International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 SJIF 2023: 8.176 ISSN: 2582-3930

TEXT-IMAGE GENERATION-STABLE DIFFUSION

Unnati kolte, Samprada Kale, Priti Yamkar, Sakshi Deshpande
Department of Computer Engineering, Savitribai Phule Pune University,
Pune – 411007

Abstract: The text to image conversion task has long been a challenging problem in the field of computer vision. Recent advancements in generative models have led to the development of various text-to-image synthesis techniques. In this project, we explore the use of the Stable Diffusion framework for text-to-image synthesis. The Stable Diffusion model is a diffusion-based generative model that can produce high-quality images with fine-grained details. We propose a novel approach to generate images from textual descriptions using the Stable Diffusion model, which involves learning a joint embedding space for the text and the image domains. Our approach enables the Stable Diffusion model to generate highly realistic images that are closely aligned with the given textual descriptions. We evaluate our proposed approach on several benchmark datasets, and our experimental results demonstrate the effectiveness of our proposed method for text-to-image synthesis.

I. INTRODUCTION

The ability to generate images from textual descriptions has been a longstanding goal in the field of computer vision. It has various practical applications, such as generating images for virtual reality, video games, and image editing. Recent advancements in deep learning and generative models have enabled the development of various text-to-image synthesis techniques. In particular, diffusion-based generative models have shown great potential for generating high-quality images with fine-grained details.

II. LITERATURE REVIEW

The Stable Diffusion framework is used in this paper which is an extension of the DDPM that can generate high-quality images with fine-grained details. Several approaches have been proposed that use diffusion-based generative models for text-to-image synthesis, but these methods have limitations. A new approach has been proposed that uses the Stable Diffusion framework for text-to-image synthesis by learning a joint embedding space for text and images, which generates high-quality images with complex image attributes and avoids mode collapse. The proposed method has been evaluated on several benchmark datasets and has demonstrated effectiveness in generating high-resolution images.

III. METHODOLOGY

This project is done by using diffusion models mainly latent diffusion model for creating art from natural language text descriptions.

Diffusion Models: Diffusion models are Generative models, data which is similar to the data the model is trained on is produced. This model works by a method of adding Gaussian Noise and then it learns to recover the data by undoing noise added.

Latent Model: In the Stable Diffusion framework, a latent model refers to the generative model that is used to produce the images from the latent space. The Stable Diffusion model operates by iteratively diffusing the noise in the latent space, which produces a sequence of intermediate representations that gradually become more structured and resemble the target image. The latent model in Stable

Diffusion typically uses a deep neural network to learn the distribution of the latent space, which is then used to sample the intermediate representations at each diffusion step. The output of the latent model is a sequence of intermediate representations that are then transformed into the final image using an inverse diffusion process. The use of the latent model in Stable Diffusion enables the model to capture complex image attributes and generate high-quality images with fine-grained details

Steps Involved:

Data collection: Collect a large dataset of paired text and image examples for training and evaluation.

Preprocessing: Preprocess the data by resizing, cropping, and normalizing the images, and tokenizing and embedding the text.

Model architecture: Design a model architecture that uses the Stable Diffusion framework for text-to-image synthesis, with a joint embedding space for the text and image domains.

Training: Train the model using the collected data, optimizing the loss function to minimize the difference between the generated images and the target images.

Evaluation: Evaluate the model on several benchmark datasets by computing various metrics such as Fréchet Inception Distance (FID) and Inception Score (IS).

Hyperparameter tuning: Tune the model hyperparameters to improve performance.

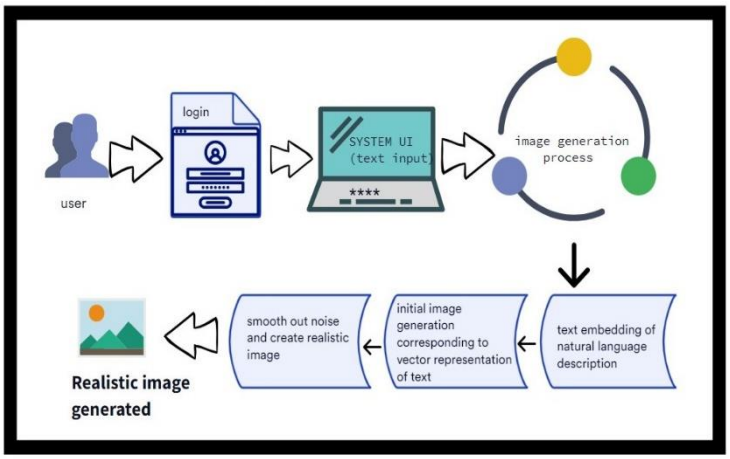
Error analysis: Analyze the errors made by the model to identify areas for improvement.

Comparison: Compare the proposed method with state-of-the-art text-to-image synthesis techniques.

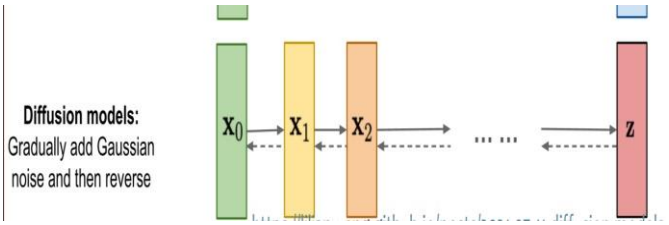
Deployment: Deploy the trained model for use in real-world applications.

Documentation: Document the methodology, code, and results for reproducibility and future reference.

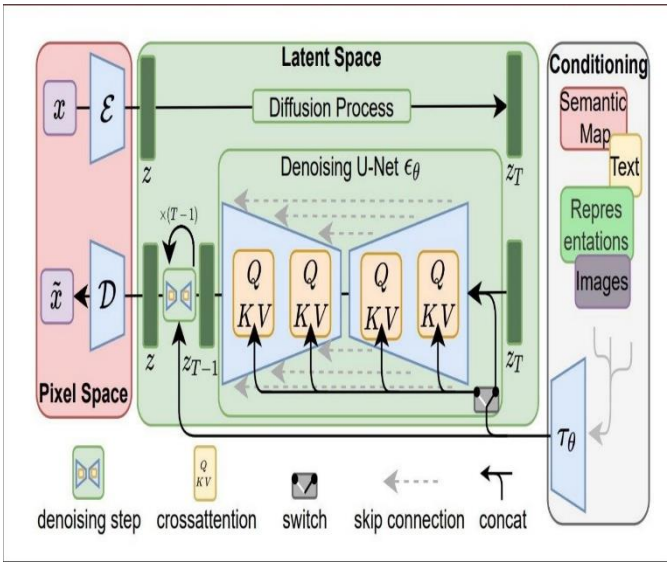
IV. SYSTEM DIAGRAM



V. SEQUENCE DIAGRAM



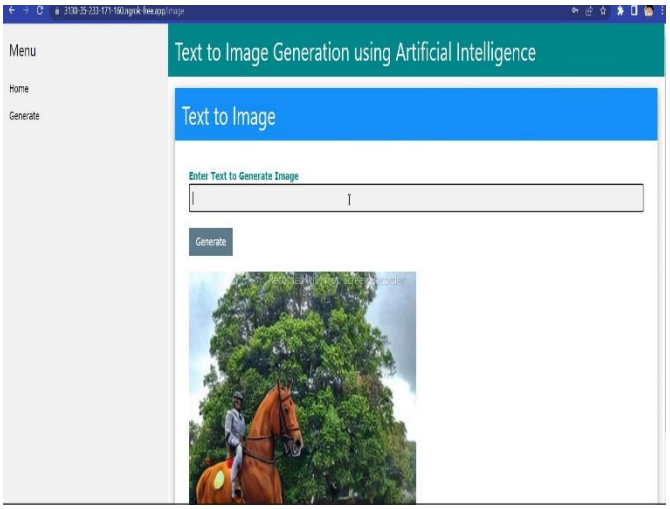
VI. MODEL ARCHITECTURE

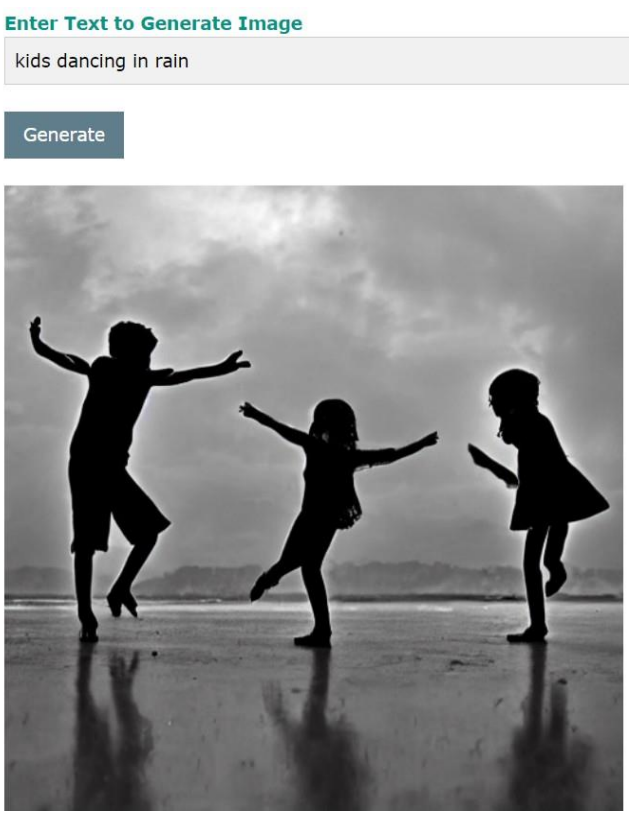


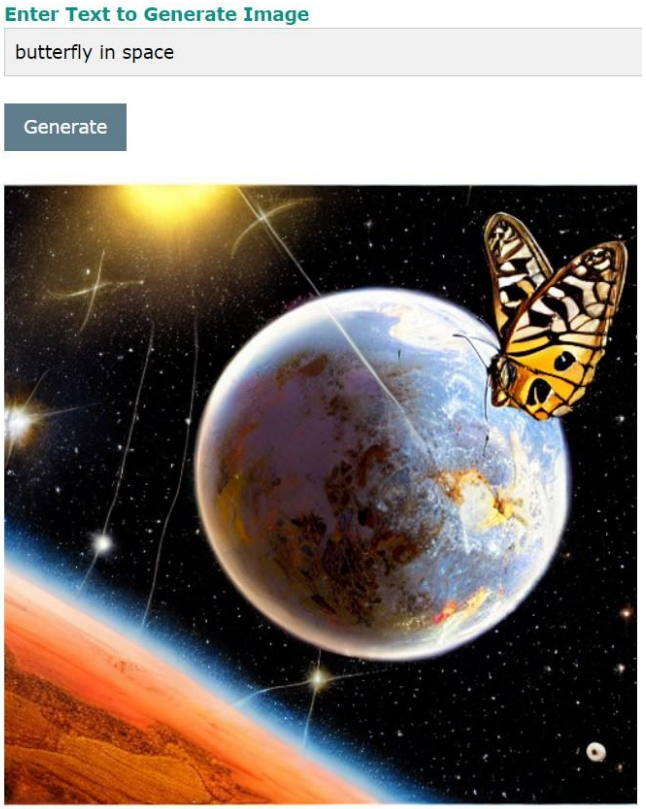
VI. RESULT

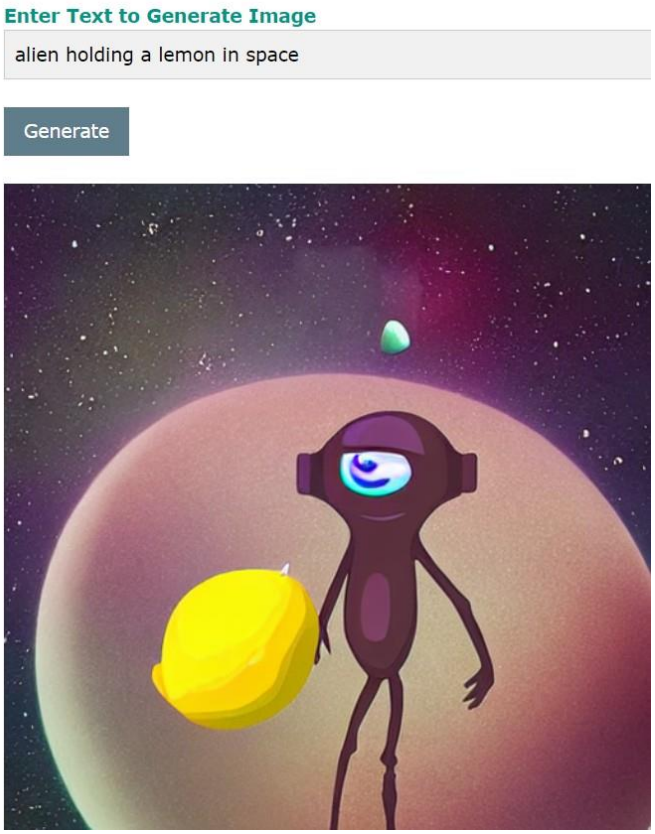
The results of text-to-image generation project using Stable Diffusion would be an image that is generated based on the input text prompt. The generated image aims to represent or depict the content described in the text.

Here are few results generated by our project









VII. CONCLUSION

In conclusion, a text-to-image generation project using Stable Diffusion combined with HTML, CSS, and Flask can provide an interactive and visually appealing way to generate images from textual input. By leveraging the power of Stable Diffusion, the project can generate high-quality images that correspond to the provided text prompts. HTML and CSS can be used to create an intuitive and user-friendly interface where users can enter their desired text prompts and view the generated images. The design and layout can be customized using CSS to enhance the overall user experience. Integrating Flask, a Python web framework, allows for seamless communication between the frontend and the backend. Flask can handle the user input, pass it to the Stable Diffusion model for image generation, and return the generated image to the frontend for display. Such a project can open up possibilities for various applications, including generating visual representations for textual data, assisting in creative design processes, or enhancing storytelling by transforming written descriptions into vivid images. However, it is important to note that the success of the project relies on various factors such as the quality of the trained Stable Diffusion model, the availability and suitability of the training data, and the implementation details. Regular updates and advancements in the field of text-to-image generation should also be considered to ensure that the project remains at the forefront of the latest research and techniques. By combining Stable Diffusion with HTML, CSS, and Flask, the text-to-image generation project can provide a valuable tool for generating visually appealing images from textual input, unlocking new possibilities in creative expression, data visualization, and user interaction.

ABSTRACT:

This project explores the application of Stable Diffusion in the field of text-to-image generation. Stable Diffusion, a powerful generative model, is employed to iteratively refine images based on textual input. The project utilizes HTML, CSS, and Flask to create an interactive web interface that allows users to input text prompts and generate corresponding images. The combination of Stable Diffusion with these web technologies enables the creation of visually captivating and contextually relevant images based on textual descriptions. The project demonstrates the potential for generating high-quality images from text, opening up opportunities for various applications in data visualization, creative design, and storytelling.

REFERENCES

Nick babich, How to generate stunning images using stable diffusion 09jan2023

Rinon Gal, Yuval Alaluf, Yuval Atzmon, An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion 2Aug2022 arXiv:2208.01618 Cornell university

Yossef hosni, Getting Started With Stable Diffusion Nov11 2022

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer, High-Resolution Image Synthesis with Latent Diffusion Models 13 Apr2022 arXiv:2112.10752