the generated video results[1]

Dataset

We will be training a personalized stable diffusion model using the PIXAR-SDXL dataset which consists of 1,851 Pixar images (Dataset Link).

Models

Model that we will train: Stable diffusion model to generate videos of Pixar art style
Pre-trained models: The various ControlNet models that will be explored to integrate control signals

Deliverables

An end-to-end generative model takes a prompt as input and outputs a cartoon-styled video while maintaining visual quality

References

[1]https://arxiv.org/pdf/2401.16224.pdf
[2]https://arxiv.org/pdf/2112.10752.pdf
[3]https://arxiv.org/pdf/2302.05543.pdf