Image Generation using Stable Diffusion and Comfy UI

A Project Report

submitted in partial fulfillment of the requirements

of

AICTE Internship on AI: Transformative Learning
with
TechSaksham – A joint CSR initiative of Microsoft & SAP

by

Shaheen Shaik, shaikfathima08390@gamil.com

Under the Guidance of

Adarsh P

ACKNOWLEDGEMENT

I am very grateful to take the to express my gratitude to all the individuals who helped me guided me selflessly during the entire project tenure.

Firstly, I wolud like to thank **Mr.Jay Rathod** whose mentorship and guidance had helped me in the successful completion of this project. His mentorship helped to learnt about Stable diffusion and Comfy UI models and enabled me to build a complete and functional implementation of them models. His constant support and advices helped me to overcome the obstacles that I had faced in the project.

I would also like to thank **Mr.Pavan kumar Sumohana** who shared all the updates regarding to the sessions which helped me to attend all the sessions. His support and timely updates greatly enhanced my learning experience.

A special thanks to **Adarsh P** Sir, his exceptional knowledge on the subject and explanation in theory classes made me to learnt about the concept very clearly and helped me grasp complex concepts with ease. I am truly grateful for your invaluable guidance and lectures. Thank you for being such a remarkable mentor.

I would like to thank **Techsaksham,Edunet foundations** for this joint csr intiative program by **Microsoft and SAP.**

Shaheen Shaik

ABSTRACT

This project primarily focuses on generating images based on user input or prompts, while also developing a robust and efficient image generation system. The core objective is to produce high-quality images with precision and clarity, ensuring an optimal balance between creativity and computational efficiency. By leveraging advanced techniques, the project aims to enhance the overall user experience and streamline the image generation process.

Objective of the project:

1) Generation of the images using advanced model like Stable Diffusion.
2) Enhancing the user experience by integrating advanced techniques for improved performance and making it more user friendly.
3) Ensure precision and clarity in the generated images while maintaining computational efficiency.

Methodology of the project:

1) Utilize **Stable Diffusion** and **ComfyUI** models for text-to-image generation.
2) Implement a structured workflow to process user prompts and translate them into high-quality images.
3) Integrate a user-friendly interface to facilitate seamless interaction with the system.

Key Results:

1) Successfully generated high-resolution images based on using various prompts.
2) Developed a stable and reliable implementation, making the system accessible and user-friendly.
3) Achieved a balance between computational efficiency and creativity, optimizing resource utilization.

Conclusion: In the conclusion I would like to add I have used and implemented a model which help us to generate the images based on the text .This project successfully demonstrates an effective approach to generating high-quality images based on user input. The results validate the effectiveness of the implemented approach, laying the foundation for future improvements and potential applications in various creative and commercial domains.

TABLE OF CONTENT

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Problem Statement:

The problem being addressed is the need for high-quality, customizable, and efficient AI-generated images using **Stable Diffusion** and **ComfyUI**. Traditional methods of digital art and image creation can be time-consuming, require specialized skills, and may not always provide the desired flexibility. AI-powered image generation, specifically using Stable Diffusion, offers a powerful alternative that allows users to generate complex and creative visuals with minimal effort.

Significance of the problem:

1. **Accessibility & Efficiency**
    - Many artists, designers, and creators lack the time or skills to manually create high-quality images. AI-powered tools provide a faster and more accessible way to generate professional visuals.
2. **Quality & Realism**
    - Ensuring that AI-generated images are free from artifacts, distortions, or unwanted elements remains a challenge. ComfyUI allows for detailed workflow adjustments to improve image quality.
3. **Ethical & Creative Concerns**
    - AI-generated content raises ethical questions about originality, copyright, and bias. By using **open-source tools like Stable Diffusion**, users can have more transparency and control over how images are generated.
4. **Scalability for Industries**
    - Industries such as game development, advertising, and film production require large volumes of visual content. AI image generation speeds up production without sacrificing quality.

1.2Motivation:

Why This Project?

This project was chosen to address the need for **high-quality, customizable, and efficient AI-generated images** using **Stable Diffusion** and **ComfyUI**. Traditional methods are time-consuming, and existing AI tools often lack flexibility. ComfyUI's **node-based workflow** gives users more control over the generation process, making AI art more accessible and powerful.

Potential Applications:

1. Digital Art & Design

- Artists can experiment with different styles, compositions, and ideas without extensive manual effort.
- Designers can quickly generate concept art, logos, and promotional materials.

2. Game Development & Animation

- Game studios can generate textures, characters, and environments efficiently.
- Animators can prototype scenes and visual elements quickly.

3. Marketing & Advertising

- Marketers can create high-quality promotional visuals without relying on stock images.
- Advertisers can generate unique and engaging content tailored to their campaigns.

4. Content Creation & Social Media

- Influencers and content creators can generate custom illustrations, thumbnails, and backgrounds.
- Video creators can use AI-generated assets to enhance storytelling.

5. Film & Storyboarding

- Film directors and screenwriters can create visual storyboards efficiently.
- Concept artists can generate realistic or stylized scenes for pre-production.

Impact of the Project

- **Democratization of Art**: Enables non-artists to create visually stunning images.
- **Faster Creative Workflows**: Reduces time needed for concept development.
- **Personalization & Customization**: Provides control over AI-generated images.
- **Cost Savings**: Eliminates the need for expensive stock images or commissioned work.
- **Innovation in AI Art**: Advances the use of AI in creative fields, pushing the boundaries of what's possible.

1.3Objective:

Objectives of the Project

1. **Develop an Efficient AI Image Generation Workflow**
    - Utilize **Stable Diffusion** and **ComfyUI** to create high-quality, customizable images.
    - Optimize the workflow for better efficiency, flexibility, and user control.
2. **Enhance Customization & Control**
    - Implement **ComfyUI's node-based approach** for precise image adjustments.
    - Allow users to fine-tune image parameters such as style, resolution, and composition.
3. **Improve Image Quality & Consistency**
    - Minimize artifacts, distortions, and inconsistencies in AI-generated images.
    - Experiment with different models, prompts, and settings for better outputs.
4. **Expand Creative Applications**
    - Enable AI-driven image generation for **art, design, gaming, marketing, and film**.
    - Provide a scalable solution for industries needing large volumes of visual content.

1.4Scope of the Project:

Scope:

1. **AI Image Generation Using Stable Diffusion & ComfyUI**
    - Utilize **Stable Diffusion** for high-quality AI-generated images.
    - Implement **ComfyUI's node-based workflow** to enhance customization and control.
2. **Customization & Workflow Optimization**
    - Allow fine-tuning of image parameters such as **style, resolution, and composition**.
    - Support for different models, prompts, and settings to improve output quality.
3. **Applications Across Multiple Fields**
    - **Art & Design**: Concept art, illustrations, and branding.
    - **Game Development & Animation**: Character, texture, and scene generation.
    - **Marketing & Advertising**: Unique campaign visuals.
    - **Content Creation**: Thumbnails, social media posts, and video assets.
    - **Film & Storyboarding**: Pre-visualization and scene planning.
4. **User Accessibility**
    - Designed for **both beginners and advanced users**.
    - Focus on making AI image generation **intuitive and flexible.**

Limitations of the project:

1. **Image Generation Speed**
    - Processing time varies based on **model size, resolution, and hardware**.
    - High-resolution or detailed images may take longer to generate.
2. **Consistency & Control Challenges**
    - Achieving **consistent character design** across multiple generations is difficult.

- Requires fine-tuning and experimentation to maintain a specific style.

3. **Ethical & Copyright Concerns**

   - AI-generated images may **replicate copyrighted styles or content** if not carefully controlled.
   - The ethical use of AI in art remains a debated topic.

4. **Limited Realism & Creativity Constraints**

   - AI-generated images may sometimes produce **unrealistic or distorted features**.
   - The model's creativity is limited by its training data and prompt effectiveness.

CHAPTER 2

Literature Survey

2.1. Review of Relevant Literature or Previous Work in This Domain

The field of image generation has experienced remarkable progress in recent years, particularly with the emergence of **diffusion-based models**. A comprehensive survey by **Zhang et al. (2023)** examines advancements and challenges in **diffusion-based image generation**, providing insights into the **mathematical foundations** of these models and their role in producing **high-quality, realistic images**.

A significant contribution by **Nikam et al. (2024)** explores the use of **Stable Diffusion** for image generation, emphasizing its **efficiency and adaptability**. Their study highlights the advantages of **latent space diffusion**, which enhances **processing speed and reduces computational overhead**, making it a viable solution for diverse applications.

2.2 Existing Models, Techniques, and Methodologies

- **DDPMs (Denoising Diffusion Probabilistic Models)** – Gradually refine noise into high-quality images (**Ho et al., 2020**).
- **LDMs (Latent Diffusion Models)** – Optimize diffusion in latent space for better efficiency (**Rombach et al., 2022**).
- **CLIP** – Helps AI understand and generate images based on text prompts (**Radford et al., 2021**).
- **ControlNet** – Adds structure-based controls for precise outputs (**Zhang & Agrawala, 2023**).
- **LoRA** – Enables quick fine-tuning for specific styles with minimal training (**Hu et al., 2021**).
- **Automatic1111 WebUI** – A simple, prompt-based UI for Stable Diffusion.

2.3 Gaps and Limitations in Existing Solutions & How This Project Addresses Them

- **High Computational Requirements** – Many existing models, including **Stable Diffusion**, require **powerful GPUs** for optimal performance. This project aims to **optimize workflows and reduce computational load** through **efficient model configurations**.
- **Limited User Control** – Traditional **prompt-based interfaces** (e.g., **Automatic1111 WebUI**) lack **granular control** over image generation. By using **ComfyUI's node-based approach**, this project will enable **precise customization** of generation parameters.
- **Steep Learning Curve** – Many AI image generation tools are complex for beginners. This project will create a **user-friendly workflow in ComfyUI**, making advanced AI techniques more accessible.
- **Lack of Structured Editing Capabilities** – Existing solutions struggle with **fine-grained control over image elements** (e.g., **pose control, structure preservation**). By integrating **ControlNet and LoRA fine-tuning**, this project will enhance **editing precision**.
- **Slow Image Generation** – Some workflows involve **inefficient sampling methods**, leading to long processing times. This project will implement **optimized diffusion parameters** to **accelerate image generation** without quality loss.

CHAPTER 3

Proposed Methodology

3.1 System Design

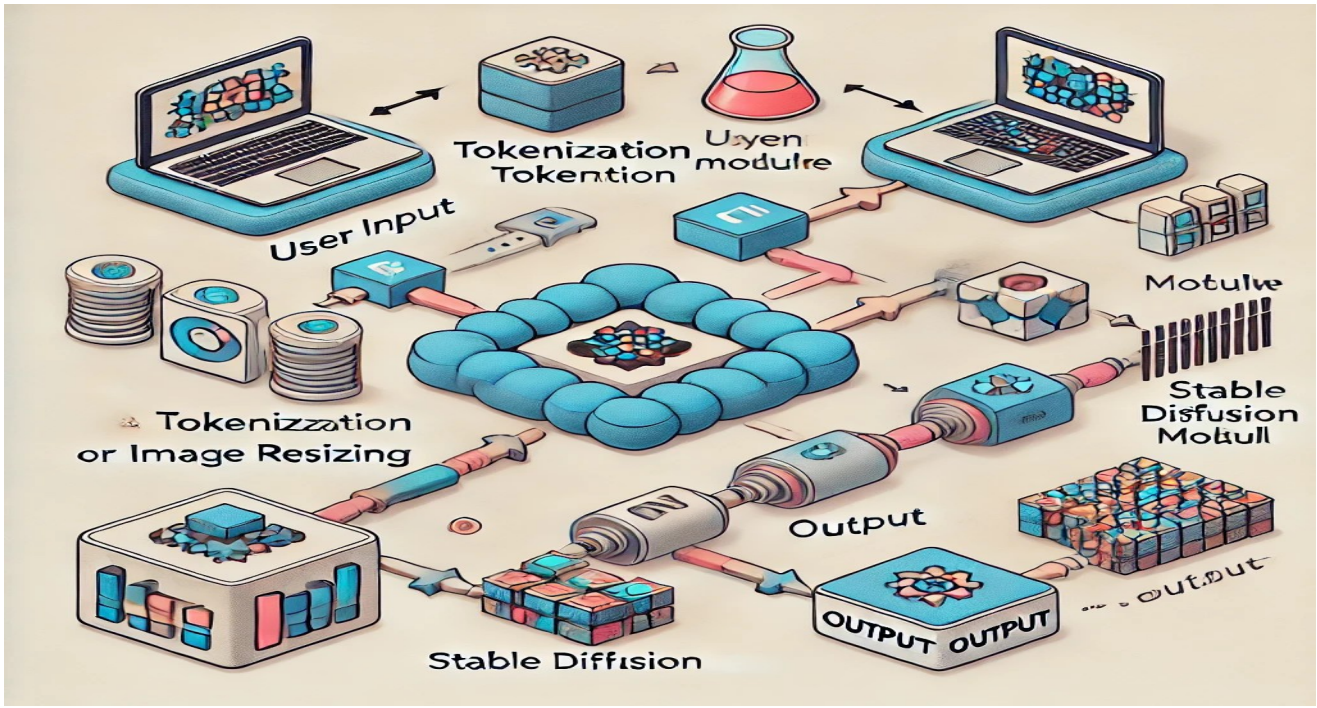

Fig 01:Working of the models

Explanation :

1. **User Input (Text or Image)**
    - The user provides input, either a text prompt or an image for further processing.
2. **Preprocessing Module (Tokenization, Image Resizing)**
    - If the input is text, it is tokenized and converted into an embedding format.
    - If the input is an image, it is resized and normalized for processing.
3. **ComfyUI Workflow (Processing Pipeline)**

- This module acts as a flexible node-based system for managing the Stable Diffusion workflow.
- It enables users to customize the image generation pipeline visually.

4. **Stable Diffusion Model (Text-to-Image or Image-to-Image Generation)**

   - This model generates images based on the processed input.
   - If using Text-to-Image, the model uses CLIP embeddings to guide the diffusion process.
   - If using Image-to-Image, the model refines the input image based on prompt and noise adjustments.

5. **Postprocessing Module (Enhancements, Filtering)**

   - Generated images undergo enhancement techniques like upscaling, noise reduction, and color correction.

6. **Output (Generated Image)**

   - The final image is displayed to the user for review and download.

3.2 Requirement Specification

The tools and technologies required to implement the solution.

3.2.1 Hardware Requirements:

1. High-Performance Server

- A **dedicated AI server or workstation** with high computational power is required to handle large datasets, training, and inference workloads.
- **Recommended:** Cloud-based AI instances (e.g., NVIDIA DGX, AWS EC2 with A100/H100 GPUs) or an enterprise-grade workstation.

2. Processor (CPU)

- A **high-performance, multi-core processor** is crucial for preprocessing, data handling, and workflow execution.
- **Recommended:**
  - Intel **Core i9-13900K** (3.5GHz+, 24 cores)

- AMD **Ryzen 9 7950X** (4.5GHz+, 16 cores)
- AMD **Threadripper PRO 5965WX** (for high-end workstations)

- **Minimum:**
  - Intel Core i7 (10th Gen or later)
  - AMD Ryzen 7 (5000 series or newer)

3. Graphics Processing Unit (GPU)

- A **powerful GPU** is essential for accelerating deep learning tasks such as Stable Diffusion model training and inference.
- **Recommended:**
  - **NVIDIA RTX 4090** (24GB VRAM)
  - **NVIDIA A100** (40GB/80GB VRAM, for enterprise AI workloads)
  - **NVIDIA H100** (for cloud-based AI applications)
- **Minimum:**
  - NVIDIA RTX **3060 (12GB VRAM)**
  - **CUDA support** is required for PyTorch and TensorFlow acceleration.

3.2.2 Software Requirements:

- **AI Models**: Stable Diffusion (v1.5, v2.1, SDXL), ControlNet, LoRA.
- **Core Tools**: ComfyUI, AUTOMATIC1111, InvokeAI.
- **ML Frameworks**: Python, PyTorch, TensorFlow, CUDA/cuDNN.
- **Image Processing**: OpenCV, Pillow, NumPy.
- **Enhancements**: GFPGAN, Real-ESRGAN, ClipSeg.
- **Deployment**: Streamlit, Gradio, FastAPI, Docker.

CHAPTER 4

Implementation and Result

4.1 Snap Shots of Result:

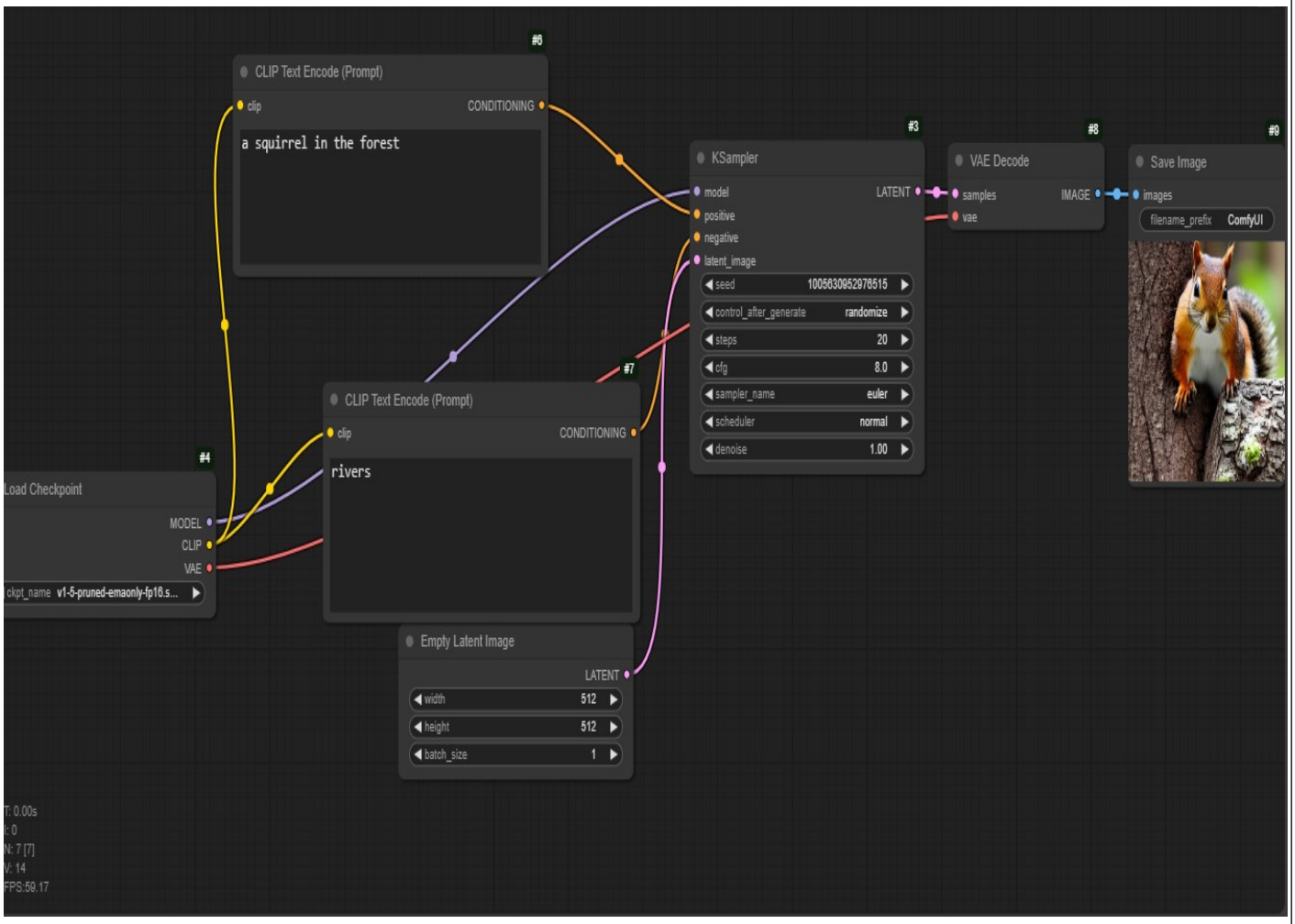

Fig 02: A squirrel in a forest

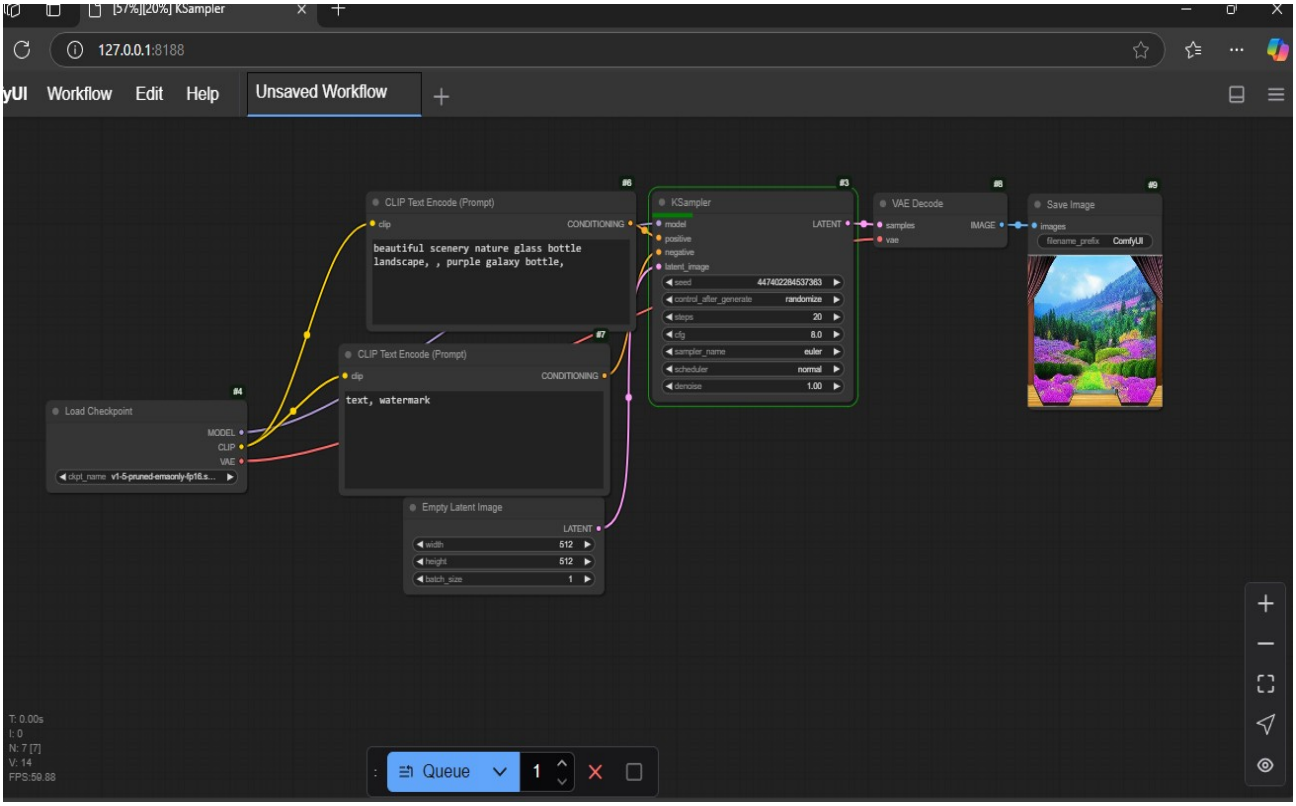

Fig 03: A Purple glass bottle

CHAPTER 5

Discussion and Conclusion

5.1 Future Work:

Future Improvements and Unresolved Issues

1. **Optimize Model Efficiency** – Implement quantization (INT8/FP16), TensorRT, and ONNX Runtime to reduce memory usage and speed up inference.

2. **Enable Multi-GPU and Distributed Processing** – Use DeepSpeed and Fully Sharded Data Parallel (FSDP) to scale training and inference across multiple GPUs.

3. **Enhance Image Quality** – Integrate advanced upscaling techniques like Real-ESRGAN and improve text-to-image alignment with larger CLIP models.

4. **Improve Prompt Adherence** – Fine-tune Stable Diffusion models to better interpret text prompts and generate more accurate images.

5. **Mitigate Bias and Ethical Issues** – Train models on diverse datasets and implement fairness-aware sampling techniques to reduce bias in outputs.

6. **Implement Content Moderation** – Use NSFW classifiers and automated prompt filtering to prevent harmful or inappropriate image generation.

7. **Enhance User Experience** – Develop AI-assisted prompt suggestions, real-time preview rendering, and a more intuitive UI in ComfyUI.

8. **Improve Workflow Customization** – Provide prebuilt templates, drag-and-drop functionalities, and better integration with ControlNet for structured image generation.

9. **Expand to Video Generation** – Research and implement real-time text-to-video synthesis using Stable Diffusion advancements and latent space interpolation.

10. **Develop Lightweight Models for Edge Devices** – Optimize diffusion models for lower VRAM usage, enabling real-time generation on mobile and embedded systems.

5.2 Conclusion:

This project significantly enhances AI-powered image generation by leveraging **Stable Diffusion** and **ComfyUI**, offering an efficient, customizable, and high-quality image synthesis pipeline. The key contributions include:

- **Improved Efficiency** – Optimized model performance through quantization, multi-GPU processing, and advanced sampling techniques.
- **High-Quality Image Generation** – Enhanced image resolution, style consistency, and better text-to-image alignment.
- **Ethical and Responsible AI** – Addressed bias, implemented content moderation, and ensured safer AI-generated visuals.
- **User-Centric Design** – Introduced interactive workflows, AI-assisted prompt refinement, and real-time previews for better usability.
- **Scalability and Integration** – Enabled API-based deployment, web integration, and future expansion into **video generation** and **lightweight edge devices**.

By advancing **efficiency, quality, usability, and ethical AI practices**, this project contributes to the **next generation of AI-driven creative tools**, empowering artists, designers, and researchers to generate stunning visuals with ease.

REFERENCES

1. CompVis, Stability AI (2023). Stable Diffusion Model Card & Documentation. https://github.com/CompVis/stable-diffusion

2. ComfyUI GitHub Repository (2023). ComfyUI: A Powerful, Node-Based Graph Interface for Stable Diffusion. https://github.com/comfyanonymous/ComfyUI

3. Nichol, A. Q., Dhariwal, P., Ramesh, A., Shyam, P., Mishkin, P., McGrew, B., … & Chen, M. (2021). GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models. https://arxiv.org/abs/2112.10741

4. Esser, P., Rombach, R., & Ommer, B. (2021). Taming Transformers for High-Resolution Image Synthesis. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021. https://arxiv.org/abs/2012.09841

5. Dhariwal, P., & Nichol, A. (2021). Diffusion Models Beat GANs on Image Synthesis. Advances in Neural Information Processing Systems (NeurIPS), 2021. https://arxiv.org/abs/2105.05233

6. Baio, G., Bhatt, U., & Kim, B. (2023). Ethical Considerations and Bias Mitigation in AI-Generated Content. AI Ethics Journal, Vol. 6, Issue 2. DOI Pending