School of
Electronics and Communication Engineering

Course Project Report

on

Image Variation Generation Using Stable Diffusion

By:

1. **Saurav Kumar** USN:01FE21BEC365
2. **Vaishnavi Agrawal** USN:01FE21BEC168

Semester: VII, 2024-2025

Under the Guidance of

Uma Mudenagudi, Ramesh Ashok Tabib

K.L.E SOCIETY'S
KLE Technological University,
HUBBALLI-580031
2024-2025

SCHOOL OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that project entitled **" Image Variation Generation Using Stable Diffusion "** is a bonafide work carried out by the student team of **"Saurav Kumar(01fe21bec365), Vaishnavi Agrawal(01fe21bec168) "**. The project report has been approved as it satisfies the requirements with respect to the course project work prescribed by the university curriculum for BE (VII Semester) in School of Electronics and Communication Engineering of KLE Technological University for the academic year 2024-2025.

Uma Mudenagudi, Ramesh Ashok Tabib	**Suneeta V Budihal**	**B. S. Anami**
Guide	**Head of School**	**Registrar**

External Viva:

Name of Examiners	**Signature with date**
1.	
2.	

ACKNOWLEDGMENT

I would like to express my heartfelt gratitude to everyone who contributed to the completion of my project on "Image Variation Generation Using Stable Diffusion." I am deeply thankful to the researchers and developers whose work on generative models and diffusion processes formed the foundation of this study.

A special thanks to the creators of Stable Diffusion for making this powerful tool accessible as well as to the open-source community for providing valuable resources and support. Your contributions were crucial to the success of this project.

I would also like to acknowledge the guidance and feedback of my professors, mentors, and peers, whose insights and encouragement greatly improved my work. Finally, I am grateful to my family and friends for their unwavering support and patience throughout this research journey.

-Saurav Kumar, Vaishnavi Agrawal

ABSTRACT

This paper explores the use of Stable diffusion for generating diverse image variations from a given input, with a focus on enhancing the quality, diversity, and realism of generated images. Leveraging the power of diffusion models, the study investigates methods for conditioning image generation on specific attributes, such as style, content, and identity, while preserving natural variation. We present a novel approach that integrates techniques like latent space manipulation and fine-tuning to allow for more controlled image variations. Through experiments, we demonstrate the ability of stable diffusion to generate high-quality, consistent image variations with minimal loss of detail or fidelity. The findings suggest that diffusion models, especially stable diffusion, hold significant potential for personalized and creative image generation across various domains, from art to realistic imagery. This work contributes to ongoing advancements in generative models and their application in dynamic image synthesis, highlighting future directions for enhancing model adaptability and efficiency in generating realistic and varied outputs.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

Stable Diffusion allows users to generate multiple image variations by adjusting aspects like color, style, and content, offering creative flexibility.

With its inpainting and style transfer capabilities, stable Diffusion enables users to refine and modify images.

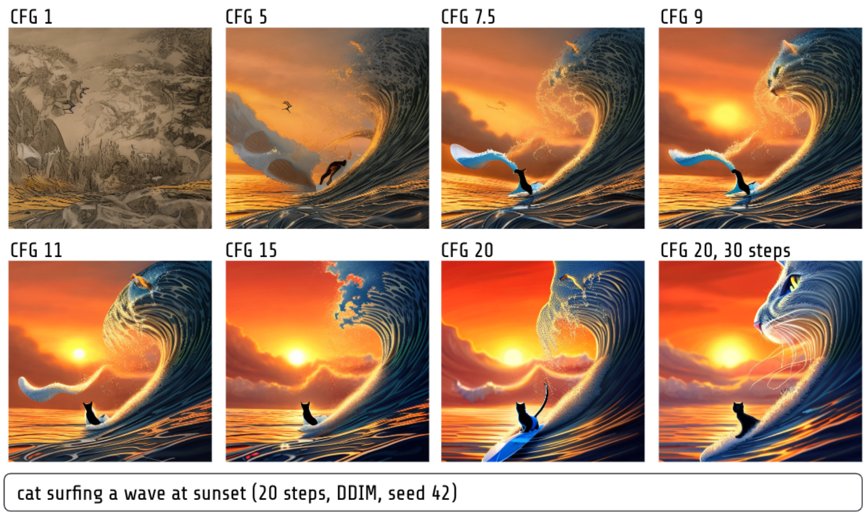

Figure 1.1: Stable Diffusion Classifier Guidance

1.2 Objectives

Generate multiple visually distinct variations of a given input image.

1.3 Literature survey

1.3.1 Versatile Diffusion: Text, Images and Variations All in One Diffusion Model

Diffusion models, like DALL-E 2 and stable diffusion, surpass traditional GANs by enabling cross-modal tasks such as text-to-image generation and image superresolution, offering greater versatility across domains.

Superior Performance: Empirical results show that VD outperforms traditional models in both qualitative and quantitative assessments, highlighting its potential for universal AI research.

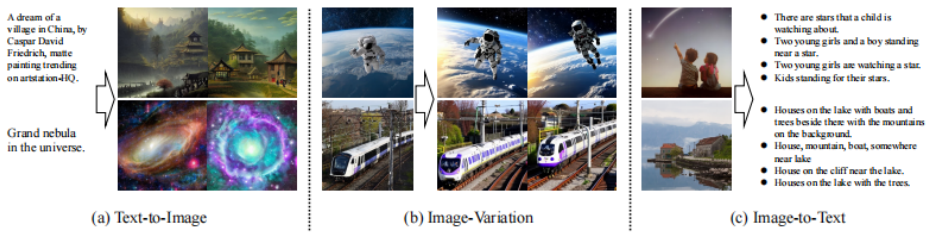

Figure 1.2: Demo results of our Versatile Diffusion (VD) framework on three out of all primary tasks.

1.3.2 When StyleGAN Meets Stable Diffusion: a W+ Adapter for Personalized Image Generation

Recent personalized text-to-image (T2I) methods embed target identities into textual spaces but struggle with balancing identity preservation and diverse facial attributes. Techniques like textual inversion and dreambooth face difficulties in separating identity features from irrelevant attributes.

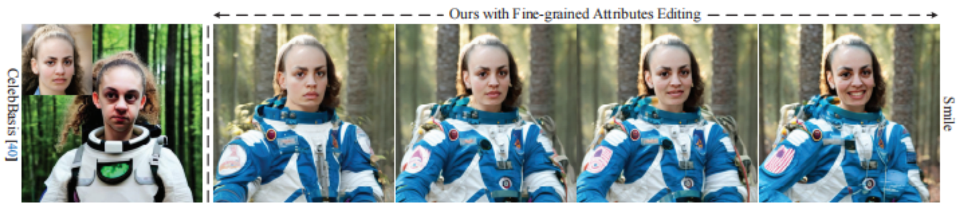

Figure 1.3: The text prompt is "a woman wearing a spacesuit in a forest."

1.4 Problem statement

Generating diverse, high-quality image variations while preserving key attributes is challenging. This research aims to improve image variation generation with stable diffusion by enhancing fidelity, diversity, and model adaptability.

1.5 Functional block diagram

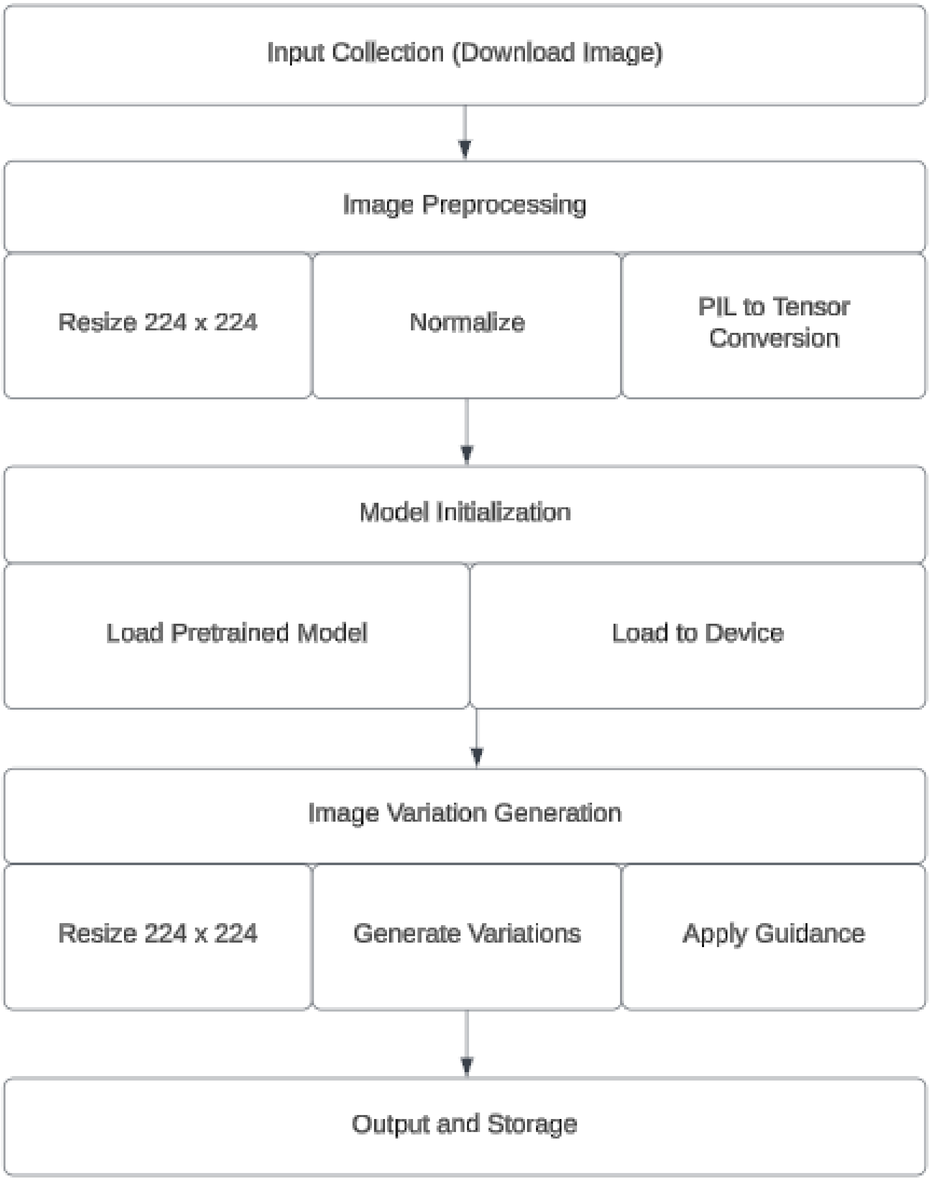

Figure 1.4: Proposed Pipeline

1.Input Collection:
Step 1: Image URL (can be replaced with other input methods like local files).
Step 2: Download the image from the URL.

2. Image Preprocessing:
Step 1: Download image from URL or load from local storage.
Step 2: Convert the image into a format that the model can process (e.g., tensor).
Step 3: Apply necessary transformations (resize, normalization, etc.) to prepare the image for input into the model.

3. Model Initialization:
Step 1: Load the StableDiffusionImageVariationPipeline from Hugging Face's diffuser library.
Step 2: Initialize the pipeline with the pre-trained model("lambdalabs/sd-image-variations-diffusers").
Step 3: Transfer the model to the appropriate device (e.g., GPU, CUDA).

4. Image Variation Generation:
Step 1: Define the number of image variations (e.g., 5 variations).
Step 2: For each iteration: Set a unique random seed for diversity in output. Pass the preprocessed image to the pipeline for image generation with a guidance scale to control the degree of adherence to the input image.
Step 3: Store the results in a list or dictionary (for internal reference) and save each image as a separate file on disk.

5. Output and Storage:
Step 1: Each generated image is saved with a unique filename.
Step 2: Log and print the successful save message for each image.

Chapter 2

Results and discussions

2.1 Result Analysis

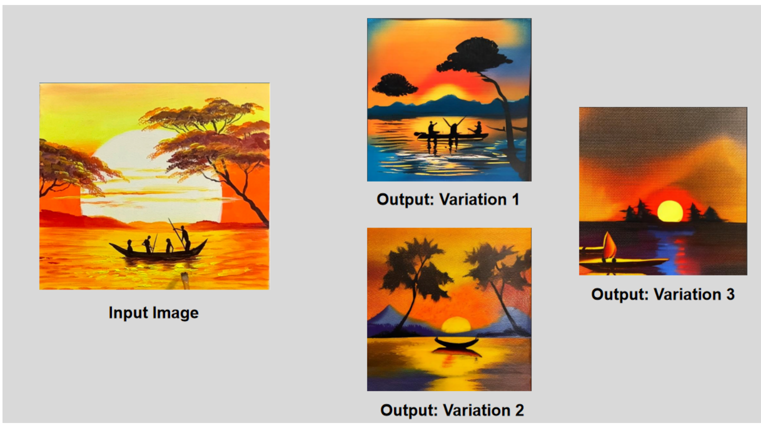

Figure 2.1: Input of the beach image and three output variations are generated.

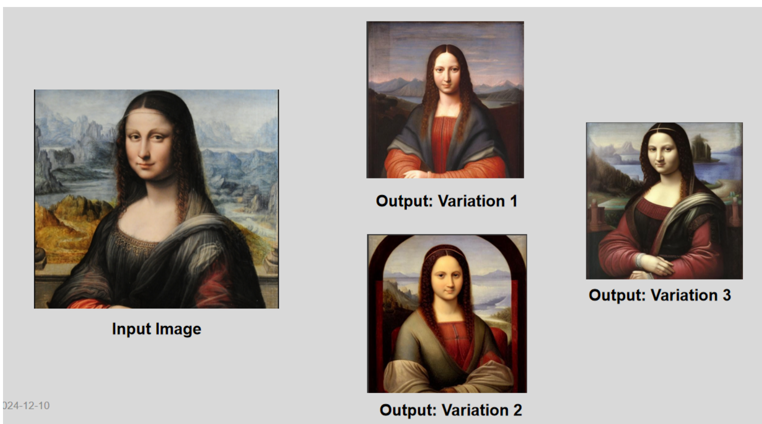

Figure 2.2: Input of Mona Lisa and three output variations are generated.

Chapter 3

Conclusions and future scope

3.1 Conclusion

In conclusion, this research highlights the significant potential of stable diffusion in generating diverse and high-quality image variations while effectively preserving key attributes such as style, content, and identity. By exploring novel techniques for controlling image diversity, we were able to improve both the fidelity and realism of generated images, demonstrating that diffusion models can successfully balance natural variation with consistency. This study contributes to the growing body of work on generative models, particularly in the context of personalized image synthesis, where maintaining identity while allowing for variation is critical. The findings suggest that stable diffusion holds great promise for diverse applications, from creative arts to realistic image generation, and opens up new possibilities for more controlled and adaptable image generation workflows. Future work could further refine these methods, addressing remaining challenges such as reducing artifacts, improving fine-grained control over attributes, and enhancing computational efficiency. Ultimately, this research paves the way for more versatile and powerful generative models that can meet the demands of complex, real-world image generation tasks.

Bibliography

[1] Xingqian Xu, Zhangyang Wang, Eric Zhang1, Kai Wang1 Humphrey Shi, *SHI Labs @ Georgia Tech UIUC Oregon, 2UT Austin Picsart AI Research (PAIR)*

[2] Xiaoming Li, Xinyu Hou, Chen Change Loy, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, pp. 2187-2196*

[3] Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C Lawrence Zitnick, and Devi Parikh. Vqa: Visual question answering. In *Proceedings of the IEEE international conference on computer vision,*, pages 2425–2433, 2015

[4] David Bau, Jun-Yan Zhu, Hendrik Strobelt, Bolei Zhou, Joshua B. Tenenbaum, William T. Freeman, and Antonio Torralba. Gan dissection: Visualizing and understanding generative adversarial networks *In Proceedings of the International Conference on Learning Representations* , (ICLR), 2019.