AVATAR GENERATOR

R Yashodara[1], Divyashree K[2],
Rachana N[3], Pragna S[4],Prakruthi G[5]

[1] *Assistant Professor,Information Science and Engineering, Don Bosco Institute of Techonology,Karnataka,India*
[2] *Assistant Professor,Information Science and Engineering, Don Bosco Institute of Techonology,Karnataka,India*
[3] *Student,Information Science and Engineering, Don Bosco Institute of Techonology,Karnataka,India*
[4] *Student,Information Science and Engineering, Don Bosco Institute of Techonology,Karnataka,India*
[5] *Student,Information Science and Engineering, Don Bosco Institute of Techonology,Karnataka,India*

ABSTRACT

This report presents the findings and outcomes of the mini project titled "Avatar Generator," conducted as part of the Mini Project under the Department of Information Science & Engineering, Visvesvaraya Technological Technology.

The purpose of the project is to create a system that produces personalized avatars utilizing user- uploaded images and text prompts, employing machine learning models to generate artistic avatars in a designated style. The main aim of this initiative is to demonstrate the role of AI in creative domains, providing an interactive and user-friendly platform for the creation of tailored avatars.

The project is developed using Python, the Streamlit web framework, and the Stable Diffusion Img2Img model, which supports image-to-image transformations with style-driven adaptations. This mini project offers important insights into the application of generative models in image processing and underscores the capabilities of AI tools in the realms of digital art and character design. The project's scope encompasses creating avatars in artistic styles but does not cover other aspects of AI-driven image processing or real-time user interactions.

The project serves as a demonstration of how AI can be utilized to blend technology with creativity, offering a unique tool for both individuals and businesses to create personalized digital representations. The outcome emphasizes the potential for expanding AI applications in the realm of digital identity, gaming, social media, and virtual environments. Moreover, the project's use of a machine learning model like Stable Diffusion showcases the power of deep learning in transforming visual input into artistic outputs, which could pave the way for further innovations in generative art.

Keyword: *Streamlit ,Stable Diffusion, Machine Learning, and Digital Identity.*

1. INTRODUCTION

This project implements an **Avatar Generator** that transforms uploaded images into stylized avatars based on a user-defined text prompt. Using the **Streamlit** framework, this web application allows users to upload an image, input a descriptive prompt, and generate a personalized avatar in a **Ghibli animation style**. The transformation is powered by the **Stable Diffusion** model, which is a state-of- the-art deep learning technique used to generate high- quality images from text and image inputs.

Key technologies and libraries employed in this project include:

- **Streamlit**: A Python library used to create interactive web applications. It serves as the front- end interface for the project, enabling image uploads, text input, and customization of avatar generation parameters (inference steps and guidance scale).
- **NumPy**: While not explicitly used in the core functionality, NumPy is a standard library for numerical operations and is useful in handling any array-based data processing that may arise in future

developments.

- **PIL (Python Imaging Library)**: PIL, or its modern version **Pillow**, is used for image processing. It helps resize uploaded images to the required dimensions (512x512) before passing them to the model for transformation.
- **PyTorch**: A deep learning framework essential for tensor computation. It powers the Stable Diffusion model and handles the underlying operations for image generation and processing.
- **Diffusers**: A library from Hugging Face that provides tools to run diffusion models. In this project, the **StableDiffusionImg2ImgPipeline** is used to adapt the image-to-image process, transforming the uploaded image into a stylized avatar based on the text prompt.
- **Hugging Face API**: This API is used to access the pre-trained **Ghibli-Diffusion** model, which generates the stylized avatars in the desired artistic style.

1.1 Aim and Objectives of the Proposed

Work Aim:

The aim is to provide an accessible tool that is user friendly and enables anyone to generatepersonalized avatars without requiring advanced artistic abilities or software.The system will leverage the Stable Diffusion Img2Img model to transform user images into unique avatars in various artistic styles, offering customizable options such as inference steps and guidance scale for greater control over the output.

Objectives:

The objective is to create an interactive and user-friendly platform for generating custom avatars with high levels of personalization.

- Technical Objectives: Implement a robust Stable Diffusion-based image generation pipeline capable of transforming user-uploaded images into stylized avatars.
- Performance Objectives: Achieve high-quality image generation with minimal computational overhead.
- Innovation Objectives: Demonstrate AI's potential in creative digital identity generation.
- Accessibility Objectives: Support multiple image formats and styles.

1.2 Proposed System

The proposed system for the "Avatar Generator" project is designed to enable users to create personalized avatars by uploading an image and providing a text-based prompt. The system utilizes machine learning techniques, specifically the Stable Diffusion Img2Img pipeline, to generate avatars in various artistic styles, including cartoonish or anime-inspired designs. The key components are:

- User Interface: Built using Streamlit, the user interface allows users to interact with the system in an intuitive manner. Users can upload an image, enter a text prompt describing the desired avatar style, and adjust key parameters such as inference steps and guidance scale to control the quality and style of the generated avatar.
- Image Transformation: The core of the system relies on the Stable Diffusion Img2Img pipeline, a deep learning model capable of transforming uploaded images into avatars. The model is trained on various artistic styles, such asthe Ghibli style, and can generate diverse and visually appealing outputs based on the input prompt.

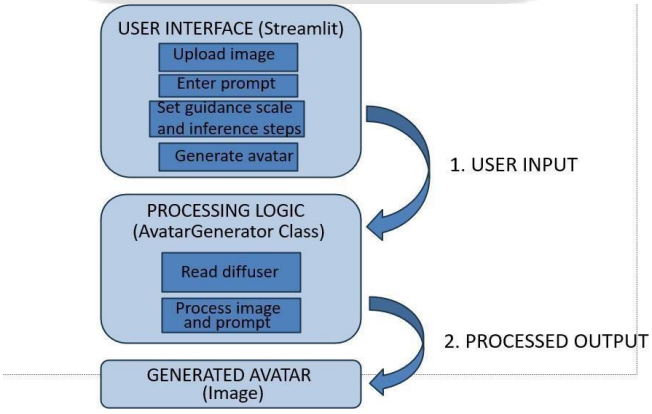

Fig -1: Block Diagram

- Customization Controls: Users have the ability to adjust important parameters:
  - Inference Steps: Controls the number of iterations the model performs to generate the image, influencing the detail and clarity of the output.
  - Guidance Scale: Allows users to control the adherence of the output to the text prompt, affecting how strongly the model follows the provided instructions.
- Avatar Generation Process: Once the user uploads an image and provides a text prompt, the system preprocesses the image as shown in the Figure 1.4.1 (resizing and converting it to the appropriate format) and feeds it into the Stable Diffusion model. The model then generates an avatar based on the input image and text, which is displayed to the user in real-time. The generated avatar can be downloaded or saved locally by the user.

2. LITRATURE SURVEY

The literature survey for this project focuses on the use of Stable Diffusion for image generation, specifically leveraging techniques like Img2Img to create custom avatars. This involves several interrelated fields of research, including Generative Adversarial Networks **(**GANs**)**, Diffusion Models, Text-to-Image Synthesis, and Applications of AI in Creative Industries. Below, we explore the key areas relevant to the project and summarize the literature surrounding them.

1. Generative Adversarial Networks (GANs) and Image Synthesis:

Generative Adversarial Networks (GANs) have been a cornerstone of generative image synthesis. GANs, introduced by Ian Goodfellow et al. in 2014, consist of two neural networks a generator and a discriminator that work together to produce high-quality images from random noise or latent vectors. GANs are capable of generating highly realistic images, such as faces, animals, and landscapes. In the context of avatar generation, GANs have been used to create stylized portraits or specific artistic renderings based on user input or pre-defined styles.[1]

2. Diffusion Models in Image Generation:

Stable Diffusion utilizes a reverse diffusion process where noise is gradually removed from an image conditioned on a given prompt or starting image. This makes it suitable for applications like Img2Img, where an initial image can be transformed according to a textual description or artistic style. The model has also been widely used in text-to-image synthesis, allowing for detailed control over the generated content, which is a key feature for the avatar generation task in this project.[2]

3. Text-to-Image Synthesis:

Text-to-image synthesis is an active area of research focused on generating images from textual descriptions. Modern models like DALL·E and Stable Diffusion have significantly advanced the field by demonstrating the ability to create photorealistic or stylized images based on complex textual prompts. These models leverage large-scale training on vast datasets and apply transformer architectures to learn the mapping between text and image spaces.In the context of avatar generation, text-to-image synthesis allows users to specify a character's attributes or stylistic preferences, such as in the project's case of generating avatars with a "Ghibli style" based on a text prompt. This approach provides significant creative control, as users can fine-tune the visual outcome through textual descriptions.[3]

4. Stable Diffusion in Img2Img Applications:

The Img2Img capability of Stable Diffusion allows for image transformation, making it particularly useful for generating avatars by applying different artistic styles to user-provided input images. In Img2Img tasks, an initial image (e.g., a user-uploaded photo) is manipulated according to a given text prompt, which could describe the desired output style (e.g., "anime- style portrait"). This technique allows for maintaining the structure and identity of the original image while transforming its visual appearance based on user specifications.[4]

5. Streamlit and Its Role in AI Application Development:

Streamlit is a popular framework used for building interactive web applications with Python, particularly for AI and machine learning models. It provides an easy way to create front-end user interfaces that interact with complex models, making it ideal for creating applications like the avatar generator in this project. Streamlit's simplicity and powerful integration with Python libraries (like PIL, PyTorch, and Diffusers) allow developers to quickly prototype and deploy machine learning applications with a focus on user interaction.[5]

2.1 Methadology

The methodology for the Avatar Generator project is based on integrating machine learning and web development technologies to create an interactive system that transforms user-uploaded images into personalized avatars using a text prompt. The following steps outline the approach taken to develop the system:

1. Problem Analysis and Requirement Gathering:

- Identifying User Needs: The system is designed to cater to users who want to create personalized avatars easily. This requires understanding user preferences in terms of image transformation and customization features.
- Defining System Requirements: Based on user needs, the software and hardware requirements were identified. This includes selecting technologies like Streamlit for the frontend, and integrating a machine learning model (Stable Diffusion Img2Img) for image generation.

2. System Design and Architecture:

- User Interface Design: The interface was designed using Streamlit, which allows for rapid development of interactive web applications. The user interface enables easy upload of images and text prompt input.
- Model Selection: The Stable Diffusion Img2Img pipeline was chosen for image transformation. This deep learning model is capable of generating high-quality images based on a given input image and a text prompt, making it ideal for generating personalized avatars.
- Customization Parameters: Options like inference steps and guidance scale were incorporated to allow users to customize the style and quality of the generated avatars.

3. Data Preprocessing:

- Image Preprocessing: Before passing the uploaded image to the model, the image is resized to 512x512 pixels tomatch the model's input requirements. The image is also converted to RGB format to ensure compatibility with the Stable Diffusion model.
- Text Prompt Processing: The user-provided text prompt is cleaned and formatted to ensure the model understands the desired style for the avatar generation.

4. Model Integration and Training:

- Stable Diffusion Img2Img Model: The project utilizes a pre-trained model from Hugging Face's model hub. The model is used in an image-to-image generation setup where it takes the uploaded image and text prompt, processes the image, and produces the output in the desired artistic style.
- Fine-Tuning Parameters: The inference steps control the number of iterations during image generation, impacting thequalityand detail of the avatar. Guidance scale isused to ensure the model follows the text prompt more closely, influencing the output style.

5. System Implementation:

- Frontend Development: Using Streamlit, the web interface was developed to handle
- image uploads and text prompt inputs. Users can upload an image file, enter a descriptive prompt, and adjust parameters like inference steps and guidance scale using sliders.
- Backend Development: The backend processes the image and prompt by running the Stable Diffusion Img2Imgmodel. After processing, thegeneratedavatar isreturned to the user's interface.

6. Avatar Generation and Output:

- Image Generation: Once the user inputs an image and a text prompt, the system processes these inputs and generates the avatar by passing them through the Stable Diffusion model.
- Customization: The system offers sliders for adjusting the inference steps (which controls how detailed the avatar is) and the guidance scale (which defines how closely the output matches the text prompt).

7. Testing and Evaluation:

- Unit Testing: Individual components, such asthe image preprocessing and model output, were tested to ensure correctness.
- System Testing: The full system was tested by simulating real-world use cases, such as varying image

inputs, text prompts, and parameter adjustments.

- Performance Evaluation: The system's response time and the quality of generated avatars were evaluated to ensure it meets user expectations for both speed and output quality.

8. Deployment and User Feedback:

- Deployment: The system was deployed locally for initial testing, and feedback was gathered fromusers to refine the interface and improve the avatar generation process.
- User Feedback: User feedback was collected to assess the ease of use, quality of generated avatars, and customization options, which were incorporated into further system iterations.

2.2 System Desgin

This flowchart as shown in the Figure 2 describes the steps involved in generating avatars using a Streamlit application. The process starts with the user uploading an image and adjusting the guidance scale. The application then checks the image for inappropriate content and prompts the user if necessary. If the image is deemed appropriate, the user can submit a button to proceed. The application then sets a seed and generates text prompts. An avatar is then generated using a chosen method and saved as an image. The user can also generate avatars using a pipeline. Finally, the generated image is opened.

Fig -2:System Design

3. RESULT

Figure 3 demonstrates the user interface page that opens after execution for the code “python -m streamlit run” in the visual studio code terminal.

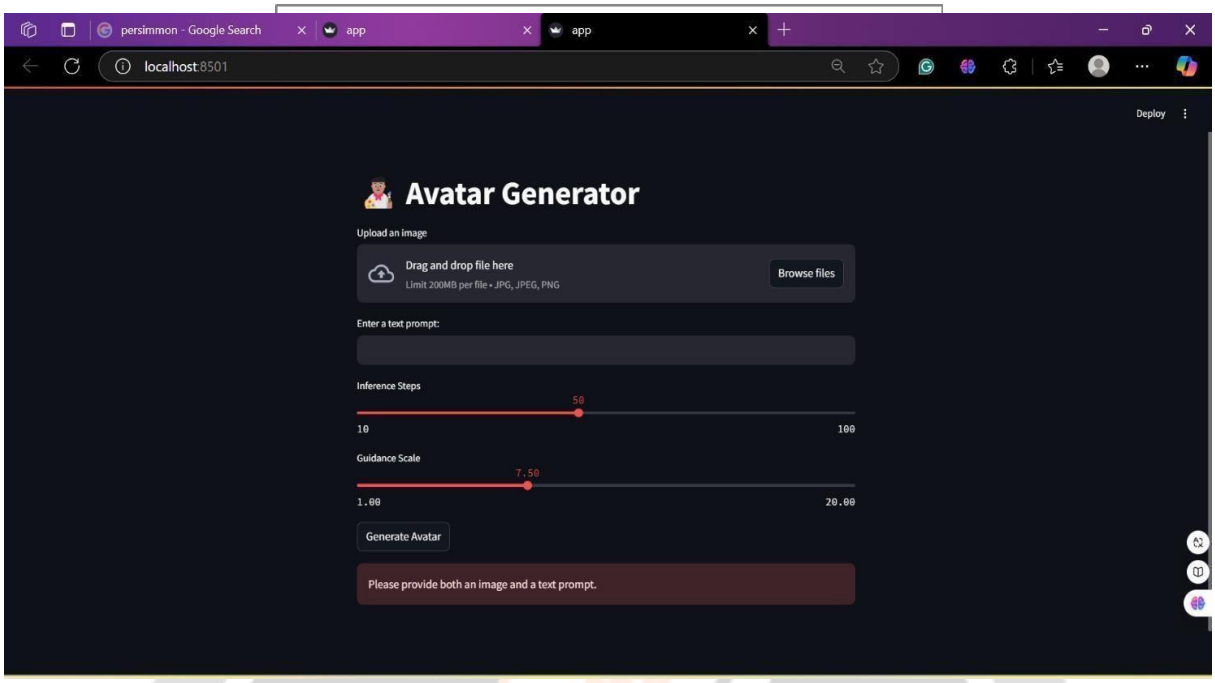

Fig -3: Start page

3.1 Input and output

The image shown in Figure 4 will be the user uploaded image and along with the text prompt stating "two boys in the library".

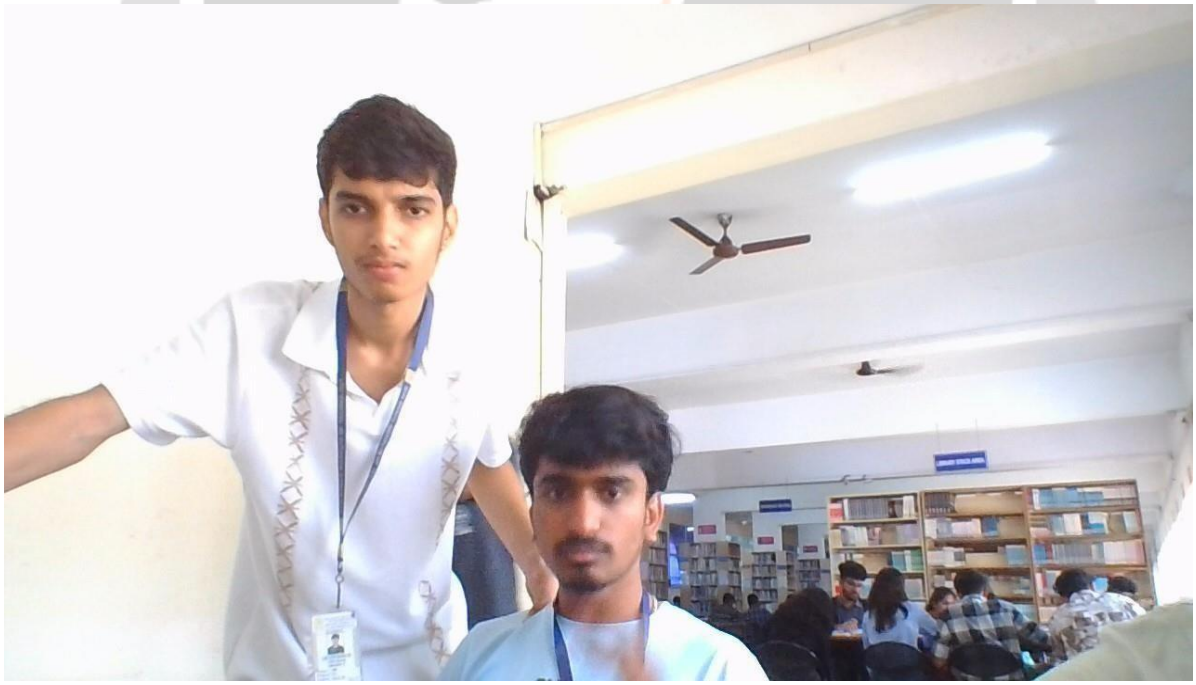

Fig -4: Original uploaded image

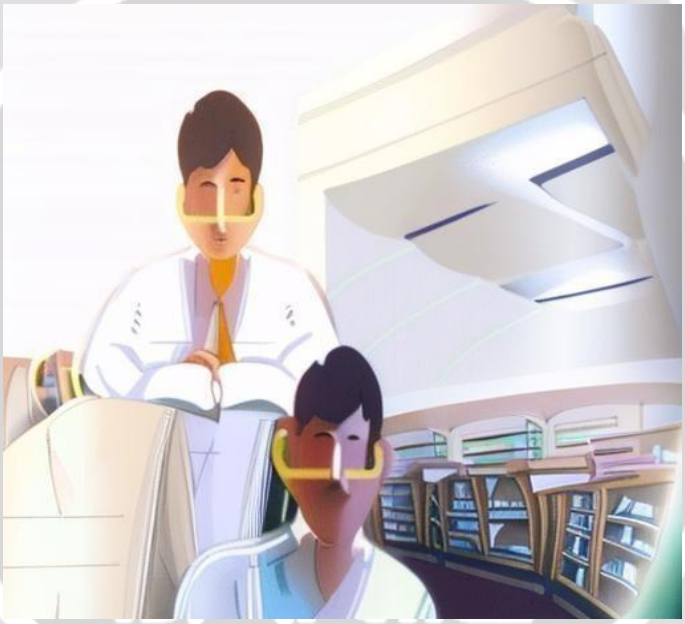

Fig -5: "nitrosocke/Ghibli-Diffusion" style avatar

Fig -6: "stabilityai/stable-diffusion-2-1" style avatar

The Figure 5 represents the Ghibli Diffusion style in animation output. Ghibli-inspired diffusion models could generate artwork or imagery resembling the aesthetic of films like Spirited Away, My Neighbor Totoro, or Princess Mononoke.The Figure 6 is one of the outputs of generated avatar in the “Stabilityai Diffusion” style of animation. It uses latent diffusion to generate highly detailed and visually stunning images based on text prompts.

4. CONCLUSIONS

The Avatar Generator project successfully demonstrates the potential of artificial intelligence in creating personalized and creative avatars based on user inputs. By leveraging the Stable Diffusion Img2Img model and the power of Streamlit for an intuitive user interface, the system allows users to upload images and provide text prompts to generate unique avatars in various artistic styles.

Throughout the development and testing of the system, it became evident that the integration of AI- based image generation technology can greatly simplify and enhance the avatar creation process. This tool not only enables personalized representation but also empowers users to express their creativity without requiring advanced graphic

design skills.

In conclusion, the Avatar Generator is a powerful tool with practical applications in gaming, social media, marketing, and digital art. With future enhancements, this system has the potential to revolutionize the way users create and interact with digital representations of themselves, paving the way for more immersive and personalized digital experiences.

5. ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crowned the efforts with success.We would like to profoundly thank Management of Don Bosco Institute of Technology for providing such a healthy environment for the successful completion of Project work. We would like to express our deepest sense of gratitude to our guide Mrs. R Yashodara, Assistant professor &Mrs. Divyashree K, Department of Information Science & Engineering for her constant support and guidance throughout the Project work.Also, we would like to thank all teaching and non-teaching staff of Department of Information Science & Engineering who have helped directly and indirectly throughout the project work.

6. REFERENCES

[1]. Streamlit Docs. "Streamlit: The Fastest Way to Build and Share Data Apps," Streamlit Inc. (2023)

[2]Rombach, R., et al. (2022). "High-Resolution Image Synthesis with Latent Diffusion Models," CVPR.

[3]Ramesh, A., et al. (2021). "Zero-Shot Text-to-Image Generation," *ICML*.

[4] Ho, J., et al. (2020). "Denoising Diffusion Probabilistic Models," NeurIPS.

[5] Goodfellow, I., et al. (2014). "Generative adversarial nets," Advances in Neural Information Processing System